BGP Analysis and Anomaly Detection

Info				Reading	Bonus	Effort			Attendance		
Name	SID	Email	Cell*	Book	Papers	Total	Dev	Comm/ Writing	In person	Zoom in real time	Recording
Sai Vamsi Reddy Kinkiri	862466036	skink001 @ucr.edu	9515365913	80	70	20	20	20	5	15	1
Devaki Kalyan Chandra Yadav Podila	862468795	pyada015 @ucr.edu	9513472546	75	75	20	20	20	5	15	1
Hemanth Paladugu	862464963	vpala021 @ucr.edu	9512019348	70	80	20	20	20	5	15	1
Harsha Vardhan Reddy Devarapalli	862467201	hdeva002 @ucr.edu	9513347756	75	75	20	20	20	5	15	1
Sri Sai Naga Venkata Adithya Swarna	82467837	sswar011 @ucr.edu	9515447715	80	70	20	20	20	5	15	1

Presentation Link: Team 05 - RiversideRouting Presentation Link
Project Code: Team 05 - RiversideRouting Project Code

Effort:

- **Sai Vamsi Reddy Kinkiri** played a crucial role in reading and synthesizing insights from relevant research papers, which laid the groundwork for our project's theoretical foundation. He was instrumental in implementing the first part of our program, focusing on parsing IP addresses from JSON files and fetching BGP updates. He significantly contributed to conducting the experiments for the first part of our project. His contributions to the report were centered around elucidating the project's problem statement and our approach to addressing it, ensuring a solid introduction and background for our work.

- **Devaki Kalyan Chandra Yadav Podila** was pivotal in the collection of BGP Update data, leveraging scripts to automate the retrieval of updates from various sources. This effort was critical in amassing the dataset required for our analysis. In addition to data collection, he actively contributed to the development of the second part of our implementation, specifically in automating the analysis of the vast datasets and ensuring the accuracy of our update classifications. His insights were also reflected in the sections of the report and presentation discussing our data collection methodologies and the significance of automated analysis in our project.

- **Hemanth Paladugu** focused his efforts on the hands-on implementation of both parts of our program. He was responsible for integrating the data collection scripts with the analysis modules, ensuring seamless data flow and processing. This integration was key to our project's success, allowing for the efficient handling of large volumes of data. His technical expertise contributed to the detailed sections of the report and presentation that covered the innovative aspects of our project, including the statistical measures and pattern identification techniques we employed.

- **Harsha Vardhan Reddy Devarapalli** took on the challenge of implementing the complex algorithms required for our project's second part. His work involved developing the logic for classifying BGP updates and identifying patterns within the data, which are crucial for understanding the dynamics of internet routing. He also played a significant role in writing the experimental results section of our report, detailing the outcomes of our analyses and the implications of our findings on the stability and efficiency of internet routing.

- **Sri Sai Naga Venkata Adithya Swarna** was deeply involved in the writing and compilation of the final report and presentation. He actively participated in the implementation phase of part 2 of our project. Additionally, Adithya contributed to the project by assisting in reading and analyzing research papers, helping the team to gain a comprehensive understanding of the domain. His contributions to the report's conclusion and future work sections were particularly noteworthy, offering insights into the potential implications of our work and avenues for further research.

- Each team member's contributions were integral to the project's success, reflecting a well-rounded and collaborative effort across all aspects of the work.

BGP Analysis and Anomaly Detection

Sai Vamsi Reddy Kinkiri
Computer Science
University of California
Riverside, CA, USA
skink001@ucr.edu

Devaki Kalyan Chandra Yadav Podila
Computer Engineering
University of California
Riverside, CA, USA
pyada015@ucr.edu

Hemanth Paladugu
Computer Science
University of California
Riverside, CA, USA
vpala021@ucr.edu

Harsha Vardhan Reddy Devarapalli
Computer Science
University of California
Riverside, CA, USA
hdeva002@ucr.edu

Sri Sai Naga Venkata Adithya
Computer Engineering
University of California
Riverside, CA, USA
sswar011@ucr.edu

A. ABSTRACT:

The main objective of this project is to demonstrate the complexities of the Internet's routing mechanism by analyzing the Border Gateway Protocol (BGP) — the postal service of the Internet that dictates how data finds its way from source to destination. With an eye on the route stability and unusual route selections, we did a long-term analysis of vast amounts of BGP data, employing advanced detection methods outlined in scholarly research. Leveraging the latest collected data, we performed experiments outlined in [1] to validate the paper's findings. We have meticulously implemented signature-based detection and time duration/statistics-based modeling [2]. We have classified the patterns in UP, DOWN, FLAT and WITHDRAWAL patterns in consecutive BGP updates collected from the RIPE servers through their APIs for the selected IP prefixes.

We modeled the patterns by harnessing the power of Python, to spotlight irregular patterns that could signal potential disruptions in Internet connectivity. The focus of the implementation is to deep dive into the world of Internet traffic control, finding abnormal outliers from huge datasets winding it down to little focus points to keep the data flowing smoothly and securely. Notably, our investigation leveraged the pivotal Meta(formerly Facebook) outage of 2021 and several critical network outages as test cases, substantiating our model's validity. The models were not only tested against these significant events but were also visualized comprehensively, demonstrating their practical utility and enhancing our understanding of BGP's behavior during Internet-scale disruptions.

B. INTRODUCTION:

The digital landscape of the 21st century is defined by connectivity, underpinned by complex routing protocols that govern the flow of data across the global Internet. At the heart of this digital ecosystem is the Border Gateway Protocol (BGP), a critical infrastructure that, despite its robust design, is vulnerable to disruptions that can ripple through the

network, manifesting as large-scale outages. The importance of maintaining BGP stability cannot be overstated, as it directly impacts economic, social, and security dimensions of the Internet.

The general scope of this project is to investigate BGP's intricate dynamics, pinpoint weaknesses, and enhance its reliability through sophisticated analysis and detection of outliers. The input involves BGP update messages from vast datasets, including events like the Meta outage of 2021 and output is a comprehensive analysis of BGP dynamics, classification of update patterns, and identification of outliers. The BGP data collected is representative of the broader Internet's health and that patterns indicative of stability or outliers can be accurately discerned from this data.

The previous research into BGP dynamics has largely focused on snapshot analyses and limited datasets, offering a fragmented understanding of the protocol's behavior. There exists a considerable gap in long-term, in-depth analyses that leverage substantial amounts of data to draw meaningful insights into BGP's operational intricacies. As the demands on the Internet grow, the need for advanced, data-driven approaches to BGP analysis becomes more pressing.
This work takes a pioneering approach to BGP analysis by employing a dual-model detection system that incorporates signature-based and time duration-based methods, validated against significant outage events. The key results and contributions of this work are:

- We conducted a long-term analysis of BGP UPDATE traffic, examining the stability of routing paths and identifying outliers within an extensive dataset.
- We utilized the Meta outage of 2021 and several crucial outages as benchmarks to validate our models, proving their effectiveness through meticulous visualizations.
- We demonstrated that our models could detect and categorize BGP outliers with a high degree of accuracy, providing insights that are critical for proactive network management and response to Internet-scale disruptions.
- Our project extends the boundaries of current BGP research by incorporating real-world events into the validation process, offering a more grounded and applied understanding of the protocol's dynamics.
- By bridging the gap between theoretical research and practical application, this project not only contributes to the academic discourse but also serves as a valuable tool for network administrators and policy-makers tasked with safeguarding the Internet's backbone.

C. BACKGROUND & PROBLEM DEFINITION:

The Internet, imagine it as a vast network of roads along which information travels, relies on something called the Border Gateway Protocol (BGP) to guide data to its destination, much like traffic signals guide cars. However, BGP isn't foolproof, sometimes, it can be misled, resulting in data getting lost or delayed, affecting websites and users worldwide. While studies have explored how BGP works, many have only scratched the surface, focusing on short periods or small sections of the Internet, and missing the bigger, more complex picture.

Our project delves deep into the world of BGP, examining a vast amount of data collected over an extended period to uncover where and how the Internet's routing system occasionally fails. We're employing advanced detection techniques to sift through this data, identifying outliers that could disrupt the Internet's smooth operation. To validate our methods, we're putting them to the test against real-life events, such as the Meta outage in 2021. This innovative approach aims to equip those managing the Internet's infrastructure with the tools and insights needed to keep data flowing securely and without interruption.

Addressing the problem of BGP's vulnerabilities requires a nuanced detection system capable of identifying and analyzing routing outliers. By combining signature-based detection, which looks for known patterns of disruption, with statistics-based detection, which flags deviations from normal routing behavior, we aim to enhance the accuracy and reliability of anomaly detection. Visualizing these outliers through graphs transforms the complex world

of BGP updates into something more understandable, helping to pinpoint exactly where the system falters. The challenge lies in processing vast amounts of data, fine-tuning our detection methods to accurately identify true outliers, and adapting to the ever-evolving landscape of the Internet. The ultimate goal of our project is not just to spot problems but to understand their causes and work towards preventing them, ensuring the Internet remains a stable and secure platform for global communication.

D. METHODOLOGY / IMPLEMENTATION:

The first part of our project introduces a novel approach to studying BGP (Border Gateway Protocol) routing behaviors over large timescales by developing and utilizing sophisticated data extraction and analysis scripts. Our contribution lies in the detailed, time-extended study of BGP path stability, leveraging real-world data to understand the dynamics of internet routing more comprehensively than previously possible.

For the second and major part of our project, we categorized each BGP update to understand the nuances of routing preferences, labeling them as UP, DOWN, WITHDRAWAL, or FLAT, depending on their characteristics. This classification, done through automated Python scripts, set the stage for identifying broader BGP patterns, known as Types B, C, and F, which reveal the stability and resilience of routing paths. Alongside, we incorporated several key measures—like Inter-Arrival Time, AS Path Changes, Update Classifications Frequency, and Average Path Length—to quantify and model the BGP update behaviors, providing a structured approach to discerning normal operations from potential disruptions.

Part 1 Implementation:

A custom python script is implemented which initiates the methodology by parsing a predefined JSON file, which contains a list of key-value pairs representing various entities (e.g., companies or services) and their corresponding IP addresses. For each IP address, the script constructs and sends a request to the RIPE Stat API [4] to retrieve BGP updates within a specified timestamp range. These updates include both announcements (A) and withdrawals (W) related to the BGP routes. The script dynamically generates and saves these updates into individual JSON files named after each entity. This process ensures a structured collection of BGP routing data, ready for in-depth analysis.

Following data collection, the main script takes over to analyze the BGP updates. It starts by reading the generated JSON files, creating prefixes (source AS, destination AS and path) and extracting crucial information such as timestamps, BGP path attributes, and the lifecycle of prefixes (from first seen to last seen timestamps). It computes several key metrics, such as the total number of distinct paths used by a prefix, the average lifetime of these prefixes, and the distribution of path lifetimes. Furthermore, the script categorizes prefixes based on their active durations and quantifies path stability by examining the consistency of path usage over the observed period.

The analysis script employs a meticulous approach to understanding BGP routing dynamics. It calculates lifetimes, identifies dominant paths, and assesses the stability and diversity of routing paths across different prefixes and Autonomous Systems (AS). This comprehensive analysis provides insights into the temporal characteristics of BGP routes, including path persistence and the prevalence of route changes.

Part 2 Implementation:

a. Classification of BGP Updates:

To understand the changes in BGP route preferences, each update from our dataset was classified based on the following criteria:

UP: An update was classified as UP when it introduced a more preferred routing path. This could involve a shorter AS_PATH, a more favorable ORIGIN type value compared to the previous update for the same route.

DOWN: An update was marked as DOWN when it presented a less preferred path, which could mean a longer AS_PATH, less favorable ORIGIN type.

WITHDRAWAL: If an update indicated that a previously available route was no longer available, it was classified as a WITHDRAWAL. This could be due to a route being removed from the BGP table.

FLAT: When consecutive updates showed no change in preference, meaning the same AS_PATH length and this update was classified as FLAT.

We automated this classification process using a Python script that parsed through the collected BGP update messages. The script assessed the changes in attributes between consecutive updates for the same route to determine the appropriate classification.

b. Identification of Patterns B, C, F:

Beyond individual classifications, we looked for specific sequences of updates that indicated certain behavior patterns in BGP updates, namely B, C, and F:

Type B - Transient Failure Followed by Fast Failover: A B-type sequence contains a WITHDRAWAL (WD) of a route within a series of updates, indicating a temporary routing disruption and subsequent rapid recovery. For instance, let's consider a route via AS_PATH [ISP1, ISP2, ISP3]. If this path is announced and then withdrawn, followed by the quick announcement of a new path [ISP1, ISP4, ISP3], this sequence would be classified as Type B, signaling a transient failure with a quick rerouting solution.

Type C - Single Preference Fluctuation: Type C is characterized by only one instance of a routing preference fluctuation without any withdrawal, meaning there is a single UP or DOWN movement within a sequence of updates. For example, if the original AS_PATH [ISP1, ISP2, ISP3] changes to [ISP1, ISP2, ISP5, ISP3] in one update, indicating a less preferred route (DOWN) due to an additional AS, and then reverts back to the original AS_PATH in a subsequent update (UP), this sequence is a C type. It suggests either a policy change or a transient network condition that briefly affected the route preference.

Type F - Steady State: The F-type pattern is observed when a series of updates all maintain the same routing preference. Imagine a scenario where consecutive updates keep showing the same AS_PATH [ISP1, ISP2, ISP3] without any changes in origin type. This scenario would be classified as Type F. It might hint at a stable route but could also indicate potential issues such as route oscillation if these FLAT updates are unexpectedly frequent.

Type	Pattern	Examples
B	A sequence of updates with WD in the middle	<D,D,W,U,U> <D,W,U,W,U>
C	A sequence of updates with only one preference fluctuation	<U,U,D,D,F> <D,D,U,F,U> <D,D,U,U,D>
F	A sequence of updates with same preference	<F,F,F,F,F>

Table 1: Pattern Identification

To identify these patterns, we created a windowed analysis mechanism in our code, where sequences of updates were examined for each BGP prefix. By iterating through these sequences, we were able to detect the presence of B, C, and F patterns, which informed us about the stability and health of BGP routes over time.

In our study, following the methodologies described in scholarly papers, we developed a set of measures to assess and model the behavior of BGP updates. These measures allowed us to statistically understand the BGP dynamics beyond the basic UP, DOWN, FLAT, and WITHDRAWAL classifications. The measures we implemented are:

Inter-Arrival Time: Inter-Arrival Time refers to the time elapsed between consecutive BGP updates for the same prefix. This metric is significant because it can indicate the stability or volatility of a route. Shorter inter-arrival times might suggest instability or frequent route changes, while longer times could imply stability.

Number of AS Path Changes: This measure counts how often the AS_PATH attribute changes within a set period, indicating the frequency of route changes across different autonomous systems. Frequent AS path changes might signal an unstable routing environment or active route optimization efforts.
Frequency of Specific Update Classifications: We tracked how often each type of update classification occurred within our dataset. This frequency analysis helped us identify which types of updates were most common and could potentially indicate normal operational behavior versus outliers.
Average Path Length: The Average Path Length measure computes the average number of autonomous systems (ASes) that BGP updates traverse. This can reflect the efficiency of the routing paths chosen over time and highlight potential deviations from expected behavior.

These measures were critical in interpreting the complex dynamics of internet routing and provided a quantitative backbone for our anomaly detection system.

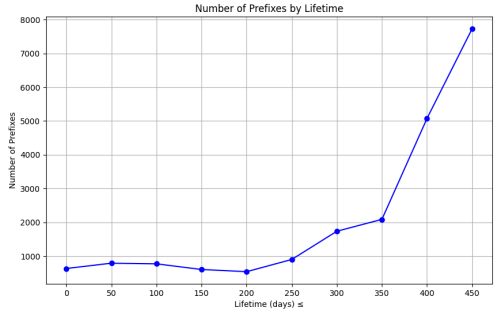
E. EXPERIMENTAL RESULTS:

In this section, we present the results and key observations derived from applying our BGP analysis methodology, particularly focusing on the classification of BGP updates during specific network events and the overall performance over a recent ten-day period.

Part 1 Results:

The methodology and analysis presented in this study are based on an extensive dataset, encompassing a total of 3.88 GB of BGP routing information, meticulously collected over a period of 658 days. This dataset incorporates information from a diverse range of IP prefixes, culminating in the acquisition of data pertaining to 22,200 unique prefixes. The findings derived from this comprehensive dataset are summarized as follows:

- The cumulative number of prefixes analyzed in this study amounts to 22,200.
- A subset of 882 prefixes, equivalent to approximately 3.97% of the total dataset, was observed to have visibility for less than 10 days. This implies that the majority of the prefixes live more than 10 days.
- A significant proportion of the dataset, comprising 10,436 prefixes or 47.01% of the total, demonstrated a consistent routing path for no less than 50% of their observable lifespan. This implies that approximately half of prefixes had only one path that followed for 50% of the prefix lifetime.
- Furthermore, 7,010 prefixes, representing 31.57% of the collected data, utilized two or fewer distinct routing paths during their period of visibility showcasing the network stability.
- An illustrative graph detailing the distribution of prefix lifetimes within the dataset is provided to offer additional insights into the temporal characteristics of BGP prefix longevity.
- This empirical analysis provides a granular view of BGP routing dynamics, offering valuable insights into the temporal stability and path diversity of internet routing prefixes over an extended observation window.



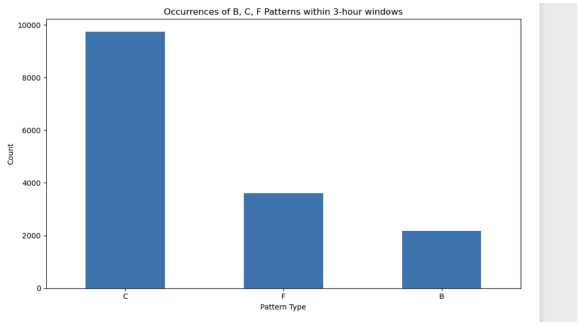
Part 2 Results:

Validation Against Meta Outage:
Our analysis during the Meta outage incident in 2021 revealed a stark contrast between the BGP update patterns observed during the outage and those under normal operational conditions. The graphs constructed from the dataset showed a significant increase in WITHDRAWAL (WD) and DOWN classifications during the outage window. This was indicative of route withdrawals and less preferred paths being announced, which aligns with the expected behavior during a large-scale network disruption.

The visualization of the data, through time-based scatter plots, clearly depicted the temporal concentration of abnormal BGP activities, with a clustering of WD and DOWN points during the outage period. In contrast, the pattern normalized post-outage, reverting to a distribution consistent with regular network operations, with more UP and FLAT classifications, indicating the restoration and stabilization of BGP routes.

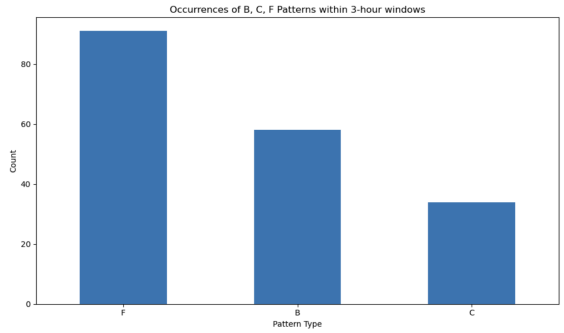
Graph 1.1 (Meta during Outage):

X-axis: Represents the different types of BGP update patterns detected ('B', 'C', 'F').
Y-axis: Shows the frequency of occurrences of each pattern type within 3-hour time frames. Graph Explanation: The graph visualizes how often each BGP update pattern occurs within short, 3-hour windows, providing insights into the temporal dynamics of routing behaviors. This step is based on the assumption that patterns occurring over a period longer than 3 hours may not be as relevant or could represent different network behavior.



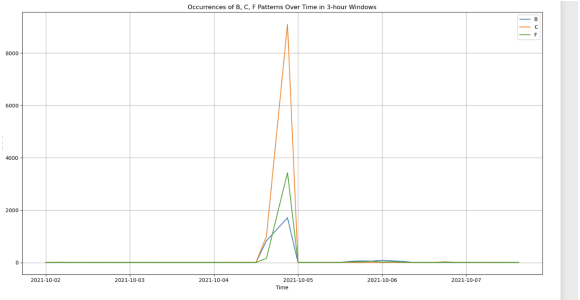
Graph 1.2 (Meta during Normalcy):

X-axis: Represents the different types of BGP update patterns detected ('B', 'C', 'F').
Y-axis: Shows the frequency of occurrences of each pattern type within 3-hour time frames. Graph Explanation: The graph visualizes how often each BGP update pattern occurs within short, 3-hour windows, providing insights into the temporal dynamics of routing behaviors. This step is based on the assumption that patterns occurring over a period longer than 3 hours may not be as relevant or could represent different network behavior.



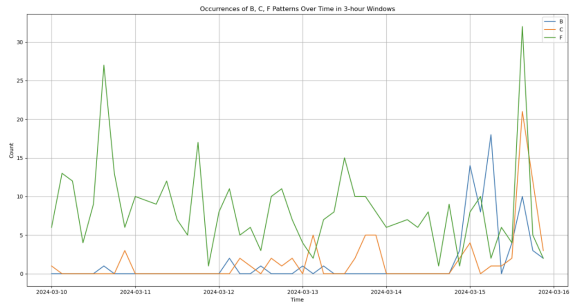
Graph 2.1 (Meta during Outage):

X-axis: Displays the timeline, indicating the date and time when the BGP update patterns were observed.
Y-axis: Shows the number of BGP update pattern occurrences recorded within each 3-hour window along the timeline. Graph Explanation: The lines of different colors represent the frequency of each BGP update pattern type ('B', 'C', 'F') as they occur over time. The sharp peaks, particularly noticeable for the 'C' pattern, suggest periods of high activity or potential routing events that were detected during those specific 3-hour windows.



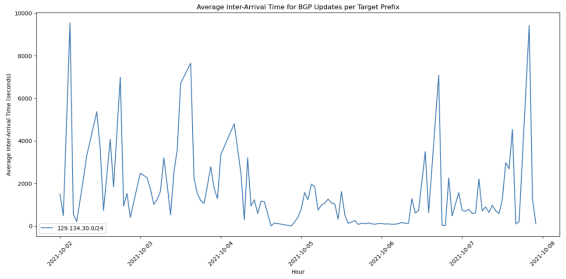
Graph 2.2 (Meta during Normalcy):

X-axis: Displays the timeline, indicating the date and time when the BGP update patterns were observed.
Y-axis: Shows the number of BGP update pattern occurrences recorded within each 3-hour window along the timeline.
Graph Explanation: The lines of different colors represent the frequency of each BGP update pattern type ('B', 'C', 'F') as they occur over time. The sharp peaks, particularly noticeable for the 'C' pattern, suggest periods of high activity or potential routing events that were detected during those specific 3-hour windows.



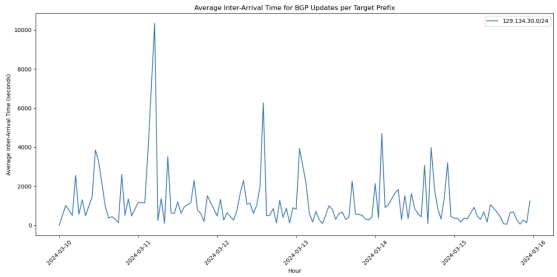
Graph 3.1 (Meta during Outage):

X-axis: Represents time, laid out by hour, across a span of days.
Y-axis: Shows the average time, in seconds, between consecutive BGP updates for a specific target prefix.
Graph Explanation: This graph illustrates the average time gaps between received BGP updates, again for a specific target prefix, and how they change over time. Large spikes might indicate periods with less frequent updates, whereas lower points suggest more frequent updates, potentially highlighting periods of network instability or routing changes.



Graph 3.2 (Meta during Normalcy):

X-axis: Represents time, laid out by hour, across a span of days.
Y-axis: Shows the average time, in seconds, between consecutive BGP updates for a specific target prefix.
Graph Explanation: This graph illustrates the average time gaps between received BGP updates, again for a specific target prefix, and how they change over time. Large spikes might indicate periods with less frequent updates, whereas lower points suggest more frequent updates, potentially highlighting periods of network instability or routing changes.

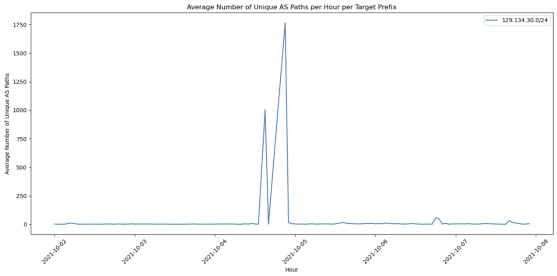


Graph 4.1 (Meta during Outage):

X-axis: Represents time, specifically hours, over a period of several days.

Y-axis: Indicates the average number of unique AS paths per hour for a specific target prefix.

Graph Explanation: The graph shows the variation in the number of unique AS paths observed for BGP updates related to a particular target prefix over time. The spikes represent hours when a significantly higher number of unique AS paths were recorded, which could suggest routing changes or instability.

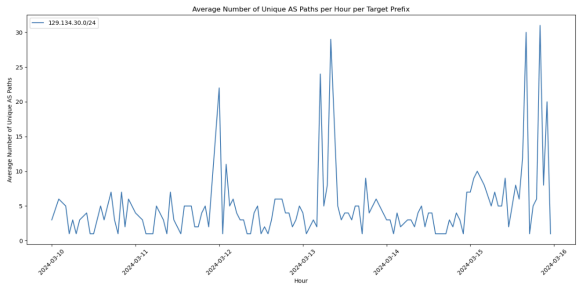


Graph 4.2 (Meta during Normalcy):

X-axis: Represents time, specifically hours, over a period of several days.

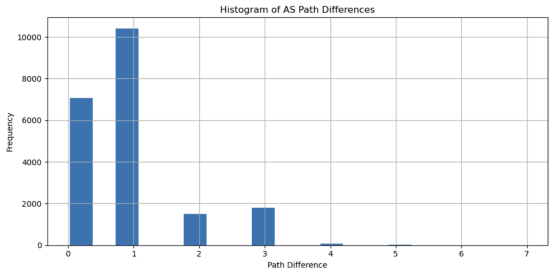
Y-axis: Indicates the average number of unique AS paths per hour for a specific target prefix.

Graph Explanation: The graph shows the variation in the number of unique AS paths observed for BGP updates related to a particular target prefix over time. The spikes represent hours when a significantly higher number of unique AS paths were recorded, which could suggest routing changes or instability.



Graph 5.1 (Meta during Outage):

X-axis: Represents the absolute difference between the length of an AS path and the average AS path length.
Y-axis: Indicates the frequency of AS path occurrences for each path length difference.
Graph Explanation: The histogram shows how often each AS path length difference occurs, providing a visual representation of the variation in AS path lengths compared to the average.

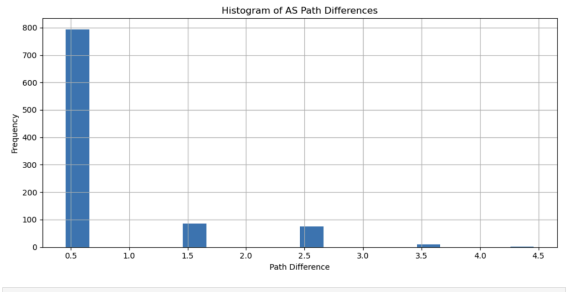


Graph 5.2 (Meta during Normalcy):

X-axis: Represents the absolute difference between the length of an AS path and the average AS path length.

Y-axis: Indicates the frequency of AS path occurrences for each path length difference.

Graph Explanation: The histogram shows how often each AS path length difference occurs, providing a visual representation of the variation in AS path lengths compared to the average.

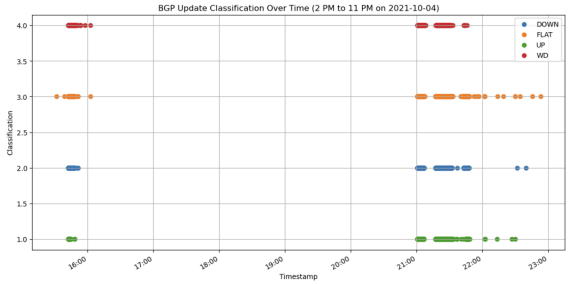


Graph 6.1 (Meta during Outage):

X-axis: Shows the time of day on October 4, 2021, from 2 PM to 11 PM.

Y-axis: Represents a numeric encoding of the BGP update classifications, with an additional category: 4 for 'WD' (withdrawal).

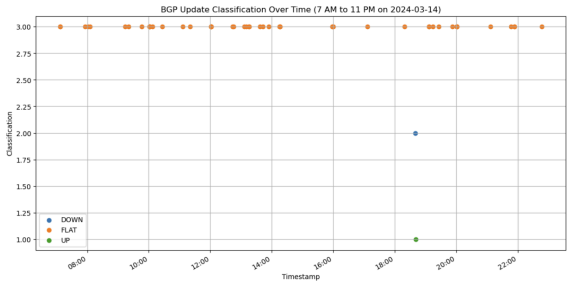
Graph Explanation:This graph indicates a more volatile network state with several 'WD' events signifying BGP withdrawal updates, which could be indicative of network issues such as outages or instabilities.



Graph 6.2 (Meta during Normalcy):

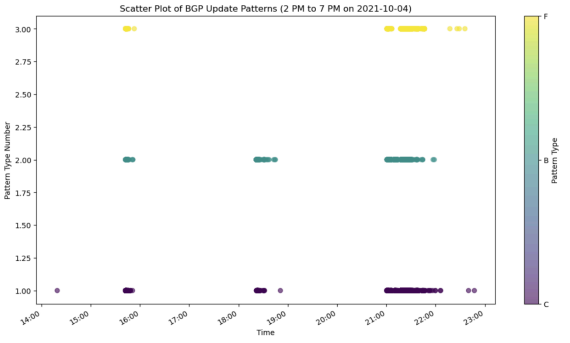
X-axis: Displays the time of day on March 14, 2024, from 7 AM to 11 PM.
Y-axis: Represents a numeric encoding of the BGP update classifications: 1 for 'UP', 2 for 'DOWN', and 3 for 'FLAT'.
Graph Explanation: This graph shows a stable pattern of BGP update classifications over the course of the day, primarily 'FLAT' with very few 'DOWN' or 'UP'

updates, suggesting a relatively stable network state with little change in BGP routing.



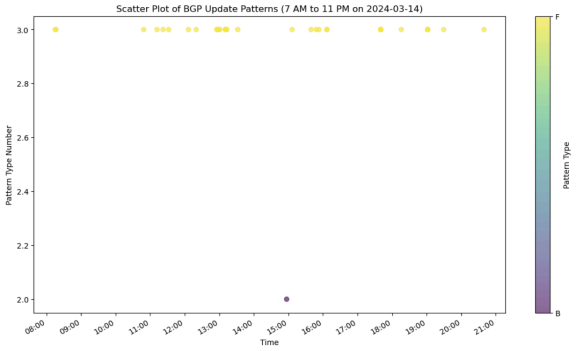
Graph 7.1 (Meta during Outage):

X-axis: Displays the time of day, showing when each BGP update pattern was observed.
Y-axis: Represents different BGP update patterns, which have been assigned numerical values for visualization purposes.
Graph Explanation: The scatter plots illustrate the occurrence of different BGP update patterns over a specified time period. The color and position of each point correspond to the type of BGP pattern and the time it occurred. The points are more scattered across different classifications, including 'WD'. The presence of 'WD' points, in particular, could indicate network instability or disruptions, such as route withdrawals which are often associated with network outages or routing issues.



Graph 7.2 (Meta during Normalcy):

X-axis: Displays the time of day, showing when each BGP update pattern was observed.
Y-axis: Represents different BGP update patterns, which have been assigned numerical values for visualization purposes.
Graph Explanation: The scatter plots illustrate the occurrence of different BGP update patterns over a specified time period. The color and position of each point correspond to the type of BGP pattern and the time it occurred. The distribution of points is mostly constant across the 'FLAT' classification, with a couple of instances of 'DOWN' and 'UP'. This indicates a generally stable network condition without significant BGP updates or disruptions.



Recent Performance of UCR BGP Updates:

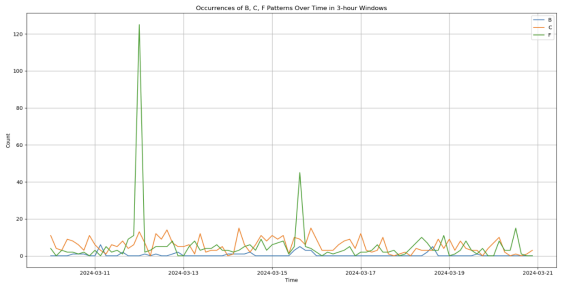
For a contemporary assessment, we analyzed the BGP updates from the UCR's network spanning the last ten days.

The experiments done on the Meta data are conducted on the UCR BGP data and the plots generated from this analysis provided a visualization of the network's BGP behavior over a continuous time frame.

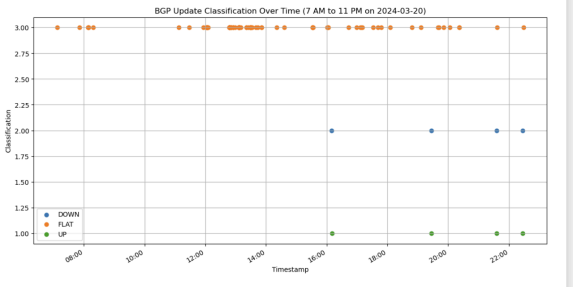
During this period, the BGP updates predominantly fell into the FLAT category, suggesting a stable and consistent routing environment. Occasional UP and DOWN classifications were noted, which is expected in a dynamic routing context and did not indicate any significant instability or network disruptions. No substantial WD activity was observed, further underscoring the stable state of the network during this window.

The dimensions of the X-axis and Y-axis remain identical to those previously established during the experiments conducted on the Meta BGP data.

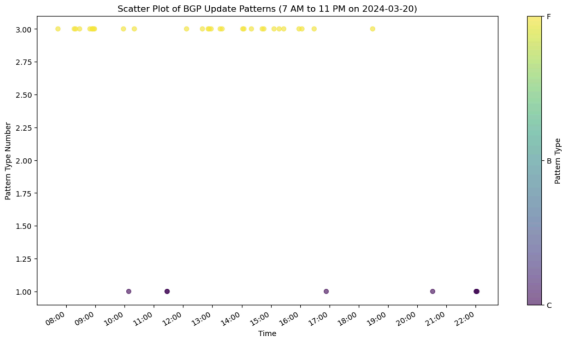
Graph 8.1 (UCR BGP):



Graph 8.2 (UCR BGP):



Graph 8.3 (UCR BGP):



F. DISCUSSION:

Our project's discussion pivots on the inferences drawn from the analysis of vast BGP datasets, focusing on route stability and anomaly detection without the use of machine learning algorithms. Our analytical framework, rooted in the implementation of signature-based detection and statistical measure based models, has provided a fertile ground for understanding the complexities of BGP dynamics.

We took a close look at how the big Meta outage in 2021 and other similar events shook up the internet's usual flow of information. During these times, we noticed a lot more instances where routes were dropped or became less effective—just like we'd expect when there's a big problem causing a stir in the network. Our system for spotting these issues really proved its worth by picking up on these changes. Plus, we made some clear graphs that showed what was going on with the network during those outages, as well as when things were running normally. This helped us show that our system is really good at noticing when and where the internet hits a snag.

In the case of the UCR network data analysis over the last ten days, the prevalence of FLAT classifications illuminated the inherent stability and robustness of the network's routing protocol. These observations, coupled with a lack of significant WITHDRAWAL events, painted a portrait of a well-managed and resilient network infrastructure.

Our methodological approach underscores the cruciality of data completeness for the integrity of the analysis. Through careful data preprocessing, including the treatment of missing values, we ensured a dataset conducive to rigorous analysis. While we did not venture into machine learning predictions, the application of our methods yielded a framework for pattern recognition and the detection of outliers which is both robust and sensitive.

We acknowledge, however, that our scope was limited to the datasets provided by RIPE servers. While these datasets are extensive, they do not encapsulate the entirety of global BGP dynamics. As such, the possibility of unobserved outliers outside these datasets presents a limitation and an opportunity for future exploration.

Looking ahead, we aspire to enhance our analysis by incorporating real-time data and expanding the breadth of our detection algorithms to keep pace with the ever-evolving landscape of BGP updates. While

this project has laid a significant groundwork in the realm of BGP analysis, it is but a stepping stone towards the larger objective of securing the internet's routing fabric, ensuring its operational integrity for all users.

In conclusion, our discussion reaffirms the efficacy of our non-machine learning analytical approach in scrutinizing BGP updates, while also acknowledging the limitations and setting the stage for future work. Our project moves the dial forward in the endeavor to create a more secure and stable internet routing environment, contributing valuable insights to both academic research and practical network management.

G. RELATED WORK:

BGP Routing: A study at Large Time Scale (2002)
This paper[1] provides an extensive analysis of BGP routing stability over several years. It highlights the characteristics of routing paths, including their prevalence, persistence, and influence on network policies and traffic engineering. The findings reveal the dynamic nature of BGP routing, focusing on the stability and changes in routing paths.

On Detection of Anomalous Routing Dynamics in BGP (2004)
The paper[2] discusses methods for identifying anomalies in BGP routing dynamics, utilizing signature-based and statistics-based detection techniques. It emphasizes learning from historical data to recognize deviations in BGP Update traffic. The research is instrumental in understanding how to pinpoint and analyze irregularities in BGP events effectively.

An Internet Routing Forensics Framework for Discovering Rules of Abnormal BGP Events (2005)
The paper[3] presents a framework to detect abnormal BGP events using data mining techniques. It leverages insights on normal BGP similar to that of paper[1] to train models to identify anomalies. It validates the framework on real BGP data during worm attacks and blackouts, showing it can accurately detect unseen events. Overall, it can be said that the paper takes the characterization of BGP from paper[1] and applies it to detect BGP anomalies.

Multi-path inter-domain routing:The impact on BGP's scalability,stability and resilience to link failures (2011)
The paper[9] explores the implementation and effects of multi-path inter-domain routing methods on BGP, focusing on scalability, stability, and resilience to failures. The paper adds up on the work done in paper[1] by evaluating how multi-path routing can potentially address some of the stability and scalability challenges identified in earlier studies, suggesting improvements over the traditional BGP approach. The study conducts simulations to evaluate the impact of multi-path routing methods on BGP's performance, offering a more experimental and predictive approach.

Modeling BGP Updates for Anomaly Detection using Machine Learning (2023)
The 2023 paper[6] by Daniel Robertson, "Modelling BGP Updates for Anomaly Detection using Machine Learning," advances the field of detecting and visualizing Border Gateway Protocol (BGP) update anomalies in real-time. The paper presents a system that can visualize router topology and detect BGP anomalies using historical and live BGP traffic data. This approach marks an improvement from the paper[2], which focused on the detection of anomalous routing dynamics in BGP using signature-based and statistics-based detection without the real-time component and advanced visualization techniques that Robertson's paper introduces.

BGP anomaly detection - a path-based approach (2023)
This paper[7] focuses on analyzing the paths of BGP updates rather than individual updates themselves, leveraging the insight that anomalies often manifest as unusual changes in routing paths. The paper builds on the foundation done in paper[2] by introducing a novel path-based approach to anomaly detection. This method combines topological information and attributes from BGP routing tables to more accurately detect anomalies. The paper suggests a different technique, including a random walk path sampling and modeling method, to understand the underlying structure of the routing table more deeply.

BGP Anomaly Detection Techniques: A Survey (2017)

The paper[8] covers a wide range of methods, including statistical analysis, machine learning, and data mining, each with its advantages and limitations. By examining the existing literature, the authors identify key trends, challenges, and future directions in BGP anomaly detection. The paper provides a broad survey of various anomaly detection techniques in BGP, expanding on the specific methods discussed in the paper[2]. It covers a wider range of techniques and methodologies that have been developed over the years.

H. WHY THIS WORK WAS NOT TRIVIAL:

Our project on analyzing Border Gateway Protocol (BGP) routing dynamics and detecting outliers was a complex endeavor, marked by several challenging aspects. Data collection was a significant hurdle, involving intricate interactions with APIs to fetch extensive datasets. This process required a deep understanding of network protocols and efficient data handling techniques.

The coding effort was considerable, spanning multiple Python scripts and a Jupyter notebook. We developed sophisticated functions for parsing, analyzing, and classifying BGP updates. Handling the vast volumes of BGP data necessitated advanced programming techniques to ensure processing efficiency.

Understanding the complexity of BGP itself was another challenge, demanding an extensive literature review and the application of this knowledge to our data analysis. Moreover, moving beyond analysis to anomaly detection introduced complexities that required innovative analytical models and machine learning techniques.

In summary, the project's challenges, from complex data collection and the need for advanced coding to understanding and analyzing BGP's intricacies, underscored its non-trivial nature. Overcoming these obstacles required skill, creativity, and dedication, highlighting the project's contribution to network analysis and its significance as a learning experience.

I. CONCLUSION:

In this project, we delved into the intricate dynamics of the Border Gateway Protocol (BGP), the backbone of Internet routing, to understand and detect anomalies that could lead to disruptions. By leveraging large-scale datasets and advanced analytical techniques, we aimed to enhance the stability and reliability of BGP operations.

Our main contributions and results include:

- Conducting a long-term analysis of BGP UPDATE traffic, examining the stability of routing paths and identifying outliers within an extensive dataset spanning multiple years.
- Implementing a dual-model detection system that combines signature-based and time duration/statistics-based methods, validated against significant real-world events like the Meta outage of 2021 and several other crucial network outages.
- Demonstrating the effectiveness of our models in accurately detecting and categorizing BGP outliers, providing valuable insights for proactive network management and response to Internet-scale disruptions.
- Bridging the gap between theoretical research and practical application by incorporating real-world events into the validation process, offering a more grounded and applied understanding of BGP dynamics.

The significance of this work lies in its potential to contribute to the academic discourse while serving as a valuable tool for network administrators and policymakers tasked with safeguarding the Internet's backbone. Our models and visualizations can aid in pinpointing potential issues and understanding the root causes of routing disruptions, ultimately contributing to a more secure and stable Internet routing environment.

For practical use, our methodology and findings can be leveraged by network operators and researchers alike. The detection models can be integrated into existing monitoring systems to provide real-time alerts and insights into BGP anomalies. Additionally, our approach can serve as a blueprint for conducting

similar analyses on different datasets or network configurations.

Looking ahead, we aim to enhance our analysis by incorporating real-time data and expanding the breadth of our detection algorithms to keep pace with the ever-evolving landscape of BGP updates. Furthermore, exploring the integration of machine learning techniques could potentially improve the accuracy and adaptability of our anomaly detection system.

In conclusion, this project represents a significant step forward in the endeavor to create a more secure and stable Internet routing environment, contributing valuable insights and tools to both academic research and practical network management.

J. BIBLIOGRAPHY:

[1] G. Siganos and M. Faloutsos, "BGP routing: a study at large time scale," Global Telecommunications Conference, 2002. GLOBECOM '02. IEEE, Taipei, Taiwan, 2002, pp. 2200-2204 vol.3, doi: 10.1109/GLOCOM.2002.1189022. keywords: {Routing;Stability analysis;Convergence;Robust stability;Guidelines;Statistics;Noise robustness},

[2] Zhang, K., Yen, A., Zhao, X., Massey, D., Wu, S.F., Zhang, L. (2004). On Detection of Anomalous Routing Dynamics in BGP. In: Mitrou, N., Kontovasilis, K., Rouskas, G.N., Iliadis, I., Merakos, L. (eds) Networking 2004. NETWORKING 2004. Lecture Notes in Computer Science, vol 3042. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-24693-0_22

[3] Li, Jun & Dou, Dejing & Wu, Zhen & Kim, Shiwoong & Agarwal, Vikash. (2005). An internet routing forensics framework for discovering rules of abnormal BGP events. Computer Communication Review. 35. 55-66. 10.1145/1096536.1096542.

[4] https://stat.ripe.net/docs/02.data-api/bgp-updates.html

[5] http://www.routeviews.org/routeviews/

[6] Robertson, D. (2023). Modeling BGP Updates for Anomaly Detection using Machine Learning. Wellington Faculty of Engineering Symposium. Retrieved from https://ojs.victoria.ac.nz/wfes/article/view/8368

[7] C. Yang and W. Jia, "BGP anomaly detection - a path-based approach," 2023 3rd Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS), Shenyang, China, 2023, pp. 408-414, doi: 10.1109/ACCTCS58815.2023.00100.

[8] B. Al-Musawi, P. Branch and G. Armitage, "BGP Anomaly Detection Techniques: A Survey," in IEEE Communications Surveys & Tutorials, vol. 19, no. 1, pp. 377-396, First Quarter 2017, doi: 10.1109/COMST.2016.2622240.

[9] Adriana Szekeres “Multi-path inter-domain routing:The impact on BGP’s scalability, stability and resilience to link failures” August, 2011