Data-Driven Impact Analysis of DNSSEC Outages

Carlos Daing
Data Analytics Engineering
cdaing@gmu.edu
George Mason University
Fairfax, VA USA

Durga Sahithi Dunaboyina
Data Analytics Engineering
ddunaboy@gmu.edu
George Mason University
Fairfax, VA USA

Leo Franzinetti
Data Analytics Engineering
lfranzin@gmu.edu
George Mason University
Fairfax, VA USA

Hadeel Idris
Data Analytics Engineering
hidris2@gmu.edu
George Mason University
Fairfax, VA USA

Beenil Jain
Data Analytics Engineering
bjain2@gmu.edu
George Mason University
Fairfax, VA USA

Tanya Pati
Data Analytics Engineering
tpati2@gmu.edu
George Mason University
Fairfax, VA USA

Abstract- The Domain Name System (DNS) is the system used by the Internet most commonly used to convert domain names such as " www.example.com " into an IP address for your computer to connect to and to communicate with each other. This system is further divided into various Zones, each with designated authorities. These authorities manage the records within their respective zones and can vary in levels, ranging from entities such as colleges to entire countries. To ensure that a person is sent to the correct address and not a malicious website, a security protocol can be used for further protection named Domain Name System Security Extension, or DNSSEC. In this study, we plan on using a dataset sourced from SecSpider, a DNSSEC monitoring system that collects data on all available DNSSEC Zones. With this dataset, we plan on developing an engine aimed at identifying, categorizing, and subclassifying errors experienced by name servers and within DNSSEC Zones. Our proposed system plans on using macro-scale analysis of name server errors from multiple DNSSEC Zones to discern patterns indicative of DNSSEC Zone occurrences of complete, partial, or no outages. Through the observation of recurrent instances within defined time frames, we can distinguish between different outage types. Furthermore, we plan on deploying a classifier to intake new data points within Zones, to improve future response times for outage resolutions and provide insights into underlying factors contributing to the outages. With computing power drastically improving each year, for good and malevolent intentions, having a more robust outage resolution system and a stronger foundation for DNSSEC is becoming imperative. Moreover, with stronger security, more redundant safety protocols and systems can be set aside to help foster and improve Internet performance.

Keywords – DNS, DNSSEC, SecSpider, DNS Outages

I. Introduction

The DNS (Domain Name System) plays a crucial role in the functionality of nearly every IP network application, including web browsing, email, multimedia applications, and more. Any attack that interrupts the DNS service or manipulates the integrity of DNS data can effectively make an application or network unacceptable. Therefore, the protection of DNS data and communications throughout the resolution process is of utmost importance. Various security vulnerabilities are within DNS and ongoing monitoring of the operating system and associated software vulnerabilities is fundamental in this realm. These security risks include Packet Interception or Spoofing, ID Guessing or Query Prediction, Name Chaining or Cache Poisoning, Zone Transfers, and Denial of Service. [1]

The DNS serves as the address book for the Internet, allowing computers to locate and communicate with each other by converting human-readable domain names into machine-readable IP addresses. But because it is open, it may also take any address sent to it without any kind of authentication. DNSSEC, a protocol that strengthens DNS security by adding an additional layer of trust through authentication procedures, is the solution. DNSSEC adds cryptographic signatures to traditional DNS records, strengthening the system's trustworthiness. Alongside common record types like A, AAAA, MX, and CNAME, DNS name servers also contain these distinctive digital signatures. DNSSEC ensures that requested DNS records are legitimate and come from the authoritative name server by verifying its accompanying signature. This helps prevent unwanted changes to DNS records while they are in transit. [2]

DNSSEC (Domain Name System Security Extensions) serves as a tool for confirming the legitimacy of original data in DNS resolutions and validating the integrity of such information. It also provides a means to verify the nonexistence of DNS data, allowing for the signing of outcomes like "not found" resolutions. This functionality

empowers the detection of potential threats such as packet interception, ID guessing, and cache poisoning attacks on both successful and unsuccessful resolutions. Utilizing asymmetric public key cryptography technology, DNSSEC ensures data origin authentication and end-to-end integrity verification. [1]

DNSSEC has become an integral part of today's internet security, however operational experience has shown that its complexity has led to negative implications in the DNS namespace. Like any technology, it is vital to consider possible problems and risks when using it and to make sure it is properly configured and managed. Mitigating these concerns requires deploying effective tools designed to recognize, diagnose, and help correct such errors. [3]

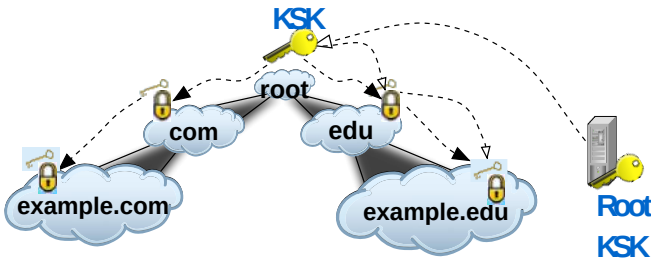

Figure 1: DNS Visualization

II. Why Internet Security Matters

The internet has become an integral part of daily life for billions of people worldwide, facilitating communication, information access, and social interaction. Sharing personal information online is a major component of internet use. Most of the time, people publish different kinds of personal information online without fully realizing the possible repercussions.

Users are required to submit personal information, including names, addresses, phone numbers, email addresses, and birthdates, on social networking platforms, e-commerce websites, and other online services. People also voluntarily provide other personal information about themselves, such as images, thoughts, hobbies, and activities. Companies frequently utilize this abundance of personal data for user analytics, content personalization, and targeted advertising. It also serves as the basis for digital identities. Although social connections and convenience are increased when personal information is shared online, there are concerns about security and privacy. In the digital era, concerns about data breaches, hacking events, and illegal access to personal information are commonplace. Cybercriminals take advantage of weaknesses in online systems to steal confidential information, which can result in financial fraud, identity theft, and other types of cybercrime.

While older folks could be more cautious, younger generations—who have grown up in a digital age—may feel more at ease revealing personal information online. The internet has completely changed the way that businesses and customers interact with one another globally. Online payment methods, digital currencies, and e-commerce platforms have completely changed how products and services are purchased and sold, spurring previously unheard-of development in the online retail sector.

Prominent online retailers like Amazon, Alibaba, and eBay control the market by providing a wide range of goods and services to customers all over the world. Secure online transactions are made possible by payment processors like Square, PayPal, and Stripe, which guarantee that private financial data is shielded from unwanted access. An increasing number of people are adopting smartphones and tablets for online shopping, which has led to the emergence of mobile commerce, or m-commerce, as a major force in the digital economy. The checkout process has been improved by mobile payment applications, digital wallets, and contactless payment technologies, which have expedited the transition to digital payments. Online purchases are efficient and convenient, but security flaws, fraud, and data breaches continue to be worries. Cybercriminals use a variety of strategies, such as virus assaults, phishing schemes, and credit card theft, to take advantage of holes in online payment systems and steal confidential financial data. Thousands of people suffer from emotional misery, financial losses, and impaired credit because of identity theft every day. [4]

Cybercriminals use a variety of strategies, such as phishing emails, social engineering techniques, malware assaults, and data breaches, to get personal information and assume the identities of their victims. Once they get access to private information like credit card numbers, social security numbers, and account passwords, identity thieves may wreak havoc on the lives of their victims by creating false accounts, making transactions without authorization, and engaging in other financial crimes.

By employing strong, one-of-a-kind passwords, turning on two-factor authentication, and steering clear of dubious websites and communications, people may safeguard themselves.

In addition, legislative frameworks like the California Consumer Privacy Act (CCPA) in the US and the General Data Protection Regulation (GDPR) in the EU mandate that enterprises protect personal data and inform consumers in the case of a data breach. Adherence to these standards reduces the likelihood of data breaches and raises customer assurance over the security of personal data.

III. Outages

A. Different Types of Outages

1) Expired Signatures

DNSSEC functions by appending cryptographic signatures to DNS records, which are kept in DNS name servers alongside standard record types like MX and AAAA. Resolvers can confirm the legitimacy of DNS answers and make sure the records haven't been altered in transit thanks to these signatures.

DNSSEC adds several record types to make signature validation easier:

- A record set's cryptographic signature is contained in the Resource Record Signature (RRSIG).
- Public signing key included in DNSKEY.
- DNSKEY record hash is contained in the DS field.
- NSEC and NSEC3: List all possible record kinds and provide a link to the next record in the zone.

DNSSEC improves security by grouping DNS entries and isolating zones, shielding domains from cache poisoning

and forgeries. There are trade-offs between security and operation when using it in several modes, such as centralized online, offline, and on-the-fly signing. DNSSEC's dependency on UDP makes it more susceptible to DDoS attacks and UDP spoofing, despite its advantages. Root-signing rituals protect the integrity of the infrastructure. Overall, DNSSEC improves DNS security, but risk mitigation requires careful implementation. [5]

2) Natural Disasters

Significant risks to internet infrastructure are posed by disasters such as hurricanes, earthquakes, fires, and strong storms. These events frequently result in extensive physical damage and service disruptions. Severe weather may bring down electricity lines and interfere with community connectivity. [6]

B. Exploring Real World Outages

October 12, 2009, was the date of the ".se" DNSSEC outage. However, as this outage occurred before IANIX.com started documenting DNSSEC outages, there is not a lot of information available about it on the site.

The whole internet of Sweden was down for about an hour in October 2009 due to a major outage that occurred during scheduled maintenance for the countries .se domain. The Internet Infrastructure Foundation, which oversees the .se domain, published an erroneous zone file, causing the outage. The cause of this error was a malfunctioning software upgrade that was missed during testing. In response, the foundation quickly released a replacement file that lacked the necessary DNSSEC signatures, momentarily impairing the availability of .se domains. Technically, the downtime lasted only one hour, but because ISPs had to clear their DNS caches, users had to wait longer than expected. To ensure that such situations never arise again, the foundation initiated an internal investigation. Given how rarely complete top-level domains perceive such extensive outages, this incident highlighted the possible influence of DNS problems on internet accessibility. [7]

C. DNSSEC Outage Policy and Standards Overview

ICANN, the Internet Corporation for Assigned Names and Numbers, is a non-profit organization with a global and multistakeholder approach. Headquartered in the United States, ICANN is responsible for coordinating and overseeing the maintenance and procedures of various databases related to the Internet's namespaces and numerical spaces. This encompasses tasks such as managing the Domain Name System (DNS) and allocating IP addresses. This entity aims at ensuring the Internet's stable and secure operation, in terms of DNSSEC, ICANN provides organizations with certain regulations to follow and adhere to for maintaining this realm. [8]

There is no standardized global policy specifically addressing DNSSEC outage incidents. Organizations, domain registrars, and internet service providers usually have incident response and outage policies, which may include procedures for dealing with DNSSEC-related issues. These response activities include incident response plans, communication protocols, coordination with DNS Operators, as well as continued monitoring and detection. They often follow best practices and standards related to DNSSEC, such as those outlined in RFCs (Request for Comments), guidelines provided by internet governance organizations and other internet security platforms. Compliance with these standards can contribute to a more robust DNSSEC operation. [9]

D. Failures/Glitches versus Outages in DNSSEC

It is important to distinguish a failure in the infrastructure versus a complete outage when combating DNSSEC operations. Accurate predictions of each respective type will aid in the proper mitigation of issues when or even before they occur. A DNSSEC failure occurs when there is a problem with the DNSSEC validation process, and the digital signatures associated with DNS data cannot be successfully verified. This can lead to an inability to establish the authenticity and integrity of DNS data, potentially exposing users to security risks like DNS spoofing or cache poisoning. It is usually a result of expired or missing signatures, key rollover issues, misconfiguration of DNSSEC-related records, or compromise of cryptographic keys. A DNSSEC outage on the other hand is a period during which the entire DNSSEC service or infrastructure is unavailable or not functioning correctly, operations are disrupted, and digital signatures cannot be effectively processed, leaving the DNS vulnerable to potential security threats. [10]

IV. SecSpider

SecSpider is a project aimed at monitoring the deployment of DNS Security Extensions (DNSSEC) globally. Its purpose is to provide insights into the size, scope, and trends of the global rollout of DNSSEC and to operate as a distributed key lookup service. The SecSpider database is proprietary to the Measurable Security Lab at George Mason University. The database contains over 20 tables and over 54 billion records. Some of the tables explored in this project include SS_Zone, SS_RRSet, SS_RRSig, SS_Nameserver, SS_Nameserver_Stats, and others. [11]

V. Previous Work

A former student's work serves as a reference and inspiration for our project. They investigated DNSSEC outages focusing specifically on those caused by expired RRSig signatures. The project examined data from the SecSpider database which was minded for instances of signature expiration outages in Top Level Domains (TLDs) from 2013 to 2017. The project cross-verified the findings from the IANIX website. Results reveal that more signature expiration outages were detected than reported by IANIX, highlighting the importance of comprehensive data analysis for understanding and addressing DNSSEC vulnerabilities. Additionally, the study uncovered instances where partial outages affect only a subset of name servers within zones, suggesting the potential for service continuity despite DNSSEC failures. These findings underscore the ongoing challenges and opportunities for enhancing DNSSEC resilience and reliability in the face of evolving security threats. A big limitation of the paper is that it focuses on a small portion of the data. Our project expands on this work. [12]

VI. Exploratory Analysis

The objective of outage exploration is to detect patterns correlated with outage occurrences holistically and make data-driven recommendations that can provide stakeholders with the means to improve management practices and protocols.

Outages can be attributed to internal and external factors. To begin mitigating these vulnerabilities, it is important to start assessing factors within the infrastructure that could be associated with these incidents. Three variables of interest were used for exploration:

Average TTL (Time to Live): refers to the amount of time that a DNS record is cached by DNS resolvers and other systems before it expires and needs to be refreshed. It indicates the update frequency of DNS records; the objective is to see whether longer vs shorter TTLs affect zone outages.

Average Signature Validity: represents the duration of valid DNSSEC signatures. These are usually set by the zone administrator; the objective is to understand how the duration variations affect zone outages.

DNSSEC: adds cryptographic security to DNS responses and might affect zone outage behavior due to additional processing requirements or potential misconfigurations, the objective is to see whether zones with DNSSEC vs without influence outages.

The methodology involved accessing the data, determining tables of interest, and preparing an appropriate subset. The data was then uploaded to AWS to prepare for further investigation, analysis, and visualization on Tableau.

Determining what constitutes the outage variable was computed as follows:

LAST_SEEN (from RRSET Table) – EXP (from RRSIG Table)

(LAST_SEEN - EXP) > 0 = 'Outage'

A subset was produced to conduct exploration based on average signature validity in days, average TTL, and DNSSEC status, to understand their effects/correlations with zone status – a column representing outage occurrences per zone in three categories: no outage, partial outage, complete outage.

It consisted of creating nine new columns derived from the existing data: *see appendix A3*.

A variable importance plot was also created to assess the importance of the features (variables) on model accuracy, which can be used for future prediction models. The plot suggested that Signature Validity Duration & Time to Live are highly important variables.

Multinomial regression was used to provide insights into the relationships between the categorical dependent variables (zone status) and independent variables (avg_sig_val, avg_ttl, and DNSSEC status), to predict the categorical outcomes and understand how these factors influence zone status.

Interpretation:

(Average Time-to-Live)

Zone status 1: The coefficient for avg_ttl is approximately -6.85e-06 – meaning that for every unit increase in avg_ttl, the log odds of the DNS zone being a complete outage decreases by approximately 6.85e-06 units.

Zone status 2: The coefficient for avg_ttl is approximately 3.549370e-06 – meaning that for every unit increase in avg_ttl, the log odds of the DNS zone being a partial outage increases by approximately 3.549370e-06 units.

(Average Signature Validity Duration):

Zone status 1: The coefficient for average_sigval_in_days is approximately -0.00478 - meaning that for every unit increase in average_sigval_in_days, the likelihood of the DNS zone being a complete outage decreases by 0.00478.

Zone status 2: The coefficient for average_sigval_in_days is approximately -0.005292581 - meaning that for every unit increase in average_sigval_in_days, the likelihood of the DNS zone being a partial outage decreases by 0.005292581.

(DNSSEC Status):

DNSSEC Status is a categorical variable, so we have coefficients for each category compared to a reference category.

Both coefficients are positive indicating that the presence of DNSSEC is associated with a higher likelihood of outage classification compared to the absence of DNSSEC.

A bar plot showcasing the impact of DNSSEC status on zone status reflected that zones with DNSSEC have significantly more outages compared to zones without.

Upon visualizing the data points on Tableau, the output revealed that although a significant variable, there was no distinct duration when it came to signature validity to determine an optimal benchmark that can be recommended to stakeholders or zone administrators. Additionally, based on the output that compared TTL against different zone types, there was no significant effect of a specific TLL value correlating with the various outage types.

VII. Expired Signature Outages

When first starting the project, we wanted to explore the easiest form of outage we could pull from SecSpider, expired signatures. Due to the simple nature of comparing the "EXP" field to the "LAST_SEEN" and if the latter was larger than the former, then we could make a simple if statement to create a dependent variable regarding whether an outage occurred or not. Then using tools in R, we created a predictive logistic regression model. By also studying the variables through correlation analysis and by making modifications to the predictive models, we were able to understand how these attributes were affecting the predictive capability of when outages were occurring. After careful exploration, we created this model:

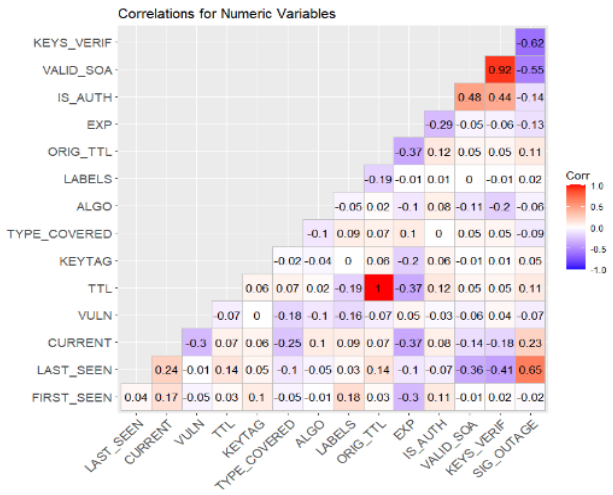

Figure 2: Signature outage correlation plot

$$f(x) = -13.2 + 9.319e^{-9}(x_1) - 1.3076e^{-10}(x_2) + 0.0407(x_3) - 0.2278(x_4)$$

$$X_{outages} = \begin{cases} 1, & f(x) \geq 0.5 \\ 0, & f(x) < 0.5 \end{cases} \quad (1)$$

***see appendix A2 for the description of variables**

From these models, we were able to see that the variables for "VALID_SOA", "KEYS_VERIF", and "IS_AUTH" from the SS_NAMESERVER_STATS table are quite impactful when making predictions for these types of outages and took them into closer consideration when looking at the other outages. As for the accuracy of the model, it had an accuracy of 99% and a specificity of around 70%, the purpose of the model was never to make predictions, as the SecSpider system itself can detect them if needed; rather, it the goal was just to learn more about the variables.

VIII. Natural Disasters

When dealing with outages caused by natural disasters, we started researching reports, or articles, that detailed when and where there were network outages. Then by running the SQL statements used to pull the dataset for the signatures, we filtered for the dates based on the "SEEN" variable in the SS_NAMESERVER_STATS table as well as the nameserver names that contained the Top-Level Domain for that area. Then using other external datasets, we matched the IP addresses to ensure that they were registered in the correct country. If they had matched and the attribute "IS_ONLINE" was marked as 0, we would have considered that server to be out. Using this dependent variable (named "FINAL_NOT_ONLINE") as well as studying the variables that had the strongest relationships with the dependent variable, we created a logistic regression model for instances of recorded natural disasters.

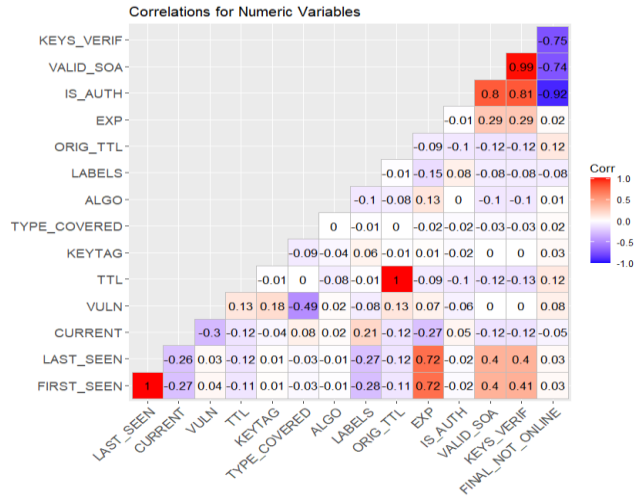

Figure 3: Natural disaster outage correlation plot

$$f(x) = 0.861889 - 0.853448(x_1) + 0.10535(x_2) - 0.108162(x_3)$$

$$X_{outages} = \begin{cases} 1, & f(x) \geq 0.5 \\ 0, & f(x) < 0.5 \end{cases} \quad (2)$$

***see appendix A2 for the description of variables**

From these models, we still see that the variables that had a high degree of importance to the signature model, "VALID_SOA", "KEYS_VERIF", and "IS_AUTH" were not only still extremely important to the predictive capabilities of this model but were the only necessary variables.

	Act: No Outage	Act: Outage
Pred: No Outage	84511	0
Pred: Outage	694	4278

Figure 4: Natural disaster outage results

Based on the results of our model, in instances of natural disasters, this logistic regression model was particularly effective at correctly predicting instances when an outage did occur, which in our opinion is a lot more important for the security of the internet compared to predicting accurately when outages did not occur.

IX. Outages as a Whole

Following the exploration of both the natural disasters and the expired signatures we decided to attempt a logistic regression model that explores both the two as one single outage. In order words, by merging the natural disaster dataset and the exploratory dataset used in the expired signatures, we then create a new dependent variable for "outages as a whole" where if at least one instance of either is present in the record, then the server is out. To also not include the timestamps as variables, we replaced them with

the duration time of a signature expiration if one was present. The variables for the logistic regression were explored through a correlation analysis as well as the impact each had on the predictive capabilities based on previous models.

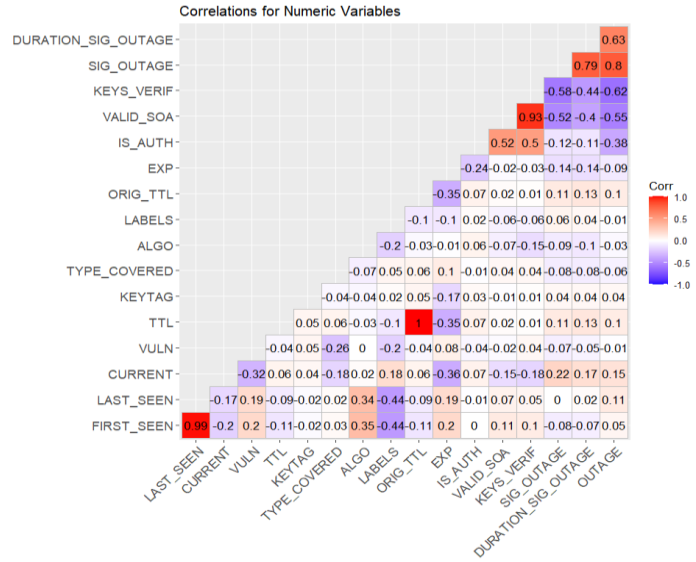

Figure 5: Outage as a Whole correlation plot

$$f(x) = 0.6757 - 6.68e^{-9}(x_1) + 1.84e^{-7}(x_2) + 3.1e^{-7}(x_3) - 5.93e^{-3}(x_4) - 0.1387(x_5) - 0.09277(x_6) - 0.2645(x_7)$$

$$X_{outages} = \begin{cases} 1, & f(x) \geq 0.5 \\ 0, & f(x) < 0.5 \end{cases} \quad (3)$$

***see appendix A2 for the description of variables**

Following the trend of the previous models, the "IS_AUTH", "VALID_SOA", and "KEYS_VERIF" variables are still quite impactive variables in the ability to establish if a server is out.

	Act: No Outage	Act: Outage
Pred: No Outage	333420	5330
Pred: Outage	42	6802

Figure 6: Outage as whole logistic regression results

Although the specificity of this model was much smaller compared to the natural outage, standing at around 56%, it was expected to follow that trend since due to the limitations and challenges occurring during our project we only were able to tackle two types of outages. In the future, if more types are added, we expect improvements in the predictive capabilities; nonetheless, the current model can help in tracking potential instances and tackle them when needed or for academic purposes.

X. Challenges

Our efforts during this project have been affected by several challenges, including some of the following:

A. Time Restrictions

With the delays in data analysis and model testing, time became an important concern. To mitigate these delays, it became imperative to prioritize our tasks, swiftly troubleshoot any difficulties that arose, and focus on setting up the classifier as much as possible.

B. Data Access/Permissions:

Access to and manipulation of data proved to be a challenging endeavor because of permissions problems. We tried to work around this barrier via SQL queries and Excel connections, but permissions were still a barrier that prevented us from extracting and uploading data to AWS. Moreover, there were difficulties with software installation and configuration, especially when it came to resource optimization on AWS servers.

C. Finding External Datasets:

It was not easy to find external datasets to support our project's classification attempts. It was still difficult to find complete and pertinent statistics, even with efforts to gather data from sources such as the Digital Attack Map. The lack of external data made our research and model creation more difficult.

D. Volume of Data:

In our project, we managed an immense volume of data, consisting of millions of records, which presented significant challenges such as the removal of duplicates from such a vast dataset. Another complication arose from the sheer size of the data, which made it difficult to isolate specific name servers associated with critical outages. Additionally, the bulk of the dataset posed logistical challenges in transferring the entire set to AWS, necessitating the adoption of a subset model for more manageable processing.

E. New Field of Study:

As a team with a background in data analytics, delving into the realm of cybersecurity presented a steep learning curve. The novelty of the project demanded extra time for understanding and addressing unclassified terms and data columns, requiring close collaboration with our sponsors and extensive research under their guidance.

Moreover, challenges such as data transfer issues between MariaDB and AWS, access denials, and handling missing values were tackled through joint efforts in classification and problem-solving, further supported by our sponsors.

F. Sampling Bias:

Generally, this affects data interpretation and distorts the representation of what is being studied, especially when certain segments of the data are overrepresented or underrepresented in the sample. In this project, there was a higher presence of 'no outages' compared to 'partial outages' or 'complete outages' which may have potentially undermined the reliability and validity of the overall findings.

In conclusion, although the project presented multiple complexities, our methodical approach in addressing each challenge not only enhanced our understanding of the cybersecurity domain but also underscored the importance of rigorous data management and security practices.

XI. Future Scope

Our current research primarily addresses the identification of non-functional nameservers. Future studies could extend beyond mere identification, aiming to prevent such outages. Initial investigations should target determining the nameservers most susceptible to failures by utilizing historical data and predictive modeling. This approach could significantly enhance our ability to maintain zone functionality by preemptively addressing potential vulnerabilities.

In addition, while this study concentrated on specific outage types—namely Expired Signature and Natural Disaster Outages—there is scope to broaden our analysis. Future research could encompass a wider array of outage causes such as Cache Poisoning, DDOS attacks, Configuration Issues, Hardware Failures, Network Problems, and TLD Failures. An expanded study in these areas could facilitate the development of a comprehensive classification model. This model would predict outage types based on the most affected features during an incident, improving our predictive capabilities and response strategies.

Furthermore, to complement these predictive models, we propose the development of more dynamic and interactive dashboards. Tools like Tableau or Power BI could be leveraged to integrate live data feeds, particularly from areas experiencing natural disasters. Such real-time visualizations would allow users to assess the immediate risk of outages in their respective zones, enhancing both awareness and preparedness.

This strategic expansion of our research and analytical tools aims not only to enhance our understanding of nameserver vulnerabilities but also to elevate the resilience of our network infrastructure against diverse outage scenarios.

XII. Conclusion

This investigation has yielded valuable insights into the vulnerabilities and operational hurdles encountered with DNSSEC implementation. Through meticulous examination of outages and their root causes, our study has understood the pressing necessity for continuous strong monitoring capabilities and refined predictive models to safeguard DNS systems against a spectrum of threats, ranging from natural calamities to cyber-attacks.

Our inquiry has brought to light the profound impacts specific outage types, such as those stemming from expired signatures and natural disasters, can inflict upon the accessibility and integrity of DNS services. These revelations accentuate the imperative for robust DNSSEC configurations and continual vigilance to mitigate potential vulnerabilities.

The Random Forest model pinpointed the useful features for the machine learning models. The features when utilized in the logistic regression classification model for expired signature outage gave us **95%** accuracy.

The Natural Disaster Outages when further analyzed using a logistic regression model gave a sensitivity value of **100%** and an overall accuracy of **99.5%.** The confusion matrix provided supports our conclusion as our model can predict which nameservers could face an outage given a natural disaster.

The overall model for classifying outages based on the features affected has an accuracy of **99.7%** and the specificity is at **56%**.

Looking ahead, the imperative lies in expanding the horizon of outage analysis and refining predictive methodologies. This entails delving into additional categories of DNSSEC-related failures and diversifying the dataset to encompass a broader spectrum of global instances. Additionally, the relentless pursuit of more dynamic and responsive tools for real-time data visualization and risk assessment will be instrumental in enhancing our capacity to proactively address and mitigate potential disruptions.

Acknowledgment

The authors would like to thank their instructor, Professor F. Brett Berlin, and their sponsors, Dr. Jean-Pierre Auffret and Dr. Eric Osterweil, for their guidance throughout the project.

References

[1] T. Rooney, "Securing DNS (Part I)," 2010.

[2] Cloudfare, Inc., "How DNSSEC Works," n.d. https://www.cloudflare.com/dns/dnssec/how-dnssec-works/

[3] T. Rooney, "Securing DNS (Part II): DNSSEC," 2010.

[4] A. T. Tunggal, “Why is Cybersecurity Important?,” UpGuard, Aug. 28, 2019. https://www.upguard.com/blog/cybersecurity-important

[5] S. Preston, “What Is DNSSEC, and How Does It Work?,” Akamai, Jun. 09, 2022. https://www.akamai.com/blog/trends/dnssec-how-it-works-key-considerations

[6] T. Plant, “Causes of Internet Outages and How to Avoid Them,” Ecessa, a product of OneNet Global, Feb. 13, 2023. https://www.ecessa.com/blog/causes-of-internet-outages-and-how-to-avoid

[7] PCMag, "DNS ERROR BREAKS SWEDEN'S INTERNET DOMAINS," Oct. 13, 2019. https://www.pcmag.com/archive/dns-error-breaks-swedens-internet-domains-244950

[8] ICANN, “What Does ICANN Do? - ICANN,” *www.icann.org*. https://www.icann.org/resources/pages/what-2012-02-25-en

[9] DNS Response Policy Zones, "DNS RPZ," 2024. https://dnsrpz.info/

[10] P. Hoffman, "DNS Security Extensions (DNSSEC)," Feb. 2023. https://www.ietf.org/archive/id/draft-ietf-dnsop-dnssec-bcp-00.html

[11] R. Gunasekarn and E. Osterweil, "SecSpider Database," Oct. 6, 2020.

[12] Q. Nguyen, "Mining DNSSEC outages using SecSpider Database," George Mason University, 2023.

APPENDIX

A1) Data Variables

The Database is set up using MariaDB a relational database management system. There are 4 different databases each containing the same tables but differing in dates of the data. The database we are focusing on contains entries from 2013 to 2020. The database contains 40 different tables. Each tables contains specific information about DNS, outages, poller, nameserver, expired signatures and etc. The database is accessed using regular SQL queries. The tables are in one-to-one, one-to-many and many-to-many relationships.
Let us understand few of the tables where we are focusing:

SS_POLLER -> Poller is used to query dns server at regular intervals to gather information. They check for multiple things like: Monitoring (to check the health and responsiveness of DNS servers. DNS pollers can monitor the uptime of DNS services and ensure that they are functioning correctly.), validation (to validate dnssec signatures), data collection (To gather DNS records and maintain up-to-date information. Pollers can be used by services like DNSSEC monitoring tools to collect data on the current state of DNSSEC deployments.), Synchronization, Dynamic DNS updates.

ID	Primary Key
Name	Name of the place where it is hosted from
Descr	Description of the place
Host	IPV4 address of the place
Port	Port Number
Online	Binary variable (1 -> Online, 0 -> Offline)
TSIG_KEY_FILE	TSIG signature file
Latitude	Latitude of the place
Longitude	Longitude of the place

SS_ZONE -> This table contains the information about the zone where all the DNS and RRSET is stored.

ID	Primary Key
Poller_ID	Reference Key to Poller table
First Seen	The instance when Zone was created
Last Seen	The instance when Zone was stopped/ceased
Name	Name of the Zone
Mon_Reason	
Avail	Binary Variable (1 -> Available, 0 -> NA)
Verif	Binary Variable (1 -> Verified, 0 -> NV)
Valid_Deleg	
Valid_Fresh	
DNSSEC	Binary Variable (1 -> Secured, 0 -> NS)

SS_RRSET -> This table contains the details about the Resource Record. A Resource Record (RR) is a basic information element that is stored in a zone file on the DNS server. Each RR defines information associated with the domain name. There are several types of DNS resource records, and each type serves a specific purpose. Here are some of the common types of RRs:
1. A Record: The "A" stands for "Address". An A record maps a domain name to its corresponding IPv4 address. For example, if you have a domain name like "example.com", the A record would tell you the actual IPv4 address (like 93.184.216.34) where "example.com" is hosted.
2. AAAA Record: Similar to the A record, but it maps a domain name to its corresponding IPv6 address, which is the newer version of IP addresses.
3. MX Record: The "MX" stands for "Mail Exchange". MX records are used to specify the mail servers responsible for accepting email messages on behalf of a domain and to prioritize mail delivery if multiple mail servers are available.
4. CNAME Record: The "CNAME" stands for "Canonical Name". CNAME records are used to alias one name to another. For example, if you have "www.example.com" and you want it to point to "example.com", a CNAME record can be used.

5. NS Record: The "NS" stands for "Name Server". NS records identify the DNS servers responsible (authoritative) for a zone. A domain like "example.com" will have at least one NS record listing the DNS server that contains the actual DNS information about the domain.
6. TXT Record: The "TXT" stands for "Text". These records hold free-form text of any type. They are often used to provide information to outside sources. For example, TXT records can hold SPF data (Sender Policy Framework) to prevent email spoofing.
7. SRV Record: The "SRV" stands for "Service". SRV records are used to define the location of servers for specified services, such as VOIP (Voice Over IP), instant messaging, and other services that require specific ports.
8. PTR Record: The "PTR" stands for "Pointer". PTR records are used for reverse DNS lookups, mapping an IP address to a domain name, which is the opposite of what A or AAAA records do.

ID	Primary Key
Zone_ID	Reference Key to Zone table
First Seen	The instance when Zone was created
Last Seen	The instance when Zone was stopped/ceased
Name	Name of the Resource Record Domain
RR_Type	Type of Resource Record
Current	Binary Variable (1 -> Currently Active, 0 -> NA)
Vuln	Vuln
NS_IP	IP address of Name Server

SS_RRSIG -> This table contains the whole details of each RR. It contains all the signature information, including the signature and the signer's name.

ID	Primary Key
SET_ID	Reference Key
TTL	Time to live (important variable)
Type Covered	Type of Resource Record Covered
SIG	Signature of the RR
Signers_Name	Name of the root/parent node

SS_DNSKEY -> A DNSKEY is a DNS record type that contains a public signing key. If you are migrating a DNSSEC signed zone to another DNS operator, you might need to see the DNSKEY records.

ID	Primary Key
SET_ID	Reference Key
TTL	Time to live (important variable)
Key_Text	Public Signature Key
ALG	Type of Algorithm
Flags	Flag Set Number

SS_NAMESERVER > This table contains information about the nameserver for every zone.

ID	Primary Key
ZONE_ID	Reference Key
NAME	Name
IP	IP Address

SS_NAMESERVER_STATS > This table has information about the nameserver such as whether the nameserver is online and whether authentication is active.

ID	Primary Key
NS_ID	Reference Key
SEEN	Time Period when server was checked
AVG_RTT_MILLIS	Average Round Trip Time
IS_ONLINE	Server is online

IS_AUTH	Authorization
SERIAL	Zone Serial Number
VERSION	Version of Server
EDNS0	Extension Mechanism for DNS
CNAME	Canonical Name
BOGUS	Bogus
VALID_SOA	Valid Start of Authorirty
KEYS_VERIF	Keys_Verif
FOUND_WHERE	Found_Where

SS_ZONE_STATS > This table has information about the zone such as DNSSEC and Delegation Signer(DS), timestamp, etc.

ID	Primary Key
ZONE_ID	Reference Key
SEEN	Instance request
PARENT	Parent Zone
EDNS0	Extension Mechanism for DNS
DNSSSEC	Boolean (0,1)
DNSSEC_VERIF	DNSSEC Verified
DS	Delegation Signer
DS_VERIF	Delegation Signer Verified
PROD	PROD
AVAIL	Zone available
VERIF	Verified
VALID_DELEG	VALID_DELEG
VALID_FRESH	VALID_FRESH
WILDCARD	Wildcard DNS record (non-existant domain)

We are using two more tables viz. **SS_EX_RRSET** and **SS_RRSET_EXP_REL,** these tables contain the expired signatures of the Resource Record.

A2) Logistic Regression Model Variables

Expired Signatures Model: x1: LAST_SEEN; x2: EXP; x3: VALID_SOA; x4: KEYS_VERIF

Natural Disaster Model: x1: IS_AUTH; x2:VALID_SOA; x3:KEYS_VERIF

Outages as a Whole Model: x1: DURATION_SIG_OUTAGE; x2:TTL; x3: KEYTAG; x4:ALGO;
x5:IS_AUTH; x6:VALID_SOA; x7:KEYS_VERIF

A3) Outage Exploration Subset Variables

Zone ID: from the SS_ZONE Table (existing)
DNSSEC Status**:** from the SS_ZONE Table (existing)
Average Outage Duration: LAST_SEEN (from SS_RRSET Table) – EXP (from SS_RRSIG Table) (new)
Average Signature Validity Duration**:** EXP (from SS_RRSIG Table) -INCEP (from SS_RRSIG Table) (new)
Average TTL**:** AVG TTL (from SS_RR Table) (new)
Total Nameserver IP**:** SUM of Distinct NS IP (from SS_RRSIG) per zone (new)
Total Outage Count: SUM of **:** LAST_SEEN (from SS_RRSET Table) – EXP (from SS_RRSIG Table) >1 (new)
Outage Percentage**:** (SUM(outage_count) / SUM(total_ns_ip)) (new)
Zone Status: (SUM(outage_count) / SUM(total_ns_ip)) * 100 = 100 THEN 1 (SUM(outage_count) / SUM(total_ns_ip)) * 100 = 0 THEN 0 ELSE 2 (new)
Average Outage in Days: (Average Outage Duration/86400) (new)
Average Signature Validity in Days**:** (Average Signature Validity Duration /86400) (new)