A Case for UDP Offload Engines in LambdaGrids

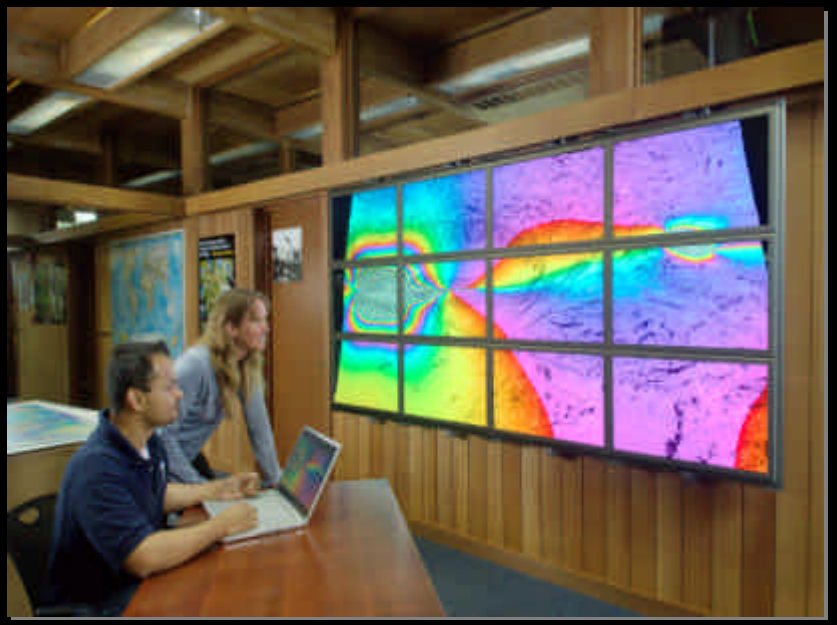
Venkatram Vishwanath, Jason Leigh
Electronic Visualization Laboratory, University of Illinois at Chicago

Pavan Balaji, Dhabaleshwar Panda
Ohio State University

Wu-chun Feng
Virginia Tech

Motivation

- Real-time interactive scientific visualization and high-definition video conferencing require high-throughput, low latency and low jitter data delivery.
- UDP is commonly used for transporting real-time streaming media.
- Trend in large-scale viz is to give users thin low-maintenance clients and have the high resolution visualizations streamed to them from remote supercomputers. E.g. TeraGrid, OptIPuter, IBM Deep Computing Visualization, Sun Ray.



Scripps Institution of Oceanography

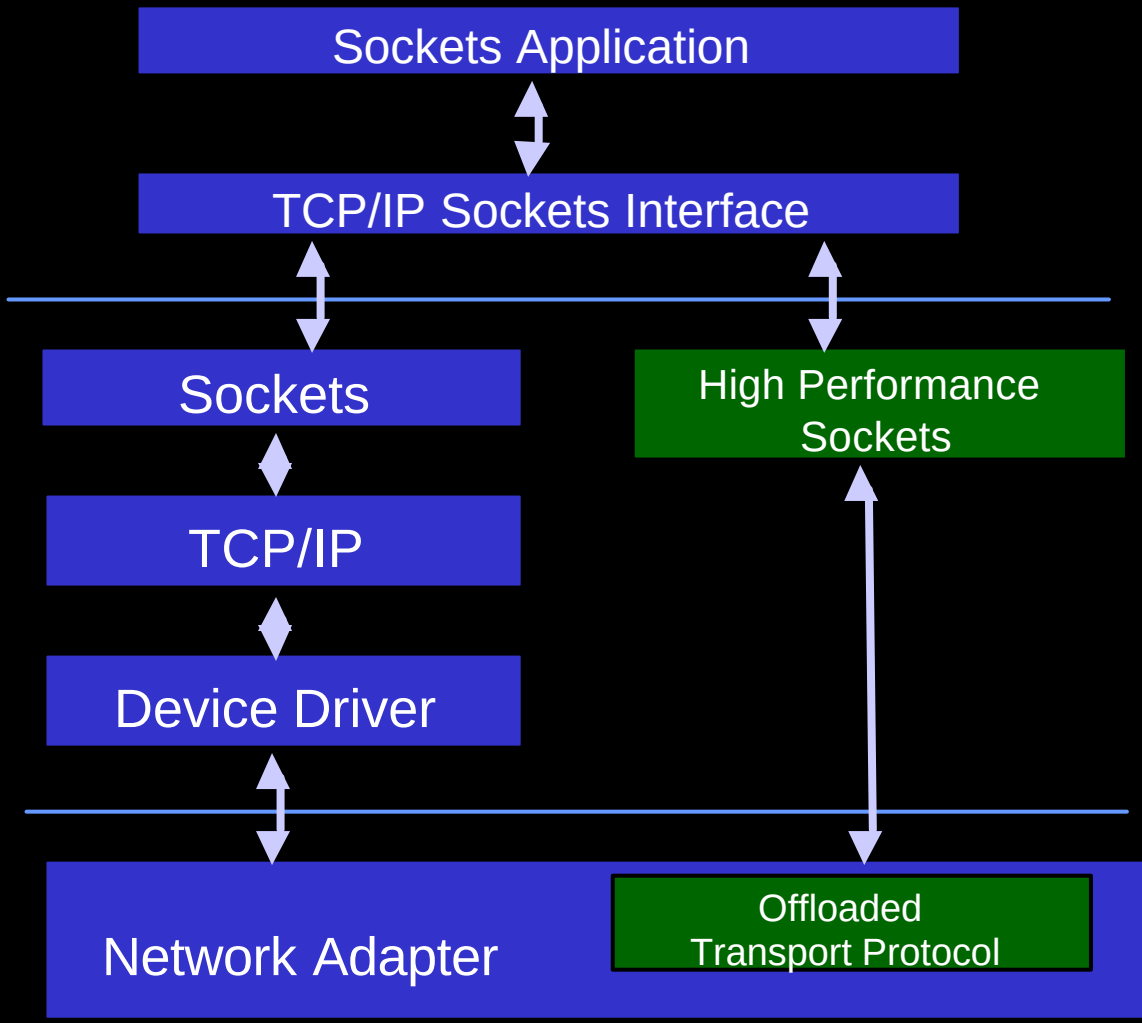
Motivation

- Reliable UDP-based transport protocols (LambdaStream, RBUDP, UDT, Tsunami) have found a home on the LambdaGrid where network paths can be provisioned by applications for sole use.
- But these protocols are CPU intensive.

- While TCP offload is available commercially, equivalent for UDP is not.

High Performance Sockets and Protocol Offload Engines

- Offload the processing of Protocols from CPU to the Network Adaptors
- Similar to a GPU (NPU)
- Currently only TCP Off-load Engines (TOE) are available.



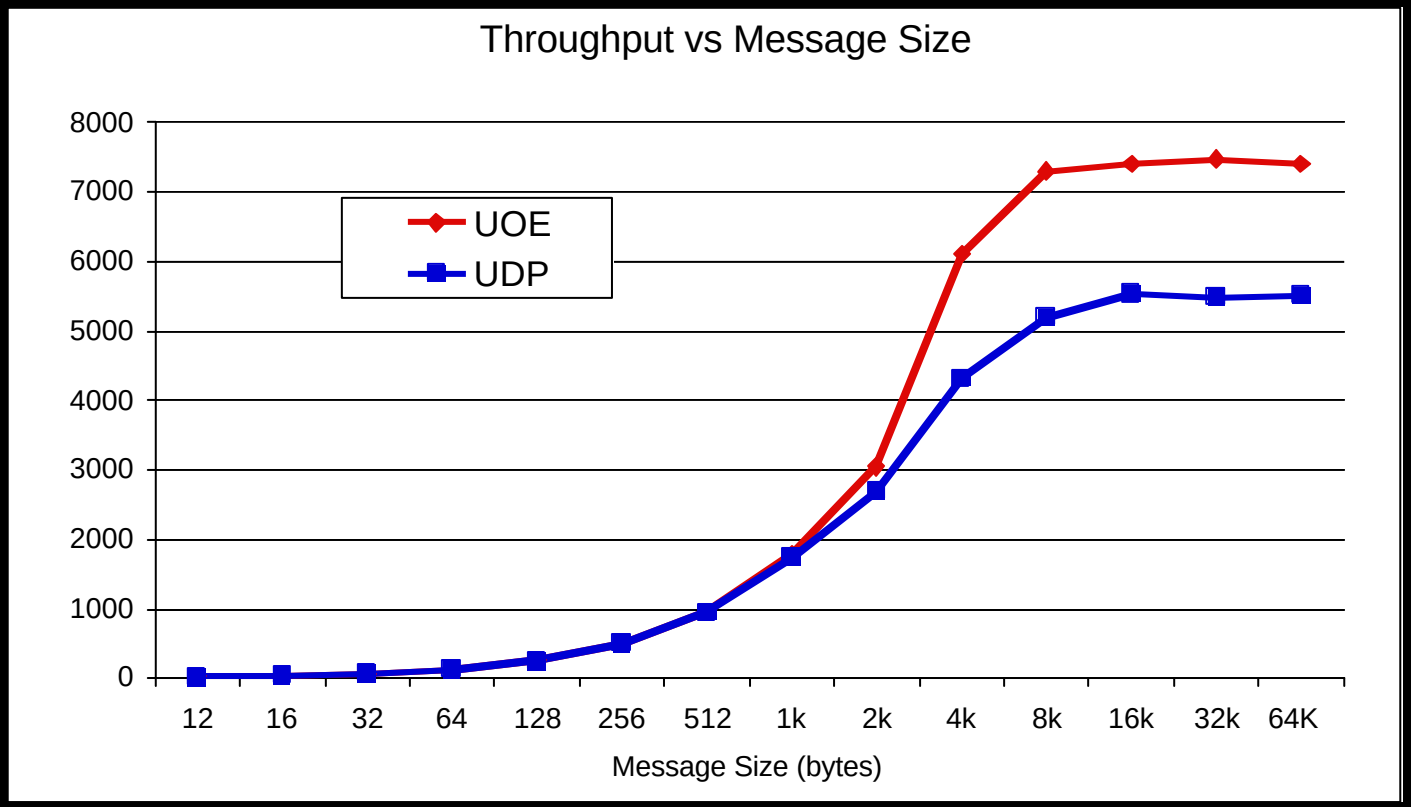
Converting a TCP Off-Load Engine (TOE) to a UDP Off-Load Engine (UOE)

- Disable Congestion Control.
- Change the driver to allow for out of order packets- whereas default TOE will compensate for it.
- Seamlessly hijack an application's UDP calls to use this "alternative":
  - Implement a Connection Management Layer to turn a connectionless protocol (such as UDP) into a connection-oriented protocol (ie TOE).
  - Implement our own Data Management Layer to deal with out of order packets.
- Eliminate TCP's need for Acknowledgments. Not able to completely disable this on Chelsio T110.
- Temporary workaround by setting very large window sizes.

Experiment Test Bed

- Local Area Network
- Dual Opterons 2.4 Ghz
  - 1MB L2 Cache.
  - 4GB 200Mhz DDR SDRAM.
  - Vanilla 2.6.6 – SMP kernel.
  - Chelsio T110
  - Chelsio Driver version 2.1.1

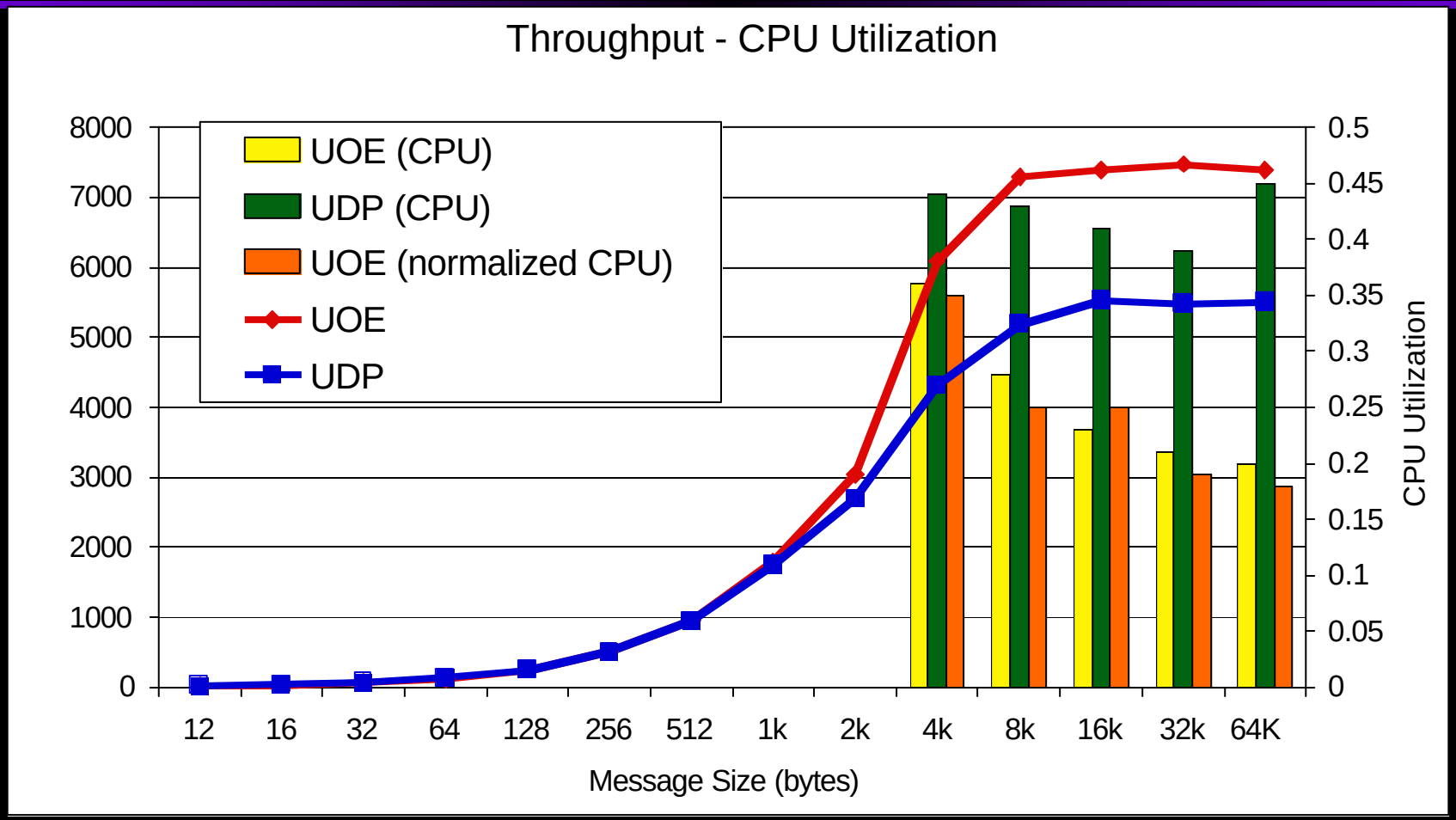
Throughput - UOE vs Traditional Host-based UDP



Initial Iperf results:

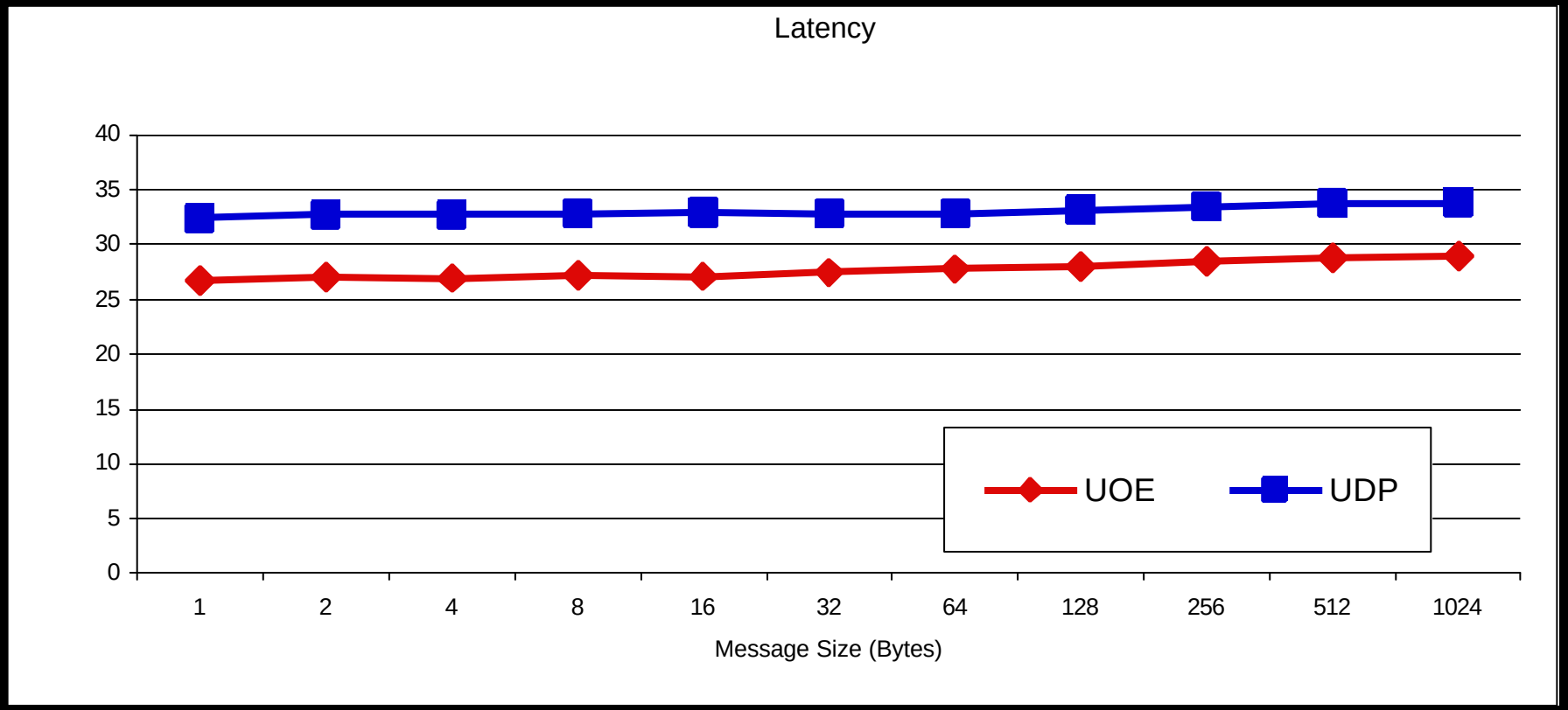
- **7.4 Gbps** Maximum throughput.
- 35% improvement over Host based UDP.

CPU Utilization - UOE vs Host-based UDP



- Up to 50% improvement in CPU utilization.
- Important because in real applications, CPU has other work to do NOT just move the data. E.g. Decode or decrypt the images / data.

Latency



17% improvement in Latency.

Conclusion

- There is a case for them for applications.
- There are real benefits in terms of performance by doing it.
- Especially useful for streaming visualization / high definition video.
- Message from Venkat to Michael Chen :^)
  - I want: T210s - replacement for T110s. It is the lower power version and much smaller.
  - When will PCI-Express NICs be available? Netereon is already available.
  - Would like to correspond more deeply with Chelsio engineers to resolve ACK issues.

Future Work

- Compare with Partial Offload and quantify / identify what is the most useful part of a stack to offload. E.g. offload checksum of packet rather than the entire UDP stack.
- Conduct MAN and WAN Area trials.
- Compare with other Partial offload NICs such as Neterion.
- Implementation & Comparison on Myrinet 10G.
- Apply UOE implementation to a currently existing UDP-based transport protocol like LambdaStream, RBUDP, UDT, Tsunami, etc..

Thank You

- Chelsio Engineers
- National Science Foundation:
  - CNS-0224306 Research Resources: Matching Visualization & Intelligent Data Mining to Experimental Networks
  - CNS-0420477 LambdaVision (Major Research Instrumentation)
  - OCI-0229642 StarLight: Strategic Technologies for Internet Discovery and Development (STI)
  - OCI-0441094 TransLight/StarLight
  - OCI-0225642 The OptIPuter
- National Institutes of Health, the State of Illinois, the Office of Naval Research
- NTT Optical Network Systems Laboratory in Japan