Specialized network services for WebRTC

TURN-based architecture proposal

Ewa Janczukowicz
Arnaud Braud
Stéphane Tuffin
Gaël Fromentoux

Orange Labs, Lannion, France
{ewa.janczukowicz, arnaud.braud, stephane.tuffin, gael.fromentoux}@orange.com

Ahmed Bouabdallah
Jean-Marie Bonnin

Institut Mines-Telecom/Telecom Bretagne,
Cesson Sévigné, France
{ahmed.bouabdallah, jm.bonnin}@telecom-bretagne.eu

Abstract

Real-time communications have substantially evolved in recent years. Consequently network operators are challenged by web companies providing Internet-wide WebRTC services that are developer and user friendly. So far WebRTC uses best-effort routing but it would benefit from better quality and specialized network services provided by network operators. Therefore this paper presents research approaches: "in-network" and "over-the-network", allowing different level of cooperation between actors. It also highlights two business models: neutral and monetized specialized network services. Later this paper identifies motivations along with technical and business challenges for network operators and WebRTC communication service providers to develop cooperative solutions. It discusses aspects of providing specialized services in access networks and at interconnection. Additionally a collaborative solution, based on TURN servers, is proposed. The advantage of this solution is that it proposes improvements to WebRTC services by using managed VoIP principles, but by assuring compatibility with current web technologies, i.e. it can be introduced incrementally.

Categories and Subject Descriptors C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design

General Terms Design

Keywords WebRTC, QoS, specialized network services, TURN

1. Introduction

Traditional telephony offered by network operators is challenged by new web actors that have entered the market [1].

The web communication ecosystem is very different from the traditional Telco model, i.e. model provided by network service providers. Web companies are not limited to a local territory and have less regulatory constraints. Additionally, they take advantage of flat rate data plans, so users do not pay directly for their service [2]. This movement is reinforced by emerging WebRTC technology that makes implementing communication services even easier [3]. It allows browser to browser communication by using native browser tools and it requires knowledge of only web technologies, so may be easily integrated with the web infrastructure and web services.

The Telco ecosystem has better reliability as it must comply with regulatory and contractual aspects, but it lacks the global reach and flexibility. Network operators need to use specialized network services for their managed Voice over IP (VoIP) solutions, because they are an alternative to traditional telephony and need to meet similar reliability requirements.

These specialized network services are needed since IP networks congestion directly impacts VoIP traffic [4]. At interconnection congestion occurs because of disagreements between different actors especially when the traffic becomes too asymmetric [5]. Furthermore, the traffic is affected by increasing number of peer-to-peer (P2P) applications and applications with video and rich media content. In wireline networks aggregated traffic has grown, since the number of devices on a single link increased. [6] In mobile networks radio resources are also limited, since they are shared by a number of users in a cell. [1]

Users are generally less demanding when using Web real-time communications, like WebRTC solutions, since they usually do not pay directly for the communication service itself. But these communications are as vulnerable to Internet congestions as Telco VoIP services. They do not benefit from any specialized network services, but WebRTC communication service providers improve the quality by adapting their applications, e.g. using adaptive codecs or browser congestion control mechanisms [16]. Therefore with growing traffic demand and user expectations, these mechanisms become insufficient. This is especially the case for WebRTC applications that are used on smartphones, so additionally they are impacted by mobility of users, radio layer mechanisms aimed at isolating upper layers from high transmission error ratios, limited spectrum resources of radio access networks and device constraints, e.g. battery usage.

WebRTC traffic would benefit from using dedicated specialized network services offered by network operators. However providing these services is not trivial since in WebRTC there is a separation of network and communication layers. Thus it is difficult to identify WebRTC flows that would be eligible for these services.

AWeS'15, April 21 2015, Bordeaux, France

ACM 978-1-4503-3477-8/15/04…$15.00
http://dx.doi.org/10.1145/2749215.2749218

This paper analyses the business and technical approaches of offering specialized network services to WebRTC communications and proposes a solution compatible with currently used technologies.

The paper is structured as follows. Section 2 identifies business models but also issues and research opportunities of offering specialized network services to Web actors. Section 3 gives an overview on current solutions and presents a proposed implementation. Section 4 focuses on flow identification mechanisms in access wireline networks. In section 5 conclusion and future research works are discussed.

2. Specialized network services

2.1 Approaches and business models

Two research approaches for offering specialized network services for WebRTC can be identified. "Over-the-network" approach allows a communication service provider to work independently from network operator and to choose the most advantageous best-effort path by using real-time network statistics. "In-network" approach, which is chosen in this paper, assumes cooperation between web companies and network operators. [1]

In "in-network" approach two business models are possible:

- Improving quality in general for everyone, i.e. by offering neutral specialized network services;
- Improving quality for given actors in a business to business to client (B2B2C) model by offering monetized specialized network services.

Each business model will have different concerns. This paper focuses on monetized specialized network services. Neutral solutions are going to be a subject of future studies.

2.2 Issues and research opportunities

Solutions and mechanisms used in Telco ecosystems cannot be directly applied to Web ecosystems since, since both ecosystems are very different. As a result, when offering specialized network services, network operators should adapt to web communication service providers. Otherwise they can end up offering a service that is not attractive for Web companies. Thus it is important to analyse issues and different blocking points of offering specialized services to web real-time communications.

Web companies when offering communication services rely on loosely coupled elements, i.e. their applications including mashups (i.e. application using web services provided by 3rd parties), but also identity providers, device-side platforms, cloud services and networks [2]. These over the top (OTT) communication services have Internet-wide reach but depend on not necessarily cooperating multiple actors.

Endpoints can belong to access networks of different types (mobile, WiFi, wireline,…) and be provided by different network service providers. So even if one end-user is eligible to benefit from a better quality, it is not sure that the other end-user can benefit from the same performances.

Endpoints used for WebRTC communication often have multiple network interfaces and can obtain additional addresses from STUN [14] and TURN [15] servers provided by different actors. The ICE mechanism [12] ensures that two endpoints have a functional media path between them, i.e. it chooses a pair of candidates to use. Currently, there are priorities assigned to different pairs of candidates, and as a result direct communication is privileged and communication with TURN servers is chosen as a last resort. This approach is static and once the candidates' pair is selected it is infrequently (practically never) changed. There are ongoing works on improvements to ICE candidate nomination [13]. The aim of these improvements is to allow the controlling endpoint to change the selected pair of ICE candidates, based on metrics like RTT (Round Trip Time) and loss rate. So far all these mechanisms are meant to be implemented directly in the browsers and there is no API allowing communication service provider to choose a preferred path.

There can be multiple interconnection paths possible:

- Direct interconnection between network operators supporting specialized services;
- Interconnection with multiple networks by using connection hubs, i.e. IPX;
- Best-effort interconnection where there can be actors in the media path that not necessarily support specialized services.

Direct interconnections and connection hubs can assure a certain service level agreement (SLA), but may introduce additional cost. However, best-effort connection can also assure sufficient performances. So far there is no mechanism allowing selection between these different types of interconnections.

Another issue is that a communication service provider, that offers a global solution, will not contact each actor in the media path to ask for specialized network service, since it becomes too complex. At this point there is no single contact point that would solve this issue.

Likewise when there are several actors implicated in the media path, it is not trivial to assure monitoring and troubleshooting, because there is no single actor, that would have a global visibility over the whole connection.

Last but not least, specialized network services may introduce additional treatment and possibly some delay. It is not sure yet at what point there are going to be beneficial when compared to best-effort routing.

All the above issues show the importance of rethinking current models and proposing new approaches applicable to current and evolving solutions.

Offering end-to-end specialized network services and QoS is complex given global aspects of web communication services. OTT traffic is difficult to predict for network operators, because they do not have control or even knowledge about different elements. However network operators can still offer some improvements in order to provide specialized network services for best-effort that would intervene in the most sensitive network segments, e.g. access and peering points.

As a result different use cases will be studied in order to improve quality where it is needed by taking into account eventual implementation cost and feasible benefits.

This study will focus on evaluation of specialized network services in access networks for different use cases, since access network prioritization is essential whenever there are limited resources, e.g. for cellular networks where resources are shared by users in a cell. Also improving ICE mechanism and offering a network enabler for path selection will be studied in detail, because current proposals stay very over the top, whereas they could benefit from the support of network operators giving more visibility over network performances. This would also allow choosing a path based on whether the SLA is assured or not.

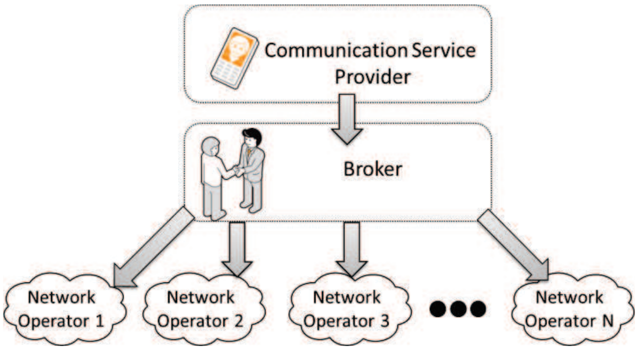

Figure 1. Broker

An appropriate relationship should be established between given network operators and a given communication service provider. That will allow offering a service that is global. However each communication provider cannot establish a relationship with each network operator. As a result a single contact point will be created. In order to achieve that, an abstraction layer, i.e. a broker, will be needed (Figure 1). So far it is not sure who could take a role of a broker (a network operator or a new 3rd party player). Also the exact responsibilities of this actor need to be defined.

This research will lead to proposing a new model of offering specialized network services to web real-time communication service providers.

3. Proposed implementation

To offer specialized network services to WebRTC communication service providers, some principles used by network operators for managed VoIP can be used. Thus, to show analogy with WebRTC, these mechanisms will be discussed below.

In managed VoIP solutions, provided by network operators, the end-to-end service is assured. It is possible because each network operator provides service for its network and interconnects with other network operators. To ensure this, network operators must control the whole ecosystem, i.e. devices, services and networks [1].

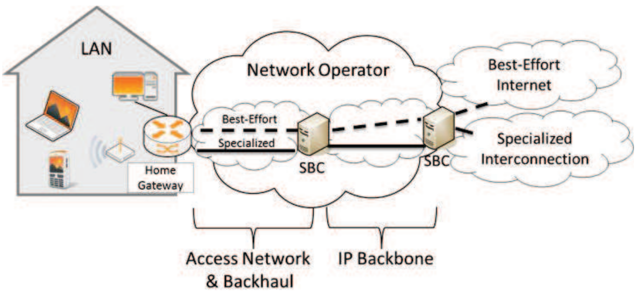

Figure 2. Managed VoIP specialized services

The common practice is presented in Figure 2 in a simplified way.

A network operator starts by prioritizing flows in access network by using mechanisms based on DiffServ IP (or distinct ATM Virtual Circuits). These flows are separated into VLANs (or ATM Virtual Paths) in the backhaul, and are later sorted and separated into IP-VPNs in the IP backbone. At interconnection either dedicated interconnections between two network operators or connection hubs are used. To control differentiated services Session Border Controllers are deployed (SBC). SBCs are media relays that sort and separate the flows at access and interconnection level. They simplify security and traffic management.

In WebRTC the media plane is separated from the signaling plane. Web companies manage everything at the application level. Also in WebRTC direct communication between hosts (P2P media) is privileged. Media relays, i.e. TURN servers are used as a last resort, so the media is not steered in any specific way and just uses best-effort routing.

SBC media steering is shown in Figure 2. It is essential in managed VoIP when it comes to coupling the network and the application layer. Using SBC is similar to using TURN servers in WebRTC. As a result both mechanisms were analysed in order to propose a future implementation.

So far TURN servers are only used to relay the traffic and they do not make any flow differentiation. The idea of the proposed solution is to offer the specialized flow treatment, meaning steering the flows through specialized network paths. That would enable coupling between the application and network layer. The subject of improving the use of TURN servers for WebRTC is a very up-to-date topic since there is a lot of ongoing work in IETF TRAM (TURN Revised and Modernized) group [7].

The solution of offering TURN servers by network operators was chosen for analysis since it solves several issues of offering specialized network services to WebRTC communication services. It addresses access and interconnection network segments. Additionally, it is compatible with existing solutions so it can be introduced incrementally. It is also in line with technological choices of web companies and with evolution of different standards and solutions. Thus it offers a collaborative approach between web companies and network operators.

3.1 Proposed architectures

When working on proposed solution, it is important to take under account a fact that TURN servers can be provided by different actors and that can be located in different network segments. Since the backward compatibility is essential, all those aspect need to be considered.

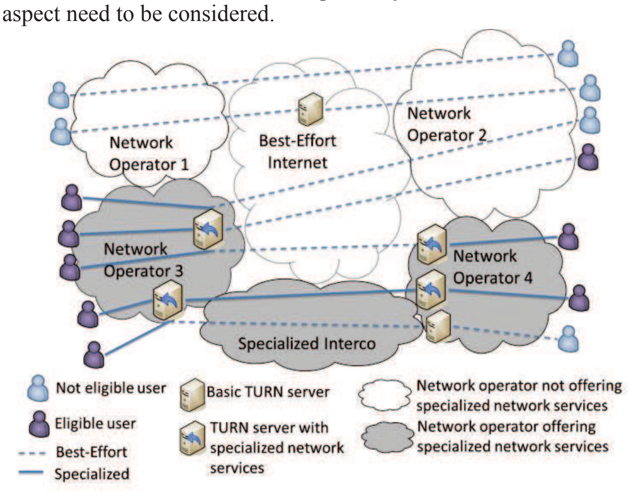

Figure 3. Exemplary communications

As it was discussed in Section 2, it is not always possible to offer end-to-end service. As a result proposed implementation gives a possibility of offering a service for different network segments.

Figure 3 gives an overview of several possible connection scenarios. If only one user is eligible to use specialized network services, a better service can still be offered in his access network. If both users are eligible, more segments can be impacted, i.e. access networks of both users and interconnection network. It is important to note that in order to assure specialized flow treatment in access networks, a TURN server needs to be in the network that a given network operator controls, i.e. access, backhaul or IP backbone and not in the open Internet. Specialized network service at interconnection will depend on interconnection type, i.e. if it is best effort or dedicated interconnection, e.g. connection hub.

The exact QoS algorithms and mechanisms are out of scope of this paper, but it can be assumed that the same technics as in managed VoIP (and quoted in Section 2) can be used.

3.1.1 Decision making mechanism at endpoint

In the first architecture two pools of servers are offered: basic and specialized (or one TURN server with multiple input interfaces and corresponding output interfaces – there is no functional difference since it is only deployment specific).

A decision about users' eligibility to the specialized network services is made when choosing a TURN server, i.e. at connection establishment. It thus relies on the TURN discovery mechanism. So when end user decides to make a call, the application verifies its eligibility to the specialized network services and attributes an appropriate TURN server (or input interface of a TURN server). Since the TURN server is known, flow prioritisation in the access network can be achieved. Accordingly, the corresponding output interface is linked either to best-effort or to prioritized path.

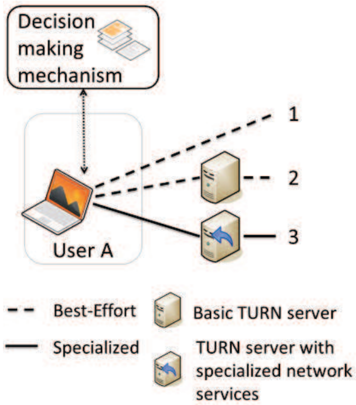

Figure 4. Multiple input and output interfaces

Figure 4 shows simplified users' flows, based on their eligibility of using specialized network services would be:

1. Sent over P2P, so directly between two hosts;
2. Relayed by using basic TURN server;
3. Relayed by using specialized TURN server.

3.1.2 Decision making mechanism at TURN server

In the second architecture additional functionality would be added to a TURN server. In existing implementations, there is the possibility to assign multiple relay addresses to a TURN server. However there is no mechanism for choosing one of them. They are equivalent for load balancing purposes [8]. In proposed architecture there would be an intelligent way of choosing relay address (output interface), i.e. a TURN server would have a single input interface and multiple output interfaces. As a result a decision about users' eligibility of using specialized network services is made by this TURN server.

Based on decision made by a TURN server access and interconnection prioritization can be offered, as it is done in the first proposed architecture.

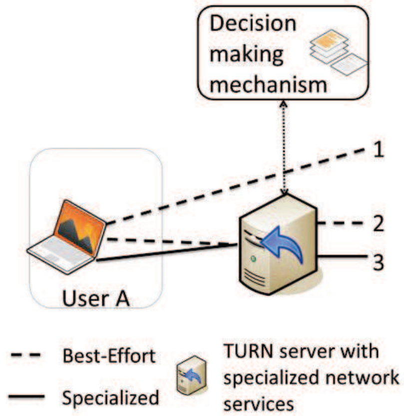

Figure 5. Single input and multiple output interfaces

Figure 5 shows simplified users' flows, based on their eligibility of using specialized network services would be:

1. Sent over P2P, so directly between two hosts;
2. Relayed by basic output interface of multi-interface TURN server;
3. Relayed by using specialized output interface of multi-interface TURN servers.

4. Flow identification

4.1 Motivation

One of the major issues, when offering specialized traffic treatment to WebRTC, is flow identification. It is important in order to distinguish WebRTC flows from normal best-effort traffic and to carry out specialized flow steering. WebRTC uses best-effort Internet, preferably P2P communication and does not provide any signalling information since the signalling plane is not standardized. As a result network operator does not have visibility over OTT traffic. Therefore flows have to be already identified in the access networks.

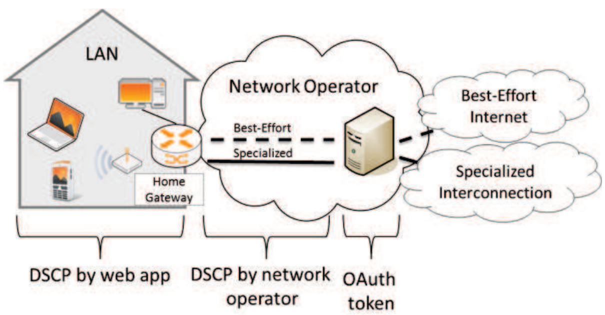

Figure 6. Wireline access networks flow identification

There are different network segments that need to be analysed since they have various constraints. In each network segment different actors are implicated (Figure 6): actors that command dedicated services, i.e. communication service providers (and users indirectly) and actors that provide these specialized network services to eligible users.

Network operators provide specialized services and allocate resources, but they do not necessarily have control over users' eligibility or authorisation. For that reason mechanisms allowing collaborative cooperation are needed. Consequently trust relationship between different actors is essential.

Home (LAN) networks cannot be fully controlled by network operators or communication services providers. Also not all LAN networks are managed in a specific way, so a default, universal mechanisms for WebRTC flows handling should be proposed. The impact of eventual specialized flow treatment is limited to a given LAN location.

Operators' networks are fully controlled, because they are shared by many users and transport different traffic and eventual errors while prioritization may impact other users. Given regulations and responsibilities of network operators, it is not possible to delegate control to a 3rd party. Therefore mechanisms that allow collaborative approach but are controlled and secure need to be used.

TURN server is typically a resource that can be provided by a network operator. However its control should be coordinated with communication service provider. It can be used for differentiating flows and allowing specialized steering. For that it needs a collaborative authorization mechanism.

Several existing mechanisms were identified and assessed. All of these mechanisms ensure flow identification, but each of them impacts different network segments and is controlled by different actors, as a result they are complementary.

4.2 DSCP marking by an application

Networks can provide per packet treatment and differentiated QoS based on Differentiated Service Code Point (DSCP). However it is important to mark packets appropriately.

IETF draft "DSCP and other packet markings for RTCWeb QoS" [9] provides recommended DSCP values for browsers. It indicates that packet marking may frequently help in environments like enterprise networks and residential networks where there is a risk of congestion on some links. The draft also specifies that even though usually DSCP is site specific, it may be needed to introduce some default values that browsers can use, when there are no site specific recommendations.

Even though packet markings by a WebRTC application (or browser) may help to avoid congestion on a given site, they have to be limited to a given area. Network operators cannot trust an over-the-top application, especially since DSCP values can be easily modified. Thus at home gateway all DSCP tags, marked by applications or network equipment in LAN networks, are discarded.

4.3 DSCP marking by a home gateway

Even though network operators do not trust end-device markings, they still use DSCP markings. The QoS management starts at the home gateway and is assured for end-devices connected directly to it. This marking can be fully trusted by network operator since it is done by its equipment and in the environment that it controls. It allows flow prioritization in network controlled by a given network operator.

In order to use appropriate DSCP values one of the following mechanisms can be used.

4.3.1 Fix address of TURN servers

In managed VoIP packets coming to and from SBC can be marked with given DSCP values since SBCs IP addresses are known to the home gateway.

The same principle can be used in presented solutions, since network operators provide TURN servers and they know their IP addresses. These addresses have to be provided to the home gateway, so it can give a specific DSCP value to packets going and coming from this TURN server. TURN server addresses can be either configured manually or by using DNS. Since manual configuration may cause scalability issues, DNS solution is preferred.

4.3.2 Heritage

Heritage mechanism is based on trusting downstream DSCP marking information. So when an end-device initiates a call, a home gateway forwards its traffic with best effort DSCP tag. The home gateway tracks the connection and if this end-device receives traffic that is marked by the network in downstream direction, its upstream traffic inherits this downstream marking.

This solution solves the scalability issue of manual configuration that could occur when using the above mechanism.

4.4 OAuth by TURN server

The OAuth authorization framework allows a 3rd party application to get limited access to an HTTP service on behalf of the resource owner [10]. As a result flows eligible to access specialized network services can be identified by using OAuth mechanism, meaning that the ephemeral tokens can be used to ensure controlled access to STUN or TURN servers [11].

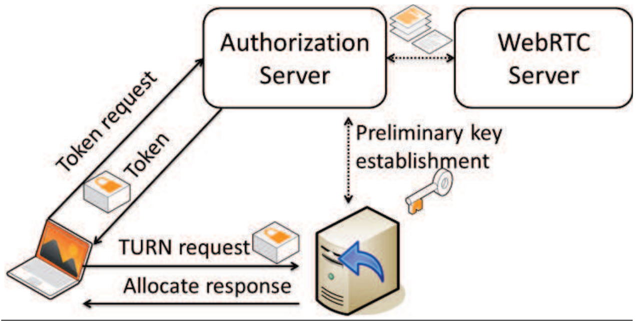

Figure 7. OAuth for WebRTC [11]

In this mechanism, the WebRTC client authorized by the WebRTC server obtains limited access to a TURN server (Figure 7). TURN server provided by network operator is a resource server that can be used by authorized clients. The authorization server provides tokens to the authorized users. It also establishes keys with TURN server that will be used for token verification and is responsible of verifying their eligibility. WebRTC server is a resource owner, meaning that it informs the authorization server which users are eligible to access TURN server and with which permissions.

As it is presented above OAuth can be used to authorize access to a TURN server and to specialized interconnection network services. As a result it indirectly offers flow identification, since it allows their differentiated treatment. Consequently its role may be enlarged when implementing architecture discussed in section 3.2.2, so it can also allow deciding if a given flow can benefit from basic or specialized treatment. This should be done by defining different permission levels, by e.g. using the scopes [10].

This mechanism needs a collaborative approach, since the resources are managed by network operator, but tokens are generated by an authorization server with permissions done by the WebRTC server, belonging to communication service provider.

The authorization server and WebRTC server may be provided by the same actor. Although, there is a possibility of creating a new profession, so that authorization server would be offered by a 3rd party. This would be especially beneficial for more complex architectures, where there are multiple TURN server providers. As a result authorization server could be offered by e.g. a broker.

This mechanism does directly enable prioritisation for access network, since it concerns flows that have already reached a TURN server. However it can have an important role in implementing interconnection prioritization, since it can be used for authorizing flows to access specialized interconnection services, e.g. direct connection between two TURN servers.

4.5 Assessment

In order to provide a full implementation, a mix of different mechanisms is needed. This would allow impacting different network segments, i.e. from end-device to TURN server. And since each network segment has different requirements and is controlled in a separate way, a trust relationship needs to be established between network operators and communication service providers.

The proposed mechanisms are known and currently used, so the work should be done mostly on interactions between different actors, especially in most complex cases, where there are several actors in a media path and when end-users do not have the same privileges.

Therefore the proposed mechanisms need to be implemented in a proof of concept in order to test various scenarios and identify eventual blocking points.

5. Conclusion

Emerging WebRTC technology simplifies creating web real-time communication services. The fact that this technology is currently under standardisation makes it even more accessible. As a result, it may lead to diminishing the importance of voice services offered by network operators.

Currently web real-time communications have to use best-effort routing. They adapt to provided network performances by using the mechanisms built in the applications. Whereas they could benefit from specialized network services offered by network operators.

In this article, different research opportunities are identified including two possible business models based on neutral or monetized specialized network services. Also motivations, challenges and issues of offering specialized network services to WebRTC communication service providers are analysed. The difficulties of providing end-to-end specialized network services are discussed.

An architecture enabling collaborative approach for monetized model is presented. Therefore flow identification mechanisms, necessary to differentiate WebRTC flows, are assessed.

The future work would consist of answering research questions that are revealed in this article.

Specialized network services for access networks will be studied in detail. The focus will be made especially on mobile networks.

Different possibilities of choosing interconnection paths will be evaluated, including assessment of paths with and without SLA. The decision making mechanism for choosing media path will be analysed in detail, in order to decide which entity assigns specialized services and if this decision is static or dynamic.

Later, interactions between different actors will be analysed in order to define a role of a broker more precisely and to propose Internet-wide APIs.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645342; project reTHINK.

References

[1] E. Janczukowicz, S. Tuffin, A.Braud, A. Bouabdallah, G. Fromentoux, J.M. Bonnin. Approaches for Offering QoS and Specialized Traffic Treatment for WebRTC. In Advances in Communication Networking, pages 59-69. Springer International Publishing, 2014.

[2] S. Bécot, E. Bertin, J.M. Crom, V. Frey, S. Tuffin. Communication Services in the Web Era. In ICIN Proceedings, 2015.

[3] WebRTC 1.0: Real-time Communication Between Browsers, http://w3c.github.io/webrtc-pc/

[4] X. Chen, C. Wang, D. Xuan, Z. Li, Y. Min, W. Zhao. Survey on QoS management of VoIP. In Computer Networks and Mobile Computing, pages 69-77. IEEE, 2003.

[5] D. Clark, S. Bauer, W. Lehr, KC. Claffy, A. Dhamdhere, B. Huffaker, M. Luckie. Measurement and Analysis of Internet Interconnection and Congestion. In TPRC Proceedings, 2014.

[6] S. Bauer, D. Clark, W. Lehr. The Evolution of Internet Congestion. In TPRC Proceedings, 2009.

[7] TRAM, https://datatracker.ietf.org/wg/tram/documents/

[8] coturn, https://code.google.com/p/coturn/

[9] DSCP and other packet markings for RTCWeb QoS, https://tools.ietf.org/html/draft-ietf-tsvwg-rtcweb-qos-03

[10] The OAuth 2.0 Authorization Framework, https://tools.ietf.org/html/rfc6749

[11] Session Traversal Utilities for NAT (STUN) Extension for Third Party Authorization, http://tools.ietf.org/html/draft-ietf-tram-turn-third-party-authz-10

[12] Interactive Connectivity Establishment (ICE): A Protocol for Network Address Translator (NAT) Traversal for Offer/Answer Protocols, https://tools.ietf.org/html/rfc5245

[13] Improvements to ICE Candidate Nomination, http://juberti.github.io/draughts/nombis/draft-uberti-mmusic-nombis.html

[14] Session Traversal Utilities for NAT (STUN), http://tools.ietf.org/html/rfc5389

[15] Traversal Using Relays around NAT (TURN), https://tools.ietf.org/html/rfc5766

[16] V Singh, A. A. Lozano, J. Ott, Performance analysis of receive-side real-time congestion control for WebRTC. in Proc. Intl. Packet Video Workshop, 2013