

SOVoIP: True Convergence of Data and Voice Network

M.J.Arif, S.Karunasekera

Department of Computer Science and Software Engineering,
The University of Melbourne
Melbourne, Australia

{marif, shanika}@csse.unimelb.edu.au

S.Kulkarni

National ICT Australia, Victoria Research
Laboratories, Australia

santoshk@csse.unimelb.edu.au

ABSTRACT

Voice Over Internet Protocol (VoIP) is the reality of future telephony provided over the Internet. In the process it is claimed that, voice and data networks are converging into one network. However, such claim only introduced new protocols for telephony on the Internet which are not truly interoperable with existing protocols. Thus we propose a service oriented VoIP architecture, Service Oriented VoIP (SOVoIP), which not only ensures inclusive convergence of Internet and telephony but also ensures Quality of Service (QoS), Enhanced 911 (E911), Communication Assistance for Law Enforcement Act (CALEA), NAT and firewall traversal issue.

Categories and Subject Descriptors

C.2.4 [Computer Systems Organization]: Computer-Communication Networks-Distributed Systems, Client/server, Distributed applications.

General Terms

Design, Experimentation, Algorithms.

Keywords

Web Service, VoIP, Peer-to-Peer.

1. INTRODUCTION

The Internet was designed as a network to support data traffic. However, as the need for different applications like voice and video grew, there was a compelling need to upgrade the Internet. As a result, the community took the challenge to make the Internet behave more like a telephony network yet retain its characteristics to support data. The outcome is an improved Internet where data, voice and video coexist. This merging has been achieved by introducing new protocols for the voice and video architecture on the Internet which is broadly known as VoIP. The functionality of VoIP has been achieved by protocols such as Session Initiation Protocol (SIP), H.323 and Media Gateway Control Protocol (MGCP) etc.

However, these protocols merely ensured multimedia operation on the Internet. There is no interoperability, between them or other existing protocols on the Internet. As a result, we are yet to experience a real converged data and voice network. Thus we propose a web services based peer-to-peer (p2p) architecture for VoIP, Service Oriented VoIP (SOVoIP). SOVoIP ensures universality for all the protocols both from telephony and Internet. Thus voice and data networks converge in its true sense.

SOVoIP supports features such as extensibility, mobility, QoS and security. Besides, it resolves NAT and firewall traversal in a simple manner. It also addresses issues such as E911 [20] and CALEA [21]. Introduction of Service Oriented Architecture (SOA) in VoIP makes organizational processes faster with less human interactions and expedite the decision making. This is because information from different source can be inserted in the decision support system by different means. In addition to that, it promises a lightweight client which consumes fewer client resources and does not require any configuration on the client side. It is also a cost-effective solution as changes in the architecture are transparent to clients. Thus, it is free from frequent upgrade costs to the client.

The rest of the paper is organized as follows: section 2 explores some existing solutions to VoIP, section 3 describes the SOVoIP architecture, section 4 illustrates features and benefits of the architecture, section 5 studies the feasibility and integration of SOVoIP to existing business models, section 6 discusses the prototype and section 7 concludes the paper.

2. RELATED WORK

This section describes some of the existing VoIP architectures.

Skype[32] is one of the most popular VoIP clients. Skype's client is self-contained and creates a p2p network. Skype maintains central login server for authentication [15]. Skype's protocol is proprietary and messaging is encrypted. Thus it does not communicate with other protocols. Our experiment suggests that Skype nodes continuously maintain UDP connections to surrounding nodes. However, TCP is used for call setup. As Skype clients need to do continuous processing, they are yet to succeed in the cell phone arena. Best of CPU and battery available today for mobile devices are not perfect match for Skype client. Skype's architecture is not suitable to provide E911 services.

H.323 [24] is an umbrella recommendation from International Telecommunication Union (ITU) that covers all aspects of

multimedia communication over the IP network. It is a part of the H.32x series of protocols that describes multimedia communication over other networks such as Integrated Services Digital Network (ISDN) and Public switched telephone network (PSTN). In addition to that with its binary encoding and requirement for deployment of extra configuration such as Gatekeepers and Multipoint Control Units (MCUs) makes it complex. It is more suitable as a protocol to interface with PSTN than to the Internet.

SIP[31] is a client-server, text based lightweight protocol that works both on UDP and TCP. SIP was developed by Internet Engineering Task Force (IETF) to setup, modify and tear down multimedia sessions over the Internet. Similar to H.323, SIP architecture requires extra hardware and software in the network such as proxy servers, redirect servers and registration servers. However, it is more affable with other Internet protocols. SIP is not a transfer protocol like HTTP, designed to carry large amounts of data. It does not define any specific mechanism for E911 service, NAT and firewall traversal. Its aim for switching to a p2p architecture [13] will increase its complexity similar to Skype. However for real voice data both H.323 and SIP use Real-Time Transport Protocol (RTP) [33].

All these standalone VoIP solutions are not interoperable. They do not provide the capability to integrate with existing data network unless all the devices interacting using the same protocol. The probable way to ensure interoperability between protocols is to deploy gateways in the network. This requires additional hardware to be deployed in the network and an increase in the software processing that leads to higher latency. They are merely different voice architectures over the Internet each with its advantages and disadvantages. SOVoIP proposes a common platform where underlying architecture and complexities are hidden from the application. In addition, different protocols can join SOVoIP, yet retain their protocol specific operation. Thus SOVoIP truly merges the voice and data network.

Previous studies [4, 6, 14] for integrating web services as an interoperability solution between protocols added extra layers in the signaling. Calls from SIP first go to Venice, their web service based project and then it is forwarded to H.323. This solution still possesses protocol specific complexities at either end. It requires interaction between two protocol specific web service interfaces in the middle. In contrast our interoperability solution uses the processing power of clients intelligently. It therefore, saves bandwidth and reduces network complexities. Moreover, this solution forces users to use their Venice project in the middle even when interoperability is not required.

3rd Generation Partnership Project (3GPP) [34] is a collaboration effort among partners in the context of mobile phone standards. The IP Multimedia Subsystem (IMS) defined by 3GPP is basically an overlay which uses SIP underneath in order to provide services over IP. However, SOVoIP exchanges simple universal messages over the Internet. Thus anyone can join in SOVoIP easily. Besides, SOVoIP architecture is such that it can fit itself in the existing Internet infrastructure and it consumes less client resources.

3. SOVoIP ARCHITECTURE

SOVoIP is a service oriented, p2p architecture for VoIP as shown

in Figure 1. There are two types of nodes: Public Nodes and Client Nodes.

Public nodes, which are available on the network, have public IP addresses, and are managed by VoIP service providers. They provide a web service interface to the clients and are responsible for most of the data processing associated with VoIP connectivity. In addition to authentication and IP look up, these nodes are responsible for managing tasks such as NAT and firewall traversal and addressing the E911 problem. Public nodes are high configuration machines that have sufficient bandwidth to support simultaneous requests from various client nodes. The concept of a public node is similar to a cell in a cellular network. Public nodes cover a specific area. We considered one public node to serve a particular post code. Public nodes have a local database which maintains information regarding clients assigned to its postcode. Public nodes form a P2P network with similar public nodes. They can provide proactive suggestion, if needed, regarding media codecs or network congestion in both sides, to the clients before call setup starts.

A client node, where the VoIP software is installed, can be a PC or a handheld VoIP enabled phone. Client nodes connect to a web service at a public node via standard web service interface for registration, authentication, user search and billing. They communicate via SOAP [10] over HTTP [25]. Each client node is associated with a public node that operates as the home node for the client. Home node for the client is selected based on the postcode information provided during the first registration. All information related to a client is stored at the client's home node. During registration a client node shares the following information with its home node: User ID, Signaling IP (the IP address of the client); Signaling Port (port client uses to listen for incoming call); Media Port (port where the actual voice data will arrive); Media Capability (encoding information of the client). In some cases it may be necessary to share the physical address as well.

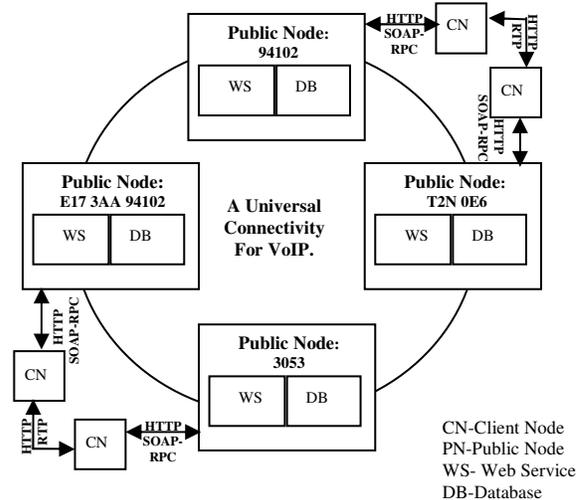


Figure 1: Web Service Based SOVoIP Architecture.

At start up, clients authenticate themselves with their home node. Once authenticated, clients communicate with the nearest public node for call setup and to access supplementary services. After authentication, the home node is responsible for redirecting the

user to the nearest public node. Communication between the calling and called party occurs directly, using HTTP in a p2p manner. Media communication between clients uses RTP. Some public nodes also play the role of a Global Public Node (GPN). These nodes are in the network with fixed addresses. In case a client node cannot locate its home node or is yet to have a home node, it will contact a GPN which will locate the home-node of the client.

If a user wants to make a call to another user, the caller sends the destination user id to the nearest public node and obtains all the information regarding the destination user. Figure 2 depicts call setup between two clients that are under different public and home nodes. PN1 and PN2 are the nearest public and home node for CN1 respectively. On the other hand, PN3 is the home node of CN2. As the figure shows, caller and callee are from two different home-nodes. Thus Domain Name System (DNS) (between step 1 to 5) is used to locate the called party's home node. SOVoIP with home node concept and a postcode based domain system experience first user search. Since the user search is based on the user's home node, the search performance does not get impacted based on the user's connection via Dynamic Host Configuration Protocol (DHCP).

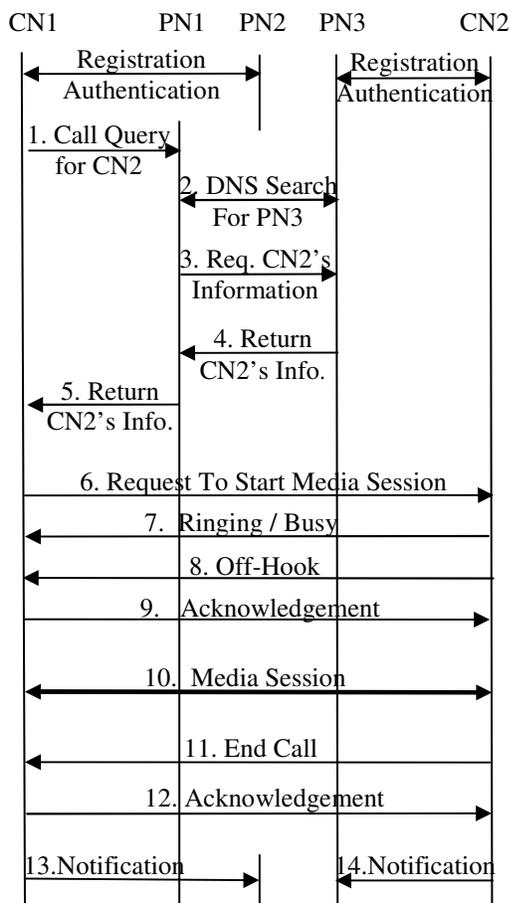


Figure 2: Call Establishment In SOVoIP Architecture.

A client listens to a HTTP port for incoming messages. When the calling party has all the information of the destination user, it sends a HTTP request directly to the destination node and the called node responds back with a busy or ringing response. Called node obtains the caller information from this HTTP request. Once these initial messages (step 6 to 9) have been exchanged they start sending real voice data over RTP directly. The call setup scenario may differ slightly when it has NAT or firewall between two clients. At the end of the call, client node will notify its corresponding home-node for billing purpose. One noticeable point is, CN1 completes its registration and authentication with its home node PN2 but for call setup it communicates directly to the nearest public node PN1.

4. FEATURES and BENEFITS

In this section, we describe the major features and benefits of SOVoIP.

SOVoIP has a thin client. Peer-to-peer nature of call setup between clients reduces the probability of failure. It is deployable to the existing hardware and software. Addition of new features are transparent to the clients. Thus it ensures extensibility and saves client upgrade cost. Text based encoding and complete modularity makes the development process much easier.

In SOVoIP, users can move to different terminals and still achieve service mobility with the help of the home node concept (for example access to the address book). GPN can further help users achieve mobility. Different terminal addresses can be used against a single user ID. Moreover, study [1] suggests smooth passages from one public node to another in a wireless network can be achieved with the help of DNS.

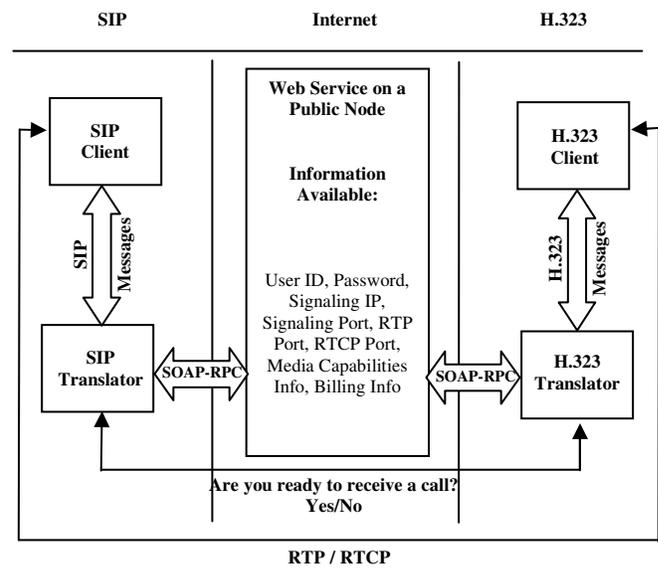


Figure 3: Solution for Interoperability in SOVoIP.

Interoperability is desired between different VoIP protocols. Our solution to address interoperability uses a simple protocol translator at the client end. This approach does not require an upgrade to the network. All the VoIP protocols in use today,

require three components of information shared during call setup. They are signaling destination address, local and remote media transport address and local and remote media capabilities [12]. The protocol translators collect these three information from the actual clients and have them for future use. Figure 3, shows a schematic diagram of how interoperability between SIP and H.323 can be achieved. The figure shows, a SIP client and a SIP translator running on one node and a H.323 client and a H.323 translator running on another node. Both SIP and H.323 client are conventional clients. On the other hand, SIP and H.323 translators are protocol translators that convert protocol specific messages to SOAP messages. These translators sit in the same node as the clients. Each translator acts like a receiver soft-phone and the actual clients share all the signaling with these translators. Both the translators talk to their respective client at one end while listening to a port for respective signaling messages. On the other end, they listen to a HTTP port for an incoming call. One of the advantages of such an approach is that users themselves can switch back to its original architecture if interoperability is not required. Thus all the users are not forced to use the same protocol in order to have interoperability. However, this time the translators setup the call rather than the clients. If the handshake between two translators from two clients completes, then the RTP stream will flow between actual clients. Another advantage of this deployed interoperability solution is that it does not require protocol dependent software and extra hardware on the network since we have used the ever growing processing power of client nodes. This results in reducing the network congestion and therefore lowers the latency.

In SOVoIP, media and signaling flow are independent of each other. SOVoIP deals with the signaling phase only. Thus introduction of SOA in the architecture does not create any performance drop for voice quality. Moreover, with the distributed public nodes it is unlikely to face bottleneck to access SOVoIP network. In addition to that client nodes talk to the public nodes for a small fraction of time and the public nodes are not involved in the entire call setup process.

In order to start a conversation between nodes and to use the web services functionality, authentication is required. Nodes can communicate between each other over a secure channel i.e. on HTTPS. As other applications can very easily hook up into this web services based architecture, it is proposed to implement Security Assertions Markup Language (SAML) [10] security protocol for authentication. The SAML assertion is then passed between various applications and used as a proof that you really are who you say you are. Moreover, SOVoIP implements CALEA transparently, where legal system can have access to the user's conversation. It appends legal system's address along with the response of user's call initiation request to the public node.

SOVoIP follows mesh topology for conferencing where every client node has a relationship with every other client node in the conference. However, we are extending it to a star topology in order to use the power of our public nodes.

NAT and firewall traversal has been a critical problem for VoIP. In SOVoIP public nodes help the client nodes overcome NAT problems. Client nodes stamp their non-routable private IP address in its packet payload. When client nodes call the registration interface of a public node, the public node checks the

IP address with the address in the packet payload. If the addresses are different, there is a NAT between them. Then the non-routable IP address is replaced with the actual global IP address and the global address is saved in the database. On the other hand, the client is notified about the public IP, so for future call establishment between two clients the public IP is used. This ensures that the call is established using a public, routable address and ports, ensuring end-to-end connectivity. In order to deal with the symmetric NAT it is proposed to use the integrated TURN server in the public nodes. Public nodes are capable of relaying both signaling and media data in case it is not possible to establish a direct connection between two client nodes. Furthermore, as SOVoIP uses HTTP as its transport protocol, it is highly unlikely to face problem with firewall. A future topic of study is to explore the need to add keep-alive feature for the clients.

E911 is a feature of the 911 emergency calling system in U.S.A that automatically associates a physical address with the calling party's telephone number. As IP has no geographical mapping it is hard to track a user in emergency calling. However in SOVoIP, as the users communicate to the nearest public node for call setup, it is always easy to track the last mile of the users. In SOVoIP architecture the geographic location of public nodes are known. Thus it is proposed, the location of the client node is approximated with technique such as IP2Geo [18] or Constraint-Based Geolocation (CBG) [7]. As a result, SOVoIP is able to redirect the user to the nearest public node and estimate user's location. Thus both fixed line and wireless users can be tracked at least up to the postcode or closer. Most importantly with SOA all the emergency organization can be hook up easily to SOVoIP. Thus, less human involvement will fasten the whole emergency service.

5. FEASIBILITY and FUTURE BUSINESS MODEL

Feasibility study of SOVoIP ascertains the likelihood of its success. Current technological advancement and available standalone voice solutions have rendered the current systems trivial. Expanding businesses with extra workload confronting customers complains about the speed and quality of work businesses provide. In addition to that competitors are always looking for opportunities to grab big market share. Recent Yahoo and MSN merge is such a consequence. On the contrary SOVoIP brings all the platforms, protocols and devices into a single stage. SOVoIP's feasibility attribute can be seen as:

- Unified model such as SOVoIP brings speed and efficiency in the business process. As people from different levels of management can access the system from anywhere, with whatever communication medium they have. That influences the customer satisfaction and ensures economical feasibility of the infrastructure.
- Unlike SIP or any other VoIP architecture SOVoIP does not require any network configuration in the client end. Only the providers need to maintain the network. Moreover its firewall friendly behavior brings simplicity to the system. Furthermore, HTTP is the most adopted protocol on the Internet. Thus universality of SOAP combined with HTTP, not only ensures interoperability between protocols but also make these protocols more open to the Internet. In addition to that QoS and security measurement ensure technical feasibility of the

architecture. Finally, SOVoIP is an application layer solution. Thus adoption to future IPV6 network will not cause any problem.

- SOVoIP's feasibility attribute, time, is perfect. SOA is maturing day by day. On the other hand SOVoIP does not require any especial hardware or software. Thus its time-to-market is predicted to beat the competition.
- Low client resource consumption and ability to cope up with changes in the architecture without client upgrade pledge extendibility and scalability. Thus organizational feasibility ensured.

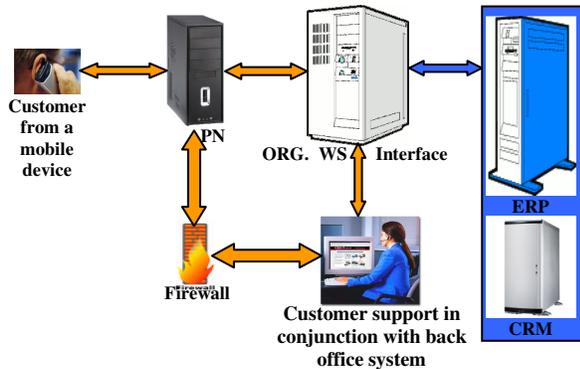


Figure 4: SOVoIP Business Model.

Figure 4 depicts a sample business model in conjunction with SOVoIP. Here a mobile customer from SOVoIP made a query to the organization. Based on the customer query it checked with the backend office system and redirected the user to the appropriate help. The aid was in outdoor and behind a firewall. Still could assist the customer from available soft-phone. During the call customer data was available in the soft-phone via the web service interface of the organization.

6. PROTOTYPE

We have deployed a prototype for SOVoIP using JAVA over the LAN. The web service is developed by JAVA Web Services Developer Pack (JWSDP) [27] and is described via a standard Web Service Definition Language (WSDL). This web service serves as a public node. We have developed a client which invokes the methods from the public node. For media processing the clients use JAVA Media Framework (JMF) [26]. However, in order to demonstrate interoperability we had to implement both SIP and H.323 protocol stack. For the SIP stack, we have chosen JAIN SIP developed by NIST [28]. For H.323, we have used OOH323C (Objective Open H.323 for C) developed by Objective Systems. Inc.[29]. Ethereal [23] is used to observe and measure traffic flows between various nodes of Skype, SIP and SOVoIP.

The primary test results are promising. As expected our client consumes approximately 400KB of memory compared to 4500KB used by Skype. Total call setup time between two clients is 1.0887975 seconds including user search. Method invocation is found to be reliable and can be on HTTPS. Two clients were able to share HTTP messages between them in a p2p manner. Once the handshake between clients went well they exchanged RTP media packet directly. As SOVoIP is focused on the signaling we

have not done any measurement on the media quality. We have used default encodings of JMF version 2.1.1e. However, we found both signaling and media performance are acceptable. Finally, our protocol interoperability solution found to be very reliable and consistent.

7. CONCLUSION and FUTURE WORK

In this paper we have presented SOVoIP, a web services based VoIP architecture. We described how SOVoIP works and how it merges the voice and data traffic seamlessly. We tried to explore its economical, technical, organizational and competitive feasibility.

We are currently focusing on interoperability with Public Switched Telephony Network (PSTN). Our E911, conferencing, NAT and firewall traversal demand improvement and real time testing. We are integrating remote fixing feature to SOVoIP. Where fixing of software faults or upgrade in the public nodes can be done remotely and without any interruption. Of course, it can only be applied to certain kinds of paradigm. We are also evaluating the choice of web services over other existing middleware. Primarily firewall friendliness and simplicity of web services influenced the decision.

However SOVoIP is an architecture which will make sure internetworking between office, home and any other external information system and accessibility to these systems via the available hand held device.

8. REFERENCES

- [1] Ahmed, A.; Reaz, S.; Atiquzzaman, M.; Fu, S. Performance of DNS as location manager. . In the *2005 IEEE International Conference on Electro Information Technology*.
- [2] Alan B. Johnston. SIP Understanding the Session Initiation Protocol
- [3] Baset, S. A., and Schulzrinne, H. An Analysis of the Skype Peer-to-Peer Internet Telephony Protocol. In *Proceedings of the INFOCOM '06* (Barcelona, Spain, Apr. 2006).
- [4] G. Zhang and M. Hillenbrand. Implementing SIP and H.323 Signaling as Web Services. In *Proceedings of the 30th Euromicro Conference 2004 (Rennes, France)*, 8,9 2004.
- [5] G. Fox, W.Wu, A. Uyar, H. Bulut, and S. Pallickara. A Web Services Framework for Collaboration and Videoconferencing. In *Proceedings of the 4th International Conference on Internet Computing 2003, Las Vegas, USA*, 6 2003.
- [6] Ge Zhang, Markus Hillenbrand, Paul Müller. Facilitating the Interoperability among Different VoIP Protocols with VoIP Web Services. In *Proceedings of the First International Conference on Distributed Frameworks for Multimedia Applications (DFMA'05)*
- [7] Gueye, B.; Ziviani, A.; Crovella, M.; Fdida, S. Constraint-Based Geolocation of Internet Hosts. *Networking, IEEE/ACM Transactions on Volume 14, Issue 6, Dec. 2006* Page(s):1219 – 1232.
- [8] H. Schulzrinne and J. Rosenberg. A Comparison of SIP and H.323 for Internet Telephony. In *Proceedings of Network and Operating System Support for Digital Audio and Video NOSSDAV (Cambridge, England)*, 7 1998.

- [9] Henry Sinnreich and Alan B. Johnston. Internet Communications Using SIP.
- [10] James Snell, Doug Tidwell and Pavel Kulchenko. Programming Web Services with SOAP.
- [11] Jorge Salas, Francisco Perez-Sorrosal, Marta Patiño-Martínez, Ricardo Jiménez-Peris. WS-Replication: A Framework for Highly Available We Services. In Proceedings of the 15th international conference on World Wide Web WWW '06.
- [12] K. Singh and H. Schulzrinne. Interworking Between SIP/SDP and H.323. In *Proceedings of the 1st IP-Telephony Workshop IPTel*, 4 2000.
- [13] Kundan Singh, Henning Schulzrinne. Peer-to-Peer Internet Telephony using SIP. In Proceedings of the International workshop on Network and operating systems support for digital audio and video NOSSDAV '05.
- [14] Markus Hillenbrand, Ge Zhang, "A Web Service Based framework for Voice over IP", *Proceeding of the 30th EUROMICRO Conference*, August 2004
- [15] Paul L. Piccard, Brian Baskin, Craig Edwards, George Spillman and Marcus H. Sachs. Securing IM and P2P Applications for the Enterprise.
- [16] Saikat Guha and Neil Daswani. An Experimental Study of the Skype Peer-to-Peer VoIP System. <http://saikat.guha.cc/pub/iptps06-skype/>
- [17] V. Pappas, D. Massey, A. Terzis, L. Zhang. A Comparative Study of the DNS Design with DHT-Based Alternatives. In the Proceedings of IEEE INFOCOM'06, Apr. 2006.
- [18] Venkata N. Padmanabhan, Lakshminarayanan Subramanian. An investigation of geographic mapping techniques for internet hosts. In the Proceedings of ACM SIGCOMM Computer Communication Review, August 2001.
- [19] Wang, S.S.; Green, M.; Malkawa, M. E-911 location standards and location commercial services. Emerging Technologies Symposium: Broadband, Wireless Internet Access, IEEE. April 2000.
- [20] 911 Services: <http://www.fcc.gov/911/>
- [21] CALEA (Communication Assistance for Law Enforcement Act): <http://www.askcalea.net/calea/>
- [22] Delivering Superior Self Service with Open Standards. VoiceXML and the Future of Service Oriented Architectures. AVAYA White Paper. May 2005. <http://www.avaya.com/>
- [23] Ethereal: <http://www.ethereal.com/>
- [24] H.323 <http://www.h323forum.org/>
- [25] HTTP <http://www.w3.org/Protocols/>
- [26] JMF (JAVA Media Framework) <http://java.sun.com/products/java-media/jmf/>
- [27] JWSDP (Java Web Services Developer Pack) <http://java.sun.com/webservices/jwsdp/index.jsp>
- [28] NIST SIP - <http://snad.ncsl.nist.gov/proj/iptel/>
- [29] OOH323C (Objective Open H.323 for C) <http://www.obj-sys.com/open/index.shtml>
- [30] Self Service Solutions for IP Telephony and Service Oriented Architectures. AVAYA White Paper November, 2005. <http://www.avaya.com/>
- [31] SIP <http://www.ietf.org/rfc/rfc3261.txt>
- [32] Skype <http://www.skype.com/>
- [33] RTP <http://tools.ietf.org/html/rfc3550>
- [34] Third Generation Partnership Project (3GPP) <http://www.3gpp.org/>