CONCURRENCY AND COMPUTATION: PRACTICE AND EXPERIENCE
Concurrency Computat.: Pract. Exper. (2010)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/cpe.1610

A two-tier framework for transmission-cost minimization of high-performance communication applications

Chia-Cheng Hu[1], Chin-Feng Lai[2], Yueh-Min Huang[2] and Han-Chieh Chao[3,4,*,†]

[1]*Department of Information Management, Naval Academy, Kaohsiung, Taiwan*
[2]*Department of Engineering Science, National Cheng-Kung University, Tainan, Taiwan*
[3]*Institute of Computer Science & Information Engineering and Department of Electronic Engineering, National Ilan University, I-Lan, Taiwan*
[4]*Department of Electrical Engineering, National Dong Hwa University, Hualien, Taiwan*

SUMMARY

In two-tier high-performance networks (HPNs), some facilities are constructed to form a powerful super-computing environment, and to alleviate server load. Then, the applications are provided by them in co-operated, parallel and distributed manners. A proper way to select facilities is crucial to the performance of two-tier HPNs The problem of selecting facilities can be regarded as a kind of the facility location problem, which is to determine an optimal subset of facilities that will be open to serve users. The traditional facility location problem aims to minimize the incurred costs between the users/servers and their assigned facilities. In two-tier HPNs, the incurred costs can be regarded as the transmission costs, e.g. transmission latency, bandwidth overhead. We observe that most of the packets are transmitted among the facilities for application servicing and framework maintaining. In this paper, we address the problem of selecting facilities in two-tier HPNs by minimizing the transmission costs from servers to users by passing through the selected facilities. Our problem is different from the traditional facility location problem, which only considers the transmission costs between the users/servers and their assigned facilities. In our problem, the transmission costs between the selected facilities are further considered. The problem is formulated as a 0/1 integer non-linear programming (0/1 INLP) and 0/1 integer linear programming (0/1 ILP). Further, a simple heuristic algorithm is proposed for obtaining a feasible solution when the network sizes increase, since solving INLPs and ILPs for large-scale problems takes long time. Copyright © 2010 John Wiley & Sons, Ltd.

Received 7 August 2009; Revised 24 November 2009; Accepted 19 April 2010

KEY WORDS: high-performance network; two-tier; facility location problem; integer programming

*Correspondence to: Han-Chieh Chao, Institute of Computer Science & Information Engineering and Department of Electronic Engineering, National Ilan University, I-Lan, Taiwan.
†E-mail: hcc@mail.niu.edu.tw

1. INTRODUCTION

A high-performance network (HPN) is a communication network that supports a large variety of applications and that is scalable. In order to support many applications, the networks must be able to transfer traffic from servers to users at high speed and with low delay, and to allocate resources in ways that match the application requirements [1]. On the other hand, with the increase in Internet users, a two-tier framework is adopted in large-scale HPNs in order to alleviate server load. In a two-tier framework, some facilities are constructed in the network to form a powerful supercomputing environment, and the applications are provided by them in co-operated, parallel and distributed manners. Without high-performance communication among them, the applications cannot exploit effectively.

Examples of such applications are cache/proxy placement, mirror server placement, wireless sensor networks, wireless broadband networks, and large-scale wireless *ad hoc* networks. Caches, proxies and mirror servers have been used for web clients to improve network and system performance by saving network bandwidth, reducing delays to clients, and alleviating server load [2–8].

In wireless sensor networks, the main challenge is to find an efficient way by gathering data from the sensors to a control station using as little energy as possible. In order to save energy, the sensors would not transmit data directly, but rather relay it from one sensor to its neighbors until it reaches the control station. Assume a network enhanced with a few dozen high-powered relay stations capable of communicating directly with the control station. The relay stations could dramatically increase the energy efficiency of the sensors. Therefore, to select an optimal subset of the relay stations to relay data is a facility location problem [9].

In wireless broadband networks, the coverage of wireless high-speed broadband service is expected to expand dramatically in the near future. The wireless broadband networks consist of two-layer devices, e.g. subscriber stations and base stations. Traffic demands from end users are aggregated at a set of subscriber stations. Then, the aggregated traffic demands at subscriber stations will be satisfied by a set of base stations. In order to satisfy the traffic demands, how to deploy the base stations is an important issue [10, 11].

Another network system can be found in large-scale wireless *ad hoc* networks. *Ad hoc* networks require no fixed framework or centralized administration, and hosts must communicate one another via packet radios in a collaborated manner. Recent advances in wireless telecommunication technologies and portable computing are continuing to drive the revolution toward large-scale networks and flexible new generation e-services. To avoid the inefficiency of transmission in large-scale *ad hoc* networks, two-tier frameworks are adopted in [12–22] in which some of hosts are selected as backbone hosts to manage the transmission Since most of the packets are initiated and processed by backbone hosts, a proper way to select backbone hosts is crucial in a large-scale *ad hoc* network

Consider a company providing services to users. Every time the company opens a new facility to serve users, the facility incurs a cost. The facility location problem is to determine an optimal subset of facilities that will be open to serve users. The set of the open facilities should minimize the incurred costs. In traditional facility location problem, an optimal subset of facilities is determined for minimizing the costs incurred by the determined facilities, and various users are assigned to the determined facilities so as to satisfy user demands. The incurred costs are treated as the distance between the users and their assigned facilities. The objective is to minimize the average distance so that a user will travel short distance to reach a facility.

The problem of constructing two-tier frameworks in [2–22] is a kind of facility location problem. Most of them regard the incurred costs as the transmission costs (e.g. transmission latency,

DOI: 10.1002/cpe

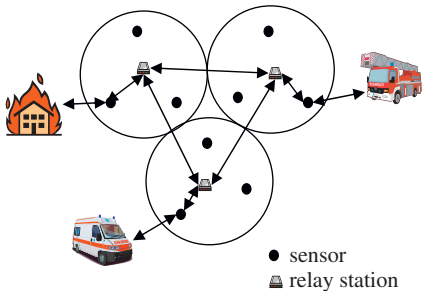

Figure 1. A two-level sensor network example in a disaster management application.

bandwidth overhead) between the users/servers and their assigned facilities, and they aim to minimize the average transmission costs. We observe that the facilities (e.g. Caches, proxies, mirror servers, relay stations, base stations and backbone hosts) determined in [2–22] are not only responsible for providing service to the clients, but also for transmitting a great deal of packets among the determined facilities to satisfy the user demands and to maintain the two-tier framework.

Figure 1 depicts a two-tier sensor network example in a disaster management application. Once a sensor detects a fire event, it transmits a packet to the sensor equipped by a fire truck. The relay stations assigned by the two sensors are responsible to relay the transmitted packet. An important issue is to minimize the transmission delay from the source sensor to the destination sensor by passing through the assigned relay stations.

In a two-tier framework, the process of path determination works in two stages: selecting facilities and determining paths. In the first stage, the most previous works adopt the strategy of selecting those facilities with low transmission cost to users. In the second stage, they minimize the transmission cost between the two selected facilities by selecting the path with the minimal transmission cost from the multiple paths, which are available for the two facilities. However, they may suffer from selecting the path with high transmission cost between the two selected facilities since the transmission cost among facilities is not considered in the first stage.

In this paper, our problem is a variant of facility location problem, in which we select facilities and assign users/servers to the selected facilities in order to minimize the average transmission cost between server–user pairs by passing through the selected facilities. Our problem is different from the traditional facility location problem in [2–22] where only the transmission costs between the users/servers and their attached facilities are taken into consideration. In our problem, the transmission costs between the assigned facilities are further taken into consideration in the first stage of path determination, since most of the packets are transmitted among the facilities for application servicing and framework maintaining.

In the next section, previous works are first reviewed Our problem is defined in Section 3. In Section 4, the problem is formulated as a 0/1 integer non-linear programming (0/1 INLP) and 0/1 integer linear programming (0/1 ILP). Then, simulations are implemented to compare the performance of the 0/1 ILP with the 0/1 INLP solving by a branch-and-bound solver. The simulation results show that the 0/1 ILP is superior to the 0/1 INLP in execution time and cost optimization. However, when the network sizes increase, not only solving INLPs but also solving ILPs for large-scale problems takes long time. Hence in Section 5, a simple heuristic algorithm is proposed for obtaining a feasible solution when the network sizes are large. Simulation is also

DOI: 10.1002/cpe