A Game Theoretic Analysis of Selfish Content Replication on Graphs

Valentino Pacifici and György Dán
ACCESS Linnaeus Center, School of Electrical Engineering
KTH, Royal Institute of Technology, Stockholm, Sweden
E-mail: {pacifici,gyuri}@kth.se

***Abstract*—Replication games are a model of the problem of content placement in computer and communication systems, when the participating nodes make their decisions such as to maximize their individual utilities. In this paper we consider replication games played over arbitrary social graphs; the social graph models limited interaction between the players due to, e.g., the network topology. We show that in replication games there is an equilibrium object placement for arbitrary social graphs. Nevertheless, if all nodes follow a myopic strategy to update their object placements then they might cycle arbitrarily long before reaching an equilibrium if the social graph is non-complete. We give sufficient conditions under which such cycles do not exist, and propose an efficient distributed algorithm to reach an equilibrium over a non-complete social graph.**

I. Introduction

Replicating content close to its users has long been used in computer architectures and in the Internet to improve system performance. In computer architectures CPU caches have been used to decrease memory access latencies [1]. In the Internet caches are employed to provide faster access to content for local customers and at the same time to decrease the amount of network traffic [2], [3], [4]. Content replication is also at the core of clean-slate information centric network architectures [5].

The problem of content replication is often modeled by a distributed replication group [6], [7], [8], [9]. A replication group consists of nodes and users located near the nodes. The nodes can replicate objects, which are accessed by local or remote users. The cost incurred by a user accessing an object depends on whether the object is replicated locally, in a remote node, or is not replicated in any node. A commonly studied problem of replication is how to minimize the total cost of users accessing the objects through either optimal node placement [6] or through the optimal allocation of objects to nodes [7].

However there is often no central authority that would be able to enforce the optimal solution on the nodes, thus they would not implement the optimal solution if they can individually fare better by deviating from it [8], [9], [10]. Instead, nodes would replicate the objects that minimize their own costs, and would update the set of replicated objects as a response to the decisions made by other nodes.

In this paper we model the problem of selfish replication as a non-cooperative graphical game. The social graph models the limited interactions among nodes defining neighborhood relationships between them, which influence the cost of accessing replicas. We show that Nash equilibria exist for arbitrary social graphs, but if the social graph is non-complete, the nodes might cycle arbitrarily long before reaching an equilibrium state. Based on our results we give a sufficient condition under which a simple and efficient distributed algorithm can be used to reach an equilibrium and illustrate the efficiency of the algorithm with numerical results.

II. System model

In the following we describe the system model, and formulate the problem of replication as a non-cooperative game.

A. The replication problem

We consider a set of nodes N and a set of objects $\mathcal{O}$. The demand for object $o \in \mathcal{O}$ at node $i \in N$ is given by the rate $w_i^o \in \mathbb{R}_+$. We consider that every object $o \in \mathcal{O}$ has unit size, $S^o = 1$, which is a reasonable simplification if objects are divisible into unit-sized chunks. Node i has integer storage capacity $K_i \in \mathbb{N}_+$, which it uses to replicate objects locally. We describe the set of objects replicated at node i with the $|\mathcal{O}|$ dimensional vector $r_i = (r_i^1, \dots, r_i^{|\mathcal{O}|})$, whose component $r_i^o \in \{0, 1\}$ is 1 if object o is replicated in node i. Due to the limited storage capacity $\sum_o r_i^o S^o \leq K_i$.

Every node is located at a vertex of an undirected graph $\Gamma(N, E)$, called the *social graph*. We denote the set of neighbors of player i by $\mathcal{N}(i)$, i.e., $\mathcal{N}(i) = \{j|(i,j) \in E\}$. The social graph allows us to consider a generalized version of the cost model described in [1], whose variations were used in [8], [9], [10]. In our model the marginal cost of accessing object o in node i is α_i if the object is replicated in node i, it is β_i if it is replicated in a node $j \in \mathcal{N}(i)$ neighboring i, and it is γ_i otherwise. We consider the practically relevant case when $\alpha_i \leq \beta_i < \gamma_i$, or equivalently

$$0 \leq \delta_i \triangleq \frac{\beta_i - \alpha_i}{\gamma_i - \alpha_i} < 1. \qquad (1)$$

To ease notation we say that an object $o \in \mathcal{O}$ is *i-available* if it is replicated by at least one of player i's neighbors in which case

$$\pi_i^o \triangleq \prod_{j \in \mathcal{N}(i)} (1 - r_j^o) = 0,$$

otherwise we say that object $o \in \mathcal{O}$ is not *i-available*.

The cost of node i due to object o is proportional to the demand w_i^o, and is a function of r_i and the replication states r_{-i}^o of the neighboring nodes

$$C_i^o(r_i^o, r_{-i}^o) = w_i^o \left(\alpha_i r_i^o + (1 - r_i^o)\left[\gamma_i \pi_i^o \;+\; \beta_i (1 - \pi_i^o)\right]\right) \tag{2}$$

B. The graphical replication game

We consider a system in which the goal of every node is to minimize its own total cost. We model this problem of selfish replication as a multiplayer non-cooperative game played on a graph, called a graphical game. The players are the nodes, the set of actions of player i is the set of feasible replication configurations $\mathcal{R}_i = \{r_i | \sum_o r_i^o S^o \leq K_i\}$, and the cost function of player i is given by $C_i(r_i, r_{-i}) = \sum_o C_i^o(r_i^o, r_{-i}^o)$. The social graph influences the cost function via the neighbor set $\mathcal{N}(i)$, i.e., the cost function of player i is entirely specified by the actions of players $j \in \mathcal{N}(i)$. The goal of player i is to choose a replication strategy r_i that minimizes its total cost given the strategy profile r_{-i} of the other players

$$\arg\min_{r_i} C_i(r_i, r_{-i}) = \arg\min_{r_i} \sum_o C_i^o(r_i^o, r_{-i}^o). \tag{3}$$

Observe that the quantity $U_i^o(r_i^o, r_{-i}^o) = C_i^o(0, r_{-i}^o) - C_i^o(r_i^o, r_{-i}^o)$ expresses the cost saving that player i achieves through object o given the other players' replication strategies. We define the utility function of player i as the sum of the cost savings $U_i(r_i, r_{-i}) = \sum_o U_i^o(r_i^o, r_{-i}^o)$. Note that finding the minimum cost is equivalent to finding the maximum utility. Consequently, the problem of replication can be modeled by the strategic game $\mathcal{G} =< N, (\mathcal{R}_i), (U_i) >$. We can express the utility $U_i^o(r_i^o, r_{-i}^o)$ of player i by substituting (2) into the definition of the cost saving

$$U_i^o(r_i^o, r_{-i}^o) = r_i^o w_i^o \left[\beta_i (1 - \pi_i^o) + \gamma_i \pi_i^o - \alpha_i\right]. \tag{4}$$

We note a property of the utility defined in (4): the utility of player i due to object o is independent of the other players' strategies if she does not replicate the object, i.e., $r_i^o = 0 \Rightarrow U_i^o(0, r_{-i}^o) = 0$. If player i replicates object o then the cost saving is

$$U_i^o(1, r_{-i}) = \begin{cases} w_i^o \left[\gamma_i - \alpha_i\right] = c_{io} & \text{if } \pi_i^o = 1 \\ w_i^o \left[\beta_i - \alpha_i\right] = \delta_i c_{io} & \text{if } \pi_i^o = 0 \end{cases} \tag{5}$$

III. Results

In this section we investigate the existence of Nash Equilibria in graphical replication games and whether the players will reach an equilibrium if they myopically update their strategies. For the proofs of all the theorems and propositions that follow, we refer to [11].

A. Existence of equilibria

The first question we address is whether every graphical replication game possesses a pure strategy *Nash equilibrium* (NE). It is known that for a complete social graph pure strategy NE exist in a replication game [9], but it is not known whether pure strategy NE exist for non-complete social graphs. In what follows we show that pure strategy NE exist for arbitrary social graphs.

We first define a *best reply* of player i as a best strategy r_i^* of player i given the other players' strategies

$$U_i(r_i^*, r_{-i}) \geq U_i(r_i, r_{-i}) \;\; \forall \; r_i \in \mathcal{R}_i. \tag{6}$$

The NE is a strategy profile r^* in which every player's strategy is a best reply to the other players' strategies

$$U_i(r_i^*, r_{-i}^*) \geq U_i(r_i, r_{-i}^*) \;\; \forall \; r_i \in \mathcal{R}_i, \; \forall \; i \in N. \tag{7}$$

Finally, we define a *best reply path* as a sequence of strategy profiles, such that in every step t there is one player that strictly increases its utility by updating her strategy to a best reply $r_i(t)$ with respect to the other players' most recent strategies $r_{-i}(t-1)$. A best reply path terminates when no player can increase its utility, in which case a NE is reached. Hence, to show the existence of NE it is enough to show that there is a particular strategy profile starting from which there is at least one finite best reply path.

Consider the strategy profile $r(0)$ that consists of the best replies that the players would play on an edgeless social graph. In this strategy profile every player i replicates the K_i objects with highest demands w_i^o. Let us consider now a best reply path starting from the strategy profile $r(0)$. For $t \leq n$ each player $i \in N$ has a chance to play her first best reply at $t = i$. For $t > n$ they play in an arbitrary order. We showed in [11] that following this dynamic the utilities of the players cannot decrease for $t > 0$. Nevertheless, every time a player updates her strategy her utility must strictly increase. Since the players' utilities cannot increase indefinitely, the best reply path must end in a Nash equilibrium. Hence we can state the following.

Theorem 1. *Every graphical replication game possesses a pure strategy Nash equilibrium.*

B. Reaching an equilibrium state

The existence of equilibrium states is important, but in a distributed system it is equally important that the nodes would be able to reach an equilibrium state using a distributed algorithm. The algorithm used to prove Theorem 1 can be adequate if the demands for the objects in the nodes w_i^o never change, so once a NE is reached, the nodes will not deviate from it. Nevertheless, the algorithm would be inefficient if the demands can change over time, as the equilibrium states for different demands are, in general, different. Hence, an important question is whether the players will reach a NE given an arbitrary initial strategy profile, e.g., a NE for past demands, and given the myopic decisions they make to update their strategies.

It is easy to show that if in every time step t every player i simultaneously updates her strategy to her best reply with respect to $r_{-i}(t-1)$, the dynamic can cycle. For example consider two nodes and $|\mathcal{O}| \geq 2$. Let $c_{i1} > c_{i2} > c_{i1}\delta_i$ and $K_i = 1$. If the initial replication strategies are $r_i(0) = (1,0)$ then in the next two steps $r_i(1) = (0,1)$ and $r_i(2) = (1,0)$ for both players, and so on.
As an alternative, consider a sequence of best reply steps as defined in Section III-A, the result of an asynchronous best reply dynamic. A natural question is whether all best reply paths are finite irrespective of the initial strategy profile $r(0)$ and the social graph. The answer is no in general: in Table I we show the players' updates that form a cycle in a best reply path. The arrangement of the players and their relevant preferences are shown in Figure 1.

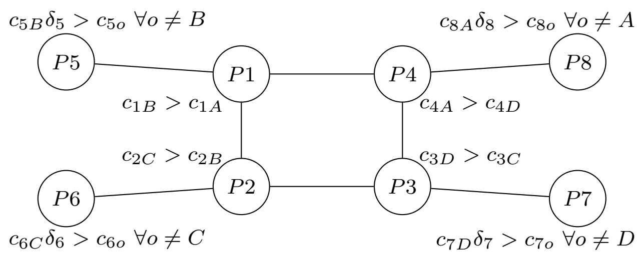

Fig. 1: Social graph and players' preferences that allow a cycle in best replies.

Player	*P1*	*P2*	*P3*	*P4*	P5	P6	P7	P8
r(0)	A	B	D	A	B	C	D	A
r(1)	A	B	↓C	A	B	C	D	A
r(2)	↓B	B	C	A	B	C	D	A
r(3)	B	B	C	↓D	B	C	D	A
r(4)	B	↓C	C	D	B	C	D	A
r(5)	↓A	C	C	D	B	C	D	A
r(6)	A	C	↓D	D	B	C	D	A
r(7)	A	↓B	D	D	B	C	D	A
r(8)	A	B	D	↓A	B	C	D	A

TABLE I: Cycle in best replies for the game played over the social graph in Figure 1. An arrow shows the best reply played at each step.

This negative example raises the question under what conditions nodes would reach an equilibrium state in a finite number of steps. In the following we distinguish between complete and non-complete social graphs.

Theorem 2. *Every best reply path in a replication game played over a complete social graph is finite.*

We know, however, that if the social graph is non-complete then this is not necessarily the case.

The case of non-complete social graph: Given that cycles might exist in the best reply paths for non-complete social graphs, an important question is whether the players might cycle forever without being able to reach an equilibrium state. In the following we show that if we consider replication games with $K_i = 1$, then from every strategy profile there exists at least one finite best reply path that leads to a NE. This property is known in the literature as *weak acyclicity in best replies.*
Investigating the restrictions on the strategy profiles which can be included in a cycle, we can prove the following

Proposition 3. *From every strategy profile of a graphical replication game with $K_i = 1\ \forall i \in N$ there exists a best reply path that leads to a NE in a finite number of steps.*

In the following we show that if we introduce a small restriction on the marginal costs of accessing objects then cycles cannot exist even if the social graph is non-complete.

We define an *improvement step* of player i at step t as an update of her strategy $r_i(t)$ to $r_i(t+1)$, such that its utility increases

$$U_i(r_i(t+1), r_{-i}(t)) > U_i(r_i(t), r_{-i}(t)). \tag{8}$$

The *best reply step* defined in (6) is special case of an improvement step. A sequence of improvement steps is called an *improvement path.* An improvement path terminates when no player can perform an improvement step, i.e., in an equilibrium.

In order to avoid cycles, we consider a subset of the set of improvement paths. We define a *lazy improvement step* of player i as an improvement step with minimal number of changes among all improvement steps that lead to the same utility. Formally, $r_i(t+1)$ is a lazy improvement step if there is no $r'_i(t+1) \neq r_i(t+1)$ for which

$$\begin{aligned} U_i(r_i(t+1), r_{-i}(t)) &= U_i(r'_i(t+1), r_{-i}(t)) \text{ and} \\ |I_i(t+1)| &> |I'_i(t+1)| \end{aligned}$$

where $I_i(t+1)$ is the *inserted set* defined as $I_i(t+1) = \{o | r_i^o(t) = 0 \wedge r_i^o(t+1) = 1\}$. If players only make lazy improvement steps then we can give a sufficient condition under which all improvement paths are finite.

Proposition 4. *In a graphical replication game with $\beta_i = \alpha_i\ \forall i \in N$ every lazy improvement path is finite.*

The case $\beta_i = \alpha_i$ was considered as a model of cooperative caching between peering ISPs in [10].

C. Fast convergence based on graph coloring

In the previous section we showed that, under certain conditions, the players always reach a Nash equilibrium if they update their strategies asynchronously. Unfortunately the implementation of the asynchronous update rule in a distributed system would require global synchronization, which can be impractical in large distributed systems. Hence, an important question is whether the players would always reach a Nash equilibrium even if some players would update their strategies simultaneously. With the following theorem we show that relaxing the requirement of asynchronicity is indeed possible.

Theorem 5. *Consider a graphical replication game with $\beta_i = \alpha_i \ \forall i \in N$. If player i makes an improvement step at time t only if no neighboring player $j \in \mathcal{N}(i)$ makes an improvement step at time t, then every lazy improvement path is finite.*

We refer to this dynamic as the *plesiochronous better reply dynamic* (PBRD), as opposed to the asynchronous dynamic (ABRD) considered in Proposition 4. In order to maximize the convergence speed of PBRD we need to find a minimum vertex coloring of Γ, i.e., we have to find the chromatic number $\chi(\Gamma)$ of graph Γ. Finding the chromatic number is NP-hard in general, but efficient distributed graph coloring algorithms exist [12], which can be used to find a coloring in a distributed system. Given a coloring, the number of steps required to reach equilibrium can be significantly smaller than for ABRD for sparse graphs. We illustrate the convergence speedup of PBRD compared to ABRD in Figure 2. The figure shows the average number of steps needed to reach equilibrium as a function of the edge probability in Erdős-Rényi random graphs with 87 vertices. For the PBRD we used the Welsh-Powell algorithm to find a coloring [13]. Each player had storage capacity $K = 5$ and we considered two scenarios, $\alpha_i = \beta_i$ and $\alpha_i \neq \beta_i$. Each data point is the average of the results obtained on 160 random graphs with the same edge probability. The figure shows the 95% confidence intervals for the case $\alpha_i = \beta_i$. We omitted the confidence intervals for $\alpha_i \neq \beta_i$ to improve readability.

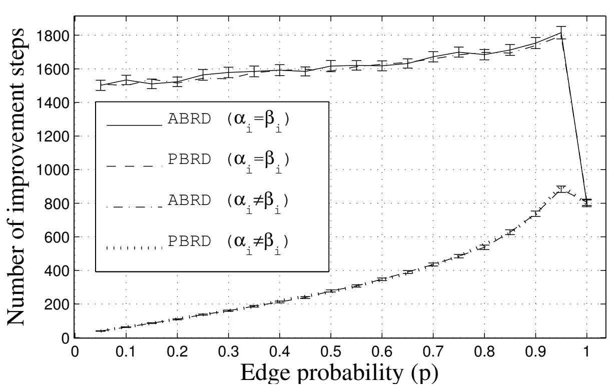

Fig. 2: Average number of improvement steps needed to reach a NE for ABRD and PBRD vs. the edge probability of the Erdős-Rényi random graphs used as social graph.

The results confirm that PBRD converges significantly faster compared to ABRD, especially over sparse social graphs. The existence of cycles when $\alpha_i \neq \beta_i$ does not make a significant impact on the results. The figure also confirms that the convergence properties are different on a complete social graph than on a sparse graph, as the number of steps necessary for ABRD to reach a NE drops for $p = 1$. This observation is in accordance with the difference in computational complexity of finding the optimal object replication strategy [1]: the problem is NP-complete on a non-complete social graph, but is polynomial in the number of players if the social graph is complete. A thorough analysis of the relationship between the complexity of finding the optimal solution and the convergence properties of improvement paths as a function of the graph topology is subject of our future work.

IV. Conclusion

In this paper we considered the problem of replication of contents by a set of selfish nodes, which replicate content to minimize their own costs. We modeled the problem as a graphical replication game, a replication game played over a social graph. We showed that independently of the social graph there always exists an equilibrium state from which no node wants to deviate, but the social graph affects the ease of reaching such an equilibrium state. Over a complete social graph the nodes can follow a simple myopic strategy and would always reach an equilibrium in a finite number of steps, but over a non-complete social graph they could cycle arbitrarily long before reaching an equilibrium. Finally, we provided a condition under which cycles do not exist, and based on this result we proposed an efficient algorithm to reach an equilibrium state over sparse social graphs.

References

[1] A. Leff, J. Wolf, and P. Yu, “Replication algorithms in a remote caching architecture,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 4, no. 11, pp. 1185–1204, 1993.

[2] S. Dykes and K. Robbins, “A viability analysis of cooperative proxy caching,” in *Proc. of IEEE INFOCOM*, Apr. 2001.

[3] A. Wolman, M. Voelker, N. Sharma, N. Cardwell, A. Karlin, and H. M. Levy, “On the scale and performance of cooperative web proxy caching,” in *Proc. of ACM Symposium on Operating Systems Principles (SOSP)*, 1999, pp. 16–31.

[4] T. Karagiannis, P. Rodriguez, and K. Papagiannaki, “Should internet service providers fear peer-assisted content distribution?” in *Proc. of Internet Measurement Conference*, 2005, pp. 63–76.

[5] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, “Networking named content,” in *Proc. of ACM CoNext*, Dec. 2009.

[6] L. Qiu, V. Padmanabhan, and G. Voelker, “On the placement of web server replicas,” in *Proc. of IEEE Infocom*, Apr. 2001.

[7] I. D. Baev and R. Rajaraman, “Approximation algorithms for data placement in arbitrary networks,” in *Proc. of ACM Symposium on Discrete Algorithms (SODA)*, Jan. 2001.

[8] B. Chun, K. Chaudhuri, H. Wee, M. Barreno, C. Papadimitriou, and J. Kubiatowicz, “Selfish caching in distributed systems: a game-theoretic approach,” in *Proc. of ACM Symposium on Principles of Distributed Computing (PODC)*, July 2004.

[9] N. Laoutaris, O. Telelis, V. Zissimopoulos, and I. Stavrakakis, “Distributed selfish replication,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 17, no. 12, pp. 1401–1413, 2006.

[10] G. Dán, “Cache-to-cache: Could ISPs cooperate to decrease peer-to-peer content distribution costs?” *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 9, pp. 1469–1482, 2011.

[11] V. Pacifici and G. Dán, “Pure nash equilibria in graphical replication games,” in *Proc. of 23rd International Teletraffic Congress (ITC 2011)*, Sep. 2011.

[12] F. Kuhn, “Weak graph colorings: Distributed algorithms and applications,” in *ACM SPAA*, Aug. 2009.

[13] D. Welsh and M. Powell, “An upper bound for the chromatic number of a graph and its application to timetabling problems,” *The Computer Journal*, vol. 10, no. 1, pp. 85–86, 1967.