Determining Optimal Update Period for Minimizing Inconsistency in Multi-server Distributed Virtual Environments

Li Yusen, Wentong Cai

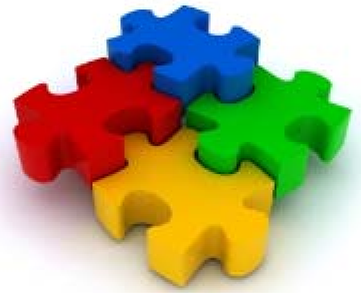
Presented by Stephen John Turner

PDCC, SCE

Nanyang Technological University, Singapore

NANYANG TECHNOLOGICAL UNIVERSITY

Overview



- Introduction
- System Model
- Problem Definition
- Optimization Method
- Performance Evaluation
- Conclusion and Future Work

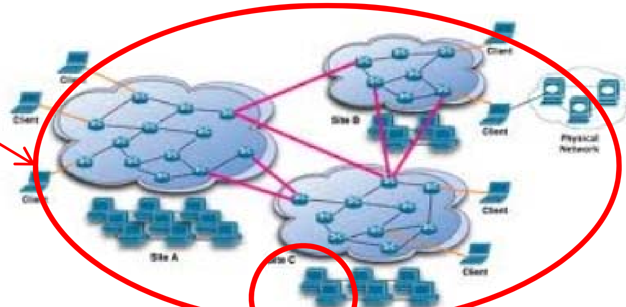
NANYANG TECHNOLOGICAL UNIVERSITY

Introduction to the Research Problem

- Distributed Virtual Environment

virtual world

To deploy on a group of computers connected via networks



NANYANG TECHNOLOGICAL UNIVERSITY

node, client/user, player

avatar

object/entity

Introduction to the Research Problem

- Fundamental goal
  - Create a common and consistent representation of the virtual world among all users
  - Any state change of an entity in the virtual world should be disseminated to all users who require it in a timely manner
- Challenges
  - Network latency
  - Resource limitations as the number of users increases (e.g., MMOG)
    - Computational power
    - Network capacity

NANYANG TECHNOLOGICAL UNIVERSITY

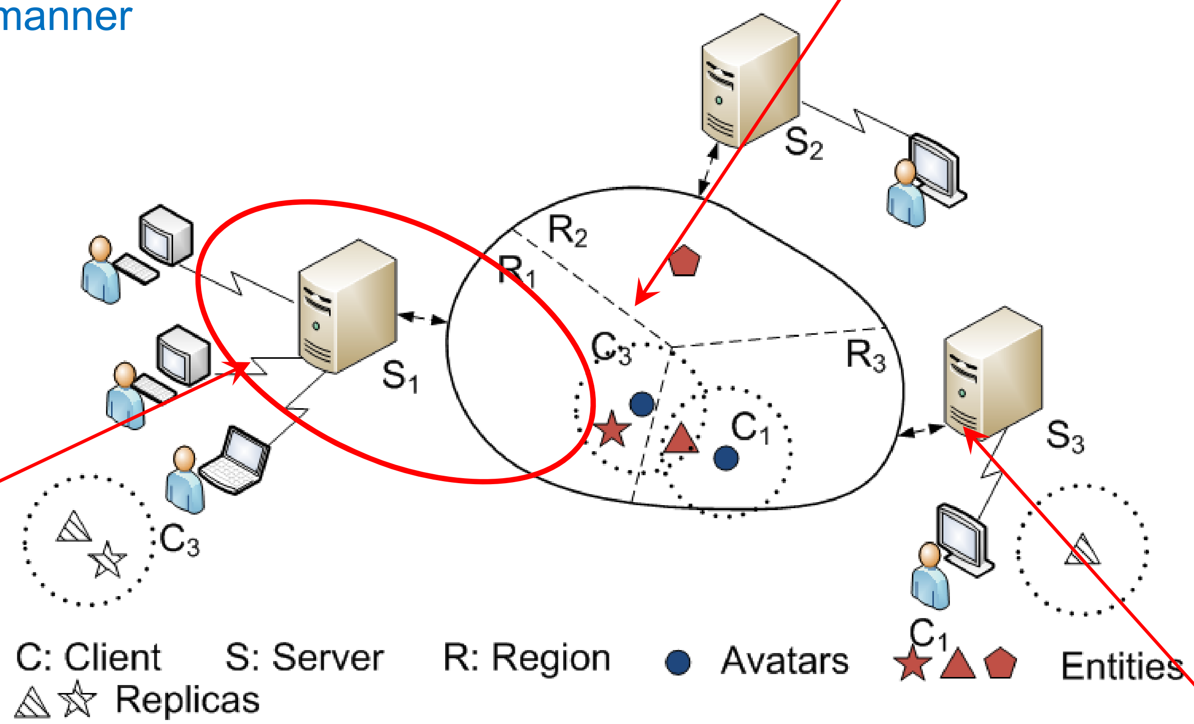
Introduction to the Research Problem

- Research Objective
  - Derive state update schedules for improving consistency in multi-server DVEs with network capacity constraints
- Contributions
  - Time-space inconsistency is used to evaluate the total inconsistency of an multi-server DVE
  - The problem of minimizing total inconsistency is formulated as an Inequality Constrained Problem (ICP)
  - Interior point method is used to solve the problem

System Model

The virtual world is partitioned into several fixed regions

Servers are connected to each other in a peer to peer manner



Each region is maintained by one server (e.g., R_1 is maintained by S_1)

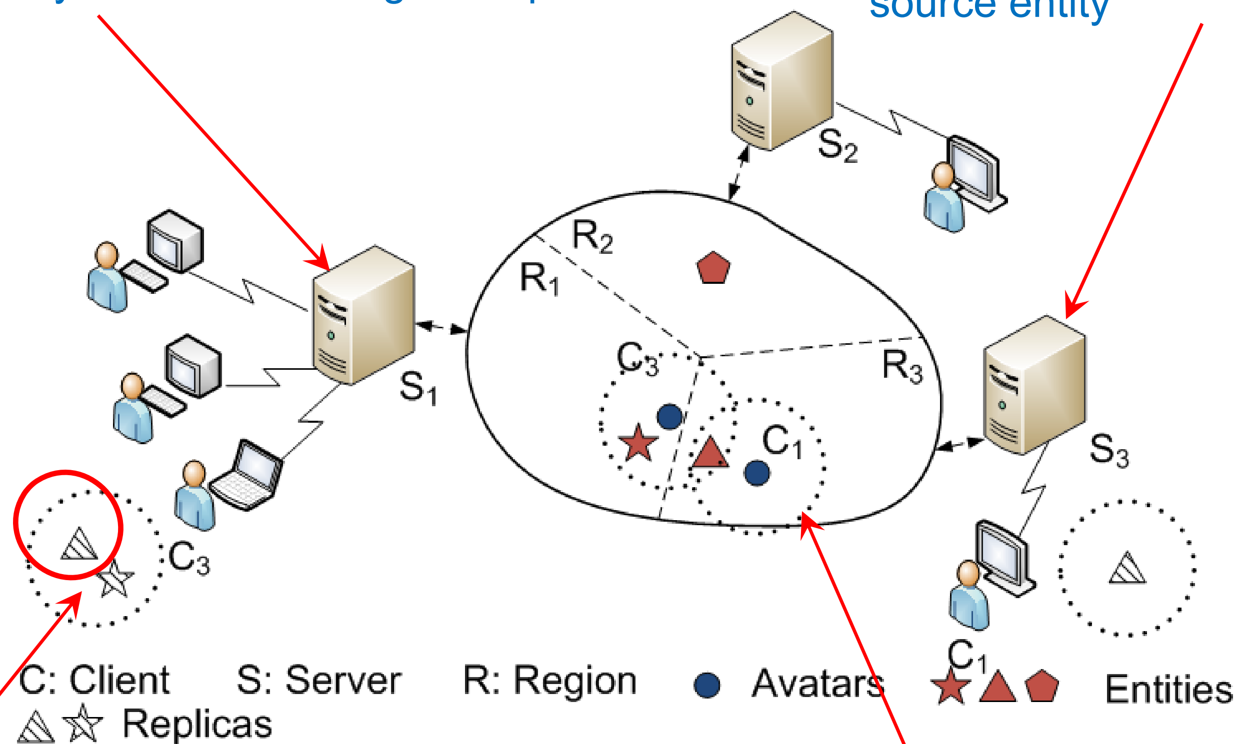
Client connects to the server if its avatar is residing in the region maintained by the server (e.g., C_1 is connected to S_3)

NANYANG TECHNOLOGICAL UNIVERSITY

System Model

The contact server is the server that is connected by the client holding the replica

For a replica (e.g., triangle), the target server is the server that maintains the source entity



The entities in an avatar's AOI will be replicated at this avatar's client side

Area of Interest (AOI) is used to define a neighborhood area for avatars

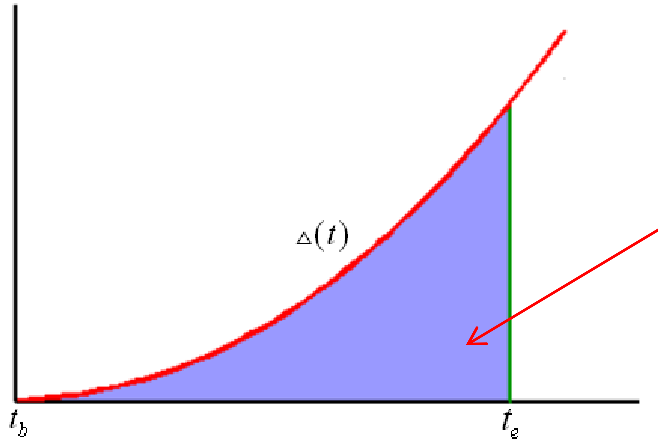
NANYANG TECHNOLOGICAL UNIVERSITY

System Model

- State Update Schema
  - Client first sends the operation on an entity to the server maintaining this entity
  - Server executes the operation and disseminates new states to all interested clients for updating the replicas
  - For a replica, if its target server and contact server are the same, target server directly disseminates state update to the replica
  - If its target server and contact server are different, target server first sends update to contact server, contact server forwards to replica

Problem Definition

- Time-space inconsistency
  - $\Delta(t)$: spatial difference between a replica and its source entity
  - Time-space inconsistency over $[t_b, t_e]$



$$\Omega(s) = \int_{t_b}^{t_e} \Delta(t)\, dt$$

Problem Definition

- Objective
  - To minimize total time-space inconsistency over all replicas with a set of servers with limited network capacity
- Assumptions
  - For each replica, assume after the replica receives a position update, the difference $\Delta(t)$ grows following an increasing function $\delta(\cdot)$, $\Delta(t) = \delta(t-(tlast+d))$
  - Configurations such as world partition, client assignment, server side bandwidth, etc. remain unchanged over a period

NANYANG TECHNOLOGICAL UNIVERSITY

Problem Definition

- Theorem
  - In multi-server DVEs, for any replica, given a fixed number of updates allowed in a period at the target server, these updates should be disseminated periodically over this period for minimizing time-space inconsistency
  - To minimize total time-space inconsistency over all replicas over a period with a set of servers with limited network bandwidth, we just need to determine the optimal update period of each replica

Problem Definition (cont.)

- Notations
  - NS the number of servers in the DVE
  - S_i the i^th server in the DVE
  - NR the number of replicas in the DVE
  - NR^i the number of replicas whose target server is S_i
  - NR^i_j the number of replicas whose target server is S_i and contact server is S_j
  - r_k the k^th replica in the DVE
  - $e(r_k)$ the entity which r_k is replicating, i.e., source entity of r_k
  - $T(r_k)$ the target server id of r_k
  - $C(r_k)$ the contact server id of r_k

NANYANG TECHNOLOGICAL UNIVERSITY

Problem Definition (cont.)

- Notations
  - R_i^T the set of replicas whose target server is S_i
  - R_i^C the set of replicas whose contact server is S_i
  - d_i the transmission delay of position update of replica r_i from target server to r_i
  - *a* bandwidth consumption for disseminating a position update
  - *b* bandwidth consumption for receiving and forwarding a position update
  - c_i the network capacity of S_i
  - f_r the update frame length of each server
  - p_k update period of replica r_k

Problem Definition

Objective function to minimize: total time-space inconsistency over all replicas over period T

Network capacity constraint for each server

$$f(p) = \sum_{i=1}^{NR} \frac{T}{p_i} \int_0^{p_i} \delta_i(t)dt, \quad (5)$$

subject to

$$g_i(p) \le 0, \quad i = 1, ..., NS, \quad (6)$$

$p = [p_1, ..., p_{NR}]$ and $g_i(p)$ is defined as

$$g_i(p) = a \cdot \left(\sum_j \frac{T}{p_j}\right) + b \cdot \left(\sum_k \frac{T}{p_k}\right) - c_i \cdot \frac{T}{f_r}, \quad (7)$$

where $j \in \{j|1 \le j \le NR, r_j \in R_i^T\}$ and $k \in \{k|1 \le k \le NR, r_k \in R_i^C\}$.

Bandwidth consumption on disseminating position updates for the replicas whose target server is this server

Bandwidth consumption on forwarding

NANYANG
TECHNOLOGICAL
UNIVERSITY

Convex Optimization

- Problem Transformation
  - Let $q_i = \frac{1}{p_i}$, the problem is converted to minimize

$$f(q) = \sum_{i=1}^{NR} T \cdot q_i \cdot \int_0^{\frac{1}{q_i}} \delta_i(t)dt, \quad (8)$$

subject to

$$g_i(q) \leq 0, \quad i = 1, ..., NS, \quad (9)$$

$q = [q_1, ..., q_{NR}]$ and $g_i(q)$ is denoted as

$$g_i(q) = a \cdot \left(\sum_j T \cdot q_j \right) + b \cdot \left(\sum_k T \cdot q_k \right) - c_i \cdot \frac{T}{f_r}, \quad (10)$$

  - Convex Property

$$\nabla^2_{qq} f(q) \geq 0.$$

NANYANG TECHNOLOGICAL UNIVERSITY

Convex Optimization

- Interior Point Method
  - The basic idea is to approximate the original problem to the following problem

$$minimize \quad f(q) + \sum_{i=1}^{NS} \left[-\left(\frac{1}{\alpha}\right) log(-g_i(q)) \right], \qquad (15)$$

  - α is a parameter that sets the accuracy of the approximation
- Solution
  - Define $L(q) = f(q) + \sum_{i=1}^{NS} \left[-\left(\frac{1}{\alpha}\right) log(-g_i(q)) \right], \qquad (16)$
  - $L(q)$ is a convex function and if $\nabla_q L(q^*) = 0$ holds, q^* is a global minimum.
  - Gradient Descent Method

NANYANG TECHNOLOGICAL UNIVERSITY

Convex Optimization

- Gradient Descent Method
  - Iterative Algorithm

$$q^{(k+1)} = q^{(k)} - t^{(k)} \cdot \nabla L(q^{(k)})$$

  - t is a constant value, can be different for each iterative step
- Values Need to Know
  - $$\frac{\partial L(q)}{\partial q_i} = \frac{\partial f(q)}{\partial q_i} - \sum_{i=1}^{NS} \left[\frac{1}{\alpha} \cdot \frac{1}{g_i(q)} \cdot \frac{\partial (g_i(q))}{\partial q_i} \right]$$
  - $$\frac{\partial f(q)}{\partial q_i} = T \cdot \left(\int_0^{\frac{1}{q_i}} \delta_i(t)dt - (\frac{1}{q_i}) \cdot \delta_i(\frac{1}{q_i}) \right)$$
  - $\delta_i(\frac{1}{q_i^k})$ and $\int_0^{1/q_i^k} \delta_i(t)dt$ need to be estimated

NANYANG TECHNOLOGICAL UNIVERSITY

Performance Evaluation

- Experimental parameters

Parameter	Default Value
DVE Dimension Number of Servers	5000x5000 (distance units)^2 25
Number of Regions	100
Number of Clients/avatars	1500
AOI Size	500x500
Average Network Latency	100ms
Variance of Latency Frame Length a, b Entity Moving Speed Network Capacity	0.95 0.025s 1 unit [0.1, 10] distance units/frame [5, 300] units

Simulation Results

- Converge Speed of Iterative Algorithm

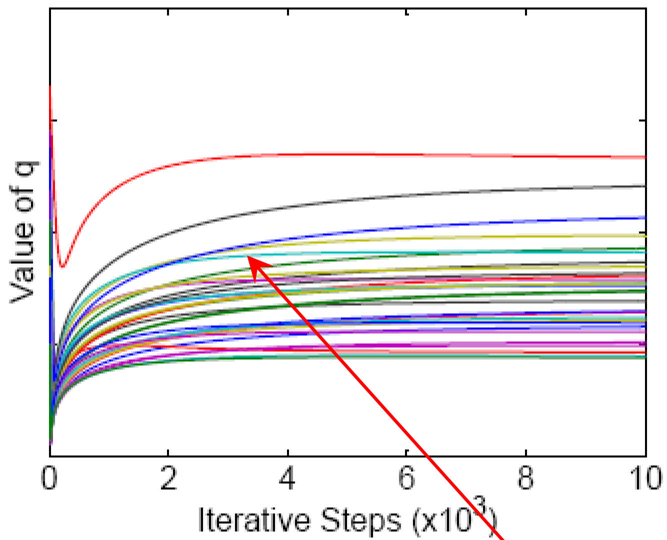
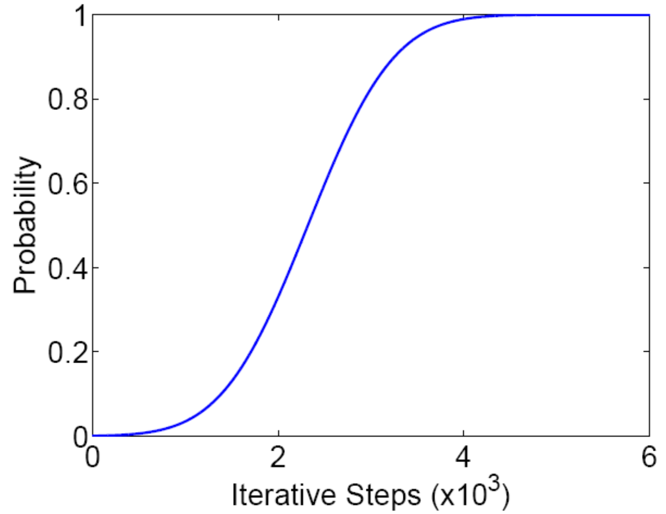


Fig. 4. Convergece of q



Most of variables converge after 3000 iterative steps

Simulation Results

- Impact of α in the Interior Point Approximation

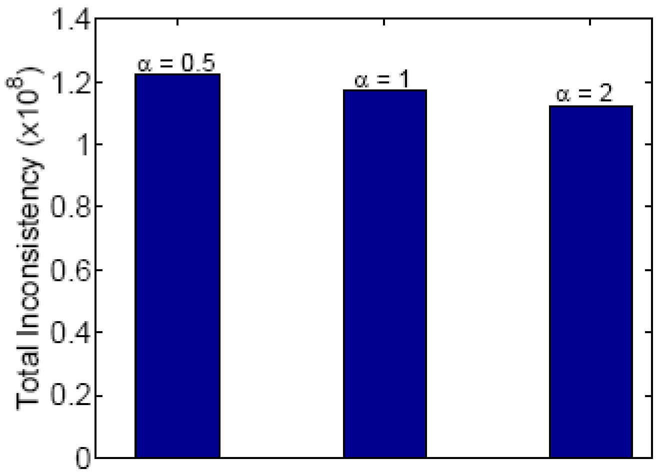


Fig. 6. Impact of α

parameter	values
Network latency	100ms
Network capacity	50
T	60s
Server Number	25

Larger α makes more accurate, but more difficult to converge

NANYANG TECHNOLOGICAL UNIVERSITY

Simulation Results

- Impact of Network Capacity

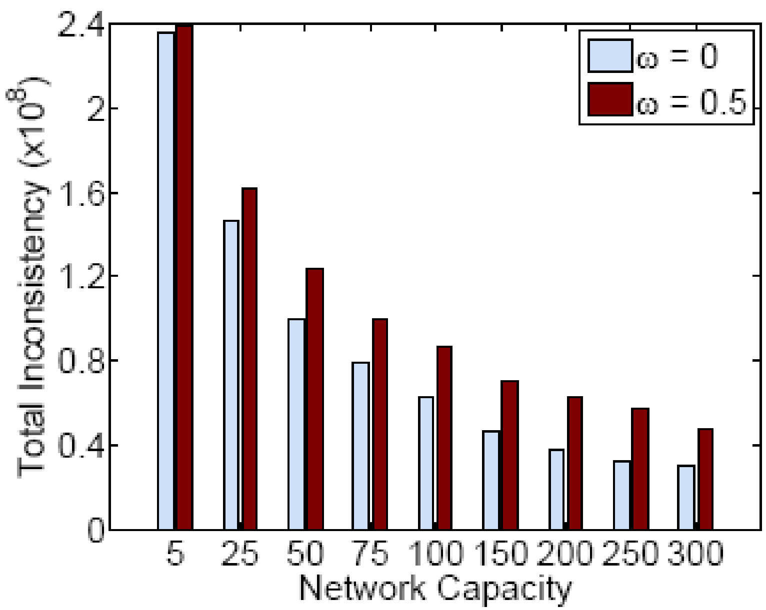


Fig. 7. Impact of Network Capacity

parameter	values
Network latency	100ms
T	60s
Server number	25

NANYANG TECHNOLOGICAL UNIVERSITY

Simulation Results

- Impact of Network Latency

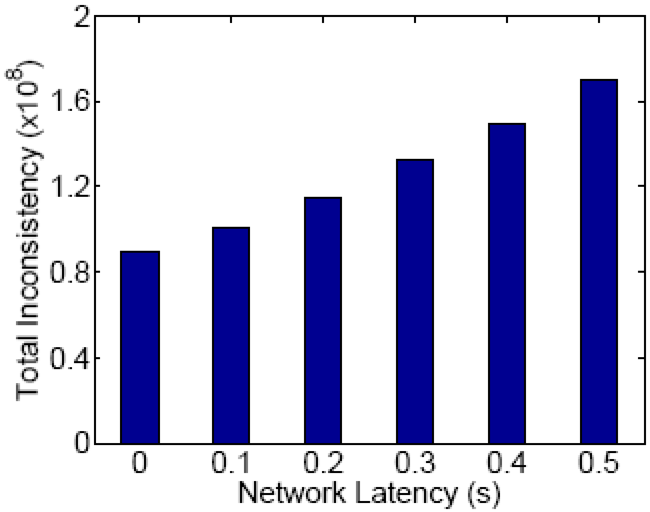


Fig. 8. Impact of Network Latency

parameter	values
Network capacity	50
T	60s
Server number	25

NANYANG TECHNOLOGICAL UNIVERSITY

Simulation Results

- Impact of Inter-server Communication

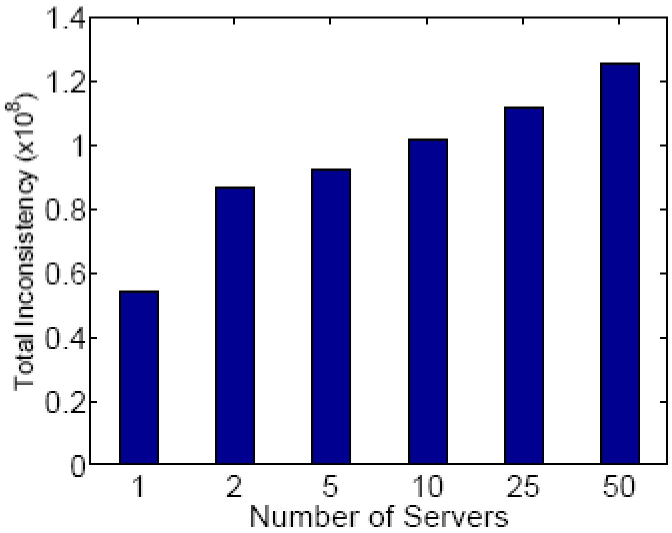


Fig. 9. Impact of Inter-server Communication

parameter	values
Network latency	100ms
Network capacity	50
T	60s

NANYANG TECHNOLOGICAL UNIVERSITY

Conclusion and Future Work

- Conclusion
  - Study the update scheduling issues in multi-server DVEs with limited network bandwidth
  - Formulate and solve the problem for an ideal situation where configurations keep unchanged
- Future Work
  - Update schedules in practical systems

NANYANG
TECHNOLOGICAL
UNIVERSITY

The End

NANYANG
TECHNOLOGICAL
UNIVERSITY