Massivizing Online Games:
High Performance Computing and
High Quality Time

Alexandru Iosup
Parallel and Distributed Systems Group
Delft University of Technology

Our team: Undergrad Adrian Lascateu, Alexandru Dimitriu (UPB, Romania), ..., Grad Siqi Shen, Yong Guo (TU Delft, the Netherlands), ..., Staff Otto Visser, Dick Epema, Johan Pouwelse, Henk Sips (TU Delft), Thomas Fahringer, Radu Prodan (U. Innsbruck), Nicolae Tapus, Vlad Posea (UPB), ...
Massivizing Online Games as an HPC Problem

Online Gaming used to be art, may now be computing

Online Gaming used to be multimedia, is now HPC

Online Gaming used to be networking, is now all HPC

Online Gaming used to be v-worlds, is now many apps
What’s in a name? MSG, MMOG, MMO, …

Over 250,000,000 active players

Massively Social Gaming = (online) games with massive numbers of players (100K+), for which social interaction helps the gaming experience

1. Virtual World Sim
Explore, do, learn, socialize, compete

2. Game Data
Player stats and relationships, others

3. Game Content
Graphics, maps, puzzles, quests, culture
MSGs are a Popular, Growing Market

- 25,000,000+ subscribed players (from 250,000,000+ active)
- Over 10,000 MSGs in operation
- Subscription market size $7.5B+/year, Zynga $600M+/year

Sources: MMOGChart, own research. Sources: ESA, MPAA, RIAA.
Zynga, an Amazon WS User

Selling in-game virtual goods: Sources: CNN, Zynga.

“Zynga made est. $270M in 2009 from.”
http://techcrunch.com/2010/05/03/zynga-revenue/

“Zynga made more than $600M in 2010 from selling in-game virtual goods.”
S. Greengard, CACM, Apr 2011
World of Warcraft, a Traditional HPC User
(since 2003)

- 10 data centers
- 13,250 server blades, 75,000+ cores
- 1.3PB storage
- 68 sysadmins (1/1,000 cores)

Bungie, Computing then Serving 1.4PB/yr.

• Halo 3 is one of the many successful games
• Halo 3 players get, in 1.4PB
  • Detailed player profiles
  • Detailed usage stats
  • Ranking

• CERN produces ~15PB/year (10x larger)
Agenda

1. What’s in a Name?

2. Three Current Challenges
   1. Platform Scalability Challenge
   2. Gaming Analytics Challenge
   3. Content Generation Challenge

3. The Next Five Years

4. Conclusion
@large Research Challenge: V-World Platform for MMOGs

Scaling quickly to millions of players
- 1M in 4 days, 10M in 2 months
- Up-front and operational costs
- Performance, Scalability, & Cost

http://www.developeranalytics.com

Massivizing Social Games: High Performance Computing and High Quality Time – A. Iosup
Impact on Game Experience

Responsive game

Unresponsive game

Massivizing Social Games: High Performance Computing and High Quality Time – A. Iosup

[Source: Nae, Iosup, and Prodan, ACM SC 2008 and IEEE TPDS 2011]
Cloud Computing

VENI – @larGe: Massivizing Online Games using Cloud Computing
Resource Provisioning and Allocation

Static vs. Dynamic Provisioning

Massivizing Social Games: High Performance Computing and High Quality Time – A. Iosup

[Source: Nae, Iosup, and Prodan, ACM SC 2008]
@large Research Challenge: Continuous Analytics for MMOGs

Analyzing the behavior of millions of players, on-time
- Data mining, data access rights, cost v. accuracy, ...
- Reduce upfront costs
- Low response time & Scalable
- Large-scale Graph Processing
@large: Social Everything!

- **Social Network** = undirected graph, **relationship** = edge
- **Community** = sub-graph, density of edges between its nodes higher than density of edges outside sub-graph

(Analytics Challenge)

*Improve gaming experience*
- Ranking / Rating
- Matchmaking / Recommendations
- Play Style/Tutoring

Self-Organizing Gaming Communities
- Player Behavior
The CAMEO Framework

1. **Address community needs**
   - Can analyze skill level, experience points, rank
   - Can assess community size dynamically

2. **Using on-demand technology: Cloud Comp.**
   - Dynamic cloud resource allocation, Elastic IP

3. **Data management and storage: Cloud Comp.**
   - Crawl + Store data in the cloud (best performance)

4. **Performance, scalability, robustness: Cloud Comp.**

A. Iosup, CAMEO: Continuous Analytics for Massively Multiplayer Online Games on Cloud Resources. ROIA, Euro-Par 2009 Workshops, LNCS 6043, (2010)
@large: Sample Analytics Results
Activity and Social Network

- **Bridge Base Online (BBO)**: 1M+ players, top free site
- **Dataset**: 100K players
  - 9K group
  - Social relationships from bridge pairing
- **Large (~10K) online social groups can coordinate**
- **Identified player behavior**
  - community builder,
  - community member,
  - random player,
  - faithful player

---

M. Balint, V. Posea, A. Dimitriu, and A. Iosup, An Analysis of Social Gaming Networks in Online and Face to Face Bridge Communities, LSAP 2011.
@large: Sample Analytics Results

Analysis of Meta-Gaming Network

• “When you play a number of games, not as ends unto themselves but as parts of a larger game, you are participating in a metagame.” (Dr. Richard Garfield, 2000)

• XFire: since 2008 (3+ years), 500K of 20M players
@large Research Challenge: Content Generation for MMOGs

Generating content on time for millions of players
- Player-customized: Balanced, Diverse, Fresh
- Up-front and operational costs
- Response time, Scalability, & Cost
Player-Customized Content
Skill Level Distribution in RuneScape

- **RuneScape**: 135M+ open accounts (world record)
- **Dataset**: 3M players (largest measurement, to date)
  - 1,817,211 over level 100
  - Max skill 2,280

- **Number of mid- and high-level players is significant**

New Content Generation Challenge

Massivizing Social Games: High Performance Computing and High Quality Time – A. Iosup
@large: Sample Analytics Results
Skill Level Distribution in RuneScape

- Runescape: 135M active accounts, 7M active (2008)
- (largest MMOG msmt.)

- Player skill: distribution changes over time

(Procedural) Game Content (Generation)

Derived Content
NewsGen, Storification

Game Design
Rules, Mechanics, ...

Game Scenarios
Puzzle, Quest/Story, ...

Game Systems
Eco, Road Nets, Urban Envs, ...

Game Space
Height Maps, Bodies of Water, Placement Maps, ...

Game Bits
Texture, Sound, Vegetation, Buildings, Behavior,
Fire/Water/Stone/Clouds

Hendricks, Meijer, vd Velden, Iosup,
Procedural Content Generation for Games: A Survey,
ACM TOMCCAP, 2012
The POGGI Content Generation Framework

Only the puzzle concept, and the instance generation and solving algorithms, are produced at development time

* A. Iosup, POGGI: Puzzle-Based Online Games on Grid Infrastructures, EuroPar 2009 (Best Paper Award)
Puzzle-Specific Considerations
Generating Player-Customized Content

Puzzle difficulty
- Solution size
- Solution alternatives
- Variation of moves
- Skill moves

Player ability
- Keep population statistics and generate enough content for most likely cases
- Match player ability with puzzle difficulty
- Take into account puzzle freshness
Agenda

1. What’s in a Name?
2. Three Current Challenges

3. The Next Five Years
   1. Cloudification
   2. Mobile Social Gaming
   3. Social Everything!
   4. Content, Content, Content

4. Conclusion
Cloudification: PaaS for MSGs

(Platform Challenge)
Build MSG platform that uses (mostly) cloud resources

• Close to players
• No upfront costs, no maintenance
• Compute platforms: multi-cores, GPUs, clusters, all-in-one!
• Performance guarantees
• Code for various compute platforms—platform profiling
• Load prediction miscalculation costs real money
• What are the services?
• Vendor lock-in?
• My data
(Platform Challenge)
Support MSGs on mobile devices

- Mobiles everywhere (2bn+ users)
- Gaming industry for mobiles is new Growing Market
- SuperServer to generate content for low-capability devices?
- Battery for 3D/Networked games?
- Where is my server? (Ad-hoc mobile gaming networks?)
- Security, cheat-prevention
Social Everything!

- **Social Network** = undirected graph, **relationship** = edge
- **Community** = sub-graph, density of edges between its nodes higher than density of edges outside sub-graph

(Analytics Challenge)

**Improve gaming experience**

- Ranking / Rating
- Matchmaking / Recommendations
- Play Style/Tutoring

**Self-Organizing Gaming Communities**

- Player Behavior
Content, Content, Content

(Content Challenge)

Produce and distribute content for 1BN people

- Game Analytics → Game statistic
- Crowdsourcing
- Storification
- Auto-generated game content
- Adaptive game content
- Content distribution/
  Streaming content
Massivizing Online Gaming

- Million-user, multi-bn market
- V-World, Content, Analytics

Current Technology

- Upfront payment
- Cost and scalability problems
- Makes players unhappy

@large: Our Vision

- HPC has to help
- Economy of scale with clouds

@large: Ongoing Work

- Content: POGGI Framework
- Platform: edutain@grid
- Analytics: CAMEO Framework

@large: The Future

- Happy players
- Happy cloud operators

Publications Gaming and Clouds

2008: ACM SC
2009: ROIA, CCGrid, NetGames, EuroPar (Best Paper Award), ...
2010: IEEE TPDS, Elsevier CCPE
2011: Book Chapter CAMEO, IEEE TPDS, IJAMC
2012: IPDPS, CCGrid, ...

Graduation (Forecast)

2012–14: 3PhD, 6Msc, 6BSc
Thank you for your attention!
Questions? Suggestions? Observations?

More Info:
- http://www.st.ewi.tudelft.nl/~iosup/research.html
- http://www.st.ewi.tudelft.nl/~iosup/research_gaming.html
- http://www.st.ewi.tudelft.nl/~iosup/research_cloud.html

Alexandru Iosup

A.Iosup@tudelft.nl
http://www.pds.ewi.tudelft.nl/~iosup/ (or google “iosup”)
Parallel and Distributed Systems Group
Delft University of Technology
• Trade-off Utility-Cost still needs investigation
• Performance and Cost are not both improved by the policies we have studied
(Variable) Blackbox Performance Engineering

- Performance Evaluation of Four Commercial Clouds
  - Amazon EC2, GoGrid, Elastic Hosts, Mosso
  - Resource acquisition
  - Single- and Multi-Instance benchmarking

- Low compute and networking performance\(^1\)

- Performance variability over time\(^2\)

---


2- Iosup et al., On the Performance Variability of Production Cloud Services, CCGrid 2011, pds.twi.tudelft.nl/reports/2010/PDS-2010-002.pdf
Multi-Resource Provisioning/Release

- Time for multi-resource increases with number of resources

GAE Dataset: Run Service

- **Fibonacci [ms]**: Time it takes to calculate the 27th Fibonacci number
- Highly variable performance until September
- Last three months have stable performance (low IQR and range)
Online Scheduling + Optimization

ExPERT

ExPERT recommended:

\[ (N = 3, T = T_{ur}, D = 2T_{ur}, M_r = 0.02) \]

in words: Send \( N = 3 \) instances to the unreliable pool during the tail phase, each timed out after twice the average task time \( (D = 2T_{ur}) \). Send the next instance after the average task time passes \( (T = T_{ur}) \). Use only one \( (\#_{ur} = 50, 50 \times M_r = 1) \) reliable machine at a time.

Performance Metrics

- Makespan very similar
- Very different job slowdown

Cost Metrics

Actual Cost

Charged Cost

- Very different results between actual and charged
  - Cloud charging function an important selection criterion
- All policies better than Startup in actual cost
- Policies much better/worse than Startup in charged cost

Single Resource Provisioning/Release

- Time depends on instance type
- Boot time non-negligible

CPU Performance of Single Resource

- ECU definition: “a 1.1 GHz 2007 Opteron” ~ 4 flops per cycle at full pipeline, which means at peak performance one ECU equals 4.4 gigaflops per second (GFLOPS)
- Real performance 0.6..0.1 GFLOPS = ~1/4..1/7 theoretical peak

- Low efficiency for parallel compute-intensive applications
- Low performance vs cluster computing and supercomputing
Performance Stability (Variability)

- High performance stability for the best-performing instances

AWS Dataset (4/4): Summary

- **All services exhibit time patterns in performance**
- **EC2**: periods of special behavior
- **SDB and S3**: daily, monthly and yearly patterns
- **SQS and FPS**: periods of special behavior
- **Read Latency [s]**: Time it takes to read a “User Group”
- Yearly pattern from January to August
- The last four months of the year exhibit much lower IQR and range
  - More stable performance for the last five months
  - Probably due to software/infrastructure upgrades

GAE Dataset (3/4): Memcache

- **PUT [ms]**: Time it takes to put 1 MB of data in memcache.
- Median performance per month has an increasing trend over the first 10 months
- The last three months of the year exhibit stable performance
GAE Dataset (4/4): Summary

- All services exhibit time patterns
  - **Run Service**: daily patterns and periods of special behavior
  - **Datastore**: yearly patterns and periods of special behavior
  - **Memcache**: monthly patterns and periods of special behavior
  - **URL Fetch**: daily and weekly patterns, and periods of special behavior

---

Massivizing Social Games: High Performance Computing and High Quality Time – A. Iosup

Focus on game content generation on grids

- Use existing middleware
- Control MMOG-specific workload demands and variability (soft guarantees for low response time by pre-generating content)

... but do not forget lessons on system design

- Add components for capacity planning and process monitoring
Continuous Analytics for MMOGs

MMOG Data = raw and derivative information from the virtual world (millions of users)

Continuous Analytics for MMOGs = Analysis of MMOG data s.t. important events are not lost

- Data collection
- Data storage
- Data analysis
- Data presentation
- ... at MMOG rate and scale
Continuous Analysis for MMOGs
Main Uses By and For Gamers

1. Support player communities
2. Understand play patterns (decide future investments)
3. Prevent and detect cheating or disastrous game exploits (think MMOG economy reset)
4. Broadcasting of gaming events
5. Data for advertisement companies (new revenue stream for MMOGs)
Other Uses for MMOG Data

Social Sciences
- The emergence and performance of ad hoc groups in contemporary society
- Emergent behavior in complex systems

Economy
- Contemporary economic behavior

Psychology
- Games as coping mechanism (minorities)
- Games as cure (agoraphobia)

Biology
- Disease spread models
The CAMEO Framework [ROIA09]
Continuous MSG Analytics on the Cloud

- Use own resources for continuous or predicted load
- Use cloud (on-demand, paid-for, guaranteed) resources for sparse or excess load
- Users (peers) may also provide service (future)

CAMEO: Analytics Capabilities

1. Various pieces of information
   - Skill level, experience points, rank

2. Single and Multi-snapshot analysis

3. Analysis functions already implemented
   - Ranking by one or more pieces of information
   - Community statistical properties for a piece of information
   - Identification of Top-K players in single/multi-snapshot
   - Evolution of (Top-)K players
   - Evolution of average community skill
   - Identification of players with special skill combos
CAMEO: Cloud Resource Management

- Snapshot = dataset for a set of players
- More machines = more snapshots per time unit
CAMEO: Exploiting Cloud Features

- Machines close(r) to server
  - Traffic dominated by small packets (latency)

- Elastic IP to avoid traffic bans
  (legalese: acting on behalf of real people)

Cost of Continuous RuneScape Analytics

Billing Statement: April 1, 2009
Billing Cycle for this Report: March 1 - March 31, 2009

<table>
<thead>
<tr>
<th>Rate</th>
<th>Usage</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon Elastic Compute Cloud</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amazon EC2 running Linux/UNIX</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$0.10 per Small Instance (m1.small) instance-hour (or partial hour)</td>
<td>2,097 Hrs</td>
<td>209.70</td>
</tr>
<tr>
<td>Amazon EC2 Bandwidth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$0.100 per GB Internet Data Transfer - all data transfer into Amazon EC2</td>
<td>611.005 GB</td>
<td>61.10</td>
</tr>
<tr>
<td>Taxes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$0.170 per GB Internet Data Transfer - first 10 TB / month data transfer out of Amazon EC2</td>
<td>507.121 GB</td>
<td>86.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>67.83</td>
</tr>
<tr>
<td>Charges due on April 1, 2009†</td>
<td>424.85</td>
<td></td>
</tr>
</tbody>
</table>

- Put a price on MMOG analytics (here, **$425/month**, or less than **$0.00015/user/month**)
- Trade-off accuracy vs. cost, runtime is constant
Performance Results: Why Choosing the Cloud Matters

- Location of machines influences MMOG analytics performance (data acquisition)