Project 1 - Correlation between Streams

1. Project description:

In this project, students are going to develop correlation algorithms of traffic packets time stamps.

By given two streams of packets arrival time stamps, students should implement proper correlation algorithm to calculate the dissimilarity score of the correlation. Two streams from the same connection chain should have low dissimilarity score. Streams from different connection chains should have significantly higher dissimilarity scores.

2. Data sets:

For data in spreadsheet "Correlation Data.xlsx", there are two sheets "part 1" and "part 2". Each sheet has 10 columns of time stamps. For each column, there are about 100 time stamps which indicate the local arrival time of that packet arrived at a host. The starting time of packets arrival at different hosts is local host based.

3. Project goal:

By implementing correlation between two parts, students should find out which streams from part2 are from the same connection chain with streams in part1. Students can evaluate different correlation algorithms, such as OSA, DTW, etc..

Project 2 – Determine whether a connection is from a long chain

1. Project description and goal:

In this project, students are going to investigate connection packets to determine if the connection is from a long chain.

By given different collected packets pcap files, students should inspect original packets first, and try to determine what is the original IP address sending from.

Students should answer if the IP address got above is the real sender's IP address.

Furthermore, students should try to investigate time stamps of those packets and try to utilize those time stamps to determine if those packets are sent from a long connection chain.

2. Data set: Using data2.

Project 3 – Analyze Jittered Data

1. Project description and goal:

In this project, students are going to extract sequence of time stamps of packets from the original "pcap" file.

Students should jitter the sequence of time stamps using several probability distribution models (uniform, Pareto, Lognormal, etc.).

Students should print out the histograms of those time stamps before and after jittering. By comparing histograms before and after jittering, what conclusion can be made based on difference?

Furthermore, students should utilize Matlab or other distribution fitting tools to fit those histograms based on certain distribution (such as Pareto or Lognormal) and estimate parameters. Students should analyze and give conclusion based on distribution fitting.

2. Data set: Using any "pcap" file from Data2.

Project 4 – Analyze Chaffed Data

1. Project description and goal:

In this project, similar to Project 3, students are going to extract sequence of time stamps of packets from the original "pcap" file.

Students should add chaffing into the sequence of time stamps using several probability distribution models.

Students should print out the histograms of those time stamps before and after adding chaffing. By comparing histograms before and after, what conclusion can be made based on difference?

Furthermore, students should utilize Matlab or other distribution fitting tools to fit those histograms based on certain distribution (such as Pareto or Lognormal) and estimate parameters. Students should give discussion and conclusion based on distribution fitting.

2. Data Set: Using any "pcap" file from Data2.