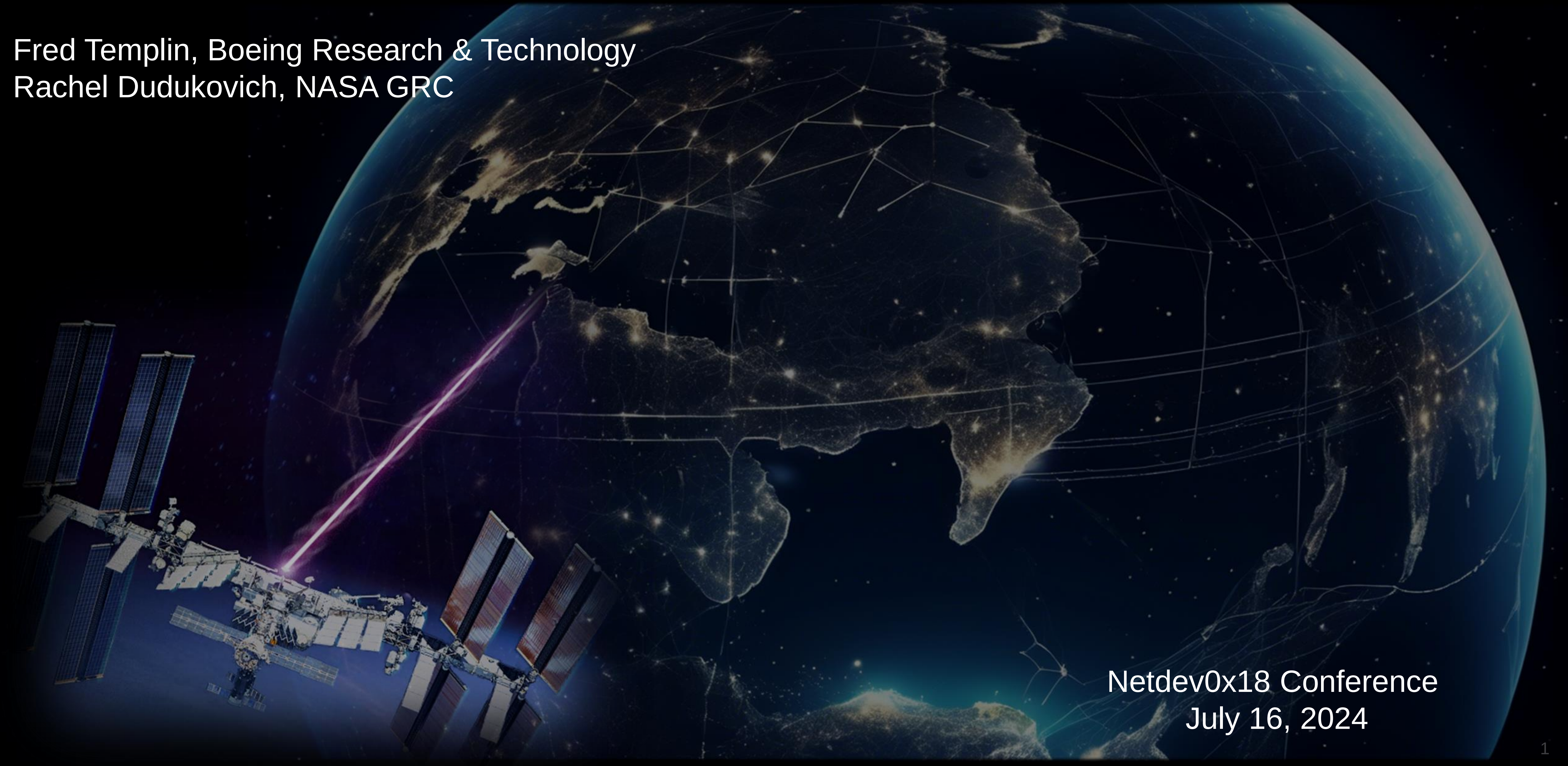
High-performance DTN Using Larger Packets with Forward Error Correction

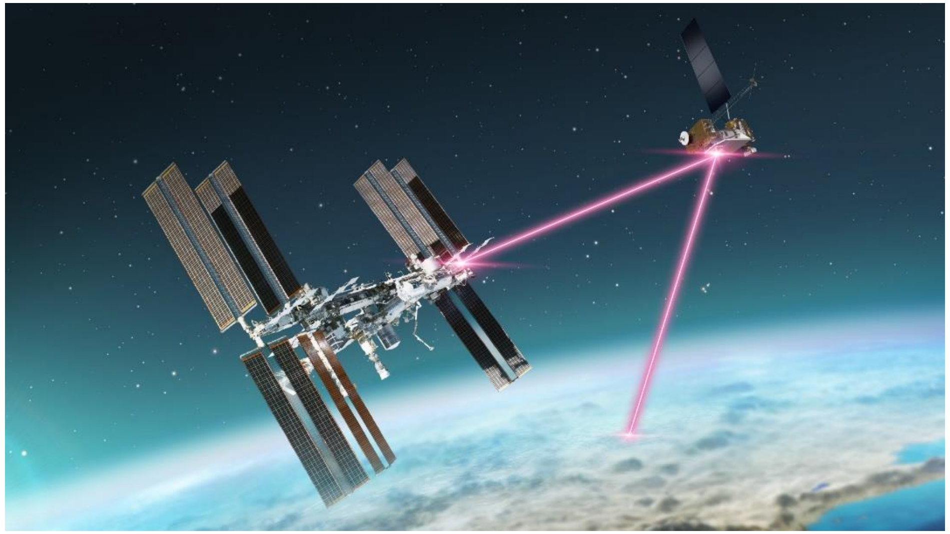
Fred Templin, Boeing Research & Technology
Rachel Dudukovich, NASA GRC

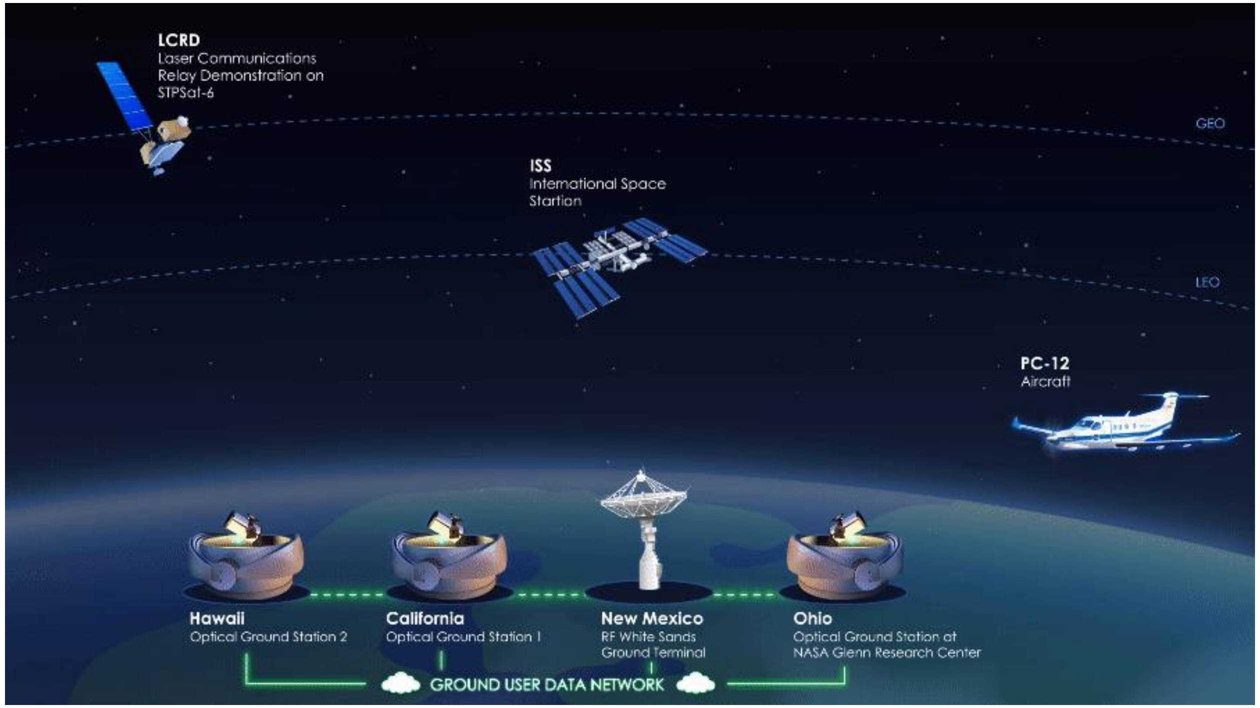


Netdev0x18 Conference
July 16, 2024

NASA Use-Cases: High-rate Optical Gateway

ISS ILT/LCRD Demo





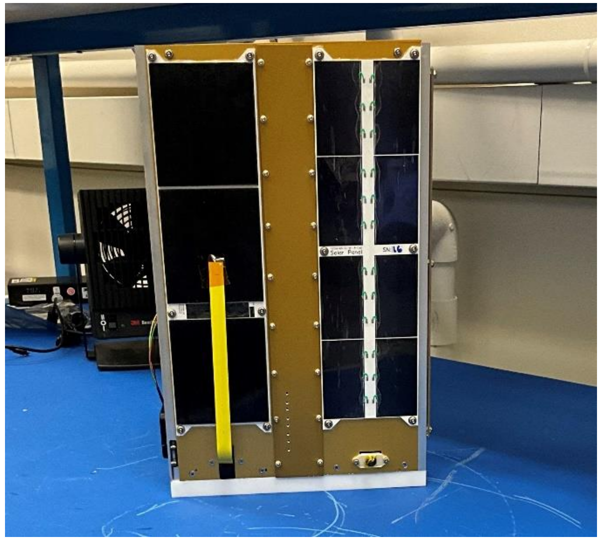
Characteristics

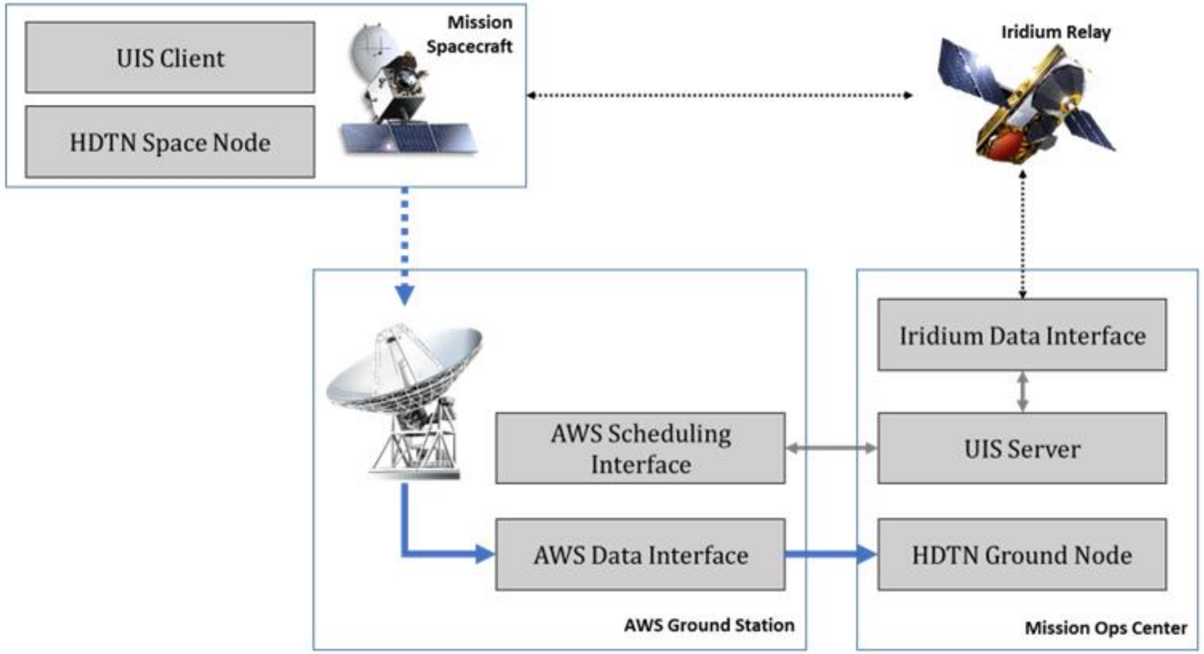
- Large-scale platform (commercial laptop)
- Gigabit per second downlink
- Bi-directional link (155 Mbps forward, 1244 Mbps return)
- Significant roundtrip time (seconds)
- Capable of multi-source/multi-destination
- Accessible to operators and reconfiguration possible after launch

- Demonstration of high-rate onboard gateway and near space ground network
- Space to ground always used LTP
- BP v6 w/wo custody transfer
- BP v7 with BPSec
- Multimedia streaming

NASA Use-Cases: Resource Constrained Platforms

TechEd Sat 11





Characteristics

- Small embedded platform
- Highly asymmetric/unidirectional communication
  - Transmission via unidirectional S-band radio
  - Command interface via Iridium short burst data service
- Very limited software reconfiguration after launch

- Small research payloads
- Low cost demonstrations
- Custom communication pipeline, “non-networked”
- Utilizing FEC rather than LTP

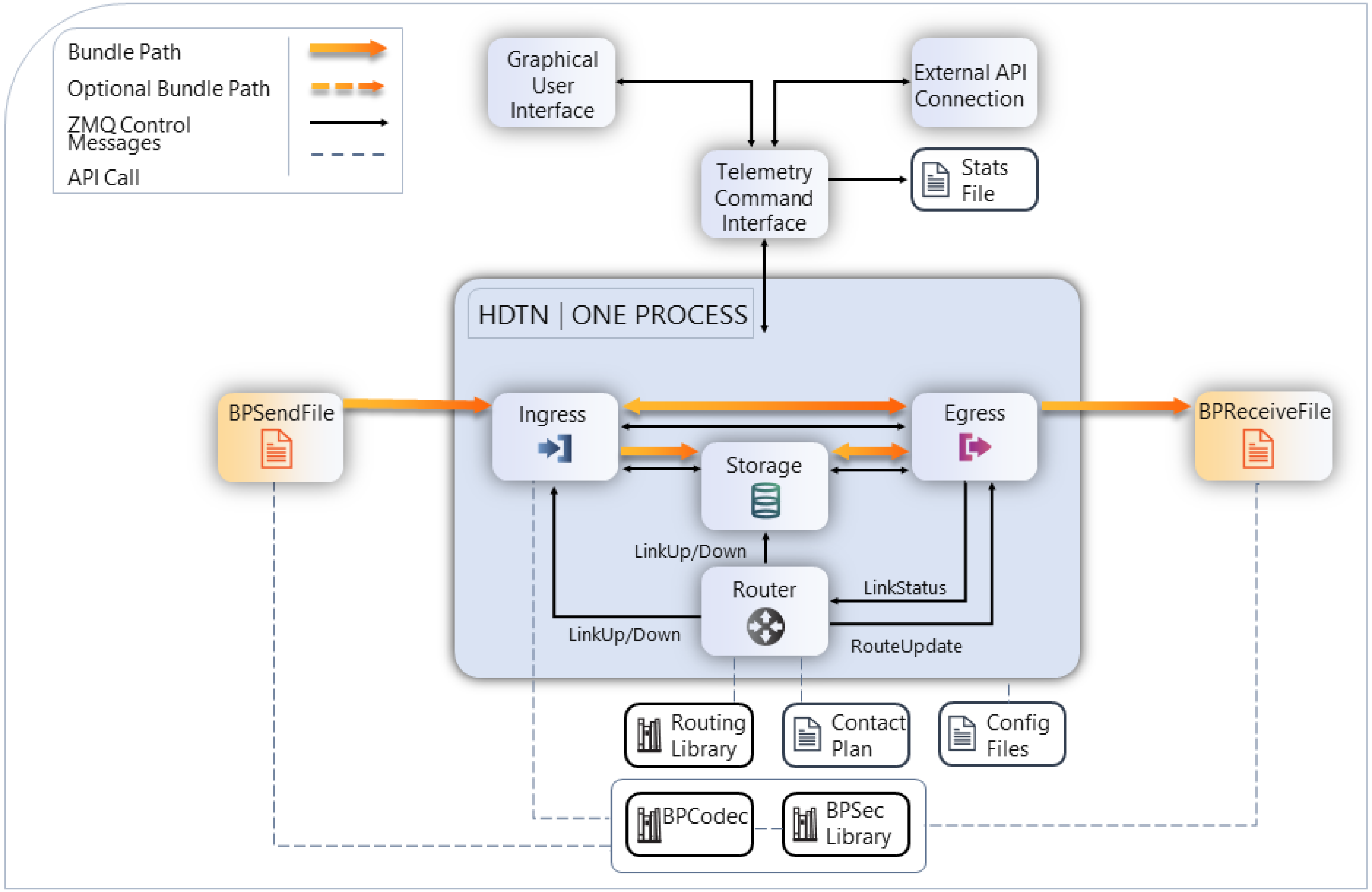
HDTN Architecture

Performance

- Message bus architecture
  - Distributed and single process modes
- Avoids semaphore and mutex locks on shared memory
- Avoids copying memory
- Asynchronous operations

Usability

- Platform independent
- Well maintained dependencies
- Fully open-source with documentation
- Graphical interface
- API and command line interface

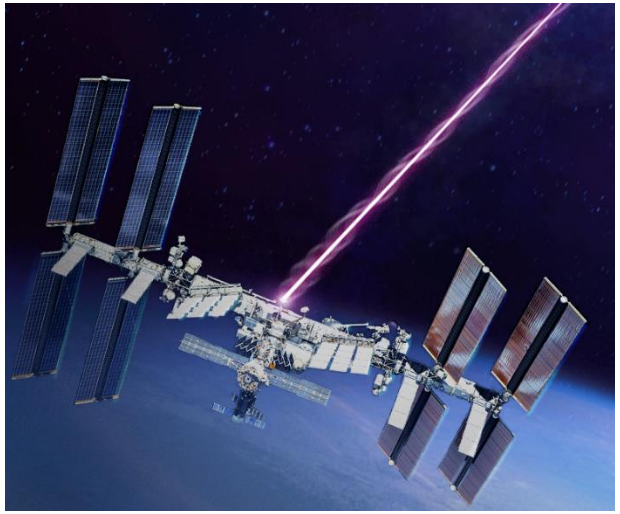


Evaluation of LTP in Multiple Environments

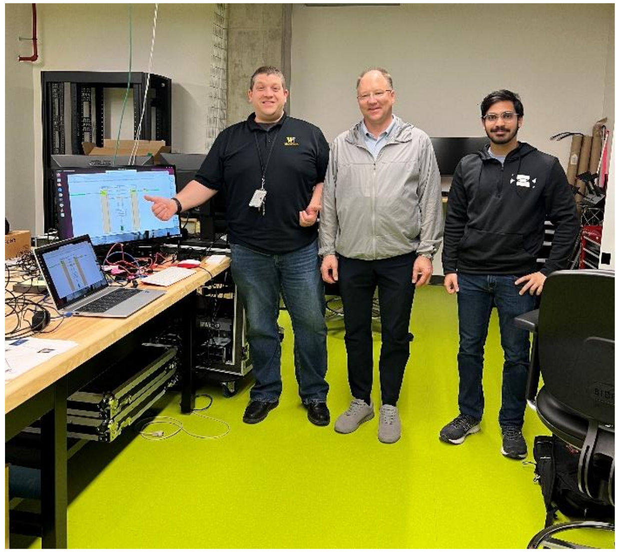
Software Defined Radio Lab w/Cesium Astro

Optical Comm PC-12 Aero Experiments

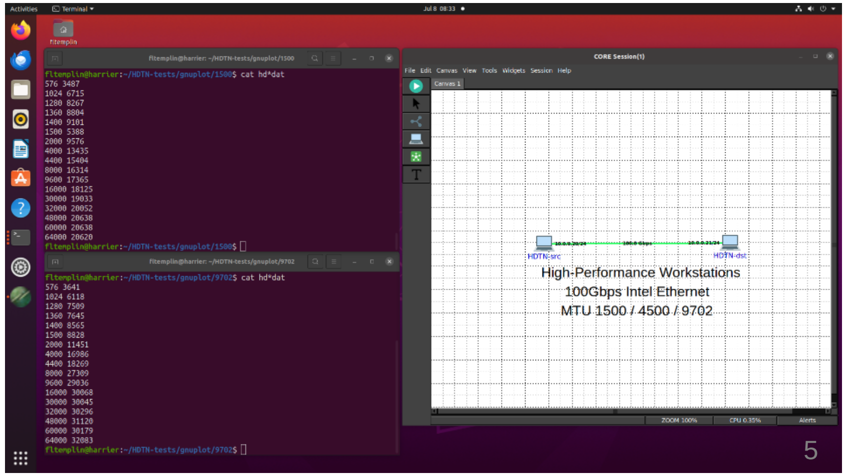
ISS ILT/LCRD



LunaNet Testbed



Boeing High-rate Testbed



Challenges and Opportunities

Challenges

- LTP protocol is complex with many parameters to configure
  - May result in errors or poor performance
- May be difficult to formally verify requirements
- Still relies on 2-way communication
  - May not be possible
  - May degrade performance

Opportunities

- Perform parameterized benchmarking and analysis
- Trade study between LTP and custody transfer
- Investigate new protocols
  - High Performance Reliability Protocol
  - Forward error correction
  - Others
- Refine protocol specification


```
"outductsConfig": {
    "outductConfigName": "myconfig",
    "outductVector": [
        {
            "name": "for egress",
            "convergenceLayer": "ltp_over_udp",
            "nextHopNodeId": 20,
            "remoteHostname": "hdtn_receiver",
            "remotePort": 1113,
            "maxNumberOfBundlesInPipeline": 50,
            "maxSumOfBundleBytesInPipeline": 50000000,
            "thisLtpEngineId": 10,
            "remoteLtpEngineId": 20,
            "ltpDataSegmentMtu": 1360,
            "oneWayLightTimeMs": 1000,
            "oneWayMarginTimeMs": 200,
            "clientServiceId": 1,
            "numRxCircularBufferElements": 100,
            "ltpMaxRetriesPerSerialNumber": 5,
            "ltpCheckpointEveryNthDataSegment": 0,
            "ltpRandomNumberSizeBits": 64,
            "ltpSenderBoundPort": 1113,
            "ltpMaxUdpPacketsToSendPerSystemCall": 15,
            "ltpSenderPingSecondsOrZeroToDisable": 15,
            "delaySendingOfDataSegmentsTimeMsOrZeroToDisable": 20,
            "keepActiveSessionDataOnDisk": false,
            "activeSessionDataOnDiskNewFileDurationMs": 2000,
            "activeSessionDataOnDiskDirectory": ".\/"
        }
```

Packet Size Issues

- LTP/UDP/IP encapsulates LTP segments in UDP/IP packets
- Most Internet and DTN links configure a **1500 byte Maximum Transmission Unit (MTU)** (largest packet size for the link)
- Smallest **link MTU** in path determines **path MTU**
- Transport layer protocols (TCP, QUIC, LTP, etc.) often limit packet sizes **to no larger than the path MTU**
- **IP fragmentation** needed for larger sizes, but:
  - “IP Fragmentation Considered Harmful” (Kent, Mogul – 1987)
  - “IP Fragmentation Considered Fragile” (IETF RFC8900 – 2020)
  - BCP: use path MTU discovery instead (IETF RFC1191, RFC8201)

Path MTU Discovery (PMTUD)

- Depends on ICMP **Packet Too Big (PTB)** messages from the network (messages may be lost or spoofed)
- PTBs always indicate packet loss; source backs off to using smaller packets for long periods of time before trying again **(not adaptive)**
- Discovering larger MTUs over arbitrary Internet paths difficult using legacy PMTUD mechanisms, but:
  - Newer packetization layer (end-to-end) active probing approaches offer possible improvements (IETF RFC4821, RFC8899)
  - New approach uses passive hop-by-hop measurements (IETF RFC9268)

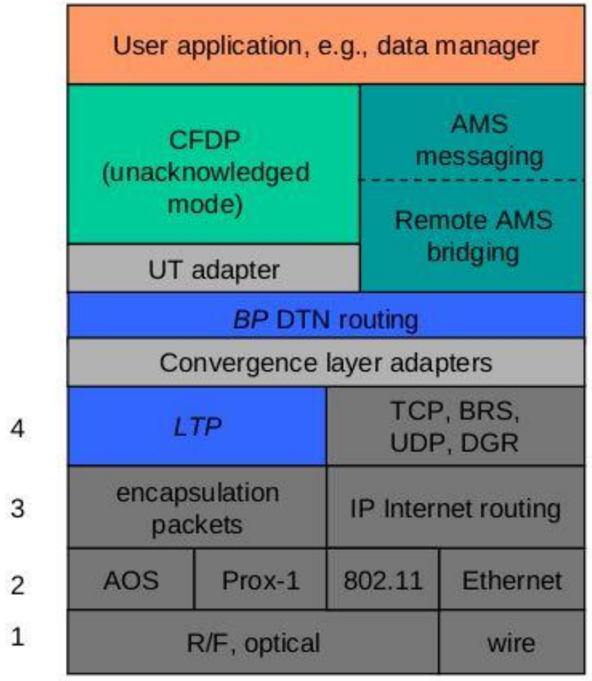
Generic Segment/Receive Offload (GSO/GRO)

- PMTUD shortcomings often cause transport protocols to use small segment sizes
- Small segment sizes can cause performance bottleneck at OS syscall interface since **small amount of data copied per call**
- GSO/GRO concatenates multiple smaller segments into larger buffer; **amortizes data copies across syscall interface**
  - Source OS fragments large GSO buffer into smaller whole packets for transmission
  - Destination OS reassembles packets into large GRO buffer for transport protocol delivery

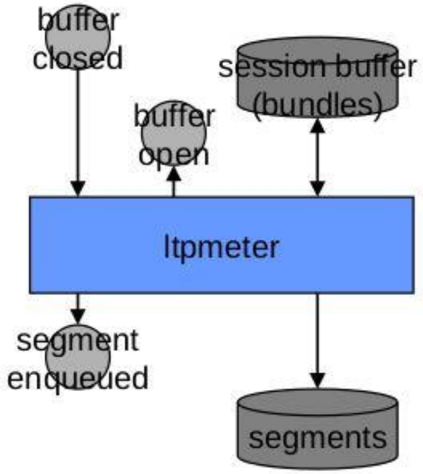
Delay Tolerant Network (DTN) Protocol Layering

- DTN Bundle Protocol (BP) introduces **new layer in architecture** below applications but above transport
- Licklider Transmission Protocol (LTP) is a **transport protocol convergence layer for BP**
- LTP breaks bundles into segments for transmission
- **Segment size affects performance**

1. Initializes session buffer, gives *buffer open* semaphore.
2. Waits for *buffer closed* semaphore (indicating that the session buffer is ready for transmission).
3. Segments the entire buffer into segments of managed MTU size – fragmentation.
4. Appends all segments to segments queue for immediate transmission.
5. Gives *segment enqueued* semaphore.



ION DTN Protocol Stack (*)



LTP Processing in ION (*)

* Excerpted from Interplanetary Overlay Network (ION) Design and Operation Guide (V4.0.1)

LTP Performance

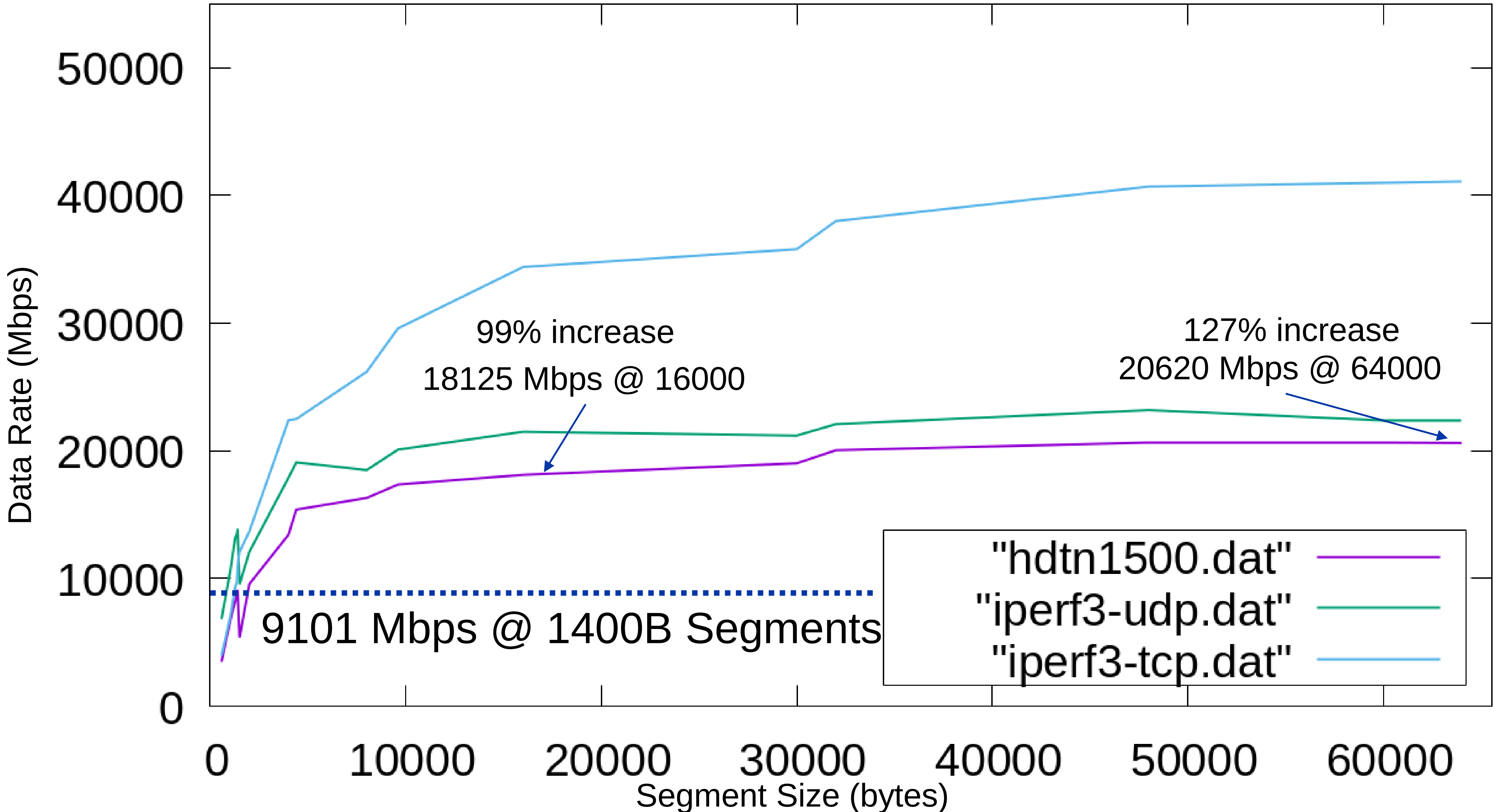
- Implemented GSO/GRO in ION DTN LTP but **saw no performance benefit**; syscall interface not a bottleneck
- Experiments with larger ION LTP segment sizes showed dramatic performance increases **even when IP fragmentation engaged**
- Larger HDTN LTP segment sizes also showed significant increases
  - For two popular DTN LTP implementations, increasing LTP segment size directly increases performance **even when IP fragmentation engaged**
  - Mirrors earlier Internet services such as NFS over UDP that saw greater performance using larger segment sizes with IP fragmentation

Performance Testbed

- Dell Precision 3660 workstations; Ubuntu 20.04 LTS operating system
- 12th Generation Intel Core I7-12700Kx20 processors; 32GB memory
- Intel E810 CQDA2 100Gbps Ethernet Network Interface Cards (NICs)
- NICs connected point-to-point with Cat 6 Ethernet cable
- NICs can accept MTU configurations up to 9702 octets
- Used 1500, 4500 and 9702 octet MTU settings in tests

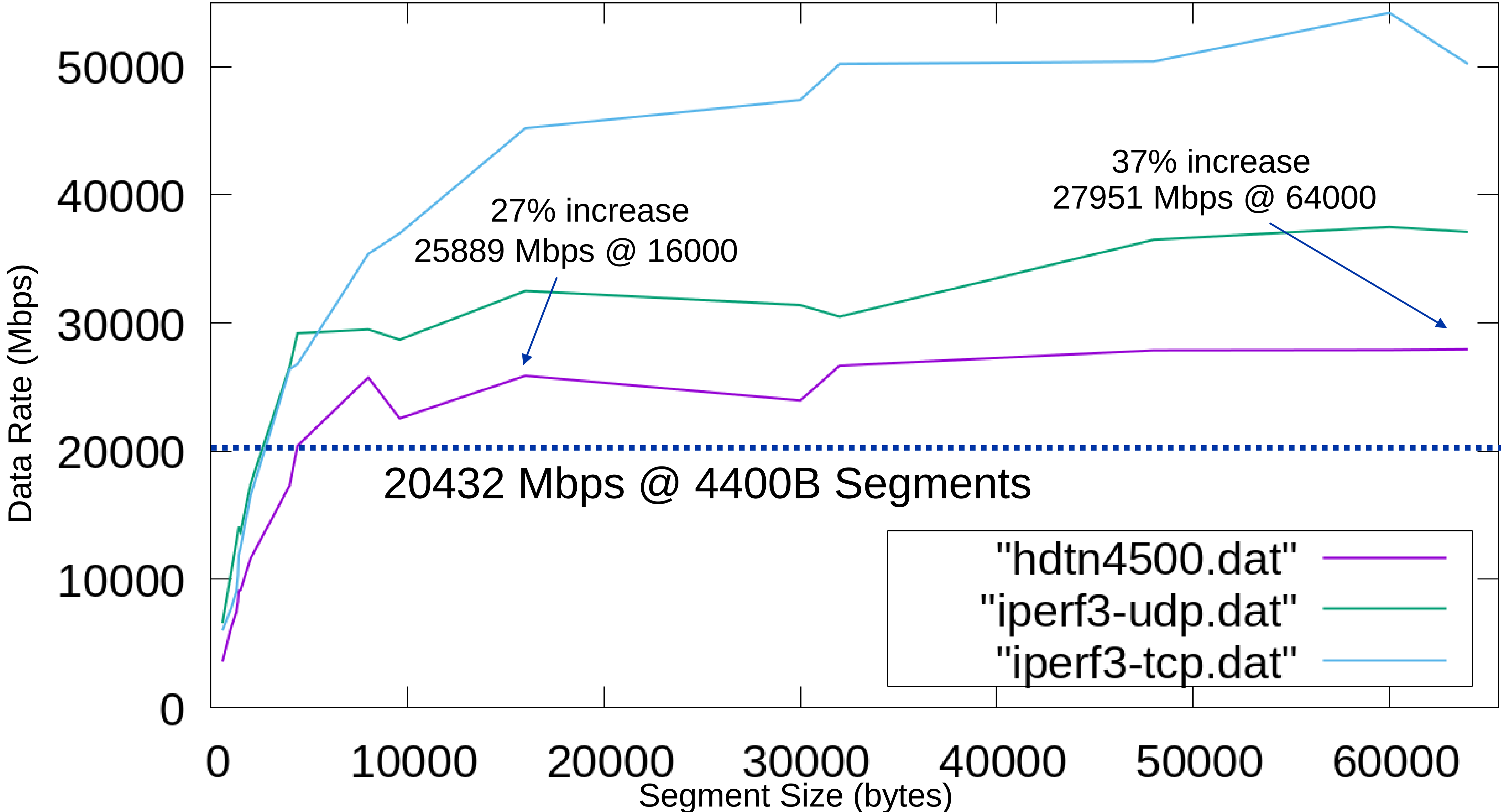
HDTN LTP Performance

100Gbps Ethernet with 1500 MTU

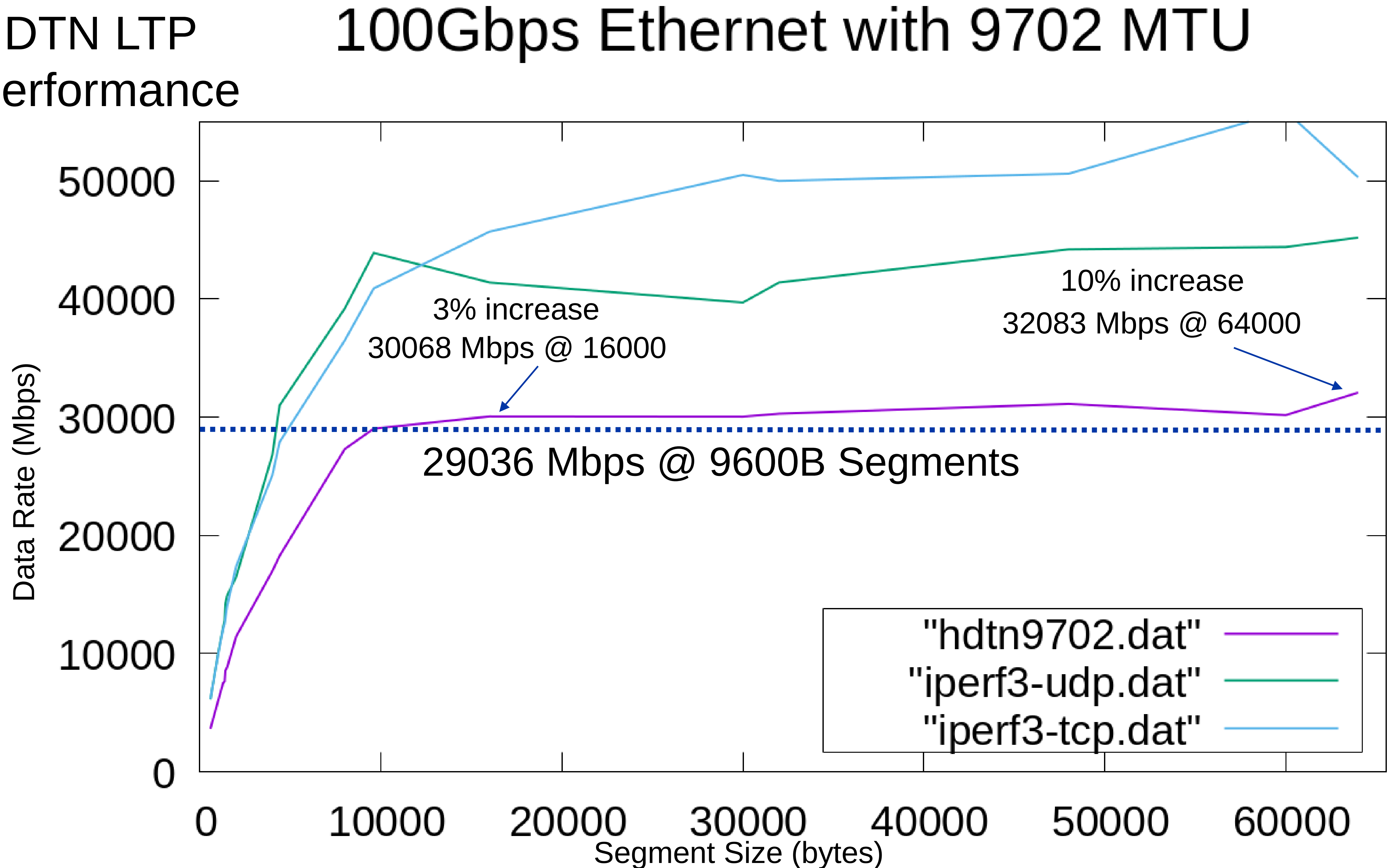


HDTN LTP Performance

100Gbps Ethernet with 4500 MTU



HDTN LTP Performance

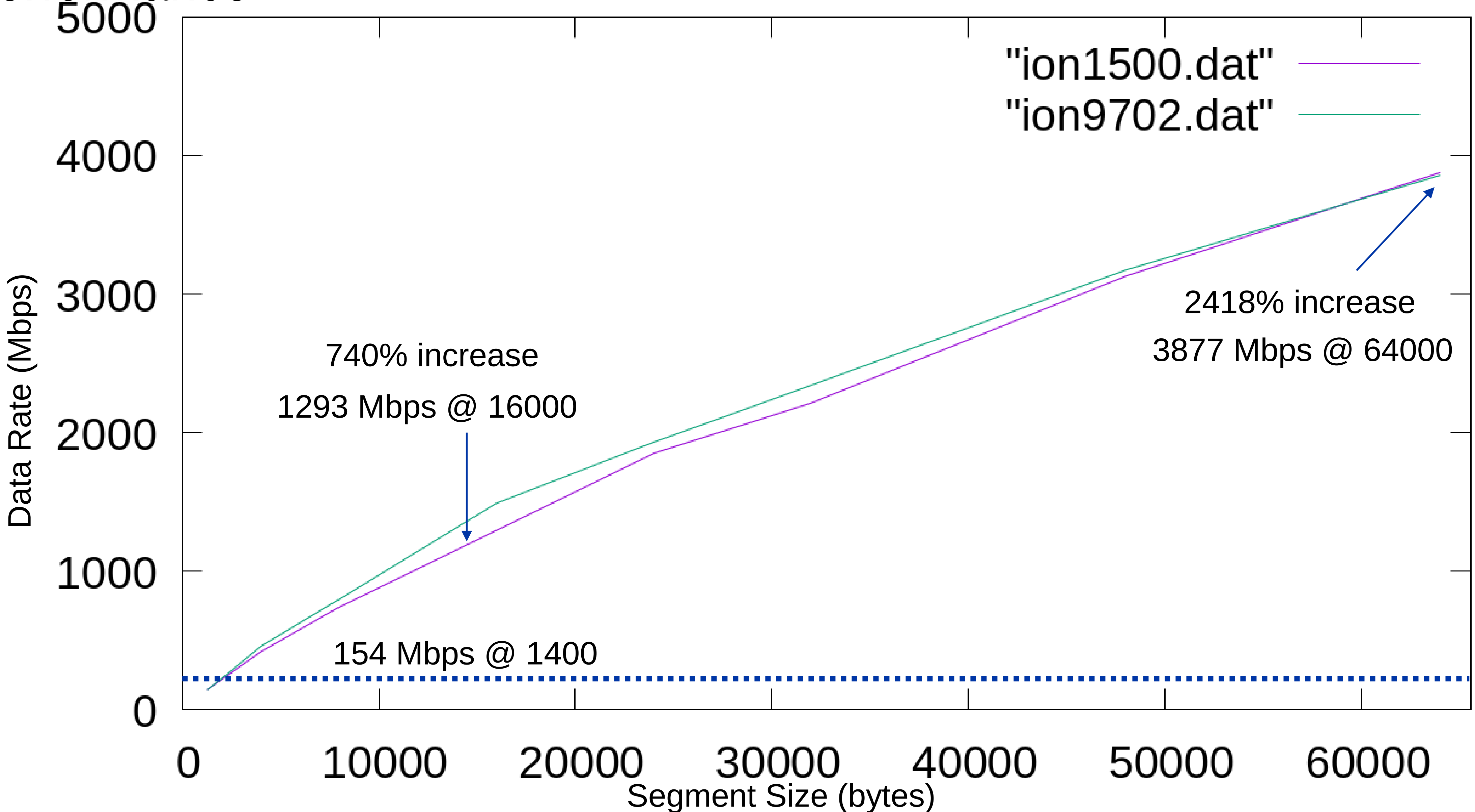


HDTN LTP Performance Implications

- Engages network at high utilization – **good fit for high data rate DTN relay over Laser links**
- For nominal path MTU (1500), performance **more than double** with larger LTP segment sizes that engage IP fragmentation
- For larger path MTUs (4500; 9702), **larger LTP segment sizes provide significant performance gains**; IP fragmentation still provides considerable gains for larger MTUs
- HDTN may benefit from “jumbo” path MTUs larger than 9702
- HDTN may benefit from GSO/GRO – to be investigated

ION DTN LTP Performance

100Gbps Ethernet with ION DTN



ION DTN LTP Performance Implications

- Does not fully engage network at nominal segment sizes, but based on a lightweight multi-processing architecture – **good fit for lower-end links and end systems such as spacecraft**
- Performance profile identical at all path MTUs up to 9702
- **Increasing LTP segment size produces linear performance gains for all sizes with IP fragmentation fully engaged**
- Maximum segment size is currently 64KB; **significant ION performance gains likely at “super-jumbo” segment sizes** (e.g., 256KB; 512KB; 1MB; 10MB, etc.)

IP Fragmentation

- For IPv4, 16-bit Identification can wrap with **reassembly errors possible** even at moderate data rates (IETF RFC4963)
- IPv6 includes 32-bit Identification field, but **this length still too small** if starting sequence number reset frequently
- IPv6 Extended Fragment Header includes **64-bit Identification** field that addresses these issues → **OMNI Interface**
- IP fragmentation only used for segment sizes up to 64KB; larger sizes require **IP Parcels or Advanced Jumbos**
- Dealing with fragment loss; reassembly congestion
  - Destination sends fragmentation report “soft errors” to source
  - Source **adaptively increases or decreases** the size of its packets
  - Supports adaptive packet sizing **on a per-flow granularity**

Adaptation Layer Fragmentation

- OMNI interface exposes an entry point into the Adaptation Layer – a layer below IP
- OMNI interface sets an “unlimited” MTU – this is the size that will be exposed to IP
- Inside the OMNI interface, encapsulation and fragmentation occur at a layer below IP to make sure packets of all sizes get through
- IP layer sees a stable interface that accepts larger packets
- Surrogate OMNI interface developed and tested in Linux kernel; performance evaluation for HDTN and ION TBD

IP Parcels and Advanced Jumbos (AJs)

- Some transport protocols may benefit from segment sizes that exceed 64KB for which fragmentation can't be used
- Peers can use IP Parcels and AJs over paths that support them
- **How large?**
  - IP Parcels include up to 64 64KB segments (4MB)
  - AJs include single segment up to 4GB
- **What about integrity?**
  - Link Layer CRC32 only useful for data sets up to ~9KB
  - Use link-layer CRC32 for headers only, with much stronger end-to-end integrity check
- **What about corruption?**
  - Forward Error Correction (FEC) - sender encodes; receiver decodes
  - End-to-End integrity check determines whether FEC was successful

Segment Size Considerations

- Segment size determines **Retransmission Unit**
  - Loss of single fragment requires retransmission of whole segment
- GSO/GRO employ MTU-sized segments even if path MTU small
  - Loss of single GSO packet requires retransmission of only single packet
- Pragmatic approach:
  - Use large segments only when loss probability small
  - Use FEC to repair damaged segments whenever possible
  - **be adaptive to accommodate changing network conditions**
- Choice between GSO/GRO and IP fragmentation can also be adaptive according to current networking conditions – both tools useful

Future Work

- Evaluate TES-11 results
- LTP analysis in GRC and Boeing labs
- LTP parameter tuning on PC-12 experiments
- Investigate High Performance Reliability Protocol
- Custody transfer versus LTP
- Experiment with Adaptation Layer fragmentation on HDTN; ION
- Experiment with sendmmsg()/recvmmsg() and GSO/GRO in HDTN
- Incorporate Forward Error Correction and large packet sizes

Collaborations and References

- **This work represents the combined efforts of our team, including:**
  - Rachel Dudukovich
  - Daniel Raible
  - Brian Tomko
  - Scott Burleigh
  - Bill Pohlchuck
  - Fred Templin
  - Bhargava Raman Sai Prakash
  - Tom Herbert
- **An earlier version of this work is published in the APNIC Blog at:**
  - https://blog.apnic.net/2024/03/25/delay-tolerant-networking-performance/