

Cost- and Quality-of-Service-Aware Network-Service Deployment

Robert Haas*, Patrick Droz*, Burkhard Stiller**,

*IBM Zurich Research Lab,

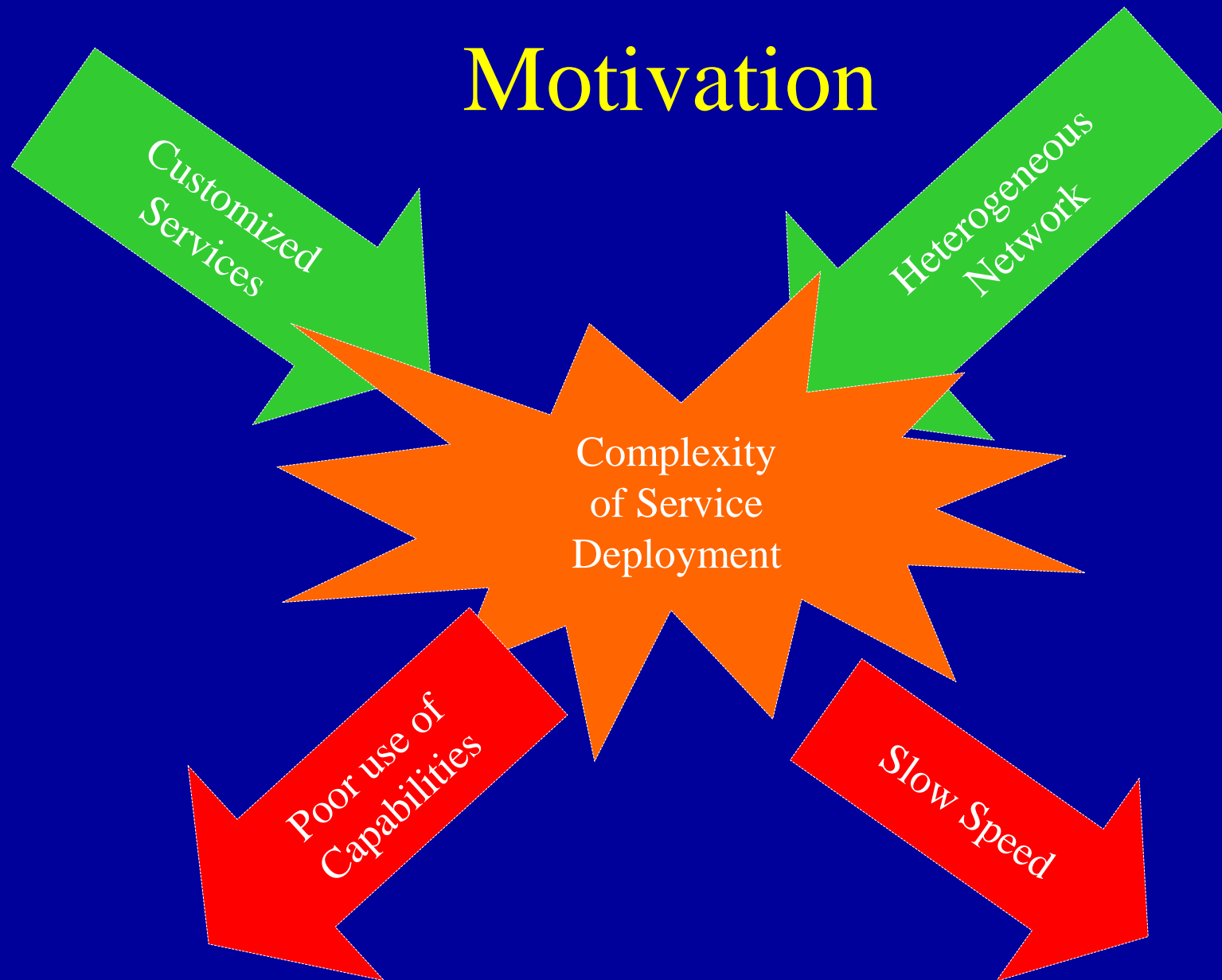
**Computer Engineering and Networks Lab, ETHZ.

ICQT'01, Vienna, September 2001

Content

- Motivation
 - General problem statement
 - Related work
- Procedure overview
 - Example of service deployment
- Extension for cost-aware deployment
 - Examples of node-only & path-and-node –based
- Conclusion

Motivation



Overview

- ➔ Provide automated service deployment
 - make "best use" of capabilities of the infrastructure
 - price-wise and performance-wise.
 - automatically organize installation and configuration
- Hierarchical deployment:
 - aggregated views of network basic information
 - necessary for large networks
- Deployment is service-generic:

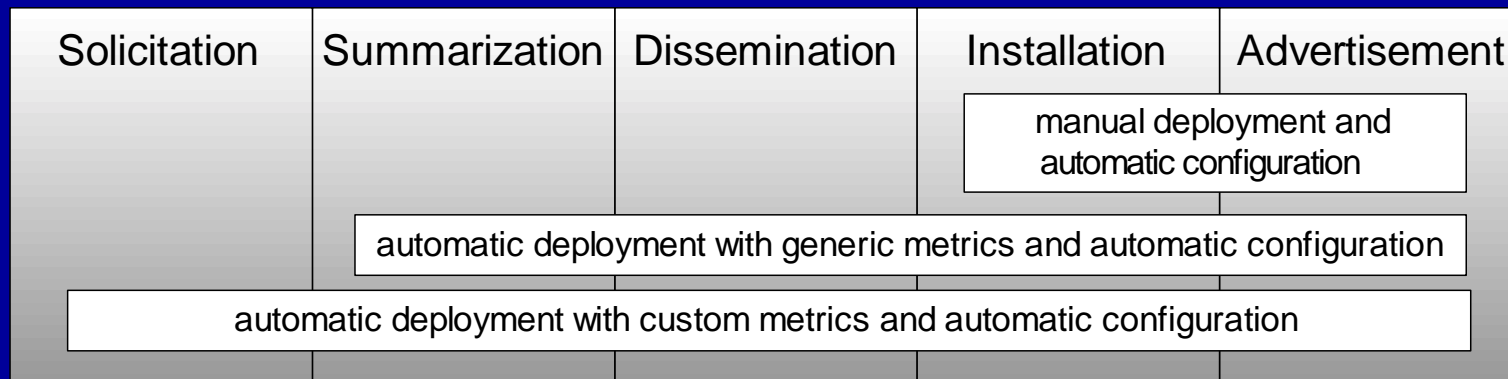
Basic Information Types
Topology
Capabilities
Performance
Cost

Service Deployment Policy
Path-based
Node-based
Path- and Node-based

Related Work

- in active networks
 - service deployed along the path taken by the capsule
 - service destined for that session
- in mobile agents
 - navigation model
- in programmable networks
 - service loaded on-demand, manual coordination
- in IN (Intelligent Network)
 - service deployed at a central point (SCP)
- in sensor networks

5-steps Procedure



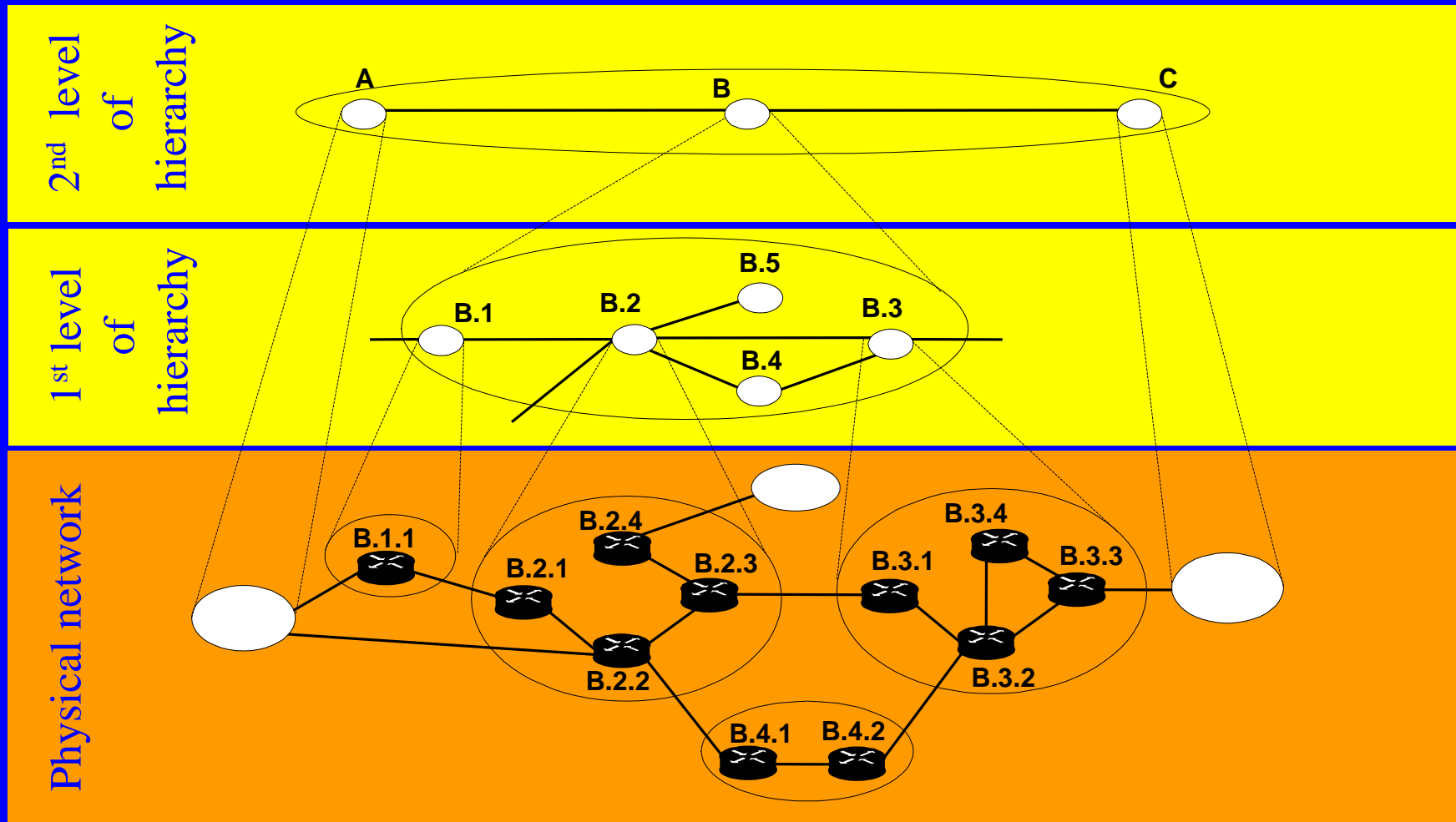
Examples of Services

	Deployment	Addressing	Signaling
Hierarchical IP Routing	Node-based	Explicit	Out-of-band
Transparent Multi-layer Proxy-Caching	Node-based	Implicit (for clients) and Explicit	In-band
VPN	Path- and Node-based	Implicit (for RSVP) and Explicit (for endpoints)	Out-of-band
Diff-Serv++	Path-based	Implicit	In-band

Solicitation ←

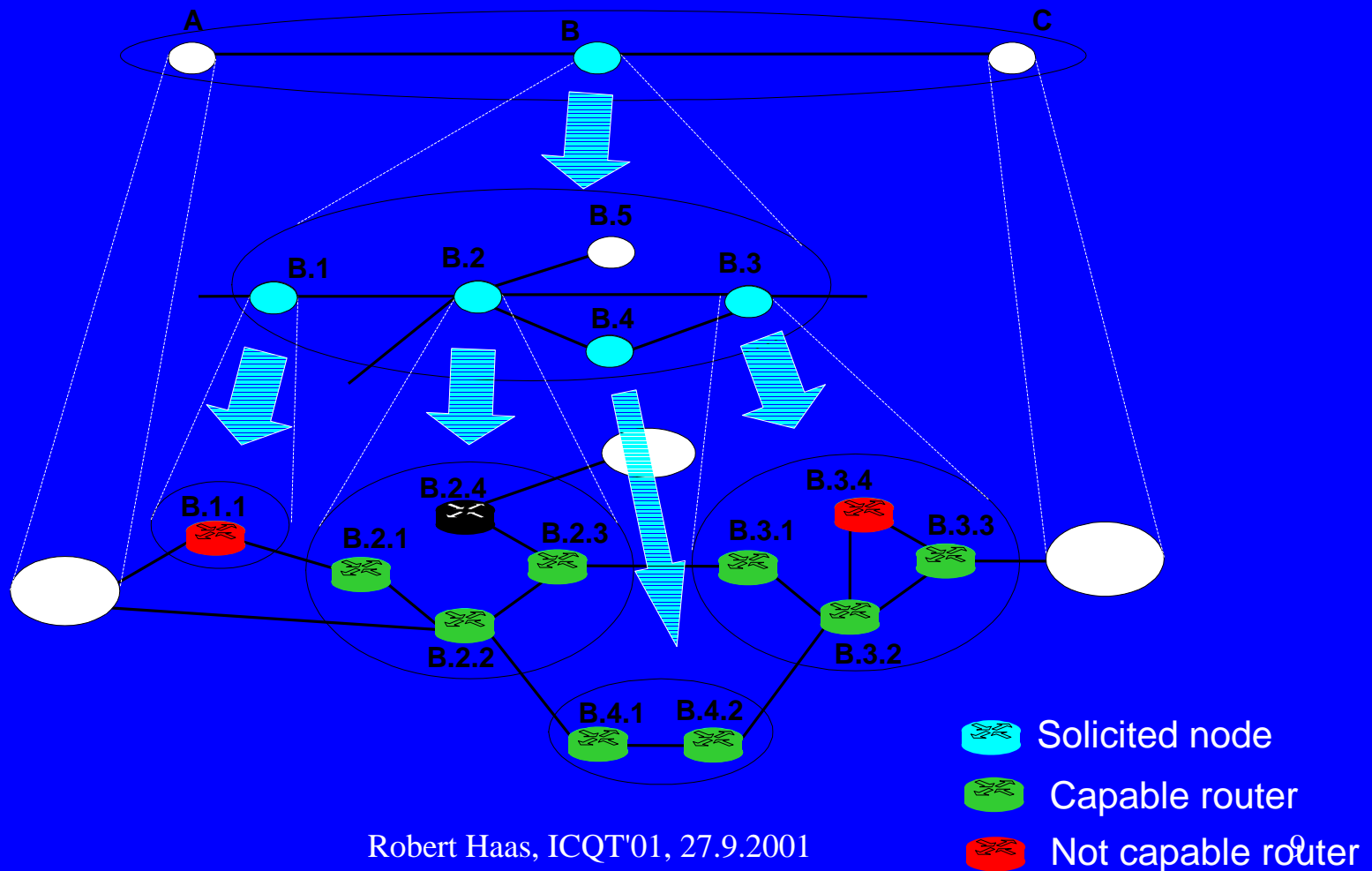
Summarization
Dissemination
Installation
Advertisement

Diff-Serv++



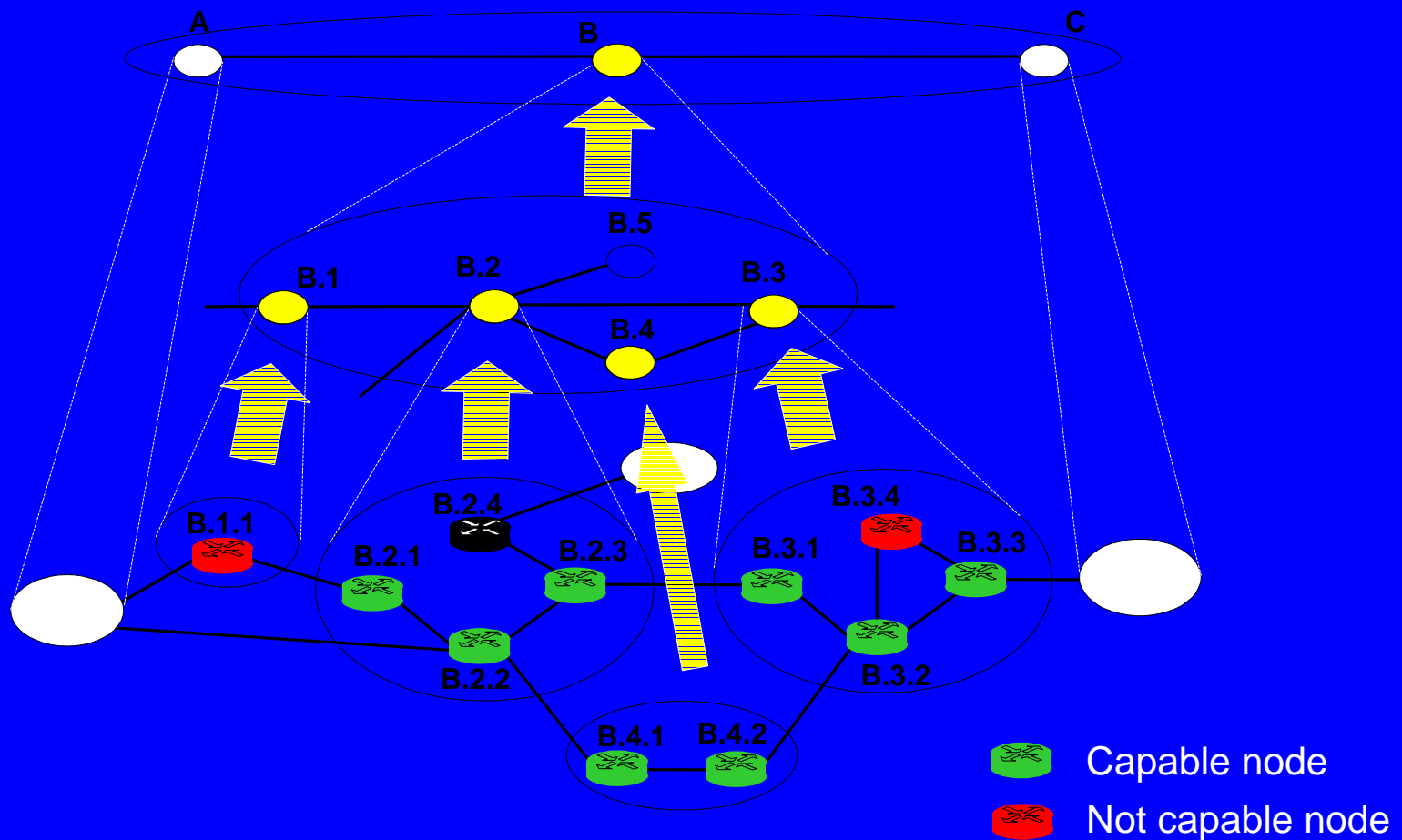
Solicitation ←
 Summarization
 Dissemination
 Installation
 Advertisement

Solicitation



Solicitation
Summarization ←
Dissemination
Installation
Advertisement

Summarization



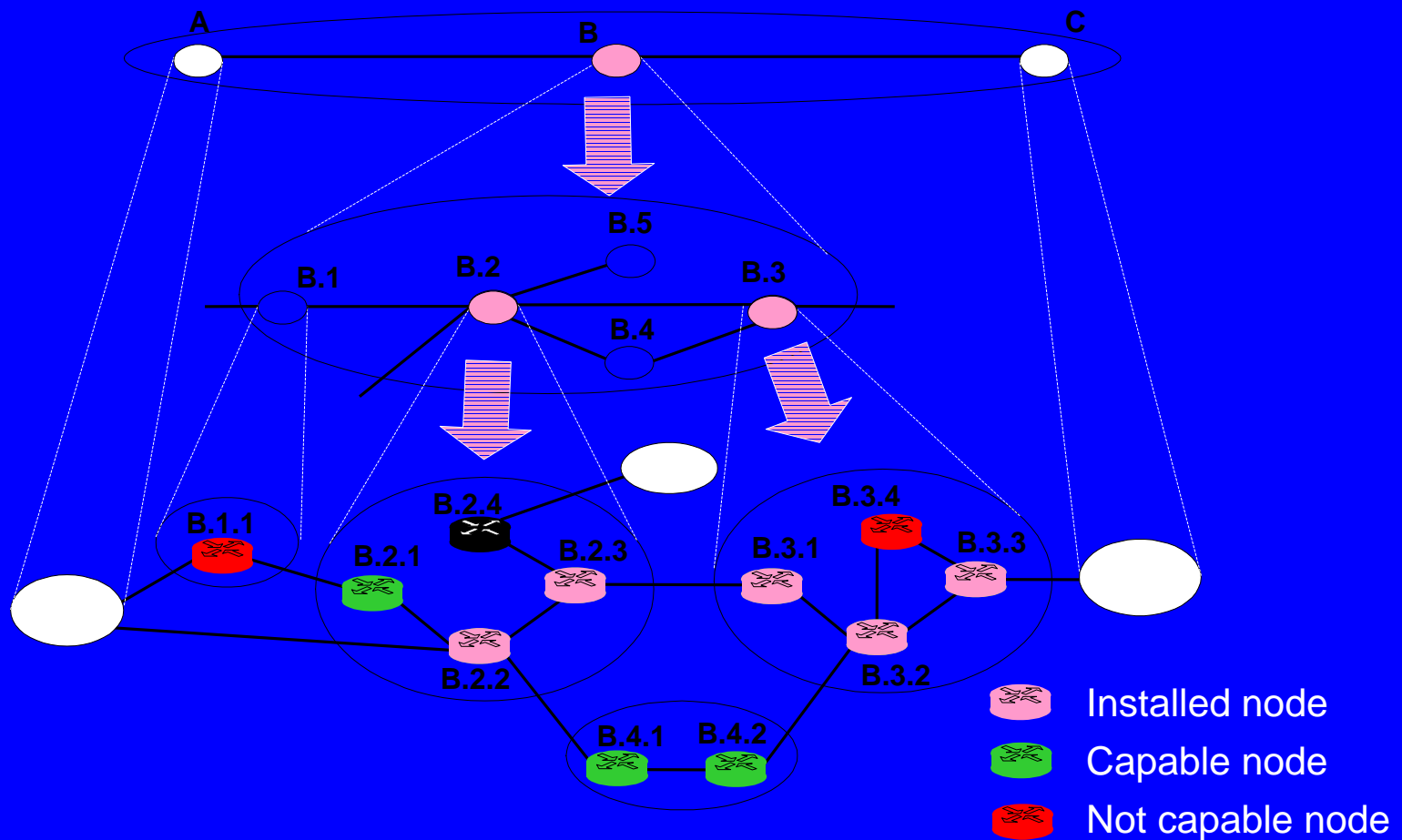
Summarized Information

- Transition matrices contain:
 - Topology
 - Node capabilities
 - Cost (of links and node capabilities)
- Used during the next step
 - Select shortest path

$$T_{B_2} = T_{B_3} = \begin{pmatrix} 1 & \dots & \\ 2 & 1 & \dots \\ 3 & 2 & 1 \end{pmatrix}; T_B = \begin{pmatrix} 0 & \dots & \\ 0 & 1 & \dots \\ 0 & 5 & 1 \end{pmatrix}$$

Solicitation
Summarization
Dissemination ←
Installation
Advertisement

Dissemination

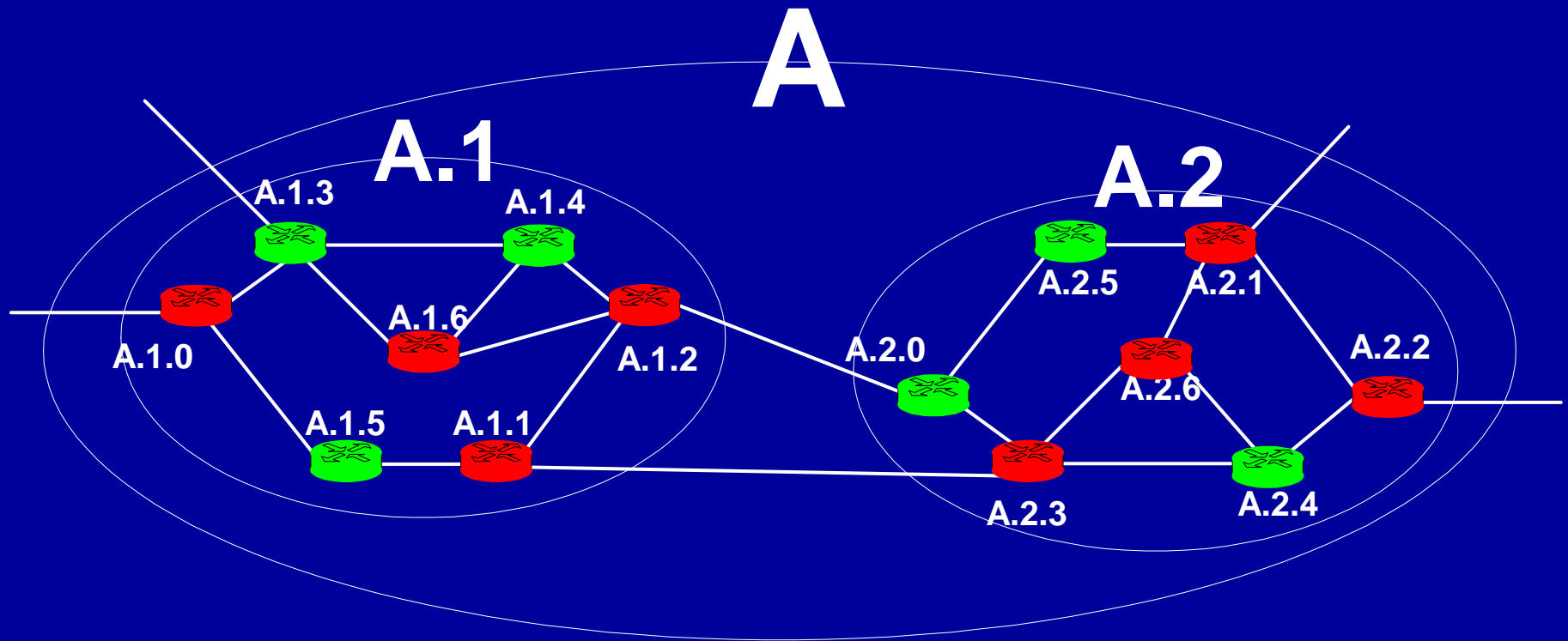


Cost-aware Deployment

- Meanings of “cost”:
 - intrinsic cost for ISP itself
 - published cost to customers
- Treated as an additional dimension in the service deployment framework
- But ... how to delay the decision of **what is the acceptable cost so to weight it against the performance gain ?**
- Illustrations:
 - Node-only deployment
 - Node and path-based deployment

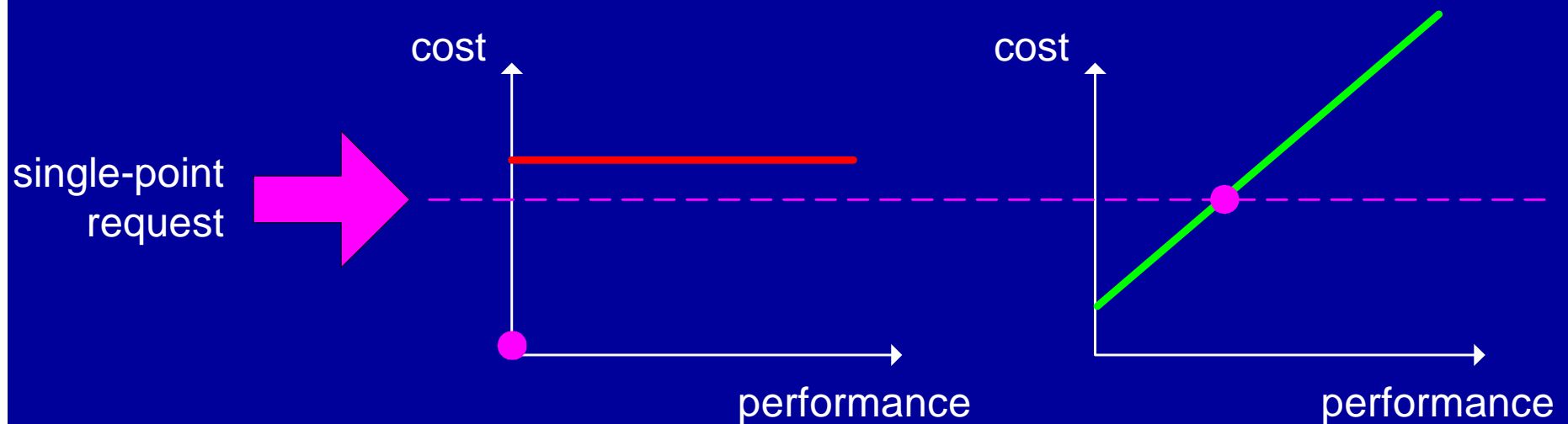
Node-only Deployment

- Web-cache: which host can provide the best performance given a certain cost in the group composed of A.1 and A.2 ?



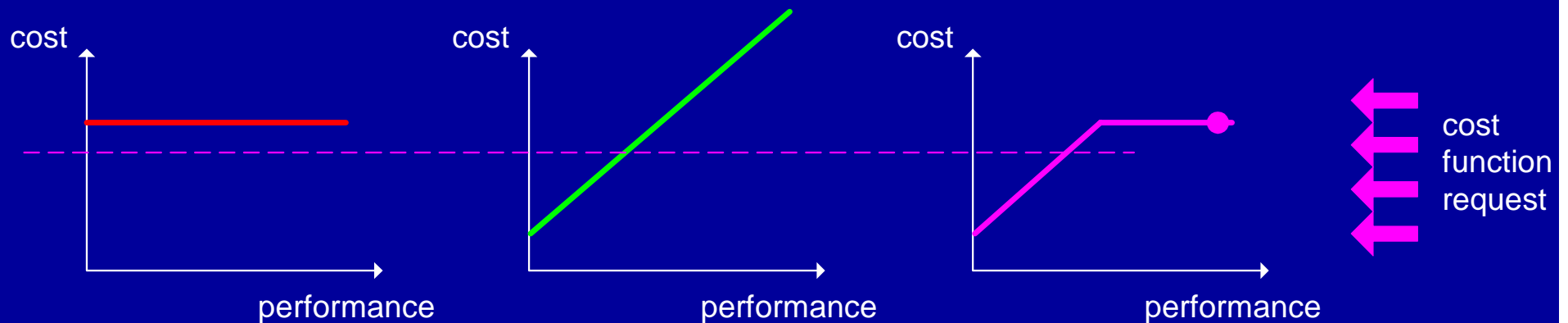
Node-only Deployment

- single-point:
 - Set the maximum cost
 - Recursively choose the host (or logical node) with the maximum performance for that cost



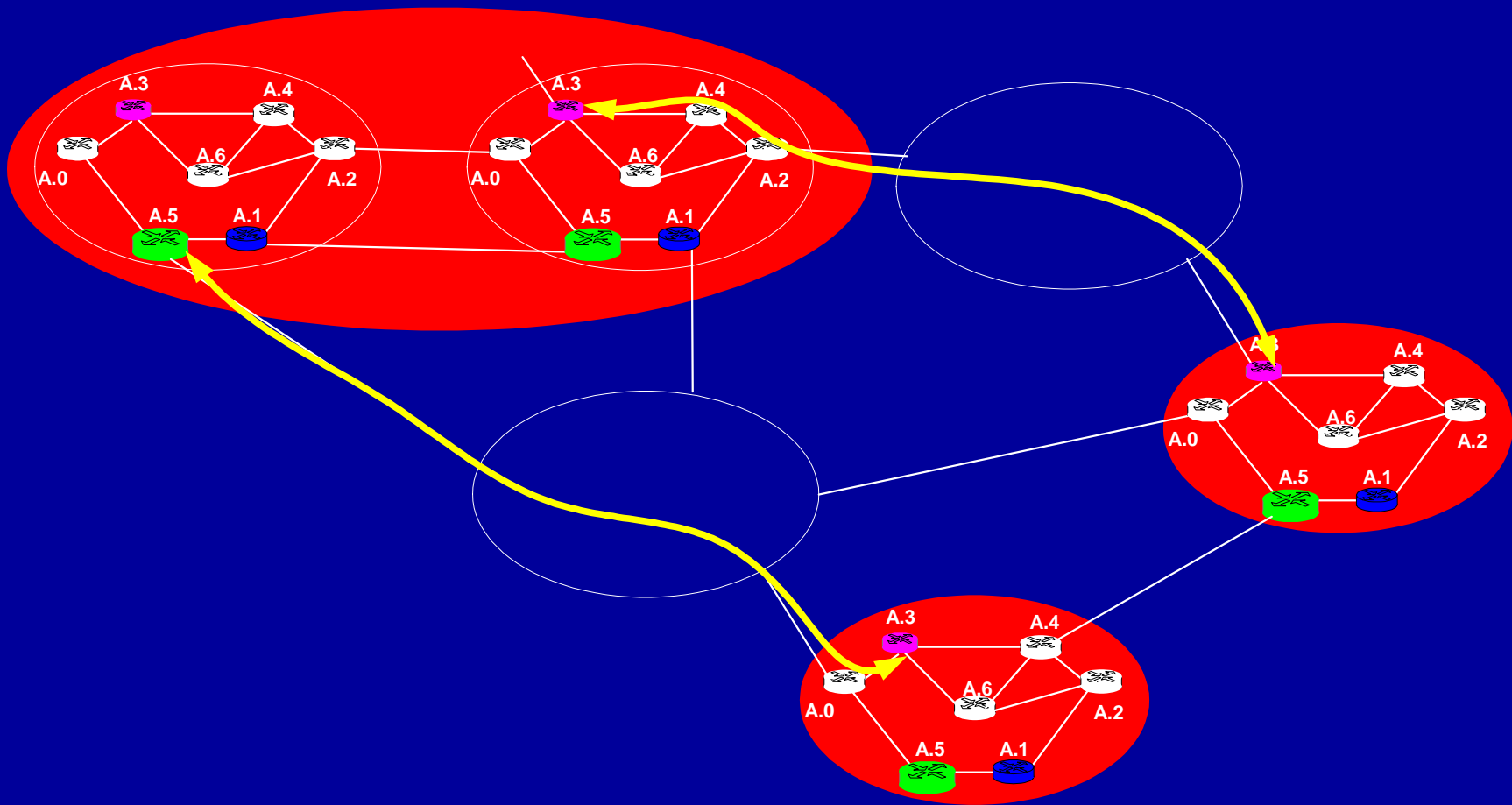
Node-only Deployment

- Cost functions
 - Don't set any maximum cost
 - Recursively summarize cost functions

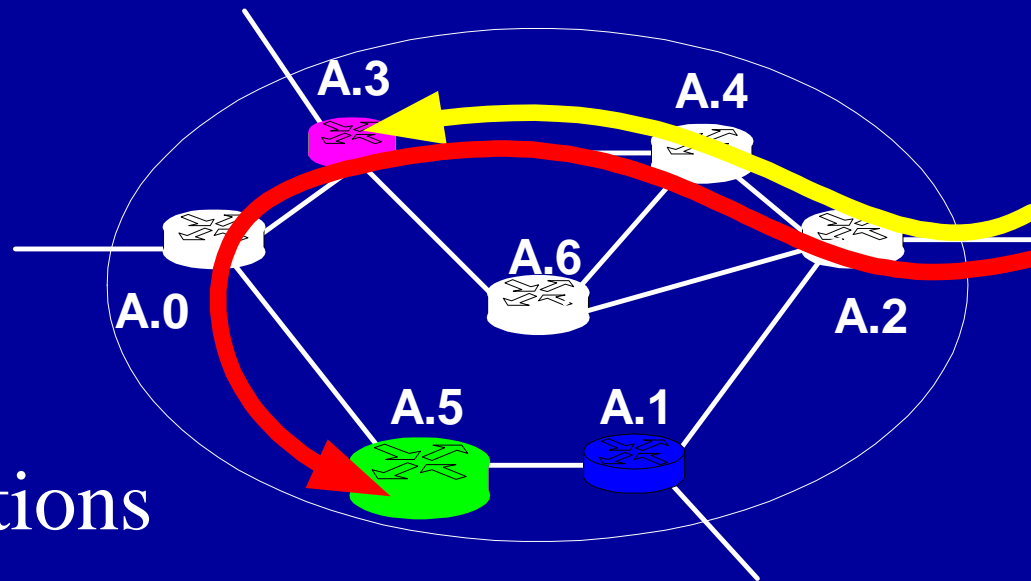


Path-and-node Deployment

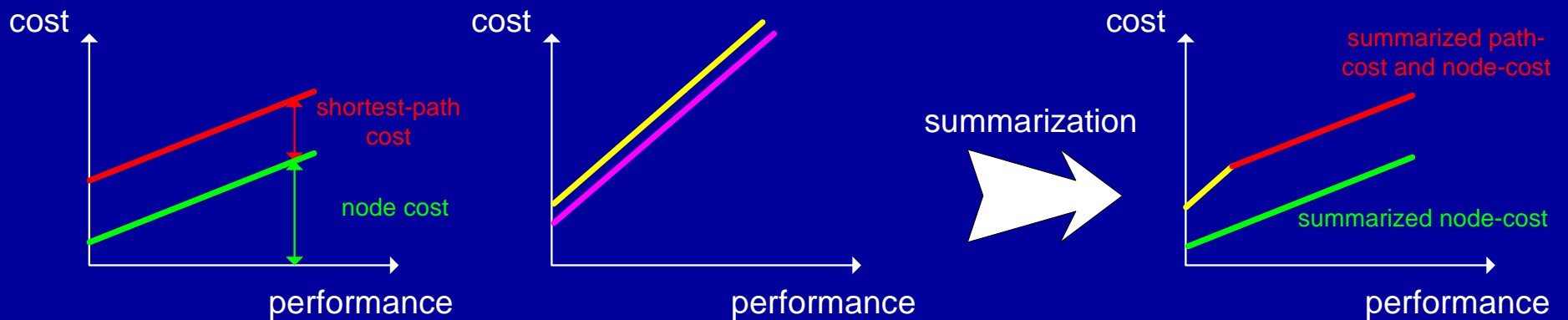
- VPN: place the end-points in the most appropriate hosts in each of the customer “islands”.



Path-and-node Deployment



- Cost functions



Cost Functions

- Extension to Service Deployment for cost-awareness
 - Additional dimension (basic information type)
 - Suitable for summarization
 - Allows cost-to-performance evaluation

Conclusion

- Internet evolves from providing pure connectivity, to high bandwidth, to custom services.
- Programmable networks bypass repetitive & lengthy protocol standardization
 - "once-for-all" standardized APIs
- Challenges for the intelligent network infrastructure:
 - rapidly deploying services
 - managing heterogeneity (in capabilities and interfaces)
 - supporting unpredicted services
 - remain scalable