Multicast Service for UltraFlow Access Networks

David Larrabeiti, Leonid Kazovsky*, Manuel Urueña, Ahmad R. Dhaini*, Shuang Yin*,
Jose A. Hernández, Pedro Reviriego, Thomas Shunrong Shen***
Universidad Carlos III de Madrid, Spain
*Stanford University, USA**
*Universidad Antonio de Nebrija***
Avda. Universidad 30, E-28911 Leganés, Madrid, Spain
Tel. (+34) 91 624 9953, Fax: (+34) 91 624 8749, e-mail: dlarra @it.uc3m.es

ABSTRACT
Optical Flow Switching (OFS) is envisaged as an efficient solution for ultra-broadband end-to-end Internet data transfers. In this paper, we investigate the possibility of providing multicast services over a recently proposed UltraFlow access network that offers two types of access service to its end-users at the same time: IP over GPON and OFS. Our focus is set on the viability of multicast in this dual-mode access concept. This paper studies several application scenarios for multicast UltraFlow access and makes a preliminary assessment of practical feasibility of this service.

Keywords: UltraFlow, Optical Flow Switching, Multicast, Optical Access Network, Passive Optical Network

1. INTRODUCTION

Optical Flow Switching (OFS) has been proposed as an alternative to hop-by-hop electronic packet switching for large data transfers, as it can feature very low end-to-end latency in the data transfer phase at ultra-broadband speeds. In OFS dedicated *lightpaths* are scheduled and dynamically allocated along the entire network path between end-systems [1]. OFS requires the Access networks to be connected to the MAN/WAN by means of optical switches instead of electronic packet switches. No buffering occurs between the source and destination as packet switching is replaced by dynamic optical circuit switching. OFS *lightpaths* are scheduled (for an envisioned time period larger than 100 ms) using centralized network edge schedulers coordinated through an electronic control plane [1-2]. Efficient OFS network design and wavelength allocation mechanisms in the Metro/Core network have been investigated [2]. However, the design of an OFS-enabled access network, which bridges the end-user premises with the OFS Metro/Core, has not been explored so far. Furthermore, the possibilities of OFS as enabler of new services in the residential access area have not been studied in depth. We claim that in the next future there will be a proliferation of networked services that will require ultra-low latency [3], and that it is technical and economically viable to deliver such ultra-low latency OFS services if cloud services are physically pushed toward the user in the access-metro area. To show this, we study the UltraFlow Access [4], a novel optical access network architecture that enables the coexistence of OFS and IP services over the same optical distribution network (ODN). To the date, only the point-to-point UltraFlow concept has been prototyped. In this paper we explore the possibilities of multicast UltraFlow in terms of potential applications within the access area and perform a preliminary analysis of viability, having into account that OF connections should be in operation the least time possible.

2. ULTRAFLOW ACCESS TESTBED AND MULTICAST SUPPORT

As illustrated in Fig. 1, The Stanford UltraFlow Access enables IP access using legacy Passive Optical Networks (*e.g.*, GPON, G-EPON), and OFS access via a set of novel Optical Flow Network Units (OFNUs), which serve as Flow users/ servers, and are connected via the same ODN as PON, to a novel Optical Flow Line Terminal (OFLT) located at the central office (CO). The OFLT serves as interface between the Flow access and metro/core networks and it performs dynamic wavelength allocation (DWA) in the Access, and provides network-monitoring functions such as power and wavelength characterization. It also connects to the legacy IP path via the PON's optical line terminal (OLT). We designed and experimented three different OFNU architectures; each of them is equipped with a 10 Gbps tunable transceiver, a Flow network interface card (NIC) and an IP NIC connecting to a PON via an Optical Network Unit (ONU). The main difference between the three OFNUs is in the optical filter design. OFNU1 has a fixed optical filter, and thus it is a colored architecture (*i.e.*, can receive on one specific wavelength only); this is cost-effective, but lacks flexibility in terms of scheduling and resource availability. A colorless OFNU design (*i.e.*, can receive/transmit on any wavelength) is more costly, but it enables efficient resource utilization and scheduling flexibility. Thus, we consider two colorless OFNU architectures: OFNU2 and OFNU3. The architecture OFNU2 has a tunable filter covering the C-band to offer wavelength selection flexibility for downstream transmission. Similarly, OFNU3 employs a circulator to separate upstream and downstream Flow traffic, which also offers full wavelength flexibility for bidirectional transmission. To separate/combine IP and downstream/upstream Flow traffic based on different wavelengths utilized in both services, we have designed and implemented 3-port and 4-port gateways (see Fig. 1).

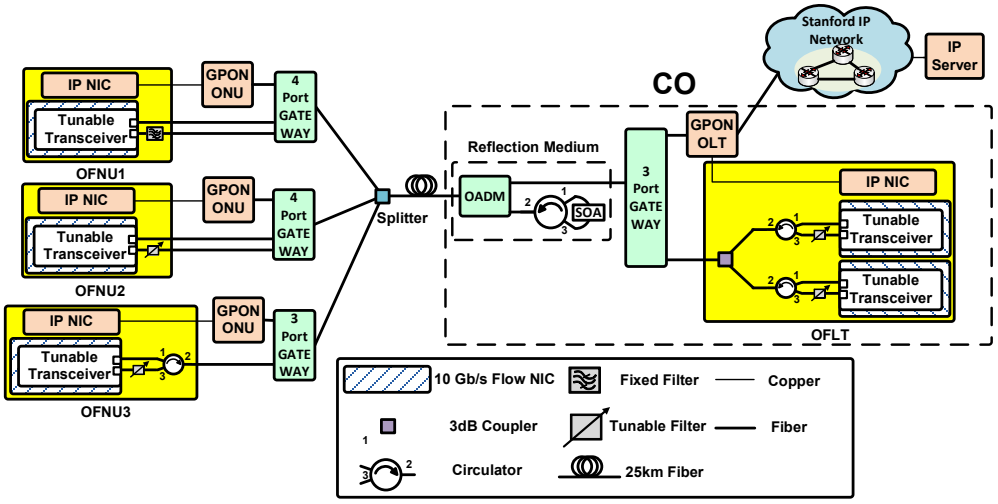

Figure 1. The UltraFlow Access testbed at Stanford University.

As can be followed from Fig. 1, two different types of communications are supported in the UltraFlow Access testbed: OFLT-OFNU and OFNU-OFNU. The OFLT-OFNU communication is the regular network to users communication (*i.e.*, between the service provider and customers' premises); whereas the OFNU-OFNU communication is between two OFNUs located in the same Access. The latter is enabled through a Reflection medium located in the CO, where a specific upstream signal is reflected back into the downstream direction via a circulator. A Semiconductor Optical Amplifier (SOA) is used to compensate for the "double-distance" fiber loss. Multicast can natively be enabled for both types of communications due to the *star* nature of the passive splitter.

In order to get a preliminary experimental assessment of the viability of the approach we conducted a multipoint experiment in the UltraFlow testbed at Stanford University. The Ethernet NIC in the testbed supports 10GBASE-SR, LR and ER, which is defined based on the plugged transceiver. In our case, it is a DWDM 10GBASE-ZR for an 80 km reach (Ethernet), which is not standardised by IEEE but it is vendor-defined. To test the feasibility of supporting multicast, we configured different ONFUs' filters to receive on the same wavelength. Experimental results demonstrated error-free communication for both the primary (bidirectional connected to the multicast sender) and secondary OFNUs (unidirectional connected to the multicast sender). Furthermore, Ethernet frames were successfully handled by the NICs at both OFNUs.

Having successfully experimented the feasibility of multicast transmission in our testbed, we now explore its envisaged application(s) and analyse its theoretical viability.

3. ULTRAFLOW MULTICAST SCENARIOS AND IMPLEMENTATION CHALLENGES

As stated in the introduction, we envision a tendency to push new very high throughput-based services closer to the user in the medium term. The reason is twofold. On one hand, the light propagation speed is an insurmountable factor that limits the perception of instantaneous interaction in a number of new applications (*e.g.*, cloud disk, cloud virtual PC, immersive worlds). On the other hand, sustainable per-user ultra-broadband rates are not feasible in shared packet networks nowadays, and WAN-wide optical circuit switching is technically complex. In this context, the support of multicast in UltraFlow can be useful in several scenarios that are described in the following.

3.1 Multicast UltraFlow scenarios:

Multicast transmission over UltraFlow networks can either be performed globally, that is between two OFNUs located in different PONs connected via the MAN/WAN, or locally (*i.e.*, between two OFNUs in the same PON). The former case requires complex optical devices/nodes in the MAN/WAN, which can split the same optical signal toward two (or more) different PONs (without any buffering). To the best of the authors' knowledge, such devices are not currently available. Therefore, we focus on the latter case where two types of multicast scenarios are envisioned to exist: User-to-User and Network-to-User.

3.1.1 User-to-User:

In this scenario, a user wishes to send multimedia content to more than one user within the same PON, which is enabled in the current Access architecture via the reflection medium installed at the CO (as illustrated in Fig. 1). The following are typical application scenarios for User-to-User multicast transmission.

- *File exchange or HD videoconference between multiple subscribers*.

- *Instant synchronization of server backups*: servers (*e.g.*, databases) located in different geographical areas (each connecting to an OFNU across the same PON), are commonly setup to share the same data for backup purposes. As known, backups are usually scheduled overnight so as not to disrupt daytime operations due to their prohibitive file transfer time.

The expected demand for this type of service is low and so is the probability of a peer being in the same PON. Most services connecting various end-users such as multi-user networked games require a server. Therefore we focus on the Network-to-User scenario. If the User-to-User service is eventually required it can be implemented by combining native UltraFlow multicast support and a multi-point unit running in the service provider's network, which converges the problem into the Network-to-User multicast scenario.

3.1.2 Network to user:

In this scenario, the content is at the CO, and is delivered via the OFLT to multiple OFNUs. This is easily enabled in the architecture via the passive splitter, and by tuning the designated OFNUs' filters to the multicast wavelength. The following are typical application scenarios for Network-to-User multicast transmission.

- *3D Ultra-HD Panoramic IPTV broadcasting:* 360º immersive video in full room walls featuring HD LED displays will be soon in place, requiring lots of bandwidth and low latency for live broadcasts (*e.g.*, sports events). Nowadays, raw transmission of 8K UHD TV requires 24 Gb/s (7680x4320pixels x12bits/pixel at 60 fps)
- *Content Delivery Networks (CDN):* flash transfer of HD video files to many user caches simultaneously. For instance, if 90% of the subscribers watch *premier* of movies, it may be convenient to preload the content locally (at the OFNU) to off-load the main video-on-demand server.
- *Flash update of set-top box software by an operator:* IPTV devices can be as complex as PCs, and their software may consist of several Terabytes, which needs periodic rewrite for security reasons and/or to deploy new services.
- *Flash simultaneous system/network update:* systems (*e.g.*, operating systems, local area network services) may have the same setup in different OFNUs, and thus a remote (directly from the operator) flash and simultaneous update of these systems is desirable.

3.2 Implementation challenges

There are several problems to solve before multicast UltraFlow services can be implemented:

- **Optical 1G, 10G-Ethernet is not designed to work in a broadcast/multicast one-way medium.**
  The OFNUs within a multicast group should be able to identify and forward the multicast frames for upper layer processing. This may be problematic for OFNUs that operate over legacy Layer 2 protocols (*e.g.*, Ethernet), which imposes restrictions on multicast addressing. Depending on the actual implementation of OFNU, different approaches may be taken to overcome this problem. If the OFNU is integrated with the end-user terminal using *off-the-shelf* Flow NICs (*e.g.* 10G-Ethernet in our testbed), a multicast group could be formed through the IP channel and a multicast MAC address is assigned accordingly. Upon receiving a Flow frame, the OFLT and OFNUs will match the destination address with the assigned multicast MAC address. The problem of this approach is that the mapping between the multicast IP and MAC address is not unique. Therefore, the OFLT and OFNUs need to forward the data to Layer 3 for IP header processing to confirm the matching of local and intended multicast address. To mitigate this issue, the OFNU may be designed and built using customized NICs that provide Layer 2 independence and in which multicast IDs can be easily embedded in the frames.
  Since there are no currently available layer 2 standards that can directly serve our purpose, secondary OFNUs can simply snoop what the primary OFNU is receiving.
- **Reliable multicast requires a return channel.**
  Even though UltraFlow is designed to feature a very low BER, and its reliability can be further improved with FEC techniques, reliable file transfer always requires a return channel. In this paper we propose to use the IP network as a return channel for all receivers except for the primary client (the one initiating the UltraFlow connection on the PON).
  A number of scalable reliable multicast protocols were proposed and studied in the literature [5]. However this scenario has specific novel characteristics that deserve further study:
  - The secondary clients can benefit from the retransmissions of packets requested by the primary client on the UltraFlow link (as they tune the same downstream lambda).
  - Clients may send retransmission requests via the IP network and retransmission replies via UltraFlow link, or alternatively use TCP/IP to recover the missing packets, or a hybrid approach: use UDP/IP or TCP/IP by default and, only if the server detects many requests for the same packet (for instance due to a fault in the feeder segment), make use of the UltraFlow link.
  - It is a single sub-network transfer mechanism, not multi-hop IP downstream. The focus is set on minimizing the individual transfer time of the best receiver, not the average group transfer latency.

There is no multicast congestion or flow control mechanism because secondary receivers actually snoop what the primary UltraFlow client is receiving in back-to-back mode.

- We need to minimise the holding time for the UltraFlow connection, as it is a scarce shared resource. Connection time must squeeze the wavelength capacity and devote as little as possible to retransmissions. Therefore we shan't allow for more than 1-2% of overhead caused by retransmissions, as an arbitrary design rule.

4. PRELIMINARY ANALYSIS OF VIABILITY

In this section we try to predict the expected performance of multicast UltraFlow as a function of the quality of the downstream UF channel of the set of receivers, as an initial step to study the whole problem. We shall leave the effect of control packet errors for further study. For the discussion, we use BER_{PON} as an estimation for the probability of a bit received erroneously by any of the receivers of the UltraFlow multicast downstream flow in our PON. Assuming a set of N receivers on the PON with the same BER conditions, the BER_{PON} can be estimated as:

$$BER_{PON} = BER_{feeder} + (1-(1-BER_{branch})^N) - BER_{feeder}(1-(1-BER_{branch})^N) \approx BER_{feeder} + (1-(1-BER_{branch})^N) \quad (1)$$

if we can measure the BER in the OLT-splitter (BER_{feeder}) and splitter-branch (BER_{branch}) segments. Alternatively, we can simply make this term 0 and measure individual OLT-OFNU BERs, assuming full independence of errors among PON branches in both cases:

$$BER_{PON} = 1 - \prod_{i=1}^{N}(1 - BER_{branch}(i)) \quad (2)$$

On the other hand, since we are going to retransmit on a per-packet basis, we need to compute the Packet Error Rate for the group or receivers (PER_{PON}) that is the *rate of packets received with error by at least one of the receivers in the group*. For the sake of simplicity we shall consider the broadcast transmission of bits as independent events, even though more realistic burst-error models exist, under the assumption that this will lead to a more conservative estimation of PER. Assuming a packet size of MTU (bits) for all packets, and no use of FEC techniques the PER_{PON} can be estimated as $PER_{PON} = 1-(1-BER_{PON})^{MTU}$. Sample values of these parameters for a $BER_{branch} = 10^{-12}$, $MTU = 12000$ bits and $N = 128$ receivers, are:
$BER_{PON} = 1.3\times10^{-10}$ and $PER_{PON} = 1.5\times10^{-6}$.

4.1 File Transfer

Now let us analyse a scenario where the return channel is IP and only downstream data errors are considered, in order to assess the impact of the multiplicity of receivers and downstream PER in isolation. The selected application scenario is File Transfer: the content of a 1-Tbyte Blue-ray disc is to be transferred from network to a set of users. This roughly takes about 800 seconds at 10Gb/s with back-to-back packets, ignoring link, network and application layer overheads. We need to estimate how long would actually take the complete transfer due to retransmissions. To this end we shall use a simple yet effective approach to achieve the best performance with the receivers that have higher quality links (see Fig. 2a). Firstly, the whole file is broadcasted. Then a retransmission cycle starts where the receivers selectively notify Negative Acknowledged (NACKed) packet sets to the server and the missing packets are re-broadcast until all receivers have acknowledged the complete file transmission. In order to prevent NACK implosion, during preliminary file transmission receivers may periodically feedback the current set of NACKed packets and the server could periodically broadcast the set of packet IDs that will be re-broadcast at the end of the current file transfer. In addition this technique can save a one-way trip time for many feedback packets. However we shall not use this type of techniques in the analysis in order to have a more conservative estimation of the additional time required for retransmissions. On the other hand, in this model, we shall not include the time required to transmit in parallel all feedback NACKs over the IP return channel.

Since it is not possible to bound the maximum file transfer time, we try to determine the expectation of the effective transfer time of a *filesize*, over an UltraFlow circuit of R b/s, using packets of MTU size, under a given PER_{PON} and an RTT (Round-Trip Time) from server to the furthest apart client. Transmissions take place in cycles: data-feedback-feedforward. Cycle one is the first bulk file transmission, which has an additional one-way delay that we shall omit here, along with the initial file transfer initiation signalling (that could take place via TCP/IP over the IP service). Then, from the instant the first bit arrives at a receiver, each cycle implies a time of: data_size/R+RTT. Thus the expectation of the whole file transfer time T_{tx} can be estimated as:

$$E(T_{tx}) = \sum_{j=1}^{\infty} p(j) \sum_{k=1}^{j} \left(\frac{filesize}{R} PER_{PON}^{k-1} + RTT\right) \quad (3)$$

where $p(j)$ stands for the probability of completing the file transfer in j transmission cycles, and can be expressed as:

$$p(j) = (1 - PER_{PON})^{n.PER_{PON}^{(j-1)}} \prod_{i=0}^{j-2} (1 - (1 - PER_{PON})^{n.PER_{PON}^{i}}) \quad (4)$$

where n is the *filesize* in number of packets units (*filesize/MTU*). Figure 2 provides a visual and quantitative estimation of the effect of retransmissions over the effective multipoint transfer time. In the figure, we show the additional percentage of time required to achieve a full transfer of a 1TB file to N receivers ($N = 64$ to 1024[1]) as a function of the individual OLT-OFNU BER. An *RTT* equal to 10 ms was assumed. The figure shows that the penalty time due to retransmissions is very low up to individual BERs of 10^{-9}, and has an acceptable value for 10^{-8} if the number of receivers is below 128 in our case.

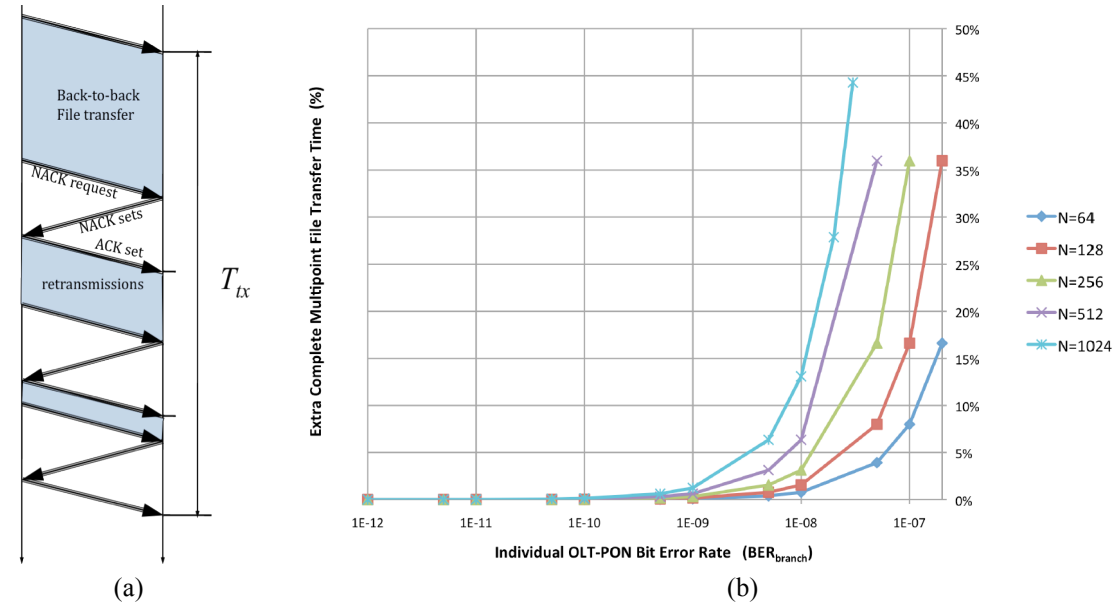

Figure 2: a) Protocol outline; b) Overhead in Multipoint File Transfer time vs. point to point OLT-OFNU BER.

We would also like to study the effect of a non-regular BER in the PON. Figure 3 shows that a single lossy ONU does not have a significant impact on file transfer overhead up to an individual BER of 10^{-6}, irrespective of the number of receivers, within our range of interest (see overheads for $N = 128$ and $N = 1024$). In the example, the rest of ONUs are assumed to work at a BER of 10^{-9}. However, the amount of lossy receivers has a strong influence on the permissible BER for those receivers, getting as low as 10^{-7} for 10 independent lossy receivers, and leaving no extra error margin (10^{-9}) for multicast sessions with 100 low quality links.

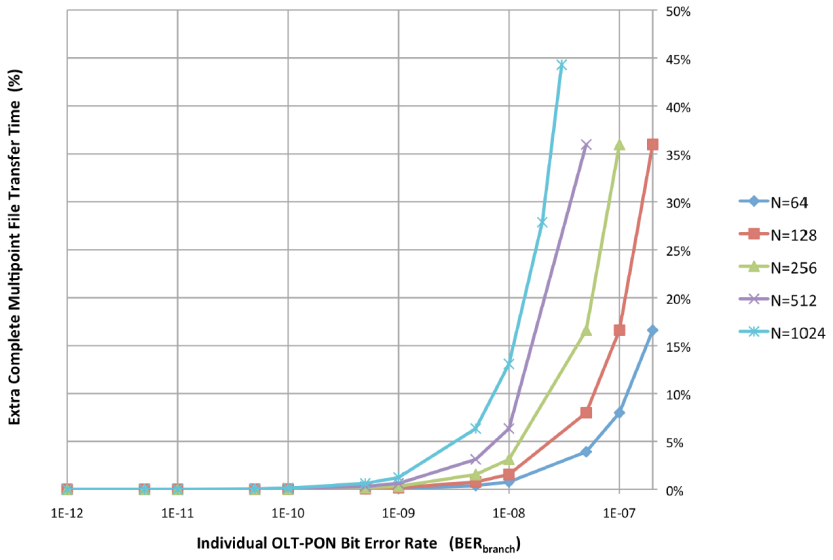

Figure 3. Overhead in Multipoint File Transfer Time with low quality receivers.

[1] Setting OFS in a theoretical PON scenario with a fan-out of 1024. It can also be assumed that 1024 receivers correspond to various smaller PONs merged at the OLFT by a local splitter/combiner.

4.2 Live HD 3D TV broadcasting

The case of future-generation immersive live broadcasting systems is also a niche of application in UltraFlow. In this case, the objective is not full reliability but the improvement of the effective PER_{PON} by selective retransmission of packets within the timing frame of the play-out buffer. In this context, the EPER, defined as Effective Packet Error Rate after in-time retransmissions, both individual and for the group depend on the size of the play-out buffer *B* (usually in the range 20-100ms), on the RTT and on the downstream capacity allocated to retransmissions *R'*. Assuming that a) this latter capacity is much greater than required by the group error rate ($R \cdot PER_{PON} << R'$), b) retransmissions requests are issued as soon as errors are detected and c) *BER*s are homogeneous across receivers, then the expression for the individual *EPER* is simply $EPER = PER^{(num\_rtx+1)}$ where $num\_rtx = \left\lfloor \frac{B}{RTT + 2\frac{MTU}{R}} \right\rfloor$ or $\left\lfloor \frac{B}{RTT + \frac{MTU}{R'}} \right\rfloor$, depending on whether retransmissions take place as soon as the current downstream packet transmission is completed or based on a bandwidth-guaranteeing scheduling algorithm.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we studied future target applications of multicast over UltraFlow networks. We have described how the system can support multicast in the Access, and presented a preliminary assessment of the impact of transmission errors on multicast transmission. The initial results are promising and show that it is viable to provide efficient multipoint file transmission to a large number of receivers given realistic transmission BERs. However, the performance of the technique is subject to the strong influence of error-prone branches. We attempted to quantify the permissible BER for lossy receivers. Very *lossy* receivers should be discarded from the use of this technique, and instead rely on TCP-based point-to-point schemes. The error-analysis presented focused on the downstream medium only, which is the main source of global reliable transfer latency. However, since we propose to use the regular IP network of the Ultraflow architecture as the return channel for secondary receivers, it is necessary to address the effect of errors in the control plane as well. The study of this effect is left for future work.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the support of the Chair of Excellence of Bank of Santander – UC3M, the National Science Foundation, NSERC and the Spanish projects CRAMnet (grant no. TEC2012-38362-C03-01), and MEDIANET.

REFERENCES

[1] V. W. S. Chan, "optical flow switching networks," *Proceedings of the IEEE*, vol. 100, no. 5, pp. 1079-1091, 2012.

[2] Z. Rosberg, J. Li, F. Li and M. Zukerman, “Flow scheduling in optical flow switched (OFS) networks under transient conditions”, *Journal of Lightwave Technology*, vol. 29, no. 21, pp. 3250-3264, 2011.

[3] L. G. Kazovsky, A. R. Dhaini, M. De Leenheer, T. S. Shen, S. Yin, and B. A. Detwiler. “UltraFlow access networks: A dual-mode solution for the access bottleneck”, in *Proc. International Conference on Transparent Optical Networks (ICTON’13)* , Cartagena (Spain). June 2013.

[4] D. Larrabeiti, J. A. Hernandez, I. Seoane, and R. Romeral, “Managing delay in the access”, in *Proc. 17th European Conference on Networks and Optical Communications (NOC)*, pp. 1-8, 2012.

[5] K. Obraczka, “Multicast transport protocols: A survey and taxonomy”, *IEEE Communications Magazine*, vol. 36, no. 1, pp. 94-102, Jan. 1998.