Software Defined Cognitive Networking: Supporting Intelligent Online Video Streaming

Mu Mu
The University of Northampton, Northampton, UK
mu.mu@northampton.ac.uk

***Abstract*—Adaptive media such as HTTP adaptive streaming (HAS) is becoming a standard tool for online video distribution. The non-cooperative competition of network resources between a growing number of adaptive streaming applications has a significant detrimental impact on the user experience and network efficiency. Existing network infrastructures often prioritise fast packet forwarding and not the quality of the delivered content. Future network management must leverage application and user-level cognitive factors to allocate scarce network resources effectively and intelligently. Our software defined cognitive networking (SDCN) project aims at incorporating new developments in human cognition, media technology and communication networks to ensure the user experience, user-level fairness and network efficiency of online adaptive media.**

I. Introduction

Online audio-visual content distribution has become a cornerstone of digital economy and creative industry. With the proliferation of video streaming services and the increasing online presence of traditional broadcasters, users are spending more time consuming media content over the Internet [3]. Meanwhile, adaptive media such as HTTP adaptive streaming (HAS) is becoming the standard tool to streamline media distribution on various types of user devices over heterogeneous networks. HAS delivers media content in adaptation sets of one or multiple representations. Each representation is a version of the same content using a different encoding scheme. User devices can choose the most suitable media representation on-the-fly based on client-side statistics such as network throughput and player buffer occupancy.

With the growing deployment of adaptive media applications, a number of fundamental challenges emerged, of which the most pressing is the unsupervised and non-cooperative competition of network resources between media applications. Such competition leads to detrimental quality degradations to the delivered media content and an inefficient share of network resources. Conventionally, over-provisioning and client-side optimisation are possible solutions to mitigate the issue. However, in the face of the vast and growing media consumption, resource over-provisioning does not resolve network bottlenecks or fluctuations especially with "content hopping" and flash-crowd during peak hours. Single-stream HAS optimisation algorithms [12], [8] are unaware of each other on the same network. As a result, the resource competition between adaptive media causes highly bursty traffic and fluctuating user experience [5], [8]. Existing network management models are not designed with adaptive media in mind and their goal is to maximise the aggregate utility based on the assumption that all applications share the same characteristics [6]. Research in video quality assessment has shown that there is a number of crucial cognitive factors affecting the user experience of online media services [1]. These factors determine the non-linear correlation between the throughput of a video stream and its actual perceptual quality [2]. For adaptive media, the switching between media representations in different frequencies and magnitudes also lead to complex and perceivable impact [10]. Furthermore, the inequality of media quality between applications on the same network, caused by unbalanced sharing of network resources, also affects a user's overall experience [11]. Therefore, resource provisioning must consider the user-level QoE fairness between applications and devices as a primary factor. Unfortunately there is a lack of systematic studies on modelling the combined impact attributed to picture fidelity, motion, switching event, content characteristics, session duration and other perceptual and psychological factors [13].

Improved resource provisioning for better QoE and fairness hinges on a fundamental premise: the networks can accurately and efficiently estimate and ensure user experience on media flows at scale. Most video quality assessment models such as Netflix's VMAF model [7], are built for offline evaluation and not suitable for synchronous network-level management. Parametric no-reference video quality evaluation on adaptive streaming is still at an early stage. Improving resource provision at edge networks must also benefit from programmable networks such as software-defined networking (SDN) to execute flow management, routing, and access control efficiently across equipment from different vendors. There have been some pioneering studies on exploiting SDN for network management in both fixed and mobile networks. Liotou et al. exploited the SDN global resource view and complementary QoE metrics to assure the desired network performance in LTE environment [9]. Bentaleb et al. introduced a SDN architecture to dynamically allocate network resource for user clients based on QoE measurements [1]. Most of the related work only focus on single-objective performance improvement. A holistic approach to model human factors and user-level fairness is absent, whereby the model is integrated as part of a network management eco-system.

II. Project aims and development plan

Our project aims at developing software defined cognitive networking (SDCN) to ensure the user experience, user-level fairness and network efficiency of online adaptive media using SDN-assisted and QoE-aware resource management. To achieve this goal, the project will develop a user-centric resource optimisation model, which will be paired with tailored software-defined networking functions to form a service for QoE-aware in-network monitoring and resource allocation. In particular, there are two strands of work in progress that underpin our objectives, as described below.

A. *Human factor modelling and multi-objective optimisation*

We will develop a cognitive model that estimates the user experience of adaptive media and optimises the allocation of network resource between applications and devices. To achieve its objectives, the model will incorporate multi-faceted human factors related to content characteristics, video codec, media encapsulation, quality adaptation, psychological effects and device capabilities. The work will define reference use scenarios and the supporting systems such as an enterprise network switch or a smart home gateway. The modelling work will be led by analytical studies and initial user tests to establish candidate human factors in adaptive media consumption and their corresponding quantitative metrics. In order to enable the cognitive model in practice, the model will employ metrics that are directly measurable or can be inferred through a network monitoring service.

$$\begin{aligned} Q_{720p} = &- 4.85r^{-0.647} + 1.011 \\ &r \text{ is the video bitrate} \end{aligned} \tag{1}$$

$$\begin{aligned} SI_i(t) = &(\Delta_{VQ})\mathrm{e}^{-0.015(t-t_i)}, \\ &\Delta_{VQ} \text{ is the change of video quality} \\ &t_i \text{ is the time of the quality switch } i \end{aligned} \tag{2}$$

$$\begin{aligned} \Im^{CT} = &\frac{\sum_{i}^{N} r_i}{\sum_{i}^{N} U'_{res_i}(r_i)} \\ &r_I \text{ is the bitrate of video stream } i \\ &U'_{res_i} \text{is the video quality utility function} \end{aligned} \tag{3}$$

Through analytical research and empirical studies, we have established a number of impact metrics related to fairness, video quality, switching impact and cost (Equation 1,2,3) [10]. However, not all supporting research behind these metrics has been conducted in the context of emerging online video services and programmable network infrastructure. We plan to take a holistic view of the problem space and employ subjective user studies to 1) verify the statistical significance of the metrics in capturing the user experience, and 2) derive a large user response dataset for modelling human perception of distortions in different forms and severities. The experiments will reflect real-world scenarios where users consume a broad range of online media content on various types of user devices over heterogeneous networks. Through the test plan, we will produce a large set of test videos (to be open to the community) that covers a spectrum of crucial configurations such as session duration, viewing condition, native/screen resolution, frame-rate, genre, compression loss, stalling, and quality fluctuations in different frequencies, directions and scales. The experiments will capture test participants' immediate and cumulative experience.

Based on analytical research and empirical data, a model will then be developed and validated using various statistical analysis and machine learning tools. Specifically, user perception will be modelled using a "sliding window" of perceptual and psychological impact in the context of online media consumption as illustrated in Figure 1. Metrics such as relative standard deviation between the cumulative QoE from different applications will be defined to quantify user-level fairness. We will also measure network efficiency based on the aggregated user experience and the consumption of network resources.

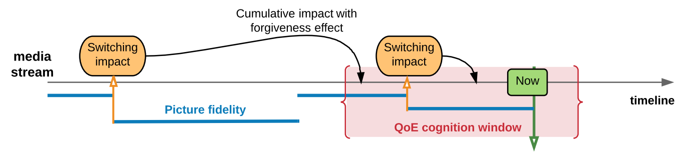

Fig. 1. Cognitive model

The work also includes a multi-objective optimisation function, which assembles measurements of QoE and network dynamics and seeks optimal solutions to provision network resources between applications for an improved balance between QoE, fairness and resource efficiency. The function must support resource allocation in a multi-household and multi-device topology (such as the one shown in Figure 2) and be able to scale up for larger networks. The function will also estimate the impact of any potential solution for resource re-allocation, hence unnecessary and detrimental adjustments will be avoided. We will also investigate the run-time complexity of the cognitive model, which is affected by the number of media applications and adaptive media representations, to facilitate real-time resource management of multiple network segments. Specifically for the algorithm that searches for optimal solutions, full-search will be used as the baseline for benchmarking. Reference supporting systems will also be defined based on different use scenarios. For a home environment, we envisage an advanced home gateway with processing power akin to Raspberry Pi 3 that supports P4 development. Other scenario will see the usage of enterprise-grade SDN switches such as HPE Aruba 3810 series.

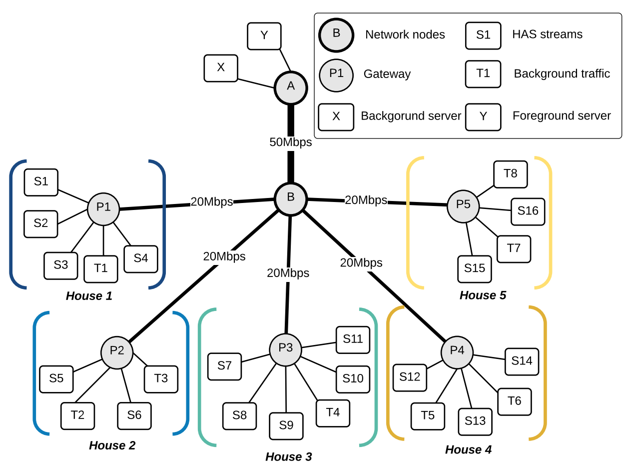

Fig. 2. Multi-household and multi-device topology

We carried out functional evaluations to study how different

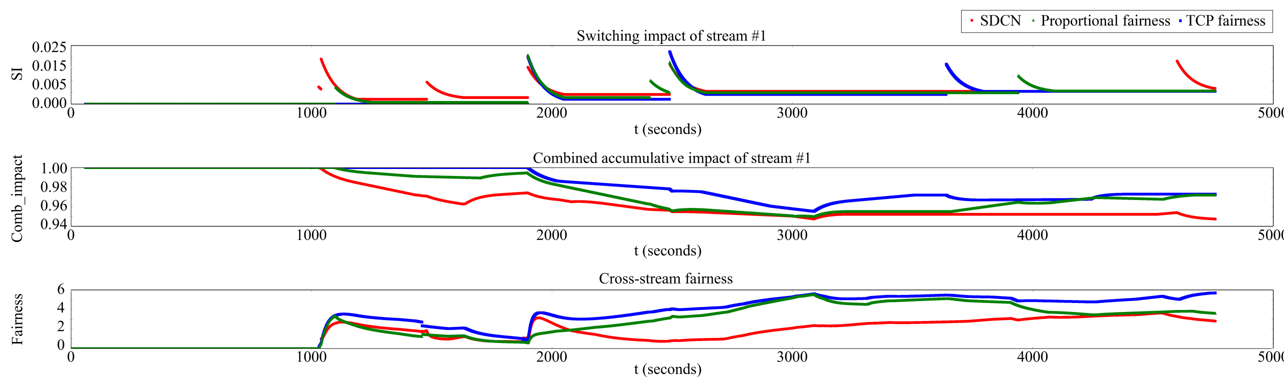

Fig. 3. Functional evaluation

QoE and fairness metrics capture the user experience over adaptive media streams. Preliminary results show that the metrics and models employed by SDCN can improve user-level fairness while reducing the impact of video quality degradation (Figure 3). In particular, our design is able to improve fairness in the face of bandwidth fluctuations (a lower fairness measure depicts better fairness), compared with TCP and Proportional fairness models. Although the analytical studies and simulations can produce very positive results, the actual model performance must be studied through comprehensive experiments that involve close-to-practice networking testbed and human participants. This is assisted by a dedicated experimentation environment.

B. Model deployment using SDN and NFV

The cognitive model will be integrated and evaluated using an experimentation environment. We aim to streamline the validation of essential functions and designs using a network and service environment that is close to practice, rather than starting from simulation. Figure 4 depicts the architecture of our SDN/NFV (Network Function Virtualisation) research and experimentation testbed [4]. The testbed encompasses OpenFlow-capable network switches and an OpenStack-based private cloud environment to instantiate and control a large number of user devices and networks.

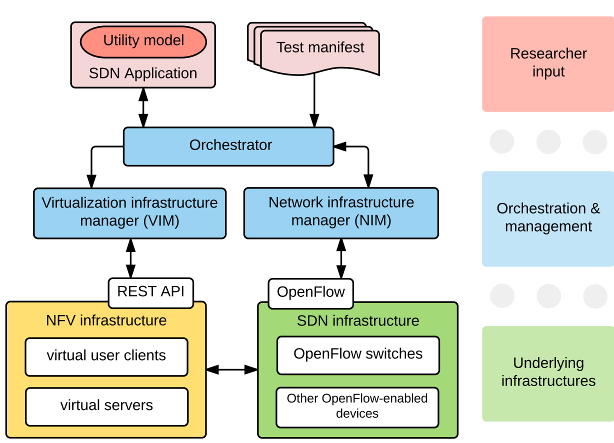

Fig. 4. Architecture of our SDN/NFV research and experimentation testbed

To provide a harness capable of supporting rapid deployment and orchestration of experiments, our experimentation platform fulfils the following requirements:

- Experiments close to practice and at scale. The system should be able to realise and manage a large number of clients and networks.
- Dynamic manipulation of the network. Rate limiting, queuing, flow redirection, and other features of SDN implementation are required to enforce decisions made by intelligent network traffic management modules.
- Configurable clients. The client's configuration should be quickly changeable after an experiment to set up for a new experiment as well as at run time.
- Rapid repeatability of experiments in a clean environment.

Our testbed consists of a three-layer architecture (Figure 4): the top layer contains components provided by the researcher including the test manifest and application/user-level functions such as our case studies: QoE and security applications. The middle layer contains the orchestrator which interfaces with, and includes, the infrastructure managers. The bottom layer contains the network and virtualisation infrastructure where the experiments are deployed.

We will focus on interoperability between the cognitive model and network-level OpenFlow monitoring and control functions. Our testbed is also equipped with a cross-layer orchestration function, which coordinates QoS metrics and corresponding OpenFlow features. We will use this function as the blueprint to develop an *QoE orchestrator* which will serve as a monitoring and control service that 1) assembles network and application statistics as the input metrics of the cognitive model, and 2) deconstructs user-level QoE requirements and resource allocation arrangements to flow, port, or device-level rules to be installed by OpenFlow-controllers. The orchestrator will be designed using virtualised network functions, allowing advanced network management to be instantiated or decommissioned when necessary. The orchestrator includes two subcomponents, the Virtual Infrastructure Manager (VIM) and Network Infrastructure Manager (NIM). VIM controls the virtualization infrastructure through a RESTful API, it launches and configures experiment nodes with information from the test manifest. NIM controls the network infrastructure and consists of a Ryu OpenFlow controller containing a metering and monitoring application. It installs meter flow mods on request from the SDN application and provides information from the network including current throughput of flows and switches. The orchestrator to define and configure network setup on-the-fly through a simple JSON formatted request. A typical request would be to report the current network traffic level for a port or previously defined flow. An example command would be to define a flow (e.g. source/destination IP pair), and request that the flow is limited to a certain level (defined in Mbps).

Experiments will be conducted to contour the overall performance of cognitive networking designs using a range of configurations that reflect real-world use scenarios. The work will use live media applications and fully functional network environment to evaluate the entire SDCN eco-system as a whole. To this end, open-source toolsets of our existing testbed will be used to configure bandwidth fluctuations, the quantity and diversity of user devices, media session flash-crowd, and typical adaptive streaming configurations. A logging function will track time-coded network, system, and application-level statistics for visualisation and further data analysis. Human participants will also be involved as part of the experiments to provide user opinion ratings as the "ground truth" references. The cognitive model will be benchmarked against relevant resource allocation models using tools such as bucket testing. We will also evaluate the scalability and interoperability of our work in different environments and over equipment from various vendors by connecting our testbed to other network testing facilities in academia and industry.

III. Conclusions

SDN is a rapidly evolving landscape. IDC estimates SDN market to experience strong growth over the next several years and reach the value of \$12.5 billion in 2020. Meanwhile, Internet video streaming and downloads will grow to more than 80% of all consumer Internet traffic by 2020 according to Cisco's estimates. An intelligent networking architecture with a deep understanding of media applications and human perception has a pivotal role in optimising content distribution for better cost efficiency, higher penetration of digital media, and resilience against flash-crowd and attacks. The development of SDCN will contribute to the growth of media and Internet sector, and cherish new markets in the area of IoT and creative media while maximising the value of any investment in networking infrastructure. With the rise of SDN, fog computing and network function virtualisation, the capability of cloud computing is being brought closer to the end user in the form of micro data centres or cloudlets. This creates computing capabilities at edge and home networks for contextual network management services, supporting the distribution of cognitive networking functions.

Our work in progress on software defined cognitive networking will capitalise on our recent work on QoE fairness and OpenFlow-assisted network management while taking on new challenges in the understanding and modelling of human factors in adaptive media experiences and its cross-layer integration with software-defined networking. The work will fill the gap of QoE-aware resource management for large-scale concurrent adaptive media. The design and prototyping of the described work will be open and modular, allowing the resultant applications to be purposed for different objectives such as network resilience and energy-aware content delivery.

IV. Acknowledgement

This work is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/P033202/1 (Software Defined Cognitive Networking: Intelligent Resource Provisioning For Future Networks).

References

[1] A. Bentaleb, A. C. Begen, and R. Zimmermann. SDNDASH: Improving QoE of HTTP Adaptive Streaming Using Software Defined Networking. In *Proceedings of the 2016 ACM on Multimedia Conference*, pages 1296–1305. ACM, 2016.

[2] G. Cermak, M. Pinson, and S. Wolf. The relationship among video quality, screen resolution, and bit rate. *IEEE Transactions on Broadcasting*, 57(2):258–262, 2011.

[3] ChildWise. Childwise Monitor Report 2016. *Research and Markets*, 2016.

[4] L. Fawcett and M. Mu. REF: Enabling Rapid Experimentation of Contextual Network Management using Software Defined Networking. *under revision for IEEE Communications Magazine*, 2017.

[5] T.-Y. Huang, R. Johari, N. McKeown, M. Trunnell, and M. Watson. A buffer-based approach to rate adaptation: Evidence from a large video streaming service. In *Proceedings of the 2014 ACM conference on SIGCOMM*, pages 187–198. ACM, 2014.

[6] F. P. Kelly, A. K. Maulloo, and D. K. Tan. Rate control for communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society, p.237–252*, 1998.

[7] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara. Toward A Practical Perceptual Video Quality Metric. *Netflix Techblog: http://techblog.netflix.com/2016/06/toward-practical-perceptual-video.html*, 2016.

[8] Z. Li, X. Zhu, J. Gahm, R. Pan, H. Hu, A. Begen, and D. Oran. Probe and Adapt: Rate Adaptation for HTTP Video Streaming At Scale. *IEEE Journal on Selected Areas in Communications*, 32(4):719–733, 2014.

[9] E. Liotou, G. Tseliou, K. Samdanis, D. Tsolkas, F. Adelantado, and C. Verikoukis. An SDN QoE-Service for dynamically enhancing the performance of OTT applications. In *International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE, 2015.

[10] M. Mu, M. Broadbent, N. Hart, A. Farshad, N. Race, D. Hutchison, and Q. Ni. A Scalable User Fairness Model for Adaptive Video Streaming over SDN-Assisted Future Networks. *IEEE Journal on Selected Areas in Communications*, 34(8), Aug 2016.

[11] M. Mu, S. Simpson, A. Farshad, and Q. Ni. User-level fairness delivered: network resource allocation for adaptive video streaming. *Proceedings of 2015 IEEE/ACM International Symposium on Quality of Service (IWQoS)*, 2015.

[12] Y. Sani, M. Mu, A. Mauthe, and C. Edwards. A Bio-inspired HTTP-based adaptive streaming player. In *IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2016.

[13] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hobfeld, and P. Tran-Gia. A survey on quality of experience of HTTP adaptive streaming. *IEEE Communications Surveys & Tutorials*, 17(1):469–492, 2015.