Server-Efficient High-Definition Media Dissemination

Philip W. Frey 1), Andreas Hasler 1), Bernard Metzler 1) and Gustavo Alonso 2)

1) Systems Department | IBM Research GmbH
2) Systems Group | Department of Computer Science | ETH Zurich

June 16, 2009
Media Consumption Trends

- **Personalized & interactive services**
  - Unicast
  - Video-on-Demand

- **High-Definition Media**
  - Significantly higher bit rates
  - Widespread flat rate broadband Internet access

[1] V. Saxena (Comcast Cable): Bandwidth drivers for 100 G Ethernet
So what is the Issue?

- **Orthogonal service requirements**
  - Throughput & latency

- **Challenges for the service providers: many users!**
  - High aggregate data throughput
  - Unpredictable user behavior

- **Most prevalent solutions fail to fully utilize available link bandwidth!**
  - **Network** is no longer bottleneck
  - Significant **I/O overhead** on server machines
    (intermediate **data copying**)
Our proposal

- **Goal:** Fully utilize 10GbE link with minimal server I/O cost
  - Serve HD media to **maximum number of clients**
  - Offer convenient **on-demand** service
  - Utilize even **faster** links in the future

- **Data transfer using** Remote Direct Memory Access (**RDMA**)
  - Minimal I/O overhead (CPU & memory bus)
  - Client-driven protocol
  - Minimal **control** overhead
Overview

- **Prominent Solutions**
  - RTP over UDP
  - HTTP over TCP

- **Reducing the I/O Overhead - RDMA Transport**
  - Background & Technology Overview
  - iWARP/RDMA-based Media Dissemination Protocol

- **Server Scalability**
  - Host overhead
  - VCR-like Media Control

- **Summary & Outlook**
Testbed

- **Network**
  - 10 Gb/s Ethernet

- **Machines**
  - IBM HS21 BladeServers (Quad Core 2.33GHz Intel Xeon, 8GB memory)
  - Linux 2.6.27 (Fedora Core 9)

- **Codec**
  - H.264
    - Bit rates: ~1 Mbps (SD) … ~9 Mbps (Full HD)

- **Server software**
  - RTP: VLC media player, Apple Darwin Streaming Server
  - HTTP: Apache web server
Prominent Solution 1 - RTP over UDP

- **UDP + RTP + RTCP + RTSP**
  - Unidirectional transport
  - Feedback channel
  - Out-of-band session control

- **Overhead**
  - Data copying
  - Packetizing
  -> **CPU & memory bus** load!

Only ~15% link utilization due to excessive server I/O overhead!
Prominent Solution 2 - HTTP over TCP

- **TCP + HTTP**
  - Bidirectional (Sockets)
  - Not designed for VoD

- **Potential Advantages**
  - Standard software
  - Firewall (port 80)

- **Overhead**
  - Data copying
  - \(~ 70\%\) link utilization
  - \(\rightarrow\) Sendfile: \(100\%\) link utilization!
User Interactivity

- **Sendfile**
  - Part of mainline Linux kernel
  - Utilize 10GbE link capacity
  - Data from HDD
  - Depending on user interactivity
  - Stateful server
  - Zero-copy only on TX side
Overview

- Prominent Solutions
  - RTP over UDP
  - HTTP over TCP

- Reducing the I/O Overhead - RDMA Transport
  - Background & Technology Overview
  - iWARP/RDMA-based Media Dissemination Protocol

- Server Scalability
  - Host overhead
  - VCR-like Media Control

- Summary & Outlook
Reducing the I/O Overhead - RDMA Transport

- Semantic Benefits
  - Direct access to main memory of remote host
  - One-sided operations
  - Asynchronous interface

- Performance Benefits
  - Zero-Copy & Kernel Bypassing
  - Minimal CPU & memory bus load
  - HW accelerated (RNIC)
  - iWARP = RDMA over Ethernet
RDMA Operations during Data Exchange

Local Process

... 
\texttt{wait( src = D, dst = D )} 
\texttt{send( src = S )} 
... 

Remote Process

... 
\texttt{receive( dst = D )} 
... 

Network
iWARP/RDMA-based Media Dissemination Protocol

- **Buffer advertisement**

- **Client reads from server**
  - Pull-based
  - One-sided RDMA Read operation
  - True remote random access

- **VCR-like media control**
  - Play, pause, skip, change media, etc.
  - In-band (no control channel)
  - ‘Stateless’ server
Overview

- **Prominent Solutions**
  - RTP over UDP
  - HTTP over TCP

- **Reducing the I/O Overhead - RDMA Transport**
  - Background & Technology Overview
  - iWARP/RDMA-based Media Dissemination Protocol

- **Server Scalability**
  - Host overhead
  - VCR-like Media Control

- **Summary & Outlook**
iWARP/RDMA Server Scalability

- **Parameters**
  - Bit rate
  - Seek rate
  - Number of different movies
  - Movie length (min)
  - Client buffer size

- **Excellent scalability**
  - Utilize full 10GbE
  - Minimal overhead
  - Maximum number of clients
Overview

- **Prominent Solutions**
  - RTP over UDP
  - HTTP over TCP

- **Reducing the I/O Overhead - RDMA Transport**
  - Background & Technology Overview
  - iWARP/RDMA-based Media Dissemination Protocol

- **Server Scalability**
  - Host overhead
  - VCR-like Media Control

- **Summary & Outlook**
iWARP-based Media Dissemination - Summary

✓ Scalable
  ✓ Maximum copy avoidance
  ✓ Minimal memory bus load
  ✓ Minimal state on server

✓ True remote random access
  ✓ Responsive and efficient VCR-like media control (for free)
  ✓ In-bound media control

✓ Easy to apply (~300 LOC)

✗ Special RDMA NIC to get full performance ($800)
✗ Data must reside in main memory

⇒ Minimal server overhead ➔ Maximum #clients

© 2009 IBM Corporation
Outlook & Future Work

- **Local buffer replacement strategies**
  - Local pyramid broadcast

- **Software iWARP stack on clients**

- **Extension to Live Streaming**
  - Special case of VoD
  - Push-based protocol using RDMA write
  - Update protocol

- **P2P Overlay (?)**
Take Home Message

- **HD media dissemination poses challenge to server infrastructure**
  - High aggregate throughput
  - Dynamic and unpredictable user interaction
  - **Massive I/O overhead!**

- **UDP no longer the obvious choice**

- **RDMA offers significant server I/O reduction**
  - Scalability: HW acceleration, copy avoidance
  - VCR-like control: One-sided, asynchronous operations