

Mälardalen University Press Dissertations  
No. 133

# **NETWORK OVERLOAD AVOIDANCE BY TRAFFIC ENGINEERING AND CONTENT CACHING**

**Henrik Abrahamsson**

**2012**



**MÄLARDALEN UNIVERSITY  
SWEDEN**

School of Innovation, Design and Engineering

Copyright © Henrik Abrahamsson, 2012

ISBN 978-91-7585-087-1

ISSN 1651-4238

Printed by Mälardalen University, Västerås, Sweden

Mälardalen University Press Dissertations  
No. 133

NETWORK OVERLOAD AVOIDANCE BY TRAFFIC  
ENGINEERING AND CONTENT CACHING

Henrik Abrahamsson

Akademisk avhandling

som för avläggande av teknologie doktorsexamen i datavetenskap vid  
Akademin för innovation, design och teknik kommer att offentligen försvaras  
onsdagen den 19 december 2012, 13.15 i Kappa, Mälardalens högskola, Västerås.

Fakultetsopponent: Doctor Luca Muscariello, Orange Telecom



Akademin för innovation, design och teknik

## Abstract

The Internet traffic volume continues to grow at a great rate, now driven by video and TV distribution. For network operators it is important to avoid congestion in the network, and to meet service level agreements with their customers. This thesis presents work on two methods operators can use to reduce links loads in their networks: traffic engineering and content caching.

This thesis studies access patterns for TV and video and the potential for caching. The investigation is done both using simulation and by analysis of logs from a large TV-on-Demand system over four months.

The results show that there is a small set of programs that account for a large fraction of the requests and that a comparatively small local cache can be used to significantly reduce the peak link loads during prime time. The investigation also demonstrates how the popularity of programs changes over time and shows that the access pattern in a TV-on-Demand system very much depends on the content type.

For traffic engineering the objective is to avoid congestion in the network and to make better use of available resources by adapting the routing to the current traffic situation. The main challenge for traffic engineering in IP networks is to cope with the dynamics of Internet traffic demands.

This thesis proposes L-balanced routings that route the traffic on the shortest paths possible but make sure that no link is utilised to more than a given level  $L$ . L-balanced routing gives efficient routing of traffic and controlled spare capacity to handle unpredictable changes in traffic. We present an L-balanced routing algorithm and a heuristic search method for finding L-balanced weight settings for the legacy routing protocols OSPF and IS-IS. We show that the search and the resulting weight settings work well in real network scenarios.

Swedish Institute of Computer Science  
Doctoral Thesis  
SICS Dissertation Series 58

Network Overload Avoidance by Traffic  
Engineering and Content Caching

Henrik Abrahamsson

2012



Swedish Institute of Computer Science  
Stockholm, Sweden



# Abstract

The Internet traffic volume continues to grow at a great rate, now driven by video and TV distribution. For network operators it is important to avoid congestion in the network, and to meet service level agreements with their customers. This thesis presents work on two methods operators can use to reduce links loads in their networks: traffic engineering and content caching.

This thesis studies access patterns for TV and video and the potential for caching. The investigation is done both using simulation and by analysis of logs from a large TV-on-Demand system over four months.

The results show that there is a small set of programs that account for a large fraction of the requests and that a comparatively small local cache can be used to significantly reduce the peak link loads during prime time. The investigation also demonstrates how the popularity of programs changes over time and shows that the access pattern in a TV-on-Demand system very much depends on the content type.

For traffic engineering the objective is to avoid congestion in the network and to make better use of available resources by adapting the routing to the current traffic situation. The main challenge for traffic engineering in IP networks is to cope with the dynamics of Internet traffic demands.

This thesis proposes L-balanced routings that route the traffic on the shortest paths possible but make sure that no link is utilised to more than a given level  $L$ . L-balanced routing gives efficient routing of traffic and controlled spare capacity to handle unpredictable changes in traffic. We present an L-balanced routing algorithm and a heuristic search method for finding L-balanced weight settings for the legacy routing protocols OSPF and IS-IS. We show that the search and the resulting weight settings work well in real network scenarios.





# Sammanfattning

Trafiken på Internet fortsätter att växa i snabb takt nu pådrivet av TV- och videodistribution över nätet. För nätverksoperatörer är det viktigt att förstå och hantera trafikbeteendet för att undvika överlast i nätet och för att kunna tillhandahålla kommunikationstjänster av god kvalitet. Den här avhandlingen handlar om två olika tillvägagångssätt för att undvika överlast i nätet: lastbalansering och lokal mellanlagring.

I den här avhandlingen undersöks användarbeteenden och efterfrågemönster för TV och video och potentialen för lokal mellanlagring. Undersökningen görs dels med simulering och dels genom analys av loggar från ett stort TV-system.

Resultaten visar att det är en liten andel av programmen som står för en stor del av efterfrågan. I många fall kan man hantera 50% av efterfrågan genom att lagra 5% av utbudet. Studien visar också att programutbud och genre har stor inverkan på efterfrågemönster och på hur snabbt programmen avtar i popularitet. Det är också stora dygnsvariationer i efterfrågan och det är viktigt att lagra rätt program för att hantera toppar i efterfrågan under kvällstid.

För lastbalansering i IP-nätverk är målet att kunna anpassa vägvalet efter den aktuella trafiksituationen och balansera trafiken över flera vägar genom nätverket om det behövs. Man kan på så vis utnyttja nätverket mer effektivt och undvika överlast. Utmaningen ligger i att Internettrafik ofta är skurig med stora variationer i trafikens mängd och riktning.

I den här avhandlingen föreslås så kallat L-balanserat vägval där trafiken skickas kortast möjliga väg men man ser till att ingen länk lastas till mer än en given nivå L. L-balanserat vägval ger en kontrollerad reservkapacitet för att hantera oförutsägbara förändringar i trafiken. Vi presenterar en L-balanserat vägvalsalgoritm samt en heuristisk sökmetod för att hitta L-balanserade vikt-sättningar i vägvalsprotokollen OSPF och IS-IS. Vi visar att sökmetoden och de resulterande viktinställningarna fungerar väl i verkliga nätverksscenarier.



# Acknowledgements

During the years that I have worked on this thesis, I have been involved in many different projects and collaborations. I have on numerous occasions discussed the research with colleagues both in Kista and Västerås, but also at project meetings, workshops and conferences in many other places: from Nanjing and Beijing to Ottawa, Boston, Madrid and Zurich, to name a few. I want to thank all research colleagues who have been involved in the discussions that have helped to shape the ideas that I present in this thesis.

I especially want to thank my advisor Mats Björkman. Without his encouragement, guidance and support this thesis would never have been finished. I also want to thank my co-advisor Bengt Ahlgren for his support and for giving me the opportunity to work in the NETS group and to do research at SICS.

I want to thank Anders Gunnar and Per Kreuger for many interesting discussions on all aspects of traffic and cache management and for great collaboration on many research papers and project deliverables. I would also like to thank Adam Dunkels for many inspiring discussions over the years.

Thanks also to Mattias Nordmark and Göran Olofsson for our collaboration on access patterns for TV-on-Demand, which also made it possible to write the final paper and complete this thesis.

Many thanks also to Ian Marsh, Laura Feeney, Anders Lindgren, Mudassar Aslam, Oliver Schwarz, and all other, current and former, colleagues at SICS for your support and for creating a stimulating work environment.

*Much of the work in this thesis has been performed within the SICS Center for Networked Systems funded by VINNOVA, KKS, SSF, ABB, Ericsson, Saab SDS, TeliaSonera, T2Data, Vendolocus and Peerialism.*



# List of publications

## Publications included in the thesis

1. Henrik Abrahamsson, Juan Alonso, Bengt Ahlgren, Anders Andersson and Per Kreuger. A Multi Path Routing Algorithm for IP Networks Based on Flow Optimisation. In *Proceedings of Third COST 263 International Workshop on Quality of Future Internet Services(QoFIS 2002)*, Zurich, Switzerland, October 2002.
2. Henrik Abrahamsson and Mats Björkman. Robust Traffic Engineering using L-balanced Weight-Settings in OSPF/ISIS. In: *Sixth International Conference on Broadband Communications, Networks, and Systems (BROADNETS 2009)*, Madrid, Spain, September 2009.
3. Henrik Abrahamsson and Mats Björkman. Simulation of IPTV caching strategies. In: *International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS'10)*, Ottawa, Canada, July 2010.
4. Henrik Abrahamsson and Mats Björkman. Caching for IPTV distribution with time-shift. In: *International conference on Computing, Networking & Communications (ICNC'13)*, January 2013, San Diego, USA.
5. Henrik Abrahamsson and Mattias Nordmark. Program popularity and viewer behaviour in a large TV-on-Demand system. In: *Internet Measurement Conference (IMC'12)*, Boston, USA, November 2012.

**Other publications**

In addition to the papers included in the thesis I have also co-authored the following papers:

1. Simon Schütz, Henrik Abrahamsson, Bengt Ahlgren, and Marcus Brunner, Design and Implementation of the Node Identity Internetworking Architecture. In *Computer Networks*, 54(7):1142-1154, May, 2010.
2. Per Kreuger and Henrik Abrahamsson, Scheduling IPTV content pre-distribution, In *Proceedings of the 9th IEEE International Workshop on IP Operations and Management (IPOM'09)*, Venice, Italy, October, 2009.
3. Henrik Abrahamsson and Per Kreuger, A Case for Resource Management in IPTV distribution, In *Proceedings of the 5th Swedish National Computer Networking Workshop (SNCNW)*, Karlskrona, Sweden, April 2008.
4. G. Leduc, H. Abrahamsson, S. Balon, S. Bessler, M. D'Arienzo, O. Delcourt, J. Domingo-Pascual, S. Cerav-Erbas, I. Gojmerac, X. Masip, A. Pescapè, B. Quoitin, S.F. Romano, E. Salvatori, F. Skivée, H.T. Tran, S. Uhlig, and H. Ümit. An open source traffic engineering toolbox , In *Computer Communications*, 29(5):593-610, March, 2006.
5. Anders Gunnar, Henrik Abrahamsson, and Mattias Söderqvist. Performance of Traffic Engineering in Operational IP-Networks - An Experimental Study, In *Proceedings of the 5th IEEE International Workshop on IP Operations and Management (IPOM'05)*, Barcelona, Spain, October, 2005.
6. M. Brunner, A. Galis, L. Cheng, J. Andres Colas, B. Ahlgren, A. Gunnar, H. Abrahamsson, R. Szabo, S. Csaba, J. Nielsen, S. Schütz A. Gonzalez Prieto, R. Stadler, and G. Molnar, Towards Ambient Networks Management, In *Proceedings of the Second International Workshop on Mobility Aware Technologies and Applications (MATA 2005)*, Montreal, Canada, October, 2005.
7. Henrik Abrahamsson and Anders Gunnar, Traffic Engineering in Ambient Networks: Challenges and Approaches , In *Proceedings of the Second Swedish National Computer Networking Workshop (SNCNW)*, Karlstad, Sweden, November, 2004.

8. M. Brunner, A. Galis, L. Cheng, J. Andres Colas, B. Ahlgren, A. Gunnar, H. Abrahamsson, R. Szabo, S. Csaba, J. Nielsen, A. Gonzalez Prieto, R. Stadler, and G. Molnar, Ambient Networks Management Challenges and Approaches, In *Proceedings of the First International Workshop on Mobility Aware Technologies and Applications (MATA 2004)*, Florianopolis, Brazil, October, 2004.
9. Henrik Abrahamsson, Olof Hagsand, and Ian Marsh. TCP over High Speed Variable Capacity Links: A Simulation Study for Bandwidth Allocation , In *Proceedings of the 7th International Workshop on Protocols For High-Speed Networks (PfHSN 2002)*, Berlin, Germany, April, 2002.
10. Henrik Abrahamsson and Bengt Ahlgren. Using empirical distributions to characterize web client traffic and to generate synthetic traffic. In *Proceedings of IEEE Globecom: Global Internet*, San Francisco, USA, November 2000.





# Contents

|          |  |           |
|----------|--|-----------|
| <b>I</b> | <b>Thesis</b>  | <b>1</b>  |
| <b>1</b> | <b>Introduction</b>  | <b>3</b>  |
| 1.1      | Internet – a network of networks . . . . .   | 3         |
| 1.2      | Traffic characteristics and access patterns . . . . .  | 4         |
| 1.3      | Television and video over IP . . . . .   | 5         |
| 1.4      | Overload avoidance . . . . .   | 8         |
| 1.4.1    | Traffic management . . . . .   | 8         |
| 1.4.2    | Caching . . . . .  | 9         |
| 1.5      | Outline of thesis . . . . .  | 10        |
| <b>2</b> | <b>Research Issues and Scientific Contributions</b>  | <b>11</b> |
| 2.1      | Robust traffic engineering . . . . .   | 11        |
| 2.1.1    | Contributions . . . . .  | 12        |
| 2.2      | Understanding TV-on-Demand access patterns and their impact on caching . . . . .             | 12        |
| 2.2.1    | Contributions . . . . .  | 12        |
| <b>3</b> | <b>Summary of the Papers and Their Contributions</b>   | <b>15</b> |
| 3.1      | Paper A: A Multi Path Routing Algorithm for IP Networks Based on Flow Optimisation . . . . . | 16        |
| 3.2      | Paper B: Robust Traffic Engineering using L-balanced Weight-Settings in OSPF/ISIS . . . . .  | 17        |
| 3.3      | Paper C: Simulation of IPTV caching strategies . . . . .                                     | 17        |
| 3.4      | Paper D: Caching for IPTV distribution with time-shift . . . . .                             | 18        |
| 3.5      | Paper E: Program popularity and viewer behaviour in a large TV-on-Demand system . . . . .    | 19        |

|           |  |           |
|-----------|--|-----------|
| <b>4</b>  | <b>Related Work</b>  | <b>21</b> |
| 4.1       | Traffic engineering in IP networks . . . . .                                     | 21        |
| 4.2       | Access patterns and potential for caching for TV and video on-demand . . . . .   | 23        |
| <b>5</b>  | <b>Conclusions and Future Work</b>   | <b>25</b> |
| 5.1       | Conclusions . . . . .  | 25        |
| 5.2       | Future work . . . . .  | 26        |
|           | <b>Bibliography</b>  | <b>31</b> |
| <b>II</b> | <b>Included Papers</b>   | <b>43</b> |
| <b>6</b>  | <b>Paper A:</b>  |           |
|           | <b>A Multi Path Routing Algorithm for IP Networks Based on Flow Optimisation</b> | <b>45</b> |
| 6.1       | Introduction . . . . .   | 47        |
| 6.2       | Architecture . . . . .   | 48        |
| 6.3       | Optimisation . . . . .   | 50        |
| 6.3.1     | Desirable Solutions . . . . .  | 51        |
| 6.3.2     | How to Obtain Desirable Solutions . . . . .                                      | 52        |
| 6.3.3     | The Result . . . . .   | 53        |
| 6.3.4     | A Generalisation . . . . .   | 54        |
| 6.3.5     | Quantitative Results . . . . .   | 54        |
| 6.4       | Multi-Path Forwarding . . . . .  | 55        |
| 6.5       | Related Work . . . . .   | 56        |
| 6.6       | Conclusions . . . . .  | 57        |
|           | Bibliography . . . . .   | 57        |
| <b>7</b>  | <b>Paper B:</b>  |           |
|           | <b>Robust Traffic Engineering using L-balanced Weight-Settings in OSPF/IS-IS</b> | <b>61</b> |
| 7.1       | Introduction . . . . .   | 63        |
| 7.2       | Traffic Engineering in IP networks . . . . .                                     | 64        |
| 7.3       | Related work . . . . .   | 65        |
| 7.4       | L-balanced solutions . . . . .   | 67        |
| 7.4.1     | Optimal l-balanced routing . . . . .   | 67        |
| 7.4.2     | Search for l-balanced weight settings . . . . .                                  | 67        |
| 7.4.3     | How to determine weight increments for a link? . . . . .                         | 69        |

|       |  |    |
|-------|--|----|
| 7.4.4 | How to determine ECMP weight settings? . . . . .             | 70 |
| 7.4.5 | Increment weight on a less utilised link in a path . . . . . | 71 |
| 7.4.6 | Comments on the search method . . . . .                      | 72 |
| 7.5   | Evaluation . . . . .   | 73 |
| 7.5.1 | Method . . . . .   | 73 |
| 7.5.2 | Static scenario: Evaluating the search method . . . . .      | 73 |
| 7.5.3 | Dynamic scenario: Evaluation of robustness . . . . .         | 78 |
| 7.6   | Conclusions . . . . .  | 78 |
|       | Bibliography . . . . .                                       | 80 |

**8 Paper C:**

|       |   |           |
|-------|---|-----------|
|       | <b>Simulation of IPTV caching strategies</b>    | <b>85</b> |
| 8.1   | Introduction . . . . .                          | 87        |
| 8.2   | IPTV and time-shifted TV . . . . .              | 88        |
| 8.3   | Simulation of IPTV . . . . .                    | 89        |
| 8.3.1 | Workload model . . . . .                        | 89        |
| 8.3.2 | Data set . . . . .                              | 89        |
| 8.3.3 | TV programs . . . . .                           | 91        |
| 8.3.4 | TV viewers . . . . .                            | 91        |
| 8.3.5 | Network model and simulation scenario . . . . . | 93        |
| 8.4   | Caching strategies . . . . .                    | 94        |
| 8.4.1 | Least Recently Used . . . . .                   | 94        |
| 8.4.2 | Least Frequently Used . . . . .                 | 94        |
| 8.4.3 | Clairvoyant . . . . .                           | 94        |
| 8.5   | Evaluation . . . . .                            | 95        |
| 8.6   | Related Work . . . . .                          | 96        |
| 8.7   | Future work . . . . .                           | 99        |
| 8.8   | Conclusions . . . . .                           | 101       |
|       | Bibliography . . . . .                          | 102       |

**9 Paper D:**

|       |  |            |
|-------|--|------------|
|       | <b>Caching for IPTV distribution with time-shift</b> | <b>107</b> |
| 9.1   | Introduction . . . . .                               | 109        |
| 9.2   | On TV viewing behaviour . . . . .                    | 110        |
| 9.2.1 | Traditional linear TV . . . . .                      | 110        |
| 9.2.2 | Time-shifted TV . . . . .                            | 111        |
| 9.3   | Simulation of IPTV with time-shift . . . . .         | 111        |
| 9.3.1 | Network model and simulation scenario . . . . .      | 113        |
| 9.3.2 | TV viewers . . . . .                                 | 113        |

|       |   |     |
|-------|---|-----|
| 9.3.3 | TV programs . . . . .                                       | 114 |
| 9.3.4 | Cache replacement policy . . . . .                          | 116 |
| 9.4   | Simulation results . . . . .                                | 117 |
| 9.4.1 | Impact of cache size and cache replacement policy . . . . . | 117 |
| 9.4.2 | Impact of on-demand time and program set size . . . . .     | 118 |
| 9.4.3 | Impact of program popularity . . . . .                      | 120 |
| 9.5   | Related Work . . . . .                                      | 121 |
| 9.6   | Discussion . . . . .  | 122 |
|       | Bibliography . . . . .                                      | 122 |

**10 Paper E:**

|        |   |            |
|--------|---|------------|
|        | <b>Program Popularity and Viewer Behaviour in a Large TV-on-Demand System</b>               | <b>127</b> |
| 10.1   | Introduction . . . . .  | 129        |
| 10.2   | The data set . . . . .  | 130        |
| 10.3   | Access patterns . . . . .   | 132        |
| 10.3.1 | Access pattern over a week . . . . .  | 132        |
| 10.3.2 | Daily and hourly change in user interest . . . . .  | 134        |
| 10.4   | Program popularity . . . . .  | 137        |
| 10.4.1 | Access patterns per program category . . . . .  | 139        |
| 10.4.2 | Access patterns for individual programs: how program popularity changes over time . . . . . | 139        |
| 10.5   | Impact on caching . . . . .   | 148        |
| 10.5.1 | Cacheability . . . . .  | 148        |
| 10.5.2 | Limited cache size . . . . .  | 150        |
| 10.6   | Related Work . . . . .  | 153        |
| 10.7   | Future Work . . . . .   | 154        |
| 10.8   | Conclusions . . . . .   | 156        |
| 10.9   | Acknowledgments . . . . .   | 156        |
|        | Bibliography . . . . .  | 156        |

# **I**

# **Thesis**



# Chapter 1

## Introduction

The Internet is a worldwide communication network that today serves billions of Internet users [1]. It is a giant infrastructure of optical fibres, copper wires and wireless connections that via packet switches connect a wide variety of end-hosts: ranging from servers in data centers to PC:s and laptop computers, to mobile phones and smaller devices embedded in our homes, in cars and in the environment around us. The Internet is also an infrastructure that supports a diversity of applications like the web, mail, file sharing, social networking services, telephony, radio, video and TV distribution, games, banking and commerce of many kinds; and where new applications constantly are developed and deployed.

Internet traffic volumes continue to grow at a great rate. For network operators it is important to avoid congestion in the network, and to meet service level agreements with their customers. This thesis presents work on two methods operators can use to reduce links loads and avoid congestion in their networks: traffic engineering and caching of video and TV content.

### 1.1 Internet – a network of networks

The Internet is a network of networks. It consists of a large number of independently managed networks of different sizes, different capacities, and under different administrations. When you click on a link in your web browser the requested webpage often travels over many different networks, sometimes worldwide, on the way to your computer. The view point in this thesis is often

from one operator network and the challenge of understanding and handling traffic demands to avoid overload in the network.

The structure of the Internet and how traffic flows between networks are changing over time [2], often driven by commercial interests and business agreements. The traditional view is that the networks that constitute the Internet are Internet Service Provider (ISP) networks connected together in a loose hierarchy. At the top there are a small number of tier-1 operators (for instance AT&T, Level 3, and TeliaSonera International carrier [3]) with large international high-capacity networks, that directly connect to each other. The tier-1 operators have peering agreements that allow data to flow between the networks without charging each other for the data transmitted. A tier-2 network is typically a regional or national network. It can have peering agreements with other tier-2 networks to exchange traffic but it is also a customer to one or more tier-1 operators and need to buy transit to reach some parts of the Internet. At the bottom of the network hierarchy are the access networks that connect the end hosts to the Internet. These are typically local telephone companies, university or company networks that in turn are customers to upper-tier networks to be able to communicate worldwide. The hierarchical network structure is also complemented by a very large number of peering connections between networks of different types at Internet exchange points (IXPs) [4, 5]. Networks make peering agreement and exchange traffic based on commercial or other interests, irrespective of network size and tier structure.

In addition to traditional ISP networks, content delivery networks (CDNs) like Akamai and Limelight are well-established since a decade back, and today deliver a large share of the Internet content [6, 7, 8]. More recently, large content providers like Google and Netflix have started to build their own content delivery networks [9, 10, 11, 12, 13].

## 1.2 Traffic characteristics and access patterns

The traffic characteristics in a network depend on when and where on the Internet the traffic is measured. The traffic behaviour in a large backbone network differs from that in a small company network, and the traffic characteristics change with new applications, new types of networks and with changing user behaviour.

The Internet traffic volumes are constantly increasing but both the growth rate and the traffic mix very much depend on where on the Internet the measurements are done. Recent measurements of traffic volumes from large ISPs,



peering routers and Internet exchange points report annual growth rates of 35-100% [2, 5, 14, 15]. Figure 1.1 shows an example of traffic volumes at the Netnod Internet exchange point in Stockholm.

The Internet traffic over the last 15 years has been dominated by web traffic (transferred with the HTTP protocol) and peer-to-peer (P2P) traffic [2, 4, 14, 15, 17, 18, 19, 20, 21]. The share of the traffic volume that is P2P or HTTP traffic differs between different parts of the Internet and has changed over time. Fifteen years ago, measurements on the Internet backbone showed that 70-75% of the traffic was web traffic [22]. After that P2P file sharing applications became popular and contributed to a large share of the traffic volume [14, 15, 17, 20, 21], but many reports from the last couple of years show that HTTP traffic is again increasing. Measurements from large ISPs and peering routers [2, 14], show a decline in the share of P2P traffic and a growth in HTTP to more than 50% of the traffic. Measurements at a large European IXP [4] also show that HTTP accounts for more than 50% of the bytes, but the amount of HTTP traffic varies greatly between different participating AS:es.

Maier et al [18], monitoring 20000 residential DSL customers in 2009, report that HTTP and not P2P dominates the traffic with 57% of the transferred bytes, while other measurements of residential user traffic show that P2P is still dominant but not growing [15, 17].

A large part of the Internet traffic is delivery of video content in different ways: P2P file sharing, P2P streaming services, and much of the increase in web traffic is video that is transferred with HTTP, for instance from sites like Youtube. Video and TV-on-Demand streaming services like Netflix are also becoming increasingly popular. There are reports that Netflix alone represents more than 30% of peak downstream traffic in the US [11, 23].

## **1.3 Television and video over IP**

Television and video distribution over IP networks is an area with fast development. There are many terms that describe slightly different aspects of the area: IPTV, Internet television, web TV, TV-on-Demand, time-shifted TV, start-over TV, restart TV, catch-up TV, and so on. Some of these terms can also have different meanings in different contexts.

Internet television is a general term that here means TV programs that are available via the Internet. This includes TV services where traditional TV broadcasters (or others) make TV programs available for on-demand viewing. It also includes live broadcasts of individual programs or entire TV channels

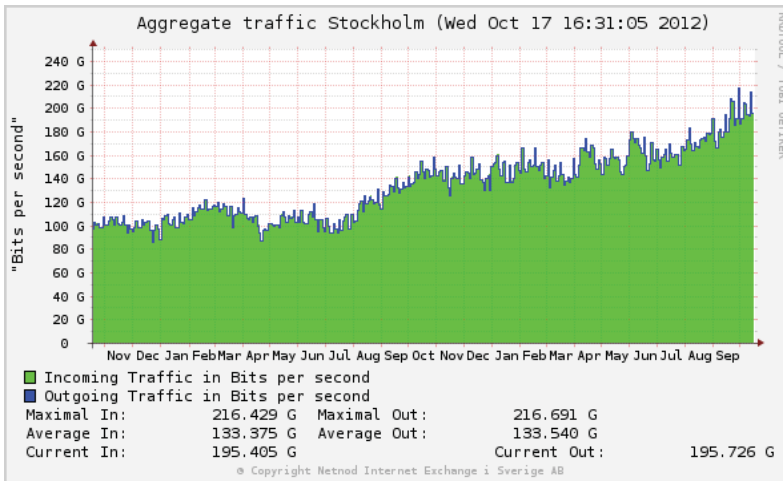
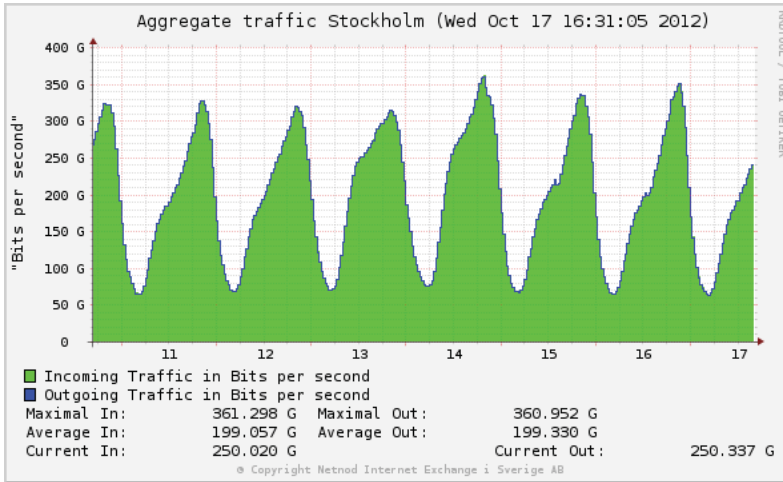


Figure 1.1: Example of Internet traffic at the Netnod Internet exchange point in Stockholm [16] (reprinted with permission). The top graph shows variation over a week (30 minute average) and the bottom graph shows how the traffic volume has increased over two years (one day average).

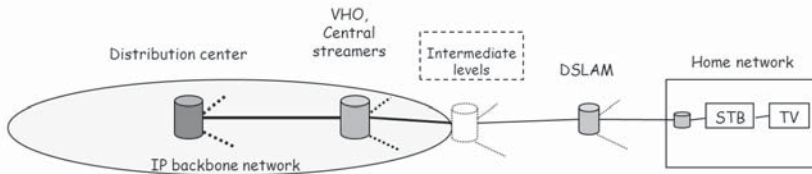


Figure 1.2: IPTV network architecture.

over the Internet.

From an Internet service provider perspective much of the TV and video (for instance from Youtube or Netflix) is so called over-the-top (OTT) content. This means that the operator just delivers the IP packets and does not control the TV and video services.

There are also operator managed services where TV is delivered over an IP network to subscribers. This is usually termed IPTV (Internet Protocol Television). The IPTV service includes traditional TV channels that usually are distributed using IP multicast. The operators often also introduce new on-demand services where viewers can control when to watch the programs. These services differ slightly depending on when the programs become available and for how long they are available. We here use the terms TV-on-Demand and time-shifted TV as general terms for programs that can be viewed decoupled from the traditional TV schedule. Start-over TV and restart TV more specifically means that the viewer can restart and choose to watch an ongoing broadcast program from the beginning. Catch-up TV usually means that programs become available for on-demand viewing some time after the broadcast. An IPTV service often have a mix of these features for different programs depending on agreements with content providers. It is also often combined with a traditional Video-on-Demand service with streaming of rental movies.

When distributing broadcast TV channels using IP multicast there is only one data stream per channel, while for TV-on-Demand there can be one stream per customer. Distributing dedicated TV streams to each viewer requires a lot of bandwidth and server capacity.

One branch of a typical IPTV architecture with a hierarchical tree-like network structure is illustrated in Figure 1.2. The TV content is delivered from content providers and comes into the network at a central distribution center from where it is transmitted to Video Hub Offices (VHO). A Video Hub Office

has storage and video streaming equipment to serve a district or a city. Under the VHO there can be intermediate levels of storage and video servers. Different operators try and use different structures of varying complexity. The figure also shows a TV subscriber with a home network where the TV and the set-top box (STB) is connected via a residential gateway to a Digital Subscriber Line Access Multiplexer (DSLAM). The TV channels are distributed using IP multicast from the distribution center to the set-top boxes. TV programs requested outside the schedule are streamed with unicast from the VHO (or from an intermediate server if available) to the set-top box.

## **1.4 Overload avoidance**

### **1.4.1 Traffic management**

Internet traffic management means handling the traffic situation in the networks; avoiding congestion and making good use of available network resources.

Traffic management involves both the end hosts and the network operators. It involves the end hosts in that they for many applications run TCP congestion control and adapt the send rate to what the network can handle. TCP increases the send rate to find out the available network capacity. When a packet is lost this is interpreted as network congestion and the transmission rate is decreased. From a network operator perspective traffic management involves monitoring and controlling the traffic behaviour in the network. It also includes traffic engineering where the routing of traffic through the network is adapted to the current traffic situation.

For network operators it is important to manage the traffic situation in the network and meet service level agreements (SLAs) made with their customers. The traffic demands in a network may fluctuate and change over time. Traffic engineering mechanisms can then be used to adapt to the changes in traffic demand and distribute traffic in order to benefit from available network resources.

The first step in the traffic engineering process is to collect the necessary information about network topology and the current traffic situation. Most traffic engineering methods need as input a traffic matrix describing the demand between each pair of nodes in the network. The traffic matrix is then used as input to the routing optimization.

Network operators today have different alternatives for coping with traffic variability: ranging from just over-dimensioning network capacity a lot, to

occasionally tuning the configuration of the routing protocols in order to postpone upgrades of network equipment, to more active use of traffic monitoring and traffic engineering mechanisms to manage the traffic situation.

One of the main alternatives for traffic engineering within an IP network [24, 25] is to use different methods for setting the link costs, and so decide upon the shortest paths, in the routing protocols OSPF (Open Shortest Path First) and IS-IS (Intermediate System to Intermediate System). These are both link-state protocols where the routing decisions are based on link costs and a shortest (least-cost) path calculation. With the equal-cost multi-path (ECMP) extension to the routing protocols the traffic can also be distributed over several paths that have the same cost. These routing protocols were designed to be simple and robust rather than to optimise the resource usage. They do not by themselves consider network utilisation and do not always make good use of network resources. The traffic is routed on the shortest path through the network even if the shortest path is overloaded and there exist alternative paths. It is up to the operator to find a configuration of the protocol, a set of link costs, that is best suited for the current traffic situation and that avoids congestion in the network.

There are also many other alternatives for how to do traffic engineering. For instance, Multi-Protocol Label Switching (MPLS) [26] has been widely used to control network traffic flows by setting up label-switched paths through the network. More recently, much focus has been on OpenFlow and Software-defined Networking (SDN) with the possibility of fine-grained, flow-based management and control, and the separation of control plane and data plane functionality [27, 28, 29, 30, 31, 32, 33].

### 1.4.2 Caching

One way to reduce the network load is to use caching, where copies of content are stored in local server nodes closer to the clients. By serving requests from the local cache instead of from a central server, repeated transfers of popular content over the network can be avoided.

Caching can be used to reduce network traffic and server load. It can also be used with other objectives: to lower access latency or to increase availability and robustness of a service.

Caching has been widely studied and used for web content [34, 35, 36, 37], for video and TV-on-demand [38, 39, 40, 41, 42] and for content distribution network [6, 7, 8, 9, 10, 11, 12]. Caching, integrated into the network architecture, is also a fundamental component in much of the long term research on future Internet architectures, like Information-centric networking [43, 44].

If we consider caching in a simple hierarchical system, as outlined in Figure 1.2 for IPTV, then a request from a client first goes to the cache, and if the program is not available there it is instead transferred from the central server. The system design parameters include: on what level in the network should the cache be placed, the size of the cache, and what caching policy to use.

The hit ratio, the share of requests that can be served by the cache, depends on the request pattern and on what content is placed in the cache. Given a limited cache size, and content that change in popularity over time, a strategy is needed to decide what should be put in the cache and what should be evicted. Many different cache replacement policies have been proposed in the literature [35, 37]. Two classic eviction policies are Least Recently Used (LRU) and Least Frequently Used (LFU). With the LRU strategy the program that has not been requested for the longest time is deleted from the cache. With LFU the program that is requested least often is discarded.

For the design of a caching system and for the choice of caching strategy, it is important to understand demand and access patterns.

## **1.5 Outline of thesis**

This thesis has two parts: an introductory part (Chapters 1 to 5) followed by a collection of five papers. Chapter 2 describes the research issues that this thesis deals with and the scientific contributions of the thesis. Chapter 3 summarizes the papers included in the thesis and their contributions. Chapter 4 discusses related work and put the research into context. In Chapter 5 there are conclusions and future work.

## **Chapter 2**

# **Research Issues and Scientific Contributions**

This thesis presents work on traffic engineering and on caching as means to avoid link overload in the network. For traffic engineering the purpose is to develop methods to control and steer the traffic. For caching the idea is to store popular content closer to the users to avoid repeated transfers of identical content. The work is done by simulation and by empirical studies and analysis of access patterns using logs from a real system.

### **2.1 Robust traffic engineering**

The objective of traffic engineering is to avoid congestion in the network and to make better use of available resources by adapting the routing to the current traffic situation. The main challenge for traffic engineering is to cope with the dynamics of traffic demands and topology. Traffic is often bursty and there can be unpredictable changes and shifts in traffic demand, for instance due to hotspots and flash crowds, or because a link goes down, there are changes in the inter-domain routing, or because traffic in an overlay is re-directed. For future networks more variability in traffic demands is also expected due to mobility of nodes and networks and more dynamic on-demand service level agreements.

The traffic variability means that, even if we could measure the current traffic situation exactly, it would not always correctly predict the near future traffic situation. Traffic engineering mechanisms need to be robust and able to

handle traffic variability and uncertainties in input traffic data.

### **2.1.1 Contributions**

The Papers A and B in this thesis cover different aspects of robust traffic engineering. We propose  $l$ -balanced routings as a way for an operator to handle traffic variability and uncertainties in input traffic data. An  $l$ -balanced solution routes the traffic on the shortest paths possible but makes sure that no link is utilised to more than a given level  $l$ . The contributions are an  $l$ -balanced routing algorithm based on multi-commodity flow optimisation and a heuristic search method for finding  $l$ -balanced weight settings for the legacy routing protocols OSPF and IS-IS.

$l$ -balanced routing gives the operator possibility to apply simple rules of thumb for controlling the maximum link utilisation and control the amount of spare capacity needed to handle sudden traffic variations. It gives more controlled traffic levels than other cost functions and more efficient routing for low traffic loads when there is no need to spread traffic over longer paths.

## **2.2 Understanding TV-on-Demand access patterns and their impact on caching**

Today video and TV distribution dominate Internet traffic. The increasing demand for high-bandwidth multimedia services put pressure on Internet service providers. It is therefore essential for traffic and cache management to understand TV program popularity and access patterns in real networks.

### **2.2.1 Contributions**

The Papers C, D and E in this thesis cover different aspects of TV-on-Demand access patterns and the potential for caching. In Papers C and D we simulate TV distribution with time-shift and investigate what impact TV program popularity, program set size, cache replacement policy and other factors have on the caching efficiency. The simulation results show that introducing a local cache close to the viewers significantly reduces the network load from TV on-demand services. By caching 4% of the program volume we can decrease the peak load during prime time by almost 50%. We also show that the TV program type and how program popularity changes over time can have a big influence on cache hit ratios and the resulting link loads.



For the models in Papers C and D we rely to a large extent on statistics from traditional scheduled TV. In Paper E we study access patterns in a real TV-on-Demand system over four months. We study user behaviour and program popularity and its impact on caching. We show how the popularity of TV-on-Demand programs changes over time. We see that the access pattern in a TV-on-Demand system very much depend on what type of content it offers. Furthermore, we find that the share of requests for the top most popular programs grows during prime time, and the change rate among them decreases. The cacheability is very high and the cache hit ratio increases during prime time when it is needed the most.



## Chapter 3

# Summary of the Papers and Their Contributions

This thesis is a collection of five papers. Papers A-B study different aspects of robust traffic engineering. Papers C-E investigate TV-on-Demand access patterns and the potential for caching. The papers are all published at refereed international conferences.

In Paper A we look at robust traffic engineering as an optimisation problem. In Paper B we build upon the work in Paper A by applying the ideas to the legacy routing protocols OSPF and IS-IS. We study search heuristics for finding weight-settings, and evaluate how different cost functions manage to handle faults in input traffic data due to traffic hotspots.

In Paper C we use an empirical IPTV workload model to simulate IPTV distribution with time-shift and investigate the benefit of introducing a local cache closer to the TV subscribers. In Paper D we extend the work by looking at how TV program popularity changes over time. For the simulations in Paper C and D we use TV schedules and statistics from linear broadcast TV. In Paper E we analyse logs from a large TV-on-Demand system over four months.

### 3.1 Paper A: A Multi Path Routing Algorithm for IP Networks Based on Flow Optimisation

Henrik Abrahamsson, Juan Alonso, Bengt Ahlgren, Anders Andersson and Per Kreuger. A Multi Path Routing Algorithm for IP Networks Based on Flow Optimisation. In *Proceedings of Third COST 263 International Workshop on Quality of Future Internet Services (QoFIS 2002)*, Zurich, Switzerland, October 2002.

**Summary:**

Intra-domain routing in the Internet normally uses a single shortest path to forward packets towards a specific destination with no knowledge of traffic demand. We present an intra-domain routing algorithm based on multi-commodity flow optimisation which enables load sensitive forwarding over multiple paths. It is neither constrained by weight-tuning of legacy routing protocols, such as OSPF, nor requires a totally new forwarding mechanism, such as MPLS. These characteristics are accomplished by aggregating the traffic flows destined for the same egress into one commodity in the optimisation and using a hash based forwarding mechanism. The aggregation also results in a reduction of computational complexity which makes the algorithm feasible for on-line load balancing. Another contribution is the optimisation objective function which allows precise tuning of the tradeoff between load balancing and total network efficiency.

**Contribution:**

There are two contributions in this paper: the modelling of the problem as an optimisation problem, and the definition of an optimisation objective function for  $l$ -balanced solutions. In the modelling of the optimisation problem we aggregate all traffic destined for a certain egress into one commodity in a multi-commodity flow optimisation. It is this definition of a commodity that both makes the computation tractable, and the forwarding simple.

$L$ -balanced solutions allows the network operator to choose a maximum desired link utilisation level. The optimisation will then find the most efficient solution, if it exists, satisfying the link level constraint. Our objective function thus enables the operator to control the trade-off between minimising the network utilisation and balancing load over multiple paths.

**My contribution:**

This is joint work with Bengt Ahlgren, Juan Alonso, Anders Gunnar and Per Kreuger. Juan Alonso did most of the mathematical work for this paper. In

discussion with Juan I contributed to the idea of only looking at the destination of the traffic when formulating the optimisation problem. I co-authored the paper.

### **3.2 Paper B: Robust Traffic Engineering using L-balanced Weight-Settings in OSPF/ISIS**

Henrik Abrahamsson and Mats Björkman. Robust Traffic Engineering using L-balanced Weight-Settings in OSPF/ISIS. In: *Sixth International Conference on Broadband Communications, Networks, and Systems (BROADNETS 2009)*, September 2009, Madrid, Spain.

**Summary:**

The focus of this work is on robust traffic engineering for the legacy routing protocols OSPF and IS-IS. The idea is to use the  $l$ -balanced solutions proposed in Paper A to make sure that there are enough spare capacity on all links to handle sudden hotspots and traffic shifts. Search heuristics are used to find the set of weights that avoid loading any link to more than  $l$  and the resulting routings are evaluated using real topologies and traffic scenarios.

**Contribution:**

The contributions are the idea of  $l$ -balanced weight-settings for robust traffic engineering, the search heuristics for finding such weight-settings, and the evaluation of how different cost functions (including  $l$ -balanced) manage to handle faults in input traffic data due to traffic hotspots.

**My contribution:**

The idea of using the  $l$ -balanced solution for robust weight-settings was mine. I implemented the search heuristics and did the evaluations and wrote most of the paper.

### **3.3 Paper C: Simulation of IPTV caching strategies**

Henrik Abrahamsson and Mats Björkman. Simulation of IPTV caching strategies. In: *International Symposium on Performance Evaluation of Computer*

and Telecommunication Systems (SPECTS'10), 11-14 July 2010, Ottawa, Canada

**Summary:**

In this paper we use an empirical IPTV workload model to simulate IPTV distribution with time-shift and investigate the benefit of introducing a local cache closer to the TV subscribers. The simulations are based on real TV schedules, and statistics about TV program popularity and viewer activity. We simulate a large number of TV viewers that, when active, request scheduled or on-demand programs and we investigate the resulting bandwidth requirements on the down link for different cache sizes and caching strategies.

**Contribution:**

The contributions of this paper are: We present an empirical IPTV workload model. We simulate a realistic scenario for IPTV distribution and compare the Least Recently Used (LRU) and Least Frequently Used (LFU) caching strategies. We show that time-shifted TV can be very capacity demanding and that considerable amounts of bandwidth can be saved by caching the most popular programs closer to the viewers.

**My contribution:**

I designed and implemented the simulator, did the evaluations and wrote most of the paper.

### 3.4 Paper D: Caching for IPTV distribution with time-shift

Henrik Abrahamsson and Mats Björkman. Caching for IPTV distribution with time-shift. In: *International conference on Computing, Networking & Communications (ICNC'13)*, 28-31 January 2013, San Diego, USA.

**Summary:**

In this paper we simulate TV distribution with time-shift and investigate what impact TV program popularity, program set size, cache replacement policy and other factors have on the caching efficiency. The simulation results show that introducing a local cache close to the viewers significantly reduces the network load from TV on-demand services. By caching 4% of the program volume we can decrease the peak load during prime time by almost 50%. We also show that the TV program type and how program popularity changes over time can

have a big influence on cache hit ratios and the resulting link loads.

**Contribution:**

In this paper we extend the work in Paper C by looking at how TV program popularity changes over time. Many programs such as news programs and weather forecasts quickly become outdated and lose their popularity when available on-demand. Other programs, typically drama TV-shows, retain interest from some viewers even a long time after their first release and initial peak in popularity. We show that the TV program type and how program popularity changes over time can have a big influence on cache hit ratio and the resulting link loads.

**My contribution:**

I did the analysis of program popularity, implemented the simulator, did the evaluations and wrote most of the paper.

### **3.5 Paper E: Program popularity and viewer behaviour in a large TV-on-Demand system**

Henrik Abrahamsson and Mattias Nordmark. Program popularity and viewer behaviour in a large TV-on-Demand system. In: *Internet Measurement Conference (IMC'12)*, 14-16 November 2012, Boston, USA.

**Summary:**

In this paper we analyse the access patterns in a large TV-on-Demand system and study the potential for caching. We characterize access patterns for different program categories, we show how program popularity changes over time and how this differs between different program types. We then use the request sequence in the data set for trace-driven simulation and study cache hit ratios for different cache sizes, cache replacement policies and population sizes.

**Contribution:**

Our contribution in this paper is three-fold. As a first-order result, we provide reconfirmation of known observations with an independent dataset. We demonstrate that there is a small set of programs that account for a large part of the requests. The program popularity conforms with the Pareto principle, or 80-20 rule. The demand follows a diurnal and weekly pattern, and there are large peaks in demand on Friday and Saturday evenings that need to be handled.

Second, we provide systematic evidence of TV-on-Demand access pattern characteristics that are intuitive yet unconfirmed in the literature. We show that news programs have a very short lifespan and are often only requested for a few hours, childrens programs are top ranked in the mornings and early evenings, and movie rentals are concentrated over weekends.

Finally, we also provide novel insights into access patterns that have not been reported previously to the best of our knowledge. We show how the popularity of TV-on-Demand programs changes over time. We see that the access pattern in a TV-on-Demand system very much depend on what type of content it offers. Furthermore, we find that the share of requests for the top most popular programs grows during prime time, and the change rate among them decreases. The cacheability is very high and the cache hit ratio increases during prime time when it is needed most.

**My contribution:**

I did the analysis with help from Mattias Nordmark. I did the simulations and I wrote the paper.



# Chapter 4

## Related Work

### 4.1 Traffic engineering in IP networks

Many different approaches for dynamic routing and traffic engineering have been proposed and used in telecommunication [45] and computer networks. For instance, the early ARPANET routing algorithms were based on measured link delay but had problems with traffic shifts and oscillations [46, 47].

The IETF Network Working Group presented a taxonomy of Internet traffic engineering methods in RFC3272 [48] in 2002. But for much of the traffic engineering research at that time the existing routing protocols were fixed. The challenge was to find configurations that adapted the routing to the current traffic situation. Traffic engineering by finding a suitable set of weights in OSPF/IS-IS is now a well studied area of research and it is described in textbooks in the area [25, 49]. When we in Paper B revisited the weight setting approach to traffic engineering we were most inspired by the pioneering works by Fortz and Thorup [50, 51] and Ramakrishnan and Rodrigues [52], in that we use a piece-wise linear cost function and search heuristics to find suitable weight settings.

Several studies [50, 53, 54, 55] have shown that even though we limit the routing of traffic to what can be achieved with weight-based ECMP shortest paths, and not necessarily the optimal weights but those found by search heuristics, it often comes close to the optimal routing for real network scenarios. How the traffic is distributed in the network very much depends on the objectives, usually expressed as a cost function, in the optimisation. An often proposed objective function is described by Fortz and Thorup [50]. Here the sum of the

cost over all links is considered and a piece-wise linear increasing cost function is applied to the flow on each link. The basic idea is that it should be cheap to use a link with small utilisation while using a link that approaches 100% utilisation should be heavily penalised. The  $l$ -balanced cost function used in Papers A and B is similar in that it uses a piecewise linear cost function to obtain desirable solutions. Additionally, it gives the operator the opportunity to set the maximum wanted link utilisation. Cost functions for traffic engineering is further investigated by Balon et al. [56]

Paper B added to existing work on weight settings by focusing on robustness and the objective of achieving a controlled spare capacity for handling unpredictable traffic shifts. For robust traffic engineering much of the focus has been on handling multiple traffic matrices and traffic scenarios [51, 57, 58, 59, 60, 61] and handling the trade-off between optimising for the common case or for the worst case. Nucci et al. [62] investigate link weight assignments that take into account SLA requirements and link failures. Xu et al. [63] describe a method to jointly solve the flow optimisation and the link-weight approximation using a single formulation resulting in a more efficient computation. Their method can also direct traffic over non-shortest paths with arbitrary percentages. Their results should also be directly applicable to our problem of providing robustness to changes, by just substituting their piece-wise linear cost function with our cost function. In a continuation on this work Xu et al. [64] propose a new link-state routing protocol. The protocol splits traffic over multiple paths with an exponential penalty on longer paths and achieves optimal traffic engineering while retaining the simplicity of hop-by-hop forwarding.

There are also several proposed traffic engineering protocols such as MATE [65], TeXCP [66] and REPLEX [67], that can balance traffic over several paths between ingress and egress nodes in the network, for instance by using MPLS [26]. Recently, much research focus has also been on OpenFlow and Software-defined Networking (SDN) with the possibility of fine-grained, flow-based management and control, and the separation of control plane and data plane functionality [27, 30, 31, 32, 33].

The advantage of optimising the weights in OSPF and IS-IS is of course easy deployment of the traffic engineering mechanism. However, the disadvantage is the difficulties and constraints imposed by using legacy routing. The general problem of finding the best way to route traffic through a network can be mathematically formulated as a multi-commodity flow (MCF) optimisation problem. In Paper A we present a routing algorithm based on multi-commodity flow optimisation. By aggregating the traffic flows destined for the same egress into one commodity in the optimisation we reduce the computational complex-

ity. The same approach was later used for instance by Sridharan et al. [68] and Fu et al. [69]. MCF optimisation is also used by many other research groups to address traffic engineering problems including [50, 70]. See also the book by Pioro and Medhi [49] and references therein.

## 4.2 Access patterns and potential for caching for TV and video on-demand

The recent growth and popularity of IPTV services have led to an increasing interest from researchers to measure and model IPTV viewing behavior. Cha et al. [71] present an extensive study of viewing behavior including channel popularity and channel switching in an operational IPTV network. Ramos et al. [72] present work on constructing an IPTV workload model capturing the way viewers change channels and watch live TV. Yu et al [73] study user activity and channel zapping in a municipal network. Qiu et al. model TV channel popularity [74] and user activities [75] in a large IPTV system and present the SimulWatch workload generator. These studies are similar to ours in that they model IPTV viewer behavior – but they study traditional live TV, and model channel popularity and not the popularity of individual programs. In Papers C and D we also simulate TV channels but our focus is on investigating time-shifted TV and the potential for caching. For this the popularity of individual programs is a fundamental part of the model. In this sense our work is closer to studies of traditional VoD systems.

Yu et al. [76] present a large measurement study of the Chinese PowerInfo Video-on-Demand system. This work is similar to ours in that they investigate many aspects of user behaviour and content access patterns. The PowerInfo system is a traditional VoD system. The videos in the library are old TV shows and movies and there are usually only a few new movies introduced to the system per day. This is different from the TV-on-Demand system that we study where there is a large inflow of new programs from the TV-schedule, time-shifted viewing, and programs with a very short life-span. Our work in Paper E is also different in other aspects in that we investigate how the access pattern depend on genre, we study cacheability and use trace-based simulation to investigate what impact the access patterns have on caching.

There are many other interesting studies of VoD systems and video popularity. Griwodz et al. [77] model long-term popularity of videos on the time scale of days based on VHS rental statistics. Lou et al. [78] give examples of the popularity evolution of video files from a Chinese television station. Tang

et al. [79] analyse and model many aspects of media server access. Avramova et al. [80] model the popularity evolution of TV-on-demand and video traces. Dan and Carlsson [81] measure and analyse BitTorrent content popularity. Guo et al. [82] study the probability distributions of Internet media workloads and analyse caching using a mathematical model. Yin et al. [83] study live VoD workloads from the 2008 Beijing Olympics. There are also many studies of Youtube and user generated videos [84, 85, 86, 87]. Szabo and Huberman [88] predict the long-term popularity of online content at Digg and Youtube based on early measurements of user accesses. Much research and many measurement studies have also focused on peer-assisted techniques for TV and VoD including [89, 90, 91, 92, 93, 94]. Ager et al. [95] study the cacheability for HTTP- and P2P-based applications.

Gopalakrishnan et al. [96] study user behaviour in a large IPTV system. This is similar to our work but their focus is on modeling the interactive user behaviour in an IPTV environment, including how users fast-forward, pause and rewind to control their viewing.

In Papers C and D we use an empirical IPTV workload model to simulate IPTV distribution and study caching. The simulations are based on real TV schedules, and statistics about TV program popularity and viewer activity. In Paper E we use trace-driven simulation, and utilize the sequence of requests in logs from a real TV-on-Demand system. There is also a lot of related work that use analytical models and simulations to study the performance of caching including [39, 40, 41, 42, 97, 98]. These studies have a more theoretical approach and is in this sense complementary to our work.

Seen in a broader perspective, a vast amount of research has been done on caching architectures, algorithms and protocols for instance for web, video and content distribution networks, as described in Section 1.4.2.

Another important issue for traffic and cache management is the interaction between traditional traffic engineering and content distribution in operator networks. What techniques and optimisations are possible here depend on the level of knowledge and control that the operator can have of the content distributed [99, 100, 101].

## Chapter 5

# Conclusions and Future Work

### 5.1 Conclusions

The Internet traffic volume continues to grow at a great rate, now pushed on by video and TV distribution in the networks. Increasing traffic volumes and the introduction of delay and loss sensitive services makes it crucial for operators to understand and manage the traffic situation in the network. More traffic also necessitate upgrades of network equipment and new investments for operators, and keep up-to-date the question of over-dimensioning network capacity versus using mechanisms for better handling the traffic.

This thesis deals with two approaches for avoiding network overload: traffic engineering and caching. We study traffic engineering mechanisms for adapting the routing to the current traffic situation and to steer traffic away from overloaded links. We study TV-on-Demand access patterns and the possible benefits of using caching mechanisms to avoid loading links with repeated transfers of popular content.

This thesis proposes  $l$ -balanced routings as a way for an operator to handle traffic variability and uncertainties in input traffic data. An  $l$ -balanced routing algorithm based on multi-commodity flow optimisation was presented in Paper A. A heuristic search method for finding  $l$ -balanced weight settings for the legacy routing protocols OSPF and IS-IS was presented in Paper B.  $L$ -balanced routing gives the operator possibility to apply simple rules of thumb for con-

trolling the maximum link utilisation and control the amount of spare capacity needed to handle sudden traffic variations. It gives more controlled traffic levels than other cost functions and more efficient routing for low traffic loads when there is no need to spread traffic over longer paths. The evaluation in Paper B shows that the search and the resulting weight settings work well in real network scenarios.

In Papers C-E we study TV-on-Demand access pattern and the potential for caching. We observe that there is a small set of programs that account for a large part of the requests. The program popularity conforms with the Pareto principle, or 80-20 rule. The demand follows a diurnal and weekly pattern, and there are large peaks in demand on Friday and Saturday evenings that need to be handled.

The popularity of rental movies, news, and TV shows changes over time in very different ways. News programs are often only requested for a few hours, movies are popular for months and increase in rank during weekends, TV shows increase in rank when the next episode is shown, and children's programs are top ranked in the mornings and early evenings. This means that programs jumps in and out of the top list of most popular programs. This can have implications for the choice of caching strategy. It is important to have the right programs in the cache in the evenings when the total demand is the highest. Another conclusion is that the access pattern in a TV-on-Demand system very much depend on what type of content it offers. We also observe that the request pattern for different episodes of the same show, and for programs within the same genre, often are very similar.

Another conclusion, from studying the cache friendliness of the TV-on-Demand workload, is that the potential for caching is high. The cacheability is very high, and in many scenarios the cache hit ratio with basic replacement policies is above 50% when caching 5% of the daily demand. We also observe that the hit ratio increases during prime time. The share of requests for the top most popular programs grows during prime time, and the change rate among them decreases.

## 5.2 Future work

It is an interesting time to work on issues related to television and video distribution over IP networks and the Internet. It is an area with fast development. Even with a view limited to Sweden in autumn 2012, as of this writing, a lot of things happen on many levels.

The way we are watching TV is slowly changing towards IP distributed television with more opportunities to choose what we want to watch, and when and where we watch TV. This thesis studies TV-on-Demand access patterns and the impact on caching. But the media consumption pattern is a moving target, it changes when the TV and video services evolve, and for future work there is a need to continuously study user behaviour and access patterns.

There is a trend towards start-over TV and TV-on-Demand where the viewer can choose to watch broadcast programs from the beginning or later after its scheduled time. Perhaps this is the beginning of a development in which the TV schedule becomes more a part of a recommendation system or a personalized playlist with a mix of live and pre-recorded content.

The devices are changing. More and more TV is watched on Internet connected smart TVs, and on smaller devices such as phones and tablets.

Another trend is that the TV and video market is changing. New players appear and compete with existing services. Many telecom and broadband operators have become TV distributors and offer new TV services in their own networks. HBO and Netflix, American providers of on-demand Internet streaming media, were launched in Sweden during the autumn 2012.

Traditional TV broadcasters are now also starting to distribute the scheduled TV via the web. It is also common with web exclusive content. One example of TV content that is often sent over the Internet today is sports events. SVT, the Swedish public service television company, showed 1600 hours from the London Olympics in two traditional broadcast channels and in six web channels [102]. A lot of the content was exclusively shown on the the web. But the change in viewing behaviour is still at an early stage and it takes time. The vast majority of viewing continues to be via traditional broadcast.

Television is a big thing. Although it is so commonplace that we might not think about it. When the way we are watching TV changes it can have a big impact on the distribution networks. In Sweden more than 70% of the population watch something on television on an average day, more than 40% of the people are watching TV during primetime, and individual TV shows can sometimes assemble 30-45% of the population [103]. If the TV viewing shifts from traditional broadcast to on-demand, personalized viewing on mobile devices, then it also gives rise to interesting future technical challenges.

The increasing demand for high-bandwidth streaming media services, both operator managed and OTT services, puts a big load on the networks. Caching seems to be a promising part of the solution. But there are many open issues for future work about caching, for instance what should be stored and on what level in the network should the caching be done.

In Papers C-E in this thesis we study many aspects of the access patterns in TV-on-Demand systems. We look at the cache friendliness of the workload in terms of cacheability and hit ratios for basic replacement policies. An immediate future work is to try to design and evaluate a caching strategy that is customized for the TV-on-Demand access patterns and investigate the extent to which it can reduce the network load.

When studying the cache friendliness of the request stream in Papers C-E we used the basic LRU and LFU cache replacement policies. With these the last requested program is always cached and the choice of what to evict from the cache is between the least recently and the least frequently requested program. A more advanced system could use more knowledge about access patterns and program popularity to decide what program to put in the cache and what program to evict.

One such strategy could be to keep track of all programs in the system, also those that are not currently in the cache. One could monitor the popularity by counting requests, let the programs age over time and for each program keep a value that describes the probability that it will be requested. There are a number of observations about the access patterns in this thesis that can be useful for such an informed caching strategy:

**Give preference to new programs** The broadcast of the traditional TV schedule has a marketing effect and with time-shifted TV ongoing scheduled programs immediately get a lot of requests. Some programs, like TV-news, also have a very short life-span. The value of a program should not have to be built up by requests over a long time.

**Categorize programs by genre to predict change in popularity over time**

We see in Papers D and E that the access pattern very much depends on the type of program. A news program that is top-ranked the first evening age quickly and has a very low probability for being requested the next evening. A rental movie however is popular for months and increase in rank during weekends. By categorizing programs by genre the probability for future requests can be predicted. The categorization of programs can also be more detailed. The request patterns for different episodes of the same show are often surprisingly similar. For a new episode of a show it is a reasonable assumption that the popularity of the program will change over time in a way similar to that of the previous episodes.

**Focus on prime time** The value of a program should reflect the probability that it will be requested during prime time. There are large peaks in



demand in the evenings and at the weekends that need to be handled. If caching is used to limit the maximum link load then it is essential to have the right programs in the cache on Friday and Saturday evenings. There are program like cartoons that are top-ranked in the mornings and early evenings that probably should never be in the cache.

The observations and the predictions outlined above can be used to optimise the caching performance. However, the basic monitoring of request frequency is still needed as a basis, and to handle unexpected changes and sudden peaks in program demand for instance due to large news events.

In Papers C-E we see that the cacheability and cache hit ratios for the TV-on-Demand workload are high even for small populations. Introducing a comparatively small local cache could significantly reduce the peak link loads. But for operators the monetary cost (both the capital expenditures and operational costs) of introducing memory into the network versus providing the bandwidth needed is essential. This is an important aspect to consider in future work.



# Bibliography

- [1] Measuring the Information Society 2012. International Telecommunication Union, report, October 2012.
- [2] Craig Labovitz, Scott Iekel-Johnson, Danny McPherson, Jon Oberheide, and Farnam Jahanian. Internet Inter-Domain Traffic. In *Proceedings of the ACM SIGCOMM'10*, New Delhi, India, 2010.
- [3] Wikipedia. Tier 1 network. [http://en.wikipedia.org/wiki/Tier\\_1\\_network](http://en.wikipedia.org/wiki/Tier_1_network), October 2012.
- [4] Bernhard Ager, Nikolaos Chatzis, Anja Feldmann, Nadi Sarrar, Steve Uhlig, and Walter Willinger. Anatomy of a Large European IXP. In *Proceedings of ACM SIGCOMM'12*, Helsinki, Finland, 2012.
- [5] Juan Camilo Cardona Restrepo and Rade Stanojevic. A History of an Internet eXchange Point. *ACM SIGCOMM Computer Communication Review*, 42(2), April 2012.
- [6] Erik Nygren, Ramesh K. Sitaraman, and Jennifer Sun. The Akamai Network: A Platform for High-Performance Internet Applications. *ACM SIGOPS Operating Systems Review*, 44(3), July 2010.
- [7] Tom Leighton. Improving Performance of the Internet. *Communications of the ACM*, 52(2), February 2009.
- [8] Ao-Jan Su, David R. Choffnes, Aleksandar Kuzmanovic, and Fabián E. Bustamante. Drafting behind akamai (travelocity-based detouring). In *Proceedings of SIGCOMM '06*, Pisa, Italy, 2006.

- [9] Ruben Torres, Alessandro Finamore, Jin Ryong Kim, Marco Mellia, Maurizio M. Munaf, and Sanjay G. Rao. Dissecting Video Server Selection Strategies in the YouTube CDN. In *Proceedings of 31st International Conference on Distributed Computing Systems (ICDCS'11)*, June 2011.
- [10] Vijay Kumar Adhikari, Sourabh Jain, Yingying Chen, and Zhi-Li Zhang. Vivisecting YouTube: An Active Measurement Study. In *Proceedings of IEEE INFOCOM'12, Mini-conference*, Orlando, USA, March 2012.
- [11] Vijay Kumar Adhikari, Yang Guo, Fang Hao, Matteo Varvello, Volker Hilt, Moritz Steiner, and Zhi-Li Zhang. Unreeling Netflix: Understanding and Improving Multi-CDN Movie Delivery. In *Proceedings of IEEE INFOCOM'12*, Orlando, USA, March 2012.
- [12] Announcing the Netflix Open Connect Network. Netflix blog, <http://blog.netflix.com/2012/06/announcing-netflix-open-connect-network.html>, June 2012.
- [13] Google and Netflix Make Land Grab On Edge Of Internet. Wired, <http://www.wired.com/wiredenterprise/2012/06/cdn/>, June 2012.
- [14] Alexandre Gerber and Robert Doverspike. Traffic Types and Growth in Backbone Networks. In *Proceedings of OFC/NFOEC*, March 2011.
- [15] Kenjiro Cho, Kensuke Fukuda, Hiroshi Esaki, and Akira Kato. Observing Slow Crustal Movement in Residential User Traffic. In *Proceedings ACM CoNEXT'08*, Madrid, Spain, December 2009.
- [16] Netnod. On–line: <http://www.netnod.se/ix/statistics>.
- [17] Jie Li, Andreas Aurelius, Viktor Nordell, Manxing Du, Åke Arvidsson, and Maria Kihl. A five year perspective of traffic pattern evolution in a residential broadband access network. In *Proceedings of Future Network & Mobile Summit 2012*, Berlin, Germany, July 2012.
- [18] Gregor Maier, Anja Feldmann, Vern Paxson, and Mark Allman. On Dominant Characteristics of Residential Broadband Internet Traffic. In *Proceedings of the 9th ACM SIGCOMM Internet measurement conference (IMC'09)*, Chicago, Illinois, USA, 2009.

- [19] Pierre Borgnat, Guillaume Dewaele, Kensuke Fukuda, Patrice Abry, and Kenjiro Cho. Seven Years and One Day: Sketching the Evolution of Internet Traffic. In *Proceedings on INFOCOM'09*, 2009.
- [20] Wolfgang John, Sven Tafvelin, and Tomas Olovsson. Trends and differences in connection-behavior within classes of internet backbone traffic. In *Proceedings of the 9th international conference on Passive and active network measurement (PAM'08)*, Cleveland, USA.
- [21] Subhabrata Sen and Jia Wang. Analyzing peer-to-peer traffic across large networks. *IEEE/ACM Transactions on Networking*, 12(2):219–232, April 2004.
- [22] G. J. Miller K. Thompson and R. Wilder. Wide-area traffic patterns and characteristics (extended version). *IEEE Network*, 11(6):10–23, 1997.
- [23] Sandvine. Global Internet Phenomena Report, 1H 2012. [http://www.sandvine.com/news/global\\_broadband\\_trends.asp](http://www.sandvine.com/news/global_broadband_trends.asp), 2012.
- [24] Bernard Fortz, Jennifer Rexford, and Mikkel Thorup. Traffic engineering with traditional IP routing protocols. *IEEE Communication Magazine*, 40, October 2002.
- [25] J. Rexford. Route optimization in IP networks. In Mauricio G.C. Resende and Panos M. Pardalos, editors, *Handbook of Optimization in Telecommunications*. Springer Science+Business Media, 2006.
- [26] Multiprotocol label switching working group . <http://datatracker.ietf.org/wg/mpls/charter/>.
- [27] Nick McKeown, Tom Anderson, Hari Balakrishnan, Guru Parulkar, Larry Peterson, Jennifer Rexford, Scott Shenker, and Jonathan Turner. OpenFlow: Enabling Innovation in Campus Networks. *SIGCOMM Computer Communication Review*, 38(2):69–74, April 2008.
- [28] Openflow @ google. Keynote by Urs Hoelzle, Open Networking Summit 2012, <http://www.opennetsummit.org>.
- [29] Going With the Flow: Googles Secret Switch to the Next Wave of Networking. *Wired*, April 2012. <http://www.wired.com/wiredenterprise/2012/04/going-with-the-flow-google/>.

- [30] Natasha Gude, Teemu Koponen, Justin Pettit, Ben Pfaff, Martín Casado, Nick McKeown, and Scott Shenker. NOX: Towards an Operating System for Networks. *SIGCOMM Computer Communication Review*, 38(3):105–110, July 2008.
- [31] Teemu Koponen, Martin Casado, Natasha Gude, Jeremy Stribling, Leon Poutievski, Min Zhu, Rajiv Ramanathan, Yuichiro Iwata, Hiroaki Inoue, Takayuki Hama, and Scott Shenker. Onix: A Distributed Control Platform for Large-scale Production Networks. In *Proceedings of the 9th USENIX conference on Operating systems design and implementation (OSDI'10)*, Vancouver, Canada, 2010.
- [32] Minlan Yu, Jennifer Rexford, Michael J. Freedman, and Jia Wang. Scalable Flow-Based Networking with DIFANE. In *Proceedings of the ACM SIGCOMM'10*, New Delhi, India, 2010.
- [33] Andrew R. Curtis, Jeffrey C. Mogul, Jean Tourrilhes, Praveen Yalagandula, Puneet Sharma, and Sujata Banerjee. DevoFlow: Scaling Flow Management for High-Performance Networks. In *Proceedings of the ACM SIGCOMM'11*, Toronto, Canada, 2011.
- [34] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker. Web Caching and Zipf-like Distributions: Evidence and Implications. In *Proceedings of IEEE INFOCOM*, 1999.
- [35] Jia Wang. A Survey of Web Caching Schemes for the Internet. *ACM SIGCOMM Computer Communication Review*, 29:36–46, 1999.
- [36] Alec Wolman, Geoffrey M. Voelker, Nitin Sharma, Neal Cardwell, Anna Karlin, and Henry M. Levy. On the scale and performance of cooperative Web proxy caching. In *Proceedings of the 17th ACM Symposium on Operating Systems Principles (SOSP '99)*, 1999.
- [37] Abdullah Balamash and Marwan Krunz. An Overview of Web Caching Replacement Algorithms. *IEEE Communication surveys*, 6:44–56, 2004.
- [38] Jiangchuan Liu and Jianliang Xu. Proxy Caching for Media Streaming Over the Internet. *IEEE Communications Magazine*, 42:88–94, 2004.
- [39] Sem Borst, Varun Gupta, and Anwar Walid. Distributed caching algorithms for content distribution networks. In *Proceedings of INFOCOM'10*, San Diego, USA, 2010.

- [40] D. De Vleeschauwer, Z. Avramova, S. Wittevrongel, and H. Brueel. Transport Capacity for a Catch-up Television Service. In *Proceedings of EuroITV'09*, pages 161–170, Leuven, Belgium, June 2009.
- [41] Danny De Vleeschauwer and Koenraad Laevens. Performance of caching algorithms for IPTV on-demand services. *IEEE Transactions on broadcasting*, 55:491 – 501, 2009.
- [42] T. Wauters, W. Van de Meerssche, F. De Turck, B. Dhoedt, P. Demeester, T. Van Caenegem, and E. Six. Co-operative Proxy Caching Algorithms for Time-Shifted IPTV Services. In *Proceedings of 32nd EUROMI-CRO Conference on Software Engineering and Advanced Applications (SEAA)*, pages 379–386, Dubrovnik, Croatia, September 2006.
- [43] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher, and B. Ohlman. A Survey of Information-Centric Networking. *IEEE Communications Magazine*, 50(7), July 2012.
- [44] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard. Networking named content. In *Proceedings of the 5th international conference on Emerging networking experiments and technologies (CoNEXT'09)*, Rome, Italy, December 2009.
- [45] Gerald Ash. *Dynamic Routing in Telecommunications Networks*. McGraw-Hill, 1997.
- [46] John M. McQuillan, Ira Richer, and Eric C. Rosen. The New Routing Algorithm for the ARPANET. *IEEE Transactions on communications*, 28(5), May 1980.
- [47] A. Khanna and J. Zinky. The revised ARPANET routing metric. In *Symposium proceedings on Communications architectures & protocols, SIGCOMM '89*, 1989.
- [48] D. Awduche, A. Chiu, A. Elwalid, I. Widjaja, and X. Xiao. Overview and principles of Internet traffic engineering. Internet RFC 3272, May 2002.
- [49] M. Pióro and D. Medhi. *Routing, Flow, and Capacity Design in Communication and Computer Networks*. Morgan Kaufmann, 2004.

- [50] B. Fortz and M. Thorup. Internet traffic engineering by optimizing OSPF weights. In *Proceedings IEEE INFOCOM 2000*, pages 519–528, Israel, March 2000.
- [51] B. Fortz and M. Thorup. Optimizing OSPF/IS-IS weights in a changing world. *IEEE Journal on Selected Areas in Communications*, 20(4):756–767, May 2002.
- [52] K.G. Ramakrishnan and M.A. Rodrigues. Optimal routing in shortest path data networks. *Lucent Bell Labs Technical Journal*, 6(1), 2001.
- [53] A. Gunnar, H. Abrahamsson, and M. Söderqvist. Performance of Traffic Engineering in Operational IP-Networks: An Experimental Study. In *Proceedings of 5th IEEE International Workshop on IP Operations and Management*, Barcelona, Spain, October 2005.
- [54] A. Sridharan, R. Guérin, and C. Diot. Achieving Near-Optimal Traffic Engineering Solutions for Current OSPF/IS-IS Networks. In *IEEE Infocom*, San Francisco, March 2003.
- [55] David Applegate and Edith Cohen. Making Intra-Domain Routing Robust to Changing and Uncertain Traffic Demands: Understanding Fundamental Tradeoffs. In *Proceedings of ACM SIGCOMM*, Karlsruhe, Germany, August 2003.
- [56] S. Balon, F. Skivee, and G. Leduc. How Well Do Traffic Engineering Objective Functions Meet TE Requirements? In *Proceedings of IFIP International Networking Conference*, Coimbra, Portugal, May 2006.
- [57] C. Zhang, Z. Ge, J. Kurose, Y. Liu, and D. Townsley. Optimal Routing with Multiple Traffic Matrices: Tradeoffs between Average Case and Worst Case Performance. In *Proceedings of ICNP 2005*, Boston, USA, November 2005.
- [58] C. Zhang, Y. Liu, W. Gong, J. Kurose, R. Moll, and D. Townsley. On Optimal Routing with Multiple Traffic Matrices. In *Proceedings of Infocom 2005*, Miami, USA, March 2005.
- [59] H. Wang, H. Xie, L. Qiu, Y. R. Yang, Y. Zhang, and A. Greenberg. Cope: Traffic engineering in dynamic networks. In *Proceedings of ACM SIGCOMM*, Pisa, Italy, September 2006.



- [60] Anders Gunnar and Mikael Johansson. Robust routing under bgp reroutes. In *Proceedings of Globecom 2007*, Washington, DC, USA, 2007.
- [61] Anders Gunnar. *Aspects of proactive traffic engineering in IP networks*. PhD thesis, KTH Royal Institute of Technology, 2011.
- [62] Antonio Nucci, Supratik Bhattacharyya, Nina Taft, and Christophe Diot. IGP Link Weight Assignment for Operational Tier-1 Backbones. *IEEE/ACM Transactions on Networking*, 15(4), August 2007.
- [63] Dahai Xu, Mung Chiang, and Jennifer Rexford. DEFT: Distributed exponentially-weighted flow splitting. In *IEEE Infocom*, Anchorage, Alaska, USA, May 6–12, 2007.
- [64] Dahai Xu, Mung Chiang, and Jennifer Rexford. Link-state routing with hop-by-hop forwarding can achieve optimal traffic engineering. *IEEE/ACM Transactions on networking*, 19(6), December 2011.
- [65] A. Elwalid, C. Jin, S. Low, and I. Widjaja. MATE: MPLS Adaptive Traffic Engineering. In *Proceedings of IEEE INFOCOM*, April 2001.
- [66] Srikanth Kandula, Dina Katabi, Bruce Davie, and Anna Charny. Walking the tightrope: responsive yet stable traffic engineering. In *Proceedings of SIGCOMM '05*, Philadelphia, Pennsylvania, USA, 2005.
- [67] Simon Fischer, Nils Kammenhuber, and Anja Feldmann. Replex: dynamic traffic engineering based on wardrop routing policies. In *Proceedings of the 2006 ACM CoNEXT conference*, CoNEXT '06, 2006.
- [68] A. Sridharan, R. Guérin, and C. Diot. Achieving Near-Optimal Traffic Engineering Solutions for Current OSPF/IS-IS Networks. *IEEE/ACM Transactions on Networking*, 13, April 2005.
- [69] Jing Fu, Peter Sjödin, and Gunnar Karlsson. Traffic engineering and routing in IP networks with centralized control. May 2008. IFIP Networking 2008.
- [70] D. Mitra and K. G. Ramakrishnan. A Case Study of Multiservice, Multipriority Traffic Engineering Design for Data Networks. In *Proceedings of Globecom'99*, Brazil, 1999.

- [71] Meeyoung Cha, Pablo Rodriguez, Jon Crowcroft, Sue Moon, and Xavier Amatriain. Watching Television Over an IP Network. In *Proceedings of Internet Measurement Conference(IMC)*, pages 71–84, Greece, October 2008.
- [72] Fernando M.V. Ramos, Fei Song, Pablo Rodriguez, Richard Gibbens, Jon Crowcroft, and Ian H. White. Constructing an IPTV Workload Model. In *Proceedings of SIGCOMM, Poster session*, Barcelona, Spain, August 2009.
- [73] Geng Yu, Tord Westholm, Maria Kihl, Iñigo Sedano, Andreas Aurelius, Christina Lagerstedt, and Per Ödling. Analysis and Characterization of IPTV user behavior. In *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB'09)*, 2009.
- [74] Tongqing Qiu, Zihui Ge, Seungjoon Lee, Jia Wang, Qi Zhao, and Jun Xu. Modeling Channel Popularity Dynamics in a Large IPTV System. In *Proceedings of SIGMETRICS*, pages 275–286, Seattle, USA, June 2009.
- [75] Tongqing Qiu, Zihui Ge, Seungjoon Lee, Jia Wang, Jun Xu, and Qi Zhao. Modeling User Activities in a Large IPTV System. In *Proceedings of Internet Measurement Conference*, pages 430–441, Chicago, USA, November 2009.
- [76] H. Yu, D. Zheng, B. Zhao, and W. Zheng. Understanding User Behavior in Large-Scale Video-on-Demand Systems. In *Proceedings of EuroSys2006*, pages 333–344, Leuven, Belgium, 2006.
- [77] C. Griwodz, Michael Bär, and L. Wolf. Long-term Movie Popularity Models in Video-on-Demand Systems or The Life of an on-Demand Movie. In *Proceedings of the fifth ACM international conference on Multimedia*, pages 349 – 357, Seattle, USA, 1997.
- [78] J. Lou, Y. Tang, M. Zhang, and S. Yang. Characterizing User Behavior Model to Evaluate Hard Cache in Peer-to-Peer Based Video-on-demand Service. In *Proceedings of MMM'07*, pages 125–134, 2007.
- [79] W. Tang, Y. Fu, L. Cherkasova, and A. Vahdat. Modeling and Generating Realistic Streaming Media Server Workloads. *Computer Networks*, 51:336–356, 2007.

- [80] Z. Avramova, S. Wittevrongel, H. Bruneel, and D. Vleeschauwer. Analysis and Modeling of Video Popularity Evolution in Various Online Video Content Systems: Power Law versus Exponential Decay. In *Proceedings of International Conference on Evolving Internet*, 2009.
- [81] György Dan and Niklas Carlsson. Power-law Revisited: A Large Scale Measurement Study of P2P Content Popularity. In *Proceedings of the 9th international conference on Peer-to-peer systems (IPTPS'10)*, San Jose, USA, April 2010.
- [82] Lei Guo, Enhua Tan, Songqing Chen, Zhen Xiao, and Xiaodong Zhang. The stretched exponential distribution of internet media access patterns. In *Proceedings of the twenty-seventh ACM symposium on Principles of distributed computing (PODC'08)*, New York, USA, 2008.
- [83] H. Yin, X. Liu, F. Qiu, N. Xia, C. Lin, H. Zhang, V. Sekar, and G. Min. Inside the Bird's Nest: Measurements of Large-Scale Live VoD from the 2008 Olympics. In *Proceedings of Internet Measurement Conference (IMC'09)*, USA, 2009.
- [84] Y. Borghol, S. Mitra, S. Ardon, N. Carlsson, D. Eager, and A. Mahanti. Characterizing and Modeling Popularity of User-generated Videos. In *Proceedings of IFIP International Symposium on Computer Performance, Modeling, Measurements and Evaluation (PERFORMANCE)*, 2011.
- [85] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon. Analyzing the Video Popularity Characteristics of Large-Scale User Generated Content Systems. *IEEE Transactions on Networking*, 17:1357–1370, October 2009.
- [86] Phillipa Gill, Martin Arlitt, Zongpeng Li, and Anirban Mahanti. YouTube Traffic Characterization: A view From the Edge. In *Proceedings of ACM SIGCOMM Internet Measurement Conference*, San Diego, USA, October 2007.
- [87] X. Kang, H. Zhang, G. Jiang, H. Chen, X. Meng, and K. Yoshihira. Measurement, Modeling, and Analysis of Internet Video Sharing Site Workload: A Case Study. In *Proceedings of IEEE Conference on Web Services*, 2008.

- [88] G. Szabo and B. Huberman. Predicting the Popularity of Online Content. *Communications of the ACM*, 53, August 2010.
- [89] C. Huang, J. Li, and K.W. Ross. Can Internet VoD be Profitable? In *Proceedings of ACM Sigcomm 2007*, pages 133–144, Kyoto, Japan, 2007.
- [90] Y. Huang, T. Fu, D. Chiu, J. Lui, and C. Huang. Challenges, Design and Analysis of a Large-scale P2P-VoD System. In *Proceedings of ACM SIGCOMM 2008*, pages 375–388, Seattle, USA, August 2008.
- [91] M. Cha, P. Rodriguez, S. Moon, and J. Crowcraft. On Next-Generation Telco-Managed P2P TV Architectures. In *Proceedings of International workshop on Peer-To-Peer Systems (IPTPS)*, 2008.
- [92] X. Hei, C. Liang, J. Liang, Y. Liu, and K.W. Ross. A measurement study of a large-scale P2P IPTV system. *IEEE Transactions on Multimedia*, 9:1672–1687, December 2007.
- [93] T. Silverstone and O. Fourmaux. Measuring P2P IPTV Systems. In *Proceedings of NOSSDAV'07*, Illinois, USA, 2007.
- [94] Thomas Silverston, Olivier Fourmaux, Alessio Botta, Alberto Dainotti, Antonio Pescapé, Giorgio Ventre, and Kavé Salamatian. Traffic analysis of peer-to-peer IPTV communities. *Computer Networks*, 53(4):470–484, March 2009.
- [95] Bernhard Ager, Fabian Schneider, Juhoon Kim, and Anja Feldmann. Revisiting Cacheability in Times of User Generated Content. In *Proceedings of 13th IEEE Global Internet Symposium*, San Diego, CA, USA, March 2010.
- [96] V. Gopalakrishnan, R. Jana, K. Ramakrishnan, D. Swayne, and V. Vaishampayan. Understanding Couch Potatoes: Measurement and Modeling of Interactive Usage of IPTV at large scale. In *Proceedings of Internet Measurement Conference (IMC'11)*, 2011.
- [97] Massimo Gallo, Bruno Kauffmann, Luca Muscariello, Alain Simonian, and Christian Tanguy. Performance Evaluation of the Random Replacement Policy for Networks of Caches. In *Proceedings of ACM SIGMETRICS'12*, London, England, UK, 2012.

- [98] John Ardelius, Björn Grönvall, Lars Westberg, and Åke Arvidsson. On the Effects of Caching in Access Aggregation Networks. In *Proceedings of the second edition of the ICN workshop on Information-centric networking*, Helsinki, Finland, 2012.
- [99] Wenjie Jiang, Rui Zhang-Shen, Jennifer Rexford, and Mung Chiang. Cooperative Content Distribution and Traffic Engineering in an ISP Network. In *Proceedings of ACM SIGMETRICS'09*, Seattle, USA, June 2009.
- [100] Ingmar Poese, Benjamin Frank, Bernhard Ager, Georgios Smaragdakis, and Anja Feldmann. Improving Content Delivery Using Provider-aided Distance Information. In *Proceedings of Internet Measurement Conference (IMC'10)*, Melbourne, Australia, November 2010.
- [101] Ingmar Poese, Benjamin Frank, Georgios Smaragdakis, Steve Uhlig, Anja Feldmann, and Bruce Maggs. Enabling Content-aware Traffic Engineering. *ACM SIGCOMM Computer Communication Review*, 42(5):21–28, October 2012.
- [102] Datorn ger OS-tittandet ett lyft. *Dagens Nyheter*, 24 July 2012. <http://www.dn.se/sport/datorn-ger-os-tittandet-ett-lyft>.
- [103] Mediamätning i Skandinavien (MMS). On-line: <http://www.mms.se>.