Arrakis

CSE 451: Operating Systems
Autumn 2013

Irene Zhang
iyzhang@cs.washington.edu
Arrakis

An **OS** that provides **low-latency, high-throughput** network I/O for **datacenter applications** using **hardware virtualization**.
The OS is a bottleneck for many datacenter applications.

Many applications are I/O-intensive,

- memcached, key-value stores
- load balancer and other middleboxes

And datacenters have high-speed networks.

- 10-100 Gb/s
- Infiniband, fiber and other technologies
Some numbers ...

Time through 100 ft of cable: \textbf{100 ns}
Time through a 10G switch: \textbf{350 ns}
Time for memcached to respond: \textbf{1 us}
Time for OS to process a packet: \textbf{6 us}
Why is the OS so slow?

The OS provides

• multiplexing

• isolation

• higher level abstractions (sockets)

which have become relatively slow.
Key Idea

Hardware virtualization provides fast multiplexing and isolation in the networking card allowing us to eliminate the OS from most network I/O.
Virtualization

- Technology that allows more than one OS to share one physical machine.
- Each OS runs in its own virtual machine.
- Virtual machine monitor multiplexes the hardware among virtual machines.
Network Virtualization

- VMM multiplexes one network card among several virtual machines.
- Each virtual machine gets one virtual network card (vNIC).
- Guest OS still runs a network driver and TCP stack.
Hardware Virtualization

- Network card directly exposes a number of vNICs.
- VMM allocates one vNIC to each VM.
- Guest OS directly talks to networking card without VMM interposition on each packet.
Arrakis OS

• OS directly exposes vNICs to applications.
• Each application runs a tiny user-level TCP stack with no multiplexing.
• No intervention from the OS (kernel crossings/copying) for each network packet.
Arrakis

Network stack: .2 us

App
libos

App
libos

Network stack: .2 us

Control Plane

Kernel

User

vNIC

vNIC

Switch

Network stack: .2 us

NIC
Some more numbers ...

Time through 100 ft of cable: \textbf{100 \text{ ns}}

Time through a 10G switch: \textbf{350 \text{ ns}}

Time for memcached to respond: \textbf{1 \text{ us}}

Time for OS to process a packet: \textbf{6 \text{ us}}

Time for Arrakis to process a packet: \textbf{200 \text{ ns}}
Interested?

Come work with us!

Or learn more:

http://arrakis.cs.washington.edu