Mobile Telemedicine using Data Grid

Sriram Kailasam, Santosh Kumar, and D. Janakiram

Abstract— **According to a survey by Indian Council of Medical Research, an abysmally low number of people living in rural India have access to specialist care and advice. Telemedicine aims to fill up this gap by enabling specialists deliver medical advice to remote areas. The recent advances in broadband technology for mobile phones add new dimensions to telemedicine by facilitating any time, anywhere access. It is no longer required for the doctor to be physically present at the hospital centre as is the case with most of the existing telemedicine systems. In this paper, we present an overview of a context-aware, P2P data grid framework for mobile telemedicine. We clearly indicate the need for context-aware scheduling in telemedicine and discuss the applicability of recent wireless technologies in large-scale telemedicine systems.**

I. Introduction

ACCORDING to a survey by Indian Council of Medical Research, an abysmally low number of people living in rural India have access to specialist care and advice. This opens up the possibility of using technology to bridge the gap and bring specialist care/advise to the door step of rural India. Telemedicine may be defined as the use of electronic information and communication technologies to provide and support health care when the patient and the specialists are remotely located. Many of the existing telemedicine solutions like [1], [2] and [3] are centralized i.e. the patient reports from the remote areas are received at the hospital center and the doctors needs to be present there to provide consultation. However, the recent advances in broadband technology for mobile phones add a new dimension to telemedicine by facilitating any time anywhere computing. Now, the specialist located anywhere in the world could use his mobile device to access the patient reports via internet and provide the required advice. WiMAX could be used to provide last-mile internet connectivity to the rural areas. Thus, there is this exciting possibility of using Internet to bring a shared object repository for collaboration between specialists located anywhere and anytime with health workers and patients located in rural India.

Many approaches like [4], [5], [6], [7] have introduced mobile telemedicine. [4], [5] describe telemedicine system for home care and patient monitoring wherein mobile phones can interact with electromedic devices like patient monitors and then transmit the vital signals via internet to the hospital for tele-consultation. [6] describes the design of a real time diagnosis system during ambulance travel. Approach followed in [7] is closest to our system. Their architecture comprises of 3 components: medical call centre, distributed specialists and allied hospitals which are linked by a service network. The requests have to be first made to the call centre who will redirect it to distributed specialists depending on the request severity. Our approach differs from that system in 2 ways. First, the data is stored on a data grid which is both scalable and robust. Next, a distributed context-aware scheduler does the scheduling automatically while supporting query load-balancing. Some attempts like [8], [9] have used advances in communication technology to facilitate multi-site collaboration among doctors using video-conferencing. Few others like [10], [11] have used grid technology to carry out large-scale medical simulations. However, the existing telemedicine solutions have overlooked the problem of context-aware scheduling of patient requests or in other words the dynamic discovery of resources (doctors) to serve the requests. With attempts like [12] to integrate vital sign sensors into textiles which can continuously monitor the patient, we believe that there will be huge amount of data generated and the problem of locating an appropriate doctor to interpret the data becomes important. Thus, in this paper we present an overview of a context-aware, P2P data grid framework for mobile telemedicine that addresses the above problem. Further, we discuss the applicability of recent wireless technologies to achieve mobility at both patients as well as doctor ends.

The remaining of this paper is organized as follows: section II explains the requirements of large-scale telemedicine. Section III explains our system model while section IV explains the choice of wireless technology at each step in the realization of the telemedicine system. Section V provides an overview of the implementation and we conclude in section VI.

II. Large scale telemedicine

A. Conventional model of telemedicine

The model followed by many of the existing telemedicine solutions is given below:

-- Tele-medicine infrastructure is set up at a big hospital and connectivity is established with different parts of the country using VSAT or other means.

-- A nodal centre (referred as telemedicine unit) is created in a local hospital in these remote areas where doctors and patients can directly interact with specialists at the big

The Telemedicine project is supported by Indo-German Project (Indian Council of Medical Research) and Intel Research.

Authors are with Department of Computer Science and Engineering, Indian Institute of Technology, Madras, India 600036. (email: ksriram@cse.iitm.ernet.in, santosh@cse.iitn.ernet.in, djram@iitm.ac.in).

hospital

-- The nodal center is equipped with different kinds of vital parameter measuring devices which connect to a desktop PC. The PC records the device readings.

-- The health worker then connects to a server at the big hospital and uploads the reports.

-- As soon as a report is received, a specialist at the big hospital would go through the report and send back comments and advice. The patient can be waiting at the other end and receive the reports personally.

B. Road blocks to scalability

As the number of telemedicine units increase, the centre at the big hospital may become a bottleneck. Hence, specialist advice may have to be provided by several such hospital centers, thereby necessitating large-scale enterprise collaboration among hospitals across the country.

Currently, the telemedicine infrastructure requires the specialist to be available at the centre whenever the data is received. This is important to ensure that the advice is sent within a few minutes, thus helping the patient who is waiting at the other end. However, this constraint on specialists to be always available may not be feasible as the number of requests increase.

Similarly, the remote telemedicine units are small in number and distributed across the country. While a remote unit helps the local population, it does not have the penetration that a mobile unit can achieve. Besides, it may not always be possible for a patient to come to a nodal centre. Instead, a doctor may go to the patient with an ECG instrument and his/her mobile device.

C. Vision of large scale telemedicine

This vision for scalable telemedicine is illustrated in Fig. 1 below.

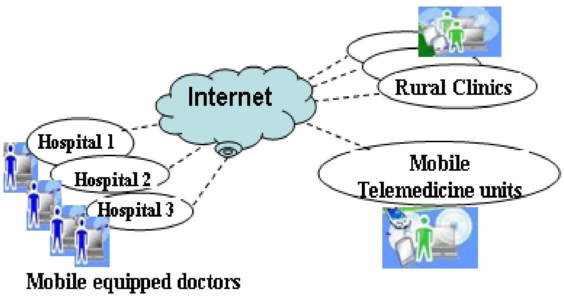

Fig. 1. Telemedicine: A giant virtual mobile enterprise

It comprises of multiple telemedicine hospitals, mobile medical specialists and rural mobile units/clinics forming a large virtual enterprise. It must support mobility at both the patient's end and the specialist's end. Support for mobility at the patient's end will result in increased penetration. In addition, since small mobile devices can be used for this purpose, it is cost effective. Due to these factors, the scale of operation of the system increases. Mobility at the end of a specialist also has several key advantages. One is flexibility - a specialist need not always be at the central server waiting for reports to be received. Another advantage is improved availability - a report will be delivered on to the handheld of a specialist, who may be located, anywhere, rather than a central server, so that he or she can immediately attend to it.

D. Other field challenges

The telemedicine system must be robust i.e. the patients must be able to depend upon the telemedicine system even during emergencies. For this, the system must guarantee service despite partial infrastructure failures.

The telemedicine system must be location and context sensitive. It must enable discovery and tele-consultation with a doctor located within shortest distances. This helps incase the patient is advised to travel to the nearest hospital. Besides, it must provide context-aware mechanisms to handle emergencies / normal cases.

The telemedicine system must be secure. It must maintain the privacy of patient data.

III. SYSTEM MODEL

A secure, context-aware P2P data grid framework integrating hospitals distributed across the country fulfils all the requirements of large-scale telemedicine. P2P systems have been shown to be scalable and robust. Adding context-awareness and security to it makes it dependable. Such a large-scale grid can provide an infrastructure for storing huge amounts of data as well as provide massive computational power. Context-aware scheduling ensures that the requests (from patients) are served by the best available resource (doctor) at that moment. The proposed architecture is available in [13]. In this paper, we only present a high level view of the mobile telemedicine system and discuss the applicability of recent wireless technologies in large-scale telemedicine systems.

A. System Overview

Fig. 2 shows the high-level view of the telemedicine system.

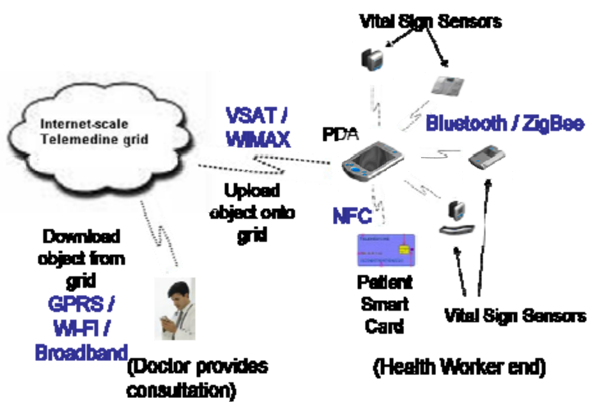

Fig. 2. Bird's eye view of the telemedicine system

The health workers are equipped with a paramedic suitcase containing different sensors like Blood Pressure meter, ECG jacket etc. They measure the vital parameters and then upload them onto the data grid. The location and context-aware grid scheduler matches the request context to a nearest available doctor and notifies him via email / SMS. He can then look at the request on his PDA and give consultation. The reply is then relayed to the patient.

As a pilot realization, vital parameters like ECG, BP, and Blood Glucose are measured for tele-consultation. As per reports published by Indian Council of Medical Research, cardio-vascular diseases are the most prevalent ones in India. 1 out of 6 in India suffers from heart diseases. However, the telemedicine framework is generic enough to support other kinds of ailments as well.

B. *Resource modeling*

The telemedicine system classifies resources into two types: data storage resources and medical resources (doctors). The data storage resources are modeled as peers in the telemedicine grid while the medical resources (doctors with mobile phones) are modeled as external entities which contact or are contacted by the grid nodes for providing/accepting services. Modeling the mobile nodes as external entities and not as peers effectively masks the resource constraints of these devices in terms of bandwidth, storage and intermittent connectivity. They are only used to store the request object while providing consultation as well as while uploading the request object. Permanent storage is provided by the grid which is realized as a persistent object space.

An object space provides an easy-to program abstraction for building applications. It allows us to model real world entities like patients, doctors as objects. The different kinds of objects used in the telemedicine system are patient profile object, treatment profile object, patient request object and doctor object. The patient profile object is used to maintain general data about the patient like name, blood group, address etc. The treatment profile object stores aggregate information about diagnosis, prescription, date of visit etc. for a particular area of treatment. The patient request object represents the current request while the doctor object stores the doctor's profile information. While the request is being served by the doctor, he may also need to see the patient history. Hence, the grid prepares a composite object consisting of the current request and the treatment profile, and schedules it to the doctor.

C. *Overlay structure*

The data storage nodes (static peers) in the telemedicine grid are grouped into proximity based zones. The zone boundaries are statically decided by considering the region population, the number of hospitals etc. The data of patients belonging to that region are replicated on nodes within the zone thereby ensuring replica proximity. The overlay structure used in the telemedicine system is shown in fig. 3. The structured network within zone uses Pastry [14] while Chord [15] is used for inter-zonal routing. Both Pastry and Chord are DHT-based routing substrates and provide lookup guarantee of O (log N) hops. The Chord routing protocol is used across zones because it ensures correctness in routing even if very few routing table entries are correct (lazy maintenance across zones is facilitated). The unique feature of the overlay is that each node is part of an independent Chord ring. This increases the number of entry points into the zone and thereby helps in the distribution of incoming requests from other zones. This overlay structure is similar to the one used in Vishwa [16], a P2P computational grid.

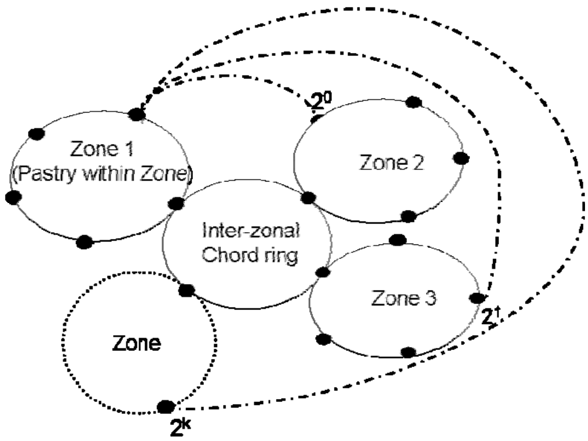

Fig. 3. Overlay structure of the telemedicine system

D. *Context-aware scheduling*

As we discussed earlier, most of the existing telemedicine systems are point to point i.e. they assume that a rural centre is mapped to a city hospital. Hence the patient requests are scheduled to doctors belonging to that hospital alone. However, by realizing the telemedicine system as a large-scale data grid and by enabling mobility at the doctor's end, scalable resource (suitable doctor) discovery is a problem. We adopt a context-aware scheduling mechanism to tackle this problem.

Context is information that can be used to characterize the situation of the entity. The telemedicine system considers the following context parameters in the patient request while scheduling: location, treatment history and language. Location-aware scheduling helps in case the patient needs to be hospitalized while doctors who have treated the patient previously (in the required area of consultation) and speak the same language as that of the patient would be able to quickly recognize and diagnose the patient. Besides these context parameters, the system must allow emergencies to be handled immediately. A redundant preemptive scheduling policy is followed in case there are no available doctors at that moment. That is the emergency request is scheduled to more than one doctor at a time and the doctor to whom the emergency request is scheduled is given an option to preempt the current request and serve the emergency. The preempt option is provided by the client application on the doctor's mobile phone.

Fig. 4 shows the architecture of context-aware scheduler in the telemedicine system. Context-aware scheduling is achieved by using zone-based scheduling. Every zone has a zonal server that acts as the patient request gateway. The zonal server also receives periodic resource advertisements from hospital nodes. For every hospital, one or more capable nodes are chosen as hospital nodes. They provide an interface to the doctors to indicate their availability for serving requests. They aggregate the doctor's availability status and compute the number of free request slots. Besides this, they also maintain capable node-set containing k-capable nodes from their neighbor list for scheduling requests. Periodically, they publish this resource information (or resource advertisements) : <number of free request slots, capable node list> with the zonal server. Thus the zonal server has aggregate information about the number of free request slots in each hospital within the zone.

When the patient request is received by the zonal server (request gateway), it first matches the GPS co-ordinates viz. the location of the patient to a nearest hospital having free request slots and redirects the patient request to a capable node registered by the corresponding hospital node. Thus the zonal server makes location as well as resource-aware decision. Now, the node to which the request is redirected considers the detailed context parameters like severity of ailment (normal / emergency), patient history, language etc. to select the best matching doctor. It then notifies the chosen doctor via SMS or email containing the uploaded object Id. If the doctor is free, he can download the request object from the grid and give his consultation which is updated on the grid and relayed to the patient.

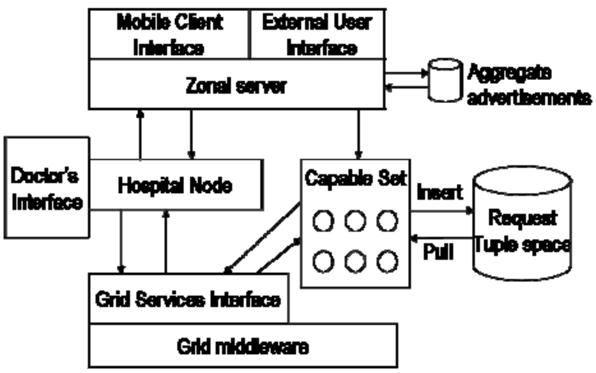

Fig. 4. Architecture of Context-aware scheduler

The decision to temporarily store the request on a capable node ensures best response time to the clients viz. the health worker while uploading the patient request, and the doctor while downloading the patient request object, and submitting his response. It should be noted that these are short-running tasks and hence it is reasonable to assume that the node remains capable for the short duration. Another important thing to be noted is that location-aware scheduling minimizes the data movement in the telemedicine grid by scheduling the request to a proximal doctor. Only when no doctor within the zone is free and the time spent by the request within the system is above a threshold, the request is made available globally so that any hospital with free request slots can serve it. This is achieved by inserting the pending request ids into a tuple space which is realized on top of DHT. Now any hospital with free request slots can pick the request from the tuple space. The exact details of the tuple space realization are beyond the scope of this paper and are discussed in [13]. Context-aware scheduling also uses a hierarchical scheduling mechanism to reduce the query load on the specialists. That is the request is first scheduled to a general doctor and only on his recommendation to a specialist. Thus context-aware scheduling policy tries to ensure timely fulfillment of requests while simultaneously minimizing the data movement by using zone-based scheduling. Finally, it empowers the non-specialists to share a good proportion of the query load by hierarchical scheduling.

E. *Security*

Security is an essential component of the telemedicine system. PKI is used to ensure security. Since the telemedicine grid comprising of data storage nodes from different hospitals is more or less a closed system, it is easy to maintain security among the different node interactions. The current version of the telemedicine system does not handle byzantine failures of data storage nodes and provides only basic security.

IV. MOBILE TECHNOLOGIES IN TELEMEDICINE

As discussed earlier, supporting mobility is an integral part of the future telemedicine enterprises. In this section, we shall review the spectrum of wireless technologies in terms of bandwidth, range and availability. Then, we describe how some of them are used in the proposed telemedicine system.

A. *Review of wireless technologies*

Wi-Fi is a wireless networking technology with network size ranging from size of a room to area covered by a building. This technology can be used in the hospitals to provide broadband internet connectivity to doctors within the hospital. WiMAX is a wireless digital communication system with range of 50km. This technology can be used to provide the "last mile" connectivity viz. providing broadband internet connectivity to the rural areas where installing GPRS is expensive. ZigBee is a wireless mesh networking standard which is intended to be simpler and cheaper than other Wireless Personal Area Networks (WPAN) technologies like Bluetooth. It also consumes less-power. Thus the target applications are those that require low data rates, long battery life, and secure networking. Bluetooth is a standard similar to ZigBee with the advantage that more devices support it. Thus, ZigBee / Bluetooth are well suited to be integrated into the medical sensors.

B. *Internet connectivity at doctor and patient ends*

The telemedicine system requires internet connectivity at the doctor and health worker ends. The doctor can connect to the internet via hospital Wi-Fi when he is in the hospital. While on the move, he can use GPRS or mobile broadband connection. Depending on where he is, he may use different modes to connect to the internet. Now, each of these will support different data download rates. This information must be captured as part of the context. Besides this, his current location and availability (free/busy) are other important context parameters.

The client application running on the doctor is pre-configured with certain gateway addresses. It connects to one of the gateways and passes on this context information to the grid. The gateway relays the current location information of the doctor to the nearest set of hospital nodes and returns their addresses to the application. The application can post future updates at these set of addresses. In the current version, this process is manual i.e. the doctor must invoke the application and update his status with the hospital node. The application needs to be integrated with GPS, so that it auto-relays the context information to the grid.

The mobile device in the health worker end can connect to the internet using WiMAX / VSAT / GPRS or any other means. The telemedicine system only expects internet connectivity. As a pilot realization only patient ECG, blood pressure and blood glucose will be used. Blood pressure and Blood Glucose are just numbers while size of ECG is around 256 KB. Hence, the data upload rate doesn't matter much. Table I indicates the estimates of digital image sizes. This gives us an idea as to where mobile telemedicine could be used.

TABLE I
DIGITAL IMAGE SIZE ESTIMATES*

Image Type	Image Resolution (spatial)	Image Size
Ultrasound	512 x 512	256 KB
Angiography, Endoscopy, Cardiology, Radiology	512 x 512	256 KB
Computed Tomography	512 x 512	384 KB
Magnetic Resonance Imaging	1024 x 1024	1.5 MB
Scanned X-Ray	1024 x 1280	1.9 MB
Digital Radiology	2048 x 2048	4 MB
High Resolution Digital Radiology	2048 x 2048	6 MB
Mammography	4096 x 4096	25 MB

*Source is Fig. 3 of reference [18]

C. *Use of ZigBee / Bluetooth between sensors and PDA*

Few medical sensors like [17] have integrated ZigBee / Bluetooth into them. They can communicate the vital measurements to a PDA. We propose to use them for our pilot realization. The PDA can then compose the vital measurements into a patient request object and upload it onto the grid. It receives the request Id as an acknowledgement. When the response arrives from the grid, the PDA uses the request Id to match the request and the response.

D. *Use of NFC enabled mobile phones by health worker*

An economic model needs to be worked out for the sustenance of the telemedicine system. As a research prototype implementation, this part has not been explored. Integrating health insurance into the telemedicine system would make patient authentication and data security very important. As a pilot implementation, smart cards could be issued to the rural patients. The smart card would store the private key for the patient. NFC enabled mobile phones used by the health worker can read the data from the smart card and authenticate the patient to the system. Once the prototype is found to be feasible, biometric smart cards could be used for better security.

V. Implementation Overview

P2P data grid middleware has been realized in java. The object schemas are represented using xml. The composite object consisting of the patient request and the treatment profile (patient history) is prepared by applying XSLT to the xml source files. The composite object URL is then sent via SMS / email to the doctor. Java servlets are used for the web interface while the mobile client application (midlet) is done using J2ME with MIDP profile. The screenshots of the prototype application using Nokia Series-40 Emulator are shown in Fig. 5. The doctor can look at the ECG on his mobile phone and submit his response to the grid. The prescription is updated on the grid and relayed to the patient.

Fig. 5. Screenshots of the prototype system at the doctor-end

Though the telemedicine application is web-based, the client-side application on the mobile is required to effectively mask node failures (exception handling) from the doctor. Besides, the mobile client application also pre-fetches the patient request object before intimating the doctor. This eliminates the waiting time of the doctor thereby improving his efficiency.

Wireless Messaging API (JSR-120) is used for receiving SMS while Clickatell SMS gateway is used to send SMS notification to the doctor. The J2ME midlet (doctor side) is registered to a PUSH registry for listening on a particular port. When an SMS arrives on that port, the midlet is automatically invoked. The midlet application processes the SMS, fetches the request object from the grid, stores it locally and then notifies the doctor about the arrival of the request. An SMS is sent to a specific port by setting the UDH (User Data Header) in the SMS gateway.

VI. CONCLUSION

In this paper, we have seen that using a data grid along with a context-aware scheduler makes the telemedicine system scalable and robust. The data grid can be seen as a nation-level distributed database of patient medical records. This data can be used for carrying out large-scale simulations for medical research. Besides, by integrating health insurance, blood banks, ambulance etc. into the grid, we can have a full-fledged health grid spanning across the country that can provide a whole lot of medical services. Such a health grid will be of immense help to the developing nations and this forms part of our future research.

ACKNOWLEDGMENT

We thank Rahul Mourya from IT-BHU for his contributions towards building the doctor side application.

REFERENCES

[1] Medintegra. Available: http://www.telemedicineindia.com/medint_web.html

[2] Center for Connected-Health at Partners Healthcare, Massachusetts General Hospital, Boston, MA, USA. Available: http://www.connected-health.org/

[3] BELGIUM-HF - the largest belgian telemedicine clinical trial on congestive heart failure http://www.belgium-hf.be/

[4] M. V. M. Figueredo and J. S. Dias, "Mobile Telemedicine System for Home Care and Patient Monitoring," *in Proc. 26th Annu. Intl. Conf. of the IEEE EMBS*, San Francisco, CA, 2004, vol.2. pp. 3387 – 3390.

[5] J. Mikael Eklund, Thomas Riisgaard Hansen, Jonathan Sprinkle and Shankar Sastry, "Information Technology for Assisted Living at Home: building a wireless infrastructure for assisted living," in *Proc. 27th Annu. Conference*, Shanghai, China, 2005.

[6] Xiao, Y. Gagliano, D. LaMonte, M. Hu, P. Gaasch, W. Gunawadane, and R. Mackenzie, C. "Design and evaluation of a real-time mobile telemedicine system for ambulance transport," *Journal Of High Speed Networks*, vol. 9, no. 1, pp. 47-56, 2000.

[7] Tepei Tang and Jeng Ku, "Mobile Care: Telemedicine Based On Medical Grid and Mobile Network" in *Intl. Conf on Wireless Communications, Networking and Mobile Computing*, 2007, pp. 3160-3163.

[8] Telemedicine on the Grid Project at University of Cambridge Available: http://www.escience.cam.ac.uk/projects/telemed/

[9] L.S. Satyamurthy, "ISRO's experience in Telemedicine with special reference to Mobile Telemedicine system," November 2007. Available: http://www.aprsaf.org/data/aprsaf14_data/day1/CSA05_APRSAF-14%20-%20Telemedicine.pdf

[10] G. Berti, S. Benkner, J. Fenner, J. Fingberg, G. Lonsdale, S. Middleton, and M. Surridge, "Medical Simulation Services via the Grid," in *17th International Parallel and Distributed Processing Symposium (IPDPS 2003)*, Nice, France, April 2003.

[11] The GEMSS Project: Grid-Enabled Medical Simulation Services, EU IST Project, IST-2001-37153. Available: http://www.gemss.de/

[12] Tae-Ho Kang, Carey Merritt, Burcak Karaguzel, John Wilson,Paul. Franzon, Behnam Pourdeyhimi, Edward Grant, and Troy Nagle, "Sensors on Textile Substrates for Home-Based Healthcare Monitoring," in *Proc. 1st Distributed Diagnosis and Home Healthcare (D2H2) Conference*, Arlington, Virginia, USA, April 2-4, 2006.

[13] Sriram Kailasam, Santosh Kumar K and D. Janakiram, "Architecture for an Internet-scale Telemedicine Grid," *Technical Report IITM-CSE-DOS-2008-03*, Distributed & Object Systems Lab, Department of Computer Science & Engineering, Indian Institute of Technology Madras, 2008.

[14] A. Rowstron and P. Druschel, "Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems," *IFIP/ACM International Conference on Distributed Systems Platforms (Middleware)*, Heidelberg, Germany, pp. 329-350, November, 2001.

[15] Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan, "Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications," in *Proc ACM SIGCOMM*, San Diego, pp. 160-177, Aug 2001.

[16] M. Venkateswara Reddy, A. Vijay Srinivas, Tarun Gopinath and D. Janakiram "Vishwa: A reconfigurable P2P middleware for Grid Computations," in *Proc Intl. Conf. on Parallel Processing (ICPP'06)*, pp. 381-390, 2006.

[17] BioComfort Health Manager. Available: http://www.biocomfort.com

[18] Ackerman, Michael, R. Craft, F. Ferrante, M. Kratz, S. Mandil, and H. Sapci, "Telemedicine Technology," National Library of Medicine, *Telemedicine Journal and e–Health*, vol. 8, no. 1, chapter 6, 2002.