Visual Analysis of Traffic Light Data and Detecting Anomalies

Ashokkumar Madanmohan Thakur, Jahanavi Sanda

Abstract—Traffic Signals are very vital for us in our daily life. These traffic signals help us monitor the flow of traffic. The anomalies in this system will be a huge threat. In our project, we represent a visual analytics approach to observe the pattern and detect anomalies in the traffic signal data. The project implemented has an exploration with the visualization techniques. Data cleaning and several preprocessing techniques for the summary of information of the data have been performed before analyzing the entire data. Through this project, we try to reveal the errors from the input data and found patterns that helped us analyze the abnormalities. We tried to compare the data of different intersections and analyze them. In summary, our project helped us understand the traffic signal control systems using a visual analytical approach.

Index Terms—Traffic Control System, Visual Analytics, controllers

1 INTRODUCTION

With the rising need for mobility, the road transportation network is an important backbone of today's society. Controlling traffic is a major difficulty in many cities, particularly expanding ones. Because traffic throughput capacity in metropolitan areas is limited, it is necessary to enhance traffic network capacity. Controlling traffic lights is an essential component of all road transportation systems, and it is utilized to handle the problem of traffic conflict at junctions. The traffic signal analysis might lead to potential attacks on traffic lights pose a significant risk to national security and local economies. If multiple traffic signals fail, the economic impact could be severe, affecting the entire city. There are several potential vulnerabilities in traffic signal control equipment, each of which can be exploited using malicious intent and hacking techniques. A Traffic Signal's Terminology In addition to the signals themselves, a typical traffic junction signal system includes sensors, controllers, and network devices [8]. Sensors are often installed on the highway to detect traffic flow. Video and other means are utilized on occasion. The sensors may be wirelessly connected or attached to the controller. These provide data to controllers with pre-programmed instructions, which decide the status of the traffic lights. Some controllers operate alone, while others are linked together to create a system to coordinate signal information, and still, others are linked to a central server that delivers orders to them. The Types of attacks that are potential for the traffic signals are: Denial of Service: This attack occurs when legitimate users are unable to access information systems, devices, or other resources. Herein the systems controller shuts down. This could be a possibility of the system failure but there is no evidence to confirm the gaps that we have found in the data. A Light Control attack is a direct attack on the system. This involves changing the controller and intercepting and changing control signals from a central server. [3] Some of the studies have shown that the traffic signal system is vulnerable to hacking and spoofed data can be fed into the system[ref]. We can see the change in the duration of light for this kind of attack. Some cyber-attacks may track the system down and send disrupted signals and attack the system. We provide a visual analytics framework for traffic light data acquired during a field test to help in the identification of possible attacks on a traffic management system. Our technique initially pre-processes the traffic light data, correcting apparent problems such as out-of-order records and detecting missing entries. It then employs the "overview first, zoom, and then details-on-demand" paradigm to facilitate user discovery.

1.1 Motivation

We started to look for anomalies in the traffic signal data, as the attacks are a threat, we can analyze the change of traffic signal patterns. We were keen to observe deviations from the traffic signal data. Finally, the data for this kind of analysis can be large, the data set used here has multiple records per second for only a few hours, resulting in over a million records.

1.2 Contribution

To find out the abnormalities in the visual analytics framework from the collected data. Initially, the method was to preprocess the traffic light data and remove the noise and filter out records as per our usage and detect missing values. We computed the statistics of traffic light data and the duration of each of the lights and after the summarization of the entire information, we visualized using standard plotting techniques, such as bar plots and time series plots, and histograms. From these plots, we observed that there are possibilities of potential attacks on this traffic light system. We have used visual analytics to process the entire data and successfully revealed several hidden errors in the raw data taken for multiple intersections. Though there was a difference in the intersections each of them depicted an anomaly. We observe the short Greenlight and large Red light in some intersections and vice versa in others.

2 RELATED WORK

Visual analytics is a growing area of study in data visualization. It is difficult to go through all of the visual analytics systems and methodologies. There are several technologies available for visual data analysis. Users are frequently required to contribute their expertise to iteratively develop the procedures. Visual analytics tools enable decision-makers to get insight into complicated situations by combining their flexibility, creativity, and previous knowledge with the massive storage and processing capacity of today's computers. Thus, the goal of visual analytics research is to turn information overload into an opportunity: decision-makers should be able to examine this massive, multidimensional, multi-source, time-varying, and frequently conflicting information stream through interactive visual representations to make effective decisions in critical situations.[7] When complicated, fascinating patterns are uncovered, it is sometimes challenging to grasp and interpret the findings in an intuitive and meaningful way [4] Custom visualization software packages for intrusion analysis, based on visual data mining of log data, have been developed using commercial packages mentioned above [5] [6]. In our method, we rely on human judgment and observation, which is aided by data visualization. We're seeking specifics in traffic signal data.

2.2 Background of the data

The data that we have considered for our analysis is from the field test associated with the Multi-Modal Intelligent Signal System. The data is taken from the state of Arizona. We have specifically focused on "Detailed Description for Signal Plans for Roadside Equipment (RSE) Data. The MMITS document provides a map of locations of different intersections where the data was collected. The MMITS Sample data had 1.5 million records, in which each of the intersections was separated. We took multiple intersections and compared them for a better analysis of the data RSE25, RSE 26, RSE 29, RSE30. The columns that we have used for our analysis from the data are CurrentIndication_MajorStreet, Duration_Indication_MajorStreet,CurrentIndication_MinorStreet, Duration_Indication_MinorStreet provided information of the color of signals 1 for Red, 3 for Green, and 4 for Yellow, and the Duration indicated for how long the signal was held. Both analyses for major roads and minor roads we to be considered.

3 METHODOLOGY

To understand and analyze our data in a better way we follow following steps.

3.1 Filtering intersection data

The data file has more than 1.5 million records with all intersection data are combined together. To analyze this data and come up with some meaning ful visalization, it is needed to segregate this data first according to different intersection points. We have filter out the 1.5 million records and group it into 6 categories according to its interaction. These six groups are RSE-25, RSE-26, RSE-27, RSE-28, RSE-29, and RSE-30. Following are the number of records that each group contains.

Group	Number of Records
RSE – 25	196141
RSE – 26	257030
RSE – 27	260952
RSE – 28	262216
RSE – 29	250344
RSE – 30	295949

3.2 Conversion of timestamp to datetime

To further analyze the records of each intersection and to get the understanding of overall data it is needed to compareit with some commo ground. Here the timestamp acted as common ground which was used in comparison and in analysis of records. Further this timestamp value is converted to data and time format for better understanding.

3.3 Grouping of relevant data fields

For the coming up with the visuaization of current signal indicators of major and minor street. It became necessary to group the data according to its state. So, for this the each intersection data is further grouped according to their their current major and minor street signal indicator as well as time. This gave the feasibility to analyze the signal state according to time.

3.4 Visualization of data

After filtering out the required data and grouping it together, visualization of this data was started.

4 RESULTS

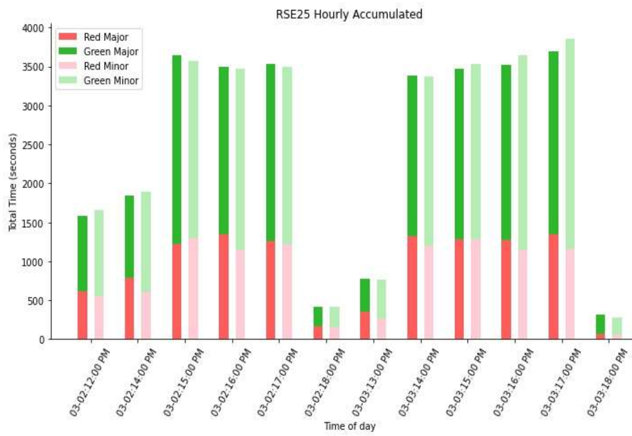


Figure 1: Hourly accumulated signal for RSE-25 major and minor street

Fig. 1 above shows the hourly accumulated signal for intersection RSE-25. Here it can be seen that both street signals are having nearly similar duration of red and green lights.

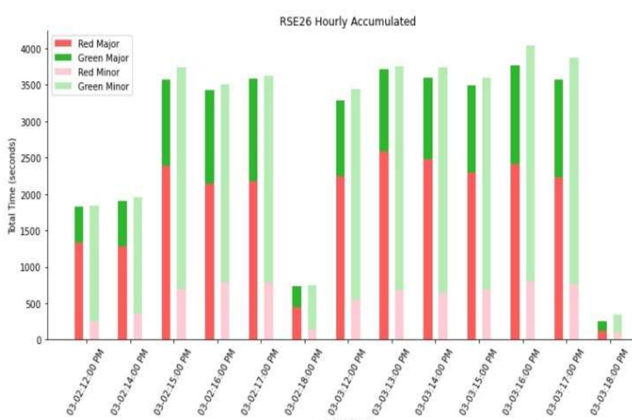
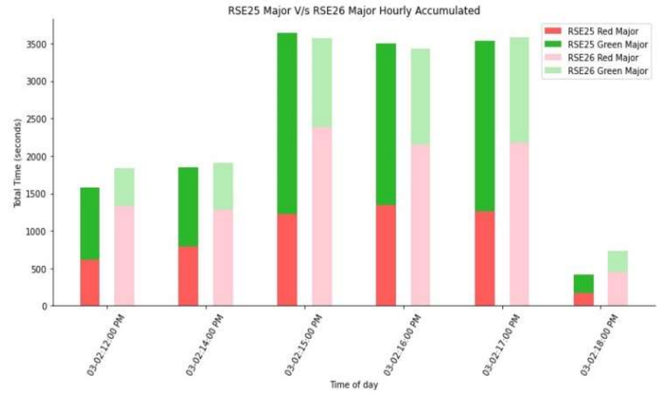


Figure 2: Hourly accumulated signal for RSE-26 major and minor streets

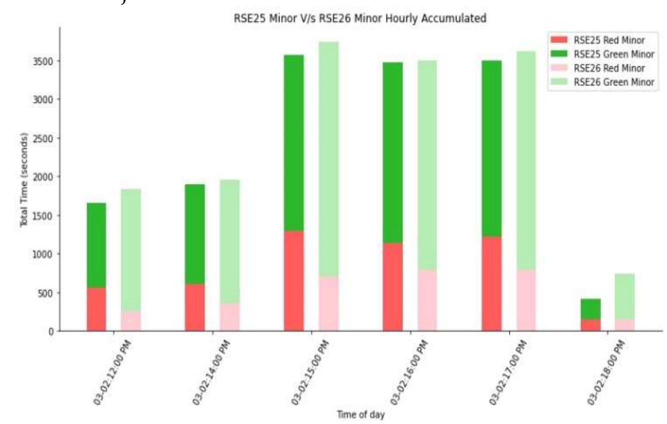
Fig 2. Represents the hourly accumulated signal for intersection RSE-26. Here, unlike RSE-25 intersection the duration of red light signal and green light signal of both streets are not matching. Instead, the red light signal of major street matches with the duration of green light of minor street and vice versa.

By looking at both the above figures it is obvious that both intersection show different behaviour of the duration of signal lights.

As both the intersection shows different behaviour a comparison plot of RSE-25 and RSE-26 was plotted.



(a) Comparison between hourly accumulated of RSE-25 and RSE-26 major streets



(b) Comparison between hourly accumulated of RSE-25 and RSE-26 minor streets

Figure 3: Hourly accumulated comparison of RSE-25 and RSE-26

The comparison between the hourly accumulated of RSE-25 and RSE-26 major streets in Fig 3. (a) shows that the duration of red signal at RSE-25 is similar to the duration of green signal at RSE-26. Both the intersection shows opposite behaviour for red and green signal durations. This can be due to some possible attack on the signal mechanism.Similar to RSE-25 and RSE-26, for this project we also studied RSE-29 and RSE-30.

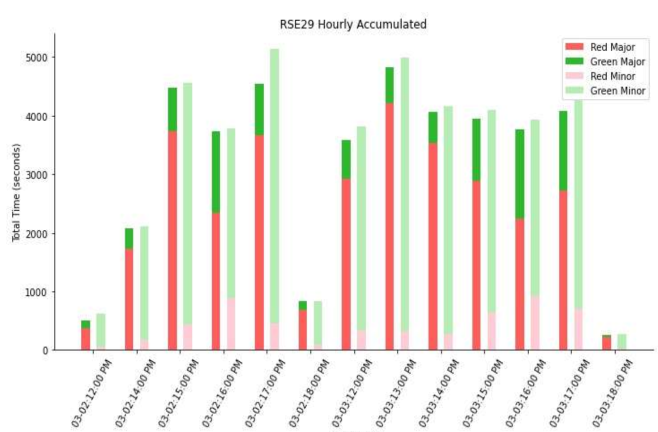


Figure 4: Hourly accumulated signal for RSE-29 major and minor streets

Unlike the other two studied intersections, RSE-29 does not show direct relation between duration of its light of major and minor street. But it can be seen that the duration of red light at major street is somewhat similar to green light of minor street.

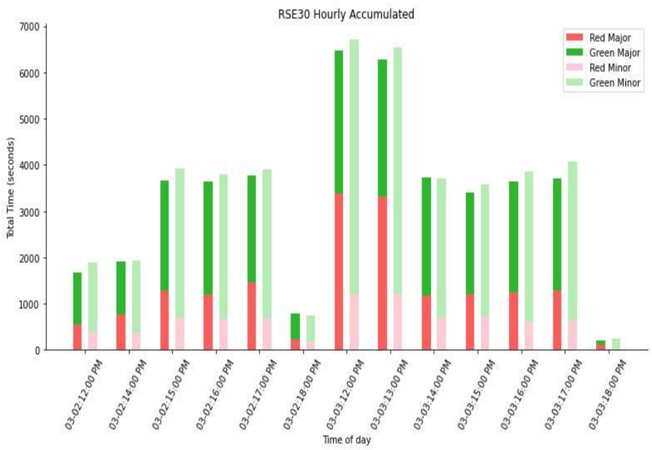
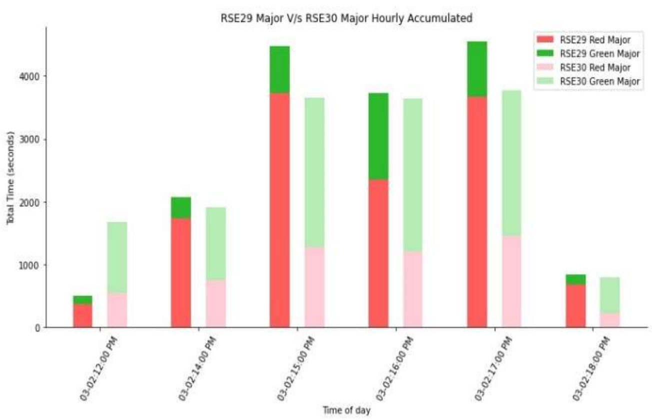
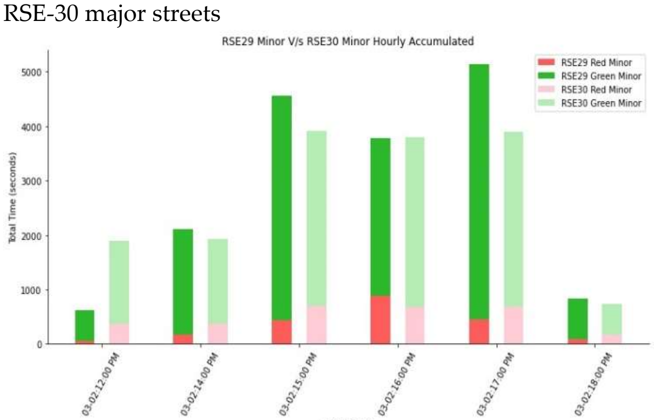


Figure 5: Hourly accumulated signal for RSE-30 major and minor streets

The hourly accumulated of RSE-30 does not have direct or ooposite relation between its major and minor lights that we are seeing till now. But it is obvious that the duration of green light is more than the duration of red light.
Similar to comparison of RSE-25 and RSE-2, there are also comparison of RSE-29 and RSE-30.



(a) Comparison between hourly accumulated of RSE-29 and RSE-30 major streets

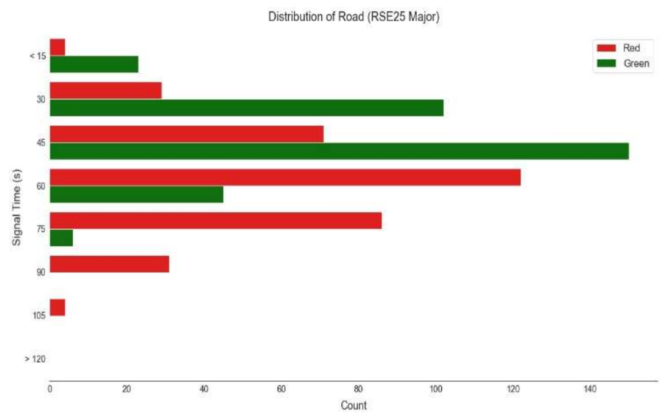


(b) Comparison between hourly accumulated of RSE-29 and RSE-30 minor streets

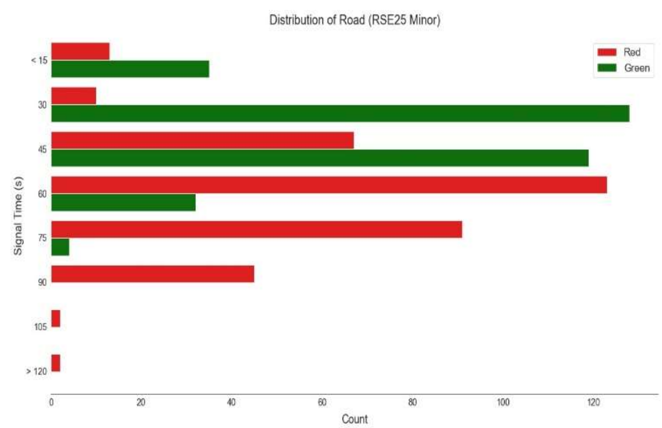
Figure 6: Hourly accumulated comparison of RSE-29 and RSE-30

Fig.6 (a) shows the comparison of hourly accumulated of RSE-29 and RSE-30 major street. This plot shows that the duration of red light at RSE-29 major is more than its green light. While the duration of green light at RSE-30 is more than the duration of its red light. While Fig. 6 (b) the duration of green light is more than the duration of red light. This behavious is observe for both intersection in Fig. 6 (b).

After analyzing the accumulated hours of different intersections as well as their comparisons, now it is required to analyze how many times a particular signal (either red or green) occurred in a time duration and how much each signal lasts.



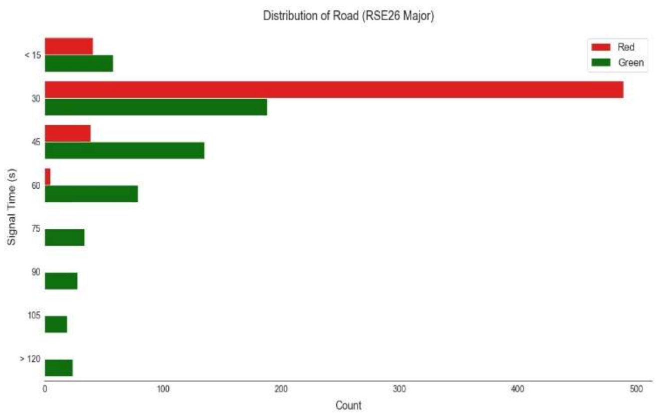
(a) Distribution of roads at major streets of RSE-25



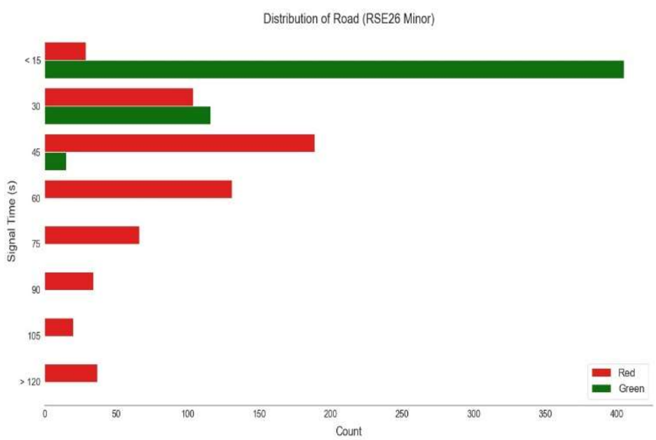
(b) Distribution of roads at minor streets of RSE-25

Figure 7: Distribution of roads at intersection RSE-25

Fig. 7 shows the counts in various bins of intersection RSE-25. From both charts it can be observed that there are more green signals with short time duration then the red signals. There are red signals on both major and minor streets with more duration (>75 s), but those are less in count.



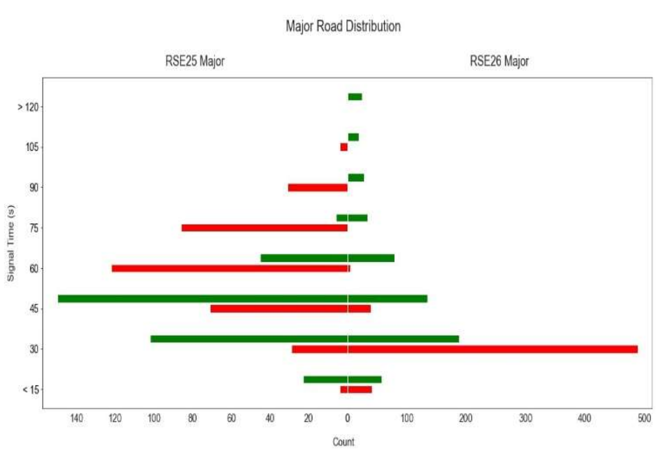
(a) Distibution of roads at major streets of RSE-26



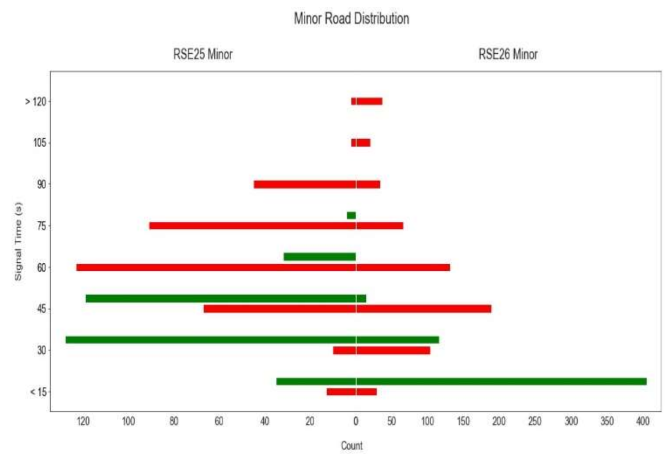
(b) Distibution of roads at minor streets of RSE-26
Figure 8: Distribution of roads at RSE-26

Fig. 8 (a) shows that there are many green light signals then the red ones at the RSE-26 major streets. The duration of these green signals is also going more tha 120 seconds.

Whereas the minor streets have more red signals then the green ones. The green signals are having more counts if they are 30 seconds or less but after that there are only red signals are present.



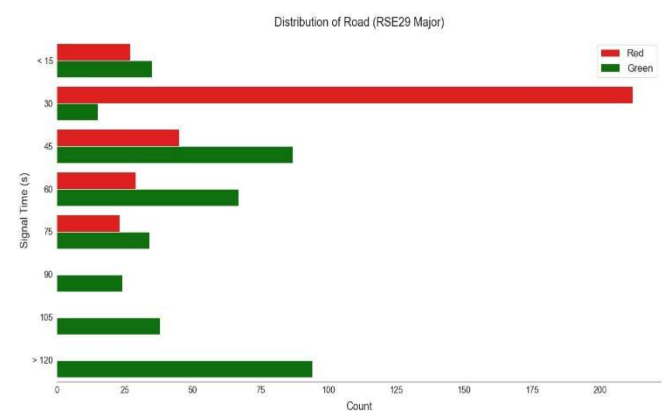
(a) Comparison of road distribution of major street of RSE-25 and RSE-26



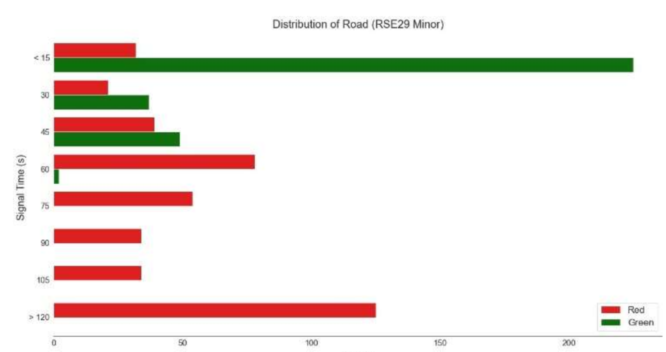
(b) Comparison of road distribution of minor street of RSE-25 and RSE-26
Figure 9: Comparison of distribution of roads at RSE-25 and RSE-26

By comparison we can see that the RSE-25 major street has more red light than the RSE-26 major. This same continues for the minor street as well where count of red signals is more in RSE-25 than RSE-26. RSE-25 also has more green signals of larger duration then RSE-26.

The similar analysis is carried on RSE-29 and RSE-30.



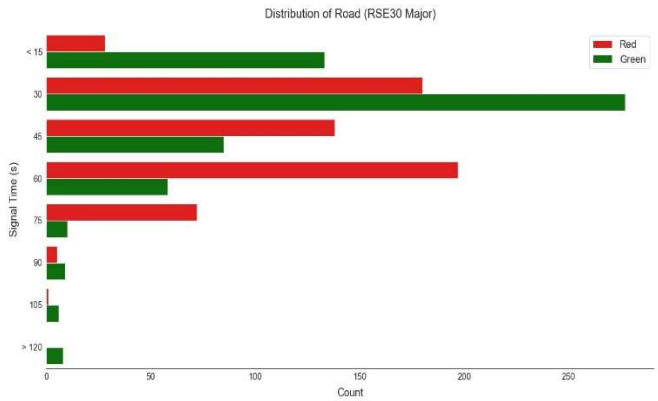
(a) Distribution of roads at major street of RSE-29



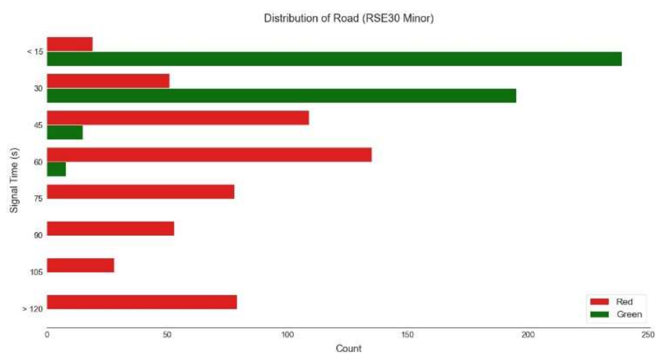
(b) Distribution of roads at minor street of RSE-29
Figure 10: Distribution of roads at RSE-29

Fig 10(a) shows that the RSE-29 major has more green signals than the red ones. Also, there are more count of red signal with duration of 30 seconds then any other signals.

Fig 10(b) indicates that for minor street in RSE-29, there are more red signals then the green signals. Also, unlike major street, minor street has count of green signals more for lower duration signals.



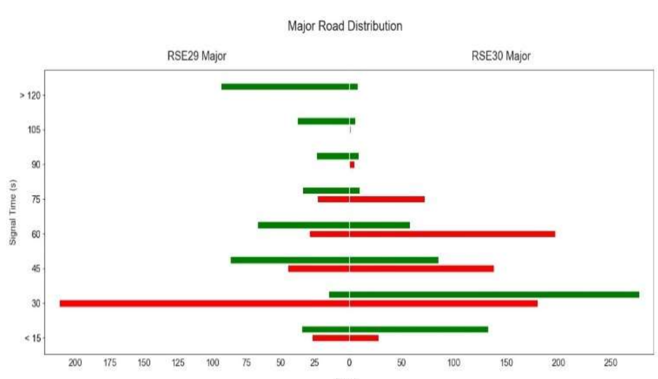
(a) Distribution of road of major street at RSE-30



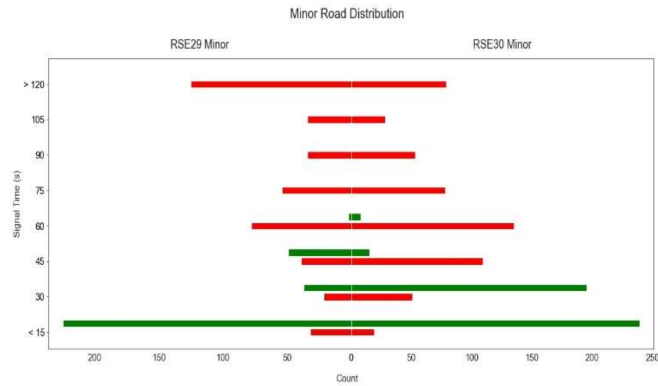
(b) Distribution of road of minor street at RSE-30
Figure 11: Distribution of roads at RSE-30

In RSE-30 major street, there are more number of green signals then the red (Fig. 11(a)), also the larger duration signals are also green one and the are very few large duration red signals.

For minor street (Fig. 11(b)), there are only lower duration green signals. For larger duration signals street only has red signals and no green ones.



(a) Comparison of distribution of roads of major street at RSE-29 and RSE-30



(b) Comparison of distribution of roads at minor street for RSE-29 and RSE-30
Figure 12: Comparison of distribution of road for RSE-29 and RSE-30

By comparing major road of RSE-29 and RSE-30 (Fig. 12(a)) it can be observed that the RSE-29 has more red signals of lower duration while for larger duration it only has green signals. While for RSE-30, green signals are only available for smaller duration, they are not present for larger duration signals.

Whereas for minor street of RSE-29, it is has opposite behaviour then its major streets. It has higher count for lower duration green signals while it only has red signals with more duration. Having so many red signals for so larger duration can be result of compromised system.

While studying the other factor for analyzing the data, we also encounter with the non-uniformity of the green-yellow-red light cycle/occurrence. For many portion of the data this transition is not uniform and its changes due to some unknow reasons.

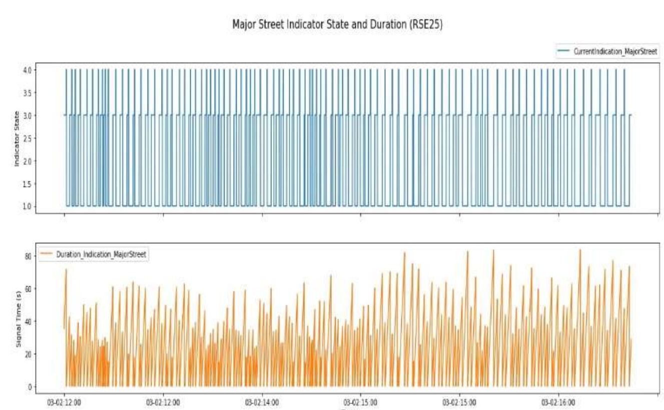


Figure 13: Non-uniform signal transitions

From Figure 13, it can be observed that the transition of signals takes non-uniform time. For some part of the region the signal transition happening fast while for some other time it is bit slow which can be observed by the sparse placement of the lines in upper plot.

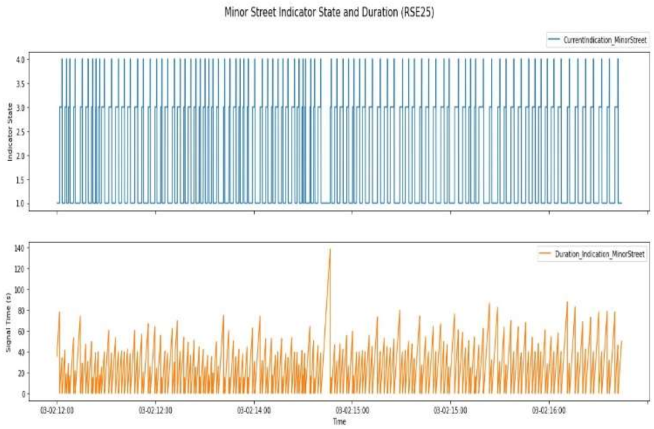


Figure 14: Minor street state and its duration

Fig. 14 represent plots between the state of minor street and duration that it taken. From this figure it can be obseverd that the for nearly whole time the signal time is unforma and same signal time was taken except for time near 15:00. The sudden spike in signal time shows that the signal state 1 (red) has taken more time then usual. This is the largest time taken bu any signal state. This can be possible outcome of attack on signal mechanisn at that time.

5 Conclusion and Future Work

Based on our data, we saw longer Red lights and shorter Green lights on large and minor roads, however, there is no evidence to support the explanation of why this occurred. Further research into cycle times may offer a clearer picture and produce better outcomes. The existing method only examined two crossroads; we could compare numerous intersections and seek interesting results. and was applied only on the traffic light data. In the future, we take the data on a larger scale and understand the causes of such abnormalities too.

References

[1] Glenn T, Guoning C, Yunpeng Z, "A Visual Analytics Approach for Anomaly Detection from a Novel Traffic Light Data"

[2] Ghena, B.; Beye, W.; Hillaker, A.; Pevarnek, J.; Halderman, J. Green lights forever: Analyzing the Security of Traffic Infrastructure WOOT'14 Proceedings of the 8th USENIX conference on Offensive Technologies

[3] Ernst, J.; Michaels, A. (2017) Framework for Evaluating the Severity of Cybervulnerability of a Traffic Cabinet Transportation Research Record 2619(1), 55–63 [Online] Available: https://doi.org/10.3141/2619-06

[4] Keim D A, Kohlhammer J, Ellis G, Mansmann F. Mastering the Information Age: Solving Problems with Visual Analytics. Florian Mansmann, 2010

[5] Etoty, R; Erbacher, r (2014) A Survey of Visualization Tools Assessed for Anomaly-Based Intrusion Detection Analysis Army Research Laboratory [Online] Available: https://apps.dtic.mil/dtic/tr/fulltext/u2/a601590.pdf

[6] McLendon, M.;Shhead, t.; Wilson, A.;Wylie, B.; Baumes, j. (2010), Network algorithms for information analysis using the Titan Toolkit 44th Annual 2010 IEEE International Carnahan Conference on Security Technology (1-10) San Diego, CA

[7] Keim D.A., Mansmann F., Stoffel A., Ziegler H. (2009) Visual Analytics. In: LIU L., ÖZSU M.T. (eds) Encyclopedia of Database Systems. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-39940-9_1122

[8] https://nakedsecurity.sophos.com/2018/03/08/smart-traffic-lights-cause-jams-when-fed-spoofed-data/

[9] https://towardsdatascience.com/a-complete-guide-to-time-series-data-visualization-in-python-da0ddd2cfb01

[10] https://www.openriskmanagement.com/21_ways_to_visualize_a_timeseries/

[11] https://catalog.data.gov/lt/dataset?tags=arizona&organization=dot-gov