Shoal: A Network Architecture for Disaggregated Racks

Vishal Shrivastav (Cornell University)
Asaf Valadarsky (Hebrew University of Jerusalem)
Hitesh Ballani, Paolo Costa (Microsoft Research)
Ki Suh Lee (Waltz Networks)
Han Wang (Barefoot Networks)
Rachit Agarwal, Hakim Weatherspoon (Cornell University)
Traditional racks in datacenters
Disaggregated racks in datacenters

- Intra-rack Network
  - NVMe
  - Accelerators (FPGA, GPU, TPU)
  - Storage
  - SoCs

- Inter-rack DC Network
  - I/O controllers
  - CPU
  - Memory
  - NIC
Disaggregated racks in datacenters

Prior works [OSDI'16] [HPCA'12] [Keeton’15]
- High compute density
- Fine-grained resource pooling and provisioning
- Seamless scaling and independent evolution of resources

Intra-rack Network

Inter-rack DC Network

I/O controllers
- CPU
- Memory
- NIC

NVMe

Storage

SoCs

Accelerators (FPGA, GPU, TPU)
Disaggregated racks in datacenters

Prior works [OSDI’16] [HPCA’12] [Keeton’15]
- High compute density
- Fine-grained resource pooling and provisioning
- Seamless scaling and independent evolution of resources
Challenges for disaggregated rack network

• Connect as many as an order of magnitude more nodes than traditional racks

- Be high performant
  - low latency / high throughput

- Be power efficient
  - to enable high compute density
Challenges for disaggregated rack network

• Connect as many as an order of magnitude more nodes than traditional racks

~15KW power budget [NSDI’16]

- Be high performant
  - low latency / high throughput

- Be power efficient
  - to enable high compute density
Challenges for disaggregated rack network

• Connect as many as an order of magnitude more nodes than traditional racks

- Intra-rack Network

- Be high performant
  - low latency / high throughput

- Be power efficient
  - to enable high compute density

~15KW power budget [NSDI’16]
Challenges for disaggregated rack network

- Connect as many as an order of magnitude more nodes than traditional racks

- Intra-rack Network
  - Be high performant
    - low latency / high throughput
  - Be power efficient
    - to enable high compute density

~15KW power budget [NSDI'16]
Challenges for disaggregated rack network

• Connect as many as an order of magnitude more nodes than traditional racks

~15KW power budget
[NSDI’16]

- Be high performant
  - low latency / high throughput

- Be power efficient
  - to enable high compute density
Potential disaggregated rack network designs

<table>
<thead>
<tr>
<th></th>
<th>Low Power consumption</th>
<th>High Performance (low latency / high throughput)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Packet-switched</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Networks</td>
<td><img src="image" alt="ToR chassis switch" /></td>
<td><img src="image" alt="Network of switches" /></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Direct-connect</strong></td>
<td><img src="image" alt="Checkmark" /></td>
<td><img src="image" alt="X" /></td>
</tr>
<tr>
<td>Networks</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Shoal is a network stack and fabric for disaggregated racks that is both low power and high performance (low latency, high throughput)
Shoal is a network stack and fabric for disaggregated racks that is both low power and high performance (low latency, high throughput)

Key feature:
Shoal network fabric comprises purely fast circuit switches that can reconfigure within nanoseconds
Shoal is a network stack and fabric for disaggregated racks that is both **low power** and **high performance** (low latency, high throughput)

**Key feature:**
Shoal network fabric comprises purely *fast circuit switches* that can reconfigure within nanoseconds
Goal 1: Low power consumption

Circuit switches
- No buffering
- No packet processing
- No serialization/de-serialization

Consumes significantly less power than packet switches
Goal 2: High network performance

**Key Challenge:**
Need to explicitly set up circuits (reconfigure) before sending packets

- **Traditional circuit-switched networks**
  - Uses switches with high reconfiguration delay, up to milliseconds
  - Uses a central controller to decide the circuits (reconfiguration algorithm)
  - Not suitable for low latency traffic

- **Shoal**
  - Leverages circuit switches with nanosecond reconfiguration delay

**Key Design Idea:**
De-centralized, traffic agnostic reconfiguration algorithm
- Inspired from LB monolithic packet switches [Comp Comm’02]
Shoal for a single circuit switch network
Shoal for a single circuit switch network

Static pre-defined schedule
Shoal for a single circuit switch network

N-1 time slots (an epoch)

Static pre-defined schedule
Shoal for a single circuit switch network

N-1 time slots (an epoch)

Static pre-defined schedule
Shoal for a single circuit switch network

A permutation of connections
N-1 time slots (an epoch)

Time slot

A permutation of connections
N-1 time slots (an epoch)

Static pre-defined schedule
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Time slot

Static pre-defined schedule
Shoal for a single circuit switch network

A permutation of connections

<table>
<thead>
<tr>
<th>Time slot</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>4</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>6</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>7</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
</tbody>
</table>

N-1 time slots (an epoch)

Static pre-defined schedule

![Diagram of Shoal for a single circuit switch network](image-url)
Shoal for a single circuit switch network

A permutation of connections  N-1 time slots (an epoch)

Time slot

Static pre-defined schedule
(a cyclic permutation)

A
B
C
D
E
F
G
H

1 2 3 4 5 6 7

A B C D E F G H
B C D E F G H A
C D E F G H A B
D E F G H A B C
E F G H A B C D
F G H A B C D E
G H A B C D E F
H A B C D E F G
Shoal for a single circuit switch network

**A permutation of connections**

**N-1 time slots (an epoch)**

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td></td>
</tr>
</tbody>
</table>

**Static pre-defined schedule**

*(a cyclic permutation)*

Uniformly load-balanced traffic
Shoal for a single circuit switch network

A permutation of connections  N-1 time slots (an epoch)

Time slot

Static pre-defined schedule (a cyclic permutation)

Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Time slot

1 2 3 4 5 6 7

A B C D E F G H
B C D E F G H A
C D E F G H A B
D E F G H A B C
E F G H A B C D
F G H A B C D E
G H A B C D E F
H A B C D E F G

Static pre-defined schedule
(a cyclic permutation)

Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Static pre-defined schedule
(a cyclic permutation)

100% throughput

Uniformly load-balanced traffic

(a cyclic permutation)
Shoal for a single circuit switch network

A permutation of connections

<table>
<thead>
<tr>
<th>Time slot</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
</tbody>
</table>

N-1 time slots (an epoch)

Static pre-defined schedule

(a cyclic permutation)

Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Time slot

1 2 3 4 5 6 7

A B C D E F G H
B C D E F G H A
C D E F G H A B
D E F G H A B C
E F G H A B C D
F G H A B C D E
G H A B C D E F
H A B C D E F G

Static pre-defined schedule (a cyclic permutation)

Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Time slot

1 2 3 4 5 6 7

A B C D E F G H
B C D E F G H A
C D E F G H A B
D E F G H A B C
E F G H A B C D
F G H A B C D E
G H A B C D E F
H A B C D E F G

Static pre-defined schedule (a cyclic permutation)

Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Time slot

1 2 3 4 5 6 7

A: B C D E F G H
B: C D E F G H A
C: D E F G H A B
D: E F G H A B C
E: F G H A B C D
F: G H A B C D E
G: H A B C D E F
H: A B C D E F G

Static pre-defined schedule
(a cyclic permutation)

Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections

\[
\begin{array}{cccccccc}
A & B & C & D & E & F & G & H \\
B & C & D & E & F & G & H & A \\
C & D & E & F & G & H & A & B \\
D & E & F & G & H & A & B & C \\
E & F & G & H & A & B & C & D \\
F & G & H & A & B & C & D & E \\
G & H & A & B & C & D & E & F \\
H & A & B & C & D & E & F & G \\
\end{array}
\]

Time slot

N-1 time slots (an epoch)

Uniformly load-balanced traffic

100% throughput

(a cyclic permutation)
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Time slot

1 2 3 4 5 6 7

A B C D E F G H
B C D E F G H A
C D E F G H A B
D E F G H A B C
E F G H A B C D
F G H A B C D E
G H A B C D E F
H A B C D E F G

Static pre-defined schedule
(a cyclic permutation)

Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Time slot

A B C D E F G H
B C D E F G H A
C D E F G H A B
D E F G H A B C
E F G H A B C D
F G H A B C D E
G H A B C D E F
H A B C D E F G

Static pre-defined schedule

(a cyclic permutation)

Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Time slot

1 2 3 4 5 6 7

A

B

C

D

E

F

G

H

A B C D E F G H
C D E F G H A
D E F G H A B
E F G H A B C
F G H A B C D
G H A B C D E
H A B C D E F
A B C D E F G

Static pre-defined schedule
(a cyclic permutation)

Arbitrary traffic pattern

Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Static pre-defined schedule (a cyclic permutation)

Arbitrary traffic pattern

Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections
N-1 time slots (an epoch)

Time slot

1 2 3 4 5 6 7

A | B | C | D | E | F | G | H
B | C | D | E | F | G | H | A
C | D | E | F | G | H | A | B
D | E | F | G | H | A | B | C
E | F | G | H | A | B | C | D
F | G | H | A | B | C | D | E
G | H | A | B | C | D | E | F
H | A | B | C | D | E | F | G

Static pre-defined schedule (a cyclic permutation)

Arbitrary traffic pattern
Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Static pre-defined schedule (a cyclic permutation)

Each node has N-1 queues (one per dst)

100% throughput

Arbitrary traffic pattern

Uniformly load-balanced traffic
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Time slot

1 2 3 4 5 6 7

A B C D E F G H
B C D E F G H A
C D E F G H A B
D E F G H A B C
E F G H A B C D
F G H A B C D E
G H A B C D E F
H A B C D E F G

Static pre-defined schedule (a cyclic permutation)

Each node has N-1 queues (one per dst)

Arbitrary traffic pattern

Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Static pre-defined schedule
(a cyclic permutation)

Time slot

1 2 3 4 5 6 7

A B C D E F G H
B C D E F G H A
C D E F G H A B
D E F G H A B C
E F G H A B C D
F G H A B C D E
G H A B C D E F
H A B C D E F G

Each node has N-1 queues (one per dst)

A -> H

Arbitrary traffic pattern

Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Each node has N-1 queues (one per dst)

Static pre-defined schedule
(a cyclic permutation)

Time slot

\begin{array}{cccccccc}
  1 & 2 & 3 & 4 & 5 & 6 & 7 \\
\end{array}

\begin{array}{cccccccc}
  A & B & C & D & E & F & G & H \\
  B & C & D & E & F & G & H & A \\
  C & D & E & F & G & H & A & B \\
  D & E & F & G & H & A & B & C \\
  E & F & G & H & A & B & C & D \\
  F & G & H & A & B & C & D & E \\
  G & H & A & B & C & D & E & F \\
  H & A & B & C & D & E & F & G \\
\end{array}

100% throughput

Arbitrary traffic pattern

Uniformly load-balanced traffic
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Time slot

1 2 3 4 5 6 7

A B C D E F G H
B C D E F G H A
C D E F G H A B
D E F G H A B C
E F G H A B C D
F G H A B C D E
G H A B C D E F
H A B C D E F G

Static pre-defined schedule *(a cyclic permutation)*

Each node has N-1 queues (one per dst)

100% throughput
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Time slot

1 2 3 4 5 6 7

A
B
C
D
E
F
G
H

A permutation of connections

N-1 time slots (an epoch)

A
B
C
D
E
F
G
H

A
B
C
D
E
F
G
H

A
B
C
D
E
F
G
H

A
B
C
D
E
F
G
H

Each node has N-1 queues (one per dst)

A -> H

A -> H

Static pre-defined schedule (a cyclic permutation)

Arbitrary traffic pattern

Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Time slot

A permutation

N-1 time slots

(an epoch)

Static pre-defined schedule

(a cyclic permutation)

Each node has
N-1 queues
(one per dst)

Arbitrary traffic pattern

Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Static pre-defined schedule (a cyclic permutation)

Each node has N-1 queues (one per dst)

Arbitrary traffic pattern

Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

A -> H
A -> H
A -> H
A -> H

Each node has N-1 queues (one per dst)

Static pre-defined schedule (a cyclic permutation)

Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
</tbody>
</table>

N-1 time slots (an epoch)

- Each node has N-1 queues (one per dst)

Static pre-defined schedule (*a cyclic permutation*)

- Arbitrary traffic pattern
- Uniformly load-balanced traffic
- 100% throughput
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Time slot

Static pre-defined schedule *(a cyclic permutation)*

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
</tbody>
</table>

Each node has N-1 queues (one per dst)

 Arbitrary traffic pattern

Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections

<table>
<thead>
<tr>
<th>Time slot</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
</tbody>
</table>

N-1 time slots (an epoch)

Each node has N-1 queues (one per dst)

Static pre-defined schedule (a cyclic permutation)

Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Static pre-defined schedule *(a cyclic permutation)*

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
</tbody>
</table>

Each node has N-1 queues (one per dst)

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
</tbody>
</table>

100% throughput

Arbitrary traffic pattern

Uniformly load-balanced traffic
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Time slot

1 2 3 4 5 6 7

A B C D E F G H
B C D E F G H A
C D E F G H A B
D E F G H A B C
E F G H A B C D
F G H A B C D E
G H A B C D E F
H A B C D E F G

Static pre-defined schedule (a cyclic permutation)

Each node has N-1 queues (one per dst)

A permutation

of connections

N-1 time slots
(an epoch)

1 2 3 4 5 6 7

A B C D E F G H
B C D E F G H A
C D E F G H A B
D E F G H A B C
E F G H A B C D
F G H A B C D E
G H A B C D E F
H A B C D E F G

 Arbitrary traffic pattern

Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Static pre-defined schedule (a cyclic permutation)

Each node has N-1 queues (one per dst)

100% throughput

Arbitrary traffic pattern

Uniformly load-balanced traffic
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Static pre-defined schedule (a cyclic permutation)

Time slot

1 2 3 4 5 6 7

A B C D E F G H
B C D E F G H A
C D E F G H A B
D E F G H A B C
E F G H A B C D
F G H A B C D E
G H A B C D E F
H A B C D E F G

Each node has N-1 queues (one per dst)

Arbitrary traffic pattern

Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections
N-1 time slots (an epoch)

Static pre-defined schedule
(a cyclic permutation)

Each node has N-1 queues (one per dst)

Arbitrary traffic pattern

Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Time slot

1 2 3 4 5 6 7

A
B
C
D
E
F
G
H

Static pre-defined schedule (a cyclic permutation)

Each node has N-1 queues (one per dst)

 Arbitrary traffic pattern

Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Static pre-defined schedule
(a cyclic permutation)

Time slot

1 2 3 4 5 6 7

A B C D E F G H
B C D E F G H A
C D E F G H A B
D E F G H A B C
E F G H A B C D
F G H A B C D E
G H A B C D E F
H A B C D E F G

Each node has N-1 queues (one per dst)

100% throughput
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Time slot:

1 2 3 4 5 6 7

A B C D E F G H
B C D E F G H A
C D E F G H A B
D E F G H A B C
E F G H A B C D
F G H A B C D E
G H A B C D E F
H A B C D E F G

Static pre-defined schedule
(a cyclic permutation)

Each node has N-1 queues
(one per dst)

Arbitrary traffic pattern

Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections
N-1 time slots (an epoch)

Static pre-defined schedule (a cyclic permutation)

Each node has N-1 queues (one per dst)

Arbitrary traffic pattern
Uniformly load-balanced traffic

100% throughput
Shoal for a single circuit switch network

A permutation of connections

N-1 time slots (an epoch)

Time slot

1 2 3 4 5 6 7
A B C D E F G H
B C D E F G H A
C D E F G H A B
D E F G H A B C
E F G H A B C D
F G H A B C D E
G H A B C D E F
H A B C D E F G

Static pre-defined schedule (a cyclic permutation)

Each node has N-1 queues (one per dst)

Arbitrary traffic pattern

Uniformly load-balanced traffic

100% throughput
50% throughput in worst-case
Extending Shoal to a network of circuit switches

<table>
<thead>
<tr>
<th>Time slot</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
</tbody>
</table>
Extending Shoal to a network of circuit switches

A non-blocking topology of circuit switches
Extending Shoal to a network of circuit switches

A non-blocking topology of circuit switches
Extending Shoal to a network of circuit switches

Requires very tight network-wide synchronization

- DTP [Sigcomm’16] + WhiteRabbit can achieve sub-nanosecond synchronization precision

A non-blocking topology of circuit switches
## Congestion in Shoal

<table>
<thead>
<tr>
<th>Time slot</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
<tr>
<td>B</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>D</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>E</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>G</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>H</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
</tbody>
</table>

The diagram illustrates the connections between different elements labeled A to H, showing how they interact with each other over time slots.
## Congestion in Shoal

### Time slot

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
</tbody>
</table>

### Diagram

- Flow to H
- Flow to H
Congestion in Shoal

<table>
<thead>
<tr>
<th>Time slot</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
<tr>
<td>B</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>D</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>E</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>G</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>H</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
</tbody>
</table>

Flow to H  Flow to H

Flow to H  Flow to H
Congestion in Shoal

<table>
<thead>
<tr>
<th>Time slot</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
</tbody>
</table>

Flow to H
Flow to H
B -> H
Congestion in Shoal

<table>
<thead>
<tr>
<th>Time slot</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>5</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>7</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>8</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
</tbody>
</table>

Flow to H  Flow to H

A -> H
Congestion in Shoal

Time slot

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
</tbody>
</table>

Flow to H
Flow to H
A -> H
B -> H
# Congestion in Shoal

<table>
<thead>
<tr>
<th>Time slot</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
</tbody>
</table>

The diagram illustrates the flow to different destinations, indicated by the arrows. For example, there is a flow from A to H and from B to H.
Congestion in Shoal

A
B
C
D
E
F
G
H

Flow to H
Flow to H

A -> H

A
B
C
D
E
F
G
H
Congestion in Shoal

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
</tbody>
</table>

Flow to H
Flow to H

A -> H
# Congestion in Shoal

## Time Slot Table

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
</tbody>
</table>

## Diagram

- **Flow to H**:
  - From A
  - From B

- **A -> H**:
- **B -> H**:

---

Note: The diagram includes nodes labeled A, B, C, D, E, F, G, and H, with connections indicating flow directions.
Congestion in Shoal

Time slot

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
</tbody>
</table>

Flow to H
Flow to H

A -> H
B -> H
Congestion in Shoal

Time slot

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>3</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>6</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>7</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
</tbody>
</table>

Flow to H

Flow to H

A -> H

A -> H
Shoal proposes a novel congestion control algorithm for a fast circuit-switched network
## Congestion control in Shoal

<table>
<thead>
<tr>
<th>Time slot</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td></td>
</tr>
</tbody>
</table>
Congestion control in Shoal

<table>
<thead>
<tr>
<th>Time slot</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
</tbody>
</table>
# Congestion control in Shoal

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong></td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td><strong>B</strong></td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td><strong>C</strong></td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td><strong>D</strong></td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td><strong>E</strong></td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td><strong>F</strong></td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td><strong>G</strong></td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td><strong>H</strong></td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
</tbody>
</table>

![Graph showing the congestion control process in Shoal](image)
## Congestion control in Shoal

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Time slot

Diagram showing the flow from A to C.
## Congestion control in Shoal

<table>
<thead>
<tr>
<th>Time slot</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td></td>
</tr>
</tbody>
</table>

A → C

Diagram showing the connections between A and C.
## Congestion control in Shoal

<table>
<thead>
<tr>
<th>Time slot</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong></td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td><strong>B</strong></td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td><strong>C</strong></td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td><strong>D</strong></td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td><strong>E</strong></td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td><strong>F</strong></td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td><strong>G</strong></td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td><strong>H</strong></td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
</tbody>
</table>

Queue for destination H at C:

- B -> H
Congestion control in Shoal

<table>
<thead>
<tr>
<th>Time slot</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
</tbody>
</table>

Queue for destination H at C

A -> H

B -> H

A -> H

B -> H
Congestion control in Shoal

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
</tbody>
</table>

Queue for destination H at C

A -> H
B -> H
A -> H
B -> H
Congestion control in Shoal

<table>
<thead>
<tr>
<th>Time slot</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
</tbody>
</table>

A

C

Queue for destination H at C

B -> H

A -> H  B -> H

A -> H
Congestion control in Shoal

<table>
<thead>
<tr>
<th>Time slot</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
</tbody>
</table>

Queue for destination H at C

A: [A -> H]
B: [B -> H]
C: [A -> H, B -> H]
D: [A -> H]
E: [A -> H]
Congestion control in Shoal

<table>
<thead>
<tr>
<th>Time slot</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>F</td>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>G</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>H</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
</tr>
</tbody>
</table>

Queue for destination H at C

at most 1 packet per source
Congestion control in Shoal

Each per-destination queue $Q_i$ corresponding to destination $i$ is bounded!

$$\text{len}(Q_i) \leq 1 + \text{incast\_degree}(i)$$ packets
Key properties of Shoal

- No central controller for reconfiguration
  - Fully de-centralized, traffic agnostic reconfiguration logic
  - Allows circuit switches to reconfigure at nanosecond timescales

- Each per-destination queue in the network is bounded

- Each packet traverses the network at most twice
  - Worst-case 50% throughput compared to an ideal packet-switched network
  - Can be compensated by allocating 2X bandwidth per node
  - Cost (Shoal) $\leq$ Cost (packet-switched network with $\frac{1}{2}$ bandwidth of Shoal)
Implementation

- Stratix V FPGA
- Bluespec System Verilog

- Implemented custom NIC and circuit switch on FPGA

Circuit switch implementation can reconfigure in < 6.4ns

Verified the queuing and throughput properties of Shoal on a 8-node testbed
Evaluation

- **Power consumption**

  For a 512-node rack

  - Packet-switched network comprises 24 64x50 Gbps packet switches
  - Shoal comprises 48 64x50 Gbps circuit switches

<table>
<thead>
<tr>
<th>Network</th>
<th>Power Consumption (KW)</th>
<th>% of Rack Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Packet-switched Network</td>
<td>8.72</td>
<td>(58%)</td>
</tr>
<tr>
<td>Shoal</td>
<td>2.55</td>
<td>(17%)</td>
</tr>
</tbody>
</table>

- Shoal consumes 3.5x less power than packet-switched network!
Evaluation

- **Power consumption**

For a 512-node rack

- Packet-switched network comprises **24 64x50 Gbps packet switches**
- Shoal comprises **48 64x50 Gbps circuit switches**

<table>
<thead>
<tr>
<th>Network</th>
<th>Power Consumption</th>
<th>Budget Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Packet-switched Network</td>
<td>8.72 KW</td>
<td>(58% of rack budget)</td>
</tr>
<tr>
<td>Shoal</td>
<td>2.55 KW</td>
<td>(17% of rack budget)</td>
</tr>
</tbody>
</table>

- Shoal consumes 3.5x less power than packet-switched network!
Evaluation

Network performance

- Packet-level simulator in C
- 512-node rack
- 5 disaggregated workload traces [OSDI’16]
- Shoal has 2X bandwidth (with comparable cost)

- Shoal performs comparable or better than several recent designs for packet-switched networks!

![Graph showing network performance comparisons](image)
## Conclusion

<table>
<thead>
<tr>
<th></th>
<th>Low Power consumption</th>
<th>High Performance (low latency / high throughput)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Packet-switched Networks</strong></td>
<td><img src="image1" alt="ToR chassis switch" /></td>
<td><img src="image2" alt="Network of switches" /></td>
</tr>
<tr>
<td><strong>Direct-connect Networks</strong></td>
<td><img src="image3" alt="Direct-connect networks" /></td>
<td><img src="image4" alt="Direct-connect networks" /></td>
</tr>
<tr>
<td><strong>Shoal (circuit-switched)</strong></td>
<td><img src="image5" alt="Shoal" /></td>
<td><img src="image6" alt="Shoal" /></td>
</tr>
</tbody>
</table>
Thank you!

Shoal FPGA prototype and simulator code is available at:

https://github.com/vishal1303/Shoal