Monday

Put together a plan

Waited around for data

Did a lot of research about the topic

This is a commonly known problem. There are multiple organizations that sell products to help prevent others from spying on network traffic. Most of them send out a uniform stream of packets. This prevents people from noticing an increase in packet generation. This is not exactly what we are looking for. We are looking for associating two different devices according to similar network traffic.

Our hypothesis is that devices that visit the same website are associated. The idea is that people working together would be visiting the same websites. Groups would have some form of repository set up that they are visiting. They would also have some form of communication going between them. The problem with this is that most people have multiple devices, and some of those devices generate distracting noise.

We started working on a program that will sort the packet information so that we can better determine patterns. The program will look for all the animals that are visiting the same websites. It will output an animal name and a list of all the animals that visited the same websites as it along with the number of times those animals visited the same website.

Tuesday

Finished sorting program

It is very complicated. We had to use a nested loop with four levels of iteration. It runs a two thousand line file in about five seconds. We have noticed that Alligator and Elephant appear to be associated, and cat goes everywhere.

We think that elephant and alligator might be us. They are heavily associated and there are only two of them. Alligator was not very active today, and I did not use my computer during the period that the collection took place. Elephant was still as active today as yesterday. Elephant is also associated with Dog, Cat, Lizard, and Opossum, so Elephant could be another group. Also, Bethany was not here today, so that could account for an active user that is not active any more.

Thursday

Found graphing software

Figured out how to weight edges

Graphs turned out to be useful

Preliminary evaluation

6 groups scored over 75%

2 groups scored about 50%

2 groups scored nope

1 group scored ?????

Narrowed down which groups did what

Friday

Reevaluated relationships

Gave up on an attempt at an adjacency matrix

Tried to get the graphs to display better information

Regrouped animals

Came up with final answer

Method

Gather data

Filter out duplicates and other columns that aren’t external and internal IPs

Run filtered file through python script that adds up all the times an animal goes to the same place as another animal and outputs it to a text file

Run text file through another script (Hazen Script) that formats that information for the graph software (Graphviz)

Run through graph software

Look for clusters in the graph

Clusters represent groups of devices

List off groups

Repeat multiple times for multiple data sets

Compare listed groups to determine the true groups

Figure out how many times a device is listed in the same group as another device

Higher correlating devices belong together

Put together final list of groups

Final groups decided

Catfish

Squirrel

Swan

Dog

Bat

Alligator

Elephant

Lemming

Cougar

Fox

Camel

Bear

Pelican

Snake

Anaconda

Stork

Tortoise

Wombat

Lynx

Dragon

Crow

Antelope

Porcupine

Baboon

Anteater

Rat

Ant

Goose

Opossum

Otter

Lizard

Cat

Crocodile

Parrot

Badger

Ferret

Stork

Lion

Eagle

Hare

Wallaby

Sheep

Sea lion

Spider

Owl

Cheetah

Cow

Hyena

Possible defenses

We noticed that we had a rather difficult time due to there being an excess of extra devices (phones/tablets). We had a hard time figuring out if groups were part of a project or just someone’s devices doing automatic things. With this in mind it might be possible to use an excess of extra devices to generate distracting traffic. This will make it more difficult to find the important stuff. You could also possibly send your data around to multiple different places rather than sending it all directly to where you want it to go. This will make it difficult to identify relationships between external IPs. You could also use a VPN. This will hide the external IP. The problem with this is that we think a VPN might be breakable. You could potentially fix this problem by sending out a constant steady stream of data. This will make it difficult to find any correlation between anything, but it takes up a lot of bandwidth and data.