

The Service Aware Network: Balancing cost and quality for Social Multimedia Communication

Ian Kegel, Doug Williams and Tim Stevens

BT Research & Innovation, Adastral Park,
Martlesham Heath, Ipswich, UK

Simon Gunkel, Pablo Cesar and Jack Jansen

CWI: Centrum Wiskunde & Informatica
Amsterdam, Netherlands

Abstract— This article explores the notion that a ‘service-aware’ network will help in the cost-effective delivery of social communication between communities, when it is enriched by high quality video and audio. While the concept of dynamically managing network components to balance cost and quality of service is not at all new, the paper explains how future plausible use cases for social multimedia communication prompt four key requirements for a new type of service-aware network. A brief summary is then provided of current research into some of the new capabilities needed to deliver these requirements: Quality of Experience Modelling, Dynamic Network Configuration, and Composition in the Network. Finally, an overview is given of a programme of experiments and trials which are being carried out to demonstrate the applicability and scalability of the service-aware network to real services based on the aforementioned use cases.

Keywords — video conferencing, social communication, network optimisation, quality of experience

I. INTRODUCTION

Social multimedia communication services that include computer-mediated interaction, social networking and multimedia content can be very rich experiences mirroring real life, but they have the potential to be both expensive (which is a problem for infrastructure providers) and complicated (which is a problem for users).

Previously [14] we have described how we believe that the way such experiences are presented to the users, on the screen and through loudspeakers, could be improved by intelligently analysing the communication in order to make the way it is represented ‘appropriate’ to the instantaneous needs of the participants. We have called the analysis ‘Communication Modelling’, the process that decides how to represent the interactions ‘Orchestration’ [17], and the process of actually representing it ‘Composition’ [9].

While Communication Modelling, Orchestration and Composition provide the potential to create social multimedia experiences under laboratory conditions, the complexity of these use cases makes them especially vulnerable to the varying capabilities of public networks when in general use.

Popular consumer video communication services today use packet loss recovery techniques, such as Forward Error Correction and Retransmission, to achieve a stable high quality experience wherever possible over public networks [21]. But such services assume use cases addressing small groups and do

not attempt to use higher-level modelling of a communication situation to influence low-level decisions about video transmission. Multi-layer video transmission and layered video codecs such as H.264 SVC [19] are also gaining popularity within video communication services, and algorithms have been proposed for the adaptive selection of video layers based on user preference [6].

The service-aware network is a potential solution to the challenge of supporting complex use cases for multimedia communication between large ad hoc groups, in which communication traffic across the network must be managed dynamically in order to maintain Quality of Experience while minimising cost. Making these ambitious forms of communication cheaper to deliver, and better experiences for the consumer, will hasten their transition from novelty to popular services.

This article describes a number of use cases for social multimedia communication in order to highlight the concomitant requirements on a service-aware network. We describe three key capabilities that will help a service-aware network to meet these requirements. Finally, we review a programme of experiments and trials which are being carried out to demonstrate the applicability and scalability of the service-aware network to real services based on these use cases.

II. USE CASES

This section describes use cases of social multimedia communication which help us discern requirements for a service-aware network.

In each case we introduce the idea with reference to an existing analogous behaviour that we anticipate could be developed to take on more of the attributes we associate with socially-aware multimedia. In all use cases the roles the participants take in any interactions are defined by soft social attributes and not by the technology set-up.

1) Family catch-up

It is traditional in many cultures for families to spend time together during key national festivals, such as Christmas. When families cannot be together they often substitute physical presence with a telephone call or even a video chat. The use case we suggest involves a shared computer-mediated ‘space’ being available and through which individuals can communicate. This sounds abstract, but in practice could be a

multi-party video conference on a large screen, connecting the living rooms of different families. We imagine this could be something that happens on special occasions initially, instead of the Christmas telephone call, but that could develop into a much more normal way for people to interact. It might even evolve into the development of a permanent shared space through which family members would be aware of each other's presence and activity. Through it they could see and talk to each other and share things, both cerebral (thoughts), emotional (feelings) and digital (photos and videos) as well as including people in their day-to-day experiences (How was your day? Do you like my new shoes?)

To an extent, this behaviour already takes place. But the degree to which it exhibits the full gamut of the characteristics of what we describe as socially-aware multimedia is limited. The telephone call does not allow people to see each other (nor their new shoes) and is normally a point-to-point connection and not a group experience. Group based video chat services are usually viewed through personal devices (tablets, mobile phones and PCs) and typically involve individuals joining a conference rather than a multi-way link between a number of spaces (living rooms) that would allow groups (family members) to be together. We note that the emergence of Skype® on Smart TVs and TV video chat devices like Biscotti® may change this.

In 2010, 9.4% of the EU population was not born in the country in which they now reside [18], it seems that connecting families that live a long way away from each other provides a significant opportunity and challenge for service providers.

2) Homework club for Massive Open Online Courses

Students work together on homework. They may work together in class, in libraries, in each other's homes or they may call each other when they get stuck on questions (or when they don't remember the homework that was set). Sometimes they are asked to work together in teams to complete joint projects. They may also discuss homework on social networking sites and even be provided social spaces on learning platforms [11] that enable interaction.

We assert that interactions could evolve from being text-based to including group video chat and suggest that extensions to social networking services such as ooVoo or Skype on Facebook could become attractive and helpful adjuncts to other forms of interaction.

In parallel with the way social interactions are developing amongst students, there is also a growing number of Massive Open Online Courses (MOOCs) in which the student community may be national or global [7]. Should the MOOC student body wish to gain the help of their peers they will often resort to online discussion. Social networking sites and learning platforms are usually the basis for such interactions but they could evolve into shared spaces which people inhabit via video links.

3) Gaming space

Poker Stars is a popular website for playing poker online. The site remains open 24 hours a day and attracts, at its peak 30,000-50,000¹ simultaneous users.

Playing poker in the flesh allows the player to see and "read" their opponents in ways not afforded by the online counterpart. Since both in-the-flesh and online alternatives have their proponents it seems likely that a form of online poker in which you could also see your opponents could also be made into an experience that could attract a significant audience. Whether or not poker is the best game for such a scenario is also moot, but the idea of playing a competitive game in which you can see and hear your opponents is plausible. This is the Gaming space use case.

III. THE SERVICE-AWARE NETWORK

In this section we show how the use cases described above provide some of the key requirements for a service-aware network. We then explain in more detail how our current research is addressing these requirements through specific investigations.

A. Requirements

The use cases described above all share the potential for a globally-distributed audience, most of whom are likely to rely on consumer-grade broadband access networks. In spite of significant investment national access-network infrastructure projects [3] in many parts of the world, customers are most likely to connect to a range of different asymmetric access technologies (such as ADSL2+, VDSL or Fibre-To-The-Premises), some of which will be very restricted in their ability to carry video upstream. The service-aware network must therefore be capable of **adapting social multimedia communication to suit heterogeneous access networks**.

Use cases such as Massive Open Online Courses and Gaming are also likely to draw a continuous global community which convenes at particular times of day in particular locations. The resulting network demand is likely to move across the planet as time progresses. While many conferencing services today employ utility 'cloud' hosting to reduce their costs [2], this behaviour suggests a requirement which goes beyond traditional concepts of scalability. The service-aware network must be capable of **dynamically scaling resources in multiple physical locations** such that they can follow demand within individual communication sessions.

The commercial viability of all the use cases described above will be influenced by the cost of providing a multimedia communication system which is more technically complex than consumer offerings today. It is therefore essential to minimise the additional cost of providing a service. The service-aware network must therefore be capable of **configuring resources for optimal associated cost** – and reconfiguring them within a session if a significant impact on cost is anticipated (eg. a cheaper route becomes available).

¹ www.pokerscout.com (viewed between June and July 2013) reported statistics for the 24hr peak number for Poker Stars and other on line poker sites.

A final dimension that is common to all the use cases is the expectation of a certain Quality of Experience (QoE). Standards-based subjective testing methods are widely used to evaluate the quality of conventional communication systems, and low-level network measurements such as delay, packet loss and jitter are sometimes used to infer Quality of Experience. While these are all of value, the use cases describe scenarios in which participants frequently join and leave and media resources frequently change. This suggests that the service-aware network must have a **comprehensive capability for introspection**, i.e. the ability to collate detailed real-time measurements and determine how effectively the technology is meeting the needs of the participants.

B. Modelling Quality of Experience

A significant challenge for the introspection capability described above is the need to determine Quality of Experience (QoE) in real time, without affecting the participants. One possible solution for this is to build a predictive model that allows us to estimate the QoE for a particular participant at a particular time [10]. To create such a model for social multimedia communication we must consider several factors that influence QoE.

Those factors need to go beyond the classical perspective of quality of service (QoS) parameters, because network parameters alone are insufficient to estimate the QoE [4]. Therefore we want to include the user, context and network into our QoE analysis [12]. This perspective on QoE derives from two popular QoE models developed by Geerts et al. [8] and by Wu et al. [20]. Further, to study the impact of those three factors towards the actual QoE of users we developed a testbed that allows extensive user trials in a fully controllable environment. It includes a platform for multi-party video conferencing, which allows measuring and altering of network and client properties [12].

With this testbed we are carrying out extensive subjective user trials to determine the balance between QoE factors, and to introduce new factors (such as conversation properties) which could potentially provide a better indication of the user's experience. One property we primarily look at is the voice activity of users, in terms of "on-off" patterns. Research in speech patterns [13] has defined a number of events which are particularly interesting to investigate. This is especially the occurrence of simultaneous speech and turn-taking. Particularly whether people start to speak at nearly the same time or if a speaker change occurred after simultaneous talk. In this way, we try to understand if certain factors influence the speech pattern in a negative way.

As a result, our goal is a model that is able to predict the QoE for a participant given their own feedback, client settings and network properties. Factors that we can fully control in our testbed and that we are currently investigating include, but are not limited to: Resolution, Frame-rate, Distortion, Delay, Jitter, Inter-stream (Audio/Video) Synchronization, Inter-participant Synchronization. Investigating those factors will result in a predictive QoE model that we can map into a transition cost table. This table combines the measureable characteristics of the network (e.g., bitrate, jitter, and delay) with the changeable

TABLE I. EXAMPLE STREAM TRANSITION COST TABLE (FOR NICK)

Activity	User	Image Size	Bitrate (Cost)	QoE Impact
Discussion	Alice	Small	LOW	MED
		Large	HIGH	MED
	Bob (speaking)	Small	LOW	MED
		Large	HIGH	HIGH
Presentation	Alice	Small	LOW	LOW
		Large	HIGH	LOW
	Bob (speaking)	Small	LOW	MED
		Large	HIGH	HIGH

parameters of the client (e.g., resolution, frame rate, and coding settings) to estimate the QoE in order to find the "best" video stream for any participant. Further the table must include the participants' context or activity which is likely to change over time.

Table I shows a very limited perspective of such a table, as the full table would be too complex to visualize here. In our example, three students (Alice, Bob and Nick) are discussing their homework together, and Table I shows Nick's simplified QoE model. It illustrates the impact of image quality and bitrate towards Nick's QoE. Bob is currently speaking; therefore showing him in a higher quality will have a bigger QoE impact than showing Alice in a high quality. Whereas, depending on the activity they engage in, it would be sufficient to show Alice either as a small image or to not show her at all, i.e. if Bob is giving a presentation. This knowledge allows us to adjust the system to scarce resources or network congestion.

In conclusion, such a transition cost table allows us to find the "best" combination of the participant's visual representations, at any given moment. Thus the Service-Aware Network can dynamically adapt the user experience and use of resources to maintain the balance between Quality of Experience and the cost of providing the service. As explained, Table I only shows a simplified example and a practical transition cost table may include a complex transition matrix with all influencing QoE factors and might vary depending on the context.

C. Dynamic Network Reconfiguration

The use cases imply multiple high-bitrate video streams with demanding low-latency and reliability requirements. As explained above, these requirements may not be constant throughout a session: participants may join and leave, meaning that the optimum network topology could change with respect to time and geography. In a simplistic solution, the network could be dimensioned to serve the most demanding situations; however it would be under-utilised most of the time. We are developing experimental components for the service-aware network which support dynamic reconfiguration in response to changing demand without interrupting communication sessions. We use the term Video Router (VR) to refer to the routing and control logic which is placed at key interconnects in the underlying network.

Consider the gaming example. The number of players may be roughly constant over a 24-hour period, but the clusters will 'follow the sun' (or moon), as people wake and sleep. Over

time, the network will move from a situation where equal numbers may be located on each side of an ocean, to one where many more are on one side than the other. Costs and delays induced by the trans-oceanic backbone imply that where possible, redundant streams should not be sent over that link. The service-aware network would identify situations where the same stream is required by multiple clients, and instruct the Video Router to only send a single copy of the stream across the link. The source-side router would do this by terminating the redundant streams and transmitting a single copy of the stream, and the destination router would regenerate the additional copies. Both would inspect and update the signalling and control protocols (such as SIP and RTCP) so that the video applications themselves were unaware of the modification.

We are investigating how this reduction/optimisation architecture may have applications at the gateways to VPNs, and for creating efficient trunks within emerging carrier technology such as IP eXchange (IPX) [15], since network providers are now beginning to extend from using IPX for voice-interconnect to video-interconnect.

D. Composition in the Network

The use cases suggest that participants may, for at least some time, be remote spectators, not being fully involved in the conversation. This provides a significant opportunity for the service-aware network to optimise cost and adapt to heterogeneous access networks. Even though the idea of applying additional media processing into the network is not new [1] [16], its benefits towards social multimedia communication are not fully clear. Thus, we are currently investigating how additional media processing components could be effectively deployed 'in the cloud'. We are especially interested in how to start such component on demand so that video composition can move dynamically between the client-side and the network.

Such a component, much like a Multipoint Conferencing Unit (MCU) within a commercial telepresence system, would have a relatively high cost in terms of processing and bandwidth connectivity when compared with a Video Router. However, by combining multiple video streams into a single composition which can be distributed to multiple 'spectating' participants, it could significantly reduce load on the client device and downstream access network. It could be used to reduce the impact of access congestion on Quality of Experience, but also importantly could enable thin clients on mobile devices to participate with acceptable quality via wireless broadband networks.

Moving composition functionality into the network and the associated additional video encode/decode cycle does have significant implications on end-to-end delay, and this is a key parameter in our ongoing investigations. Thus, our main goal of this investigation is to identify under which circumstances a pure client-side processing or a cloud-based approach is more beneficial.

IV. CONCLUSION AND FUTURE WORK

In this article we have introduced the service-aware network as an enabler for social multimedia communication

services based on real-world use cases where customer experience must be balanced against the cost of implementation and operation. Our practice-based research is focused on a component-based technology platform whose capabilities are tested, evaluated and iteratively improved through experiments and trials.

Our "Homework Club" use case is being developed in close collaboration with SAPO Campus, a social media product for the education sector developed by Portugal Telecom's Internet business, SAPO. We are using the results of extensive user studies and lab experiments with students and teachers to implement a public trial integrated with the SAPO Campus product in Summer 2014.

In addition to user evaluations and trials, we are investigating how a service-aware network could operate at scale by developing simulations based on credible demand and cost models. This should identify the potential benefits (such as cost savings) which could be attributed to the capabilities described earlier in this article.

In a world which has already embraced computer-mediated interaction, social networking and multimedia content, we believe that the service-aware network can reduce the barriers to the wider commercial adoption of rich, socially-aware multimedia experiences.

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. ICT-2011-287760.

REFERENCES

- [1] Amir, E., McCanne, S., Zhang, H., "An application level video gateway", in *Proceedings of the third ACM international conference on Multimedia (MULTIMEDIA '95)*. ACM, New York, NY, USA, 255-265.
- [2] BlueJeans/Frost & Sullivan, "Overcoming the Challenges of Pervasive Video: Moving Video Collaboration to the Cloud", <http://pages.bluejeans.com/frost-sullivan-moving-video-collaboration-to-the-cloud.html> (accessed 10th July 2013)
- [3] Broadband Commission, "National Broadband Policies", *BroadbandCommission.org*, 2012. http://www.broadbandcommission.org/Documents/NationalBBPolicies_2012.pdf (accessed 7th February 2013)
- [4] Bly, S.A., Harrison, S.R., Irwin, S., "Media spaces: bringing people together in a video, audio, and computing environment". *Commun. ACM*, 36, 1 (Jan. 1993), 28-46.
- [5] Cano, M.-D., Cerdan, F., "Subjective QoE analysis of VoIP applications in a wireless campus environment", *Telecommunication Systems*, 49, 1 (Jan. 2012), pp5-15
- [6] Chen, M., Su, G., Wu, M., "Dynamic Resource Allocation for Robust Distributed Multi-Point Video Conferencing", *IEEE Transactions on Multimedia*, Vol. 10, No. 5, August 2008
- [7] Daniel, John. "Making Sense of MOOCs: Musing in a Maze of Myth, Paradox and Possibility", *Journal of Interactive Media in Education*, 2012.
- [8] Geerts, D., De Moor, K., Ketyko, I., Jacobs, A., Van den Bergh, J., Joseph, W., Martens, L., De Marez, L., "Linking an integrated framework with appropriate methods for measuring QoE", *QoMEX'10*, 2010, pp. 158-163.
- [9] Kegel, I., Cesar, P., Jansen, J., Bulterman, D.C.A., Stevens, T., Kort, J. and Färber, N., "Enabling Togetherness in High-Quality Domestic

Video Conferencing”, *Proceedings of the 20th ACM International Conference on Multimedia*. 2012, pp159–168

- [10] Khan, Shoaib, Duhovnikov, S., Steinbach, E., Kellerer, W., “MOS-Based Multiuser Multiapplication Cross-Layer Optimisation for Mobile Multimedia Communication”, *Advances in Multimedia 2007*, pp1–11.
- [11] Pedro, L., Santos, C., Almeida S. and Kock-Grunberg T. “Building a Shared Personal Learning Environment with SAPO Campus”, *PLE Conference*. 2012. pp3-8.
- [12] Schmitt, M., Gunkel, S., Cesar, P., Hughes, P., “A QoE Testbed for Socially-Aware Video-Mediated Group Communication”, *Proceedings of ACM Workshop on Socially-Aware Multimedia*, 2013
- [13] Sellen, A.J. “Remote conversations: the effects of mediating talk with technology” *Hum.-Comput. Interact.* 10, 4 (Dec. 1995), 401–444.
- [14] Stevens, T., Kegel, I., Williams, D., Cesar, P., Kaiser, R., Färber, N., Torres, P., Stenton, P., Ursu, M., Falelakis, M., “Video Communication for Networked Communities: Challenges and Opportunities”, *Proceedings of the 16th International Conference on Intelligence in Next Generation Networks*. 2012.
- [15] Sybase 365 Mobile Services, “IPX: The Second IP Revolution”, 2012, http://www.gsma.com/membership/wp-content/uploads/2012/03/Sybase365_IPX_wp.pdf (accessed 10th July 2013)
- [16] Tsang Ooi, W., van Renesse, Robbert., “Distributing media transformation over multiple media gateways.” *Proceedings of the ninth ACM international conference on Multimedia (MULTIMEDIA '01)*, ACM, New York, NY, USA, 159-168.
- [17] Ursu, M., Torres, P., Frantzis, M., Zsombori, V., Kaiser, R., “Socialising through orchestrated video communication”, *Proceedings of the 19th ACM International Conference on Multimedia*. 2011, pp1526-1530
- [18] Vasileva, K., “Population and Social Conditions”. *Eurostat - Statistics in Focus*, 2011
- [19] Wainhouse Research, “A Ready Market: Introducing H.264-SVC”, sponsored white paper, 2006.
- [20] W. Wu, A. Arefin, R. Rivas, K. Nahrstedt, R. Sheppard, and Z. Yang, “Quality of experience in distributed interactive multimedia environments: toward a theoretical framework,” *Proc. of ACM MM'09*, New York, NY, USA, 2009, pp. 481–490
- [21] Xu, Y., Yu, C., Li, J. and Liu, Y., “Video Telephony for End-consumers: Measurement Study of Google+, iChat, and Skype”, in *Proc. of ACM IMC'12*, Boston, Massachusetts, USA, 2012.



Ian Kegel is Head of Future Content Research at British Telecommunications plc. Having studied Electrical and Information Sciences at the University of Cambridge, Ian has worked in both the defence and telecommunications industries on projects ranging from radar signal processing to multimedia delivery, and has spent over 10 years leading research projects on digital media production and multimedia communication.



Doug Williams started his career in BT developing optical fibre amplifiers and switches. Since 2001 he has spent his time researching applications and services that may profitably occupy the data carrying capacity these fibres offer. Much of his work recent work has focused on TV related themes and has included projects on interactive narrative media; on using the TV to support group based games; on improving communication between groups and on calculating the aggregate bandwidth required when such new services are delivered to both consumers and businesses.



Tim Stevens has been with BT since the late 1980s. Recently, he has developed distributed software for real-time video capture and switching as part of the EU FP7 TA2 and VConect projects, and is currently researching media formats and architectures for unicast and multicast content distribution. Earlier work included writing software for both embedded and interactive systems. Following this, he moved into research in digital media and metadata, for which he gained several patents. Tim holds a B.Sc. in physics and is a Chartered Engineer & member of the U.K.'s Institution of Engineering and Technology.



Simon Gunkel received his Master of Science in 2012 both from the Technical University Berlin in Germany and KAIST in Daejeon, Republic of Korea. Since 2012, he is working as a PhD Student at the CWI (The National Research Institute for Mathematics and Computer Science in the Netherlands) in the Distributed and Interactive Systems group. His research interests include multimedia communication, video conferencing, and quality of experience.



Pablo Cesar leads the Distributed and Interactive Systems group at CWI (The National Research Institute for Mathematics and Computer Science in the Netherlands). He has (co)-authored over 50 articles about multimedia systems and infrastructures, social media sharing, interactive media, multimedia content modelling, and user interaction. He has given tutorials about multimedia systems in prestigious conferences such as ACM Multimedia, CHI, and the WWW conference.

<http://homepages.cwi.nl/~garcia>



Jack Jansen is a researcher at Centrum Wiskunde & Informatica (CWI), with over 25 years of experience in multimedia and distributed systems. Empowering people to put available technology to a use they themselves envision is his driving principle. This results in activities ranging from languages, such as Python, via web standardization work (SMIL, Rich Web Application Backplane) to implementing systems for accessible and reusable multimedia (Ambulant). In the recent past he was the main programmer with Oratrix, a startup that aimed to bring structured multimedia content to the Web. More recently, he is one of the main architects of the Vconnect project. <http://homepages.cwi.nl/~jack/>