Service Placement in Network-aware Cloud Infrastructures

“Visions of Future Generation Networks” (EuroView2009)

Andreas Reifert
andreas.reifert@ikr.uni-stuttgart.de
July, 28th 2009

Universität Stuttgart
Institute of Communication Networks
and Computer Engineering (IKR)
Prof. Dr.-Ing. Andreas Kirstädter

Agenda

Motivation

- Can the Cloud host telecommunication services?
- Integrated infrastructure

Service and infrastructure model

Service placement

- Algorithms
- Evaluation methodology

Results

Conclusion

Motivation

Telecommunication service requirements

Maximal network delay guarantees / Maximal response time guarantees

- Due to
  - Interactivity
  - State lookup
  - Data synchronization

→ Particular locations of components

→ Specialized components

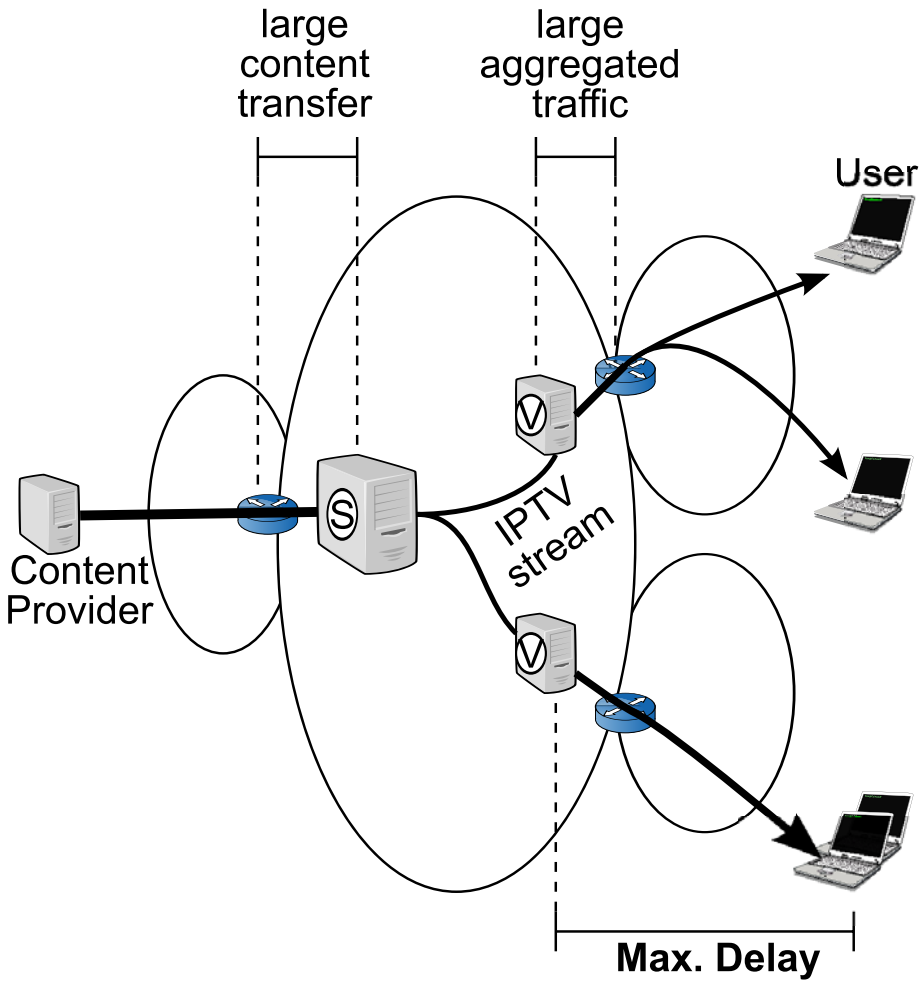
→ Replicated components

Bandwidth guarantees

- Due to
  - Stream transfers
  - Large content transfers

→ Particular locations of components

→ Specialized components



Motivation

Telecommunication service requirements

Maximal network delay guarantees / Maximal response time guarantees

- Due to
  - Interactivity
  - State lookup
  - Data synchronization

→ Particular locations of components

→ Specialized components

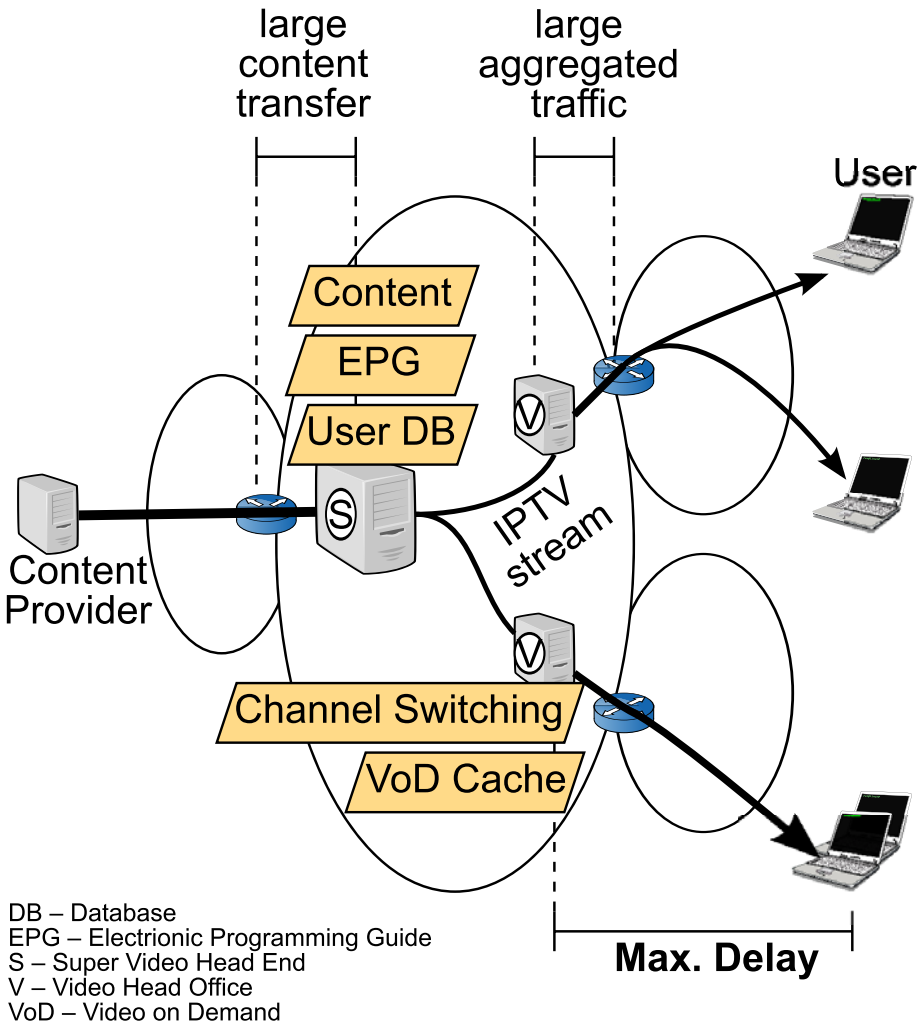
→ Replicated components

Bandwidth guarantees

- Due to
  - Stream transfers
  - Large content transfers

→ Particular locations of components

→ Specialized components



Motivation

Telecommunication services in the Cloud?

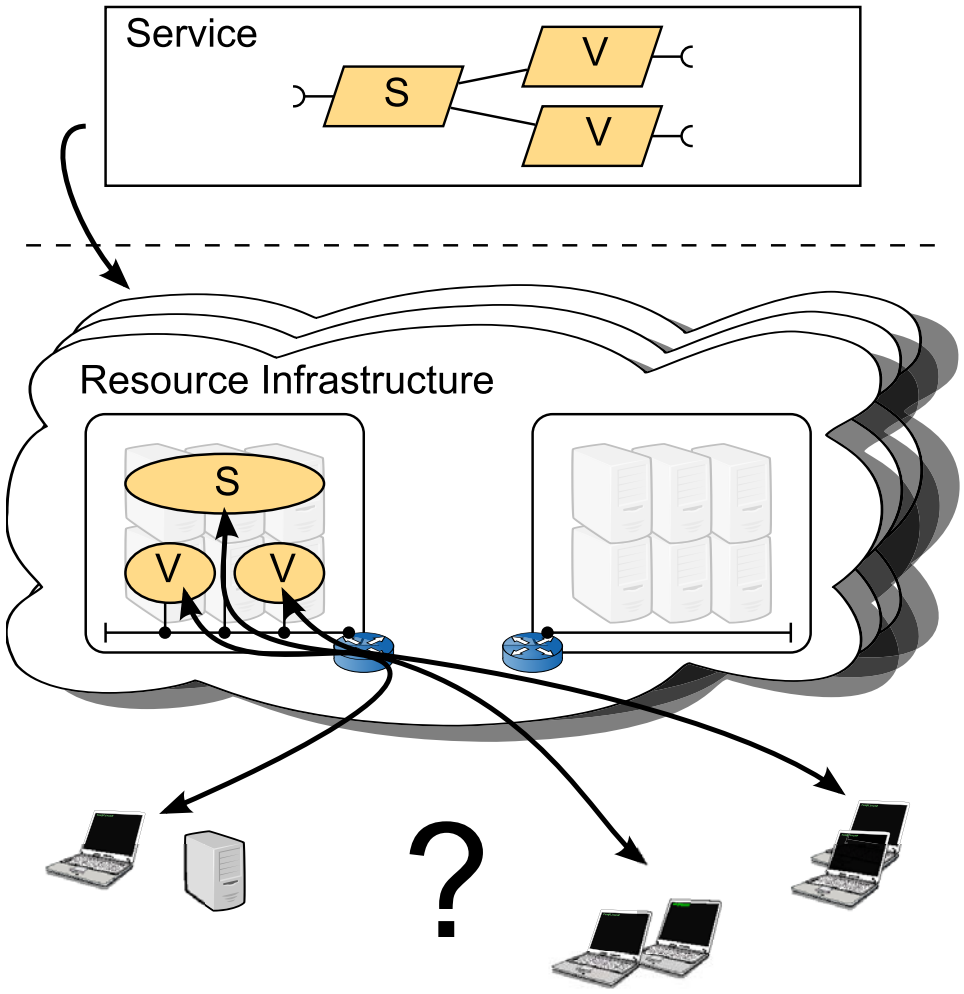
Cloud = Infrastructure as a Service Cloud

Decoupling of

- Service Provider

  is customer of

- Resource Infrastructure Provider (Cloud Provider)



Components

- Loosely coupled
- Placement / location not relevant
- Usually one location
  - Sufficient bandwidth
  - Low inter-component delay

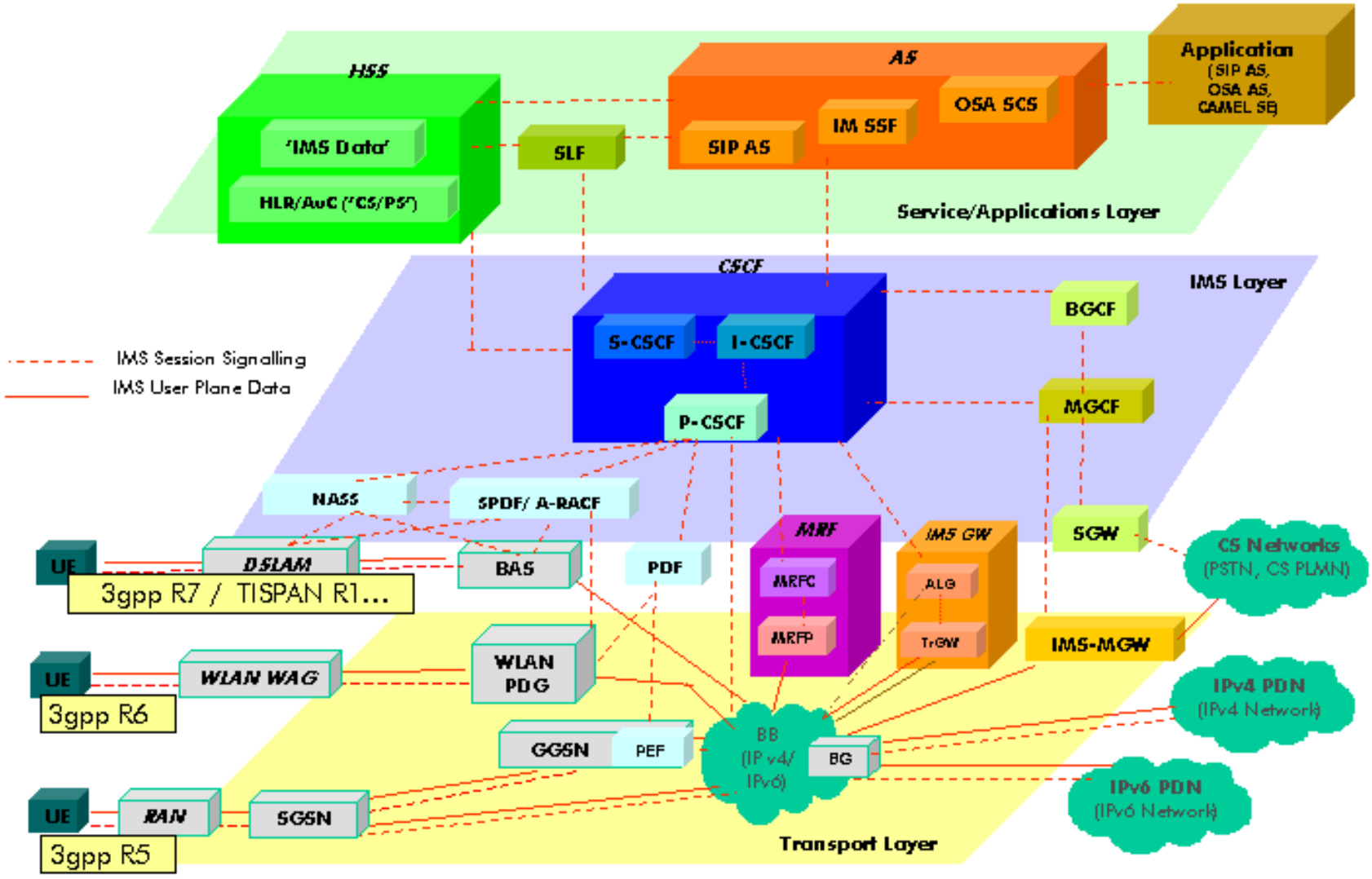
No Network View

→ **No delay guarantees**

→ **No bandwidth guarantees**

Motivation

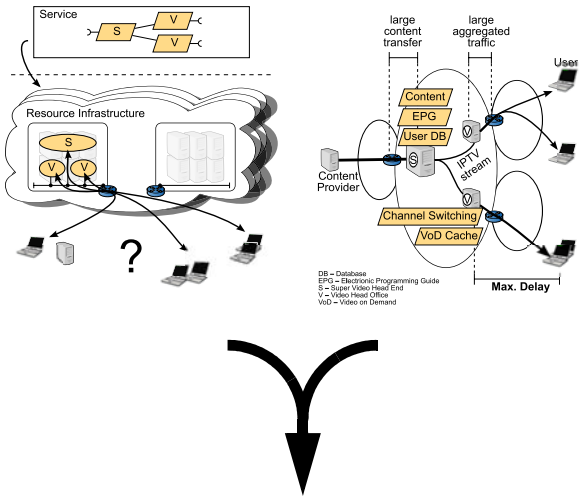
IMS – Another real world telecommunication service



Source: Wikipedia

Motivation

Integrated Infrastructure and Network



BMBF Project MAMS/MAMSplus

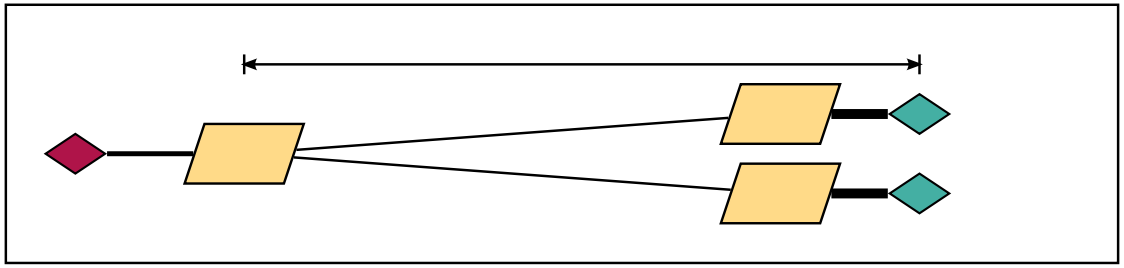
- Simple communication service creation and execution environment for non-experts
- …, Intelligent Service Oriented Network Infrastructure, …
- Concepts and prototype

Service AND resource management

Integrated view of services, infrastructure, and network necessary

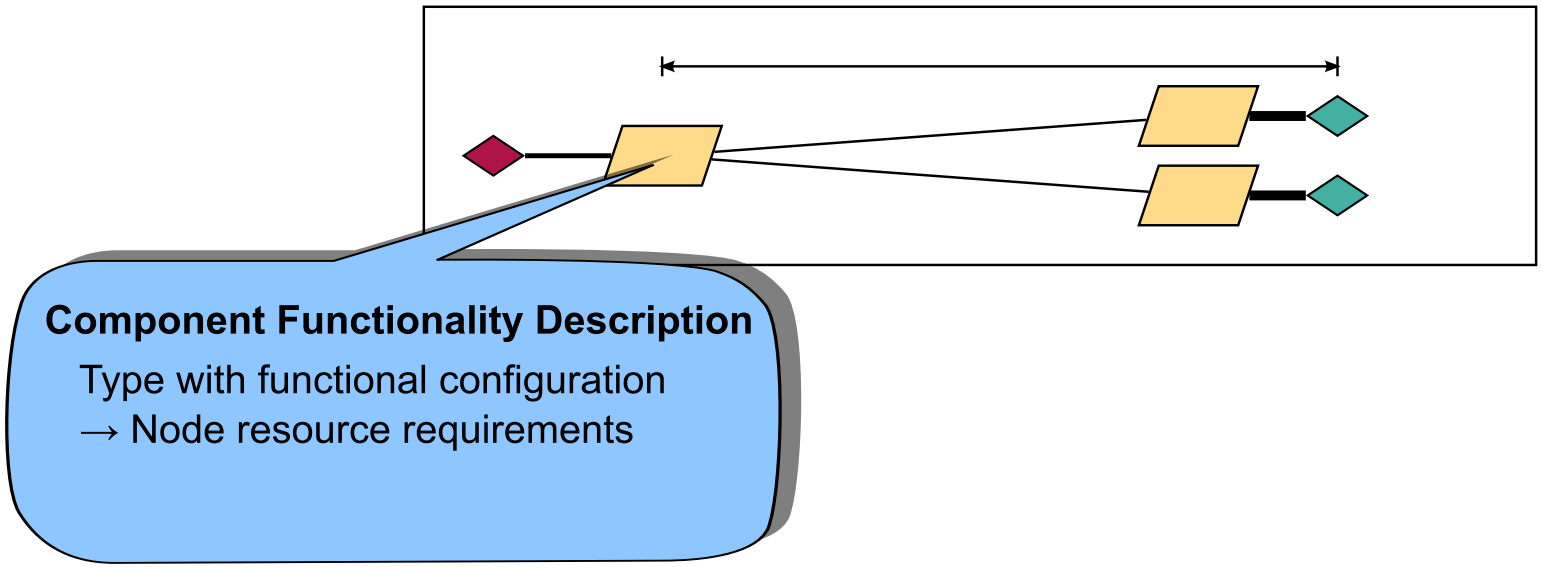
Model

Service Description



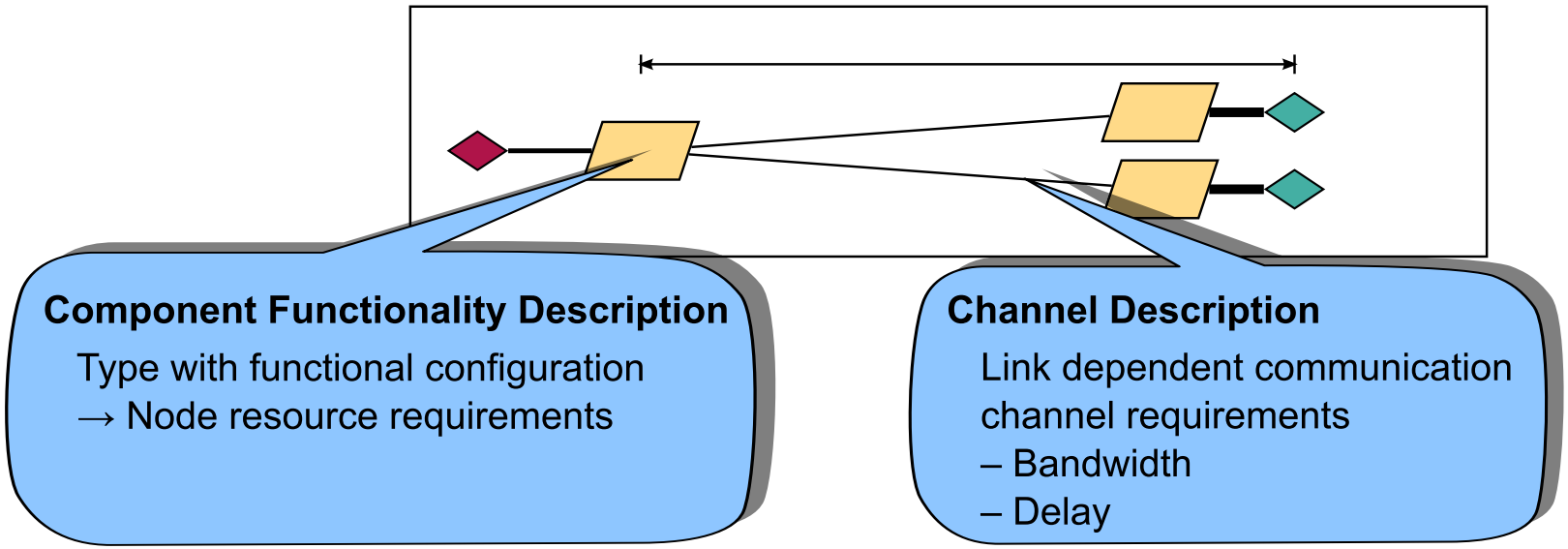
Model

Service Description



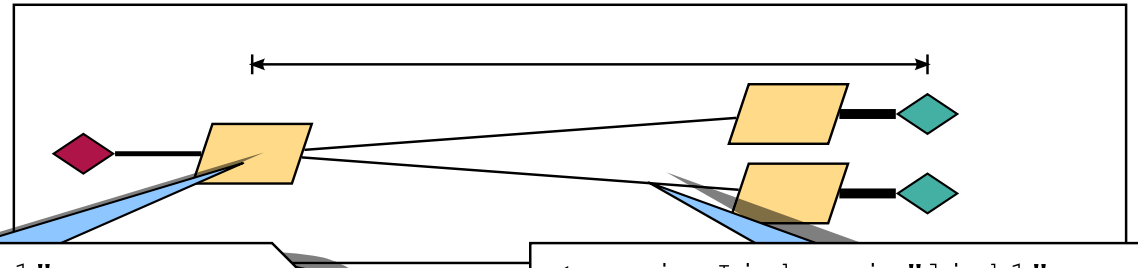
Model

Service Description



Model

Service Description

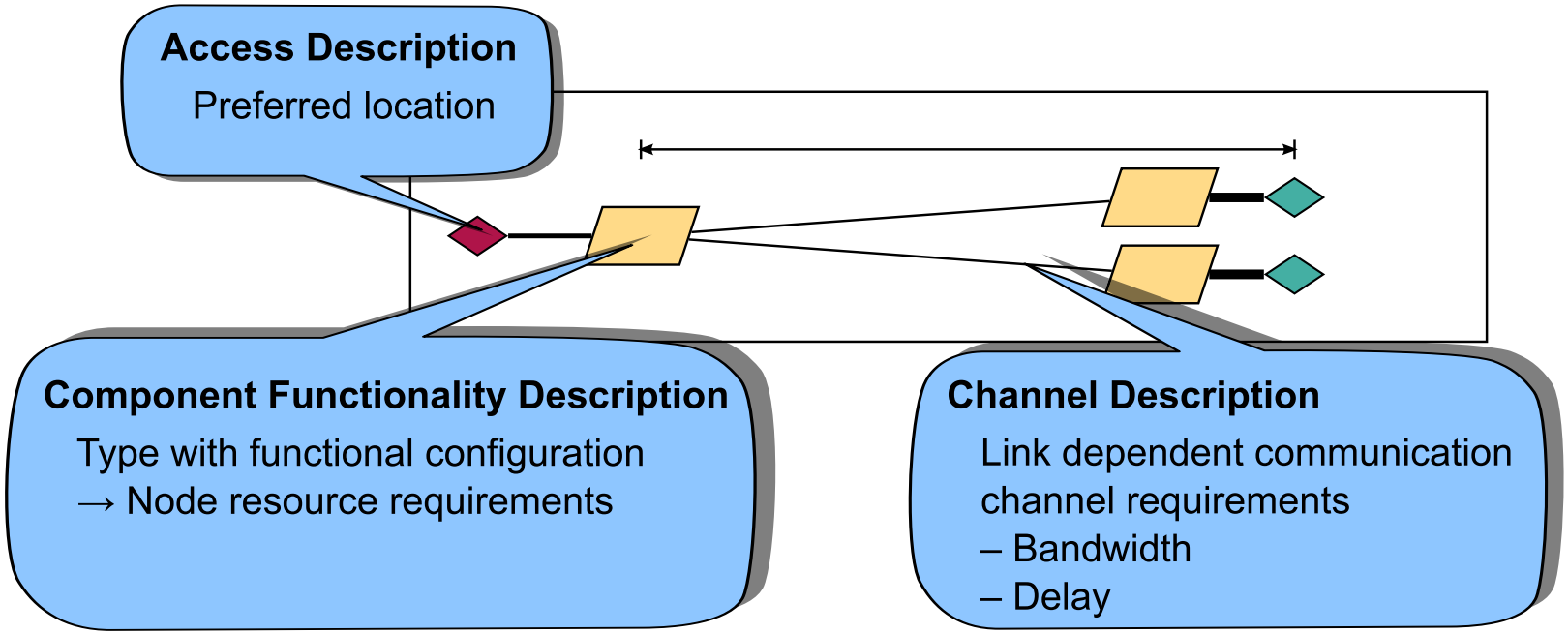


```
<WiSC uri="instance1"
      type="functional">
  <endpoint>10.0.1.1</endpoint>
  <protocol name="SuperVideoHeadEnd"
            version="1.0"
            contextType="Iptv">
    <methodProfile>standard
    </methodProfile>
  </protocol>
  <argument ref="VoDContent"
            unit="M">
    <value>10</value>
  </argument>
  <argument ref="LiveContent">
    <value>2000</value>
  </argument>
</WiSC>
```

```
<serviceLink uri="link1"
             type="bidirectional">
  <endpoint>10.0.1.1</endpoint>
  <endpoint>10.0.1.2</endpoint>
  <resourceRequirement uri="1">
    <ofClass
      ref="link/bidirectional" />
    <capacity ref="bandwidth"
              unit="Mbps">
      <value>300</value>
    </capacity>
    <specialProperty ref="delay"
                     unit="ms">
      <maxvalue>200</maxvalue>
    </specialProperty>
  </resourceRequirement>
</serviceLink>
```

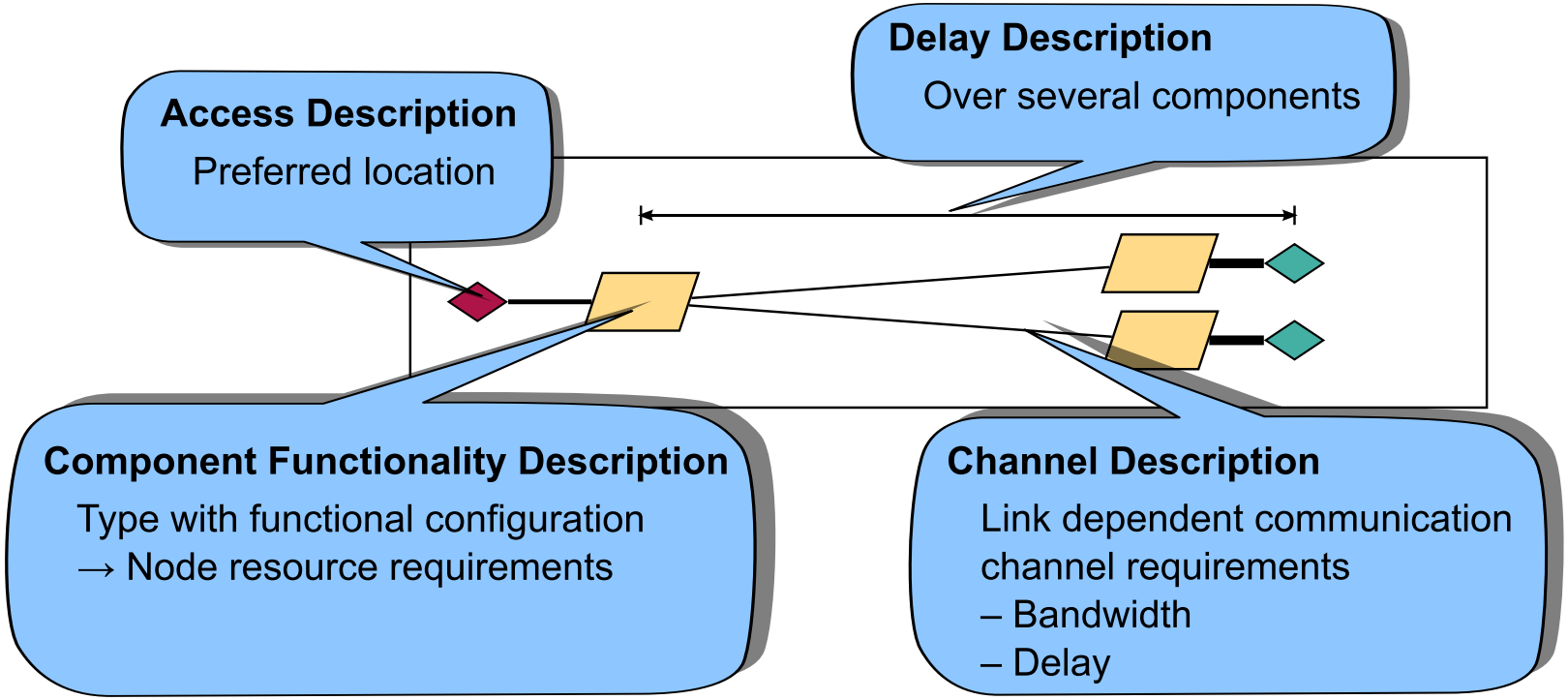
Model

Service Description



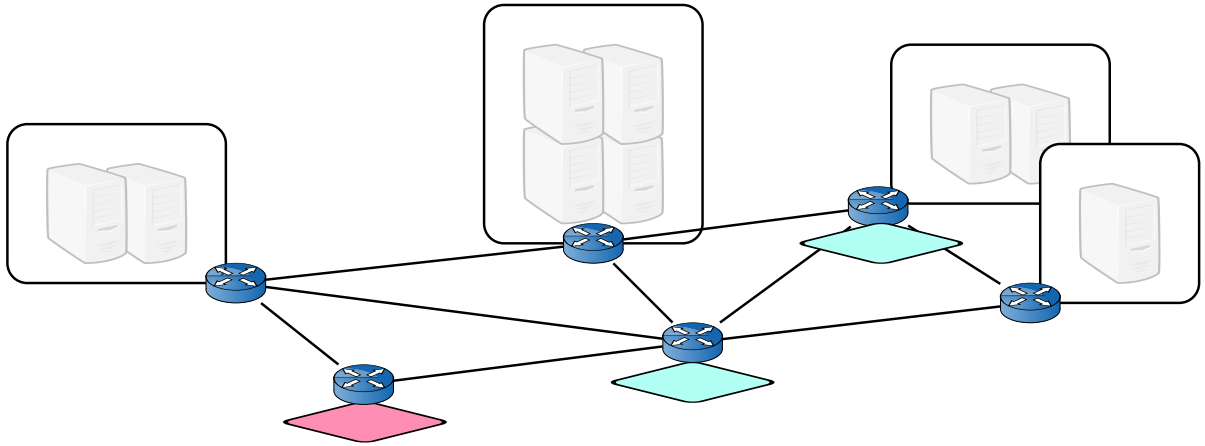
Model

Service Description



Model

Resource Infrastructure / Network

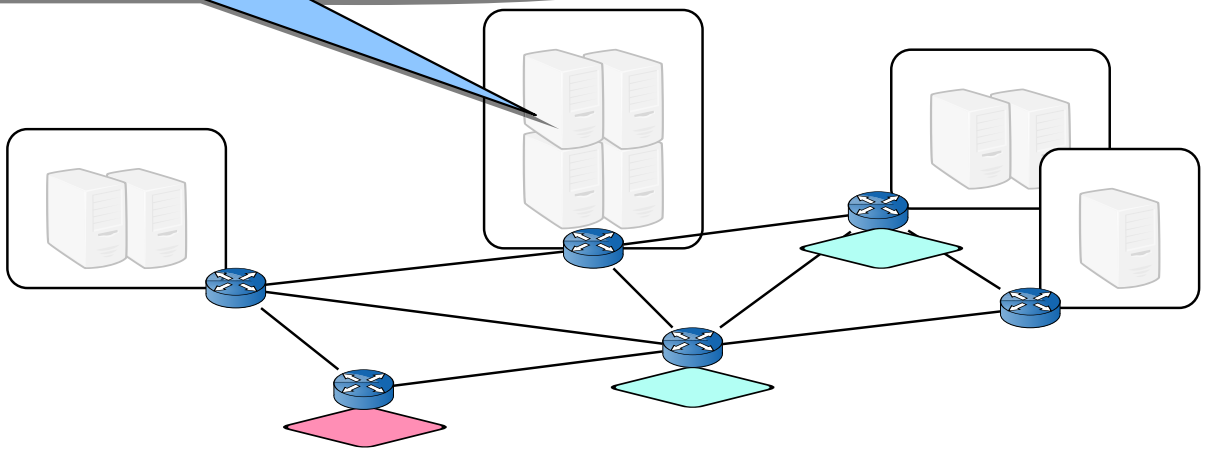


Model

Resource Infrastructure / Network

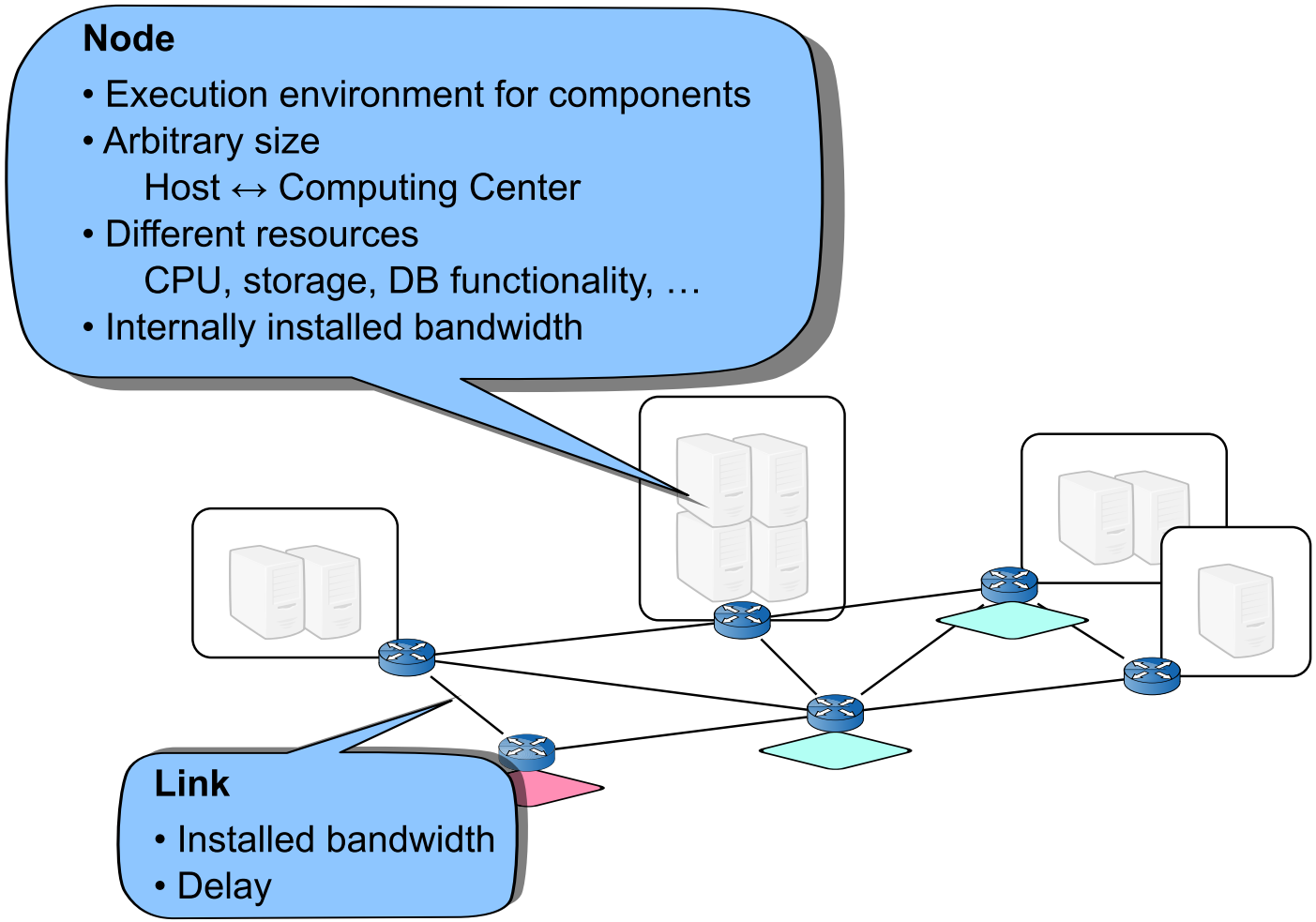
Node

- Execution environment for components
- Arbitrary size
  - Host ↔ Computing Center
- Different resources
  - CPU, storage, DB functionality, …
- Internally installed bandwidth



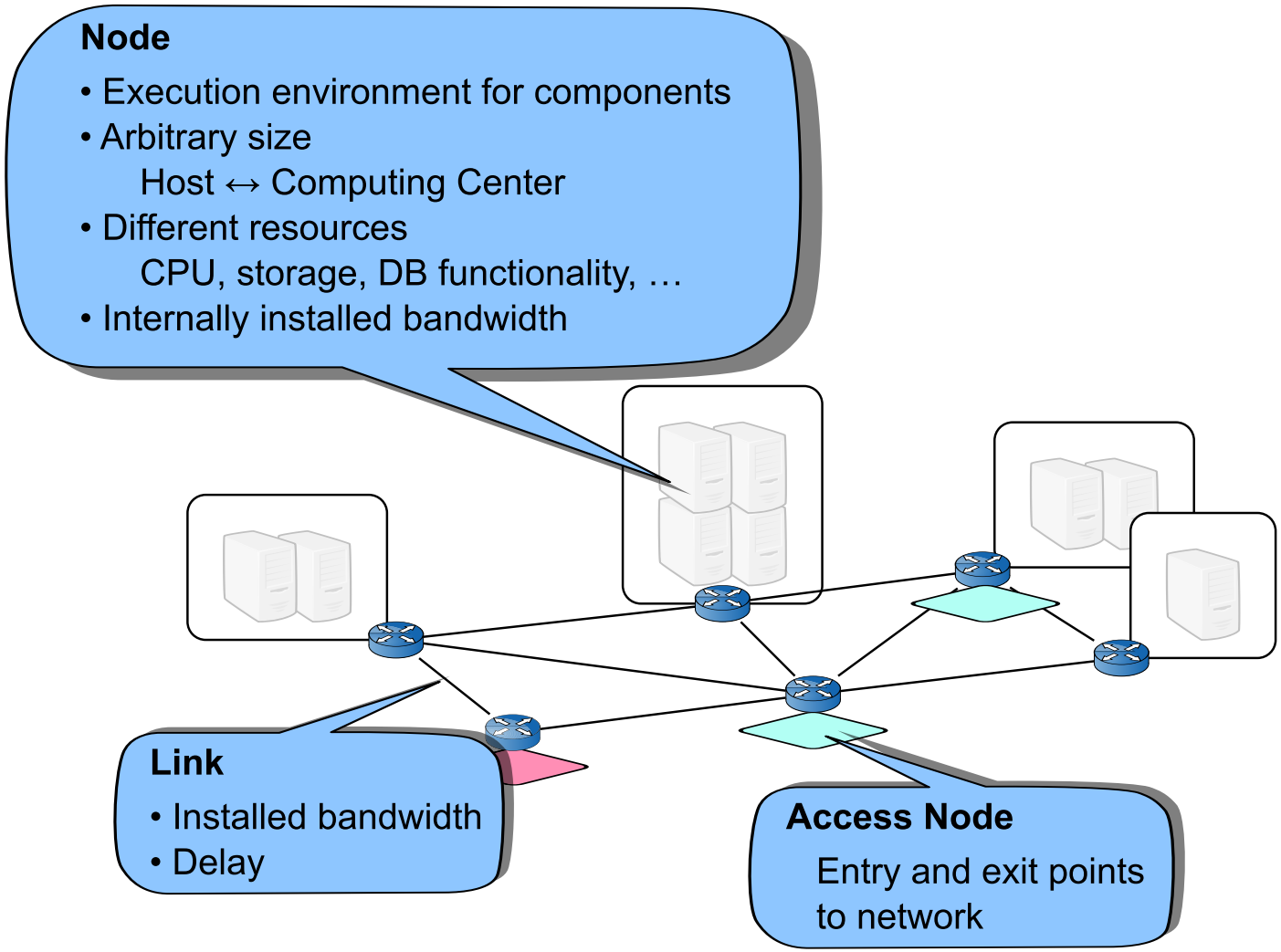
Model

Resource Infrastructure / Network



Model

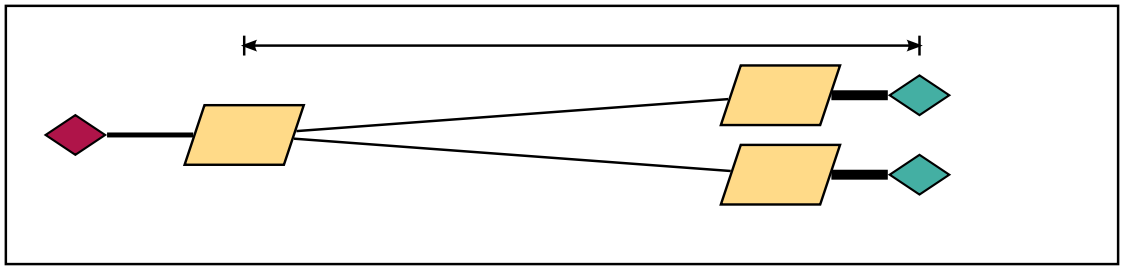
Resource Infrastructure / Network



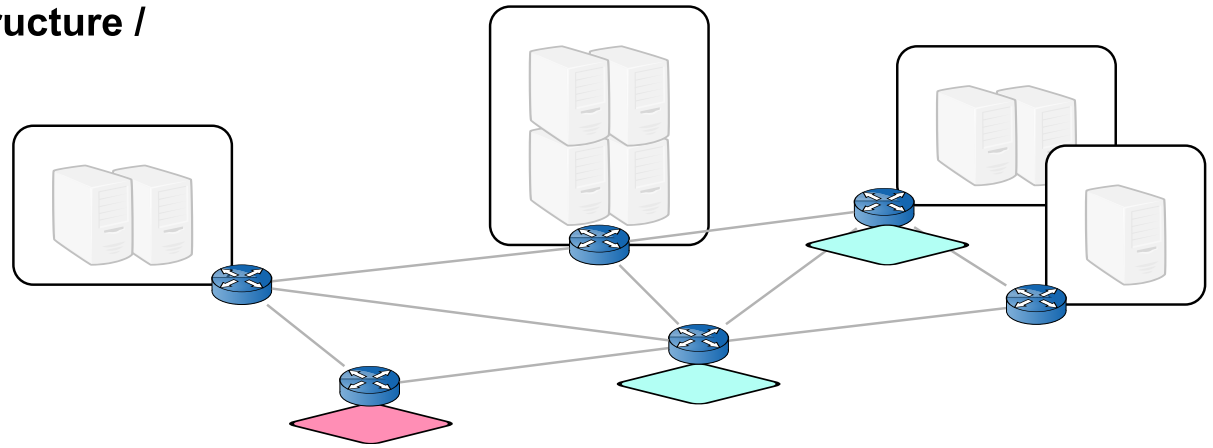
Model

Placement

Service Description

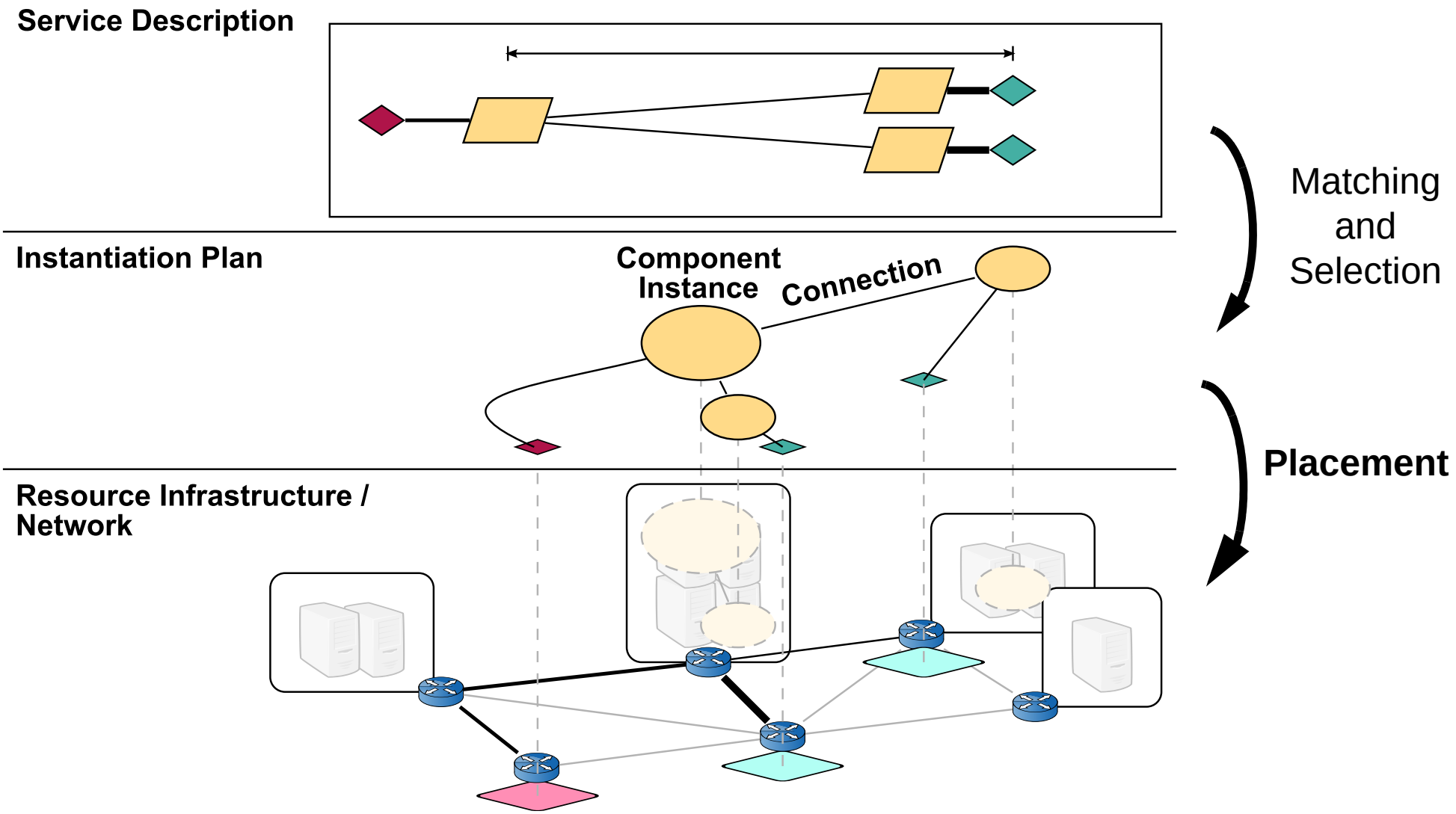


Resource Infrastructure / Network



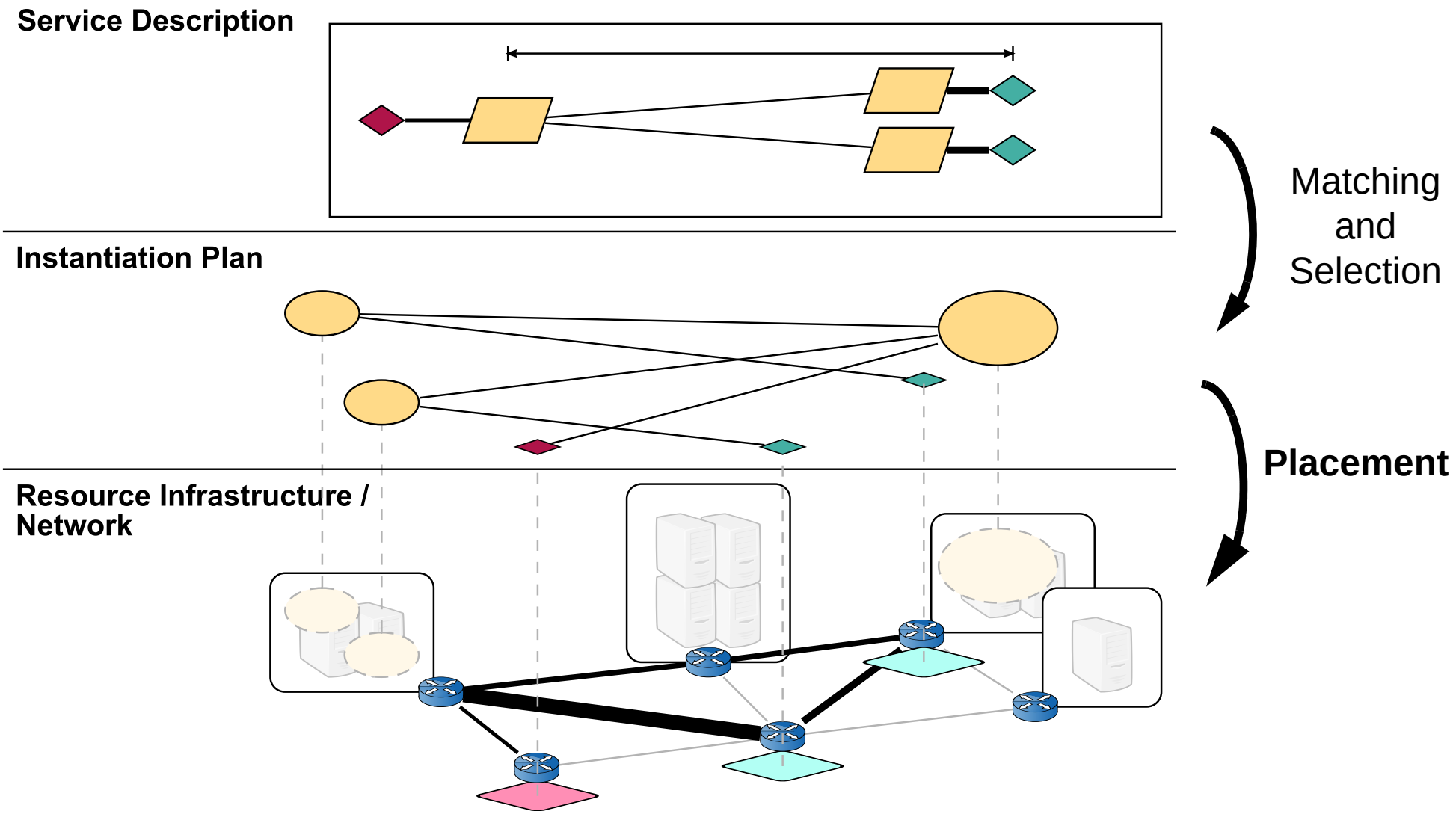
Model

Placement



Model

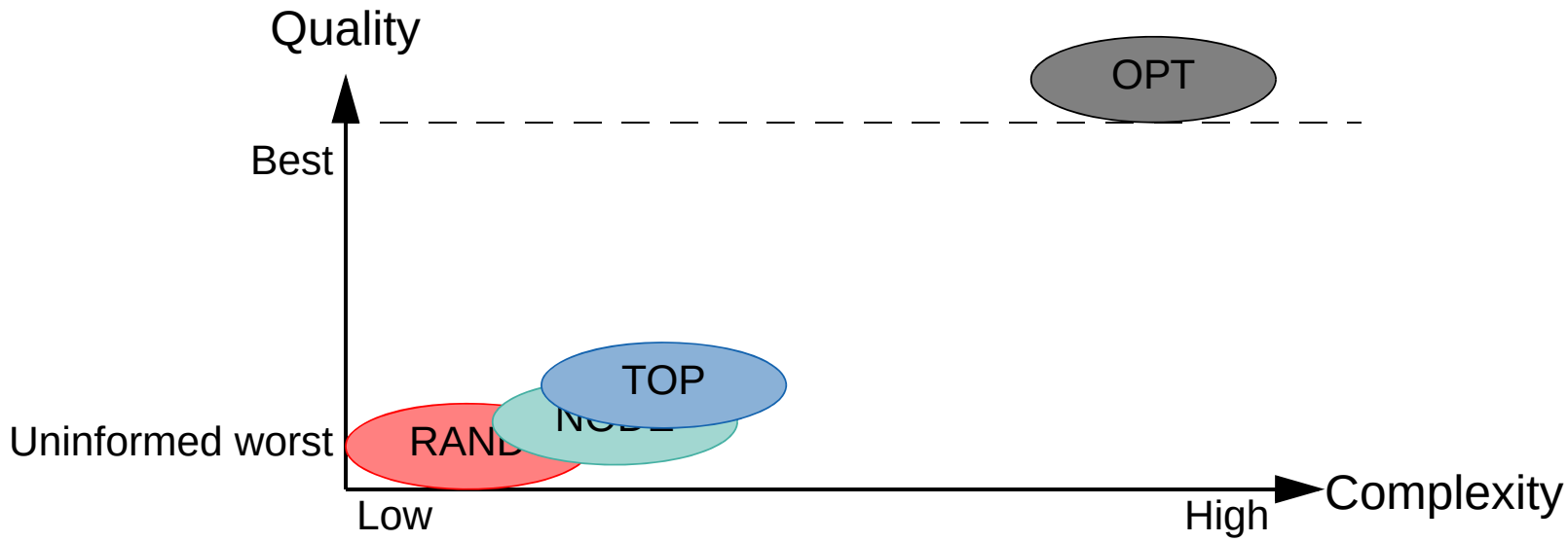
Placement



Placement

Strategies

	RAND	NODE	TOP	OPT
Type	Random	Greedy	Greedy	Optimal
Principle	Uninformed	Node-based (Only Node resources)	Topology-based (Service and Network inc. resources)	MILP (Mixed Integer Linear Program)

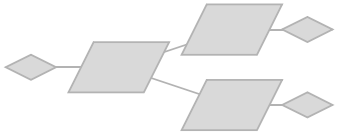


Placement

Evaluation Methodology

Monte Carlo Simulation

- Independent samples with random service placed on random infrastructure/network
- Parameters
  - Allocated resources in infrastructure/network 0% (Empty) ◄——► 100% (Full)
  - Characteristic of service

		Topology	
		Centralized	**Distributed**
Total Resource Demand	**Heavy-weight 80%**		
	Light-weight 20%		

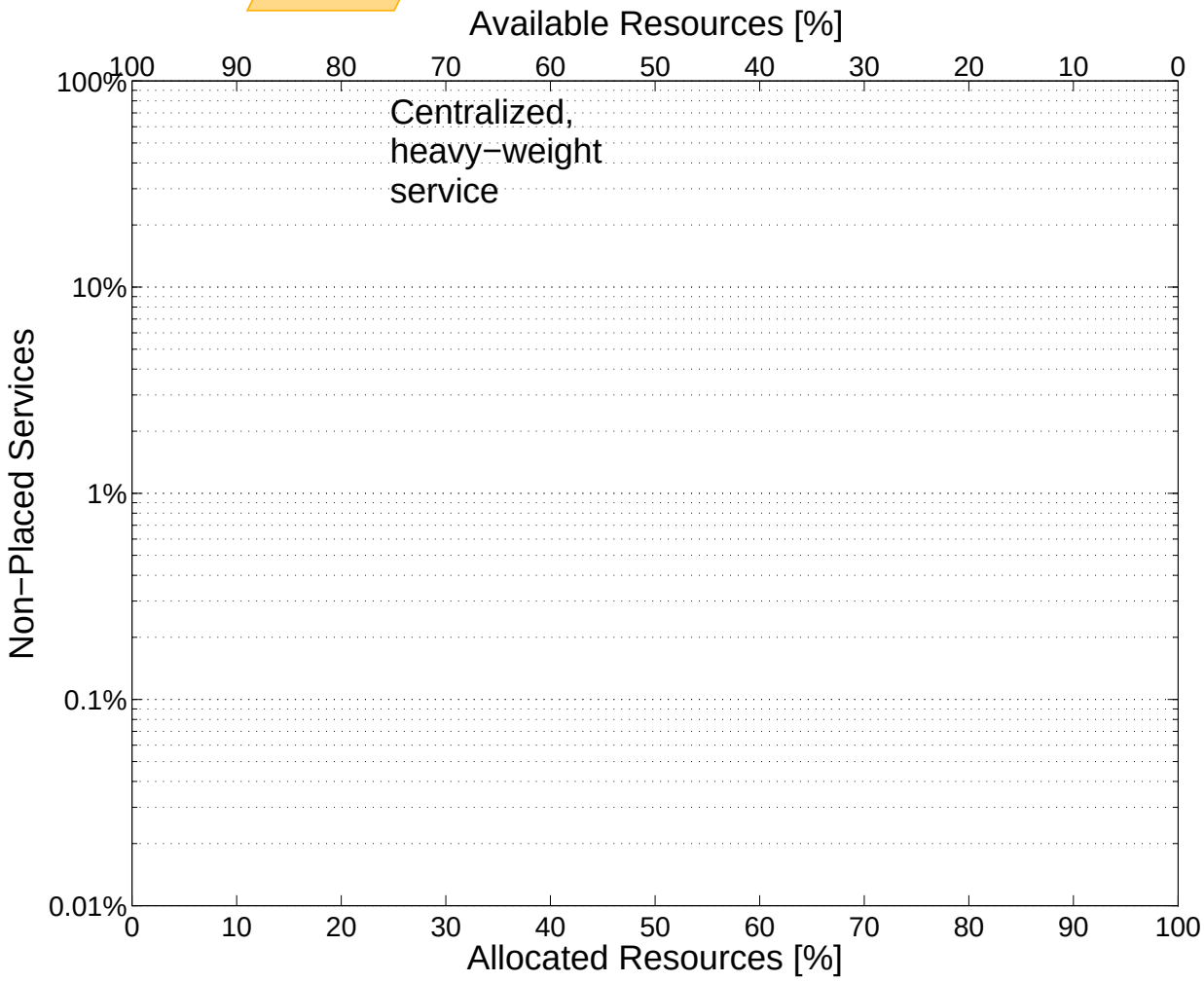
Performance Metrics

- Ability Fraction of non-placed services (Rejection)
- Quality Comparison of total link bandwidth allocation w.r.t. OPT allocation

Results

Ability to Find Placement – Centralized, heavy-weight service

Centralized heavy-weight service

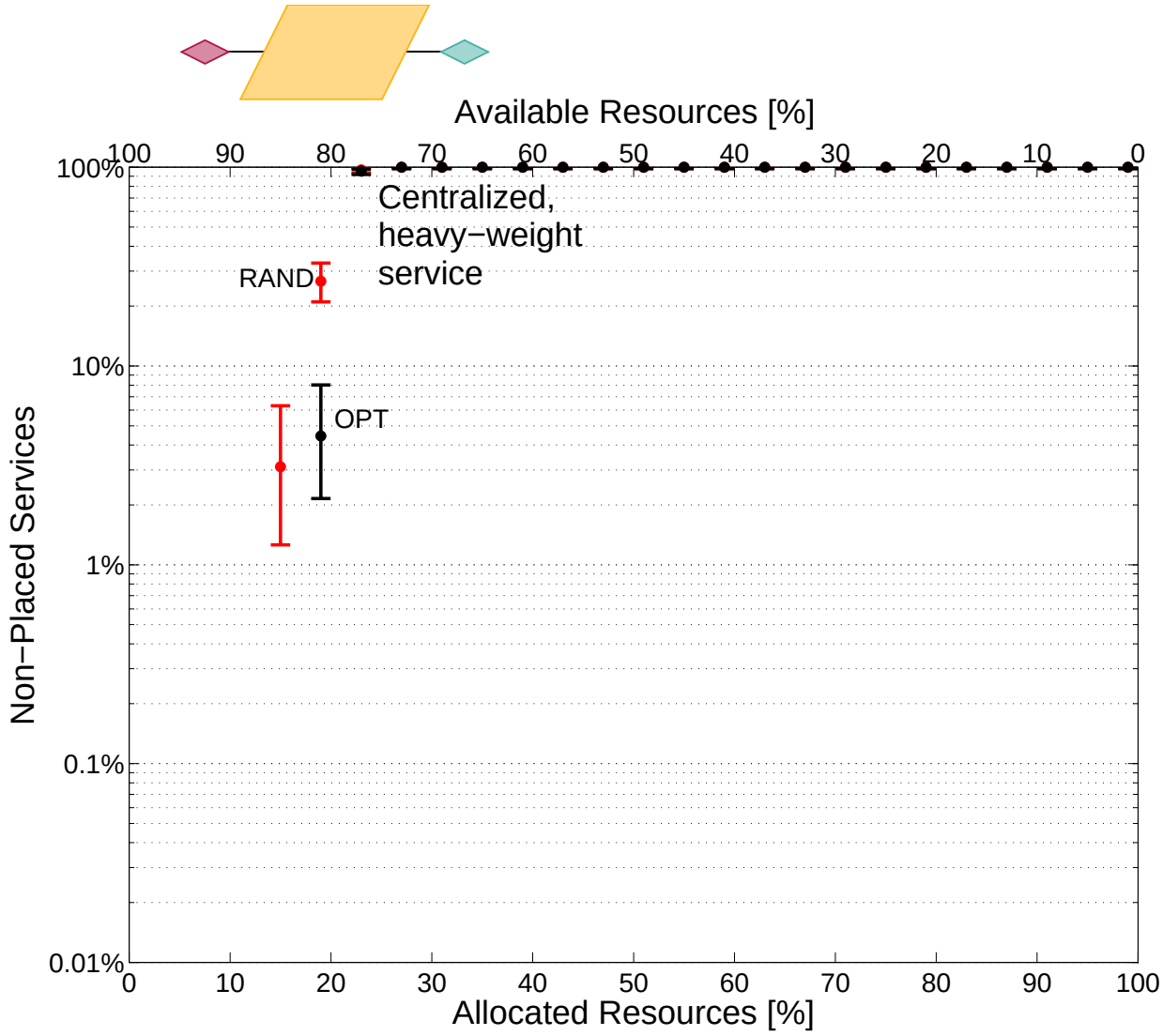


Results

Ability to Find Placement – Centralized, heavy-weight service

Centralized heavy-weight service

- Low optimization potential

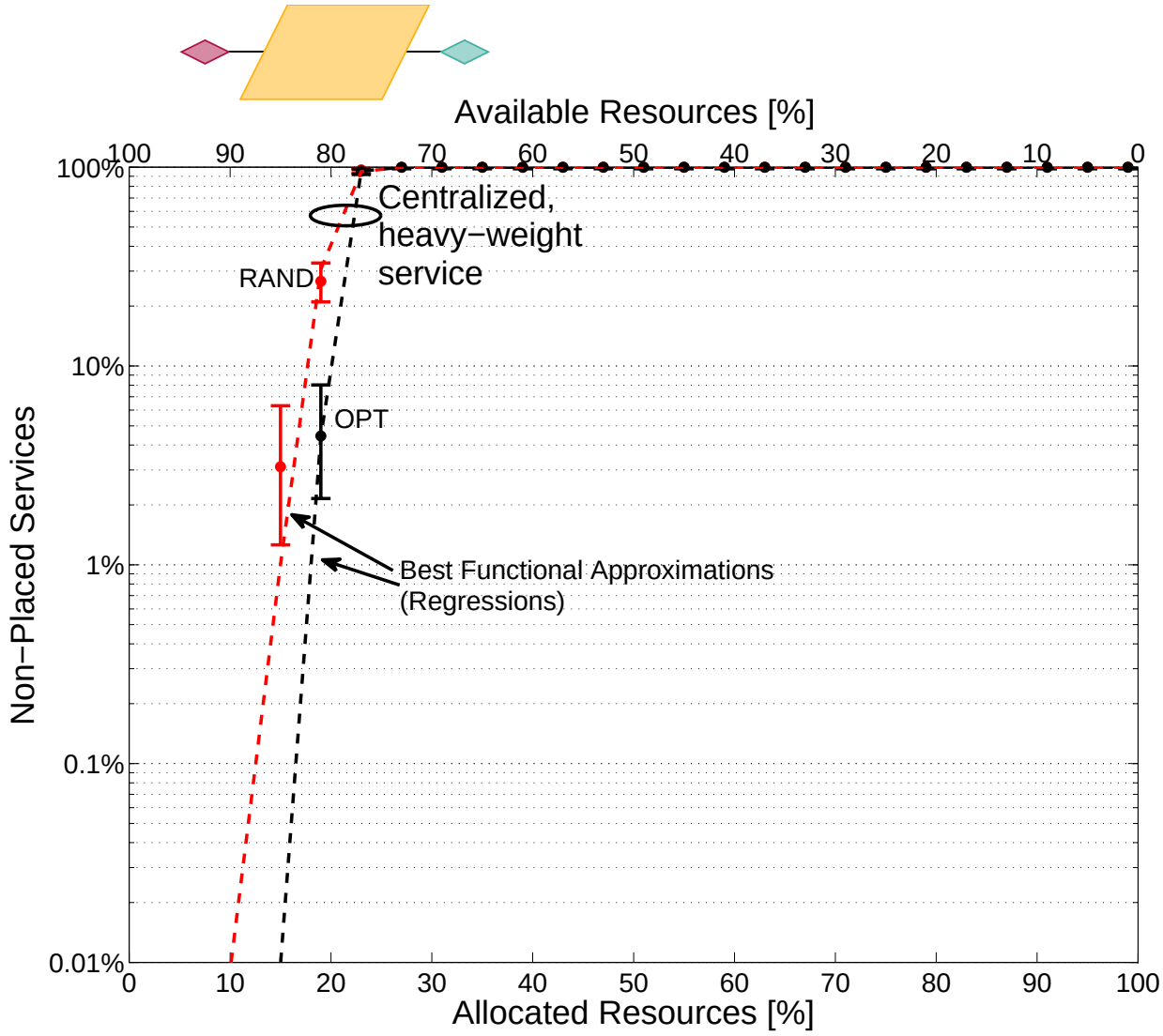


Results

Ability to Find Placement – Centralized, heavy-weight service

Centralized heavy-weight service

- Low optimization potential



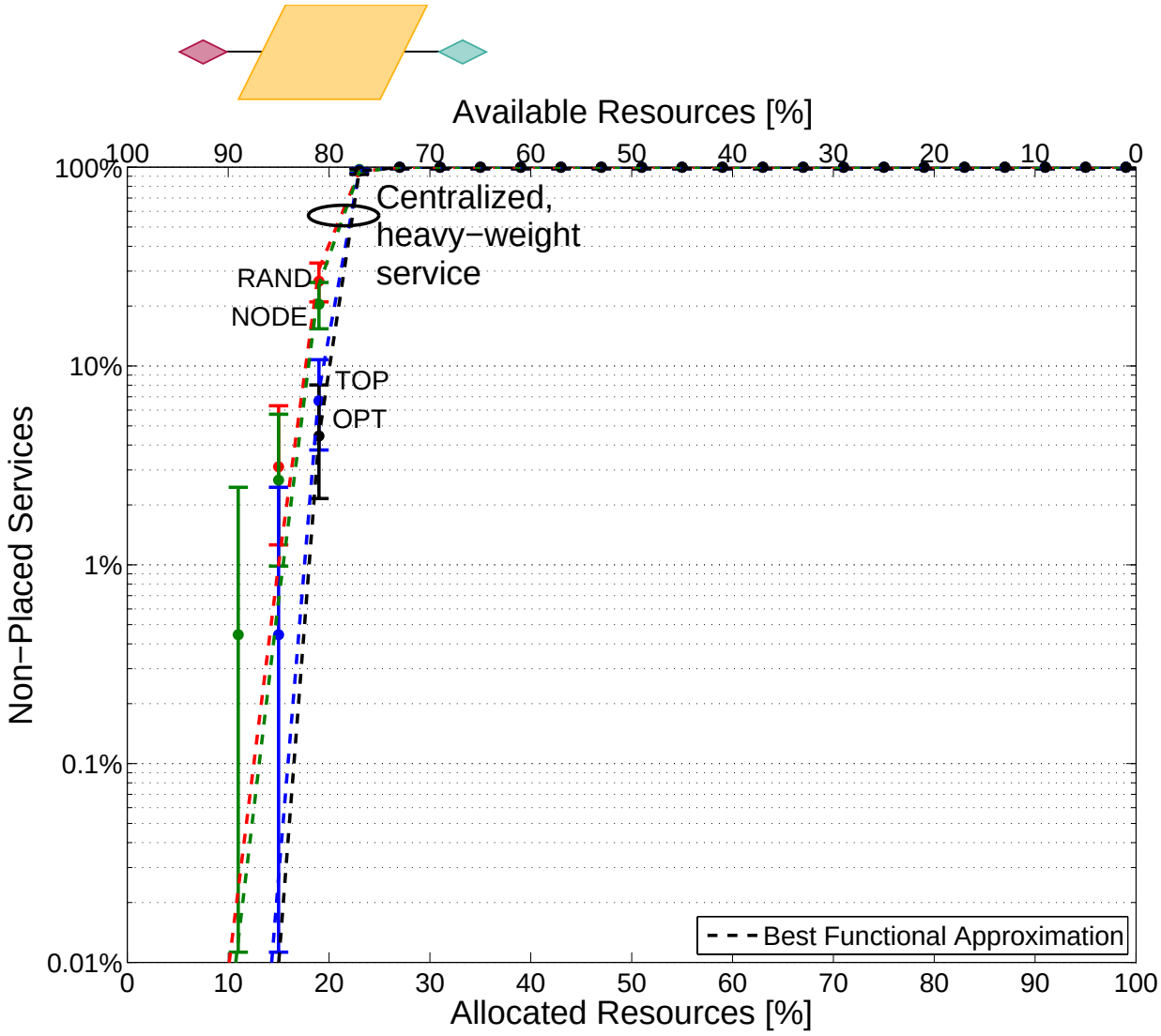
Results

Ability to Find Placement – Centralized, heavy-weight service

Centralized heavy-weight service

- Low optimization potential
- TOP close to OPT

→ **Service- and network-topology matter**



Results

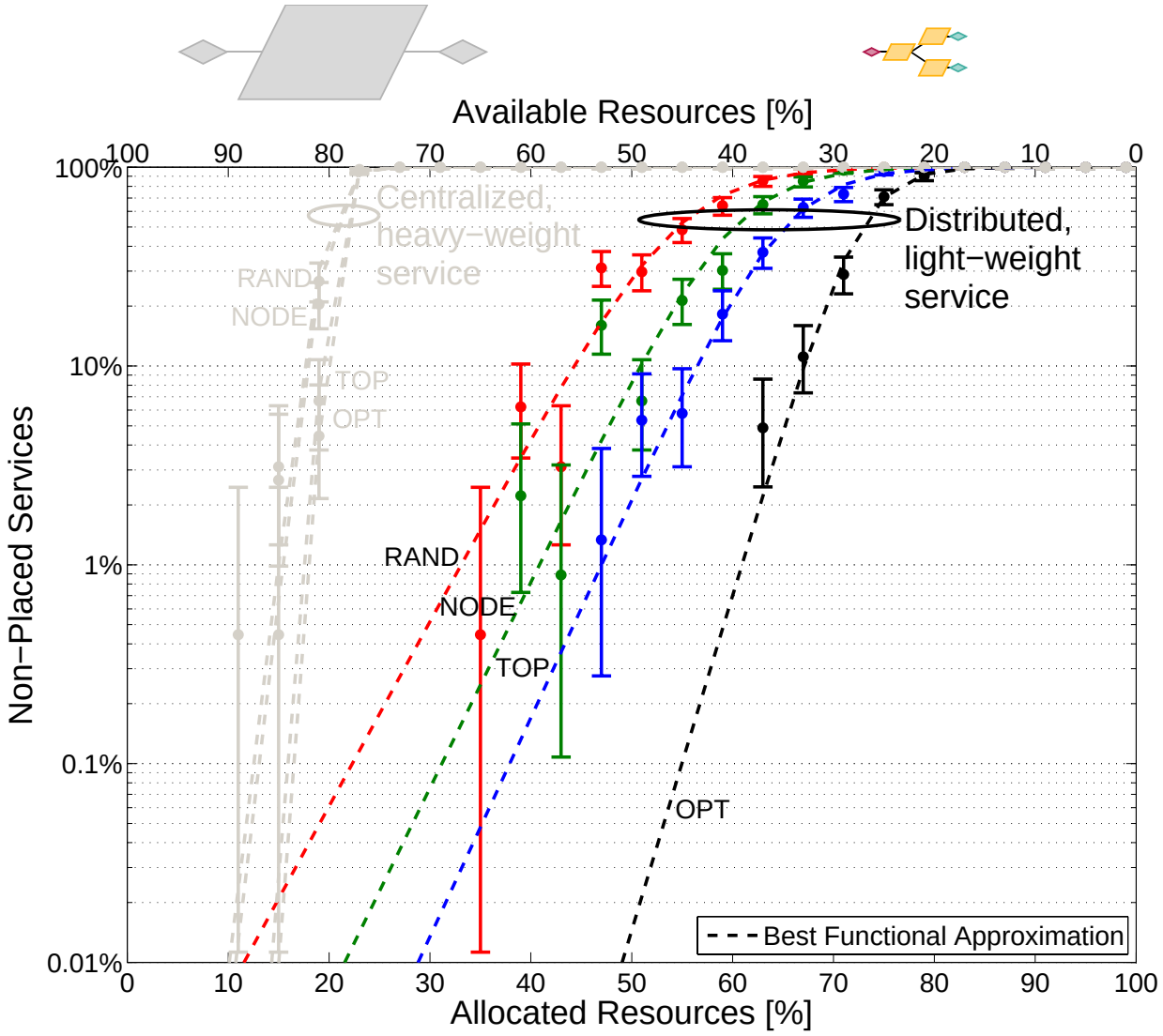
Ability to Find Placement – Distributed, light-weight service

Distributed, light-weight service

- Significant observed differences in algorithmic behavior
- High optimization potential
  - Up to several orders of magnitude
  - Even between TOP and OPT

→ **Service- and network-topology matter**

→ **Simple algorithms leave significant room for improvement**



Results

Quality of Found Placement – PRELIMINARY RESULTS

Centralized, heavy-weight service

- TOP almost optimal

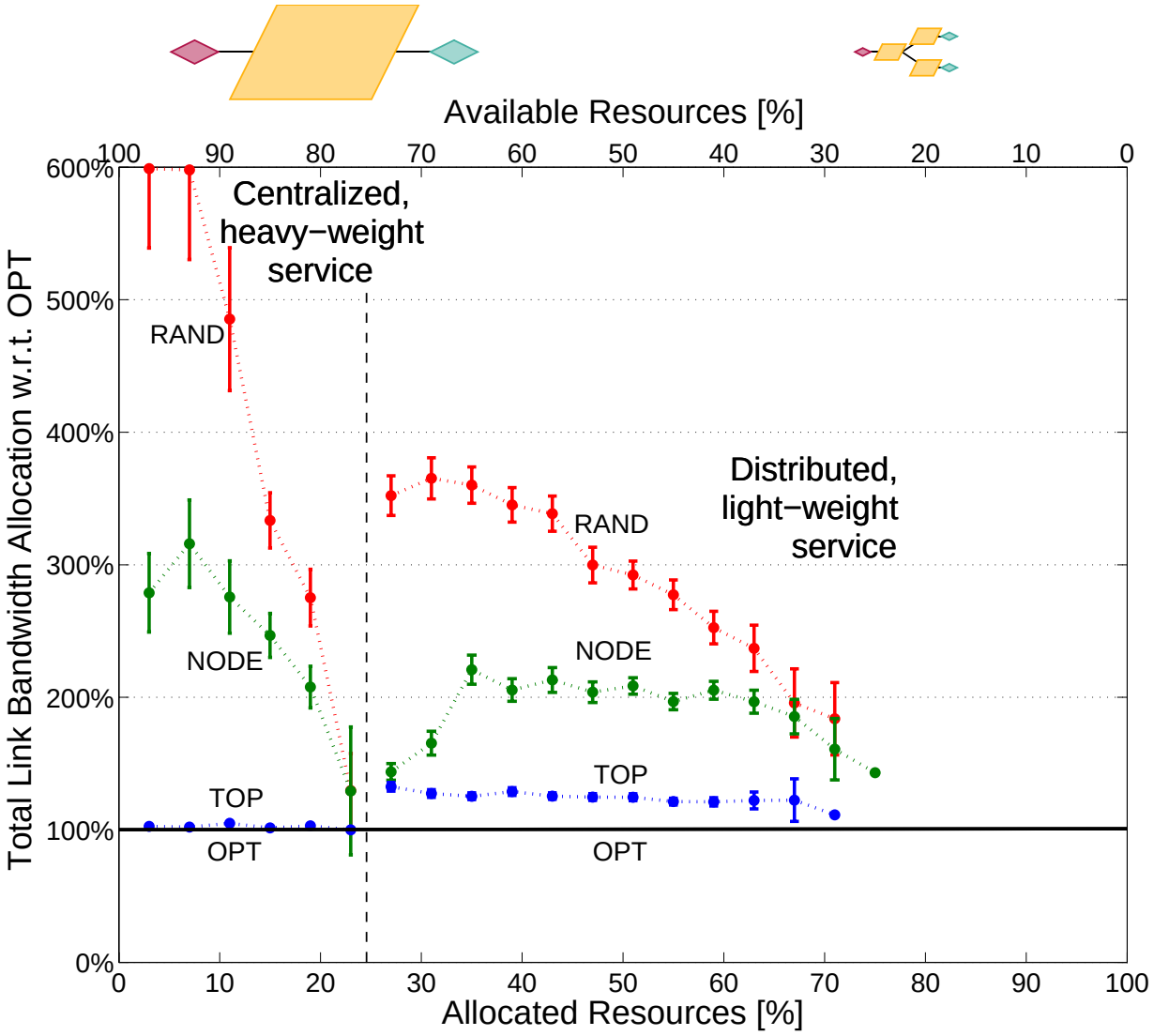
Distributed, light-weight service

- TOP with “acceptable” placements

  If found!

- Behavior of NODE not yet understood
- Improved performance in high occupancy region due to few possible placements

Improvements without modification to routing!



Conclusion

- Current IaaS Clouds not prepared for telecommunication services
- Network view essential for channels between components and towards end-systems
  - Delay requirements
  - Required bandwidth guarantees
- Integrated view “Service/Infrastructure/Network” necessary for system management

  → Detailed model

- Placement of components has significant impact on
  - Number of running services
  - Bandwidth consumption
- Good placement algorithms must match service- and network-topology
  - Especially for distributed services
  - TOP leaves room for improvement