Estimating Original Flow Characteristics from Sampled Traffic Statistics

Tarem Ahmed and Mark Coates
tarem.ahmed@mail.mcgill.ca, coates@ece.mcgill.ca

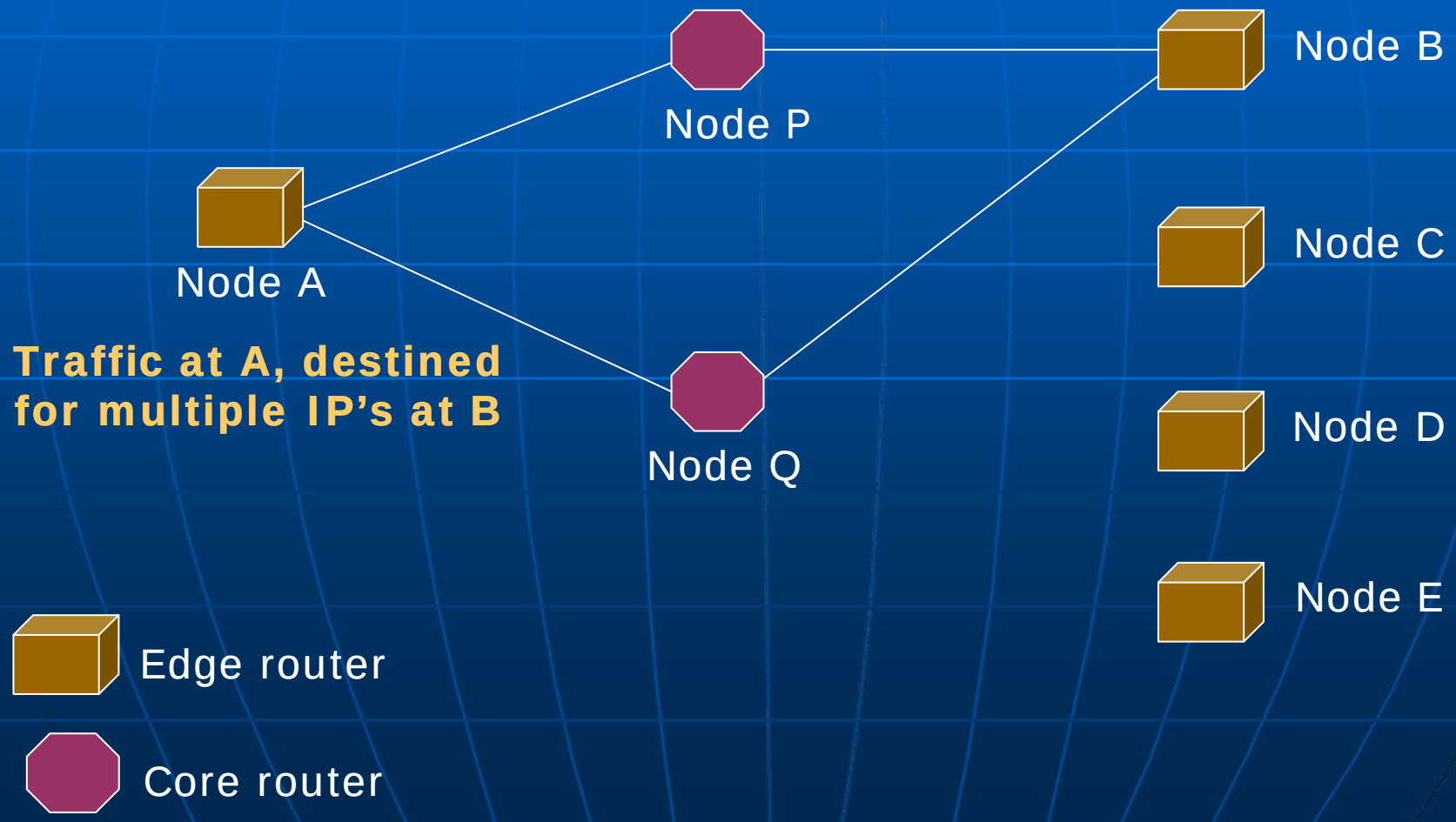
April 07, 2004

Motivation/Application to AAPN

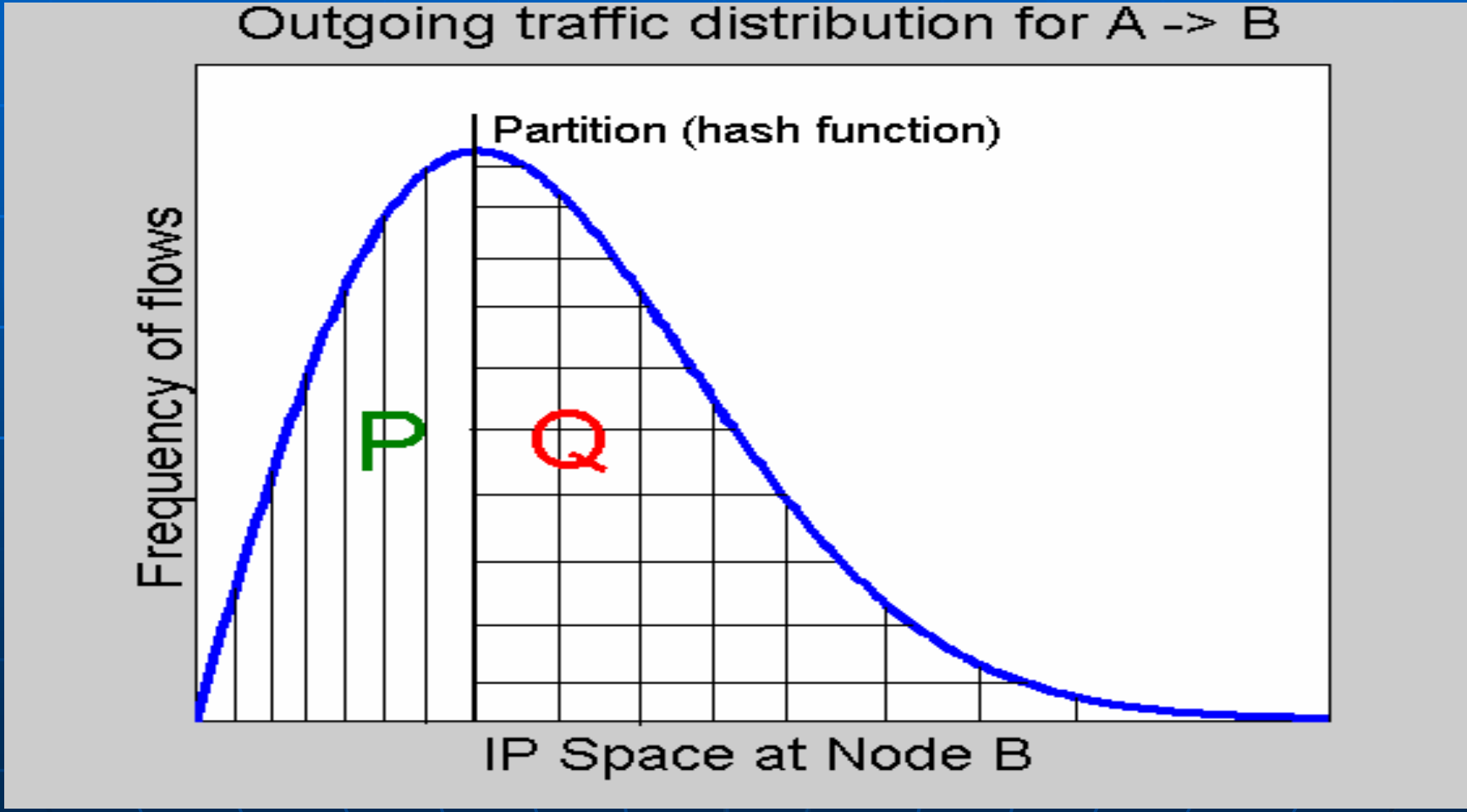
- Routers cannot keep record of every packet
- Thus sampling, and estimation of original flow characteristics, required to
  - analyze utilization of router processing power
  - determine need for web proxies
  - **perform load balancing**
  - **make dynamic routing decisions**

Route Flow Through P or Q?

Load Balancing

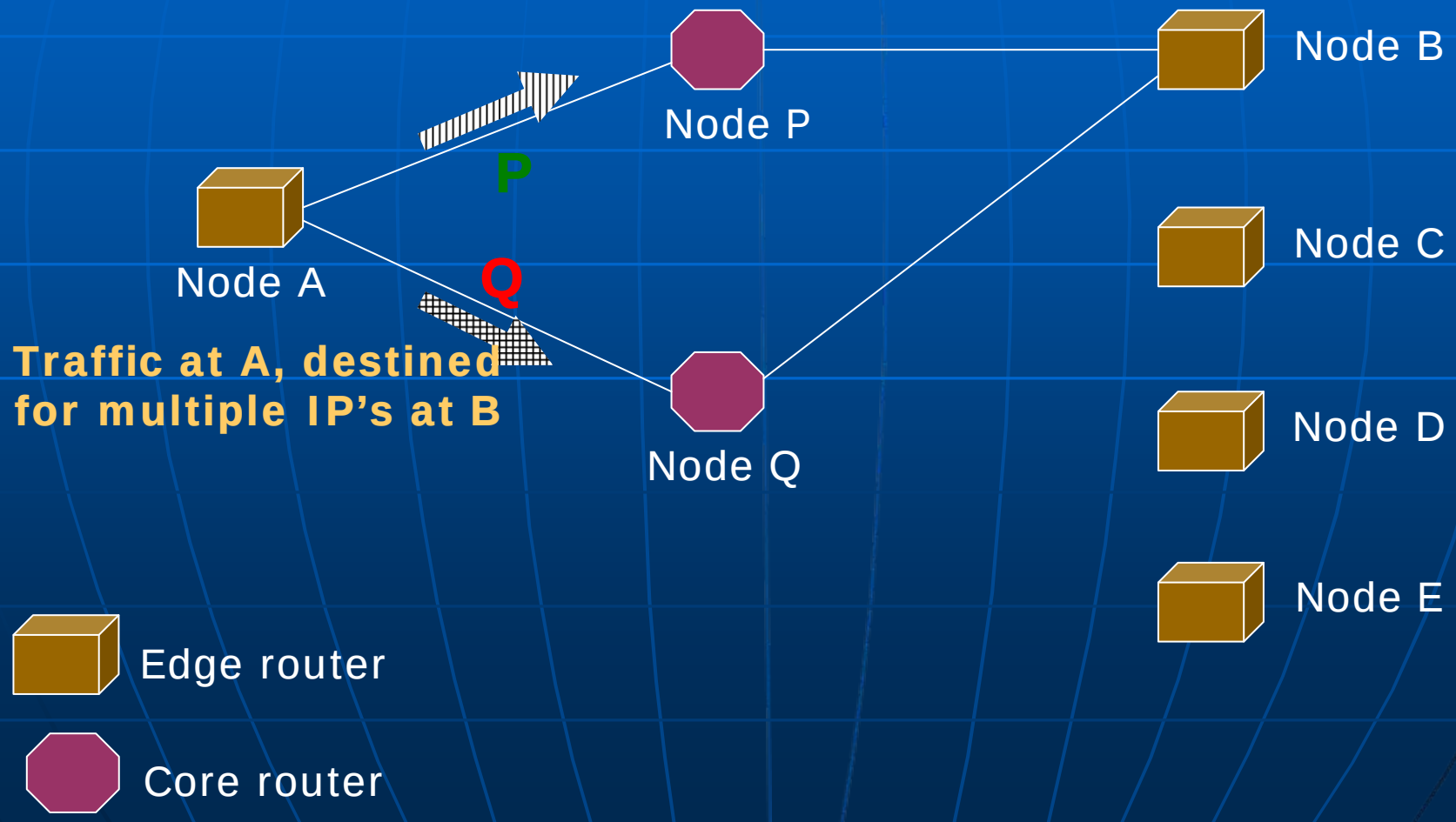


Distribution of Destination IP Addresses



Route Flow Through P or Q?

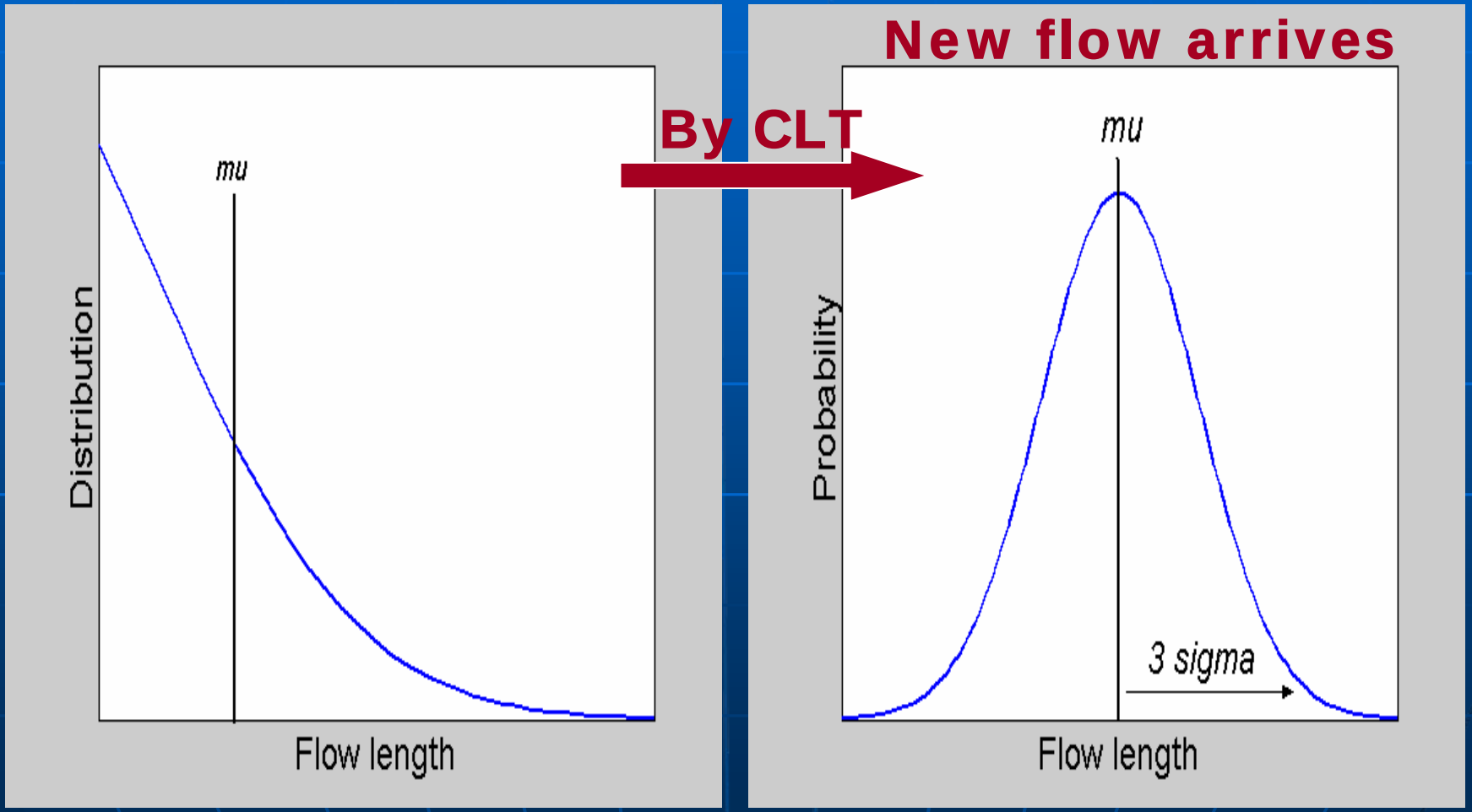
Load Balancing



Motivation/Application to AAPN

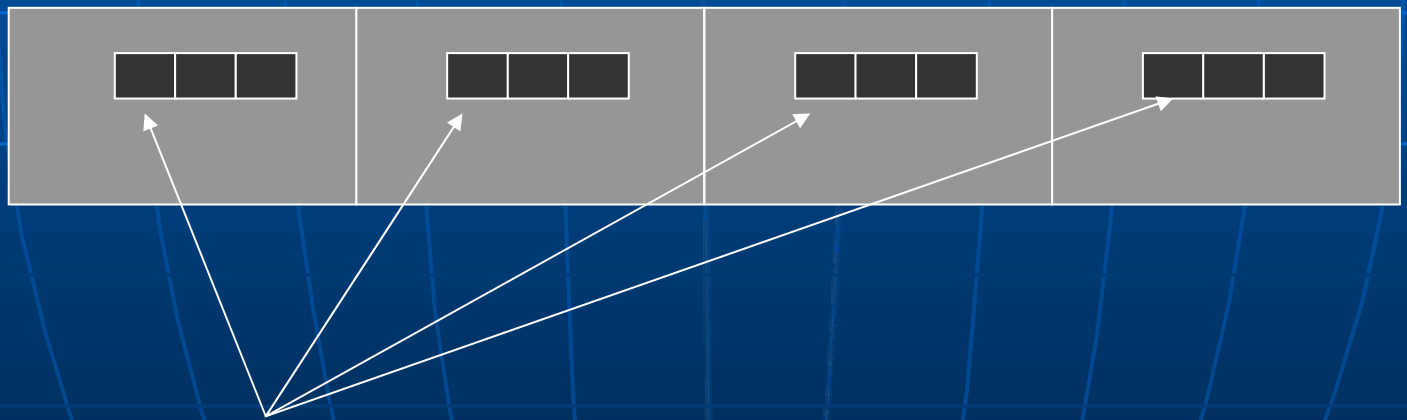
- Routers cannot keep record of every packet
- Thus sampling, and estimation of original flow characteristics, required to
  - analyze utilization of router processing power
  - determine need for web proxies
  - perform load balancing
  - make dynamic routing decisions
  - **reserve resources for new flows**

Expected flow length



Once New Flow Arrives

- Predict length of flow and reserve time slots in future frames, accordingly



mu + 3*sigma* time slots

=> 99% confidence

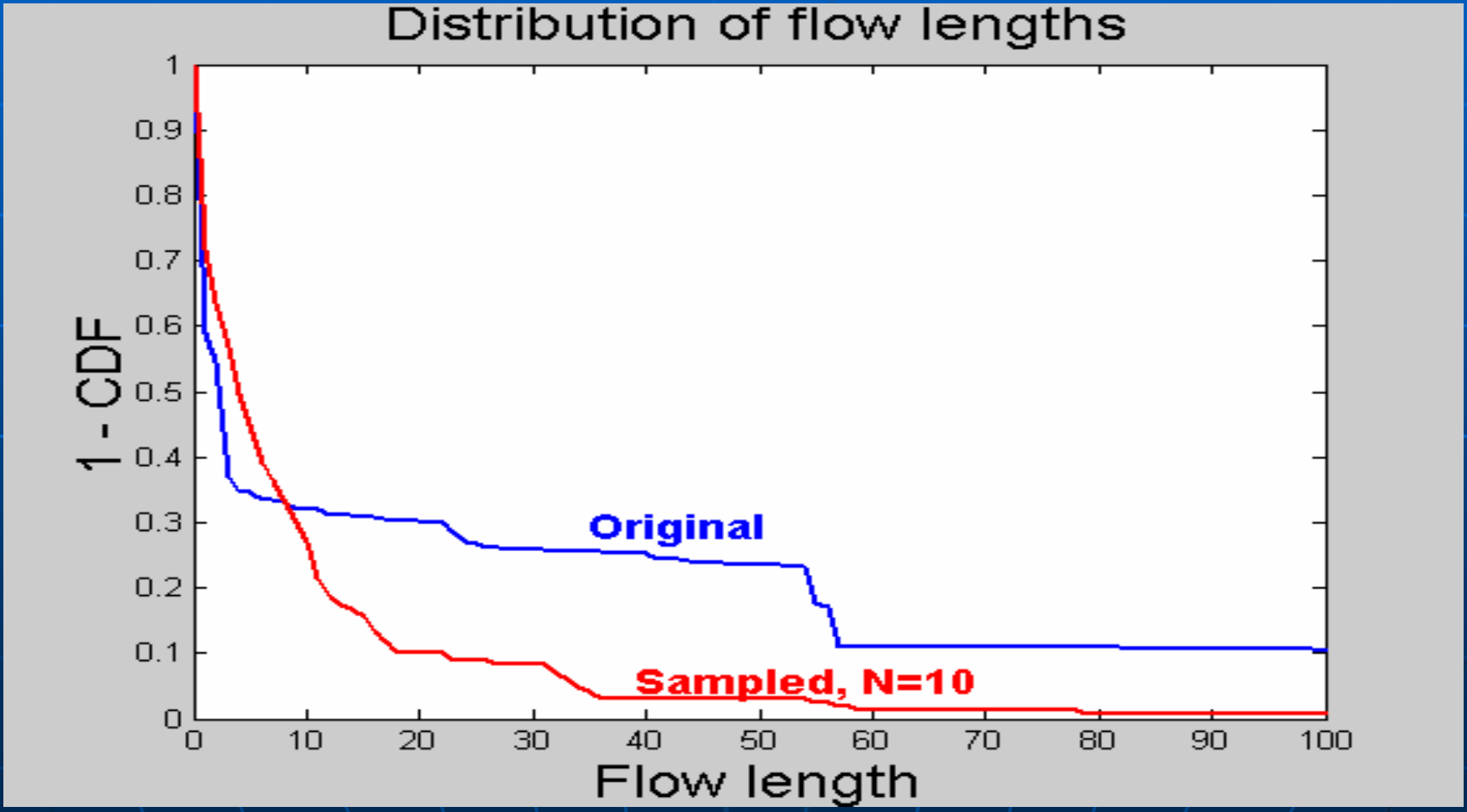
Immediate Research Objectives

- Estimate from a sample
  - the number of distinct flows
  - the number of packets in each flow (ie flow length)
  - the number of bytes in these packets (ie packet length)

Problems/Challenges

- Tempting to:
  - sample every *N*'th pkt
  - attribute original length of *NL* to sampled flow length of *L*
- Problems:
  - some smaller flows are not sampled at all
  - larger flows are more likely to be sampled

Simple extrapolation does not work!



Sampling Strategies

- Periodic sampling of every *N*th pkt
  - introduces sampling correlations
  - biases against selection of multiple closely-spaced packets
- Independent sampling with probability $p = 1/N$
- Multirate sampling: vary the sampling rate over time
  - start with high value of p, and then gradually reduce
  - so as not to lose information contained in smaller flows

Implementation Method

- Assume independent binomial distribution
- Obtain log-likelihood function
- Numerically maximize using Expectation-Maximization algorithm

References

- “Estimating Flow Distributions from Sampled Flow Statistics”, Duffield, Lund & Thorup, SIGCOMM’03, August 25-29, 2003, Karlsruhe, Germany
- http://pma.nlanr.net/Traces