Computer Networks

Assignment 4: Internet Measurement

Overview

- You will be analyzing public data to observe patterns and properties of the internet.
- You can use any programming language to analyze the data given to you for this assignment.
  - Data comes from routers managing traffic for locations and monitoring BGP, stored in CSV files.
- Will be answering questions on the metrics, graphs, and patterns your asked to generate or find and will compile your solutions in a PDF.

Why Measure The Internet?

- Understand performance of protocols, topology
- Maintenance and repairs
- Billing

Traffic Measurement

- Data collected using Netflow in routers.
  - Allows routers to examine traffic as it comes in.
  - Does so by aggregating packets into flows.
  - Data about the flows can be saved which can be used to profile traffic and measure things like congestion

Understanding Netflow Data

Date first seen	Time first seen (m:s)	Date last seen	Time last seen (m:s)	Duration (s)	Protocol
10/29/15	04:48.9	10/29/15	04:48.9	0	TCP

Src IP addr	Src port	Dst IP addr	Dst port	Packets	Bytes	Flags	Input interface	Output interface
116.211.0.90	52704	128.112.186.67	9095	1	40	....S.	120	70

BGP Measurement

- Data is collected using monitoring routers
  - These routers only monitor changes in BGP across the internet (relative to them), do not participate in handling traffic
  - RouteViews is a project that monitors BGP updates amongst different ISPs, and all their sessions' data is publicly archived on their site
- Will be analyzing both a snapshot of BGP Routing Information Base (RIBS, stores routes for destinations IPs) and a series of BGP update messages leading to that snapshot.

Understanding RouteViews Data

BGP RIBs

TIME	NEXT_HOP	FROM	ASPATH	PREFIX	ORIGIN
03/28/21 04:00:00	85.114.0.217	194.153.0.253 AS5413	286 3257 13335	1.0.0.0/24	IGP

BGP Updates

TIME	FROM	ASPATH	NEXT_HOP	COMMAND	ORIGIN
45:00.3	208.51.134.246 AS3549	3549 3356 8820 8820 8820 8820	208.51.134.246	ANNOUNCE 46.41.11.0/24	IGP

CDFs

- Some questions ask for you to create cumulative distribution probability functions (CDFs) of the data.
  - They represent the probability $P(X \leq t)$ where X is the random variable and t is the value we want the probability of X being less then or equal to.
  - Can solve them by using summations (or integrals): $\mathrm{P}(\mathrm{X} \leq t) = \sum_{t_i \leq t} P(X = t_i)$

CDFs (cont'd)

- For example: create a CDF for the following probability distribution

$$P(X = t) = \frac{5 - t}{10}, t = 1,2,3,4$$

  - For each $t_i \in \{1,2,3,4\}$, calculate $P(X = t_i)$
    - $P(X = 1) = \frac{5-1}{10} = \frac{4}{10}$
    - $P(X = 2) = \frac{5-2}{10} = \frac{3}{10}$
    - $P(X = 3) = \frac{5-3}{10} = \frac{2}{10}$
    - $P(X = 4) = \frac{5-4}{10} = \frac{1}{10}$
  - For each $t_i \in \{1,2,3,4\}$, calculate $\sum_1^{t_i} P(X = t_i)$
  - Show distribution of $P(X \leq t)$

CDFs (cont'd)

- For example: create a CDF for the following probability distribution

$$P(X = t) = \frac{5 - t}{10}, t = 1,2,3,4$$

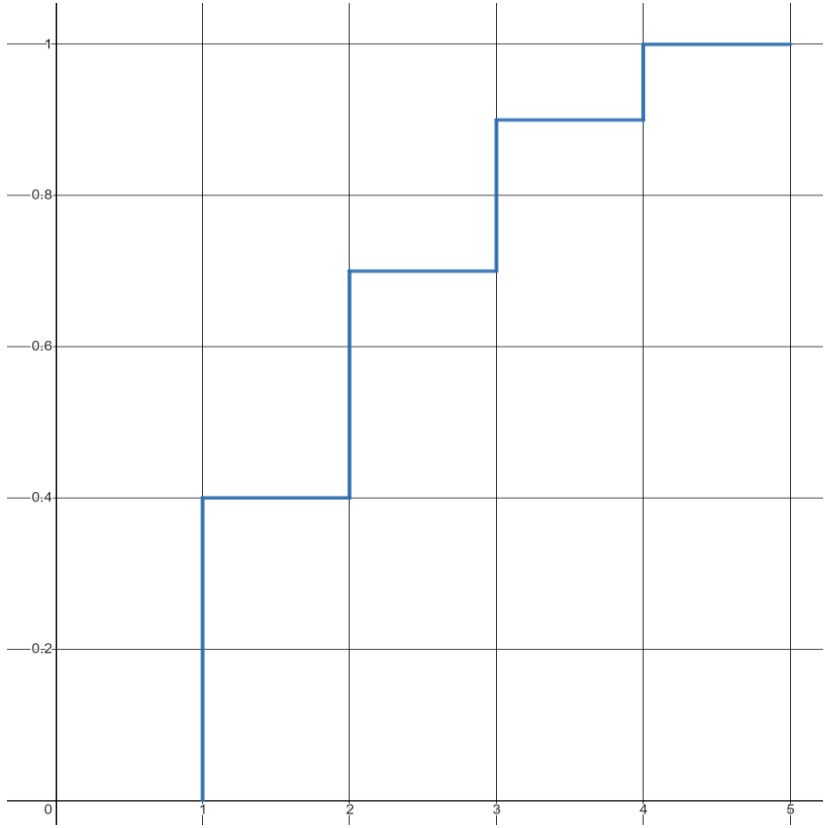
  - For each $t_i \in \{1,2,3,4\}$, calculate $P(X = t_i)$
  - For each $t_i \in \{1,2,3,4\}$, calculate $\mathrm{P(X \leq t_i)} = \sum_{j=1}^{t_i} P(X = j)$
    - $P(X \leq 1) = P(X = 1) = \frac{4}{10}$
    - $P(X \leq 2) = P(X = 1) + P(X = 2) = \frac{4}{10} + \frac{3}{10} = \frac{7}{10}$
    - $P(X \leq 3) = P(X = 1) + P(X = 2) + P(X = 3) = \frac{4}{10} + \frac{3}{10} + \frac{2}{10} = \frac{9}{10}$
    - $P(X \leq 4) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = \frac{4}{10} + \frac{3}{10} + \frac{2}{10} + \frac{1}{10} = 1$
  - Show distribution of $P(X \leq t)$

CDFs (cont'd)

- For example: create a CDF for the following probability distribution

$$P(X = t) = \frac{5 - t}{10}, t = 1,2,3,4$$

  - For each $t_i \in \{1,2,3,4\}$, calculate $P(X = t_i)$
  - For each $t_i \in \{1,2,3,4\}$, calculate $\mathrm{P(X \leq t_i)}$
  - Show distribution of $P(X \leq t)$
    - X-axis is $t_i \in \{1,2,3,4\}$, Y-axis is $P(X \leq t_i)$



What About The PDF?

- Every question should be answered in this document (no credit for answers not here).
- Have a section explaining how your code works, what platform you coded it on, and what are some interesting methods you are using to calculate your solutions (basically a README).
- Answers and observations should be full sentences, written well, clearly answering the question being asked.
- PDF should be formatted nicely.

Suggestions

- Recommend using a programming language that makes manipulation of data, generating graphs painless.
  - Python have a lot of libraries that make these kind of things easy (Pandas, Numpy, Matplotlib).
- Make sure visuals are well-labeled and are easy to understand.

Any Questions?