Network Coding for Next Generation Personal Satellite Converged Services

Daniel E. Lucani[1,*] and Marie-José Montpetit[2]

[1] Instituto de Telecommunicações, DEEC, Faculdade de Engenharia da Universidade do Porto, Porto 4200-465, Portugal
[2] Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA
`dlucani@fe.up.pt`, `mariejo@mit.edu`

Abstract. Seeking to meet the resilience, efficiency, and quality of experience challenges of personal and converged satellite services, we present a new approach that leverages the benefits of network coding. The salient features of this strategy are (i) source nodes manage and control the transmission of linear combinations of data packets through heterogeneous communication routes, and (ii) intermediate nodes at each route can generate new coded packets with opportunistic storage. Hence, the amount of data and redundancy sent through each route can meet the required performance of the different sessions. In particular it can help surmount varying channel conditions and correct erasures but also adapt to the different delays and bandwidth that are features of the converged PSAT networks of the future. The main technical challenge is to choose adequate, coding-aware policies to leverage heterogeneous networks based on content and user requirements. We present preliminary analysis that illustrates that exploiting the routes jointly can be performed seamlessly using network coding even with limited feedback capabilities.

Keywords: Convergence, Heterogeneous Networks, Network Coding, Satellite Communications.

1 Introduction

Following well-established trends in the terrestrial networks, personal satellite services (PSATS) will provide end users with a variety of personalized services focused on rich media content. While there is still a difference in terms of spectrum between fixed and mobile satellite systems, the same convergence that was experienced in terrestrial networks in terms of services is happening in the satellite realm. The aim there is to provide services that are independent of the satellite system that is serving them and leverage terrestrial networks when appropriate. Advances in terrestrial and space segments' technology already allow personalized multimedia and broadband services to be offered from a single satellite system, such as the Inmarsat Global Xpress [1], and there are technical and

---

[*] This work was supported in part by the Instituto de Telecommunicações COHERENT Project, under the FCT project PEst-OE/EEI/LA0008/2011.

P. Pillai, R. Shorey, and E. Ferro (Eds.): PSATS 2012, LNICST 52, pp. 17–25, 2013.

commercial incentives to continue integrating satellite and terrestrial networks, as demonstrated by systems like TerreStar [2]. Recent advances in networking technologies offer novel approaches for this seamless platform integration. In this paper, network coding will be presented as a key enabler for i) improving the performance of IP-based protocols over satellite, but also for ii) empowering network combining itself be it satellite-satellite of the fixed or mobile category or satellite-terrestrial when appropriate. We infer that the resulting services strategies are instrumental in securing the position of satellite networks in the next generation Internet.

Network coding (NC) considers digital traffic as algebraic entities not just data that needs to be transported [3]. Because these entities can be multiplied by constants and linearly combined, some of the usual networking constraints such as state information and independent use of network resources can be lifted. In addition since the coding parameters can be simple random numbers over a fairly short Galois field [4] the coding and decoding process may remain simple [5]. While network coding is not specific to a single type of network, it has shown promise on long delay networks such as satellite and underwater networks [6].

In the remainder of the paper we present our approach to satellite network convergence based on NC. We will show that it allows soft service-focused handover and heterogeneous network combining with overall improved performance. Section 2 reviews some aspects of satellite system convergence leading to our proposed architectures, outlined in Section 3. Section 4 provides a very short introduction to network coding that clarifies aspects of the analysis presented in Section 5. This analysis is based on heterogeneous path analysis and probabilistic principles. Section 6 concludes with some look into the future especially referring to the position of satellite systems for content dissemination in the Future Internet.

2 Convergence in Satellite Networks

Satellite networks are impacted by the disruptions in the communications industry, driven by an emphasis on Internet based services, machine-to-machine communications, and ubiquitous multimedia. In this environment, the converged nature of many services is driving innovation including systems that combine satellite and terrestrial 3G/4G to guarantee Internet access everywhere. The successful integration of satellite networks into the Internet has been progressing for over 15 years and has lead to implementable standards such as those defined in the IETF TCPSAT [7] and IPDVB [8] as well as DVB-RCS [9] and the ETSI-BSM [10].

Moving further, converged networking architectures are characterized by the unbundling of services and underlying transport functions (i.e. network technologies) which enables the definition of service-driven network policies independently of the underlying infrastructure. One architectural embodiment of the principle is provided by the Next-Generation Networks (NGN) architecture, one that has already been adapted to broadband services via fixed satellite systems [11] [12]. While these are major steps towards satellite-terrestrial hybrid networks, there remains open challenges to respond in terms of capacity, packet erasures, and delay mismatches between systems. Full convergence of the satellite network themselves has not being fully investigated but is an active field of research. There are many services, from infrastructure support to social television and emergency management, which require revisiting the way satellites and other networks can work together. The new challenge is to develop mechanisms to exploit collaboration and interaction across the different domains beyond exchanges of signaling information.

This is where NC can clearly make a difference. In order to profit to the maximum of the use of NC we need to investigate the architecture of the network that will support it from the source nodes that implement the content mixing to the judicious placement of intermediate nodes, which without the need of decoding, can remix different flows and guarantee they will be delivered together to the receiver. The decoding can be naturally pushed all the way to the end user device. This architecture is presented in the next section.

3 Architecture and Derived Scenarios

A hybrid architecture that combines both space and ground segments is considered in this paper. Figure 1 presents an overview of the targeted multi-system and converged ecosystem. The space segment can be one or more fixed and mobile systems and the ground segments, while in principle of any type, will be assumed to be mostly wireless and broadband in nature. Specific scenarios can be carved from this generic architecture that match service characteristics or operational requirements. In all our scenarios, we will however assume that the NC elements remain in the terrestrial domain. While the operations needed to perform network coding are linear in nature and require little processing, their implementation in a regenerative satellite are beyond the scope of this paper.

Hybrid architectures are however plagued by the consequences of heterogeneity. For example, different paths to the same destination may experience differing delays, fading and erasures, not all network capacity comes at the same operational cost, etc. NC will reduce the complexity of managing and operating this hybrid network by enabling mixes of data to be sent through different paths and be recombined at the end without having to keep packet trackers and otherwise state information. As will be seen in the next section, even without implementing a feedback loop to manage the mixing of packets, significant gains can be made when compared with non-network coded system. With NC the capabilities of each system, such as capacity or cost of the bandwidth, can be traded, combined and optimized. In our architecture, the source nodes (the data sources) implement the initial network coding and the information can be sent via a combination of terrestrial and satellite paths who will act as partners in the delivery of the information. Intermediate nodes, like satellite ground stations, receivers and wireless nodes can act as decoders or as re-encoders based on the instantaneous conditions they experience but also delay requirements or user equipment capabilities.

Use cases can be defined based on the elements of Figure 1. The first one is system combining. With network coding, a satellite system can complement another one in times of fading. These could be of the fixed-fixed, fixed-mobile or mobile-mobile category depending on equipment, business models and regulatory environments. Even if the second satellite is of lower capacity, the combination of the two systems can allow a session/terminal to survive a momentary fading and loss of capacity without significant QoS degradation and facilitate a handover if appropriate. As will be seen in the next section there is a high probability that the rich mix of content coming from a secondary path will contain enough mixed packets (degrees of freedom) to recuperate lost or delayed information on the primary path. There is evidence coming from terrestrial networks that only a few packets borrowed from a secondary path can greatly improve overall performance [13].

Another use case could combine a wireless/mobile terrestrial network while leveraging a fixed satellite network; this can add interactivity and support rich video applications. It will be seen in the analysis (Section 5) that with network coding this can be implemented with little complexity. Since, network coding relaxes the need for maintaining session state and data pointers in relaying nodes, it supports this combination of fixed satellite and terrestrial networks. As can be foreseen from these two very simple use cases, NC can provide additional the coverage to increase the number of satisfied customers.

4 Network Coding Primer

Network coding (NC) offers exciting possibilities for the efficient transmission of video over wireless and bottleneck networks by considering traffic as algebraic information and sending linear combinations of packets. Overall NC allows to reduce the required number of transmissions to complete a file or stream operation over noisy or unreliable networks. This results in increased throughput but also, since the smaller completion time will reduce application layer timeouts, of goodput which is a strong measure of Quality of Experience (QoE). While NC adds some complexity to both source and destination nodes since it involves performing linear operations these are not too demanding and have been successfully implemented in hand-held devices [5]. Also the use of codes defined on small Galois fields (GF), and of codes that are systematic, i.e., packets are initially sent uncoded followed by a number of mixed packets, can significantly reduce coding operation complexity [14].

We illustrate the idea of network coding in a simple multiple route scenario in Figure 2 and compare it to a more traditional technique. A classical system trying to include additional redundancy could transmit by assigning resources in both routes and sending the same packets through each of the beams, as in Figure 2 (a). This decision is agnostic to the underlying channel conditions and is meant to provide additional reliability when a single packet transmission does not fulfill the required reliability. In a network coded system, we have the flexibility to send a different fraction of the data through each beam (as long as enough coded packets are sent) and to choose the desired level of redundancy. In Figure 2 (b) the system chooses to send 3 coded packets through one route and 2 through the other due to channel constraints and/or system load.

Although the system of Figure 2 (a) sends one packet more than its coded counterpart of Figure 2 (b), it is simple to see that the coded system provides

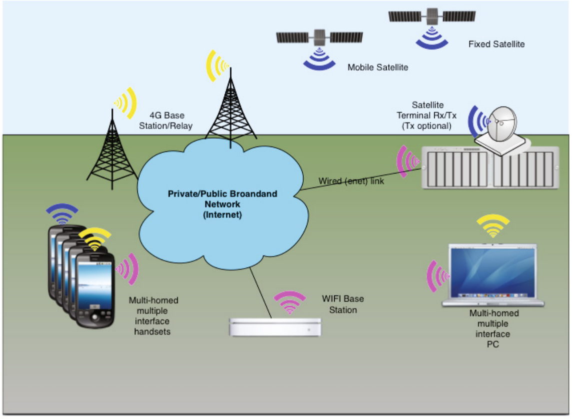

Fig. 1. Overview

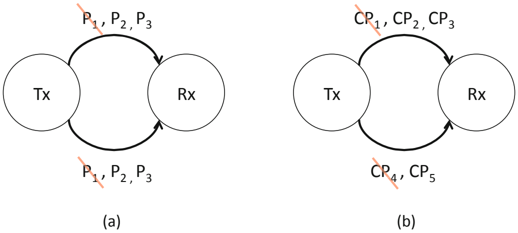

Fig. 2. A simple example of multiple routes. (a) Example of a transmission that sends the same 3 data packets through each route for increased reliability. (b) Example of a network coding approach where coded packets are sent through each route.

higher resiliency to packet losses. In our example, both cases may sustain up to two packets being lost. However, without coding when the same packet is lost in both routes, that packet is simply not recovered. Our coded example does not share this problem as different linear combinations can be sent through each route, guaranteeing resilience to exactly 2 packet transmissions, since the receiver only requires 3 independent linear combinations (out of 5 that were generated) to recover the original data.

From the example of Figure 2 (b) it is clear that a coded mechanism can provide additional guarantees of recovering all data packets. However, if more losses occur it may impede the receiver from recovering any single packet. In RLNC, coding across M packets requires M coded packets to recover any information. The key is to find a trade-off between partial recovery of the data by using a sparser code (RLNC is a dense code) and the inherent loss in performance due to the sparser nature of the code (RLNC is delay-optimal in our example for large enough field). A simple solution is to use a systematic structure, i.e., original packets are sent without coding once, while all additional packets are sent coded with RLNC [14]. Systematic network coding provides no degradation in performance while ensuring i) partial recovery of the packets, and ii) a reduction in decoding complexity, as shown in [14].

5 Motivating Analysis

Consider the problem of transmission of data packets from a source to a destination in a time-slotted system, where two independent channels are available. At each time slot, the source can transmit random linear network coded packets through both channels (different coded packet in each), one channel, or can decide to not transmit in that time slot.

We assume an independent Gilbert-Elliott model for the channels, where $p_{(g,i)}$ and $p_{(b,i)}$ are the corresponding packet erasure (loss) probabilities on the i-th link for the good and bad channel, respectively. The probability of link i to remain in state $c \in \{b, g\}$ is given by $p_c^{(i)}$.

We assume that a genie indicates the state $C = (c_1, c_2)$ of the two channels, i.e., the probabilities of packet loss in each channel, at each time slot. However, the event of a packet loss is not known *a priori* to the genie.

Let us model the general state of the system at the n-th time slot as $\mathcal{S}(n) = (Q(n), c_1(n), c_2(n), \mathcal{P}(n))$, where $Q(n)$ represents the number of independent linear combinations (or degrees of freedom) missing at the receiver at time slot n, $c_1(n)$ and $c_2(n)$ represent the state of the Gilbert-Elliott channels, and $\mathcal{P}(n)$ constitutes the source's policy. We assume an online NC approach as in [15].

We define a policy $\mathcal{P} \in \big\{\emptyset, S_1, S_2, \{S_1, S_2\}\big\}$ as a function that schedules the transmission of each packet based on the channel states of the available routes, where S_i represents the event of the source transmitting through channel i.

5.1 Optimal Policy in Terms of Reduction of Channel Utilization

In the following we define an optimal policy in terms of the reduction of the mean channel utilization, using a similar technique to [16]. Let us first define the *effective* mean erasure probability of S_i, $\bar{p}_{\text{eff}}(i)$ as the mean erasure probability seen by the transmitter as a result of the scheduling policy. We define $\alpha_{(i,C)}$ and $Pr_{(i,C)}$ as the fraction of the rate of S_i and the probability of transmission through channel i during the channel state C, respectively. Thus,

$$\bar{p}_{\text{eff}}(i) = \sum_{C=(c_1,c_2)\in\{g,b\}^2} p_{(c_1,i)} Pr_{(i,C)} \pi_C, \tag{1}$$

where π_C constitutes the stationary probability of the channel state C, which can be easily determined through standard finite Markov chain techniques. The channel utilization of channel i in our system is given by

$$U_i([Pr_{(i,C)}]) = \sum_C Pr_{(i,C)} \pi_C. \tag{2}$$

The optimization problem is stated as:

$$\begin{aligned}
&\min_{[\Pr_{(i,C)}]} \sum_i U_i\left(Pr_{(i,C)}\right),\\
&\text{subject to}\\
&\Pr_{(i,C)} \in [0,1], \quad \forall C \in \{g,b\}^2, i \in \{1,2\}\\
&\sum_{i\in\{1,2\},C\in\{g,b\}^2} \alpha_{(i,C)} = 1,\\
&(1-p_{(c_1,1)})Pr_{(1,C)}\pi_C = \lambda\alpha_{(1,C)}, \quad \forall C \in \{g,b\}^2,\\
&(1-p_{(c_2,2)})Pr_{(2,C)}\pi_C = \lambda\alpha_{(2,C)}, \quad \forall C \in \{g,b\}^2.
\end{aligned}$$

In the first line, we have used the right hand side of (1) as the argument of $U_i(\cdot)$. The last two conditions capture the fact that the probability of S_k transmitting in a given channel state is linked to the mean usage of the channel during that state, e.g., $\lambda\alpha_{(1,C)}/(1-p_{(c_1,1)})$ for channel 1. We emphasize that to achieve this throughput performance using no NC would require a feedback mechanism that signals the correct reception of packets sent at each time slot. On the other hand, NC can achieve this performance naturally without requiring such an intensive feedback mechanism. In fact, feedback is required for other practical purposes, for example, i) to maintain reasonably sized queues at the sender while reducing decoding complexity [15], or ii) to enforce a required delay for decoding the packet at the receiver [17].

The optimal policy for a given channel state C and source rate λ is given by the vector $[Pr_{(i,C)}]$ that results of this optimization. The optimal policy $\mathcal{P}_{\text{opt}}(C) \in \big\{\emptyset, S_1, S_2, \{S_1,S_2\}\big\}$ can be stated as

Policy 1

$$\mathcal{P}_{opt}(C) = \begin{cases} \{S_1,S_2\} & w.p.\ Pr_{(1,C)}Pr_{(2,C)}\\ S_1 & w.p.\ Pr_{(1,C)}(1-Pr_{(2,C)})\\ S_2 & w.p.\ (1-Pr_{(1,C)})Pr_{(2,C)}\\ \emptyset & otherwise \end{cases}$$

where S_i indicates the event of transmitting through channel i. Note that the probability of transmitting through channel 1 and channel 2 is independent, thus transmission over two channels or no channels at each time slot is possible.

This policy is probabilistic in nature and relies on the fact that the queues are assumed to be of infinite size. The latter allows the system to store the (coded) packets while awaiting a good channel. In practice, deterministic algorithms that link the queue lengths and channel-awareness may be more relevant since they may guarantee better delay performance, albeit with a possibly degraded throughput region.

6 Conclusion

This paper presents network coding as a key enabler for personal satellite converged services. It emphasizes an architecture for these satellite networks that maximizes the benefits of network coding by exploiting multiple available routes in the space and terrestrial segments and opportunistic coding at various locations in the network. The crux of the solution lies in choosing coding-aware policies that leverage (i) statistics from the available, time-varying channels, and (ii) content and user requirements to allocate the right level of redundancy at each transmission route as a means to control the performance of each session. A case study with Markovian channels is reported; it emphasizes an optimal probabilistic policy that performs transmission decisions based on a joint state of the routes available to the transmission. This result represents a very promising starting point for more in-depth evaluation of the potential benefits of network coding for converged satellite services. Future research will analyze the different available channels jointly as well as adding ancillary information to the policies, e.g, statistics of the channels, knowledge of traffic statistics, feedback etc, to provide practical mechanisms that can be seamlessly implemented in current and future systems. By enabling novel combination of systems and fixed-mobile as well as satellite-terrestrial convergence NC will allow operators to flexibly define new services and refine current offerings for optimized performance in terms of cost, delay and other quality of service parameters that are required by their customers.

Acknowledgements. The authors would like to acknowledge the ETSI Satellite Earth Station and Systems (SES) Working Group for initiating the discussion on fixed-mobile satellite networks and supporting work on satellite-terrestrial integration in the NGN, in particular R. Mort, R. Goodings, A. Noerpl, and N. Chuberre. We would also like to thank Marcus Vilaça and David Bath of Inmarsat, and Dave Bettinger of iDirect for their encouragement in pursuing this work. Finally we would like to thank our anonymous reviewers for their comments leading to this version of the paper.

References

1. Global Xpress, `http://www.inmarsat.com/About/Newsroom/00037147.aspx?language=EN&textonly=False`
2. TerreStar, `http://www.terrestar.com/`
3. Koetter, R., Médard, M.: An Algebraic Approach to Network Coding. IEEE/ACM Transactions on Networking 11(5) (2003)
4. Ho, T., Koetter, R., Médard, M., Karger, D.R., Effros, M.: The Benefits of Coding over Routing in a Randomized Setting. In: IEEE International Symposium on Information Theory (2003)
5. Heide, J., Pedersen, M.V., Fitzek, F.H.P., Larsen, T.: Network Coding for Mobile Devices - Systematic Binary Random Rateless Codes. In: IEEE International Conference on Communications (ICC) - Workshop on Cooperative Mobile Networks, Dresden, Germany (2009)
6. Lucani, D.E., Médard, M., Stojanovic, M.: On Coding for Delay - New Approaches Based on Network Coding in Networks with Large Latency. In: Proc. ITA Workshop, San Diego (2009)
7. IETF TCP over Satellite (TCPSAT), `http://datatracker.ietf.org/wg/tcpsat/charter/`

8. IETF IP over DVB - IPDVB Working Group, `http://datatracker.ietf.org/wg/ipdvb/charter/`
9. ETSI EN 301 790 v.1.4.1: Digital Video Broadcasting (DVB); Interaction channel for satellite distribution systems (2005-09)
10. ETSI TR 101 984 v1.1.1: Satellite Earth Stations and Systems (SES); Broadband satellite multimedia; Services and Architectures (2002-2011)
11. ETSI ES 282 001 v3.4.1: Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); NGN Functional Architecture (2009)
12. ETSI TS 102 855 v.1.1.1: Satellite Earth Stations and Systems (SES); Broadband satellite multimedia; Interworking and Integration of BSM in Next Generation Networks (NGNs) (2011-03)
13. Kulkarni, A., Heindlmaier, M., Traskov, D., Medard, M., Montpetit, M.J.: Network Coding with Association Policies in Heterogeneous Networks. In: Proc. NC-Pro 2011 (May 2011)
14. Lucani, D.E., Médard, M., Stojanovic, M.: Systematic network coding for time-division duplexing. In: Proc. IEEE International Symposium on Information Theory, Austin, TX, USA, pp. 2403–2407 (2010)
15. Sundararajan, J.K., Shah, D., Médard, M.: ARQ for network coding. In: IEEE International Symposium on Information Theory, Toronto, Canada, pp. 1651–1655 (2008)
16. Lucani, D.E., Kliewer, J.: On the Delay and Energy Performance in Coded Two-Hop Line Networks with Bursty Erasures. In: Proc. IEEE International Symposium on Wireless Communication Systems, Russia (2011)
17. Lucani, D.E., Médard, M., Stojanovic, M.: Online Network Coding for Time-Division Duplexing. In: IEEE Global Telecommunications Conference, Miami, FL, USA, pp. 1–6 (2010)