A Distributed Peer-to-Peer Platform for Synchronized Group Collaboration and Knowledge Sharing

Jo-Yew Tham, *Member, IEEE*; Seng-Luan Lee; Choon-Ee Tan, Roger; and Leong-Chiang Tee
Centre for Industrial Mathematics*, Department of Mathematics,
National University of Singapore
2 Science Drive 2, Singapore 117543
Email: thamjy@ieee.org; {matleesl; scitance; mattlc}@nus.edu.sg Tel.: +65 6874 3183

ABSTRACT

This paper presents new research and development of an integrated peer-to-peer (P2P) platform comprising of a network of distributed and decentralized peer devices connected directly with one another in an ad-hoc virtual group manner. The platform is built upon an extended version of the Sun Microsystem's Project JXTA and the Jabber's XMPP (Extensible Messaging and Presence Protocol) protocols. The proposed P2P platform has a comprehensive set of application programming interfaces (APIs) that provide a high-level encapsulation of many core P2P platform services, security control and policies, online presence management, data management and transfer, etc. to the application layer. By using these platform APIs, we have developed a number of essential tools targeting distributed e-Education and e-Collaboration environments, such as integrated secure chat, digital asset management, sharing, searching and retrieval, synchronized calendaring and contacts management, and scalable multimedia communications. The core vision and strategy here is to enable a truly distributed means for multiparty communications and collaboration in an ad-hoc peer group eco-system without the need for centralized systems, file servers, databases and corporate networks setup (such as extranets and virtual private networks). The platform's flexible plug-in and XML web services architecture allows easy development and integration of many new applications and services modules. By employing a 100% Java implementation, the platform is OS-independent and has been shown to work well on Windows, Macintosh, and Linux.

Keywords: Distributed peer-to-peer infrastructure, distributed database, multimedia, and scalable compression, presence management, synchronized group e-Collaboration, Project JXTA, XML Web Services.

1. INTRODUCTION

Consider a world of directly connected networks whereby each online user can actively and conveniently engage in a unified interactive environment to securely exchange his or her knowledge as well as to share, search and retrieve digital assets directly with one another. Such interactions can happen at any time and place using any networked devices, regardless of whether the user is on the public Internet, behind some corporate firewalls, or on mobile networks. In addition, this ad-hoc environment offers state-of-the-art cryptographic security and grants authorized user the total freedom to create ad-hoc protected collaborative groups *without* the regulated central control and setup of secured extranets or virtual private networks. Hence, this results in near-zero administrative and operational costs due to the absence of expensive centralized servers, databases, and file storage systems, while fully leveraging existing computing hardware resources as the peer devices in this P2P eco-system.

Although there exists a number of client-server solutions that attempt to create these collaborative environments, such a centralized computing solution does not truly add value nor fit into the natural workflow lifestyle demanded by many users. Consider the scenario of knowledge collaboration (i.e., organize, classify, share, search, retrieve, co-edit, post, survey, group chat, etc.) across different corporations or among multiple geographically dispersed offices of a large multi-national company. It is important to note that both the digital assets (data, documents, graphics, audio and video files, presentations, spreadsheets, etc.) as well as the knowledge base of the professionals are naturally originating and residing at their respective computing systems and not automatically pooled together at any centralized servers. By forcing the users to upload all digital assets from the local systems to a centralized repository, just to enable searching and sharing, clearly does not fit into a natural and productive workflow process. More importantly, the knowledge know-how and skill-sets of a professional are intangible assets that cannot simply be stored in centralized locations for later searching and retrieval.

In today's many client-server solutions, it has become habitual that almost nobody shares anything by explicitly uploading his or her assets from the local computers to a central server. At the very most, some assets may be shared but they are oftentimes not the latest versions. Hence, the inability to find the right updated information and the difficulty to share knowledge can become very costly and unproductive. According to the Meta Group, workers spend approximately 25-35% of their time searching for the information they need, rather than working on strategic projects and business opportunities. IDC Research further states that Fortune 500 companies will lose $31.5 billion by 2003 due to rework and the inability to find information. With a unified ad-hoc P2P platform, all peers are directly connected and the most updated shared assets can be directly accessible given sufficient authorizations. In addition, a fully distributed and decentralized P2P eco-system is not plagued with some common client-server problems, such as:

- *Scalability*: A centralized computing system requires regular infrastructure upgrading to support a growing client base. Otherwise, it will become the processing, storage and bandwidth bottlenecks as the transaction traffic increases, thus leading to slow or even stalled interactivities. Centralization also presents the

* This work was supported in part by the InfoComm and InfoTech Initiative (ICITI) project grants.

possibility for a single point of failure of the entire eco-system.

- *Flexibility*: A centralized system dictates and restricts the choice of application tools for the users. All setups and configurations are managed and controlled by the administrators or service providers. While this may have some benefits, a P2P system, on the other hand, provides a natural and flexible environment for users to collaborate and share information directly with one another in an ad-hoc manner at anytime, anywhere using the most appropriate application tools.
- *Complex and High Costs*: The setup of extranets and virtual private networks (VPNs) to enable cross-enterprise collaborations encompassing multiple corporate firewall boundaries can be complex and time-consuming. It may require additional server hardware and software licensing as well as IT manpower for their operations, maintenance, and supports.

With a vision to overcome these constraints and high costs of a centralized solution for distributed ad-hoc group communications and collaboration, a number of P2P-based systems have been developed. Among these, Groove Networks [3] is one of the most established platforms in the market. However, it may not be truly scalable in an ad-hoc sense due to its reliance on some centrally managed relay servers for peer discovery and connection establishment (especially when the connecting peers are behind some firewalls). The proposed P2P ad-hoc platform based on Project JXTA aims to overcome such shortcomings.

2. THE PROPOSED AD-HOC P2P PLATFORM

Overview

Figure 1 below depicts a typical online P2P community using the proposed distributed platform. Secured ad-hoc virtual peer groups (VPG) can be easily formed for multiparty communications and knowledge collaboration among members sharing a common topic of interest.

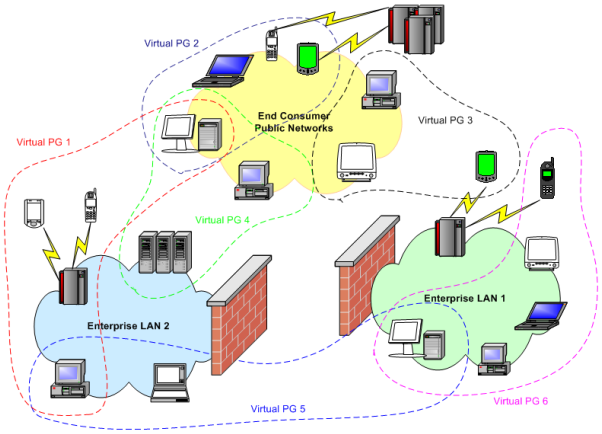

Fig. 1: *A typical fully distributed and decentralized P2P eco-system without the needs and constraints of any centralized control, thus enabling the total freedom of forming secured ad-hoc virtual peer groups (VPG) comprising members sharing a common topic of interest.*

Architectural Overview of Ad-Hoc P2P Platform

With the vision to develop an extensible P2P platform that supports disparate vertical applications, we have ensured that the design incorporates a good set of core APIs encapsulating the Project JXTA [7], Jabber XMPP [8], and XML Web Services [9] standards. Figure 2 shows a high-level architectural overview of the platform. It provides some core platform services such as authentication, file management and transfer, basic console GUI widgets, messaging, synchronization, multimedia streaming, etc. Third-party adapters to existing backend systems (e.g., LDAP, CRM, ERP, etc.) can also be developed, thus enabling the P2P platform to act as the middleware glue with other systems. A plethora of applications can then be developed and plugged into the proposed platform and made interoperable with other tools to create a unified and integrated peer console.

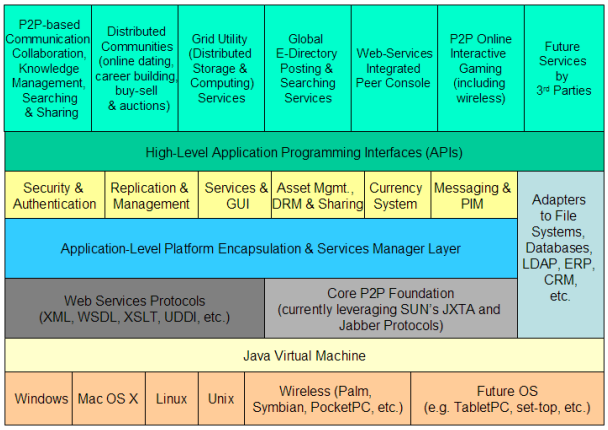

Fig. 2: *Architectural overview of the proposed distributed and decentralized P2P platform. Its flexible plug-in and web services-enabled design allows easy integration of new applications and services modules via a comprehensive set of the platform APIs.*

From a research and development standpoint, we focus on a number of strategic multi-disciplinary areas encompassing:

- Information security in a distributed ad-hoc environment, which includes finding better elliptic curves and developing more efficient cryptosystems; identity-based encryption and its application to simplify the deployment of public key infrastructure; data protection and access control via digital rights management; peer handshaking protocols for secure exchange of encryption keys, setup of secured peer-to-peer sessions; and localized (non-centralized) peer authentication and digital certificate/signature verification for peer identification and non-repudiation.
- Optimization algorithms for efficient global load balancing, and distributed data storage. Advanced information dispersal algorithms (IDA), which were first considered by mathematician, Michael O. Rabin [5], are also researched for developing fault-tolerant and secured data replication, transmission, searching and retrieval. Implementation of secured collaborative access via a P2P-based secret key sharing and management, and the development of feedback-free P2P communication protocols for parallel data retrieval and group data multicasting are also of great interest.
- Real-time communications and collaboration via lossy and lossless multimedia compression technologies for efficient media storage as well as online voice/video chat and messaging. This collaborative tool leverages our advanced wavelet video codec [1], [2], whereby its

highly scalable compressed bit stream feature are fully exploited in a P2P environment for adaptive rate-control in multiparty conferencing over heterogeneous networks and peer devices. Scalable live video multicasting [4] and broadcasting in peer groups is also being researched for development.

- Knowledge management (KM) engine for lightweight peer-side knowledge mining, processing and classification. These KM algorithms will greatly facilitate automated and intelligent indexing of information on each local peer for faster and more accurate deep searching and fast parallel data retrieval.

Handshaking Protocol for Secured Peer Session

Figure 3 illustrates one proposed peer handshaking protocol between two trusted peers who have previously authenticated with one another's digital fingerprints when they first joined the same VPG. The **PeerA-Hello** and **PeerB-Hello** messages negotiate a suitable cipher suite (e.g., *ECDH-ECDSA-RC4-SHA*). Peer A (the session initiator) then generates a master keys (i.e., the actual symmetric keys that will be used for encrypting subsequent data once the secured session is set up) and shares it with Peer B via **PeerA-MasterKey**. The use of CHALLENGE data and CONN_ID data are used here to circumvent possible *replay attack* by a third-party who could sniff at the channel pipes. By utilizing both Peers A and B's public keys, this protocol also prevents *man-in-the-middle attack* by forcing the engaging peers to decrypt using their respective private keys. **PeerA-Complete** and **PeerB-Verify** also serve to verify the correctness of the exchanged secret keys. Once Peer A has verified the CHALLENGE data, Peer B can then complete the handshaking via **PeerB-Complete** with the encrypted SESSION_ID. This session ID can be cached by each peer and be reused should they decide to restart a secured session within a short period of time after the termination of the previous session. This greatly speeds up the entire process by avoiding the key generation and verification steps and reusing the already exchanged secret keys for the new secured session (if such a security level is acceptable by the application of interest).

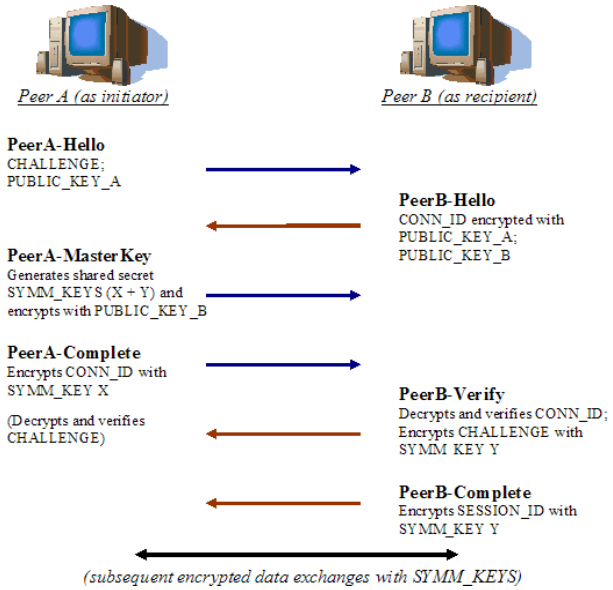

Fig. 3: *Proposed peer handshaking protocol between two trusted/verified peers for secured key exchange and setup of a secured peer session using a symmetric encryption key.*

Extension work related to identity-based encryption (IBE) [6] for public-key cryptography can also be employed. In this scenario, Alice can simply send a secure message to Bob by encrypting her message using one of Bob's unique identifiers (but publicly known to the other peers) such as the public key (e.g., bob@email.com or the unique peer ID). Now, there is *no need* for Alice to obtain Bob's public key certificate, and Bob does *not* need to pre-register with any Certificate Authority before Alice can send the secure message to Bob.

3. IMPLEMENTATION OF BASIC P2P PLATFORM APPLICATIONS

This section presents a few basic but useful application tools that are developed for the proposed P2P platform. They include ad-hoc virtual peer group (VPG) management with online peer presence monitoring using an integrated buddy roster, secured asset management and distributed sharing, searching and retrieval, synchronization of shared calendars and contacts, and scalable multimedia communications and collaboration in multiparty ad-hoc environments.

Secure Ad-Hoc Peer Grouping and Presence Management

In this P2P eco-system, a peer can create a new VPG and protect it with a password. Alternatively, the VPG can also be linked to a centralized LDAP server for membership authentication. The group owner will have full control over the membership, roles and rights of each member in the protected (can be either a public or private) VPG. In a private group, invitation to join the group can be issued to some selected peers who can then join by supplying the correct authentication credentials. Once joined, the online presence of all members within the VPG can be monitored instantly via an integrated buddy roster. As the platform also supports the standard Jabber's XMPP (Extensible Messaging and Presence Protocol), the same unified roster also displays online presence of buddies in other popular instant messaging (IM) gateways such as Yahoo!, MSN, AOL, ICQ, IRC, and Jabber. The primary focus here is to enable secured communications (both one-to-one and group chats) on all IM gateways via our unified peer consoles. Figure 4 illustrates a screenshot of the integrated buddy roster with secured chat and peer asset browsing.

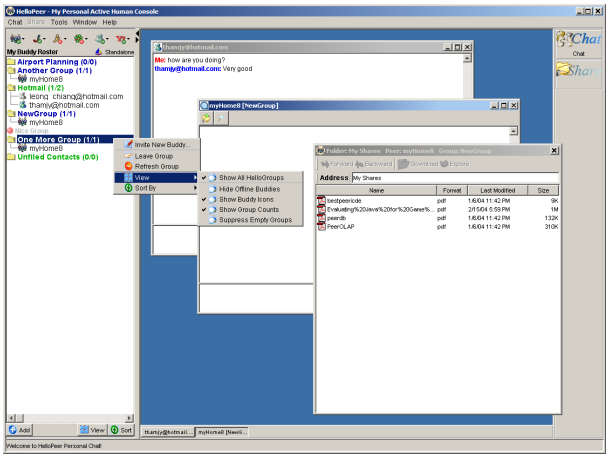

Fig. 4: *A screenshot of an integrated buddy roster that provides online presence monitoring across multiple IM gateways and a unified interface for easy access to other applications, such as secured chat and peer asset browsing and retrieval.*

Distributed Asset Sharing and Retrieval

Each member in a VPG can also conveniently manage and organize all his or her digital assets into hierarchical albums in a shareable library, and then apply specific security policies to govern the visibility, accessibility, and usage rights of the shared assets. Known types of digital assets (such as .doc, .pdf, .jpg, .gif, .mpg, .avi, .ppt, .xml, etc.) as well as the entire directory can be selected from either the peer's local hard-disk or a mapped networked drive for sharing. For example, a confidential financial document can be encrypted and protected with a password before it is made shareable with other members. The document owner can further specify the time period within which the document is available for searching and download, and the particular group of peers who can have visibility and access to the shared document.

For each shared asset, the owner can specify free-text metadata fields to describe it so that other peers with the authorized permission can search for the asset via search keywords. We have conducted research and compared two different approaches to this peer file sharing: (i) file replication and synchronization with shared asset advertisement published; and (ii) file residing on local peer and indexed locally, and searching is then performed locally too in response to distributed queries. We concluded that approach (ii) is very efficient with fast searching and greatly reduces bandwidth utilization needed for file replication. However, approach (ii) requires that the asset owner's peer console be online in order for the requesting peer to directly download the shared asset from. Extension using ideas on information dispersal algorithms for file splitting and replication on multiple peers are being considered for improved performance on parallel file retrieval and high data availability [5]. However, security protection and access control issues still require careful investigation and implementation.

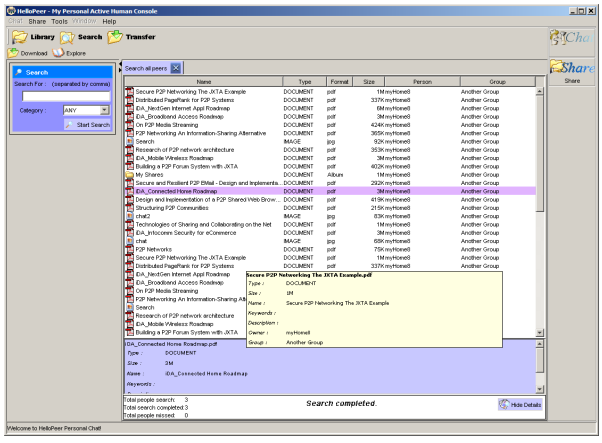

Fig. 5: *A screenshot of the peer console interface which provides a simple but powerful search capability for all shared assets by other members in all joined virtual peer groups. It also allows easy asset management in a shareable library with flexible access control policies.*

Once the assets are organized for sharing, a requesting peer in a particular joined peer group can then conduct a search by specifying some keywords, asset types, and/or the selected peer groups to be searched. The search request will then be dispatched to all the relevant peers in the P2P eco-system. Each peer who receives this search query will then perform a local search (if the requesting peer has the permission) and return the search results (if any) to the requesting peer. The requesting peer will then collate and present the search results in a friendly GUI which displays some attributes of each matching asset (see Figure 5). The requesting peer may then download one or more of these files (if he/she has the permission as specified by the asset owner while creating the shared library) directly from the owner peer's device. Of course, if the owner peer is currently online, the requesting peer can also initiate a new chat session. More importantly, we ensure that all file transmissions are encrypted for secured access and content integrity checking.

Synchronized Sharing of Calendars and Contacts

Each ad-hoc VPG has a common group calendar that is shared and automatically synchronized with other members in the peer group. A group calendar can also be created without being associated with a VPG. Members in a particular VPG having the appropriate rights can add, edit, delete, publish, or subscribe to the appointments in the calendar. The creator of an appointment can apply specific publication security access rules to determine the scope of visibility and editing rights of the appointment within the VPG. Once the appointments are published, other members having the appropriate access rights can (automatically or manually) subscribe for their latest updates via real-time online peer synchronization. From a graphical user interface perspective, a user can view multiple overlaid group calendars simultaneously on the peer console. Furthermore, since members in a VPG are directly connected, a meeting organizer can also browse for the time availability of each member in the group in order to schedule the best time slot for the meeting. Meeting invitations can then be sent out to all attendees and their acceptance statistics can be collated easily. Such a group scheduling capability will become very useful in such an ad-hoc group whereby the collaborating members straddle disparate corporate boundaries.

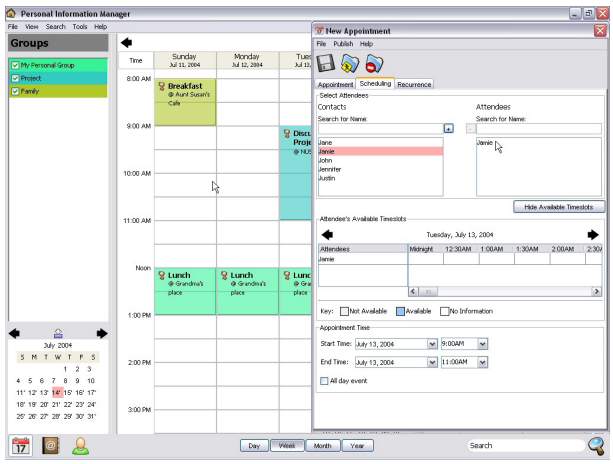

Fig. 6: *A screenshot of the secured group calendar scheduling application. The meeting organizer can browse for the attendees' available time slots. Changes in the meeting appointments can then be automatically synchronized with the other members within the VPG.*

In addition to scheduling group appointments, this application module also includes a shareable contacts management capability. Each peer can organize his or her contacts into multiple categories, and may choose to automatically

subscribe to receive the latest updates of the contacts from the remote peers once changes have been made. The contacts owner may selectively publish only particular portions of his or her full contacts information via the creation of different virtual cards. Each virtual card is assigned with flexible publication rules to govern its subscription rights by other members. Similar to group calendaring, a peer can publish public virtual cards which can be subscribed by any peers on the P2P eco-system.

Scalable Multimedia Communications and Ad-Hoc Group Collaboration

This application module focuses on adding real-time group communications and collaboration capabilities by exploiting our highly scalable multiwavelet-based image and video compression system [1], [2]. An earlier version of the scalable video codec has been employed for the development of a layered video multicasting system to heterogeneous wired and wireless networks over the IP multicast backbone [4]. In this module, we focus on exploiting the scalable video codec for adaptive real-time video communications over an ad-hoc P2P network environment. In this implementation, different layers of the multi-scalable compressed video bit stream are multicast to multiple media sub-groups, each carrying a particular layer of the compressed video. Each peer can then selectively tune into the best possible number of media sub-groups, as limited only by the capabilities of the peer device and network connection bandwidth. In this case, the collaborating peers with different peer devices on diverse networks can still communicate seamlessly, while each peer is dynamically selecting his or her best combinations of display resolution, playback frame rate, color depth, encoding/decoding processing complexity, and streaming bit rate, which are originating from the same single source of compressed video. For example, a wireless peer may selectively tune into a low resolution and low frame rate version, while a broadband peer can receive a high-fidelity version of the scalable video multicast. Multiparty voice conferencing among peers in ad-hoc VPG is also a part of this module.

Fig. 7: *A screenshot of the 3-D graphics modeling and navigation software of SGI VizServer. This tool will be integrated into the P2P platform to develop a group collaborative visualization application.*

In addition to voice and video streaming, the scalable codec can also be applied for progressive image browsing of shared image libraries within the VPGs. Members having the appropriate access rights can quickly and progressively browse through a large collection of compressed image archives. During the browsing process, the peer can choose to browse a low-resolution thumbnail version of the images or to progressively stream a blur-to-fine version of the original images. With this module, peers can easily share compressed images and graphics in an ad-hoc community manner with full security access control. In addition, multiple collaborating peers within a VPG can also co-navigate and co-edit a large image together. Some useful imaging applications here include collaboration on large medical images, graphical designs and modeling, satellite images and maps, etc. Extension for collaborative real-time graphics modeling is being explored too. Here, the peers are co-navigating and interacting in real-time to render a complex 3-D model via a centralized high-performance graphics server using the SGI VizServer (see Figure 7). Depending on the collaborating peers' input commands, the VizServer will render the 3-D model in real-time, and then the graphic image sequences are compressed using the scalable image codec before being delivered to the peers. Similarly, all communications are encrypted and the collaborative group session can easily be set up in a truly ad-hoc manner.

4. CONCLUSIONS

In this paper, we have proposed and developed a comprehensive and extensible peer-to-peer platform which directly connects all peers into secured ad-hoc virtual peer groups. Coupling this with different useful application tools such as integrated buddy roster and chat, secure file sharing, searching and retrieval, synchronized shared calendars and contacts, and scalable multimedia group communication and interactive collaboration, the distributed P2P platform is poised for many strategic vertical industries, such as e-Education, e-Sharing, e-Business, and e-Collaboration applications. Extensions to developing other XML web services tools such as online shopping, P2P auctioning, and universal searching via the Amazon's and Googles' Web Services APIs can also be well suited for developing a more unified distributed ad-hoc P2P eco-system. Other useful tools include whiteboard sharing, survey and polling, threaded discussion groups, and bulletin boards. In a nutshell, this platform presents a practical and visionary starting point for delving into many challenging and open research problems related to P2P security, distributed and reliable databases and file storage systems, optimization of ad-hoc network routing and quality of service, multimedia streaming in ad-hoc networks, grid computing, and many more.

5. REFERENCES

[1] J. Y. Tham, "Multiwavelets and Scalable Video Compression," Ph.D. Dissertation, Department of Electrical and Computer Engineering, National University of Singapore, 2002. (Downloadable from: http://wavelet.cwaip.nus.edu.sg/thamjy/)

[2] J. Y. Tham, S. Ranganath, A. K. Kassim, "Highly Scalable Wavelet-Based Video Codec for Very Low Bit Rate Environment," *IEEE Journal on Selected Areas in Communications -- Special Issue on Very Low Bit-rate Video Coding*, vol. 16, no. 1, pp. 12-27, Jan. 1998.

[3] Groove Networks at http://www.groove.net/

[4] Scalable Video Multicast Over Diverse Networks at http://www.eng.nus.edu.sg/EResnews/1003/rd/rd_9.html

[5] M. O. Rabin, "Efficient dispersal of information for security, load balancing, and fault tolerance," Journal of the ACM, vol. 36, no. 2, pp. 335-348, April 1989.
[6] D. Boneh, and M. Franklin, "Identity based encryption from the Weil pairing," SIAM Journal of Computing, vol. 32, no. 3, pp. 586-615, 2003.
[7] Project JXTA, Sun Microsystems at http://www.jxta.org/
[8] Jabber XMPP Protocols at http://www.jabber.org/protocol/
[9] XML Web Services Standards at http://www.w3.org/2002/ws/

Tham Jo-Yew is currently a Research Fellow with the Centre for Industrial Mathematics, National University of Singapore (NUS). He was a recipient of the ASEAN scholarship, and obtained his bachelor degree and Ph.D. in Electrical and Computer Engineering, NUS, in 1995 and 2002 respectively. Between 1996 and 1999, he was a Research Associate with the NUS Centre for Wavelets, Approximation and Information Processing. He later spent three years with two multimedia high-tech startup companies in the Silicon Valley, USA, as Senior R&D Manager and Chief Streaming Architect/Chief Technology Officer. His current research focus includes distributed peer-to-peer and grid computing, distributed information security infrastructure, scalable multimedia compression, group collaboration, and knowledge mining, sharing and rights protection. He is also a member of MENSA and IEEE.