

In-Network Processing for Streaming Data

Martin Swany

Professor, School of Informatics and Computing
Executive Associate Director, Center for Research in Extreme
Scale Computing (CREST)
Indiana University

Network Logistics

- Logistics - *the management of the flow of resources from the point of origin to the point of consumption*
 - **OED: big data** *n. Computing* (also with capital initials) data of a very large size, typically to the extent that its manipulation and management present significant logistical challenges
- Don't treat the network as opaque
- Latency / proximity matters
- Processing and storage in the network at strategic locations
- Question of router buffering (bufferbloat vs the control loop)
 - How about an Internet *FOR* Things?

Network Measurement

- In trying to make logistical decisions, we measure
- High-frequency data challenges
- Over-summarization loses important behavior
- Windows of data dumped on triggers

In-network Processing

- As the scale increases, in-network processing will matter
- Something between the edge and the data center
- Hierarchical control
- In-network fusion, reduction

Network Processors

- NFV / SDN
- Network processors are arrays of tightly-coupled cores
- The same programmability that allows deep packet inspection implies that we can route and process based on contents
 - Parallelize by routing into parallel pipelines (with e.g. a hash function)

Where and how to program?

- Where is the real network bottleneck?
 - Often at the uplink, not the access layer
- Flange is our language to program network flows
 - SDN but also insertion of processing kernels on network processors or network-enabled FPGAs

