Journal of Computational Information Systems 10: 17 (2014) 7507–7516
Available at http://www.Jofcis.com

Research on Media Intercommunications Between WebRTC and IMS Network

Kai SHUANG*, Guoshuai ZHAO

State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract

WebRTC is a real-time audio and video communication technology, which is natively supported by the browser and is not dependent on third-party plug-ins. And developers only need to call the API provided by browsers to achieve the media communications between browsers. The intercommunication of the WebRTC system and the IMS network has several problems in signaling and media plane. Combining the characteristics of heterogeneous networks, this paper puts forward a complete solution for intercommunication between WebRTC and IMS network, designs the system architecture and processes of intercommunication, and carries out a detailed analysis about the key module named as the media gateway. In addition, this paper presents a distributed idea to solve bandwidth and CPU bottlenecks of the media gateway and ensures the scalability of the system performance and parallelism of different sessions.

Keywords: WebRTC; IMS; SIP; Intercommunication; Media Gateway

1 Introduction

Generally, the traditional real-time communication services always use private technologies, and different audio and video communication systems use private communication protocols so that intercommunications between different systems are almost impossible. In the implementation level, a variety of audio and video communication methods need to install clients or plug-ins to achieve communication, and there does not exist a universal, high-quality, integral audio and video communication solution. These defects in audio and video communication systems result in high development costs, and the poor versatility.

WebRTC is a browser-based real-time audio and video communication technology [1]. Compared to the traditional means of real-time communication which are based on clients or plug-ins

*Project supported by the Fundamental Research Funds for the Central Universities (No. 2013RC1102), National Key Basic Research Program of China (973 Program) (No. 2011CB302506) and Important national science & technology specific projects:Next-generation broadband wireless mobile communications network (2012ZX03002008-002-03).

*Corresponding author.
Email address: shuangk@bupt.edu.cn (Kai SHUANG).

DOI: 10.12733/jcis11499
September 1, 2014

in browsers, WebRTC integrates modules of audio and video processing (capturing, encoding, decoding), network transmission, session control into browsers and enables developers to get real-time audio and video communication capabilities through simple JavaScript API calls [2]. Moreover, WebRTC integrates with best-of-breed voice and video engines, contains NAT and firewall traversal technology using STUN, ICE, TURN, RTP-over-TCP, and is a universal, high-quality, full audio and video communication solution.

For traditional telecom operators, WebRTC not only brings challenges, but also brings opportunities [3, 4]. On the one hand, WebRTC technology has an impact on real-time multimedia communication services of the IMS network. On the other hand, achieving the combination of WebRTC technology and IMS, telecom operators can employ the advantages of WebRTC which include the huge potential user base, lower development and maintenance costs and rapid deployment in businesses of IMS. And the combination can provide users with richer user experience, promote traditional users to migrate to the IMS network, and create tremendous business value. Therefore, in addition to internet companies, telecom operators and telecom equipment manufacturers are concerned about the progress of standardization of WebRTC technology [5].

There are many problems in the intercommunication process of WebRTC clients and SIP terminals in the IMS network [6]. In the signaling plane, the conversion of signaling protocols in different networks needs to be achieved. In the media plane, the conversion of real-time transport protocols and media codec types in different networks needs to be achieved. Acquiring media streams from terminals of different networks is the prerequisite for conversion, so we have to establish communications in P2P mode with terminals of different networks. In addition, since the processing of media session requires a lot of bandwidth and CPU resources, the architecture of the intercommunication system needs to ensure the scalability of the system performance.

2 Issues in Intercommunication

In order to achieve the media intercommunication between WebRTC clients and SIP terminals, the problem of mapping the signaling protocols needs to be solved. In the process of signaling interaction between WebRTC clients and the WebRTC server, RTCWeb (Real-Time Communication in WEB-browsers) work group formulates JSEP (JavaScript Session Establishment Protocol) [7] as an offer/answer mechanism. JSEP is not a signaling protocol, and it provides a way to maintain the signaling state machine. So the actual signaling protocols can be used on JSEP, such as ROAP [8], SIP and XMPP. SIP is a signaling control protocol developed by the IETF for creating, modifying and releasing one or more participants in the session. In the process of intercommunication, a signaling gateway needs to be designed to achieve the media negotiation and specify the mapping between the signaling protocols used by WebRTC clients and SIP terminals.

In order to ensure the safety of media sessions, WebRTC clients use SRTP [9] as media data transport protocol. However, SIP terminals use RTP [10] as media data transport protocol. The real-time transport protocols are incompatible, so a media gateway needs to be designed to achieve the conversion of SRTP and RTP, which works as a media relay.

The media gateway needs to acquire media stream from WebRTC clients and SIP terminals before it makes the conversion between SRTP and RTP. To achieve this functionality, the media gateway needs to perform an ICE [11] client function. ICE is a private network traversal solution widely used in VOIP systems. Through ICE, both sides of communications can collect their own candidate addresses, then get the other peer's candidate addresses, finally do connectivity checking. The P2P communications will be established if the connectivity checking is successful. In the intercommunication process, the media gateway will establish communications in P2P mode with WebRTC clients and SIP terminals separately to acquire media streams from both sides.

In order to achieve media intercommunications between WebRTC clients and SIP terminals, they must support at least one common audio codec and video codec. If both sides do not have a common codec, the media gateway needs to perform an audio and video codec conversion function. It should decode the media stream from one side, encode it according to the codec of the other side, and then send encoded media stream to the other side. In the choice of audio codec, g.711 can be used as the preferred codec, which is widely used in VoIP systems and is also supported by WebRTC clients. Using g.711 as the audio codec in intercommunications can reduce the pressure on the media gateway, which doesn't need to decode and encode audio stream. In the choice of video codec, WebRTC clients are internet applications, while SIP terminals are telecommunication network terminals. They support different video codecs due to different network environments and patent issues, so the media gateway needs to make a conversion.

As a bridge between two heterogeneous networks, the media gateway needs to handle the problems of transforming between SRTP and RTP, decoding and encoding video media streams and establishing communications in P2P mode with WebRTC clients and SIP terminals. There are high demands on processing power and network bandwidth of the media gateway so that the media gateway supports a limited number of media sessions at the same time. In addition, the intercommunication requires that the different media sessions can be processed in parallel, and the system architecture can be extended easily by deploying more media gateways to deal with a large number of media session requests and achieve load balancing [12].

3 Architecture and Processes of the Intercommunication System

3.1 Architecture of the intercommunication system

The WebRTC/SIP intercommunication system consists of browsers, WebRTC clients, the WebRTC server, the WebRTC gateway, the IMS network environment and SIP terminals.

WebRTC clients are running in the web browser environment, which consist of the business logic module and the interface UI module. The business logic module is responsible for management and control of media sessions, which consists of JSEP protocol implemented in JavaScript and the signaling protocol stack (ROAP, SIP, XMPP). The interface UI module is responsible for interacting with users. WebRTC clients use universal web technologies such as HTML, CSS and JavaScript, and have good compatibility and portability.

The WebRTC server is responsible for management of WebRTC sessions in the signaling plane, which has the functions of routing signaling messages, maintaining session state machine, and handling session exceptions. When a message is received, the WebRTC server will route the message to a local WebRTC client connecting with the WebRTC server or a SIP terminal in IMS network according to the identifier of the callee.

The WebRTC gateway includes the signaling module and the media module. The signaling module is responsible for the conversion of signaling protocol between WebRTC clients and SIP terminals, and the routing of signaling message. The media module is responsible for establishing P2P communications with WebRTC clients and SIP terminals, transforming between SRTP and RTP, decoding and encoding media stream according to the codecs of both sides, and sending the processed media stream from one side to the other side.

IMS (IP Multimedia Subsystem) is a new multimedia service network based on IP technology, and is able to meet the needs of more innovative, more diversified multimedia businesses. In order to conform to the trend of IP network, IMS uses SIP as call control protocol. Accessing SIP terminals to the IMS network can realize the management and control of sessions. And SIP terminals can be any devices or software terminals complying with the standard protocol of SIP.

As shown in Fig. 1, the signaling module and the media module work as a whole, and they can interact with each other easily. The signaling module achieves the intercommunication in signaling level, and the media module achieves the intercommunication in media level. Inadequacies of this architecture are that the processing capability for the multimedia session can't be extended, and the media module takes up a lot of CPU and bandwidth resources and become the bottleneck in the process of intercommunications. To solve this problem, the gateway has been separated into the two parts. One is named as the signaling gateway, and the other is named as the media gateway. As shown in Fig. 2, the new architecture leads to better scalability of system performance, and has more clear logical framework.

The media gateways establish socket connections with the signaling gateway, and the signaling gateway maintains sockets of all media gateways. When a new session request arrives, the signaling gateway evenly distributes sessions to different media gateway, so that load balancing of the media gateway for media session can be achieved. In addition, the deployment of media gateways has high scalability. When the load of current media gateways is too large, we can dynamically increase the number of media gateways by the way of deployment to improve the processing capability of system for media sessions. When a media gateway is down and disconnects with the signaling gateway, the signaling gateway will no longer distribute a new session to the media gateway.

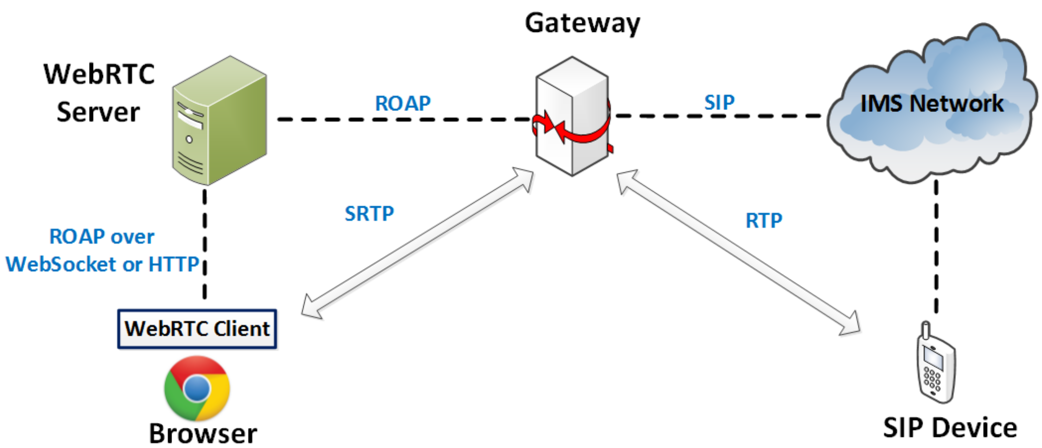

Fig. 1: Centralized architecture of the signaling and media gateways

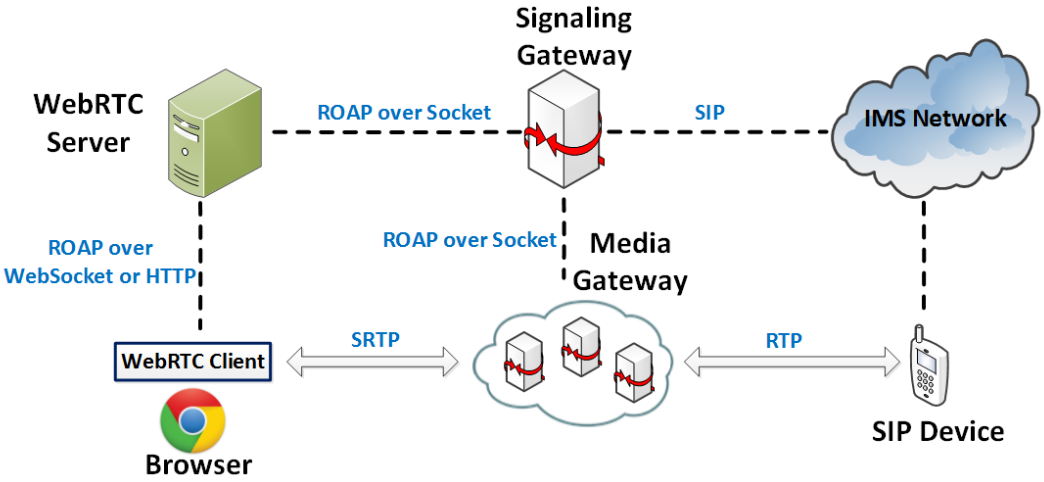

Fig. 2: Separated architecture of the signaling gateway and media gateways

In terms of signaling protocols, ROAP (RTCWeb Offer/Answer Protocol) is proposed by RTCWeb working group and is firstly used by WebRTC as the signaling protocol. ROAP defines the minimum set of message semantics and message formats for the media negotiation between WebRTC clients on the basis of offer/answer model. Subsequently RTCWeb Working Group only focuses on standardization of JSEP rather than on standardization of ROAP, because JSEP develops a general mechanism of signaling state management and can support multiple signaling protocols. Since ROAP is a lightweight signaling protocol, using ROAP in the WebRTC system is simpler and more efficient. Both JSEP and SIP use the offer/answer model to achieve media negotiation, and use SDP (Session Description Protocol) [13] to provide multimedia session description for session notification, session invitation, and other forms of multimedia session initiation. Another advantage of using the ROAP protocol is that the message format of ROAP is designed to be a subset of the SIP. Since there is a clear mapping between two kinds of protocol messages, the conversion between ROAP and SIP messages is convenient. For example, the offer message of ROAP needs to be mapped into the invite message of SIP, and the answer message of ROAP needs to be mapped into the 200 message of SIP.

3.2 Analysis of intercommunication processes

Taking the session request from WebRTC clients to SIP terminals for example, the following describes the process of media negotiation and media intercommunication.

The process of initiating a session request from WebRTC clients to SIP terminals is as follows.

(1) The offer message from WebRTC clients to SIP terminals is firstly sent to the WebRTC server. The request carries the media negotiation information which describes the media codecs supported by WebRTC clients, candidates information for the private network traversal, and key information used by SRTP.

(2) The WebRTC server receives the offer message, and sends it to the signaling gateway according to the identifier of the callee.

(3) The signaling gateway forwards the request to the media gateway, which will replace the ICE candidates in the request with those collected by itself for communicating with SIP terminals, replace the media description information with that supported by SIP terminals and the media gateway, and save the candidates of WebRTC clients and key information of SRTP. Then the media gateway will return the processed offer message to the signaling gateway, and this moment the media gateway will listen on the ports used to communicate with SIP terminals in P2P mode.

(4) The signaling gateway transforms the offer message received from the media gateway into invite message, and send to the IMS network. Then the message is routed to the corresponding SIP terminal.

(5) After receiving the invite message, the target SIP terminal will record media negotiation information, and use the candidates in the message to establish P2P communication with the media gateway according to ICE protocol.

The process of sending a response from SIP terminals to the WebRTC clients is as follows.

(1) SIP terminals send the response of 200 message to the IMS network, which carries the media negotiation information which describes the media codecs supported by SIP terminals, and candidates information for the private network traversal.

(2) The IMS network routes the 200 message to the signaling gateway.

(3) The signaling gateway transforms the 200 message sent by SIP terminals into answer message, and send to the media gateway, which will replace the ICE candidates with those collected by itself for communicating with WebRTC clients, replace the media description information with that supported by WebRTC clients and the media gateway, and save the candidates of SIP terminals. Because SIP terminals use RTP as media data transport protocol and there is no key information of SRTP in the message, the media gateway needs to generate the key information randomly, places the information in the answer message, and return it to the signaling gateway. This moment the media gateway will listen on the ports used to communicate with WebRTC clients in P2P mode.

(4) The signaling gateway routes the answer message to the WebRTC server.

(5) The WebRTC server routes the answer message to WebRTC clients according to the identifier of the caller.

(6) After receiving the response message, the target WebRTC client will use the candidates in the message to establish P2P communication with the media gateway according to ICE protocol.

Through the above processes, the WebRTC client and the SIP terminal complete the process of media negotiation separately with the media gateway. In terms of audio codec, because both WebRTC clients and SIP terminals support g.711, the media gateway can make them use g.711 as audio codec in communication, and does not need to decode or encode media streams. In terms of video codec, WebRTC clients and SIP terminals do not support one same code, so the media gateway should negotiate codec with them respectively to determine two kinds of codecs, then it will do the decoding and encoding work.

The process of handling media streams from WebRTC clients and SIP terminals is as follows.

(1) The media gateway and the WebRTC client have got each other's candidates, so do the media gateway and the SIP terminal. Then they will do connectivity checking according to ICE protocol. After the success, the WebRTC client and the SIP terminal will send their local media stream to the media gateway.

(2) While receiving the media stream from the WebRTC client, the media gateway transforms the media transport protocol from SRTP into RTP according to the key information coming from the WebRTC client, transforms the media codec according to the media negotiation information, and send the processed stream to the SIP terminal.

(3) While receiving the media stream from the SIP terminal, the media gateway transforms the media transport protocol from RTP into SRTP according to the key information generated by itself, transforms the media codec according to the media negotiation information, and send the processed stream to the WebRTC client.

4 Structure of the Media Gatway

In the process of intercommunication, there are many requirements for the media gateway. The media gateway needs to solve the following problems: parallel management of multiple media sessions, realization of private network traversal, conversion of SRTP/RTP and conversion of media codecs.

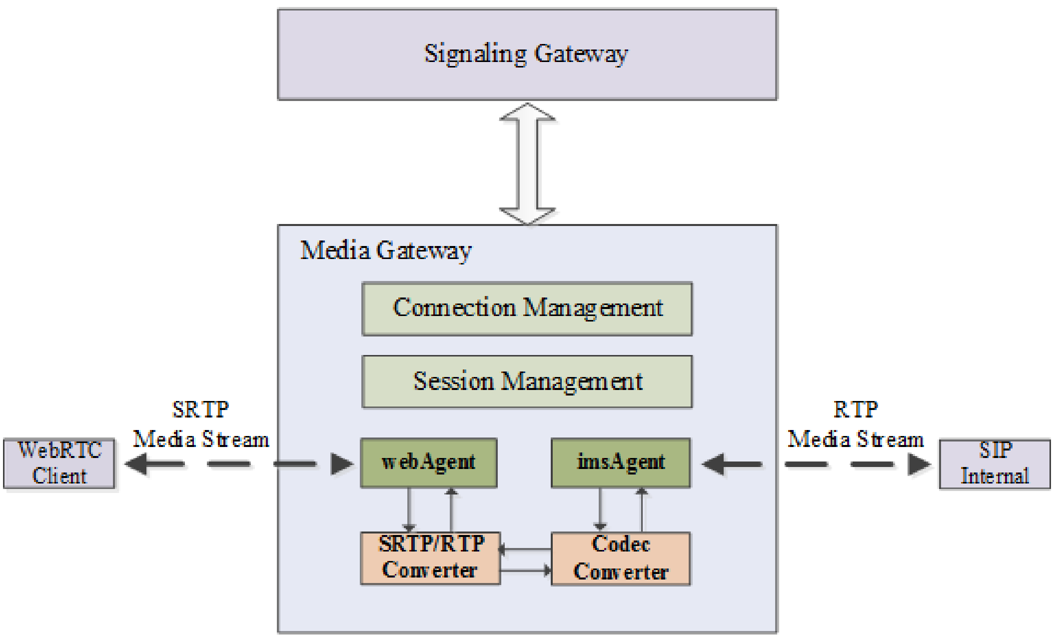

Fig. 3: Structure of the media gateway

As shown in Fig. 3, the media gateway consists of 5 modules. They are the connection management module, the session management module, the ICE agent module used to establish P2P communications, the SRTP/RTP conversion module and the media codec conversion module.

The connection management module is responsible for maintaining the connection established with the signaling gateway, resolving the signaling message, and sending back the processed message to the signaling gateway.

The session management module packages media sessions. Every session management object is an independent instance, which represents a media session and does not disturb others.

As an independent thread, the ICE agent module is responsible for establishing P2P communication with peers. Every session management object contains two ICE agent objects, which will do connectivity checking and establish P2P communication with WebRTC clients and SIP terminals respectively. And every ICE agent object saves the corresponding media stream's SRTP key information.

The SRTP/RTP conversion module is responsible for transforming between SRTP and RTP, which is called by the ICE agent module.

The codec conversion module is responsible for transforming different codecs, which is also called by the ICE agent module. It decodes the media stream, and then encodes the media stream according to the media codec determined by negotiation.

When the connection management module receives a new session request, it instantiates a session management object, resolves the request message, and gives the media description information and key information of SRTP to the session management object. If there is no key information in the message, we can conclude that the request comes from SIP terminals, and the connection management will generate the key information randomly. Then the session management object instantiates two ICE agent objects, and gives the media description information and key information of SRTP to the corresponding ICE agent object according to the message sender. One ICE agent encodes the media stream which is sent to WebRTC clients according to the random key information, and WebRTC clients should get the key information to decode the received media stream. The other ICE agent decodes the media stream which is sent to SIP terminals according to the key information coming from WebRTC clients. Before sending back the message to the signaling gateway, the session management object modifies the session description information in the message. For example, it replaces media codecs information with those supported by the codec conversion module and the message receiver, replaces candidates with those owned by the ICE agent which will establish P2P communications with the message receiver, and add SRTP key information into the message if necessary.

5 Implementation and Analysis

In the realization of the media gateway, the function of SRTP/RTP conversion is achieved based on the open source library named as libsrtp, which uses the SRTP key information to transform media streams transport protocols. The function of codec conversion is achieved base on the open source library named as ffmpeg, which is used to decode and encode the media stream in real time. The function of ICE agent is achieved based on the open source library named as libnice. Using industry-recognized open source libraries ensure the high clarity of the media gateway's architecture and the high efficiency of the system's processing capacity. In terms of the IMS network, we use openIMSCore as experimental IMS environment, and select xlite as the SIP softphone. When a new session is initiated, the signaling gateway will complete the conversion of signaling messages and the media negotiation; The media gateway will realize establishment of P2P communications with the WebRTC client and the SIP terminal, transform between SRTP and RTP, transform media codecs, and route the media stream.

When the running media gateways are overloaded, new media gateways can be deployed into the system conveniently benefited from the scalable system architecture. The new media gateways will establish connections with the signaling gateway, which can distribute session requests to lightly loaded media gateways to improve the carrying capacity of the entire system.

In order to verify the effect of load balancing, we assume that there exist 10 media sessions and calculate the average upload bandwidth of the media gateways when N(from 1 to 6) media gateways are all running in dependent servers with single CPU core. The result of the average upload bandwidth of every media gateway is shown in Fig. 4. With the increase in the number of media gateways, the average upload bandwidth reduces. Fig. 5 shows the result of the average CPU utilization of every media gateway. The average CPU utilization also reduces when more media gateways are deployed.

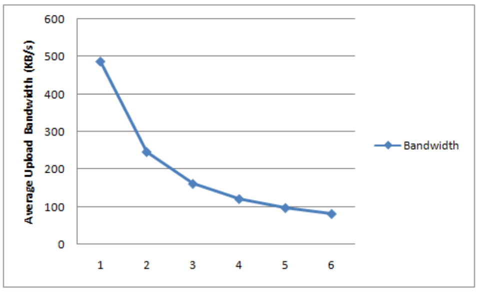

Fig. 4: The average upload bandwidth

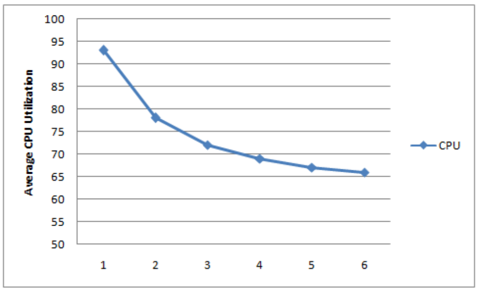

Fig. 5: The average CPU utilization

6 Conclusion

This paper analyzes the features of WebRTC and IMS network, summarizes the key issues in the process of intercommunication, and presents a complete audio and video solution for intercommunication. The solution has the following advantages on the basis of realizing the basic functions. Firstly, the deployment of media gateways has high scalability, which is easy to implement load balancing. Secondly, using multi-threading technology to manage media sessions ensures parallelism of different sessions.

For traditional telecom operators and VoIP manufacturers, the appearance of WebRTC not only brings challenges, but also brings opportunities. As a free technology, WebRTC allows users establish real-time communications via web without installing specific software, and reduces the threshold for development of rich media applications. Moreover, WebRTC clients can intercommunicate with peers in heterogeneous networks through gateways, which brings new applications and scenarios for the audio and video services and open up the market for audio and video communications. Achieving the intercommunications between WebRTC clients and SIP terminals in the IMS network, the advantages of WebRTC which include the huge potential user base, lower development and maintenance costs and rapid deployment can be transformed into powerful driving force of the IMS businesses development. To seize the opportunities, telecom operators need to develop positive attitude to deal with WebRTC technology, pay more attention to the value of WebRTC applications and services, and form a broad ecosystem of rich media services.

References

[1] Alvestrand H. Overview: Real Time Protocols for Brower-based Applications [J]. 2013.

[2] Bergkvist A, Burnett D C, Jennings C, et al. WebRTC 1.0: Real-time communication between browsers [J]. Working draft, W3C, 2012, 91.

[3] Romain C. Some WebRTC opportunities for RCS: And some inner challenges to overcome [C]. Intelligence in Next Generation Networks (ICIN), 2013 17th International Conference on. IEEE, 2013: 31-38.

[4] Bertin E, Cubaud S, Tuffin S, et al. WebRTC, the day after: What's next for conversational services? [C]. Intelligence in Next Generation Networks (ICIN), 2013 17th International Conference on. IEEE, 2013: 46-52.

[5] Loreto S, Romano S P. Real-Time Communications in the Web: Issues, Achievements, and Ongoing Standardization Efforts [J]. IEEE Internet Computing, 2012, 16 (5).

[6] Alessandro Amirante, Tobia Castaldi, Lorenzo Miniero, Simon Pietro Romano. On the Seamless Interaction between WebRTC Browsers and SIP-Based Conferencing Systems [J]. IEEE Communications Magazine, 2013, 51 (4): 42-47.

[7] Uberti J, Jennings C. Javascript Session Establishment Protocol [J]. draft-ietf-rtcweb-jsep-05, 2013.

[8] Jennings C, Rosenberg J, Jesup R. RTCWeb Offer/Answer Protocol (ROAP) [J]. 2011.

[9] Baugher M, McGrew D, Naslund M, et al. The secure real-time transport protocol (SRTP) [J]. 2004.

[10] Frederick R, Jacobson V, Design P. RTP: A transport protocol for real-time applications [J]. IETF RFC3550, 2003.

[11] Rosenberg J. Interactive connectivity establishment (ICE): A protocol for network address translator (NAT) traversal for offer/answer protocols [R]. RFC 5245, April, 2010.

[12] MI W, ZHANG C, QIU X. SLBA: A Security Load-balancing Algorithm for Structured P2P Systems? [J]. Journal of Computational Information Systems, 2012, 8 (7): 2751-2760.

[13] Andreasen F, Baugher M, Wing D. Session description protocol (SDP) security descriptions for media streams [J]. RFC4568 July, 2006.