# Collection of Traffic Traces on the Global NDN Testbed

Sankalpa Timilsina, Davide Pesavento, Junxiao Shi, Susmit Shannigrahi, Lotfi Benmohamed



# Importance of traffic traces

- Reproducible results (e.g., in simulations)
- To enable apples-to-apples comparisons of different protocols and designs
- To help design forwarding strategies, caching policies, etc.
- Performance measurements for optimization and tuning
- Network troubleshooting and debugging

# Approaches to produce an NDN traffic trace

## Convert from other traces that contain hierarchical names (HTTP, NFS)

- Source traces are hard to obtain, harder to publish
- Source traces can be obfuscated but must preserve the name hierarchy
- Cannot capture the full semantics of NDN
- May not represent how native NDN apps interact

## Synthesize

- What parameters?
- Chicken-and-egg problem

## Capture from a real/realistic NDN deployment

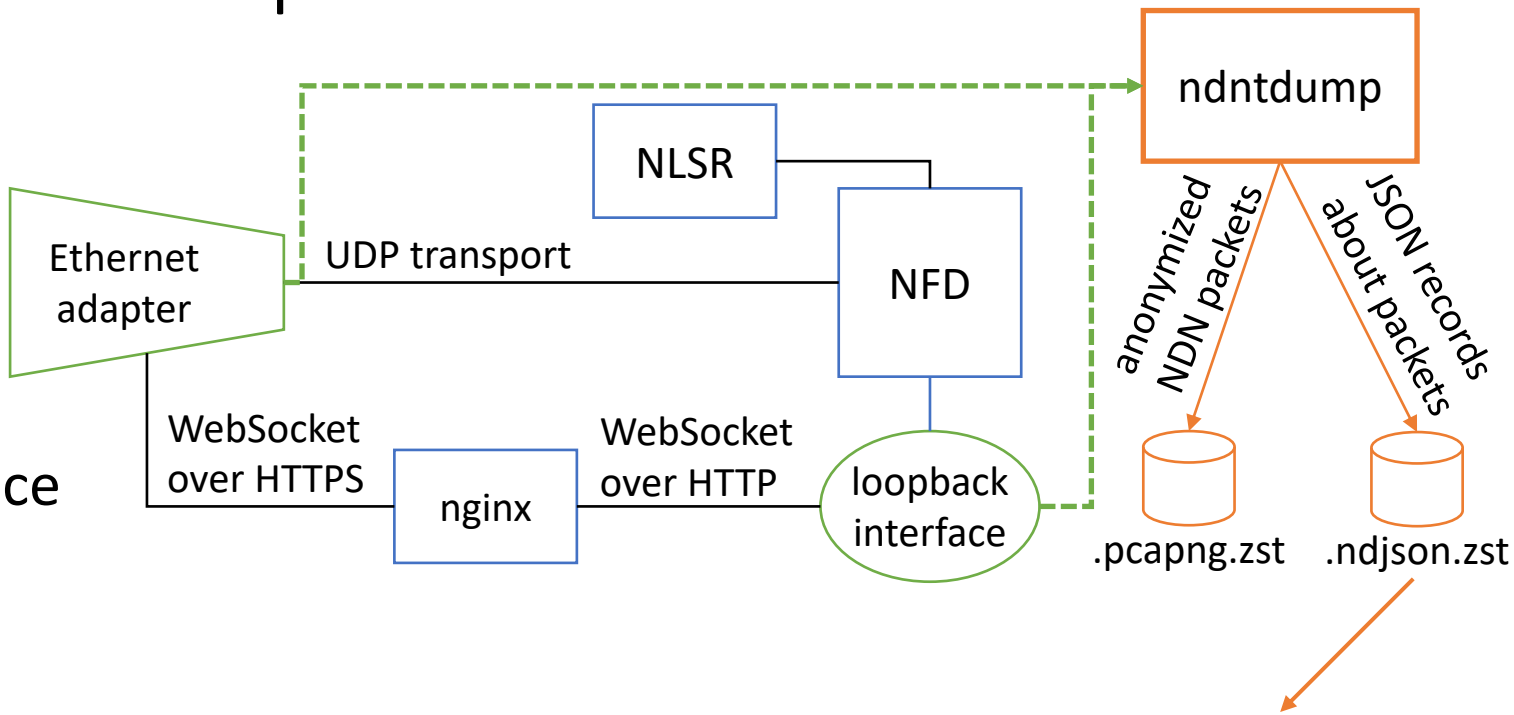
- Not many NDN applications
- Real users vs. synthetic users

# Goals

- **Dataset** of traffic traces
- Software **toolkit**
  - Automated traffic capture, anonymization, and compression
  - Replay the traces in ndnSIM and Mini-NDN
  - Analyze the traces to extract basic metrics and trends
  - Dissect the traces and take only a subset, e.g., only routing traffic
- Each piece is **modular** and can be used independently

# Trace collection

# *ndntdump*: NDN traffic dumper




- Live capture from network interface
  - Or read an existing trace from file
- Save in *pcap* or *pcapng* format
  - With on-the-fly gzip or zstd compression
- Best-effort anonymization features
  - Replace *Content* and *ApplicationParameters* with zeros
    - Also increases the compression efficiency
  - LSB of IPv4/IPv6/MAC address is XORed with a random key
  - More advanced anonymization techniques to be developed

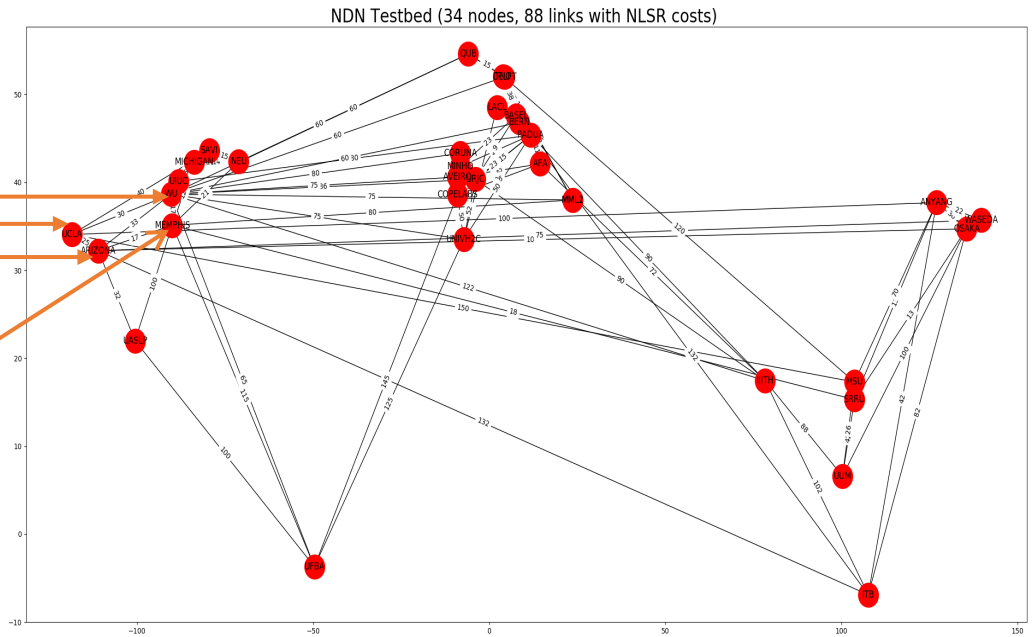
| | |
|---|---|
| Packet direction | Incoming or Outgoing |
| Flow key | String identifying UDP/TCP flow |
| Packet type | Interest, Data, Nack, or Fragment |
| Packet size | At NDNLPv2 layer and at network layer |
| Interest packet fields | Name, CanBePrefix, MustBeFresh, ForwardingHint, InterestLifetime, HopLimit |
| Data packet fields | Name, ContentType, FreshnessPeriod, FinalBlockId |
| Nack packet fields | NackReason, plus the fields from the enclosed Interest |

# Trace collection

We utilized the NDN testbed and the FABRIC testbed for data collection

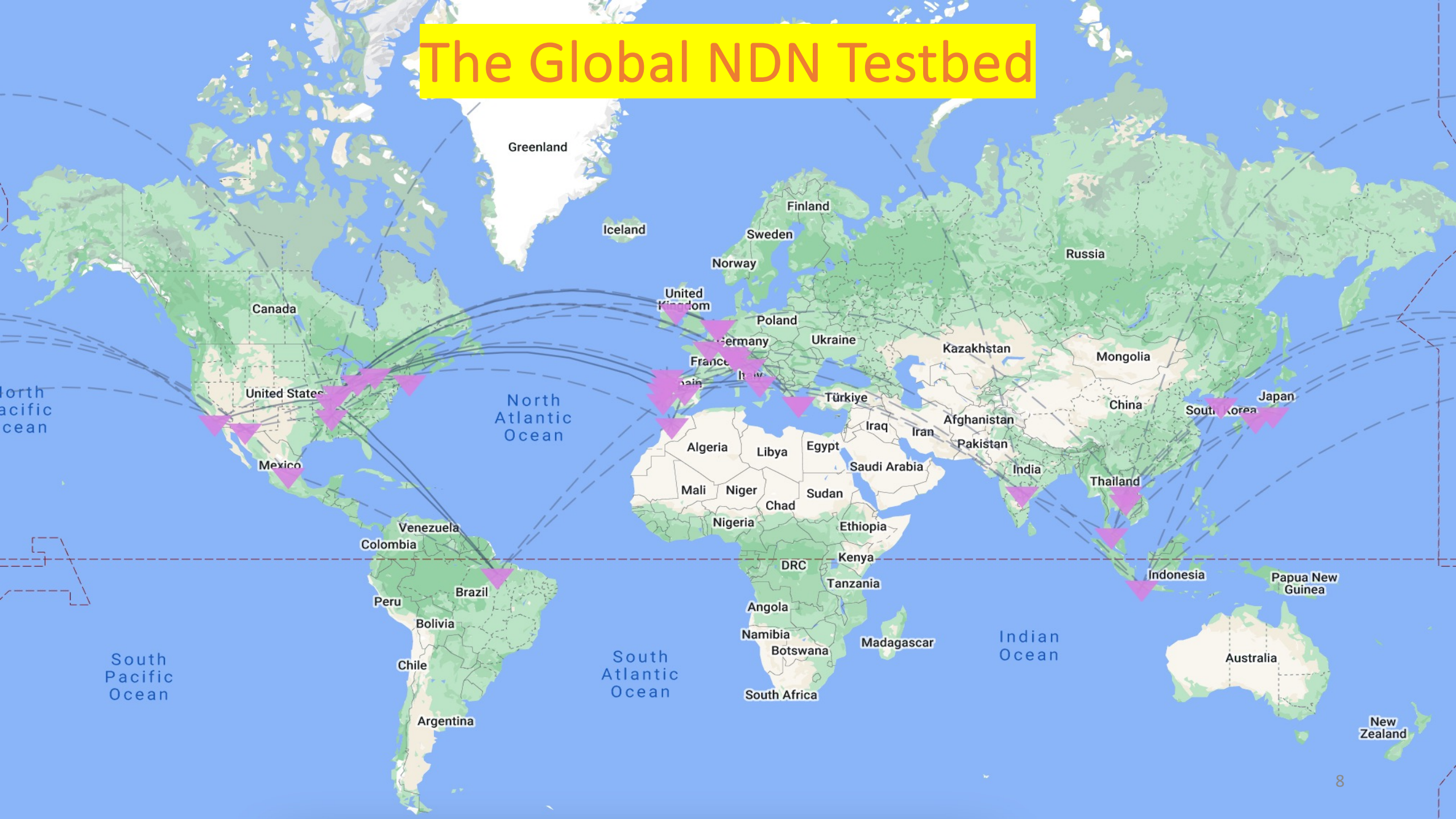
- Traffic was dumped from 4 NDN routers in North America
  - ➢ The testbed has ~30 nodes but only 16 were online at the time of the experiment in Q2 2023
- NDN applications generating additional traffic were deployed on FABRIC


**FABRIC**

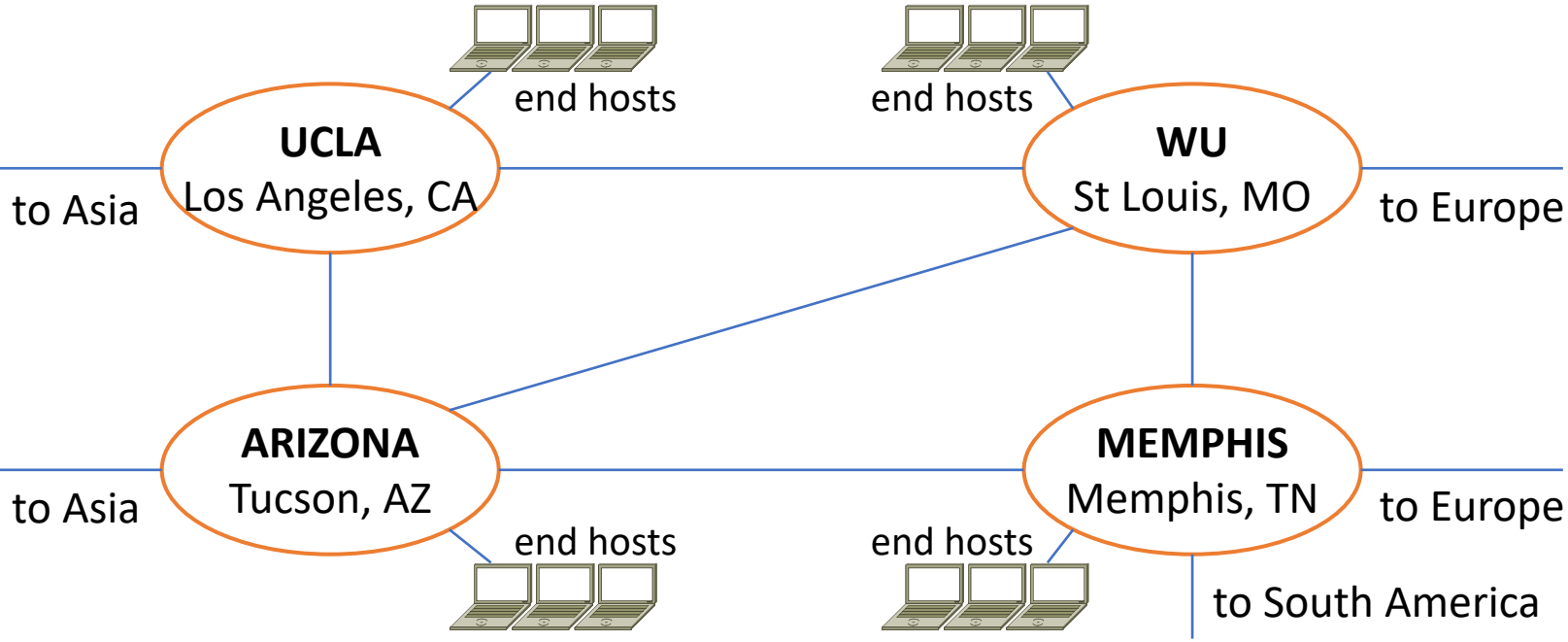



**NDN TESTBED**

# The Global NDN Testbed



# Trace collection




- Traffic dumpers placed on four testbed routers in North America
- End hosts (on FABRIC) running two kinds of NDN applications:
  - File transfer
  - Video streaming
- Duration: 3 hrs/day x 7 days

# Published dataset

- In total, we captured 320 GiB of traffic (13 GiB compressed)
  - 108 million packets
- Full dataset available at https://github.com/tntech-ngin/ndn-traffic-traces

❖ **7 different traffic scenarios:**

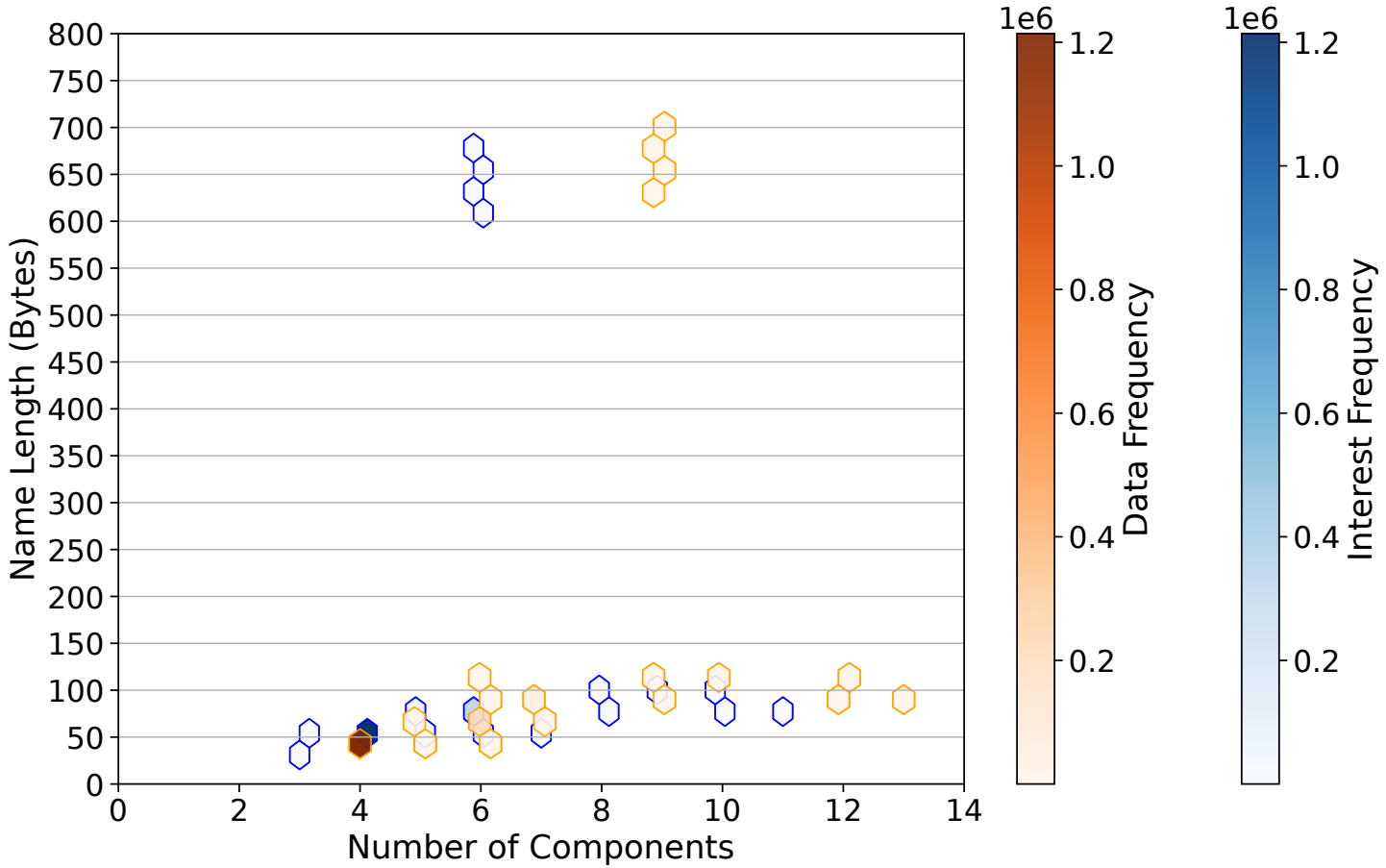
A. Baseline traffic from NLSR and other testbed users
B. File transfer only
C. Video streaming only
D. Both applications, one producer, one consumer
E. Both applications, one producer, 3 consumers
F. Both applications, 3 producers, one consumer
G. Same as D, but end hosts are spread across multiple FABRIC sites

| | Total packet count | | | Size | CBP | MBF |
|---|---|---|---|---|---|---|
| | Interests | Data | Nacks | (MiB) | (%) | (%) |
| A | 149,683 | 84,617 | 10,699 | 62 | 59.55 | 81.26 |
| B | 10,041,803 | 9,906,857 | 35,350 | 60,604 | 1.23 | 1.56 |
| C | 1,533,686 | 1,074,899 | 11,388 | 3,987 | 24.02 | 25.95 |
| D | 10,352,712 | 10,152,325 | 10,360 | 60,858 | 2.24 | 2.74 |
| E | 6,353,608 | 5,908,883 | 15,645 | 34,487 | 3.87 | 4.81 |
| F | 14,823,353 | 14,535,375 | 6,899 | 87,706 | 1.59 | 1.94 |
| G | 10,330,511 | 13,271,744 | 7,015 | 80,548 | 1.59 | 2.18 |

# Preliminary analysis of the traces

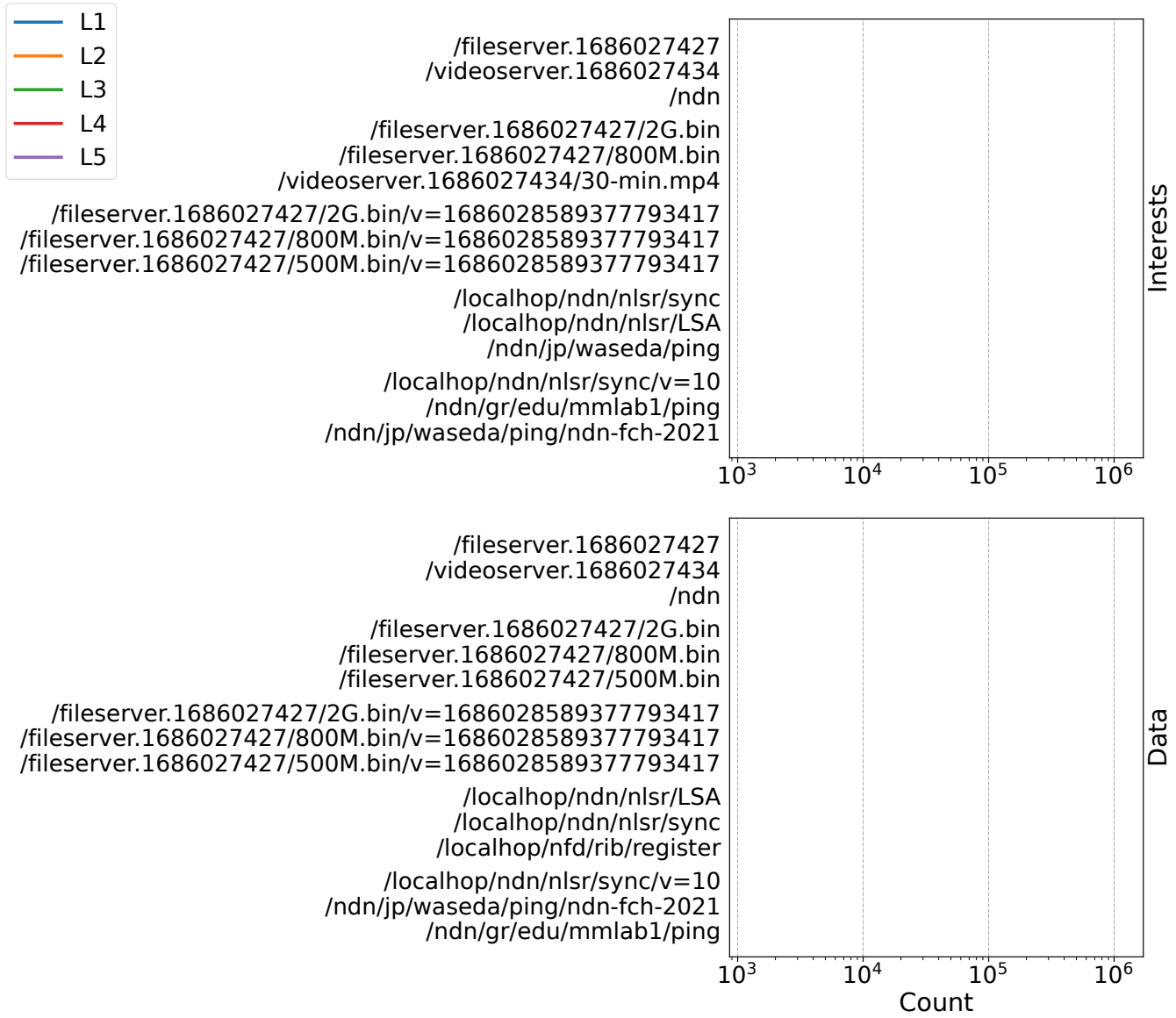
# Name components

- All names have between 3 and 13 components
  - Most have either 4 or 6 components
- Typically 25-130 bytes
  - Some names can go up to 700 bytes



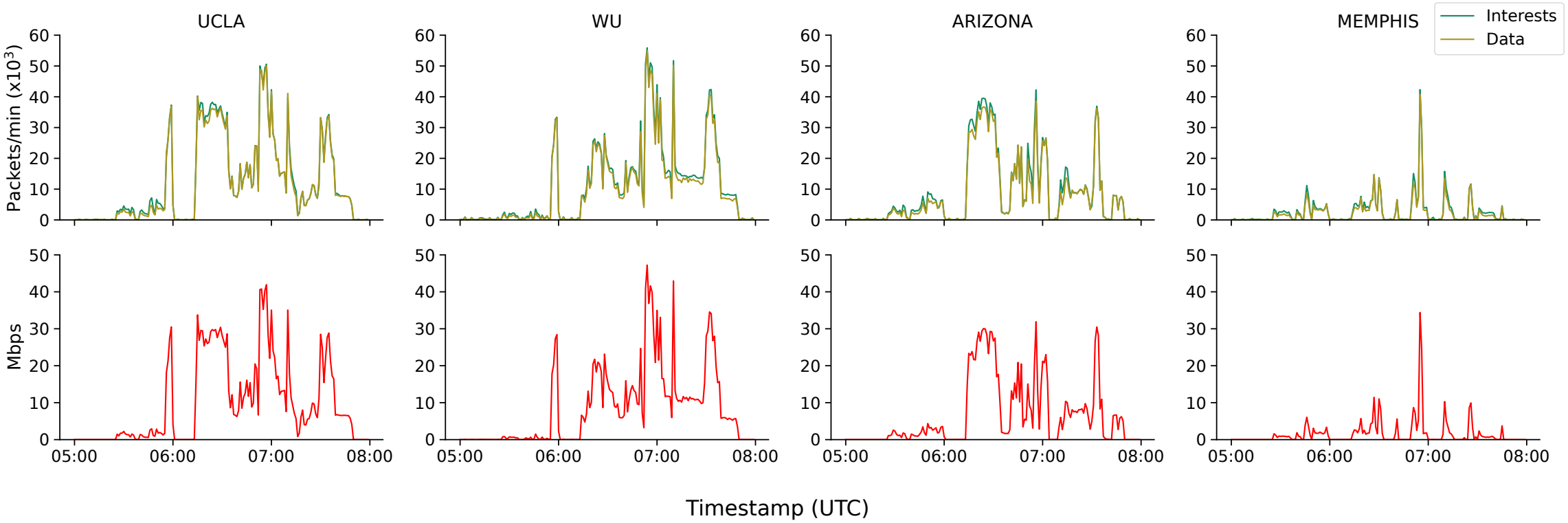
# Name prefixes

- The majority of the traffic is generated by the file transfer and video streaming applications
- Smaller amounts of traffic from ndnping and NLSR



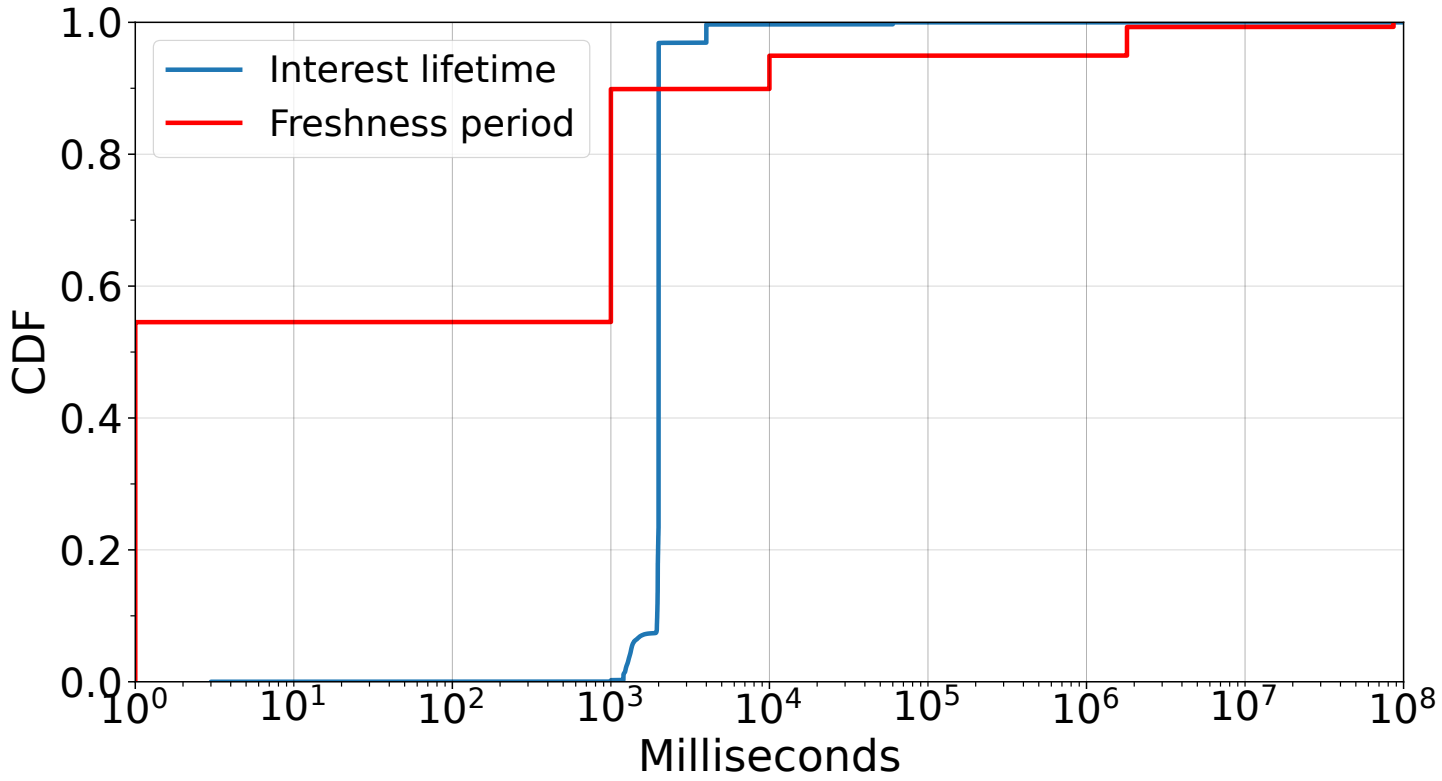
# Throughput and caching

- Producer: UCLA
- Consumers: WU, ARIZONA, Memphis
- UCLA received fewer Interests than cumulatively sent by consumers
  - Implies some caching/aggregation benefits



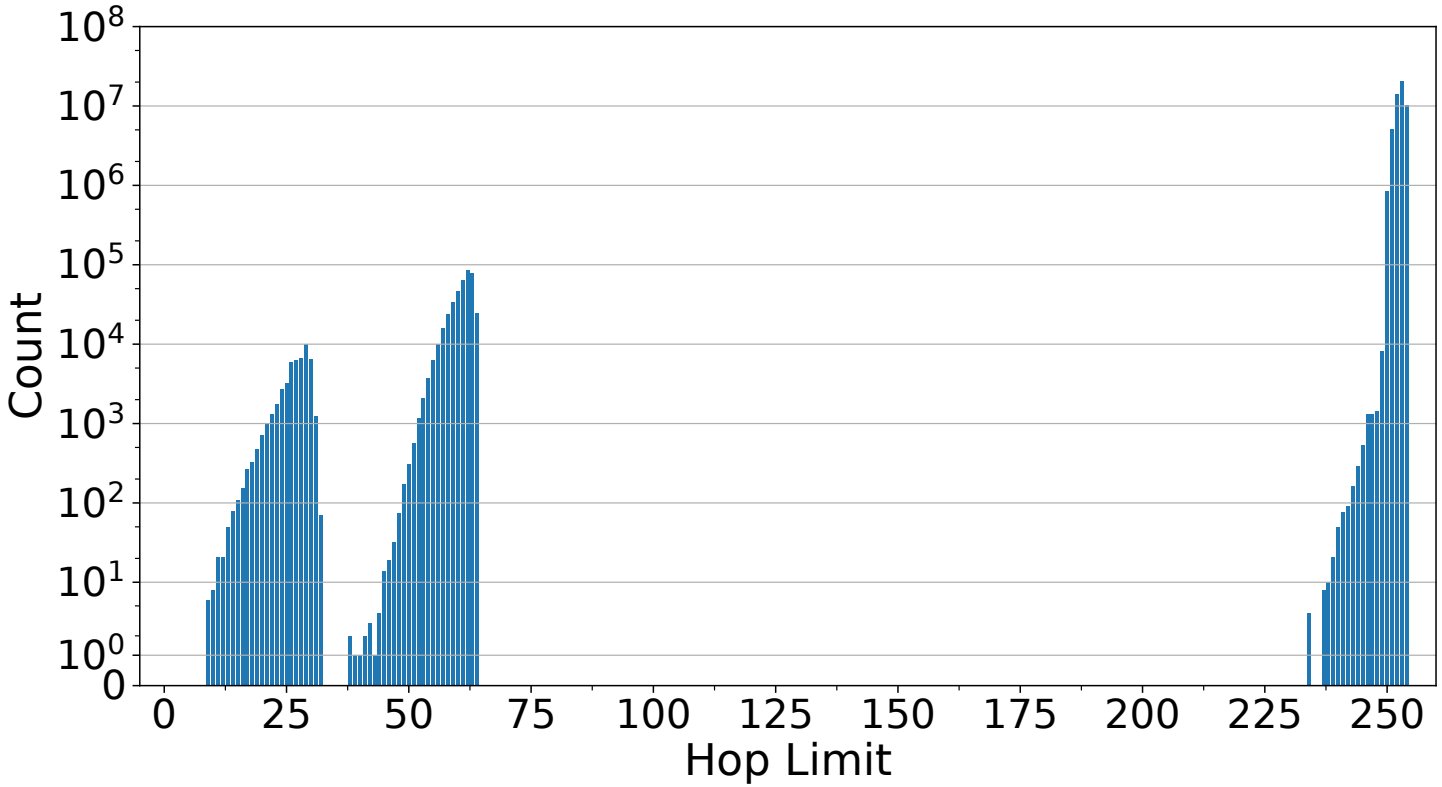
# Interest Lifetime and Data Freshness

- 2000 ms is the most commonly used Interest lifetime
- 97% of Data packets do not have a freshness period
  - Not shown in the plot



# Hop Limit

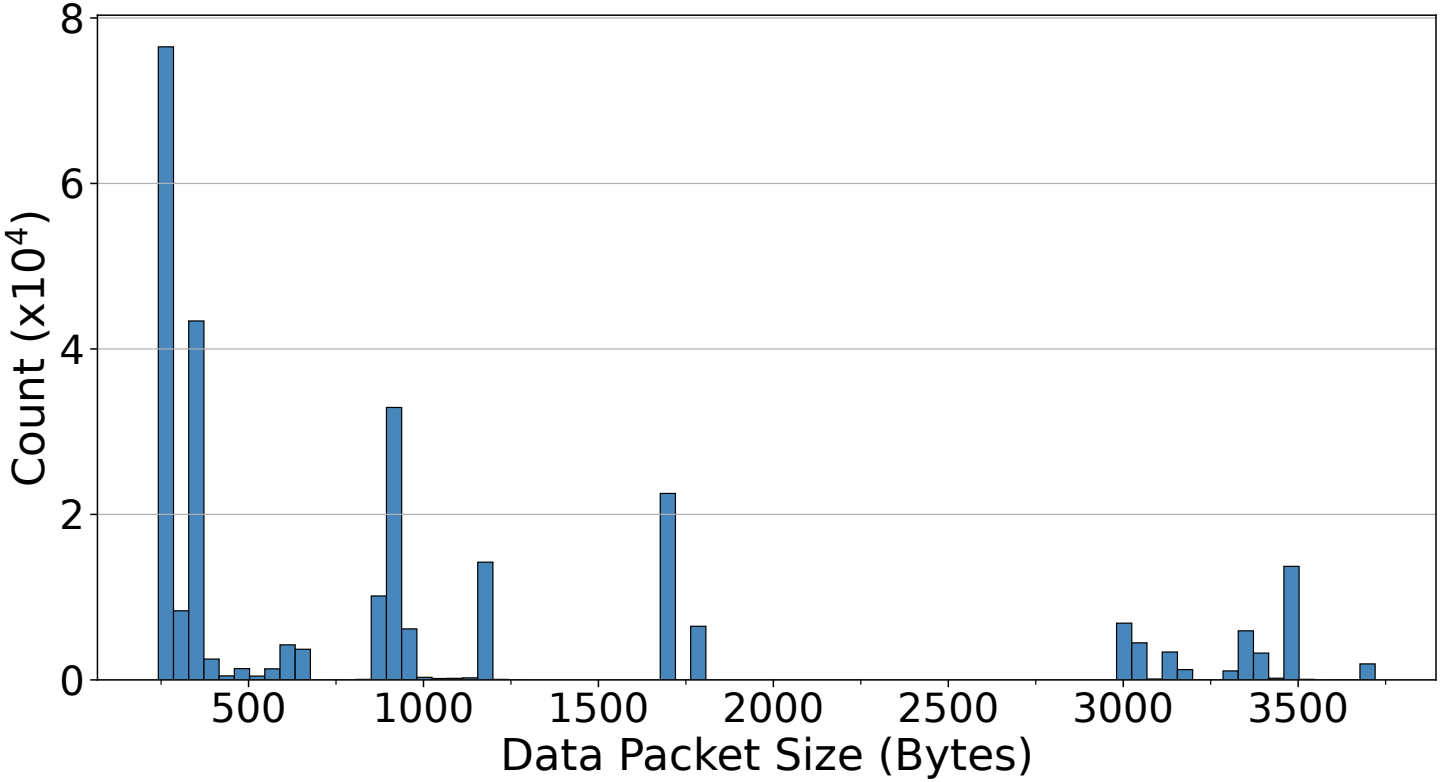
- Three peaks at and below 32, 64, and 255
  - Attributed to the default settings in various NDN implementations



# Routing protocol traffic (NLSR)

Clearly shows 4 different types of NLSR packets:

- HELLO
- PSYNC
- LSDB coordinates
- LSDB names



# Caveats

- Traffic mix is probably not representative of a real NDN deployment
  - But this is the best we can provide at the moment
- Traffic anonymization is a simple XOR
  - Researchers can plug in better algorithms
- Testbed scale
  - Current dataset: 4 nodes, U.S. centric
  - Added 10 more collection points in the past month
- Only packet traces are collected
  - May be hard to reconstruct and understand the forwarder's behavior

# What’s next?

# Potential next steps

On-demand traffic collection, filtered by name prefix

Expand data collection to more than just packet traces

- Data plane telemetry
- Control plane telemetry

Analysis of additional metrics, e.g., RTT

Real-time troubleshooting of the testbed

- “Where did my packet go?”

Better traffic anonymization techniques

# How the community can help

## Currently the community lacks real datasets

- Use the provided dataset and let us know what else is needed
- Use our toolkit to collect (and publish) your own NDN traces

## The current dataset lacks application diversity

- Use the NDN testbed for your research / pilot deployments

## Expand the testbed

- It is currently shrinking

# References

- Project website with all current and future datasets: https://tntech-ngin.net/datasets
- Dataset: https://github.com/tntech-ngin/ndn-traffic-traces
- Traffic capture scripts: https://github.com/tntech-ngin/ndn-traffic-capture-scripts
- Plotting scripts: https://github.com/tntech-ngin/ndn-traffic-plotting-scripts
- Trace replay tool for ndnSIM (proof of concept): https://github.com/tntech-ngin/ndn-traffic-replay-ndnSIM
- NDN traffic dumper: https://github.com/usnistgov/ndntdump
- ACM ICN 2023 paper: https://www.nist.gov/publications/capture-and-analysis-traffic-traces-wide-area-ndn-testbed