

# Requirements of a Unified Transport Protocol for In-Network Computing in Support of RPC-based Applications

draft-song-inc-transport-protocol-req-01

Haoyu Song (Futurewei)

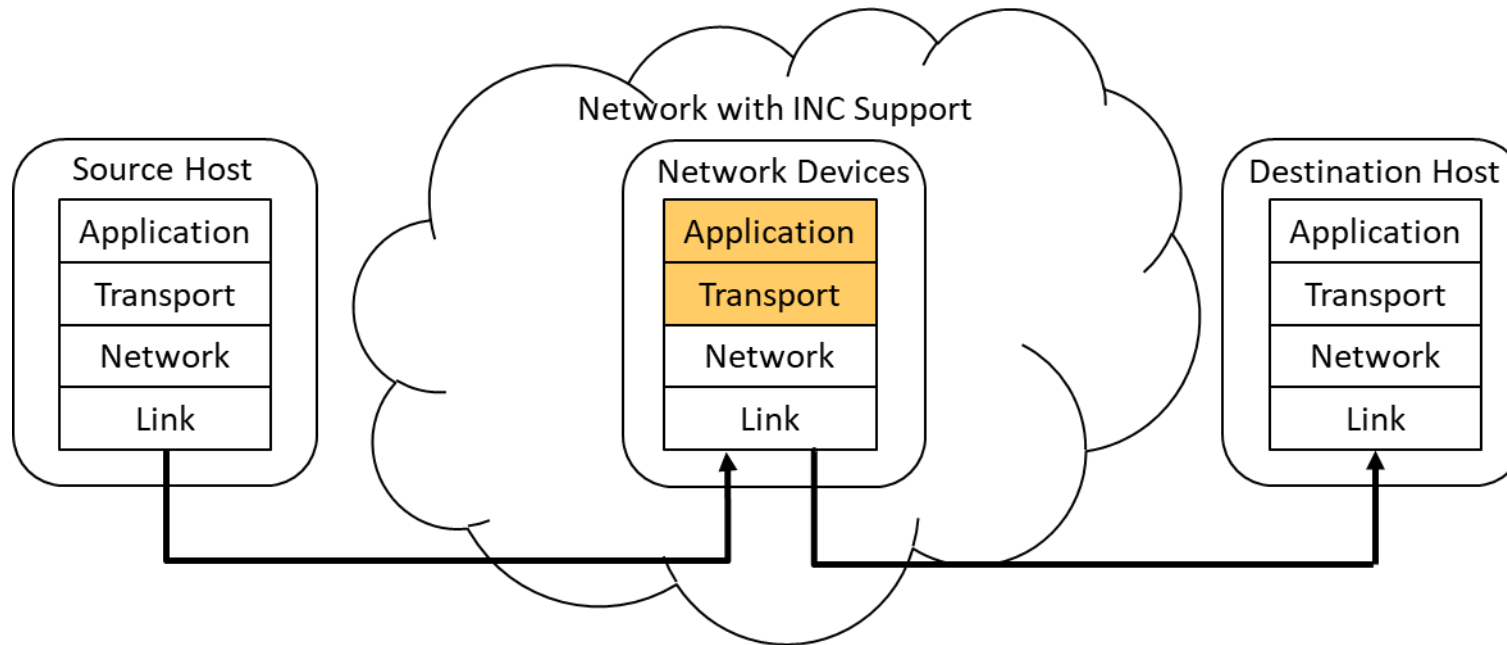
Wenfei Wu (Peking University)

Dirk Kutscher (HKUST)

# Background

- In-network computing (INC) is a new paradigm that involves the programmable network switches to improve the application performance (e.g., lower latency, higher throughput) or reduce the system cost (e.g., lower power consumption, fewer servers)
- The most promising INC applications (aggregation, caching, agreement) follows the similar pattern
  - Each packet is an individual message that can be processed
  - Idempotent Remote Procedure Call (RPC)
  - The computing can be solely or partially done by network devices in the network, and the result can be replied by a network device or a server
  - No data streaming is required, so TCP is not needed
  - More reliability is required, so UDP is not enough

# INC Network Protocol Layer Model



- Break the E2E assumption of the transport layer
- A new E2M2E model (M stands for middle point)

# INC Service Model and Communication Pattern

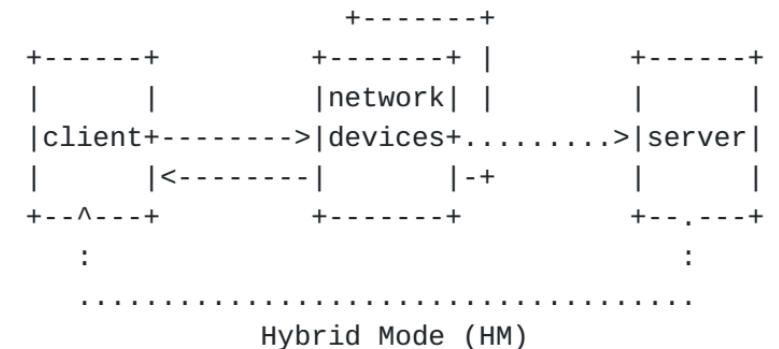
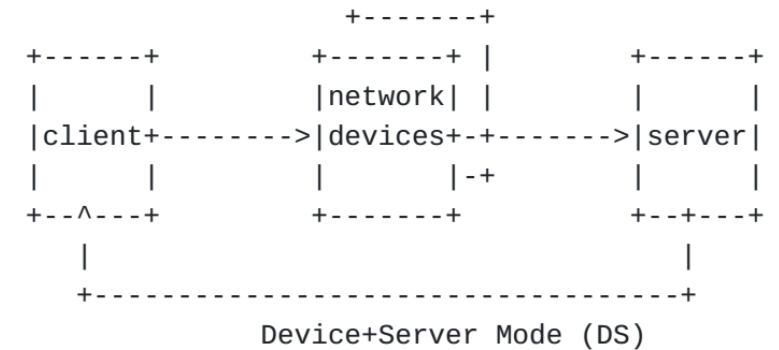
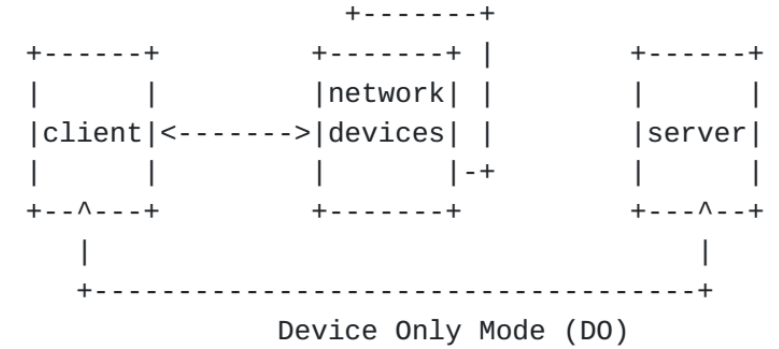
|                         | DO | DS | HM |
|-------------------------|----|----|----|
| Sync Collaboration(SC)  | x  | x  | x  |
| Async Collaboration(AC) |    | x  |    |
| Individual Request(IR)  | x  |    | x  |

Example:

SC: AllReduce

AD: MapReduce

IR: NetCache



# INC Transport Protocol Function and Requirements

- Main function:
  - Signaling
    - Request in-network processing and get acknowledged
    - Allow falling back to server processing
  - Reliability
    - Packet loss detection and retransmission
  - Congestion control
    - Window-based
    - ECN
- Requirement:
  - Keep complex logic out of switch, only done in end servers
  - General to all the INC applications with the RPC pattern
  - Extensible for future enhancements
- Advantage:
  - Transparent to algorithms for packet loss detection (in switch and in server) and congestion control (in server)
  - Transparent to application logic

# Existing Transport Protocols (do not work for INC)

- TCP
- UDP
- QUIC
- MTP
- RDMA
- HOMA
- Falcon
- ...

# Discussion

- Will collective communication require in-network computing for acceleration?
- If so, what functions will be supported by INC?
- What are the CCO specific requirements for the transport protocol?