Performance Analysis of an Adaptive Media Content Streaming System

Marius Vochin[1,*], Eugen Borcoci[1], Serban Georgica Obreja[1], Jordi Mongay Batalla[2], Daniel Negru[3]

[1]University POLITEHNICA of Bucharest, Telecommunications, Bucharest, Romania
[2]National Institute of Telecommunications, Warsaw, Poland
[3]LaBRI Lab, University of Bordeaux Bordeaux, France

[*]Corresponding author (E-mail: marius.vochin@elcom.pub.ro)

***Abstract* — Adaptive content streaming is frequently used as an efficient and cheap solution to achieve a good quality for media streaming, in systems having light Over-the–Top architectures. The streaming system developed here makes an initial optimized server selection based on multi-criteria algorithms and then in-session media adaptation. The focus of this paper is on performance analysis (based on real-life testbed experiments) as part of the overall system validation.**

***Keywords* - media content delivery, server and path selection, quality of services, quality of experience, dynamic adaptive streaming, monitoring.**

I. INTRODUCTION

The media content services are more and more used in the current and will be in the future Internet. Recently, cheap "over-the-top" (OTT) solutions are developed, for media/content delivery, where the Service Providers (SP) offers services delivered on top of the current Internet – but without strong Quality of Services (QoS) guarantees. The OTT SP could exist as a separate entity from traditional Internet Service Provider (ISP). The OTT–like architectures are much cheaper and less complex than Information Centric Networking (ICN) ones [1][2], or Content Delivery Network (CDN) [3].

The European *DISEDAN* Chist-Era project *(**DI**stributed **SE**lection of content streaming source and **D**ual **A**daptatio**N**,* 2014-2015) [4], proposes a novel OTT light architecture, for media content streaming systems, working over the current IP multi-domain networks. Several business entities/actors are involved. The *Service Provider (SP)* offers content services to the users. It might be (but not mandatory in OTT model), owner and manager of a transportation network. The *End Users (EU)* consumes the content; they are using *End User Terminals* (EUT) to request and then consume media content. A distinct *Content Provider (CP)* entity could exist as owner of several *Content Servers (CS).*

The DISEDAN project is focused on assuring a good Quality of Experience (QoE) when delivering content streams in real-time for the EUs. So, it does not deal with contractual CP-SP relationships; we assume that CSs are owned and managed by the SP. Two main problems should be solved: optimal initial content *server selection* and optimised *media stream delivery* during the session. A novel solution is proposed in DISEDAN for the best content source (server) selection (this is a multi-criteria-hard problem) by considering as input parameters the user context, servers' availability and requested content. The novelty consists in: (1) initial *two-step server selection mechanism* (at SP and at EUT sides), using multi-criteria decision algorithms that consider context- and content-awareness; (2) *dual adaptation mechanism*, running during media session. The latter consists of *media flow adaptation* and/or *content source adaptation* (i.e., *streaming server switching*) when the quality observed by the user suffers degradation during the media session.

For in-session media adaptation, the *Dynamic Adaptive Streaming over Hypertext Transfer Protocol- HTTP (DASH)* technology is used [5][6].

The DISEDAN solution could be rapidly deployed in the market, given its simpler architecture in comparison to complex ones like Content Oriented Networking[1][2] or CDNs [3].

The specific contribution of this paper is to present a part of the performance evaluation tests (focused on the server selection phase) of the DISEDAN implementation, executed on a testbed including the functional blocks defined in the system architecture.

The paper structure is the following: Section II is a short overview of related work. Section III outlines the overall DISEDAN architecture and main design blocks. Section IV contains the paper main contributions, focused on DISEDAN performance evaluation, executed on a real life testbed. Section V contains conclusions and future work outline.

II. RELATED WORK

The final overall QoE observed by an end user for a content streaming service depends on two main elements: the *servers* (sources) and *transportation* (thorough the network). These problems are solved in ICN solution by defining a novel architecture content-oriented [1][2] which replaces the host-to-host (*where*) communication paradigm with content-objects (*what*) based one. The ICN problem is the high complexity and cost introduced at network level (high processing and storage resources needed in routers). The Content Delivery Networks (CDN) solution improves the quality by replicating the same content object in several replica servers, geographically situated close to groups of users [3] and thus shortening the network segments. The CDN complexity and cost is high. That is why, DISEDAN adopted an OTT style and targeted the above two aspects in a combined and adaptive way.

The initial server selection can be based on optimization algorithms like *Multi-Criteria Decision Algorithms (MCDA)* or *Evolutionary Multi-objective Optimization algorithm* (EMO), [8][9]. The solution chosen for DISEDAN has been MCDA reference level selection algorithm, whose details are described in [9].

During the media session, real-time adaptation in content streaming can be used, to solve the fluctuations in QoE/QoS. One can classify adaptation as acting on Media (flow) and/or on CS. The *Media adaptation* is a main research innovation area in media streaming applications [5][6]. *CS adaptation* means a new content server selection (during the session) and switching, depending on the consumer device capabilities, consumer location, content servers state and/or network state. The DISEDAN novel architecture [4][7], combines the initial server selection (result of cooperation between SP and EU) with session-time dual adaptation, in a single solution.

DASH is a recent standard, for delivery of high quality multimedia content over the Internet, by using conventional HTTP Web servers [5][6]. It minimizes server processing power and is video codec agnostic. A DASH client continuously selects the highest possible video representation quality that ensures smooth play-out, in the current downloading conditions. This selection is performed on-the-fly, during video play-out, from a pre-defined discrete set of available video rates and with a pre-defined granularity (according to video segmentation). The standard ISO/IEC 23009-1 [5] defines the DASH-Metrics client reference model, composed of *DASH access client (DAC)*, followed by the *DASH-enabled application (DAE)* and *Media Output* (MO) module. The DAC issues HTTP requests (for DASH data structures), and receives HTTP request responses.

Consequently three observation points (interfaces – I/F) can be identified (see Figure 1):

- O1 *at network-DAC I/F*: a set of TCP connections, each defined by its destination IP address, initiation, connect and close times; a sequence of transmitted HTTP requests, each defined by its transmission time, contents, and the TCP connection on which it is sent; and for each HTTP response, the reception time and contents of the response header and the reception time of each byte of the response body.

- *O2 at DAC-DAE I/F*: consists of encoded media samples. Each encoded media sample is defined as: media type; decoding time; presentation time; the *@id* of the Representation from which the sample is taken; the delivery time.

- *O3 at DAE-MO I/F*: consists of decoded media samples. Each decoded media sample is defined as: the media type; the presentation timestamp of the sample (media time); the actual presentation time of the sample (real time); the @id of the Representation from which the sample is taken (the highest dependency level if the sample was constructed from multiple Representations).

III. DISEDAN ARCHITECTURE AND DESIGN

The definition and details of the system architecture and design are already given in [3][7][10]. This section presents a summary, to support the understanding of the paper.

The main business actors have been mentioned in Section I: SP, EU, CS. The connectivity between CSs and EU Terminals (EUT) is assured by traditional *Internet Services Providers (ISP) / Network Providers (NP)* - operators. The ISP/NPs do not enter explicitly in the business relationships set considered by DISEDAN, neither in the management architecture (DISEDAN works in OTT style).

The system can work over the traditional TCP/IP mono and/or multi-domain networks, but can be also applied in more complex business models, e.g., involving Cloud Providers, CDN providers, etc. The relationships between SP and such entities could exist, but their realization is out of scope of this study. While Service Level Agreements (SLAs) might be agreed between SP and ISPs/NPs, related to connectivity services offered by the latter to SP, such SLAs are not directly visible at DISEDAN system level.

The work [7] has identified all system requirements coming for EU, SP, CP and defined the system architecture (Figure 1). The functional blocks correspond respectively to SP, EUT and CS. Note that only blocks relevant to DISEDAN are shown.

The Service Provider (SP) includes in its Control Plane: *MPD File generator* – dynamically generates Media Presentation Description (MPD) XML file, containing media segments information (video resolution, bit rates, etc.), ranked list of recommended CSs and, optionally - current CSs state information and network state (if applicable); *Selection algorithm* – runs Step 1 of server selection process (it exploits *MCDA* [8][9] to rank the CSs and media representations); *Monitoring module* – collects information from CSs and processes it to estimate the current state of each CS.

The End User Terminal (EUT) includes the modules: *Data Plane: DASH (access and application)* – parses the MD file received from SP and handles the download of media segments from CS; *Media Player* – playbacks the downloaded media segments (see [5][6] for details); *Control Plane: Content Source Selection and Adaptation engine* – implements the dual adaptation mechanism; *Selection algorithm* – performs the Step 2 of server selection process. It can also exploit MCDA, or other algorithms to select the best CS from those recommended by SP; *Monitoring module* – monitors changing (local) network and server conditions.

The CS entity includes the modules: *Data Plane: Streaming module* – sends media segments requested by End Users; *Monitoring module* – monitors CS performance metrics (CPU utilization, network interfaces utilization, etc.).

Figure 1 shows also the main functional steps: (1) EUT issues to SP a media file request. (2) SP analyzes the status of the CSs and runs the CS selection algorithm. (3) SP returns to EUT an ordered list of candidates CS (SP proposal, embedded in a MD - xml) file. (4) The EUT finally selects the CS, by running its own algorithm. (5) EUT starts asking segments from the selected CS. During media session, the EUT makes quality and context measurements. Continuous media flow DASH adaptation is applied, or, (6) CS switching is decided. The steps 1-2-3-4 correspond to initial serve selection phase and 5-6 to the media session phase. During the receipt of consecutive chunks, the user's application can automatically change the rate of the content stream (based on DASH measurements, which are out of scope in this paper) and/or also can switch to another CS.

Following the architecture definition, the design and implementation of the system have been performed. Details on the Control Plane design are given in [10].

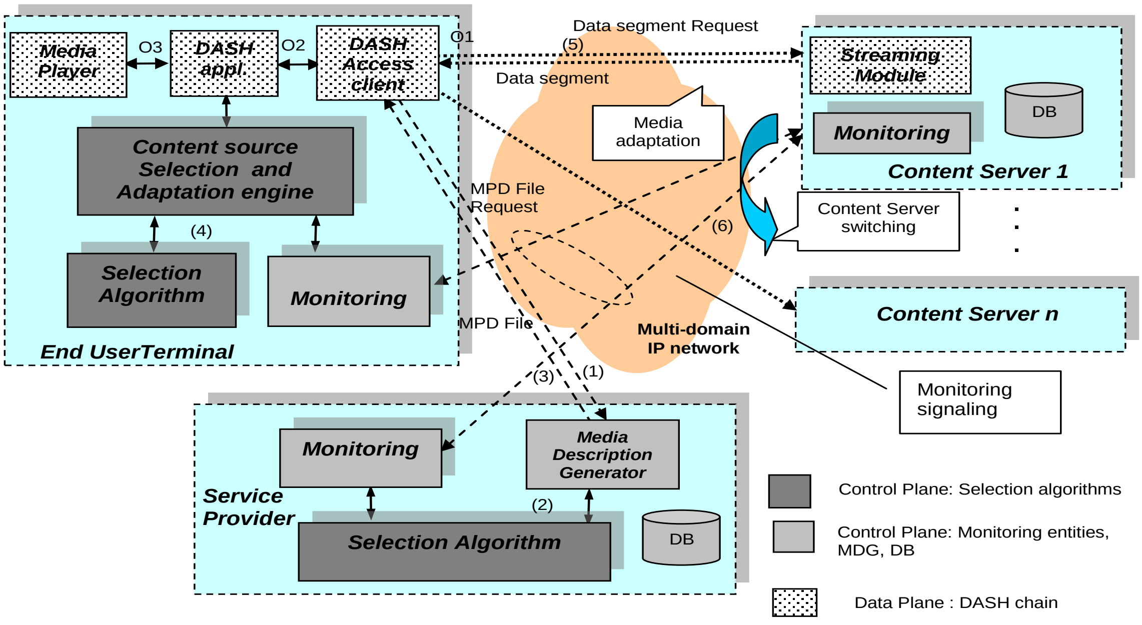

Figure 1 DISEDAN architecture

IV. SYSTEM PERFORMANCE EVALUATION

An experimental testbed has been built for DISEDAN functionalities validation (Figure 2). The system comprises three independent IP network domains, equipped with several core and edge/border routers (Linux based). No QoS technologies are active in these networks. Several DISEDAN entities are connected to this network: SP, EU, and CS through some access networks. Note that the presence of the access networks in the overall system is not essential, given the OTT-style of work for DISEDAN.

Several tests were run in order to investigate Disedan End User Terminal (EUT) performance aspects, for both Linux and Android implementations. The MCDA algorithm is computed in order to select the best available server when EUT client is launched. The selection process takes into account parameters likes hop count, RTT and available server bandwidth, obtained from the monitoring system. Summary example selected from the tests is given here.

In the first test, EUT client software was installed on Android v4.2 device with an Intel Atom Dual-core 2.0 GHz CPU, 2 GB of RAM memory and a dual-band Wi-Fi n connectivity. Wi-Fi network connection was established through a 450Mbps N dual-band access point in 10.2.5.* subnet. MCDA algorithm has selected CS2 as best server available, and ping RTT sent from EUT to e126 was 17ms with ttl 61. Best server determination phase duration was 28 seconds, and after 2 more seconds the movie displayed. During MCDA calculation phase, the SP python process running on e130 cpu peak load usage was of 0.2% for about 3 seconds, and MCDA calculation time was 0.2 microseconds. At the server monitoring side, the SP total communication time is 9.17 seconds (with 3 monitoring agents enabled), and at the agent side *nodejs* process maximal cpu peak load was of 0.2% for about 5 seconds.

The movie flow bandwidth was of 24KB, with an average of 19 pps, and android EUT cpu load was ranging between 0.22 to 0.38%; apache streaming server determined load was 0.1%.

In the second test, EUT client was installed on Linux Debian 6.0, running on a 2.6GHz quad-core Intel Xeon CPU, with 4GB of DDR 3 RAM memory, and 1000Mbps Ethernet link. The MCDA calculation time phase was about 15 seconds, and EUT client software player (VLC with *libdash*) topped about 14 to 16% of CPU usage.

V. CONCLUSION

A complete set of functional and performance–oriented tests have been defined in the project. The results validated the architecture, and implementation demonstrating the proof of concepts. Specific sample results presented in this paper illustrate the basic functionalities of the DISEDAN system's modules and the operating mode of the MCDA algorithm in the first phase of the content server's selection process, in different content servers and network load conditions.

More detailed results on final validation are given in [11].

Future work can include the extension of DISEDAN concepts for content delivery systems in wireless and mobile environment, where the server selection and handover is fit to the DISEDAN solution.

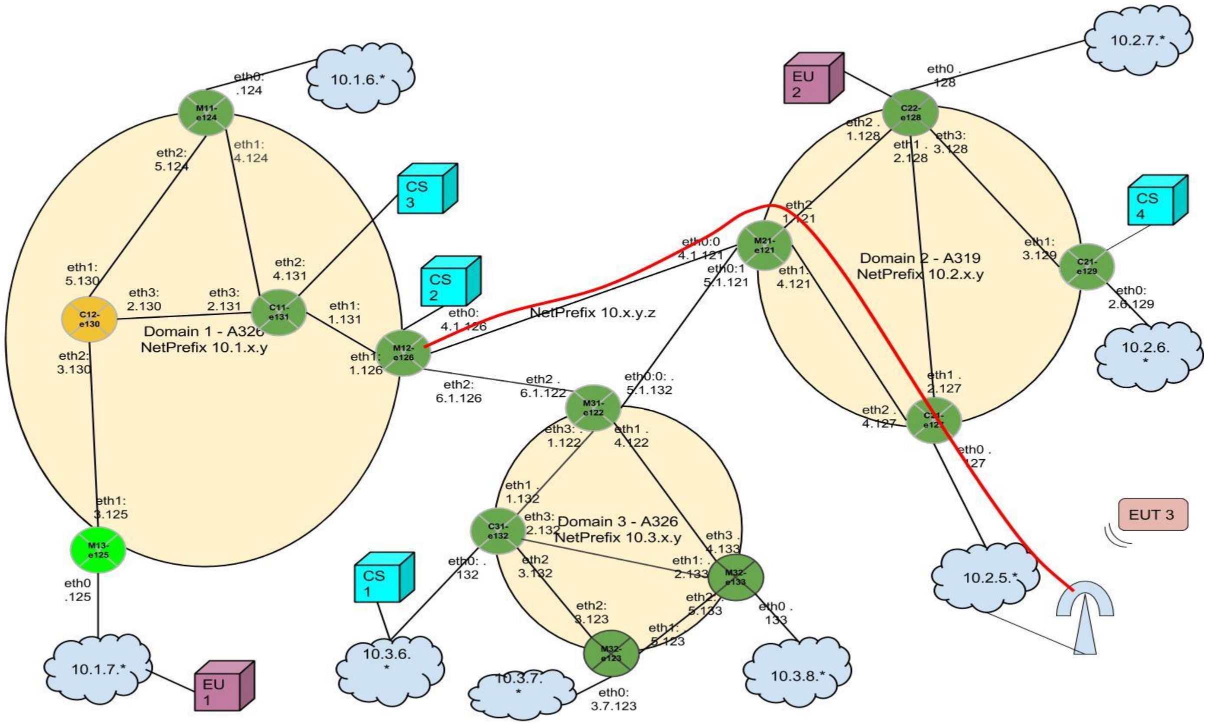

Figure 2 Disedan Experimental testbed

Acknowledgment

This work has been partially supported by the Research Project DISEDAN, No.3-CHIST-ERA C3N, 2014- 2015.

References

[1] J. Choi, J. Han, E. Cho, T. Kwon, and Y. Choi, "A Survey on Content-Oriented Networking for Efficient Content Delivery", IEEE Communications Magazine, March 2011, pp. 121-127.

[2] ***, "Information-Centric Networking-3", Dagstuhl Seminar, July 13-16 2014, Available from: http://www.dagstuhl.de/en/program/calendar/semhp/?seminar=14291

[3] P. A. Khan and B. Rajkumar, "A Taxonomy and Survey of Content Delivery Networks". Department of Computer Science and Software Engineering, University of Melbourne. Australia : s.n., 2008. www.cloudbus.org/reports/CDN-Taxonomy.pdf.

[4] *** http://wp2.tele.pw.edu.pl/disedan/ [retrieved:May, 2015]

[5] ISO/IEC 23009-1, "Information technology -- Dynamic adaptive streaming over HTTP (DASH) -- Part 1: Media presentation description and segment formats," ISO/IEC, Geneve, second edition, 2014.

[6] I. Sodagar, "The MPEG-DASH Standard for Multimedia Streaming Over the Internet," MultiMedia, IEEE, vol. 18, no. 4, 2011, pp. 62 - 67.

[7] E. Borcoci, ed., et al., "D2.1 System requirements and comparative analysis of existing solutions for media content server selection and media adaptation", DISEDAN Project, July 2014, http://wp2.tele.pw.edu.pl/disedan.

[8] A. Beben, J. M. Batalla, W. Chai, and J. Sliwinski, "Multi-criteria decision algorithms for efficient content delivery in content networks", Annals of Telecommunications - annales des telecommunications, Springer, vol. 68, Issue 3, 2013, pp. 153-165.

[9] E. Borcoci, M. Vochin, M. Constantinescu, J. M. Batalla, and D. Negru, "On Server and Path Selection Algorithms and Policies in a light Content-Aware Networking Architecture", ICSNC 2014 Conference, http://www.elcom.pub.ro/disedan/docs/ICSNC%202014%20Conf.pdf.

[10] E. Borcoci, C. Cernat, and R. Iorga, "Control Plane Design for a Content Streaming System with Dual Adaptation", AICT 2015 Conference, https://www.thinkmind.org/index.php?view=article&articleid=aict_2015_6_40_10154

[11] D. Negru, J. Bruneau-Queyreix (LABRI), J. M. Batalla, A. Bęben, P. Krawiec, P. Wiśniewski (WUT), S. Obreja, C. Cernat, R. Iorga, M. Vochin (UPB), D4.2 "Final Validation of the DISEDAN integrated Pilot", Dec. 2015, http://www.elcom.pub.ro/disedan/docs/