UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

PEDRO DE BOTELHO MARCOS

Towards a dynamic Internet interconnection ecosystem for improved wide-area traffic delivery

Thesis presented in partial fulfillment
of the requirements for the degree of
Doctor of Computer Science

Advisor: Prof. Dr. Marinho Pilla Barcellos

Porto Alegre
September 2019

CIP — CATALOGING-IN-PUBLICATION

Marcos, Pedro de Botelho

Towards a dynamic Internet interconnection ecosystem for improved wide-area traffic delivery / Pedro de Botelho Marcos. – Porto Alegre: PPGC da UFRGS, 2019.

98 f.: il.

Thesis (Ph.D.) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR–RS, 2019. Advisor: Marinho Pilla Barcellos.

1. Internet. 2. Wide-area traffic delivery. 3. Interconnection. 4. Peering. 5. Peering Infrastructures. 6. Internet eXchange Point. 7. Colocation Facility. I. Barcellos, Marinho Pilla. II. Título.

"Forever trusting who we are
and nothing else matters."
— METALLICA

ACKNOWLEDGEMENTS

This thesis is the end of a journey that I would not be able to reach without the support of amazing people. I am incredibly thankful to my advisor, Marinho Barcellos, for his guidance in the past four years. Marinho has introduced me to the world of computer networks research and taught me to pursue impactful research that addresses significant challenges. Marinho was always open to discuss ideas, provide his feedback, and to support me even when he was not entirely convinced about my claims, which were essential to my growth as a researcher. He also was invariably ready to incentivize me to attend conferences (and in helping finding support for it) and to prospect collaborations that would foster our research. While now it is time to starting flying by myself, I expect that we can continue collaborating for the upcoming years.

I am also very grateful to Marco Canini, which co-advised this thesis. Canini started collaborating with us at the early stages of my Ph.D. and was fundamental in helping set the right direction of our research. In 2017 Marco hosted me at KAUST in what I would describe as one of the most relevant experiences in my life as a researcher and as a human being. While some paperwork prevented us from officially giving Marco deserved recognition as co-advisor of this research, it is impossible to imagine that this work would have achieved all these results without Canini's contributions.

I am also super thankful to Marco Chiesa, who I have met in CoNEXT 2016 and was the starting point of the amazing collaboration that resulted in this thesis. His expertise and knowledge about the Internet ecosystem were crucial to guide this research. Marco started as a collaborator and ended this journey as my third advisor, alongside with Marinho and Canini. I was also very fortunate to collaborate and co-author papers with a group of talented researchers. So, thank you Lucas Müller (UFRGS), Christoph Dietzel (DE-CIX), Pradeeban Kathiravelu (ULisboa), Yousef Alowayed (KAUST), and Alexandre Wermann (UFRGS) for your collaboration.

I am very thankful to Ignacio Castro (QMUL) for the fruitful discussions about the interconnection ecosystem we had in the past years in all the conferences (IMC 2017, SIGCOMM 2018, CoNEXT 2018) we have met. His expertise and insightful comments were fundamental to shape this thesis. I am also thankful to Leandro Bertholdo (UTwente), Raul Sejas (Telefónica), Philippe Duguet (Hopus), Eric Loos (BICS), Ankit Singla (ETH Zurich), Somesh Chaturmohta (Microsoft), Nikolaos Laoutaris (IMDEA), and Josh Bailey (Google) for their excellent feedback and discussions that helped to improve this work. I am also thankful to all network operators and peering coordinators for taking part in our survey. I also would like to thank Alberto Dainotti (CAIDA/UCSD), Italo Cunha (UFMG), Lisandro Granville (UFRGS), and Weverton Cordeiro (UFRGS) for their valuable feedback as members of the committees of my thesis and thesis proposal.

I could not be more grateful to my fiance (and love of my life) Debora, for her support during this journey. Thanks for always believing in me, for encouraging me to do things that I would probably not do without your incentive, and for all your patience and understanding during the Ph.D. I am what I am today because I have an amazing woman by my side. I love you! I am also very thankful to my parents Celso and Mariangela, and my brother Gabriel, for always being there for me and guaranteeing that I have had everything that I needed to reach the end of this endeavor.

I was also very fortunate to have shared this journey with such a wonderful group of colleagues. Many thanks to Rodrigo Oliveira, Miguel Neves, Fabricio Mazzola, Rodrigo Dal Ri, Lucas Leal, Lucas Müller, and Leandro Mondin for your friendship and the countless hours of exciting discussions about the most diverse themes (including scientific ones :)). I am also thankful to Amedeo Sapio, Arnaud Dethise, and Sebastian Voll for being such good friends (alongside with Pradeeban) during the time I was visiting KAUST. Being away from home is not an easy thing sometimes, but your friendship made it a delightful time.

Finally, I would like to thank FURG, UFRGS, and KAUST for all the support during my Ph.D. I also would like to acknowledge ACM for the travel support to attend conferences. This research was in part supported by European Union's Horizon 2020 research and innovation program under the ENDEAVOUR project (grant agreement 644960), by the project Mapping Interconnection in the Internet: Colocation, Connectivity, and Congestion (NSF CNS-1414177 grant), by CNPq Grant 310408/2017-2 and by CAPES/Brazil - Finance Code 001.

ABSTRACT

Traffic delivery is a fundamental aspect of the Internet today as most of the traffic relates to applications such as streaming and gaming, which have strict service requirements. As a consequence, in recent years, peering infrastructures such as Internet eXchange Points and colocation facilities have emerged as crucial elements of the Internet topology. Today, these facilities provide physical interconnection for hundreds to thousands of Autonomous Systems, creating rich connectivity environments. Despite the potential benefits, any pair of Autonomous Systems needs first to agree on exchanging traffic. By interviewing and surveying more than 100 network operators, we discovered that most interconnection agreements are established through ad-hoc and lengthy processes heavily influenced by personal relationships and brand image. As such, Autonomous Systems often prefer long-term contracts at the expense of a potential mismatch between actual delivery performance and current traffic dynamics. Autonomous Systems may also miss interconnection opportunities because their operators do not have a personal relationship or do not have information to build their opinion about the other AS. We argue that peering infrastructures have a considerable unexplored potential to allow network operators to be more responsive to the Internet traffic dynamics and improve wide-area traffic delivery performance, since, in addition to the rich connectivity, their peering ports have substantial amounts of spare capacity during most time. To unleash this potential, we propose Dynam-IX, a Dynamic Interconnection eXchange. Dynam-IX allows network operators to build trust cooperatively and implement traffic engineering policies to exploit the plentiful interconnection opportunities at peering infrastructures quickly while preserving the privacy of peering policies. Dynam-IX offers a protocol to automate the interconnection process, an intent abstraction to express interconnection policies, a legal framework to digitally handle contracts, and a distributed tamper-proof ledger to create trust among Autonomous Systems. To evaluate our approach, we build a prototype and show that an AS can establish tens of agreements per minute with negligible storage and network overheads for Autonomous Systems and the peering infrastructure. We also show that Dynam-IX scales to the size of the largest peering infrastructures, demonstrating its practicality.

Keywords: Internet. Wide-area traffic delivery. Interconnection. Peering. Peering Infrastructures. Internet eXchange Point. Colocation Facility.

Rumo a um Ecossistema de Interconexão Mais Dinâmico para Entrega de Tráfego Aprimorada na Internet

RESUMO

Entrega de tráfego é um aspecto fundamental da Internet atual uma vez que a maior parte do tráfego está relacionada à aplicações como *streaming* e jogos, as quais possuem requisitos estritos de serviço. Em consequência, nos últimos anos, infraestruturas de *peering* como pontos de troca de tráfego e pontos de interconexão emergiram como elementos cruciais na topologia da Internet. Hoje em dia, estes locais fornecem conecitividade física para centenas ou até mesmo milhares de sistemas autônomos, criando ambientes ricos em conecitividade. Apesar do benefícios, qualquer par de sistemas autônomos precisa primeiro concordar em trocar tráfego. Através de entrevistas e uma pesquisa com mais de 100 operadores de rede nós descobrimos que a maior parte dos acordos de interconexão são estabelecidos através de um processo manual e demorado que é largamente influenciado por relações pessoais e reconhecimento de marcas. Desta forma, sistemas autônomos seguidamente preferem estabelecer contratos de longa duração ao custo de uma potencial discrepância entre a efetiva entrega de tráfego e a dinâmica atual de tráfego. Sistemas autônomos podem também perder oportunidades de interconexão porque seus operadores não possuem relações pessoais ou não possuem informações para construir sua opinião a respeito do outro sistema autônomo. Nós argumentamos que infraestruturas de *peering* possuem um considerável potencial ainda não explorado para permitir que operadores de rede sejam mais responsivos à dinâmica de tráfego da Internet e possam com isso aprimorar a entrega de tráfego na Internet, uma vez que, além da rica conecitividade, as infraestruturas de *peering* também possuem uma quantidade considerável de capacidade ociosa durante a maior parte do tempo. Para habilitar este potencial nós propomos o Dynam-IX, um ponto de interconexão dinâmico. O Dynam-IX permite que operadores de rede construam confiança de forma cooperativa e implementem políticas de engenharia de tráfego que exploram as oportunidades de interconexão das infraestruturas de *peering* de forma rápida enquanto mantém a privacidade das políticas de interconexão. O Dynam-IX oferece um protocolo para automatizar o processo de interconexão, uma abstração de intenções para expressar políticas de interconexão, um arcabouço de contratos para lidar com aspectos legais digitalmente e uma base de dados distribuída e não-adulterável para criar confiança entre os sistemas autônomos. Para avaliar nossa proposta nós construímos um protótipo e mostramos que um sistema autônomo pode estabelecer dezenas de acordos de interconexão em um minuto com sobrecarga negligível de armazenamento e de rede para os sistemas autônomos e para a infraestrutura de *peering*. Nós também mostramos que o Dynam-IX escala para o tamanho das maiores infraestruturas de *peering*, demonstrando assim a sua viabilidade.

Palavras-chave: Internet, Entrega de Tráfego na Internet, Interconexão, Peering, Infraestrutura de Peering, Ponto de Troca de Tráfego, Ponto de Interconexão.

LIST OF ABBREVIATIONS AND ACRONYMS

ACM	Association for Computing Machinery
AFNOG	African Network Operators Group
AFRINIC	African Network Information Center
AMS-IX	Amsterdam Internet eXchange
ANRW	Applied Network Research Workshop
API	Applicaion Programming Interface
APNIC	Asia Pacific Network Information Co
ARIN	American Registry for Internet Numbers
AS	Autonomous System
ASN	Autonomous System Number
AWS	Amazon Web Services
BGP	Border Gateway Protocol
CAIDA	Center for Applied Internet Data Analysis
CCR	Computer Communication Review
CDN	Content Delivery Network
CIR	Committed Information Rate
CoNEXT	Conference on emerging Networking EXperiments and Technologies
DDoS	Distributed Denial of Service
DENOG	German Network Operators Group
DE-CIX	Deutscher Commercial Internet eXchange
DNS	Domain Name System
DoS	Denial of Service
Dynam-IX	Dynamic Interconnection eXchange
EA	Eletronic Arts

EC2	Elastic Compute
ET	East Time
Euro-IX	European Internet Exchange Association
FIFA	Fédération Internationale de Football Association
France-IX	France Internet eXchange
GB	Gigabyte
Gbps	Gigabits per second
GCC	Google Cloud Computing
HLF	HyperLedger Fabric
iOS	iPhone Operating System
IP	Internet Protocol
IRR	Internet Routing Registry
ISOC	Internet Society
IFIP	International Federation for Information Processing
ISP	Internet Service Provider
IXP	Internet eXchange Point
IX.br	Internet eXchange Brazil
IX-Forum	Brazilian Internet eXchange Forum
IXP-EU	European Internet eXchange Point
IXP-LA	Latin American Internet eXchange Point
KAUST	King Abdullah University of Science and Technology
LACNIC	Latin America and Caribbean Network Information Centre
LACNOG	Latin America and Caribbean Network Operators Group
LINX	London Internet eXchange
LLF	Lightweight Legal Framework
L2	Layer 2

L3	Layer 3
MB	Megabyte
Mbps	Megabits per second
MINT	Market for Internet Transit
MPLS	Multi Protocol Label Switching
MS	Microsoft
NANOG	North America Network Operators Group
NDA	Non-disclosure Agreement
NL-IX	Neutral Internet eXchange
NOC	Network Operations Center
NOG	Network Operators Group
PCH	Packet Clearing House
PeA	Peering Analytics
PeeringDB	Peering Database
Ph.D.	Philosophiae Doctor
PINT	Peering Intent
PM	Post Meridiem
PoP	Point of Presence
QoE	Quality of Experience
QoS	Quality of Service
RAM	Random Access Memory
RIPE	Réseaux IP Européens Network Coordination Centre
RIR	Regional Internet Registry
ROV	Route Origin Validation
RS	Route Server
SDX	Software Defined eXchange

SGX	Software Guard eXtension
SIGCOMM	Special Interest Group on Data Communication
SLA	Service Level Agreement
SP-IX	São Paulo Internet eXchange
SSL	Secure Sockets Layer
Tbps	Terabits per second
TPB	Transactions per Block
vCPU	Virtual Central Processing Unit
VLAN	Virtual Local Area Network

LIST OF SYMBOLS

α	Fraction of Time
c	Estimated port capacity
e	Event identifier
l	Event length
m	Event magnitude
p	Highest value in the sequence R
R	Sequence of 5-minute long measurements
s	Spare capacity ratio
t	Capacity threshold
u	Port capacity utilization ratio

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 INTRODUCTION

A challenging scenario. It is 1 PM on a Wednesday. Network operators of a large Internet Service Provider (ISP) observe a sudden increase in traffic volumes, reaching three times the average daily levels of gaming traffic, which results in worsened network performance. After some analysis, they identify the cause: it is the release of the public beta of a famous video game. Although this may all sound too hypothetical, it is not. The demo release of EA Sports' FIFA 16, with its average download size of over 4 GB, caused the total traffic on this network to surge to levels that were almost 8% higher than the usual (SANDVINE, 2015). Traffic surges are not a rare phenomenon either. Several past major software releases of Apple's iOS caused traffic spikes of almost one *terabit* per second for several hours and anecdotal evidence in blogs, news, and Internet forums suggest that these spikes led to slowdowns and errors for users (ARTHUR, 2011; BRODKIN, 2013). They also caused challenges for network operators. One operator in North America noted that "upon release at 1 PM ET, Apple Updates immediately became almost 20% of total network traffic, and continued to stay above 15% of total traffic into the evening peak hours" (SANDVINE, 2013).

Traffic surges are just one example of the challenges faced by network operators on the Internet today. Another one relates to the large volumes of traffic and the strict service requirements from modern Internet applications (GHASEMI et al., 2016). In one side, end-users want to watch videos without interruptions and play their online games without lag, while on the other side content and service providers are seeking for alternatives to accommodate the traffic inside their infrastructures. Such a condition has led large content providers such as Google and Facebook, two of the largest content producers, to deploy mechanisms (SCHLINKER et al., 2017; YAP et al., 2017) to measure performance metrics and dynamically adapt the routes overriding BGP decisions. Additionally, outages (GIOTSAS et al., 2017) are also impacting traffic delivery, requiring operators to reroute part of their traffic using other links.

The rise of peering infrastructures. To deal with these scenarios, networks operators have been increasing the connectivity of their ASes. As a consequence, peering infrastructures, such as Internet eXchange Points (IXPs) and colocation facilities, became the high-speed physical crossroads of Internet traffic and key elements of the Internet topology. Today there are over 800 IXPs (TELEGEOGRAPHY, 2018) and more than 3000 colocation facilities (PeeringDB, 2018) spread worldwide, with the largest ones carrying multiple terabits of traffic per second and enabling Autonomous Systems (ASes) to reach hundreds of other networks directly. By increasing connectivity, peering infrastructures contribute to improving the quality of Internet traffic delivery with lower latency and higher throughput (AHMED et al., 2017), and to keep traffic local (CHATZIS et al., 2013), avoiding trombone paths and surveillance issues (EDMUNDSON et al., 2018). Such state of affairs has led to substantial changes in the Internet topology over the past decade, resulting in a topology that is now richly connected and flattened (DHAMDHERE; DOVROLIS, 2010; CHIU et al., 2015).

Interconnecting is a cumbersome process. Peering infrastructures provide high-speed physical connectivity (i.e., L2) among any pair of members. Nevertheless, before exchanging any traffic, the operators of two ASes first need to agree on the terms and configure L3 information. Even as of today, the interconnection process continues to be an unstructured effort mostly dependent on personal relationships and brand image. As a consequence, it is not uncommon in mailing lists of Network Operators Groups (NOG) messages asking for information about the person responsible for peering/interconnection in some AS, previous interconnection experiences, or reporting issues such as route leaks, prefix hijackings, and route flappings caused by a given AS (NANOG, 2019d; NANOG, 2019a; NANOG, 2019b; NANOG, 2019c). To better understand the limitations of the interconnection ecosystem, we conducted interviews with operators from peering infrastructures and a tier-1 ISP, and a survey of 100+ network operators and peering coordinators. We found that before an AS attempts any routing change, much of the process relies on human interaction including in-person meetings, trust and reputation, billing and payment arrangements, and possibly lengthy legal negotiations, resulting in a process that can take weeks or even months to complete.

Technical and human factors affect today's ability to leverage the rich connectivity diversity of peering infrastructures. In the case of settlement-free peering, ASes may adopt multi-lateral agreements using a single BGP session with a Route Server (RS). Even though this may simplify the negotiation of interconnection, it comes with undesired technical limitations. Multi-lateral peering reduces control over routing decisions because Route Servers propagate only the best route. Such a condition is undesirable as it negates opportunities to optimize traffic engineering in response to downstream congestion (SCHLINKER et al., 2017; YAP et al., 2017) and to quickly reroute traffic under failures (BAKKER et al., 2016). Furthermore, ASes may not be willing to disclose their peering policies to the other ASes or the peering infrastructure (CHIESA et al., 2017b).

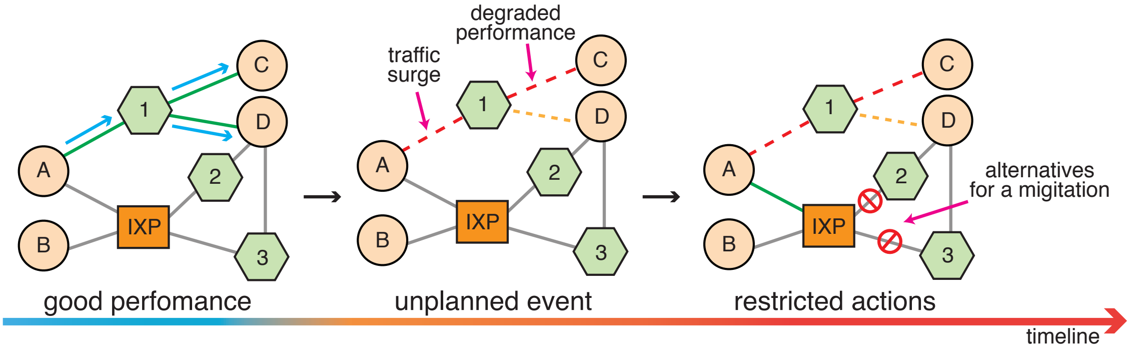
In contrast, bi-lateral agreements enable ASes to retain control over routing decisions and preserve the privacy of their policies. Recent studies suggest that the majority of the traffic at peering infrastructures traverses bi-lateral agreements (RICHTER et al., 2014; CHIESA et al., 2016; CHIESA et al., 2017b). However, because establishing agreements is cumbersome, the current practices are to form medium- or long-term contracts (e.g., a year or longer). As such, they ignore opportunities to dynamically adapt routing to reflect new (short-term) trends or to account for unplanned events, such as traffic surges (ARTHUR, 2011; BRODKIN, 2013; SANDVINE, 2015; MCGEE-ABE, 2017) and link failures (DUNCAN, 2017). Our survey results further corroborate the operators' desire for fast interconnection procedures; a majority of the respondents (56%) states that being able to interconnect in short time frames is a relevant improvement for the interconnection ecosystem.

In both cases (multi-lateral and bilateral), in addition to the technical limitations (e.g., privacy, route control), human factors contribute to limit the ability to improve wide-area traffic delivery. During our interviews, network operators highlighted that personal relationships and brand image play an important role when deciding whether or not to interconnect. The lack of methods to identify reliable peering partners might result in ASes not interconnecting because they do not trust each other (MEIER-HAHN, 2017), even if doing so would benefit both networks.

Motivating example. To illustrate this point, we provide an example in Figure 1.1. A is a stub AS connected to ISP_1 and an IXP. Then, a traffic surge towards C starts, congesting the link between A and ISP_1. Such congestion will affect the performance of all the traffic originated on A and going through the connection with ISP_1, represented in the example by the traffic going to D. A mitigation alternative for A would be to send the traffic to D via IXP link. However, this is not possible for A since it has agreements neither with ISP_2 nor with ISP_3. Additionally, A is not able to identify if ISP_2 and ISP_3 are trustworthy to interconnect. Thus, even having physical capacity available and L2 connectivity with potential providers, A is unable to mitigate the impacts of the traffic surge due to limitations of the current interconnection process.

An underutilized interconnection ecosystem. We posit that peering infrastructures have a largely unexplored potential to improve wide-area traffic delivery performance, as they (i) provide physical connectivity among hundreds or even thousands of ASes (rich path diversity) and (ii) their peering ports have a substantial spare capacity that ASes could utilize for new interconnection agreements. Concerning the former, previous work has

Figure 1.1: Limited dynamism (circle - stub, polygon - ISP).



Source: the authors.

shown that 40% of the IPv4 addresses is directly accessible by connecting to five IXPs only and that 91% is reachable considering the 1-hop customers of the facilities members (KOTRONIS et al., 2016). Alas, this path diversity could be used to recover up to 60% of connectivity during failures caused by emergencies (e.g., earthquakes) (Hu et al., 2012). Regarding the latter, interconnection links of most ASes have spare capacity due to conservative network planning (FEAMSTER, 2016; CHIESA et al., 2016). To confirm these conditions, we analyzed traffic data from a medium-sized and a large IXP. We found that more than 50% of IXP ports have at least 80% unused bandwidth for 50% of the time.

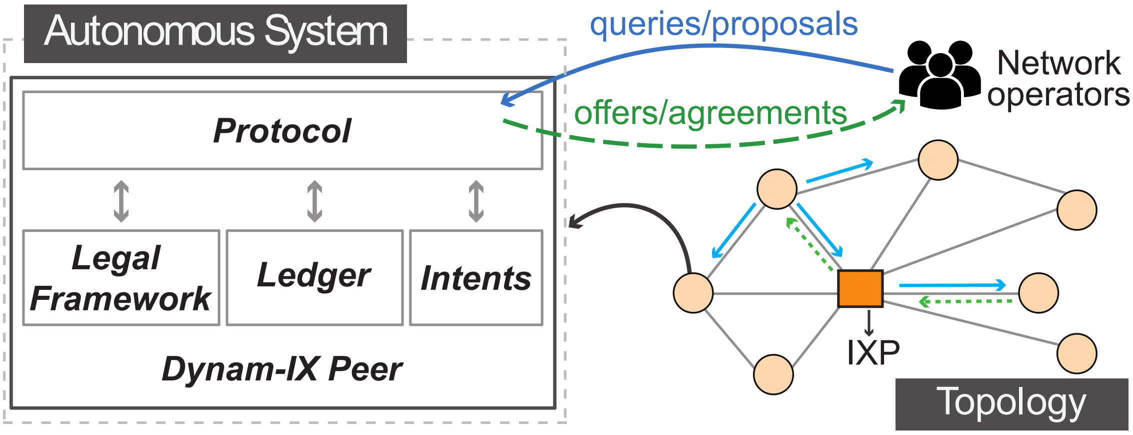
Research questions. Unleashing the massive unexplored potential of peering infrastructures to improve wide-area traffic delivery poses several research questions: (i) *how to allow ASes to express their interconnection interests and to interconnect in short time frames?* (ii) *how to offer alternatives to identify which networks are reliable to route traffic?*, and (iii) *how to guarantee the privacy of peering policies and the AS control over its routing decisions?* First, as of today, there is no well-defined method to express interconnection negotiation procedures. Second, there are currently no means to discover interconnection opportunities systematically[1]. Also, the knowledge provided by personal relationships is crucial but slow to acquire. Thus the interconnection ecosystem would benefit from a reliable mechanism to identify ASes deemed reliable to interconnect with, but this is missing. Third, the existing mechanisms (e.g., Route Servers) to speed-up the interconnection process require ASes to disclose peering policies and remove from ASes their control over routing decisions. Effectively addressing these questions will enable a series of improvements in the interconnection ecosystem.

A long-standing open problem. There have been efforts since the early 2000s to com-

[1] While PeeringDB (PeeringDB, 2018) may offer information about potential peers, the data about peering infrastructure members can be outdated or missing (LODHI et al., 2014).

moditize the bandwidth market (GONCHAROFF, 1998) and enable short-term agreements (GIOVANNETTI; RISTUCCIA, 2005; FUSARO; MILLER, 2002). These early efforts failed eventually due to lack of proper methods to specify interconnection properties and inability to control traffic more than one hop away (FERREIRA; MINDEL; MCKNIGHT, 2004), but now conditions are different. The Internet topology is now flatter, and a recent survey highlights that ∼99% of the over 1.9 M surveyed *peering agreements* were established without any formal contract (WOODCOCK; FRIGINO, 2016), indicating that operators are willing to avoid lengthy bureaucratic discussions. The emergence of new connectivity services, such Epsilon (EPSILON, 2017), MegaPort (MEGAPORT, 2017), PacketFabric (PACKETFABRIC, 2017), and Console Connect (CONSOLE, 2017), is another indication that conditions now are different. IXPs such as France-IX and NL-IX, are also offering alternatives to ease the interconnection process (FRANCE-IX, 2019; NL-IX, 2019). In the academic front, proposals include MINT (VALANCIUS et al., 2008), ChoiceNet (WOLF et al., 2014), and RouteBazaar (CASTRO et al., 2015). Both industry and academic proposals, however, suffer from two important limitations: they require ASes to disclose their interconnection policies, do not provide intuitive approaches to specify interconnection intents, and do not offer proper methods to assess the quality of peering partners.

Figure 1.2: Overview of Dynam-IX.



Source: the authors.

Our novel systematic approach. To effectively benefit from the rich connectivity of peering infrastructures to improve wide-area traffic delivery, we propose Dynam-IX, a Dynamic Interconnection eXchange. Figure 1.2 provides an overview of our proposal. A Dynam-IX peer is a node that interacts with other ASes through a *protocol* to offer and to query interconnection opportunities. To describe their interconnection interests and peering policies (e.g., pricing, SLA, duration) efficiently, operators use a *high-level interconnection intent abstraction.* To preserve the privacy of the interconnection policies, Dynam-IX is decentralized and keeps sensitive information stored locally at Dynam-IX peers. To mitigate potential losses and disputes, agreements are processed through a *legal framework* that handles the necessary steps to generate the contractual terms (based on standardized legal templates) and digitally sign contracts. During the interconnection process, ASes can query the *ledger* to get information regarding previous interconnection agreements. This information collates feedback that ASes leave periodically and helps operators decide about the quality of a potential peer. The ledger works in a distributed manner, and it is tamper-proof, offering a trustworthy manner to build trust among networks that do not necessarily rely on each other. We note that, alternatively, one could store all information in a centralized system (e.g., at the facility). Nevertheless, the business-sensitive nature of the agreements among peering infrastructure members is something that would discourage the facilities from operating such services. Once two ASes establish an agreement, they store the respective information on the ledger, and (automatically) inject via BGP the routes reflecting the new agreement and start exchanging traffic.

We summarize our contributions as follows.

1. We show the opportunity to improve wide-area traffic delivery performance (Chapter 4 and Chapter 5). We support our findings with interviews with peering coordinators from two large ISPs, a network operator from one of the top five IXPs worldwide, a survey of over 100 network operators, and analyses of traces from two relevant IXPs.

2. We design Dynam-IX, a framework that realizes such improvements in wide-area traffic delivery performance by allowing network operators to establish interconnections easily and to build trust cooperatively without solely depending on personal relationships and brand recognition (Chapter 6).

3. We evaluate a prototype implementation of Dynam-IX (Chapter 7). Our results show that an AS can establish tens of interconnection agreements within a minute while requiring negligible bandwidth (smaller than 0.2%) and storage resources from ASes and peering infrastructures.

We organize the remainder of this thesis as follows. In Chapter 2 we present the fundamentals of the interconnection ecosystem. In Chapter 3 we discuss the related approaches to improve wide-area traffic delivery and to offer new services at peering infrastructures. In Chapter 4 we present a survey with 100+ network operators showing the current practices of the interconnection ecosystem and opportunities to evolve it, while on Chapter 5 we demonstrate that peering infrastructures offer a great underutilized opportunity for improving wide-area traffic delivery. In Chapter 6 we design a solution to unleash the unexplored potential of peering infrastructures for wide-area traffic delivery and in Chapter 7, we evaluate our approach. Finally, in Chapter 8, we summarize our findings and conclusions and elaborate on an agenda towards an evolved interconnection ecosystem.

2 INTERNET INTERCONNECTION 101

This chapter presents the fundamentals of the Internet interconnection ecosystem. We first describe the basics of peering infrastructures, followed by details about the agreement models, the billing methods commonly used, and the interconnection process. The reader who is already familiar with these concepts can skip this chapter and continue on Chapter 3.

2.1 Peering infrastructures

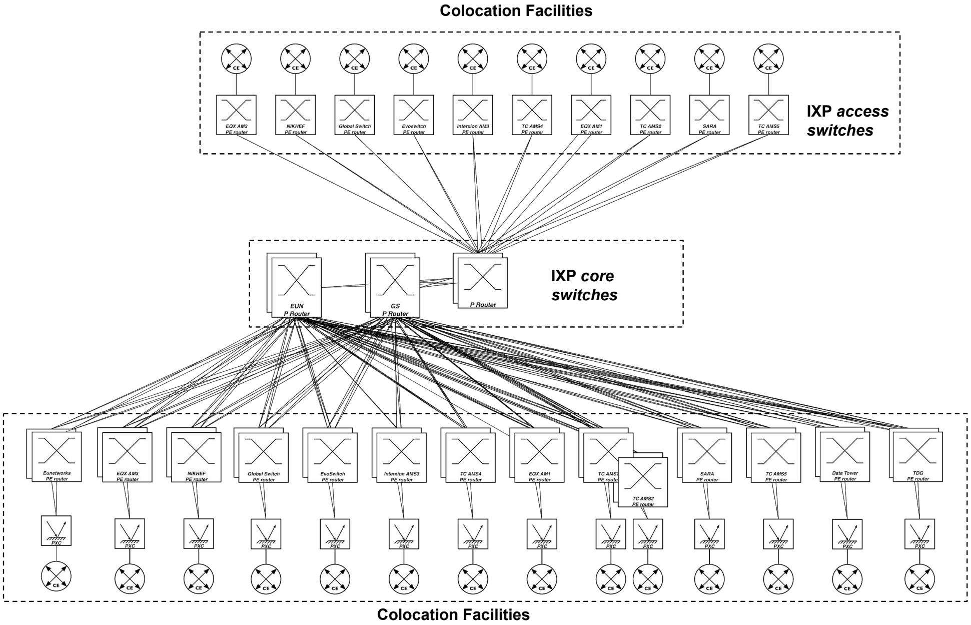
Peering infrastructures such as IXPs and colocation facilities are layer-2 switching fabrics where an AS can interconnect to exchange traffic with other AS members of the facility (AGER et al., 2012). We detail them below.

Colocation facilities. These facilities provide essential infrastructure services such as power, cooling, rack space, and physical security. ASes can then use these (paid) services to deploy their networking pieces of equipment. By using colocation facilities, ASes (especially small and medium providers) can reduce their costs and grow their business (KOTRONIS et al., 2017). ASes using the colocation facility can directly connect to other facility members or to members in other facilities operated by the same company. Colocation facilities can also offer connectivity to IXPs, cloud/content providers, and transit networks (GIOTSAS et al., 2015).

Internet eXchange Points. IXPs provide a shared switching fabric where participating networks can interconnect their routers. The switching fabric carries the traffic related to all interconnection agreements of its member ASes. Each IXP can have one or more core switches in the shared fabric for redundancy. They may also associate with several colocation facilities and install access switches to reach city-level interconnection with other networks (GIOTSAS et al., 2015). Figure 2.1 illustrates the architecture of an IXP whose members can connect from multiple colocation facilities.

Today there are over 800 IXPs (TELEGEOGRAPHY, 2018) and more than 3000 colocation facilities (PeeringDB, 2018) spread worldwide, with the largest ones carrying multiple terabits of traffic per second and enabling Autonomous Systems (ASes) to reach hundreds of other networks directly. Examples of large peering infrastructures include DE-CIX (Frankfurt), AMS-IX (Amsterdam), LINX (London), and SP-IX (São Paulo) in-

Figure 2.1: Architecture of an IXP with multiple colocation facilities.



Source: the authors.

terconnect more than 1,500 members and carry multiple terabits of traffic per second during the peak hours (DE-CIX, 2018; AMS-IX, 2018; LINX, 2018; IX.BR, 2018a). Given their characteristics as crucial elements of the Internet topology, peering infrastructures are offering to their members free value-added services specially intended to enhance the security of traffic exchange and to help to attract new members to the facility. Examples of such functions include mitigation of Denial of Services (DoS) attacks with blackholing (DIETZEL; FELDMANN; KING, 2016) and Route Origin Validation (ROV) at the Route Servers (KONSTANTARAS, 2018).

2.1.1 Interconnection types

We can classify the interconnection types in two different perspectives. The first classification relates to how a given AS connects to the facility. The second specifies how a pair of ASes connect to each other inside the peering infrastructure. We describe them below.

Connecting to the peering infrastructure. To benefit from the connectivity of the peering infrastructure, an AS needs first to interconnect to it. There are two options: local or remote. The former option is to connect directly to the facility. Such a method is used mostly by ASes located close to the metro area of the peering infrastructure. Alternatively, ASes can connect to the peering facilities remotely (a.k.a. as *remote peering*). In this approach, the AS willing to connect to the infrastructure utilizes the services from specialized companies that are connected to the target facility and offer remote peering as a service. ASes commonly use such an option when located far from the metro area of the peering infrastructure. According to recent studies (NOMIKOS et al., 2018), up to 40% of the ASes of European IXPs are making using of remote peering services.

Connecting to other ASes. ASes can establish interconnection agreements using two different approaches. The first is to establish a *bi-lateral* BGP session. In this case, the routers of both ASes will directly speak to each other to exchange routing information. While this method allows ASes to keep the control and the privacy of their routing policies, it also incurs in higher complexities due to the increasing number of members connected to peering infrastructures. Alternatively, to simplify interconnection, ASes can use a Route Server (RICHTER et al., 2014) to exchange BGP information. A common default setup allows a free traffic exchange (e.g., with no monetary compensation) with all other connected networks. Members making use of such services establish a single *multilateral* BGP session to the route server, which is then used to exchange routes with other members in the Route Server. In such a case, however, ASes lose control over routing decisions and need to disclose information about their routing policies to the peering infrastructure.

When establishing bi-lateral agreements, ASes can choose between two options. One alternative is to interconnect using the public infrastructure of the facility. We call it *public peering*. Its main advantages are the ability to establish multiple BGP sessions using the same switch port, and allow ASes to establish peering sessions without requiring any intervention from the operators of the peering infrastructure (NORTON, 2014). Alternatively, ASes can directly interconnect through a cross-connect (a.k.a. golden jump). We refer to this method as *private peering*. Such a method offers better monitoring capabilities; it is more secure and guarantees the availability of the capacity of the interconnection. However, to establish a private peering session, ASes need first to ask the operators of the peering infrastructure to set up the environment at the switching fabric.

2.1.2 Business models

To connect their networks to a peering infrastructure, ASes are usually required to pay a monthly fee based on the capacity of their interconnection ports. Its cost (and its existence) depends on the business model of the peering infrastructure, which we can classify as follows.

- **For-profit.** These peering infrastructures charge fees from their members to make a profit. Typically, a single entity (e.g., an ISP) controls the infrastructure. Common examples include colocation facilities and most IXPs in the United States of America.
- **Not-for-profit with fees.** These peering infrastructures charge fees from their members to keep the infrastructure working properly. Examples of this model include most IXPs in Europe and the Brazilian IXPs of São Paulo and Rio de Janeiro.
- **Not-for-profit without fees.** These peering infrastructures do not charge fees from their members. All costs associated with the facility are the responsibility of an external entity (e.g., government). The most popular example of this model is the Brazilian ecosystem of IXPs, where 30 of the 32 IXPs do not require any fee to interconnect.

2.2 Interconnection agreement models

Before effectively interconnecting the ASes need to decide the model of the interconnection agreement. Today, there are two well-known alternatives, namely, *transit* and *peering*, each of them offering two sub-types. ASes that interconnect in multiple locations can use a combination of the models (CASTRO; GORINSKY, 2012; GIOTSAS et al., 2014). Below we describe these methods focusing on their main differences: *reachability*, and *monetary compensation*.

2.2.1 Transit

There are two types of transit agreements. The most traditional one is the *full transit* model. In such a case an AS (e.g., an ISP) provides connectivity to the entire

Internet and charges its customers on peak-hour traffic (e.g., through the 95th-percentile basis). A variant of this model is the *partial transit* agreement. This model differs from the full transit as its reachability is limited to a subset of Internet routes (normally the customers and the peers of the provider) and it has a lower price (due to its limited reachability) (GIOTSAS et al., 2014).

2.2.2 Peering

Peering agreements have emerged a model to reduce interconnection costs, achieve lower latencies, and offer more control over routing decisions (NORTON, 2014). In a peering agreement, two ASes interconnect directly (no transit provider involved) and agree to exchange traffic originated/destined from/to their networks or their cone of customers (LUCKIE et al., 2013) reciprocally. We can classify peering agreements in two types, namely, *settlement-free peering* and *paid peering*. Both types provide the same reachability but differ in the existence of monetary compensation. As the name suggests, in a settlement-free peering agreement, there is no monetary compensation involved. Alternatively, in paid-peering agreements, one of the ASes should pay the other. In recent years, the number of paid-peering is growing (ARCEP, 2018). The decision between the two models depends if one of the AS will benefit the most from the agreement, e.g., due to traffic imbalance or route diversity from a larger ISP. Table 2.1 shows a comparison of the interconnection models.

Table 2.1: Comparison of interconnection agreement models.

Models	Reachability	Monetary compensation
Full Transit	Internet	Yes
Partial Transit	Customer Cone and Peers	Yes
Settlement-free Peering	Customer Cone	No
Paid Peering	Customer Cone	Yes

Source: the authors.

2.3 Billing methods

Typically, interconnection agreements require monetary compensation paid by one of the ASes to the other. The exception is settlement-free peering agreements. On the other cases, ASes may employ methods that vary from more traditional ones such as Committed Information Rate (CIR) and 95th-percentile to more complex ones where the

AS pays to send traffic and earn revenue when receiving traffic. We detail them in the following subsections.

2.3.1 Committed Information Rate - CIR

With CIR, the customer AS will pay a fixed amount based on the capacity of the interconnection independently of its actual usage. While this model does not incur in any need for constantly measuring the link utilization, its fixed-capacity model could make capacity planning decision harder. For example, an AS that has peak traffic much higher than the average traffic during the remaining time would need to decide between one of the following alternatives. It could establish an interconnection agreement for the peak capacity at the expense of higher costs, or it can prefer a lower capacity agreement and eventually face congestion scenarios. While the second alternative seems more conservative, it can harm the revenues of the AS since some of its customers might decide to change the provider due to the congestions. Nonetheless, according to a survey from 2014 with network operators (GILL; SCHAPIRA; GOLDBERG, 2013), 39% of the ASes were still using a billing model different than the 95th-percentile. Similarly, in a preliminary survey we did with 25 network operators, 44% reported that they use CIR in their interconnection agreements.

2.3.2 95th-Percentile

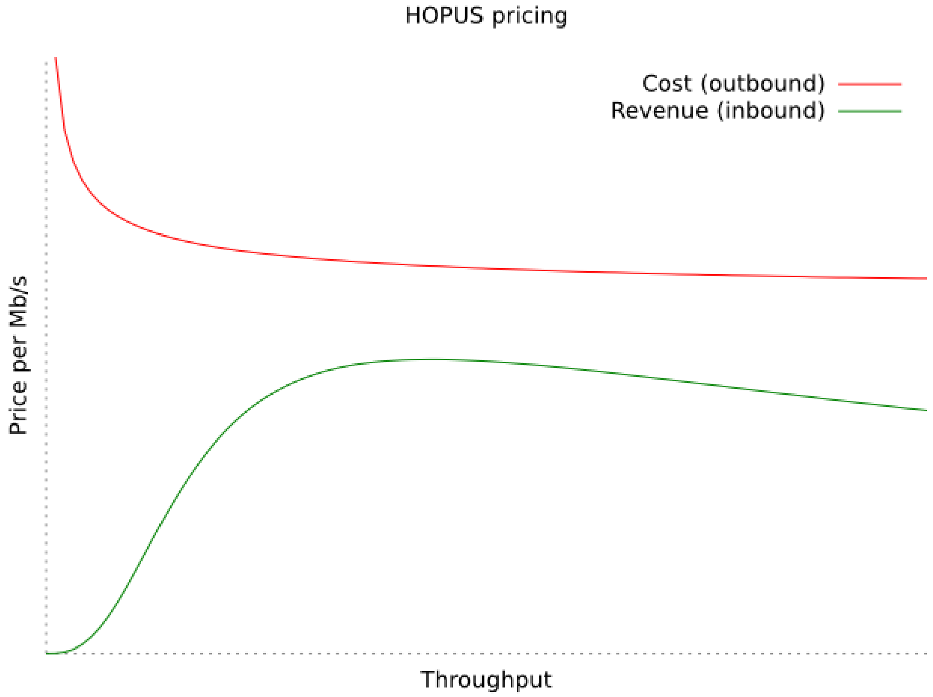
In the 95th-percentile model, the provider AS measures the traffic volume at every interval (e.g., 5 minutes) during a billing period (e.g., one month) (DIMITROPOULOS et al., 2009). The traffic volume is either the sum of inbound and outbound traffic or the maximum between them. At the end of the period, the provider will charge the customer by the 95th-percentile peak rate. This method simplifies the capacity planning process as the amount paid by ASes depends on the actual usage of the interconnection capacity. Also, ASes are not charged by their peak utilization, which allows ASes to handle some traffic bursts without requiring additional payments. There is an alternative method that combines the 95th-percentile with the CIR. In that case, the ASes agree on a minimum billing capacity for each period and, if traffic exceeds this capacity, the provider charges the customer using the 95th percentile method.

2.3.3 The Hopus model

Hopus (HOPUS, 2018a) is a company that offers a peering infrastructure (with monthly port fees) where all ASes that join agree to follow the same price model for exchanging traffic, independently of their size and traffic volumes. Differently from the traditional methods where a (customer) AS pays to send/receive traffic to/from its transit provider, in the Hopus model an AS sending traffic will pay for it, but it will also earn revenue when receiving traffic.

Traffic is measured monthly following the 95th-percentile and considering all ports of the AS. Prices (in Mbps) are calculated using a continuous function where the outbound traffic follows a decreasing function providing degressive prices (similarly to IP transit), and the inbound traffic price increases progressively in order to support the development of access networks' infrastructures then slowly decreases to accommodate high capacity transport. Figure 2.2 illustrates Hopus pricing functions.

Figure 2.2: Hopus pricing model.



Source: Hopus.

The rationale for this model is to incentivize ASes to connect and allow them to exchange traffic with all other members without needing to negotiate specific agreements with each AS. Today Hopus interconnects more than 60 ASes across 14 peering facilities.

2.4 Interconnection process

Establishing interconnection agreements is mostly an ad-hoc process without a set of well-defined steps to follow. Nonetheless, based on our interviews with network operators and previous work (NORTON, 2014) focused in the establishment of peering agreements, we have identified four main phases, namely (i) finding an interconnection partner, (ii) discussing the interconnection properties, (iii) legal procedures, and (iv) configuring the infrastructure to reflect the new arrangement, that comprise the entire process. In the following subsections, we detail them.

2.4.1 Finding an interconnection partner

Once an AS has identified the need to establish a new interconnection agreement (e.g., to avoid congestion, to reduce latency) it needs to find other ASes that can offer alternatives to address the issues (e.g., provide connectivity to a given destination). As of today, personal relationships and brand image largely influence the interconnection process (MEIER-HAHN, 2017). Thus, it is common to observe, for example, messages in mailing lists of NOGs asking for information about ASes that could provide the requested service (e.g., transit in a specific location) or related to previous interconnection experiences with a given AS (NANOG, 2019d; NANOG, 2019a; NANOG, 2019b; NANOG, 2019c).

Websites also help operators to identify potential peering partners. Peering infrastructures such as IXPs frequently provide a list of connected members in their websites. Additionally, services such as PeeringDB (PeeringDB, 2018) and Inflect (INFLECT, 2019) offer information related to interconnection alternatives. PeeringDB is a database maintained by the network operators' community containing information about the ASes exchanging traffic in each peering infrastructure, and Inflect offers a service to identify alternatives to reach a given AS from a specific interconnection facility. Finally, it is worth noting that entities such as peering infrastructures, Regional Internet Registries (RIRs), and other associations/federations (e.g., IX-Forum, Euro-IX, NOGs) organize face-to-face meetings (FORUM-IX, 2016) whose primary purpose is to introduce people to each other and foster interconnection among ASes.

2.4.2 Discussing the interconnection properties

After identifying a potential interconnection partner, the ASes in question need to discuss the properties of the agreement. Typically, network operators perform this step by e-mail or in face-to-face meetings. We can classify the attributes into data plane, control plane, and operation information. We describe them below.

- **Data plane information:** includes attributes such as bandwidth, latency, packet loss.
- **Control plane information:** parameters mostly related to the AS-Paths that operators will use to forward the traffic.
- **Operation information:** aspects that are indirectly related to traffic forwarding such as NOC availability, escalation path, and incident response time.

The set of parameters that operators discuss depends on the type of agreement (e.g., transit, peering) and on the AS peering policy (e.g., open, selective, restrictive). Due to their nature, transit agreements tend to require the discussion of all three set of properties, while for peering interconnections the parameters may vary. For settlement-free peering interconnections, in most cases, there will be no discussion about data and control plane parameters. ASes discuss operation information typically when one of them has a selective or restrictive peering policy. In such cases, the AS with selective peering policy usually requires the other to prove that it can satisfy some requirements (e.g., traffic volumes, 24x7 NOC). In paid-peering agreements, in addition to the operation information, operators can also discuss data and control plane parameters.

2.4.3 Legal procedures

The third step of the interconnection process is dependent on the type of interconnection agreement. When establishing peering agreements, network operators typically do not require any legal procedure and merely configure their infrastructure after converging on its properties. In fact, a recent survey from PCH has shown that more than 99% of 1.9 M peering interconnection agreements tend not to be formalized in contracts (WOODCOCK; FRIGINO, 2016). These are mainly settlement-free peerings, which are agreements that do not involve monetary aspects. In these, if an AS is not satisfied anymore with the interconnection, it can end the BGP session, and ASes will stop exchanging traffic. Peering agreements that require legal procedures usually are the ones involving money.

For transit agreements ASes tend to require legal procedures since such kind of agreement involves a customer AS paying a provider AS. Depending on the length of the interconnection, the monetary compensation, and the obligations of each AS, such a step can be time-consuming as it may require several rounds of meetings among lawyers and generate documents that can vary from a few pages to more than 30 (INC., 2019; COMCAST, 2019). According to our survey respondents, legal procedures is one of the most lengthy steps when interconnecting.

2.4.4 Configuring the infrastructure to reflect the new agreement

Once the ASes converged on the interconnection properties and have signed a contract (in some cases), it is necessary to configure their infrastructure to reflect the new interconnection agreement. The only mandatory procedure is configuring their BGP border routers to establish a BGP session and start exchanging routes. Other configurations, such as creating VLANs, setting MPLS tunnels, might be required and are directly related to the AS internal procedures. Depending on the complexity of the interconnection (e.g., single x multiple interconnection points), ASes can easily automate this step using tools such as Kees (COLOCLUE, 2019).

3 RELATED WORK

To evolve the interconnection ecosystem and improve wide-area traffic delivery, Industry and Academia have been proposing different alternatives. These proposals focus on innovating various parts of the ecosystem, including the inter-domain routing, the peering infrastructures, and the interconnection process. Below we detail and discuss how these solutions compare to ours (staring from the ones less related to ours).

3.1 Innovating inter-domain routing

One way to advance inter-domain routing is to address specific limitations of BGP by adding new features. Examples include multi-protocol extensions (KATZ et al., 2007), BGP extended (REKHTER; SANGLI, 2006), the use of large communities (HEITZ et al., 2017) to carry (more) meta-information, and BGP session culling (HARGRAVE et al., 2018), to mitigate negative impact on networks resulting from maintenance. While these represent important steps forward, the innovation and impact at large are questionable. Due to the difficulty of modifying BGP itself (SAMBASIVAN et al., 2017), researchers and engineers try to overcome the limitations with external systems. Edge Fabric (SCHLINKER et al., 2017) and Espresso (YAP et al., 2017) strive to improve traffic engineering by considering multiple routes and monitoring available bandwidth. While they ultimately increase interconnection utilization, this is mainly beneficial for the operators of these proprietary solutions. In contrast, Dynam-IX aims to benefit all ASes physically connected to the peering infrastructure.

3.2 Peering infrastructures as service enablers

Due to their nature as convergence points of hundreds of ASes, peering infrastructures have been advocated as places to spur innovation and promote new services for network operators. Control eXchange Point (KOTRONIS et al., 2016) proposes the use of IXPs to establish paths with QoS guarantees, by stitching together inter-domain links at IXPs. The introduction of SDN at IXPs (SDXs) (GUPTA et al., 2014) and several refinements and extensions (GUPTA et al., 2016; CHIESA et al., 2016; ANTICHI et al., 2017) aim to offer operators more fine-grained control over their routing policies. They also simplify more complex usage scenarios, e.g., improved traffic engineering or advanced DDoS mitigation (DIETZEL et al., 2017) and allow the use of economic aspects in the policy configuration at IXPs (GRIFFIOEN; WOLF; CALVERT, 2016). These proposals can work together with Dynam-IX and enable network operators to optimize route configuration after the establishment of the interconnection agreement.

3.3 Evolving the interconnection process

To properly evolve the interconnection process proposals should achieve a set of requirements, including (i) the ability to interconnect in short time frames; (ii) preserve the privacy of peering policies; (iii) offer a method to identify reliable peering partners; and (iv) provide an expressive interface for ASes to express their interconnection policies and interests. Academia and Industry have been proposing alternatives to provide a fast and more flexible interconnection process. We survey them below.

MINT - A Market for Internet Transit. One of the earlier approaches for automating the interconnection process is MINT (VALANCIUS et al., 2008). Its goal is to allow ASes to compose end-to-end paths by combining multiple interconnection links. A centralized entity called *mediator* is responsible for storing information related to the available interconnection links. ASes interested in offering connectivity services can register offers on the mediator. Each offer consists of a source AS, a destination AS, the available bandwidth, and the minimum leasing time. When an AS wants to compose a path to reach a given network, it sends a request to the mediator containing the destination, the amount of bandwidth, and the length of the interconnection. The mediator will then try to compose a path with the minimum price that satisfies the bandwidth requirements. Once it finds a proper alternative, the mediator will inform (no details provided) the involved ASes to configure the respective links.

While MINT may achieve the goal of composing end-to-end paths, several aspects hinder its practicality. First, by relying on a centralized entity to store offers and perform the matching among available interconnection links and requests MINT is forcing the ASes to disclose their interconnection policies, which is something that operators prefer not to do (CHIESA et al., 2017b). Second, relying on the mediator to perform the matching between a request and an offer prevents the ASes (both customers and providers) from selecting their interconnection partners. Not being able to choose the peering partners could result in undesired situations such as interconnecting with a competitor or sending traffic through a country with surveillance. Alas, MINT does not offer a method to identify reliable peering partners neither defines an interface for network operators to express their interconnection policies and interests. Finally, the list of interconnection links attributes that MINT considers is narrow as it only allows the specification of price, bandwidth, and leasing time. Parameters such as latency, packet loss, link availability are absent in MINT.

ChoiceNet - An Economy Plane for the Internet. ChoiceNet (WOLF et al., 2014) follows an approach similar to MINT. The main difference lies in the fact that in ChoiceNet, the ASes can select the other ASes that will provide the connectivity, thus gaining control over interconnection decisions. However, ChoiceNet keeps all the other limitations present on MINT, such as revealing ASes interconnection policies to other ASes and the operator of ChoiceNet, not offering a method to identify reliable peering partners, and the absence of an interface for ASes to express interconnection policies and interests.

Route Bazaar - Automatic Interdomain Contract Negotiation. Route Bazaar (CASTRO et al., 2015) proposes a high-level design of a distributed approach to allow ASes to compose end-to-end paths. In Route Bazaar, ASes rely on a public ledger to store information about interconnection offers, existing interconnection agreements, and pieces of data describing the "performance" of the ASes during the interconnection agreement as well as proofs of payments related to the contracts.

In Route Bazaar, an AS intending to announce an interconnection link (pathlet (GODFREY et al., 2009)) registers on the public ledger a piece of information containing the advertising AS, the destination AS, the price, latency, and bandwidth of the link. Other ASes can then compose end-to-end paths by combining multiple pathlets announcements. After choosing the desired pathlets, the AS asks the "owners" of the pathlets to establish an interconnection agreement. ASes will then agree on an identifier for each interconnection (one per selected pathlet) and commit to the ledger this information and the properties of the interconnection (latency and bandwidth). During the agreement (or after) the pathlet owners will publish on the ledger forwarding proofs indicating that they are respecting (or not) the agreed terms. Similarly, at the end of the agreement, the customer AS will store on the ledger a proof of payment. Other ASes can then use these pieces of information (forwarding proofs and proof of payment) to decide whether or not an AS is reliable for interconnecting. While this approach removes the centralized entity, it still exposes the interconnection policies of the ASes. Differently from MINT and ChoiceNet, Route Bazaar discusses alternatives to provide information to help ASes decide whether or not to interconnect with another AS. Such a method, however, leaks information about the interconnection agreement properties. It also reveals the interconnection policies of the ASes since the offers are publicly available on the ledger. Alas, Route Bazaar does not define an abstraction that ASes can use to express their interconnection policies and interests.

On-demand connectivity companies. Recently, interconnection companies such as Megaport (MEGAPORT, 2017), Packet Fabric (PACKETFABRIC, 2017), Epsilon (EPSILON, 2017), and ConsoleConnect (CONSOLE, 2017) emerged offering on-demand connectivity to cloud providers (e.g., Amazon AWS, MS Azure, Google GCC) for networks connected to their Points of Presence (PoPs). In these approaches, a single AS (the interconnection company) offers connectivity to other ASes. When an AS wants to interconnect to another AS, it uses a web interface or an API to request to the interconnection company to establish an interconnection agreement with the intended destination.

We can compare such alternatives to Dynam-IX in two different perspectives. If we consider the interconnection company as the one providing the connectivity to the desired destination, it will be the only AS able to announce connectivity (and profit from it). In contrast, our solution allows all ASes connected to the peering infrastructures to offer connectivity to other parts of the Internet. If we consider the interconnection companies as the ones responsible for providing the physical connectivity among the ASes, the ASes interested in interconnecting need to carry a separate (external) negotiation before using the services of the interconnection company. In Dynam-IX, instead, ASes wishing to interconnect benefiting from the connectivity of the peering infrastructure can negotiate directly using our interconnection intent abstraction.

Regarding the build of trust, in the former case, ASes need to trust the company providing the connectivity services, while in the latter case a method to allow any pair of ASes to build trust is absent. Additionally, in the second scenario, ASes need to reveal their interconnection policies to the companies offering the connectivity service. Table 3.1 shows a comparison among the existing alternatives to evolve the interconnection process and ours.

Given this state of affairs, we posit that an approach to effectively unleash the vast unexplored potential of peering infrastructures for improving wide-area traffic delivery is missing. Thus, to help the design of this approach, we interviewed and surveyed more than 100 network operators to learn the current interconnection practices and how network

Table 3.1: Comparison of existing approaches for evolving the interconnection process.

Approach	Architecture	Partner Selection	Expresive Interface	Privacy-preserving	Trust
MINT	Centralized	Mediator	No	No	No
ChoiceNet	Centralized	ASes	No	No	No
Route Bazaar	Distributed	ASes	No	No	Yes
On-demand connectivity companies	Centralized	ASes	Partial	No	No
Dynam-IX	Distributed	ASes	Yes	Yes	Yes

Source: the authors.

operators would benefit from a solution for dynamic interconnection. We discuss the results of our survey in the next chapter.

4 A SURVEY ON THE CURRENT INTERCONNECTION PRACTICES

While the Internet topology has evolved significantly, it is unclear whether if the interconnecting continues to be driven by personal relationships, brand image, and word-of-mouth in a highly ad-hoc and lengthy manner. A first intuition indicates that the interconnection process has also changed. A recent survey by Packet Clearing House (PCH) indicates that more than 99% of the peering agreements are handshake-based (WOODCOCK; FRIGINO, 2016), hinting that the interconnection process may be faster today. Additionally, in recent years, on-demand connectivity companies and proposals from academia have emerged promising the provisioning of interconnection agreements in short-time frames (EPSILON, 2017; MEGAPORT, 2017; PACKETFABRIC, 2017; CONSOLE, 2017; VALANCIUS et al., 2008; WOLF et al., 2014; CASTRO et al., 2015). While these indicate that things might have changed, recent studies spotted that at peering infrastructures, where a large portion of the interconnection agreements takes place today and exists a Route Server (RS) to ease the exchange of reachability information (i.e., L3), operators are preferring not to use the features of the RS due to security and routing control issues (CHIESA et al., 2017b). In these cases, operators continue to follow a process that consists of identifying the need for a new interconnection (e.g., through monitoring systems) and a set of possible candidates (e.g., checking PeeringDB (PeeringDB, 2018)), then they need to discuss the properties of the interconnection agreement (e.g., Service Level Agreement, length, peering or transit), eventually sign a legal contract, and finally execute the agreement by configuring their border routers (NORTON, 2014).

To shed light on the current interconnection practices, its limitations, and how the interconnection ecosystem can benefit from an improved interconnection process, we conducted a series of interviews and surveyed more than 100 network operators and peering coordinators. We start this chapter describing the survey methodology and characterizing the survey respondents (§4.1). Next, we present the results of the questions about the current practices of the interconnection ecosystem (§4.2). Finally, we show and discuss the network operators' perceptions about limitations and possible improvements in the interconnection process (§4.3).

4.1 Survey methodology and respondents characterization

Methodology. To elaborate the survey, we first interviewed a peering coordinator from a tier-1 ISP, and three network operators from peering infrastructures, one from a large IXP, one from a medium-sized IXP, and one from a medium-sized colocation facility. Based on the interviews, we proposed a set of questions about the current practices of interconnection ecosystem and a list of the most relevant use cases (and consequences) for improving the interconnection process. We then asked a subset of network operators to provide feedback about the structure and clarity of our survey. Then, we circulated the survey (as a Google Form[1]) to several mailing lists of Network Operators Groups, including NANOG, RIPE, AusNOG, DENOG, LACNOG, APNIC-talk, ARIN-tech, IX.br, AFRINIC, AFNOG, between December 2017 and January 2018. We also have sent the survey to the AS members of a large European IXP and made it available through a blog post at <ipspace.net> website. We collected 106 answers from network operators and peering coordinators from ASes. Below we classify the respondents according to their *Region*, *AS Type*, and *AS Size*. Since we told the survey respondents that we will not reveal their identities, we will not present results correlating multiple dimensions (e.g., a large content provider in Brazil).

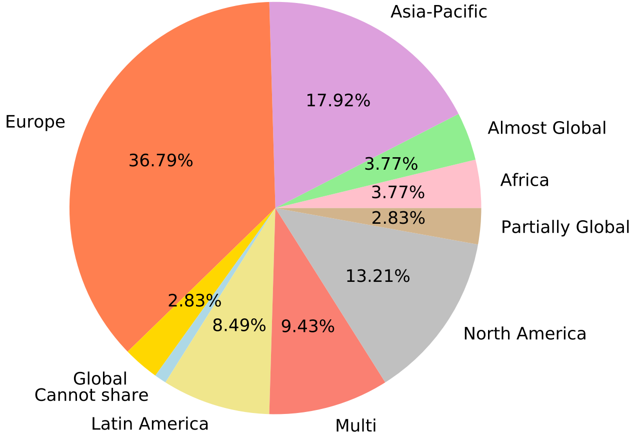
Region. We asked the respondents to select the region(s) in which their AS is present. The participants could select among the five regions with Regional Internet Registries (RIRs). We classified the ASes that are present in two or more regions in the following groups: *multi* for ASes in two regions, *partially global* for ASes in three regions, *almost global* for ASes in four regions, and *global* for ASes present in all regions. The respondents could also select not to share the region of their ASes. Figure 4.1 presents the distribution of respondents by region.

Among the respondents, 80.2% are present in a single region, 18.9% correspond to ASes with infrastructure in multiple regions, and 0.9% did not share their location. We note that while on a different scale, the regional distribution of respondents resembles the proportion of ASes registered in each of the five RIRs. To illustrate, as of today, 38.06% of the ASes are in the RIPE region, 32.18% are in the ARIN region, 17.49% are in the APNIC region, LACNIC accounts for the 10.38% of the ASes, and AFRINIC 1.89% (MAIGRON, 2019).

AS Type. To characterize the type of the ASes, we asked the survey participants to

[1] Available at <https://bit.ly/2kQtsNg>

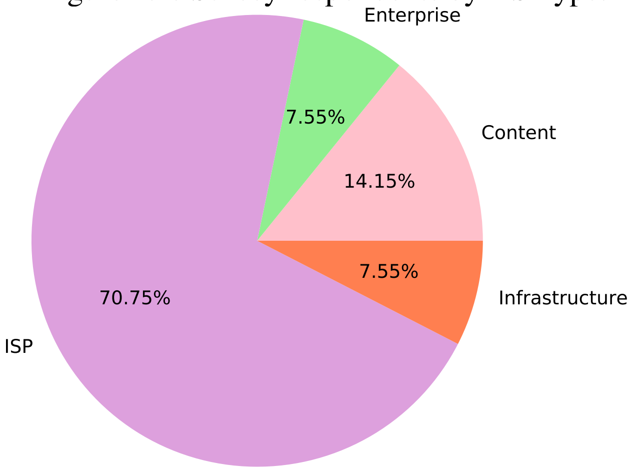
Figure 4.1: Survey respondents by region.



Source: the authors.

select the terms that best describe their ASes. We classify the ASes into four different types, namely, Internet Service Providers (ISPs), Content Providers, Enterprise Networks (e.g., companies, Universities, Research and Education Networks), and Infrastructure Providers (e.g., IXPs, RIRs, DNS Providers). Figure 4.2 shows the AS types of survey respondents.

Figure 4.2: Survey respondents by AS Type.

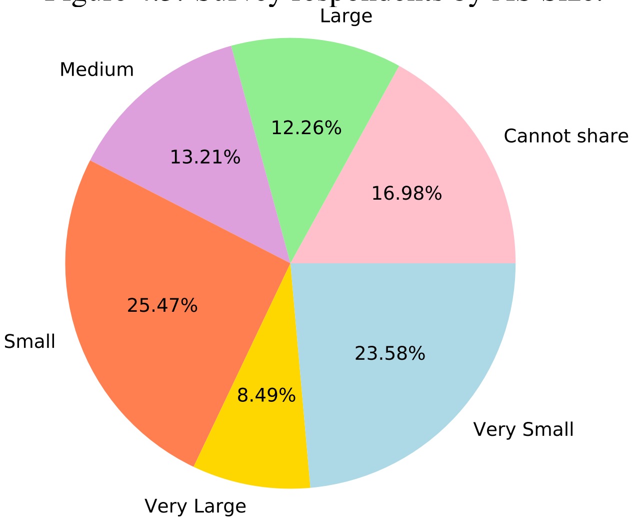


Source: the authors.

The vast majority (70.8%) of the respondents operate ISPs, while 14.2% are Content Providers, 7.5% are Entreprises, and 7.5% represent Internet Infrastructure ASes. Regarding the ISPs, we can classify them into three different categories, namely, tier-1 (6.7%), transit (64%), and access networks (29.3%).

AS Size. Finally, to obtain a view of the size of the respondents ASes, we asked the operators to indicate an estimate of the number of end-users of their ASes. Options include five different sizes (very small, small, medium, large, and very large) and the possibility not to share this information. Figure 4.3 shows the characterization by AS size.

Figure 4.3: Survey respondents by AS Size.



Source: the authors.

The respondents of our survey are heterogeneous when considering their size in terms of end-users.[2] 23.6% are very small networks with up to 1,000 end-users, 25.5% are small providing services to 1,001-10,000 users, 13.2% are medium-sized networks that offer connectivity to 10,001-100,000 customers, and 12.3% are large networks with 100,001-1,000,000 users. Finally, 8.5% represent very large networks with more than 1,000,000 end-users. 17% of the survey respondents preferred not to share the size of their networks.

4.2 Current interconnection practices

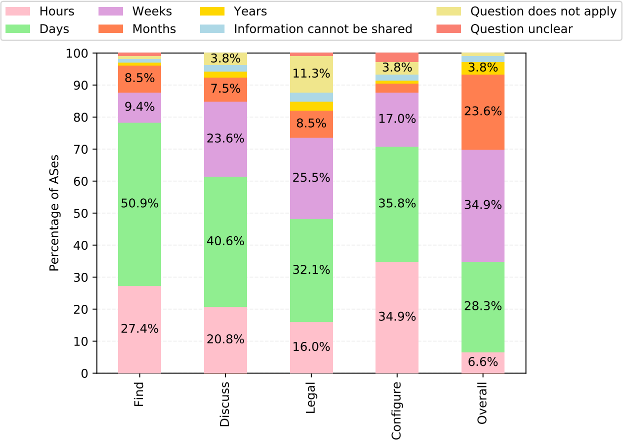
To learn the current interconnection practices, we asked the respondents four questions (i) *how long does it take on average to establish a bilateral agreement?*; (ii) *what are the reasons to establish an interconnection agreement?*; (iii) *what are the reasons to renegotiate the parameters of an existing interconnection agreement?*; and (iv) *what parameters network operators use to discuss when establishing an interconnection agreement?*. Below we describe the overall results and highlight the main differences (if they exist) when comparing the answers by region, AS type, and AS size.

How long does it take on average to establish a bilateral agreement? We divided the interconnection process into four steps: *identifying a potential peering partner*, *discussing the properties of the interconnection agreement*, *legal procedures*, and *configuring the agreement* (see Chapter 2). We then asked network operators to specify the

[2] Or AS members in the case of ASes operating IXPs.

average time (e.g., hours, days, weeks) required in each of these steps and to complete the overall process. Valid answers also included *question does not apply* (e.g., an AS may never require legal procedures), *question unclear*, and *information cannot be shared*. Figure 7.2(a) shows the summary of the answers.

Figure 4.4: Average time to establish an agreement.



Source: the authors.

Among the different steps, finding a peering partner is the shortest one, requiring hours or at most days for the majority (78.3%) of ASes. Possible reasons include the facts that ASes tend to have monitoring systems that indicate candidate ASes and because PeeringDB (PeeringDB, 2018) and websites from IXPs and colocation facilities usually provide information about their members, including contact details, which helps to accelerate the process. It is not uncommon, however, to observe in mailing lists NOGs, operators asking for contact information of a given AS (NANOG, 2019b; NANOG, 2019c), which may indicate that sometimes is not trivial to find the person responsible for establishing agreements in some ASes. Discussing the properties (e.g., technical and pricing aspects) of the interconnection agreement demands mostly days (40.6%) or weeks (23.6%) to complete. Regarding the legal phase, 11.3% of the ASes answered that this step does not apply for them. This scenario is possibly related to the fact that these ASes may have an open peering policy, thus interconnecting to all ASes interested in exchange traffic. A recent survey by Packet Clearing House (PCH) shows that more than 99% of the peering agreements are handshake-based, thus no legal contracts involved. After completing the legal step, ASes need to configure their routers (equipment) to reflect the new settlement. Such a procedure tends to require hours (34.9%) or days (35.9%) to complete. Finally, the respondents indicated that the overall interconnection process usually requires days (6.6%), weeks (28.3%), or even months (34.9%) to complete. When analyzing the results by specific properties (e.g., region, AS size, AS type), we can identify the following patterns.

ASes located in Africa and Latin America have longer interconnection times. The same condition holds for very small and very large ASes. These results support our case that interconnecting continues to be a lengthy process.

What are the reasons to establish an interconnection agreement? To identify the leading factors, we asked the network operators to select one or more alternatives that represent reasons to establish a new interconnection agreement. Possible answers were *increase capacity*, *decrease capacity*, *reduce costs*, *reduce number of hops*, *reduce latency*, *increase revenue*, *improve resilience*, and *other*. Table 4.1 presents the results.

Table 4.1: Reasons to establish and to renegotiate an interconnection agreement.

	Establish	Renegotiate
Latency	81.1%	29.2%
Costs	66%	53.8%
Resilience	64.2%	41.5%
Increase Capacity	59.4%	50.9%
Reduce Hops	56.6%	20.8%
Revenue	12.3%	8.5%
Reduce Capacity	0.9%	3.8%
Other	8.5%	5.7%
QNA	0%	12.3%

Source: the authors.

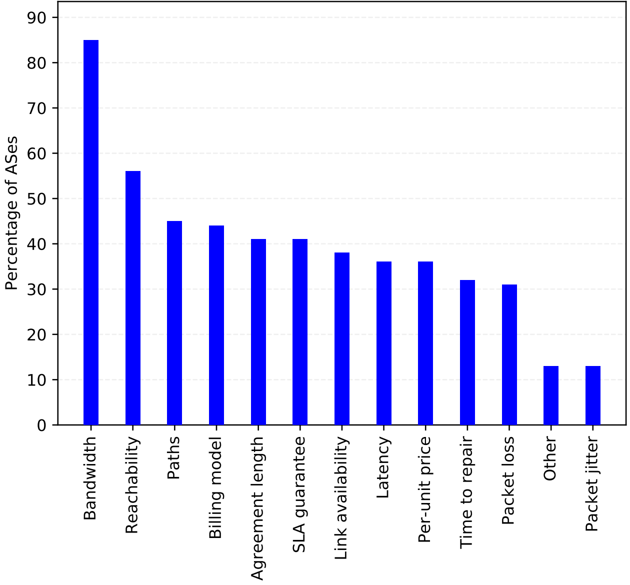
Reducing latency emerges as the most common reason (81.1% of network operators) to establish a new interconnection agreement. Other main reasons are reducing costs (66%), improving resilience (64%), increasing capacity (59.4%), and reducing the number of hops (56.6%). This order is consistent when analyzing the answers by region, AS size, and AS type. Nine ASes, mostly the largest ones, also consider *other* reasons to interconnect, such as commercial aspects, to increase their footprint, to improve peering stats (e.g., AS-Rank), and to build political relationships.

What are the reasons to renegotiate the parameters of an existing interconnection agreement? Similarly, we asked the operators to identify the reasons to renegotiate the properties of an interconnection agreement. The set of alternatives is the same as the previous question. Table 4.1 shows the results.

The five main reasons to renegotiate an interconnection agreement are the same as the ones to establish a new agreement. However, for this scenario, they appear in a different order. The top motivation is to reduce the interconnection costs (53.8%), followed by increasing capacity (50.1%), and improving resilience (41.5%). Reducing latency (29.2%) and the number of hops (20.7%) complete the list. Interestingly (but not surprisingly), the main reason to renegotiate an agreement is an economic one, while the number-one motivation to interconnect with a new partner is performance-related. Other takeaways from the answers are the higher number (when compared to the previous question) of ASes that renegotiate an interconnection agreement to decrease capacity, and the 11.3% of ASes that selected *question does not apply*, probably indicating that they do not renegotiate interconnection agreements. Finally, when analyzing the answers by region and AS size, we identify that ASes with a global footprint and large ASes are the ones interested in increasing their revenues by renegotiating interconnection agreements.

What parameters network operators use to discuss when establishing an interconnection agreement? Finally, we asked network operators to indicate the parameters that they use to discuss when interconnecting. Figure 4.5 shows the results.

Figure 4.5: Parameters discussed before establishing an interconnection agreement.



Source: the authors.

Bandwidth is the most common attribute discussed by the survey respondents (86%). Other relevant aspects are reachability (56.6%), the paths to steer traffic during the agreement (45.3%), the billing model (44.3%), the guarantee of the SLA properties (41.5%), and the agreement length (41.5%). There is no significant difference when comparing the results by region, AS size, and AS type.

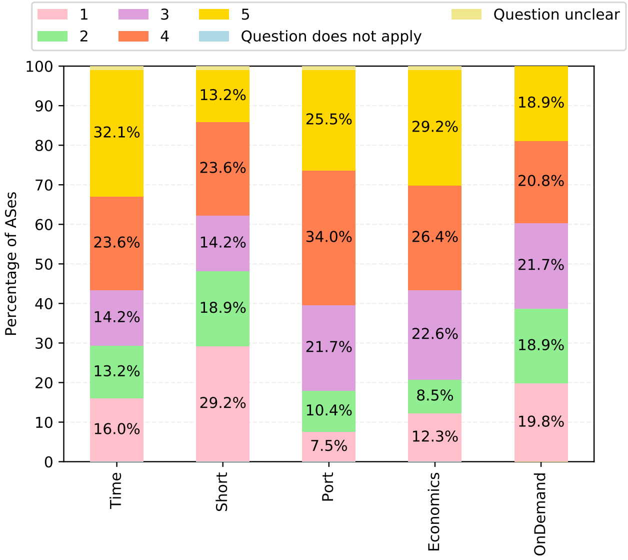
4.3 Perceptions on evolving the interconnection process

To understand the network operators interests in evolving the interconnection process, we inquired them about their interest in reducing the interconnection setup time and in possible use cases enabled by a more dynamic interconnection process. Additionally, to help us design Dynam-IX, we asked the survey respondents about their perceptions of existing on-demand connectivity companies and potential issues caused by their utilization.

4.3.1 Possible improvements

We aimed at identifying the most relevant aspects of the interconnection ecosystem that could evolve. Based on the interviews we presented to the operators a set of improvements including a) *reducing the interconnection setup time*, b) *short-time interconnection for traffic engineering*, c) *increasing peering port utilization*, d) *benefiting from new economic opportunities*, and e) *ordering network services on-demand*. Then, we asked *how important would be to evolve each of these aspects?*. We requested the operators to give a score on a scale from 1 to 5, indicating the relevance of each aspect. Figure 4.6 shows a summary of the answers.

Figure 4.6: Operators' perception about possible improvements in the interconnection ecosystem.



Source: the authors.

Reducing the interconnection setup time. Most operators (56%) have indicated that reducing the interconnection setup is a valid improvement (scores 4 and 5). Not surprisingly, today we are observing several efforts (EPSILON, 2017; MEGAPORT, 2017; PACKETFABRIC, 2017; CONSOLE, 2017; VALANCIUS et al., 2008; WOLF et al., 2014; CASTRO et al., 2015), both from Industry and Academia, offering alternatives to speed up the interconnection process. Interestingly, 80% of the content providers consider reducing interconnection setup time valuable. These results strengthen our case for evolving the interconnection process. Being able to interconnect faster can provide different benefits, ranging from removing the burden from network operators to improving wide-area traffic delivery performance as we describe below.

Enhanced traffic engineering. Network operators continuously perform inter-domain traffic engineering to optimize traffic flow in response to events such as topology and traffic demand changes. If operators could quickly establish short time interconnection agreements, there would be a richer set of possibilities for traffic engineering. Such additional capacity is desirable to cope with sudden traffic surges, congested paths, routes with high latencies, and link failures. In all these cases, an operator would benefit from a short-term interconnection agreement to improve performance (thus user experience) or to restore connectivity after a link failure. Roughly 37% of the survey respondents considered the traffic engineering use case relevant (scores 4 and 5) for their operations while 14% were neutral. When focusing our analysis per AS size, we identify that the majority (56%) of the very small ASes consider this a significant improvement. Similarly, when analyzing by AS type, approximately 47% of the content providers are interested in this feature.

Increasing peering port utilization. The majority (60%) of the survey respondents have indicated this as a relevant improvement. Likewise, Deutsche Telekom recently reported that increasing resource utilization by 1% or 2% could result in saving millions of dollars in future infrastructure investments (BORNSTAEDT, 2017). The ability to establish interconnection agreements in short time frames especially helps in the following cases. First, it reduces the time until an AS starts using a new, recently deployed port. While a Route Server can assist in quickly connecting to networks with an open peering policy, ASes cannot leverage RSes when they must implement other types of peering policies or need more control over their routes. Second, the ability to establish short-term interconnections and route traffic using the IXP port can help steer traffic from a congested transit link to one with spare capacity, increasing its utilization. Otherwise, the AS would need to go through a potentially lengthy process to add capacity to the transit link.

Economics. Increasing revenue or reducing interconnection costs is also an essential improvement for 56% of the survey respondents. The ability to establish interconnection agreements in short time frames can generate novel business opportunities, increased revenues for ISPs, and cost-saving alternatives for ASes. Consider the following examples. First, before establishing long-term agreements, a customer AS wants to try an interconnection for a short period (e.g., one month) before effectively committing on it. This technique allows customers to accurately assess the level of service and detect any adverse impact stemming from this new agreement (LODHI et al., 2015). In the second example, consider an eyeball AS facing congestion on one of its upstream links. To resolve this situation quickly, finding another ISP offering connectivity to the congested destination and establishing a short-term interconnection agreement would be critical. This operation brings benefits to both customer and providers. Customers can save money if its (congested) transit provider charges them at the 95th-percentile, in which a sudden increase in traffic may drastically increase their costs. Instead, the customer operators could establish short-term interconnection agreements, which may be cheaper than paying for the extra capacity to accommodate the traffic surge at the 95th-percentile. Finally, providers can increase their revenues by serving more customers.

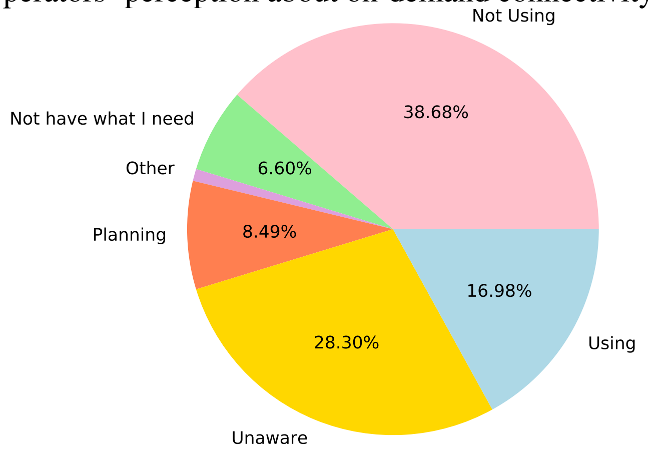
Ordering network services on-demand. Distributed Denial of Service (DDoS) is one of the most frequent attacks against infrastructures and services on the Internet. Recent examples are attacks of 1.3 Tbps and 1.7 Tbps against service providers (MORALES, 2018; NEWMAN, 2018). DDoS can be devastating, especially for networks that do not own the appropriate infrastructure to absorb or withstand the increased traffic volumes seamlessly. Ordering services on-demand was considered essential (scores 4 and 5) by 42% of the respondents. Within this group, roughly 93% of the respondents said that it takes in the order of days or weeks to set up an agreement, thus hindering the operators' ability to mitigate the effects of such attacks quickly. In contrast, those operators would need to establish proper levels of connectivity with anti-DDoS companies or scrubbing centers that peer at the facility quickly. Operators can also order direct access to cloud infrastructures, and to network analytics solutions.

4.3.2 On-demand connectivity solutions

On-demand connectivity appears as one of the most prominent evolutions of the interconnection ecosystem, especially due to the lengthy nature of the current process. As a consequence, we have been observing emerging companies such as MegaPort, PacketFabric, Epsilon Infiny, and ConsoleConnect offering such services. To understand the operators' perception about such services we asked them the following questions: (i) *what is your perception of the emerging companies offering on-demand connectivity?*, (ii) *do* *you envision that the existence of on-demand connectivity alternatives might cause a negative impact on the Internet or in the way that networks do business?*, and (iii) *would you mind if information about your interconnection agreements would be disclosed to other ASes in a solution for on-demand connectivity agreements?*.

What is your perception of the emerging companies offering on-demand connectivity? To answer this question, we offered a set of alternatives that include *I never heard of them*, *I know what they offer, but I do not need them*, *I know what they offer but they do not have what I need*, *I am planning to use them in my organization*, and *I am using them in my organization*. Operators could also specify *other* impression. Figure 4.7 shows the answers.

Figure 4.7: Operators' perception about on-demand connectivity companinies.



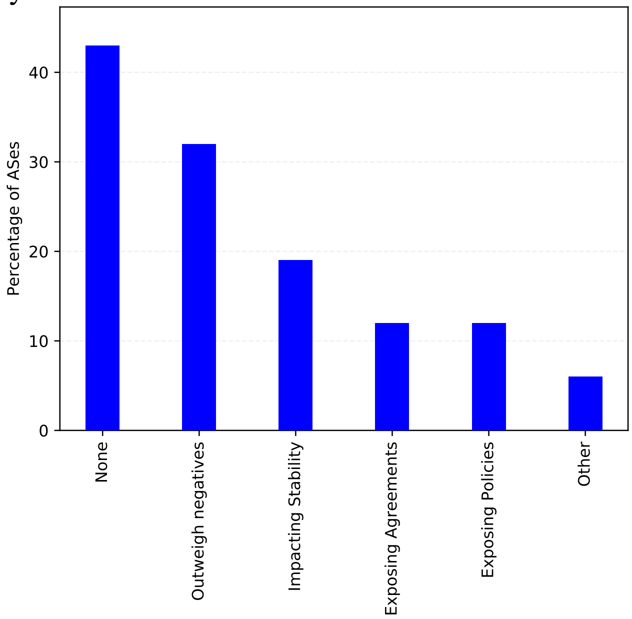
Source: the authors.

While on-demand connectivity companies are relatively new in the interconnection ecosystem, network operators are already aware of their existence, with only 28.3% of the respondents reporting not being aware of such companies. However, while the operators' community is aware of these companies, their utilization is not widespread yet, with only 16.9% using them and 8.4% planning to use. A significant fraction (45.3%) knows about such services but is not using them because they do not need (38.7%) or because the companies do not offer the service required by the ASes (6.6%). Interestingly, 0.9% of the respondents selected *other* and indicated that they are one of the on-demand connectivity companies. Analyzing the answers by region, we identify that most ASes that reported being using such services are located in Asia-Pacific, Europe, and North America. None of the respondents from Africa and Latin America and Caribbean are utilizing on-demand connectivity companies. These results are aligned with the location of the Points of Presence of the connectivity companies.

As a complementary-related question, the survey respondents could also relate why they are not using/interested in using the current on-demand connectivity services. We divided the reported reasons into three groups: pricing, infrastructure, and interconnection process. Two respondents mentioned that prices are too high and that these companies do not have a transparent pricing model. Regarding infrastructure, one operator said that the reason not to use those services is that they are not present in its region, while others said these companies, in general, offer remote peering solutions, which is something that operators prefer to avoid (NOMIKOS et al., 2018). Finally, with regards to the interconnection process, two operators reported that they prefer to continue interconnecting without depending on a man-in-the-middle entity that will increase the complexity of the process.

Do you envision that the existence of on-demand connectivity alternatives might cause a negative impact on the Internet or in the way that networks do business? To learn possible concerns of the operators' community about the utilization of on-demand connectivity services, we asked the survey respondents if they have any concerns about potential impacts on the Internet. Valid alternatives include *none*, *I expect the positives to outweigh any negatives*, *Impacting Internet routing stability*, *Exposing the existence of agreements between networks*, and *Exposing network business policy*. As in previous questions, operators could also specify *other* aspects. Each operator could select one or more alternatives. Figure 4.8 shows the answers distribution.

Figure 4.8: Operators' perception about possible impacts on the Internet due to on-demand connectivity.

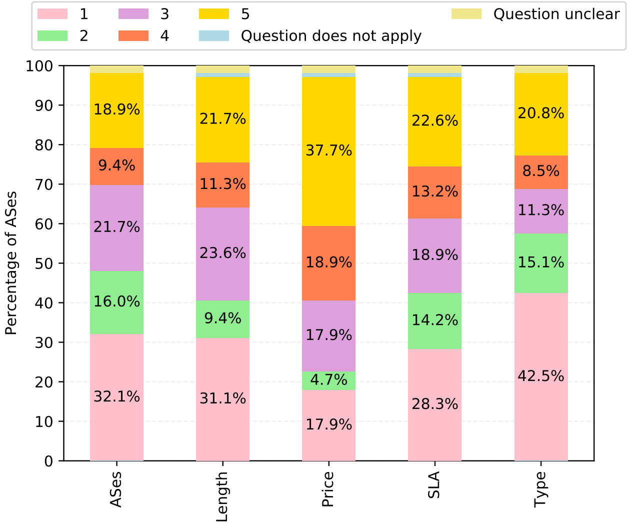


Source: the authors.

We see that 43.3% of the network operators do not expect any negative impact on the Internet. A lower fraction, 32.1%, has indicated they acknowledge that on-demand connectivity might somehow impact the Internet but expect the benefits to outweigh possible issues. The main concern (19.8%) for network operators relates to impact Internet stability since the ability to interconnect in short time frames might result in disruptions if agreements are too short. Additionally, operators are equally concerned (12.2%) about exposing the existence of interconnection agreements and exposing details about peering policies. Such an issue relates to the fact that companies for on-demand connectivity are acting as intermediaries in the interconnection process and might end-up having access to sensitive information. Finally, 6.6% of the respondents declared to have other concerns.

Would you mind if information about your interconnection agreements would be disclosed to other networks in a solution for on-demand connectivity agreements? Given that ASes might not be willing to share information about their interconnection agreements, we asked them to indicate their degree of concern in sharing each piece of information. To that end, we divided the attributes in five different sets, namely *agreement type (e.g., peering, transit)*, *agreement length (e.g., 3 months)*, *pricing (e.g., $1 per Mbps)*, *SLA (e.g., latency, bandwidth)*, *ASes involved (e.g., A has an agreement with B)*, and asked the operators to give a score from 1 to 5, where 1 indicates that information is not much sensitive and 5 represents data that is very sensitive. Figure 4.9 presents the results.

Figure 4.9: Operators' perception in sharing information about their interconnection agreements.



Source: the authors.

The most sensitive information is pricing, which 56% of the survey respondents indicated a score of 4 or 5, followed by length (33%) and SLA (35.8%). The identity of the ASes involved and information about the agreement model are less sensitive according to the network operators. We expected that because while the former three are strongly related to the AS' interconnection policy and are hard or impossible to infer, the latter two are less revealing and easier to infer by analyzing public routing information. Analyzing the results by region, we identify that ASes in North America are more open to sharing such pieces of information when compared to ASes in other regions. We observe the same condition when comparing ISPs to Content Providers, Infrastructure, and Enterprise ASes.

5 AN UNDERUTILIZED ECOSYSTEM

Enabling the set of improvements discussed in Chapter 4 requires two main conditions. The first is that there should exist physical connectivity between ASes. The second one is that there should be spare capacity available at the link between the ASes. Peering infrastructures such as IXPs and colocation facilities satisfy the first condition since they provide physical connectivity among hundreds or even thousands of AS members (AMS-IX, 2018; DE-CIX, 2018; IX.BR, 2018b). Concerning the underutilized link, previous studies have hinted the lack of congestion (FANOU; VALERA; DHAMDHERE, 2017) and existence of spare capacity in IXP ports (CHIESA et al., 2016) and interconnection links (FEAMSTER, 2016). However, they analyzed small or heavily aggregated datasets. To confirm the availability of spare capacity, we analyze the utilization of peering ports from two commercial IXPs.

5.1 Methodology

To understand the availability of capacity at IXP peering ports, we collected datasets consisting of traffic traces from two IXPs and used them to answer the following questions: (i) *how much spare capacity do IXP ports typically have?* (ii) *for how long (continuously) is the capacity available?* (iii) *how does the port capacity utilization differ by hour of the day?* (iv) *how does the availability of spare capacity have changed over the years?* and (v) *for how long (continuously) is a given fraction of the port capacity unavailable?*

Datasets. We collected traces from two IXPs: IXP-EU, one of the largest IXPs worldwide located in Europe, and IXP-LA, a medium-sized IXP situated in Latin America, both transporting high volumes of traffic per second among hundreds of members (over 5 Tbps and 100 Gbps, respectively). We are not authorized to disclose the identities of these IXPs. While insufficient to allow generalizations, they are enough to provide useful insights. We note that obtaining access to IXP datasets, carrying commercial traffic, is challenging. In IXP-LA, we captured flow summaries between mid-October 2015 and mid-October 2016, with a sample rate of 1:32768 packets. In IXP-EU, we collected flow summaries during a total of 9 week-long periods in 2016 and 2017 sampled with a rate of 1:10000 packets. In both IXPs the measurements are aggregated in 5-minute intervals.

Assumptions. To reason about the port capacity availability, we make the following assumptions. First, for each IXP port, we assumed its capacity as the highest observed *peak* (ingress or egress, 5-min average) utilization. We use this premise because the information about specific provisioned capacity is too sensitive and not part of our dataset. However, we note that a previous study (CHIESA et al., 2016) has hinted that for 63% of the ports of a large IXP their maximum utilization during seven days did not exceed 50% of their actual capacities. Thus, our assumption is conservative. Second, we consider a port as "active" in the period between its first and last observed non-nil 5-min measurement. We make this assumption because not all ports are active during the measurement period. While the granularity of our dataset does not capture micro-bursts of traffic, we observe that *i)* the respondents to our survey also acknowledge low port utilization and *ii)* Internet traffic tends to exhibit a low level of (micro) burstiness (FRALEIGH; TOBAGI; DIOT, 2003), much smaller than, for instance, those observed in data center networks.

Metrics. We use the following metrics in our analyses:

- **Spare capacity ratio** (s)**.** Let R be a sequence $\{R_0, R_1, \ldots, R_n\}$ of 5-minute long measurements of the traffic forwarded through a peering port during a specific time interval. The port spare capacity $s(\alpha)$ in $[0, 1]$ represents the minimum fraction of the estimated port capacity c that is available for a fraction α of the time, i.e., there exists a fraction α of the measurements in R where $1 - \frac{R_i}{c}$ is at least $s(\alpha)$. To illustrate, $s(0.5) = 0.8$ means that at least 80% of the port capacity is available during 50% of the time.

- **Port capacity utilization ratio** (u)**.** Let R be a sequence $\{R_0, R_1, \ldots, R_n\}$ of 5-minute long measurements of the traffic forwarded through a peering port during a specific time interval and c the estimated port capacity. We define the port capacity utilization u in $[0, 1]$ for a given instant i from R as $\frac{R_i}{c}$.

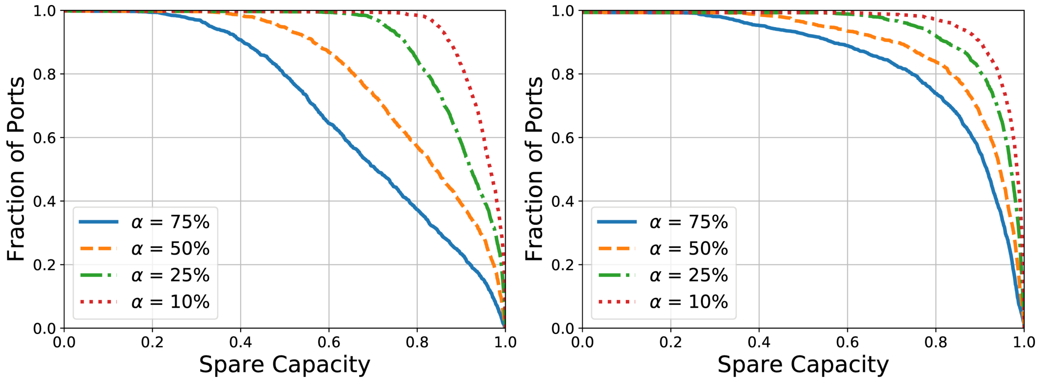
- **Event magnitude** (m)**.** Let R be a sequence $\{R_0, R_1, \ldots, R_n\}$ of consecutive 5-minute long measurements of the traffic forwarded through a peering port during a specific time interval that has exceeded a given capacity threshold t, let p be the highest value of R, and c the estimated port capacity. We define the magnitude m in $[0, 1]$ of a given event e as $\frac{p}{c} - t$. To illustrate, if the port capacity c is $1Gbps$, the event peak p is $0.7Gbps$, and the threshold t is 0.5 (50% of the port capacity), the event magnitude m will be 0.2, which means that in this event exceeded the threshold in 20% of the port capacity.

- **Event length** (l)**.** Let R be a sequence $\{R_0, R_1, \ldots, R_n\}$ of consecutive 5-minute long measurements of the traffic forwarded through a peering port during a specific time interval that is above or below a given capacity threshold t. We define the length l of a given event e as the size of $R * 5$, since each measurement sample represents a 5-minute long unit.

5.2 How much spare capacity do IXP ports have?

To analyze the availability of spare capacity, we use the IXP-LA dataset. Figure 5.1 shows the spare capacity for different values of α for the ports present at IXP-LA. As the actual port capacities may have changed during our measurement campaign, we divide the dataset into twelve windows each of one-month length to avoid over-estimating the ports capacities. Such a decision leads to up to twelve (see assumptions) different port capacity estimations instead of a single one spanning the entire year. As the ports are full-duplex, we present the ingress and the egress utilization ratio for each port. We consider the AS as the reference point for the direction of the traffic.

Figure 5.1: Monthly spare capacity for IXP-LA.



(a) Ingress ports

(b) Egress ports

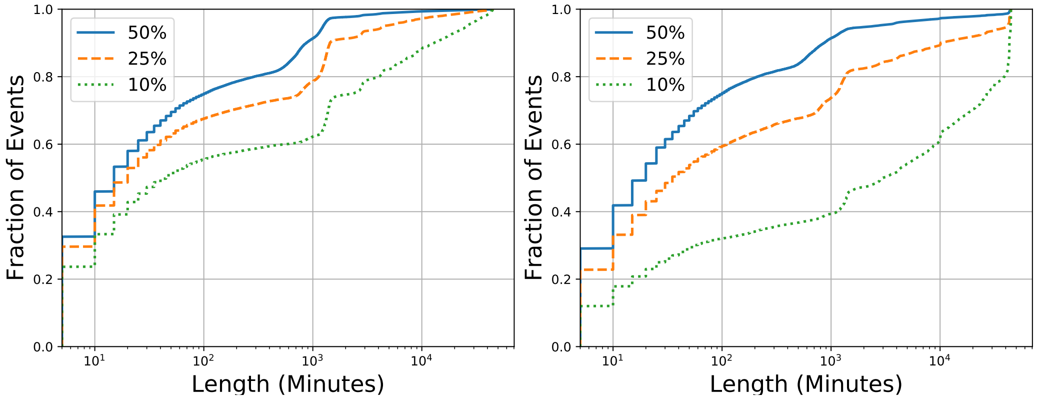
Source: the authors.

IXP-LA presents a significant amount of spare capacity in both directions. For example, 60% of the ingress ports (Fig. 5.1(a)) have at least 78% of spare capacity during 50% of the time. As for the egress direction (Fig. 5.1(b)), 60% of the ports have at least 92% of their capacities available during 50% of the time. The spare capacity is higher in the egress direction because the majority of ASes connected to IXP-LA are access networks. These results are the first indication that the conditions to deploy Dynam-IX are favorable.

5.3 For how long (continuously) is the capacity available?

While the availability of substantial amounts of spare capacity hints the possibility of leveraging this capacity for new interconnection agreements to improve wide-area traffic delivery, we need to ascertain that the capacity is available for a continuous amount of time. If the availability of spare capacity persists for small periods (e.g., 10 minutes), ASes cannot effectively use it for new agreements without impacting other traffic using the same port. We measure the length of the events where at least a fraction α of the capacity is available and consider as actual candidates for new interconnection agreements events that lasted for at least one hour.[1] Figure 5.2 shows the results for all ports (egress and ingress) of IXP-LA considering three different values of α.

Figure 5.2: Length of all periods with spare capacity for IXP-LA.



(a) Ingress ports

(b) Egress ports

Source: the authors.

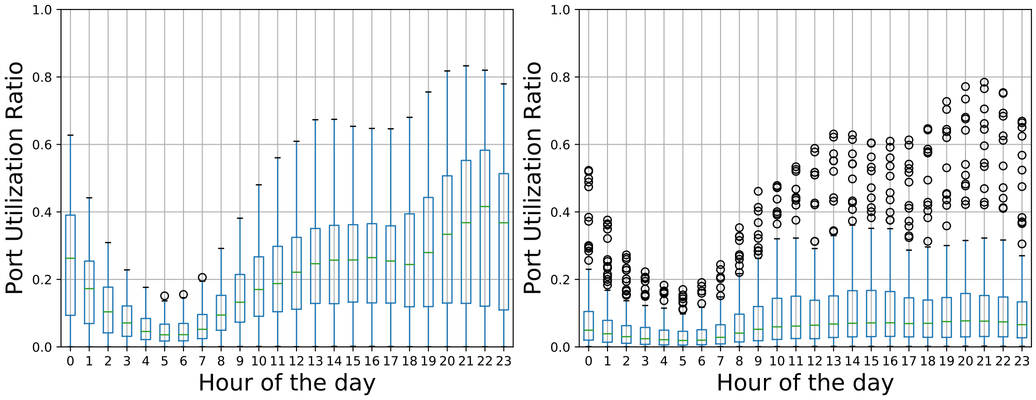
We have identified 115,117 events where at least 50% of the ingress capacity was available (Figure 5.2(a)). 29.3% (33,822) of them lasted for at least one hour, 22% (25,308) for three hours or more, and 19.3% (22,229) for at least six hours. While the number of events that are actual candidates for new interconnection agreements is approximately one-third of the cases, they correspond to 97.62% of the spare time of the peering ports. For the egress ports, we have spotted 43782 events where at least 50% of the spare capacity was available. Among these, 30.24% (13,240) lasted for more than one hour, 20.97% (9,184) for at least 3 hours, and 17.89% (7,834) for more than 6 hours. For the egress ports, the events that last for at least one hour correspond to 99.11% of the time that ports had at least 50% of available capacity. These results allow us to confirm the practicality of using the spare capacity for new interconnection agreements.

[1] In our interviews with network operators they mentioned that while shorter agreements would be feasible, in their opinion, one hour would be the minimum required length for a practical agreement.

5.4 How does the port capacity utilization differ by the hour of the day?

Our third analysis seeks to understand the usage of the port capacity for each hour of the day. The rationale for this analysis is that having available capacity in non-peak hours would not be so useful. Thus, using the IXP-LA dataset, we measured the median port utilization for each hour of the day for each port in both directions. Figure 5.3 shows the median port utilization both for the ingress and egress traffic for each hour of the day.

Figure 5.3: Median port utilization for each hour of the day for all ports of IXP-LA.



(a) Ingress ports

(b) Egress ports

Source: the authors.

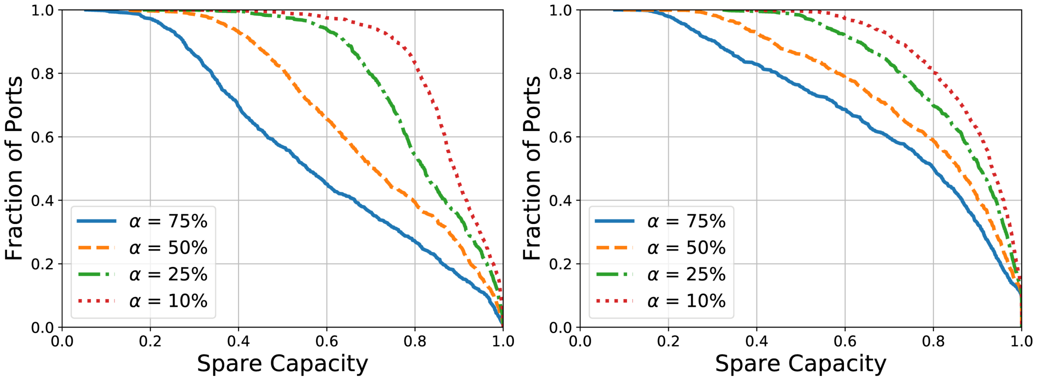
We can observe that even during the peak hours, more than 50% of cases have a median port utilization below 50% of the estimated port capacity in both egress and ingress directions. We also note that a few ports present higher median utilization ratios (e.g., > 0.7) during the peak hours, which can indicate that these ports might not be suitable for the establishment of new interconnection agreements. While we acknowledge this, we also remember that our estimate of the port capacities tends to be very conservative (see §5.1). Effectively confirming the practicality of using these ports for new agreements would require knowledge about the actual port capacities, which is a piece of sensitive information that is not part of our datasets. Based on this evaluation, we can conclude that most ports have capacity available even during peak hours.

5.5 How does the availability of spare capacity change over the years?

To understand if the identified spare capacity is consistently available, or even growing, we analyze the IXP-EU port utilization over time. We note a consistent pattern of available capacity at both ingress and egress ports for all seven weeks of our dataset.

This behavior has persisted even with an increase in the number of members and the traffic volume they generated during the time. Furthermore, the difference between the snapshots with the largest and the smallest spare capacity availability is negligible (less than 1% difference). A possible reason for that consistency over time relates to the fact that ASes tend to upgrade the capacity of their interconnection links when the traffic reaches a given utilization threshold. To illustrate the spare capacity ratio at IXP-EU port, we zoom in a single week of 2017 (Fig. 5.4).

Figure 5.4: Weekly spare capacity for IXP-EU.



(a) Ingress ports

(b) Egress ports

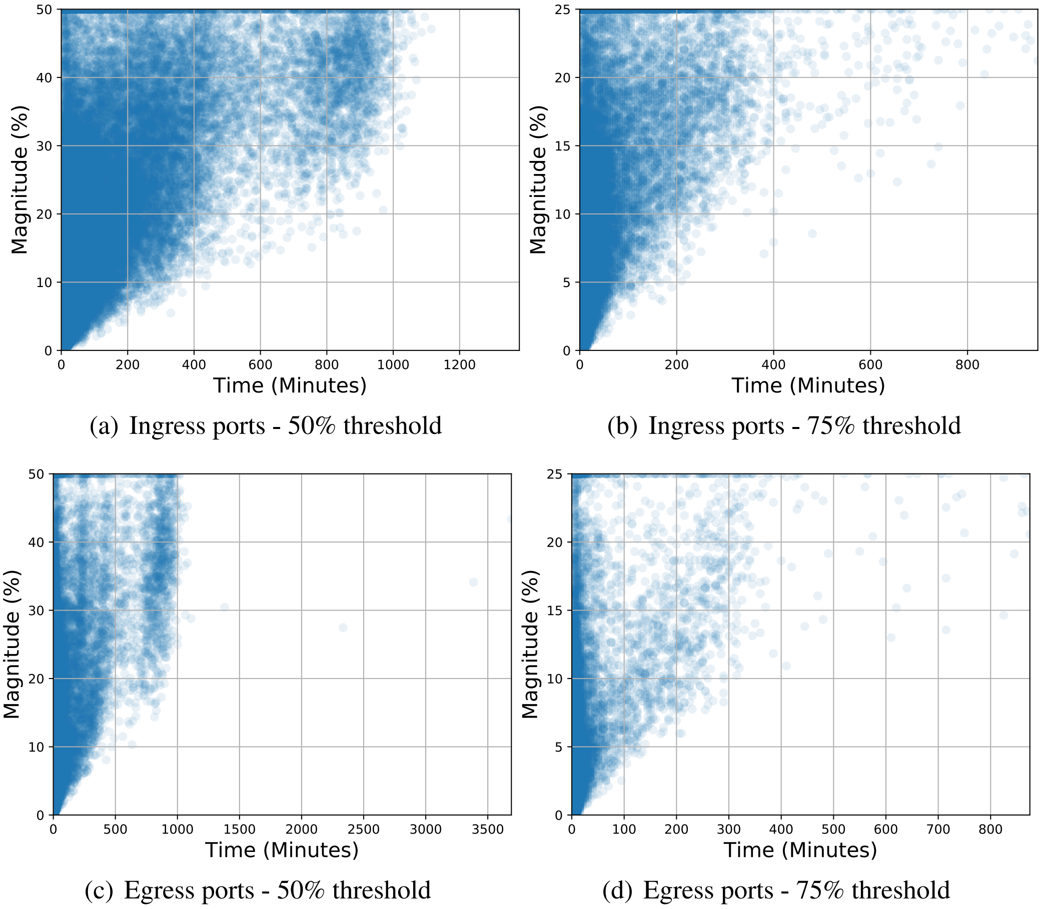
Source: the authors.

We observe that during 50% of the time about 60% of the ingress ports have (Fig. 5.4(a)) at least 63% of spare capacity. For egress ports, in turn, 60% show 80% of available resources (Fig. 5.4(b)). The existence of such a condition in a second IXP indicates that our premises could be valid for other peering infrastructures. When comparing the availability of spare capacity at the ports of IXP-EU to the ones from IXP-LA, we observe that the ports from the Latin American IXP have more spare capacity available. One possible reason for that could be the fact that while at IXP-EU ASes need to pay a fee based on the capacity of their interconnection ports, members of IXP-LA were not required to do so during the period we captured the traces. As a result, some ASes might order ports with higher capacities than their actual bandwidth requirements. Another reason can be related to how much ASes from Latin America rely on IXPs to exchange traffic.

5.6 For how long (continuously) is a given fraction of the port capacity unavailable?

Lastly, we analyze the magnitude and the length of events that exceed a given port capacity utilization. Similarly to the previous analysis that show the length of the cases where there is spare capacity available, the current analysis seeks to learn the properties of the events when the capacity is not available. Figure 5.5 shows the magnitude and the length of the events that exceeded 50% and 75% of the port capacity utilization for all ports of IXP-LA.

Figure 5.5: Magnitude and length of events exceeding a given threshold for all ports of IXP-LA.



(a) Ingress ports - 50% threshold

(b) Ingress ports - 75% threshold

(c) Egress ports - 50% threshold

(d) Egress ports - 75% threshold

Source: the authors.

We observe 114,266 events exceeded 50% of the ingress port capacity utilization. 50% of them lasted for up to 10 minutes, and their median magnitude was 4.9% of the port capacity (meaning that up to 45.1% of the capacity was still available). For the egress ports, 42,406 events require more than 50% of the port capacity. Their median length is five minutes, and their median magnitude is 5.41% (port capacity utilization of 55.41%). Analyzing the events that exceed 75% of the port capacity, we identify 43,251 cases for the ingress ports and 11666 for the egress ports. Their median lengths are 5 and 10 minutes for egress and ingress ports, respectively, while their median magnitudes were 3.75% (egress) and 4.09% (ingress). It is possible to infer that most events are short-lived and of small magnitude, which indicates that while the regular traffic of the peering port might get impacted by the new agreement, the impacts will probably not last long.

5.7 Discussion

We have identified that peering ports have substantial amounts of spare capacity, that it is constantly available, even during peak hours. We also learned that most of the availability of spare capacity persists for periods of at least one hour and that most unavailability events are short-lived and of small magnitude. These analyses hint the practicality of designing a solution to allow ASes to improve wide-area traffic delivery performance by exploiting the rich connectivity of peering infrastructures. We also acknowledge that the reasons for the existence of spare capacity may vary, such as (i) AS inability to produce/attract traffic to use the available capacity; (ii) peering infrastructures offering ports with more capacity than the current needs of ASes; and (iii) to accommodate traffic micro-bursts and the natural traffic growth. Cases (i) and (ii) represent scenarios where the AS can leverage the spare capacity without affecting the rest of its traffic. Case (iii) would require network operators to make planned decisions to avoid negatively impacting traffic delivery.

6 DYNAM-IX- A DYNAMIC INTERCONNECTION EXCHANGE

In the previous chapters, we have shown that network operators network operators consider important to evolve the interconnection process, and that peering infrastructures offer a vast unexplored potential to improve wide-area traffic delivery. In this chapter, we present the design of Dynam-IX, a Dynamic Interconnection eXchange. We start discussing the requirements that we should satisfy for a practical solution. Then, we justify our design choices and detail the Dynam-IX architecture. Finally, we end this chapter discussing practical aspects to deploy our proposal and its possible limitations.

6.1 Requirements

Adoption is the underlying requirement for any practical approach to inter-domain routing. To facilitate adoption, we design our solution to complement the existing practices in the area, leading us to the following high-level requirements:

- **structured process**: there should exist a structured process for network operators to find and establish interconnection agreements and to express interconnection negotiation procedures;

- **expressive interface**: an operator should be able to specify its business interconnection policies, including the traditional interconnection models (e.g., transit and settlement-free peerings) as well as future ones;

- **confidentiality**: unauthorized parties should not have access to information considered private about an interconnection agreement (e.g., business policies, interconnection terms). Our survey shows that network operators are reluctant to sharing interconnection policy-related information with third parties, thus confirming the findings in (CHIESA et al., 2017a);

- **mechanism to build trust**: network operators should be able to identify partners deemed reliable (by the community) systematically. In fact, while today operators drive their peering business decisions based on personal relationships and brand recognition, we argue that these approaches must be *complemented* with a more systematic and automated technique that improves the operator's ability to engage in interconnection with the ever-growing number of members at peering infrastructures.

Besides, secondary requirements are the ability to scale with the largest peering infrastructures; interoperability with both the current network protocols, processes for establishing interconnection agreements, and operators' mindset for administering peerings; and providing benefits upon incremental deployment.

6.2 Design Choices

A straightforward approach to our goal would be to provision peering infrastructures to offer a service where ASes can query and advertise interconnection opportunities. Unfortunately, there are two main issues with such a centralized solution.

The first issue is that a centralized service, which intermediates interconnection agreements, must be trusted with confidential information (i.e., interconnection requests and offers). Given the competitive nature of the interconnection ecosystem, this scenario seems plausible only for open, settlement-free peering. An alternative to preserve confidential information could be engineered to guarantee strong security properties (e.g., using secure multi-party computations (GOLDREICH; MICALI; WIGDERSON, 1987) or trusted execution environments such as Intel SGX (COSTAN; DEVADAS, 2016)). However, this raises the complexity of the solution and incurs processing overheads. This approach also introduces a third-party service, making ASes dependent on its availability and impartiality.

Instead, we design Dynam-IX based on a distributed *protocol* that works in conjunction with a *legal framework* to preserve confidentiality while avoiding processing overheads and the need for trusted entities. In Dynam-IX, interconnection policies are expressed using a *high-level interconnection intent abstraction*. Such an abstraction provides a powerful interface that allows operators to easily query and offer interconnection opportunities while removing the need for human interaction to discuss the properties of an interconnection agreement. A high-level interconnection intent abstraction provides a natural way to express their interconnection intents (as opposed to, say, low-level routing configurations) and can be intuitive for people without a programming background.

The second issue with a centralized solution relates to the mechanism to build trust. Unlike interconnection negotiations, this mechanism may not require confidential information. However, we argue that a centralized solution is not practical because of market incentives. Peering infrastructures are disincentivized to interfere in the business decisions of ASes, which are customers of the facility. Moreover, peering infrastructures would face the burden of dealing with disputes should ASes question the information collected by the facility-operated mechanism. Our contacts in the peering infrastructure operation community confirmed these concerns, making a centralized solution offered by them (peering infrastructures) impractical. To overcome these limitations, Dynam-IX uses a distributed tamper-proof ledger to enable ASes to build trust cooperatively. The tamper-proof property is necessary to prevent a malicious AS from tampering with the information to gain a benefit or to harm another AS.

Designing a distributed approach allows ASes to keep their independence to interconnect.[1] Alas, such an alternative allows incremental deployment without depending on the willing of the peering infrastructure to offer such kind of service. Finally, a distributed approach also allows the ASes to benefit from it without needing to pay extra fees to the peering infrastructure. We detail each of the Dynam-IX components (Figure 1.2) in the following sections.

6.3 Protocol

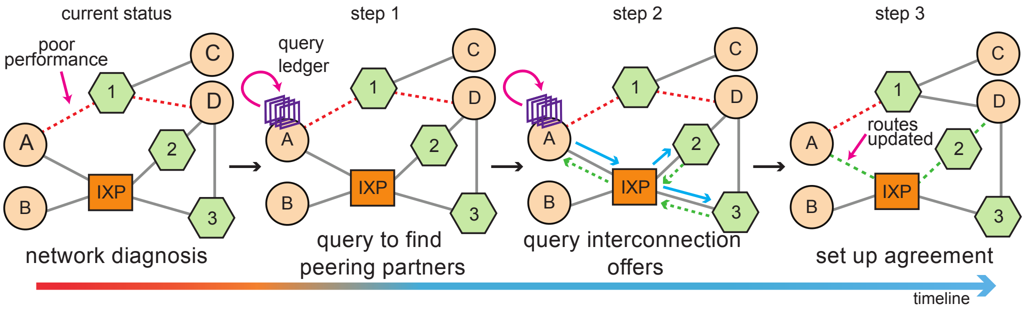
The protocol is the core component of Dynam-IX. We define it to resemble the current process for establishing interconnection agreements. The protocol allows network operators to automate the interconnection process by providing to them well-defined methods to query and offer interconnection proposals as well as to settle agreements.

To start using Dynam-IX, the network operator must first initialize a Dynam-IX peer and connect to the distributed ledger. Additionally, the AS needs to make available to the other ASes information about its IP address, the port of the Dynam-IX peer, public key, and a description of the services the AS is offering. For the sake of detailing the protocol, we assume here that ASes store such information on the ledger; however, this is not mandatory, and ASes can use other storage systems for this specific information (see §6.7). The ledger also contains information about the past performance of each Dynam-IX member as a customer and a provider of an interconnection agreement (§6.6). After connecting to the Dynam-IX ledger, the AS can start using the protocol. To illustrate

[1] Remember that some survey respondents declared that they prefer not to use on-demand connectivity for this specific reason (Chapter 4).

how the protocol works, we use the example presented in Chapter 1, where an AS A is facing a traffic surge towards C and congesting the traffic going to D. For the sake of our example, we consider that ASes store on the ledger scores reporting their experience in the agreements. In Section 6.6, we discuss the different pieces of information that ASes can store on the ledger. Figure 6.1 illustrates all the protocol steps.[2]

Figure 6.1: Dynam-IX protocol.



Source: the authors.

Identifying potential peering partners. After diagnosing the need for an interconnection agreement, AS A queries the ledger to identify providers (based on the service description field) that may offer connectivity to the intended destination (AS D). The ledger returns a list of providers and the essential information to allow the customer to contact each provider. In the example, ISPs 2 and 3 can reach the desired destination. AS A can use the score information to filter out ASes that might not be reliable for interconnecting (e.g., a low score from previous interconnection agreements).

Obtaining interconnection offers. Second, the AS A submits a request to each reliable provider for interconnection proposals to reach a target (e.g., AS D) with specific desired interconnection properties (e.g., minimum bandwidth and maximum latency SLAs). To protect the confidentiality, any communication between a customer and a provider is encrypted using standard SSL (Secure Sockets Layer) and authenticated using public keys of each AS. When a provider receives a query for an interconnection offer, it decides whether to answer or not. The decision could be made automatically by an algorithm or delegated to a human, and based on both the provider business policy and any information available on the ledger. In either case, our approach provides a structure for the negotiation process. For the sake of the explanation, assume the provider does answer. The provider matches the customer request against its interconnection intents and, if a valid match exists, it composes an interconnection offer and sends it to the customer. Other-

[2] We use the terms *customer* and *provider* as a reference to identify the protocol roles. In our scheme, an AS can be both a customer and a provider.

wise, the provider notifies the customer that no match exists. Offers are digitally signed so ASes can verify their authenticity at any time.

Establishing an agreement. Third, the AS A selects (according to its policy) one of the offers. If there were none, the protocol needs to be restarted, possibly with different desired interconnection properties to match the current interconnection conditions. Assume the AS A has chosen the offer from ISP 2. The customer sends an agreement proposal to the provider of the selected offer. The provider verifies that the proposal corresponds to a valid offer (each offer has an expiry date). Given a legitimate offer, the provider creates a legal contract (see §6.5), digitally signs it, and sends it to the customer. The customer verifies the provider signature and contract terms. If the signature is valid and the contract terms are as expected, the customer digitally signs the contract and sends it to the provider. In turn, the provider verifies the customer signature and, assuming it is valid, proceeds to register the interconnection agreement on the ledger (see §6.6). Once AS stores the information in the ledger, both ASes update their BGP configurations and start exchanging traffic.

Ending an agreement. When an interconnection agreement ends, besides tearing down the BGP configuration, both ASes store on the ledger a piece of information reflecting their experience (§6.6) during the agreement. ASes use this information to help whether or not a network is reliable to interconnect.

6.4 Interconnection Intent Abstraction

Network operators and peering coordinators need a simple and easy-to-use abstraction. We define an *interconnection intent abstraction* as the relevant technical and business information associated with an interconnection offer. To design the intention abstraction on practical grounds, we rely on the interviews with peering coordinators and network operators, as well in our survey. We note that all parameters mentioned by three or more operators (among 100+) are present in the abstraction. We also observe that the abstraction is easily extendable, thus more parameters we can add in the future in response to the specific needs of the operators.

The interconnection intent abstraction. Each intent consists of a *target*, i.e., the traffic destination considered within the intent, and a set of attributes that describes information about the interconnection offer.

Table 6.1: Summary of Intent Abstraction Attributes.

Category	**Attribute**	**Description**
`Routing`	as_path	List of ASes on the path
SLA	bwidth	Available bandwidth (Mbps)
	latency	Expected latency (milliseconds)
	pkt_loss	Expected loss (percentage)
	jitter	Expected jitter (milliseconds)
	repair	Expected repair time (minutes)
	guarantee	SLA guarantee (% of time)
	availability	Link availability (% of time)
`Pricing`	billing	Billing method
	ingress	Per-unit price function
	egress	Per-unit price function
`Time`	length	Agreement length (hours)

Source: the authors.

```
target: {
routing: { attributes }
sla: { attributes }
pricing: { attributes }
time: { attributes }
}
```

The target of the intent is used to identify the type of traffic for which the intent holds. Valid targets are IP prefixes (e.g., 8.8.0.0/16), which can be used to negotiate connectivity towards a specific IP prefix destination, 0.0.0.0/0, which can be used to acquire transit connectivity, and ASNs (e.g., ASN12345), which can be used for peering agreements or to reach all prefixes of an AS. Table 6.1 presents a summary of the intent abstraction attributes. These are divided into four categories: `routing`, Service Level Agreement (`SLA`), `pricing`, and `time`.

The *routing* category contains one or more *as_path(s)* that will be used to reach the prefix (target) of the intent. The *Service Level Agreement* category attributes describe the expected performance and availability properties of the intent. The *pricing* group specifies the *billing method*, which models the traditional flat-rate or 95th percentile, but also *per-unit price functions* related to the ingress and egress traffic. Per-unit price functions allow network operators to support on-demand connectivity where a network pays for the egress traffic and profits from the ingress traffic. Several ASes connected to Hopus (HOPUS, 2018b) already use this billing method. Finally, the *time* metric specifies the time granularity (in hours) of the interconnection agreement. A period of one hour means that the duration of an interconnection agreement for that intent must be an integer multiple of one hour.

Specifying prices as functions of time or bandwidth. Several connectivity providers (e.g., (HOPUS, 2018a)) specify their (ingress/egress) per-unit costs as a function of the bandwidth and the length of the interconnection agreements (HOPUS, 2018b). Operators often want to incentivize their users to commit for longer periods and higher bandwidth by offering lower per-unit prices. As a simple example, an operator could specify the per-unit price as follows, which decreases as the bandwidth and time commitments increase.

```
pricing: {
"ingress": e^(1/(sla.bwidth*time.length))-1
}
```

Sharing properties among intents. As a straightforward optimization, ASes can group multiple intents that share common properties using intent profiles, which serve as a template for actual intents. Profiles are identified by *prof-id*, where *id* in a unique identifier for the profile. A profile can be associated with a target as follows.

```
target: {
profile: prof-id
}
```

Intents without strict guarantees. Network operators have reported in the survey and interviews that some parameters, such as the ones defined in the SLA category, may not be taken into consideration when establishing certain interconnection agreements (e.g., in settlement-free peering). In these cases, ASes can use the *wildcard* character '*' as the attribute value.

Querying ASes for an interconnection agreement. The intent abstraction defines a function called *query*, allowing an AS to retrieve interconnection agreement proposals.

```
query(ASN, target, [properties])
```

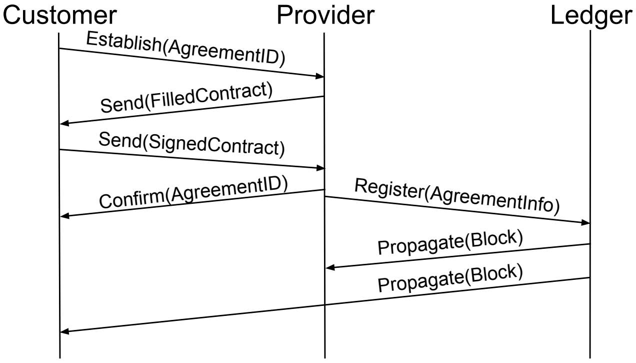
The *ASN* specifies the AS to which the customer will send the query. The *target* parameter defines the traffic of interest for the issuer of the query, i.e., IP prefix destinations. Finally, the properties of the query map to the attributes of the intents. *Properties* specify the requested conditions of the interconnection agreement. These are specified as a conditional expression over the attributes (e.g., `sla.latency == 15 && sla.bwidth > 1000`). When performing a query, the only mandatory property is the expected length of the agreement (time attribute). Provider ASes do not consider unspecified properties during the query operation. A single query (to a single AS) can provide interconnection offers for multiple targets, but, if so, all targets will share the same interconnection properties.

6.5 Lightweight Legal Framework

Discussions among legal offices and lawyers can represent a crucial phase before ASes establish an agreement. Contractual terms and conditions should be carefully stated to legally protect parties in possible future disputes, which are not uncommon in the Internet ecosystem (BAFNA; PANDEY; VERMA, 2014). Our survey findings reveal that legal matters require hours (19%), days (37%), weeks (30%), months (10%), or even years (4%).

Although operators are more open to handshake agreements (WOODCOCK; FRIGINO, 2016), providing legal protection to the agreements (especially the ones involving monetary compensation) would spur adoption, but doing so with lengthy legal procedures would severely hinder the efficiency of Dynam-IX. We overcome this problem by adopting a *Lightweight Legal Framework* (LLF). It can protect networks when signing contracts without incurring lengthy delays to set up an interconnection agreement. LLF involves defining one (or more) *general contract template(s)* that is (are) stored on the ledger, and digitally signed by every AS that joins an instantiation of Dynam-IX. A contract template contains standard clauses related to interconnection agreement and empty fields to be completed with the specific properties when ASes establish an interconnection agreement. Figure 6.2 shows how the Dynam-IX legal framework works.

Figure 6.2: LLF steps.



Source: the authors.

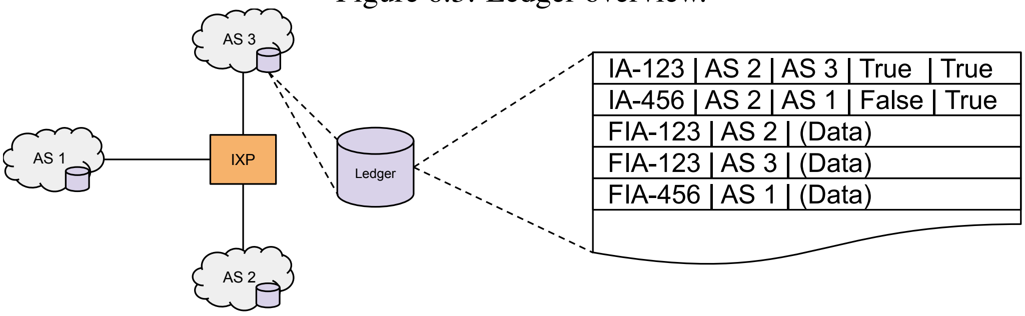
When a customer and a provider AS are establishing an agreement, the provider fills the general template with the specific properties of the interconnection agreement and submits it to the customer, along with the digitally signed hash of contract. Then, the customer receives the contract and checks both its properties and the provider signature. At this point, the customer may confirm that it agrees with the terms by sending the provider a digitally signed hash of the contract. When confirming an agreement, the customer also stores a local copy of the signed contract. The provider will check the customer signature and then store a local copy of the contract. By storing a local copy of the signed contract, both ASes can handle future disputes related to the agreement. Once the agreement is confirmed, the provider registers the agreement on the ledger, which will eventually create a new block and propagates it to all Dynam-IX peers.

The general template can be reviewed and updated by the members of Dynam-IX at any time. In such a case, the new contract must be published on the ledger and digitally signed by the members. Thus, LLF requires lawyers only when an AS joins Dynam-IX or when the template is updated.

6.6 Tamper-proof Distributed Ledger

The protocol, the legal framework, and the interconnection intent abstraction provide answers for three of our research questions. The last aspect that we need to address is to empower ASes with a mechanism to allow the identification of reliable peering partners. To that end, Dynam-IX uses a tamper-proof distributed ledger. The ledger goal is to act as a trusted place where ASes can store and retrieve information about previous interconnection agreements. Figure 6.3 illustrates an overview of the ledger.

Figure 6.3: Ledger overview.



IA-123	AS 2	AS 3	True	True
IA-456	AS 2	AS 1	False	True
FIA-123	AS 2	(Data)		
FIA-123	AS 3	(Data)		
FIA-456	AS 1	(Data)		

Source: the authors.

Once two ASes establish an agreement, a mutually digitally signed piece of information is stored in the distributed ledger to indicate that they have settled an agreement. This piece of information is necessary to associate feedback information with valid interconnection agreements only. Each record is a 5-tuple containing: a unique identifier of the agreement; the ASNs of the two connected networks; and two attributes to control the update of the feedback of corresponding networks. We note that most survey respondents indicated that information about the ASes involved in the agreement is not sensitive (Chapter 4). When the agreement ends (or periodically), both ASes store on the ledger information about their experience during the interconnection agreement. The customer feedback indicates the "quality" of the service offered by the provider AS during the agreement. While the provider feedback indicates if the customer AS is a "good player". Other ASes can then use this information before interconnecting to decide whether or not a given AS is reliable for exchanging traffic.

Initially, ASes do not have any information associated with them. Not having information does not mean an AS cannot be trusted. We envision at least two different possibilities for the type of information that network operators can share: scores and reports. We discuss them below.

Feedback Scores. One alternative would be to use scores to represent the ASes perceptions about the interconnection agreement. In such a case, when an interconnection agreement ends or periodically, the provider invokes a method to store the agreement score on the ledger and update the customer score. A similar process is executed to update the provider's score. After verifying (based on the contract terms) that the AS provided (or not) the adequate service, the customer invokes the procedure to store its agreement score and update the provider's score. ASes can rate each other following any previously agreed algorithm such as the one presented by Alowayed et al (ALOWAYED et al., 2018). Dynam-IX will then provide two approaches for ASes to use the scores: (i) relying on the aggregated per-AS scores automatically computed and stored on the ledger or (ii) locally aggregating the individual per-agreements scores.

- **Ledger trust score computation.** When invoking the procedure, the AS needs to provide the ID of the interconnection agreement (encrypted with its private key) and whether the provider score should be updated either positively or negatively. The procedure will then verify if the AS was part of the interconnection agreement and that the AS has not sent more than one score per scoring period. If such conditions are valid, the score is updated. To avoid benefiting the AS that submits its score second, we use an approach based on the coin flipping problem (BLUM, 1983), which allows two parties to commit to their values before revealing them, thus ensuring fairness. Each party encrypts the score using its private key. Then, each network generates a random nonce n that is used to create a unique hash. Next, the participants hash their nonce and encrypted score and add it to the ledger. Nonces are used to avoid leaking information, which would give an advantage to the participant going second. Once the two ASes publish their scores on the ledger, they can publish the decrypted scores and reveal the nonces. The overall score is updated only if the decrypted and encrypted score match.

- **Local trust score computation.** Based on some personal information, an AS may not trust all the scores stored on the ledger. In this case, an AS locally computes per-ASes scores based on both the individual per-agreement scores stored on the ledger and its level of trust concerning these scores. This approach is allowed in Dynam-IX but requires more effort on the AS side to specify its trust policies.

Feedback Reports. Alternatively, instead of using non-objective scores that might not work for every AS since different network operators may have different perceptions about what is a “good” or a “bad” network, ASes can store any report about the interconnection agreement. For example, the customer AS can store on the ledger a report containing traffic measurements of the interconnection agreement and the agreed SLA. Similarly, the provider can store, for instance, a report showing that the customer effectively has paid for the interconnection. Other ASes can then check the reports to build their perceptions about the other ASes in a more objective way.

However, using feedback reports may result in exposing sensitive information. While survey respondents indicated that sharing information about the ASes involved in the agreement and the type of interconnection is not a significant concern, providing details about pricing and SLA aspects can be a problem. While Dynam-IX purpose is to guarantee the properties of data (e.g., tamper-proof) independently of the type of information stored by the ASes, we note that ASes could overcome these limitations using approaches such as the one offered in Zkledger (NARULA; VASQUEZ; VIRZA, 2018), that allows the storage of sensitive information without revealing it.[3]

6.7 Limitations and Practical Considerations

Dynam-IX design provides answers to all questions presented in Chapter 1, thus

[3] We have contacted the authors of Zkledger asking for their source-code in order to incorporate it on Dynam-IX, but they never replied to our message.

enabling ASes with an approach to benefit from the vast unexplored potential of peering infrastructures to improve wide-area traffic delivery. Here we discuss existing limitations in Dynam-IX and the main practical aspects for its adoption.

6.7.1 Limitations

Single round negotiation. Compared to the current process of establishing interconnection agreements, the present specification of Dynam-IX does not offer a method for operators to negotiate prices. There is an inherent trade-off between the desire to being responsive to traffic changes and negotiating terms, and Dynam-IX is biased towards enabling the former. A primary measure is for a customer to send new queries with lower requirements until the desired target price strikes. Remember that Dynam-IX goal is to complement the existing practices. Thus, Dynam-IX could help to quickly identify reliable partners in an automated manner through short-term agreements. Humans could then conduct any further price negotiation.

Intents to define how and when to interconnect. While Dynam-IX intent abstraction allows network operators to specify their interconnection intents, its current design misses some features. For instance, there is no method yet for ASes to specify *when* it needs to establish a new interconnection agreement nor *how* to reply for interconnection queries. Adding such features is in our roadmap towards a more dynamic interconnection ecosystem.

Single legal template. In the current design, all interconnection agreements must follow the same terms. While ideal, this scenario might prevent ASes to use Dynam-IX as they cannot define the terms of their agreements. Although having general conditions for the users of a given service is a common practice, we can extend the legal framework to allow ASes to create their contract templates and store them on the ledger. In such case, the other ASes can proactively agree to the terms of the template by digitally signing it or doing this on the first agreement (which may require a lawyer to check the clauses).

6.7.2 Practical Considerations

Connecting to the distributed ledger. Only ASes that are members of the peering infrastructure are allowed to join the Dynam-IX ledger. An alternative to performing the admission process would be to require that all ASes connect to a specific local network at the peering infrastructure, relying on the facility to authenticate the members or to allow Dynam-IX members to authorize a new member to join. We note that if the entity responsible for the admission control stops working (in the case of using the facility to perform the admission), ASes already connected to the ledger would not be affected.

Finding peering partners. Information about the ASes can be made available in different ways, including the Dynam-IX ledger or external sources (e.g., the AS' website). Independently of the source, each AS must provide information that eases the querying process, such as ASN, $(IP, port)$ endpoints where this AS runs its Dynam-IX peer and the AS public key. Moreover, ASes can decide what business information should be made public in the attribute containing a description of the services offered by the AS (e.g., transit provider).

Deploying an interconnection agreement. Once ASes establish an interconnection agreement, they need to update their BGP routes to benefit from the new agreement. This process can be done manually by a network operator or using network automation tools. We observe that the provisioning of specific configurations by the peering infrastructure (e.g., VLANs) is not mandatory in Dynam-IX.

Incentives. Dynam-IX offers incentives for the different types of ASes connected to the peering infrastructure. *Eyeball* networks can benefit from the enhanced responsiveness to improve traffic delivery and increase the satisfaction of their access clients. Similarly, *content providers* can establish agreements to enhance the Quality of Experience faced by their subscribers. *Network providers* serve requests from eyeball networks and content providers, which would not be possible without a framework to establish interconnection agreements in short time frames. Finally, note that facilities may indirectly profit from our solution, increasing their revenue, since new ASes may connect to the peering infrastructure to benefit from dynamic interconnection, and the current members may decide to increase their port capacities.

Specifying and updating intents. Manually specifying interconnection intents and keeping them updated is an error-prone task. As an example, according to CAIDA AS-

Rank (CAIDA, 2018), Telia Company AB (ASN1299) has more than 250,000 prefixes in its cone (the set of ASes an AS can reach using customer links (LUCKIE et al., 2013)), which would probably require a substantial amount of time to specify the respective intents. A similar situation would occur to ensure that the intents attributes have been updated as, for example, AS paths change over time. Inspired by existing BGP automation tools (e.g., IRR-based filtering (6connect, Inc., 2018)), we envision that ASes can use the same principle to parse BGP updates to intents whose SLA parameters are provided by network monitoring tools, requiring the operators only to specify profiles and associate them with the intents.

BGP routing stability. Enabling ASes to establish short-term interconnection agreements can impact BGP routing stability due to the potential increase in the number and frequency of route changes. We note, however, that our approach has not introduced this problem and that it is up to the network operators to use the features of Dynam-IX properly.

Threats. Malicious ASes can try to abuse Dynam-IX in different ways. One possibility would be for two ASes two collude to create fake interconnection agreements between them and store positive feedback about them on the ledger.[4] While this attack is impossible to detect when an AS is verifying if another AS is reliable, it can verify how many different ASes have interconnected to the one under analysis. Another threat concerns to ASes offering fake services, for example, announcing one path and using a different one to route traffic. In such a case, the AS that did not get the proper service can store feedback on the ledger about its experience. If an AS consistently offers fake services, other ASes would probably decide not to interconnect with it after verifying the information on the ledger. Regarding inferring interconnection policies, an AS can try to send multiple queries to a competitor AS to get insights about its business practices. Nevertheless, since Dynam-IX design allows the ASes to decide whether or not to reply to a query, if the same AS is sending multiple queries, the AS can not answer them. It can also decide not the reply because it may know that the one querying is a competitor. Lastly, an AS that receives a query may try to redistribute it to learn potential offers of its competitor and make a cheaper one. Similarly to the previous case, if the AS suspects that a given query is not legit, it can decide not to reply the query.

[4] Sybil ASes are unlikely to happen since its rare that a given organization will have two ASes connected to the same peering infrastructure.

7 EVALUATION

Dynam-IX design matches all the requirements for a proposal to unleash the large unexplored potential of peering infrastructures to improve wide-area traffic delivery. In this chapter, we evaluate our design to understand its benefits and overheads better. We first describe the methodology of our evaluation; then, we present the results of our experiments. We end the chapter with a discussion about the results.

7.1 Methodology

Our evaluation seeks to answer four main questions: (i) *how long does it take to establish an interconnection agreement?* (ii) *how does the ledger size grow?* (iii) *how does the type of the information impact the ledger growth?* and (iv) *what are the bandwidth requirements of Dynam-IX?* With the first question, we aim to quantify the benefits of having an approach for establishing interconnection agreements and demonstrate practical feasibility, while with the other three questions we investigate the possible scalability limits of the proposed solution.

Implementation. We built a prototype implementation[1] of Dynam-IX using Hyperledger Fabric 1.0.5 (HLF) (ANDROULAKI et al., 2018), a permissioned blockchain, as the distributed tamper-proof ledger.[2] The prototype was developed in Python, Node.js, and Go in approximately 1200 lines of code. A blockchain is a tamper-proof distributed ledger consisting of a growing number of blocks securely chained together, each block comprising of several records or transactions and a hash of the content of the previous block. Permissioned blockchains are resource-efficient and easy to maintain or upgrade as they avoid the need for large amounts of resources spent on achieving consensus, by limiting the numerous untrusted entities that can write to the blockchain. Blockchain implementations support self-enforcing codes called *smart contracts*. Hyperledger Fabric provides all the components needed to run a permissioned blockchain, including smart contracts (a.k.a. chaincodes) and an ordering system for block creation. In our prototype, all procedures related to associating scores/reports and agreements and, updating the ledger are implemented using smart contracts. A recent study (SOUSA; BESSANI; VUKOLIC, 2017)

[1] Source code, documentation, and reproducibility scripts available at <https://github.com/dynam-ix/dynam-ix>.

[2] Note that using any other implementation of a tamper-proof distributed ledger is also possible.

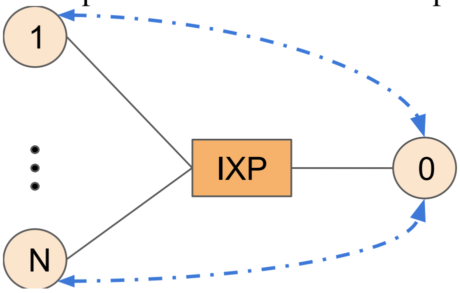
shows that Hyperledger Fabric is capable of achieving more than 10k transactions per second, well beyond our needs in a peering infrastructure context. To put things in perspective, today's route servers at one of the largest IXPs worldwide process an average of roughly four BGP routes per second (CHIESA et al., 2017b).

7.2 Interconnection time

To determine whether Dynam-IX will let operators establish agreements in short time frames, we measure the time needed to perform a query and the time required to establish an agreement. The query time is the elapsed time between an AS sending a query to a potential provider and the response with an interconnection offer. For the sake of the evaluation, we assumed that the provider will always reply to a query with an offer and that the ASes store their contact information on the ledger, which in practice may not always be the case (see §6). As queries on the ledger are local, the overhead of this operation is negligible. We measure the establishment time as the moment an AS sends an interconnection proposal (based on an offer from a provider) to the moment the AS stores the agreement related information on the ledger.

We determine the limits of Dynam-IX with a throughput test: N ASes flooding a single AS with queries and establishing interconnection agreements proposals. A case like this can happen in practice when a large number of ASes use the same congested path to reach a given prefix p. Thus, all these ASes may try to establish an interconnection agreement with a single AS that is offering connectivity towards p. We assumed that the AS receiving requests has 10 interconnection intents. Figure 7.1 illustrates this scenario.

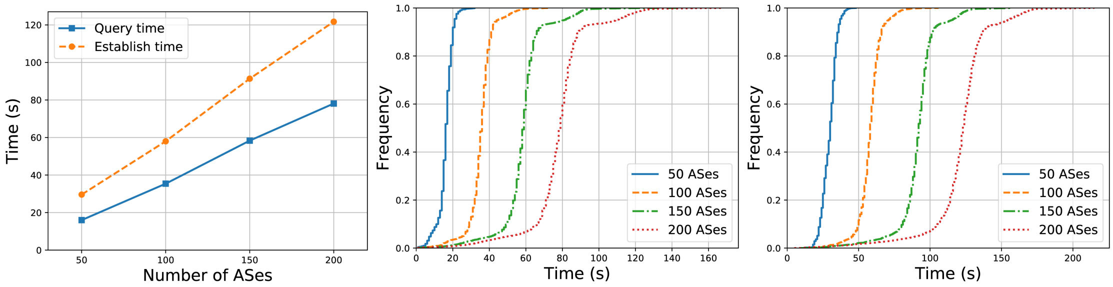
Figure 7.1: Experiment to evaluate the response time.



Source: the authors.

To evaluate this scenario, we use up to 200 AWS EC2 cloud instances (AMAZON, 2018b), each one hosting a single AS. The ASes sending the queries are instantiated in t2.micro instances (i.e., 1 vCPU, 1 GB RAM), while the AS receiving the requests is running on a c4.4xlarge instance (16 vCPU, 30 GB RAM). Besides, a c4.4xlarge instance is used to run the ordering system, which is responsible for grouping transactions into blocks, of the blockchain implementation. During the experiment, each t2.micro instance repeats the complete Dynam-IX protocol (see §6.3) 30 times.

Figure 7.2: Response time for different number of ASes.



(a) Average response time (b) Query response time (c) Establish response time

Source: the authors.

Figure 7.2(a) presents the average response times in the number of ASes. Both query and agreement times grow linearly (0.41s per additional AS) with an average response time of 120 seconds when establishing 200 agreements simultaneously. While Dynam-IX performs well even under high artificial loads, we observe that under more relaxed conditions, it can establish a single agreement in less than 10 seconds. As a reference, MegaPort, a company that uses a *centralized* approach for establishing interconnection agreements, on-demand claims that they can provide an agreement in less than 60 seconds after a network orders it (MEGAPORT, 2017).

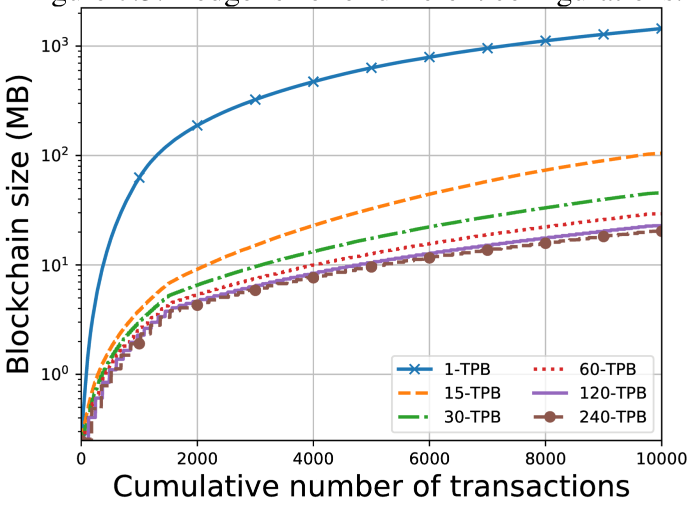
The response time of each query/proposal is approximately constant, as observed in Figure 7.2(b) and Figure 7.2(c). Even with response times in the order of a few dozens of seconds, the average number of established agreements per second (goodput) is 2.4 (for 50 ASes) and 1.4 (for 200 ASes), meaning that a single AS can establish more than 80 interconnection agreements within a minute. While we note that it is unlikely that a single AS would ever require to establish so many agreements in a small period, these results show the practicality of Dynam-IX.

7.3 Storage requirements

Every Dynam-IX peer keeps a local copy of the ledger containing the registers of all agreements and their feedback reports. To assess the storage impact for an AS using our approach, we estimate the growth of the ledger under different operations and transaction conditions. Operations provided by Dynam-IX comprise manipulation of data stored on the ledger: register an AS, register an interconnection agreement, provide the feedback about the interconnection agreement, and update AS information (e.g., the service description). We created a workload with 10k transactions combining these operations as follows: 1500 AS registrations, 2750 agreement registrations, 5500 feedback reports (2 per agreement, 1 for the customer and 1 for the provider), and 250 AS information updates (e.g., public key).

In our experiments, the growth of the ledger depends on the number of transactions grouped in each block of the blockchain (the higher the transactions per block, the higher the storage savings). This quantity depends on the arrival rate of transactions, the maximum number of transactions per block, and the timeout to create a new block. As there is no prior history of the behavior of ASes establishing agreements in short time frames, we decided to use the average number of BGP updates in the route server of a large IXP, i.e., 4 per second (CHIESA et al., 2017b). We used BGP updates as a guideline because they also relate to the establishment or withdrawal of reachability on the Internet, and we relied on data from an IXP route server because Dynam-IX is intended to run in such environments (peering infrastructures). This experiment is entirely local to each AS and is not affected by resource contention, allowing us to run it on a single computer with all the necessary Hyperledger Fabric components (peer and ordering system). We divide our experiments into two parts. The first one uses scores as the feedback information, and the second one uses reports from different sizes. Besides, to understand the storage requirements, our goal is to learn the impact of the information size in the resource demands.

Figure 7.3: Ledger size for different configurations.



Source: the authors.

Scores as feedback information. Figure 7.3 presents the results (in log scale). The worst case relates to a scenario where each block stores only one transaction (1-TPB). Such a situation happens when the interval between two consecutive transactions is longer than the block creation timeout. If the transaction rate is low, so will be the ledger growth (and therefore less likely an issue). Otherwise, if the transaction rate is high, the block limit tends to be reached well before the timeout, triggering the creation of the block. While increasing the timeout may help saving storage in low transaction rates, it may delay the agreement confirmation up to the timeout duration.

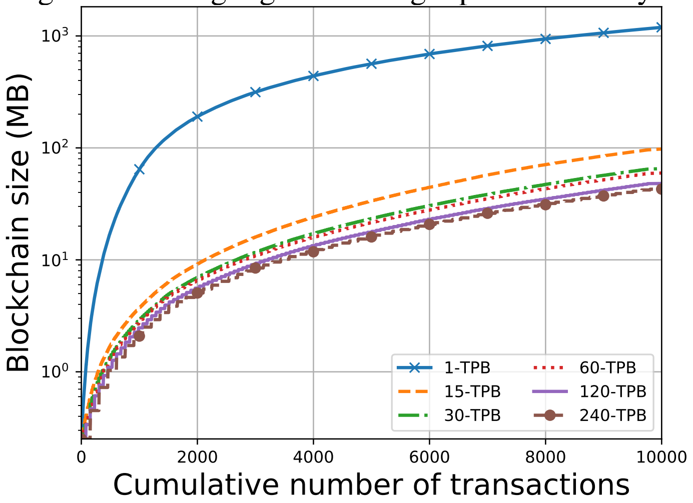
We evaluate the blockchain growth for three different block creation timeouts (15, 30, and 60 seconds) and two transaction rates, 1 (the minimum number reported in (CHIESA et al., 2017b)) and 4 per second. The number of cases is the combination of timeouts and rates (e.g., with four transactions per second and a 30s timeout, we have 120 transactions per block - TPB). As expected, the size of the blockchain grows more slowly with longer timeouts since the ledger implementation combines more transactions into a single block, thus sharing the same block data structure (hashes and pointers to the previous block on the chain). Specifically, with a timeout of 60 seconds, the size of the ledger after 10k transactions is between 29.47 MB and 20.34 MB, for 1 and 4 transactions per second, respectively.

To further illustrate, consider the same scenario with four transactions per second on average and the ledger configured with a 60-second block creation timeout. The ledger will reach 100 GB after 30 million transactions, in the worst case. Such size corresponds to approximately 10 million interconnection agreements (as each one consists of three transactions). For a peering infrastructure with 1500 members (size of the largest IXP in terms of members (IX.BR, 2018b)) and one year, it means that each AS could establish 20 unique interconnection agreements every day. We note that the block creation timeout does not impact the time to establish an interconnection agreement since the agreement is valid once the two ASes digitally have signed the contract.

Reports as feedback information. To assess if using more detailed information about the interconnection agreement will impact the storage requirements, we repeated the previous experiment but instead of using scores, we used pieces of information of 128 bytes to represent the feedback of each AS about the interconnection agreement (as discussed in Chapter 6). Figure 7.4 shows the results of this experiment.

For a timeout of 60 seconds, the size of the ledger after 10k transactions using pieces of information of 128 bytes is 59.63 MB and 43.01 MB, for 1 and 4 transactions

Figure 7.4: Ledger growth using reports of 128 bytes.



Source: the authors.

per second, respectively. While using reports of 128 bytes requires approximately twice the storage when compared to the case using scores, it allows operators to provide more details regarding their experience during the interconnection agreements. In this case, the ledger will achieve a size of 100 GB after 15 million transactions, which for a peering infrastructure with 1,500 members, represents ten unique agreements per AS every day during an entire year. We note that, as discussed in Chapter 6, there is a tradeoff between the details provided by the ASes in their report and possible privacy issues. Dynam-IX design focuses on guaranteeing the properties of the data stored on the ledger, leaving the decision about which information to store to the operators.

7.4 Bandwidth requirements

Dynam-IX is designed to operate alongside the infrastructure of a peering infrastructure. Such a condition raises the question of how much bandwidth is needed by Dynam-IX to operate. To assess the impact on regular traffic, we measure the peak bandwidth (during the experiments to measure the interconnection time) as reported by Amazon CloudWatch (AMAZON, 2018a). We consider the three different instance roles: (i) regular ASes, (ii) the AS receiving all the requests, and (iii) the ordering system of the blockchain implementation. Table 7.1 shows the peak ingress and egress bandwidth requirements during our experiments with 50 and 200 ASes.

The peak traffic related to the regular ASes was the same in all experiments, with individual peaks of 4.8 Mbps (ingress traffic) and 0.8 Mbps (egress traffic). The reason for that is because the number of messages (and their size) exchanged by each regular AS

Table 7.1: Comparison of the peak bandwidth requirements of Dynam-IX for a different number of ASes.

	50 ASes - Ing.	50 ASes - Egr.	200 ASes - Ing.	200 ASes - Egr.
Regular ASes	4.8 Mbps	0.8 Mbps	4.8 Mbps	0.8 Mbps
AS recv. reqs.	8.1 Mbps	9.3 Mbps	9.4 Mbps	12.4 Mbps
Ordering syst.	8.7 Mbps	224.1 Mbps	18.8 Mbps	931.1 Mbps

Source: the authors.

is the same in both experiments. The AS receiving all the requests presents ingress peaks between 8.1 Mbps and 9.4 Mbps (for experiments with 50 and 200 ASes respectively) and egress peaks between 9.3 Mbps and 12.4 Mbps likewise. While the number of ASes in the experiment with 200 ASes is four times the number of the other experiment, the ingress and the egress bandwidth increased by 16% and 13.5%, respectively. Unlike a regular AS, the one receiving all the requests needs to handle more interconnection requests and also register more agreements on the ledger, thus, resulting in higher bandwidth requirements. The traffic generated by these components is mainly related to the Dynam-IX protocol, used to query for offers, establish agreements, and interact with the ledger.

The ordering system presents egress peaks between 224.1 Mbps and 931.1 Mbps (experiments respectively with 50 and 200 ASes) while the ingress peaks remain between 8.7 Mbps and 18.8 Mbps. These numbers represent an increase of 116% on the ingress peak bandwidth and of 315% for egress bandwidth peak demands. Such traffic is directly related to the creation and distribution of the blocks containing the transactions. Since every AS keeps a local copy of the ledger, the bandwidth required by the blockchain ordering system grows proportionally to the number of ASes. We observe that a different ledger implementation/model might require less traffic.

Casting these values to a large peering infrastructure with 1500 members, the cumulative highest peak of traffic for the entire Dynam-IX framework is around 7 Gbps (in the throughput test). Considering that large peering infrastructures such as DE-CIX (DE-CIX, 2018) and AMS-IX (AMS-IX, 2018) carry an aggregated traffic of more than 5 Tbps, the demands of Dynam-IX from the peering infrastructure will be approximately 0.14% of the total traffic and can be rate limited if necessary, illustrating the practicality of deploying Dynam-IX at peering infrastructures. We note that the scenario of the throughput experiment describes an extreme case (considering the peak bandwidth requirements) and in normal conditions, the bandwidth requirements tend to be much smaller than 7 Gbps of traffic for 1500 ASes.

7.5 Discussion

Tens of interconnection agreements per minute. Our experiments show that a single AS can establish tens of interconnection agreements within a minute. While such a condition is unlikely to happen, we note that this capacity could be useful when an AS connects to the peering infrastructure for the first time and needs to interconnect with several ASes (see Chapter 4).

Impact of the number of interconnection intents. In our experiments, we assumed that the AS receiving all the requests has ten interconnection intents (i.e., is offering connectivity to 10 different prefixes). We argue that such a decision is not a limiting factor in our evaluation for two reasons. First, as of today, 86% of the ASes have at most ten prefixes in their customer cone (CAIDA, 2019). Second, the lookup process uses the longest prefix matching algorithm that requires logarithmic time, which is the same used by AS routers that handle several requests per second efficiently.

Ledger bandwidth requirements. Dynam-IX bandwidth requirements are strongly related to our choice of using a permissioned blockchain as the ledger to store information about the interconnection agreements in our prototype. We note, however, that our prototype establishes an upper bound of the bandwidth requirements. Using a different implementation of a distributed tamper-proof ledger such as Ethereum might offer the same benefits with smaller traffic requirements. Relaxing the properties of the ledger (e.g., not being tamper-proof) tends to reduce the bandwidth requirements also. However, such a case would not offer the same guarantees as Dynam-IX, which can hinder the adoption of the solution.

8 CONCLUSIONS

Dynam-IX is a step towards a more dynamic interconnection ecosystem. In the following sections, we summarize our findings and the main takeaways of this thesis and discuss the possibilities to continue this research.

8.1 Remarks

Unleashing the vast unexplored potential of peering infrastructures. Dynam-IX provides a well-defined process to establish interconnection agreements and a method to build trust cooperatively. Our approach allows operators to improve wide-area traffic delivery performance by exploiting the rich connectivity opportunities at peering infrastructures. By interviewing and surveying more than 100 network operators and peering coordinators, we have identified the current interconnection practices and confirmed that interconnecting continues to be an ad-hoc and lengthy process driven by personal relationships and brand image. In our survey operators also confirmed their desire for an improved interconnection process (Chapter 4), which could lead to several improvements in the interconnection ecosystem, including more responsiveness to the traffic dynamics, increased utilization of peering ports, and new economic opportunities. Nevertheless, to be able to provide a better interconnection process, we need to find a place offering physical connectivity among ASes and with spare capacity available. It is well-known that peering infrastructures offer rich connectivity among any pair of members. To confirm the availability of spare capacity, we collected and analyzed traces from two relevant IXPs and confirmed the existence of proper conditions to achieve this goal (Chapter 5).

To proper achieve the requirements of a practical solution, we relied upon three key insights, namely, a distributed architecture, a high-level interconnection intent abstraction, and a tamper-proof ledger. First, while the intuitive approach would be to design a centralized solution, the distributed architecture allowed us to guarantee the privacy of ASes interconnection policies in a simple manner. In Dynam-IX, each AS locally stores its interconnection policies, thus keeping control when and with which ASes it is interested in sharing the policies. The distributed approach also offers benefits upon incremental deployment if at least two ASes decide to start using Dynam-IX. Second, the high-level interconnection abstraction defines an expressive interface that ASes can use to specify their peering policies and interconnection desires. Third, the tamper-proof ledger complements the existing practices to build trust among ASes, offering a place where ASes can find (reliable) information about potential peering partners.

We then demonstrated through a prototype and set of experiments that Dynam-IX successfully achieves its goal by allowing a single AS to establish tens of interconnections agreements within a minute without imposing significant overheads neither for the ASes or the peering infrastructure. Our experiments also have shown that our approach scales to the size of the largest peering infrastructures, which was also one of the requirements for an effective solution.

Interest from the Industry. During the time of this Ph.D., we have interacted with potential industry partners regularly, resulting in discussions to effectively deploy a Dynam-IX-like solution. First, a large peering infrastructure has approached us interested in benefiting from our interconnection protocol and high-level interconnection intent abstraction to allow their members to interconnect faster. They were also interested in extending the functionalities of the distributed tamper-proof ledger to integrate the agreement payment in the process. The discussions evolved with the signature of an NDA (Non-Disclosure Agreement), but later the company alleged that while they see the benefits of interconnecting faster, the market (their customers) would not be ready yet for such a change. Later, a large content provider contacted us intending to create a consortium of ASes to start using a solution like Dynam-IX. They were mostly interested in benefiting from the information that ASes would store on the ledger. AS a large content provider, they are interested in interconnecting with as many ASes as possible. However, doing that might create security (e.g., prefix hijacking, route leaks) and performance issues (e.g., poor QoE). Thus, the information on the ledger would help them to decide whether or not they should interconnect with a given AS. Similarly to the discussions with the other company, these discussions have not evolved into an actual deployment of Dynam-IX. In this case, however, the limiting factor to start were bureaucratic aspects. While none of our discussions with potential Industry partners have evolved into an effective deployment of Dynam-IX, their interest in our proposal confirms the relevance of the problems addressed in this thesis.

Publications (and submissions). The main contributions of this thesis resulted in a conference paper and a journal submission. Dynam-IX design and evaluation have been published at ACM CoNEXT'18 (MARCOS et al., 2018), while the detailed results of our survey with network operators are under submission at ACM CCR. Before being accepted at CoNEXT'18, we published the preliminary results of Dynam-IX as a poster at ACM SIGCOMM'18 and won the 3rd place of the ACM Student Research Competition. We also gave talks of our approach in different events such as RIPE 76, the European Peering Forum 2018, and the ACM/ISOC ANRW 2018. Additionally, we published a paper at ANRW'18 discussing a semi-fair protocol to score ASes on the ledger (ALOWAYED et al., 2018). Finally, during my time as a visiting researcher at KAUST, I was able to participate in a project not related to Dynam-IX that resulted in a paper published at IFIP Networking 2018 (Kathiravelu et al., 2018).

8.2 Future work

Dynam-IX allows ASes to effectively benefit from the rich connectivity of peering infrastructures to improve wide-area traffic delivery. Such a feature strongly aligns with the current trends of self-driving networks (FEAMSTER; REXFORD, 2017). We envision a set of components that can act alongside Dynam-IX towards an even more dynamic and responsive interconnection ecosystem. We discuss them below.

Objectively measuring the "quality" of a peering partner (TrustAS). While Dynam-IX offers a distributed tamper-proof ledger for ASes to store information about their interconnection agreements, we believe that having a method that allows ASes to objectively measure the quality of a potential peering partner would be of great value. We believe that ASes could leverage different vantage points that are publicly available (e.g., RIPE Atlas Probes) to issue a series of measurements towards a given AS and combine them to identify the "quality" of the potential peering partner. The envisioned solution should provide answers to the following questions (i) how to identify a set of representative vantage points to perform the measurements? (ii) what metrics are the most representative for inferring the (forwarding) "quality" of a given AS? (iii) how to combine measurements from multiple vantage points into a representative piece of information to help ASes identifying potential peering partners?

Enhanced interconnection intent abstraction (PINT). The current intent abstraction of Dynam-IX allows ASes to specify their interconnection desires and the properties of a given offer. We plan to extend our abstraction to empower ASes with an abstraction to configure their peering behavior. Such an abstraction should allow ASes to specify when they should try to find a new peering partner (e.g., if packet loss is higher than a given threshold) and how to reply to interconnection requests (e.g., only reply if average link utilization is below a given threshold). Main research challenges are (i) verifying that policies are not conflicting, and (ii) defining a representative abstraction that covers network operators' needs.

Peering analytics (PeA). Operators need to identify what is happening in their networks and what are the best actions to be taken to be responsive to Internet traffic dynamics. Actions can be related to long-term (e.g., natural traffic growth, new service) or short-term decisions (e.g., traffic surges, link failures, congestion). To make these decisions, operators need to sample and analyze control and data plane information. These tasks have become challenging due to traffic growth and the increasing need for responsiveness to meet the strict service requirements of modern Internet services. Challenges include: how to sample the control and data plane information without losing granularity? How to efficiently store the data? How to allow operators to quickly query and process the stored data? Given the network policy, how to identify the need for a new interconnection agreement? How to perform such a task in real-time?

SmartPeering. An approach to ease the establishment of interconnection agreements enables a series of interconnection opportunities that were not possible before. In this case, it will be essential to have a method to identify among the interconnection alternatives and the AS current requirements, which one would be the best. We envision that such a tool will combine the information from the ledger, with the requirements of the AS, and the offers other ASes have sent, to pick the one that will provide more benefits. Research questions include: (i) how different goals such as performance, robustness, and security correlate? (ii) how to check that a given decision will not violate the AS' peering policies? (iii) how can different types of ASes (e.g., content provider, CDN, transit, eyeball) optimize their traffic?

Combining the above solutions could lead to a self-driving network that improves wide-area traffic delivery, increases the resilience of the Internet, and reduces operational costs (which are the biggest in networking). Nevertheless, empowering ASes with a significant level of responsiveness to the Internet traffic dynamics can impact Internet routing stability if not used properly. Understanding these impacts are also in our research roadmap. We note, however, that the route instability problem was not created by Dynam-IX (or any of the envisioned extensions). In fact, they have been a focus of the research community for several years.

REFERENCES

6connect, Inc. *IRR Power Tools – A utility for managing Internet Routing Registry (IRR) filters*. 2018. Available at <https://github.com/6connect/irrpt>.

AGER, B. et al. Anatomy of a Large European IXP. In: **SIGCOMM**. [S.l.: s.n.], 2012.

AHMED, A. et al. Peering vs. Transit: Performance Comparison of Peering and Transit Interconnections. In: **ICNP**. [S.l.: s.n.], 2017.

ALOWAYED, Y. et al. Picking a Partner: A Fair Blockchain Based Scoring Protocol for Autonomous Systems. In: **ANRW**. [S.l.: s.n.], 2018.

AMAZON. **Amazon CloudWatch**. 2018. Available at <https://aws.amazon.com/cloudwatch>.

AMAZON. **Amazon EC2**. 2018. Available at <https://aws.amazon.com/ec2/>.

AMS-IX. **AMS-IX**. 2018. Available at <https://ams-ix.net>.

ANDROULAKI, E. et al. Hyperledger Fabric: A Distributed Operating System for Permissioned Blockchains. In: **EuroSys**. [S.l.: s.n.], 2018.

ANTICHI, G. et al. ENDEAVOUR: A Scalable SDN Architecture for Real-World IXPs. **IEEE Journal on Selected Areas in Communications**, v. 35, n. 11, Nov 2017.

ARCEP. **The State of Internet in France**. 2018. Available at <https://archives.arcep.fr/uploads/tx_gspublication/report-state-internet-2018_conf050618-ENG.pdf>.

ARTHUR, C. **iOS 5 update causes massive Internet traffic spike - to users' frustration**. [S.l.]: The Guardian, 2011. Available at <https://www.theguardian.com/technology/2011/oct/13/ios-5-update-internet-traffic-spike>.

BAFNA, S.; PANDEY, A.; VERMA, K. Anatomy of the Internet Peering Disputes. **CoRR**, abs/1409.6526, 2014. Available from Internet: <http://arxiv.org/abs/1409.6526>.

BAKKER, N. et al. **Internet Exchange BGP Route Server (RFC7947)**. 2016. Available at <https://tools.ietf.org/html/rfc7947>.

BLUM, M. Coin Flipping by Telephone a Protocol for Solving Impossible Problems. **ACM SIGACT News**, v. 15, n. 1, 1983.

BORNSTAEDT, F. **New Levels of Cooperation between Eyeball ISPs and OTT/CDNs**. 2017. Available at <https://ripe75.ripe.net/archives/video/126/> – starting at 7min50s.

BRODKIN, J. **iOS 7 downloads consumed 20 percent of an ISP's traffic on release day**. [S.l.]: arstechnica, 2013. Available at <https://arstechnica.com/information-technology/2013/11/ios-7-downloads-consumed-20-percent-of-an-isps-traffic-on-release-day/>.

CAIDA. **CAIDA AS-Rank**. 2018. Available at <http://as-rank.caida.org>.

CAIDA. **AS-Rank**. 2019. Available at <https://as-rank.caida.org/>.

CASTRO, I.; GORINSKY, S. T4P: Hybrid interconnection for cost reduction. In: **INFOCOM Workshops 2012**. [S.l.: s.n.], 2012.

CASTRO, I. et al. Route Bazaar: Automatic Interdomain Contract Negotiation. In: **HotOS**. [S.l.: s.n.], 2015.

CHATZIS, N. et al. There is More to IXPs Than Meets the Eye. **SIGCOMM Comput. Commun. Rev.**, 2013.

CHIESA, M. et al. **Internet Routing Privacy Survey**. 2017. Available at <https://six-pack.bitbucket.io/media/privacy-survey-2017.pdf>.

CHIESA, M. et al. SIXPACK: Securing Internet eXchange Points Against Curious onlooKers. In: **CoNEXT**. [S.l.: s.n.], 2017.

CHIESA, M. et al. Inter-domain Networking Innovation on Steroids: Empowering IXPs with SDN Capabilities. **IEEE Communications Magazine**, 2016.

CHIU, Y.-C. et al. Are We One Hop Away from a Better Internet? In: **IMC**. [S.l.: s.n.], 2015.

COLOCLUE. **KEES - The Coloclue Network Automation Toolchain**. 2019. Available at <https://github.com/coloclue/kees>.

COMCAST. **Service Level Agreement for Wholesale Dedicated Internet**. 2019. Available at <https://portals.comcasttechnologysolutions.com/sites/default/files/service_level_agreement_for_wholesale_dedicated_internet_sla07292014.pdf>.

CONSOLE. **Console - The Cloud Connection Company**. 2017. Available at <https://www.consoleconnect.com/>.

COSTAN, V.; DEVADAS, S. **Intel SGX Explained**. 2016. Cryptology ePrint Archive, Report 2016/086. <http://ia.cr/2016/086>.

DE-CIX. **DE-CIX**. 2018. Available at <https://www.de-cix.net>.

DHAMDHERE, A.; DOVROLIS, C. The Internet is Flat: Modeling the Transition from a Transit Hierarchy to a Peering Mesh. In: **CoNEXT**. [S.l.: s.n.], 2010.

DIETZEL, C. et al. SDN-enabled Traffic Engineering and Advanced Blackholing at IXPs. In: **SOSR**. [S.l.: s.n.], 2017.

DIETZEL, C.; FELDMANN, A.; KING, T. Blackholing at IXPs: On the Effectiveness of DDoS Mitigation in the Wild. In: KARAGIANNIS, T.; DIMITROPOULOS, X. (Ed.). **PAM 2016**. [S.l.: s.n.], 2016.

DIMITROPOULOS, X. et al. On the 95-percentile billing method. In: **PAM**. [S.l.: s.n.], 2009.

DUNCAN, S. **Australian Internet slows to a crawl after undersea cable cut**. 2017. Available at <http://www.dailymail.co.uk/news/article-5146795/Aussie-internet-slows-crawl-undersea-cable-cut.html>.

EDMUNDSON, A. et al. Nation-State Hegemony in Internet Routing. In: . [S.l.: s.n.], 2018. (COMPASS).

EPSILON. **Epsilon Telecommunications Limited – Connectivity made simple**. 2017. Available at <www.epsilontel.com/>.

FANOU, R.; VALERA, F.; DHAMDHERE, A. Investigating the Causes of Congestion on the African IXP Substrate. In: **IMC**. [S.l.: s.n.], 2017.

FEAMSTER, N. Revealing Utilization at Internet Interconnection Points. **CoRR**, abs/1603.03656, 2016.

FEAMSTER, N.; REXFORD, J. Why (and How) Networks Should Run Themselves. **CoRR**, 2017.

FERREIRA, P.; MINDEL, J.; MCKNIGHT, L. Why Bandwidth Trading Markets Have not Matured? Analysis of Technological and Market Issues. **International Journal of Technology, Management and Policy**, v. 2, 2004.

FORUM-IX. **Beer, Gear and Peer at IX Forum**. 2016. Available at <http://forum.ix.br/en/index.html>.

FRALEIGH, C.; TOBAGI, F. A.; DIOT, C. Provisioning IP Backbone Networks to Support Latency Sensitive Traffic. In: **INFOCOM**. [S.l.: s.n.], 2003.

FRANCE-IX. **Marketplace FranceIX's**. 2019. Available at <https://www.franceix.net/en/solutions/marketplace/>.

FUSARO, P. C.; MILLER, R. M. **What went wrong at Enron: Everyone's guide to the largest bankruptcy in US history**. [S.l.]: John Wiley & Sons, 2002.

GHASEMI, M. et al. Performance Characterization of a Commercial Video Streaming Service. In: **IMC**. [S.l.: s.n.], 2016.

GILL, P.; SCHAPIRA, M.; GOLDBERG, S. A Survey of Interdomain Routing Policies. **SIGCOMM Comput. Commun. Rev.**, 2013.

GIOTSAS, V. et al. Detecting Peering Infrastructure Outages in the Wild. In: **SIGCOMM**. [S.l.: s.n.], 2017.

GIOTSAS, V. et al. Inferring Complex AS Relationships. In: **IMC**. [S.l.: s.n.], 2014.

GIOTSAS, V. et al. Mapping Peering Interconnections to a Facility. In: **CoNEXT**. [S.l.: s.n.], 2015.

GIOVANNETTI, E.; RISTUCCIA, C. A. Estimating market power in the Internet backbone. Using the IP transit Band-X database. **Telecommunications Policy**, Elsevier, 2005.

GODFREY, P. B. et al. Pathlet Routing. In: **SIGCOMM**. [S.l.: s.n.], 2009.

GOLDREICH, O.; MICALI, S.; WIGDERSON, A. How to Play any Mental Game or A Completeness Theorem for Protocols with Honest Majority. In: **STOC'87**. [S.l.]: ACM, 1987. p. 218–229.

GONCHAROFF, K. **Bandwidth's New Bargaineers**. [S.l.]: MIT Technology Review, 1998. Available at <https://www.technologyreview.com/s/400275/bandwidths-new-bargaineers/>.

GRIFFIOEN, J.; WOLF, T.; CALVERT, K. L. A Coin-Operated Software-Defined Exchange. In: **ICCCN**. [S.l.: s.n.], 2016.

GUPTA, A. et al. An Industrial-Scale Software Defined Internet Exchange Point. In: **NSDI**. [S.l.: s.n.], 2016.

GUPTA, A. et al. SDX: A Software Defined Internet Exchange. In: **SIGCOMM**. [S.l.: s.n.], 2014.

HARGRAVE, W. et al. **Mitigating Negative Impact of Maintenance through BGP Session Culling (draft-ietf)**. 2018. Available at <https://tools.ietf.org/html/draft-ietf-grow-bgp-session-culling-05>.

HEITZ, J. et al. **BGP Large Communities Attribute (RFC8092)**. 2017. Available at <https://tools.ietf.org/html/rfc8092>.

HOPUS. **HOPUS - meet networks that matter**. 2018. Available at <http://hopus.net>.

HOPUS. **HOPUS Pricing scheme**. 2018. Available at <http://hopus.net/price>.

Hu, C. et al. A Measurement Study on Potential Inter-Domain Routing Diversity. **IEEE Transactions on Network and Service Management**, 2012.

INC., G. C. **Sample Business Contracts**. 2019. Available at <https://contracts.onecle.com/limelight-networks/global-crossing-services.shtml>.

INFLECT. **Find Data Center and Colocation Connectivity | Build Hybrid Cloud | Inflect**. 2019. Available at <https://inflect.com/>.

IX.BR. **IX.br**. 2018. Available at <http://ix.br/>.

IX.BR. **São Paulo IXP members list**. 2018. Available at <http://ix.br/particip/sp>.

Kathiravelu, P. et al. Moving Bits with a Fleet of Shared Virtual Routers. In: **Networking**. [S.l.: s.n.], 2018.

KATZ, D. et al. **Multiprotocol Extensions for BGP-4 (RFC4760, updated by 7606)**. 2007. Available at <https://tools.ietf.org/html/rfc4760>.

KONSTANTARAS, S. **Implementation of RPKI and IRR filtering on the AMS-IX platform**. 2018. Available at <https://www.ripe.net/support/training/ripe-ncc-educa/presentations/use-cases-stavros-konstantaras.pdf>.

KOTRONIS, V. et al. Stitching Inter-Domain Paths over IXPs. In: **SOSR**. [S.l.: s.n.], 2016.

KOTRONIS, V. et al. Shortcuts Through Colocation Facilities. In: **IMC**. [S.l.: s.n.], 2017.

LINX. **LINX**. 2018. Available at <https://linx.net/>.

LODHI, A. et al. Complexities in Internet Peering: Understanding the “Black” in the “Black Art”. In: **INFOCOM**. [S.l.: s.n.], 2015.

LODHI, A. et al. Using peeringdb to understand the peering ecosystem. **SIGCOMM Comput. Commun. Rev.**, ACM, New York, NY, USA, v. 44, n. 2, p. 20–27, abr. 2014. ISSN 0146-4833. Available from Internet: <http://doi.acm.org/10.1145/2602204.2602208>.

LUCKIE, M. et al. AS Relationships, Customer Cones, and Validation. In: **IMC**. [S.l.: s.n.], 2013.

MAIGRON, P. **Regional Internet Registries Statistics**. 2019. Available at <https://www-public.imtbs-tsp.eu/~maigron/RIR_Stats/RIR_Delegations/ByRIR/Stats-ByRIR.html>.

MARCOS, P. et al. Dynam-IX: A Dynamic Interconnection eXchange. In: **CoNEXT**. [S.l.: s.n.], 2018.

MCGEE-ABE, J. **Apple devices behind DE-CIX Frankfurt 5.88Tbps data traffic rate**. 2017. Available at <http://www.capacitymedia.com/Article/3751343/Apple-devices-behind-DE-CIX-Frankfurt-588Tbps-data-traffic-rate>.

MEGAPORT. **Megaport - A Better way to connect**. 2017. Available at <http://megaport.com/>.

MEIER-HAHN, U. **The Internet was Built on Trust. But What Does it Run On?** 2017. Available at <https://labs.ripe.net/Members/uta_meier_hahn/the-internet-was-built-on-trust-but-what-does-it-run-on>.

MORALES, C. **NETSCOUT Arbor Confirms 1.7 Tbps DDoS Attack; The Terabit Attack Era Is Upon Us**. 2018. Available at <https://asert.arbornetworks.com/netscout-arbor-confirms-1-7-tbps-ddos-attack-terabit-attack-era-upon-us/>.

NANOG. **AS4134/AS4847 - Appear to be hijacking some ip space**. 2019. Available at <https://mailman.nanog.org/pipermail/nanog/2019-April/100415.html>.

NANOG. **ATT contact**. 2019. Available at <https://mailman.nanog.org/pipermail/nanog/2019-May/100752.html>.

NANOG. **Comcast contact for wholesale ethernet/local loop**. 2019. Available at <https://mailman.nanog.org/pipermail/nanog/2019-March/099895.html>.

NANOG. **IP Route Hijacking Bad Actor: AS57129/RU-SERVERSGET-KRSK, RU/Optibit LLC**. 2019. Available at <https://mailman.nanog.org/pipermail/nanog/2019-August/102774.html>.

NARULA, N.; VASQUEZ, W.; VIRZA, M. zkLedger: Privacy-Preserving Auditing for Distributed Ledgers. In: **NSDI**. [S.l.: s.n.], 2018.

NEWMAN, L. H. **GitHub Survived the Biggest DDoS Attack Ever Recorded**. 2018. Available at <https://www.wired.com/story/github-ddos-memcached/>.

NL-IX. **The technology behind the route server configurator**. 2019. Available at <https://www.nl-ix.net/feed/blog/technology-behind-route-server-configurator/>.

NOMIKOS, G. et al. O Peer, Where Art Thou?: Uncovering Remote Peering Interconnections at IXPs. In: **IMC**. [S.l.: s.n.], 2018.

NORTON, W. B. **The Internet Peering Playbook: Connecting to the Core of the Internet**. [S.l.]: DrPeering Press, 2014.

PACKETFABRIC. **PacketFabric**. 2017. Available at <https://www.packetfabric.com/>.

PeeringDB. **PeeringDB**. 2018. Available at <https://www.peeringdb.com>.

REKHTER, Y.; SANGLI, S. R. **BGP Extended Communities Attribute (RFC4360, updated by 7153, 7606)**. 2006. Available at <https://tools.ietf.org/html/rfc4360>.

RICHTER, P. et al. Peering at Peerings: On the Role of IXP Route Servers. In: **IMC**. [S.l.: s.n.], 2014.

SAMBASIVAN, R. R. et al. Bootstrapping Evolvability for Inter-domain Routing with D-BGP. In: **SIGCOMM**. [S.l.: s.n.], 2017.

SANDVINE. **Traffic Spotlight: iOS 7 Launch**. 2013. Available at <http://www.internetphenomena.com/2013/09/traffic-spotlight-ios-7-launch/>.

SANDVINE. **FIFA 16 - The Beautiful Game?** 2015. Available at <http://www.internetphenomena.com/2015/09/fifa-16-the-beautiful-game/>.

SCHLINKER, B. et al. Engineering Egress with Edge Fabric: Steering Oceans of Content to the World. In: **SIGCOMM**. [S.l.: s.n.], 2017.

SOUSA, J.; BESSANI, A.; VUKOLIC, M. A Byzantine Fault-Tolerant Ordering Service for the Hyperledger Fabric Blockchain Platform. **CoRR**, abs/1709.06921, 2017.

TELEGEOGRAPHY. **Internet Exchange Map**. 2018. Available at <https://www.internetexchangemap.com/>.

VALANCIUS, V. et al. MINT: A Market for INternet Transit. In: **ReArch**. [S.l.: s.n.], 2008.

WOLF, T. et al. ChoiceNet: Toward an Economy Plane for the Internet. **SIGCOMM Comput. Commun. Rev.**, 2014.

WOODCOCK, B.; FRIGINO, M. 2016 Survey of Internet Carrier Interconnection Agreements. **Packet Clearing House**, nov. 2016.

YAP, K.-K. et al. Taking the Edge off with Espresso: Scale, Reliability and Programmability for Global Internet Peering. In: **SIGCOMM**. [S.l.: s.n.], 2017.