UC San Diego

CSE 291: Operating Systems in Datacenters

Amy Ousterhout

Oct. 10, 2023

UC San Diego

Agenda for Today

- Announcement and reminders
- Introduction to RDMA and RPCs
- FaRM discussion
- Where do research ideas come from?

Announcements and Reminders

- Course Feedback #FinAid
  - UCSD requirement
  - Due Friday 10/13
  - See Canvas -> Assignments
- Warm-up assignment
  - Due Monday 10/16 at 11:59 pm
- Research Project
  - Start thinking about:
    - Who you want to work with (1-3 people)
    - What topics interest you?

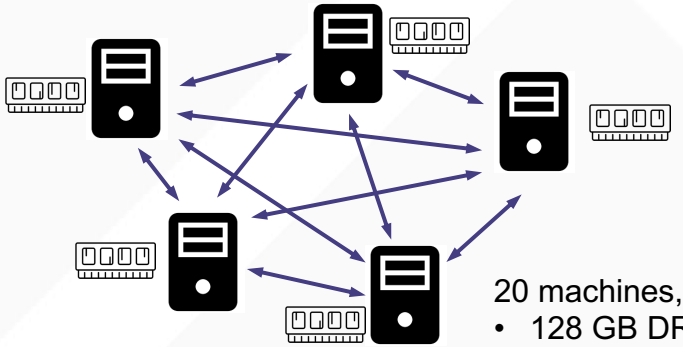
UC San Diego

RDMA and RPCs

The Shift Towards Storing Data in Memory

- Disk is a poor fit for modern datacenter applications
  - Disk is much slower to access than memory (10 ms vs. 100 ns)
  - Datacenter workloads require random access
- RAM (random-access memory) is becoming much cheaper
- Feasible to store a significant fraction (or all of) your app's data in memory, distributed across a cluster

- 500 GB disk
- 800 Mbps
- 10 ms latency

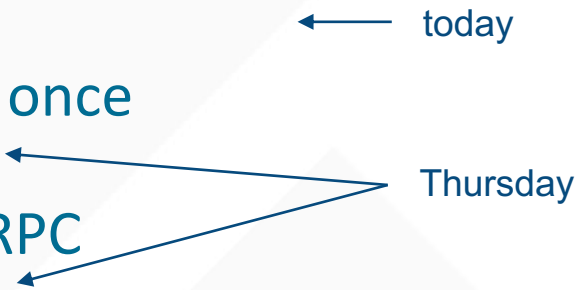


20 machines, each with:
- 128 GB DRAM
- 40 Gbps
- 10 µs latency

Higher throughput, lower latency

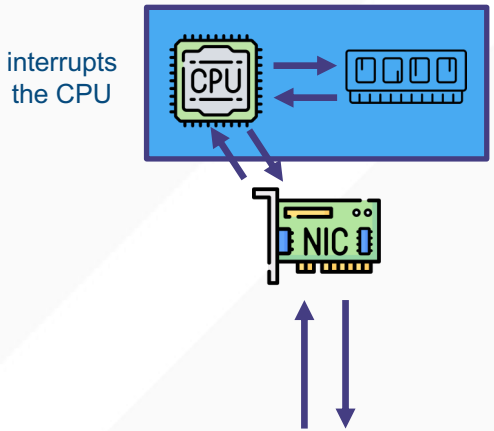
How Should Programs Access Remote Memory?

- Access data one word at a time, similar to local memory
  - `movl remote_addr %eax`
- Access a chunk of data at once (e.g., 64 bytes, 1 KB)
  - FaRM
- Access multiple dependent chunks of data at once
  - PRISM
- Execute a function on the remote server via RPC
  - eRPC



CPU-Based Memory Access vs. RDMA

CPU-Based Memory Access

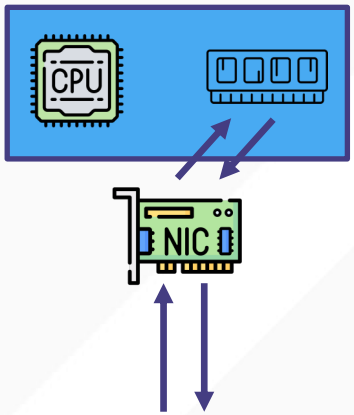


read data using TCP
(e.g., Linux, IX, XDP)

"RPC" or
"two-sided RDMA"

Remote Direct Memory Access

- Access memory directly from the NIC



doesn't involve
the CPU!

read data using RDMA

"one-sided RDMA"

RDMA – An Old Technology

- First proposed in 1993
- Used in super computers (HPC) for many years
- Relied on Infiniband in the 2000s
  - Lossless network
  - Expensive
- RoCE (~2010)
  - RDMA over Converged Ethernet, pronounced “rocky”
  - Provides a reliable network and enables RDMA over regular Ethernet
  - Cheaper than Infiniband
  - Made it easier to adopt RDMA in datacenters

How RDMA is Used Today

- In private datacenters
  - By 2015 Microsoft was using RDMA in clusters with latency-sensitive services
- In public clouds
  - Alibaba uses RDMA within its storage clusters
  - Azure uses RDMA to communicate between VMs and storage clusters within regions (about 70% of Azure traffic)
- In multi-tenant settings
  - Google deployed their own variant of RDMA (1RMA)

UC San Diego

FaRM Discussion