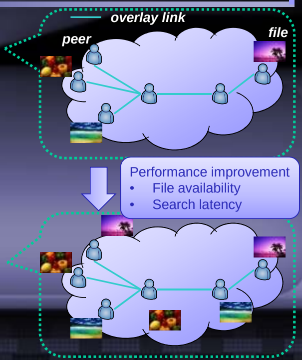
A Caching Algorithm using Evolutionary Game Theory in a File-Sharing System

Masahiro Sasabe,
Naoki Wakamiya, Masayuki Murata
Osaka University, Japan
sasabe@comm.eng.osaka-u.ac.jp

Contents

P2P file-sharing system

- A logical network is constructed by establishing logical links among nodes
  - Files and messages travel over the logical network
- Each node caches files and shares them with others
  - Files are original or retrieved from others
- Cooperatively caching a file on multiple nodes enhances the system performance



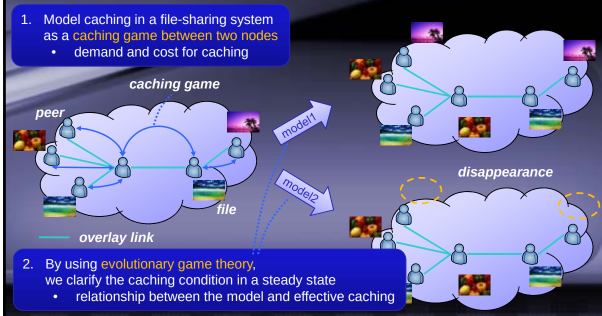
Problem of cooperative caching

- Caching a file incurs cost
  - processing load, storage capacity
- Some nodes may not be cooperative to cache files
  - Other nodes cannot use the file due to disappearance
- A mechanism to accomplish effective caching is required
  - It is difficult to monitor and manage all nodes
  - Can the effective caching emerge from autonomous and selfish node behavior?

Research object

- Proposal of a mechanism to accomplish effective caching in file-sharing systems
  - We model caching in a file-sharing system as a caching game
    - We take into account the cost and benefit of caching as selfishness of users
  - We clarify the relationship between the model and the number of cache files
    - How does individual behavior affect system performance?
    - Is it possible that selfish node behavior leads to cooperative caching?
    - Evolutionary game theoretic approach

Research overview



Evolutionary game theory

- What is evolutionary game theory?
  - It tries to figure out the evolutionary process of organisms by game theory
  - Superior genes yield much offspring
  - Superior strategies acquire more payoffs
    - Individual strategy = behavior defined by a gene
    - Payoffs acquired by a strategy = number of offspring
- Evolutionary game theory provides us with a framework to investigate the following characteristics
  - Strategy distribution in a steady state
  - Condition to reach the steady state

General game

- A game between two players is defined as a payoff matrix

Table 1. payoff matrix (general)

player 1 \ player 2	cooperator	defector
cooperator	(R, R)	(S, T)
defector	(T, S)	(P, P)

  - ex) $(S,\ T)$
    - A defector exploiting a cooperator gets T
    - The exploited cooperator receives S
  - Well known games
    - Prisoner's Dilemma Game: $T>R>P>S$
    - Snowdrift Game: $T>R>S>P$

Caching game

- We model caching in a file-sharing system as a caching game between two neighboring nodes
  - Two strategies (caching, no caching)
    - We assume a file-sharing system consists of multiple caching games each of which deals with a single file
  - Benefit for caching
    - Satisfaction obtained by the use of the file : b
  - Cost for caching
    - Processing load: cl (Tab.2)
    - Storage capacity: cs (Tab.3)

Table 2. payoff matrix (caching game, cost: load)

player 1 \ player 2	caching	no caching
caching	$(b-c_l, b-c_l)$	$(b-2c_l, b)$
no caching	$(b, b-2c_l)$	$(0,0)$

Table 3. payoff matrix (caching game, cost: storage)

player 1 \ player 2	caching	no caching
caching	$(b-c_s, b-c_s)$	$(b-c_s, b)$
no caching	$(b, b-c_s)$	$(0,0)$

Derivation of strategy distribution in a steady state

- Replicator dynamics
  - Mathematical model of a phenomenon where a superior strategy increases
  - It can be applicable when the number of individuals is relatively large and the network among them is mean-field like
    - full-mesh network, global information
- Agent-based dynamics
  - Simulation-based model of a phenomenon that a superior strategy spreads over the network in a hop-by-hop manner
    - various network, local information

Replicator dynamics
- overview -

- The number of individuals increases in proportional to the superiority of the strategy
  - superiority=average payoffs acquired by the strategy – average payoffs
- Derivation process in a general game (Tab. 1)
  - ratio of cooperators: x
  - ratio of defectors: $1-x$

Table 1. payoff matrix (general)

player 1 \ player 2	cooperator	defector
cooperator	(R, R)	(S, T)
defector	(T, S)	(P, P)

Replicator dynamics
- derivation process -

- Average payoffs of cooperators
  - $Rx+S(1-x)$
- Average payoffs of defectors
  - $Tx+P(1-x)$
- Average payoffs of the whole nodes
  - $x\{Rx+S(1-x)\}+(1-x)\{Tx+P(1-x)\}$
- Superiority of cooperators
  - $\{(R+P-T-S)x+S-P\}(1-x)$
- Replicator dynamics
  - $\dot{x}=\{(R+P-T-S)x+S-P\}(1-x)x$

Table 1. payoff matrix (general)

player 1 \ player 2	cooperator	defector
cooperator	(R, R)	(S, T)
defector	(T, S)	(P, P)

Growth rate depends on the strategy's superiority

Replicator dynamics

- equilibria -

- Replicator dynamics
  - $\dot{x}=\{(R+P-T-S)x+S$
- Equilibria
  - By substituting $\dot{x}=0$ ➔ $x=0,1,\frac{P-S}{R+P-T-S}$
- Stability of equilibria in the caching game
  - $0\le x\le 1\rightarrow(1-x)x\ge 0$
  - From the definition of the caching game

$$(R+P-T-S)x+S-P>0 \quad \text{if } x<\frac{P-S}{R+P-T-S}$$

$$(R+P-T-S)x+S-P<0 \quad \text{if } x>\frac{P-S}{R+P-T-S}$$

  - Stable equilibria (steady state) $x=\frac{P-S}{R+P-T-S}$

2007/7/3 ISCC 2007 13

Replicator dynamics

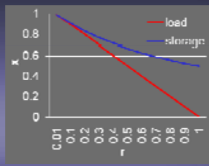
- Characteristics of caching games -

- Cost-to-benefit ratio of mutual cooperation: ***r***
- Relationship between ***r*** and ***x***
  - cost: load

$$x=\frac{b-2c_l}{b-c_l},\ r=\frac{c_l}{b-c_l} \quad ➔ \quad x=1-r$$

  - cost: storage

$$x=\frac{b-c_s}{b},\ r=\frac{c_s}{b-c_s} \quad ➔ \quad x=\frac{1}{1+r}$$



  - characteristics
    - ***x*** deteriorates with the increase of ***r*** (decrease of demand ***b***)
    - cost model of load causes file disappearance when ***r*** = 1
    - cost model of storage achieves high file-availability

2007/7/3 ISCC 2007 14

Agent-based dynamics

- Node ***i*** plays a game once with all neighboring nodes (one generation)
- It determines the strategy of the next generation based on average ***Ai*** of payoffs acquired
  1. Select Node ***j*** randomly from neighboring nodes for comparison of their payoffs
  2. Decide the next strategy
     - If ***Aj>Ai,*** node ***i*** imitates the strategy of node ***j*** with the following probability

$$P_A(i,j)=\frac{A_j-A_i}{T-P}$$

     - Otherwise, it does not change its strategy

Strategy is imitated in accordance with its superiority

2007/7/3 15

Simulation model

- Number of nodes: 1000
- Network topologies
  - scale-free network based on BA model
  - waxman random network (α=0.15, β=0.2)
  - ***m***=2, 4
- Caching Costs ***cl*** and ***cs*** are set to 1
  - From the definition of ***r***, benefit ***b*** is $b=\frac{1+r}{r}$
- Initial strategy distribution is fifty-fifty
- Simulation results are average over twenty simulations in steady states

2007/7/3 ISCC 2007 16

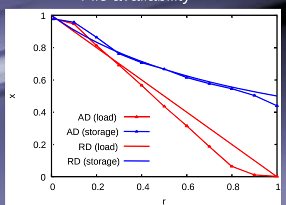
Evaluation criteria

- File availability
  - Ratio of nodes taking the strategy "caching:" ***x***
    - Number of cache files is product of ***x*** and the number of nodes
- Search latency
  - Average hop count between a node to its closest provider including itself
  - If a file disappears from the system, we alternatively use the maximum hop count between two arbitrary nodes

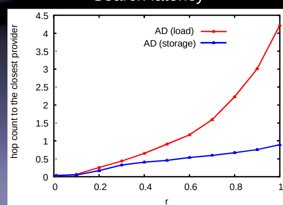
2007/7/3 ISCC 2007 17

Impact of cost model

File availability



Search latency



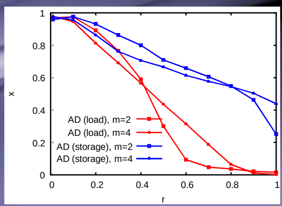
File availability
•Cost model of storage achieves high file-availability

Search latency
•Search latency increases with growth of ***r***
•Cost model of storage suppress the increase of search latency

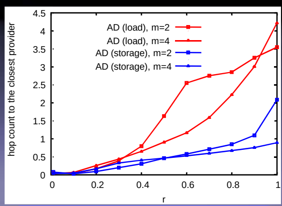
2007/7/3 ISCC 2007 18

Impact of network structure
- average degree -

File availability



Search latency

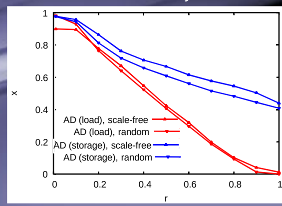


File availability
- •Smaller *m* promotes to increase high-demand files
  - •High degree nodes tends to acquire more payoffs and be imitated
  - •The impact of high degree nodes is accelerated by their surrounding low degree nodes

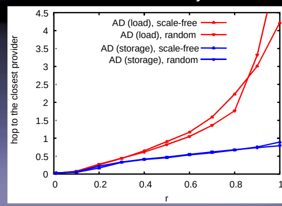
Search latency
- •In the cost model of storage, the difference of *m* does not affect so much

Impact of network structure
- degree distribution -

File availability



Search latency



File availability
- •Scale-free network slightly encourages nodes to cooperatively behave

Search latency
- •Search latency is not much affected by the difference of degree distribution

Conclusions and future work

- Conclusions
  - We investigated effective caching in a file-sharing system by using evolutionary game theory
    - We first modeled caching in a file-sharing system as a caching game between two nodes
    - We derived the relationship between models and the number of cache files in a steady state
      - Theoretical approach by replicator dynamics
      - Simulation-based approach by agent-based dynamics
    - The cost model of storage could achieve a file-sharing system with high file-availability and low search-latency independently of the network structure
- Future work
  - Caching game taking into account combination of multiple costs including dynamic costs