# Load Balancing

Martin Haidn & Nikolaus Schrack

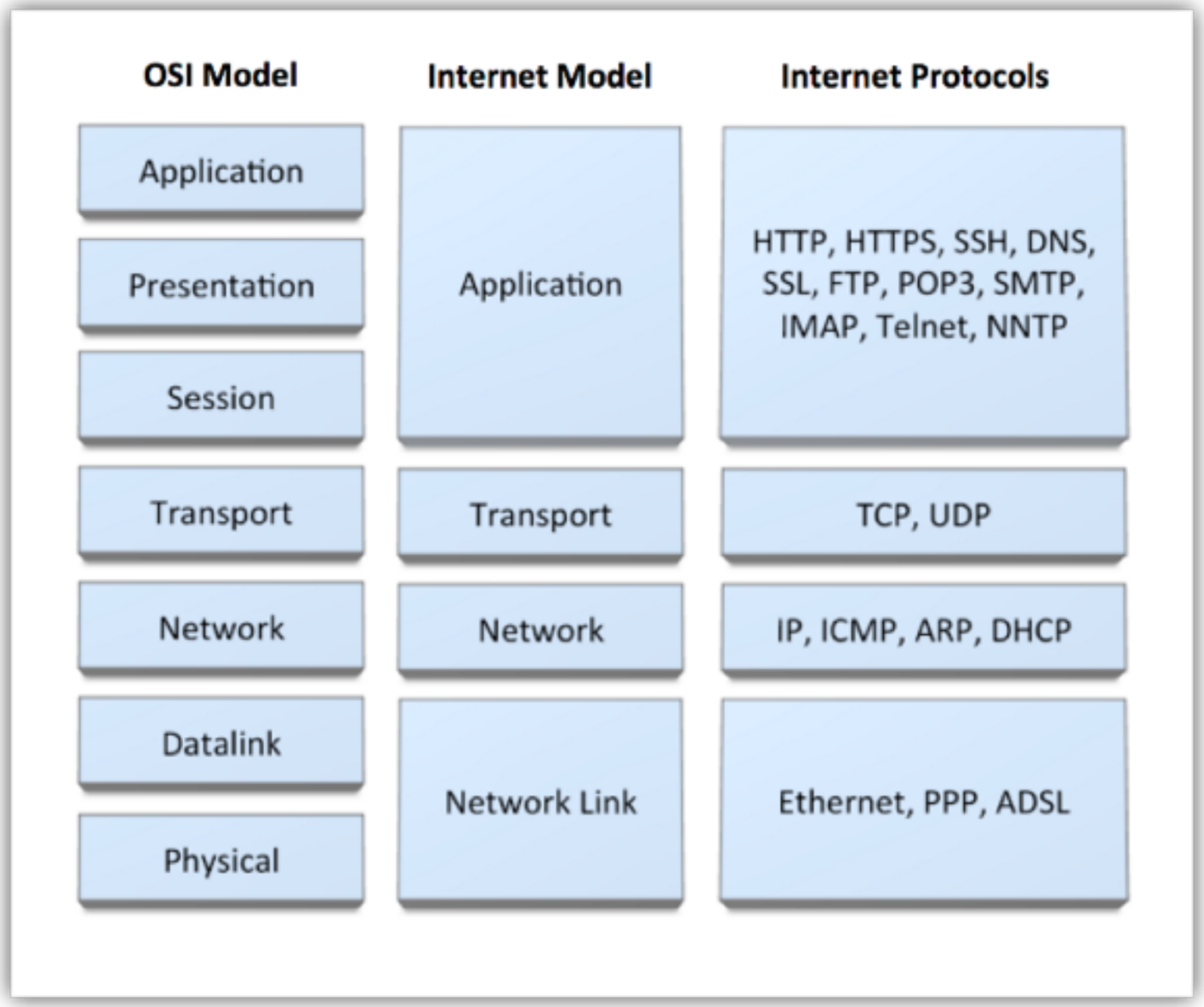
# Need and Goals

- Scalability
- Availability
- Performance
- Manageability
- Security
- Costs

# Products

- Software Load Balancing
- Appliances
- Switches

# OSI-Model




http://vichargrave.com/

# Fundamentals

- Low Layer (1-4)
- IP-/ MAC- Addresses
- Making a decision
- Distribute load based on required performance

# Layer-7 LB

- Content based decisions
- Application Delivery Controllers
- Parsing application data
- LB Policy
- Heavy processing time

# Advanced Concepts

- Session Persistance
  - TCP-SYN Packet
  - Application Request
- URL Switching
  - Separating content
  - Cookie Switching
- SSL Termination

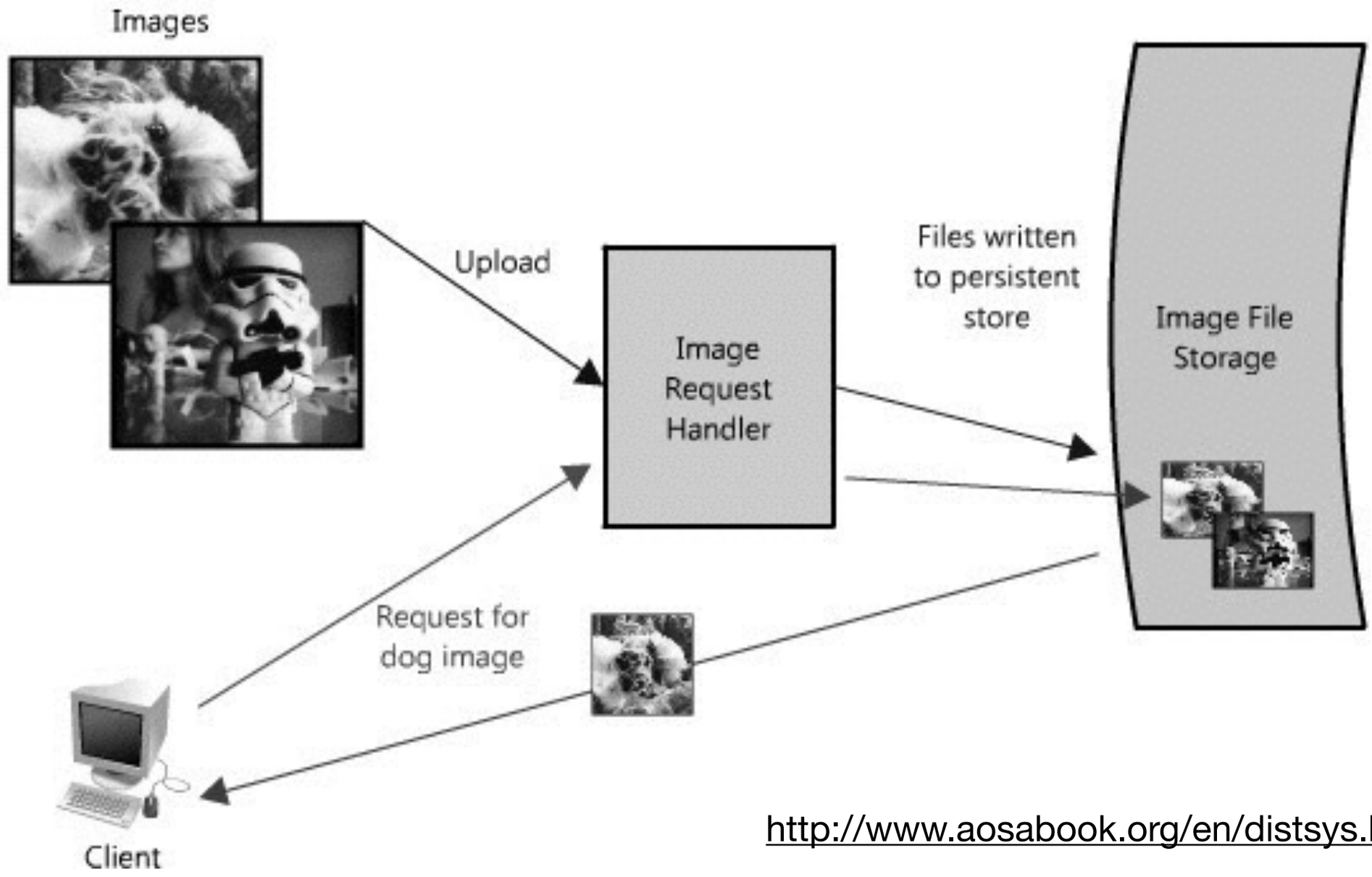
# Web-App Design

- Same goals as load balancing
- Content
- Service
- Redundancy
- Partitions

# Content

- Distribute request handler
- Shorten querying time
- Raise scalability

# Content

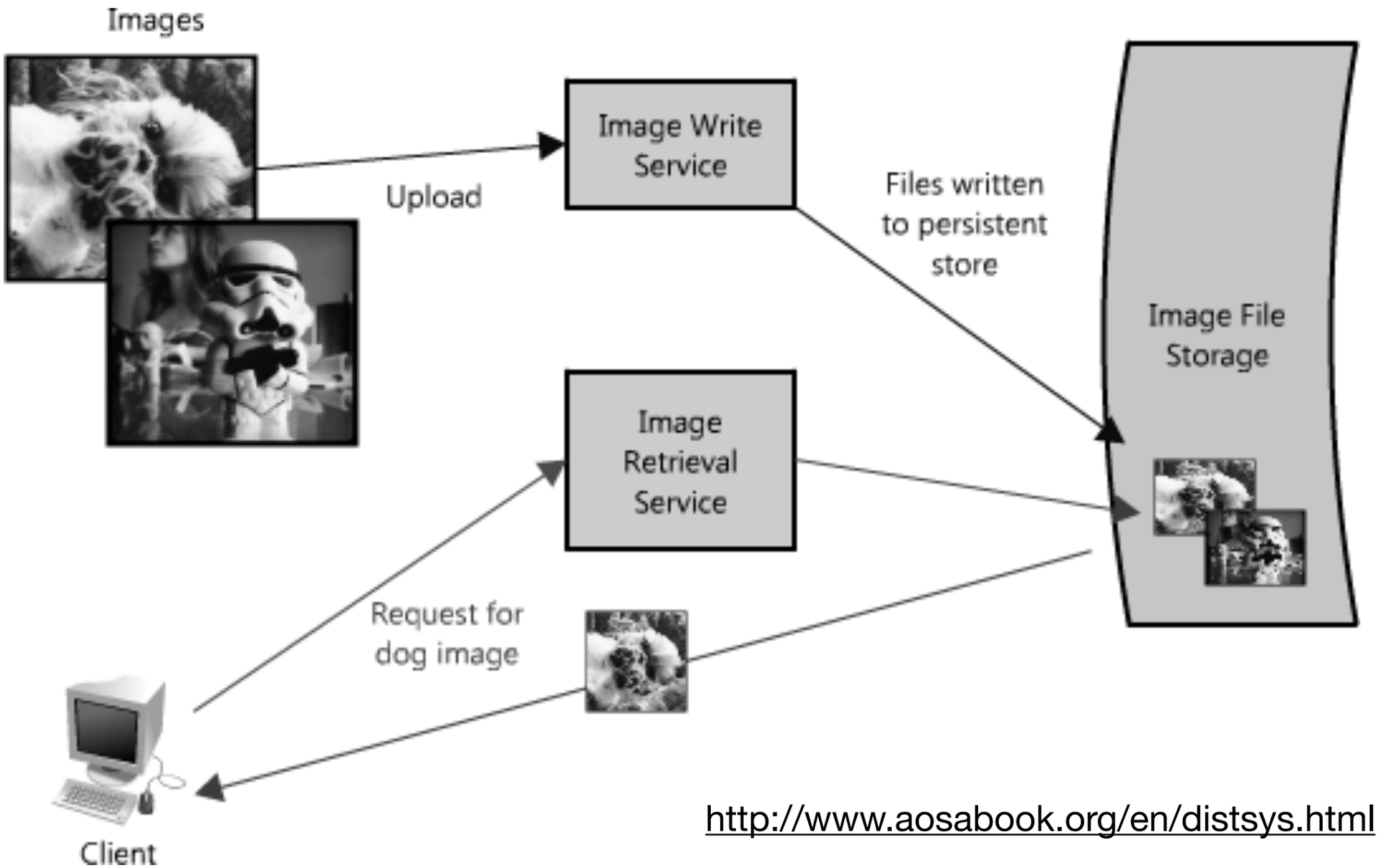



http://www.aosabook.org/en/distsys.html

# Service

- See every component as own service
- Define clear interfaces
- Service Oriented Architectures (SOA)
- Public- Facing API’s

# Service




http://www.aosabook.org/en/distsys.html

# Scheduling Algorithm

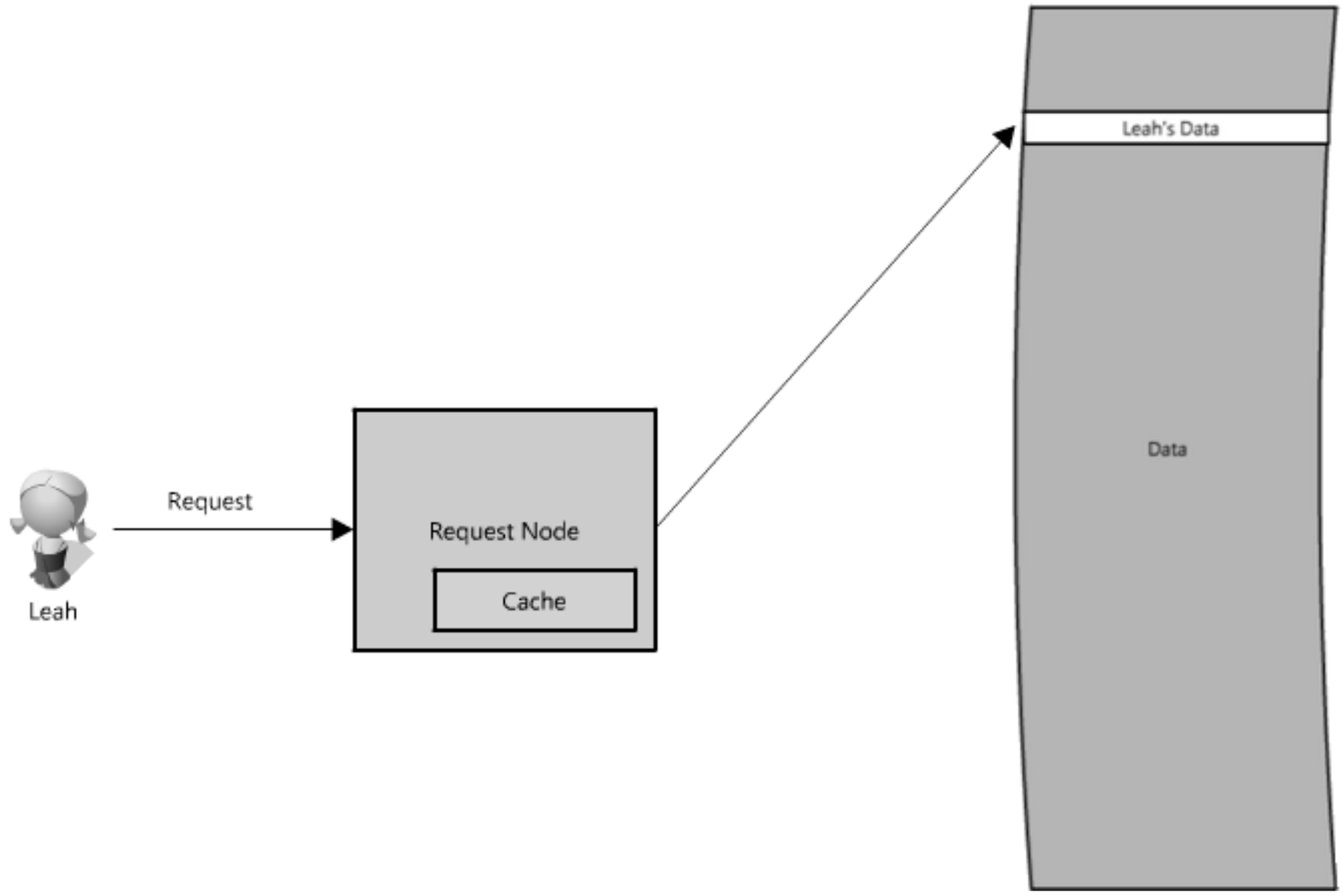
- Round Robin
- Weighed Round Robin
- Least Connection
- Least Connected Slow Start Time
- Weighed Least Connected
- Agent Based Adaptive

# Scheduling Algorithm

Agent Based Adaptive

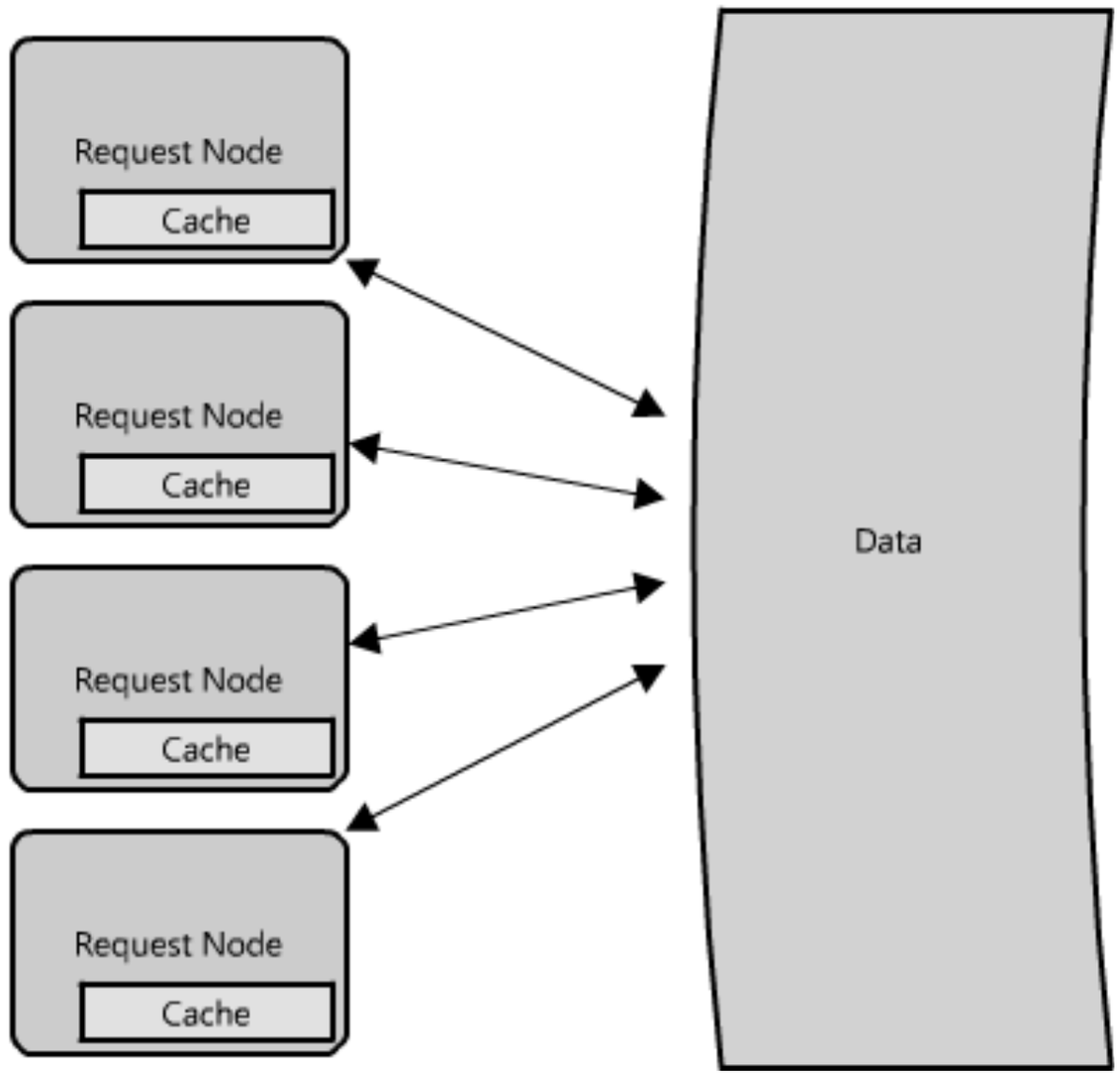
- adaptive logic
- periodically file
- numerical value
- low traffic

# Caches



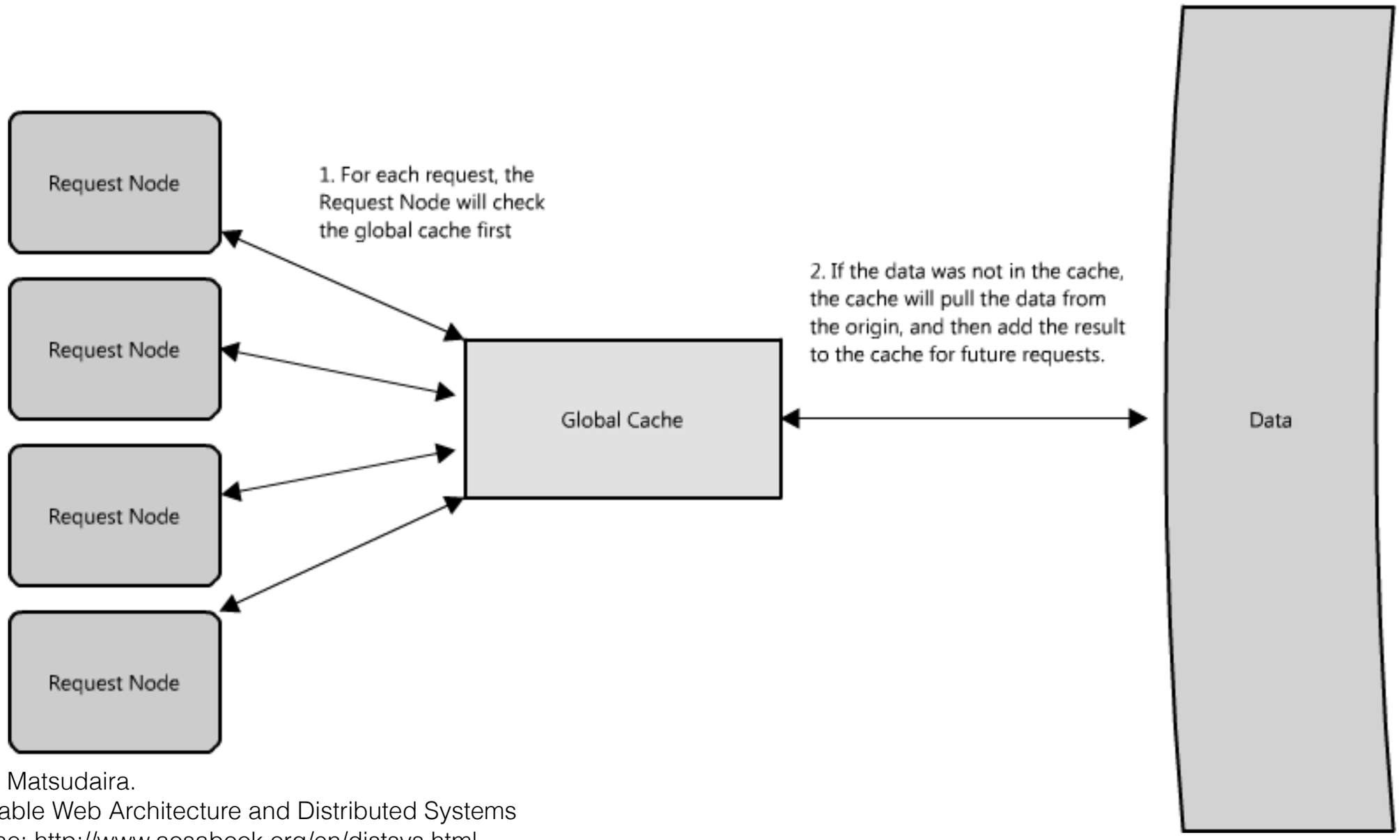

Kate Matsudaira.
Scalable Web Architecture and Distributed Systems
Online: http://www.aosabook.org/en/distsys.html
zuletzt aufgerufen: 04.12.2014"

# Caches



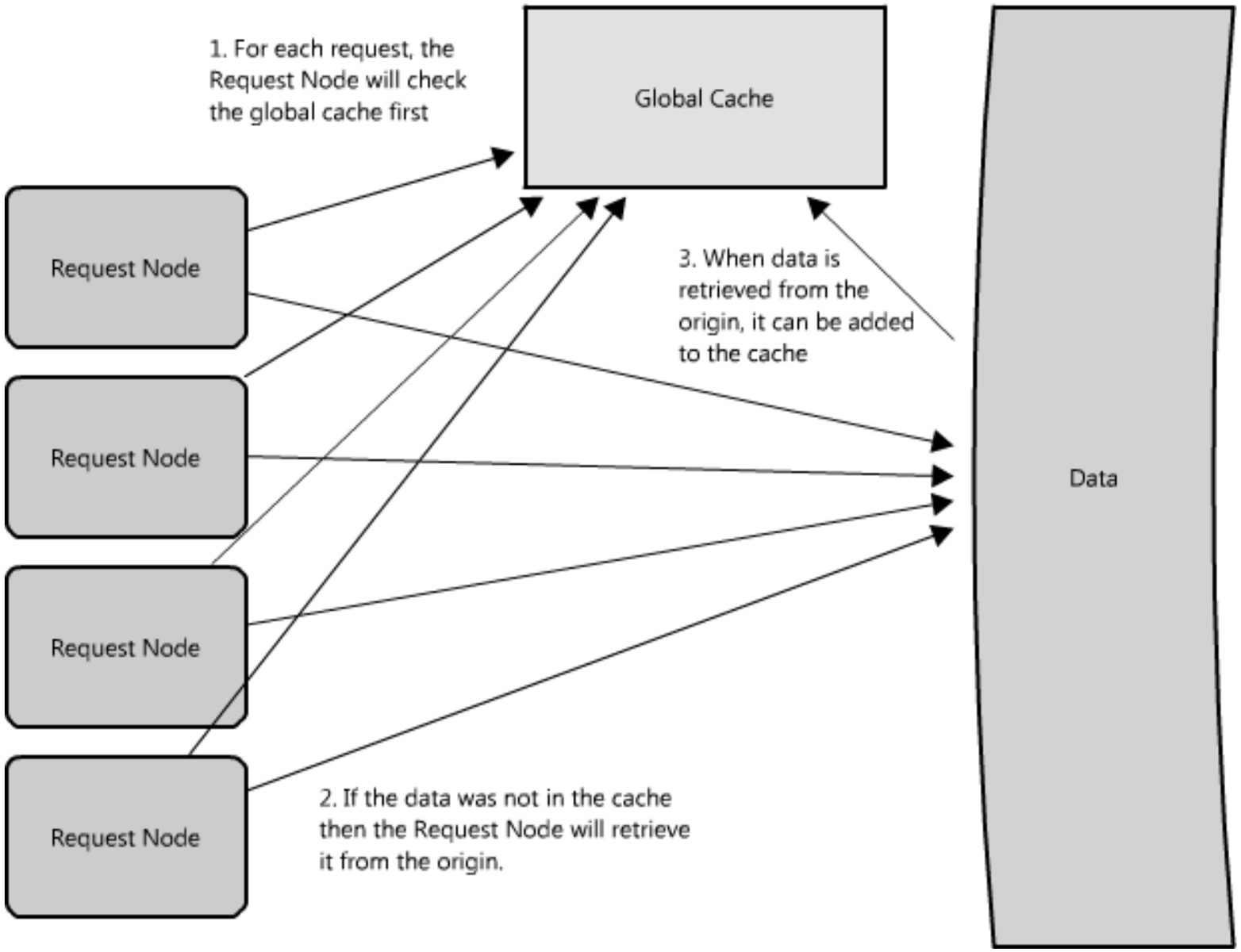

Kate Matsudaira.
Scalable Web Architecture and Distributed Systems
Online: http://www.aosabook.org/en/distsys.html
zuletzt aufgerufen: 04.12.2014"

# Global Caches



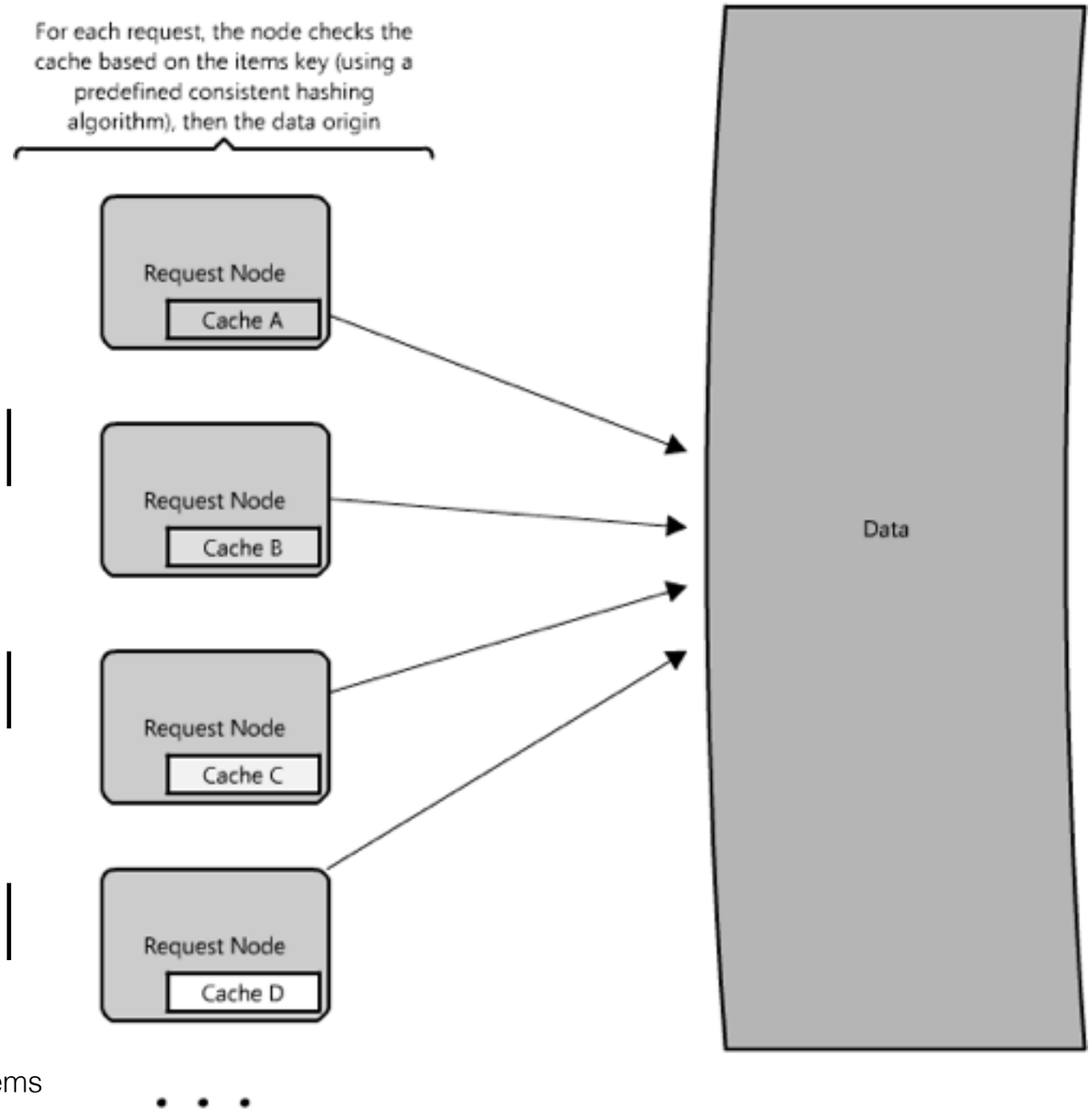

Kate Matsudaira.
Scalable Web Architecture and Distributed Systems
Online: http://www.aosabook.org/en/distsys.html
zuletzt aufgerufen: 04.12.2014"

# Global Cache



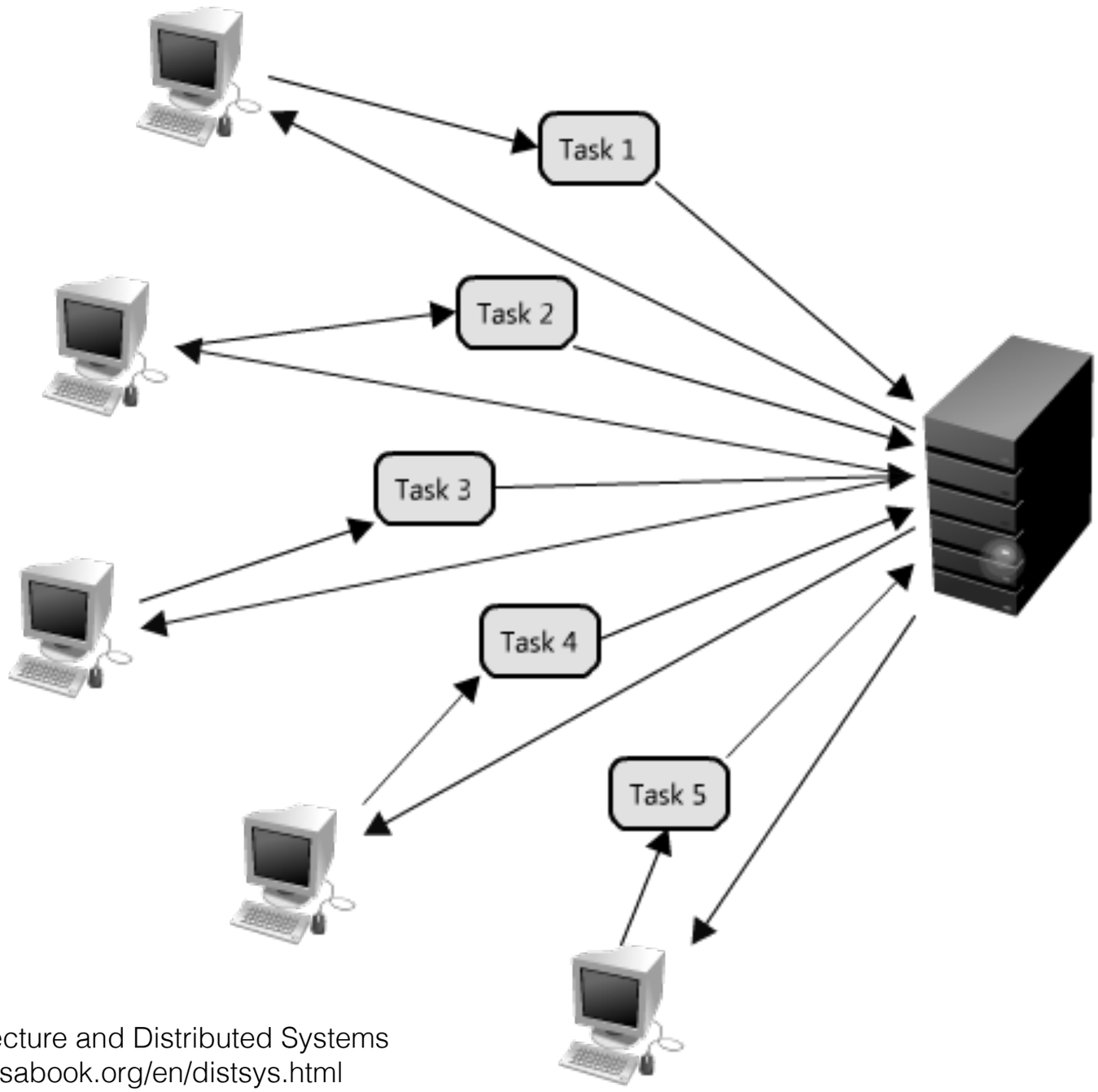

Kate Matsudaira.
Scalable Web Architecture and Distributed Systems
Online: http://www.aosabook.org/en/distsys.html
zuletzt aufgerufen: 04.12.2014"

# Distributed Cache



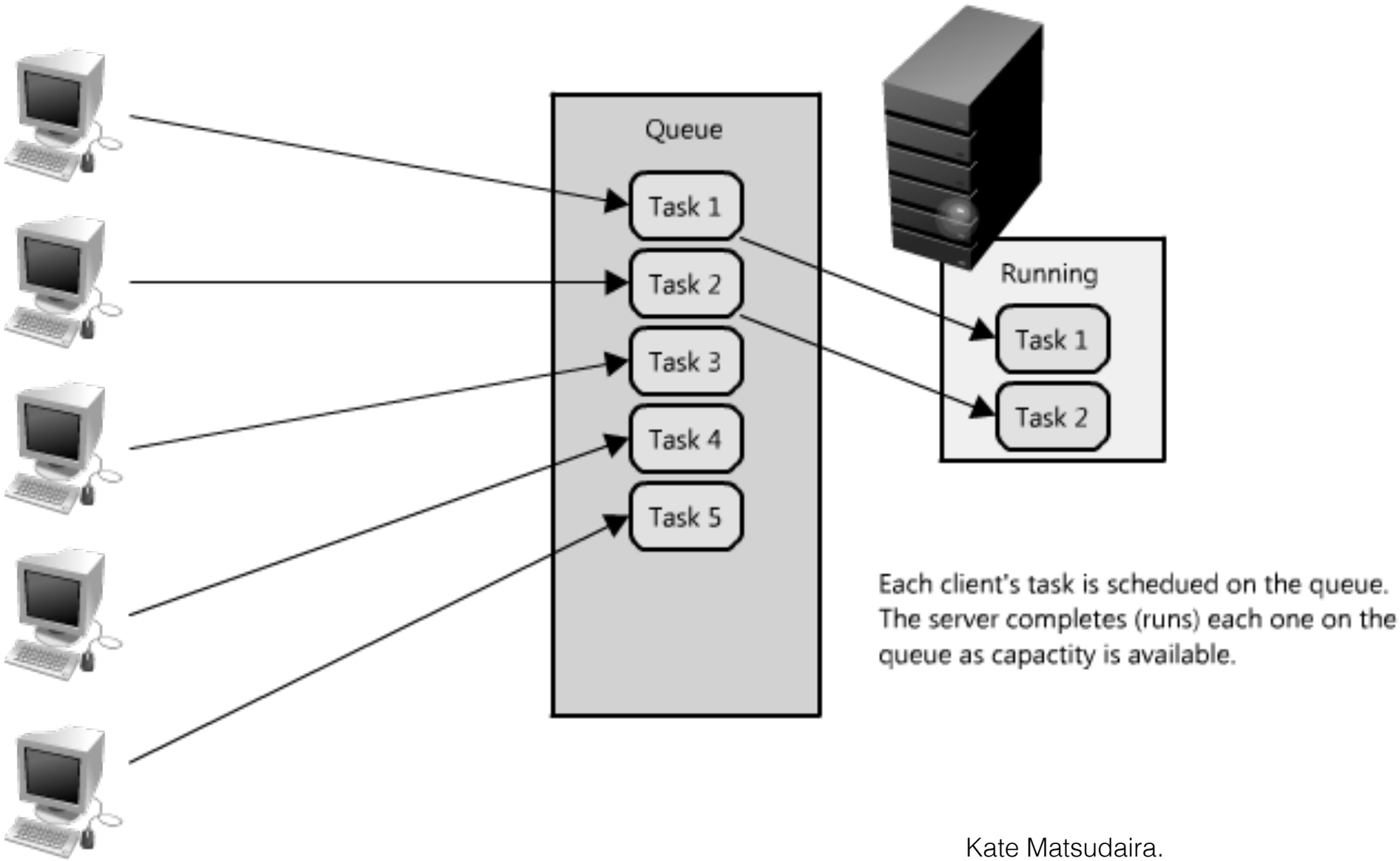

Kate Matsudaira.
Scalable Web Architecture and Distributed Systems
Online: http://www.aosabook.org/en/distsys.html
zuletzt aufgerufen: 04.12.2014"

# Queue



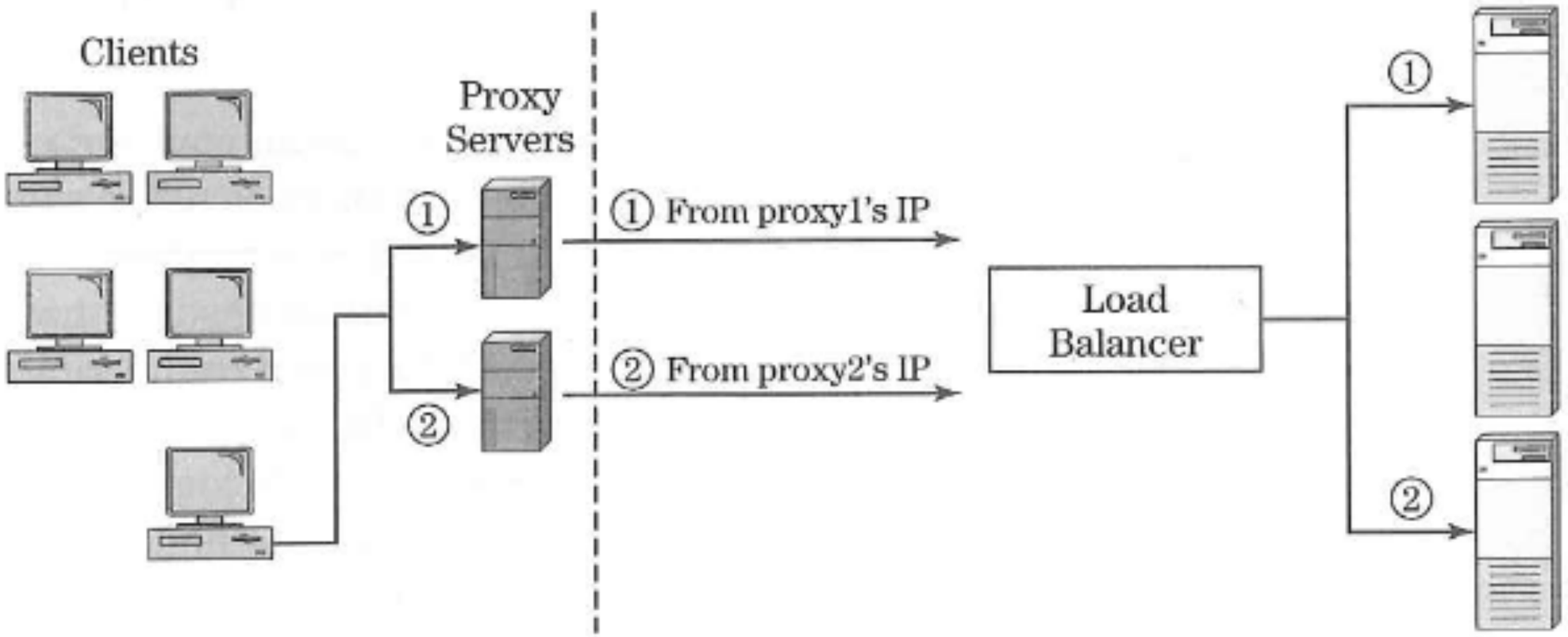

Kate Matsudaira.
Scalable Web Architecture and Distributed Systems
Online: http://www.aosabook.org/en/distsys.html
zuletzt aufgerufen: 04.12.2014"

# Queue




Kate Matsudaira.
Scalable Web Architecture and Distributed Systems
Online: http://www.aosabook.org/en/distsys.html
zuletzt aufgerufen: 04.12.2014"

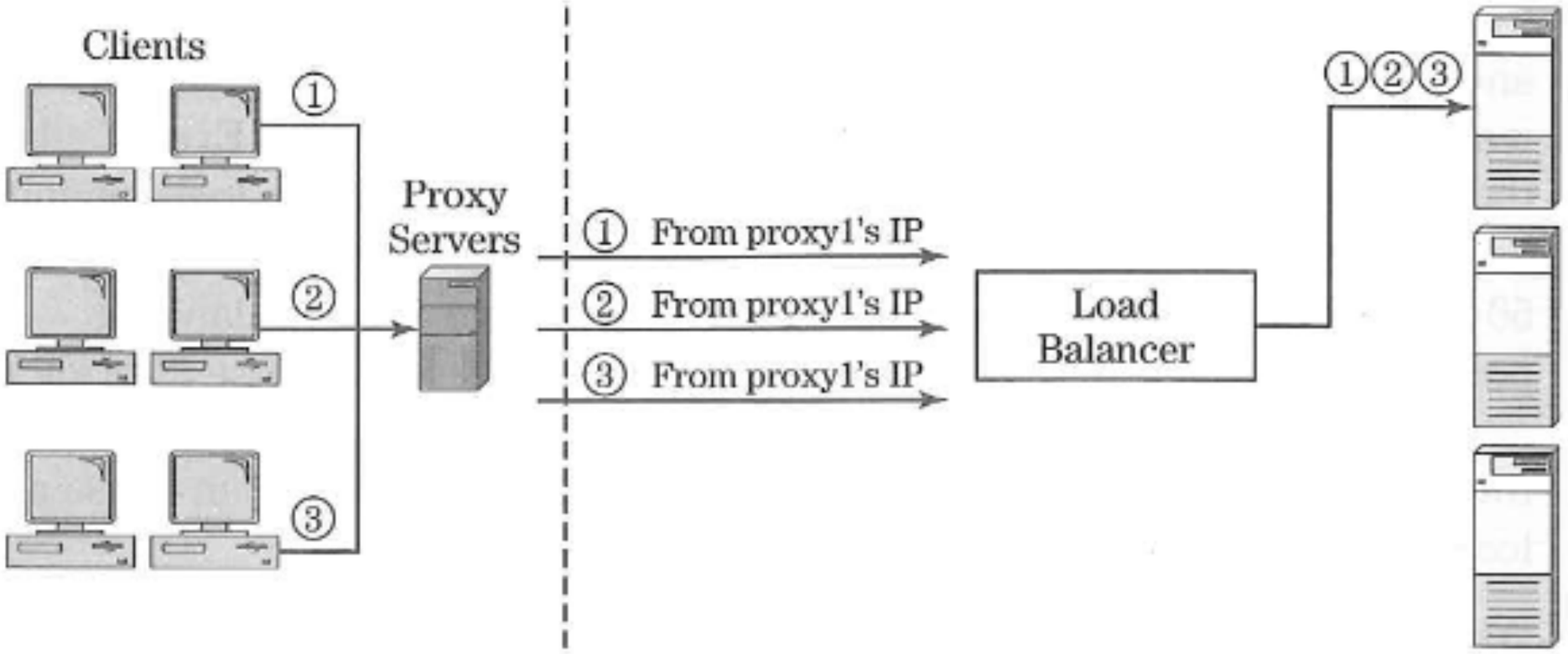
# Mega Proxy Problem




Internal network for an enterprise or ISP with multiple proxy servers

The second connection comes from proxy2's IP address, causing the load balancer to assign the connection to a server based on load rather than on session persistence.

# Mega Proxy Problem




Internal network for an enterprise or ISP with a proxy server serving thousands of users

Load balancer sends all connections to the same server to assure session persistence, because all users through proxy server have proxy's IP address as the source IP. This causes uneven load balancing.

# Sources

Chandra Kopparapu, Load Balancing Servers,

http://networksandservers.blogspot.co.at/2011/03/balancing-iii.html

http://www.loadbalancer.org/load_balancing_methods.php

http://www.liquidweb.com/kb/understanding-load-balancing/

http://www.loadbalancerblog.com

http://www.aosabook.org/en/distsys.html

http://vichargrave.com/network-programming-design-patterns-in-c/

# Thanks For Your Attention