Real-Time Multipath Multimedia Traffic in Cellular Networks for Command and Control Applications

Manlio Bacco*, Pietro Cassará*, Alberto Gotta*, Vincenzo Pellegrini†

*CNR, Institute of Information Science and Technologies, Pisa, Italy
†Ingegneria dei Sistemi (IDS), Pisa, Italy

***Abstract*—This work describes a real testbed for enabling Unmanned Aerial Vehicles (UAVs)-to-ground real-time video streaming. The aim is in providing a video feed to the pilot on the ground for Beyond Visual Line of Sight (BVLoS) operations exploiting cellular networks in urban/suburban areas. Towards this aim, multipath communications are used in a multi-operator setup to counteract the intermittent network coverage in urban and suburban ares. The main requirements are: low latency, and a continuous video stream of reasonable quality. We rely on both upper-layer Forward Error Correction (FEC) techniques and link diversity, so to increase the probability of a fluid video playback with acceptable quality. We report per-link statistics, collected during field trials, of three different cellular operators, to analyse the impact of using a set of physical links as a single logical one on a RTP-based video streaming. In our tests, such a setup has provided a good level of performance.**

***Index Terms*—multipath, UAV, smart cities, command and control, video quality, PSNR, SSIM**

I. Introduction

UAVs are more and more affordable from both cost and easiness of use viewpoints, thus opening to their use in a wide spectrum of applications and services. Being connected flying objects that can carry things, connect to networks or provide connectivity, monitor areas, people, and buildings, UAVs can prove to be very versatile, fast-moving, and in a large range of sizes. At this time, their use in heavily populated areas, as urban ones, is strictly limited, if not forbidden, to avoid any dangerous situations that can occur due to close proximity to buildings, things, people, and also airports. Anyway, it is likely that in the future the regulation will change, also thanks to improved safety mechanisms, and to the introduction of the so-called Unmanned Aircraft System Traffic Management (UTM) framework. UTM will manage the airspace where also UAVs, or even large Flying Ad-Hoc Networks (FANETs), will semi-autonomously fly thanks to BVLoS control. To push towards those scenarios, active investigations are required because of the great challenge in providing a real-time Command and Control (C2) link to UAVs via public 4G/5G networks in real scenarios. At now, UAVs have been proposed to provide connectivity services, by acting as moving cells [1], [2] and providing additional on-demand capacity. Afar from coverage extension, UAVs can also support environmental monitoring [3] and agriculture applications [4], [5], among others. At this time, civil applications require that the human pilot, still needed for security reasons, has Visual Line of Sight (VLoS) with the UAV, thus somewhat limiting the scope of their use because of the feasible operating range.

corresponding author: Manlio Bacco, e-mail: manlio.bacco@isti.cnr.it

In this work, we explore the possibility of using an UAV-to-ground video feed for providing visual context to a remote pilot on the ground in BVLoS conditions. The video feed is supposed complemented by telemetry data coming from the UAV to the Ground Control Station (GCS). This activity is framed within the SCIADRO (Drone Swarms) research project [6] funded by Tuscany region, in Italy. The feasibility of using a 4G link to control an UAV has been subject of studies in the literature [7], [8], in order to characterize the quality of signals received at different flying altitudes. Results show that a 4G link can be effectively used to provide wide-area wireless connectivity to small UAVs [7], even if limitations should be taken into account, like the rapid decrease of the received signal level as altitude increases [8]. In this work, we exploit multipath techniques, i.e., the possibility of *aggregating* multiple physical links into a single logical one, in order to increase the robustness of the downlink video channel. In fact, some precautions are necessary to ensure the continuity of video connectivity, in order to properly support C2 operations in BVLoS conditions via public cellular networks. Further than continuity, the *playout delay*, i.e., the time delay after which a video chunk is played w.r.t. its generation at the source, must be strictly limited to actually provide real-time visual context. Based on those premises, this work reports the performance evaluation of a real testbed in what follows, providing insights in: (*i*) the advantages brought by the opportunistic use of several cellular networks to effectively deliver a video stream of good quality by exploiting multipath techniques; (*ii*) the impact of cross-correlation among the packet arrivals from the cellular networks jointly used, which is one of the main factors impacting the achievable performance.

The rest of this paper is organized as follows: Section II describes related works, focusing on similar scenarios and on the use of multipath techniques for delivering multimedia live streams. Section III describes the overall system setup in use during the testbed; Section IV reports the performance evaluation, considering both network statistics and video quality. Section V draws the conclusions and opens to future works.

II. Related Works

A work closely related to the case of using cellular networks for the C2 operations of an UAV is presented in [9]. The authors test the possibility of using Long-Term Evolution (LTE) networks for piloting an UAV in a rural area in Denmark. The preliminary measurement campaigns, aimed at characterizing the path loss model, are then backed by simulation results confirming that LTE networks can be used for such a purpose. Furthermore, the authors show that Interference Cancellation (IC) techniques can reduce the impact of interference, along with network diversity, to improve both coverage and reliability of the downlink C2 channel. When discussing network diversity, i.e., the joint use of multiple physical links, two seminal works should be taken into account. The first one discusses MultiPath TCP (MP-TCP) [10], allowing the simultaneous use of multiple Network Interface Cards (NICs) by a TCP connection. The second one, focused on multimedia delivery, is presented in [11], proposing MultiPath Real-time Transport Protocol (RTP) (MP-RTP). The latter is a milestone towards the objective of this paper. The main idea is that disjoint paths, between a sender and a receiver, can be used as a single logical one to deliver data. Such an architecture can provide failover capabilities, or can aggregate the overall capacity to increase the achievable Quality of Experience (QoE), as shown in [11] by means of simulation results. The authors in [11] highlight how bandwidth increase should be pursued by carefully choosing the links to be used in order to respect any latency constraints; in fact, heterogeneous networks may exhibit different statistics. The maximum playout delay is set to 500 ms in [11], a fair value for online interactivity.

Multihoming has been studied in the literature in several contexts. In [12], multiple LTE and WiFi interfaces are used to support real-time data streaming services through a middleware, namely GreenBag. The authors compare GreenBag with MP-TCP and MP-RTP, concluding that MP-TCP had still open issues with congestion control procedures at that time, and that MP-RTP does not take into account energy requirements. Also live streaming has been proposed as MP-TCP-based, i.e., using elastic protocols that have not been designed with such a purpose in mind. Typically, real-time multimedia streaming occurs over UDP because retransmissions are not necessary, if not even detrimental in terms of playback delay. MP-TCP-based live streaming is proposed in [13]. The authors propose a Dynamic MPath-streaming (DMP-streaming) scheme, highlighting how homogeneous paths (i.e., links showing comparable network statistics) represent a condition for satisfactory performance. When considering mobility as well as the use of multiple paths, the features provided by MP-TCP are explored in [14]. The authors conclude that MP-TCP can naturally shield the application layer from the multiple handoffs occurring at lower layers in mobility conditions.

At this time, standard approaches such as MP-TCP and MP-RTP should be preferred to custom solutions, because they are transparent to the sender and receiver application layers. Both protocols use feedback mechanisms, thus requiring a return channel: in the case of MP-TCP, the return connection is mandatory for reliability purposes; in the case of MP-RTP, the return connection is optional, but typically employed with the purpose of providing out-of-band information about the experienced Quality of Service (QoS). The RTP Control Protocol (RTCP) carries the periodic reports, and its multipath variant is referred to as MP-RTCP.

Fig. 1: The testbed platform in use: the UAV on the left, the RPi on top of a large capacity battery on the right, with a mounted camera, and three LTE modems.

We studied an alternative approach for WebRTC-based multimedia delivery in [15], focusing on energy efficiency. The scenario under consideration herein does not explicitly consider energy efficiency, even if part of the testbed has been performed on a Raspberry Pi (RPi), providing low power consumption and acceptable video quality [15]. The main focus is on video quality, and real-time data delivery. The use of network diversity is here tested in a real urban/suburban environment, thus we do not provide simulation results as in [9], [11]. Additionally, because of the potential magnitude of the negative consequences if the pilot cannot intervene as soon as necessary, we consider as acceptable a maximum delay of 200 ms, a stricter value than that in [11].

III. System Configuration

In this section, we describe the system configuration in use in our testbed. The same configuration has been deployed on two different hardware setups: the first one is based on the use of a laptop for experimental purposes and for data collecting, the second one relies on a RPi as a lightweight payload for the UAV. The latter is visible in Figure 1, as well as three LTE modems, the RPi equipped with a camera, and two USB WiFi interfaces in addition to the one integrated in the RPi. Each WiFi NIC is connected to a modem, so that three cellular connections have been contemporarily used. The receiver side is the same for both hardware setups: a desktop PC, and the open-source software *QGroundControl* as GCS for both telemetry data and the live video stream. The three LTE modems have three SIMs of Italian operators: Vodafone, Tim, and Wind-Tre.

A. Application Setup

GStreamer is considered to be one of the reference frameworks for video streaming applications: it is an open source multimedia platform, available for the most common operating systems and embedded platforms, like RPi. The use of QGroundControl, as the software part of the GCS, provides

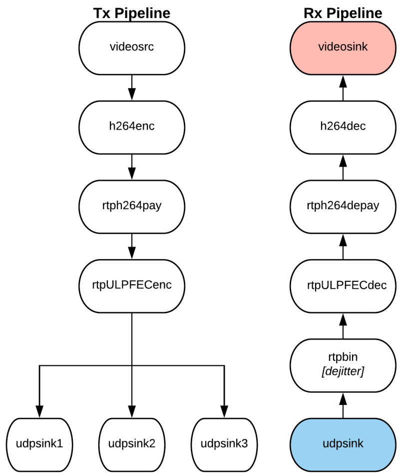

Fig. 2: Logical diagrams of the transmitting (tx) and receiving (rx) Gstreamer-based pipelines in use.

a solution that does not require any modifications on the receiver side. Furthermore, it is portable to desktop PCs, smartphones, and RPis. This means that our development efforts have been concentrated on the transmitter side, i.e. the platform on board of the UAV. A Gstreamer-based application is based on a pipeline of software modules, called *plugins*, implementing the needed functional blocks. Figure 2 shows the transmitting (tx) and the receiving (rx) logical blocks of the pipelines we used. Note the three *udpsink* elements on the left corresponding to the three cellular interfaces in use at the sender. At the receiver side, there is a single *udpsink* element, corresponding to a single high-bandwidth interface; this setup relies on the proper configuration of *iptables* at the sender side, in order to have a 3:1 ratio for NICs coupling. In details, the video stream is captured through the camera, scaled to a resolution of 1024x768 pixels at 10 fps, and then compressed with an hardware-accelerated H.264 encoder. The video stream is encapsulated into RTP packets, then a FEC module is used to increase the robustness of the video flow [16], [17], the latter finally sent to the UDP layer for transmission via the three cellular networks. The Uneven Level Protection FEC (ULPFEC) plugin [18], which is a XOR-based FEC technique, has been implemented as a *repetition code* with rate 1/3 (corresponding to the aforementioned 3:1 ratio), providing a lightweight implementation to be used on constrained devices. We consider as acceptable a maximum playout delay of $PD_{max} = 200$ ms. At the receiver side, a *dejitter* module (in Figure 2 as part of the *rtpbin* plugin) has been used to reorder and remove duplicate RTP packets within PD_{max}. Table I shows the values of the parameters in use.

IV. Performance Evaluation

This section describes the performance evaluation from two viewpoints: in Section IV-A, the network traffic is analysed; then, in Section IV-B, we assess the resulting video quality.

Plugin	Parameters
videosrc	video/x-raw, width=1024, height=768
h264enc	Group of Pictures (GOP) size = 90
rtpULPFEC	pt=101, percentage=50, 100, 200
udpsink	src_port=5004, 5006, 5008; dst_port=5000
dejitter	PD_{max}=200[ms], max-dropout-time=60[s], max-disorder-time=2[s]

TABLE I: System parameters of the GStreamer pipelines.

A preliminary set of tests was conducted in safe and delimited flight area with an UAV. The aim was in verifying the feasibility of the BVLoS operations with the described configuration. The FEC module was added after those tests to counteract link failures, and to increase the robustness of the video streaming. This test was conducted with an RPi on board of the UAV in Figure 1, so to have a lightweight payload. After this successful preliminary test, we scaled up the testbed: three cellular operators were considered instead of a single one. At this time, BVLoS operations are strictly prohibited in urban areas and surroundings in Italy, thus we resorted using a car instead of the UAV. The larger testbed we designed involves both urban and suburban areas, which are shown in Figure 3. The aim is in considering the different density of the Evolved Nodes B (eNBs) in the urban part (see Figure 3a) w.r.t. the suburban part (see Figure 3b). Furthermore, the suburban part comprises uphill location, thus providing cellular connectivity in Line of Sight (LoS) conditions with several distant eNBs, exposing the sender to inter-cell interference and frequent handovers [7]. The RPi sender-side has been substituted with a laptop for the purpose of collecting richer statistics, both at network and application layers. The traffic at the transmitter and at the receiver have been dumped through Tshark, an open-source network protocol analyzer.

A. Traffic Analysis

This section analyses the performance of the three network operators by using dumps collected before the dejitter buffer (i.e., at the blue block in Figure 2). The cross-correlation of the packet arrivals from the three operators (see Table II) is calculated by means of the Pearson's correlation coefficient, verifying if a given packet, with a given sequence ID, sent on a given link, has been successfully received or not. It is worth noting that such a correlation is almost negligible in the urban part, while the same cannot be said in the suburban part. From

	Vodafone	Wind-Tre	TIM
(U) Vodafone	1	-0.03	0.14
(U) Wind-Tre	-0.03	1	0.1
(U) TIM	0.14	0.1	1
(SU) Vodafone	1	0.71	0.016
(SU) Wind-Tre	0.71	1	0.010
(SU) TIM	0.016	0.010	1

TABLE II: Cross-correlation of the packet arrival process at the receiver in both scenarios: urban (U) and suburban (SU).

a logical viewpoint, in order to take as much advantage of possible of multipath techniques, the links should be ideally uncorrelated. Otherwise, the expected *gain* decreases as the

(a) suburban fraction of the testbed (Pisa's surroundings, Italy)

(b) urban fraction of the testbed (Pisa, Italy)

Fig. 3: The maps of the real testbed performed by car. The path is marked in yellow, the eNBs in blue.

correlation increases. This is further discussed in Section IV-B, looking at the achieved video quality.

Table III reports the network statistics. In the first column, the packet loss (PL) rate is shown: while the overall values are rather consistent in both fractions of the testbed, the TIM operator suffers from noticeable higher rates than the other two cellular operators. The second column shows the error burst length (EBL). In the urban scenario, the EBLs are rather variable; thanks to the low cross-correlation, the EBL in the aggregated case is comparable to the best achieved value. This improvement is almost absent in the suburban case because of the aforementioned cross-correlation. The third column shows the probability of having out of sequence packets. In the urban case, the Vodafone operator has the larger probability, while the other two operators have comparable statistics. In the suburban case, the comparable values are those of the two cross-correlated operators. Out of sequence packets are handled by the dejitter module (see Figure 2), which provides reordering functionalities. However, packets reordering is done at the expense of the delay, an important constraint for real-time video, as a C2 scenario. Our setup tolerates a certain probability of having out of sequence packets, according to the dejitter buffer size (see Table I): its reduced length satisfies real-time constraints, but paying some corrupted video frames, as we show in Section IV-B by means of metrics on video quality. Finally, the last column of Table III shows the average one-way delay, along with its variance, of the three network operators. The urban case shows slightly lower average values than the suburban one.

Operator	**PL rate**	**EBL**	**out of seq. probability**	**one-way delay** *avg [ms]*	*var* E-03
(U) Vodafone	2.1 E-05	1.00	10 E-02	44	0.51
(U) Wind-Tre	3.0 E-02	30.9	5.0 E-03	46	2.2
(U) TIM	19 E-02	10.6	5.0 E-03	39	0.57
(U) Aggregated	2.0 E-05	1.00	9.5 E-03	40	0.17
(SU) Vodafone	1.5 E-05	1.00	4.3 E-02	47	0.48
(SU) Wind-Tre	2.1 E-05	1.01	5.1 E-02	55	0.69
(SU) TIM	5.8 E-02	5.17	1.9 E-04	50	0.56
(SU) Aggregated	1.5 E-05	1.00	1.4 E-02	48	0.23

TABLE III: Network statistics in both scenarios.

B. Video Quality

In this section, we provide the evaluation of the perceived video quality at the GCS, according to Peak signal-to-noise ratio (PSNR) and Structural SIMilarity (SSIM) metrics, and according to the perceived QoE of generic users, and of an UAV pilot (i.e., a specialized technician). The pilot is more interested in video continuity than in the overall video quality, i.e., the pilot tolerates better than generic users the loss of details if the stream suffers from less interruptions. In more words, being able to always have an usable (i.e., continuous) video feedback is more important than having an optimum video feedback with even few interruptions. Those results are reported in Table IV, and discussed in the following. Note that the results in this section are calculated after the dejitter buffer (at the red block in Figure 2).

During the testbed, the live video was that resulting from the aggregation of the three flows. In order to assess the improvement provided by multipath techniques w.r.t. to using a single operator, we later reconstructed offline the streaming by using network dumps for each cellular operator. Looking at Table IV, the poor performance of the TIM operator, if used alone, can be read in the third row of the urban scenario; better results are provided by the other two operators. Thanks to the negligible cross-correlation shown in Table II, the quality of the aggregated case is ranked as excellent by the pilot. Looking at the video quality in the 2nd, 3rd, and 4th column, it can be noted that the aggregated case, in the forth row, has a number of low quality frames comparable to the best single performance (i.e., 7% vs 6.9%), medium quality is improved (0.3% vs 0.7%), and high quality is slightly improved as well (92.7% vs 92.4%). As a result, the SSIM of the aggregated case shows an improvement ranging from 5% to 33% on the one hand; on the other hand, such an improvement has the cost of using three links instead of a single one. Anyway, flight safety is the foremost requirement to be met, thus this setup provides noticeable advantages. The suburban scenario shows a more consistent performance among the three operators, and the lower improvement of the SSIM/PSNR must be read as an effect of the larger cross-correlation.

Operator	low vq	medium vq	high vq	avg PSNR	avg SSIM	subj. eval. (GUs)	subj. eval. (P)
(U) Vodafone	7.1%	2.4%	90.0%	46.85	0.8605	good	good
(U) Wind-Tre	6.9%	0.7%	92.4%	45.49	0.8778	good	good
(U) Tim	41.8%	0.7%	57.5%	33.23	0.5919	poor	poor
(U) Aggregated	7.0%	0.3%	92.7%	47.6	0.9301	good	excellent
(SU) Vodafone	7.0%	2.3%	90.7%	46.93	0.9045	good	good
(SU) Wind-Tre	11.9%	0.1%	88.0%	47.41	0.9206	good	excellent
(SU) Tim	18.0%	1.0%	81.0%	43.14	0.8181	fair	good
(SU) Aggregated	6.3%	0.1%	93.6%	47.46	0.9203	good	excellent

TABLE IV: Statistics on the video quality (vq) based on PSNR [dB] and SSIM metrics for the three cellular operators and for the aggregated case. We assume PSNR $\geq$ 50 dB as high vq, low vq if PSNR $\leq$ 20 dB, medium vq otherwise. Subjective evaluation is shown in the last two columns, according to generic users (GUs) and to an UAV pilot (P).

V. CONCLUSIONS

In this paper, we described a real testbed for enabling BVLoS flight conditions exploiting public cellular networks. The large testbed we conducted has provided us various insights on the performance of a multipath real-time video streaming via 4G networks, which we reported in this work. The results show how the use of multipath techniques provides an increase of the video quality in C2 scenarios, i.e., in the presence of strict real-time requirements. When using uncorrelated links, the performance increase is really significant (up to 33% in our tests). In future works, we plan on extending such an analysis in order to analytically model the impact of both cross-correlation among network links and network statistics on the achievable video quality.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Marco Colucci and Dr. Andrea Berton for their efforts in this work. This work has been supported by the Tuscany Region (DRGT 758/2013) in the framework of the SCIADRO (SCIAmi di DROni - Drone Swarms) project (PAR-FAS 2007-2013 FAR-FAS 2014 Action Line B).

REFERENCES

[1] I. Bor-Yaliniz and H. Yanikomeroglu, "The new Frontier in RAN Heterogeneity: Multi-Tier Drone-Cells," *IEEE Communications Magazine*, vol. 54, no. 11, pp. 48–55, 2016.

[2] M. Alzenad, A. El-Keyi, F. Lagum, and H. Yanikomeroglu, "3D Placement of an Unmanned Aerial Vehicle Base Station (UAV-BS) for Energy-Efficient Maximal Coverage," *IEEE Wireless Communications Letters*, vol. 6, no. 4, pp. 434–437, 2017.

[3] M. Bacco, F. Delmastro, E. Ferro, and A. Gotta, "Environmental Monitoring for Smart Cities," *IEEE Sensors Journal*, vol. 17, no. 23, pp. 7767–7774, 2017.

[4] M. Bacco, A. Berton, A. Gotta, and L. Caviglione, "IEEE 802.15.4 Air-Ground UAV Communications in Smart Farming Scenarios," *IEEE Communications Letters*, pp. 1–4, July 2018.

[5] M. Bacco, A. Berton, E. Ferro, C. Gennaro, A. Gotta, S. Matteoli, F. Paonessa, M. Ruggeri, G. Virone, and A. Zanella, "Smart Farming: Opportunities, Challenges and Technology Enablers," in *IoT Vertical and Topical Summit on Agriculture-Tuscany (IOT Tuscany)*, 8-9 May 2018, pp. 1–6.

[6] M. Bacco, S. Chessa, M. Di Benedetto, D. Fabbri, M. Girolami, A. Gotta, D. Moroni, M. A. Pascali, and V. Pellegrini, "UAVs and UAV Swarms for Civilian Applications: Communications and Image Processing in the SCIADRO Project," in *International Conference on Wireless and Satellite Systems*. Springer, 2017, pp. 115–124.

[7] X. Lin, V. Yajnanarayana, S. D. Muruganathan, S. Gao, H. Asplund, H.-L. Maattanen, M. Bergstrom, S. Euler, and Y.-P. E. Wang, "The Sky is not the Limit: LTE for Unmanned Aerial Vehicles," *IEEE Communications Magazine*, vol. 56, no. 4, pp. 204–210, 2018.

[8] B. Van der Bergh, A. Chiumento, and S. Pollin, "LTE in the Sky: Trading off Propagation Benefits with Interference Costs for Aerial Nodes," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 44–50, 2016.

[9] H. C. Nguyen, R. Amorim, J. Wigard, I. Z. Kovacs, and P. Mogensen, "Using LTE Networks for UAV Command and Control Link: A Rural-Area Coverage Analysis," in *Vehicular Technology Conference (VTC-Fall), 2017 IEEE 86th*. IEEE, 2017, pp. 1–6.

[10] S. Barré, C. Paasch, and O. Bonaventure, "Multipath TCP: from Theory to Practice," in *International Conference on Research in Networking*. Springer, 2011, pp. 444–457.

[11] V. Singh, S. Ahsan, and J. Ott, "MPRTP: Multipath Considerations for Real-Time Media," in *Proceedings of the 4th ACM Multimedia Systems Conference*. ACM, 2013, pp. 190–201.

[12] D. H. Bui, K. Lee, S. Oh, I. Shin, H. Shin, H. Woo, and D. Ban, "Greenbag: Energy-efficient Bandwidth Aggregation for Real-Time Streaming in Heterogeneous Mobile Wireless Networks," in *Real-Time Systems Symposium (RTSS), 2013 IEEE 34th*. IEEE, 2013, pp. 57–67.

[13] B. Wang, W. Wei, Z. Guo, and D. Towsley, "Multipath Live Streaming via TCP: Scheme, Performance and Benefits," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 5, no. 3, p. 25, 2009.

[14] C. Raiciu, D. Niculescu, M. Bagnulo, and M. J. Handley, "Opportunistic Mobility with Multipath TCP," in *Proceedings of the sixth international workshop on MobiArch*. ACM, 2011, pp. 7–12.

[15] M. Bacco, M. Catena, T. De Cola, A. Gotta, and N. Tonellotto, "Performance Analysis of WebRTC-based Video Streaming over Power Constrained Platforms," in *Global Communications Conference (GLOBECOM), 9-13 December 2018, Abu Dhabi UAE*. IEEE, 2018, pp. 1–6.

[16] A. Gotta and P. Barsocchi, "Experimental Video Broadcasting in DVB-RCS/S2 with Land Mobile Satellite Channel: a Reliability Issue," in *Satellite and Space Communications, 2008. IWSSC 2008. IEEE International Workshop on*. IEEE, 2008, pp. 234–238.

[17] P. Barsocchi, A. Gotta, F. Potortì, F. J. González-Castaño, F. Gil-Castiñeira, J. I. Moreno, and A. Cuevas, "Experimental Results with Forward Erasure Correction and Real Video Streaming in Hybrid Wireless Networks," in *Wireless Communication Systems, 2005. 2nd International Symposium on*. IEEE, 2005, pp. 662–666.

[18] A. Li, "RFC 5109 - RTP Payload Format for Generic Forward Error Correction," IETF, Tech. Rep., 2007.