Dynamic SLA, QoS and invoicing for ROIA in edutain@grid

ROIA workshop tutorial, GECON 2009
Delft, 2009
Stuart E. Middleton
IT Innovation Centre, University of Southampton
Tutorial Overview

Key decisions for ROIA provisioning

Business models
  edutain@grid value chain

Business architecture in edutain@grid
  service architecture, two example edutain@grid ROIA

Quality of service measurement
  SLA metrics, SLA representation

SLA management
  SLA negotiation
  edutain@grid ROIA runtime business context

Load management
  Load management strategies, edutain@grid zone management

Invoice models
  edutain@grid invoicing

Conclusions, Questions

© University of Southampton, IT Innovation Centre 2009
24/08/2009, edutain@grid, ROIA workshop tutorial, GECON 2009, Delft
Six key decisions to get the right business context

- Which value chain?
- What SLA metrics should you measure?
- How will your SLA be represented?
- What SLA negotiation strategy?
- What load management strategy?
- What invoice model?
Key decisions for ROIA provisioning

Six key decisions to get the right business context

- Which value chain?
- What SLA metrics should you measure?
- How will your SLA be represented?
- What SLA negotiation strategy?
- What load management strategy?
- What invoice model?
Business models

Identify the value chains for your business model

- Identify the key actors in the value chain
- Customers or consumers? Brokers or distributors?
- Single or multi-hoster ROIA provision?
- Competition or cooperation between hosters?

Reward behaviour that promotes desired objectives

- Payment for service provider hardware, QoS, QoE
- Penalties for loss of services
- Shared costs / rewards
AMIS commercial model for ROIA service provision

- Single hoster (e.g. AMIS)
- Written fixed duration SLA, payment for hardware
- Penalties if hardware QoS targets missed (e.g. network hardware bandwidth, server availability)
Enhanced business model

**edutain@grid value chain**

- **Actors**: customer(s), coordinator(s), hoster(s)
- **QoS**: based on service provision metrics not just hardware
- **Coordinator ‘broker’**: allows flexible provisioning models
- **On demand**: electronic SLA’s based on a template

**Customer(s)**
- **Pay per play**
- e.g. Game player

**Coordinator(s)**
- Pay per time on a
  - dedicated server
  - shared server
- e.g. Game provider
- Penalties if QoS breached
- SLA contract(s)
- e.g. Service provider

**Hoster(s)**
- Pay for time on a
- e.g. Service provider
Scalable SLA architecture

Customer - Coordinator
- User accounts, out of band payments (e.g. PayPal)

Coordinator - Hoster
- Trade account at each hoster
- Electronic SLA, QoS metrics, invoicing

Bipartite SLA agreements provide scalability
- ROIA start small, need low cost small scale hosting
- As ROIA attracts players, more hosters are needed
- Value networks that scale gradually lowers the entry cost
- Attracting smaller ROIA could expand the overall market
Scalable value networks

Customers  ROIA  Coordinators  Bipartite SLA's  Hosters
Possible new business models

Bi-directional value relationships

Customers want a good quality of experience (QoE)

Good QoE = regular players + new players = more revenue

Coordinators could share customer revenue with hosters

Hoster payment could be based on customer revenue over a time period
SLA’s could define QoE metrics in addition to QoS metrics
e.g. number of new players, number of repeat plays, player feedback

Hosters and coordinators would then have an incentive to work together on cross-hoster load management

These concepts are explore in the ROIA workshop paper

Stuart E. Middleton, Mike Surridge, Bassem I. Nasser, Kevin Yang

Bipartite dynamic SLA as a business framework to support cross-hoster load management of real-time online applications, GECON ROIA workshop 2009
Service oriented architecture for ROIA

- Real-time & management layers provides ROIA support
  - Connections (UDP, TCP, DTLS), zone replication, zone migration
  - QoS measurement, load prediction

- Business layer services for Coordinators
  - User account service (customer authentication)
  - Global session service (session management, customer access control)
  - QoS monitor service (cross-hoster load management)

- Business layer services for Hosters
  - Trade account service (hoster invoices to coordinator)
  - SLA service (SLA templates and instances)
  - Local session service (ROIA sessions)

- Based on the open source GRIA middleware [www.gria.org](http://www.gria.org)
Business architecture in edutain@grid

Business workflow: Customer (e.g. game player)

- Customer(s)
- Coordinator(s)
- Hoster(s)

**Register**

**Login**

- **Trade account service**
- **SLA service**
- **Local session service**
- **ROIA (server)**

**X.509 certificate SAML token**

- **Setup secure connection to session**

  - **User account service**
  - **Management client**
  - **QoS monitor service**
  - **Global session service**

**Run ROIA via runtime connection**

© University of Southampton, IT Innovation Centre 2009
24/08/2009, edutain@grid, ROIA workshop tutorial, GECON 2009, Delft
Business workflow: Coordinator (e.g. game provider)

- **Customer(s)**
  - Setup account
  - Check invoice
  - Negotiate SLA
  - Monitor QoS
  - Make new session
  - Zone migration & replication (can be cross-hoster)

- **Coordinator(s)**
  - User account service
  - Management client
  - QoS monitor service
  - Global session service
  - ROIA (client)
  - ROIA finishes

- **Hoster(s)**
  - Trade account service
  - Update invoice
  - SLA service
  - QoS measurements
  - Local session service
  - ROIA (server)
edutain@grid ROIA : Hunter online game
- Real-time 3D first person perspective shoot-up
- Game world scales to number of players
- Single signon, multi-zone, UDP
- Key QoS metric is packet latency & frame rate

edutain@grid ROIA : eLearning application
- eLearning shell running training simulations
- Coastguard Search and Rescue app
- Single signon, VOIP, DTLS
- Key QoS metric is data throughput / bandwidth
Key decisions for ROIA provisioning

Six key decisions to get the right business context

- Which value chain?
- What SLA metrics should you measure?
- How will your SLA be represented?
- What SLA negotiation strategy?
- What load management strategy?
- What invoice model?
Quality of service measurement

Which QoS metrics should you measure?

Do you trust client-side metrics?
Client machines are not under coordinator or hoster control – can you trust them?
Network speed, reliability, client frame rate
Customer satisfaction, feedback, complaints

Which QoS metrics are going to ensure customer QoE?
Hardware specification (CPU, memory, network)
Number of client connections, availability, reliability
Data throughput, packet loss
Server frame rate, zone migration/replication delay

How can measurements be taken?
Live measurements can be from ROIA (client, server), middleware, third party instruments
Offline measurements include user feedback, statistics etc.

Think about the value chain and the customer experience
edutain@grid metrics are measured server side

- Real-time layer sample period is 10 seconds
- QoS recorded per session
- QoS metrics measured by the real-time layer
  - Client connection count [count]
  - Data throughput [bytes/second]
  - Server tick time [milliseconds] which is the time to process a server frame
  - Average packet latency [milliseconds]
  - Packet loss [percentage]
QoS measurement workflow

- Trade account service
- SLA service
  - QoS values
  - Local session service
  - Metrics to measure
    - Smootherd QoS sample values
    - Raw QoS values
  - Management layer
  - Real-time layer
    - QoS report
    - Agreed SLA
    - Invoice
  - Coordinator
    - Bill
    - RTF logging
  - Hosted
- Coordinator
- Invoice
- RTF logging
- ROIA
- QoS report
- Coordinator
- Invoice
- RTF logging
- ROIA
- QoS report
Key decisions for ROIA provisioning

Six key decisions to get the right business context

- Which value chain?
- What SLA metrics should you measure?
- How will your SLA be represented?
- What SLA negotiation strategy?
- What load management strategy?
- What invoice model?
SLA representation

Which SLA representation to use?

- Need a machine readable SLA template
  - Date of agreement, Duration of SLA
  - Metric: Metric name, unit, upper/lower threshold values
  - Measurement type: Scalars, peak values, averages, rate of change
  - Pricing: Fixed costs, variable costs, penalties
  - Static details: hardware expected, other agreements

- Standards for SLA’s
  - WSLA, SLAng - failed to gain traction in community
  - WS Agreement - popular, focus on protocol, lacks QoS metrics, constraints, penalties
  - OWL-S, WSMO - web service ontology standards (metadata etc), not SLA specific
  - XML - maximum flexibility, non-standard

- edutain@grid uses XML
  - Allows innovation without restriction
  - Can be relatively easily transferred into a standards-compliant schema if required
**edutain@grid SLA example**

### Billing period – every 90 days
- **SLA duration**: start, end
  - (can be incremental or at end of SLA)

### Pricing term – ‘Maximum’
- **Metric**: RTF server data throughput
- **Unit**: bytes / second
- **Cost**: price 0.1 € * QoS peak value
- **QoS summed across local sessions**

### Pricing term – ‘Penalty’
- **Metric**: Average client packet latency
- **Unit**: milliseconds
- **Upper threshold**: 0.01 ms
- **Breach cost**: price -2 €
- **QoS per local session**

### Pricing term – ‘Duration’
- **Metric**: SLA duration
- **Cost**: 1 € every 1 day

---

Prices defined in Euro's
Key decisions for ROIA provisioning

Six key decisions to get the right business context

- Which value chain?
- What SLA metrics should you measure?
- How will your SLA be represented?
- What SLA negotiation strategy?
- What load management strategy?
- What invoice model?
Which SLA negotiation strategy is best?

+ Decide many parties
  Generally the actors in value chain are negotiating parties

+ Decide how many iterations
  Offer, counter offer etc
  Depends on the negotiation / action strategy adopted (e.g. English auction)

+ Where is there a human in the decision loop?
  Is financial SLA cost low enough to delegate final agreement to a CPU?
  Humans are slow, do performance time constraints allow for human decisions?
  A human budget holder will normally set some sort of boundary cost constraints
  A control UI, notification and/or confirmation callbacks are needed if semi-automatic

+ Standard management protocols?
  SLA negotiation protocols (e.g. WS Agreement, Auction protocols)
  Game theory / Agent communities have well defined protocols for machine negotiation

+ Long negotiations get good deals but cost I/O, CPU & time
edutain@grid SLA management

- Actors: coordinator, hoster
- Discrete offer strategy: simple, efficient
  - Hoster agrees a SLA template with coordinator offline
  - Coordinator creates template instances on-demand
- Human SLA proposal step
- ROIA run in local sessions under a SLA business context

- Setup a new SLA template
- Setting up sessions and zones
- Setup a user account
Key decisions for ROIA provisioning

Six key decisions to get the right business context

- Which value chain?
- What SLA metrics should you measure?
- How will your SLA be represented?
- What SLA negotiation strategy?
- What load management strategy?
- What invoice model?
Cross-hoster load measurement

How will load be managed between hosters?

+ Review single-hoster resource level scheduling strategy
  Each hoster will manage its own local resources (servers, network etc)
  Do your hosters support load prediction, load scheduling, reservations?

+ Cross-hoster management - manual or automated?
  The coordinator controls hoster load when creating new sessions
  Can hoster load be monitored? Does hoster report if new load would compromise QoS?
  Coordinator UI needed for manual overview and control
  Hoster feedback & predictions needed for automation of cross-hoster load management

+ Will there be penalties for QoS failure?
  Penalties may make hosters more conservative
  Hoster safety margins will maintain QoS but might limit hoster availability

+ Decide on the load management strategy
edutain@grid cross-hoster load management

Single-hoster load management
Hosters model load and generate 2 minute predictions (neural network)
Inter-server scheduling to micro-manage single-hoster load
Hoster load predictions made available to coordinator

Coordinator monitors QoS levels cross-hoster
UI to display graphs of measured QoS over time
SLA, session and zone overview for all hosters

Zone management allows load to be moved cross-hoster
Zones can be moved between sessions (i.e. between hosters)
Coordinator UI controls this process

Automated zone management
2 minute load prediction drives decision to move zones away from a loaded hoster
Simple move to least-loaded hoster strategy (proof-of-concept)
Cross-hoster load mgt workflow

New hosters can be added also, just need to agree an SLA with them

Local session service [hoster 3,4,5]

Zone migration / replication

Local session service [hoster 2]

> Pull point <<

QoS values

SLA metrics

QoS measurements from management layer

Coordinator

Hoster(s)

Monitor QoS, Zone management

Global session service

SLA service [hoster 1]

Live QoS measurement reports (inc predictions)

Decision: move zone

QoS monitor service

New hosters can be added also, just need to agree an SLA with them.
Six key decisions to get the right business context

- Which value chain?
- What SLA metrics should you measure?
- How will your SLA be represented?
- What SLA negotiation strategy?
- What load management strategy?
- What invoice model?
Choosing the right invoice model

+ **Hoster** will invoice coordinator for services rendered
  
  SLA defines pricing terms

+ **Fixed costs**
  
  Hardware costs? Cost per session? Cost per duration unit?

+ **Variable costs**
  
  Cost per QoS value? Average values? Cumulative values? Peak values?

+ **Penalty costs**
  
  QoS threshold values? Upper / Lower bounds
  
  Cost of overall SLA breach

+ **Banded pricing**
  
  Variable cost depends on some usage bands

+ **Hoster accounts** are needed to store invoices

+ **Cost models** should reward desired behaviour
  
  e.g. penalties can encourage minimum levels of customer QoE
edutain@grid invoice model

+ Pricing terms defined in SLA template
+ Fixed cost based on SLA duration
+ Variable cost proportional to QoS measurements
  Peak and accumulated values used in Edutain
  Banded pricing supported also if needed
+ Penalty cost based on QoS min/max thresholds
+ Hoster trade account service store invoices once SLA’s terminate
Summary of edutain@grid business layer

+ **Extended value chains**
  Customer(s) -> Coordinator(s)
  Coordinator(s) -> Hoster(s)

+ **SLA template and metrics**
  XML SLA template, metrics based on server-side QoS measurements
  Packet latency, Packet loss, Throughput, Frame rate, Client connection count

+ **Simple fast SLA negotiation**
  SLA template agreed by hoster, SLA instance created by coordinator

+ **Cross-hoster load management for ROIA**
  UI for QoS monitoring, zone migration & replication
  2 min predictions allow automated decisions for zone management

+ **Invoice model**
  SLA template agreed by hoster, SLA instance created by coordinator