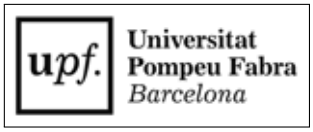
Architecture Design of Real-Time Communication for Organizations with WebRTC

Pedro Vílchez

TFG UPF / YEAR 2015

DIRECTOR/S OF THE TFG:
Miquel Oliver, Victor Pascual
DEPARTMENT:
Departament de Tecnologies de la Informació i les Comunicacions (DTIC)



Dedicated to my family.

Acknowledgments

Special thanks to Victor Pascual and Miquel Oliver for his mentorship. Thanks to Victor Oncins and Angel Elena (craem) for his feedback and help.

Thanks to Daniel Pocock for its work on rtcquickstart.org. Thanks to webrtchacks.com and all its team for the useful articles.

Thanks to all the people that works for the democratization of communications

Thanks for reading. Thanks for your time.

Abstract

The present project introduces the disrupting technology WebRTC (Web Real-Time Communication), that supports browser-to-browser applications without need of third party plugins. It is detailed how, since its release by Google in 2011, it is evolving and changing the way communications are understood. How to materialise a Real Time Communications in organizations with WebRTC and the use case of video and audio calls, taking as example Guifi.net and the opportunities that it offers: requirements, architecture design, component selection, implementation and demo.

Resum

Aquest projecte introdueix la tecnologia disruptiva WebRTC (comunicació web en temps real), que suporta aplicacions de navegador a navegador sense la necessitat de complements adicionals. Es detalla com, des de que va ser alliberat per Google al 2011, està evolucionant i canviant la forma en que les comunicacions són enteses. Com materialitzar les comunicacions en temps real en organitzacions amb WebRTC i el cas d'ús de trucades de veu i vídeo, prenent com exemple Guifi.net i les oportunitats que ofereix: requeriments, disseny d'arquitectura, selecció de components, implementació i demostració.

Resumen

Este proyecto introduce la tecnología disruptiva WebRTC (comunicación web en tiempo real), que soporta aplicaciones de navegador a navegador sin necesidad de complementos adicionales. Se detalla cómo, desde que fue liberado por Google en el 2011, está evolucionando y cambiando la forma en que son entendidas las comunicaciones. Cómo materializar las comunicaciones en tiempo real en organizaciones con WebRTC y el caso de uso de llamadas de voz y vídeo, tomando como ejemplo Guifi.net y las oportunidades que ofrece: requerimientos, diseño de la arquitectura, selección de componentes, implementación y demostración.

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

A project presentation, motivation, objectives and outline of the chapters.

1.1 Motivation

In my previous Bachelor's thesis [Vílchez, 2014] I developed two ideas: *together we can build a commons network* and *it is simple to start*. This project explores how to provide Real-Time Communications (RTC) to a commons network such as Guifi.net. Very few large cities have a commons network. Barcelona's commons network is called Guifi.net. However, the majority of Barcelona's population has no knowledge nor understanding of Guifi.net. Since the communications market is already highly commercial and competitive, users have a very wide choice of providers and therefore no need to think any further. In areas such as smaller cities and towns where large telecommunication companies have not invested in infrastructure, there is a far greater incentive to participate in a commons network.

Inside a commons network, the implementation of other services besides Internet inherently adds value to the network and could help people understand its advantages. There are services based on contents which are difficult to maintain. But RTC as a service can be easily applied on a large scale. Once RTC has been set up, it is merely a question of time before people start using it. For various reasons a commons network generally functions within a narrow radius, and an RTC system could enhance social cohesion within the community.

To apply an RTC system to Guifi.net's commons network is a challenge. Attempts to implement an RTC system have failed. WebRTC is a new opportunity to try again.

1.2 Objectives

- Free and secure communication between users via an RTC system and commons network infrastructure.
- Backward compatibility with VoIP[1] network. Hence, users can communicate to other VoIP operators from inside and/or outside Guifi.net.
- Designing RTC architecture to fit the commons network scenario.
- Ease of installation and use of RTC.

1.3 Outline

This document is organised in 5 chapters.

Chapter 1 Introduction introduces the project. The motivation to start this project and the general objectives to achieve.

Chapter 2 Fundamentals of RTC presents the basics of the concepts communications and real-time quality measurements.

Chapter 3 State of the art discusses the state of the art of WebRTC and its associated technologies.

Chapter 4 Methodology contains a description of the processes involved to achieve this project.

Chapter 5 Contributions and Results develops the theory and practice of the objectives proposed by this project

Chapter 6 Conclusions and Future Work provides a general assessment of the project

[1] Voice over IP. The use of telephone adapted to the Internet network.

Chapter 2

FUNDAMENTALS OF RTC

The generic concepts and infrastructures that make possible real-time communications and important quality parameters to consider.

2.1 Communications

In order to design communications in a system, the different elements must know what to do with the messages. That is why it is needed protocols. A protocol defines the format and the order of messages exchanged between two or more communicating entities, as well as the actions taken on the transmission and/or receipt of a message or other event.

Real-time communications are even more complex, the information has to arrive fast and it is convenient to prepare the receiver of the communication. Signalling is the process of sending control information over networks to monitor, control, route, and set up sessions between devices. These sessions include video and audio conference calls, data sessions, video calls, and mobile and landline telephone calls. Signalling is also used to set up instant messaging and chat sessions. A signalling protocol, is a protocol that needs signalling features.

There are different approaches to achieve real-time systems, that is why it is needed compatibility between the systems. A gateway allows equipment with different protocols to communicate with one another. For example, gateways are used when incompatible video systems are used for a video conference.

It is also important how to transport the data, there are two fundamental approaches: circuit switching and packet switching.

Packet switching and the Internet. The messages are divided in appropriate sized chunks of data known as packets. Packets travel from the source through a network, when they arrive at its destination, the messages are reconstructed. The network makes its best effort to deliver packets in a timely manner, but it does not

make any guarantees. Because the resources of the network are used on demand, not reserved, as a consequence, a packet may have to wait for access a communication link. Each packet could be routed in a different network path depending on the network state. Some of them could be lost if there is a congestion (a lot of traffic, a large number of packets). The network equipment needed to build a packet switching infrastructure is affordable, flexible and can manage easily a large amount of data (throughput). The internet is an example of a network of networks that consists of lots of private, public, academic, business, and government networks of local to global scope. The technical operation and standardization of the Internet is done with the RFC's[1] (Request for Comments) and is an activity of the IETF (Internet Engineering Task Force). The protocol stack of the Internet is the TCP/IP stack, the name was given due to the importance of this two protocols on the system.

Circuit switching and traditional telephony. The resource (a network path composed of communication channels between source and destination) is reserved for the duration of the communication session between the endpoints. In one hand, is a guaranteed constant transmission rate, good link quality, in the other hand, is an expensive resource that is being wasted in silent periods. One of the techniques used to reduce the silent periods is the multiplexing of the circuits in time (Time-Division Multiplexing, TDM) or frequency (Frequency-Division Multiplexing, FDM). The Public Switched Telephone Network (PSTN) is an example of an aggregate network operated by national, regional and local telephony operators that used this type of network paradigm in the past. Now it is moving towards a packet switching network. The technical operation of the PSTN uses the standards created by the ITU (International Telecommunication Union[2]). The protocol stack of PSTN network is the SS7 stack (Signalling System No. 7). Later, it was added the possibility to include generic data to its networks with ISDN (Integrated Services Digital Network).

IETF and ITU did efforts to adapt its networks to the requested uses: transport of generic data and real-time data. Internet is becoming the standard way to transport any kind of data. IETF did operations to include traditional telecommunication operators inside Internet, for example SIGTRAN[3] family of protocols (compatibility with SS7 and ISDN stacks). There is also the Quality of Service (QoS) concept, that gives priority to chosen packets to arrive faster. That is why Internet can manage real-time data with reasonable delay.

SS7, ISDN and TCP/IP stacks are based on the Open Systems Interconnection model (OSI model[4]). Each protocol belongs to one layer. Each layer provides its

[1] In this document each IETF protocol is accompanied by its RFC.
[2] Formerly the International Telegraph Union.
[3] Derived from signalling transport, working group of IETF.
[4] ISO/IEC 7498-1:1994.

service by performing certain actions within that layer and by using the services of the layer directly below it, this is called the service model. Table 2.1 shows the difference between the and OSI model and the TCP/IP stack, each row is a layer. In TCP/IP, physical layer is implementated with hardware, for example with NIC (Network Interface Controller) or WNIC (Wireless Network Interface Controller), generally uses the IEEE 802 family, and its standardization is managed by IEEE (Institute of Electrical and Electronics Engineers.). Link layer is typically implemented by an element called switch, they provide close connectivity. Network layer is typically implemented by an element called router, they provide far connectivity. Session and presentation layers has to be implemented by the developer of the application (↗).

Table 2.1: OSI, TCP layers and its description.

OSI layer	Description	TCP/IP layer
Application	Network and application services	Application
Presentation	Data format	↗
Session	Signalling of data	↗
Transport	Connection establishment side to side	Transport
Network	Logic addressing (far)	Network
Link	Physical addressing (close)	Link
Physical	Binary signal and transmission	Physical

Table 2.2 shows the name of each Protocol Data Unit (PDU), the most important protocols and the organization that standarizes it. At link layer the most common PDU is the Ethernet frame, with its associated MAC address (Media Access Control). At the network layer the datagram, with its associated IP address (Internet Protocol). The version 4 of IP, IPv4 [Postel, 1981a], contains 2^{32} (4.3 billion) addresses, as this is not enough, that is why Internet network is upgrading to IPv6 [Deering and Hinden, 1998] with 2^{128} (more than $7.9 \cdot 10^{28}$ times as many as IPv4). Meanwhile, there are basically two types of IPv4 addresses: private IP [Rekhter et al., 1996], not reachable via Internet, and public IP, reachable via Internet. It is difficult to remember IP addresses to access different locations, that is why there is DNS (Domain Name Server, [Mockapetris, 1987]), it associates human readable addresses (domain) to IP's. An IP adddress includes a port in the range of 0 to 65535, its main purpose is to share a single physical connection to another place (host), for example an IP destination can has more than one different services available; the different services is usually associated with a specific port. At the transport layer, one unreliable and fast protocol, UDP (User Datagram Protocol, [Postel, 1980]) and another reliable, TCP (Transmission Control Protocol, [Postel, 1981b]). For the real-time topic, UDP is massively used in the real-time

communication, because TCP favors reliability over timeliness. TCP is usually used for the signalling. At the application layer there are many possibilities, an example is the DNS, briefly described above.

Table 2.2: PDU associated with each layer in TCP/IP model

TCP/IP layer	PDU	Standardization
Application	Message	IETF
Transport	Datagram (UDP), Segment (TCP)	IETF
Network	Datagram or Packet (IP)	IETF
Link	Frame (MAC)	IEEE

Every PDU has two types of fields: header and payload. A PDU is encapsulated with the layer below (see Figure 2.1). If the header part is greater than payload part is called overhead. If it occurs, the communication is not very efficient, specially if they are sent lots of packets. Signalling data is overhead in the sense that it is not user data, but is useful to establish a call.

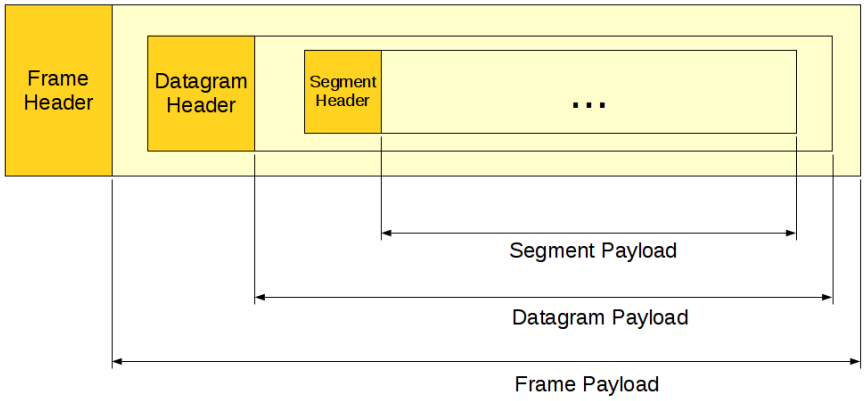


Figure 2.1: PDU encapsulation with IP and TCP protocols, application is not specified.

Before initializing a call is a requirement to convey media details to the participants, SDP (Session Description Protocol, [Handley et al., 2006]) provides a standard representation for such information.

During a call, there is the RTP (Real-time Transport Protocol, [Schulzrinne et al., 2003]). The majority of the RTP implementations are build on top of UDP. RTP is a protocol used to encapsulate multimedia content (audio, video). Its header contains a sequence number and timestamp. The sequence number increments by one for each RTP data packet sent, and may be used by the receiver to detect packet loss and to restore packet sequence. The timestamp registers the time when the RTP

packet has been generated, it helps with the synchronization and jitter[5] calculations. The RTP Control Protocol (RTCP) periodically sends packets with statistics information to participants in a streaming multimedia RTP session. RTP and RTCP provide information on the quality of communication but do nothing to fix or improve it. An application may use this information to control quality of service parameters. RTCP itself does not provide encryption[6] or authentication methods. If they are needed, it is available the SRTP (Secure Real-time Transport Protocol, [Baugher et al., 2004]).

More information about packet switching can be found at [Kurose and Ross, 2013] and about general telecommunications and signalling at [Dodd, 2012]. Some parts of that sources were included in this section.

2.2 Real-time quality parameters

The network is defined as the router, switch and link components that interconnect all the path between receiver and its destination (them included). The real-time communication depends on how the network is affected by the following parameters.

The bandwidth is the number of bits[7] that can arrive from the source to the destination through the network in one unit of time (for example, MB/s). A greater bandwidth delivers higher definition of the media content. A packet can be lost if it is never received at the destination. A communication suffers jitter when different packets arrive with different delays at its destination. It is defined delay as the time between transmitting the packet and arriving at its destination. Due to the nature of packet switching, packets can be delivered out of order. The arrived packet can present a corruption of data. The prior parameters presented affect negatively to the communication. Its undesirable effects might be silent lapses and interruptions of the communication.

Real-time data transport is managed with UDP. This protocol does not perform retransmission, it means that the communication should be fixed for the next upcoming packets. Real-time communications require low bandwidth and delay. This two requirements can be achieved easily with the technologies nowadays.

Some parts of [Velázquez, 2010] and [Frank, 2004] were included in this section. This sources are about voice and its quality, but it can be extended to certain forms of real-time video because of the improved technology nowadays.

[5] Jitter is briefly defined in the next section.

[6] An encryption system guarantees privacy, only authorized participants can read the content.

[7] A bit can be 0 or 1. A stream of this symbols represents information. In telecommunications the information unit is the bit, expressed with the unit b. In computer science the Byte, expressed with the unit B. 1 Byte is 8 bits.

Chapter 3

STATE OF THE ART

Important mechanisms and different technologies that use WebRTC and other associated technologies such as SIP and XMPP

3.1 Internet's evolution

The Internet is in constant evolution, and different mechanisms are necessary for solving problems affecting RTC.

Apart from IPv4 addresses to destination hosts, people can sign up with devices onto host services and be reachable via Internet. A clear example of this is the email system; bob@example.com means that the user, bob, is registered in the domain example.com that is reachable via Internet. DNS needs an additional parameter to reach bob; in the case of email the parameter is MX record (Mail eXchanger record, detailed in the SMTP[1] standard [Klensin, 2008]). This is a resource record that specifies a mail server responsible for accepting email messages on behalf of a recipient's domain. To advertise other network services such as SIP and XMPP, the SRV record (SeRVice record, [Gulbrandsen et al., 2000]) is necessary. The NAPTR record (Name Authority PoinTeR, [Mealling, 2002]) allows a simple contact address such as bob@example.com to contain the available forms of contact and its priority to different services.

In the current upgrade from IPv4 to IPv6, NAT (Network Address Translation, [Senie, 2002]) has been standarized in order to alleviate the scarcity and depletion of IPv4 addresses. NAT allows a public IP address to be reused among many different private networks (and its private IP addresses). It functions by connecting private local IP addresses and port tuples to one or more public IP addresses and port tuples. Unfortunately this introduces connectivity problems,

[1] Simple Mail Transfer Protocol, defines the electronic mail (e-mail) transmission.

especially in peer-to-peer connections[2] (p2p). There are two server utilities which can be configured into a device to solve these problems. Both are required to be on the public network. A STUN server (Session Traversal Utilities for NAT, [Rosenberg et al., 2008]) allows a device to find out its public IP address if the device is located behind a NAT. STUN service keeps the NAT binding alive. A TURN server (Traversal Using Relays around NAT, [Mahy et al., 2010]) acts as a transport address intermediary between the two people wishing to communicate. Unlike STUN, TURN is resource-intensive for the provider. The problem with these two techniques is that they are optimal in some network topologies but a poor choice in others. That is why the most interesting solution for NAT traversal is ICE (Interactive Connectivity Establishment, [Rosenberg, 2010]). ICE exchanges, using the offer/answer SDP model, a multiplicity of IP addresses and ports, that includes STUN and TURN, which are then tested for connectivity by peer-to-peer connectivity checks. The best candidate transport address is then selected.

The World Wide Web[3], the Web, is the most common way to use the Internet. It has a client-server distributed architecture model[4]. Web servers act as providers of resources or services and browsers as client requesters. It means that it follows a request-response communication model: client requests and server responses. The browsers navigates between different web pages provided by the web servers. The web navigation could be unencrypted with HTTP[5] (HyperText Transfer Protocol, [Fielding et al., 1999]) requests or encrypted with HTTPS (HTTP Secure, [Rescorla, 2000]). The success of the Web is promoting the use of web applications for all possible services. Two important features of web applications are that they are cross-platform, the browser is a default application in most of the operating systems, and easy to update or upgrade, the browsers only has to refresh the page to obtain a new version. An important problem of the request-response communication model was its complexity associated to the use of bidirectional communication or streaming [Loreto et al., 2011], for example to call through a browser. The solution is provided below (the Websocket protocol).

As Internet is such a vast, open and widely distributed network, it is therefore also open to undesirable consequences. For example, devices reachable via Internet are constantly attacked by other devices trying to exploit its weaknesses in computer systems and computer networks and thus can affect the security of an organization. For this and other reasons, its administrators deny by default any service which is not explicitly important to the organization. One of the compo-

[2] A connection that goes directly from a host to another host, without an intermediate server.

[3] An information system for interlinked hypertext documents and other digital resources that are accessed via the Internet

[4] `http://www.w3.org/Proposal.html`

[5] HTTP is a coordinated effort by the IETF and the World Wide Web Consortium (W3C).

nents used to control the incoming and outgoing network traffic on an applied rule set is the firewall. But the firewall itself can also have undesirable consequences; in large organizations it can be less flexible. Take as an example a university with students interested in computers who are experimenting with services related to courses or projects; they may be having connectivity problems due to the restrictive nature of the firewall.

Another common mechanism to control the networks is the proxy. A proxy is a network entity that acts as an intermediary between two communication entities. A web proxy satisfies HTTP requests on the behalf of an origin web server. A proxy has the ability to inspect traffic flowing through the service, and can ensure no traffic flows through to the Internet without inspection or control.

Firewalls and web proxies commonly admit traffic via ports 80 (HTTP) and 443 (HTTPS). The Websocket protocol [Fette and Melnikov, 2011] introduces bidirectional communication between a browser and a web server via ports 80, 443 or any other specified port. This provides websocket-based services an ability to traverse firewalls and proxies, and also prepares the way for a new generation of bidirectional applications, such as RTC.

3.2 Voice, video calls and instant messaging

There are two strong open standard alternatives to supply the use cases of voice, video calls and instant messaging: SIP and XMPP.

SIP (Session Initiation Protocol, [Rosenberg et al., 2002]) is a protocol used to establish, modify and terminate multimedia sessions in the Internet. It is one of the VoIP[6] available services, an alternative to H.323 from the ITU. It is text-based, on a request/response transaction model from HTTP and SMTP; for example encrypted SIP is SIPS, and it has similarities to the relation between HTTP and HTTPS. SIP itself provides a TCP or UDP signalling channel, SDP to agree a media session and a RTP media channel. SIMPLE[7] is a concluded workgroup at IETF that extended SIP with messaging and presence, this source offers an overview of the work done [Rosenberg, 2013]. Important telecommunication enterprises provides SIP trunking[8] such as Telefonica[9] and AT&T[10]. SIP have hardware client implementations called SIP-based VoIP phones, and software client

[6] Voice over IP, allows telephone calls via Internet.

[7] SIP for Instant Messaging and Presence Leveraging Extensions `https://tools.ietf.org/wg/simple/`.

[8] SIP trunking means interconnection to an operator through SIP.

[9] `https://www.globalsolutions.telefonica.com/en/wholesale/products-services/global-voice/sip-trunking/`.

[10] `http://www.business.att.com/enterprise/Service/voice-services/null/sip-trunking`.

implementations called softphones. Old telephones can have a SIP conversion with ATA (Analog Telephone Adapter).

XMPP[11] (Extensible Messaging and Presence Protocol), is a protocol to manage instant messaging and presence of your contact list, called roster in its terminology. It is defined in the IETF [Saint-Andre, 2011] and the XEP's (XMPP Extension Protocols) are worked by the XMPP Standards Foundation[12] (XSF). It is one of the IM (Instant Messaging) applications available. It encrypts its traffic with TLS such as SIP and HTTP. It is text-based streaming XML (Extensible Markup Language) data in close to real-time. XMPP itself provides a signalling channel, typically via the TCP. Jingle [Ludwig et al., 2009] is a signalling protocol that extends XMPP with audio and video calls and is designed to interwork with SIP through gateways. Massively used services provides XMPP-based solutions such as WhatsApp Messenger with 800 million monthly active users[13]. The XMPP clients usually are implemented as a software device.

Because of firewalls and proxies, clients usually have connectivity problems connecting to SIP (port 5060) and XMPP services (port 5222). XMPP started using BOSH (Bidirectional-streams Over Synchronous HTTP, [Paterson et al., 2014]), this had the problems of using HTTP for bidirectional communication. Now, with the standardization of Websocket protocol, there are SIP over websocket [Castillo et al., 2014] and XMPP over websocket [Stout et al., 2014].

3.3 WebRTC

WebRTC is the technology able to perform calls through the browser and real-time communications in general. Here is presented WebRTC through the works of the two organizations working on its standardization: IETF and W3C. The presentation of the different concepts it is inspired by the IETF Internet-Draft *Overview: Real Time Protocols for Browser-based Applications*[14].

WebRTC is a browser-embedded media engine released by Google in 2011. A media engine packages multimedia processing components into an optimized software solution for smoother integration and better performance. It is also an effort from IETF and W3C to add standardized RTC capabilities into browsers. IETF RTCWEB[15] working group is producing architecture and requirements for selection and profiling of the on the wire protocols. W3C WEBRTC[16] working

[11] Formerly the protocol was originally named Jabber.

[12] Formerly the Jabber Software Foundation.

[13] `https://www.facebook.com/jan.koum/posts/10153230480220011`.

[14] `https://tools.ietf.org/html/draft-ietf-rtcweb-overview-13`

[15] `https://tools.ietf.org/wg/rtcweb/`

[16] `http://www.w3.org/2011/04/webrtc/`

group is defining browser APIs to enable Real-Time Communications in Web browsers. The WebRTC effort is still a work in progress. Figure 3.1 links the browser model with the work in progress of IETF and W3C, the source[17] is from a Victor Pascual's presentation of WebRTC.

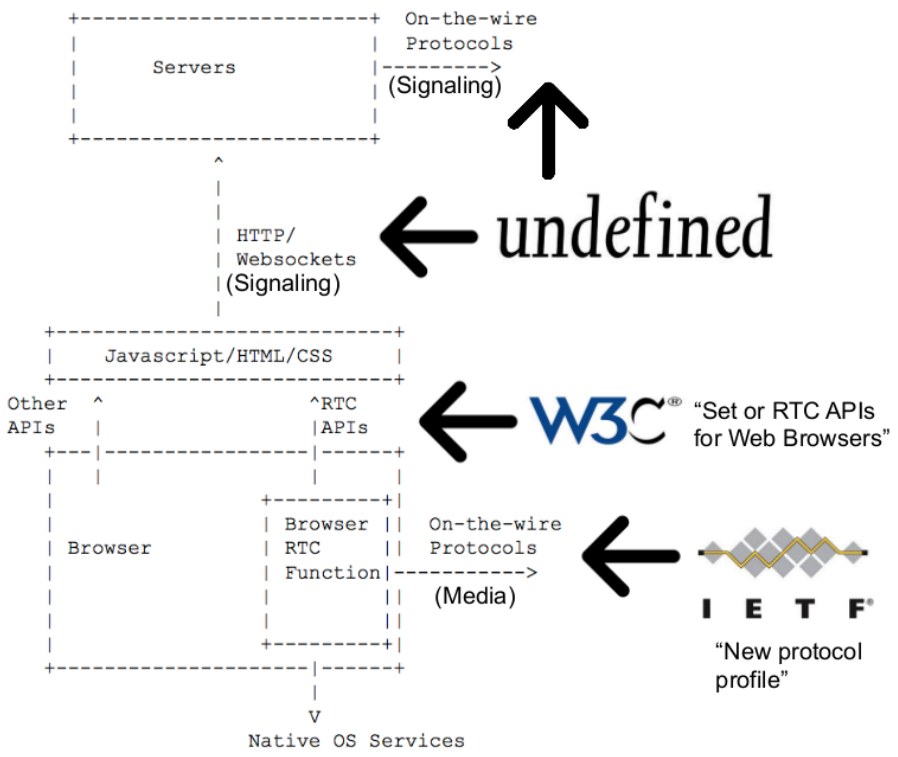


Figure 3.1: WebRTC browser model

The WebRTC trapezoid architecture model (see figure 3.2) shows the elements that make possible the communication between browsers. JS/HTML/CSS are the components to do a web application, it includes the WebRTC W3C APIs. It is required bidirectional communication channel between the web browser and the web server, it can be used an application-defined over Websocket Protocol such as: XMPP or SIP. The signalling path is composed by the communication channels used between entities participating in signalling. Once signalling is transferred, it starts a media path, a communication channel where multimedia content follows from one WebRTC device to another.

[17] http://www.slideshare.net/victorpascual/webrtc-standards-update-october-2014/4

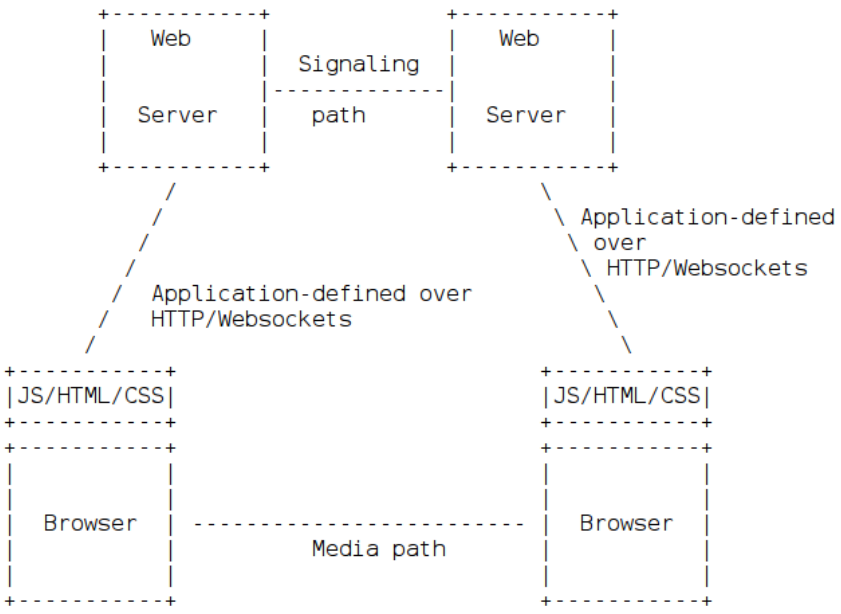


Figure 3.2: WebRTC trapezoid architecture model. Source: IETF WebRTC Overview

Data transport. IETF Internet-Draft *Transports for WebRTC*[18]

Data framing and securing (RTP, SRTP). It MUST IMPLEMENT IETF Internet-Draft *Web Real-Time Communication (WebRTC): Media Transport and Use of RTP*[19] IETF Internet-Draft *Security Considerations for WebRTC*[20]. IETF Internet-Draft *WebRTC Security Architecture*[21]. For non RTP data, datachannels (SCPT over DTLS over UDP) IETF Internet-Draft *WebRTC Data Channels*[22], IETF Internet-Draft *WebRTC Data Channel Establishment Protocol*[23].

Data formats. IETF Internet-Draft *WebRTC Audio Codec and Processing Requirements*[24]. IETF Internet-Draft *WebRTC Video Processing and Codec Requirements*[25]. Codecs, Audio: G.711, Opus; Video: H264, VP8.

Connection management. No signalling is specified. Use of SDP (JSEP is

[18] https://tools.ietf.org/html/draft-ietf-rtcweb-transports-06
[19] https://tools.ietf.org/html/draft-ietf-rtcweb-rtp-usage-16
[20] https://tools.ietf.org/html/draft-ietf-rtcweb-security-07
[21] https://tools.ietf.org/html/draft-ietf-rtcweb-security-arch-10
[22] https://tools.ietf.org/html/draft-ietf-rtcweb-data-channel-11
[23] https://tools.ietf.org/html/draft-ietf-rtcweb-data-protocol-07
[24] https://tools.ietf.org/html/draft-ietf-rtcweb-audio-05
[25] https://tools.ietf.org/html/draft-ietf-rtcweb-video-00

the evolution). IETF Internet-Draft *Javascript Session Establishment Protocol*[26]

Presentation and control. Peer Connection API. W3C Working Draft *WebRTC 1.0: Real-time Communication Between Browsers*[27]. Media Capture API. W3C Working Draft *Media Capture and Streams*[28]

Local system support functions

- Echo cancellation should be good enough to achieve the suppression of acoustical feedback loops below a perceptually noticeable level.
- Privacy concerns MUST be satisfied; for instance, if remote control of camera is offered, the APIs should be available to let the local participant figure out who's controlling the camera, and possibly decide to revoke the permission for camera usage.
- Automatic gain control, if present, should normalize a speaking voice into a reasonable dB range.

The requirements on WebRTC systems with regard to audio processing are found in `https://tools.ietf.org/html/draft-ietf-rtcweb-audio-05` (previous); the proposed API for control of local devices are found in Media Capture API.

- Media Management
  - MCU
  - SFU
- **Security** problems hiding public IP
- Blocks
  - Application
  - Signalling
  - Gateway
    - Media Gateway
    - Signalling Gateway
  - Transport

Some parts of [**?**] and [**?**] were included in this section.

[26] `https://tools.ietf.org/html/draft-ietf-rtcweb-jsep-07`
[27] `http://www.w3.org/TR/2015/WD-webrtc-20150210/`
[28] `http://www.w3.org/TR/2015/WD-mediacapture-streams-20150414/`

Chapter 4

METHODOLOGY

4.1 SWOT analysis of WebRTC

An analysis of Strengths, Weaknesses, Opportunities and Threats (SWOT) will help the decision-making and tasks for the project.

- Strengths
  - Ease of use: real-time communication is supported without the need for additional applications or plug-ins.
  - It helps to solve connectivity problems caused by NAT, Firewall, etc.
  - Solved the problem of selection of video and audio codecs.
  - It is based on open standards and open source implementations.
  - It has well general acceptance in both worlds: enterprise and community.
  - The communication between peers is bidirectional and can be P2P
  - WebRTC standard does not specify signalling: it can be used in very different scenarios.
  - The communication channel between peers is encrypted
- Weaknesses
  - It is not implemented in all browsers.
  - The different browsers that implement WebRTC could have incompatibilities.
  - WebRTC has incompatibility at transport level with SIP, a gateway is needed.

  - Security compromised when using split VPN-tunnels, the IP address before the VPN-tunnel is exposed.

- Opportunities
  - WebRTC can be used in web based softphone for VoIP. Easy to install, easy to update.
  - A WebRTC audio call could be routed to traditional telephony.
  - It uses javascript as programming language, this language has the widest developer's community.
  - It encourages a new generation of web applications using its strenghts.

- Threats
  - WebRTC standard do not specify signalling: this can produce a positive or negative fragmentation of projects. Positive fragmentation: different projects for different applications. Negative fragmentation: divided effort.
  - Is a work in progress technology, it is being changed.

4.2 Scope

There are lots of RTC systems for different purposes. This project focuses in the work of IETF organization and Internet. Guifi.net is part of Internet, and has additional constraints to take in account.

The following topics are worked only in its fundamentals: RTC standard systems from ITU[3], referred to in the 2, and XMPP, referred to in the 3.

In general it has chosen technologies based on open standards, open source implementations and royalty free patent.

4.3 Resources

There are costs related to the activity of this project in terms of equipment and human effort.

Table 4.1 shows the equipment resources and its economic estimation. Observations:

- Guifi.net connectivity to Barcelona, a reachable IPv4 10.0.0.0/8[1] has not direct cost.

[1]Range of IP's used by Guifi.net and private networks.

- Nearly all software involved is open source and has no direct cost.
- Usually the cost of installation it's greater or equal than the cost of equipment.

Table 4.1: Equipment resources

Material	Estimated cost (euro)
Guifi.net equipments in my home	200
PC with virtualization capabilities [home]	1000
Guifi.net equipments in university	1000
PC with Internet public IPv4 [university]	300
Laptop	400
ATA x 2	60
Old phone x 2	2
Total	2962

The human effort part was financed by the university in the form of a grant to the author, representing a cost of 2800 euro. A bachelor's thesis corresponds in Europe to 500 hours of work.

This implies a total cost of approximately 6000 euro

4.4 Planning

The project can be separated in two phases. The first phase is a long preamble of studying VoIP and WebRTC. The second phase is an agile plan. Figure 4.1 shows the two phases in a gantt chart.

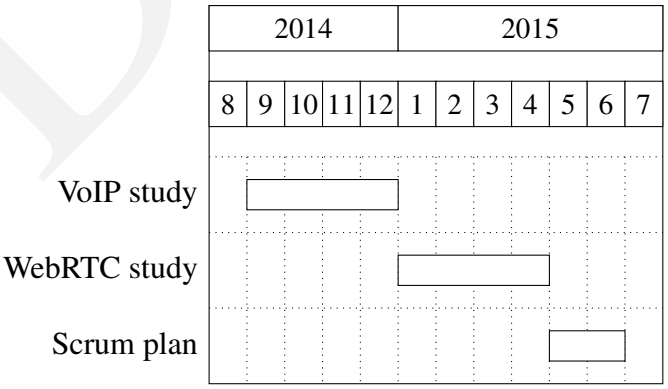


Figure 4.1: General gantt chart

In the first phase, while studying VoIP the intention was to work about VoIP and Guifi.net. But Miquel Oliver encouraged me to do it about WebRTC. He presented me Victor Pascual, a SIP and WebRTC expert. It was hard to realise a convenient project, because this technology involves lots of protocols, other technologies, and it's being modified now. In this phase It were settled the necessary concepts to start the project.

The second phase is an agile plan, inspired by the Scrum methodology. Scrum is one of the Agile methods[2] used for software development. The important fact is that promotes adaptive planning and flexible response to change. Scrum, particularly, is a general method that should be adapted to a concrete scenario.

The Scrum Team consists of a Product Owner, the Development Team, and a Scrum Master. The work of the scrum team according to the Scrum Guide[3] is *deliver products iteratively and incrementally, maximizing opportunities for feedback. Incremental deliveries of "Done" product ensure a potentially useful version of working product is always available*. The roles are

- Product Owner: *is responsible for maximizing the value of the product and the work of the Development Team*
- Development Team: *consists of professionals who do the work of delivering a potentially releasable Increment of "Done" product at the end of each Sprint*
- Scrum Master: *is responsible for ensuring Scrum is understood and enacted*

*The heart of Scrum is a **Sprint**, a time-box of one month or less during which a "Done", usable, and potentially releasable product Increment is created*

4.4.1 Scrum plan

It is necessary to adapt the different concepts that comprise the scrum methodology for this particular project.

Roles:

- Product Owner (in some way, stakeholders): Mentors, University, people interested in the project. The author is interested in the output of the project because is volunteer in Guifi.net.

[2]There are different metodologies grouped into agile. The process started with the write of the Agile Manifesto (12 principles) `http://agilemanifesto.org/iso/en/principles.html`. Since February 2001, this manifesto remains unchanged.

[3]`http://www.scrumguides.org/docs/scrumguide/v1/scrum-guide-us.pdf`.

- Development Team: assumed by the author
- Scrum Master: assumed by the author, optionally could be assumed by mentors.

This means that the author has to see the project with different points of view.

The Sprint time is approximately one week, because it is assumed that the minimum time-box possible to do a release of the product is one week. The product comprise two major tasks: the theory (documentation, memory) and practice (how this theory is fitted to the real world experiments). The tasks are explained with more detail in the next section.

Figure 4.2 shows the Scrum plan with the different sprint phases (s1, s2, s3, s4) and important milestones:

- d1: project charter and tasks, delivery to mentors
- d2: first consistent draft memory, delivery to mentors
- d3: set title and abstract to the thesis, delivery to university
- d4: thesis, delivery to assigned tribunal

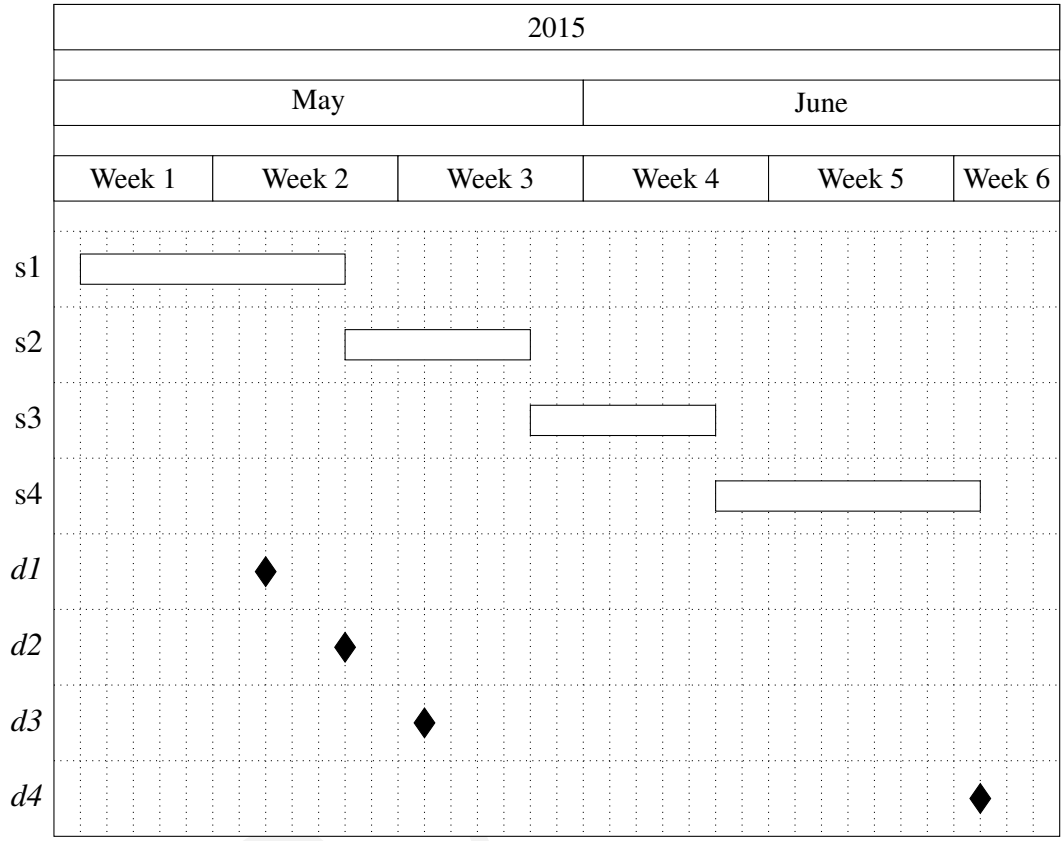


Figure 4.2: Scrum plan gantt chart

4.4.2 Metatools

To ensure the scrum plan and the project, different tools were used:

- Emacs orgmode: is a plain text syntax and software that facilitates different operations
  - nested concepts: It is possible to fold and unfold nested concepts different parts. This brings facilities to take different points of view of the project.
  - write the memory of the project and export to UPF publication constraints.
  - diary: used as autoevaluation tool. Time spent in some operations. Place to record when was discovered something.
  - tasks: to write things to do and mark them as TODO and DONE. To see overall progress of the project.

- Git: is a distributed version control system that helps to ensure the work is not lost. It can has a local and remote copy of all different states (commits) of the project. It is very flexible to do changes and apply.

- Github[4] repository: is a social network that uses git and has the largest community. A place to host and share open source projects. This project is hosted as a repository in `https://github.com/pedro-nonfree/guifi-webrtc`. Featured files:

  - diary.org: record of activity in time
  - tasks.org: parts to do for the project
  - doc directory: independent parts written before starting the memory, or that needs isolation
    - doc/index.org: organise the different files of this directory
  - latexbuild directory: place where emacs orgmode thesis file is exported to latex and compiled to PDF
    - thesis.org: source code of memory
    - thesis.pdf: memory

4.5 Tasks and work style

The tasks for this project are divided in two components: theory and practice.

Inside **theory**, there is:

- Documentation
  - Things to say/explain: what should be said that at the moment is missing (checklist)
  - Parts to fill: developed parts that are missing few details (checklist)
  - Parts to fix: developed parts that are incorrect and should be fixed (checklist)
  - Questions: related to the writing or the theory part, that it is needed an answer (checklist, done when answered)
  - Review: concepts that should be reviewed again, after a scheduled date (checklist)

[4]The web implementation is proprietary software, but it can be easily migrated to other open source tools such as `http://gitlab.com` or `http://gogs.io/`.

  - Memory document has tools to track the state of different sections. For theory, it will be specially important:
    * Fundamentals
    * State of the Art
    * Result and Contributions

- Search of information
  - What things should be read (checklist).

Inside **practice**, there is:

- WebRTC POC: what WebRTC Proof Of Concepts that have been executed, and wishlist (checklist). What signalling was used. The POCs are web applications that have library linking with signalling. Interested in SIP (jssip) and XMPP (strophe) signalling.
- Tested components: what specific components that have been executed, and wishlist (checklist). LDAP Authentication, SSL/TLS certificates, STUN/TURN server, DNS.

Figure 4.3 shows the **work style**, how the objectives will be accomplished and its quality. Stable means that it should be clear and complete its content; best effort that it will work in the best way possible but with less priority:

- Requirements: use cases, constraints needed for the chosen organization. (quality: stable)
- Design: arquitecture design that fits the requirements. (quality: stable)
- Implementation: component selection, protocols (quality: best effort)
- PoC: applications that shows some of the results (quality: best effort)

4.6 Architecture Design

- network architecture
  - structure: nodes, connections, relationships
  - behavior: protocols, functionality
  - production and deployment plans

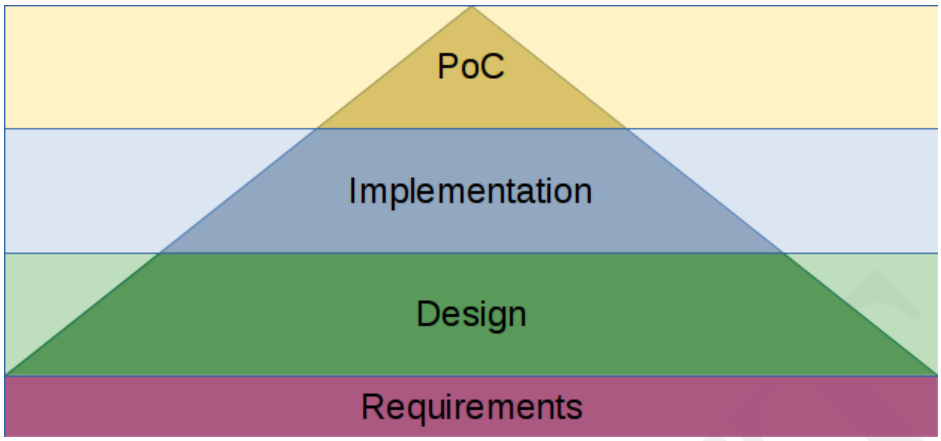


Figure 4.3: Work style diagram

- architecture features
  - centralization vs. distribution
  - open vs closed
  - availability
  - scalability
    - extendability
  - control
    - security
- network architectures (guifi)
- signalling types

Chapter 5

CONTRIBUTIONS AND RESULTS

It is wanted to define the architecture of a general RTC service, as set of different services and applications.

5.1 Architecture of Guifi.net

5.2 Requirements

5.2.1 Network requirements

number and math justification?

- DNS (multiple servers)
- Quality of Service (QoS): put priority to real-time traffic
- Throughput (more used than Bandwidth)
- Delay
- Jitter
- Packet Loss
- Congestion

5.2.2 Generic use cases

The use cases developed are those that are bold.

Definitions:

- Actors or Roles have a **user** and/or **admin** account. The user has minimum permissions in the application, the admin has all the permissions in the application. It can be defined one or more middle actors that have more permissions than user and less permissions than admin.
- A **call** refers to an audio or video call, bidirectional communication with video and/or audio channel.
- A user is **available** if is connected to the service and busy.

A general RTC service could be defined as it follows:

1. **Send calls: a user calls another user with an audio channel. Optional channels of communication if available: video and chat.**
    - Access to the service through an application.
    - Authentication.
    - Decision of which user it is wanted to call.
    - The call is accepted by the other user.
    - Bidirectional communication.
    - One of the two users stop the communication.
2. **Receive calls: a user receives a call only if is connected to the service with at least one device and is available.**
    - Access to the service through an application.
    - Authentication.
    - A lapse of random time until a call is received. If there are more devices of the same user, all of them receive the call, but only one can accept it.
    - The call is accepted.
    - Bidirectional communication.
    - One of the two users stop the communication.
3. A user can subscribe or unsubscribe to the RTC service.
4. Call history: a user can see the calls sent, received, missed. He can delete it.
5. Missed call notification

- When a user calls to another user that is not available. A ”missed call” notification is generated to be delivered to the other user.
- A user has been notified by a missed call if the device is compatible with this service and he is available.

6. Contact list
   - A user can see the status (online, offline, busy, etc.) of another user in its contact list if he is allowed by the other user.
   - A user can add or remove another user from the contact list.
7. Chat rooms
   - A user can be in a public place where there are rooms and people talk openly.
   - A user can speak privately to the users connected to this place.
   - The identity in the chat room is the same as in the contact list.
8. User preferences
   - User can set its own photo, nickname and description.
   - Users can set if a room is able to record a history conversation (and files) such that users that connect and disconnect can follow the conversation.
   - User can change password of its account.
9. Share advanced media
   - User can share its screen with another user.
   - User can share files of limited size in a room or privately to another user. The data is temporarily stored.
   - Share N streams to N users (Multiuser bidirectional videoconference).
   - Share one stream to N users (Streaming).
10. Administration
    - User can only change its settings. Admin can change configuration of all users.
    - Users can report other users because of a social conflict, admin is notified.

11. **Integration: all the services are integrated and is the same account.**

Guifi.net service is defined as it follows:

1. A user can connect to a server if he could reach it with good quality, if not, he can easily install it in its zone.
2. If a server reach another, the users of a server can communicate to the users of another server.
3. The service is compatible with VoIP Guifi.net project.

5.2.3 Required components

This section presents the components required for the architecture design. Concretely the requirements to send, receive calls, subscribe, unsubscribe and have integration are described now. For a WebRTC scenario, the components are distributed.

- Application interface: gives appropriate interaction to the actors in order to perform the different operations. Is distributed, a web server (WSRV) offering a page, and a client executing the web application (WAPP) through the web browser.
- Authentication service (AUTH): restricts access of service only to permitted users. Differentiate available operations depending on if is user or admin. Single sign on, after the web page is accessed, the user can operate.
- Signalling protocol (SIG): manages the side to side connections and logic to establish the call. The two peers must use a compatible signalling protocol.
- Transport protocol (TP): used between users and between user and server. Compatible, secure if possible.
- Database (DB): stores and encrypts personal information or preferences for a particular user. Accessible through web application if succeed in authentication.
- Connectivity solver (CS): a set of tools to avoid common communication problems that appear in networking scenarios. The most common problems are NAT and firewall.
- Gateway (GW): adapts or converts the communication to work with different communication systems. The most important difference between communication systems is the signalling protocol.

5.3 Architecture design

general diagrams

Each link has associated transport protocol

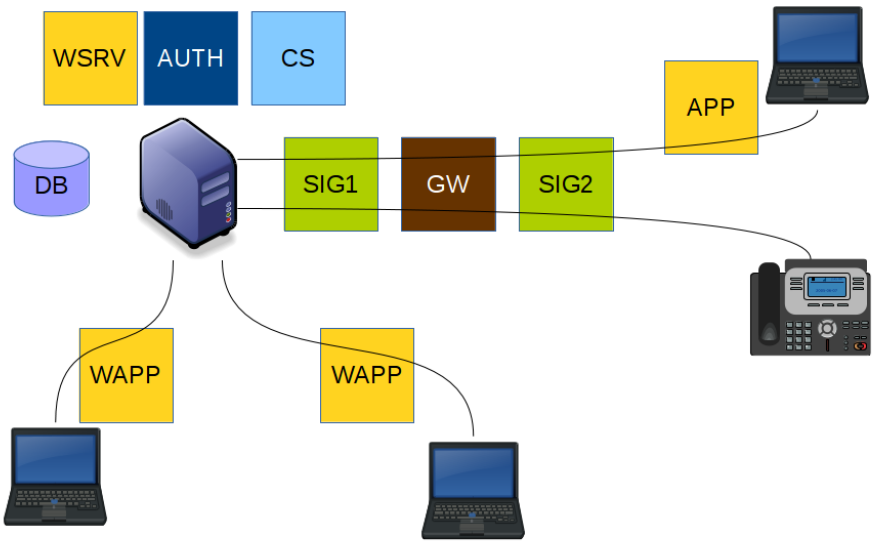


Figure 5.1: Component Diagram

flow chart communication

5.3.1 WebRTC to SIP case

Authentication *authentication diagrams*

Gateway *gateway diagrams*

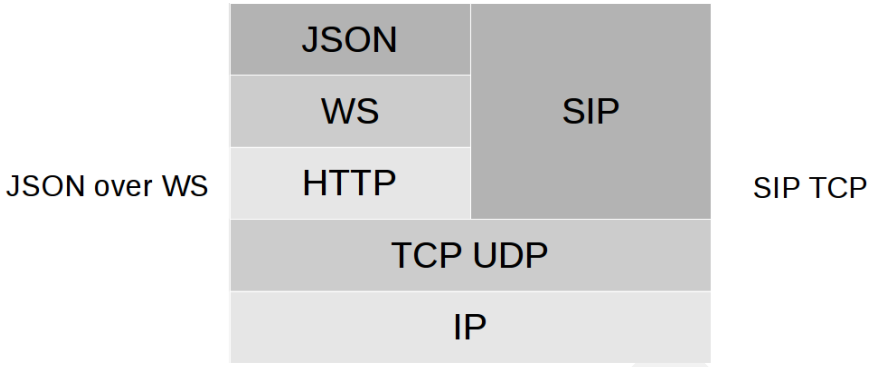


Figure 5.2: SDP exchange during SIP signalling

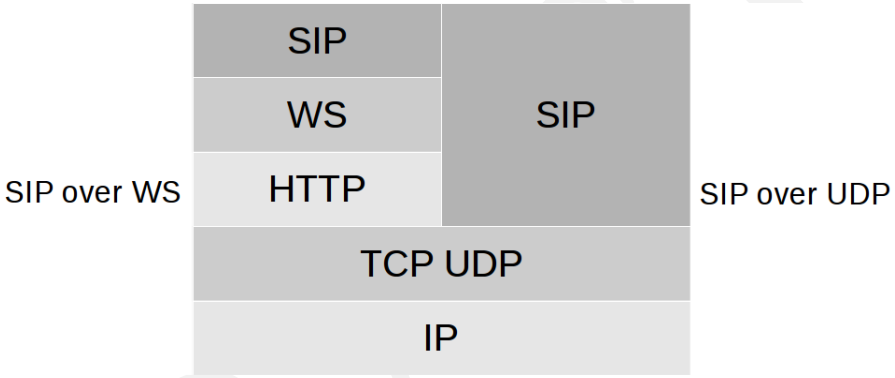


Figure 5.3: SIP transport

5.4 Component selection

5.5 Applications available

5.5.1 POCs

5.6 Implementation

5.7 *Demo*

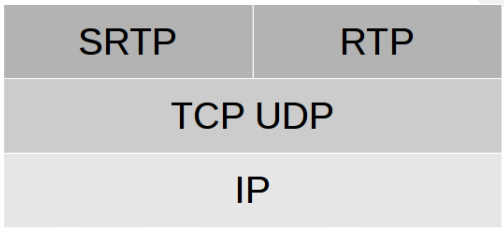


Figure 5.4: Transport

Chapter 6

CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

6.2 Future Work

CUSAX [Ivov et al., 2013]

Chapter 7

FAKE

[To be removed]
Fake section to put the last latex code of the document

Bibliography

[Baugher et al., 2004] Baugher, M., McGrew, D., Naslund, M., Carrara, E., and Norrman, K. (2004). The Secure Real-time Transport Protocol (SRTP). RFC 3711 (Proposed Standard). Updated by RFCs 5506, 6904.

[Castillo et al., 2014] Castillo, I. B., Villegas, J. M., and Pascual, V. (2014). The WebSocket Protocol as a Transport for the Session Initiation Protocol (SIP). RFC 7118 (Proposed Standard).

[Deering and Hinden, 1998] Deering, S. and Hinden, R. (1998). Internet Protocol, Version 6 (IPv6) Specification. RFC 2460 (Draft Standard). Updated by RFCs 5095, 5722, 5871, 6437, 6564, 6935, 6946, 7045, 7112.

[Dodd, 2012] Dodd, A. Z. (2012). *The essential guide to telecommunications*. Prentice Hall, fifth edition.

[Fette and Melnikov, 2011] Fette, I. and Melnikov, A. (2011). The WebSocket Protocol. RFC 6455 (Proposed Standard).

[Fielding et al., 1999] Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., and Berners-Lee, T. (1999). Hypertext Transfer Protocol – HTTP/1.1. RFC 2616 (Draft Standard). Obsoleted by RFCs 7230, 7231, 7232, 7233, 7234, 7235, updated by RFCs 2817, 5785, 6266, 6585.

[Frank, 2004] Frank, O. (2004). *Voice over 802.11*. Artech House, first edition.

[Gulbrandsen et al., 2000] Gulbrandsen, A., Vixie, P., and Esibov, L. (2000). A DNS RR for specifying the location of services (DNS SRV). RFC 2782 (Proposed Standard). Updated by RFC 6335.

[Handley et al., 2006] Handley, M., Jacobson, V., and Perkins, C. (2006). SDP: Session Description Protocol. RFC 4566 (Proposed Standard).

[Ivov et al., 2013] Ivov, E., Saint-Andre, P., and Marocco, E. (2013). CUSAX: Combined Use of the Session Initiation Protocol (SIP) and the Extensible Messaging and Presence Protocol (XMPP). RFC 7081 (Informational).

[Klensin, 2008] Klensin, J. (2008). Simple Mail Transfer Protocol. RFC 5321 (Draft Standard).

[Kurose and Ross, 2013] Kurose, J. F. and Ross, K. W. (2013). *Computer Networking. A Top-Down Approach*. Pearson, sixth edition.

[Loreto et al., 2011] Loreto, S., Saint-Andre, P., Salsano, S., and Wilkins, G. (2011). Known Issues and Best Practices for the Use of Long Polling and Streaming in Bidirectional HTTP. RFC 6202 (Informational).

[Ludwig et al., 2009] Ludwig, S., Beda, J., Saint-Andre, P., McQueen, R., Egan, S., and Hildebrand, J. (2009). Jingle. XEP-0166 (Standards Track).

[Mahy et al., 2010] Mahy, R., Matthews, P., and Rosenberg, J. (2010). Traversal Using Relays around NAT (TURN): Relay Extensions to Session Traversal Utilities for NAT (STUN). RFC 5766 (Proposed Standard).

[Mealling, 2002] Mealling, M. (2002). Dynamic Delegation Discovery System (DDDS) Part Three: The Domain Name System (DNS) Database. RFC 3403 (Proposed Standard).

[Mockapetris, 1987] Mockapetris, P. (1987). Domain names - concepts and facilities. RFC 1034 (INTERNET STANDARD). Updated by RFCs 1101, 1183, 1348, 1876, 1982, 2065, 2181, 2308, 2535, 4033, 4034, 4035, 4343, 4035, 4592, 5936.

[Paterson et al., 2014] Paterson, I., Smith, D., Saint-Andre, P., Moffitt, J., Stout, L., and Tilanus, W. (2014). Bidirectional-streams Over Synchronous HTTP (BOSH). XEP-0124 (Standards Track).

[Postel, 1980] Postel, J. (1980). User Datagram Protocol. RFC 768 (INTERNET STANDARD).

[Postel, 1981a] Postel, J. (1981a). Internet Protocol. RFC 791 (INTERNET STANDARD). Updated by RFCs 1349, 2474, 6864.

[Postel, 1981b] Postel, J. (1981b). Transmission Control Protocol. RFC 793 (INTERNET STANDARD). Updated by RFCs 1122, 3168, 6093, 6528.

[Rekhter et al., 1996] Rekhter, Y., Moskowitz, B., Karrenberg, D., de Groot, G. J., and Lear, E. (1996). Address Allocation for Private Internets. RFC 1918 (Best Current Practice). Updated by RFC 6761.

[Rescorla, 2000] Rescorla, E. (2000). HTTP Over TLS. RFC 2818 (Informational). Updated by RFCs 5785, 7230.

[Rosenberg, 2010] Rosenberg, J. (2010). Interactive Connectivity Establishment (ICE): A Protocol for Network Address Translator (NAT) Traversal for Offer/Answer Protocols. RFC 5245 (Proposed Standard). Updated by RFC 6336.

[Rosenberg, 2013] Rosenberg, J. (2013). SIMPLE Made Simple: An Overview of the IETF Specifications for Instant Messaging and Presence Using the Session Initiation Protocol (SIP). RFC 6914 (Informational).

[Rosenberg et al., 2008] Rosenberg, J., Mahy, R., Matthews, P., and Wing, D. (2008). Session Traversal Utilities for NAT (STUN). RFC 5389 (Proposed Standard). Updated by RFC 7350.

[Rosenberg et al., 2002] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., and Schooler, E. (2002). SIP: Session Initiation Protocol. RFC 3261 (Proposed Standard). Updated by RFCs 3265, 3853, 4320, 4916, 5393, 5621, 5626, 5630, 5922, 5954, 6026, 6141, 6665, 6878, 7462, 7463.

[Saint-Andre, 2011] Saint-Andre, P. (2011). Extensible Messaging and Presence Protocol (XMPP): Core. RFC 6120 (Proposed Standard).

[Schulzrinne et al., 2003] Schulzrinne, H., Casner, S., Frederick, R., and Jacobson, V. (2003). RTP: A Transport Protocol for Real-Time Applications. RFC 3550 (INTERNET STANDARD). Updated by RFCs 5506, 5761, 6051, 6222, 7022, 7160, 7164.

[Senie, 2002] Senie, D. (2002). Network Address Translator (NAT)-Friendly Application Design Guidelines. RFC 3235 (Informational).

[Stout et al., 2014] Stout, L., Moffitt, J., and Cestari, E. (2014). An Extensible Messaging and Presence Protocol (XMPP) Subprotocol for WebSocket. RFC 7395 (Proposed Standard).

[Velázquez, 2010] Velázquez, T. (2010). Sistema de servidores voip del proyecto guifi.net. Bachelor's thesis. Universitat Autònoma de Barcelona (UAB).

[Vílchez, 2014] Vílchez, P. (2014). Starting, contributing and empowering community networks in cities. Bachelor's thesis. Universitat Pompeu Fabra (UPF).