Posrednici umreženih sustava

Prof. dr. sc Siniša Srbljić

Dr. sc. Ivan Benc

Fakultet elektrotehnike i računarstva,
Zavod za elektroniku, mikroelektroniku, računalne i inteligentne sustave

Quality of Service in Middleware Systems

Dr. sc. Daniel Skrobo

Fakultet elektrotehnike i računarstva,
Zavod za elektroniku, mikroelektroniku, računalne i inteligentne sustave

Outline

- Introduction
- Quality of Service
- Quality of Experience
- Service Level Agreements
- Conslusion

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Introduction

- **Enforcing quality in computer systems is a dificult task that requires effort in a range of software engineering activities**

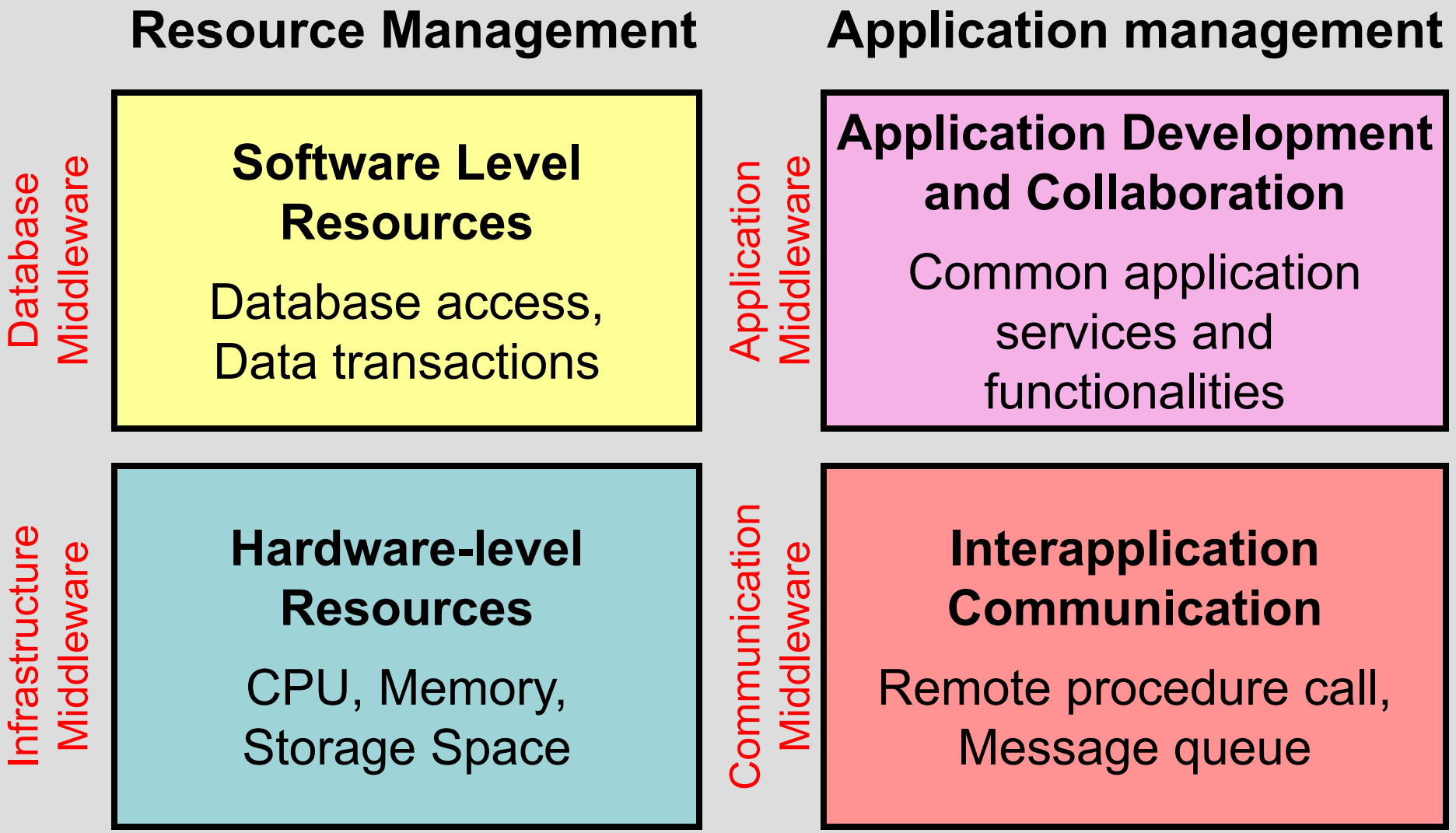
- **Some of the key features of software quality are**
  - Correctness
  - Completeness
  - Scalability
  - Fault-tolerance
  - Extensibility
  - Maintainability
  - Documentation

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Quality in Middleware Systems

- **Middleware systems must be designed with respect to quality assurance expectations of users**
  - Middleware includes open services and functionalities
  - Quality assurance is a basis for enforcing application-specific middleware services and functionalities
  - Service providers depend on quality assurance to deliver their end services to users over middleware platform

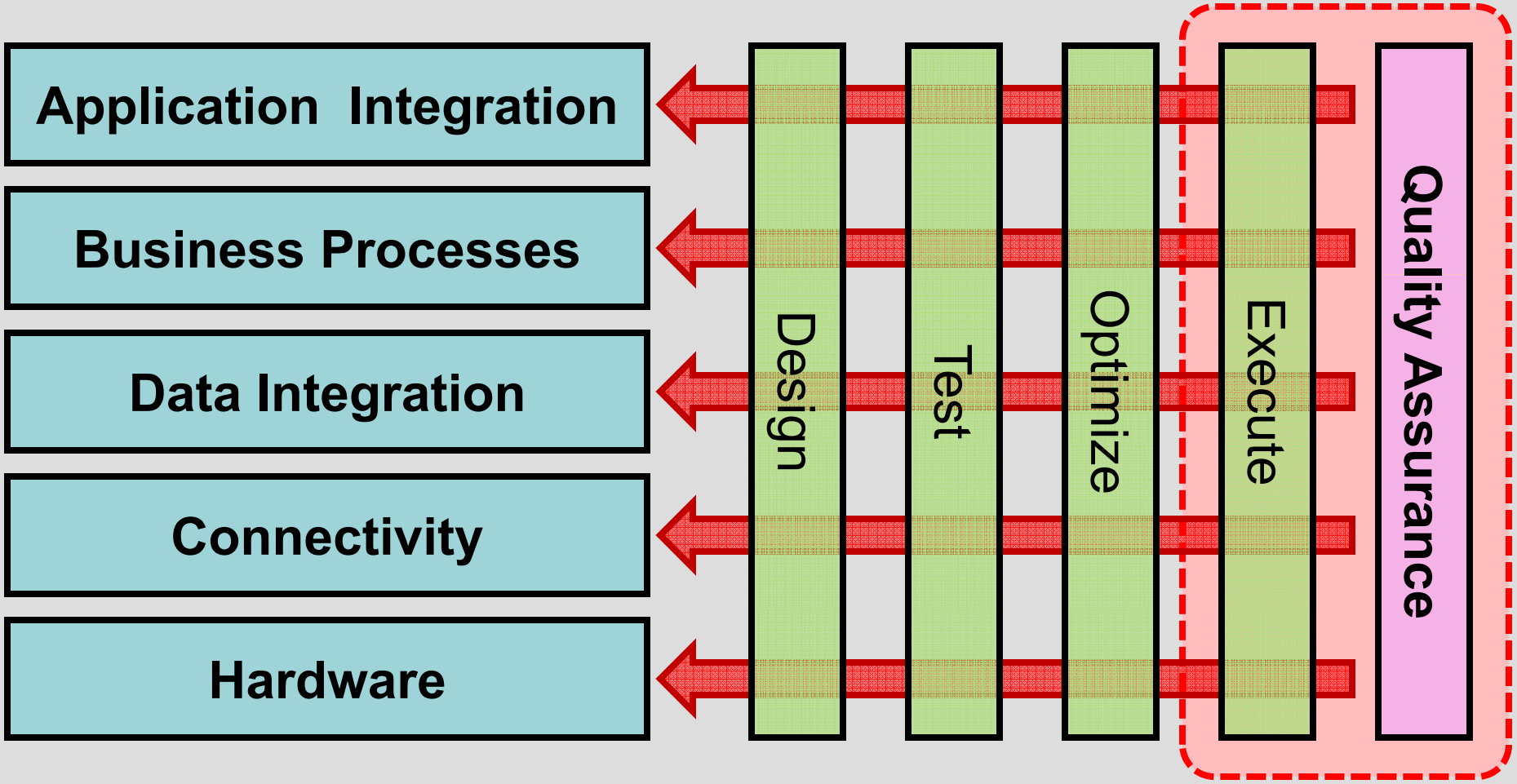
Middleware Classification



Quality in Middleware Layers

- **Differentiated classes of quality assurance**
  - Each type of middleware system requires unique quality of service assurance level
- **Combined quality of middleware platform**
  - Quality of each middleware layer provides support for the grand total quality of the entire middleware platform

Quality Assurance in Context



Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Quality Assurance in Context

- **Quality assurance in the engineering process**
  - Design and implementation (Code quality)
  - Test (Functional quality)
  - Optimization (Nonfunctional quality)

- **Quality assurance in the deployment and execution**
  - Execution (Quality of Service and Quality of Experience)

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Quality of Service and Experience

- **Quality of Service (QoS)**
  - Objective measure of characteristics of the service delivered provider to user

- **Quality of Experience (QoE)**
  - User's subjective impression on the quality of the used service

Effective Quality = Deliverables - Expectations

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Quality of Service

"It is the quality of our work which will please God and not the quantity"

~Mahatma Gandhi

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Quality of Service

- **Quality of Service (QoS) defines the functional and nonfunctional characteristics of a computing system that are delivered as a service to end-user**

- **QoS aspects**
  - Resources define the hardware and software artifacts that are served to the user
  - Metrics define the objective and consistent way to measure the quality of the delivered resources

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

QoS Mechanisms Enforcement

- **QoS resources and metrics**
  - Computing power [CPUh]
  - Storage space size [MB]
  - Network bandwidth [Mbps]
  - Network Delay [s]
  - Data quantity and quality [MB, %]
  - Computation precision [%]
  - Application usage time [s]

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

QoS Mechanisms Enforcement

- **QoS mechanisms enforcement**
  - Completeness
  - Type of Guarantees
  - Performance Type or Level
  - Connections
  - Degree of Support
  - Flexibility and Availability

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

QoS Mechanisms Enforcement

- **Completeness**
  - Is it a total QoS solution or just a component of QoS scheme?
- **Type of guarantees**
  - Are guarantees strict (always fulfilled) or dynamic (mostly fulfilled) ?
- **Performance type or level**
  - Performance level is constant (delay < 20ms) or dynamic (mostly delay < 20ms) ?

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

QoS Mechanisms Enforcement

- **Connections**
  - Is service packet-oriented (messages) or connection-oriented (virtual connection negotiation) ?
- **Degree of Support**
  - Does the scheme support local or End-to-End QoS ?
- **Flexibility and Adaptability**
  - Is the scheme applicable to systems of variable sizes (LAN, MAN, WAN, Internet)?

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

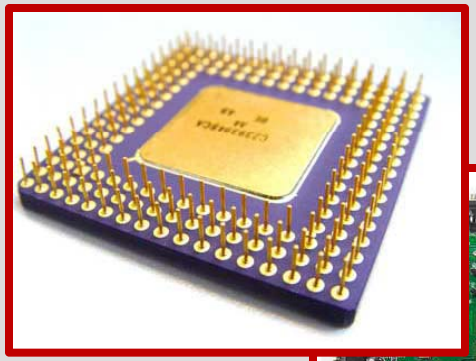
QoS Enforcement Levels

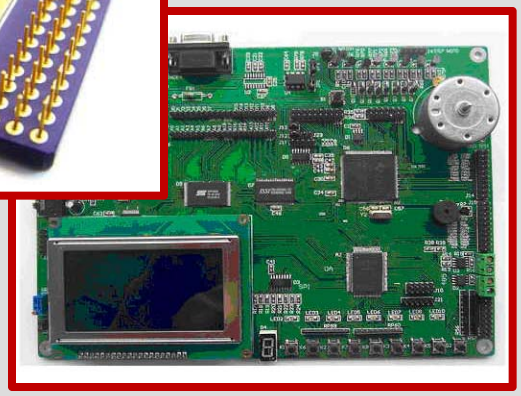
- **Hardware-Level QoS**
- **Connectivity-Level QoS**
- **Data-Level QoS**
- **Application-Level QoS**

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Hardware-level QoS

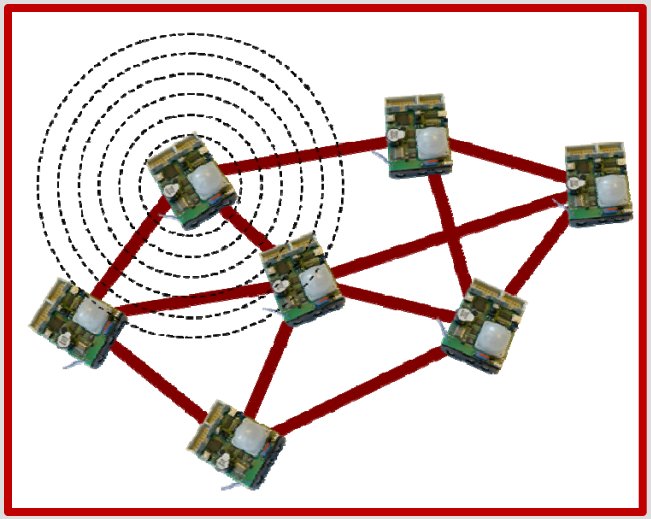
- **Utilization of hardware resources in compliance with resource availability and system status**
  - Battery lifetime
  - Processing power
  - Communication range
  - Data throughput





Hardware-level QoS

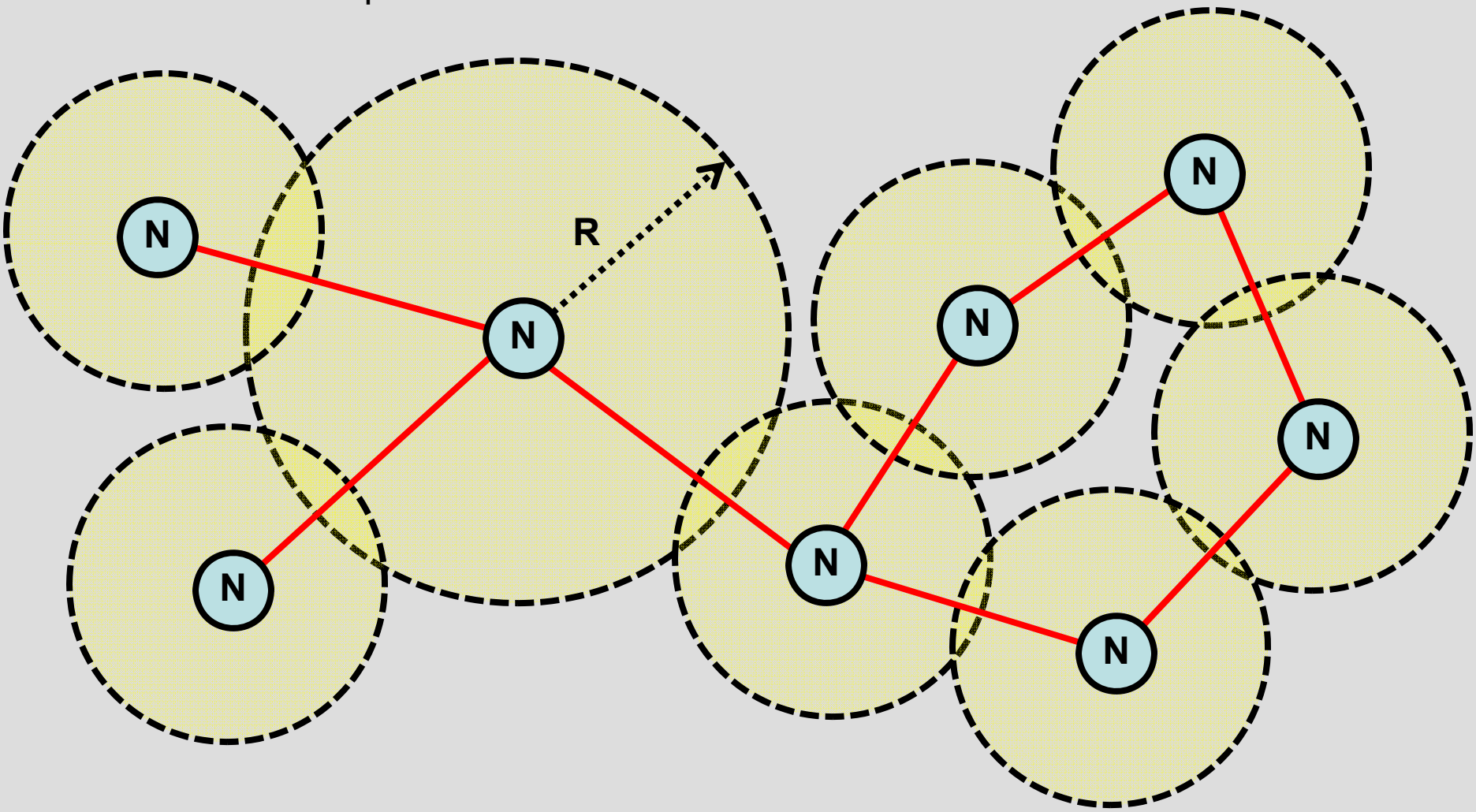
- **Optimization hardware resources utilization in MANET (Mobile Ad Hoc Networks) systems**
  - Collection of independent computing nodes
  - Communicating through wireless network
- **MANET QoS management factors**
  - Topology management
  - Packet Routing
  - Data acquisition and processing
  - Battery life



Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Hardware-level QoS

N – Wireless Network node
R – Communication perimeter



Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

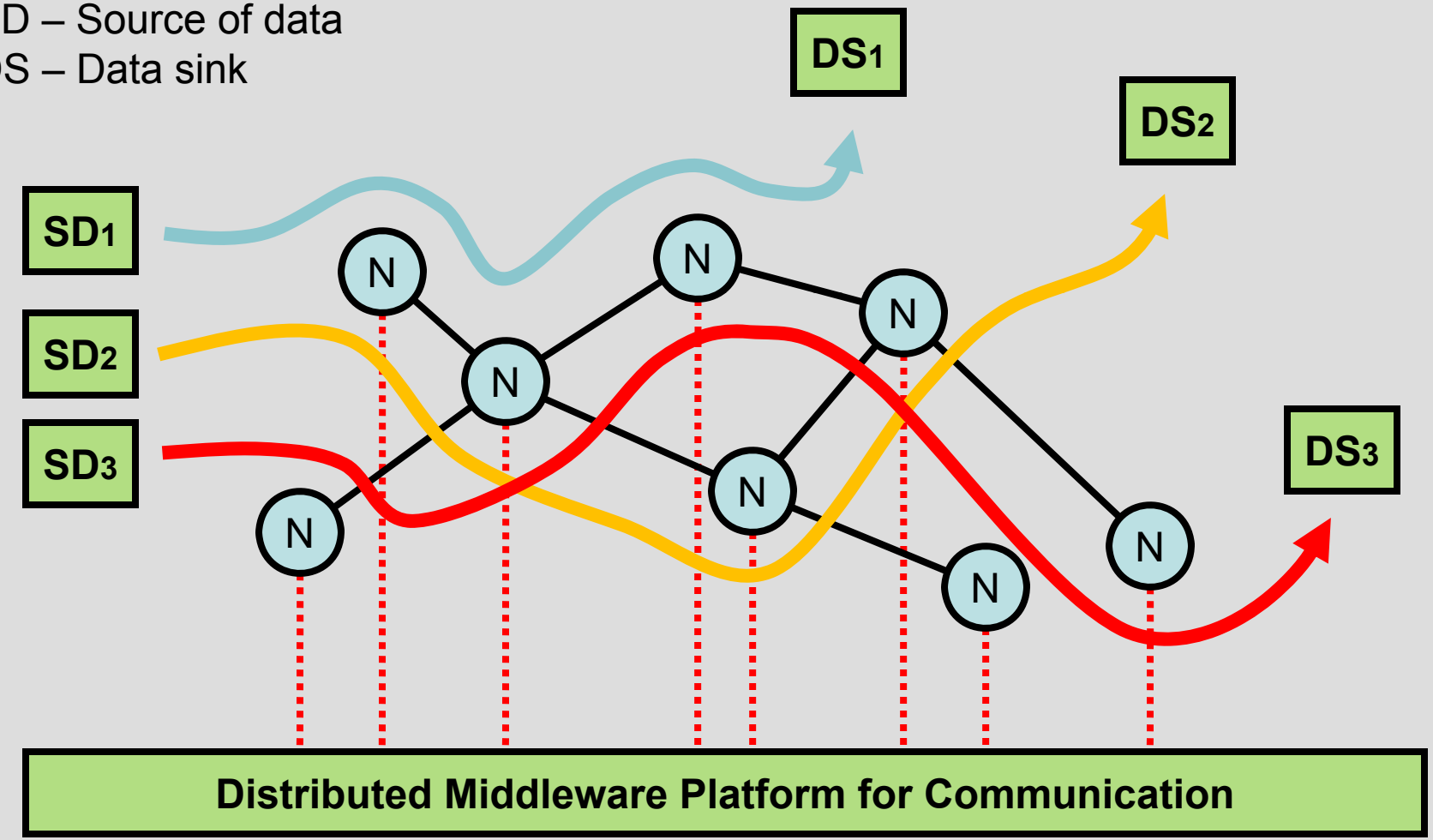
Connectivity-level QoS

- **Messages sent between sources and sinks in a network receive a differentiated level of service**
  - Network throughput
  - Network latency
- **Classes of network packet streams**
  - Best effort traffic (File transfer)
  - Normal traffic (Document fetch)
  - Real-time traffic (Audio and Video Multimedia)
  - Management traffic (Monitoring and Control)

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Connectivity-level QoS

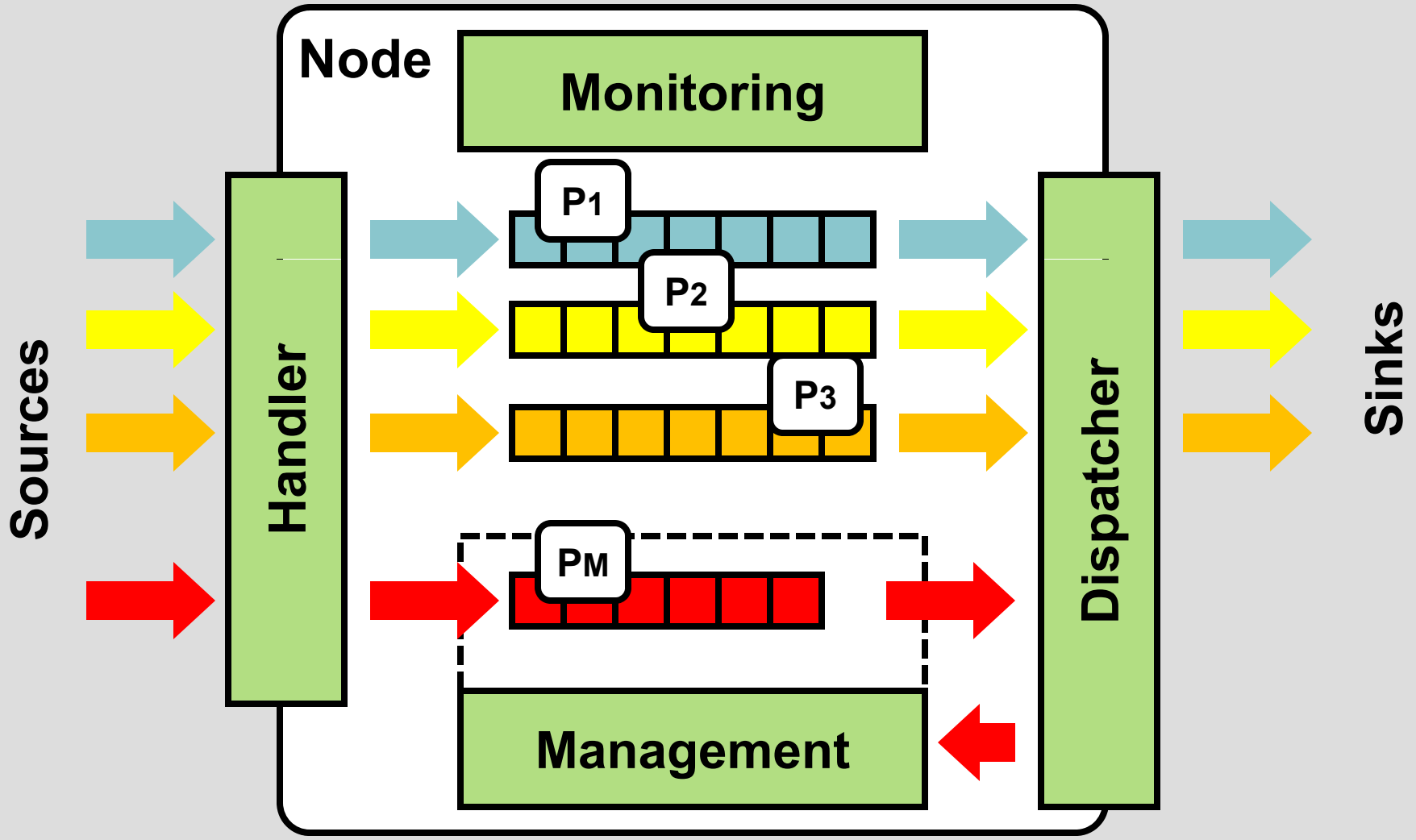
N – Interconnection node
SD – Source of data
DS – Data sink



Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Connectivity-level QoS

P1 – Best Effort, P2 – Normal Priority, P3 – Real-time Priority, PM – Management



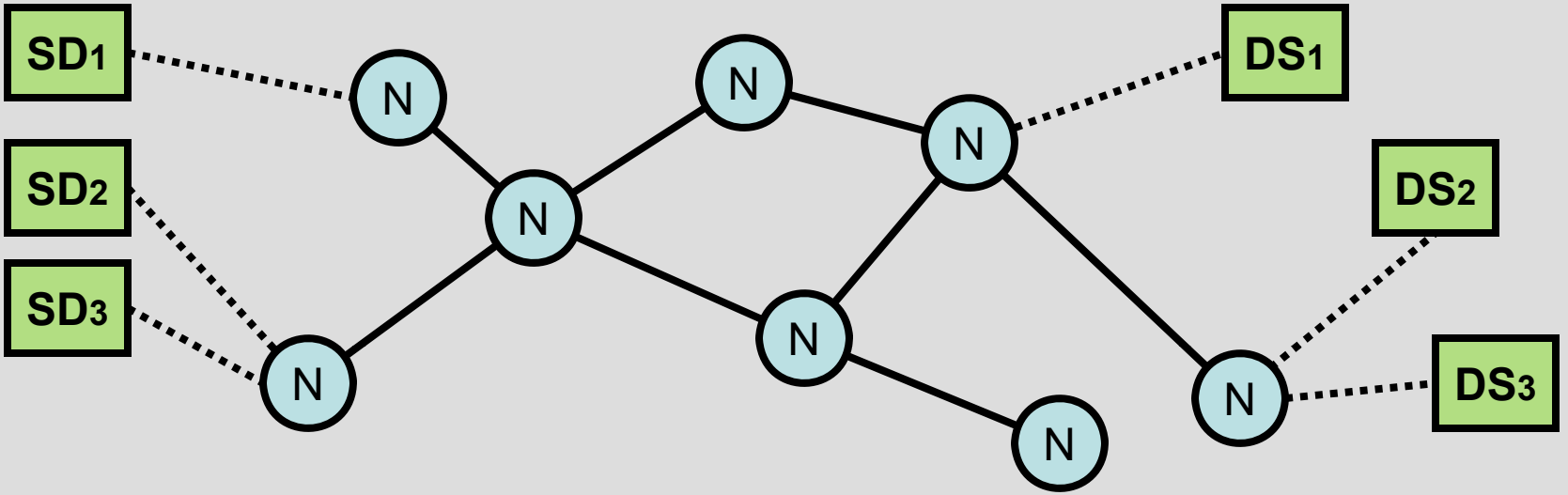
Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

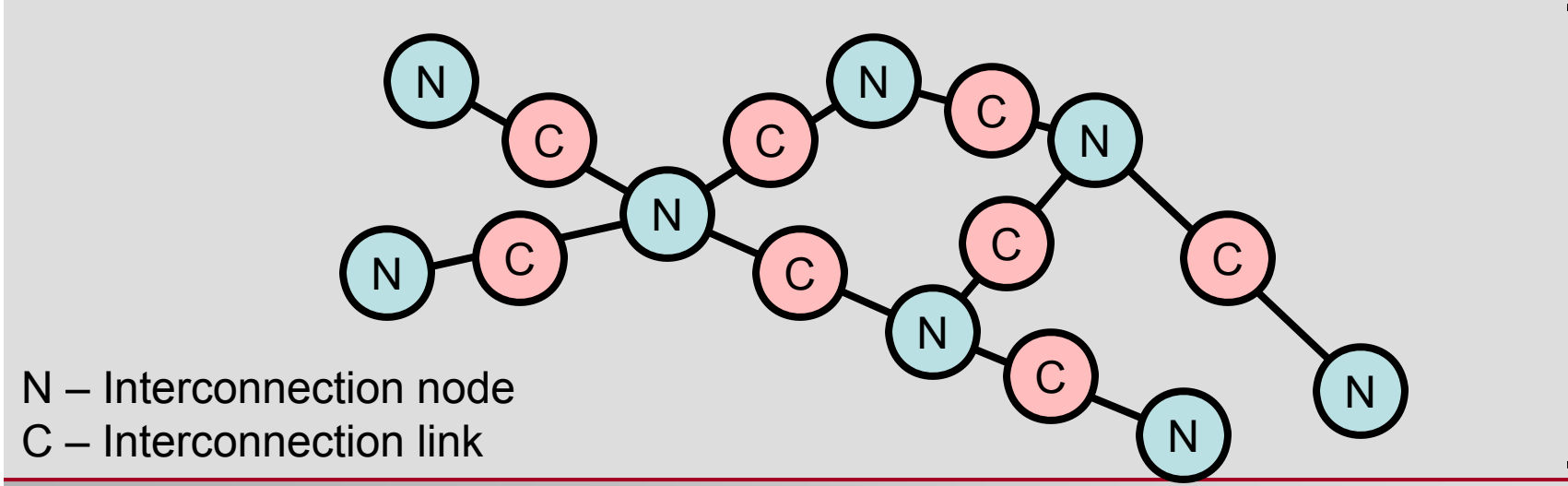
Computing Connectivity-Level QoS

- **Network is modeled as an extended graph structure**
  - Nodes represent network connections and interconnection nodes
  - Nodes are assigned delay and throughput characteristics
- **Packet routing decisions based on algorithms for path length optimization**
  - Solve network for data streams with respect to their allocated QoS characteristics

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Network Model with QoS Guarantees

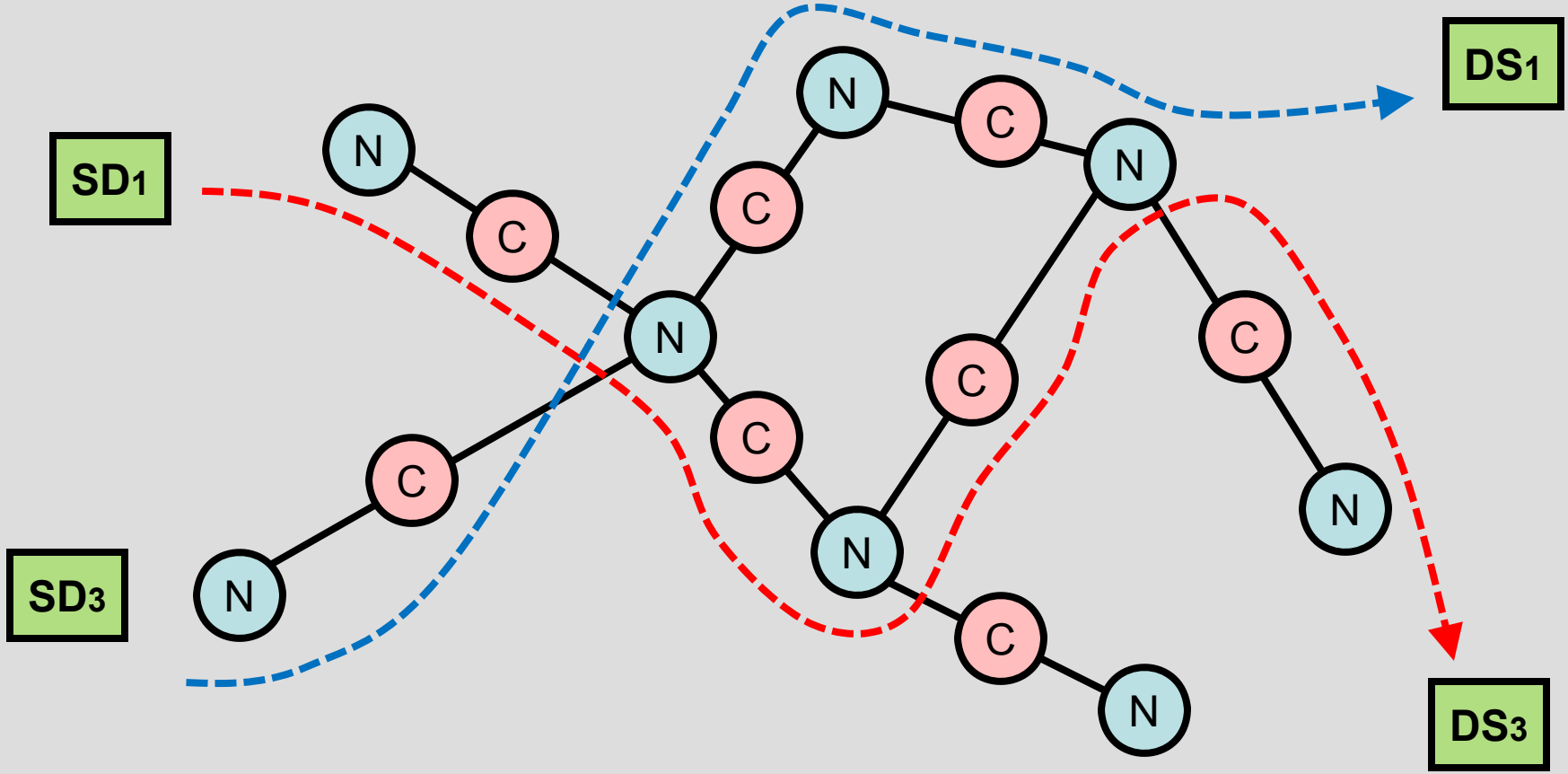




N – Interconnection node
C – Interconnection link

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Network Model with QoS Guarantees



N – Interconnection node
C – Interconnection link

Connectivity-level QoS

- **Communication end-to-end delay**

$$D_P = \sum_{c \in Pc} D_C + \sum_{n \in Pn} D_n(\theta)$$

  - D_C – transmission delay for connection node C
  - $D_N(\Theta)$ – transmission delay for interconnection node N as a function of existing load

- **QoS constraints for set of streams**

$$D_P \leq D_P^{Max}$$

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

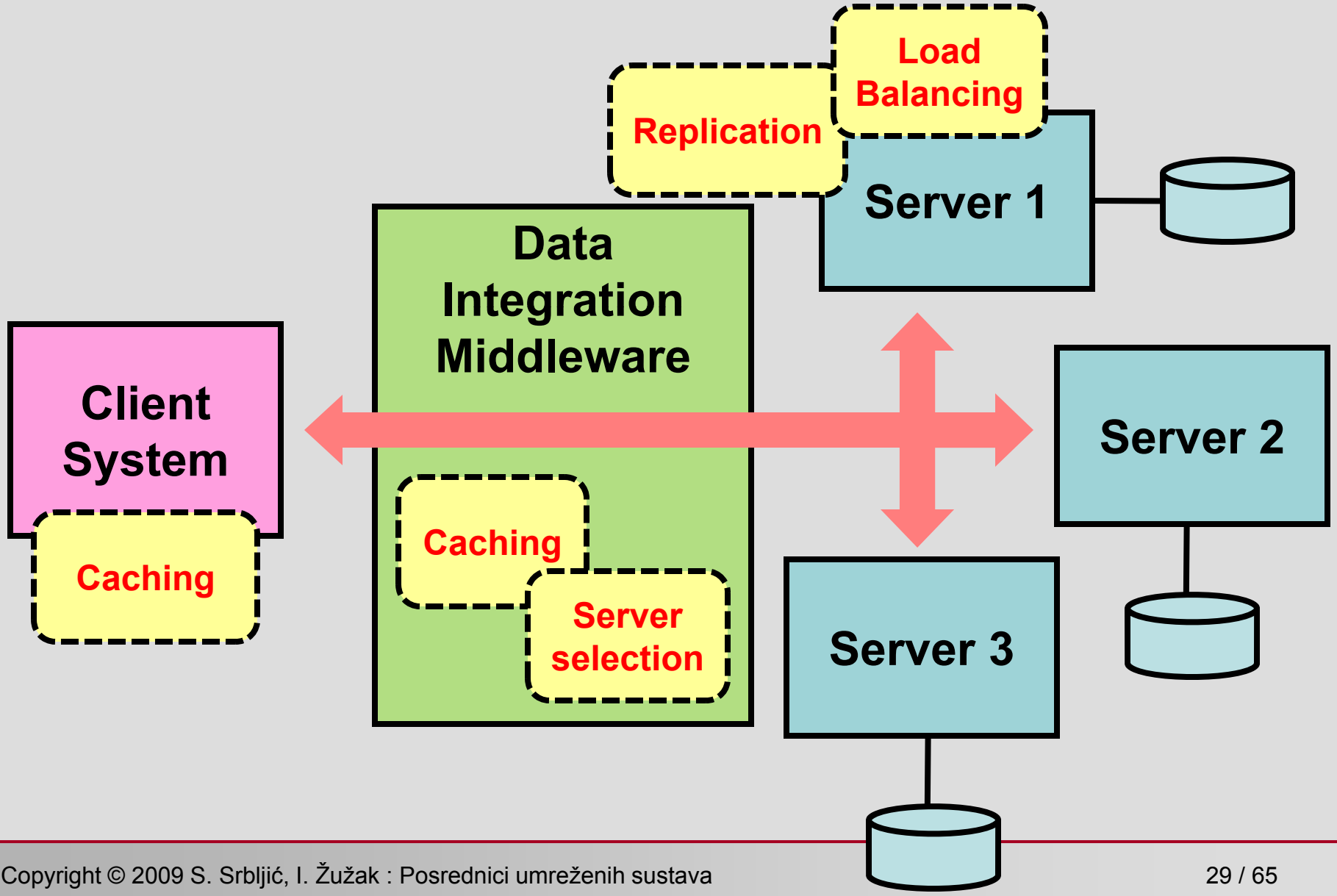
Data-level QoS

- **The performance perceived by the users of a web service depends on the efficiency of protocols that operate between web clients and servers**

- **QoS Issues**
  - Data Retrieval Latency
  - Data Availability

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Data-Level QoS Enforcement



Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

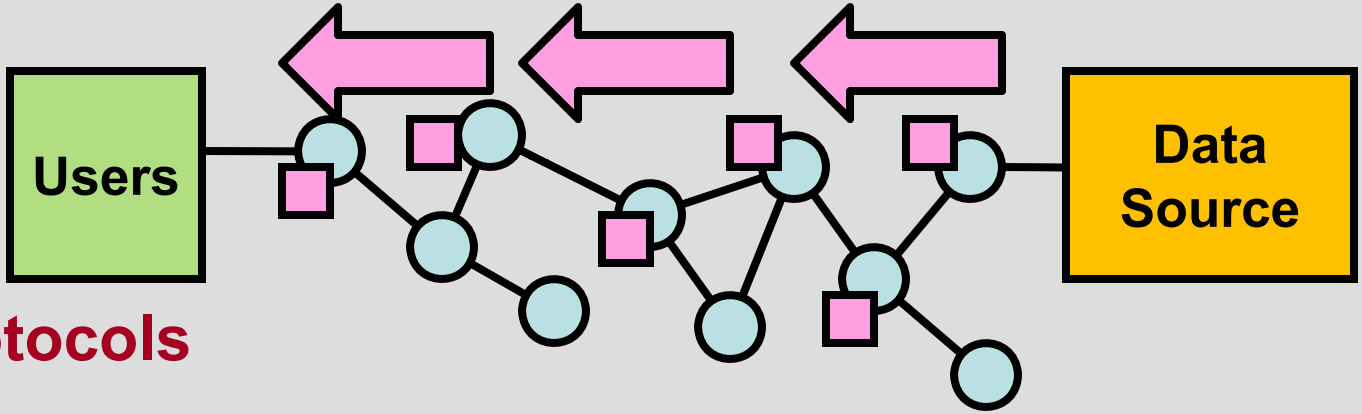
Data-level QoS Enforcement

- **Client-side**
  - Data Caching (Data pull, push, and hybrid protocols)
- **Server-side**
  - Data Replication (Consistency protocols)
  - Request Load Balancing (Request forwarding policies)
- **Integration Middleware**
  - Data Caching (Data pull, push, and hybrid protocols)
  - Server Selection (Request forwarding policies)

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Data Caching

- **Data is moved close to clients to reduce response time and save bandwidth**

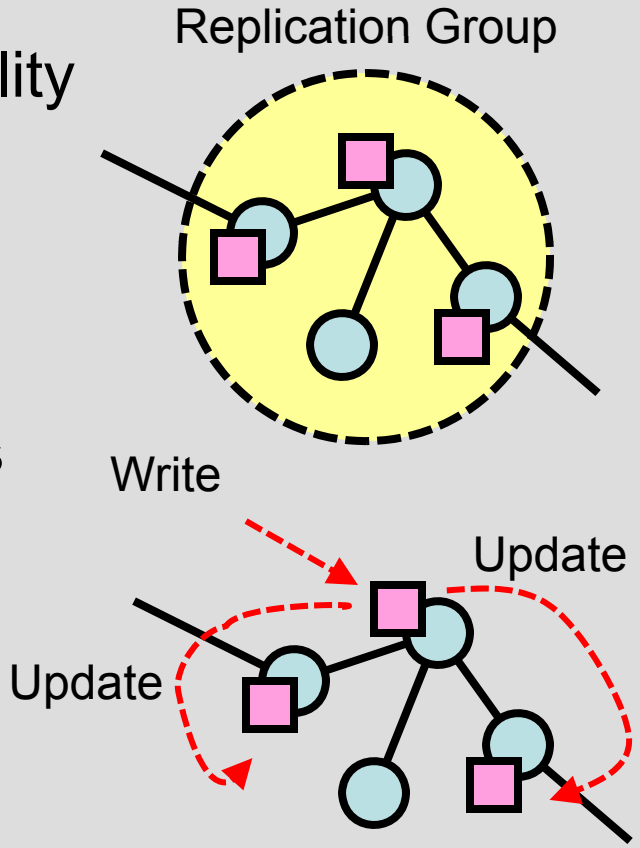


- **Data Pull Protocols**
  - Data is periodically refreshed from the source
- **Data Push Protocols**
  - Data is transferred on change from source to clients
- **Hybrid Protocols**

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Data Replication

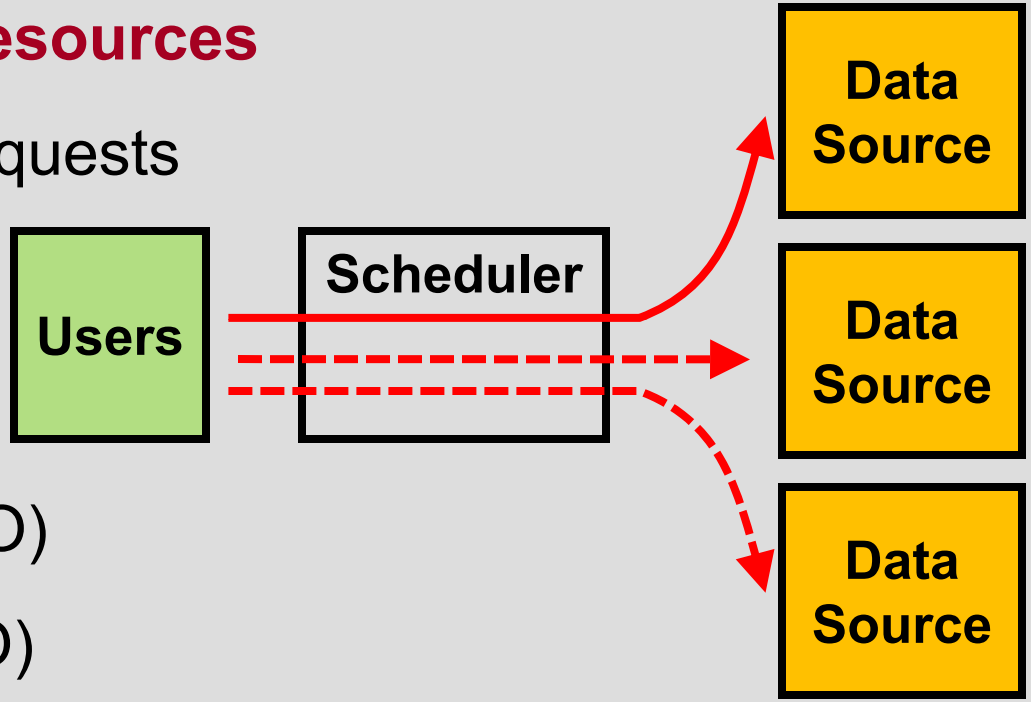
- **Data is kept in multiple copies in the system**
  - Enforce fault tolerance and availability
- **Consistency protocols**
  - Propagate changes among replicas
- **Load balancing mechanism**
  - Uniformly utilize data sources



Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Server Selection Strategies

- **Uniformly utilize data resources**
  - Distribute the client requests
- **Scheduling strategies**
  - First-In-First-Out (FIFO)
  - Last-In-Last-Out (LILO)
  - Least-Recently-Used (LRU)
  - Equal-Load (EL)



Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Application-level QoS

- **Management of applications**
  - Service execution duration (D)
  - Service cost (C)
  - Service reliability (R)
  - Service availability (A)
  - Service reputation (E)
- **Business activity monitoring systems**
  - Monitor QoS of application integration deployment
  - Facilitate change to sustain desired QoS

Application-level QoS

- **Deployment of QoS-aware application through service composition**
  - Select basic services
  - Interconnect services
  - Deploy and execute service composition
- **Deployed Application QoS**
  - End-product of basic services QoS characteristics

Application-level QoS

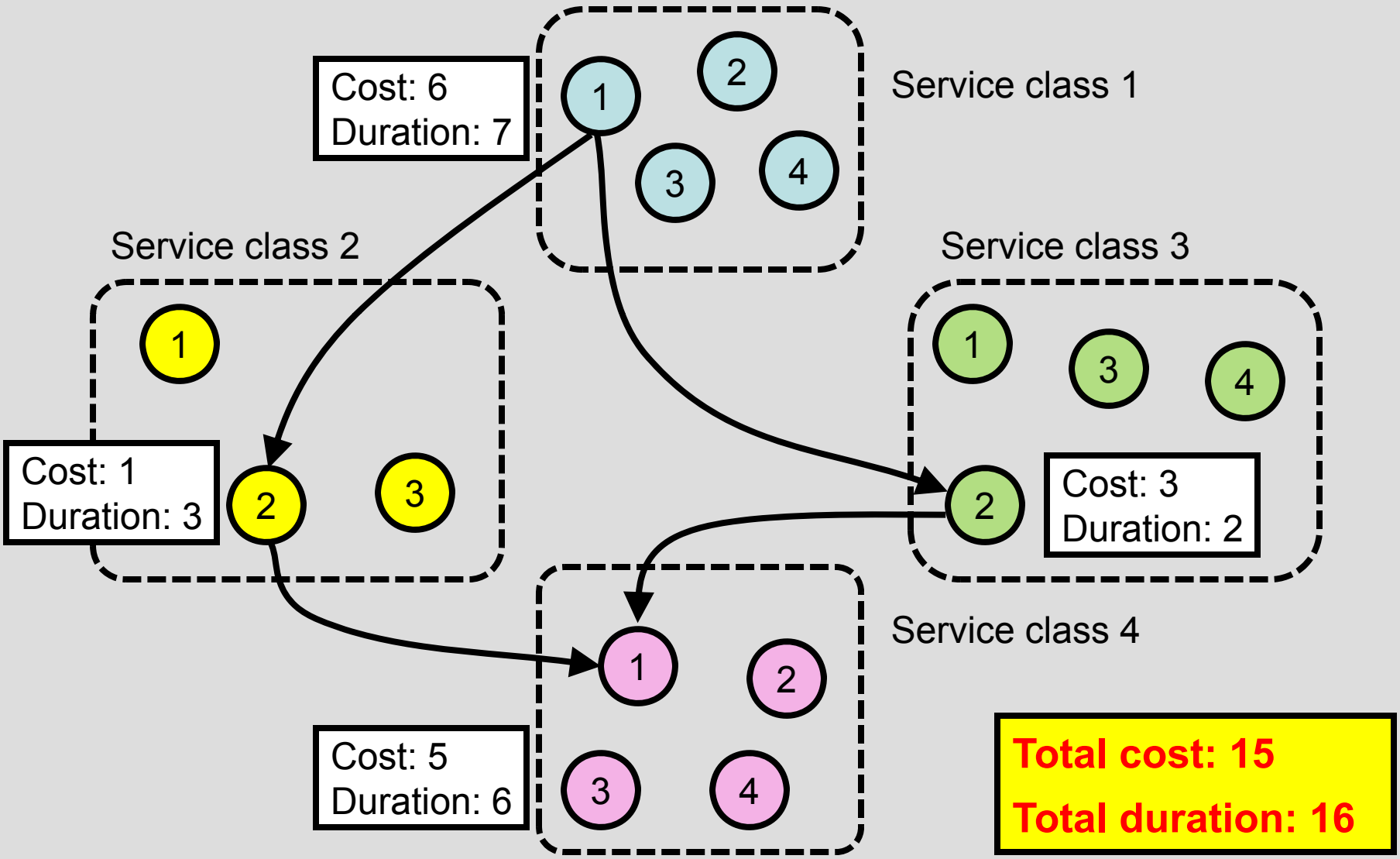
- **Goal functions for QoS-aware application composition**
  - Minimize response time (D)
  - Minimize total price (P)
  - Maximize reliability (R)
  - Maximize availability (A)
  - Maximize reputation (F)

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Example: QoS-Aware Composition

- **Application is built by composing services**
- **Basic services are grouped into service classes**
- **Service classes**
  - Services have the same functional properties (purpose, function, usage context and pattern)
  - Services have differentiated nonfunctional properties (cost, performance, availability, reliability, reputation)

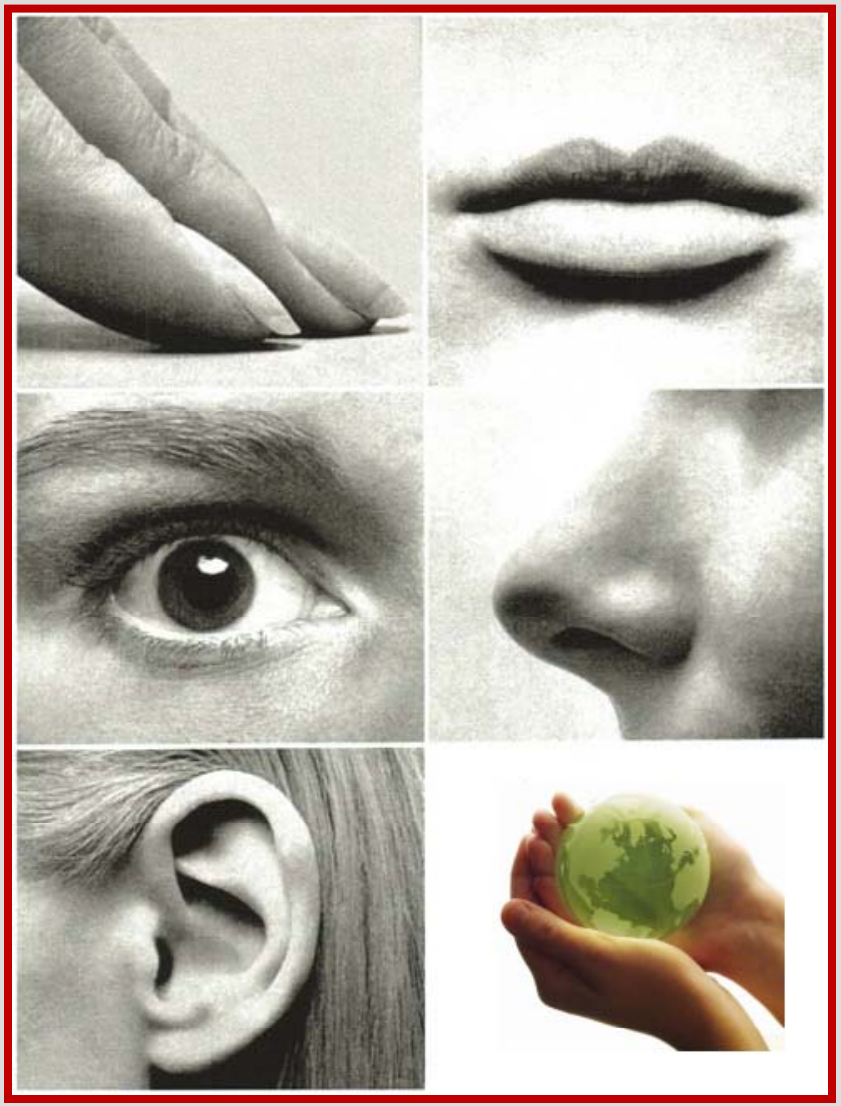
Example: QoS-Aware Composition



Quality of Experience

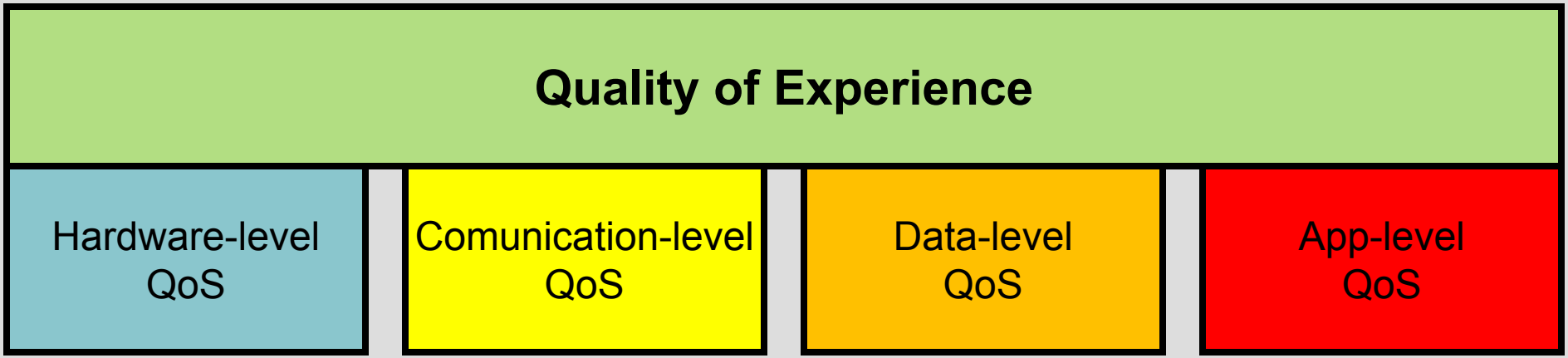
"Information is pretty thin stuff unless mixed with experience"

~Clarence Day, The Crow's Nest



Quality of Experience

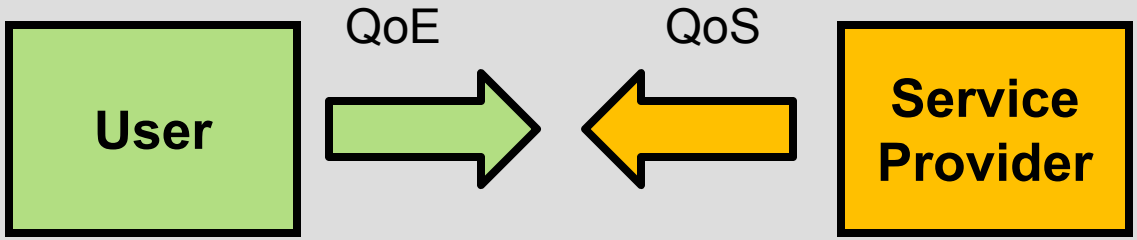
- **How does effective service design provides end-users with a successful and satisfying experience?**
- **How should be QoS features combined and in what ways in order to get a successful end-user service?**



Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

QoE and QoS

- QoS is an objective measure of service delivery characteristics and how well the provider fulfills the terms of service



- Provider may obey the terms of a contract's language, thus rating high in QoS, but, the users may be very unhappy, thus causing a low QoE

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Cab Ride QoS and QoE Example

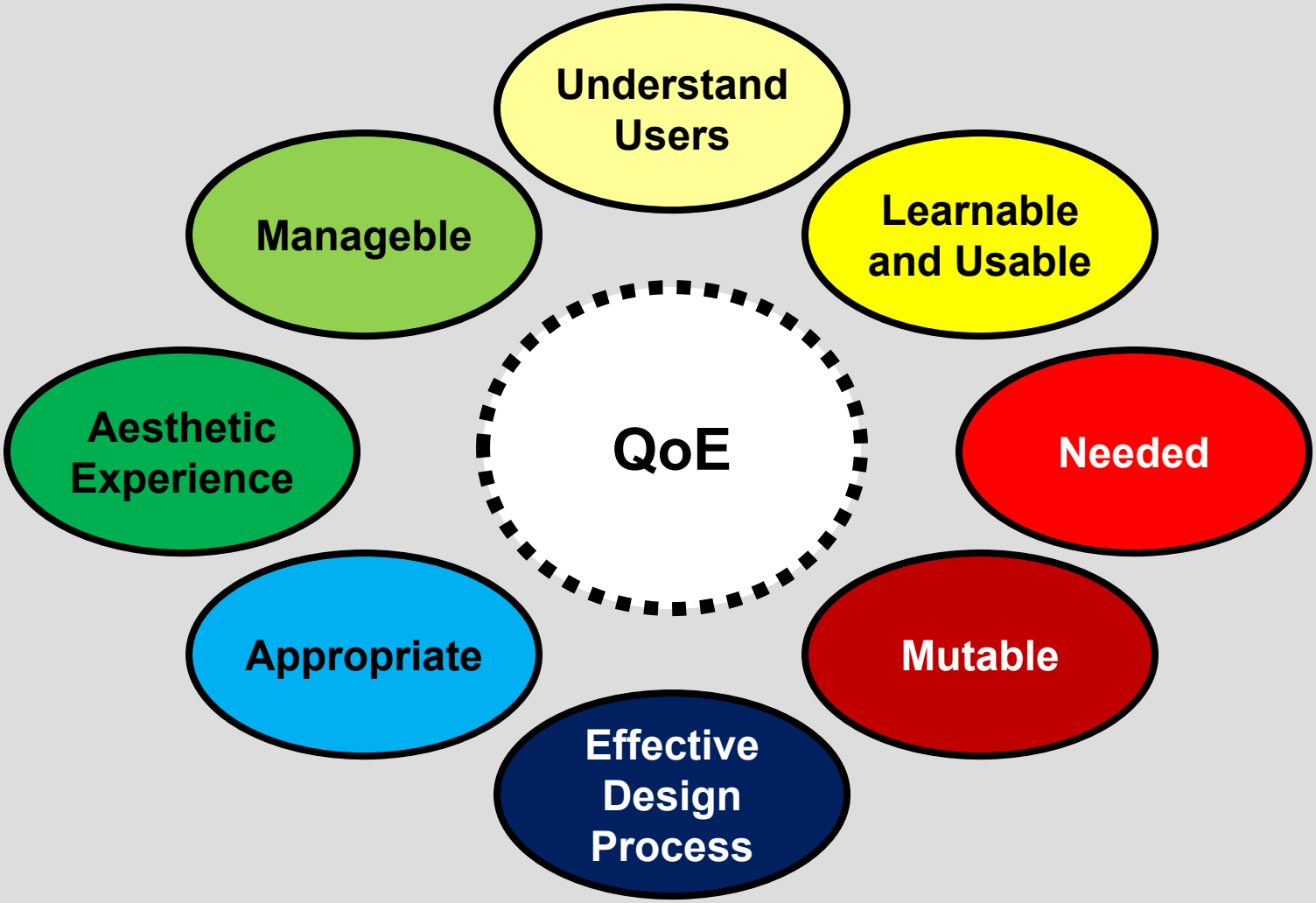
- **Safety is of the utmost customer expectation**
- **QoS Decision (High safety)**
  - The driver drives with an average speed of 10 mph
- **QoE Impact (Long ride)**
  - The ride takes 2 hours to travel 20 miles to the destination

QoS ⇧

⇩ **QoE**

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

QoS Engineering Factors



Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Quality of Experience Elements

- **Understand users**
  - How well do we understand the target users of service?
- **Effective design process**
  - Is the service a result of a well-known delivery process?
- **Needed**
  - What need does the service satisfy?
- **Learnable and usable**
  - Can the service be easily used?

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Quality of Experience Elements

- **Appropriate**
  - Does the service provide needed user experience?
- **Aesthetic experience**
  - Is the use of service pleasing and sensually satisfying?
- **Mutable**
  - Can the service be personalized to individual users?
- **Manageable**
  - Can the service be controlled effectively?

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

QoE Example: Media Streaming

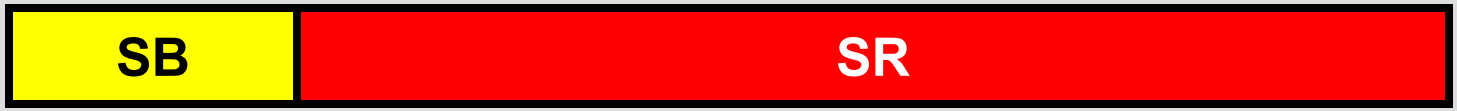
- **Streaming of multimedia video content on mobile devices**

- **Issues related to QoE**
  - Picture quality
  - Sound quality
  - Prefetching delay
  - Media continuity

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

QoE Example: Media Streaming

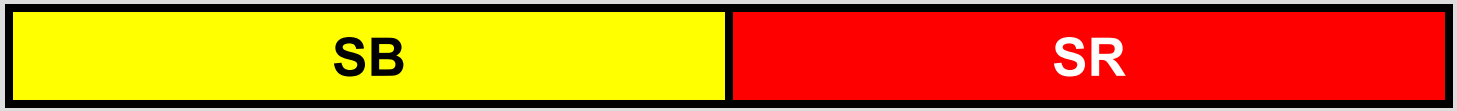
- **Device streaming capabilities**
  - Network throughput BT
- **Video clip**
  - Duration D, size S, play throughput BP = S [Mb] / D
- **Playback features**
  - Realtime download size SR = BT * D [s]
  - Buffered size SB = S – SR
  - Total time T = TB + D, Effectiveness E = D / T



Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

QoE Example: Media Streaming

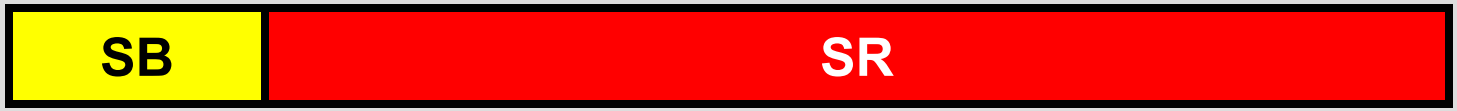
- **Device streaming capabilities**
  - BT = 1.8 Mbps (3G HSDPA)
- **Video clip**
  - D = 2 min, S = 50 MB (lossless cmpr), BP = 3.5 Mbps
- **Playback features**
  - SR = BT * D [s] = 25.75 MB, SB = 24.25 MB
  - Total time T = TB + D = SB/BT + D = 113 + 120 = **233 s**
  - Effectiveness E = D / T = **0.51 (51 %)**



Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

QoE Example: Media Streaming

- **Device streaming capabilities**
  - BT = 1.8 Mbps (3G HSDPA)
- **Video clip**
  - D = 2 min, S = 30 MB (cmpr w. loss), BP = 2.4 Mbps
- **Playback features**
  - SR = BT * D [s] = 25.75 MB, SB = 4.25 MB
  - Total time T = TB + D = SB/BT + D = 20 + 120 = **140 s**
  - Effectiveness E = D / T = **0.86 (86 %)**

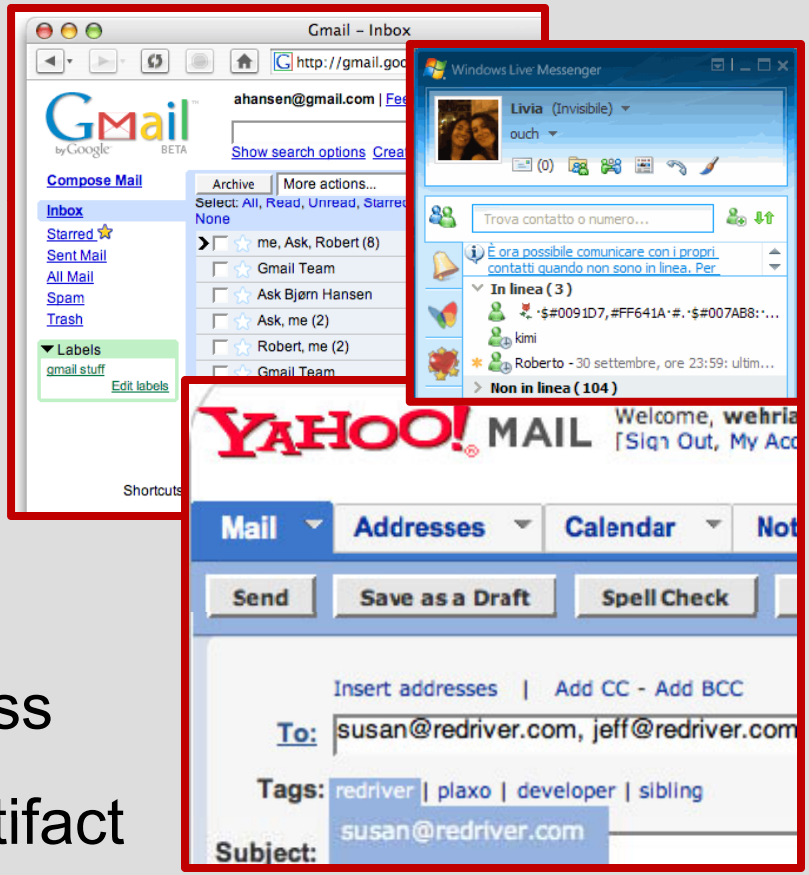


Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

QoE Example: UI Interaction

- **Design of UI elements for software and hardware artifacts**

- **Goals, Operators, Methods, and Selection rules (GOMS)**
  - A framework for quantitative evaluation of user experience during the UI interaction process with a software of hardware artifact



QoE Example: UI Interaction

- **Goals**
  - User purpose and intentions during interaction (send mail, download file)
- **Operations**
  - Actions performed to reach the desired goal (mouse clicks, typing in text, move object)
- **Methods**
  - Sequence of operations that users do to accomplish the goal (Select menu → Type name → Click search)

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

QoE Example: UI Interaction

- **Selection rules**
  - Criteria used by users to select the appropriate method for accomplishing the desired goal (minimal time, minimal effort, maximal precision, maximal control)

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

QoE Example: UI Interaction

- **System UI Evaluation 1 (eg. gMail)**
  - Goal type 1: 3 methods, 5 operations per method
    - (eg. Send mail)
  - Goal type 2: 2 methods, 6 operations per method
    - (eg. Search mail)
- **System UI Evaluation 2 (eg. Yahoo! Mail)**
  - Goal type 1: 4 methods, 12 operations per method
    - (eg. Send mail)
  - Goal type 2: 4 methods, 10 operations per method
    - (eg. Search mail)

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

QoE Example: UI Interaction

- **QoE weight factors**
  - w_g – goal weight, w_m – method weight, w_o – operation weight
- **QoE total measure**
  - Weighted sum over goals, methods, and operations

$$QoE_{System} = \sum_{g \in Goals} w_g \left(\sum_{m \in Methods_g} w_m \left(\sum_{o \in Operations_m} w_o \right) \right)$$

Service Level Agreements

"My idea of an agreeable person is a person who agrees with me"

~ Benjamin Disraeli

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Service Level Agreements

- **A formal contract which defines the terms under which service provider and consumer engage in interaction with the purpose of delivering and consuming a service**
- **Service Level Agreement regulates**
  - Priorities
  - Guarantees
  - Responsibilities
  - Warranties

Service Level Agreements

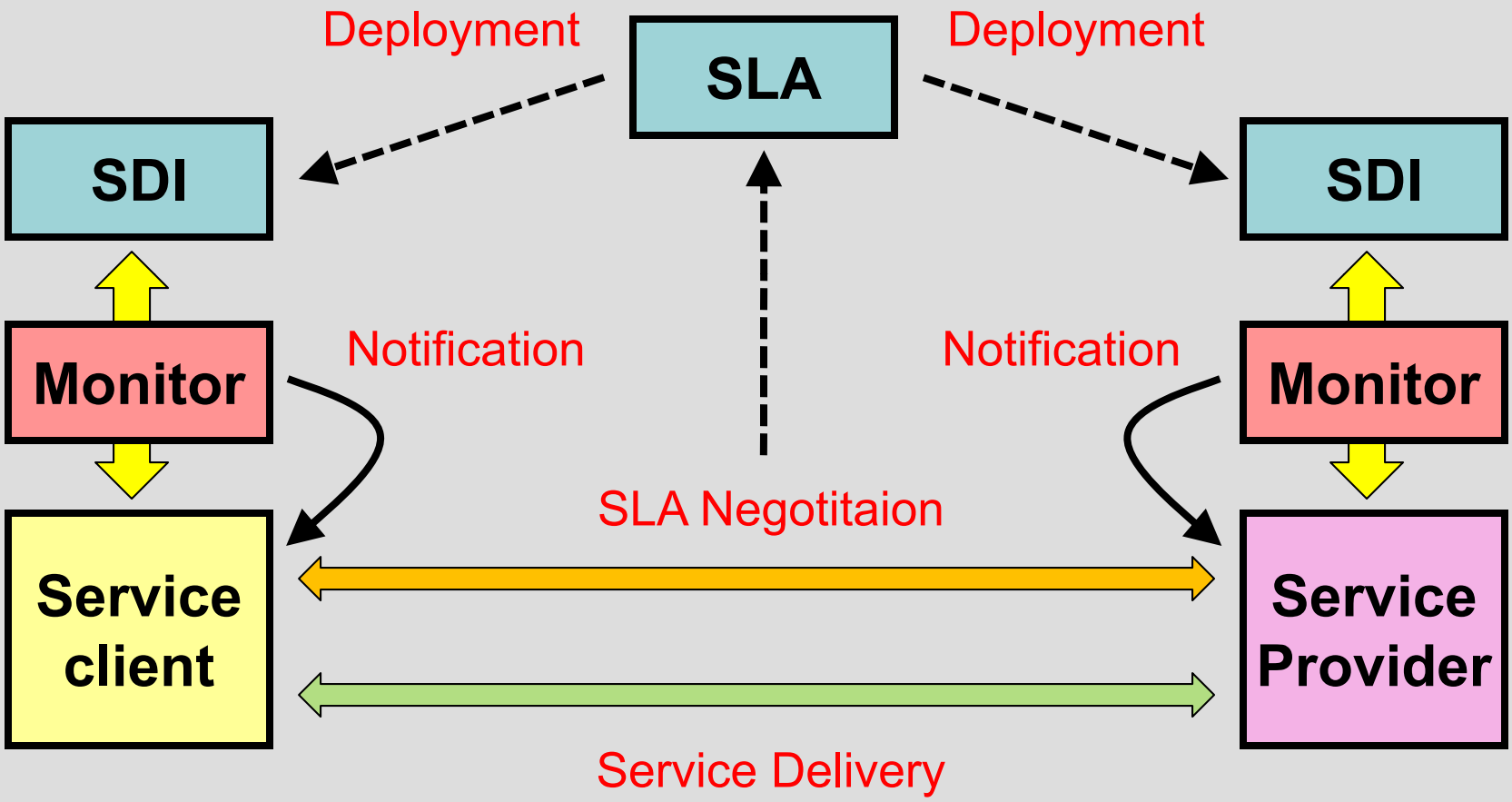
- **Service Level Agreement consists of a set of statements defining the level of service**
  - Minimum service level
  - Target service level
- **Service level metrics**
  - Availability (%, 24/7 - 1)
  - Performance (MIPS, Mbps)
  - Turn around time (s)
  - Data volume (MB)
  - Custom metrics (hits per request)

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Service Level Agreements

- **SLA Enforcement Standrads**
  - Web Service Level Agreement Language (WSLA)
    - *SLA Contract Document Specification*
  - WSLA Service Deployment Information (WSLA SDI)
    - *SLA Enforcement Policies*
  - WS-MetadataExchange
    - *SLA negotiation protocol*

SLA Architecture



SLA Example – Parties

```
<Parties>
  <ServiceProvider name="provider">
    <Contact>
      <POBox>P.O.Box 218</POBox>
      <City>Yorktown, NY 10598, USA</City>
    </Contact>
    ....
  </ServiceProvider>
  <ServiceCustomer name="customer">
    <Contact>
      <Street>30 Saw Mill River RD</Street>
      <City>Hawthorne, NY 10532, USA</City>
    </Contact>
    ....
  </ServiceCustomer>
  <SupportingParty name="ms" role="MeasurementServiceProvider">
    <Contact>
      <Street>Saeumerstrasse 4</Street>
      <City>CH-8803 Ruschlikon, Switzerland</City>
    </Contact>
  </SupportingParty>
</Parties>
```

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

SLA Example – Operation QoS

```
<ServiceDefinition name="StockQuoteService">
  <Operation>
    <WSDLFile>http://www.uddi.com/StockQuoteService.wsdl</WSDLFile>
    <SOAPBindingName>SOAPNotificationBinding</SOAPBindingName>
    <SOAPOperationName>getQuote</SOAPOperationName>

    <SLAParameter name="AverageResponseTime"
                  type="float"
                  unit="seconds">
      <Metric>averageResponseTime</Metric>
    </SLAParameter>

    <Schedule name="MainSchedule">
      <Period>
        <Start>2001-11-30T14:00:00.000-05:00</Start>
        <End>2001-12-31T14:00:00.000-05:00</End>
      </Period>
      <Interval>
        <Minutes>2</Minutes>
        <Seconds>30</Seconds>
      </Interval>
    </Schedule>
```

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

SLA Example

```
...
<Obligations>
  <ServiceLevelObjective name="g1" serviceObject="WSDLSOAPGetQuote">
    <Obliged>provider</Obliged>
    <Validity>
      <StartDate>2001-08-15:1400</StartDate>
      <EndDate>2001-09-15:1400</EndDate>
    </Validity>
    <Expression>
      <Predicate xsi:type="wsla:Less">
        <SLAParameter>AverageResponseTime</SLAParameter>
        <Value>5</Value>
      </Predicate>
    </Expression>
  </ServiceLevelObjective>
...
```

SLA Example

```
...
  <ActionGuarantee name="g2">
    <Obliged>ms</Obliged>
    <Expression>
      <Predicate xsi:type="wsla:Violation">
        <ServiceLevelObjective>g1</ServiceLevelObjective>
      </Predicate>
    </Expression>
    <QualifiedAction>
      <Party>customer</Party>
      <Action actionName="notification" xsi:type="Notification">
        <NotificationType>Violation</NotificationType>
        <CausingGuarantee>g1</CausingGuarantee>
        <SLAParameter>AverageResponseTime</SLAParameter>
      </Action>
    </QualifiedAction>
    ....
</ActionGuarantee>
...
```

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

Conclusion

- **Quality of Service**
  - Delivery of service with not only functional characteristics but also with the desired nonfunctional properties
- **Quality of Experience**
  - Service quality as perceived by the service user
- **Service Level Agreement**
  - Contract specifying the required properties of service exchanged between clients and providers

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava

References

- H.A. Duran-Limon, G.S. Blair, G. Coulson, **"Adaptive Resource Management in Middleware: A Survey"**, IEEE Dist. Sys. Online, 2004.
- T.F. Abdelzaher, K.G. Shin, N. Bhatti, **"User Level QoS-Adaptive Resource Management in Server End-Systems"**, IEEE Trans. on Comp., 2003.
- D.E. Taylor, **"Survey and Taxonomy of Packet Classification Techniques"**, ACM Comp. Surv., 2005.
- L. Zeng, B. Benatallah, A.H.H. Ngu, M. Dumas, J. Kalagnanam, H. Chang, **"QoS-Aware Middleware for Web Services Composition"**, IEEE Trans. on SW Eng., 2004.
- A.W. Bragg, **"Quality of Service: Old Idea, New Options"**, IT Pro, 1999.
- M.P. Papazoglou, W.-J. van den Heuvel, **"Web Services Management: A Survey"**, 2005.
- M. Conti, M. Kumar, S.K. Das, B.A. Shirazi, **"Quality of Service Issues in Internet Web Services"**, IEEE Trans. on Comp., 2002.
- S.-N. Chuang, A.T.S. Chan, **"Dynamic QoS Adaptation for Mobile Middleware"**, IEEE Trans. on SW Eng., 2008.

Copyright © 2009 S. Srbljić, I. Žužak : Posrednici umreženih sustava