Overview

• Goal:
  • To explain how bandwidth limits cause scalability problems. In non-trivial environments its simply not possible to communicate all states to all parties.

• Topics:
  • Management of awareness
  • Interest specification
  • Server partitioning
Interest Specification

- Users are not omniscient beings and thus they can’t be interested in every event in a non-trivial scene
  - Plausibility needs to be maintained
- Systems thus model the user’s awareness so that they can only deliver a conservative approximation to the necessary events so that the user’s illusion of a shared virtual environment is maintained
Awareness Categories

• Primary awareness
  • Those users you are collaborating with
  • Typically near by, typically highest bandwidth available

• Secondary awareness
  • Those users that you might see in the distance
  • Can in principle interact with them within a few seconds by movement

• Tertiary awareness
  • All other users accessible from same system (e.g. by teleporting to them)
System Goals

• Attempt to keep
  • overall system utilization to a manageable level
  • client inbound bandwidth at a manageable level
  • client outbound bandwidth to a manageable level

• To do this
  • Have clients discard received information
  • Have the system manage awareness
  • Have clients generate information at different levels of detail
Managing Awareness

• A complex distributed problem
• Users’ expressions of interest in receiving information balanced against system’s and other clients’ capabilities
• Awareness scheme is partly dependent on the networking architecture, but most awareness management schemes can be applied to different architectures
• Spatial layout is the primary moderating factor on awareness
Filtering traffic

Application
Filter on Send
Network Routing

Message Routing
Network Routing

Network Routing
Message Filtering
Network Routing

Application
Filter on Receive
Network Routing
Spatial Partitions

- Global Partitions
  - Static Grid
  - Hierarchical Grid
  - Locales
- Local Partitions
  - Aura / nearest neighbours
  - Visibility
Global Partitions: Static Cells

- A static partition into regular cells
- Players only communicate with other players in the same cell
Global Partitions: Static Cells

- A slightly more sophisticated partitioning
- Each player receives information from 7 nearest cells
- As they move they change the cells they receive from
- No longer abrupt changes across borders
Global Partitions: Irregular

Two irregular partitionings
Spatial Partitions: Auras / Nearest Neighbours

- Aura focus nimbus model from Benford, Greenhalgh, et al.
- Network connections are set up if users are close to each other and “looking” or “listening” in their direction.
Spatial Partitions: Local Visibility
Spatial Partitions: Local Visibility
Practical Systems

- A system such as Second Life™ utilizes a regular grid layout with one server per region
  - Regions are laid out on a mostly-contiguous map
- However is a game session, far too many players want to access a specific game content
- A game *shard* is a complete copy of a system, you connect to one system and see one player cohort
- A game *instance* is similar, but is replication of a particular area (e.g. dungeon) to support one group of players within a cohort. Often created on demand.
Game Shards

Master Server

New Process

Server_A

Server_B
Game Regions

Master Server

New Process

Server_A

Server_B

Server_C

Server_D

1 2 3
Game Regions & Instances
Summary

- Scalability depends on a choice of awareness mechanism
  - Requires a logical scalability mechanism based on what is most relevant for the users
  - Needs to consider bottlenecks at several points
  - Most common strategy is to partitioning users