行政院國家科學委員會專題研究計畫　成果報告

結合第三代行動通訊與無線網路之開放式多媒體服務平台

計畫類別：個別型計畫
計畫編號：NSC94-2213-E-029-026-
執行期間：94 年 08 月 01 日至 95 年 07 月 31 日
執行單位：東海大學資訊工程與科學系

計畫主持人：張文貴
共同主持人：呂芳懌

報告類型：精簡報告
報告附件：出席國際會議研究心得報告及發表論文
處理方式：本計畫可公開查詢

中　華　民　國 95 年 10 月 23 日

行政院國家科學委員會專題研究計畫成果報告

結合第三代行動通訊與無線網路之開放式多媒體服務平台
A Wireless Grid Service Platform Using SIP and Agents

計畫編號：NSC 94-2213-E-029-026-

執行期限：94 年 8 月 1 日至 95 年 7 月 31 日

主持人：張文貴 教授　　東海大學資訊工程與科學系

共同主持人：呂芳懌 副教授　　東海大學資訊工程與科學系

計畫參與人員：陳又新 賴瑞鈞 東海大學資訊工程與科學系

Abstract

Recently, Grid computing infrastructures have been developed, but most of them connect nodes with hard-wired links. Seldom adopt wireless technology and mobile devices. However, if we can integrate a Grid system with wireless network devices, techniques and resources to form a wireless Grid infrastructure to provide different kinds of terminal devices, Grid applications will soon be popularly and widely deployed in different domains to serve many more people in the world. In this article, we propose a service platform, named Wireless Grid-based mobile agent system (WGMAS) which integrates mobile agent techniques and wireless environment with a Grid system, to provide a wireless accessible mobile agent system to users.

Keywords: *Grid, wireless, mobile agent system, wireless Grid platform*

1. Introduction

Recently, many commercial and industry wireless communication standards have been developed. By the transmission distance and application domain, they can be classified into several types, including: Wireless Wide Area Network (WWAN), Wireless Metro Area Network (WMAN), Wireless Local Area Network (WLAN) and Wireless Personal Area Network (WPAN), among which WLANs have been widely constructed around us. Users worldwide can access the Internet through them. This occurrence provides us with a convenient network environment to enrich our daily life.

Besides, Grid computing infrastructures have been quickly developed, but most of them are used in scientific computing or to provide huge storage to their users only, not very helpful in serving ordinary people. Also, they connect nodes with hard-wired links. Seldom adopt wireless technology and mobile devices. In fact, if we can integrate wireless network devices, techniques and resources with a Grid system to form a wireless accessible infrastructure to provide different kinds of terminal devices, Grid will soon be popularly and widely deployed in different domains and applications to serve many more people in the world. This occurrence will make Grid become much more practical and useful.

Moreover, wireless networks and access points (APs) have been extensively built. Peripheral prices are also significantly reducing. They are now entering users' life, education and entertainment speedily. Although the computing capability and storage capacity of mobile devices have become more powerful and larger respectively, day by day, the limited resolution of screen size, unacceptable power consumption and less multi-tasking functionality make it difficult replace PC, laptop and work station overall in the near future. However, integrating these wireless resources with Grid to bridge users and a Grid infrastructure will make mobile computing be pervaded as other network devices.

Additionally, a mobile agent (MA) is a software program of which the main purpose is moving among nodes within a network to perform its predefined task. It can process a given task based on its designed logic, and then return the result to users. Autonomy and mobility are its key features. Numerous mobile agent systems (MASs) have been developed and applied to many domains, such as information retrieval and percolation, e-commerce, mobile communication and distributed computing.

Basically, communication is one of the key issues in an MAS. An agent has to communicate with others before it can properly cooperate with them to finish their given tasks. Therefore, a reliable communication protocol is required. However, “reliable” often induces a sophisticate validation process or mechanism which often degrades the performance of an MAS, and even makes MAs unable to accomplish the desired mission.

In this article, we propose a distributed sendbox scheme, which is developed on a Grid platform, to deal with problems depicted above. Each Grid node in the platform has a sendbox to temporarily buffer and send messages for agents, and some access points to support wireless access to Grid resources. Our scheme can efficiently forward messages to the receiver agent R even R migrating frequently.

The rest of this article is organized as follows. Section 2 introduces the relevant background and related work. Section 3 describes our system architecture. The service platform and the experimental results are presented in sections 4 and 5, respectively. Section 6 concludes this article.

2. Related Work

Wireless Grid is a newly popular concept in recent years. McKnight et al. [1] proposed a resource sharing model for small and nomadic devices with which they simultaneously recorded a series of mono sounds and then mixed them into stereo sounds. However, this approach did not address communication stability.

Srinivasan [2] proposed a new suppose: Wireless devices were primary integration devices. All peripherals could use short distance wireless communication (e.g., Bluetooth, Ultra Wide Band (UWB), Zigbee) to form a PLAN to share resources, such as monitor, disk, or input devices. However, the UWB is still in draft and Bluetooth is widespread but too slow to transfer a huge amount of information.

MEFS [3] proposed Chasing Message Register and Over-speed Agent Waiting mechanism. A message was considered as a chasing message after C times of failed delivery. When the migration frequency of an agent was over a threshold V, it stopped moving and checked if any chasing message was stored in the Chasing Message Register. Some problems on the other hand were raised, e.g. the configuration of system parameter V.

Resending-based protocol [4] could provide reliable communication by using sliding-window for error control. When some messages were lost, S resent them. After several trails, S requested the receiver’s new location from the server and sent messages again. However, there was no upper limit of resending times. For frequently migrating agents, this protocol suffered from communication overhead.

Other schemes such as Mailbox-based scheme [7], ARP [8], forwarding-based MAS [5-8] and resending-based approaches [3, 4], all had their own problems and drawbacks to be solved.

3. System Architecture

In this article, we propose a Wireless Grid System, named Wireless Grid-based Mobile Agent System

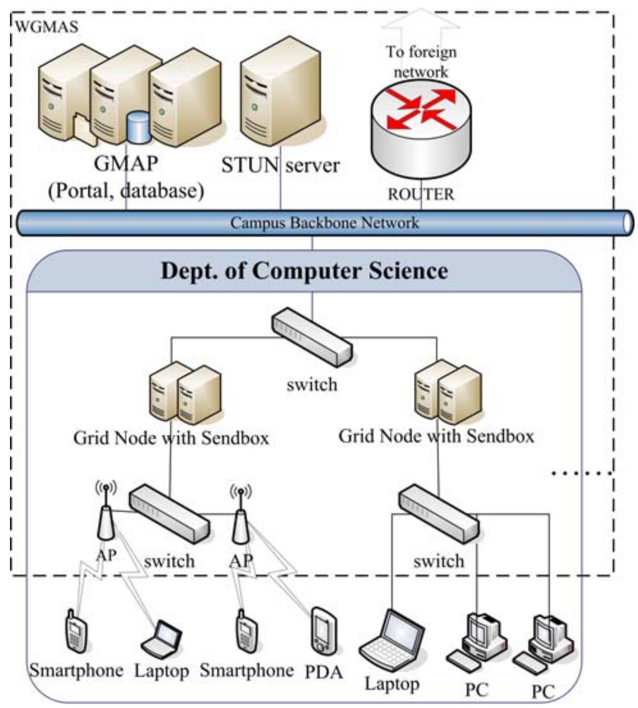

Fig. 1 WGMAS architecture

(WGMAS), which is a hard-wired Grid integrated with wireless network and a Session Initiate Protocol [9] based (SIP-based) communication mechanism to establish a wireless Grid environment. Users can access Grid resources through wired or wireless networks. Figure 1 illustrates the WGMAS architecture in which Grid Mobile Agent Proxy (GMAP) is the management center taking charge of agent login, registration, dispatch, and recall. STUN (Simple Traversal of UDP through NATs), as a NAT Traversal solution of SIP, makes packets routable when users are now inside a Network Address Translation (NAT) environment. Sendbox, as one of a storage subsystem of WGMAS, stores messages Ms, that are unable to be safely delivered to their destinations and forwards Ms when receivers are ready. It can be used to solve the message chasing problem.

Users can establish a wired or wireless connection and register their SIP to GMAP through Grid Mobile Agent (GMA) which is an MA implemented on Grid. After that, they can check node status and request Grid services and/or other SIP-compatible services. All the accessed contents are sent to users in synchronously or asynchronously way.

To date, most Grid systems are equipped with high speed wired networks to perform their communication and computation. If a communication between a Grid node and a mobile device is required, its communication stability, cost and session continuity should be the major concerns. The goals are that the "online" time should be reduced and messages should be successfully sent and received.

3.1. GMAP

In our scheme, before moving to another site S, a GMA must un-register to GMAP. After arriving at S, it must register to GMAP which can then keep track the GMA's position and status in order to minimize the probability of communication failure. The registration information includes the GMA's ID, S's IP, arrival time, and task (such as, querying node status or requesting services offered).

Before sending packets to GMA B, GMA A should query B's location from GMAP. The query information consists of B's ID, query code (such as the code for querying B's location L), and query time. The replied message from GMAP comprises B's ID, B's state (such as, waiting or moving) and query answer (e.g., B's location L). GMAP is also responsible for commanding GMAs, e.g., order a GMA to access some types of documents or stop accessing then when an error occurs.

3.2. GMAP Database

The GMAP database is to save GMA's information, such as: GMA's SIP, status (online or offline), location (staying at some Grid node or moving), port number (must be provided while the device is under a NAT environment), registration time, device type (laptop, personal digital assist, smart phone, etc). The database like others also provides query interface so that users can conveniently access the contents.

3.3. Inter-agent Communication Protocols

In our previous work [11], we proposed the distributed sendbox scheme. If reciver agent is staying on a node, the sendbox could send messages to it directly, Direct mode communication so-called. Otherwise, Forward communication mode is used. Generally, direct mode handles all the communication between agents. If the receiver agent is moving all the time, the message can not be safely delivered, WGMAS switches to forward mode to asynchronously forward messages.

3.4. Sendbox

A Grid node is equipped with a sendbox to store messages Ms to be re-sent. Accompanied with M should be sender's SIP, receiver's SIP, Time Of Arrival (TOA), Current Life Time (CLT), Max Life Time (MLT), message priority, and contents. If CLT reaches the MLT set by the sender or system administrator, Sendbox Monitor (SM), which is a rule-based sub-system that manages sandbox, will delete M to avoid "Sendbox fully filled" caused by receiver being offline for a long time. "If the free space is less than 30%, SM will delete the oldest message first based on the CLT field, otherwise, delete the messages of whose CLT > 3hr." is a rule example.

3.5. Wired/Wireless Connectivity

To achieve a high usability of GMAP and the goal of pervasive computing, both GMAP's wired and wireless connections must be reliably maintained. We propose an architecture that uses Grid nodes as local gateways to improve network connectivity and hence accessibility.

3.5.1. Wireless Connection. We connect those wireless APs with no routing function to their nearby Grid nodes (GNs), which provide the DHCP (Dynamic Host Configuration Protocol) service, through direct wired links. Using this type of APs, we can not only significantly reduce system construction cost, but also effectively manage multiple APs' IP assignment without modifying their firmwares. The resource usage of a Grid node can be markedly maximized as well.

3.5.2. Wired Connection. Similar to wireless connection, when a user U connects his/her terminal T to the Internet by directly wiring T to a switch or hub, T will then receives a temporary IP from DHCP server so as to communicate with others.

3.6. Traversal through NAT

WGMAS uses the STUN [10] protocol to make packets "routable" especially when they are tunneled by an NAT. The process is as follows. GMA A inside a NAT sends a packet P to the STUN server, which is outside the

NAT, to query P's "source port #". The server replies with the port number N to A, which will send N to GMA B. Thus, with N, GMA B can directly communicate with A as in a normal situation.

4. Service Platform

We propose an Open Grid Service Platform (OGSP) to intermediate the web services, which are installed on GN, to enable users to search or publish their service descriptions which are indexes and can be queried by users. Users in WGMAS are classified into ordinary users and Grid Service Providers (GSPs). All services that GSPs provide must be indexed by OGSP before they can be searched or used by users. Desiring to register to GMAP, users can deploy GMAs to connect to it through wired/wireless network. In this platform, information or data will be delivered with direct mode or forward mode mentioned before to raise the communication quality and form a pervasive computing environment.

4.1. Components

The OGSP, as shown in Fig. 2, consists of Service Repository (SR), Subscription Repository (SuR), Service Content Index (SCI) and Temporary Information Holder (TIH). SR stores the service contents, advertisements or announcements. SuR keeps the information concerning users' subscription. SCI indexes open service providers and service contents to speed up users browsing. TIH is implemented on sandbox to temporary hold service information.

4.1.1. SR and SCI. SR is equipped on every GN as the service container. After users upload their contents to SR, OGSP indexes the content automatically and save the indexes to SCI to briefly introduce services provided. Besides, users can also upload their indexes to SCI and keep service contents in their servers.

4.1.2. SuR and TIH. Users can subscribe their services to OGSP periodically or occasionally. Subscription information includes providers' name, service name, update frequency (e.g. daily or weekly) and subscription period (e.g. one week or one month). After requesting a service, a user may go offline temporarily before receiving the service information Inf. The OGSP will store Inf into the TIH of the underlying node. When users go online again, TIH delivers Inf to users directly.

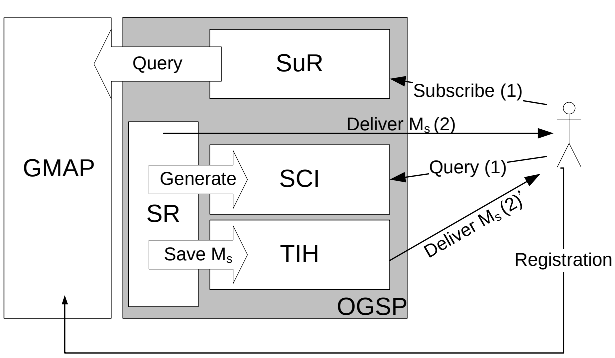

Fig. 2 Components of OGSP

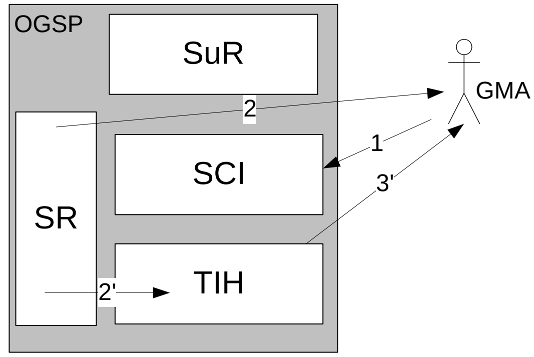

Fig. 3 IQS

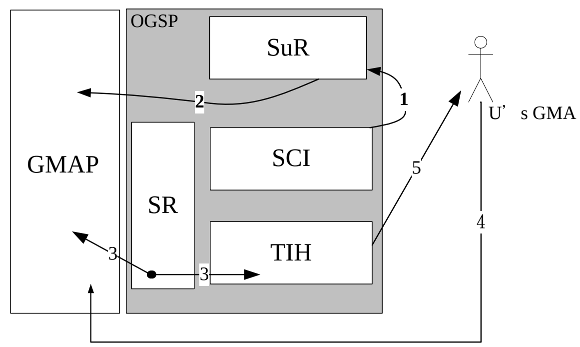

Fig. 4 SNS

4.2. Service types

To improve the quality of service of WGMAS, we propose two main access mechanisms: Instant Query Service (IQS) and Subscription and Notify Service (SNS). The former is an interactive mode. The later delivers service information indirectly.

4.2.1. Instant Query Service (IQS). IQS is the basic user service mode of this system. As shown in Fig. 3. Wanting to query Grid node GN_1's status or request services, a user dispatches a GMA to GN_1 first. SR sends the contents to user devices also through a GMA. IQS is implemented by direct mode. However, if the user device becomes offline suddenly, the system will buffer the remaining messages in TIH and switch to SNS mode.

4.2.2. Subscription-and-Notify Service (SNS). OGSP allows a mobile user U to subscribe his/her required information Inf since it is inconvenient for a user to periodically or frequently self-browse a system to access new information (e.g., checking a node status) and/or new resources (e.g., searching a new article). All the operations, including subscription, notification, and communication, are performed by GMAs which first search the SCI, collect newly updated or published Inf, and check the SuR to determine which users subscribed Inf. After that, GMAs send Inf to the TIH, or deliver the Inf to U if U is still online. After sending a subscription, U may move or go offline immediately. GMAs perform the same operations but notify U with a short message service through his/her cell phone. When U goes online again, two GMAs will be generated. One notifies the TIH to resend Inf. The other on U's device receives Inf and shows it on the screen. Then, U will never lose his/her desired information even U is busy or moving frequently, or the receiving device is temporality offline.

5. Experiments

In the following, we analyze the two inter-agent communication modes and present their communication performance.

5.1. Utilization

There are two major factors affecting the two modes' utilization probabilities, *UP(Dir)* and *UP(Fwd)*, including, migration cost μ and the period R stays on a node, T_{stay}. If an agent moves frequently, it spends much more time to migrate. *UP(Fwd)* will be higher. While network is congested or receiver R moves more frequently, an agent takes longer migration time μ to migrate and to negotiate with GMAP. This occurrence also increases *UP(Fwd)*. *UP(Dir)* and *UP(Fwd)* of an agent are defined as:

$$UP(Dir) = \frac{\sum_{i=0}^{k} T_{stay}}{\sum_{i=0}^{k} (T_{stay} + \mu)} \quad \cdots(1)$$

$$UP(Fwd) = 1 - UP(Dir) \quad \cdots(2)$$

where k is the number of nodes an agent pass through along its journey, the term μ includes times of unregistering to GMAP, mobile code transportation, java object serialization and registering to GMAP.

5.2. Communication Cost

The components of communication costs are shown as follows:

T_{inv}: the duration from the time point when an agent sends an Invite to the time point when it receives either Moved Temporarily or Temporarily Unavailable

T_{msg}: the duration from the time point when an agent sends Message to receiver to the time point when it receivers an OK response

T_{not}: the time needed to deliver Notify to an agent from GMAP

T_{sub}: the duration from the time point when an agent sends Subscribe to GMAP to the time point when it receivers an OK

T_{mig}: the duration from the time point when an agent leaves a GN to the time point when it reaches another GN

The parameters include computation time and network transmission time between platforms.

The communication cost of the direct model is:

$$Cost(Dir) = T_{inv} + T_{msg} \quad \cdots(3)$$

In the forward mode, the communication cost is:

$$Cost(Fwd) = T_{inv} + T_{sub} + T_{mig}(R) + T_{not} + T_{msg} \quad \cdots(4)$$

The average cost is:

$$Avg(\cos t) = T_{inv} + UP(Dir) * T_{msg}(Dir) + UP(Fwd)(T_{sub} + T_{mig} + T_{not} + T_{msg}(Fwd)) \quad \cdots(5)$$

5.3. Testing Results

The WGMAS prototype and our communication protocols are developed with Java. The test-bed of the WGMAS system is implemented with: 1. our campus LAN and 5 Grid nodes as the mobile agent network; 2. 100MBit/s Ethernet interface to interconnect Grid nodes and 802.11g 54 Mbps WLAN AP to serve the wireless network connection; 3. IBM R40 laptop as the mobile device. The agent size is 20 Kb and the size of M_{SR} is 10 KB.

R roams in the WGMAS network, visiting nodes randomly, staying at a node for a given time T_{stay}, and then migrating to other node. S sends M_{SR} to R periodically, once every l_{msg} time interval.

In the environment, μ ranges from 50~130ms, avg(μ)=90ms. The experimented results on different l_{msg}, T_{stay} and the corresponding utilization rates of direct mode *UP(Dir)* are shown in Fig. 7 from which we concluded 1. the experimental results do follow the expected tendency; 2. the more frequently receivers move, the higher *UP(Fwd)* (1-*UP(Dir)*) will be. *UP(Dir)* will be lower relatively; 3. l_{msg} does not significantly affect the utilization rate.

Some experiments concerning the communication time are performed. The results are shown in Figs. 8 and 9. The communication time of direct mode varies between 14~41ms, where as that of forward mode ranges between

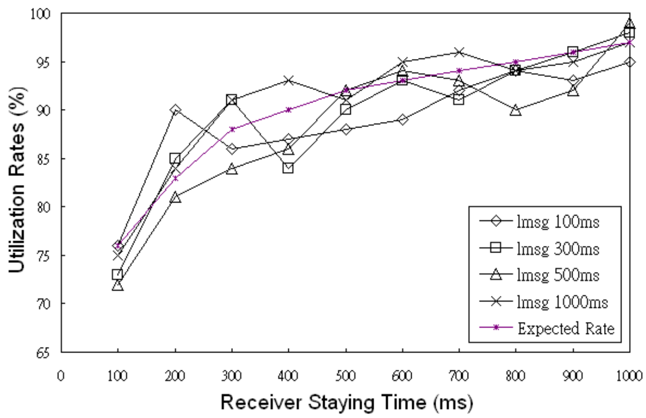

Fig. 7 Direct mode utilization rates

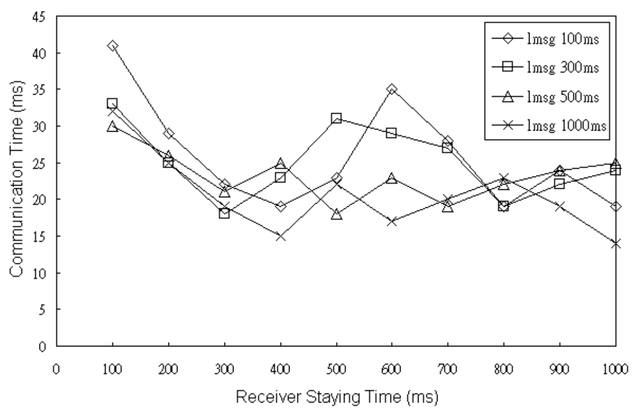

Fig. 8 Communication time of Direct mode

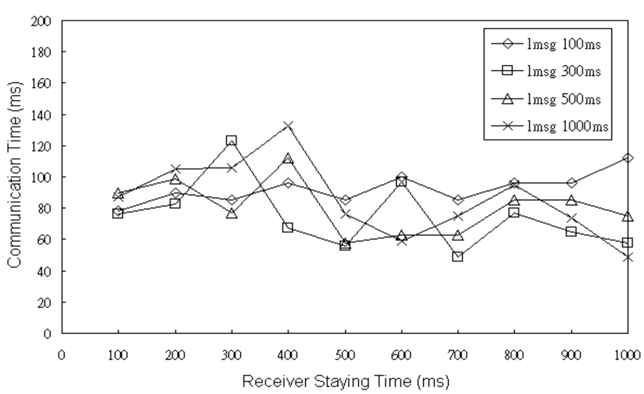

Fig. 9 Communication time of Forward mode

49~133ms. We can see, in the forward mode, the sendbox spends more time for buffering and sending messages.

6. Conclusion and Future Work

In this article, we integrate a distributed sendbox scheme with Grid platform which delivers agent messages based on SIP to solve communication failure and message chasing issues. While receiver agents are stationary, direct mode is used to handle inter-agent communications. Forward mode is invoked while recipients are unregistered, and messages are forwarded indirectly and distributively by sendboxes. The two modes could eventually delivery all messages to their receivers. We also propose an architecture to make the Gird node wirelessly accessible to maximize the resource utilization for each node. Moreover, we develop a service platform, which integrates agents, services, brokers and resources, to make this proposed environment more robust and reliable.

Communication performance and utilization probability are the two factors we want to be prospected. We will compare the performance of our SIP-based mechanism with others existing protocols, and port various feasible applications, especially those of mobile computing domain, to WGMAS and OGSP. Furthermore, we are planning to integrate the sensors into our mechanism, to make our service platform become “location sensible” and “location awardable”.

References

[1] S.H. Srinivasan, “Pervasive wireless grid architecture,” In Proc. of the International Conference on Wireless On-demand Network Systems and Services, 2005, pp. 83 – 88.

[2] L.W. McKnight, J. Howison and S. Bradner, “Guest Editors' Introduction: Wireless Grids--Distributed Resource Sharing by Mobile, Nomadic, and Fixed Devices,” IEEE Internet Computing, vol. 8, 2004, pp. 24 – 31.

[3] J.Y. Zhou, Z.Y. Jia and D.X. Chen, “Designing Reliable Communication Protocols for Mobile Agents,” In Proc. of the International Conference on Distributed Computing Systems Workshops, 2003, pp. 484-487.

[4] M. Ranganathan, M. Beddnarek and D. Montgomery, “A Reliable Message Delivery Protocol for Mobile Agents,” In Proc. of the Joint International Symposium ASA/MA’2000.

[5] ObjectSpace Inc., “Objectspace voyager”, http://www.objectspace.com/voyager, 1997.

[6] J. Desbiens, M. Lavoie and F. Renaud, “Communication and Tracking Infrastructure of a Mobile Agent System,” In Proc. of the International Conference on System Science, vol.7, 1998, pp. 54-63.

[7] X.Y. Feng, J.N. Cao, et al., “An Efficient Mailbox-Based Algorithm for Message Delivery in Mobile

Agent Systems,”G. P. Picco (Ed.), Mobile Agents, LNCS 2240, Spintger-Verlag, 2001, pp. 135-151.
[8] J.N. Cao, X.Y. Feng, et al., “Design of Adaptive and Reliable Mobile Agent Communication Protocols,” In Proc. of International Conference on Distributed Computing Systems, 2002, pp. 471-472.
[9] Internet Engineering Task Force (IETF),”RFC3261 (SIP: Session Initiation Protocol),” 1999, http://www.cs.columbia.edu/sip/drafts.html
[10] Internet Engineering Task Force (IETF), “RFC 3489 (STUN: Simple Traversal of UDP through NATs)”, 2003, http://www.faqs.org/rfcs/rfc3489.html.
[11] Hsiao-Hong Tsai, Fang-Yie Leu and Wen-Kui Chang, ”Distributed SendBox Scheme for Mobile Agent Communication,” In Proc. of International Conference on Mobile Business, 2005, pp. 545-550.