Massivizing Social Games: Today and the Next Five Years

Our team: Undergrad Adrian Lascateu, Alexandru Dimitriu (UPB, Romania), ..., Grad Vlad Nae (U. Innsbruck, Austria), Siqi Shen, Nezih Yigitbasi (TU Delft, the Netherlands), Staff Dick Epema, Johan Pouwelse, Henk Sips (TU Delft), Thomas Fahringer, Radu Prodan (U. Innsbruck), Nicolae Tapus, Mihaela Balint, Vlad Posea (UPB), Guillaume Pierre (Vrije U.).

Alexandru Iosup
Parallel and Distributed Systems Group
Delft University of Technology
MSGs are a Popular, Growing Market

- 25,000,000 subscribed players (from 200,000,000+ active)
- Over 10,000 MSGs in operation
- Market size 7,500,000,000$/year

Sources: MMOGChart, own research. Sources: ESA, MPAA, RIAA.
FarmVille, a Massively Social Game

Zynga Game

FarmVille

Average Number of Active Players

27 million daily / 75 million monthly

Player Profile

N/A average age
60% female, 40% male

Most Popular Time to Play (EST)

8-9 AM

Average Session

33 minutes

That's a lot

FarmVille boasts 118 million total installs. It has more monthly active users than the population of France.

Sources: CNN, Zynga.

Source: InsideSocialGames.com
What’s in a name?

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

1. Virtual world
   Explore, do, learn, socialize, compete
   
2. Content
   Graphics, maps, puzzles, quests, culture
   
3. Game data
   Player stats and relationships
Agenda

1. What’s in a Name?

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

3. The Next Five Years

4. Conclusion
Computational Model for the Server

- Single sequential loop
- 3 steps in each loop:
  1. Game-world state update
  2. Entity interaction computation (dominant for MSGs)
  3. Entity state updates
- Load generated by (2) non-deterministic ← human factor

[Source: Nae, Iosup, and Prodan, ACM SC 2008]
Model for Entity Interaction Computation

- Player to Player/Player to Environment interaction
  - **Low** interaction: $O(n)$, e.g. RTS
  - **Medium** interaction: $O(n \cdot \log(n))$ – RPG quest maps
  - **High** interaction: $O(n^2)$ – RPG war maps, FPS
  - **Very High** interaction: $O(n^2 \cdot \log(n))/O(n^3)$, unit-target matching, team path-finding, maxflow alg., …
Load Impact on Game Experience

Responsive game

Unresponsive game

[Source: Nae, Iosup, and Prodan, ACM SC 2008]
Research Challenge: Solve the Platform Problem of MMOGs

The Platform Problem of MMOGs
Scaling quickly to millions of players
- 1M in 4 days, 10M in 2 months
- Up-front and operational costs
- Performance, Scalability, & Cost
Online games hosting model

• Generic Online Games (non-MM)
  • **Static**: dedicated isolated single servers

• MMOGs
  • **Static**: dedicated clusters - using parallelization techniques

• Problems with these approaches
  1. Large amount of over-provisioning
  2. Non-efficient coverage of the world for the provided service
Background on Cloud Computing

• “The path to abundance”
• On-demand capacity
• Pay what you use
• Great for web apps (EIP, web crawl, DB ops, I/O)

VS

• “The killer cyclone”
• Not so great performance for compute- or data-intensive applications
• Long-term perf. variability


Proposed hosting model: dynamic

- Using data centers for dynamic resource allocation

  - Main advantages:
    1. Significantly lower over-provisioning
    2. Efficient coverage of the world is possible

[Source: Nae, Iosup, and Prodan, ACM SC 2008]
MSG Ecosystem: Model

- **Game operators**
  - Past player activity/business model → **Predicted load → requests**

- **Data centers**
  - **Local** time-space **renting policy → offers**
  - Time-Space renting policy, e.g., 1 node-hour
  - Resource allocation: central request-offer matching
  - Rules for ranking request-offer match:
    1. The offer size and type vs. the request
    2. The geographical proximity offer-request
    3. The finer grained resources (quantity & time)

[Source: Nae, Iosup, and Prodan, ACM SC 2008]
Experimental Setup [1/3] Discrete-Event Simulator

• **Input**
  - Traces from RuneScape, a real top-5 MMOG
    - 7 countries, 3 continents
    - More than 130 game worlds
  - Consisting of
    - Geographical location
    - Number of clients
    - Over 10,000 samples at 2 min. interval, 2 weeks

• **Output** (for every time-step)
  - Resource allocation decisions
  - Resource allocation
  - Performance metrics

[Source: Nae, Iosup, and Prodan, ACM SC 2008]
Experimental Setup [2/3]

Environment

- 1 game operator
- 17 data centers
- 11 data center time-space renting policies

<table>
<thead>
<tr>
<th>Location</th>
<th>Data Centers</th>
<th>Machines (total)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Europe</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finland</td>
<td>2</td>
<td>8 machines</td>
</tr>
<tr>
<td>Sweden</td>
<td>2</td>
<td>8 machines</td>
</tr>
<tr>
<td>U.K.</td>
<td>2</td>
<td>20 machines</td>
</tr>
<tr>
<td>Netherlands</td>
<td>2</td>
<td>15 machines</td>
</tr>
<tr>
<td><strong>North America</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S. (West)</td>
<td>2</td>
<td>35 machines</td>
</tr>
<tr>
<td>Canada (West)</td>
<td>1</td>
<td>15 machines</td>
</tr>
<tr>
<td>U.S. (Central)</td>
<td>1</td>
<td>15 machines</td>
</tr>
<tr>
<td>U.S. (East)</td>
<td>2</td>
<td>32 machines</td>
</tr>
<tr>
<td>Canada (East)</td>
<td>1</td>
<td>10 machines</td>
</tr>
<tr>
<td><strong>Australia</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>2</td>
<td>8 machines</td>
</tr>
</tbody>
</table>

[Source: Nae, Iosup, and Prodan, ACM SC 2008]
Experimental Setup [3/3]  
Performance Metrics

• **Resource over-allocation [%]**
  - The wasted resources vs. optimal allocation at each simulation time step for all utilized machines (cumulative)

• **Resource under-allocation [%]**
  - The amount of resources needed but not allocated, for all machines (computed individually)

• **Significant under-allocation events (count)**
  - Number of times the resource under-allocation is >1%, for a period of 2 minutes → **people leave**

[Source: Nae, Iosup, and Prodan, ACM SC 2008]
Q: What is the penalty for static vs. dynamic allocation?

250%  
25%  

[Source: Nae, Iosup, and Prodan, ACM SC 2008]
Impact of Load Prediction Accuracy

Q: How does the prediction accuracy impact the resource allocation? A: Good prediction matters.

[Source: Nae, Iosup, and Prodan, ACM SC 2008]
**Impact of Interaction Compute-Intensiveness**

Q: How are different MMOG types handled under dynamic resource provisioning?

- **Interaction models**
  - Low ~ O(n). Medium. High. Very High ~ O(n^3)

A: Over-allocation, Under-allocation worse with increase in interaction compute-intensiveness

[Source: Nae, Iosup, and Prodan, ACM SC 2008]
Agenda

1. What’s in a Name?

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

3. The Next Five Years

4. Conclusion
(Procedural) Game Content (Generation)

- 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

Derived Content
- NewsGen, Storification


Massivizing Social Games: Today and the Next Five Years – A. Iosup
Research Challenge: Solve the Content Problem of MMOGs

The Content Problem of MMOGs

Generating content on time for millions of players
- Player-customized: Balanced, Diverse, Fresh
- Up-front and operational costs
- Response time, Scalability, & Cost
The New Content Generation Process*

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)
The POGGI Framework

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
Workflow Execution Engine for Puzzle Instance Generation

Generic engine for puzzle generation

- Can plug-in different puzzles
- Can plug-in different solvers
- Can plug-in different policies for instance generation

Reduce execution overheads

- By-pass RMS (similar to Condor glide-ins, Falkon/Swift, etc., but for WFs instead of tasks)
- Execute on single resource (current implementation, simplicity)
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
Why Not Let Players Generate Puzzles?

How to control production pipeline?
After all, game developers sell content not technology.

How to select content?
Ranking problems, diversity problems.

How to avoid game exploits?
Virtual currency = Real currency

<table>
<thead>
<tr>
<th>server</th>
<th>faction</th>
<th>100 Gold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aegwynn</td>
<td>Alliance</td>
<td>$0.57</td>
</tr>
<tr>
<td>Horde</td>
<td></td>
<td>$0.45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Amount</th>
<th>Price</th>
<th>Quick link</th>
<th>Shop profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>5mn Gold</td>
<td>$15.90</td>
<td>Buy now!</td>
<td>OgPal</td>
</tr>
</tbody>
</table>

Source: mmobux.com, Aug 2009

User-generated content is clearly an interesting research area, but that’s another story.
Agenda

1. What’s in a Name?

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

3. The Next Five Years

4. Conclusion
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
Research Challenge: Solve the Analytics Problem of MMOGs

The Analytics Problem of 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
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)
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)
Sample Game Analytics Results
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
Sample MMOG Analytics Results (2/2)
Skill Level Distribution in RuneScape

- Dataset 2: **3,531,478 players** (largest MMOG msmt.)
  - 3,239,089 over level 100
  - Max skill 2,488

- Distribution changed over time

* A. Iosup, POGGI: Puzzle-Based Online Games on Grid Infrastructures EuroPar 2009 (Best Paper Award)
Sample Game Analytics Results
BBO 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

*Massivizing Social Games: Today and the Next Five Years – A. Iosup*
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><strong>Amazon Elastic Compute Cloud</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>View/Edit Service</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Amazon EC2 running Linux/UNIX</strong></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><strong>Amazon EC2 Bandwidth</strong></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>$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><strong>Taxes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>67.83</td>
</tr>
<tr>
<td><strong>Charges due on April 1, 2009+</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>424.85</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)
Agenda

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

3. The Next Five Years
   1. Social Everything!
   2. Cloudification
   3. Mobile Social Gaming
   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
Mobile Social Gaming and the SuperServer

(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
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
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
Summary

Massive Social Gaming

- Million-user, multi-bn market
- Sim, Content, Analytics

Current Technology

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

Our Vision

- Scalability & Automation
- Economy of scale with clouds

Ongoing Work

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

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

Graduation (Forecast)
2010/2011: 1PhD, 2Msc, 4BSc
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

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