Cache Allocation in CDN: An Evolutionary Game Generalized Particle Model

Xiang Feng[1], Francis C.M. Lau[2], and Daqi Gao[1]

[1] East China University of Science and Technology, Shanghai 200237, China
[2] The University of Hong Kong, Hong Kong
xfeng@ecust.edu.cn, fcmlau@cs.hku.hk, gaodaqi@ecust.edu.cn

Abstract. Content distribution networks (CDNs) increasingly have been used to reduce the response times experienced by Internet users through placing surrogates close to the clients. This paper presents an object replacement approach based on an evolutionary game generalized particle model (G-GPM). We first propose a problem model for CDNs. The CDN model is then fit into a gravitational field. The origin servers and surrogates are regarded as two kinds of particles which are located in two force-fields. The cache allocation problem is thus transformed into the kinematics and dynamics of the particles in the annular and the round force-fields. The G-GPM approach is unique in four aspects: 1) direct viewing of individual and overall optimization; 2) parallel computing (lower time complexity); 3) multi-objective solution; and 4) being able to deal with some social interactions behaviors.

Keywords: Content Delivery Networks (CDN), cache resource allocation, evolutionary game generalized particle model (G-GPM), placement algorithm, distributed and parallel algorithm.

1 Introduction

Content delivery networks (CDNs) were developed to overcome performance problems, such as network congestion and server overload, that arise when many users access popular contents simultaneously. CDNs improve end-user performance by caching popular contents on edge servers located close to the users. CDNs help prevent server overload, since the replicated contents can be delivered to users from edge servers. Furthermore, since contents are delivered from the closest edge server and not from the origin server, the request response time is reduced, and so are the probability of packet loss, and the total network resource usage.

The placement/replacement strategy is a key component of any CDN which has a direct and significant impact on the CDN's performance. Recent research in replacement strategies starts to use analytical methods to augment simulation-based evaluation. For instance, in [1], a precise analytical model is developed to evaluate the LRU policy and its variations. The user Web access behavior is modeled using a system of differential equations which explicitly take into

J. Zhou (Ed.): Complex 2009, Part II, LNICST 5, pp. 1226–1237, 2009.

consideration the ages of documents (the time elapsed since these documents were last accessed). The exact expressions for the hit rate and expected network latency can then be derived. The benefits of this kind of analytical work are twofold. First, it provides a theoretical framework to evaluate known policies and can generate technical insights as to why certain policies are effective in practice and outperform others. Second, it can lead to the development of new policies based on the deepened understanding of the structure of the caching problems and the interactions among various components of the caching system.

In this paper, an analytical approach based on the evolutionary game generalized particle model to the content placement problem is proposed.

Our study differs from the previous work in certain aspects, as follows:

1. We map the origin server-surrogate cache resource allocation problem to the kinematics and dynamics of particles in two dual particle fields by analytical mathematical modeling.
2. We embody the game relation between the origin servers and the surrogates in the process of content delivery.
3. We make use of agents (origin servers and surrogates) and their dynamic equations for the evolutionary game generalized particle model to solve content placement problem and overcome some shortcomings of traditional methods in terms of the ability to deal with interactions such as competition, collaboration and unilateral behavior among agents. In previous work, there is no unilateral behavior among the origin servers and among the surrogates being assumed.
4. We embody the dynamic policy and autonomy behavior of the origin servers and the surrogates in the CDN.
5. We consider a market-based price mechanism.
6. We can achieve a high degree of parallelism and scalability. The origin servers and the surrogates can compute and update in parallel their respective policies according to the dynamic environment.

2 CDN Model

The objective of the content distribution problem is to maximize the reduction in average client latency by the use of surrogate caches, while the total cost of service is less than the highest investments the origin servers (publishers) are willing to make for the caching resources. Fig. 1 illustrates the model of the content delivery problem. There are I origin servers on the outer circle, and J surrogates on the inner circle. Origin server nodes and surrogate nodes are evenly distributed at an even radian along the outer circle and the inner circle, respectively. The center of the two concentric circles denotes the clients. Related definitions and expressions in the problem model are explained as follows.

Definitions and Expressions

1. $O_i, S_j, \lambda_i^{j,n}, \lambda_i^j, \alpha_i, q$:

Assume that there are I different origin servers (O_i) and J surrogates (S_j) present in the network. Client requests arrive from N different client LANs. Let

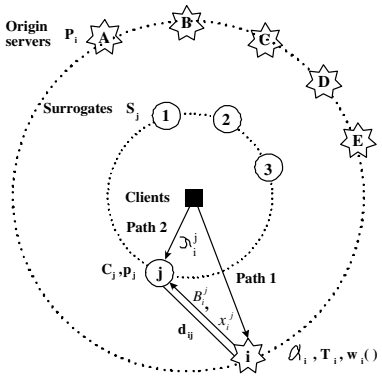

Fig. 1. The CDN model

$\lambda_i^{j,n}$ denote the total request arrival rate from LAN n at surrogate j for the content in the ith origin server. Let $\lambda_i^j = \sum_n \lambda_i^{j,n}$ be the total arrival rate to the surrogate j for the content in origin server i. The clients' interest for the objects of the origin servers is distributed according to the $Zipf$ distribution and α_i is the distribution characteristic of the origin server i ($0 < \alpha_i < 1$) [2]. Let c be the standard constant, then the probability that the client x being interested in the objects of the origin server i is $q(x) = c/h^{\alpha^i}$.

2. $B_i^j, B_i, T_i, C_j, x_i, r$:
Let B_i^j be the investment of the ith origin server in the jth surrogate. Let $B_i = \sum_j B_i^j$ be the total investment of the ith origin server. It is assumed that the information stored in the servers is continuous and can be replicated continuously to a surrogate. Total information available at the origin server i is T_i. Let C_j be the cache space of surrogate j. The origin server replicates its most popular part of the content to the surrogates so that the cache hit probability is maximized. Assuming that x_i units of cache space are allocated to the origin server i, the probability that an incoming client request is satisfied at the surrogate is $r = \int_0^{x_i} q(t)dt = \int_0^{x_i} \frac{c}{t^{\alpha_i}} dt = \int_0^{x_i} \frac{(1-\alpha_i)/t^{1-\alpha_i}}{t^{\alpha_i}} dt = \left(\frac{x_i}{T_i}\right)^{1-\alpha_i}$.

3. p_j, d_{ij}, x_i^j :
Let p_j denote the price of the unit cache space in surrogate j. Let the pricing policy, $\mathrm{P} = (p_1, p_2, \cdots, p_J)$, denote the set of prices for unit cache space of all the surrogates in the network. Let d_{ij} denote the additional average delay that a user request forwarded from surrogate j to the origin server of origin server i will experience. Let x_i^j ($x_i = \sum_{j=1}^{J} x_i^j$) be the cache space allocated to origin server i in surrogate j. If the ith origin server's investment in the jth surrogate is B_i^j ,

then the total cache space allocated to the content of origin server i in surrogate j is $x_i^j = \frac{B_i^j}{p_j}$.

4. $w_i(d_{ij}), m_j, \beta_i^j, U_i$:
Let $w_i(d_{ij})$ denotes the benefit received by origin server i when the delay of service of user request is reduced by d_{ij} units. Thus, we consider a generalized version of the content delivery problem, where each origin server receives varying degree of benefit from the use of surrogates as reflected by the benefit function $w_i()$. Assume that $w_i(d)$ is a concave function. The average reduction in the user delay or, equivalently, the average net benefit that origin server i generates by B_i^j investment in surrogate j, is $m_j = \lambda_i^j w_i(d_{ij}) \left(\frac{x_i^j}{T_i}\right)^{1-\alpha_i}$. Define $\beta_i^j = \lambda_i^j w_i(d_{ij}) \Big/ (T_i)^{1-\alpha_i}$ as the gain factor for origin server i from surrogate j. The utility function, i.e., the total additional average benefit U_i of origin server i is $U_i = \sum_{j=1}^{J} \beta_i^j (x_i^j)^{1-\alpha_i}$.

3 Evolution of the G-GPM

With our CDN model above, we can now examine the evolutionary model that can mathematically describe the G-GPM for the content placement problem. We can identify two types of optimization problems in the CDN model: origin server's revenue maximization and surrogate's revenue maximization.

The theory of evolution is a dynamical theory. The evolutionary dynamics will drive the G-GPM to the equilibrium state.

Definition 1. *Cache and prices dynamic equations of G-GPM are defined, respectively, by*

$$x(t+1) = x(t) + \Delta x(t) \tag{1}$$

$$p(t+1) = p(t) + \Delta p(t) \tag{2}$$

The two dynamic equations are interpreted as "G-GPM evolution" by fictitious agents (origin server particles and surrogate particles), which set the allocated cache and prices in motion until an equilibrium is reached.

For fictitious agents—origin server particles (O) and surrogate particles (S), there are three factors related to allocated cache (x) and prices (p).

- personal utility (u);
- minimal personal utility (to realize max-min fair allocation and to increase the overall utility) (F);
- interaction among particles (to realize the main object) (I).

According to "differential equation theory", the variable's increment to make it maximum is equal to the sum of negative items from related factors differentiating the variable. So we have the following definitions.

Definition 2. *The increments of cache and prices are defined, respectively, by*

$$\Delta x \approx \frac{dx}{dt} = -\lambda_1 \frac{\partial u_O}{\partial x} - \lambda_2 \frac{\partial F_O}{\partial x} - \lambda_3 \frac{\partial I_O}{\partial x} \tag{3}$$

$$\Delta p \approx \frac{dp}{dt} = -\gamma_1 \frac{\partial u_S}{\partial p} - \gamma_2 \frac{\partial F_S}{\partial p} - \gamma_3 \frac{\partial I_S}{\partial p} \tag{4}$$

$\lambda_1, \lambda_2, \lambda_3, \gamma_1, \gamma_2, \gamma_3$ are coefficients.

Definition 3. *Three factor functions for origin server particles and surrogate particles are defined, respectively, by*

$$u_{Oi} = \exp(-U_i) = \exp\left(-\sum_{j=1}^{J} \beta_i^j (x_i^j)^{1-\alpha_i}\right) \tag{5}$$

$$F_O = k^2 \mathrm{In} \sum_{i=1}^{I} \exp[u_{Oi}^2 / 2k^2] \tag{6}$$

$$I_O = a_1 \sum_{i=1}^{I} \left\| \sum_{j=1}^{J} x_i^j p_j - B_i \right\| + a_2 \sum_{j=1}^{J} \left\| \sum_{i=1}^{I} x_i^j + C_j \right\| \tag{7}$$

$$u_{Sj} = \exp\left(-\sum_{i=1}^{I} x_i^j \cdot p_j\right) \tag{8}$$

$$F_S = k'^2 \mathrm{In} \sum_{i=1}^{J} \exp[u_{Sj}^2 / 2k'^2] \tag{9}$$

$$I_S = b_1 \left\| \sum_{i=1}^{I} x_i^j - C_j \right\| \tag{10}$$

k, a_1, a_2, k', b_1 are coefficients.

In order to realize that the smaller Equations 5 and 8 the better, the definitions of Equations 5 and 8 are changed to some extent.

Definition 4. *The utility functions of origin servers and surrogates are defined, respectively, by*

$$U_{Oi} = 1 - u_{Oi} = 1 - \exp(-U_i) = 1 - \exp\left(-\sum_{j=1}^{J} \beta_i^j (x_i^j)^{1-\alpha_i}\right) \tag{11}$$

$$U_{Sj} = 1 - u_{Sj} = 1 - \exp\left(-\sum_{i=1}^{I} x_i^j \cdot p_j\right) \tag{12}$$

We can therefore obtain the iteration velocity of origin server particle and surrogate particle by the equations, respectively.

$$v_i = du_{Oi}/dt = \frac{\partial u_{Oi}}{\partial x_i^j}\frac{dx_i^j}{dt} \tag{13}$$

$$v_j' = du_{Sj}/dt = \frac{\partial u_{Sj}}{\partial p_j}\frac{dp_j}{dt} \tag{14}$$

Theorem 1. *If k is very small, the increase of F_O will cause an increase of the origin servers' minimal utility. (Likewise, If k' is very small, an increase of F_S will cause an increase of the surrogates' minimal utility.)*

Proof. Supposing that

$M(t) = \min_i U_{Oi}^2(t) = \max_i [u_{Oi}^2(t)]$,

Because

$M(t) = \max_i u_{Oi}^2(t) \le \sum_{i=1}^{I} u_{Oi}^2(t) \le I \cdot \max_i u_{Oi}^2(t) = I \cdot M(t)$,

we then have

$$\left[e^{\frac{M(t)}{2k^2}}\right]^{2k^2} \le \left[\sum_{i=1}^{I} e^{\frac{u_{Oi}^2(t)}{2k^2}}\right]^{2k^2} \le \left[I \cdot e^{\frac{M(t)}{2k^2}}\right]^{2k^2}.$$

Simultaneously taking the logarithm of each side of the equation above will lead to

$$M(t) \ge 2k^2 \mathrm{In} \sum_{i=1}^{I} e^{\frac{u_{Oi}^2(t)}{2k^2}} \ge M(t) + 2k^2 \mathrm{In} I,$$

$$2k^2 \mathrm{In} \sum_{i=1}^{I} e^{\frac{u_{Oi}^2(t)}{2k^2}} \le M(t) \le 2k^2 \mathrm{In} \sum_{i=1}^{I} e^{\frac{u_{Oi}^2(t)}{2k^2}} - 2k^2 \mathrm{In} I,$$

$$2F_O(t) \ge \max_i u_{Oi}(t) \ge 2F_O(t) - 2k^2 \mathrm{In} I.$$

$$2F_O(t) \ge \min_i U_{Oi}^2(t) \ge 2F_O(t) - 2k^2 \mathrm{In} I.$$

Since I is the number of origin servers, $2k^2 \mathrm{In} I$ is constant.

It turns out that$F_O(t)$ at time t represents the minimum among $U_{Oi}(t)$ obtained by the origin server O_i, namely, the minimum of the personal profit obtained by an origin server at time t. Hence decreasing $F_O(t)$ implies an increase of the minimal utility of the origin server.

Definition 5. *(Max-min Fairness). [3] A feasible allocation of cache space x is max-min fair if and only if an increase of any cache space within the domain of feasible allocations must be at the cost of a decrease of some already smaller cache space. Formally, for any other feasible allocation y_i^j, if $y_i^j > x_i^j$ then there must exist some i' such that $x_{i'}^j \le x_i^j$ and $y_{i'}^j < x_{i'}^j$.*

Definition 6. *The set of unit cache space prices of all the surrogates in the network except the jth one is defined by*

$$\mathrm{P}^{-j} = (p_1, p_2, \cdots, p_{j-1}, p_{j+1}, \cdots, p_J) \tag{15}$$

Theorem 2. *Surrogate revenue* $r_i(p_j) = \sum_j x_i^j(p_j)p_j$ *is maximized under a given fixed pricing policy* P^{-j}*, when* p_j *satisfies* $\sum_{i=1}^{I} x_i^j(p_j) = C_j$.

It means that the surrogate revenue is maximized when the total origin server demand is equal to the surrogate cache space. A consequence of this theorem is that $r_i(p_j)$ achieves an interior maximum. This is due to the fact that $r_i(p_j)$ tends to zero both as p_j goes to zero and as p_j goes to infinity. Whether the function $r_i(p_j)$ is maximized at the price that completely allocates surrogate cache or at a higher price will depend on whether C_j is above or below this extremum. However, in many practical cases, the cache capacity is much lower than the total information available in the network. Thus, the surrogate will be able to sell all of its capacity without setting a price that is approaching zero.

Theorem 3. *The behavior of the origin server* O_i *that is related to the second term of the Eq. (3) will always bring about the increase of the minimal profit obtained by an origin server, and the increment of the minimal profit is directly proportional to the coefficient vector* λ_2*. (Likewise, The behavior of the surrogate* S_j *that is related to the second term of the Eq. (4) will always bring about the increase of the minimal profit obtained by a surrogate, and the increment of the minimal profit is directly proportional to the coefficient vector* γ_2*.)*

Theorem 4. *The behavior of the origin server* O_i *that is related to the first term of the Eq. (3) will always result in the increase of the personal profit of origin server* O_i*, and the increment of its personal profit is related to coefficient vectors* λ_1*. (Likewise, The behavior of the surrogate* S_j *that is related to the first term of the Eq. (4) will always result in the increase of the personal profit of surrogate* S_j*, and the increment of its personal profit is related to coefficient vectors* γ_1*.*

Theorem 5. *The behavior of the origin server* O_i *that is related to the third term of the Eq. (3) will decrease the potential interaction energy function* I_O*, with the intensity of the decrease being proportional to coefficient vector* λ_3*. (Likewise, The behavior of the surrogate* S_j *that is related to the third term of the Eq. (4) will decrease the potential interaction energy function* I_S*, with the intensity of the decrease being proportional to coefficient vector* γ_3*.*

Theorem 6. *(Max-min fair allocation). Max-min fair allocation can be obtained by the mathematical model for the CDN cache optimization problem as defined by Eqs. (1–12).*

The proofs of Theorems 2–6 are omitted.

Convergence Analysis

Until now, we have discussed the optimal strategies of the origin servers and the surrogates given that the system is in steady state. However, we have not discussed whether such a steady state exists or not. Note that, when a surrogate reevaluates its pricing policy according to the pricing policies of the rival

surrogates, the rest of the surrogates will do the same. For each different pricing policy, the origin servers' optimal investments will be different as well.

Theorem 7. *The origin server-surrogate distribution game has at least one Nash Equilibrium solution.*

The profit for surrogate j is $r_j(\mathrm{p}) - c_j$, where c_j is the cost of the surrogate j's cache. We assume that there exists some price $\hat{p}_j$ at which demand for the cache space of surrogate j is zero, regardless of the prices of other surrogates. As an example, we have the revenue of surrogate j increases until a certain price p_j^*, beyond which it starts to decrease again. Then, we may limit p_j to the interval $[0, \hat{p}_j]$ and still be able to cover the complete range of payoff function. Thus, p_j is convex and compact.

The profit of each surrogate is bounded from below by zero and, since the total investment of all origin servers is limited, the profit can never exceed $\sum_i B_i - c_j$. Based on the Cournot behavioral assumption [4], the surrogate takes its rivals actions as given, supposing that they will remain constant, and chooses its own best course of action accordingly. The payoff function under this assumption is given by $r_j(\mathrm{p})$. There is a unique equilibrium prices for the origin server-surrogate distribution game; that is, there is a unique best reply function, $R_j(\mathrm{p}) = \arg\max_{p_j} \{r_j(\mathrm{p})\}$ for surrogate j, which is also continuous. This result shows that the game has a Nash equilibrium.

When there is a unique equilibrium for the origin server-surrogate distribution game, the equilibrium prices solve the origin server optimization problem for all origin servers $i = 1, \cdots, I$, i.e., the solution is globally Pareto optimal. Assume that each surrogate uses Eq. (4) to update its price. The best price for surrogate j given the pricing policy P^{-j} is calculated from $\sum_i x_i^j = C_j$. At the equilibrium, this condition is satisfied as well. Furthermore, the origin servers use Eq. (3) to calculate x_i^j, which guarantees local optimality of the solution and the feasibility of the two conditions, $\sum_{j=1}^{J} x_i^j p_j \leq B_i$ and $\sum_{i=1}^{I} x_i^j \leq C_j$. Uniqueness of the equilibrium guarantees that the feasible locally optimum solution is also the global optimum. Under these conditions, the outcome of the origin server-surrogate game is a solution to the origin server optimization problem.

Pareto optimality is a relevant criterion in a multiobjective problem setting. At the Pareto optimum, one can find no other feasible solution that can increase some objectives while not hurting at least another objective.

4 The Parallel G-GPM Algorithm

The results given in the previous sections suggest that we may use a distributed and parallel evolutionary game generalized particle model approach to solve the two-stage origin server-surrogate cache resource placement problem. We consider the following algorithm for this purpose.

G-GPM algorithm:

1. Surrogates announce in parallel a set of initial prices $\mathrm{P}^{(0)} = (p_1^{(0)}, p_2^{(0)}, \cdots, p_J^{(0)})$; the origin servers use Eq. (3) to determine their resource (cache) demands, $x_i^{j(0)}$, $i = \overline{1,I}, j = \overline{1,J}$, according to these initial prices as well as the request rates and the observed delays from the surrogates; initialize the desired thresholds z_O, z_S for checking the state equilibrium of origin server particles and surrogate particles.

2. At iteration k, calculate in parallel the iteration speeds, $v_i^{(k)}$, $v'^{(k)}_j$, according to Eqs. (13)(14) of every origin server particle and surrogate particle, respectively.

3. If any $v_i^{(k)} \leq z_O$ and any $v'^{(k)}_j \leq z_S$, $i = \overline{1,I}, j = \overline{1,J}$, all the origin server particles and surrogate particles reach their equilibrium state, then finish with success; otherwise, by Eq. (4), each surrogate j updates in parallel its price, $\mathrm{P}^{(k)} = (p_1^{(k)}, p_2^{(k)}, \cdots, p_J^{(k)})$, according to the origin server demands; by Eq. (3), each origin server i calculates in parallel its optimal cache demand for surrogate j, $x_i^{j(k)}$, according to the surrogate prices; then go to *step* 2.

5 Physical Meaning of G-GPM

G-GPM places emphasis on

- direct view of individual and overall optimizations;
- parallel computing (lower time complexity);
- multi-objective solution;
- being able to deal with some social interactions behaviors.

The mathematical model of G-GPM proposed mention above has its physical meaning.

From now on suppose that the origin server particle O_i is so arranged in the big ring origin server particle-field that it can only move along the radial line that makes the angle of $i \cdot 2\pi/I$ radian to the horizontal coordinate, where I is the number of origin servers in CDN. And assume that the distance between the inside circumference to the outside circumference in the origin server particle field is normalized to 1. u_{Oi} can denote the radial distances from the origin server particle O_i to the outside circumference of the gravitational field in the origin server particle field. As mentioned in the above section, the smaller the profit acquired by the origin server particle O_i, the larger the value of u_{Oi} will be.

We suppose that the surrogate particle S_j is so arranged in the small round surrogate particle-field that it can only move along the radial line that makes the angle of $j \cdot 2\pi/J$ radian, where J is the number of surrogates in CDN. And assume that the radius of the circumference of gravitational field in the surrogate particle field is normalized to 1. u_{Sj} can denote the radial distances from the surrogate particle S_j to the circumference of the gravitational field in the vertical surrogate particle field. As mentioned in the above section, the smaller the profit acquired by the surrogate particle S_j, the larger the value of u_{Sj} will be.

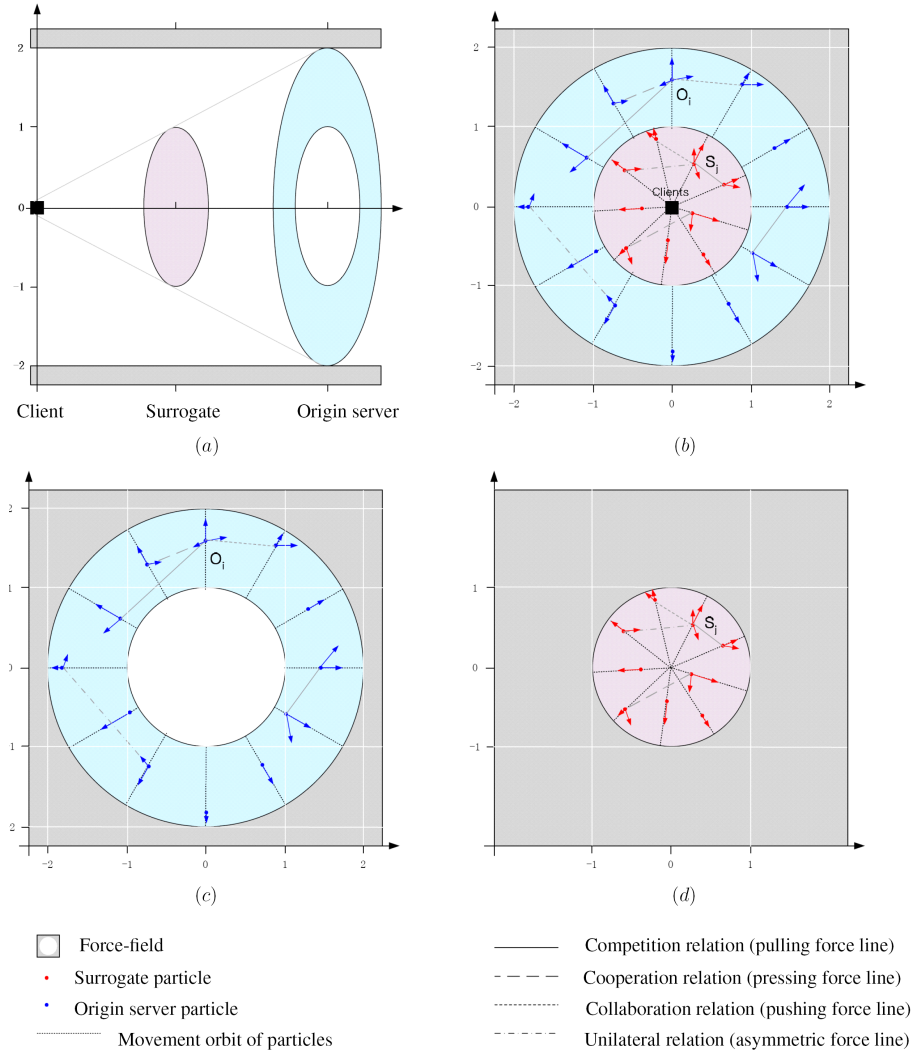

Fig. 2. The architecture of an evolutionary game generalized particle model for cache allocation of distribution subsystem in a CDN with 12 origin servers and 9 surrogate: (a) a 3D global view of the G-GPM where the rectangles represent a gravitational field; (b) a left view of the G-GPM, where two force-fields are parallel to each other and are surrounded by the same gravitational field; (c) the origin server force-field; (d) the surrogate force-field, where the different types of lines represent different types of interaction forces

We can therefore obtain the radial velocity of the origin server particles and the surrogate particles along their radial orbit to the circumference of the particle field by the iteration speeds v_i, v'_j.

Thus, the evolutionary game generalized particle model is composed of two dual particle fields, which is illustrated in Fig. 2. In this figure, small round and

big annular particle fields correspond to the origin servers and the surrogates. There is no direct connection between the round and the annular particle fields; however, they will exert their influence on each other through the origin servers' optimal caching strategy and the surrogates' optimal pricing strategy. Thus, the two particle fields constitute a reciprocal dual game particle field. A game occurs not only among the same kind of particles in each particle field but also between the two particle fields.

The big annular particle field (origin server particle field) is composed of particles and vectors, which represent the origin servers and their social interactions, respectively. All the origin server particles are surrounded by the outside and inside circumferences in the big ring, and spread at random around the ring by equal radial angles. Each origin server particle is exerted simultaneously by the gravitational field of the outside circumference, and by the forces that represent the interactions with other origin server particles, where movements along the radial orbit from the inside circumference to the outside circumference are allowed in this ring. F_O can denote the energy function of the gravitational field. I_O can denote the interaction energy function. The distance from an origin server particle to the inside circumference is proportional to the personal profit acquired by the origin server under the current situation in the CDN. If an origin server particle has social interaction with respect to another origin server particle, then there is a linking line between them, whose force strength and force-displacement property depend upon the interaction occurring between them. Note that the force can be asymmetric for both sides. As for the small round particle-field corresponding to the surrogates, the organization is similar. The distance from a surrogate particle to the center of the circumference in the small round plane is proportional to the personal profit acquired by the surrogate. Like origin servers, the surrogates also interact in various manners with each other in order to decrease their personal overhead as much as possible.

It turns out that the behavior of both an origin server and a surrogate to pursue the maximal personal benefit is embodied in the origin server particle in the big ring particle field, and the surrogate particle in the small round particle field would try to move along the radial to their corresponding circumferences as near as possible. We should indicate that we can make use of an appropriate force to easily represent any kind of social interactions based on mind-reading among the particles (origin servers and surrogates), including the competition, cooperation, enticement, deception, avoidance, exploitation, coalition, reciprocation, interference, collaboration, habituation, compromise, preference, etc. And a force with an asymmetric property between two particles can describe the unilateral interactions such as enticement, avoidance, exploitation, deception, etc. Moreover, we can easily design a force that has the time-varying non-linear force-change property to outline the more complicated social interaction among the autonomous particles if necessary. In the simplest case, for example, a pulling force can represent the competition between two particles; and a pushing force can describe the cooperation between two particles, with the force strength reflecting the interaction intensity. The larger the resultant forces along the radial

towards the circumference, the faster the movement along the radial towards the circumference of the particle. When the resultant force on a particle is equal to zero, the particle will stop moving, being at an equilibrium.

6 Conclusion

In this paper, we propose an analytical approach based on a novel evolutionary game generalized particle model to the content placement problem. The approach maps the origin server-surrogate cache resource placement problem to the movement of particles in two dual particle fields by mathematical modeling. All particles move according to certain rules defined by mathematical model until reaching a stable state; then the solution to the origin server-surrogate cache resource placement problem is obtained by anti-mapping the stable state.

The evolutionary game generalized particle model approach has embodied what might be the mysteries of the particle field: movement and interaction of particles. Moreover, the functors of generation and annihilation, and the relation of nontrivial commutation and anti-commutation between these functors will be addressed in our future work.

To conclude, we give a summary of the key terms that characterize our analytical evolutionary game generalized particle model placement approach.

- The CDN model and the G-GPM architecture
- The movement and interaction of particles
- Interesting game
- Distributed and parallel performance

Content delivery networks accelerate the provision of information in the Internet. The novel analytical evolutionary game generalized particle model proposed in this paper can help solve a key problem of content placement in the Internet—the two-stage origin server-surrogate cache resource placement problem.

Acknowledgements. This project is supported by the Hong Kong General Research Fund under Grant No. HKU 713708E, the National Science Foundation of China (NSFC) under Grant No. 60575027, the High-Tech Development Program of China (863) under Grant No. 2006AA10Z315, and the Specialized Research Fund for the Doctoral Program of Higher Education under Grant No. 20060251013.

References

1. Mookerjee, V.S., Tan, Y.: Analysis of a least recently used cache management policy for web browsers. Oper. Res. 50, 345–357 (2002)
2. Zipf, G.K.: Human Behavior and the Principle of Least Effort. Addison-Wesley, Cambridge (1949)
3. Jean-yves, L.B.: Rate adaptation, Congestion Control and Fairness: A Tutorial (2005)
4. Alexander, G.J.: Advanced Microeconomic Theory. Prentice-Hall, Englewood Cliffs (1991)