CHALLENGES OF MOBILE AD-HOC GRIDS AND THEIR APPLICATIONS IN E-HEALTHCARE

Zhuoqun Li, Lingfen Sun and Emmanuel C. Ifeachor
University of Plymouth, UK

ABSTRACT

The growing availability of networked mobile devices has created a vast collective potential of unexploited resources. Grid computing with its model of coordinated resource sharing may provide a way to utilize such resources that are normally distributed throughout a mobile ad-hoc network. In this paper, we explore the challenging operation of Grid computing in mobile ad-hoc networks which are characterized by their unreliability of topologies and heterogeneity of mobile nodes. By integrating Grid functionalities with ad-hoc networking protocols, we envisage a lightweight architecture of mobile ad hoc Grids as the synergy between the resource sharing potential of Grid computing and the flexible, ubiquitous availability of mobile ad-hoc networks. We also discuss the general challenges in implementing Grid functionalities (e.g. service discovery, job scheduling and Quality of Service (QoS) provisioning) in the mobile environment and the specific issues arise from realistic application scenarios, i.e. the e-healthcare emergency.

Keywords: Mobile Ad-hoc Networks, Grid Computing, E-Healthcare

1 INTRODUCTION

Advances in wireless mobile networking technologies have engendered a new communication paradigm, the mobile ad-hoc network, which has yielded devices with a high degree of heterogeneity. These mobile devices range from relatively powerful computing systems carried by a vehicle, to very tiny, low-power sensors that can be implanted in the human body. Although they may know little about the identities and capabilities of each other, a group of mobile devices are able to organize a highly dynamic and infrastructure-less ad-hoc network, in which nodes can communicate in a hop-by-hop manner.

Grid computing [1] [2] is an important developing computing initiative that involves the aggregation of network-connected computers to form a large-scale, distributed system for coordinated problem solving and resource sharing. By spreading computing workload across this distributed system of computers, Grid users can take advantage of enormous computational, storage, and bandwidth resources that would otherwise only be available within traditional multiprocessor supercomputers.

The resource sharing potential of Grid computing opens up the possibilities of new applications that involve the integration of Grid technology with new areas such as mobile ad-hoc networks. The availability of wirelessly connected mobile devices has grown considerably in recent years, creating an enormous collective unexploited potential for resource utilization. It is a natural idea to integrate the resource aggregation model of Grid with mobile ad-hoc networks, so as to build a mobile ad-hoc Grid platform that can be instantly constructed anytime, anywhere. Having been constructed from a group of mobile devices, an ad-hoc grid would allow the networked devices to accomplish a specific mission that maybe beyond an individual's computing or communication capacity. Examples of applications of mobile ad-hoc Grids can be disaster management, wildfire fighting, and e-healthcare emergency, etc.

However, it is not trivial for the conventional Grid Computing architecture to operate in the mobile ad-hoc network, in which the Grid computers have limited local resources, the network topology is unstable and changes over time. Although current architecture of Grid Computing has not addressed these problems extensively, issues related to mobile Grid computing have been addressed and corresponding Grid architectures proposed in recent work [3] [4] [5] [6]. Nevertheless, there is little discussion about the challenging problem of how to make the Grid components to function or to guarantee the Quality of Service (QoS) requirements of Grid applications in the mobile ad-hoc environments.

In this survey, we explore the challenges in providing Grid computing in the dynamic, infrastructure-less mobile ad-hoc networks. By integrating Grid functionalities with the ad-hoc networking protocols to increase efficiency and adaptability, we outline a lightweight architecture of mobile ad hoc Grids as the synergy between the resource sharing potential of Grid computing and the flexible, ubiquitous availability of mobile ad-hoc networks. We also discuss the general challenges in implementing Grid functionalities (e.g. service discovery, job scheduling and QoS provisioning) in the mobile environment and the specific problems raised by a realistic application scenario (e.g. the e-healthcare emergency service in a car accident).

The rest of this paper is organized as follows: Section II starts our survey from a brief introduction of the characteristics of mobile ad-hoc networks and the envisioned protocol architecture of mobile ad-hoc Grids. Section III, section IV and section V describe investigations into the challenges of service discovery, job scheduling and Quality of Service provisioning in the mobile ad-hoc Grid paradigm, respectively. An application scenario and the associated problems of utilizing mobile ad-hoc Grids for e-Healthcare emergency are discussed in section VI. Section VII concludes this survey.

2 FROM MOBILE AD-HOC NETWORKS TO MOBILE AD-HOC GRIDS

2.1 the Attributes of Mobile Ad-hoc Networks

Mobile ad-hoc networks are a collection of mobile nodes forming temporary networks without the support of any static infrastructures or centralized network backbones. Fig.1 shows an example of a mobile ad-hoc network consisting of wireless PCs and Laptops. Communications between one node and another normally require multiple hops as the transmission range of a mobile node is limited. Intermediate nodes that lie in the communication paths act as routers and forward packets towards the hosts. Diverse mobility patterns of different nodes make the situation even more complicated when several partitions appear in the ad-hoc network. The general attributes of mobile ad-hoc networks may be summarized as follows:

Fig. 1: an example of mobile ad-hoc network

- *Bandwidth and Power constraints*: The wireless channels over which mobile nodes communicate can not easily provide as much bandwidth as that of a wired connection, due to the adverse conditions of the wireless channel (e.g. fading, shadowing, interference, collisions, etc). Generally, mobile nodes are battery-powered mobile devices composed of energy-efficient components. In a network of such mobile nodes, node failures may occur if the computing and communication activities in a mobile device consumed too much energy before recharging.

- *Multi-hop delivery*: In mobile ad-hoc networks, the direct transmission range of a node is limited by its transmitting signal strength. When packets are sent out from a source node to its destination outside the direct transmission range, one or more intermediate nodes are required to relay the packets in a hop-by-hop manner.

- *Network Partitioning*: Network partitioning is another unique challenge in mobile ad-hoc networks. The main reason for network partitioning is the group mobility behaviour [7], in which mobile nodes with similar mobility patterns could eventually form mobility groups exhibiting distinct movement patterns. In the worst cases, network partitioning would break a connected network topology into several separate, disconnected partitions.

- *Infrastructure unpredictability*: Since mobile nodes can join and leave the ad-hoc network at will, the connectivity of links among mobile nodes may vary with time leading to the unpredictability of the network topology. In order to locate destination nodes they wish to communicate with, mobile nodes have to dynamically update their routing tables among themselves as they move around so as to refresh their global vision of the network.

2.2 Mobile Ad-hoc Grids: Integrating Ad-hoc Networking with Grid Functionalities

Building a Grid computing platform within the unreliable and resource-limited wireless mobile environment is a challenging task, but it provides a way to exploit the enormous collective potential of the networked mobile devices.

Recent studies [3] [5] have proposed conceptual models of mobile Grid computing for resource aggregation and collaborative problem solving. The communication architecture of ad-hoc Grids presented in [3] is based on a virtual backbone constructed from powerful mobile nodes and maintained with a proactive protocol. This Grid architecture deploys intelligent agents and relies on the virtual backbone to organize the network scale resources for a specific mission. MobiGrid [5], on the other hand, is a Grid architecture that integrates ad-hoc networking with functionalities of peer to peer systems to support distributed data management, i.e. indexing, searching and sharing.

By extending these conceptual models of mobile Grid computing further, we envision an efficient and adaptive mobile Grid architecture with integrated Grid functionalities and ad-hoc networking for coordinated resource sharing as well as problem solving. As shown in Fig.2, in the outlined protocol architecture of mobile ad-hoc Grids, the Grid functionalities are integrated with and distributed through the ad-hoc networking protocol stack. For example, the service discovery and job scheduling component will be interacting with the ad-hoc routing in the network layer. The proposed architecture would introduce much complexity into the present TCP/IP protocol stack. However, the beneficial outcome is that the mobile Grid nodes are freed from the inter-layer communications between the Grid middleware and the networking protocols and become more sensitive to network dynamics. In addition to the MAC and IP level security mechanisms developed

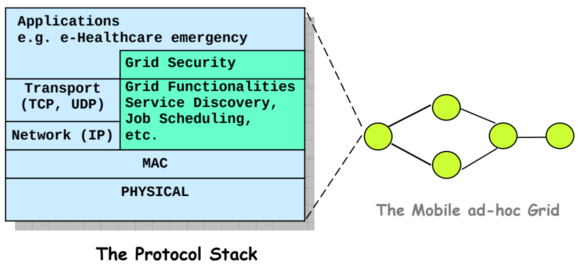


Fig. 2: the protocol stack of a node in the mobile ad-hoc Grid

for wireless/IP networking, the Grid security and authentication schemes still reside in application level so as to support application-specified security profiles.However, there are some new challenges posed by the integration of these two paradigms:

- Energy consumption is crucial for the lifetime of battery-limited mobile devices as well as the whole system. To make Grid functionalities feasible in the network of mobile devices, energy efficient should be the major concern for their implementation in each node.
- Unlike traditional Grid computing platforms, the network topology is neither stable nor predictable. The wireless connections between Grid nodes are normally bandwidth-limited and vulnerable to adverse signal quality.
- Without a reliable central infrastructure or backbone, the organization of Grid resources and execution of Grid applications would require decentralized operations.
- Fault tolerance is an important characteristic of Grid computing. The frequent failures of mobile nodes or communication links would always negatively impact the performance of the whole system.
- Since the system is built upon a wireless and pervasive network, it will be more difficult to perform authentications and to provide general security mechanisms.

The details of implementation and design challenges of this architecture of mobile ad-hoc Grids (e.g. service discovery, Job scheduling and QoS provisioning) are given in the following sections.

3 SERVICE DISCOVERY

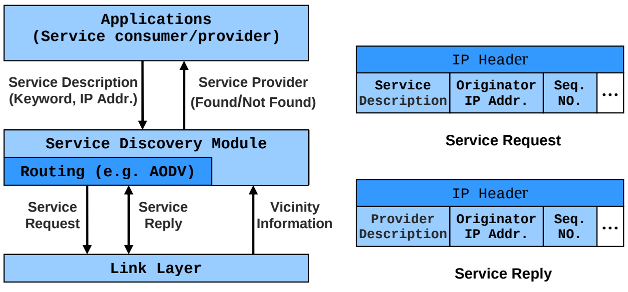
The concept of resource sharing in Grid computing is realized by service discovery mechanisms which transparently and seamlessly locates available resources/services throughout the Grid infrastructure upon request. By services, we mean software and hardware entities such as a computational routing, a segment of data file, a communication channel or storage, etc. In the basic paradigm of service discovery, service consumers advertise requests containing attributes representing the services they need. A few service providers who want to share their resources keep listening on a specified interface for service requests and reply to those matching the services they hold. A more complicated scheme may involve service brokers or directory agents who response to service requests on behalf of a set of service providers.

There are a number of service discovery standards that can be easily implemented for infrastructured Grid computing, e.g. Jini [8], UDDI [9]. These standards are developed with the assumption of fixed network topology and stable nodes acting as central service registry. Recent studies (e.g. Chord [10] and LANE [11]) make use of the Distributed Hash Tables (DHTs) to support distributed service lookup in large scale peer-to-peer networks. Distributed peer to peer service discovery can be also supported by UPnP [12] and SLP [13], in which nodes can choose to send service requests on demand to the local multicast address of service providers. The common problems of these distributed protocols are that they rely on underlying routing protocols to handle the network dynamics and construct fully routed topology for their service lookups. To enhance the its adaptability to frequent topology changes and to reduce inter-layer communications, recent work [14] [15] [16] have suggested to integrate service discovery with networking technologies (e.g. routing protocols).

Fig. 3a shows the protocol stack in which we integrated service discovery with an ad-hoc routing protocol (e.g. AODV) for mobile Grid computing. In this networking architecture, the conventional routing module is integrated into the service discovery module, which accepts a description of services demanded by applications, packetizes them into service request packets and then pushes them to the link layer. The service discovery module is also responsible for reporting the service availability to the applications according to the service replies it receives. If one of the applications is a service provider, it should also respond to requests and send back service reply packets to the requesting nodes. As shown in Fig.3b, the format of service request/reply packets are extensions of network routing request/reply packets, i.e. the destination IP address is replaced by the service description, which can be keywords or port number representing the requesting services. The service discovery process would reduce to be a normal routing discovery if the service description only contains an IP address.

3.1 Current Research on Ad-hoc Service Discovery/Routing

Existing research on network layer service discovery/routing protocols can be broadly classified as purely



(a) (b)

Fig. 3: Network level service discovery (a) the protocol stack with message flows, (b) the format of service request/reply packets.

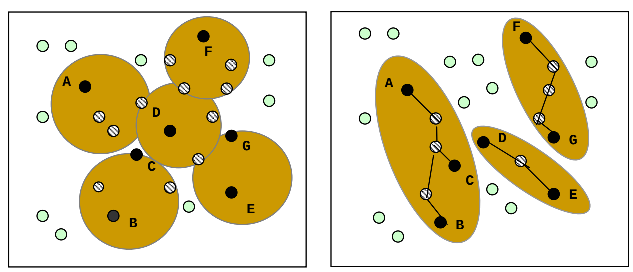
reactive [14] [17] [18], virtual backbone [15] [19] and hybrid reactive/proactive [20] [21] [22] [23].

- *Reactive service discovery*: In pure reactive service discovery approaches, e.g. EDSR [14], AODV [17], GOSSIP [18], nodes advertise service requests reactively so that periodical exchange of control message or service information is not necessary. Only when a node needs to know about nodes holding the resources it wants, that it floods service requests throughout the ad-hoc network. Response messages are sent back by nodes which provide the very resources being requested or cached information about routes to the right destination. Efforts to maintain proactive areas are saved in such an on-demand discovery scheme. However, flooding the whole network for every query is inefficient, especially when the service provider can be easily located at a node's neighbourhood.

- *Virtual backbone*: Approaches in the category of virtual backbone [15] [19] [24] try to maintain a virtual hierarchical architecture in the ad-hoc network to fit directory based service discovery protocols. The virtual hierarchy provides a way for the implementation of service discovery approaches that were developed for wired networks. The drawbacks are management overhead that constantly consumes extra bandwidth and its dependence on virtual backbone nodes that are prone to link failures.

- *Hybrid service discovery*: Hybrid approaches such as CARD [20], ZRP [21] and its variations [22] [23] try to exploit the advantages of both proactive and reactive strategies. These protocols limit periodic exchanges of route updates to a node's neighbourhood, i.e. zones or vicinities of nodes several hops away. Resources inside the proactive area will be ready on demand. For resources not reachable in a node's own zone or vicinity, these protocols would reactively initiate service discoveries by unicasting queries to remote 'contacts' [20] or bordercasting to neighbouring zones [21] [22] [23]. Common drawback of these approaches is that substantial bandwidth overhead need to be spent on maintaining the proactive area even in the absence of service queries or data traffic.

As part of a research programme to develop an efficient architecture of Mobile ad-hoc Grids, our recent work on the *Track*-based hybrid service discovery [25] sought to minimize the proactive traffic by its light-weight organization of the hybrid structure. It differs from Zone-based hybrid approaches [21] [22] [23] in that it maintains one dimensional track-like structure as proactive areas instead of two dimensional zones. As illustrated in Fig.4 (a), proactive areas in the Zone-based service discovery are circular zones consisting of nodes several hops away. Maintenance overhead for such a two dimensional topology is very expensive and increasing with the node density of the ad hoc network. On the contrary, as depicted in Fig.4 (b) the one dimensional structure of proactive areas in the *Track*-based scheme requires minimal maintenance overhead, as each node in a track only needs to keep trace of its predecessor and its successor in the same track.



(a) (b)

Fig. 4: Instances of proactive areas. (a) the Zone-based service discovery, (b) the *Track*-based service discovery.

3.2 More Design Challenges

In addition to the bandwidth efficient approaches that we described above, various issues of service discovery in mobile ad-hoc environments have been addressed recent studies. Nevertheless, we can still list hereafter more design challenges of mobile ad-hoc service discovery that demand further exploitations.

- *Energy Efficiency*: In a network of mobile devices, energy constraint may not be a problem for those with enough power supply, e.g. a vehicular mobile node. However, communications of powerful nodes could be affected by energy-limited devices if these devices are exhausted while acting as intermediate nodes. Energy-efficiency has been a major concern in recent work for networking in mobile ad-hoc networks. A common technique to reduce energy consumptions of service discovery is to limit the scale of flooding/query targets [26]. However, visible future directions probably lies in building a cross-layer power management system for ad-hoc Grids as described in [3].

- *Latency and Collision*: Service discovery latency is referred as the time required between when a service consumer issues a request and when it gets a positive response. Latency has been considered in designing service discovery protocols [15] [27]. It has been shown that latency can be reduced by a proper design of the service discovery protocol. Nevertheless, there is always a trade-off between the latency and other aspects of the system performance. Packet collisions can be also a serious problem, as a communication channel is shared by a number of nodes and packet transmissions are random. Although Medium Access Control (MAC) is the common way to reduce packet collisions, an inappropriately designed service discovery scheme would lead to more collisions than that controllable by the MAC protocol.

- *Continuous Service Provisioning in Network Partitioning*: As discussed in Section II, network partition happens when several parts of the ad-hoc network get disconnected completely due to node mobility. In network partitioning, some services/resources located in a partition would not be accessible by service consumers from other partitions. To guarantee a global scale service coverage, the service discovery component of mobile ad-hoc Grids have to predict further network partitioning according to present pattern of group-based node movements, and to replicate services in the predicted partitions before they get detached. Even though the problem of service provisioning in network partitioning has been addressed in some recent studies (e.g. [28] [29]), more realistic models that can accurately characterize the current pattern of group-based movement remain open for further research.

4 JOB SCHEDULING

The task of job scheduling in Grid computing is to allocate resources to a certain type of jobs for a certain amount of time. Typical resources that the scheduling system deals with depend on the requests of the applications, and they can range from processing nodes, data to network bandwidth, signal filter, etc. Such requests are defined as ”jobs”, which normally consists of information about the requirements on the resources that are necessary to execute this particular job. For example, a computational job may indicate that it can only be solved by a floating point processor. It is the task of a scheduling system to decide where and when to execute a job.

4.1 System Architectures of Job Scheduling

Common architectures for job scheduling in Grid computing includes [30],

- *Centralized Scheduling*: In centralized scheduling, all parallel jobs are scheduled by a central scheduler. The central scheduler has to collect information of network-wide resources and decide when and on which site to execute the submitted jobs. According to whether a job can be split to several parts for parallel execution at different sites, this centralized strategy can be further distinguished as Single-site scheduling and Multi-site scheduling. The centralized scheduling architecture obviously does not scale well with increasing size of the Grids. The central scheduler may also become a bottleneck in case of node failures. As an advantage, such centralized structure doses not require bandwidth overhead and latency for communications among multiple schedulers.

- *Hierarchical scheduling*: In hierarchical scheduling, there are low-level schedulers responsible for jobs submitted from a cluster of computers. The central scheduler is kept at the top level coordinating globally for low-level schedulers. The main advantage is that different policies can be used for local and global job scheduling. The availability of the whole system would not rely on a single scheduler.

- *Decentralized Scheduling*: In decentralized scheduling systems, distributed schedulers interact with each other and schedule jobs at remote nodes for execution. No central scheduler is responsible for global job scheduling. Local schedulers can submit/accept jobs to/from each other through direct communications or a central job pool. The advantages of distributed scheduling are better fault-tolerance and reliability, as the failure of a single component will not affect the whole scheduling system. However, without a central scheduler, a single scheduler usually made sub-optimal schedules, as it is not aware of system information of the global scale.

4.2 Challenges of Job Scheduling in Mobile Ad-hoc Grids

For scheduling of jobs in mobile ad-hoc Grids, decentralized scheduling seems to be the best candidate among the scheduling systems that were developed for wired Grids. However, the unique challenges of mobile environment demand more than what the available job scheduling techniques can provide. Apart from being light-weight to fit for the energy-constrained mobile devices, the following properties are worth being considered in designing a distributed job scheduling system for mobile ad-hoc Grids.

- *Alert for Node Mobility*: An important factor that influences the success rate and execution time of the scheduled tasks is the mobility of the devices. If the devices in the network are highly mobile during the lifetime of a distributed application running on the Grid, the application may be disrupted due to the reconfiguration of the network topology and previously established connections. Therefore, in mobile ad-hoc Grids, it is no longer sufficient to permanently select specific execution sites for distributed task. The scheduling system should be able to continuously monitor status of the selected execution sites and replace unavailable ones to resume the Grid application. If the rescheduling of distributed tasks and replacement of their execution sites occur too frequently, the overall execution time of an application would be delayed and at the worst case be interrupted if time runs out.

- *Robust in Network Partitioning*: When mobility causes network partitions or disconnections, schedulers and execution sites may not be able to communicate with each other if the partitions break all connections among them [31]. In such cases, the schedulers have to select new resources for submitted tasks and reschedule their executions. To avoid too much of rescheduling on network partitioning, future job scheduling systems are expected to be equipped with partition detection abilities.

- *Fault-tolerant upon Data loss*: Data loss can be a critical issue for job scheduling, especially when schedulers accept job submissions from applications or exchange job status with execution sites. For distributed Grid applications loss of data can result in potentially erroneous outcomes, i.e. chain reactions between scheduled successive jobs, in which some jobs in queue have to delay their own processing if the outcomes of their predecessors get lost and need to be retransmitted. Such a situation would require the scheduling system to be fault-tolerant while it is in distributed operation.

5 QUALITY OF SERVICE PROVISIONING IN MOBILE AD-HOC GRIDS

The primary issue in the applications of mobile ad-hoc Grid is the lack of application-aware QoS frameworks designed for the special challenges of the mobile ad-hoc environment. QoS is usually defined as a set of service requirements that the network must meet while transporting a stream of packets from a source to its destination [32]. The QoS framework of a network is a system of technologies which allow the network to guarantee a set of performance criterias representing the service requirements, e.g. delay, packet loss ratio and jitter for a normal IP network.

Quality of Service provisioning is more complex in the autonomous mobile ad-hoc Grid than in a wired Grid. For example, the communication quality of an end-to-end channel connecting two mobile hosts is difficult to control as the route may change over time. The unique characteristics of mobile ad-hoc networks demand new QoS frameworks to tackle the challenges that do not appear in the wired networks. In order to measure the service performance of mobile ad-hoc environments, it is also necessary to deploy new QoS metrics, e.g. power consumption and service coverage. Example QoS frameworks for mobile ad-hoc environments are INSIGNIA [33] and ASAP [34]. INSIGNIA is an adaptive QoS framework specifically designed to guarantee QoS for real-time traffic in mobile ad-hoc networks. The key component of INSIGNIA is its in-band signalling system, which allows fast service differentiation and resource reservation in response to the network dynamics. ASAP is a more efficient QoS framework with its two signalling messages, e.g. SR message to periodically setup/repair paths and collect QoS information, whereas HR to update the reservation state of the flow and to inform the sender upon QoS adaptations. However, current QoS frameworks for mobile ad-hoc networks limit themselves in more of the conventional end-to-end communications and do not cover the service requirements of those related to Grid computing, i.e. group communication/collaboration, job scheduling and service discovery. Furthermore, the applications running on top of a mobile ad-hoc Grid would impose their own perspective of QoS requirements (e.g. of data, networking and the Grid platform) that more than what present QoS frameworks of Grids/mobile networks can sufficiently support. For instance, given the safety-critical nature of healthcare, any serious deployments of mobile ad-hoc Grids for e-healthcare applications would not be possible before the special requirements of e-healthcare services are addressed in the QoS framework.

In view of the diversity and complexity of the application and networking requirements, an ideal QoS framework in a mobile ad-hoc Grid would integrate network awareness, application adaptations, and resource allocation/scheduling controls. Although not designed for the mobile ad-hoc environment, GARA [35] is a typical example of an integrated QoS framework for Grid computing and applications. It allows dynamic interactions between network level resource reservations and application level adaptations resulting in a more efficient QoS control framework. Combined with recent work on QoS frameworks for mobile ad-hoc networks (e.g. INSIGNIA), GARA could provide valuable hints to build an application aware QoS framework for mobile ad-hoc Grids. However, a simple combination of these works would not sufficient to deploy mobile ad-hoc Grids in realistic e-healthcare scenarios if the complex issues associated with healthcare are not covered.

6 DEPLOYMENT OF MOBILE AD-HOC GRIDS FOR E-HEALTHCARE APPLICATIONS

The perspective of mobile ad-hoc Grids would benefit applications such as the e-healthcare emergency. Considering a medical emergency taking place in a hostile environment where fixed computing/communication infrastructures are not available, a mobile ad-hoc Grid can be built instantly from a group of heterogeneous mobile devices to accomplish medical missions (e.g. diagnosis and treatment) that are crucial in this situation.

6.1 An Application Scenario of e-Healthcare Emergency and the QoS Requirements

We present a scenario of healthcare emergency to illustrate the issues regarding the deployment of mobile ad-hoc Grids in e-healthcare. As depicted in Fig.5(a), a car accident has happened in the middle of a highway. A man has been seriously injured and become unconscious. Police and ambulance have been called and coming to the scene of accident. By the time when an ambulance arrives, a policeman has arrived and collected the victim's personal or family information in his PDA. As soon as the ambulance and the paramedics arrive, a mobile ad-hoc Grid is constructed on a group of facilities (e.g. the police's PDA, the physical sensors attached to the injurer, a portable medical exam kit and the biomedical system in the ambulance) to process the collected information and to collaborate for prehospitcal treatment suggestions.

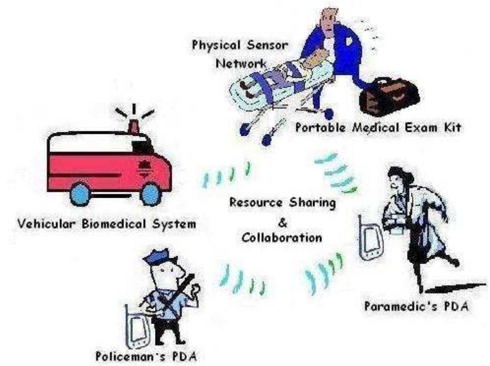
As shown in Fig.5(b), the e-healthcare applications can utilize the mobile ad-hoc Grid for distributed data (e.g. vital signals from the biosensors attached to the patient) discovery or job scheduling (e.g. to analyzed received vital signals in several medical devices). However, these operations impose strict requirements on the infrastructure and protocols in use due to the safety-critical nature of healthcare. Below are a set of basic QoS requirements that we believe are crucial in realizing the scenario described above.

- *Latency, Error Rate and Bandwidth of the data in transmission*

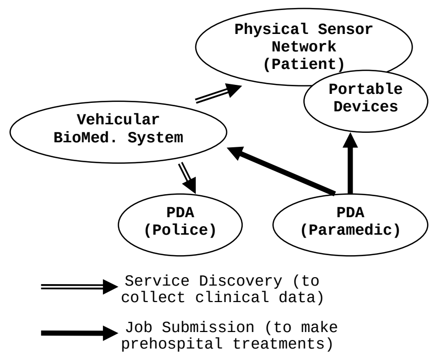
  In realtime transmissions, medical data would normally require constrained delivery delay for smooth one-way/two-way communications and acceptable bit/packet error rate that would not lead to misjudgement. The bandwidth requirements depend on the type of data being transmitted (e.g. an accurate 12 lead over-sampled ECG set would require a 64kbps UMTS channel [36]. However, the medical data would demand a guaranteed baseline bandwidth in frequent link fluctuations.

- *Reliable and adaptable communication*

  The process of diagnosis and treatment in the e-healthcare emergency is expected to be carried out on a stable communication infrastructure. The application would require the communication system to be adaptable upon link failures. For example, in transmitting sequences of biosignals or medi-



(a)



(b)

Fig. 5: Utilizing the mobile ad-hoc Grid for e-Healthcare emergency. (a) the car accident scenario, (b) the resource sharing model of the mobile ad-hoc Grid constructed for the accident.

cal images, the data transmitters should be able to switch to low bandwidth channels by transcoding when the high bandwidth ones are not available due to weak signal strength.

- *Secure authentications and privacy protections*

  As the data are transmitted in the public media (e.g wireless channels), they are more vulnerable to unauthorized access. To protect the privacy of medical data, the infrastructure should strictly authenticate users of the system, i.e. to only authorize access from medical devices, involved doctors and paramedics, etc.

6.2 Challenges of Deploying Mobile Ad-hoc Grids for e-Healthcare Emergency

The provision of the QoS requirements for the e-healthcare emergency is of crucial importance to the patient, as ”the appropriate medical intervention during the golden-window immediately following an accident significantly improves the chance of recovery” [36]. However, it is a challenging task for the Grid infrastructure constructed by portable devices to work in the wireless environment. The research on utilizing Grid technologies in healthcare sector have been reported in [37] [38] [39]. These work have addressed the ideas and issues arisen from deploying high performance infrastructured Grid computing for medical and biological applications. Recent work [36] [40] [41] have also discussed the QoS issues of e-healthcare emergency or teletrauma. Their work are mainly about the requirements and the performance of teletrauma systems based on current mobile networks. While they are expected to benefit the e-healthcare emergency with the coordinated resource sharing in pervasive environments, mobile ad-hoc Grids raised unique research problems on satisfying the requirements of e-healthcare emergency.

- *On-demand access to distributed/roaming data*

  The e-healthcare application would expect transparent access to network wide data disregarding their location or their status, e.g. roaming or static. This would require efficient organization of distributed data replications that are easily accessible and adaptive resource discovery mechanisms that can promptly responses to the resource requests even in high mobility.

- *Manageable jobs priorities*

  In the healthcare emergency, some types of medical data are more important than others for a quick diagnosis and should be processed accordingly. For example, Heart rate as a trend sign of ECG has a higher priority to be analyzed than ECG. The job schedulers of the mobile Grids should be application-aware and arrange the executing sequences of jobs according their priorities. In case of job-queue overflows, the low-priority ones may even be dropped.

- *Simultaneous transmission of data streams*

  When a group of medical devices collecting and exchanging information simultaneously in the ad-hoc Grid, they are likely to incur congestions and collisions. The situation could be alleviated by an distributed QoS framework which can reserve resources along the path of data streams and schedule their transmission in real-time with the latency and bandwidth constraints.

- *Fault-tolerant infrastructures with fault detection, logging and recovery*

  The mobile ad-hoc Grid should also have the capability of fault detection and self-recovery so as to isolate network failures from the healthcare applications for a smooth diagnosis and treatment procedure.

7 CONCLUSION

In this survey, we investigate the challenging operation of Grid computing in the infrastructure-less mobile ad-hoc networks constructed on a group of mobile devices. By integrating the Grid functionalities with ad-hoc networking protocols (e.g. routing) for better efficiency and adaptability, we outline a lightweight architecture of mobile ad hoc Grids to exploit the resource potentials created by the networked mobile devices. As a base for their design considerations, we discuss the general challenges of implementing Grid functionalities (e.g. service discovery, job scheduling and QoS provisioning) in the mobile ad-hoc environment and the specific issues raised by a real application scenario, i.e. the e-healthcare emergency service in a car accident. Our survey suggest that the key technologies of Grid computing need to evolve new capabilities to deal with the multiple challenges imposed by the mobile ad-hoc networking, Grid computing and QoS-sensitive applications (e.g. the e-healthcare).

ACKNOWLEDGMENTS

The work reported here is supported in part by the BIOPATTERN EU Network of Excellence (EU Contract 508803).

REFERENCES

[1] I. Foster, C. Kesselman, and S. Tuecke, “The anatomy of the grid, enabling scalable virtual organizations,” *International Journal on Supercomputer Applications*, vol. 15, no. 3, 2001.

[2] I. Foster, C. Kesselman, J. M. Nick, and S. Tuecke, "The physiology of the grid, an open grid services architecture for distributed systems integration," *Open Grid Service Infrastructure WG, Global Grid Forum*, June 2002.

[3] D. C. Marinescuand Gabriela M Marinescuand Yongchang Ji, L. Boloni, and H. J. Siegel, "Ad hoc grids: Communication and computing in a power constrained environment," *Workshop on Energy-Efficient Wireless Communications and Networks (EWCN), Phoenix, USA*, 2003.

[4] T. Phan, L. Huang, and C. Dulan, "Challenge: Integrating mobile wireless devices into the computational grid," *the 8th ACM International Conference on Mobile Computing and Networking (Mobicom 2002)*, September 2002.

[5] A. Datta, "Mobigrid: Peer-to-peer overlay and mobile ad-hoc network rendezvous - a data management perspective," *the 15th Conference On Advanced Information Systems Engineering, Austria*, June 2003.

[6] A. Litke, D. Skoutas, and T. Varvarigou, "Mobile grid computing: Changes and challenges of resource management in a mobile grid environment," *in Proc. of Practical Aspects of Knowledge Management (PAKM 2004), Austria*, December 2004.

[7] K. H. Wangand Baochun Li, "Group mobility and partition prediction in wireless ad-hoc networks," *in Proc. of IEEE ICC 2002*, October 2002.

[8] "JINI™ architectural overview, technical white paper," *Sun Microsystem*, January 1999.

[9] "UDDI version 3.0.2," *UDDI Spec Technical Committee Draft*, October 2004.

[10] I. Stoica, R. Morris, D. Liben-Nowell, D. R. Karger, M. F. Kaashoek, F. Dabek, and H. Balakrishnan, "Chord: A scalable peer-to-peer lookup protocol for internet applications," *IEEE/ACM Transactions on Networking*, vol. 11, no. 1, February 2003.

[11] M. Klein, B. Konig-Ries, and P. Obreiter, "Lanes - a lightweight overlay for service discovery in mobile ad hoc networks," *in Proc. of 3rd Workshop on Applications and Services in Wireless Networks (ASWN2003)*, July 2003.

[12] "Universal plug and play (upnp™) device architecture version 1.0," *UPnP™Forum*, June 2000.

[13] J. Veizades, E. Guttman, C. Perkins, and S. Kaplan, "Service location protocol," *RFC 2165*, June 1997.

[14] I. Gruber, R. Schollmeier, and W. Kellerer, "Performance evaluation of the mobile peer-to-peer protocol," *Proc. of Fourth International Workshop on Global and Peer-to-Peer Computing*, April 2004.

[15] U. C. Kozat and L. Tassiulas, "Network layer support for service discovery in mobile ad hoc networks," *in Proc. of IEEE Infocom 2003, San Francisco, USA*, March 2003.

[16] C. Frank and H. Karl, "Consistency challenges of service discovery in mobile ad hoc networks," *Proc. of the 7th ACM international symposium on Modeling, analysis and Simulation of WIreless and Mobile systems*, October 2004.

[17] C. E. Perkins, E. M. Belding-Royer, and S. Das, "Ad-hoc on-demand distance vector (AODV) routing," *RFC 3561*, July 2003.

[18] Z. J. Haas, J. Y. Halpern, and L. Li, "Gossip-based ad hoc routing," *in Proc. of IEEE Infocom 2002*, June 2002.

[19] J. C. Liu, Q. Zhang, W. Zhu, and B. Li, "Service locating for large-scale mobile ad hoc networks," *Kluwer International Journal of Wireless Information Networks*, 2002.

[20] A. Helmy, S. Garg, P. Pamu, and N. Nahata, "Contact Based Architecture for Resource Discovery (CARD) in large scale MANETs," *Proc. of Int'l IEEE/ACM Parallel and Distributed Processing Symp.*, April 2003.

[21] Z. J. Haas and M. R. Pearlman, "The Zone Routing Protocol (ZRP) for ad hoc networks," *Internet Draft*, November 1997.

[22] V. Ramasubramanian, Z. J. Haas, and E. G. Sirer, "SHARP: A hybrid adaptive routing protocol for mobile ad hoc networks," *in Proc. of ACM MobiHoc Conference 2003*, pp. 303–313, June 2003.

[23] L. Wang and S. Olariu, "A two-zone hybrid routing protocol for mobile ad hoc networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 15, no. 12, December 2004.

[24] I. Rubin, A. Behzad, H.-J. Ju, R. Zhang, X. Huang, Y. Liu, and R. Khalaf, "Ad hoc wireless networks with mobile backbones," *IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC 2004)*, 2004.

[25] Z. Li, L. Sun, and E. Ifeachor, "Track-based service discovery in mobile ad-hoc grids," *Technical Report, SPMC, School of Computing, Communications and Electronics, University of Plymouth, UK*, December 2004.

[26] A. Helmy, "Contact-extended zone-based transactions routing for energy-constrained wireless ad hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 1, January 2005.

[27] G. Pei, M. Gerla, and X. Hong, "Lanmar: Landmark routing for large scale wireless ad hoc networks with group mobility," *in Proce. of IEEE/ACM MobiHOC 2000*, Augst 2000.

[28] K. H. Wang and B. Li, "Efficient and guaranteed service coverage in partitionable mobile ad-hoc networks," *in Proc. of IEEE Infocom 2002*, June 2002.

[29] B. Li and K. H. Wang, "Nonstop: Continuous multimedia streaming in ireless ad hoc networks with node mobility," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 10, December 2003.

[30] V. Hamscher, U. Schwiegelshohn, A. Streit, and R. Yahyapour, "Evaluation of job-scheduling strategies for grid computing," *Workshop Grid 2000 at 7th International Conference on High Performance Computing (HiPC 2000), India*, pp. 191–202, 2000.

[31] P. Basu, W. Ke, and T. D. C. Little, "Dynamic task based anycasting in mobile ad hoc networks,"

ACM/Kluwer Mobile Ad-hoc Networks, vol. 8, no. 5, October 2003.

[32] P. Mohapatra,, C. Gui, and J. Li, "Group communications in mobile ad hoc networks," *IEEE Computer Magazine*, February 2004.

[33] S.-B. Lee, G.-S. Ahn, X. Zhang, and A. T. Campbell, "Insignia: An ip-based quality of service framework for mobile ad hoc networks," *Journal of Parallel and Distributed Computing*, vol. 60, no. 4, pp. 374–406, April 2000.

[34] J. Xue, P. Stuedi, and G. Alonso, "ASAP: An adaptive QoS protocol for mobile ad hoc networks," *IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC 2003)*, 2003.

[35] I. Foster, A. Roy, and V. Sander, "a Quality of Service architecture that combines resource reservation and application adaptation," *8th International Workshop on Quality of Service (IWQoS 2000)*, June 2000.

[36] I. Widya, A. van Halteren, V. Jones, R. Bults, D. Konstantas, P. A. M. Vierhout, and J. Peuscher, "Telematic requirements for a mobile and wireless healthcare system derived from enterprise models," *Proc. 7th Int. Conference on Telecommunications (ConTEL 2003)*, 2003.

[37] C. Chen and B. Schmidt, "Performance analysis of computational biology applications on hierarchical grid systems," *4th IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGrid 2004)*, 2004.

[38] R. McClatchey, et al, "The mammogrid project grids architecture," *Procedings of the 10th Int. Conf. on Computing for High Energy Physics (CHEP'03)*, 2003.

[39] V. Breton, A. E. Solomonides, and R. H. McClatchey, "A perspective on the healthgrid initiative," *Second International Workshop on Biomedical Computations on the Grid in conjunction with IEEE/ACM CCGrid 2004*, April 2004.

[40] Y. Chu and A. Ganz, "A mobile teletrauma system using 3g networks," *IEEE Transaction on Information Technology in Biomedicine*, vol. 8, no. 4, December 2004.

[41] V. Jones, R. Bults, D. Konstantas, and P. A. Vierhout, "Healthcare PANs: Personal area networks for trauma care and home care," *Proc. of Fourth International Symposium on Wireless Personal Multimedia Communications (WPMC)*, 2001.