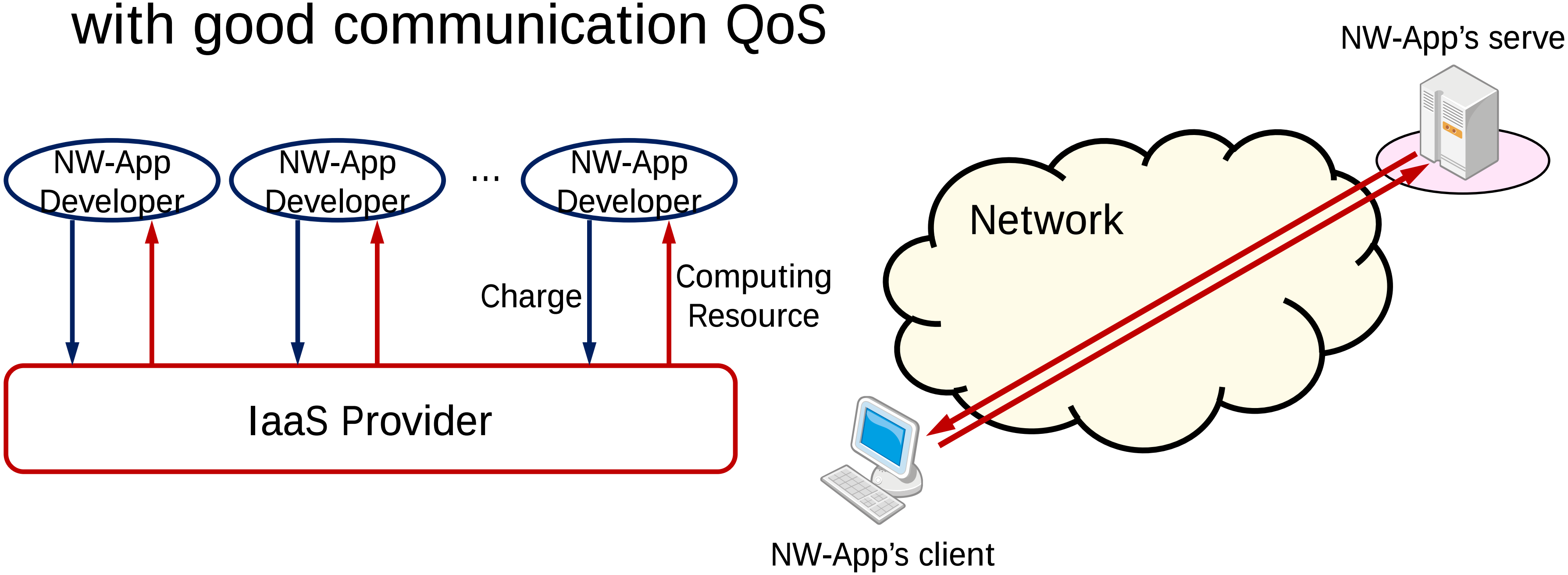
Destination Selection Algorithm in a Server Migration Service

Asato Yamanaka[1], Yukinobu Fukushima[1],
Tutomu Murase[2], Tokumi Yokohira[1] and Tatsuya Suda[3]

The Graduate School of Natural Science and Technology,
Okayama Univirsity, Japan[1]
Cloud System Research Laboratories, NEC Corporation, Japan[2]
University Netgroup Inc, USA[3]

Background (1/2)

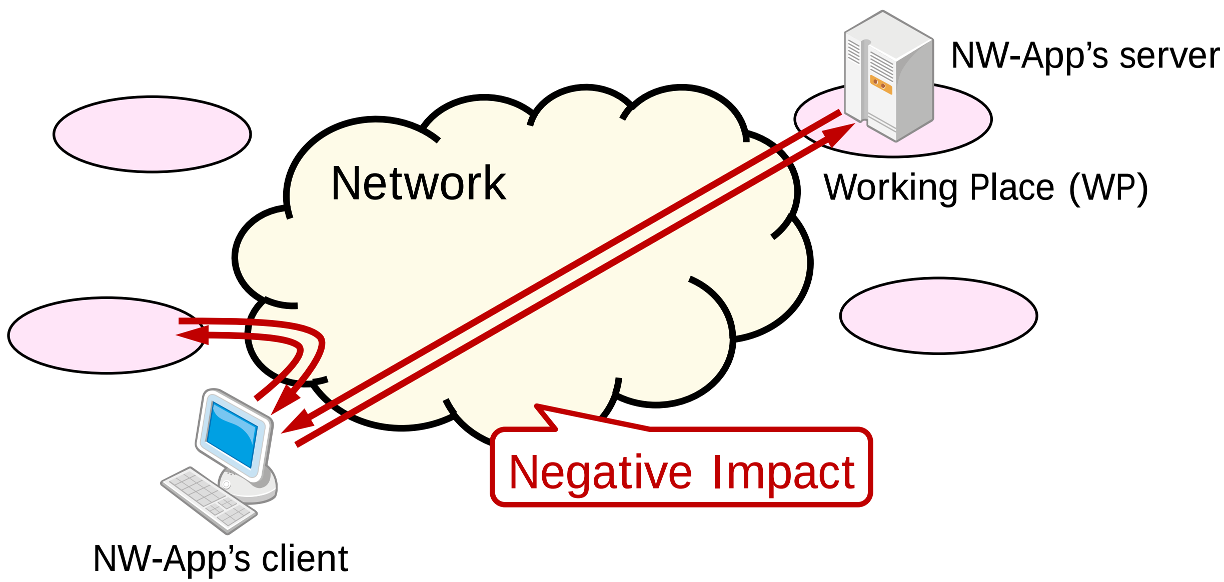
- IaaS cloud service is attracting attention
  - NW-Apps (e.g., online games application) can operate their servers at a data center without any initial cost
- Location of server is fixed at a data center
  ⇨ It is difficult for an IaaS provider to provide NW-Apps with good communication QoS



Background (2/2)

- Server migration service (SMS) can improve communication QoS in IaaS cloud service
- In SMS, we need to
  - Satisfy SLA of as many clients as possible
  - Decrease the migrating server's negative impact (network impact) on its background traffic

by appropriately determining where to locate servers



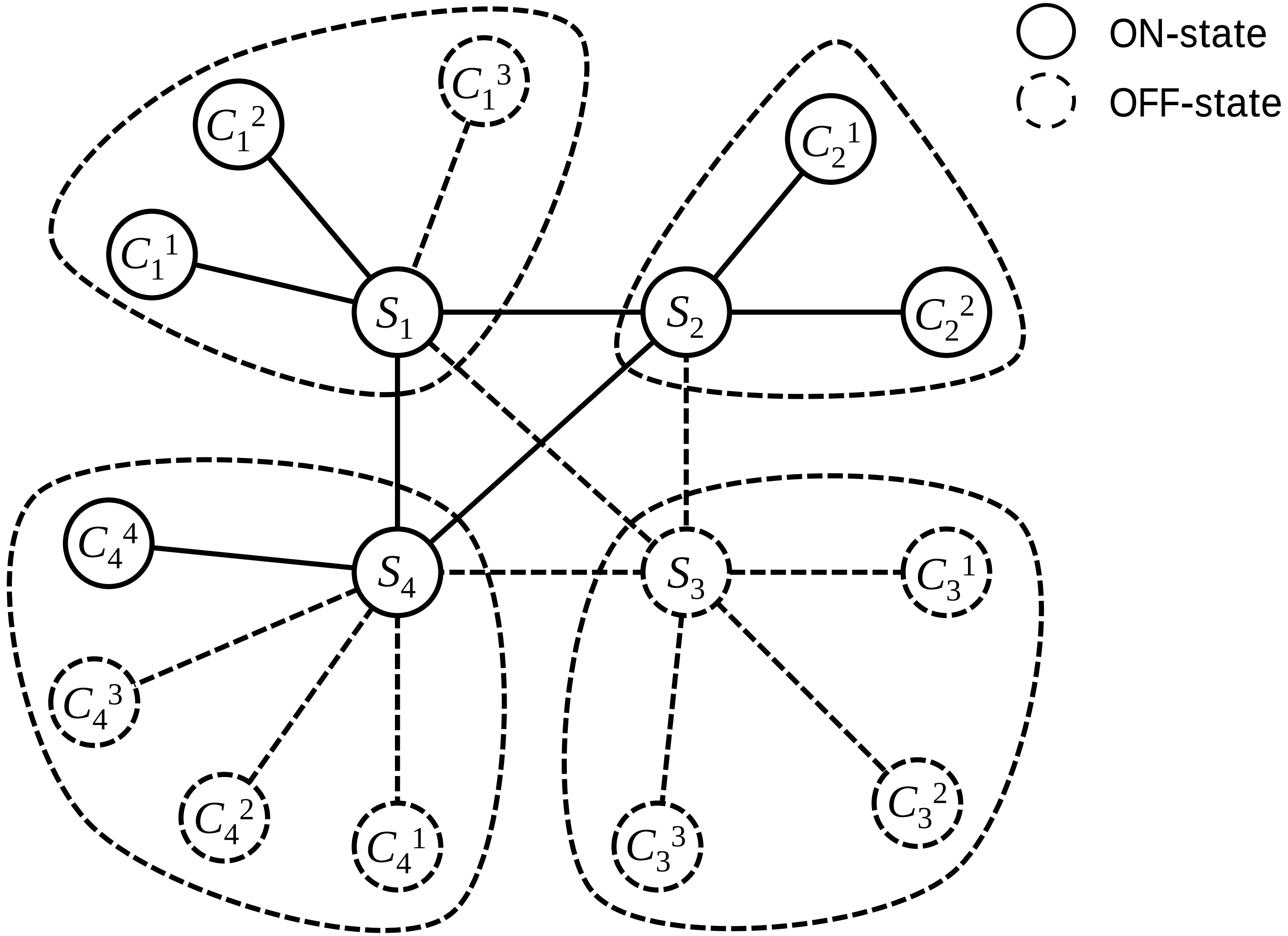
Research Objective

- Previous research of server migration
  - Live migration of VMs [1]
    - Decreasing downtime of a migrating server
    - No consideration of network impact
    - Server's destination is determined in advance

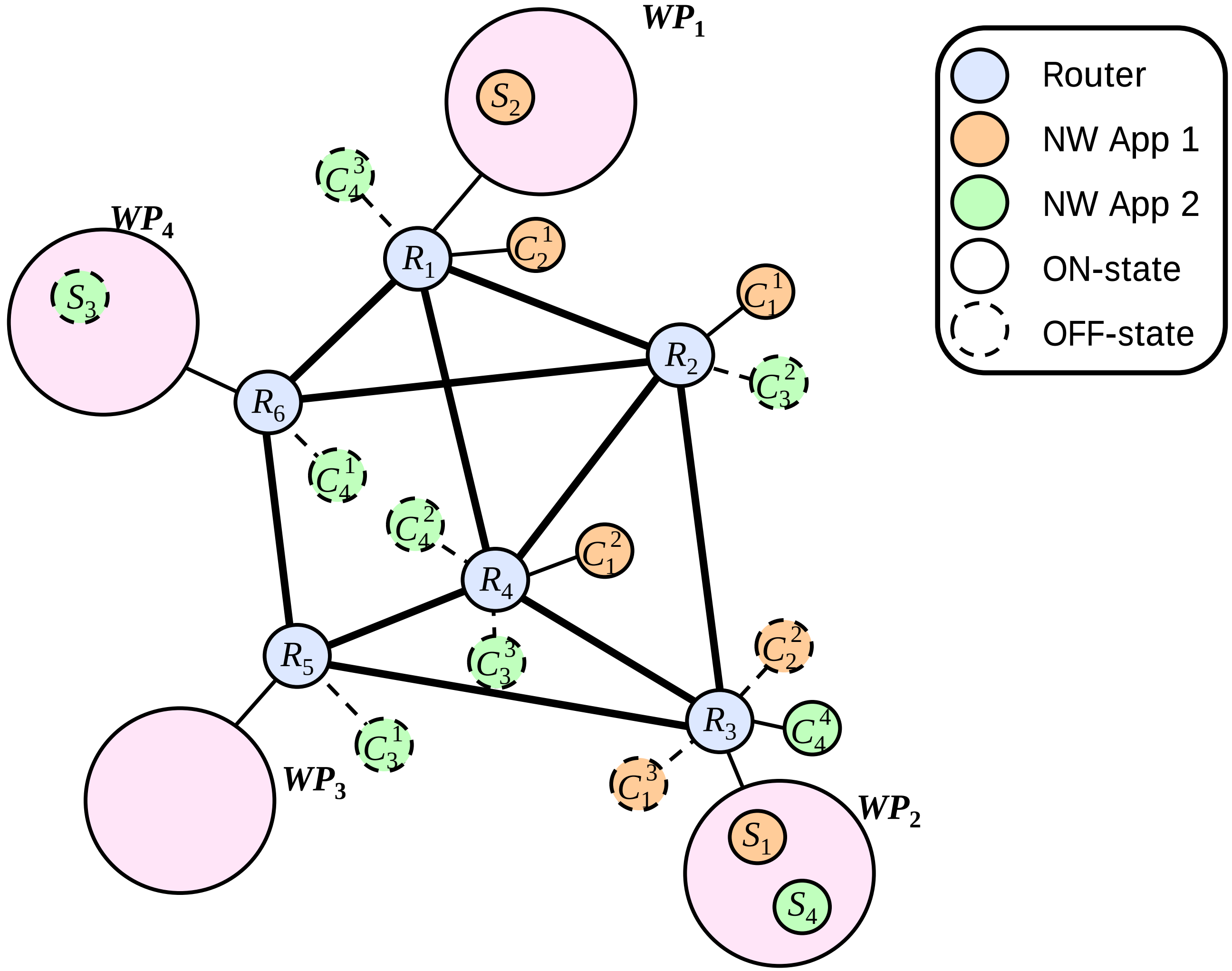
- Research objective
  - Proposing destination selection algorithms that try to decrease network impact while satisfying SLA of as many clients as possible

[1] C. Clark, K. Fraser, S. Hand, J. G. Hansen, E. Jul, C. Limpach, I. Pratt, and A. Wareld, "Live Migration of Virtual Machines," in Proceedings of NSDI, May 2005.

Model of Network Application (NW-App)



Model of Network Environment



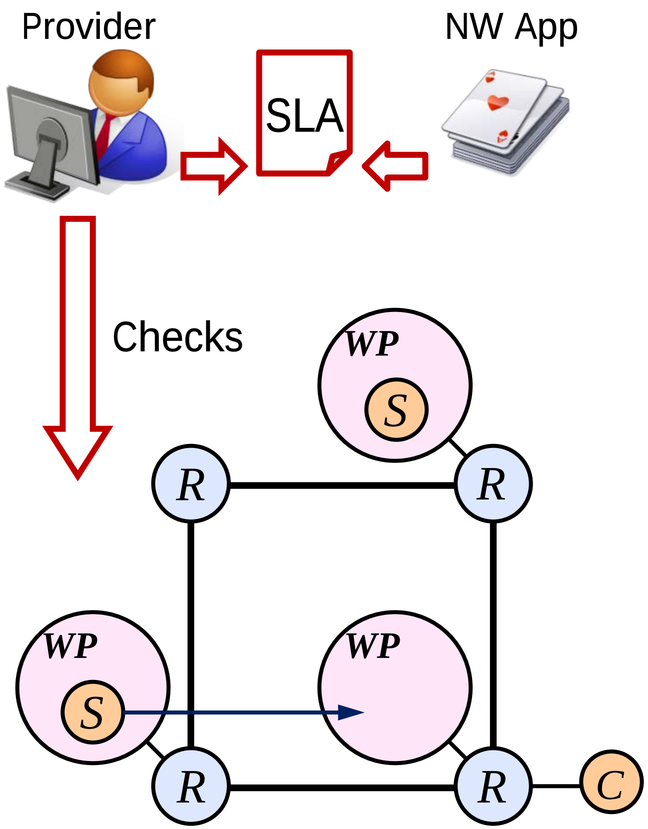
Procedures of Server Migraion Service

NW-App developer and the server migration service provider make an SLA about communication QoS

The provider always checks whether the SLAs (delays, throughputs, packet loss, etc) are satisfied or not

Not satisfied

The provider executes destination selection algorithm, determines destination WP and moves server in the new WP (Only one server can migrate)

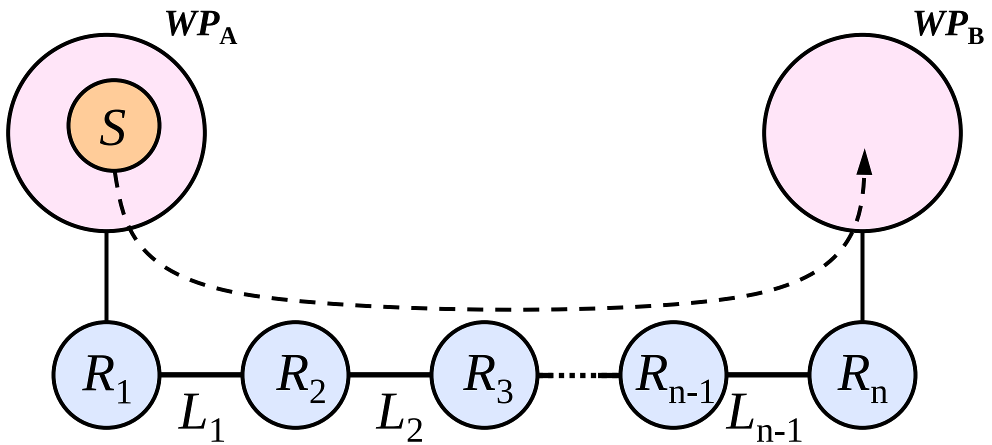


Quantification of Network Impact

- Definition
  - Number of bits of background traffic that suffer negative effect for the period in which server migration traffic is being transmitted

$$NI = \sum_{i=1}^{n-1} A_i T_i$$

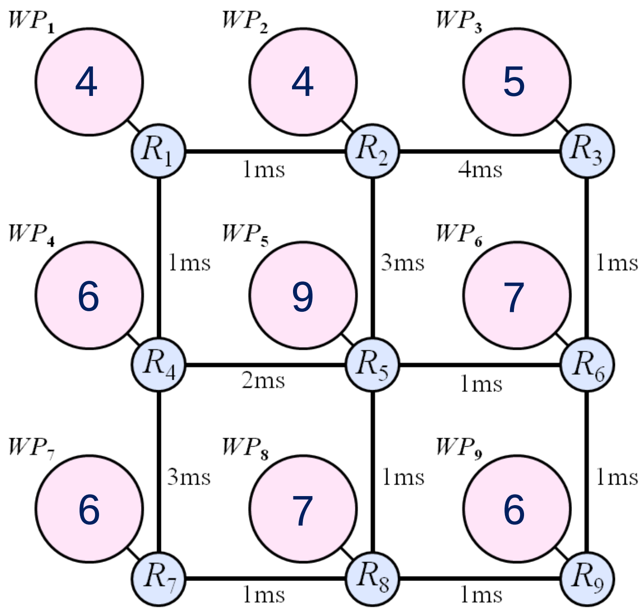
A_i: Bit rate of background traffic on L_i [bps]
T_i: Migration time on link L_i [s]



Destination Selection Algorithms

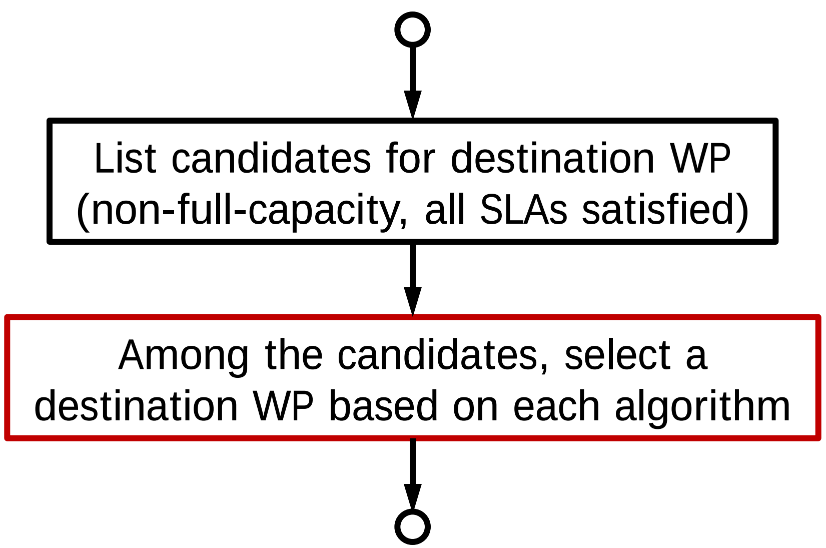
Algorithm	Destination WP	Expected effect
Minimum Impact Algorithm (MIA)	Migrate to WP with minimum impact	Decrease total NW impact
Maximum Remaining space Algorithm (MRA)	Migrate to WP with the maximum remaining space	Increase the migration success rate in the future
Maximum Covering Algorithm (MCA)	Migrate to WP with the maximum coverage	Decrease the number of migrations in the future

Coverage: The number of routers which the communication QoS is satisfying the SLA between router and WP



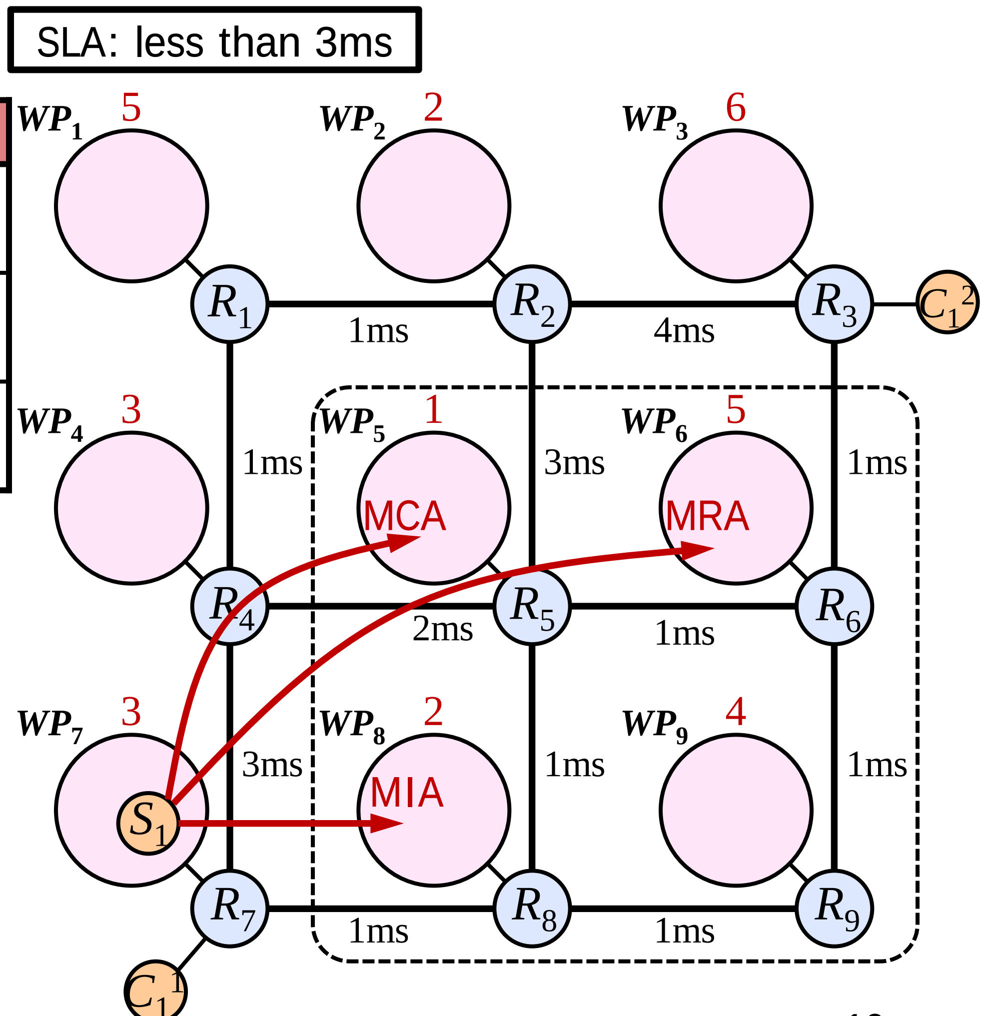
SLA: less than 3ms

Selection flow

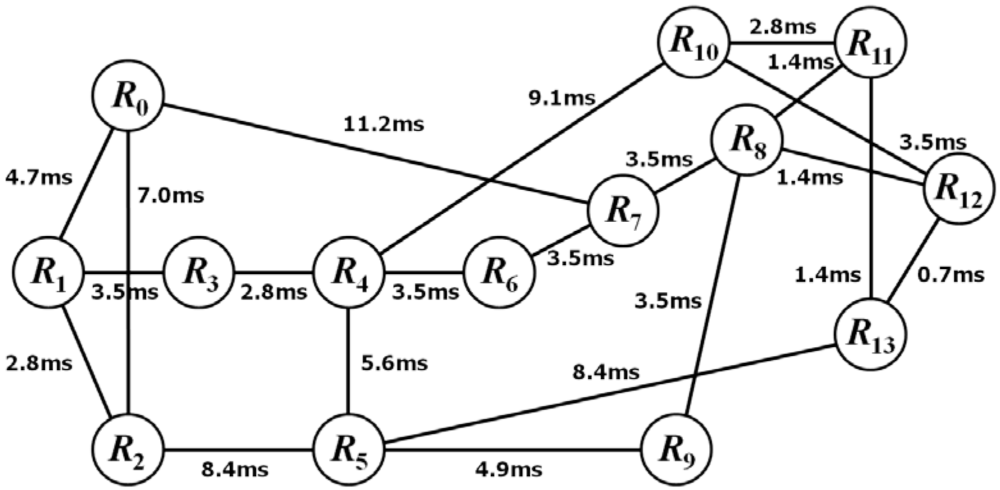


Example of Server Migration

Algorithm	Destination WP
Minimum Impact Algorithm (MIA)	Migrate to WP with minimum impact
Maximum Remaining space Algorithm (MRA)	Migrate to WP with the maximum remaining space
Maximum Covering Algorithm (MCA)	Migrate to WP with the maximum coverage



Simulation Models



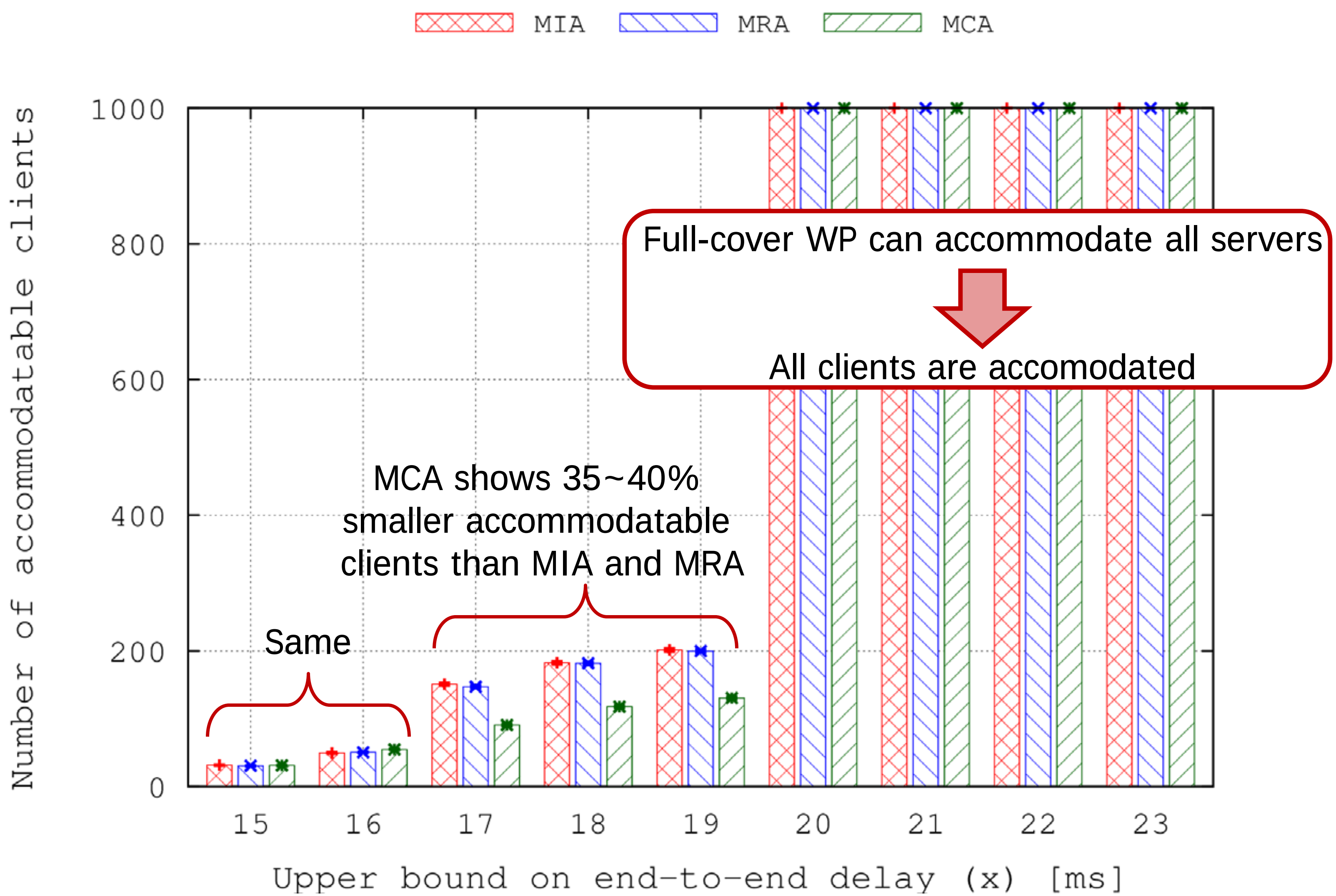
Parameter	Value
Number of routers	14
Number of WPs	14
Capacity of WP	4
Number of servers	7
Number of clients	1000
Number of servers in NW App 1	4
Number of servers in NW App 2	3
SLA	15~23ms
Link bandwidth	10Gbps
Background traffic	1Gbps
Server size	500Mbyte

Evaluation Index

1. number of *accommodatable clients
2. Total network impact divided by number of accommodatable clients

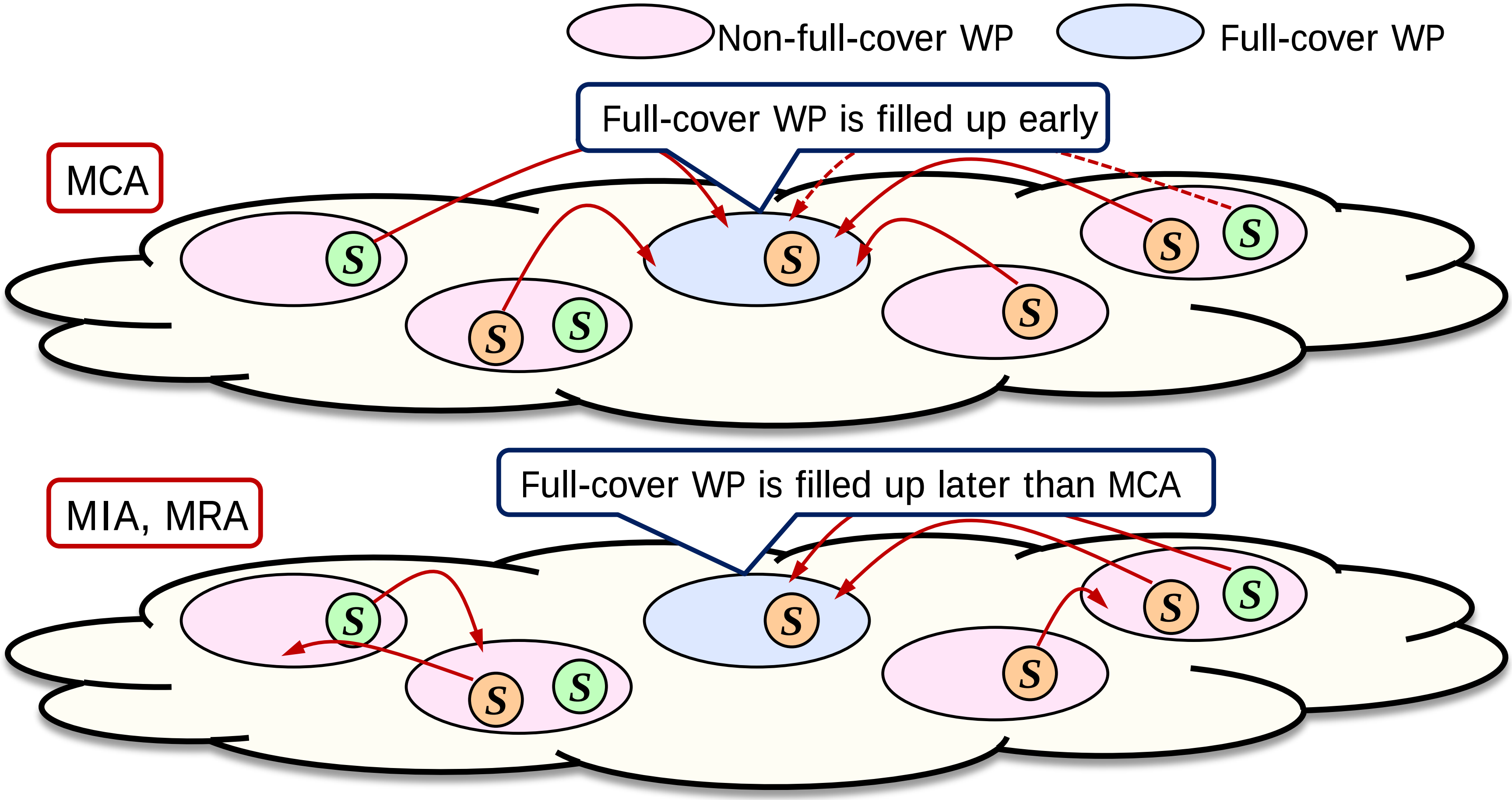
*Accommodatable clients: The number of clients whose SLA are satisfied

Evaluation Result (Accommodatable Clients)

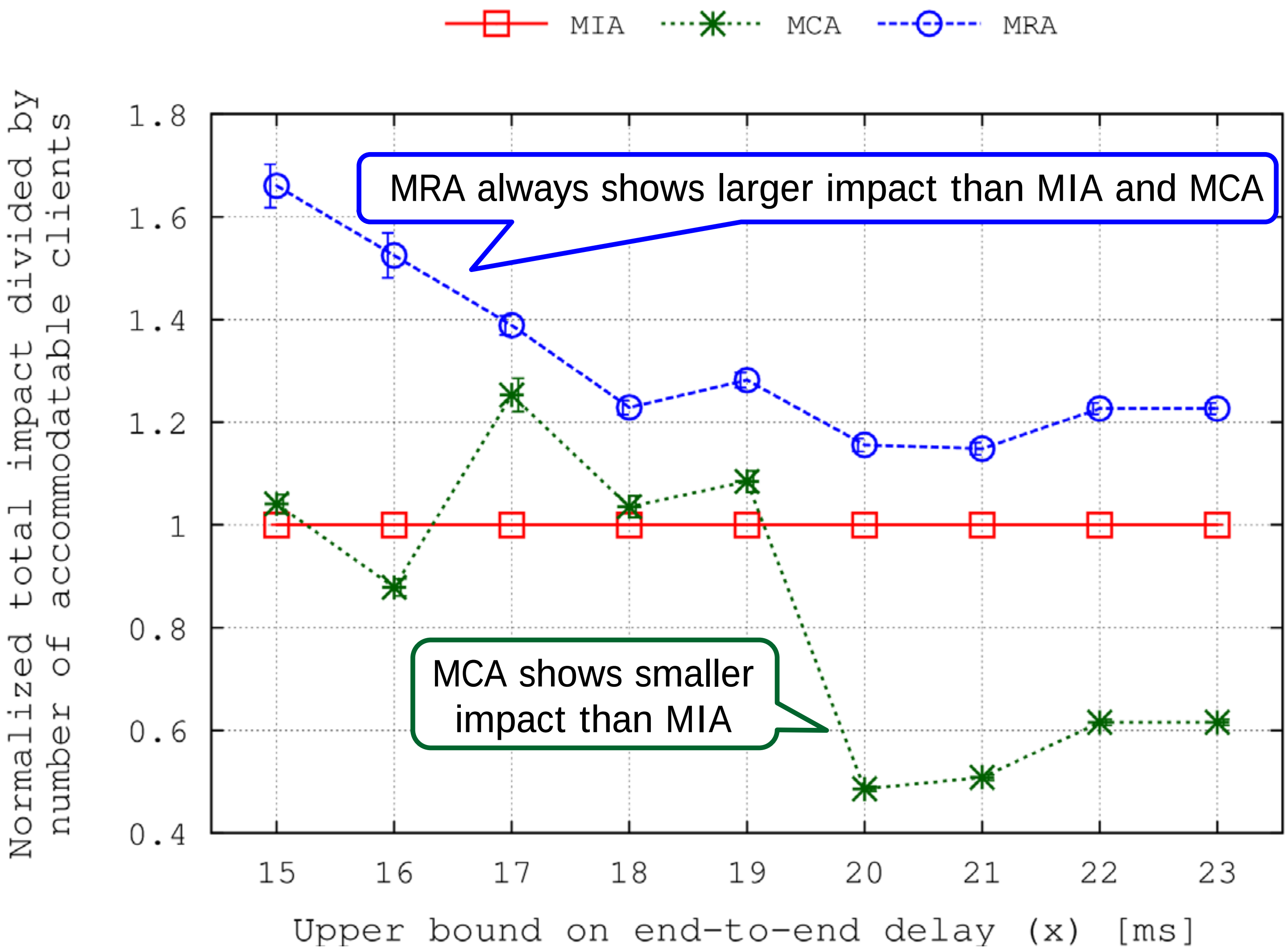


Evaluation Result (Explanation)

- Why MCA shows smaller accommodatable clients than MIA and MRA



Evaluation Result (Network Impact)



Conclusions and Future Works

- Conclusions
  - Propose and evaluate destination selection algorithms
  - When the capacity of the full-cover WP is smaller than the total number of servers
    - MIA shows the best performance
  - When the capacity of the full-cover WP is equal to or larger than the total number of servers
    - All the algorithm accommodates almost the same number of clients
    - MCA shows the best performance as to network impact
- Future Works
  - Realization of push-out function
  - Performance evaluate in terms of downtime of NW-Apps
  - Design of an algorithm for server replication