COS 720 Assignment Report

## Group 3

Attributes Identification

The following are the attributes that we identified to detect any malicious and abnormal behaviours and the reasons why we chose these attributes.

Message-ID

Message-ID can be used to check if and email has been spoofed by checking if it has been repeated or if it is valid.

Date

A fake timestamp can be used by hackers to try make their fake email look like it was sent at a certain time.

From

An email can be spoofed to look like it was sent from a certain person when it was not.

To

An employee could send sensitive information to an email that looks similar to a valid work email but in reality is different and could be received by someone outside the company.

Subject

The subject could reveal what the email is about, which could possibly violate company policy.

Content-Type

Some content types can be used to send malware which could infect a user’s computer when the email is opened.

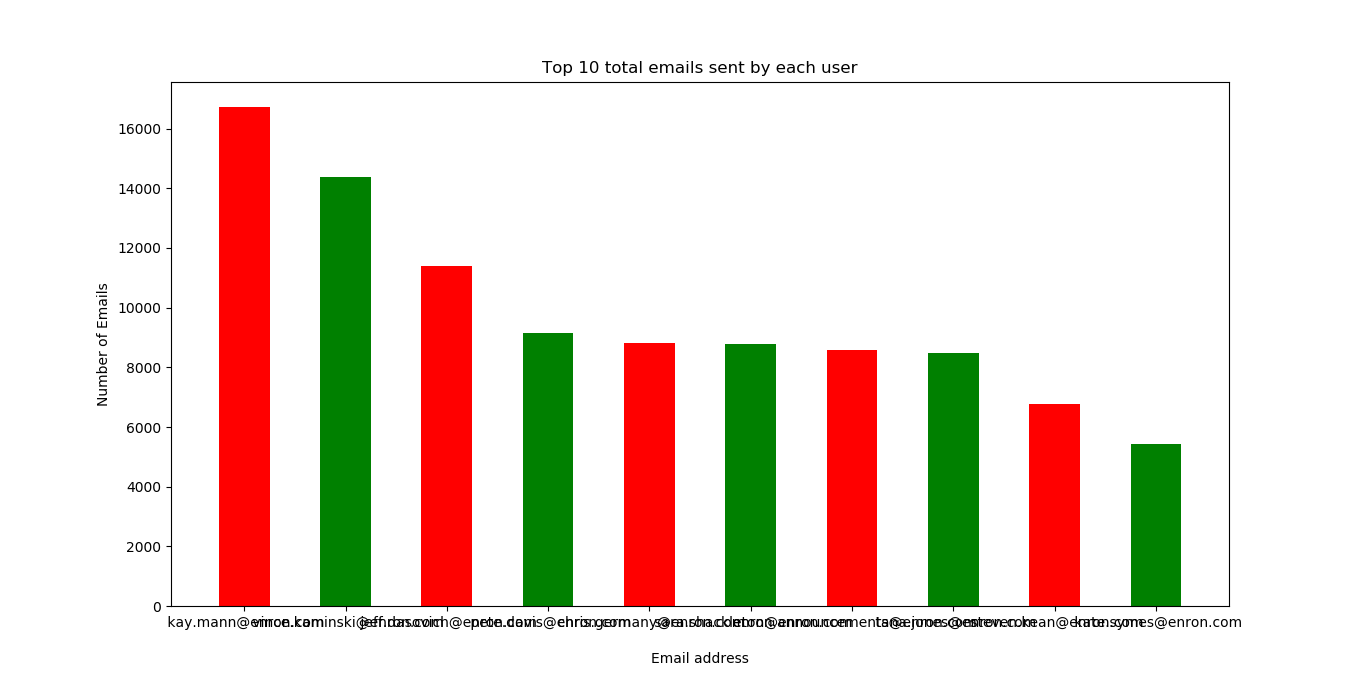
X-from, X-To and X-Origin

The X-headers are made by the Enron email server and are completely trustworthy. They can be used to compare the To and From fields in the email header to check if an email has been spoofed.

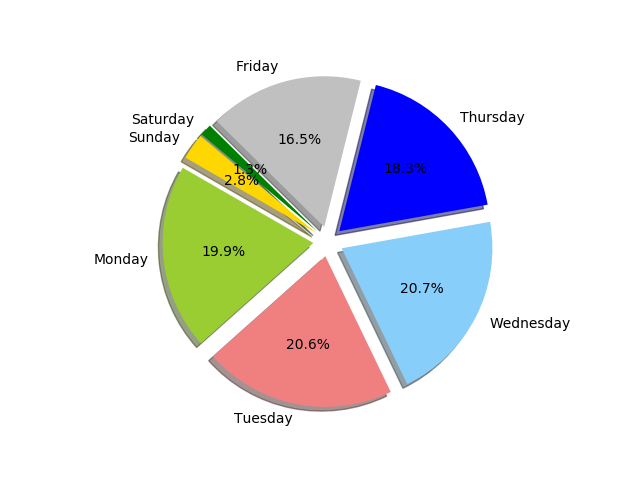
X-Filename

The .nsf file that is used for the X-Filename header is known to be used insert documents that are infected with malware.

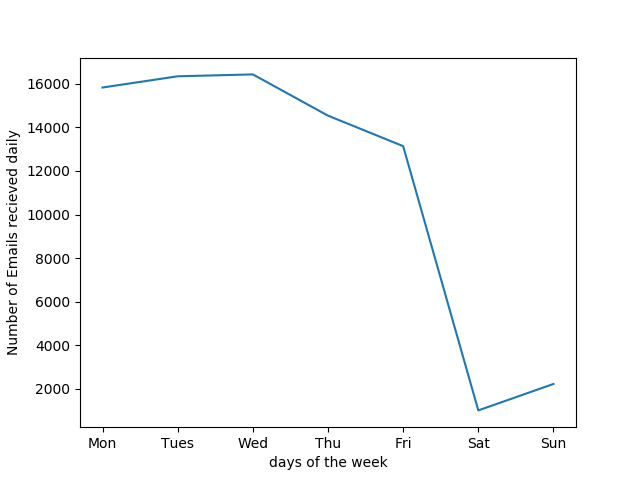
Exploratory data analysis

****

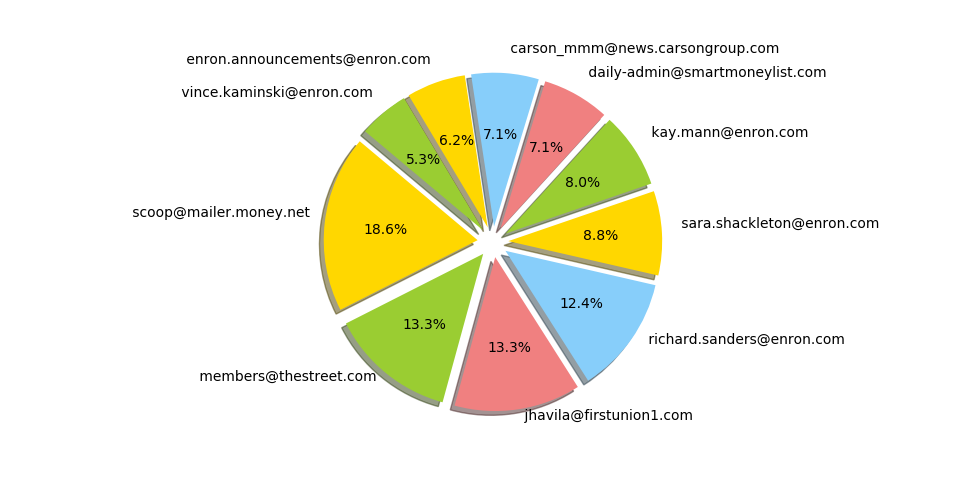
This graph show the top ten enron email addresses that sent the most emails.



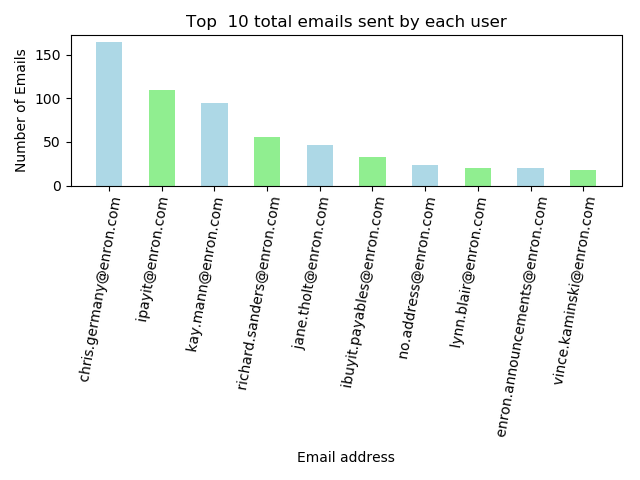
This pie chart shows how emails were received during the week. They start off to increase gradually from Monday and they peak on Wednesday. Friday less emails were received during days.

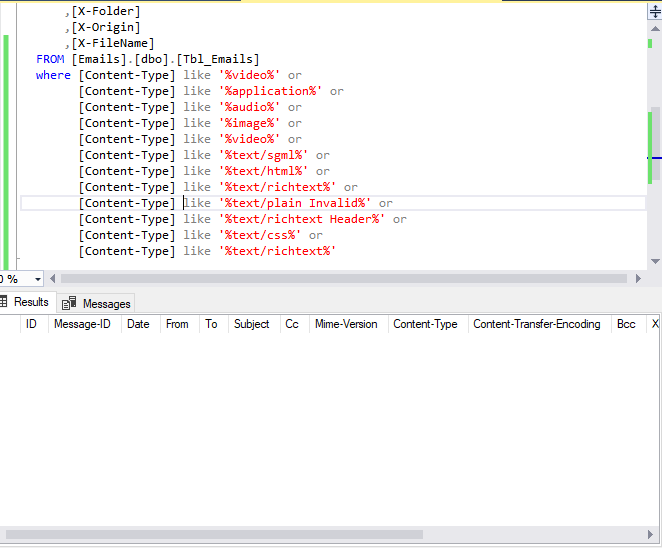


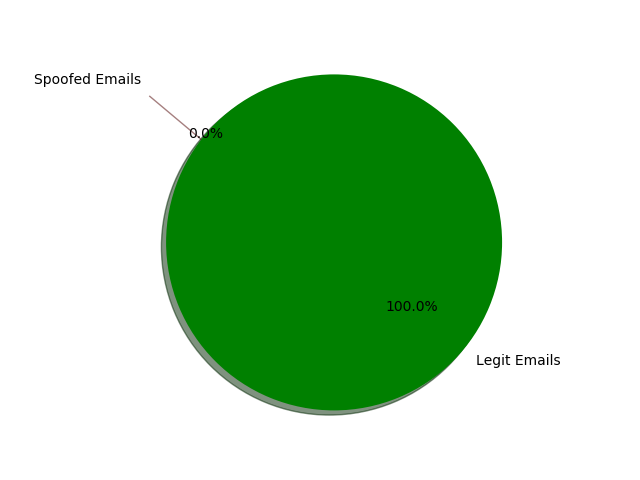
Feature Engineering



The pie chart shows the emails addresses that frequently use the word “Money”, which is a spam keyword, in their subject. The top 3 emails indicates signs of phishing attacks.



Invoices



Cyber Criminal Profiling

One  traits that constainlty appear in email