

DISS. ETH. NO. 29768

**MODEL-BASED INSIGHTS
ON THE FUTURE INTERNET**

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES
(Dr. sc. ETH Zurich)

presented by
SIMON SCHERRER
MSc ETH in Computer Science, ETH Zurich
born on 05.02.1994

accepted on the recommendation of
Prof. Dr. Adrian Perrig (doctoral thesis supervisor)
Prof. Dr. Laurent Vanbever (co-examiner)
Prof. Dr. Stefan Schmid (co-examiner)
Prof. Dr. Michael Schapira (co-examiner)

All theory depends on assumptions which are not quite true.
That is what makes it theory.
The art of successful theorizing is to make the inevitable simplifying assumptions
in such a way that the final results are not very sensitive.

– ROBERT M. SOLOW [239]

ABSTRACT

In the Internet, billions of endpoints exchange data, and thereby create a global traffic pattern, characterized by a time-variant distribution of traffic volume over inter-domain paths. This global traffic pattern is increasingly reshaped by two mechanisms for data-traffic control that have experienced real-world deployment in recent years. First, end-host path selection allows end-hosts to allocate traffic on inter-domain paths that are under-utilized or particularly suitable for a given application. Second, modern congestion-control algorithms (CCAs) promise to simultaneously achieve high utilization and low congestion, namely with a latency-sensitive approach that can withstand competition from legacy algorithms.

However, despite extensive research into end-host path selection and modern congestion control, the research community does not yet completely understand how these two control mechanisms will affect the Internet traffic pattern under global deployment. To further such an understanding, this thesis follows a model-based approach, i.e., it describes end-host path selection and congestion-control algorithms in a mathematical fashion, and leverages these descriptions to investigate the effects of the emerging mechanisms. Thanks to such formal representation, the model-based analysis in this thesis can rely on diverse mathematical tools, including game theory and control theory, and therefore provide provable guarantees.

The model-based analysis in this thesis is separated into two parts, both representing aspects of the intricate dynamics in Internet traffic:

Part I: Network performance. In the first part of this thesis, I find that load-adaptive path selection by selfish end-hosts admits near-optimal equilibria, but may not converge to these equilibrium distributions. Instead, this path selection may cause oscillation, which demonstrably hampers network performance, but which can be suppressed by network operators. Moreover, I present a highly accurate fluid model of the BBR congestion-control algorithm, and prove that BBR is stable in homogeneous settings, but unstable in competition with loss-based CCAs.

Part II: ISP economics. In the second thesis part, I discover that end-host path selection alters the cooperation among ISPs, by allowing novel interconnection agreements that substantially boost path diversity, and the competition among ISPs, by motivating investments that improve transmission quality.

In summary, the model-based approach in this thesis yields numerous complex insights, and thus contributes towards understanding the future Internet.

ZUSAMMENFASSUNG

Der Datenaustausch durch Milliarden von End-Hosts im Internet führt dazu, dass sich der globale Datenverkehr ständig neu über über domänenübergreifende Pfade verteilt. Dieses globale Datenverkehrsmuster steht zunehmend unter dem Einfluss von zwei Mechanismen zur Datenverkehrskontrolle, welche in den letzten Jahren praktisch verfügbar geworden sind. Der erste Mechanismus, die Pfadwahl durch End-Hosts, erlaubt es, Datenverkehr auf Pfade zu verschieben, welche schwach ausgelastet oder für eine bestimmte Anwendung besonders gut geeignet sind. Der zweite Mechanismus, namentlich moderne Algorithmen zur Datenstaukontrolle, verfolgen das Ziel hoher Kapazitätsauslastung bei geringer Überlastung, indem sie das Sendetempo auf Basis von staubedingten Übertragungsverzögerungen festlegen und es im Wettbewerb mit älteren Algorithmen nur bedingt reduzieren.

Trotz ausgedehnter Forschung über End-Host-Pfadwahl und moderne Datenstaukontrolle fehlt allerdings immer noch ein genaues Verständnis davon, wie sich diese zwei Kontrollmechanismen auf das globale Datenverkehrsmuster auswirken werden, sobald sie im gesamten Internet verwendet werden. Um dieses Verständnis zu verbessern, verwendet die vorliegende Dissertation einen modellbasierten Ansatz, welcher die Pfadwahl durch End-Hosts und die modernen Algorithmen zur Datenstaukontrolle auf eine mathematische Weise beschreibt. Diese mathematischen Beschreibungen ermöglichen es, die Effekte der erwähnten Mechanismen theoretisch zu untersuchen. In einer solchen Untersuchung können modellbasierte Analysen auf mathematische Methoden zurückgreifen, so etwa auf die Spieltheorie und die Kontrolltheorie, und beweisbare Garantien abgeben.

Die modellbasierten Analysen in der vorliegenden Dissertation sind in zwei Teile aufgegliedert, welche unterschiedlichen Aspekten der vielschichtigen Dynamik des Internetverkehrs entsprechen:

Teil I: Netzwerkleistung. Im ersten Teil der Dissertation komme ich zum Ergebnis, dass eine auslastungsorientierte Pfadwahl durch eigennützige End-Hosts zwar quasi-optimale spieltheoretische Gleichgewichte zulässt, aber nicht zwingend diesen Gleichgewichtsverteilungen entgegenstrebt. Die Oszillation, die stattdessen einsetzt, wirkt sich nachweislich negativ auf die Netzwerkleistung aus, kann allerdings von Netzwerkbetreibern unterdrückt werden. Darüber hinaus präsentiere ich ein äußerst genaues Fluidmodell des BBR-Algorithmus zur Datenstaukontrolle. Basierend auf diesem Modell lässt sich beweisen, dass BBR in einem homoge-

nen Umfeld stabil ist, im Wettbewerb mit anderen, verlustbasierten Algorithmen allerdings ein instabiles Verhalten zeigt.

Teil II: ISP-Ökonomie. Im zweiten Teil der Dissertation weise ich nach, dass die Pfadwahl durch End-Hosts die Zusammenarbeit zwischen ISPs beeinflusst, nämlich indem neue Vereinbarungen zur Vernetzung ermöglicht werden, welche die Pfaddiversität im Internet erheblich verbessern. Ebenso beeinflusst die Pfadwahl durch End-Hosts auch den Wettbewerb zwischen ISPs: Genauer gesagt verstärkt die Pfadwahl die Anreize, die Übertragungsqualität im Internet durch Investitionen zu verbessern.

Insgesamt gestattet der modellbasierte Ansatz dieser Dissertation eine Vielzahl komplexer Einsichten und trägt damit in wertvoller Weise zum Verständnis des zukünftigen Internets bei.

PUBLICATIONS

In my doctorate, I coordinated and contributed to various research projects, resulting in the accepted publications and conference submissions presented here.

The following articles (in chronological order) were written under my leadership, and serve as the basis of the dissertation at hand:

- [1] **The Value of Information in Selfish Routing**
Simon Scherrer, Adrian Perrig, and Stefan Schmid
SIROCCO, Paderborn, Germany (virtual), 2020
- [2] **Incentivizing Stable Path Selection in Future Internet Architectures**
Simon Scherrer, Markus Legner, Adrian Perrig, and Stefan Schmid
PERFORMANCE, Milan, Italy (virtual), 2020
- [3] **Enabling Novel Interconnection Agreements with Path-Aware Networking Architectures**
Simon Scherrer, Markus Legner, Adrian Perrig, and Stefan Schmid
DSN, Taipei, Taiwan (virtual), 2021
- [4] **An Axiomatic Perspective on the Performance Effects of End-Host Path Selection**
Simon Scherrer, Markus Legner, Adrian Perrig, and Stefan Schmid
PERFORMANCE, Milan, Italy (virtual), 2021
- [5] **Model-Based Insights on the Performance, Fairness, and Stability of BBR**
Simon Scherrer, Markus Legner, Adrian Perrig, and Stefan Schmid
IMC, Nice, France, 2022 (IRTF APPLIED NETWORKING RESEARCH PRIZE)
- [6] **Quality Competition Among Internet Service Providers**
Simon Scherrer, Seyedali Tabaeiaghdaei, and Adrian Perrig
PERFORMANCE, Evanston, IL, USA, 2023
- [7] **A Dynamic Model-Based Perspective on BBR Fairness**
Simon Scherrer, Adrian Perrig, and Stefan Schmid
Under submission, 2023

The following articles (in chronological order) also contain research contributions of mine, but are not relevant to the topic of this dissertation:

[8] **Speed Records on Network Flow Measurement on FPGA**

Arish Sateesan, Jo Vliegen, Simon Scherrer, Hsu-Chun Hsiao, Adrian Perrig, and Nele Mentens
FPL, Dresden, Germany (virtual), 2020

[9] **Low-Rate Overuse Flow Tracer (LOFT): An Efficient and Scalable Algorithm for Detecting Overuse Flows**

Simon Scherrer, Che-Yu Wu, Yu-Hsi Chiang, Benjamin Rothenberger, Daniele Asoni, Arish Sateesan, Jo Vliegen, Nele Mentens, Hsu-Chun Hsiao, and Adrian Perrig
SRDS, Chicago, IL, USA (virtual), 2021

[10] **Bayesian Sketching for Volume Estimation in Data Streams**

Francesco Da Dalt, Simon Scherrer, and Adrian Perrig
VLDB, Vancouver, Canada, 2023

[11] **Carbon-Aware Global Routing in Path-Aware Networks**

Seyedali Tabaeiaghdaei, Simon Scherrer, Jonghoon Kwon, and Adrian Perrig
e-Energy, Orlando, FL, USA, 2023 (BEST-PAPER CANDIDATE)

[12] **ALBUS: a Probabilistic Monitoring Algorithm to Counter Burst-Flood Attacks**

Simon Scherrer, Jo Vliegen, Arish Sateesan, Hsu-Chun Hsiao, Nele Mentens, and Adrian Perrig
SRDS, Marrakesh, Morocco, 2023

[13] **GLIDS: Towards Global Latency Transparency**

Cyrill Krähenbühl, Seyedali Tabaeiaghdaei, Simon Scherrer, and Adrian Perrig
Unpublished, 2023

CURRICULUM VITAE

Simon Scherrer

born on 5 February 1994 in Rorschach SG, Switzerland

EDUCATION

ETH Zürich, Switzerland

Doctor of Sciences (Dr. sc. ETH) in Computer Science *2019 – 2023*

Network Security Group, supervised by Prof. Dr. Adrian Perrig

Thesis defense on 23 October 2023

Master of Science (MSc) in Computer Science *2017 – 2019*

Focus: Information Security

Bachelor of Science (BSc) in Computer Science *2013 – 2016*

Sichuan University, Chengdu, China

Chinese language program for overseas students

Kantonsschule am Burggraben, St.Gallen, Switzerland *2008 – 2012*

Swiss high-school diploma (Latin profile)

CONTENTS

Publications	ix
Curriculum Vitae	xi
1 Introduction	1
1.1 Motivating Trends	3
1.2 Research Questions	5
1.3 Thesis Approach: Modelling	9
1.4 Contributions	11
1.5 Methodology	18
 I NETWORK PERFORMANCE	
Part Overview	25
2 Information in End-Host Path Selection	29
2.1 Introduction	29
2.2 Model and First Insights	31
2.3 The Benefits of Information	39
2.4 The Drawbacks of Information	41
2.5 Case Study: Abilene Network	44
2.6 Related Work	46
2.7 Conclusion	48
3 Incentives in End-Host Path Selection	49
3.1 Introduction	49
3.2 Oscillation Model	51
3.3 Limits of Stable Strategies	59
3.4 Stabilization Mechanisms	62
3.5 The FLOSS Mechanism	64
3.6 The CROSS Mechanism	69
3.7 Practical Application	74
3.8 Related Work	80
3.9 Conclusion	81
4 Impact of Unstable Path Selection	83
4.1 Introduction	83
4.2 Model and Assumptions	85
4.3 Lossless Equilibria	92

4.4	Lossy Equilibria	98
4.5	Axioms	101
4.6	Axiom-Based Insights	106
4.7	Related Work	114
4.8	Conclusion	115
5	Modelling BBR Dynamics	117
5.1	Introduction	117
5.2	Network Fluid Model	119
5.3	BBR Fluid Model	122
5.4	Experimental Validation	131
5.5	Theoretical Analysis	140
5.6	Insights and Discussion	149
5.7	Related Work	151
5.8	Conclusion	153
6	Modelling BBR/CUBIC Competition Dynamics	155
6.1	Introduction	155
6.2	Background on BBR	157
6.3	The Steady-State Perspective	157
6.4	The Dynamic Perspective	164
6.5	Modelling Dynamic Competition	168
6.6	Fairness under Oscillation	175
6.7	Conditions for Oscillation	177
6.8	Preventing Oscillation	180
6.9	Related Work	184
6.10	Conclusion	186
II ISP ECONOMICS		
	Part Overview	189
7	ISP Cooperation: Interconnection Agreements	191
7.1	Introduction	191
7.2	The Relevance of GRC for BGP and PANs	194
7.3	Modelling Interconnection Economics	195
7.4	Optimization of Mutuality-Based Agreements	200
7.5	Mechanism-Assisted Negotiation	204
7.6	Effect on Path Diversity	215
7.7	Related Work	220
7.8	Conclusion	222
8	ISP Competition: Quality Attributes	225

8.1	Introduction	225
8.2	Model and First Insights	227
8.3	Theoretical Analysis	233
8.4	A Model Instance Based on Real-World Data	246
8.5	Simulation	252
8.6	Related Work	256
8.7	Conclusion	258
 III OUTLOOK AND CONCLUSION		
9	Future Work	263
9.1	Network Performance	263
9.2	ISP Economics	265
10	Conclusion	267
10.1	Network Performance	267
10.2	ISP Economics	272
10.3	Final Remarks	274
 IV Appendix		
A	Appendix of Chapter 2	279
A.1	Parallel Links: End-Host Optimum	279
A.2	Parallel Links: PI Equilibrium	280
A.3	Ladder Network: PI Equilibrium	281
A.4	Ladder Network: Proof of Theorem 2.3	282
B	Appendix of Chapter 3	289
B.1	Example of PSS Equilibrium Analysis	289
B.2	Proof of Insight 3.1: No PSS Equilibrium by Convergent Strategy	291
B.3	Proof of Insight 3.2: No PSS Equilibrium by MATE	291
B.4	CROSS Stability Analysis	293
C	Appendix of Chapter 4	295
C.1	Analysis of the Continuity-Time Distribution	295
C.2	Approximation Accuracy	298
C.3	Logical Consistency of P -Step Oscillation	300
C.4	Additional Figures	302
D	Appendix of Chapter 5	305
D.1	Loss-Based CCA Models	305
D.2	Aggregate Validation for Short RTT	307
D.3	Proofs for Stability Analysis	307
E	Appendix of Chapter 6	315

E.1	Proof of Theorem 6.1: Short-Term Equilibrium	315
E.2	Proof of Theorem 6.2: Stability of the Short-Term Equilibrium	321
E.3	Proof of Theorem 6.3: BBR-CUBIC Oscillation	332
E.4	Proof of Theorem 6.4: Maximally Intensive Oscillation	338
E.5	Deriving Update Function from Experiment Samples	342
F	Appendix of Chapter 8	345
F.1	Proof of Theorem 8.1: Best-Response Attribute	345
F.2	Proof of Theorem 8.2: Homogeneous Nash Equilibrium	349
F.3	Proof of Theorem 8.3: Stability of Homogeneous Equilibrium	352
F.4	Proof of Theorem 8.4: Suboptimal Homogeneous Equilibrium	355
F.5	Proof of Theorem 8.5: Homogeneous Competition (Attributes)	357
F.6	Proof of Theorem 8.6: Homogeneous Competition (Profits)	360
F.7	Proof of Theorem 8.7: Optimum on Heterogeneous Paths	362
F.8	Proof of Theorem 8.8: Equilibrium on Heterogeneous Paths	365
F.9	Proof of Theorem 8.9: Suboptimal Heterogeneous Equilibrium	366
F.10	Proof of Theorem 8.10: Two-Path Heterogeneous Equilibrium	366
F.11	Proof of Theorem 8.11: Stability of Heterogeneous Equilibrium	374
F.12	Proof of Theorem 8.12: Heterogeneous Competition (Improvement)	379
F.13	Proof of Theorem 8.13: Heterogeneous Competition (Degradation)	381

BIBLIOGRAPHY

Doctoral Research (Covered)	383
Doctoral Research (Not Covered)	385
References	387

INTRODUCTION

In 2023, around 5 billion users and 30 billion connected devices are expected to exchange 1.8 zettabytes of data over the Internet [62, 178], making the Internet arguably one of the most complex systems of human creation. This complexity is not least expressed in the *patterns of traffic flow* over the Internet, i.e., the data volume exchanged by endpoints in the Internet, the Internet paths traversed by this data, and the variation of this distribution over time.

This traffic-flow pattern is highly relevant to the effectiveness of the Internet as a communication tool, and to its attractiveness as a business environment: If these traffic flows consistently exceeded capacity limits or fluctuated wildly, the Internet would be substantially less reliable, usable, and profitable.

The concern for well-managed traffic-flow dynamics has regained urgency in recent years, due to a number of trends that might fundamentally reshape Internet traffic patterns. More precisely, these trends might affect Internet traffic dynamics in ways that are hard to predict, because traffic dynamics are an intricate phenomenon with temporal, probabilistic, performance-related, and economic aspects. However, *model-based research* can yield valuable insights into the traffic-flow dynamics of the modern Internet, as I argue in this thesis.

To make the objective of this thesis more concrete, I describe the trends affecting Internet traffic flows (§1.1), I outline the research questions provoked by these trends (§1.2), I discuss why model-based research gives valuable answers to the research questions (§1.3), I present my contributions (§1.4), and I provide an overview of the methods used in this thesis (§1.5). The overall structure of this thesis is also visualized in Fig. 1.1.

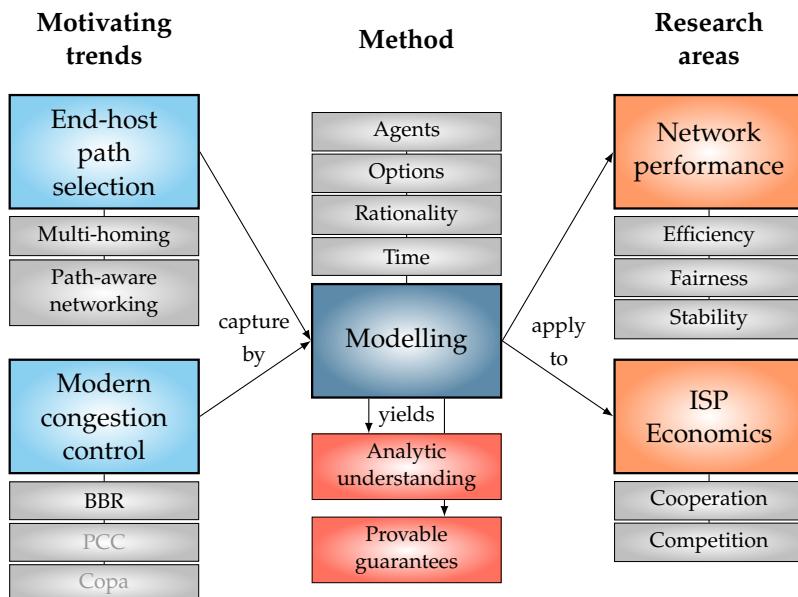


Figure 1.1: Structural overview of the dissertation.

1.1 MOTIVATING TRENDS

The Internet is experiencing multiple technological trends with the potential to reshape the dynamics of its traffic flows. In particular, I identify two such impactful trends, namely (i) the emergence of path-selection mechanisms for end-hosts, and (ii) the deployment of modern congestion-control algorithms. In the following, I describe these trends and their manifestation.

1.1.1 End-Host Path Selection

End-host path selection describes the capability of end-hosts to reach a given destination over different network paths, where paths can be selected by the end-host. While these paths could be arbitrarily constructed in traditional source routing, the modern and more practical form of end-host path control leaves the construction of paths or path segments to network operators, who then offer these predefined paths to the end-hosts.

Such path control is increasingly available in the Internet through two means, namely through (i) multi-homing, and (ii) path awareness.

Multi-homing. In multi-homing, an end-host has access to the Internet via multiple interfaces that are associated with different Internet paths. Such multi-homing already enables a limited degree of path control in today's Internet, and can be implemented with an increasing number of technological tools. For example, on the user side, mobile devices can nowadays access the Internet simultaneously via high-performance cellular networks and residential Wi-Fi connections [93]. On the side of content providers, resources can be made available under multiple IP addresses that are routed differently, e.g., with multiple ISP subscriptions or by using multiple cloud egresses [132]. However, in all of these approaches based on multi-homing, information about the available paths is generally not provided in a systematic manner.

Path awareness. This provision of path information by the network allows path awareness of end-hosts [245]. Such path awareness is the guiding principle of the SCION Internet architecture [60], which is currently undergoing practical deployment [106, 155] and standardization at the IETF [138]. SCION enables endpoints to select an Internet forwarding path from a set of pre-defined paths. This selection is then enforced by including a secure representation of the selected path

in packet headers. As such, SCION offers both path information and path control at a finer granularity than multi-homing-based path selection.

While multi-homing and path-aware architectures differ in the technical approach and the degree of path control, both paradigms give rise to new traffic dynamics in the Internet:

Multi-path usage. For example, while all traffic between two Internet endpoints traditionally used a single path at a time, this traffic might now traverse several paths simultaneously, based on application requirements and path characteristics. This decomposition of traffic flows might affect the severity of congestion at bottlenecks or alter the economic calculus of ISPs.

Load-adaptive path selection. Moreover, while end-hosts traditionally could only alleviate network congestion by reducing their sending rate, end-hosts with path-selection capabilities can circumvent congested links altogether. However, this load-adaptive path selection might itself lead to congestion or to unstable traffic patterns, and might thus negatively affect network performance, if performed poorly.

1.1.2 Modern Congestion Control

In transmission over data networks, congestion control refers to mechanisms aimed at avoiding the overload of network resources (e.g., links or router buffers), as opposed to avoiding overload of receiver capacity (i.e., flow control). Such congestion control is challenging because network resources are typically shared between multiple independent senders, and because congestion control should ideally be achieved by end-host algorithms without network support, according to the end-to-end principle [226]. These constraints give rise to the central question in congestion-control research: How can network feedback (e.g., loss, delay, and delivery rate) be interpreted to achieve efficient, fair, and stable bandwidth sharing?

Traditional congestion control. This question is notoriously difficult: While congestion-control research famously averted the congestion collapse of the Internet in the 1980s [130], the subsequent history of Internet congestion control has been marked by slow innovation. For a long time, the most consequential innovation was the adoption of CUBIC as the default congestion-control algorithm (CCA) of the Linux kernel in 2006 [117], securing CUBIC dominance until today [180]. Importantly, CU-

BIC still follows the loss-based paradigm of Reno, the original congestion-control algorithm [82].

New algorithms. In recent years, however, congestion-control research has produced a number of new congestion-control algorithms that are inspired by novel approaches and have been deployed in the Internet. For example, Google’s BBR algorithm [47] constructs a model of the network path, and has become the congestion-control algorithm used by YouTube [48]. PCC [75] uses an online-learning approach, and is used today by a major OTT streaming provider in the US [154]. Copa [29] is based on the previously known concept of delay targeting, but makes it robust to competition from loss-based CCAs, and has been adopted by Facebook for video upload [92]. Notably, all of these algorithms are latency-sensitive, while avoiding starvation in competition with traditional loss-based CCAs – unlike previous latency-based CCAs such as Vegas [39].

While these new approaches in recent CCAs improve performance in target scenarios, these approaches come at the cost of increased complexity, and make it more difficult to predict CCA behavior in the wider Internet, which is marked by enormous heterogeneity in network infrastructure, workloads, and CCA adoption. Moreover, this complexity in congestion control only grows due to the parallel trend of end-host path selection (§1.1.1), as little is known about the interaction of the two control processes.

1.2 RESEARCH QUESTIONS

As the preceding section illustrates, the technological trends of end-host path selection and modern congestion control promise to improve the performance and reliability of the Internet. At the same time, these trends raise numerous critical questions regarding their impact on users and network operators. While investigating this impact in all its facets is beyond the scope of this thesis, a model-based approach can illuminate the impact in essential areas. In particular, this thesis considers the impact of the mentioned trends on (i) the performance of the Internet as a communication tool (§1.2.1), and (ii) the economic interactions among Internet service providers (ISPs) (§1.2.2).

1.2.1 Network Performance

Since path-selection routines and congestion-control algorithms prescribe how networks should be used, these processes must be evaluated based on their impact on network performance, i.e., on the effectiveness of the Internet as a communication tool for users. Among the numerous aspects of network performance, I consider the essential aspects of efficiency, fairness, and stability in this thesis:

Efficiency. A key aspect of network performance is efficiency, which is understood here as the degree of optimality of transmission speed given the constraints of the network. On the one hand, this efficiency increases with link utilization: The utilization of the available network capacity should be complete, in order to maximize the network throughput, and thus to minimize the duration of data transmission. On the other hand, the network capacity should not be over-utilized, as such overloading has again negative consequences: Mild overload causes queuing in buffers, and thereby excess latency and jitter, whereas heavy overload results in packet loss and expensive retransmissions.

Fairness. To make practical statements about performance, efficiency considerations must be combined with fairness considerations, relating to the uniformity with which the individual end-hosts in the network benefit from the efficiency. Conventionally, fairness is considered perfect if the sharing of network resources is fully equal. However, this vague condition has proven surprisingly hard to translate into concrete fairness metrics, so far preventing a universally accepted precise definition of fairness.

Stability. Finally, efficiency and fairness must be characterized over time, as both path selection and congestion control involve adaptive and dynamic processes. From this dynamic perspective, a desirable performance property is stability, understood as the absence of persistent large variance of efficiency and fairness over time.

Such stability requires that the flow dynamics satisfy two conditions. First, the flow dynamics must possess equilibria (i.e., steady states), from which the flow dynamics do not spontaneously deviate. Second, the flow dynamics must eventually converge to these equilibria, which guarantees that fairness and efficiency will eventually be relatively static.

All of these three aspects are affected by the two recent trends of end-host path selection and congestion control, which provokes the following research questions:

End-host path selection. End-host path selection has implications on efficiency, namely via the traffic distribution across paths: If this traffic distribution over-utilized some paths while under-utilizing others, efficiency would suffer. Therefore, the traffic distributions caused by end-host path selection must be predicted and rated with respect to efficiency. In this prediction, it is central to assume that the selecting end-hosts in an Internet context are self-interested and uncoordinated:

Q 1 Path-Selection Efficiency. *How efficient are the traffic distributions across networks paths that can be expected from path selection by self-interested and uncoordinated end-hosts?*

In addition to efficiency, the stability of selfish end-host path selection is an important concern:

Q 2 Path-Selection Stability. *Under which conditions does path selection by selfish and uncoordinated end-hosts converge to steady traffic distributions?*

This concern for stability conditions is motivated by the unclear performance cost of instability:

Q 3 Performance under Unstable Path Selection. *How does unstable end-host path selection affect the efficiency and fairness in networks?*

Finally, for the sake of performance, the traffic distributions and the convergence behavior of end-host path selection should not only be understood, but also potentially shaped by network operators through information and incentives:

Q 4 Path-Selection Shaping. *How can network operators inform or incentivize selfish end-hosts to perform path selection in an efficient and stable manner?*

Congestion control. While an influencing factor in end-host path selection, congestion control in itself has important performance implications. Given the congestion-control innovations in recent years (§1.1.2), it is therefore natural to ask:

Q 5 Congestion-Control Performance. *How efficient, fair, and stable are the new congestion-control algorithms?*

Importantly, these performance implications must be understood under both homogeneous and heterogeneous CCA adoption, i.e., for competition among flows adopting the same CCA or different CCAs, respectively. Answering this question provokes a deeper question regarding the tools for CCA evaluation:

Q 6 Congestion-Control Methodology. What kind of evaluation methods can provide reliable insights into CCA behavior under the highly diverse conditions found in the Internet?

1.2.2 ISP Economics

The traffic flow patterns in the Internet are not only relevant for Internet users, but also for Internet service providers (ISPs) who generate profit from data forwarding. This profit is determined by the amount of traffic that ISPs can attract, plus the prices they can charge and the costs they incur for forwarding that traffic. In other words, the monetary transfers among ISPs follow the paths taken by traffic, and the increasingly end-host-based selection of these paths is thus highly relevant for ISP economics. In contrast, congestion control may affect the business calculus of an ISP by influencing resource utilization and thereby the provisioning of costly network equipment.

In this thesis, ISP economics are investigated with respect to ISP *interaction* given end-host path selection. Economic interaction among ISPs takes place in two fundamental dimensions, namely in (i) ISP cooperation, and (ii) ISP competition.

ISP Cooperation. If end-hosts have path control, the ISP decisions regarding forwarding directions do not exclusively determine the Internet traffic distribution. Forwarding decisions are typically implemented by means of interconnection agreements, in which neighboring ISPs cooperatively determine the conditions of mutual traffic forwarding, e.g., forwarding paths and prices. These interconnection agreements may thus change in scope and structure if end-host path selection loosens the connection between path provision and actual on-path traffic:

Q 7 Interconnection Agreements. How should interconnection agreements among ISPs be structured and negotiated under end-host path selection in the Internet?

Furthermore, if end-hosts can better avoid low-quality paths, the economic performance of ISPs might become more strongly dependent on the quality of provided paths. Notably, the quality of an Internet path is also the result of ISP cooperation, but cooperation of a fickle sort: Potentially multiple on-path ISPs jointly determine path quality, e.g., by individually making costly investments such as installing bandwidth; however, each ISP is self-interested and optimizes only its own profit. This distributed decision making by self-interested on-path ISPs raises the question:

Q 8 Intra-Path Cooperation. *How is the quality of Internet paths affected by cooperation among self-interested on-path ISPs if end-hosts select paths based on quality?*

ISP Competition. Like ISP cooperation, also the competition among ISPs is a determining factor of Internet path quality. Since ISP competition is essentially created by path selection, the transfer of path control to end-hosts provokes the following question:

Q 9 Inter-Path Competition. *How is the quality of Internet paths affected by ISP competition if end-hosts select paths based on quality?*

1.3 THESIS APPROACH: MODELLING

The research questions from the previous section allow analysis in various ways. For example, the performance effects of path selection and congestion control (Q 1–6) might be investigated with experiments in network test-beds, whereas the economic response to path selection (Q 7–9) could be investigated by interviewing ISP strategists and learning about their coping strategies.

In this thesis, however, I opt to investigate all questions with an approach based on *models*, i.e., mathematical descriptions of the agents, processes, and dependencies relevant to each investigated phenomenon. Such models have a number of methodological strengths that make them suitable for the objective of this thesis (§1.3.1). Simultaneously, model-based research faces a number of challenges that deserve special attention (§1.3.2).

1.3.1 Strengths

A model-based approach complements previous literature thanks to the following strengths:

Improve analytic understanding. Networks are highly complex systems, in which system components influence each other in numerous and subtle forms. Understanding these dependencies is significantly helped by models, as models make these dependencies explicit in mathematical formulas. Such analytic transparency is especially important for the study of effects (e.g., of path selection), as it allows to trace the causality chain from a root cause to other parts of the complex system,

thereby enabling a comprehensive understanding. To name a concrete example, describing a congestion-control algorithm (CCA) in a model allows to identify, quantify, and explain the effects of a delay measurement in the first order, e.g., the corresponding rate reduction prescribed by the CCA, and in higher orders, e.g., the impact on competitor flows. In contrast to modelling, experiments in general do not explain equally clearly how the experiment results came about, and might further complicate the closer investigation of observations by reproducibility issues.

In some cases, models are not only more effective at generating insights than other methods, but are in fact the only applicable method under practical constraints. These cases typically involve global effects in large-scale networks, which can be predicted by models, but barely with other approaches. For example, an experimental evaluation of performance effects in large-scale networks requires the physical construction of such networks, which is out of reach for many researchers. Similarly, interviewing real-world ISPs might illuminate rational business strategies from the perspective of an individual ISP, but does not predict the global effects that arise from interacting individual strategies. In contrast, these macroscopic effects can be captured by models, both by theoretical analysis and model-based simulation.

Derive proven guarantees. The analytical transparency of models does not only yield explainable insights, but also provides high confidence in these insights. This high confidence can be created by applying mathematical tools to the model, which then enables rigorous proofs of model properties. For example, applying stability theory to the fluid model of a CCA allows to prove that the CCA is stable, i.e., that the CCA is guaranteed to converge. Such guarantees are especially valuable in the context of the Internet, as Internet disruptions may be caused under circumstances that are difficult to identify with an intuitive or experiment-based approach.

1.3.2 Challenges

In addition to methodological strengths, models also encounter a number of methodological challenges, which receive special attention throughout this dissertation:

Over-abstraction. Models typically represent a simplified version of the targeted system, as capturing the full system complexity usually makes analysis intractable for both theoretical investigation and simulation. Such abstraction is a powerful approach if it reduces complexity while retaining the aspects that most significantly determine the observations of the system. However, if abstraction is excessive, the

model becomes unrealistic, and the model-based insights lose relevance for real-world systems. Hence, modelling research must strike the right balance between complexity reduction and realism.

Sensitivity to assumptions. As mentioned above, modelling virtually always relies on simplifying assumptions, namely that certain internal system dependencies are negligible and need not be represented in the model. Another set of assumptions is introduced in the numerical instantiation of models, where model variables are assigned a number to allow quantitative predictions. Both sets of assumptions may lead to unreliable predictions under excessive model sensitivity, i.e., if model predictions change considerably even if the assumptions are only slightly modified. Therefore, models should be developed with the goal of *robustness*, i.e., the model predictions should have limited sensitivity to the assumptions made.

1.4 CONTRIBUTIONS

After championing a model-based approach in the previous section, I demonstrate in this section that the model-based approach effectively enables a number of important research contributions. Concretely, the following sections lists the research insights of my dissertation, gained through the application of modelling to the research questions from §1.2. These research questions are naturally segmented into the domains of network performance (§1.4.1) and ISP economics (§1.4.2). Both these domains are affected by changes in traffic patterns (cf. §1.2) and are associated with Parts i and ii of this thesis, respectively.

1.4.1 Network Performance

During my doctoral studies, I investigated the performance effects of end-host path selection and modern congestion control, and arrived at the following insights:

1.4.1.1 Information in End-Host Path Selection

As end-hosts gain path control in the Internet, their self-interested and uncoordinated path-selection decisions will increasingly determine the traffic distribution across Internet paths.

State of the art. To investigate the efficiency of these distributions, previous work builds on the classic game-theoretic model of *selfish routing* [224]. In this model,

end-hosts play a congestion game by selecting the paths with minimal cost, where path cost depends on path load. The Nash equilibria of this game then represent the traffic distributions that can be expected from self-interested path selection. These equilibrium distributions are compared to socially optimal traffic distributions in order to compute the *Price of Anarchy* [148].

Contribution. To allow more specific insights on end-host path selection in the Internet, I refine the selfish-routing model in two main ways. First, I introduce two different notions of social optima, capturing the perspectives of end-hosts and network operators, respectively. Second, I newly distinguish two degrees of path-information detail that end-hosts possess, and thereby investigate how network operators could shape path selection by providing information. Intriguingly, I find that (i) different degrees of information induce different equilibrium traffic distributions, (ii) more detailed information is not always superior to less detailed information with respect to the efficiency of the associated equilibria, both for end-hosts and network operators, and (iii) near-optimal traffic distributions tend to arise even if end-hosts only have basic latency information that they can measure themselves.

Research questions. The insights in Chapter 2 are relevant for Research Questions Q 1 (Path-Selection Efficiency) and Q 4 (Path-Selection Shaping).

1.4.1.2 Incentives in End-Host Path Selection

In addition to path information, incentives play a crucial role regarding the performance effects of end-host path selection.

State of the art. Previous research has suggested incentives in the form of prices in order to improve the efficiency of traffic-distribution equilibria [232]. However, the role of incentives is curiously under-explored in the context of *dynamic* path selection by end-hosts, which is relevant for stability. In dynamic path selection, the lack of concern for incentives has produced a situation where numerous path-selection strategies are known to be stable [79, 85], but no strategy is known to be *rational*, i.e., suitable for adoption by self-interested end-hosts.

Contribution. Reasoning about self-interest in dynamic path selection requires a new game-theoretic framework, which is presented in Chapter 3. To develop that framework, I again leverage the selfish-routing model (as in Chapter 2), but

additionally consider the dynamic perspective where end-hosts select paths over time based on outdated load information. In this environment, Chapter 3 evaluates the stable path-selection strategies from previous research on their rationality under universal adoption, and finds that these strategies are in fact irrational. Hence, if a stable path-selection strategy is universally adopted, it becomes worthwhile to adopt an unstable strategy. As a result, the stable strategy will quickly be abandoned by self-interested end-hosts, and load oscillation will result from the newly adopted unstable strategy.

To counter this selfishness-induced instability, Chapter 3 suggests that network operators take an active role in creating incentives for stability. More concretely, Chapter 3 presents two stabilization mechanisms, which are provably incentive-compatible and can be applied by network operators with limited overhead.

Research questions. The insights in Chapter 3 are relevant for Research Questions Q 2 (Path-Selection Stability) and Q 4 (Path-Selection Shaping).

1.4.1.3 Impact of Unstable Path Selection

While Chapter 3 demonstrates why oscillatory path selection arises and how it can be disincentivized, the analysis in Chapter 3 does not show why oscillatory path selection is actually undesirable, i.e., whether and how this oscillation hurts network performance.

State of the art. In fact, the performance effects of oscillatory path selection are discussed by surprisingly little previous research, and are generally not well understood. To analyze these effects, any analysis must take into account that end-hosts are running congestion-control algorithms in addition to path selection, as congestion control is a crucial network process that is potentially affected by oscillation. Representing congestion control requires a departure from the selfish-routing model used in Chapters 2 and 3, as this model assumes that the sending rate of any single end-host is infinitesimal and thus effectively static. This limitation is overcome by fluid models of multi-path congestion-control algorithms [144, 204, 259], such as LIA and OLIA. However, these models investigate fairness and stability concerns of specific multi-path congestion-control algorithms, rather than investigating the impact of instability on other network-performance metrics.

Contribution. To close this research gap, Chapter 4 presents a model-based approach to characterize network performance under unstable path selection. Instead,

Chapter 4 relies on a model proposed for an axiomatic analysis of congestion control [268], and extends it with simultaneous path selection. This extension also has the benefit that the axioms related to the original model can be used for evaluation in our context, albeit after some adaptation. These axioms correspond to performance metrics, rating the equilibrium behavior of joint path selection and congestion control under a variety of parameters.

This axiomatic approach yields a number of valuable insights. For example, oscillatory path selection creates a fundamental trade-off between efficiency and loss avoidance on one side and fairness and responsiveness on the other side. Furthermore, the model enables to compare network performance under oscillatory path selection to network performance without any path selection, and suggests for example that oscillatory path selection may reduce packet loss in some scenarios.

Research questions. The insights in Chapter 4 are relevant for Research Questions Q 3 (Performance under Unstable Path Selection) and Q 6 (Congestion-Control Methodology).

1.4.1.4 BBR Dynamics

Among the modern CCAs with real-world deployment (cf. §1.1.2), Google’s BBR algorithm is most widely used with an estimated adoption by 40% of Internet downstream traffic [180].

State of the art. Despite this significance for Internet performance, BBR performance has been incompletely described by previous research, due to methodological limitations. For one, experiment-based studies test BBR implementations in real networks, and thus face a trade-off regarding the number and scale of evaluated settings: Building more diverse and larger test networks achieves more generalizable results, but also incurs higher experiment cost. In contrast, model-based research avoids the need to build physical networks. However, the models in previous research exclusively describe the steady state of BBR, and therefore cannot describe transient phenomena such as short flows or BBR convergence.

Contribution. These gaps in BBR understanding call for a fluid-model approach, which is pursued for the first time in Chapter 5. More precisely, Chapter 5 presents the first fluid model of both BBRv1 and BBRv2, consisting of differential equations that describe CCA behavior over time. This fluid model lends itself to both efficient simulation and theoretical stability analysis, and therefore enables important in-

sights into the efficiency, fairness, and stability of BBR. Of course, the correctness of these insights relies on the accuracy of the fluid model, which is, however, confirmed by (i) an experimental validation of fluid-model predictions in diverse network scenarios, and (ii) the fluid-model reproduction of numerous insights from prior research. Besides these previous insights, the fluid model also yields new insights, for example regarding (i) a bufferbloat issue in BBRv2, (ii) unfairness of BBRv2 towards loss-based CCAs under a random-early-drop (RED) queuing discipline, and (iii) the provable convergence of BBRv1 and BBRv2.

Research questions. The insights in Chapter 5 are relevant for Research Questions Q 5 (Congestion-Control Performance) and Q 6 (Congestion-Control Methodology).

1.4.1.5 BBR/CUBIC Competition

Among the numerous performance considerations regarding BBR, the issue that has received most attention is the fairness of BBR towards traditional loss-based CCAs in general, and the CUBIC algorithm in particular.

State of the art. This competition between BBR and CUBIC has previously been investigated by both experiment-based analyses [119, 231] and model-based research [181, 254], and has even inspired a new harm-based fairness notion [255]. However, previous model-based research is exclusively based on steady-state models, i.e., models that identify equilibria of the competition dynamics and make fairness predictions based on these equilibria.

Contribution. However, this fairness is more accurately predicted by dynamic fluid models than by previous steady-state models, as Chapter 6 demonstrates. Intriguingly, dynamics are important for fairness prediction because the competition of BBR flows and CUBIC flows may suffer from persistent load oscillation, i.e., alternating up- and down-swings of CCA-specific throughput, and frequently do not even converge to equilibria. This oscillation is not only predicted by fluid-model simulation, but also explained and quantified by a theoretical stability analysis based on the fluid model. In particular, the fluid model illustrates that the oscillation is created by the periodic RTT measurements by the BBR flows, which is distorted by the competing flows: If the CUBIC flows heavily utilize the buffer when the BBR flows are measuring their RTT, the resulting high RTT measurements cause a high BBR rate and a low CUBIC rate; if the BBR flows then probe the RTT again, the CUBIC flows do not heavily utilize the buffer anymore, and the sending rates

evolve in the reverse direction. While this oscillation cause could also be identified by a qualitative algorithm analysis, the fluid-model analysis additionally identifies quantitative network conditions under which the oscillation provably arises. More formally, under some network conditions, the joint dynamics of BBR and CUBIC flows are unstable, which Chapter 6 proves with a combination of Lyapunov theory, center-manifold theory, and fixed-point iteration. Unfortunately, these network conditions are frequently satisfied in common networks, and the oscillation issue is not easily remedied by BBR modifications, as Chapter 6 demonstrates.

Research questions. The insights in Chapter 6 are relevant for Research Questions Q 5 (Congestion-Control Performance) and Q 6 (Congestion-Control Methodology).

1.4.2 ISP Economics

Besides the insights on network performance in §1.4.1, my doctoral research also yields insights on ISP economics. Concretely, I made the following contributions towards understanding how Internet path selection affects economic cooperation and competition among ISPs:

1.4.2.1 Interconnection Agreements

Cooperation among ISPs is most concretely embodied by interconnection agreements, in which ISPs record the conditions for mutual traffic forwarding.

State of the art. Clearly, these interconnection agreements are only concluded if the agreement conditions are viable in both technical and economic terms: The agreements must neither lead to routing instability nor to monetary loss for any agreement party. To respect both of these concerns, interconnection agreements in today's Internet are restricted by the so-called Gao-Rexford conditions in the traditional Internet [91]. These conditions rule out intermediation between providers or peers, guarantee BGP convergence, and ensure that any ISP is compensated for its transit service.

Contribution. Given this significance of the Gao-Rexford conditions, Chapter 7 re-evaluates these conditions in the context of emerging path-aware Internet architectures, and arrives at an interesting observation: Intriguingly, path-aware Internet architectures render the Gao-Rexford conditions obsolete from a technical perspective, because these architectures achieve convergence of path discovery by design.

However, path-aware Internet architectures still need to respect the economic self-interest of ISPs, which is protected by the Gao-Rexford conditions today. Therefore, path-aware Internet architectures enable interconnection agreements that violate the Gao-Rexford conditions and thereby increase path diversity – that is, if these agreements can be made economically viable.

To tackle this challenge, Chapter 7 proposes agreement structures in which peers provide each other with access to their respective peers or providers, and thereby allow mutual benefit from the agreement. To not only allow, but also ensure mutual benefit, Chapter 7 extends this basic agreement idea with either (i) volume targets for transit flows or (ii) measures for financial compensation. Based on these conditions, efficient and fair agreements can be constructed, but only if the agreement parties truthfully report the utility that they derive from the agreement. However, this truthfulness cannot be expected in inter-ISP negotiation, especially as ISPs usually benefit from understating their agreement utility. To actively incentivize truthfulness, Chapter 7 also proposes a bargaining mechanism that structures the negotiation and thereby makes the conclusion of efficient and fair agreements more likely. Such facilitation of agreements is very valuable in terms of path diversity, which is also shown by an Internet topology analysis in Chapter 7: With universal conclusion of the proposed agreements, 50% of autonomous-system pairs would reduce latency between them, and 35% of AS pairs would obtain more bandwidth between them, compared to today.

Research questions. The insights in Chapter 7 are relevant for Research Questions Q 7 (Interconnection Agreements) and Q 8 (ISP Cooperation).

1.4.2.2 ISP Quality Competition

Importantly, the ISP cooperation mentioned above is embedded in an environment of competition. ISP competition arises from path selection, where this path selection is performed by autonomous systems (ASes) during route discovery, and additionally by end-hosts during data forwarding in future path-aware Internet architectures. Path-selecting entities of any kind base their selection at least in part on the *quality* of paths. Broadly understood, this quality may encompass aspects related to performance (e.g., bandwidth, latency, and jitter), security (e.g., presence of security middleboxes), sustainability (e.g., carbon emission), and legal matters (e.g., traversed jurisdictions). Clearly, these quality attributes can be improved by on-path ISPs in order to attract traffic and increase revenue, but these improvements also come at a cost. Hence, ISPs face the complex challenge of dimensioning their

quality investments, in the context of investment decisions by both competing ISPs on alternative paths and other ISPs on the same paths.

State of the art. While multiple models for ISP competition already exist, these previous models fail to represent important aspects of ISP competition. For example, previous models neglect that (i) transmission quality in the wider sense is determined by a combination of multiple attributes, (ii) quality attributes may have different impacts on fixed and variable cost of an ISP, and (ii) multiple self-interested ISPs jointly provide the paths to the path-selecting entities.

Contribution. To better inform quality-investment decisions and better understand the global effects of ISP quality competition, Chapter 8 presents a new game-theoretic model of ISP quality competition in networks. This model reflects the important competition aspects mentioned above, among other aspects neglected by previous models. Based on this model, Chapter 8 yields a variety of interesting insights, e.g., regarding the conditions under which ISP competition is beneficial or detrimental to path quality, regarding the profit impact of ISP competition, and regarding the prisoner's dilemma among ISPs on the same path. These insights are confirmed by both a theoretical analysis, which identifies asymptotically stable competition equilibria, and a simulation-based case study, which instantiates the model based on real-world data.

Research questions. The insights in Chapter 8 are relevant for Research Questions Q 8 (ISP Cooperation) and Q 9 (ISP Competition).

1.5 METHODOLOGY

While the preceding section provides an overview of the insights gained during my doctoral research, the section at hand provides an overview of the methods applied to gain these insights. In particular, this section presents the modelling choices made in the individual chapters, which are summarized in Table 1.1.

As Table 1.1 suggests, model design involves a number of fundamental choices that determine the model suitability for specific research questions, and the model complexity. In this thesis, the fundamental model-design questions include:

Agents. What real-world entities are *agents* in the model, i.e., decision makers?

Options. What options do agents in the model have in their decisions?

Rationality. To which extent are the model agents rational in a game-theoretic sense, i.e., self-interested and able to optimize decisions?

Time. How is time represented in the model?

In the following, I will discuss each model-design question in detail, and justify the model-design decisions made in the individual chapters.

1.5.1 Agents

Agents, i.e., decision-making units, come in various forms throughout this thesis:

End-hosts. In all models in Chapters 2–4, decisions are made by end-hosts, i.e., systems that sit at the edge of the network, and exclusively use the network for reaching other end-hosts, rather than contributing to the network via data forwarding. End-hosts are the natural agents in modelling end-host path selection, which is the subject of Chapters 2–4.

Flows. The models in Chapters 5 and 6 allow for decisions to be made by network flows, understood here as the system at an end-host, controlling the data transmission over a single path towards another end-host. Given this definition, flows are more limited agents than end-hosts, but have sufficient agency to express the (single-path) congestion-control functionality investigated in Chapters 5 and 6.

ISPs. Since Part ii covers ISP economics, the agents in the corresponding models are ISPs, i.e., organizations that optimize profit, and have both technical abilities (e.g., forwarding traffic to neighbors) and economic abilities (e.g., agreement negotiation, monetary transfers, investments).

1.5.2 Options

The agents from the preceding sections have a variety of decision options, depending on the research question:

Path. In the models for end-host path selection in Chapters 2–4, the end-host agents in the models naturally decide on the path used for data forwarding. This path decision is either one-off in the static model of Chapter 2, or varied over time in the dynamic models of Chapters 3 and 4.

Rate. For agents in congestion-control models (Chapters 4–6), the key decision is how to adapt the sending rate over time. This decision is intertwined with path selection in the model of Chapter 4, and is independent of path selection in the models of Chapters 5 and 6.

Negotiation offer. Chapter 7 models the negotiation of interconnection agreements between self-interested and intransparent ISPs. Hence, the ISP agents in this model control the offer with which they enter the negotiation, where an offer may contain (i) a set of forwarding paths, or (ii) demands for traffic-volume limits or monetary compensation.

Quality attributes. Finally, the ISP agents in Chapter 8 try to optimize their profit by adapting various aspects of transmission quality, based on attractiveness and cost of these quality indicators. Hence, these quality attributes represent the space of decision options in the model of Chapter 8.

1.5.3 Rationality

When deciding on options for action, agents follow a rationality notion:

Algorithm compliance. The congestion-control chapters in Chapters 4–6 of this work do not involve self-interested agents which evaluate decision options for their own benefit. Instead, the agents (flows) in the respective models comply with the rate adaption prescribed by the congestion-control algorithm.

Full rationality. If agents are considered self-interested, the standard assumption is that these agents are fully rational, i.e., the agents have the capability to identify the optimal decision option from their individual perspective. This assumption of full rationality is attractive for its intrinsic simplicity, and the simplifying effects on theoretical analysis. Hence, full rationality is assumed in the theoretical analyses of Chapters 2, 3, 7, and 8.

Bounded rationality. The alternative to the full-rationality assumption is given by bounded rationality, describing a limited capacity to identify the decision option that best satisfies self-interest. While such bounded rationality is potentially more realistic, the bounding of rationality requires that the limitations of agent rationality are specified and justified. In this thesis, bounded rationality is assumed in two chapters for varying reasons. In Chapter 4, the agents have bounded rationality

because they follow a simple greedy strategy for path selection, although this strategy is sub-optimal. However, this strategy is guaranteed to lead to load oscillation, which is the object of investigation in Chapter 4. In contrast, Chapter 8 uses a bounded-rationality assumption for a large-scale simulation, where the ISP agents can only identify a strategy that improves their profit compared to their current strategy, but not identify the profit-optimal strategy. Limiting the agents in this manner enables tractability of the simulation, as identifying a profit-improving strategy is computationally cheaper than identifying the profit-optimal strategy.

1.5.4 Time

The final model choice concerns the representation of time:

No time. Static models do not represent the temporal evolution of agent decisions at all, which is sufficient for some research questions. In particular, the model in Chapter 2 is supposed to predict the equilibrium traffic distributions under end-host path selection, but does not need to represent how end-hosts arrive at these distributions. Similarly, the model in Chapter 7 is used to construct interconnection agreements that are acceptable to the negotiating parties, which does not require to understand the process of negotiation over time.

Continuous time. If a model should describe dynamic phenomena, continuous time is the most realistic time notion possible. However, representing continuous time requires specialized mathematical tools in the models: In this thesis, a notion of continuous time is achieved whenever the model relies on differential equations, i.e., in Chapters 3, 5, 6, and 8.

Discrete time. When constructing a model, various reasons may make it preferable to represent time in a discrete fashion, i.e., as a series of time steps, rather than continuously. In this thesis, a discrete-time notion is applied in Chapters 4, 6, and 8, for different reasons. In Chapter 4, time is discretized because the model in the chapter extends a previous axiomatic model [268], which uses a time-stepped model of congestion-control algorithms for simplicity. In Chapter 6, part of the model relies on fixed-point iteration over discrete time because the underlying real-world process in BBR actually proceeds in discrete steps. Finally, Chapter 8 employs a discrete-time notion in its simulation part, which relies on step-wise better-response dynamics to accelerate convergence.

Chapter	Agent	Option	Rationality	Time
Part I: Network Performance				
Chapter 2: Information in End-Host Path Selection	End-host	Path	Full	–
Chapter 3: Incentives in End-Host Path Selection	End-host	Path	Full	Continuous
Chapter 4: Interaction of End-Host Path Selection and Congestion Control	End-host	Path, Rate	Bounded	Discrete
Chapter 5: Modelling BBR Dynamics	Flow	Rate	–	Continuous
Chapter 6: Modelling BBR/CUBIC Competition	Flow	Rate	–	Continuous, Discrete
Part II: ISP Economics				
Chapter 7: ISP Cooperation: Interconnection Agreements	ISP	Negotiation offer	Full	–
Chapter 8: ISP Competition: Quality Attributes	ISP	Quality attributes	Full (Theory), Bounded (Simulation)	Continuous (Theory), Discrete (Simulation)

Table 1.1: Overview of modelling choices.

Part I

NETWORK PERFORMANCE

PART OVERVIEW

In recent years, Internet traffic patterns have been increasingly reshaped by two emerging paradigms that promise to improve network performance, namely end-host path selection and competitive latency-sensitive congestion control. These new paradigms affect Internet traffic patterns in two main aspects. First, end-host path selection affects the distribution of traffic across the Internet topology, as multi-homing and emerging path-aware network architectures such as SCION [138, 149] allow end-hosts to select between inter-domain forwarding paths. Second, modern congestion-control algorithms increasingly govern the evolution of traffic volumes over time, where the most prominent algorithms BBR [47, 50, 52], PCC [75], and Copa [29] are all latency-sensitive while still being competitive with traditional loss-based algorithms.

Both end-host path selection and modern congestion-control algorithms have been experimentally evaluated in controlled settings, showing significant performance benefits. However, the performance implications of both approaches are much less studied on an analytical level, i.e., using mathematical methods that can provide proven guarantees. However, such a theoretical understanding is essential for a responsible global deployment of new Internet technologies, as an analytic perspective might highlight performance issues that are not revealed by experiments in managed environments.

In this thesis part, I employ an analytic approach to investigate the following performance-relevant aspects of end-host path selection and congestion control:

Information in end-host path selection. Given path control, end-hosts can perform load-adaptive path selection in a selfish fashion, which influences the traffic distribution across the network topology. The efficiency of this distribution depends on the path information provided to end-hosts, which begs the question: What kind of path information leads to globally desirable traffic distributions? To answer this question, Chapter 2 extends the classic selfish-routing literature by introducing differing degrees of path-information detail, and by contrasting the equilibrium traffic distributions induced by different information degrees. Surprisingly, information that can be directly measured by end-hosts themselves (i.e., path latency) induces near-optimal traffic distributions in many realistic scenarios, whereas more detailed information even tends to cause less efficient traffic distributions.

Incentives in end-host path selection. The selfish and strategic behavior of end-hosts does not only affect the equilibrium traffic distribution, but also the path-selection dynamics outside of this steady state. In particular, selfish load-adaptive path selection causes load oscillation unless the end-hosts have an incentive to adopt a stable path-selection strategy. Interestingly, Chapter 3 demonstrates that end-hosts generally have no such incentive, because deviating end-hosts improve their performance with unstable strategies if all other end-hosts adopt stable strategies. Hence, network operators have to *create* an incentive for adopting stable path-selection strategies by means of incentive-compatible stabilization mechanisms, also introduced by Chapter 3.

Impact of unstable path selection. While Chapter 3 demonstrates how unstable path selection can be prevented by network operators, it does not demonstrate the performance benefits of this stabilization. To quantify such potential benefits, the performance impact of unstable path selection must first be properly understood. To that end, Chapter 4 presents a model that captures both the dynamics of oscillatory path-selection and the dynamics of parallel congestion control, as the interaction of these two control loops is significant for performance. The model then allows to quantify the performance effects of this oscillatory path selection by leveraging an axiomatic approach, suggesting that oscillatory path selection introduces fundamental trade-offs between key objectives of network performance.

BBR dynamics. While Chapter 4 discusses congestion control only as a complicating factor for path selection, congestion control is a crucial subject for the future Internet in its own right, especially since multiple novel congestion-control algorithms have recently been proposed and deployed. Among these algorithms, Google’s BBR algorithm is the most prominent example, as BBR governs around 40% of downstream Internet traffic. This wide deployment makes understanding the behavior and the limitations of BBR consequential for the Internet. Chapter 5 contributes to this understanding by presenting a new fluid model for both BBR versions, based on novel modeling techniques, and by leveraging the model for both extensive simulation and theoretical stability analysis. This two-fold investigation confirms findings from previous research (e.g., BBR unfairness in shallow buffers), yields new insights into BBR behavior (e.g., bufferbloat in BBRv2, asymptotic stability), and thereby underlines the methodological utility of fluid models.

BBR/CUBIC competition dynamics. Based on the fluid model in Chapter 5, Chapter 6 presents further analysis of a crucial aspect of BBR, namely its fairness towards

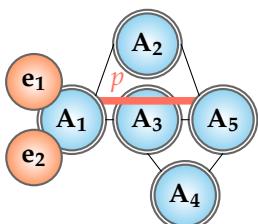
CUBIC, the de-facto standard CCA in the Internet. Since the fluid model allows to represent the competition of BBR and CUBIC over time, the model reveals that BBR/CUBIC competition suffers from persistent load oscillation. This susceptibility to oscillation is also rigorously proven by means of control theory, and is not easily remediable without further drawbacks in term of fairness or performance.

INFORMATION IN END-HOST PATH SELECTION

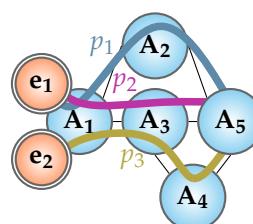
2.1 INTRODUCTION

Traditional Internet infrastructure is based on a forwarding mechanism that grants almost exclusive control to the network and almost no control to users (or *end-hosts*). In fact, all communication from a given end-host to another end-host takes place over the *single AS-level path* that BGP (Border Gateway Protocol) converged on. In the upcoming paradigm of *end-host path selection* [245], however, network operators supply end-hosts with a pre-selected set of paths to a destination, enabling end-hosts to select a forwarding path themselves (cf. Fig. 2.1).

End-host path selection allows end-hosts to select paths tailored to their application requirements, or to quickly switch to an alternative path upon link failures. However, a widely shared concern about end-host path selection regards the loss of control by network operators, which fear that the traffic distribution resulting from individual user decisions might impose considerable cost on both themselves and their customers. Another concern is that end-hosts require path-load information



(a) Network-based path selection.



(b) End-host path selection.

Figure 2.1: Network-based vs. end-host path selection. Decision makers (ASes A_i or end-hosts e_i) in the respective paradigms are marked with a double border.

in order to perform path selection effectively, necessitating complex and expensive systems for the dissemination of network-state information.

To investigate the validity of these concerns, we refine and extend concepts from the selfish-routing literature in this chapter. This literature on selfish routing is based on the observation that the uncoordinated path selection by selfish agents can produce sub-optimal traffic allocations in networks [221, 222, 224]. While seminal work on such game-theoretic analyses dates back to Wardrop [253], especially the notion of *Price of Anarchy*, coined by Koutsoupias and Papadimitriou [148], has received much attention: the Price of Anarchy compares the worst possible outcome of individual decision making, i.e., the worst Nash equilibrium, to the global optimum, by taking the corresponding cost ratio. The Price of Anarchy in network path selection is typically measured in terms of *latency*.

In this chapter, we revisit these concepts to investigate one key aspect which has been less explored in the literature so far and is highly relevant for newly emerging path-aware network architectures, namely the *impact of information* that end-hosts possess about the congestion state of network links. A fundamental design question of network architectures concerns which information about the network state should be shared with end-hosts, beyond the latency information that can be observed by the end-hosts directly.

2.1.1 Contributions

This chapter is based on my publication at SIROCCO 2020 [1]. It presents a game-theoretic model (§2.2) that quantifies not only the Price of Anarchy experienced by end-hosts, but also accounts for the network operators. Furthermore, we use our model to explore how end-host information about the network state affects the Price of Anarchy.

We find that different levels of information indeed lead to different Nash equilibria and thus also to different Prices of Anarchy. Intriguingly, we find that while more information can improve the efficiency of selfish routing in networks with few end-hosts (§2.3), more information tends to induce a *higher* Price of Anarchy in more general settings (§2.4). Indeed, near-optimal outcomes are typically achieved if end-hosts select paths based on simple latency measurements of different paths. These theoretical results suggest that end-host path selection cannot only achieve a good network performance in selfish contexts, but can be realized in a fairly light-weight manner, avoiding the need to distribute much information about the network state. This insight is validated with a case study on the Abilene topology (§2.5).

2.2 MODEL AND FIRST INSIGHTS

In this section, we present a network model of path-aware networking architectures (§2.2.1), define the social optima from both the end-host perspective and the network-operator perspective (§2.2.2), formalize different degrees of information (§2.2.3) and characterize Nash equilibria based on these different information assumptions (§2.2.4). Finally, we leverage the concept of the Price of Anarchy (§2.2.5) to establish a new concept that we call the *Value of Information* (§2.2.6).

2.2.1 Model

Wardrop Model. As in previous work on selfish routing [85, 224], our model is inspired by the classic Wardrop model [253]. In this model, the network is abstracted as a graph $G = (A, L)$, where the edges $\ell \in L$ between the nodes $A_i \in A$ represent links. Every link $\ell \in L$ is described by a link-cost function $c_\ell(f_\ell)$, where f_ℓ is the amount of load on link ℓ , i.e., a *link flow*. Typically, link-cost functions are seen as describing the latency behavior of a link. To reflect queuing dynamics, link cost functions are convex and non-decreasing. For every node pair (A_i, A_j) , there is a set of paths $P(A_i, A_j)$ that contains all non-circular paths between A_i and A_j . Between any node pair (A_i, A_j) , a demand d is shared by infinitely many agents, where each agent is controlling an infinitesimal share of traffic.

End-host-in-AS Model. However, the traditional Wardrop model is not suitable to analyze traffic dynamics in an Internet context, where *autonomous systems* (ASes) accommodate *end-hosts*. We thus adapt the Wardrop model into a more realistic model as follows. First, an AS $A_i \in A$ is represented by a node in the network graph G . The AS contains a set of end-hosts, which are the players in the path-selection game. Differently than in the Wardrop model, we allow for non-negligible, heterogeneous demand between end-host pairs in order to accommodate the variance of demand in the Internet. For example in origin-destination pair $od = (e_s, e_t) \in OD$ (short: (s, t)), an end-host $e_s \in A_i$ can have a demand $d_{s,t} \geq 0$ towards another end-host $e_t \in A_j$. We also deviate from the Wardrop model by considering a multi-path setting, where the demand $d_{s,t}$ of one agent can be arbitrarily distributed over all paths $p \in P(A_i, A_j)$. The amount of flow from end-host e_s to end-host e_t on path $p \in P(A_i, A_j)$ is denoted as a *path flow* $F_{(s,t),p}$, which must be non-negative, with $\sum_{p \in P(A_i, A_j)} F_{(s,t),p} = d_{s,t}$. The set $\Pi(e_s, e_t) \subseteq \Pi$ contains all end-host paths of the form $\pi = [(s, t), p]$, where e_s, e_t are end-hosts connected by the AS-level path p . All path flows $F_{(s,t),p}$ for an origin-destination pair (e_s, e_t) are collected in a

Table 2.1: Notation used in our model in alphabetic order.

Symbol	Description
$A_i \in A$	Set of autonomous systems (ASes)
$C_\pi(\mathbf{F})$	Cost of path π (based on path-flow pattern \mathbf{F})
$C^*(\mathbf{F})$	Social cost to end-hosts
$C_{(e)}^*(\mathbf{F})$	Total cost to end-host e
$C^\#(\mathbf{F})$	Social cost to network operators
$c_\ell(f_\ell)$	Cost of link ℓ (based on link flow f_ℓ)
$d_{s,t}$	Demand of end-host e_s towards end-host e_t
\mathbf{F}	Path-flow pattern (Demand distribution onto paths)
\mathbf{F}^*	Social optimum for end-hosts
$\mathbf{F}^\#$	Social optimum for network operators
\mathbf{F}^0	Nash equilibrium under the LI assumption
\mathbf{F}^+	Nash equilibrium under the PI assumption
F_π	Flow on path π
f_ℓ	Flow on link ℓ
$\ell \in L$	Set of inter-AS links
$od = (e_s, e_t) \in OD$	Set of end-host origin-destination pairs
PoA^{*0}	Price of Anarchy to end-hosts under the LI assumption
PoA^{*+}	Price of Anarchy to end-hosts under the PI assumption
$PoA^{\#0}$	Price of Anarchy to network operators under the LI assumption
$PoA^{\#+}$	Price of Anarchy to network operators under the PI assumption
$p \in P(A_i, A_j)$	Set of AS-level paths between ASes A_i and A_j
$\pi = [(s, t), p] \in \Pi$	Set of paths between end-hosts
VoI^*	Value of Information to end-hosts
$VoI^\#$	Value of Information to network operators

path-flow vector $\mathbf{F}_{s,t} \in \mathbb{R}^{|\Pi(e_s, e_t)|}$. All such path-flow vectors $\mathbf{F}_{s,t}$ are collected in the global *path-flow pattern* $\mathbf{F} \in \mathbb{R}^{|\Pi|}$. A link flow f_ℓ for link $\ell \in L$ is the sum of the path flows in \mathbf{F} that refer to end-host paths π containing link ℓ , i.e., $f_\ell = \sum_{\pi \in \Pi: \ell \in \pi} F_\pi$.

Cost functions. The cost C_π of an end-host path π given a certain path-flow pattern \mathbf{F} is the sum of the cost of all links in the path: $C_\pi(\mathbf{F}) = \sum_{\ell \in \pi} c_\ell(f_\ell)$. The cost to end-hosts $C^*(\mathbf{F})$ from a path-flow pattern \mathbf{F} is the latency experienced by all end-hosts on all the paths to all of their destinations, weighted by the amount of traffic that goes over a given path. This term can be simplified as follows:

$$C^*(\mathbf{F}) = \sum_{(s,t) \in OD} \sum_{\pi \in \Pi(s,t)} F_\pi \cdot C_\pi(\mathbf{F}) = \sum_{\pi \in \Pi} F_\pi \cdot \sum_{\ell \in \pi} c_\ell(f_\ell) = \sum_{\ell \in L} f_\ell \cdot c_\ell(f_\ell) \quad (2.1)$$

Existing work on selfish routing [222, 224] usually defines total cost in the above sense. However, when analyzing end-host path selection architectures, the network-operator perspective on social cost is essential. Therefore, we also introduce a social cost function relating to the perspective of network operators.

The basic idea of the network-operator cost function $C^\#$ is to treat links as investment assets. Thus, the business performance of a link ℓ is given by a function $p_\ell^\#(f_\ell) = b_\ell^\#(f_\ell) - c_\ell^\#(f_\ell)$, where $b_\ell^\#$ and $c_\ell^\#$ are the benefits and costs of a link, respectively. As we investigate effects on the aggregate of network operators, we model the network-operator cost function as follows:

$$C^\#(\mathbf{F}) = \sum_{\ell \in L} -p_\ell^\#(f_\ell) = \sum_{\ell \in L} c_\ell^\#(f_\ell) - b_\ell^\#(f_\ell) = \sum_{\ell \in L} c_\ell(f_\ell) \quad (2.2)$$

We justify this formulation as follows. Concerning link costs $c_\ell^\#$, a central insight is that network-operator costs mostly stem from heavily used links. In volume-based interconnection agreements, excessive usage of a link induces high charges, whereas in peering agreements, excessive usage violates the agreement and triggers expensive renegotiation. Moreover, heavy usage necessitates expensive capacity upgrades. As the latency function $c_\ell(f_\ell)$ indicates the congestion level on link ℓ , we approximate $c_\ell^\# \approx c_\ell$. The link benefit $b_\ell^\#$ captures the link revenue, both revenue from customer ASes and customer end-hosts. In the aggregate, the monetary transfers between ASes (charges paid and received) sum up to zero. Given a fixed market size, the revenue from end-hosts sums up to a constant in the aggregate. Hence, the global benefit $\sum_{\ell \in L} b_\ell^\#$ is constant and can be dropped, as the absolute level of the network-operator cost is irrelevant for our purposes. This convex formulation of $C^\#$ allows theoretical analysis.

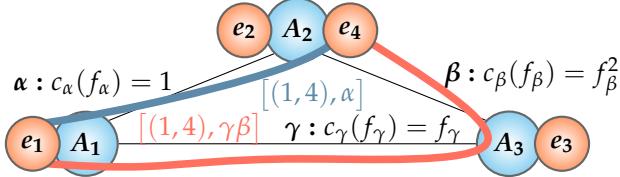


Figure 2.2: Example network illustrating the end-host path selection model.

2.2.2 Social Optima

According to Wardrop [67, 253], a socially optimal traffic distribution is reached if and only if the total cost cannot be reduced by moving traffic from one path to another. In the optimum, the cost increase on an additionally loaded path at least outweighs the cost reduction from a relieved path. Because the cost functions are convex and non-decreasing, it suffices that this condition holds for an infinitesimal traffic share. Adding an infinitesimal amount to the argument of a cost function imposes a *marginal cost*, given by the derivative of the cost function. A socially optimal traffic distribution is thus reached if and only if the marginal cost of every alternative path is not smaller than the marginal cost of the currently used paths [67]:

Definition 2.1 Social Optimum. A path-flow pattern \mathbf{F} represents a social optimum w.r.t. cost function C if and only if for every origin-destination pair $od \in OD$, the paths $\pi_1, \dots, \pi_i, \pi_{i+1}, \dots, \pi_{|\Pi(od)|} \in \Pi(od)$ stand in the following relationship:

$$\frac{\partial}{\partial F_{\pi_1}} C(\mathbf{F}) = \dots = \frac{\partial}{\partial F_{\pi_i}} C(\mathbf{F}) \leq \frac{\partial}{\partial F_{\pi_{i+1}}} C(\mathbf{F}) \leq \dots \leq \frac{\partial}{\partial F_{\pi_{|\Pi(od)|}}} C(\mathbf{F}) \quad (2.3)$$

$$F_\pi > 0 \quad \text{for } \pi = \pi_1, \dots, \pi_i, \quad F_\pi = 0 \quad \text{for } \pi = \pi_{i+1}, \dots, \pi_{|\Pi(od)|}.$$

In this work, we refine the conventional notion of the social optimum by distinguishing two different perspectives on social cost: The *end-host optimum* \mathbf{F}^* satisfies the above conditions with respect to the function C^* , whereas the *network-operator optimum* $\mathbf{F}^\#$ satisfies the above conditions with respect to function $C^\#$.

Interestingly, the end-host optimum \mathbf{F}^* and the network-operator optimum $\mathbf{F}^\#$ can differ substantially. Assume that end-host e_1 in Figure 2.2 has a demand of $d_{1,4} = 1$ towards end-host e_4 and that there is no other traffic in the network. The network-operator cost function $C^\#(\mathbf{F})$ is $1 + F_\beta^2 + F_\gamma \beta$ and is minimized by

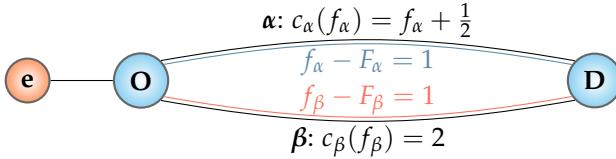


Figure 2.3: Example illustrating the different degrees of end-host information.

$\mathbf{F}^{\#} = (1, 0)^{\top}$, i.e., by sending all traffic over link α . In contrast, the end-host cost function is $F_{\alpha} + F_{\gamma\beta}^3 + F_{\gamma\beta}^2$ and is minimized by $\mathbf{F}^* = (2/3, 1/3)^{\top}$, i.e., by sending two thirds of traffic over link α and the remaining third over path $\gamma\beta$.

2.2.3 Degrees of Information

In this chapter, we consider the following two assumptions on the network information possessed by end-hosts:

- *Latency-only information (LI)*: End-hosts know the latency of every path to a destination.
- *Perfect information (PI)*: End-hosts know not only the latency of different paths, but also how the latency of the network links depends on the current load, i.e., the *latency functions*. Moreover, the end-hosts know the current link utilization, i.e., the background traffic.

The LI assumption hence reflects a scenario where end-hosts have to rely solely on latency measurements of paths, i.e., through RTT measurements from their own device. The LI assumption is the standard model traditionally considered in the selfish routing literature [88, 148, 224].

In this work, we extend the standard model by introducing the concept of *perfect information (PI)*. The PI assumption reflects a scenario where end-hosts can always take the *best* traffic-allocation decision in selfish terms. More specifically, the PI assumption allows end-hosts to compute the *marginal cost* of a path. In path-aware networking, supplying end-hosts with perfect information is possible, as such information is known by network operators and can be disseminated along with path information.

Figure 2.3 illustrates the difference between the LI assumption and the PI assumption. Assume that end-host e , residing in AS O , has a demand of $d = 1$ to a destination in AS D . End-host e can split its traffic between two paths α and β , both

consisting of a single link with the cost functions c_α (linear) and c_β (constant). The background traffic (traffic not from end-host e) is 1 on both paths. Assuming the traffic allocation of end-host e is $(F_\alpha, F_\beta) = (0.5, 0.5)$, the path-latency values are given by $c_\alpha(0.5 + 1) = 2$ and $c_\beta(0.5 + 1) = 2$. Given the LI assumption, end-host e performs no traffic reallocation, as there is no lower-cost alternative path which traffic could be shifted to. Moreover, there is no method for predicting the path costs for a different traffic allocation. However, such a prediction is possible with perfect information (PI): under the PI assumption, end-host e knows the cost functions and the background traffic such that it can optimize the objective

$$\begin{aligned} C_{(e)}^*(F_\alpha) &= F_\alpha \cdot c_\alpha(F_\alpha + 1) + (d - F_\alpha) \cdot c_\beta(d - F_\alpha + 1) \\ &= F_\alpha \cdot (F_\alpha + 1 + \frac{1}{2}) + (1 - F_\alpha) \cdot 2. \end{aligned} \tag{2.4}$$

As a result, end-host e discovers the optimal traffic assignment $(0.25, 0.75)$. Intriguingly, the more detailed perfect information (PI) enables end-host e to detect an optimization that it cannot directly observe with latency values only (LI).

2.2.4 Nash Equilibria

In general, uncoordinated actions of selfish end-hosts do not result in socially optimal traffic allocations. Instead, the only stable states that arise in selfish path selection are *Nash equilibria*, i.e., situations in which no end-host perceives an opportunity to reduce its selfish cost by unilaterally reallocating traffic. However, as shown in §2.2.3, the degree of available information (LI or PI) strongly influences the optimization opportunities that an end-host perceives. Therefore, different information assumptions induce different types of Nash equilibria:

LI equilibrium. An end-host restricted to latency measurements will shift traffic from high-cost paths to low-cost paths whenever there is a cost discrepancy between paths, and will stop reallocating traffic whenever there is no lower-cost path anymore which the traffic could be shifted to. In the latter situation, an end-host under the LI assumption cannot perceive any way of reducing its selfish cost. We thus define a Nash equilibrium under the LI assumption (short: LI equilibrium):

Definition 2.2 LI Equilibrium. A path-flow pattern \mathbf{F} represents an LI equilibrium \mathbf{F}^0 if and only if for every origin-destination pair $od \in OD$, the paths $\pi_1, \dots, \pi_i, \pi_{i+1}, \dots, \pi_{|\Pi(od)|} \in \Pi(od)$ have the following relationship:

$$C_{\pi_1}(\mathbf{F}) = \dots = C_{\pi_i}(\mathbf{F}) \leq C_{\pi_{i+1}}(\mathbf{F}) \leq \dots \leq C_{\pi_{|\Pi(od)|}}(\mathbf{F}) \quad (2.5)$$

$$F_\pi > 0 \quad \text{for } \pi = \pi_1, \dots, \pi_i \quad F_\pi = 0 \quad \text{for } \pi = \pi_{i+1}, \dots, \pi_{|\Pi(od)|}$$

Traditionally, the selfish-routing literature [88, 221, 224] considers a Nash equilibrium in the sense of the LI equilibrium, namely an equilibrium defined by the cost equality of all used paths to a destination. Under this classical definition, selfish routing is an instance of a *potential game* [227].

PI Equilibrium. We contrast the classical equilibrium (LI equilibrium) with a different equilibrium definition that builds on our new concept of perfect information (PI). As explained in §2.2.3, the PI assumption states that end-hosts do not only possess cost information of available paths to a destination, but are informed about the cost *functions* of all links in the available paths, as well as the background traffic on these links, i.e., the arguments to the cost functions. An end-host can thus calculate the selfish cost of a specific traffic reallocation and find the path-flow pattern that minimizes the end-host's selfish cost.

The selfish cost $C_{(e)}^*(\mathbf{F})$ of end-host e is given by the cost of all paths to all desired destinations, weighted by the amount of flow relevant to end-host e :

$$C_{(e)}^*(\mathbf{F}) = \sum_{\ell \in L} f_{\ell,(e)} \cdot c_\ell(f_\ell) \quad (2.6)$$

where $f_{\ell,(e)}$ is the flow volume on link ℓ for which e is origin or destination.

Similar to the end-host social cost function C^* of which it is a partial term, $C_{(e)}^*$ has a minimum that is characterized by a marginal-cost equality. An equilibrium under the PI assumption is thus given if and only if all end-hosts are at the minimum of their respective selfish cost functions, given the traffic by all other end-hosts:

	LI equilibrium	PI equilibrium
End-host perspective	$PoA^{*0} = \frac{C^*(\mathbf{F}^0)}{C^*(\mathbf{F}^*)}$	$PoA^{*+} = \frac{C^*(\mathbf{F}^+)}{C^*(\mathbf{F}^*)}$
Network-operator perspective	$PoA^{\#0} = \frac{C^\#(\mathbf{F}^0)}{C^\#(\mathbf{F}^\#)}$	$PoA^{\#+} = \frac{C^\#(\mathbf{F}^+)}{C^\#(\mathbf{F}^\#)}$

Table 2.2: Different versions of the Price of Anarchy.

Definition 2.3 PI Equilibrium. A path-flow pattern \mathbf{F} represents a PI equilibrium \mathbf{F}^+ if and only if for every origin-destination pair $od = (e, -) \in OD$, the paths $\pi_1, \dots, \pi_i, \pi_{i+1}, \dots, \pi_P \in \Pi(od)$ stand in the following relationship:

$$\frac{\partial}{\partial F_{\pi_1}} C_{(e)}^*(\mathbf{F}) = \dots = \frac{\partial}{\partial F_{\pi_i}} C_{(e)}^* \leq \frac{\partial}{\partial F_{\pi_{i+1}}} C_{(e)}^* \leq \dots \leq \frac{\partial}{\partial F_{\pi_{|\Pi(od)|}}} C_{(e)}^* \quad (2.7)$$

$$F_\pi > 0 \quad \text{for } \pi = \pi_1, \dots, \pi_i \quad F_\pi = 0 \quad \text{for } \pi = \pi_{i+1}, \dots, \pi_{|\Pi(od)|}$$

2.2.5 Price of Anarchy

A natural way of analyzing the efficiency of selfish routing is to compare the social optima and the equilibria in a network. Typically, such a comparison involves computing the *Price of Anarchy* (*PoA*), i.e., the ratio of the equilibrium cost and the optimal cost. By definition of the optimal cost, this ratio is never lower than 1.

In our model, the classical Price of Anarchy from the existing literature reflects a comparison of the end-host cost C^* of the LI equilibrium \mathbf{F}^0 and the end-host cost C^* of the end-host optimum \mathbf{F}^* . With the additional versions of social optima and equilibria established in the preceding sections, a total of four different variants of the Price of Anarchy are possible, one for each combination of equilibrium (LI or PI) and perspective (end-hosts or network operators). These Prices of Anarchy are presented in Table 2.2.

2.2.6 Value of Information

To compare different equilibria for different information assumptions, we introduce the *Value of Information* (*VoI*). For a given perspective, the Value of Information is

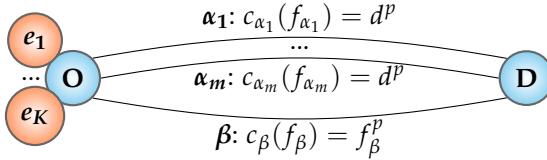


Figure 2.4: Example network with beneficial impact of end-host information.

the difference between the Prices of Anarchy under the LI and PI assumptions, denominated by the Price of Anarchy under the LI assumption:

$$Vol^* = \frac{PoA^{*0} - PoA^{*+}}{PoA^{*0}} \quad Vol^\# = \frac{PoA^{\#0} - PoA^{\#+}}{PoA^{\#0}} \quad (2.8)$$

A positive Value of Information reflects a situation where the equilibrium under the PI assumption is closer to the social optimum than the equilibrium under the LI assumption. We identify and analyze scenarios with a positive impact of information in §2.3. A negative Value of Information reflects the counter-intuitive scenario where additional information makes the equilibrium more costly (cf. §2.4).

2.3 THE BENEFITS OF INFORMATION

In this section, we will show that information is beneficial in the artificial network settings traditionally considered in the literature [221]. More precisely, we show that in this setting, the PI equilibrium induces a lower Price of Anarchy than the LI equilibrium such that the Value of Information is positive. This is intuitive: if end-hosts possess more information, end-host path selection is more efficient.

In the network of Figure 2.4, K end-hosts e_1, \dots, e_K reside in AS O . Each end-host has a demand of d/K towards a destination in AS D . ASes O and D are connected by m links $\alpha_1, \dots, \alpha_m$ with a constant cost function $c_{\alpha_i}(f_{\alpha_i}) = d^p$ and one link β with a load-dependent cost function $c_\beta(f_\beta) = f_\beta^p$, where $p \geq 1$.

Such networks of parallel links are of special importance in the theoretical selfish-routing literature. In particular, Roughgarden [221] proved that the network in Figure 2.4 reveals the worst-case Price of Anarchy for any network with link cost functions limited to polynomials of degree p . The intuition behind this result is that the Price of Anarchy relates to a difference of steepness between cost functions of competing links: the link β allows to reduce the cost of traffic from AS O to AS D if used modestly, but loses its advantage over the links α_i if fully used, i.e., if $f_\beta = d$.

However, in selfish routing, end-hosts will use link β until the link is fully used, as it is always a lower-cost alternative path if not fully used. Therefore, the end-hosts overuse link β compared to the optimum.

Roughgarden's result refers to the classical Price of Anarchy, i.e., the Price of Anarchy PoA^{*0} to end-hosts under the LI assumption. In this section, we will show how this result is affected by additionally introducing the network-operator perspective and the PI assumption. In particular, we prove the following theorem:

Theorem 2.1 Benefits of Information. *In a network of parallel links, a higher degree of information (PI assumption) is always more socially beneficial compared to a lower degree of information (LI assumption), both from the perspective of end-hosts and network operators:*

$$PoA^{*+} \leq PoA^{*0} \quad PoA^{\#+} \leq PoA^{\#0} \quad (2.9)$$

2.3.1 Social Optima

The end-host optimum \mathbf{F}^* has social cost (cf. Appendix A.1)

$$C^*(\mathbf{F}^*) = d^{p+1} \left(1 - \frac{p}{(p+1)^{(p+1)/p}} \right). \quad (2.10)$$

The network-operator optimum $\mathbf{F}^\#$ is simple to derive: Since the cost of the links α_i is independent of the flow on these links in contrast to the cost of link β , any flow on link β increases the cost $C^\#$ to network operators. The minimal cost to network operators is thus simply $C^\#(\mathbf{F}^\#) = m \cdot d^p$.

2.3.2 LI Equilibrium

Under the LI assumption, a network is in equilibrium if for every end-host pair, all used paths have the same cost and all unused paths do not have a lower cost. Applied to the simple network in Figure 2.4, this condition is satisfied if and only if $f_\beta^0 = d$ and $f_{\alpha_i}^0 = 0 \forall f_{\alpha_i}$, implying $c_\beta(f_\beta^0) = d^p = c_{\alpha_i}(f_{\alpha_i}^0)$. The path-flow pattern \mathbf{F}^0 with $F_{(k,D),\beta} = d/K$ and $F_{(k,D),\alpha_i} = 0$ therefore represents the LI equilibrium. The cost C^* of the LI equilibrium \mathbf{F}^0 to end-hosts is simply $C^*(\mathbf{F}^0) = d^{p+1}$. The Price of Anarchy to end-hosts under the LI assumption is thus

$$PoA^{*0} = \frac{C^*(\mathbf{F}^0)}{C^*(\mathbf{F}^*)} = \frac{1}{1 - p/(p+1)^{(p+1)/p}}. \quad (2.11)$$

The cost $C^\#$ of the LI equilibrium \mathbf{F}^0 to network operators is given by $C^\#(\mathbf{F}^0) = d^p + \sum_{\alpha_i} d^p = (m+1) \cdot d^p$. The Price of Anarchy to network operators under the LI assumption is thus

$$PoA^{\#0} = \frac{C^\#(\mathbf{F}^0)}{C^\#(\mathbf{F}^\#)} = \frac{m+1}{m} \stackrel{m=1}{\leq} 2. \quad (2.12)$$

The Price of Anarchy to network operators in networks of parallel links is thus upper-bounded by 2 whereas the Price of Anarchy to end-hosts is unbounded for arbitrary p .

2.3.3 PI Equilibrium

If the end-hosts e_1, \dots, e_K are equipped with perfect information, they are in equilibrium if and only if the *selfish* marginal cost of every path to AS D is the same for every end-host. Under this condition, the cost term C^* of the PI equilibrium \mathbf{F}^+ to end-hosts can be derived to be $C^*(\mathbf{F}^+) = d^{p+1} (1 - (p/K)/(p/K+1)^{(p+1)/p})$ (cf. Appendix A.2). The Price of Anarchy to end-hosts under the PI assumption is

$$PoA^{*+} = \left(1 - \frac{p/K}{(p/K+1)^{(p+1)/p}}\right) \cdot PoA^{*0} \stackrel{K \rightarrow \infty}{\leq} PoA^{*0}. \quad (2.13)$$

The cost $C^\#$ of the PI equilibrium \mathbf{F}^+ to network operators is $C^\#(\mathbf{F}^+) = (m+1/(p/K+1)) \cdot d^p$ and the corresponding Price of Anarchy to network operators is

$$PoA^{\#+} = \frac{m+1/(p/K+1)}{m} \stackrel{K \rightarrow \infty}{\leq} \frac{m+1}{m} = PoA^{\#0}. \quad (2.14)$$

Based on Eqs. (2.13) and (2.14), Theorem 2.1 holds. However, the Prices of Anarchy PoA^{*+} and $PoA^{\#+}$ under the PI assumption are dependent on K , which is the number of end-hosts in the network. If K is very high, as it is in an Internet context, the Prices of Anarchy under the PI assumption approximate the Prices of Anarchy under the LI assumption. Thus, for scenarios of heterogeneous parallel paths to a destination, the benefit provided by perfect information is undone in an Internet context. In fact, the effect of additional information can even turn negative when considering more general networks, as we will show in the next section.

2.4 THE DRAWBACKS OF INFORMATION

We will now show that in more general settings, more information for end-hosts can *deteriorate* outcomes of selfish routing. Such a case is given by the general *ladder*

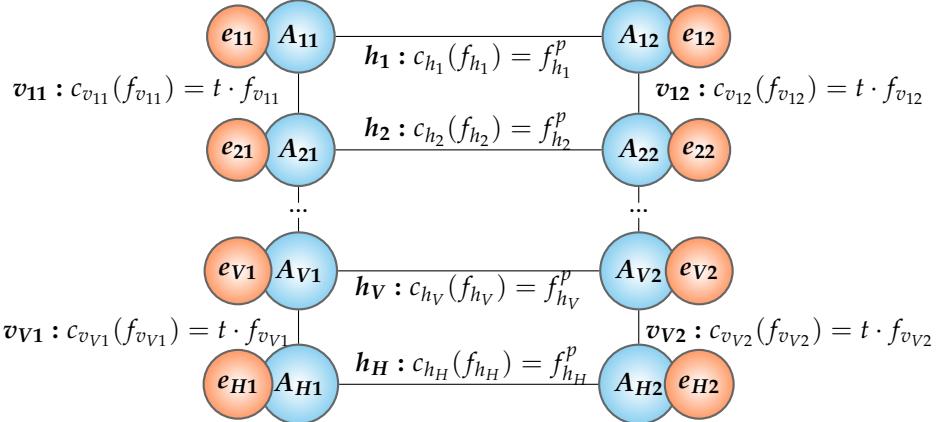


Figure 2.5: Example network with harmful impact of end-host information ($V = H - 1$).

network in Figure 2.5, a natural generalization of the simple topology considered above and a traditional ISP topology [167].

A ladder network of height H contains H horizontal links h_1, \dots, h_H , which represent the rungs of a ladder and have the cost function $c_{h_i}(f_{h_i}) = f_{h_i}^p$. Each horizontal link h_i connects an AS A_{i1} to AS A_{i2} , which accommodate the end-hosts e_{i1} and e_{i2} , respectively. Every end-host e_{i1} has the same demand d towards the corresponding end-host e_{i2} . Neighboring rungs of a ladder are connected by vertical links v_{ij} , $i \in \{1, \dots, V = H - 1\}$, $j \in \{1, 2\}$, where the vertical link v_{ij} connects the ASes A_{ij} and $A_{i+1,j}$ and has the linear cost function $c_{v_{ij}}(f_{v_{ij}}) = t \cdot f_{v_{ij}}$ with $t \geq 0$. We denote a ladder network with this structure and a choice of parameters H, p, d , and t by $\mathcal{L}(H, p, d, t)$.

By comparing optima and equilibria, we will prove the following theorem in the following subsections:

Theorem 2.2 Drawbacks of Information. *For any ladder network $\mathcal{L}(H, p, d, t)$, the Value of Information for both end-hosts and network operators is **negative**, i.e., $VoI^* < 0$ and $VoI^\# < 0$.*

2.4.1 Social Optima

Both the end-host optimum \mathbf{F}^* and $\mathbf{F}^\#$ are equal to the *direct-only* path-flow pattern \mathbf{F}^\sim that is defined as follows: For every end-host e_{i1} , $F_{(i1,i2),h_i}^\sim = d$ and $F_{(i1,i2),q}^\sim = 0$ where q is any other path between A_{i1} and A_{i2} than the direct path over link h_i .

Simple intuition already confirms the optimality of this path-flow pattern. The social cost from the horizontal links is minimized for an equitable distribution of the whole-network demand Hd onto the H horizontal links. In contrast, the cost from vertical links v_{ij} can be minimized to 0 by simply abstaining from using vertical links. In fact, every use of the vertical links is socially wasteful.

More formally, if $f_{h_i} = d$ for $i \in \{1, \dots, H\}$ and $f_{v_{i1}} = f_{v_{i2}} = 0$ for $i \in \{1, \dots, V\}$, the marginal costs of the direct path and every indirect path can be easily shown to equal $(p+1)d^p$, given end-host cost function C^* . Concerning network-operator cost $C^\#$, the direct and indirect paths have marginal costs $p \cdot d^{p-1}$ and $p \cdot d^{p-1} + 2yt \forall y \in \mathbb{N}_{\geq 1}$, respectively. The used direct paths thus do not have a higher marginal cost than the unused indirect paths.

2.4.2 LI Equilibrium

Also the LI equilibrium path-flow pattern \mathbf{F}^0 is equal to the direct-only path-flow pattern \mathbf{F}^\sim . For \mathbf{F}^\sim , the costs of the direct path π and the indirect path are

$$\begin{aligned} C_\pi(\mathbf{F}^\sim) &= F_{(i1,i2),\pi}^p = d^p, \\ C_{\pi'}(\mathbf{F}^\sim) &= f_{h'}^p + \sum_{v \in W_{\pi'}} f_v = d^p + 0 = d^p, \end{aligned} \tag{2.15}$$

where π' contains the remote horizontal link h' and the vertical links $v \in W_{\pi'}$. Thus, the LI equilibrium conditions of cost equality are satisfied by \mathbf{F}^\sim .

As the LI equilibrium is equal to the social optimum both from the end-host perspective and the network-operator perspective, both variants of the Price of Anarchy under the LI assumption are optimal, i.e., $PoA^{*0} = PoA^{\#0} = 1$.

2.4.3 PI Equilibrium

Differently than under the LI assumption, the direct-only flow distribution \mathbf{F}^\sim is not stable under the PI assumption. An end-host e_i can improve its individual cost by allocating some traffic to an indirect path π_k (involving the horizontal link h_k) and interfering with another end-host e_k . This reallocation decision will increase

the social cost for end-hosts and network operators. In particular, the end-host e_k that previously used the link h_k exclusively will see its selfish cost increase. In turn, the harmed end-host e_k will reallocate some of its traffic to an indirect path in order to reduce its selfish cost $C_{(e_k)}$, leading to a process where all end-hosts in the network interfere with each other until they reach a PI equilibrium with a suboptimal social cost for end-hosts and network operators.

Similar to §2.3.3, we use the condition of marginal selfish cost equality in order to derive the Price of Anarchy under the PI assumption for a ladder network with $H = 2$. This derivation, as performed in Appendix A.3, yields the following results for the Price of Anarchy to end-hosts and network operators:

$$PoA_{H=2}^{*+}(p) = 1 + p/12 \quad PoA_{H=2}^{\#+}(p) = 1 + p/3 \quad (2.16)$$

Since the LI equilibrium is optimal and the PI equilibrium is generally suboptimal on the considered ladder networks, Theorem 2.2 holds. This finding is confirmed by a case study of the Abilene network (cf. §2.5), which structurally resembles a ladder topology. The case study also reveals that the negative impact of information is especially pronounced if path diversity is high.

Interestingly, there is an upper bound of the Price of Anarchy to network operators for a *general* ladder network. This bound is given by the following theorem and proven in Appendix A.4:

Theorem 2.3 Bounded Price of Anarchy. *For every ladder network $\mathcal{L}(H, p, d, t)$, the Price of Anarchy $PoA^{\#+}$ to network operators is lower than the following upper bound $PoA_{\max}^{\#+}$:*

$$PoA^{\#+} \leq PoA_{H,\max}^{\#+} = 1 + \frac{2(H-1)}{3H} p \leq PoA_{\max}^{\#+} = 1 + \frac{2}{3} p \quad (2.17)$$

2.5 CASE STUDY: ABILENE NETWORK

To verify and complement our theoretical insights, we conduct a case study with a real network: we consider the well-known Abilene network, for which topology and workload data is publicly available [145, 146]. We accommodate the Abilene topology into our model as follows. For the demand d between the 11 points-of-presence, which we consider ASes, we rely on the empirical traffic matrix from the dataset. Concerning the link-cost functions c_ℓ , we model the latency behavior of a link by a function $c_\ell(f_\ell) = f_\ell^2 + \delta_\ell$, where f_ℓ^2 captures the queueing delay and δ_ℓ

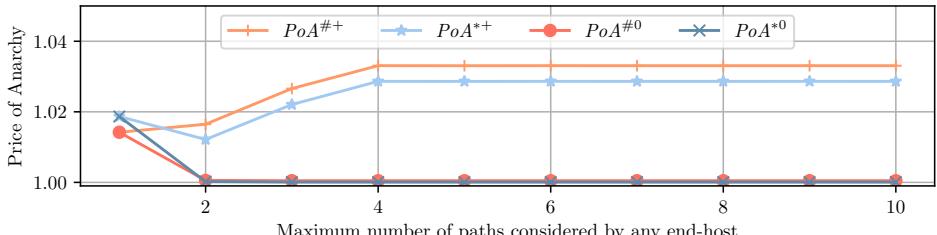


Figure 2.6: Abilene network results.

is a constant quantity depending on the geographical distance between the two end-points of link ℓ , approximating the link's propagation delay.

In order to study the effect of both end-host information and multi-path routing on the Price of Anarchy, we perform the following simulation experiment. First, we compute the social optima F^* and $F^{\#}$ for the Abilene network. Second, we simulate the convergence to the Nash equilibria F^0 and F^+ for different degrees of multi-path routing, represented by the maximum number of shortest paths that end-hosts consider in their path selection. Once converged, we compute the social cost of the equilibrium traffic distributions and the corresponding Prices of Anarchy.

The experiment results in Figure 2.6 offer multiple interesting insights. Most prominently, if simple shortest-path routing represents the baseline of network-controlled path selection, end-host path selection with latency-only information improves the performance of the network (up to a near-optimum), which confirms findings of prior work [210]. In contrast, path selection with perfect information deteriorates performance, especially for a higher degree of multi-path routing. Therefore, the potential performance benefits of end-host path selection with multi-path routing are conditional on the amount of information possessed by end-hosts, where a higher degree of information is associated with lower performance. However, while an increasing degree of multi-path routing is associated with worse performance under perfect information, the resulting inefficiency is bounded at a modest level of less than 4 percent for both end-hosts and network operators. The near-optimality of latency-only information in terms of performance and the bounded character of the Price of Anarchy under perfect information reflect the findings from §2.4 about ladder topologies, which resemble the Abilene topology. Thus, the experiment results not only show that end-host path selection can be a means to improve the performance of a network but also confirm the practical relevance of our theoretical findings.

2.6 RELATED WORK

Selfish routing. Inefficiency arising from selfish behavior in networks is well-known to exist in transportation networks and has been thoroughly analyzed with the framework of the Wardrop model [67, 253]. The most salient expressions of this inefficiency is given by the Braess Paradox [38].

Literature on selfish routing is often concerned with the discrepancy between optimum and Nash equilibrium: the *Price of Anarchy* [76, 148]. The Price of Anarchy was initially studied for network models (see Nissan et al. [197] for an overview), but literature now covers a wide spectrum, from health care to basketball [223]. Our work has a closer connection to more traditional research questions, such as bounds on the Price of Anarchy for selfish routing. An early result has been obtained by Koutsoupias and Papadimitriou [148], who formulated routing in a network of parallel links as a multi-agent multi-machine scheduling problem.

A different model has been developed by Roughgarden and Tardos [224] who build on the Wardrop model [253] for routing in the context of computer networks. The Price of Anarchy in the proposed routing game is the ratio between the latency experienced by all users in the Wardrop equilibrium and the minimum latency experienced by all users. For different classes of latency functions, the authors derive explicit high bounds on the resulting Price of Anarchy. In a different work, they show that the worst-case Price of Anarchy for a function class can always be revealed by a simple network of parallel links and that the upper bound on the Price of Anarchy depends on the growth rate of the latency functions [221].

The relatively loose upper bounds on the Price of Anarchy of previous works [148, 224] have been qualified by subsequent research. It was found that problem instances with high Prices of Anarchy are usually artificial. By introducing plausible assumptions to make the routing model more realistic, upper bounds on the Price of Anarchy can be reduced substantially. For instance, Friedman [88] shows that the Price of Anarchy is lower than the mentioned worst-case derived by Roughgarden and Tardos [224] if the Nash equilibrium cost is not sensitive to changes in the demand of agents. By computing the Price of Anarchy for a variety of different latency functions, topologies, and demand vectors, Qiu et al. even show that selfish routing is nearly optimal in many cases [210].

Convergence. Convergence to Nash equilibria has been studied in the context of congestion games [218] and, in a more abstract form, in the context of potential games [185, 227]. Sandholm [227] showed that selfish player behavior in potential games leads to convergence to the Nash equilibrium and, under some conditions,

even to convergence to the social optimum. As the question of equilibrium convergence is traditionally studied separately from the question of equilibrium cost, we address convergence issues in Chapter 2.

Incomplete information. The study of the effect of incomplete information also has a long tradition [113], but still poses significant challenges [223]. Existing literature in this area primarily focuses on scenarios where players are uncertain about each others' payoffs, studying alternative notions of equilibria such as Bayes-Nash equilibria [237], which also leads to alternative definitions of the price of anarchy such as the Bayes-Nash Price of Anarchy [160, 223] or the price of stochastic anarchy [61]. A common observation of many papers in this area is that less information can lead to significantly worse equilibria [223]. There is also literature on the impact on the Price of Anarchy in scenarios where interacting players only have *local* information, e.g., the evolutionary price of anarchy [230].

However, much less is known today about the role of information in games related to *routing*. In this context, one line of existing literature is concerned with the recentness of latency information. Most prominently, research on the damage done by stale information in load-balancing problems [68, 184] has been applied to routing games by Fischer and Vöcking [85]. This work investigates whether and how rerouting decisions converge onto a Wardrop equilibrium if these rerouting decisions are based on obsolete latency information. Other recent work about the role of information in routing games investigates how the amount of topology information possessed by agents affects the equilibrium cost [14].

Differentiation of our work. Existing work on the subject of path-selection efficiency differs from our work in two important aspects. First, to the best of our knowledge, all existing work on the subject defines the social optimum as the traffic assignment that minimizes the total cost experienced by users, which is indeed a reasonable metric. However, our work additionally investigates the total cost experienced by *links*, i.e., the network operators. Since cost considerations by network operators are a decisive factor in the deployment of architectures offering end-host path selection, the Price of Anarchy to network operators is an essential metric. Second, although existing work on the topic has investigated the role played by the *recentness* of congestion information or the degree of *topology* information, it does not investigate the role played by the *degree of congestion information* that agents possess. Indeed, a major contribution of our work is to highlight the effects of perfect information, i.e., information that allows agents to perfectly minimize their selfish cost. Latency-only

information, which agents are assumed to have in existing work, does not enable agents to perform perfect optimization.

2.7 CONCLUSION

Motivated by the emerging paradigm of end-host path selection, we refine and extend the Wardrop model in order to study the implications of end-host path selection. Our analysis provides several interesting insights with practical relevance. First, the cost of selfish routing to network operators differs from the cost experienced by users. Since network operators are central players in the adoption of path-aware networking, research on the effects of selfish routing thus needs to address the network-operator perspective separately. However, we proved upper bounds on the Price of Anarchy which suggest that selfish routing imposes a low cost on network operators. Second, we found that basic latency information, which can be measured by the end-hosts themselves, leads to near-optimal traffic allocations in many cases. Selfish routing thus causes modest inefficiency even if end-hosts have only imperfect path information and network operators do not disseminate detailed path-load information.

3

INCENTIVES IN END-HOST PATH SELECTION

3.1 INTRODUCTION

End-host path control in inter-domain networking empowers end-hosts to avoid congested links, i.e., enables *load-adaptive path selection*, and therefore has the potential to cause a relatively even traffic distribution. However, load-adaptive path selection creates new challenges, including the potential of sub-optimal traffic distributions (discussed in the preceding chapter), but also the introduction of instability. Instability due to load-adaptive path selection typically appears in the form of *oscillation*, i.e., periodic up- and downswings of link utilization, leading to a large variance of the traffic load in a short time span. This oscillation represents a central obstacle to the deployment of path-aware network architectures according to a recent RFC by the IETF [72]. Indeed, such oscillation can be shown to occur if path-selection decisions are taken on the basis of outdated (stale) load information [85, 233], which is the case in any real system.

Such oscillation is undesirable for many reasons, both from the perspective of the end-hosts and the perspective of the network operator. From the perspective of end-hosts, oscillation is associated with unpredictable performance: If oscillation occurs when a link is near its capacity limit, there is a danger of queue build-up, jitter, and even loss. This unpredictable performance has especially severe consequences on throughput if packet loss forces the congestion-control algorithms to recurring restarts. From the perspective of network operators, oscillation leads to an inefficient utilization of network resources: Due to the large variance of the load level over time, network operators have to perform substantial overprovisioning of link capacities, which is undesirable from a business perspective. Moreover, oscillation of inter-domain traffic imposes additional overhead for intra-domain traffic engineering (e.g., MPLS circuit setup), as oscillating inter-domain flows may constantly switch between inter-AS interfaces.

To avoid these damaging effects, researchers have devised numerous schemes that aim to guarantee stability of load-adaptive path selection. However, to the best of our knowledge, no scheme so far has aimed at providing stability for end-host path control in the *inter-domain* context. Many systems have been designed under the assumption of network-based path selection, i.e., hop-by-hop forwarding according to decisions taken by intermediate routers [84, 100, 152, 177]. These systems achieve convergence by appropriately adjusting how much traffic is forwarded to each next hop towards a destination. Therefore, these systems cannot be used if packets must be sent along paths selected by end-hosts. Other systems allow end-point path selection, but are targeted to an intra-domain context where the end-points (typically ingress and egress routers) are under the control of a network operator [79, 85, 136, 137, 140, 196]. In an intra-domain context, network operators are able to prescribe arbitrary path-selection policies that generate stability. Conversely, in an inter-domain context, the end-points are not under control of network operators and can thus not be forced to adopt a non-oscillatory path-selection strategy. Instead, as end-hosts must be assumed to be selfish, they can only be expected to adopt path-selection strategies that optimize performance *from their individual perspective*.

By performing a game-theoretic analysis, we show in this chapter that the non-oscillatory path-selection strategies traditionally proposed in the literature on stable end-host path selection [79, 85, 136, 137, 140, 196] are incompatible with the self-interest of end-hosts: Assuming that such non-oscillatory path-selection strategies are universally adopted, an end-host can increase its utility by deviating in favor of a strategy that is oscillatory (§3.3). Therefore, stability of load-adaptive path selection in an inter-domain context cannot be achieved by relying only on the rationality of path-selecting end-hosts. Instead, network operators have to *incentivize* end-hosts to adopt a convergent path-selection strategies with *stabilization mechanisms*. These mechanisms have to be *incentive-compatible*, i.e., the mechanisms must create an incentive structure such that adopting a non-oscillatory path-selection strategy is in an end-host's self-interest (§3.4). In this chapter, we present two such stabilization mechanisms, FLOSS and CROSS, and formally prove their incentive compatibility. These mechanisms employ different techniques to disincentivize oscillatory switching between paths, namely limiting the migration rate between paths (FLOSS, §3.5) and imposing a cost on switching between paths (CROSS, §3.6). To complement our mainly theoretical work, we also discuss how our findings could be practically applied (§3.7).

3.1.1 Contribution

This chapter is based on my publication at PERFORMANCE 2020 [2]. In the chapter, we revisit the theoretical study of the dynamic effects of end-point path selection, for the first time focusing the analysis on inter-domain networks where the end-points are selfish and uncontrolled. We present a game-theoretic model that allows us to investigate which path-selection strategies will be adopted by selfish end-hosts. In particular, we introduce the notion of equilibria to path-selection strategies (PSS equilibria). Moreover, we formally show that the non-oscillatory path-selection strategies proposed in the existing literature do not form such PSS equilibria. Thus, we provide evidence towards the hypothesis that stability in load-adaptive path selection over multiple domains cannot be achieved by exclusively relying on end-hosts' path-selection behavior. To remedy this problem, we leverage insights from mechanism design to devise two incentive-compatible stabilization mechanisms enforced by network operators. While these mechanisms build on existing insights from intra-domain traffic engineering, their methods of incentivization represent a novel approach to achieve stability in inter-domain networks with load-adaptive path selection. We formally prove the incentive compatibility of both mechanisms and discuss their practical application.

3.2 OSCILLATION MODEL

3.2.1 Model Overview

Wardrop model. In order to study oscillation in network architectures with end-host path selection, we build on the well-established Wardrop model [253], which is the standard model for studying the interactions of selfish agents in computer networks [210, 221, 224]. In the Wardrop model, an infinite number of end-hosts, each controlling an infinitesimal traffic share, select one path π among multiple paths Π between two network nodes. Every path π has a load-dependent cost, where the path-cost function c_π is typically interpreted as latency. The end-hosts' path-selection decisions form a congestion game, where the path-selection decisions of end-hosts both determine and follow the load f_π on every path π [59, 121, 218].

Dynamicity. In this work, we analyze congestion games with a temporal component, i.e., end-hosts take path-selection decisions over time based on currently available information. More precisely, an end-host performs an average of $r > 0$ re-evaluations per unit of time. The aggregate re-evaluation behavior is uniform over

Table 3.1: Notation used in our model in alphabetic order.

Symbol	Description
A	Oscillation amplitude
$\mathbf{A}_0 \in [0, 1]^{ \Pi }$	Initial load imbalance across paths Π
$C(\sigma O)$	Cost of applying path-selection strategy σ in parallel-path system O
$C(\sigma, t)$	Cost of applying path-selection strategy σ at time t
$c(\sigma, t \pi')$	Cost of applying path-selection strategy σ at time t while on path π'
$c_{\mathcal{M}}(\pi, t)$	Cost imposed by mechanism \mathcal{M} for using path π at time t
$c_{\pi}(f_{\pi})$	Cost of path π (based on path load f_{π})
$c_u(\pi, t)$	Cost incurred by choosing path π at time t
$f_{\pi}(t)$	Load on path π at time t
p	Steepness of path-cost function $c_{\pi}(f_{\pi}) = f_{\pi}^p$
$\pi \in \Pi$	Set of paths
$\tilde{\pi}$	Alternative path to path π (in a two-path system)
R	Expected time between path re-evaluations
r	Number of path re-evaluations per time unit
σ_a	Anticipating path-selection strategy (Always switch to costlier path)
σ_c	Converging path-selection strategy (according to Fischer [85])
σ_g	Greedy path-selection strategy (Always switch to cheaper path)
T	Delay of path-load information before reaching agents
$u(\pi, t \pi')$	Probability of selecting path π at time t while on path π'
$v(\sigma)$	Share of agents adopting path-selection strategy σ
W	Time until migration direction is reversed in oscillation
$y(\pi, t \sigma)$	Probability of being on path π at time t given path-selection strategy σ

time, i.e., when dividing time into intervals of length $\epsilon \in (0, 1]$, $r\epsilon$ re-evaluations are performed in any interval.

Parallel-path systems. Whenever an end-host performs a re-evaluation, it chooses one path π to its destination according to a freely chosen path-selection strategy σ . We thus formalize the environment of congestion games as parallel-path systems:

Definition 3.1 Parallel-Path Systems. A parallel-path system O is characterized by a tuple (Π, r, p, T, A_0, v) , where:

- a set of parallel paths Π jointly accommodates a total demand normalized to 1;
- $r > 0$ is the average number of re-evaluations per end-host and unit of time;
- $p \geq 1$ is the steepness of the path cost as a function of the load (i.e., $c_\pi = f_\pi^p$);
- $T \geq 0$ is the average time that it takes for cost information to reach the agents;
- $A_0 \in [0, 1]^{|\Pi|}$ is the initial load matrix, where the entry $A_{0\pi} = f_\pi(0)$; and
- v is the strategy profile, defining for every available path-selection strategy σ the share $v(\sigma) \in [0, 1]$ of end-hosts that permanently apply strategy σ .

Wardrop equilibria. Every congestion game possesses at least one Wardrop equilibrium $\mathbf{f}^* = \{f_\pi^*\}_{\pi \in \Pi}$, consisting of a traffic distribution where no single agent can reduce its cost by selecting an alternative path [218]. Clearly, this Wardrop equilibrium is achieved at equal load for the parallel-path systems described above, i.e., for all $\pi \in \Pi$, the Wardrop equilibrium path load f_π^* is $1/|\Pi|$ in our setting.

Stability and oscillation. If a Wardrop equilibrium \mathbf{f}^* is converged upon by the end-hosts, we consider a parallel-path system stable at equal load:

Definition 3.2 Stability at Equal Load. A system O is stable at equal load if

$$\Delta^* = \lim_{t \rightarrow \infty} \Delta(t) = 0 \quad \text{where} \quad \Delta(t) = \sum_{\pi \in \Pi} \left| f_\pi - \frac{1}{|\Pi|} \right|. \quad (3.1)$$

Notably, a parallel-path system can also be stable at *unequal* load if the end-hosts converge to an unequal load distribution, i.e., $\Delta^* \in \mathbb{R}$ and $\Delta^* \neq 0$. However, this type of stability requires the adoption of irrational path-selection strategies by at least some end-hosts, as the imbalance is only preserved if some end-hosts do not take advantage of the consistently cheaper path.

More interestingly, the rationality of strategy profiles is not obvious if the system is not stable at all:

Definition 3.3 *Oscillation.* A parallel-path system O experiences oscillation if there exists no limit $\Delta^* \in \mathbb{R}$.

Two-path systems. In this chapter, we study parallel-path systems with two paths α and β (i.e., $|\Pi| = 2$), but our insights directly generalize to more paths. Due to total demand normalization, it holds that $f_\beta(t) = 1 - f_\alpha(t)$ for all $t \geq 0$. Thus, the unique Wardrop equilibrium in a two-path oscillation-prone system is given by $f_\alpha^* = f_\beta^* = 1/2$. Moreover, we assume w.l.o.g. that the initial imbalance A_0 exists with the higher load on path α : $f_\alpha(0) = A_0 = \mathbf{A}_{0\alpha} > 1/2$. For this system of two parallel paths, $\tilde{\pi}$ denotes the alternative path to path π , i.e., $\tilde{\alpha} = \beta$ and $\tilde{\beta} = \alpha$.

3.2.2 Path-Selection Strategies

Strategy form. In a congestion game, end-hosts select paths according to freely adopted path-selection strategies. In order to enable a theoretical treatment, we follow Fischer and Vöcking [85] in assuming that path-selection strategies are memory-less, i.e., not dependent on anything else than currently observable information. Therefore, any path-selection strategy σ can be fully characterized by two elements, $\sigma = (R, u)$, which we will describe in the following.

Re-evaluation periods. Every strategy is characterized by the expected time R between re-evaluations of an end-host. The expected re-evaluation period R reflects the reallocation behavior of end-hosts that non-deterministically re-evaluate the costs of path options, decide for one option based on the perceived costs, and keep sending on the selected path until the next re-evaluation is due. The expected re-evaluation period R has to be in accordance with the parameter r of the parallel-path system, which describes the average number of re-evaluations per end-host and unit of time. Hence, $R = 1/r$.

Path-selection functions. Every strategy σ is based on a path-selection function $u(\pi, t | \pi')$, which gives the probability for selecting path π at time t if the currently used path is π' . Given universal adoption of a strategy σ and re re-evaluations per interval of length ϵ , the number of end-hosts on path π changes by the amount $\Delta_\epsilon f_\pi(t) = -re \cdot u(\tilde{\pi}, t | \pi) \cdot f_\pi(t) + re \cdot u(\pi, t | \tilde{\pi}) \cdot f_{\tilde{\pi}}(t)$ within an interval starting

at time t , given a two-path system. If ϵ is chosen to be infinitesimal, we obtain the rate of change:

$$\frac{\partial f_\pi(t)}{\partial t} = \lim_{\epsilon \rightarrow 0} \frac{\Delta_\epsilon f_\pi(t)}{\epsilon} = -r \cdot u(\tilde{\pi}, t | \pi) \cdot f_\pi(t) + r \cdot u(\pi, t | \tilde{\pi}) \cdot f_{\tilde{\pi}}(t) \quad (3.2)$$

Throughout the rest of the chapter, we describe oscillation dynamics by such differential equations.

Greedy path selection. An example of a path-selection strategy is the greedy path-selection strategy σ_g , which selects the path perceived as cheaper:

$$u_g(\pi, t | \tilde{\pi}) = \begin{cases} 1 & \text{if } c_\pi(t - T) < c_{\tilde{\pi}}(t - T) \\ 0 & \text{otherwise} \end{cases} \quad (3.3)$$

Conversely, the probability of staying on a path is $u_g(\tilde{\pi}, t | \tilde{\pi}) = 1 - u_g(\pi, t | \tilde{\pi})$. At time t , the number of end-hosts on a more expensive path π thus changes with rate $-r \cdot f_\pi(t)$.

3.2.3 Examples of Oscillation and Stability

Whether a parallel-path system in fact experiences oscillation entirely depends on the path-selection strategies adopted by end-hosts. In this section, we present the example of a parallel-path system that experiences oscillation for some path-selection strategy, but converges to stability for a different strategy.

Example of oscillation. For every $T > 0$, oscillation occurs in a system in which all agents adopt a *greedy* path-selection strategy σ_g presented in the previous section. The dynamics of a system with universal adoption of the greedy strategy are given by the partial differential equation:¹

$$\frac{\partial f_\alpha(t)}{\partial t} = \begin{cases} -r \cdot f_\alpha(t) & \text{if } c_\alpha(t - T) > c_\beta(t - T) \\ r \cdot f_\beta(t) & \text{if } c_\alpha(t - T) < c_\beta(t - T) \\ 0 & \text{otherwise} \end{cases} \quad (3.4)$$

¹ An analogous equation holds for f_β .

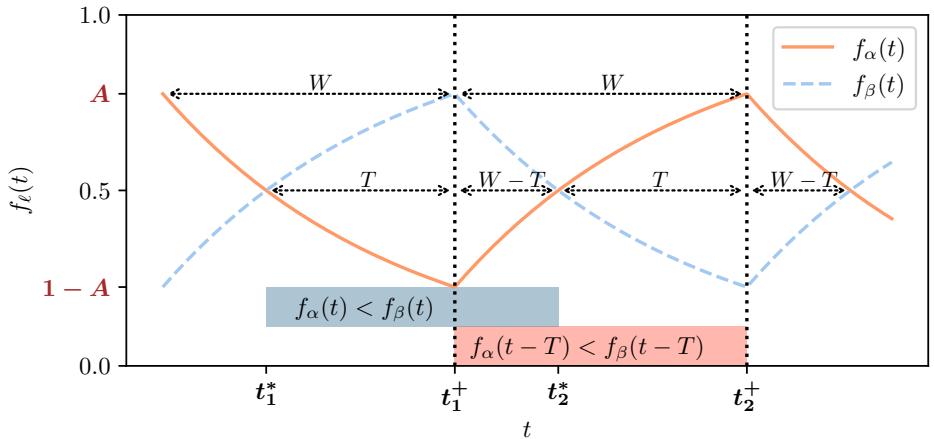


Figure 3.1: Oscillation structure for parallel-path system.

We henceforth refer to *turning points* as all points in time t^+ where $c_\alpha(t^+ - T) = c_\beta(t^+ - T)$, as $f_\alpha(t)$ switches between increasing and decreasing at these moments. In particular, we write $t^+(t)$ for the most recent turning point $t^+ < t$.

Solving the differential equation piece-wise yields a recursive function:²

$$f_\alpha(t) = \begin{cases} e^{-r \cdot (t - t^+(t))} \cdot f_\alpha(t^+(t)) & \text{if } f_\alpha(t - T) \geq \frac{1}{2} \\ 1 - e^{-r \cdot (t - t^+(t))} \cdot f_\beta(t^+(t)) & \text{otherwise} \end{cases} \quad (3.5)$$

Since T is constant, $f_\alpha(t)$ is periodic after the first turning point t_1^+ irrespective of the initial imbalance A_0 . Therefore, the oscillation can be described by the non-recursive function:

$$f_\alpha(t) = \begin{cases} e^{-r \cdot (t - t^+(t))} \cdot A & \text{if } \frac{t^+(t)}{W} \text{ is even,} \\ 1 - e^{-r \cdot (t - t^+(t))} \cdot A & \text{otherwise,} \end{cases} \quad (3.6)$$

where

$$W = \frac{\ln(2e^{rT} - 1)}{r}, \quad A = 1 - \frac{1}{2e^{rT}}, \quad (3.7)$$

and $t^+(t) = t - (t \bmod W)$ is a multiple of W . Figure 3.1 shows an example of $f_\alpha(t)$ for the oscillation-prone system $O = (\Pi = \{\alpha, \beta\}, r = 0.3, p \geq 1, T = 2, A_0 =$

² In the two-path system, $f_\alpha \geq \frac{1}{2}$ is equivalent to $c_\alpha \geq c_\beta$.

$A, v = \{\sigma_g \mapsto 1\}$), where A_0 has been chosen as A in order to skip the irregular starting phase. Figure 3.1 also highlights the time interval during which path α is the cheaper path (in blue, between t_1^* and t_2^*) and the time interval during which path α is perceived to be the cheaper path (in red, between t_1^+ and t_2^+). Clearly, the discrepancy between reality and perception of path costs is the source of oscillation, as the discrepancy leads to increasing load on a path even when it is no longer the cheaper path (i.e., path α between t_2^* and t_2^+). Due to the periodicity of this phenomenon, there exists no limit Δ^* of load difference and the oscillation-prone system experiences oscillation. An interesting observation is that both amplitude (A) and oscillation period ($2W$) increase with the staleness of the information (T); any $T > 0$ leads to oscillations, only $T = 0$ ensures stability.

Example of stability. If the strategy profile v contains different path-selection strategies, a parallel-path system may experience stability. In particular, the parallel-path system above is stable if a sufficient number of end-hosts anticipate the greedy strategy σ_g with an *anticipating* strategy σ_a . An end-host adopting the anticipating strategy always selects the path with the higher perceived cost, speculating that the seemingly cheaper path will soon be overloaded by greedy-strategy players:

$$u_a(\pi, t | \tilde{\pi}) = \begin{cases} 1 & \text{if } c_\pi(t - T) > c_{\tilde{\pi}}(t - T) \\ 0 & \text{otherwise} \end{cases} \quad (3.8)$$

Conversely, $u_a(\tilde{\pi}, t | \tilde{\pi}) = 1 - u_a(\pi, t | \tilde{\pi})$.

In a parallel-path system with strategy profile $v = \{\sigma_g \mapsto q, \sigma_a \mapsto 1 - q\}$ and initial imbalance $A_0 > 1/2$, the initial dynamics of the system are

$$\begin{aligned} \frac{\partial f_\alpha(t)}{\partial t} &= -r \cdot q \cdot f_\alpha(t) + r \cdot (1 - q) \cdot (1 - f_\alpha(t)) \\ f_\alpha(0) = A_0 \implies f_\alpha(t) &= (A_0 + q - 1)e^{-rt} + (1 - q). \end{aligned} \quad (3.9)$$

For $q \leq 1/2$, we see that $f_\alpha(t) > f_\beta(t)$ for all $t \geq 0$, since $\lim_{t \rightarrow \infty} f_\alpha(t) = 1 - q \geq 1/2$, $f_\alpha(0) = A_0 > 1/2$, and $f_\alpha(t)$ is monotonic. Using the definitions from §3.2.1, the oscillation-prone system is *stable at equal load* for $q = 1/2$.

3.2.4 Equilibria on Path-Selection Strategies

In general, Nash equilibria refer to strategy profiles that do not allow for beneficial selfish strategy changes by individual agents. In the context of path-selection

strategies, a Nash equilibrium is thus given if every end-host cannot improve its utility by switching to an alternative path-selection strategy. More formally, a Nash equilibrium on path-selection strategies can be defined as follows:

Definition 3.4 PSS Equilibrium. A strategy profile v^* is a Nash equilibrium on path-selection strategies (PSS equilibrium) in a parallel-path system $O = (\Pi, r, p, T, A_0, v^*)$ if and only if all strategies σ with $v^*(\sigma) > 0$ have cost $C(\sigma | O) = C^*$ and all strategies σ' with $v^*(\sigma') = 0$ have cost $C(\sigma' | O) \geq C^*$.

It remains to formally define the cost $C(\sigma | O)$ of a strategy σ in a parallel-path system O with global strategy profile v . First, we note that a global strategy profile v , together with an initial strategy-adoption distribution for each path, uniquely defines the flow dynamics $f(t) = (f_\alpha(t), f_\beta(t))$ in oscillation-prone systems with two paths. As the flow share controlled by each agent is assumed to be negligible in the Wardrop model, the flow dynamics $f(t)$ are not affected by the choice of σ when varying σ for a single agent. The basic costs of the two path options α and β at any moment t are thus given by $c_\alpha(t)$ and $c_\beta(t)$, both uniquely defined by a parallel-path system $O = (\Pi, r, p, T, A_0, v)$.

Given expected re-evaluation periods of duration R , an end-host deciding for path π at time t incurs the *usage cost*

$$c_u(\pi, t) = \frac{1}{R} \int_t^{t+R} c_\pi(s) \, ds. \quad (3.10)$$

At time t , the cost $c(\sigma, t)$ of applying a strategy σ is

$$c(\sigma, t | \pi') = \sum_{\pi \in \Pi} u_\sigma(\pi, t | \pi') \cdot c_u(\pi, t), \quad (3.11)$$

where π' is the current path of the end-host before the decision at time t , and $u_\sigma(\pi, t | \pi')$ is the probability that strategy σ prescribes the selection of path π at time t given current path π' .

Furthermore, the strategy σ also determines the probability distribution $y(\pi', t | \sigma)$ that defines the probability of the current path being π' at time t . The expected cost for applying a strategy σ at time t is thus given as follows:

$$C(\sigma, t) = \sum_{\pi' \in \Pi} y(\pi', t | \sigma) \cdot c(\sigma, t | \pi') \quad (3.12)$$

The expected cost of applying a strategy σ in general can be derived as the average time-dependent strategy cost during a certain *relevant time span* $[t_0, t_1]$:

$$C(\sigma | O) = \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} C(\sigma, t) \, dt \quad (3.13)$$

For systems that converge to stability at equal load, the relevant time span extends from $t_0 = 0$ until time t_δ when the system has converged according to some criterion $\delta > 0$, i.e., $\forall t > t_\delta. \Delta(t) < \delta$. The time after convergence does not have to be considered as all strategies have the same cost for a system with equal path costs. For periodic oscillating systems, the relevant time span is defined as every interval that contains the periodically repeated sub-function. For an example of a PSS equilibrium analysis, see Appendix B.1.

3.3 LIMITS OF STABLE STRATEGIES

In this section, we investigate whether the stability-inducing path-selection strategies proposed in the literature form PSS equilibria. In this investigation, the central question is whether an end-host can minimize its cost with a stability-oriented strategy if that strategy is universally adopted.

We perform this investigation by means of two case studies. In §3.3.1, we analyze the convergent rerouting policies designed by Fischer and Vöcking [85] and show that such rerouting policies are not compatible with the selfishness of end-hosts. In §3.3.2, we analyze the MATE algorithm [79] and show its equivalence to the rerouting policies discussed in §3.3.1.

3.3.1 Rerouting Policies by Fischer & Vöcking

A typical example of a convergent path-selection strategy has been proposed by Fischer and Vöcking [85]. The proposed path-selection strategy, which we henceforth refer to as the convergent strategy σ_c , works as follows: If an end-host discovers a path with lower cost according to stale information, the end-host switches to that path with a probability that is a linear function of the perceived latency difference. More formally, the probability $u(\pi, t | \tilde{\pi})$ to switch from path π_t to path $\pi \neq \tilde{\pi}$ at time t is:

$$u(\pi, t | \tilde{\pi}) = \begin{cases} \mu \cdot \frac{c_{\tilde{\pi}}(t-T) - c_{\pi}(t-T)}{\Delta_{\max}} & \text{if } c_{\pi}(t-T) < c_{\tilde{\pi}}(t-T), \\ 0 & \text{otherwise,} \end{cases} \quad (3.14)$$

Here, μ is a parameter in $[0, 1]$ and the latency difference is normalized by Δ_{\max} , which is 1 in parallel-path systems as defined in §3.2.1. The dynamics of a two-path

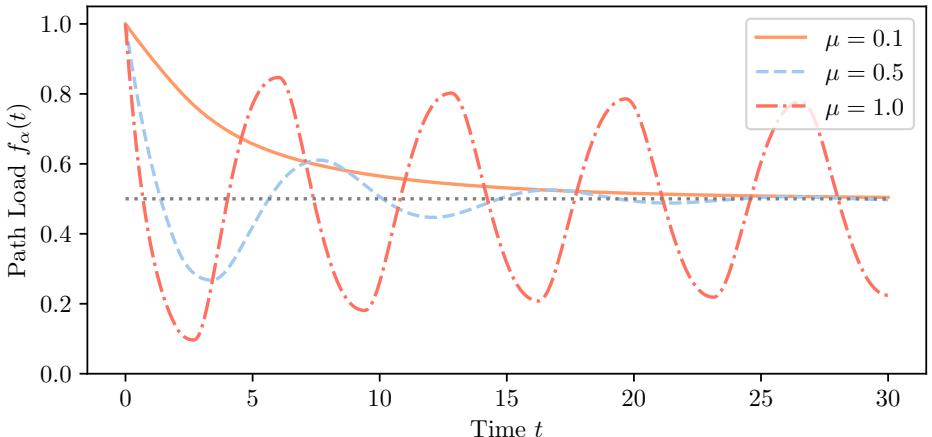


Figure 3.2: Dynamics produced by universal adoption of strategy σ_c with different μ in parallel-path system $O = (\{\alpha, \beta\}, r = 1, p = 1, T = 2, A_0 = 1, v = \{\sigma_c \mapsto 1\})$.

parallel-path system where strategy σ_c is universally adopted can thus be described by the delay-differential equation (DDE)

$$\frac{\partial f_\alpha}{\partial t} = \begin{cases} r \cdot \mu \cdot \Delta c(t - T) \cdot f_\alpha(t) & \text{if } \Delta c(t - T) \leq 0, \\ r \cdot \mu \cdot \Delta c(t - T) \cdot f_\beta(t) & \text{otherwise,} \end{cases} \quad (3.15)$$

where $\Delta c(t - T) = c_\beta(t - T) - c_\alpha(t - T)$. This DDE describes a damped oscillator with delayed feedback and does not have an explicit solution [46]. However, we can numerically compute a solution using the method of steps [81].

As Fig. 3.2 shows, the choice of the parameter μ is critical for the strategy to actually lead to convergence. For high values of μ , such as 1, the strategy fails to produce convergence and yields undamped periodic oscillations. For low values of μ , such as 0.1, the system monotonically approaches the equilibrium without overshooting, i.e., it is *overdamped* (or, if nearly avoiding overshooting, *critically damped*). For values in-between, such as 0.5, the system eventually converges to stability at equal load, but only after overshooting, i.e., it is *underdamped*. However, for both the overdamped and the underdamped convergent strategies, we can make the following observation:

Insight 3.1 No PSS Equilibrium by Convergent Strategy. Universal adoption of the convergent path-selection strategy σ_c does not represent a PSS equilibrium, neither in its underdamped nor in its overdamped variant.

In the case of the overdamped strategy (e.g., σ_c with $\mu = 0.1$), the link loads monotonically approach each other and thus the greedy strategy allows an end-host to make use of a cheaper path sooner, making it the best-response strategy given universal adoption of σ_c . In the case of the underdamped convergent strategy (e.g., σ_c with $\mu = 0.5$), it is not obvious that the strategy is not a PSS equilibrium strategy in general. However, we can show that there exist alternative strategies to the underdamped rerouting policy that reduce a deviant agent's cost, see Appendix B.2.

3.3.2 MATE Algorithm

The MATE algorithm [79] was designed for the intra-domain context, where an ingress router has to distribute its demand d between multiple label-switched paths to a given egress router. As these ingress routers are under control of the domain operator, the MATE algorithm pursues convergence to the socially optimal traffic distribution, which minimizes latency from a global perspective, but is generally not preserved by selfish end-hosts. In the context of inter-domain networks, the MATE algorithm is instantiated such that it converges to a Wardrop equilibrium, which is preserved by selfish agents.

However, while the equilibrium produced by the MATE algorithm is compatible with the selfishness of agents, actually following the MATE algorithms is not necessarily rational from an end-host's perspective. Namely, an end-host in a parallel-path two-path system would execute the MATE algorithm as follows. In every re-evaluation, the end-host selfishly optimizes its traffic allocation (F_α, F_β) , where $F_\alpha = d - F_\beta$. In order to conform to the Wardrop model, the demand d is negligible from a global perspective. A MATE optimization step is defined as follows:

$$\begin{pmatrix} F'_\alpha \\ F'_\beta \end{pmatrix} = \left[\begin{pmatrix} F_\alpha \\ F_\beta \end{pmatrix} - \gamma \cdot \begin{pmatrix} c_\alpha(t-T) \\ c_\beta(t-T) \end{pmatrix} \right]^+ \quad (3.16)$$

Here, $[\mathbf{F}]^+$ represents a projection of allocation vector \mathbf{F} to the feasible allocation set defined by $F_\alpha + F_\beta = d$ with $F_\alpha, F_\beta \geq 0$. In order to reach convergence despite stale information, the coefficient γ has to conform to a certain upper bound [79].

As we show in Appendix B.2, the dynamics of an parallel-path system with universal adoption of the MATE algorithm are described by the following differential equation:

$$\frac{\partial f_\alpha}{\partial t} = \begin{cases} r \cdot \frac{\gamma}{2} \cdot \Delta c(t-T) \cdot f_\alpha(t) & \text{if } \Delta c(t-T) \leq 0 \\ r \cdot \frac{\gamma}{2} \cdot \Delta c(t-T) \cdot f_\beta(t) & \text{otherwise} \end{cases} \quad (3.17)$$

This equation is clearly equivalent to Eq. (3.15) for a choice of $\mu = \gamma/2$. A parallel-path system with universal adoption of σ_c and a system with universal adoption of the MATE algorithm thus exhibit the same flow dynamics, which allow for beneficial deviation:

Insight 3.2 No PSS Equilibrium by MATE. *The path-selection strategy as prescribed by the MATE algorithm is equivalent to the path-selection strategy σ_c . Thus, universal adoption of the MATE algorithm neither constitutes a PSS equilibrium.*

3.3.3 Conclusion

In summary, the kind of convergent path-selection strategies proposed in the literature cannot be assumed to be adopted by selfish end-hosts, as deviating from these strategies (e.g., by switching faster than prescribed by the strategy) is beneficial to an end-host.

Stability in a path-aware network architecture with selfish end-hosts can thus not be guaranteed by non-oscillatory path-selection strategies that prescribe a maximum rate of change to be respected by end-hosts. Instead, the network could employ mechanisms that incentivize end-hosts to follow non-oscillatory path-selection strategies. This finding reflects a similar result [16, 99] in the context of congestion control, namely that socially desirable behavior of end-hosts can only be enforced with network support.

3.4 STABILIZATION MECHANISMS

As argued in the previous section, rational end-hosts in networks with unrestricted path choice are unlikely to adopt convergent path-selection strategies. Therefore, network operators require mechanisms to incentivize the adoption of path-selection strategies that induce stability at equal load, i.e., *incentive-compatible* stabilization mechanisms. First, we integrate the concept of traffic-steering mechanisms into our

game-theoretic model (§3.4.1). Second, we specify in §3.4.2 the conditions under which these mechanisms are incentive-compatible.

3.4.1 Traffic-Steering Mechanisms

In order to affect the path-selection decisions of end-hosts in a parallel-path system O , a traffic-steering mechanism \mathcal{M} needs to alter the strategy cost $C(\sigma|O)$ for at least one path-selection strategy σ . A mechanism \mathcal{M} thus defines a function $c_{\mathcal{M}}(\pi, t)$ that quantifies the mechanism-imposed cost for using path π at time t . This cost is imposed onto the user of a path π in addition to the load-dependent path cost.

If a mechanism \mathcal{M} is active, the usage cost $c_u^{\mathcal{M}}$ extends the standard usage cost c_u from Equation (3.10) as follows:

$$c_u^{\mathcal{M}}(\pi, t) = c_u(\pi, t) + c_{\mathcal{M}}(\pi, t). \quad (3.18)$$

The cost formulas $c^{\mathcal{M}}(\pi, t|\tilde{\pi})$, $C^{\mathcal{M}}(\sigma, t)$, and $C^{\mathcal{M}}(\sigma|O)$ can be constructed from $c_u^{\mathcal{M}}(\pi, t)$, analogously to §3.2.4.

3.4.2 Incentive Compatibility

In general, incentive-compatible mechanisms are mechanisms that incentivize a certain form of desirable behavior. In our context, we consider traffic-steering mechanisms to be incentive-compatible if these mechanisms incentivize the desirable behavior of adopting a non-oscillatory path-selection strategy. In other words, an incentive-compatible mechanism creates a PSS equilibrium, i.e., a situation where every end-host minimizes its cost by adopting a non-oscillatory path-selection strategy, given that all other end-hosts do so:

Definition 3.5 Incentive-Compatible Traffic Steering. A traffic-steering mechanism \mathcal{M} is an incentive-compatible stabilization mechanism for a parallel-path system O if there is a strategy profile v^* such that

- (i) v^* leads to stability at equal load, and
- (ii) v^* represents a PSS equilibrium with respect to the cost function $C^{\mathcal{M}}(\sigma|O)$.

In the following two sections, we present two instances of stabilization mechanisms, namely FLOSS and CROSS, and prove their incentive compatibility. The two

mechanisms differ in the methods for achieving stability: Whereas FLOSS reduces the imbalance between two paths by regulating the migration rate between the paths, CROSS achieves stability by repetitive reshuffling of flows between paths and increasing the cost of path migration.

3.5 THE FLOSS MECHANISM

In this section, we present the FLOSS mechanism (*Flow-Loyalty Oscillation-Suppression System*).

3.5.1 Overview

As shown in §3.3, convergent path-selection strategies are characterized by careful path-switching behavior: An end-host only switches to a seemingly cheaper path with a modest probability that depends on the measured latency difference, translating into a relatively low migration rate between paths. It is well known that system stability can be achieved by limiting the rate of change (also known as the *system gain* [140]). However, the challenge is to develop methods that achieve this change-rate limitation in the face of selfish, uncontrolled end-hosts. Such a method is given by FLOSS.

As selfish end-hosts do not voluntarily conform to a modest path-migration rate, the path-migration rate has to be *regulated* by network operators. The FLOSS mechanism performs such regulation by rewarding end-hosts that are *loyal* to a certain path and by restricting arbitrary path migration of oscillating end-hosts.

In order to regulate path migration, the FLOSS mechanism makes use of *registrations* and proceeds in *intervals*. Figure 3.3, which shows a simulation of the FLOSS mechanism in a two-path system, illustrates the FLOSS approach. Initially, the FLOSS mechanism announces at time t' that all end-hosts are required to obtain a registration for one path π of their choice. This registration allows an end-host to use path π during a future time interval $I_0 = [t_0, t_1)$ with $t' < t_0 < t_0 + T < t_1$. End-hosts that use path π without a registration are punished in the interval (e.g., by dropping packets).

This call for registration produces a distribution of flows over the two paths, which is stable during the interval as no end-host can switch to the path which it is not registered for. However, this load distribution is unlikely to be perfectly equal. The FLOSS mechanism iteratively reduces this imbalance: In every following time interval, a small set of flows are allowed to migrate from the more expensive path to the cheaper path. This allowance is enforced by selectively granting registrations:

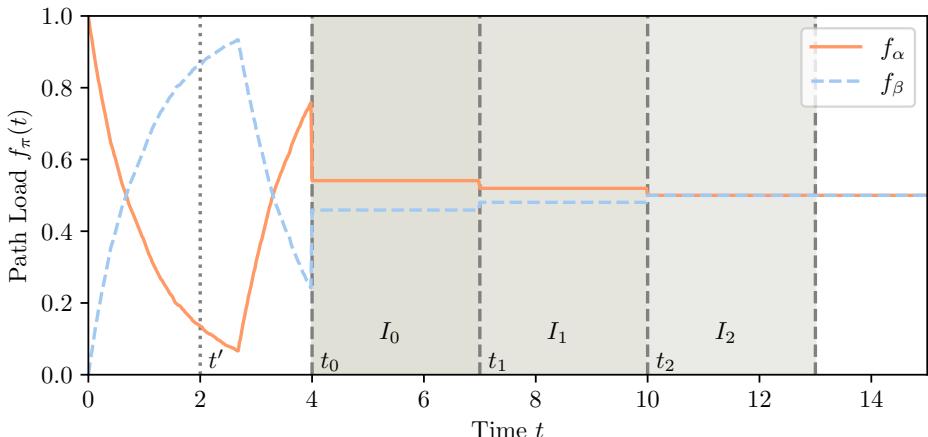


Figure 3.3: Simulation of FLOSS enforcement in an oscillation-prone system $O = (\Pi = \{\alpha, \beta\}, r = 1, p = 1, T = 2, A_0 = 1, v = \{\sigma_F \mapsto 1\})$.

Whereas end-hosts with a pre-existing registration for a path (loyal end-hosts) always obtain a registration for that path, end-hosts without a pre-existing registration are not always allowed to register. Once the imbalance is sufficiently small, the end-hosts do not have an incentive anymore to switch paths, at which point the enforcement of the mechanism can be suspended (e.g., at the end of interval I_2 in Figure 3.3).

Theorem 3.1 Incentive Compatibility of FLOSS. *The FLOSS mechanism is an incentive-compatible stabilization mechanism.*

As defined in §3.4.2, incentive compatibility implies the existence of a strategy profile that leads to stability at equal load *and* is a PSS equilibrium during mechanism enforcement. For FLOSS, such a strategy profile is given by universal adoption of the FLOSS-compliant path-selection strategy σ_F . The strategy σ_F prescribes to use the path with the lowest expected cost which the end-host is entitled to use. Our incentive-compatibility proof thus builds on the following two concrete lemmas, which are proven in §3.5.2 and §3.5.3, respectively:

Lemma 3.1 Stability of FLOSS. *Universal adoption of the FLOSS path-selection strategy σ_F leads to stability at equal load.*

Lemma 3.2 PSS Equilibrium under FLOSS. Universal adoption of the FLOSS path-selection strategy σ_F represents a PSS equilibrium during enforcement of the FLOSS mechanism.

3.5.2 Stability Analysis

In order to prove Lemma 3.1, we assume universal adoption of path-selection strategy σ_F , i.e., an end-host always uses the path with the lower expected cost provided that the end-host is entitled to use that path.

Initial interval. When registering before the initial interval, all end-hosts simultaneously decide for one path to use during the upcoming interval $[t_0, t_1]$. Confronted with such a choice, each end-host aspires to commit to the path π that will be selected by fewer other end-hosts, i.e., the path π with $f_\pi(t_0) < f_{\hat{\pi}}(t_0)$. In absence of inherent differences between the two choices, the only Nash equilibrium of such a speculative game is given if every end-host commits to each path π with probability 1/2.

Idealized execution. In expectation, the load on both paths α and β is thus $\mathbb{E}[f_\alpha(t_0)] = \mathbb{E}[f_\beta(t_0)] = 1/2$. Since no migration occurs during the interval $[t_0, t_1]$, the load distribution is expected to remain equal during the interval, i.e., $\mathbb{E}[f_\alpha(t)] = \mathbb{E}[f_\beta(t)] = 1/2 \forall t \in [t_0, t_1]$. If this expectation were actualized without variance, the mechanism enforcement could end at time t_1 , and the equal load distribution would be preserved even if the end-hosts were free again to arbitrarily select paths after time t_1 . Since $t_0 + T < t_1$, any end-host performing a re-evaluation after t_1 perceives the Wardrop equilibrium $c_\alpha(t - T) = c_\beta(t - T)$ and will thus not switch paths. Therefore, the system is stable at equal load even when the mechanism is not enforced anymore.

Realistic execution. In reality, however, variance makes it likely that the load on paths α and β is not perfectly equalized at t_0 . In that case, the FLOSS mechanism attempts to eliminate the remaining load difference $\Delta f(t_0) = |f_\alpha(t_0) - f_\beta(t_0)| > 0$ as follows. Starting from $t'' = t_0 + T$, the end-hosts can again register on paths for an upcoming interval $[t_1, t_2]$. At t'' , all end-hosts correctly perceive the cost difference between a cheaper path π and a more expensive path $\hat{\pi}$, as for every path $\hat{\pi}$, $c_{\hat{\pi}}(t'' - T) = c_{\hat{\pi}}(t_0) = c_{\hat{\pi}}(t'')$ due to the constant load in $[t_0, t'']$. The core idea of the FLOSS mechanism is to enforce a *migration allowance* $\rho_\pi(t_1)$, which is an

upper bound on the amount of end-hosts that are allowed to migrate from path $\tilde{\pi}$ to path π at time t_1 .

Migration allowance. Importantly, $\rho_\pi(t_1)$ is chosen such that

$$f_\pi(t_0) + \rho_\pi(t_1) \cdot f_{\tilde{\pi}}(t_0) \leq (1 - \rho_\pi(t_1)) \cdot f_{\tilde{\pi}}(t_0), \quad (3.19)$$

which implies $c_\pi(t_1) \leq c_{\tilde{\pi}}(t_1)$ (i.e., the cheaper path π will remain the cheaper path in the next interval even if a share $\rho_\pi(t_1)$ of end-hosts on the more expensive path $\tilde{\pi}$ migrate to path π). This choice of $\rho_\pi(t_1)$ ensures the correct incentives for the end-hosts. Given such an assurance, end-hosts registered on the cheaper path π during $[t_0, t_1]$ minimize their cost by remaining on path π . Since these end-hosts are considered *loyal* to path π , their registration at path π will be renewed for the upcoming interval $[t_1, t_2]$. Conversely, all end-hosts registered on the more expensive path $\tilde{\pi}$ would minimize their cost by migrating to the cheaper path π . However, the FLOSS mechanism restricts this migration by only granting a registration for π to a share $\rho_\pi(t_1)$ of end-hosts on $\tilde{\pi}$. The non-migrating end-hosts on path $\tilde{\pi}$ are considered loyal on path $\tilde{\pi}$ and are thus allowed to renew their registration at $\tilde{\pi}$.

Therefore, exactly $\rho_\pi(t_1) \cdot f_{\tilde{\pi}}(t_0)$ migrate from path $\tilde{\pi}$ to path π at time t_1 , which reduces the difference in load and cost between the paths π and $\tilde{\pi}$. By repetitive mechanism application with appropriately chosen migration allowances, the FLOSS mechanism can arbitrarily minimize the cost differential between the paths π and $\tilde{\pi}$. When the cost difference becomes so small that end-hosts perceive a Wardrop equilibrium, the mechanism has achieved stability at equal load that continues to hold even without mechanism enforcement.

3.5.3 PSS Equilibrium Analysis

We now prove Lemma 3.2, i.e., we show that path-selection strategy σ_F is the optimal strategy for an end-host given that all other end-hosts have adopted σ_F . Concretely, we show that the FLOSS mechanism induces a PSS equilibrium $v^* = \{\sigma_F \mapsto 1\}$, where σ_F is the universally adopted path-selection strategy with the following path-selection function:

$$u_F(\pi, t | \tilde{\pi}) = \begin{cases} 1/2 & \text{if } t = t_0, \\ 1 & \text{if } t > t_0 \text{ and } E_e(\pi, t) \\ & \quad \text{and } c_\pi(t - T) < c_{\tilde{\pi}}(t - T), \\ 0 & \text{otherwise} \end{cases} \quad (3.20)$$

where $E_e(\pi, t)$ is true if and only if end-host e is entitled to use path π at time t . We assume that an end-host always knows whether it is entitled to use a path. For the initial interval, every path is selected with equal probability $1/2$. For all subsequent intervals, a path π is selected if the path is perceived to be cheaper than the current path $\tilde{\pi}$ and end-host e is entitled to use path π . For remaining on a path $\tilde{\pi}$, it holds that $u_F(\tilde{\pi}, t | \tilde{\pi}) = 1 - u_F(\pi, t | \tilde{\pi})$.

Mechanism cost. The FLOSS mechanism makes strategy σ_F the equilibrium strategy by imposing the additional cost $c_M(\pi, t)$ for using path π at time t . End-host e incurs a cost c_a for attempting to register and a penalty cost c_p for using a path without a registration. We assume $c_p = \infty$, i.e., the penalty cost makes a path unusable. Let $A_e(\pi, t)$ be true if and only if end-host e applies to register for using path π at time t and let $R_e(\pi, t)$ be true if and only if end-host e obtained a registration for using path π at time t , i.e., $R_e(\pi, t) = A_e(\pi, t) \wedge E_e(\pi, t)$. Using these predicates, the cost imposed by the FLOSS mechanism can be expressed as

$$c_M(\pi, t | A_e, R_e) = [A_e(\pi, t)] \cdot c_a + [\neg R_e(\pi, t)] \cdot c_p, \quad (3.21)$$

where $[P] = 1$ if the predicate P is true and 0 otherwise.

A selfish end-host e chooses its actions such that its cost from the mechanism is minimized. Therefore, an end-host e requests a registration if and only if the end-host is entitled to the registration, as there is no benefit of a registration request that will be refused. Thus, the relevant mechanism-imposed cost for end-host e is

$$c_M(\pi, t) = \min_{A_e} c_M(\pi, t | A_e, R_e) = \begin{cases} c_a & \text{if } E_e(\pi, t), \\ c_p & \text{otherwise.} \end{cases} \quad (3.22)$$

Initial interval. Concerning the initial interval with start t_0 , both paths α and β have expected cost $c_\pi(t_0) = 1/2^p$ if all other end-hosts choose each path with probability $u_F(\pi, t | \tilde{\pi}) = 1/2$. As both paths have the same cost and both paths require a registration, the usage cost of both paths is $c_u^M(\pi, t_0) = 1/2^p + c_a$.

Independent of the current path $\tilde{\pi}$, the cost of applying strategy σ_F at time t_0 is thus $c^M(\sigma_F, t_0 | \tilde{\pi}) = 1/2^p + c_a$ for any choice of $u(\pi, t_0 | \tilde{\pi})$. Therefore, end-host e cannot reduce its cost by choosing another path-selection probability than $u_F(\pi, t_0 | \tilde{\pi}) = 1/2$, which makes σ_F an equilibrium strategy for the initial interval.

Subsequent intervals. Concerning subsequent intervals with start $t_i > t_0$, we have to distinguish two cases for the current path π' of end-host e , namely whether end-host e is on the cheaper path π or on the more expensive path $\tilde{\pi}$.³

1. If end-host e is on the cheaper path π , the cost of remaining on π is $c_u^M(\pi, t_i) = c_\pi(t_i) + c_a$, whereas the cost of switching to $\tilde{\pi}$ is $c_u^M(\tilde{\pi}, t_i) = c_{\tilde{\pi}}(t_i) + c_a$ if $E_e(\tilde{\pi}, t_i)$ and $c_{\tilde{\pi}}(t_i) + c_p$ otherwise. As always $c_u^M(\pi, t_i) < c_u^M(\tilde{\pi}, t_i)$, the current path π must be selected with probability $u(\pi, t | \pi) = 1$ to minimize the end-host's cost.
2. If end-host e is on the more expensive path $\tilde{\pi}$, the cost of remaining on $\tilde{\pi}$ is $c_u^M(\tilde{\pi}, t_i) = c_{\tilde{\pi}}(t_i) + c_a$, whereas the cost of switching to π is $c_u^M(\pi, t_i) = c_\pi(t_i) + c_a$ if $E_e(\pi, t_i)$ and $c_\pi(t_i) + c_p$ otherwise. Thus, $c_u^M(\pi, t_i) < c_u^M(\tilde{\pi}, t_i)$ if $E_e(\pi, t_i)$, but $c_u^M(\tilde{\pi}, t_i) < c_u^M(\pi, t_i)$ otherwise. If end-host e is entitled to use the cheaper path π , the cheaper path π must thus be selected with probability $u(\pi, t | \tilde{\pi}) = 1$ to minimize the end-host's cost, and with probability 0 otherwise.

In summary, for all intervals with start $t_i > t_0$, an end-host e optimizes its cost by switching to an alternative path π if and only if path π is cheaper than the current path $\tilde{\pi}$ and end-host e is entitled to use path π . This path-switching behavior is exactly captured by the path-selection function $u_F(\pi, t | \tilde{\pi})$. Therefore, path-selection strategy σ_F is an equilibrium strategy for both the initial interval and the subsequent intervals of the mechanism, which proves Lemma 3.2.

3.6 THE CROSS MECHANISM

In this section, we present a second stabilization mechanism called CROSS (for *Computation-Requiring Oscillation Suppression System*).

³ Thanks to the load being constant in subsequent intervals, the cost $c_{\tilde{\pi}}(t)$ of a path $\tilde{\pi}$ at registration time t is equal to the known stale cost $c_{\tilde{\pi}}(t - T)$. Therefore, any end-host can correctly identify the cheaper and the more expensive path.

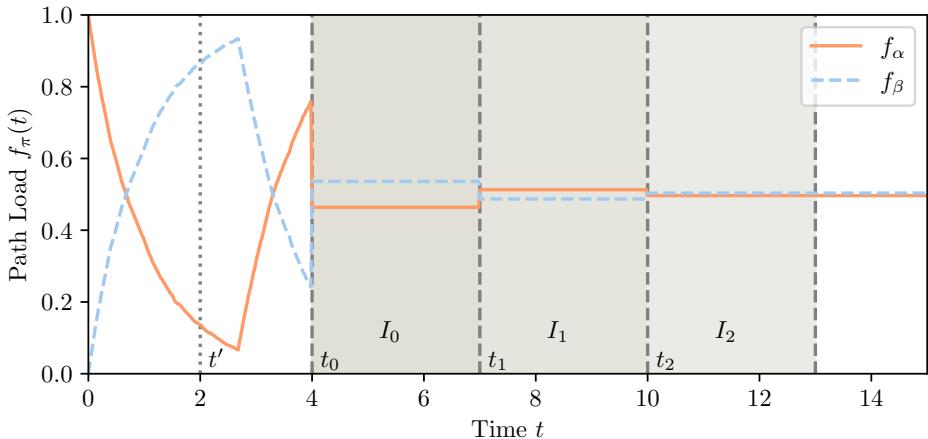


Figure 3.4: Simulation of CROSS enforcement in an oscillation-prone system $O = (\Pi = \{\alpha, \beta\}, r = 1, p = 1, T = 2, A_0 = 1, v = \{\sigma_C \mapsto 1\})$, with $\epsilon = 0.01$.

3.6.1 Overview

Motivation: Backup paths. While the FLOSS mechanism (cf. §3.5) deterministically achieves stability at equal load, its strict enforcement of the migration allowance represents a problem in case of path failures. When a path fails, an end-host on that path is not allowed to switch to an alternative path immediately. Only when the path failure is detected after some time by the mechanism, enforcement of the mechanism can be stopped and the end-hosts can be allowed to use an alternative path. For highly critical transmissions, such inflexibility is undesirable.

The CROSS mechanism allows end-hosts to obtain an *insurance* against such cases of path failure. Basically, the CROSS mechanism works similarly to the initial interval of the FLOSS mechanism: End-hosts are required to register for one path of their choice, which in general cannot be changed during the upcoming interval. Unlike FLOSS, however, the CROSS mechanism offers the possibility of registration for a second path that can be immediately used in case of a path failure, even if the path failure is not yet confirmed.

Design challenge: Avoid oscillation. However, the question is how to avoid that end-hosts always register for both paths and, if on the more expensive path, falsely claim to be affected by a path failure and switch to the cheaper path. Such opportunistic behavior would cause oscillation. To solve this problem, the idea of

the CROSS mechanism is that end-hosts must prove that they need the immediate-switching option for insurance against path failures, not simply for opportunistic cost reduction. End-hosts can prove their truthfulness by paying a price for the immediate-switch option. This price must be higher than any cost gain that can be achieved by switching to a cheaper path in a scenario without path failure. An end-host that paid this price thus only switches to the backup path if a path failure has occurred; if no path failure occurred, the end-host would not trade its insurance option against the cost gain, as the insurance option is more valuable to the end-host than any cost gain. Immediate switching during the interval can thus be allowed to the end-hosts with a backup-path registration. Moreover, immediate switching behavior by those end-hosts is an indication of path failure, which means that all other end-hosts must be allowed to migrate as well.

Cryptographic puzzles. As a price for the backup path registration, the CROSS mechanism requires the solution to a computationally hard puzzle. This puzzle is structured such that only end-hosts with a sufficiently high valuation of the backup path will obtain a solution. More precisely, each puzzle \mathcal{E} is associated with a cryptographic hash function $h : \{0,1\}^* \mapsto [0,1]$ and a difficulty level $\delta \geq 0$. An end-host e can solve a puzzle $\mathcal{E}(\pi)$ for registering at a backup path π by finding a value s such that $h(\pi, t_i, e, s) \leq 2^{-\delta}$, where t_i is the start of the next interval. Given a secure hash function, a puzzle $\mathcal{E}(\pi)$ can only be solved by brute force, i.e., varying s in a series of hash computations. By finding an appropriate s , an end-host can obtain a backup-path registration.

Probabilistic stability guarantees. Also unlike FLOSS, the CROSS mechanism allows end-hosts to register at a path of their choice not only for the initial interval, but for every interval. Therefore, even if the path failure is not detected for some reason (e.g., because no end-host obtained a backup registration), the end-host can use the alternative path in the interval after a path failure. The CROSS mechanism thus has a non-deterministic approach for achieving stability: Intervals in CROSS serve as *balancing trials* and are repeated until the load imbalance is small enough that end-hosts do not switch paths anymore. Since the end-hosts select each path with probability 1/2 in any balancing trial, the probability that an approximately equal load distribution results after a few balancing trials is substantial. Still, the additional flexibility of CROSS results in a relaxation of convergence guarantees: Instead of convergence to an equal-load distribution, the CROSS mechanism only guarantees convergence to a traffic distribution with approximately equal load. A

simulation of CROSS enforcement is visualized in Figure 3.4, which also shows the convergence produced by the CROSS approach.

Theorem 3.2 Incentive Compatibility of CROSS. *The CROSS mechanism is an incentive-compatible stabilization mechanism that achieves stability at approximately equal load, i.e., for every $\epsilon > 0$, $\lim_{t \rightarrow \infty} \Delta(t) < \epsilon$.*

The CROSS mechanism achieves stability at approximately equal load by incentivizing the universal adoption of path-selection strategy σ_C , which prescribes that end-hosts only use a path if they have a corresponding registration and only use a backup in case of path failures. More formally, Theorem 3.2 directly follows from Lemmata 3.3 and 3.4:

Lemma 3.3 Stability of CROSS. *Universal adoption of the CROSS path-selection strategy σ_C leads to stability at approximately equal load.*

Lemma 3.4 PSS Equilibrium under CROSS. *Universal adoption of the CROSS path-selection strategy σ_C represents a PSS equilibrium given enforcement of the CROSS mechanism.*

While the proof of Lemma 3.3 is intuitive and can thus be found in Appendix B.4, Lemma 3.4 is proven below.

3.6.2 PSS Equilibrium Analysis

CROSS PSS. In this section, we prove Lemma 3.4 by showing that universal adoption of path-selection strategy σ_C is a PSS equilibrium, i.e., if all other end-hosts adopt σ_C , σ_C is the optimal strategy for a single end-host e . The path-selection strategy σ_C is characterized by the following path-selection function for $\pi \neq \tilde{\pi}$:

$$u_C(\pi, t | \tilde{\pi}) = \begin{cases} 1/2 & \text{if } t = t_i \wedge \neg R'_e(\pi, t), \\ 1 & \text{if } c_{\tilde{\pi}}(t - T) = \infty \wedge R'_e(\pi, t), \\ 0 & \text{otherwise,} \end{cases} \quad (3.23)$$

where t_i is the start time of any balancing trial, $c_{\tilde{\pi}}(t - T) = \infty$ designates a path failure and $R'_e(\pi, t)$ is true if and only if end-host e has a *backup* registration for path π at time t . Moreover, $u_C(\tilde{\pi}, t | \tilde{\pi}) = 1 - u_C(\pi, t | \tilde{\pi})$.

Mechanism cost. As in FLOSS, registering has cost c_a , whereas using a path without registration imposes a penalty cost $c_p = \infty$. Additionally, an end-host incurs cost by solving puzzles, where each hashing operation has cost c_h . To an end-host with valuation ω of a backup path, a hash operation has the expected utility $\mathbb{E}[U_h](\delta, \omega) = 2^{-\delta}\omega - c_h$.

Optimal PSS. Given puzzle-difficulty level δ , an end-host thus solves a puzzle if and only if it has a backup valuation ω such that $\mathbb{E}[U_h](\delta, \omega) > 0$. If an end-host does not solve a puzzle, it simply obtains a regular registration for one path at cost c_a , where every path is selected with probability $1/2$. Obtaining no registration and using any path would incur a much higher penalty cost $c_p \gg c_a$ and is thus not rational. Therefore, an end-host with a registration for one path uses this path from the start t_i of the balancing trial. If an end-host solves a puzzle, the end-host obtains a backup registration for the path corresponding to the puzzle and obtains a regular registration for the other path at cost c_a . Since CROSS enforces that an end-host can only switch once to its backup path and never switch back during the balancing trial, every end-host with a backup-path registration starts by using the path with its regular registration at time t_i . In summary, the optimal path-selection function for all $t = t_i$ is $u_C(\pi, t|\pi') = 1/2$ if $\neg R'_e(\pi, t)$.

During the balancing trial, no reallocation decisions are taken before $t_i + T$, as the expected path costs during $[t_i, t_i + T]$ is $\mathbb{E}[c_\alpha] = \mathbb{E}[c_\beta] = 1/2^p$. Only at $t_i + T$, the actual imbalance $\Delta(t) = |f_\pi(t_i) - f_{\tilde{\pi}}(t_i)|$ between a more expensive path $\tilde{\pi}$ and a cheaper path π becomes visible to the end-hosts. If the end-hosts on path $\tilde{\pi}$ with a backup registration for path π switched at that point, they would save

$$\Delta \bar{C} = \int_{t_i+T}^{t_{i+1}} (c_{\tilde{\pi}}(t) - c_\pi(t)) dt < \Delta \bar{C}_{\max} = t_{i+1} - t_i - T. \quad (3.24)$$

However, such a switch would erase the backup value ω of path π for the end-host, which is why an end-host with a backup registration for path π only switches to path π if $\omega < \Delta \bar{C}$. In order to disincentivize such migration and keep the load distribution constant, the CROSS mechanism chooses the puzzle-difficulty level δ such that $\mathbb{E}[U_h](\delta, v) > 0$ if and only if $\omega > \Delta \bar{C}_{\max}$. This choice of δ leads to a situation where the end-hosts with a backup registration will only switch to the backup path in case of a path failure, as these end-hosts value the backup option higher than any cost reduction obtainable without path failure. In case of a path failure, however, trading the backup value ω of path π against the infinite cost of failed path $\tilde{\pi}$ is rational and the end-hosts with a backup registration switch the paths. In summary, the optimal path-selection function for end-host e and for

all $t \neq t_i$ is thus $u_C(\pi, t | \tilde{\pi}) = 1$ if $R'_e(\pi, t)$ and $c_{\tilde{\pi}}(t - T) = \infty$, and $u_C(\pi, t | \tilde{\pi}) = 0$ otherwise. Thereby, path-selection strategy σ_C has been established as the PSS equilibrium strategy.

3.7 PRACTICAL APPLICATION

While the focus of this chapter is on the theoretical exploration of selfish path selection and stabilization mechanisms, this section lays out a pathway toward practical application of our findings. First, we discuss practical requirements for inter-domain stabilization mechanisms in §3.7.1. In §3.7.2, we present a mechanism-enforcement architecture that conforms to these requirements. In §3.7.3 and §3.7.4, we outline how the FLOSS and CROSS mechanisms could be practically implemented.

3.7.1 Requirements

If a stabilization mechanism is to be practically applied by network operators in an inter-domain architecture, the mechanism must conform to the following requirements:

Limited overhead. The stabilization mechanism must only induce a small overhead on network operators. In particular, the genuine function of AS border routers (forwarding traffic at line rate) must not be compromised by expensive mechanism-enforcement tasks. Note that both mechanisms only need to be enforced by routers in case of oscillation and until stabilization is achieved; however, the mechanisms should induce little overhead even during this short time span.

No explicit inter-AS coordination (coordination-freeness). The stabilization mechanism must not rely on explicit inter-AS coordination. Such explicit coordination may not be feasible or scalable, as the domains that perceive the same oscillation pattern may be mutually unknown, mutually distrusted, or very distant from each other.

3.7.2 Mechanism-Enforcement Architecture

To enforce an incentive-compatible stabilization mechanism, an AS operator needs the means to detect, inform, and punish the selfish entities that employ an oscillatory path-selection strategy. In this section, we describe a mechanism-enforcement

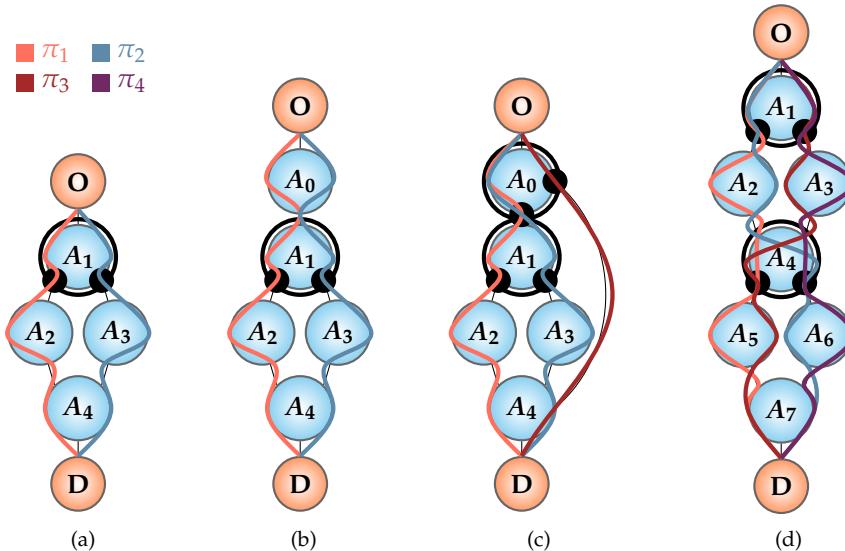


Figure 3.5: Oscillation patterns.

architecture that provides these means to an AS operator while conforming to the requirements in §3.7.1.

Inter-domain architecture. From an inter-domain perspective, the most important architectural question concerns *coordination*, i.e., how each AS perceiving an oscillation pattern contributes to oscillation suppression. As explicit inter-AS coordination is undesirable, an implicit method for *responsibility assignment* is necessary.

We leverage a fundamental property of paths in inter-domain networks as a natural way to assign responsibility for inter-domain oscillation suppression: For every pair of paths connecting the same origin and destination ASes, there is at least one AS (henceforth: the *splitting AS*) in which the paths split, i.e., the paths contain different egress interfaces out of the AS. For every oscillation between two paths, at least one AS perceives the oscillation as an oscillation of traffic between egress interfaces, not only as periodic upswings and downswings in the load at one egress interface. Such splitting ASes are the natural candidates for a leading role in inter-domain oscillation suppression, as these ASes are both best informed about the oscillation and in the best position to manage the oscillating traffic.

For illustration of the path-splitting property, Fig. 3.5 shows different types of oscillation patterns for paths connecting an origin end-host O and a destination end-host D . In the simplest cases, the oscillation may be perceived at the origin AS (AS A_1 in Fig. 3.5a) or at one intermediate AS (AS A_1 in Fig. 3.5b). However, the oscillation may be perceived at multiple splitting ASes. The different paths may pass through a different number of egress interfaces at which the mechanism is enforced. For example, path π_3 in Fig. 3.5c only passes through one critical egress interface (at AS A_0), whereas paths π_1 and π_2 pass through two critical egress interfaces. Conversely, each path in Fig. 3.5d passes through two egress interfaces at which a load-balancing mechanism is enforced. Any stabilization mechanism may thus be applied repeatedly and with different frequency to flows belonging to the same parallel-path system.

Intra-domain architecture. In the intra-domain context, the mechanism-enforcement architecture envisages a centralized oscillation-suppression service (OSS) in each AS. The OSS is capable of interacting with the border routers at the egress interfaces. For a splitting AS, this OSS functions as displayed in Fig. 3.6. By collecting aggregate load statistics from the border routers, the OSS in the splitting AS can identify the egress interfaces between which oscillation occurs (through correlation). As the presence of such oscillation means that the AS is obliged to enforce a stabilization mechanism, the OSS equips every oscillation-perceiving border router r_i with data M_i that is necessary to enforce the mechanism (e.g., start time of the next interval). By further collecting load statistics from the egresses, the OSS monitors and continuously adapts the execution of the mechanism. The border routers communicate with the origins of the oscillating flows by appending mechanism-relevant information to passing packets.

3.7.3 FLOSS in Practice

In the following, we discuss how the FLOSS mechanism could be applied by the mechanism-enforcement architecture from §3.7.2, while under the requirements from §3.7.1, namely limited overhead and coordination-freeness.

3.7.3.1 Limited Overhead

Registration on routers. In order to signal that end-hosts must register for an upcoming time interval, a border router appends the start time t_i of the next interval to passing packets. If an end-host witnesses such a call for registrations

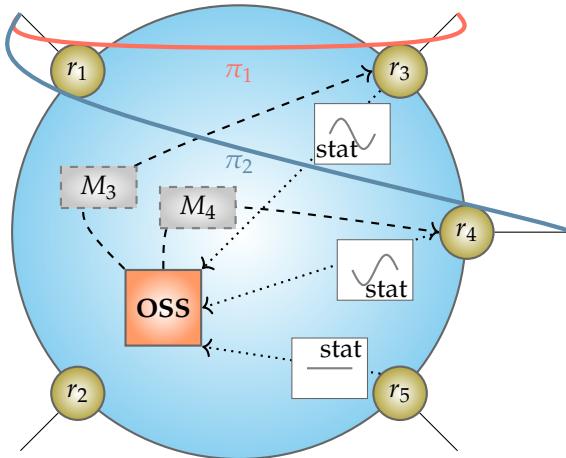


Figure 3.6: Mechanism-enforcement architecture (within the splitting AS).

in its packets, it can send a packet with a registration request over the desired egress. A border router can keep track of registrations using a Bloom filter, which approximates a set of flow IDs. A Bloom filter offers constant complexity for both lookup and insertion, although suffering from false positives. When checking for registrations, false positives result in unregistered flows being able to send over an egress and being rewarded like loyal flows. However, the enforced migration rate ρ can simply be discounted by the false-positive rate of the Bloom filter such that the desired migration rate is enforced despite the presence of lucky unregistered flows.

Enforcement of single registration. In order to avoid that an end-host registers on multiple egresses, a border router forwards all registrations to the OSS, which keeps track of egress-specific registration by flows and can therefore spot multiple registrations by the same flow. If multiple registrations are detected, the OSS pushes a blacklist update for the malicious flow ID to the border routers. In order to avoid introducing DoS attacks where a malicious actor provokes the blacklisting of an end-host by sending multiple registrations, we assume some form of lightweight source authentication, which is typically offered by path-aware Internet architectures [219].

Selective admission of migrating flows. Border routers need an efficient way to decide whether to grant registration applications to flows that are willing to switch paths, while preserving the property that a maximum share ρ of flows migrates. Such

selective admission can be implemented using a publicly known hash function h , which maps the flow ID f to the interval $[0, 1]$. If $h(t_i|f) < \rho$, the registration is granted, where t_i is the beginning time of the next registration-enforcement interval. This construction has the advantage that an end-host can locally check whether it will be accepted on the alternative ingress, as h , t_i , and f are known to the end-host. Therefore, the border router is not bothered by registration requests from end-hosts that would be rejected. Furthermore, it is important to choose the flow ID f based on attributes that the source end-host cannot easily influence without compromising its communication, e.g., source and destination IP, but not source or destination port.

Small traffic allowance for unregistered flows. While unregistered end-hosts should not be able to properly use an egress, these end-hosts should be able to send a few packets over the egress to measure the latency of the corresponding path. Also, short flows, e.g., DNS requests, should not be required to obtain a registration. Such a limited traffic allowance can be efficiently achieved by applying the mechanism only to a subset of packets, e.g., by sub-sampling. If registrations are only checked for a sub-set of packets, even an unregistered flow has a high chance of getting a few packets through the egress, while still experiencing severe disruption when sending a large number of packets over the egress. Due to the structure of congestion-control algorithms, sub-sampling rates as low as 1% already cause enough packet drops to make a path completely unusable for unregistered flows [168]. Moreover, sub-sampling reduces the workload on border routers.

Addition of new flows. In reality, new flows appear during the execution of the mechanism. Clearly, these flows cannot register in advance for an enforcement interval, as these flows do not exist beforehand. Therefore, new flows are also allowed to register at one path of their choice *during* an enforcement interval. In order to distinguish new flows from flows that merely pretend to be new, the FLOSS mechanism samples the active flows at both egresses in every interval and inserts them into a Bloom filter. These previously active flows are supposed to have a registration in the subsequent interval. In contrast, truly new flows can be identified with a lookup failure in the mentioned Bloom filter. Due to false positives, a truly new flow might be mistaken for a previously active flow and thus be denied a retroactive registration. However, given a small false-positive probability, the registration at one path should always be possible in practice. As all new flows (except the false-positive new flows) must be expected to flock to the cheaper path, the migration allowance must be discounted by the birth rate of flows.

3.7.3.2 Coordination-Freeness

If there is one splitting AS for a parallel-path system, there are no unintended effects due to distributed application of the mechanism. However, as explained in §3.7.2, there may be multiple mechanism-enforcing ASes along a path. If n_i is the number of splitting ASes along path π_i , the costs for obtaining a registration for π_i and for using π_i without a registration are $n_i \cdot c_a$ and $n_i \cdot c_p$, respectively. In cases where n_i is the same for every path π_i of an oscillation pattern (such as in Figure 3.5d), the incentives for the end-hosts thus do not change compared to a single-application scenario. However, if n_i is different for the paths π_i in the parallel-path system (such as in Figure 3.5c), the registration cost for different paths may be different. For example, the registration cost for obtaining a registration of path π_3 in Figure 3.5c is c_a , whereas the corresponding cost for paths π_1 and π_2 is $2c_a$. Since $c_p = \infty > n_i c_a$ for all finite n_i , registering for a path is still worthwhile. However, an equilibrium between the two egresses of AS A_0 is only reached if $(f_{\pi_1} + f_{\pi_2})^p + 2c_a = f_{\pi_3}^p + c_a$, which implies stability at *unequal* load. However, since the cost c_a for obtaining a registration is modest (just a single packet as explained in §3.7.3.1), the resulting load imbalance between the ASes is also modest. Therefore, no explicit inter-AS coordination is needed.

3.7.4 CROSS in Practice

In this section, we discuss the CROSS mechanism with respect to the two practicality requirements, i.e., limited overhead and coordination-freeness.

3.7.4.1 Limited Overhead

Compared to FLOSS, the only additional piece of functionality needed for CROSS is puzzle verification. Efficient puzzle-solution verification on border routers is performed by a hash function evaluation with the appropriate arguments, among which is the solution value provided by the data packet (cf. §3.6.2).

3.7.4.2 Coordination-Freeness

Like FLOSS, CROSS suffers from the minor issue that some paths may require more registrations than other paths. Concerning backup registrations, multiple applications of the mechanism do not constitute a problem, as an end-host always has to solve only one puzzle to obtain a backup registration. For example, an end-host in the network of Figure 3.5c could insure against path failure as follows.

At AS A_0 , the end-host would obtain a normal registration for π_3 and a backup registration for π_1 and π_2 . Such a combined backup registration is possible by including only the respective egress of AS A_0 in the puzzle solution, not the specific path. At AS A_1 , the end-host can then obtain a normal registration for one of these paths, e.g., π_1 . If the end-host desires an additional insurance against failure of path π_1 , the end-host can solve a puzzle to obtain a backup registration for π_2 at AS A_1 . Since only one puzzle per backup path is needed, no explicit inter-AS coordination is necessary to preserve the incentives of the CROSS mechanism.

3.8 RELATED WORK

Hop-by-hop forwarding. Due to the traditional paradigm of network-controlled path selection, most traffic-engineering tools assume that packet forwarding is performed by series of decisions taken by the hops along a path. Systems such as AMP [100], ReplEx [84], Homeostasis [152], and HALO [177] thus prescribe how routers along a path should take forwarding decisions, mostly by adapting traffic-splitting ratios based on network information. If packets must be forwarded along a path chosen by the end-host, these schemes cannot be used.

End-host path selection. An alternative line of work is generally compatible with the emerging paradigm of end-point path selection. Assuming path selection by the source, this flavor of research prescribes path-selection strategies that lead to convergence. However, such convergent path-selection strategies are always designed for an intra-domain context, i.e., for path selection within a domain where end-points are under control of the network operator. Due to the selfishness of end-hosts in the inter-domain context, these schemes are thus impractical. For example, Proportional Sticky Routing [196] relies on self-restraint of end-points, which leads to persistent preference of shortest paths over alternative paths even when alternative paths are more attractive. The convergence of MATE [79] and the rerouting strategy designed by Kelly and Voice [140] is built on the assumption that the end-points restrain themselves to a maximum speed when reallocating traffic on cheaper paths, which cannot be expected from selfish end-hosts. In TeXCP [137], end-points are expected to comply with maximum traffic-reallocation allowances dynamically set by the network. Similarly, the rerouting policies designed by Fischer and Vöcking [85] require that end-hosts do not exceed a certain probability for switching to a cheaper path. Finally, OPS [136] also demands behavior from end-hosts that is irrational in a game-theoretic sense, in particular the probabilistic usage of sub-optimal paths.

Inter-domain traffic engineering. Inter-domain traffic engineering by means of incentives has only been studied in context of the BGP ecosystem, thus not accounting for path choice by end-hosts. Given rational ASes, there are different methods to achieve stability for inter-domain traffic: incentive-compatible yet oscillation-free BGP policies [83, 266], egress-router selection under QoS constraints [118], cooperative traffic-engineering agreements between ASes reached by Nash bargaining [235], and the use of prices as traffic-steering incentives [186].

3.9 CONCLUSION

In this chapter, we develop a game-theoretic framework that allows to test path-selection strategies on their viability for selfish end-hosts, i.e., to show whether it is rational for an end-host to adopt a path-selection strategy, given that all other end-hosts use said path-selection strategy. Only strategies that form such equilibria may be adopted in an Internet environment, where end-hosts are self-interested and uncontrolled.

Using this framework, we show that the non-oscillatory path-selection strategies traditionally proposed in the literature are not rational strategies and thus cannot be expected to be adopted by selfish, unrestricted end-hosts. This insight suggests that end-hosts must be incentivized to abstain from oscillatory path selection by means of stabilization mechanisms. We present two such stabilization mechanisms, and prove their incentive compatibility.

This incentive compatibility is a desirable property of a system, for two reasons. First, the selfishness assumption for end-hosts is a pessimistic assumption for networked systems, and designing systems under this assumption limits the possible damage. Second, a system in which users can gain by rule-breaking may be seen as inherently unfair, which limits the adoption of the system. With this work, we make a first step in addressing these concerns regarding path-aware network architectures, which hopefully sparks further research into load-adaptive path selection and its impact on these architectures.

4

IMPACT OF UNSTABLE PATH SELECTION

4.1 INTRODUCTION

As discussed in the preceding chapter, end-host path selection introduces a risk of load oscillation, which is expressed as an alternating grow-and-shrink pattern of traffic volumes on links, and is perceived as an obstacle to the deployment of path-aware networks [72]. Interestingly, this oscillation is targeted by a wide range of solutions for oscillation suppression [79, 85, 140, 196, 2], but relatively little is known about *how exactly* and *by how much* instability from path selection deteriorates network performance. In other words, the solution to the oscillation problem is much clearer than both the impact vectors and the magnitude of this oscillation problem.

In this chapter, we therefore aim at qualifying and quantifying the effects of oscillatory path selection on various metrics of a network. To tackle this challenge, we must take into account that end-points in real path-aware networks employ algorithms which jointly perform path selection and congestion control (CC), i.e., multi-path congestion-control (MPCC) algorithms. In this work, we will focus on MPCC algorithms that are inspired by greedy, myopic path-selection behavior and thus simultaneously produce and react to oscillation.

To analyze these algorithms, we require an analytical approach that (i) captures the congestion-window fluctuations that represent the oscillation, and (ii) is general enough to deliver fundamental insights into the nature of CC-assisted end-host path selection. Alas, fluid models [112, 140, 142, 204, 253] are well suited to represent equilibria in terms of the rough traffic distribution on a network; these models, however, either completely disregard congestion-control dynamics (such as the classic Wardrop model [253]) or fail to capture the small-scale dynamics of congestion-window fluctuations (as noted by Peng et al., who themselves operate with a fluid model [204]). More applied approaches, as employed in the design of

multi-path TCP (MPTCP) [144, 214, 259], can capture oscillatory phenomena (e.g., the ‘flappiness’ of protocols [144]), but these approaches rely on ad-hoc reasoning from stylized network examples and experimental validation, which reduces their viability as generic analytic tools.

We argue that a so-called *axiomatic* approach recently initiated by Zarchy et al. [268] offers both the right analytical resolution and the required generality for the question at hand. This approach is axiomatic in a sense borrowed from economics and game theory, where properties with obvious desirability (e.g., the acyclicity of preferences [36] or the fairness of a bargaining outcome [194]) are formulated as axioms. Zarchy et al. apply this approach to congestion control by capturing desirable properties of CC protocols such as efficiency, fairness, and stability in axioms. The approach allows to analytically rate protocols with respect to these axioms and highlight the fundamental trade-offs between them. In our work, we further extend Zarchy et al.’s model to a multi-path context with the goal of characterizing fundamental properties of joint algorithms for path selection and congestion control.

4.1.1 Contribution

This chapter is based on my publication at PERFORMANCE 2021 [4]. Our work uses a theoretical model to investigate how network performance is affected by the instability due to greedy end-point path selection. In contrast to earlier theoretical models, we develop a model that is able to capture both path-selection dynamics and congestion-window fluctuations in §4.2. Within this model, we identify and formalize different classes of dynamic equilibria (in §4.3 and §4.4) to which the flow dynamics can be expected to converge exponentially fast. These equilibria are essential for the analytical rating of MPCC protocols: In §4.5, we rate these dynamic equilibria with respect to a number of performance metrics (the axioms). These metrics are inspired by the recently developed axiomatic approach to CC [268], but are extended to accommodate path selection. This equilibrium formalization allows to derive the following insights in §4.6:

No trade-off between efficiency, convergence and loss avoidance. Through appropriate protocol tuning, the metrics efficiency, loss avoidance, and convergence can be simultaneously optimized. Hence, there is no trade-off between these properties in theory. Unfortunately, however, such optimization requires knowledge about system parameters such as the number of end points and the bottleneck capacities, making them hard to determine in most practical settings.

Trade-off with fairness and responsiveness. There is, however, a fundamental trade-off between the above metrics and the fairness and the responsiveness of a MPCC protocol. In particular, higher responsiveness makes a protocol less efficient, but more fair.

Effects of introducing end-host path selection. By contrasting the axiomatic performance ratings for a general network with and without path selection, we obtain a multifaceted formalization of the performance impact of introducing end-host path selection. This formalization allows to interpret and quantify how unstable path selection affects network performance depending on network parameters. The insights gained from this approach show that there are both benefits and drawbacks of end-host path selection in an unstable form.

4.2 MODEL AND ASSUMPTIONS

4.2.1 Discrete Model

We leverage the analytical model of congestion control proposed by Zarchy et al. [268] and extend it to a multi-path context with path selection, as illustrated in Fig. 4.1.

Agents, paths, and flows. In summary, N agents (denoted by set $A = [N] := \{0, \dots, N - 1\}$) compete for bandwidth on the bottleneck links of P parallel paths from set Π . Each agent $i \in A$ maintains a congestion window with size $cwnd_i$, which evolves over time t . At each moment $t \in \mathbb{N}_0$ in discrete time, any path $\pi \in \Pi$ accommodates a set $A_\pi(t)$ of agents that use path π at moment t , and carries load $f_\pi(t) = \sum_{i \in A_\pi(t)} cwnd_i(t)$.

Agent behavior. In each time step t , every agent i takes two actions. First, agent i performs *congestion control*, i.e., adapts its congestion-window size $cwnd_i(t)$ according to a chosen CC protocol $CC_i(t, cwnd_i(t))$, resulting in congestion-window size $cwnd_i(t + 1)$. Second, agent i performs *path selection*, i.e., determines the path π such that $i \in A_\pi(t)$ and $cwnd_i(t)$ is included in $f_\pi(t)$, according to a given path-selection strategy. In Fig. 4.1 as well as in our following analysis, agents implement probabilistic greedy path selection, i.e., switch to the path carrying the lowest load in the last time step with a given probability m . Finally, in order to investigate different behaviors for congestion-window adaptation upon path switches,

Table 4.1: Notation used in our model in alphabetic order.

Symbol	Description
$A = [N]$	Set of agents in network
$A_\pi(t)$	Set of agents using path π at time t
$a_\pi(t)$	Number of agents using path π at time t
$\alpha(\tau)$	Additive increase given continuity time τ
β	Multiplicative-decrease parameter
C	Total bottleneck capacity of network
C_π	Bottleneck capacity of path π
$cwnd_i(t)$	Congestion-window size of agent i at time t
$f(t)$	Combined congestion-window size of all agents at time t
$f_\pi(t)$	Combined congestion-window size of all agents using path π at time t
$M_\pi(t)$	Set of agents who migrate away from path π at time t
m	Responsiveness (Probability of switching to more attractive path in each time step)
N	Number of agents in the network
P	Number of paths in the network
Π	Set of paths in the network
$\pi_i(t)$	Path used by agent i at time t
$\pi_{\min}(t)$	Path with lowest utilization at time t
r	Reset softness (Multiplicative decrease of congestion-window size on path switch)
$\text{rank}(\pi, t)$	Rank of path π at time t (Number of paths with higher utilization than π at time t)
τ	Continuity time (time since last loss or path switch)
$\tau_i(t)$	Continuity time of agent i at time t
$z(a_\pi(t), N)$	Scaling factor for extrapolating on-migration flow volume from path flow

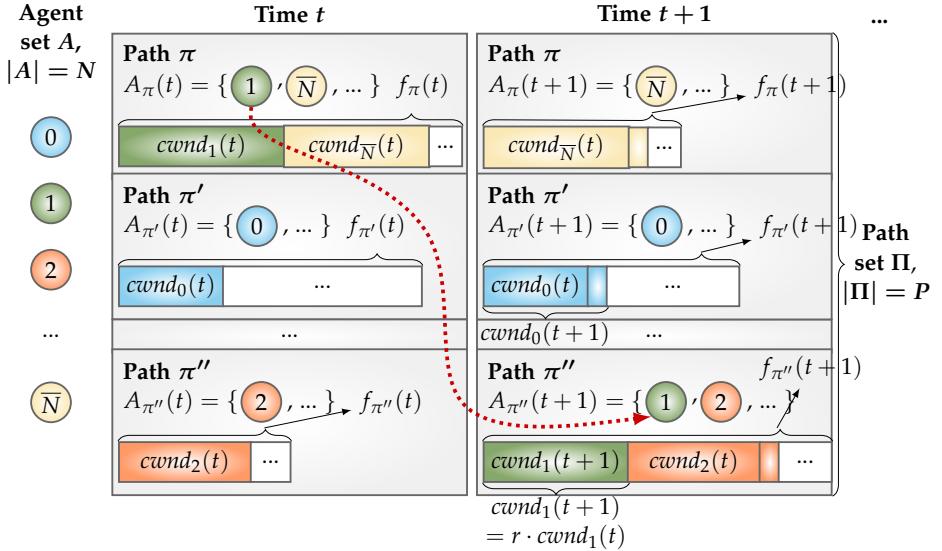


Figure 4.1: Illustration of discrete model (Notation: $\bar{N} = N - 1$). The dotted arrow visualizes path migration by agent 1 from path π to path π'' .

we introduce a reset-softness parameter $r \in [0, 1]$ that determines the extent of congestion-window reduction for path-switching agents (e.g., agent 1 in Fig. 4.1).

The agents are further constrained by path capacities C_π , $\pi \in \Pi$, where C_π is the amount of data in maximum segment size (MSS) that can be transmitted on path π during one round-trip time (RTT). If the capacity C_π of path π is exceeded by the flow $f_\pi(t)$, the agents $A_\pi(t)$ experience packet loss and take this loss into account in their congestion-control protocol.¹ For example, the TCP Reno protocol, with a multiplicative decrease of 0.5 as a reaction to loss and an additive increase of 1 otherwise, is modelled as follows for an agent i using path π at time t :

$$\text{TCPReno}(t, cwnd_i(t)) = \begin{cases} cwnd_i(t) + 1 & \text{if } f_\pi(t) \leq C_\pi \\ 0.5 \cdot cwnd_i(t) & \text{otherwise} \end{cases} \quad (4.1)$$

¹ We note that this loss modelling is a simplification in three respects. First, loss may already occur when $f_\pi(t) > s_\pi$, namely if all agents send out all traffic $f(t)$ in a burst that exceeds the buffer size s_π . Second, even if $f_\pi(t) > C_\pi$, the loss may not be perceived by all agents. Third, CC algorithms may react differently depending on the number of recent losses.

4.2.2 Scenario of Interest and Assumptions

Since the goal of this work is to characterize the worst-case effects of oscillatory path selection, our analysis throughout the chapter will focus on a network scenario that maximizes the severity of load oscillation. This scenario has the following properties, which henceforth serve as assumptions:

Greedy load-adaptive path selection. Oscillation is caused by greedy, myopic path selection behavior [2], which dynamically determines the number $a_\pi(t) = |A_\pi(t)|$ of agents on path π . In any time step t , agents select the path $\pi_{\min}(t)$ with the lowest bottleneck utilization $f_\pi(t)/C_\pi$ and hence the lowest latency (assuming roughly equal propagation delay of all paths as stated below) and lowest loss rate. Since monitoring the state of alternative paths and switching paths consume resources, agents may not consider a path change in every time step. Instead, the path-selection behavior is regulated by a *path-migration probability* $m \in (0, 1]$, i.e., the probability with which an agent switches to a more attractive path in any time step. Alternatively, m is interpreted as a measure for the *responsiveness* of agents.

Sequential multi-path usage. The intensity of oscillations grows with the size of shifted flow volume per time unit. In order to maximize oscillation, we therefore assume that a path-switching agent completely stops using its previously used path and exclusively sends on the newly selected path. This coarse-granular migration behavior produces sequential instead of concurrent usage of multiple paths. This mode of sequential multi-path usage approximates the actual behavior of real-world algorithms such as MPTCP, which tends to use only the most attractive path for data transmission and sends a negligible amount of probing traffic over the alternative paths [140, 144, 259]. Moreover, the average utility improvement per user that is possible by concurrently using multiple paths instead of a single selected path vanishes for a high number of agents [252]. Sequential multi-path usage implies that $\sum_{\pi \in \Pi} a_\pi(t) = N \forall t$.

Disjoint and similar paths. We investigate a network consisting of paths that are parallel (i.e., disjoint) and equal in terms of latency D_π and bottleneck capacity $C_\pi = C/P$, where C is the bottleneck capacity of the complete network. Such a network, while being a simplification of general networks, is likely to bring out the worst-case effects of myopic, greedy path selection, which are the subject of this chapter. In particular, load oscillations are strongest if the actions of the sending agents are strongly correlated because they react to the same (potentially misleading)

feedback signals (i.e., path loss and latency) simultaneously [2]. If agents sharing a link react to different feedback signals or at different times, e.g., because they are using different paths with different round-trip latencies, their actions are less strongly correlated and the flow dynamics are likely to oscillate less. The feedback synchronization by equal path RTTs also ensures that the discrete time steps of the model have consistent duration across all paths.

4.2.3 Stochastic Dynamics

In summary, a multi-path congestion-control protocol $MPCC(CC, m, r)$ is a combination of a CC protocol $CC(t)$, a responsiveness parameter m , and a reset-softness parameter r . In a network with path selection, the MPCC dynamics are thus represented by a pair of functions $(a_\pi(t), f_\pi(t))$ for any path $\pi \in \Pi$. Since the path-selection behavior is probabilistic (regulated by responsiveness parameter m), the MPCC dynamics are not uniquely determined by initial conditions, but need to be modeled as a stochastic process. In particular, the MPCC dynamics under universal adoption of $MPCC(CC, m, r)$, are given by

$$a_\pi(t+1) = \begin{cases} a_\pi(t) - |M_\pi(t)| & \text{if } \pi \neq \pi_{\min}(t) \\ a_\pi(t) + \sum_{\tilde{\pi} \neq \pi} |M_{\tilde{\pi}}(t)| & \text{otherwise} \end{cases} \quad (4.2a)$$

$$f_\pi(t+1) = \begin{cases} \text{if } \pi \neq \pi_{\min}(t): \\ \quad f_\pi(t) - \sum_{i \in M_\pi(t)} cwnd_i(t) + \sum_{i \in A_\pi(t) \setminus M_\pi(t)} \Delta cwnd_i(t) \\ \text{otherwise:} \\ \quad f_\pi(t) + \sum_{\tilde{\pi} \neq \pi} \sum_{j \in M_{\tilde{\pi}}(t)} r \cdot cwnd_j(t) + \sum_{i \in A_\pi(t)} \Delta cwnd_i(t), \end{cases} \quad (4.2b)$$

where $M_\pi(t)$ is a random subset of $A_\pi(t)$, which contains the agents who leave path π at time t , and $\Delta cwnd_i(t) = cwnd_i(t+1) - cwnd_i(t)$ is the growth in congestion-window size of agent i . Intuitively, the flow on a more congested path π is reduced by the congestion windows of all agents M_π that leave the path, and increased by the congestion-window growth of the remaining agents $A_\pi(t) \setminus M_\pi(t)$. In contrast, the flow on the least congested path π is increased by the post-reset congestion-window sizes $r \cdot cwnd_j(t)$ of the agents $j \in M_{\tilde{\pi}}$ who migrate to path π , and by the congestion-window growth of the previously present agents $A_\pi(t)$.

4.2.4 Expected Dynamics

While the formulations in Eq. (4.2) capture the evolutionary dynamics of an MPCC system, their discrete and probabilistic nature hinders analytic treatment. However, as we investigate large-scale systems with a high number of agents, the law of large numbers allows that the probabilistic elements in Eq. (4.2) can be well approximated by their expected values and traffic randomness can be largely ignored. This insight is also frequently used in mean-field analysis [19].

For the remainder of this chapter, we therefore consider the *expected* MPCC dynamics, where the recursion on the random variables $(a_\pi(t), f_\pi(t))$ is approximated with a recursion on the expectations $(\hat{a}_\pi(t), \hat{f}_\pi(t))$ (where we write $\hat{x} := \mathbb{E}[x]$ for any function x). The accuracy of this approximation is validated with simulations in Appendix C.2.

Path selection. Concerning the agent dynamics in Eq. (4.2a), we note that $\mathbb{E}[|M_\pi(t)|] = m \cdot \hat{a}_\pi(t)$ for any path $\pi \neq \pi_{\min}(t)$. Moreover, the expected volume of flow associated with the agents in M_π in Eq. (4.2b) is a proportional share of the expected total flow $\hat{f}_\pi(t)$ on path π :

$$\mathbb{E} \left[\sum_{i \in M_\pi(t)} cwnd_i(t) \right] = m \cdot \hat{f}_\pi(t). \quad (4.3)$$

By the same argument, it holds that

$$\mathbb{E} \left[\sum_{\tilde{\pi} \neq \pi} \sum_{j \in M_{\tilde{\pi}}(t)} r \cdot cwnd_j(t) \right] = m \cdot r \cdot \sum_{\tilde{\pi} \neq \pi} \hat{f}_{\tilde{\pi}}(t) \quad (4.4)$$

for $\pi = \pi_{\min}(t)$. However, in order to make the second case of Eq. (4.2b) independent of flows $f_{\tilde{\pi}}$ on alternative paths, we additionally make the following approximation:

$$\sum_{\tilde{\pi} \neq \pi} \hat{f}_{\tilde{\pi}}(t) \approx \frac{N - \hat{a}_\pi(t)}{\hat{a}_\pi(t)} \cdot \hat{f}_\pi(t) = z(\hat{a}_\pi(t), N) \cdot \hat{f}_\pi(t), \quad (4.5)$$

where $z(\hat{a}_\pi(t), N)$ is henceforth referred to as the *extrapolation factor*. In this approximation, the flow on path π is scaled proportionally to the number of agents $N - \hat{a}_\pi(t)$ on other paths. This approximation can be justified on the grounds that in a steady state, imbalances in path load are likely to stem from imbalances in the number of agents between paths, not from imbalances in the average congestion-window size between paths.

Congestion control. Finally, in order to arrive at the expected flow dynamics $\hat{f}_\pi(t)$, the expected combined congestion-window change $\mathbb{E}[\sum_{k \in A_\pi(t)} \Delta cwnd_k(t)]$ (or for $A_\pi(t+1)$, respectively) must be formalized. Of course, this change depends on the CC protocols employed by the agents. In order to maximize the generality of our analysis, we rely on the following generic form of a loss-based CC protocol employed by each agent i , where $\pi_i(t)$ denotes the path that agent i uses at time t :

$$CC_i(t, cwnd_i(t)) = \begin{cases} cwnd_i(t) + \alpha(\tau_i(t)) & \text{if } f_{\pi_i(t)}(t) \leq C_{\pi_i(t)} \\ \beta \cdot cwnd_i(t) & \text{otherwise} \end{cases} \quad (4.6)$$

Here, $\tau_i(t)$ is the so-called *continuity time* of agent i , i.e., the number of time steps in which agent i has already been on its current path without experiencing packet loss. This continuity time is the argument to a function α , which determines the additive increase to the congestion window in absence of loss. This formulation allows to mimic the window-growth behavior in classic TCP Reno [82], in the widely deployed TCP CUBIC [110], in the slow-start phase of many TCP protocols [243], or in more theoretical MIMD protocols [20]. Finally, $\beta \in [0, 1]$ is a parameter that determines the multiplicative decrease of the congestion-window size in the case of packet loss, which is the predominant practice in loss-based CC protocols.

Based on the probability distribution for the continuity time $\tau_i(t)$ of any agent $i \in A_\pi(t)$ at time t from Appendix C.1, we can calculate the average congestion-window increase per agent, conditioned on the path π used by the agent at time t :

$$\hat{\alpha}_\pi(t) = \sum_{\tau=0}^{\infty} \mathbb{P}[\tau_i(t) = \tau \mid i \in A_\pi(t)] \cdot \alpha(\tau). \quad (4.7)$$

This average congestion-window increase then allows to obtain the aggregate additive increase in absence of loss. In contrast, loss reduces the expected flow volume $\hat{f}_\pi(t)$ through multiplicative decrease β , complementing the effects of out-migration (for $\pi \neq \pi_{\min}(t)$) or in-migration (for $\pi_{\min}(t)$).

Complete expected dynamics. In summary, under universal adoption of a protocol $MPCC(CC, m, r)$, the expected dynamics therefore are:

$$\hat{a}_\pi(t+1) = \begin{cases} (1 - m) \cdot \hat{\alpha}_\pi(t) & \text{if } \pi \neq \pi_{\min}(t) \\ (1 - m) \cdot \hat{\alpha}_\pi(t) + m \cdot N & \text{otherwise} \end{cases} \quad (4.8a)$$

$$\hat{f}_\pi(t+1) = \begin{cases} \text{If } \pi \neq \pi_{\min}(t) \wedge \hat{f}_\pi(t) \leq C_\pi : \\ \quad (1-m) \cdot \hat{f}_\pi(t) + \hat{\alpha}_\pi(t) \cdot (1-m) \cdot \hat{a}_\pi(t) \\ \text{If } \pi = \pi_{\min}(t) \wedge \hat{f}_\pi(t) \leq C_\pi : \\ \quad (1+m \cdot r \cdot z(\hat{a}_\pi(t), N)) \cdot \hat{f}_\pi(t) + \hat{\alpha}_\pi(t) \cdot \hat{a}_\pi(t) \\ \text{If } \pi \neq \pi_{\min}(t) \wedge \hat{f}_\pi(t) > C_\pi : \\ \quad \beta \cdot (1-m) \cdot \hat{f}_\pi(t) \\ \text{If } \pi = \pi_{\min}(t) \wedge \hat{f}_\pi(t) > C_\pi : \\ \quad (\beta + m \cdot r \cdot z(\hat{a}_\pi(t), N)) \cdot \hat{f}_\pi(t) \end{cases} \quad (4.8b)$$

4.2.5 Limitations

While our model presents a tractable approach to analyze oscillatory MPCC dynamics, our investigation and the resulting insights have clear limitations worth addressing in future research. In particular, as our network model is an extension of the network model by Zarchy et al. [268], our work inherits some limitations noted by Zarchy et al., most importantly the assumption of synchronized feedback, and the focus on a specific type of network. However, it is noteworthy that our work addressed the previously identified challenge concerning randomized protocols through the concept of expected dynamics. In general, the comprehensiveness of our analysis would benefit from relaxing the worst-case conditions elicited in §4.2.2, most prominently the assumption of disjoint and similar paths, and from introducing latency-based and model-based CC protocols.

4.3 LOSSLESS EQUILIBRIA

In order to rate MPCC protocols, we focus on the *equilibria* that these protocols induce, i.e., stable load patterns to which the MPCC dynamics from Eq. (4.8) eventually converge. In this section, we characterize one class of equilibria that are attained before the capacity limit of any bottleneck link is exceeded, i.e., these equilibria are *lossless*. Equilibria without this lossless property, i.e., lossy equilibria, are presented in §4.4. All of these equilibria are *dynamic* equilibria, i.e., periodic patterns of the number of agents and the load on the different paths. Note that the insights

regarding equilibria only apply to the theoretical construct of expected dynamics in an exact sense, and only approximately apply to actual MPCC dynamics.

4.3.1 Structure of Lossless Equilibria

In order to characterize lossless equilibria, we need to investigate whether the expected MPCC dynamics tend to exhibit a certain pattern in the case where capacity limits are disregarded. Unfortunately, even this simplified discrete dynamical system (determined by Eq. (4.8) without the two last cases of Eq. (4.8b)) is analytically intractable due to the presence of case distinctions in the evolution functions [89]. Therefore, we use a hybrid approach, similar to previous work [16]: By performing simulations as in Fig. C.2, we arrive at the following two observations about MPCC dynamics with greedy, myopic agents sharing parallel and similar paths (cf. §4.2.2), which serve as a basis for further analytical investigation:

In-migration is utilization-maximizing. Whenever path π with minimal utilization, i.e., $\hat{u}_\pi(t) = \hat{f}_\pi(t)/(C/P)$, experiences in-migration according to the second case of Eq. (4.8b), this path tends to become the most utilized path.²

Out-migration is order-preserving: If two paths π and $\tilde{\pi}$ with $\hat{u}_\pi(t) > \hat{u}_{\tilde{\pi}}(t)$ experience out-migration according to the first case of Eq. (4.8b), it tends to hold that $\hat{u}_\pi(t+1) > \hat{u}_{\tilde{\pi}}(t+1)$.

If the expected dynamics consistently conform to these two observations, they exhibit the following pattern:

Definition 4.1 P-Step Oscillation. MPCC dynamics exhibit P-step oscillation if there exists a time $t_0 \geq 0$ such that

$$\forall T \in \mathbb{N} \geq 0. \quad \text{rank}(\pi, t_0) = p \implies \text{rank}(\pi, t_0 + T) = (p + T) \bmod P, \quad (4.9)$$

where $\text{rank}(\pi, t)$ ranks all paths $\pi \in \Pi$ in descending order according to their utilization at time t :

$$\text{rank}(\pi, t) = p \iff |\{\tilde{\pi} \mid \hat{u}_{\tilde{\pi}}(t) > \hat{u}_\pi(t)\}| = p. \quad (4.10)$$

In P-step oscillation, the assignment of the rank to paths changes in a round-robin fashion, i.e., in any time step t , every path π rises by one rank, except the path with

² This observation suggests that myopic, greedy load-adaptive path selection is not a Nash equilibrium strategy, which has also been demonstrated in Appendix B.1.

rank $P - 1$ (i.e., with the lowest expected utilization), which obtains the lowest rank 0 at time $t + 1$. After P time steps, a path reaches its original place in the ranking order, i.e., $\text{rank}(\pi, t) = \text{rank}(\pi, t + P)$ for all $t \geq t_0$. We present an argument for the prevalence of P -step oscillation in §4.3.3.

4.3.2 Lossless Agent Equilibrium

As this P -step oscillation uniquely determines the least congested path in any time step $t \geq t_0$, this pattern also determines the agent-migration dynamics. Starting from an agent distribution $\{a_\pi(t_0)\}_{\pi \in \Pi}$ at time t_0 , all the paths π with $\text{rank}(\pi, t_0) \neq P - 1$ will experience an outflow of agents (according to case 1 in Eq. (4.8a)) and only the path with rank $P - 1$ experiences an inflow of agents (according to case 2 in Eq. (4.8a)). In a single round of P -step oscillation with start time t_0 , the path $\pi^{(0)}$ with $\text{rank}(\pi^{(0)}, t_0) = 0$ will thus first experience agent outflow for $P - 1$ times and then once experience agent inflow. Hence, the following difference equation characterizes the discrete dynamical system for a granularity of P time steps:

$$\begin{aligned}\hat{a}_{\pi^{(0)}}(t_0 + P) &= (1 - m) \cdot \left((1 - m)^{P-1} \cdot \hat{a}_{\pi^{(0)}}(t_0) \right) + m \cdot N \\ &= (1 - m)^P \cdot \hat{a}_{\pi^{(0)}}(t_0) + m \cdot N,\end{aligned}\tag{4.11}$$

To find an equilibrium of the dynamic system for the agent dynamics on $\pi^{(0)}$, we identify a fixed point of the difference equation in Eq. (4.11), i.e., we solve

$$\hat{a}^{(0)} = (1 - m)^P \cdot \hat{a}^{(0)} + m \cdot N \iff \hat{a}^{(0)} = \frac{m \cdot N}{1 - (1 - m)^P},\tag{4.12}$$

where $\hat{a}^{(0)}$ is the *equilibrium value* for any $\hat{a}_\pi(t)$ with $\text{rank}(\pi, t_0) = 0$, which generalizes as follows:

Insight 4.1 Convergence to Unique Dynamic Agent Equilibrium. Under P -step oscillation, the expected agent dynamics $\{\hat{a}_\pi(t)\}_{\pi \in \Pi}$ of an MPCC system asymptotically converge to a unique dynamic equilibrium, i.e., a cyclic series of states. This dynamic equilibrium of the agent dynamics consists of P states in each of which the rank- p path accommodates the corresponding equilibrium amount of agents $\hat{a}^{(p)}$, i.e.,

$$\hat{a}_\pi(t) = \hat{a}^{(\text{rank}(\pi, t))}, \text{ where } \hat{a}^{(p)} = \frac{(1 - m)^p \cdot m \cdot N}{1 - (1 - m)^P}.\tag{4.13}$$

This convergence can be shown by finding a *trajectory function*:

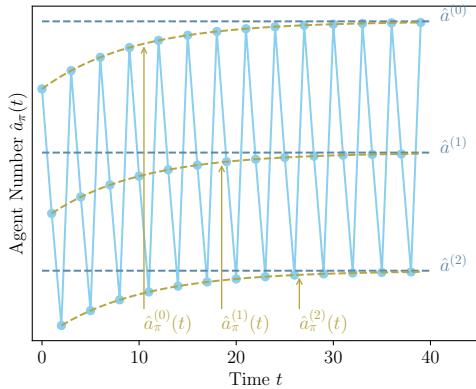


Figure 4.2: Convergence to lossless agent equilibrium (for $P = 3$ and $m = 0.1$).

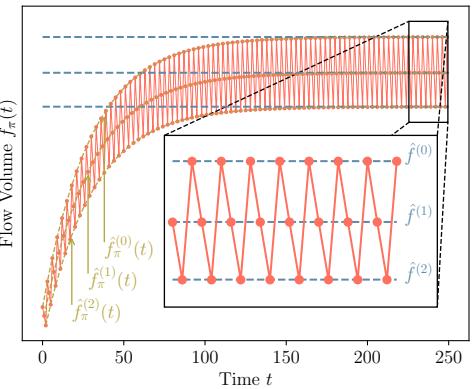


Figure 4.3: Convergence to lossless flow equilibrium (for $P = 3$, $m = 0.1$, and $r = 0.5$).

Definition 4.2 Trajectory Function. A trajectory function $x_\pi^{(p)}(t)$ is an explicit interpolation function that yields the correct value of path-specific dynamics $x_\pi(t)$ at all moments where path π has rank p :

$$\forall k \in \mathbb{N}_{\geq 0}. \quad x_\pi^{(p)}(t_{\pi p} + k \cdot P) = x_\pi(t_{\pi p} + k \cdot P), \quad (4.14)$$

where $t_{\pi p} = \min\{t \mid t \geq t_0 \wedge \text{rank}(\pi, t) = p\}$ is the first time step in which path π has rank p , and t_0 is the start time of P -step oscillation.

For the agent dynamics $\hat{a}_\pi(t)$, such a trajectory function is given by

$$\hat{a}_\pi^{(p)}(t) = (\hat{a}_\pi(t_{\pi p}) - \hat{a}^{(p)}) \cdot (1 - m)^{t - t_{\pi p}} + \hat{a}^{(p)}. \quad (4.15)$$

As $\lim_{t \rightarrow \infty} \hat{a}_\pi^{(p)}(t) = \hat{a}^{(p)}$, the trajectory functions converge to the equilibrium found above exponentially fast. Figure 4.2 visualizes the asymptotic convergence to the dynamic equilibrium $\{\hat{a}^{(p)}\}_{p \in [P]}$ (highlighted in blue) along the trajectory functions.

4.3.3 Lossless Flow Equilibrium

After identifying the agent equilibrium in §4.3.2, we identify the equilibria of the MPCC flow dynamics $\{\hat{f}_\pi(t)\}_{\pi \in \Pi}$ in this section. We first consider hypothetical

equilibria, which are equilibria of the flow dynamics under the assumption that the capacity of each path is never exceeded. In a second step, we will show under which conditions these hypothetical equilibria are actual equilibria.

4.3.3.1 Hypothetical Flow Equilibria

Rank-based dynamics. To find the hypothetical equilibria of the flow dynamics, we can simplify the flow dynamics from Eq. (4.8b) by disregarding the capacity limit C/P . In addition, we insert the equilibrium agent levels $\hat{a}^{(p)}$ from §4.3.2 and the expected additive increase $\hat{\alpha}^{(p)}$ derived in Appendix C.1 to arrive at the following formulation:

$$\hat{f}_\pi(t+1) = \begin{cases} (1-m) \cdot (\hat{f}_\pi(t) + \hat{\alpha}^{\text{rank}(\pi,t)} \cdot \hat{a}^{\text{rank}(\pi,t)}) & \text{if } \text{rank}(\pi,t) \neq P-1, \\ (1+m \cdot r \cdot z(m,P)) \cdot \hat{f}_\pi(t) + \hat{\alpha}^{(P-1)} \cdot \hat{a}^{(P-1)} & \text{if } \text{rank}(\pi,t) = P-1, \end{cases} \quad (4.16)$$

where the extrapolation factor z is only dependent on m and P given the agent equilibrium, i.e., $z(m,P) = N/\hat{a}^{(P-1)} - 1 = (1 - (1-m)^{P-1})/(m \cdot (1-m)^{P-1})$.

Rank-based fixed points. Similar to Eq. (4.11), we set up a first-order difference equation for the dynamics for the path that has rank p at time t_0 (where the P -step oscillation starts) and find a fixed point that is attained every P time steps, for example for ranks 0 and $P-1$:

$$\hat{f}^{(0)} = \frac{((1+m \cdot r \cdot z(m,P)) \cdot (\sum_{p=0}^{P-2} \hat{\alpha}^{(p)}) + \hat{\alpha}^{(P-1)}) \cdot \hat{a}^{(P-1)}}{1 - (1+m \cdot r \cdot z(m,P)) \cdot (1-m)^{P-1}}, \quad (4.17a)$$

$$\hat{f}^{(P-1)} = \frac{(\sum_{p=0}^{P-2} \hat{\alpha}^{(p)} + \hat{\alpha}^{(P-1)} \cdot (1-m)^{P-1}) \cdot \hat{a}^{(P-1)}}{1 - (1+m \cdot r \cdot z(m,P)) \cdot (1-m)^{P-1}}. \quad (4.17b)$$

The fixed point for a general rank p can be derived analogously and expressed by a similar (albeit quite complicated) term $\hat{f}^{(p)}$ shown in Eq. (C.11) in Appendix C.3. These fixed points $\{\hat{f}^{(p)}\}_{p \in [P]}$ constitute the hypothetical equilibrium, i.e., if a rank- p path carries flow volume $\hat{f}^{(p)}$, the path will carry this flow volume again P time steps later, where it is again the rank- p path.

Insight 4.2 Hypothetical Dynamic Flow Equilibrium. If capacity limits of links are disregarded, the dynamic equilibrium of the flow dynamics $\{\hat{f}_\pi(t)\}_{\pi \in \Pi}$ consists of P states in each of which the rank- p path accommodates flow volume $\hat{f}^{(p)}$.

In order for such an equilibrium to be valid, it must be consistent with P -step oscillation, i.e., it must hold that $\hat{f}^{(p)} > \hat{f}^{(p+1)}$ for all $p \in [P - 1]$. Interestingly, if a certain parameter combination is associated with an invalid equilibrium, it follows that P -step oscillation is fundamentally impossible for that parameter combination. However, we show in Appendix C.3 that only a small part of the parameter space, containing rather extreme parameters, is inconsistent with P -step oscillation.

Convergence. Similarly as in §4.3.2, convergence to this equilibrium can be proven using a trajectory function (cf. Definition 4.2). The following trajectory function yields the correct flow volume in all time steps where path π has rank p again:

$$\hat{f}_\pi^{(p)}(t) = (\hat{f}_\pi(t_{\pi p}) - \hat{f}^{(p)}) \cdot ((1 + m \cdot r \cdot z(m, P)) \cdot (1 - m)^{P-1})^{\frac{t-t_{\pi p}}{P}} + \hat{f}^{(p)}. \quad (4.18)$$

The limit of this trajectory function for $t \rightarrow \infty$ is the equilibrium value $\hat{f}^{(p)}$, which establishes convergence;³ this is illustrated in Fig. 4.3.

4.3.3.2 Actual Flow Equilibrium

Intuitively, this hypothetical equilibrium given by $\{\hat{f}^{(p)}\}_{p \in [P]}$ is an actual equilibrium of the MPCC dynamics if the convergence is not disturbed by the capacity limit C/P on any path π , i.e., if the trajectory functions for all ranks consistently remain below C/P . We therefore require an upper bound on all trajectory functions $\{\hat{f}_\pi^{(p)}(t)\}_{p \in [P]}$. Thanks to the structure of P -step oscillation, it holds that $\hat{f}^{(p)} > \hat{f}^{(p+1)} \forall p \in [P - 1]$. Therefore, in the hypothetical equilibrium, $\hat{f}^{(0)}$ represents an upper bound on the flow dynamics. We speak of flow dynamics with *consistent trajectories* if such an ordering not only holds on the equilibrium values $\hat{f}^{(p)}$, but also on the trajectory functions $\hat{f}_\pi^{(p)}(t)$ for all paths π :

³ Note that $\hat{f}^{(0)}$ from Eq. (4.17a) is undefined for $r = 1$, as the flow dynamics do not converge to a fixed point in that case. Given $r = 1$, the trajectory function for rank 0 can be expressed with the following linear function, which has no limit:

$$\hat{f}_\pi^{(0)}(t) = \left[(1 - m)^{1-P} \cdot \left(\sum_{p=0}^{P-2} \hat{a}^{(p)} \right) + \hat{a}^{(P-1)} \right] \cdot \hat{a}^{(P-1)} \cdot P^{-1} \cdot (t - t_{\pi 0}) + \hat{f}_\pi(t_{\pi 0}). \quad (4.19)$$

Definition 4.3 Consistent Trajectories. The flow dynamics $\{\hat{f}_\pi(t)\}_{t \geq 0}$ have consistent trajectories at time point t' if on every path $\pi \in \Pi$, the rank-specific trajectory functions $\{\hat{f}_\pi^{(p)}(t)\}_{p \in [P]}$ satisfy the following condition:

$$\forall p \in [P-1], t \geq t'. \quad \hat{f}_\pi^{(p)}(t) > \hat{f}_\pi^{(p+1)}(t) \quad (4.20)$$

As trajectories are always eventually consistent, it holds that:

$$\begin{aligned} \exists t'. \forall t > t'. \quad \forall p \in [P]. \quad & \hat{f}_\pi^{(0)}(t) \geq \hat{f}_\pi^{(p)}(t) \\ \forall \pi \in \Pi. \quad & \hat{f}_\pi^{(0)}(t) \geq \hat{f}_\pi(t) \end{aligned} \quad (4.21)$$

In other words, the rank-0 trajectory function for any path is eventually an upper bound on all rank-specific trajectory functions and by consequence also an upper bound on the flow dynamics.

As $\hat{f}_\pi^{(0)}(t)$ is monotonic, its function values will not exceed C/P if $\hat{f}_\pi(t_{\pi p}) \leq C/P$ and $\hat{f}^{(0)} \leq C/P$. Due to the introduction of capacity limits, it is necessary to alter the definition of $t_{\pi p}$ to be the first point in time after oscillation began (at t_0) where $\text{rank}(\pi, t_{\pi p}) = p$ (as before) and additionally $\hat{f}_\pi(t_{\pi p}) \leq C/P$.⁴ Therefore, we arrive at the following insight:

Insight 4.3 Dynamic Lossless Flow Equilibrium. The hypothetical equilibrium (disregarding capacity limitations) from Insight 4.2 is an actual, lossless equilibrium (taking capacity limits into account) for the flow dynamics $\{\hat{f}_\pi(t)\}_{\pi \in \Pi}$ if and only if $\hat{f}^{(0)} \leq C/P$, i.e., the maximum flow-equilibrium level does not exceed the bottleneck capacity of any path.

4.4 LOSSY EQUILIBRIA

In this section, we characterize *lossy* equilibria, i.e., dynamic equilibria where $\hat{f}^{(0)} > C/P$ and the flow dynamics therefore periodically exceed bottleneck capacities.

4.4.1 Structure of Lossy Equilibria

Lossy equilibrium types. In order to identify the typical structure of lossy equilibria, we again rely on simulations similar to §4.3.1. Based on these simulations,

⁴ Such a $t_{\pi p}$ always exists as any reasonable CC's reaction to loss reduces $\hat{f}_\pi(t)$ below C/P eventually.

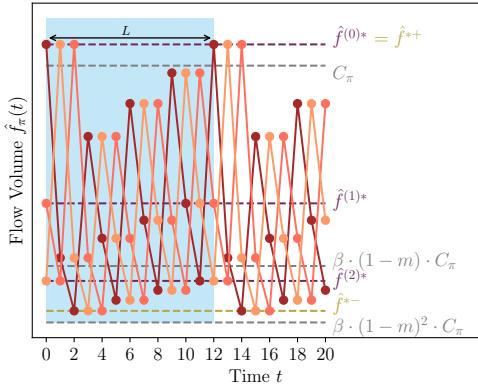


Figure 4.4: Type-1 lossy equilibrium for $P = 3$, $m = 0.45$, and $r = 0.9$ (One period is highlighted in light-blue).

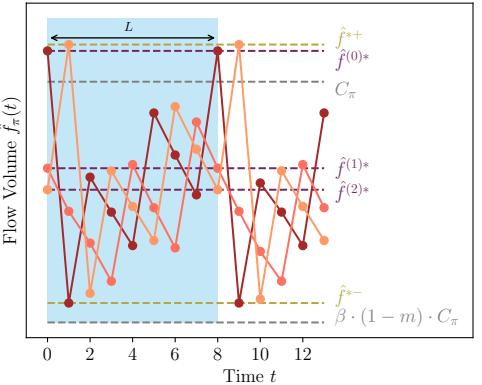


Figure 4.5: Type-2 lossy equilibrium for $P = 3$, $m = 0.1$, and $r = 1$ (One period is highlighted in light-blue).

we can distinguish two types of lossy equilibria, illustrated in Figs. 4.4 and 4.5. Note that both of these lossy-equilibrium types are characterized by flow volumes $\{\hat{f}^{(p)*}\}_{p \in [P]}$, each carried by the path with rank p in the state that is designated as the initial state of the lossy equilibrium ($t = 0$ in the figures) and is periodically revisited every L time steps. Moreover, the boundary points, i.e., the largest and smallest flow volume arising in a lossy equilibrium, are denoted by \hat{f}^{*+} and \hat{f}^{*-} , respectively.

Type distinction. The main distinguishing property of type-1 lossy equilibria (cf. Fig. 4.4) is that these lossy equilibria are consistent with P -step oscillation despite the occasional multiplicative decrease β on rank-0 paths. In contrast, type-2 lossy equilibria (cf. Fig. 4.5) temporarily deviate from P -step oscillation whenever there is packet loss on a path. In that case, the rank-0 path with loss directly becomes the rank- $(P - 1)$ path in the subsequent time step. However, even in type-2 lossy equilibria, P -step oscillation eventually resumes, e.g., at $t = 2$ in Fig. 4.5. Type-1 equilibria typically appear for a relatively high migration rate m , whereas type-2 equilibria tend to appear for lower migration rates.

4.4.2 Flow Equilibria

While the lossy equilibria cannot be characterized as simply as the lossless equilibria in §4.3.3, it is feasible to determine the flow-volume bounds for the presented types of lossy equilibria.

Upper bound. Both for type-1 and type-2 lossy equilibria, we note that the upper-boundary point \hat{f}^{*+} is reached after one round of P -step oscillation starting from a flow volume below the capacity limit. Hence, an upper bound on \hat{f}^{*+} can be represented as follows:

$$\hat{f}^{*+} \leq \hat{f}_\pi^{(0)}(t_{\pi 0} + P) \text{ where } \hat{f}_\pi^{(p)}(t_{\pi 0}) = C/P. \quad (4.22)$$

The trajectory function from Eq. (4.18) (or from Eq. (4.19) for $r = 1$) is used to calculate the effects of one round of P -step oscillation on flow volume C/P , which is the highest flow volume from which ordinary P -step oscillation can proceed. Note that this trajectory function is only usable if the agent dynamics are in equilibrium according to §4.3.2. Type-1 lossy equilibria preserve P -step oscillation and thus also the corresponding agent equilibria. For type-2 lossy equilibria, however, P -step oscillation is occasionally disturbed, which can result in agent dynamics out of equilibrium. However, multiple rounds of P -step oscillation precede the moment of reaching \hat{f}^{*+} and the convergence to the agent equilibrium is exponential. Hence, we observe that the agent dynamics are close to the agent equilibrium and the trajectory function can therefore be used to obtain an approximate upper bound for type-2 lossy equilibria.

Lower bound. In type-1 lossy equilibria, the lower-boundary point \hat{f}^{*-} is reached after $P - 1$ time steps with agent outflow on an overloaded rank-0 path. Combined with the multiplicative decrease β in the first of these $P - 1$ time steps, we can thus formulate a lower bound on \hat{f}^{*-} for type-1 lossy equilibria:

$$\hat{f}^{*-} > \beta \cdot (1 - m)^{P-1} \cdot C_\pi, \quad (4.23)$$

given a rank-0 path that is only infinitesimally overloaded and $\alpha(\tau) > 0, \forall \tau \in \mathbb{N}_{>0}$. For type-2 lossy equilibria, this lower bound is too pessimistic, as the combination of multiplicative decrease and agent out-migration directly transforms the overloaded rank-0 path into the least utilized path and there are no further consecutive time steps with agent out-migration on this path. Hence, $\beta \cdot (1 - m) \cdot C_\pi$ suffices as a lower bound for type-2 lossy equilibria. For a validation of these lower bounds by simulation, consult Fig. C.5 in Appendix C.4.

4.5 AXIOMS

In this section, we use an axiomatic approach inspired by Zarchy et al. [268] to derive insights regarding the effects of oscillatory path selection. We adapt a number of their axioms, which were formulated for a single-path context, to a multi-path context in §4.5.1. In §4.5.2, we evaluate the equilibria from Sections 4.3 and 4.4 with respect to these axioms.

4.5.1 List of Axioms

In our axiomatic approach to multi-path congestion control, axioms correspond to desirable properties that MPCC protocols should possess. However, as these properties refer to general and vague concepts (e.g., efficiency or fairness), the conditions for possessing these properties are usually not well-defined. Therefore, the axioms here are formalized as metrics for rating an MPCC protocol with respect to a certain property, instead of binary indicators of whether the protocol possesses the given property. Concretely, we consider the following axioms in this work:

Axiom 4.1 Efficiency. An MPCC protocol is ϵ -efficient if under universal adoption of this protocol, the bottleneck utilization of every path π with capacity C/P is never lower than a share ϵ after some time t' :

$$\exists t'. \quad \forall t \geq t', \pi \in \Pi. \quad \frac{P \cdot \hat{f}_\pi(t)}{C} \geq \epsilon \quad (4.24)$$

Larger values of ϵ are better, and we consider an ϵ -efficient protocol optimal if $\epsilon \geq (C - s)/C$, where s is the buffer size.^a

^a In terms of latency, $(C - s)/C$, i.e., empty buffers, would even be preferable to higher values of ϵ . This latency effect could be captured by an additional axiom, which we do not introduce in this work.

Axiom 4.2 Loss avoidance. An MPCC protocol is λ -loss-avoiding if under universal adoption, the loss rate on any path π with capacity C/P never exceeds λ after some time t' :

$$\exists t'. \quad \forall t > t', \pi \in \Pi. \quad \frac{\hat{f}_\pi(t)}{C/P} - 1 \leq \lambda \quad (4.25)$$

Thus, smaller values of λ are better, and a 0-loss-avoiding protocol is optimal.

Axiom 4.3 Convergence. An MPCC protocol is γ -convergent if under universal adoption, the flow volume $\hat{f}_\pi(t)$ on every path π lies consistently within a range $[\gamma \cdot \hat{f}_\pi^+, \hat{f}_\pi^+]$ below a path-specific maximum level \hat{f}_π^+ after some time t' :

$$\exists t' > 0, f_\pi^+. \quad \forall t > t', \pi \in \Pi. \quad \gamma \cdot \hat{f}_\pi^+ \leq \hat{f}_\pi(t) \leq \hat{f}_\pi^+ \quad (4.26)$$

Thus, larger values of γ are better, and a 1-convergent protocol is optimal.

Axiom 4.4 Fairness. An MPCC protocol is η -fair if under universal adoption, the variance of congestion-window sizes of all agents $i \in A$ in the network never exceeds η after some time t' .^a

$$\exists t' > 0. \quad \forall t > t'. \quad \text{Var}_{i \in A} [cwnd_i(t)] \leq \eta \quad (4.27)$$

Thus, smaller values of η are better, and a 0-fair protocol is optimal.

^a Zarchy et al. [268] formalize fairness with the ratio of the smallest to the largest congestion-window size in the steady state. Given path selection, this ratio is always potentially 0, e.g., if an agent migrates in every time step.

For any axiom metric μ , we write $\mu(\text{MPCC})$ for the most desirable value of metric μ that the protocol MPCC can be rated with.

4.5.2 Axiomatic Characterization of Equilibria

The axioms in §4.5.1 refer to characteristics which are eventually attained and then persistently preserved by the flow dynamics. Hence, a natural way to axiomatically rate an MPCC protocol is to evaluate the equilibria (i.e., stable states) of this protocol (cf. Sections 4.3 and 4.4).

Efficiency (Axiom 4.1). We distinguish lossless and lossy flow equilibria. If there is a lossless equilibrium ($\hat{f}^{(0)} \leq C/P$), the minimal flow volume ever carried by any path π is the equilibrium value for rank $P - 1$, i.e., $\hat{f}^{(P-1)}$.⁵ The network-wide efficiency level is therefore $\epsilon = P \cdot \hat{f}^{(P-1)}/C$ (Case ①). In contrast, for lossy equilibria, the efficiency level is the lower bound on the lower boundary point \hat{f}^{*-} according to §4.4.2. Depending on the lossy-equilibrium type, this lower bound is given by $\beta \cdot (1-m)^{P-1}$ or $\beta \cdot (1-m)$, respectively. Since the lower bound is never higher for type 1 than for type 2, we consider $\beta \cdot (1-m)^{P-1}$ to be the minimum flow volume for lossy equilibria (Case ②). While this lower bound is too pessimistic for lossy equilibria of type 2, these type-2 lossy equilibria mostly appear for low values of m , where the difference between the two bounds is small.

$$\epsilon(\text{MPCC}(\alpha, \beta, m, r)) \geq \begin{cases} P \cdot \hat{f}^{(P-1)}/C & \text{① if } \hat{f}^{(0)} \leq C/P, \\ \beta \cdot (1-m)^{P-1} & \text{② otherwise.} \end{cases} \quad (4.28)$$

Loss avoidance (Axiom 4.2). If all paths are in lossless equilibrium ($\hat{f}^{(0)} \leq C/P$), it is clear that the maximum loss rate in the whole network is 0 (Case ①). If the network is in lossy equilibrium, the maximum loss rate is determined by the upper boundary point \hat{f}^{*+} (cf. §4.4.2). As shown in Eq. (4.22), this boundary point is maximal at $\hat{f}_\pi^{(0)}(t_{\pi 0} + P)$, where $\hat{f}_\pi^{(0)}$ is the rank-0 trajectory function for an arbitrary path π and is anchored at $\hat{f}_\pi^{(0)}(t_{\pi 0}) = C/P$. For $r \neq 1$ (Case ②) and $r = 1$ (Case ③), this trajectory function is given by Eq. (4.18) and Eq. (4.19), respectively. In summary, the maximum loss rate λ is

$$\lambda(\text{MPCC}(\alpha, \beta, m, r)) \leq \begin{cases} \text{① If } \hat{f}^{(0)} \leq C/P: 0 \\ \text{② If } \hat{f}^{(0)} > C/P \wedge r \neq 1: \\ q(m, r, P) \cdot (1-m)^{P-1} - 1 + \\ (q(m, r, P) \cdot \sum_{p=0}^{P-2} \hat{\alpha}^{(p)} + \hat{\alpha}^{(P-1)}) \cdot \frac{\hat{\alpha}^{(P-1)} P}{C} \\ \text{③ Otherwise:} \\ ((1-m)^{1-P} \cdot \sum_{p=0}^{P-2} \hat{\alpha}^{(p)} + \hat{\alpha}^{(P-1)}) \cdot \frac{\hat{\alpha}^{(P-1)} P}{C} \end{cases} \quad (4.29)$$

where we use the abbreviation $q(m, r, P) := (1 + m \cdot r \cdot z(m, P))$.

⁵ To be precise, the asymptotic convergence to $\hat{f}^{(P-1)}$ permits that $\hat{f}_\pi(t)$ for $\text{rank}(\pi, t) = P - 1$ is consistently below $\hat{f}^{(P-1)}$. However, since this shortfall is infinitesimal, we treat the equilibrium as completely reached instead of only asymptotically approached.

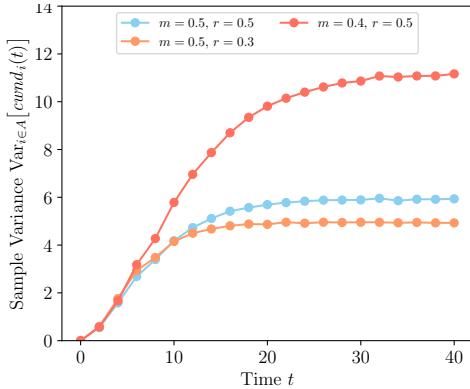


Figure 4.6: Computation of variance in congestion-window size according to the lossless Markov process in Eq. (4.32).

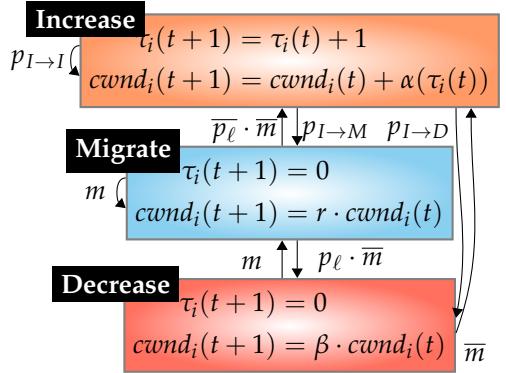


Figure 4.7: Markov process for congestion-window size given lossy equilibria (Notation: $\bar{p} = 1 - p$).

Convergence (Axiom 4.3). If the network is in a lossless equilibrium ($\hat{f}^{(0)} \leq C/P$), the convergence behavior of the flow dynamics can be derived from the boundaries $\hat{f}^{(0)}$ and $\hat{f}^{(P-1)}$ of the hypothetical flow equilibrium (Case ①). Given a lossy equilibrium, we can build on the range between the upper boundary point \hat{f}^{*+} and the lower boundary point \hat{f}^{*-} , for which we have derived an upper and a lower bound, respectively (Case ②). From these ranges, the derivation of the convergence indicator γ and the maximum level \hat{f}_π^+ is straightforward:

$$\gamma(\text{MPCC}(\alpha, \beta, m, r)) \geq \begin{cases} \frac{\hat{f}^{(P-1)}}{\hat{f}^{(0)}} & \text{① if } \hat{f}^{(0)} \leq \frac{C}{\bar{p}}, \\ \frac{\beta \cdot (1-m)^{P-1}}{\lambda(\text{MPCC}(\alpha, \beta, m, r)) + 1} & \text{② otherwise.} \end{cases} \quad (4.30)$$

Fairness (Axiom 4.4). We consider the variance of congestion-window sizes in the equilibrium as a metric for the fairness of an MPCC algorithm:

$$\text{Var}_{i \in A} [cwnd_i(t)] = \mathbb{E}_{i \in A} [cwnd_i(t)^2] - \mathbb{E}_{i \in A} [cwnd_i(t)]^2 \quad (4.31)$$

As the congestion-window evolution of a single agent is a probabilistic process where any state transition only depends on the current state, we approxi-

mate $cwnd_i(t)$ for the case of lossless equilibria by means of the following Markov process with two state variables:

If $\tau_i(t) \neq P - 1$:

$$\tau_i(t+1), cwnd_i(t+1) = \begin{cases} 0, r \cdot cwnd_i(t) & \text{prob. } m \\ \tau_i(t) + 1, cwnd_i(t) + \alpha(\tau_i(t)) & \text{prob. } 1 - m \end{cases} \quad (4.32)$$

Else:

$$\tau_i(t+1), cwnd_i(t+1) = \tau_i(t) + 1, cwnd_i(t) + \alpha(\tau_i(t))$$

where the initial state is given by $\tau_i(0) = cwnd_i(0) = 0$.

Computationally tractable computation of the congestion-window size variance can be done by averaging many simulation samples of the Markov process from Eq. (4.32), which has only linear complexity in t and yields the expectation of the congestion-window size by the central limit theorem. Fig. 4.6 illustrates that the variance of $cwnd_i(t)$ has a limit for $t \rightarrow \infty$.

Regarding lossy equilibria, the Markov process from Eq. (4.32) must be adapted as shown in Fig. 4.7, where the following properties hold:

$$p_{I \rightarrow I} = \text{if } \tau_i(t) \bmod P \neq (P - 1) \text{ then } \bar{p}_\ell \cdot \bar{m} \text{ else } \bar{p}_\ell \quad (4.33)$$

$$p_{I \rightarrow M} = \text{if } \tau_i(t) \bmod P \neq (P - 1) \text{ then } m \text{ else } 0 \quad (4.34)$$

$$p_{I \rightarrow D} = \text{if } \tau_i(t) \bmod P \neq (P - 1) \text{ then } p_\ell \cdot \bar{m} \text{ else } p_\ell \quad (4.35)$$

In particular, we assume that every path encounters loss with probability p_ℓ in any time step, except if the path has experienced loss in the previous time step (as there are no consecutive loss events on the same path in the lossy equilibria in §4.4). If the agent is using a lossy path, but does not leave the path, the congestion-window size is multiplicatively decreased as shown in transition *Decrease* in Fig. 4.7. Like for lossless equilibria, a simulation-based approach enables to efficiently compute the variance in congestion-window size (cf. Fig. C.6 in the appendix). This figure suggests that the variance limit for lossy equilibria is decreasing in loss probability p_ℓ . Moreover, since the lossy Markov process in Fig. 4.7 is equivalent to the lossless Markov process in Eq. (4.32) for $p_\ell = 0$, the variance of the lossless Markov process represents an upper bound on the variance of the lossy Markov process. Therefore, we henceforth exclusively rely on the lossless Markov process.

4.6 AXIOM-BASED INSIGHTS

In this section, we derive fundamental insights into the nature of end-host path selection on the basis of the axioms presented in the previous section. First, we investigate in §4.6.1 how the performance characteristics of a network change if end-host path selection is introduced. Second, we show in §4.6.2 that there are fundamental trade-offs when applying end-host path selection.

4.6.1 Performance Effects of Introducing End-Host Path Selection

4.6.1.1 Evaluation Method

General approach. In order to analyze how end-host path selection affects the performance characteristics of a network, we use a comparative approach: First, we characterize the performance of a network without end-host path selection based on the axioms from §4.5.1 (henceforth: Scenario (I)). Afterwards, we compare the axiomatic ratings of the network without path selection to the axiomatic characterization of the MPCC equilibria (cf. §4.5.2) that arise in the same network given end-host path selection (henceforth: Scenario (II)).

Evaluation network. We base the comparison on a network with N agents and a total bottleneck capacity C distributed over P paths with equal bottleneck capacity C/P . All agents adopt the same CC protocol $CC_i(\alpha, \beta)$ (cf. Eq. (4.6)) in Scenario (I), whereas they employ a multi-path version $MPCC_i(\alpha, \beta, m, r)$ of this CC protocol in Scenario (II). Moreover, while the agent distribution on paths is dynamically determined in Scenario (II), the agent distribution in Scenario (I) is static: To identify the worst-case effects of end-host path selection, let this static agent distribution be optimal, i.e., $a_\pi = N/P$. Finally, we consider two different additive-increase functions: We distinguish a constant additive-increase function $\alpha_1(\tau) = 1$ and an additive-increase function α_S in the style of TCP Slow Start: $\alpha_S(\tau) = 2^\tau$ if $\tau < 5$ and $\alpha_S(\tau) = 1$ otherwise.

Evaluation figures. The changes in evaluation metrics are also visualized in Figs. 4.8 and 4.9: For any m and every equilibrium class (lossless or lossy), the possible range of the metric change is shown for the two additive-increase functions and an example network. The range associated with each value of m is $[\min_{r \in R(m)} \Delta(m, r), \max_{r \in R(m)} \Delta(m, r)]$, where Δ is the difference metric as a func-

tion of m and r , and $R(m)$ contains all values of r that produce a valid equilibrium of the given class (lossless or lossy) in the example network given m .

4.6.1.2 Efficiency (Axiom 4.1).

Given that the employed protocol $CC_i(\alpha, \beta)$ eventually exhausts the capacity of any path, the efficiency level is given by the lowest possible flow volume that results from loss. This lower bound is determined by the multiplicative decrease β applied to a flow volume that is infinitesimally above the capacity limit:

$$\forall \pi \in \Pi. \quad \epsilon(CC_i(\alpha, \beta)) = \frac{\beta \cdot C/P}{C/P} = \beta \quad (4.36)$$

We now compare this efficiency level to the MPCC efficiency levels from Eq. (4.28) and analyze the efficiency change $\Delta\epsilon = \epsilon(MPCC(\alpha, \beta, m, r)) - \epsilon(CC(\alpha, \beta))$ that is due to the introduction of end-host path selection. For a visualization of this efficiency change, consider Fig. 4.8a.

Lossless equilibria. If the efficiency level of the MPCC dynamics is determined by a lossless equilibrium, then $\Delta\epsilon$ is given by $P \cdot \hat{f}^{(P-1)} / C - \beta$. As $\hat{f}^{(P-1)}$ is a decreasing function of the migration rate m and an increasing function of the reset softness r , end-host path selection is more likely to negatively affect ϵ for high migration rates and hard resets on path switch:

Insight 4.4 Efficiency Effects of Path Migration and Resets in Lossless Equilibria. *The more likely agents are to migrate away from a path at any single point in time, the further the bottleneck-link utilization can drop, and if agents start out with a small congestion window every time they switch to a new path, utilization (and therefore efficiency) are relatively low.*

Nonetheless, it is possible that the introduction of end-host path selection leads to a higher level of efficiency. The computations for the example network, visualized in Fig. 4.8a, show that for low values of m and high values of r , introducing end-host path selection can increase efficiency.

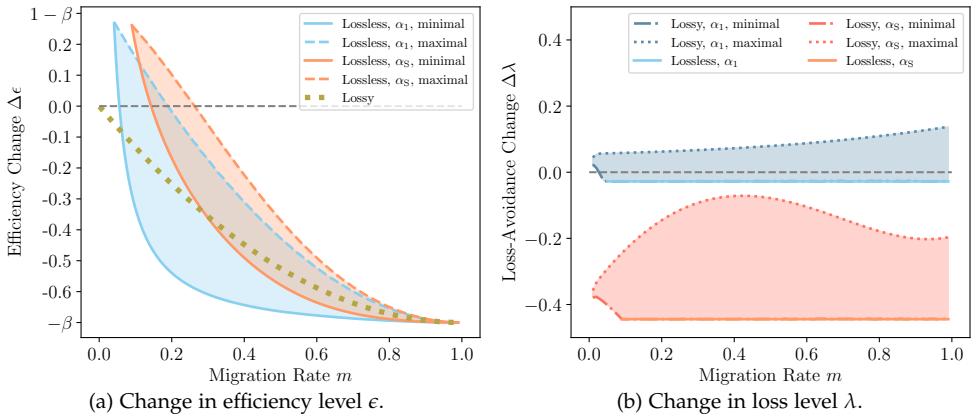


Figure 4.8: Effects of end-host path selection for an example network with $P = 3$, $\beta = 0.7$, $N = 1000$, and $C/P = 12000$.

Lossy equilibria. In contrast, if the MPCC efficiency level is determined by a lossy equilibrium, then $\Delta\epsilon$ is given by $\beta \cdot (1-m)^{P-1} - \beta$, which is bound to be negative. This fact allows the interpretation that end-host path selection strictly lowers the efficiency in case of loss, as emigration from a path reinforces the utilization plunge created by the CC loss reaction, i.e., the multiplicative decrease β . As Fig. 4.8a shows, such less efficient lossy equilibria are bound to exist for low values of m , for which there is no value of r such that a lossless equilibrium can arise. This insight points to a subtle relationship between migration rates and efficiency:

Insight 4.5 Inefficient Equilibria due to Low Migration. While lowering the migration rate can increase the efficiency of end-host path selection, very low migration rates necessarily lead to inefficient (lossy) equilibria, which make end-host path selection detrimental to efficiency compared to a scenario without path selection.

4.6.1.3 Loss avoidance (Axiom 4.2).

In Scenario I, the worst-case loss rate occurs if flow f_π on path π is exactly at the capacity limit C_π , and there is an additional increase by all agents on the path:

$$\forall \pi \in \Pi. \quad \lambda(CC_i(\alpha, \beta)) = \frac{\alpha^{\max} \cdot a_\pi}{C_\pi} = \frac{\alpha^{\max} \cdot N}{C}, \quad (4.37)$$

where $\alpha^{\max} = \max_{\tau \in \mathbb{N}_{\geq 0}} \alpha(\tau)$ to represent the maximum possible loss.

Lossless equilibria. In case of lossless equilibria of the MPCC dynamics, it is clear that $\Delta\lambda$ (defined analogously to $\Delta\epsilon$) is negative, i.e., the loss rate can be reduced (to 0). This improvement in $\Delta\lambda$ is shown in Fig. 4.8b for all values of m for which there is a value of r such that a lossless equilibrium arises.

Lossy equilibria. If a lossy equilibrium is present, the effects of end-host path selection are more ambivalent. In that case, the maximum loss rate in the path-aware network is proportional to $\hat{f}^{(0)}$: the larger the hypothetical limit value $\hat{f}^{(0)}$ of the trajectory function, the stronger the increase of the trajectory function at level C_π and thus the higher the loss rate. As $\hat{f}^{(0)}$ is proportional to r and effectively infinite for $r = 1$, the highest loss rate for every value of m is achieved for $r = 1$, which yields the following intuitive insight:

Insight 4.6 Loss Effects of Soft Resets. *If agents only perform soft resets of the congestion-window size when switching paths, high loss may arise on the newly selected path.*

In contrast, if m and r are such that the equilibrium is only marginally lossy, i.e., $\hat{f}^{(0)}$ is only infinitesimally larger than C_π , then the maximum loss rate in a lossy equilibrium is arbitrarily close to 0 (similar to a lossless equilibrium). However, a value of r that achieves $\hat{f}^{(0)} \approx C_\pi$ may not exist given a (low) value of m ; in this case, the reduction of the loss rate to 0 is not possible. Therefore, we arrive at a counter-intuitive insight that mirrors Insight 4.5:

Insight 4.7 Loss Effects of Low Migration. *Loss is not minimized by minimizing the migration rate m , as low migration rates may prohibit the emergence of completely lossless equilibria.*

Loss effects of additive-increase functions. Fig. 4.8b allows another non-obvious insight:

Insight 4.8 Loss Effects of Path Selection with Variable Additive-Increase Functions. *The benefits of end-host path selection in terms of loss are particularly large if additive-increase functions with high inherent variability (such as α_S in Fig. 4.8) are used by the agents.*

In that case, end-host path selection may reduce loss because it leads to desynchronization of the continuity time τ between agents: If all agents tend to

have the same continuity time τ , there is a chance that many agents have continuity time τ_{\max} with $\alpha(\tau_{\max}) = \alpha^{\max}$ at the same time, resulting in high loss. In contrast, agent migration due to path selection causes more heterogeneity in τ and therefore leads to an averaging of $\alpha(\tau)$, which reduces the aggregate additive increase and therefore the maximum possible loss. While this observation may first seem like an unfair comparison of a maximum to an average, the averaging of additive increases is exactly a fundamental feature of path selection. In particular, the averaging reduces the possible maximum of aggregate additive increase compared to a scenario without path selection.

4.6.1.4 Convergence (Axiom 4.3).

The convergence level γ is determined by the minimum and the maximum possible flow volume, as derived above:

$$\gamma(CC_i(\alpha, \beta)) = \frac{\beta \cdot C}{C + \alpha^{\max} \cdot N} \quad (4.38)$$

Lossless equilibria. In the case of lossless equilibria, end-host path selection can increase stability if

$$\frac{\hat{f}^{(P-1)}}{\hat{f}^{(0)}} = \frac{\sum_{p=0}^{P-2} \hat{\alpha}^{(p)} + \hat{\alpha}^{(P-1)} \cdot (1-m)^{P-1}}{(1+m \cdot r \cdot z(m, P)) \cdot \sum_{p=0}^{P-2} \hat{\alpha}^{(p)} + \hat{\alpha}^{(P-1)}} > \gamma(CC_i(\alpha, \beta)), \quad (4.39)$$

which is unsurprisingly true for a low migration rate m and hard resets $r \approx 0$.

Lossy equilibria. However, analogously to efficiency and loss, convergence surprisingly suffers from very low migration rates m , as this causes lossy equilibria, which are inferior to lossless equilibria in terms of convergence (cf. Fig. 4.9a).

The convergence in these lossy equilibria benefits from low migration rates and hard resets, without the exception for very low migration rates that exists for lossless equilibria. While such lossy equilibria might be inferior to lossless equilibria in terms of convergence, lossy equilibria of end-host path selection might still be preferable to no end-host path selection at all, as Fig. 4.9a suggests for the lossy equilibria for α_S . Similar to Insight 4.8, the reason for this improvement is the de-synchronization of the continuity time brought about by agent migration, which reduces the variance of the aggregate additive increase and thus the flow-volume fluctuations. Contrary to the widespread belief that end-host path selection necessarily hurts stability (in the sense of the convergence axiom), our analysis thus shows that network stability can in fact benefit from end-host path selection.

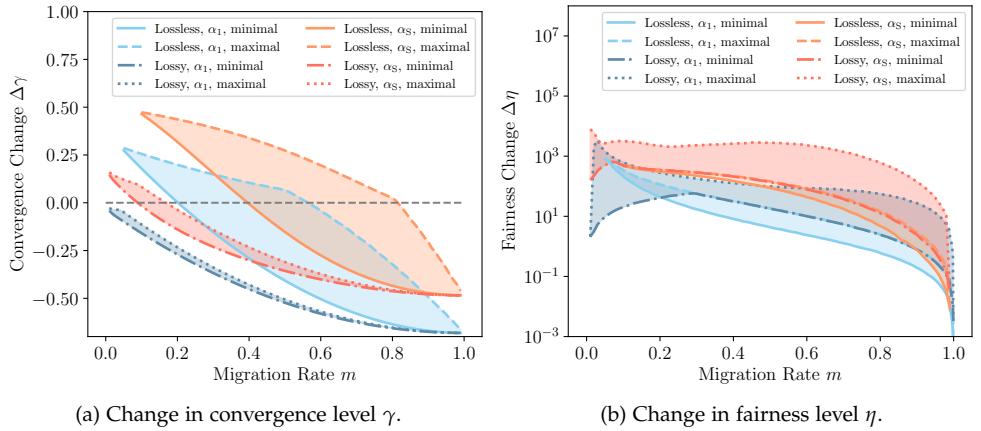


Figure 4.9: Illustration of effects of end-host path selection on the basis of the same example network as in Fig. 4.8.

4.6.1.5 Fairness (Axiom 4.4).

Fairness without path selection. Given simultaneous sending start and no path selection, perfect synchronization implies that all agents always have exactly the same congestion-window size, i.e., $\eta = 0$. Moreover, Zarchy et al. show that even if some agents start sending after others, CC protocols generally tend to come close to perfect fairness [268]. To find the worst-case effects of end-host path selection, we thus assume perfect fairness in the scenario without path selection:

$$\eta(CC_i(\alpha, \beta)) = 0 \quad (4.40)$$

Fairness effects of migration. Hence, the fairness change $\Delta\eta$ due to end-host path selection is equal to the unfairness level η of the MPCC dynamics, which has been computed as a function of the migration rate m in Fig. 4.9b for two different additive-increase functions. In Fig. 4.9b, the lowest values for η , i.e., the highest fairness, is achieved for very high migration rates $m \approx 1$, which leads to the following insight:

Insight 4.9 Fairness Effects of Path Migration. *In a system with greedy end-host path-selection, a very high migration rate m leads to optimal fairness.*

This phenomenon can be intuitively explained as follows: If the migration rate is high, any agent is likely to reset its congestion-window size in any time step,

which results in a compact distribution of the congestion-window size. Under a low migration rate, some agents may reach a high congestion-window size due to uninterrupted growth, while a few agents per time step perform a reset, which leads to a high variance of the congestion-window size distribution.

Fairness effects of resets. The effects of the reset softness r on η are more nuanced. As Fig. 4.9b shows, the unfairness metric η is generally higher for lossy equilibria, which appear for high reset softness, than for lossless equilibria, i.e., lossless equilibria are fairer. However, as mentioned in §4.5.2, the fairness metric for lossy equilibria is computed for infinitesimal loss probability p_ℓ ; for any higher p_ℓ , η is lower, which complicates a comparison to lossless equilibria. Also, for a low migration rate m , lossy equilibria with a high reset softness are associated with lower η than lossless equilibria. The reason behind this phenomenon is that soft resets reduce the difference in congestion-window size between the agents that have not migrated in a long time (and therefore have a large congestion-window) and the agents that have recently migrated and reset their congestion-window size.

Fairness over time. Finally, while end-host path selection seems to reduce fairness as captured by η , we note that η only represents the fairness at any single point in time. However, under low migration rates, there may still be very high *inter-temporal* fairness. If the migration probability is low, any agent has a high probability to uninterruptedly grow its congestion window for a long time. If the congestion-window sizes of any agent were averaged over a certain time span, the distribution of such average congestion-window sizes would have low variance. We leave this more complex fairness analysis as an interesting task for future work.

4.6.2 Fundamental Trade-Offs

In Sections 4.5.2 and 4.6.1, the dependency of the MPCC dynamics on the migration rate (or responsiveness) m and the reset softness r has been qualified and quantified.

Trade-off regarding migration rate. These characterizations allow to observe the following trade-off in the design of systems with end-host path selection:

Insight 4.10 Efficiency and Convergence vs Fairness and Responsiveness. Efficiency ϵ and convergence γ are more favorable under low migration rates, whereas fairness η and responsiveness m are more favorable under high migration rates, implying a fundamental trade-off between these axioms.

However, we note that this trade-off is only valid *within* equilibrium classes, e.g., for comparing lossless equilibria among each other, but not *across* equilibrium classes: Lowering the migration rate below a certain (low) level restricts the set of possible equilibria to lossy equilibria, which are worse in terms of efficiency and convergence than lossless equilibria (cf. Insight 4.5).

Migration and loss. Regarding loss avoidance, the effect of migration rates depends on the remaining network parameters. If resets are hard ($r \approx 0$), higher migration rates are associated with lower loss rate (as higher migration rates make lossless equilibria more likely, which are optimal in terms of loss). In contrast, if resets are soft ($r \approx 1$), lossless equilibria are impossible and the effects of the migration rate on the loss rate are unclear in general, because the migration rate non-monotonically affects the aggregate additive increase (cf. the curve for maximal $\Delta\lambda$ given α_S and lossy equilibria in Fig. 4.8b). However, this unpredictable effect vanishes for constant additive-increase functions (such as α_1 from §4.6.1). For constant-increase functions, a higher migration rate leads to a *higher* loss rate given soft resets. This finding underlines the relevance of congestion-window adaptation on path switch:

Insight 4.11 Reset Dependence of Migration Effect on Loss. While a high migration rate decreases loss under hard resets, it increases loss under soft resets.

Optimal MPCC parameters. Despite this subtle relationship of migration rates and the axiomatic metrics, we can identify parameters m and r that are optimal with respect to all the metrics efficiency, loss, and convergence simultaneously. These parameters are given by the lowest m such that a lossless equilibrium is still possible given a complete reset $r = 0$. These parameters yield a lossless equilibrium with high efficiency and convergence (cf. Insight 4.10).

Insight 4.12 No Trade-Off between Efficiency, Loss Avoidance and Convergence. Since there exist protocol parameters that are optimal with respect to efficiency, loss avoidance, and convergence simultaneously, there exists no fundamental trade-off between these metrics.

Unfortunately, determining these optimal parameters requires knowledge about specific and variable properties of the target network, i.e., the number of agents $|A|$ and the path-bottleneck capacities C_π in the network, making it unattainable in most practical settings.

4.7 RELATED WORK

End-host path selection. Traditionally, the effects of end-host path selection have been theoretically studied in the literature on *selfish routing*. In this line of research, the classic Wardrop model [253] is used to characterize stable traffic distributions (equilibria) that result from uncoordinated path selection by self-interested agents. These equilibria have been thoroughly investigated with respect to their existence [218, 222], their efficiency (typically termed *Price of Anarchy* [148, 210, 221, 224]), and their convergence properties [85, 227]. However, the Wardrop model cannot represent congestion-control dynamics appropriately, which we consider important for characterizing the impact of end-host path selection on network performance.

The effects of end-host path selection have also been characterized by Wang et al. [252], whose ‘cost of not splitting in routing’ captures the difference in network utility between a scenario where end-hosts select a single path and a scenario where multiple paths can be selected. However, this work differs from ours in investigating static rate allocations instead of dynamic rate evolution, in evaluating a single metric (utility) instead of multiple axioms, and in contrasting different modes of end-host path selection instead of contrasting path selection with path pre-determination.

Methods for MPCC analysis. In research about multi-path congestion control, there has been widespread use of fluid models which can better represent congestion-control dynamics [112, 140, 142, 204]. However, also these models focus on representing equilibria in terms of approximate traffic distributions on networks and do not capture small-scale dynamics such as congestion-window fluctuations. More applied approaches rely on reasoning from network examples and experimental validation and have been used in the design of MPTCP algorithms such as LIA [214],

259] and OLIA [144]. These approaches are rather suited for the design of concrete protocols than for the elicitation of fundamental properties of end-host path selection. Moreover, MPTCP research typically only investigates the effects of path selection by scrutinizing friendliness concerns between single-path and multi-path TCP users in the same network, not by looking at the impact that the introduction of end-host path selection has on aggregate performance based on various metrics.

Axiomatic reasoning. In contrast, the axiomatic approach used in this chapter allows to qualify and quantify the performance impact of path selection on a fundamental level while taking congestion-control dynamics into account. Thanks to this power, the axiomatic perspective has been applied to various topics beyond game theory: In computer science, for example, research on congestion control [268], routing protocols [161], and recommendation systems [21] have benefited from axiom-based approaches.

4.8 CONCLUSION

Motivated by a stability concern about end-host path selection, we qualify and quantify the performance impact of such path-selection-induced instability in this chapter. More precisely, we analyze a general network in which end-hosts employ greedy load-adaptive path selection and characterize the resulting traffic pattern with respect to five metrics of interest (“axioms”): efficiency, loss avoidance, convergence, fairness and responsiveness. Through this analysis, we show how the performance impact of end-host path selection depends on the path-migration behavior, the underlying congestion-control protocol, and the structure of the network. Among the dependencies that we present and explain, there are both intuitive, well-known dependencies (e.g., high migration rates decrease efficiency) and non-intuitive, more complex dependencies (e.g., very low migration rates increase loss). Moreover, we show that there are fundamental limitations such that no multi-path congestion-control protocol can optimize all metrics simultaneously.

MODELLING BBR DYNAMICS

5.1 INTRODUCTION

To this day, ever changing applications, traffic patterns, network capacities, and path types prompt research into new and better congestion-control algorithms (CCAs). Most prominent in recent years was Google’s introduction of BBR [47], which was promptly enabled in 2017 for some Google services and thus widely deployed in the public Internet [48]. Since then, several theoretical and experimental studies of the behavior of BBR [75, 119, 231, 246, 254] have identified issues with this first version of BBR, relating to both fairness (especially towards loss-based CCAs) and efficiency (e.g., excessive queue buildup). As a result, BBRv2 [50] has been proposed, triggering another series of evaluation studies [101, 143, 193, 240].

Still, the characterization of BBR performance remains incomplete. Experiment-based studies [101, 119, 143, 193, 231, 246], by their nature, allow statements relating to the concrete network settings in the experiments. Given the variety of scenarios in which a CCA might be deployed, such experimental investigations could only be

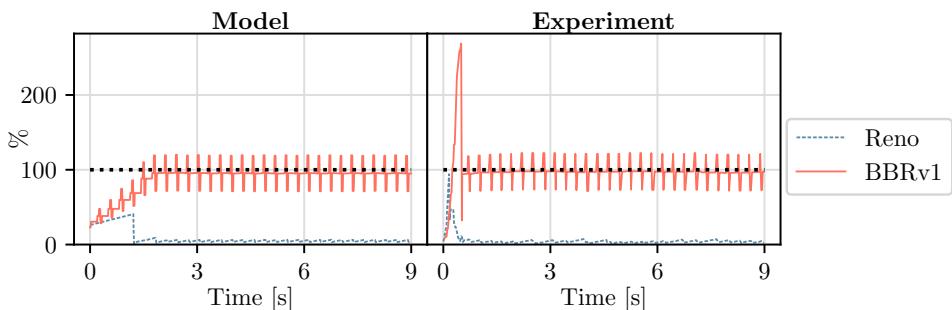


Figure 5.1: Competition of sending rates (in % of link bandwidth) between a Reno flow and a BBRv1 flow.

made exhaustive with great effort and large-scale testbeds which only a minority of researchers has access to. Previous model-based studies contain BBR steady-state models that are valuable for specific settings (e.g., deep buffers [255] or wireless links [264]); however, a deep theoretical understanding requires a model that allows investigation of general settings and the convergence process.

In this chapter, we complement previous approaches to BBR analysis with a classic approach in CCA research: fluid models consisting of differential equations [86, 163, 166, 182, 215, 241, 249]. Such fluid models are unique in their suitability for *both* efficient simulation and theoretical stability analysis. Enabling efficient simulation is critical because the model must be simulated under a plethora of configurations, including settings that are expensive to build. Enabling theoretical stability analysis is crucial because the equilibria (i.e., steady states) of the CCA dynamics are only relevant for performance characterization if stable in a control-theoretic sense, i.e., if the dynamics actually converge to the equilibria.

While a fluid model for BBR is thus well-suited to complement previous work, constructing such a model is challenging because BBR does not naturally fit into the existing fluid-model framework for loss-based CCAs [166, 249]. In fact, BBR does not exclusively rely on a congestion window affected by loss, but includes traffic pulses for capacity probing and measurement-driven state transitions. By using new techniques, e.g., by mimicking the probing pulses with sigmoid functions, this work establishes the first highly accurate and highly general model of BBR, both for versions 1 and 2.

Our fluid model predicts BBR behavior with high accuracy, which we validate with experiments with the network emulator mininet [157]. The validated model confirms BBR performance issues from previous studies and yields new insights. Moreover, we apply dynamical-system analysis (i.e., Lyapunov method) to our fluid model to identify asymptotically stable equilibria of the BBR dynamics.

5.1.1 Contributions

This chapter is based on my publication at IMC 2022 [5]. Its main contributions are the following:

BBR fluid model. We introduce the first general fluid model for BBR (versions 1 and 2), using new techniques such as sigmoid pulses and mode variables.

Validated BBR analysis. We present extensive, systematic model-based calculations, experimentally validate the results of these calculations, and provide profound

insights into fundamental metrics of BBR. For example, we analytically confirm the previous insight that BBRv1 can lead to unfair bandwidth allocations—especially when it competes against loss-based CCAs, but also in competition with itself given unequal RTTs. Moreover, we confirm that BBRv2 eliminates most of the undesirable behavior of BBRv1, but we also identify settings in which BBRv2 leads to bufferbloat and unfairness.

Stability analysis. We analytically identify asymptotically stable equilibria of BBRv1 and BBRv2.

5.2 NETWORK FLUID MODEL

In this section, we present our network fluid model, which closely follows the work by Low et al. [166]. However, we have made several improvements to the network model, which we will highlight in the following. We denote function $f(t)$ by f and its derivative by \dot{f} , unless the argument differs from the default time variable t .

In our model, the network consists of links ℓ with capacity C_ℓ , buffer size B_ℓ , and propagation delay d_ℓ .

Link-arrival rate. The arrival rate $y_\ell(t)$ at link ℓ is

$$y_\ell = \sum_{i \in U_\ell} x_i \left(t - d_{i,\ell}^f \right), \quad (5.1)$$

where U_ℓ is the set of agents using link ℓ , $x_i(t)$ is the sending rate of agent i at time t , and $d_{i,\ell}^f$ is the propagation delay from agent i to link ℓ . Together with Low et al. [166], we neglect queuing delay and packet losses previous to link ℓ .

Queue length. In general, the queue length grows or shrinks according to the discrepancy between combined arrival rate y_ℓ and the transmission capacity C_ℓ at the respective link [166], but never exceeds its buffer size B_ℓ :

$$\dot{q}_\ell = (1 - p_\ell) \cdot y_\ell - C_\ell, \quad q_\ell(t) \in [0, B_\ell], \quad (5.2)$$

where $p_\ell(t)$ is the loss probability of link ℓ at time t (cf. §5.2). We refine the model by Low et al. to additionally capture the effect of packet drops on the queue length.

Table 5.1: Notation used in network model.

Symbol	Description
B_ℓ	Buffer size at link ℓ
C_ℓ	Link capacity of link ℓ
d_ℓ	Propagation delay of link ℓ
$d_{i,\ell}^f$	Forward propagation delay from agent i to link ℓ
$d_{i,\ell}^b$	Backward propagation delay from link ℓ to agent i (over destination and return segment of path)
d_i^p	Total propagation delay for agent i ; $d_i^p = d_{i,\ell}^f + d_{i,\ell}^b$
π_i	Path used by agent i ; modeled as set of links
$\pi_{i,\ell}$	Path from agent i to link ℓ
U_ℓ	Set of agents using link ℓ
$p_\ell(t)$	Loss probability at link ℓ at time t
$p_\pi(t)$	Loss probability on path π at time t
$q_\ell(t)$	Queue length at link ℓ at time t
$\tau_\ell(t)$	Latency (propagation and queueing delay) on link ℓ
$\tau_\pi(t)$	Round-trip latency (propagation and queueing delay) of path π
$\tau_\ell(t)$	Latency of link ℓ
$w_i(t)$	Congestion-window size of agent i at time t
$x_i(t)$	Sending rate of agent i at time t
$y_\ell(t)$	Arrival rate at link ℓ at time t

Latency. The link latency is the fixed link propagation delay plus the queuing delay, which depends on queue size $q_\ell(t)$. The latency of a path is the sum of link latencies:

$$\tau_{\pi_i} = \sum_{\ell \in \pi_i} \tau_\ell = \sum_{\ell \in \pi_i} d_\ell + \frac{q_\ell}{C_\ell}. \quad (5.3)$$

Loss probability and queuing disciplines. Without active queuing discipline, loss occurs if the buffer of a link is full. Given such a simple *drop-tail* policy, the loss probability is given by the relative excess rate whenever the queue is full, and is 0 otherwise [258]. To facilitate analytical treatment, we refine previous models by a smooth approximation:

$$p_\ell(t) = \sigma(y_\ell(t) - C_\ell) \cdot \left(1 - \frac{C_\ell}{y_\ell}\right) \cdot \left(\frac{q_\ell}{B_\ell}\right)^L \quad (5.4)$$

where $L \gg 1$ and $\sigma(v)$ is a relatively sharp sigmoid function:

$$\sigma(v) = \frac{1}{1 + e^{-K \cdot v}} \quad (5.5)$$

with $K \gg 1$ controlling the sharpness of the increase at $v = 0$.

In contrast to drop-tail, the loss probability under the RED queuing discipline moves synchronously with the queue size. More precisely, RED keeps the drop probability at 0 if the queue size is below a configurable threshold q_0 , increases the drop probability linearly to a configurable value p_1 for queue sizes up to q_1 , and drops all packets for larger queue sizes. We approximate the RED behavior as follows, representing the general idea of RED:

$$p_\ell = \frac{q_\ell}{B_\ell} \in [0, 1]. \quad (5.6)$$

which corresponds to a RED configuration with $q_0 = 0$, $q_1 = B_\ell$, and $p_1 = 1$. The extension of the fluid model to other RED configurations is straightforward.

Regarding the loss probability of paths, link-specific loss probabilities are assumed to be small enough such that the following approximations regarding loss hold:

$$p_{\pi_i}(t) = 1 - \prod_{\ell \in \pi_i} \left(1 - p_\ell(t + d_{i,\ell}^f)\right) \approx \sum_{\ell \in \pi_i} p_\ell(t + d_{i,\ell}^f). \quad (5.7)$$

Congestion window and sending rate. For the window-based congestion-control algorithms Reno and CUBIC (cf. Appendix D.1), the sending rate of agent i is determined by the congestion-window size w_i and round-trip latency:

$$x_i = \frac{w_i}{\tau_i}. \quad (5.8)$$

5.3 BBR FLUID MODEL

In this section, we introduce the first fluid model for BBR, both for BBRv1 [47] and BBRv2 [50]. Interestingly, the fluid-model techniques used for the loss-based CCAs cannot reflect essential BBR features, in particular its phases with different behavior. Hence, we construct our BBR model using new techniques, i.e., periodic probing pulses and mode variables (for simulating the BBR state machine). In the following, we first describe the behavior of BBR for both versions 1 and 2. Then, we present our fluid model for BBR by means of a basic fluid model, which can be concretized for each version.

5.3.1 Description of BBR

Fundamentally, BBR continuously performs measurements to estimate two core properties of the network path, namely the bottleneck bandwidth BtlBw and the minimal round-trip time (RTT) RTprop (i.e., propagation delay). To estimate these properties, BBR constantly switches between two states, namely the `ProbeBW` state and the `ProbeRTT` state. While the `ProbeBW` state consumes most of flow lifetime and is considerably different across the two BBR versions, the `ProbeRTT` state is only infrequently and briefly entered and is mostly identical across both BBR versions.

ProbeRTT state. BBR enters the `ProbeRTT` state if no smaller round-trip time than the existing `RTprop` estimate is observed for 10 seconds. To discover the propagation delay, the `ProbeRTT` state tries to eliminate queuing delay by restricting the data in flight (*inflight* in BBR terminology) to a small volume during 200 ms. In BBRv1, this small volume has a fixed size of 4 segments; since this volume has been found to be too conservative, the `ProbeRTT` *inflight* limit in BBRv2 has been chosen to half the estimated bandwidth-delay product, i.e., half the product of BtlBw and RTprop .

ProbeBW state in BBRv1. The `ProbeBW` state aims at measuring the bottleneck bandwidth of the network path, and includes a periodic probing strategy with the

pacing rate as the primary control of the sending rate. In this probing strategy, each period consists of 8 phases with the duration of RTprop. In one phase randomly chosen from the first 7 phases, BBRv1 sets its pacing rate to $5/4 \cdot \text{BtlBw}$ to find the capacity limit of the path. In the subsequent phase, BBRv1 decreases its pacing rate to $3/4 \cdot \text{BtlBw}$ to drain the queues potentially built up during the aggressive previous phase. In the other 6 phases of the period, BBRv1 paces at rate BtlBw. At the period end, the maximum delivery rate from the period is then considered the new bottleneck bandwidth estimate and thus serves as a base pacing rate for the next period.

BBRv1 congestion window. BBRv1 also maintains a congestion window, which amounts to twice the estimated BDP and was intended as a safeguard against ‘common network pathologies’ such as delayed ACKs [47]. Contrary to design intention, this inflight limit by the congestion window is the essential constraint on the sending rate of BBRv1 when competing with loss-based CCAs given large buffers [119, 254], letting BBRv1 degenerate into a window-based CCA in these circumstances.

ProbeBW state in BBRv2. This unintentional relevance of the inflight limit in some circumstances, plus the unfairness towards loss-based CCAs in shallow buffers, led Google to revise the ProbeBW mechanism for BBRv2. This revision mainly aimed at making BBR less aggressive, through increasing its sensitivity to loss and ECN signals (where we henceforth only consider loss for simplicity), less frequent probing, and a persistent coupling between inflight limits and the sending rate. To be precise, BBRv2 tries to obtain additional bandwidth only every few seconds, where the time between such probings is given by the minimum of 62 estimated RTTs (chosen for fairness reasons) and a random value between 2 and 3 seconds. In this probing, BBRv2 first paces at the rate given by BtlBw for one RTprop, with the goal to achieve an inflight corresponding to the bandwidth-delay product. Then, BBRv2 sets its pacing rate to $5/4 \cdot \text{BtlBw}$ and increases the inflight until it reaches $5/4$ of the estimated BDP or the loss rate exceeds 2%. At this point, the bottleneck-bandwidth estimate BtlBw is updated to the maximum delivery rate from the last two ProbeBW periods. Moreover, BBRv2 also records the maximum tenable inflight in state variable `inflight_hi`, which tracks the observed inflight, but is reduced by a multiplicative decrease β if the exponential-increase phase has been terminated by excessive loss. Afterwards, BBRv2 chooses a pacing rate of $3/4 \cdot \text{BtlBw}$ until the inflight is reduced to an arguably safe level, which corresponds to the minimum of the estimated BDP and 85% of the previously measured `inflight_hi` (where the

erased 15% are termed **headroom** in BBRv2). Once the inflight has been reduced to that level, BBRv2 enters into *cruising* mode. In cruising mode, BBRv2 aims to keep its inflight on a safe level by introducing the additional inflight bound **inflight_lo**, which is activated if packet loss occurs: **inflight_lo** starts from the congestion-window size at the moment of loss and is reduced by β upon packet loss. In contrast to **inflight_hi**, which serves as a *long-term* inflight bound, **inflight_lo** serves as a *short-term* inflight bound and is therefore reset at the end of the bandwidth-probing period. In summary, at any point in time, the congestion-window size of a BBRv2 flow is the minimum of the general BBR congestion window of two BDP, the long-term bound **inflight_hi** (discounted by **headroom** in cruising mode), and the short-term bound **inflight_lo** (if activated).

5.3.2 Basic fluid model for BBR

We rely on a skeleton fluid model that captures the common properties of BBRv1 and BBRv2.

MinRTT estimate. As mentioned in the previous section, the two versions of BBR are mostly similar regarding the estimation of the minimum RTT given by **RTprop**, which we represent with variable $\tau_i^{\min}(t)$ for the **RTprop** estimate of agent i at time t . The variable τ_i^{\min} is continuously adjusted downwards upon encountering smaller RTTs:

$$\dot{\tau}_i^{\min} = -\Gamma \left(\tau_i^{\min}(t) - \tau_i(t - d_i^p) \right) \quad (5.9)$$

where $\Gamma(v)$ is a differentiable function approximating the ReLU function $\max(0, v)$. Such a function can be constructed using the sigmoid function from Eq. (5.5):

$$\Gamma(v) = v \cdot \sigma(v). \quad (5.10)$$

In Eq. (5.9), this formulation of Γ leads to a proportional decrease in minimum RTT estimate τ_i^{\min} if the currently observed delay $\tau_i(t - d_i^p)$ is below the previously observed minimum τ_i^{\min} , i.e., if the argument of Γ exceeds 0. Otherwise, τ_i^{\min} is preserved.

ProbeRTT state. To describe that BBR is in ProbeRTT state, we use a *discrete mode variable* m_i^{prt} , which is 1 if BBR is in ProbeRTT state and 0 otherwise. In both BBR

Table 5.2: Notation used in BBR fluid model (in alphabetical order).

Symbol	Description
Γ	Continuous approximation of ReLU function
m_i^{crs}	Indicator of flow i being in cruising mode (BBRv2)
m_i^{dwn}	Indicator of flow i currently reducing inflight volume (BBRv2)
m_i^{prt}	Indicator of flow i being in ProbeRTT mode
$\Phi_i(t, \phi)$	Pulse function activated in bandwidth-probing phase ϕ of flow i
σ	Sigmoid function
T_i^{pbw}	Duration of bandwidth-probing period for flow i
t_i^{prt}	Time spent by flow i in current bandwidth-probing period
T_i^{prt}	Time between entries/exits of ProbeRTT mode for flow i
t_i^{prt}	Time spent by flow i since last entry/exit of ProbeRTT mode
τ_i^{\min}	Minimum-RTT estimate of flow i
v_i	Inflight data volume of flow i
w_i^-	Inflight-draining target of flow i (BBRv2)
w_i^{hi}	inflight_hi limit of flow i (BBRv2)
w_i^{lo}	inflight_lo limit of flow i (BBRv2)
w_i^{pbw}	Inflight limit of flow i in ProbeBW mode
w_i^{prt}	Inflight limit of flow i in ProbeRTT mode
x_i^{btl}	Bottleneck-bandwidth estimate of flow i
x_i^{dlv}	Delivery rate measured by flow i
x_i^{\max}	Maximum delivery rate measured by flow i during probing period
x_i^{pbw}	Sending rate of flow i in ProbeBW mode
x_i^{pcg}	Pacing rate of flow i

versions, the ProbeRTT mode is switched on or off upon time-out of the ProbeRTT timer t_i^{prt} :

$$\Delta m_i^{\text{prt}} = \sigma \left(t_i^{\text{prt}} - T_i^{\text{prt}} \right) \cdot \left((1 - m_i^{\text{prt}}) - m_i^{\text{prt}} \right) \quad (5.11)$$

where T_i^{prt} is the time period between entries and exits of the ProbeRTT state for agent i . Note that Eq. (5.11) represents an update rule for simulations rather than a differential equation, as m_i^{prt} is discrete. Upon time-out of the current ProbeRTT timer (i.e., $\sigma \approx 1$), Eq. (5.11) leads to an inversion of m_i^{prt} :

$$m_i^{\text{prt}} + \Delta m_i^{\text{prt}} = \begin{cases} 1 + 1 - 2 \cdot 1 = 0 & \text{if } m_i^{\text{prt}} = 1 \\ 0 + 1 - 2 \cdot 0 = 1 & \text{if } m_i^{\text{prt}} = 0. \end{cases} \quad (5.12)$$

The two time-related variables in Eq. (5.11) behave as follows:

$$T_i^{\text{prt}} = m_i^{\text{prt}} \cdot 0.2 + (1 - m_i^{\text{prt}}) \cdot 10 \quad (5.13)$$

$$t_i^{\text{prt}} = 1 - \sigma \left(t_i^{\text{prt}} - T_i^{\text{prt}} \right) \cdot t_i^{\text{prt}} - \sigma \left(\tau_i^{\min} - \tau_i(t - d_\ell) \right) \cdot t_i^{\text{prt}} \quad (5.14)$$

The constants in Eq. (5.13) cause BBR to remain in ProbeRTT state for 0.2 seconds and to wait 10 seconds before re-entering the state after exiting it. Eq. (5.14) causes a reset of the ProbeRTT timer to 0 if the timer has reached the limit T_i^{prt} or a lower RTT has been measured, and to tick up otherwise.

State-specific sending rate. In ProbeRTT state, the sending rate is limited by a version-dependent inflight limit $w_i^{\text{prt}}(t)$:

$$x_i = m_i^{\text{prt}} \cdot \frac{w_i^{\text{prt}}}{\tau_i} + (1 - m_i^{\text{prt}}) \cdot x_i^{\text{pbw}} \quad (5.15)$$

where the ProbeBW sending rate x_i^{pbw} follows the relevant constraint (congestion window or pacing rate):

$$x_i^{\text{pbw}} = \min \left(\frac{w_i^{\text{pbw}}}{\tau_i}, x_i^{\text{pcg}} \right) \quad (5.16)$$

ProbeBW state. Similar to the ProbeRTT state, we also introduce the two time-related variables T_i^{pbw} and t_i^{pbw} for the ProbeBW state, where T_i^{pbw} is the duration of a ProbeBW period and t_i^{pbw} is the time within the current period. While T_i^{pbw} is version-dependent, t_i^{pbw} grows with time and is reset to 0 when exceeding the period duration for both BBR versions:

$$t_i^{\text{pbw}} = 1 - \sigma(t_i^{\text{pbw}} - T_i^{\text{pbw}}) \cdot t_i^{\text{pbw}} \quad (5.17)$$

In the ProbeBW state, the bottleneck-bandwidth estimation is based on measurements of the delivery rate x_i^{dlv} (with link ℓ being the bottleneck link of agent i):

$$x_i^{\text{dlv}} = \frac{x_i(t - d_i^{\text{p}})}{y_{\ell}(t - d_{i,\ell}^{\text{b}})} \cdot \begin{cases} C_{\ell} & \text{if } q_{\ell}(t - d_{i,\ell}^{\text{b}}) > 0 \\ y_{\ell}(t - d_{i,\ell}^{\text{b}}) & \text{otherwise} \end{cases} \quad (5.18)$$

where d_i^{p} is the propagation delay of flow i , and $d_{i,\ell}^{\text{b}}$ is the propagation delay from link ℓ to sender i (via the destination host). As a result, the fraction in Eq. (5.18) denotes the share of flow i 's traffic, emitted one RTT before time t , among the aggregate traffic simultaneously arriving at link ℓ .

We accommodate the recorded maximum delivery rate $x_i^{\max}(t)$ per ProbeBW period as follows:

$$\dot{x}_i^{\max} = \Gamma(x_i^{\text{dlv}} - x_i^{\max}) - \sigma(0.01 - t_i^{\text{pbw}}) \cdot x_i^{\max} \quad (5.19)$$

where the second term provokes a reset of x_i^{\max} in the first ten milliseconds of the period. The mechanism for adjusting the bottleneck-bandwidth estimate x_i^{btl} (corresponding to BtlBw) to x_i^{\max} is specific to each BBR version.

Inflight volume. Finally, we choose the following natural formulation to model the inflight volume $v_i(t)$:

$$\dot{v}_i = x_i - x_i^{\text{dlv}} \quad (5.20)$$

5.3.3 BBRv1 Fluid Model

Given the basic BBR fluid-model framework, the biggest challenge in modelling BBRv1 is to model the randomized probing behavior with varying pacing rates.

BtlBw update. As described in §5.3.1, BBRv1 proceeds in bandwidth-probing periods that are 8 phases long, where each phase has a duration of τ_i^{\min} , i.e., $T_i^{\text{pbw}} = 8 \cdot \tau_i^{\min}$. The bottleneck-bandwidth estimate x_i^{btl} is updated to the maximum delivery rate x_i^{\max} at the end of the period, which we formalize as follows:

$$\dot{x}_i^{\text{btl}} = \sigma(t_i^{\text{pbw}} - T_i^{\text{pbw}} + 0.01) \cdot (x_i^{\max} - x_i^{\text{btl}}) \quad (5.21)$$

Pacing rate. In general, BBRv1 prescribes a pacing rate x_i^{pcg} equal to x_i^{btl} in each phase, but increases x_i^{pcg} to $5/4 \cdot x_i^{\text{btl}}$ in one randomly chosen phase and decreases it to $3/4 \cdot x_i^{\text{btl}}$ in the subsequent phase. To restrict a given behavior to a certain phase $\phi \in \{0, \dots, 7\}$, we introduce the following *pulse* function Φ , which is 1 if BBRv1 is in phase ϕ and 0 otherwise:

$$\Phi_i(t, \phi) = \sigma(t^{\text{pbw}}(t) - \phi \cdot \tau_i^{\min}) \cdot \sigma((\phi + 1) \cdot \tau_i^{\min} - t^{\text{pbw}}) \quad (5.22)$$

This pulse function allows to model the pacing behavior of an agent i that employs the augmented pacing rate in phase ϕ_i :

$$x_i^{\text{pcg}} = x_i^{\text{btl}} \cdot \left(1 + \frac{1}{4} \cdot \Phi_i(t, \phi_i) - \frac{1}{4} \cdot \Phi_i(t, \phi_i + 1)\right) \quad (5.23)$$

In the implementation of BBRv1, the phase ϕ_i is randomly chosen from $\{0, \dots, 6\}$ every time BBRv1 switches from ProbeRTT state back to ProbeBW state. Since such randomness is incompatible with the determinism of fluid models, we mimic the randomness of ϕ_i by choosing it as $i \bmod 6$, where we assume the agent identifier i to be a natural number. This agent-dependent choice of ϕ_i desynchronizes the pacing-rate variation of agents i on paths with equal RTT, which is the central goal of the randomization, without sacrificing the determinism of the fluid model. The interplay of BBRv1 variables in pacing-based mode is visualized in Fig. 5.2a.

Congestion window. The basic BBR fluid-model allows a straightforward integration of the state-dependent inflight limits of BBRv1:

$$w_i^{\text{prt}} = 4 \quad w_i^{\text{pbw}} = 2 \cdot \bar{w}_i = 2 \cdot x_i^{\text{btl}} \cdot \tau_i^{\min} \quad (5.24)$$

where \bar{w}_i denotes the BDP estimated by agent i .

5.3.4 BBRv2 Fluid Model

BBRv2 mostly differs from BBRv1 with regard to the structure of the bandwidth-probing phase in several ways.

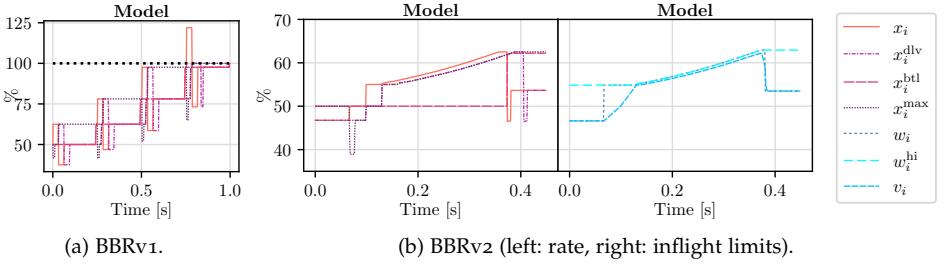


Figure 5.2: Visualization of BBR fluid-model variables link capacity normalized to 100%, single flow)

ProbeBW duration. In BBRv2, a bandwidth-probing period is considerably longer than in BBRv1: The duration of bandwidth-probing periods in BBRv2 is given by the minimum of 62 estimated RTTs and a random value between 2 and 3 seconds. This randomness in BBRv2 poses a similar challenge as the randomness in BBRv1, such that we again use an approach based on the agent identifier to achieve the central goal of agent desynchronization without sacrificing determinism:

$$T_i^{\text{pbw}} = \min \left(62 \cdot \tau_i^{\min}, 2 + \frac{i}{N} \right) \quad (5.25)$$

ProbeBW behavior. The behavior in the bandwidth-probing phases of BBRv2 differs from BBRv1. To model the BBRv2 phases, we introduce two additional mode variables, namely $m_i^{\text{dwn}}(t)$, which indicates whether agent i is attempting to reduce its inflight at time t , and $m_i^{\text{crs}}(t)$, which indicates whether agent i is cruising at time t . The mode variable m_i^{dwn} affects the pacing rate x_i^{pcg} as follows:

$$x_i^{\text{pcg}} = x_i^{\text{btl}} \cdot \left(1 + \frac{1}{4} \cdot \sigma \left(t_i^{\text{pbw}} - \tau_i^{\min} \right) \cdot \left(1 - m_i^{\text{dwn}} \right) - \frac{1}{4} \cdot m_i^{\text{dwn}} \right) \quad (5.26)$$

where m_i^{dwn} increases the pacing rate to $5/4 \cdot x_i^{\text{btl}}$ if $m_i^{\text{dwn}} = 0$ (and one RTT has passed in the bandwidth-probing period), and decreases the pacing rate to $3/4 \cdot x_i^{\text{btl}}$ if $m_i^{\text{dwn}} = 1$.

ProbeBW phase transition. While we modelled phase transitions in BBRv1 as purely dependent on time t_i^{pbw} , the phase transitions in BBRv1 are triggered by probing observations. In particular, the inflight-reducing mode m_i^{dwn} is activated if the

inflight v_i exceeds $5/4 \cdot \bar{w}_i$ or loss p_{π_i} exceeds 2%, and is disabled once the reduced pacing rate has reduced the inflight v_i to the draining target w_i^- , i.e.,

$$w_i^- = \min \left(\bar{w}_i, 0.85 \cdot w_i^{\text{hi}} \right), \quad (5.27)$$

where w^{hi} is the variable accommodating the long-term bound `inflight_hi`:

$$\begin{aligned} \Delta m_i^{\text{dwn}} = & (1 - m_i^{\text{crs}}) \cdot (1 - m_i^{\text{dwn}}) \cdot \sigma(t_i^{\text{pbw}} - \tau_i^{\text{min}}) \\ & \cdot \min(\sigma(v_i - 5/4 \cdot \bar{w}_i) + \sigma(p_{\pi_i} - 0.02), 1) \\ & - m_i^{\text{dwn}} \cdot \sigma(w_i^- - v_i) \end{aligned} \quad (5.28)$$

Moreover, the disabling of m_i^{dwn} automatically leads to the activation of m_i^{crs} , which is then disabled again when a new bandwidth-probing period starts:

$$\Delta m_i^{\text{crs}} = -\Delta m_i^{\text{dwn}} - \sigma(t_i^{\text{pbw}} - T_i^{\text{pbw}}) \cdot m_i^{\text{crs}} \quad (5.29)$$

BtlBw update. The fourth difference concerns the adjustment of the bottle-neck-bandwidth estimate x_i^{btl} . In BBRv2, x_i^{btl} is adjusted to the maximum delivery rate from the last two probing periods when the inflight-growing phase has stopped:

$$\dot{x}_i^{\text{btl}} = m_i^{\text{dwn}} \cdot \left(\max(x_i^{\text{max}}, x_i^{\text{max}}(t - T^{\text{pbw}})) - x_i^{\text{btl}} \right) \quad (5.30)$$

Additional state variables. BBRv2 operates with another two additional state variables, namely `inflight_hi` and `inflight_lo`, which we accommodate in our fluid model with w_i^{hi} and w_i^{lo} , respectively. The upper inflight bound w^{hi} is exponentially adjusted upwards when it represents the relevant constraint on the sending rate ($v_i = w_i^{\text{hi}}$) during the aggressive probing phase and no excessive loss occurs. In contrast, `inflight_hi` is reduced by a multiplicative decrease of 30% if encountering loss exceeding 2%. To be precise, the BBRv2 implementation applies this multiplicative decrease at most once per bandwidth-probing period. We approximate this behavior with a reduced multiplicative decrease in presence of excessive loss:

$$\begin{aligned} \dot{w}_i^{\text{hi}} = & (1 - m_i^{\text{crs}}) \cdot \sigma(t_i^{\text{pbw}} - \tau_i^{\text{min}}) \cdot \sigma(v_i - w_i^{\text{hi}}) \cdot 2^{t_i^{\text{pbw}} / \tau_i^{\text{min}}} \\ & - \sigma(p_{\pi_i} - 0.02) \cdot \frac{0.3}{\tau_i^{\text{min}}} \cdot w_i^{\text{hi}} \end{aligned} \quad (5.31)$$

Outside of cruising mode, the lower inflight bound w_i^{lo} is unset (which we represent with an assimilation to w_i^-). In cruising mode, w_i^{lo} is also decreased by 30% per RTT upon encountering loss:

$$\dot{w}_i^{\text{lo}} = (1 - m_i^{\text{crs}}) \cdot (w_i^- - w_i^{\text{lo}}) - m_i^{\text{crs}} \cdot \sigma(p_{\pi_i}) \cdot \frac{0.3w_i^{\text{lo}}}{\tau_i^{\min}}. \quad (5.32)$$

In summary, the congestion-window size in ProbeBW state is given as follows in BBRv2:

$$w_i^{\text{pbw}} = \min \left(2 \cdot \bar{w}_i, (1 - m_i^{\text{crs}}) \cdot w_i^{\text{hi}} + m_i^{\text{crs}} \cdot w_i^{\text{lo}} \right) \quad (5.33)$$

ProbeTT congestion window. A final difference between BBRv1 and BBRv2 concerns the congestion-window size in ProbeRTT mode. Instead of using a fixed congestion-window size of 4 segments, BBRv2 cuts the congestion window to half the estimated BDP in this mode:

$$w_i^{\text{prt}} = \frac{\bar{w}_i}{2} = \frac{2 \cdot x_i^{\text{btl}} \cdot \tau_i^{\min}}{2}. \quad (5.34)$$

The interplay of the variables in the BBRv2 fluid model is visualized by means of an example in Fig. 5.2b.

5.4 EXPERIMENTAL VALIDATION

In this section, we experimentally validate our BBR fluid model, building on the network emulator mininet [157].

5.4.1 Validation Set-up

5.4.1.1 Model-Based Computations

The implementation of the fluid models uses NumPy [199] and is available online [229]. Differential equations are solved with the method of steps [81, §1.1.2] with a step size of $10\ \mu\text{s}$.

5.4.1.2 Experiments

To compare the model output with implementation behavior, we perform experiments using the network emulator mininet [157]. In mininet, we use OvS to emulate

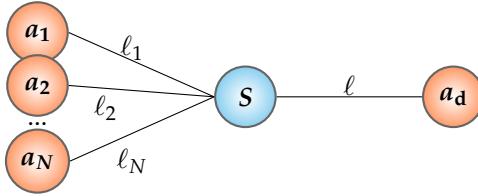


Figure 5.3: Dumbbell topology.

switches [87]. The emulated hosts send traffic by using iPerf [175]. All experiments are run with an Intel Core Intel Xeon E5-2695 v4 CPU.

5.4.1.3 Topology

As usual in the literature [101, 119, 143, 193, 231, 240, 254], we consider the dumbbell topology in Fig. 5.3. In this topology, N agents $a_i, i \in \{1, \dots, N\}$, communicate with a destination host a_d via a switch S . In all paths, the shared link ℓ between switch S and destination host a_d constitutes the bottleneck link. The links $\ell_i, i \in \{1, \dots, N\}$, which connect the individual senders to switch S , are never saturated and therefore do not affect the sending rates. The propagation delays of these non-shared links are heterogeneous (randomly selected from a given range) such that the individual senders experience different RTTs. Switch S is equipped with a buffer, the size of which is measured in bandwidth-delay product (BDP) of the bottleneck link ℓ .

5.4.2 Validation of Trace Results

Network parameters. Using the validation set-up described in §5.4.1, we first verified traces as predicted by the fluid models regarding their similarity with traces from experimental measurements. The concrete network setting in this validation included a single sender, a bottleneck link ℓ with a rate of $C_\ell = 100$ Mbps (as recommended by mininet [156]) and with a propagation delay of $d_\ell = 10$ ms, a non-bottleneck link ℓ_1 with a delay of $d_{\ell_1} = 5.6$ ms, and a switch buffer of 1 BDP.

Validation figures. Figs. 5.4 and 5.5 visualize the comparison of the thus obtained traces for the BBRv1 and BBRv2, respectively, where each CCA was tested under both a drop-tail and a RED queuing discipline (A validation for Reno and CUBIC can be found in Appendix D.1). All measurements have been normalized: The sending rate is given in percent of the bottleneck-link rate, the queue in percent of buffer volume, loss in percent of traffic volume, the RTT as relative excess delay, and

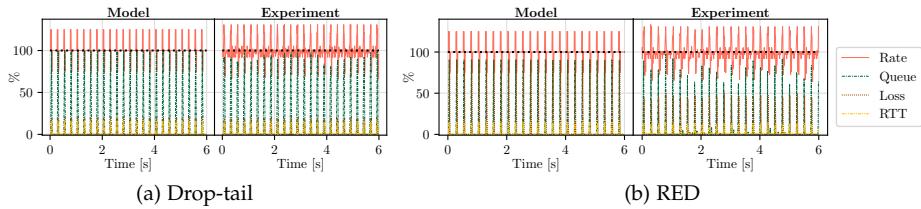


Figure 5.4: BBRv1 trace validation.

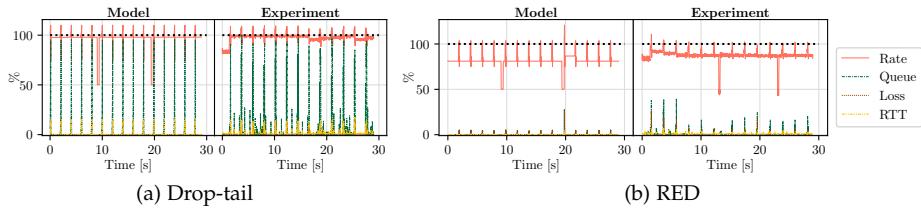


Figure 5.5: BBRv2 trace validation.

the congestion window in percent of the path BDP. The comparisons highlight both important commonalities and differences between the model and the experiments.

Similarities. The fluid models are highly predictive regarding the rate patterns over time. In addition, the fluid models correctly capture that the loss-sensitive BBRv2 lead to considerably smaller loss (barely visible) than BBRv1, which is almost insensitive to loss. Finally, the fluid model correctly predicts that the sending rate of the loss-sensitive BBRv2 barely exceeds the bottleneck rate under RED, while RED has no impact on the loss-insensitive BBRv1. The model thus also reflects the relatively low buffer usage of BBRv2 under RED, although it slightly overestimates the buffer usage of BBRv2 under RED (cf. Fig. 5.5b).

Difference: RED idealization. The difference above is due to the idealization of RED: In the model, the queue size affects the loss probability instantly; in reality, RED relies on outdated and averaged measurements of the queue size, causing some lag between queue build-ups and loss surges. In effect, a larger queue can build up until the increased drop probability stabilizes the queue, which translates into larger buffer usage in the experiments.

Difference: ProbeRTT state in BBRv2. The BBRv2 flow in the model simulation for drop-tail regularly enters the ProbeRTT state, unlike in the corresponding experiment. This observation can be explained as follows. In the model, BBRv2 regularly

manages to drain the queue, therefore discovers the propagation delay early, and cannot detect a lower RTT afterwards; hence, it enters the ProbeRTT state every 10 seconds. In the experiment, however, BBRv2 never fully uncovers the propagation delay, and experiences random fluctuations in the RTT measurements. Hence, BBRv2 occasionally observes RTTs that fall short of the current minimum-RTT estimate, which keeps it from entering the ProbeRTT state.

In summary, the fluid models capture the differences among CCAs and queuing disciplines with high accuracy, especially relatively (e.g., which CCAs lead to lower buffer usage) and to a lesser degree also absolutely (e.g., level of buffer usage).

5.4.3 Validation of Aggregate Results

The trace validations in the previous subsection indicate that the presented CCA fluid models yield reasonable predictions for single senders. The more important question, however, is whether these fluid models can acceptably predict network-performance metrics given interacting senders.

Aggregate-validation metrics. To test the fluid models in this metric-oriented aspect, we compare aggregate results from model computations and experiments for a wide variety of network parameters, in particular with respect to Jain fairness (Fig. 5.6), packet loss (Fig. 5.7), buffer occupancy (Fig. 5.8), bottleneck-link utilization (Fig. 5.9) and jitter, i.e., packet-delay variation (Fig. 5.10).

Experiment settings. All metrics were obtained from the aggregation of 5-second traces, where the experiment results are averaged over 3 runs. In contrast, fluid models are deterministic and do not require averaging. The network setting was based on the topology in Fig. 5.3, $N = 10$ senders, a bottleneck-link rate of $C_\ell = 100$ Mbps, a bottleneck-link propagation delay $d_\ell = 10$ ms and total RTTs randomly selected between 30 and 40ms. For heterogeneous CCAs, each CCA was employed by $N/2 = 5$ senders. To strengthen our validation, we conduct the same analysis for shorter delays, which confirms our results (cf. Appendix D.2).

5.4.3.1 Fairness

BBRv1 vs loss-based CCAs. Regarding fairness (cf. Fig. 5.6), we first observe that the least fairness arises when a loss-sensitive CCA (Reno, CUBIC or BBRv2) competes with BBRv1 in shallow buffers, which has already been well documented in previous research [231, 254]. This unfairness is the result of the loss insensitivity of BBRv1,

which maintains its rate despite loss while loss-sensitive CCAs practically stop sending in reaction to the loss caused by BBRv1.

Starting at buffer sizes from 4 BDP, however, the fairness in these settings increases for two reasons. First, these large buffers reduce the occurrence of loss, which prevents the back-off of loss-sensitive CCAs. Second, in large buffers, the inflight limit of the congestion window restricts the sending rate of BBRv1 and allows competing flows to obtain a higher share of bandwidth than in shallow buffers. Given a RED queue, however, the fairness of BBRv1 towards loss-sensitive CCAs is consistently low because RED (1) increases loss and (2) restricts the buffer build-up such that the inflight of the BBRv1 flows is substantially below their inflight limit.

BBRv2 vs loss-based CCAs. The fairness issues of BBRv1 have been largely resolved in BBRv2, as the fluid model and the experiment results show. However, BBRv2 is still unfair towards loss-based CCAs in RED buffers, where the higher loss sensitivity of loss-based CCAs is revealed.

Homogeneous BBRv1 fairness. One substantial difference between the fluid-model predictions and the experiment results is the decreasing fairness of BBRv1 in homogeneous settings in deep drop-tail buffers, which only appears in the fluid model. The fluid model reveals the RTT unfairness of BBRv1, which has indeed been experimentally confirmed [231, 254], although for higher RTT differences than used in our network setting. This RTT unfairness stems from the inflight limit of BBRv1, which becomes relevant in deep buffers: Since flows with a lower RTT are estimating a lower BDP and hence maintain a smaller congestion window, lower-RTT flows are more severely restricted by the inflight limit. Our fluid model can reveal this effect. In deep buffers with large queues, each BBRv1 sender i is restricted by its congestion-window size w_i , resulting in the following sending rate:

$$x_i = \frac{w_i}{\tau_i} = \frac{2\tau_i^{\min} x_i^{\text{btl}}}{\tau_i} = \frac{2d_i^{\text{p}}}{d_i^{\text{p}} + q_\ell/C_\ell} x_i^{\text{btl}}, \quad (5.35)$$

where ℓ is the bottleneck link shared among all flows, and we assume $\tau_i^{\min} = d_i^{\text{p}}$ (i.e., the propagation delay has been successfully uncovered). With these sending rates, the delivery rate for sender i is according to Eq. (5.18):

$$x_i^{\text{dlv}} = \frac{x_i C_\ell}{x_i + \sum_{j \neq i} x_j}. \quad (5.36)$$

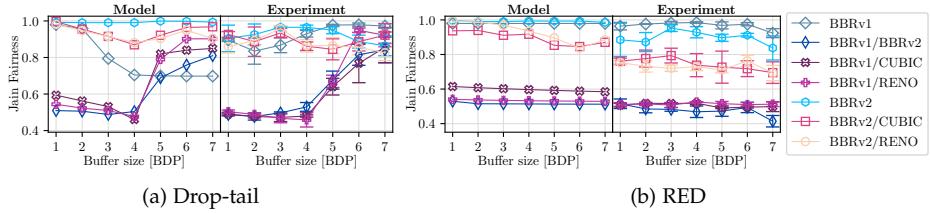


Figure 5.6: Fairness validation

Moreover, since the sending rates are static (because the varying pacing rate is overruled by the congestion-window constraint), the delivery rates are static across the bandwidth-probing interval as well, resulting in $x_i^{\max} = x_i^{\text{dly}}$. This maximum delivery rate x_i^{\max} is monotonically increasing in propagation delay d_i^p of sender i :

$$\frac{\partial x_i^{\max}}{\partial d_i^p} = \frac{2x_i^{\text{btl}} \sum_{j \neq i} x_j}{\left(2d_i^p x_i^{\text{btl}} + (d_i^p + q_\ell/C_\ell) \sum_{j \neq i} x_j\right)^2} > 0 \quad (5.37)$$

Hence, flows with a higher RTT have a larger congestion window, can thus send at a higher rate, measure a higher maximum delivery rate, and in turn estimate a higher bottleneck bandwidth. This bottleneck-bandwidth estimate then increases the congestion-window size, leading to a positive feedback loop. However, the representation of the delivery rate in Eq. (5.36) idealizes the noisy relationship between sending rates and delivery rates in real-world buffers. This noise can eliminate the difference in measured delivery rates for flows with small RTT differences, and thus break the positive feedback loop in that case. Hence, the described effect only appears for relatively large RTT differences in reality.

Fairness conclusion. In conclusion, the fluid models correctly predict fairness effects from a qualitative perspective, i.e., they rank CCA settings correctly according to their fairness, and approximately also from a quantitative perspective. Interestingly, the fluid model also predicts RTT unfairness among BBRv1 flows in deep drop-tail buffers, which does not appear in the corresponding experiments. However, this RTT unfairness is a real issue in more extreme settings than tested in this validation, i.e., for higher RTT differences between the senders. Since the role of a fluid model is to reveal problematic CCA features, the exaggeration of an existing problem barely weakens the methodological value of the BBR fluid model.

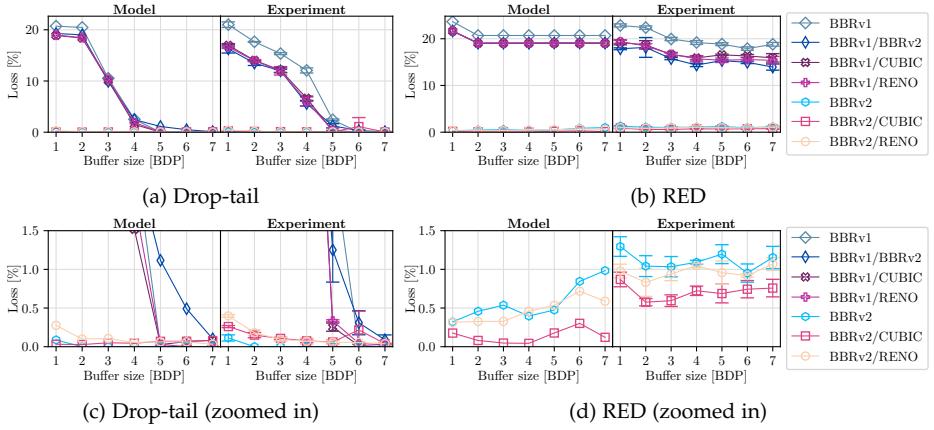


Figure 5.7: Loss validation

5.4.3.2 Loss

Predicted phenomena. Fig. 5.7 suggests that fluid models are highly suitable to predict loss rates for different CCAs, both in homogeneous settings and heterogeneous settings and both qualitatively and quantitatively. Our model correctly predicts (1) that the loss rate of loss-sensitive CCAs (Reno, CUBIC, BBRv2 and combinations thereof) in drop-tail buffers is below 1% and goes to 0% for increasing buffer size, (2) that BBRv1 leads to considerable loss of at most 20%, where the loss rate is indirectly proportional to the buffer size for drop-tail queuing, and (3) that a RED queuing discipline keeps loss rates roughly consistent across buffer sizes.

Prediction error. One obvious prediction error of the fluid model is the underestimation of loss rates for loss-sensitive CCAs given RED in Fig. 5.7d. This underestimation stems again from an idealization of the RED queue in the model, which determines the loss rate based on the current queue length. In contrast, real RED tracks the queue length with a moving average and hence reacts to queue build-up with delay. Since the queue has more time to accumulate until stabilization by RED, the queue length is higher than given an instantaneously reacting RED algorithm (cf. also Fig. 5.8b). Moreover, since the RED dropping probability is proportional to the queue size, the loss rate given delayed RED is slightly higher than for instantaneous RED. However, since this underestimation only amounts to 0.5 percentage points, we still consider our fluid model highly predictive with respect to loss.

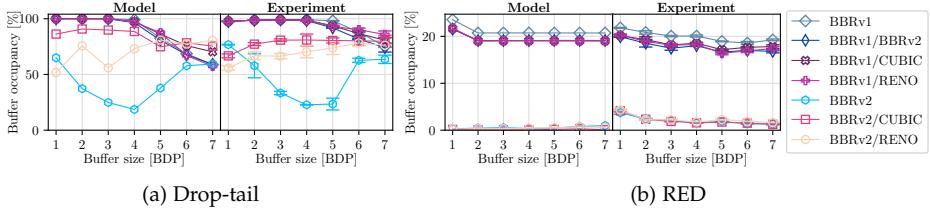


Figure 5.8: Queuing validation

5.4.3.3 Queueing

Fig. 5.8 shows the average queue size as a share of buffer capacity.

Bufferbloat by loss-based CCAs. Notably, the fluid model captures the effect that the traditional loss-based CCAs Reno and CUBIC cause *bufferbloat*, i.e., lead to consistently high buffer utilization. This effect is visible in combined settings of BBRv2 with loss-based CCAs, as the generally lower buffer usage of BBRv2 in a homogeneous setting demonstrates that BBRv2 is not responsible for the bufferbloat.

Buffering by BBRv1. Interestingly, BBRv1 leads to even more intense buffer usage than loss-sensitive CCAs, whether in homogeneous or in combined settings. Under drop-tail (Fig. 5.8a), BBRv1 uses most of the buffer independent of buffer size, where the relative buffer usage is only moderately reduced in large buffers. This effect is surprising, as a major design goal of BBR is exactly to avoid the bufferbloat caused by traditional loss-based CCAs [47].

Buffering by BBRv2. The validation analysis reveals another unexpected phenomenon, which concerns the buffer utilization of BBRv2 in homogeneous settings given a drop-tail queuing discipline. In particular, BBRv2 leads to constant absolute buffer usage for buffer sizes up to 4 BDP, which is visible as decreasing relative buffer usage in Fig. 5.8a. In these scenarios, the adjustments to BBR appear to have resolved the issue of bufferbloat in BBRv1. In large buffers, however, the buffer utilization increases again with buffer size. Through trace inspection, we found that this phenomenon is caused by initial measurements of `inflight_hi` during the start-up phase of BBRv2: Given large buffers, this initial `inflight_hi` bound may be set too high or not set at all. Moreover, `inflight_lo` may never be set either, because large buffers prevent loss, which would activate `inflight_lo`. In absence

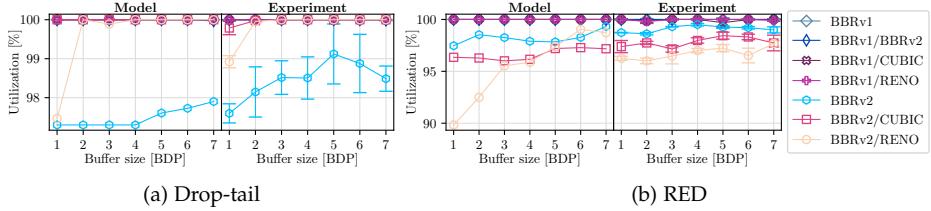


Figure 5.9: Utilization validation

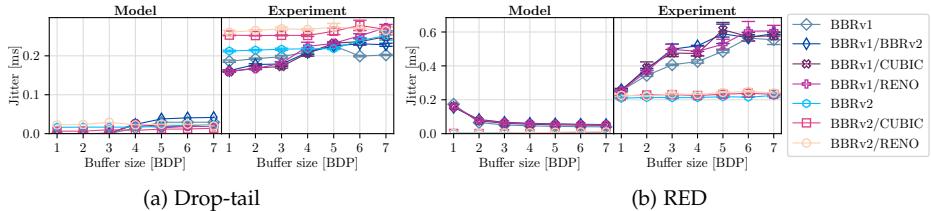


Figure 5.10: Jitter validation

of stringent bounds given by `inflight_hi` and `inflight_lo`, BBRv2 falls back on the standard BBR congestion-window size of 2 estimated BDP (cf. Eq. (5.33)). In comparison with the empirically found `inflight_hi` and `inflight_lo`, this congestion-window size is a loose bound that allows higher sending rates of BBRv2 and thus causes more intense buffering. To the best of our knowledge, this behavior of BBRv2 has not been publicly documented so far. While our BBRv2 fluid model does not model the start-up phase which causes this issue, the same effect can be observed in the model when choosing the initial condition of the differential equation for w_i^{hi} (cf. Eq. (5.31)) dependent on the buffer size. Therefore, we note that fluid models have to be evaluated under a variety of initial conditions to reveal design issues.

5.4.3.4 Utilization

The fluid model captures three important aspects of link utilization (cf. Fig. 5.9).

Full utilization by BBRv1. First, the fluid model correctly predicts that BBRv1 (or combinations with BBRv1) lead to full utilization of the bottleneck link, both under drop-tail and RED. This high utilization by BBR is unsurprising given the aggressiveness of the CCA, which also manifests in high loss (Fig. 5.7) and intense queuing (Fig. 5.8).

Utilization by loss-based CCAs. Second, the fluid model mirrors the increasing link utilization by loss-sensitive CCAs for increasing buffer size under drop-tail: Loss-based CCAs grow their rate while the buffer is filling up and cut it by a constant factor when the buffer is full, so larger buffers imply higher rates and thus higher utilization.

Utilization by BBRv2. Third, the fluid model reflects that BBRv2 yields the lowest utilization given drop-tail among all CCAs, although the wasted capacity only amounts to 3% at most. This incomplete utilization stems from the ProbeRTT state of BBRv2, in which the inflight is reduced to half the estimated BDP. Under RED, BBRv2 does not enter the ProbeRTT phase for 10 senders; neither does BBRv1 for both queueing disciplines.

Prediction error. The only major prediction error of the model is the underestimation of utilization by the BBRv2/Reno combination in shallow RED buffers. This result again points to the idealization of instantly reacting RED queues: The Reno flows in the model back off as soon as the arrival rate exceeds the bottleneck capacity, whereas this back-off is delayed in the experiments. Hence, the arrival-rate evolution moves on a lower level in the model than in the experiment, leading to an underestimation of utilization.

5.4.3.5 Jitter

Jitter corresponds to the mean delay difference between consecutive packets. The experimental jitter results in Fig. 5.10 are calculated in this manner. As the fluid model misses a notion of packets, we compute the jitter for fluid-model traces by sampling the RTT at a virtual packet rate, i.e., every $g \cdot N/C_\ell$ seconds, where g is a given packet size.

However, as Fig. 5.10 makes clear, also this makeshift calculation unsurprisingly fails to predict jitter: Fluid models intentionally abstract from small-scale fluctuations and describe only the macroscopic tendency of network indicators with smooth curves. Nonetheless, fluid models could be combined with packet-level models aimed at modeling jitter [69, 70]; we leave this challenge for future research.

5.5 THEORETICAL ANALYSIS

In this section, we analyze the BBR fluid models to characterize the stability of these CCAs.

5.5.1 BBRv1 Stability Analysis

While the fluid model in §5.3.3 is suitable for simulation, we have to simplify it to a high-level model for analysis (§5.5.1.1). In §5.5.1.2, we investigate the existence, form and stability of BBRv1 equilibria.

5.5.1.1 Model Reduction

Abstraction of ProbeRTT state. The first simplification step is given by disregarding the ProbeRTT state, which lets flow discover their propagation delay and generally achieves this goal. Hence, we assume that

$$\tau_i^{\min} = \sum_{\ell \in \pi_i} d_\ell =: d_i. \quad (5.38)$$

Since this minimum-RTT measurement is present in the congestion-window size, we simplify:

$$w_i^{\text{pbw}} = 2\tau_i^{\min} x_i^{\text{btl}} = 2d_i x_i^{\text{btl}}. \quad (5.39)$$

Apart from affecting the congestion-window size, the ProbeRTT state has no lasting effects and can hence be omitted for the purpose of stability analysis.

Simplification of delivery-rate evolution. The second step involves understanding the evolution of the maximum measurement x_i^{\max} . This maximum measurement is the maximum delivery rate x_i^{dlv} over a period, which in turn depends on the sending rates of all flows and the queue length at the bottleneck link (cf. Eq. (5.18)). The sending rate of a sender i is

$$x_i = \begin{cases} \min(w_i^{\text{pbw}} / \tau_i, 5/4x_i^{\text{btl}}) & \text{if flow } i \text{ is probing,} \\ \min(w_i^{\text{pbw}} / \tau_i, 3/4x_i^{\text{btl}}) & \text{if flow } i \text{ is draining,} \\ \min(w_i^{\text{pbw}} / \tau_i, x_i^{\text{btl}}) & \text{otherwise.} \end{cases} \quad (5.40)$$

Hence, the maximum delivery rate depends on the concurrent behavior of the other flows on the bottleneck link, which may be probing or draining during the measurements of flow i . For many flows, the probing and draining flows can be expected to offset each other such that the total background-traffic volume is similar as if

$$\forall j \in U_{\ell_i}, j \neq i. \quad x_j^{\text{pbw}} = \min(w_j^{\text{pbw}} / \tau_j, x_j^{\text{btl}}). \quad (5.41)$$

Given this background traffic, a sender i measures the maximum delivery rate in the probing phase. Hence, the maximum measurement is determined as follows:

$$x_i^{\max} = \begin{cases} \frac{\min(5/4, \Delta_i) \cdot x_i^{\text{btl}} \cdot C_{\ell_i}}{\min(5/4, \Delta_i) \cdot x_i^{\text{btl}} + \sum_{j \neq i} \min(1, \Delta_j) \cdot x_j^{\text{btl}}} & \text{if } q_{\ell_i} > 0 \\ \min(5/4, \Delta_i) \cdot x_i^{\text{btl}} & \text{otherwise} \end{cases} \quad (5.42)$$

where

$$\Delta_i = \frac{2d_i}{d_i + \sum_{\ell \in \pi_i} \frac{q_{\ell}}{C_{\ell}}}, \quad (5.43)$$

and ℓ_i is flow i 's bottleneck link.

Simplification of bottleneck-estimate update. The final simplification step concerns the adaptation of the bottleneck-bandwidth estimate x_i^{btl} . Over a long duration, this regular update can be approximated by a continuous assimilation:

$$\dot{x}_i^{\text{btl}} = x_i^{\max} - x_i^{\text{btl}} \quad (5.44)$$

5.5.1.2 BBRv1 Stability Analysis

Equilibria. To characterize stability of a CCA, we first need to identify its *equilibria*, i.e., configurations from which the fluid-model dynamics cannot depart. In the case of BBRv1, the network state may change if the maximum measurement x_i^{\max} by some flow i differs from the bottleneck-bandwidth estimate x_i^{btl} , or if the queue length q_{ℓ} of some link ℓ grows; both events may lead to subsequent rate changes. To formalize this condition, we henceforth consider N senders which share a single bottleneck link ℓ^* . Moreover, we first assume that buffer capacities do not constrain the dynamics, and modify this assumption later.

Definition 5.1 BBRv1 Equilibrium Conditions. N BBRv1 senders sharing a bottleneck link ℓ^* are in equilibrium if and only if $\{x_i^{\text{btl}}\}_{i \in U_{\ell^*}}$ and q_{ℓ^*} satisfy:

$$\sum_{i \in U_{\ell^*}} \min(1, \Delta_i) \cdot x_i^{\text{btl}} = C_{\ell^*} \quad \forall i \in U_{\ell^*}. x_i^{\text{btl}} = x_i^{\max} \quad (5.45)$$

The first condition keeps the aggregate rate y_{ℓ^*} at line rate C_{ℓ^*} and hence ensures a static queue length. The remaining constraints rule out rate adaptations. These conditions imply the following equilibria according to the proof in Appendix D.3.1:

Theorem 5.1 BBRv1 Equilibrium. N BBRv1 senders sharing a bottleneck link ℓ^* are in equilibrium if and only if propagation delay equals queuing delay for every sender, i.e.,

$$\forall i \in U_{\ell^*}. d_i = \sum_{\ell \in \pi_i} \frac{q_\ell}{C_\ell}.$$

Interestingly, Theorem 5.1 suggests that the equilibria of BBRv1 in single-bottleneck scenarios (with non-limiting buffers) can be arbitrarily unfair as long as $\sum_{i \in U_{\ell^*}} x_i^{\text{btl}} = C_{\ell^*}$. Furthermore, we note that the BBRv1 equilibrium requires equal path propagation delay d for all senders if all senders only encounter a non-empty queue at the bottleneck link ℓ^* .

Stability. For our stability analysis, we focus on that case, i.e., a scenario where the queue lengths on all involved links except the bottleneck link ℓ^* are zero, which is a scenario frequently investigated in the literature [15, 120, 273]. In this case, we can prove asymptotic stability of BBRv1 with the indirect Lyapunov method, meaning that initial configurations exist for which the BBRv1 dynamics converge to the equilibrium.

Theorem 5.2 BBRv1 Stability. *In a single-bottleneck network with a queue exclusively at the bottleneck, the BBRv1 equilibrium from Theorem 5.1 is asymptotically stable.*

PROOF. In the scenario under consideration, it holds that $q_\ell = 0 \ \forall \ell \neq \ell^*$. Given Theorem 5.1, it thus holds that the equilibrium is valid only for equal RTTs:

$$\forall i \in U_{\ell^*}. d_i = \frac{q_{\ell^*}}{C_{\ell^*}} =: d. \quad (5.46)$$

Hence, we can simplify: $\Delta_i = \Delta(q_{\ell^*}) := 2d/(d + q_{\ell^*}/C_{\ell^*})$. As a result, the equilibrium requires that $\Delta(q_{\ell^*}) = 1 \iff q_{\ell^*} = dC_{\ell^*}$.

We now consider a configuration where the senders are out of equilibrium and constrained by the congestion-window limit, i.e., $\Delta(q_{\ell^*}) < 1 \iff q_{\ell^*} > dC_{\ell^*}$. The dynamics of the bottleneck-bandwidth estimates by sender i are then given by:

$$\dot{x}_i^{\text{btl}} = \frac{\Delta(q_{\ell^*})x_i^{\text{btl}}C_{\ell^*}}{\Delta(q_{\ell^*})\sum_{k \in U_{\ell^*}} x_k^{\text{btl}}} - x_i^{\text{btl}} = \frac{x_i^{\text{btl}}C_{\ell^*}}{\sum_{k \in U_{\ell^*}} x_k^{\text{btl}}} - x_i^{\text{btl}} \quad (5.47)$$

Moreover, the dynamics of the bottleneck-link queue q_{ℓ^*} are:

$$\dot{q}_{\ell^*} = y_{\ell^*} - C_{\ell^*} = \Delta(q_{\ell^*}) \sum_{i \in U_{\ell^*}} x_i^{\text{btl}} - C_{\ell^*} \quad (5.48)$$

where y_{ℓ^*} is the arrival rate at bottleneck link ℓ^* . Based on these dynamics, we derive the dynamics of the arrival rate y_{ℓ^*} :

$$\begin{aligned} \dot{y}_{\ell^*} &= \dot{\Delta}(q_{\ell^*}) \sum_{i \in U_{\ell^*}} x_i^{\text{btl}} + \Delta(q_{\ell^*}) \sum_{i \in U_{\ell^*}} \dot{x}_i^{\text{btl}} \\ &= -\frac{1}{C_{\ell^*}(d + \frac{q_{\ell^*}}{C_{\ell^*}})} y_{\ell^*}^2 + \left(\frac{1}{d + \frac{q_{\ell^*}}{C_{\ell^*}}} - 1 \right) y_{\ell^*} + \Delta(q_{\ell^*}) C_{\ell^*} \end{aligned} \quad (5.49)$$

Building on this formalization, we can define a classic non-linear dynamic system with y_{ℓ^*} and q_{ℓ^*} as state variables, and \dot{y}_{ℓ^*} and \dot{q}_{ℓ^*} as entries of the vector-valued function f describing the dynamics. To characterize the stability of that system, we can then employ the indirect Lyapunov method [209]. This method states that a system is locally asymptotically stable if the Jacobian matrix of the system dynamics f has eigenvalues with exclusively negative real parts when evaluated at the equilibrium. The Jacobian matrix $J_f \in \mathbb{R}^{2 \times 2}$ has the following entries:

$$\frac{\partial \dot{y}_{\ell^*}}{\partial y_{\ell^*}} = -\frac{2}{C_{\ell^*}(d + \frac{q_{\ell^*}}{C_{\ell^*}})} y_{\ell^*} + \frac{1}{d + \frac{q_{\ell^*}}{C_{\ell^*}}} - 1 \quad (5.50)$$

$$\frac{\partial \dot{y}_{\ell^*}}{\partial q_{\ell^*}} = \frac{y_{\ell^*}^2}{C_{\ell^*}^2(d + \frac{q_{\ell^*}}{C_{\ell^*}})^2} - \frac{y_{\ell^*}}{C_{\ell^*}(d + \frac{q_{\ell^*}}{C_{\ell^*}})^2} - \frac{2d}{(d + \frac{q_{\ell^*}}{C_{\ell^*}})^2} \quad (5.51)$$

$$\frac{\partial \dot{q}_{\ell^*}}{\partial y_{\ell^*}} = 1 \quad \frac{\partial \dot{q}_{\ell^*}}{\partial q_{\ell^*}} = 0 \quad (5.52)$$

Evaluating this Jacobian matrix at the equilibrium, i.e., $y_{\ell^*} = C_{\ell^*}$ and $q_{\ell^*} = dC_{\ell^*}$, yields a matrix for which the maximum eigenvalue λ^+ can be found via the characteristic equation:

$$\begin{aligned} J_f(C_{\ell^*}, dC_{\ell^*}) &= \begin{pmatrix} -\frac{1}{2d} - 1 & -\frac{1}{2d} \\ 1 & 0 \end{pmatrix} \\ \implies \lambda^+ &= -(\frac{1}{4d} + \frac{1}{2}) + \frac{1}{2d}(\pm(d - \frac{1}{2})) \end{aligned} \quad (5.53)$$

Performing a case distinction on d confirms that the maximum eigenvalue λ^+ is always negative:

$$\begin{aligned} d \leq \frac{1}{2} : \quad \lambda^+ &= \frac{-(\frac{1}{2d} + 1) - \frac{1}{d}(d - \frac{1}{2})}{2} = -1 < 0 \\ d > \frac{1}{2} : \quad \lambda^+ &= \frac{-(\frac{1}{2d} + 1) + \frac{1}{d}(d - \frac{1}{2})}{2} = -\frac{1}{2d} < 0 \end{aligned} \tag{5.54}$$

Hence, we observe that the Jacobian matrix \mathbf{J}_f has consistently negative eigenvalues, which by the indirect Lyapunov method proves the asymptotic stability of the dynamics. \square

Shallow buffers. As the proof of Theorem 5.2 makes clear, the BBRv1 equilibrium from Theorem 5.1 is only viable if the bottleneck-link buffer capacity B_{ℓ^*} permits the equilibrium queue length $q_{\ell^*} = dC_{\ell^*}$. Intuitively, the equilibrium is valid for a bottleneck buffer that is large enough for the congestion-window constraint Δ_i to have an impact. To analytically investigate the shallow-buffer case where the congestion-window limit is not effective, we assume that the bottleneck queue length q_{ℓ^*} is restricted by the buffer size B_{ℓ^*} such that the congestion-window limit has no effect for any flow i , i.e., $\Delta_i \geq 5/4$ for all $i \in U_{\ell^*}$ (cf. Eq. (5.42)). With this assumption, we find a different equilibrium for BBRv1, proven in Appendix D.3.2:

Theorem 5.3 Stable BBRv1 Equilibrium in Shallow Buffers. *N BBRv1 senders sharing a bottleneck link ℓ^* that has a shallow buffer (i.e. $\Delta_i > 5/4 \forall i \in U_{\ell^*}$) are in equilibrium if and only if each flow i has the following bottleneck-bandwidth estimate x_i^{btl} :*

$$x_i^{\text{btl}} = \frac{5C_{\ell^*}}{4N + 1}.$$

This equilibrium is perfectly fair and asymptotically stable.

Theorem 5.3 thus implies that without an effective congestion-window limit, the aggregate rate y_{ℓ^*} in equilibrium consistently exceeds the link capacity C_{ℓ^*} , except for $N = 1$. As a result, multiple BBRv1 senders fill the shallow bottleneck-link buffer, eventually incurring a loss rate equal to the excess sending rate (20% for $N \rightarrow \infty$). While this consistent packet loss does not reduce the rate of loss-insensitive BBRv1 senders, the loss is fatal for loss-based CCAs on the same bottleneck link, which produces high inter-CCA unfairness. Among each other,

BBRv1 flows must converge to perfect fairness in shallow buffers, whereas such fairness is only possible, but not required in deep buffers (cf. Theorem 5.1).

5.5.2 BBRv2 Stability Analysis

This section again presents a condensed version of the BBRv2 fluid model from §5.3.4, which is then used for stability analysis.

5.5.2.1 Model Reduction

Thanks to the shared foundation of BBRv1 and BBRv2, our reduced fluid model for BBRv2 largely matches the reduced model for BBRv1 (cf. §5.5.1.1) such that we only discuss the simplification of the maximum measured delivery rate, which is different from §5.5.1.1.

Sending rate during pulse. In particular, the specific probing process of BBRv2 affects the maximum measurement x_i^{\max} . This probing process is centered around a traffic pulse, which raises the pacing rate to $5/4 \cdot x_i^{\text{btl}}$ and the inflight volume to $5/4 \cdot \bar{w}_i$, except the loss exceeds 2%. Since we limit our analysis to networks with buffers large enough to prevent loss, the traffic-pulse rate is:

$$x_i^{\text{pls}} = 5/4 \cdot \min(1, \delta_i) \cdot x_i^{\text{btl}}, \quad \text{where} \quad \delta_i = \frac{d_i}{d_i + \sum_{\ell \in \pi_i} q_\ell / C_\ell} = \frac{\Delta_i}{2}. \quad (5.55)$$

Background traffic during bandwidth probing. In addition to this pulse rate, the background traffic co-determines the maximum delivery rate x_i^{\max} . This background traffic can be assumed to consist of flows in cruising mode, in which any BBRv2 flow spends the vast majority of its lifetime. In cruising mode, a BBRv2 flow i sets its pacing rate to x_i^{btl} and keeps its inflight volume at the minimum of the estimated BDP \bar{w}_i and 85% of the upper inflight bound `inflight_hi` (w_i^{hi}). Since we can exclude packet loss, w_i^{hi} corresponds to the maximum measured inflight from the probing pulse, which is $5/4 \cdot \bar{w}_i$. Since the estimated BDP \bar{w}_i is consistently smaller than $0.85 \cdot 5/4 \cdot \bar{w}_i = 1.0625 \cdot \bar{w}_i$, the sending rate in cruising mode is:

$$x_i = \min(1, \delta_i) \cdot x_i^{\text{btl}} \quad (5.56)$$

Evolution of maximum delivery rate. Based on the sending rates of pulses and the cruising mode, the evolution of the maximum delivery rate x_i^{\max} is approximated as follows (with ℓ_i as the bottleneck link):

$$x_i^{\max} = \begin{cases} \frac{5/4 \cdot \min(1, \delta_i) \cdot x_i^{\text{btl}} \cdot C_{\ell_i}}{5/4 \cdot \min(1, \delta_i) \cdot x_i^{\text{btl}} + \sum_{j \neq i} \min(1, \delta_j) \cdot x_j^{\text{btl}}} & \text{if } q_{\ell_i} > 0 \\ 5/4 \cdot \min(1, \delta_i) \cdot x_i^{\text{btl}} & \text{otherwise} \end{cases} \quad (5.57)$$

5.5.2.2 Stability Analysis

Equilibria. For BBRv2, the equilibrium conditions match the equilibrium conditions for BBRv1 (cf. Definition 5.1), with Δ_i substituted by $\delta_i = \Delta_i/2$. However, the modified adaptation rule for x_i^{btl} induces a different equilibrium for BBRv2:

Theorem 5.4 BBRv2 Equilibrium Conditions. *N BBRv2 senders sharing a bottleneck link ℓ^* are in a perfectly fair equilibrium if propagation delay and queuing delay for each flow have the following relation:*

$$\forall i \in U_{\ell^*}. \frac{N-1}{4N+1} \cdot d_i = \sum_{\ell \in \pi_i} \frac{q_\ell}{C_\ell}. \quad (5.58)$$

Importantly, the above equilibrium is not necessarily the only equilibrium for BBRv2, which may thus induce unfair equilibria like BBRv1. Nevertheless, the above BBRv2 equilibrium conditions has an inter-dependency with the rate *distribution*, in contrast to the BBRv1 equilibrium conditions from Theorem 5.1. Similar to the BBRv1 equilibrium, however, the above equilibrium implies equal path propagation delay d for all senders if only the bottleneck link has a non-empty queue.

Stability. For our stability analysis, we thus again focus on a scenario where the queue lengths on all involved links except the bottleneck link ℓ^* are zero:

Theorem 5.5 *In a single-bottleneck network with a queue exclusively at the bottleneck, the BBRv2 equilibrium from Theorem 5.4 is asymptotically stable.*

If a network has a queue exclusively at the bottleneck link, Theorem 5.4 implies equilibrium queue length $q_{\ell^*} = \frac{N-1}{4N+1}dC_{\ell^*}$. In comparison with BBRv1, BBRv2 thus reduces buffer utilization by at least 75% (for $N \rightarrow \infty$), assuming the buffer is large enough to accommodate the BBRv1 equilibrium queue.

5.5.3 Summary of Theoretical Results

To provide a practical interpretation of our theoretical findings, we summarize the key takeaways from the preceding sections in the following.

BBRv1 in deep buffers. In deep buffers, BBRv1 converges to rate distributions with no fairness guarantees (cf. Theorems 5.1 and 5.2). Indeed, previous research has already demonstrated that unfairness can arise among BBRv1 flows with strongly different RTTs, with longer-RTT flows obtaining higher bandwidth shares [119, 254]. However, we find that RTT diversity is not necessary for unfairness: Even flows with equal RTTs can obtain different bandwidth shares as long as the propagation delay of each flow equals the cumulative queuing delay on the used path. Hence, if flows with equal RTT d share a bottleneck link ℓ^* with capacity C_{ℓ^*} and only encounter a queue at that bottleneck link, the steady-state queue length of BBRv1 is the product of path propagation delay and bottleneck capacity, i.e., $d \cdot C_{\ell^*}$.

BBRv1 in shallow buffers. BBRv1 converges to different steady states (i.e., equilibria) in shallow buffers than in deep buffers (cf. Theorem 5.3). In particular, the steady states that are attained given shallow buffers involve perfectly fair rate distributions. However, the aggregate equilibrium rate for shallow buffers necessarily exceeds the bottleneck capacity for more than one concurrent flow; hence, BBRv1 causes permanently full buffers and enduring loss (of up to 20% for a high number of flows) in shallow buffers.

BBRv2 improvements. Our analysis illustrates the improvements of BBRv2 over BBRv1 for the scenario where all flows have equal propagation delay and only encounter a queue at the bottleneck link. In that case, BBRv2 necessarily converges to a steady state (cf. Theorems 5.4 and 5.5), which is preferable to the BBRv1 steady state in two respects. First, the steady state that is attained by BBRv2 necessarily involves a perfectly equitable rate distribution (unlike the steady states of BBRv1, which can be arbitrarily unfair). Second, the steady state involves a bottleneck queue length which is shorter than the BBRv1 equilibrium queue length by at least 75% and is even 0 for a single sender. Our theoretical analysis thus confirms that BBRv2 improves upon BBRv1 in the essential aspects of fairness and buffer utilization.

5.6 INSIGHTS AND DISCUSSION

In this section, we summarize the most interesting insights from our experimental validation (§5.4), and our theoretical analysis (§5.5). These insights reflect properties of CCAs (§5.6.1) and properties of the fluid-model methodology (§5.6.2).

5.6.1 *Insights into CCA Performance*

In the following, we will distinguish previously known insights that were confirmed by our fluid model, and novel insights that our fluid model disclosed. For this distinction, we indicate the type of insight by (P) (for previous) and (N) (for novel).

One of the most consistent findings in the previous sections relates to the packet loss caused by different types of CCAs:

Insight 5.1 *Loss Rates of CCAs.* *BBRv1 causes considerable loss of up to 20% of traffic under drop-tail, while the loss-sensitive CCAs Reno, CUBIC, and BBRv2 cause loss rates of around 1% (P). The same behavior is observed for a RED queuing discipline (N).*

While such a difference between BBRv1 and loss-based CCAs is not surprising given different loss sensitivity, the large extent of the loss caused by BBRv1 is unexpected. Importantly, while the loss insensitivity of BBRv1 does not lead to throughput reductions of BBRv1, the high packet loss will still lead to unsatisfactory application performance. Moreover, in competition with loss-sensitive CCAs, the loss insensitivity of BBRv1 poses a fairness concern, which is also reflected in previous work [231, 254]:

Insight 5.2 *BBRv1's Unfairness Towards Loss-Based CCAs.* *BBRv1 is highly unfair towards loss-sensitive CCAs, leading to near starvation of loss-based flows in shallow buffers (given a drop-tail queuing discipline) (P) or buffers of any size (given a RED queuing discipline) (N). In large drop-tail buffers, the congestion window of BBRv1 becomes effective, leading to improvements in fairness towards loss-sensitive CCAs (P).*

The aggressiveness of BBRv1 has two additional effects:

Insight 5.3 Utilization and Buffer Usage of BBRv1. In all investigated settings, BBRv1 (also in combination with other CCAs) achieves full link utilization, but also significant bufferbloat independent of the queuing discipline (P).

Many of these insights gained from our fluid model have already been identified by previous, experiment-based analyses. In response to documentations of these issues, Google has begun to develop BBRv2, which can be characterized as follows:

Insight 5.4 Performance of BBRv2. BBRv2 mostly achieves the redesign goals of reduced buffer usage, avoiding excessive loss, and preserving fairness to loss-based CCAs (P).

However, we have identified two settings in which BBRv2 does not achieve its design goal:

Insight 5.5 BBRv2 in Large Drop-Tail Buffers. In drop-tail buffers with a size exceeding five BDP, BBRv2 causes higher buffer utilization than for smaller buffers, caused by distortions in an initial `inflight_hi` estimate in the start-up phase (N).

Insight 5.6 BBRv2 in RED Buffers. When competing with loss-based CCAs (Reno and CUBIC) under a RED queuing discipline, BBRv2 is unfair towards the loss-based CCAs. The reason for this unfairness is that on high-capacity links, the loss sensitivity of loss-based CCAs is markedly higher than the loss sensitivity of BBRv2 (N).

5.6.2 Insights into Fluid Models

The preceding sections not only yield valuable insights into the performance characteristics of BBR, but also illustrates the strengths and limitations of fluid models as an analysis tool. We assess the predictive power of fluid models as follows:

Insight 5.7 Qualitative Accuracy of Fluid Models. Fluid models are highly predictive from a qualitative perspective, i.e., they accurately capture the direction of correlations between CCA performance and network parameters as well as the ranking of different CCAs according to performance metrics.

Insight 5.8 Quantitative Accuracy of Fluid Models. The accuracy of the quantitative predictions by fluid models depends on the metric: While the quantitative predictions of fluid models are highly accurate regarding loss and fairly accurate regarding buffer usage, the quantitative predictions regarding fairness and utilization are only partially accurate.

Despite their overall high predictive power, fluid models yield misleading results in some cases. We identified the following sources of potentially inaccurate predictions:

Insight 5.9 Sources of Inaccuracy. Inaccurate predictions by fluid models can result from at least three sources:

- idealizations, e.g., assuming instantly reacting RED queues (cf. §5.4.3.2);
- difficulty of capturing discrete phenomena, e.g., jitter (cf. §5.4.3.5);
- and negligence of the start-up phase, e.g., BBRv2 has to be simulated with varying initial conditions to find issues arising from the start-up phase (cf. §5.4.3.3).

If the developers of CCAs are aware of the above pitfalls of fluid models, they can interpret the fluid-model results in the context of these caveats.

5.7 RELATED WORK

Congestion management. While our focus on this chapter is on congestion-control algorithms (CCAs), we note that there exist several orthogonal approaches to deal with congestion, for example, related to buffer management [25, 58, 63], scheduling [18, 122, 205, 206], and bandwidth reservation [43, 97].

Congestion control. Since the seminal work by Jacobson [130], a wide range of CCAs have been proposed and analyzed [30, 120, 124, 134, 141, 208]. While traditional CCAs are based on loss (timeout) signals, more recent protocols leverage explicit congestion notification (ECN) [17, 248, 272] or delay [15, 115, 151, 159, 183] to react in a more informed and fine-grained manner. With BBR [47], recently another flavor of CC has been introduced, which is often referred to as model-based.

BBR. BBR has been studied in a number of papers. In particular, Hock et al. [119] present a first independent study of BBRv1 and found fairness issues, and that

multiple BBR flows operate at their in-flight cap in buffer-bloated networks. This work led to several interesting follow-up works [75, 231, 246]. In particular, Scholz et al. [231] show that BBRv1 flows are robustly able to claim a disproportionate share of the bandwidth. Ware et al. [254] complement these empirical studies by presenting a first analytical model (although not based on differential equations) capturing BBR's behavior in competition with loss-based CCAs in deep buffers. This model has recently been refined by Mishra et al. [181]. Yang et al. [264] devise a simple fluid model for Adaptive-BBR, i.e., their adaptation of BBR specialized for wireless links. Neither of these model-based works possesses the generality of our fluid model, nor do they include a rigorous convergence analysis. BBRv2 has been investigated by a number of experiment-based studies [101, 143, 193, 240], finding mostly that BBRv2 resolves the most serious issues of BBRv1, but also identifying problematic facets of BBRv2 behavior, although not the ones found by the analysis in this chapter.

Fluid models. Fluid models (also known as differential-equation models) provide a particularly powerful framework for an analytical understanding of CC protocols and their equilibria [241], and have been widely used in the literature [86, 163, 166, 182, 215, 249]. These models are attractive for their flexibility (e.g., supporting different topologies and queuing disciplines), and for allowing fast initial analyses. In general, the models come in different flavors and can for example be analyzed using dynamical-systems techniques [215]. In one prominent work [182], a dynamic model of TCP behavior is proposed using a fluid-flow and stochastic differential-equation analysis. Using the Runge-Kutta algorithm, the fluid model also allows efficient time-stepped network simulations [163].

Recent approaches in congestion-control research. Recently, CCA research methodology has experienced innovation with promising proposals for an axiomatic approach [268] and a formal-verification approach [27]. These approaches are complementary to the fluid-model approach: While the axiomatic approach allows to identify fundamental design constraints and the formal-verification approach allows to identify network configurations in which CCA performance does not conform to specifications, neither of them is equally well-suited as the fluid-model approach to reveal the qualitative and quantitative effects of network settings and competing CCAs on CCA performance.

5.8 CONCLUSION

In this chapter, we take a deep dive into the recent CCA proposals of BBRv1 and BBRv2 by complementing previous analyses with an approach based on *fluid models*. Fluid models are a classic but lately seldom employed approach to evaluating CCA properties, and are unique in their ability to allow both theoretical stability analysis and efficient simulation for a wide range of network scenarios.

We devise such a fluid model for both BBR versions by using new modelling techniques such as sigmoid pulses and mode variables, and perform an experiment-based validation to show that the model is highly predictive regarding performance and fairness properties. We further leverage the model for both an extensive simulation and a theoretical stability analysis. This investigation confirms previously found issues in BBRv1, but also yields new insights, e.g., regarding the structure and asymptotic stability of BBR equilibria, as well as regarding bufferbloat and inter-CCA unfairness in BBRv2.

While our model is accurate and general, we understand our analysis as a first step in exploring the investigation opportunities that our fluid model opens up. Indeed, the next chapter leverages our BBR fluid model to analyze the competition between BBR and CUBIC using theoretical stability analysis. Furthermore, it will be interesting to evaluate the BBR fluid models in multiple-bottleneck scenarios, both through simulations and further theoretical analysis.

6

MODELLING BBR/CUBIC COMPETITION DYNAMICS

6.1 INTRODUCTION

Besides providing insights by simulation, the BBR fluid model from the preceding chapter can also enhance the analytical understanding of a frequent phenomenon in the modern Internet, namely the *competition between BBR and CUBIC flows*. This form of inter-CCA competition is relevant for estimating the performance impact of BBR deployment on legacy applications, and for estimating to which extent BBR will replace CUBIC as the de-facto standard CCA in the Internet [181].

In previous model-based research on BBR/CUBIC competition, all analysis is based on *steady-state models* [181, 254]. These models identify a steady state of the inter-CCA dynamics, and predict the long-term fairness of BBR towards CUBIC based on the sending-rate distribution in this steady state. Steady-state models thus assume that BBR and CUBIC converge to a steady rate distribution, without rigorously justifying convergence.

In fact, as we show in this chapter, the convergence assumption in BBR/CUBIC competition is violated in many common settings. Instead, BBR/CUBIC competition often suffers from persistent *oscillation* of sending rates, as we demonstrate by both test-bed experiments and dynamic fluid-model simulations [5, 249]. This oscillation involves regular change patterns of CCA-specific throughput, repeating in periods of 20–30 seconds. This oscillation regularly results in extreme rate distributions, in which flows using different CCAs obtain highly uneven bandwidth shares.

Since oscillation therefore affects BBR fairness, this chapter presents a dynamic model of BBR/CUBIC competition, which describes the inter-CCA dynamics over time and therefore analytically confirms the oscillation. More precisely, the model reflects the complex interplay between the RTT probing in BBR and the CUBIC congestion-window adaptation, which drives the oscillation. As the model allows to prove, BBR and CUBIC flows do converge towards an equilibrium as long as

minimum-RTT estimate of the BBR flows is fixed. However, since this estimate regularly changes, the long-term dynamics of BBR/CUBIC competition oscillate under common conditions, which we also prove based on the model. The model also allows to compute (i) worst-case bounds for transitory unfairness during the oscillation, and (ii) network parameters under which oscillation occurs.

6.1.1 Contributions

In summary, we present the following contributions:

Documentation of instability. We show that steady-state models predict BBR fairness towards CUBIC less accurately than dynamic models (§6.3), due to their assumption that BBR/CUBIC competition converges to an equilibrium. In reality, convergence is often prevented by persistent sending-rate oscillation, which we demonstrate and explain (§6.4).

Dynamic BBR/CUBIC competition model. We develop a dynamic model of BBR/CUBIC competition that explains the observed oscillation in a mathematically rigorous fashion (§6.5). The model separates BBR/CUBIC competition into short-term dynamics, which are continuous and asymptotically stable, and long-term dynamics, which are discrete and unstable under mild conditions. These stability properties are proven with various techniques, including Lyapunov theory, center-manifold theory, and fixed-point iteration.

Insights into temporal fairness. Besides confirming the possibility of oscillation, the model provides insights on the fairness implications of BBR/CUBIC oscillation, yielding worst-case fairness bounds that we experimentally validate (§6.6). We also gain insights on the prevalence of oscillation, as the model allows to identify the space of network parameters conducive to oscillation (§6.7).

Countermeasure evaluation. We evaluate multiple adaptations of BBR regarding their effectiveness in suppressing oscillation (§6.8). These redesigned versions include several ad-hoc modifications of BBR, which we design and implement, and the official release of BBRv2. However, none of these redesigns eliminates oscillation without reducing fairness or responsiveness of BBR.

6.2 BACKGROUND ON BBR

In this section, we provide a short overview of the BBR functionality that is relevant for this chapter. A detailed description of BBR is presented in §5.3.1.

Fundamentally, a BBR flow maintains a *bottleneck-bandwidth estimate* x^{btl} and a *minimum-RTT estimate* τ^{\min} . These two state variables determine the sending rate and are continuously adjusted by the following probing processes.

Bandwidth probing. To continuously adjust x^{btl} , a BBR flow cycles through periods of eight phases, where each phase lasts for the duration of the minimum-RTT estimate τ^{\min} . In six of these eight phases, the BBR flow sends at pacing rate x^{btl} . However, in one phase (`ProbeBW_UP`), the BBR flow raises the sending rate to $5/4 \cdot x^{\text{btl}}$ to discover whether more bandwidth is available. In the subsequent phase, the pacing rate is decreased to $3/4 \cdot x^{\text{btl}}$ to eliminate any potentially built-up queues. At the end of the eight-phase period, the bottleneck-bandwidth estimate x^{btl} is updated to the maximum measured delivery rate (ACK rate) during the period.

RTT probing. Since the minimum-RTT estimate τ^{\min} should ideally match the path propagation delay, τ^{\min} tracks the minimum measured RTT. If τ^{\min} is not adjusted for 10 seconds, BBR performs an *RTT-probing step*: Namely, the flow resets τ^{\min} and tries to drain the buffer by drastically reducing its sending rate for 200 milliseconds. This rate reduction is achieved by limiting the BBR congestion window to 4 segments. At any other time, the BBR congestion-window size amounts to twice the estimated BDP, i.e., $2\tau^{\min}x^{\text{btl}}$.

6.3 THE STEADY-STATE PERSPECTIVE

In previous research, the competition between BBR and CUBIC flows has been analyzed by means of steady-state models, namely by Ware et al. [254] and Mishra et al. [181]. These steady-state models have recently been complemented by our fluid model from Chapter 5, where we analyze BBR/CUBIC competition by means of fluid-model simulation. While such simulations yield time series that describe the competition over time in specific networks, these simulations cannot yield general analytical insights or provable guarantees.

To bridge the gap between analytic steady-state modelling and numerical fluid-model simulation, we derive a *fluid-equilibrium model* of BBR/CUBIC competition in §6.3.1. This model identifies the steady state of the BBR/CUBIC dynamics by deriving the equilibrium of the joint BBR/CUBIC fluid model, and serves as a

basis for our dynamic model in §6.5. All models are evaluated w.r.t. their predictive power in §6.3.2.

6.3.1 Fluid-Equilibrium Model

In the following, we combine our reduced BBR fluid model from §5.5 with the CUBIC fluid model by Vardoyan et al. [249], and derive the equilibrium of the new joint fluid model.

6.3.1.1 Basic Model

We consider N_B BBR flows in set F_B and N_C CUBIC flows in set F_C , all of which share a single bottleneck link ℓ .

BBR flows. As described in §5.5.1, each BBR flow i generally sends at rate $x_i^B = \beta_i x_i^{\text{btl}}$, where x_i^{btl} is the bottleneck-bandwidth estimate of flow i , and β_i is the flow's *strength*. This strength β_i enforces the congestion-window constraint on the BBR sending rate. The BBR congestion-window size is $2\tau_i^{\min} x_i^{\text{btl}}$, where τ_i^{\min} is the minimum-RTT estimate of BBR flow i . Hence, the strength β_i is

$$\beta_i = \min \left(1, 2\tau_i^{\min} / \tau_i \right), \quad (6.1)$$

where τ_i is the current delay experienced by BBR flow i .

CUBIC flows. In contrast, each CUBIC flow k sends at rate $x_k^C = w_k / \tau_k$, where w_k is the current congestion window of CUBIC flow k and τ_k is the currently experienced delay of flow k . The congestion-window size w_k is determined by the CUBIC window-growth function $W(w^{\max}, s)$:

$$w_k = W(w_k^{\max}, s_k) = w_k^{\max} + c \left(s_k - \sqrt[3]{\frac{w_k^{\max} b}{c}} \right)^3. \quad (6.2)$$

In this function, w_k^{\max} is the recorded congestion-window size at the time of the last loss, s_k is the time since the last loss, and $b = 0.3$ and $c = 0.4$ are CUBIC parameters [217].

Link-arrival rate. The total load y_ℓ on link ℓ is thus given by:

$$y_\ell = \sum_{i \in F_B} x_i^B + \sum_{k \in F_C} x_k^C = \sum_{i \in F_B} \beta_i x_i^{\text{btl}} + \sum_{k \in F_C} \frac{w_k}{\tau_k}. \quad (6.3)$$

Table 6.1: Notation used in model (in alphabetical order).

Symbol	Description
α_i	Probing strength of BBR flow i
$\hat{\alpha}$	Probing strength that minimizes BBR flow i
B_ℓ	Buffer volume at link ℓ
b, c	CUBIC parameters
β_i	Strength of BBR flow i
C_ℓ	Rate capacity of link ℓ
F_B	Set of BBR flows
F_C	Set of CUBIC flows
N_B	Number of BBR flows
N_C	Number of CUBIC flows
p_ℓ	Loss rate at link ℓ
q_ℓ	Queue length at link ℓ
s_k	Time since last loss of CUBIC flow k
τ_j	RTT experienced by flow j
τ_j^p	Propagation delay experienced by flow j
τ_i^{\min}	Minimum-RTT estimate of BBR flow i
w_k	Congestion-window size of CUBIC flow k
w_k^{\max}	Congestion-window size at time of last loss for CUBIC flow k
$w^\leftarrow(w)$	CUBIC congestion-window size 10 seconds after RTT-probing step with congestion-window size w
x_i^B	Sending rate of BBR flow i
x_i^{btl}	Bottleneck-bandwidth estimate of BBR flow i
x_k^C	Sending rate of CUBIC flow k
χ	Lower limit on bottleneck-bandwidth estimate for BBR flows
y_ℓ	Total arrival rate at link ℓ
Ω	Unstable neighborhood of CUBIC congestion-window size in long-term equilibrium
$\bar{\cdot}$	Variable in long-term equilibrium
$\tilde{\cdot}$	Variable in short-term equilibrium

Loss rate. The loss rate p_ℓ is based on the excess sending rate of y_ℓ w.r.t. the link capacity C_ℓ :

$$p_\ell = \begin{cases} \frac{y_\ell - C_\ell}{y_\ell} & \text{if } y_\ell > C_\ell \wedge q_\ell = B_\ell, \\ 0 & \text{otherwise,} \end{cases} \quad (6.4)$$

where q_ℓ is the bottleneck-queue length and B_ℓ is the bottleneck-buffer size.

6.3.1.2 CUBIC Equilibrium

The behavior of a CUBIC flow k is captured by the two variables w_k^{\max} and s_k , which evolve as follows according to Vardoyan et al. [249]:

$$\dot{w}_k^{\max} = (w_k - w_k^{\max}) \cdot x_k^C \cdot p_\ell, \quad (6.5)$$

$$\dot{s}_k = 1 - s_k \cdot x_k^C \cdot p_\ell, \quad (6.6)$$

Intuitively, Eq. (6.5) describes that the maximum recorded window w^{\max} is adjusted towards the current window for each lost segment. Similarly, Eq. (6.6) captures the fact that the window-growth duration s is reset to 0 for each lost segment, but grows linearly in absence of loss.

Given the CUBIC dynamics in Eqs. (6.5) and (6.6), the CUBIC equilibrium conditions correspond to:

$$\forall k \in F_C. \quad \dot{w}_k^{\max} = (\bar{w}_k - \bar{w}_k^{\max}) \cdot \bar{x}_k^C \cdot \bar{p}_\ell = 0 \quad (6.7)$$

$$\dot{s}_k = 1 - \bar{s}_k \cdot \bar{x}_k^C \cdot \bar{p}_\ell = 0 \quad (6.8)$$

Eq. (6.8) implies that none of \bar{s}_k , \bar{x}_k^C , and \bar{p}_ℓ can be zero in equilibrium. Using this insight on Eq. (6.7) implies:

$$\dot{w}_k^{\max} = 0 \xrightarrow[\bar{x}_k^C, \bar{p}_\ell \neq 0]{(6.7)} \bar{w}_k = \bar{w}_k^{\max} \xrightleftharpoons{(6.2)} \bar{w}_k = \bar{w}_k^{\max} = \frac{c}{b} \bar{s}_k^3. \quad (6.9)$$

Hence, in the fluid equilibrium, the rate of the CUBIC flows is kept constant over time by persistent loss. Note that fluid models average network metrics over time; in reality, the loss actually occurs for specific packets rather than continuously, and thereby causes multiplicative-decrease fluctuations in the CUBIC sending rate. Inserting $\bar{x}_k^C = \bar{w}_k^{\max}/\bar{\tau}_k$ into Eq. (6.8) yields:

$$\bar{p}_\ell = \frac{b\bar{\tau}_k}{c\bar{s}_k^4} > 0. \quad (6.10)$$

This *CUBIC-stabilizing loss* \bar{p}_ℓ must correspond to the actually occurring loss, which in turn depends on the sending rates of *both* CUBIC and BBR flows:

Lemma 6.1 CUBIC Equilibrium Conditions:

$$\forall k \in F_C. \quad \frac{b\bar{\tau}_k}{c\bar{s}_k^4} = \frac{\bar{y}_\ell - C_\ell}{\bar{y}_\ell} \quad (6.11)$$

Since $\bar{p}_\ell > 0$ follows from Eq. (6.10), Eq. (6.11) implies $\bar{y}_\ell > C_\ell$, i.e., an equilibrium can only arise on a congested link, and $\bar{q}_\ell = B_\ell$, i.e., congestion fills the whole buffer and loss occurs. For that reason, the delays in equilibrium become:

$$\forall j \in F_B \cup F_C. \quad \bar{\tau}_j = \tau_j^p + \frac{B_\ell}{C_\ell}, \quad (6.12)$$

where τ_j^p is the propagation delay experienced by flow j .

6.3.1.3 BBR Equilibrium

The bottleneck estimate x_i^{btl} of BBR flow i tracks the maximum delivery rate x_i^{dlv} produced by bandwidth probing (cf. §5.5.1.1), i.e., $x_i^{\text{btl}} = x_i^{\text{dlv}} - x_i^{\text{btl}}$ with

$$x_i^{\text{dlv}} = \begin{cases} \frac{\alpha_i x_i^{\text{btl}} C_\ell}{y_\ell + (\alpha_i - \beta_i)x_i^{\text{btl}}} & \text{if } y_\ell + (\alpha_i - \beta_i)x_i^{\text{btl}} \geq C_\ell, \\ \alpha_i x_i^{\text{btl}} & \text{otherwise,} \end{cases} \quad (6.13)$$

where α_i denotes the *probing strength* of the BBR flow: In its 8-RTT bandwidth-probing cycle, a BBR flow raises its pacing rate during a single RTT (`ProbeBW_UP` phase) to discover whether additional bandwidth is available. In this phase, the BBR flow sets the pacing rate to $5/4x_i^{\text{btl}}$, while being constrained by the congestion-window size of $2\tau_i^{\min}x_i^{\text{btl}}$:

$$\alpha_i = \min \left(\frac{5}{4}, 2\tau_i^{\min}/\tau_i \right). \quad (6.14)$$

We now extend previous work by identifying the BBR/CUBIC equilibrium (outside of the RTT-probing steps). Since $\bar{y}_\ell > C_\ell$ (implied by Lemma 6.1) and $\alpha_i \geq \beta_i$ (Eqs. (6.1) and (6.14)), Eq. (6.13) implies the BBR equilibrium condition:

$$\forall i \in F_B. \quad \frac{\bar{\alpha}_i \bar{x}_i^{\text{btl}} C_\ell}{\bar{y}_\ell + (\bar{\alpha}_i - \bar{\beta}_i)\bar{x}_i^{\text{btl}}} - \bar{x}_i^{\text{btl}} = 0. \quad (6.15)$$

These conditions always admit the equilibrium $\bar{x}_i^{\text{btl}} = 0$ for any i . However, $\bar{x}_i^{\text{btl}} = 0$ is an artificial equilibrium that cannot arise in practice: The BBR implementation never sets the bottleneck-bandwidth estimate x^{btl} completely to zero, even under 100% packet loss, as such zeroing would prohibit any recovery of the sending rate [131]. Effectively, the bottleneck-bandwidth estimate is thus lower-bounded by some small number $\chi > 0$. Hence, we arrive at the following BBR equilibrium conditions:

Lemma 6.2 BBR Equilibrium Conditions: $\forall i \in F_B$.

$$\bar{x}_i^{\text{btl}} = \max \left(\chi, C_\ell - \frac{1}{\bar{\alpha}_i} \left(\sum_{j \in F_B \setminus \{i\}} \bar{\beta}_j \bar{x}_j^{\text{btl}} + \sum_{k \in F_C} \bar{x}_k^C \right) \right) \quad (6.16)$$

To determine the strengths $\bar{\alpha}_i$ and $\bar{\beta}_i$ in equilibrium, we require the equilibrium minimum-RTT estimate $\bar{\tau}_i^{\min}$. To determine $\bar{\tau}_i^{\min}$, we first note that the CUBIC equilibrium conditions in Lemma 6.1 imply a consistently full buffer. Hence, the BBR flows never *spontaneously* observe decreasing RTT samples in equilibrium, and must perform an RTT-probing step every 10 seconds. In an RTT-probing step, a BBR flow i reduces its congestion window to 4 segments, with the goal of uncovering the path propagation delay τ_i^P . Intriguingly, BBR flows synchronize their RTT-probing steps (cf. §6.4.2) such that the propagation delay is indeed uncovered if only BBR flows compete. However, in our scenario, the bottleneck link ℓ is shared with CUBIC flows, which do not participate in this buffer-draining behavior. Hence, the equilibrium minimum-RTT estimate $\bar{\tau}_i^{\min}$ is inflated by the minimal *back-off queue length* \bar{q}_ℓ^- remaining in the RTT-probing step:

$$\bar{q}_\ell^- = \left[4N_B + (1-b) \sum_{k \in F_C} \bar{w}_k - \tau_\ell^P C_\ell \right]_0^{B_\ell} \implies \bar{\tau}_i^{\min} = \tau_i^P + \frac{\bar{q}_\ell^-}{C_\ell}. \quad (6.17)$$

Since \bar{q}_ℓ^- can neither be negative nor exceed the bottleneck-buffer capacity B_ℓ , we use the notation $[\cdot]_0^{B_\ell}$ for projection to the interval $[0, B_\ell]$. Intuitively, the remaining queue volume \bar{q}_ℓ^- contains the inflight volume of all flows when the BBR flows are in an RTT-probing step and the CUBIC flows are minimal, i.e., back off because of loss (factor $1-b$). In case of congestion, this inflight data tends to accumulate on the bottleneck link ℓ . Hence, this inflight data is discounted by the volume that fits in the pipe, i.e., the BDP $\tau_\ell^P C_\ell$ (τ_ℓ^P being the propagation delay of bottleneck link ℓ).

The equilibrium conditions in Lemma 6.1 (CUBIC flows) and in Lemma 6.2 (BBR flows) form a system of $N = N_B + N_C$ nonlinear equations with N variables. This equation system can then be solved to compute an equilibrium rate distribution. While such solutions are difficult in general, the starvation of BBR flow i is easy to derive given $\bar{\alpha}_i \leq 1$:

Lemma 6.3 Sufficient Condition for BBR Starvation:

$$\bar{\alpha}_i \leq 1 \implies \bar{x}_i^{\text{btl}} = \chi \quad (6.18)$$

PROOF. From Lemma 6.1, we know that $\bar{y}_\ell > C_\ell$, and hence:

$$\begin{aligned} \bar{y}_\ell > C_\ell &\xrightleftharpoons[\bar{\alpha}_i \leq 1]{\bar{x}_i \leq 1} \bar{y}_\ell > \bar{\alpha}_i C_\ell &\xrightleftharpoons[-1]{/\bar{y}_\ell} 0 > \frac{\bar{\alpha}_i C_\ell}{\bar{y}_\ell} - 1 &\xrightleftharpoons[\bar{\alpha}_i = \bar{\beta}_i]{\cdot \bar{x}_i^{\text{btl}}} 0 > \frac{\bar{\alpha}_i \bar{x}_i^{\text{btl}} C_\ell}{\bar{y}_\ell + (\bar{\alpha}_i - \bar{\beta}_i) \bar{x}_i^{\text{btl}}} - \bar{x}_i^{\text{btl}} = \dot{x}_i^{\text{btl}} \\ 0 > \frac{\bar{\alpha}_i \bar{x}_i^{\text{btl}} C_\ell}{\bar{y}_\ell + (\bar{\alpha}_i - \bar{\beta}_i) \bar{x}_i^{\text{btl}}} - \bar{x}_i^{\text{btl}} &= \dot{x}_i^{\text{btl}} \end{aligned} \quad (6.19)$$

Therefore, if $x_i^{\text{btl}} = \chi$, a BBR flow i would reduce x_i^{btl} , which is impossible. Hence, $\bar{x}_i^{\text{btl}} = \chi$ is the unique equilibrium. \square

6.3.2 Model Evaluation

In this section, we experimentally evaluate the two previous steady-state models [181, 254], the simulated fluid model [5], and the fluid-equilibrium model from §6.3.1.

6.3.2.1 Experiment Setting

As usual in the literature [101, 119, 143, 193, 5, 231, 240, 254], we consider a dumbbell topology with a single bottleneck link ℓ . In our experiment, this bottleneck link ℓ has capacity $C_\ell = 100\text{Mbps}$ and propagation delay $\tau_\ell^P = 10\text{ms}$, which is in line with the capabilities of the network emulator Mininet [156]. Moreover, each of the N senders is connected to the bottleneck link ℓ via an individual non-shared link, also with a propagation delay of 10ms, which raises the overall RTT propagation delay to $\tau^P = 40\text{ms}$ for each flow. The non-shared links and the bottleneck link are intermediated by a buffer with a size of 1.5 path BDP (750KB), as buffer sizes above 1 BDP are especially interesting for BBR/CUBIC fairness [231, 254]. Each

experiment runs for 120 seconds, and is repeated three times. Unless otherwise stated, these experiment settings are relevant for the remainder of the chapter.

6.3.2.2 Results

We consider all possible BBR/CUBIC combinations of a network with $N = 10$ flows, and run all models plus the Mininet experiment for each combination. Then, we compute the obtained capacity share of all BBR flows. For the steady-state models and the fluid-equilibrium model, the capacity share is derived from the equilibrium rate distribution. For the simulated fluid model and the experiments, the capacity share is calculated as the average capacity share over time. The results in Fig. 6.1 yield two important insights.

BBR unfairness. First, the BBR flows consistently obtain a disproportionately large share of the capacity, especially if they are in a minority. Hence, the fairness from the balanced scenario with 5 flows for each CCA is not generalizable.

Accuracy by dynamicity. Second and more importantly for our purpose, we note that the predictive power of the models grows with their fidelity regarding the time dimension of the competition: Previous steady-state models (referred to as SteadyState I [254] and SteadyState II [181]) have no notion of time, but SteadyState II at least captures that the CUBIC flows regularly back off upon loss and is thus more accurate. All steady-state models (including the fluid-equilibrium model) are less accurate than the predictions from fluid-model simulation (*Fluid (Sim)*). The reduced accuracy of the fluid-equilibrium model (*Fluid (Eq)*) is especially interesting, given that it is derived from the time-aware fluid model. This observation suggests that the steady-state perspective on BBR/CUBIC competition fails to capture important phenomena that unfold over time. Put differently, the steady-state models predict the outcomes of BBR/CUBIC competition from a steady state, and thus implicitly assume that the competition dynamics *converge* to this steady state. However, this assumption is often questionable, as we will demonstrate in the following sections.

6.4 THE DYNAMIC PERSPECTIVE

In §6.3, we concluded that BBR/CUBIC competition must involve important time-varying phenomena that reduce the accuracy of predictions from steady-state models. In fact, the detailed results from the fluid-model simulations and the

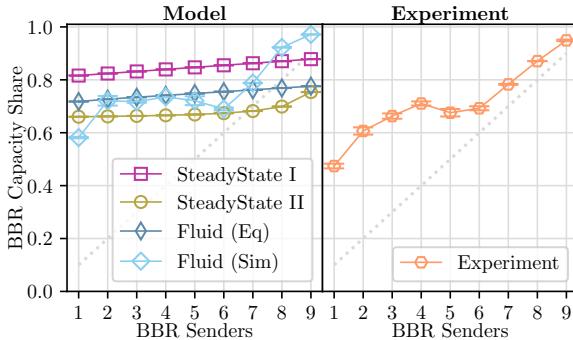
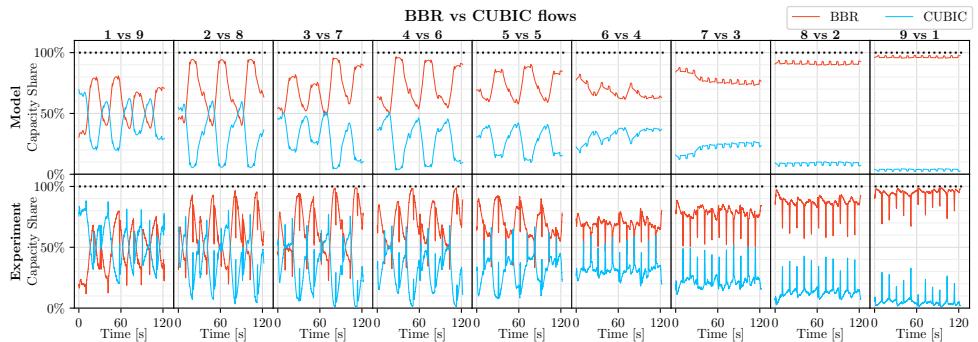


Figure 6.1: Model evaluation for 10 flows.

Figure 6.2: BBR/CUBIC competition over time for different CCA combinations
(Aggregate capacity share by CCA).

experiments (shown in Fig. 6.2) confirm that BBR/CUBIC competition often suffers from persistent *sending-rate oscillation*.

This oscillation of sending rates results from the dynamic interplay between the minimum-RTT estimate of BBR flows (τ^{\min}) and the congestion-window size of CUBIC flows (w). This interplay is explained in its basic form (i.e., for one BBR flow and one CUBIC flow) in §6.4.1, and extended to an arbitrary number of flows in §6.4.2.

6.4.1 Basic Oscillation Mechanism

Fundamentally, oscillation in BBR/CUBIC competition is created by the alternation between (i) the *RTT probing step* of BBR, and (ii) the *short-term dynamics*, which are

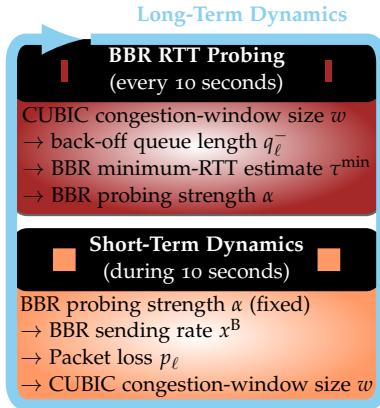


Figure 6.3: Basic two-level mechanism behind oscillation in BBR/CUBIC competition.

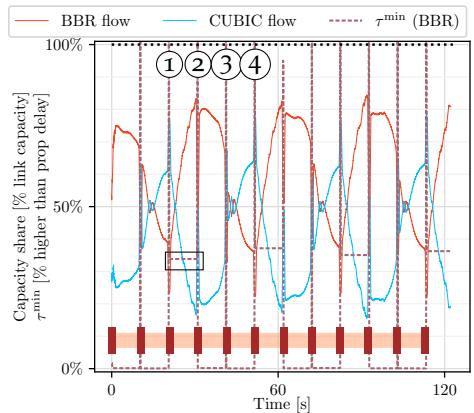


Figure 6.4: Oscillation of sending rates for competition between one BBR flow and one CUBIC flow.

determined by the preceding RTT-probing step and last until the next RTT-probing step. This alternation forms the *long-term dynamics*, which is explained in Fig. 6.3, and demonstrated with an experiment in Fig. 6.4.

Inflation of τ^{\min} . In the RTT-probing step, the BBR flow reduces its sending rate to almost zero, with the goal of emptying the buffer and discovering the propagation delay. However, the CUBIC flow does not participate in this reduction, and may thus prevent the complete clearance of the buffer if its contemporary congestion-window size w is relatively large. The CUBIC congestion-window size w thus affects the back-off queue length q_{ℓ}^- during the RTT-probing step, and hence also the BBR minimum-RTT estimate τ^{\min} that is computed from probing measurements (cf. Fig. 6.3). For example, the CUBIC congestion-window size w in RTT-Probing Step ① in Fig. 6.4 is relatively large (cf. high CUBIC rate); as a result, the buffer is not completely cleared during RTT probing and the minimum-RTT estimate τ^{\min} after the probing step (in interval [①, ②] in Fig. 6.4) is around 35% higher than the actual propagation delay (highlighted with rectangle).

Effect on sending rates. This inflated minimum-RTT estimate τ^{\min} is relevant because it co-determines the congestion-window size of the BBR flow, and thus the probing strength α of the BBR flow (Eq. (6.14)). In particular, an excessive

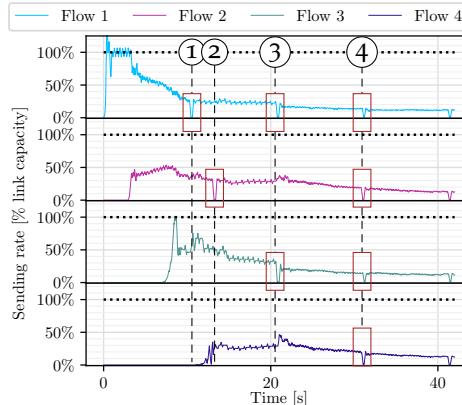


Figure 6.5: Self-synchronization of RTT probing (highlighted with red rectangles) among 4 BBR flows.

minimum-RTT estimate τ^{\min} (such as in time interval $\textcircled{1}, \textcircled{2}$ in Fig. 6.4) results in a high bandwidth-probing strength of the BBR flow, which increases the BBR sending rate x^B and reduces the CUBIC congestion-window size w by causing packet loss.

Temporarily fixed probing strength. Crucially, this high probing strength is fixed after RTT probing. Since the BBR flow and the CUBIC flow make intensive use of the buffer, they prevent a downward revision of the minimum-RTT estimate τ^{\min} during the short-term dynamics. Hence, these short-term dynamics continue until the minimum-RTT estimate τ^{\min} times out, i.e., for 10 seconds.

Circular stepwise process. At the next RTT probing step, the modified CUBIC congestion-window size w determines a new probing strength α : For example, the reduced window size w at Step $\textcircled{2}$ in Fig. 6.4 reduces the minimum-RTT estimate τ^{\min} to the propagation delay and thus induces a relatively low probing strength α . Hence, the congestion-window size w increases again until Step $\textcircled{3}$. At this Probing Step $\textcircled{3}$, w is still low enough to cause a correct estimate of the propagation delay and thus another growth phase of w . However, at Probing Step $\textcircled{4}$, τ^{\min} is inflated again and the oscillation pattern is restarted.

6.4.2 Generalization to Multiple Flows

The oscillation mechanism explained above applies to the case with a single flow per CCA. However, as Fig. 6.2 illustrates, we observe oscillation also for multiple flows per CCA. This observation raises the question how and why exactly the mechanism in Fig. 6.3 generalizes to multiple flows.

CUBIC flows. The generalization to multiple CUBIC flows is straightforward: The explanation in §6.4.1 is equally valid for multiple CUBIC flows, with the slight adaptation that the *aggregate* congestion-window size of all CUBIC flows affects the minimum-RTT estimate τ^{\min} of a BBR flow.

BBR flows. Interestingly, multiple BBR flows also behave analogously to a single BBR flow for the purpose of oscillation because BBR flows synchronize their RTT-probing steps. This synchronization has been documented by previous work [49, 231], and is experimentally confirmed in Fig. 6.5. In that experiment, 4 BBR flows are initiated sequentially with 4-second delays, and compete with 6 CUBIC flows (not pictured). The BBR flow 1 performs the first RTT probing at Time ①, which shrinks the queue and reduces the RTT. In fact, the RTT is reduced enough such that Flow 3 measures a new minimum RTT. Hence, Flow 1 and Flow 3 start the reset timers of their minimum-RTT estimates at the same time, and therefore simultaneously probe the RTT at Time ③ = ① + 10 seconds. At this Time ③, also the remaining two flows 2 and 4 become synchronized with flows 1 and 3 such that all flows perform a simultaneous RTT probing at Time ④.

Absence of oscillation. Finally, note that oscillation does not necessarily arise in a multiple-flow scenario if the CCA distribution is very unequal, e.g., for 9 BBR flows and 1 CUBIC flow in Fig. 6.2. In this case, the single CUBIC flow never manages to inflate the minimum-RTT estimate of the BBR flows, and hence the rate distribution is largely static.

6.5 MODELLING DYNAMIC COMPETITION

While §6.4 provides an intuitive explanation why oscillation arises in our experiment setting, we are now interested in a rigorous mathematical characterization of this oscillation. With this characterization, we can identify the conditions under which BBR/CUBIC oscillation occurs. In the next two sections (§6.7 and §6.6), we will then apply the model to gain practical insights.

6.5.1 Model Overview

We distinguish short-term and long-term dynamics (Fig. 6.3).

Short-term dynamics. The short-term dynamics describe the continuous BBR/CUBIC competition *between* the RTT probing steps of BBR. Between these probing steps, the BBR probing strength α is fixed, and fully determines the equilibria (§6.5.2) and the convergence behavior (§6.5.3) of the short-term dynamics.

Long-term dynamics. Building on these short-term dynamics, we derive the long-term dynamics, describing the BBR/CUBIC competition *across* RTT-probing steps (§6.5.4). In each RTT-probing step, the CUBIC congestion-window size w determines the probing strength α and thus the short-term dynamics for the next 10 seconds; these short-term dynamics then determine the congestion-window size w at the next RTT-probing step. Hence, the long-term dynamics form a discrete process.

In the following, we consider the case for one BBR flow and one CUBIC flow ($N_B = 1$, $N_C = 1$). For simplicity, we thus eliminate the flow-specific subscripts i and k where the association is obvious, e.g., only the BBR flow i has a probing strength $\alpha_i = \alpha$. As discussed in §6.4.2, the scenarios for multiple flows per CCA are qualitatively similar for a majority of CCA compositions.

6.5.2 Short-Term Equilibria

In §6.4, we have observed that the minimum-RTT estimate τ^{\min} of the BBR flow i is only periodically updated in competition with CUBIC. To investigate the dynamics between these updates, we can thus treat τ^{\min} as fixed, and by extension also the probing strength α (cf. Eq. (6.14)). This fixed α then determines the equilibrium $\tilde{\sigma}(\alpha)$ of the short-term dynamics:

Theorem 6.1 Unique Short-Term Equilibrium. Given probing strength α , the short-term equilibrium $\tilde{s}(\alpha) = (\tilde{x}^{\text{btl}}(\alpha), \tilde{w}^{\max}(\alpha), \tilde{s}(\alpha))$ for competition between a BBR flow i and a CUBIC flow k on a bottleneck link ℓ has the unique values:

$$\begin{aligned}\alpha &\geq \hat{\alpha} : \tilde{s}(\alpha) \text{ s.t. } \tilde{S}_1(\tilde{s}) = 0, \quad \tilde{x}^{\text{btl}}(\alpha) = C_\ell - \frac{\tilde{w}^{\max}(\alpha)}{\alpha \bar{\tau}_k} \\ \alpha &< \hat{\alpha} : \tilde{s}(\alpha) \text{ s.t. } \tilde{S}_2(\tilde{s}) = 0, \quad \tilde{x}^{\text{btl}}(\alpha) = \chi\end{aligned}\tag{6.20}$$

$$\tilde{w}^{\max} = \frac{c}{b} \tilde{s}(\alpha)^3, \quad \frac{\hat{\alpha}^4(\hat{\alpha} - 1)^3}{(\chi + \hat{\alpha}(C_\ell - \chi))^3} = \frac{c}{b \bar{\tau}_k (C_\ell - \chi)^7},\tag{6.21}$$

$$\tilde{S}_1(s) = \frac{(\alpha - 1)c^2}{\alpha b \bar{\tau}_k} s^7 - \frac{(\alpha - 1)c}{\alpha} s^3 - b C_\ell \bar{\tau}_k, \quad \text{and}\tag{6.22}$$

$$\tilde{S}_2(s) = \frac{c^2}{b \bar{\tau}_k} s^7 - c(C_\ell - \alpha \chi) s^4 - c s^3 - \alpha b \bar{\tau}_k \chi.\tag{6.23}$$

PROOF SKETCH. The full proof of Theorem 6.1 is provided in Appendix E.1. The proof first considers the case $\alpha \leq 1$, which implies $\tilde{x}^{\text{btl}}(\alpha) = \chi$ according to Lemma 6.3. Inserting this value into the CUBIC equilibrium conditions from Lemma 6.1 yields the condition $\tilde{S}_2(s) = 0$. The corresponding septic polynomial can be shown to have a unique root, which guarantees a unique equilibrium $\tilde{s}(\alpha)$. Second, the proof considers the case $\alpha > 1$, and distinguishes the sub-cases $\hat{x}^{\text{btl}} \geq \chi$ and $\hat{x}^{\text{btl}} < \chi$, where $\hat{x}^{\text{btl}} = C_\ell - \tilde{x}^{\text{C}}/\alpha$ is the equilibrium BBR bottleneck-bandwidth estimate without restriction to the domain $[\chi, \infty)$. If $\hat{x}^{\text{btl}} \geq \chi$, combining $\hat{x}^{\text{btl}} = \tilde{x}^{\text{btl}}$ with the CUBIC equilibrium conditions from Lemma 6.2 yields $\tilde{S}_1 = 0$, which again can be confirmed to have a unique solution. For the other sub-case where $\hat{x}^{\text{btl}} < \chi$, we identify the probing strength $\hat{\alpha} > 1$ such that χ is an unrestricted equilibrium, i.e., $\hat{x}^{\text{btl}} = \tilde{x}^{\text{btl}} = \chi$. For any probing strength $\alpha < \hat{\alpha}$, it holds that $\hat{x}^{\text{btl}} < \chi$ and thus $\tilde{x}^{\text{btl}}(\alpha) = \chi$. \square

6.5.3 Stability of Short-Term Equilibria

For the short-term equilibria to be relevant, the short-term dynamics have to converge to these equilibria. In particular, this attractiveness is also necessary for the BBR/CUBIC dynamics to converge to a new short-term equilibrium rate distribution if the minimum-RTT estimate τ^{\min} is updated and α thus changes. To provide evidence for the attractiveness of the short-term equilibria, we prove the asymptotic stability of these equilibria, meaning that the competition converges to the short-term equilibrium if the initial rate distribution is close enough:

Theorem 6.2 Stability of Short-Term Equilibrium. *In the competition between one BBR flow and one CUBIC flow, the short-term equilibrium $\tilde{\sigma}(\alpha) = (\tilde{x}^{\text{btl}}(\alpha), \tilde{w}^{\max}(\alpha), \tilde{s}(\alpha))$ of the joint dynamics is asymptotically stable.*

PROOF SKETCH. We provide the full rigorous proof in Appendix E.2, and provide a high-level overview in the following.

Importantly, a straightforward stability proof via linearization of the dynamic system fails, for the reasons already noted by Vardoyan et al. [249] who considered a CUBIC flow in isolation. Namely, when the BBR/CUBIC dynamics approach the equilibrium $\tilde{\sigma}(\alpha)$, the evolution of the CUBIC maximum window w^{\max} is exclusively determined by high-order terms that are not captured by the linearization.

However, since the short-term equilibrium is not explicitly unstable with respect to the linearized dynamics, we can rely on the *center-manifold emergence theorem* [128]: This theorem states that the dynamics converge exponentially fast to the so-called center manifold and then approximately follow a slow (i.e., sub-exponential) trajectory along the center manifold. The center manifold itself contains the equilibrium point, and thus potentially allows convergence to the equilibrium. Hence, the dynamics along the center manifold determine the overall short-term dynamics, and thus the stability properties of the short-term equilibrium. These center-manifold dynamics have lower dimension and thus allow a more tractable stability analysis.

Figure 6.6 illustrates this convergence behavior for the dynamics between one BBR flow and one CUBIC flow. As these dynamics involve the three variables x^{btl} (for the BBR flow), w^{\max} , and s (both for the CUBIC flow), the state space of the competition dynamics is three-dimensional. The system evolution from any starting point σ_n is given by a path through the state space, which is determined by the differential equations governing the three variables, and is visualized by a curve $\Gamma_e(\sigma_n)$ in the three-dimensional space. This evolution curve $\Gamma_e(\sigma_n)$ approaches the center manifold exponentially quickly. In our proof, this center manifold can be expressed by means of the single variable w^{\max} , and thus corresponds to another curve Γ_c .

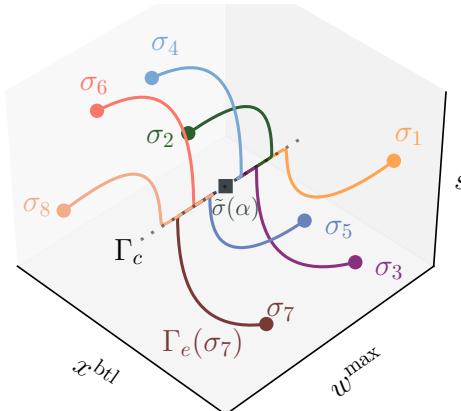


Figure 6.6: Two-step convergence in short-term dynamics: First exponential convergence towards the center manifold Γ_c (a function of w^{\max}), then sub-exponential along Γ_c towards the short-term equilibrium $\tilde{\sigma}(\alpha)$.

Moreover, the system-evolution path $\Gamma_e(\sigma_n)$ tracks the center-manifold curve Γ_c in the direction towards the equilibrium $\tilde{\sigma}(\alpha)$, indicating that the short-term equilibrium is asymptotically stable. \square

6.5.4 Long-Term Dynamics

Given the insights above, we know that the BBR/CUBIC dynamics converge to a short-term equilibrium $\tilde{\sigma}(\alpha)$ for a fixed minimum-RTT estimate τ^{\min} and fixed probing strength α . While τ^{\min} is indeed fixed for a certain amount of time, it is also regularly adjusted, at least every 10 seconds. This readjustment of τ^{\min} and α is based on the CUBIC congestion-window size w at the time of RTT probing. At that time, the CUBIC congestion window w is evolving towards the window size $\tilde{w}(\alpha') = \tilde{w}^{\max}(\alpha')$ (cf. Eq. (6.9)) of the short-term equilibrium $\tilde{\sigma}(\alpha')$ based on the previous $\alpha = \alpha'$. Importantly, however, this convergence may not be completed at the time of RTT probing. Hence, an intermediate window size $w' \neq \tilde{w}(\alpha')$ determines the readjustment of α .

As a result, the long-term evolution of the CUBIC window size w is described by a discrete-time process:

$$\forall t \in \mathbb{N}, t \geq 0. \quad w(t+1) = w^\leftarrow(w(t)), \quad (6.24)$$

where the discrete-time moments t are the RTT-probing steps of BBR, i.e., i.e., 10 seconds apart. Intuitively, $w^\leftarrow(w(t))$ is the CUBIC congestion-window size that results from the short-term dynamics (cf. §6.5.3) between the probing steps t and $t + 1$. These short-term dynamics are determined by the BBR probing strength α in that interval, which in turn results from the CUBIC window size $w(t)$ at the start time of the interval, i.e., when α was updated.

In fact, the dynamic process in Eq. (6.24) is guaranteed to have a unique *long-term equilibrium window size* \bar{w} , corresponding to the equilibrium in the fluid-equilibrium model (§6.3.1). However, this equilibrium is *unstable*, i.e., not attractive to the dynamics, under the following mild conditions:

Theorem 6.3 Instability of Long-Term Equilibrium. *The dynamics between a BBR flow and a CUBIC flow have a unique long-term equilibrium at CUBIC window size \bar{w} , which is unstable if the equilibrium has a neighborhood Ω in which the window-update function $w^\leftarrow(w)$ decreases fast enough in w :*

$$\exists \Omega = [\omega_0, \omega_1], \omega_0 < \bar{w} < \omega_1. \quad \forall \omega \in \Omega. \quad \frac{\partial w^\leftarrow(\omega)}{\partial w} < -1. \quad (6.25)$$

PROOF SKETCH. The proof in Appendix E.3 can be summarized by means of the geometric interpretation visualized in Fig. 6.7. The graph in Fig. 6.7 contains the *actual* window-update function $w^\leftarrow(w)$, which yields the CUBIC window size after an interval that started with CUBIC window size w . Moreover, the graph contains the *equilibrium* window-update function $\tilde{w}^\leftarrow(w)$, which yields the short-term equilibrium window size \tilde{w} for an interval that started with CUBIC window size w . If the short-term dynamics always completely converged to their equilibrium during the interval, then it would hold that $w^\leftarrow = \tilde{w}^\leftarrow$. The function \tilde{w}^\leftarrow is fully known, has a finite value range $[w_{\min}, w_{\max}]$, and is computed from network parameters as documented in Appendix E.3.

In contrast, the window-update function $w^\leftarrow(w)$ is mostly unknown, except for the property that $w^\leftarrow(w)$ is always between the initial window size w and the short-term equilibrium window size $\tilde{w}^\leftarrow(w)$. This last insight follows from the asymptotic stability of the short-term equilibrium (proven in Theorem 6.2), meaning that the CUBIC window always evolves towards the short-term equilibrium window size $\tilde{w}(\alpha)$ while α is fixed; however, this convergence is not necessarily complete when α is updated. In Fig. 6.7, w^\leftarrow is chosen arbitrarily for visualization purposes.

Given these functions, the updates of the discrete-time process in Eq. (6.24) can be visualized with the fixed-point diagram in Fig. 6.7. Let each window size w

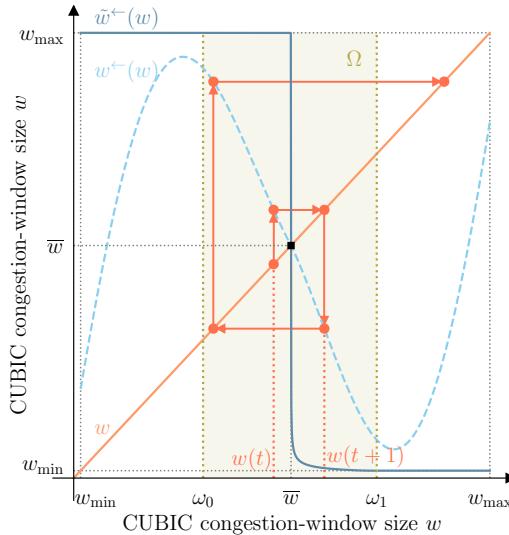


Figure 6.7: Instability of long-term dynamics: Under the condition from Theorem 6.3, the dynamics suffer from persistent oscillation as the evolution trajectory moves away from the CUBIC long-term equilibrium window \bar{w} .

correspond to point (w, w) on the identity line w . Starting with a window size $w(t)$, the evolution of w during an interval corresponds to a *vertical* projection of the point $(w(t), w(t))$ onto w^\leftarrow . The new window size $w(t+1)$ is reached by the CUBIC flow after the interval $[t, t+1]$ starting at window size $w(t)$. Graphically, this new window size $w(t+1)$ then corresponds to point $(w(t+1), w(t+1))$, which is achieved by a subsequent *horizontal* projection of $(w(t), w(t+1))$ onto the identity line. A series of such alternating projections thus visualizes the evolution trajectory of the discrete-time system.

Given this analogy, it becomes clear that the intersection point \bar{w} of the window-update function w^\leftarrow and the identity line w constitutes an equilibrium. This equilibrium is unique because $w^\leftarrow(w)$ can be shown to be monotonic, and hence the functions only have one intersection.

The stability properties of the long-term equilibrium \bar{w} depend on the update function w^\leftarrow around the equilibrium. Intuitively, the update function w^\leftarrow must be decreasing steeply enough around the equilibrium such that the evolution trajectory spirals away from the long-term equilibrium \bar{w} . Hence, under the condition on w^\leftarrow from Theorem 6.3, we can find a neighborhood Ω around the equilibrium \bar{w} such

that all trajectories entering Ω are guaranteed to leave Ω again eventually. Hence, such a neighborhood prevents asymptotic convergence to the equilibrium, and thus causes the equilibrium to be unstable. \square

6.6 FAIRNESS UNDER OSCILLATION

We now consider the *fairness implications* of BBR/CUBIC oscillation. In particular, the long-term BBR/CUBIC dynamics from Eq. (6.24) involve CUBIC congestion-window sizes $\{w(t)\}_{t \in \mathbb{N}, t \geq 0}$. Given this association between any oscillation state $w(t)$ and sending rates, fairness under oscillation depends on the fairness of the *oscillation pattern*, i.e., the fairness of all rate distributions associated with $\{w(t)\}_{t \in \mathbb{N}, t \geq 0}$. This oscillation pattern has a worst-case form with respect to fairness, which is found in §6.6.1. From this worst-case oscillation pattern, fairness bounds are computed and experimentally validated in §6.6.2.

6.6.1 Bounding the Oscillation Pattern

In particular, the fairness of the oscillation pattern can be lower-bounded by maximizing the *amplitude* of the oscillation pattern, i.e., the variance of $x^C(t)$ and $x^B(t)$ over time:

Theorem 6.4 Worst-Case Oscillation Pattern. *Given oscillation with maximum amplitude, the CUBIC flow oscillates between the congestion-window sizes $\hat{w}_0 = \tilde{w}^\leftarrow(w_{\max})$ and $\hat{w}_1 = \tilde{w}^\leftarrow(w_{\min})$ when competing with a BBR flow ($\hat{w}_0 < \hat{w}_1$).*

PROOF SKETCH. The proof in Appendix E.4 rests on two observations regarding the oscillation amplitude. First, this oscillation amplitude corresponds to the window-size changes in the update intervals, i.e., the difference between the window-size w at the start of the interval and the convergence result $w^\leftarrow(w)$ at the end of the interval (cf. Fig. 6.7). This difference $|w - w^\leftarrow(w)|$ is upper-bounded by $|w - \tilde{w}^\leftarrow(w)|$, i.e., the difference for complete convergence to the short-term equilibrium. Hence, the amplitude of the oscillation is maximized by assuming $w^\leftarrow = \tilde{w}^\leftarrow$. Second, the oscillation amplitude correlates with the size of the unstable neighborhood Ω around the long-term equilibrium \bar{w} , as the oscillation pattern typically involves a window-size change across Ω , e.g., from $w(t) < \omega_0$ to $w(t+1) > \omega_1$. In the feasible maximum, Ω covers the entire decreasing part of \tilde{w}^\leftarrow .

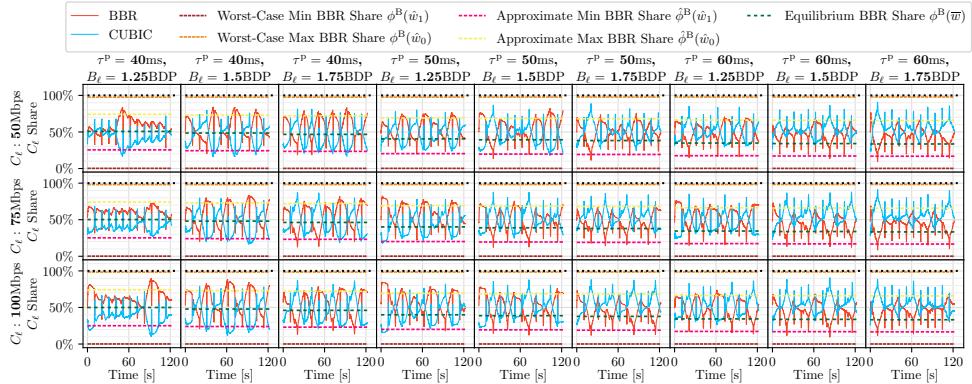


Figure 6.8: Experiments involving one flow per CCA, used for the experimental validation of fairness bounds (cf. §6.6) and for the empirical estimation of the window-update function w^\leftarrow (cf. §6.7).

Under these two assumptions, the oscillation corresponds to a *limit cycle*, cyclically revisiting the window sizes $\hat{w}_0 = \tilde{w}^\leftarrow(w_{\max})$ and $\hat{w}_1 = \tilde{w}^\leftarrow(w_{\min})$. This limit cycle is *sound* for the process in Eq. (6.24), i.e., $\tilde{w}^\leftarrow(\hat{w}_0) = \hat{w}_1$ and $\tilde{w}^\leftarrow(\hat{w}_1) = \hat{w}_0$. Moreover, the limit cycle is *attractive*, i.e., the process in Eq. (6.24) eventually enters the limit cycle. \square

6.6.2 Computing Fairness Bounds

Given the worst-case oscillation pattern from §6.6.1, we now compute fairness bounds for a selection of network configurations, and validate these bounds by means of experiments.

In particular, we test a range of 27 configurations, which are variations of the experiment settings in §6.3.2.1 regarding bottleneck-link capacity C_ℓ , path-propagation delay τ^P and buffer capacity B_ℓ . These experiments result in 120-second traces that involve two flows and are depicted in Fig. 6.8.

Bound computation. For each of these configurations, we find the worst-case oscillation pattern according to Theorem 6.4. In this oscillation pattern, the CUBIC congestion-window size alternates between two values \hat{w}_0 and \hat{w}_1 at every RTT-probing step. These congestion-window sizes constitute the *extreme points* of the

oscillation: \hat{w}_0 is the smallest congestion-window size attained by the CUBIC flow during the oscillation, whereas \hat{w}_1 is the largest window size. Hence, $x^C(\hat{w}_0) = \hat{w}_0/\bar{\tau}_k$ is the minimal attained CUBIC rate, whereas $x^C(\hat{w}_1)$ is maximal. For BBR, the maximum sending rate is (symmetric for $x^B(\hat{w}_1)$):

$$x^B(\hat{w}_0) = \beta(\hat{w}_1) \cdot \max \left(\chi, C_\ell - \frac{x^C(\hat{w}_0)}{\alpha(\hat{w}_1)} \right), \quad (6.26)$$

which is based on the BBR equilibrium conditions from Lemma 6.2, because \hat{w}_0 and \hat{w}_1 represent short-term equilibria. Note that the minimum-RTT estimate (and thus also the strengths α and β) of a BBR flow is based on the CUBIC congestion-window size at the previous RTT-probing step, which is \hat{w}_1 when the current congestion-window size is \hat{w}_0 .

Based on these sending rates, we also identify the maximum BBR capacity share:

$$\phi^B(\hat{w}_0) = \frac{x^B(\hat{w}_0)}{x^C(\hat{w}_0) + x^B(\hat{w}_0)}. \quad (6.27)$$

The minimum BBR share $\phi^B(\hat{w}_1)$ is found analogously.

Bound validation. To validate these bounds, we compare the theoretically derived bounds to the experimentally found oscillation patterns in Fig. 6.8. For bound correctness, the oscillation pattern of the BBR flow is supposed to be contained within these bounds, which is clearly true. However, the worst-case bounds are relatively loose, as they are derived based on worst-case assumptions. These assumptions do not hold in the experiments in Fig. 6.8, but may hold in other configurations.

Approximate empirical bound. In addition, we empirically find the following bound, which approximates the maximum BBR share more accurately:

$$\hat{\phi}^B(\hat{w}_0) = \frac{1}{2} \left(\phi^B(\bar{w}) + \phi^B(\hat{w}_0) \right), \quad (6.28)$$

where \bar{w} is the CUBIC congestion-window size in the long-term equilibrium (which is not converged upon). Intuitively, Eq. (6.28) states that the maximum BBR capacity share is roughly half-way between the BBR share $\phi^B(\bar{w})$ that would be achieved in the long-term equilibrium, and the worst-case maximal BBR share $\phi^B(\hat{w}_0)$.

6.7 CONDITIONS FOR OSCILLATION

After leveraging the oscillation model to predict fairness under oscillation, we now apply it again to predict the *prevalence* of oscillation, i.e., to predict which concrete

network parameters lead to oscillation. An abstract condition for oscillation is already contained in Theorem 6.3; however, to make this condition practically applicable, we need to instantiate the window-update function w^\leftarrow used in Eq. (6.24). This function is instantiated in §6.7.1 based on experimental results. Based on this instantiation, we can predict oscillation for a wide range of network parameters in §6.7.2.

6.7.1 Approximating the Update Function

The window-update function w^\leftarrow cannot be derived from the fluid-model alone, as w^\leftarrow depends on the explicit *rate of convergence* of the short-term dynamics in §6.5.3. However, this rate of convergence is only known in terms of order (i.e., exponential vs sub-exponential) rather than in explicit form. Hence, we estimate the rate of convergence and thus characterize w^\leftarrow based on the empirical observations from the experiments in Fig. 6.8.

Samples of w^\leftarrow . For each configuration γ listed in Fig. 6.8, we perform 3 experiments $E(\gamma)$. In each experiment e , we observe that the BBR/CUBIC dynamics contain a series of 12 window sizes $\{w_e(t)\}_{t \in \{1, \dots, 12\}}$, which are attained every 10 seconds, i.e., when the BBR flow performs RTT probing. Given the observations from experiment e , we know:

$$\forall t \in \{1, \dots, 11\}. \quad w_e(t+1) = w^\leftarrow(w_e(t)), \quad (6.29)$$

as w^\leftarrow describes the result of the CUBIC congestion-window evolution between probing steps. Hence, the window sizes $\{w_e(t)\}_t$ represent samples of w^\leftarrow . To fit w^\leftarrow to these samples, we use the estimation process in Appendix E.5. In this estimation, we rely on the insights that (i) the CUBIC congestion-window size always evolves towards its short-term equilibrium, and (ii) the CUBIC congestion-window size decreases faster than it increases (cf. Fig. 6.4).

Illustration. For an illustration of w^\leftarrow -fitting, consider the fixed-point diagram of Fig. 6.9, which was introduced in §6.5.4. This diagram contains the window-update function w^\leftarrow , which has been estimated with the approach in Appendix E.5 for the experiment in Fig. 6.4. Moreover, the window sizes $\{w_e(t)\}_{t \in \{1, \dots, 12\}}$ from that experiment are also located in the fixed-point diagram. These window sizes are consecutively attained via the congestion-window evolution, which can be visualized as a series of movements with alternating directions (visualized by arrows). Ideally, these movements should correspond to alternating projections

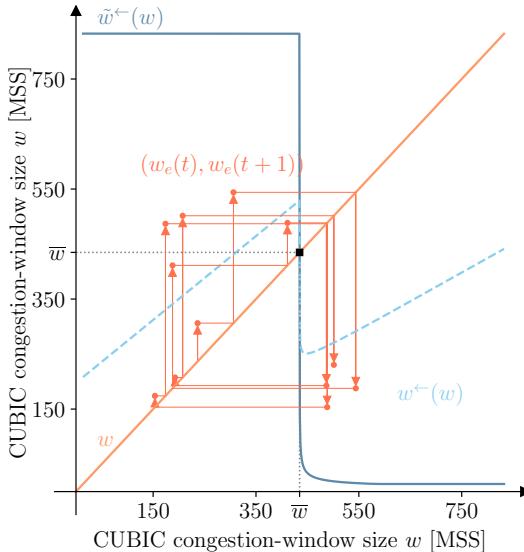


Figure 6.9: The empirically estimated window-update function w^\leftarrow is approximately consistent with the oscillation trajectory from experiment e , i.e., $w^\leftarrow(w_e(t)) \approx w_e(t+1)$.

onto the estimated w^\leftarrow and onto the identity line w . Achieving this consistency is the goal of the estimation approach in Appendix E.5. However, since w^\leftarrow has been estimated based on a function template and multiple experiments, it is only approximately consistent with the actual trajectory from the experiment in Fig. 6.4.

6.7.2 Identifying Oscillatory Networks

With an empirically grounded instance of w^\leftarrow , we can predict whether a given network will experience oscillation. For this prediction, we check whether the parameters of the given network satisfy the condition in Theorem 6.3. To evaluate the strictness of this condition, we test a range of parameter combinations similar to the experiment configurations from Fig. 6.8. In particular, we simultaneously vary two critical parameters of the configuration, where a configuration includes a bottleneck capacity $C_\ell \in [1, 200]$ Mbps (default: 100 Mbps), a path propagation delay $\tau^P \in [1, 100]$ ms (default: 40 ms), the bottleneck-link propagation delay τ_ℓ^P in percent of τ^P (default: 25%), and the bottleneck buffer capacity $B_\ell \in [0.1, 3]$ as a multiple of path BDPs (default: 1.5). Since these parameters are sufficiently similar

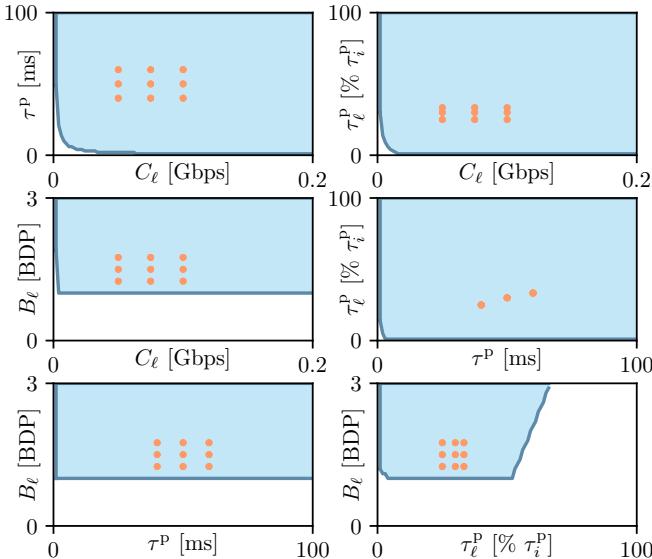


Figure 6.10: Network-parameter subspace (blue) that satisfies the oscillation condition from Theorem 6.3, based on an empirical w^\leftarrow . The dots mark the experiment configurations from Fig. 6.8.

to the experiment configurations in Fig. 6.8, we take caution not to over-generalize the validity of the empirical w^\leftarrow , which is based on these experiments.

The results of this parameter exploration are depicted in Fig. 6.10, indicating that oscillation is fostered by large network parameters C_ℓ , τ^P , and B_ℓ . Moreover, the demarcation of the oscillatory parameter sub-space is confirmed by the experiments from Fig. 6.8, as the configurations of these oscillatory experiments are all within the oscillatory subspace (red dots in Fig. 6.10).

6.8 PREVENTING OSCILLATION

The preceding sections confirm that rate oscillation in BBR/ CUBIC competition is a frequent problem with severe fairness implications. Therefore, we now consider methods to prevent this oscillation. To that end, we discuss potential countermeasures and their drawbacks in §6.8.1. In §6.8.2, we evaluate whether and how BBRv2 avoids oscillation.

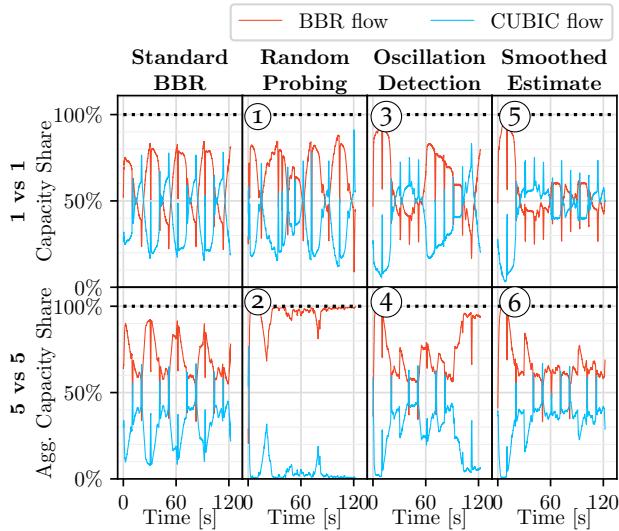


Figure 6.11: Evaluation of strategies for oscillation suppression. ‘Random Probing’ desynchronizes RTT-probing steps, ‘Smoothed Estimate’ makes min-RTT estimate a moving average, ‘Oscillation Detection’ freezes min-RTT estimate in case of high estimate variance.

6.8.1 Evaluating Countermeasures

As explained in §6.4, BBR/CUBIC oscillation happens because the BBR flows simultaneously perform RTT probing by briefly, but sharply reducing their rate. Therefore, the BBR flows regularly estimate a relatively low minimum RTT, especially if the CUBIC flows are small at the time of probing. This low minimum-RTT estimate then decreases the BBR sending rate (via the BBR congestion window), increases the CUBIC sending rate, and thus leads to a higher minimum-RTT estimate at the next RTT probing. If that next minimum-RTT estimate is high enough, the evolution of the sending rates is reversed, causing oscillation.

This causal chain may be disrupted by a number of possible modifications to the BBR algorithm, which we discuss and experimentally evaluate in the following.

Randomize RTT probing. As described above, oscillation requires that all BBR flows simultaneously perform RTT probing, namely every 10 seconds. If these RTT-probing periods were randomized, each flow would probe the RTT when other BBR

flows still contribute to the queue. Hence, the high minimum-RTT estimates would be consistently high rather than varying over time. While such randomization might avoid oscillation, it is undesirable for three reasons.

First, the synchronization of the RTT probing among BBR flows is a conscious feature of BBR, enabling BBR flows to discover the path propagation delay in a pure BBR scenario; this discovery would be prevented by randomization. Second, randomization does not prevent oscillation given a single BBR flow, as visible in Fig. 6.11 ①. Given a single BBR flow, no other BBR flows exist that could inflate the minimum-RTT estimate. Hence, the rate of the randomized single BBR flow still oscillates, although not in 10-second steps anymore, but in intervals with varying duration. Third, randomization can suppress oscillation only at the cost of even lower fairness, as the experiment in Fig. 6.11 ② demonstrates: Since randomization causes consistently high minimum-RTT estimates, it also causes consistently high BBR rates and near-starvation of the CUBIC flows.

Detect oscillation and freeze. To suppress oscillation in a more targeted fashion, we envision that a BBR flow (i) keeps a recent history of its minimum-RTT estimates, (ii) maintains the mean μ and the standard deviation σ of these estimates, and (iii) concludes that oscillation is ongoing if the standard deviation σ exceeds a configured share κ of the mean μ . In case of oscillation, the BBR flow then considers μ its minimum-RTT estimate.

Unfortunately, this oscillation-suppression strategy is self-defeating: When oscillation is suppressed for a sufficiently long time, the variance of minimum-RTT estimates decreases, the oscillation-suppression mechanism is deactivated, and oscillation resumes. Hence, this mechanism does not eliminate the oscillation, but only prolongs the oscillation period (see experiments ③ and ④ in Fig. 6.11). Moreover, the mechanism relies on a suitable value κ to distinguish structural oscillation from acceptable fluctuation of the minimum-RTT estimates, which might be difficult to find in practice.

Restrict update of minimum-RTT estimate. Since the variance of minimum-RTT estimates over time causes oscillation, oscillation might be suppressed by restricting the adjustment of these minimum-RTT estimates. For example, the minimum-RTT estimate might be slowly adjusted to new RTT measurements using a moving average. Indeed, this smoothing of the minimum-RTT estimate can dampen, but not completely eliminate oscillation (experiments ⑤ and ⑥ in Fig. 6.11). While being the most promising among the evaluated countermeasures, this smoothing

approach reduces responsiveness when a path change alters the propagation delay, which may be acceptable in fixed networks, but is problematic in mobile networks.

6.8.2 Evaluating BBRv2

The previous section demonstrates that suppressing oscillation in BBR/CUBIC competition with ad-hoc changes to the BBR algorithm is difficult and comes with several drawbacks. In this section, we show that also a more fundamental redesign of BBR represented by the BBRv2 algorithm [50] eliminates oscillation, but introduces new problems.

Fundamentally, BBRv2 eliminates oscillation by reducing the negative impact of the BBR congestion window. Remember that this congestion window has a size of two estimated BDPs, where the BDP is estimated based on minimum-RTT estimates. While this congestion window still exists in BBRv2, BBRv2 has introduced additional limits on the inflight volume that are activated on packet loss and are stricter than the BBR congestion window. In particular, BBRv2 has a long-term inflight bound (*inflight_hi*), which is at most $5/4$ estimated BDPs, and a short-term inflight bound (*inflight_lo*), which is at most one estimated BDP; moreover, both these bounds are below maximum if loss occurs frequently, as in the competition with CUBIC. Hence, the inflight volume of a BBRv2 flow is considerably less sensitive to minimum-RTT estimates. As a result, varying minimum-RTT estimates do not translate into large changes of the sending rate, which prevents oscillation. This insight is confirmed by the experiments in Fig. 6.12, which also includes predictions by the BBRv2 fluid model by Scherrer et al. [5].

However, the BBRv2 features that suppress oscillation have two main drawbacks. First, BBRv2 flows might obtain less than the fair share of bandwidth, e.g., all cases with more than 4 BBRv2 flows in Fig. 6.12 show sub-proportional throughput for BBRv2. This lack of assertiveness might hamper the adoption of BBRv2 [181]. Second, the tight and relatively static inflight bounds of BBRv2 hurt its ability to quickly increase its sending rate when additional bandwidth becomes available, as demonstrated by previous work on BBRv2 [263].

In summary, we conclude that avoiding oscillation in BBR/CUBIC competition is hard to achieve without damaging fairness or responsiveness of the BBR algorithm. Designing an oscillation-free BBR version without major performance drawbacks is an interesting task for future work.

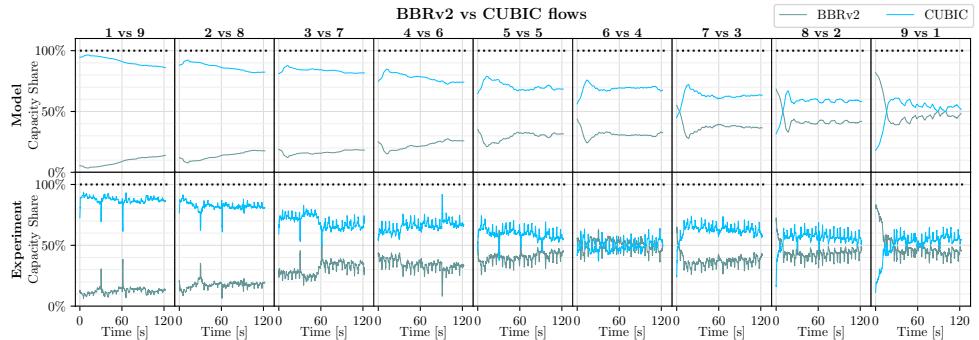


Figure 6.12: BBRv2/CUBIC competition for different CCA combinations (Aggregate capacity share by CCA).

6.9 RELATED WORK

Most relevant related work. As apparent throughout the chapter, our work has a close connection to the two previous steady-state models of BBR/CUBIC competition [181, 254], and the fluid models for CUBIC [249] and BBR [5]. In addition, we also emphasize the relevance of recent work by Arun et al. [26], who investigate the delay oscillations caused by delay-bounding CCAs (such as BBR) and loss-based CCAs (such as CUBIC) in homogeneous settings. In particular, they find that the amplitude of these delay oscillations should exceed the random, non-congestive jitter that is expected, in order to avoid starvation of flows measuring a distorted RTT. Our work complements this insight by showing that another type of oscillation can result from the *competition* of CCAs. Interestingly, we also identify a model condition for BBR starvation involving delay measurements (i.e., $\alpha \leq 1$ from Lemma 6.3), but demonstrate that BBR can recover from this starvation in the long term when competing with CUBIC.

Congestion-control models. Congestion-control algorithms (CCAs) have been analyzed with a wide range of modeling techniques [201, 241].

For example, *steady-state models* describe only the equilibrium of CCA execution in terms of network metrics. Most influentially, Mathis et al. [171] and Padhye et al. [203] provide closed-form functions yielding long-term Reno throughput based on RTT and packet loss; their methodology has been extended to short-lived flows and other CCAs [30, 51, 150].

In contrast, *dynamic models* represent the full CCA behavior over time. This evolution is sometimes represented in discrete time [16, 28, 207, 268], but more often in continuous time by fluid models. Specific fluid models exist for Reno [166, 182], Vegas [37], CUBIC [249], and both BBR versions [5]. While these models have been applied to analyze fairness and stability [135, 139, 141, 163], our work is the first to analytically investigate inter-CCA fairness with a dynamic fluid model.

BBR. Motivated by the excessive queuing of loss-based CCAs (e.g., Reno [82], CUBIC [110]) and competitiveness issues of latency-based CCAs (e.g., Vegas [39]), Google proposed the BBR CCA in 2017 [49]. BBR was enabled for YouTube soon afterwards [48], and was used by around 40% of Internet traffic in 2019 [180]. Given that BBR competes with other CCAs in the Internet, its fairness towards other CCAs has received much attention. In a first independent study, Hock et al. [119] demonstrate that BBR is over-aggressive against loss-based CCAs for settings with small buffers, and under-aggressive given large buffers. In these large buffers, the BBR congestion window restricts the BBR sending rate, which is confirmed by experiment-based follow-up work [75, 231, 246, 262]. This insight about BBR fairness triggered the release of BBRv2 [50], which improves inter-CCA fairness thanks to higher loss sensitivity, but is less responsive than BBR version 1 [143, 193, 240, 263].

Congestion-control fairness. In evaluating CCAs, much attention is traditionally devoted to *fairness*, i.e., the equality of resource sharing under distributed CCA execution. This fairness is typically measured by some aggregation (e.g., Jain fairness index [133]) of the throughput-share distribution across flows on a bottleneck link [49, 75, 119, 181, 231, 254]. However, researchers have argued that fairness measures should instead focus on flow-completion times [77] or quality of experience (QoE) [123], emphasize compatibility with CCAs already deployed in the Internet [255], or avoid flows as central entities of the fairness definition [41]. Regarding fairness between different CCAs, TCP friendliness (i.e., fairness towards the Reno CCA) has been shown to require a fundamental trade-off with other goals such as throughput and responsiveness [43, 268]. In this work, we also demonstrate that fairness has an important time component: If throughput shares are averaged over a long enough time, high transitory inequality might be obscured.

6.10 CONCLUSION

To better understand BBR/CUBIC competition over time, we provide a dynamic perspective on BBR fairness in this chapter. In this dynamic perspective, we document persistent sending-rate oscillation in BBR/CUBIC competition, and leverage fluid models to provide a mathematically rigorous explanation for this oscillation. Moreover, we show that this oscillation occurs frequently and causes recurring unfair rate distributions. While these extreme rate distributions are transient and (loosely) bounded in their unfairness, they substantially differ from the long-term average rate distribution, and thus matter for the fairness among short flows.

Furthermore, this sending-rate oscillation can have severe ramifications beyond fairness. In particular, the continuous change in flow-specific throughput is challenging to handle for throughput-sensitive applications. For example, video-conferencing applications that use adaptive bitrate (ABR) algorithms might experience quality-of-experience oscillation if the video quality is adapted to the oscillation in flow throughput. Crucially, such ABR algorithms might even intensify the sending-rate oscillation if the application increases demand when throughput increases. For future work, it will be interesting to investigate the effects of BBR/CUBIC oscillation on the performance of real-world applications.

Part II
ISP ECONOMICS

PART OVERVIEW

In the first part of this thesis, the focus is on end-hosts and their traffic-sending behavior, expressed by algorithms for end-host path selection and congestion control. For the following thesis part, I change the focus regarding both the agent type and the area of interaction: More concretely, the following two chapters are related to Internet Service Providers (ISPs) and their economic strategies.

However, the model-based analysis of ISP economic strategies is made both interesting and challenging by the same features as end-host network usage. First, ISP economics are shaped by the underlying technical mechanisms in the Internet, in particular the mechanisms for routing and forwarding. As a result, ISP economics are impacted in manifold ways by the emerging paradigm of path-aware networking. Second, ISP economics has strategic and dynamic aspects, and therefore requires to combine game theory with stability theory. Third, ISP economics are inherently networked, i.e., the underlying network topology constrains which ISPs can interact with each other other in which manner.

This interaction among ISPs can take the forms of *cooperation* and *competition*, which are discussed as follows:

ISP Cooperation: Interconnection Agreements. Cooperation among ISPs is most concretely embodied by interconnection agreements, in which ISPs determine how each agreement party should forward traffic from other parties. Today, these interconnection agreements are subject to the Gao-Rexford conditions, i.e., technical constraints that ensure BGP convergence, and therefore become irrelevant in path-aware networking. These new opportunities for interconnection agreements are explored in Chapter 7. In particular, Chapter 7 illustrates (i) how to structure interconnection agreements that are beneficial to all involved parties, (ii) how to efficiently negotiate these agreements despite private information and incentives for dishonesty, and (iii) how these novel agreements will affect path diversity in the Internet.

ISP Competition: Quality Attributes. While ISPs cooperate in order to provide Internet paths to customers, these customers usually have the choice between multiple paths towards a destination, which gives rise to ISP competition. Under this competitive pressure, ISPs might differentiate themselves by investing in the

quality of their offered paths, e.g., in bandwidth, latency, security, or ecological concerns. Importantly, these quality investments should optimize profit in the context of alternative offers by other ISPs. To inform quality-investment decisions and analyze these competitive dynamics, Chapter 8 presents a game-theoretic competition model that is applicable to the traditional BGP-based Internet as well as to a future path-aware Internet. The model provides various insights on competition effects on path quality and ISP profits, including the impacts of competition intensity, ISP heterogeneity, and ISP collusion.

ISP COOPERATION: INTERCONNECTION AGREEMENTS

7.1 INTRODUCTION

In contrast to today's Internet, *path-aware networks* (PANs) enable end-hosts to choose their forwarding path at the level of autonomous systems (ASes), which is then embedded in the header of data packets. As such, end-hosts are not limited to using a single path between a pair of ASes, but can use multiple paths simultaneously. This multi-path availability offers new opportunities and challenges for both end-hosts and network operators. For end-hosts, PANs present new questions regarding *path selection*, some of which are discussed in Chapters 2–4 of this thesis. For network operators, PANs modify the economic environment in the Internet by expanding the opportunities for cooperation and competition among ISPs.

In this chapter, we focus on ISP cooperation in PANs, as embodied by *interconnection agreements* between ASes possible in such architectures. These interconnection agreements are highly relevant for both network operators and end-hosts. From the perspective of ASes, interconnection agreements determine the economic opportunities offered by PANs, which are critical to PAN adoption. From the perspective of end-hosts, interconnection agreements play an essential role for path diversity; the extent of path diversity, in turn, influences the magnitude of the resilience and efficiency improvements of multi-path availability.

In this context, we observe that PAN architectures enable new types of interconnection agreements that are not possible in today's Internet. Nowadays, interconnection agreements are heavily influenced by the Gao–Rexford conditions (henceforth: GRC) [64, 91], which prescribe that traffic from peers and providers must not be forwarded to other peers or providers. It is important to distinguish between two aspects of the GRC which refer to independent concerns: a stability aspect and an economic aspect. Regarding stability, the GRC provably imply route convergence of the Border Gateway Protocol (BGP) [91]. Regarding economics,

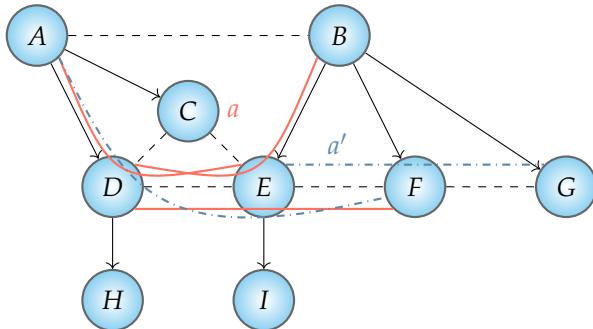


Figure 7.1: AS topology with interconnection agreements a and a' (discussed in §7.3.2).
Peering links are shown as dashed lines, provider–customer links as “provider → customer”.

the GRC signify that an AS only forwards traffic if the cost of forwarding can be directly recuperated from customer ASes or end hosts.

However, PAN architectures no longer require the GRC for providing stability. While paths in PAN architectures are discovered similarly as in BGP, namely by communicating path information to neighboring ASes, data packets are forwarded along a path selected by the packet source, which is embedded in the packet headers. Thus, PAN architectures directly achieve convergence in the sense of achieving a consistent view of the used forwarding paths, as we will explain in §7.2.

PAN architectures therefore present the exciting opportunity to create and use GRC-violating paths—if such paths can be made economically viable. In particular, we observe that PAN architectures may no longer require the GRC for reasons of stability, but must still respect the economic logic that makes the GRC a rational forwarding policy. For example, while the creation of GRC-violating path ADE by D in Fig. 7.1 may not lead to convergence problems, the path still is economically undesirable for D , because D would incur a charge from its provider A for forwarding traffic of E , which it cannot recuperate due to E 's status as a peer.

In this chapter, we tackle this challenge by proposing new interconnection agreements based on *mutuality*, a concept that is already present in peering agreements today, but can be leveraged to set up more complex and flexible agreements. Concretely, mutuality means that the mentioned example path ADE could be rendered economically viable for D by requiring a *quid pro quo* from E , the main beneficiary of the path. For example, E could offer path DEB to D such that both D and E could save transit cost for accessing ASes B and A , respectively, but incur additional

transit cost for forwarding their peer's traffic to their respective provider. Moreover, ASes *D* and *E* might as well provide each other with access to their peers *C* and *F*, thereby saving transit cost while experiencing additional load on their network. If the flows over the new path segments are properly balanced and especially if the new path segments allow ASes *D* and *E* to attract additional revenue-generating traffic from their customers, such unconventional agreements can be mutually beneficial. Hence, PANs offer opportunities for profit maximization which are not present in today's Internet.

7.1.1 Contributions

This chapter is based on my publication at DSN 2021 [3], and contains the following contributions:

Agreement conditions based on AS business model. Concluding mutuality-based agreements affects revenue and cost of the AS parties in many ways, which requires a careful structuring of such agreements. We envisage that such agreements contain conditions that must be respected in order to preserve the positive value of the agreement for both parties. To this end, we further present a formal model of AS business calculations and AS interconnections that allows to derive two different types of agreement conditions, namely conditions based on flow volumes and conditions based on cash compensation. Moreover, we show how to shape mutuality-based agreements to maximize the utility (i.e., the profit) obtained by both parties.

BOSCO negotiation mechanism. When designing an optimal structure for mutuality-based interconnection agreements, we demonstrate that this optimization relies on private information of the AS parties, and that an AS party benefits from unilateral dishonesty. Furthermore, if both parties are dishonest, negotiations of mutuality-based agreements may become inefficient in the sense that the agreements are not concluded even if mutually beneficial. To limit this inefficiency, we present BOSCO, a negotiation mechanism with a game-theoretic foundations and a range of provable desirable properties.

Path-diversity effects. Finally, we investigate the effect of mutuality-based agreements on path diversity, building on a combination of several publicly available datasets [53–55, 173]. Our results underpin the benefits of mutuality-based agreements, which provide ASes access to thousands of additional paths, many of

which are considerably more attractive regarding latency and bandwidth than the previously available paths.

7.2 THE RELEVANCE OF GRC FOR BGP AND PANS

To clarify why the GRC are needed in a BGP/IP-based Internet but not in PAN architectures, we compare their convergence requirements using the example topology of Fig. 7.1.

The fundamental issue with convergence in BGP is the next-hop principle: ASes can only select a next-hop AS for their traffic and thus rely on that AS to forward the traffic along the route that was originally communicated via BGP. If this assumption is violated—even temporarily—routing loops can arise.

For example, suppose that ASes *D* and *E* forwarded routes from their respective providers *A* and *B* to each other, which violates the GRC. Assuming both *D* and *E* prefer routes learned from peers, this results in a (slightly extended) instance of the classical DISAGREE example [105], which does converge with BGP but non-deterministically. The non-determinism of such topologies, which are also known as “BGP wedgies” [103], is undesirable, but it does not constitute a fundamental problem for convergence in BGP. However, adding a single additional AS *C*, which concludes similar agreements with both *D* and *E*, this topology leads to the famous BAD GADGET problem, which causes persistent route oscillations [105].

This susceptibility to oscillations is also worrisome because seemingly benign topologies and policies may easily reduce to the BAD GADGET in case one network link fails [105]. This shows that GRC-violating policies need to be implemented very carefully and with coordination among all involved parties to ensure routing stability (e.g., using BGP communities). As a consequence, “sibling” agreements in which two ASes provide each other access to their respective providers (as presented above) generally only exist between ASes controlled by a single organization.

Unlike IP, PANs forward a packet along the path encoded in its header. Thus, there is no uncertainty about the traversed forwarding path after the next-hop AS and routing loops can be prevented. For example, if a source in *D* would encode path *DEBA* in packets sent to a receiver in *A*, *E* would not send these packets back to *D*. Precautions like the GRC are therefore not required for stability in a PAN and ASes have substantially more freedom in deciding which interconnection agreements to conclude and which paths to authorize.

7.3 MODELLING INTERCONNECTION ECONOMICS

In this section, we describe our model of the economic interactions between ASes in the Internet, which allows to derive quantitative conditions that must be fulfilled by interconnection agreements.

7.3.1 AS Business Calculation

Network topology. We model the Internet as a mixed graph $\mathcal{G} = (\mathcal{A}, \mathcal{L}_{\leftrightarrow}, \mathcal{L}_{\uparrow})$, where the nodes \mathcal{A} correspond to ASes, the undirected edges $\mathcal{L}_{\leftrightarrow}$ correspond to peering links, and the directed edges \mathcal{L}_{\uparrow} correspond to provider–customer links. An edge $(X, Y) \in \mathcal{L}_{\uparrow}$ corresponds to a link from provider X to customer Y . An AS node $X \in \mathcal{A}$ is connected to a set of neighbor ASes that can be decomposed into a provider set $\pi(X)$, a peer set $\varepsilon(X)$, and a customer set $\gamma(X)$. For simplicity of notation, we denote the customer end-hosts of X as a virtual stub AS $\Gamma_X \in \gamma(X)$, connected over a virtual provider–customer link ℓ' .

Flows. In our model, f_{ℓ} for $\ell \in (\mathcal{L}_{\leftrightarrow} \cup \mathcal{L}_{\uparrow})$ denotes the traffic volume carried by link ℓ , and f_X for $X \in \mathcal{A}$ denotes the traffic volume carried by node X . Furthermore, let the flow f_{XY} be the share of f_X that also flows directly to or from X 's neighbor Y . These sub-flows are represented in vector \mathbf{f}_X , i.e., $(\mathbf{f}_X)_Y = f_{XY}$. Analogously, f_{XYZ} is the flow volume on the path segment consisting of ASes X , Y , and Z in that order, but independent of direction.

Link pricing. Each provider–customer link $\ell = (X, Y) \in \mathcal{L}_{\uparrow}$ has a corresponding pricing function $p_{\ell}(f_{\ell})$, yielding the amount of money X receives from Y given flow volume f_{ℓ} on link ℓ . This flow volume f_{ℓ} can be interpreted as is appropriate for the pricing function, e.g., as the median, average, or 95th percentile of traffic volume over a given time period. Each pricing function $p_{\ell}(f_{\ell})$ is of the form

$$p_{\ell}(f_{\ell}) = \alpha_{\ell} f_{\ell}^{\beta_{\ell}}, \quad (7.1)$$

where $\alpha_{\ell} \geq 0$ and $\beta_{\ell} \geq 0$ are pricing-policy parameters. For example, $\beta_{\ell} = 0$ corresponds to flat-rate pricing with flow-independent fee α_{ℓ} , $\beta_{\ell} = 1$ corresponds to pay-per-usage pricing with traffic-unit cost α_{ℓ} , and $\beta_{\ell} > 1$ results in a superlinear pricing function, e.g., as given in congestion pricing. For simplicity, we assume that all peering links $\ell' \in \mathcal{L}_{\leftrightarrow}$ are settlement-free, as usual in the literature [125]. Paid-peering links can be represented in the model as provider–customer links. We write $p_{XY} = p_{(X,Y)}$ for brevity.

Table 7.1: Notation used in model (in alphabetical order).

Symbol	Description
\mathcal{A}	Set of autonomous systems (ASes)
c_X	Cost of AS X
Γ_X	Stub AS containing end-host customers of AS X
$\gamma(X)$	Set of customer ASes of AS X
$\Delta f_P^{(a)}$	Volume of customer traffic on path segment P newly attracted thanks to agreement a
$\epsilon(X)$	Set of peer ASes of AS X
\mathbf{f}_X	Traffic-volume pattern of AS X with respect to its neighbor ASes
$\mathbf{f}_X^{(a)}$	Traffic-volume pattern of AS X with respect to its neighbor ASes, given agreement a
$f_P^{(a)\downarrow}$	Volume of previously forwarded traffic newly allocated to path segment P , given agreement a
f_X	Traffic volume forwarded by AS X
f_{XYZ}	Traffic volume forwarded over path segment composed of ASes X , Y , and Z (in that order, but in both directions)
i_X	Internal forwarding cost of AS X
$\mathcal{L}_{\leftrightarrow}$	Set of peering links
\mathcal{L}^{\uparrow}	Set of provider-customer links
r_X	Revenue of AS X
p_ℓ	Price of link ℓ (based on link load f_ℓ)
$\Pi_{X \leftarrow Y}$	Monetary transfer sum paid by AS X to AS Y
$\pi(X)$	Set of provider ASes of AS X
\mathcal{U}_X	Profit of AS X
$u_X(a)$	Utility (profit impact) of agreement a to AS X

Internal forwarding cost. In addition to charges defined by pricing functions, an AS X incurs an internal cost according to an internal-cost function $i_X(f_X)$, which is non-negative and monotonously increasing in the flow f_X through X.

Utility. The utility (or profit) $\mathcal{U}_X(\mathbf{f}_X) = r_X(\mathbf{f}_X) - c_X(\mathbf{f}_X)$ of an AS X is the difference between the revenue $r_X(\mathbf{f}_X)$ obtained and the costs $c_X(\mathbf{f}_X)$ incurred by traffic distribution \mathbf{f}_X :

$$r_X(\mathbf{f}_X) = \sum_{Y \in \gamma(X)} p_{XY}(f_{XY}), \quad (7.2a)$$

$$c_X(\mathbf{f}_X) = i_X(f_X) + \sum_{Y \in \pi(X)} p_{YX}(f_{XY}). \quad (7.2b)$$

This simple model can already formalize some important insights. For example, consider ASes A, D, and H in Fig. 7.1, connected by provider–customer links (A, D) and (D, H). For D to make a profit, i.e., $\mathcal{U}_X(\mathbf{f}_X) > 0$, it must hold that $r_D(\mathbf{f}_D) > c_D(\mathbf{f}_D)$. This in turn implies

$$p_{DH}(f_{DH}) + p_{D\Gamma_D}(f_{D\Gamma_D}) > p_{AD}(f_{AD}) + i_D(f_D), \quad (7.3)$$

i.e., the revenue from H and D’s customer end-hosts must cover the cost induced by charges from A as well as internal cost.

7.3.2 Interconnection Agreements

We denote an interconnection agreement a between two ASes X and Y in terms of the respective neighbor ASes to which X and Y gain new paths thanks to the agreement:

$$a = [X(\uparrow \pi'_X, \rightarrow \varepsilon'_X, \downarrow \gamma'_X); Y(\uparrow \pi'_Y, \rightarrow \varepsilon'_Y, \downarrow \gamma'_Y)] \quad (7.4)$$

Here $\pi'_X \subseteq \pi(X)$, $\varepsilon'_X \subseteq \varepsilon(X)$ and $\gamma'_X \subseteq \gamma(X)$ are the providers, peers, and customers of AS X, respectively, to which Y obtains access through the agreement (analogously for π'_Y , ε'_Y , γ'_Y). Furthermore, we introduce the notation $a_X = \pi'_X \cup \varepsilon'_X \cup \gamma'_X$, and an analogous notation for Y.

Next, we formalize the utility of interconnection agreements. Let the utility $u_X(a)$ of agreement a to X be the difference in \mathcal{U}_X produced by changes in flow composition due to agreement a , i.e.,

$$u_X(a) = \mathcal{U}_X(\mathbf{f}_X^{(a)}) - \mathcal{U}_X(\mathbf{f}_X) = \Delta r_X - \Delta c_X, \quad (7.5)$$

where $f_X^{(a)}$ is the distribution of traffic passing through X if agreement a is in force, and Δr_X and Δc_X are agreement-induced changes in revenue and cost of X , respectively.

7.3.2.1 Example of Peering Agreement

Consider the negotiation of a classic peering agreement between ASes D and E in Fig. 7.1, which so far have been connected by providers A and B . We assume that ASes D and E are pure transit ASes, i.e., there are no customer end-hosts within these ASes.

Formalization. In a classic peering agreement, both ASes provide each other access to all of their respective customers. Using the notation introduced above, this agreement is formalized as $a_p = [D(\downarrow \{H\}); E(\downarrow \{I\})]$. The change in revenue for D , namely

$$\Delta r_D = p_{DH} \left(f_{DH}^{(a_p)} \right) - p_{DH}(f_{DH}), \quad (7.6)$$

results from changes in flows to D 's customer H , driven by the new peering link $\ell'(D, E)$. The changes in cost to D ,

$$\Delta c_D = i_D \left(f_D^{(a_p)} \right) - i_D(f_D) + p_{AD}(f_{AD} - f_{DABE}) - p_{AD}(f_{AD}), \quad (7.7)$$

result from changes in internal and provider cost. In Eq. (7.7), it is assumed that all traffic f_{DABE} originally reaching AS E via ASes A and B is reallocated to the direct link (D, E) made usable by the agreement, and hence does not contribute to charges by AS A anymore. The utility of a peering agreement to D is then $u_D(a_p) = \Delta r_D - \Delta c_D$.

Interpretation. The strongest rationale for peering agreements is that the agreement leads to considerable cost decrease, i.e., a strongly negative Δc_D , as provider A can be avoided for any traffic f_{DE} . The new peering link may also attract additional traffic from customer H (e.g., due to the lower latency of the new connection), thus increasing D 's revenue. If $\Delta r_D > \Delta c_D$, agreement a_p has positive utility $u_D(a_p) > 0$ and is worth concluding from D 's perspective. However, D may also experience a substantial increase in internal cost (Δi_D) due to peering, with little savings in provider cost and no extra income from the additionally attracted traffic (e.g., due to flat-rate fees). In such a case, $u_D(a_p)$ is negative, and the agreement is not attractive

to D . For agreement a_p to be concluded, both $u_D(a_p)$ and $u_E(a_p)$ need to be non-negative (or, if cash transfers as in paid peering [267] are used, $u_D(a_p) + u_E(a_E)$ would need to be non-negative such that one party could compensate the other party and still benefit from the agreement).

7.3.2.2 Example of Novel Mutuality-Based Agreement

As discussed in §7.2, the GRC are not necessary for stability in a PAN, which allows for new types of interconnection agreements.

Formalization. In the example topology of Fig. 7.1, the following agreement a could not be concluded in today's BGP-based Internet, but could be concluded in a path-aware inter-domain network:

$$a = [D(\uparrow \{A\}); E(\uparrow \{B\}, \rightarrow \{F\})] \quad (7.8)$$

In this agreement a , D offers E access to its provider A , whereas E in return provides D with access to its provider B and its peer F .

The agreement utility $u_D(a)$ of agreement a to D can be derived similarly to the peering-agreement example in §7.3.2.1. Namely, the changes in revenue and cost of D are

$$\Delta r_D = \sum_{X \in \gamma(D)} p_{DX}(f_{DX}^{(a)}) - p_{DX}(f_{DX}), \quad (7.9a)$$

$$\Delta c_D = i_D(f_D^{(a)}) - i_D(f_D) + \sum_{Y \in \pi(D)} p_{YD}(f_{DY}^{(a)}) - p_{YD}(f_{DY}), \quad (7.9b)$$

where

$$f_{DY}^{(a)} = f_{DY} + f_{EDY}^{(a)} - \sum_{Z \in a_E} f_{DY}^{\uparrow}(Z, E) \quad (7.9c)$$

is the flow from D to one of its providers, Y , after conclusion of the agreement. This flow towards the provider is increased by the traffic $f_{EDY}^{(a)}$ that D transfers to Y for E in accord with the agreement. Simultaneously, flow f_{DY} is decreased by the flow $f_{DY}^{\uparrow}(Z, E)$ to the destinations $Z \in a_E$ that was previously forwarded via provider Y , but is newly forwarded over E thanks to agreement a . As the new paths over agreement partner E are the reason for newly attracted traffic from D 's customers, we assume here that all such newly attracted traffic is forwarded over the agreement partner, not over D 's providers.

Interpretation. Clearly, agreement a is not per se attractive to ASes D and E . For D , the higher the amount of flow from E that newly must be forwarded to a provider AS, the less attractive the agreement to D , i.e., the higher Δc_D . In contrast, the higher the amount of flow offloaded to E , the higher the utility that D can derive from agreement a . Vice versa, the agreement utility for E is conversely affected by the size of these new flows. Thus, the agreement a must be qualified. Necessarily, these qualifications must guarantee positive agreement utility to both parties. Furthermore, it is desirable that the qualifications achieve Pareto-optimal [73] and fair agreement utility, i.e., no party's utility could be increased without decreasing the other party's utility, and the utility obtained by both parties is as similar as possible. In §7.4, we propose two different types of agreement qualifications to achieve these goals.

7.3.2.3 Extension of Agreement Paths

Thanks to agreement a , ASes D and E obtain access to the new path segments DEB and DEF (for D), or EDA (for E). As the motivation behind the agreement is the attraction of additional customer traffic, the agreement parties would provide access to the new path segments only to their respective customers. For example, D would extend the new path segment DEB to $HDEB$, but not to $ADEB$ or $CDEB$.

However, the new path segments can themselves become the subject of other agreements. For example, in an agreement a' between E and F , E could provide F with access to path segment EDA if F in return provides access to its peer G . Note that agreement a' must be negotiated such that the conditions defined in agreement a can still be respected, as these agreements are interdependent.

7.4 OPTIMIZATION OF MUTUALITY-BASED AGREEMENTS

The novel mutuality-based agreements should achieve Pareto-optimal and fair utility in order to be attractive to both agreement parties. Moreover, a necessary economic condition to conclude the agreement is the guarantee of non-negative agreement utility for both parties. Hence, defining an optimal interconnection agreement between two ASes D and E corresponds to solving the nonlinear program

$$\begin{aligned} & \text{maximize} && u_D(a) \cdot u_E(a) \\ & \text{subject to} && u_D(a) \geq 0, \quad u_E(a) \geq 0, \end{aligned} \tag{7.10}$$

where the objective is given by the Nash product [35, 194], which is only optimized for Pareto-optimal and fair values of $u_D(a)$ and $u_E(a)$. Hence, if the Nash bargaining product is optimized, no party can increase its utility without decreasing the other party's utility, and the utility of both parties is as similar as possible.

In the following, we present two methods to solve the nonlinear program in Eq. (7.10), i.e., two ways to qualify agreement a such that the constrained optimization problem is solved: The optimization method in §7.4.1 is based on defining flow-volume targets, which offers better predictability, whereas the method in §7.4.2 is based on cash transfers between the agreement parties, which offers more flexibility.

7.4.1 Optimization via Flow-Volume Targets

The constrained optimization problem in Eq. (7.10) can be solved by determining volume limits for the flows that traverse the new path segment created by agreement a .

Concretized optimization problem. Concretely, the general optimization problem can be instantiated by the nonlinear program

$$\begin{aligned} \max \quad & u_D\left(\mathbf{f}^{(a)}, \Delta \mathbf{f}^{(a)}\right) \cdot u_E\left(\mathbf{f}^{(a)}, \Delta \mathbf{f}^{(a)}\right) \\ \text{s.t. } & \Delta r_D\left(\mathbf{f}^{(a)}, \Delta \mathbf{f}^{(a)}\right) \geq \Delta c_D\left(\mathbf{f}^{(a)}, \Delta \mathbf{f}^{(a)}\right) \quad (\text{I-D}) \\ & \forall X \in a_E. \quad f_{DEX}^{(a)} \geq \sum_{Z \in \gamma(D)} \Delta f_{ZDEX}^{(a)} \quad (\text{II-D}) \quad (7.11) \\ & \forall X \in a_E. \forall Z \in \gamma(D). \quad \Delta f_{ZDEX}^{(a)} \leq \Delta f_{ZDEX}^{\max} \quad (\text{III-D}) \\ & + \text{constraints (I-E), (II-E), (III-E) where } D \leftrightarrow E. \end{aligned}$$

Here, $f_P^{(a)}$ refers to the total flow volume on a new path segment P allowed by the agreement, and $\Delta f_P^{(a)}$ is the volume of newly attracted customer traffic on path segment P after agreement conclusion. Hence, the flow volume on new path segment P that consists of rerouted existing traffic is at most $f_P^{(a)\downarrow} = f_P^{(a)} - \Delta f_P^{(a)}$.

Constraints. The constraints (I-D) and (I-E) capture the fact that the agreement must be economically viable for both parties. The constraints (II-D) and (II-E) capture the requirement that all the agreement-induced additional traffic from customers has to be accommodated within the flow allowances defined in the

agreement. Finally, as any agreement could be made viable by attracting enough additional customer traffic, the constraints (III-D) and (III-E) express that there is a limit Δf_P^{\max} to customer demand for new path segment P . The optimization problem can be solved by appropriately adjusting $f^{(a)}$, i.e., the total allowance for flows on the new path segments, and $\Delta f^{(a)}$, i.e., the amount of additionally attracted customer traffic on the new path segments. The resulting $f^{(a)}$ can then be included into the agreement as flow-volume targets, whereas the resulting $\Delta f^{(a)}$ can be used by each AS to optimally allocate the flow-volume allowance among its customers. If agreement paths are extended as discussed in §7.3.2.3, additional constraints may hold; in this chapter, we do not investigate these constraints.

Example optimization. We illustrate the optimization via flow-volume targets at the example of the mutuality-based agreement in §7.3.2.2. Hence, we attempt to find volume targets for the new flows $f_{EDA}^{(a)}$, $f_{DEB}^{(a)}$, and $f_{DEF}^{(a)}$. To simplify the derivation of flow-volume targets, we assign pricing functions to the links in the topology of Figure 7.1. Let pricing function p_ℓ of any provider–customer link in Figure 7.1 be a linear function, i.e., $\beta_\ell = 1$, with $\alpha_\ell > 0$. Moreover, let the internal-cost function i_X of an AS X be a linear function with unit cost $j_X > 0$, i.e., $i_X(f_X) = j_X f_X$. For ASes D and E , the constraint $\Delta r > \Delta c$ produces the following concrete constraints on the new sub-flows $f_{EDA}^{(a)}$, $f_{DEB}^{(a)}$, and $f_{DEF}^{(a)}$:

$$f_{EDA}^{(a)} \leq \frac{(\alpha_{DH} - j_D) \Delta f_{HDE}^{(a)} + \alpha_{AD} (f_{DEB}^{(a)\uparrow} + f_{DEF}^{(a)\uparrow})}{\alpha_{AD} + j_D} \quad (7.12a)$$

$$f_{DEB}^{(a)} \leq \frac{(\alpha_{EI} - j_E) \Delta f_{IED}^{(a)} + \alpha_{BE} f_{EDA}^{(a)\uparrow}}{\alpha_{BE} + j_E} - \frac{j_E}{\alpha_{BE} + j_E} f_{DEF}^{(a)} \quad (7.12b)$$

Intuitively, the flow that is newly forwarded for the agreement partner (e.g., $f_{EDA}^{(a)}$ in the case of D) is determined by the agreement-induced additional revenue from customers, the transit-cost savings from rerouting traffic, and the unit cost incurred by both internally and externally forwarding the partner’s traffic. Furthermore, the size of multiple new forwarded flows must be coordinated, e.g., as in case for E with Eq. (7.12b). As the new flows must accommodate the newly attracted traffic from customers, there are the additional constraints

$$f_{EDA}^{(a)} \geq \Delta f_{IEDA}^{(a)}, \quad f_{DEB}^{(a)} \geq \Delta f_{HDEB}^{(a)}, \quad \text{and} \quad f_{DEF}^{(a)} \geq \Delta f_{HDEF}^{(a)}. \quad (7.13)$$

Finally, the limits of customer demand for additional traffic must be respected:

$$\Delta f_{IEDA}^{(a)} \leq \Delta f_{IEDA}^{\max}, \quad \Delta f_{HDEB}^{(a)} \leq \Delta f_{HDEB}^{\max}, \quad \text{and} \quad \Delta f_{HDEF}^{(a)} \leq \Delta f_{HDEF}^{\max}. \quad (7.14)$$

Solving an optimization problem with the constraints in Eqs. (7.12)–(7.14) allows to determine the Pareto-optimal, fair size of volume limits for the new flows. As an example, we assume the following relations between the parameters:

$$\begin{aligned} \alpha_{AD} = \alpha_{BE} = j_D = j_E = \frac{1}{2} & \quad \alpha_{DH} = \alpha_{BE} = 3 \\ \Delta f_{HDEB}^{\max} = \Delta f_{HDEF}^{\max} = \frac{1}{4} & \quad \Delta f_{IEDA}^{\max} = \frac{1}{2} \end{aligned} \quad (7.15)$$

With these parameters, the optimal sizes of flow-volume targets are

$$\begin{aligned} f_{EDA}^{(a)} = \frac{1}{2}, & \quad f_{DEB}^{(a)} = \frac{1}{4}, & \quad f_{DEF}^{(a)} = \frac{3}{8}, \\ \Delta f_{IEDA}^{(a)} = \frac{1}{2}, & \quad \Delta f_{HDEB}^{(a)} = \frac{1}{4}, & \quad \Delta f_{HDEF}^{(a)} = \frac{1}{4}. \end{aligned} \quad (7.16)$$

Hence, ASes D and E can satisfy the complete demand of their customers for new path segments, and D can even forward up to

$$f_{DEF}^{(a)\downarrow} = f_{DEF}^{(a)} - \Delta f_{HDEF}^{(a)} = \frac{3}{8} - \frac{1}{4} = \frac{1}{8} \quad (7.17)$$

of pre-existing traffic over the new path segment DEF .

7.4.2 Optimization via Cash Compensation

Instead of fixing flow-volume targets, a non-technical approach could be based on cash transfers between the agreement parties. The idea of such an agreement structure is to abstain from limiting flow volumes, but to agree upon a cash payment π for compensating the party that benefits less or even stands to lose from the agreement.

Concretized optimization problem. Formally, negotiating an agreement between ASes D and E is equivalent to defining a cash sum $\Pi_{D \rightarrow E}$ from D to E (for negative $\Pi_{D \rightarrow E}$, E pays D) that solves the optimization problem

$$\begin{aligned} \max \quad & (u_D(a) - \Pi_{D \rightarrow E})(u_E(a) + \Pi_{D \rightarrow E}) \\ \text{s.t.} \quad & u_D(a) - \Pi_{D \rightarrow E} \geq 0, \quad u_E(a) + \Pi_{D \rightarrow E} \geq 0. \end{aligned} \quad (7.18)$$

In negotiation, the utilities $u_D(a)$ and $u_E(a)$ are estimated based on the expected volume of the newly enabled flows.

The optimization problem in Eq. (7.18) has a solution if and only if $u_D(a) + u_E(a) \geq 0$, i.e., one party gains at least as much as the other party loses and

can thus compensate the losing party while still benefiting from the agreement. If $u_D(a) + u_E(a) \geq 0$, Eq. (7.18) is always solved based on the Nash Bargaining Solution [194]:

$$\Pi_{D \rightarrow E} = u_D(a) - \frac{u_D(a) + u_E(a)}{2}. \quad (7.19)$$

7.4.3 Comparison of Optimization Methods

Profit safety. The main advantage of flow-volume targets over cash transfers is their higher predictability: As flow-volume agreements allow the agreement parties to enforce volume limits, they are more likely to guarantee positive agreement utility than cash-compensation agreements. The latter depend on ex-ante estimates of newly attracted customer traffic that might be incorrect, in which case the stipulated cash sum might not respect the constraints in Eq. (7.18).

Flexibility. Besides being easier to compute, an important advantage of cash compensation over volume targets is its larger flexibility, which translates into higher probability of the agreement being concluded as well as higher achievable joint utility. In certain settings where the revenues and costs of two ASes are very dissimilar, the flow-volume optimization problem in Eq. (7.12a) has a solution where all flow-volume targets are zero, i.e., the agreement cannot be concluded. In contrast, a cash-compensation agreement can always be concluded as long as the joint utility is positive.

Negotiation inefficiency. A common difficulty of both agreement structures is that they depend on private information of the negotiating parties, namely the charges from their respective providers, their internal forwarding cost, and the pricing for their customers, which determine the utility each party derives from the agreement. It cannot be assumed that the parties are willing to truthfully reveal this private information, as false claims about the cost structure strengthen a party's bargaining position. In §7.5, we show how the inefficiency arising from such private information can be limited by means of a bargaining mechanism.

7.5 MECHANISM-ASSISTED NEGOTIATION

In this section, we present a bargaining mechanism that we have designed to allow two interested parties to negotiate a mutuality-based interconnection agreement

Table 7.2: Notation used in our negotiation model (in alphabetical order).

Symbol	Description
$m_X(v_X)$	Probability of agreement conclusion given claim v_X
$\mathcal{N}(u_X, u_Y, v_X, v_Y)$	Nash bargaining product given true utility values u_X, u_Y , and utility claims v_X, v_Y
$\Pi_{X \leftarrow Y}$	Monetary transfer sum paid by party X to party Y
$q_X(v_X)$	Expected compensation of party X from party Y, given utility claim v_X
σ_X	Bargaining strategy of party X
σ_X^+	Best-response bargaining strategy of party X
σ_X^{++}	Nash-equilibrium bargaining strategy of party X
$\sigma_X(u_X)$	Utility claim of party X given bargaining strategy σ_X and true utility u_X
$\sigma_X^+ V_X$	Set of possible utility claims for party X, constructed by BOSCO
v_X	Utility claim by party X
\mathbb{U}_X	Probability distribution of utility for party X
u_X	Utility (profit impact) of agreement to party X
\bar{u}_X	Utility of agreement to party X, after monetary transfers

in an automated fashion, while reducing the negotiation inefficiency arising from bargaining under private information. However, while there are considerable advantages to using such a bargaining mechanism, there is no inherent necessity to use it; mutuality-based agreements might as well be negotiated by classic offline negotiations similar to classic peering agreements.

7.5.1 Problem Statement

When negotiating a mutuality-based interconnection agreement a , the agreement parties X and Y must agree on flow-volume targets or a cash sum transferred between the parties. The core difficulty of such negotiations is that the determination of the agreement conditions relies on $u_X(a)$ and $u_Y(a)$, i.e., the amount of utility that either party derives from the agreement, which is unknown to the respective other party. The presence of such private information allows each party to falsely

report a lower agreement utility than it really obtains, which leads to more favorable terms of the agreement for the dishonest party.

For example, when negotiating a cash-compensation agreement, the after-negotiation utility \bar{u}_X of party X is determined by

$$\bar{u}_X = u_X - \Pi_{X \rightarrow Y} = u_X - \frac{v_X - v_Y}{2}, \quad (7.20)$$

where v_X and v_Y are the values of the utility which X and Y *claim* to obtain from the agreement and which are used for determining the cash-compensation sum $\Pi_{X \rightarrow Y}$. To simplify our notation, we drop the reference to *a* here and in the remainder of the section, as we always consider a single agreement. Clearly, party X can increase \bar{u}_X by decreasing v_X , i.e., its utility claim. However, if both parties follow such a dishonest strategy, the apparent utility surplus $v_X + v_Y$ of the agreement tends to become negative, in which case the agreement seems to be not worth concluding, the negotiation breaks down, and both parties derive zero utility. Hence, the challenge in negotiating mutuality-based agreements (as for paid-peering agreements in today's Internet [267]) is the classic problem of *non-cooperative bilateral bargaining* [190].

In the game-theoretic literature, this problem is tackled by *mechanism design*, i.e., by structuring the negotiation in a way that minimizes the inefficiency of the result. Especially for inter-AS negotiation, such mechanisms have the additional advantage that they enable the automation and mathematical characterization of negotiations which nowadays are often informal and risky [45].

In this section, we present a bargaining mechanism with multiple desirable properties. We focus on negotiating cash-compensation agreements (cf. §7.4.2); adapting the mechanism for flow-volume agreements (cf. §7.4.1) is an interesting challenge for future work.

7.5.2 Desirable Mechanism Properties

Typically, desirable properties of bilateral-bargaining mechanisms include the following properties [190]:

P1 *Individual rationality.* Participation in the mechanism should be associated with non-negative utility in expectation (weak individual rationality) or in any outcome (strong individual rationality) for any party such that no party must be forced to take part in the mechanism.

P2 *Ex-post efficiency.* The mechanism should lead to conclusion of the agreement if and only if the utility surplus is non-negative.

P3 *Incentive compatibility.* The mechanism should structure the negotiation such that it is in a party's self-interest to be honest about its valuation of the agreement.

P4 *Budget balance.* The mechanism should neither require external subsidies nor end up with left-over resources (e.g., money) that are not ultimately assigned to the negotiating parties [195].

According to the famous Myerson–Satterthwaite theorem, no mechanism can satisfy the requirements P1, P2, and P4 simultaneously [190, 191]. The prominent Vickrey–Clarkes–Grove (VCG) mechanism, for example, guarantees individual rationality (P1) and ex-post efficiency (P2), but violates budget balance (P4) [65, 107, 251]. Absent government intervention, individual rationality and budget balance are necessary conditions for an inter-AS negotiation mechanism; we therefore sacrifice perfect ex-post efficiency and instead aim at maximizing the Nash bargaining product:

$$\mathcal{N}(u_X, u_Y, v_X, v_Y) = \begin{cases} (u_X - \Pi_{X \rightarrow Y})(u_Y + \Pi_{X \rightarrow Y}) & \text{if } v_X + v_Y \geq 0, \\ 0 & \text{otherwise,} \end{cases} \quad (7.21)$$

where the cash transfer is $\Pi_{X \rightarrow Y} = (v_X - v_Y)/2$.

While there are bargaining mechanisms that offer individual rationality, budget balance and incentive compatibility (according to the notion of Bayes–Nash incentive compatibility (BNIC) [188]), there are three arguments for relaxing the incentive-compatibility requirement as well.

Privacy. First, incentive compatibility might not be desired, because an AS might not want to disclose its true utility from an agreement for privacy reasons (e.g., not to hamper its prospects in future negotiations) and an incentive-compatible mechanism would allow the other party to learn the utility of a party X from the mechanism-induced truthful claim $v_X = u_X$.

Restrictiveness. Second, incentive compatibility is in general unnecessary to achieve an optimal Nash bargaining product: For a viable agreement (i.e., $u_X + u_Y \geq 0$), the Nash bargaining product is optimized for all v_X, v_Y with $v_X - u_X = v_Y - u_Y$ (i.e., equal dishonesty) and $v_X + v_Y \geq 0$. Hence, while truthfulness, i.e., $v_X - u_X = 0 = v_Y - u_Y$, is a sufficient condition for an optimal Nash bargaining product, it is not a necessary condition.

Inefficiency. Third, while incentive compatibility can be guaranteed with mechanisms, this guarantee often comes at the cost of introducing inefficiency, in the sense

of negotiation-failure probability induced by the mechanism itself. For example, the randomized-arbitration mechanism by Myerson [189] introduces a relatively high probability of negotiation cancellation such that even agreements with large surplus often cannot be concluded. Counter-intuitively, mechanisms which allow small deviations from truthfulness might thus be more effective at facilitating agreements than perfectly incentive-compatible mechanisms.

In the following subsection, we present a mechanism that sacrifices incentive compatibility in order to increase the probability of negotiation conclusion, while preserving individual rationality and budget balance.

7.5.3 BOSCO: Bargaining in One Shot with Choice Optimization

In this subsection, we present the BOSCO mechanism, which we have designed to enable automated negotiation of inter-AS agreements with high bargaining efficiency. The core idea of the BOSCO mechanism is as follows: The negotiating parties play a simple bargaining game supervised by a BOSCO service, in which each party has a *set of choices* defined by the mechanism. Each combination of such choice sets is associated with at least one Nash equilibrium, i.e., a combination of strategies (here: utility-choice mappings) in which no party can profitably deviate from the strategy assigned to it. In turn, each such Nash equilibrium can be rated with respect to a bargaining-efficiency metric.

The benefit of the mechanism is thus realized by the BOSCO service, which appropriately constructs the choice sets and picks an associated Nash equilibrium such that a high bargaining-efficiency results. In the following, we will present and formalize the components of the mechanism.

7.5.3.1 Utility Distributions

For executing the BOSCO mechanism, the two agreement parties X and Y communicate the content of a mutuality-based agreement to an BOSCO service. While the BOSCO service does not know the true utility u_X and u_Y that either party derives from the agreement, we assume (as usual in bargaining-mechanism design) that the BOSCO service can estimate a *utility distribution* $\mathbb{U}_Z(u)$, which yields the probability that party $Z \in \{X, Y\}$ derives utility u from the agreement. We envision that such an estimation can be performed on the basis of heuristics, taking standard transit and network-equipment prices into account.

7.5.3.2 Choice Sets

After deriving $\mathbb{U}_Z(u)$ for each agreement party Z , the BOSCO mechanism constructs a *choice set* V_Z of possible claims for each agreement party Z . For BOSCO, these choice sets correspond to finite discrete sets with cardinality $W_Z = |V_Z|$. To guarantee strong individual rationality, each choice set always contains the option $-\infty$, with which any party can cancel the negotiation. Moreover, let there be an ordering $v_{Z,1}, \dots, v_{Z,W_Z}$ on the choices such that $v_{Z,i} < v_{Z,j}$ for all $1 \leq i < j \leq W_Z$.

7.5.3.3 Bargaining Game

The utility distributions and the choice sets represent the basis of a *bargaining game*. In this bargaining game, each party Z picks a suitable choice $v_Z \in V_Z$ and commits it to the BOSCO service. The BOSCO service then checks whether the apparent utility surplus is non-negative, i.e., $v_X + v_Y \geq 0$. If yes, the mechanism prescribes the conclusion of the agreement with cash compensation $\Pi_{X \rightarrow Y} = (v_X - v_Y)/2$, resulting in after-negotiation utility $\bar{u}_X = u_X - \Pi_{X \rightarrow Y}$ and $\bar{u}_Y = u_Y + \Pi_{X \rightarrow Y}$. If not, the mechanism cancels the negotiation, resulting in the after-negotiation utility $\bar{u}_X = \bar{u}_Y = 0$.

7.5.3.4 Bargaining Strategies

In the bargaining game, the *bargaining strategy* $\sigma_Z(u_Z)$ of party Z is a function which yields a choice $v_Z \in V_Z$, given the true utility u_Z of party Z . The bargaining strategy σ_Y of a counter-party Y is especially relevant when party X evaluates the attractiveness of a choice v_X based on the *expected after-negotiation utility*:

$$\mathbb{E}[\bar{u}_X](u_X, v_X) = \sum_{\substack{v_Y \in V_Y \\ v_Y \geq -v_X}} \mathbb{P}[v_Y] \cdot \left(u_X - \frac{v_X - v_Y}{2} \right), \quad (7.22)$$

where

$$\mathbb{P}[v_Y] = \int_{-\infty}^{\infty} \mathbb{U}_Y(u_Y) \cdot [\sigma_Y(u) = v_Y] \, du_Y, \quad (7.23)$$

and $[\![P]\!]$ is 1 if statement P is true and 0 otherwise.

This expected after-negotiation utility $\mathbb{E}[\bar{u}_X](u_X, v_X)$ can also be represented in the following form:

$$\mathbb{E}[\bar{u}_X](u_X, v_X) = m_X(v_X)u_X + q_X(v_X), \quad (7.24)$$

where

$$m_X(v_X) = \mathbb{P}[\sigma_Y(u_Y) \geq -v_X], \quad (7.25)$$

$$q_X(v_X) = \sum_{v_Y \geq -v_X} \mathbb{P}[v_Y] \cdot \frac{v_Y - v_X}{2} \quad (7.26)$$

are the *choice slope* and the *choice intercept*, respectively. Intuitively, $m_X(v_X)$ denotes the probability of agreement conclusion given claim v_X , and $q_X(v_X)$ denotes the expected monetary transfer that party X obtains from party Y given claim v_X . These values are henceforth abbreviated as $m_{X,i} = m_X(v_{X,i})$ and $q_{X,i} = q_X(v_{X,i})$ based on our choice-set enumeration. Notably, it holds that claim slopes grow together with the claims, i.e., $m_X(v_{X,i}) \geq m_X(v_{X,j})$ for all $i > j$ by nature of the CDF in Eq. (7.25).

This formulation allows a natural definition of the *best-response strategy* σ_Z^+ :

Definition 7.1 Best-Response Strategy. In the negotiation between two parties X and Y, the best-response strategy $\sigma_X^+(u_X)$ of party X picks the choice $v_X \in V_X$ with the highest expected after-negotiation utility given its true utility u_X :

$$\sigma_X^+(u_X) = \arg \max_{v_X \in V_X} \mathbb{E}[\bar{u}_X](u_X, v_X). \quad (7.27)$$

In the following, we provide insight into the functional structure of σ_X^+ . In particular, we will identify the true-utility values u_X for which each choice $v_{X,i}$ is the best choice.

Fundamentally, $v_{X,i}$ is the best choice for true utility u_X if and only if $v_{X,i}$ has higher expected utility than all other choices $v_{X,j}$, $j \neq i$, i.e.,

$$\forall j \neq i. \quad m_{X,i}u_X + q_{X,i} \geq m_{X,j}u_X + q_{X,j}. \quad (7.28)$$

Notably, a choice $v_{X,i}$ may not be optimal for any u_X : If another choice $v_{X,j}$ has identical slope $m_{X,j} = m_{X,i}$ and higher intercept $q_{X,j} > q_{X,i}$, the choice $v_{X,i}$ is never the best choice, and does not need to be considered further. Furthermore, if another choice $v_{X,j}$ has identical slope $m_{X,j} = m_{X,i}$ and identical intercept $q_{X,j} = q_{X,i}$, the two choices are completely equivalent, and can be considered a single choice.

For all other cases, we can thus assume distinct slopes $m_{X,i}$ and $m_{X,j}$. This assumption allows the following transformation of the pairwise inequality in Eq. (7.28):

$$u_X \begin{cases} \geq \tilde{u}_X(i, j) & \text{if } m_{X,i} > m_{X,j}, \\ \leq \tilde{u}_X(i, j) & \text{if } m_{X,i} < m_{X,j}, \end{cases} \quad \text{where} \quad \tilde{u}_X(i, j) = \frac{q_{X,j} - q_{X,i}}{m_{X,i} - m_{X,j}}. \quad (7.29)$$

In this formulation, we first consider the choice $v_{X,1}$ with the minimal $m_{X,1}$. This minimal choice $v_{X,1}$ is optimal for true utility u_X ranging from $-\infty$ to the lowest value $\tilde{u}_X(1, j)$ for any j . More formally, the upper end of the optimality range for minimal choice $v_{X,i}$ is at $\tilde{u}_X^+(1)$, where

$$\tilde{u}_X^+(i) = \tilde{u}_X(i, j^+(i)) \quad \text{where} \quad j^+(i) = \arg \min_{j: m_{X,j} > m_{X,i}} \tilde{u}_X(i, j). \quad (7.30)$$

In turn, the choice $v_{X,j^+(1)}$ is optimal for true utility u_X ranging from $\tilde{u}_X^+(1)$ to $\tilde{u}_X^+(j^+(1))$, where choice $v_{X,j^+(j^+(1))}$ becomes optimal. This process can be continued until the choice $v_{X,k}$ with maximal slope $m_{X,k}$, yielding a finite recursive series J :

$$J_1 = 1 \quad \text{and} \quad \forall n > 1. \quad J_n = j^+(J_{n-1}). \quad (7.31)$$

Hence, the best-response strategy σ_X^+ can be represented as follows:

$$\sigma_X^+(u_X) = \begin{cases} v_{X,1} & \text{if } u_X \leq \tilde{u}_X^+(1), \\ v_{X,J_n} & \text{if } u_X \in [\tilde{u}_X^+(J_{n-1}), \tilde{u}_X^+(J_n)), \\ v_{X,k} & \text{otherwise.} \end{cases} \quad (7.32)$$

7.5.3.5 Nash Equilibria

A *Nash equilibrium* $\sigma^* = (\sigma_X^*, \sigma_Y^*)$ in the bargaining game is a set of two bargaining strategies, each of which is the best-response strategy to the other strategy. We compute such an equilibrium by assuming arbitrary σ_X and σ_Y , and then computing the best-response strategies in an alternating fashion until the best-response strategy of any party is their existing strategy. While it can be shown that the considered bargaining game is not a potential game [185] (which would guarantee convergence to an equilibrium by alternating unilateral optimization), the best-response dynamics always converged to an equilibrium in our diverse simulations.

7.5.3.6 Bargaining Efficiency

Given a Nash equilibrium, a natural question arises concerning the efficiency of such an equilibrium σ^* . Clearly, if the BOSCO service knew u_X and u_Y , it could simply compute the associated Nash bargaining product $\mathcal{N}(u_X, u_Y, \sigma_X^*(u_X), \sigma_Y^*(u_Y))$ and compare it to the optimal Nash bargaining product $\mathcal{N}(u_X, u_Y, u_X, u_Y)$ that arises under universal truthfulness. However, as the BOSCO service has only probabilistic

knowledge about the true utility of the agreement, it must evaluate the efficiency of an equilibrium σ^* by computing the *expected Nash bargaining product* $\mathbb{E}[\mathcal{N}|\sigma^*]$ for this strategy, which is

$$\mathbb{E}[\mathcal{N}|\sigma^*] = \iint_{-\infty}^{\infty} \mathbb{U}(u_X, u_Y) \mathcal{N}(u_X, u_Y, \sigma_X^*(u_X), \sigma_Y^*(u_Y)) du_Y du_X \quad (7.33)$$

where \mathbb{U} is the joint utility distribution for parties X and Y. The optimal expected Nash bargaining product is given by $\mathbb{E}[\mathcal{N}|\sigma^\top]$ where $\sigma_Z^\top(u_Z) = u_Z$ is the truthful strategy for party Z. Similar to a Price of Anarchy formulation [148], we thus formalize the efficiency of an equilibrium with the following metric of *Price of Dishonesty* (*PoD*):

$$PoD(\sigma^*) = 1 - \frac{\mathbb{E}[\mathcal{N}|\sigma^*]}{\mathbb{E}[\mathcal{N}|\sigma^\top]}, \quad (7.34)$$

which is always in $[0, 1]$ for reasons laid out in §7.5.4. Note that *PoD* is undefined if $\mathbb{E}[\mathcal{N}|\sigma^\top] = 0$, i.e., if the agreement would be consistently non-viable even under honesty, which is an uninteresting case that we henceforth disregard.

In summary, the BOSCO service is thus tasked with estimating \mathbb{U}_X and \mathbb{U}_Y and constructing choice sets V_X and V_Y such that the thereby defined bargaining game has an equilibrium σ^* with a low *PoD*. After the BOSCO service finds such a configuration, it communicates the mechanism-information set $(\mathbb{U}_X, \mathbb{U}_Y, V_X, V_Y, \sigma^*)$, to the communicating parties, which can verify that σ^* is indeed a Nash equilibrium. As a result, the parties then follow the strategy that is assigned to them in the equilibrium. Hence, each party Z plays the bargaining game by applying the equilibrium strategy σ_Z^* to its true utility value u_Z and commits the resulting claim to the BOSCO service, which then decides on agreement conclusion and, in case of negotiation success, on the exchanged cash compensation $\Pi_{X \rightarrow Y}$.

7.5.4 BOSCO Properties

After describing the BOSCO mechanism in the previous section, we now analyze the mechanism with respect to the properties listed in §7.5.2. First of all, it is clear that the BOSCO mechanism is budget-balanced, as the cash transfer paid by one party is exactly the cash transfer received by the other party. Other important properties are listed and justified with proofs below.

Theorem 7.1 Strong Individual Rationality. *The BOSCO mechanism offers strong individual rationality:*

$$\forall u_X, u_Y. \quad \bar{u}_X \geq 0 \text{ and } \bar{u}_Y \geq 0. \quad (7.35)$$

PROOF. Given utility u_X of party X, the Nash-equilibrium strategy choice $\sigma_X^*(u_X)$ is the best choice for party X. If u_Y is such that $\sigma_Y^*(u_Y) < -\sigma_X^*(u_X)$, then the agreement is not concluded and $\bar{u}_X = \bar{u}_Y = 0$. Conversely, if $\sigma_Y^*(u_Y) \geq -\sigma_X^*(u_X)$, the agreement is concluded and $\sigma_Y^*(u_Y)$ enters into $\mathbb{E}[\bar{u}_X](u_X, \sigma_X^*(u_X))$ from Eq. (7.22) with non-zero weight.

Now assume for contradiction that $\bar{u}_X = u_X - (v_X - \sigma_Y^*(u_Y))/2 < 0$. In that case, $\bar{u}_X(v_Y) = u_X - (v_X - v_Y)/2$ is also negative for any lower $v_Y \in V_Y$ with $v_Y \leq \sigma_Y^*(u_Y)$. To avoid that the negative $\bar{u}_X(v_Y)$ for these lower v_Y enter into $\mathbb{E}[\bar{u}_X]$ with non-zero weight, party X could choose a lower $v'_X < v_X \in V_X$ with $-v'_X > \sigma_Y^*(u_Y) \geq -v_X$. With such a choice, all the mentioned v_Y are inactivated in the expected-utility term. This inactivation increases expected utility as the summands associated with these inactivated v_Y only contribute negative values and the summands associated with the remaining choices of party Y increase. Thanks to $-\infty \in V_X$, such a choice is always possible.

However, this non-optimality of $\sigma_X^*(u_X)$ contradicts the best-response character of σ_X^* , which invalidates the assumption of a negative \bar{u}_X . Hence, if an agreement is concluded, $\bar{u}_X \geq 0$ for any u_X and u_Y (The case for \bar{u}_Y is analogous).

In summary, non-negativity of after-negotiation utility holds in any case (non-conclusion and conclusion), establishing strong individual rationality. \square

Theorem 7.2 Soundness. *The BOSCO mechanism is sound, i.e., it never leads to conclusion of a non-viable agreement:^a*

$$\forall u_X, u_Y. \quad \sigma_X^*(u_X) + \sigma_Y^*(u_Y) \geq 0 \implies u_X + u_Y \geq 0 \quad (7.36)$$

^a In order to be ex-post efficient, the mechanism would additionally need to be *complete* in the sense that all viable agreements are concluded. However, as noted in §7.5.2, ex-post efficiency is not achievable given other desirable properties.

PROOF. If $\sigma_X^*(u_X) + \sigma_Y^*(u_Y) \geq 0$, the agreement is concluded and $\bar{u}_X = u_X - \Pi_{X \rightarrow Y}$ and $\bar{u}_Y = u_Y + \Pi_{X \rightarrow Y}$. By strong rationality (Theorem 7.1), it holds that $\bar{u}_X \geq 0$ and $\bar{u}_Y \geq 0$, which implies $u_X \geq \Pi_{X \rightarrow Y}$ and $u_Y \geq -\Pi_{X \rightarrow Y}$. By addition of the

inequalities, $u_X + u_Y \geq \Pi_{X \rightarrow Y} - \Pi_{Y \rightarrow X} = 0$. Hence, $\sigma_X^*(u_X) + \sigma_Y^*(u_Y) \geq 0 \implies u_X + u_Y \geq 0$. \square

Theorem 7.3 Privacy Preservation. *The BOSCO mechanism is privacy-preserving, i.e., an exact reconstruction of the true utility of a party from its choice is impossible:*

$$\forall v_X \in V_X. \quad |\{u_X | \sigma_X^*(u_X) = v_X\}| \neq 1 \quad (7.37)$$

PROOF. In terms of privacy, the only problematic case would arise if the range $[\tilde{u}_X^+(J_n), \tilde{u}_X^+(J_{n+1}))$ contained only one value, which would allow the derivation of that value u_X from the associated choice $v_{X,i}$. However, since half-open intervals on the real numbers cannot contain a single value, this problematic case cannot arise. \square

While exact reconstruction of the true utility is thus impossible, it might still be possible to predict the true utility with high precision if the interval associated with a choice is very small. Hence, the degree to which an equilibrium preserves privacy could be quantified (e.g., by the length of the shortest non-empty interval) and then taken into account by the BOSCO service when picking an equilibrium.

7.5.5 Choice-Set Construction

It remains to analyze how the choice sets should be constructed such that equilibria with a low Price of Dishonesty result. Surprisingly, we have found that random generation of such choice sets works reasonably well in practice. In particular, the choice set V_X for any party X can be constructed by sampling a high enough number of choices v_X from the utility distribution U_X . With multiple trials of such random choice-set generation, choice sets with a relatively low Price of Dishonesty can be found.

In Fig. 7.2, for example, we analyze the resulting *PoD* from random choice-set generation for two uniform utility distributions, namely $U^{(1)}$, which is a uniform distribution of (u_X, u_Y) on $[-1, 1] \times [-1, 1]$, and $U^{(2)}$, which is a uniform distribution on $[-\frac{1}{2}, 1] \times [-\frac{1}{2}, 1]$. For each choice-set cardinality $W_X = W_Y$, we generate 200 random choice-set combinations and find the mean and the minimum of the associated *PoD* values. Clearly, a higher number of choices generally helps to reduce the Price of Dishonesty, but given around 50 choices, adding more choices does not improve the mechanism efficiency.

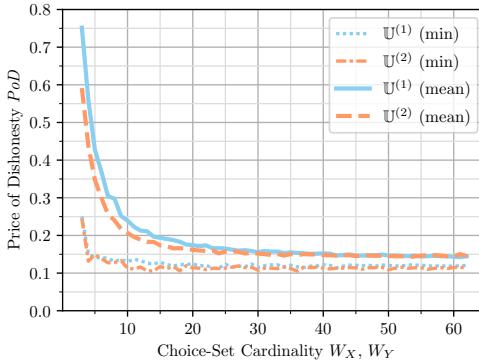


Figure 7.2: Price of Dishonesty (minimum and mean) guaranteed by BOSCO depending on the number of choices $W_X = W_Y$ for two different utility distributions $\mathbb{U}^{(1)}$ and $\mathbb{U}^{(2)}$.

Interestingly, we also observe that the number of *equilibrium choices* (i.e., choices which have a non-empty associated interval in the equilibrium strategy) for each party reaches 4 at that point and is not further increased for more possible choices.

Hence, the BOSCO service can increase the number of possible choices until the resulting *PoD* values do not substantially decrease anymore. With this procedure, the BOSCO mechanism could guarantee a Price of Dishonesty of around 10% for both $\mathbb{U}^{(1)}$ and $\mathbb{U}^{(2)}$ in the example at hand, meaning that the negotiation can be expected to be 10% less efficient than under the unrealistic assumption of perfect honesty.

7.6 EFFECT ON PATH DIVERSITY

In this section, we attempt to quantify the effect of mutuality-based agreements on path diversity in the Internet.

7.6.1 Method

Agreement prediction. Starting from the CAIDA AS-relationship dataset [54], we construct a network of ASes where a provider–customer or peering relationship results in a single provider–customer or peering link, respectively. In this graph, we generate all possible mutuality-based agreements (MAs) for the whole topology: For every pair (A, B) of peers, we generate an MA in which A gives B access to

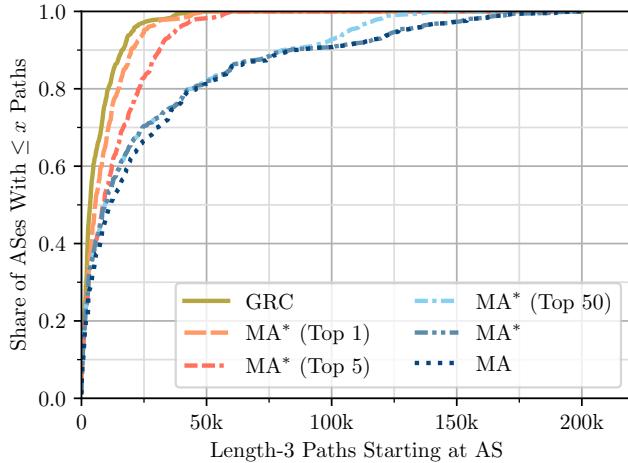


Figure 7.3: Distribution of ASes with respect to the number of length-3 paths starting at the AS, given different degrees of MA conclusion.

all its providers and peers which are not customers of B , and vice versa. As MAs consist of an AS A giving its peer B access to a provider or another peer, these agreements enlarge the set of paths with 3 AS hops and 2 inter-AS links (henceforth: length-3 paths) for B , as well as the set of ASes that B can reach with such length-3 paths (henceforth: nearby destinations).

Impact analysis. Given this graph and these MAs, we perform the following analysis for 500 randomly chosen ASes. First, we find the GRC-conforming length-3 paths starting at the given AS. Then, we find the MA-created length-3 paths for which the given AS is an end-point. The number of these additional paths and the number of additional nearby destinations are thus metrics for the increased path diversity that the given AS enjoys thanks to MAs.

7.6.2 Number of Paths and Nearby Destinations

Paths: Complete conclusion. Figure 7.3 shows the substantial increase in the number of length-3 paths that are potentially available to these ASes thanks to mutuality-based agreements: For example, whereas none of the analyzed ASes have more than 45,000 GRC-conforming paths with length 3, 20% of the analyzed ASes have more than 45,000 length-3 paths if all MAs are concluded (CDF for ‘MA’). Note

that for a fixed source and a fixed destination, all length-3 paths are disjoint by definition.

Paths: Partial conclusion. Since the conclusion of all possible MAs is an extreme case (although MAs could be negotiated in an automated fashion with the mechanism presented in §7.5), we analyze the effects of partial agreement conclusion.

Initially, we note that an MA can provide an AS with new paths in two manners. First, an AS can *directly* gain an MA path by concluding an MA that includes the path (e.g., as AS D gains the path DEB in Fig. 7.1 from the MA with AS E). Second, an AS can *indirectly* gain an MA path by being the subject of an MA that includes the path (e.g., as AS B or AS F gain paths to AS D from the MA between AS D and AS E in Fig. 7.1). Interestingly, most additional MA paths are directly gained paths, as the similarity of the CDFs for all MA paths (MA) and directly gained MA paths (MA^*) in Fig. 7.3 suggests. Hence, the ASes bearing the negotiation effort of an MA have a strong incentive to negotiate that MA despite the effort, because they typically are its biggest beneficiaries. Moreover, we find that an AS already tends to obtain substantial gains in path diversity with only a handful of MAs. This point is demonstrated by the results for the scenarios where an AS only concludes the n MAs which provide it with the most new paths (annotated with ‘ MA^* (Top n)’ in Fig. 7.3): Even if an AS only concludes the single most attractive agreement from its perspective, it stands to gain several thousands of new paths.

Destinations. Figure 7.4 also illustrates that mutuality-based agreements enlarge the set of destinations reachable with paths of length 3: For example, whereas 40% of the analyzed ASes can reach more than 5,000 destinations over length-3 paths, 57% of ASes can reach more than 5,000 destinations over such paths if all MAs are concluded. Interestingly, very few MAs per AS already suffice to reap most of these benefits, as the results for non-comprehensive agreement conclusion demonstrate.

Distribution of benefits. For the set of analyzed ASes, the average number of additional length-3 paths thanks to mutuality-based agreements is 22,891 paths (maximum: 196,796 paths), and the average number of additionally reachable destinations over length-3 paths is 2,181 ASes (maximum: 7,144 ASes). Interestingly, the gains in terms of additionally reachable destinations are more broadly distributed than the gains regarding paths. The explanation for this phenomenon is that mutuality-based agreements in very densely connected regions of the Internet lead to a high number of additional length-3 paths, but have little impact on the number of ASes reachable over such paths.

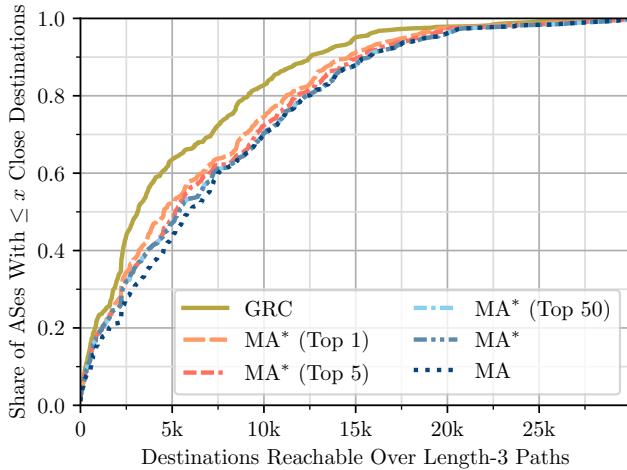


Figure 7.4: Distribution of ASes with respect to the number of destination reachable over length-3 paths, given different degrees of MA conclusion.

7.6.3 Geodistance

In order to gain a more qualitative perspective on the additional paths enabled by MAs, we also investigate the geographical length (henceforth: geodistance) of these new paths. Such geodistance is an important determinant of path latency [238], which is typically considered a core aspect of path diversity.

Geodistance calculation. As the CAIDA AS-relationship dataset [54] does not directly contain the necessary information, we additionally build on the CAIDA prefix-to-AS dataset [55], the GeoLite2 database [173], and the CAIDA geographic AS-relationship dataset [53]. In particular, we determine the geolocation of any AS by finding the IP prefixes associated with the AS number in the prefix-to-AS dataset, determining the geolocation of the IP prefixes via the GeoLite2 database, and averaging the resulting coordinates to obtain the center of gravity of the AS. With such averaging, the potentially considerable intra-AS latency of geographically distributed top-tier ASes is automatically taken into account. Moreover, we obtain the geolocation of an AS interconnection from the CAIDA geographic AS-relationship dataset. The geodistance of a length-3 path $\pi = (A_1, \ell_{12}, A_2, \ell_{23}, A_3)$, where A_i are ASes and ℓ_j are inter-AS links, is then computed as $d(\pi) = d(A_1, \ell_{12}) + d(\ell_{12}, \ell_{23}) + d(\ell_{23}, A_3)$, where $d(X, Y)$ is the geodistance between two points. If there are multi-

ple known AS interconnections, the geodistance of the AS-level path (A_1, A_2, A_3) is computed for ℓ_{12} and ℓ_{23} that minimize $d(\pi)$.

Impact analysis. Using this measure of path geodistance, we again compare the set of paths that conform to the GRC and the set of paths enabled by novel MAs. For every analyzed AS pair connected by at least one length-3 GRC path, we determine the maximum, median, and minimum geodistance given the length-3 GRC paths connecting the AS pair. In a next step, we determine the geodistance of the additional MA paths and check for each MA path whether it is lower than the maximum, median, or minimum GRC geodistance, respectively. Each AS pair is then characterized by the number of MA paths below these comparison thresholds. The aggregate results of this comparison method are presented in Fig. 7.5a.

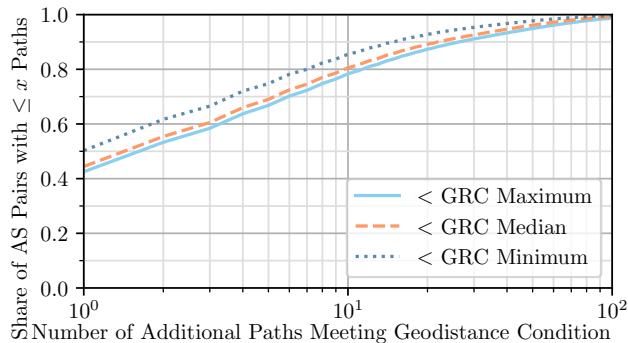
Results. Figure 7.5a shows that through MAs, around 50% of AS pairs gain at least 1 path with a lower geodistance than the minimum-geodistance GRC path, suggesting that inter-AS latency can be decreased by MAs in these cases. Around 25% of AS pairs even gain at least 5 paths that improve upon the minimum GRC geodistance, and at least 7 and 8 paths that improve upon the median and maximum GRC geodistance, respectively. Another 20% of AS pairs only gain MA paths with a higher geodistance than the maximum GRC geodistance (or no new paths at all); however, also these additional paths have value in terms of reliability. Regarding the AS pairs experiencing a reduction in minimum geodistance, Fig. 7.5b illustrates the considerable extent of the geodistance reduction for these AS pairs: For example, 50% of AS pairs that experience a geodistance reduction obtain a reduction of more than 24%.

7.6.4 Bandwidth

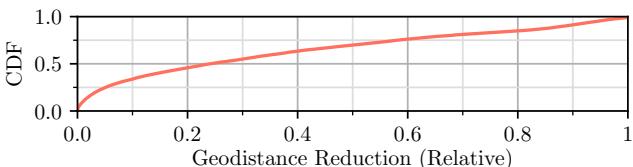
We perform an analogous analysis as in the preceding section with respect to the bandwidth of additional paths.

Bandwidth estimation. To infer the bandwidth of inter-AS links, we employ a *degree-gravity model* [225] which endows each link with a capacity value proportional to the product of the node degrees of the link end-points. The path bandwidth is then the minimum such computed link bandwidth of all links in the path.

Results. With such an analysis, we find that 35% of all investigated AS pairs obtain a new MA path that has more bandwidth than the corresponding maximum-



(a) Distribution of AS pairs with respect to geographical length of additional paths from mutuality-based agreements, distinguished by satisfied geodistance-comparison condition.



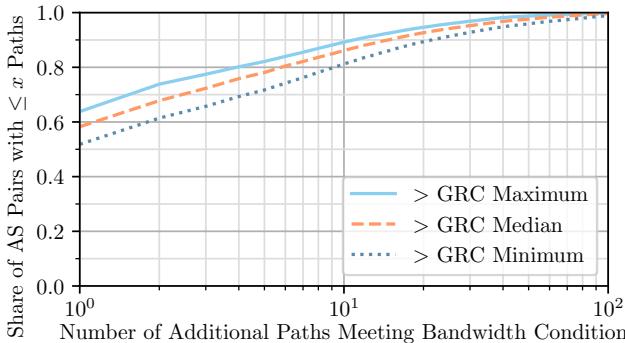
(b) Distribution of relevant AS pairs (i.e., AS pairs which experience a geodistance reduction thanks to MAs) with respect to the relative geodistance reduction from mutuality-based agreements.

Figure 7.5: Results of geodistance analysis.

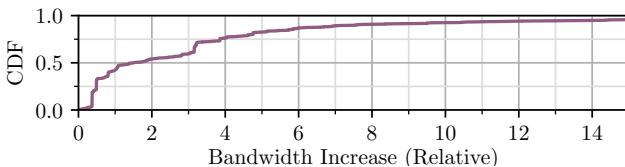
bandwidth GRC path (cf. Fig. 7.6a). Of these benefiting AS pairs, 50% gain an MA path with at least 150% more bandwidth than the respective maximum-bandwidth GRC path (cf. Fig. 7.6b).

7.7 RELATED WORK

Traditional BGP interconnection agreements. After the growth of the Internet had led to considerable BGP stability problems in the late 1990s [102, 153], the research of AS interconnection agreements, their stability properties, and their optimal structures received significant interest. The commercial reality of the Internet has been shown to mainly contain two basic types of agreements that determine route-forwarding policies, namely provider–customer agreements and peering agreements [125, 198]. The relative exclusiveness of these two agreement types was reinforced by the



(a) Distribution of AS pairs with respect to bandwidth of additional paths from mutuality-based agreements, distinguished by satisfied bandwidth-comparison condition.



(b) Distribution of relevant AS pairs (i.e., AS pairs which experience a bandwidth increase thanks to MAs) with respect to the relative bandwidth increase from mutuality-based agreements.

Figure 7.6: Results of bandwidth analysis.

important result of Gao and Rexford, showing that BGP route convergence is guaranteed if ASes stick to these two forms of route-forwarding policies [91].

Non-standard BGP interconnection agreements. However, it is well-known that such strict BGP policies reduce path quality: For a majority of routes selected in BGP, there exists a route that is more attractive with respect to metrics such as bandwidth, latency, or loss rate [94, 96, 109, 147, 211, 228]. Motivated by such improvements, many ASes do not always follow the Gao–Rexford conditions [94]. For example, some ISPs use alternative paths to reach content distribution networks such as Akamai [24, 95, 174], other ISPs prefer the peer route through their Tier-1 neighbor over a longer customer route [24], and so on. Still, these deviations are narrow in scope, with most non-GRC policies being explainable by “sibling” ASes, which belong to the same organization and provide mutual transit services [24, 90], and partial/hybrid provider relationships [95]. This is due to the fact that more complex

policies would threaten the convergence of the routing process unless they are supported through multi-AS coordination efforts. These restrictions have also been acknowledged by previous efforts to provide multipath routing based on BGP such as MIRO [261].

Non-BGP interconnection agreements. Interconnection agreements in PAN architectures [22, 33, 60, 98, 212, 213, 245, 265, 269] do not need to follow the guidelines devised to achieve BGP stability. In that respect, PANs confirm an insight by Haxell and Wilfong [114], who showed that a fractional relaxation of the stable-paths problem of BGP [104] guarantees a solution and that a more flexible routing paradigm can thus defuse BGP stability issues. In contrast to this work, however, our focus lies on the structure of interconnection agreements under this new paradigm, the extent to which these agreements increase path diversity, and their economic rationality and bargaining aspects.

Economic aspects of interconnection agreements. Interconnection agreements have to respect the economic self-interest of ASes. The Gao–Rexford guidelines for BGP policies have been proven to be rational in that sense [83]. Notable proposals for agreement structures that attempt to satisfy both AS self-interest and global efficiency include Nash peering [74, 267], where the cooperative surplus of the agreement is shared among the parties according to the Nash bargaining solution [194], and ISP-settlement mechanisms based on the Shapley value [169]. It is important to note that unlike traditional source routing and similar to MIRO [261], PANs still offer transit ASes control over the traffic traversing their network, and hence to maximize their revenue. In contrast to MIRO, however, PANs guarantee path stability.

7.8 CONCLUSION

This chapter shows that PAN architectures enable novel types of interconnection agreements, thereby substantially improving path diversity in the Internet and creating new business opportunities. Such new possibilities exist in PAN architectures as they do not rely on the nowadays essential route-forwarding policy guidelines formulated by Gao and Rexford [91] for route convergence.

Our results show that path diversity in the Internet benefits enormously by enabling paths beyond the Gao–Rexford constraints: By using previously impossible path types, the median AS can reach thousands of new destinations with 3-hop paths and benefit from tens of thousands of additional paths, some of which have

more desirable characteristics than the previously available paths. There is thus a largely unknown advantage to PAN architectures: Not only do these architectures enable end-hosts to *select* a forwarding path, they also allow network operators to *offer* new (and often shorter) forwarding paths. As PANs are not limited to using a single path between a pair of ASes, all these paths can be used simultaneously according to the requirements of end-hosts and their applications (e.g., low latency for voice over IP and high bandwidth for file transfers). These direct benefits to end-hosts in turn incentivize providers to explore new interconnection agreements and offer diverse paths to attract new customers.

To enable such new interconnection agreements, we present two methods for designing agreements that are Pareto-optimal, fair, and thus attractive to both parties. We also show that, assisted by an appropriate bargaining mechanism, the negotiation of such agreements can lead to efficient agreements, even though necessary information is private.

8

ISP COMPETITION: QUALITY ATTRIBUTES

8.1 INTRODUCTION

Whereas the preceding chapter investigated opportunities for ISP *cooperation* in terms of interconnection agreements, the chapter at hand focuses on ISP *competition*, which arises from path-selection decisions: Both today's Border Gateway Protocol (BGP) and emerging path-aware network architectures (PANs) supply ISPs with potentially multiple paths towards an IP prefix. When selecting among these paths, ISPs and end-hosts decide on the basis of price and quality of the available paths. This path quality is determined by multiple *quality attributes* of potentially multiple on-path ISPs. Such quality attributes may include conventional performance metrics (e.g., bandwidth, latency, loss rate, jitter) or security features (e.g., presence of security middleboxes), but also properties that traditionally receive less attention, e.g., environmental, social, and governmental (ESG) properties such as carbon emission from data transmission [274] or geopolitical concerns regarding on-path ISPs [71]. Transit ISPs invest in their attributes, communicate them in path announcements, and thereby attract traffic from selecting entities (ISPs and end-hosts). However, improving these attributes comes at a cost, which may exceed the revenue from attracted traffic, especially if ISPs on competing paths also raise their quality level.

Given this competitive setting, ISPs today face complex strategic questions when optimizing profit: What quality attributes should be invested in, and to what extent? How should prices be determined? And how are these decisions affected by ISPs on competing paths *and* ISPs elsewhere on the provided paths?

Well-informed strategic decisions thus require a fundamental understanding of ISP competition under path selection, not only on an intuitive, but also on a rigorous analytical level. While such an understanding has been furthered by previous academic research [170, 192, 234, 244, 270], many open questions of practical relevance remain, e.g., regarding the multi-attribute nature of path quality,

the dependence of fixed and variable ISP cost on provided quality, the feasibility of cooperation among ISPs on the same path, and the impact of differing degrees of competition intensity (cf. §8.6).

To address these questions, we present a new game-theoretic model, enabling a rigorous investigation of quality competition among ISPs. We perform this investigation through theoretical analysis and simulation:

Theoretical analysis. We conduct an extensive theoretical analysis to systematically understand the effect of ISP competition on path quality, and ISP profits (Path price constitutes a quality attribute in our model). In particular, we identify closed-form solutions for the Nash equilibria of the competition dynamics, prove the stability of these equilibria, and contrast them for varying degrees of path diversity and ISP heterogeneity. On the one hand, this theoretical analysis confirms intuitive insights, namely that competition tends to raise the prevalence of valuable attributes. On the other hand, our model reveals counter-intuitive insights, namely that the cooperation between ISPs on the same path suffers from a prisoner's dilemma, that ISP profits can increase under intensified competition, and that additional paths may decrease the prevalence of quality attributes if unchangeable path attributes are starkly different.

Simulation-based case study. To determine which competition effects are significant in practice, we leverage our model for a simulation-based case study. In this case study, we investigate the competition dynamics in the Internet core with respect to two attributes (internal bandwidth and the share of clean energy used by an ISP). This simulation requires a numerical instantiation of the model, based on real-world data. For this model instance, our simulations yield robust evidence that competition raises the prevalence of valuable attributes, the quality of available paths, and the profits of most ISPs.

8.1.1 Contributions

In summary, this chapter includes the following contributions:

Game-theoretic competition model. Our new ISP-competition model (§8.2) departs from previous competition models by representing both inter-path competition and intra-path cooperation, accommodating a multi-faceted notion of path quality, revealing the effect of path diversity, and reflecting realistic ISP cost structures (§8.6).

Theoretical analysis. We conduct a rigorous theoretical analysis by reasoning from basic competition scenarios that showcase the fundamental effects in ISP competition (§8.3). In particular, we contrast monopolistic and competitive scenarios in ISP path selection, investigate networks with varying similarity in ISP profit functions, and identify asymptotically stable equilibria and social optima of the competition. Our analysis suggests that ISP competition has nuanced effects on ISP profits and path quality, going beyond the predictions of basic economic theory.

Large-scale simulation. We demonstrate how to instantiate our model based on real-world data, with the goal of predicting competition effects in the Internet core (§8.4). These predictions are generated with simulations, which rely on randomization to achieve robust results, represent the competition behavior with better-response dynamics, and are executed for varying path diversity. The simulation results suggest that competition, induced by path diversity, has positive effects for a majority of ISPs on multiple tiers of the Internet, i.e., raises ISP profits and path quality (§8.5).

8.2 MODEL AND FIRST INSIGHTS

In the following, we present a game-theoretic model, which we employ to investigate the competition dynamics under attribute-oriented path selection. While our model reflects common characterizations of inter-domain network economics, it is more general than previous models (cf. §8.6).

Network and paths. We abstract the network as a set N of ISPs, which represent the players in the competition game. Each ISP $n \in N$ is assumed to be fully rational. The ISPs form paths, where each path $r \subseteq N$ is a set of ISPs. All usable paths in a network are collected in the path set R , and all usable paths between selecting ISP n_1 and destination ISP n_2 constitute the set $R(n_1, n_2)$. Throughout this work, we study how ISPs affect the quality of paths as given by path set R , not how ISPs strategically adapt the set R of usable paths via interconnection agreements and announcements, which is a related but distinct problem [176, 3].

Attributes. We consider a network with a set K of ISP attributes, $|K| \geq 1$, that are relevant in path selection. Hence, each ISP n is associated with an attribute vector $\mathbf{a}_n \in \mathbb{R}_{\geq 0}^K$, where $a_{nk} \in \mathbb{R}_{\geq 0}$ denotes the prevalence of attribute $k \in K$ in ISP n . As a player in the competition game, each ISP n strives to choose its attributes \mathbf{a}_n in order to optimize its profit (see below). Since the lowest possible degree of attribute

Table 8.1: Notation used in our model (in alphabetical order).

Symbol	Description
$\mathbf{A} \in \mathbb{R}_{\geq 0}^{ N \times K }$	Attributes of all ISPs N
$\mathbf{A}^+ \in \mathbb{R}_{\geq 0}^{ N \times K }$	Nash-equilibrium attributes of all ISPs N
$\mathbf{A}^\circ \in \mathbb{R}_{\geq 0}^{ N \times K }$	Socially optimal attributes (Nash bargaining solution) of all ISPs N
$\mathbf{a}_n \in \mathbb{R}_{\geq 0}^{ K }$	Attributes of ISP n
a_{nk}	Attribute $k \in K$ of ISP $n \in N$
$a_{nk}^*(\mathbf{A}_{-nk})$	Best-response attribute k of ISP n given all other attributes
$\hat{a}_{nk}^*(\mathbf{A}_{-nk})$	Unrestricted best-response attribute k of ISP n given all other attributes (potentially complex or negative and thus invalid)
α_{rnk}	Weight of attribute k of ISP n in valuation of path r
α_{r0}	Base valuation of path r
Γ_n	Demand-independent cost of ISP n
γ_{nk}	Weight of attribute k in demand-independent cost of ISP n
γ_{n0}	Attribute-independent demand-independent cost of ISP n
D_n	Demand volume forwarded by ISP n
$d_{(n_1, n_2)}$	Limit of demand of ISP n_1 towards ISP n_2
δ_r	Demand volume forwarded over path r
Φ_n	Demand-dependent cost of ISP n
ϕ_{nk}	Weight of attribute k in demand-dependent cost of ISP n
ϕ_{n0}	Attribute-independent demand-dependent cost of ISP n
K	Set of attributes
N	Set of ISPs
p_r	Selection probability of path r
π_n	Profit of ISP n
$R(n_1, n_2)$	Set of paths between ISPs n_1 and n_2
\mathcal{R}_n	Revenue of ISP n
$r \subseteq N$	Path
ρ_n	Sensitivity of revenue on attracted demand for ISP n
$V(\mathbf{A})$	Total valuation of network given attributes \mathbf{A}
v_r	Valuation of path r

prevalence is attained if an ISP does not possess an attribute at all, we restrict the attribute values to non-negative real numbers: $\forall n \in N, \forall k \in K. a_{nk} \geq 0$. For convenience of notation, we also define an attribute matrix $\mathbf{A} \in \mathbb{R}_{\geq 0}^{|N| \times |K|}$, with the n -th row being \mathbf{a}_n .

Path valuations. The attributes of an ISP n determine the attractiveness of using paths including that ISP. Hence, we define the attractiveness of available options on the level of paths, specifically by *valuation functions* $\{v_r\}_{r \in R}$. The valuation v_r for path r then depends on all attributes \mathbf{a}_n of all on-path ISPs $n \in r$. Since we consider desirable attributes in our model, every function that is monotonically increasing in all attribute values is a suitable valuation function. Throughout this chapter, we use affine functions:

$$v_r(\mathbf{A}) = \sum_{n \in r} \sum_{k \in K} \alpha_{rnk} a_{nk} + \alpha_{r0}, \quad (8.1)$$

where each $\alpha_{rnk} > 0$ determines how strongly attribute k of ISP n affects the valuation of path r , and $\alpha_{r0} \geq 0$ is the *base valuation* of path r . This formulation captures several real-world aspects of path valuations, as not all attributes are equally important and not all ISPs on a path equally affect the valuation, e.g., ISPs providing a large segment of the path might be more relevant for the valuation. The linear formulation might be counter-intuitive given that the marginal utility of attribute prevalence is likely decreasing; this property, however, is reflected in the formulation for the path-selection probability below, which predicts that the volume of attracted demand on a path is sub-linear in path attributes. Moreover, we show by simulation that the model predictions do not strongly rely on the affine formulation (cf. §8.5.2.3).

Path-selection probability. Path valuations inform the path selection at the selecting ISP, and thus determine the probability of each path being selected. More precisely, when a selecting ISP n_1 or an end-host in n_1 selects a path towards a prefix hosted by ISP n_2 , each path r among the available paths $R(n_1, n_2)$ is selected for transit with probability $p_r(\mathbf{A})$. Inspired by the popular logit-demand model [31], we consider the selection probability p_r to be proportional to the *relative attractiveness* of path r compared to alternative paths:

$$\forall (n_1, n_2) \in N \times N, r \in R(n_1, n_2). \quad p_r(\mathbf{A}) = \frac{v_r(\mathbf{A})}{1 + \sum_{r' \in R(n_1, n_2)} v_{r'}(\mathbf{A})}. \quad (8.2)$$

Crucially, the addition term 1 in the fraction denominator captures *demand elasticity*, i.e., the path selector might not select any path in $R(n_1, n_2)$ at all if the available

paths are generally unattractive. Instead, selecting ISP n_1 might not offer its customers any path to n_2 , create a new path to n_2 by concluding a peering agreement, or obtain the desired content from another destination ISP than n_2 .

Demand. The path-selection probabilities above determine the demand volume D_n that is obtained by any ISP n , which we formalize by:

$$D_n(\mathbf{A}) = \sum_{r \in R, n \in r} \delta_r(\mathbf{A}) = \sum_{\substack{r \in R, n \in r \\ r \in R(n_1, n_2)}} p_r(\mathbf{A}) \cdot d_{(n_1, n_2)}. \quad (8.3)$$

Due to the elasticity of demand, actual total demand (i.e., $\delta_r(\mathbf{A})$ summed over all $r \in R(n_1, n_2)$) is strictly below the *demand limit* $d_{(n_1, n_2)}$.

The practical interpretation of path demand δ_r in Eq. (8.3) depends on the transit behavior of the selecting ISP n_1 . If n_1 is a stub AS or operates in a path-aware Internet, then traffic can be split across multiple paths towards a given prefix. If n_1 thus selects multiple paths, δ_r denotes the actual demand allocated to path r . In contrast, if ISP n_1 is a transit AS in a BGP-based Internet, the traffic transited by n_1 must follow the single path announced by ISP n_1 to neighboring ISPs, as BGP forwarding loops might arise otherwise. If n_1 thus selects only a single path for transit, δ_r corresponds to the *expected* demand allocated on path r .

Profit. Given the demand model, an ISP n can affect attracted demand D_n with an appropriate choice of \mathbf{a}_n . However, the profit π_n depends not only on the volume of attracted demand, but also on the cost for provision of the attributes. We thus consider the profit function π_n of ISP n to have three components. First, the ISP profit is increased by a revenue function \mathcal{R}_n , which is a function of D_n . Second, the profit is reduced by a demand-dependent cost function Φ_n , which as well depends on D_n , but also directly on \mathbf{a}_n , as the cost of transmitting a unit of demand depends on the chosen attributes. Third, the profit is reduced by a demand-independent cost function Γ_n , which depends only directly on the chosen attributes \mathbf{a}_n and thus corresponds to the ‘fixed cost’ of possessing certain attributes.

While these component functions could in principle be any monotonically increasing function, we use the following formulations for the component functions throughout this chapter:

$$\begin{aligned} \mathcal{R}_n(\mathbf{A}) &= \rho_n \cdot D_n(\mathbf{A}) \\ \Phi_n(\mathbf{A}) &= \left(\sum_{k \in K} \phi_{nk} a_{nk} + \phi_{n0} \right) \cdot D_n(\mathbf{A}) \end{aligned} \quad (8.4)$$

$$\Gamma_n(\mathbf{a}_n) = \sum_{k \in K} \gamma_{nk} a_{nk} + \gamma_{n0}$$

where the coefficient $\rho_n > 0$ corresponds to the per-unit base revenue of ISP n (transit prices are additionally subsumed within quality attributes, as described below). Furthermore, the coefficients $\phi_{nk} \geq 0$ and $\gamma_{nk} \geq 0$ determine the attribute-specific increase in demand-dependent and demand-independent cost, respectively, and the intercepts $\phi_{n0} \geq 0$ and $\gamma_{n0} \geq 0$ express the attribute-independent basic values for the respective cost terms. Throughout the chapter, we assume $\rho_n \geq \phi_{n0}$ for all ISPs $n \in N$, as a rational ISP that loses money by attracting demand even with the most cost-saving strategy (i.e., $\mathbf{a}_n = \mathbf{0}$) would in fact go out of business. The affine formulations of Φ_n and Γ_n predict qualitatively similar competition effects as quadratic functions, as we demonstrate by simulation in §8.5.2.3.

In summary, the profit function π_n is of the following form in our investigations:

$$\begin{aligned} \pi_n(\mathbf{A}) &= \mathcal{R}_n(\mathbf{A}) - \Phi_n(\mathbf{A}) - \Gamma_n(\mathbf{a}_n) \\ &= D_n(\mathbf{A}) \cdot \left(\rho_n - \sum_{k \in K} \phi_{nk} a_{nk} - \phi_{n0} \right) - \sum_{k \in K} \gamma_{nk} a_{nk} - \gamma_{n0}. \end{aligned} \quad (8.5)$$

Undesirable attributes. So far, our model formulation assumes *desirable* attributes, which increase path attractiveness in high quantities and are costly to increase, e.g., bandwidth. However, many relevant ISP attributes are *undesirable* in high quantities and are challenging to decrease, e.g., transit price or latency. To accommodate undesirable attributes in the model, a naive approach would consist of allowing negative coefficients α_{rnk} , ϕ_{nk} , and γ_{nk} for any undesirable attribute k . However, such negativity would entail nonsensical model predictions, such as potentially negative path-selection probabilities from Eq. (8.2), or infinite profit given an undesirable attribute (n, k) with $a_{nk} = \infty$ and $\gamma_{nk} < 0$ (cf. Eq. (8.5)).

To avoid such nonsensical predictions and preserve model tractability, we suggest to convert undesirable attributes into their desirable counterparts. For example, the actual transit-price attribute $a'_{nk} \in [0, P_n^{\max}]$ could be translated to a non-negative *cheapness* attribute $a_{nk} = P_n^{\max} - a'_{nk} \in [0, P_n^{\max}]$, where P_n^{\max} is the maximum price that ISP n can reasonably ask. The cheapness attribute then formally contributes to the costs in Φ and Γ , while actually quantifying foregone revenue in the profit function π .

Nash equilibrium. The competition dynamics in attribute-oriented path selection can be characterized by their Nash equilibria. In our setting, such a Nash equilibrium is a choice of attributes in which each ISP has optimal attributes (w.r.t. its profit) given the attributes of all other ISPs:

Definition 8.1 *Nash Equilibrium.* A choice of attributes \mathbf{A}^+ form a Nash equilibrium if and only if

$$\forall n \in N. \mathbf{a}_n^+ = \arg \max_{\mathbf{a}_n \in \mathbb{R}_{\geq 0}^{|K|}} \pi_n(\mathbf{a}_n \oplus_n \mathbf{A}_{-n}^+) \quad (8.6)$$

where \oplus_n combines the attribute choice \mathbf{a}_n of ISP n with the equilibrium attribute choices \mathbf{A}_{-n}^+ of the remaining ISPs.

In this abstract form, the Nash equilibria offer little opportunity for analytical characterization. However, if the attributes a_{nk} are restricted to $[0, a_{\max}]$ rather than to $\mathbb{R}_{\geq 0}$ (e.g., if there is an upper bound a_{\max} on all attribute values), the existence of Nash equilibria is guaranteed by Brouwer's fixed-point theorem [42]. To gain a deeper understanding of Nash equilibria beyond that special case, we concretize equilibria in this work, and investigate these equilibria with respect to existence, uniqueness, stability, and efficiency.

Social Optimum. To assess the efficiency of Nash equilibria, we compare these equilibria to *social optima*. In our setting, such a social optimum optimizes a metric that aggregates the perspectives of all agents involved in the competitive dynamics. Our model contains two types of agents, namely *path selectors* (selecting ISPs or end-hosts) and *transit ISPs*, with non-aligned interests, which warrants two different formalizations of the social optimum.

First, path selectors are interested in path quality. Hence, the social efficiency for path selectors is simply measured by the aggregate valuation V of all paths in the network, given a choice of attributes \mathbf{A} :

$$V(\mathbf{A}) = \sum_{r \in R} v_r(\mathbf{A}). \quad (8.7)$$

Since the valuation functions v_r are assumed to be linear and therefore unbounded in this chapter, a finite social optimum for path selectors only exists if all attributes are restricted to a finite domain.

Second, transit ISPs are interested in profit. To characterize the social optimum from the perspective of transit ISPs, we rely on the conditions of the *Nash bargaining*

solution (NBS), i.e., the conditions that a global attribute choice \mathbf{A} would have to fulfill if ISPs had to agree on this attribute choice through cooperative bargaining [194]. The two most important NBS conditions are *Pareto-optimality*, i.e., no ISP can increase its profit without any other ISP experiencing a decrease in its profit, and *symmetry*, i.e., among Pareto-optimal profit distributions, the fairest distribution is preferred. These conditions are achieved if the attributes \mathbf{A} optimize the *Nash bargaining product*:

Definition 8.2 Social Optimum for Transit ISPs. A choice of attributes \mathbf{A}° forms a social optimum from the perspective of transit ISPs if it corresponds to the *Nash bargaining solution* (NBS), i.e.,

$$\mathbf{A}^\circ = \arg \max_{\mathbf{A} \in \mathbb{R}_{\geq 0}^{|N| \times |K|}} \prod_{n \in N} \pi_n(\mathbf{A}). \quad (8.8)$$

8.3 THEORETICAL ANALYSIS

In this section, we theoretically analyze the competition dynamics in path selection. For that purpose, we focus on an individual *market* in isolation, i.e., the competition between transit ISPs for traffic between a single source-destination pair (n_1, n_2) . As a result, we write $R = R(n_1, n_2)$ and $d = d_{(n_1, n_2)}$ throughout this section.

8.3.1 Optimal Attribute

To analyze the competition dynamics, we first investigate how any single ISP n should choose its attribute k in response to the attribute choices of all other ISPs. This optimal attribute is given by the following closed-form solution:

Theorem 8.1 Best-Response Attribute. In an individual market with arbitrarily overlapping paths, the optimal admissible attribute a_{nk}^* of ISP n given the remaining attributes \mathbf{A}_{-nk} is

$$a_{nk}^*(\mathbf{A}_{-nk}) = \begin{cases} \hat{a}_n^*(\mathbf{A}_{-nk}) & \text{if } \hat{a}_n^*(\mathbf{A}_{-nk}) \in \mathbb{R} \text{ and } \hat{a}_n^*(\mathbf{A}_{-nk}) \geq 0, \\ 0 & \text{otherwise,} \end{cases} \quad (8.9)$$

where $\hat{a}_n^*(\mathbf{A}_{-nk})$ is the optimal unrestricted (i.e., potentially complex or negative) attribute:

$$\begin{aligned} \hat{a}_{nk}^*(\mathbf{A}_{-nk}) = & \frac{1}{\alpha_{nk}} \left(\sqrt{\frac{d(1 + v_{-r(n)}(\mathbf{A}_{-nk}))}{d\phi_{nk} + \gamma_{nk}}} \right. \\ & \sqrt{\phi_{nk}(1 + v_{-nk}(\mathbf{A}_{-nk})) + \alpha_{nk}(\rho_n - \Phi_{-nk}(\mathbf{A}_{-nk}))} \\ & \left. - (1 + v_{-nk}(\mathbf{A}_{-nk})) \right) \end{aligned} \quad (8.10)$$

Eq. (8.10) uses the following abbreviations:

Valuation parameter for attribute (n, k) over all paths:

$$\alpha_{nk} = \sum_{\substack{r \in R \\ n \in r}} \alpha_{rnk} \quad (8.11)$$

Valuation for all paths not including ISP n :

$$v_{-r(n)}(\mathbf{A}) = \sum_{\substack{r' \in R \\ n \notin r'}} v_{r'}(\mathbf{A}) \quad (8.12)$$

Valuation for all paths including ISP n , without attribute (n, k) :

$$v_{-nk}(\mathbf{A}) = \sum_{r \in R} \alpha_{r0} + \sum_{\substack{(n', k') \in N \times K \\ (n', k') \neq (n, k)}} \alpha_{n'k'} a_{n'k'} \quad (8.13)$$

Demand-dependent cost of ISP n , without attribute (n, k) :

$$\Phi_{-nk}(\mathbf{A}) = \sum_{k' \in K \setminus \{k\}} \phi_{nk'} a_{nk'} + \phi_{n0} \quad (8.14)$$

The proof of Theorem 8.1 is provided in Appendix F.1. To provide intuition about the formula in Eq. (8.10), we transform it to a simplified version:

$$\hat{v}_{r(n)}^* = \sqrt{\left(\underbrace{\frac{d\phi_{nk}(1 + v_{-nk})}{d\phi_{nk} + \gamma_{nk}}}_{\textcircled{1}} + \underbrace{\frac{d\alpha_{nk}(\rho_n - \Phi_{-nk})}{d\phi_{nk} + \gamma_{nk}}}_{\textcircled{2}} \right) \underbrace{(1 + v_{-r(n)})}_{\textcircled{3}} - \underbrace{(1 + v_{-r(n)})}_{\textcircled{3}}} \quad (8.15)$$

where

$$\hat{v}_{r(n)}^* = \sum_{\substack{r \in R \\ n \in r}} \hat{v}_r^* \quad (8.16)$$

contains the sum of unrestricted valuations of all paths containing ISP n that would be optimal for n given \mathbf{A}_{-nk} .

This $\hat{v}_{r(n)}^*$ (and thus also \hat{a}_{nk}) correlates positively with term $\textcircled{1}$, which relates to the share of demand-dependent cost ($\propto d\phi_{nk}$) among total cost ($\propto d\phi_{nk} + \gamma_{nk}$) with respect to attribute (n, k) . This correlation suggests that ISPs should champion attributes with low demand-independent cost.

Moreover, $\hat{v}_{r(n)}^*$ correlates with term $\textcircled{2}$, which relates to the revenue from attribute (n, k) per unit of cost from the attribute, i.e., the ‘return’ on attribute (n, k) .

Term $\textcircled{3}$, which describes the attractiveness of paths avoiding n , can have a positive effect on $\hat{v}_{r(n)}^*$ up to a point, as competition incentivizes ISP n to raise its attribute values. However, from a certain point onwards, term $\textcircled{3}$ has a negative effect on $\hat{v}_{r(n)}^*$, as detracting traffic from highly attractive alternatives becomes too costly compared to the achievable revenue.

For an individual market, the equilibrium condition from Definition 8.2 can thus be concretized based on Theorem 8.1: A choice of attributes \mathbf{A}^+ is a Nash equilibrium if and only if

$$\forall n \in N, \forall k \in K. \quad a_{nk}^+ = a_{nk}^*(\mathbf{A}_{-nk}^+). \quad (8.17)$$

For the general case, we find that deriving equilibria based on this condition is intractable. For example, when considering a market with two disjoint paths, a single attribute, and a single ISP with arbitrary parameters on each path, the equilibrium must be found by solving a quartic equation, which impedes an analysis even for that simple network. Fortunately, we identify two types of markets that allow deriving closed-form equilibria and therefore analytic insights, while still capturing the fundamental characteristics of ISP competition, i.e., inter-path

competition, intra-path cooperation, and ISP heterogeneity. Concretely, we separately analyze homogeneous markets (cf. §8.3.2) and heterogeneous markets with attribute-independent traffic-unit cost (cf. §8.3.3), both with disjoint paths.

8.3.2 Homogeneous Markets

By homogeneous markets, we refer to topologies of $Q := |R| > 0$ disjoint paths in competition, each of which accommodates the same number $I = |N|/Q$ of ISPs. All ISPs are identical and all attributes are identically valuable and costly, i.e., for all ISPs $n \in N$, it holds that $\rho_n = \rho$ and $\phi_{n0} = \phi_0$, and $\forall k \in K$, it holds that $\alpha_{nk} = \alpha_1$, $\phi_{nk} = \phi_1$, and $\gamma_{nk} = \gamma_1$. Moreover, the path-valuation functions for all paths are identical as well, i.e., $\forall r \in R$. $\alpha_{r0} = \alpha_0$. While artificial, such competition among completely equal goods (here: paths) and firms (here: ISPs) is common in competition models, as homogeneity allows isolating pure competition effects that are not due to variety in offers [66, 78].

In our case, the homogeneity also permits to identify the Nash equilibria of the competition dynamics (§8.3.2.1), to investigate the convergence to these equilibria (§8.3.2.2), to compare these equilibria to social optima (§8.3.2.3), and to evaluate the effect of competition intensity (§8.3.2.4).

8.3.2.1 Equilibria

The symmetry of the homogeneous markets allows finding an equilibrium:

Theorem 8.2 Nash Equilibrium in Homogeneous Markets. *The Nash equilibrium of a homogeneous market is given by an attribute sum a^+ such that $\sum_k a_{nk}^+ = a^+ \forall n \in N$, where $a^+ = \max(0, \hat{a}^+)$ with*

$$\hat{a}^+ = \frac{\sqrt{T_2^2 - 4T_1T_3} - T_2}{2T_1}. \quad (8.18)$$

Eq. (8.18) uses the following abbreviations:

$$T_1 = Q^2 I^2 \alpha_1^2 - \frac{d}{d\phi_1 + \gamma_1} (QI - 1)(Q - 1) I \alpha_1^2 \phi_1, \quad (8.19)$$

$$T_2 = 2QI\alpha_1 (1 + Q\alpha_0) - \frac{d}{d\phi_1 + \gamma_1}. \quad (8.20)$$

$$\left(\alpha_1 \phi_1 (QI - 1) (1 + (Q - 1)\alpha_0) + I\alpha_1 (Q - 1) (\phi_1 (1 + Q\alpha_0) + \alpha_1 (\rho - \phi_0)) \right), \text{ and} \quad (8.21)$$

$$T_3 = (1 + Q\alpha_0)^2 - \frac{d}{d\phi_1 + \gamma_1}. \quad (8.22)$$

$$(1 + (Q - 1)\alpha_0)(\phi_1(1 + Q\alpha_0) + \alpha_1(\rho - \phi_0)).$$

The proof of Theorem 8.2 is provided in Appendix F.2. Note that the equilibrium in Theorem 8.2 is only unique with respect to the attribute sum a^+ of any ISP and hence also with respect to path valuations, as the equilibrium path valuation is $v_r(\mathbf{A}^+) = I\alpha_1 a^+ + \alpha_0$. However, the equilibrium is not necessarily unique with respect to the individual attribute values a_{nk} .

8.3.2.2 Stability

The Nash equilibrium from Theorem 8.2 is an interesting fixed point of the competitive dynamics in homogeneous markets. However, the equilibrium is only relevant if the distributed profit optimization by the ISPs converges to it. Hence, the equilibrium must be additionally investigated with respect to its *stability*, i.e., its attractive effect on the competition dynamics. To investigate this stability, we formally describe the competition by the following system of ordinary differential equations (ODEs):

$$\forall n \in N. \quad \dot{a}_n(t) = a_n^*(\mathbf{A}_{-n}(t)) - a_n(t) \quad (8.23)$$

Intuitively, one ODE in this system describes the behavior of an ISP n which continuously adjusts its attribute-sum value a_n towards the optimal choice a_n^* given the contemporary attribute values of all other ISPs. Given this dynamic process, we can show the following property:

Theorem 8.3 Stability of Homogeneous Equilibrium. *The Nash equilibrium from Theorem 8.2 is an asymptotically stable equilibrium of the competition dynamics in Eq. (8.23).*

The proof of Theorem 8.3 is provided in Appendix F.3.

8.3.2.3 Intra-Path Dynamics

The equilibrium formalization from Theorem 8.2 also applies to the case where an ISP pair is only connected by a single usable path. Such a single-path scenario

represents a monopoly in economic terms. Crucially, the ISPs on the same single path are supposed to *cooperate* rather than compete, as the decisions by each ISP contribute to path attractiveness, which in turn benefits all ISPs. For that single-path case, we can make the following interesting observation about the cooperation among on-path ISPs:

Theorem 8.4 Suboptimality of Homogeneous Equilibrium. *On a single path with I identical ISPs, the equilibrium attribute sum a^+ is generally lower than the NBS attribute sum a° , i.e., $\forall I \in \mathbb{N}, I \geq 1. a^+ \leq a^\circ$.*

The proof of Theorem 8.4 is provided in Appendix F.4.

Intuitively, Theorem 8.4 states that the cooperation by on-path ISPs suffers from inefficiency caused by individual selfishness, similar to a prisoner's dilemma [216]. More precisely, the NBS attribute sum a° , which would optimize every ISP's profit if chosen universally, is not a rational choice for an individual ISP. In particular, if an ISP n chooses $a_n = a^\circ$, ISP n enables another ISP m to optimize its profit π_m by choosing a lower attribute sum $a_m < a^\circ$, and thus to free-ride on the path attractiveness created by ISP n . Because of this selfish deviation from the global optimum, the on-path ISPs converge to the equilibrium attribute prevalence a^+ , which is generally lower than the NBS attribute sum a° , prevents the transit ISPs from reaping optimal profit, and also reduces the path attractiveness for the path selectors.

8.3.2.4 Competition Effects

After investigating intra-path cooperation in the preceding section, we now investigate the effect of *inter-path competition* on attribute prevalence. In particular, we are interested in the dependence of the equilibrium attribute a^+ on the number of available paths between an origin-destination pair.

To characterize this dependence, we compare the Nash equilibria in two homogeneous markets. First, we consider the competition-free network \mathcal{N}_1 in Fig. 8.1a, which is partitioned between Q origin-destination pairs $\{(m_{q1}, m_{q2})\}_{q=1,\dots,Q}$, each connected by a single path with I ISPs and obtaining the same demand limit d' . This network is a set of homogeneous markets, each with one available path and no competition. Second, we consider the competitive network \mathcal{N}_2 in Fig. 8.1b, where each of the Q origin-destination pairs can use all Q available paths.

By identifying the equilibrium attribute value $a^+(\mathcal{N})$ for each network \mathcal{N} , we find that competition has a consistently positive effect on attribute prevalence:

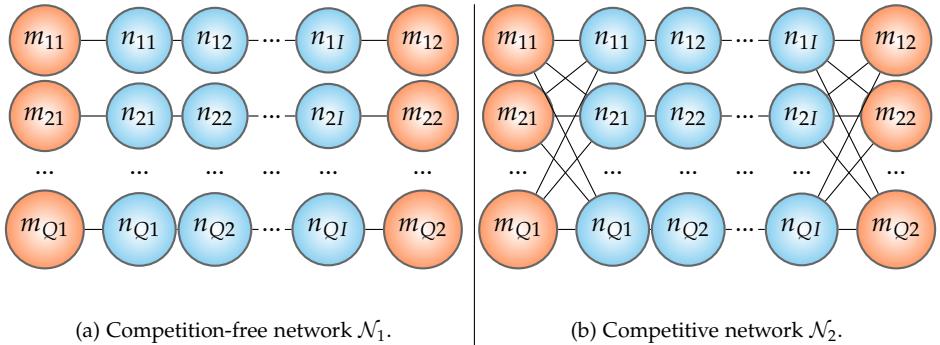


Figure 8.1: Homogeneous markets with and without inter-path competition.

Theorem 8.5 Attribute Improvement under Homogeneous Competition. *The equilibrium attribute prevalence is never lower in the competitive network \mathcal{N}_2 than in the competition-free network \mathcal{N}_1 , i.e., $a^+(\mathcal{N}_2) \geq a^+(\mathcal{N}_1)$ for all $Q \in \mathbb{N} \geq 1$.*

The proof of Theorem 8.5 is provided in Appendix F.5.

Surprisingly, the higher attribute values in the competitive equilibrium do not necessarily come at the cost of lower ISP profits. Instead, the profit of the ISPs may even increase under competition, which is important because the ISPs partially control whether they engage in competition at all, namely through link setup and path announcements. For example, an increase in equilibrium profits through competition happens if competition causes only a modest increase in attribute values:

Theorem 8.6 Profit Improvement under Homogeneous Competition. *The equilibrium profit $\pi^+(\mathcal{N}_2)$ in the competitive network preserves or exceeds the equilibrium profit $\pi^+(\mathcal{N}_1)$ of the uncompetitive network if $a^+(\mathcal{N}_2) \in [a^+(\mathcal{N}_1), a^\circ(\mathcal{N}_1)]$, i.e., the equilibrium attribute sum $a^+(\mathcal{N}_2)$ from the competitive network is between the equilibrium attribute sum $a^+(\mathcal{N}_1)$ of the uncompetitive network and the corresponding NBS attribute sum $a^\circ(\mathcal{N}_1)$.*

The proof of Theorem 8.6 is provided in Appendix F.6. In summary, we conclude that inter-path competition in homogeneous markets is always desirable from the perspective of path selectors, and potentially desirable from the perspective of transit ISPs.

8.3.3 Heterogeneous Markets

The homogeneous markets discussed in the previous section can reflect competition dynamics among arbitrarily many paths. However, these market models cannot represent differences between paths that go beyond attribute values. In reality, on-path ISPs may differ in their importance for the path valuation, in their revenue per traffic unit, and in their attribute-specific costs. In this section, we study competition among such heterogeneous ISPs, i.e., every ISP n has arbitrary parameters $\alpha_{nk} \forall k \in K$, ρ_n , ϕ_{n0} , $\gamma_{nk} \forall k \in K$, and γ_{n0} . To achieve tractability despite this additional complexity, we restrict our analysis to markets with at most two paths. Moreover, since traffic-unit cost is commonly considered negligible for ISPs [250], we consider especially the attribute-dependent part of this traffic-unit cost to be negligible, i.e., $\phi_{nk} = 0 \forall n \in N, k \in K$.

8.3.3.1 Intra-Path Dynamics

To understand the attribute-choice dynamics among ISPs in a heterogeneous market, we first consider a single path in isolation, i.e., a monopoly scenario. As before, ISPs on a path collectively determine the attractiveness of the path, but optimize only their individual profit. This selfishness may lead to a sub-optimal global outcome, both regarding ISP profits and path valuations. To quantify this shortfall, we first identify the Nash bargaining solution (NBS) for the attribute choices, i.e., the attribute values that all on-path ISPs would agree on if they were bound by the result of a collective negotiation. This Nash bargaining solution represents the global optimum with respect to the ISP profits.

Theorem 8.7 Profit Optimum on Heterogeneous Paths. On a path r with heterogeneous ISPs, the attributes \mathbf{A}° form a Nash bargaining solution if and only if

$$\mathbf{A}^\circ = \arg \max_{\mathbf{A} \in \mathbb{R}_{\geq 0}} \prod_{n \in r} \pi_n(\mathbf{A}) \quad (8.24)$$

subject to

$$v_r(\mathbf{A}) = v_r^\circ \quad \forall (n, k) \in K_r^\circ. a_{nk} \geq 0 \quad \forall (n, k) \notin K_r^\circ. a_{nk} = 0 \quad (8.25)$$

where $v_r^\circ = \max \left(\alpha_{r0}, \max_{(n', k') \in r \times K} \sqrt{\frac{\alpha_{n'k'}}{\gamma_{n'k'}}} \sqrt{d \sum_{n \in r} (\rho_n - \phi_{n0})} - 1 \right), \quad (8.26)$

and $K_r^\circ = \left\{ (n, k) \mid (n, k) = \arg \max_{(n', k') \in r \times K} \frac{\alpha_{n'k'}}{\gamma_{n'k'}} \right\}. \quad (8.27)$

The proof of Theorem 8.7 is provided in Appendix F.7. More informally, the attributes \mathbf{A}° form a Nash bargaining solution if and only if these attributes optimize the product of all ISP profits while

1. leading to the NBS path valuation v_r° ,
2. containing non-zero attribute values only for attributes (n, k) with maximal ratio $\frac{\alpha_{nk}}{\gamma_{nk}}$,
3. and containing zero attribute values for all other attributes.

To optimize aggregate profit, the on-path ISPs should thus only upgrade the attribute(s) with maximal ‘return’ α_{nk}/γ_{nk} while minimizing the prevalence of all other attributes. This return ratio α_{nk}/γ_{nk} yields the valuation for attribute k of ISP n , compared to the cost that ISP n incurs for adopting that attribute. The return ratio also correlates with the optimal path valuation (cf. Eq. (8.26)).

However, aggregate profit is not the objective of selfish ISPs when determining attribute values. Instead, selfish ISPs optimize their individual profit, and eventually arrive at the following equilibrium by their non-aligned optimization behavior:

Theorem 8.8 Nash Equilibrium on Heterogeneous Paths. On a path r with heterogeneous ISPs, the attributes \mathbf{A}^+ form a Nash equilibrium if and only if

$$v_r(\mathbf{A}^+) = v_r^+ \quad \forall (n, k) \in K_r^+ \cdot a_{nk}^+ \geq 0 \quad \forall (n, k) \notin K_r^+ \cdot a_{nk}^+ = 0 \quad (8.28)$$

where $v_r^+ = \max \left(\alpha_{r0}, \max_{(n', k') \in r \times K} \sqrt{\frac{\alpha_{n'k'}}{\gamma_{n'k'}}} \sqrt{d(\rho_{n'} - \phi_{n'0})} - 1 \right), \quad (8.29)$

and $K_r^+ = \left\{ (n, k) \mid (n, k) = \arg \max_{(n', k') \in r \times K} \frac{\alpha_{n'k'}(\rho_{n'} - \phi_{n'0})}{\gamma_{n'k'}} \right\}. \quad (8.30)$

The proof of Theorem 8.8 is provided in Appendix F.8. More informally, the attributes \mathbf{A}^+ form a Nash equilibrium if and only if these attributes

1. lead to the equilibrium path valuation v_r^+ ,
2. contain non-zero attribute values only for attributes (n, k) with maximal ratio $\frac{\alpha_{nk}(\rho_n - \phi_{n0})}{\gamma_{nk}}$,
3. and contain zero attribute values for all other attributes.

Interestingly, the equilibrium in Theorem 8.8 is similar to the Nash-bargaining solution in Theorem 8.7, but contains one crucial difference: The return ratio associated with cultivated attributes includes the net revenue per unit of traffic $\rho_n - \phi_{n0}$ of ISP n (Eq. (8.27) vs. Eq. (8.30)). This inclusion reflects that each ISP n optimizes its individual profit rather than the aggregate profit: When optimizing an attribute (n, k) for individual profit, an ISP n only considers its individual net revenue per traffic unit, not the *aggregate* net revenue per traffic unit, which would be relevant for aggregate profit.

This difference, albeit subtle, generally leads to different attribute choices in equilibrium than postulated by the Nash-bargaining solution, meaning that the transit ISPs generate sub-optimal profits. Unfortunately, also the path selectors suffer from this selfishness, as the individual-profit optimization leads to less valuable paths:

Theorem 8.9 Suboptimality of Heterogeneous Equilibrium. On a path r with heterogeneous ISPs, the equilibrium path valuation $v_r(\mathbf{A}^+)$ never exceeds the NBS path valuation $v_r(\mathbf{A}^\circ)$, i.e., $v_r(\mathbf{A}^+) \leq v_r(\mathbf{A}^\circ)$.

The proof of Theorem 8.9 is provided in Appendix F.9.

8.3.3.2 Two-Path Equilibria

In the preceding section, the social optimum and the Nash equilibrium are characterized for a single-path scenario, which is exclusively informed by (failing) intra-path cooperation among selfish ISPs. Since we are also interested in the effect of inter-path competition, we now consider heterogeneous markets in which the path selectors can select between two disjoint paths. For these networks, the single-path equilibrium in Theorem 8.8 can be adjusted as follows:

Theorem 8.10 Nash Equilibrium in Heterogeneous Markets. *In a two-path heterogeneous market, the attribute values \mathbf{A}^+ form a Nash equilibrium if and only if the attribute values \mathbf{A}^+ satisfy the conditions from Theorem 8.8, but with modified equilibrium path valuation v_r^+ :*

$$v_r^+ = \max(\alpha_{r0}, \hat{v}_r^* (\max(\alpha_{\bar{r}0}, \hat{v}_{\bar{r}}^+))) \quad (8.31)$$

where \bar{r} is the alternative path to r ,

$$\hat{v}_r^*(v_{\bar{r}}) = \psi_r \sqrt{d} \sqrt{1 + v_{\bar{r}}} - (1 + v_{\bar{r}}), \quad (8.32)$$

$$\hat{v}_r^+ = \frac{\psi_r^3 \psi_{\bar{r}}}{(\psi_r^2 + \psi_{\bar{r}}^2)^2} \left(\sqrt{d (\psi_r^2 + \psi_{\bar{r}}^2)} + \frac{1}{4} \psi_r^2 \psi_{\bar{r}}^2 d^2 + \frac{d}{2} \psi_r \psi_{\bar{r}} \right) \quad (8.33)$$

$$- \frac{\psi_{\bar{r}}^2}{\psi_r^2 + \psi_{\bar{r}}^2}, \text{ and}$$

$$\psi_r = \max_{\substack{n \in r \\ k \in K}} \sqrt{\frac{\alpha_{nk}(\rho_n - \phi_{n0})}{\gamma_{nk}}}. \quad (8.34)$$

The proof of Theorem 8.10 is provided in Appendix F.10. We note that ψ_r from Eq. (8.34) is the square root of the maximum individual return ratio discussed in the previous section, albeit only among the attributes of path r . In the following, we refer to ψ_r as the *characteristic ratio* of path r .

Similar to §8.3.2, we are again interested in the stability of the equilibrium w.r.t. the process:

$$\forall n \in N, k \in K. \quad \dot{a}_{nk}(t) = a_{nk}^*(\mathbf{A}_{-nk}(t)) - a_{nk}(t). \quad (8.35)$$

However, stability analysis in the case of heterogeneous two-path networks is complicated by the fact that the equilibrium from Theorem 8.10 is only unique in

the path valuations $\{v_r\}_{r \in R}$, but not necessarily unique in the attribute choices \mathbf{A} by the ISPs. Therefore, if the equilibrium is not unique in \mathbf{A} , no single equilibrium \mathbf{A}^+ is asymptotically stable in a narrow sense, as the process in Eq. (8.35) does not converge to \mathbf{A}^+ from $\mathbf{A}(t)$ in case $\mathbf{A}(t)$ already represents a different equilibrium.

Therefore, we focus on the stability of unique equilibria:

Theorem 8.11 Stability of Heterogeneous Equilibrium. *The Nash equilibrium \mathbf{A}^+ from Theorem 8.10 is an asymptotically stable equilibrium of the competition dynamics in Eq. (8.35) if the equilibrium \mathbf{A}^+ is unique, i.e., if there is only one attribute on every path which has potentially non-zero prevalence ($|K_r^+| = |K_{\bar{r}}^+| = 1$).*

The proof of Theorem 8.11 is provided in Appendix F.11.

8.3.3.3 Competition Effects

Based on the equilibria for single-path and two-path markets, we now investigate the effect of inter-path competition in heterogeneous markets. For this investigation, we use a similar approach as in §8.3.2.4: We contrast a competition-free network \mathcal{N}_3 , which consists of two paths r and \bar{r} , each connecting one origin-destination pair, with a competitive network \mathcal{N}_4 , where both origin-destination pairs are connected by both paths. The origin-destination pair connected by path r in the competition-free network \mathcal{N}_3 has demand limit d_r ; hence, the total demand limit $d = d_r + d_{\bar{r}}$ is distributed over both paths in the competitive network \mathcal{N}_4 . The networks \mathcal{N}_3 and \mathcal{N}_4 thus differ in the same manner as the networks \mathcal{N}_1 and \mathcal{N}_2 from Fig. 8.1, except that the different paths may have different length in ISPs, each ISP may have different parameters, and each origin-destination pair may have a different demand limit. When contrasting these two networks, we gain the following insight:

Theorem 8.12 Attribute Improvement under Heterogeneous Competition. *For any competition-free network \mathcal{N}_3 and the corresponding competitive network \mathcal{N}_4 , a demand limit d exists such that the competitive network \mathcal{N}_4 has a higher equilibrium valuation than the competition-free network \mathcal{N}_3 independent of the demand distributions $(d_r, d_{\bar{r}})$, i.e.,*

$$\exists d \text{ s.t. } \forall d_r, d_{\bar{r}} \text{ with } d_r + d_{\bar{r}} = d. V^+(\mathcal{N}_4) \geq V^+(\mathcal{N}_3) \quad (8.36)$$

The proof of Theorem 8.12 is provided in Appendix F.12.

In simplified terms, inter-path competition thus affects the attribute values and the path valuations positively for high-enough demand, given the remaining network parameters. This condition on demand, however, raises the question whether competition reduces the network valuation in some circumstances. Indeed, we find that such a counter-intuitive effect can arise at every demand level if the remaining network parameters are unfavorable:

Theorem 8.13 Attribute Decline under Heterogeneous Competition. *For every demand distribution $(d_r, d_{\bar{r}})$, there exist characteristic ratios $(\psi_r, \psi_{\bar{r}})$ and path base valuations $(\alpha_{r0}, \alpha_{\bar{r}0})$ such that the competitive network \mathcal{N}_4 has a lower equilibrium valuation than the competition-free network \mathcal{N}_3 , i.e.,*

$$\forall d_r, d_{\bar{r}}. \exists \psi_r, \psi_{\bar{r}}, \alpha_{r0}, \alpha_{\bar{r}0} \text{ s.t. } V^+(\mathcal{N}_4) < V^+(\mathcal{N}_3). \quad (8.37)$$

The proof of Theorem 8.13 is provided in Appendix F.13.

To understand this effect intuitively, we note that an ISP n optimizes its profit by balancing the *marginal revenue* and the *marginal cost* with respect to attribute prevalence, i.e., adjusts attribute prevalence as long as the adjustment generates more revenue than cost. In the competition-free scenario of \mathcal{N}_3 , the marginal revenue and cost of an ISP n with respect to attribute (n, k) are:

$$\frac{\partial \mathcal{R}_n}{\partial a_{nk}} = \frac{d_r \alpha_{nk}}{(1 + v_r)^2} \cdot \rho_n \quad \frac{\partial}{\partial a_{nk}} (\Phi_n + \Gamma_n) = \frac{d_r \alpha_{nk}}{(1 + v_r)^2} \cdot \phi_{n0} + \gamma_{nk} \quad (8.38)$$

In contrast, the corresponding terms for the competitive scenario in network \mathcal{N}_4 are as follows:

$$\frac{\partial \mathcal{R}_n}{\partial a_{nk}} = \frac{d \alpha_{nk} \cdot (1 + v_{\bar{r}})}{(1 + v_r + v_{\bar{r}})^2} \cdot \rho_n \quad \frac{\partial}{\partial a_{nk}} (\Phi_n + \Gamma_n) = \frac{d \alpha_{nk} \cdot (1 + v_{\bar{r}})}{(1 + v_r + v_{\bar{r}})^2} \cdot \phi_{n0} + \gamma_{nk} \quad (8.39)$$

On the one hand, competition has a positive effect on marginal revenue $\partial \mathcal{R}_n / \partial a_{nk}$ by increasing the total amount of attractable demand from d_r to $d > d_r$. On the other hand, the new competition embodied by the alternative-path valuation $v_{\bar{r}}$ has a negative effect on marginal revenue. The negative effect predominates if the alternative-path valuation $v_{\bar{r}}$ is relatively large and unresponsive to competition, as the proof of Theorem 8.13 demonstrates. If marginal revenue in fact decreases, marginal cost decreases less strongly as $\rho_n \geq \phi_{n0}$. Given negative marginal profit, the profit of ISP n is thus optimized by a lower attribute prevalence a_{nk} , which translates into decreasing path value.

8.4 A MODEL INSTANCE BASED ON REAL-WORLD DATA

In this section, we demonstrate how to instantiate our competition model from §8.2 to investigate a large-scale network containing multiple intertwined markets. To that end, we construct a topology approximating the Internet core and a corresponding traffic matrix in §8.4.1. Furthermore, we consider two ISP attributes in the competitive dynamics, namely *internal bandwidth* and *clean-energy share*, and estimate appropriate model parameters in §8.4.2 and §8.4.3, respectively. Attribute-independent parameters are estimated in §8.4.4.

Importantly, we note that estimating highly realistic parameters for the model goes beyond the scope of this chapter, as the scarcity of publicly available data and the complexity of real-world business practices considerably complicates this estimation. Therefore, the goal of the following parameter estimation is to place the parameters in the right order of magnitude, especially in relation to each other, rather than to determine each parameter highly realistically. Interestingly, our sensitivity analysis in §8.5 suggests that such an approximate estimation might be sufficient to yield useful predictions.

8.4.1 Network Topology and Demand

To investigate the effects of competition in practically interesting, large-scale settings while keeping the complexity of the simulation manageable, we extract a network topology that roughly approximates the Internet core from a public dataset. In particular, we rely on a CAIDA dataset containing 12 300 autonomous systems (ASes), their economic relationships, and the geolocation of their interconnections (i.e., inter-domain interfaces) [53]. From this dataset, we extract the topology of the 2000 most interconnected ASes by iteratively removing the lowest-degree ASes.

In this reduced topology, we aim at finding the 5 shortest paths between every origin-destination pair of ASes. For scalability, we can only consider AS paths with at most 4 AS hops, which is not a strong limitation: The paths in our topology only represent the core-traversing segments of whole Internet paths, which have an average length of around 5 hops (and decreasing) [126]. Moreover, for both scalability and practical relevance, we only consider paths that are Gao-Rexford-compliant [91], i.e., are compatible with the economic self-interest of ASes regarding monetization of traffic. With these constraints, we can identify 5 paths for $\sim 52.4\%$ of AS pairs in the topology.

While only a subset of all AS pairs, these pairs of closely located ASes are disproportionately relevant for the competition dynamics, as they account for a

substantial share of traffic given the gravity-like nature of Internet traffic [220]. Generally, gravity models predict that the traffic demand $d_{(n_1, n_2)}$ between two ISPs n_1 and n_2 is proportional to the product of the ‘masses’ $m_1 \cdot m_2$ of the two ISPs divided by the squared distance r_{12}^2 between the ISPs:

$$d_{(n_1, n_2)} \propto G_{12} = \frac{m_1 \cdot m_2}{r_{12}^2}. \quad (8.40)$$

In order to synthesize a traffic matrix for our purpose, we concretize this gravity model as follows. First, we calculate the mass m_n of an AS n as the number of distinct IPs in all prefixes owned by AS n and by the ASes in the customer cone of AS n . This information is available via the datasets ‘Routeviews Prefix-to-AS Mapping’ [55] and ‘AS Relationships’ [54], both from CAIDA. Second, we determine the distance r_{12} for each AS pair (n_1, n_2) as the average number of hops in the 5 paths connecting the AS pair. Third, we calculate the gravity G_{12} according to Eq. (8.40) for every AS pair (n_1, n_2) . Finally, we allocate the total Internet traffic volume of 170 Tbps [178] to the AS pairs (n_1, n_2) according to the relative size of G_{12} .

8.4.2 Attribute 1: Internal Bandwidth

To instantiate the model, we define the ISP attributes K that are affected by the competitive dynamics, and the corresponding model parameters. As an intuitive example of desirable ISP attributes, we consider the *internal bandwidth* of an ISP (in Gbps) the first such attribute ($k = 1$). If the ISPs along a path have a large bandwidth capacity, these ISPs are likely able to absorb sudden traffic surges, tolerate equipment failures, handle large traffic flows, and in general deliver a high quality of service; hence, the internal bandwidth of on-path ISPs correlates with the attractiveness of the given path.

8.4.2.1 Valuation

This valuation by path-selecting ISPs is quantified by the valuation function v_{r1} , quantifying the valuation of a bit traversing path r given the internal bandwidth of on-path ISPs. This valuation function v_{r1} is characterized by the parameters α_{rn1} , giving the valuation of a bit traversing ISP n on path r for a unit of the internal bandwidth of ISP n . For this quantification, we rely on two empirical findings. First, the average US consumer transmits 536.3 GB of data per month [202], and is willing to pay 94 USD per month for a 1Gbps connection [162]. Hence, we arrive at

a monthly willingness-to-pay of around $w = 0.17$ USD per GB at the quality of a 1Gbps connection. With this willingness-to-pay w , we determine the bandwidth valuation parameters α_{rn1} , namely by defining $\alpha_{rn1} = w/(|r| \cdot m_n)$, where $|r|$ is the number of ASes on path r (averaging the internal bandwidth across on-path ISPs) and m_n is the number of IPs in the customer cone of AS n (correcting for the number of end-points sharing the bandwidth). Multiplied with the internal bandwidth a_{n1} , these valuation parameters thus approximate the valuation per bit traversing ISP n given the internal bandwidth of ISP n . Furthermore, the bandwidth valuation function v_{r1} is also characterized by the base valuation α'_{r0} of path r . However, since a path only has value in terms of bandwidth if the on-path ASes have non-zero internal bandwidth, we choose $\alpha'_{r0} = 0$.

8.4.2.2 Cost

Apart from increasing valuation by path selectors, providing bandwidth also has a cost. However, it is difficult to quantify the cost of providing a Gbps of internal bandwidth, as this cost heavily depends on the way of provision (leasing or physically installing new capacity), on the necessary installation procedures (e.g., length of cables to be newly laid), on the location where capacity should be added, and on numerous other aspects. Hence, we rely on the simple insight that the cost of providing a Gbps of connectivity is likely lower than the corresponding willingness-to-pay by consumers (94 USD per Gbps per month [162]), as ISPs would go out of business otherwise. Hence, we randomly vary the cost parameter γ_{n1} between 0 and 94 USD per Gbps per month in our simulations, for all $n \in N$. Importantly, the provision of bandwidth only affects the demand-independent cost Γ_n of an ISP n , as providing a certain bandwidth capacity causes the same cost independent of the actually experienced demand. Hence, we can also define the demand-dependent cost parameter for the bandwidth attribute $k = 1$: $\phi_{n1} = 0$ for all $n \in N$.

8.4.2.3 Attribute Bounds.

Using internal bandwidth as one of multiple attributes leads to an implausible model prediction in the case where all ASes on a path r have zero internal bandwidth ($a_{n1} = 0 \forall n \in r$), but some non-zero values for other attributes. In that case, the valuation function v_r might still assign some non-zero valuation and some demand to path r , although the zero-bandwidth path r is clearly worthless. To avoid this implausible case of the model, we place a lower bound on the bandwidth attribute $a_{n1} \forall n \in N$. This lower bound is given by 10% of the demand experienced

by AS n if the demand of every origin-destination pair was equally distributed among the available 5 paths:

$$\forall n \in N. a_{n1} \geq \frac{0.1}{5} \cdot \sum_{\substack{r \in R. n \in r \\ r \in R(n_1, n_2)}} d_{(n_1, n_2)} \quad (8.41)$$

8.4.3 Attribute 2: Clean-Energy Share

Path-selection preferences are not exclusively related to transmission performance (such as internal bandwidth of on-path ISPs), but may also reflect ESG considerations [57, 164]. For example, in carbon-intelligent routing [11, 274], path selection takes into account the carbon emission that results from data transmission. More precisely, the path-specific *transmission carbon intensity*, i.e., the volume of carbon emission per bit of transmitted data on a given path, affects path selection. To investigate the effect of competition on this carbon intensity, we choose the *share of clean energy* used by an ISP (in percent) as the second attribute ($k = 2$) for our simulations, i.e., $a_{n2} \in [0, 1] \forall n \in N$.

8.4.3.1 Transmission carbon intensity

The clean-energy share attributes of on-path ASes determine the carbon intensity of a path as follows. First, any AS-level path r must be transformed into a router-level path s_r , which is possible by means of the CAIDA ITDK dataset [56]. For simplicity, we assume that the intra-AS router-level path s_{rn} in AS n is the shortest router-level path between the two AS interconnections derived from the AS-level path r . For any intra-AS path s_{rn} , we determine the energy intensity e_{rn} , i.e., the amount of consumed electricity per bit transmitted on path s_{rn} . This energy intensity e_{rn} can be calculated from the number of routers and the covered distance of path s_{rn} , given by the CAIDA ITDK dataset, and the energy-intensity values for various devices, as reported by Heddeghem et al. [116]. Then, we calculate the maximum transmission carbon intensity $c_{rn,\max}$ of any intra-AS path s_{rn} by multiplying the corresponding energy intensity e_{rn} with the energy carbon intensity c_{\max} of the most carbon-intensive electricity, namely 875 gCO₂/kWh for coal-generated electricity [127]. This maximum transmission carbon intensity $c_{rn,\max}$ thus quantifies the carbon emission associated with the transmission of a bit across path s_{rn} if AS n used maximally carbon-intensive electricity. Finally, we derive the *actual* transmission carbon intensity c_{rn} of any intra-AS path s_{rn} as the product of the maximum transmission carbon intensity $c_{rn,\max}$ and the dirty-energy share of ISP n ,

i.e., $1 - a_{n2}$. The carbon intensity c_r of a path r is the sum of carbon-intensity values c_{rn} of the constituting intra-AS paths $s_{rn} \forall n \in r$:

$$c_r(\mathbf{A}) = \sum_{n \in r} c_{rn}(\mathbf{A}) = \sum_{n \in r} c_{rn,\max} \cdot (1 - a_{n2}) = \sum_{n \in r} e_{rn} \cdot c_{\max} \cdot (1 - a_{n2}). \quad (8.42)$$

8.4.3.2 Valuation

This carbon-intensity calculation also informs the valuation v_{r2} , which quantifies the valuation of path r exclusively with respect to carbon emissions. In fact, we understand v_{r2} as an affine function of the *negative* carbon intensity of path r :

$$\begin{aligned} v_{r2}(\mathbf{A}) &= \sum_{n \in r} \alpha_{rn2} a_{n2} + \alpha''_{r0} = - \sum_{n \in r} p_{CO_2} c_{rn}(\mathbf{A}) + q_r \\ &= \sum_{n \in r} (p_{CO_2} \cdot c_{rn,\max} \cdot a_{n2} - p_{CO_2} \cdot c_{rn,\max}) + q_r, \end{aligned} \quad (8.43)$$

where p_{CO_2} is the cost of emitted CO₂, chosen as 90 USD per ton according to the EU emission-trading scheme [80], and q_r is a constant that ensures the non-negativity and comparability of the valuation (see below). From Eq. (8.43), we can determine the valuation parameters α_{rn2} , describing the valuation of ISP n 's clean-energy share on path r , as $p_{CO_2} \cdot c_{rn,\max}$. The base valuation α''_{r0} is determined based on two considerations. First, the valuation function v_{r2} must be consistently non-negative. Second, the valuation function must allow a meaningful comparison between paths $R(n_1, n_2)$ connecting the same AS pair (n_1, n_2) : For example, if all ISPs use zero clean energy, a path with higher energy intensity should still be valued less than a path with lower energy intensity. Conversely, if all ISPs use perfectly clean energy, all paths should be valued identically. To achieve these properties, we determine α''_{r0} as follows:

$$\begin{aligned} \alpha''_{r0} &= - \sum_{n \in r} p_{CO_2} \cdot c_{rn,\max} + q_r \\ &= - \sum_{n \in r} p_{CO_2} \cdot c_{rn,\max} + \max_{\substack{r' \in R(n_1, n_2) \\ r \in R(n_1, n_2)}} \sum_{n' \in r'} p_{CO_2} \cdot c_{r'n',\max}. \end{aligned} \quad (8.44)$$

With such determined v_{r2} , we can formalize the complete path-valuation v_r as the sum of the attribute-specific valuation functions v_{r1} and v_{r2} . Since $\alpha_{rnk} a_{nk}$ yields a valuation per bit for both attributes $k \in \{1, 2\}$, the attribute-specific valuation functions are compatible.

8.4.3.3 Cost

To estimate the costs associated with the clean-energy share of an ISP n , we rely on the analysis of the levelized cost of energy (LCOE) of different electricity-generation technologies, performed by Lazard [158]. According to the Lazard analysis, electricity from low-carbon sources (solar, wind, nuclear) is on average $g = 3.375$ USD per MWh more expensive than electricity from high-carbon sources (coal, gas). This cost penalty, together with the average energy intensity of all intra-AS paths in AS n , yields the parameter ϕ_{n2} relevant for demand-dependent cost:

$$\phi_{n2} = g \cdot \frac{1}{|R(n)|} \cdot \sum_{r \in R(n)} e_{rn}, \quad (8.45)$$

where $R(n) = \{r \in R \mid n \in R\}$. Multiplied with the clean-energy share attribute a_{n2} , the parameter ϕ_{n2} yields the extra cost per transported bit that AS n incurs by using clean energy.

Regarding demand-independent cost, we note that the idle-power requirement of network devices plays an important role, as this requirement generates electricity bills even in absence of demand. The idle-power consumption u_n of a complete AS n can be estimated from the number of devices in AS n [56], the power consumption of network devices [116], and an average idle-power requirement of 85% [129]. This idle-power consumption u_n , together with the extra cost g for clean energy, determines the parameter γ_{n2} relevant for demand-independent cost: $\gamma_{n2} = g \cdot u_n$.

8.4.4 Attribute-Independent Parameters

In addition to the attribute-specific parameters in Sections 8.4.2 and 8.4.3, the attribute-independent parameters ρ_n , ϕ_{n0} and γ_{n0} also appear in our model.

The parameter ρ_n quantifies the revenue per transported bit of AS n . To estimate this parameter, we use a top-down approach: We divide the global annual revenue of wholesale Internet backbone providers (45.2 billion USD in 2019 [256]) by the amount of global annual Internet traffic (433 exabyte in 2019 [178]), and arrive at an average revenue of $\rho = 0.104$ USD per GB. For simplicity, we use this ρ as revenue parameter ρ_n for every ISP $n \in N$.

The parameter ϕ_{n0} describes the marginal cost of AS n per transported bit, excluding extra marginal cost due to clean-energy usage (cf. §8.4.3). As this marginal cost is commonly understood to be ‘essentially zero’ [250], we determine $\phi_{n0} = 0 \forall n \in N$.

Conversely, the attribute-independent fixed cost γ_{n0} of AS n can be quite substantial. However, since we are mainly interested in the attribute-optimization behavior of ASes under competition, and γ_{n0} does not affect this optimization behavior, we abstain from estimating γ_{n0} , i.e., use $\gamma_{n0} = 0$ in our simulations. As a result, the absolute value of the profit function π_n is not meaningful, which we take into account for the result discussion in §8.5.2.

8.5 SIMULATION

Section 8.3 theoretically illustrates the diverse results of quality competition among ISPs. These results include both positive and negative effects of competition on attribute prevalence and profits, depending on the concrete topologies of competing paths in the considered markets. In this section, we investigate which types of effect are observable if competition is introduced in a large-scale topology where transit ASes (autonomous systems, corresponding to ISPs) simultaneously compete in multiple markets. To that end, we run simulation experiments described in §8.5.1 for the instance of the competition model constructed in §8.4, and discuss the results in §8.5.2.

8.5.1 Experiments

Since the parameters estimated in §8.4 are afflicted with considerable uncertainty, we conduct our experiments without being overly reliant on the exact estimated parameter values. More precisely, we generate 10 different sets of model parameters by randomly modifying each model parameter y such that it lies between 0 and $2y$ in virtually all cases. We achieve this modification by randomly sampling each model parameter y' from the following restricted normal distribution, based on the corresponding estimated parameter y :

$$y' \sim \max \left(\mathcal{N} \left(y, \frac{y^2}{9} \right), 0 \right). \quad (8.46)$$

For each random sample of parameters, we investigate the effect of increasing *competition intensity* on the attribute-value choice of transit ASes. In our experiments, the intensity of competition corresponds to the number of usable paths between any AS pair. This number of usable paths is varied between 1 and 5 across experiments, where the case of 1 path corresponds to a monopoly scenario. In each simulation experiment, we thus simulate the competitive dynamics given a set of randomly sampled model parameters and given a certain intensity of competition.

Each simulation experiment amounts to computing round-robin better-response dynamics [44], where all ASes consecutively adjust their attribute values in the direction which increases their profit. This discrete process is an approximation of the continuous ODE process in Eq. (8.23). Moreover, the process can be understood as reflecting bounded rationality [236], as we assume that ASes can only identify profit-improving rather than profit-optimal attribute values. The simulation is terminated once the competitive dynamics have converged upon an equilibrium, i.e., each round only causes negligible change in the attribute values \mathbf{A} . The attribute values \mathbf{A}^+ in the equilibrium then represent the results of the experiment.

8.5.2 Results

The results of the experiments described in §8.5.1 are visualized in Figs. 8.2–8.4. The error bar of any data point in these figures illustrates the variance of the respective aggregate result across the 10 random parameter samples. Interestingly, the variance of the aggregate results is limited, although the variance in individual parameters is considerable. This observation indicates that our results are not highly sensitive to the model parameters from §8.4, and suggests that an approximate estimation of model parameters might be sufficient to yield useful predictions.

The central question in our analysis is: How does the intensity of competition affect the attribute values and the ISP profits? Our theoretical investigation in §8.3 indicates that the competitive dynamics can both increase and decrease these indicators compared to a monopoly scenario, depending on network properties. Hence, we investigate which type of effect is predominant for the large-scale network from §8.4.

In this analysis, we distinguish three groups of ASes that differ in their topology rank, namely tier-1 ASes (ASes that have no provider), tier-2 ASes (ASes that have only tier-1 providers), and tier-3 ASes (ASes that have only tier-1 and tier-2 providers). Since these AS groups differ in their market power, the effect on competition on attributes and profits for these groups may be different.

8.5.2.1 Attribute Prevalence

Regarding competition effects on attribute prevalence, we observe the following:

Effects on transit ASes. Fig. 8.2 confirms that an increasing number of options for the path-selecting ASes intensifies competition between transit ASes, which then improve their attribute values in response. In particular, about half of ASes

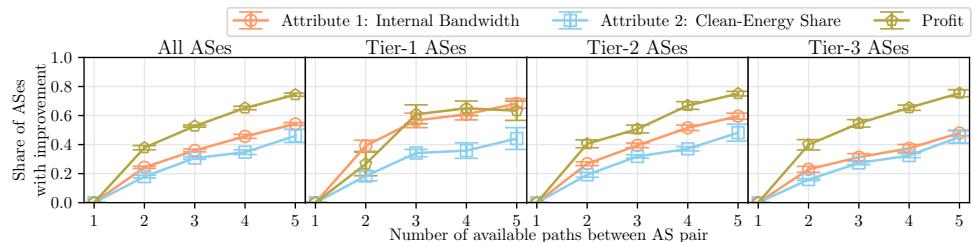


Figure 8.2: Under increasing competition, more ASes improve their attribute values or their profit compared to a single-path scenario, independent of their topology level.

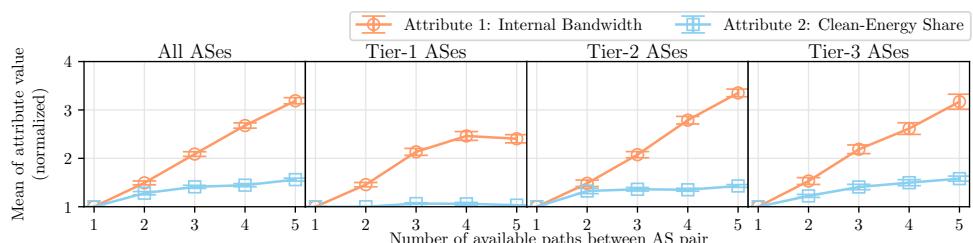


Figure 8.3: Competition tends to raise the mean attribute values across ASes (for all AS tier groups), compared to a monopoly situation. However, attributes may be affected differently by competition due to differences in valuation and cost.

improve both their attribute values given 5 available paths, whereas only 20% improve their attributes in a duopoly scenario (compared to a monopoly scenario). Note that some ASes may decrease their attribute values under competition for the counter-intuitive reasons explained in §8.3.3.

Moreover, the general level of attribute prevalence is raised by competition, which is implied by the increasing global average of attribute values in Fig. 8.3. The internal-bandwidth attribute is more strongly affected by this average gain than the clean-energy-share attribute, because (i) the bandwidth attribute is not upper-bounded, and (ii) the bandwidth attribute has zero demand-dependent cost.

Effects on path selectors. The average improvement in attributes translates into a more attractive offer for path selectors, i.e., source ASes and end-hosts. More precisely, the most attractive path between each AS pair tends to become more attractive as competition increases: Fig. 8.4 shows that 75% of AS pairs obtain access to a more valuable path if two paths instead of a single path are available (increasing maximum valuation), irrespective of the tier of the path-selecting AS.

Notably, we would expect that around 50% of AS pairs obtain a second path of higher quality in the absence of dynamic competition effects. Hence, the increasing maximum valuation is a combined effect of multi-path availability and competition.

Moreover, in absence of competition effects, a second path can only decrease, but not raise the minimum valuation across available paths. However, we observe that for 40% of AS pairs, both paths in a two-path scenario are more highly valued than the single path in a monopoly scenario, which suggests that competition raises the value of the previously monopolistic path. However, as the number of available paths increases, the tendency of additional paths to decrease the minimum quality becomes more visible.

Finally, these offer improvements materialize for all tiers of path-selecting ASes.

Differences in market power. Intriguingly, the higher market power of tier-1 ASes is not visible in Fig. 8.2, as tier-1 ASes are equally likely as lower-tier ASes to improve their attributes. However, the market power of tier-1 ASes becomes apparent in Fig. 8.3, which indicates that tier-1 ASes in competition improve their attributes to a lower extent than ASes on lower tiers.

8.5.2.2 Profits

Regarding competition effects on profits, we make the following observations:

Effects on transit ASes. Increasing path diversity and competition lead to increasing profits for a substantial share of ASes (cf. Fig. 8.2). At 5 available paths per AS pair, 75% of ASes increase their profits. This insight is surprising, given that competition is traditionally expected to increase consumer welfare and to reduce producer profits. In an ISP market, however, profits may increase because competition is modulated by path diversity. Such path diversity not only allows selecting ASes to select more different paths, but also allows transit ASes to attract and monetize traffic from more selecting ASes, increasing profit. Importantly, such an increase in attracted demand for an ISP does not necessarily reduce the attracted demand of another ISP, as the volume of total demand is elastic in our model.

Differences in market power. Interestingly, the profit increase under competition among 5 paths is more pronounced for tier-2 and tier-3 ASes than for tier-1 ASes. The reason is that the tier-1 ASes become more likely to be circumvented as path-selecting ASes obtain additional path options, and lower-tier ASes can thus attract more demand.

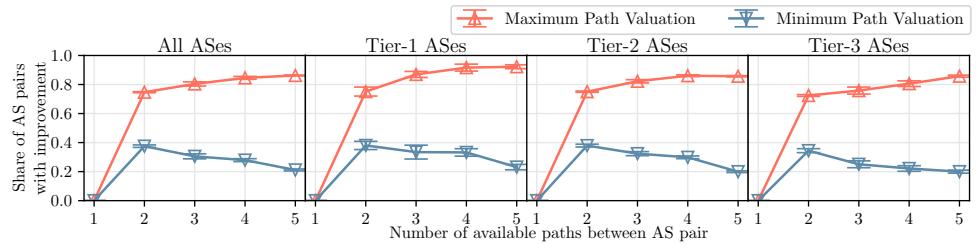


Figure 8.4: On average, competition raises the the attractiveness of the most *and* the least attractive path that connect two ASes, independent of the source-AS tier.

8.5.2.3 Sensitivity to Model Functions

In our model, we assume affine functions for both the path-valuation functions v_r and the cost functions Φ_n and Γ_n . To assess the impact of this assumption, we rerun the simulations by replacing linear dependencies within these functions. In particular, we replace a_{nk} in v_r by sub-linear $\sqrt{a_{nk}}$ (cf. Eq. (8.1)), and replace a_{nk} in Φ_n and Γ_n by super-linear a_{nk}^2 (cf. Eq. (8.4)). The results are presented in Fig. 8.5.

Intriguingly, the results for the non-affine functions closely match the results for affine functions in a qualitative sense, i.e., competition improves the attribute values, profits, and path options for a large share of ASes. Quantitatively, the largest difference concerns the mean increase in attribute values (cf. Fig. 8.5b vs. Fig. 8.3), which is considerably lower for the bandwidth attribute for the non-affine functions. However, this effect is to be expected because the non-affine functions make attributes both less valuable and more costly (if $a_{nk} > 1$, as for the bandwidth attribute), and thus less attractive to invest in.

8.6 RELATED WORK

General competition models. Internet competition is frequently studied by means of the three fundamental competition models from the economic literature. First, Cournot competition [66] describes multiple firms that produce the same homogeneous good, individually determine the quantity to be produced, and thus indirectly determine the market price. Cournot competition suggests that in comparison with a monopoly, a duopoly increases the quantity of the good and reduces its price, indicating that competition benefits consumers. In the second competition model, Bertrand competition [78], firms also produce a homogeneous good, but directly set a price instead of a quantity. Moreover, the firm with the lowest price acquires

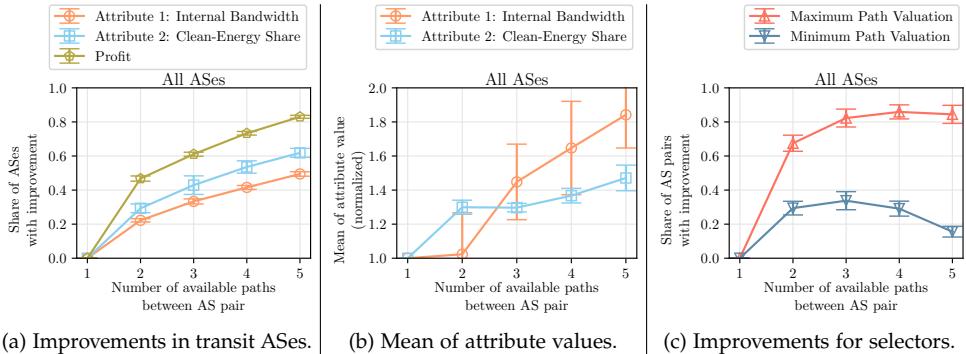


Figure 8.5: Equilibrium results for competitive dynamics with non-affine functions for valuation and cost.

the whole market. Hence, Bertrand competition is considered more suitable to analyze highly competitive markets. Lastly, Stackelberg competition [242] is similar to Cournot competition, but is suitable for hierarchical markets in which follower firms determine their production quantity after observing the quantity produced by a leader firm. All competition models have been adapted to *networked markets*, i.e., markets in which each segment of consumers can only be served by a corresponding subset of firms [23, 34, 187].

Our competition model is more strongly inspired by the logit-demand model [31], which originates from econometrics, can more directly represent competition between goods with different characteristics, and has been used in research on Internet transit pricing [247]. Still, the market in our model is networked, as determined by the network topology.

Internet competition models. To study Internet competition in particular, Shakkottai and Srikant [234] leverage Bertrand and Stackelberg competition to theoretically analyze the effect of competition in different levels of the Internet, i.e., for tier-1, tier-2, and local ISPs. Their model shows that competition may exert downwards pressure on prices, and an assimilating pressure on the quality-of-service (QoS) levels offered by different ISPs. These insights are confirmed by a subsequent line of simulation-based research [170, 244]. These studies analyze the ISP competition induced by virtualized access networks and by the ChoiceNet proposal [260], which describes a marketplace for transit services. Using both theoretical analysis and simulation, Nagurney and Wolf [192] expand on a study by Zhang et al. [270]

to investigate the intertwined competition dynamics among service providers (in Bertrand competition) and among network providers (in Cournot competition). In this analysis, the offers of service providers and network providers differ in both price and quality, and converge to an equilibrium describable by variational-inequality conditions.

In this chapter, we extend the previous work in a number of aspects. First, our model acknowledges that path quality may be determined collectively by multiple selfish on-path ISPs, and reveals the inefficient cooperation within a path due to selfishness. In contrast, previous work assumed that network service is bought from a single access/transit provider, and is thus effectively limited to one-hop paths. Second, path quality in our model depends on multiple underlying attributes, whereas previous work abstracts path quality in a single attribute. This fine-grained view of quality attributes is valuable, as it reveals how different attributes are affected differently by competition (cf. §8.3.3). Third, our model represents the internal cost structure of ISPs in a detailed manner, as it (i) distinguishes demand-dependent and demand-independent cost, and (ii) formalize the cost dependence on quality attributes, unlike previous work. Fourth, we make an effort to find realistic parameters for our large-scale simulations, whereas the parameters in previous simulation-based works are arbitrary. Finally, we explicitly investigate the differences between differing degrees of competition, and find network examples in which more intense competition leads to previously undocumented effects, namely increasing profits and decreasing path quality.

ISP cooperation. The economic dynamics between network entities that collectively provide connectivity has been studied with the lens of cooperative game theory [40], i.e., assuming that agents within a group choose rules which are enforced thereafter. Such considerations can inform financial settlement among ISPs in a coalition, where settlement mechanisms based on the Shapley value [169] or ISP characteristics [271] have been proposed. In our work, we discuss intra-path dynamics using non-cooperative game theory, as setting up a binding contract among the ISPs on every path is difficult in practice. Moreover, our model also reflects that multiple coalitions (paths) are in competition, which is missing from previous work.

8.7 CONCLUSION

ISPs determine the quality attributes of their connectivity offer (e.g., performance metrics, security features, sustainability properties) in line with their profit objective

and alternative offers by other ISPs. The presence of such alternative offers (i.e., competition) tends to improve path quality, as we demonstrate in this chapter. We provide evidence for this positive effect of ISP competition with an extensive theoretical analysis, based on a new game-theoretic model, and a large-scale simulation, based on empirical data. Our theoretical analysis suggests that an augmented path choice incentivizes transit ISPs to improve path quality, especially if ISPs have similar cost structures or traffic demand is high (Theorems 8.5 and 8.12). Interestingly, this higher investment in quality attributes does not necessarily reduce transit ISP profits, as entering competition (by connecting to new customers) also allows transit ISPs to attract revenue-generating traffic from new customer segments (Theorem 8.6). While these positive effects do not materialize in unfavorable circumstances (Theorem 8.13), our simulation-based case study indicates largely positive effects of competition in practice.

Our analysis does not only reveal the macroscopic effects of competition, but also formalizes the rational attribute choice for ISPs, which can inform ISP business strategies. In particular, we obtain three main recommendations for quality investment from our analysis:

- *Invest in attributes with low fixed cost:* Theorem 8.1 suggests that the optimal extent of a quality attribute correlates with the ratio of demand-dependent attribute-specific cost to total attribute-specific cost. Hence, the lower the demand-independent (fixed) cost of an attribute, the higher the optimal investment in the attribute. For example, renting internal bandwidth on-demand tends to improve profit more than a fixed bandwidth installation.
- *Invest exclusively in attributes with high return:* Theorem 8.1 also shows a correlation between the optimal attribute extent and the attribute return, i.e., the attribute-specific net revenue per traffic unit, divided by the attribute-specific cost. Theorem 8.10 even suggests *specialization* in heterogeneous markets with negligible demand-dependent cost, i.e., only the path attribute with the highest return should be invested in, while all less attractive attributes should be abandoned.
- *Engage in competition and on-path cooperation:* Both our theoretical analysis (Theorem 8.6) and our simulations (§8.5) show that engaging in competition by connecting to new customers tends to increase transit ISP profit, as the revenue from newly attracted traffic generally outweighs the costs of competing in the new markets. Furthermore, transit ISPs should also engage in cooperation with other on-path ISPs by coordinating attribute investment. Such coordination leads to achievement of the Nash bargaining solution, and

therefore to higher profits and higher path quality, which also benefits path users (Theorems 8.4 and 8.9). However, to achieve stable cooperation among selfish ISPs, additional work based on mechanism design is needed.

Part III

OUTLOOK AND CONCLUSION

9

FUTURE WORK

In this chapter, I discuss possibilities to extend and refine the research work in this thesis, oriented around the areas of network performance (§9.1) and ISP economics (§9.2).

9.1 NETWORK PERFORMANCE

The models and insights in Part i naturally raise a number of follow-up research questions:

Influence of network topology. In Chapter 2, the inefficiency of self-interested load-oriented path selection is investigated in two different topologies, namely a network of parallel links and a ladder topology. These two topologies are crucial examples, as they illustrate that detailed path-load information can have positive and negative effects on the global efficiency, respectively. Still, the analysis of these two topologies does not provide a rigorous understanding of the topology properties that determine the inefficiency of self-interested path selection and the effects of path-load information. Achieving such an understanding is an interesting task for future work, as topology properties that induce inefficient traffic distributions are valuable in predicting the performance effects of end-host path selection in the Internet.

Different path-selection behaviors. In both Chapters 2 and 3, path selection is studied under the assumption of rational agents that select paths exclusively based on path load. While this assumption is standard in the literature on selfish routing, future work might uncover relevant insights by modifying this assumption. For example, end-hosts may only possess bounded rationality, e.g., they may select the minimal-load path in a subset of available paths, or may show byzantine behavior, e.g.,

they may select random paths. Alternatively, end-hosts may also select paths based on load-independent properties, e.g., security-relevant or geographic properties. In this respect, it will be especially interesting to combine the enhanced selfish-routing models for end-host path selection from Chapters 2 and 3 with the quality-competition model from Chapter 8 that reflects the economic reaction of network operators to that path selection.

Path-selection performance beyond the worst case. Chapter 4 aims at identifying the performance damage of load oscillation in the worst-case, and therefore focuses on the worst-case scenario of greedy load-adaptive path selection between disjoint and identical paths with unsplittable traffic. While this worst case is relevant to obtain performance guarantees, the worst-case performance effects may be substantially more negative than the performance effects under *realistic* conditions. For example, the performance effects of end-host path selection would likely improve given stable path-selection strategies, latency-based congestion-control algorithms, or heterogeneous path options. These more friendly settings may be studied in future work to obtain insights into path-selection performance beyond the worst case.

Time bounds for congestion-control convergence. Both in Chapters 5 and 6, the convergence properties of BBR are investigated with a stability analysis based on fluid models. This stability analysis yields information regarding the existence of convergence and regarding the *asymptotic* convergence time, e.g., the fluid dynamics may converge exponentially fast. However, the stability analysis as performed in this work does not provide bounds on the convergence time in explicit terms, which would be valuable in practice. To estimate such a convergence, future work will benefit from additionally integrating concepts such as finite-time stability [32].

Combination of congestion-control models. While Chapters 5 and 6 demonstrates that fluid models are a powerful tool in congestion-control analysis, fluid models might benefit from combination with different models that compensate the abstraction of packet-level phenomena. As mentioned in Chapter 5, the notion of discrete packets is crucial to capture jitter, which is not captured by the smooth evolution curves present in fluid models. In contrast, models from queuing theory (e.g., M/M/1) represent packet arrivals as stochastic processes, and thus capture the fluctuation of the queue length that results in jitter. As such, queuing-models might be more suitable to analyze the performance improvements by the constant bit-rate traffic that is usable in presence of bandwidth-reservation systems. As these improvements are likely based on jitter reduction, combining queuing models and fluid models

is a promising approach for more comprehensive congestion-control modelling in future work.

Application-demand modelling. Chapters 5 and 6 of this thesis follow the entire literature on congestion control in assuming that the flows employing the congestion-control algorithms have infinite data to send. As a result, congestion-control algorithms are never application-limited in the academic analysis, although application limits determine to which extent the sending rate prescribed by congestion control is actualized. Therefore, application limits have significant performance implications, and would ideally also be integrated into congestion-control models. However, such an integration requires models for the evolution of application demand over time, which are challenging to create and may vary from application to application. Still, approximate application-demand models might be feasible for prominent application types, which is an interesting task for future work.

Further novel congestion-control algorithms. This thesis is motivated in part by the emergence of the novel congestion-control algorithms (CCAs) BBR, PCC, and Copa. All these algorithms are based on the idea of latency-sensitive congestion control which does not yield to competing flows employing a loss-based CCA (unlike earlier latency-sensitive CCAs such as Vegas). Since these algorithms thus share a fundamental similarity, this thesis focuses on BBR as an example of this new generation of CCAs. Still, BBR is different from PCC and Copa in its mechanisms; therefore, the insights into BBR performance likely do not fully apply to the other two algorithms. Hence, it will be interesting to investigate also PCC and Copa with the fluid-model approach refined in this thesis.

9.2 ISP ECONOMICS

Part ii presents my model-based work on ISP economics, which could be extended in the following ways:

Path-aware agreements with multiple objectives. Chapter 7 demonstrates how to structure interconnection agreements in path-aware Internet architectures in order to pursue the main ISP goal of profit optimization. However, ISPs have secondary goals beyond profit generation, e.g., to maximize link utilization, to attain quality-of-service (QoS) benchmarks, or to enable secure communication. Satisfying these objectives might require further clauses to interconnection agreements, going beyond the flow-volume targets and the monetary compensation proposed in Chapter 7.

Designing such agreement conditions related to non-profit objectives is an exciting avenue for future work.

Assessment of different competition models. As Chapter 8 mentions, the model for quality competition in the chapter departs from previous competition models with a more fine-grained formalization of quality attributes, ISP cost structures, and the hybrid cooperation-competition environment in the Internet. While this detailed modelling plausibly increases the accuracy of model predictions, rigorously assessing the predictive accuracy of the various competition models is a complex endeavor in itself. Notably, ISP competition models predict the prices, quality offers, and profits of ISPs based on information about the network topology and costs. Hence, if model predictions should be compared to ground truth, ground-truth information about actual transit prices, transmission quality, and ISP profits is required; however, such information is not available in a systematic and detailed form. This limited information requires future work on model-validation approaches that require little ground-truth data.

Generalization of competition model. Interestingly, the new competition model in Chapter 8 is not only applicable to ISP competition, but in general to settings in which coalitions of selfish entities stand in competition. While ISPs and paths represent the entities and coalitions in the ISP market, competition in other markets arises between firms that form value chains. Our model allows investigating complex economic phenomena such as the interaction between firms along a value chain, or the effect of overlapping value chains. Hence, it is an intriguing opportunity for future work to investigate whether the findings for the ISP market in Chapter 8 translate to other economic sectors.

10

CONCLUSION

To conclude this thesis, I revisit the research questions initially presented in §1.2, and formulate an answer to these questions based on the implications of my doctoral research. These answers are again organized by the topic areas of network performance (§10.1), and ISP economics (§10.2). The thesis is then concluded by some final remarks in §10.3.

10.1 NETWORK PERFORMANCE

Regarding network performance, the introduction poses the following research questions:

Q 1 *Path-Selection Efficiency.* How efficient are the traffic distributions across networks paths that can be expected from path selection by self-interested and uncoordinated end-hosts?

As Chapter 2 demonstrates, the efficiency of traffic-distribution equilibria produced by self-interested path selection depends both on the perspective and on the characteristics of the network. Regarding the perspective, end-hosts and network operators are affected by traffic distributions in different ways. Hence, end-hosts and network operators experience a different Price of Anarchy, i.e., the different stakeholders rate the efficiency of traffic-distribution equilibria differently when comparing to the stakeholder-specific social optimum.

However, while the perspectives of end-hosts and network operators are not identical, the findings in Chapter 2 suggest that the interests of the stakeholder groups are well-aligned, because both groups are similarly affected by changes in network characteristics. For example, if the traffic-distribution efficiency improves for end-hosts thanks to more detailed load information about paths, network operators also experience an efficiency improvement. This similarity also applies to networks

where more detailed load information surprisingly induces less efficient traffic equilibria than the basic latency information obtainable from RTT measurements. Moreover, end-hosts and network operators are also similarly affected by another network property, namely the scale of the network: In large-scale networks such as the Internet, the efficiency induced by detailed path-load information approaches the efficiency induced by basic latency information, for both end-hosts and network operators. Fortunately, Chapter 2 indicates that the efficiency associated with this basic latency information is near-optimal in realistic network topologies, as confirmed by both theoretical worst-case bounds and simulation results.

However, this equilibrium-based efficiency is only relevant under the condition that the path-selection dynamics in fact converge to traffic-distribution equilibria. This caveat underlines the relevance of the next research question:

Q 2 Path-Selection Stability. *Under which conditions does path selection by self-interested and uncoordinated end-hosts converge to steady traffic distributions?*

This stability question is answered by the new notion of equilibria on path-selection strategies (PSS equilibria), introduced in Chapter 3: A path-selection strategy, which prescribes path selection over time based on network observations, is adopted by a self-interested end-host only if no other path-selection strategy allows the end-host to improve performance, given the path-selection strategies adopted by all other end-hosts. Hence, if a path-selection strategy is (i) such a PSS equilibrium strategy under universal adoption, and (ii) a stable strategy that produces convergence to the equilibrium, then stable path selection can be expected from self-interested end-hosts. However, the game-theoretic analysis in Chapter 3 demonstrates that the stable path-selection strategies, which have been proposed in previous research to achieve convergence under outdated load information, are not rational strategies under universal adoption.

In a narrow game-theoretic understanding, path selection by self-interested end-hosts is thus bound to produce load oscillation, unless network operators take active measures to stabilize path selection. However, these active measures introduce overhead, which must be compared to the performance cost of the load oscillation itself. This insight naturally begets the next research question:

Q 3 Performance under Unstable Path Selection. *How does unstable end-host path selection affect the efficiency and fairness in networks?*

This question is tackled in Chapter 4 with an axiomatic method. This method allows to analytically rate the performance of congestion-control algorithms with

and without coupled path selection, where the path selection is greedy and therefore unstable. This comparison-based approach reveals that the performance effects of unstable path selection significantly depend on the specifics of the congestion-control and path-selection routines. For example, highly responsive path selection, which quickly switches to a supposedly more attractive path, can reduce link utilization and increase packet loss compared to no path selection at all; in contrast, moderately responsive path selection actually improves these metrics compared to a path monopoly. Conversely, highly responsive path selection is preferable to moderately responsive path selection in terms of fairness. Moreover, these performance effects depend on the congestion-window adaptation on path switch: If the size of the congestion window is significantly reduced when switching paths, packet loss is reduced, but link utilization suffers.

Hence, unstable end-host path selection does not necessarily reduce any single performance indicator compared to path pre-determination, but leads to (i) undesirable trade-offs between performance objectives, and (ii) performance unpredictability for network operators who may not know the specific parameters underlying end-host path selection. Therefore, avoiding instability in path selection appears to be a worthwhile goal, which is pursued by the next research question:

Q 4 Path-Selection Shaping. *How can network operators inform or incentivize selfish end-hosts to perform path selection in an efficient and stable manner?*

Judging from the previous research questions, this research question is answered differently for the goals of efficiency and stability.

Efficiency. Regarding efficiency, the findings related to Q 1 suggest that the efficiency of self-interested path selection cannot be substantially improved by the provision of detailed path-load information, at least in Internet-scale networks. However, the urgency of improving this efficiency is also quite low, as self-interested path selection possesses equilibrium traffic distributions with near-optimal efficiency. To close this small efficiency gap, a more effective tool than information provision might be congestion pricing [232]. However, given that traffic distributions already tend to be near-optimal without pricing, any efficiency benefits of such smart data pricing are limited, and will thus likely be considered insufficient to tackle the practical and economic challenges that have traditionally held back price-based traffic shaping [200].

Stability. Regarding stability, the answer to Q 2 suggests that stability is unlikely to spontaneously arise in self-interested path selection. Instead, stability can be

incentivized with the incentive-compatible stabilization mechanisms from Chapter 3. These stabilization mechanisms suppress load oscillation by presenting rewards or imposing costly requirements for using certain ingresses of autonomous systems (ASes). These incentivization mechanisms are designed with limited overhead to be applicable under the high traffic speeds in the Internet core.

While such stabilization mechanisms are thus capable to suppress oscillation, their necessity can be questioned from two angles.

Rationality assumption. First, stabilization mechanisms are only necessary if self-interested end-hosts do not adopt stable, but irrational path-selection strategies. This rationality assumption is central to the game-theoretic perspective, but might be considered overly pessimistic in practice. This skepticism is supported by the observation that end-hosts in various areas of networking comply with prescribed actions that are desirable but irrational from an individual perspective, e.g., in congestion control [16] or file-sharing protocols [165]. Indeed, the rationality of users or application developers might be bounded in practice, as deviating to more attractive actions is prevented by reasons that are not commonly accommodated in game-theoretic models, e.g., inertia, incompetence, or idealism. However, the limited understanding of these factors, which violate rationality in a narrow sense, introduces considerable risk when relying on these factors. Notably, non-compliance for self-interest is definitely observable in the Internet, e.g., in the tendency of web browsers to open multiple flows to circumvent per-flow fairness [41] or in the increased aggressiveness of vendor-specific congestion-control implementations [179]. In this regard, the rationality assumption of game-theoretic models is a worst-case assumption, and incentive-compatible stabilization mechanisms are a uniquely viable tool in a worst case.

Intra-domain adaptation. Second, stabilization mechanisms would not be required if suppressing oscillation was more costly than only alleviating the consequences of said oscillation. More precisely, instead of avoiding that traffic oscillates between AS ingress interfaces, an AS might employ traffic engineering to dynamically rebalance the traffic flows across its internal network infrastructure [84, 177]. Thereby, an AS would continuously adapt to an oscillating traffic matrix rather than shape this traffic matrix. However, dynamic rebalancing by traffic engineering is limited by both its response time and by the capacity of paths through the internal network. Hence, intra-domain traffic engineering likely benefits from being complemented by traffic-steering mechanisms that stabilize the traffic matrix.

While strategies for path selection thus play an important role for network performance, strategies for the sending-rate adaptation, i.e., congestion-control algorithms are equally essential. This relevance motivates the following research questions:

Q 5 Congestion-Control Performance. *How efficient, fair, and stable are the new congestion-control algorithms?*

Q 6 Congestion-Control Methodology. *What kind of evaluation methods can provide reliable insights into CCA behavior under the highly diverse conditions found in the Internet?*

To investigate Q 6, this thesis explores and applies three different evaluation methods for congestion control. First, Chapter 4 describes congestion-control dynamics (with and without path selection) as a discrete-time process. Second, Chapter 5 and Chapter 6 describe congestion-control dynamics by means of fluid models. Third, Chapters 5 and 6 also use experimental evaluations based on network emulation, mostly to validate the insights from the fluid models.

Each of these evaluation methods has strengths and shortcomings. For one, the discrete-time model is straightforward and allows theoretical analysis, in particular the identification of dynamic equilibria. However, the discrete-time assumption in the model is artificial, the feedback in the model is coarse-grained (e.g., binary packet-loss indicator per time step), and queuing dynamics are neglected. All of these shortcomings are remedied by the continuous-time fluid model, which also allows theoretical stability analysis and provides a basis for efficient simulation. However, the fluid model introduces higher complexity than the discrete-time model, and still cannot represent packet-level phenomena such as jitter due to the smoothness of fluid traffic. The full spectrum of congestion-control performance, including packet-level phenomena, is of course revealed by experimental evaluations. Unfortunately, experimental evaluations are expensive to conduct for Internet conditions, as the necessary test-bed must include a large-scale topology and high-throughput hardware.

Among these methods, this thesis applies a combination of fluid modelling and small-scale experimental validation to answer Q 5, i.e., to evaluate the new congestion-control algorithms (CCAs). In particular, this combination of techniques yields robust and relevant insights into the behavior of the BBR algorithm. On the one hand, some of these insights confirm previous research, e.g., regarding the high loss caused by BBRv1 in shallow buffers, the resulting unfairness towards

loss-based CCAs, and the successful elimination of this issue by BBRv2 under a drop-tail queuing discipline. On the other hand, some insights yielded by fluid modelling are previously undocumented, e.g., the unfairness of BBRv2 towards loss-based CCAs under RED, buffer-size-dependent bufferbloat behavior by BBRv2, and the rate oscillation in BBRv1 when competing with loss-based CCAs. All these insights are not only revealed by fluid-model simulation and confirmed by experiments, but also rigorously proven by means of theoretical stability analysis.

These observations allow two conclusions. First, BBR – as the most prominent example of the new CCAs – is not free from performance issues, but is continuously improved by new releases. This ongoing refinement is likely to continue, as evidenced by the release of BBRv3, which is imminent at the time of writing [52]. Second, fluid models are a powerful yet under-explored tool for analyzing the performance of contemporary CCAs, as these models allow both accurate simulation and theoretical stability analysis.

10.2 ISP ECONOMICS

Regarding ISP economics, Chapter 1 poses the following research question:

Q 7 *Interconnection Agreements.* *How should interconnection agreements among ISPs be structured and negotiated under end-host path selection in the Internet?*

As described in Chapter 7, ISP interconnection agreements benefit from relaxed technical constraints if end-host path selection is implemented through path-aware networking. More precisely, if forwarding directives in data packets eliminate the need for BGP convergence, ISP interconnection agreements can deviate from the Gao-Rexford conditions (GRC), which are necessary for BGP convergence. To make such GRC-violating agreements acceptable to self-interested ASes, such agreements might additionally include restrictions on exchanged traffic volumes or procedures for monetary compensation, as some peering agreements today already do. If such additional conditions can make these agreements viable, Internet path diversity benefits to an enormous extent, as our topology analysis in Chapter 7 shows: When introducing mutuality agreements in a BGP-based Internet, the median AS gains access to three times the number of short paths (with 3 AS hops), 50% of ASes gain a path with lower latency, and 35% of ASes gain a path with higher bandwidth.

While ISP interconnection agreements thus benefit from new possibilities in path-aware networks, the negotiation of agreements based on path awareness is complicated by a game-theoretic issue that already hampers ISP negotiation today.

In particular, ISP agreements can only be optimized with respect to fairness and efficiency if the revenue structure and the cost structure of the involved ISPs are known. Since this information is private, an ISP can strategically misrepresent how an agreement would affect its revenue and cost, and thereby aim at agreement terms that are more favorable from an individual perspective, but inefficient and unfair from a global perspective. This problem can be remedied by our automated bargaining mechanism BOSCO, which intelligently restricts the offers that the negotiating parties can make.

While this exploration of new interconnection opportunities in Chapter 7 assumes a path-aware Internet, the analysis also provides relevant insights on today's Internet. First, the topology analysis in Chapter 7 quantifies the cost of the Gao-Rexford conditions in terms of path diversity. Second, the BOSCO bargaining mechanism is not only applicable to novel agreements in path-aware networks, but can also be applied to classic interconnection agreements in today's Internet, which also confront the same game-theoretic challenge of private information.

Still, the study of interconnection agreements in Chapter 7 covers only one aspect of ISP cooperation, namely the collaborative construction of paths by ISPs. Once these paths are announced, ISPs also need to contribute to the transmission quality of the paths together with other on-path ISPs, especially as the provided paths stand in competition with other paths. This hybrid environment of cooperation and competition gives rise to the last two research questions:

Q 8 Intra-Path Cooperation. *How is the quality of Internet paths affected by cooperation among self-interested on-path ISPs if end-hosts select paths based on quality?*

Q 9 Inter-Path Competition. *How is the quality of Internet paths affected by ISP competition if end-hosts select paths based on quality?*

These research questions are tackled in Chapter 8, which proposes a game-theoretic model of ISP quality competition that is applicable under path selection by both ISPs (as in BGP) and end-hosts. This model captures how ISPs adapt their quality attributes to optimize profit in path-based competition, where quality attributes are understood broadly as performance, security, environmental, and economic ISP properties. Notably, ISPs can affect each other's profit with their quality-attribute investments both positively (if the ISPs are on a shared path) and negatively (if the ISPs are on competing paths). This distributed optimization

by interdependent ISPs leads to a number of interesting phenomena that are revealed by the model through rigorous proofs. First, competition between paths has an improving effect on quality attributes if ISPs have similar cost structures or demand is sufficiently high, but may have a negative effect if the competing ISPs have sufficiently different characteristics. Second, under common conditions, ISPs optimize their profit by exclusively investing in the single quality attribute with the best cost-benefit ratio. Third, the cooperation among ISPs on a shared path suffers from a prisoner's dilemma: If all on-path ISPs make the attribute investments that optimize the aggregate profit on the path, some ISPs can increase their profit by reducing their attribute investment. As a result, the equilibrium attribute investments are lower than the profit-optimal investments, which not only causes sub-optimal global profit, but also damages the users of the path, i.e., end-hosts.

In summary, the analysis from Chapter 8 focuses on the quality competition among ISPs under path selection, independently of whether the path selection is performed by ISPs or by end-hosts. Still, the model can capture the increasingly prevalent end-host path selection through appropriate parameter selection. First, the model includes valuation parameters, which determine how strongly any quality attribute is valued by the path-selecting entities, i.e., how strongly the attribute attracts traffic. These valuation parameters are likely different if paths are selected by end-hosts rather than ISPs. Second, if end-host path selection is enabled in a strong form by path-aware networks, the different mode of path discovery substantially increases the number of competing paths between any two ASes. Since the number and the composition of paths are represented as model parameters, a shift to path-aware networking can be accommodated by changes to the underlying topology in the model.

10.3 FINAL REMARKS

This thesis presents an analysis of the impact of two contemporary Internet trends, namely end-host path selection and novel congestion-control algorithms, on traffic patterns in the Internet. This shift in traffic patterns has numerous effects on network performance and ISP economics, where these effects may be subtle and even counter-intuitive. In this thesis, many of these effects are identified, characterized, and quantified with a model-based approach, which provides a unique perspective and mathematical guarantees in all investigated areas. In particular, my modelling underlines the relevance of (i) inefficient equilibria and (ii) stability issues of

networked dynamics, which play an essential role for both network performance and ISP economics in a future Internet.

Part IV
APPENDIX

A

APPENDIX OF CHAPTER 2

A.1 PARALLEL LINKS: END-HOST OPTIMUM

In the end-host optimum \mathbf{F}^* , it holds that for every end-host e_k , $k \in \{1, \dots, K\}$, the marginal cost of the path over the link β is equal to the marginal cost of any path over a link α_i , $i \in \{1, \dots, m\}$. Using this insight, the end-host optimum can be derived by the solution of the following equation:

$$\frac{\partial}{\partial F_{(k,D),\beta}} C^* = \frac{\partial}{\partial F_{(k,D),\alpha_i}} C^* \quad (\text{A.1})$$

As it holds that

$$\frac{\partial}{\partial F_{(k,D),\ell}} C^* = \frac{\partial}{\partial f_\ell} C^*, \quad (\text{A.2})$$

the optimal link-flow pattern \mathbf{f}^* is obtained by solving:

$$\frac{\partial}{\partial f_\beta} (f_\beta \cdot f_\beta^p) = \frac{\partial}{\partial f_{\alpha_i}} (f_{\alpha_i} \cdot d^p) \iff (p+1) \cdot f_\beta^p = d^p, \quad (\text{A.3})$$

which yields $f_\beta^* = d / \sqrt[p]{p+1}$.

Any path-flow pattern \mathbf{F}^* that produces a link-flow pattern with $f_\beta = f_\beta^*$ and $\sum_{\alpha_i} f_{\alpha_i} = d - f_\beta^*$ is thus optimal from the perspective of end-hosts. The total cost of such an optimal path-flow pattern \mathbf{F}^* is given by

$$C^*(\mathbf{F}^*) = f_\beta^* \cdot f_\beta^{*p} + (d - f_\beta^*) \cdot d^p = d^{p+1} \left(1 - p/(p+1)^{(p+1)/p} \right) \quad (\text{A.4})$$

A.2 PARALLEL LINKS: PI EQUILIBRIUM

In the network from Figure 2.4, the selfish cost function of an end-host e_k can be simplified to

$$C_{(e_k)}^*(\mathbf{F}) = F_{(k,D),\beta} \cdot f_\beta^p + \sum_{\alpha_i} F_{(k,D),\alpha_i} \cdot d^p \quad (\text{A.5})$$

Since it holds that $\partial/\partial F_{(k,D),\beta}$ $f_\beta = 1$, the marginal selfish costs for the paths over link β and α_i are given by

$$\begin{aligned} \partial/\partial F_{(k,D),\beta} C_{(e_k)}^*(\mathbf{F}) &= f_\beta^p + F_{(k,D),\beta} \cdot p \cdot f_\beta^{p-1} \\ \partial/\partial F_{(k,D),\alpha_i} C_{(e_k)}^*(\mathbf{F}) &= d^p. \end{aligned} \quad (\text{A.6})$$

An equilibrium under the PI assumption is characterized by the equality of these selfish marginal costs.

Note that for every end-host e_k , the marginal selfish cost of every path over a link α_i is the same, namely d^p . By marginal selfish cost equality, the marginal selfish costs $\partial/\partial F_{(k,D),\beta}(C_{(e_k)}^*)$ must be equal to d^p for every end-host e_k and thus also equal across all end-hosts. This condition is only satisfied if every end-host e_k has the same flow on the path over link β , i.e., $F_{(m,D),\beta} = F_{(n,D),\beta}$ for all end-hosts e_m, e_n . As $f_\beta = \sum_{e_k} F_{(k,D),\beta}$, the fact $\sum_{e_k} F_{(k,D),\beta} = K \cdot F_{(k,D),\beta}$ implies $F_{(k,D),\beta} = f_\beta/K$ for every end-host e_k .

This knowledge about $F_{(k,D),\beta}$ allows to simplify the selfish marginal cost equation and to compute the PI equilibrium. By inserting f_β/K for $F_{(k,D),\beta}$, the selfish marginal cost equality reads

$$f_\beta^p + f_\beta/K \cdot p \cdot f_\beta^{p-1} = d^p. \quad (\text{A.7})$$

This equation yields the PI equilibrium link flow $f_\beta^+ = d / \sqrt[p]{p/K+1}$. The PI equilibrium \mathbf{F}^+ is thus given by every path-flow pattern that satisfies the following conditions for every end-host e_k :

$$F_{(k,D),\beta} = f_\beta^+ / K \quad \wedge \quad \sum_{\alpha_i} F_{(k,D),\alpha_i} = (d - f_\beta^+) / K. \quad (\text{A.8})$$

The cost term C^* of the PI equilibrium \mathbf{F}^+ to end-hosts is thus

$$C^*(\mathbf{F}^+) = d^{p+1} \left(1 - (p/K) / (p/K + 1)^{(p+1)/p} \right). \quad (\text{A.9})$$

A.3 LADDER NETWORK: PI EQUILIBRIUM

In order to compute the Price of Anarchy under the PI assumption for a general ladder network, we start by computing the Price of Anarchy for the simple ladder network of $H = 2$. Conforming to the PI equilibrium conditions in Definition 2.3, the PI equilibrium is given by the solution of the following equilibrium equation system \mathcal{E}_2 that formalizes the marginal cost equality:

$$\begin{aligned} \mathcal{E}_2 &= \begin{cases} \partial/\partial F'_1 \left(F'_1 \cdot f_{h_1}^p \right) = \partial/\partial F_1 \left(F_1 \cdot \left(f_{h_2}^p + t(f_{v_{11}} + f_{v_{12}}) \right) \right) \\ \partial/\partial F'_2 \left(F'_2 \cdot f_{h_2}^p \right) = \partial/\partial F_2 \left(F_2 \cdot \left(f_{h_1}^p + t(f_{v_{11}} + f_{v_{12}}) \right) \right) \end{cases} \\ &= \begin{cases} f_{h_1}^p + F'_1 p f_{h_1}^{p-1} = f_{h_2}^p + t(f_{v_{11}} + f_{v_{12}}) + F_1(p \cdot f_{h_2} + 2t) \\ f_{h_2}^p + F'_2 p f_{h_2}^{p-1} = f_{h_1}^p + t(f_{v_{11}} + f_{v_{12}}) + F_2(p \cdot f_{h_1} + 2t) \end{cases} \end{aligned} \quad (\text{A.10})$$

where $F'_1 = F_{(11,12),h_1}$ and $F_1 = F_{(11,12),v_{11}h_2v_{12}}$ and the analogous abbreviations have been made by F'_2 and F_2 for the direct and indirect path flow of end-host e_{21} . Due to demand constraints, it holds that $F'_1 = d - F_1$ and $F'_2 = d - F_2$. Due to the symmetry of the equation system, it is possible to conclude that $F'_1 = F'_2$ and $F_1 = F_2$. Since $f_{h_1} = F'_1 + F_2$ and $f_{h_2} = F'_2 + F_1$, we obtain

$$f_{h_1} = F'_1 + F_2 \stackrel{\text{Demand}}{=} F'_1 + d - F'_2 \stackrel{\text{Symmetry}}{=} F'_2 + d - F'_1 = d, \quad (\text{A.11})$$

and a symmetric derivation for $f_{h_2} = d$. Furthermore, the flow on the vertical links v_{11} and v_{12} can be expressed as follows: $f_{v_{11}} = f_{v_{22}} = F_1 + F_2 = 2F_1$. The equilibrium equation system \mathcal{E}_2 can thus be reduced to the single equation:

$$\begin{aligned} d^p + (d - F_1) \cdot \lambda &= d^p + 2t \cdot (2F_1) + F_1 \cdot (\lambda + 2t) \\ \iff \lambda d - (6t + 2\lambda)F_1 &= 0 \iff F_1 = \frac{\lambda d}{6t + 2\lambda} \end{aligned} \quad (\text{A.12})$$

where $\lambda = p \cdot d^{p-1}$.

Based on this solution for the path flow $F_1 = F_{(11,12),v_{11}h_2v_{12}}$, all other path flows can be derived, which yields the following terms for the two perspectives on the Price of Anarchy:

$$\begin{aligned} PoA_{H=2}^{*+}(d, t, p) &= \frac{2 \left((d - F_1) \cdot f_{h_1}^p + F_1 \cdot \left(f_{h_2}^p + t \cdot (f_{v_{11}} + f_{v_{12}}) \right) \right)}{2d \cdot f_{h_1}} \\ &= \frac{(d - F_1) \cdot d^p + F_1 \cdot (d^p + t \cdot 4F_1)}{d \cdot d^p} \\ &= \frac{d^{p+1} + 4t(\lambda d / (6t + 2\lambda))^2}{d^{p+1}} \\ PoA_{H=2}^{\#+}(d, t, p) &= \frac{2 \left(f_{h_1}^p + t \cdot (f_{v_{11}} + f_{v_{12}}) \right)}{2 \cdot f_{h_1}^p} \\ &= \frac{d^p + t \cdot (4F_1)T}{d^p} = \frac{d^p + 4t\lambda d / (6t + 2\lambda)}{d^p} \end{aligned} \quad (\text{A.13})$$

The Price of Anarchy for all ladder networks with $H = 2$ is obtained by computing an upper bound in terms of demand d and parameter t :

$$PoA_{H=2}^{*+}(p) = 1 + p/12 \quad PoA_{H=2,\max}^{\#+}(p) = 1 + p/3 \quad (\text{A.14})$$

A.4 LADDER NETWORK: PROOF OF THEOREM 2.3

We start by observing that $PoA_{H,\max}^{\#+}$ is given by the limit in t and is only dependent on the flows on vertical links v :

$$PoA_{H,\max}^{\#+} = 1 + \lim_{t \rightarrow \infty} (t \cdot \sum_v f_v) / (H \cdot d^p), \quad (\text{A.15})$$

where we used that $\lim_{t \rightarrow \infty} \sum_h f_h^p = H \cdot d^p$ as vertical links become infinitely expensive. We need only characterize the sum of vertical link flows $f_V = \sum_v f_v$, for which we use an argument based on the structure of the equilibrium equation system \mathcal{E}_H .

For setting up \mathcal{E}_H , we consider Figure A.1 which illustrates numerically computed equilibria for some $H > 2$. The figure shows the traffic distribution on the horizontal links of a ladder network, where different-color flow shares correspond to flows of different end-hosts. Figure A.1 shows two insights that are relevant for setting up \mathcal{E}_H . First, the path-flow pattern \mathbf{F}^+ only contains non-zero flows on paths that deviate at most one ladder level from the originating end-host (for high enough t).

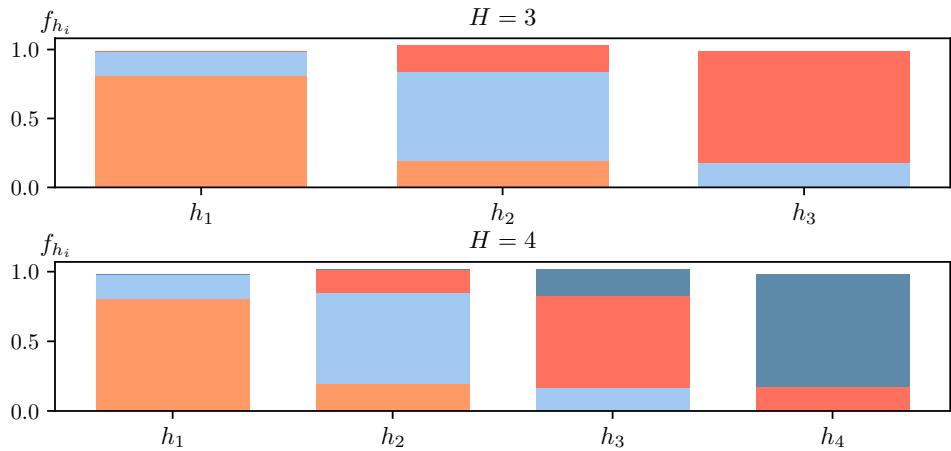


Figure A.1: Traffic distribution over horizontal links of a ladder network in PI equilibrium \mathbf{F}^+ (for $p = 2, t = 1, H = 3, 4$).

Second, the path-flow pattern is symmetric with respect to the horizontal axis of the ladder network. The variables in \mathcal{E}_H can thus be assigned to the indirect path flows as displayed in Figure A.2. Variable assignments for higher H work analogously to Figure A.2a (for odd H) and Figure A.2b (for even H).

With these variables and the knowledge about the equilibrium traffic distribution, the equation system \mathcal{E}_H can be set up for all values for H . In the following, we demonstrate the construction of the equation systems for $H = 3, H = 4$, and the general case.

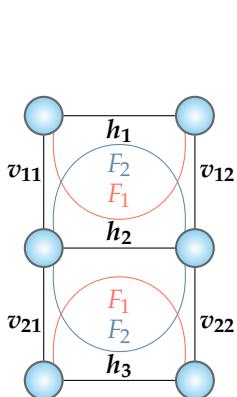
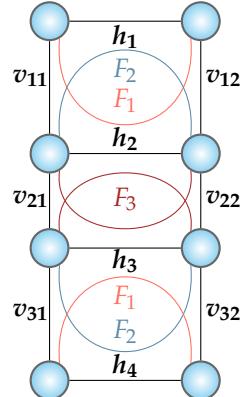
(a) $H = 3$ (b) $H = 4$

Figure A.2: Variable assignments to indirect path flows in a ladder network for computing the PI equilibrium.

$H = 3$. In this case, the PI equilibrium is given by the solution of the following equilibrium equation system \mathcal{E}_3 encoding the marginal cost equality:

$$\begin{aligned} \mathcal{E}_3 = & \begin{cases} \partial/\partial F'_1 \left(F'_1 \cdot f_{h_1}^p \right) = \partial/\partial F_1 \left(F_1 \cdot \left(f_{h_2}^p + t(f_{v11} + f_{v12}) \right) \right) \\ \partial/\partial F'_2 \left(F'_2 \cdot f_{h_2}^p \right) = \partial/\partial F_2 \left(F_2 \cdot \left(f_{h_1}^p + t(f_{v11} + f_{v12}) \right) \right) \end{cases} \\ = & \begin{cases} f_{h_1}^p + F'_1 p f_{h_1}^{p-1} = f_{h_2}^p + t(f_{v11} + f_{v12}) + F_1(p \cdot f_{h_2} + 2t) \\ f_{h_2}^p + F'_2 p f_{h_2}^{p-1} = f_{h_1}^p + t(f_{v11} + f_{v12}) + F_2(p \cdot f_{h_1} + 2t) \end{cases} \\ = & \begin{cases} d^p + (d - F_1) \lambda = d^p + 2t(F_1 + F_2) + F_1(\lambda + 2t) \\ d^p + (d - 2 \cdot F_2) \lambda = d^p + 2t(F_1 + F_2) + F_2(\lambda + 2t) \end{cases} \\ = & \begin{cases} \lambda d - (4t + 2\lambda) \cdot F_1 - 2t \cdot F_2 = 0 \\ \lambda d - (4t + 3\lambda) \cdot F_2 - 2t \cdot F_1 = 0 \end{cases} \quad (\text{A.16}) \end{aligned}$$

$H = 3$. In this case, the equilibrium equation system \mathcal{E}_4 is given by the following equation system, where the demand between end-hosts e_{21} and e_{22} is composed as $F_2 + F'_2 + F_3 = d$:

$$\begin{aligned} \mathcal{E}_4 &= \begin{cases} \partial/\partial F'_1 \left(F'_1 \cdot f_{h_1}^p \right) = \partial/\partial F_1 \left(F_1 \cdot \left(f_{h_2}^p + t(f_{v_{11}} + f_{v_{12}}) \right) \right) \\ \partial/\partial F'_2 \left(F'_2 \cdot f_{h_2}^p \right) = \partial/\partial F_2 \left(F_2 \cdot \left(f_{h_1}^p + t(f_{v_{11}} + f_{v_{12}}) \right) \right) \\ \partial/\partial F'_2 \left(F'_2 \cdot f_{h_2}^p \right) = \partial/\partial F_3 \left(F_3 \cdot \left(f_{h_3}^p + t(f_{v_{21}} + f_{v_{22}}) \right) \right) \end{cases} \\ &= \begin{cases} f_{h_1}^p + F'_1 p f_{h_1}^{p-1} = f_{h_2}^p + t(f_{v_{11}} + f_{v_{12}}) + F_1(p \cdot f_{h_2} + 2t) \\ f_{h_2}^p + F'_2 p f_{h_2}^{p-1} = f_{h_1}^p + t(f_{v_{11}} + f_{v_{12}}) + F_2(p \cdot f_{h_1} + 2t) \\ f_{h_2}^p + F'_2 p f_{h_2}^{p-1} = f_{h_3}^p + t(f_{v_{21}} + f_{v_{22}}) + F_3(p \cdot f_{h_3} + 2t) \end{cases} \quad (\text{A.17}) \\ &= \begin{cases} d^p + (d - F_1) \lambda = d^p + 2t(F_1 + F_2) + F_1(\lambda + 2t) \\ d^p + (d - F_2 - F_3) \lambda = d^p + 2t(F_1 + F_2) + F_2(\lambda + 2t) \\ d^p + (d - F_2 - F_3) \lambda = d^p + 2t(2F_3) + F_3(\lambda + 2t) \end{cases} \\ &= \begin{cases} \lambda d - (4t + 2\lambda) \cdot F_1 - 2t \cdot F_2 = 0 \\ \lambda d - (4t + 3\lambda) \cdot F_2 - 2t \cdot F_1 - \lambda \cdot F_3 = 0 \\ \lambda d - (6t + 2\lambda) \cdot F_3 - \lambda \cdot F_2 = 0 \end{cases} \end{aligned}$$

General case. Table A.1 lists the equation systems \mathcal{E}_H for all H . All equations in a system \mathcal{E}_H are of the form $E_{a,b}(F_c, F_d, F_f) = 0$, where

$$E_{a,b}(F_c, F_d, F_f) = \lambda d - (at + b\lambda) \cdot F_c - 2t \cdot F_d - \lambda \cdot F_f. \quad (\text{A.18})$$

The set $E(\mathcal{E}_H)$ contains all left-hand side terms $E_{a,b}$ of the equations in \mathcal{E}_H . Let the sum $\Sigma(\mathcal{E}_H)$ of an equation system \mathcal{E}_H be the equation $\sum_{E(\mathcal{E}_H)} E_{a,b} = 0$. It holds that for all H , $\Sigma(\mathcal{E}_H)$ is the equation

$$(H - 1)\lambda d - (6t + 2\lambda) \cdot F_1 - (6t + 3\lambda) \sum_{2 \leq u \leq H-1} F_u = 0. \quad (\text{A.19})$$

Solving equation $\Sigma(\mathcal{E}_H)$ for $F_V = \sum_{1 \leq u \leq H-1} F_u$, we obtain

$$F_V = \frac{(H - 1)\lambda d}{\rho(t) \cdot (6t + 3\lambda)} \quad (\text{A.20})$$

Ladder height H	Equation system \mathcal{E}_H
$H = 2$	$\begin{cases} E_{6,2}(F_1, 0, 0) = 0 \end{cases}$
$H = 3$	$\begin{cases} E_{4,2}(F_1, F_2, 0) = 0 \\ E_{4,3}(F_2, F_1, 0) = 0 \end{cases}$
Even $H \geq 4$ Odd $j, 3 \leq j \leq H - 3$	$\begin{cases} E_{4,2}(F_1, F_2, 0) = 0 \\ E_{4,2}(F_2, F_1, F_3) = 0 \\ E_{4,2}(F_j, F_{j+1}, F_{j-1}) = 0 \\ E_{4,2}(F_{j+1}, F_j, F_{j+2}) = 0 \\ E_{6,2}(F_{H-1}, 0, F_{H-2}) = 0 \end{cases}$
Odd $H \geq 5$ Odd $j, 3 \leq j \leq H - 4$	$\begin{cases} E_{4,2}(F_1, F_2, 0) = 0 \\ E_{4,2}(F_2, F_1, F_3) = 0 \\ E_{4,2}(F_j, F_{j+1}, F_{j-1}) = 0 \\ E_{4,2}(F_{j+1}, F_j, F_{j+2}) = 0 \\ E_{4,2}(F_{H-2}, F_{H-1}, F_{H-3}) = 0 \\ E_{4,3}(F_{H-1}, F_{H-2}, 0) = 0 \end{cases}$

Table A.1: Equation systems \mathcal{E}_H characterizing PI equilibrium for all H .

where $\lim_{t \rightarrow \infty} \rho(t) = 1$.

Due to the horizontal and vertical symmetry of the PI equilibrium on the ladder network, it holds that $f_V = 4 \cdot F_V$. Inserting f_V into $PoA_{H,\max}^{\#+}$ yields

$$1 + \lim_{t \rightarrow \infty} \frac{t}{H \cdot d^p} \cdot \frac{4(H-1)\lambda d}{\rho(t)(6t+3\lambda)} = 1 + \frac{2(H-1)}{3H} \cdot p. \quad (\text{A.21})$$

Taking the limit of this term for $H \rightarrow \infty$ results in

$$PoA_{\max}^{\#+} = 1 + \frac{2}{3} \cdot p. \quad (\text{A.22})$$

B

APPENDIX OF CHAPTER 3

B.1 EXAMPLE OF PSS EQUILIBRIUM ANALYSIS

In this section, we illustrate the calculation of strategy costs of the form set out in §3.2.4 by investigating whether the strategies described in §3.2.3 form PSS equilibria. Proving that a strategy profile is not a PSS equilibrium amounts to finding a deviant strategy that reduces an end-host's cost. Indeed, there exist such deviant strategies for the strategy profile v with $v(\sigma_g) = q$ and $v(\sigma_a) = 1 - q$ for all $q \in [0, 1]$.

$q \leq 1/2$. For the case $q \leq 1/2$, there is no inversion of link costs and a deviant agent can always assume that $f_\pi(t) > f_{\tilde{\pi}}(t)$ if the agent perceives $f_\pi(t - T) > f_{\tilde{\pi}}(t - T)$. The best strategy given such a strategy profile thus consists of switching to the cheaper path $\tilde{\pi}$ in a deterministic and immediate fashion, as in the greedy strategy σ_g presented in §3.2.3. Every delay of switching simply translates into more time needlessly spent on a strictly more expensive path. As the greedy strategy σ_g allows an end-host to reduce its cost, $v(\sigma_g)$ would quickly rise from q as more end-hosts adopt this strategy. Therefore, any strategy profile with $q \leq 1/2$ is not a PSS equilibrium.

$q > 1/2$. For $q > 1/2$, the periodic dynamics are structured as

$$f_\alpha(t) = \begin{cases} (A + q - 1) \cdot e^{-rt'} + 1 - q & \text{if } \frac{t^+(t)}{W} \text{ is even,} \\ -(A + q - 1) \cdot e^{-rt'} + q & \text{otherwise,} \end{cases} \quad (\text{B.1})$$

where $t' = t - t^+(t)$,

$$W = \frac{\ln(2e^{rT} - 1)}{r}, \quad \text{and} \quad A = \left(\frac{1}{2} - q\right) e^{-rT} + q. \quad (\text{B.2})$$

For showing that the anticipating strategy σ_a allows an end-host to improve its cost if $q \in (1/2, 1]$, we construct a mixed strategy $\sigma_p(q')$. This strategy $\sigma_p(q')$ plays the greedy strategy σ_g with probability q' and the anticipating strategy σ_a with probability $1 - q'$. We show that an end-host minimizes its cost by choosing $q' = 0$ given $q \in (1/2, 1]$, i.e., the anticipating strategy $\sigma_p(0) = \sigma_a$ is the better strategy than the greedy strategy $\sigma_p(1) = \sigma_g$.

As mentioned in §3.2.4, the cost of a strategy in periodic oscillating systems is computed over a single periodic interval. For the dynamics above, it is even sufficient to calculate the strategy cost between two turning points t_0^+ and t_1^+ , as the costs of the paths α and β would simply be reversed in the subsequent turning-point interval. Without loss of generality, we thus operate on a turning-point interval $[t_0^+, t_1^+]$ during which path α is perceived to be the cheaper path and $f_\alpha(t_0^+) < f_\beta(t_0^+)$.

The time-dependent strategy cost $C(\sigma_p(q'), t)$ for the deviant agent is calculated based on a linear combination of the two path costs, weighted by q' :

$$C(\sigma_p(q'), t) = \frac{1}{R} \int_t^{t+R} [q' \cdot c_\alpha(s) + (1 - q') \cdot c_\beta(s)] \, ds \quad (\text{B.3})$$

We further assume $R \leq W$, as any choice of higher R forces an agent to select a path that is sub-optimal during at least time $R - W$. Using this rationality constraint, it is possible to derive a formula for the strategy cost $C(\sigma_p(q')|O)$ that is a linear function of q' ,

$$C(\sigma_p(q')|O) = m \cdot q' + \gamma \quad (\text{B.4})$$

where γ is constant w.r.t. q' and the slope m is

$$m = \frac{R[(2q - 1)(W - R) + \frac{2a}{r}(e^{-rW} + 1)] + \frac{4a}{r^2}(e^{-rR} - 1)}{RW}, \quad (\text{B.5})$$

using the abbreviation $a = A + q - 1$. The cost function steepness is assumed to be $p = 1$, as the integral in Eq. (B.3) is not tractable otherwise.

The slope m can be shown to be positive for all $R > 0$, $r \in [0, 1]$, and $T \geq T(R)$, where $T(R)$ is such that $W = R$. Showing this property is feasible in a two-step proof, where we first show $m(T) > 0$ for $T = T(R)$ and then $\partial/\partial T m(T) > 0$ for all $T > T(R)$. The positiveness of m implies that the minimum of the strategy cost $C(\sigma_p(q')|O)$ is achieved for $q' = 0$, i.e., the anticipating strategy σ_a .

Given a strategy profile with $q > 1/2$, the adoption rate q of the greedy strategy would thus quickly decrease in favor of the anticipating strategy σ_a . Therefore, no strategy profile v with $q > 1/2$ represents a PSS equilibrium.

B.2 PROOF OF INSIGHT 3.1: NO PSS EQUILIBRIUM BY CONVERGENT STRATEGY

We can numerically show that there exist parallel-path systems where the greedy strategy σ_g ensures a lower cost than an underdamped convergent strategy σ_c , given universal adoption of the convergent strategy. In fact, the parallel-path system O assumed in Fig. 3.2 is such a parallel-path system where the strategy σ_c in an underdamped fashion does not yield the optimal cost. Using the definition of strategy cost introduced in §3.2.4, we calculate both $C(\sigma_c|O)$ and $C(\sigma_g|O)$.

In the calculation of $C(\sigma_c|O)$, we choose $u_c(\pi, t | \pi_t)$ as defined in Eq. (3.14). Furthermore, we can assume that $y(\pi_t, t | \sigma_c) = f_{\pi_t}(t)$, because of universal adoption of σ_c : An agent applying strategy σ_c allocates its traffic in accordance with all other agents and its probability distribution of being on a certain path is equivalent to the general traffic distribution over the paths. As for the calculation of $C(\sigma_g|O)$, we know that

$$u_g(\pi, t | \tilde{\pi}) = \begin{cases} 1 & \text{if } c_\pi(t - T) < c_{\tilde{\pi}}(t - T), \\ 0 & \text{otherwise,} \end{cases} \quad (\text{B.6})$$

and

$$y(\pi, t | \sigma_g) = \begin{cases} 1 & \text{if } c_\pi(t - T) = \min_{\pi'} c_{\pi'}(t - T), \\ 0 & \text{otherwise.} \end{cases} \quad (\text{B.7})$$

In Fig. B.1, the comparison of strategy costs for σ_c and σ_g are shown for all $R \in [0, 1]$ and the mentioned parallel-path system O . Clearly, given the parallel-path system O where agents universally apply an underdamped convergent strategy σ_c , any single agent would have an incentive to switch to a greedy strategy σ_g . The underdamped convergent strategy σ_c is thus not a PSS equilibrium.

B.3 PROOF OF INSIGHT 3.2: NO PSS EQUILIBRIUM BY MATE

The flow-allocation vector \mathbf{F}^\sim before projection is given by (using the abbreviation c_π for $c_\pi(t - T)$):

$$\mathbf{F}^\sim = \begin{pmatrix} F_\alpha - \gamma \cdot c_\alpha \\ F_\beta - \gamma \cdot c_\beta \end{pmatrix}. \quad (\text{B.8})$$

The projection on the feasible allocation set is the intersection of the line describing the feasible set $F'_\beta = d - F'_\alpha$ and the line through \mathbf{F}^\sim which is orthogonal to the

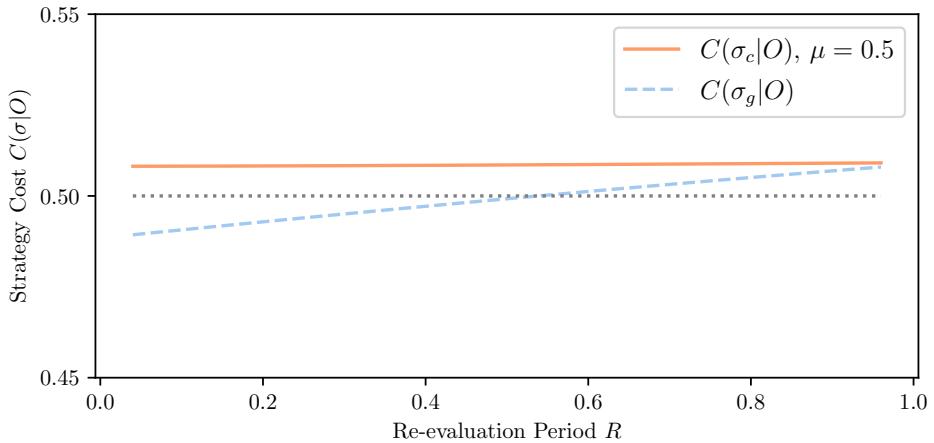


Figure B.1: The underdamped convergent strategy σ_c may be an inferior strategy (Parallel-path system $O = (\{\alpha, \beta\}, r = 1, p = 1, T = 2, A_0 = 1, v = \{\sigma_c \mapsto 1\})$).

feasibility line. If representing this latter line in the form $F'_\beta = m \cdot F'_\alpha + q$, we know that $m = 1$ because of orthogonality, and that $F_\beta - \gamma \cdot c_\beta = m \cdot (F_\alpha - \gamma \cdot c_\alpha) + q$, which implies the line:

$$F'_\beta = F'_\alpha + (F_\beta - F_\alpha - \gamma(c_\beta - c_\alpha)). \quad (\text{B.9})$$

This intersection is at $F'_\alpha = 1/2 \cdot (d - F_\beta + F_\alpha + \gamma(c_\beta - c_\alpha))$. The change in an end-host's flow on path α is thus

$$F'_\alpha - F_\alpha = \gamma/2 \cdot (c_\beta(t-T) - c_\alpha(t-T)). \quad (\text{B.10})$$

If path α appears to be the more expensive path, this change is performed by the re-evaluating end-hosts on path α , and otherwise by the re-evaluating end-hosts on path β . Multiplying by the number of re-evaluating end-hosts thus yields the aggregate dynamics

$$\frac{\partial f_\alpha}{\partial t} = \begin{cases} r \cdot \frac{\gamma}{2} \cdot \Delta c(t-T) \cdot f_\alpha(t) & \text{if } \Delta c(t-T) \leq 0 \\ r \cdot \frac{\gamma}{2} \cdot \Delta c(t-T) \cdot f_\beta(t) & \text{otherwise} \end{cases} \quad (\text{B.11})$$

where $\Delta c(t-T) = c_\beta(t-T) - c_\alpha(t-T)$.

B.4 CROSS STABILITY ANALYSIS

To prove Lemma 3.3, we show that stability at approximately equal load arises given universal adoption of path-selection strategy σ_C , i.e., end-hosts use a path if they have a registration for that path and only use a backup path in case of a path failure.

Approximate Wardrop equilibrium. For stability at approximately equal load with parameter ϵ , we assume that an end-host does not reallocate traffic at time t if the imbalance between paths $\Delta f(t - T) = |f_\alpha(t - T) - f_\beta(t - T)|$ is less than ϵ and thus the perceived cost difference is too small to justify path migration. If the imbalance $\Delta f(t)$ can be kept below ϵ for a period of length T , i.e., $\Delta f(t) < \epsilon$ for all $t \in [\tilde{t}, \tilde{t} + T]$, there will be no reallocation during the following interval $[\tilde{t} + T, \tilde{t} + 2T]$ and, by extension, also none in all subsequent intervals.

Balancing trials. In any balancing trial with start t_i , there will result a traffic imbalance $\Delta f(t_i) = |f_\alpha(t_i) - f_\beta(t_i)|$. This imbalance remains constant during time $[t_i, t_i + T]$, as the end-hosts only perceive the imbalance at time $t_i + T$. Thus, if $\Delta f(t_i) < \epsilon$, stability at approximately equal load is reached and enforcement of the mechanism can be suspended. However, if $\Delta f(t_i) \geq \epsilon$, stability is not achieved and the balancing trials are repeated until $\Delta f(t_i) < \epsilon$.

Since an end-host selects each path with probability $1/2$, the distribution of $f_\alpha(t_i)$ on $[0, 1]$ can be approximated with a normal distribution \mathcal{N} possessing mean $\mu = 1/2$ and variance σ^2 that depends on the number of end-hosts. If $\Phi(f_\alpha)$ is the CDF of \mathcal{N} , then the probability that $\Delta f(t_i) < \epsilon$ is $p_{<\epsilon} = \Phi((1 + \epsilon)/2) - \Phi((1 - \epsilon)/2) > 0$. With an increasing number of balancing trials over time t , the probability that $\Delta f(t_i) < \epsilon$ goes to 1 for $t \rightarrow \infty$. Therefore, for $t \rightarrow \infty$, it also holds that $\Delta f(t) < \epsilon$, which is stability at approximately equal load. Lemma 3.3 thus holds.

CROSS without puzzles. Indeed, the CROSS mechanism eventually achieves stability at approximately equal load even without relying on the computational puzzles mentioned in §3.6.1. However, it is desirable that oscillation can already be avoided during the execution of the mechanism. In particular, if a balancing trial fails and $\Delta f(t_i) \geq \epsilon$, no oscillation should take place until the start of the next balancing trial, i.e., during time $[t_i + T, t_{i+1}]$. If the imbalance $\Delta f(t_i)$ becomes visible to end-hosts at time $t_i + T$, the end-hosts on path $\tilde{\pi}$ with a backup registration for path π could migrate. However, since CROSS ensures that an end-host with a backup registration only uses its backup path in case of a path failure (§3.6.2), no migration takes place

during $[t_i + T, t_{i+1})$. Therefore, in absence of a path failure, the load distribution remains constant during the whole duration $[t_i, t_{i+1})$ of a balancing trial.

C

APPENDIX OF CHAPTER 4

C.1 ANALYSIS OF THE CONTINUITY-TIME DISTRIBUTION

The agent dynamics involved in P -step oscillation (Definition 4.1) allow to estimate how long the agents on a path have already been using that path without a packet loss, i.e., allow to characterize the distribution of the *continuity time* introduced above. For the following analysis, we introduce the notation $\theta(\pi, t)$, which shall denote the time since the most recent loss event on path π at time t .

We now derive a probability distribution $\mathbb{P}[\tau_i(t) = \tau]$, giving the probability that agent i on path $\pi_i(t)$ has continuity time $\tau \in \mathbb{N}$ at time t . This distribution will later be used to determine the expected congestion-window increase $\hat{\alpha}_\pi(t)$ in Eq. (4.8b). We consider an arbitrary agent $i \in A$ at an arbitrary time t , residing on path $\pi_i(t)$. Clearly, agent i must have continuity time $\tau_i(t) = 0$ right after a loss event, i.e., whenever $\theta(\pi_i(t), t) = 0$, irrespective of the rank of $\pi_i(t)$:

$$\forall t \text{ s.t. } \theta(\pi_i(t), t) = 0, p \in [P]. \quad \mathbb{P} [\tau_i(t) = 0 \mid \text{rank}(\pi_i(t), t) = p] = 1 \quad (\text{C.1})$$

However, in the subsequent time steps, where $\theta(\pi_i(t), t) > 0$, the continuity-time distribution of agent i on path $\pi_i(t)$ depends on the rank of that path. If $\text{rank}(\pi_i(t), t) = 0$ or, equivalently, $\text{rank}(\pi_i(t-1), t-1) = P-1$, all the $\hat{\alpha}^{(P-1)}$ agents that were on path π in the last time step $t-1$ have remained on the path and have increased their continuity time by 1, but their *relative share* is reduced by on-migration from other paths:

$$\begin{aligned} & \forall t \text{ s.t. } \theta(\pi_i(t), t) > 0, \tau > 0. \\ & \mathbb{P} [\tau_i(t) = \tau \mid \text{rank}(\pi_i(t), t) = 0] \\ &= \mathbb{P} [\tau_i(t-1) = \tau - 1 \mid \text{rank}(\pi_i(t-1), t-1) = P-1] \cdot \hat{\alpha}^{(P-1)} / \hat{\alpha}^{(0)} \\ &= \mathbb{P} [\tau_i(t-1) = \tau - 1 \mid \text{rank}(\pi_i(t-1), t-1) = P-1] \cdot (1-m)^{P-1} \end{aligned} \quad (\text{C.2})$$

All the $m \cdot (N - \hat{a}^{(P-1)})$ agents that migrated from the other paths have continuity time 0:

$$\forall t \text{ s.t. } \theta(\pi_i(t), t) > 0.$$

$$\mathbb{P}[\tau_i(t) = 0 \mid \text{rank}(\pi_i(t), t) = 0] = \frac{m \cdot (N - \hat{a}^{(P-1)})}{\hat{a}^{(0)}} = 1 - (1 - m)^{P-1}. \quad (\text{C.3})$$

If path $\pi_i(t)$ has $\text{rank}(\pi_i(t), t) \neq 0$, this path has experienced out-migration in the last time step, which reduces the number of agents on the path, but does not in itself affect the continuity-time distribution among the remaining agents. In fact, the continuity-time distribution is only shifted up by 1 in the last time step, but is otherwise unaffected:

$$\forall t \text{ s.t. } \theta(\pi_i(t), t) > 0, p \in [P] \setminus \{0\}, \tau > 0.$$

$$\begin{aligned} \mathbb{P}[\tau_i(t) = \tau \mid \text{rank}(\pi_i(t), t) = p] \\ = \mathbb{P}[\tau_i(t-1) = \tau - 1 \mid \text{rank}(\pi_i(t-1), t-1) = p-1] \end{aligned} \quad (\text{C.4})$$

These recursive characterizations of the probability distribution shall now be translated to closed-form expressions. To that end, we investigate (i) the case where the continuity time $\tau_i(t)$ equals the time since last loss $\theta_i(\pi_i(t), t)$, and (ii) the case where the continuity time $\tau_i(t)$ is lower than the time since the last loss $\theta_i(\pi_i(t), t)$; a continuity time higher than the last-loss time is impossible.

$\tau_i(t) = \theta(\pi_i(t), t)$. Fundamentally, the continuity time of an agent i , being on path $\pi_i(t)$ at time t , can only correspond to the last-loss time $\theta(\pi_i(t), t)$ of that path if the agent has never migrated since the loss. Clearly, the share of such persistent agents on path $\pi_i(t)$ is down-scaled every time path $\pi_i(t)$ experiences in-migration, according to Eq. (C.2). This in-migration happens whenever path $\pi_i(t)$ switches from rank $P-1$ to rank 0. Such a critical rank transition has never happened if the current path rank $p = \text{rank}(\pi_i(t), t)$ is larger or equals the last-loss time $\theta(\pi_i(t), t)$. In contrast, if $\theta(\pi_i(t), t) > p$, a critical rank transition has happened at least once, namely p time steps ago at time $t-p$. Moreover, additional critical rank transitions may have happened earlier, i.e., at the previous time steps $t-p-n \cdot P$, $n \in \mathbb{N}$, depending on how large $\theta(\pi_i(t), t)$ is. In summary, a critical rank transition has happened for $\lceil \theta(\pi_i(t), t) - p \rceil / P$ times, which holds both if $\theta(\pi_i(t), t) \leq p$ (always 0 times) and if $\theta(\pi_i(t), t) > p$ (at least one time). This insight suggests that

$$\mathbb{P}[\tau_i(t) = \theta(\pi_i(t), t) \mid \text{rank}(\pi_i(t), t) = p] = (1 - m)^{\lceil \frac{\tau_i(t) - p}{P} \rceil (P-1)}. \quad (\text{C.5})$$

$\tau_i(t) < \theta(\pi_i(t), t)$. In that case, it is clear that agent i migrated to its current path $\pi_i(t)$ only after this path experienced loss. Let that time of migration be denoted by t' , where $t' = t - \tau_i(t)$. At this time t' , agent i and its fellow migrants had a continuity time of $\tau_i(t') = 0$ and represented a share of $1 - (1 - m)^{P-1}$ among the total agents on path $\pi_i(t) = \pi_i(t')$, according to Eq. (C.3). In a next step, we are interested in how this share of migrants has evolved between time steps t' and t . Crucially, at time t' , path $\pi_i(t)$ must have had rank 0, as only rank-0 paths have experienced recent in-migration. Hence, at time t , the path must have rank $p = (t - t') \bmod P = \tau_i(t) \bmod P$; put differently, given rank p of a path, only a continuity time $\tau_i(t)$ that satisfies $\tau_i(t) \bmod P = p$ is present on the path. Furthermore, since migration time t' , path $\pi_i(t)$ has experienced $\lfloor (t - t')/P \rfloor = \lfloor \tau_i(t)/P \rfloor$ critical rank transitions from rank $P - 1$ to rank 0; these critical rank transitions have reduced the share of the agents that migrated at time t' according to Eq. (C.2). In total, we find for the case $\tau_i(t) < \theta(\pi_i(t), t)$ that

$$\begin{aligned} & \mathbb{P}[\tau_i(t) = \tau \mid \text{rank}(\pi_i(t), t) = p] \\ &= \begin{cases} (1 - (1 - m)^{P-1}) \cdot (1 - m)^{\lfloor \frac{\tau}{P} \rfloor (P-1)} & \text{if } \tau \bmod P = p, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \quad (\text{C.6})$$

In summary, the recursive characterizations of the continuity time are equivalent to the following explicit definition of the continuity-time distribution, which is visualized in Fig. C.1:

Insight C.1 Explicit Continuity-Time Distribution. At time t , the probability that an agent i on a path $\pi_i(t)$ with rank p and time since last loss $\theta(\pi_i(t), t)$ has continuity time τ is

$$\begin{aligned} & \mathbb{P}[\tau_i(t) = \tau \mid \text{rank}(\pi_i(t), t) = p] \\ &= \begin{cases} (1 - m)^{\lceil \frac{\tau-p}{P} \rceil (P-1)} & \text{if } \tau = \theta(\pi_i(t), t), \\ (1 - (1 - m)^{P-1}) \cdot (1 - m)^{\lfloor \frac{\tau}{P} \rfloor (P-1)} & \text{if } \tau < \theta(\pi_i(t), t) \\ & \wedge \tau \bmod P = p, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \quad (\text{C.7})$$

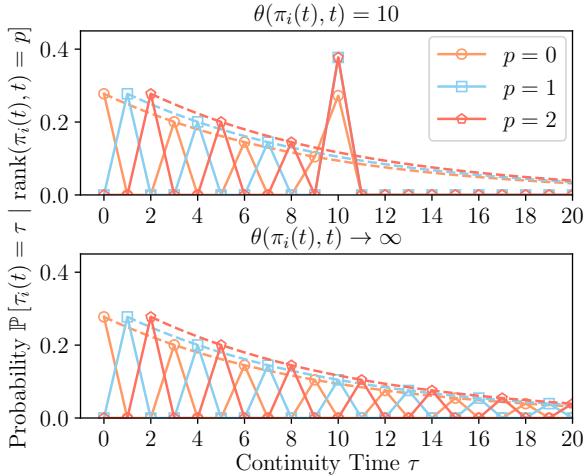


Figure C.1: Continuity-time distribution for $P = 3$, $m = 0.15$, and different θ . The dashed lines represent the function $(1 - (1 - m)^{P-1}) \cdot (1 - m)^{(\tau-p)/P \cdot (P-1)}$.

On a path π with rank p at time t , the expected additive increase $\hat{\alpha}_\pi(t)$ at time t is therefore:

$$\hat{\alpha}_\pi(t) = (1 - m)^{\lceil \frac{\theta-p}{P} \rceil (P-1)} \cdot \alpha(\theta) \quad (\text{C.8})$$

$$+ \sum_{k=0}^{\lceil (\theta-p)/P \rceil - 1} (1 - (1 - m)^{P-1}) \cdot (1 - m)^{k(P-1)} \cdot \alpha(Pk + p) \quad (\text{C.9})$$

where $\theta = \theta(\pi, t)$.

For increasing time since the last loss ($\theta \rightarrow \infty$), the expected average additive increase on a path with rank p converges to the following quantity, which can be easily computed for any additive-increase function α :

$$\hat{\alpha}^{(p)} = \sum_{k=0}^{\infty} (1 - (1 - m)^{P-1}) \cdot (1 - m)^{k(P-1)} \cdot \alpha(Pk + p) \quad (\text{C.10})$$

C.2 APPROXIMATION ACCURACY

The expected MPCC dynamics in Eq. (4.8) are an approximation of the actual probabilistic MPCC dynamics in Eq. (4.2), which are unsuitable for analytic investigation. In order to demonstrate the accuracy of this approximation, we present

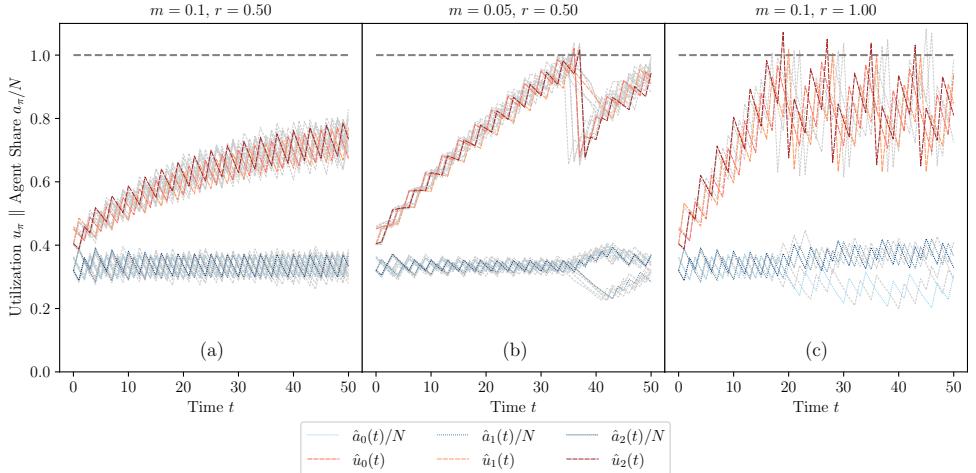


Figure C.2: Simulated MPCC dynamics $\{(a_\pi(t), f_\pi(t))\}_{\pi \in \Pi}$ (gray dashed) and expected MPCC dynamics $\{(\hat{a}_\pi(t), \hat{f}_\pi(t))\}_{\pi \in \Pi}$ (in color) for $P = 3$, $N = 1000$, $\alpha(\tau) = 1$, $\beta = 0.7$, and $C_\pi = 12\,000$ for every $\pi \in \Pi$.

a comparison between the actual dynamics and the expected dynamics for a selection of parameters in Fig. C.2. In each sub-figure, the actual MPCC dynamics from Eqs. (4.2a) and (4.2b) are simulated and shown with light gray lines, and the expected dynamics are computed and drawn with colored lines (agent dynamics in dotted lines, flow dynamics in dashed lines).

$P = 3$, $r \neq 1$. In Fig. C.2(a) and Fig. C.2(b), the expected dynamics are compared with results from 5 simulation runs of the actual dynamics. The expected dynamics appropriately capture the structure of both the agent dynamics and the flow dynamics, in particular the curvature, the convergence behavior and the reaction to loss (e.g., at $t = 35$ in Fig. C.2(b)). As the actual dynamics are realizations of a random variable, their values deviate from the expectation; however, the variance is modest.

$P = 3$, $r = 1$. In Fig. C.2(c), the actual flow dynamics look more different from the expected flow dynamics than for $r \neq 1$. This difference is due to loss events at different points in time, which can even result in case of low variance, but make the dynamics look quite different. However, the pattern of recurring loss is well

captured by the expected dynamics. In order to make this similarity visible, only one simulation run of the actual dynamics is shown.

Varying configurations. The analysis above is repeated for more paths and a non-constant additive-increase function in Fig. C.3. In particular, we repeat this analysis for constant additive increase, but with $P = 5$ (cf. Fig. C.3a), as well as with an additive-increase function that mimics TCP slow-start behavior ($\alpha_S(\tau) = 2^\tau$ if $\tau < 5$ else 1) for both $P = 3$ (cf. Fig. C.3b) and $P = 5$ (cf. Fig. C.3c).

C.3 LOGICAL CONSISTENCY OF P -STEP OSCILLATION

General rank-based flow volume. In §4.3.3, we have shown that given P -step oscillation and without capacity limits, the flow dynamics exponentially converge to a dynamic equilibrium where the rank- p path carries flow volume $\hat{f}^{(p)}$ in every time step. The general rank- p equilibrium flow volume is given by the following term:

$$\begin{aligned} \hat{f}^{(p)} = & \frac{\sum_{p'=0}^{p-1} (1-m)^{p-p'} \hat{\alpha}(p') \cdot \hat{a}^{(p')} + (1-m)^p \cdot \hat{\alpha}^{(p-1)} \cdot \hat{a}^{(p-1)}}{1 - (1 + m \cdot r \cdot z(m, P)) \cdot (1-m)^{P-1}} \\ & + \frac{(1 + m \cdot r \cdot z(m, P)) \cdot \sum_{p'=p}^{P-2} (1-m)^{P-1+p-p'} \cdot \hat{\alpha}(p') \cdot \hat{a}^{(p')}}{1 - (1 + m \cdot r \cdot z(m, P)) \cdot (1-m)^{P-1}} \end{aligned} \quad (\text{C.11})$$

Internal consistency of P -step oscillation. Interestingly, analyzing the equilibrium flow volumes $\{\hat{f}^{(p)}\}_{p \in [P]}$ allows to draw conclusions about the occurrence of P -step oscillation for a certain parameter combination, which works by logical contraposition: If P -step oscillation occurs for a certain parameter combination, then P -step oscillation produces the equilibrium flow volumes $\{\hat{f}^{(p)}\}_{p \in [P]}$. However, if the equilibrium flow volumes are themselves inconsistent with P -step oscillation, i.e., if $\hat{f}^{(p)} < \hat{f}^{(p+1)}$ for some $p \in [P-1]$, then the equilibrium cannot exist and there is a contradiction. This contradiction suggests that P -step oscillation is fundamentally impossible for the given parameter combination, as P -step oscillation would have produced the equilibrium flow volumes if it had occurred.¹

Inconsistent parameter subspace. Based on this reasoning, we can find a parameter sub-space for which P -step oscillation is impossible. More precisely, given any parameter combination, we can compute the equilibrium flow volumes $\{\hat{f}^{(p)}\}_{p \in [P]}$

¹ Note that the inverse is not true: The absence of a contradiction does not mean that P -step oscillation necessarily occurs for a given parameter combination.

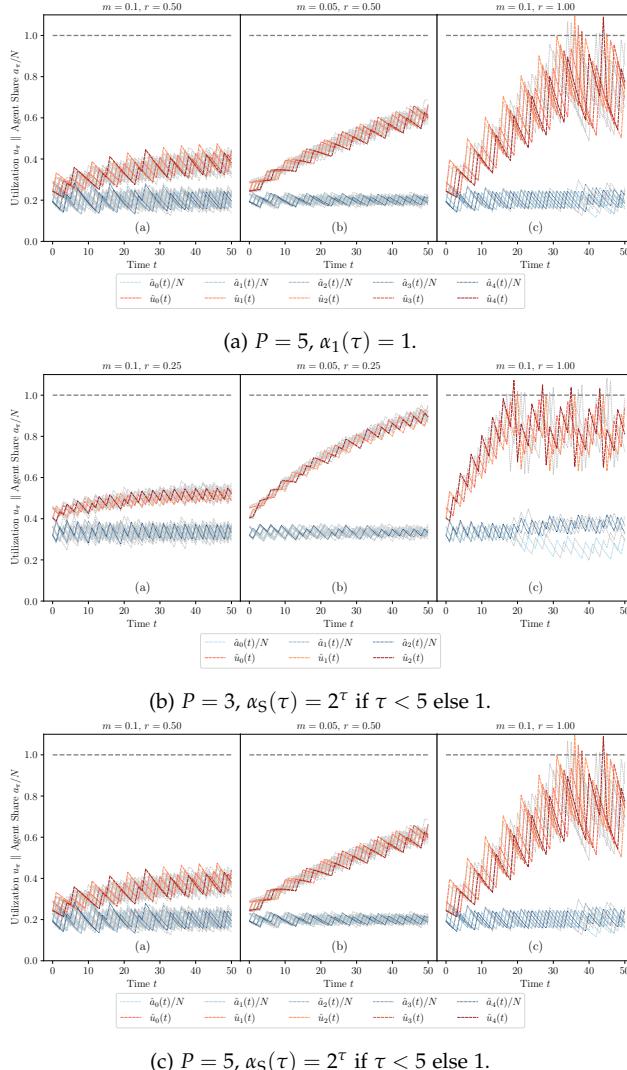


Figure C.3: Comparison of model and simulations to demonstrate approximation accuracy.

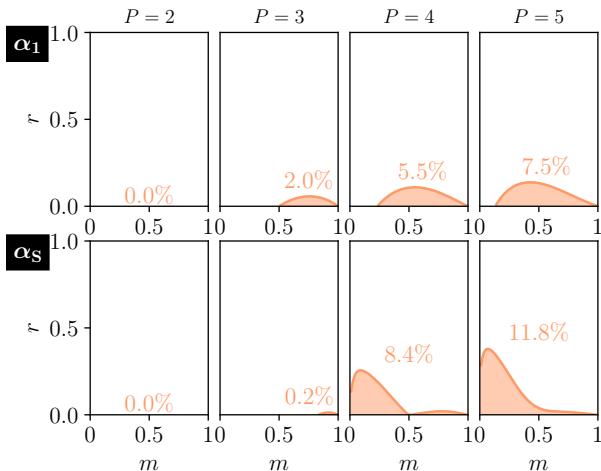


Figure C.4: Visualization of parameter sub-space that is inconsistent with P -step oscillation for different additive-increase functions $\alpha_1(\tau) = 1$ and $\alpha_S(\tau) = (2^\tau \text{ if } \tau < 5 \text{ else } 1)$.

and check if $\hat{f}^{(p)} < \hat{f}^{(p+1)}$ for any $p \in [P - 1]$. As Eq. (C.11) shows, the parameter space for the equilibrium flow volumes consists of the migration rate m , the reset softness r , the number of paths P , the additive-increase function $\alpha(\tau)$, and the number of agents N (appearing in $\hat{a}^{(p)}$). Luckily, as N is a linear coefficient of $\hat{f}^{(p)}$ and $N > 0$, N can be eliminated in the inequality $\hat{f}^{(p)} < \hat{f}^{(p+1)}$. We performed such an exploration of the parameter space with a focus on m and r , yielding the results in Fig. C.4. These results indicate that for the two analyzed additive-increase functions, P -step oscillation is never logically inconsistent for 2 paths and only rarely logically inconsistent for higher number of paths. While not a definitive proof for the prevalence of P -step oscillation, these results suggest that the notion of P -step oscillation is a sound concept for most parameter combinations.

C.4 ADDITIONAL FIGURES

This appendix section contains additional figures that illustrate concepts presented in the main body of Chapter 4. Fig. C.5 presents a simulation-based validation of the lower bounds on the flow volume in lossy equilibria, derived in §4.4.2.

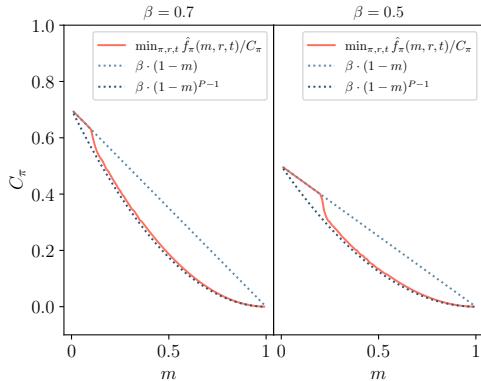


Figure C.5: Simulation-based validation of lower bounds on efficiency ϵ for lossy equilibria as derived in §4.4.2. Simulation parameters of interest include $C_\pi = 12000$, $N = 1000$, and $\alpha(\tau) = 1$.

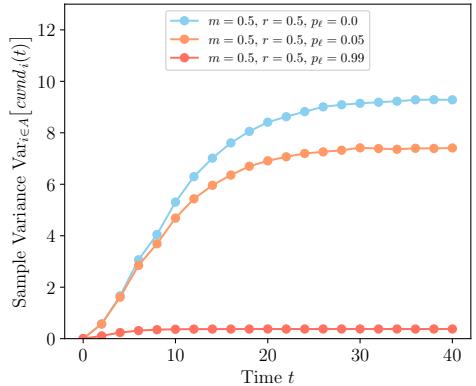


Figure C.6: Simulation-based computation of variance in congestion-window size according to the lossy Markov process in Fig. 4.7 for different values of loss probability p_ℓ .

Fig. C.6 presents the variance in congestion-window size given a lossy equilibrium, computed from simulation of the lossy Markov process in Fig. 4.7 in §4.5.2.

D

APPENDIX OF CHAPTER 5

D.1 LOSS-BASED CCA MODELS

In this section, we discuss the existing CCA models for Reno and CUBIC, which have been used in our simulations. We also validate these CCA fluid models in isolation

D.1.1 *Reno*

In its congestion-avoidance phase, TCP Reno increases the congestion-window size by $1/w$ upon successful transmission (signaled by an ACK) and cuts it in half upon loss. This adaptation logic is approximated by the following differential equation for the congestion-window size $w_i(t)$ of agent i using path π_i [166]:

$$\begin{aligned}\dot{w}_i &= x_i(t - d_i^p) \cdot (1 - p_i(t - d_i^p)) \cdot \frac{1}{w_i} \\ &\quad - x_i(t - d_i^p) \cdot p_{\pi_i}(t - d_i^p) \cdot \frac{w_i}{2}\end{aligned}\tag{D.1}$$

D.1.2 *CUBIC*

In contrast, TCP CUBIC cannot directly be described with a differential equation for the congestion-window size. Instead, Vardoyan et al. [249] suggest to track two instrumental variables in CUBIC, namely the time since last loss of agent i , s_i , and the congestion-window size at that moment of loss, w_i^{\max} :

$$\dot{s}_i = 1 - s_i \cdot x_i(t - d_i^p) \cdot p_i(t - d_i^p),\tag{D.2a}$$

$$\dot{w}_i^{\max} = (w_i - w_i^{\max}) \cdot x_i(t - d_i^p) \cdot p_i(t - d_i^p).\tag{D.2b}$$

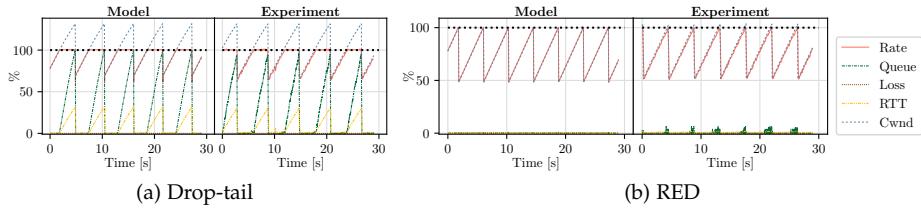


Figure D.1: Reno trace validation

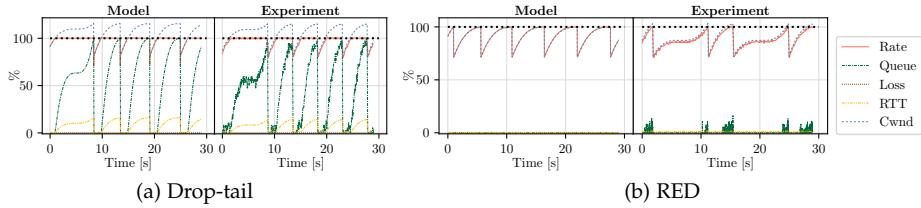


Figure D.2: CUBIC trace validation

The intuition behind Eq. (D.2a) is that s_i is increased by 1 in absence of loss ($p_{\pi_i} = 0$) and reduced to 0 when a loss occurs. Equation (D.2b) describes that $w_i^{\max}(t)$ should be updated to $w_i(t)$ in presence of loss. Knowing s_i and w_i^{\max} , the congestion-window size w can be determined by the CUBIC window-growth function [110],

$$w_i = c \cdot \left(s_i - \sqrt[3]{\frac{w_i^{\max} \cdot b}{c}} \right)^3 + w_i^{\max}, \quad (\text{D.3})$$

where c and b are configurable parameters with standardized values of 0.4 and 0.7, respectively [217]. Moreover, the CUBIC implementation in the Linux kernel uses a time unit of around 1 second for $s_i(t)$ [111].

D.1.3 Trace Validation of Models

Figs. D.1 and D.2 present a comparison of single-sender traces obtained from running both model simulation and mininet experiments. The fluid model correctly predicts that the rate growth of Reno and CUBIC decouples from the congestion-window growth as soon as the buffer fills up. In addition, the fluid models correctly capture that Reno and CUBIC lead to considerably smaller loss (barely visible) than BBRv1, which is insensitive to loss (cf. §5.4.2). Finally, the fluid model correctly predicts that the sending rate of loss-based CCAs never exceeds the bottleneck rate

under RED, while the congestion windows can temporarily exceed the network BDP under a drop-tail queuing discipline. As a result, the smaller buffer usage under RED is also reflected in the model, although the difference between RED and drop-tail is more pronounced in the model. This last difference is due to the idealization of the RED algorithm in the model.

D.2 AGGREGATE VALIDATION FOR SHORT RTT

Figures D.3, D.4, D.5, D.6, and D.7 extend the validation, performed in §5.4.3, of fluid models regarding the aggregate metrics Jain fairness, loss rates, buffer occupancy, utilization and jitter, respectively. In contrast to the validation in the body of Chapter 5, the fluid-model predictions are experimentally validated for a bottleneck-link delay of 5 milliseconds and total RTTs between 10 and 20 milliseconds.

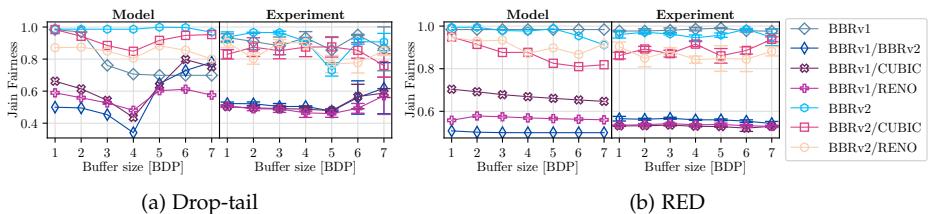


Figure D.3: Fairness validation

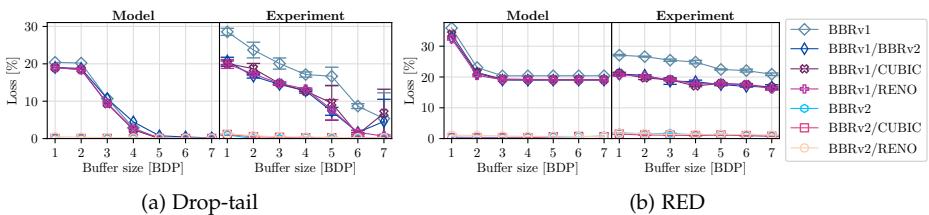


Figure D.4: Loss validation

D.3 PROOFS FOR STABILITY ANALYSIS

This appendix section contains the proofs of the theorems in §5.5.

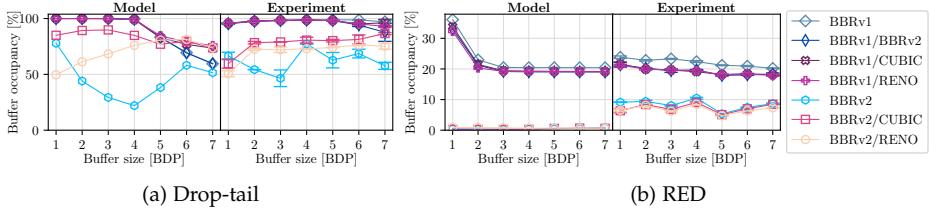


Figure D.5: Queuing validation

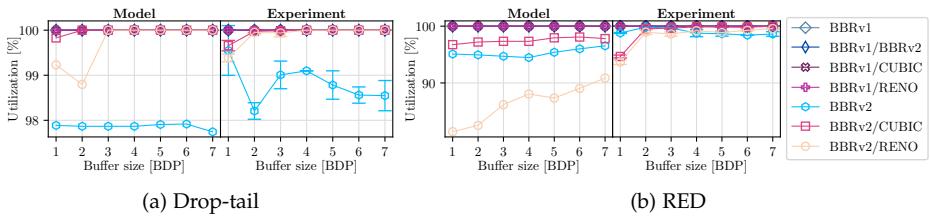


Figure D.6: Utilization validation

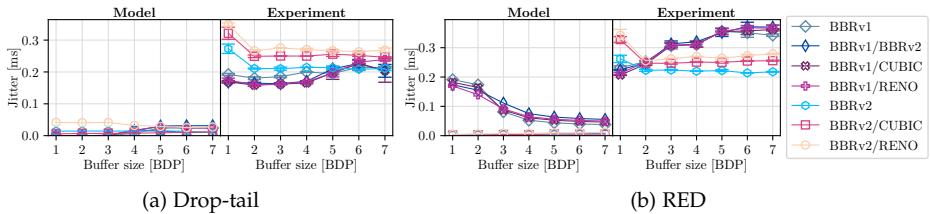


Figure D.7: Jitter validation

D.3.1 Proof of Theorem 5.1

Equilibrium validity. First, we consider the case for $q_{\ell^*} = 0$. In that case, the second condition in Eq. (5.45) implies:

$$\forall i \in U_{\ell^*}. \quad x_i^{\text{btl}} = x_i^{\max} \stackrel{(5.42)}{=} \min(5/4, \Delta_i) \cdot x_i^{\text{btl}}. \quad (\text{D.4})$$

For this equation, the only solution in terms of $\{\Delta_i\}_{i \in U_{\ell^*}}$ is $\Delta_i = 1 \forall i \in U_{\ell^*}$, which shows that the network is in equilibrium for $q_{\ell^*} = 0$ and $d_i = \sum_{\ell \in \pi_i} q_\ell / C_\ell \forall i \in U_{\ell^*}$.

For $q_{\ell^*} > 0$, we note that the conditions in Eq. (5.45) can be transformed into the following conditions for each x_i^{btl} :

$$x_i^{\text{btl}} = \max(1, 1/\Delta_i) \cdot \left(C_{\ell^*} - \sum_{j \neq i} \min(1, \Delta_j) \cdot x_j^{\text{btl}} \right) \quad (\text{D.5})$$

$$x_i^{\text{btl}} = C_{\ell^*} - \max(4/5, 1/\Delta_i) \cdot \sum_{j \neq i} \min(1, \Delta_j) \cdot x_j^{\text{btl}} \quad (\text{D.6})$$

Clearly, the previously found solution $\Delta_i = 1 \forall i \in U_{\ell^*}$ is also a solution to the conditions in Eqs. (D.5) and (D.6). Hence, we have proven that the network is in equilibrium if $d_i = \sum_{\ell \in \pi_i} q_{\ell} / C_{\ell} \forall i \in U_{\ell^*}$.

Equilibrium uniqueness. It remains to prove that the previously mentioned equilibria are the only possible equilibria if $q_{\ell^*} > 0$. To confirm the uniqueness of these equilibria, we first assume an equilibrium where $\exists i \in U_{\ell^*}. \Delta_i > 1$. For that agent i , this assumption implies

$$\begin{aligned} x_i^{\text{btl}} &\stackrel{(\text{D.5})}{=} \underbrace{\max(1, 1/\Delta_i)}_{=1} \cdot \left(C_{\ell^*} - \sum_{j \neq i} \min(1, \Delta_j) \cdot x_j^{\text{btl}} \right) \\ &< C_{\ell^*} - \underbrace{\max(4/5, 1/\Delta_i)}_{<1} \sum_{j \neq i} \min(1, \Delta_j) \cdot x_j^{\text{btl}} \stackrel{(\text{D.6})}{=} x_i^{\text{btl}}, \end{aligned} \quad (\text{D.7})$$

which is a contradiction and rules out an equilibrium. Conversely, if assuming an equilibrium where $\exists i \in U_{\ell^*}. \Delta_i < 1$, this assumption implies

$$\begin{aligned} x_i^{\text{btl}} &\stackrel{(\text{D.5})}{=} \underbrace{\max(1, 1/\Delta_i)}_{=1/\Delta_i > 1} \cdot \left(C_{\ell^*} - \sum_{j \neq i} \min(1, \Delta_j) \cdot x_j^{\text{btl}} \right) \\ &> C_{\ell^*} - \underbrace{\max(4/5, 1/\Delta_i)}_{=1/\Delta_i > 1} \sum_{j \neq i} \min(1, \Delta_j) \cdot x_j^{\text{btl}} \stackrel{(\text{D.6})}{=} x_i^{\text{btl}}, \end{aligned} \quad (\text{D.8})$$

which again leads to contradictory equations. Hence, no equilibria other than the equilibrium with $\forall i \in U_{\ell^*}. \Delta_i = 1$ are possible, which concludes the proof.

D.3.2 Proof of Theorem 5.3

Equilibrium rate. Given $\Delta_i \geq 5/4$ for all $i \in U_{\ell^*}$, the equilibrium condition on $\{x_i^{\text{btl}}\}_{i \in U_{\ell^*}}$ is:

$$\forall i \in U_{\ell^*}. x_i^{\text{btl}} \stackrel{(5.45)}{=} x_i^{\max} \stackrel{(5.42)}{=} \frac{5/4x_i^{\text{btl}} C_{\ell^*}}{5/4x_i^{\text{btl}} + \sum_{j \neq i} x_j^{\text{btl}}} = C_{\ell^*} - 4/5 \sum_{j \neq i} x_j^{\text{btl}}. \quad (\text{D.9})$$

This equation system requires all $\{x_i^{\text{btl}}\}_{i \in U_{\ell^*}}$ to be equal, which allows a straightforward solution:

$$\forall i \in U_{\ell^*}. x_i^{\text{btl}} \stackrel{(\text{D.9})}{=} C_{\ell^*} - 4/5(N-1) \cdot x_i^{\text{btl}} = \frac{5/4x_i^{\text{btl}} C_{\ell^*}}{(N+1/4)x_i^{\text{btl}}} = \frac{5C_{\ell^*}}{4N+1}. \quad (\text{D.10})$$

Stability. It remains to show that this equilibrium is asymptotically stable, for which we employ the indirect Lyapunov method. We apply this method to a non-linear dynamic process with $\{x_i^{\text{btl}}\}_{i \in U_{\ell^*}}$ as state variables and $\{x_i^{\text{btl}}\}_{i \in U_{\ell^*}}$ as vector-valued evolution function f . The Jacobian matrix J_f has the following entries, which we evaluate at the equilibrium:

$$\frac{\partial \dot{x}_i}{\partial x_i} = \frac{5/4C \sum_{j \neq i} x_j^{\text{btl}}}{(5/4x_i + \sum_{j \neq i} x_j)^2} - 1 \stackrel{(\text{D.10})}{=} -\frac{5}{4N+1} =: J_{ii} \quad (\text{D.11})$$

$$\frac{\partial \dot{x}_i}{\partial x_j} = -\frac{5/4Cx_i}{(5/4x_i + \sum_{j \neq i} x_j)^2} \stackrel{(\text{D.10})}{=} -\frac{4}{4N+1} =: J_{ij} \quad (\text{D.12})$$

(D.13)

The eigenpairs (λ, \mathbf{v}) of J_f at the equilibrium satisfy the following conditions:

$$\forall i \in U_{\ell^*}. (J_{ii} - \lambda)v_i + J_{ij} \sum_{j \neq i} v_j = 0 \quad (\text{D.14})$$

The first type of solution for this equation system is given by $\lambda = J_{ii} - J_{ij} < 0$ and every \mathbf{v} with $\|\mathbf{v}\|_1 = 0$, i.e., with the eigenvector entries summing up to 0. The second type of solution is found by assuming $\lambda \neq J_{ii} - J_{ij}$, which implies equal $v_i \forall i \in U_{\ell^*}$ and hence (together with $v_i \neq 0$) $\lambda = J_{ii} + (N-1)J_{ij} < 0$. Since the eigenvalues of the Jacobian are thus consistently negative, the indirect Lyapunov method suggests that the dynamics are asymptotically stable.

D.3.3 Proof of Theorem 5.4

The equilibrium conditions can be translated given $q_{\ell^*} > 0$:

$$\sum_{i \in U_{\ell^*}} \min(1, \delta_i) \cdot x_i^{\text{btl}} = C_{\ell^*} \quad (\text{D.15})$$

$$\iff x_i^{\text{btl}} = \max(1, 1/\delta_i) \cdot (C_{\ell^*} - \sum_{j \neq i} \min(1, \delta_j) \cdot x_j^{\text{btl}}) \quad (\text{D.16})$$

$$x_i^{\text{btl}} = x_i^{\max} \stackrel{(5.57)}{=} \frac{5/4 \cdot \min(1, \delta_i) \cdot x_i^{\text{btl}} \cdot C_{\ell_i}}{5/4 \cdot \min(1, \delta_i) \cdot x_i^{\text{btl}} + \sum_{j \neq i} \min(1, \delta_j) \cdot x_j^{\text{btl}}}$$

$$\iff x_i^{\text{btl}} = C_{\ell^*} - 4/5 \cdot \max(1, 1/\delta_i) \cdot \sum_{j \neq i} \min(1, \delta_j) \cdot x_j^{\text{btl}} \quad (\text{D.17})$$

For the special case $N = 1$, the conditions above simplify to:

$$x_i^{\text{btl}} = \max(1, 1/\delta_i) \cdot C_{\ell^*} \quad \text{and} \quad x_i^{\text{btl}} = C_{\ell^*}, \quad (\text{D.18})$$

which clearly implies $\delta_i = 1$.

For $N > 1$, these constraints potentially admit multiple equilibria. However, the equilibrium from Theorem 5.4 is a special equilibrium for which δ_i is equal across all $i \in U_{\ell^*}$, i.e., $\delta_i = \delta$. Substituting δ for all δ_i , and equating Eq. (D.16) with Eq. (D.17), we obtain:

$$\begin{aligned} \forall i \in U_{\ell^*}. \quad & \max(1, 1/\delta) \cdot C_{\ell^*} - \sum_{j \neq i} x_j^{\text{btl}} = C_{\ell^*} - 4/5 \sum_{j \neq i} x_j^{\text{btl}} \\ \implies & \sum_{j \neq i} x_j^{\text{btl}} = 5 \cdot (\max(1, 1/\delta) - 1) \cdot C_{\ell^*}. \end{aligned} \quad (\text{D.19})$$

This equation system requires that all $x_i^{\text{btl}} \forall i \in U_{\ell^*}$ equal the same value, and hence we obtain the following perfectly fair equilibrium from Eq. (D.15):

$$\forall i \in U_{\ell^*}. \quad x_i^{\text{btl}} = \max(1, 1/\delta) \cdot \frac{C_{\ell^*}}{N}. \quad (\text{D.20})$$

Furthermore, by inserting x_i^{btl} from Eq. (D.20) into Eq. (D.17), we obtain:

$$x_i^{\text{btl}} \stackrel{(\text{D.20})}{=} \max(1, 1/\delta) \cdot \frac{C_{\ell^*}}{N} \stackrel{(\text{D.17})}{=} C_{\ell^*} - 4/5 \cdot \max(1, 1/\delta) \cdot \frac{N-1}{N} \cdot C_{\ell^*}. \quad (\text{D.21})$$

Given this equation, we can show that $\delta \leq 1$ by producing a contradiction when assuming $\delta > 1 \iff \max(1, 1/\delta) = 1$:

$$\frac{1}{N} \cdot C_{\ell^*} = \left(1 - 4/5 \cdot \frac{N-1}{N}\right) \cdot C_{\ell^*} \stackrel{/C_{\ell^*}}{\iff} 4/5 \cdot \frac{N-1}{N} = \frac{N-1}{N}, \quad (\text{D.22})$$

which is a contradiction given $N > 1$.

In contrast, solving that equation given $\delta \leq 1 \iff \max(1, 1/\delta) = 1/\delta$ yields:

$$\begin{aligned} \frac{1}{\delta N} \cdot C_{\ell^*} &= \left(1 - \frac{4}{5} \cdot \frac{N-1}{\delta N}\right) \cdot C_{\ell^*} \\ \iff \frac{\delta N - 1 - 4/5(N-1)}{\delta N} &= 0 \iff \delta = \frac{4N+1}{5N}, \end{aligned} \tag{D.23}$$

which is equivalent to the condition in Theorem 5.4.

D.3.4 Proof of Theorem 5.5

In the scenario under consideration, the equilibrium requires that the propagation delay d_i is equal for all senders. Moreover, it holds that $q_\ell = 0 \forall \ell \neq \ell^*$. Hence, we can simplify: $\delta_i = \delta(q_{\ell^*}) := d / (d + q_{\ell^*} / C_{\ell^*})$, where d is the propagation delay experienced by all agents. As a result, the equilibrium requires that

$$\delta(q_{\ell^*}) = \frac{4N+1}{5N} \iff q_{\ell^*} = \frac{N-1}{4N+1} d C_{\ell^*}. \tag{D.24}$$

Dynamic process. We translate the reduced model from §5.5.2.1 into a nonlinear dynamic process with the sending rates $\{x_i\}_{i \in U_{\ell^*}}$ and the queue length q_{ℓ^*} as state variables. The evolution of these state variables is given by vector-valued function f with the following entries:

$$\dot{x}_i = \dot{\delta}(q_{\ell^*}) x_i^{\text{btl}} + \delta(q_{\ell^*}) \dot{x}_i^{\text{btl}} \tag{D.25}$$

$$= \left(\frac{C_{\ell^*} - \sum_{k \in U_{\ell^*}} x_k}{C_{\ell^*}(d + q_{\ell^*} / C_{\ell^*})} + \frac{5/4\delta C}{5/4x_i + \sum_{j \neq i} x_j} - 1 \right) \cdot x_i$$

$$\dot{q}_{\ell^*} = \sum_{i \in U_{\ell^*}} x_i - C_{\ell^*} \tag{D.26}$$

Jacobian matrix. The corresponding Jacobian matrix \mathbf{J}_f is composed of the following entries:

$$\frac{\partial \dot{x}_i}{\partial x_i} = \frac{C_{\ell^*} - 2x_i - \sum_{j \neq i} x_j}{C_{\ell^*}(d + q_{\ell^*} / C_{\ell^*})} + \frac{5/4\delta C \sum_{j \neq i} x_j^{\text{btl}}}{(5/4x_i + \sum_{j \neq i} x_j)^2} - 1 \tag{D.27}$$

$$\frac{\partial \dot{x}_i}{\partial x_j} = -\frac{x_i}{C_{\ell^*}(d + q_{\ell^*} / C_{\ell^*})} - \frac{5/4\delta C x_i}{(5/4x_i + \sum_{j \neq i} x_j)^2} \tag{D.28}$$

$$\frac{\partial \dot{x}_i}{\partial q} = \frac{1}{d + \frac{q_{\ell^*}}{C_{\ell^*}}} \left(\frac{C_{\ell^*} - \sum_{k \in U_{\ell^*}} x_k}{C_{\ell^*}^2 \left(d + \frac{q_{\ell^*}}{C_{\ell^*}} \right)} - \frac{\frac{5}{4}\delta C}{\frac{5}{4}x_i + \sum_{j \neq i} x_j} \right) x_i \quad (\text{D.29})$$

$$\frac{\partial \dot{q}}{\partial x_i} = 1 \quad \frac{\partial \dot{q}}{\partial q} = 0 \quad (\text{D.30})$$

Evaluating the Jacobian matrix at the equilibrium point from Theorem 5.4 yields the following matrix \mathbf{J} :

$$\frac{\partial \dot{x}_i}{\partial x_i} = -\frac{4N+1}{5N^2d} - \frac{5}{4N+1} =: J_{ii} \quad \frac{\partial \dot{q}}{\partial x_j} = 1 \quad (\text{D.31})$$

$$\frac{\partial \dot{x}_i}{\partial x_j} = -\frac{4N+1}{5N^2d} - \frac{4}{4N+1} =: J_{ij} \quad \frac{\partial \dot{q}}{\partial q} = 0 \quad (\text{D.32})$$

$$\frac{\partial \dot{x}_i}{\partial q} = -\frac{4N+1}{5N^2d} =: J_{iq} \quad (\text{D.33})$$

Lyapunov analysis. By Lyapunov's indirect method, the above Jacobian matrix must have exclusively negative eigenvalues in order for the equilibrium to be asymptotically stable, i.e., for every pair (λ, \mathbf{v}) with $\mathbf{J}\mathbf{v} = \lambda\mathbf{v}$, the eigenvalue λ must be lower than 0. To verify this property \mathbf{J} , we concretize the eigenvalue condition:

$$\forall i \in U_{\ell^*}. \quad J_{ii}v_i + J_{ij} \sum_{j \neq i} v_j + J_{iq}v_q = \lambda v_i \quad (\text{D.34})$$

$$\sum_{i \in U_{\ell^*}} v_i = \lambda v_q \quad (\text{D.35})$$

By solving these equations for v_q and equating the resulting terms, we obtain the following conditions:

$$\forall i \in U_{\ell^*}. \quad \sum_{k \in U_{\ell^*}} v_k = \frac{\lambda}{J_{iq}} \left((\lambda - J_{ii})v_i - J_{ij} \sum_{j \neq i} v_j \right) \quad (\text{D.36})$$

This equation allows two types of solutions. First, for $\lambda = J_{ii} - J_{ij} = -1/(4N+1) < 0$, the set of valid eigenvectors \mathbf{v} is only constrained by a condition on $\sum_{k \in U_{\ell^*}} v_k$; more importantly for the proof, λ is negative. Second, for $\lambda \neq J_{ii} - J_{ij}$, the values $v_i \forall i \in U_{\ell^*}$ must be equal such that the equation system from Eq. (D.36) can be collapsed into a single quadratic equation, which yields the maximum eigenvalue λ^+ :

$$N \cdot v_i = \frac{\lambda}{J_{iq}} ((\lambda - J_{ii})v_i - J_{ij}(N-1)v_i) \implies \lambda^+ = -1. \quad (\text{D.37})$$

Since the maximum eigenvalue λ^+ is negative, all eigenvalues of J are negative, which by the indirect Lyapunov method proves that the dynamic process defined by f is asymptotically stable.

E

APPENDIX OF CHAPTER 6

E.1 PROOF OF THEOREM 6.1: SHORT-TERM EQUILIBRIUM

In the following, we consistently write \tilde{s} for $\tilde{s}(\alpha)$, as α is considered fixed throughout the proof (analogously for \tilde{w}^{\max} and \tilde{x}^{btl})

The CUBIC maximum window \tilde{w}^{\max} in equilibrium (and by extension also the CUBIC equilibrium rate \tilde{x}^C) is directly determined by the CUBIC window-growth duration \tilde{s} in equilibrium:

$$\tilde{w}^{\max} \stackrel{(6.9)}{=} \frac{c}{b} \tilde{s}^3 \iff \tilde{x}^C = \frac{c \tilde{s}^3}{b \tilde{\tau}_k}. \quad (\text{E.1})$$

The short-term equilibrium delay $\tilde{\tau}_k$ also matches the general equilibrium delay $\bar{\tau}_k$ from Eq. (6.12) because also the short-term equilibrium includes non-zero loss, which implies a full buffer. To characterize the equilibrium completely, it remains to determine \tilde{s} and \tilde{x}^{btl} , which we achieve with a case distinction on α .

E.1.1 Low Probing Strength: $\alpha \leq 1$

We first consider the case $\alpha \leq 1$. According to Lemma 6.3, \tilde{x}_i^{btl} must equal the lower bound χ in this case. For $\tilde{x}_i^{\text{btl}} = \chi$, the CUBIC equilibrium condition in Lemma 6.1 suggests:

$$\begin{aligned} \frac{b \tilde{\tau}_k}{c \tilde{s}^4} &\stackrel{(6.11)}{=} 1 - \frac{C_\ell}{\tilde{y}_\ell} \stackrel{(6.3)}{=} 1 - \frac{C_\ell}{\beta \tilde{x}^{\text{btl}} + \tilde{x}^C} \stackrel{\text{E.1}}{=} 1 - \frac{C_\ell}{\alpha \chi + \frac{c \tilde{s}^3}{b \tilde{\tau}_k}} \iff \\ &\frac{c^2}{b \tilde{\tau}_k} \tilde{s}^7 - c(C_\ell - \alpha \chi) \tilde{s}^4 - c \tilde{s}^3 - \alpha b \tilde{\tau}_k \chi = 0, \end{aligned} \quad (\text{E.2})$$

where the polynomial in the last equation corresponds to $\tilde{S}_2(\tilde{s})$ in Eq. (6.23). Note that $\beta = \alpha$ is implied by $\alpha \leq 1$ (cf. Eqs. (6.1) and (6.14)). The septic equation $\tilde{S}_2(s) = 0$ has a unique solution for positive s , which follows from two arguments:

1. *Negativity at o*: $\tilde{S}_2(0) = -\alpha b \tilde{\tau}_k \chi < 0$.
2. *Strict monotonic increase after monotonic decrease*:

$$\begin{aligned} \exists s' \geq 0 \quad \text{s.t.} \quad \forall s \in [0, s']. \tilde{S}'_2(s') \leq 0 \quad \text{and} \\ \forall s > s'. \tilde{S}'_2(s) > 0. \end{aligned} \quad (\text{E.3})$$

The existence of such a unique turning point s' is proven in Appendix E.1.3.

E.1.2 High Probing Strength: $\alpha > 1$

Conversely, if $\alpha > 1$, \hat{x}^{btl} may exceed χ . For this case, we consider the *unrestricted* equilibrium \hat{x}^{btl} , i.e., the equilibrium bottleneck-bandwidth estimate \hat{x}^{btl} according to Lemma 6.2, but without restriction to the domain $[\chi, \infty)$:

$$\hat{x}^{\text{btl}} \stackrel{(6.16)}{=} C_\ell - \frac{\tilde{x}^C}{\alpha} \stackrel{(\text{E.1})}{=} C_\ell - \frac{c \tilde{s}^3}{\alpha b \tilde{\tau}_k}. \quad (\text{E.4})$$

We now distinguish the cases $\hat{x}^{\text{btl}} \geq \chi$ and $\hat{x}^{\text{btl}} < \chi$.

$\hat{x}^{\text{btl}} \geq \chi$. In this case, the actual equilibrium \hat{x}^{btl} matches the unrestricted equilibrium \hat{x}^{btl} . Plugging $\hat{x}^{\text{btl}} = \hat{x}^{\text{btl}}$ into the CUBIC equilibrium condition from Eq. (6.11), we obtain:

$$\begin{aligned} \frac{b \tilde{\tau}_k}{c \tilde{s}^4} \stackrel{(6.11)}{=} 1 - \frac{C_\ell}{\beta \hat{x}^{\text{btl}} + \tilde{x}^C} \stackrel{(\text{E.1})}{=} \beta=1 1 - \frac{C_\ell}{\hat{x}^{\text{btl}} + \frac{c \tilde{s}^3}{b \tilde{\tau}_k}} \stackrel{(\text{E.4})}{\iff} \\ \frac{(\alpha - 1)c^2}{\alpha b \tilde{\tau}_k} \tilde{s}^7 - \frac{(\alpha - 1)c}{\alpha} \tilde{s}^3 - b C_\ell \tilde{\tau}_k = 0, \end{aligned} \quad (\text{E.5})$$

where the polynomial in the second equation corresponds to $\tilde{S}_1(\tilde{s})$ in Eq. (6.22). Note that $\beta = 1$ is implied by $\alpha > 1$. Moreover, the uniqueness of the solution to $\tilde{S}_1 = 0$ can be shown similarly as before:

1. *Negativity at o*: $\tilde{S}_1(0) = -b C_\ell \tilde{\tau}_k < 0$.
2. *Strict monotonic increase after monotonic decrease*:

$$\tilde{S}'_1(s) > 0 \iff \frac{7(\alpha - 1)c^2}{\alpha b \tilde{\tau}_k} s^6 - 3 \frac{(\alpha - 1)c}{\alpha} s^2 > 0 \quad (\text{E.6})$$

$$\Leftrightarrow \frac{7(\alpha - 1)c^2}{\alpha b \tilde{\tau}_k} s^4 - \frac{3(\alpha - 1)c}{\alpha} > 0 \Leftrightarrow s > \sqrt[4]{\frac{3b\tilde{\tau}_k}{7c}} > 0.$$

$\hat{x}^{\text{btl}} < \chi$. This case arises if the equilibrium window-growth duration \tilde{s} is sufficiently high:

$$\begin{aligned} \hat{x}^{\text{btl}} = C_\ell - \frac{c\tilde{s}^3}{\alpha b \tilde{\tau}_k} &< \chi \Leftrightarrow \frac{c\tilde{s}^3}{b\tilde{\tau}_k} > \alpha(C_\ell - \chi) \\ \Leftrightarrow \tilde{s} &> \sqrt[3]{\frac{\alpha b \tilde{\tau}_k}{c}(C_\ell - \chi)} =: \hat{s}(\alpha), \end{aligned} \quad (\text{E.7})$$

where $\hat{s}(\alpha)$ is the window-growth duration that leads to

$$\hat{x}^{\text{btl}} = C_\ell - \frac{c\hat{s}(\alpha)^3}{\alpha b \tilde{\tau}_k} = \chi \stackrel{\geq \chi}{=} \tilde{x}^{\text{btl}}, \quad (\text{E.8})$$

and to a CUBIC equilibrium rate

$$\tilde{x}^C \stackrel{(\text{E.4})}{=} \alpha(C_\ell - \hat{x}^{\text{btl}}) \stackrel{(\text{E.8})}{=} \alpha(C_\ell - \chi), \quad (\text{E.9})$$

given probing strength α . In order for $\hat{s}(\alpha)$ to be an equilibrium (i.e., $\tilde{s} = \hat{s}(\alpha)$), $\hat{s}(\alpha)$ has to satisfy the CUBIC equilibrium conditions from Lemma 6.1:

$$\begin{aligned} \frac{b\tilde{\tau}_k}{c\hat{s}(\alpha)^4} &\stackrel{(\text{6.11})}{=} 1 - \frac{C_\ell}{\beta\tilde{x}^{\text{btl}} + \tilde{x}^C} \stackrel{(\text{E.8})}{=} 1 - \frac{C_\ell}{\chi + \alpha(C_\ell - \chi)} \\ \stackrel{(\text{E.7})}{\Leftrightarrow} \frac{\alpha^4(\alpha - 1)^3}{(\chi + \alpha(C_\ell - \chi))^3} &= \frac{c}{b\tilde{\tau}_k(C_\ell - \chi)^7} \end{aligned} \quad (\text{E.10})$$

Let $\hat{\alpha}$ be the solution in α to Eq. (E.10), which cannot be found analytically in general. However, it must hold that $\hat{\alpha} > 1$, as $\alpha = 1$ yields a zero LHS in Eq. (E.10), which cannot match the non-zero RHS.

At probing strength $\hat{\alpha}$, it holds that $\hat{s}(\hat{\alpha}) = \tilde{s}$, i.e., in equilibrium, \tilde{s} is such that \tilde{x}^{btl} is exactly moved to χ even without restriction to $[\chi, 0]$. Since $\hat{s}(\alpha)$ is an increasing function of α according to Eq. (E.7), any $\alpha < \hat{\alpha}$ also leads to $\hat{s}(\alpha) < \tilde{s}$, and thus to $\hat{x}^{\text{btl}} < \chi = \tilde{x}^{\text{btl}}$. Hence, for $\alpha < \hat{\alpha}$, \tilde{x}^C is found by solving \tilde{S}_2 , i.e., as for $\alpha \leq 1$. We thus arrive at the condition on α in Theorem 6.1.

E.1.3 Uniqueness of Turning Point for \tilde{S}_2

We note that the turning point s' of \tilde{S}_2 in Eq. (E.2) should be a unique root of the first derivative \tilde{S}'_2 , where

$$\tilde{S}'_2(s) = \frac{7c^2}{b\tilde{\tau}_k} s^6 - 4c(C_\ell - \alpha\chi)s^3 - 3cs^2. \quad (\text{E.11})$$

E.1.3.1 Strict Convexity above Critical Value s'''

To find the area in which \tilde{S}'_2 is strictly convex, we solve the following inequality for the second derivative of \tilde{S}'_2 :

$$\tilde{S}'''_2(s) > 0 \iff \frac{210c^2}{b\tilde{\tau}_k} s^4 - 24c(C_\ell - \alpha\chi)s > 0 \quad (\text{E.12})$$

$$\iff \frac{210c^2}{b\tilde{\tau}_k} s^3 - 24c(C_\ell - \alpha\chi) > 0 \quad (\text{E.13})$$

The division by s is admissible because we only consider $s > 0$. To identify s''' , we note that the LHS in Eq. (E.13) increases from non-positive to positive with s if $C_\ell > \alpha\chi$, and is consistently positive for all $s > 0$ in the rare case where $C_\ell \leq \alpha\chi$. Hence, we arrive at the following value for s''' , marking the start of the convex area of \tilde{S}'_2 :

$$s''' = \begin{cases} \sqrt[3]{\frac{b\tilde{\tau}_k}{c} \cdot \frac{4}{35}(C_\ell - \alpha\chi)} & \text{if } C_\ell > \alpha\chi \\ 0 & \text{if } C_\ell \leq \alpha\chi. \end{cases} \quad (\text{E.14})$$

E.1.3.2 Non-Positivity at Critical Value s'''

The function \tilde{S}'_2 yields the following non-positive value at the start point s''' of the convex area:

$$\tilde{S}'_2(s''') = \begin{cases} -0.11(C_\ell - \alpha\chi)^2\tilde{\tau}_k - 3cs'''^{\frac{2}{3}} < 0 & \text{if } C_\ell > \alpha\chi \\ 0 & \text{if } C_\ell \leq \alpha\chi. \end{cases} \quad (\text{E.15})$$

E.1.3.3 Non-Negativity in Convex Area

To find an argument s at which \tilde{S}'_2 is non-negative, we again distinguish the cases $C_\ell > \alpha\chi$ and $C_\ell \leq \alpha\chi$.

$C_\ell > \alpha\chi$. We consider the following function $\Psi^-(s)$, which constitutes a lower bound on \tilde{S}'_2 , i.e., $\forall s > 0$. $\Psi^-(s) \leq \tilde{S}'_2(s)$:

$$\Psi^-(s) = \begin{cases} \frac{7c^2}{b\tilde{\tau}_k} s^6 - 4c(C_\ell - \alpha\chi)s^3 - 3cs^3 & \text{if } s \geq 1 \\ \frac{7c^2}{b\tilde{\tau}_k} s^6 - 4c(C_\ell - \alpha\chi)s^2 - 3cs^2 & \text{if } s < 1 \end{cases} \quad (\text{E.16})$$

The highlights in Eq. (E.16) mark the differences of Ψ^- and \tilde{S}'_2 from Eq. (E.11).

The non-zero root of Ψ^- is $s^- = \max(\sqrt[3]{s_0}, \sqrt[4]{s_0})$, where

$$s_0 = \frac{b\tilde{\tau}_k}{c} \left(\frac{4}{7}(C_\ell - \alpha\chi) + \frac{3}{7} \right). \quad (\text{E.17})$$

Since Ψ^- is a lower bound on \tilde{S}'_2 , it holds that

$$\tilde{S}'_2(s^-) \geq \Psi^-(s^-) = 0. \quad (\text{E.18})$$

The non-negativity point s^- is in the convex area if $s''' < s^-$, which demonstrably holds:

$$\begin{aligned} s''' &\stackrel{(\text{E.14})}{=} \sqrt[3]{\frac{b\tilde{\tau}_k}{c} \cdot \frac{4}{35}(C_\ell - \alpha\chi)} < \sqrt[3]{\frac{b\tilde{\tau}_k}{c} \cdot \frac{4}{7}(C_\ell - \alpha\chi)} \\ &< \sqrt[3]{\frac{b\tilde{\tau}_k}{c} \left(\frac{4}{7}(C_\ell - \alpha\chi) + \frac{3}{7} \right)} \leq s^-. \end{aligned} \quad (\text{E.19})$$

$C_\ell \leq \alpha\chi$. In this rare case, the lower bound function Ψ^- and its root s^- are more simply expressed:

$$\Psi^-(s) = \frac{7c^2}{b\tilde{\tau}_k} s^6 + 4c(C_\ell - \alpha\chi)s^2 - 3cs^2 \quad (\text{E.20})$$

$$s^- = \sqrt[4]{\frac{b\tilde{\tau}_k}{c} \left(-\frac{4}{7}(C_\ell - \alpha\chi) + \frac{3}{7} \right)} \quad (\text{E.21})$$

Note that Ψ^- is a *strict* lower bound (i.e., $\Psi^-(s) < \tilde{S}'_2(s)$) if $C_\ell < \alpha\chi$, and equals \tilde{S}'_2 for $C_\ell = \alpha\chi$. Hence, the following property holds on s^- :

$$\tilde{S}'_2(s^-) \begin{cases} > \Psi^-(s^-) & \text{if } C_\ell < \alpha\chi \\ = \Psi^-(s^-) & \text{if } C_\ell = \alpha\chi \end{cases} = 0 \quad (\text{E.22})$$

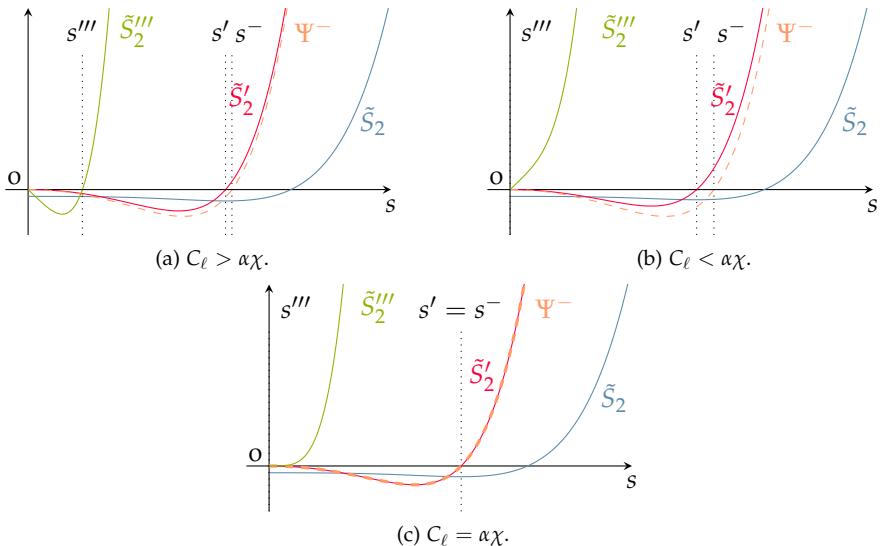


Figure E.1: Case distinction for determination of unique turning point in §E.1.3.4.

The convex-area membership of $s^- > s'''$ is demonstrated analogously to the previous case:

$$s''' \stackrel{(E.14)}{=} 0 \stackrel{C_\ell \leq \alpha\chi}{<} \sqrt[4]{\frac{b\tilde{\tau}_k}{c} \left(-\frac{4}{7}(C_\ell - \alpha\chi) + \frac{3}{7} \right)} \stackrel{(E.21)}{=} s^-. \quad (E.23)$$

E.1.3.4 Combination of Arguments

In summary, the function \tilde{S}'_2 evolves in a strictly convex fashion from $\tilde{S}'_2(s''') \leq 0$ at s''' to $\tilde{S}'_2(s^-) \geq 0$ at $s^- > s'''$. To demonstrate that these conditions imply a unique root with subsequent increasing behavior, we consider the relevant cases separately and visualize them in Fig. E.1 to simplify understanding:

$C_\ell > \alpha\chi$ (Fig. E.1a). In this case, we observe that $\tilde{S}'_2(s''') < 0$ (Eq. (E.15)) and $\tilde{S}'_2(s^-) \geq 0$ (Eq. (E.18)). Given this property, the function \tilde{S}'_2 has at least some increasing part between s''' and s^- , and at least the first root is in such an increasing part. Moreover, since \tilde{S}'_2 is strictly convex, the function keeps increasing once it is increasing, proving uniqueness of the root s' and increasing behavior after s' .

$C_\ell < \alpha\chi$ (Fig. E.1b). In this case, it holds that $\tilde{S}'_2(s''') = 0$ (Eq. (E.15)) and $\tilde{S}'_2(s^-) > 0$ (Eq. (E.22)). Strict convexity allows either (i) monotonic increase between $s''' = 0$ and $s^- > 0$, which creates a unique root at $s''' = 0$ such that $\tilde{S}'_2(s) > 0 \forall s > 0$, or (ii) a decrease followed by an increase, causing a unique root $s' \in (s''', s^-) = (0, s^-)$ above which $\tilde{S}'_2(s) > 0$.

$C_\ell = \alpha\chi$ (Fig. E.1c). In this last case, the function \tilde{S}'_2 has the property $\tilde{S}'_2(s''') = \tilde{S}'_2(s^-) = 0$ (Eqs. (E.15) and (E.22)). Given strict convexity, the function \tilde{S}'_2 can only return to 0 at s^- if it first decreases and then increases forever, causing a unique root $s' = s^-$ such that $\tilde{S}'_2(s) > 0 \forall s > s'$.

E.2 PROOF OF THEOREM 6.2: STABILITY OF THE SHORT-TERM EQUILIBRIUM

Our proof proceeds in three main steps, namely:

1. Stability investigation of the full linearized dynamics
2. Dimension reduction by characterization of the center manifold
3. Stability investigation of the lower-dimensional dynamics

E.2.1 *Stability Investigation of the Full Linearized Dynamics*

E.2.1.1 *Centering the Dynamics*

Fundamentally, we consider the dynamic system $\dot{\sigma} = f(\sigma)$, where

$$\sigma = \begin{bmatrix} \tilde{x}^{\text{btl}} \\ \tilde{w}^{\max} \\ \tilde{s} \end{bmatrix}, \quad f = \begin{bmatrix} \dot{\tilde{x}}^{\text{btl}} \\ \dot{\tilde{w}}^{\max} \\ \dot{\tilde{s}} \end{bmatrix}, \quad \text{and} \quad \tilde{\sigma} = \begin{bmatrix} \tilde{x}^{\text{btl}} \\ \tilde{w}^{\max} \\ \tilde{s} \end{bmatrix} \quad (\text{E.24})$$

are the state variables, the adaptation function, and the equilibrium, respectively. Hence, it holds that $\tilde{\sigma} = f(\tilde{\sigma})$. To show the asymptotic stability of the equilibrium $\tilde{\sigma}$, we first center the dynamic system around the equilibrium: We transform the adaptation function f to f° such that $\tilde{\mathbf{z}} = \mathbf{0} = [0, 0, 0]^\top$ is an equilibrium of the dynamic system $\dot{\mathbf{z}} = f^\circ(\mathbf{z})$, i.e., $\mathbf{0} = \tilde{\mathbf{z}} = f^\circ(\tilde{\mathbf{z}})$. In our case, the dynamics in the

differential equations for \dot{x}^{btl} , \dot{w}^{\max} (Eq. (6.5)) and \dot{s} (Eq. (6.6)) are centered as follows:

$$\dot{z}_1 = \frac{\alpha(z_1 + \tilde{x}^{\text{btl}})C}{\alpha(z_1 + \tilde{x}^{\text{btl}}) + \frac{1}{\tilde{\tau}_k}W^\circ(z_2, z_3)} - (z_1 + \tilde{x}^{\text{btl}}) \quad (\text{E.25})$$

$$\dot{z}_2 = (W^\circ(z_2, z_3) - (z_2 + \tilde{w}^{\max})) \cdot \frac{W^\circ(z_2, z_3)}{\tilde{\tau}_k} \cdot p^\circ(\mathbf{z}) \quad (\text{E.26})$$

$$\dot{z}_3 = 1 - (z_3 + \tilde{s}) \cdot \frac{W^\circ(z_2, z_3)}{\tilde{\tau}_k} \cdot p^\circ(\mathbf{z}) \quad (\text{E.27})$$

where both the CUBIC window-growth function W° and the loss-rate function p° are centered as well:

$$W^\circ(z_2, z_3) \stackrel{(6.2)}{=} (z_2 + \tilde{w}^{\max}) + c \left((z_3 + \tilde{s}) - \sqrt[3]{\frac{(z_2 + \tilde{w}^{\max})b}{c}} \right)^3 \quad (\text{E.28})$$

$$p^\circ(\mathbf{z}) \stackrel{(6.4)}{=} \begin{cases} 1 - \frac{C_\ell}{\beta(z_1 + \tilde{x}^{\text{btl}}) + \frac{1}{\tilde{\tau}_k}W^\circ(z_2, z_3)} & \text{if } y_\ell(\mathbf{z}) > C_\ell \wedge q_\ell = B_\ell \\ 0 & \text{otherwise} \end{cases} \quad (\text{E.29})$$

E.2.1.2 Linearizing the Dynamics

We now linearize the adaptation function f° of this centered dynamic system around the equilibrium $\tilde{\mathbf{z}} = \mathbf{0}$ with the first-order Taylor expansion:

$$\mathcal{L}_{\mathbf{0}}[f^\circ](\mathbf{z}) = f^\circ(\mathbf{0}) + \mathbf{J}_{f^\circ}(\mathbf{0})\mathbf{z} = \mathbf{J}_{f^\circ}(\mathbf{0})\mathbf{z}, \quad (\text{E.30})$$

where $\mathcal{L}_{\mathbf{0}}[f^\circ]$ indicates the linearization of the adaptation function f° around the new equilibrium $\mathbf{0}$, $\mathbf{J}_{f^\circ}(\mathbf{0})$ is the Jacobian matrix of f° evaluated at the equilibrium $\mathbf{0}$. Since $f^\circ(\mathbf{0}) = \mathbf{0}$, the dynamics around the equilibrium are dominated by $\mathbf{J}_{f^\circ}(\mathbf{0})\mathbf{z}$. The stability properties of the dynamic system thus depend on $\mathbf{J}_{f^\circ}(\mathbf{0})$, especially on its eigenvalues.

To find these eigenvalues, we need to characterize $\mathbf{J}_{f^\circ}(\mathbf{0})$, which is considerably simplified by the following identities:

$$\begin{aligned} W^\circ(0, 0) &= \tilde{w}^{\max}, & \frac{\partial W^\circ}{\partial z_2}(0, 0) &= 1, \\ \frac{\partial W^\circ}{\partial z_3}(0, 0) &= 0, & \text{and} \quad \frac{\partial p^\circ}{\partial z_3}(\mathbf{0}) &= 0. \end{aligned} \quad (\text{E.31})$$

Moreover, we know the following properties of the equilibrium from §6.3.1:

$$\begin{aligned} \tilde{s} > 0, \quad \tilde{w}^{\max} = \frac{c}{b} \tilde{s}^3 > 0, \quad p^\circ(\mathbf{0}) = \frac{b\tilde{\tau}_k}{c\tilde{s}^4} \in (0, 1], \\ \text{and } \tilde{x}^{\text{btl}} = \max \left(\chi, C_\ell - \frac{1}{\alpha} \frac{\tilde{w}^{\max}}{\tilde{\tau}_k} \right) > 0. \end{aligned} \quad (\text{E.32})$$

Using these equalities, $\mathbf{J}_{f^\circ}(\mathbf{0})$ in our case is:

$$\mathbf{J}_{f^\circ}(\mathbf{0}) = \begin{bmatrix} J_{11} & J_{12} & 0 \\ 0 & 0 & 0 \\ J_{31} & J_{32} & J_{33} \end{bmatrix} \quad (\text{E.33})$$

$$J_{11} = \frac{\alpha C_\ell \frac{1}{\tilde{\tau}_k} W^\circ(0, 0)}{\left(\alpha \tilde{x}^{\text{btl}} + \frac{1}{\tilde{\tau}_k} W^\circ(0, 0) \right)^2} - 1 \quad (\text{E.34})$$

$$= \frac{\alpha C_\ell \frac{1}{\tilde{\tau}_k} \tilde{w}^{\max}}{\left(\alpha \tilde{x}^{\text{btl}} + \frac{1}{\tilde{\tau}_k} \tilde{w}^{\max} \right)^2} - 1 < 0$$

$$J_{12} = - \frac{\alpha C_\ell \tilde{x}^{\text{btl}}}{\tilde{\tau}_k \left(\alpha \tilde{x}^{\text{btl}} + \frac{1}{\tilde{\tau}_k} W^\circ(0, 0) \right)^2} \cdot \frac{\partial W^\circ}{\partial z_2}(0, 0) \quad (\text{E.35})$$

$$= - \frac{\alpha C_\ell \tilde{x}^{\text{btl}}}{\tilde{\tau}_k \left(\alpha \tilde{x}^{\text{btl}} + \frac{1}{\tilde{\tau}_k} \tilde{w}^{\max} \right)^2} < 0$$

$$J_{31} = - \tilde{s} \frac{W^\circ(0, 0)}{\tilde{\tau}_k} \frac{\beta C}{\left(\beta \tilde{x}^{\text{btl}} + \frac{1}{\tilde{\tau}_k} W^\circ(0, 0) \right)^2} \quad (\text{E.36})$$

$$= - \frac{\beta C_\ell \tilde{s} \tilde{w}^{\max}}{\tilde{\tau}_k \left(\beta \tilde{x}^{\text{btl}} + \frac{1}{\tilde{\tau}_k} \tilde{w}^{\max} \right)^2} < 0$$

$$J_{32} = - \frac{\tilde{s}}{\tilde{\tau}_k} \cdot \frac{\partial W^\circ}{\partial z_2}(0, 0) \cdot \left(1 - \frac{C_\ell}{\beta \tilde{x}^{\text{btl}} + \frac{1}{\tilde{\tau}_k} W^\circ(0, 0)} \right) \quad (\text{E.37})$$

$$- \tilde{s} \cdot \frac{W^\circ(0, 0)}{\tilde{\tau}_k} \cdot \frac{C_\ell}{\tilde{\tau}_k (\beta \tilde{x}^{\text{btl}} + \frac{1}{\tilde{\tau}_k} W^\circ(0, 0))^2} \cdot \frac{\partial W^\circ}{\partial z_2}(0, 0)$$

$$= - \frac{\tilde{s}}{\tilde{\tau}_k} \cdot p^\circ(\mathbf{0}) - \frac{\tilde{s} \tilde{w}^{\max}}{\tilde{\tau}_k^2} \frac{C_\ell}{\left(\beta \tilde{x}^{\text{btl}} + \frac{1}{\tilde{\tau}_k} \tilde{w}^{\max} \right)^2} < 0$$

$$\begin{aligned}
J_{33} &= - \frac{W^\circ(0,0)}{\tilde{\tau}_k} \cdot p^\circ(\mathbf{0}) - \frac{\tilde{s}}{\tilde{\tau}_k} \cdot \frac{\partial W^\circ}{\partial z_3}(0,0) \cdot p^\circ(\mathbf{0}) \\
&\quad - \tilde{s} \cdot \frac{W^\circ(0,0)}{\tilde{\tau}_k} \cdot \frac{\partial p^\circ}{\partial z_3}(\mathbf{0}) \\
&= - \frac{1}{\tilde{s}} < 0
\end{aligned} \tag{E.38}$$

Hence, all entries of $\mathbf{J}_{f^\circ}(\mathbf{0})$ are negative. Among the above matrix entries, the bounding of the entry J_{11} in Eq. (E.34) is not trivial and requires a case distinction on χ :

$\chi \leq C_\ell - \tilde{w}^{\max}/(\alpha \tilde{\tau}_k)$. In this case, the bounding is straightforward:

$$\tilde{x}^{\text{btl}} \stackrel{(E.32)}{=} \max \left(\chi, C_\ell - \frac{\tilde{w}^{\max}}{\alpha \tilde{\tau}_k} \right) \stackrel{\text{Case}}{=} C_\ell - \frac{\tilde{w}^{\max}}{\alpha \tilde{\tau}_k} \tag{E.39}$$

$$\implies \alpha \tilde{x}^{\text{btl}} + \frac{\tilde{w}^{\max}}{\tilde{\tau}_k} = \alpha C_\ell \implies \tag{E.40}$$

$$J_{11} = \frac{\alpha C_\ell \frac{\tilde{w}^{\max}}{\tilde{\tau}_k}}{(\alpha C_\ell)^2} - 1 = \frac{\frac{\tilde{w}^{\max}}{\tilde{\tau}_k}}{\alpha C_\ell} - 1 \stackrel{\text{Case}}{\leq} \frac{\alpha(C_\ell - \chi)}{\alpha C_\ell} - 1 < 0 \tag{E.41}$$

$\chi > C_\ell - \tilde{w}^{\max}/(\alpha \tilde{\tau}_k)$. For this case, we note that \tilde{x}^{btl} is at the minimum χ of its domain:

$$\tilde{x}^{\text{btl}} \stackrel{(E.32)}{=} \max \left(\chi, C_\ell - \frac{\tilde{w}^{\max}}{\alpha \tilde{\tau}_k} \right) = \chi \tag{E.42}$$

Moreover, we obtain a lower bound on the CUBIC equilibrium rate \tilde{x}^C for this case:

$$\chi > C_\ell - \frac{\tilde{w}^{\max}}{\alpha \tilde{\tau}_k} \iff \tilde{x}^C = \frac{\tilde{w}^{\max}}{\tilde{\tau}_k} > \alpha(C_\ell - \chi). \tag{E.43}$$

Moreover, we obtain another lower bound \hat{x}^C on the CUBIC equilibrium rates that lead to a negative Jacobian entry J_{11} :

$$J_{11}(\tilde{x}^C) \stackrel{(E.34)}{=} \frac{\alpha C_\ell \tilde{x}^C}{(\alpha \chi + \tilde{x}^C)^2} - 1 < 0 \tag{E.44}$$

$$\iff \tilde{x}^{C2} + \alpha(2\chi - C_\ell)\tilde{x}^C + \alpha^2\chi^2 =: \Psi(\tilde{x}^C) > 0 \tag{E.45}$$

$$\iff \tilde{x}^C > \alpha \frac{\sqrt{C_\ell(C_\ell - 4\chi)} + (C_\ell - 2\chi)}{2} =: \hat{x}^C. \tag{E.46}$$

We note that the lower bound \hat{x}^C in Eq. (E.46) only exists if $4\chi \leq C_\ell$. In contrast, the non-existence of \hat{x}^C means that Ψ from Eq. (E.45) has no root. Since Ψ is strictly convex, and a strictly convex function with no roots is always positive, $\Psi > 0$ from Eq. (E.45) holds for all \hat{x}^C , and J_{11} is always negative in this case.

Conversely, if $4\chi \leq C_\ell$ and \hat{x}^C therefore exists, we need to verify that $\hat{x}^C \leq \alpha(C_\ell - \chi)$ such that all $\hat{x}^C > \alpha(C_\ell - \chi)$ (i.e., all CUBIC equilibrium rates possible according to Eq. (E.43)) lead to a negative J_{11} . Given \hat{x}^C from Eq. (E.46), we obtain:

$$\begin{aligned} \alpha \frac{\sqrt{C_\ell(C_\ell - 4\chi)} + (C_\ell - 2\chi)}{2} &\leq \alpha(C_\ell - \chi) \iff \\ -4\chi C_\ell \leq 0 &\iff \chi \geq 0 \iff T. \end{aligned} \quad (\text{E.47})$$

In summary, $J_{11} < 0$ thus holds for all $\tilde{w}^{\max}/\tilde{\tau}_k$ and all χ .

E.2.1.3 Finding the Eigenpairs

Finding the eigenvalues and eigenvectors of $\mathbf{J}_{f^\circ}(0)$ means finding (λ, \mathbf{v}) such that $\lambda \in \mathbb{C}$ is an eigenvalue, $\mathbf{v} \in \mathbb{C}^3$ is the corresponding eigenvector and must be non-zero ($\mathbf{v} \neq 0$), and $\mathbf{J}_{f^\circ}(0)\mathbf{v} = \lambda\mathbf{v}$. Hence, any solution (λ, \mathbf{v}) satisfies the following system of equations:

$$(J_{11} - \lambda)v_1 + J_{12}v_2 = 0 \quad (\text{E.48})$$

$$\lambda v_2 = 0 \quad (\text{E.49})$$

$$J_{31}v_1 + J_{32}v_2 + (J_{33} - \lambda)v_3 = 0 \quad (\text{E.50})$$

Clearly, Eq. (E.49) implies that λ or v_2 must be zero.

$\lambda = 0$. First, we check whether $\lambda = 0$ is an eigenvalue of $\mathbf{J}_{f^\circ}(0)$. Assuming $\lambda = 0$, the equation system reduces to two equations:

$$J_{11}v_1 + J_{12}v_2 = 0 \quad J_{31}v_1 + J_{32}v_2 + J_{33}v_3 = 0 \quad (\text{E.51})$$

Given this equation system, we can identify the following eigenvector $v^{(1)}$ for the eigenvalue $\lambda^{(1)} = 0$:

$$v_1^{(1)} \in \mathbb{C} \setminus \{0\} \quad v_2^{(1)} = -\frac{J_{11}}{J_{12}}v_1^{(1)} \quad v_3^{(1)} = \frac{J_{11}J_{32} - J_{12}J_{31}}{J_{12}J_{33}}v_1^{(1)} \quad (\text{E.52})$$

Note that $v_1^{(1)}$ must be non-zero because $v_1 = 0$ would imply that also $v_2 = 0$ and $v_3 = 0$, which would be an invalid eigenvector. This eigenvector exists because the denominators are non-zero given $J_{12} < 0$ and $J_{12}J_{33} > 0$.

$v_2 = 0$. To find further eigenvalues, we now assume $v_2 = 0$, leading again to a reduced equation system:

$$(J_{11} - \lambda)v_1 = 0 \quad J_{31}v_1 + (J_{33} - \lambda)v_3 = 0 \quad (\text{E.53})$$

For this equation system, we can perform a case distinction on v_1 , the first entry of the eigenvector:

- $v_1 \neq 0$: Assuming $v_1 \neq 0$ yields the eigenpair $(\lambda^{(2)}, \mathbf{v}^{(2)})$:

$$\lambda^{(2)} = J_{11} < 0 \quad v_1^{(2)} \in \mathbb{C} \setminus \{0\} \quad v_2^{(2)} = 0 \quad v_3^{(2)} = \frac{J_{31}}{J_{11} - J_{33}}v_1^{(2)} \quad (\text{E.54})$$

- $v_1 = 0$: In this case, the equation system collapses to the single equation $(J_{33} - \lambda)v_3 = 0$, where $v_3 \neq 0$ because not all entries of an eigenvector can be zero. Hence, $\mathbf{J}_{f^\circ}(\mathbf{0})$ has the following eigenpair $(\lambda^{(3)}, \mathbf{v}^{(3)})$:

$$\begin{aligned} \lambda^{(3)} &= J_{33} \stackrel{\text{(E.38)}}{=} -\frac{1}{\tilde{s}} < 0 & v_1^{(3)} &= 0 \\ v_2^{(3)} &= 0 & v_3^{(3)} &\in \mathbb{C} \setminus \{0\} \end{aligned} \quad (\text{E.55})$$

Summary of eigenpairs. The Jacobian matrix $\mathbf{J}_{f^\circ}(\mathbf{0})$ thus has zero and negative eigenvalues. For convenience, we categorize the eigenvectors into sets corresponding to zero and negative eigenvalues, respectively:

$$V^0 = \{\mathbf{v}^{(1)}\} \quad V^- = \{\mathbf{v}^{(2)}, \mathbf{v}^{(3)}\} \quad (\text{E.56})$$

The presence of both zero and negative eigenvalues means that the stability properties of the original nonlinear dynamic system $\dot{\mathbf{z}} = f^\circ(\mathbf{z})$ cannot be derived from the linearized system $\dot{\mathbf{z}} = \mathbf{J}_{f^\circ}(\mathbf{0})\mathbf{z}$ [249]. Instead, higher-order terms need to be investigated with respect to stability.

E.2.2 Dimension Reduction via the Center Manifold

While the linearized system is inconclusive about the desired stability properties, it allows some insight into the dynamics of the nonlinear system when using center-manifold theory.

E.2.2.1 Center-Manifold Properties

To introduce the center manifold, we note that the subspace spanned by the eigenvectors in V^- is the *stable subspace* of the Jacobian $J_{f^0}(\mathbf{0})$, and the subspace spanned by the eigenvectors in V^0 is the *center subspace* [257]. Each of these subspaces is associated with a manifold that has the same dimension as the corresponding subspace, and is tangential to the corresponding subspace at the equilibrium [108].

Since our system has only has a stable manifold and a center manifold (and no unstable manifold, which would be associated with positive eigenvalues), we can use the *center-manifold emergence theorem* [128]. This theorem states that given a starting point sufficiently close to the center manifold, the dynamics converge exponentially quickly to the center manifold, and thus approach a trajectory on the center manifold. The overall dynamics of the nonlinear system can thus be approximated by the dynamics on the center manifold, which have lower dimension and thus allow a more tractable analysis.

E.2.2.2 Center-Manifold Dynamics

To derive the dynamics on the center manifold, we start by decoupling the system state \mathbf{z} along subspaces, i.e., we transform it onto a different basis such that every variable only effects a change along either the center or the stable subspace. We achieve this by a coordinate transformation using the eigenbasis:

$$\mathbf{z} = \mathbf{T}\zeta = [\mathbf{v}^{(1)} \ \mathbf{v}^{(2)} \ \mathbf{v}^{(3)}]\zeta = \begin{bmatrix} T_{11} & T_{12} & 0 \\ T_{21} & 0 & 0 \\ T_{31} & T_{32} & T_{33} \end{bmatrix} \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{bmatrix} \quad (\text{E.57})$$

$$\iff \zeta = \mathbf{T}^{-1}\mathbf{z},$$

where ζ_1 is the variable associated with the center subspace, and ζ_2 and ζ_3 are the variables associated with the stable subspace. From the structure of \mathbf{T} , we see that the center variable ζ_1 can be expressed exclusively by z_2 (corresponding to w^{\max}):

$$z_2 = T_{21}\zeta_1 \iff \zeta_1 = \frac{z_2}{T_{21}}. \quad (\text{E.58})$$

According to the center-manifold existence theorem, a manifold Γ_c exists with the following properties around the equilibrium $\mathbf{0}$ [257]:

$$\Gamma_c = \{(\zeta_1, \zeta_2, \zeta_3) \mid \zeta_2 = h_2(\zeta_1), h_2(0) = 0, h_2'(0) = 0\}, \quad (\text{E.59})$$

$$\zeta_3 = h_3(\zeta_1), h_3(0) = 0, h_3'(0) = 0 \}$$

where both h_2 and h_3 are one-dimensional functions of the form:

$$h(\zeta_1) = \sum_{i=2}^{\infty} a_i \zeta_1^i. \quad (\text{E.60})$$

Intuitively, in our case, the center manifold Γ_c corresponds to a curve in three-dimensional space, which is tangential to the center subspace at the equilibrium $\mathbf{0}$ and is fully describable by the center variable ζ_1 . The overall dynamics of the nonlinear system move along that curve, although it is not clear yet whether towards or away from the equilibrium. To identify this direction, we only need to consider the following one-dimensional dynamics along the center manifold:

$$\dot{\zeta}_1 \stackrel{(\text{E.58})}{=} \frac{1}{T_{21}} \dot{z}_2(\mathbf{z}) \stackrel{(\text{E.57})}{=} \frac{1}{T_{21}} \dot{z}_2(\mathbf{T}\zeta') \quad (\text{E.61})$$

where

$$\zeta' \stackrel{(\text{E.59})}{=} \begin{bmatrix} \zeta_1 \\ h_2(\zeta_1) \\ h_3(\zeta_1) \end{bmatrix} \stackrel{(\text{E.58})}{=} \begin{bmatrix} z_2/T_{21} \\ h_2(z_2/T_{21}) \\ h_3(z_2/T_{21}) \end{bmatrix} \quad (\text{E.62})$$

E.2.3 Stability Investigation of the Lower-Dimensional Dynamics

So far, we have reduced the full dynamics to the lower-dimensional center-manifold dynamics, which we can now investigate.

E.2.3.1 Taylor Expansion of Center-Manifold Dynamics

The one-dimensional center-manifold dynamics in Eq. (E.61) suggests that the dynamics of \dot{z}_2 need to be investigated. These dynamics are already given in Eq. (E.26). However, these dynamics must be investigated on the center manifold, which requires expressing z_1 and z_3 as center-manifold functions of z_2 :

$$\hat{\mathbf{z}}(z_2) = \mathbf{T}\zeta' \stackrel{(\text{E.62})}{=} \begin{bmatrix} T_{11} \frac{z_2}{T_{21}} + T_{12} h_2\left(\frac{z_2}{T_{21}}\right) \\ z_2 \\ T_{31} \frac{z_2}{T_{21}} + T_{32} h_2\left(\frac{z_2}{T_{21}}\right) + T_{33} h_3\left(\frac{z_2}{T_{21}}\right) \end{bmatrix}. \quad (\text{E.63})$$

Hence, we investigate the following dynamics:

$$\dot{z}_2 = (W^\circ(z_2, \hat{z}_3(z_2)) - (z_2 + \tilde{w}^{\max})) \cdot \frac{W^\circ(z_2, \hat{z}_3(z_2))}{\tilde{\tau}_k} \cdot p^\circ(\hat{z}(z_2)), \quad (\text{E.64})$$

where

Around the equilibrium $\tilde{z}_2 = 0$, the evolution function \dot{z}_2 can be approximated via a Taylor expansion of third degree:

$$\dot{z}_2(z'_2) \approx \dot{z}_2(0) + \frac{\partial \dot{z}_2}{\partial z_2}(0) \cdot z'_2 + \frac{\partial^2 \dot{z}_2}{\partial z_2^2}(0) \cdot \frac{{z'_2}^2}{2!} + \frac{\partial^3 \dot{z}_2}{\partial z_2^3}(0) \cdot \frac{{z'_2}^3}{3!}. \quad (\text{E.65})$$

By using the Symbolic Math Toolbox of Matlab [172], this approximated function can be reduced to the following expression:

$$\dot{z}_2(z'_2) \approx \frac{c\tilde{w}^{\max}T_{31}^3 p^\circ(\mathbf{0})}{\tilde{\tau}_k T_{21}^3} \cdot {z'_2}^3 = K{z'_2}^3. \quad (\text{E.66})$$

It is easy to see that asymptotic stability of this system requires a negative K : If $z'_2 < 0$ (below the equilibrium), z'_2 would be positively affected (multiplication of negative numbers) and thus drawn closer to the equilibrium at 0, whereas if $z'_2 > 0$, z'_2 would be negatively affected, and thereby also attracted to the equilibrium at 0. To prove asymptotic stability, we thus have to show:

$$K < 0 \iff \frac{c\tilde{w}^{\max}T_{31}^3 p^\circ(\mathbf{0})}{\tilde{\tau}_k T_{21}^3} < 0 \iff \frac{c\tilde{w}^{\max}v_3^{(1)3} p^\circ(\mathbf{0})}{\tilde{\tau}_k v_2^{(1)3}} < 0 \quad (\text{E.67})$$

where $v^{(1)}$ fulfills the condition in Eq. (E.52).

Without loss of generality, we set $v_1^{(1)} < 0$, which implies (together with $J_{11} < 0$ and $J_{12} < 0$):

$$v_2^{(1)} \stackrel{(\text{E.52})}{=} -\frac{J_{11}}{J_{12}} v_1^{(1)} > 0. \quad (\text{E.68})$$

Since c , \tilde{w}^{\max} , $p^\circ(\mathbf{0})$, and $\tilde{\tau}_k$ are all known to be positive, asymptotic stability depends on the following condition:

$$v_3^{(1)3} < 0 \iff v_3^{(1)} < 0 \stackrel{(\text{E.52})}{\iff} \frac{J_{11}J_{32} - J_{12}J_{31}}{J_{12}J_{33}} v_1^{(1)} < 0$$

$$\iff \frac{J_{11}J_{32} - J_{12}J_{31}}{J_{12}J_{33}} > 0. \quad (\text{E.69})$$

Since $J_{12} < 0$ and $J_{33} < 0$, we can simplify this condition even further to:

$$J_{11}J_{32} > J_{12}J_{31} \iff J_{11} < \frac{J_{31}}{J_{32}}J_{12}. \quad (\text{E.70})$$

To check whether this condition holds, we again perform a case distinction regarding χ , i.e., we distinguish the cases $\chi \leq C_\ell - \tilde{w}^{\max}/(\alpha\tilde{\tau}_k)$ and $\chi > C_\ell - \tilde{w}^{\max}/(\alpha\tilde{\tau}_k)$.

E.2.3.2 $\chi \leq C_\ell - \tilde{w}^{\max}/(\alpha\tilde{\tau}_k)$

In this case, the expanded form of Eq. (E.70) can be simplified by the finding in Eq. (E.40):

$$\begin{aligned} J_{11} &\stackrel{(\text{E.34})}{=} \frac{\alpha C_\ell}{\tilde{\tau}_k \left(\alpha \tilde{x}^{\text{btl}} + \frac{1}{\tilde{\tau}_k} \tilde{w}^{\max} \right)^2} \cdot \tilde{w}^{\max} - 1 \stackrel{(\text{E.70})}{<} \\ &- \frac{J_{31}}{J_{32}} \frac{\alpha C_\ell}{\tilde{\tau}_k \left(\alpha \tilde{x}^{\text{btl}} + \frac{1}{\tilde{\tau}_k} \tilde{w}^{\max} \right)^2} \cdot \tilde{x}^{\text{btl}} \stackrel{(\text{E.35})}{=} \frac{J_{31}}{J_{32}} J_{12} \end{aligned} \quad (\text{E.71})$$

$$\stackrel{(\text{E.40})}{\iff} \frac{1}{\tilde{\tau}_k \alpha C_\ell} \cdot \tilde{w}^{\max} - 1 < - \frac{J_{31}}{J_{32}} \frac{1}{\tilde{\tau}_k \alpha C_\ell} \cdot \tilde{x}^{\text{btl}} \quad (\text{E.72})$$

$$\stackrel{\cdot \tilde{\tau}_k \alpha C_\ell / \tilde{x}^{\text{btl}}}{\iff} (\tilde{w}^{\max} - \tilde{\tau}_k \alpha C_\ell) \cdot \frac{1}{\tilde{x}^{\text{btl}}} < - \frac{J_{31}}{J_{32}} \quad (\text{E.73})$$

$$\stackrel{(\text{E.39})}{\iff} (\tilde{w}^{\max} - \tilde{\tau}_k \alpha C_\ell) \cdot \frac{1}{C_\ell - \frac{\tilde{w}^{\max}}{\tilde{\tau}_k \alpha}} < - \frac{J_{31}}{J_{32}} \quad (\text{E.74})$$

$$\stackrel{\text{expand}}{\iff} (\tilde{w}^{\max} - \tilde{\tau}_k \alpha C_\ell) \cdot \frac{\tilde{\tau}_k \alpha}{\tilde{\tau}_k \alpha C_\ell - \tilde{w}^{\max}} < - \frac{J_{31}}{J_{32}} \quad (\text{E.75})$$

$$\stackrel{\text{cancel}}{\underset{-1}{\iff}} \tilde{\tau}_k \alpha > \frac{J_{31}}{J_{32}} \stackrel{1/}{\iff} \frac{J_{32}}{J_{31}} > \frac{1}{\tilde{\tau}_k \alpha}. \quad (\text{E.76})$$

In turn, expanding the ratio J_{32}/J_{31} yields:

$$\begin{aligned} \frac{J_{32}}{J_{31}} &\stackrel{(\text{E.37})}{=} - \left(\frac{\tilde{s}}{\tilde{\tau}_k} \cdot p^\circ(\mathbf{0}) + \frac{\tilde{s}\tilde{w}^{\max}}{\tilde{\tau}_k^2} \frac{C_\ell}{\left(\beta \tilde{x}^{\text{btl}} + \frac{1}{\tilde{\tau}_k} \tilde{w}^{\max} \right)^2} \right). \\ &\quad \left(- \frac{\tilde{\tau}_k \left(\beta \tilde{x}^{\text{btl}} + \frac{1}{\tilde{\tau}_k} \tilde{w}^{\max} \right)^2}{\beta C_\ell \tilde{s}\tilde{w}^{\max}} \right) \end{aligned} \quad (\text{E.77})$$

$$= \frac{p^\circ(\mathbf{0}) \left(\beta \tilde{x}^{\text{btl}} + \frac{1}{\tilde{\tau}_k} \tilde{w}^{\max} \right)^2}{\beta C_\ell \tilde{w}^{\max}} + \frac{1}{\tilde{\tau}_k \beta} > \frac{1}{\tilde{\tau}_k \alpha}. \quad (\text{E.78})$$

The bound holds because $\beta \leq \alpha$ and the first term on the LHS in Eq. (E.78) is strictly positive. Hence, also Eq. (E.70) holds.

E.2.3.3 $\chi > C_\ell - \tilde{w}^{\max}/(\alpha \tilde{\tau}_k)$

In this case, $\tilde{x}^{\text{btl}} \stackrel{(\text{E.42})}{=} \chi$ and the expanded form of Eq. (E.70) is thus:

$$\frac{\alpha C_\ell}{\tilde{\tau}_k \left(\alpha \chi + \frac{1}{\tilde{\tau}_k} \tilde{w}^{\max} \right)^2} \cdot \tilde{w}^{\max} - 1 \stackrel{(\text{E.70})}{<} - \frac{J_{31}}{J_{32}} \frac{\alpha C_\ell \chi}{\tilde{\tau}_k \left(\alpha \chi + \frac{1}{\tilde{\tau}_k} \tilde{w}^{\max} \right)^2} \quad (\text{E.79})$$

$$\iff \frac{J_{32}}{J_{31}} > \frac{\alpha C_\ell \tilde{\tau}_k \chi}{\tilde{w}^{\max 2} + \alpha(2\chi - C_\ell) \tilde{\tau}_k \tilde{w}^{\max} + \alpha^2 C_\ell^2 \chi^2} \quad (\text{E.80})$$

Expanding the ratio J_{32}/J_{31} yields:

$$\begin{aligned} & \frac{J_{32}}{J_{31}} \stackrel{(\text{E.37})}{=} - \left(\frac{\tilde{s}}{\tilde{\tau}_k} \cdot p^\circ(\mathbf{0}) + \frac{\tilde{s} \tilde{w}^{\max}}{\tilde{\tau}_k^2} \frac{C_\ell}{\left(\beta \chi + \frac{1}{\tilde{\tau}_k} \tilde{w}^{\max} \right)^2} \right). \\ & \left(- \frac{\tilde{\tau}_k \left(\beta \chi + \frac{1}{\tilde{\tau}_k} \tilde{w}^{\max} \right)^2}{\beta C_\ell \tilde{s} \tilde{w}^{\max}} \right) \\ & = \frac{\tilde{\tau}_k \left(\beta \chi + \frac{1}{\tilde{\tau}_k} \tilde{w}^{\max} \right)^2}{\beta C_\ell \tilde{s} \tilde{w}^{\max 2}} + \frac{1}{\beta \tilde{\tau}_k} > \frac{1}{\beta \tilde{\tau}_k}. \end{aligned} \quad (\text{E.81})$$

Thanks to this neat lower bound of J_{32}/J_{31} , Eq. (E.80) is implied by the following stronger condition:

$$\frac{J_{32}}{J_{31}} \stackrel{(\text{E.81})}{>} \frac{1}{\beta \tilde{\tau}_k} > \frac{\alpha C_\ell \tilde{\tau}_k \chi}{\tilde{w}^{\max 2} + \alpha(2\chi - C_\ell) \tilde{\tau}_k \tilde{w}^{\max} + \alpha^2 C_\ell^2 \chi^2} \quad (\text{E.82})$$

$$\begin{aligned} & \iff \frac{1}{\beta \tilde{\tau}_k} \tilde{w}^{\max 2} + \frac{\alpha(2\chi - C_\ell)}{\beta} \tilde{w}^{\max} + \frac{\tilde{\tau}_k \alpha^2 \chi^2}{\beta} - \tilde{\tau}_k \alpha C_\ell \chi \\ & =: \Psi(\tilde{w}^{\max}) > 0. \end{aligned} \quad (\text{E.83})$$

To verify Eq. (E.83) for all $\tilde{w}^{\max} > \tilde{\tau}_k \alpha (C_\ell - \chi)$, we first note that Ψ is convex. Hence, if Ψ has no roots, the convexity of Ψ implies that $\Psi(\tilde{w}^{\max}) > 0$ holds for

any \tilde{w}^{\max} . In contrast, if Ψ has roots, $\Psi(\tilde{w}^{\max}) > 0$ holds for any \tilde{w}^{\max} above the upper root ψ of Ψ , which is:

$$\psi = \frac{\tilde{\tau}_k}{2} \left(\alpha(C_\ell - 2\chi) + \sqrt{\alpha^2 C_\ell^2 + 4\alpha C_\ell (\beta - \alpha)\chi} \right) \quad (\text{E.84})$$

The truth of $\Psi(\tilde{w}^{\max}) > 0 \forall \tilde{w}^{\max} > \tilde{\tau}_k \alpha(C_\ell - \chi)$ is confirmed by the fact that ψ is below the relevant area of argument \tilde{w}^{\max} :

$$\begin{aligned} \psi &\leq \tilde{\tau}_k \alpha(C_\ell - \chi) \stackrel{2/\tilde{\tau}_k}{\iff} \\ &\alpha(C_\ell - 2\chi) + \sqrt{\alpha^2 C_\ell^2 + 4\alpha C_\ell (\beta - \alpha)\chi} \leq 2\alpha(C_\ell - \chi) \\ &\stackrel{-\alpha(C_\ell - 2\chi)}{\iff} \sqrt{\alpha^2 C_\ell^2 + 4\alpha C_\ell (\beta - \alpha)\chi} \leq \alpha C_\ell \\ &\stackrel{(\cdot)^2}{\iff} 4\alpha C_\ell (\beta - \alpha)\chi \leq 0 \stackrel{\beta \leq \alpha}{\iff}_{\chi > 0} \top. \end{aligned} \quad (\text{E.85})$$

E.2.3.4 Conclusion

In conclusion, the relation between the Jacobian entries J_{11} , J_{12} , J_{31} , and J_{32} as given in Eq. (E.70) ensures a negative third entry $v_3^{(1)}$ in the center eigenvector, which is required for a negative coefficient in the third-order Taylor expansion of the center-manifold dynamics in Eq. (E.67). The negativity of this coefficient guarantees asymptotic stability of the center-manifold dynamics in Eq. (E.64), and thus of the full dynamic system.

E.3 PROOF OF THEOREM 6.3: BBR-CUBIC OSCILLATION

E.3.1 Update Functions

We start the proof by characterizing the update function w^\leftarrow used in the discrete-time process in Eq. (6.24). Conceptually, this function $w^\leftarrow(w)$ yields the CUBIC window size after an interval of the short-term dynamics given BBR probing strength α , which in turn results from CUBIC window size w at the beginning of the interval. We denote this resulting probing strength by $\alpha^\leftarrow(w)$, and characterize $\alpha^\leftarrow(w)$ in §E.3.1.1 before characterizing w^\leftarrow in §E.3.1.2.

E.3.1.1 α -Update Function α^{\leftarrow}

From Eq. (6.17) in the fluid-equilibrium model in §6.3.1, we know that α is determined by the queue length q_{ℓ}^- that remains when the BBR flow backs off:

$$\alpha^{\leftarrow}(w) = \min(5/4, \hat{\alpha}^{\leftarrow}(w)) \quad (\text{E.86})$$

$$\text{where } \hat{\alpha}^{\leftarrow}(w) = \frac{2}{\tau_i} \left(\tau_i^P + \frac{q_{\ell}^-(w)}{C_{\ell}} \right) \quad (\text{E.87})$$

$$= \frac{2}{\tau_i} \left(\tau_i^P + \frac{[4 + (1-b)w - \tau_{\ell}^P C_{\ell}]_0^{B_{\ell}}}{C_{\ell}} \right). \quad (\text{E.88})$$

This function $\alpha^{\leftarrow}(w)$ achieves its minimum value α_{\min} if $q_{\ell}^- = 0$, which is achieved for all CUBIC window sizes $w \leq w_0$, where:

$$w_0 = \frac{1}{1-b} (C_{\ell} \tau_{\ell}^P - 4) \implies \alpha_{\min} = \alpha^{\leftarrow}(w_0) = \frac{2\tau_i^P}{\tau_i}. \quad (\text{E.89})$$

Conversely, the maximum α_{\max} is not achieved for $q_{\ell}^- = B_{\ell}$, as in this case $\hat{\alpha}^{\leftarrow} = 2\tau_i/\tau_i = 2$. Instead, the maximum value α_{\max} is achieved for all window sizes $w \geq w_1$, where

$$\begin{aligned} w_1 &= \frac{1}{1-b} \left(C_{\ell} \left(\frac{5}{8} \tau_i + \tau_{\ell}^P - \tau_i^P \right) - 4 \right) \implies \hat{\alpha}^{\leftarrow}(w_1) = 5/4 \\ &\implies \alpha^{\leftarrow}(w_1) = 5/4 = \alpha_{\max}. \end{aligned} \quad (\text{E.90})$$

Given the above value range, we rewrite the function α^{\leftarrow} such that its constant and linear pieces become obvious:

$$\alpha^{\leftarrow}(w) = \begin{cases} \alpha_{\min} = \frac{2\tau_i^P}{\tau_i} & \text{if } w \leq w_0, \\ \hat{\alpha}(w) = \frac{2}{\tau_i} \left(\tau_i^P + \frac{w+4-\tau_{\ell}^P C_{\ell}}{C_{\ell}} \right) & \text{if } w \in (w_0, w_1), \\ \alpha_{\max} = 5/4 & \text{if } w \geq w_1. \end{cases} \quad (\text{E.91})$$

This function α^{\leftarrow} is illustrated in Fig. E.2.

E.3.1.2 Window-Update Function w^{\leftarrow}

The window-update function $w^{\leftarrow}(w)$ yields the CUBIC window size after an interval with initial window size w and corresponding BBR probing strength $\alpha =$

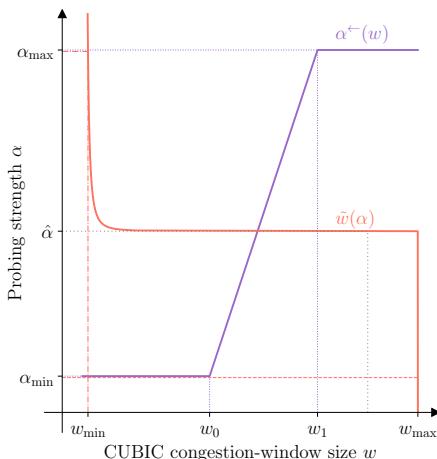


Figure E.2: Illustration of functions $\alpha^{\leftarrow}(w)$ and $\tilde{w}(\alpha)$.

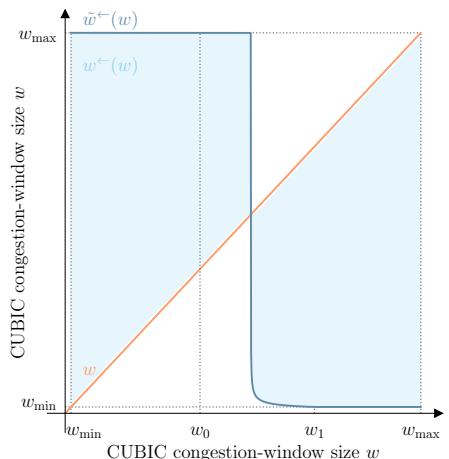


Figure E.3: Illustration of function $\tilde{w}^{\leftarrow}(w)$, and value range for $w^{\leftarrow}(w)$.

$\alpha^{\leftarrow}(w)$. Since this new CUBIC window size is the result of convergence towards the short-term equilibrium window size $\tilde{w}(\alpha)$, we first revisit $\tilde{w}(\alpha)$ based on Theorem 6.1. While this function does not have a closed-form representation, we can still conclude that $\tilde{w}(\alpha)$ has a two-part structure. In particular, we know that an inflection point $\hat{\alpha} > 1$ exists such that $\tilde{w}(\alpha)$ is found by solving the equation $\tilde{S}_2(s) = 0$ (cf. Eq. (6.23)) for all $\alpha < \hat{\alpha}$, and found by solving $\tilde{S}_1(s) = 0$ (cf. Eq. (6.22)) for all $\alpha \geq \hat{\alpha}$ (cf. Appendix E.1). Crucially, $\tilde{w}(\alpha)$ can be further confirmed to be strictly monotonically decreasing in α (cf. SE.3.1.3).

Since $\tilde{w}(\alpha)$ is thus monotonically decreasing on its complete domain $[0, \alpha_{\max}]$, the value range of \tilde{w} for that domain is bounded to $[w_{\min}, w_{\max}]$, where

$$w_{\min} = \tilde{w}(\alpha_{\max}) \stackrel{(E.90)}{=} \tilde{w}(\alpha^{\leftarrow}(w_1)) = \tilde{w}^{\leftarrow}(w_1) \quad (E.92)$$

$$w_{\max} = \tilde{w}(\alpha_{\min}) \stackrel{(E.89)}{=} \tilde{w}(\alpha^{\leftarrow}(w_0)) = \tilde{w}^{\leftarrow}(w_0) \quad (E.93)$$

The function $\tilde{w}(\alpha)$ is visualized in Fig. E.2.

Note that the above equation also introduces the function $\tilde{w}^\leftarrow(w) = \tilde{w}(\alpha^\leftarrow(w))$, which is of great importance throughout the rest of the proof. This function \tilde{w}^\leftarrow can be explicitly represented based on Eq. (E.91):

$$\tilde{w}^\leftarrow(w) = \begin{cases} w_{\max} & \text{if } w \leq w_0, \\ \tilde{w}(\alpha^\leftarrow(w)) & \text{if } w \in (w_0, w_1), \\ w_{\min} & \text{if } w \geq w_1, \end{cases} \quad (\text{E.94})$$

where w_0 and w_1 are such that \tilde{w}^\leftarrow is *strictly* monotonically decreasing for arguments $w \in [w_0, w_1]$. In summary, the function $\tilde{w}^\leftarrow(w)$ is monotonically decreasing for all w :

$$\forall w \in \mathbb{R}. \quad \frac{\partial \tilde{w}^\leftarrow(w)}{\partial w} \leq 0. \quad (\text{E.95})$$

Based on Eq. (E.94), we can finally represent $w^\leftarrow(w)$. For every interval, w^\leftarrow describes the convergence from the interval-start CUBIC window size w towards its associated equilibrium $\tilde{w}(\alpha^\leftarrow(w)) = \tilde{w}^\leftarrow(w)$, which follows from the asymptotic stability of any $\tilde{w}(\alpha)$ proven in Theorem 6.2:

$$w^\leftarrow(w) \in \begin{cases} (w, \tilde{w}^\leftarrow(w)] & \text{if } w < \tilde{w}^\leftarrow(w), \\ [w] = [\tilde{w}^\leftarrow(w)] & \text{if } w = \tilde{w}^\leftarrow(w), \\ [\tilde{w}^\leftarrow(w), w) & \text{if } w > \tilde{w}^\leftarrow(w). \end{cases} \quad (\text{E.96})$$

The function \tilde{w}^\leftarrow and the value ranges for $w^\leftarrow(w)$ are illustrated in Fig. E.3.

E.3.1.3 Strict Monotonic Decrease of \tilde{w}

To confirm that $\tilde{w}(\alpha)$ is strictly monotonically decreasing for all α , we distinguish the cases $\alpha \geq \hat{\alpha}$ and $\alpha < \hat{\alpha}$.

$\alpha \geq \hat{\alpha}$. First, we note that $\tilde{w} = \tilde{w}^{\max}$ is strictly monotonically increasing in \tilde{s} (Eq. (6.9)). In turn, $\tilde{s}(\alpha)$ is strictly monotonically decreasing in α if all α_0 and α_1 with $\hat{\alpha} \leq \alpha_0 < \alpha_1$ satisfy:

$$\tilde{s}(\alpha_0) > \tilde{s}(\alpha_1) \iff \forall s > s_0(\tau_k). \tilde{S}_1(s, \alpha_0) < \tilde{S}_1(s, \alpha_1), \quad (\text{E.97})$$

where $s_0(\tau_k)$ is the lower bound on all roots of polynomial \tilde{S}_1 from Theorem 6.1, given delay τ_k for the CUBIC flow. This lower bound is found by setting $C_\ell = 0$, which yields:

$$\tilde{S}_1(s_0) = \frac{\alpha - 1}{\alpha} \left(\frac{c^2}{b\tau_k} s_0^7 - cs_0^3 \right) \stackrel{!}{=} 0 \iff s_0 = \sqrt[4]{\frac{b\tau_k}{c}}. \quad (\text{E.98})$$

Returning to Eq. (E.97), we note that

$$\begin{aligned} & \tilde{S}_1(s, \alpha_0) < \tilde{S}_1(s, \alpha_1) \\ \iff & \frac{\alpha_0 - 1}{\alpha_0} \left(\frac{c^2}{b\tau_k} s^7 - cs^3 \right) < \frac{\alpha_1 - 1}{\alpha_1} \left(\frac{c^2}{b\tau_k} s^7 - cs^3 \right) \end{aligned} \quad (\text{E.99})$$

holds if the s -polynomial in the parentheses is strictly positive, which is again true for all $s > s_0(\tau_k)$. This insight implies $\tilde{s}(\alpha_0) > \tilde{s}(\alpha_1)$ and also $\tilde{w}(\alpha_0) > \tilde{w}(\alpha_1)$, i.e., $\tilde{w}(\alpha)$ is strictly monotonically decreasing for all $\alpha \geq \hat{\alpha}$.

$\alpha < \hat{\alpha}$. The demonstration of strict negative monotonicity for $\alpha < \hat{\alpha}$ works analogously to the previous case, but is based on \tilde{S}_2 instead of \tilde{S}_1 . Intriguingly, we again find $s_0(\tau_k)$ as in Eq. (E.98), marking both the lower bound on possible roots and the lower bound on roots whose position is a strictly monotonically decreasing function of α . Hence, \tilde{w} is a strictly monotonically decreasing function for all α .

E.3.2 CUBIC Long-Term Equilibrium Window Size \bar{w}

Given the update functions from Appendix E.3.1, we can characterize the equilibrium of the discrete dynamic process in Eq. (6.24). In particular, we will prove the existence and the uniqueness of an equilibrium \bar{w} such that $\bar{w} = w^\leftarrow(\bar{w})$. From Eq. (E.96), we learn that $w^\leftarrow(w)$ only returns the argument w as a value if $w = \tilde{w}^\leftarrow(w)$. Hence, the equilibrium condition can only hold if

$$\bar{w} = \tilde{w}^\leftarrow(\bar{w}). \quad (\text{E.100})$$

Since \tilde{w}^\leftarrow is monotonically decreasing (Eq. (E.95)), we know that its value range $\mathcal{R}(\tilde{w}^\leftarrow)$ given argument range $[w_{\min}, w_{\max}]$ is

$$\mathcal{R}(\tilde{w}^\leftarrow) = [\tilde{w}^\leftarrow(w_{\max}), \tilde{w}^\leftarrow(w_{\min})] \subseteq [w_{\min}, w_{\max}] \quad (\text{E.101})$$

where the subset relation holds for the following reason (symmetric for w_{\min} and w_0):

$$\text{If } w_{\max} < w_1 : \quad \tilde{w}^\leftarrow(w_{\max}) \stackrel{\text{Eq. (E.95)}}{\geq} \tilde{w}^\leftarrow(w_1) \stackrel{\text{Eq. (E.92)}}{=} w_{\min}$$

$$\text{If } w_{\max} \geq w_1 : \quad \tilde{w}^{\leftarrow}(w_{\max}) \stackrel{\text{Eq. (E.94)}}{=} w_{\min} \quad (\text{E.102})$$

In a next step, we consider the fixed-point function $w^*(w) = w - \tilde{w}^{\leftarrow}(w)$. Since any equilibrium \bar{w} must satisfy $\bar{w} = \tilde{w}^{\leftarrow}(\bar{w})$, it must satisfy $w^*(\bar{w}) = 0$. Note that $w^*(w)$ is now *strictly monotonically increasing*. The value range $\mathcal{R}(w^*)$ for argument range $[w_{\min}, w_{\max}]$ is therefore:

$$\mathcal{R}(w^*) = [w_{\min} - \tilde{w}^{\leftarrow}(w_{\min}), w_{\max} - \tilde{w}^{\leftarrow}(w_{\max})]. \quad (\text{E.103})$$

Eq. (E.101) suggests that $w_{\min} \leq \tilde{w}^{\leftarrow}(w_{\max}) \leq \tilde{w}^{\leftarrow}(w_{\min})$ and $w_{\max} \geq \tilde{w}^{\leftarrow}(w_{\min}) \geq \tilde{w}^{\leftarrow}(w_{\max})$. As a result, we find that $w_{\min} - \tilde{w}^{\leftarrow}(w_{\min}) \leq 0$ and $w_{\max} - \tilde{w}^{\leftarrow}(w_{\min}) \geq 0$. This finding, together with the continuousness and strict monotonicity of w^* , imply (by the intermediate-value theorem) that a unique $\bar{w} \in [w_{\min}, w_{\max}]$ exists such that $w^*(\bar{w}) = 0 \iff \bar{w} = \tilde{w}^{\leftarrow}(\bar{w})$. This \bar{w} is thus a unique equilibrium which is guaranteed to exist.

E.3.3 Convergence Trajectories

Given the unique equilibrium \bar{w} , we now investigate the stability properties of this equilibrium. In other words, we characterize the convergence trajectories determined by the discrete process in Eq. (6.24), and elicit a condition under which these trajectories may or may not lead to the equilibrium. This stability property depends on the neighborhood of the equilibrium \bar{w} with respect to the update function w^{\leftarrow} .

In particular, we now show that the equilibrium \bar{w} is unstable if an equilibrium neighborhood Ω exists such that the window-update function w has a slope of less than -1 in the complete neighborhood:

$$\exists \Omega = (\omega_0, \omega_1), \omega_0 < \bar{w} < \omega_1. \quad \forall \omega \in \Omega. \frac{\partial w^{\leftarrow}(\omega)}{\partial w} < -1 \quad (\text{E.104})$$

In other words, we will show that all evolution trajectories from a window size $\omega \in \Omega$ in this neighborhood lead out of the neighborhood. In that case, even if the dynamics outside the neighborhood converge into the neighborhood, the trajectory then leaves the neighborhood again and the equilibrium is not converged upon.

In particular, we consider a state $w(t)$ with $w(t) \in \Omega$ and $w(t) < \bar{w}$ (The proof for $w(t) > \bar{w}$ is symmetric). Since both $w(t)$ and \bar{w} are in Ω , w^{\leftarrow} is strictly monotonically decreasing between $w(t)$ and \bar{w} , and therefore

$$w(t) < \bar{w} \implies w(t+1) = w^{\leftarrow}(w(t)) > w^{\leftarrow}(\bar{w}) \stackrel{\text{(E.100)}}{=} \bar{w}. \quad (\text{E.105})$$

At this point, $w(t+1)$ might already be outside of Ω if $w(t+1) > \omega_1$. Otherwise, if $w(t+1) \in \Omega$, we know by the same argument that $w(t+2) = w^\leftarrow(w(t+1)) < \bar{w}$. Again, $w(t+2)$ might be out of the neighborhood Ω if $w(t+2) < \omega_0$.

Importantly, w will *eventually* leave the neighborhood if for all any t , $w(t+2) < w(t)$, i.e., the window state moves away from the equilibrium \bar{w} and will eventually fall below ω_0 . This condition can be reformulated:

$$\begin{aligned} w(t+2) < w(t) &\stackrel{\text{Expand}}{\iff} w^\leftarrow(w^\leftarrow(w(t))) < w(t) \stackrel{-w^\leftarrow(w(t))}{\iff} \\ w^\leftarrow(w^\leftarrow(w(t))) - w^\leftarrow(w(t)) &< w(t) - w^\leftarrow(w(t)) \\ \stackrel{\text{RHS}}{\iff} \frac{w^\leftarrow(w^\leftarrow(w(t))) - w^\leftarrow(w(t))}{w(t) - w^\leftarrow(w(t))} &> 1 \\ \stackrel{(E.105)}{\iff} \frac{w^\leftarrow(w^\leftarrow(w(t))) - w^\leftarrow(w(t))}{w^\leftarrow(w(t)) - w(t)} &< -1. \end{aligned} \quad (\text{E.106})$$

The last inequality in Eq. (E.106) can be understood as a condition on the average slope of w^\leftarrow in $[w(t), w^\leftarrow(w(t))]$. Since we are considering the case where both $w(t)$ and $w^\leftarrow(w(t))$ are in the neighborhood Ω , the derivative condition in Eq. (E.104) implies this average-slope condition.

E.4 PROOF OF THEOREM 6.4: MAXIMALLY INTENSIVE OSCILLATION

E.4.1 Maximum Oscillation Amplitude

In order to find plausible bounds of the flow-size distribution in BBR-CUBIC competition, we consider the worst case in terms of oscillation amplitude. To characterize this worst-case oscillation amplitude, we make two observations.

Convergence speed. We note that the oscillation amplitude is proportional to the convergence speed of the short-term dynamics in each update interval: In interval $[t, t+1]$, these short-term dynamics involve a CUBIC window-size change from the initial CUBIC window size $w(t)$ to window size $w(t+1) = w^\leftarrow(w(t))$, where $w^\leftarrow(w(t))$ is between the initial window size $w(t)$ and the short-term equilibrium window size $\tilde{w}^\leftarrow(w(t))$. Hence, a high similarity between w^\leftarrow and \tilde{w}^\leftarrow indicates a high convergence speed, quick window-size changes, and thus a large amplitude of the oscillation. In the following proof, we therefore assume $w^\leftarrow = \tilde{w}^\leftarrow$ to maximize the oscillation amplitude.

Neighborhood size. We note that the oscillation amplitude is proportional to the size of the unstable neighborhood $\Omega = [\omega_0, \omega_1]$ of the long-term equilibrium window size \bar{w} . Since the process $\{w(t)\}_{t \in \mathbb{N}, t \geq 0}$ evolves away from this equilibrium \bar{w} when in neighborhood Ω , the oscillation mostly involves window sizes $w \neq \Omega$ left and right of the neighborhood. Moreover, many steps of the oscillation also need to cross Ω , e.g., change the window size from $w(t) < \omega_0$ to $w(t+1) > \omega_1$. Hence, $\omega_1 - \omega_0$ is proportional to the window-size changes within the update periods, and thus to the oscillation amplitude. Since the update function w^\leftarrow (in our case: \tilde{w}^\leftarrow) must be strictly decreasing in Ω , the maximally large neighborhood $\Omega = (\omega_0, \omega_1)$ corresponds to (w_0, w_1) , i.e., the decreasing part of \tilde{w}^\leftarrow (cf. Eq. (E.94)).

E.4.2 Existence of Limit Cycle

With the two assumptions from Appendix E.4.1, we now prove that the BBR-CUBIC oscillation has a limit cycle of CUBIC window sizes $\tilde{w}^\leftarrow(w_{\min})$ and $\tilde{w}^\leftarrow(w_{\max})$, i.e., the CUBIC window size persistently alternates between these two values.

Clearly, such a limit cycle only exists given an unstable long-term equilibrium \bar{w} . As noted in Appendix E.3.3, an equilibrium \bar{w} is only unstable if $\bar{w} \in (w_0, w_1)$ with $w_0 < w_1$, i.e., $\bar{w} = \tilde{w}^\leftarrow(\bar{w})$ is a value of \tilde{w}^\leftarrow in the decreasing part of \tilde{w}^\leftarrow . Since the values of \tilde{w}^\leftarrow are restricted to $[w_{\min}, w_{\max}]$ (cf. Eqs. (E.92) and (E.93)), it holds that $\bar{w} \in ([w_{\min}, w_{\max}] \cap (w_0, w_1)) =: W_\cap$. Since \bar{w} is guaranteed to exist, we know that $W_\cap \neq \emptyset$, and thus

$$w_0 < w_{\max} \quad \text{and} \quad w_1 > w_{\min}. \quad (\text{E.107})$$

In the following, we consider all possible cases for W_\cap , and verify whether the limit cycle is sound, i.e.,:

$$\begin{aligned} \tilde{w}^\leftarrow(w_{\min}) &= \tilde{w}^\leftarrow(\tilde{w}^\leftarrow(w_{\max})) \\ \tilde{w}^\leftarrow(w_{\max}) &= \tilde{w}^\leftarrow(\tilde{w}^\leftarrow(w_{\min})) \end{aligned} \quad (\text{E.108})$$

E.4.2.1 $W_\cap = (w_0, w_1)$

Given the intersection range, we know that $w_{\min} \leq w_0 < w_1 \leq w_{\max}$. This interleaving implies that

$$\tilde{w}^\leftarrow(w_{\min}) \stackrel{(\text{E.94})}{=} w_{\max} \quad \text{and} \quad \tilde{w}^\leftarrow(w_{\max}) \stackrel{(\text{E.94})}{=} w_{\min}. \quad (\text{E.109})$$

Clearly, the limit cycle is sound in this case:

$$\begin{aligned}\tilde{w}^{\leftarrow}(w_{\min}) &\stackrel{(E.109)}{=} \tilde{w}^{\leftarrow}(\tilde{w}^{\leftarrow}(w_{\max})) \\ \tilde{w}^{\leftarrow}(w_{\max}) &\stackrel{(E.109)}{=} \tilde{w}^{\leftarrow}(\tilde{w}^{\leftarrow}(w_{\min}))\end{aligned}\quad (E.110)$$

E.4.2.2 $W_{\cap} = [w_{\min}, w_1]$

Given the intersection range, we know that $w_0 < w_{\min} < w_1 \leq w_{\max}$. Hence, we know that

$$\tilde{w}^{\leftarrow}(w_{\min}) \stackrel{(E.94)}{<} w_{\max} \quad \text{and} \quad \tilde{w}^{\leftarrow}(w_{\max}) \stackrel{(E.94)}{=} w_{\min}. \quad (E.111)$$

With this knowledge, the first soundness condition in Eq. (E.108) holds:

$$\tilde{w}^{\leftarrow}(w_{\min}) \stackrel{(E.111)}{=} \tilde{w}^{\leftarrow}(\tilde{w}^{\leftarrow}(w_{\max})) \quad (E.112)$$

The second condition, however, is not directly satisfied, but can be converted to:

$$\tilde{w}^{\leftarrow}(w_{\max}) = \tilde{w}^{\leftarrow}(\tilde{w}^{\leftarrow}(w_{\min})) \stackrel{(E.111)}{\iff} \quad (E.113)$$

$$w_{\min} = \tilde{w}^{\leftarrow}(\tilde{w}^{\leftarrow}(w_{\min})) \stackrel{(E.94)}{\iff} \quad (E.114)$$

$$\tilde{w}^{\leftarrow}(w_{\min}) \geq w_1 \stackrel{(E.92)}{\iff} \tilde{w}^{\leftarrow}(\tilde{w}^{\leftarrow}(w_1)) \geq w_1 \quad (E.115)$$

$$\stackrel{-\tilde{w}^{\leftarrow}(w_1)}{\iff} \tilde{w}^{\leftarrow}(\tilde{w}^{\leftarrow}(w_1)) - \tilde{w}^{\leftarrow}(w_1) \geq w_1 - \tilde{w}^{\leftarrow}(w_1)$$

$$\stackrel{(E.94)}{=} w_1 - w_{\min} \stackrel{(E.107)}{>} 0 \quad (E.116)$$

$$\stackrel{\text{RHS}}{\iff} \frac{\tilde{w}^{\leftarrow}(\tilde{w}^{\leftarrow}(w_1)) - \tilde{w}^{\leftarrow}(w_1)}{\tilde{w}^{\leftarrow}(w_1) - w_1} \leq -1 \quad (E.117)$$

In turn, the last condition on w_1 is ensured by the derivative condition in neighborhood $\Omega = [w_0, w_1]$, i.e., by Eq. (E.104).

E.4.2.3 $W_{\cap} = (w_0, w_{\max}]$

This case is symmetric to the case for $W_{\cap} = [w_{\min}, w_1]$ since $w_{\min} \leq w_0 < w_{\max} < w_1$.

E.4.2.4 $W_{\cap} = [w_{\min}, w_{\max}]$

From the intersection range, we derive $w_{\min} > w_0$ and $w_{\max} < w_1$. These conditions can be shown to be incompatible with an unstable equilibrium (again via the derivative condition):

$$w_{\max} - w_{\min} < w_1 - w_0 \xrightleftharpoons[-1]{\text{RHS}} \frac{\tilde{w}^{\leftarrow}(w_0) - \tilde{w}^{\leftarrow}(w_1)}{w_0 - w_1} > -1. \quad (\text{E.118})$$

Hence, this case is outside the scope of this proof.

E.4.3 Convergence to Limit Cycle

In the following, we show that the long-term dynamics of the CUBIC window size w converge to the limit cycle constituted by $\tilde{w}^{\leftarrow}(w_{\min})$ and $\tilde{w}^{\leftarrow}(w_{\max})$. For this purpose, we revisit all relevant cases from Appendix E.4.2.

E.4.3.1 $W_{\cap} = (w_0, w_1)$

Without loss of generality, we assume an initial state $w(t) < \bar{w}$, which allows the following case distinction:

- $w(t) \leq w_0$: We know that $w(t+1) = \tilde{w}^{\leftarrow}(w(t)) \stackrel{(\text{E.94})}{=} w_{\max}$. Hence, the limit cycle is entered in the first step, as w_{\max} is part of the limit cycle by Eq. (E.109).
- $w(t) \in (w_0, \bar{w})$: We note that \tilde{w}^{\leftarrow} is strictly monotonically decreasing between $w(t) > w_0$ and $\bar{w} < w_1$ by Eq. (E.94), and thus

$$w(t+1) = \tilde{w}^{\leftarrow}(w(t)) > \tilde{w}^{\leftarrow}(\bar{w}) \stackrel{(\text{E.100})}{=} \bar{w}. \quad (\text{E.119})$$

- $w(t+1) \geq w_1$: The limit cycle is entered at $w(t+2) = \tilde{w}^{\leftarrow}(w(t+1)) = w_{\min}$, as w_{\min} is part of the limit cycle by Eq. (E.109).
- $w(t+1) \in (\bar{w}, w_1)$: We again note that \tilde{w}^{\leftarrow} is strictly monotonically decreasing between $\bar{w} > w_0$ and $w(t+1) < w_1$, and thus $w(t+2) = \tilde{w}^{\leftarrow}(w(t+1)) < \bar{w}$. Moreover, the window size makes progress towards the limit cycle if $w(t+2) < w(t)$, which can once more be guaranteed via the derivative condition.

E.4.3.2 $W_{\cap} = [w_{\min}, w_1]$

In this case, the limit cycle L is composed of w_{\min} and $\tilde{w}^{\leftarrow}(w_{\min}) < w_{\max}$ (Eq. (E.111)). Without loss of generality, we assume an initial state $w(t) < \bar{w}$, which allows the following case distinction:

- $w(t) \leq w_0$: In that case, $w(t+1) = \tilde{w}^{\leftarrow}(w(t)) = w_{\max}$ by Eq. (E.94). Since $w_{\max} \geq w_1$ in the current case, the limit cycle is entered at $w(t+2) = \tilde{w}^{\leftarrow}(w_{\max}) = w_{\min}$.
- $w(t) \in (w_0, w_{\min})$: We note that \tilde{w}^{\leftarrow} is strictly monotonically decreasing between $w(t) > w_0$ and $w_{\min} \leq w_1$ (Eq. (E.94), where $w(t) < w_{\min}$). This property implies that

$$w(t+1) = \tilde{w}^{\leftarrow}(w(t)) > \tilde{w}^{\leftarrow}(w_{\min}) \stackrel{(E.116)}{\geq} w_1. \quad (E.120)$$

Since $w(t+1) \geq w_1$, the limit cycle is then entered at $w(t+2) = \tilde{w}^{\leftarrow}(w(t+1)) = w_{\min}$.

- $w(t) \in [w_{\min}, \bar{w}]$: This case only arises if $w_{\min} < \bar{w}$. Hence, \tilde{w}^{\leftarrow} is again strictly monotonically decreasing between $w_{\min} > w_0$ and $\bar{w} < w_1$, and therefore $w(t+1) > \bar{w}$ from Eq. (E.119) holds again.
 - $w(t+1) \geq w_1$: The limit cycle is entered at $w(t+2) = \tilde{w}^{\leftarrow}(w(t+1)) = w_{\min}$ by Eq. (E.94).
 - $w(t+1) \in (\bar{w}, w_1)$: Based on arguments symmetric to the above analysis, plus the derivative condition from Eq. (E.104), we find that $w(t+2) \in [w_{\min}, w(t))$, which implies that the limit cycle is eventually entered.

E.4.3.3 $W_{\cap} = (w_0, w_{\max}]$

This case is symmetric to the case in §E.4.3.2 (directly above).

E.5 DERIVING UPDATE FUNCTION FROM EXPERIMENT SAMPLES

To fit w^{\leftarrow} to the samples $\{w_e(t)\}_{t \in \{1, \dots, 12\}}$ from experiment e , we require an ansatz. To that end, we observe from the experiment in Fig. 6.4 that the evolution speed of

the CUBIC window size w seems asymmetric: The window shrinks substantially faster than it grows. Hence, we use the following ansatz for w^\leftarrow :

$$w^\leftarrow(w) = \begin{cases} (1 - \rho^\uparrow)w + \rho^\uparrow \tilde{w}^\leftarrow(w) & \text{if } w \leq \tilde{w}^\leftarrow(w), \\ (1 - \rho^\downarrow)w + \rho^\downarrow \tilde{w}^\leftarrow(w) & \text{if } w > \tilde{w}^\leftarrow(w). \end{cases} \quad (\text{E.121})$$

where \tilde{w}^\leftarrow is the update function under complete convergence to the short-term equilibrium \tilde{w} . Since \tilde{w} also requires a value for the minimum BBR bottleneck-bandwidth estimate χ , we choose $\chi = 1\text{MSS}/\tau = 1500 \text{ byte}/\tau$. Moreover, $\rho^\uparrow, \rho^\downarrow \in (0, 1]$ are the upwards and downwards convergence speeds, respectively.

With this ansatz and the experiment data, we can estimate the upwards convergence speed ρ_γ^\uparrow for each configuration γ :

$$\rho_\gamma^\uparrow = \text{MEAN}_{t \in T^\uparrow(E(\gamma))} \left[\frac{w_e(t+1) - w_e(t)}{\tilde{w}^\leftarrow(w_e(t)) - w_e(t)} \right], \quad (\text{E.122})$$

where

$$T^\uparrow(E) = \{t \mid |w_e(t+1) - w_e(t)|\}. \quad (\text{E.123})$$

The downwards convergence speed is estimated analogously.

To obtain a general approximation of w^\leftarrow , we average the convergence speeds over all configurations:

$$\begin{aligned} \rho^\uparrow &= \text{MEAN}_\gamma [\rho_\gamma^\uparrow] \approx 23.7\% \\ \rho^\downarrow &= \text{MEAN}_\gamma [\rho_\gamma^\downarrow] \approx 47.8\% \end{aligned} \quad (\text{E.124})$$



APPENDIX OF CHAPTER 8

F.1 PROOF OF THEOREM 8.1: BEST-RESPONSE ATTRIBUTE

F.1.1 Unrestricted Best Response in Eq. (8.10)

For an individual source-destination pair (n_1, n_2) , we can simplify Eq. (8.5) as follows:

$$\pi_n = \frac{dv_{r(n)}}{1 + \sum_{r' \in R} v_{r'}} \cdot \left(\rho_n - \sum_{k \in K} \phi_{nk} a_{nk} - \phi_{n0} \right) - \sum_{k \in K} \gamma_{nk} a_{nk} - \gamma_{n0} \quad (\text{F.1})$$

where the argument \mathbf{A} has been omitted, and $v_{r(n)}$ denotes the sum of valuations for all paths including node n :

$$v_{r(n)} = \sum_{r \in R, n \in r} v_r. \quad (\text{F.2})$$

Differentiating Eq. (F.1) with respect to attribute a_{nk} of ISP n yields:

$$\frac{d\pi_n}{da_{nk}} = \frac{d\alpha_{nk} (1 + v_{-r(n)})}{(1 + \sum_{r' \in R} v_{r'})^2} \cdot \left(\rho_n - \sum_{k \in K} \phi_{nk} a_{nk} - \phi_{n0} \right) - \frac{d\phi_{nk} v_{r(n)}}{1 + \sum_{r' \in R} v_{r'}} - \gamma_{nk} \quad (\text{F.3})$$

where the abbreviations from Eq. (8.14) have been used. Setting Eq. (F.3) to 0 allows the following rewriting:

$$\begin{aligned} & d\alpha_{nk} (1 + v_{-r(n)}) \cdot \left(\rho_n - \sum_{k' \in K} \phi_{nk'} a_{nk'} - \phi_{n0} \right) - d\phi_{nk} v_{r(n)} \cdot \left(1 + \sum_{r' \in R} v_{r'} \right) \\ & - \gamma_{nk} \cdot \left(1 + \sum_{r' \in R} v_{r'} \right)^2 = 0 \end{aligned} \quad (\text{F.4})$$

In the LHS, we can substitute

$$\begin{aligned} v_{r(n)} + v_{-r(n)} &= \sum_{r' \in R} v_{r'} = \alpha_{nk} a_{nk} + v_{-nk} = \alpha_{nk} a_{nk} + v_{r(n), -nk} + v_{-r(n)} \\ \sum_{k' \in K} \phi_{nk'} a_{nk'} + \phi_{n0} &= \phi_{nk} a_{nk} + \Phi_{-nk} \end{aligned} \quad (\text{F.5})$$

and obtain:

$$-d\alpha_{nk}\phi_{nk} \left(1 + v_{-r(n)}\right) \cdot a_{nk} + d\alpha_{nk} (\rho_n - \Phi_{-nk}) \left(1 + v_{-r(n)}\right) \quad (\text{F.6})$$

$$-d\phi_{nk} \left(\alpha_{nk} a_{nk} + v_{r(n), -nk}\right) (\alpha_{nk} a_{nk} + 1 + v_{-nk})$$

$$-\gamma_{nk} \cdot (\alpha_{nk} a_{nk} + 1 + v_{-nk})^2$$

$$\iff -d\alpha_{nk}\phi_{nk} \left(1 + v_{-r(n)}\right) \cdot a_{nk} + d\alpha_{nk} (\rho_n - \Phi_{-nk}) \left(1 + v_{-r(n)}\right) \quad (\text{F.7})$$

$$-d\phi_{nk} \cdot \alpha_{nk}^2 a_{nk}^2$$

$$-d\phi_{nk} \cdot \alpha_{nk} \cdot \left(v_{r(n), -nk} + (1 + v_{-nk})\right) a_{nk}$$

$$-d\phi_{nk} \cdot v_{r(n), -nk} \cdot (1 + v_{-nk})$$

$$-\gamma_{nk} \cdot \left(\alpha_{nk}^2 a_{nk}^2 + 2\alpha_{nk} (1 + v_{-nk}) a_{nk} + (1 + v_{-nk})^2\right)$$

$$\iff \left(-d\phi_{nk} \alpha_{nk}^2 - \gamma_{nk} \alpha_{nk}^2\right) \cdot a_{nk}^2 \quad (\text{F.8})$$

$$+ \left(-d\alpha_{nk}\phi_{nk} \left(1 + v_{-r(n)}\right) - d\alpha_{nk}\phi_{nk} \left(1 + v_{r(n), -nk} + v_{-nk}\right)\right) \cdot a_{nk}$$

$$+ (-2\alpha_{nk}\gamma_{nk} (1 + v_{-nk})) \cdot a_{nk}$$

$$+ d\alpha_{nk} (\rho_n - \Phi_{-nk}) \left(1 + v_{-r(n)}\right)$$

$$-d\phi_{nk} \cdot v_{r(n), -nk} \cdot (1 + v_{-nk}) - \gamma_{nk} \cdot (1 + v_{-nk})^2$$

$$\iff -\alpha_{nk}^2 (d\phi_{nk} + \gamma_{nk}) \cdot a_{nk}^2 \quad (\text{F.9})$$

$$-2d\alpha_{nk}\phi_{nk} (1 + v_{-nk}) \cdot a_{nk} - 2\alpha_{nk}\gamma_{nk} (1 + v_{-nk}) \cdot a_{nk}$$

$$+ d\alpha_{nk} (\rho_n - \Phi_{-nk}) - d\phi_{nk} \cdot v_{r(n), -nk} \cdot (1 + v_{-nk}) - \gamma_{nk} \cdot (1 + v_{-nk})^2$$

$$\iff -\alpha_{nk}^2 (d\phi_{nk} + \gamma_{nk}) \cdot a_{nk}^2 \quad (\text{F.10})$$

$$-2\alpha_{nk} (d\phi_{nk} + \gamma_{nk}) (1 + v_{-nk}) \cdot a_{nk}$$

$$+ d\alpha_{nk} (\rho_n - \Phi_{-nk}) \left(1 + v_{-r(n)}\right)$$

$$-d\phi_{nk} \cdot v_{r(n), -nk} \cdot (1 + v_{-nk}) - \gamma_{nk} \cdot (1 + v_{-nk})^2$$

$$\iff T_1 a_{nk}^2 + T_2 a_{nk} + T_3 = 0 \quad (\text{F.11})$$

Hence, we obtain a quadratic equation in a_{nk} . This quadratic equation has the solutions:

$$a_{nk} = \frac{-T_2 \pm \sqrt{T_2^2 + 4T_1 T_3}}{2T_1} \quad (\text{F.12})$$

$$= -\frac{2\alpha_{nk}(d\phi_{nk} + \gamma_{nk})(1 + v_{-nk})}{2\alpha_{nk}^2(d\phi_{nk} + \gamma_{nk})} \quad (\text{F.13})$$

$$\pm \frac{\sqrt{4\alpha_{nk}^2(d\phi_{nk} + \gamma_{nk})^2(1 + v_{-nk})^2 + 4\alpha_{nk}^2(d\phi_{nk} + \gamma_{nk}) \cdot T_3}}{2\alpha_{nk}^2(d\phi_{nk} + \gamma_{nk})} \quad (\text{F.14})$$

$$= -\frac{1 + v_{-nk}}{\alpha_{nk}} \pm \frac{\sqrt{(d\phi_{nk} + \gamma_{nk})(1 + v_{-nk})^2 + T_3}}{\alpha_{nk}\sqrt{d\phi_{nk} + \gamma_{nk}}}, \quad (\text{F.15})$$

where we expand the term under the root as follows:

$$(d\phi_{nk} + \gamma_{nk})(1 + v_{-nk})^2 + T_3 \quad (\text{F.16})$$

$$= d\phi_{nk} \cdot (1 + v_{-nk})^2 + \underline{\gamma_{nk}(1 + v_{-nk})^2} \quad (\text{F.17})$$

$$\begin{aligned} &+ d\alpha_{nk}(\rho_n - \Phi_{-nk}) \left(1 + v_{-r(n)}\right) \\ &- d\phi_{nk} \cdot v_{r(n),-nk} \cdot (1 + v_{-nk}) \underline{-\gamma_{nk} \cdot (1 + v_{-nk})^2} \\ &= d\phi_{nk} \cdot (1 + v_{-nk})^2 - d\phi_{nk} \cdot v_{r(n),-nk} \cdot (1 + v_{-nk}) \end{aligned} \quad (\text{F.18})$$

$$+ d\alpha_{nk}(\rho_n - \Phi_{-nk}) \left(1 + v_{-r(n)}\right) \quad (\text{F.19})$$

$$\stackrel{(\text{F.20})}{=} d\phi_{nk} \left(1 + v_{-r(n)}\right) (1 + v_{-nk}) + d\alpha_{nk}(\rho_n - \Phi_{-nk}) \left(1 + v_{-r(n)}\right),$$

where we have made use of the following equality in the last step:

$$\begin{aligned} &(1 + v_{-nk})^2 - v_{r(n),-nk} (1 + v_{-nk}) \\ &= \left(1 + v_{r(n),-nk} + v_{-r(n)}\right) (1 + v_{-nk}) - v_{r(n),-nk} (1 + v_{-nk}) \\ &= \left(1 + v_{-r(n)}\right) (1 + v_{-nk}) \end{aligned} \quad (\text{F.20})$$

By reinserting Eq. (F.19) in Eq. (F.15), we obtain:

$$a_{nk} = \frac{1}{\alpha_{nk}} \left(\pm \sqrt{\frac{d(1 + v_{-r(n)})}{d\phi_{nk} + \gamma_{nk}} (\phi_{nk}(1 + v_{-nk}) + \alpha_{nk}(\rho_n - \Phi_{-nk}))} - (1 + v_{-nk}) \right)$$

(F.21)

where only the upper solution (i.e., with the positive coefficient of the square-root term) is potentially valid given the non-negativity of attributes. Hence, we arrive at $\hat{a}_{nk}^*(\mathbf{A}_{-nk})$ as in Eq. (8.10).

F.1.2 Confirmation of Maximum

This solution is a maximum of π_n if π_n from Eq. (F.1) is concave in a_{nk} , which can be demonstrated by means of the second derivative:

$$\frac{\partial^2 \pi_n}{\partial a_{nk}^2} = \frac{-2d\alpha_{kn}^2 (1 + v_{-r(n)})}{(1 + \sum_{r' \in R} v_{r'})^3} \cdot \left(\rho_n - \sum_{k \in K} \phi_{nk} a_{nk} - \phi_{n0} \right) - \frac{2d\alpha_n \phi_{nk} (1 + v_{-r(n)})}{(1 + \sum_{r' \in R} v_{r'})^2} \quad (\text{F.22})$$

Clearly, π_n may only be non-concave under the following condition:

$$\frac{\partial^2 \pi_n}{\partial a_{nk}^2} > 0 \implies \rho_n - \sum_{k \in K} \phi_{nk} a_{nk} - \phi_{n0} < 0 \quad (\text{F.23})$$

However, if the condition in Eq. (F.23) is true for some $\mathbf{a}_n \in \mathbb{R}_{\geq 0}^K$, then the profit function has a negative slope at that point (cf. Eq. (F.3)). Hence, the profit function has no extrema in the non-concave regions, and thus any extremum, in particular $\hat{a}_{nk}^*(\mathbf{A}_{-nk})$, is guaranteed to be located in the concave regions and to be a maximum.

F.1.3 Restricted Best Response in Eq. (8.9)

Since attribute values must be non-negative, we now investigate how the unrestricted best-response \hat{a}_{nk}^* informs the best-response a_{nk}^* on the restricted domain $\mathbb{R}_{\geq 0}$. Clearly, if $\hat{a}_{nk}^* \geq 0$, then $a_{nk}^* = \hat{a}_{nk}^*$. Otherwise, if $\hat{a}_{nk}^* < 0$, the boundary point $a_{nk} = 0$ constitutes a local maximum on the restricted domain $\mathbb{R}_{\geq 0}$. To confirm this statement, we have to distinguish the cases $\phi_{nk} > 0$ and $\phi_{nk} = 0$.

If $\phi_{nk} > 0$, we note that the profit function can only be non-concave for high enough a_{nk} :

$$\rho_n - \sum_{k \in K} \phi_{nk} a_{nk} - \phi_{n0} < 0 \iff a_{nk} > \frac{\rho_n - \Phi_{-nk}}{\phi_{nk}} =: \bar{a}_{nk} \quad (\text{F.24})$$

Hence, π_n is guaranteed to be concave for all $a_{nk} \leq \bar{a}_{nk}$. Furthermore, as argued in Appendix F.1.2, π_n is strictly decreasing for all $a_{nk} > \bar{a}_{nk}$. Hence, $a_{nk} = 0$ is the maximum on the restricted domain independent of \bar{a}_{nk} , given that $\hat{a}_{nk}^* < 0$: Either the boundary point $a_{nk} = 0$ is in the concave region (if $\bar{a}_{nk} \geq 0$) or in the decreasing region (if $\bar{a}_{nk} < 0$).

If $\phi_{nk} = 0$, π_n can only be non-concave if $\rho_n - \Phi_{-nk} < 0$, independent of a_{nk} . Hence, π_n is either guaranteed to be consistently concave in a_{nk} (if $\rho_n - \Phi_{-nk} \geq 0$), or guaranteed to be decreasing for all $a_{nk} \geq 0$. In both cases, the boundary point $a_{nk} = 0$ constitutes a local maximum.

Finally, we investigate the case where the unrestricted best response \hat{a}_{nk}^* is *undefined* on \mathbb{R} , i.e., if the term under the square root is negative. A necessary condition for this negativity is that $\rho_n - \Phi_{-nk}$ is negative, which according to Eq. (F.24) implies that $\bar{a}_{nk} < 0$. Hence, in this case π_n is decreasing for $a_{nk} \geq 0$, which again makes $a_{nk} = 0$ a local maximum on the restricted domain $\mathbb{R}_{\geq 0}$.

F.2 PROOF OF THEOREM 8.2: HOMOGENEOUS NASH EQUILIBRIUM

F.2.1 Homogeneous Profit Function

In a homogeneous network, all attributes are equally valuable and costly, i.e., α_{nk} , ϕ_{nk} and γ_{nk} are equal across all attributes $k \in K$. This homogeneity allows to understand the profit function π_n of a ISP n as a function of the attribute sums $a_n = \sum_{k \in K} a_{nk}$:

$$\pi_n(\mathbf{A}) = \frac{\alpha_1 a_n + \alpha_1 \sum_{n' \in r(n) \setminus \{n\}} a_{n'} + \alpha_0}{1 + \alpha_1 a_n + \alpha_1 \sum_{n' \in N \setminus \{n\}} a_{n'}} d(\rho - \phi_1 a_n - \phi_0) - \gamma_1 a_n - \gamma_0 \quad (\text{F.25})$$

In the following, we thus treat the attribute sum a_n like a single attribute of ISP n .

F.2.2 Unrestricted Equilibrium in Eq. (8.18)

The equilibrium conditions in Eq. (8.17) suggest that the equilibrium for a homogeneous parallel-path network satisfies the following equation for every $n \in N$:

$$a_n^+ = \max(0, \hat{a}_n^*(\mathbf{A}_{-n}^+)) \quad (\text{F.26})$$

where

$$\hat{a}_n^*(\mathbf{A}_{-n}^+) = \frac{1}{\alpha_1} \left(\sqrt{\frac{d(1 + v_{-r(n)}(\mathbf{A}_{-n}^+))}{d\phi_1 + \gamma_1}} (\phi_1(1 + v_{-n}(\mathbf{A}_{-n}^+)) + \alpha_1(\rho - \phi_0)) \right) - (1 + v_{-n}(\mathbf{A}_{-n})) \quad (\text{F.27})$$

$$v_{-r(n)}(\mathbf{A}_{-n}^+) = \alpha_1 \sum_{n' \in N \setminus r(n)} a_{n'}^+ + (Q - 1)\alpha_0 \quad (\text{F.28})$$

$$v_{-n}(\mathbf{A}_{-n}^+) = v_{-r(n)}(\mathbf{A}_{-n}^+) + \alpha_1 \sum_{n' \in r(n) \setminus \{n\}} a_{n'}^+ + \alpha_0 \quad (\text{F.29})$$

Note that we effectively consider a single attribute in the style of the attribute sum, which simplifies $\Phi_{-nk} = \phi_0 \leq \rho$. Hence, the undefined case from Theorem 8.1 does not arise because the term under the square root in Eq. (F.26) is always non-negative.

To solve the equation system Eq. (F.26) $\forall n \in N$, we first consider the unrestricted system (i.e., without required non-negativity of solutions) in $\hat{\mathbf{A}}^+$:

$$\forall n \in N. \quad \hat{a}_n^+ = \frac{\sqrt{\frac{d(1 + v_{-r(n)}(\hat{\mathbf{A}}_{-n}^+))}{d\phi_1 + \gamma_1}} (\phi_1(1 + v_{-n}(\hat{\mathbf{A}}_{-n}^+)) + \alpha_1(\rho - \phi_0))}{\alpha_1} - \frac{(1 + v_{-n}(\hat{\mathbf{A}}_{-n}^+))}{\alpha_1} \quad (\text{F.30})$$

This equation system can be transformed such that the LHS is constant for all n :

$$\forall n \in N. \quad 1 + \alpha_1 \sum_{r \in R} \sum_{n' \in r} \hat{a}_{n'}^+ + Q\alpha_0 = \sqrt{\frac{d(1 + v_{-r(n)}(\hat{\mathbf{A}}_{-n}^+))}{d\phi_1 + \gamma_1}} (\phi_1(1 + v_{-n}(\hat{\mathbf{A}}_{-n}^+)) + \alpha_1(\rho - \phi_0)) \quad (\text{F.31})$$

which implies that all $a_n^+ \forall n \in N$ are equal to a value \hat{a}^+ . This value \hat{a}^+ can be found by solving the following single equation:

$$1 + QI\alpha_1\hat{a}^+ + Q\alpha_0 = \sqrt{\frac{d(1 + (Q - 1)(I\alpha_1\hat{a}^+ + \alpha_0))}{d\phi_1 + \gamma_1}} (\phi_1(1 + (QI - 1)\alpha_1\hat{a}^+ + Q\alpha_0) + \alpha_1(\rho - \phi_0)) \quad (\text{F.32})$$

This equation is solved by \hat{a}^+ as defined in Theorem 8.2.

F.2.3 Restricted Equilibrium

It remains to show how the solution \hat{a}^+ of the unrestricted system can be used to derive the actual solution a^+ of the restricted system. If $\hat{a}^+ \geq 0$, the unrestricted-system solution \hat{a}^+ is clearly also a solution to the restricted system, i.e., $a^+ = \hat{a}^+$. However, if $\hat{a}^+ < 0$, the attribute values as suggested by the unrestricted system are negative, which is invalid for the restricted system. In this case, we can show that a solution of the restricted system is given by $a^+ = 0$, i.e., $0 = a_n^*(\mathbf{0})$ (where a_n^* is the optimal choice in single-market networks according to Theorem 8.1). To show this property, we first observe that the condition for $\hat{a}^+ < 0$ can be simplified to a condition on sub-term T_3 :

$$\hat{a}^+ < 0 \implies \frac{\sqrt{T_2^2 - 4T_1 T_3} - T_2}{2T_1} < 0 \quad (\text{F.33})$$

$$\text{For } T_1 > 0: \quad T_2^2 - 4T_1 T_3 < T_2^2 \implies -4T_1 T_3 < 0 \implies T_3 > 0 \quad (\text{F.34})$$

$$\text{For } T_1 < 0: \quad T_2^2 - 4T_1 T_3 > T_2^2 \implies -4T_1 T_3 > 0 \implies T_3 > 0 \quad (\text{F.35})$$

$$\begin{aligned} \implies T_3 &= (1 + Q\alpha_0)^2 - \frac{d(1 + (Q-1)\alpha_0)}{d\phi_1 + \gamma_1} (\phi_1(1 + Q\alpha_0) + \alpha_1(\rho - \phi_0)) \\ &> 0 \end{aligned} \quad (\text{F.36})$$

Furthermore, the inequality on T_3 allows the following conclusion:

$$(1 + Q\alpha_0)^2 - \frac{d(1 + (Q-1)\alpha_0)}{d\phi_1 + \gamma_1} (\phi_1(1 + Q\alpha_0) + \alpha_1(\rho - \phi_0)) > 0 \iff \quad (\text{F.37})$$

$$(1 + Q\alpha_0) - \sqrt{\frac{d(1 + (Q-1)\alpha_0)}{d\phi_1 + \gamma_1} (\phi_1(1 + Q\alpha_0) + \alpha_1(\rho - \phi_0))} > 0 \quad (\text{F.38})$$

due to the equivalence $x^2 - y > 0 \iff x^2 > y \iff x > \sqrt{y} \iff x - \sqrt{y} > 0$ (if $x, y \geq 0$).

Eq. (F.38) has a striking similarity to $\hat{a}_n^*(\mathbf{0})$ for a homogeneous parallel-path network:

$$\hat{a}_n^*(\mathbf{0}) = \frac{\sqrt{\frac{d(1 + (Q-1)\alpha_0)}{d\phi_1 + \gamma_1} (\phi_1(1 + Q\alpha_0) + \alpha_1(\rho - \phi_0))} - (1 + Q\alpha_0)}{\alpha} \quad (\text{F.39})$$

More precisely, Eq. (F.38) implies that the unrestricted best response $\hat{a}_n^*(\mathbf{0})$ is always below 0 if the raw-equilibrium choice $\hat{a}^+ < 0$, and that the restricted best response $a_n^*(\mathbf{0})$ thus always equals 0 for $\hat{a}^+ < 0$, yielding a restricted equilibrium choice of $a^+ = 0$. This insight concludes the proof.

F.3 PROOF OF THEOREM 8.3: STABILITY OF HOMOGENEOUS EQUILIBRIUM

F.3.1 Linearization of Dynamic System

In order to prove asymptotic stability of the given Nash equilibrium, we leverage the indirect Lyapunov method [209]. This method requires that the equilibrium of an ODE system is asymptotically stable if the Jacobian matrix of the ODE system, evaluated at the equilibrium point, has exclusively negative eigenvalues. More formally, given the Jacobian matrix $\mathbf{J}(\mathbf{A}^+) \in \mathbb{R}^{|N| \times |N|}$, it must hold that $\forall \lambda \in \mathbb{R}$ that $\exists \mathbf{x} \in \mathbb{R}^{|N|}, \mathbf{x} \neq \mathbf{0}. \mathbf{J}(\mathbf{A}^+) \mathbf{x} = \lambda \mathbf{x} \implies \lambda < 0$. This matrix $\mathbf{J}(\mathbf{A}^+)$ is defined as follows for the dynamic system from Eq. (8.23):

$$\begin{aligned} J_{nn} &= \frac{\partial \dot{a}_n}{\partial a_n}(\mathbf{A}^+) = -1 \\ n \neq m, r(n) = r(m) : J_{nm} &= \frac{\partial \dot{a}_n}{\partial a_m}(\mathbf{A}^+) = \begin{cases} \frac{T_4}{T_5} - 1 & \text{if } \hat{a}^+ \geq 0 \\ 0 & \text{otherwise} \end{cases} \\ n \neq m, r(n) \neq r(m) : J_{nm} &= \frac{\partial \dot{a}_n}{\partial a_m}(\mathbf{A}^+) = \begin{cases} \frac{T_4 + T_6}{T_5} - 1 & \text{if } \hat{a}^+ \geq 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

where \hat{a}^+ is the unrestricted equilibrium attribute value according to Theorem 8.2, and

$$T_4 = d\phi_1(1 + v_{-r(n)}(\hat{\mathbf{A}}^+)), \quad (\text{F.43})$$

$$T_5 = 2(d\phi_1 + \gamma_1) \sqrt{\frac{d(1 + v_{-r(n)}(\hat{\mathbf{A}}_{-n}^+))}{d\phi_1 + \gamma_1}}. \quad (\text{F.44})$$

$$\sqrt{\phi_1(1 + v_{-n}(\hat{\mathbf{A}}_{-n}^+)) + \alpha_1(\rho - \phi_0)}, \text{ and} \quad (\text{F.45})$$

$$T_6 = d(\phi_1(1 + v_{-n}(\hat{\mathbf{A}}_{-n}^+)) + \alpha_1(\rho - \phi_0)). \quad (\text{F.46})$$

F.3.2 Case 1: Non-Negative Unrestricted Equilibrium ($\hat{a}^+ \geq 0$)

We first consider the case of a non-negative unrestricted equilibrium value \hat{a}^+ such that $\mathbf{A}^+ = \hat{\mathbf{A}}^+$. In that case, the eigenvalue condition induces the following system of equations:

$$\begin{aligned} \forall n \in N. \quad & (-\lambda - 1)x_n + \left(\frac{T_4}{T_5} - 1 \right) \sum_{n' \in r(n) \setminus \{n\}} x_{n'} + \\ & \left(\frac{T_4 + T_6}{T_5} - 1 \right) \sum_{n' \in N \setminus r(n)} x_{n'} = 0 \end{aligned} \quad (\text{F.47})$$

This system has a number of solutions (λ, \mathbf{x}) .

$\lambda = -T_4/T_5$. For $\lambda_1 = -T_4/T_5$, the first two terms in Eq. (F.47) obtain the same coefficient, and the equation system is reduced from $|N|$ ISP-specific to $|R|$ path-specific equations:

$$\begin{aligned} \forall r \in R. \quad & \left(\frac{T_4}{T_5} - 1 \right) X_r + \left(\frac{T_4 + T_6}{T_5} - 1 \right) \sum_{r' \in R \setminus \{r\}} X_{r'} = 0 \\ \text{where } X_r = \sum_{n' \in r} x_{n'} \end{aligned} \quad (\text{F.48})$$

Equation systems of this form may have three types of solutions in \mathbf{x} . For $T_6 = 0$ and $T_4 = T_5$, any \mathbf{x} is a solution, as the coefficients of the variables X_r are 0. For $T_6 = 0$ and $T_4 \neq T_5$, any \mathbf{x} with entries summing up to 0 is a solution, as the sum of all X_r has a single non-zero coefficient. For $T_6 \neq 0$, any \mathbf{x} with $X_r = 0 \forall r \in R$ is a solution. More importantly for the proof, $\lambda_1 = -T_4/T_5$ is consistently negative given the parameter ranges, except for the case where $\phi_1 = 0$ and hence $\lambda_1 = 0$. The case of $\phi_1 = 0$ is indeed an interesting special case for which the equilibrium is not unique, as we will show in §8.3.3; for this special case, $\lambda_1 = 0$ describes the fact that the dynamics do not converge to a specific equilibrium if they have already converged onto another equilibrium.

$\lambda \neq -T_4/T_5$. After discovering the first eigenvalue $\lambda_1 = -T_4/T_5$, we now consider the case where $\lambda \neq -T_4/T_5$. In this case, the symmetric structure of the equation system implies that all eigenvector entries x_n associated with the same

path r are equal, and thus $x_n = X_r / I$. Hence, a reduction of the equation system to $|R|$ equations is possible again:

$$\forall r \in R. \quad \left(\frac{-\lambda - 1}{I} + \frac{I - 1}{I} \left(\frac{T_4}{T_5} - 1 \right) \right) X_r + \left(\frac{T_4 + T_6}{T_5} - 1 \right) \sum_{r' \in R \setminus \{r\}} X_{r'} = 0 \quad (\text{F.49})$$

Again, this equation system admits different types of solutions.

The first type is obtainable by assuming $X_r = -\sum_{r' \in R \setminus \{r\}} X_{r'}$, and is associated with the following eigenvalue:

$$\frac{-\lambda_2 - 1}{I} + \frac{I - 1}{I} \left(\frac{T_4}{T_5} - 1 \right) = \frac{T_4 + T_6}{T_5} - 1 \implies \lambda_2 = -\frac{T_4 + IT_6}{T_5} \quad (\text{F.50})$$

Importantly in this case, λ_2 is consistently negative.

The second type of solution for the equation system in Eq. (F.49) is obtainable by assuming equal X_r across all paths $r \in R$, and is associated with the following eigenvalue:

$$\begin{aligned} \frac{-\lambda_3 - 1}{I} + \frac{I - 1}{I} \left(\frac{T_4}{T_5} - 1 \right) &= -(Q - 1) \left(\frac{T_4 + T_6}{T_5} - 1 \right) \\ \implies \lambda_3 &= QI \left(\frac{T_4 + T_6}{T_5} - 1 \right) - \frac{T_4 + IT_6}{T_5} \end{aligned} \quad (\text{F.51})$$

By inspection of λ_3 , we confirm that the maximum λ_3 is negative:

$$\max_{\substack{\alpha_1, \alpha_0, \phi_1, \phi_0, \\ \gamma_1, \rho, d, Q, I}} \lambda_3 = \max_{\phi_1, Q} \lim_{\substack{\alpha_1, \alpha_0, \rho, d \\ \rightarrow \infty}} \lim_{\phi_0, \gamma_1 \rightarrow 0} \lim_{I \rightarrow 1} \lambda_3 < -\frac{1}{2} \quad (\text{F.52})$$

Hence, all eigenvalues of $\mathbf{J}(\mathbf{A}^+)$ for $\hat{\alpha}^+ > 0$ are negative, i.e., the equilibrium is asymptotically stable in this case.

F.3.3 Case 2: Negative Unrestricted Equilibrium ($\hat{\alpha}^+ < 0$)

It remains to show that the equilibrium \mathbf{A}^+ is also asymptotically stable for the case $\hat{\alpha}^+ < 0$ such that $\mathbf{A}^+ = \mathbf{0}$. This part of the proof is trivial: For $\hat{\alpha}^+ < 0$, the Jacobian $\mathbf{J}(\mathbf{A}^+)$ corresponds to the negative identity matrix, which only has the negative eigenvalue $\lambda = -1$. Hence, the proof is concluded.

F.4 PROOF OF THEOREM 8.4: SUBOPTIMAL HOMOGENEOUS EQUILIBRIUM

F.4.1 NBS Attribute

To characterize the NBS attribute a° , we first require an understanding of the aggregate profit of ISPs on the path:

$$\sum_{n \in r} \pi_n(\mathbf{A}) = d \frac{\alpha_1 \sum_{n \in r} a_n + \alpha_0}{1 + \alpha_1 \sum_{n \in r} a_n + \alpha_0} (I(\rho - \phi_0) - \phi_1 \sum_{n \in r} a_n) - \gamma_1 \sum_{n \in r} a_n - \gamma_0 \quad (\text{F.53})$$

The aggregate profit can thus be considered a function of the sum a_r of ISP attributes on path, i.e., $a_r := \sum_{n \in r} a_n$. By Theorem 8.1, the unrestricted optimal attribute sum a_r° is $a_r^\circ = \max(0, \hat{a}_r^\circ)$, where:

$$\hat{a}_r^\circ = \frac{\sqrt{\frac{d}{d\phi_1 + \gamma_1} (\phi_1(1 + \alpha_0) + I\alpha_1(\rho - \phi_0))} - (1 + \alpha_0)}{\alpha_1}. \quad (\text{F.54})$$

Clearly, the Nash bargaining attributes $\{a_n^\circ\}_{n \in r}$ must sum to a_r° in order to be optimal. Moreover, the Nash bargaining solution is fair for the cooperating entities, requiring equal profit for all ISPs in our context. As a result, the Nash bargaining solution stipulates a single attribute value a° , which is adopted by all ISPs. This NBS attribute a° is $a^\circ = \max(0, \hat{a}^\circ)$, where $\hat{a}^\circ = \hat{a}_r^\circ / I$.

F.4.2 Equilibrium Attribute

The equilibrium attribute a^+ is defined as in Theorem 8.2, but can be considerably simplified for the case $Q = 1$. In particular, the equilibrium attribute a^+ for $Q = 1$ is $a^+ = \max(0, \hat{a}^+)$, where:

$$\hat{a}^+ = \frac{\sqrt{T_2^2 - 4T_1T_3 - T_2}}{2T_1}, \quad (\text{F.55})$$

$$T_1 = I^2\alpha_1^2, \quad (\text{F.56})$$

$$T_2 = 2I\alpha_1(1 + \alpha_0) - \frac{d}{d\phi_1 + \gamma_1} \alpha_1 \phi_1 (I - 1), \text{ and}$$

$$T_3 = (1 + \alpha_0)^2 - \frac{d}{d\phi_1 + \gamma_1} (\phi_1(1 + \alpha_0) + \alpha_1(\rho - \phi_0)). \quad (\text{F.57})$$

F.4.3 Comparison of Attributes

We show that $a^+ \leq a^\circ$ by showing that $\hat{a}^+ \leq \hat{a}^\circ$. This inequality can be rewritten as

$$\frac{\sqrt{T_2^2 - 4T_1T_3} - T_2}{2T_1} \leq \hat{a}^\circ \iff T_2\hat{a}^\circ \geq -(T_1\hat{a}^{\circ 2} + T_3) \quad (\text{F.58})$$

The two sides of the second inequality in Eq. (F.58) expand to:

$$T_2\hat{a}^\circ = -2(1 + \alpha_0)^2 + 2(1 + \alpha_0) \cdot \sqrt{\frac{d}{d\phi_1 + g_1}(\phi_1(1 + \alpha_0) + I\alpha_1(r - \phi_0))} \\ - \frac{d}{d\phi_1 + g_1}(I - 1)\alpha_1\phi_1\hat{a}^\circ \quad (\text{F.59})$$

$$-(T_1\hat{a}^{\circ 2} + T_3) = -2(1 + \alpha_0)^2 + 2(1 + \alpha_0) \cdot \sqrt{\frac{d}{d\phi_1 + g_1}(\phi_1(1 + \alpha_0) + I\alpha_1(r - \phi_0))} \\ - \frac{d}{d\phi_1 + g_1}(I - 1)\alpha_1(r - \phi_0) \quad (\text{F.60})$$

Since these terms lend themselves to considerable simplification, Eq. (F.58) becomes:

$$\rho - \phi_1\hat{a}^\circ - \phi_0 \geq 0 \quad (\text{F.61})$$

Interestingly, \hat{a}° is guaranteed to satisfy this inequality. To see why, assume the opposite for the sake of contradiction: $\rho - \phi_1\hat{a}^\circ - \phi_0 < 0$. If $\phi_1 = 0$, this inequality conflicts with the model assumption $\rho - \phi_0 \geq 0$. If $\phi_1 > 0$, the same model assumption indicates that $\hat{a}^\circ > (\rho - \phi_0)/\phi_1 \geq 0$. Hence, the profit function of any ISP n is negative:

$$\pi_n(\hat{\mathbf{A}}^\circ) = d \underbrace{\frac{\alpha_1 I \hat{a}^\circ + \alpha_0}{1 + \alpha_1 I \hat{a}^\circ + \alpha_0}}_{>0} \underbrace{(\rho - \phi_1\hat{a}^\circ - \phi_0)}_{<0} \underbrace{- \gamma_1\hat{a}^\circ - \gamma_0}_{\leq 0} \quad (\text{F.62})$$

However, this negative profit could be strictly improved by choosing the lower attribute value $a' = (\rho - \phi_0)/\phi_1 < \hat{a}^\circ$. This profit improvement is a contradiction to the character of \hat{a}° as the profit-optimizing attribute value. Hence, Eq. (F.61) holds, and therefore also the proposition $\hat{a}^+ \leq \hat{a}^\circ$ holds. This insight concludes the proof.

F.5 PROOF OF THEOREM 8.5: HOMOGENEOUS COMPETITION (ATTRIBUTES)

F.5.1 Equilibrium for Competitive Network \mathcal{N}_2

We begin the proof by characterizing the equilibrium for the competitive network \mathcal{N}_2 , in which every ISP n optimizes the following profit function π_n :

$$\pi_n(a_n) = d' \left(\sum_{q=1}^Q \frac{v_{r(m_{q1}, m_{q2}, n)}}{1 + \sum_{r' \in R(m_{q1}, m_{q2})} v_{r'}} \right) (\rho - \phi_1 a_n - \phi_0) - \gamma_1 a_n - \gamma_0 \quad (\text{F.63})$$

where $r(m_{q1}, m_{q2}, n)$ denotes the unique path connecting (m_{q1}, m_{q2}) via ISP n . In the unrestricted equilibrium $\hat{\mathbf{A}}^+$, every ISP n has the optimal attribute value \hat{a}_n^+ given competitor attributes $\hat{\mathbf{A}}_{-n}$, which can be found by setting $\partial\pi_n/\partial a_n = 0$:

$$\begin{aligned} & d' \left(\sum_{q=1}^Q \frac{\alpha_1 (1 + v_{-r(m_{q1}, m_{q2}, n)})}{(1 + \sum_{r' \in R(m_{q1}, m_{q2})} v_{r'})^2} \right) (\rho - \phi_1 a_n - \phi_0) \\ & - d' \phi_1 \left(\sum_{q=1}^Q \frac{v_{r(m_{q1}, m_{q2}, n)}}{1 + \sum_{r' \in R(m_{q1}, m_{q2})} v_{r'}} \right) - \gamma_1 = 0 \end{aligned} \quad (\text{F.64})$$

Since this equation is equivalent for every ISP n , the equilibrium \hat{a}_n^+ is identical for all ISPs n , i.e., $\hat{a}_n^+ = \hat{a}^+$. This simplification allows the following transformation of Eq. (F.65):

$$\begin{aligned} & d' Q \frac{\alpha_1 (1 + (Q - 1)(I\alpha_1 \hat{a}^+ + \alpha_0))}{(1 + Q(I\alpha_1 \hat{a}^+ + \alpha_0))^2} (\rho - \phi_1 a_n - \phi_0) \\ & - d' Q \phi_1 \frac{I\alpha_1 \hat{a}^+ + \alpha_0}{1 + Q(I\alpha_1 \hat{a}^+ + \alpha_0)} - \gamma_1 = 0 \end{aligned} \quad (\text{F.65})$$

This equilibrium condition is identical to the equilibrium condition for a homogeneous parallel-path network with a single origin-destination pair and demand limit $d = d'Q$. Hence, the unrestricted equilibrium value \hat{a}^+ from Theorem 8.2 (with $d'Q$ substituted for d) also applies to the competitive network \mathcal{N}_2 .

F.5.2 Equilibrium for Competition-Free Network \mathcal{N}_1

Moreover, we note that a single sub-network (for one origin-destination pair) of the competition-free network \mathcal{N}_1 is equivalent to the network \mathcal{N}_2 for $Q = 1$. Since

the identical, isolated sub-networks of the competition-free network \mathcal{N}_1 do not influence each other, the equilibrium attribute value \hat{a}^+ is equal in that whole network.

F.5.3 Comparison of Equilibria

Hence, if $\hat{a}^+(Q)$ is considered the equilibrium attribute for the competitive network \mathcal{N}_2 , we can prove the proposition $a^+(\mathcal{N}_2) \geq a^+(\mathcal{N}_1)$ for $Q \geq 1$ by showing $\hat{a}^+(Q) \geq \hat{a}^+(1)$ for $Q \geq 1$. To show this property, we solve the following inequality:

$$\hat{a}^+(Q) - \hat{a}^+(1) \geq 0 \quad (\text{F.66})$$

$$\iff \frac{\sqrt{T_2(Q)^2 - 4T_1(Q)T_3(Q)} - T_2(Q)}{2T_1(Q)} - \hat{a}^+(1) \geq 0 \quad (\text{F.67})$$

$$\iff \sqrt{T_2(Q)^2 - 4T_1(Q)T_3(Q)} \geq T_2(Q) + 2T_1(Q)\hat{a}^+(1) \quad (\text{F.68})$$

$$\iff T_2(Q)^2 - 4T_1(Q)T_3(Q) \geq (T_2(Q) + 2T_1(Q)\hat{a}^+(1))^2 \quad (\text{F.69})$$

In Eq. (F.67), the equilibrium constituent terms T_1 , T_2 , and T_3 from Theorem 8.2 are considered functions of Q . In Eq. (F.68), the transformation is possible by the fact that $T_1(Q) > 0$ for $Q \geq 1$:

$$\begin{aligned} T_1(Q) &= Q^2 I^2 \alpha_1^2 - \frac{Qd'}{Qd'\phi_1 + \gamma_1} (QI - 1)(Q - 1) I \alpha_1^2 \phi_1 > 0 \\ \iff Q &> \frac{d'\phi_1}{(d'\phi_1(I + 1) + I\gamma_1)} \end{aligned} \quad (\text{F.70})$$

where the RHS in the last inequality is consistently below 1, and $T_1(Q) > 0$ thus holds for all $Q \geq 1$. Using lengthy rewriting, the inequality in Eq. (F.69) can then be transformed into the following inequality containing a quadratic equation of Q :

$$T_7 Q^2 + T_8 Q + T_9 \leq 0 \quad (\text{F.71})$$

where

$$\begin{aligned} T_7 &= \hat{a}^+(1)^2 \alpha_1^2 I (I\gamma_1 + d\phi_1(I + 1)) \\ &\quad + \hat{a}^+(1) \alpha_1 (2I\alpha_0\gamma_1 + (2I + 1)d\phi_1\alpha_0 - d(\rho - \phi_0)I\alpha_1) \\ &\quad + \alpha_0^2 (d\phi_1 + \gamma_1) - d\alpha_0\alpha_1(\rho - \phi_0), \end{aligned} \quad (\text{F.72})$$

$$T_8 = -\hat{a}^+(1)^2 d I \alpha_1^2 \phi_1 \quad (\text{F.73})$$

$$\begin{aligned} & + \hat{a}^+(1) \alpha_1 (2I\gamma_1 + d\phi_1(I+1) - d\phi_1\alpha_0 + d(\rho - \phi_0) I\alpha_1) \\ & + 2\alpha_0\gamma_1 + (\alpha_0 - 1)d\alpha_1(\rho - \phi_0) + d\alpha_0\phi_1, \text{ and} \end{aligned}$$

$$T_9 = \gamma_1. \quad (\text{F.74})$$

To solve Eq. (F.71), we make use of the following two properties.

- $Q = 1$ is a root of $\hat{a}^+(Q) - \hat{a}^+(1)$, which implies:

$$T_7 + T_8 + T_9 = 0 \iff T_7 + T_8 = -T_9. \quad (\text{F.75})$$

- The inspection of T_7 yields the following insight, which we derived by means of the symbolic algebra system in MATLAB:

$$T_7 \leq \lim_{d,\alpha_0 \rightarrow 0} T_7 = \gamma_1 = T_9 \quad (\text{F.76})$$

Given the lower root \underline{Q} and the higher root \overline{Q} of the quadratic function in Eq. (F.71) (which are guaranteed to exist at least at $Q = 1$ and are identical if $T_7 = 0$), the inequality is solved by the following Q :

$$Q \in \begin{cases} [\underline{Q}, \overline{Q}] & \textcircled{1} \text{ if } T_7 > 0 \\ (-\infty, \underline{Q}] \cup [\overline{Q}, \infty) & \textcircled{2} \text{ if } T_7 < 0 \\ (-\infty, \underline{Q}] & \textcircled{3} \text{ if } T_7 = 0 \wedge T_8 > 0 \\ [\overline{Q}, \infty) & \textcircled{4} \text{ if } T_7 = 0 \wedge T_8 < 0 \\ (-\infty, \infty) & \textcircled{5} \text{ if } T_7 = 0 \wedge T_8 = 0 \wedge T_9 \leq 0 \\ \emptyset & \textcircled{6} \text{ if } T_7 = 0 \wedge T_8 = 0 \wedge T_9 > 0 \end{cases} \quad (\text{F.77})$$

This area of Q (leading to non-positive values of the quadratic function in Eq. (F.71)) includes $[1, \infty)$ in all cases:

1. $T_7 \neq 0$. (Eq. (F.77)① and ②): For $T_7 \neq 0$, the property in Eq. (F.75) facilitates finding the solutions $(\underline{Q}, \overline{Q})$:

$$(\underline{Q}, \overline{Q}) = \frac{-T_8 \pm \sqrt{T_8^2 - 4T_7T_9}}{2T_7} = \frac{-T_8 \pm \sqrt{(2T_7 + T_8)^2}}{2T_7} = \frac{-T_8 \pm (2T_7 + T_8)}{2T_7} \quad (\text{F.78})$$

a) $T_7 > 0$ (Eq. (F.77)(1)): For $T_7 > 0$, we note that

$$2T_7 + T_8 \stackrel{(F.75)}{=} T_7 - \gamma_1 \stackrel{(F.76)}{\leq} 0. \quad (\text{F.79})$$

Hence, the solutions from Eq. (F.78) are:

$$\begin{aligned} \underline{Q} &= \frac{-T_8 + (2T_7 + T_8)}{2T_7} = 1 \\ \overline{Q} &= \frac{-T_8 - (2T_7 + T_8)}{2T_7} = \frac{\gamma_1}{T_7} \geq 1 \end{aligned} \quad (\text{F.80})$$

where the higher solution \overline{Q} is spurious and has been introduced by the squaring operation in Eq. (F.69).

b) $T_7 < 0$: (Eq. (F.77)(2)): For $T_7 < 0$ (Eq. (F.77)(2)), the solutions are as follows:

$$\begin{aligned} \underline{Q} &= \frac{-T_8 - (2T_7 + T_8)}{2T_7} = \frac{\gamma_1}{T_7} < 0 \\ \overline{Q} &= \frac{-T_8 + (2T_7 + T_8)}{2T_7} = 1 \end{aligned} \quad (\text{F.81})$$

2. $T_7 = 0$ (Eq. (F.77)(3)–(6)): For $T_7 = 0$, it holds that $T_8 = -\gamma_1 - T_7 = -\gamma_1 \leq 0$ and $\overline{Q} = -T_9/T_8 = (-\gamma_1)/(-\gamma_1) = 1$.

- a) $T_8 > 0$ (Eq. (F.77)(3)): The case $T_8 > 0$ thus cannot arise.
- b) $T_8 < 0$ (Eq. (F.77)(4)): For $T_8 < 0$, the proposition clearly holds.
- c) $T_8 = 0$ (Eq. (F.77)(5) and (6)): For $T_8 = 0$, the equality $T_7 + T_8 = -T_9$ from Eq. (F.75) implies $T_9 = 0$. Hence, the case in Eq. (F.77)(5) always arises if $T_7 = T_8 = 0$, whereas the case in Eq. (F.77)(6) never arises.

Since Eq. (F.66) thus always holds for $Q \geq 1$, the proposition is proven.

F.6 PROOF OF THEOREM 8.6: HOMOGENEOUS COMPETITION (PROFITS)

To start the proof, we note that both the equilibrium attribute sum $a^+(\mathcal{N}_1)$ and the NBS attribute sum $a^\circ(\mathcal{N}_1)$ for the competition-free network are found by analyzing a single path, since the isolated sub-paths in the competition-free network do not influence each other. Hence, $a^+(\mathcal{N}_1)$ and $a^\circ(\mathcal{N}_1)$ are as in Theorem 8.4, which

relates to the single-path context and thus states that $a^+(\mathcal{N}_1) \leq a^\circ(\mathcal{N}_1)$. Therefore, the interval $[a^+(\mathcal{N}_1), a^\circ(\mathcal{N}_1)]$ is never empty.

From the proof of Theorem 8.5, we know that the proposition $\pi^+(\mathcal{N}_2) \geq \pi^+(\mathcal{N}_1)$ is equivalent to the proposition $\Delta\pi = \pi(Q, a^+(\mathcal{N}_2)) - \pi(1, a^+(\mathcal{N}_1)) \geq 0$, where $\pi(Q, a)$ is defined as follows:

$$\pi(Q, a) = Qd' \frac{I\alpha_1 a + \alpha_0}{1 + Q(I\alpha_1 a + \alpha_0)} (\rho - \phi_1 a - \phi_0) - \gamma_1 a - \gamma_0. \quad (\text{F.82})$$

Clearly, $\pi(Q, a^\circ(\mathcal{N}_1))$ is optimal for $Q = 1$, i.e., the NBS attribute sum is optimal in the competition-free network. Hence, it also holds that $\pi(1, a^+(\mathcal{N}_1)) \leq \pi(1, a^\circ(\mathcal{N}_1))$, i.e., the equilibrium profit in the competition-free network is generally sub-optimal. Moreover, since $\pi(Q, a)$ is consistently concave in a in the relevant regions, the assumption $a^+(\mathcal{N}_2) \in [a^+(\mathcal{N}_1), a^\circ(\mathcal{N}_1)]$ implies

$$\pi(1, a^+(\mathcal{N}_1)) \leq \pi(1, a^+(\mathcal{N}_2)). \quad (\text{F.83})$$

Given Eq. (F.83), we can lower bound the profit difference:

$$\begin{aligned} \Delta\pi &= \pi(Q, a^+(\mathcal{N}_2)) - \pi(1, a^+(\mathcal{N}_1)) \\ &\geq \pi(Q, a^+(\mathcal{N}_2)) - \pi(1, a^+(\mathcal{N}_2)) =: \underline{\Delta\pi} \end{aligned} \quad (\text{F.84})$$

Hence, if $\underline{\Delta\pi} \geq 0$ holds, the proof proposition $\Delta\pi \geq 0$ follows. At this point, we also note that $a^+(\mathcal{N}_2) \in [a^+(\mathcal{N}_1), a^\circ(\mathcal{N}_1)]$ is only a sufficient, but not a necessary condition for $\underline{\Delta\pi} \geq 0$; hence, profit increases might also happen if $a^+(\mathcal{N}_2) \notin [a^+(\mathcal{N}_1), a^\circ(\mathcal{N}_1)]$.

We can reformulate the lower bound $\underline{\Delta\pi}$ on the profit difference as follows:

$$\begin{aligned} \underline{\Delta\pi} &= \pi(Q, a^+(\mathcal{N}_2)) - \pi(1, a^+(\mathcal{N}_2)) \\ &= d' \left(\frac{Q(I\alpha_1 a^+(\mathcal{N}_2) + \alpha_0)}{1 + Q(I\alpha_1 a^+(\mathcal{N}_2) + \alpha_0)} - \frac{I\alpha_1 a^+(\mathcal{N}_2) + \alpha_0}{1 + I\alpha_1 a^+(\mathcal{N}_2) + \alpha_0} \right) \Psi(a^+(\mathcal{N}_2)) \\ &= d' \left(\frac{(Q-1)(I\alpha_1 a^+(\mathcal{N}_2) + \alpha_0)}{(1 + Q(I\alpha_1 a^+(\mathcal{N}_2) + \alpha_0))(1 + I\alpha_1 a^+(\mathcal{N}_2) + \alpha_0)} \right) \Psi(a^+(\mathcal{N}_2)) \end{aligned} \quad (\text{F.85})$$

Given $d' > 0$ and $Q \geq 1$, the first and second factor of $\underline{\Delta\pi}$ in Eq. (F.85) are non-negative. Hence, $\underline{\Delta\pi} \geq 0$ is equivalent to

$$\Psi(a^+(\mathcal{N}_2)) = (\rho - \phi_1 a^+(\mathcal{N}_2) - \phi_0) \geq 0. \quad (\text{F.86})$$

This latter condition also always holds, which is demonstrable by contradiction. Let $\rho - \phi_1 a^+(\mathcal{N}_2) - \phi_0 < 0 \iff a^+(\mathcal{N}_2) > (\rho - \phi_0)/\phi_1$, which makes the

minuend in $\pi(1, a^+(\mathcal{N}_2))$ negative (cf. Eq. (F.82)). In that case, all $a' > a^+(\mathcal{N}_2)$ would lead to lower profit $\pi(1, a')$. This observation contradicts the optimality of the NBS attribute sum $a^\circ(\mathcal{N}_1)$ regarding $\pi(1, a')$, as $a^\circ(\mathcal{N}_1) \geq a^+(\mathcal{N}_2)$.

Hence, since $\underline{\Delta\pi} \geq 0$, it holds that $\Delta\pi \geq 0$ and the theorem proposition follows.

F.7 PROOF OF THEOREM 8.7: OPTIMUM ON HETEROGENEOUS PATHS

In order to be a Nash bargaining solution, the attribute values \mathbf{A}° should both optimize the aggregate profit function $\pi(\mathbf{A}) = \sum_{n \in r} \pi_n(\mathbf{A})$, and create a maximally equitable profit distribution across the ISPs $n \in r$. This maximum fairness is achieved by optimizing the Nash bargaining product, i.e.:

$$\mathbf{A}^\circ = \arg \max_{\mathbf{A} \in \mathbb{R}_{\geq 0}} \prod_{n \in r} \pi_n(\mathbf{A}) \quad (\text{F.87})$$

This optimization of the Nash bargaining product must be performed subject to the constraints in Theorem 8.7 that are associated with optimal aggregate profit. In the following, we characterize this aggregate-profit function, and show that the conditions stated in Theorem 8.7 are both sufficient and necessary in order for \mathbf{A}° to satisfy aggregate-profit optimality.

F.7.1 Aggregate-Profit Function

The aggregate profit $\pi(\mathbf{A})$ in our setting is:

$$\pi(\mathbf{A}) = \sum_{n \in r} \pi_n(\mathbf{A}) = d \frac{v_r(\mathbf{A})}{1 + v_r(\mathbf{A})} \left(\sum_{n \in r} \rho_n - \phi_{n0} \right) - \sum_{n \in r} \left(\sum_{k \in K} \gamma_{nk} a_{nk} + \gamma_{n0} \right) \quad (\text{F.88})$$

This aggregate-profit function has the following first and second derivative in any a_{nk} :

$$\frac{\partial}{\partial a_{nk}} \pi(\mathbf{A}) = d \frac{\alpha_{nk}}{(1 + v_r(\mathbf{A}))^2} \left(\sum_{n \in r} \rho_n - \phi_{n0} \right) - \gamma_{nk} \quad (\text{F.89})$$

$$\frac{\partial^2}{\partial a_{nk}^2} \pi(\mathbf{A}) = -d \frac{2\alpha_{nk}^2}{(1 + v_r(\mathbf{A}))^3} \left(\sum_{n \in r} \rho_n - \phi_{n0} \right) \quad (\text{F.90})$$

As the second derivative is non-positive for all $\mathbf{A} \in \mathbb{R}_{\geq 0}$, the aggregate-profit function is consistently concave in any a_{nk} on the valid domain $\mathbb{R}_{\geq 0}$. Therefore, if the first derivative $\partial/\partial a_{nk} \pi(\mathbf{A})$ is negative for any a_{nk} , all reductions of a_{nk}

increase aggregate profit, and all increases of a_{nk} reduce the aggregate profit (The reverse holds for a positive first derivative). This condition on the first derivative is equivalent to the following condition, which is central for the proof:

$$\forall n \in r, k \in K. \quad \frac{\partial}{\partial a_{nk}} \pi(\mathbf{A}) < 0 \iff v_r(\mathbf{A}) > \sqrt{\frac{\alpha_{nk}}{\gamma_{nk}}} \sqrt{d \sum_{n \in r} (\rho_n - \phi_{n0})} - 1 \quad (\text{F.91})$$

F.7.2 Sufficiency of Conditions

After this characterization of the aggregate-profit function, we now demonstrate that the conditions in Theorem 8.7 are *sufficient*, i.e., any \mathbf{A}° with the conditions optimizes the aggregate profit. Sufficiency can be demonstrated by performing the following case distinction:

$$1. \quad \forall (n, k) \in r \times K. \quad \alpha_{r0} > \sqrt{\frac{\alpha_{nk}}{\gamma_{nk}}} \sqrt{d \sum_{n \in r} (\rho_n - \phi_{n0})} - 1$$

According to Theorem 8.7, all optimal attribute values \mathbf{A}° must be $\mathbf{0}$ in this case:

$$v_r(\mathbf{A}^\circ) = v_r^\circ \stackrel{(8.26)}{=} \alpha_{r0} \iff \mathbf{A}^\circ = \mathbf{0} \quad (\text{F.92})$$

Moreover, the first derivative $\partial/\partial a_{nk} \pi(\mathbf{A}^\circ)$ for all $n \in r, k \in K$ must be negative:

$$\begin{aligned} \forall n \in r, k \in K. \quad v_r(\mathbf{A}^\circ) = \alpha_{r0} &> \sqrt{\frac{\alpha_{nk}}{\gamma_{nk}}} \sqrt{d \sum_{n \in r} (\rho_n - \phi_{n0})} - 1 \\ \stackrel{(\text{F.91})}{\iff} \quad \frac{\partial}{\partial a_{nk}} \pi(\mathbf{A}^\circ) &< 0 \end{aligned} \quad (\text{F.93})$$

Hence, the only way to further increase the aggregate profit π would be by reductions in any a_{nk} . However, since every $a_{nk} = 0$, such reductions are not possible given the restricted domain $\mathbb{R}_{\geq 0}$. Hence, $\mathbf{A}^\circ = \mathbf{0}$ is optimal.

$$2. \quad \exists (n, k) \in r \times K. \quad \alpha_{r0} \leq \sqrt{\frac{\alpha_{nk}}{\gamma_{nk}}} \sqrt{d \sum_{n \in r} (\rho_n - \phi_{n0})} - 1$$

In this case, the attributes $(n^\circ, k^\circ) \in r \times K$ with the maximal ratio $\alpha_{n^\circ k^\circ} / \gamma_{n^\circ k^\circ}$ play a special role according to Theorem 8.7. We denote the set of these attributes by K_r° :

$$K_r^\circ = \left\{ (n^\circ, k^\circ) \mid (n^\circ, k^\circ) = \arg \max_{(n, k) \in r \times K} \frac{\alpha_{nk}}{\gamma_{nk}} \right\}. \quad (\text{F.94})$$

This maximal ratio also determines the optimal path valuation $v_r(\mathbf{A}^\circ)$ according to Theorem 8.7:

$$\forall (n^\circ, k^\circ) \in K_r^\circ. v_r(\mathbf{A}^\circ) = \sqrt{\frac{\alpha_{n^\circ k^\circ}}{\gamma_{n^\circ k^\circ}}} \sqrt{d \sum_{n \in r} (\rho_n - \phi_{n0})} - 1 \quad (\text{F.95})$$

In contrast, for all attributes $(n^\circ, k^\circ) \notin K_r^\circ$, the following condition holds:

$$\begin{aligned} \forall (n^\circ, k^\circ) \notin K_r^\circ. v_r(\mathbf{A}^\circ) &= \sqrt{\frac{\alpha_{n^\circ k^\circ}}{\gamma_{n^\circ k^\circ}}} \sqrt{d \sum_{n \in r} (\rho_n - \phi_{n0})} - 1 \\ &> \sqrt{\frac{\alpha_{n^\circ k^\circ}}{\gamma_{n^\circ k^\circ}}} \sqrt{d \sum_{n \in r} (\rho_n - \phi_{n0})} - 1 \\ \iff \frac{\partial}{\partial a_{n^\circ k^\circ}} \pi(\mathbf{A}^\circ) &< 0 \end{aligned} \quad (\text{F.96})$$

Hence, the only way to increase the aggregate profit π would be by reductions in any $a_{n^\circ k^\circ}$ for $(n^\circ, k^\circ) \notin K_r^\circ$. However, since $a_{n^\circ k^\circ} = 0 \ \forall (n^\circ, k^\circ) \notin K_r^\circ$ by Theorem 8.7, such reductions are not possible, and hence \mathbf{A}° is optimal.

F.7.3 Necessity of Conditions

After demonstrating that the conditions in Theorem 8.7 are sufficient for optimal aggregate profit, we now demonstrate that the conditions are also *necessary*, i.e., no choice of attribute values \mathbf{A}° that violates these conditions can be optimal. For the sake of contradiction, we assume that some attribute values \mathbf{A}° are optimal while satisfying the following conditions:

$$\exists (n^\circ, k^\circ) \notin K_r^\circ. a_{n^\circ k^\circ} > 0. \quad (\text{F.97})$$

A contradiction can be produced in all cases of the following case distinction:

$$1. \forall (n, k) \in r \times K. \alpha_{r0} > \sqrt{\frac{\alpha_{nk}}{\gamma_{nk}}} \sqrt{d \sum_{n \in r} (\rho_n - \phi_{n0})} - 1$$

Since $a_{n^\circ k^\circ} > 0$ for the fixed attribute (n°, k°) , the optimal path valuation $v_r(\mathbf{A}^\circ)$ exceeds α_{r0} , and hence:

$$\begin{aligned} \forall (n, k) \in r \times K. v_r(\mathbf{A}^\circ) &> \alpha_{r0} > \sqrt{\frac{\alpha_{nk}}{\gamma_{nk}}} \sqrt{d \sum_{n \in r} (\rho_n - \phi_{n0})} - 1 \\ \implies \frac{\partial}{\partial a_{n^\circ k^\circ}} \pi(\mathbf{A}^\circ) &< 0 \end{aligned} \quad (\text{F.98})$$

The aggregate profit can thus be increased by reducing $a_{n^0 k^0}^\circ$, and such a reduction is also possible since $a_{n^0 k^0} > 0$. Hence, the attribute values \mathbf{A}° are not optimal, which causes a contradiction.

$$2. \exists (n, k) \in r \times K. \quad \alpha_{r0} \leq \sqrt{\frac{\alpha_{nk}}{\gamma_{nk}}} \sqrt{d \sum_{n \in r} (\rho_n - \phi_{n0})} - 1$$

In that case, we perform a sub-case distinction on the value of $v_r(\mathbf{A}^\circ)$:

$$a) v_r(\mathbf{A}^\circ) \leq \sqrt{\frac{\alpha_{n^0 k^0}}{\gamma_{n^0 k^0}}} \sqrt{d \sum_{n \in r} (\rho_n - \phi_{n0})} - 1$$

Since $a_{n^0 k^0} / \gamma_{n^0 k^0} < a_{n^0 k^0} / \gamma_{n^0 k^0}$, it follows that:

$$v_r(\mathbf{A}^\circ) < \sqrt{\frac{\alpha_{n^0 k^0}}{\gamma_{n^0 k^0}}} \sqrt{d \sum_{n \in r} (\rho_n - \phi_{n0})} - 1 \iff \frac{\partial}{\partial a_{n^0 k^0}} \pi(\mathbf{A}) > 0, \quad (\text{F.99})$$

which implies that the profit can be increased by increasing the value of $a_{n^0 k^0} \forall (n, k) \in K_r^\circ$, which contradicts the assumption that \mathbf{A}° is optimal.

$$b) v_r(\mathbf{A}^\circ) > \sqrt{\frac{\alpha_{n^0 k^0}}{\gamma_{n^0 k^0}}} \sqrt{d \sum_{n \in r} (\rho_n - \phi_{n0})} - 1$$

This condition implies:

$$\frac{\partial}{\partial a_{n^0 k^0}} \pi(\mathbf{A}^\circ) < 0. \quad (\text{F.100})$$

Hence, the profit can be increased by reducing $a_{n^0 k^0}$, which is possible given $a_{n^0 k^0} > 0$. Therefore, we again produce a contradiction to the optimality of \mathbf{A}° .

We have thus identified the conditions on \mathbf{A}° that are sufficient and necessary for optimal aggregate profit. Thereby, the theorem is proven.

F.8 PROOF OF THEOREM 8.8: EQUILIBRIUM ON HETEROGENEOUS PATHS

Since the equilibrium conditions in Theorem 8.8 are highly similar to the optimality conditions in Theorem 8.7, the proof of Theorem 8.8 is analogous to the proof of Theorem 8.7. The proof is analogous because the derivatives of the individual profit functions π_n have equivalent properties to the derivatives of the aggregate-profit function π from Eq. (F.88). In particular, every first derivative satisfies:

$$\forall n \in r, k \in K. \quad \frac{\partial \pi_n(\mathbf{A})}{\partial a_{nk}} < 0 \iff v_r(\mathbf{A}) > \sqrt{\frac{\alpha_{nk}}{\gamma_{nk}}} \sqrt{d(\rho_n - \phi_{n0})} - 1 \quad (\text{F.101})$$

Moreover, every individual profit function π_n is consistently concave in any relevant attribute a_{nk} :

$$\forall n \in r, k \in K. \frac{\partial^2}{\partial a_{nk}^2} \pi_n(\mathbf{A}) = -d \frac{2\alpha_{nk}^2}{(1 + v_r(\mathbf{A}))^3} (\rho_n - \phi_{n0}) \quad (\text{F.102})$$

Building on these properties, the equilibrium conditions can be shown to be sufficient and necessary analogously to Theorem 8.7.

F.9 PROOF OF THEOREM 8.9: SUBOPTIMAL HETEROGENEOUS EQUILIBRIUM

In order to show that $v_r(\mathbf{A}^+) \leq v_r(\mathbf{A}^\circ)$, it is enough to show that:

$$\sqrt{\frac{\alpha_{n^+k^+}}{\gamma_{n^+k^+}}} \sqrt{d(\rho_{n^+} - \phi_{n^+0})} - 1 \leq \sqrt{\frac{\alpha_{n^\circ k^\circ}}{\gamma_{n^\circ k^\circ}}} \sqrt{d \sum_{n \in r} (\rho_n - \phi_{n0})} - 1 \quad (\text{F.103})$$

where $(n^+, k^+) \in K_r^+$ and $(n^\circ, k^\circ) \in K_r^\circ$. This inequality can be transformed into the following form:

$$\sqrt{\frac{\alpha_{n^+k^+}}{\gamma_{n^+k^+}}} \sqrt{\frac{d(\rho_{n^+} - \phi_{n^+0})}{d \sum_{n \in r} (\rho_n - \phi_{n0})}} \leq \sqrt{\frac{\alpha_{n^\circ k^\circ}}{\gamma_{n^\circ k^\circ}}}. \quad (\text{F.104})$$

Thanks to the following two insights, this inequality is always satisfied:

$$\sqrt{\frac{d(\rho_{n^+} - \phi_{n^+0})}{d \sum_{n \in r} (\rho_n - \phi_{n0})}} \leq 1 \quad \frac{\alpha_{n^+k^+}}{\gamma_{n^+k^+}} \leq \frac{\alpha_{n^\circ k^\circ}}{\gamma_{n^\circ k^\circ}} = \max_{(n,k) \in r \times K} \frac{\alpha_{nk}}{\gamma_{nk}} \quad (\text{F.105})$$

Hence, the theorem holds.

F.10 PROOF OF THEOREM 8.10: TWO-PATH HETEROGENEOUS EQUILIBRIUM

F.10.1 Analogy to Theorem 8.8

To start off, we once more characterize the derivatives of the individual-profit functions π_n for any attribute value a_{nk} :

$$\forall n \in N, k \in K, r = r(n). \quad (\text{F.106})$$

$$\frac{\partial \pi_n(\mathbf{A})}{\partial a_{nk}} = d \frac{\alpha_{nk} (1 + v_s(\mathbf{A}))}{(1 + v_r(\mathbf{A}) + v_s(\mathbf{A}))^2} (\rho_n - \phi_{n0}) - \gamma_{nk} \quad (\text{F.107})$$

$$\frac{\partial^2 \pi_n(\mathbf{A})}{\partial a_{nk}^2} = -d \frac{2\alpha_{nk}^2 (1 + v_s(\mathbf{A}))}{(1 + v_r(\mathbf{A}) + v_s(\mathbf{A}))^3} (\rho_n - \phi_{n0}) \quad (\text{F.108})$$

For better readability, this proof denotes the alternative path \bar{r} to path r by s .

Since the second derivative is never positive, every profit function π_n is consistently concave in the attribute values controlled by ISP n . Hence, a negative first derivative $\partial \pi(\mathbf{A}) / \partial a_{nk} < 0$ indicates that a_{nk} must be reduced if the profit is to be increased. The case of a negative first derivative in a_{nk} can be expressed as follows (for $r = r(n)$):

$$\frac{\partial \pi_n(\mathbf{A})}{\partial a_{nk}} < 0 \iff v_r(\mathbf{A}) > \sqrt{\frac{\alpha_{nk}}{\gamma_{nk}}} \sqrt{d(\rho_n - \phi_{n0})} \sqrt{1 + v_s(\mathbf{A})} - (1 + v_s(\mathbf{A})) \quad (\text{F.109})$$

When thinking of Eq. (F.109) as an extension of Eq. (F.91) with v_s as a fixed quantity, an analogous proof to the proof of Theorem 8.7 can be performed. The extension by fixed v_s does not change the finding that only attributes $(n^\circ, k^\circ) \in K_r^\circ$ might have non-zero values in equilibrium. However, the extension by v_s changes the equilibrium path valuation v_r^+ from Eq. (8.29) to:

$$v_r^+ = \max \left(\alpha_{r0}, \sqrt{\frac{\alpha_{n^\circ k^\circ}}{\gamma_{n^\circ k^\circ}}} \sqrt{d(\rho_{n^\circ} - \phi_{n^\circ 0})} \sqrt{1 + v_s} - (1 + v_s) \right). \quad (\text{F.110})$$

Crucially, this condition simultaneously holds for both paths r and s in a two-path scenario, creating an interdependence of the equilibrium path valuations:

$$\begin{aligned} \forall r \in R. \quad v_r^+ &= \max \left(\alpha_{r0}, \sqrt{\frac{\alpha_{n^\circ k^\circ}}{\gamma_{n^\circ k^\circ}}} \sqrt{d(\rho_{n^\circ} - \phi_{n^\circ 0})} \sqrt{1 + v_s^+} - (1 + v_s^+) \right) \\ &= \max \left(\alpha_{r0}, \psi_r \sqrt{d} \sqrt{1 + v_s^+} - (1 + v_s^+) \right) \\ &= \max \left(\alpha_{r0}, \hat{v}_r^*(v_s^+) \right), \end{aligned} \quad (\text{F.111})$$

where $\hat{v}_r^*(v_s)$ is the *unrestricted best-response valuation* for path r given competing-path valuation v_s . The *characteristic ratio* ψ_r is reflected in Eq. (8.34).

The remainder of the proof illustrates how to derive the equilibrium path valuations v_r^+ and v_s^+ .

F.10.2 Unrestricted Equilibrium \hat{v}_r^+

Considering a relaxed setting in which the constraint $\mathbf{A}^+ \in \mathbb{R}_{\geq 0}^{|N| \times |K|} \iff v_r^+ \geq \alpha_{r0}$ is ignored, the unrestricted equilibrium path valuations \hat{v}_r^+ satisfy the following system of equations:

$$\forall r \in R. \quad \hat{v}_r^+ = \hat{v}_r^*(\hat{v}_s^+). \quad (\text{F.112})$$

In this relaxed setting, this system of two equations can be conventionally solved for the unrestricted equilibrium path valuations $\hat{v}_r^+, r \in R$, resulting in the unique solution denoted in Eq. (8.33).

Moreover, we make the following important observation:

$$\hat{v}_r^*(\hat{v}_s^*(v_r)) \leq v_r \iff v_r \geq \hat{v}_r^+. \quad (\text{F.113})$$

F.10.3 Restricted Equilibrium v_r^+

We now rely on the equilibrium gained by relaxation to characterize the equilibrium under the re-introduced constraint $v_r^+ \geq \alpha_{r0} \forall r \in R$. In particular, we want to show that the calculation provided in Theorem 8.10 is correct:

$$\forall r \in R. \quad v_r^+ = \max(\alpha_{r0}, \hat{v}_r^*(\max(\alpha_{s0}, \hat{v}_s^+))) \quad (\text{F.114})$$

In other words, v_r^+ as calculated by Eq. (F.114) should satisfy the equilibrium conditions on v_r^+ in Eq. (F.111). To satisfy these conditions, we consider all cases regarding \hat{v}_r^+ and \hat{v}_s^+ :

1. $\hat{v}_r^+ \geq \alpha_{r0}$:

a) $\hat{v}_s^+ \geq \alpha_{s0}$: In that case, Eq. (F.114) suggests that

$$\begin{aligned} v_r^+ &\stackrel{1.(a)}{=} \max(\alpha_{r0}, \hat{v}_r^*(\hat{v}_s^+)) \stackrel{(\text{F.112})}{=} \max(\alpha_{r0}, \hat{v}_r^+) \stackrel{1.}{=} \hat{v}_r^+, \text{ and} \\ v_s^+ &\stackrel{1.}{=} \max(\alpha_{s0}, \hat{v}_s^*(\hat{v}_r^+)) \stackrel{(\text{F.112})}{=} \max(\alpha_{s0}, \hat{v}_s^+) \stackrel{1.(a)}{=} \hat{v}_s^+. \end{aligned} \quad (\text{F.115})$$

These values satisfy the equilibrium conditions in Eq. (F.111):

$$\begin{aligned} v_r^+ &\stackrel{(\text{F.111})}{=} \max(\alpha_{r0}, \hat{v}_r^*(v_s^+)) \stackrel{(\text{F.115})}{=} \max(\alpha_{r0}, \hat{v}_r^*(\hat{v}_s^+)) \\ &\stackrel{(\text{F.112})}{=} \max(\alpha_{r0}, \hat{v}_r^+) \stackrel{1.}{=} \hat{v}_r^+, \end{aligned} \quad (\text{F.116})$$

and symmetrically for v_s^+ .

b) $\hat{v}_s^+ < \alpha_{s0}$: In that case, Eq. (F.114) suggests that

$$\begin{aligned} v_r^+ &\stackrel{1.(b)}{=} \max(\alpha_{r0}, \hat{v}_r^*(\alpha_{s0})) \\ v_s^+ &\stackrel{1.}{=} \max(\alpha_{s0}, \hat{v}_s^*(\hat{v}_r^+)) \stackrel{(F.112)}{=} \max(\alpha_{s0}, \hat{v}_s^+) \stackrel{1.(b)}{=} \alpha_{s0}. \end{aligned} \quad (\text{F.117})$$

For that case, we perform another level of sub-case distinctions:

i. $\hat{v}_r^*(\alpha_{s0}) \geq \alpha_{r0}$: In that case, Eq. (F.117) is simplified to

$$\begin{aligned} v_r^+ &\stackrel{(F.117)}{=} \max(\alpha_{r0}, \hat{v}_r^*(\alpha_{s0})) \stackrel{1.(b).i}{=} \hat{v}_r^*(\alpha_{s0}) \\ v_s^+ &\stackrel{(F.117)}{=} \alpha_{s0} \end{aligned} \quad (\text{F.118})$$

Using Eq. (F.113) and case condition 1.(b), we can also deduce:

$$\hat{v}_s^*(\hat{v}_r^*(\alpha_{s0})) < \alpha_{s0}. \quad (\text{F.119})$$

Then, we can again verify that these values satisfy the equilibrium conditions from Eq. (F.111):

$$\begin{aligned} v_r^+ &\stackrel{(F.111)}{=} \max(\alpha_{r0}, \hat{v}_r^*(v_s^+)) \stackrel{(F.118)}{=} \max(\alpha_{r0}, \hat{v}_r^*(\alpha_{s0})) \stackrel{1.b.i}{=} \hat{v}_r^*(\alpha_{s0}) \\ v_s^+ &\stackrel{(F.111)}{=} \max(\alpha_{s0}, \hat{v}_s^*(v_r^+)) \stackrel{(F.118)}{=} \max(\alpha_{s0}, \hat{v}_s^*(\hat{v}_r^*(\alpha_{s0}))) \stackrel{(F.119)}{=} \alpha_{s0}. \end{aligned} \quad (\text{F.120})$$

ii. $\hat{v}_r^*(\alpha_{s0}) < \alpha_{r0}$: In that case, Eq. (F.117) is simplified to

$$\begin{aligned} v_r^+ &\stackrel{(F.117)}{=} \max(\alpha_{r0}, \hat{v}_r^*(\alpha_{s0})) \stackrel{1.(b).ii}{=} \alpha_{r0} \\ v_s^+ &\stackrel{(F.117)}{=} \alpha_{s0} \end{aligned} \quad (\text{F.121})$$

Moreover, as proven in F.10.3.1, the current case implies

$$\hat{v}_s^*(\alpha_{r0}) \leq \alpha_{s0}. \quad (\text{F.122})$$

Based on these findings, the equilibrium conditions from Eq. (F.111) are satisfied:

$$\begin{aligned} v_r^+ &\stackrel{(F.111)}{=} \max(\alpha_{r0}, \hat{v}_r^*(v_s^+)) \stackrel{(F.121)}{=} \max(\alpha_{r0}, \alpha_{r0}) = \alpha_{r0} \\ v_s^+ &\stackrel{(F.111)}{=} \max(\alpha_{s0}, \hat{v}_s^*(v_r^+)) \\ &\stackrel{(F.121)}{=} \max(\alpha_{s0}, \hat{v}_s^*(\alpha_{r0})) \stackrel{(F.122)}{=} \alpha_{s0}. \end{aligned} \quad (\text{F.123})$$

2. $\hat{v}_r^+ < \alpha_{r0}$:

a) $\hat{v}_s^+ \geq \alpha_{s0}$: This case is symmetric to case 1.(b).

b) $\hat{v}_s^+ < \alpha_{s0}$: In that case, Eq. (F.114) suggests that

$$\begin{aligned} v_r^+ &\stackrel{2.(b)}{=} \max(\alpha_{r0}, \hat{v}_r^*(\alpha_{s0})) \\ v_s^+ &\stackrel{2.}{=} \max(\alpha_{s0}, \hat{v}_s^*(\alpha_{r0})) \end{aligned} \quad (\text{F.124})$$

Hence, we again need to perform another level of sub-case distinctions:

i. $\hat{v}_r^*(\alpha_{s0}) \geq \alpha_{r0}$: For that case, we show in F.10.3.2 that

$$\hat{v}_s^*(\alpha_{r0}) \leq \alpha_{s0}. \quad (\text{F.125})$$

Hence, Eq. (F.124) simplifies to:

$$\begin{aligned} v_r^+ &\stackrel{2.(b)}{=} \max(\alpha_{r0}, \hat{v}_r^*(\alpha_{s0})) \stackrel{2.(b).i}{=} \hat{v}_r^*(\alpha_{s0}) \\ v_s^+ &\stackrel{2.}{=} \max(\alpha_{s0}, \hat{v}_s^*(\alpha_{r0})) \stackrel{(\text{F.125})}{=} \alpha_{s0} \end{aligned} \quad (\text{F.126})$$

Moreover, using Eq. (F.113) and case condition 2.(b), we can again deduce:

$$\hat{v}_s^*(\hat{v}_s^*(\alpha_{r0})) < \alpha_{s0}. \quad (\text{F.127})$$

Based on these findings, the equilibrium conditions from Eq. (F.111) are satisfied:

$$\begin{aligned} v_r^+ &\stackrel{(\text{F.111})}{=} \max(\alpha_{r0}, \hat{v}_r^*(v_s^+)) \\ &\stackrel{(\text{F.126})}{=} \max(\alpha_{r0}, \hat{v}_r^*(\alpha_{s0})) \stackrel{2.(b).i}{=} \hat{v}_r^*(\alpha_{s0}) \\ v_s^+ &\stackrel{(\text{F.111})}{=} \max(\alpha_{s0}, \hat{v}_s^*(v_r^+)) \\ &\stackrel{(\text{F.126})}{=} \max(\alpha_{s0}, \hat{v}_s^*(\hat{v}_r^*(\alpha_{s0}))) \stackrel{(\text{F.127})}{=} \alpha_{s0} \end{aligned} \quad (\text{F.128})$$

ii. $\hat{v}_r^*(\alpha_{s0}) < \alpha_{r0}$: In that case, Eq. (F.124) simplifies to:

$$\begin{aligned} v_r^+ &\stackrel{2.(b)}{=} \max(\alpha_{r0}, \hat{v}_r^*(\alpha_{s0})) \stackrel{2.(b).ii}{=} \alpha_{r0} \\ v_s^+ &\stackrel{2.}{=} \max(\alpha_{s0}, \hat{v}_s^*(\alpha_{r0})) \end{aligned} \quad (\text{F.129})$$

To further simplify Eq. (F.129), we perform another sub-case distinction:

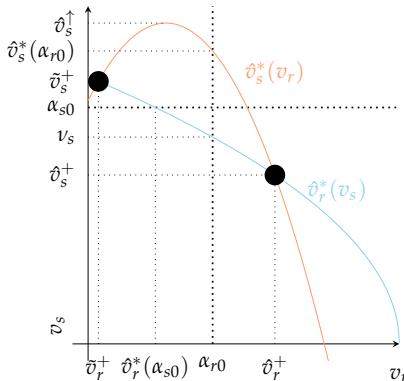


Figure F.1: Visualization of non-unique equilibrium $(\hat{v}_r^+, \hat{v}_s^+)$ in F.10.3.1.

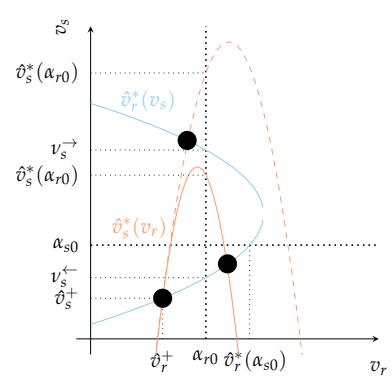


Figure F.2: Visualization of non-unique equilibrium $(\hat{v}_r^+, \hat{v}_s^+)$ in F.10.3.2.

A. $\hat{v}_s^*(\alpha_{r0}) \geq \alpha_{s0}$: This case is symmetric to case 2.(b).i.

B. $\hat{v}_s^*(\alpha_{r0}) < \alpha_{s0}$: In that case, Eq. (F.129) directly simplifies to:

$$\begin{aligned} v_r^+ &\stackrel{2.(b)}{=} \max(\alpha_{r0}, \hat{v}_r^*(\alpha_{s0})) \stackrel{2.(b).ii}{=} \alpha_{r0} \\ v_s^+ &\stackrel{2.}{=} \max(\alpha_{s0}, \hat{v}_s^*(\alpha_{r0})) \stackrel{2.(b).ii.B}{=} \alpha_{s0} \end{aligned} \quad (\text{F.130})$$

Clearly, the equilibrium conditions from Eq. (F.111) are satisfied by these values:

$$\begin{aligned} v_r^+ &\stackrel{(\text{F.111})}{=} \max(\alpha_{r0}, \hat{v}_r^*(v_s^+)) \\ &\stackrel{(\text{F.130})}{=} \max(\alpha_{r0}, \hat{v}_r^*(\alpha_{s0})) \stackrel{2.(b).ii}{=} \alpha_{r0} \\ v_s^+ &\stackrel{(\text{F.111})}{=} \max(\alpha_{s0}, \hat{v}_s^*(v_r^+)) \\ &\stackrel{(\text{F.130})}{=} \max(\alpha_{s0}, \hat{v}_s^*(\alpha_{r0})) \stackrel{2.(b).ii.B}{=} \alpha_{s0} \end{aligned} \quad (\text{F.131})$$

Since the values calculated according to Eq. (F.114) always satisfy the restricted-equilibrium conditions from Eq. (F.111), the theorem is proven.

F.10.3.1 Upper Bound on $\hat{v}_s^*(\alpha_{r0})$ for Case 1.(b).ii

We can show that case 1.(b).ii implies

$$\hat{v}_s^*(\alpha_{r0}) \leq \alpha_{s0}. \quad (\text{F.132})$$

In particular, let us assume the opposite for the sake of contradiction, i.e., we assume

$$\hat{v}_s^*(\alpha_{r0}) > \alpha_{s0}. \quad (\text{F.133})$$

For this proof, we first investigate the functions \hat{v}_r^* and \hat{v}_s^* more thoroughly, and then produce a contradiction by showing the existence of a second unrestricted equilibrium $(\tilde{v}_r^+, \tilde{v}_s^+) \neq (\hat{v}_r^+, \hat{v}_s^+)$. The proof idea is visualized in Fig. F.1.

\hat{v}_r^* . In case 1.(b).ii, we can more precisely characterize the function \hat{v}_r^* based on the case conditions. In particular, we know that \hat{v}_r^* evolves from value $\hat{v}_r^*(\hat{v}_s^+) = \hat{v}_r^+ \geq \alpha_{r0}$ (1.) at argument \hat{v}_s^+ down to value $\hat{v}_r^*(\alpha_{s0}) < \alpha_{r0}$ (1.(b).ii) at argument $\alpha_{s0} > \hat{v}_s^+$ (1.(b)). Hence, the intermediate-value theorem and the concavity of \hat{v}_r^* stipulate

$$\exists v_s \in [\hat{v}_s^+, \alpha_{s0}] . \quad \hat{v}_r^*(v_s) = \alpha_{r0} \quad \text{and} \quad \forall v_s > v_s . \quad \hat{v}_r^*(v_s) < \alpha_{r0}. \quad (\text{F.134})$$

\hat{v}_s^* . The assumption in Eq. (F.133) suggests that \hat{v}_s^* reaches a value above α_{s0} at argument α_{r0} . Hence, the maximum of \hat{v}_s^* is also at least α_{s0} :

$$\hat{v}_s^\uparrow = \max_{v_r} \hat{v}_s^*(v_r) \stackrel{(\text{F.134})}{\geq} \alpha_{s0} > v_s. \quad (\text{F.135})$$

Contradiction. Based on these properties of \hat{v}_r^* and \hat{v}_s^* , we now show that there exist $\tilde{v}_r^+ < \alpha_{r0}$ and $\tilde{v}_s^+ > v_s$, with the unrestricted-equilibrium properties $\tilde{v}_r^+ = \hat{v}_r^*(\tilde{v}_s^+)$ and $\tilde{v}_s^+ = \hat{v}_s^*(\tilde{v}_r^+)$. To verify the existence of these valuations, note that the value \tilde{v}_s^+ satisfies the condition:

$$\tilde{v}_s^+ = \hat{v}_s^*(\tilde{v}_r^+) = \hat{v}_s^*(\hat{v}_r^*(\tilde{v}_s^+)) \iff \hat{v}_s^{**}(\tilde{v}_s^+) := \tilde{v}_s^+ - \hat{v}_s^*(\hat{v}_r^*(\tilde{v}_s^+)) = 0. \quad (\text{F.136})$$

We now evaluate the ‘reflector’ function \hat{v}_s^{**} at two arguments v_s , namely v_s and \hat{v}_s^\uparrow . For $v_s = \hat{v}_s^\uparrow$, it holds that

$$\hat{v}_s^{**}(v_s) \stackrel{(\text{F.136})}{=} v_s - \hat{v}_s^*(\hat{v}_r^*(v_s)) \stackrel{(\text{F.134})}{=} v_s - \hat{v}_s^*(\alpha_{r0}) \stackrel{(\text{F.133})}{<} v_s - \alpha_{s0} \stackrel{(\text{F.134})}{<} 0 \quad (\text{F.137})$$

For $v_s = \hat{v}_s^\uparrow > v_s$, it holds that

$$\hat{v}_s^{**}(\hat{v}_s^\uparrow) \stackrel{(\text{F.136})}{=} \hat{v}_s^\uparrow - \hat{v}_s^*(\hat{v}_r^*(\hat{v}_s^\uparrow)) \stackrel{(\text{F.135})}{=} 0. \quad (\text{F.138})$$

Since \hat{v}_s^{**} is continuous, the intermediate-value theorem stipulates that a $\tilde{v}_s^+ \in (v_s, \hat{v}_s^\uparrow]$ exists that satisfies $\hat{v}_s^{**}(\tilde{v}_s^+) = 0$. Then, since $\tilde{v}_s^+ > v_s$, it follows from Eq. (F.134) that

$$\tilde{v}_r^+ = \hat{v}_r^*(\tilde{v}_s^+) < \alpha_{r0}. \quad (\text{F.139})$$

Since the unrestricted equilibrium valuations $(\hat{v}_r^+, \hat{v}_s^+)$ are unique, it must hold that $\hat{v}_r^+ = \tilde{v}_r^+$. However, the case condition $\hat{v}_r^+ \geq \alpha_{r0}$ conflicts with the derived condition $\tilde{v}_r^+ < \alpha_{r0}$ from Eq. (F.139). We thus arrive at a contradiction, which invalidates Eq. (F.133) and confirms Eq. (F.132).

F.10.3.2 Upper Bound on $\hat{v}_s^*(\alpha_{r0})$ for Case 2.(b).i

The following proof is similar in structure and goal as the proof in F.10.3.1, namely to prove $\hat{v}_s^*(\alpha_{r0}) \leq \alpha_{s0}$ by assuming

$$\hat{v}_s^*(\alpha_{r0}) > \alpha_{s0}. \quad (\text{F.140})$$

The arguments of the proof are visualized in Fig. F.2.

\hat{v}_r^* . In case 2.(b).i, the strict concavity of \hat{v}_r^* , together with the knowledge of $\hat{v}_r^*(\alpha_{s0}) \geq \alpha_{r0}$ (2.(b).i), imply:

$$\begin{aligned} \exists v_s^\leftarrow, v_s^\rightarrow. \quad \alpha_{s0} \in [v_s^\leftarrow, v_s^\rightarrow] \text{ and } \hat{v}_r^*(v_s^\leftarrow) = \alpha_{r0} \text{ and } \hat{v}_r^*(v_s^\rightarrow) = \alpha_{r0} \text{ and} \\ \forall v_s \in [v_s^\leftarrow, v_s^\rightarrow]. \hat{v}_r^*(v_s) \geq \alpha_{r0} \text{ and} \\ \forall v_s \notin [v_s^\leftarrow, v_s^\rightarrow]. \hat{v}_r^*(v_s) < \alpha_{r0}. \end{aligned} \quad (\text{F.141})$$

\hat{v}_s^* . Given Eq. (F.140), we find the maximum of \hat{v}_s^* :

$$\hat{v}_s^\uparrow = \max_{v_r} \hat{v}_s^*(v_r) \geq \hat{v}_s^*(\alpha_{r0}) \stackrel{(\text{F.140})}{>} \alpha_{s0}. \quad (\text{F.142})$$

Contradiction. Similar as in F.10.3.1, we show the existence of $(\tilde{v}_r^+, \tilde{v}_s^+) \neq (\hat{v}_r^+, \hat{v}_s^+)$, which satisfy the unrestricted-equilibrium properties $\tilde{v}_r^+ = \hat{v}_r^*(\tilde{v}_s^+)$ and $\tilde{v}_s^+ = \hat{v}_s^*(\tilde{v}_r^+)$, and thus contradict the uniqueness of the unrestricted equilibrium valuations $(\hat{v}_r^+, \hat{v}_s^+)$. To that end, we again introduce a reflector function \hat{v}_s^{**} with

$$\hat{v}_s^{**}(\tilde{v}_s^+) = \tilde{v}_s^+ - \hat{v}_s^*(\hat{v}_r^*(\tilde{v}_s^+)) = 0. \quad (\text{F.143})$$

However, unlike in F.10.3.1, we additionally have to consider the relative position of v_s^\rightarrow from Eq. (F.141) and $\hat{v}_s^*(\alpha_{r0})$.

- $\hat{v}_s^*(\alpha_{r0}) \geq v_s^\rightarrow$: For that case, we evaluate the reflector function \hat{v}_s^{**} at arguments v_s^\rightarrow from Eq. (F.141) and $\hat{v}_s^\uparrow \geq v_s^\rightarrow$:

$$\begin{aligned} \hat{v}_s^{**}(v_s^\rightarrow) &\stackrel{(\text{F.143})}{=} v_s^\rightarrow - \hat{v}_s^*(\hat{v}_r^*(v_s^\rightarrow)) \stackrel{(\text{F.141})}{=} v_s^\rightarrow - \hat{v}_s^*(\alpha_{r0}) \leq 0 \\ \hat{v}_s^{**}(\hat{v}_s^\uparrow) &\stackrel{(\text{F.143})}{=} \hat{v}_s^\uparrow - \hat{v}_s^*(\hat{v}_r^*(\hat{v}_s^\uparrow)) \stackrel{(\text{F.142})}{\geq} 0 \end{aligned} \quad (\text{F.144})$$

By the intermediate value theorem, there must thus exist a $\tilde{v}_s^+ \in [\nu_s^\rightarrow, \hat{v}_s^\uparrow]$ with $\hat{v}_s^{**}(\tilde{v}_s^+) = 0$. Since $\tilde{v}_s^+ \in [\nu_s^\rightarrow, \hat{v}_s^\uparrow] \subset [\nu_s^\leftarrow, \nu_s^\rightarrow]$, Eq. (F.141) implies that $\tilde{v}_s^+ \geq \nu_s^\rightarrow \geq \alpha_{s0}$, which conflicts with $\tilde{v}_s^+ < \alpha_{s0}$ from case condition 2.(b) and the uniqueness of the unrestricted equilibrium.

- $\hat{v}_s^*(\alpha_{r0}) < \nu_s^\rightarrow$: For that case, we evaluate the reflector function \hat{v}_s^{**} at arguments ν_s^\leftarrow from Eq. (F.141) and \hat{v}_s^\uparrow from Eq. (F.142):

$$\begin{aligned} \hat{v}_s^{**}(\nu_s^\leftarrow) &\stackrel{(F.143)}{=} \nu_s^\leftarrow - \hat{v}_s^*(\hat{v}_r^*(\nu_s^\leftarrow)) \stackrel{(F.141)}{=} \nu_s^\leftarrow - \hat{v}_s^*(\alpha_{r0}) \\ &\stackrel{(F.140)}{<} \nu_s^\leftarrow - \alpha_{s0} \stackrel{(F.141)}{\leq} 0 \\ \hat{v}_s^{**}(\hat{v}_s^\uparrow) &\stackrel{(F.143)}{=} \hat{v}_s^\uparrow - \hat{v}_s^*(\hat{v}_r^*(\hat{v}_s^\uparrow)) \stackrel{(F.142)}{\geq} 0 \end{aligned} \quad (F.145)$$

By the intermediate value theorem, there must thus exist a $\tilde{v}_s^+ \in [\nu_s^\leftarrow, \hat{v}_s^\uparrow]$ with $\hat{v}_s^{**}(\tilde{v}_s^+) = 0$. Since $\tilde{v}_s^+ \in [\nu_s^\leftarrow, \hat{v}_s^\uparrow] \subset [\nu_s^\leftarrow, \nu_s^\rightarrow]$, Eq. (F.141) implies that $\tilde{v}_r^+ = \hat{v}_r^*(\tilde{v}_s^+) \geq \alpha_{r0}$, which conflicts with $\tilde{v}_r^+ < \alpha_{r0}$ from case condition 2 and the uniqueness of the unrestricted equilibrium.

F.11 PROOF OF THEOREM 8.11: STABILITY OF HETEROGENEOUS EQUILIBRIUM

F.11.1 Proof Idea

To confirm the asymptotic stability of an equilibrium \mathbf{A}^+ , we demonstrate that the Jacobian matrix $\mathbf{J}(\mathbf{A})$ of the process in Eq. (8.35) is negative definite when evaluated at the equilibrium \mathbf{A}^+ . On a high level, the Jacobian matrix $\mathbf{J}(\mathbf{A})$ is defined as follows:

$$\forall(n, k), (n', k') \in N \times K. \quad J_{I(n,k), I(n',k')}(\mathbf{A}) = \frac{\partial}{\partial a_{n'k'}} (a_{nk}^*(\mathbf{A}_{-nk}) - a_{nk}) \quad (F.146)$$

where $I(n, k)$ is an index corresponding to attribute (n, k) . The derivatives of the restricted best-response a_{nk}^* in any attribute prevalence $a_{n'k'}$ are as follows:

$$\frac{\partial}{\partial a_{n'k'}} a_{nk}^*(\mathbf{A}) = \frac{\partial}{\partial a_{n'k'}} \max(0, \hat{a}_{nk}^*(\mathbf{A})) = \begin{cases} \frac{\partial}{\partial a_{n'k'}} \hat{a}_{nk}^*(\mathbf{A}) & \text{if } \hat{a}_{nk}^*(\mathbf{A}) \geq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (F.147)$$

To show this negative definiteness of $\mathbf{J}^+ = \mathbf{J}(\mathbf{A}^+)$, we demonstrate that every eigenvalue λ of \mathbf{J}^+ has a negative real part $\operatorname{Re}(\lambda)$. To find the eigenvalues λ of \mathbf{J}^+ ,

we solve the equation system $\mathbf{J}^+ \mathbf{x} = \lambda \mathbf{x}$ for $\lambda \in \mathbb{C}$, $\lambda \neq 0$, and $\mathbf{x} \in \mathbb{C}^{|N||K|}$, $\mathbf{x} \neq \mathbf{0}$. This equation system can be represented in the following form:

$$\forall (n, k) \in N \times K. \quad \left(J_{I(n,k), I(n,k)}^+ - \lambda \right) x_{I(n,k)} + \sum_{\substack{(n', k') \in N \times K. \\ (n', k') \neq (n, k)}} J_{I(n,k), I(n', k')}^+ x_{I(n', k')} = 0 \quad (\text{F.148})$$

F.11.2 Simplification of Eq. (F.148)

To concretize Eq. (F.148), we instantiate the Jacobian matrix \mathbf{J}^+ . As a_{nk}^* is independent of a_{nk} , the diagonal entries of \mathbf{J}^+ are:

$$\forall (n, k) \in N \times K. \quad J_{I(n,k), I(n,k)}^+ \stackrel{(\text{F.146})}{=} \frac{\partial}{\partial a_{n'k'}} (a_{nk}^* (\mathbf{A}_{-nk}) - a_{nk}) = -1, \quad (\text{F.149})$$

Now, we consider the entries not on the diagonal of \mathbf{J}^+ , i.e., $J_{I(n,k), I(n', k')}^+$ for all $(n, k) \neq (n', k')$. First, we specifically consider the rows of \mathbf{J}^+ associated with attributes $(n, k) \in L^+$, where L^+ is the set of attributes which must have zero prevalence $a_{nk}^+ = 0$ in equilibrium:

$$L^+ = N \times K \setminus (K_r^+ \cup K_{\bar{r}}^+). \quad (\text{F.150})$$

The following inequality holds on the equilibrium path valuation v_r^+ for each path r (cf. Theorems 8.8 and 8.10):

$$\begin{aligned} \forall (n, k) \in L^+. \quad v_{r(n)}^+ &\geq \psi_r \sqrt{d} \sqrt{1 + v_{\bar{r}(n)}^+} - (1 + v_{\bar{r}(n)}^+) \\ &> \sqrt{\frac{\alpha_{nk}}{\gamma_{nk}} d(\rho_n - \phi_{n0})} \sqrt{1 + v_{\bar{r}(n)}^+} - (1 + v_{\bar{r}(n)}^+) \end{aligned} \quad (\text{F.151})$$

Then, remember the following equivalence from Theorem 8.8 for any attribute (n, k) :

$$v_{r(n)}^+ > \sqrt{\frac{\alpha_{nk}}{\gamma_{nk}} d(\rho_n - \phi_{n0})} \sqrt{1 + v_{\bar{r}(n)}^+} - (1 + v_{\bar{r}(n)}^+) \iff \frac{\partial \pi_n(\mathbf{A}^+)}{\partial a_{nk}} < 0. \quad (\text{F.152})$$

Together with the concavity of π_n , we thus note that the attribute value a_{nk} needs to be decreased to optimize the profit π_n in a_{nk} . Given $a_{nk}^+ = 0$, we note that the unrestricted best-response attribute \hat{a}_{nk}^* for $(n, k) \in L^+$ is thus negative in the equilibrium:

$$\forall (n, k) \in L^+. \quad \hat{a}^*(\mathbf{A}_{-nk}^+) < 0 \quad (\text{F.153})$$

Given the definition of the Jacobian entries in Eq. (F.146) and the derivative of the restricted best response a_{nk}^* in Eq. (F.147), we note that:

$$\forall (n, k) \in L^+. \forall (n', k') \neq (n, k). J_{I(n,k), I(n',k')}^+ = 0. \quad (\text{F.154})$$

The eigenvalue equation system in Eq. (F.148) can thus be written as:

$$\forall (n, k) \in K_r^+ \cup K_{\bar{r}}^+. -(\lambda + 1)x_{I(n,k)} \quad (\text{F.155})$$

$$+ \sum_{\substack{(n', k') \in N \times K, \\ (n', k') \neq (n, k)}} J_{I(n,k), I(n',k')}^+ x_{I(n',k')} = 0 \quad (\text{F.156})$$

$$\forall (n, k) \in L^+. -(\lambda + 1)x_{I(n,k)} = 0 \quad (\text{F.157})$$

Interestingly, the equation system in Eqs. (F.156) and (F.157) can be considerably simplified in our proof, which can be shown by a case distinction on $L^+ = (N \times K) \setminus (K_r^+ \cup K_{\bar{r}}^+)$, i.e., the set of attributes that certainly have zero prevalence in the equilibrium.

- $L^+ = \emptyset$: In this case, the equation system can be simplified in two respects. First, we note that $L^+ = \emptyset$ implies that no equations in the form of Eq. (F.157) exist in the equation system. Second, we note that we only investigate networks with a unique equilibrium, i.e., non-zero equilibrium prevalence is possible for only one attribute on each path ($|K_r^+| = |K_{\bar{r}}^+| = 1$). Hence, if $L^+ = \emptyset$, we know that $K_r^+ \cup K_{\bar{r}}^+$ covers both of the two attributes of the network, one on each path:

$$K_r^+ \cup K_{\bar{r}}^+ = N \times K = \{(n(r), k(r)), (n(\bar{r}), k(\bar{r}))\} \quad (\text{F.158})$$

where $(n(r), k(r))$ is the single attribute with possibly non-zero equilibrium prevalence on path r . These insights allow to reduce the equation system in Eq. (F.156) to only two equations (No equations like Eq. (F.157) exist):

$$-(\lambda + 1)x_r + J_{\bar{r}}^+ x_{\bar{r}} = 0 \quad -(\lambda + 1)x_{\bar{r}} + J_r^+ x_r = 0 \quad (\text{F.159})$$

where we have abbreviated:

$$J_r^+ = J_{I(n(r), k(r)), I(n(\bar{r}), k(\bar{r}))}^+ \quad x_r = x_{I(n(r), k(r))} \quad (\text{F.160})$$

Since the eigenvector \mathbf{x} in the current case only has the two entries x_r and $x_{\bar{r}}$, we require $\mathbf{x} = (x_r, x_{\bar{r}})^\top \neq \mathbf{0}$. We find that $\lambda = -1$ is an eigenvalue of the system in Eq. (F.159) if and only if $J_r^+ = 0$ or $J_{\bar{r}}^+ = 0$, i.e., at least one of the two relevant Jacobian entries is zero. Since $\lambda = -1$ would preserve negative definiteness of \mathbf{J}^+ , we do not need to consider this case further.

- $L^+ \neq \emptyset$: If $L^+ \neq \emptyset$, the equation system contains equations in the form of Eq. (F.157). Then, $\lambda = -1$ may be an eigenvalue of J^+ , which would preserve negative definiteness of J^+ ; hence, this case is not further considered. Conversely, if $\lambda = -1$ is not a solution of the system, the equations in the form of Eq. (F.157) imply that $x_{I(n,k)} = 0$ for all $(n, k) \in L^+$. This insight then again allows the simplification to the equation system in Eq. (F.159). Crucially, since $x_{I(n,k)} = 0$ for all $(n, k) \in L^+$, it must hold that $(x_r, x_{\bar{r}})^\top \neq 0$ such that $x \neq 0$, i.e., such that x is a valid eigenvector.

F.11.3 Solution of Eq. (F.159)

In summary, we only need to consider the equation system in Eq. (F.159) and the case $\lambda \neq -1$. Furthermore, we require $(x_r, x_{\bar{r}})^\top \neq 0$. Without loss of generality, let r be the path with $x_r \neq 0$. Then, we can perform the following transformation:

$$-(\lambda + 1)x_r + J_{\bar{r}}^+ x_{\bar{r}} = 0 \implies \lambda + 1 = J_{\bar{r}}^+ \frac{x_{\bar{r}}}{x_r} \quad (\text{F.161})$$

$$-(\lambda + 1)x_{\bar{r}} + J_r^+ x_r = 0 \implies x_{\bar{r}} = \frac{J_r}{\lambda + 1} x_r \quad (\text{F.162})$$

Inserting Eq. (F.162) into Eq. (F.161) yields a quadratic equation in λ :

$$(\lambda + 1)^2 - J_r^+ J_{\bar{r}}^+ = 0 \implies \lambda_{1,2} = -1 \pm \sqrt{J_r^+ J_{\bar{r}}^+} \quad (\text{F.163})$$

If $J_r^+ J_{\bar{r}}^+ = 0$, then $\lambda_{1,2} = -1$ produces a contradiction to the assumption $\lambda \neq -1$, which implies that no eigenvalue $\lambda \neq -1$ exists.

If $J_r^+ J_{\bar{r}}^+ \neq 0$, then we know that

$$\operatorname{Re}(\lambda_{1,2}) < 0 \iff \operatorname{Re}\left(\sqrt{J_r^+ J_{\bar{r}}^+}\right) < 1 \iff J_r^+ J_{\bar{r}}^+ < 1. \quad (\text{F.164})$$

F.11.4 Bounding of $\lambda_{1,2}$

To verify that the condition in Eq. (F.164) always holds, we first find J_r^+ for any path r :

$$J_r^+ = \begin{cases} \frac{\alpha_{n(\bar{r})k(\bar{r})}}{\alpha_{n(r)k(r)}} \left(\frac{\psi_r \sqrt{d}}{2\sqrt{1+v_r^+}} - 1 \right) & \text{if } \hat{a}_{n(r)k(r)}^* \left(\mathbf{A}_{-n(r)k(r)}^+ \right) \geq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (\text{F.165})$$

Given $J_r^+ J_{\bar{r}}^+ \neq 0$, stability requires:

$$J_r^+ J_{\bar{r}}^+ \stackrel{(F.165)}{=} \left(\frac{\psi_r \sqrt{d}}{2\sqrt{1+v_{\bar{r}}^+}} - 1 \right) \left(\frac{\psi_{\bar{r}} \sqrt{d}}{2\sqrt{1+v_r^+}} - 1 \right) \stackrel{(F.164)}{<} 1 \quad (F.166)$$

$$\iff \psi_r \psi_{\bar{r}} d - 2\sqrt{d} \left(\psi_r \sqrt{1+v_{\bar{r}}^+} + \psi_{\bar{r}} \sqrt{1+v_r^+} \right) < 0 \quad (F.167)$$

Moreover, $J_r^+ J_{\bar{r}}^+ \neq 0$ implies that the restricted equilibrium valuation v_r^+ for each path r corresponds to the unrestricted equilibrium valuation \hat{v}_r^+ from Theorem 8.10:

$$\begin{aligned} \forall r \in R. J_r^+ J_{\bar{r}}^+ \neq 0 &\implies \forall r \in R. \hat{a}_{n(r)k(r)}^*(\mathbf{A}_{-n(r)k(r)}^+) \geq 0 \\ \implies \forall r \in R. \hat{v}_r^+ &\geq \alpha_{r0} \implies \forall r \in R. v_r^+ = \hat{v}_r^+. \end{aligned} \quad (F.168)$$

Hence, we can expand (symmetrically for $\psi_{\bar{r}} \sqrt{1+v_{\bar{r}}^+}$):

$$\begin{aligned} &\psi_r \sqrt{1+v_r^+} \\ \stackrel{(F.168)}{=} &\psi_r \sqrt{1+\hat{v}_r^+} \\ \stackrel{Th8.10}{=} &\psi_r \sqrt{\frac{\psi_r^3 \psi_{\bar{r}}}{(\psi_r^2 + \psi_{\bar{r}}^2)^2} \left(\sqrt{d (\psi_r^2 + \psi_{\bar{r}}^2) + \frac{1}{4} \psi_r^2 \psi_{\bar{r}}^2 d^2} + \frac{d}{2} \psi_r \psi_{\bar{r}} \right) + \frac{\psi_r^2}{\psi_r^2 + \psi_{\bar{r}}^2}} \quad (F.169) \\ = &\psi_r^2 \sqrt{\frac{\psi_r \psi_{\bar{r}}}{(\psi_r^2 + \psi_{\bar{r}}^2)^2} \left(\sqrt{d (\psi_r^2 + \psi_{\bar{r}}^2) + \frac{1}{4} \psi_r^2 \psi_{\bar{r}}^2 d^2} + \frac{d}{2} \psi_r \psi_{\bar{r}} \right) + \frac{1}{\psi_r^2 + \psi_{\bar{r}}^2}} \end{aligned}$$

We use this equality to rewrite the inequality in Eq. (F.167):

$$\begin{aligned}
 & \psi_r \psi_{\bar{r}} d - 2\sqrt{d} \left(\psi_r \sqrt{1 + v_{\bar{r}}^+} + \psi_{\bar{r}} \sqrt{1 + v_r^+} \right) \\
 \stackrel{(F.169)}{=} & \psi_r \psi_{\bar{r}} d - 2\sqrt{d} \left(\psi_r^2 + \psi_{\bar{r}}^2 \right) \cdot \\
 & \sqrt{\frac{\psi_r \psi_{\bar{r}}}{(\psi_r^2 + \psi_{\bar{r}}^2)^2} \left(\sqrt{d (\psi_r^2 + \psi_{\bar{r}}^2) + \frac{1}{4} \psi_r^2 \psi_{\bar{r}}^2 d^2} + \frac{d}{2} \psi_r \psi_{\bar{r}} \right) + \frac{1}{\psi_r^2 + \psi_{\bar{r}}^2}} \quad (F.170) \\
 = & \psi_r \psi_{\bar{r}} d - \sqrt{\left(\frac{\psi_r \psi_{\bar{r}}}{(\psi_r^2 + \psi_{\bar{r}}^2)^2} \sqrt{d (\psi_r^2 + \psi_{\bar{r}}^2) + \frac{1}{4} \psi_r^2 \psi_{\bar{r}}^2 d^2} + \frac{1}{\psi_r^2 + \psi_{\bar{r}}^2} \right)} \\
 < & (1 - \sqrt{2}) d \psi_r \psi_{\bar{r}} < 0
 \end{aligned}$$

The last upper bound holds because we can exclude $\psi_r = 0$ for any path r in the current case $v_r^+ = \hat{v}_r^+$, as $\psi_r = 0$ produces the contradiction $v_r^+ < v_r^+$:

$$\begin{aligned}
 v_r^+ & \stackrel{(F.168)}{=} \hat{v}_r^+ \stackrel{(F.112)}{=} \psi_r \sqrt{d} \sqrt{1 + \hat{v}_{\bar{r}}^+} - (1 + \hat{v}_{\bar{r}}^+) \\
 & \stackrel{\psi_r=0}{=} -(1 + \hat{v}_{\bar{r}}^+) \stackrel{(F.168)}{=} -(1 + v_{\bar{r}}^+) \\
 & \stackrel{\text{Th8.10}}{\leq} -(1 + \alpha_{\bar{r}0}) \stackrel{\alpha_{\bar{r}0} \geq 0}{<} 0 \stackrel{\alpha_{r0} \geq 0}{\leq} \alpha_{r0} \stackrel{\text{Th8.10}}{\leq} v_r^+.
 \end{aligned} \quad (F.171)$$

In summary, we now have shown that $J_r^+ J_{\bar{r}}^+ < 1$, which ensures a negative real part $\text{Re}(\lambda_{1,2}) < 0$ (Eq. (F.164)) of the eigenvalues $\lambda_{1,2}$ from Eq. (F.163), and thus confirms that J^+ is negative definite. Since J^+ is negative definite, the equilibrium \mathbf{A}^+ from Theorem 8.10 is asymptotically stable with respect to the process in Eq. (8.35), which concludes the proof.

F.12 PROOF OF THEOREM 8.12: HETEROGENEOUS COMPETITION (IMPROVEMENT)

For the competitive network \mathcal{N}_4 , the unrestricted equilibrium network valuation \hat{V}^+ is given by

$$\hat{V}^+(\mathcal{N}_4) = \hat{v}_r^+ + \hat{v}_{\bar{r}}^+ = \frac{\psi_r \psi_{\bar{r}}}{\psi_r^2 + \psi_{\bar{r}}^2} \left(\sqrt{d (\psi_r^2 + \psi_{\bar{r}}^2) + \frac{1}{4} \psi_r^2 \psi_{\bar{r}}^2 d^2} + \frac{d}{2} \psi_r \psi_{\bar{r}} \right) - 1, \quad (F.172)$$

where \hat{v}_r^+ and $\hat{v}_{\bar{r}}^+$ are as in Theorem 8.10. The restricted equilibrium network valuation $V^+(\mathcal{N}_4)$ is equal to $\hat{V}^+(\mathcal{N}_4)$ if $\hat{v}_r^+ \geq \alpha_{r0}$ and $\hat{v}_{\bar{r}}^+ \geq \alpha_{\bar{r}0}$. These unrestricted equilibrium path valuations are monotonically increasing in the demand limit d (cf. Theorem 8.10). Hence, if d is high enough, it holds that $V^+(\mathcal{N}_4) = \hat{V}^+(\mathcal{N}_4)$.

For the competition-free network \mathcal{N}_3 , the unrestricted equilibrium network valuation \hat{V}^+ is

$$\hat{V}^+(\mathcal{N}_3) = \psi_r \sqrt{d_r} - 1 + \psi_{\bar{r}} \sqrt{d_{\bar{r}}} - 1. \quad (\text{F.173})$$

Among all demand distributions $(d_r, d_{\bar{r}})$ with $d_r + d_{\bar{r}} = d$, the demand distribution maximizing $\hat{V}^+(\mathcal{N}_3)$ can be found as follows:

$$\frac{\partial}{\partial d_r} \left(\psi_r \sqrt{d_r} + \psi_{\bar{r}} \sqrt{d - d_r} - 2 \right) = 0 \iff d_r = \frac{\psi_r^2}{\psi_r^2 + \psi_{\bar{r}}^2} d, \quad (\text{F.174})$$

where the maximum character of this value is ensured by a consistently non-positive second derivative of $\hat{V}^+(\mathcal{N}_3)$ in d_r . In the following, we thus consider only the maximum unrestricted equilibrium network valuation $\hat{V}^+(\mathcal{N}_3)$:

$$\hat{V}^+(\mathcal{N}_3) = \psi_r \sqrt{\frac{\psi_r^2}{\psi_r^2 + \psi_{\bar{r}}^2} d} + \psi_{\bar{r}} \sqrt{\frac{\psi_{\bar{r}}^2}{\psi_r^2 + \psi_{\bar{r}}^2} d} - 2 = \sqrt{d(\psi_r^2 + \psi_{\bar{r}}^2)} - 2. \quad (\text{F.175})$$

Again, for $V^+(\mathcal{N}_3) = \hat{V}^+(\mathcal{N}_3)$, d must be high enough such that

$$\psi_r \sqrt{\frac{\psi_r^2}{\psi_r^2 + \psi_{\bar{r}}^2} d} - 1 \geq \alpha_{r0} \quad \text{and} \quad \psi_{\bar{r}} \sqrt{\frac{\psi_{\bar{r}}^2}{\psi_r^2 + \psi_{\bar{r}}^2} d} - 1 \geq \alpha_{\bar{r}0}. \quad (\text{F.176})$$

If d is high enough, the difference of the equilibrium network valuations is thus:

$$\begin{aligned} \Delta V^+ &= V^+(\mathcal{N}_4) - V^+(\mathcal{N}_3) \\ &= \frac{\psi_r \psi_{\bar{r}}}{\psi_r^2 + \psi_{\bar{r}}^2} \left(\sqrt{d(\psi_r^2 + \psi_{\bar{r}}^2)} + \frac{1}{4} \psi_r^2 \psi_{\bar{r}}^2 d^2 + \frac{d}{2} \psi_r \psi_{\bar{r}} \right) \\ &\quad - \sqrt{d(\psi_r^2 + \psi_{\bar{r}}^2)} + 1 \end{aligned} \quad (\text{F.177})$$

Clearly, this difference is eventually positive when increasing the demand limit d , meaning that $V^+(\mathcal{N}_4)$ exceeds $V^+(\mathcal{N}_3)$ for high enough d :

$$\lim_{d \rightarrow \infty} \Delta V^+ = \infty. \quad (\text{F.178})$$

This last insight proves the theorem.

F.13 PROOF OF THEOREM 8.13: HETEROGENEOUS COMPETITION (DEGRADATION)

The following proof is constructive, i.e., we demonstrate how to choose $(\psi_r, \psi_{\bar{r}})$ and $(\alpha_{r0}, \alpha_{\bar{r}0})$ such that $V^+(\mathcal{N}_4) < V^+(\mathcal{N}_3)$ holds given demand distribution $(d_r, d_{\bar{r}})$. In this construction, the goal is to create a scenario where the competitive network \mathcal{N}_4 will be at minimum valuation $\alpha_{r0} + \alpha_{\bar{r}0}$, but the competition-free network has a path r with an equilibrium valuation v_r^+ exceeding the minimum path valuation α_{r0} .

Regarding the path-characteristic ratios $(\psi_r, \psi_{\bar{r}})$, our first step consists of choosing the ratios such that the competitive network is at minimum valuation, i.e., such that $v_r^+ = \alpha_{r0}$ and $v_{\bar{r}}^+ = \alpha_{\bar{r}0}$. To do so, we first determine $\psi_{\bar{r}}$ such that $\hat{v}_{\bar{r}}^+ \leq \alpha_{\bar{r}0}$ for all ψ_r , which is achieved by $\psi_r = 0$:

$$\lim_{\psi_{\bar{r}} \rightarrow 0} \hat{v}_{\bar{r}}^+ \stackrel{\text{Th8.10}}{=} -1 \stackrel{\alpha_{\bar{r}0} \geq 0}{<} \alpha_{\bar{r}0}. \quad (\text{F.179})$$

Having selected $\psi_{\bar{r}}$ such that $\hat{v}_{\bar{r}}^+ \leq \alpha_{\bar{r}0}$, it holds that $v_{\bar{r}}^+ = \alpha_{\bar{r}0}$ by Theorem 8.10. As a result, the equilibrium path valuation for path r in the competitive network \mathcal{N}_4 is:

$$v_r^+(\mathcal{N}_4) \stackrel{\text{Th8.10}}{=} \max \left(\alpha_{r0}, \psi_r \sqrt{d_r} \sqrt{1 + \alpha_{\bar{r}0}} - (1 + \alpha_{\bar{r}0}) \right). \quad (\text{F.180})$$

To ensure that v_r^+ is minimal (i.e., equals α_{r0}), the following condition must hold:

$$\alpha_{r0} \geq \psi_r \sqrt{d_r} \sqrt{1 + \alpha_{\bar{r}0}} - (1 + \alpha_{\bar{r}0}) \iff \psi_r \leq \frac{1 + \alpha_{r0} + \alpha_{\bar{r}0}}{\sqrt{d_r} \sqrt{1 + \alpha_{\bar{r}0}}} \quad (\text{F.181})$$

If ψ_r is chosen according to Eq. (F.181), the equilibrium network valuation in the competition-free network is minimal, i.e., $V^+(\mathcal{N}_4) = \alpha_{r0} + \alpha_{\bar{r}0}$.

To let $V^+(\mathcal{N}_3)$ of the competition-free network exceed $V^+(\mathcal{N}_4)$ of the competitive network, we further need to choose ψ_r such that $\hat{v}_r^+(\mathcal{N}_3) > \alpha_{r0}$. This condition can be transformed in the following fashion:

$$\psi_r \sqrt{d_r} - 1 > \alpha_{r0} \iff \psi_r > \frac{1 + \alpha_{r0}}{\sqrt{d_r}} \quad (\text{F.182})$$

To allow a selection of ψ_r that achieves $v_r^+(\mathcal{N}_4) = \alpha_{r0}$ but $v_r^+(\mathcal{N}_3) > \alpha_{r0}$, it must thus hold that

$$\frac{1 + \alpha_{r0}}{\sqrt{d_r}} < \frac{1 + \alpha_{r0} + \alpha_{\bar{r}0}}{\sqrt{d_r} \sqrt{1 + \alpha_{\bar{r}0}}} \iff \alpha_{r0} < \frac{\sqrt{d_r} \alpha_{\bar{r}0}}{\sqrt{d_r} \sqrt{1 + \alpha_{\bar{r}0}} - \sqrt{d_r}} - 1, \quad (\text{F.183})$$

This condition always holds when choosing $\alpha_{r0} = 0$ and $\alpha_{\bar{r}0} > d_{\bar{r}}/d_r$:

$$\begin{aligned}
 & \frac{\sqrt{d_r}\alpha_{\bar{r}0}}{\sqrt{d}\sqrt{1+\alpha_{\bar{r}0}} - \sqrt{d_r}} - 1 \stackrel{\alpha_{\bar{r}0} > d_{\bar{r}}/d_r}{>} \frac{\sqrt{d_r}\frac{d_{\bar{r}}}{d_r}}{\sqrt{d}\sqrt{1+\frac{d_{\bar{r}}}{d_r}} - \sqrt{d_r}} - 1 \\
 & \stackrel{d_r+d_{\bar{r}}=d}{=} \frac{\sqrt{d_r}\frac{d_{\bar{r}}}{d_r}}{\sqrt{d}\sqrt{\frac{d}{d_r}} - \sqrt{d_r}} - 1 \stackrel{\frac{\sqrt{d_r}}{d_r}=\frac{1}{\sqrt{d_r}}}{=} \frac{\frac{d_{\bar{r}}}{\sqrt{d_r}}}{\frac{d}{\sqrt{d_r}} - \frac{d_r}{\sqrt{d_r}}} - 1 \\
 & \stackrel{d_r+d_{\bar{r}}=d}{=} \frac{\frac{d_{\bar{r}}}{\sqrt{d_r}}}{\frac{d_{\bar{r}}}{\sqrt{d_r}}} - 1 = 0 \stackrel{\alpha_{r0}=0}{=} \alpha_{r0}
 \end{aligned} \tag{F.184}$$

Hence, $(\psi_r, \psi_{\bar{r}})$ and $(\alpha_{r0}, \alpha_{\bar{r}0})$ can be chosen such that $V^+(\mathcal{N}_3) > V^+(\mathcal{N}_4)$, which concludes the proof.

DOCTORAL RESEARCH (COVERED)

- [1] Simon Scherrer, Adrian Perrig, and Stefan Schmid. "The Value of Information in Selfish Routing." In: *Proceedings of the International Colloquium on Structural Information and Communication Complexity (SIROCCO)*. July 2020.
- [2] Simon Scherrer, Markus Legner, Adrian Perrig, and Stefan Schmid. "Incentivizing Stable Path Selection in Future Internet Architectures." In: *Proceedings of the International Symposium on Computer Performance, Modeling, Measurements and Evaluation (PERFORMANCE)*. Aug. 2020.
- [3] Simon Scherrer, Markus Legner, Adrian Perrig, and Stefan Schmid. "Enabling Novel Interconnection Agreements with Path-Aware Networking Architectures." In: *Proceedings of the International Conference on Dependable Systems and Networks (DSN)*. June 2021.
- [4] Simon Scherrer, Markus Legner, Adrian Perrig, and Stefan Schmid. "An Axiomatic Perspective on the Performance Effects of End-Host Path Selection." In: *Proceedings of the International Symposium on Computer Performance, Modeling, Measurements and Evaluation (PERFORMANCE)*. Aug. 2021.
- [5] Simon Scherrer, Markus Legner, Adrian Perrig, and Stefan Schmid. "Model-Based Insights on the Performance, Fairness, and Stability of BBR." In: *Proceedings of the Internet Measurement Conference (IMC)*. Nov. 2022.
- [6] Simon Scherrer, Seyedali Tabaeiaghdaei, and Adrian Perrig. "Quality Competition Among Internet Service Providers." In: *Proceedings of the International Symposium on Computer Performance, Modeling, Measurements and Evaluation (PERFORMANCE)*. Nov. 2023.
- [7] Simon Scherrer, Adrian Perrig, and Stefan Schmid. "A Dynamic Model-Based Perspective on BBR Fairness." In: *Under submission*. 2023.

DOCTORAL RESEARCH (NOT COVERED)

- [8] Arish Sateesan, Jo Vliegen, Simon Scherrer, Hsu-Chun Hsiao, Adrian Perrig, and Nele Mentens. "Speed Records in Network Flow Measurement on FPGA." In: *Proceedings of the International Conference on Field-Programmable Logic (FPL)*. Aug. 2021.
- [9] Simon Scherrer, Che-Yu Wu, Yu-Hsi Chiang, Benjamin Rothenberger, Daniele Asoni, Arish Sateesan, Jo Vliegen, Nele Mentens, Hsu-Chun Hsiao, and Adrian Perrig. "Low-Rate Overuse Flow Tracer (LOFT): An Efficient and Scalable Algorithm for Detecting Overuse Flows." In: *Proceedings of the Symposium on Reliable Distributed Systems (SRDS)*. Sept. 2021.
- [10] Francesco Da Dalt, Simon Scherrer, and Adrian Perrig. "Bayesian Sketching for Volume Estimation in Data Streams." In: *Proceedings of the International Conference on Very Large Databases (VLDB)*. Aug. 2023.
- [11] Seyedali Tabaeiaghdaei, Simon Scherrer, Jonghoon Kwon, and Adrian Perrig. "Carbon-Aware Global Routing in Path-Aware Networks." In: *Proceedings of ACM International Conference on Future Energy Systems (e-Energy)*. June 2023.
- [12] Simon Scherrer, Jo Vliegen, Arish Sateesan, Hsu-Chun Hsiao, Nele Mentens, and Adrian Perrig. "ALBUS: a Probabilistic Monitoring Algorithm to Counter Burst Flood Attacks." In: *Proceedings of the International Symposium on Reliable Distributed Systems (SRDS)*. Sept. 2023.
- [13] Cyrill Krähenbühl, Seyedali Tabaeiaghdaei, Simon Scherrer, and Adrian Perrig. "GLIDS: Toward Global Latency Transparency." In: *Unpublished*. Sept. 2023.

REFERENCES

- [14] Daron Acemoglu, Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar. "Informational Braess' paradox: The effect of information on traffic congestion." In: *Operations Research* 66 (2018).
- [15] Vamsi Addanki, Oliver Michel, and Stefan Schmid. "PowerTCP: Pushing the Performance Limits of Datacenter Networks." In: *Proceedings of the USENIX Symposium on Networked Systems Design and Implementation (NSDI)*. 2022.
- [16] Aditya Akella, Srinivasan Seshan, Richard Karp, Scott Shenker, and Christos Papadimitriou. "Selfish behavior and stability of the Internet: a game-theoretic analysis of TCP." In: *ACM SIGCOMM Computer Communication Review (CCR)*. 2002.
- [17] Mohammad Alizadeh, Albert Greenberg, David A Maltz, Jitendra Padhye, Parveen Patel, Balaji Prabhakar, Sudipta Sengupta, and Murari Sridharan. "Data center TCP (DCTCP)." In.
- [18] Mohammad Alizadeh, Shuang Yang, Milad Sharif, Sachin Katti, Nick McKeown, Balaji Prabhakar, and Scott Shenker. "pfabric: Minimal near-optimal datacenter transport." In: *ACM SIGCOMM Computer Communication Review (CCR)* 43 (2013).
- [19] Sebastian Allmeier and Nicolas Gast. "Mean Field and Refined Mean Field Approximations for Heterogeneous Systems: It Works!" In: *Proceedings of the ACM SIGMETRICS Conference* (2022).
- [20] Eitan Altman, KE Avrachenkov, and BJ Prabhu. "Fairness in MIMD congestion control algorithms." In: *Telecommunication Systems* (2005).
- [21] Reid Andersen, Christian Borgs, Jennifer Chayes, Uriel Feige, Abraham Flaxman, Adam Kalai, Vahab Mirrokni, and Moshe Tennenholtz. "Trust-based recommendation systems: an axiomatic approach." In: *Proceedings of the International Conference on World Wide Web (WWW)*. 2008.
- [22] Tom Anderson et al. "The NEBULA Future Internet Architecture." In: *The Future Internet*. 2013. DOI: [10.4236/ghnnfx](https://doi.org/10.4236/ghnnfx).

- [23] Elliot Anshelevich and Shreyas Sekar. "Price competition in networked markets: How do monopolies impact social welfare?" In: *International Conference on Web and Internet Economics (WINE)*. 2015.
- [24] Ruwaifa Anwar, Haseeb Niaz, David Choffnes, Ítalo Cunha, Phillipa Gill, and Ethan Katz-Bassett. "Investigating interdomain routing policies in the wild." In: *Proceedings of the ACM Internet Measurement Conference (IMC)*. 2015.
- [25] Maria Apostolaki, Laurent Vanbever, and Manya Ghobadi. "FAB: Toward flow-aware buffer sharing on programmable switches." In: *Proceedings of the Workshop on Buffer Sizing*. 2019.
- [26] Venkat Arun, Mohammad Alizadeh, and Hari Balakrishnan. "Starvation in end-to-end congestion control." In: *Proceedings of the ACM SIGCOMM Conference*. 2022.
- [27] Venkat Arun, Mina Arashloo, Ahmed Saeed, Mohammad Alizadeh, and Hari Balakrishnan. "Formally Verifying Congestion Control Performance." In: *Proceedings of SIGCOMM 2021*. 2021.
- [28] Venkat Arun, Mina Tahmasbi Arashloo, Ahmed Saeed, Mohammad Alizadeh, and Hari Balakrishnan. "Toward formally verifying congestion control behavior." In: *Proceedings of the ACM SIGCOMM Conference*. 2021.
- [29] Venkat Arun and Hari Balakrishnan. "Copa: Practical Delay-Based congestion control for the internet." In: *Proceedings of the USENIX Symposium on Networked Systems Design and Implementation (NSDI)*. 2018.
- [30] Wei Bao, Vincent WS Wong, and Victor CM Leung. "A model for steady state throughput of TCP CUBIC." In: *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM)*. 2010.
- [31] David Besanko, Sachin Gupta, and Dipak Jain. "Logit Demand Estimation under Competitive Pricing Behavior: An Equilibrium Framework." In: *Management Science* 44 (1998).
- [32] Sanjay P Bhat and Dennis S Bernstein. "Finite-time stability of continuous autonomous systems." In: *SIAM Journal on Control and Optimization* 38 (2000).
- [33] Bobby Bhattacharjee, Ken Calvert, Jim Griffioen, Neil Spring, and James P. G. Sterbenz. "Postmodern internetwork architecture." In: *NSF Nets FIND Initiative* (2006).
- [34] Kostas Bimpikis, Shayan Ehsani, and Rahmi Ilkılıç. "Cournot competition in networked markets." In: *Management Science* 65 (2019).

- [35] Ken Binmore, Ariel Rubinstein, and Asher Wolinsky. "The Nash bargaining solution in economic modelling." In: *The RAND Journal of Economics* (1986).
- [36] Douglas H Blair and Robert A Pollak. "Acyclic collective choice rules." In: *Econometrica: Journal of the Econometric Society* (1982).
- [37] Thomas Bonald. "Comparison of TCP Reno and TCP Vegas via fluid approximation." PhD thesis. INRIA, 1998.
- [38] Dietrich Braess. "Über ein Paradoxon aus der Verkehrsplanung." In: *Unternehmensforschung* 12 (1968).
- [39] Lawrence S Brakmo, Sean W O'Malley, and Larry L Peterson. "TCP Vegas: New techniques for congestion detection and avoidance." In: *Proceedings of the ACM SIGCOMM Conference*. 1994.
- [40] Rodica Branzei, Dinko Dimitrov, and Stef Tijs. *Models in cooperative game theory*. Springer Science & Business Media, 2008.
- [41] Bob Briscoe. "Flow rate fairness: Dismantling a religion." In: *ACM SIGCOMM Computer Communication Review (CCR)* (2007).
- [42] LEJ Brouwer. "Über Abbildungen von Mannigfaltigkeiten." In: *Mathematische Annalen* (1911).
- [43] Lloyd Brown, Ganesh Ananthanarayanan, Ethan Katz-Bassett, Arvind Krishnamurthy, Sylvia Ratnasamy, Michael Schapira, and Scott Shenker. "On the Future of Congestion Control for the Public Internet." In: *Proceedings of the ACM Workshop on Hot Topics in Networks (HotNets)*. 2020.
- [44] Antonio Cabrales and Roberto Serrano. "Implementation in adaptive better-response dynamics: Towards a general theory of bounded rationality in mechanisms." In: *Games and Economic Behavior* 73 (2011).
- [45] Kenneth L Calvert, James Griffioen, Anna Nagurney, and Tilman Wolf. "A Vision for a Spot Market for Interdomain Connectivity." In: *Proceedings of the International Conference on Distributed Computing Systems (ICDCS)*. 2019.
- [46] Sue Ann Campbell, Jacques Belair, Toru Ohira, and John Milton. "Complex dynamics and multistability in a damped harmonic oscillator with delayed negative feedback." In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* (1995).
- [47] Neal Cardwell, Yuchung Cheng, C Stephen Gunn, Soheil Hassas Yeganeh, and Van Jacobson. "BBR: Congestion-based congestion control." In: *Queue* 14 (2016).

- [48] Neal Cardwell, Yuchung Cheng, C Stephen Gunn, Soheil Hassas Yeganeh, Ian Swett, Jana Iyengar, Victor Vasiliev, and Van Jacobson. "BBR Congestion Control: IETF 99 Update." In: *Presentation in ICCRG at IETF 99*. 2017. URL: <https://datatracker.ietf.org/meeting/99/materials/slides-99-iccrг-iccrг-presentation-2-00.pdf>.
- [49] Neal Cardwell, Yuchung Cheng, C. Stephen Gunn, Soheil Hassas Yeganeh, and Van Jacobson. "BBR: Congestion-Based Congestion Control." In: *Communications of the ACM* (2017).
- [50] Neal Cardwell, Yuchung Cheng, S Hassas Yeganeh, Ian Swett, Victor Vasiliev, Priyanjan Jha, Yousuk Seung, Matt Mathis, and Van Jacobson. "BBRv2: A model-based congestion control." In: *Presentation in ICCRG at IETF 104th meeting*. 2019.
- [51] Neal Cardwell, Stefan Savage, and Thomas Anderson. "Modeling TCP latency." In: *Proceedings of IEEE INFOCOM 2000*. Vol. 3. 2000.
- [52] Neal Cardwell et al. "BBRv3: Algorithm Bug Fixes and Public Internet Deployment." In: *Presentation in CCWG at IETF 117*. 2023. URL: <https://datatracker.ietf.org/meeting/117/materials/slides-117-ccwg-bbrv3-algorithm-bug-fixes-and-public-internet-deployment>.
- [53] Center for Applied Internet Data Analysis. *The CAIDA AS Relationships Dataset with Geographic Annotations*. <https://data.caida.org/datasets/as-relationships-geo/>. 2016.
- [54] Center for Applied Internet Data Analysis. *The CAIDA AS Relationships Dataset*. <http://data.caida.org/datasets/as-relationships/serial-2/20200401.as-rel2.txt.bz2>. 2020.
- [55] Center for Applied Internet Data Analysis. *The CAIDA Routeviws Prefix-to-AS Dataset*. <http://data.caida.org/datasets/routing/routeviws-prefix2as/>. 2020.
- [56] Center for Applied Internet Data Analysis (CAIDA). *Macroscopic Internet Topology Data Kit (ITDK)*. <https://www.caida.org/data/internet-topology-data-kit/> (last visited Jan 10, 2023), archived at <https://perma.cc/CU8X-7GRU>. 2022.
- [57] Niangjun Chen, Xiaoqi Ren, Shaolei Ren, and Adam Wierman. "Greening multi-tenant data center demand response." In: *Proceedings of the ACM SIGMETRICS Conference* (2015).

- [58] Abhijit K Choudhury and Ellen L Hahne. "Dynamic queue length thresholds for shared-memory packet switches." In: *IEEE/ACM Transactions on Networking (ToN)* 6 (1998).
- [59] George Christodoulou and Elias Koutsoupias. "The price of anarchy of finite congestion games." In: *Proceedings of the Symposium on Theory of Computing (STOC)*. 2005.
- [60] Laurent Chuat, Markus Legner, David Basin, David Hausheer, Samuel Hitz, Peter Müller, and Adrian Perrig. *The Complete Guide to SCION*. 2022.
- [61] Christine Chung, Katrina Ligett, Kirk Pruhs, and Aaron Roth. "The price of stochastic anarchy." In: *International Symposium on Algorithmic Game Theory*. 2008.
- [62] Cisco. *Cisco Annual Internet Report (2018–2023) White Paper*. <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>. 2020.
- [63] Cisco. *Nexus 9000 Series Switches*. <https://www.cisco.com/c/en/us/products/collateral/switches/nexus-9000-series-switches/white-paper-c11-738488.html>. 2021.
- [64] Luca Cittadini, Giuseppe Di Battista, Thomas Erlebach, Maurizio Patrignani, and Massimo Rimondini. "Assigning AS relationships to satisfy the Gao-Rexford conditions." In: *Proceedings of the IEEE Conference on Network Protocols (ICNP)*. 2010.
- [65] Edward H. Clarke. "Multipart pricing of public goods." In: *Public Choice* 11 (1971).
- [66] Antoine-Augustin Cournot. *Recherches sur les principes mathématiques de la théorie des richesses*. 1838.
- [67] Stella C Dafermos and Frederick T Sparrow. "The traffic assignment problem for a general network." In: *Journal of Research of the National Bureau of Standards B* 73 (1969).
- [68] Michael Dahlin. "Interpreting stale load information." In: *Transactions on Parallel and Distributed Systems (TPDS)* 11 (2000).
- [69] Hamza Dahmouni, André Girard, and Brunilde Sansò. "An analytical model for jitter in IP networks." In: *annals of telecommunications - annales des télécommunications* 67 (2012).

- [70] Edward J Daniel, Christopher M White, and Keith A Teague. "An interarrival delay jitter model using multistructure network delay characteristics for packet networks." In: *Proceedings of the Asilomar Conference on Signals, Systems & Computers (Asilomar SS& C)*. 2003.
- [71] Alex Davidson, Matthias Frei, Marten Gartner, Hamed Haddadi, Adrian Perrig, Jordi Subirà Nieto, Philipp Winter, and François Wirz. "Tango or Square Dance? How Tightly Should We Integrate Network Functionality in Browsers?" In: *Proceedings of the ACM Workshop on Hot Topics in Networks (HotNets)*. 2022.
- [72] Spencer Dawkins. *Path Aware Networking: Obstacles to Deployment (A Bestiary of Roads Not Taken)*. RFC 9049. June 2021. DOI: [10.17487/RFC9049](https://doi.org/10.17487/RFC9049). URL: <https://www.rfc-editor.org/info/rfc9049>.
- [73] Gerard Debreu. "Valuation equilibrium and Pareto optimum." In: *Proceedings of the US National Academy of Sciences*. 1954.
- [74] Amogh Dhamdhere, Constantine Dovrolis, and Pierre Francois. "A value-based framework for Internet peering agreements." In: *Proceedings of the International Teletraffic Congress (ITC)*. 2010.
- [75] Mo Dong, Tong Meng, Doron Zarchy, Engin Arslan, Yossi Gilad, Brighten Godfrey, and Michael Schapira. "PCC Vivace: Online-learning congestion control." In: *Proceedings of the USENIX Symposium on Networked Systems Design and Implementation (NSDI)*. 2018.
- [76] Pradeep Dubey. "Inefficiency of Nash equilibria." In: *Mathematics of Operations Research* 11 (1986).
- [77] Nandita Dukkipati and Nick McKeown. "Why flow-completion time is the right metric for congestion control." In: *ACM SIGCOMM Computer Communication Review (CCR)* (2006).
- [78] Francis Y Edgeworth. "The pure theory of monopoly." In: *Papers relating to political economy* 1 (1925).
- [79] Anwar Elwalid, Cheng Jin, Steven Low, and Indra Widjaja. "MATE: Multipath Adaptive Traffic Engineering." In: *Computer Networks* 40 (2002).
- [80] Ember Climate. *EU Carbon Price Tracker*. <https://ember-climate.org/data/data-tools/carbon-price-viewer/> (last visited Jan 10, 2023). 2022.
- [81] Thomas Erneux. *Applied delay differential equations*. Springer Science & Business Media, 2009.

- [82] Kevin Fall and Sally Floyd. "Simulation-based comparisons of Tahoe, Reno and SACK TCP." In: *ACM SIGCOMM Computer Communication Review (CCR)* (1996).
- [83] Joan Feigenbaum, Vijay Ramachandran, and Michael Schapira. "Incentive-Compatible Interdomain Routing." In: *Proceedings of the ACM Conference on Electronic Commerce (EC)*. 2006.
- [84] Simon Fischer, Nils Kammhuber, and Anja Feldmann. "REPLEX: Dynamic Traffic Engineering Based on Wardrop Routing Policies." In: *Proceedings of the International Conference on Emerging Networking Experiments and Technologies (CoNEXT)*. 2006.
- [85] Simon Fischer and Berthold Vöcking. "Adaptive routing with stale information." In: *Theoretical Computer Science* 410 (2009).
- [86] Nicky van Foreest, Michel Mandjes, and Werner Scheinhardt. "Analysis of a feedback fluid model for heterogeneous TCP sources." In: *Stochastic Models* (2003).
- [87] Linux Foundation. *OVS: Open vSwitch*. <https://www.openvswitch.org/>. 2021.
- [88] Eric J Friedman. "Genericity and congestion control in selfish routing." In: *Proceedings of the IEEE Conference on Decision and Control (CDC)*. 2004.
- [89] Oded Galor. *Discrete dynamical systems*. 2007.
- [90] Lixin Gao. "On inferring autonomous system relationships in the Internet." In: *IEEE/ACM Transactions on Networking (ToN)* 9 (2001).
- [91] Lixin Gao and Jennifer Rexford. "Stable Internet routing without global coordination." In: *IEEE/ACM Transactions on Networking (ToN)* (2001).
- [92] Nitin Garg. *Evaluating COPA congestion control for improved video performance*. <https://engineering.fb.com/2019/11/17/video-engineering/copa/>. 2019.
- [93] Tomer Gilad, Neta Rozen-Schiff, P Brighten Godfrey, Costin Raiciu, and Michael Schapira. "MPCC: Online learning multipath transport." In: *Proceedings of the International Conference on Emerging Networking Experiments and Technologies (CoNEXT)*. 2020.
- [94] Phillipa Gill, Michael Schapira, and Sharon Goldberg. "A survey of interdomain routing policies." In: *ACM SIGCOMM Computer Communication Review (CCR)* (2013).

- [95] Vasileios Giotsas, Matthew Luckie, Bradley Huffaker, and KC Claffy. "Inferring complex AS relationships." In: *Proceedings of the ACM Internet Measurement Conference (IMC)*. 2014.
- [96] Vasileios Giotsas, Shi Zhou, Matthew Luckie, and Kc Claffy. "Inferring multilateral peering." In: *Proceedings of the International Conference on Emerging Networking Experiments and Technologies (CoNEXT)*. 2013.
- [97] Giacomo Giuliani, Dominik Roos, Marc Wyss, Juan Angel Garcia-Pardo, Markus Legner, and Adrian Perrig. "Colibri: A Cooperative Lightweight Inter-Domain Bandwidth-Reservation Infrastructure." In: *Proceedings of the International Conference on Emerging Networking Experiments and Technologies (CoNEXT)*. 2021.
- [98] P. Brighten Godfrey, Igor Ganichev, Scott Shenker, and Ion Stoica. "Pathlet routing." In: *ACM SIGCOMM Computer Communication Review (CCR)* (2009).
- [99] P. Brighten Godfrey, Michael Schapira, Aviv Zohar, and Scott Shenker. "Incentive Compatibility and Dynamics of Congestion Control." In: *Proceedings of the ACM SIGMETRICS Conference*. 2010.
- [100] Ivan Gojmerac, Thomas Ziegler, Fabio Ricciato, and Peter Reichl. "Adaptive Multipath Routing for Dynamic Traffic Engineering." In: *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM)*. 2003.
- [101] Jose Gomez, Elie Kfouri, Jorge Crichigno, Elias Bou-Harb, and Gautam Srivastava. "A performance evaluation of TCP BBRv2 alpha." In: *Proceedings of the International Conference on Telecommunications and Signal Processing (TSP)*. 2020.
- [102] Ramesh Govindan and Anoop Reddy. "An analysis of Internet inter-domain topology and route stability." In: *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*. Vol. 2. 1997.
- [103] Timothy G. Griffin and Geoff Huston. *BGP Wedgies*. RFC. IETF, 2005.
- [104] Timothy G. Griffin, F. Bruce Shepherd, and Gordon Wilfong. "The stable paths problem and interdomain routing." In: *IEEE/ACM Transactions on Networking (ToN)* 10 (2002).
- [105] Timothy G. Griffin and Gordon Wilfong. "An analysis of BGP convergence properties." In: *ACM SIGCOMM Computer Communication Review (CCR)* (1999).

- [106] SIX Group. *Secure Swiss Finance Network: For secure, flexible and resilient data communication.* <https://www.six-group.com/en/products-services/banking-services/ssfn.html>. 2022.
- [107] Theodore Groves. "Incentives in teams." In: *Econometrica: Journal of the Econometric Society* (1973).
- [108] John Guckenheimer and Philip Holmes. *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields (Section 3.2)*. Springer Science & Business Media, 2013.
- [109] Arpit Gupta, Laurent Vanbever, Muhammad Shahbaz, Sean P Donovan, Brandon Schlinker, Nick Feamster, Jennifer Rexford, Scott Shenker, Russ Clark, and Ethan Katz-Bassett. "SDX: A Software Defined Internet Exchange." In: *ACM SIGCOMM Computer Communication Review (CCR)* 44 (2015).
- [110] Sangtae Ha, Injong Rhee, and Lisong Xu. "CUBIC: a new TCP-friendly high-speed TCP variant." In: *ACM SIGOPS Review* (2008).
- [111] Sangtae Ha, Injong Rhee, and Lisong Xu. *TCP CUBIC Implementation in Linux Kernel.* https://github.com/torvalds/linux/blob/master/net/ipv4/tcp_cubic.c. 2008.
- [112] Huaizhong Han, Srinivas Shakkottai, Christopher V Hollot, Rayadurgam Srikant, and Don Towsley. "Multi-path TCP: a joint congestion control and routing scheme to exploit path diversity in the internet." In: *IEEE/ACM Transactions on Networking (ToN)* (2006).
- [113] John C Harsanyi. "Games with incomplete information played by "Bayesian" players." In: *Management Science* 14 (1967).
- [114] Penelope Evelyn Haxell and Gordon Wilfong. "A Fractional Model of the Border Gateway Protocol (BGP)." In: *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2008.
- [115] David A Hayes and Grenville Armitage. "Revisiting TCP congestion control using delay gradients." In: *Proceedings of the International Conference on Research in Networking (NETWORKING)*. 2011.
- [116] Ward Heddeghem, Filip Idzikowski, Willem Vereecken, Didier Colle, Mario Pickavet, and Piet Demeester. "Power Consumption Modeling in Optical Multilayer Networks." In: *Photonic Network Communications* (2012).
- [117] Stephen Hemminger and David S. Miller. [TCP]: make cubic the default. <http://bit.ly/3ZjkUgY>. 2006.

- [118] Kin-Ho Ho, George Pavlou, Michael Howarth, and Ning Wang. "An Incentive-Based Quality of Service Aware Algorithm for Offline Inter-AS Traffic Engineering." In: *Proceedings of the IEEE International Workshop on IP Operations and Management (IPOM)*. 2004.
- [119] Mario Hock, Roland Bless, and Martina Zitterbart. "Experimental evaluation of BBR congestion control." In: *Proceedings of the IEEE Conference on Network Protocols (ICNP)*. 2017.
- [120] Christopher V Hollot, Vishal Misra, Don Towsley, and Wei-Bo Gong. "A control theoretic analysis of RED." In: *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*. 2001.
- [121] Ron Holzman and Nissan Law-Yone. "Strong equilibrium in congestion games." In: *Games and Economic Behavior* 21 (1997).
- [122] Chi-Yao Hong, Matthew Caesar, and P Brighten Godfrey. "Finishing flows quickly with preemptive scheduling." In: *ACM SIGCOMM Computer Communication Review (CCR)* 42 (2012).
- [123] Tobias Hoßfeld, Lea Skorin-Kapov, Poul E Heegaard, and Martin Varela. "Definition of QoE fairness in shared systems." In: *IEEE Communications Letters* 21 (2016).
- [124] Xiaomeng Huang, Chuang Lin, and Fengyuan Ren. "Generalized modeling and stability analysis of highspeed TCP and scalable TCP." In: *IEICE Transactions on Communications* 89 (2006).
- [125] Geoff Huston. "Interconnection, Peering and Settlements." In: *The Internet Protocol Journal* 2 (1999).
- [126] Geoff Huston. *BGP in 2020 – The BGP Table*. <https://labs.apnic.net/?p=1395> (last visited Jan 10, 2023), archived at <https://perma.cc/UMJ9-Q4QA>. 2021.
- [127] International Energy Agency (IEA). *Average CO₂ intensity of power generation from coal power plants, 2000-2020*. <https://www.iea.org/data-and-statistics/charts/average-co2-intensity-of-power-generation-from-coal-power-plants-2000-2020> (last visited Jan 10, 2023). 2022.
- [128] Gérard Iooss and Moritz Adelmeyer. *Topics in Bifurcation Theory and Applications*. Vol. 3. World Scientific, 1998.
- [129] Romain Jacob and Laurent Vanbever. "The Internet of tomorrow must sleep more and grow old." In: *Proceedings of the Workshop on Sustainable Computer Systems Design and Implementation (HotCarbon)*. 2022.

- [130] Van Jacobson. "Congestion Avoidance and Control." In: *ACM SIGCOMM Computer Communication Review (CCR)* (1988).
- [131] Van Jacobson, Neal Cardwell, Yuchung Cheng, and Soheil Hassas Yeganeh. *BBRv1 Source Code*. https://github.com/torvalds/linux/blob/master/net/ipv4/tcp_bbr.c. Accessed on 2023-03-28. 2023.
- [132] Paras Jain, Sam Kumar, Sarah Wooders, Shishir G Patil, Joseph E Gonzalez, and Ion Stoica. "Skyplane: Optimizing Transfer Cost and Throughput Using Cloud-Aware Overlays." In: *Proceedings of the USENIX Symposium on Networked Systems Design and Implementation (NSDI)*. 2023.
- [133] Raj Jain, Arjan Durresi, and Gojko Babic. "Throughput fairness index: An explanation." In: *ATM Forum contribution*. Vol. 99. 1999.
- [134] Cheng Jin, David X Wei, and Steven H Low. "FAST TCP: Motivation, Architecture, Algorithms, Performance." In: *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*. 2004.
- [135] Ramesh Johari and David Kim Hong Tan. "End-to-end congestion control for the Internet: Delays and stability." In: *IEEE/ACM Transactions on Networking (ToN)* 9 (2001).
- [136] Baptiste Jonglez and Bruno Gaujal. "Distributed and Adaptive Routing Based on Game Theory." In: *Proceedings of the International Teletraffic Congress (ITC)*. 2017.
- [137] Srikanth Kandula, Dina Katabi, Bruce Davie, and Anna Charny. "Walking the Tightrope: Responsive yet Stable Traffic Engineering." In: *ACM SIGCOMM Computer Communication Review (CCR)*. 2005.
- [138] Corine de Kater, Nicola Rustignoli, and Adrian Perrig. *SCION Overview*. Internet-Draft. Work in Progress. Internet Engineering Task Force, Mar. 2023. 33 pp. URL: <https://datatracker.ietf.org/doc/draft-dekater-panrg-scion-overview/03/>.
- [139] Frank Kelly. "Fairness and stability of end-to-end congestion control." In: *European Journal of Control* 9 (2003).
- [140] Frank Kelly and Thomas Voice. "Stability of end-to-end algorithms for joint routing and rate control." In: *ACM SIGCOMM Computer Communication Review (CCR)* (2005).
- [141] Frank P Kelly, Aman K Maulloo, and David KH Tan. "Rate control for communication networks: shadow prices, proportional fairness and stability." In: *Journal of the Operational Research Society (JORS)* 49 (1998).

- [142] Peter Key, Laurent Massoulié, and Don Towsley. "Path selection and multi-path congestion control." In: *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*. 2007.
- [143] Elie F Kfouri, Jose Gomez, Jorge Crichigno, and Elias Bou-Harb. "An emulation-based evaluation of TCP BBRv2 alpha for wired broadband." In: *Computer Communications* 161 (2020).
- [144] Ramin Khalili, Nicolas Gast, Miroslav Popovic, and Jean-Yves Le Boudec. "MPTCP is not Pareto-optimal: Performance issues and a possible solution." In: *IEEE/ACM Transactions on Networking (ToN)* (2013).
- [145] Simon Knight, Hung X Nguyen, Nickolas Falkner, Rhys Bowden, and Matthew Roughan. "The internet topology zoo." In: *IEEE Journal on Selected Areas in Communications* (2011).
- [146] Eric D Kolaczyk. *Statistical analysis of network data (Datasets)*. 2009. URL: <http://math.bu.edu/people/kolaczyk/datasets.html>.
- [147] Vasileios Kotronis, Rowan Kloti, Matthias Rost, Panagiotis Georgopoulos, Bernhard Ager, Stefan Schmid, and Xenofontas Dimitropoulos. "Stitching Inter-Domain Paths over IXPs." In: *Proceedings of the Symposium on SDN Research (SOSR)*. 2016.
- [148] Elias Koutsoupias and Christos Papadimitriou. "Worst-case equilibria." In: *Proceedings of the Annual Symposium on Theoretical Aspects of Computer Science (STACS)*. 1999.
- [149] Cyrill Krähenbühl, Seyedali Tabaeiaghdaei, Christelle Gloor, Jonghoon Kwon, Adrian Perrig, David Haasheer, and Dominik Roos. "Deployment and scalability of an inter-domain multi-path routing infrastructure." In: *Proceedings of the International Conference on Emerging Networking Experiments and Technologies (CoNEXT)*. 2021.
- [150] Anurag Kumar. "Comparative performance analysis of versions of TCP in a local network with a lossy link." In: *IEEE/ACM Transactions on Networking (ToN)* 6 (1998).
- [151] Gautam Kumar, Nandita Dukkipati, Keon Jang, Hassan MG Wassel, Xian Wu, Behnam Montazeri, Yaogong Wang, Kevin Springborn, Christopher Alfeld, Michael Ryan, et al. "Swift: Delay is simple and effective for congestion control in the datacenter." In: *Proceedings of the ACM SIGCOMM Conference*. 2020.

- [152] Amund Kvalbein, Constantine Dovrolis, and Chidambaram Muthu. "Multipath Load-Adaptive Routing: Putting the Emphasis on Robustness and Simplicity." In: *Proceedings of the IEEE Conference on Network Protocols (ICNP)*. 2009.
- [153] Craig Labovitz, Abha Ahuja, and Farnam Jahanian. "Experimental study of Internet stability and backbone failures." In: *Proceedings of the International Symposium on Fault-Tolerant Computing*. 1999.
- [154] Compira Labs. *Compira Labs Boosts Quality of Experience for Leading US Service Provider*. <https://www.compiralabs.com/case-study-download>. 2022.
- [155] Leah Ladines. *HIN Trust Circle, now powered by SCION*. <https://www.anapaya.net/blog/hin-trust-circle-now-powered-by-scion>. 2023.
- [156] Bob Lantz, Nikhil Handigol, Brandon Heller, and Vimal Jeyakumar. *Introduction to Mininet*. <https://github.com/mininet/mininet/wiki/Introduction-to-Mininet>. 2021.
- [157] Bob Lantz, Brandon Heller, and Nick McKeown. "A Network in a Laptop: Rapid Prototyping for Software-Defined Networks." In: *Proceedings of the ACM Workshop on Hot Topics in Networks (HotNets)*. 2010. ISBN: 9781450304092.
- [158] Lazard. *Levelized Cost of Energy Analysis: Version 14.0*. 2020.
- [159] Changhyun Lee, Chunjong Park, Keon Jang, Sue Moon, and Dongsu Han. "Accurate latency-based congestion feedback for datacenters." In: *Proceedings of the USENIX Annual Technical Conference (ATC)*. 2015.
- [160] Renato Paes Leme and Eva Tardos. "Pure and Bayes-Nash price of anarchy for generalized second price auction." In: *Proceedings of the Annual Symposium on Foundations of Computer Science (FOCS)*. 2010.
- [161] Omer Lev, Moshe Tennenholtz, and Aviv Zohar. "An axiomatic approach to routing." In: *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*. 2015.
- [162] Yu-Hsin Liu, Jeffrey Prince, and Scott Wallsten. "Distinguishing bandwidth and latency in households' willingness-to-pay for broadband internet speed." In: *Information Economics and Policy* 45 (2018).
- [163] Yong Liu, Francesco Lo Presti, Vishal Misra, Don Towsley, and Yu Gu. "Fluid models and solutions for large-scale IP networks." In: *Proceedings of the ACM SIGMETRICS Conference*. 2003.

- [164] Zhenhua Liu, Minghong Lin, Adam Wierman, Steven H Low, and Lachlan LH Andrew. "Greening geographical load balancing." In: *Proceedings of the ACM SIGMETRICS Conference* (2011).
- [165] Thomas Locher, Patrick Moore, Stefan Schmid, and Roger Wattenhofer. "Free riding in BitTorrent is cheap." In: *Proceedings of the ACM Workshop on Hot Topics in Networks (HotNets)*. 2006.
- [166] Steven H Low, Fernando Paganini, and John C Doyle. "Internet congestion control." In: *IEEE Control Systems Magazine* 22 (2002).
- [167] Marcelo Caggiani Luizelli, Leonardo Richter Bays, Luciana Salete Buriol, Marinho Pilla Barcellos, and Luciano Paschoal Gaspary. "Characterizing the impact of network substrate topologies on virtual network embedding." In: *Proceedings of the International Conference on Network and Service Management (CNSM)*. 2013.
- [168] Thomas Lukaseder, Leonard Bradatsch, Benjamin Erb, Rens W Van Der Heijden, and Frank Kargl. "A comparison of TCP congestion control algorithms in 10G networks." In: *Proceedings of the IEEE Conference on Local Computer Networks (LCN)*. 2016.
- [169] Richard TB Ma, Dah Ming Chiu, John CS Lui, Vishal Misra, and Dan Rubenstein. "Internet Economics: The use of Shapley value for ISP settlement." In: *IEEE/ACM Transactions on Networking (ToN)* (2010).
- [170] Andres Marentes, Thiago Teixeira, and Tilman Wolf. "Exploring economic dynamics in an internet with service choices." In: *Proceedings of the IEEE International Conference on Communications (ICC)*. 2015.
- [171] Matthew Mathis, Jeffrey Semke, Jamshid Mahdavi, and Teunis Ott. "The macroscopic behavior of the TCP congestion avoidance algorithm." In: *ACM SIGCOMM Computer Communication Review (CCR)* 27 (1997).
- [172] MathWorks. *Symbolic Math Toolbox*. <https://ch.mathworks.com/products/symbolic.html>. 2023.
- [173] MaxMind. *GeoLite2 Free Downloadable Databases*. <https://dev.maxmind.com/geoip/geoip2/geolite2/>. 2020.
- [174] Riad Mazloum, Marc-Olivier Buob, Jordan Auge, Bruno Baynat, Dario Rossi, and Timur Friedman. "Violation of interdomain routing assumptions." In: *Proceedings of the International Conference on Passive and Active Network Measurement (PAM)*. 2014.

- [175] Robert McMahon. *iPerf*. <https://sourceforge.net/projects/iperf2/>. 2021.
- [176] Eli A Meirom, Shie Mannor, and Ariel Orda. “Network formation games with heterogeneous players and the internet structure.” In: *Proceedings of the ACM Conference on Electronic Commerce (EC)*. 2014.
- [177] Nithin Michael and Ao Tang. “Halo: Hop-by-hop adaptive link-state optimal routing.” In: *IEEE/ACM Transactions on Networking (ToN)* (2014).
- [178] Jayne Miller. 2021 *Global Internet Map Tracks Global Capacity, Traffic, and Cloud Infrastructure*. <https://blog.telegeography.com/2021-global-internet-map-tracks-global-capacity-traffic-and-cloud-infrastructure> (last visited Jan 10, 2023). Feb. 2021.
- [179] Ayush Mishra, Sherman Lim, and Ben Leong. “Understanding speciation in QUIC congestion control.” In: *Proceedings of the 22nd ACM Internet Measurement Conference*. 2022.
- [180] Ayush Mishra, Xiangpeng Sun, Atishya Jain, Sameer Pande, Raj Joshi, and Ben Leong. “The great internet TCP congestion control census.” In: *Proceedings of the ACM SIGMETRICS Conference* (2019).
- [181] Ayush Mishra, Wee Han Tiu, and Ben Leong. “Are we heading towards a BBR-dominant Internet?” In: *Proceedings of the ACM Internet Measurement Conference (IMC)*. 2022.
- [182] Vishal Misra, Wei-Bo Gong, and Don Towsley. “Fluid-based analysis of a network of AQM routers supporting TCP flows with an application to RED.” In: *Proceedings of the ACM SIGCOMM Conference*. 2000.
- [183] Radhika Mittal, Vinh The Lam, Nandita Dukkipati, Emily Blem, Hassan Wassel, Monia Ghobadi, Amin Vahdat, Yaogong Wang, David Wetherall, and David Zats. “TIMELY: RTT-based congestion control for the datacenter.” In: *ACM SIGCOMM Computer Communication Review (CCR)* 45 (2015).
- [184] Michael Mitzenmacher. “How useful is old information?” In: *IEEE Transactions on Parallel and Distributed Systems (TPDS)* 11 (2000).
- [185] Dov Monderer and Lloyd S Shapley. “Potential games.” In: *Games and Economic Behavior* (1996).
- [186] Richard Mortier and Ian Pratt. “Incentive Based Inter-Domain Routing.” In: *Proceedings of the International Workshop on Networked Group Communication (NGC)*. 2003.

- [187] Mahdi Motalleb, Pierluigi Siano, and Reza Ghorbani. "Networked Stackelberg Competition in a Demand Response Market." In: *Applied Energy* 239 (2019).
- [188] Rudolf Müller, Andrés Perea, and Sascha Wolf. "Weak monotonicity and Bayes–Nash incentive compatibility." In: *Games and Economic Behavior* 61 (2007).
- [189] Roger B. Myerson. "Incentive compatibility and the bargaining problem." In: *Econometrica: Journal of the Econometric Society* (1979).
- [190] Roger B. Myerson and Mark A. Satterthwaite. "Efficient mechanisms for bilateral trading." In: *Journal of Economic Theory* 29 (1983).
- [191] John Nachbar. "The Myerson-Satterthwaite Theorem." In: *Notes for Game Theory Course* (2017).
- [192] Anna Nagurney and Tilman Wolf. "A Cournot–Nash–Bertrand game theory model of a service-oriented Internet with price and quality competition among network transport providers." In: *Computational Management Science* 11 (2014).
- [193] Aarti Nandagiri, Mohit P Tahiliani, Vishal Misra, and KK Ramakrishnan. "BBRv1 vs BBRv2: Examining performance differences through experimental evaluation." In: *Proceedings of the IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN)*. 2020.
- [194] John F Nash Jr. "The bargaining problem." In: *Econometrica: Journal of the Econometric Society* (1950).
- [195] Swaprava Nath and Tuomas Sandholm. "Efficiency and budget balance in general quasi-linear domains." In: *Games and Economic Behavior* 113 (2019).
- [196] Srihari Nelakuditi, Zhi-Li Zhang, Rose P Tsang, and David Hung-Chang Du. "Adaptive Proportional Routing: a Localized QoS Routing Approach." In: *IEEE/ACM Transactions on Networking (ToN)* (2002).
- [197] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V Vazirani. *Algorithmic game theory*. Cambridge University Press, 2007.
- [198] William B Norton. *The Internet peering playbook: connecting to the core of the Internet*. 2011.
- [199] NumPy Project. *NumPy*. <https://numpy.org/>. 2021.
- [200] Andrew Odlyzko. *Will smart pricing finally take off?* 2014.

- [201] Jörgen Olsén. "Stochastic modeling and simulation of the TCP protocol." PhD thesis. 2003.
- [202] OpenVault. *Broadband Insights Report (OVBI), Q4/2021*. https://openvault.com/wp-content/uploads/2022/03/0VBI_4Q21_Report_FINAL-1.pdf (last visited Jan 10, 2023). 2021.
- [203] Jitendra Padhye, Victor Firoiu, Don Towsley, and Jim Kurose. "Modeling TCP throughput: A simple model and its empirical validation." In: *Proceedings of the ACM SIGCOMM Conference*. 1998.
- [204] Qiuyu Peng, Anwar Walid, and Steven H Low. "Multipath TCP algorithms: theory and design." In: *Proceedings of the ACM SIGMETRICS Conference* (2013).
- [205] Jonathan Perry, Hari Balakrishnan, and Devavrat Shah. "Flowtune: Flowlet control for datacenter networks." In: *Proceedings of the USENIX Symposium on Networked Systems Design and Implementation (NSDI)*. 2017.
- [206] Jonathan Perry, Amy Ousterhout, Hari Balakrishnan, Devavrat Shah, and Hans Fugal. "Fastpass: A centralized "zero-queue" datacenter network." In: *Proceedings of the ACM SIGCOMM Conference*. 2014.
- [207] Sudheer Poojary and Vinod Sharma. "Analytical model for congestion control and throughput with TCP CUBIC connections." In: *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM)*. 2011.
- [208] Sudheer Poojary and Vinod Sharma. "An asymptotic approximation for TCP CUBIC." In: *Queueing Systems* 91 (2019).
- [209] Chutiphon Pukdeboon. "A review of fundamentals of Lyapunov theory." In: *The Journal of Applied Science* 10 (2011).
- [210] Lili Qiu, Yang Richard Yang, Yin Zhang, and Scott Shenker. "On selfish routing in internet-like environments." In: *Proceedings of the ACM SIGCOMM Conference*. 2003.
- [211] Sophie Y Qiu, Patrick D McDaniel, and Fabian Monroe. "Toward valley-free inter-domain routing." In: *Proceedings of the IEEE International Conference on Communications (ICC)*. 2007.
- [212] Barath Raghavan and Alex C. Snoeren. "A system for authenticated policy-compliant routing." In: *ACM SIGCOMM Computer Communication Review (CCR)* 34 (2004).

- [213] Barath Raghavan, Patric Verkaik, and Alex C. Snoeren. "Secure and Policy-Compliant Source Routing." In: *IEEE/ACM Transactions on Networking (ToN)* 17 (2009).
- [214] Costin Raiciu, Mark Handley, and Damon Wischik. *Coupled congestion control for multipath transport protocols*. RFC 6356. 2011.
- [215] Gaurav Raina and Damon Wischik. "Buffer sizes for large multiplexers: TCP queueing theory and instability analysis." In: *Proceedings of the Conference on Next Generation Internet Networks (NGI)*. 2005.
- [216] Anatol Rapoport. "Prisoner's dilemma." In: *The New Palgrave Dictionary of Economics*. Palgrave Macmillan UK, 1989.
- [217] Injong Rhee, Lisong Xu, Sangtae Ha, Alexander Zimmermann, Lars Eggert, and Richard Scheffenegger. *CUBIC for Fast Long-Distance Networks*. RFC 8312. IETF, 2018. 18 pp. doi: [10.17487/RFC8312](https://doi.org/10.17487/RFC8312).
- [218] Robert W Rosenthal. "A class of games possessing pure-strategy Nash equilibria." In: *International Journal of Game Theory* 2 (1973).
- [219] Benjamin Rothenberger, Dominik Roos, Markus Legner, and Adrian Perrig. "PISKES: Pragmatic Internet-Scale Key-Establishment System." In: *Proceedings of the ACM Asia Conference on Computer and Communications Security (ASIACCS)*. 2020.
- [220] Matthew Roughgarden. "Simplifying the synthesis of Internet traffic matrices." In: *ACM SIGCOMM Computer Communication Review (CCR)* 35 (2005).
- [221] Tim Roughgarden. "The price of anarchy is independent of the network topology." In: *Journal of Computer and System Sciences (JCSS)* 67 (2003).
- [222] Tim Roughgarden. "Routing games (Chapter 18)." In: *Algorithmic Game Theory*. 2007.
- [223] Tim Roughgarden. "The price of anarchy in games of incomplete information." In: *ACM Transactions on Economics and Computation (TEAC)* 3 (2015).
- [224] Tim Roughgarden and Éva Tardos. "How bad is selfish routing?" In: *Journal of the ACM (JACM)* 49 (2002).
- [225] Lorenzo Saino, Cosmin Cocora, and George Pavlou. "A toolchain for simplifying network simulation setup." In: *SimuTools* 13 (2013).
- [226] Jerome H Saltzer, David P Reed, and David D Clark. "End-to-end arguments in system design." In: *ACM Transactions on Computer Systems (TOCS)* 2 (1984).

- [227] William H Sandholm. "Potential games with continuous player sets." In: *Journal of Economic Theory* 97 (2001).
- [228] Stefan Savage, Andy Collins, Eric Hoffman, John Snell, and Thomas Anderson. "The end-to-end effects of Internet path selection." In: *ACM SIGCOMM Computer Communication Review (CCR)*. Vol. 29. 1999.
- [229] Simon Scherrer and James Dermelj. *BBR Fluid-Model Simulator and Mininet Experiment Suite*. <https://github.com/simonschdev/imc22-bbr-fluid-model>. Uploaded on 2022-09-01. 2022.
- [230] Laura Schmid, Krishnendu Chatterjee, and Stefan Schmid. "The Evolutionary Price of Anarchy: Locally Bounded Agents in a Dynamic Virus Game." In: *Proceedings of the International Symposium on Distributed Computing (OPODIS)*. 2019.
- [231] Dominik Scholz, Benedikt Jaeger, Lukas Schwaighofer, Daniel Raumer, Fabien Geyer, and Georg Carle. "Towards a deeper understanding of TCP BBR congestion control." In: *Proceedings of the IFIP Networking Conference*. 2018.
- [232] Soumya Sen, Carlee Joe-Wong, Sangtae Ha, and Mung Chiang. "A survey of smart data pricing: Past proposals, current plans, and future trends." In: *ACM Computing Surveys (CSur)* 46 (2013).
- [233] Anees Shaikh, Jennifer Rexford, and Kang G Shin. "Evaluating the Impact of Stale Link State on Quality-of-Service routing." In: *IEEE/ACM Transactions on Networking (ToN)* (2001).
- [234] Srinivas Shakkottai and Rayadurgam Srikant. "Economics of network pricing with multiple ISPs." In: *IEEE/ACM Transactions on Networking (ToN)* 14 (2006).
- [235] Gireesh Shrimali, Aditya Akella, and Almir Mutapcic. "Cooperative Interdomain Traffic Engineering using Nash Bargaining and Decomposition." In: *IEEE/ACM Transactions on Networking (ToN)* (2009).
- [236] Herbert A Simon. "Bounded rationality." In: *Utility and Probability*. Palgrave Macmillan UK, 1990.
- [237] Satinder Singh, Vishal Soni, and Michael Wellman. "Computing approximate Bayes-Nash equilibria in tree-games of incomplete information." In: *Proceedings of the ACM Conference on Electronic Commerce (EC)*. 2004.
- [238] Ankit Singla, Balakrishnan Chandrasekaran, P Brighten Godfrey, and Bruce Maggs. "The Internet at the speed of light." In: *Proceedings of the ACM Workshop on Hot Topics in Networks (HotNets)*. 2014.

- [239] Robert M Solow. "A contribution to the theory of economic growth." In: *The Quarterly Journal of Economics* 70 (1956).
- [240] Yeong-Jun Song, Geon-Hwan Kim, Imtiaz Mahmud, Won-Kyeong Seo, and You-Ze Cho. "Understanding of BBRv2: Evaluation and comparison with BBRv1 congestion control algorithm." In: *IEEE Access* 9 (2021).
- [241] Rayadurgam Srikant. *The mathematics of Internet congestion control*. Springer Science & Business Media, 2004.
- [242] Heinrich von Stackelberg. *Marktform und Gleichgewicht*. J. Springer, 1934.
- [243] Wright Stevens. *TCP slow start, congestion avoidance, fast retransmit, and fast recovery algorithms*. RFC 2001. 1997.
- [244] Thiago Teixeira, Andres Marentes, and Tilman Wolf. "Economic incentives in virtualized access networks." In: *Proceedings of the IEEE International Conference on Communications (ICC)*. 2017.
- [245] Brian Trammell, Jean-Pierre Smith, and Adrian Perrig. "Adding Path Awareness to the Internet Architecture." In: *IEEE Internet Computing* 22 (2018).
- [246] Belma Turkovic, Fernando A Kuipers, and Steve Uhlig. "Fifty shades of congestion control: A performance and interactions evaluation." In: *arXiv preprint arXiv:1903.03852* (2019).
- [247] Vytautas Valancius, Cristian Lumezanu, Nick Feamster, Ramesh Johari, and Vijay V Vazirani. "How many tiers? Pricing in the Internet transit market." In: *Proceedings of the ACM SIGCOMM Conference*. 2011.
- [248] Balajee Vamanan, Jahangir Hasan, and TN Vijaykumar. "Deadline-aware datacenter TCP (d2tcp)." In: *ACM SIGCOMM Computer Communication Review (CCR)* 42 (2012).
- [249] Gayane Vardoyan, C. V. Hollot, and Don Towsley. "Towards Stability Analysis of Data Transport Mechanisms: A Fluid Model and Its Applications." In: *IEEE/ACM Transactions on Networking (ToN)* (2021). doi: [10.1109/TNET.2021.3075837](https://doi.org/10.1109/TNET.2021.3075837).
- [250] Hal R Varian and JK MacKie Mason. "Some FAQs about usage-based pricing." In: *Okerson [Oke95]* (1994).
- [251] William Vickrey. "Counterspeculation, auctions, and competitive sealed tenders." In: *The Journal of Finance* 16 (1961).
- [252] Meng Wang, Chee Wei Tan, Weiyu Xu, and Ao Tang. "Cost of not splitting in routing: Characterization and estimation." In: *IEEE/ACM Transactions on Networking (ToN)* 19 (2011).

- [253] John Glen Wardrop. "Some theoretical aspects of road traffic research." In: *Proceedings of the Institution of Civil Engineers* (1952).
- [254] Ranysha Ware, Matthew K Mukerjee, Srinivasan Seshan, and Justine Sherry. "Modeling BBR's interactions with loss-based congestion control." In: *Proceedings of the ACM Internet Measurement Conference (IMC)*. 2019.
- [255] Ranysha Ware, Matthew K. Mukerjee, Srinivasan Seshan, and Justine Sherry. "Beyond Jain's Fairness Index: Setting the Bar For The Deployment of Congestion Control Algorithms." In: *Proceedings of the ACM Workshop on Hot Topics in Networks (HotNets)*. 2019, pp. 17–24. ISBN: 9781450370202. DOI: [10.1145/3365609.3365855](https://doi.org/10.1145/3365609.3365855). URL: <https://doi.org/10.1145/3365609.3365855>.
- [256] Cindy Whelan. *Wholesale Forecast: Backbone Services Prove Resilient, Offsetting Voice Declines*. <https://apacblog.lumen.com/wholesale-forecast-backbone-services-prove-resilient-offsetting-voice-declines/> (last visited Jan 13, 2023). 2020.
- [257] Stephen Wiggins and Martin Golubitsky. *Introduction to applied nonlinear dynamical systems and chaos (Chapter 18)*. 2003.
- [258] D. Wischik. "Queueing theory for TCP." In: *Presentation at the Conference on Stochastic Networks*. 2006.
- [259] Damon Wischik, Costin Raiciu, Adam Greenhalgh, and Mark Handley. "Design, Implementation and Evaluation of Congestion Control for Multipath TCP." In: *Proceedings of the USENIX Symposium on Networked Systems Design and Implementation (NSDI)*. 2011.
- [260] Tilman Wolf, James Griffioen, Kenneth L Calvert, Rudra Dutta, George N Rouskas, Ilya Baldin, and Anna Nagurney. "ChoiceNet: toward an economy plane for the Internet." In: *ACM SIGCOMM Computer Communication Review (CCR)* 44 (2014).
- [261] Wen Xu and Jennifer Rexford. "MIRO: Multi-path Interdomain Routing." In: *Proceedings of the ACM SIGCOMM Conference*. 2006.
- [262] Xiaokun Xu and Mark Claypool. "Measurement of cloud-based game streaming system response to competing TCP cubic or TCP BBR flows." In: *Proceedings of the ACM Internet Measurement Conference (IMC)*. 2022.
- [263] Furong Yang, Qinghua Wu, Zhenyu Li, Yanmei Liu, Giovanni Pau, and Gaogang Xie. "BBRv2+: Towards balancing aggressiveness and fairness with delay-based bandwidth probing." In: *Computer Networks* 206 (2022).

- [264] Ming Yang, Peng Yang, Chaozhun Wen, Qiong Liu, Jingjing Luo, and Li Yu. "Adaptive-BBR: Fine-grained congestion control with improved fairness and low latency." In: *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*. 2019.
- [265] Xiaowei Yang, David Clark, and Arthur W. Berger. "NIRA: A New Inter-Domain Routing Architecture." In: *IEEE/ACM Transactions on Networking (ToN)* (2007).
- [266] Yang Richard Yang, Haiyong Xie, Hao Wang, Avi Silberschatz, Arvind Krishnamurthy, Yanbin Liu, and Li Erran Li. "On Route Selection for Interdomain Traffic Engineering." In: *IEEE network* (2005).
- [267] Doron Zarchy, Amogh Dhamdhere, Constantine Dovrolis, and Michael Schapira. "Nash-peering: A new techno-economic framework for Internet interconnections." In: *INFOCOM Workshop*. 2018.
- [268] Doron Zarchy, Radhika Mittal, Michael Schapira, and Scott Shenker. "Axiomatizing congestion control." In: *Proceedings of the ACM SIGMETRICS Conference* (2019).
- [269] Lixia Zhang, Kevin Fall, and David Meyer. *Report from the IAB Workshop on Routing and Addressing*. RFC 4984. IETF, 2007.
- [270] Zhi-Li Zhang, Papak Nabipay, Andrew Odlyzko, and Roch Guerin. "Interactions, competition and innovation in a service-oriented Internet: An economic model." In: *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*. 2010.
- [271] Weihua Zhou, Yaqi Pu, Hongyan Dai, and Qingwei Jin. "Cooperative interconnection settlement among ISPs through NAP." In: *European Journal of Operational Research* 256 (2017).
- [272] Yibo Zhu, Haggai Eran, Daniel Firestone, Chuanxiong Guo, Marina Lipshteyn, Yehonatan Liron, Jitendra Padhye, Shachar Raindel, Mohamad Haj Yahia, and Ming Zhang. "Congestion control for large-scale RDMA deployments." In: *ACM SIGCOMM Computer Communication Review (CCR)* 45 (2015).
- [273] Yibo Zhu, Manya Ghobadi, Vishal Misra, and Jitendra Padhye. "ECN or Delay: Lessons Learnt from Analysis of DCQCN and TIMELY." In: *Proceedings of the International Conference on Emerging Networking Experiments and Technologies (CoNEXT)*. 2016.

- [274] Noa Zilberman, Eve M Schooler, Uri Cummings, Rajit Manohar, Dawn Nafus, Robert Soulé, and Rick Taylor. "Toward Carbon-Aware Networking." In: *Proceedings of the Workshop on Sustainable Computer Systems Design and Implementation (HotCarbon)*. 2022.