# TRAFFIC OFFLOADING IN A CONVERGED SATELLITE AND TERRESTRIAL NETWORK

A thesis submitted for the degree of
Doctor of Philosophy

by
Christian Niephaus

Department of Computer Science
Brunel University London

May 2020

# Abstract

Rural and remote areas all over the world still suffer from the digital divide and limited access to broadband Internet connections. Despite ambiguous goals of many institutions, such as the European Union, operators are reluctant to deploy the same broadband infrastructure in such difficult to serve areas due to economical constraints. Simultaneously, satellite networks evolved significantly during the last decade, so that they can virtually provide broadband connectivity everywhere. Geostationary satellites, however, introduce a high amount of latency in each connection due to their altitude of 36 000 km, leading to a negative effect on the end user's experience. Hence, a promising solution is to combine narrow-band terrestrial connection with high capacity satellite links to form a converged satellite and terrestrial network.

The contribution of this thesis rests in the investigation, design and evaluation of offloading techniques, that allow for distributing traffic in a converged satellite and terrestrial network, so that the end user experience in rural and remote areas can be increased.

Throughout this thesis, firstly the problem space has been investigated, so that appropriate research questions could be derived. Secondly, solutions based on these questions have been proposed and, finally, evaluated. The results show that a converged satellite and terrestrial network can be formed, in which traffic is effectively offloaded from the limited terrestrial link to the satellite connection. With such an approach, the overall network performance in rural and other underserved areas benefits significantly, so that eventually the end users' experience will increase.

# Acknowledgements

Looking back across eight years since embarking upon this PhD, I was very lucky to count on the enormous support from friends and colleagues.

First of all, I would like to thank my principal supervisor, Gheorghita Ghinea, for his support, feedback and, most importantly, for his constant motivation during the whole journey.

Mathias Kretschmer and Jens Moedeker also deserve a very big 'Thank you' for reviewing publications and for the very valuable and fruitful discussions during the whole time. Furthermore, I would like to thank Karl Jonas, who encouraged me to start with the PhD in the first place, for his support. Last but definitely not least a very special *Thank you* goes to my fiancée Sandra for her love, support and understanding throughout the whole work.

# List of Publications

## Journals

- C. Niephaus and G. Ghinea, "Evaluation of offloading approaches in converged satellite and terrestrial networks" in *IEEE Transactions on Broadcasting*, under review
- C. Niephaus, J. Mödeker and G. Ghinea, "Toward Traffic Offload in Converged Satellite and Terrestrial Networks" in *IEEE Transactions on Broadcasting*, vol. 65, no. 2, pp. 340-346, June 2019.
- C. Niephaus, M. Kretschmer and G. Ghinea, "QoS Provisioning in Converged Satellite and Terrestrial Networks: A Survey of the State-of-the-Art," in *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2415-2441, Fourthquarter 2016.
- C. Niephaus, O. G. Aliu, M. Kretschmer, S. Hadzic and G. Ghinea, "Wireless Back-haul: a software defined network enabled wireless Back-haul network architecture for future 5G networks," in *IET Networks*, vol. 4, no. 6, pp. 287-295, 11 2015.

## Peer-reviewed Conferences and Workshops

- C. Niephaus, G. Ghinea, O. G. Aliu, S. Hadzic and M. Kretschmer, "SDN in the wireless context - Towards full programmability of wireless network elements," *2015 1st IEEE Conference on Network Softwarization (NetSoft)*, London, 2015, pp. 1-6.
- C. Niephaus, O. G. Aliu, M. Kretschmer, S. Hadzic and G. Ghinea, "WiBACK: A Back-haul network architecture for 5G networks," *International Conference on Frontiers of Communications, Networks and*

*Applications (ICFCNA 2014 - Malaysia)*, Kuala Lumpur, 2014, pp. 1-6.

- C. Niephaus, M. Kretschmer, G. Ghinea and S. Hadzic, "Enabling Converged Satellite and Terrestrial Access Networks", *e-Infrastructure and e-Services for Developing Countries - 6th International Conference AFRICOMM 2014*, Kampala, 2014, pp. 43-52

- C. Niephaus, M. Kretschmer and G. Ghinea, "Towards QoS-aware load distribution in heterogeneous networks," *2013 IEEE 11th Malaysia International Conference on Communications (MICC)*, Kuala Lumpur, 2013, pp. 151-156.

# Contents

# List of Figures

# List of Tables

# Acronyms

**ACM** adaptive coding and modulation

**APN** Access Point Name

**APSK** Amplitude and phase-shift keying

**ARP** Address Resolution Protocol

**BATS** Broadband Access via integrated Terrestrial and Satellite systems

**BE** best-effort

**BGP** Border Gateway Protocol

**BS** Base Station

**CDN** Content Delivery Network

**CPE** Customer premises equipment

**CRA** constant rate assignment

**CSMA** Carrier-Sense Multiple Access

**DAMA** Demand Assigned Multiple Access

**DASH** Dynamic Adaptive Streaming over HTTP

**Diffserv** Differentiated Services

**DNS** Domain Name Service

**DPI** Deep Packet Inspection

**DSCP** Differentiated Services Code Point

**DSL** Digital Subscriber Line

**DSR** Design Science Research

**DVB-GSE** DVB Generic Stream Encapsulation

**DVB-RCS** DVB Return Channel via Satellite

**DVB-RCS2** DVB Return Channel via Satellite - Second Generation

**DVB-S2** Digital Video Broadcasting-Satellite - Second Generation

**eNodeB** Evolved Node B

**EPC** evolved Packet Core

**ESA** European Space Agency

**ETSI** European Telecommunications Standards Institute

**FAMA** Fixed Assigned Multiple Access

**FCA** free capacity allocation

**FEC** Forward Error Correction

**FSS** Fixed Satellite Service

**FTP** File Transfer Protocol

**GEO** Geostationary Earth Orbit

**HeNB** Home eNB

**HTML** Hypertext Markup Language

**HTTP** Hypertext Transfer Protocol

**ICT** Information and Communication Technology

**IETF** Internet Engineering Task Force

**IoT** Internet of Things

**IP** Internet Protocol

**ISL** Inter-satellite link

**ISP** Internet Service Provider

**KPI** Key Performance Indicator

**LAN** Local Area Network

**LEO** Low Earth orbit

**LSP** Label-Switched Path

**LTE** Long Term Evolution

**MAC** Media Access Control

**MCS** Modulation and Coding Scheme

**MEO** Medium Earth Orbit

**MIMO** multiple-input and multiple-output

**ML** Machine Learning

**MOS** Mean Opinion Score

**MPLS** Multi Protocol Label Switching

**MPTCP** Multipath Transmission Control Protocol

**MSS** Mobile Satellite Services

**MT** Mobile Terminal

**MTC** Machine-Type-Communication

**NAT** Network Address Translation

**NCC** network control center

**NFV** Network Function Visualization

**NRTV** Near Real-Time Video

**NS-3** Network Simulator 3

**NSIS** Next Steps in Signaling

**OSPF** Open Shortest Path First

**P2P** peer-to-peer

**PCE** Path Computation Element

**PEP** Performance Enhancement Proxy

**PMIPv6** Proxy Mobile IPv6

**PPP** Point-to-Point Protocol

**PPPoA** Point-to-Point Protocol over ATM

**PPPoE** Point-to-Point Protocol over Ethernet

**QoE** Quality-of-Experience

**QoS** Quality-of-Service

**RAN** Radio Access Network

**RBDC** rate-based dynamic capacity

**RFC** Request for Comments

**RSVP** Resource Reservation Protocol

**RTT** round-trip time

**SANSA** Shared Access terrestrial-satellite backhaul Network enabled by Smart Antennas

**SAT5G** Satellite and Terrestrial Network for 5G

**SDN** Software Defined Network

**SDR** Software Defined Radio

**SIPTO** Selected IP Traffic Offload

**SNR** Signal-to-Noise Ratio

**SNS3** Satellite Network Simulator 3

**SOHO** Small Office/Home Office

**TBTP** terminal burst time plan

**TCP** Transmission Control Protocol

**TDM** Time Division Multiplex

**TE** Traffic Engineering

**UDP** User Datagram Protocol

**VBDC** volume-based dynamic capacity

**VITAL** VIrtualized hybrid satellite-TerrestriAl systems for resilient and fLexible future networks

**VNF** Virtual Network Function

**VoIP** Voice over IP

**VR** Virtual Reality

**WAN** Wide Area Network

**WLAN** Wireless Local Area Network

**xDSL** X-Digital Subscriber Line

# Chapter 1

# Introduction

Broadband Internet connectivity has become increasingly important during the last years and is nowadays considered "a crucial factor to realize economic growth", which enables the development of new services and applications, as described in the European Digital Agenda (European Commission - Directorate-General for Communication Networks, Content and Technology (CONNECT) 2012, European Commission 2016). Hence, in this agenda the European Commission sets the objective to enable broadband Internet connections of at least 30 Mbit/s to be available to all EU citizens and 100 Mbit/s to at least half of European households by the end of 2020, which seems unlikely to be achieved. According to (Cisco Cooperation 2020) globally only 39% of broadband connections will be at least 100 Mbit/s.

Moreover, current and emerging networks need to be able to cope with a tremendous increase of traffic volume over the next years. For example, by 2023 each person will have on average 3.6 networking devices, compared to 2.4 by 2018 (Cisco Cooperation 2020). Furthermore, IP video traffic will account for the majority of Internet traffic and increase the global Internet traffic in the so-called Zettabyte Era (Cisco Cooperation 2017*b*).

This significantly higher amount of traffic volume will pose a major challenge for operators, especially in rural or other difficult-to-serve areas. For example, in the Back-haul segment, the deployment of nowadays' typically-used technologies, such as optical fibers, microwave radio links or copper connections (Briggs et al. 2010) is prevented by economical constraints (Henkel et al. 2011) leading to many underserved areas.

In order to address this issue, a promising approach is to rely on satellite systems. Bidirectional satellite networks recently regained the attention of both the scientific and the industry communities since the next generation of GEO fixed satellite systems, which is on the horizon, are targeting the Terabit/s aggregated capacity (Evans et al. 2011, Kyrgiazos et al. 2014, Ortíz-Gómez et al. 2018, Chong et al. 2018). These systems will lower the cost per bit significantly mainly by transmitting on different frequencies, i.e. Ka-band, and by using multiple but relatively small spots. Those spots have a size of a few hundred kilometers (instead of e.g a single spot for Europe) and, thus, allow for more spatial frequency re-use (Botta & Pescape 2013). Furthermore, satellite links provide ubiquitous and resilient services as well as broadband coverage, and they are able to deliver high throughput connections and their capacity wherever it is needed, on a very flexible basis.

However, compared to most terrestrial network technologies, both wired and wireless, satellite links have highly different characteristics in terms of latency, burstiness or link stability. Even though approaches exist, which aim at providing triple-play services over broadband GEO satellites, it is de facto impossible to achieve a similar Quality-of-Experience (QoE) perceived by users compared to broadband terrestrial networks, when real-time and interactive application are being used (Chen et al. 2004, ITU-T 2019, ITU-T 2011). The higher latency on satellite connections as a result of the high signal propagation time inevitably means the user's QoE is lower compared to a broadband terrestrial link for many applications, e.g. browsing the web might feel somewhat more sticky.

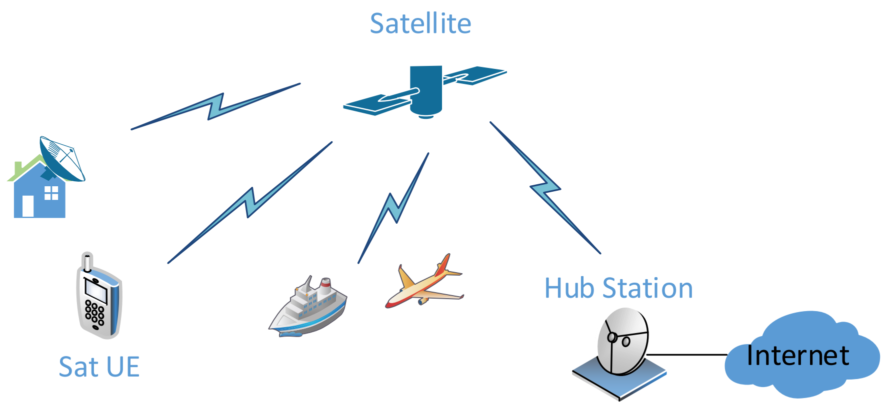
Thus, using solely new satellite systems to provide broadband Internet access in rural and remote areas would not solve the problem since the Quality-of-Service (QoS) and QoE demands, as expected by the users, independent of their location cannot be met due to the high latency, i.e. a user in a remote area expects the same high quality service as a user living in an urban region. Instead, satellite networks needs to be integrated as a native component into existing terrestrial infrastructures, so that they can be used seamlessly for the user and mitigate the effect of a narrow band terrestrial network, as already acknowledged previously (Artiga et al. 2016, Evans et al. 2005, Khodashenas et al. 2019).

Hence, the goal of this study is *to investigate, design and evaluate offloading techniques that allows for shifting network traffic from a narrowband terrestrial last-mile connection to a high bandwidth satellite network* *when beneficial, so that the user's available bandwidth and QoE in underserved areas is improved compared to a sole terrestrial link.* Eventually this leads to network architecture that enables the integration of satellite connections into a terrestrial last-mile network architecture, in order to form a converged satellite and terrestrial network.

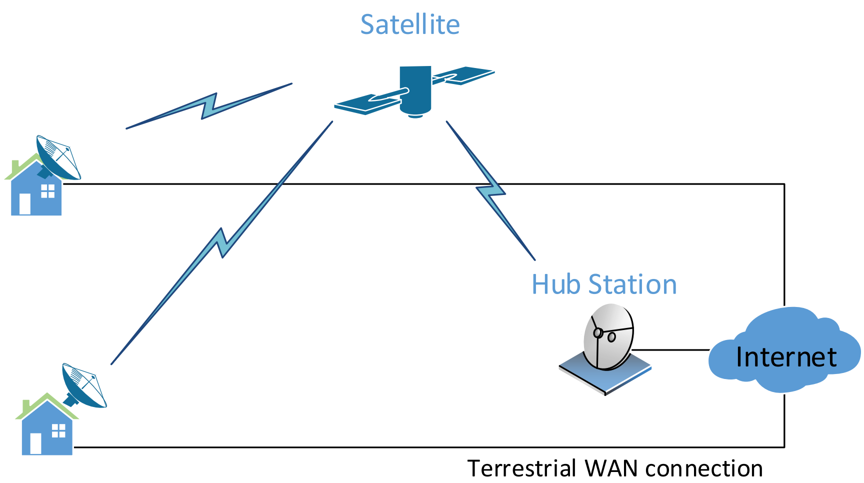
## 1.1 Converged satellite and terrestrial networks

Historically satellite networks are often used to provide connectivity to areas without any terrestrial connection, as shown in Figure 1.1a. These scenarios typically include providing connectivity to very remote households and premises or moving locations, such as a vessel or an airplane (Breiling et al. 2014). Typically, terrestrial networks are connected to one or both edges of the satellite link. For instance, in a remote household typical end user devices are connected locally via a (W)LAN router connected to a satellite modem, which establishes the satellite connection, and on the other edge the operator's core network provide the connectivity to other networks and the Internet. However, it is important to note that the satellite link is the sole connection between the terrestrial edges of the network, so that the traffic cannot be routed differently. Such a setup usually aims at providing general connectivity. It should be noted that many publications consider such a network also as an integrated satellite and terrestrial network, as the satellite network extends a terrestrial network.

Various scenarios and use cases following this setup have already been widely discussed and surveyed in the literature, e.g. (Breiling et al. 2014, Kota & Marchese 2003, Zampognaro 2009, Gayraud & Berthou 2007, Botta & Pescape 2013, Audah et al. 2011, Akyildiz & Jeong 1997). Hence, in this thesis the focus is on scenarios where the satellite network provides an additional connection in parallel to an existing terrestrial one, as shown in Figure 1.1b. A high capacity satellite network supplements an existing terrestrial connection in order to increase the performance of a terrestrial network. Such a scenario makes sense if the terrestrial link performs insufficiently and is not able to cope with the traffic demands. In contrast to the previous setup (see Figure 1.1a), the satellite link provides not the sole connection but an alternative one. Thus, traffic needs to be distributed properly onto the terrestrial and the satellite network.




(a) Satellite network extending a access terrestrial network




(b) Satellite network supplements a terrestrial access network

Figure 1.1: Comparison of satellite link integration options

In the following, the characteristics of both parts of the network and their individual challenges are described in more detail.

### 1.1.1 Satellite network part

In this study, multi-beam bi-directional GEO satellite connections based on Digital Video Broadcasting-Satellite - Second Generation (DVB-S2) (ETSI EN302 307 2013) and DVB Return Channel via Satellite (DVB-RCS) (ETSI EN301 790 2009) are considered. That is, the satellite can be used for up- and down-link traffic. Moreover, even though several aspects discussed in this work are also applicable for Mobile Satellite Services (MSS), Fixed Satellite Service (FSS) networks are assumed, given that MSS networks have to deal with handovers and other mobility related issues, which are out of scope of this thesis.

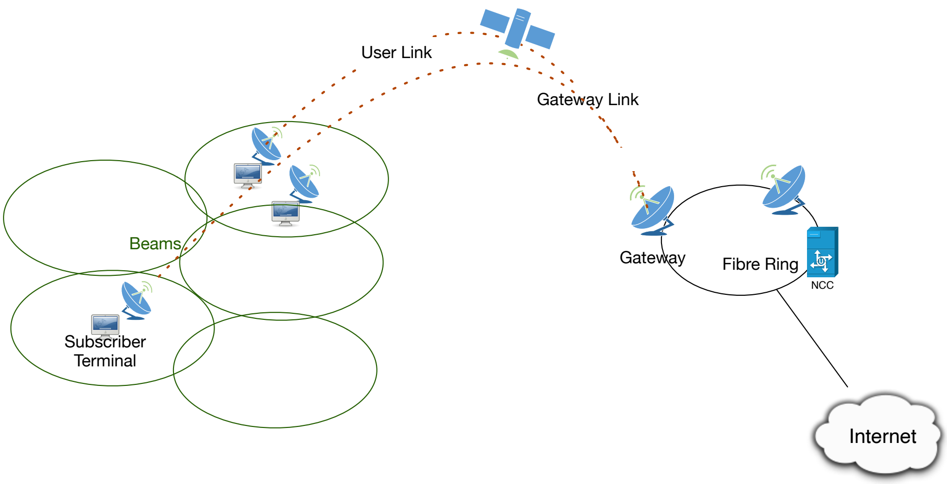



Figure 1.2: Typical architecture of a satellite network

The typical architecture of a satellite network is depicted in Figure 1.2. It usually consists of one or multiple so-called Gateways or hub stations, which provide on the one hand connection to other networks, i.e. the Internet, and on the other hand establish the link to the satellite, which are called Gateway Links. The Gateways are also interconnected by high speed links, so that they can fail-overs can be done between Gateways, e.g. in case of heavy rain at one Gateway location. The so-called User Terminals or subscriber terminals form the other edge of the satellite network. They receive the downlink data from the satellite, which is referred to as user link, but also send back the uplink packets. The end user clients, i.e. the User Equipment, such as Laptops or Smartphones, are connected by Local

Area Network (LAN) technologies, e.g. Ethernet or Wireless Local Area Network (WLAN) to the User Terminals. Usually co-located with one of the operator's Gateways is the network control center (NCC), which has, among others, the responsibility to assign time slots to the User Terminals when they can send data. This is only required for the uplink, as the User Terminals are multiple distributed senders, which compete for the same media.

Moreover, it is assumed that the satellite network provides a transparent Internet Protocol (IP) connection. Since satellites are primarily used for broadcasting content, while terrestrial networks rather transport unicast data, both worlds need to be brought together. That is, the IP domain and its concepts and mechanisms, which are usually used on terrestrial links, must become usable on satellite networks. In order to do this, for instance DVB Generic Stream Encapsulation (DVB-GSE) (ETSI TS102 606-1 2014) has been standardized by the European Telecommunications Standards Institute (ETSI). DVB-GSE provides an efficient encapsulation method for IP packets over variable length Layer 2 fragments, which are then directly scheduled on the physical layer into base-band frames and, hence, make the satellite connection transparent to IP.

### Terabit/second satellite systems

In order to lower the cost per bit on a communication system, one option is to increase the throughput of system. To increase the throughput on a satellite system, the following factors are relevant:

- available bandwidth (MHz),
- spectral efficiency (bits/Hz), i.e. the amount of data transmitted error-free in a given channel bandwidth to a user,
- the amount of beams, in order to allow for frequency reuse.

As described in (Kyrgiazos et al. 2014) GEO satellite systems have operated for several years in the C (3.7 GHz - 6.4 GHz) or Ku (10 GHz - 18 GHz) bands. With the advent of the Ka band, i.e. 18 GHz - 30 GHz, more spectrum becomes available, so that eventually more bandwidth per receiver is available (Gayrard 2009). Furthermore, better carrier to noise ratios that


Figure 1.3: Different beam sizes

allow for more efficient modulation and coding schemes as well as emerging new waveforms on the physical layer foster an increased spectral efficiency.

Current satellites, such as die ViaSat-2 operating in Ka-Band, already provide more than 300 Gbit/s system capacity. ViaSat-3, which is expected to start service in 2022, aims at delivering 100 Mbit/s residential Internet service (Giambene et al. 2018).

Moreover, the coverage footprint of a satellite is also reduced from so-called global or regional beams to spot beams. That is, as shown in Figure 1.3, instead of using a single beam that covers, e.g. Europe, a multiple smaller spot beams is applied, so that frequencies can be reused. While large beams are appropriate for broadcast applications, more and smaller beams are more optimal for the individual content requested by nowadays users. The size of the beams are decreased from the order of 1000s of km to 100s of km. The higher frequencies in the Ka band also ease the architecture of smaller beams. By using these optimizations, the capacity of current satellite systems has already increased from approximately 10 Gbit/s up to 100 Gbit/s. Current Ka band satellite systems generate already 80-100 spot beams to e.g. cover Europe. To provide Terabit/Second High Throughput Satellite system, 200 and more beams towards the User Terminals are required.

Moreover, in order to have more available spectrum, on the feeder link, i.e. the connection between the hub stations and the satellite, even higher frequencies in the Q/V bands are considered. In order to mitigate impact of rain and other factors that influence the attenuation, efficient adaptive coding and modulation (ACM) becomes necessary.

In order to increase the efficiency even more and to achieve the terabit/second satellite system, upcoming optimization of the air interface, such as turbo codes, are required as well. To conclude, as evaluated by (Kyrgiazos et al. 2014, Ortíz-Gómez et al. 2018, Chong et al. 2018) terabit/second satellite systems are on the horizon.

### High latency on satellite connections

As already mentioned, a major difference between satellite and terrestrial systems is the latency. More precisely, GEO satellite links have a high fixed latency, as depicted in Figure 1.4. Dynamic latency consists of the time required to serialize data, process a packet at a network entity, potential queuing and buffering delays as well as the time required to get access to the medium. In contrast to that the fixed part of the overall latency is the actual signal propagation time that a packet experiences in any case when being transported over a network medium. It is a physical characteristic, mainly restricted by the speed of light. Moreover, the dynamic latency highly depends on the available capacity on the link and can be reduced and controlled by prioritization, packet dropping or admission control to avoid congested links. Hence, in terrestrial networks low latency and low jitter values can be ensured by just avoiding congested links and, thus, preventing extensive queuing (Xiao & Ni 1999).

GEO satellites operate at a height of 36 000 km. Therefore, the fixed latency on a satellite link is in the order of magnitude of a few hundred milliseconds leading to a high overall latency even if sufficient capacity on the satellite link is available.

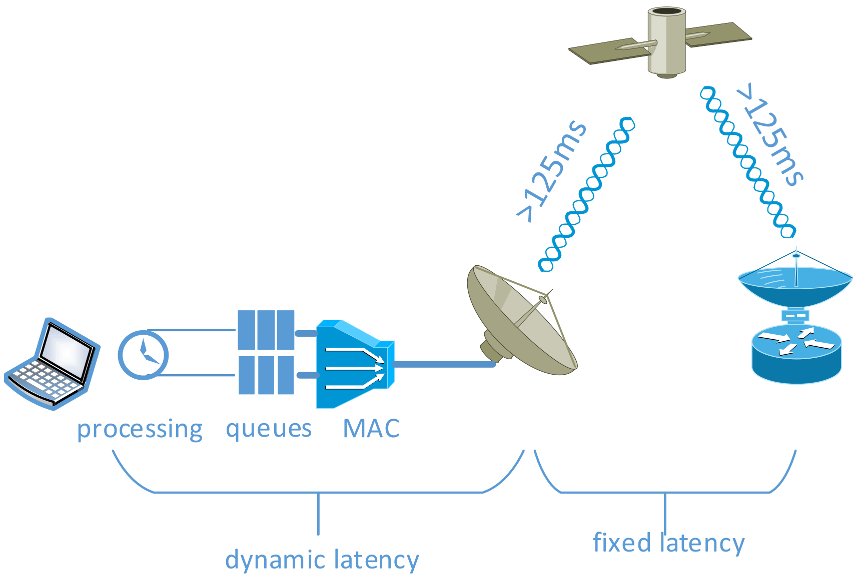



Figure 1.4: Latency model of a GEO satellite link

Obviously, the impact of these fixed latencies on the total transmission time depends on the amount of data to be sent. That is, during the transmission of a large amount of data the high fixed latency has a proportionally minor impact compared to the transmission of small amounts of data due to the length of the overall transmission. Moreover, given that the higher bandwidth of the satellite link, a theoretical threshold can be calculated when the overall transmission time is faster in spite of the high fixed latency.

This is illustrated by an example depicted in Figure 1.5: The theoretical time required to transmit data over different hypothetical links is calculated. Up to a certain threshold, which depends on the link speed, the total transmission time of a narrow-band hypothetical terrestrial link with negligible signal propagation time is faster than a high bandwidth satellite link. For an exemplary 2Mb/s terrestrial link this threshold is already reached when 60 kB are in the queue. For a 10 Mbit/s terrestrial link around 500 kB of queued data is required so that the satellite outperforms the terrestrial link. It should be noted that the MAC overhead is not taken into account in this figure and that 250 ms and 20 ms latencies are assumed for the satellite and

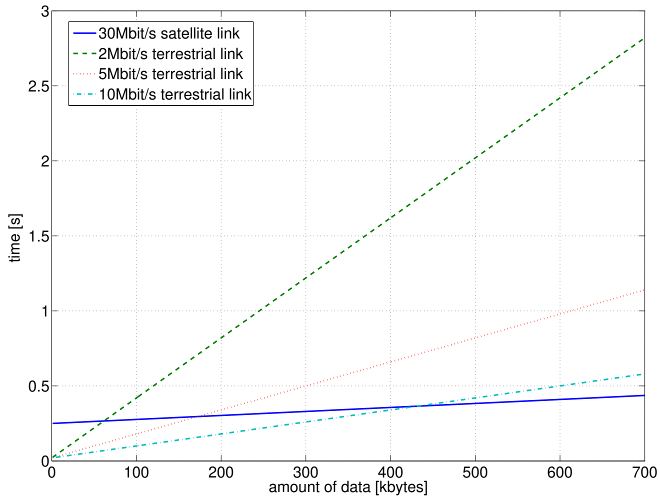



Figure 1.5: Transmission time comparison of different terrestrial link speeds and 30 Mbit/s satellite

the terrestrial link, respectively.

This issue is further illustrated by the following example: A messenger pigeon carrying a 4 GB USB drive over 63 km takes around two hours (*The Pigeon versus the Computer: A Surprising Win For All* n.d.). This leads to a bandwidth of approximately 4 Mbit/s, which might be even more than a current X-Digital Subscriber Line (xDSL) connection in some rural areas. Furthermore, while the QoS provided by this kind of network is most likely sufficient for large file transfer applications, it is absolutely unacceptable for any kind of application requiring interactivity.

As can be seen from this extreme example, but also from Figure 1.5, bandwidth and latency constraints need to be looked at together in satellite environments. While the high capacity of satellite links can be a benefit for exchanging large amounts of data, the additional delay has a negative impact on the QoS and on applications provided via satellite that are not tolerant against high latencies. An example of these are real-time or interactive applications. Furthermore, protocol state machines and specific functionalities may also need to be adopted to cope with the higher latency. For example, the (standard) congestion control mechanisms of the Transmission Control Protocol (TCP)(Allman et al. 2009) rely on the round-trip time (RTT) and thus the TCP throughput suffers significantly from the high latency on the satellite links.

In order to mitigate this, specialized TCP algorithms, such as (Papadimitriou et al. 2011), or TCP-Performance Enhancement Proxies (PEPs) (Border et al. 2001) are commonly used. PEPs are TCP proxies, which break the TCP end-to-end connection at the satellite modem and/or the hub station, so that the satellite portion of the end-to-end path is separated from the rest. That is, instead of a single end-to-end TCP connection up to three intermediate connections are established, namely between the end host initiating the connection to the first edge of the satellite link, between the satellite modem and the hub, and finally between the other edge of the satellite link the and the destination host. This allows for running an adapted TCP protocol or a completely different transport layer protocol optimized for the high latency.

#### Medium access coordination and resource reservation

Another major difference in satellite networks is the coordination of the uplink connections (towards the core network), in order to provision QoS. In satellite networks on the downlink, only a single sender, namely the hub station, exists, while on the uplink potentially many highly distributed senders access the shared medium. That is, the senders compete for transmission on the same frequency. One way of solving the problem is to assign fixed sending slots and, thus, capacity to a given User Terminal, regardless of the actual required capacity at the particular User Terminal. This approach is called Fixed Assigned Multiple Access (FAMA). While such an approach renders a Media Access Control (MAC) protocol virtually needless and therefore significantly reduces the overhead, it leads to wasted capacity and inefficient use of the available resources in situation with arbitrary and unpredictable traffic patterns.

In order to avoid this, the DVB Return Channel via Satellite - Second Generation (DVB-RCS2) standard (ETSI TS301 545-2 2014) specifies transmission-slots during which they can transmit. These are assigned to each sender by a central entity called NCC located at the hub station. The NCC periodically broadcast the assignment of transmission-slots, which are either based on dynamic requests from a sender depending on the current traffic demands. This is called Demand Assigned Multiple Access (DAMA). Depending on the traffic, a satellite terminal can request different categories of capacity. For services requiring a higher priority or real-time services typically constant rate assignment (CRA) or rate-based dynamic capacity (RBDC) are used, which provide capacity guarantees. In contrast, for lower priority and best-effort (BE) traffic, volume-based dynamic capacity (VBDC) and free capacity allocation (FCA) capacities are requested. Those provide fewer guarantees. The NCC grants permissions to transmit by sending a terminal burst time plan (TBTP). Since those requests and the allocation responses traverse the satellite link as well as the actual data, the overall latency increases even further. Thus, higher priority, real-time and more interactive traffic should be mapped onto constantly assigned slots. However, capacity, which is fixed and assigned to a station, is wasted if the station has nothing to send in a particular slot. Hence, there is a trade-off between the fixed allocated capacity to certain stations and the capacity dynamically allocated.

Moreover, to properly prioritize traffic on the satellite link, the QoS mechanisms of layer 3 and layer 2 need to work jointly together. The approach described in (ETSI TS101 545-3 2014) relies on the Differentiated Services (Diffserv) model (Blake et al. 1998) to classify and mark packets by using the IP Differentiated Services Code Point (DSCP) field. The packets of the same class are assigned to so-called Behavior Aggregates. All packets belonging to the same Behavior Aggregate share the same network behavior. That is, they experience the same queuing and scheduling treatment. Hence, a Behavior Aggregate can be compared to a Bearer in an LTE network or a Multi Protocol Label Switching (MPLS) Label-Switched Path (LSP). Depending on the Behavior Aggregate, the DAMA process responsible for the medium access control requests and allocates capacity. A concrete implementation for QoS provisioning over DVB-S2 is presented for example by (Angeles Vazquez Castro & Vieira 2012). By concatenating different round-robin schedulers and also considering the different Modulation and Coding Scheme (MCS) the authors implement fairness policies while giving QoS guarantees. Different other approaches exist, e.g. (de la Cuesta et al. 2013, ETSI TS101 545-3 2014, ETSI TS102 462 2006, ETSI TS101 545-5 2014).

It should be noted, however, that the DVB-RCS2 standard does define the process of how the User Terminals asks for capacity, but not define how different traffic classes are prioritized or otherwise treated and how the NCC distributes the available capacity onto the competing requests. Hence, all implementations in solutions on the market are typically proprietary.

### 1.1.2 Terrestrial network part

The network architecture of the terrestrial network part considered in this study is a typical modern all-IP network. That is, on the data plane the regular IP protocol is used. It is assumed that clients are connected to a Customer premises equipment (CPE) that corresponds to the User Terminal in the satellite network. The connection to the operator network is assumed to be realized via a tiered network architecture. That is, a last mile tier and one or multiple aggregations networks form the backhaul part to the operator's network. There the connection towards the Internet is established. This architecture is depicted in Figure 1.6.

It should be noted that for the reminder of this study the exact architecture of the terrestrial network is not significantly relevant, as long as it is an all-IP network.

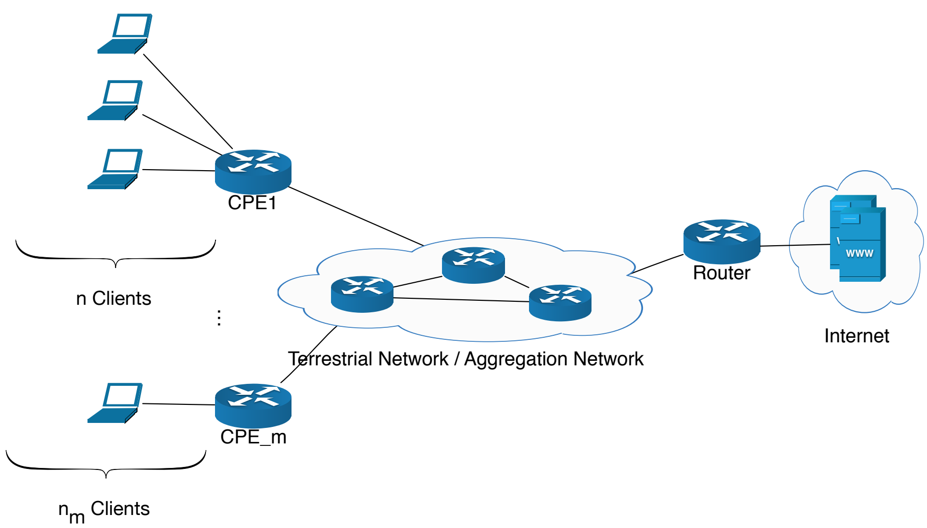



Figure 1.6: Terrestrial WAN architecture

Furthermore, it is assumed that the capacity in terrestrial network, which is available at the edge of the network, is rather limited, as commonly occurs in underserved areas. A cost benefit analysis conducted by the EU BATS research project (EU BATS Project 2014), found out that a converged satellite and terrestrial scenario can play a role if the terrestrial fixed line speed for a household is between 0-8Mbit/s. Above that threshold, the benefit provided by the additional satellite link is rather low, since the performance of the terrestrial link increases.

### 1.1.3 Use-Case

In the following an exemplary scenario is presented, which benefits from an additional satellite link, namely a remote household, connected with a poor last mile connection, such as a slow xDSL connection, as previously described.

Many areas all over the world, particularly in rural and remote regions, lack terrestrial high-speed broadband connections, which are able to cope with future traffic demands. Moreover, a low population density and large areas often render upgrading existing narrow-band terrestrial connections economically unappealing. In this case, the fast and flexible deployment characteristics of satellite connections can be exploited to establish additional satellite links, which do not replace but rather supplement the terrestrial connection (Peters et al. 2013), as shown in Figure 1.7.

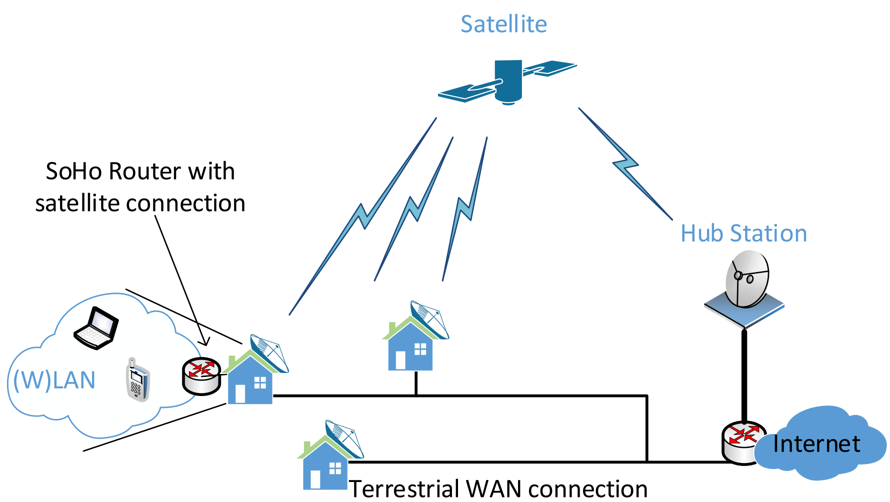



Figure 1.7: Overview remote household scenario

Individual houses and premises are equipped with an additional satellite connection in order to increase the available overall capacity. Both, the terrestrial Wide Area Network (WAN), such as a Digital Subscriber Line (DSL) or cable link, as well as the satellite connection terminate at a typical Small Office/Home Office (SOHO) router connected to or equipped with a satellite modem. End user devices are connected to this router using typical (W)LAN connections, such as Ethernet or IEEE 802.11 WLAN.

It should be noted that for example in Germany in rural regions around 70% are connected with 30 Mbit/s and only 63% with more than 50 Mbit/s (German Federal Ministry for Economic Affairs and Energy 2018). Similarly, in the USA in 2016 less than 70% of the remote population is equipped with fixed broadband connectivity with a higher bandwidth than 25 Mbit/s (Federal Communications Commission 2018).

## 1.2 Challenges for converged satellite and terrestrial networks

In the following specific challenges associated with converged satellite and terrestrial networks are discussed.

### 1.2.1 Path selection and offloading decision

When a satellite network supplements terrestrial infrastructures the network topology becomes more complex, since alternative parallel paths are available between two nodes in the network. However, unlike parallel links in plain homogeneous network, the different links in a converged satellite and terrestrial environment differ extremely in their characteristics, i.e. the difference is not only limited to available capacity, as described in the previous section. These characteristics are typically not considered by regular routing protocols such as Open Shortest Path First (OSPF) or Border Gateway Protocol (BGP) which in most IP networks are responsible to form a loop-free IP-WAN network. The metrics used are usually only the amount of hops or weights based on the capacity of a link.

An example is depicted in Figure 1.8. Two different paths exist between the User Equipment and the Internet. Hence, for the uplink traffic what needs to be decided is if the satellite or the terrestrial link should be used and the same decision needs to be made for the downlink traffic. Given that, traffic with the same source and destination IP address might take different routes depending on its requirements and the current situation in the network. Even though at first glance this might seem trivial given the amount of matured IP routing protocols, the unique characteristics of satellite links tremendously increase the complexity. For example, given the characteristics of a GEO satellite link, even in non-congested networks a real-time and interactive application, e.g. Voice over IP (VoIP), will rather use the terrestrial link due to the latency aspects, whereas a video stream benefits from the high capacity on the satellite link yielding potentially to a higher video resolution.

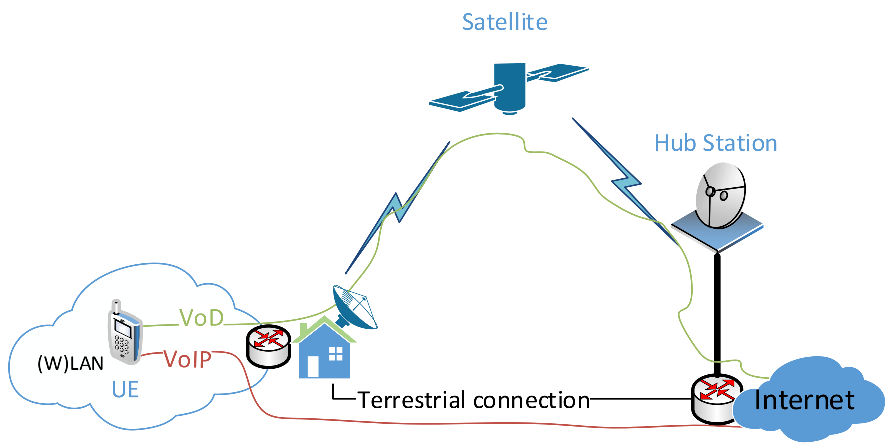



Figure 1.8: Data paths in a integrated satellite/terrestrial network

As implied in this simplistic figure, ideally the path selection is not static or simply based on destination IP addresses like regular IP routing, but rather dynamic and on a fine-granular basis, since traffic of different services exchanged between the same source and destination might need to be routed differently. That is, the decision if certain traffic flows should be offloaded to the satellite is potentially based on various factors, which are depicted in Figure 1.9.

Firstly, the characteristics of traffic needs to be considered, i.e. are the

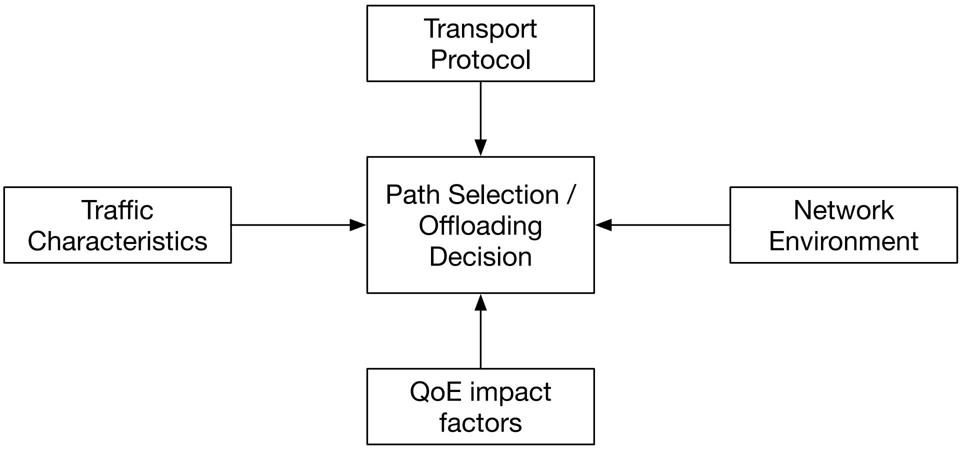



Figure 1.9: Factors impacting the path selection

packets generated burstily or uniformly, how much amount of data is generated in which period of time? If the amount of data ready to be sent exceeds the capacity of the terrestrial link, the packets needs to be queued and the dynamic latency increases. Thus, it might be faster to send the data via the satellite connection despite the higher fixed latency. Secondly, the most relevant network related factors impacting the QoE might be taken into account. For example, for regular web-browsing it is beneficial to reduce the Page Load Time and, hence, simply select the path with the shortest delivery time. In contrast, for video streaming it is important to have a high bandwidth, since modern video streaming methods, e.g. Dynamic Adaptive Streaming over HTTP (DASH) adjust the resolution depending on the throughput. Thirdly, the behavior of the used transport protocol becomes important and might impact the path selection. For example, while User Datagram Protocol (UDP) is basically not affected by reordering, packet loss or transmission time, TCP performance is highly impacted by these issues. Finally, the network capabilities, such as current link utilization, available capacity or fixed latency, has to be considered so that it can be determined when a link is overloaded and the dynamic latency increases.

Hence, it needs to be decided when and which portion of the traffic should be offloaded to the satellite, in order to ultimately increase the end user's QoE.

### 1.2.2 QoS requirement identification

Given that some traffic is more tolerant to latency than other traffic and, thus, more suitable to be sent over a satellite connection, identifying traffic in terms of its application type or QoS requirements is essential in a converged satellite and terrestrial network. If, for example, a real-time service requires a guaranteed latency below 100 ms, in order to achieve a good QoE, the service might not be usable over a GEO satellite connection, regardless of the available bandwidth on that link. This is a major difference compared to pure terrestrial networks, where capacity management and congestion avoidance might be sufficient, without the need for determining the actual traffic QoS requirements, since high latencies or jitter values as well as packet losses can often be avoided by preventing a congestion in the network.

It should be noted, however, that this becomes increasingly important in today's operator networks independent of the satellite network integration, since many application providers require special services for their applications, particularly rich media and pervasive video applications, which have enormous QoE expectations as revealed by Internet Engineering Task Force (IETF) activities (Fan et al. 2014). Moreover, in order to perform adequate TE, knowledge of traffic requirements, such as required bandwidth or latency demands is also worthwhile.

However, becoming aware of the these is not a trivial task. Various methods are available to identify applications or application types just by monitoring the traffic. The examination of the transport-layer port numbers, which has been the preferred method for several years, has proven to be inaccurate in networks (Bernaille et al. 2006), e.g. due to re-usage of well-known ports or higher layer tunneling. This leads to more complex methods, such as Deep Packet Inspection (DPI) or machine learning approaches (Nguyen & Armitage 2008). An alternative to identify IP traffic by monitoring and analyzing it, is the usage of meta-information sent by the application or application-related protocols. Such meta-information could contain application type or even the traffic QoS requirements of the application. Examples of such approaches, are the IP DSCP bits, Resource Reservation Protocol (RSVP) (Braden (Ed.) et al. 1997) or Next Steps in Signaling (NSIS) (Manner et al. 2010). Even the usage of different Access Point Names (APNs) for different applications can serve the same purpose.

Given the complexity of these, the challenge is acknowledged as relevant, yet it is out of scope of this work.

### 1.2.3 Network architecture

In order to actually use the additional satellite link and send a portion of traffic through it, the routing behavior of the involved network devices needs to adapted. As the traffic is ideally offloaded on fine-granular basis, e.g. on a flow or cluster of packets basis and not like regular IP routing based on destination IP, the routing needs to operate on the same level. That is, if a decision to offload a certain portion of the traffic onto the satellite link has been made, it needs to be ensured that this decision is being executed in the network.

Various means exist to modify the routing and switching behavior in a network, starting from typical IP-related methods, such as NAT, Mobile-IP or extended routing protocols, through to novel networking concepts, such as Software Defined Network (SDN) or Proxy Mobile IPv6 (PMIPv6).

Moreover, the location and function of the interconnection points between satellite and terrestrial networks also need to be discussed and evaluated, in order to route traffic effectively through both networks depending on the outcome of the path selection decision. That includes, how tight or loose both networks are coupled with each other.

## 1.3 Differentiation of existing commercial solutions

Several regular satellite Internet Service Providers (ISPs) exist that provide Internet connectivity in various, mainly rural ares, e.g. Filiago[1], Tooway[2] or SES Broadband[3]. These ISPs provide general connectivity to customers solely over a satellite connection, as explained in the previous chapter (see Figure 1.1a). Since they are based on novel technologies, e.g. DVB-S2 and DVB-RCS, Ka-Band, etc., the available capacity is comparatively high. However, as they are relying on GEO satellites their customer will experience the inevitable latency caused by the high altitude of the satellites.

It should be noted that this is different to setups where the satellite connection is only used for traffic towards the end user, which is referred

[1] https://www.filiago.de
[2] http://www.tooway.co.uk/
[3] https://www.ses.com/networks/telecom/broadband

to as downlink, and the terrestrial connection for traffic towards the operator's network, i.e. the uplink. These setups, which have been commercially available, only require a unidirectional satellite connection and distribute the traffic statically based on its direction. Obviously, the traffic is trivial to route, as there is no decision to take. However, the user will still experience a high latency on the downlink. Moreover, the uplink/downlink ratio in current networks is also changing. Due to the increasing usage of social networks and cloud services, it can be assumed that the amount of uplink traffic will also increase, so that in terrestrial narrow-band networks not only the downlink might be not able to fulfill the demands but the uplink too.

Moreover, Deutsche Telekom for example is offering its triple play package with regular DSL and a satellite connection. The satellite connection, however, is only used for the television broadcast part. Similarly other approaches that are often sold as e.g. Hybrid DSL are typically combining a terrestrial and a mobile network, i.e. LTE, but not a satellite link. Even though these approaches partially share similar challenges, like distributing traffic onto available connections, the characteristics in terms of latency of both connections in such hybrid approaches differ not in the same order of magnitude than a satellite connection. Hence, the approaches used in these solutions are only partially applicable to converged satellite and terrestrial networks.

It should also be noted, that Internet and data services can obviously also be provided by Medium Earth Orbit (MEO) or Low Earth orbit (LEO) satellites, which are located on a lower orbit than GEO satellites. The main advantage of those systems is a much lower latency, due to the lower altitude compared to GEO satellites (see 1.2). LEO satellites orbit the earth between 500 and 1600 km. However, the lower altitude reduces the area of the earth the satellite can cover. Hence, typically so-called constellations, i.e. a network of satellites, are required, to cover larger areas. Moreover, as the position of these satellites is not fixed related to the earth, handover mechanisms are needed to maintain connectivity when a satellite gets out of sight and another one has to take over. Hence, commercial operators using LEO satellites to provide connectivity, such as Iridium[4], usually target areas without any existing infrastructures. Their goal is to provide connectivity virtually everywhere and independent of local infrastructures as well as re-

[4]https://www.iridium.com/

silient to all kinds of local disasters. Given that, such approaches are mainly used for use cases that require limited bandwidth, such as Internet of Things (IoT) or voice communication. Moreover, other approaches like OneWeb[5] provide in total over 10Tbps. However, since they are not GEO, their satellites cover quite often oceans during their orbit around the earth, leading to a waste of capacity, which makes the business case for LEO constellations at least questionable (EU Project: Broadband Access via Integrated Terrestrial & Satellite Systems 2016).

Obviously, as soon as a LEO satellite constellation can provide a similar high throughput to emerging GEO satellites, converged satellite and terrestrial networks as considered in this work become obsolete. However, even though it might be technology-wise feasible to create such a constellation with a sufficient amount of satellites, it can be assumed that such an approach would be significantly too expensive to be realized by any operator, as explained previously.

## 1.4 Summary and Problem statement

As can be seen, GEO satellite networks have unique characteristics and behavior compared to many terrestrial networks. Particularly their high fixed latency is a difficult to address challenge. However, the ability of satellite networks to provide connectivity virtually everywhere makes them very suitable for rural and other underserved areas. Due to recent developments in the satellite domain leading to high capacity connections renders them as one option to address the goals of the EU's Digital Agenda and other initiatives to provide at least 30 Mbit/s for each household and ubiquitous broadband.

Unfortunately, the quality of certain services suffers tremendously from the unique characteristics of satellite links in particular the high fixed latency. First and foremost real-time and interactive applications will be perceived quite badly by the user compared to a broadband terrestrial connection with significantly less latency. Given that, using solely satellite connections to provide the demanded bandwidth will lead to reduced QoE. Instead, integrating high-capacity satellite links into existing narrow-band terrestrial networks, *to form a converged satellite and terrestrial network* is

[5] http://onewebsatellites.com

a promising approach to provide high quality connectivity in underserved areas. However, in order to realize such an approach, sophisticated traffic offloading techniques need to be available that shift traffic from terrestrial connections onto a satellite, while considering the specifics of satellite links.

Given that, as well as the aforementioned challenges, the following research questions can be derived:

- What are the key functional building blocks to enable offloading in converged satellite and terrestrial networks?
- How to decide when and which traffic should be offloaded to the satellite network?
- How can the architecture of a converged satellite and terrestrial network look like?
- How to enable the usage of satellite links for the traffic that has been selected for offloading?

By addressing these research questions, this thesis aims at designing and evaluating a traffic offloading approach for converged satellite and terrestrial networks, so that the end users' QoE in undeserved areas can be increased. In order to achieve this, the following objectives are defined:

- to investigate and to define key functional building block required to enable offloading in converged satellite and terrestrial networks,
- to evaluate existing network architectures from other domains with respect to their applicability in the considered scenario,
- to design and to evaluate different traffic offloading methods that can be used in the considered scenario,
- to design and to evaluate a network architecture that allows for executing offloading decisions in converged satellite and terrestrial networks.

Given these objectives, the reminder of this thesis is structured as follows: Firstly, in this Introduction chapter, the relevant scenarios as well as the scope of this thesis is presented. This includes a differentiation to existing commercial approaches and to other forms of satellite and terrestrial network integration, given that this term is not clearly defined in existing work. The second and third chapter primarily describe existing work. While the second chapter extensively evaluates existing network architectures in other domains, such as mobile networks or load balancing in IP networks, the third chapter surveys the scientific literature in the specific context of converged satellite and terrestrial networks. This differentiation is made to show different viewpoints: The existing network architectures in other domains are analyzed primarily with the respect to their applicability to control, steer and offload traffic in the network, as well as their feasibility to be used in the given scenario. In contrast to that, the third chapter presents contemporary research projects and other scientific work in the domain of satellite and network convergence. It should be noted that in the second chapter also key functional building blocks are defined, in order to structure the problem space and to foster a systematic analysis. In the fourth chapter, the used methodology throughout this thesis is presented. This chapter also outlines the methods with which the evaluation of different offloading approaches and network architecture designs, which follows in the next chapters, has been done. Particularly, it described how meaningful results are obtained by the used simulation approach. Accordingly, in the fifth chapter concrete offloading approaches are designed and evaluated by simulations. Finally, in the sixth chapter, the outcome of the fifth chapter is used as input to design a potential network architecture for a converged satellite and terrestrial network that enables the offloading of traffic onto the satellite. Finally, the thesis is concluded and an outlook for future work is given in the seventh chapter.

# Chapter 2

# Related work in other network domains

This chapter presents related work that has relevance to converged satellite and terrestrial networks. More precisely, existing network architectures from other domains are analyzed with respect to their suitability to be used in the considered scenario. This seems reasonable, given that several protocols and approaches exist in other areas that are used to offload traffic from a "primary" network to an alternative network in order to relieve the primary network capacity-wise or to enhance the QoS. Such approaches differ mainly in terminology but share similar ideas and challenges. For example, traffic offloading is also considered in mobile networks. In this regard, firstly mechanisms and techniques that might be used in any IP network are discussed, since "finding a proper path in the network" is the task of IP and IP routing protocols in the first place. These include typical TE approaches, (flow) mobility mechanisms and enhancements for IP routing protocols. Secondly, approaches that are used in mobile networks are explored. Given that radio spectrum is a scare resource in mobile networks, many concepts exists that offload traffic from the primary network to alternative networks. These approaches are discussed here. Thirdly, emerging SDN concepts are shown, as these seem to be a promising solution to allow for a high flexibility and a high dynamic in modern networks.

In order to structure the discussion and to systematically analyze the various approaches, at the beginning of this chapter functional building blocks are identified and described, which encompass the functionalities to achieve the objectives defined in the previous chapter. All considered existing protocols and approaches are mapped against these generic functional building in order to evaluate their applicability for the considered scenario. This extensive analysis is conducted, in order to analyze the applicability of existing approaches or parts of it in the satellite domain and the considered scenario.

Finally, this chapter also elaborates on changes in nowadays Internet traffic, such as the level of encrypted traffic or its overall composition. This serves as an input for the evaluation conducted in Chapters 5 and 6. It should also be noted that the current research activities in the satellite domain are presented in the Chapter 3.

## 2.1 Functional building blocks

The challenges presented in Section 1.2 can be, and to a certain extent already have been, addressed and solved independently for other use cases than integrating satellite networks with terrestrial ones. For instance, distributing traffic onto multiple network paths is dealt with extensively in network load balancing research activities (Prabhavat et al. 2012) or productively used in modern Equal Cost Multipath (ECMP) routing or TE techniques. Similarly, protocols are available that exploit multi-path networks, e.g. Multipath Transmission Control Protocol (MPTCP) or offload traffic in mobile networks, such as Selected IP Traffic Offload (SIPTO).

Moreover, the challenges are also shared among multiple disciplines of computer network related research. For example, they can be seen as being a port of a network routing problem, a load balancing problem or to a certain extent even a network mobility problem. Furthermore, depending on the actual domain, different approaches are implemented on different layers in the Open Systems Interconnection (OSI) model.

However, even though most of the challenges identified previously seem to be solved individually in terrestrial networks, using (some of) them jointly to form a converged satellite and terrestrial networks, as described in Section 1.1, is still a challenging task, which has not yet been realized. This is mainly due to two reasons: firstly, not all mechanisms, which are designed for terrestrial networks, can cope with the specifics of satellite links and secondly, a holistic architecture is missing, which provides clear functional blocks and well-defined interfaces among them.

In order to do that, it is necessary to define key functional building blocks, which logically structure the problem space. This is required to break down and compare existing approaches, which often are a monolithic block.

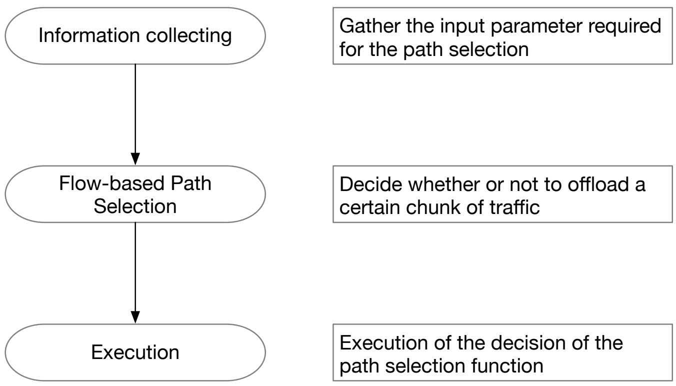



Figure 2.1: Functional building blocks

Given the identified challenges, the following required functions can be defined, as depicted in Figure 2.1. First and foremost, a path selection function is essential. It needs to decide whether to send a portion of the traffic via a non-default route or not. More precisely, in the case of a converged satellite and terrestrial network, whether to offload a certain chunk of traffic to the satellite link or not. In order to perform this decision certain input parameters are required. Depending on the concrete approach, this might be the availability of an alternative link, its capacity, but also the kind of traffic or its priority. Hence, a functional block collecting this information and providing it to the path selection functional block is required. Finally, a functional block executing the decision of the path selection function is needed, e.g. by modifying routing tables in a decentralized approach or or by signaling the decision throughout the network, when centralized approaches are used.

In the following the presented techniques and protocols of the related work are matched against these functional building blocks, in order to analyze the suitability of these for the scenario considered in this thesis.

## 2.2 IP-related techniques

For many years now different mechanisms and technologies have been added to the IP ecosystem perform path selection, to deal with QoS, guaranteed services as well as to perform other forms of TE or mobility. Figure 2.2 gives a non-exhaustive overview of this environment, which is discussed in the following. It should be noted, that as ”standard IP” routing, regular IP protocol, which is used jointly with a typical Interior gateway protocol (IGP) routing protocol, such as OSPF, is considered.

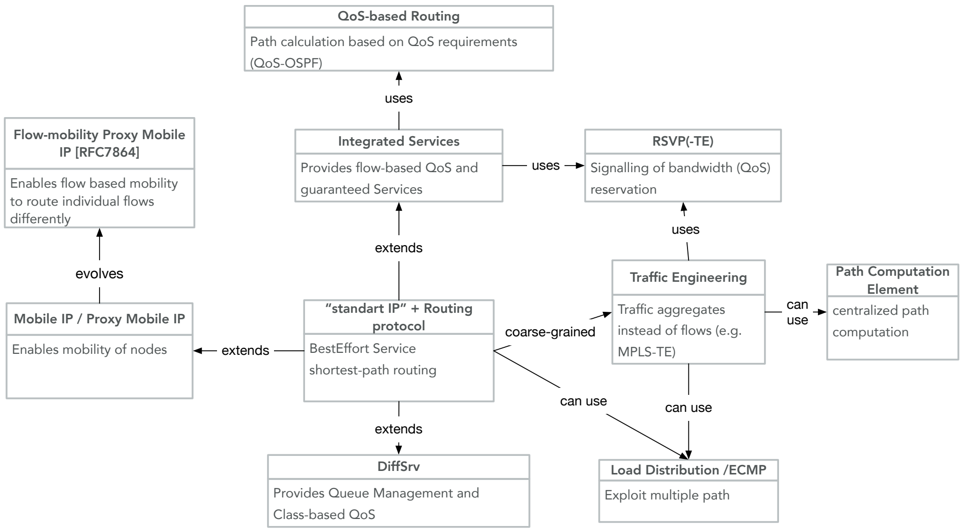



Figure 2.2: IP TE environment

Also this ”standard IP” routing can be matched against the functional building blocks as presented in Section 2.1, it can be seen that all three functional building blocks exist. That is, a routing protocol gathers information if a link is up or down and selects the path between based on the shortest hop count on a per-packet basis. Finally, by modifying the routing tables of the nodes in the network this path selection decision is executed. However, it is obvious that even though all building blocks are presented, they do not provide sufficient functionality, so that this approach is not suitable for the considered scenario as explained in Section 1.2.1.

Hence, extensions to ”standard IP” routing, which are presented in the following, are discussed.

### 2.2.1 QoS-based routing and Integrated Services

The Integrated services (IntServ) architecture (Braden et al. 1994) aims at providing QoS guarantees in a network on a per-flow level. In order to do that the RSVP (Braden (Ed.) et al. 1997, Wroclawski 1997) is commonly used to signal resource reservation along the path. Applications on the end hosts systems (or intermediate router) can issue RSVP requests to gain specific qualities for a certain flow. These requests contain a flow description, e.g. source IP and port, and the desired QoS, and are routed from source to destination along the path calculated by the routing protocol. Each intermediate RSVP-enabled router on the path can then accept or reject the RSVP depending on its policies and available resources. Further details on RSVP can be found in (Pana & Put 2013).

In order to take care of the information distribution and path selection block IGP routing is used. One example in this regard is Quality-of-Service Open Shortest Path First (QoS-OSPF) (Apostolopoulos et al. 1999), which is an extension to OSPF (Moy 1998). It has been specified already more than two decades ago and was defined to enable QoS in OSPF. It performs constraint-based routing, which in contrast to regular IP routing aims at finding a path in the network which fulfills certain bandwidth requirements, such as a minimum available bandwidth. This extension aims at enabling QoS routing in IP networks by distributing, along with the link states in the network, the available bandwidth for each link and by considering these in the routing decision. It should be noted that (Apostolopoulos et al. 1999) does not consider any other QoS requirement apart from bandwidth. Even though propagation delay of a link is considered, it is only used for pruning the topology tree of the network, which is created by OSPF in order to calculate the paths. The path calculation algorithm itself uses a modified Bellman-Ford algorithm which allows for determining paths of maximum available bandwidth for all hop counts by exploiting the fact that the Bellman-Ford algorithm progresses by increasing the hop count. Hence, among the paths which support the requested bandwidth, a path with the minimum number of hops can be selected without increasing the complexity. This limitation on the available bandwidth and hop count is caused by a general problem of constraint-based routing, namely its complexity. More complex metrics or a combination of metrics easily result in an NP-hard or even NP-complete problem. For example, selecting a path based on multiple independent QoS-constraints such as delay and cost has already proven to be a NP-complete problem, if the QoS metrics are real numbers or unbounded integers (Chen & Nahrsted 1998, Kuipers et al. 2002, Yuan 2002).

In order to transport the information on the bandwidth of a link across the QoS-OSPF routers in a network an extension to the OSPF link state advertisements is used, so that each router has the required information to calculate the routes.

It should be noted that IntServ can be used without QoS-OSPF but with any regular routing protocol. This, however, might lead to more denied requests if capacity constraints occur in the network, since the routing is performed unaware of the available resources on the router.

These approaches have some major drawbacks. As it is working stateful, scalability can easily become an issue (Vali et al. 2004), regardless of the presence of satellite links in the network or not. Hence, neither IntServ nor QoS-OSPF are widely deployed in larger networks. (Apostolopoulos et al. 1999), for example, still targets the experimental state, which means that it is lacking interest in implementing this Request for Comments (RFC). Furthermore, the main idea behind these approaches is generally to have enough capacity available for critical traffic flows, since latency, jitter and loss usually increase if the available bandwidth is exceeded (Xiao & Ni 1999) or to prioritize real-time and other high-priority traffic by configuring the packet scheduler properly. Both are also important in converged satellite and terrestrial networks. However, those approaches do not consider the impact of the high fixed latency of satellite links, when choosing a path for certain traffic.

Even though (Shenker et al. 1997) describe how network elements, i.e. routers, should behave to provide services with a guaranteed end-to-end delay and bandwidth, determining a path, which fits latency requirements is out of scope. Hence, satellite links are not treated properly, since a significant part of the latency is independent of the link's load. Moreover, while QoS-OSPF distributes link information within the network, it does not provide any means to gain information on the link characteristics, besides the link status, i.e. up or down.

### 2.2.2 Traffic Engineering and Traffic-Aggregate based QoS

Given the scalability issues of flow-based approaches, current TE concepts usually operate on traffic aggregates. Already in the last century Diffserv (Nichols et al. 1998, Blake et al. 1998) has been standardized, which provides traffic classification and therefore allows for prioritization of IP traffic by using the IP DSCP field. While various queuing strategies and optimizations exist (e.g. (Li & Mao 2004, Liao & Campbell 2004)), Diffserv provides QoS only by prioritizing packets of higher priority traffic. It is only mentioned here for the sake of completeness, since pure prioritization is insufficient for converged satellite and terrestrial networks with parallel links since such an approach only optimizes the terrestrial connection but does not offload traffic onto the satellite.

Furthermore, current TE approaches often rely on MPLS (Rosen et al. 2001) and LSPs (Awduche et al. 1999) in order to become independent of the link layer technology. Instead of calculating a proper route for each flow, routes are calculated for an aggregate level on a larger-time scale. Hence, more complex path computation algorithms can be used compared to e.g. QoS-OSPF. Typically, RSVP-Traffic Engineering (RSVP-TE) (Awduche et al. 2001) is used to establish these MPLS LSPs in the network and to block the proper resources (Awduche et al. 2001, Claise (Ed.) 2008). For different QoS classes different MPLS LSPs can be established, which introduce the challenge to map the traffic onto the LSPs.

Moreover, to further enhance the path computation, the IETF has created the so-called Path Computation Element (PCE) architecture (Farrel et al. 2006). Among other reasons, one motivation for PCE is to offload computational intensive path calculation, such as the path selection based on multiple independent constraints, which can easily overload a single network entity, to a more powerful machine. So-called Path Computation Clients (PCCs), which might be an edge router of a network domain, can request a PCE to calculate a path using the Path Computation Element Protocol (PCEP), as depicted in Figure 2.3. The PCE consists of a path computation module, a Traffic Engineering Database (TED) and a communication module. While the TED collects and tracks the current network status including the network topology, link costs and the link utilization, the path computation module calculates the requested paths based on the information from the TED and selected algorithms and policies. Unfortunately, it is not defined how the topology information is gathered, apart from that it should be learned from a routing protocol (Farrel & King 2014). Moreover, with the PCE communication protocol and the communication module, a well-defined interface is available which enables PCCs to request path calculation between two arbitrary nodes in a standardized way. That is, a PCC, e.g. the first router, sends a PCReq message to a PCE optionally including the required bandwidth or other metrics. The PCE replies with a PCRep message that includes the path, in case of a successful calculation, which, in turn, can be used by another signaling protocol, i.e. RSVP to set up a path in the network accordingly.

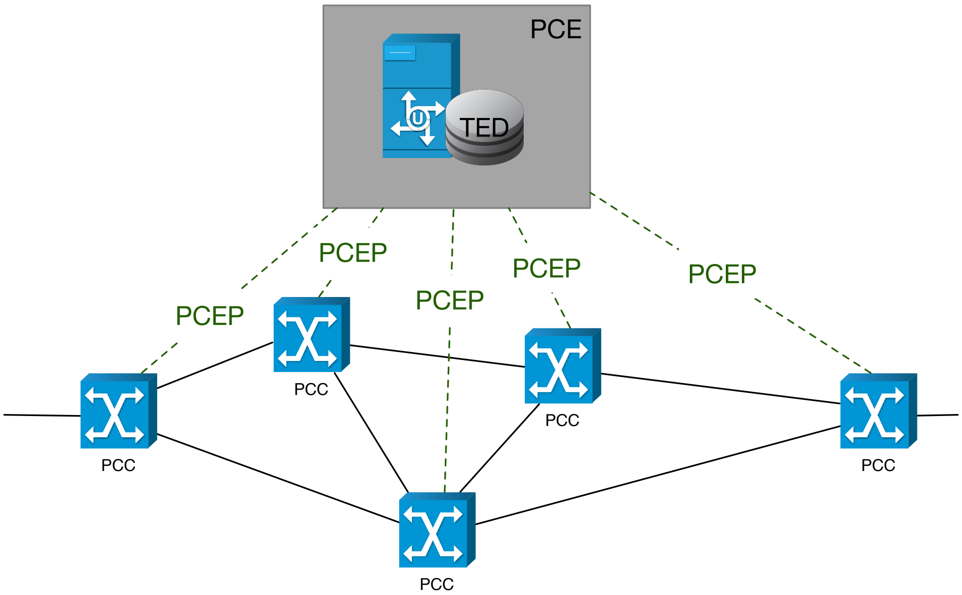



Figure 2.3: Centralized Path Computation

It should be noted that PCE is clearly designed with terrestrial networks in mind. Its major goal is to solve bandwidth issues in a single or multi-domain environment. For example, the constraints based on which the path is calculated are connectivity, available bandwidth and link costs.

### 2.2.3 Load Balancing

Furthermore, multiple LSPs might exist between any source and destination. In such a case the load must be distributed onto these paths. While there are specific MPLS approaches, such as (Foteinos et al. 2014) or (Elwalid et al. 2001), general algorithms exist as well. These can also be used with a regular IP routing protocol in case multiple paths of the same cost to a certain destination exist, which is referred to as ECMP. The purpose of ECMP is to distribute traffic onto more than one path in order to balance the load and to avoid congestion, so that eventually a higher total throughput can be achieved (Kandula et al. 2007, Fernandez et al. 2009, Brassil 2005). A main differentiation between load distribution approaches is inter- and intra-connection parallelism. The traffic belonging to the same connection (or application) can be either sent over multiple available connections in parallel (intra-connection) or only over a single one and another connection might use a different parallel path (inter-connection). This is also often referred to as per-Packet distribution and per-Flow distribution, respectively. Per-packet approaches usually allow on the one hand for fine-grained and flexible path selection, since the link is selected for each packet independently, but on the other hand might cause re-ordering of packets. Re-ordering will occur particularly if the latency of the multiple links differ heavily, which for example impacts the performance of TCP (Kandula et al. 2007) or introduces extra time needed to recover the packet order again (Prabhavat et al. 2011). In contrast, approaches providing inter-connection parallelism do not introduce re-ordering of packets of a single flow but in the case of very different flow characteristics might introduce over- or underloaded paths.

A further classification of load distribution into adaptive and non-adaptive models is given in (Prabhavat et al. 2012). While the non-adaptive models distribute the links statically and thus cannot react on dynamically changing conditions, adaptive models are able to react to variations in traffic- or network conditions. Typical examples for non-adaptive models are round-robin approaches or simple hash-based techniques.

Adaptive Load Balancing can be classified into Traffic-Condition-Based and Network-Condition-Based approaches. While Network-Condition-Based models can adapt to changing network conditions, such as delivery time or network utilization in terms of packets/s or bytes/s, Traffic-Condition-Based Models consider traffic characteristics, e.g. flow- or packet size, packet arrival time, etc., when performing load distribution. It should be noted that those classes are not mutually exclusive.

The Adaptive Flow-Level Load Control Scheme for Multipath Forwarding (Lee & Choi 2001) is an example of a Traffic-Condition-Based approach.

The authors make use of the fact that in IP networks one can generally differentiate between two kinds of flows, namely long-lived and short lived-flows. Short-lived flows, so-called transient flows, occur more frequently and have greater variation in packet arrival time than long-lived flows, while the long-lived flows carry the major part of the traffic load and, hence, are referred to as base flows. In order to detect a base flow the number of packets $X$ per flow within a given time $Y$ is measured. The idea of (Lee & Choi 2001) is, if two links are available, to send all base flows over one path and all transient flows via the other one. By adapting $X$ and $Y$ the classification into base and transient flows can be dynamically changed and, thus, the assigned load can be controlled. The authors also propose a simple Load Control Algorithm, which adapts the $X$ value based on the load ratio on the primary path. This is measured in packets sent via the primary path over the total number of packets.

An example for the Network-Condition-Based method is the Earliest Delivery Path First algorithm proposed in (Chebrolu et al. 2005) and (Chebrolu & Rao 2006). The authors' approach requires multiple connections between a mobile terminal and a counterpart, a so-called Network Proxy, in the network. By estimating the delivery time of each packet for each of the available paths, it schedules packets so that they are delivered as early as possible. The calculation takes into account the available bandwidth, the packet size and the queuing delay. This approach aims at increasing the usable bandwidth to the same performance a single link with the same aggregated bandwidth would have. However, (Chebrolu & Rao 2006) explicitly ignores the signal propagation delay, i.e. the fixed latency, since it is designed for terrestrial networks. Moreover, the Earliest Delivery Path First algorithm approach assumes that changes in the available capacity on each link and delay variations are only minor, which might not be an accurate assumption for satellite connections.

As can be seen, load distribution can be used to avoid congested links by balancing the load, even if the underlying networks are not homogeneous in terms of their capacity. Clearly, non-adaptive models are not suitable for the considered scenario, as explained in (Niephaus et al. 2013). However, also adaptive models do not provide traffic differentiation in terms of QoS requirements. Hence, load distribution needs to be done per traffic class. For further details please refer to (Prabhavat et al. 2012), which provides a comprehensive study and comparison of different adaptive algorithms.

### 2.2.4 Mobile-IP

The primary goal of Mobile IP (MIP) is obviously to allow for mobility of IP nodes in the network. However, given that a mobile node in an MIP enabled network might also have access to multiple networks this leads to similar challenges as in the considered scenario. Furthermore, extensions to MIP also deal with individual flow routing in a network. Hence, MIP is analyzed here as well.

One of the main issues with IP is that it acts as locator of node as well as the identifier for the higher layer. That is, the IP address is used to find the path from a given source to a destination and also to identify the destination node. For stationary nodes this approach seams appealing as identifying the node as well as finding its location is basically the same problem and is solved by the same mechanisms, namely IP routing (Soto et al. 2010). However, it makes mobility difficult to realize. Hence, MIP has been standardized to enable mobility in IP networks.

MIP and Mobile IPv6 (MIPv6) are long-standing standards, i.e. (Perkins (Ed.) 2010) and (Perkins (Ed.) et al. 2011), respectively. The main idea of both is to allow for roaming of devices between different networks while maintaining their IP address. Therefore, each device is equipped with a so-called home address and a care-of address. While the first is a permanent address, the latter identifies its current location of the device. In order to enable to maintain connections, while on the move, an IP tunnel is established between the moving device and an entity in the home network, i.e. the home agent. The home agent is a router in the home network of the device. Its responsibility is twofold. On the one hand it keeps track of the current location of the mobile device, and on the other hand it is responsible to deliver the packets to the mobile node, while in a foreign network, as depicted in Figure 2.4.

The counterpart to the home agent in the foreign network is the Foreign Agent. It usually acts as default router for the Mobile Node and tunnels the datagrams received from the Home Agent, if not sent directly to the Mobile Node. Furthermore, the Mobile Node is suppose to discover the agents and to detect whether it is in the home network or a foreign network, by relying on Agent advertisements sent by the agents. In addition to that, the Mobile

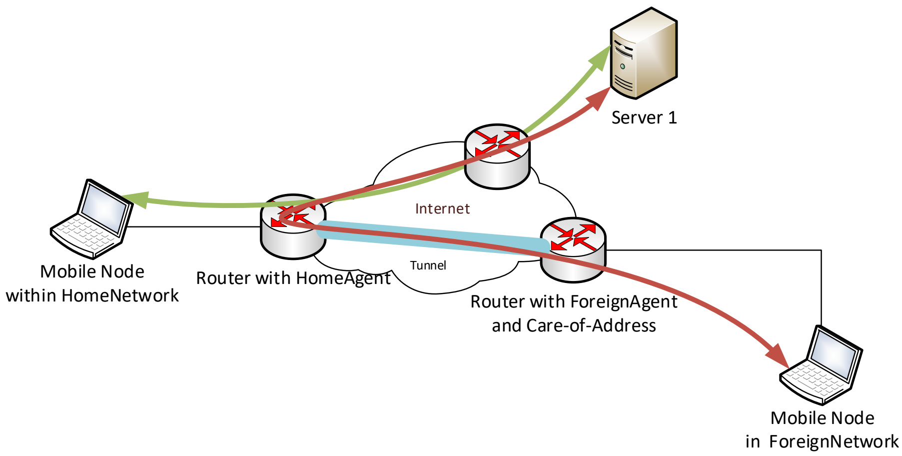



Figure 2.4: Mobile IP

Node is responsible for registering itself with the new IP address.

Mobile IP operates on a per device basis. That is, all traffic is impacted if Mobile IP mangles with the routing process and not just a single flow is routed differently, which is not suitable in the considered scenario.

### Proxy Mobile IP

Regular Mobile IP requires modifications on the end user devices in order to detect being in a foreign network and to communicate the new care-of address to the home agent. Hence, this kind of mobility is called host-based mobility, as the end host must support it and must signal its current point of attachment to the network. This requires complex security mechanisms to ensure message integrity and interoperability, which often limits broad usage (Kempf (Ed.) 2007). Therefore, PMIPv6 (Gundavelli (Ed.) et al. 2008) transfers all mobility functions in the network, so that no modification on the host systems are required.

In order to do this, PMIPv6 introduces the concept of a Local Mobility Domain (LMD), Local Mobility Anchor (LMA) and Mobile Access Gateway (MAG), as shown in Figure 2.5. LMD refers to the localized area in which PMIPv6 is supported. That is, while a mobile node is moving within the LMD, the network keeps track of its current location, i.e. point of attachment. The actual mobility functions are realized by the MAG and LMA.

The MAG is responsible for signaling the movements of mobile node on its behalf. If a node attaches to a specific MAG, it signals it to the LMA. The MAG is, thus, usually the first-hop router. The LMA keeps track of the MAG each mobile node is attached to. Hence, the LMA acts as the anchor point in the network.

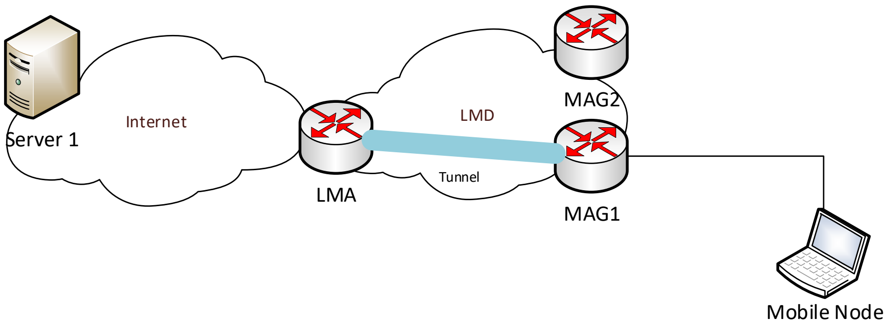



Figure 2.5: Proxy Mobile IPv6 Entities Overview

In general, PMIPv6 works as follows: As soon as a mobile node attaches to a certain MAG, the MAG signals a so-called Proxy Binding Update to the LMA, that associates the MAG address with the identity of the mobile node. Upon reception, the LMA assigns an IPv6 network prefix to this particular mobile node and signals this prefix back to the MAG with a Proxy Binding Acknowledgment message. Moreover, a bidirectional IP tunnel between the LMA and the MAG is established. Afterwards, by using regular IPv6 stateless auto-configuration, the MAG sends a regular Router Advertisement (Thomson et al. 2007) message to the mobile node, that includes the prefix the LMA has assigned before. This process is depicted in Figure 2.6.

If a mobile node now starts to move to another MAG, the current MAG sends a de-registration message to the LMA and the new MAG signals the new attachment, similar to the attachment in the first place via a Proxy Binding Update to the LMA. The LMA tears down the tunnel to the old MAG and established a tunnel to the new MAG. It should be noted that the

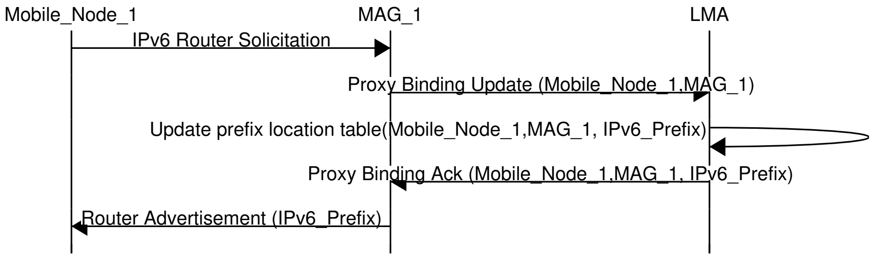



Figure 2.6: PIMPv6 Node Attachment

IPv6 prefix, which is send by the LMA to the MAG via the Proxy Binding Acknowledgement remains the same for a certain nodes. Hence, the mobile node keeps its IP address. Hence, PMIPv6 works only within a LMD, as previously mentioned.

It should also be noted that regular PMIPv6 is, similar to MIP operating on a per IP device level, which make it likewise not suitable for the considered scenario.

#### Flow Mobility for PIMPv6

In (Melia et al. 2011) the authors discuss also Flow Mobility for PMIPv6. That is, that the mobile node is equipped with two or more interface and is not necessarily moving to a different MAG but is in the range of multiple MAG, so that individual flows can be moved to a different MAG, i.e. to a different network. The authors' main motivation is to distribute traffic generated by a mobile phone between the cellular network and a local WLAN. In order to enhance regular PMIPv6 with this kind of mobility requires several modifications.

The first issue that is identified with regular PMIPv6 (Gundavelli (Ed.) et al. 2008) is that PMIPv6 typically operates on an interface level. That is, a prefix is actually assigned to an interface and not to a node. If a node is connected with multiple interfaces to the same LMA it receives different prefixes. An example is depicted in Figure 2.7. In this case, the mobile node has both, a WLAN and a cellular interface. With regular PMIPv6, both interfaces will get an individually assigned prefix from each MAG. More importantly, also the state maintained at the LMA needs to be extended compared to regular PMIPv6. On the one hand, mobility bindings for the

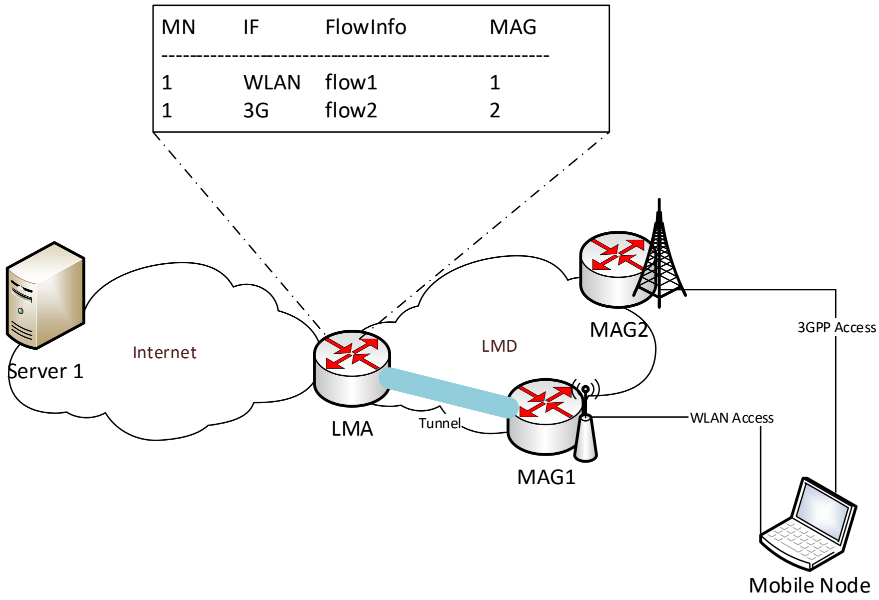



Figure 2.7: Flow mobility in PMIPv6

same nodes but from different interfaces needs to be grouped, and on the other hand, a flow state needs be kept as well, in order to allow for the required flow mobility. Given that in such a situation the mobile node has theoretically also two options to route the traffic and, hence, also needs to maintain a flow table, (Melia et al. 2011) propose to implement the decision process only in the downlink. For the uplink traffic the Mobile node should just replicate the decision that was made by the LMA. The proposed ideas of (Melia et al. 2011) have to a considerable extent materialized in (Bernardos (Ed.) 2016). However, this approach requires the mobile node, i.e. the end user system, to become aware of the two networks.

Furthermore, even though the architecture itself is well-defined, it is open to the operator of such a network to define the policies based on which flow mobility rules are applied, i.e. which traffic is routed using which MAG. Examples given in (Soto et al. 2010) that trigger a flow handover are a complete interface shutdown, which handovers all flows to the other interface, as well as QoS changes, which are simulated by bandwidth degradation.

### 2.2.5 Summarization of IP-related techniques

To conclude, besides performing routing, i.e. finding a path from an arbitrary source to a destination, IP-related TE approaches are highly optimized to perform TE in terrestrial networks as they aim at avoiding congested links, balancing the load, prioritizing certain traffic and managing the available capacity properly. Even mobile networks and their requirements and use cases, e.g. flow specific traffic routing to a WLAN, are to a certain extant addresses by PMIPv6 and flow mobility. However, the unique characteristics of satellite links are not taken into account - more precisely, the high fixed latency and the varying link conditions prohibit the integration of satellite links. Moreover, in terms of the aforementioned functional building blocks, particularly for the PMIPv6, the clear focus is on the execution function. That is, offloading can be realized in a defined way, but which traffic is offloaded (path selection block) is out of scope or up to a network operator to define proper policies.

## 2.3 Traffic steering in LTE networks

Obviously operators of mobile networks also need to deal with the significantly increasing amount of traffic in nowadays' networks. In order to do this, many approaches exist. Besides increasing the capacity by designing better radio connections, many of them require more complex traffic steering than in legacy mobile networks. Thus, certain mechanisms and techniques seem applicable to the scenario considered in this thesis as well. Firstly, with the advent of small cells and LTE heterogeneous networks (HetNets) (Andrews et al. 2012), traffic needs to be distributed between the macro and the small cell. Offloading users to small cells might actually increase interferences in the macro cells, which needs to be traded off. Indeed, a lot of research work in this area focuses on spectrum sharing issues, which is obviously important for this kind of offloading. Secondly, the appearance of LTE Dual Connectivity as well as LTE in unlicensed bands and the joint usage of LTE and WLAN by a single User Terminal also requires distributing traffic in a controlled manner among the available channels. Finally, approaches exist that aim to relieve the costly links in both the Radio Access Network (RAN) part as well as the mobile packet core by rerouting traffic to more cost-effective paths or by enabling Device-to-device (D2D) communication (Maallawi et al. 2015).

For the sake of completeness, it should be noted that there also other types of offloading, such as local caching, which have no relevance in the scenarios considered in this work. These are not listed in the following part.

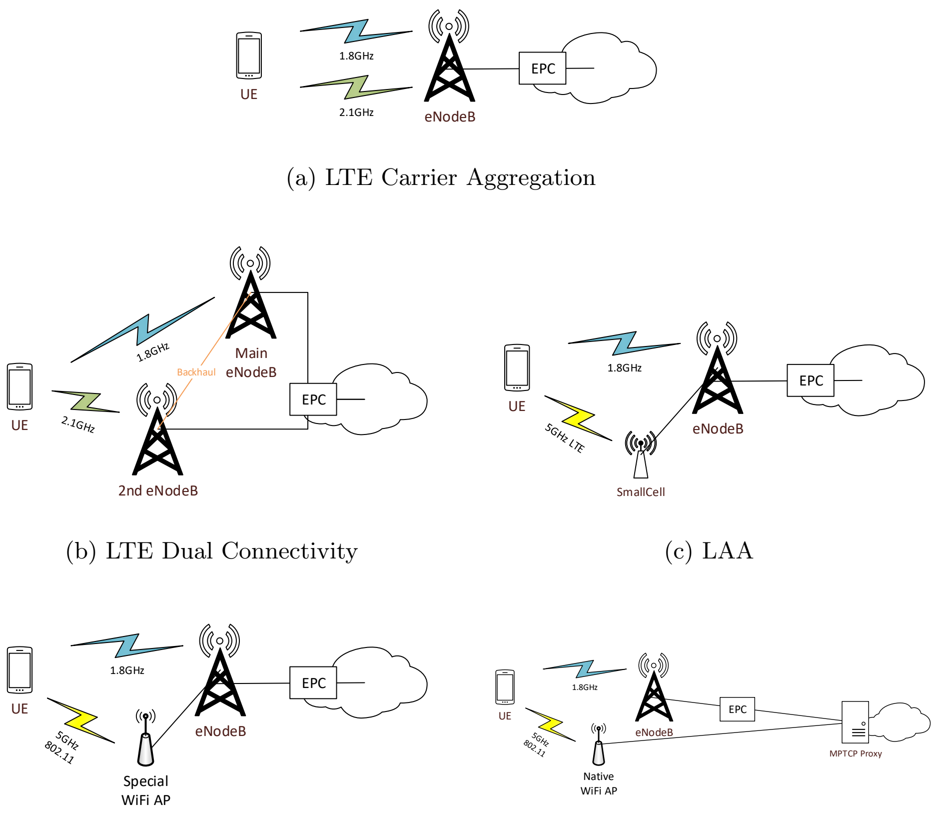
### 2.3.1 LTE heterogeneous networks

In so-called HetNets, besides typical macro-cells, which cover larger areas, additional smaller cells exist to allow for further spatial re-usage of spectrum and to eventually increase the user's performance. In such scenarios, a major challenge is that the spectrum resources needs to be shared and transmissions need to be coordinated in order to avoid interference between the macro and the small cells. Several approaches exist that aim at assigning Mobile Terminals (MTs) to either macro or small cell, while considering the interference between them. Various approaches exist, which use different algorithms and optimization parameters. For example, in (Chen et al. 2015) the authors use model-free reinforcement learning to minimize the overall energy consumption, while still maintaining a certain QoS level. More precisely, their approach decides, depending on the traffic load, how many additional small cells have to be enabled in order to minimize on the one hand the overall energy consumption for the cells as much as possible, while on the other hand maintain a high service quality. That is, if a macro cell is heavily loaded additional small cells needs to be turned on in order to achieve a high QoS. However, once the load deceases these small cells should be switched off again, in order to reduce energy consumption. Similarly, (Kong & Karagiannidis 2016) define an algorithm to offload MTs to small cells. The authors proposed a mechanism to allocate time slices for transmission to macro and small cells, while considering the interference between them. By doing so, they aim at maximizing the overall throughput. Compared to other work in the regards, the novelty in their approach is that they assume that the small cells are not connected via an ideal connection but with a capacity constrained backhaul. However, the scope here is to manage the spectrum and usually to offload the traffic of a complete MT, if this is beneficial. These approaches are agnostic to individual traffic flows. For such a scenario this a valid approach, since more usable capacity in both RAN and backhaul eventually leads to less latency for the user and, thus, increases QoS and QoE.

Another variant of HetNets is considered when D2D communication is used, e.g. in (Sciancalepore et al. 2016) and (Cheng & Lin 2015). Typical applications for this kind of offloading are streaming. Given that in these opportunistic networks, it needs to be traded off between load in the cellular network and the additional latency being introduced by this kind of D2D communication. It should be noted, that data can only be exchanged when the User Terminals are in proximity. The authors in (Sciancalepore et al. 2016) present their so-called HYPE approach, which is based on the content distribution to a couple of nodes via the regular mobile network, which are then responsible to transmit it to the other receivers. However, the assumption of the authors is that a particular data chunk needs to be received by every node in the network. Moreover, their focus is to optimize the number of data chunks injected via the mobile path into the network, so that both delay constraints can be met but also as much traffic as possible is offloaded. Similarly, (Cheng & Lin 2015) deal with the selection of initial nodes and the opportunistic forwarding strategy, while also considering the user preferences of its current and future local contacts to further optimize the distribution. To conclude, offloading traffic from mobile network to opportunistic D2D networks usually deals with the issue of content distribution, i.e. particular content needs to be received by multiple receivers. The challenges that arise from these questions are typically on how many and on which nodes should the content be "injected" by the mobile network, so that it can quickly be distributed, and when is the amount of MTs high enough that offloading traffic to D2D communication makes sense. Given satellites are primed for broad- and multicasting, since they can reach many receivers with a single transmission, most of the challenges becomes obsolete.

### 2.3.2 Channel aggregation

With the advent of usage of LTE in unlicensed bands, i.e. LTE-Unlicensed (LTE-U), as well as Carrier Aggregation and Dual Connectivity (Jha et al. 2014), traffic steering becomes also important in mobile networks, as traffic needs to be distributed between either multiple LTE base station, i.e. Evolved Node Bs (eNodeBs) or a WLAN access point and the eNodeB, as depicted in Figure 2.8. The aggregation of multiple links in order to increase the performance can be done on different layer and with different technologies. That is, LTE Carrier Aggregation (Figure 2.8a) basically uses two or




(a) LTE Carrier Aggregation

(b) LTE Dual Connectivity

(c) LAA

(d) LWA

(e) MPTCP

Figure 2.8: LTE aggregation technologies

more LTE channels to increase the throughput towards the User Equipment. In contrast to that, in LTE Dual connectivity (LTE DC) (Figure 2.8b) the User Equipment is being served from two different eNodeBs. Once unlicensed 5 GHz bands are considered in a LTE network, two approaches are standardized, namely Licensed-Assisted Access (LAA) and LTE-WLAN Aggregation (LWA). While in LAA the LTE radio technology is used on the 5 GHz band (see Figure 2.8c), using LWA a special WLAN access point is used to provide a second connection to the User Equipment, as depicted in Figure 2.8d. Additionally, traffic can be send via two channels, i.e. LTE and WLAN using transport layer techniques, such as MPTCP (see Figure 2.8e). All the aforementioned approaches share the same challenge to distribute traffic onto two paths.

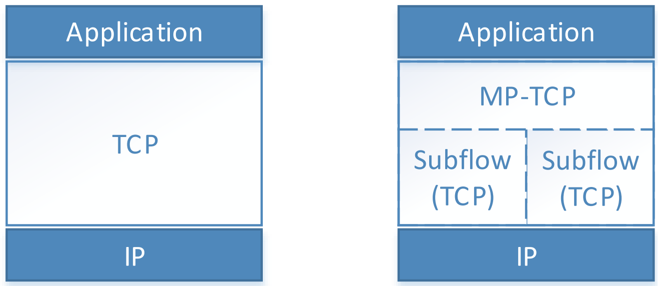
Several papers already deal with the traffic control, i.e. the traffic distribution between the two paths, e.g. (Lopez-Perez et al. 2016, Laselva et al. 2018). Typically, these approaches use a capacity metric or the expected delay to distribute the traffic. Again, this seems reasonable in a terrestrial environment or mobile network environment. However, high fixed latencies are not considered. Often also an ideal Back-haul connection is assumed, which does not have any latency. It should be noted, that in (3GPP TR36.842 2013) it is recommended that the Back-haul latency is between 5 ms and 30 ms in order to have an adequate performance. More precisely, according to (Jha et al. 2014) latencies of 20 ms and more already have a negative impact on the performance, so that splitting the traffic onto both paths is not useful. The (fixed) latency as occurs on satellite links will have even a significantly higher impact.

Moreover, LWA aggregates LTE and WLAN on the Packet Data Convergence Protocol (PDCP) layer, i.e. on layer 2. Given that PDCP ensures in-order delivery to the higher layer, it is obvious that a high latency on one of the links leads to significant buffering in order to have a consistent order.

### MPTCP

The current TCP (Postel 1981) supports only connections between a pair of IP addresses, even though nowadays devices are often multihomed and, hence, both source and destination end hosts might have multiple IP addresses and different paths between them might exist, which are not exploited. In order to address this, MPTCP (Ford et al. 2013) has been defined by the IETF, which enables end hosts to use multiple, but not necessarily disjoint, paths and IP addresses while providing the same interface to both the application and the network layer as regular TCP and, thus, does not require any modifications neither in the application nor in the IP protocol. By using multiple paths in parallel, MPTCP aims at increasing throughout, reliability and resilience.

Figure 2.9 compares the protocol stack of regular TCP and MPTCP. As can be seen, MPTCP is logically located on top of regular TCP connection and utilizes one or more TCP instances. In general MPTCP behaves like ordinary TCP, i.e. the connection setup, data transfer and connection termination phase are nearly identical. The most significant difference is that additional TCP sessions, so called MPTCP *subflows*, can be established and combined with the existing connection if multiple paths exists. Those sub-




(a) TCP (b) MP-TCP

Figure 2.9: Comparison of TCP and MP-TCP protocol stack

flows can be used as regular path or as backup path, which will only be used if no regular path is available anymore.

In order to reliably deliver TCP segments without reordering over multiple subflows while additional subflows might be established or existing subflows might be terminated at any time, MPTCP introduces an additional 64-bit data sequence number, which counts all data send over an MPTCP connection. In order to ensure in-order delivery within a single subflow the regular TCP 32-bit sequence number is used. Due to a 64-bit data sequence number, lost or corrupted segments can be retransmitted over another subflow.

It should be noted that (Ford et al. 2013) do not provide any algorithm to distribute data to the available subflows. The authors leave this to the implementer or the host policy to decide. Moreover, the usage of multiple, potentially asymmetric paths might cause MPTCP to deliver data with a higher jitter and more bursty than regular TCP if the data is distributed among multiple subflows and multiple, heterogeneous paths. In particular real-time applications might be affected by that (Scharf & Ford 2013).

It should be noted that TCP and likewise MPTCP is intended to be an end-to-end protocol, i.e. establishing a connection between the end systems. Hence, in order to use MPTCP both end systems need to support MPTCP. To enable MPTCP on just a part along the end-to-end path, proxy servers need to be used, similar to PEP (see Section 1.1.1).

### 2.3.3 Traffic Break-out

Another kind of offloading standardized for mobile networks is the so-called traffic break-out. While channel aggregation techniques mainly focus on the exploitation of additional RAN capacity by using multiple transmissions, traffic break-out approaches aim at routing traffic as early as possible to high capacity networks. This also includes routing traffic from or to a MT through customer's fixed line connection by e.g. using WLAN. Obviously this use case overlaps or even competes with LWA approaches, yet the motivation is somewhat different. Traffic break-out approaches have been designed to offloading certain traffic via low-cost fixed access networks to the Internet without routing in through the operator's core network. This seems an appealing approach to reduce the load in operator networks (Samdanis et al. 2012). One motivation for the operator to perform this kind of offloading arises from the appearance of so-called Over-the-top (OTT) content provider, such as YouTube or Netfllix. These OTT services are typically hosted outside of the operator's network and, thus, increase the traffic transit costs (Maallawi et al. 2015).

Typical methods are SIPTO and Local IP Access (LIPA). LIPA aims at routing traffic to private networks directly instead of the regular data path through the evolved Packet Core (EPC). It requires for this purpose a local femto cell, i.e. a Home eNB (HeNB), which acts as the so-called break-out point. A typical use-case is to use a local printer or other resources in the local network from a MT. SIPTO enables a break-out also at an HeNB or at/above a regular eNodeB, i.e. at the macro cell. While LIPA aims at making local resources securely and efficiently accessible, the goal of SIPTO is to ease the load on the operator's core network, e.g. Packet Data Network Gateway (PDN-GW) or Serving Gateway (S-GW) by breaking out selective traffic closer to the edge, as depicted in Figure 2.10.

As described in (3GPP TS22.805 2012), typically three main strategies exist, namely user or Service Level Agreement (SLA)-based, application-based or traffic-type based. A user or SLA-based offloads traffic solely based on the user, which generates the traffic, regardless of the traffic characteristics or used application. Contrary, an application-based approach considers the application or application type to control the traffic offloading. For example, Video on Demand (VoD) applications are being offloaded. Likewise, the traffic-type approach considers the actual nature of the traffic. In

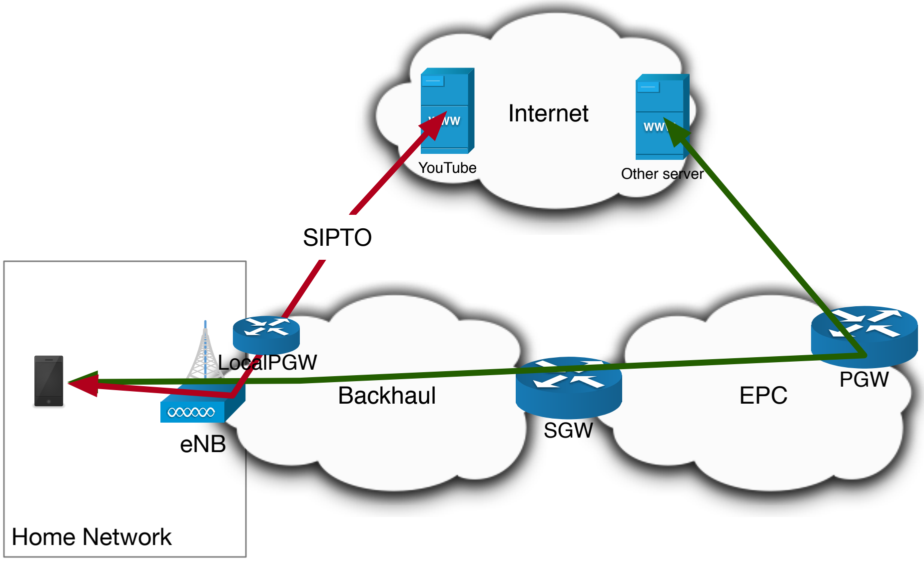



Figure 2.10: SIPTO approach

contrast to the application-based approach, however, traffic-type acts more fine-grained. Given that many applications involve different types of traffic, only certain traffic of an application might be effected. For example, a VoD application usually consists of two types of traffic: When the user is browsing through the menus in order to select a movie, the traffic resemble typical web-traffic. Once the video playback is started it changes completely to a typical stream traffic pattern. Both traffic types have significantly different requirements in terms of QoS they expect from the network.

Technically, one option to implement SIPTO is to use NAT. A local Offloading Processing Module at e.g. the HeNB decides, based on policies, to change IP addresses and therefore allows for a different routing. Another option is that a local PDN-GW is used and MTs establish connections to multiple PDN-GW. Hence, traffic that should be offloaded can be distinguished by using different APNs.

In general, SIPTO enables a relatively static offloading of traffic, i.e. based on the MT or the APN, the application protocol/port or DNS/IP addresses. That is, traffic matching certain IP addresses are offloaded. These policies needs to be available at the break-out point, e.g. using TR-69. (Samdanis et al. 2012) also suggest to enhance the Domain Name Service

(DNS) system with a DNS proxy at the local gateways to transport additional flags and offload traffic based on this.

It should be noted that the majority of publications made in this regard, actually deal with integrating the offloading approaches into the existing LTE infrastructure and enabling a proper IP routing, e.g. (Sou 2013). That is, architectures and reference models that enable offloading in LTE networks are available. However, the decision which traffic should be offloaded is usually up to the operator.

### 2.3.4 Summarization of offloading techniques in LTE networks

The term offloading is widely used in LTE networks and relates to different techniques and use cases. In HetNets offloading often deals with assigning MTs to either small cells or macro cells most optimally, in order to decrease the interference between macro and small cells. Hence, the goal is to optimize the spectrum usage and hence, increasing the overall throughput.

With respect to offloading of specific traffic flows, as required in this work, typically rather static approaches exist that perform a policy based offloading decision. These policies are typically defined by an operator and are e.g. SLA based, so that customers paying less are offloaded first onto a WLAN, while other customers are kept in the LTE network, Traffic-class based, i.e. Best-Effort bearer is offloaded, while voice traffic uses the LTE network, or destination based, e.g. traffic to the Content Delivery Network (CDN), where all the YouTube videos are, is offloaded. These are usually used in the context of SIPTO.

In the context of LWA and LAA more dynamic approaches appear. Most of them distribute the traffic even on a per packet granularity and in order to decide which link to use, these approaches usually use the delay on both links as one input parameter. This works, since LTE and WLAN has similar delay characteristics. (Even the LTE standard recommends less than 20 ms delay in the Back-haul from LTE to WLAN network.) However, this is not a given in a satellite scenario.

## 2.4 Software-defined Network

It is generally believed that extended SDN approaches increase the flexibility of networks and enable novel concepts to address challenges in contemporary networks. With the advent of SDN (Open Networking Foundation 2012) a paradigm change in networking architecture started, shifting from monolithic network devices, which combine control, monitoring, management and data-forwarding functions in a single entity, towards a clear separation of control and data planes. That is, the decision making processes, such as routing of traffic, firewalling, spanning-tree protocols, etc., are clearly separated from the pure data forwarding methods. This allows for a more flexible management of the network, as the control functions can be run centralized. Further details on the differences between traditional networking and SDN are presented in (Nunes et al. 2014).

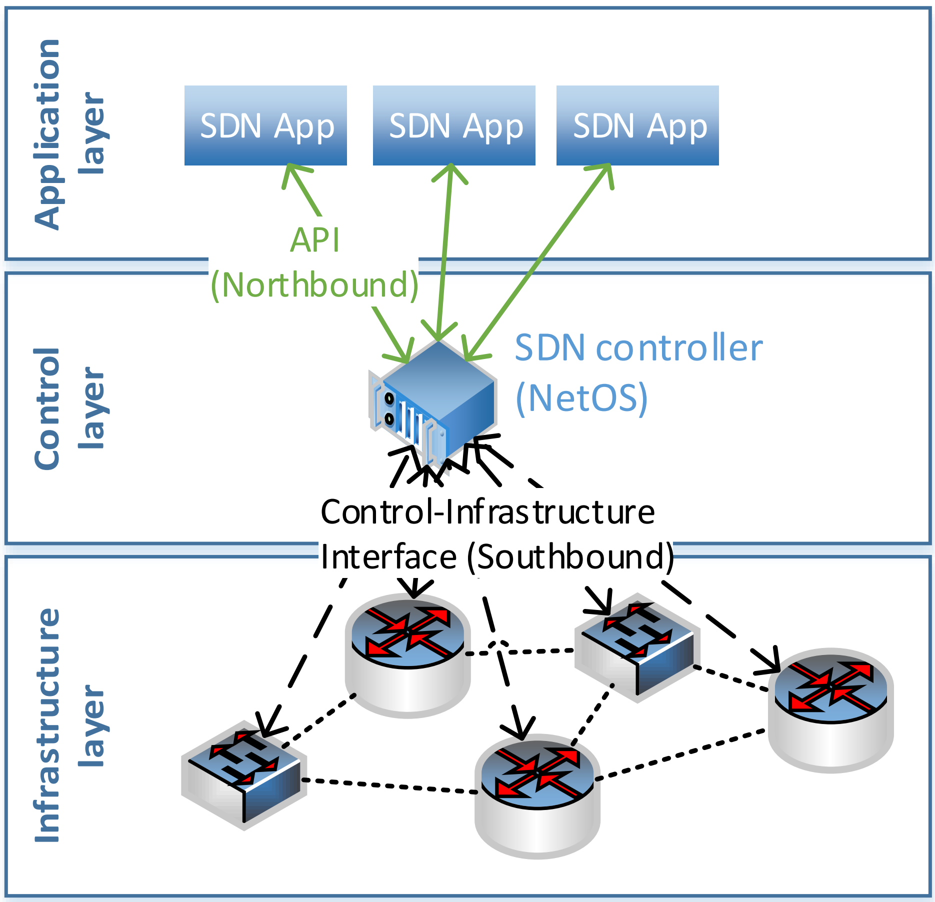



Figure 2.11: SDN architecture

SDN-enabled networks are mainly characterized by two things, first the decoupling of control- and data-plane, and second, programmability (Xia et al. 2014). Figure 2.11 shows the general SDN architecture (Xia et al. 2014). At the lowest layer, the infrastructure layer, the actual data forwarding devices are located. Their main task is to perform any kind or packet processing based on the rules that the SDN-controller, which is located in the middle layer, provides. A single controller typically controls multiple network devices.

The most commonly used protocol between the SDN controller and the devices on the infrastructure layer is currently OpenFlow (Open Networking Foundation 2015). This interface is also often referred to as the Southbound interface. It is used to push rules to the infrastructure layer, to request monitoring information and statistics or to transmit packets, for which none of the rules apply to, back to the controller. More precisely, the OpenFlow protocol is used to configure the flow tables of an OpenFlow -enabled switch. An entry in such a flow table consists, among others, of match fields and instructions. The first describes the fields against a packet is matched. The OpenFlow standard supports various fields on different layer to match against, e.g. source and destinations addresses, ports, protocols, but also MPLS header fields. If a packet matches the filter rules defined in the match fields, the actions in the instructions field of the flow table are executed. These can be e.g. to forward a packet, rewrite certain headers fields or to drop it.

Furthermore, the control layer provides an application programming interface (API), the so-called Northbound interface, to the application layer, which contains the so called network applications. An application might be as simple such as a centralized Dynamic Host Configuration Protocol (DHCP) server or more complex services like parental control for certain User Terminals or seamless mobility. It should be noted that so far there is no standardized Northbound interface.

Recently OpenFlow has been exploited for typical many network challenges, in order to either increase the flexibility or to avoid using proprietary protocols. For example, several approaches exist to realize QoS provisioning by exploiting SDN concepts. An automatic QoS management mechanism, which utilizes SDN concepts to automatically configure switches to provide queue management QoS is proposed by (Wendong et al. 2014). The authors introduced a QoS control module acting as an SDN application. A context manager is responsible for gathering and aggregating information on the network, such as switch utilization or packet loss rate, as well as on the traffic flow. Based on these parameters, rule decision components check if QoS requirements can be satisfied and, if so, proper rules are implemented. These rules realize the enforcement of QoS by adjusting the scheduling queue on each affected switch as well as by properly classifying the packets. Unfortunately, (Wendong et al. 2014) lacks a description on how information on the network status is gained from each device, since this is not part of the OpenFlow specification.

Similarly, (Egilmez et al. 2012) exploit SDN concepts to deliver end-to-end QoS. The main idea of the authors is to differentiate between multimedia flows, which usually have stringent QoS requirements, and BE flows. While the first are routed using a dynamic QoS routing approach aiming at reducing the delay, the latter are routed using a typical shortest path first algorithm. SDN is used in particular to ease the route calculation by having the required information in a central point, namely the SDN controller, in (Egilmez et al. 2012). However, similar to (Wendong et al. 2014), the publication lacks mechanisms to determine latency, available bandwidth, etc. on each link.

Matching SDN approaches against the function building blocks, it is obvious that mainly the Execution block is covered by the SDN architecture. Particularly the path selection and offloading decision function needs to be implemented as an application in the SDN context.

## 2.5 Traffic considerations

Besides the fact that the amount of traffic that nowadays networks need to carry is constantly increasing as described in the very beginning of this work, several other shifts in Internet traffic characteristics are known, which are described in the following, since they impact the architecture of a converged satellite and terrestrial network.

### 2.5.1 Encrypted traffic

The share of encrypted traffic is increasing. Statistics of Let's encrypt shows a significant increase in recent years, as depicted in Figure 2.12[1]. The Fig-

[1] https://letsencrypt.org/stats/#percent-pageloads

ure shows the percentage of websites accessed by Firefox browsers using an encrypted connection. As can be seen the percentage rate is around 80%.

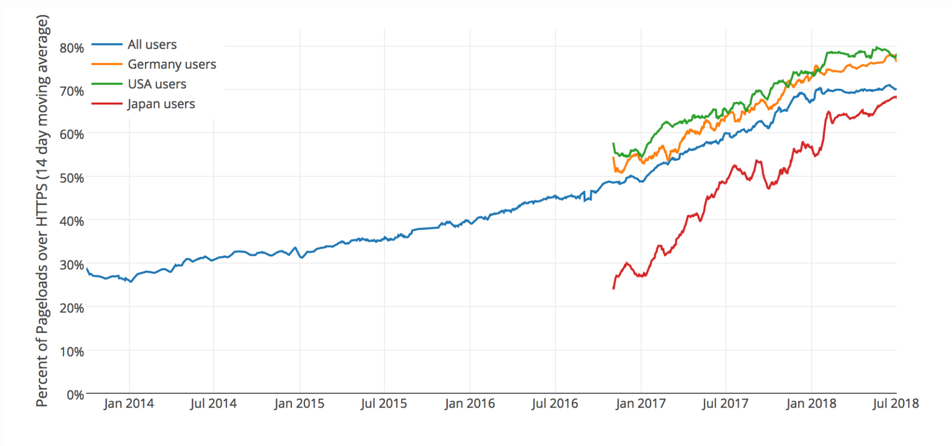


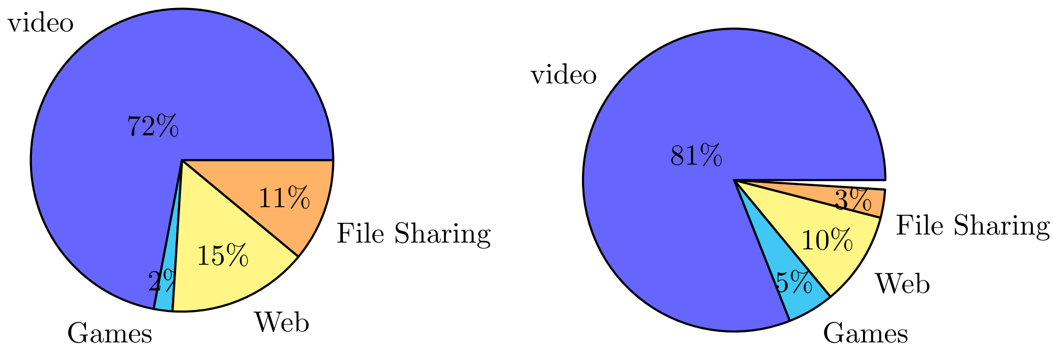

Figure 2.12: Percentage of Web Pages Loaded by Firefox Using HTTPS

Hence, the challenge of identifying the application or application class to determine the QoS requirements (see Section 1.2.2 is becoming significantly more complex, since analyzing the application layer in an IP packet, i.e. data above TCP layer, to identify the application or application class is not possible for the majority of the traffic. Thus, methods relying on DPI methods will suffer tremendously and will decrease accuracy.

Moreover, reliably identifying traffic classes of encrypted traffic is a challenging task. Some approaches exist for specific applications. For example, (Gu et al. 2011) uses machine learning algorithms to identify Skype traffic. The authors proposed an approach that can lead to an accuracy of over 90% after 10 packets of the traffic. However, generally identifying traffic classes of encrypted traffic is most likely a research topic on its own and out of scope of this thesis. Even though it is acknowledged that encrypted traffic imposes a strong challenge for the offloading procedure, it is believed that options exist to perform this task.

## 2.6 Traffic Composition

A good indicator of the actual traffic composition in terms of application type distribution can be found in (Cisco Cooperation 2017*a*). General effects are that the amount of global IP traffic is increasing and will reach 3.3 ZB in 2021, which is 127 times the amount of traffic of 2005. Moreover, traffic generated by smart phones will exceed the traffic generated by regular PCs in 2021. It is also important to node that the traffic in the busy hours of the Internet is increasing more rapidly than the traffic in the non-busy hours. It should also be noted, that CDNs will carry 71% of the global Internet traffic in 2021, which was only 52% in 2016.




(a) 2016 (b) 2021

Figure 2.13: Traffic composition in 2016 and the prediction for 2021

In addition to these general statistics, it is absolutely clear that video traffic is the predominant traffic in the Internet. As shown in Figure 2.13, video traffic accounts for most of the consumer Internet traffic. According to (Cisco Cooperation 2011) already in 2016 approximately 72% of the traffic was Internet video. The Compound annual growth rate (CAGR) is at 31% also the second largest. Only online gaming has a CAGR with 62%. However, the share of online gaming on the total Internet traffic is still only approximately 4% in the year 2021. It is also interesting to note, that the amount of file sharing traffic is not growing. While its share in 2016 was approximately 11%, it might actually decrease to approximately 3% in 2021, as the total amount remains approximately the same while the total Internet traffic increases. The reason therefore is a decrease in Peer-to-peer file sharing. Other file transfers, such as regular downloads are actually slightly increasing.

Of particular interest for this thesis is the development of video traffic, since a GEO satellite is well suited for many kinds of video traffic. In (Cisco Cooperation 2011) it is differentiated between video, which accounts for video calling, long- and short-term videos, live Internet TV, etc., and

Internet video to TV. The latter considers the video traffic that is generated by e.g. Internet-enabled set-top boxes or Internet-enabled TVs. Both have a CAGR of 32% and 27%, respectively. That is, approximately 60% of total consumer Internet traffic accounts for video traffic and 22% for Internet video to TVs. This trend is also supported by the increasing importance of CDN providers, which might even become the core of the Internet (Carisimo et al. 2018).

Moreover, due to the increasing video resolution, i.e. High Definition (HD) and Ultra-High Definition (UHD), the bandwidth occupied by VoD and other video applications is highly increasing. It should be noted that an Standard Definition (SD) video stream requires approximately2 Mbit/s, whereas an HD stream requires approximately 7.2 Mbit/s and an UHD stream even 18 Mbit/s . Given that by 2021, 56% of the connected TVs will be able to support UHD videos, the high CAGR becomes obvious.

One factor that might limit the relative increase of video traffic is online gaming. In particular newer gaming consoles like the PlayStation or the XBox tend to enable the download of games rather than selling DVDs or other media. Hence, large files are commonly downloaded to those devices. According to (Cisco Cooperation 2017*b*) the gaming download might reach up to 8% of the peak hour's traffic.

Furthermore, as described in (Sandvine - Intelligent Broadband Networks 2016), during the network peak period, Netflix, YouTube and Amazon Video jointly account for approximately 56% (in 2016) of the downstream traffic. Even though these figures are valid for North America, the trend is obvious. The authors even assume that by 2020 North America will be the first region where more than 80% of the downstream traffic will be generated by streaming applications. This trend of the predominance of video traffic is further supported by the growth rate of 9% of connected TVs, and TVs with set-top boxes or other digital media adapters (Cisco Cooperation 2017*b*). In addition to that, so-called "cut the cord" households, which do not have a regular TV broadcasting connection, i.e. cable or satellite will increase leading again to more video traffic. Particularly given that Internet video is typically distributed as unicast stream to the end user's devices, compared to typical broadcasting technologies where basically just a single stream is sent.

#### Summarization on traffic composition

The obvious trend that video traffic and in particular video streaming applications will account for the largest portion of Internet traffic can be verified by recent publications. It is also important to node that the share of video traffic is growing even further. Given the relaxed latency requirements of streaming video traffic, it becomes obvious that a converged satellite and terrestrial network can help to relieve the capacity problems in terrestrial networks by offloading this traffic to the satellite.

## 2.7 Conclusion of literature review and gap analysis

This chapter addresses the first two objectives defined in Section 1.4. Firstly, it develops functional building blocks, which are essential to achieve the overall goal of the thesis. Secondly, existing network architectures, which to a certain extend share similar challenges and provide adequate solutions in their particular domain, are mapped against these functional building blocks, in order to evaluate how applicable they are in the scenario considered in this thesis. The outcome of this analysis shows that existing approaches are either operating under the assumption that optimizing the available capacity is sufficient as this will lead to a low latency and a good user experience, which does not work in networks with a high fixed latency, or are lacking a decision functional block. Thus, the selection of traffic that should be treated differently is for the operator to identify by e.g. defining proper policies. Further insights attained from this analysis lead to Table 2.1 and are summarized in the following:

Standard IP routing performs the path selection on a hop count and the execution is on a per IP destination basis, which is insufficient as either the decision metric nor the granularity allows for routing individual flows differently. With the advent of IntServ, QoS-OSPF and RSVP this granularity changes, however, the decision metric is is solely the the available bandwidth, leading to biased decisions in favor of the satellite, without considering its delay. Obviously, the decision metric for MIP is the location of the node. However, the major gap is surely the granularity of the execution function, which works on a per device level. This changes with the introduction of flow-mobility and PMIPv6. This enhancement to MIP allows for routing individual flows differently, leading to a suitable execution functional block.

However, the path selection decision is virtually out of scope. An operator needs to define and implement certain policies. Modern TE mechanisms using MPLS and RSVP mainly share the same gaps as IntServ and QoS-OSPF routing with respect to considered scenario. The main difference is that these techniques usually operate at a different granularity level. That is, while IntServ works on a flow-level, MPLS-based TE approaches bundle multiple flows into a traffic aggregate in order to overcome scalability issues. PCE enhances the execution function is this approach, as it provides a defined centralized architecture to control the network and push decisions onto all involved network elements. However, again the path selection function that decides how traffic is routed is out-of-scope. Approaches used in the mobile RAN networks to connect a MT typically assume a lower latency in the backhaul systems than a satellite can provide, in order to allow effective offloading of traffic onto a secondary link. Besides that, typical metrics are signal strength and available bandwidth. Most of the interest for the considered scenario is the SIPTO mechanism used in mobile networks. However, similar to PMIPv6 it provides virtually only the execution functional block, as the decision which traffic should be offloaded is defined rather statically by polices, based on IP addresses, APNs or DNS-Names. This gap is also shared with the recently emerging SDN approaches. They provide sophisticated mechanisms to dynamically change traffic flows in a network, but they lack applications to decide.

Finally in this chapter, recent developments in the overall Internet traffic has been analyzed, in order to design the evaluation of different offloading approaches conducted in Chapter 5 more accurate. For the sake of completeness, it should be noted that the recent developments in this regard and in particular the increasing predominance of video traffic, shows that the integration of satellite networks can provide a huge benefit, since high latency has only a limited impact on streaming video applications. Particular in rural and remote areas, the additional bandwidth a satellite can provide can relieve the high load on narrow-band terrestrial link.

Table 2.1: Overview of network architectures with respect to converged satellite/terrestrial networks

| Architecture | Path Selection / Offloading Decision | Execution | Information collecting |
|---|---|---|---|
| Standard IP | based on hop count | modification of routing table on each node per destination | link up/down |
| IntServ with QoS-OSPF and RSVP | available bandwidth | modification of routing table on each node | link up/down, available bandwidth, required bandwidth per flow |
| Mobile IP | location based | IP tunnel configuration per device | node location changed |
| Flow-mobility with PMIPv6 | Policy-based | modification of per-flow routing table | node movement, new flow established |
| Traffic Engineering (MPLS, RSVP) | aggregate-based, policy or bandwidth metric | establishes LSPs | link up/down |
| PCE | Flexible, typically based on connectivity, bandwidth and link costs | ✗ | ✗ |
| LTE channel aggregation | not part of the standard | Aggregation on PDCP layer | signal strength, available bandwidth |
| SIPTO | Policy-based | NAT or DNS modifications | ✗ |
| SDN | ✗ | Updating of flow tables | Flow- and port statistics |

# Chapter 3

# Current State-of-the-art with respect to converged satellite and terrestrial networks

In the following chapter current research projects and other research work focusing on the convergence of satellite and terrestrial networks are presented. In contrast to the previous chapter, which deals with related work in other network domains, in this chapter research activities are addressed that share the same topic and (some of) the same research questions. The purpose of this chapter is twofold: On the one hand it underlines the fact that the topic of this thesis has not been completely addressed in previous work and the corresponding research questions of this thesis still need to be solved. On the other hand it serves as an input for Chapters 5 and 6 and their corresponding objectives.

The idea of integrating satellite into terrestrial networks has been present in articles for several years. Already in 2005 (Evans et al. 2005) discussed business and market observations for satellite networks in general. The conclusion is that integration of satellite networks and terrestrial networks is the key to success for satellite networks and that there is a mutual benefit for both the terrestrial and satellite world. In the meantime several projects and studies also deal with this topic. Interestingly, with the advent of 5G, satellite networks regained attention in the mobile world as well. Several projects and papers discuss how high-speed satellite connections can play a role in the 5G ecosystem, such as (Giambene et al. 2018, Gopal &

BenAmmar 2018, Boero et al. 2018). Form a 5G perspective satellite networks are relevant for several use cases: Firstly, as an opportunistic Back-haul network, i.e. to provide additional network capacity when it's needed. This is tremendously important in underserved areas. Secondly, satellites can increase service continuity in case of terrestrial network failures. Due to the always connected paradigm in 5G such an additional backup connectivity might become handy, also in urban areas. Thirdly, satellites are considered to provide 5G connectivity on aircraft, on vessels or in unserved remote areas, where no alternative connectivity is available. Given the increasing amount of IoT devices, autonomously driving vehicles and other devices that require permanent network connectivity, these aspects will become even more important (Gopal & BenAmmar 2018). Finally, another potential use case arises from edge caching and edge computing ideas, which are widely discussed in the 5G context. In this regard, satellites can support the required content distribution to the edges of the network (Wang et al. 2018, Kalantari et al. 2017). It should be noted that several of these use cases have a different focus than providing high quality last-mile access as considered in this thesis, however, several challenges overlap.

Hence, in the following, firstly existing research projects are discussed that deal with the convergence. Afterwards, other research work in this area is presented. Finally, a summary of the key aspects is provided.

## 3.1 Existing research projects

The SANSA project (Ziaragkas et al. 2017) aims at increasing mobile Back-haul capacity and resilience in both rural and urban scenarios. In order to realize these objectives SANSA envisages a self-configurable hybrid terrestrial-satellite Back-haul network that allows for a seamless integration of the satellite connections into the terrestrial segment. In this regard, they consider a shared spectrum between satellite and terrestrial segments. The focus of SANSA is clearly on the physical layers. By using intelligent spectrum sharing between a satellite and a terrestrial wireless Back-haul network, the spectral efficiency should be increased, which eventually should materialize in more Back-haul capacity and in a higher user data rate. It should be noted that SANSA considers a wireless terrestrial Back-haul network, as depicted in Figure 3.1 [1], which allows for dynamic reconfiguration of the topology. Link failures and congestion should be timely detected and mitigated by reconfiguration of the satellite and the terrestrial Back-haul resources. The assumption is that this will eventually result in a higher QoE.

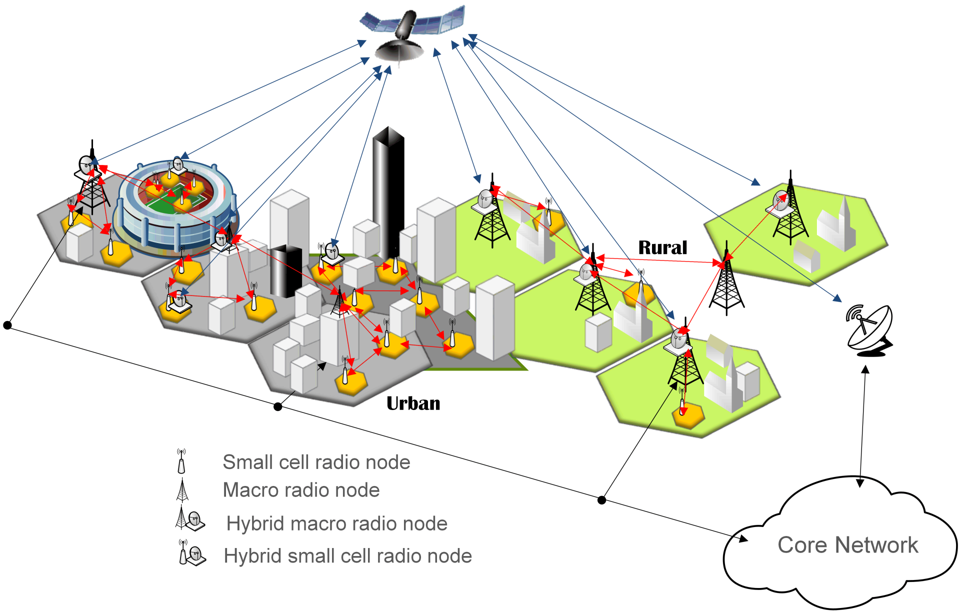



Figure 3.1: SANSA project network vision (Ziaragkas et al. 2017)

Architecture-wise SANSA adopts a typical LTE network architecture (3GPP TS36.300 2018), consisting of the RAN part with User Terminals and its eNodeB or HeNBs, the transport network and the core part, where the EPC is located. SANSA enhances this base architecture with additional components, namely the so-called Intelligent Back-haul node and the Hybrid Network Manager. The first is responsible for the routing decision in the transport network. This includes distributing the traffic between the terrestrial transport Back-haul network and the satellite network. It is also responsible for energy management and traffic classification. It usually connects terrestrially, i.e. wirelessly, to other nodes or via the satellite to the EPC. The latter is responsible for configuring the satellite resources and the topology formed by the Intelligent Back-haul nodes, i.e. by re-configuring

[1] https://sansa-h2020.eu/overview

the terrestrial radio modems and antennas. It has a global view on the network in order to manage the topology properly. It also holds the satellite ground segment and it must also react on link failures and re-configure the topology accordingly. To conclude, SANSA targets a different area than considered in this thesis, since a clear focus of the SANSA project lies in the physical layers, i.e. managing and mitigating interferences that are caused by the joint usage of spectrum between the satellite and terrestrial system or with other system, or smart beam-forming of antennas. With respect to the latter, SANSA exploits recent development in antenna beam-forming, first and foremost, so-called 3D beam-forming, which allows also for controlling the beam in elevation and azimuth. Moreover, the project also deals with full-duplex wireless transmission (Sharma et al. 2018) and energy-efficiency in Back-haul networks.

The Satellite and Terrestrial Network for 5G (SAT5G) project (Liolis et al. 2018) has different objectives and a more general focus on how satellites can play a role in 5G deployments. Several use cases with individual scenarios are identified. Among others, these include hybrid scenarios, where existing xDSL connections are complemented with a satellite connection to allow for better user experiences or edge caching scenarios, where local caches are filled by satellites connections to eventually lower the load in the terrestrial network. The SAT5G project identified challenges associated with these scenarios and with the successful integration of satellites into a 5G network. According to (Liolis et al. 2018) these are mainly the virtualization of the components of the satellite network, leveraging the 5G features in the satellite domain and the efficient exploitation of the satellite's broad- and multi-cast features to more efficiently offload traffic. The project has not yet been finished. However, recent publications, such as (Ge et al. 2019) show promising results with respect to delivering modern video services, particularly live streaming events, over a satellite link within a 5G network architecture. Unfortunately, this approach only exploits the satellite and neglects an existing terrestrial link to end users.

Similarly, an European Space Agency (ESA) study in 2015 on Service delivery over integrated satellite and terrestrial networks has been conducted that analyzes opportunities emerging from the integration of satellite and mobile networks. The study identified SDN as a key element for the successful integration of satellite and terrestrial networks, since interoperability and compatibility across the different segments is essential (Kapovits et al. 2014). Moreover, one particular use case identified by the study, which can benefit from high speed satellite networks, is connecting 5G RANs (Watts & Aliu 2014). The authors identified aligning the SDN northbound interfaces and SDN adoptions in the satellite work as work that needs to be done. However, the goal of this study was to identify open issues and potential next steps, without developing concrete solutions.

Integration of satellite and terrestrial networks to operate jointly is also discussed in an ETSI technical report (ETSI TR103 272 2015), which is an outcome from the BATS project (Peters et al. 2013). Its main goal is to increase the end user's QoE by using both terrestrial and satellite connections. The report discusses a general architecture how an integrated network might be structured. It described two devices, namely the Intelligent User Gateway and the Intelligent Network Gateway that are responsible for distributing traffic onto the available links, enforcing policies and traffic classification. While the Intelligent Network Gateway is in the operator's network, the Intelligent User Gateway remains in the home network of a customer, as depicted in Figure 3.2. The present approach maps different Class of Service onto QoS parameters and assume a resulting QoE.

For example, one factor, which impacts QoE for for streaming video flows, is the packet loss ratio. Hence, an optimal QoE can be achieved only with a loss ratio below 0.3%, an acceptable QoE level with a loss ration between 0.3 and 1.3% and a loss ratio above 1.3% will result in a not acceptable QoE. Obviously, he QoE of streaming video is also impacted by many other factors. Thus, a similar mapping needs to be done for other parameter as well as different Class of Service and other QoS parameters, i.e. latency and jitter. According to these technical reports, the distribution of traffic onto the path as well as other QoS operations, such as shaping or prioritization, should be done taking this mapping into account. Unfortunately, the report lacks concrete implementations. However, in (Ekmekcioglu et al. 2018) more details on the BATS implementation are given. Their approach is based on MPTCP proxies that are running on the Intelligent User Gateway and the Intelligent Network Gateway. The main idea is to exploit MPTCP to utilize the multiple path. A regular end-to-end TCP connection is broken up at Intelligent User Gateway and Intelligent Network Gateway, so that MPTCP can be used between these two devices. Moreover, based on the inter-arrival time of two consecutive packets of the same flow, packets are grouped into so-called objects. The proposed distribution algorithm then uses the link with the highest available bandwidth for objects larger than a certain threshold and the connection with the lowest RTT. Unfortunately, the results provided in (Ekmekcioglu et al. 2018) only show the impact of the BATS approach on the Mean Opinion Score (MOS), i.e. subjective tests that have been conducted with a small set of real users, but no other technical results. Furthermore, the tests conducted during the BATS project does not only rely on a narrow-band DSL connection and a broadband satellite, but also a third LTE connection. On the one hand, distributing traffic onto three connections increases the complexity even further, particularly given that mobile networks are a shared medium without dedicated bandwidth for each user. However, on the other hand, two terrestrial networks provide potentially more low-latency performance.

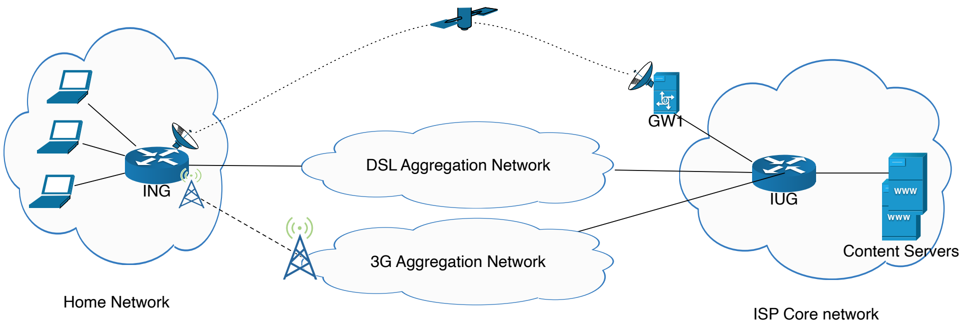



Figure 3.2: BATS network architecture (Peters et al. 2013)

The VIrtualized hybrid satellite-TerrestriAl systems for resilient and fLexible future networks (VITAL) project also focuses more on the layers above the physical layer. The goal of the project is, similar to SAT5G, to bring novel Network Function Visualization (NFV) concepts into the satellite domain as well as enabling software defined networking of satellite networks, so that eventually a satellite-network-as-a-service can be provided. That is, baseband and gateway functions, such as PEP, caching or compression, as well as other control functions required to operate a satellite network are virtualized and may run as virtualized entities on a cloud infrastructure. One challenge in this regard is to identify the functions that can be virtualized and those which cannot. Given that a lot of the functionality required to run a satellite network is often deployed on vendor specifics and not open standards, this challenge becomes even more difficult, but also quite important to e.g. allow for multi-vendor deployments (Ferrús et al. 2016). Moreover, the authors argued that SDN in hybrid satellite and terrestrial networks allow for a better media and content distribution than legacy technologies. That is, value added services, such as applying transcoders to certain media flows, can be easily deployed. Compared to SAT5G the scope is more focused on purely bringing SDN and NFV technologies into the satellite domain, while SAT5G also tackles challenges like edge caching.

Also, the ESA CloudSat project deals with SDN and NFV technologies with respect to its usage in satellite networks, as explained in (Gardikis et al. 2017) in detail. The authors present in their work the outcome of the project, which, among others, aims at determining the applicability of SDN and NFV technology in the satellite domain. They defined an integrated software-based satellite and terrestrial architecture and identified use cases, where SDN/NFV-enabled satellite networks can provide additional values. These are, for example, media distribution and CDN scenarios, bandwidth on demand and dynamic backhauling with edge processing. The latter assumes the extension of a terrestrial Back-haul with satellite capacity. According to (Gardikis et al. 2017), additional edge processing capability might allow for uses cases in the 5G context where operators might deploy (parts of) the EPC on the edge node. Moreover, the architecture proposed by the authors highly consists of individual orchestration components for the terrestrial and the satellite part. A Federated Manager brings both worlds together. The presented architecture enables the realization of the aforementioned use cases, yet the paper lacks implementation of required algorithms, e.g. when is traffic being re-routed or QoS classes applied. It should also be noted that the CloudSat project conducted a cost benefit analysis. According to this analysis, the introduction of SDN and NFV technologies in satellite networks can bring cost reductions of approximately 19%.

## 3.2 Other existing research work

Besides the work evolved from the aforementioned funded research projects, other studies exist, which is presented in the following.

Already in 2011, (Taleb et al. 2011) deal with scenarios of converged satellite and terrestrial networks. The authors identify associated issues, which include transmission efficiency, resource allocation and mobility management. Therefore, they present enhanced TCP methods as well as a cross-layer bandwidth allocation approach, which address the issue of varying physical layer capacity on satellite links caused by ACM.

An overview of issues related to integrated and hybrid satellite and terrestrial networks is provided in (Kota et al. 2011). The authors consider integrated networks, when satellite and terrestrial networks are operating on the same frequency, while hybrid satellite/terrestrial networks are interconnected but operate independently. The paper gives an overview of typical issues in such systems. These include physical layer issues, such as multiple-input and multiple-output (MIMO), but also resource management, handover issues or QoS problems. Finally, TE concepts are being discussed, which show that the satellite network can reduce the blocking on terrestrial links. However, in this regard the focus is on bandwidth management, without considering other QoS requirements of the traffic.

In (Zhang et al. 2014) the advantages of the convergence of broadcast and unicast (cellular) networks are discussed. Given that a lot of content requested by users in a network is actually the same, e.g. top 10 YouTube videos, the authors claim that it is often advisable to broadcast the content rather than transmitting it individually. This way, the load on unicast transmission networks can be reduced. However, in order to allow for broadcasting, the content needs to be requested relatively simultaneously. The authors identified technical challenges in this regard, which have not yet been solved: Firstly, correlated content that can actually be broadcasted needs to be detected. Secondly, situations when it is worth broadcasting content, i.e. the number of receivers is high enough, must be identified, and, finally, the content transmission needs to be synchronized.

In (Gopal & BenAmmar 2018) multi modem User Terminals are considered, i.e. by using Software Defined Radio (SDR) the modem can either send via the satellite or the terrestrial link on the physical layer, depending on the cost, system utilization and/or QoS. Moreover, (Gopal & BenAmmar 2018) also advocate for a convergence on all layer of 5G and satellite systems. For instance, the Xn interface in mobile networks that interconnects two eNodeBs to e.g. enhance the handover behavior can also be used to interconnect a satellite ground station with an eNodeB. Even physical and MAC might be compatible in many ways according to the authors. Similarly, (Du et al. 2018) considers the spectrum sharing aspect between terrestrial cellular and satellite networks. Moreover, the authors argue that the usage of SDN is a crucial design decision to overcome the slow configuration and inflexible traffic engineering of satellite networks and to realize their auction-based traffic offloading mechanism. The main idea is that satellite beams can "buy" the cooperation of mobile base station, which then avoid sending for a certain time slot on frequencies that interfere with the satellite communication. This approach requires a tight integration between the mobile and the satellite network operator. Unfortunately, the authors solely focus on optimizing the overall bandwidth that the system can transport, but do not consider different traffic classes and their requirements.

In (Boero et al. 2018) the authors focus on the integration between satellite and terrestrial systems from a networking point of view. That is, emerging 5G networks have a clear visualization paradigm, enabling network slicing and multi-tenancy. The authors argue that satellite networks have similar concepts. However, according to (Boero et al. 2018), in the satellite domain, these are typically proprietary and closed systems. More precisely, the authors advocate for employing NFV ideas from the 5G context also in the satellite domain, in order to enable flexible satellite network function in network. Moreover, (Boero et al. 2018) identifies several open issues with respect to the usage of SDN in satellite networks. Among others, these are the high latency and missing pieces in the OpenFlow protocol, which is considered as the de facto SDN standard. That is, in satellite networks additional information might be required to be transported via OpenFlow to a controller, which have not yet been considered. An use case that could benefit from such an approach is a handover in the satellite network.

Also the work of (Bertaux et al. 2015) acknowledges satellite and terrestrial network integration as a use case that can benefit from SDN approaches, since these ease capacity aggregation, i.e. multi-link transmission, and load balancing. Moreover, the authors highlight data flow identification, link monitoring as well as dynamic forwarding rules generation and update as essential requirements for an SDN-based satellite/terrestrial integrated network solution. Unfortunately, the publication only outlines use-cases where SDN (and NFV) can be beneficial but concrete solutions are out of scope.

The authors in (Giambene et al. 2018) present their envisioned integrated terrestrial-satellite architecture based on SDN and NFV approaches. They also argue that most of the components required to run a satellite network can be Virtual Network Functions (VNFs). Some remaining functions cannot be virtualized, which run on the satellite gateways. Moreover, typical control functions, such as routing or policy definitions should be realized within an SDN controller. One of the main differences to other work is that (Giambene et al. 2018) consider a satellite link that terminates on the MT. Given that the authors main contribution is a Network Coding approach that allows for splitting the packets of a single flow onto both the terrestrial and the satellite part, this seems reasonable since such an approach can then create more benefit. The network coding can be implemented as a shim layer between TCP and IP layer. By using network coding, the packets plus some additional redundant information are sent via both paths. Thus, the probability of successful transmission of a packet that can be received by the transport layer increases and also the load of the links is distributed. However, as all the traffic is split and no differentiation between flows is made, all flows experiences at least to a certain extent the high latency of a satellite link.

The idea of bringing SDN and NFV technologies into the satellite domain is also discussed in (Li et al. 2018). Similar to the VITAL project, the authors argue that satellite systems suffer from slow configuration and inflexible traffic engineering and that these issues can be addressed by the aforementioned technologies. Hence, they proposed to use SDN and NFV technologies to overcome this issues. However, even though a terrestrial connection exists in their discussions, the focus is clearly on large satellite constellations consisting of LEOs, MEOs and GEO satellites. Hence, the actual routing solutions the authors presented focus only on the routing within such a satellite network using Inter-satellite links (ISLs).

## 3.3 Conclusion of the state-of-the-art

Most of the recent research work dealing with convergence of satellite and terrestrial networks can be structured in two mayor classes. One class of publications deals with physical layer issues. This includes interference management and spectrum re-use between satellite and mobile or wireless

Back-haul networks. This also includes novel SDR-based User Terminals or eNodeBs that are able to communicate simultaneously to a satellite and a terrestrial mobile network. The second class is driven by the emerging SDN and NFV technologies. These publications analyze use-cases when the adoption of SDN and NFV in the satellite is beneficial to converge both kinds of networks. They share a common view that SDN and NFV adoption are essential for the integration of satellite networks into emerging 5G mobile networks. However, the research projects and the publications are either at an early stage, e.g. SAT5G project, or focuses on very specific parts, such as (Li et al. 2018) or (Giambene et al. 2018). Important questions often remain unanswered or are part of the future work, such as the location of a potential SDN controller or a clear definition of the required interface definitions or enhancements. A somewhat isolated yet comprehensive work emerged from the BATS project. A non-SDN-based architecture has been defined and tested. Unfortunately, the test lack some important reference values and the approach only works for TCP traffic.

What should also be noted, however, is that the amount of research activities dealing with the convergence of satellite and terrestrial networks in very recent 5G research projects shows the relevance of the integration. The lack of concrete solutions, however, also shows that major pieces towards a successful integration are still missing. In particular this hold true for the scenario considered in this thesis to exploit both parallel satellite and terrestrial connections.

Given that, that main insight gained from the analysis of this chapter with respect to this thesis and its objectives is that particularly the BATS work needs to be considered further in the next chapters. Moreover, also the fact that SDN is seen as an well-suited enabler by much of the existing work is an essential input to the design of the network architecture as presented in Chapter 6.

# Chapter 4

# Methodology

This chapter describes the methodology used throughout this thesis, in order to achieve its objectives. As described in Section 1.4 the main objective of this work is to design and evaluate a traffic offloading approach for converged satellite and terrestrial networks, so that the end users' QoE in undeserved areas can be increased.

Hence, this chapter firstly describes generally the Design Science Research (DSR) method, which is commonly used in Information and Communication Technology (ICT)-related research activities, and why it makes sense to use it throughout this work. Secondly, the concrete adoption of DSR for this thesis is presented and how knowledge has been gained. Finally, the chapter focuses on the technical methodology of the evaluation used in Chapters 5 and 6. Particularly, it explains why simulation methods are used and how the simulation environment is set up.

## 4.1 Design Science Research

For this study DSR methodologies has been adopted, which are described in-depth in (Vaishnavi & Kuechler 2015). While research in general can be defined as an activity that helps to understand a phenomenon, DSR considers by definition only phenomena that are artificially created and not naturally occurring. A phenomenon, in turn, can be defined as a set of behaviors of some entities and "understanding" in this context means to have *valid* knowledge, which allow for further predictions. Given that, research must lead to the contribution of new and valid knowledge.

Furthermore, design science can be clearly distinguished from natural science. While natural sciences deal with objects that are naturally occurring, e.g. in nature or in society, design science is about the design of man-made objects and phenomena, i.e. artifacts, which are made for a certain goal. Hence, design science can be considered as the knowledge to create artifacts that satisfy a given set of requirements and, thus, DSR can be defined as the process, which creates new knowledge in this regard using design, analysis, reflection and abstraction. Moreover, according to (Vaishnavi & Kuechler 2015), DSR can be distinguished from routine design efforts by the intellectual risk and the number of unknowns, i.e. missing knowledge, in the proposed design. That is, a design is innovative if the knowledge to create an artifact is not available. For such innovative designs, DSR methods can be applied to conduct the required research and close the knowledge gap. Hence, DSR can be defined as learning through building, which makes it highly applicable to the field of ICT and in particular Engineering and Computer Science disciplines as shown in (Hevner et al. 2004).

The typical DSR model is depicted in Figure 4.1. The first phase of the DSR process according to (Vaishnavi & Kuechler 2015) is the awareness of a problem, which might arise from multiple sources, e.g. new developments in industry or other research areas. The typical output of this phase is a research proposal that can be formulated formally or informally.

The second phase is the Suggestions phase, which is tightly connected to the fist phase. During this phase suggestions to solve the problem are discussed and collected, in order to develop a tentative design or an idea for the problem solution.

In the third phase, the Development phase, the actual artifact is developed. It should be noted that, according to (Vaishnavi & Kuechler 2015), the implementation of the artifact itself does not require beyond state-of-the-art methods. The novelty comes from the design and not the construction of the artifact.

Consequently, an Evaluation phase follows the Development phase. In this phase the artifact is tested and evaluated according to criteria that are either implicit or are defined during the first phase. Information gained from the evaluation of the artifact as well as information gained during the development in the previous phase are fed back to the problem awareness phase and help to refine the tentative design and hypothetical predication, which

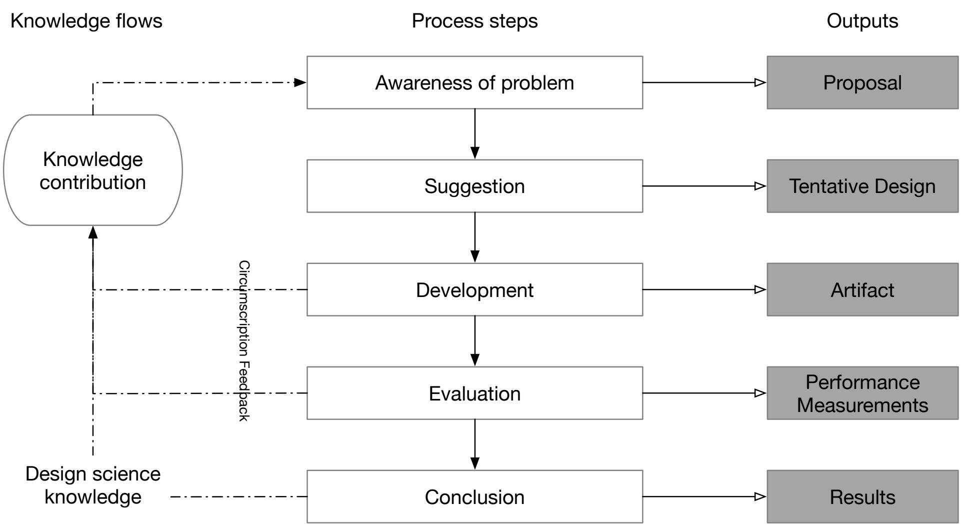



Figure 4.1: General DSR process model (Vaishnavi & Kuechler 2015)

have been made, in another iteration. Hence, new knowledge is created.

The final phase is the Conclusion phase in which the results of multiple iterations of the aforementioned process are considered as good enough. That is, either the gained knowledge is considered as firm or anomalous behavior has been identified that will be the subject of further research.

## 4.2 Adoption of DSR methodology

In the following the adoption of the aforementioned DSR methodology in this thesis and how it is used throughout this work is presented. In Figure 4.2 the adoption of the process is visually depicted.

During the first phase of Problem Awareness several research questions have been formulated in Section 1.4 and a general problem statement has been developed, which has also been published and presented at a research conference, in order to show its relevance (Niephaus et al. 2013). During the Suggestion phase an extensive review of the related work in other network domains (see Chapter 2) as well as the state-of-the-art (see Chapter 3) has been done. Based on this, firstly a tentative design for an offloading algorithm has been designed and during the following consecutive phases developed until a first evaluation. During the evaluation phase sophisti-

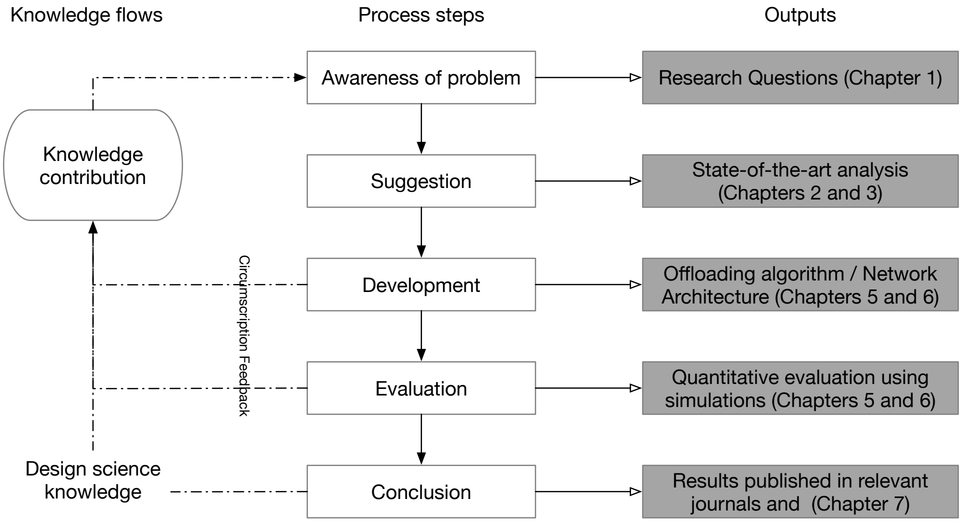



Figure 4.2: Adoption of the DSR process model for this thesis

cated simulations methods, as explained in Section 4.4, are used to gain results on its behavior. These findings yield to tentative conclusions and in an iterative process until the results are considered firm. The final outcome of this loop are different artifacts presented in Chapter 5, which are evaluated against each other, as well as submitted publications. Analogously, the DSR process has been used for the development of the network architecture artifact, yielding in Chapter 6 and (Niephaus et al. 2019). It should be noted that results gained from intermediate outcomes of the DSR process loop have been published and presented at scientific conferences (e.g. (Niephaus, Kretschmer, Ghinea & Hadzic 2015, Niephaus, Ghinea, Aliu, Hadzic & Kretschmer 2015, Niephaus et al. 2014)), so that the feedback of discussions during these conferences has also been fed back into process.

## 4.3 Relation to the BATS project

Considerable parts of the work leading to this thesis has been performed within the EU FP7 BATS project (Peters et al. 2013). Hence, the overall goal of BATS to optimize the users' QoE by combining limited capacity terrestrial links with Ultra High Throughput Satellite Systems as well as the goals of this thesis are aligned to a certain extent. Due to the amount of partners that participated in the project, the overall scope of the BATS project was broader and included several additional aspects, such as optimizing the physical layer of the satellite links, carbon footprint minimization or developing sustainable business model for integrated satellite and terrestrial networks.

It should be noted that the author's role in the project focused on the design of the network architecture as well as the development of the offloading approaches. As part of this work, the gap analysis of existing network architectures and mechanisms to be used in the considered scenario, as presented in Chapter 2, has been contributed, as well as significant input leading to the design of the BATS offloading approach, as presented in Chapter 3.1 and evaluated in Chapter 5.

The project developed this approach even further to operate with two terrestrial links, i.e. a DSL and a LTE connection, which obviously introduces further challenges. In contrast to that the focus of this study shifted towards evaluating and extending policy-based offloading approaches, since policy-based approaches, such as SIPTO, seemed to be a promising approach as well. This is is also shown in Chapter 5. Moreover, due to a strong participation of network operators in the project, the consideration regarding integration of terrestrial and satellite world in BATS strongly focused on high level management functions, such as provisioning or maintenance processes. This study has taken a different viewpoint and looks more into lower level functions and discusses in Chapter 6 how offloading decisions can be executed in a network. It should also be noted that the BATS project clearly focused on evaluating the effect of all the project's findings on the QoE by conducting subjective tests with real users. Instead of such a more macroscopic MOS evaluation, in this thesis specific and objectively measurable Key Performance Indicators (KPIs) are considered to validate different offloading techniques, so that the offloading effects on individual traffic flows can be more detailedly evaluated.

## 4.4 Quantitative Evaluation

Typically, designing and engineering ICT systems including protocols, algorithms and architectures requires at some point to understand their behavior on a micro and macro level, as well as to estimate their performance. In general, this can be done by using three different methodologies. The first option is to implement or prototype the systems in the real world and conduct experiments. The second option is to analytically evaluate the systems and the final option is to run simulations (Wehrle et al. 2010).

Given that satellite system are considered in this study, which haven't yet been deployed, and also given the overall size of the scenario considered in this work, an implementation of the artifact in a real satellite network and other larger scale terrestrial networks is unfeasible. Hence, other methods are required to evaluate the designed artifacts.

Moreover, various evaluation parameters, the high dynamism of the systems, etc., also renders the usage of analytical evaluations unfeasible, since the properties are often difficult to capture mathematically. This might lead to an unacceptable complexity of the analytic model, particularly in larger scale evaluations. Hence, simulations or emulations are the most promising option to perform a quantitative evaluation. Even though such methods inevitably introduce an abstraction when assessing an artifact, meaningful results can be created, if used properly (Tan et al. 2011). Large sets of parameters and different scenarios can be evaluated with a manageable amount of effort. Obviously, the accuracy of the underlying models in simulations and the used randomness highly impact the simulation the credibility of the results, as in-depth explained in (Pawlikowski et al. 2002, Sarkar & Gutierrez 2014). Particularly a good pseudo randomness is important when stochastic simulations are used. Besides that, also a suitable statistically analyses of the simulation output is necessary.

Moreover, one must distinguish between simulations and emulations. Both relay on models that mimic the real-world, yet differences exist. While simulations run along a simulated time axes, which allows for simulating long time periods in much shorter time, emulations are executed in real time. Emulations, on the one hand, typically makes the integration of real-world code fragments or traffic easier, but on the other hand might be impacted by side effects, such as a high CPU load on the emulation system, due to the real time approach (Horstmann et al. 2011). Both are more deeply discussed in the following.

### 4.4.1 Evaluation alternatives

Using emulations environments such as Linux NetEm (Jurgelionis et al. 2011) or NetEmu (Horstmann et al. 2011) seem to be one option that mimic a scenario required to evaluate the developed artifacts. Both allow for emulating to a certain extent channel conditions on a real or virtual link that behave like a satellite or e.g. a DSL link. More precisely, a certain fixed latency can be configured as well as certain packet loss patterns. However, what cannot be simulated easily is the complex media access patterns used on the satellite return link (see Section 1.1.1). Also, as previously mentioned, emulations do not allow for evaluating longer periods of time, as they do not have a simulated time but operate in real-time.

Simulators such as Network Simulator 3 (NS-3)[1], OPNET++ or OMNet are also frequently used nowadays to evaluate network protocols and architectures (Sarkar & Gutierrez 2014). All of them are Discrete-Event Network Simulators (Banks et al. 2009), which are typically used in the field of computer networks. In a discrete-event simulation the state of the simulation can only change at discrete points in time, i.e. the events. This simulation paradigm fits very well to computer networks, e.g. a packet is send in the network can be imitated as an event and a state change in the simulation (Gross & Güneş 2010). Obviously, discrete-event simulations also require appropriate analysis of the used models as well as the simulation output to provide credible results.

For this study NS-3 has been chosen for several reasons: First and foremost, a model for satellite networks is available. This model, which is called Satellite Network Simulator 3 (SNS3)[2], is the outcome of the ESA project "Development of an Open-Source, Modular and Flexible Satellite Network Simulator". It has been published and presented at several conferences, e.g. in (Puttonen et al. 2015, Puttonen et al. 2014, Hytönen et al. 2014) so that a certain level accuracy and sophistication for the underlying model can be assumed. Moreover, NS-3 is being used in many research papers to evaluate approaches, protocols or network architectures.

As explained in depth in (Riley & Henderson 2010), NS-3 provides simulation models for typical elements of a computer network: That is, network *nodes* simulate routers, switches and end-systems of a computer network. Network *devices* represent the physical connection of a node to a communication channel, e.g. the network interface card for a Carrier-Sense Multiple Access (CSMA) network or WLAN. A *channel* mimics the medium used

[1]https://www.nsnam.org
[2]https://www.sns3.org/

to send data between the devices, e.g. the wireless spectrum or a point-to-point copper cable. Furthermore, *packets*, *protocols* and *headers* implement various RFCs or other protocols. Moreover, its open architecture allows for adding easily additional functionalities and models.

## 4.5 Simulation Environment

In the following, the individual components of the used simulation environment are described in detail. The goal of simulations during this thesis is to eventually compare several system designs, i.e. the iteratively developed artifices of the converged satellite and terrestrial network architecture.

### 4.5.1 Satellite network simulation model

The satellite model provided by SNS3 encompasses all modules required to run a multi-spot beam GEO satellite network, which implements the DVB-S2 and DVB-RCS2 standards. That is, the satellite itself, the operator's Gateway, the User Terminals, as well as the NCC.

As explained in-depth in (Puttonen et al. 2014), SNS3 implements a reference satellite system that consists of a single satellite that covers Europe with 72 spot-beams using the Ka-band frequencies served by 5 Gateways. Two frequency bands of 2 GHz each are available for the communication between Gateways and satellite, up- and downlink. Due to assumed frequency reuse, each Gateway can use the same 2 GHz. Moreover, two frequency band of 500 MHz each are allocated for the communication between User Terminals and the satellite.

On the forward link, i.e. from a Gateway to the User Terminals, DVB-S2 Time Division Multiplex (TDM) is used. 1 GHz available bandwidth is divided into 16 carriers, each of which has a bandwidth of 125MHz and is mapped statically to a certain beam. A 32-Amplitude and phase-shift keying (APSK) modulation is simulated that is the most efficient MCS in DVB-S2 (ETSI EN302 307 2013). Specifically, the achievable data rate $DR$ on the down link if only a single User Terminal is served and no transmission errors occur can be calculated as follows:

$$DR = n \cdot CR \cdot SR \tag{4.1}$$

where $CR$ is the Forward Error Correction (FEC) rate, $SR$ represents the symbol rate and $n$ the number of bits per symbol, which depend on the MCS.

The symbol rate $SR$ can in turn be calculated as follows:

$$SR = \frac{BW}{1 + \text{roll-off-factor}} \tag{4.2}$$

where $BW$ represents the available bandwidth in Hz. The so-called roll-off factor describes how much of the "edges" of the actual spectrum can be used, given that the signal is filtered to avoid interference. The DVB-S2 standard specifies Roll-Off Factors of 0.2.

Moreover, assuming the specified FEC rate of 4/5, i.e. 0.8, and 32 APSK modulation, which encodes 5 bit/symbol can be used, the bandwidth is as follows:

$$SR = \frac{125\,\text{MHz}}{1 + 0.2} = 104,\overline{6} \cdot 10^6 symbols/s \tag{4.3}$$

$$DR = 5\,\text{bit} \cdot 0.8 \cdot 104,\overline{6} \cdot 10^6 symbols/s \approx 416\,\text{Mbit/s} \tag{4.4}$$

It should be noted that this data rate is a theoretical upper limit given the physical constraints. For practical use, additional protocol overhead, such as TCP and transmission errors need also be considered, which significantly reduces the bandwidth.

The satellite itself is modeled as a transparent bent-pipe. That is, the feeder link and the user link are directly mapped to each other, so that the signal is only amplified but not otherwise processed, i.e. the simulated satellite does not contain an onboard unit, which is common for most nowadays satellites.

It should be noted that for this thesis, an accurate satellite link is important, yet the focus is on the impact the satellite connection has on layer 3 and above. That is, the model must mimic at least the high fixed latency, the variance in medium access and the overall capacity. An accurate modeling of the interference of, for example, neighboring spot beams is not highly crucial.

However, SNS3 provides realistic antenna patterns, Signal-to-Noise Ratio (SNR) error files, positions of User Terminals, etc., in order to provide a realistic satellite system.

#### Return link

As described in Section 1.1.1, complex medium access control is required on the return link, i.e. from the end user towards the operator's edge, since multiple and highly distributed senders compete for the access on the same medium. Hence, bi-directional satellite systems based on DVB-RCS2 implement a DAMA process. As the standard only defines a basic structure and leaves the precise implementation of the capacity assignment to the implementer, real-world implementations are typically proprietary. Hence, in the simulations, the DAMA approach described in (de la Cuesta et al. 2013), which has been developed in the academic world, has been chosen, given that it is a well-known and frequently cited work, such as (Lee & Park 2019, Puttonen & Kurjenniemi 2016, Lee & Kim 2015).

However, even though (de la Cuesta et al. 2013) define the algorithm that assigns capacity based on the requests, they do not define which traffic is mapped onto which request class. More precisely, three request classes are defined, namely real-time, critical data and BE. While the handling of these different request classes are explained, the mapping of individual traffic flows (or IP packets) onto these classes is missing. Hence, in order to allow for at least some prioritization of traffic, the used SNS3 model performs the queuing as follows: First an IP packet is mapped onto one of four flow IDs based on its IP DSCP values (Babiarz et al. 2006). That is, the DSCP default value is mapped onto flow ID *BE*, all DSCP Assured Forwarding values, as well as the class selector values 1-4, which represent low-priority data, Operations Administration and Maintenance data, broadcast video and real-time interactive data, are mapped onto flow ID $AF$ and finally the DSCP Expedited Forwarding classes as well as class selector values 5-6, i.e. signaling and network control data, are mapped onto flow ID *EF*. These flow IDs, in turn, represent eventually different queues.

### 4.5.2 Terrestrial network simulation model

In order to mimic a terrestrial network, standard NS-3 simulation methods are used. More precisely, it is relied upon simulated point-to-point connections, running the well-known Point-to-Point Protocol (PPP) protocol (Simpson (Ed.) 1994). These simulated terrestrial connections can be configured with an arbitrary latency and bandwidth. It should be noted that the bandwidth can be configured for up- and downlink individually. If not otherwise mentioned, the bandwidth is configured to 1 Mbit/s on the downlink and 125 kbit/s for the uplink. This is important, given that asymmetric capacities between up and down-link holds true for most of the DSL and cable connections used in the last mile context (ITU-T 2009, Cable Television Laboratories 2019).

Although NS-3 makes some abstraction compared to a real link using the PPP protocol, such as missing authorization of data framing, these can be neglected for this particular use case, since they do not impact the actual characteristics of the link. Moreover, using a point-to-point connection to simulate the terrestrial traffic is reasonable, since a typical DSL connection relies on either Point-to-Point Protocol over Ethernet (PPPoE) or Point-to-Point Protocol over ATM (PPPoA) on the last mile.

### 4.5.3 Traffic simulation models

As explained in Section 2.6, modern networks carry traffic of various applications, predominantly video traffic and web traffic. It is therefore essential to model and generate such traffic accurately in simulations, in order to gain credible results when evaluating the performance of network protocols and architectures. This, however, requires that the relevant characteristics of the traffic are represented in the simulation model. Hence, in the following the different kinds of simulated network traffic used throughout the simulations are presented.

#### Web traffic

Web traffic, i.e. traffic generated by users browsing through the Internet, is obviously highly important and impacts significantly the end user's experience. Hence, it is crucial that the developed artifacts are tested with a traffic mix, which encompasses web traffic. In order to model web traffic, NS-3 relies on the Hypertext Transfer Protocol (HTTP) traffic model defined in various standards and guideline documents (NGMN Alliance 2008, 3GPP TR25.892 2004, IEEE 802.16m 2009). HTTP traffic is simulated by the so-called ON/OFF model. That is, during ON periods web traffic is requested by a client and transferred from the server. In contrast, the OFF period simulates times when users are reading the website. Furthermore, with each request clients first receive a main object, which can be considered as the main Hypertext Markup Language (HTML) content of a website. Afterwards, additional so-called embedded objects are requested by clients, as depicted in Figure 4.3. This simulates the structure of a real website, which usually consist of a main website and several embedded objects such as images, IFrames, JavaScripts or layout information.

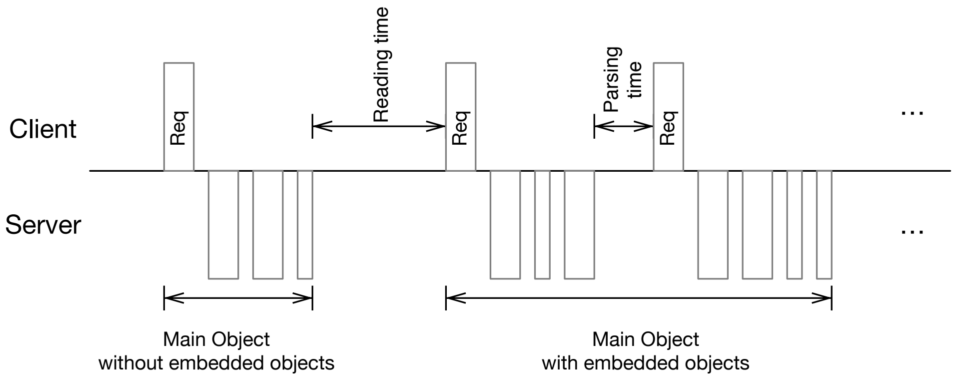
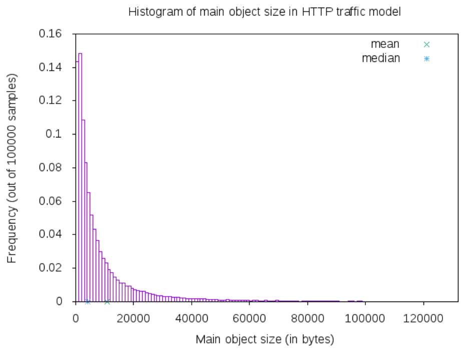



Figure 4.3: Simulated HTTP traffic pattern

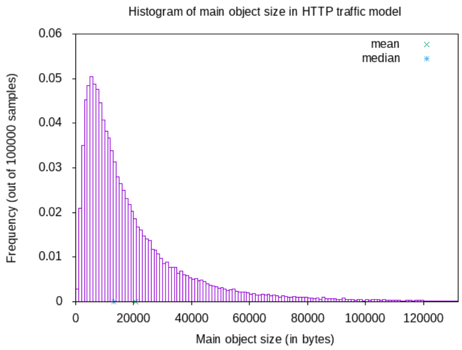
In Figure 4.4 the used random distribution of object sizes for the main object (4.4a) and the embedded objects (4.4c), as well as the number of embedded objects per simulated web-site request (4.4e) is depicted. The exact random parameters are explained in detail in the aforementioned publication and are shown in the Appendix A.1.

As can be seen in Figure 4.4 the mean size of the main object is approximately 10 kB, while the mean size of embedded objects is 7.7 kB. Moreover, the mean amount of embedded objects per website is 3.95. However, given that the used model is already a few years old, it can be assumed that modern web sites encompasses more and larger objects. In 2014 the authors of (Butkiewicz et al. 2014) have already measured that half of the web sites request over 40 objects. Even more recent figures are provided by the HTTP Archive[3], a non-profit organization that aims at providing performance information of web sites, such as page sizes or utilized technologies. These

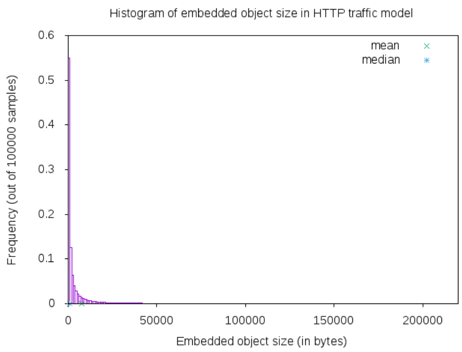
[3]https://httparchive.org



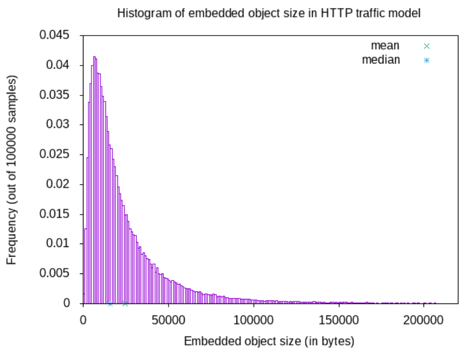

(a) Main object size histogram



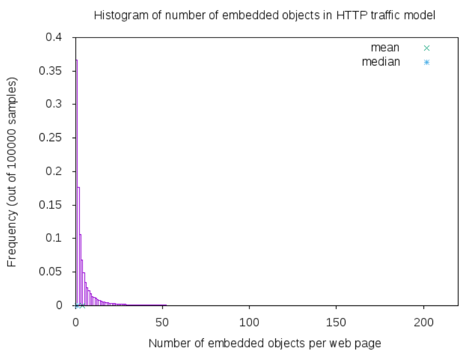

(b) Main object size modified histogram



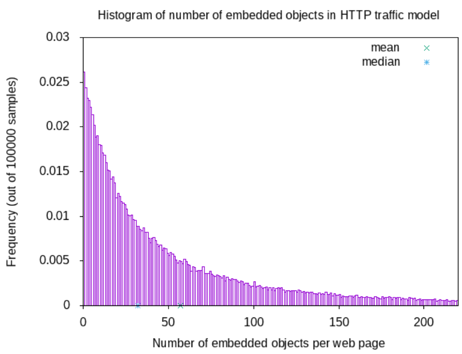

(c) Embedded object size histogram




(d) Embedded object size modified histogram




(e) Number of embedded objects histogram



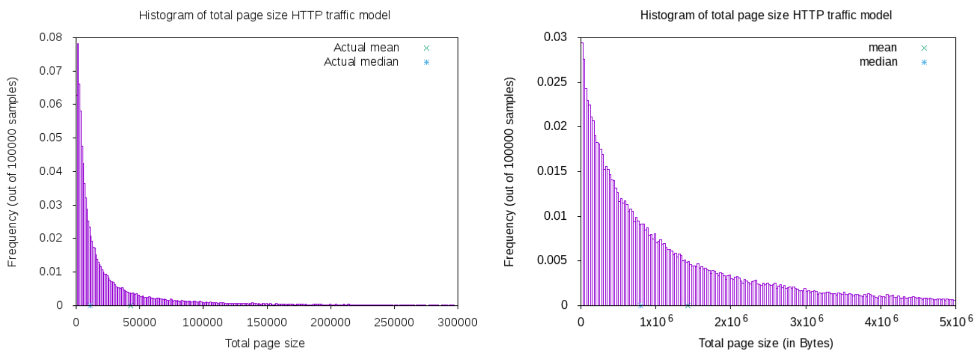

(f) Number of embedded objects modified histogram

Figure 4.4: Object size distribution

reports also show a significant increase of the page sizes. While in 2011, including all embedded objects, only 50% of web sites were larger than 467 kB, in 2019 50% are already larger than 1862 kB.

Unfortunately, HTTP Archive and other publications do not report any numbers on the size per object in general. Hence, all of the aforementioned random parameters, i.e. main object size, embedded object size and number of embedded objects, have been modified based on observations from the HTTP Archive. The resulting random distributions are depicted in Figures 4.4b, 4.4d and 4.4f, respectively. It should be noted, that these changes are to a certain extent just an educated guess. The main reason to adopt the traffic model is to reflect that modern web traffic consumes more bandwidth. However, since the focus is not to evaluate the performance of web traffic over the network but rather the performance of the offloading approach, this inaccuracy can be accepted. Furthermore, since multiple baseline simulations are conducted, as explained in the next chapter, not absolute figures, but relative changes between individual approaches are assessed.

Figure 4.5 shows the resulting total page size distribution for simulated web traffic. Similarly, to the previous figure, 4.5a shows the distribution with the parameters defined in the standard in the first place, while 4.5b presents the distribution of the page sizes with the modified parameters.




(a) Total page size (b) Total page size modfied parameter

Figure 4.5: Total page size distribution

### Web traffic KPIs

In order to evaluate the quality with which the web traffic is received, the Page Load Time is chosen as the major KPI for web traffic and is used throughout the simulations conducted in Chapter 5. That is, the time required to transmit the complete website, including its embedded object is measured. According to (ITU-T 2011, Google 2017) a low Page Load Time is crucial for a good user experience. Moreover, as explained in (Google 2017) the probability of user bounce from the web site increases by 37% if the Page Load Time is between 1 s and 3 s, while it increases by 123% if the Page Load Time is between 1 s and 10 s.

### Video traffic

Of a similar importance to web traffic in nowadays networks, given its amount, is video traffic. As explained in Section 2.6, this applies in particular for so-called Near Real-Time Video (NRTV) traffic, as generated by e.g. Netflix, Hulu or YouTube. In order to mimic this kind of traffic, the traffic simulator of the SNS3 module is used, which is also based on the same standards (NGMN Alliance 2008, 3GPP TR25.892 2004, IEEE 802.16m 2009). As described in (Puttonen et al. 2016), it relies on the TCP protocol and generates traffic between server and client upon request of the client. After a successful transmission of a video the server closes the connection and the client has to request a new video.

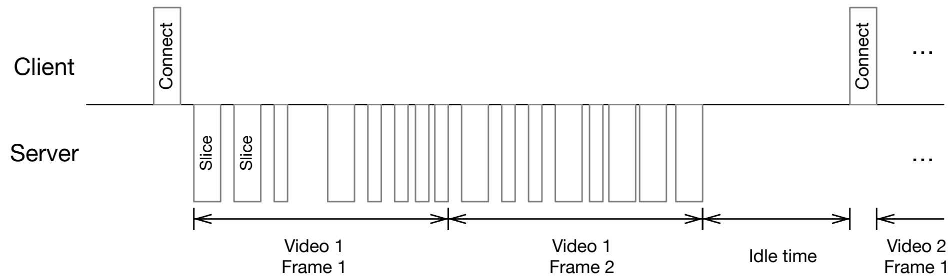
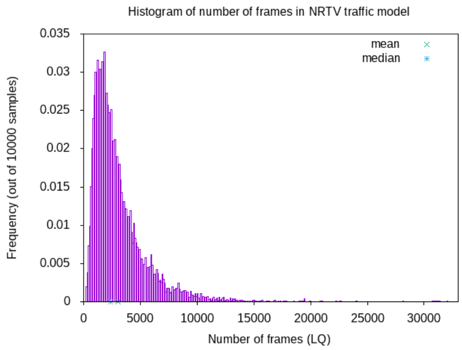


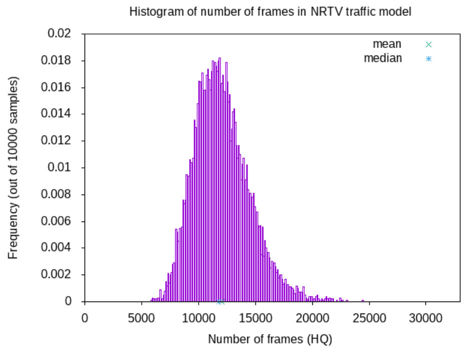

Figure 4.6: Simulated NRTV traffic pattern

In Figure 4.5.3 the structure of NRTV traffic is depicted. A simulated video consists of an arbitrary number of frames. Frames have a fixed length, which is configured in the aforementioned standards to 100 ms, so that the traffic is equivalent to a video of 10 frames per second. Each frame, in turn, consist of 8 packets, which are called slices. The size of a slice is randomly distributed.

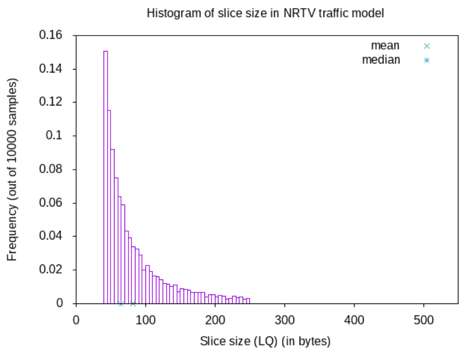
Furthermore, Figure 4.7 shows the random distributions used to determine number of frames per video, (4.7a), slice sizes (4.7c) and idle times (4.7e) between two consecutive videos, following the aforementioned model. Analogously to HTTP traffic, these parameters can be considered outdated.



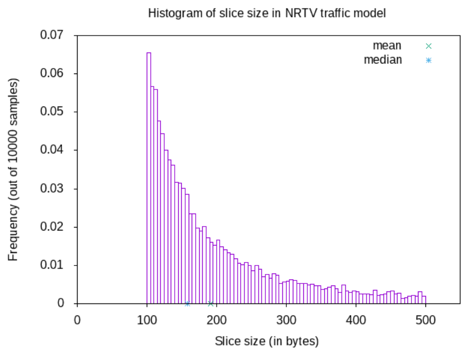

(a) Number of frames



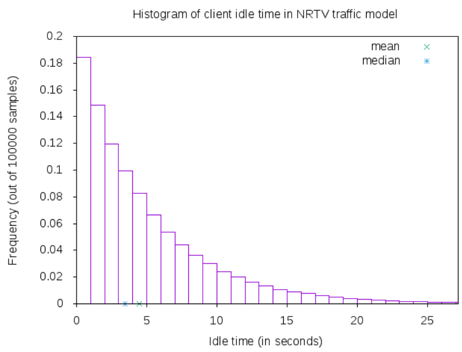

(b) Number of frames with modified paramters




(c) Slice sizes




(d) Slice sizes with modified parameters



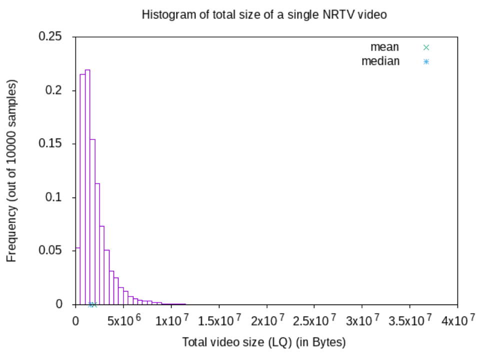

(e) Idle time

Figure 4.7: Simulated video parameter

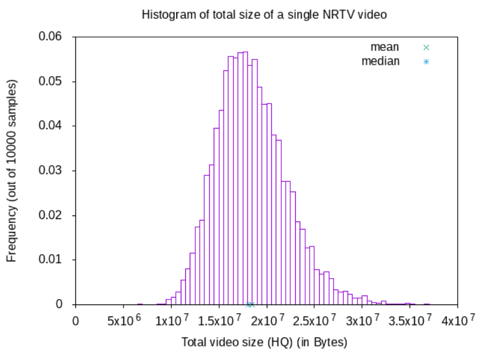
Due to increases in video resolutions, different codecs and more available capacity, the frame rate and slice size increased, which is also explained in Section 2.6. Hence, modifications have been made here as well to reflect the higher bandwidth required by modern videos.

The resulting distribution for the used random parameters are also shown in Figure 4.7, i.e. Figure 4.7b and Figure 4.7d depict the changes in the random distribution for number of frames per video and the size of slices, respectively. It should be noted, that the idle time parameter has not been modified. Moreover, also the length of frames has been changed from 100 ms to 40 ms, leading to a rate of 25 frames per second, which can occur in modern videos.

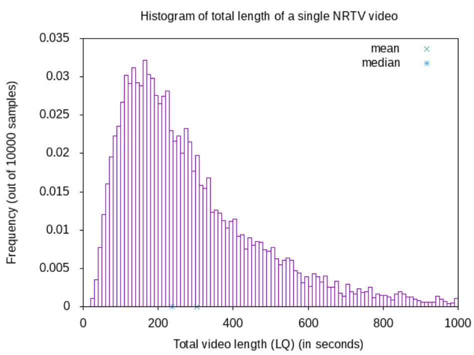
Using these random parameters, the following histograms have been created and depicted in Figure 4.8 to show the distribution of both length (4.8c and 4.8d) as well as size (4.8a and 4.8b) of the simulated videos.



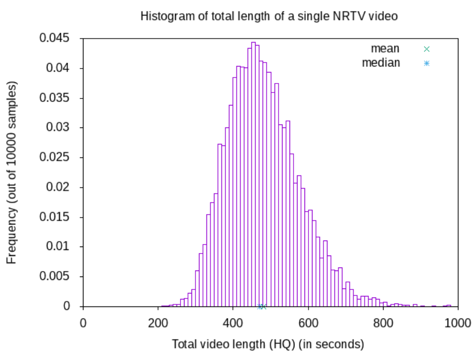

(a) Total video size




(b) Total video size with modfied parameter




(c) Total video length




(d) Total video length with modfied parameter

Figure 4.8: Video size and length distribution

Also, likewise to web traffic, the modification made in the random parameters of the simulated traffic are not necessarily most accurate. Given that here the focus is also not to evaluate the performance of the video traffic but rather the behavior of the offloading approach, the inaccuracy can be accepted here as well. However, nowadays video quality is changed dynamically depending on the users bandwidth. In order to reflect this behavior, the video model has been extended. That is, it can be switched dynamically between the modified and unmodified parameters, leading to low and high quality simulated videos.

Moreover, in addition to the aforementioned changes to the model, the concept of a playout buffer, i.e. a cache, has been added to the model as well, given that such buffers are quite common and allow for mitigating the effects of jitter in the network. The buffer stores the amount of frames successfully transmitted and "plays out" a frame at every frame interval, i.e. every 40 ms in the high quality stream and every 100 ms in the low quality stream. That is, the number of frames in the buffer is decremented by one every frame interval. Hence, the video buffer that is regularly used with contemporary video playback can also be simulated. The resulting finite state machine of the video traffic model is depicted in Figure 4.9. The exact used random parameters are shown in Appendix A.2.

Furthermore, the filling level of the buffer can be monitored and traced in the simulation, so that empty buffer events can be tracked. In the real world such an event would lead to a temporarily stopped video until the buffer has been filled up again. It should be noted, that the simulated video playback starts after configurable period of the video is cached.

To conclude, even though some abstraction compared to traffic of real video services such as You-Tube or Netflix has been made, however, the relevant aspects are identical in the traffic model used in the simulations, namely the TCP transport model and that the video is cached before the playback starts and the quality of the video can be changed.

#### Video traffic KPIs

Similar to the Page Load Time definition as KPI for web traffic, KPIs need to be defined for video traffic as well. Obviously, video traffic is not primarily impacted by the time that is required to transmit the video to the user, but rather a high video quality and a constant playback. Hence, as KPI the number of cache empty events is defined, since in such an event the video playback stops. It is acknowledged that the QoE of real video applications is actually impacted by many other factors as well. Starting from the size and the appearance of the video screen to network related parameters, such

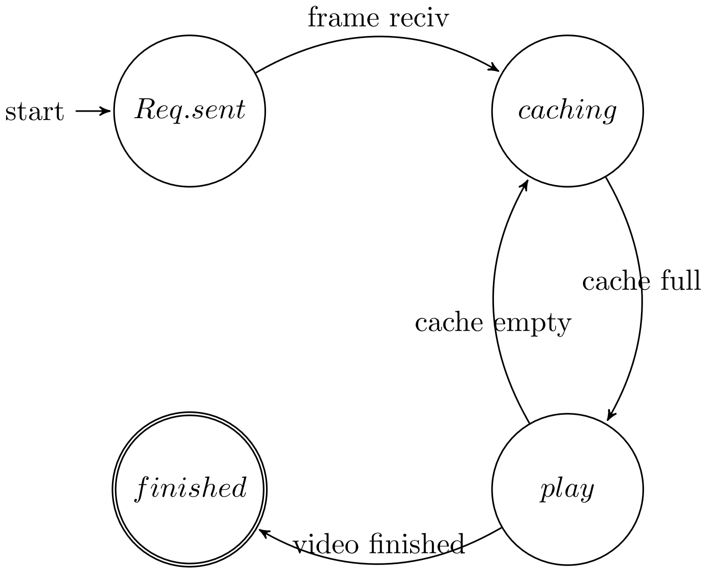


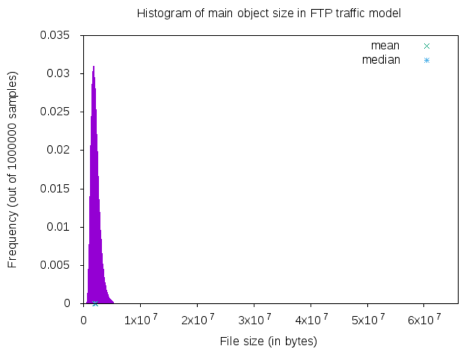

Figure 4.9: video traffic model state machine

as amount and duration of empty cache events or even the time it takes to start the video. However, in order to allow a systematic comparison of different offloading approaches, in this work only the number of cache empty events, i.e. re-caching events, is considered, as an unintentional stop of a video strongly decreases the QoE.

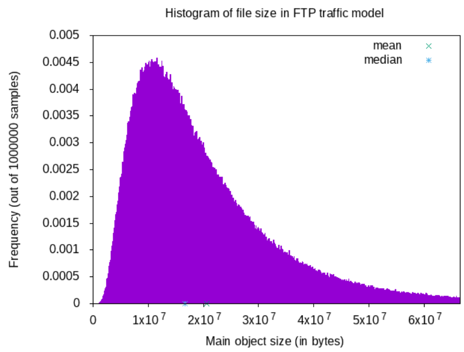
### File-transfer traffic

In addition to video and web traffic, also file-transfer traffic needs to be simulated, given its share in the total amount of Internet traffic, as explained in Section 2.6. This kind of traffic should mimic downloads of large files, which typically occur when software updates, etc. are performed. For example, if a new iPhone, Android or Windows version is available. Again, as a basis the model described in (NGMN Alliance 2008) is used. File-transfer traffic is modeled similar to HTTP traffic, i.e. it is also modeled as an On-Off application. The major difference is that a file-transfer only consists of a single file object that is being transmitted.

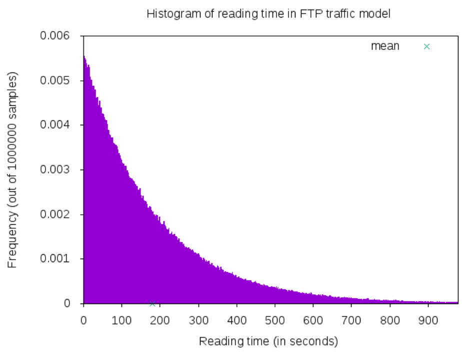
Again, the study described in this thesis has relied on (NGMN Alliance 2008, 3GPP TR25.892 2004, IEEE 802.16m 2009), which also specified parameters for File Transfer Protocol (FTP) traffic models. The characteristic of file-transfer traffic is defined by two parameters, namely the size of the transmitted files and the time between two file transfers, the so-called reading time. The random distributions used to simulate these parameters are depicted in Figure 4.10. Similar to the previously presented web and video traffic, modifications have been made to file size parameter to mimic contemporary traffic more realistically. The impact of these modifications is depicted in Figure 4.10b, which shows the resulting random distribution. The reading time has also been extended and its random distribution is shown in Figure 4.10c and 4.10d, respectively. The exact random distributions and parameters are listed in Appendix A.3.



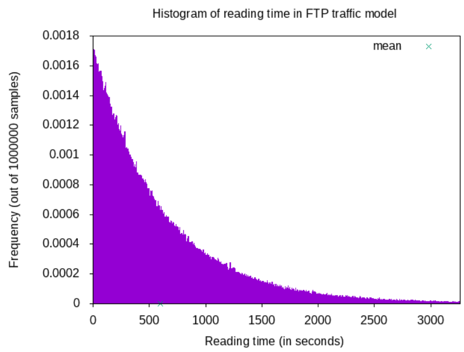

(a) File size distribution




(b) File size distribution with modified paramters




(c) Idle time




(d) Idle time with modified paramters

Figure 4.10: Simulated file-transfer parameter

Identical to the used web and video traffic models, the modifications might not be absolutely accurate in reflecting contemporary file transfers, but are sufficient for the aforementioned reasons, given that offloading behavior should be investigated.

#### File-transfer KPIs

Likewise, as for web and video traffic, KPIs need to be defined for file-transfer traffic as well. It is obvious that the time required to transfer the file is the most important KPI, assuming the transmission does not arbitrarily break down. Hence, the file-transfer time is used as a metric, which is the time period that starts when the client send its request for a file and ends once the file has been completely transmitted from the server to the client.

### 4.5.4 Structure of simulations

Figure 4.11 gives a detailed overview on an exemplary NS-3 simulation used in this work. A configurable amount of clients, representing the end user systems, are connected via a simulated CSMA network to an extended Customer premises equipment. This extended Customer premises equipment provides the connection between the terrestrial and satellite network. The connection between client and extended Customer premises equipment is configured to have no latency and a bandwidth of 100 Gbit/s in order to avoid any side effects in the results due to congestion in the home network. These components basically form the household site as explained in Section 1.1.3. The other edge of the simulation, the operator site, consists of the simulated server nodes that deliver the applications, i.e. generate the traffic. Similar to the clients, these servers are connected via a CSMA network to a Concentrator node. The concentrator is the counterpart to the extended Customer premises equipment and also provides connectivity to satellite and terrestrial network. The connection to the satellite network is via a simulated satellite gateway. It should be noted that the relevance of such a Concentrator node is further discussed in Section 6.1. Since the satellite is just a bent-pipe it appears in this simulation as the beam channel between the extended Customer premises equipment and the satellite Gateway. Both the network connecting the Gateway with the Concentrator as well the one connecting the Concentrator with the servers are also simulated Ethernet connections with no latency and 100 Gbit/s bandwidth, again in order to avoid side effects. Moreover, the extended Customer premises equipment are also directly connected to the Concentrator via a peer-to-peer (P2P) channel, in order to simulate the terrestrial network.

Between clients and servers web video and file-transfer traffic can be simulated, as previously described. The amount of parallel running applications can be configured. Each client node exchanges just one kind of traffic with a dedicated server. Thus, if e.g. three web traffic clients and two video clients should be active, in total five client nodes and five server nodes are created during the simulation.

Such a setup is adequate for the further simulation conducted in the course of this thesis, since it focuses on the relevant components and relies on designs used already in other research work (Puttonen et al. 2019, Abdelsalam et al. 2017, Artiga et al. 2016). Moreover, it can for example be easily extended to simulate more households or use different parameter for terrestrial or satellite network. That is, larger scale setups can also be simulated using this setup.

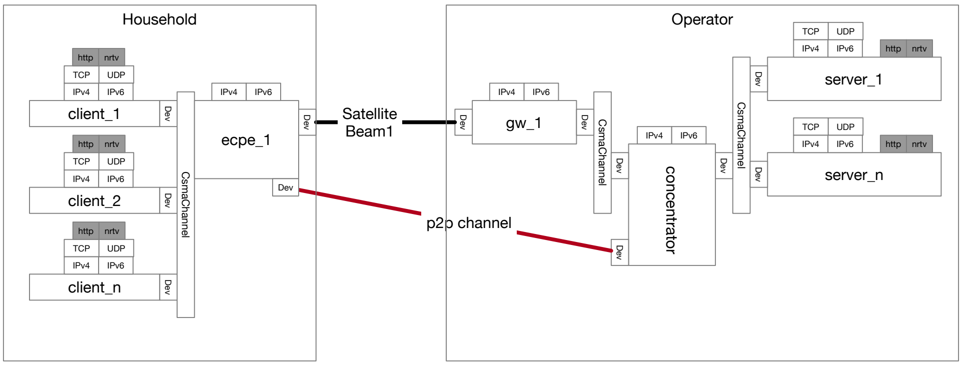



Figure 4.11: Detailed overview of simulation architecture

### 4.5.5 Randomness in start of applications

Multiple real users will not start all their application at exactly the same point in time, but rather arbitrarily. In order to mimic this fact, if multiple clients and traffic streams are simulated, randomness during the start-up is introduced. That is, if say four video traffic clients are simulated, they start not at the same time, but with a random jitter. The length of the jitter is exponentially distributed, depending on the length of the overall simulation, i.e. the mean is 10% of the simulation time and the upper bound is 80% of the simulation time. The random distribution of the jitter during the application start is depicted in Figure 4.12 for a simulation time of one hour.

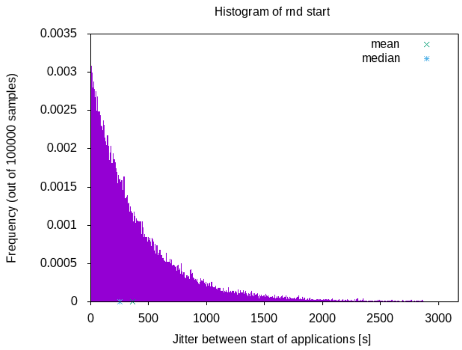



Figure 4.12: Random start jitter

### 4.5.6 Overview of the used models and limitations of the simulation

To conclude this chapter, the relevant used simulation models and their limitations are summarized in the following:

- The Point-to-point link model is used to model the terrestrial connection. Its limitation is mainly missing authentication mechanisms, which are not relevant for this thesis.
- The Csma Network model is used to simulate the connection network at the edges of the network. The model has certain limitations. First and foremost, it does not allow for full duplex transmissions. To mitigate these limitations the capacity has been configured to 10 Gbit/s with no latency. Hence, as the focus is on the satellite and terrestrial network, any impacts on the results from his model should be avoided, which is ensured by the configuration.
- The Satellite model simulates the satellite as a geostationary satellite and transparent bent pipe. The limitations are that it misses IPv6 support and configured reference system that covers Europe. Both limitations are not relevant in this context.
- The traffic models, namely the HTTP, video and file-transfer models are used to generate simulated traffic pattern. All models are based on well-known, yet relatively old standards or normative documents, leading to a difference between the generated traffic patterns and current traffic. However, the main difference is an increased number of

packets or higher amount of data. In order to mitigate this, the underlying random distributions have been adopted, so that the actual method how to generate traffic and, thus its characteristic, is kept but the amount of data is increased.

It should also be noted that the considered KPIs do not represent the QoE perceived by the users, but are rather major impact factors of it. However, those KPIs have been selected as they can be measured by simulations, whilst more comprehensive metrics, such as the MOS, often require real users and standardized testing environments, which goes beyond the scope of this thesis.

In the following chapters, the outlined simulation and testing environment are used to conduct thorough tests of various offloading approaches.

## 4.6 Conclusion of the used methodology

In the first part of this chapter, DSR and its adoption throughout this thesis has been explained. It is shown why it make sense to use it and how knowledge is gained in this work, namely by iteratively designing and evaluating new artifacts. In second part of the chapter it is shown that simulation are an adequate method to evaluate the artifacts and designs developed in the following chapters. Moreover, it also also explained that the used simulation environment and models are sufficient to provide meaningful results, so that (particularly the last two) objectives presented in Section 1.4 can be achieved.

# Chapter 5

# Evaluation of Offloading algorithms

Following the objectives of the thesis, in this chapter different offloading approaches that aim at distributing traffic between the terrestrial and the satellite link are designed and evaluated. Throughout this chapter, the evaluation methodology as described in Chapter 4 is used. Furthermore, input from Chapters 2 and 3 is also feed into the design of the offloading approaches.

More precisely, three offloading approaches are evaluated: Firstly, a static approach that simply directs all traffic from the Internet towards the end user (downlink) to the satellite connection, secondly the approach developed by the BATS project, as presented in Section 3.1, and finally a novel approach utilizing queue sizes and policies to decide with traffic should be offloaded. The first has been selected as it is a rather simple approach, which can be implemented relatively easy and was already commercially available. The second has been chosen due to the outcome of the state of the art analysis in Chapter 3, as the most promising approach,while the third is designed based on a policy-based offloading approach, as it is being used in e.g. SIPTO or PMIPv6. This seems reasonable based on the gap analysis shown in Chapter 2.

Furthermore, first baseline simulations are conducted in scenarios without any offloading, in order to have results to compare the offloading approaches to. Afterwards, three offloading approaches are tested.

Both baseline tests as well as the tests with each offloading approach encompass two phases. During the first phase, only a single application, i.e. traffic class, is used, but the amount of clients is varied. In the second phase, multiple applications running in parallel are tested.

## 5.1 Baseline simulations

As explained in Chapters 2 and 3, besides the BATS approach (see Section 3.1) currently there is neither a commercial or academic solution available that performs offloading of traffic to a parallel satellite link and existing approaches from other domains e.g. mobile networks cannot be simply used in such a scenario. Even the findings from the BATS projects with its aligned goals cannot be used. The project focused on performing subjective test and evaluating the impact on the MOS, which is a different approach as explained in Chapter 4.3.Hence, only very limited findings are available that can serve as a baseline to compare any results measured in simulation to. Given that, the performance and behavior of different traffic patterns over different connections is simulated, so that baseline results are available and the impact of traffic offloading approaches can be evaluated.

Firstly, the behavior of each kind of traffic over a simulated sole broadband terrestrial connection (100 Mbit/s), a sole narrow-band terrestrial connection (1 Mbit/s) and a sole satellite connection is simulated, as described in Section 4.5.1, will be tested individually. This way, the simulation results of more complex scenarios can be better assessed. In particular, absolute numbers need not be evaluated but relative changes between the individual approaches can be determined, which is important due to the assumptions being made in simulated traffic generation, as explained in Section 4.5.3. Moreover, in the first phase, only individual traffic types, e.g. just web traffic or just video traffic, are tested. Afterwards two traffic scenarios, which are later explained in Section 5.1.4 are defined and tested.

Furthermore, in all scenarios the satellite link is configured to be ideal. That is, the satellite connection is free of transmission errors and a MCS of 32 APSK with FEC overhead of 4/5 can be used, so that approximately 400 Mbit/s for a single User Terminal per spot beam is possible, without the overhead created by higher layer protocols, such as TCP or UDP. The terrestrial connection is also configured to provide an error-free link. Its down-link bandwidth is configured to 1 Mbit/s or 100 Mbit/s and its up-link bandwidth to 0.125 Mbit/s and 10 Mbit/s, respectively.

Moreover, during these simulations the TCP buffer sizes are optimized, in order to compensate for the lack of a PEP in the simulations. That is, the size of the sending and receiving buffer of each server and client is configured to the bandwidth-delay product of the corresponding link. For example, the satellite link has a bandwidth of 400 Mbit/s and a RTT of approximately 560 ms. Hence, the TCP buffer sizes in this setup are as follows:

$$\begin{aligned} \text{buf} &= 400\,\text{Mbit/s} \cdot 560\,\text{ms} && (5.1)\\ &= 50\,\text{MB/s} \cdot 560\,\text{ms} && (5.2)\\ &= 28\,\text{MB}. && (5.3) \end{aligned}$$

Furthermore, the simulations are set up to simulate a single household with an increasing number of clients creating traffic, i.e. starting from a single client up to 15 clients. The simulation time is one hour, in order to have a reasonable number of traffic samples. By focusing on a single household instead of considering an entire network, the evaluation can be focused on the actual offloading algorithm. Furthermore, 15 clients acting in parallel in a household seems also reasonable, given the considered scenario described in Section 1.1.3.

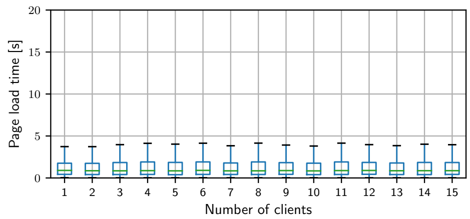
Finally, it should be noted that each simulation is repeated five times, every time with a different random seed to allow for independent replications of the same experiment.

### 5.1.1 Web traffic scenario

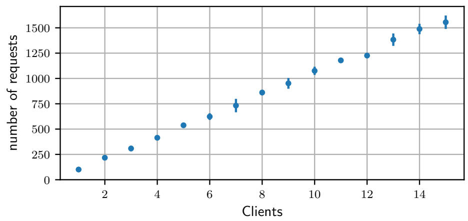
First, the performance of the web traffic is evaluated. That is, as explained in Section 4.5.3, Page Load Time, as a major KPI for web traffic is investigated.

In Figure 5.1 the results of the so-called baseline test, i.e. tests in which only a sole link is used are depicted. Box-plots of the measured Page Load Times are shown in Figures 5.1a, 5.1c and 5.1e for the different amount of clients. It should be noted that the Interquartile ranges of these box-plots are 1.5 and outliers are not shown for the sake of y-axis scaling. In Figures 5.1b, 5.1d and 5.1f the total number of requests that could be made by the clients are shown.

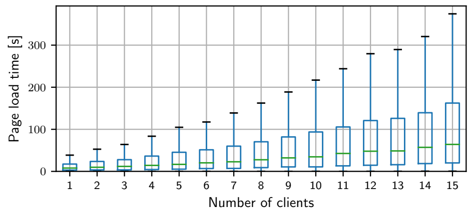
The results gained from this simulation match the expectations. That is,



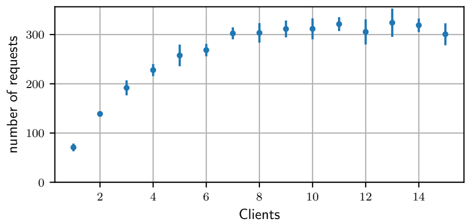

(a) PLTs ( 100 Mbit/s terrestrial link)



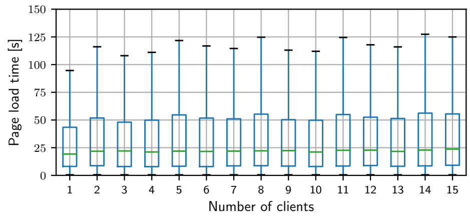

(b) number of requests (100 Mbit/s terrestrial link)



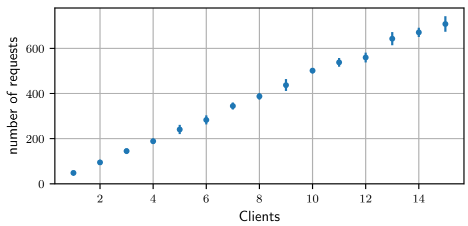

(c) PLTs (1 Mbit/s terrestrial link)




(d) number of requests (1 Mbit/s terrestrial link)




(e) PLTs (satellite link)




(f) number of requests (satellite link)

Figure 5.1: PLTs for different setups

a 100 Mbit/s terrestrial broadband connections performs significantly better than both the narrow-band terrestrial connection as well as the satellite connection. This can easily seen in the Page Load Time box-plots.

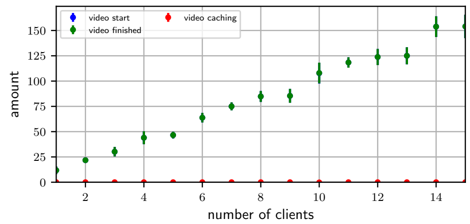
While with the broadband connection all simulated web pages are transmitted within 9 s and the median is 0.8 s, independent of the amount of clients, the median Page Load Time over the satellite link is between 14 s and 29 s. This is obviously due to the high fixed latency on the satellite link. Moreover, the narrow-band terrestrial link is clearly overloaded, particularly if the amount of clients increases, as can be seen in Figure 5.1d: while the amount of total web page request being made within the simulation time frame increases linearly with the number of clients in the broadband and the satellite case, the number runs into in upper limit for the narrow-band case. Also the difference between a single client and 15 clients shown in the Page Load Time box-plot is obvious, which shows that the amount of requests/clients impact the performance. Moreover, while over the narrow-band terrestrial connection the performance is clearly decreasing with the increasing number of clients, it remains constant over the satellite link. That is, while the Page Load Times are significantly higher for 15 clients than for a single client, when the narrow-band terrestrial link is used, there is not much difference between 1 and 15 clients over the satellite link, given that on the latter bandwidth is sufficiently available and only the latency negatively impact the Page Load Time. However, due to that, the performance for a very limited amount of clients is better in this test over a narrow-band terrestrial link than the broadband satellite link.

### 5.1.2 Video traffic scenario

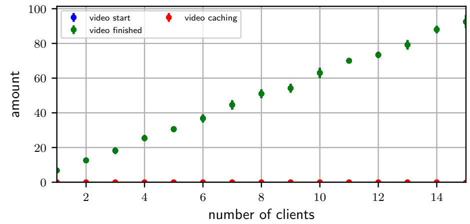
Similar to the test with the web traffic, similar tests have been conducted to evaluate the behavior of NRTV traffic. The simulation setup is identical to the web traffic as described in the previous section. However, the evaluated KPIs are different, since obviously Page Load Times do not represent a meaningful KPI for video traffic, as explained in Section 4.5.3. Thus, the unplanned video stop events are evaluated. Such events are generated when the buffer on the client has run empty and needs to be filled again during a regular playback.

The simulations have been conducted with both low video quality and high video quality (see Section 4.5.3). The video cache size has been configured to five seconds, so that some jitter can be compensated for, but the buffer can be relatively fast re-populated if e.g. the user jumps to a certain position in the video.

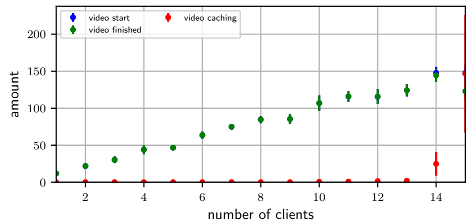
The results, depending on the amount of requesting clients, are shown in Figure 5.2. The red graphs represent the number of events when the video cache has run empty, which leads to a stop in the video playback and a negative user experience. The green graphs show the amount of videos successfully finished in the given simulation time, while the blue graphs depict the number of videos started during that period. A difference between both graphs indicates that a video could not be successfully finished before the next video is requested. The error bars indicate the standard deviation.



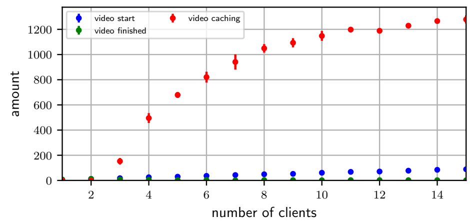

(a) Low quality (100 Mbit/s terrestrial link)



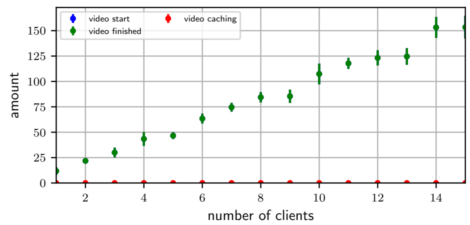

(b) High quality (100 Mbit/s terrestrial link)



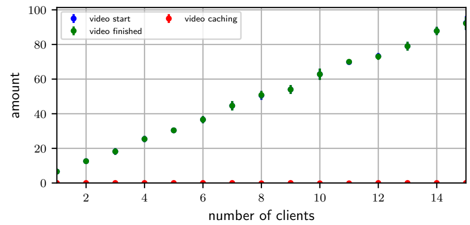

(c) Low quality (1 Mbit/s terrestrial link)




(d) High quality (1 Mbit/s terrestrial link)




(e) Low quality (satellite link)



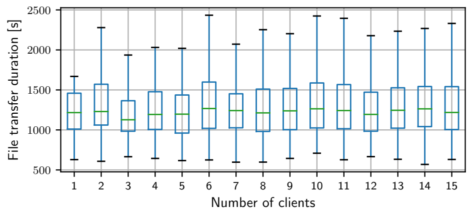

(f) How quality (satellite link)

Figure 5.2: Video stop events

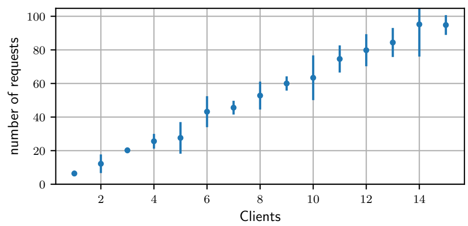
As shown in Figure 5.2a, a broadband terrestrial connection is able to cope with up to 15 clients watching video simultaneously, regardless whether high or low video quality is used. In contrast, the narrow-band terrestrial connection, which just provides 1 Mbit/s in the down-link, clearly shows its limits. During the simulation empty cache events already occur with 14 or more clients of low quality videos are used and 3 or more clients for the high quality videos, respectively. This is again due to the limited bandwidth. In contrast, the satellite performs in this scenario similarly to the broadband terrestrial connection. Due to the modified TCP buffers, the available bandwidth can be used and the video cache of five seconds is able to compensate for jitter, which might occur.

### 5.1.3 File-transfer scenario

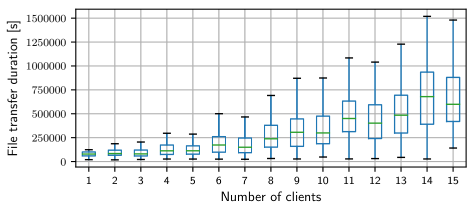
Similar to the previous simulations, a file-transfer application is tested as well. The relevant KPI measured for this application is the file-transfer duration. It represents time required for the client to send a request to the server and completely receive the requested data, which is basically the available bandwidth. Again, the simulation setup is the same as used in the previous sections.



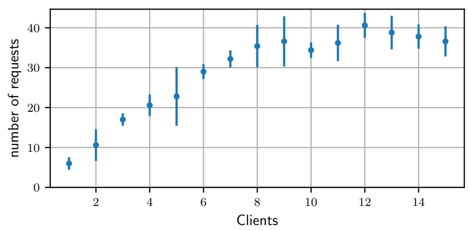

(a) File-transfer duration (100 Mbit/s terrestrial link)



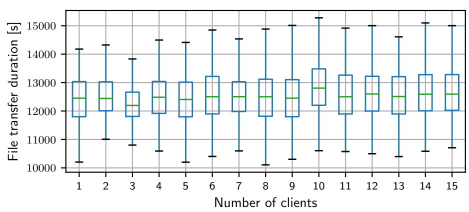

(b) transmitted files (100 Mbit/s terrestrial link)



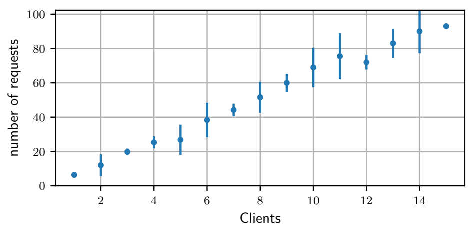

(c) File-transfer duration (1 Mbit/s terrestrial link)




(d) transmitted files(1 Mbit/s terrestrial link)




(e) File-transfer duration (satellite link)




(f) transmitted files (satellite link)

Figure 5.3: File-transfer duration and number of requests for different setups

The measured results for the single connection tests are depicted in Figure 5.3. It should be noted that the x-axis in Figure 5.3a, Figure 5.3c and Figure 5.3e, as well as the y-axis in Figure 5.3b, Figure 5.3d and Figure 5.3f have a different scale.

The measured results for the file-transfer simulation match expectations. As can be seen, the performance of narrow-band terrestrial link is the worst. This is obvious, given the amount of available bandwidth. The satellite performance is considerably better and the user experience is much higher over the satellite, as the time required to download a file is only a fraction compared to the narrow-band terrestrial link. It can also be seen that the difference between a single client requesting a file and 15 clients is marginally better than the satellite, while the file-transfer duration significantly increases with increasing amount of client when the narrow-band terrestrial link is used. However, even though more bandwidth on the satellite down-link is available compared to the broadband terrestrial connection, the broadband connection still outperforms the satellite. The reason for this is the high latency. It should be noted that the TCP acknowledgments, which are sent back from the client to the sever, experience some additional jitter due to the satellite up-link coordination and the DAMA process. Thus, the overall performance suffers.

### 5.1.4 Traffic scenarios

Compared to the precious section, where only a single traffic class has been used during the simulations, more complex scenarios are evaluated in the following phase. This is still a part of the baseline simulation, since no offloading approach is used. Again, the rationale is to measure and visualize how the quality over the individual links will be, if more applications are used in parallel.

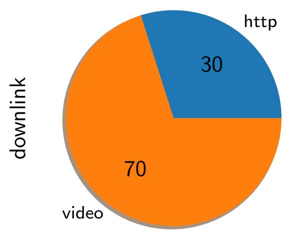
More precisely, two scenarios are defined: A *small scenario* consisting of

- a single HTTP client
- one high quality video client,

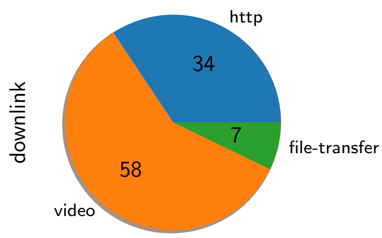
as well as a *large scenario* that encompasses

- a single HTTP client,
- two high quality video clients,
- one low quality video client,
- one file-transfer client.

The traffic pattern has been designed such that the predominant traffic is video traffic, given its predominance in the real world, as explained in Section 2.6. That is, in the small scenario the portion of the video traffic is approximately 70%, while the remaining 30% are used by web-traffic. In the large scenario approximately 58% of the used bandwidth is consumed by video traffic, 34% by web-traffic and the remaining 7% is used for file-transfer, as shown in Figure 5.4.




(a) Small scenario




(b) Large scenario

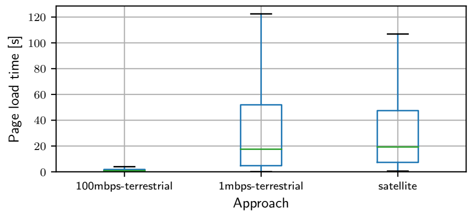
Figure 5.4: Distribution of applications in user traffic patterns

As can be seen, in both scenarios, more than half of the created traffic is video traffic. It should be noted that this Figure shows the portion of the down-link traffic.

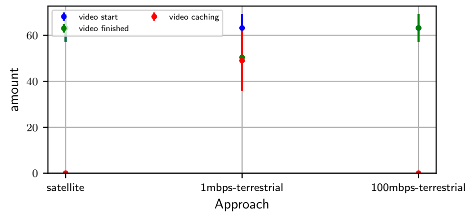
In the following simulations, these two scenarios are used to assess the performance over the various connections.

### Small scenario

Similar to the previous simulations, the same setup has now been used to test the *small scenario*. Again, the simulations have been repeated five times, each time with different random streams. The results are shown in Figure 5.5. As can be seen, the performance of the narrow-band terrestrial link is the lowest for both, video and web-traffic, as constantly more bandwidth than available is required. However, the satellite performance for the web-traffic is only marginally better (see Figure 5.5a). Clear advantages of the satellite over a narrow-band terrestrial connection are only visible for the video traffic in such a scenario.




(a) PLTs

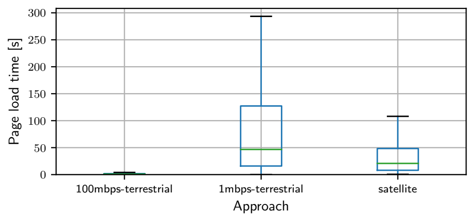



(b) Video stop events

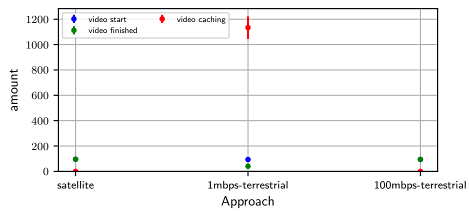
Figure 5.5: PLT and video stop events for Small Scenario

## Large scenario

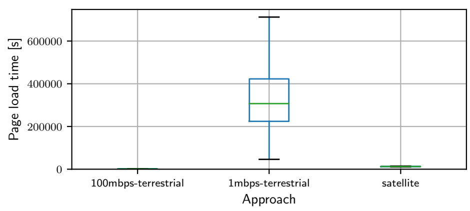
When running the *large scenario*, the negative performance impact of limited bandwidth becomes massively obvious, as shown in Figure 5.6. A high user experience is virtually impossible over a narrow-band terrestrial link. All three applications suffer severe performance impacts. Particularly, Page Load Times (see Figure 5.6a) of the two web clients renders a low QoE and the amount of cache empty events during the simulated video playback (see Figure 5.6b) will virtually prohibit the video playback. Compared to this, the sole satellite connection already brings a significant performance boost and, besides the web-traffic performance, can cope with the broadband terrestrial connection.




(a) PLTs




(b) Video stop events




(c) File-transfer duration

Figure 5.6: PLT video stop events and file-transfer duration for Large Scenario

### 5.1.5 Conclusion of baseline tests

In general, the conducted baseline tests matches expectations. Without any offloading, the broadband terrestrial connection has a clear advantage compared to both the satellite and a narrow-band terrestrial link. The measured KPIs are for several applications and scenarios orders of magnitude better for the broadband connection. Moreover, these results also confirm the initial statement that satellites alone cannot overcome the digital divide and provide rural and remote areas with the same service quality as a broadband terrestrial link. As can be seen, the negative impact of the high latency on interactive applications is severe. Of the tested applications only video traffic achieves similar results. However, the satellite connection provides a better quality than the narrow-band terrestrial link.

## 5.2 Offloading approaches

As previously mentioned, in this section different offloading approaches are evaluated. Firstly, a static approach was examined, which distributed all up-link traffic (clients to server) to the terrestrial connection and the complete down-link traffic (servers to clients) on the satellite link. Such an approach is fairly easy to implement, since no dynamic decision making is involved. Secondly, the approach, which has been developed in the BATS project, is looked at. That is, distributing the traffic based on the time interval between consecutive packets and the amount of data generated in a given time frame (see Section 3.1). Finally, a policy-based approach, so to a certain extent comparable to the SIPTO approach, which offloads traffic based on the kind of traffic. However, besides the policies, this approach has been extended to also take the load on the terrestrial connection into account. It should be noted that information such as traffic class as well as the load information can be easily used in a simulation environment but might be difficult to obtain in a real world deployment. Hence, this approach is called an all-knowing approach, as it assumes that all required information is available to the offloading decision functional block.

In the following, the same simulations as during the baseline tests are conducted - however, this time with one of the three offloading approaches active. That is, first the simulations with just a single traffic class are performed and afterwards the two scenarios with multiple traffic streams of different classes.

### 5.2.1 Down-link offloading

A simple method to utilize both terrestrial and satellite link is to offload the traffic to the satellite based on its direction. That is, as shown in Figure 5.2.1, all traffic directed towards the end user is sent via the satellite link, while for all traffic from the end users to the Internet as such the terrestrial link is used. The rationale behind this approach is that typically end users receive a much higher portion of data than they send and therefore the satellite link, which provides the higher bandwidth, is more beneficial than the terrestrial link, in spite of its lower latency. It should be noted that the increasing usage of web conferencing, cloud services or social media allowing people to easily share video and pictures, might change the ratio in the future. Furthermore, by using the terrestrial up-link, the overall RTT can be reduced compared to the satellite link, which improves the bandwidth-delay product and, thus, the TCP performance.

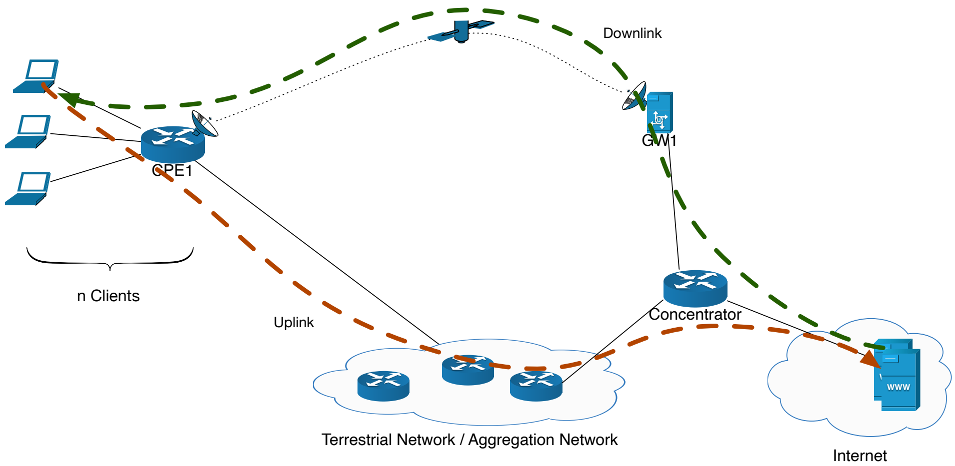


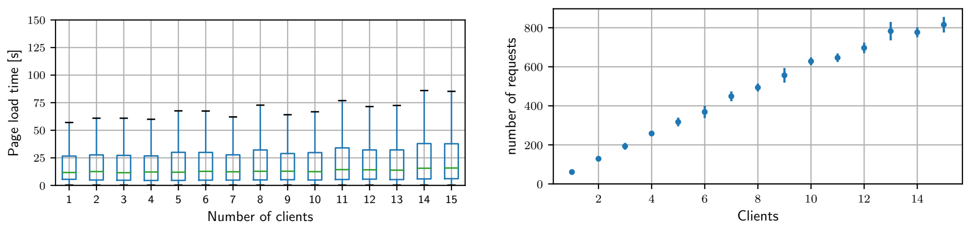

Figure 5.7: Offloading of download traffic

As previously outlined, such an approach can be easily implemented, as the offloading decision is not made dynamically but statically and no coordination is required. Finally, even the satellite connection can be simpler. On the one hand, the end user's satellite equipment does not send any data but just receives them, which allows for cheaper equipment and easier installation, and on the other hand, the complex coordination between the different senders for satellite up-link coordination (see Section 1.1.1) can be neglected.

### Individual traffic class evaluation

It should be noted that the TCP buffers can be adjusted for the offloading approach as well, given that the fixed part of the delay remains constant, as the offloading is done statically.

First, HTTP traffic is used. The corresponding results are shown in Figure 5.2.1.



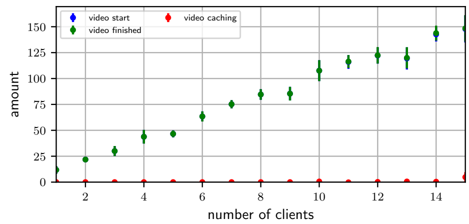

(a) PLT (b) total number of requests

Figure 5.8: PLT - Down-link offloading approach

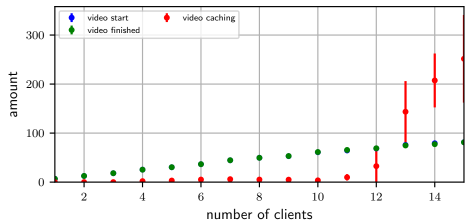
As can be seen in Figure 5.2.1, using the terrestrial link for the up-link traffic, increases the performance compared to a pure satellite link (see Figure 5.1e). That is, the median Page Load Time is slightly lower for all numbers of clients. This is caused by the lower RTT and jitter due to the terrestrial return link, which positively impacts the TCP performance. Compared to the narrow-band terrestrial link the performance is significantly better in terms of Page Load Time, particularly for a higher number of clients. The reason here is the higher available bandwidth

Secondly, the down-link offloading methods for video traffic have been simulated. As the results depicted in Figure 5.9 show, the number of empty buffer events during the video playback are significantly better for both, high and low quality videos, compared to the narrow-band terrestrial connection (see Figure 5.2), yet the performance is not better than a sole satellite link. As also shown in the figure, particularly when the high quality videos are used, more video stop events occur when a higher number of clients is active.

Further analysis reveals that, even though only TCP acknowledgments are sent on the return link, the required bandwidth on the terrestrial (up-




(a) Low quality




(b) High quality video

Figure 5.9: Video stop events - Down-link offloading approach

link) link is not sufficient. This is confirmed by the measured traffic statistics at the concentrator, which are presented in Figure 5.10.

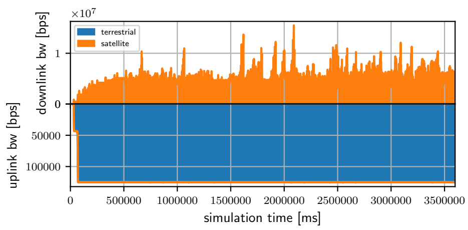
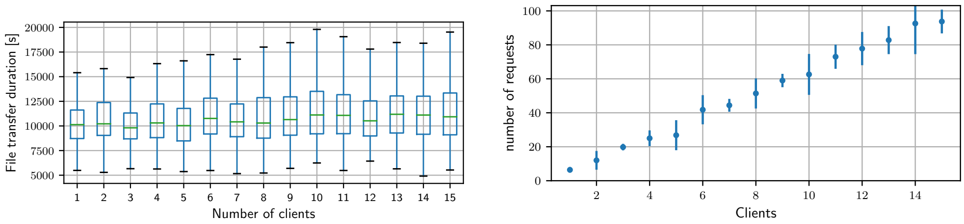



Figure 5.10: Interface traffic statistics for 15 clients

As can be seen, the up-link bandwidth stays constant during the whole simulation at 125 kbit/s, which is represented by the blue graph in the figure. The up-link direction is shown on the lower half, while the down-link direction is presented in the upper part. According to the offloading approach, only the satellite is used for the down-link, while only the terrestrial link is used for the up-link. As previously mentioned, 125 kbit/s is the configured up-link capacity on the terrestrial link. Hence, the whole capacity is used and, thus, due to the TCP behavior, the speed of the down-link is negatively impacted.

Finally, the performance of this offloading approach for file-transfer traffic is simulated. The results of this test are depicted in Figure 5.11. Similar to the video traffic, the performance in terms of file-transfer time is significantly better than the narrow-band terrestrial connection and marginally better than a sole satellite connection. More precisely, the median is between 10 s and 11 s, with this offloading approach, compared to between 12 s and 13 s for the satellite connection and even up to 500 s for the narrow-band terrestrial link.



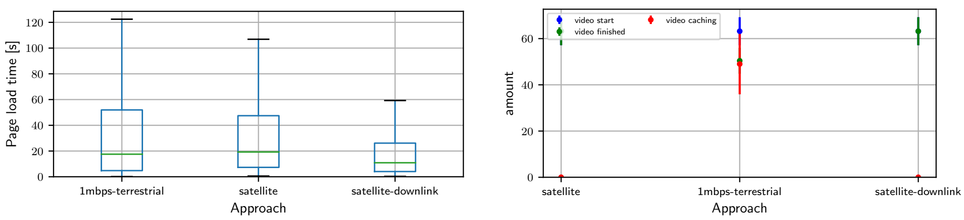

(a) File-transfer duration (b) total number of transmitted files

Figure 5.11: File-transfer results - Down-link offloading approach

These results seem obvious given the higher bandwidth, which is quite relevant for larger file transfers. The slightly better performance of this offloading approach compared to the sole satellite link can be explained by the lower RTT and the less jitter in the uplink.

### Traffic scenarios

In the following the behavior of the down-link offloading approach is evaluated in both of the previously defined mixed traffic scenarios In Figure 5.12 the results of the static down-link offloading method during the small traffic scenario is shown. To allow for a more easy comparison, again the results of the satellite connection and the narrow-band terrestrial connection with the same traffic scenario are included in the plots.

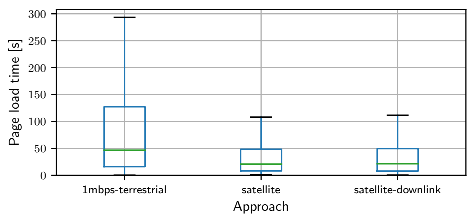



(a) PLTs (b) Video stop events

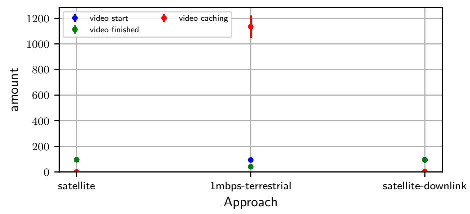
Figure 5.12: PLT, and video stop events for Small Scenario - Down-link offloading approach

As the results show, the static offloading method performs better for the HTTP traffic in the traffic mix by approximately bisecting the Page Load Times, as shown in Figure 5.12a. The video traffic (see Figure 5.12b) of this approach brings a tremendous improvement compared with the narrow-band terrestrial connection, given that no empty cache events occur, which is the same, if only the satellite is used.

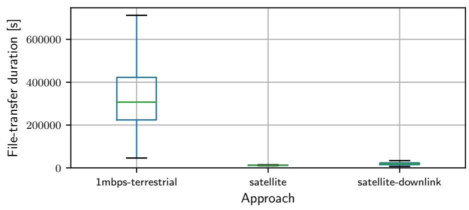
Interestingly, these results change when the amount of traffic increases, as can be seen by the measurements during the *Large scenario* simulation, which are presented in Figure 5.13. Again, in order to allow to compare the results, the measurements of the narrow-band terrestrial as well as the sole satellite connection and integrated within the plots.




(a) PLTs




(b) Video stop events




(c) File-transfer duration

Figure 5.13: PLT video stop events and file-transfer duration for Large Scenario - Down-link offloading approach

Even though the performance of all applications still clearly outperforms the narrow-band terrestrial link, the performance compared to the sole satellite link is similar, i.e. for web and video traffic as shown in Figure 5.13a and 5.13b, the file-transfer duration are slightly longer in average in this particular setup, if the whole down-link traffic is offloaded to the satellite (see Figure 5.13c). Just like in the case of the video traffic scenario, monitoring the up-link terrestrial traffic rate reveals the reason. The available 125 kbit/s are completely consumed by the TCP acknowledgments.

To conclude, this offloading approach, which sends statically all downlink traffic via the satellite connection, is practically in all tested scenarios better, i.e. it can provide a high user experience, than that yielded by a 1 Mbit/s narrow-band terrestrial connection, despite the additionally introduced fixed latency of the satellite. However, compared to a sole satellite connection, a positive performance impact can only be measured if the overall amount of traffic is limited. Otherwise, the resulting amount of up-link traffic in terms of TCP acknowledgments overloads the terrestrial link and in turn has a negative impact. The advantage of this approach is that the satellite can be designed cheaper and more simple, as it is only required for down-link traffic, i.e. the User Terminals only need to receive traffic but are not required to send.

### 5.2.2 Network Object distribution offloading approach

The network object offloading approach has been developed in the EU BATS project. As explained in Section 3.1, the main idea is to make an offloading decision based on so-called network objects. Such a network object is a number of consecutive packets belonging to the same (unidirectional) flow, for which the time between two packets is lower than a certain threshold. This packet interarrival time is measured at the locations in the network where the offloading decision needs to be executed, namely extended Customer premises equipment on the end user side and a concentrator on the network operator edge. The offloading decision itself is then based on the size (in bytes) of the object. While small objects are routed terrestrially, large objects are sent via the satellite, as depicted in Figure 5.14.

If it is larger than another administratively configured threshold, i.e. the object size threshold, it is considered as large and otherwise as small. That is, all objects are first considered as small and, thus, the initial packets of each object are always routed via the terrestrial link. Once a sufficient amount of data packets belonging to the same flow arrive at extended Customer premises equipment or concentrator and are considered to the same object, the traffic is routed via the high bandwidth connection. It should be noted that this happens "on-the-fly" and there is no cache or buffer, neither at the extended Customer premises equipment or concentrator. That is, flows might experience massive change in RTT during their lifetime. The process in summarized in the flow chart shown in Figure 5.15

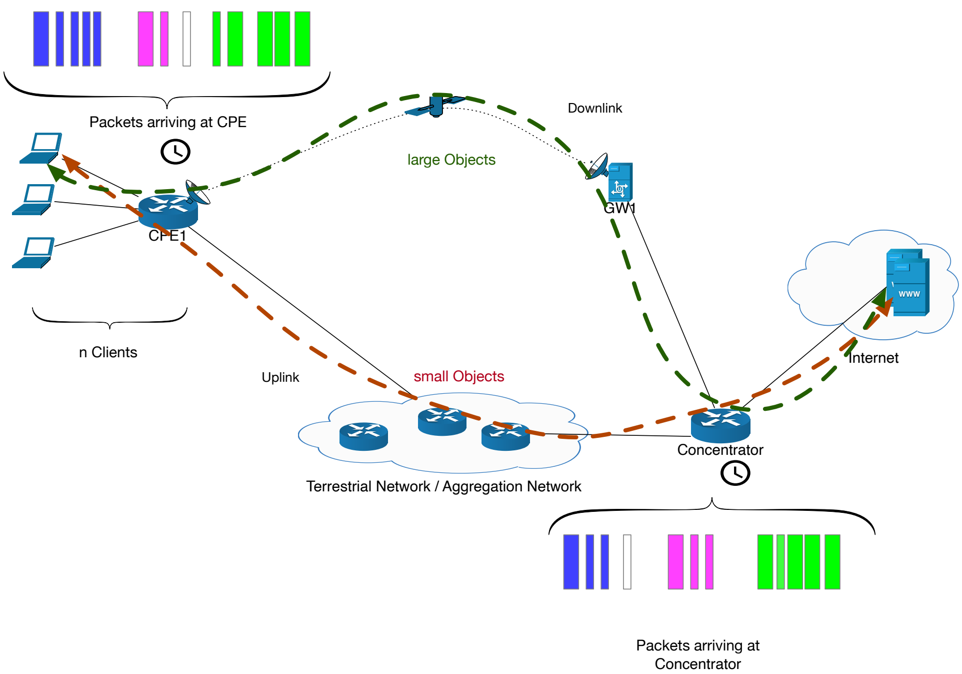



Figure 5.14: Offloading based on network objects

### Individual traffic class evaluation

Again, similarly to the previous offloading approach, the behavior of web-traffic using the BATS offloading approach is evaluated first. It should be noted, that the maximum time period between two consecutive packets of the same flow until which packets are considered as the same object, i.e. the maximum packet interarrival time, is not defined in the corresponding publications and neither is the object size threshold. Hence, reasonable values have to be chosen.

In Figure 5.16 the simulation results for varying object size thresholds are shown. The interarrival time between two packets belonging to the same object is configured to 300 ms. The object size threshold has been configured to 30 kB, 60 kB and 200 kB.

As can be seen in Figure 5.16a, 5.16c and 5.16e, with increasing threshold the average Page Load Times increases. In Figure 5.16b, 5.16d and 5.16f the corresponding traffic distribution onto both connections depending on the number of clients is shown. The red and the green graphs represent the up-link traffic (in percent) for the terrestrial and the satellite link, respectively.

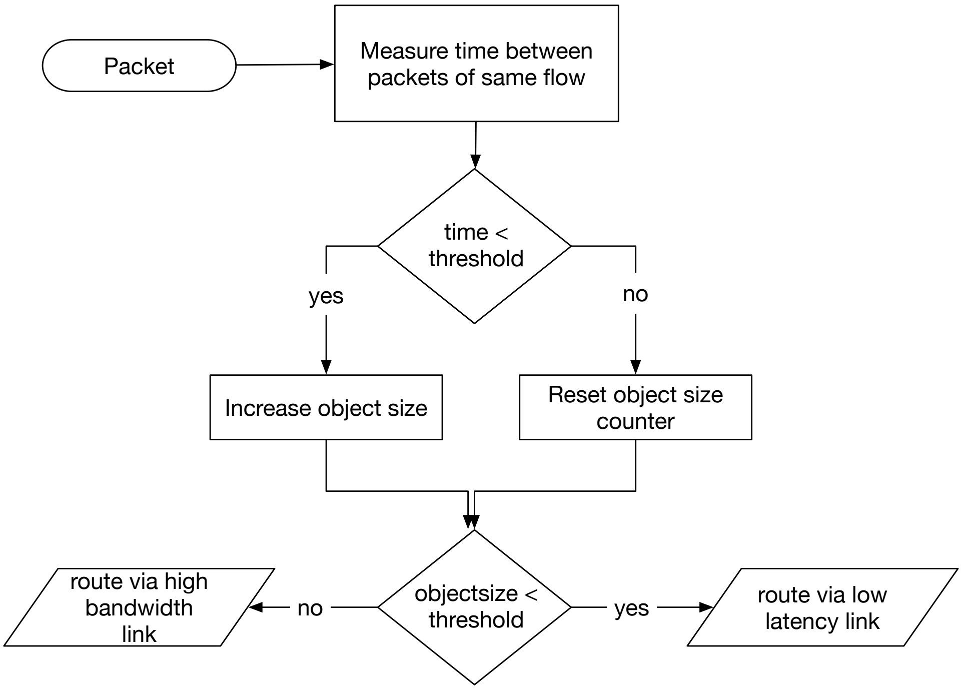
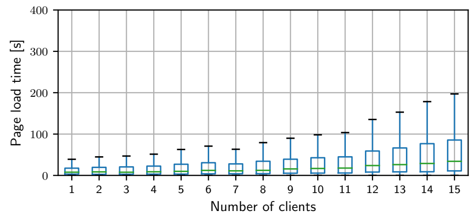


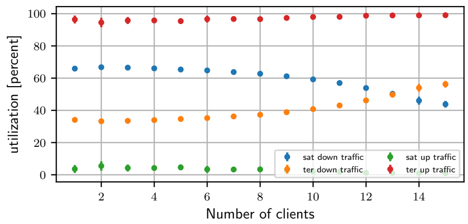

Figure 5.15: Offloading based on network objects flow chart

Accordingly, the orange and blue graphs show the distribution between the terrestrial and satellite down-link. Obviously, with an increasing object size threshold, less objects are classified as large and, thus routed via the terrestrial link. For example, depending on the amount of clients, with an object size threshold of 30 kB between 63% and 40% of the traffic is sent over the satellite, between 57% and 20% for an object size threshold of 60 kB and only between 23% and 3%, if the object size threshold is 200 kB. Moreover, almost 100% of the up-link traffic is sent terrestrially, as the TCP acknowledgments of an individual flow typically do not exceed the object size threshold. With an increasing amount of clients, it can also be observed that the percentage of traffic routed via satellite link decreases. This is caused by the increased jitter with more traffic. Due to this, two consecutive packets are unlikely to be considered as a single object within the proper interarrival time.

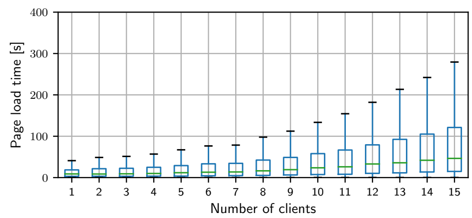
Following the test of different object size thresholds, the impact of different interarrival times is tested. Thus, the test has been repeated but with different interarrival times, namely with interarrival times of 150 ms, 450 ms



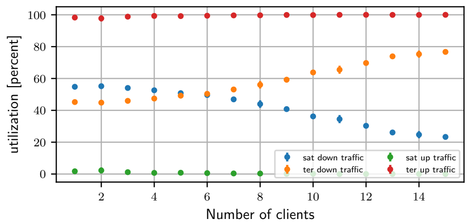

(a) PLT for object size 30 kB



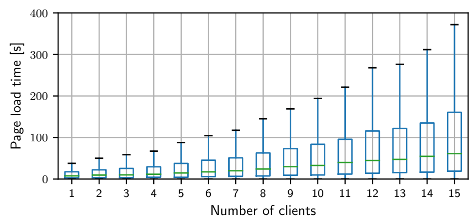

(b) taffic distribution for object size 30 kB



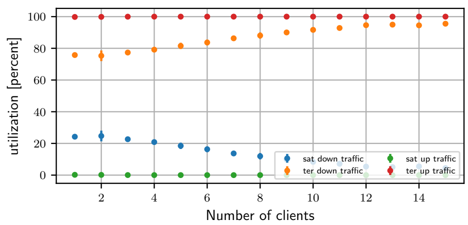

(c) PLT for object size of 60 kB




(d) taffic distribution for object size of 60 kB




(e) PLT for object size of 200 kB

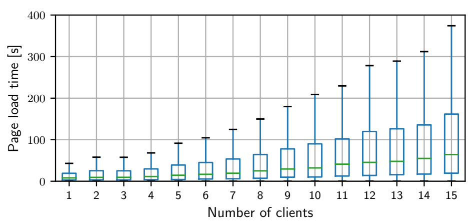



(f) taffic distribution for object size of 200 kB

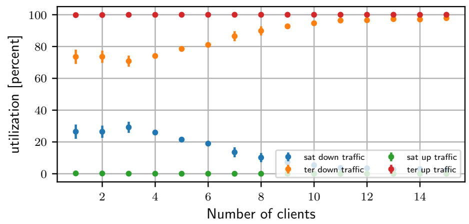
Figure 5.16: PLTs and traffic distribution for different object size threshold with BATS approach

and 800 ms. For this test, the object size threshold has been configured to 60 kB.

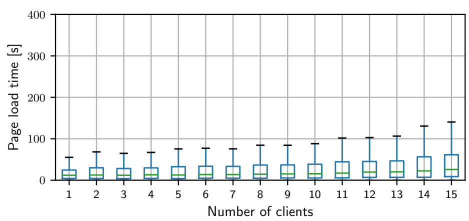
As can be seen in Figure 5.17, with increasing interarrival times, the performance increases, i.e. the Page Load Times decreases. That is, the median Page Load Time decreases from 65 s down to 18 s. Moreover, as shown in Figure 5.17b, 5.17d and 5.17f, the utilization of the satellite link also increases. More precisely, with an interarrival time of 150 ms the majority of the down-link traffic, i.e. between 78% and 100%, is sent terrestrially. Higher interarrival times leads to a usage of between 20% and 30% for an interarrival time of 450 ms and even only for approximately 5% for an interarrival time of 800 ms. The higher usage of the satellite link with its higher capacity reduces the measured Page Load Times. Consequently, the differ-



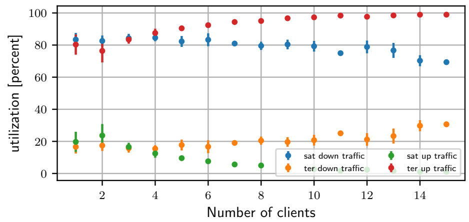

(a) PLT for object size for interarrival time of 150 ms



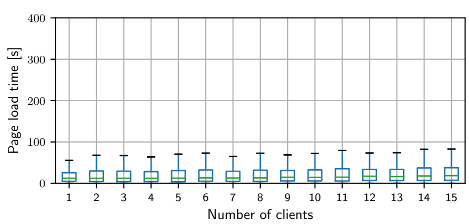

(b) taffic distribution for interarrival time of 150 ms



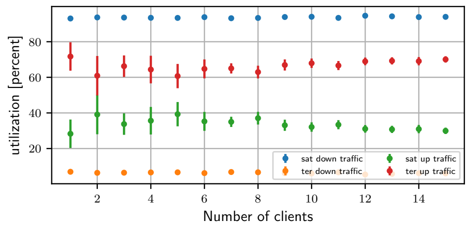

(c) PLT for object size for interarrival time of 450 ms




(d) taffic distribution for interarrival time of 450 ms




(e) PLT for object size for interarrival time of 800 ms



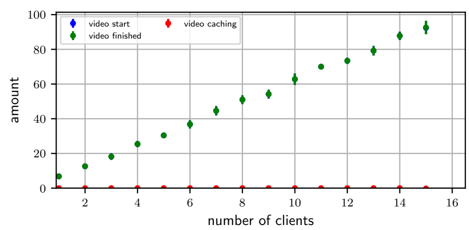

(f) taffic distribution for interarrival time of 800 ms

Figure 5.17: PLTs and traffic distribution for different interarrival times - BATS approach

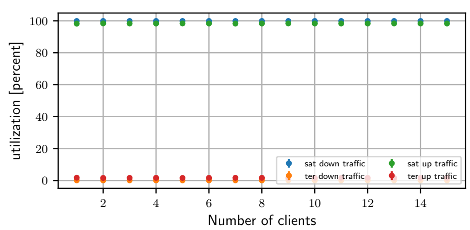
ence in terms of Page Load Time between a single client and 15 clients is marginal for an interarrival time of 800 ms, while a strong increase is visible for an interarrival time of just 150 ms. It should also be noted that with interarrival times of 450 ms and 800 ms the object size threshold of 60 kB is also sometimes exceeded for the uplink traffic, so that a portion of the up-link traffic is sent via the satellite as well.

However, compared to the previously simulated static down-link offloading and/or the sole satellite, the performance of the BATS offloading approach, even with optimized object size threshold and interarrival time values, does not achieve significantly lower Page Load Times in the tested scenarios. That is, the best median Page Load Time of the tested BATS approaches is approximately 18 s for 15 clients, compared to 23 s for a sole satellite connection and 16 s for the down-link offloading approach.

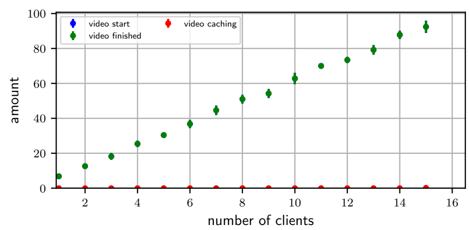
Secondly, the behavior of the BATS approach is simulated when only video traffic is used. The simulation result with an interarrival time of 300 ms and object size threshold of 30 kB, as well as an interarrival time of 1 s and object size threshold of 60 kB is shown in Figure 5.18. These configuration values have been chosen, in order to test faster and slower adoption offloading for the video traffic. All clients have been configured to use the high quality video configuration and a video buffer of 5 s, as already used in the previous configurations.



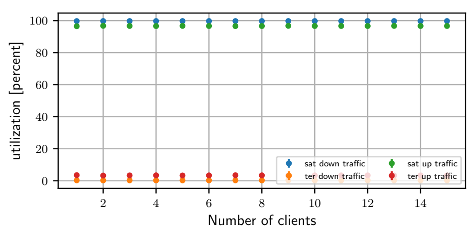

(a) Video stop events for interarrival time = 300 ms and object size threshold = 30 kB




(b) taffic distribution for interarrival time = 300 ms and object size threshold = 30 kB




(c) Video stop events for interarrival time = 1 s and object size threshold = 60 kB



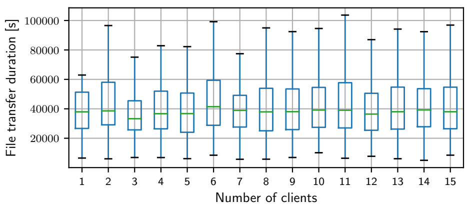

(d) taffic distribution for interarrival time = 1 s and object size threshold = 60 kB

Figure 5.18: Video stop events and traffic distribution for different interarrival times and object size threshold - BATS approach

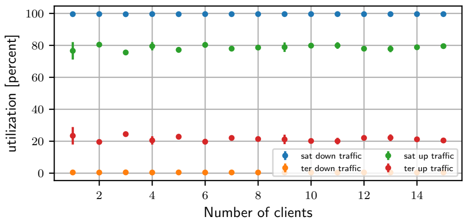
As can be seen in the figure, no video stop events occur with either of the two configurations. Moreover, with both configurations, virtually the complete traffic is sent via the satellite link. This holds true for both up- and down-link Hence, the overall performance of this offloading method in this scenario is basically identical to a sole satellite connection. It obviously outperforms the single narrow-band terrestrial connection, due to the satellite usage. However, it also outperforms the static down-link offloading approach for 11 and more clients, since the BATS approach in this scenario also uses the satellite up-link connection and, thus, avoids the congestion on the terrestrial up-link, which created the video stop events for the other offloading method.

For the sake of completeness, it should be noted that this simulation has also been conducted with low quality video traffic. However, since the results do not reveal additional insights, the corresponding plots are not depicted here.

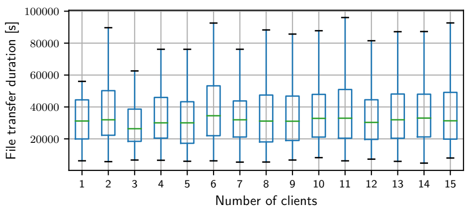
Finally, the behavior of file-transfer traffic is also simulated, when the BATS approach is active. The results of this simulation are shown in Figure 5.19. Again, this time two different configurations of the BATS offloading approach have been tested, which are again an interarrival time of 300 ms and object size threshold of 30 kB, as well as an interarrival time of 1 s and object size threshold of 60 kB.



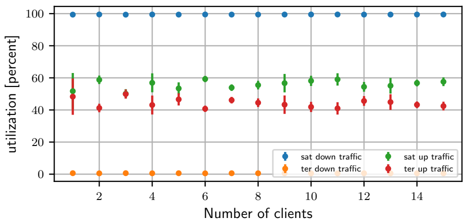

(a) File-transfer duration for interarrival time = 300 ms and object size threshold = 30 kB




(b) Traffic link distribtuion for interarrival time = 300 ms and object size threshold = 30 kB




(c) File-transfer duration for interarrival time = 1 s and object size threshold = 60 kB




(d) Traffic link distribtuion for interarrival time = 1 s and object size threshold = 60 kB

Figure 5.19: File-transfer results - BATS approach

As can be seen in Figure 5.19a and 5.19c, the file-transfer duration is lower for the second configuration. That is, while the median is approximately 660 min for the first configuration, it is 500 min for the larger object size threshold and interarrival time. The corresponding figures showing the distribution of traffic onto both links reveals the main reason for this, namely that the configuration change has virtually only impacted the up-link traffic (see Figure 5.19b and 5.19d). Both configurations of the BATS approach already send all down-link traffic via the satellite, as the file-transfer traffic always exceeds the object size threshold easily. The up-link traffic, which encompasses virtually only TCP acknowledgments, however, utilizes the terrestrial link for a portion of approximately 20% in the first configuration between 40% and 50% in the second configuration, which eventually leads to a higher TCP throughput.

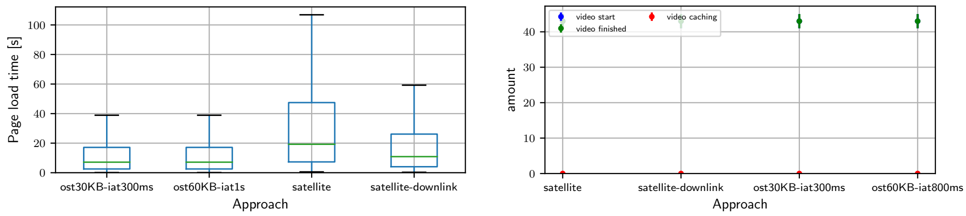
More significantly is the difference to the previously tested static down-link offloading approach, which achieves a median file-transfer duration between 165 min and 184 min in this particular test. A reason for this behavior might be the sudden change in RTT each time a flow is switched from the terrestrial link to the satellite or vice versa, which might occur more than just once for such long-living flows.

#### Traffic scenarios

Similar to the other approaches, also the BATS offloading method has been tested with both mixed traffic scenarios. Again, two configurations of the BATS approach have been tested, i.e. an interarrival time of 300 ms and object size threshold of 30 kB, as well as an interarrival time of 1 s and object size threshold of 60 kB.

The results of the *small scenario* are shown in Figure 5.20. As can be seen, the performance is virtually identical for both configurations, i.e. the same average Page Load Times values and no video stop events. This is consistent with the previous simulations of the BATS approach, where the various configurations only had a relevant impact if the amount of clients increases. For a single client, as used in this test, the differences in configuration are only marginal.

Furthermore, compared to the static down-link offloading approach, in the small scenario the results of the BATS scenario are slightly better. That is, the average Page Load Time is lower, while the video stop events do not equally occur. It is also worth mentioning that the performance compared to a single narrow-band terrestrial link is tremendously better and this approach also outperforms a sole satellite link, i.e.a median Page Load Time

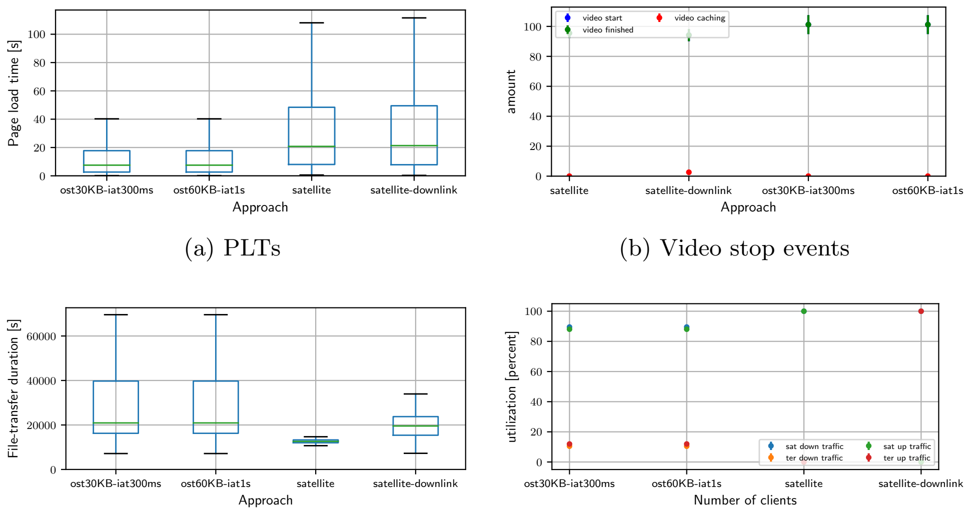



(a) PLTs (b) Video stop events

Figure 5.20: PLT and video stop events for Small Scenario - BATS approach

of 10 s compared to a median Page Load Time of 20 s.

This behavior is also confirmed in the simulation of the *large scenario*. The corresponding results are presented in Figure 5.21. As shown in the Figure, the web-traffic benefits from the BATS offloading approach compared to the satellite down-link and sole-satellite scenario (see Figure 5.21a). However, the file-transfer duration increases (see Figure 5.21c).




(a) PLTs (b) Video stop events

(c) File-transfer duration (d) Traffic link distribtuion

Figure 5.21: PLT, video stop events and file-transfer duration for Large Scenario - BATS approach

In Figure 5.21d the traffic distribution on the available connections is shown. As can be seen, the majority of traffic, approximately 90% is routed via the satellite link. while the remaining 10% use the terrestrial connection. This holds true for both, up- and down-link traffic.

To conclude, the BATS offloading method can provide performance increase and a better user experience in some of the simulated scenarios. More specifically, in the *large scenario* lower Page Load Times can be achieved. Particularly, the interactive web traffic benefits from the BATS approach compared to the the sole satellite link or the static down-link offloading approach. More or less only the file-transfer with its long lasting flows suffers from the dynamic offloading to the satellite and the inevitably significant RTT change. It should be noted that interarrival time and object size threshold need to be configured properly.

### 5.2.3 All-knowing approach

A third offloading approach, a context-aware method is evaluated. That is, this approach takes all the context information into account, which might be relevant, such as traffic class or fill levels of the queues. It is called an all-knowing approach, since it explicitly assumes all relevant information is available everywhere in the network. It should be noted that this can be implemented relatively easy in a simulation environment, but in real-world networks it might be hard to gain the information and to make them timely available in the locations where the offloading decision is done. The purpose of this is to evaluate the benefit that more complicated offloading can bring, so that the higher implementation and processing overhead can be traded off against a potential higher performance.

The implemented offloading algorithm operates on a flow-basis and has two basic steps. During the first step, it is checked if the flow belongs to so-called "latency tolerant" traffic or bulk traffic, such as file-transfer traffic or video. If so, the corresponding traffic flow is immediately offloaded to the satellite, given that such flows benefit from higher bandwidth. This first step is comparable to the SIPTO mechanisms as described in Section 2.3.3 or PMIPv6 with flow mobility as described in Section 2.2.4. Both approaches are designed to work in mobile network and actually focus on the execution of an offloading decision in the network, but not the offloading decision itself. However, a typical method to implement either SIPTO or PMIPv6 would be to define policies that describe which traffic should be offloaded, e.g. YouTube video traffic. The major difference is that the SIPTO and PMIPv6 with flow mobility policies are typically based on IP addresses or DNS names rather than traffic classes. Again, it should be emphasized that classifying traffic as "latency tolerant", so that the usage of such a policy is enabled, is not a trivial task. In Section 1.2.2 different options are briefly described. In general, traffic can be either tagged by the application or the kind of traffic needs to be detected by other means. However, given the complexity involved, this is out of scope of this thesis.

Moreover, if the flow belongs to a different traffic class, the second step is performed, during which the queue of the terrestrial link is checked. If the queue is more than 30% full, the offloading of flow starts based on the priority of the flow with the flows of lowest priority. If the queue is more than 50% full, the next higher priority flows are offloaded and finally, once the queue is more than 70% full, all flows are started to be offloaded. If there are multiple flows with the same priority, an arbitrary flow is picked to be offloaded. Again, defining the priority of a flow can be defined as a policy. Here, the control traffic, such TCP acknowledgments are treated with the highest priority. It should be noted that TCP acknowledgments are treated like latency intolerant traffic. This is particularly important on the up-link.

As can be seen, such an algorithm requires extensive knowledge about the corresponding application, in order to determine the flow's priority or whether or not it is a background flow. As previously mentioned, while in a simulation environment this information can be determined easily, in the real world deployment it can be challenging.

### Individual traffic class evaluation

This offloading approach is also tested in a similar fashion to the previous approaches. That is, the same simulations are conducted. Consequently, first the behavior with web-traffic is tested with an increasing amount of clients. The results of this test are show in Figure 5.23.

As can be seen by the Page Load Times (see Figure 5.23a) the performance, in terms of Page Load Times, is similar to both the static down-link offloading as well as the BATS approach, with the best object size threshold and interarrival time values. Moreover, Figure 5.23b reveals the distribution of traffic between terrestrial and satellite link. As can be seen, the majority of down-link traffic is routed over the satellite, i.e. approx 98%. On the up-link, however, only between 20% and 40% is is sent via the satellite. Thanks to the higher variance in HTTP traffic, also the distribution shows a higher variance.

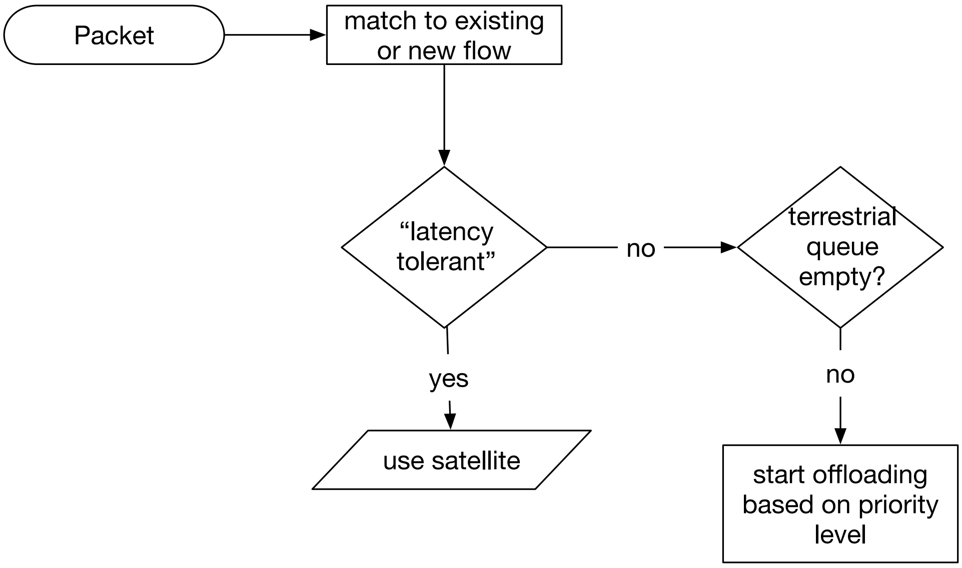
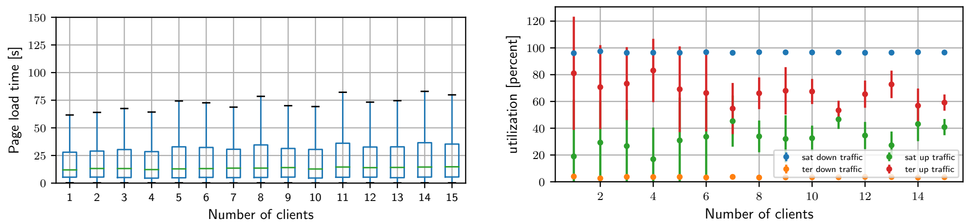



Figure 5.22: All-knowing offload approach



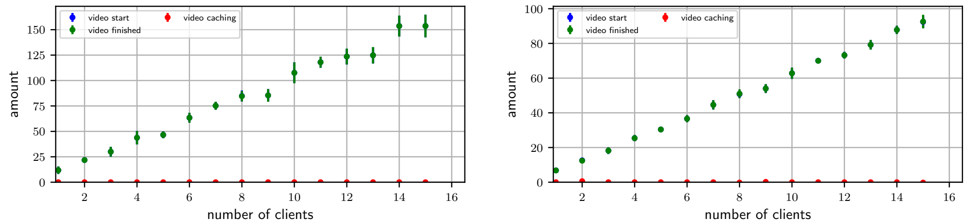

(a) PLT (b) Interface traffic distribution

Figure 5.23: Web-traffic - All-knowing offloading approach

The results of the simulation experiment with the video traffic are shown in Figure 5.24a. The measured KPI, i.e. the amount of video stop events, is zero for both high and low quality videos.

Given that down-link video traffic is instantly offloaded to the satellite in this scenario, only the distribution of the up-link is of interest. As shown in Figure 5.25, with a small number of clients, only the terrestrial link is used for up-link traffic. With an increasing amount of clients, the portion of satellite up-link traffic increases to 83%.

Due to the offloading of a portion of the up-link traffic, this approach avoids the queue overrun that occurs in the static down-link offloading approach and can also serve a high number of clients without any quality




(a) Low quality (b) High quality video

Figure 5.24: Video stop events - all-knowing approach

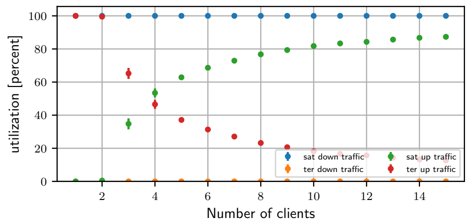
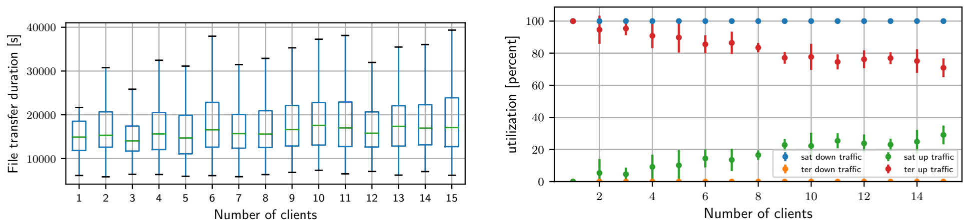



Figure 5.25: Interface traffic distribution - all-knowing approach

losses.

Finally, the results of the file-transfer scenario are shown in Figure 5.26. Compared to the BATS offloading approach, the results are better, given



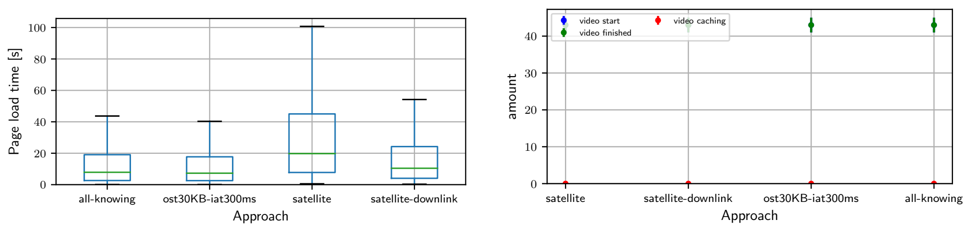

(a) File-transfer duration (b) Interface traffic distribution

Figure 5.26: File-transfer results - all-knowing approach

that this approach immediately selects the satellite for this kind of traffic, while the BATS approach first sends data over the terrestrial connection and switches to satellite once object size threshold is reached. It can be noticed that the file-transfer duration is longer on average compared to the satellite down-link approach. For this minor decrease, the fact remains that with such an offloading approach, the TCP queues cannot be adopted, as the RTT might change. In a real world deployment, this effect will be compensated by a PEP. Again, the distribution of traffic between satellite and terrestrial link (see Figure 5.26b), shows that the satellite up-link usage is constantly increasing with the amount of clients. For the sake of completeness, it should be noted that, similar to video traffic, file-transfer (down-link) traffic is immediately offloaded to the satellite.

### Traffic scenarios

In the following the behavior of this offloading approach in the traffic mix scenarios is tested. First, the simulations with the small scenario are conducted. The results are presented in Figure 5.27. In order to allow for easy comparison, the results of the BATS approach, the static down-link offloading approach, as well as sole satellite connection without any offloading are included in the figures. As can be seen, this approach performs in the *Small*

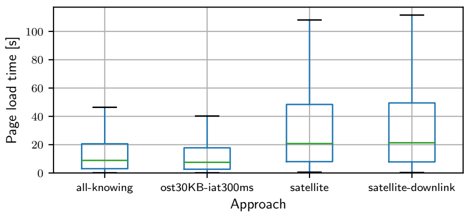



(a) PLTs (b) Video stop events

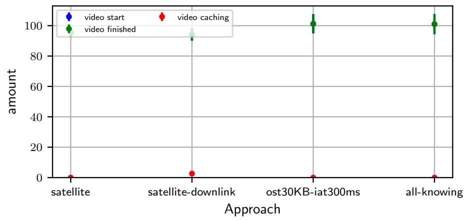
Figure 5.27: PLT and video stop events for Small Scenario - all-knowing approach

*Scenario* similar to the aforementioned BATS approach. With respect to video traffic, no stop events occur. This is obvious, given the policy to offload all video down-link traffic to the satellite. It should be noted that the other approaches perform similarly, as shown in the previous simulations. Differences, however, become visible with respect to the web traffic. While the BATS approach and the all-knowing approach perform similarity in terms of Page Load Times, the static down-link offloading approach accounts for slightly higher values. The sole satellite connection more or less doubles the Page Load Time values. It should be noted that the results for a sole narrow-band terrestrial connection is significantly higher, i.e. a median Page Load Time of 50 s, compared to approximately 10 s for the BATS and all-knowing approach.

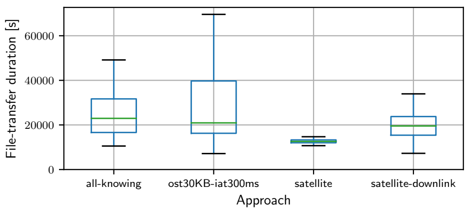
Finally, the all-knowing approach has been tested in the *Large Scenario* simulation. The corresponding results are presented in Figure 5.28.



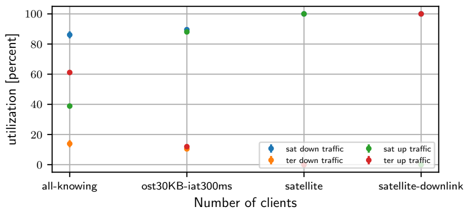

(a) PLTs




(b) Video stop events




(c) File-transfer duration




(d) Traffic link distribtuion

Figure 5.28: PLT, video stop events and file-transfer duration for Large Scenario - all-knowing approach approach

With this scenario too, results are confirm the insight already gained from the *Small scenario*. Offloading based on policies, traffic information and queue fill level achieves a similar performance than the BATS offloading process with respect to web-traffic and video traffic in the large scenario. The performance of the file-transfer with the measured KPI of file-transfer duration shows slightly better results for this approach compared to the BATS approach. More precisely, this approach yields a smaller variance, leading to less long file-transfer duration. It should be noted that both the sole satellite link as well as the static down-link offloading approach lead to much higher Page Load Times, but a lower file-transfer duration. The latter can be explained by the fact that the TCP buffers could be better optimized, as the traffic paths are clear upfront With respect to the traffic to link distribution, it can be seen that the all-knowing approach utilizes the terrestrial link more than the BATS approach, as shown in Figure 5.28d.

It can be concluded that such an approach can help optimize more interactive traffic. It effectively utilizes both the terrestrial link as well as the satellite connection, without overutilizing the terrestrial connection. Moreover, compared to the BATS approach, adaptation to future killer applications can be done easily by adjusting the policies. However, obtaining the information required to implement the algorithm in a real-world deployment, might become difficult, as explained earlier.

## 5.3 Conclusion on evaluation

In the aforementioned evaluation, three offloading approaches are compared against each other and against single network connections with different configurations. In order to assess their different behavior, these connections and offloading approaches have been tested with various simulated traffic patterns.

The main conclusion, which can be drawn from the findings of these simulations, is that offloading traffic from a narrow-band terrestrial connection, as it is common in many rural and remote areas, onto a satellite connection, which relieves the used bandwidth on the terrestrial link, can significantly improve the performance and the user's experience. However, even though the satellite provides considerable bandwidth, achieving the same quality than a broadband terrestrial link is virtually impossible, due to the impact of the high latency on transport protocols and interactive applications, which is shown in significant relative differences in the measured KPIs.

Moreover, offloading approaches, compared to sole satellite connection, also bring performance improvements, particularly for more interactive traffic, such as web-traffic, as they exploit both connections. Out of the three compared approaches, it can be said that the static down-link offloading approach is easy to implement, yet it has clear limitations, particularly if the amount of traffic increases. In such a case, the terrestrial-up-link might become overloaded, so that the overall performance is negatively impacted. As there is no traffic differentiation, particularly interactive web-traffic suffers if e.g. large a file-transfer is started. Furthermore, the BATS approach and the designed all-knowing approach, which basically takes the idea of policy-based offloading from existing work of mobile networks, namely (Bernardos (Ed.) 2016) and (Maallawi et al. 2015), and adopt them for converged satellite and terrestrial networks, have evaluated. Both achieve similar results in many test. Particularly, web- and video traffic results indicate that both achieve a similar quality. Hence, the results confirms and extends the findings of (Ekmekcioglu et al. 2018) in the sense that the BATS approach improves the overall QoE. By using other KPIs than solely the MOS as well as a different network and traffic setup than (Ekmekcioglu et al. 2018), the benefit of the BATS offloading approach has been validated from another viewpoint. However, it has also be shown the the all-knowing approach performs significantly better for file-transfers than the BATS approach. That is, the file-transfer duration is lower with the all-knowing approach for both, the individual file-transfer test as well as the file-transfer results in the mixed scenarios. Hence, the conducted simulations show that dynamic policy-based offloading approaches can also have a benefit in converged satellite and terrestrial networks. Moreover, the BATS approach operates on traffic rates, independent of the kind of application, while the all-knowing approach implements defined policies, which eventually allows for more control, such as prioritization of users, offloading certain IP-addresses or novel upcoming applications. This outcome is considered in the next chapter during the design of an overall network architecture that allows executing such an offloading approach in converged satellite and terrestrial networks.

# Chapter 6

# Converged Network Architecture

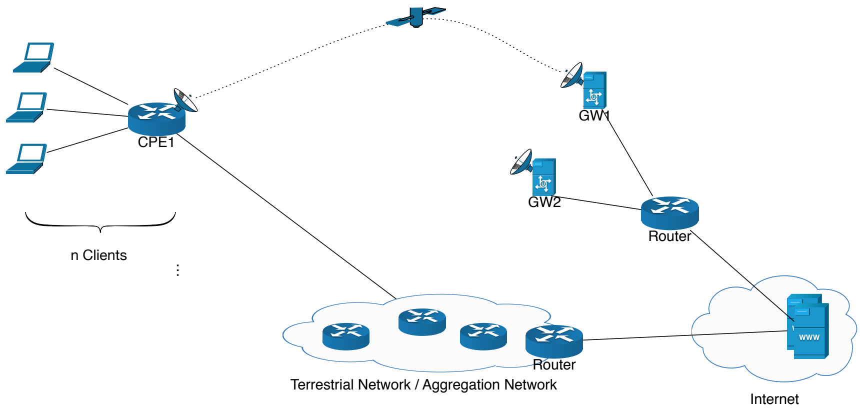
The following chapter deals with the design of a network architecture that allows for executing traffic offloading in converged satellite and terrestrial networks. It virtually defines how the outcome of the previous chapter can be enabled in a network, i.e. what kind of network architecture is required, and, thus, addresses the last objective as defined in Section 1.4.

Following the functional building blocks as defined in Section 2.1, this encompasses the execution block that ensures the offloading decision is executed in the network. Obviously, the design of this execution block highly depend on the actual path selection block, i.e. the traffic offloading method that is used. For example, the static down-link offloading approach only requires the information on the traffic direction to decide if traffic should be offloaded or not and the execution. The execution block can be designed similarly simple, given that only regular IP routes need to be modified. However, more effective offloading methods, such as the BATS approach or the designed policy-based all-knowing approach, require more sophisticated execution functions, due to their offloading granularity. Hence, in order to design a overall network architecture, the outcome of previous chapter is considered.

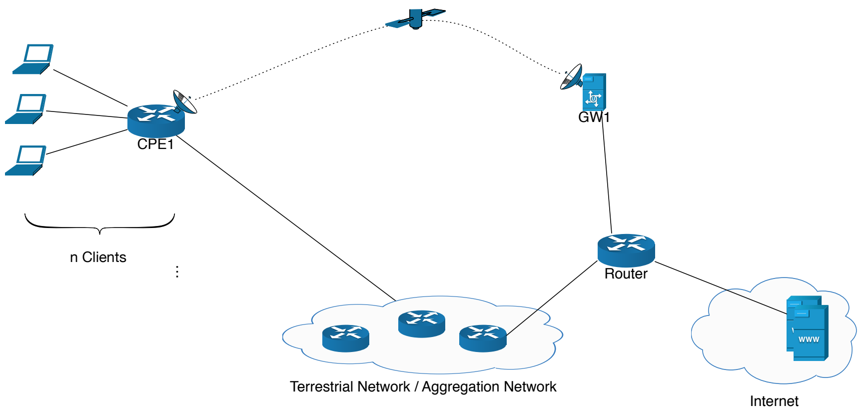
Given that, in this chapter firstly possible network architectures are discussed and evaluated. Secondly, an SDN-based execution function is designed and tested.

## 6.1 Satellite network integration

Satellite and terrestrial networks can be integrated in various ways. The integration itself can either be very loose, so that both networks exist and operate virtually independently of each other; alternatively it can be more tight, i.e. a common anchor node on the operator's edge on the network. Such an anchor node, which is referred to concentrator in the previous chapters, interconnects both the terrestrial and the satellite network.




(a) Converged network architecture without common anchor node




(b) Converged network architecture with common anchor node

Figure 6.1: Common anchor point

The difference is depicted in Figure 6.1. Either both networks are operating independently of each other locating the decision point if traffic should be offloaded to the CPEs, or both networks have a common point on the path.

Even though in the previous chapter a tight integration with a common anchor node between both networks is assumed, in the following implications of both designs are discussed.

### 6.1.1 Loose network integration

A loose network integration is characterized by a virtually independent operation of both the terrestrial and the satellite network. That is, as shown in Figure 6.1a, the connection point of both networks is on the end user's premises, i.e. the extended Customer premises equipment, but there is no connection on the operator's edge. Eventually a user cloud has independent contracts with two different operators. This has obviously implications on the execution of the offloading decision but also on the decision process itself.

First of all, the offloading decision can solely be executed on the extended Customer premises equipment. While at the extended Customer premises equipment the up-link traffic can be directly controlled, enforcing a offloading decision for the down-link traffic can only happen indirectly, as there is no concentrator node. The well-established for this purpose is the usage of NAT, which is also used for the SIPTO method in mobile networks. In Figure 6.2 the exact implementation is depicted. When a client initiates a new connection to e.g. a web-server or requests a video, and the packets are passing the extended Customer premises equipment, it continuously checks if the flow should be offloaded. If not, the packets are forwarded towards the default terrestrial path. In contrast to that, if the flow should be offloaded, the extended Customer premises equipment can perform NAT and change the source IP address of the packets and send them via the satellite provider. The content server sends back the return packets to the satellite IP of the extended Customer premises equipment. As can be seen, the offloading of the down-link traffic is also controlled by the extended Customer premises equipment without the need of a concentrator on the other edge of the network. Moreover, no direct connection between the satellite and the terrestrial operator is required.

However, the loose network integration poses some limitation on the offloading decision function. That is, due to the changed IP address the client appears to the content server as a different client, once the traffic offloaded. As a result, offloading of traffic within the same flow will lead to failure, since the client initiates the connection with its default IP address

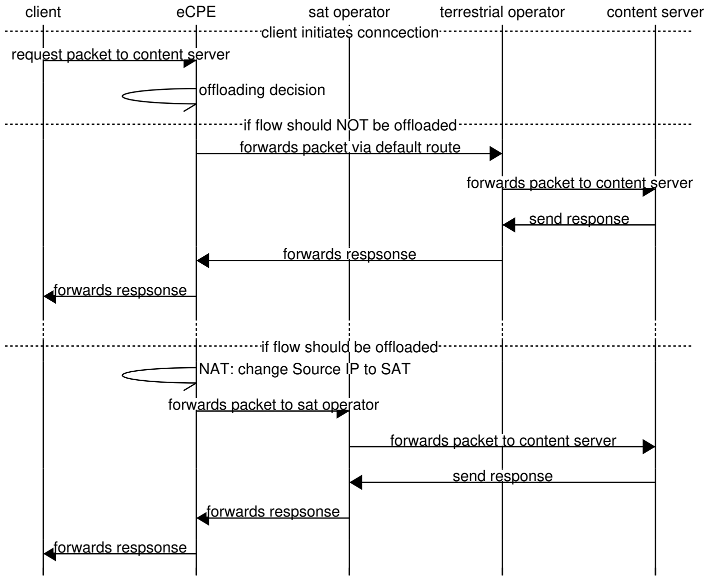



Figure 6.2: Usage of NAT to execute offloading decision

and switches to its satellite IP address, as shown in Figure 6.3. To the content server this appears as a new or incomplete TCP session. Hence, the BATS offloading approach will not work and the policy-based all-knowing approach needs to be adopted to decide if a flow should be offloaded only at the connection initialization, which will decrease its performance.

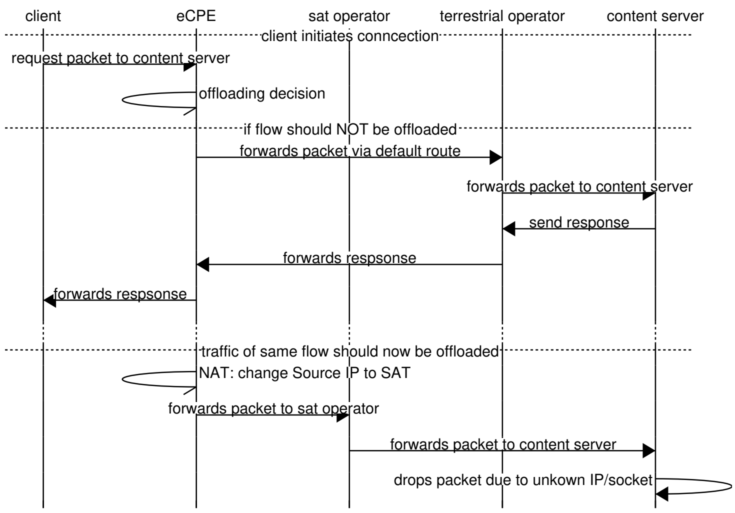



Figure 6.3: Usage of NAT to execute offloading decision within a single flow

### 6.1.2 Tight network integration

Compared to the loose network integration approach, the tight network integration is characterized by a common anchor node on the operator's edge of the network, as presented in Figure 6.1b. This so-called concentrator node requires that either both the satellite and the terrestrial operator share a common infrastructure or both networks are operated by the same entity. For example, a network operator that provides terrestrial fixed-line Internet services but owns also a satellite connection for more rural and remote areas.

The advantage of such a tighter integration is up-link and down-link traffic can be treated separately, as there are two logical points in the network at which the offloading decision can be enforced. Moreover, as there is no NAT required, also traffic within the same flow can be offloaded, which increases the flexibility of the offloaded function. In contrast to the aforementioned loose integration approach, for the content server the client always appears as a single client, independent of the use connection. The traffic can be offloaded just be selecting the proper link.

Moreover, given the concentrator node, even connections that are not initiated by the client but by a server outside of the network, can be offloaded, if required.

### 6.1.3 Discussion

Obviously both integration methods have their advantages and disadvantages. While the loose integration does not require a concentrator node and, thus, an interconnection of the satellite and terrestrial connection, it imposes strong limitations on the offloading decision functional block, since it allows only for offloading of complete flows. In contrast to that, the tight integration approach requires a common operator for both networks or at least a close integration between a satellite and a terrestrial operator.

However, the higher flexibility with respect to the offloading function seems preferable, since it allows for a more effective offloading, which ultimately makes a converged satellite and terrestrial network interesting for operators to properly integrate with each other.

## 6.2 Offloading Execution

As previously mentioned, the offloading decision needs to be executed somehow in the network. That is, in order to enable offloading of traffic from the terrestrial connection, the packet flow between end user clients and the operator's network needs to be changed from the default terrestrial path towards the satellite path for certain traffic flows. That is, the packet flow within the concentrator and/or extended Customer premises equipment needs to be dynamically modified, so that the alternative route is used by the packets of specific flows or parts of a specific flow. Furthermore, as shown in Chapter 5, effective offloading requires that the execution function operates quickly and flexibly. Given the outcome of the review of the related work and the state-of-the-art in Chapters 2 and 3, relying on emerging SDN techniques seems a promising approach for the offloading execution function, as explained by (Watts & Aliu 2014, Kapovits et al. 2014, Ferrús et al. 2016, Niephaus et al. 2019). Other potential options would be to utilize the PCE architecture (see Section 2.2.2) or PMIPv6 (see Section 2.2.4). However, while the former's main focus is to manage capacity based on traffic aggregate, the latter is limited to the IP layer. Hence, an SDN approach provides a higher flexibility, by its clear separation of control and data plane Moreover, it is shown in (Niephaus, Ghinea, Aliu, Hadzic & Kretschmer 2015, Niephaus et al. 2014, Niephaus, Hadzic, Aliu, Ghinea & Kretschmer 2015) that an SDN approach can strongly support wireless networks in general, including the satellite.

Furthermore, following the general SDN architecture inevitably leads to a centralized approach. That is, even though for resilience purposes multiple SDN controllers might exist, logically each network device is controlled by a single controller. On the one hand this increases the control traffic overhead in the network, but on the other hand offloading policies need only to be implemented in a single place, so that they can be quickly changed or adopted to new application.

In the following, the designed SDN-based execution function is described.

### 6.2.1 Integration of SDN into the overall network architecture

In order to integrate SDN (Xia et al. 2014) concepts into the network architecture, certain devices need to become SDN-enabled. Since both extended Customer premises equipment and concentrator are the interconnection nodes between the satellite and the terrestrial network, it seems logical that these devices become SDN-enabled, i.e. logically they become OpenFlow network elements. OpenFlow has been chosen, since it is an open standard an commonly used in SDN environments.

Moreover, relying on SDN creates inevitably a dedicated control plane, as depicted in Figure 6.4, where the required SDN controller is located. The controller uses the OpenFlow protocol to configure the OpenFlow devices, i.e. all extended Customer premises equipments and concentrator devices. In contrast to that, network elements, i.e. extended Customer premises equipment devices as well as a Concentrator device, are located in the Data Plane. However, by just integrating SDN devices as well as a SDN controller, traffic offloading is not enabled per se, since on the one hand integration with the offloading function needs to be made, and on the other hand, the decision needs to be transformed into specific SDN-flow rules, which can be implemented into the flow tables of the SDN devices.

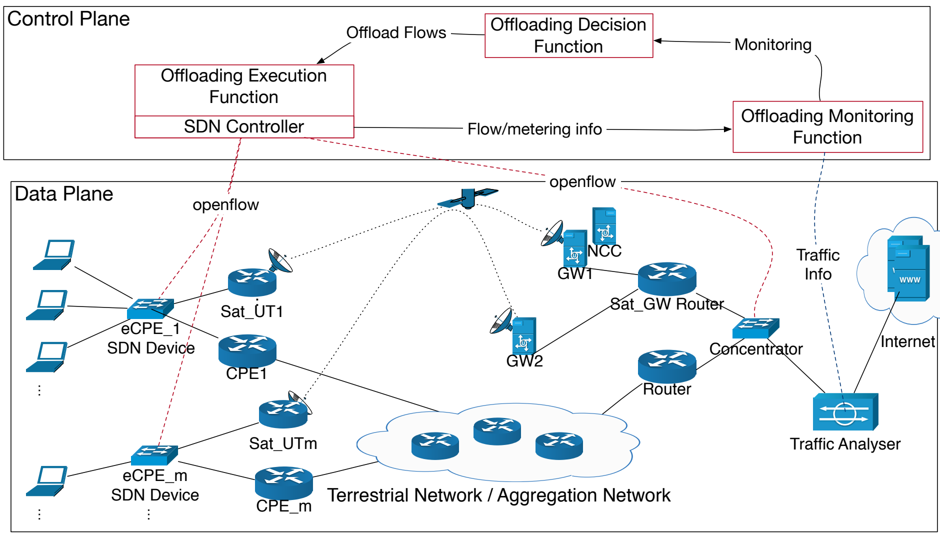



Figure 6.4: Converged network architecture

Hence, besides the SDN controller, additional components are located in the control plane, which are required to allow traffic offloading. These components can be considered as network applications in the SDN terminology. Firstly is the Offloading Decision Function , which implements the path selection and offloading function, as described previously. The Offloading Decision Function communicates its decision to an Offloading Execution Function that is responsible for executing it. That is, the Offloading Execution Function receives a 5-tuple flow description of one or multiple flows, potentially including some wild-cards, that should be offloaded to the satellite link. Secondly, the Offloading Execution Function generates proper OpenFlow rules, which are sent to the SDN controller, so that they can be forwarded to the devices using the regular OpenFlow protocol. More precisely, the created rules instructs extended Customer premises equipment and concentrator to modify the layer 2 destination address as well as the outgoing port of the packets belonging to the flow, which should be offloaded, so that packets are sent towards the satellite link. Finally, the SDN controller feeds back monitoring information, such as queue fill levels, to a Offloading Monitoring Function. The information is not provided directly to the Offloading Decision Function, in order to allow for combining it with knowledge from other sources, such as ACM, which might change the Also, an additional traffic analyzer can perform DPI or other techniques to investigate the traffic in order to identify application types. As previously mentioned in Section 1.2.2, for the scope of this work, such a traffic analyzer is seen as a black box that provides the required information for the offloading decision algorithm to work.

The drawback of such an approach is that collecting statistics on the flows from the extended Customer premises equipments generates a lot of additional traffic that needs to be transmitted in-band over the WAN connection and, thus, reduces the bandwidth that is available for transmitting user data. Moreover, as control packets are typically rather time-critical and small it makes most sense to use the terrestrial connection for the communication between the SDN controller and the controlled devices, i.e. extended Customer premises equipments and concentrator. In order to reduce this, the Offloading Monitoring Function dynamically request this information only if needed, since e.g. in many cases the statistics available at the Concentrator are sufficient. Similar to the Offloading Execution Function, the Offloading Monitoring Function creates the proper OpenFlow rules to request the flow statistics from the Concentrator and, if required, also from extended Customer premises equipments. These OpenFlow messages are sent via the SDN controller to the devices. The information is then provided to the Offloading Decision Function by the Offloading Monitoring Function. Thus, ultimately a control loop is formed that consists of a decision, execution and monitoring component which are implemented by Offloading Decision Function, Offloading Execution Function and Offloading Monitoring Function, respectively.

### 6.2.2 Control Plane Message Exchange

The exact message exchange is depicted in Figure 6.5. As can be seen, the concentrator and extended Customer premises equipment notify the SDN controller about a packet that belongs to a new flow, i.e. a new flow 5-tuple. Due to the centralized approach, the exact identification based on the typical 5-tuple is required, since both devices detect the same flow and will notify this to the controller. The SDN controller then immediately instructs the notifying devices to forward the flow terrestrially, in order to avoid a long delay in the startup phase of a flow. The Offloading Monitoring Function analyzes the messages, joins them and forwards them to the Offloading Decision Function. Moreover, if required by the Offloading Monitoring Function, the SDN controller also requests regularly from concentrator as well as extended Customer premises equipment queue and flow statistics, which are again forwarded to the Offloading Decision Function. In addition to that information on the traffic itself that are gained by the traffic analyzer is also forwarded via the Offloading Monitoring Function to the Offloading Decision Function.

The Offloading Decision Function in turn takes all this information into account and performs the offloading decision. If so, it sends the corresponding 5-tuple to the Offloading Execution Function. Upon reception, the Offloading Execution Function generates proper OpenFlow rules that modifies the flow tables of the involved devices as mentioned previously to send the traffic to towards the satellite link. These rules are sent by the Offloading Execution Function to the SDN controller that pushes these rules onto the devices.

### 6.2.3 Amount of control traffic overhead

Due to the centralized approach, the amount of control traffic that is being exchanged between the extended Customer premises equipments and the

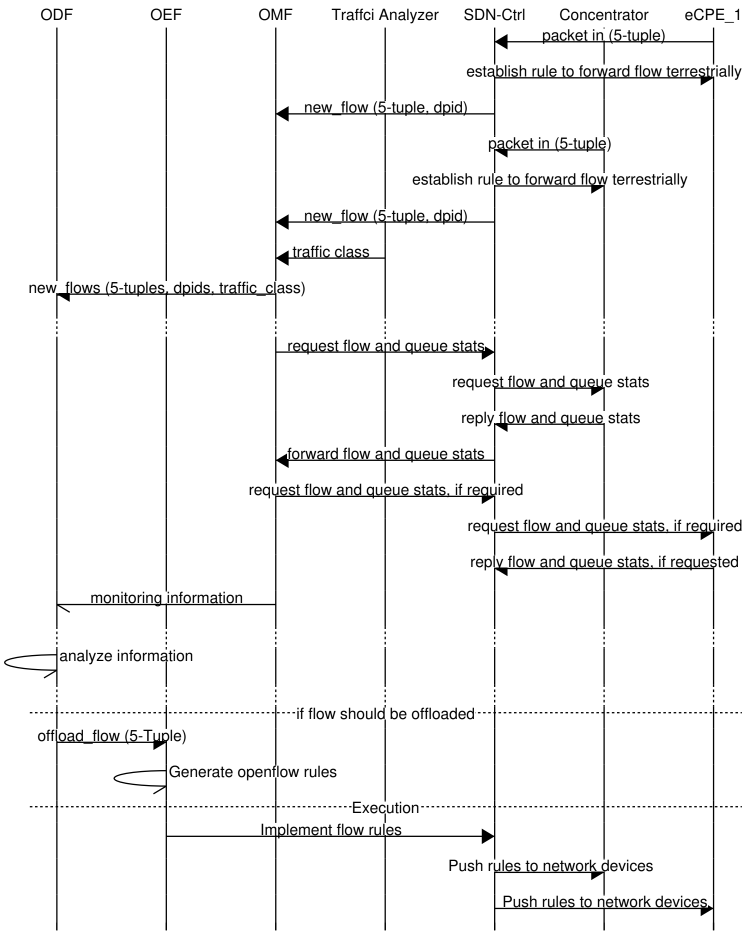



Figure 6.5: Offloading sequence chart

SDN controller is crucial. As previously mentioned, given its time criticality it needs to be transmitted terrestrially. Hence, the amount of control traffic that is being exchanged needs to be evaluated. In order to do that, the simulation environment, which has been described in Section 4.5 has used in Chapter 5, is extended with OpenFlow functionality (Chaves et al. 2016), so that extended Customer premises equipment and Concentrator effectively become OpenFlow switches. The delay introduced to simulate the average flow table search time is estimated as $k * \log_2(n)$, where $k$ is the constant attribute set to the time for a single hardware operation and $n$ represents the current number of flow entries in the pipeline. In our simulations $k$ is configured to $20\mu s$, which is the default value for (Chaves et al. 2016). The *small scenario* traffic mix has been used in this case, as it contains video and web-traffic, and the amount of control traffic, which is exchanged between the extended Customer premises equipment and SDN controller is measured. It should be noted that the amount of clients generating traffic is constantly increasing from a single client up to 29 client.

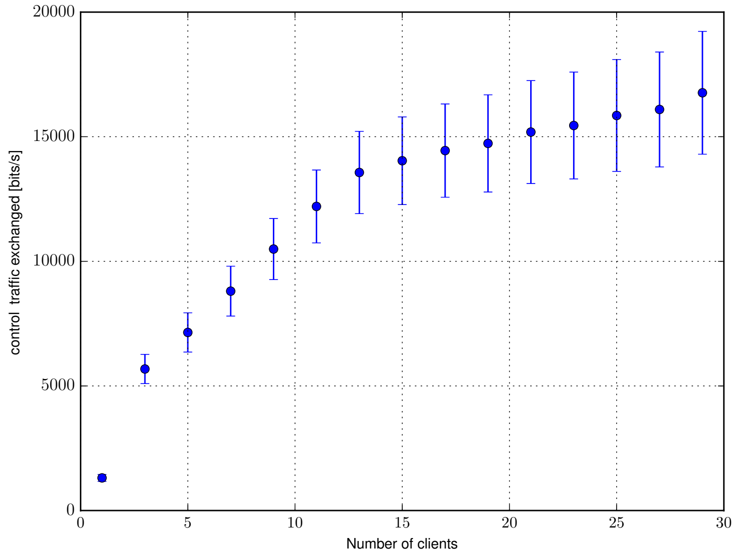



Figure 6.6: Amount of control traffic between extended Customer premises equipments and controller depending on number of clients

In Figure 6.6 the measured amount of SDN control traffic exchanged between the extended Customer premises equipments and controller depending on number of clients is shown. As previously explained, this traffic should be limited, since it will use the terrestrial WAN connection. As can be seen, the amount of SDN control traffic increases with the amount of clients. Given that each new HTTP request uses a new source port, this is obvious, since every flow is considered as a new flow by the SDN controller and the Offloading Decision Function. Hence, each flow requires a control traffic exchange between controller and extended Customer premises equipment as well as controller and concentrator. However, the rate of increase flattens out with a higher amount of clients. This can be explained by the fact that rules for typical IP control traffic, such as Address Resolution Protocol (ARP), are created just once and cover all clients, so that no additional control traffic is generated.

It should also be noted that the flow timeout is set to unlimited and the Concentrator sends the first 128 B of each packet to the controller, in case no flow rules exist. As show in Figure 6.6, the amount of control traffic exchanged between extended Customer premises equipment and SDN controller is approximately 16 kbit/s if 30 clients are active which is reasonable, even if the terrestrial up-link of 0.125 Mbit/s is used.

Hence, the impact on the flow timeout value on the amount of control traffic is tested. Therefore the same simulation is conducted again but with a fixed amount of 10 clients. Instead of changing the amount of clients, the time until an entry in the flow table of extended Customer premises equipments and concentrator times out is changed, starting from 1 s up to 30 s.

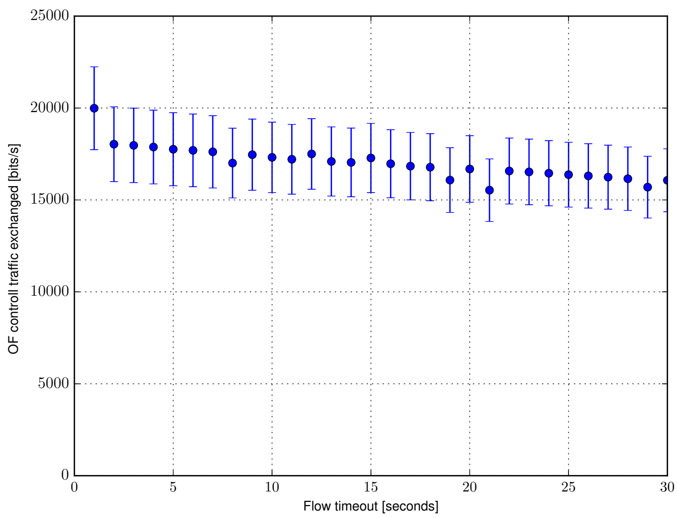



Figure 6.7: Amount of control traffic between extended Customer premises equipments and controller depending on the flow timeout

As shown in Figure 6.7, with an increasing flow timeout the control traffic rate decreases. This is mainly caused by the fact that the video flows but also the some of the HTTP flows are living longer than a short flow timeout. Hence, the message exchange between extended Customer premises equipments and the SDN controller is reduced for the same flow. This, however, leads to a larger flow table on the extended Customer premises equipments.

## 6.3 Conclusion on Network Architecture

While the previous chapter aims at designing and evaluating different traffic offloading algorithms that can be used in the considered scenario, this chapter focuses on the network architecture that allows for executing these algorithms in converged satellite and terrestrial networks. That is, the required input information for the algorithm to work needs to be provided and the decision needs to be executed in the network. Hence, this chapter first analyses two integration methods between the satellite and the terrestrial network. Due to the higher flexibility, a tight integration is selected, which implies the usage of a common anchor node between both networks. Moreover, based on the state of the art analysis of Chapter 3 a centralized SDN-based architecture has been designed and evaluated that allows for implementing the all-knowing offloading approach, which was the outcome Chapter 5. Given that keeping a low overhead of control traffic is vital for centralized approaches, the required amount of exchanged data between the extended Customer premises equipment and the SDN controller has been evaluated, which results in the finding that the designed centralized SDN-based approach, and which in turn has been extended with a Offloading Decision Function, Offloading Monitoring Function and Offloading Execution Function that can be used in the considered scenario.

In conclusion, findings of this chapter show that the designed and evaluated SDN-based network architecture enables offloading of traffic onto the satellite with the required low overhead. Moreover, this approach extends the ideas of (Kapovits et al. 2014), given that they lack a concrete implementation on how satellite and terrestrial networks can work jointly. It also confirms the authors analysis that SDN-based networks are a key element for the successful integration of satellite and terrestrial networks. The developed network architecture supports the work of (Gardikis et al. 2017) as well, which claim that SDN is suitable and required in future satellite networks.

# Chapter 7

# Conclusion

The main goal of this work is to design and evaluate a traffic offloading approach for converged satellite and terrestrial networks, so that the end users' QoE in undeserved areas can be increased. In order to achieve this goal, the following crucial research questions have been identified:

- What are the key functional building blocks to enable offloading in converged satellite and terrestrial networks?
- How to decide when and which traffic should be offloaded to the satellite network?
- How can the architecture of a converged satellite and terrestrial network look like?
- How to enable the usage of satellite links for the traffic that has been selected for offloading?

It should be noted that the prior related work has been extensively studied during the course of this thesis, so that eventually a design could be presented and the artifacts evaluated.

## 7.1 Contribution to science

Concluding this thesis, it can be stated that major key building block to form a converged satellite and terrestrial network have been successfully defined and designed, allowing the offloading of traffic onto the satellite link, in order to ultimately increase the end user's QoE. In the following, it is described how the contributions of this work answer the aforementioned research questions that are initially defined in Chapter 1.

### 7.1.1 What are the key functional building blocks to enable offloading in converged satellite and terrestrial networks?

Based on the identified challenges in Section 1.2, three major key building blocks, which are crucial in order to form a converged satellite and terrestrial network, have been defined. These are a Path Selection function that is responsible for performing the offloading decision, a Execution function that enforces the path selection and offloading in the network and a Information collecting function, which obtains the required information to enable the Path Selection function. With the definition of these key functional building blocks existing (monolithic) network architectures and methods, from various domains could be broken down and analyzed with respect to their suitability to be used in the considered scenario. The result is a thorough gap analysis, which reveals limitations with all considered approaches. Hence, a problem specific solution needs to be designed.

### 7.1.2 How to decide when and which traffic should be offloaded to the satellite network?

Given the outcome of Chapter 2 and 3, a proper state of the art offloading approach for the considered scenario is not available. Hence, several offloading methods have been designed and investigated. While a static approach that offloads all down-link traffic onto the satellite can be easily implemented, it has been shown that more flexible approach, operating more fine grained can provide a better performance. Eventually it can be shown that the designed approach, which extends policy-based mechanisms with dynamic information, such as queue fill-level achieves the best results and clearly provides a benefit to narrow-band terrestrial connection, which are common in rural and remote ares.

### 7.1.3 How can the architecture of a converged satellite and terrestrial network look like?

Basically two integration options are possible, namely a loose and a tight integration. While the first does not require any coordination between the satellite and the terrestrial network, the latter demands a common anchor node on the operator's edge of the network. However, it is shown that a loose integration imposes constraints on the actual offloading decision and also allows for less flexibility. Hence, a tight integration is the preferable option. Moreover, a tight integration is also required to support the more flexible offloading approaches, which provides the better performance, as shown in Chapter 5.

### 7.1.4 How to enable the usage of satellite links for the traffic that has been selected for offloading?

In order to enable the usage of the offloading approach designed and evaluated in Chapter 5 in a network, a proper execution methods needs to be available. Based on the outcome of Chapter 3, adopting SDN seems to be a reasonable approach. However, SDN itself does not provide a ready-to-use solution. Instead, Offloading Decision Function, Offloading Monitoring Function and Offloading Execution Function need to be defined and implemented. Eventually, it is shown that such a SDN-based architecture can offload selected traffic from the terrestrial network onto the satellite connection, while keeping a low control overhead, which is crucial in centralized architectures. Moreover, the designed approach provides a great flexibility and novel policies can be easily integrated, in order to e.g. adopt for new applications.

## 7.2 Further Considerations

During the course of this study, it has become obvious that narrow-band terrestrial networks, which are common in rural and remote areas, can be significantly enhanced by an additional satellite connection. Using proper offloading methods the performance of typical applications can be massively increased. This is shown by evaluation proper KPIs which highly impact the end user's QoE, such as the Page Load Time for web-traffic. However, it has also become obvious that a similar QoE as a broadband terrestrial connection does nowadays provide, cannot be achieved by a converged satellite and terrestrial network.

## 7.3 Limitations

The benefit of converging satellite and terrestrial networks by offloading certain traffic onto a high bandwidth satellite links has been evaluated in this thesis. However, throughout this study several assumptions have been made, which lead to certain limitations. These are discussed in the following.

### 7.3.1 Traffic class identification

As discussed in Section 1.2.2 identifying the class of certain traffic in a real network is a challenging and time-consuming task to solve. Hence, it is not considered in this thesis, yet it is been relied upon when designing and testing the all-knowing offloading approach. Since the evaluations in Chapter 5 have been done using simulations techniques, an implemented solution that preforms the task of detecting the traffic class was not required, given that the traffic class was provided through the simulation software. Obviously, such an approach will not work in real implementations. However, promising DPI or Machine Learning (ML) solutions to solve this issue exist or are currently being developed by the research community.

### 7.3.2 Network assumptions

Even though available terrestrial bandwidth in rural and remote areas is often limited as explained in Chapter 1, the actual available capacity range can vary. Moreover, a cost benefit analysis conducted by the EU BATS research project found out that converged satellite and terrestrial networks are economically reasonable if the terrestrial bandwidth is in the range of 0 Mbit/s to 8 Mbit/s. However, throughout the evaluation of the different offloading algorithms in Chapter 5 a down-link bandwidth of 1 Mbit/s has been assumed. With a different terrestrial bandwidth, the configuration of the offloading approaches might need to be adjusted, e.g. the threshold of queue-fill level, when traffic is being offloaded.

Furthermore, the evaluation of the different offloading approaches has been done with a single household. More realistically, a satellite spot beam severs a couple of hundred households. Hence, more jitter on the satellite connection can be expected. The same holds true for the assumption being made in Chapter 5 that the satellite link is ideal and free of losses. Both effects might also require adjustments to the offloading approaches.

### 7.3.3 Traffic assumption

In this thesis, basically three different traffic classes are considered during the evaluation of the offloading algorithms, namely web traffic, streaming video traffic and file-transfer traffic. These classes have been chosen, since they cover a huge portion of current and future Internet traffic as explained in Section 2.5. However, obviously other kinds of traffic exist as well and emerging applications might generate completely novel traffic patterns. Good examples in this regard are most likely online gaming, IoT, Augmented reality or Virtual Reality (VR) applications. These might lead to further evaluation of the offloading algorithms and maybe further adjustments.

## 7.4 Future work

Following-up this work, different aspects might be looked at. Given the high amount of 5G activities aiming at integration satellite connections naively into a 5G network, offloading traffic from base stations in mobile network seem promising. Moreover, defining traffic identification methods, which are out of scope of this work is another area of interest.

### 7.4.1 Remote mobile base station

Just like remote houses, mobile BSs, particularly in rural and remote areas, can also be equipped with an additional satellite connection, as depicted in Figure 7.1, in order to increase the performance of the Back-haul connection.

Future networks are expected to not only provide BE services. Instead, guaranteed services need to also be supported in order to enable novel and emerging applications. Examples of these applications include high definition video streaming, cloud-based applications, web conferencing or even Machine-Type-Communication (MTC), all of which have different requirements in terms of latency, required bandwidth, jitter and reliability (*5G radio network architecture* 2014) that need to be considered in order to satisfy the user's demands and achieve a high QoE (NetWorld2020 ETP 2014).

Given that, it is beneficial to offload certain traffic to the satellite network if the terrestrial Back-haul connection does not provide a sufficient performance. However, the additional Back-haul connection via the satellite needs to be transparent to the end users and their MTs, so that no changes on these devices are necessary.

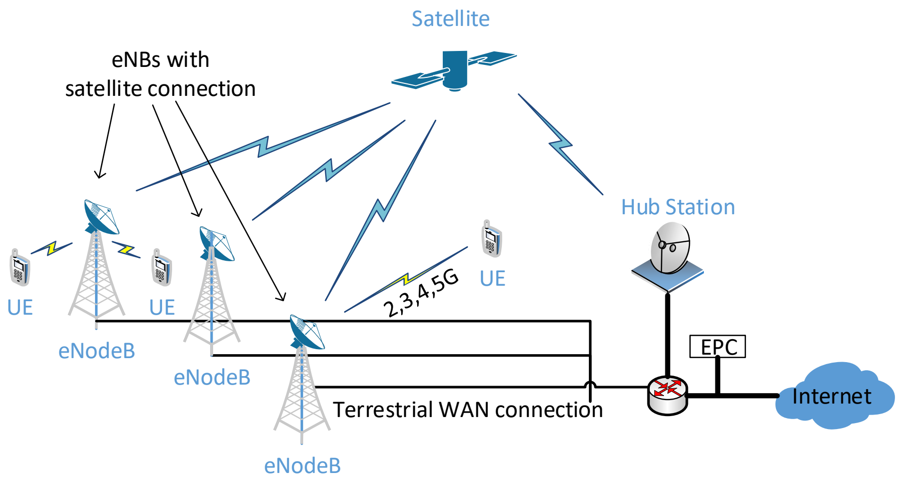



Figure 7.1: Overview remote mobile BSs scenario

With the emergence of 5G networks, this use case becomes even more appealing.On the one hand scenarios like edge computing, video distribution or application software distribution are considered typical 5G use cases that require high capacities in the network. On the other hand, 5G aims at providing ubiquitous connectivity. Both can be enabled by the integration of satellite links.

### 7.4.2 Integration of ACM

The quality of the satellite link is affected by weather, such as heavy rain, much in the same way as many other technologies transmitting wirelessly, and therefore changes more frequently. In order to avoid wasting spectral efficiency due to significant SNR buffers, satellite connections implement an ACM mechanism. ACM adapts to these changing conditions by modifying the used MCS. That is, if e.g. heavy rain appears, the modulation is changed to a more robust schema, based on feedback the receiver sends to the sender. On the other hand, the increased robustness comes with the additional cost of reduced capacity, as more redundancy is added to the signal. Once the rain stops, a more efficient MCS can be selected again. Thus, the link capacity that can be used by upper layers changes too. Furthermore, given that ACM is designed to quickly adapt to changing wireless channel conditions, these changes might occur frequently without prior notification.

As described in Chapter 6, additional information can be gathered by Offloading Monitoring Function. The changes caused by ACM can be processed here and the Offloading Decision Function can adopt its offloading decision.

# Appendix A

# Random variable used in simulations

In the following, the used random distributions and parameters are listed.

## A.1 Web traffic

| | (NGMN Alliance 2008) | Adopted parameters |
|---|---|---|
| Distribution | Truncated Lognormal distribution | Truncated Lognormal distribution |
| Mean | 10710 bytes | 25710 bytes |
| Standard deviation | 25032 bytes | 25032 bytes |
| Min | 100 bytes | 100 bytes |
| Max | 2000000 bytes | 2000000 bytes |

Table A.1: Main object size

| | (NGMN Alliance 2008) | Adopted parameters |
|---|---|---|
| Distribution | Truncated Lognormal distribution | Truncated Lognormal distribution |
| Mean | 7758 bytes | 25000 bytes |
| Standard deviation | 126168 bytes | 126168 bytes |
| Min | 50 bytes | 50 bytes |
| Max | 2000000 bytes | 2000000 bytes |

Table A.2: Embedded object size

| | (NGMN Alliance 2008) | Adopted parameters |
|---|---|---|
| Distribution | Truncated Pareto | Truncated Pareto |
| Shape | 1.1 | 1.5 |
| Scale | 2 | 60 |
| Min | 0 | 0 |
| Max | 55 | 400 |

Table A.3: Number of embedded objects per page

| | (NGMN Alliance 2008) | Adopted parameters |
|---|---|---|
| Distribution | Exponential | Exponential |
| Mean | 30s | 30s |

Table A.4: Reading Time

| | (NGMN Alliance 2008) | Adopted parameters |
|---|---|---|
| Distribution | Exponential | Exponential |
| Mean | 0.13 | 0.13s |

Table A.5: Parsing Time

## A.2 Video traffic

| | (NGMN Alliance 2008) (Low quality) | Adopted parameters (High quality) |
|---|---|---|
| Distribution | Log-normal | Log-normal |
| Mean | 3000 | 12000 |
| Standard deviation | 2400 | 2400 |
| Min | 200 | 200 |
| Max | 36000 | 144000 |

Table A.6: Number of video frames

## A.3 File-transfer traffic

| | (NGMN Alliance 2008) (Low quality) | Adopted parameters (High quality) |
|---|---|---|
| Distribution | Truncated Pareto | Truncated Pareto |
| Shape | 1.2 | 1.2 |
| Scale | 40 | 100 |
| Min | 200 | 200 |
| Max | 250 | 500 |

Table A.7: Slice size

| | (NGMN Alliance 2008) (Low quality) | Adopted parameters (High quality) |
|---|---|---|
| Distribution | Exponential | Exponential |
| Mean | 5s | 5s |

Table A.8: Idle time

| | (NGMN Alliance 2008) | Adopted parameters |
|---|---|---|
| Distribution | Truncated Lognormal distribution | Truncated Lognormal distribution |
| Mean | 2097152 bytes | 10485760 bytes |
| Standard deviation | 757072 bytes | 5242880 bytes |
| Min | 204800 bytes | 204800 bytes |
| Max | 5242880 bytes | 104857600 bytes |

Table A.9: File size

| | (NGMN Alliance 2008) | Adopted parameters |
|---|---|---|
| Distribution | Exponential | Exponential |
| Mean | 180s | 600s |

Table A.10: Reading Time

# References


3GPP TR25.892 (2004), *Feasibility Study for Orthogonal Frequency Division Multiplexing (OFDM) for UTRAN enhancement (Release 6)*.

3GPP TR36.842 (2013), *Technical Specification Group Radio Access Network; Study on Small Cell enhancements for E-UTRA and E-UTRAN; Higher layer aspects (Release 12)*.

3GPP TS22.805 (2012), *Technical Specification Group Services and System Aspects;Feasibility study on user plane congestion management (Release 23)*.

3GPP TS36.300 (2018), *LTE; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2* .

*5G radio network architecture* (2014), Whitepaper.

Abdelsalam, A., Luglio, M., Roseti, C. & Zampognaro, F. (2017), TCP connection management through combined use of terrestrial and satellite IP-based links, *in* '2017 40th International Conference on Telecommunications and Signal Processing (TSP)', IEEE.

Akyildiz, I. & Jeong, S.-H. (1997), 'Satellite atm networks: a survey', *IEEE Commun. Mag* **35**(7), 30–43.

Allman, M., Paxson, V. & Blanton, E. (2009), 'TCP Congestion Control', RFC 5681 (Draft Standard).
URL: https://www.rfc-editor.org/rfc/rfc5681.txt

Andrews, J. G., Claussen, H., Dohler, M., Rangan, S. & Reed, M. C. (2012), 'Femtocells: Past, present, and future', *IEEE Journal on Selected Areas in Communications* **30**(3), 497–508.

Angeles Vazquez Castro, M. & Vieira, F. (2012), 'Dvb-s2 full cross-layer design for qos provision', *IEEE Commun. Mag* **50**(1), 128–135.

Apostolopoulos, G., Kama, S., Williams, D., Guerin, R., Orda, A. & Przygienda, T. (1999), 'QoS Routing Mechanisms and OSPF Extensions', RFC 2676 (Experimental).
URL: `https://www.rfc-editor.org/rfc/rfc2676.txt`

Artiga, X., Nunez-Martinez, J., Perez-Neira, A., Vela, G. J. L., Garcia, J. M. F. & Ziaragkas, G. (2016), Terrestrial-satellite integration in dynamic 5g backhaul networks, *in* '2016 8th Advanced Satellite Multimedia Systems Conference and the 14th Signal Processing for Space Communications Workshop (ASMS/SPSC)', IEEE.

Audah, L., Sun, Z. & Cruickshank, H. (2011), Qos evaluation of multiservice applications over integrated satellite-terrestrial networks, *in* 'Proc. 3rd Int. Congr. on Ultra Modern Teleccommun. and Control Syst. and Workshops', pp. 1–8.

Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V. & Swallow, G. (2001), 'RSVP-TE: Extensions to RSVP for LSP Tunnels', RFC 3209 (Proposed Standard). Updated by RFCs 3936, 4420, 4874, 5151, 5420, 5711, 6780, 6790, 7274.
URL: `https://www.rfc-editor.org/rfc/rfc3209.txt`

Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M. & McManus, J. (1999), 'Requirements for Traffic Engineering Over MPLS', RFC 2702 (Informational).
URL: `https://www.rfc-editor.org/rfc/rfc2702.txt`

Babiarz, J., Chan, K. & Baker, F. (2006), 'Configuration Guidelines for DiffServ Service Classes', RFC 4594 (Informational). Updated by RFC 5865.
URL: `https://www.rfc-editor.org/rfc/rfc4594.txt`

Banks, J., Carson S II, J., L Nelson, B. & M Nicol, D. (2009), *Discrete-Event System Simulation*.

Bernaille, L., Teixeira, R., Akodkenou, I., Soule, A. & Salamatian, K. (2006), 'Traffic classification on the fly', *SIGCOMM Comput. Commun. Rev.* **36**(2), 23–26.

Bernardos (Ed.), C. (2016), 'Proxy Mobile IPv6 Extensions to Support Flow Mobility', RFC 7864 (Proposed Standard).
URL: `https://www.rfc-editor.org/rfc/rfc7864.txt`

Bertaux, L., Medjiah, S., Berthou, P., Abdellatif, S., Hakiri, A., Gelard, P., Planchou, F. & Bruyere, M. (2015), 'Software defined networking and virtualization for broadband satellite networks', *IEEE Commun. Mag* **53**(3), 54–60.

Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z. & Weiss, W. (1998), 'An Architecture for Differentiated Services', RFC 2475 (Informational). Updated by RFC 3260.
URL: `https://www.rfc-editor.org/rfc/rfc2475.txt`

Boero, L., Bruschi, R., Davoli, F., Marchese, M. & Patrone, F. (2018), 'Satellite networking integration in the 5g ecosystem: Research trends and open challenges', *IEEE Network* **32**(5), 9–15. 5G-PPP MATILD.

Border, J., Kojo, M., Griner, J., Montenegro, G. & Shelby, Z. (2001), 'Performance Enhancing Proxies Intended to Mitigate Link-Related Degradations', RFC 3135 (Informational).
URL: `https://www.rfc-editor.org/rfc/rfc3135.txt`

Botta, A. & Pescape, A. (2013), New generation satellite broadband internet services: Should adsl and 3g worry?, *in* 'IEEE 32nd Int. Conf. Comput. Commun. Workshops', pp. 399–404.

Braden (Ed.), R., Zhang, L., Berson, S., Herzog, S. & Jamin, S. (1997), 'Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification', RFC 2205 (Proposed Standard). Updated by RFCs 2750, 3936, 4495, 5946, 6437, 6780.
URL: `https://www.rfc-editor.org/rfc/rfc2205.txt`

Braden, R., Clark, D. & Shenker, S. (1994), 'Integrated Services in the Internet Architecture: an Overview', RFC 1633 (Informational).
URL: `https://www.rfc-editor.org/rfc/rfc1633.txt`

Brassil, J. (2005), An empirical study of bonded heterogeneous wan access links, *in* 'Proc. 14th IEEE Workshop on Local and Metropolitan Area Netw.', pp. 6 pp.–6.

Breiling, M., Zia, W., Sanchez de la Fuente, Y., Mignone, V., Milanesio, D., Fan, Y. & Guta, M. (2014), Lte backhauling over meo-satellites, *in* 'Proc. 7th Adv. Satellite Multimedia Syst Conf. and 13th Signal Process. for Space Commun. Workshop', pp. 174–181.

Briggs, P., Chundury, R. & Olsson, J. (2010), 'Carrier ethernet for mobile backhaul', *IEEE Commun. Mag.* **48**(10), 94–100.

Butkiewicz, M., Madhyastha, H. V. & Sekar, V. (2014), 'Characterizing web page complexity and its impact', *IEEE/ACM Transactions on Networking* **22**(3), 943–956.

Cable Television Laboratories (2019), *The DOCSIS (Data-Over-Cable Service Interface Specifications) Physical Layer Specification.*

Carisimo, E., Selmo, C., Alvarez-Hamelin, J. & Dhamdhere, A. (2018), Studying the Evolution of Content Providers in the Internet Core, *in* 'Traffic Monitoring and Analysis Workshop (TMA)'.

Chaves, L. J., Garcia, I. C. & Madeira, E. R. M. (2016), Ofswitch13: Enhancing ns-3 with openflow 1.3 support, *in* 'Proceedings of the Workshop on ns-3', ACM, pp. 33–40.

Chebrolu, K., Raman, B. & Rao, R. R. (2005), 'A network layer approach to enable tcp over multiple interfaces', *Wirel. Netw.* **11**(5), 637–650.

Chebrolu, K. & Rao, R. (2006), 'Bandwidth aggregation for real-time applications in heterogeneous wireless networks', *IEEE Trans. Mobile Comput.* **5**(4), 388–403.

Chen, S. & Nahrsted, K. (1998), 'An overview of quality of service routing for next-generation high-speed networks: problems and solutions', *IEEE Netw.* **12**(6), 64–79.

Chen, X., Wu, J., Cai, Y., Zhang, H. & Chen, T. (2015), 'Energy-efficiency oriented traffic offloading in wireless networks: A brief survey and a learning approach for heterogeneous cellular networks', *IEEE Journal on Selected Areas in Communications* **33**(4), 627–640.

Chen, Y., Farley, T. & Ye, N. (2004), 'Qos requirements of network applications on the internet', *Inf. Knowl. Syst. Manag.* **4**(1), 55–76.

Cheng, H. & Lin, K. C. (2015), 'Source selection and content dissemination for preference-aware traffic offloading', *IEEE Transactions on Parallel and Distributed Systems* **26**(11), 3160–3174.

Chong, C. K., Layman, D. A., McGeary, W. L., Menninger, W. L., Ramay, M. L. & Zhai, X. (2018), 'Q/v-band high-power uplink helix TWT for future high-data-rate communications', *IEEE Transactions on Electron Devices* **65**(6), 2201–2205.

Cisco Cooperation (2011), 'Cisco visual networking index: Global mobile data traffic forecast update, 2011-2016'.
URL: https://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf

Cisco Cooperation (2017*a*), 'Cisco visual networking index: Forecast and methodology, 2016-2021'.

Cisco Cooperation (2017*b*), 'The Zettabyte Era-Trends and Analysis', Whitepaper.

Cisco Cooperation (2020), 'Cisco Annual Internet Report 2018-2023', Whitepaper.

Claise (Ed.), B. (2008), 'Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information', RFC 5101 (Proposed Standard). Obsoleted by RFC 7011.
URL: https://www.rfc-editor.org/rfc/rfc5101.txt

de la Cuesta, B., Albiol, L., Aguiar, J. M., Baladrón, C., Carro, B. & Sánchez-Esguevillas, A. (2013), 'Innovative DAMA algorithm for multimedia DVB-RCS system', *EURASIP Journal on Wireless Communications and Networking* **2013**(1).

Du, J., Jiang, C., Zhang, H., Ren, Y. & Guizani, M. (2018), 'Auction design and analysis for SDN-based traffic offloading in hybrid satellite-terrestrial networks', *IEEE Journal on Selected Areas in Communications* **36**(10), 2202–2217.

Egilmez, H., Dane, S., Bagci, K. & Tekalp, A. (2012), Openqos: An openflow controller design for multimedia delivery with end-to-end quality of service over software-defined networks, *in* 'Proc. Asia-Pacific Signal and Informat. Process. Assoc. Annu. Summit and Conf.', pp. 1–8.

Ekmekcioglu, E., Evans, B., Kondoz, A., Thompson, P., Ottavj, L., Gamberini, F. & Watts, S. (2018), 'Context-aware hybrid satellite-terrestrial broadband access', *International Journal of Satellite Communications and Networking* .

Elwalid, A., Jin, C., Low, S. & Widjaja, I. (2001), Mate: Mpls adaptive traffic engineering, *in* 'Proc. 20th Annual Joint Conf. of the IEEE Comput. and Commun. Societies', Vol. 3, pp. 1300–1309 vol.3.

ETSI EN301 790 (2009), *Interaction channel for Satellite Distribution Systems.*

ETSI EN302 307 (2013), *Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications (DVB-S2).*

ETSI TR103 272 (2015), *Technical Report;Satellite Earth Stations and Systems (SES); Hybrid FSS satellite/terrestrial network architecture for gigh pseed boradband access.*

ETSI TS101 545-3 (2014), *DVB-RCS2 Higher Layer Satellite Specification.*

ETSI TS101 545-5 (2014), *Guidelines for the Implementation and Use of TS 101 545-3 (DVB-RCS2 Higher Layer).*

ETSI TS102 462 (2006), *Satellite Earth Stations and Systems (SES); Broadband Satellite Multimedia (BSM); QoS Functional Architecture.*

ETSI TS102 606-1 (2014), *Digital Video Broadcasting (DVB); Generic Stream Encapsulation (GSE); Part 1: Protocol.*

ETSI TS301 545-2 (2014), *DVB-RCS2 Lower Layer Satellite Specification.*

EU BATS Project (2014), 'Cost benefit analysis', Deliverable.
URL: http://cordis.europa.eu/docs/projects/cnect/3/317533/080/deliverables/001-Ares20174277027317533BATSSubmissionofD52.pdf

EU Project: Broadband Access via Integrated Terrestrial & Satellite Systems (2016), 'Business case'.
URL: https://cordis.europa.eu/docs/projects/cnect/3/317533/080/deliverables/001-Ares20174277752317533BATSdeliverablesD54.pdf

European Commission (2016), 'Connectivity for a competitive digital single market - towards a european gigabit society'.

European Commission - Directorate-General for Communication Networks, Content and Technology (CONNECT) (2012), 'Digital agenda for europe scoreboard 2012'.

Evans, B., Thompson, P., Castanet, L., Bousquet, M. & Mathiopoulos, T. (2011), Concepts and technologies for a terabit/s satellite, *in* 'Proc. 3rd Int. Conf. on Advances in Satellite and Space Communications', Budapest, Hungary.

Evans, B., Werner, M., Lutz, E., Bousquet, M., Corazza, G., Maral, G. & Rumeau, R. (2005), 'Integration of satellite and terrestrial systems in future multimedia communications', *IEEE Wireless Commun.* **12**(5), 72–80.

Fan, P., Deng, H., Boucadair, M., Reddy, T., Eckel, C. & Williams, B. (2014), 'Application Enabled Collaborative Networking: Problem Statement', draft-conet-aeon-problem-statement-01.
URL: https://tools.ietf.org/html/draft-conet-aeon-problem-statement-01

Farrel, A. & King, D. (2014), 'Unanswered Questions in the Path Computation Element Architecture', RFC 7399 (Informational).
URL: https://www.rfc-editor.org/rfc/rfc7399.txt

Farrel, A., Vasseur, J.-P. & Ash, J. (2006), 'A Path Computation Element (PCE)-Based Architecture', RFC 4655 (Informational).
URL: https://www.rfc-editor.org/rfc/rfc4655.txt

Federal Communications Commission (2018), 'Broadband deployment report'.

Fernandez, J., Taleb, T., Guizani, M. & Kato, N. (2009), 'Bandwidth aggregation-aware dynamic qos negotiation for real-time video streaming in next-generation wireless networks', *IEEE Trans. Multimedia* **11**(6), 1082–1093.

Ferrús, R., Koumaras, H., Sallent, O., Agapiou, G., Rasheed, T., Kourtis, M.-A., Boustie, C., Gélard, P. & Ahmed, T. (2016), 'SDN/NFV-enabled satellite communications networks: Opportunities, scenarios and challenges', *Physical Communication* **18**, 95–112.

Ford, A., Raiciu, C., Handley, M. & Bonaventure, O. (2013), 'TCP Extensions for Multipath Operation with Multiple Addresses', RFC 6824 (Experimental).
URL: `https://www.rfc-editor.org/rfc/rfc6824.txt`

Foteinos, V., Tsagkaris, K., Peloso, P., Ciavaglia, L. & Demestichas, P. (2014), 'Operator-friendly traffic engineering in ip/mpls core networks', *IEEE Trans. Netw. Service Manag.* **11**(3), 333–349.

Gardikis, G., Koumaras, H., Sakkas, C. & Koumaras, V. (2017), 'Towards SDN/NFV-enabled satellite networks', *Telecommunication Systems* **66**(4), 615–628.

Gayrard, J.-D. (2009), Terabit satellite: Myth or reality?, *in* '2009 First International Conference on Advances in Satellite and Space Communications', IEEE.

Gayraud, T. & Berthou, P. (2007), 'A qos architecture for dvb-rcs next generation satellite networks.', *EURASIP J. Wireless Commun. and Netw.* **2007**.

Ge, C., Wang, N., Selinis, I., Cahill, J., Kavanagh, M., Liolis, K., Politis, C., Nunes, J., Evans, B., Rahulan, Y., Nouvel, N., Boutin, M., Desmauts, J., Arnal, F., Watts, S. & Poziopoulou, G. (2019), 'QoE-assured live streaming via satellite backhaul in 5g networks', *IEEE Transactions on Broadcasting* **65**(2), 381–391.

German Federal Ministry for Economic Affairs and Energy (2018), 'Bericht zum breitbandatlas teil 1: Ergebnisse'.

Giambene, G., Kota, S. & Pillai, P. (2018), 'Satellite-5g integration: A network perspective', *IEEE Network* **32**(5), 25–31. 5G satellite working group of the 5G IEEE Roadmap initiative.

Google (2017), 'Find out how you stack up to new industry benchmarks for mobile page speed', Online.
URL: `https://think.storage.googleapis.com/docs/mobile-page-speed-new-industry-benchmarks.pdf`

Gopal, R. & BenAmmar, N. (2018), 'Framework for unifying 5g and next generation satellite communications', *IEEE Network* **32**(5), 16–24. No Project acks.

Gross, J. & Güneş, M. (2010), Introduction, *in* 'Modeling and Tools for Network Simulation', Springer Berlin Heidelberg, pp. 1–11.

Gu, C., Zhang, S. & Sun, Y. (2011), 'Realtime encrypted traffic identification using machine learning', **6**, 1009–1016.

Gundavelli (Ed.), S., Leung, K., Devarapalli, V., Chowdhury, K. & Patil, B. (2008), 'Proxy Mobile IPv6', RFC 5213 (Proposed Standard). Updated by RFCs 6543, 7864.
URL: `https://www.rfc-editor.org/rfc/rfc5213.txt`

Henkel, D., Englander, S., Kretschmer, M. & Niephaus, C. (2011), Connecting the unconnected - economic constraints and technical requirements towards a back-haul network for rural areas, *in* 'Proc. Global Commun. Conf. Workshops', pp. 1039–1044.

Hevner, A., R, A., March, S., T, S., Park, Park, J., Ram & Sudha (2004), 'Design science in information systems research', *Management Information Systems Quarterly* **28**, 75–.

Horstmann, T., Kretschmer, M., Modeker, J., Niephaus, C. & Sauer, S. (2011), Development framework for prototyping heterogeneous multi-radio wireless networks, *in* 'Computer Communications and Networks (ICCCN), 2011 Proceedings of 20th International Conference on', pp. 1–5.

Hytönen, V., Herman, B., Puttonen, J., Rantanen, S. & Kurjenniemi, J. (2014), Satellite network emulation with network simulator 3, *in* 'Ka and Broadband Communications, Navigation and Earth Observation Conference'.

IEEE 802.16m (2009), 'Evaluation Methodology Document (EMD)', IEEE 802.16 Broadband Wireless Access Working Group.

ITU-T (2009), *Recommendation G.711: Asymmetric digital subscriber line transceivers 2 (ADSL2)*.

ITU-T (2011), *Recommendation Y.1541:Network performance objectives for IP-based services*.

ITU-T (2019), *Recommendation Y.1540: Internet protocol data communication service - P packet transfer and availability performance parameters*.

Jha, S. C., Sivanesan, K., Vannithamby, R. & Koc, A. T. (2014), Dual connectivity in lte small cell networks, *in* '2014 IEEE Globecom Workshops (GC Wkshps)', pp. 1205–1210.

Jurgelionis, A., Laulajainen, J.-P., Hirvonen, M. & Wang, A. I. (2011), An empirical study of NetEm network emulation functionalities, *in* '2011 Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN)', IEEE.

Kalantari, A., Fittipaldi, M., Chatzinotas, S., Vu, T. X. & Ottersten, B. (2017), Cache-assisted hybrid satellite-terrestrial backhauling for 5g cellular networks, *in* 'GLOBECOM 2017 - 2017 IEEE Global Communications Conference', IEEE.

Kandula, S., Katabi, D., Sinha, S. & Berger, A. (2007), 'Dynamic load balancing without packet reordering', *SIGCOMM Comput. Commun. Rev.* **37**(2), 51–62.

Kapovits, A., Ververidis, C., Siris, V., Covaci, S. & Guta, M. (2014), Advanced topics in service delivery over integrated satellite terrestrial networks, *in* 'Proc. 7th Adv. Satellite Multimedia Syst Conf. and 13th Signal Process. for Space Commun. Workshop', pp. 92–98.

Kempf (Ed.), J. (2007), 'Problem Statement for Network-Based Localized Mobility Management (NETLMM)', RFC 4830 (Informational).
URL: `https://www.rfc-editor.org/rfc/rfc4830.txt`

Khodashenas, P. S., Khalili, H., Guija, D. & Siddiqui, S. (2019), TALENT: Towards integration of satellite and terrestrial networks, *in* '2019 European Conference on Networks and Communications (EuCNC)', IEEE.

Kong, P. & Karagiannidis, G. K. (2016), 'Backhaul-aware joint traffic offloading and time fraction allocation for 5g hetnets', *IEEE Transactions on Vehicular Technology* **65**(11), 9224–9235.

Kota, S., Giambene, G. & Kim, S. (2011), 'Satellite component of ngn: Integrated and hybrid networks', *Int. J. of Satellite Commun. and Netw.* **29**(3), 191–208.

Kota, S. & Marchese, M. (2003), 'Quality of service for satellite ip networks: a survey', *Int. J. of Satellite Commun. and Netw.* **21**(4-5), 303–349.

Kuipers, F., Van Mieghem, P., Korkmaz, T. & Krunz, M. (2002), 'An overview of constraint-based path selection algorithms for qos routing', *IEEE Commun. Mag* **40**(12), 50–55.

Kyrgiazos, A., Evans, B., Thompson, P., Mathiopoulos, P. T. & Papaharalabos, S. (2014), 'A terabit/second satellite system for european broadband access: a feasibility study', *International Journal of Satellite Communications and Networking* **32**(2), 63–92.

Laselva, D., Lopez-Perez, D., Rinne, M. & Henttonen, T. (2018), '3gpp lte-wlan aggregation technologies: Functionalities and performance comparison', *IEEE Communications Magazine* **56**(3), 195–203.

Lee, K.-H. & Kim, J.-H. (2015), 'Multi-way relay system with network coding in multi-spot beam satellite networks', *Wireless Networks* **23**(1), 205–217.

Lee, K.-H. & Park, K. Y. (2019), 'Overall design of satellite networks for internet services with QoS support', *Electronics* **8**(6), 683.

Lee, Y. & Choi, Y. (2001), An adaptive flow-level load control scheme for multipath forwarding, *in* P. Lorenz, ed., 'Networking - ICN 2001', Vol. 2093 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 771–779.

Li, J.-S. & Mao, C.-S. (2004), 'Providing flow-based proportional differentiated services in class-based diffserv routers', *IEE Proceedings Communications* **151**(1), 82–88.

Li, T., Zhou, H., Luo, H. & Yu, S. (2018), 'SERvICE: A software defined framework for integrated space-terrestrial satellite communication', *IEEE Transactions on Mobile Computing* **17**(3), 703–716.

Liao, R.-F. & Campbell, A. (2004), 'Dynamic core provisioning for quantitative differentiated services', *IEEE/ACM Trans. Netw.* **12**(3), 429–442.

Liolis, K., Geurtz, A., Sperber, R., Schulz, D., Watts, S., Poziopoulou, G., Evans, B., Wang, N., Vidal, O., Jou, B. T., Fitch, M., Sendra, S. D., Khodashenas, P. S. & Chuberre, N. (2018), 'Use cases and scenarios of 5g integrated satellite-terrestrial networks for enhanced mobile broadband: The SaT5g approach', *International Journal of Satellite Communications and Networking* .

Lopez-Perez, D., Laselva, D., Wallmeier, E., Purovesi, P., Lunden, P., Virtej, E., Lechowicz, P., Malkamaki, E. & Ding, M. (2016), 'Long term evolution-wireless local area network aggregation flow control', *IEEE Access* **4**, 9860–9869.

Maallawi, R., Agoulmine, N., Radier, B. & Meriem, T. B. (2015), 'A comprehensive survey on offload techniques and management in wireless access and core networks', *IEEE Communications Surveys Tutorials* **17**(3), 1582–1604.

Manner, J., Bless, R., Loughney, J. & Davies (Ed.), E. (2010), 'Using and Extending the NSIS Protocol Family', RFC 5978 (Informational).
URL: `https://www.rfc-editor.org/rfc/rfc5978.txt`

Melia, T., Bernardos, C. J., de la Oliva, A., Giust, F. & Calderon, M. (2011), 'Ip flow mobility in pmipv6 based networks: Solution design and experimental evaluation', *Wireless Personal Communications* **61**(4), 603–627.
URL: `https://doi.org/10.1007/s11277-011-0423-3`

Moy, J. (1998), 'OSPF Version 2', RFC 2328 (Internet Standard). Updated by RFCs 5709, 6549, 6845, 6860, 7474, 8042.
URL: `https://www.rfc-editor.org/rfc/rfc2328.txt`

NetWorld2020 ETP (2014), '5G: Challenges, Research Priorities, and Recommendations', Whitepaper.
URL: http://networld2020.eu/wp-content/uploads/2015/01/Joint-Whitepaper-V12-clean-after-consultation.pdf

NGMN Alliance (2008), 'Next generation mobile networks radio access performance evaluation methodology', White Paper.

Nguyen, T. & Armitage, G. (2008), 'A survey of techniques for internet traffic classification using machine learning', *IEEE Commun. Surveys Tuts.* **10**(4), 56 –76.

Nichols, K., Blake, S., Baker, F. & Black, D. (1998), 'Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers', RFC 2474 (Proposed Standard). Updated by RFCs 3168, 3260.
URL: https://www.rfc-editor.org/rfc/rfc2474.txt

Niephaus, C., Ghinea, G., Aliu, O. G., Hadzic, S. & Kretschmer, M. (2015), Sdn in the wireless context - towards full programmability of wireless network elements, *in* 'Network Softwarization (NetSoft), 2015 1st IEEE Conference on', pp. 1–6.

Niephaus, C., Hadzic, S., Aliu, O. G., Ghinea, G. & Kretschmer, M. (2015), 'Wireless back-haul: a software defined network enabled wireless back-haul network architecture for future 5g networks', *IET Networks* **4**(6), 287–295.

Niephaus, C., Kretschmer, M., Aliu, O. G., Ghinea, G. & Hadzic, S. (2014), WiBACK: a back-haul network architecture for 5g networks, *in* 'International Conference on Frontiers of Communications, Networks and Applications (ICFCNA 2014 - Malaysia)', Institution of Engineering and Technology.

Niephaus, C., Kretschmer, M. & Ghinea, G. (2013), Towards qos-aware load distribution in heterogeneous networks, *in* 'Proc. IEEE 11th Malaysia Int. Conf. on Commun.', pp. 151–156.

Niephaus, C., Kretschmer, M., Ghinea, G. & Hadzic, S. (2015), Enabling converged satellite and terrestrial access networks, *in* A. Nungu, B. Pehrson & J. Sansa-Otim, eds, 'e-Infrastructure and e-Services for

Developing Countries', Vol. 147 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, Springer International Publishing, pp. 43–52.
URL: http://dx.doi.org/10.1007/978-3-319-16886-9_5

Niephaus, C., Modeker, J. & Ghinea, G. (2019), 'Toward traffic offload in converged satellite and terrestrial networks', *IEEE Transactions on Broadcasting* **65**(2), 340–346.

Nunes, B. A. A., Mendonca, M., Nguyen, X.-N., Obraczka, K. & Turletti, T. (2014), 'A survey of software-defined networking: Past, present, and future of programmable networks', *IEEE Commun. Surveys Tuts.* **16**(3), 1617–1634.

Open Networking Foundation (2012), 'Software-Defined Networking: The New Norm for Networks', Whitepaper.
URL: http://www.opennetworking.org/images/stories/downloads/sdn-resources/white-papers/wp-sdn-newnorm.pdf

Open Networking Foundation (2015), *OpenFlow Switch Specification 1.5.1*.

Ortíz-Gómez, F., Rodríguez-Osorio, R. M., Salas-Natera, M. & Landeros-Ayala, S. (2018), Adaptive resources allocation for flexible payload enabling VHTS systems: Methodology and architecture, *in* '36th International Satellite Communications Systems Conference (ICSSC2018)', Institution of Engineering and Technology.

Papadimitriou, D., Welzl, M., Scharf, M. & Briscoe, B. (2011), 'Open Research Issues in Internet Congestion Control', RFC 6077 (Informational).
URL: https://www.rfc-editor.org/rfc/rfc6077.txt

Pana, F. & Put, F. (2013), 'A survey on the evolution of rsvp', *IEEE Commun. Surveys Tuts.* **15**(4), 1859–1887.

Pawlikowski, K., Jeong, H.-D. & Lee, J.-S. (2002), 'On credibility of simulation studies of telecommunication networks', *IEEE Communications Magazine* **40**(1), 132–139.

Perkins (Ed.), C. (2010), 'IP Mobility Support for IPv4, Revised', RFC 5944 (Proposed Standard).
URL: https://www.rfc-editor.org/rfc/rfc5944.txt

Perkins (Ed.), C., Johnson, D. & Arkko, J. (2011), 'Mobility Support in IPv6', RFC 6275 (Proposed Standard).
URL: https://www.rfc-editor.org/rfc/rfc6275.txt

Peters, G., Perez-Trufero, J., Watts, S. & Wall, N. (2013), The bats project, *in* 'Proc. Future Netw. and Mobile Summit', pp. 1–8.

Postel, J. (1981), 'Transmission Control Protocol', RFC 793 (Internet Standard). Updated by RFCs 1122, 3168, 6093, 6528.
URL: https://www.rfc-editor.org/rfc/rfc793.txt

Prabhavat, S., Nishiyama, H., Ansari, N. & Kato, N. (2011), 'Effective delay-controlled load distribution over multipath networks', *IEEE Trans. Parallel Distrib. Syst.* **22**(10), 1730–1741.

Prabhavat, S., Nishiyama, H., Ansari, N. & Kato, N. (2012), 'On load distribution over multipath networks', *IEEE Commun. Surveys Tuts.* **14**(3), 662 –680.

Puttonen, J., Herman, B., Rantanen, S., Laakso, F. & Kurjenniemi, J. (2015), Satellite network simulator 3, *in* 'Workshop on Simulation for European Space Programmes', ESA-ESTEC, Noordwijk, The Netherlands.

Puttonen, J. & Kurjenniemi, J. (2016), Radio resource management in DVB-RCS2 satellite systems, *in* '34th AIAA International Communications Satellite Systems Conference', American Institute of Aeronautics and Astronautics.

Puttonen, J., Nihtilä, T., amd Riku Järvinen, M. I., Kurjenniemi, J., Hytönen, V. & Flink, T. (2019), Cloud simulation platform for satellite and terrestrial network simulations, *in* 'Ka and Broadband Communications Conference'.

Puttonen, J., Rantanen, S., Laakso, F., Kurjenniemi, J., Aho, K. & Acar, G. (2014), Satellite model for network simulator 3, *in* 'Proceeding of

7th International ICST Conference on Simulation Tools and Techniques (SIMUtools), Lisbon'.

Puttonen, J., Rantanen, S., Laakso, F., Kurjenniemi, J., Herman, B., Alanen, O. & Aho, K. (2016), Simulator architecture and design definition, Technical report, ESA.

Riley, G. F. & Henderson, T. R. (2010), The ns-3 network simulator, *in* 'Modeling and Tools for Network Simulation', Springer Berlin Heidelberg, pp. 15–34.

Rosen, E., Viswanathan, A. & Callon, R. (2001), 'Multiprotocol Label Switching Architecture', RFC 3031 (Proposed Standard). Updated by RFCs 6178, 6790.
URL: `https://www.rfc-editor.org/rfc/rfc3031.txt`

Samdanis, K., Taleb, T. & Schmid, S. (2012), 'Traffic offload enhancements for eutran', *IEEE Communications Surveys Tutorials* **14**(3), 884–896.

Sandvine - Intelligent Broadband Networks (2016), 'Global internet phenomena', https://www.sandvine.com/hubfs/downloads/archive/2016-global-internet-phenomena-report-latin-america-and-north-america.pdf.

Sarkar, N. I. & Gutierrez, J. A. (2014), 'Revisiting the issue of the credibility of simulation studies in telecommunication networks: highlighting the results of a comprehensive survey of IEEE publications', *IEEE Communications Magazine* **52**(5), 218–224.

Scharf, M. & Ford, A. (2013), 'MPTCP Application Interface Considerations', draft-ietf-mptcp-api-07.
URL: `http://tools.ietf.org/html/draft-ietf-mptcp-api-07`

Sciancalepore, V., Giustiniano, D., Banchs, A. & Hossmann-Picu, A. (2016), 'Offloading cellular traffic through opportunistic communications: Analysis and optimization', *IEEE Journal on Selected Areas in Communications* **34**(1), 122–137.

Sharma, S. K., Bogale, T. E., Le, L. B., Chatzinotas, S., Wang, X. & Ottersten, B. (2018), 'Dynamic spectrum sharing in 5g wireless networks with full-duplex technology: Recent advances and research challenges', *IEEE Communications Surveys & Tutorials* **20**(1), 674–707.

Shenker, S., Partridge, C. & Guerin, R. (1997), 'Specification of Guaranteed Quality of Service', RFC 2212 (Proposed Standard).
URL: `https://www.rfc-editor.org/rfc/rfc2212.txt`

Simpson (Ed.), W. (1994), 'The Point-to-Point Protocol (PPP)', RFC 1661 (Internet Standard). Updated by RFC 2153.
URL: `https://www.rfc-editor.org/rfc/rfc1661.txt`

Soto, I., Bernardos, C. J., Calderón, M. & Melia, T. (2010), 'Pmipv6: A network-based localized mobility management solution', *The Internet Protocol Journal* **13**(3).

Sou, S.-I. (2013), 'Mobile data offloading with policy and charging control in 3gpp core network', *IEEE Transactions on Vehicular Technology* **62**(7), 3481–3486.

Taleb, T., Hadjadj-Aoul, Y. & Ahmed, T. (2011), 'Challenges, opportunities, and solutions for converged satellite and terrestrial networks', *IEEE Wireless Commun.* **18**(1), 46–52.

Tan, K., Wu, D., Chan, A. J. & Mohapatra, P. (2011), 'Comparing simulation tools and experimental testbeds for wireless mesh networks', *Pervasive and Mobile Computing* **7**(4), 434–448.

*The Pigeon versus the Computer: A Surprising Win For All* (n.d.).
URL: `http://www.pigeonrace2009.co.za/`

Thomson, S., Narten, T. & Jinmei, T. (2007), 'IPv6 Stateless Address Autoconfiguration', RFC 4862 (Draft Standard). Updated by RFC 7527.
URL: `https://www.rfc-editor.org/rfc/rfc4862.txt`

Vaishnavi, V. K. & Kuechler, Jr., W. (2015), *Design Science Research Methods and Patterns: Innovating Information and Communication Technology*, 2nd edn, Auerbach Publications, Boston, MA, USA.

Vali, D., Paskalis, S., Merakos, L. & Kaloxylos, A. (2004), 'A survey of internet qos signaling', *IEEE Commun. Surveys Tuts.* **6**(4), 32–43.

Wang, N., Nouwell, N., Ge, C., Evans, B., Rahulan, Y., Boutin, M., Desmauts, J., Liolis, K., Politis, C., Votts, S. & Poziopoulou, G. (2018), Satellite support for enhanced mobile broadband content delivery in 5g, *in* '2018 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)', IEEE.

Watts, S. & Aliu, O. (2014), 5g resilient backhaul using integrated satellite networks, *in* 'Proc. 7th Adv. Satellite Multimedia Syst Conf. and 13th Signal Process. for Space Commun. Workshop', pp. 114–119.

Wehrle, K., Gnes, M. & Gross, J. (2010), *Modeling and Tools for Network Simulation*, 1st edn, Springer Publishing Company, Incorporated.

Wendong, W., Qinglei, Q., Xiangyang, G., Yannan, H. & Xirong, Q. (2014), 'Autonomic qos management mechanism in software defined network', *China Commun.* **11**(7), 13–23.

Wroclawski, J. (1997), 'The Use of RSVP with IETF Integrated Services', RFC 2210 (Proposed Standard).
URL: https://www.rfc-editor.org/rfc/rfc2210.txt

Xia, W., Wen, Y., Foh, C., Niyato, D. & Xie, H. (2014), 'A survey on software-defined networking', *IEEE Commun. Surveys Tuts.* **PP**(99), 1–1.

Xiao, X. & Ni, L. (1999), 'Internet qos: a big picture', *IEEE Netw.* **13**(2), 8–18.

Yuan, X. (2002), 'Heuristic algorithms for multiconstrained quality-of-service routing', *IEEE/ACM Trans. Netw.* **10**(2), 244–256.

Zampognaro, F. (2009), Satellite and terrestrial network integration, PhD thesis, University of Rome Tor Vergata.

Zhang, Q., Chen, J. & Zhu, H. (2014), 'Network convergence: theory, architectures, and applications', *IEEE Wireless Commun.* **21**(6), 48–53.

Ziaragkas, G., Poziopoulou, G., Núñez-Martínez, J., Baranda, J., Moreno, I., Tsinos, C., Maleki, S., Sharma, S. K., Alodeh, M. & Chatzinotas,

S. (2017), 'SANSA-hybrid terrestrial-satellite backhaul network: scenarios, use cases, KPIs, architecture, network and physical layer techniques', *International Journal of Satellite Communications and Networking* **35**(5), 379–405.