Analysis of Space-Space-Space Clos-Network Packet Switch

Eiji Oki*, Nattapong Kitsuwan*, and Roberto Rojas-Cessa**
*The University of Electro-Communications
**New Jersey Institute of Technology
Outline

- Introduction
- Previous works on Clos-network switches
- Random selection algorithm with multiple iterations
- Analytical model and assumption
- Analysis
  - One iteration
  - Multiple iterations
- Summary
Introduction

- High-speed IP routers highly demanded
  - Exponential growth of Internet traffic
  - DWDM carries tremendous traffic
- Requirement for high-speed IP routers
  - Support a large number of ports
  - Support high port speed (e.g., >= 40 Gbit/s)
- Switch architecture
  - Single stage
  - Multiple stages
Single stage vs. multiple stages

- Single stage: less control complexity, but more components
  - Switch elements are proportional to $N^2$ ($N$: switch size).

- Multiple stage: less components, but more control complexity
  - Control complexity depends on where buffers are allocated.
  - Route selection function from 1$^{\text{st}}$ stage to 2$^{\text{nd}}$ stage.
Previous works on Clos-network switch

- **Memory-Memory-Memory (MMM)**
  - [Chaney, Infocom97] etc.
  - All stages have buffers.
  - Use random dispatching
  - Require re-sequence function at output port: bottleneck

- **Memory-Space-Memory (MSM)**
  - [Chiussi, ComMag97], [Oki, TON02] etc.
  - Re-sequence function is not required.
  - As input and output port speed increases, implementing memory can be bottleneck.
Previous works on Clos-network switch (cont’d)

- Space-Space-Space ($S^3$)
  - [Chao, ComMag03], [Li, ICC05], [Pun, Globecom02] etc.
  - No buffer needs to be implemented in any stage.
  - Route selection from 1$^{st}$ stage to 2$^{nd}$ stage is required to avoid contention.
  - Adopting an appropriate route selection algorithm is a key to succeed $S^3$. 
Route selection algorithm for $S^3$

- Static round-robin [Pun, globecom02]
  - Maximal matching
  - 100% throughput under uniform traffic
- Random section [Li, ICC05]
  - One of the maximal matching algorithms that can be studied as it can be modeled theoretically.
  - Li et al. analyzed switch throughput with single iteration
    - $(1-1/e)^2 = 39\%$
  - However,
    - The derived formula gives a coarse approximation and underestimates the switch throughput.
    - No theoretical analysis on the switch throughput with multiple iterations.
- We introduce a theoretical throughput analysis on $S^3$ with multiple iteration.
S\textsuperscript{3} switch model with Virtual Output Queue (VOQ)
Staged random selection with multiple iterations

Phase 1
Matching within IM

Phase 2
CM matching

Phase 3
OM matching

$i$ iterations

$j$ iterations
Example of staged random selection

Phase 1
Example of staged random selection

Phase 1 → Phase 2 → Phase 3

Iteration $i = 1$

VOQ(1, 1, 1, 1)

VOQ(1, 3, 1, 1)

VOQ(2, 3, 3, 1)

VOQ(3, 2, 2, 2)

Iteration $i = 2$

VOQ(2, 3, 3, 2)

VOQ(3, 1, 2, 1)

Iteration $j = 3$

VOQ(1, 3, 1, 2)

VOQ(1, 1, 1, 2)

VOQ(2, 3, 3, 2)

VOQ(3, 2, 2, 2)

Iteration $j = 2$

VOQ(1, 2, 1, 1)

VOQ(1, 2, 3, 3)

VOQ(2, 2, 3, 3)

VOQ(3, 3, 2, 3)

Iteration $j = 2$

VOQ(2, 1, 3, 1)

VOQ(2, 1, 3, 1)

VOQ(2, 1, 3, 1)
Terminology

IM output-link load \( \rho_i(n,m) \)

CM output-link load \( \rho_C(n,m,i) \)

OM output-link load \( \rho_O(n,m,i,j) \)

\( \rho_I(n,m) \) is the switch throughput.

Phase 1: Matching within IM

Phase 2: CM matching

Phase 3: OM matching

\( i \) iterations

\( j \) iterations
Our analytical model

- Our assumptions
  - Each IM, CM, or OM behaves independently of the other modules
  - Each output-link load depends on the previous-stage output link load.
- [Li, ICC05] assumed that contention at each stage occurs independently.

\[ \text{IM}(i), \text{CM}(i), \text{OM}(i) \]
Analysis on switch throughput with one iteration

\[ \rho_c(n,m,k,1) = 1 - \left(1 - \frac{\rho_I(n,m)}{k}\right)^k \]

\[ \rho_I(n,m) = \frac{n}{m} \]

\[ \rho_O(n,m,k,1,1) = 1 - \left(1 - \frac{\rho_C(n,m,k)}{n}\right)^m \]

Phase 1
Matching within IM

Phase 2
IM-CM matching

Phase 3
OM matching
Switch throughput with one iteration

- The switch throughput approaches \((1-1/e) = 0.632\) as \(m/n\) increases.
  - [Li, ICC05] underestimates the throughput as \((1-1/e)^2 = 0.400\).
- Our model provides the same result as simulation.
Analysis on switch throughput with multiple iteration

\[
\rho_o(n,m,k,i,j) = \sum_{n_m(1)=0}^{n_u(0)} g(i,1,n_u(0),n_m(1)) \times \\
\left[ \frac{n_m(1)}{n_u(0)} + \sum_{n_m(2)=0}^{n_u(1)} g(i,2,n_u(1),n_m(2)) \times \\
\left[ \frac{n_m(2)}{n_u(0)} + \sum_{n_m(3)=0}^{n_u(2)} g(i,3,n_u(2),n_m(3)) \frac{n_m(3)}{n_u(0)} + \\
\cdots \\
\left[ \frac{n_m(i)}{n_u(0)} + \sum_{n_m(i)=1}^{n_u(i-1)} g(i,j,n_u(j-1),n_m(j)) \frac{n_m(i)}{n_u(0)} \right] \\
\cdots \right] \\
\right]
\]
Switch throughput with multiple iterations

- The switch throughput increases with the number of OM iterations, \( j \).
- All observed differences between the analytical and simulation results are smaller than 1%.
Summary

- A theoretical throughput analysis of the staged random selection algorithm with multiple iterations in a $S^3$ Clos-network switch architecture is introduced.

- The model showed that, by increasing the number of IM-CM and OM iterations, the switch throughput achieves 100% without the internal switch expansion under uniform traffic.

- Our analytical model was confirmed to be valid for a throughput analysis, via simulation.