The Complexity of Network Traffic Traces

Chen Griner(BGU)

Collaboration: Chen Avin(BGU), Manya Ghobadi(MIT), Stefan Schmid (UW)

Why should we care about traffic complexity?

- Most network topologies are designed with worst case scenarios in mind.
- Newer datacenter technologies offer online reconfigurability of links.
- For these technologies to be helpful, predictable patterns need to exist in network traffic.
- The more "structure" a trace has, the less "complex" it is.
- We offer a simple way to analyze real and synthetic network traffic to detect complexity.

What might it tell us?

- ❖ Lower complexity means optimization opportunities.
- ❖ Identify and quantify different types of locality.
  - ❖ Temporal locality?
  - ❖ Spatial locality?
  - ❖ Others?
- ❖ Compare different traces?
- ❖ Differentiate between different workloads?

Traffic as a Network Trace

- ❖ Our main building block is the network traffic trace.
- ❖ A time ordered list of source-destination pairs.

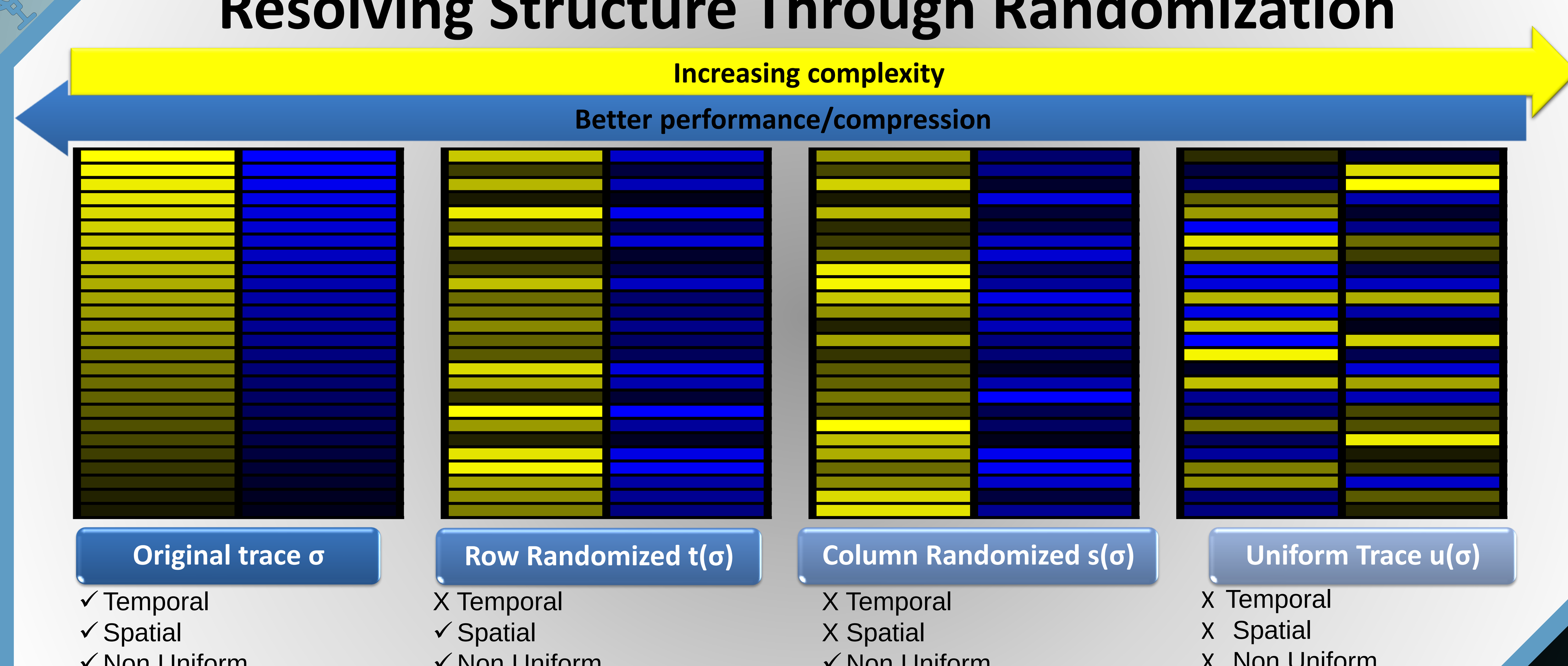
source	destination
192.168.1.3	192.168.1.1
192.168.1.2	192.168.1.5
192.168.1.3	192.168.1.1
192.168.1.5	192.168.1.3
192.168.1.42	192.168.1.59
192.168.1.55	192.168.1.5
192.168.1.4	192.168.1.3

Time

Compression Based Methodology

- ❖ Lempel-Ziv, detects different structures in an information source.
- ❖ Break different structures by sequentially randomizing a trace.
- ❖ Ratio of compressed file sizes used to calculate complexities.

Resolving Structure Through Randomization



Calculating Complexity

- ❖ If c(σ) is the size of a compressed trace σ:
- ❖ Define:

Temporal complexity: $\tau(\sigma) = \frac{c(\sigma)}{c(t(\sigma))}$

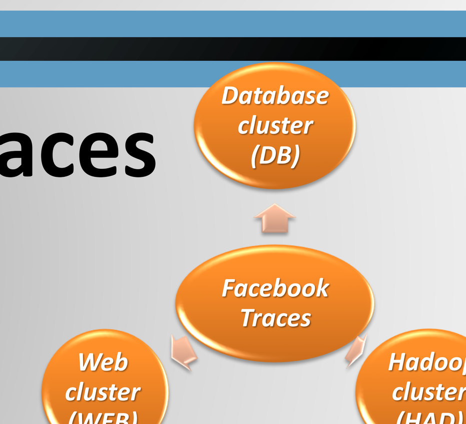
Non-temporal complexity: $n\tau(\sigma) = \frac{c(t(\sigma))}{c(u(\sigma))}$

Total complexity: $\psi(\sigma) = \frac{c(\sigma)}{c(u(\sigma))}$

- ❖ Trace complexity similar to normalized entropy measures

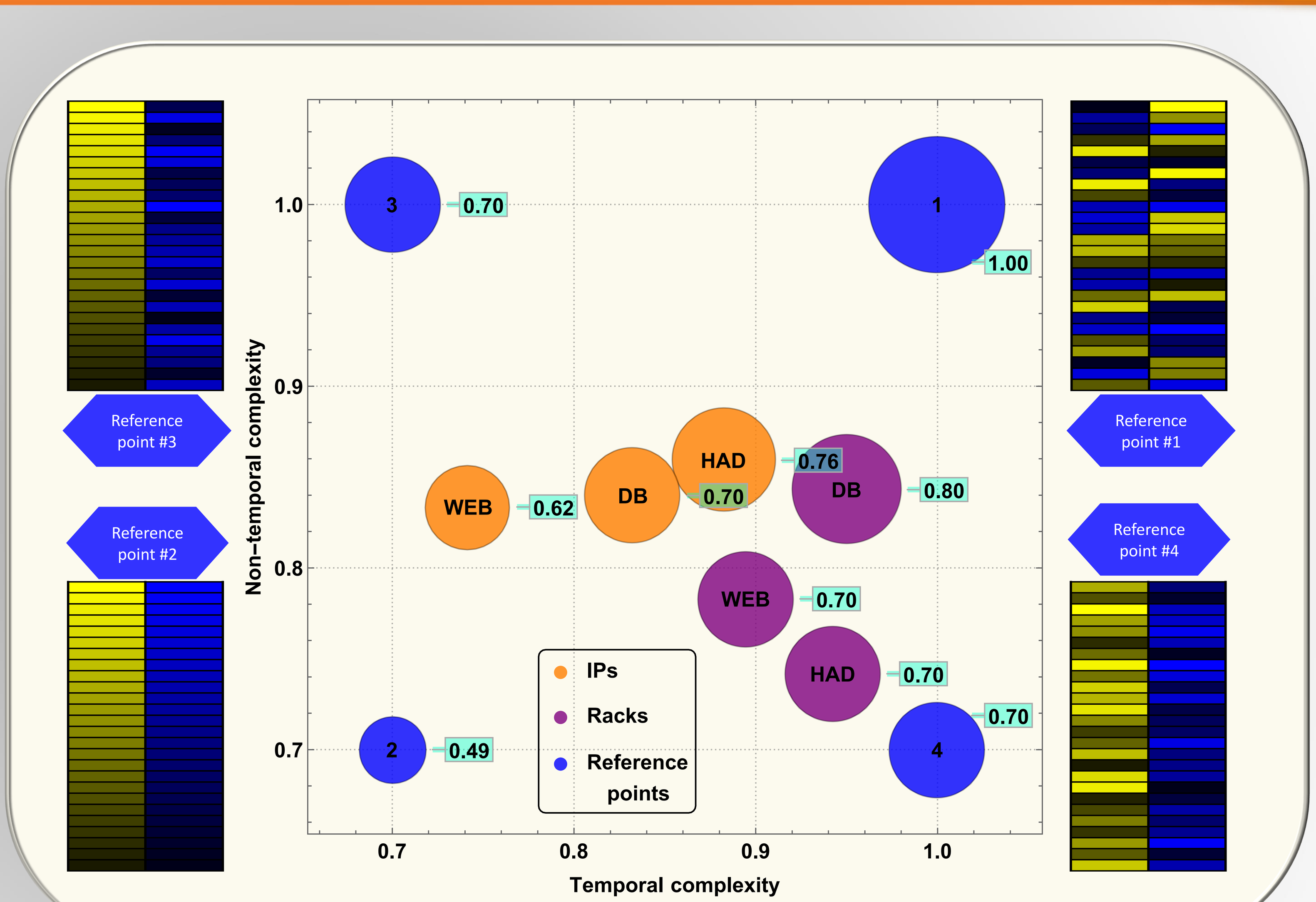
Case Study: Facebook Traces

- ❖ Real data [sigcomm15]
- ❖ Three clusters: DB,WEB, HAD.
- ❖ Two aggregation levels, Racks and IPs.



Complexity Map

- ❖ X axis: temporal complexity
- ❖ Y axis: non-temporal complexity.
- ❖ Size(value): total complexity.
- ❖ Rule of thumb: Closer to the axis's origin, lower complexity



Early Results & open Questions

- ❖ Different clusters, different complexities.
- ❖ Aggregation levels matter.
- ❖ IP has more temporal structure than rack.
- ❖ Traces show low complexity, with room for optimization.
- ❖ Is Hadoop traffic not fractal?
- ❖ What are the other types of complexity?
- ❖ How to add interarrival times to a trace?

What is the complexity of your traffic?