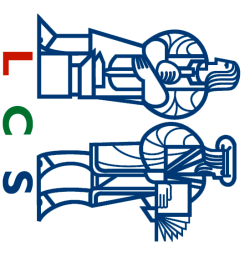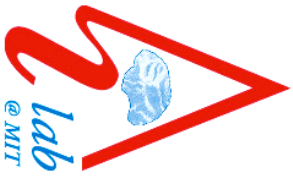Erik D. Demaine

Project Overview

- **Goal:** Gather statistics on network traffic as it streams by router
  — Develop models of net traffic & computation
  — Design and analyze algorithms & data structures
    • Use little time per packet
    • Use little space overall
  — Prove lower bounds

Network Router

Statistics gathering

Limited store

Estimated properties of stream

Confidence

Erik D. Demaine
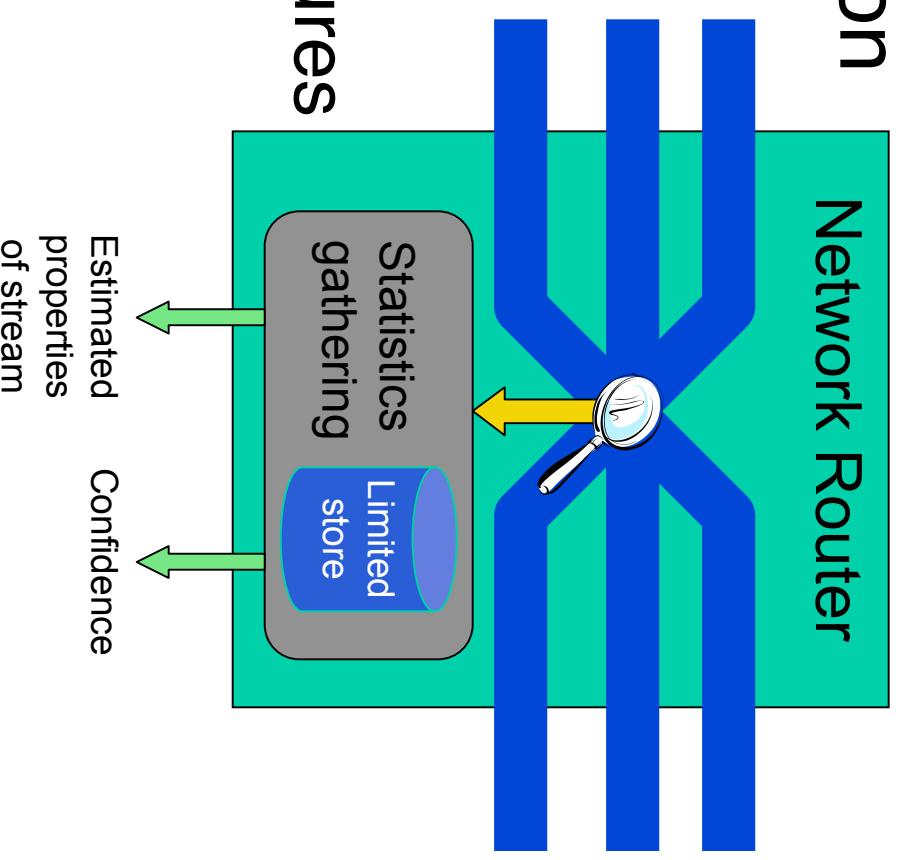
Progress Through June 2002

- Paper published at *10th Annual European Symposium on Algorithms (Sept 2002)* with Prof. J. I. Munro and Prof. A. López-Ortiz (U. Waterloo)

- **Problem:** Find the *k* most popular **categories** of packets (e.g., destination address, source address)

  – Even in the **worst case**, all categories that occur at least ~1/*m* of the time can be identified by a fast algorithm that can record statistics using just *m* counters

  – In the **stochastic model** (arbitrary network traffic but randomly permuted), all categories that occur at least ~1/sqrt(*m n*) can be identified for an *n*-packet stream, and the **frequencies can be estimated to desired accuracy**

  – Both results are the best possible in each model

Erik D. Demaine
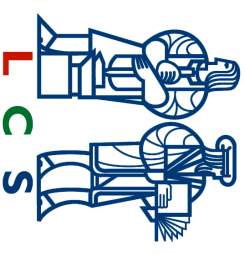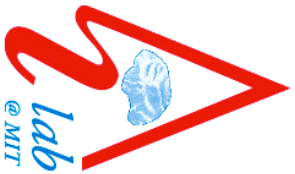
Research Plan for the Next Six Months

- Implement the algorithms to evaluate their performance and effect of models on real data

  – Begun discussions to gather Internet traffic data at LCS

- Key theoretical open problem:

  Can randomization help in the worst case (**oblivious-adversary model**)?

  – We have shown that randomization helps significantly in the average case

  – **Idea:** Take this algorithm and randomly perturb the sample windows to thwart adversary