Data Analysis Lab 4: converting pcap files to csv, and analyze using Pandas

Justin Sallese

April 14, 2017

Panda Code Origin

Github Code

Purpose:

The purpose of this report is to determine the location of the orgin of the emails that are harrasing the proffesor.

Preperation

```
setwd("~/pcap")
library(readr)
nitroba <- read_csv("nitroba")
head(nitroba)
```

```
## # A tibble: 6 × 7
##     No.     Time       Source  Destination Protocol Length
##   <int>    <dbl>        <chr>        <chr>    <chr>  <int>
## 1     1 0.000000 192.168.1.64 74.125.19.83      TCP     70
## 2     2 0.008450 74.125.19.83 192.168.1.64      TCP     70
## 3     3 0.019619 192.168.1.64 74.125.19.19     HTTP   1421
## 4     4 0.044170 74.125.19.19 192.168.1.64      TCP     70
## 5     5 0.224402 74.125.19.19 192.168.1.64     HTTP   1284
## 6     6 0.226712 192.168.1.64 74.125.19.19      TCP     70
## # ... with 1 more variables: Info <chr>
```

Data Description:

The first column is the packet number.

The second column is the time since the start of the data capture.

The third and fourth column is the source and destination IP address.

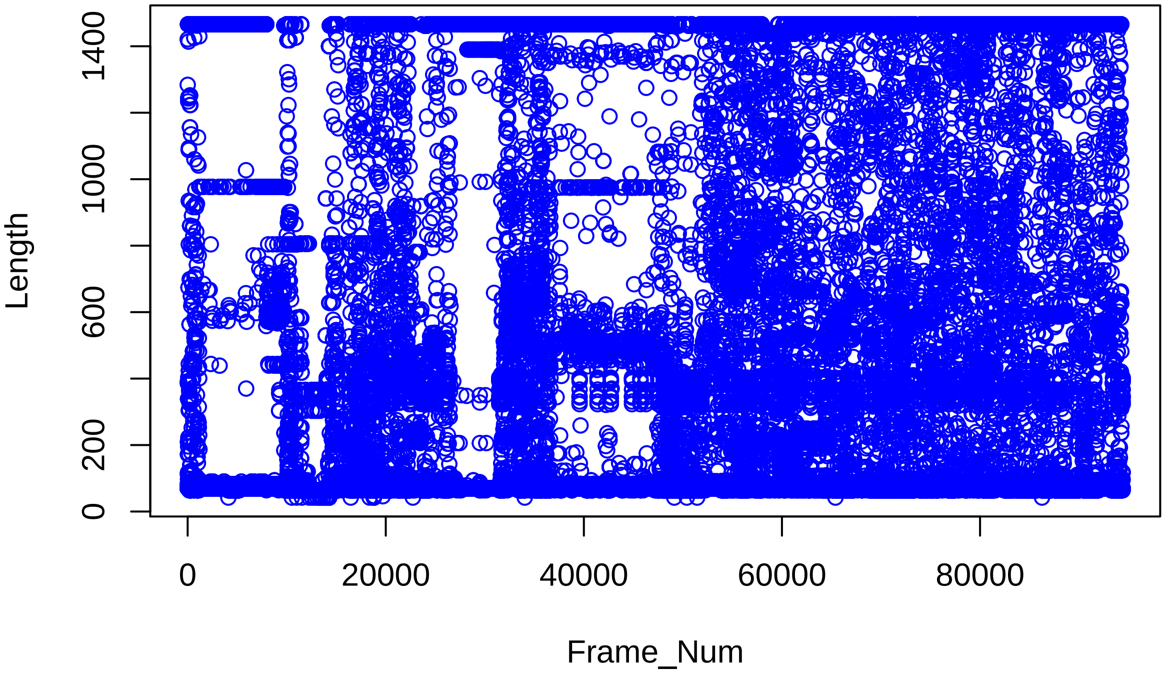
The fifth column is the packet protocol type.

The sixth and seventh column are the length of the packet and the extra info of the packet.

Nitroba University Harassment Scenario

Packet Length Distribution

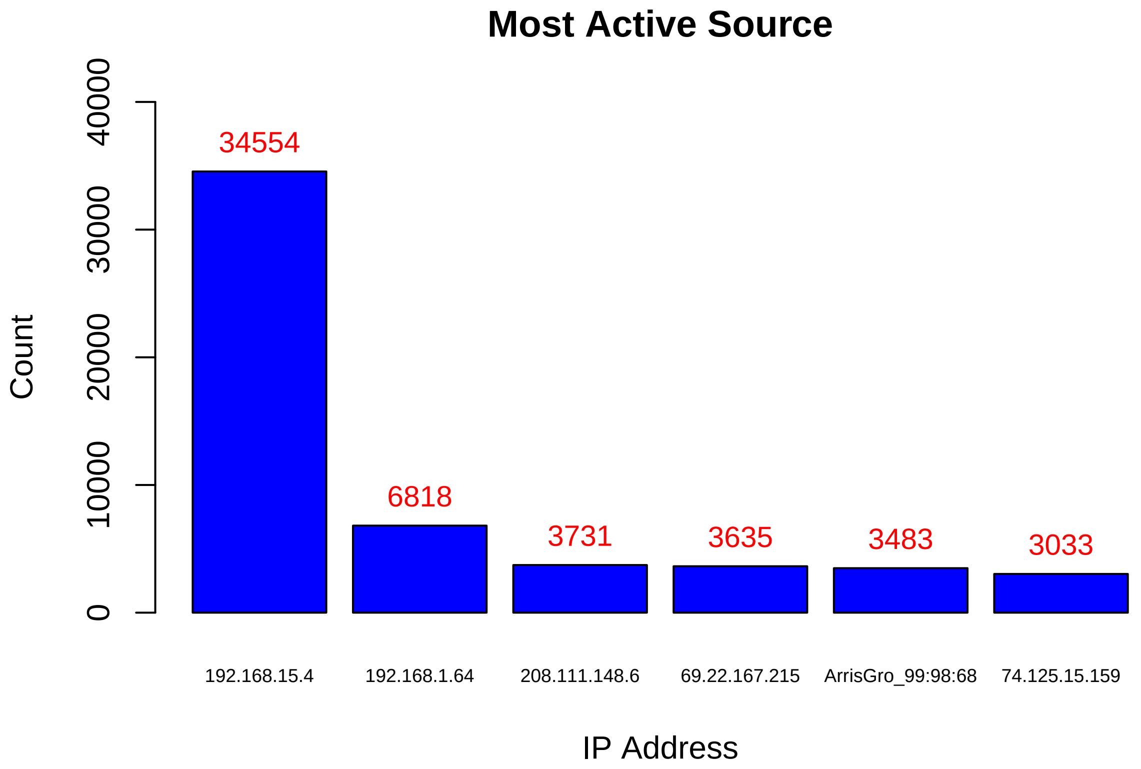
```
Length = nitroba$Length
Frame_Num = nitroba$No.
plot(Frame_Num,
     Length,
     col = "BLUE")
```



The above plot shows the distribution of the framelength throught the progression of the entirety of the data. This plot shows that there is a quite of bit of data above 1400, some below 400 and quite a bit below 200. What this can tell us is that for certain types of packets have certain fixed lengths, so if one knows the length of the type of packets you can know what type of packets where used at a glance. This plot is the same as the one genderated in the pandas code but with R code.

```
source = data.frame(head(summary(factor(nitroba$Source))))
colnames(source) <- c("count")
 q =barplot(source$count,
        main = "Most Active Source",
        ylim = range(0:40000),
        names.arg = rownames(source),
        col = "BLUE",
        width = 1,
        cex.names = .60,
        xlab = "IP Address",
        ylab = "Count")

 text(x = q,
     y = source$count,
     label = source$count,
     pos = 3, cex = 0.9, col = "red")
```

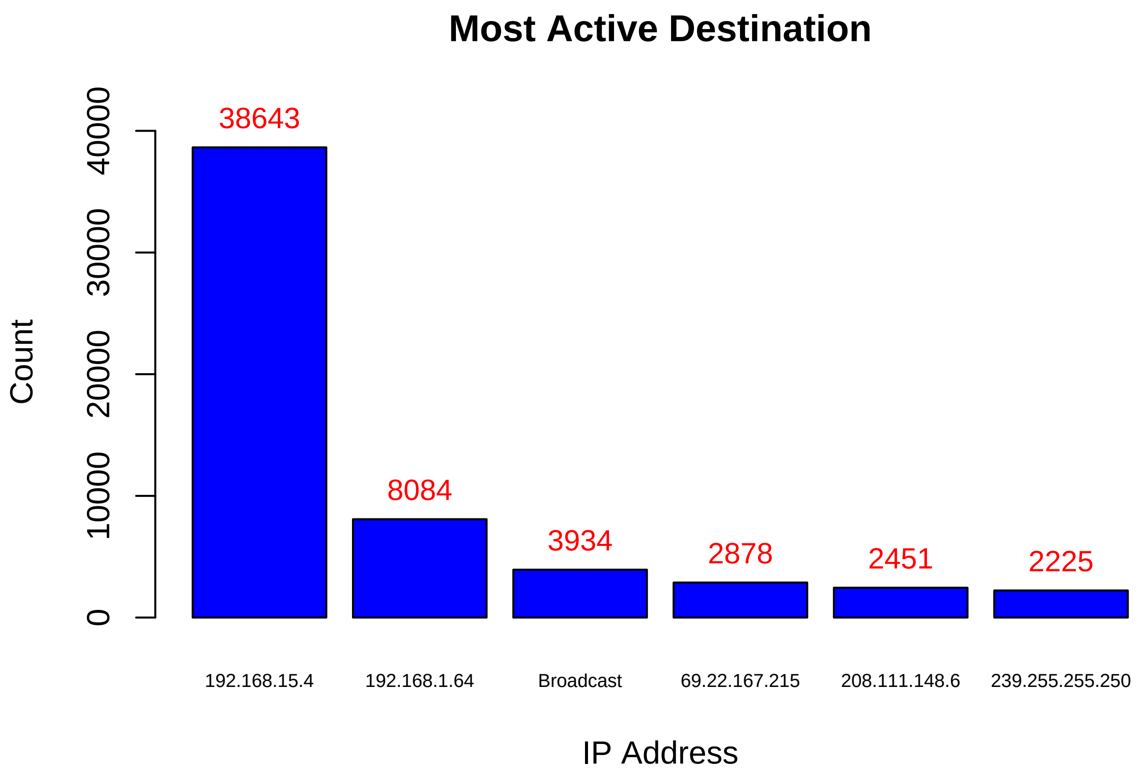


Above is a graph of the top six most active source IP addresses.By looking at the first two source IP addresses if the data is being captured during the time that the malicious emails are sent the first two most active source IP addresses are probably the attacker and the victim.

```
destination = data.frame(head(summary(factor(nitroba$Destination))))

colnames(destination) <- c("count")
 q =barplot(destination$count,
        main = "Most Active Destination",
        ylim = range(0:42000),
        names.arg = rownames(destination),
        col = "BLUE",
        width = 1,
        cex.names = .60,
        xlab = "IP Address",
        ylab = "Count")

 text(x = q,
     y = destination$count,
     label = destination$count,
     pos = 3, cex = 0.9, col = "red")
```



Above is the most active destination IP addresses. By looking at the most active distination IP addresses and comparing them to the most active source IP addresses one can see that the first two most active IP addresses are the same. So if this where an official report to the proffesor that was getting harrased then the next steps that I would reccoment would be to first find out what was the proffesors IP address at the time and then investigate the next possible IP address and to gather data on the location and workspace of the prepetrator to catch them easier.

Conclusion: Assuming that the email harrasment has happended during the data capture the attacker either has an Ip address of 192.168.15.4 or an Ip address of 192.168.1.64. Furthur investigation other applications such as nmap and network miner are required.