Internet Interdisciplinary Institute - Open University of Catalonia
(IN3 - UOC)

ECONOMICS OF INTERNET INTERDOMAIN INTERCONNECTIONS

Ph.D. Thesis

Author: Ignacio Castro, Open University of Catalonia & IMDEA Networks Institute
Director: Sergey Gorinsky, IMDEA Networks Institute

Information and Knowledge Society Doctoral Programme

Barcelona (Spain), 2015

Economics of Internet Interdomain Interconnections

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Prepared by

Ignacio Castro, Open University of Catalonia & IMDEA Networks Institute

Under the advice of

Sergey Gorinsky, IMDEA Networks Institute

Date: July, 2015

Web/contact: ignacio.decastro@imdea.org

This work has been supported by IMDEA Networks Institute.

Ph.D. Thesis

ECONOMICS OF INTERNET INTERDOMAIN INTERCONNECTIONS

Author: Ignacio Castro, Open University of Catalonia & IMDEA Networks Institute
Director: Sergey Gorinsky, IMDEA Networks Institute

Acknowledgements

Many people helped and accompanied me during my Ph.D. studies. I would like to briefly acknowledge those people without whom this thesis would be impossible.

I would like to thank Sergey Gorinsky for his assistance and support, Scott Shenker for his feedback and guidance, my collaborators (Aurogit Panda, Barath Raghavan, Juan Camilo Cardona Restrepo, Pierre François and Rade Stanojević), the members of IMDEA Networks Institute, its Network Economics Group (particularly to Pradeep Banguera for collecting the RedIRIS traffic data used in this thesis), the members of the Open University of Catalonia (UOC) and the Internet Interdisciplinary Institute (IN3), my Ph.D. committee members (Ricard Ruiz de Querol and Eduard Aibar Puentes), and the member of the UC Berkeley NetSys Lab and International Computer Science Institute (ICSI), for being so welcoming and engaging.

Abstract

The Internet is an evolving ecosystem where a multitude of interconnected networks, or Autonomous Systems (ASes), support global connectivity of end users. By providing economic incentives for routing traffic on behalf of other networks, interconnection agreements between ASes are a cornerstone of the Internet. However, rapid Internet adoption, unrelenting traffic growth, and increasing demands for quality and performance are challenging to cope with, provoke recurrent conflicts over the economic settlement of interconnections, and question the capacity of the Internet to provide critical services. Furthermore, with their capital-intensive network infrastructure, the ASes' need to recover costs complicates interconnection negotiations and settlements. Overcoming the limitations and bottlenecks of the evolving Internet ecosystem, requires understanding the economics of how networks interconnect and how traffic is routed through them.

This thesis studies the economic aspects of the interconnections between ASes, identifies challenges hampering the future of the Internet, and proposes solutions to resolve them. We begin by presenting the first analytical and empirical study on remote peering, an emerging type of interconnections that relaxes the geographical constraints of ASes and also facilitates interconnections at a lower cost. Then we introduce Cooperative IP Transit (CIPT) and Transit for Peering (T4P), two novel interconnection arrangements that reduce traffic delivery costs for the ASes. However, some of the limitations are inherent to the current Internet architecture. To overcome those constraints, we present Route Bazaar, a new Internet architecture that, inspired by the use of the block chain mechanism and cryptographic tools in cryptocurrencies, provides a contractual framework for flexible interconnections with rich policies.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Underlying the skyrocketing growth of the Internet, there is a complex and evolving ecosystem where a multitude of interconnected networks or Autonomous Systems (ASes) support global connectivity of end users. While social, cultural, and political factors are central to the Internet [35], this thesis studies the economic aspects of the physical interconnection of the networks that compose it. To achieve universal end-to-end connectivity, an interdomain protocol, Border Gateway Protocol (BGP) [92], allows networks to route traffic through intermediary networks. However, it is rather the economic relationships between ASes than the physical connectivity that dictate routing in the Internet. As a consequence, actual routes frequently deviate from the shortest paths making the physical topology of the Internet a weak guesstimate of how the Internet actually works.

In routing traffic on behalf of other networks, ASes seek compensation for the use of their resources. In the dynamic Internet ecosystem, ASes deal with potentially untrusted networks and changing traffic patterns. To cope with this uncertain environment, ASes negotiate interconnection agreements to secure payments for the use of their infrastructure as well as service conditions and compensations in case of infringement.

1.1 Background: Internet interconnections

Overcoming the limitations and bottlenecks of the evolving Internet interconnection ecosystem requires understanding the economics of how networks interconnect and how traffic is routed through them. Two types of interconnection agreements dominate the Internet landscape: Internet Protocol (IP) [155] transit and peering. Only a handful of huge ASes can access the entire Internet without paying anyone for the reachability. For the vast majority of the other ASes, the universal connectivity comes at the price of IP transit, or simply transit. Transit is a bilateral arrangement where the customer pays the provider for connectivity to the global Internet. Transit providers typically charge customers for the peak of its traffic with prices decreasing in quantity. Subadditive transit prices reflect the economies of scale present in traffic transport.

Transit costs are a significant part of the overall costs of ASes [60,74,105] because the decline of transit prices per Mbps is accompanied by the fast growth of transit traffic [186]. The problem of reducing the transit costs has attracted notable solutions including Internet eXchange Points (IXPs) [15, 55, 58], IP multicast [18, 28, 54, 82, 83], Content Delivery Networkss (CDNs) [146], Peer-to-Peer (P2P) localization [49], and traffic smoothing [63, 101, 114]. One property that these proposals share is their objective to reduce the amount of traffic that traverses transit links. Intuitively, the less traffic of an AS flows through those links, the lower the cost is for the AS.

Settlement-free peering is a cost-effective alternative to transit. If two ASes exchange their traffic via a transit provider, their payments to the provider significantly exceed the cost of communicating the same traffic over a settlement-free peering link. The costs of the peering are mostly related to the infrastructure and labor of maintaining the physical interconnection, either as a direct link or through an IXP. However the potential of settlement-free peering to reduce the costs

for both peers does not mean that the ASes will indeed establish and sustain such a relationship. For instance, the ASes might view each other as competitors and be unwilling to reduce the costs of the counterpart. Furthermore, the costs of each party depend on the peering-link traffic and AS sizes. Additionally, political or security related concerns could apply, for instance an AS might prefer to avoid specific regions or networks that could compromise the integrity of the traffic.

The early commercial Internet was essentially a hierarchy where smaller ASes paid bigger ASes for the universal Internet reachability via transit links. Subsequent massive emergence of peering enabled many ASes to exchange their customer traffic over a more economical settlement-free peering links [56, 78]. The evolution kept increasing the diversity of inter-AS connection types and introduced partial-transit and paid-peering links [70, 191]. In contrast to the full transit, a partial-transit link offers access to only a fraction of the global Internet address space. With paid peering, one of the peering ASes pays the other peer for exchanging their customer traffic.

In parallel to the interconnection evolution [57], the types of ASes have evolved as well. Some ASes run eyeball networks that primarily serve residential users. Other ASes concentrate on providing Internet access for content providers such as Yahoo or YouTube [70, 125]. While popular content providers are the major sources of Internet traffic, an eyeball network acts mostly as a traffic sink. Peering between content and eyeball ASes has been problematic not only because their traffic flows are unbalanced but also due to the heterogeneity of network types. The differences between content and eyeball networks complicate the issue of whether and how much one network should pay the other for their peering. Because the costs associated with last-mile infrastructures are typically high for the eyeball AS (and significantly higher than the infrastructure costs of the content AS), the eyeball AS can view the high costs as a just cause for demanding a compensation from the content AS. Moreover, since these high costs represent substantial barriers to entry in the eyeball-network market, the eyeball AS can try leveraging its significant market power when negotiating a peering agreement with the content AS [70, 125, 140]. On the other hand, the content AS can be reluctant to compensate the eyeball AS and even perceive such compensation demands as a violation of network neutrality [127]. The lack of clarity about proper conditions for eyeball-content peering has led to so-called peering wars [10, 25, 26, 154] which disrupted the Internet connectivity and ultimately lead to the net-neutrality debate [55].

While the Internet has proven to be extremely scalable, an ever growing traffic volume, increasing demands for quality and performance, and recurrent conflicts over the economic settlement of interconnections arrangements depict a challenging landscape for the future Internet. The traffic growth is a long-term trend [47, 110], even though the main application fueling the growth has been changing from web browsing [65] to P2P file sharing [181] to video streaming [152]. The growing popularity of delay-sensitive applications such as video streaming or Voice over IP (VoIP), together with the growing use of the Internet to provide critical services such as online banking or health services, puts additional pressure on the Internet structure.

1.2 Contributions

This thesis dwells on the economic aspects of the ASes' interconnections challenging the future of the Internet and how to overcome them. In the context of unrelenting traffic growth and with most networks dependent on transit providers to attain global Internet reachability, despite falling transit prices, transit charges are a substantial fraction of the costs for most ASes.

Even though peering allowed networks to reduce their transit costs, peering requires ASes to be physically colocated. As expanding the costly network infrastructure to colocate with other ASes is only affordable for ASes with large volumes of traffic, the initial wave of pervasive peering was lead by large content providers such as Google. In contrast to the ubiquitous presence of these large ASes, most networks had a narrower IXP presence restricted to their geographical footprint. Chapter 2 presents the first empirical and analytical study on the emerging phenomenon of *remote peering*, which enables ASes to overcome geographical barriers. Remote peering is an interconnection where a remote network reaches and peers with other networks via an intermediary called a remote-peering provider. By buying a remote-peering service, networks can peer without extending their own infrastructures to a shared location. Because remote peering is not observable with the typical data used for the understanding of Internet topologies, remote peering has been largely unnoticed by the academic community despite its wide adoption.

Chapter 3 proposes *Cooperative IP Transit (CIPT)*, a multilateral cooperative interconnection mechanism that helps ASes to deal with the financial burden of transit costs. CIPT reduces the price of transit per Mbps: by jointly purchasing the IP transit, two or more ASes reduce the transit prices per Mbps for each AS involved in the CIPT.

With the specialization of ASes into content providers and eyeball networks, different cost structures and rising traffic imbalances have resulted in the net neutrality debate and sparked recurrent conflicts over peering settlements. To alleviate these tensions, Chapter 4 proposes and evaluates *Transit for Peering (T4P)*, a hybrid bilateral AS relationship that continues the Internet trend towards more flexible interconnections at lower costs. With a T4P interconnection, one AS compensates the other AS for their peering by providing this other AS with a partial-transit service. In comparison to paid peering, T4P is able to reduce the combined transit/peering costs of an AS. By reducing traffic imbalances at a reduced cost, T4P has a potential to relax the ongoing tensions between content and eyeball ASes.

While innovations such as CIPT, T4P, and remote peering make the interdomain ecosystem more flexible, many limitations in the current Internet are inherent to its interconnection framework. Without explicit means for direct coordination among multiple networks, suboptimal routing and routes oscillation is frequent, and reaction to traffic-demand changes and infrastructure failures is problematic and slow. To ease these drawbacks, Chapter 5 departs from the current Internet interconnection framework and proposes instead *Route Bazaar*, a new Internet architecture inspired in the blockchain mechanism and cryptographic tools employed by cryptocurrencies mechanisms. Finally, Chapter 6 discusses related work, Chapter 7 considers future research, and Chapter 8 concludes this thesis.

Chapter 2

Remote Peering

While the Internet economic structure is greatly important for security, reliability, and other aspects of Internet design and operation, the structure remains poorly understood. Whereas layer-2 protocols deals with linking two neighboring nodes, layer-3 protocols provide end-to-end connectivity [107]. The Internet economic structure is typically modeled on layer 3 of Internet protocols, partly because of the ability to infer economic relationships from BGP and IP measurements.

In particular, BGP identifies ASes on announced paths, enabling inference of layer-3 structures where ASes act as economic entities interconnected by transit or peering relationships [76]. ASes are imperfect proxies of organizations, e.g., multiple ASes can be owned by a single organization and act as a single unit. Nevertheless, AS-level topologies [40,183] have proved themselves useful for reasoning about Internet connectivity, routing, and traffic delivery. Despite the usefulness, layer-3 models struggle to detect and correctly classify a significant portion of relationships in the dynamic economic structure.

Internet flattening refers to a validated evolutionary trend where major content providers expand their networks to directly connect with eyeball networks, that primarily serve residential users, and thereby reduce the number of intermediaries on end-to-end paths [30, 56, 78]. Flattening opposes the hierarchical structure characterizing the early commercial Internet, where smaller networks paid bigger ones for the universal reachability via transit links. The emergence of peering allowed networks to bypass transit providers, weakening the role of the latter in the Internet structure.

The flattening trend is commonly conflated with a trend towards more peering. Indeed, direct interconnections between content and eyeball networks are mostly peering relationships. Furthermore, cost reductions offered by peering serve as economic incentives for content providers to expand their networks. The peering is typically done at IXPs [1, 15, 32]. These layer-2 infrastructures keep growing in the number of their members and amount of peering traffic.

This chapter presents the first empirical and analytical study on an emerging phenomenon of remote peering that separates the trends of increasing peering and Internet flattening. Remote

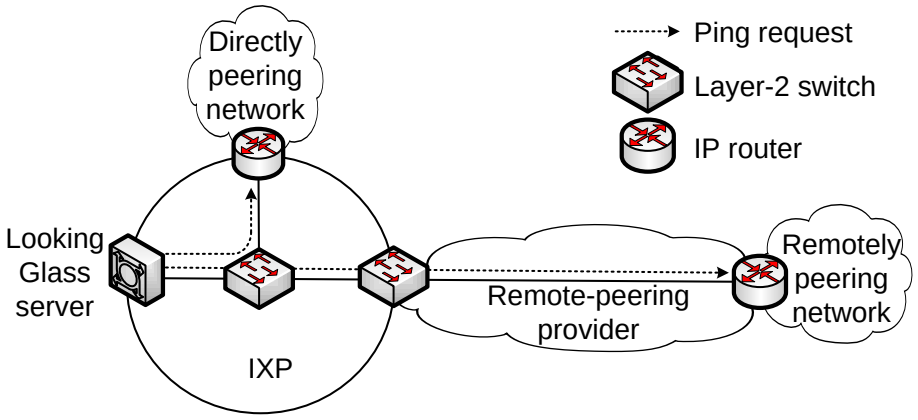


Figure 2.1: Directly and remotely peering networks, and probing them from an LG server

peering is an interconnection where a remote network reaches and peers with other networks via a layer-2 intermediary called a remote-peering provider. By buying a remote-peering service, networks can peer without extending their own infrastructures to a shared location. While remote peering enables additional peering, the increase in peering does not flatten the economic structure. Remote-peering providers include not only new companies, such as IX Reach [99] and Atrato IP Networks [12], but also traditional transit providers that leverage their traffic-delivery expertise to act as remote-peering intermediaries. Hence, remote peering means more peering without Internet flattening.

This chapter reports two measurement studies and a mathematical model that generalizes the empirical findings. First, we develop a ping-based method that conservatively estimates the spread of remote peering. We apply the method in 22 IXPs worldwide and detect remote peers in more than 90% of the studied IXPs, with remote peering by up to 20% of the members at an IXP. Our second study evaluates how remote peering can affect traffic patterns. Based on ground truth traffic in a research and education network, we estimate the amount of transit traffic that this network might offload via remote peering at 65 IXPs. The results show significant offload potential, around 25% of the traffic in some scenarios. While the measurements reveal diminishing marginal utility of remote peering at additional IXPs, we generalize this property in the mathematical model and derive conditions for economic viability of remote peering.

By demonstrating the wide spread and significant traffic offload potential of remote peering, our results challenge the research community's reliance on layer-3 topologies in representing the Internet economic structure. Because remote-peering providers are layer-2 entities, they are invisible on layer 3. Oblivious of the layer-2 intermediaries, layer-3 perspectives do not distinguish between remote peering and direct peering. Thus, layer-3 models fail to expose that remote peering separates the trends of increasing peering and Internet flattening. Our findings identify a need to reflect the presence of layer-2 entities in the Internet economic structure.

The wide spread of remote peering has broader implications for Internet research. The presence of intermediaries that are invisible on layer 3 adds to the security concerns as remote peering introduces invisible intermediaries that could monitor traffic or deliver it through undesired geographies. For Internet accountability, it is a challenge to associate an action with the responsible invisible entity. When a provider offers transit and remote peering, buying both might not yield reliable multihoming. As a new economic option, remote peering opens a whole new ballgame for connectivity, routing, and traffic distribution, e.g., via newly enabled IXPs [119]. To sum up, this chapter makes the following main contributions:

- This chapter reports the first systematic study of remote peering. The work illuminates the emerging phenomenon that many in the research community are unaware of. Even those who already know about remote peering benefit from our quantification of its wide spread and significant traffic offload potential.

- Our work reveals separation between the trends of increasing peering and Internet

flattening. While remote-peering providers enable additional peering, they act as intermediaries in the Internet economic structure.

- The results call for a rethink of modeling the Internet economic structure as layer-3 topologies. There is a need to reflect the increasing presence of layer-2 entities in the Internet structure.

- The demonstrated prominence of remote peering has broader implications for security, accountability, reliability, economics, and other aspects of Internet research.

The rest of this chapter is organized as follows. Section 2.1 provides background on Internet economic interconnections. Section 2.2 empirically studies the spread of remote peering. Section 2.3 estimates a network's potential to offload transit traffic to remote peering. Section 2.4 analyzes economic viability of remote peering versus transit and direct peering. Section 5.3 discusses broader implications of our findings. Finally, Section 2.6 sums up the chapter.

2.1 Interconnection landscape

We start by providing relevant background on economic relationships between networks in the Internet.

2.1.1 Transit

Transit refers to a bilateral interconnection where the customer pays the provider for connectivity to the global Internet. In a common setting, transit traffic is metered at 5-minute intervals and billed on a monthly basis, with the charge computed by multiplying a per-Mbps price and the 95th percentile of the 5-minute traffic rates [61, 180]. In the early commercial Internet, traffic flowed mostly through a hierarchy of transit relationships, with a handful of tier-1 networks at the top of the hierarchy.

2.1.2 Peering

Peering is an arrangement where two networks exchange traffic directly, rather than through a transit provider, and thereby reduce their transit costs. The exchange is commonly limited to the traffic belonging to the peering networks and their customer cones, i.e., their direct and indirect transit customers. To reduce costs further, peering is typically done at IXPs.

Networks differ in their policies for recognizing another network as a potential peer. The peering policies are typically classified as open, selective, and restrictive [122, 151]. An open policy allows the network to peer with every network. A network with a selective policy peers only if certain conditions are met. A restrictive policy has stringent terms that are difficult to satisfy.

Costs of peering and transit have different structures. Peering involves a number of traffic-independent costs, e.g., IXP membership fees and equipment maintenance expenses at the IXP. Peering also has traffic-dependent costs, e.g., IXP ports for higher traffic rates are more expensive. Over the years, peering relationships have proven themselves as cost-effective alternatives to transit.

Partly due to the lower costs, peering has spread widely, with the IXPs growing into major hubs for Internet traffic. Peering relationships bypass layer-3 transit providers and thus make the Internet flatter, at least on layer 3.

In this chapter, *direct peering* at an IXP refers to peering by a network that has IP presence in the IXP location. If a network is not co-located with the IXP already, the network can establish its IP presence at the IXP by contracting an IP transport service or extending its own IP infrastructure to reach the IXP location.

2.1.3 Remote peering

Remote peering constitutes an emerging type of interconnection where an IP network reaches and peers at a distant IXP via a layer-2 provider [24]. The remote-peering provider delivers traffic between the layer-2 switching infrastructure of the IXP and remote interface of the customer. On the customer's behalf, the remote-peering provider also maintains networking equipment at the IXP to enable the remote network to peer with other IXP members. Figure 2.1 depicts a typical setting for the remote-peering relationship.

Remote peering provides a smaller connectivity scope than transit. Instead of global Internet access, this service limits the connectivity to the reached IXP members and their customer cones. Technologically, remote peering can be implemented with standard methods, such as those used in layer-2 MPLS (MultiProtocol Label Switching) VPNs (Virtual Private Networks). The main innovation of remote peering lies in its economics.

Remote peering has both traffic-dependent and traffic-independent costs. In comparison to direct peering, the traffic-independent cost is lower, and the traffic-dependent cost is higher: the remote-peering provider has multiple customers and reduces its per-unit costs due to traffic aggregation and acquisition of IXP resources in bulk. Compared to transit, remote peering has lower traffic-dependent costs. Thus, from the cost perspective, remote peering represents a trade-off between direct peering and transit.

IXPs and remote peering are highly symbiotic. IXPs benefit from remote peering because the latter brings extra traffic to IXPs, enriches geographical diversity of IXP memberships, and strengthens the position of IXPs in the Internet economic structure. To promote remote peering, AMS-IX (Amsterdam Internet Exchange), DE-CIX (German Commercial Internet Exchange), LINX (London Internet Exchange), and many other IXPs establish partnership programs that incentivize distant networks to peer remotely at the IXP. For example, some IXPs reduce membership fees for remotely peering networks. AMS-IX started its partnership program around year 2003. According to our personal communications with AMS-IX staff, about one fifth of the

AMS-IX members were remote peers at the time of our study.

Implications of remote peering for transit providers are mixed. On the one hand, remote peering gives transit customers alternative means for reaching distant networks. On the other hand, remote peering is a new business niche where transit providers can leverage their traffic-delivery expertise.

According to anecdotal evidence, remote peering successfully gains ground and satisfies diverse needs in the Internet ecosystem. In this chapter, we focus on usages where remote peering at IXPs is purchased by distant networks or other IXPs. For example, AMS-IX Hong Kong and AMS-IX interconnect their infrastructures via remote peering to create additional peering opportunities for their members [178]. We do not consider an alternative usage where remote peering at an IXP is bought by a local network to benefit from cost reductions that remote peering provides even over short distances [43].

2.2 Spread of remote peering

In this section, we report measurements that conservatively estimate the spread of remote peering in the Internet.

2.2.1 Measurement methodology

Because remote peering is provided on layer 2, conventional layer-3 methods for Internet topology inference are unsuitable for the detection of remote peering. For instance, traceroute and BGP data do not reveal IP addresses or ASNs (AS Numbers) of remote-peering providers.

The basic idea of our methodology for detecting a remotely peering network at an IXP is to measure propagation delay between the network and IXP. Specifically, we use the ping utility to estimate the minimum RTT (Round-Trip Time) between the IXP location and the IP interface of the network in the IXP subnet. If the minimum RTT estimate exceeds a threshold, we classify the network as remotely peering at the IXP.

While our ping-based method is intuitive, the main challenges lie in its careful implementation and include: identification of probed interfaces, selection of vantage points, adherence to straight routes, sensitivity to traffic conditions, identification of networks, choice of IXPs, threshold for remoteness, IXPs with multiple locations, impact of blackholing, and measurement overhead. We discuss these challenges below.

Identification of probed interfaces: The targets of our ping probes are the IP interfaces of the IXP members in the IXP subnet. IXP members do not typically announce the IP addresses of these interfaces via BGP. To determine the IP addresses of the targeted interfaces, we look up the addresses on the websites of PeeringDB [151], PCH (Packet Clearing House) [150], and IXP itself.

Selection of vantage points: The ping requests need be launched into the IXP subnet from within the IXP location so that the requests take the direct route from the IXP location to the

probed interface. We send the ping requests from LG servers that PCH and RIPE NCC (Réseaux IP Européens Network Coordination Centre) [164] maintain at IXP locations. Figure 2.1 depicts our probing of IP network interfaces from an LG server at an IXP.

Adherence to straight routes: With our choice of the vantage points, the ping requests and ping replies are expected to stay within the IXP subnet. It is important to keep the probe routes straight because otherwise the RTT measurements might be high even for a directly peering network. Potential dangers include an unexpected situation where the device of a probed IP interface replies from one of its other IP interfaces and thereby sends the ping reply through an indirect route with multiple IP hops. A more realistic danger is that some of our targeted IP addresses are actually not in the IXP subnet because the respective website information is incorrect. To protect our method from such dangers, we examine the TTL (Time To Live) field in the received ping replies. When ping replies stay within the layer-2 subnet, their TTL values stay at the maximum set by the replying interface [195]. When the path of a ping reply includes an extra IP hop, the TTL value in the reply decreases. Therefore, we discard the ping replies with different TTL values than an expected maximum. We refer to this discard rule as a *TTL-match filter*. For the expected maximum TTL, our experiments accept two typical values of 64 and 255 hops. Although ping software might set the maximum TTL to other values (e.g., 32 or 128 hops), these alternative settings are relatively infrequent, and ignoring them does not significantly increase the number of discarded ping replies in our experiments. Also, different ping replies from the same interface might arrive with different TTL values, e.g., because the replying interface changes its maximum TTL. Whereas we are interested in a conservative estimate for the extent of remote peering, we discard all replies from an IP interface if their TTL value changes during the measurement period. We call this rule a *TTL-switch filter*.

Sensitivity to traffic conditions: Even if a probe stays within the IXP subnet, RTT might be high due to congestion. To deal with transient congestion, we repeat the measurements at different times of the day and different days of the week for each probed IP interface, and record the minimum RTT observed for the interface during the measurement period. This minimum RTT serves as a basis for deciding whether the interface is remote. Again to be on the conservative side, we exclude an IP interface from further consideration if we do not get at least 8 TTL-accepted ping replies from this interface for each probing Looking Glass (LG) server. We call this rule a *sample-size filter*. The limit of 8 replies and other parameter values in our study are empirically chosen to obtain reliable results while keeping the measurement overhead low. If less than 4 of the collected ping replies have RTT values within the maximum of 5 ms and 10% of the minimum RTT, i.e., below $\text{RTT}_{min} + \max\{5\text{ ms}, 0.1 \cdot \text{RTT}_{min}\}$, we apply an *RTT-consistent filter* to disregard the interface. For an IXP that has both PCH and RIPE NCC servers, we probe each IP interface from both LG servers and exclude the interface from further consideration if the larger of the two respective minimum RTTs is not within the maximum of 5 ms and 10% of the smaller one. We refer to this rule as an *LG-consistent filter*.

Identification of networks: To identify the network that owns a probed IP interface, we

IXP acronym	IXP name	Location		Peak traffic (Tbps)	Number of members	Number of analyzed interfaces
		City	Country			
AMS-IX	Amsterdam Internet Exchange	Amsterdam	Netherlands	5.48	638	665
DE-CIX	German Commercial Internet Exchange	Frankfurt	Germany	3.21	463	535
LINX	London Internet Exchange	London	UK	2.60	497	521
HKIX	Hong Kong Internet Exchange	Hong Kong	China	0.48	213	278
NYIIX	New York International Internet Exchange	New York	USA	0.46	132	239
MSK-IX	Moscow Internet eXchange	Moscow	Russia	1.32	367	218
PLIX	Polish Internet Exchange	Warsaw	Poland	0.63	235	207
France-IX	France-IX	Paris	France	0.23	230	201
PTT	PTTMetro São Paolo	São Paolo	Brazil	0.30	482	180
SIX	Seattle Internet Exchange	Seattle	USA	0.53	177	175
LoNAP	London Network Access Point	London	UK	0.10	142	166
JPIX	Japan Internet Exchange	Tokyo	Japan	0.43	131	163
TorIX	Toronto Internet Exchange	Toronto	Canada	0.28	177	161
VIX	Vienna Internet Exchange	Vienna	Austria	0.19	121	134
MIX	Milan Internet Exchange	Milan	Italy	0.16	133	131
TOP-IX	Torino Piemonte Internet Exchange	Turin	Italy	0.05	80	91
Netnod	Netnod Internet Exchange	Stockholm	Sweden	1.34	89	71
KINX	Korea Internet Neutral Exchange	Seoul	South Korea	0.15	46	71
CABASE	Argentine Chamber of Internet	Buenos Aires	Argentina	0.02	101	68
INEX	Internet Neutral Exchange	Dublin	Ireland	0.13	63	66
DIX-IE	Distributed Internet Exchange in Edo	Tokyo	Japan	N/A	36	56
TIE	Telx Internet Exchange	New York	USA	0.02	149	54

Table 2.1: Properties of the 22 IXPs in our measurement study on the spread of remote peering

use the network's ASN. We map the IP addresses to ASNs through a combination of looking up PeeringDB, using the IXPs' websites and LG servers, and issuing reverse DNS (Domain Name System) queries. If the ASN of an IP interface changes during the measurement period, we exclude the IP interface from further consideration. This exclusion rule is called an *ASN-change filter*.

Choice of IXPs: In choosing IXPs, we strive for a global scope surpassing the regional focuses of prior IXP studies. Our choice is constrained to those IXPs that have at least one LG server. Under the above constraints, we select and experiment at 22 IXPs in the following 4 continents: Asia, Europe, North America, and South America. After manually crawling the websites of the IXPs in January 2014, we collect data on their location, peak traffic, and number of members. Table 2.1 sums up these data. While information at IXP websites is often incomplete, out of date, or inconsistent in presenting a property (e.g., peak traffic), our measurement method does not rely on these data. We report this information just to give the reader a rough idea about the geography and size of the studied IXPs. For each studied IXP, Table 2.1 also includes the *number of analyzed interfaces*, i.e., interfaces that stay in our analyzed dataset after applying all 6 aforementioned filters. Across all the 22 IXPs, we apply the filters in the following order: sample-size, TTL-switch, TTL-match, RTT-consistent, LG-consistent, and ASN-change. After the filters discard 20, 82, 20, 100, 28, and 5 interfaces respectively, we have a total of 4,451 analyzed interfaces. The high count of TTL-switch discards is likely due to operating system changes during our measurements.

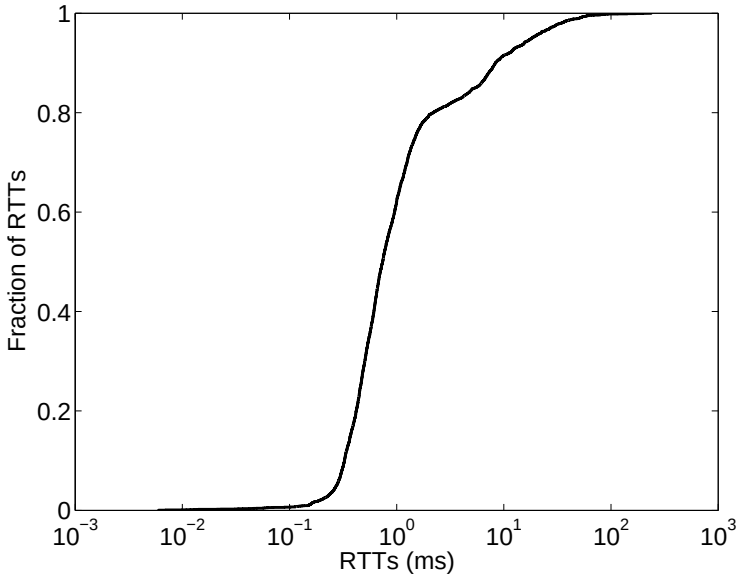


Figure 2.2: Empirical CDF of the analyzed interfaces according to their minimum RTTs

Threshold for remoteness: We classify a network as remotely peering at an IXP if the minimum RTT observed for its IP interface at the IXP exceeds a threshold. Despite the redundancy of our RTT measurements, the minimum RTT might still include non-propagation delays, e.g., due to persistent congestion of the IXP subnet or probe processing in the network devices. To minimize the possibility that such extra delays trigger an erroneous classification of a directly peering network as remote, the threshold should be sufficiently high. Figure 2.2 plots the cumulative distribution of the minimum RTTs for all the 4,451 analyzed interfaces. A majority of the analyzed interfaces have minimum RTTs distributed almost uniformly between 0.3 and 2 ms. This is a pattern expected for directly peering networks. The likelihood of a network being a direct peer declines as the minimum RTT increases. Our manual checks do not detect any directly peering network with the minimum RTT exceeding 10 ms. Thus, we set the remoteness threshold in our study to 10 ms. While this relatively high threshold value comes with a failure to recognize some remotely peering networks as remote peers, the false negatives do not constitute a significant concern because we mostly strive to avoid false positives in estimating the spread of remote peering conservatively.

IXPs with multiple locations: If an IXP operates interconnected switches in multiple locations, probes from an LG server at one location to an IP interface at another location might have a large RTT. The chosen remoteness threshold of 10 ms is sufficiently high to avoid false positives in cases where all locations of the IXP are in the same metropolitan area. False positives are possible if the geographic footprint of the IXP is significantly larger, e.g., spans multiple countries. We do not observe such situations in our experiments. In a more common scenario, two partner IXPs from different regions, e.g., AMS-IX Hong Kong and AMS-IX, interconnect by buying layer-2 connectivity from a third party. Our methodology correctly classifies such scenarios as remote

peering.

Impact of blackholing: If a probed interface intentionally blackholes or accidentally fails to respond to ping requests, the IP interface might be excluded from our analyzed data due to a low number of ping replies for the interface, as discussed above. In a hypothetical (not observed in our experiments) scenario where the probed interface forwards the probe to another machine that sends a ping reply on the interface's behalf, the ping reply is discarded by our TTL-match filter and does not affect accuracy of our RTT measurements.

Measurement overhead: While our method relies on probing from public LG servers, it is important to keep the measurement overhead low. The probes are launched through HTML (HyperText Markup Language) queries to the servers. The LG servers belonging to RIPE NCC and PCH react to an HTML query by issuing respectively 3 and 5 ping requests. For any LG server, we submit at most one HTML query per minute and generally spread the measurements over 4 months. The maximum number of ping replies received from any probed IP interface is 21 and 54 for respectively RIPE NCC and PCH servers.

We conducted the measurements during the 4 months from October 2013 to January 2014.

2.2.2 Experimental results

Figure 2.3 classifies all 4,451 analyzed interfaces across the 22 IXPs according to the minimum RTT measured for each interface. Despite using the high value of 10 ms for the remoteness criterion, the classification does not reveal remote interfaces in only two IXPs (DIX-IE and CABASE), i.e., our conservative estimate finds remote peering in 91% of the studied IXPs. While the numbers of remote interfaces are large in the 3 biggest IXPs (AMS-IX, DE-CIX, and LINX), these numbers are also large at smaller IXPs such as France-IX in France, PTT in Brazil, JPIX in Japan, and TOP-IX in Italy. Hence, our method independently confirms wide presence of remote peering in the Internet economic structure.

The classification in Figure 2.3 looks at the remote interfaces in greater detail by considering the following 3 ranges for the minimum RTT: [10 ms; 20 ms), [20 ms; 50 ms), and [50 ms; ∞) which roughly correspond to intercity, intercountry, and intercontinental distances. We detect the intercontinental-range peering at 12 IXPs, i.e., a majority of the studied IXPs. For example, Italian network E4A remotely peers at both TIE and TorIX, based in the USA and Canada respectively. Brazilian networks comprise most of the remote peers at PTT, the largest among the 21 IXPs of the PTTMetro project in Brazil. The high fraction of remote interfaces at the Turin-based TOP-IX likely results from the IXP's interconnections with VSIX and LyonIX, two other Southern European IXPs located in Padua and Lyon respectively.

Switching the perspective from the interfaces to the networks that own them, we apply our network identification method (described in Section 2.2.1) to determine ASNs for 3,242 out of the 4,451 analyzed interfaces. While a network might have interfaces at multiple IXPs, we identify a total of 1,904 networks. We refer to the number of the studied IXPs where a network peers as an *IXP count* of the network. Figure 2.4a presents the distribution of the IXP counts for all the 1,904

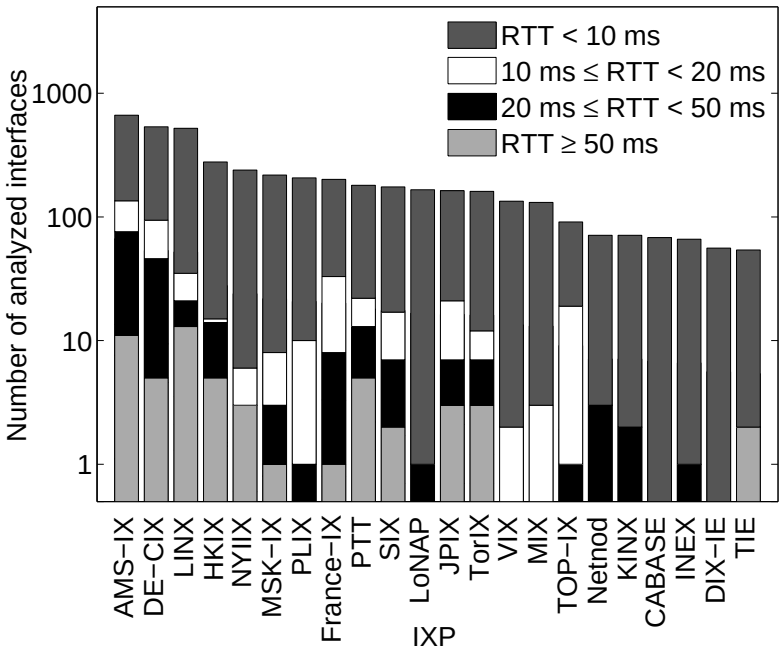
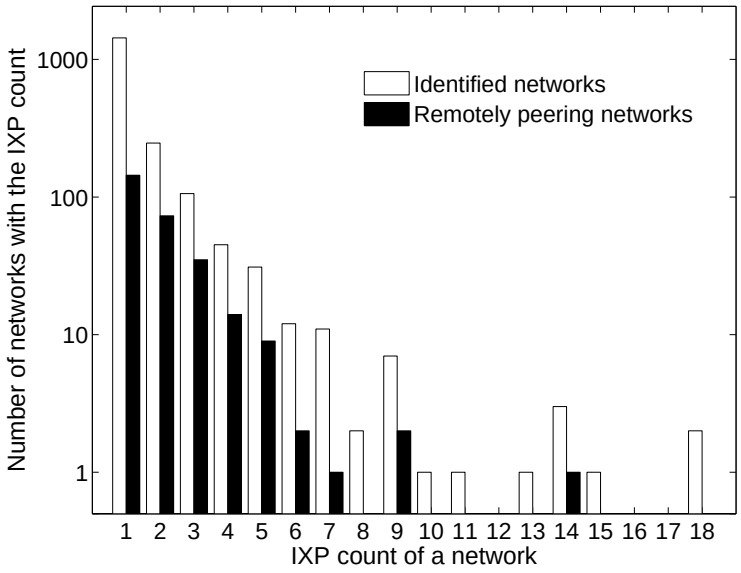


Figure 2.3: Classification of the analyzed interfaces according to their minimum RTTs

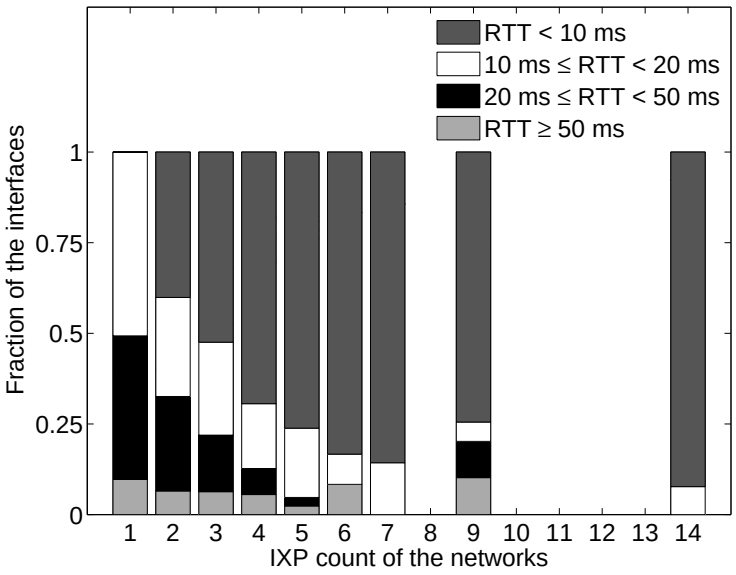
identified networks. While a majority of the networks connect to only one IXP, some networks peer at as many as eighteen IXPs.

285 of the identified networks have a remote interface at a studied IXP. Business services offered by the remotely peering networks are diverse and include transit (e.g., Türk Telecom), access (e.g., E4A and Invitel), and hosting (e.g., Trunk Networks). Figure 2.4a also plots the distribution of the IXP counts for all the 285 remotely peering networks. Both distributions in Figure 2.4a are qualitatively similar, suggesting that the choice of IXPs for a network to peer is relatively independent of whether the network peers directly or remotely.

We also examine the remotely peering networks with respect to the minimum RTTs of their analyzed interfaces. For each IXP count, we consider all the analyzed interfaces of the remotely peering networks with this IXP count and separate the interfaces into the following 4 categories according to their minimum RTT: [0 ms; 10 ms), [10 ms; 20 ms), [20 ms; 50 ms), and [50 ms; ∞). Figure 2.4b depicts the fractions of these 4 categories. By definition, the remote peering networks with the IXP count of 1 have no interfaces with the minimum RTT below 10 ms. As the IXP count increases, the fraction of the remote interfaces tends to decline because some interfaces of the remotely peering networks are used for direct peering. E4A exemplifies networks with a large number of remote interfaces: 6 of its 9 analyzed interfaces are classified as remote.



(a) Distributions of the IXP counts



(b) Interfaces of all the 285 remotely peering networks

Figure 2.4: IXP-count distributions and interface classifications for identified networks

2.2.3 Method validation

While our methodology employs a series of filters and high remoteness threshold to avoid false positives, this section reports how we validate the method and its conservative estimates of remote peering.

First, we use ground truth at TorIX, an IXP located in Toronto. TorIX staff confirmed that their members classified as remotely peering networks in our study are indeed remote peers. In one case, the TorIX staff initially thought that a network identified as a remote peer by our method

was rather a local member with a direct peering connection. Nevertheless, a closer examination showed that throughout our measurement period this local member conducted maintenance of its Toronto PoP (Point of Presence) and connected to TorIX from its remote PoP via a contracted layer-2 facility.

Then, we take a network-centric perspective, focus on the E4A and Invitel networks, and validate our classifications of their interfaces as remote. Based on the measurements, our method classifies the E4A interfaces at DE-CIX, France-IX, LoNAP, TorIX, and TlE as remote. We confirm that E4A indeed peers remotely at these 6 IXPs through private conversations as well as IXPs public information [6, 123]. Our method identifies Invitel as a remote peer at AMS-IX and DE-CIX, with RTTs of 22 and 18 ms respectively. Our private inquiries indicate that Invitel uses remote-peering services of Atrato IP Networks to reach and peer at AMS-IX and DE-CIX.

Finally, we receive an independent confirmation that our RTT measurement methodology is accurate. On our request, the TorIX staff measured minimum RTTs between the TorIX route server and member interfaces. Their results for our analyzed interfaces closely match our RTT measurements from the local PCH LG server. The mean and variance of the differences are respectively 0.3 and 1.6 ms.

2.3 Traffic offload potential

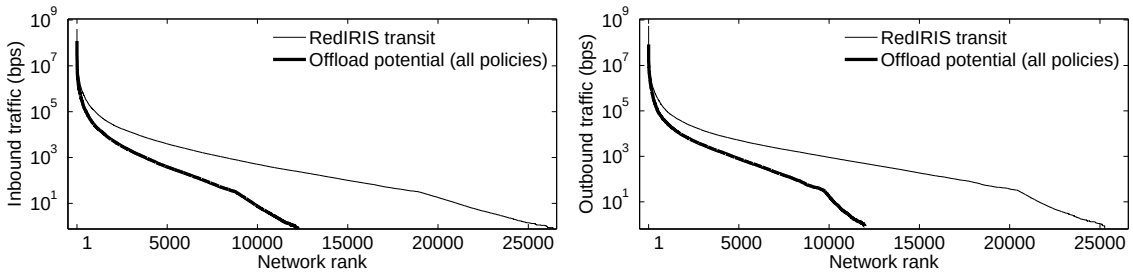
While Section 2.2 demonstrates that remote peering is widespread, we now evaluate how much transit traffic a network can offload to remote peering. Based on ground truth traffic, we estimate the traffic offload potential and study its sensitivity to peering policies.

2.3.1 Traffic data

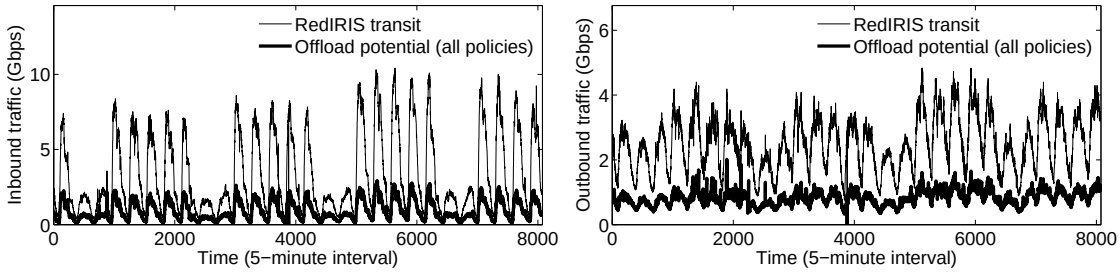
The main basis for our estimation effort is traffic in RedIRIS, the NREN (National Research and Education Network) in Spain. RedIRIS is connected to GÉANT (backbone for European NRENs), buys transit from two tier-1 providers, peers with major CDNs (Content Delivery Networks), and has memberships at two IXPs: CATNIX in Barcelona and ESpanix in Madrid. In February 2013, we used NetFlow to collect one month of traffic data at the 5-minute granularity in the ASBRs (Autonomous System Border Routers) of RedIRIS.

Our interest is limited to the traffic flows on the transit-provider links of RedIRIS. We classify each such flow as *inbound traffic* or *outbound traffic* depending on whether RedIRIS respectively receives the flow from its transit providers or sends the flow to them. The collected dataset identifies networks by their ASNs and contains records for 29,570 networks that are origins of the inbound traffic or destinations of the outbound traffic.

Utilizing the BGP routing tables in the ASBRs, we determine the AS-level path and traffic rate for each of the traffic flows. While a network can be associated with a traffic flow in the role of a traffic origin, destination, or intermediary, we classify the traffic flows associated with



(a) Network contributions to the transit traffic and offload potential



(b) Time series of the transit traffic and offload potential over the measurement period

Figure 2.5: Contributions by networks to the RedIRIS transit traffic and offload potential in scenario 4

a network as its *origin traffic* (originated in the network), *destination traffic* (terminated in the network), and *transient traffic* (passing through the network).

To illustrate the contributions of the 29,570 networks to RedIRIS' transit-provider traffic, we examine how much traffic each individual network contributes as an origin of the inbound traffic and destination of the outbound traffic. Figure 2.5a plots the average traffic rates for the respective inbound and outbound contributions by the individual networks during the measurement period. The figure ranks the networks in the decreasing order of the contributions. While a few networks make huge contributions close to the Gbps mark, most networks contribute little. In the range where the networks are ranked about 20,000 and contribute average traffic rates around 100 bps, the distributions of the inbound and outbound traffic exhibit a similar change in the qualitative profile of the decreasing individual contributions, a bend toward a faster decline. Whereas the raw data exhibit the bend as well, an explanation of the bend is an interesting question for future work. Figure 2.5b reveals daily and weekly fluctuations in RedIRIS' transit-provider traffic, which are especially clear for the inbound traffic.

2.3.2 Offload scenarios

RedIRIS cannot offload all of its transit-provider traffic. The offload potential depends on the set of IXPs that the network is able to reach via remote peering. Also, the memberships of the reached IXPs do not include all the networks that contribute to the transit-provider traffic of

RedIRIS. Finally, not all the members of the reached IXPs are likely to peer with RedIRIS.

For the set of IXPs that RedIRIS might be able to reach, we consider the Euro-IX association formed, as of February 2013, by 65 IXPs from all the continents [67]. The considered 65 IXPs are a superset of the 22 IXPs studied in Section 2.2, with the set enlargement made feasible by removing the constraint of having LG servers in the IXPs. Based on Euro-IX data from February 2013, we limit potential peers of RedIRIS to the members of these 65 IXPs.

We further trim the group of potential peers by excluding the networks that are highly unlikely to peer with RedIRIS. First, we do not consider the transit providers of RedIRIS as its potential peers because transit providers typically do not peer with their customers. It is worth noting that no network sells transit to these two tier-1 providers, and thus no such network needs to be excluded due to its transitive transit relation with RedIRIS. Second, since RedIRIS already has memberships in CATNIX and ESpanix, the other members of these two IXPs are disregarded as candidates for remote peering with RedIRIS. In particular, we exclude all the other tier-1 networks because they have memberships in ESpanix. Third, due to the cost-effective interconnectivity that comes with the GÉANT membership, we do not consider the other GÉANT members as potential peers of RedIRIS. After applying the above three rules, the group of potential remote peers of RedIRIS reduces to 2,192 networks. Even after eliminating the highly unlikely peers, there remains a significant uncertainty as to which of the 2,192 networks might actually peer with RedIRIS.

To deal with the remaining uncertainty about potential peers, we examine a range of *peer groups*, i.e., groups of networks that might peer with RedIRIS. Using PeeringDB which reports peering policies of IXP members [122, 151], we compose the following 4 peer groups so that the peering policies of their members comprise:

[peer group 1] *all open policies*;

[peer group 2] *all open and top 10 selective policies*,
which adds to peer group 1 the 10 networks that have the largest offload potentials among the networks with selective policies;

[peer group 3] *all open and selective policies*;

[peer group 4] *all policies*, i.e., all open, selective, and restrictive policies.

Peer group 4 constitutes our upper bound on the likely peers of RedIRIS. When RedIRIS reaches all the 65 IXPs, this peer group 4 includes all the aforementioned 2,192 networks. Peer group 1 represents a lower bound on the networks that might actually peer with RedIRIS. It is common for such open-policy networks to automatically peer with any interested IXP member via the IXP route server [165].

For each peer group, we determine the offload potential of RedIRIS by fully shifting to remote peering the traffic that the networks of this peer group and their customer cones contribute to the transit-provider traffic of RedIRIS. While RedIRIS is in control of its outbound transit traffic, we

assume that the networks of the peer group shift the inbound transit traffic of RedIRIS to remote peering as well.

In addition to studying sensitivity of the offload potential to the peer groups, we also evaluate its sensitivity to the choice of reached IXPs. Specifically, our evaluation varies the set of reached IXPs from a single IXP to all the 65 IXPs in the Euro-IX data.

2.3.3 Evaluating the offload potential

This section looks at the networks that contribute to RedIRIS' transit-provider traffic (as origins of the inbound traffic or destinations of the outbound traffic) and are able to shift their traffic contributions to the remote-peering links between RedIRIS and its potential peers. In scenario 4 (all policies), there are 12,238 such networks, including the 2,192 potential peers of RedIRIS. Separately for the inbound and outbound directions of the transit-provider traffic, Figure 2.5a plots the average offload potential for each of the 12,238 contributing networks, ranking them in the decreasing order of the traffic contributions. Remote peering enables RedIRIS to offload around 27% and 33% of its total inbound and outbound traffic respectively. While the inbound traffic dominates the outbound traffic during the data collection period, Figure 2.5b also shows that the peaks of the transit traffic and offload potential consistently coincide, implying that the offloading can indeed reduce the transit bill which is typically determined by the traffic peak.

Figure 2.6 zooms in on the top 30 potential peers with the largest contributions of the combined inbound and outbound offload traffic, where the contributions include origin, destination, and transient traffic. The top 30 contributors include Microsoft, Yahoo, and CDNs, suggesting that content-eyeball traffic greatly contributes to the offload potential. For a majority of the top contributors, the origin and destination traffic dominates the transient traffic.

Limiting the offload to only one IXP, we now examine the offload potential in each of the 4 scenarios. Figure 2.7 plots the results for the top 10 IXPs under this constraint, with the top 10 ranking according to the full offload potential at an IXP. The top 4 of the IXPs include the big European trio (AMS-IX, LINX, and DE-CIX) and Terremark from Miami, USA. In either of the 4 scenarios, the offload potentials at the large European IXPs remain close to each other because these IXPs have a lot of common members. On the other hand, the peering policies make a very different impact on the offload potential at Terremark. Terremark's numerous members from South and Central America [162] contribute significantly to the RedIRIS transit traffic but do not have their own presence in Europe.

To evaluate the utility of traffic offload at an additional IXP, we consider a situation where RedIRIS already realizes its full offload potential at one IXP and decides to peer remotely at another IXP. If the two IXPs have common members that contribute to the RedIRIS transit-provider traffic, the full realization of the offload potential at the first IXP reduces the amount of traffic that RedIRIS can offload at the second IXP. For scenario 4 (all peering policies), Figure 2.8 illustrates this effect when AMS-IX, LINX, DE-CIX, and Terremark act as either the first or the second IXP. When LINX and AMS-IX act as the first and second IXPs respectively, the offload

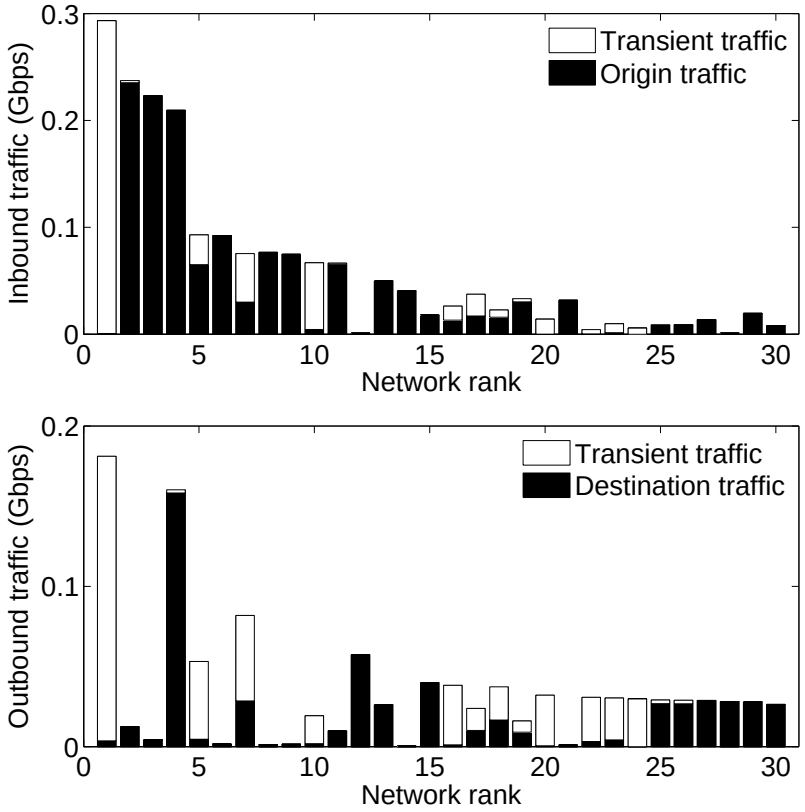


Figure 2.6: Origin and destination traffic vs. transient traffic for top contributors to the offload potential

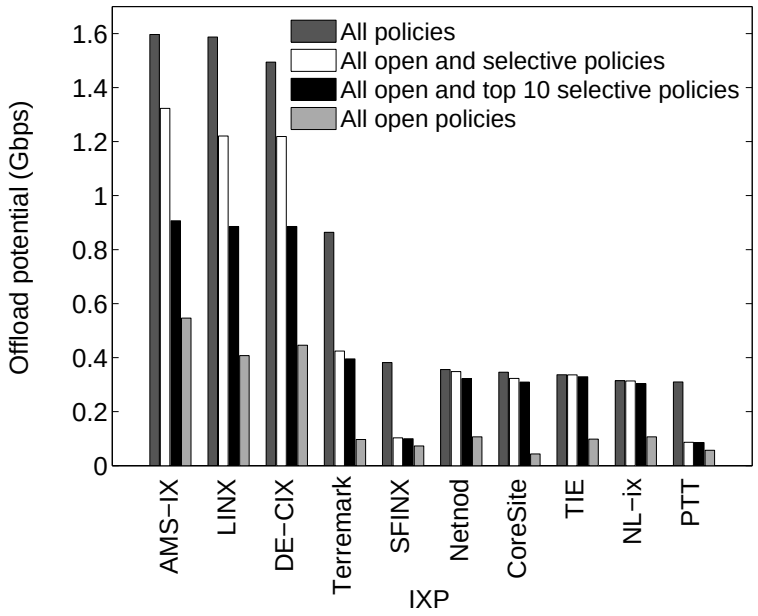


Figure 2.7: Offload potential at a single IXP

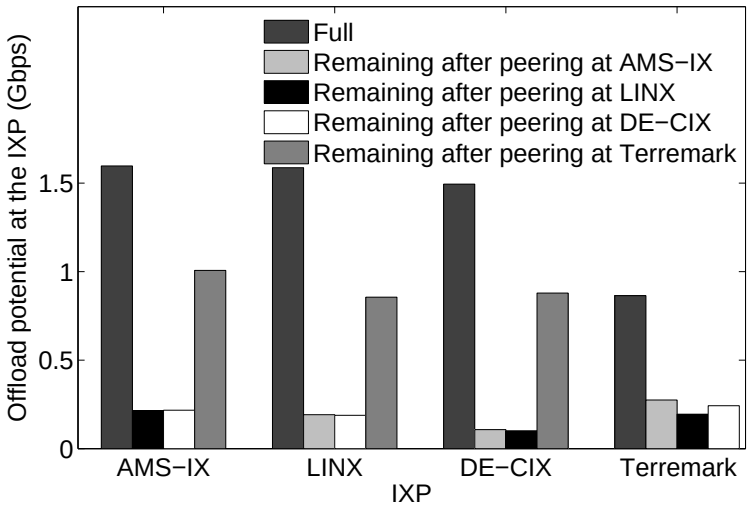


Figure 2.8: Full offload potential at an IXP and marginal utility of traffic offload at an additional IXP

potential remaining at AMS-IX is 0.2 Gbps, which is much lower than the 1.6-Gbps full potential at AMS-IX. When Terremark acts as the second IXP, the decrease in its offload potential is less pronounced because Terremark shares only about 50 of its 267 members with either of the three largest European IXPs.

Generalizing the above, we determine marginal utility of remote peering at multiple extra IXPs by expanding the peering reach of RedIRIS by one IXP at a time. The expansion is in the decreasing order of the offload potential remaining at an IXP. For example, the expansion order in scenario 4 is AMS-IX, Terremark, DE-CIX, CoreSite, ... For all the 4 scenarios, Figure 2.9 plots the remaining transit traffic of RedIRIS as a function of the IXPs where RedIRIS peers remotely. The overall reduction in transit traffic varies from 8% in scenario 1 (all open policies) to 25% in scenario 4 (all policies). Figure 2.9 shows that the marginal utility of peering at additional IXPs diminishes exponentially and that peering at a few IXPs enables RedIRIS to realize most of its total offload potential.

While the previous results are specific to RedIRIS, we now switch the metric from traffic to the number of reachable IP interfaces and present empirical results suggesting that the property of diminishing marginal utility of reaching an additional IXP holds in general. Like in the RedIRIS experiments, we expand a set of IXPs one at a time. The expansion is in the decreasing order of the number of additional IP interfaces reachable through an IXP. Figure 2.10 plots the number of IP interfaces reachable only through transit providers as the IXP set expands. With no IXPs in this set, i.e., without any peering at all, around 2.6 billion IP interfaces are reachable through the transit hierarchy. When the set includes the first IXP in scenario 4 (all policies), about 1 billion IP interfaces remain reachable only through transit providers. The marginal utility of including additional IXPs consistently declines in all 4 scenarios. Note that this perspective of reachable IP interfaces does not depend on the particulars of RedIRIS or another network. Even though the

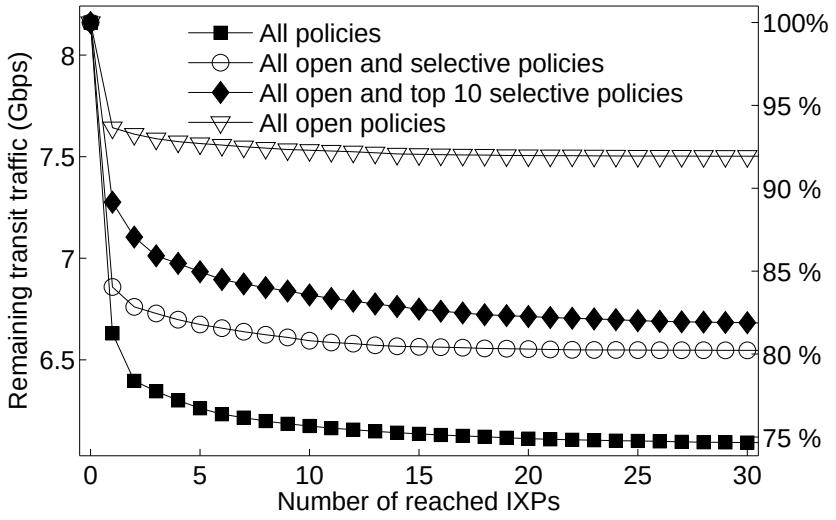


Figure 2.9: Marginal utility of RedIRIS' remote peering at multiple additional IXPs

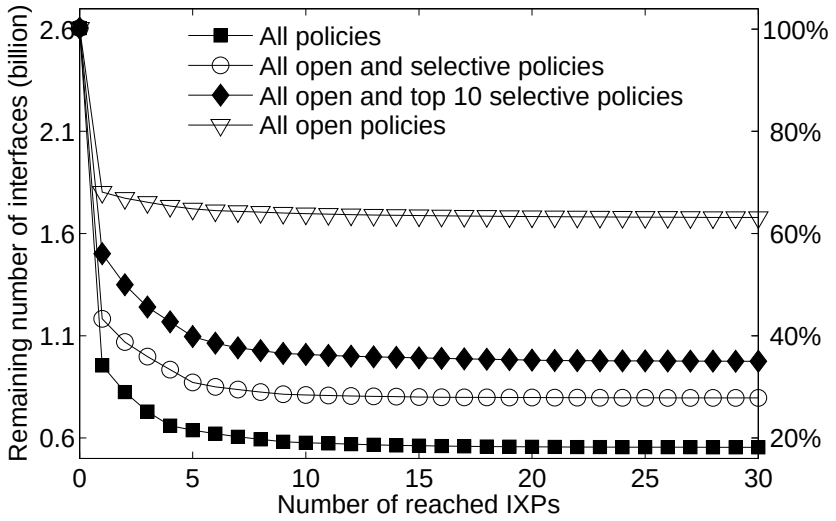


Figure 2.10: General marginal utility (measured in reachable IP interfaces) of reaching multiple additional IXPs

declining profiles in figures 2.9 and 2.10 are quantitatively different, their qualitative similarity indicates that diminishing marginal utility of reaching one more IXP is a general property.

2.4 Economic viability

While Section 2.3 showed significant traffic offload potential and diminishing marginal utility of remote peering at additional IXPs, we now generalize the empirical results in the mathematical model and derive conditions for economic viability of remote peering versus transit and direct peering.

2.4.1 Model

In our model, a network delivers its global traffic via 3 options: (1) transit, (2) expansion of its own infrastructure for direct peering at n IXPs, and (3) remote peering at m IXPs. The respective traffic fractions are denoted as t, d, and r:

$$t + d + r = 1. \tag{2.1}$$

Total cost C of the traffic delivery consists of transit, direct-peering, and remote-peering components C_t, C_d, and C_r:

$$C = C_t + C_d + C_r. \tag{2.2}$$

While Section 2.3.3 shows diminishing marginal utility of reaching one more IXP, we fit the RedIRIS data to exponential decay and model the transit traffic fraction as the following function of the number of IXPs where the network peers either directly or remotely:

$$t = e^{-b\cdot(n+m)}. \tag{2.3}$$

Equation 2.3 generalizes our empirical results via parameter b that controls how quickly the transit traffic fraction declines. While $b = 0$ represents networks that cannot reduce its transit traffic by peering at distant IXPs, $b = \infty$ enables offload of all transit traffic by reaching a single IXP. Low values of b are characteristic for networks with mostly global traffic, e.g., Google and other networks with highly distributed traffic. The results in Figure 2.4a suggest that networks with high b values are more common. With parameter p denoting the normalized transit price, we model the transit cost as

$$C_t = p \cdot t = p \cdot e^{-b\cdot(n+m)}. \tag{2.4}$$

The direct-peering cost depends on both the number of the reached IXPs and traffic delivered through them:

$$C_d = g \cdot n + u \cdot d. \tag{2.5}$$

While parameter g accounts for membership fees and other traffic-independent costs of the network in the distant IXPs, parameter u reflects traffic-dependent costs.

The remote-peering cost has a similar structure with traffic-independent and traffic-dependent parts:

$$C_r = h \cdot m + v \cdot r. \tag{2.6}$$

The per-IXP traffic-independent cost for remote peering is typically lower than for direct peering:

$$h < g, \tag{2.7}$$

and the per-unit traffic-dependent cost for remote peering is larger than for direct peering but

smaller than for transit:

$$u < v < p. \tag{2.8}$$

Combining equalities 2.2, 2.4, 2.5, and 2.6, we express the total traffic-delivery cost of the network as

$$C = p \cdot e^{-b\cdot(n+m)} + g \cdot n + u \cdot d + h \cdot m + v \cdot r. \tag{2.9}$$

2.4.2 Analysis

Seeking to minimize its total cost, the network might first consider only transit and direct peering at distant IXPs without purchase of remote peering, i.e., $m = 0$ and $r = 0$. Under this strategy, the total cost is

$$C = (p - u) \cdot e^{-b\cdot n} + u + g \cdot n, \tag{2.10}$$

and the network minimizes the cost by reaching $\tilde{n}$ IXPs to offload traffic fraction $\tilde{d}$ via direct peering:

$$\tilde{n} = \frac{\log\left(\frac{b\cdot(p-u)}{g}\right)}{b} \text{ and } \tilde{d} = 1 - e^{-b\cdot\tilde{n}}. \tag{2.11}$$

Continuing from the above solution, the network might widen its strategy to include remote peering, with the total cost becoming

$$\begin{aligned} C \ &= (p - v) \cdot e^{-b\cdot(\tilde{n}+m)} \\ &\quad + (v - u) \cdot e^{-b\cdot\tilde{n}} + g \cdot \tilde{n} + u + h \cdot m. \end{aligned} \tag{2.12}$$

The network minimizes the cost in equality 2.12 by remote peering at $\tilde{m}$ extra IXPs:

$$\tilde{m} = \frac{\log\left(\frac{g\cdot(p-v)}{h\cdot(p-u)}\right)}{b}, \tag{2.13}$$

Inequality $\tilde{m} \geq 1$ means that remote peering at one or more IXPs reduces the total cost. Thus, we establish the following condition for economical viability of remote peering:

$$\frac{g \cdot (p - v)}{h \cdot (p - u)} \geq e^{b}. \tag{2.14}$$

Our analysis shows that remote peering is more viable economically for networks with lower b values, i.e., networks carrying global traffic. Networks carrying big volumes of traffic such as Google or other large content providers do have an incentive to be present in many IXPs. However the large volumes of traffic carried justify the costs of their physical presence at multiple IXPs. Large content providers have previously underpinned the observed trend towards a flatter Internet. Differently, remote peering enables smaller networks with global traffic to benefit from peering even if the smaller volumes of traffic do not justify the costs of physical peering. Remote peering gives such networks a cost-effective alternative to the costly expansion of their own network

infrastructures and alters the trend towards a flatter Internet by introducing an intermediary in peering. For instance, networks such as Invitel, E4A or LinkedIn carry geographically distributed but not the volumes of traffic that would justify physical peering at a large number of IXPs. Instead, these networks rely on remote peering providers to extend its presence. For instance E4A peers at 9 IXPs, at 6 of them remotely.

The economical viability condition contains g/h, i.e., the ratio of the per-IXP traffic-independent costs for direct and remote peering. In regions such as Africa, h tends to be much smaller than g because local IXPs offer little opportunities to offload traffic, and transit is expensive [69, 87]. Thus, our analytical model explains why remote peering is economically attractive for African networks.

2.5 Discussion

The emergence of remote peering exposes that increasing peering and Internet flattening are separate trends. The two trends are typically conflated because the expansion of content-network infrastructures bypasses transit providers through physical peering reducing the number of intermediaries in their connections. Complementing direct peering, remote peering enables additional peering as well. However, this increase in peering involves a remote-peering provider that acts a middleman. Furthermore, the intermediary that sells the remote-peering service can be a traditional transit provider. Hence, remote peering increases peering without flattening the Internet economic structure.

The observed separation of the two trends questions the usage of AS-level topologies for representing the Internet economic structure. With remote-peering services provided on layer 2, layer-3 models of the Internet structure omit the remote-peering providers and fail to distinguish remote peering from direct peering and hence ignore the presence of remote peering provider intermediaries. Below, we elaborate on various dangers posed by this omission of the intermediary economic entities.

While it is common to use AS-level models for reasoning about Internet security, the hidden presence of layer-2 intermediaries undermines security assurances. The invisible intermediaries might be unwanted entities, e.g., those associated with problematic governments. The risks include monitoring or alteration of traffic by the intermediaries and exposure of traffic to other parties, e.g., by delivering it through undesired geographies.

The reliance on layer-3 models also compromises accountability. Whereas a layer-2 intermediary might delay or discard traffic, attribution of responsibility for such performance disruptions is complicated because the middleman is invisible on layer 3.

Layer-3 topologies can make the Internet structure look more reliable than it is. When a company employs the same physical infrastructure to provide transit and remote-peering services, buying both might not translate the redundancy into higher reliability for the multihomed customers.

Because remote peering has different economics than transit and direct peering, the omission of the layer-2 intermediaries from layer-3 models weakens the economic understanding of the Internet. In developing markets such as Africa, remote peering becomes a cost-effective alternative for reaching well-connected areas in Europe and North America [87]. Since remote peering has a smaller connectivity scope than transit, adoption of remote peering necessitates new strategies for traffic distribution. IXPs greatly benefit from remote peering: existing IXPs gain members, and new IXPs are enabled by bringing together a critical mass of traffic [119]. Ignoring the remote-peering providers distorts substantially the Internet economic landscape.

Thus, our results call for alternative models of the Internet structure that explicitly represent layer-2 entities. The relevant additions include not only remote-peering providers but also other layer-2 economic entities such as IXPs. With the growing prominence of IXPs and remote-peering connectivity to them, integrated modeling of the Internet structure on layers 2 and 3 becomes increasingly important for understanding the Internet.

The refined mapping of the Internet economic structure on layers 2 and 3 will likely require novel methods for inference of economic entities and their relationships. These methodological innovations constitute an interesting direction for future work.

2.6 Conclusion

This chapter presented the first empirical and analytical study on remote peering. Using careful measurements of RTTs at 22 IXPs worldwide, our ping-based method exposed wide spread of remote peering, with remote peering in more than 90% of the examined IXPs and peering on the intercontinental scale in a majority of them. Based on real traffic in RedIRIS, we also estimated how much transit traffic the network can offload via remote peering at 65 IXPs. The assessment showed significant traffic offload potential, around 25% in some cases, and exhibited diminishing marginal utility of remote peering at additional IXPs. After generalizing this property in the mathematical model, we derived conditions for economic viability of remote peering versus transit and direct peering.

While important in itself as an emerging factor in the Internet ecosystem, remote peering was showed to have broader implications for Internet research. Remote peering revealed separation between the trends of increasing peering and Internet flattening which had been commonly conflated. With remote peering provided on layer 2, the omission of remote-peering providers from traditional layer-3 representations of the Internet topology compromises research on Internet security, accountability, reliability, and economics. We called for refined modeling of the Internet economic structure on both layers 2 and 3.

Chapter 3

Cooperative IP Transit (CIPT)

The previous chapter showed how the rapidly evolving Internet ecosystem generates new types of interconnections, such as remote peering, to cope with new needs and challenges. Continuing this trend toward greater diversity of Internet interconnections, this chapter proposes a novel arrangement that address the problem of high transit costs.

The Internet ecosystem involves thousands of ASes linked in a more or less hierarchical manner to support universal connectivity of Internet users. While a small number of ASes can reach the entire Internet without paying any other network for the reachability, the immense majority of ASes need to pay transit providers to attain global reachability.

In spite of the steady decline of IP transit prices, the IP transit costs remain high due to the traffic growth. Over the previous decades a number of solutions have been suggested to reduce these IP transit costs by reducing the volume of billed transit traffic. Proposed solutions reduce the transit traffic volume and include settlement-free [30, 56, 78] or paid peering [58], IXPs [15, 55], IP multicast [18, 28, 54, 82, 83], CDNs [146], P2P localization [49] and traffic smoothing [63, 101, 114].

This chapter proposes *CIPT (Cooperative IP Transit)*, a different approach to reducing the cost of IP transit. Instead of altering the traffic that flows through the transit links, CIPT reduces the price of transit per Mbps: by jointly purchasing the IP transit, two or more ASes reduce the transit prices per Mbps for each AS involved in the CIPT.

Tuangou[1] (group buying) has been highly successful in other domains [117]. While tuangou succeeds primarily due to subadditivity of prices [37, 38, 186], the benefits of CIPT depend also on burstable billing [61], different methods to account for bidirectional traffic, and other complex factors.

To illustrate the CIPT concept, Figure 3.1 plots real traffic profiles of three ASes. If the ASes form a CIPT and purchase transit jointly, the total cost is smaller than when purchasing it separately because: (a) Transit billing is burstable, i.e., the buyer is billed for the peak of its traffic; because the peaks of the traffic of the ASes are not completely coincident, the peak of the combined traffic is smaller than the sum of the separate peaks; (b) Transit prices are subadditive, i.e., prices *per Mbps* decrease as the purchased amount increases. Consequently, the traffic aggregation enables CIPT to reduce costs of its partners.

The novelty of CIPT in the Internet ecosystem lies in the cooperative essence of the arrangement. While CIPT reduces costs by transit traffic aggregation, the latter is common in the Internet. Most transit providers are transit resellers that profit from lower rates resulting from transit aggregation. As early as in 1990s, The Little Garden (TLG) [187] pooled traffic of small customers together to obtain cheaper transit rates. Government-promoted IXPs also lower transit costs through national transit traffic aggregation, e.g., in Bahrain [16, 161] and other developing countries [4, 5]. CIPT is substantially different from previous transit-aggregation schemes in the following aspects:

[1] Tuangou (pronounced "twangoo"), a term originating in China, loosely translates as group buying, http://en.wikipedia.org/wiki/Tuangou.

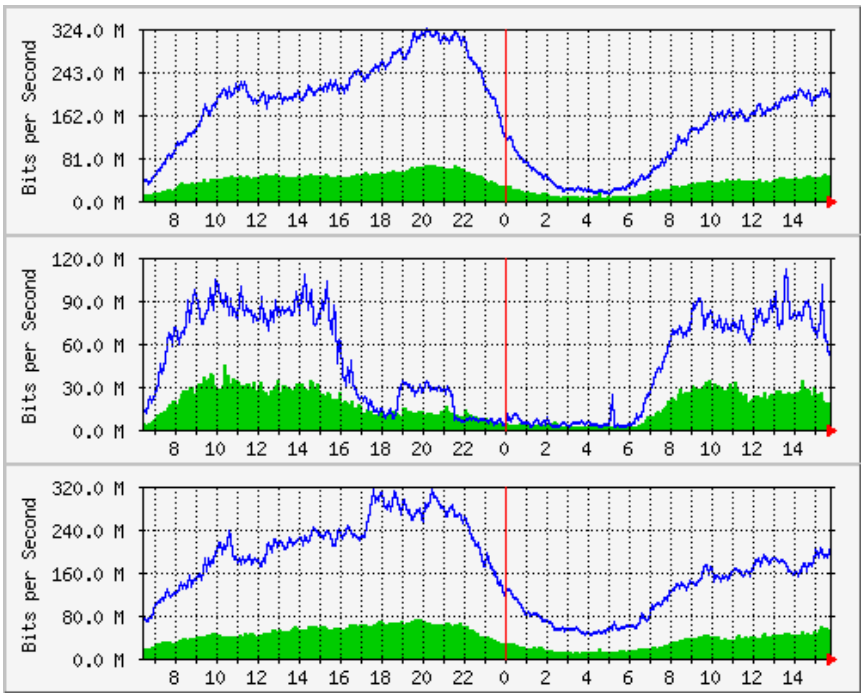


Figure 3.1: Demand statistics for partners P_1 (top), P_2 (middle), and P_3 (bottom) in the motivating example: the x-axes are in hours; the y-axes are in $Mbps$; the filled (green) areas depict the upstream traffic; the (blue) lines represent the downstream traffic.

- Cooperation of transit buyers;
- Mechanisms for distribution of the benefits of transit aggregation.

Beyond the new application of tuangou in the domain of IP transit, a major contribution of this chapter lies in its measurement and evaluation methods. Relying on real inter-domain traffic and transit pricing, this chapter estimates the gains from CIPT. We also propose Shapley value as a basis for sharing the gains among the CIPT partners so that to provide each partner with a strong economic incentive for the cooperation. Our evaluation of the aggregate and individual gains involves collection of the visual traffic statistics from 6 public IXPs with 264 participating ASes, transformation of the visual images into a numeric format, and public-data validation of the property that peering and transit traffic have similar temporal profiles. Our analysis suggests that the expected relative savings of CIPT are in the range of 8-56% for the IXP-wide coalitions; in absolute terms, each of the partners may expect annualized savings from one thousand US$ for very small ASes to several hundred thousand US$ for the few large ASes. We also show that much smaller coalitions, with a half a dozen of members, can offer close-to-maximum savings. The main contributions of this chapter are as follows:

- We propose CIPT, a simple cooperative strategy to reduce costs by purchasing IP

transit jointly.

- We show that CIPT can be modeled as a cooperative game and that Shapley value provides an intuitive mechanism for cost sharing in CIPT.

- We use public IXP data to infer the traffic time series for several hundred (mostly regional and national) ASes and use this information to assess the potential cost benefits of CIPT.

While our results on the CIPT cost reduction validate the potential of CIPT to become a new viable element of the Internet ecosystem, the practical viability of CIPT also depends on other strategic and organizational issues. For example, if two ASes are already engaged in a transit relationship, they are unlikely to agree on buying IP transit jointly from a third party. Also, the transit provider can strategically respond to CIPT by charging the coalition at higher prices per Mbps than the prices offered to an individual AS. On the other hand, big transit providers might strategically accept CIPT to squeeze out smaller transit providers. By aggregating transit traffic, CIPT might become an attractive customer for large transit providers bypassing transit resellers. It is quite possible that CIPT will not grow into the dominant mechanism for IP transit cost reduction. On the other hand, earlier success in cost reduction via transit aggregation [4,5,16,187] suggest that CIPT is certainly feasible and can gain a broad presence in the Internet ecosystem, from small websites in a hosting facility to the level of nation-wide ASes. Data-driven assessment of all these additional issues lies beyond the scope of this thesis. Similarly, while we propose Shapley value as a means for cost sharing in CIPT, evaluation of alternative solutions to CIPT cost sharing is a topic for future work.

3.1 Background and motivation

The geographic location affects significantly the cost of IP transit [135]. The IP transit prices per $Mbps$ per month range usually from \$5 to \$100 (we use \$, US\$ or USD to refer to U.S. dollars throughout the rest of this thesis): the wholesale IP transit is typically priced under \$10 per $Mbps$ in most European and North American hubs but can exceed \$100 per $Mbps$ in Australia, Latin America and other remote regions of the Internet [5, 186].

Regardless of the geographic location, IP transit is subject to economies of scale and is priced subadditively: the prices per $Mbps$ are smaller for larger quantities of IP transit [186]. Table 3.1 presents real (as of January 2011) transit pricing rates of Voxel, a middle-size transit provider in North America. The table reports the prices for different levels of Committed Data Rate (CDR), the minimum amount charged by the provider. For example, an AS with IP transit needs of 300 $Mbps$ commits at the 100-$Mbps$ CDR level and pays pro rata \$3000 to the transit provider, but an AS with IP transit needs of 700 $Mbps$ finds it more cost-effective to commit at the 1000-$Mbps$ CDR level and pays \$5000.

Committed Data Rate, $Mbps$	Price per $Mbps$ per month
10	\$25
50	\$15
100	\$10
1000	\$5
10000	\$4

Table 3.1: IP transit pricing rates

Burstable billing is another important aspect of IP transit pricing [61, 134]. To calculate the IP transit cost, the most commonly used method is to calculate the $peak$ usage (typically through the 95th-percentile rule [61,134]) and then the price function f is applied to the observed $peak$ to calculate the resulting payment. The $peak$ value is usually calculated separately for the upstream and downstream directions, and either sum or maximum of the two is used for billing. We refer to these two pricing models as sum and max models. Intuitively, the max model offers a larger opportunity for savings in cooperation because two ASes with their traffic peaks in opposite directions can mutually benefit from the less utilized directions of each other. Consequently, results for the sum model can be considered as a conservative estimate of CIPT gains (Figure 3.6 in Section 3.4.1.2) confirms this intuition).

To illustrate the potential of CIPT, we consider a simple scenario of three partners[2] P_1, P_2, and P_3 interested in purchasing IP transit from the same provider. We assume the transit pricing rates as in Table 3.1, 95th-percentile burstable billing, sum model of accounting for bidirectional traffic, and traffic profiles plotted in Figure 3.1.

If the three partners purchase the IP transit separately, the individual traffic peaks (computed as the sum of the peaks in both directions) of P_1, P_2, and P_3 are at 379 $Mbps$, 130 $Mbps$, and 362 $Mbps$ respectively, and each of the partners commits at the 100-$Mbps$ CDR level. Thus, partners P_1, P_2, and P_3 pay respectively \$3790, \$1300, and \$3620 with the aggregate transit cost of \$8710.

On the other hand, if P_1, P_2 and P_3 use CIPT to buy the IP transit together, their aggregate peak traffic is 712 $Mbps$. By committing at the 1000-$Mbps$ CDR level, the CIPT pays \$5000. Thus, the cooperation reduces the aggregate transit cost of the partners by \$3710, or 43%. This significant cost reduction comes from two different sources:

1. Burstable billing – the 712-$Mbps$ peak of the aggregate traffic is lower than the 871-$Mbps$ sum of the individual traffic peaks; hence, the aggregate transit cost would decrease even if the pricing function were additive;

2. Subadditive pricing – the upgrade from the 100-$Mbps$ CDR level to the 1000-$Mbps$ one provides a lower price per $Mbps$ and thereby reduces the aggregate transit cost even further.

[2]We interchangeably use terms partner and player to refer to any AS, hosting provider or any other entity interested in purchasing IP transit.

3.2 Cooperative IP transit

While we have already sketched the main idea of the CIPT, this section provides more details, describes the concept of cooperative (or coalitional) games, and models CIPT as a cooperative game.

Cooperative IP Transit (CIPT) refers to any cooperative mechanism in which two or more subjects purchase the IP transit jointly as a means for cost reduction. The subject interested in CIPT can be any Internet entity that buys IP transit; such entities include websites and hosting providers, as well as access, nonprofit, and content ASes.

The main incentive for forming a CIPT coalition is financial: each partner reduces its individual IP transit bill. The typical IP transit pricing makes it virtually impossible for a set of potential partners to increase their aggregate transit cost by buying the IP transit jointly. However, CIPT needs a reasonable mechanism to distribute the aggregate cost savings among all the CIPT partners. Furthermore, the aggregate and individual IP transit costs of the CIPT partners strongly depend on a number of factors such as the IP transit pricing function, number of partners, their size, and temporal patterns of their traffic demands.

3.2.1 CIPT as a cooperative game

Formally, a cooperative game is characterized by set $\mathcal{N}$ of involved players and a cost function that maps the partitive[3] set of $\mathcal{N}$ to a cost value: $c : 2^{\mathcal{N}} \rightarrow R$. In the context of CIPT, set $\mathcal{N}$ is the set of subjects interested in purchasing IP transit. The cost function maps an arbitrary subset $S \subset \mathcal{N}$ to the cost of the IP transit that the coalition of players from S would pay. An important property of the IP transit model is that the price per $Mbps$ is a non-increasing function of the $peak$, due to the subadditive nature of the pricing model.

CIPT is formed by a set of N partners. Each partner i of the CIPT has upstream and downstream IP transit traffic demands represented respectively by time series $u_i(t)$ and $d_i(t)$ where $i \in \{1, 2, \ldots, N\}$, and time t is measured in fixed-size time intervals with a typical interval duration of 5 minutes. The cost that subject i pays for the transit, without participation in CIPT, is the function of these demand series:

$$C_i = F(u_i(\cdot), d_i(\cdot)).$$

After bundling of N subjects, the aggregate upstream/downstream demands are the sum of the corresponding individual demands:

$$u(t) = \sum_{i=1}^{N} u_i(t) \text{ and } d(t) = \sum_{i=1}^{N} d_i(t),$$

[3]For set $\mathcal{N}$, the partitive set of $\mathcal{N}$ is the set of all subsets of $\mathcal{N}$ and is usually denoted as $2^{\mathcal{N}}$.

and the aggregate cost of the IP transit is

$$C = F(u(\cdot), d(\cdot)).$$

The 95th-percentiles of the upstream ($peak^{(up)}$) and downstream ($peak^{(down)}$) traffic are calculated, and the $peak$ value used for billing is either the sum or max of these two values, as described in Section 3.1. The transit cost of the coalition of these N players is then

$$C = F(u(\cdot), d(\cdot)) = f(peak)$$

where f is the pricing function decided by the IP transit provider. This pricing function is typically subadditive, Table 3.1 provides an example of such pricing function.

Additionally, for virtually any real-world subjects interested in purchasing IP transit, the peak traffic of the union of two subjects is smaller than the sum of the peaks of these two subjects. In case of measuring the peak as the maximal traffic, this is an obvious consequence of the fact that the maximum of the sum of two nonnegative functions (over the same domain) is not greater than the sum of the maximums of these two functions. If the peak is measured through the 95th-percentile method, there may be some irregular cases[4] in which the sum of the 95th-percentiles is smaller than the 95th-percentile of the union of the traffic of the two subjects. Nevertheless, these situations are extremely unlikely to happen in regular setups. We demonstrate this in Figure 3.2 by plotting the cumulative distribution for the ratio of the 95th-percentile of the union to the sum of the 95th-percentiles across *all* the pairs of ASes at Budapest Internet Exchange (BIX) in both sum and max models. BIX and several other IXPs publish traffic statistics that each of their members (mostly regional ASes) exchanges at the IXP, and this information represents valuable and useful proxy for estimating the traffic patterns (volume, peak-hour, peak-to-valley ratio, up/downstream traffic ratio, etc.) for the involved ASes.

Observation 1. The traffic patterns of subjects interested in CIPT are such that for (almost) all pairs of coalitions S_1 and S_2 of these subjects, the peak value of the union of the two coalitions is smaller than the sum of the peak values of these two coalitions.

As we elaborate above, Observation 1 is intuitive and can be empirically validated for available data of traffic patterns. From now on, we assume that subjects involved in CIPT are such that this observation is true. In that case, cost function $c(\cdot)$ is indeed subadditive:

$$c(S_1) + c(S_2) \geq c(S_1 \cup S_2), \text{ for any } S_1, S_2 \subset \mathcal{N}. \tag{3.1}$$

Hereby, virtually always the overall IP transit cost of CIPT is strictly smaller than the sum of

[4]For example, two subjects consuming 100 $Mbps$ 4% of the time each, one in the morning the other over night, and using 1 $Mbps$ the remaining 96% of the time will have their 95th-percentile equal to 1 $Mbps$, while their union would have 95th-percentile equal to 100 $Mbps$.

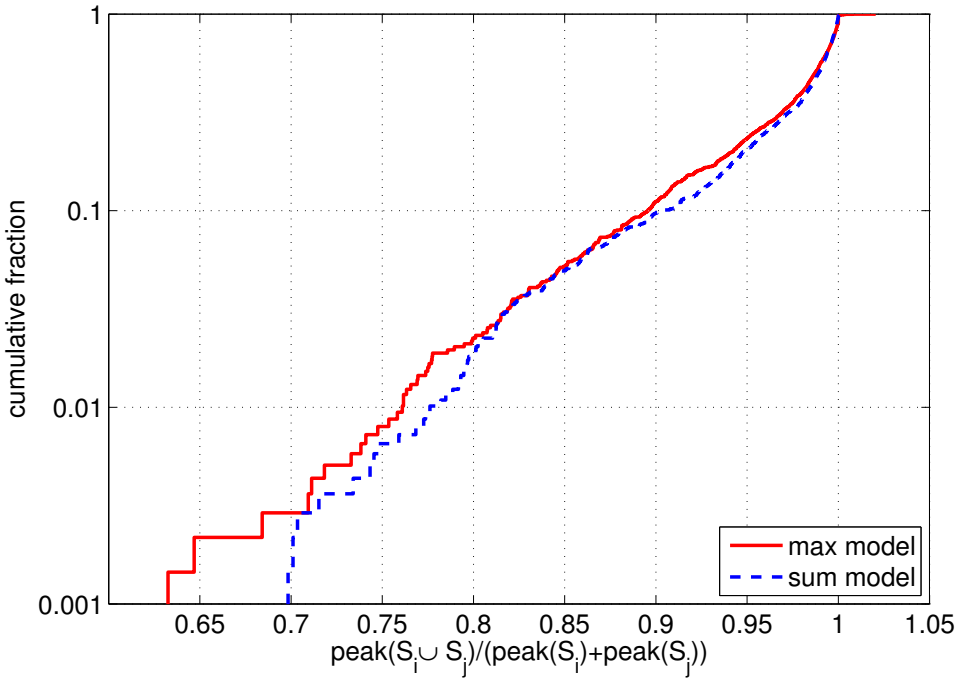


Figure 3.2: The distributions for the ratio of the 95th-percentile of the union to the sum of the 95th-percentiles across all the pairs of ASes at Budapest Internet Exchange

individual IP transit costs of all involved players:

$$\rho = \frac{C}{\sum_{i=1}^{N} C_i} < 1.$$

The relative savings $(1 - \rho)$ of the CIPT are influenced by several factors, with the two dominant being: (1) the subadditivity of the price function and (2) burstable billing through the 95th-percentile method. Namely, the subadditive pricing leads to savings for the involved players because prices (per $Mbps$) are lower for larger quantities. Additionally, with the burstable billing, when two or more players have non-overlapping peak hours, their coalition would have the peak value strictly smaller than the sum of the peak values of the involved players. While players that serve similar user bases have similar temporal usage patterns (e.g., residential networks peak in evening hours, government/academic networks peak in early afternoon), the networks of different types experience their peaks in times that are far apart, which in turns allows for additional savings in top of bundling and buying-in-bulk.

3.3 Cost sharing in CIPT

A key question in any cooperation scheme created for cost reduction reasons is how to split the aggregate costs of cooperation. As we saw in Section 3.2.1, CIPT can be abstracted as a

cooperative game which puts us in a position to use the rich set of analytic tools for solving the problem of cost sharing. There are many solution concepts for cost sharing in cooperative games, including the core, kernel, nucleolus, and Shapley value [199]. While other solution concepts have attractive features, in the context of CIPT we find particularly appealing to use the Shapley value since it has several distinct important properties, i.e., the Shapley value: (1) *exists* for any cooperative game and is *uniquely* determined, (2) satisfies basic *fairness* postulates [173, 199], and (3) is *individually rational*, i.e., each player in CIPT receives a lower Shapley value cost than what it would be if it did not participate in CIPT. One potential deficiency of the Shapley value is that in general it is computationally hard to calculate it exactly. However, state-of-the-art techniques provide simple and accurate methods for Shapley value approximation, as discussed in Section 3.3.2.

3.3.1 Shapley value: definition

For a cooperative game defined over set $\mathcal{N}$ of N players and each subset (coalition) $S \subset \mathcal{N}$, let $c(S)$ be the cost of coalition S. Thus, if coalition S of players agrees to cooperate, then $c(S)$ determines the total cost for this coalition.

For given cooperative game $(\mathcal{N}, c(\cdot))$, the Shapley value is a (unique) vector $(\phi_1(c), \ldots, \phi_N(c))$ defined below, for sharing the cost $c(\mathcal{N})$ that exhibits the coalition of all players. It is a "fair" cost allocation in that it satisfies four intuitive properties: efficiency, symmetry, additivity and null-player; see [173, 199] for exact definitions of these properties and more details. The Shapley value of player i is precisely equal to i's expected marginal contribution if the players join the coalition one at a time, in a uniformly random order. Formally it is determined by:

$$\phi_i(c) = \frac{1}{N!} \sum_{\pi \in S_N} \left(c(S(\pi, i)) - c(S(\pi, i) \setminus i) \right) \tag{3.2}$$

where the sum is taken across all permutations (or arrival orders), π, of set $\mathcal{N}$ and $S(\pi, i)$ is the set of players arrived in the system not later than i. In other words, player i is responsible for its marginal contribution $c(S(\pi, i)) - c(S(\pi, i) \setminus i)$ averaged across all $N!$ arrival orders π. Note that the Shapley value defined by Equation 3.2 indeed satisfies the *efficiency* property:

$$\sum_{i \in \mathcal{N}} \phi_i(c) = c(\mathcal{N}).$$

3.3.2 Estimation of the Shapley value in CIPT

While the Shapley value can be computed in a rather straightforward manner using Equation 3.2, it is not practically feasible to employ Equation 3.2 for $N > 30$. A number of methods have been suggested for accurate estimation of the Shapley value, and in this chapter we use a simple Monte Carlo method [118] as follows.

Instead of calculating the exact Shapley value as the average cost contribution across all $N!$

arrival orders, we estimate the Shapley value as the average cost contribution over set Π_k of K randomly sampled arrival orders:

$$\hat{\phi}_i(c) = \frac{1}{K} \sum_{\pi \in \Pi_K} (c(S(\pi, i)) - c(S(\pi, i) \setminus i)) \tag{3.3}$$

Parameter K determines the error between the real Shapley value and its estimate: the higher K the lower the error. So basically, one can control the accuracy of the estimator by increasing the number of sample permutation orders. We observe in our datasets of traffic demands that the value of $K = 1000$ provides errors of under 1% across all the CIPT players, and in the rest of this chapter we use $K = 1000$ for the computation of the Shapley value.

3.4 Evaluation

In this section we quantify various factors that impact CIPT by using traffic information from 264 (mainly national and regional) ASes. In Section 3.4.1, we describe the dataset and pricing model(s). In Section 3.4.2, we evaluate the potential savings of CIPT on country-wide (IXP-wide) collaborations and show that significant savings could be expected both in relative and absolute terms. In Section 3.4.3, we augment this analysis by empirically showing that even small single-digit coalitions can yield close-to-optimal savings, by demonstrating a law of diminishing returns for the savings as a function of the coalition size. Section 3.4.4 analyzes the per-player savings and shows somewhat expectable trends that the larger the player is, the larger are its absolute savings, but the smaller its relative savings are. Finally, in Section 3.4.5, we analyze the effects of collaboration between geo-diverse players and present an analytical upper bound on the savings as a function of the time difference in their peak-hour periods.

3.4.1 Dataset description

Although data for the traffic patterns of many ASes is often kept confidential, some public IXPs report upstream and downstream demand time series for the traffic exchanged by every member of the IXPs. Those that do it are listed in Table 3.2. This traffic statistics data is typically given in the form of mrtg images [188], similar to those shown in Figure 3.1. Overall we collected the information for 264 ASes, with the traffic peak distribution as shown in Figure 3.3. While the information about the traffic exchanged at the public IXPs is obviously a valuable piece of information, it is not straightforward how to use this information to estimate the transit usage of the ASes. In Section 3.4.1.2, we use a small set of ASes that make their detailed traffic information public, to show that the IXP related traffic is a good proxy for estimating the transit part of the interdomain traffic, at least for some ASes. Before that, we elaborate on the data collection below.

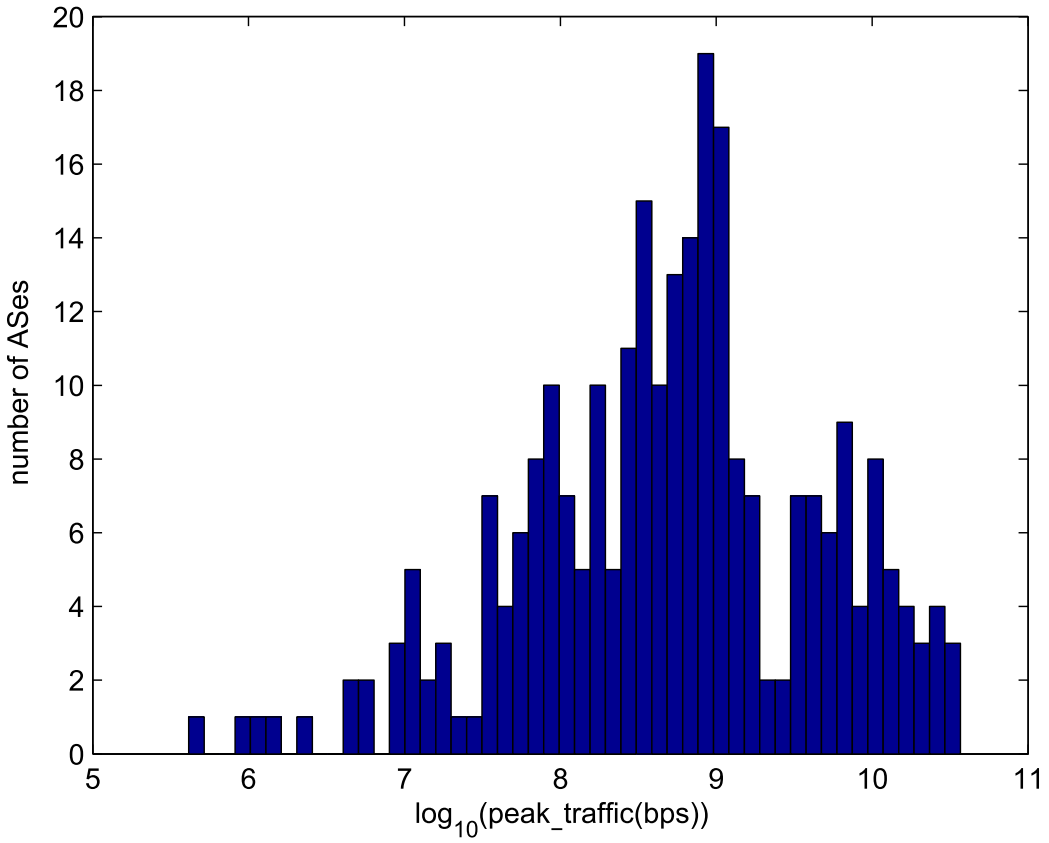


Figure 3.3: The distribution of the peak traffic rates across all 264 ASes: median: 560 $Mbps$; mean: 2.9 $Gbps$

IXP acronym	members (# of)	peak ($Gbps$)	average ($Gbps$)	95th-pct effect		subadditive effect		skewness
				sum	max	sum	max	
NIX	54	116	76	4.3%	29.1%	95.7%	70.9%	0.76
SIX	52	42	23	15.4%	44.9%	84.6%	55.1%	0.27
IIX	17	2.1	1.38	14.3%	40.6%	85.7%	59.4%	0
FICIX	25	32	19	6.7%	23.1%	93.3%	76.9%	0.48
InterLAN	63	22	11	14.3%	37.8%	85.7%	62.2%	0.12
BIX	53	152	92	3.6%	27.8%	96.4%	72.2%	0.84

Table 3.2: Basic stats on the used IXPs

3.4.1.1 Dataset collection

We started by manually inspecting the webpages of medium-size and large IXPs at [67]. A majority of these IXPs publish their aggregate traffic statistics, summed across all the members, but some also make public the detailed traffic statistics of their members. We identified several IXPs that do so. Table 3.2 lists them. We then crawled the websites of these IXPs and collected

per-member traffic information. This per-member traffic data is typically given in the form of visual images, similar to those in Figure 3.1, produced as the outputs of the standard tools for traffic visualization: mrtg/rrdtool [188]. To convert the information into a numeric form, we built a piece of software that takes as input a mrtg/rrdtool image and outputs the numeric array representing the upstream/downstream traffic time series. This operation of transforming the .png images to numeric data required serious effort in the domain of optical character and function recognition.

3.4.1.2 From IXP data to IP transit traffic

Most ASes consider the data of their networks as confidential and are reluctant to share it with third parties. However, some ASes publicly share large amounts of operational information. In particular, several European ASes serving academic institutions have publicly shared on their websites detailed pictures of both their network infrastructure and utilization of their networks. Those that we identified are HEANET (Ireland) [95], SANET (Slovak Republic) [168], CESNET (Czech Republic) [39], GRNET (Greece) [86]. We inspected the peering and transit traffic for those four ASes and found that the peering traffic pattern is a good first-order indicator of the transit traffic. In those 4 ASes, peering corresponds to 35-40% of the total traffic, with the remaining 60-65% being transit. Additionally, we observe that peering and transit traffic follow very similar temporal patterns: their growth and decay periods coincide, they peak at the same time, have similar peak-to-valley ratios, etc. In some sense, such behavior is not very surprising: given that the demand is predominantly created by humans, both transit and peering traffic demand are driven by the same end-user activities.

AS	$sim(T_{up}, P_{up})$	$sim(T_{down}, P_{down})$
HEANET	0.988	0.965
SANET	0.996	0.991

Table 3.3: The $cosine$-similarity between the transit (T) and peering (P) time series (both downstream and upstream directions)

The empirical evidence of two academic ASes that publish their network load information, HEANET and SANET, suggest that γ belongs to $[1.5, 2]$ for medium-size European countries with one dominant IXP. In Figure 3.4, we depict the peering and transit traffic for both ASes on Thursday, 13th January 2011. One can observe that the peering and transit traffic profiles are rather similar. To quantify the similarity of the demand patterns, we use the $cosine$-similarity between the corresponding demand time series: $X = (x_1, \ldots, x_T)$ and $Y = (y_1, \ldots, y_T)$:

$$sim(X, Y) = \frac{\sum_{i=1}^{T} X_i Y_i}{\sqrt{\sum_{i=1}^{T} X_i^2} \sqrt{\sum_{i=1}^{T} Y_i^2}}.$$

The value of $sim(X, Y)$ is equal to the cosine of the angle between the vectors X and Y in

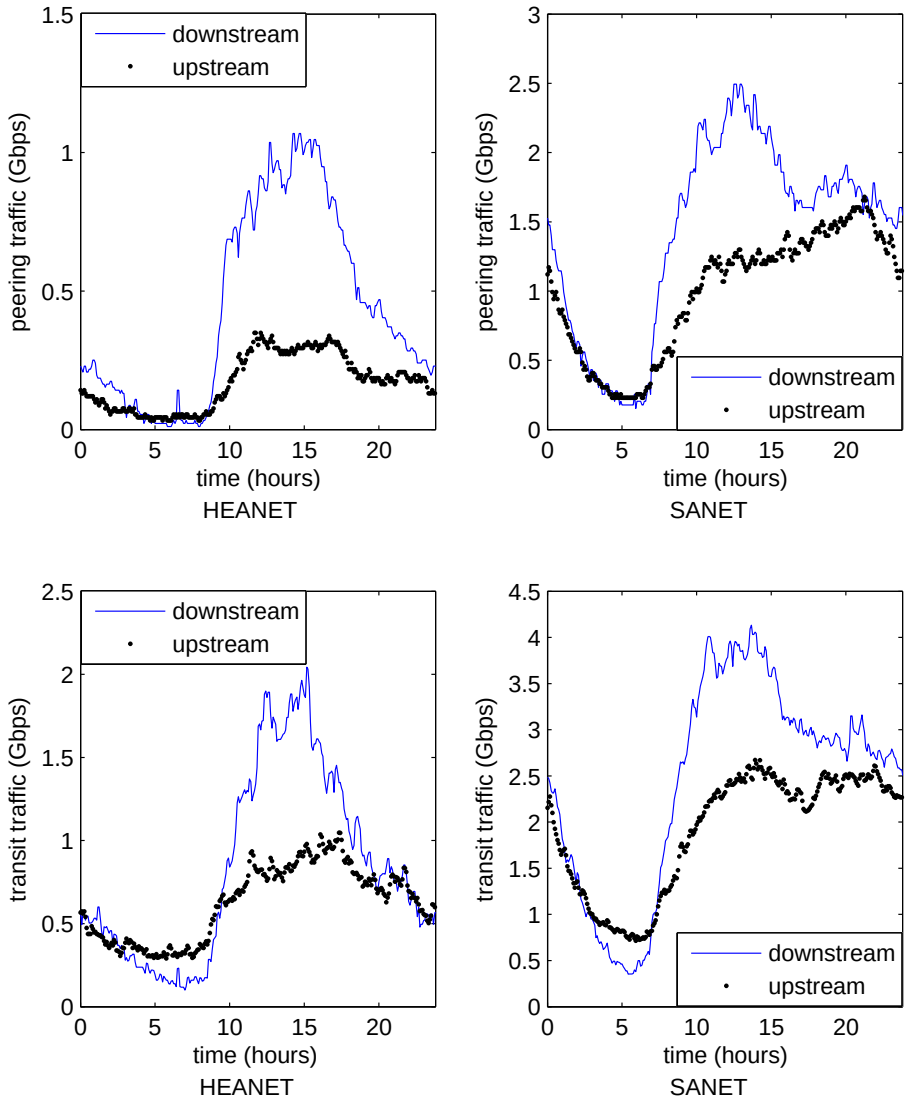


Figure 3.4: The transit and peering traffic in two national ASes: HEANET and SANET

the T-dimensional Euclidian space. Thus $sim(X, Y) = 1$ if $X = \alpha Y$ for a scalar α; otherwise $sim(X, Y) < 1$. Table 3.3 reports the values of $cosine$-similarity for the upstream and downstream time series for the both ASes.

Consequently in our analysis, we approximate the transit traffic of ASes (belonging to corresponding IXPs) with their peering traffic (information that is publicly available) multiplied by a factor γ that determines the relative weight of the transit vs. peering traffic. In Section 3.4.2, we describe expectable savings of CIPT for $\gamma \in [0.5, 4]$. However, in Sections 3.4.3 through 3.4.5 (which analyze the cost-sharing, coalition size, and geo-diversity), we fix γ at 1.5, which corresponds to the transit/peering traffic ratio of $60/40$ suggested by our empirical analysis for medium-size European countries with a single dominant IXP (the case of our 6 IXPs).

While this approximation is rather crude, it nevertheless captures the main features of the

AS: relative size, peak-hour period, upstream-to-downstream ratio, etc. For example, $\gamma = 0.5$ corresponds to the case where the peering traffic amounts to $1/(1+\gamma) = 2/3$ of all the traffic of the AS (as in Japan [47] and other localized markets), while $\gamma = 4$ corresponds to the case where $1/(1+\gamma) = 20\%$ of the total AS traffic is exchanged at the IXP, and the remaining 80% is transferred through transit (this situation is common in small markets [5]).

3.4.1.3 Pricing model

In the following evaluation, we use the pricing model (described in Section 3.1) with prices given in Table 3.1 and upstream/downstream traffic billed with either sum or max model. In Section 3.4.2, we describe the results of a comparative study of both sum and max models. In Sections 3.4.3 through 3.4.5, we focus on the sum pricing model (the more conservative one in terms of cost reduction) for the analysis of cost sharing, coalition size and geo-diversity.

3.4.2 Aggregate savings

In this section, we evaluate the aggregate potential savings of the IP transit costs for the coalitions consisting of *all* members within each IXP listed in Table 3.2. Following the discussion in Section 3.4.1.2, we approximate the IP transit traffic patterns by the traffic exchanged at these IXPs multiplied by constant $\gamma \in [0.5, 4]$; this constant represents the ratio between the transit and peering traffic volumes.

We stress again that the purpose of this evaluation is to shed some light on the potential savings of CIPT rather than computing accurate bounds of the savings. Such exact saving estimates strongly depend on various factors and should be calculated on a case-by-case basis.

For each of the 6 studied IXPs, Figure 3.5 reports the expected savings on the IP transit bill, both relative and absolute, in both the sum and max models. We see that the relative savings are in the range of 5-70$\%$ depending on the relative size of the IXPs and several other factors. These relative savings are strongly impacted by the size distribution of the involved ASes. Namely, for those IXPs that have several large ASes that dominate the traffic (and the costs), the relative savings of CIPT are low because these large ASes already receive the lowest price per $Mbps$. To illustrate that this is indeed the case, we define the *skewness* factor as the fraction of the traffic generated by the players with the peak traffic greater than 10 $Gbps$. Table 3.2 shows that the expected relative savings are considerably higher for the IXPs with a low *skewness* of under 0.3 (SIX, IIX, and InterLAN).

Remember that the savings of CIPT come from two properties of the IP transit model: price subadditivity and 95th-percentile billing. To quantify the effects that these two properties have on the CIPT savings, we identified what the relative savings would be without the subadditivity of the prices, i.e., if the price per $Mbps$ would be constant independent of the usage level. Such savings would come exclusively from the reduction in the 95th-percentile, the rest of the savings would hence correspond to the subadditivity effect. Table 3.2 presents these results in columns

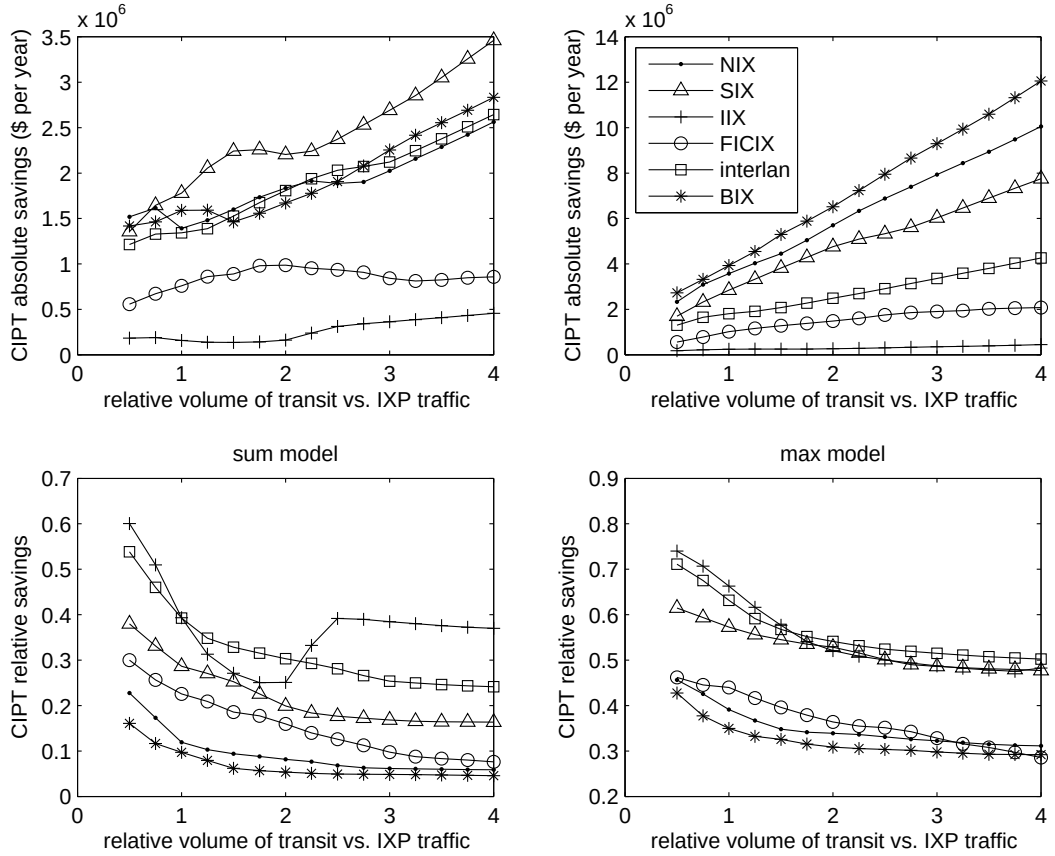


Figure 3.5: The absolute and relative savings as a function of the ratio between the transit and IXP traffic volumes

95th-pct effect and *subadditive effect* respectively. From this table, we can conclude that both properties (price subadditivity and 95th-percentile billing) influence the total savings.

The decreasing trend of relative savings can be observed in both sum and max pricing models. The decrease happens because the players with large volumes have smaller opportunities for large relative savings by CIPT (as they already experience a low price per $Mbps$). Nevertheless, the relative savings are bounded from below by the quantity of the *95th-pct effect* reported in Table 3.2 for both sum and max pricing models. Figure 3.6 replots the above findings to directly demonstrate that the sum model is indeed more conservative than the max model with respect to the attained CIPT gains.

We conclude this analysis with an observation that the 6 (medium-size European) countries hosting these IXPs have such traffic locality that around 40% of the traffic stays inside the country and is exchanged by peering (mainly through the dominant IXP), while the remaining 60% of the traffic uses IP transit. This corresponds to value γ of 1.5. Using this value of γ, we conclude that the expected relative savings in IP transit costs for the IXP-wide CIPT coalitions are in the range of 8-35% (in the sum model) and 32-56% (in the max model).

While we rely on the pricing function of a middle-size transit provider, the pricing function of

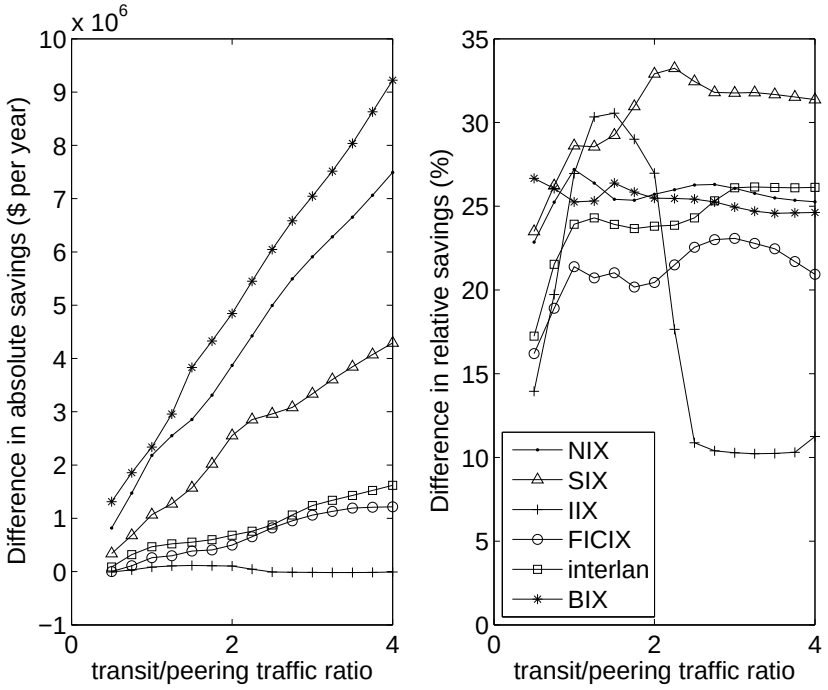


Figure 3.6: Differences in CIPT savings with the max vs. sum models

a larger provider can yield further quantity discounts: if a transit provider attracts large customers, the provider can offer discounts on larger volumes than alternative smaller transit providers, i.e., there are greater economies of scale with a large transit provider than with a smaller one.

On the other hand, regardless of how large the transit provider is, the additional discounts are finite. Therefore, starting from some huge traffic volume, CIPT cannot benefit from further discounts. At such traffic volumes, CIPT gains arise due to the 95th-percentile effect rather than the subadditive effect.

3.4.3 Coalition size

In the previous section, we analyzed the potential savings of coalitions that include *all* members of the corresponding IXPs. While such coalitions offer significant savings in terms of IP transit costs, coordination of such large coalitions may be cumbersome. In this section, we show that much smaller coalitions can offer savings comparable to those of the large coalitions. We take the Slovak Internet eXchange (SIX) with $N = 52$ members, and for each $k \in \{1, 2, \ldots, N\}$ we analyze the per-player savings from participating in the coalition of k random members of SIX. The pricing model is sum, and γ is set to 1.5. The results for other IXPs, max pricing model and other choices of γ are qualitatively similar; hence, we omit them for brevity.

In Figure 3.9, we report the median, 5th-percentile, and 95th-percentile savings, relative to the savings obtainable from the grand coalition of all $N = 52$ members. Since analyzing the statistics across all 2^{52} subsets is infeasible, we report the results obtained by sampling: for

each member i and each coalition size k, we pick random 100 subsets of size k that contain member i. From Figure 3.9 we can observe the law of diminishing returns: relatively small coalitions provide savings very close to the savings of the large coalitions, and, by adding more members to the coalition, the incremental savings are decreasing. In particular, even with as few as $k = 3$ members, one can expect savings that are half as large as the savings obtainable by the coalition of all $N = 52$ members. With $k \geq 10$ members, the median CIPT savings are greater than 80% of the savings obtainable by the grand coalition.

Note that the savings grow as the coalitions become larger. This is the consequence of the basic property of the CIPT cooperative game: the cost function is subadditive, as seen in Inequality 3.1. In other words, by adding a member, the coalition is better off. Also, note that for some ASes, participating in some smaller coalitions may be more beneficial than participating in the grand coalition (the relative savings exceed 1).

We stress that the results of this section are for random coalitions. By careful cherry-picking the most appropriate partners, one can obtain even higher savings, as the 95th-percentile of the savings in Figure 3.9 suggests. However, such optimization is out of scope for the present thesis.

3.4.4 Per-partner savings

In this section, we look at the per-member savings for each of the involved ASes when it participates in the IXP-wide CIPT. Following the reasoning described in Section 3.4.1.2, the γ factor used for scaling of the transit traffic is set to 1.5, and the pricing model is the more conservative sum model. As we elaborate in Section 3.3, each member of the coalition is assigned a cost equal to its Shapley value. The CIPT costs (across all ASes) are depicted in Figure 3.7 against the original IP transit annual costs. Figure 3.8 shows the absolute annual savings (the difference between the original IP transit costs and CIPT costs) for all ASes in these 6 IXPs.

We can observe two trends in Figures 3.7 and 3.8. First, the absolute savings typically grow with the size of the AS. This is a consequence of the fact that having a large AS in a coalition typically implies lower per $Mbps$ costs which in turn increases the contribution of the AS to the coalition, as reflected by the computation of the Shapley value in Equation 3.2. Therefore, a large AS can benefit from joining a coalition because the gains are computed as a total and then redistributed using the Shapley value, even if such large AS does not obtain a further price discount, other ASes do generate gains of which the large AS benefits.

In contrast to this increasing trend of the absolute savings, there is another interesting property of the CIPT cost allocation. Namely, the relative savings of CIPT (the ratio of the absolute savings of CIPT to the original IP transit costs) typically see a decreasing trend as a function of the AS size. This feature (decreasing trend of the relative savings) is strongly connected with the nature of the Shapley value as a cost allocation strategy but arises also because peak time of the aggregate traffic is predominantly determined by the large ASes. This means that ASes joining already larger coalitions (those that reached a close-to-minimum price per $Mbps$) bring lower relative benefits to the coalitions, consequently implying low relative-gains for these ASes.

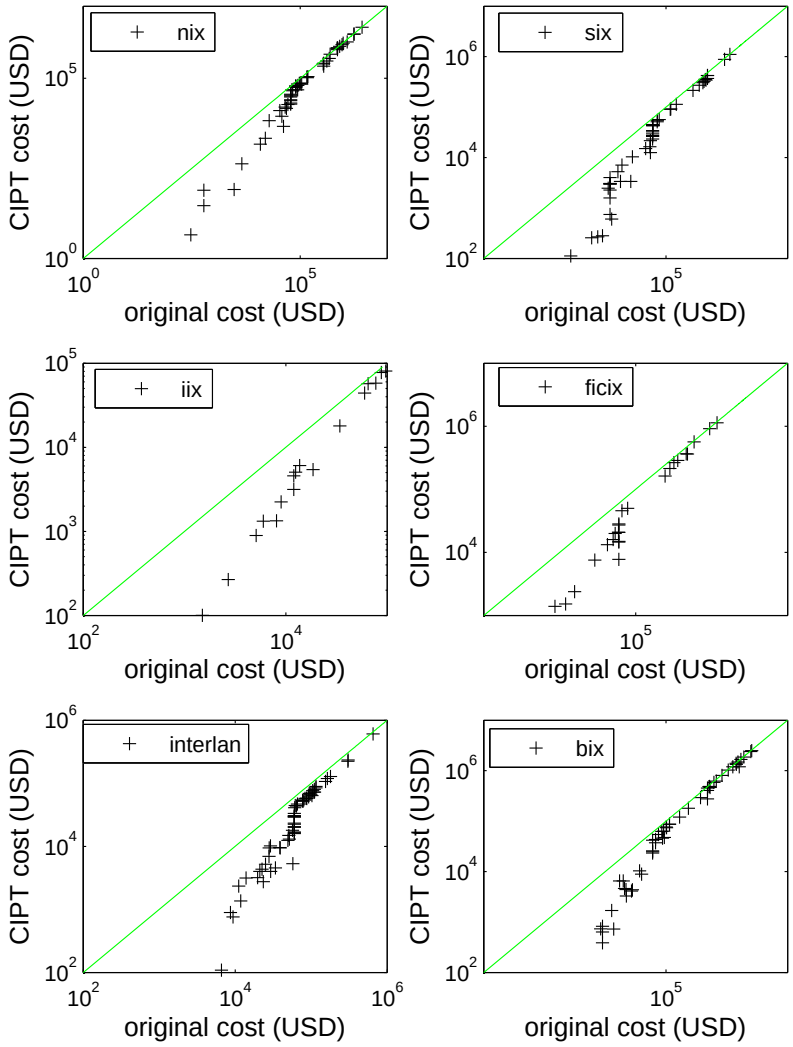


Figure 3.7: The original annual costs versus CIPT costs (Shapley value) across all the ASes from the 6 IXPs

While the Shapley value computes the expected contribution of an AS regardless of when it joins the coalition, absolute gains growth exhibits a decreasing trend. Consequently, relative gains decrease as the AS size grows.

3.4.5 Cooperation between remote subjects

So far, our analysis was concerned with the ASes operating in the same geographic area, and consequently having close peak hours. In such scenarios, the savings are mainly impacted by the price subadditivity rather than the burstable billing. In this section, we investigate potential savings of collaboration between geographically distant players. Because the remote collaboration

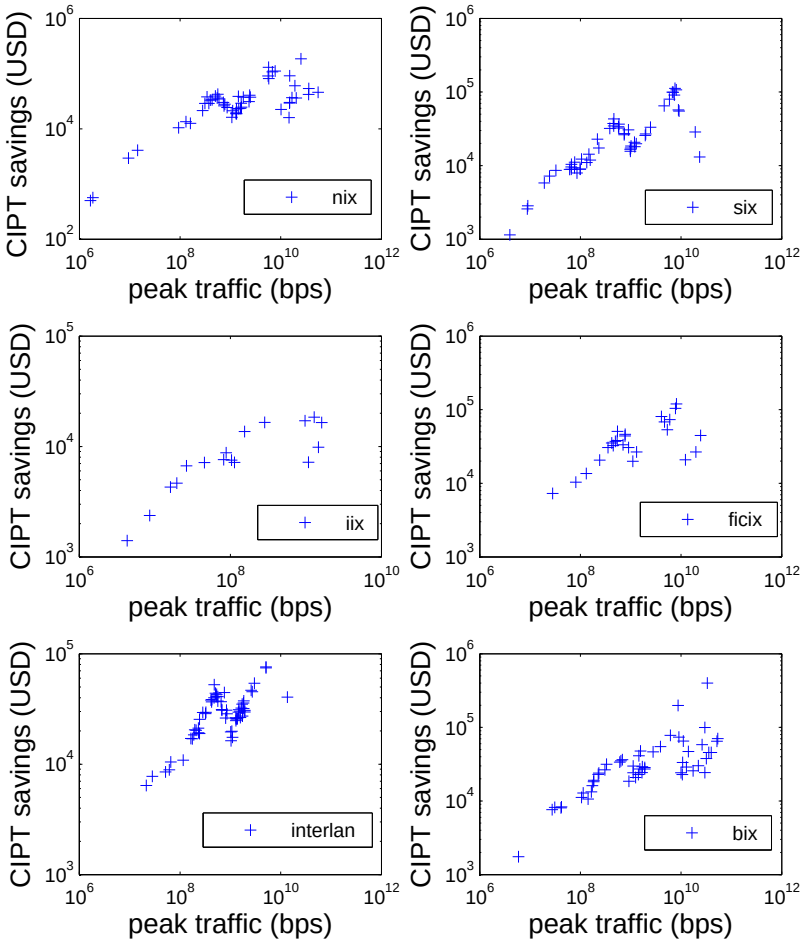


Figure 3.8: The absolute annual savings for all the ASes from the 6 IXPs

involves IP transport costs, it is possible only for large players. Only then the long-distance transport becomes cheap enough to make the CIPT economically viable [186]. Such long-distance transport to major (cheap) Internet hubs is not uncommon method for AS cost optimization. For example, each of the four largest IXPs – German Commercial Internet Exchange (DE-CIX), Amsterdam Internet Exchange (AMS-IX), London Internet Exchange (LINX), New York International Internet Exchange (NYIIX) – host ASes from more than 40 different countries.

Additionally, cooperation between very remote subjects (say, more than 6 time zones), may strongly impact the performance in terms of increased propagation delays. Some delay-sensitive applications (voice, gaming, etc.) may find such increase in delay unacceptable. Therefore, CIPT between very remote subjects is reasonable only for the traffic that is not delay sensitive (content, P2P, etc.) which indeed represents the majority of the Internet traffic [101, 114, 134].

While identifying and separating delay-tolerant traffic from non-delay-tolerant traffic is not trivial, the respective technical challenges have already been addressed [114, 134]. Even though

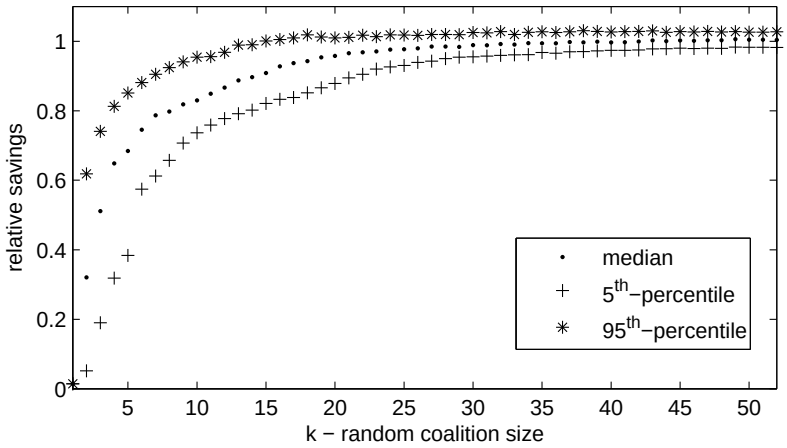


Figure 3.9: Relative (as fraction of the savings obtained in the grand coalition) per-player savings for smaller coalitions

traffic separation might be viewed as a network-neutrality violation, it might be also regarded as acceptable for performance reasons [53]. Since our goal is to show the economic attractiveness of CIPT, we focus on potential gains from CIPT between remote subjects, rather than on its technical implementation or net-neutrality aspects.

To analyze the potential savings in such a setup, we look at the potential savings of collaborations with *two* partners. Once *all* the partners are large enough to receive the minimum per-$Mbps$ price, the coalitions with more than two partners are not bringing large marginal benefits in terms of price reduction. Thus, we here focus on 2-partner coalitions. To assess the potential savings in such cases, we take all $M = 93$ ASes from our 6 IXPs with the peak traffic greater than 1 $Gbps$ and shift each of them for a (uniformly) random number of time zones. For each of the $M(M-1)/2$ pairs, we evaluate the relative savings of the coalition: $1-cost(CIPT(i,j))/(cost(i)+cost(j))$ and plot them against the time difference in Figure 3.10. One can observe the following trend: the further away the two partners are, the greater the opportunity is for the CIPT savings. In Figure 3.10, we also depict the bound

$$g(\psi) = \frac{1 - |\cos\frac{\psi}{2}|}{2}, \tag{3.4}$$

where $\psi = \frac{time\_difference}{24} 2\pi$ is the scaled time difference. We prove the upper bound on the relative savings in a simple model where the demand curves are modeled as sin-waves (see below). One can observe that the relative reduction in the 95th-percentile for a coalition of two partners is in the range of $[0, 0.5]$, in line with the model predictions. However, the expected savings appear to be larger as the time difference grows, and peak when two ASes are 12 time zones apart. To explain and quantify this property, we employ a simple trigonometric model where the demand pattern of the AS is modeled as a sin-wave function. The following proposition characterizes the expected reduction in the peak traffic from CIPT collaboration between two partners with

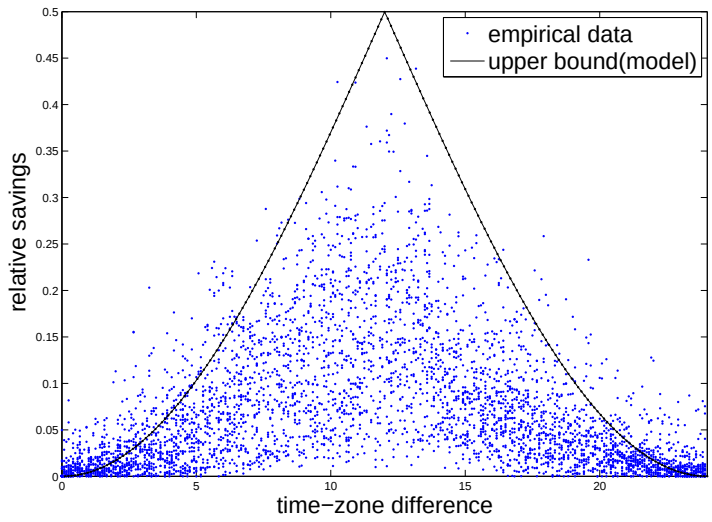


Figure 3.10: Relative savings between large remote subjects coming from the 95th-percentile subadditivity

non-coinciding peak hours:

Proposition 1. *Let two players have demand given by*

$$D_i(t) = A_i \cos(2\pi \frac{t - M_i}{24}) + B_i, \;\; t \in [0, 24) \textit{ hours}.$$

where B_i is the mean traffic intensity, $A_i + B_i$ is the peak traffic intensity, and M_i is the peak hour of partner i. By creating a CIPT coalition between these two partners, the relative reduction in the peak is equal to:

$$G_{12} = 1 - \frac{B_1 + B_2 + \sqrt{A_1^2 + A_2^2 - 2cos\psi A_1 A2}}{B_1 + B_2 + A_1 + A_2} \leq g(\psi),$$

for $\psi = \frac{M_1 - M_2}{24} 2\pi$, the scaled time-zone difference, and $g(\psi)$ defined in Equation 3.4.

3.5 Implementation and deployment issues

Section 3.4.4 presented a compelling evidence that CIPT with Shapley-value sharing of transit costs offers significant benefits to the CIPT partners. While the economic incentives are crucial for CIPT being viable, the viability is a topic with multiple dimensions. Without pretending to be comprehensive, this section discusses other aspects of CIPT such as its organizational embodiment, physical infrastructure, performance, traffic confidentiality, and interdomain routing.

Organizational embodiment: CIPT is an innovative mechanism for reducing transit costs. Among other cost-reduction mechanisms, peering is similar to CIPT in its cooperative nature and

commonly organized as a nonprofit IXP. In our vision for CIPT as an organization, a typical arrangement is also a nonprofit organization. The nonprofit status of a CIPT promotes a valuable marketplace image of its neutrality and fair treatment for all its partners. In such an organization, partnership fees are used only to recover the technical and management overhead costs of operating the CIPT and expected to be insignificant in comparison to the transit cost reductions provided by the CIPT. In a future study, we plan to quantify the technical and economic overhead. While the nonprofit arrangement looks the most suitable, deviations are quite possible and even likely; as with some existing IXPs, some CIPTs might operate as government or commercial organizations. Finally, a single AS may choose to participate in multiple CIPTs in order to increase the provider diversity.

Physical infrastructure: The physical implementation is another issue where CIPTs can benefit from the IXP experience. For buying IP transit in bulk, a CIPT needs to concentrate traffic of multiple ASes in one location. The physical infrastructure of any IXP already supports such concentration for peering purposes. Moreover, some IXPs diversify their service portfolio by offering access to transit providers. For example, Vancouver Transit Exchange is an IXP that also hosts transit providers and thereby enables an AS to satisfy its peering and transit needs at the same location [96]. A CIPT can be implemented as a further diversification of the IXP service portfolio. By leveraging the physical infrastructure of an existing IXP, the CIPT can keep its operational costs low.

Performance: A CIPT and its transit provider sign a contract for IP transit. The contract is expected to be of the same type as existing contracts between an individual AS and its transit provider. In particular, the contract includes a Service-Level Agreement (SLA) stating the maximum outage duration, packet delay, jitter, and loss rate for the CIPT traffic. The SLA also specifies financial compensations by the provider if the latter fails to provide the CIPT with the agreed performance. In reality, SLA violations are likely to be rare. Whereas the performance levels of traditional inter-provider SLAs are very similar, having a single SLA for the multiple-partner CIPT is not problematic. Also, the typical SLA metrics of packet delay, jitter, and loss rate are such that the traffic of individual CIPT partners can inherit the performance levels of the CIPT aggregate traffic without any special technical support. Furthermore, the CIPT and its individual partner can sign a separate bilateral agreement on performance issues.

Traffic confidentiality: While it is feasible to formalize traffic metering and billing for a CIPT by means of bilateral agreements between the CIPT and each of its individual partners, the bill of a partner depends on the traffic of the other partners. Some academic ISPs – such as the aforementioned HEANET, SANET, GRNET and CESNET – reveal their transit and peering traffic. However, a typical commercial AS tends to be more secretive and does not disclose its traffic patterns. To alleviate the privacy concerns, a CIPT can keep the traffic profiles of its partners confidential and incorporate an internal audit system for verifying the correctness of traffic metering and billing for each partner.

Interdomain routing: With BGP being a de facto standard protocol for routing between ASes, we see no technical complications with CIPTs from the interdomain routing perspective. A CIPT can acquire a separate AS number for inclusion into its BGP path announcements. Alternatively, as in the case of some IXPs, the partners of a CIPT can agree to use the individual AS number of one (typically, prominent) partner in all BGP announcements by the CIPT.

Multihoming and traffic engineering: Both are feasible with CIPT. A CIPT partner can buy transit outside the CIPT as well. Also, a CIPT can buy transit from multiple providers. While multihoming might increase costs, CIPT can reduce these costs due to price subadditivity and burstable billing.

Social impact: The overall social impact of CIPTs appears positive. In particular, CIPTs are beneficial for narrowing the digital divide between the developed countries and poorer world which lies on the Internet edges and does not own a transit infrastructure for reaching the Internet core. In places like Africa, IP transit (and IP transport) is more expensive but the ability to pay for it is lower. Like with IXPs that have positively affected Africa by exchanging its traffic locally rather than through North America or Europe, CIPTs can benefit Africa and other developing regions by making the access to the Internet and its information more affordable [4, 5].

3.6 CIPT: a strategic perspective

In this section, we analyze potential strategic reactions to and within CIPT. While a CIPT coalition can include members with different market power, more powerful members can try to gain extra benefits by leveraging their stronger bargaining position against weaker members of the coalition. Besides, CIPT participation depends on existing or potential transit and peering relationships. Section 3.6.2 examines such issues related to CIPT formation and participation. Strategic CIPT issues are also relevant to ASes that are not directly involved in CIPT relationships. Other individual customer ASes can react by forming their own CIPT coalitions. More interestingly, both the transit provider of the CIPT members and its competitors can adjust their strategic behaviors in response to the CIPT emergence. Section 3.6.1 studies the reactions of transit providers.

3.6.1 Transit providers

Whereas transit customers form a CIPT for the simple reason of reducing their costs, the reaction of transit providers to CIPT is a multifaceted issue. Somewhat counterintuitively, the transit providers can favor CIPT for a number of reasons.

One potential incentive for an interest of transit providers in CIPT lies in transit traffic elasticity [191]. By decreasing the transit costs of individual buyers, CIPT increases their future demands. While we had no access to reliable data on transit elasticity, this chapter quantified the benefits of CIPT conservatively without these extra gains. Also, regardless of whether a transit

provider is a monopolist, CIPT increases overall demand by turning prospective buyers into actual customers via aggregation of their individual demands. Moreover, CIPT traffic aggregation can enable the transit provider to bypass resellers of its transit service and serve small customers directly through the CIPT. Finally, if the transit provider is not a monopolist, it can adopt CIPT contracts to attract new customers from its competitors.

Traffic aggregation can allow small customers to pool their traffic together and become attractive customers for transit providers. More generally, direct provisioning of transit to small customers is sometimes unattractive for big ASes. Instead, mid-size networks resell transit of big ASes to small customers. By aggregating traffic of multiple small members, a CIPT can reach an acceptable size for direct transit sales by a big AS. Such outcome can be mutually beneficial for both the CIPT and transit provider. While this chapter already elaborated on the benefits for the CIPT members, the transit provider benefits as well by selling the same traffic at higher per-Mbps prices than through the intermediary. Even though the bypassed intermediary does not find the CIPT beneficial, the reseller does not have effective means or clear grounds to oppose the direct relationship between the CIPT and big AS.

Additionally, in situations where the transit market is competitive, a transit provider can try adopting CIPT to increase its revenues at the expense of its competitors which, in their turn, can try doing the same. The competition for CIPT contracts drives per-Mbps CIPT prices down. In Bertrand competition model for homogeneous goods, such competitions converge to the so-called Bertrand paradox where the competitors offer prices that match their costs, i.e., yield no profit [189]. In practice, while transit providers are sufficiently heterogeneous (e.g., with respect to geographic coverage, service quality, and cost structure) to avoid the extreme no-profit outcome, their actual prices are still likely to be attractive for CIPT coalitions.

In theory, transit providers can also benefit from CIPTs due to a variety of other economic factors that include transaction efficiency, traffic uncertainty, customer heterogeneity, and production postponement [7]. In the current context of IP transit, these factors do not appear to be strong enablers of CIPT. Thus, we view the traffic aggregation and inter-provider competition as the two main reasons for the CIPT feasibility from the transit-provider perspective.

3.6.2 Strategic issues within the CIPT coalition

Strategic issues exist within a CIPT coalition as well. One specific issue is CIPT formation, i.e., which ASes join the coalition. Another interesting issue is CIPT cost sharing, i.e., whether and how the CIPT members can leverage the Shapley-value cost sharing mechanism for their individual advantages.

Peering and transit relationships of CIPT members, as well as their position in the transit hierarchy, are relevant to CIPT formation. Both peering and CIPT are mechanisms for transit-cost reductions. By reducing transit costs, CIPT can decrease the value of peering. Due to this effect, ASes with established peering relationships can be reluctant to join CIPTs. For the same reason, CIPT members can be reluctant to enter peering relationships. This tension between CIPT and

peering can increase demand for traditional transit and thereby serve as an additional incentive for transit providers to support CIPT.

As we discuss in Section 3.6.1, CIPT makes a negative impact on bypassed transit resellers. To compensate for the diminished revenues, a bypassed reseller can itself join a CIPT in order to minimize the losses.

So far, our analysis considered static situations only. CIPT dynamics broaden the scope of potential strategic behaviors. For example, if an AS joins a CIPT coalition with a certain traffic contribution and later communicates at a different rate, the AS traffic change affects the gains achieved by other CIPT members. To deal with such future traffic uncertainties, CIPT coalitions can adopt a mechanism that requires each member to commit to an expected traffic level for some time period.

3.7 Conclusions

In spite of the steady decline of IP transit prices, the IP transit costs remain high due to the traffic growth. Over the previous decades a number of solutions have been suggested to reduce these IP transit costs, including settlement-free or paid peering, IP multicast, CDNs, and P2P localization. In this chapter, we proposed an alternative cost-reduction technique of Cooperative IP Transit (CIPT) that, in contrast to the existing solutions, does not alter the traffic. Namely, CIPT utilizes tuangou, or group buying, for IP transit. The savings in CIPT come from two distinct yet ubiquitous properties of the IP transit pricing model: price subadditivity and burstable billing. Our data-driven analysis suggested that significant savings can be expected from using CIPT. We are confident that the potential savings of CIPT, combined with its simplicity, will encourage many Internet entities to engage in CIPT partnerships.

Chapter 4

Transit for Peering (T4P)

While the previous chapter leveraged the economies of scale in interdomain interconnections to reduce transit bills of ASes, this chapter goes beyond transit. We study the evolved interconnection ecosystem where transit co-exists with peering. This chapter examines how the economies of scale can be leveraged in this diversified ecosystem.

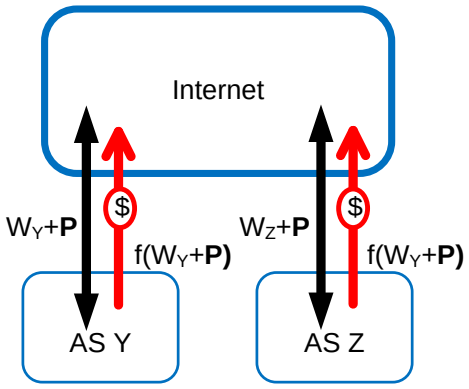
In its early years, the Internet ecosystem was essentially a hierarchy where smaller ASes paid bigger ASes for the universal Internet reachability via transit links. Subsequent massive emergence of peering enabled many ASes to exchange their customer traffic over cost-effective settlement-free peering links, leading to the observed Internet flattening phenomena as described in Chapter 2. The evolution kept increasing the diversity of inter-AS connection types and introduced partial-transit, paid-peering links [70, 191], and remote peering. In contrast to the full transit, a partial-transit link offers access to only a fraction of the global Internet address space. With paid peering, one of the peering ASes pays the other peer for exchanging their customer traffic.

To a large extent, the driving forces behind the interconnection evolution are economic. For example, if two ASes exchange their traffic via a transit provider, their payments to the provider significantly exceed the cost of communicating the same traffic over a settlement-free link. The costs of the peering are mostly related to the infrastructure and labor of maintaining the physical interconnection, either as a direct link or through an IXP [15,55,58]. The potential of settlement-free peering to reduce the costs for both peers does not mean that the ASes will indeed establish and sustain such a relationship. For instance, the ASes might view each other as competitors and be unwilling to reduce the costs of the counterpart. Furthermore, the costs of each party depend on the peering-link traffic and AS sizes. Agreements for settlement-free peering commonly stipulate that the traffic flows in the two directions of the peering link should be balanced within a certain ratio (e.g., ratio 2:1) and that the geographic scopes of the peering networks should be similar [56]. Loosening the above requirements, a paid-peering interconnection enables peering of diverse ASes through monetary payments, e.g., by allowing one AS to send more traffic and pay the other AS a monetary compensation for the traffic imbalance.

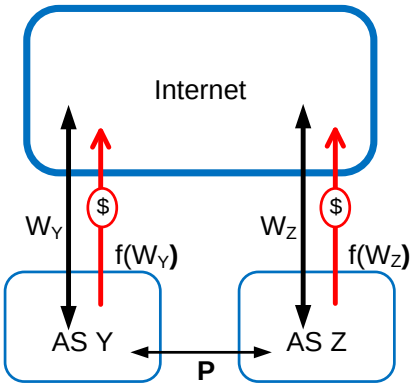
Accompanying the interconnection evolution [57], ASes specialized into providing Internet access to content (i.e., content networks) or residential users (i.e., eyeball networks) [70, 125]. As previously discussed in Section 1, recurrent conflicts over peering settlements [10,25,26,154] and the net-neutrality debate [127] resulted from the divergent interests of content and eyeball networks.

In the one hand, asymmetric traffic patterns intensify the cost recovery need of the eyeball networks, who bear a capital intensive last-mile infrastructure. On the other hand, the high cost of the last-mile infrastructure represent a significant barrier of entry, that in turn limits competition in the eyeball-network market and enhances the market power when negotiating a peering agreement with the content AS [70, 125, 140].

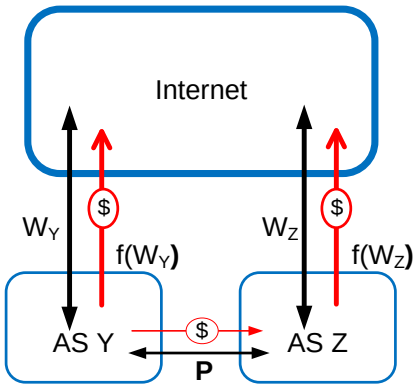
In this chapter, we propose *T4P (Transit for Peering)*, a new type of hybrid bilateral AS interconnection that can reduce interdomain traffic costs and strengthen the Internet connectivity.



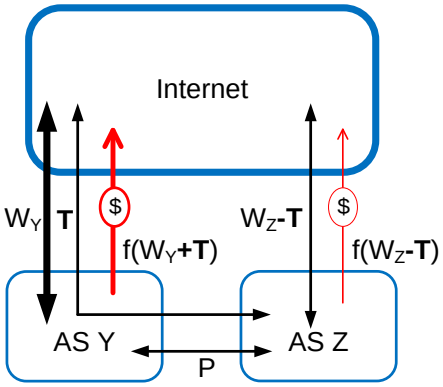
(a) only indirect interconnection through the Internet



(b) settlement-free peering



(c) paid peering



(d) T4P interconnection

Figure 4.1: Different types of interconnection between ASes Y and Z: while double-arrow lines depict traffic flows, single-arrow lines show monetary flows

In T4P, the link between two ASes Y and Z carries not only peering but also some transit traffic. In particular, AS Z communicates a portion of its transit traffic not through its transit providers but over the T4P link and through the transit providers of AS Y. While AS Y effectively becomes a partial-transit provider for this traffic of AS Z, the partial transit serves as an in-kind traffic-delivery compensation paid by AS Y to AS Z for the peering. Unlike paid peering, T4P does not employ monetary payments. In comparison to paid peering, T4P is able to reduce the combined transit/peering costs of an AS due to the subadditivity of transit billing. This chapter uses real traffic data to demonstrate the cost-reduction potential of T4P.

The main contributions of the chapter are the following ones:

- We propose T4P, a novel type of hybrid AS interconnections where partial transit serves as a compensation for peering.

- Using real traffic data available at several IXPs, we quantify the financial incentives

of ASes to adopt T4P.

The rest of the chapter is structured as follows. Section 4.1 provides additional background and motivation for the studied problem. Section 4.2 presents the concept of T4P in more detail. Section 4.3 reports the economic model and analyzes the AS incentives for T4P adoption. Finally, Section 4.4 evaluates T4P using the real IXP data.

4.1 Increasing diversity of interconnections

This section presents the background information on the Internet interconnection evolution that sets the stage for the T4P proposal.

Transit is the interconnection type that was heavily predominant in the early years of the commercial Internet. A transit link connects two ASes called a provider and customer. As it was described in the previous chapter, in transit the customer pays the provider for the traffic communicated in both directions of the interconnection. In exchange for the monetary payments, the provider offers the customer access to the global Internet. In a typical transit relationship, the provider is a larger network with a broader geographic scope. Figure 4.1 a shows two examples of transit interconnections, with ASes Y and Z acting as customers.

Transit billing, presented in the previous chapter (see Section 3.1) is typically subadditive. In this section we focus in the *max* transit billing model: the 95th-percentile traffic rate is calculated for each of the two link directions and the largest of the two 95th-percentile traffic rates serves as the billed traffic rate. Then, a subadditive price function is applied to the billed traffic rate to compute the monetary settlement. With the subadditive billing, larger traffic rates are usually billed at lower transit prices per Mbps.

Settlement-free peering has emerged as a cost-effective interconnection where two ASes Y and Z exchange their customer traffic directly without any monetary compensation. Whereas the peering link carries only traffic of own customers, a vast majority of ASes still needs transit links to reach the global Internet. Nevertheless, by reducing the traffic on the transit links, settlement-free peering reduces the transit costs for both ASes Y and Z. Figure 4.1b depicts settlement-free peering.

Despite its definite potential for cost reduction, settlement-free peering has struggled to fully accommodate the increasingly diverse population of ASes. In particular, the Internet evolution produced the two AS types of content and eyeball networks with very different profiles in regard to the number and sizes of customers, dominant direction of traffic flows, cost structure, and market power. Eyeball networks receive more traffic than they send, serve more users, incur higher traffic-delivery costs, and enjoy a stronger bargaining position because vendor lock-in is arguably easier with residential users than with content-providing customers [70, 125]. These differences make eyeball ASes hesitant to peer with content ASes on the settlement-free basis. For example, after content provider Netflix became a customer of Level 3, the traffic imbalance on the peering link between Level 3 and eyeball-network Comcast increased, and Comcast threatened to de-peer

(i.e., terminate the peering agreement) with Level 3. Although Level 3 offered to resolve the conflict by upgrading its communication infrastructure and making its routing more beneficial for Comcast, the latter rejected the offer and de-peered.

Paid peering is a more flexible interconnection that allows AS Y to monetarily compensate AS Z for their peering. With respect to the traffic flows, paid and settlement-free peering are identical. Figure 4.1c illustrates paid peering.

4.2 T4P concept

While Section 4.1 exposed the increasing interconnection diversity as well as economic factors behind this evolution, our T4P proposal continues the diversification trend to find an economically viable niche in the evolving Internet ecosystem. T4P is a novel type of hybrid bilateral AS interconnection between diverse ASes, such as content and eyeball networks. Unlike with paid peering, the T4P interconnection does not involve any monetary settlement. Instead, T4P employs in-kind compensations in the form of partial traffic transit. Specifically, AS Y compensates AS Z for their peering by providing a transit service for some traffic of AS Z. Figure 4.1d depicts the T4P interconnection between ASes Y and Z.

T4P is fundamentally different from paid peering not only because of eliminating any monetary compensation between ASes Y and Z but also due to changes in the traffic flows. While paid peering is identical to settlement-free peering in restricting the link between ASes Y and Z to own customer traffic, the hybrid T4P link combines the peering traffic with transit traffic. Hence, the T4P interconnection affects both transit routes and traffic rates along these routes.

The subadditive nature of transit billing is the reason why T4P is able to reduce the traffic costs of both ASes Y and Z in comparison to paid peering. Although the overall transit traffic of ASes Y and Z does not change with T4P, serving some transit traffic of AS Z through AS Y can decrease the overall transit costs of the two ASes due to the billing subadditivity.

Whereas the subadditive billing can reduce the combined transit/peering costs of ASes Y and Z, the attractiveness of T4P vs. paid peering for an individual AS depends on the monetary settlement in the paid-peering interconnection. In particular, AS Z finds T4P more attractive than paid peering only if the transit bill reduction for AS Z with T4P is at least as large as the monetary settlement of paid peering.

4.3 Incentive analysis

In this section, we expand and formalize the incentives of ASes to adopt T4P. While we envision T4P as a cost-effective interconnection between diverse ASes, paid peering serves as a natural baseline in our analysis.

The lack of real data on paid-peering settlements makes it problematic for our model to explicitly represent the monetary compensation paid by AS Y to AS Z in the paid-peering inter-

connection. Instead of treating the monetary compensation as an explicit parameter, our analysis relies on the key insight that the combined traffic costs of ASes Y and Z are independent from this monetary compensation: the compensation paid by AS Y is the same as the compensation received by AS Z. Hence, the first step of our analysis focuses on the overall economic efficiency of T4P vs. paid peering for the two ASes together.

Our model accounts for the subadditive billing of transit services. The subadditive billing method is a positive factor for T4P because adding the partial-transit traffic of AS Z to the own transit traffic of AS Y reduces the price paid by AS Y per Mbps of the aggregated transit. We consider the method variant that bills the two directions of a link together by applying function f to the sum of the 95th-percentile traffic rates in the individual directions. In our model, billing function f always selects the CDR value that yields the smallest possible per-Mbps price for any traffic pattern. Using i to denote either AS Y or AS Z, we express transit cost C_i of AS i as

$$C_i = f\left(A_i\right) \tag{4.1}$$

where A_i represents the billed bidirectional traffic pattern.

To represent the transit costs of ASes Y and Z interconnected with T4P, let T denote the partial-transit traffic on the T4P link, and W_i be the own transit traffic of AS i on the link with its normal transit provider. Then, transit costs F_Y and F_Z of ASes Y and Z with the T4P interconnection are

$$F_Y = f\left(W_Y + T\right) \text{ and } F_Z = f\left(W_Z - T\right). \tag{4.2}$$

With paid peering, transit costs N_i of AS i equal

$$N_i = f\left(W_i\right). \tag{4.3}$$

Then, *transit cost reduction* R_i provided by T4P to AS i is

$$R_i = N_i - F_i. \tag{4.4}$$

With the subadditive billing, f is a non-decreasing function. Thus, Equations 4.2, 4.3, and 4.4 imply that T4P does not decrease the transit costs for AS Y (i.e., $R_Y \leq 0$) and does not increase the transit costs for AS Z (i.e., $R_Z \geq 0$).

We define *aggregate T4P gain* G as

$$G = R_Y + R_Z. \tag{4.5}$$

Combining Equations 4.2, 4.3, 4.4, and 4.5, we express the aggregate T4P gain as

$$G = f\left(W_Y\right) + f\left(W_Z\right) - f\left(W_Y + T\right) - f\left(W_Z - T\right). \tag{4.6}$$

Due to the subadditive billing, the aggregate T4P gain is non-zero in general and depends on traffic T that AS Z transits through AS Y.

In comparison to paid peering, the overall economic efficiency of T4P for the two ASes together is better when $G > 0$. Hence, we have proved the following theorem:

Theorem 1. *For the two ASes together, T4P is economically more attractive than paid peering when*

$$f(W_Y) + f(W_Z) > f(W_Y + T) + f(W_Z - T). \tag{4.7}$$

Switching from the aggregate gain to the individual perspectives of ASes Y and Z, one can think of R_Z as a monetary equivalent of the in-kind traffic-delivery compensation provided to AS Z with T4P. Hence, AS Z favors T4P when R_Z is larger than monetary compensation x paid by AS Y to AS Z in the paid-peering relationship. Even when R_Z is strictly greater than x, AS Y also finds T4P more attractive as long as the transit cost increase ($-R_Y$) of AS Y with T4P is smaller than x. There is a continuum of such mutually beneficial settings when aggregate gain G is positive.

4.4 Data-driven evaluation

In Section 4.3, we analyzed AS incentives for adopting T4P. To quantify the economic attractiveness of the T4P, we evaluate it using real traffic data and real transit pricing. Section 4.4.1 presents our evaluation methodology. Section 4.4.2 illustrates the potential benefits of T4P with an example. Section 4.4.3 evaluates T4P in more detail.

4.4.1 Evaluation methodology

Similarly to the previous chapter, our evaluation relies on real traffic at the six IXPs presented in Table 3.4.1.1). Each of the six IXPs in table 3.2 reports peering traffic for its AS members in the form of network-management images, such as the one in Figure 3.1. Obtained by applying Optical Character Recognition (OCR) to such images, our numeric data for the peering traffic serve as a basis for approximating the transit traffic of the member ASes. As in the previous chapter (see Section 3.4.1.2), we scale up the peering traffic of a member AS with the factor of 1.5 to represent the transit traffic of the AS (real traffic data available at some academic ASes validate such correspondence between the peering and transit traffic patterns). To evaluate T4P instances, we consider all possible pairs of ASes at each IXP.

The 95th-percentile billing for transit services utilizes the Voxel prices in Table 3.1 (in Section 3.4.1.3 of the previous chapter). Transit providers tend to treat their prices as confidential. Voxel, a North American AS, is a rare exception and publishes its transit pricing [194]. Table 3.1 sums up the transit prices of Voxel with respect to the CDR chosen by the customer. The transit payment is calculated as the product of the price per Mbps for the chosen CDR and either 95th-percentile traffic rate or CDR when the latter is larger.

Table 4.1: Illustrative example of T4P relationships for two ASes at NIX

Interconnection type	Parameter	AS Y	AS Z
Paid peering	Billed transit traffic rate, Gbps	7.7	5.6
	Transit costs	\$38K	\$28K
T4P	Transit traffic rate, Gbps	12.2	1.1
	Transit costs	\$49K	\$6K
	Transit cost reduction	−\$11K	\$22K
	Maximum compensation to AS Z	N/A	\$11K
	Maximum compensation reduction for AS Y	\$11K	N/A

4.4.2 Illustrative example

In this section, we present an example that illustrates the potential benefits of T4P for a pair of ASes Y and Z at Neutral Internet eXchange (NIX). The peering traffic of the ASes is imbalanced with AS Y sending more traffic at the ratio of 4:1.

We examine the T4P instance that maximizes the aggregate T4P gain by shifting 80% of the transit traffic of AS Z to the T4P link. As shown in table 4.1, T4P decreases the transit costs of AS Z by \$22K but raises the transit costs of AS Y by \$11K. Thus, the maximum aggregate T4P gain attained by T4P is \$11K. AS Y can provide up to a \$11K compensation to AS Z without increasing its own overall traffic costs. This amount of $G = \$11\text{K}$ represents the budget of the T4P transit-cost benefits that can be distributed between ASes Y and Z in various ways. The other extreme of this continuum is financially equivalent to settlement-free peering for AS Z: the latter does not see any changes in its traffic costs but AS Y decreases its overall traffic costs by \$11K.

4.4.3 Evaluation results

Now, we take a detailed look at the T4P relationships between all ASes at the six IXPs. Figure 4.2 shows the distributions of aggregate T4P gain G. Such gains are in addition to those resulting from peering and arise due to the subadditive billing and the aggregation of transit traffic. At FICIX, 60% of its AS pairs gain at least \$1K, with almost 5% of its AS pairs gaining beyond

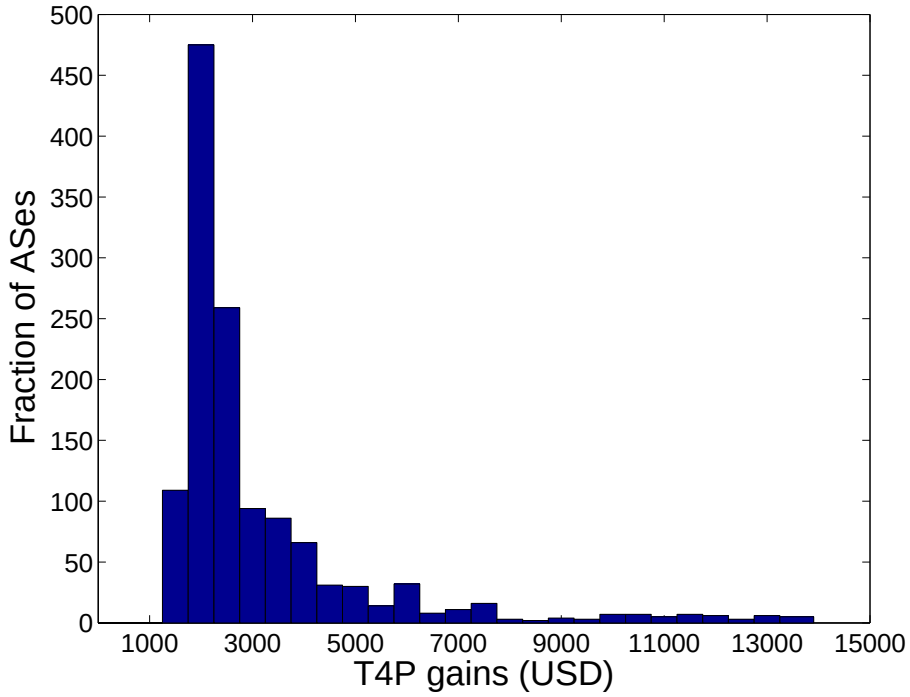


Figure 4.2: Distributions of aggregate T4P gain G for T4P at the six IXPs

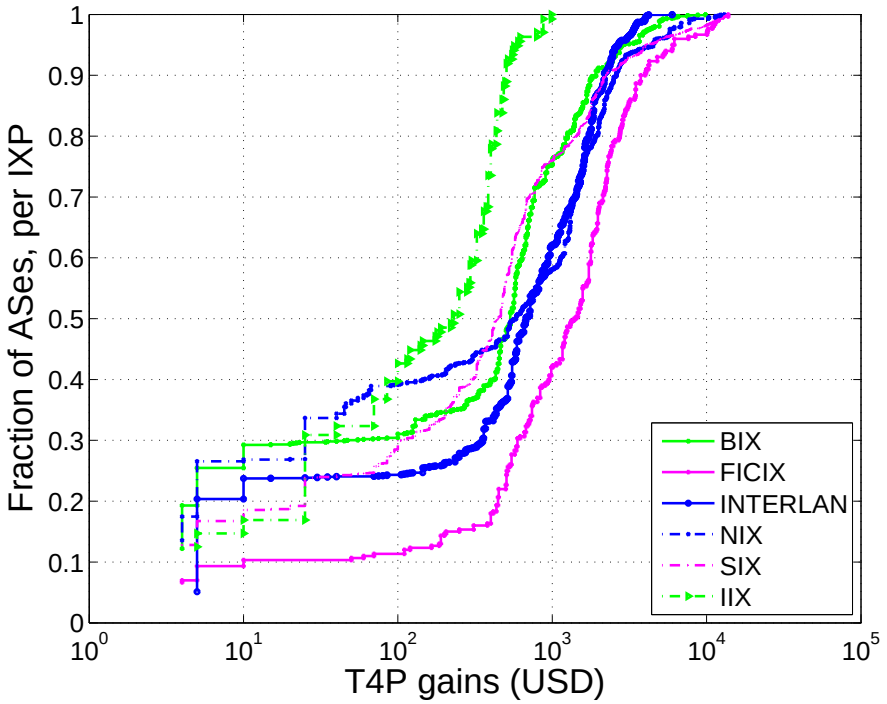


Figure 4.3: Top 20% of the AS pairs with the T4P relationships that have the biggest aggregate gains across all six IXPs

\$10K. The percentage of AS pairs with G above \$1K is greater than 40% for NIX and InterLAN, and larger than 20% for Budapest Internet eXchange (BIX) and SIX. Israeli Internet eXchange (IIX) provides the lowest gains among the 6 IXPs, with less than 5% of its AS pairs gaining more than \$1K. Figure 4.3 focuses on the top 20% of the AS pairs with the biggest G across all 6 IXPs. Around 150 AS pairs gain a larger G than \$5K, and about 45 AS pairs gain more than \$10K.

4.5 Conclusion

This chapter proposed and evaluated T4P (Transit for Peering), a new type of hybrid bilateral AS interconnection. In T4P, one AS compensates the other AS for their peering by providing this other AS with a partial-transit service. Leveraging the subadditive billing of transit services, T4P offers economic benefits over paid peering and thereby opens new opportunities for AS interconnections. We modeled and analyzed T4P relationships with paid peering as the natural comparison baseline. Then, we evaluated T4P based on the real Voxel transit pricing as well as real traffic data available at the six European IXPs. Our results confirmed the potential of T4P to expand and strengthen the Internet connectivity through the more flexible cost-effective interconnection.

Chapter 5

Route Bazaar

Whereas the previous chapters looked at solutions within the existing interconnection framework, this chapter first identifies inherent limitations of the current framework and then proposes a novel Internet architecture to overcome those drawbacks.

While the Internet became synonymous with quick transformation of everyday life, network interconnection contracts in the Internet of 2015 are established in a very old-fashioned way. Critical for new interconnections are personal contacts, legal teams, and social venues where taking care of the participants' subconscious needs [144] is a key aspect.

The current situation with interconnection contracts reflects stagnant practices that have become entrenched over a period of decades. After the infrastructure privatization in the early 1990s, the Internet came to consist of multiple independent networks or ASes. In the Internet infrastructure owned by multiple parties, global end-to-end reachability requires interconnection of individual networks. In the multi-party Internet, delivering the traffic of another network is a service that has its costs and can be provided with different levels of Quality of Service (QoS). As private for-profit entities, ASes seek compensation for their traffic delivery services and sign contacts that include an SLA (see Chapter 3). To establish trust and accountability, the SLA determines the interconnection type, service conditions, compensation arrangements, and penalties for contractual violations.

During the same transition in the 1990s, the main protocol for interdomain routing changed from the centralized Exterior Gateway Protocol (EGP) [167] to the distributed BGP [92]. Using BGP, an AS independently decides which traffic-delivery routes the AS offers to its neighbors. The AS announces the routes to only those neighbors with whom the AS has bilateral interconnection agreements, even when the available physical connectivity is richer.

While the distributed BGP realization of rigid bilateral contracts has promoted the Internet growth, the traditional interdomain routing suffers from numerous drawbacks exacerbated by the rapid Internet evolution. These drawbacks include route instability, convergence slowness, policy inflexibility, and configuration complexity [193, 202]. Although extensive research efforts have mitigated some of these issues through protocol improvements, new contract types, and new interconnection facilities, the foundations of Internet routing remain mostly the same.

While BGP has been the subject of much controversy and has suffered many modifications, we contend that routing challenges in the Internet are fundamental. We contend that some fundamental problems of Internet interdomain routing arise due to its contractual model, rather than its BGP realization. Furthermore, the bulk of work on interdomain routing focuses on improvements in protocols, rather than in the contractual system that the protocols realize. In the Internet of many independent infrastructure providers, efficient end-to-end traffic delivery needs coordination among multiple ASes who may not trust each other.

The current contractual system deals with the problem of trust by relying on rigid bilateral contracts between neighboring ASes. However, the transitive trust of the bilateral contracts comes at the expense of routing flexibility and efficiency. Without explicit means for direct coordination among multiple networks, the realized routing tends to utilize the Internet communication capa-

city suboptimally, oscillate needlessly, and react inappropriately to traffic-demand changes and infrastructure failures.

A potential alternative to this status quo is a centralized system, where a trustworthy entity determines optimal paths based on AS policies. However, this leads to a few concerns: why should independent ASes support such a centralized system? What is the incentive for ASes to sacrifice their routing independence?

This chapter proposes *Route Bazaar*, a contractual system where ASes and their customers agree on QoS-aware routes without explicit or implicit trust between the networks. The only trusted entity in Route Bazaar is a public ledger, which is assumed to be trustworthy in the absence of large corrupted coalitions. While prior work on cryptocurrencies showed how to construct a decentralized public ledger without a single trusted component, our use of the ledger enables networks to check the previous record of each participant in a route before agreeing on the route. Because the public ledger of Route Bazaar allows anyone to verify the trustworthiness of a network as a routing provider or customer, the verification mechanism incentivizes honest behavior by all networks and thereby supports effective interdomain routing in an untrusted environment.

5.1 Background

5.1.1 Routing

Current **interdomain routing** relies on bilateral contracts between ASes, which typically establish either transit or peering relationships. In a transit relationship, a customer network pays a provider for reaching the global Internet. Networks can also bypass transit providers and instead interconnect directly through a peering relationship. In peering, two networks obtain reachability to a restricted set of Internet addresses: peers only exchange traffic destined to their own networks and to networks for which they provide transit.

An AS uses BGP to exchange reachability information with neighbors with which it has bilateral contracts. An AS has limited control over the routes of inbound traffic; for outbound traffic, when multiple neighbors offer paths to the same destination, the AS can choose the neighbor through which to forward. Most ASes prefer to send outbound traffic through interconnections that generate revenue (*i.e.,* through transit customers), and avoid using transit providers if an alternate path through a customer or peer exists.

BGP is fully decentralized and is thus extremely scalable. However, this scalability comes at the expense of flexibility and responsiveness. In particular, events such as routing policy changes, misconfigurations, and infrastructure failures can trigger path oscillations and forwarding loops [90], and it is nearly impossible to verify that selected paths conform with global routing policies [81]. For example, BGP hijacking has in the past been used to redirect YouTube's traffic to a fake destination in Pakistan [19, 166], and to redirect Spamhaus's traffic to a hacker group [174]; in both of these cases a more global view would have alerted ASes to these problems.

The rapid growth of the Internet, both in terms of users and traffic volumes, and the increasing diversity of Internet services have strained the existing routing framework. To cope with these challenges, new types of interconnection contracts have emerged (*e.g.,* partial transit [191], paid peering [57] and remote peering [36]), IXPs (switching facilities to reduce the costs of peering) have mushroomed, and BGP has slowly evolved.

However, many of the limitations of Internet routing stem from two root causes. First, contracts must be explicitly negotiated by humans before two ASes can exchange traffic. Thus, in stark contrast to the dynamic nature of the Internet itself, interconnection negotiations are carried out at human-time scales (sometimes via social events for engineers as mentioned earlier). Second, these contracts are applied recursively: traffic that an AS sends to its neighbor is then governed by the contracts of that neighbor. The local (and thus recursive) routing decisions made by BGP are globally suboptimal due to limited visibility [124].

5.1.2 Cryptocurrencies

Cryptocurrencies, like Bitcoin [141], are secure, decentralized, anonymous digital currencies. These currencies are often built on a public ledger, commonly referred to as a block chain. The public ledger records transactions and enables checking of the account balance for each user. Cryptocurrencies have been adopted to new non-monetary uses [45,73,77] that leverage the public ledger as a decentralized and hard-to-corrupt log.

The public ledger needs to be resilient (*i.e.,* incorruptible) even in the presence of untrusted participants. Bitcoin's public ledger is therefore built using a consensus algorithm that is capable of solving the Byzantine consensus problem [13]. Byzantine consensus is impossible to solve in asynchronous systems in general and requires at least two thirds of participants be honest. Moreover, it assumes that the set of participants is well known. Bitcoin solves this latter problem (that of limiting who can participate) by requiring each participant in the consensus algorithm to solve a puzzle before voting, and attach a proof-of-work [137] to each vote. Generating this proof requires the participant to solve a sufficiently hard algorithmic problem. Since generating this proof requires computational resources, it ensures that the number of votes a malicious user can generate is in proportion to the computing power they control. Since voting multiple times is hard, a malicious user needs to instead convince a majority of participants to affect the result of the consensus algorithm. The lack of aligned interests among malicious parties therefore allows all users to trust the values stored in the block chain.

5.2 Route Bazaar

In this section, we present Route Bazaar, a novel system for flexible Internet connectivity. Inspired by cryptocurrencies, Route Bazaar uses a decentralized public ledger to allow mutually distrustful ASes and customers to establish dynamic, end-to-end QoS-aware paths as overlays on the existing Internet. In Route Bazaar, the information contained in the public ledger allows

Provider	Pathlet	Destination	Price	Latency SLO	Throughput SLO
AS1	f48d4c4	AS2	$50	5 ms (99.9%)	3 Gbps (99.9%)
AS1	d7228c5	AS3	$45	5 ms (99.9%)	3 Gbps (99.9%)
AS2	97dbd13	AS9	$10	10 ms (99.9%)	1 Gbps (99.9%)
AS3	ca22b8a	AS9	$20	8 ms (99.9%)	2 Gbps (99.9%)

Table 5.1: Pathlet advertisements in the public ledger.

each participant to verify any participant's conformance with previous path agreements, while simultaneously keeping the path agreements private. By relying on a public ledger, Route Bazaar establishes trust among participants which can compute the likelihood that another participant will honor a path agreement. Route Bazaar uses standard cryptographic tools to ensure privacy and existing techniques to establish overlays; our innovation lies in creating a trustworthy environment for the announcement, selection, and verification of end-to-end paths. In what follows, we assume that all communications with the public ledger are carried out through authenticated and encrypted channels, *i.e.,* entities cannot impersonate each other. Several existing protocols (*e.g.,* TLS [59]) can be used to meet this requirement.

In Route Bazaar, a *provider* is an AS that advertises connectivity over a *pathlet* (*i.e.,* a path fragment [80]). A *path* is formed by composing pathlets leading from a *source* to a *destination*. A *customer* in Route Bazaar is an entity paying for the end-to-end connectivity provided by a path. A customer may not be an AS, and might be either the source or destination of the path.

An important aim for Route Bazaar's design is to minimize the amount of information leaked about end-to-end paths and customer policies. Our design limits public information to the minimum required to support flexible routing via multilateral contracts. First, providers advertise pathlets using the public ledger. Then, customers compose end-to-end paths by combining the advertised pathlets. Finally, providers confirm or reject the agreement. We envision that these decisions are made dynamically by automated agents acting on behalf of customers and providers. Participants can hence enforce sophisticated contractual and routing policies. For instance, Route Bazaar allows these policies to exploit the historical records about forwarding performance (to choose providers) and likelihood of payment (to accept a customer) that are maintained in the public ledger. When all participants agree on a path, the public ledger records an agreement between the customer and each provider participating in the path.

As the traffic is forwarded along a computed path, the source, destination and each provider record machine-readable forwarding proofs in the public ledger. These proofs can then be used to verify that each provider delivered the desired level of service. Customers also record proofs showing that they have paid providers in the public ledger. The public ledger therefore allows potential customers and providers to compare previous performance and payment history when deciding whether or not to trust each other.

Provider	Tag	Latency SLO	Throughput SLO
AS1	enc_m(f48d4c4)	5 ms (99.9%)	1 Gbps (99.9%)
AS2	enc_m(97dbd13)	10 ms (99.9%)	1 Gbps (99.9%)

Table 5.2: Pathlet commitments table in the public ledger. $enc_m(x)$ here represents the value output by a PRF with key m and value x.

5.2.1 Routing

Providers advertise *pathlets* in the public ledger (Table 5.1). Each pathlet advertisement specifies a tag (used to refer to the pathlet), the source, destination, price and Service Level Objective (SLO) for throughput and latency. For ease of exposition, here we assume that the pathlet provider is also the source, however Route Bazaar supports pathlets where the source and provider differ.

A customer composes an end-to-end path between a source and destination by choosing from the set of advertised pathlets. The customer can enforce routing policies by filtering out policy-incompatible pathlets. For example, a customer can exclude pathlets advertised by providers who have previously not met their SLOs. Before a path can be used, each provider must agree to route traffic along the path; a provider can thus disallow the use of policy-incompatible paths. For example, a provider might deny service to customers who are unlikely to pay, or reject paths involving untrusted providers. Policies are enforced by automated agents, who act on behalf of providers and customers (and are hence aware of their policies) and can exploit the information contained in the public ledger to judge other participants.

Once the customer and all pathlet providers have agreed on a path, the participants use a symmetric key generated via key agreement (*e.g.,* Elliptic Curve Diffie-Hellman (ECDH) [22]) and the pathlets' tags as input to a cryptographic pseudorandom function (PRF) to generate an anonymous tag that is valid for only this specific path. Each provider then publishes a pathlet agreement which includes this encrypted tag, the provider's identity and the SLO offered by the pathlet to a pathlet commitments table (Table 5.2) in the public ledger. We use path agreement to refer to the collection of all pathlet agreements that allow routing along a path. The customer and providers also agree on a second key that is used to generate an anonymized payment tag (again derived from the pathlet tag) and prices that are used to record a set of payment agreements (Table 5.3) between the customer and providers.

This mechanism can accommodate a variety of end-to-end routing models including multi-path routing [197], source routing [158], opportunistic routing [158] and route repositories [29]. Route Bazaar also allows customers to outsource path computation to trusted third parties, *i.e.,* routing as a service [112]. Finally, Route Bazaar supports contractual flexibility, *e.g.,* it can accommodate both cases where the sender pays for connectivity and cases where the receiver pays for connectivity.

To illustrate how customers can use Route Bazaar to form an end-to-end path, consider a case where Alice wants to route traffic from a source in AS1 to a destination in AS9. Alice uses

Customer	Pathlet	Payment
Alice	enc_n(f48d4c4)	enc_n($50)
Alice	enc_n(97dbd13)	enc_n($10)

Table 5.3: Payment commitments in the public ledger. $enc_n(x)$ here represents the value output by a PRF with key n and value x.

Pathlet	Packet	Hash	Timestamp	Throughput
enc_m(source|A6)	50	0a9f136	420 ms	1.2 Gbps
enc_m(f48d4c4)	50	0a9f136	424 ms	1.2 Gbps
enc_m(97dbd13)	50	0a9f136	433 ms	1.1 Gbps
enc_m(destination|A9)	50	0a9f136	433 ms	1.1 Gbps

Table 5.4: Forwarding proofs in the public ledger. $enc_m(x)$ here represents the value output by a PRF with key m and value x.

the pathlet announcements in Table 5.1 to find two possible paths that provide this connectivity: AS1-AS2-AS9 and AS1-AS3-AS9. While the path through AS2 is cheaper ($60 vs $65), the path through AS3 offers better latency (13ms vs 15ms). In this example, Alice's policy favors the cheapest path[1], and she therefore decides to route along path AS1-AS2-AS9. If AS1 and AS2 agree to form a path, the three (Alice and both ASes) use ECDH to compute keys m and n. AS1 and AS2 use a PRF with key m to generate anonymized tags, and then update the pathlet commitments table in the public ledger (Table 5.2). Similarly, Alice uses key n to compute anonymized payment tags, and encrypt prices, and updates the payment commitment table (Table 5.3).

Note that while in the previous example, Route Bazaar allows Alice to exclude paths going through AS3, this is not generally possible in BGP, where all traffic originating at AS1 and destined to AS9 follows the same path (which might in fact go through AS3). Route Bazaar thus provides Alice with additional routing flexibility, allowing her to choose paths based on a richer set of policies.

5.2.2 Forwarding

Once an end-to-end path has been agreed, providers update the data plane as required. Route Bazaar provides the mechanisms to form and agree on paths as well as to verify that forwarding conforms with the agreed paths. To verify path conformance, Route Bazaar generates *forwarding proofs* that are recorded in the public ledger. Customers, providers and intermediate ASes along a path use existing techniques for traffic sampling at routers to periodically generate a forwarding proof. The forwarding proof for a pathlet includes the path-specific anonymized forwarding tag (a pathlet might be used by several paths), a sample of the traffic, the timestamp indicating when the sample was captured, and the throughput averaged over the time since the last sample. In our current design, we envision that each provider sets up Generic Route Encapsulation (GRE)

[1] Alice could have decided using other policies, *e.g.,* prior history if available, or any other reasons.

Pathlet	Paid
enc_n(f48d4c4)	Yes
enc_n(97dbd13)	Yes

Table 5.5: Payment proofs in the public ledger. $enc_n(x)$ here represents the value output by a PRF with key n and value x.

tunnels [91] across each pathlet (to ensure in-order packet transit) and samples a particular packet (*e.g.,* the 50^{th} packet in Table 5.4). The hash of this packet is used as a traffic sample for the forwarding proof.

Note that the inclusion of timestamps allows participants to compute the latency reported by a pathlet's ingress and egress neighbors. Furthermore, the participants in a path (*i.e.,* the customer, source, destination and pathlet providers) can use their path key to discover bottlenecks in the path by observing where a (sampled) packet was dropped. To preserve the anonymity of a path, this information is not available to non-participants.

When a path's agreement concludes[2], each provider is paid by the customer, and the provider registers a payment conformation in the public ledger (Table 5.5). The payment proof includes the pathlet's anonymized payment tag and a field indicating whether the payment was made. Alternately, the customer can record its unwillingness to pay in the public ledger, indicating that appropriate connectivity was not provided.

The payment proofs in the public ledger enable anyone to check customers' payment history. These records can also be used for offline arbitration of payment disputes. During such arbitration, the entities involved in the contract can present the arbitrator with a deanonymized version of the forwarding and payment proofs.

In the example above, the participants, *i.e.,* the source in AS1, pathlet providers AS1 and AS2, and the destination in AS9, sample every 50th packet and publish forwarding proofs as shown in Table 5.4. These ASes rely on NTP (Network Time Protocol) for clock synchronization to ensure reported times are comparable. Once the path agreement has concluded, Alice pays AS1 and AS2, and they record her payment in the public ledger as shown in Table 5.5.

5.2.3 Privacy

The privacy offered by Route Bazaar is comparable to BGP. Similar to BGP, Route Bazaar reveals available paths (as all possible path compositions). This information is identical to what is available in public repositories, *e.g.,* CAIDA [52]. Furthermore, Route Bazaar does not require providers or customers to reveal routing preferences and policies.

However, unlike existing interdomain routing solutions, Route Bazaar also reveals anonymized forwarding and payment proofs. If deanonymized (*e.g.,* due to a compromised participant) these proofs expose the precise paths used by customers and the volume of transferred

[2]In our current design, *path agreements* are for a fixed volume of traffic.

traffic. Similar information can be revealed today by sufficiently powerful adversaries (*e.g.,* governments or large ASes). Existing mechanisms, *e.g.,* Tor [62] and Unblock [169], for anonymizing source and destination addresses can be used on top of Route Bazaar to provide stronger anonymity.

Finally, the forwarding commitments table in Route Bazaar leaks information about which pathlets are popular, and the amount of overall traffic transmitted across a pathlet. Similarly, the payments commitment table leaks information about the number of agreements made by each customer. Route Bazaar can be extended to anonymize this information, so that contracts are established and verified out-of-band, with Route Bazaar merely serving as a record of past performance. Studying this extension and its properties is left to future work.

5.3 Discussion

Performance Overhead. Route Bazaar imposes minimal overhead on the data plane, requiring only that routers periodically sample traffic. This feature is commonly supported in most routers (to aid in debugging), and Route Bazaar does not require the samples to be transmitted in real-time. The communication overhead imposed by requiring ASes to record forwarding proofs with the public ledger is relatively small and can be controlled by adjusting the sampling rate. Furthermore, our current design also requires the use of GRE tunnels, these are supported by most existing interdomain routers.

Because decision making in Route Bazaar is not local, routers merely forward traffic and are not responsible for control-plane decisions. Instead, control-plane decisions can be done externally by computers, or at cloud computing facilities. The operations required to access and update Route Bazaar today is comparable to what is performed by a modern web browser when connecting to a website over HTTPS [163]. The primary computation overhead during path computations is therefore a function of the policy complexity, and Route Bazaar's control plane therefore imposes modest performance overheads.

Sybil Attacks. Participants can circumvent the trust mechanisms of the public ledger by creating pseudonymous identities, *i.e.,* they can perform a Sybil attack [64]. While ASes and large organizations (who are the main participants in Route Bazaar) are unlikely to jeopardize their reputation by forging their identities, Route Bazaar can protect against Sybil attacks by adopting existing solutions, *e.g.,* SybilGuard [203].

Rich routing policies. Route Bazaar can further enrich its supported routing policies, *e.g.,* by linking pathlet prices to the customers' requested traffic volume, payment history, or other conditions. To attract customers, the pathlets can also expose specific salient features of the provided connectivity, *e.g.,* its Software Defined Network (SDN) implementation. While Route Bazaar separates routing from forwarding, the main challenge in enriching the routing policies is not their storage or processing but their expression in a machine-readable language.

Backward compatibility. Because Route Bazaar can operate on top of today's Internet, it is

backward compatible with traditional bilateral contracts and BGP routing. Still, Route Bazaar diversifies contractual options, *e.g.,* by enabling IXP members to exchange traffic not only through traditional peering agreements but also with contracts formed dynamically via the public ledger.

5.4 Conclusions

The current Internet relies on explicitly negotiated bilateral agreements that are recursively applied via BGP, leading to rigid functionality and suboptimal routing behavior. In this chapter, we propose Route Bazaar, an alternative that learns from cryptocurrencies to solve the decentralized trust problem inherent in connectivity contracts. Route Bazaar forms contracts for end-to-end Internet connectivity orders of magnitude faster and supports highly flexible routing.

Chapter 6

Related Work

The previous chapters already mentioned the essential background information. This chapter takes a broader look at related work.

6.1 Cost reduction techniques

A perennial challenge for ASes has been the unrelenting traffic growth [47, 110] characteristic for the Internet since its inception. Complicating things further, the main application responsible for this growth has been changing from web browsing [65] to P2P file sharing [181] to video streaming [152]. To reduce the transit costs of this skyrocketing traffic, a large number of approaches have emerged. The existing approaches for reducing the transit costs include AS peering, IXPs, IP multicast, CDNs, P2P localization, and traffic smoothing.

Peering [15, 56, 120] enables two ASes to exchange their traffic directly, rather than through a transit provider at a higher cost. IXPs facilitate peering [1, 15, 32, 42, 43] by providing a common infrastructure for the traffic exchange. To disseminate data to multiple receivers, IP multicast [18, 28, 54, 82, 83] duplicates packets in IP routers and thereby reduces transit traffic. While IP multicast requires router support from transit providers, CDNs [94, 100, 116, 146, 156, 182] and P2P systems duplicate data on the application level. Whereas a single company controls a CDN, a P2P system [149] consists of independent hosts, and P2P localization [49,113,201] strives to reduce transit traffic without undermining the system performance. Even if the transit traffic preserves its volume but is redistributed within the billing period to peak at a lower value, the transit costs decrease due to the burstable billing [61]. An AS can do such traffic smoothing [101] with rate limiting [134] or in-network storage for delay-tolerant traffic [114]. Remote peering (Chapter 2) is an emerging approach to deal with high transit costs that allows distant networks to interconnect at the main Internet hubs.

6.2 Cost sharing

CIPT (Chapter 3) and T4P (Chapter 4) are two novel interconnection types that leverage the economies of scale in traffic delivery in general and IP transit in particular. While CIPT reduces the transit costs without altering traffic, T4P reduces the transit costs by redistributing the traffic over the transit links of the two ASes involved.

We view CIPT as a coalition and use the Shapley value [173] for sharing CIPT costs. Shair [97] and [33] are cooperative systems that enables phone users to share the committed but unused minutes and broadband capacity, respectively. Cooperative approaches have also been studied for cost sharing in IP multicast [11, 72] and interdomain routing [132, 175, 176]. The game-theoretic analyses of the Shapley-value mechanism [11, 72, 139] highlight its group-strategyproofness and other salient properties but identify its high computational complexity. Despite the computational complexity, various proposals of traffic billing between ASes [126, 127], incentives in P2P systems [138], and charging individual users by access ASes [180] rely on the

Shapley value. Our evaluation of CIPTs uses the Monte Carlo method to estimate the Shapley value accurately at a reduced computational cost [118].

6.3 Internet structure

Remote peering, CIPT, and T4P should be viewed in the context of the evolving Internet ecosystem. With the traffic rising, ASes stablish peering interconnections and bypass transit providers to reduce transit costs. The Internet has hence developed from a fundamentally hierarchical net and flattened due to the pervasive peering [56, 78, 121]. In parallel to the growing diversity of ASes [70], new forms of interconnections have emerged, e.g., partial transit [191], paid peering [57], and remote peering [36]. While previous research studies only mention remote peering [1, 43, 87, 159, 178], we are the first to closely examine this emerging type of network interconnection. Our results show wide spread, significant traffic offload potential, and conditions for economic viability of remote peering.

While resource allocation and corresponding cost recovery in a distributed environment is a complex matter in general [48], the Internet structural evolution complicated the situation even further. Although tussles affected the Internet from its early days [51], the growing diversity together with the widespread peering resulted in frequent disagreements over peering cost allocations. Traffic imbalances [143] in peering relationships led to demands of monetary compensations [58] and to a large set of techniques for minimizing costs and maximizing revenues [20, 145, 180]. The tensions were specially acute between networks that primarily connect residential users and those networks that connect content providers. Such rifts caused so-called peering wars [10, 25, 26, 154]. These tussles resulted in heated debates about network neutrality and the role of paid peering in it [53, 55, 75, 88, 125, 129, 184, 200, 205]. Instead of allowing content provides to subsidize the access of users to the Internet [128], or imposing a fee based on the different revenue or cost structure as in [125, 140], T4P exploits the different cost structure of heterogeneous ASes and succeeds in reducing overall costs without monetary payments.

The Internet structure is highly important for understanding these tussles as well as network accountability [8], multihoming [3], routing security [79], traffic delivery economics [20, 21], Internet governance [178] and various problems in content delivery via overlay systems [31, 66, 94, 100, 104, 130, 138, 182, 204]. However, understanding the Internet structure is a complex task. For instance, while the arguments that the Internet structure becomes flatter are multifaceted [57, 78, 110], remote peering reveals separation between the trends of increasing peering and Internet flattening. Also, while analyses of interconnection options are typically restricted to networks that share a location [41, 172], remote peering enables distant networks to peer over a layer-2 intermediary. While the layer-3 level topologies are known to be inaccurate [147, 198], remote peering stresses the need for more realistic topologies.

6.4 Measurement methods

To complicate things further, real interdomain data is difficult to obtain. Because network operators do not publicly disclose connectivity of their networks, the research community relies on measurements and inference to characterize the Internet structure [89]. A prominent means for the topology discovery is the traceroute tool that exposes routers on IP delivery paths [40,44]. For example, iPlane [131] and Hubble [103] use traceroute to generate and maintain annotated Internet maps. Paris traceroute enhances traceroute with the ability to discover multiple paths [14]. A complementary approach is to utilize BGP traces [9, 76, 93, 179, 183].

Chapter 2 uses active probing in the data plane to understand the role of remote peering in the Internet structure. Delay measurements are common in Internet studies, e.g., to study the evolution of Internet delay properties [115] or Internet penetration into developing regions [69,87]. We use delay measurements to investigate how geography affects peering of networks and real traffic data from RedIRIS [160] to evaluate the potential impact of remote peering at the network level. To evaluate CIPT and T4P, Chapters 3 and 4 take a different approach and approximate transit traffic by using the peering traffic data obtained by transforming mrtg images published by six IXPs.

6.5 Routing architectures

The lack of path diversity [50, 190], slow convergence of routing protocols at both the intradomain [23, 185] and interdomain level [68, 85, 108, 109, 133, 148], and the routing table growth [27, 98, 136] undermine the scalability of Internet routing [202] and threaten the ability of the Internet architecture to support future innovations.

Both researchers and practitioners grew increasingly concerned and frustrated with these and other rigidities inherent to the Internet architecture and proposed alternative frameworks. Route Bazaar belongs to the recent body of work on novel Internet architectures that overcome fundamental shortcomings of the current Internet.

Pathlet routing [80], ICING [142, 170] and Platypus [157] are the proposals most closely related to Route Bazaar. Similarly to segment routing [171], Pathlet routing proposes a mechanism where the source can construct paths by composing path segments called pathlets. Route Bazaar builds on the concept of composable pathlets, but also provides a discovery mechanism. ICING and Platypus provide different mechanisms for verification of path validity. These mechanisms can be used to provide proofs in Route Bazaar. As opposed to such schemes, the use of the public ledger allows Route Bazaar to be used by untrusted networks without any data-plane modifications.

Similar to SDN, Route Bazaar also decouples the control and data planes. Previous works, including RCP [71] and 4D [84], suggested such separation for interdomain routing. As opposed to these proposals, Route Bazaar allows networks to retain private control of their control plane

and policies. Route Bazaar also enables previous proposals such as routing as a service [111], opportunistic routing [158], multipath routing [197], and route repositories [29].

Finally, Kadupul [177] proposes a cryptocurrency-based mechanism to enable the routing of delay-sensitive traffic in wireless mesh networks. Kadupul relies on a proof-of-work mechanism implemented with time-locked puzzles to incentivize nodes to forward traffic. Unlike Route Bazaar, Kadupul does not attempt to provide mechanisms for path-discovery and cannot be easily extended to account for QoS.

Chapter 7

Future work

We see two general directions for future work. First, while previous chapters examined remote peering, CIPT, T4P, and Route Bazaar, a further analysis of these proposals is worthwhile. Second, whereas this thesis studied the economic aspects of Internet engineering at the interconnection level, these issues are intertwined with other social, cultural, and political factors that need to be considered. The given chapter discusses these two general directions for potential extensions to our prior research.

Remote peering challenged the reliance on layer-3 methods and data to infer the economic structure of the Internet. Future research will strive to improve topological studies of the Internet and refine the identification of remote peers. By integrating both layer-2 and layer-3 perspectives, we will provide a more realistic understanding of the interconnection structure of the Internet. Geolocation can refine our identification of remote peers as well as clarify the regional patterns of interconnections and content hosting [2]. However geolocation of remote peers is not straightforward. Aggravating the general lack of accuracy of standard IP geolocation databases [153], such databases fail to identify the actual location of remotely peering networks. Instead, the IP interface of a remote AS in the IXP subnet appears colocated with such IXP. Triangulation techniques combining topological and delay data [102] constitute a promising technique to reveal the real location of remote peers.

CIPT proposed a transit-cost reduction mechanism and our future research will look at its implementation and implications. While this thesis envisioned a simple organizational arrangement where the Shapley-value mechanism redistributes CIPT gains, future work will quantify the overhead created by CIPT and explore alternative solutions for distribution of CIPT gains. We will examine whether CIPT can help to reduce the costs of transit in developing countries. Regions located far away from the main hubs of the Internet typically face high transit prices [69]. To make things worse, predominant consumption of non-local content increases the transit costs further. Despite the emergence of IXPs in developing countries [4], the low volumes of local traffic decrease the potential benefit of peering [87]. By aggregating transit traffic, CIPT can improve Internet access in the poorly connected areas.

T4P can alleviate the causes leading to peering wars and net-neutrality debates. Net-neutrality is a complex issue [53] with multi-dimensional implications. Regional specifics such as population density, existing laws and infrastructure crucially impact the formulation of the issue. As the Internet evolves, changing traffic patterns or lower costs of the last-mile infrastructure, might reduce the tensions between content and eyeball networks. As online video games, live-streaming, social media, the Internet of Things (IoT), and cloud storage services gain popularity, the asymmetry between eyeball and content networks might decrease. New technologies might reduce last-mile costs and increase competition in the network-eyeball interconnection market. Our future work will study the causes and arguments around the net-neutrality debate. We will also evaluate the extent to which T4P can ease the tensions and whether regulatory solutions are required.

Route Bazaar is an outline of a novel Internet architecture and certainly deserves further

research elaborating on its design and implementation. This thesis presented a mechanism for verification of contracted paths in this new architecture. Future work will develop the rest of the architectural elements, such as the distributed ledger, and integrate this mechanism in it. An incentive structure and cryptographic tools to implement the architecture will be also part of our future work. For backwards compatibility, we will design Route Bazaar as an overlay on the top of the current Internet architecture.

While the aforementioned extensions of this thesis dwell in the realm of the economic and technical aspects of the Internet, the emergence of the Internet affects many other areas and is part of a historical process where social, political, and cultural factors are also fundamental [35]. These issues must be taken into account to have a holistic understanding of the Internet evolution and its implications. For instance, social and cultural aspects determine to a great extent traffic flows: e.g., whereas Japan consumes mostly local content, most African countries consume content created and located abroad [2, 69, 87]. This lack of local content leads to a vicious circle of inadequate incentives for infrastructure deployment, Internet provision improvement, and local content creation and hosting. Accordingly, the resulting Internet access inequality has a complex mix of factors [46] that requires interdisciplinary studies [34, 196].

Also, the change in the applications fueling the traffic growth have social, cultural, and political underpinnings and implications. As these applications shifted from web browsing, to P2P file sharing, video streaming and more recently to online video games, live-streaming, social media, and cloud storage, the traffic patterns and costs changed, and new agents emerged. This dynamism challenges the current Internet architecture. Exploring how the bottom-up approach of the Internet governance [178] can match this dynamism is also a multi-dimensional endeavor that spans several disciplines. We look forward to integrating these views in future research.

Chapter 8

Conclusions

This thesis analyzes how networks interconnect in the current Internet and which limitations its interconnection framework has. In doing so, the thesis looks at the evolution of the Internet economic structure and investigates four different but related topics: remote peering, CIPT, T4P, and Route Bazaar.

The early Internet was a publicly owned infrastructure with a relatively small number of connected networks and end users. Since its inception in the 1960s, the Internet has developed into a massive decentralized ecosystem providing end-to-end connectivity for a multitude of diverse users and services. As the traffic grew, the ASes' attempt to curb rising transit costs transformed the Internet. Evolving from a predominantly hierarchical structure where few networks provided universal connectivity, the Internet became a flatter structure with direct peering connections bypassing transit providers. Because expanding the network infrastructure to connect through peering links was financially burdensome, large content providers pioneered this trend.

Remote peering emerged as a cost-efficient solution where an AS peers with distant networks without having to bear the costs of physical presence at a common location. In revealing the spread of remote peering, this thesis exposed a long-term trend of the Internet ecosystem towards more flexible interconnections. At the same time, by unveiling the wide adoption of remote peering, we demonstrated the shortcomings of the typical layer-3 modeling of the Internet.

With more and more networks joining the Internet, interdomain interconnections struggled with increasingly heterogeneous interconnection needs and unrelenting traffic growth. CIPT and T4P are two novel interconnection types that continue the trend of interconnection diversification.

CIPT presents an alternative way to reduce transit costs. While most interdomain interconnections are bilateral, CIPT is a multilateral and cooperative approach that reduces transit costs for all the ASes involved. CIPT savings result from two properties of the IP transit model: price subadditivity and burstable billing. To redistribute the savings fairly we use the game-theoretic concept of Shapley value. By benefiting all the ASes participating in a CIPT, this interconnection can appeal to ASes of many different types.

T4P is a mechanism that leverages the heterogeneity of networks to reduce the interconnection costs. The difficulties to satisfy the frequently diverging needs for the plethora of actors comprising the Internet, triggered recurrent economic conflicts over the settlement of interconnections. In particular, the highly asymmetric traffic patterns between eyeball and content networks sparked disputes about whether one network should compensate the other for their peering. This conflict provoked frequent peering disruptions and ultimately resulted in the network neutrality debate. T4P alleviates the pressures leading to this conundrum: by redistributing the traffic of two ASes over their respective transit links, T4P reduces the overall interconnection costs of the two networks. The cost reduction can then be shared or allocated to the network demanding a compensation for the peering.

While remote peering, CIPT, and T4P strengthen Internet connectivity and reduce frictions and financial burdens, some of the limitations of the Internet are inherent to its current architecture. The flexibility of the current Internet architecture is limited by the very same features that

enabled its scalability: distributed routing via local decision making (BGP) and rigid interconnection contracts.

Route Bazaar takes a radical departure to overcome the drawbacks of the current Internet architecture. Inspired by the use of block chains and cryptographic tools in cryptocurrencies, Route Bazaar proposes a new Internet architecture that provides flexible interconnections, supports rich interconnection policies, and accommodates automatic multilateral contract establishment, termination and enforcement.

This thesis sheds light on the continuing Internet evolution with respect to the interconnection diversification and overall structure changes. While remote peering is a phenomenon that has been massively emerging in the Internet over the past decade, CIPT and T4P are two novel interconnection types that continue the diversification trend toward more flexible interconnections. Route Bazaar goes one step further and tackles the inherent limitations of the current interconnection framework by replacing it with a novel Internet architecture. However remote peering, CIPT, T4P, and Route Bazaar do not exhaust the possibilities for addressing the challenges faced by the Internet. For instance, novel interconnections might appear, and SDN-enabled routing [17,106,192] or new technologies (e.g., 5G) might ease some of the existing drawbacks. At the same time, some current challenges might fade away while new ones emerge. As the Internet evolution keeps changing the applications responsible for the traffic growth, the Internet economic structure will likely evolve to adapt to it. For example, online video games, live-streaming, social media, IoT, and cloud storage services can reduce the asymmetry between eyeball and content networks. As new challenges appear, and the Internet evolves to cope with them, new techniques to understand those changes and new solutions to address the bottlenecks will be necessary. While the challenges are many and have political, social, cultural, and economic aspects, this thesis contributed to the understanding of the current Internet economic structure and proposed solutions to alleviate the existing drawbacks. Looking into political, social and cultural aspects will be necessary to address the new and existing challenges.

References

[1] B. Ager, N. Chatzis, A. Feldmann, N. Sarrar, S. Uhlig, and W. Willinger. Anatomy of a Large European IXP. In *Proceedings of SIGCOMM*, 2012.

[2] B. Ager, W. Mühlbauer, G. Smaragdakis, and S. Uhlig. Web Content Cartography. In *Proceedings of IMC*, 2011.

[3] A. Akella, B. Maggs, S. Seshan, A. Shaikh, and R. Sitaraman. A Measurement-based Analysis of Multihoming. In *Proceedings of SIGCOMM*, 2003.

[4] C. Amega-Selorm, M. Mureithi, D. Pater, and R. Southwood. Internet Exchange Points. Their Importance to Development of the Internet and Strategies for their Deployment, The African Example. Global Internet Policy Initiative, 2004.

[5] C. Amega-Selorm, M. Mureithi, D. Pater, and R. Southwood. Impact of IXPs - A Review of the Experiences of Ghana, Kenya and South Africa. Open Society Institute, 2009.

[6] Amsterdam Internet Exchange (AMS-IX). `https://www.ams-ix.net`.

[7] K. S. Anand and R. Aron. Group Buying on the Web: a Comparison of Price-discovery Mechanisms. *Management Science*, 49(11):1546–1562, 2003.

[8] D. G. Andersen, H. Balakrishnan, N. Feamster, T. Koponen, D. Moon, and S. Shenker. Accountable Internet Protocol (AIP). In *Proceedings of SIGCOMM*, 2008.

[9] D. G. Andersen, N. Feamster, S. Bauer, and H. Balakrishnan. Topology Inference from BGP Routing Dynamics. In *Proceedings of IMC*, 2002.

[10] N. Anderson. Peering Problems: Digging into the Comcast/Level 3 Grudge-match. `http://arstechnica.com/tech-policy/news/2010/12/comcastlevel3.ars`, 2011.

[11] A. Archer, J. Feigenbaum, A. Krishnamurthy, R. Sami, and S. Shenker. Approximation and Collusion in Multicast Cost Sharing. *Games and Economic Behavior*, 47(1):36–71, 2004.

[12] Atrato IP Networks. `https://www.atrato.com`.

[13] C. Attiya, D. Dolev, and J. Gil. Asynchronous Byzantine Consensus. In Proceedings of PODC, 1984.

[14] B. Augustin, X. Cuvellier, B. Orgogozo, F. Viger, T. Friedman, M. Latapy, C. Magnien, and R. Teixeira. Avoiding Traceroute Anomalies with Paris Traceroute. In *Proceedings of SIGCOMM*, 2006.

[15] B. Augustin, B. Krishnamurthy, and W. Willinger. IXPs: Mapped? In *Proceedings of SIGCOMM*, 2009.

[16] Bahrain Internet Exchange. `http://www.bix.bh`.

[17] J. Bailey, R. Clark, N. Feamster, D. Levin, J. Rexford, and S. Shenker. SDX: A Software Defined Internet Exchange. In *Proceedings of SIGCOMM*, 2014.

[18] T. Ballardie, P. Francis, and J. Crowcroft. Core Based Trees (CBT). *CCR*, 23(4):85–95, 1993.

[19] P. Bangera and S. Gorinsky. Impact of Prefix Hijacking on Payments of Providers. In *COMSNETS*, 2011.

[20] P. Bangera and S. Gorinsky. Impact of Prefix Hijacking on Payments of Providers. In *Proceedings of COMSNETS*, 2011.

[21] P. Bangera and S. Gorinsky. Economics of Traffic Attraction by Transit Providers. In *Proceedings of Networking*, 2014.

[22] E. Barker, D. Johnson, and M. Smid. Recommendation for Pair-wise Key Establishment Schemes Using Discrete Logarithm Cryptography. *NIST Special Publication*, pages 800–56A, 2007.

[23] Z. Ben Houidi, M. Meulle, and R. Teixeira. Understanding Slow BGP Routing Table Transfers. In *Proceedings of IMC*, 2009.

[24] S. Biernacki. Remote Peering at IXPs. EURONOG 1, 2011.

[25] M. A. Brown, C. Hepner, and A. Popescu. Internet Captivity and the De-peering Menace. In *NANOG 45*, 2009.

[26] M. A. Brown, A. Popescu, and E. Zmijewski. Peering Wars: Lessons Learned from the Cogent-Telia De-peering. In *NANOG 43*, 2008.

[27] T. Bu, L. Gao, and D. Towsley. On Characterizing BGP Routing Table Growth. *Computer Networks*, 45(1):45–54, 2004.

[28] J. Byers, M. Luby, and M. Mitzenmacher. A Digital Fountain Approach to Asynchronous Reliable Multicast. *JSAC*, 20(8):1528–1540, 2002.

[29] M. Caesar, M. Casado, T. Koponen, J. Rexford, and S. Shenker. Dynamic Route Recomputation Considered Harmful. *CCR*, 40(2):66–71, 2010.

[30] M. Calder, X. Fan, Z. Hu, E. Katz-Bassett, J. Heidemann, and R. Govindan. Mapping the Expansion of Google's Serving Infrastructure. In *Proceedings of IMC*, 2013.

[31] F. Cantin, B. Gueye, D. Kaafar, and G. Leduc. Overlay Routing Using Coordinate Systems. In *Proceedings of CoNEXT*, 2008.

[32] J. C. Cardona and R. Stanojevic. IXP Traffic: A Macroscopic View. In *Proceedings of LANC*, 2012.

[33] J. C. Cardona, R. Stanojevic, and N. Laoutaris. Collaborative Consumption for Mobile Broadband: A Quantitative Study. In *Proceedings of CoNEXT*, 2014.

[34] M. Castells. *The Internet Galaxy: Reflections on the Internet, Business, and Society*. Oxford University Press, Inc., 2001.

[35] M. Castells. *The Rise of the Network Society: the Information Age: Economy, Society, and Culture*, volume 1. John Wiley & Sons, 2011.

[36] I. Castro, J. C. Cardona, S. Gorinsky, and P. Francois. Remote Peering: More Peering without Internet Flattening. In *Proceedings of CoNEXT*. ACM, 2014.

[37] I. Castro and S. Gorinsky. T4P: Hybrid Interconnection for Cost Reduction. In *Proceedings of NetEcon*, 2012.

[38] I. Castro, R. Stanojevic, and S. Gorinsky. Using Tuangou to Reduce IP Transit Costs. *ToN*, 22(99):1415–1428, 2014.

[39] CESNET. `http://www.ces.net/`.

[40] H. Chang, S. Jamin, and W. Willinger. Inferring AS-level Internet Topology from Router-Level Path Traces. In *Proceedings of ITCOM*, 2001.

[41] H. Chang, S. Jamin, and W. Willinger. To Peer or Not to Peer: Modeling the Evolution of the Internet AS-level Topology. In *Proceedings of INFOCOM*, 2006.

[42] N. Chatzis, G. Smaragdakis, J. Böttger, T. Krenc, A. Feldmann, and W. Willinger. On the Benefits of Using a Large IXP as an Internet Vantage Point. In *Proceedings of IMC*, 2013.

[43] N. Chatzis, G. Smaragdakis, A. Feldmann, and W. Willinger. There Is More to IXPs than Meets the Eye. *CCR*, 43(5):19–28, 2013.

[44] K. Chen, D. R. Choffnes, R. Potharaju, Y. Chen, F. E. Bustamante, D. Pei, and Y. Zhao. Where the Sidewalk Ends: Extending the Internet AS Graph Using Traceroutes from P2P Users. In *Proceedings of CoNEXT*, 2009.

[45] L. Chen and K. Chen. BitBill: Scalable, Robust, Verifiable Peer-to-Peer Billing for Cloud Computing. In *Proceedings of HotCloud*, 2014.

[46] M. Chinn and R. Fairlie. The Determinants of the Global Digital Divide: a Cross-country Analysis of Computer and Internet Penetration. *Oxford Economic Papers*, 59(1):16, 2007.

[47] K. Cho, K. Fukuda, H. Esaki, and A. Kato. The Impact and Implications of the Growth in Residential User-to-user Traffic. *CCR*, 36(4):207–218, 2006.

[48] S.-W. Cho and A. Goel. Pricing for Fairness: Distributed Resource Allocation for Multiple Objectives. *Algorithmica*, 57(4):873–892, 2010.

[49] D. R. Choffnes and F. E. Bustamante. Taming the Torrent: A Practical Approach to Reducing Cross-ISP Traffic in Peer-to-peer Systems. In *Proceedings of SIGCOMM*, 2008.

[50] J. Choi, J. H. Park, P.-c. Cheng, D. Kim, and L. Zhang. Understanding BGP Next-hop Diversity. In *Proceedings of Global Internet Symposium*, 2011.

[51] D. Clark, J. Wroclawski, K. R. Sollins, and R. Braden. Tussle in Cyberspace: Defining Tomorrow's Internet. *ToN*, 13(3):462–475, 2005.

[52] Cooperative Association for Internet Data Analysis (CAIDA). `http://www.caida.org/`.

[53] J. Crowcroft. Net Neutrality: the Technical Side of the Debate: a White Paper. *CCR*, 37(1):49–56, 2007.

[54] S. Deering. *Multicast Routing in a Datagram Internetwork*. PhD thesis, Stanford University, 1991.

[55] A. Dhamdhere and C. Dovrolis. Can ISPs Be Profitable Without Violating Network Neutrality? In *Proceedings of NetEcon*, 2008.

[56] A. Dhamdhere and C. Dovrolis. The Internet is Flat: Modeling the Transition from a Transit Hierarchy to a Peering Mesh. In *Proceedings of CoNEXT*, 2010.

[57] A. Dhamdhere and C. Dovrolis. Twelve Years in the Evolution of the Internet Ecosystem. *ToN*, 19(5):1420–1433, 2011.

[58] A. Dhamdhere, C. Dovrolis, and P. Francois. A Value-based Framework for Internet Peering Agreements. In *Proceedings of ITC*, 2010.

[59] T. Dierks and E. Rescorla. The Transport Layer Security (TLS) Protocol Version 1.2, *RFC 5246*, 2008.

[60] Internet Traffic and Economics. DigiWorld Summit, 2010.

[61] X. Dimitropoulos, P. Hurley, A. Kind, and M. Stoecklin. On the 95-percentile Billing Method. *PAM*, 2009.

[62] R. Dingledine, N. Mathewson, and P. Syverson. Tor: the Second-generation Onion Router. Technical report, DTIC Document, 2004.

[63] M. Dischinger, M. Marcon, S. Guha, K. P. Gummadi, R. Mahajan, and S. Saroiu. Glasnost: Enabling End Users to Detect Traffic Differentiation. In *Proceedings of NSDI*, 2010.

[64] J. R. Douceur. The Sybil Attack. In *IPTPS*, 2002.

[65] F. Douglis, A. Feldmann, B. Krishnamurthy, and J. C. Mogul. Rate of Change and other Metrics: a Live Study of the World Wide Web. In *USENIX*, 1997.

[66] Z. Duan, Z.-L. Zhang, and Y. T. Hou. Service Overlay Networks: SLAs, QoS, and Bandwidth Provisioning. *ToN*, 11(6):870–883, 2003.

[67] Euro-IX. `https://www.euro-ix.net`.

[68] A. Fabrikant, U. Syed, and J. Rexford. There's Something about MRAI: Timing Diversity Can Exponentially Worsen BGP Convergence. In *Proceedings of INFOCOM*, 2011.

[69] R. Fanou, P. Francois, and E. Aben. On the Diversity of Interdomain Routing in Africa. In *Proceedings of PAM*, 2015.

[70] P. Faratin, D. Clark, P. Gilmore, S. Bauer, A. Berger, and W. Lehr. Complexity of Internet Interconnections: Technology, Incentives and Implications for Policy. In *Proceedings of TPRC*, 2007.

[71] N. Feamster, H. Balakrishnan, J. Rexford, A. Shaikh, and J. Van Der Merwe. The Case for Separating Routing from Routers. In *Proceedings of FDNA*, 2004.

[72] J. Feigenbaum, C. H. Papadimitriou, and S. Shenker. Sharing the Cost of Multicast Transmissions. *Journal of Computer and System Sciences*, 63(1):21–41, 2001.

[73] Filecoin. `http://filecoin.io/`.

[74] B. Fletcher. Internet Transit Sales: 2005-10. `http://www.renesys.com/blog/2010/10/internet-transit-sales-2005-10.shtml`, 2010.

[75] R. M. Frieden. Network Neutrality or Bias? Handicapping the Odds for a Tiered and Branded Internet. *Hastings Comm. & Ent. LJ*, 29:171, 2006.

[76] L. Gao. On Inferring Autonomous System Relationships in the Internet. *ToN*, 9(6):733–745, 2001.

[77] M. Ghosh, M. Richardson, B. Ford, and R. Jansen. A TorPath to TorCoin: Proof-of-Bandwidth Altcoins for Compensating Relays. In *Proceedings of HotPETs*, 2014.

[78] P. Gill, M. Arlitt, Z. Li, and A. Mahanti. The Flattening Internet Topology: Natural Evolution, Unsightly Barnacles or Contrived Collapse? In *Proceedings of PAM*, 2008.

[79] P. Gill, M. Schapira, and S. Goldberg. Let the Market Drive Deployment: A Strategy for Transitioning to BGP Security. In *Proceedings of SIGCOMM*, 2011.

[80] P. Godfrey, I. Ganichev, S. Shenker, and I. Stoica. Pathlet Routing. *CCR*, 39:111–122, 2009.

[81] S. Goldberg, M. Schapira, P. Hummon, and J. Rexford. How Secure are Secure Interdomain Routing Protocols? In *SIGCOMM*, 2010.

[82] S. Gorinsky, S. Jain, H. Vin, and Y. Zhang. Design of Multicast Protocols Robust Against Inflated Subscription. *ToN*, 14(2):249–262, 2006.

[83] S. Gorinsky, S. Jain, and H. M. Vin. Multicast Congestion Control with Distrusted Receivers. In *Networked Group Communication*, pages 19–26. Citeseer, 2002.

[84] A. Greenberg, G. Hjalmtysson, D. A. Maltz, A. Myers, J. Rexford, G. Xie, H. Yan, J. Zhan, and H. Zhang. A Clean Slate 4D Approach to Network Control and Management. 2005.

[85] T. G. Griffin and B. J. Premore. An Experimental Analysis of BGP Convergence Time. In *Proceedings of ICNP*, 2001.

[86] GRNET. `http://mon.grnet.gr/`.

[87] A. Gupta, M. Calder, N. Feamster, M. Chetty, E. Calandro, and E. Katz-Bassett. Peering at the Internet's Frontier: A First Look at ISP Interconnectivity in Africa. In *Proceedings of PAM*, 2014.

[88] L. Gyarmati, N. Laoutaris, K. Sdrolias, P. Rodriguez, and C. Courcoubetis. From Advertising Profits to Bandwidth Prices: A Quantitative Methodology for Negotiating Premium Peering. *PER*, 42(3):29–32, 2014.

[89] H. Haddadi, M. Rio, G. Iannaccone, A. Moore, and R. Mortier. Network Topologies: Inference, Modeling, and Generation. *Communications Surveys & Tutorials*, 2008.

[90] A. Haeberlen, I. Avramopoulos, J. Rexford, and P. Druschel. NetReview: Detecting when Interdomain Routing Goes Wrong. In *NSDI*, 2009.

[91] S. Hanks, T. Li, D. Farinacci, and P. Traina. Generic Routing Encapsulation (GRE), RFC 1701, 1994.

[92] S. Hares, Y. Rekhter, and T. Li. A Border Gateway Protocol 4 (BGP-4), RFC 4271, 2006.

[93] S. Hasan and S. Gorinsky. Obscure Giants: Detecting the Provider-free ASes. In *Proceedings of Networking*, 2012.

[94] S. Hasan, S. Gorinsky, C. Dovrolis, and R. Sitaraman. Trade-offs in Optimizing the Cache Deployments of CDNs. In *Proceedings of INFOCOM*, 2014.

[95] HEANET. `http://www.hea.net/`.

[96] M. Hrybyk. The Transit Exchange-A New Model for Open, Competitive Network Services. In *Proceedings of PTC*, 2007.

[97] P. Hui, R. Mortier, K. Xu, J. Crowcroft, and V. O. Li. Sharing Airtime with Shair Avoids Wasting Time and Money. In *Proceedings of HotMobile*, 2009.

[98] G. Huston. Analyzing the Internet's BGP Routing Table. *The Internet Protocol Journal*, 4(1):2–15, 2001.

[99] IX Reach. `http://www.ixreach.com`.

[100] W. Jiang, S. Ioannidis, L. Massoulié, and F. Picconi. Orchestrating Massively Distributed CDNs. In *Proceedings of CoNEXT*, 2012.

[101] P. Kanuparthy and C. Dovrolis. ShaperProbe: End-to-end Detection of ISP Traffic Shaping Using Active Methods. In *Proceedings of SIGCOMM*, 2011.

[102] E. Katz-Bassett, J. P. John, A. Krishnamurthy, D. Wetherall, T. Anderson, and Y. Chawathe. Towards IP Geolocation Using Delay and Topology Measurements. In *Proceedings of IMC*, 2006.

[103] E. Katz-Bassett, H. V. Madhyastha, J. P. John, A. Krishnamurthy, D. Wetherall, and T. E. Anderson. Studying Black Holes in the Internet with Hubble. In *Proceedings of NSDI*, 2008.

[104] R. Keralapura, N. Taft, C.-N. Chuah, and G. Iannaccone. Can ISPs Take the Heat from Overlay Networks. In *Proceedings of HotNets*, 2004.

[105] E. Kreifeldt. IT Business Edge Interview. `http://www.itbusinessedge.com/`, 2010.

[106] D. Kreutz, F. M. Ramos, P. Verissimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig. Software-defined Networking: a Comprehensive Survey. *IEEE*, 103(1):14–76, 2015.

[107] J. F. Kurose. *Computer Networking: a Top-down Approach Featuring the Internet*. Pearson Education, 2005.

[108] C. Labovitz, A. Ahuja, A. Bose, and F. Jahanian. Delayed Internet Routing Convergence. *ToN*, 9(3):293–306, 2001.

[109] C. Labovitz, A. Ahuja, R. Wattenhofer, and S. Venkatachary. The Impact of Internet Policy and Topology on Delayed Routing Convergence. In *Proceedings of INFOCOM*, 2001.

[110] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian. Internet Inter-domain Traffic. In *Proceedings of SIGCOMM*, 2010.

[111] K. Lakshminarayanan, I. Stoica, S. Shenker, and J. Rexford. *Routing as a Service*. Computer Science Division, University of California, 2004.

[112] K. Lakshminarayanan, I. Stoica, S. Shenker, and J. Rexford. Routing as a Service. Technical Report UCB/EECS-2006-19, *UC Berkeley*, 2006.

[113] N. Laoutaris, G. Smaragdakis, K. Oikonomou, I. Stavrakakis, and A. Bestavros. Distributed Placement of Service Facilities in Large-scale Networks. In *Proceedings of INFOCOM*, 2007.

[114] N. Laoutaris, G. Smaragdakis, P. Rodriguez, and R. Sundaram. Delay Tolerant Bulk Data Transfers on the Internet. In *Proceedings of SIGMETRICS*, 2009.

[115] D. Lee, K. Cho, G. Iannaccone, and S. Moon. Has Internet Delay Gotten Better or Worse? In *Proceedings of CFI*, 2010.

[116] D. Lee, J. Mo, and J. Park. ISP vs. ISP+ CDN: Can ISPs in Duopoly Profit by Introducing CDN Services? *PER*, 40(2):46–48, 2012.

[117] C. Li, K. Sycara, and A. Scheller-Wolf. Combinatorial Coalition Formation for Multi-item Group-buying with Heterogeneous Customers. *Decision Support Systems*, 49(1):1–13, 2010.

[118] D. Liben-Nowell, A. Sharp, T. Wexler, and K. Woods. Computing Shapley Value in Cooperative Supermodular Games. *Computing and Combinatorics*, 7434:568–579, 2012.

[119] LINX NoVA. `https://www.linx.net/service/publicpeering/nova`.

[120] A. Lodhi, A. Dhamdhere, and C. Dovrolis. Analysis of Peering Strategy Adoption by Transit Providers in the Internet. In *Proceedings of NetEcon*, 2012.

[121] A. Lodhi, A. Dhamdhere, and C. Dovrolis. GENESIS: An Agent-based Model of Interdomain Network Formation, Traffic Flow and Economics. In *Proceedings of INFOCOM*, 2012.

[122] A. Lodhi, N. Larson, A. Dhamdhere, C. Dovrolis, and K. Claffy. Using PeeringDB to Understand the Peering Ecosystem. *CCR*, 44(2):20–27, 2014.

[123] London Internet Exchange (LINX). `https://www.linx.net`.

[124] A. Lutu. *A System for the Detection of Limited Visibility in BGP*. PhD thesis, Carlos III University of Madrid, Spain, 2014.

[125] R. Ma, D. Chiu, J. Lui, V. Misra, and D. Rubenstein. Interconnecting Eyeballs to Content: A Shapley Value Perspective on ISP Peering and Settlement. In *Proceedings of NetEcon*, 2008.

[126] R. Ma, D. Chiu, J. Lui, V. Misra, and D. Rubenstein. Internet Economics: The Use of Shapley Value for ISP Settlement. *ToN*, 18(3):775–787, 2010.

[127] R. Ma, D. Chiu, J. Lui, V. Misra, and D. Rubenstein. On Cooperative Settlement Between Content, Transit, and Eyeball Internet Service Providers. *ToN*, 19(3):802–815, 2011.

[128] R. T. Ma. Subsidization Competition: Vitalizing the Neutral Internet. In *Proceedings of CoNEXT*, 2014.

[129] R. T. Ma and V. Misra. The Public Option: a Nonregulatory Alternative to Network Neutrality. *ToN*, 21(6):1866–1879, 2013.

[130] H. V. Madhyastha, T. Anderson, A. Krishnamurthy, N. Spring, and A. Venkataramani. A Structural Approach to Latency Prediction. In *Proceedings of IMC*, 2006.

[131] H. V. Madhyastha, T. Isdal, M. Piatek, C. Dixon, T. Anderson, A. Krishnamurthy, and A. Venkataramani. iPlane: An Information Plane for Distributed Services. In *Proceedings of OSDI*, 2006.

[132] R. Mahajan, D. Wetherall, and T. Anderson. Negotiation-Based Routing Between Neighboring ISPs. In *NSDI*, 2005.

[133] Z. M. Mao, R. Govindan, G. Varghese, and R. H. Katz. Route Flap Damping Exacerbates Internet Routing Convergence. *CCR*, 32(4):221–233, 2002.

[134] M. Marcon, M. Dischinger, K. P. Gummadi, and A. Vahdat. The Local and Global Effects of Traffic Shaping in the Internet. In *Teletraffic Congress (ITC)*, pages 1–8. IEEE, 2010.

[135] Matthew Prince (Cloudflare CEO). The Relative Cost of Bandwidth Around the World. `https://blog.cloudflare.com/the-relative-cost-of-bandwidth-around-the-world/`.

[136] X. Meng, Z. Xu, B. Zhang, G. Huston, S. Lu, and L. Zhang. IPv4 Address Allocation and the BGP Routing Table Evolution. *CCR*, 35(1):71–80, 2005.

[137] A. Miller and J. J. LaViola Jr. Anonymous Byzantine Consensus from Moderately-Hard Puzzles: A Model for Bitcoin. Technical Report CS-TR-14-01, *University of Central Florida*, 2014.

[138] V. Misra, S. Ioannidis, A. Chaintreau, and L. Massoulié. Incentivizing Peer-Assisted Services: a Fluid Shapley Value Approach. In *Proceedings of SIGMETRICS*, 2010.

[139] H. Moulin and S. Shenker. Strategyproof Sharing of Submodular Costs: Budget Balance versus Efficiency. *Economic Theory*, 18(3):511–533, 2001.

[140] J. Musacchio, G. Schwartz, and J. Walrand. A Two-Sided Market Analysis of Provider Investment Incentives with an Application to the Net-Neutrality Issue. *Review of Network Economics*, 8(1), 2009.

[141] S. Nakamoto. Bitcoin: A Peer-to-Peer Electronic Cash System. `http://bitcoin.org/bitcoin.pdf`, 2008.

[142] J. Naous, M. Walfish, A. Nicolosi, D. Mazieres, M. Miller, and A. Seehra. Verifying and Enforcing Network Paths with ICING. In *Proceedings of CoNEXT*, 2011.

[143] W. B. Norton. Internet Service Providers and Peering. In *NANOG 45*, 2001.

[144] W. B. Norton. The Ideal Peering Forum. `http://goo.gl/yqtt7O`, 2010.

[145] W. B. Norton. *The Internet Peering Playbook: Connecting to the Core of the Internet*. DrPeering Press, 2012.

[146] E. Nygren, R. K. Sitaraman, and J. Sun. The Akamai Network: a Platform for High-performance Internet Applications. *OSR*, 44(3):2–19, 2010.

[147] R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang. The (in)Completeness of the Observed Internet AS-level Structure. *ToN*, 18:109–122, 2010.

[148] R. Oliveira, B. Zhang, D. Pei, and L. Zhang. Quantifying Path Exploration in the Internet. *ToN*, 17(2):445–458, 2009.

[149] J. S. Otto, M. A. Sánchez, D. R. Choffnes, F. E. Bustamante, and G. Siganos. On Blind Mice and the Elephant. In *Proceedings of SIGCOMM*, 2011.

[150] Packet Clearing House. `https://www.pch.net`.

[151] Peering Database. `https://www.peeringdb.com`.

[152] M. Podlesny and S. Gorinsky. Leveraging the Rate-delay Trade-off for Service Differentiation in Multi-provider Networks. *JSAC*, 29(5):997–1008, 2011.

[153] I. Poese, S. Uhlig, M. A. Kaafar, B. Donnet, and B. Gueye. IP Ceolocation Databases: Unreliable? *CCR*, 41(2):53–56, 2011.

[154] A. Popescu and T. Underwood. D(3)peered: Just the Facts Ma'am. In *NANOG 35*, 2005.

[155] J. Postel. Internet Protocol DARPA Internet Program Protocol Specification. RFC 791, 1981.

[156] L. Qiu, V. Padmanabhan, and G. Voelker. On the Placement of Web Server Replicas. In *Proceedings of INFOCOM*, 2001.

[157] B. Raghavan and A. C. Snoeren. A System for Authenticated Policy-compliant Routing. In *CCR*, volume 34, pages 167–178, 2004.

[158] B. Raghavan, P. Verkaik, and A. C. Snoeren. Secure and Policy-compliant Source Routing. *ToN*, 17(3):764–777, 2009.

[159] A. H. Rasti, N. Magharei, R. Rejaie, and W. Willinger. Eyeball ASes: from Geography to Connectivity. In *Proceedings of IMC*, 2010.

[160] Rediris. `http://www.rediris.es`.

[161] Renesys. Bahrain's Internet Ecosystem. `http://www.tra.org.bh/en/pdf/Renesys_Study.pdf`, 2009.

[162] D. M. (Renesys). "Crecimiento" in Latin America. `http://www.renesys.com/2013/05/crecimiento-in-latin-america/`, 2013.

[163] E. Rescorla. HTTP Over TLS, *RFC 2818*, 2000.

[164] Réseaux IP Européns Network Coordination Centre (RIPE NCC). `http://www.ris.ripe.net/cgi-bin/lg/index.cgi`.

[165] P. Richter, G. Smaragdakis, A. Feldmann, N. Chatzis, J. Boettger, and W. Willinger. Peering at Peerings: On the Role of IXP Route Servers. In *Proceedings of IMC*, 2014.

[166] RIPE. YouTube Hijacking: A RIPE NCC RIS Case Study. `http://goo.gl/jKX6Bz`, 2008.

[167] E. C. Rosen. Exterior Gateway Protocol (EGP). RFC 827, 1982.

[168] SANET. `http://samon.cvt.stuba.sk/`.

[169] W. Scott, R. Cheng, J. Li, A. Krishnamurthy, and T. Anderson. Blocking-resistant Network Services Using Unblock, 2012.

[170] A. Seehra, J. Naous, M. Walfish, D. Mazieres, A. Nicolosi, and S. Shenker. A Policy Framework for the Future Internet. In *Proceedings of HotNets*, 2009.

[171] Segment Routing. `http://www.segment-routing.net/home/ietf`.

[172] S. Shakkottai and R. Srikant. Economics of Network Pricing with Multiple ISPs. *ToN*, 14(6):1233–1245, 2006.

[173] L. S. Shapley. A Value for n-Person Games. *Annals of Mathematics Study*, (28), 1953.

[174] A. Shaw. Spam? Not Spam? Tracking a Hijacked Spamhaus IP. `http://goo.gl/GE0c05`, 2013.

[175] F. B. Shepherd and G. T. Wilfong. Multilateral Transport Games. In *Proceedings of INOC*, 2005.

[176] G. Shrimali, A. Akella, and A. Mutapcic. Cooperative Interdomain Traffic Engineering Using Nash Bargaining and Decomposition. *ToN*, 18(2):341–352, 2010.

[177] M. Skjegstad, A. Madhavapeddy, and J. Crowcroft. Kadupul: Livin' on the Edge with Virtual Currencies and Time-Locked Puzzles. 2014.

[178] J. H. Sowell. Empirical Studies of Bottom-Up Internet Governance. In *Proceedings of TPRC*, 2012.

[179] N. Spring, R. Mahajan, D. Wetherall, and T. Anderson. Measuring ISP Topologies with Rocketfuel. *ToN*, 12(1):2–16, 2004.

[180] R. Stanojevic, N. Laoutaris, and P. Rodriguez. On Economic Heavy Hitters: Shapley Value Analysis of 95th-percentile Pricing. In *Proceedings of IMC*, 2010.

[181] I. Stoica, R. Morris, D. Karger, M. Kaashoek, and H. Balakrishnan. Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications. *CCR*, 31(4):149–160, 2001.

[182] A.-J. Su, D. R. Choffnes, A. Kuzmanovic, and F. E. Bustamante. Drafting Behind Akamai: Inferring Network Conditions Based on CDN Redirections. *ToN*, 17(6):1752–1765, 2009.

[183] L. Subramanian, S. Agarwal, J. Rexford, and R. Katz. Characterizing the Internet Hierarchy from Multiple Vantage Points. In *Proceedings of INFOCOM*, 2002.

[184] J. Tang and R. T. Ma. Regulating Monopolistic ISPS without Neutrality. In *Proceedings of ICNP*, pages 374–384. IEEE, 2014.

[185] R. Teixeira and J. Rexford. Managing Routing Disruptions in Internet Service Provider Networks. *IEEE Communications Magazine*, 44(3):160–165, 2006.

[186] Telegeography. Global Internet Geography. `http://www.telegeography.com/`, 2015.

[187] The Little Garden (TLG). `http://www.wps.com/J/TLG/TLG.html`.

[188] The Multi Router Traffic Grapher. `http://oss.oetiker.ch/mrtg`.

[189] J. Tirole. *The Theory of Industrial Organization*. MIT Press, 2000.

[190] S. Uhlig and S. Tandel. Quantifying the BGP Routes Diversity Inside a Tier-1 Network. *NETWORKING*, pages 1002–1013, 2006.

[191] V. Valancius, C. Lumezanu, N. Feamster, R. Johari, and V. Vazirani. How Many Tiers? Pricing in the Internet Transit Market. In *Proceedings of SIGCOMM*, 2011.

[192] S. Vissicchio, L. Vanbever, and O. Bonaventure. Opportunities and Research Challenges of Hybrid Software Defined Networks. *CCR*, 44(2):70–75, 2014.

[193] S. Vissicchio, L. Vanbever, C. Pelsser, L. Cittadini, P. Francois, and O. Bonaventure. Improving Network Agility with Seamless BGP Reconfigurations. *ToN*, 21(3):990–1002, 2013.

[194] Voxel dot net, pricelist. Accessed in January 2011.

[195] H. Wang, C. Jin, and K. G. Shin. Defense Against Spoofed IP Traffic Using Hop-count Filtering. *ToN*, 15(1):40–53, 2007.

[196] M. Warschauer. *Technology and Social Inclusion: Rethinking the Digital Divide*. MIT Press, 2004.

[197] D. Wendlandt, I. Avramopoulos, D. G. Andersen, and J. Rexford. Don't Secure Routing Protocols, Secure Data Delivery. In *Proceedings of HotNets*, 2006.

[198] W. Willinger and M. Roughan. Internet Topology Research Redux. *SIGCOMM eBook: Recent Advances in Networking*, 2013.

[199] E. Winter. *The Shapley Value*. North-Holland, 2002.

[200] T. Wu. Network Neutrality, Broadband Discrimination. *Journal on Telecommunications and High Technology*, 2:141–141, 2003.

[201] H. Xie, Y. R. Yang, A. Krishnamurthy, Y. G. Liu, and A. Silberschatz. P4P: Provider Portal for Applications. In *Proceedings of SIGCOMM*, 2008.

[202] M. Yannuzzi, X. Masip-Bruin, and O. Bonaventure. Open Issues in Interdomain Routing: a Survey. *IEEE Network*, 19(6):49–56, 2005.

[203] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman. Sybilguard: Defending Against Sybil Attacks via Social Networks. In *Proceedings of SIGCOMM*, 2006.

[204] M. Yu, W. Jiang, H. Li, and I. Stoica. Tradeoffs in CDN Designs for Throughput Oriented Traffic. In *Proceedings of CoNEXT*, 2012.

[205] M. Yuksel, K. Ramakrishnan, S. Kalyanaraman, J. Houle, and R. Sadhvani. Class-of-service in IP Backbones: Informing the Network Neutrality Debate. In *Proceedings of SIGMETRICS*, 2008.