Decentralized Replica Selection in Cloud Based on Bandwidth Capacity and Energy Consumption

by

Bo Li

A Thesis Proposal Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer Science

Supervised by

Ivona Bezakova
Department of Computer Science
B. Thomas Golisano College of Computing and Information Sciences
Rochester Institute of Technology
Rochester, New York
Feb 2012

Approved By:

Ivona Bezakova
Professor, Department of Computer Science
Primary Adviser

Alan Kaminsky
Professor, Department of Computer Science
Reader
Contents

Abstract 1

1 Introduction 2
   1.1 Replica Selection .................................................. 2
   1.2 Distributed Optimization ......................................... 2

2 Proposal Overview and Related Work 4
   2.1 Proposal Overview .................................................... 4
   2.2 Related Work .......................................................... 4

3 Methodology 6
   3.1 System Architecture and Assumptions ............................. 6
   3.2 Algorithms ............................................................ 7
      3.2.1 Consensus Based Projected Subgradient Method .............. 7
      3.2.2 Lagrangian Dual Decomposition ............................... 8
      3.2.3 Examples .......................................................... 8

4 Proposed Work 10
   4.1 Algorithms and Server Side Implementation ..................... 10
   4.2 Trace-based Experiment ............................................. 10

5 Deliverables 11

References 12
Abstract

Replica selection for the cloud services has been studied as an important inter-data center problem considering performance, cost, and capability on both the client side and the data center side. And solving this problem in a distributed manner becomes almost necessary considering the scalability to more and more clients, reliability, and security of the system. We present a new approach of solving the replica selection problem in the cloud by considering energy consumption of data intensive applications and the bandwidth capacity of each data center. This is motivated by the fact that energy consumption at the data centers constitutes a very significant part of the operating costs. A distributed system architecture is proposed to be used for implementing such service. We formulate the replica selection problem as a quadratic optimization problem with linear equality and inequality constraints by assuming a relationship between traffic load and energy consumption. In order to solve the global optimization problem in a distributed manner, we present two different algorithms and implement them in the real system. These two algorithms are based on the consensus based projected subgradient method and the Lagrangian dual decomposition method. By using a trace-based experiment to mimic clients’ requests of downloading files from replicas, we will evaluate the algorithm performance and how its optimization affects the replicas in terms of energy consumption. The conclusion will be summarized based on the analysis of the experimental results.
1 Introduction

In my thesis, I propose to solve the replica selection problem for the data intensive applications in the cloud services. The purpose of this work is to reduce energy consumption cost, which takes significant part of the cost in data center, by splitting client’s request to one replica into portions to all the replicas. In this section, I am going to introduce briefly the replica selection problem in the cloud and the approaches of solving global optimization problem in parallel. The work can be summarized as two major parts. The first part of my thesis’ work is to formulate replica selection as an optimization problem to minimize energy consumption cost of all the clients’ requests subject to the bandwidth constraint of each replica. The second part of my thesis’ work is to solve the global optimization problem in a distributed manner rather than using a simply centralized coordinator. The distributed system architecture brings the system some important features. It solves the issues of single point of failure, scalability, security, etc. Also, the performance of the decentralized system architecture could be competitive or even better than centralized system with increasing of clients’ requests. Another feature of the proposed system is that it is user transparent that, the client can send request to any replica and get the solution of how to split its request into all the replicas.

1.1 Replica Selection

In the cloud, some services are replicated in geographically different data centers. The client who wants to use such services needs to specify one or more data centers to connect with. The problems of telling clients how to choose from those data centers are called replica selection. In my thesis, I propose to implement control to the selection of replicas for data intensive applications in order to minimize the clients’ total energy cost. The selection also needs to be feasible in terms of bandwidth capacity in each replica. The replica selection is formulated as an optimization problem of minimizing energy cost subject to the bandwidth capacity constraint. In order to calculate the energy consumption corresponding to the clients’ requests, I consider the energy consumption of data center coming from two major parts. One is from server nodes in the data center, and the other is from Ethernet devices such as switch. Based on the relationship between energy consumption of these two parts and the clients’ requested traffic load, I can formulate the total energy consumption as a function of traffic load from client.

1.2 Distributed Optimization

In order to solve the global optimization problem of replica selection, we can choose either a centralized coordinator to implement such control, such as DNS service and HTTP redirection, or a distributed system playing the same functionality as the centralized architecture. Given the advantages of distributed architecture, I proposed to solve the replica selection by using a distributed system where the global optimization problem can be solved in a distributed manner. In terms of distributed optimization, every distributed agent collaborates
with the other agents by network communication and solving local optimization problem so that an optimal solution of the global optimization problem can be achieved. In order to solve the optimization problem in a distributed manner, several approaches may be appropriate. They are consensus based distributed gradient methods, the Lagrangian dual decomposition method, etc.

In Section 2, a formal statement of the problem in my thesis is briefly described with introduction of related work. In Section 3, the distributed system architecture is presented with assumptions of the problem. With these assumptions, the replica selection problem in my thesis is formulated as an quadratic optimization problem with linear equality and inequality constraints. Then the algorithms based on two different methods are implemented for solving the global optimization in parallel under the distributed system architecture. Section 4 and 5 are the proposed work and proposed deliverables of my thesis.
2 Proposal Overview and Related Work

2.1 Proposal Overview

In my thesis, I propose to solve the replica selection problem by minimizing the total energy cost of all the clients' requests subject to the bandwidth capacity of each replica. The services focus on the data intensive applications, such as online video download. Once the client wants to see an online video (e.g., from Youtube), it can either choose randomly downloading from one of the replicas working as video server, or choose one or more replicas for such services following certain rules. For the problem solving in my thesis, the rule is minimizing the total energy cost of replicas within the capacity of replica bandwidth. The distributed system architecture of such service is necessary in terms of the scalability, reliability, security, etc. In order to solve the global optimization problem in parallel, I am going to propose a simple distributed architecture to realize such service control, and implement the algorithms based on the Lagrangian dual decomposition method and consensus based projected subgradient method. Once either of the algorithms is implemented, the system can provide the clients with the service that tell which replicas to use and how to divide its requested amount into different replicas in order to minimize energy cost.

2.2 Related Work

Some existing work towards replica selection aims at minimizing network distance [1], latency [2]. Some other work [3] relies the selection on the machine readable description of replica capabilities, such as features, performance, cost etc. Access frequency is also one of the important factors that determines the selection of replicas in some recent research. [4] Aside from these factors, energy consumption is also considered as an important issue based on the fact that energy cost takes a significant part of the cost of data center. According to the EPA report in 2007, data centers consumed 61 billion kwh which is 1.5% of the total U.S. electricity in 2006. The significant cost of data centers would certainly increase the cost of cloud services for clients. It is necessary for the clients to considering the cost of energy for data-intensive applications in order to reduce the payment to the cloud service provider. Some research work [5] has been carried out in the area of reducing the cost of energy for data centers. The scope of such work include server level [6, 7], data center level [8, 9, 10], and cloud level [1, 11].

For data intensive applications in the cloud, the traffic load affects the energy consumption of data center. Based on the current research work, the proportional relationship between load and energy consumption is a major influential factor for data center energy efficiency [12]. However, the majority of network devices in the data center, comparing with single server, are far from being energy proportional [13, 14]. In order to minimize energy cost of data centers, Liu [15] takes workload and number of active servers in each data center into consideration. However, it assumes that the single server energy consumption does not depend on the load. It does not follow the principle of energy management of data center. Rao’s work [11] minimizes the electricity cost by guaranteeing quality of service with the con-
sideration of time variant electricity price. Some other work does not consider traffic load, which is a very important feature for data intensive applications, in their model neither.

Usually the optimization problem for replica selection is nonlinear as the problem in my thesis. Some distributed methods for solving nonlinear optimization have been used in the recent published work. Liu’s work [15] presents Gauss-Seidel method and gradient projection method to solve the global optimization problem in a distributed manner. Wendell [1] proposes a Lagrangian dual decomposition method working well in a decentralized system to solve the global optimization problem. Also Newton method is a good candidate for optimization in parallel. A consensus-based projected subgradient method [16] can also solve the global optimization problem by the collaboration of distributed agents.
3 Methodology

3.1 System Architecture and Assumptions

In order to solve the replica selection problem in a distributed manner, we define all the data centers as distributed nodes. Each replica contains a group of individual nodes. $C$ represents a collection of all the active clients. The system architecture is shown in Figure 1.

Each data center, which we call replica, can accept requests from clients in $C$. By collaborating with other replicas at runtime, the distributed method can inform the clients in $C$ how to split the requests’ load among all the replicas. The objective is to minimize the total energy consumption of replicas by applying scalable distributed algorithm to adjust the distribution of the workload. Some important notations are summarized in Table 1.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$</td>
<td>Set of all clients</td>
</tr>
<tr>
<td>$N$</td>
<td>Set of replicas</td>
</tr>
<tr>
<td>$P_{cn}$</td>
<td>Traffic load mapped from client $c$ to replica $n$</td>
</tr>
<tr>
<td>$B_n$</td>
<td>Bandwidth capacity on replica $n$</td>
</tr>
<tr>
<td>$R_c$</td>
<td>Traffic load of the request from client $c$</td>
</tr>
</tbody>
</table>

Based on the problem description above, we consider the energy consumption coming from two major parts in data center, server nodes and ethernet devices. A cubic relationship between traffic load and energy consumption corresponds to the popular energy reduction methods like DVS and DFS [17]. Such relationship has been validated through Ethernet interface cards [18]. It is feasible to apply the proportional relationship at nodal level between

---

Figure 1: Distributed Services in Replicas
traffic load and energy consumption, and cubic relationship at network device level. Therefore, we can get a combination of linear and cubic relationship between energy consumption and network traffic load in our model. The total energy consumption of all the replicas is:

\[ E_g = \sum_n (\alpha_n \sum_c P_{cn} + \beta_n (\sum_c P_{cn})^3) \]

where \( \alpha_n \) and \( \beta_n \) are the weight scalars for the energy consumption of computer devices and network devices in replica \( n \). Based on this assumption, we can formulate the optimization problem as:

\[
\begin{align*}
\text{minimize} & \quad E_g = \sum_n E_n \\
\text{subject to} & \quad f(P) = \sum_c P_{cn} - B_n \leq 0, \forall n \in N \\
& \quad h(P) = \sum_n P_{cn} - R_c = 0, \forall c \in C
\end{align*}
\]

where \( E_n = \alpha_n \sum_c P_{cn} + \beta_n (\sum_c P_{cn})^3 \) is the energy consumption in replica \( n \), \( f(P) \) is the bandwidth capacity constraints of all the replicas, \( h(P) \) is the constraints of each of all the clients’ requests. The problem turns to be with a cubic objective function and several linear equality and inequality constraints. Since it is a convex optimization problem, we can use both consensus based gradient algorithm and dual decomposition method to solve it in a distributed manner for better scalability and reliability.

### 3.2 Algorithms

#### 3.2.1 Consensus Based Projected Subgradient Method

The consensus based projected subgradient method is discussed in Nedic’s paper [16]. For each distributed replica \( n \) in our problem, the solution can be achieved iteratively by applying the following equation:

\[
P^n_{X_n}(k + 1) = P_{X_n}[\sum_{j=1}^{N} a^j_n \cdot P^j(k) - \alpha_k \cdot g_n(k)]^+
\]

where \( X_n \) is the constraint set of replica \( n \), \( a^j_n \) is the weights of all the replicas, \( \alpha_k > 0 \) is the step size, and \( g_n(k) \) is the subgradient on its objective function \( E_n \). Since the objective function of our problem is twice differentiable, we could use gradient instead of subgradient as \( g_n(k) \). The symbol \( P_{X_n}[\cdot]^+ \) denotes the operation of projection. We have:

\[
P_{X_n}[P^*]^+ = \arg\min_{P \in X_n} \|P^* - P\|
\]
3.2.2 Lagrangian Dual Decomposition

The Lagrangian dual decomposition method is used to remove the coupling variables in the original problem and make all the variables private to the subproblems. By using Lagrangian Multiplier \((\lambda, v)\), we can get the Lagrangian \(L(P, \lambda, v)\):

\[
L(P, \lambda, v) = \sum_{n=1}^{N} E_{n} + \sum_{i=1}^{N} \lambda_{i} \cdot f_{i}(P) + \sum_{i=1}^{C} v_{i} \cdot h_{i}(P)
\]

\(\lambda\) and \(v\) can be updated by solving the Lagrangian dual problem:

\[
\text{maximize } g(\lambda, v) = \inf_{P} L(P, \lambda, v) \\
\text{subject to } \lambda \geq 0
\]

Gradient method can be used to solve the linear dual problem.

3.2.3 Examples

I am going to use a simple example problem to illustrate how these two algorithms work for solving optimization problems. The problem is stated as below:

\[
\begin{aligned}
\text{minimize } & \quad f(X) = f_{1}(x_{1}) + f_{2}(x_{2}) \\
\text{subject to } & \quad h(x_{1}, x_{2}) \leq 0 \\
& \quad x_{1} \in C_{1} \\
& \quad x_{2} \in C_{2}
\end{aligned}
\]

In order to solve this problem using consensus based projected subgradient method, problem can be divided into two sub problems \(f_{1}\) and \(f_{2}\) sharing the same constraint set \(C\). We express the solution of them as \(X_{1}\{x_{1}, x_{2}\}\) and \(X_{2}\{x_{1}, x_{2}\}\). \(k\) is the iteration number of this method.

So for the sub problem \(f_{1}\), \(V^{k+1} = a_{1} \cdot X_{1}^{k} + a_{2} \cdot X_{2}^{k} - \alpha_{k} \cdot g_{k}\) where \(a_{1} + a_{2} = 1\), \(\alpha_{k}\) is the step size which is usually set to be \(1/k\), and \(g_{k}\) is the gradient of \(f_{1}\) at the point \(a_{1} \cdot X_{1}^{k} + a_{2} \cdot X_{2}^{k}\).

If \(V^{k+1}\) is feasible, it is the solution of iteration \(k + 1\).

\[X_{1}^{k+1} = V^{k+1}\]

If \(V^{k+1}\) is not feasible, the solution of iteration \(k + 1\) can be achieved by projecting \(V^{k+1}\) to the constraint set to find the its nearest solution in terms of Euclidean distance. To sum these two conditions together, we can achieve the rule of projection as:

\[X_{1}^{k+1} = \arg\min_{X \in C} \|V^{k+1} - X\|\]

Meanwhile, sub problem \(f_{2}\) also does the same thing. After certain iterations, \(X_{1}\) and \(X_{2}\) can converge to the same solution through the consensus procedure.
The Lagrangian dual decomposition method can also be used to solve \( f(X) \). We can formulate the Lagrangian function \( L(X, \lambda) \) by using the Lagrangian multiplier \( \lambda \):

\[
L(X, \lambda) = f_1(x_1) + f_2(x_2) + \lambda \cdot (h_1(x_1) + h_2(x_2))
\]

So the problem is constitute with two sub problems:

\[
\begin{align*}
& \text{minimize } f_1(x_1) + \lambda_1 \cdot h_1(x_1) \\
& \text{subject to } x_1 \in C_1
\end{align*}
\]

and

\[
\begin{align*}
& \text{minimize } f_2(x_2) + \lambda_2 \cdot h_2(x_2) \\
& \text{subject to } x_2 \in C_2
\end{align*}
\]

The value of \( \lambda \) can be determined by solving

\[
\inf_X (L(X, \lambda))
\]

which is a simple linear programming problem.
4 Proposed Work

4.1 Algorithms and Server Side Implementation

- In my thesis, I will implement two algorithms for the replica selection problem. They will be based on the consensus based projected subgradient method and the Lagrangian dual decomposition method.

- I am going to use 3 computer nodes to work as three replicas listening to the requests of client. Each node implements the algorithm and collaborates with other 2 nodes to solve the optimization problem in parallel.

- In order to solve the global optimization problem, each node needs to solve its local optimization problem. In order to do so, a high performance software for large scale LP, QP, SOCP and MIP called mosek [19] will be used with its Java interface.

- Also, the server side application is proposed to be developed under Java 1.6 runtime environment.

4.2 Trace-based Experiment

After implementing the server side program, I will develop the client side program to realize the trace-based experiment for testing the algorithm performance and energy consumption in the real systems. Clients’ requests will be generated automatically and send to different replicas once in a while. This procedure will last for a period of about 12 hours.

- The performance will be compared by considering the convergence rate of these two algorithms. After accepting the client’s request, the faster the system reaches the stable status, the better the algorithm is.

- The energy consumption will be calculated at the replica side for both the system that implements the energy optimization and the system that just randomly assigns the replica to the client. We expect to see the difference before and after the optimization in terms of energy consumption. PowerPack [20] nodes from Virginia Polytechnic Institute and State University. will be used as the tool for collecting energy consumption data.
5 Deliverables

There will be several major deliverables from my thesis. The distributed system architecture I proposed in the thesis is a general model for implementing distributed cloud services. In additional to this, the algorithms and the implementation work make the replica selection problem solved under the proposed system architecture. The evaluation of the approaches relies much on the trace-based experiment. The design of experiment and its affect to the performance result will be discussed in the thesis. Finally, a conclusion of solving replica selection problem in a distributed manner and the appropriate method used to solve optimization problem in parallel will be summarized at the end of the thesis.
References


