Dynam-IX: a Dynamic Interconnection eXchange

https://dynam-ix.github.io

Pedro Marcos
UFRGS and FURG

Marco Chiesa
KTH Royal Institute of Technology

Lucas Müller
UFRGS and CAIDA/UCSD

Pradeeban Kathiravelu
INESC-ID and
Université catholique de Louvain

Christoph Dietzel
DE-CIX/TU Berlin

Marco Canini
KAUST

Marinho Barcellos
UFRGS

ABSTRACT

Autonomous Systems (ASes) can reach hundreds of networks via Internet eXchange Points (IXPs), allowing improvements in traffic delivery performance and competitiveness. Despite the benefits, any pair of ASes needs first to agree on exchanging traffic. By surveying 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, ASes prefer long-term agreements at the expense of a potential mismatch between actual delivery performance and current traffic dynamics. ASes also miss interconnection opportunities due to trust reasons. To improve wide-area traffic delivery performance, we propose Dynam-IX, a framework that allows operators to build trust cooperatively and implement traffic engineering policies to exploit the rich interconnection opportunities at IXPs quickly. 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 ASes. We build and evaluate a Dynam-IX prototype and show that an AS can establish tens of agreements per minute with negligible overhead for ASes and IXPs.

CCS CONCEPTS

• **Networks** → **Network management**;

KEYWORDS

Peering, Internet eXchange Point, wide-area traffic delivery

ACM Reference Format:
Pedro Marcos, Marco Chiesa, Lucas Müller, Pradeeban Kathiravelu, Christoph Dietzel, Marco Canini, and Marinho Barcellos. 2018. Dynam-IX: a Dynamic Interconnection eXchange: https://dynam-ix.github.io. In *The 14th International Conference on emerging Networking EXperiments and Technologies (CoNEXT '18), December 4–7, 2018, Heraklion, Greece.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3281411.3281419

CoNEXT '18, December 4–7, 2018, Heraklion, Greece

ACM ISBN 978-1-4503-6080-7/18/12…$15.00
https://doi.org/10.1145/3281411.3281419

1 INTRODUCTION

The rise of IXPs. The Internet topology has changed greatly over the past decade: it is now richly connected and flattened [28, 31]. The change was mostly driven by the popularization of Internet eXchange Points (IXPs), which became the high-speed physical crossroads of Internet traffic. There are over 800 IXPs spread worldwide [9], and the largest ones carry multiple terabits of traffic per second. IXPs enable Autonomous Systems (ASes) to reach hundreds of other networks [12] directly. By increasing connectivity, IXPs contribute to improving the quality of Internet traffic delivery with lower latency and higher throughput [13].

Interconnecting is a cumbersome process. IXPs provide high-speed physical connectivity (i.e., L2) among any pair of IXP members. 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 process of negotiating an interconnection agreement is largely a manual and unstructured effort that takes days or even weeks to complete. To better understand the limitations of the interconnection ecosystem, we conducted several interviews and a survey [57] of 100+ network operators and peering coordinators. Put simply, before any routing change is even attempted, much of the process relies on human interaction including in-person meetings, trust and reputation, billing and payment arrangements, and possibly lengthy legal negotiations.

Technical and human factors affect today's ability to leverage the rich IXP connectivity diversity. In case of settlement-free peering, ASes may adopt multi-lateral agreements using a single BGP session with a route server. Even though this may simplify the negotiation of agreements, it comes with undesired technical limitations. Multi-lateral peering reduces control over routing decisions because route servers propagate only the best route. This is undesirable as it negates opportunities to optimize traffic engineering in response to downstream congestion [68, 74] and to quickly reroute traffic under failures [21]. Furthermore, ASes may not be willing to disclose their peering policies to the IXP [26].

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 IXPs traverses bi-lateral agreements [26, 27, 65]. 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 [1–3, 67] and link failures [4]. Our survey results further corroborate the operators' desire for fast interconnection procedures; a majority of the respondents (56%) states slow interconnection times hinder their ability to achieve high port utilization.

In both cases, 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, even if doing so would benefit both networks.

An underutilized interconnection ecosystem. We posit that IXPs have a large unexplored potential to improve wide-area traffic delivery performance, as they (*i*) provide physical connectivity among hundreds of ASes and (*ii*) their peering ports have a substantial spare capacity. Due to conservative network planning, spare capacity is found in interconnection links of most ASes [27, 36]. We confirmed this empirically, by analyzing 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 (§4).

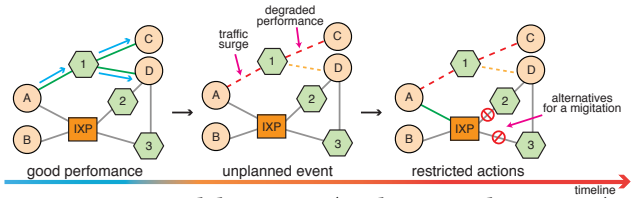


Figure 1: Limited dynamism (circle - AS, polygon - ISP).

Motivating example. To illustrate this point, we provide an example in Figure 1. A is an 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 would be to send the traffic to D via IXP link. However, this is not possible for A has agreements neither with ISP_2 nor with ISP_3. A has L2 connectivity with potential providers but is unable to establish an agreement in a short time frame.

Challenges. Facilitating interconnection via IXPs poses two major challenges: (*i*) How to quickly negotiate an agreement? (*ii*) How to decide which networks can be trusted to route traffic? First, there are currently no means to discover interconnection opportunities systematically [1], nor there exists a well-defined method to express interconnection negotiation procedures. Second, knowledge provided by personal relationships is crucial but slow to acquire; the interconnection ecosystem would benefit from a trustworthy mechanism to identify ASes deemed reliable to interconnect with, but this is missing.

A long-standing open problem. There have been efforts since the early 2000s to commoditize the bandwidth market [44] and enable short-term agreements [39, 41]. These early efforts failed as the required technology and standardization were missing [37], but now conditions are different. A recent survey highlights that ~99% of the over 1.9 M surveyed *peering agreements* were established without any formal contract [73], indicating that operators are willing to avoid lengthy bureaucratic discussions. The emergence of new connectivity services, such Epsilon [34], MegaPort [58], PacketFabric [62], and Console Connect [29], is another indication that conditions now are different. In the academic front, proposals include MINT [70], ChoiceNet [71], and RouteBazaar [24]. Both industry and academic proposals suffer from two important limitations: they require ASes to disclose their interconnection policies and do not offer proper methods to assess the quality of peering partners.

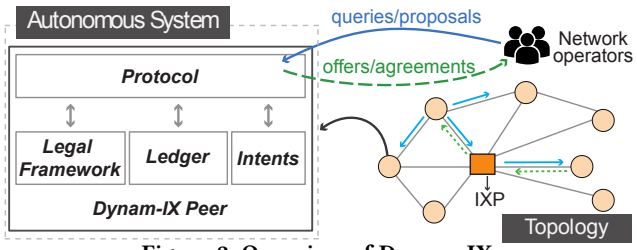


Figure 2: Overview of Dynam-IX.

Our novel systematic approach. We propose Dynam-IX, a Dynamic Interconnection eXchange, approach for facilitating interconnection agreement establishment. By complementing today's human-based practices, Dynam-IX allows network operators to leverage the rich connectivity diversity of IXPs, ultimately improving wide-area traffic delivery performance. Figure 2 provides an overview. A Dynam-IX peer is a node that interacts with other ASes through a *protocol* to offer and to query interconnection opportunities. These are described using a *high-level interconnection intent abstraction* that lets operators easily express interconnection policies and properties (e.g., pricing, SLA, duration). 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., IXP). Yet, the business-sensitive nature of the agreements among IXP members is something that would discourage IXPs from operating such services. Once an agreement

[1] While PeeringDB [63] may offer information about potential peers, the data about IXP members can be outdated or missing [55].

is established, its information is stored on the ledger, and ASes (automatically) inject via BGP the routes reflecting the new agreement and start exchanging traffic.

Our contributions are as follows:

(1) We show the opportunity to improve wide-area traffic delivery performance (§3 and §4). Our findings are supported by extensive interviews, a survey of over 100 network operators, and an analysis 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 (§5).
(3) We evaluate a prototype implementation of Dynam-IX (§6). 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 IXPs.

2 INTERCONNECTION ECOSYSTEM

Internet eXchange Points. IXPs are switching fabrics that typically provide agreement opportunities to hundreds of member ASes [12]. To connect their networks to an IXP, ASes are usually required to pay (to the IXP) a monthly fee based on the capacity of their interconnection ports. Once connected to the IXP fabric, members manually look for partners (e.g., searching into PeeringDB [63]) and implement the agreements by configuring BGP peering sessions to steer traffic accordingly. A route server (RS) [65] can be used to help members 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 connected members.

Interconnection agreement models. The current models can be grouped in two types of business relationships: *transit* and *peering*. A transit agreement is the one where an ISP provides connectivity to the entire Internet and charges its customers on peak-hour traffic (e.g., through the 95th-percentile [33]) basis. A variant of this model, called *partial transit*, provides limited reachability (just a subset of routes available) for a lower price. In contrast, in peering agreements, two ASes agree to reciprocally exchange traffic originated/destined from/to their networks or their cone of customers [56]. Peering interconnections can be settlement-free or paid, depending on who benefits the most from the agreement, e.g., due to traffic imbalance or route diversity from a larger ISP.

Interconnection process. The process of establishing interconnection agreements starts with the search for potential peering partners. As of today, it heavily relies on personal relationships and brand recognition. To foster the process, entities (e.g., IXPs, RIRs) organize periodic face-to-face meetings. Next, the operators involved discuss the properties (e.g., SLA, pricing) of the agreement. Then, depending on the interconnection model, the legal departments of ASes get involved in formalizing the terms of the agreement. In general, settlement-free peering agreements tend not to be legally formalized [73]. Finally, operators configure their border routers and start exchanging traffic.

3 COMPELLING APPLICATIONS

Is reducing the interconnection setup time significant? If, so why? What are realistic, compelling use cases that benefit from fast agreement establishment? To answer these questions, we interviewed several network operators and used the resulting feedback to prepare a survey [57] carefully to be circulated among a large number of network operators.

Our survey collected over 100 unique responses, of which 56% reported a score of 4 or 5 (on a scale from 1 to 5) on the relevance of reducing the agreement setup time. In fact, roughly 6%, 29%, 36%, 25%, and 4% of the network operators reported times to establish agreements in the order of hours, days, weeks, months, and years, respectively.

Further, our survey aimed at identifying the most relevant applications considered by peering coordinators and network operators, should they be able to benefit from reduced interconnection times. The preliminary interviews were used to determine a broader set of use cases. The survey indicated the most relevant ones: enhanced traffic engineering, improved resource utilization, new economic opportunities, and ordering network services on-demand. No other aspect was mentioned more than once by survey respondents.

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 restricting our focus to ISP operators only, 72% of the respondents deem this use case as an essential one.

Increasing peering port utilization. This has been indicated as a relevant improvement by the majority (60%) of the survey respondents. Likewise, Deutsche Telekom recently reported that increasing resource utilization by 1% or 2% could result in saving millions of dollars in future infrastructure investments [23]. The ability to establish interconnection agreements in short time frames especially helps in the following cases. First, it reduces the time until a new, recently deployed port starts being used. 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 [54]. In the second example, consider an eyeball network 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. In fact, customers can save money in case they are billed by their (congested) transit provider 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. Instead, providers 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 [59, 61]. 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 IXPs quickly. Operators can also order direct access to cloud infrastructures, and to network analytics solutions.

4 ENABLING CONDITIONS

There are two conditions for deploying the compelling applications presented: the involved networks need physical connectivity and an underutilized link between them. It is well known that IXPs such as AMS-IX and DE-CIX interconnect large numbers of ASes (700+) [6, 8]. With respect to the underutilized link, previous studies have hinted the existence of spare capacity [27, 35, 36], but analyzed small or heavily aggregated datasets. To ascertain this condition, we look at the spare port capacity available in two commercial IXPs: IXP-EU, one of the largest IXPs worldwide located in Europe and IXP-LA, a medium-sized IXP situated in Latin America, transporting high volumes of traffic per second among hundreds of members (over 5 Tbps and 100 Gbps, resp.). 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 this section, we verify whether the enabling conditions are present, by answering the following questions: (*i*) how much spare capacity do IXP ports typically have? (*ii*) How does the availability of spare capacity have changed over the years?

Datasets. We collected datasets consisting of traffic traces from the two IXPs containing source and destination MAC addresses. In IXP-LA, we captured flow summaries during 12 months (mid-October 2015 to mid-October 2016) with a sample rate of 1:32768 packets. In IXP-EU, we captured flow summaries during a total of 9 week-long periods in 2016 and 2017 sampled with rate 1:10000 packets. In both IXPs the measurements are aggregated in 5-minute intervals.

Assumptions. To reason about the port spare capacity availability, we make the following assumptions. First, port capacity is the highest observed *peak* (ingress or egress, 5-min average) utilization. We need the premise because the information about specific provisioned capacity is too sensitive and not part of our dataset. Second, a port is deemed "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 [38], much smaller than, for instance, those observed in data center networks.

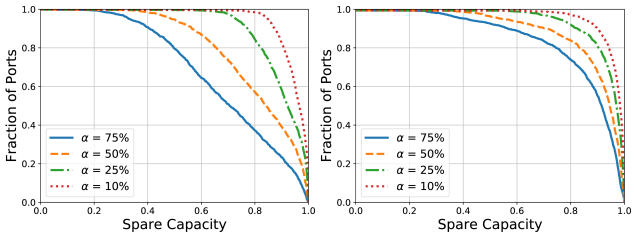
Methodology. We perform two analyses. The first one determines how much spare capacity is available at the ports of IXP-LA; the second, based on the IXP-EU dataset, looks at the availability of spare capacity over the years. In both cases, we estimate the port capacity availability using a *spare capacity* metric, denoted as s and defined as follows.

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 maximum fraction of the port capacity that is available for a fraction α of the time, i.e., there exists a fraction α of the measurements in R where $1 - \frac{R_i}{peak}$ 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.

How much spare capacity do IXP ports have? Figure 3 shows the spare capacity for different values of α for the ports present at IXP-LA. To avoid over-estimating the ports' capacities, we divide the dataset into twelve windows each of one-month length. This leads to up to twelve (see assumptions) different port capacity estimations instead of a single one spanning a year. As the ports are full-duplex, we present the ingress and the egress utilization ratio for each port.[2]

IXP-LA presents a significant amount of spare capacity in both directions. For example, 60% of the ingress ports (Fig. 3(a)) have at least 78% of spare capacity during 50% of the time. As for the egress direction (Fig. 3(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. The high spare capacity ratio of the IXP ports indicates that conditions to deploy Dynam-IX are favorable.

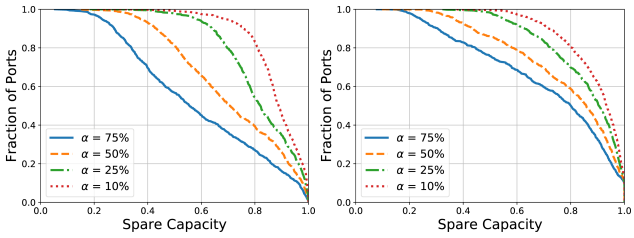
[2]The direction of the traffic is considered taking the AS as the reference.



(a) Ingress ports (b) Egress ports

Figure 3: Monthly spare capacity for IXP-LA.

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. By performing a longitudinal analysis, we note a consistent pattern of spare capacity at both ingress and egress ports. Moreover, the behavior is preserved 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). Considering this, we zoom into the spare port capacity of a single week of 2017 (Fig. 4). We observe that during 50% of the time about 60% of the ingress ports have (Fig. 4(a)) at least 63% of spare capacity. For egress ports, in turn, 60% show 80% of available resources (Fig. 4(b)).



(a) Ingress ports (b) Egress ports

Figure 4: Weekly spare capacity for IXP-EU.

Discussion. The availability of spare capacity in both datasets hints that the necessary conditions to improve wide-area traffic delivery performance by exploiting the rich connectivity of IXPs exist. 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*) IXPs 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 its traffic. Case (*iii*) would require network operators to make planned decisions to avoid negatively impacting the traffic delivery. Providing solutions to decide whether or not to establish an interconnection agreement and what AS is the appropriate peering partner are out of the scope of the paper.

5 DYNAM-IX

The goal of our work is to enable network operators to improve wide-area traffic delivery performance by leveraging the rich connectivity diversity at IXPs. We first identify the requirements to achieve this goal and then present the design of Dynam-IX and its components. Finally, we discuss the practical aspects and limitations of our approach.

5.1 Requirements

The underlying requirement for any practical approach to inter-domain routing is adoption. To facilitate adoption, we design our solution to complement the existing practices in the area. This leads 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**: no information considered private about an interconnection agreement (e.g., business policies, interconnection terms) should be leaked to unauthorized parties. Our survey shows that network operators are reluctant to sharing interconnection policy-related information with third parties [57], thus confirming the findings in [25];
- **mechanism to build trust**: network operators should be able to identify partners deemed reliable (by the community) systematically. In fact, while today's 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 IXPs.

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

5.2 Design Choices

A straightforward approach to our goal would be to provision IXPs 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 the 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. To preserve confidential information, the service could be engineered to guarantee strong security properties (e.g., using secure multi-party computations [43] or trusted execution environments such as Intel SGX [30]). However, this raises the complexity of the solution and incurs processing overheads. This approach also introduces a third-party service whose availability and impartiality ASes need to depend on.

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 abstraction provides an expressive 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 does not require confidential information. However, we argue that a centralized solution is not practical because of market incentives. In fact, IXPs are disincentivized to interfere in the business decisions of ASes, which are customers of IXPs. Moreover, IXPs would face the burden of dealing with disputes should ASes question the information collected by the IXP-operated mechanism. Our contacts in the IXP operation community confirmed these concerns make a centralized solution offered by IXPs 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 to tamper with the information to gain a benefit or to harm another AS. We now detail each of the Dynam-IX components (Figure 2) in the following subsections.

5.3 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, port of the Dynam-IX peer, public key, and a description of the services being offered. For the sake of detailing the protocol, we assume here that such information is stored on the ledger; however, this is not mandatory, and other storage systems can be used for this specific information (see §5.7). The ledger also contains a score about the past performance of each Dynam-IX member as a customer and a provider of an interconnection agreement (§5.6). After connecting to the Dynam-IX ledger, the AS can start using the protocol. To illustrate how the protocol works, we use the example presented in Section 1, where an AS *A* is facing a traffic surge towards *C* and congesting the traffic going to *D*. Figure 5 illustrates all the protocol steps.[3]

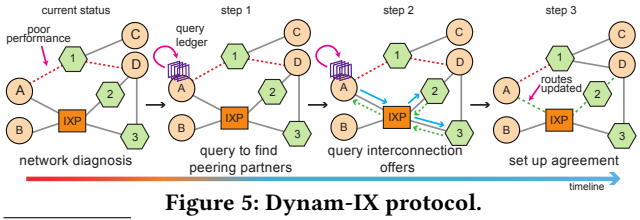


Figure 5: Dynam-IX protocol.

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

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 basic 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 interconnect (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 confidentiality, any communication between a customer and a provider is encrypted using standard SSL 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 be delegated to a human, and be based on both the provider's business policy and any information available on the ledger. In either case, our approach provides a structure for the negotiation process. Assume the provider does answer. The provider matches the customer's request against its interconnection intents and, if a valid match exists, it composes an interconnection offer that is sent to the customer. Otherwise, the customer is notified that no match exists. Offers are digitally signed so their authenticity can be verified 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 §5.5), digitally signs it, and sends it to the customer. The customer verifies the provider's signature and the 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's signature and, assuming it is valid, proceeds to register the interconnection agreement on the ledger (see §5.6). Once the information is stored 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 score reflecting their experience (§5.6). This information is used as the basis to compute an overall score rank for each participating AS to help ASes decide whether or not a network is reliable to interconnect.

5.4 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

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)

Table 1: Summary of Intent Abstraction Attributes.

with an interconnection offer. To design the intention abstraction on practical grounds, we interviewed peering coordinators and network operators, as well as conducted a survey, asking which parameters were considered when establishing interconnection agreements. We note that all parameters mentioned by three or more operators (among 100+) were included in the abstraction. We observe that more parameters can be added 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.

```
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 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. This type of billing method is currently used by several real-world connectivity providers, e.g., Hopus [51]. 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 defined as an integer multiple of one hour.

Specifying prices as functions of time or bandwidth. Several connectivity providers (e.g., [5]) specify their (ingress/egress) per-unit costs as a function of the bandwidth and the length of the interconnection agreements [51]. In fact, 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, multiple intents that share common properties can be grouped as 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). This can be accommodated by replacing the value in the attribute by the *wildcard* character '*'.

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 query will be sent. 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). The unspecified properties are not considered 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.

5.5 Lightweight Legal Framework

Discussions among legal offices and lawyers can represent a crucial phase before an agreement can be established. Formal terms and conditions should be carefully stated to legally protect parties in possible future disputes, which are not uncommon in the Internet ecosystem [20]. Our survey findings reveal that, for 56% of the surveyed networks, legal matters are settled within hours (19%), days (37%), weeks (30%), months (10%), or even years (4%). In contrast, response times for Dynam-IX are generally in the order of seconds (see §6).

Although operators are more open to handshake agreements [73], 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 an interconnection agreement is established.

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's 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's 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.

The general contract can be reviewed and updated by the members of Dynam-IX at any time. In such 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.

5.6 Tamper-proof Distributed Ledger

Dynam-IX uses a distributed tamper-proof ledger where the ASes store historical performance information about past interconnection agreements. Once two ASes establish an agreement, a mutually digitally signed piece of information is stored on the distributed ledger to indicate that an agreement has been settled. This piece of information is necessary to associate performance scores with valid interconnection agreements only. Each record is a 5-tuple containing: a unique identifier of the agreement; the ASNs of the two involved networks; and two attributes to control the update of the score of corresponding networks.

Each AS participating in Dynam-IX has two scores, for customer and provider actions, stored on the ledger. The customer score indicates if the AS is a "good player", while the provider one indicates the "quality" of the service offered to customers. By querying the ledger, an AS can verify the scores of another AS instead of only relying on personal relationships and brand image. Initially, ASes do not have any score information associated with them. Not having a score does not mean an AS cannot be trusted. Scores are initialized as soon as ASes start establishing interconnection agreements.

In Dynam-IX, ASes build trust scores of other ASes in two ways: *i)* relying on the aggregated per-AS scores automatically computed and stored on the ledger or *ii)* locally aggregating the individual per-agreements scores.

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 [14].

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 no more than one score per scoring period is sent. 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 [22], 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 own level of trust with respect to these scores. This approach is allowed in Dynam-IX but requires more efforts on the AS side to specify its trust policies.

5.7 Practical Considerations

Connecting to the distributed ledger. Only ASes that are members of the IXP are allowed to join the Dynam-IX ledger. The admission process can be performed by requiring that all ASes connect to a specific local network at the IXP, relying on the IXP 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 IXP to perform the admission), ASes already connected to the ledger can continue benefiting from Dynam-IX.

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 an interconnection agreement is established, the ASes 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 resources by the IXP is not mandatory in Dynam-IX as networks can already reap the benefits of faster and rich interconnectivity regardless of such provisioning.

Incentives. Dynam-IX offers incentives for the different types of ASes connected to the IXP. *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 IXPs may indirectly profit from our solution, increasing their revenue, since ASes may be attracted and also connect to the IXP.

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 [7], Telia Company AB (ASN1299) has more than 250 thousand prefixes in its cone (the set of ASes an AS can reach using customer links [56]), 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 [11]), we envision that BGP updates can be automatically parsed to intents whose SLA parameters are provided by network monitoring tools, requiring from the operators that they only specify profiles and associate them with the intents.

BGP routing stability. We are aware that 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. Recent work shows that the Internet's routing system is by and large resilient [40] and future research should address its robustness (similar to SWIFT [50]) to solve deficiencies not inherent to Dynam-IX.

5.8 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 basic measure is for a customer to send new queries with higher requirements until the desired target price strikes. Remember that Dynam-IX goal is to complement the existing practices. Thus, Dynam-IX could ease to quickly identify reliable partners in an automated manner through short-term agreements. Any further price negotiation could then be conducted by humans.

Legal framework. 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 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).

Intent abstraction attributes. Dynam-IX intent abstraction design may not be expressive enough to support all existing agreements. We argue that i) our intent abstraction may easily be extended in the future and ii) Dynam-IX complements existing human-based processes and does not replace them.

6 EVALUATION

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

Implementation. We built a prototype[4] of Dynam-IX using Hyperledger Fabric 1.0.5 (HLF) [17], a permissioned blockchain, as the distributed tamper-proof ledger.[5] The prototype was developed in Python, Node.js, and Go in approximately 1200 lines. 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 and agreements and, updating the ledger are implemented using smart contracts. A recent study [69] shows that Hyperledger Fabric is capable of achieving more than 10k transactions per second, well beyond our needs in an IXP 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 [26].

6.1 How long does it take to establish an interconnection agreement?

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 evaluation, we assumed that the provider will always reply to a query with an offer and that the ASes' contact information

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

[5]We note that Dynam-IX can be instantiated with any other implementation of a tamper-proof distributed ledger.

is stored on the ledger, which in practice may not always be the case (see §5). As queries on the ledger are local, the overhead of this operation is negligible. The establishment time is measured from the moment an AS sends an interconnection proposal (based on an offer from a provider) to the moment the agreement related information is published on the ledger.

We first determine the limits of Dynam-IX with a throughput test: N ASes flooding a single AS with queries and establishing interconnection agreements proposals. Such a case 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 different AS offering connectivity towards p.

We evaluate this scenario using up to 200 AWS EC2 cloud instances [16], each 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). In addition, 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 §5.3) 30 times.

Figure 6(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 Dynam-IX 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 provision an agreement in less than 60 seconds after a network orders it [58].

The response time of each query/proposal is approximately constant, as observed in Figure 6(b) and Figure 6(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.

6.2 How fast does the ledger grow?

Every Dynam-IX peer keeps a local copy of the ledger. 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, update the reliability score of an AS, 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 reliability score updates (2 per agreement, 1 for the customer and 1 for theprovider), and 250 AS information updates (e.g., public key).

In our experiments, the growth of the ledger depends on the number of transactions that are grouped in each block of the blockchain (the higher the transactions per block, the higher the storage saving). 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 [26]. 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. 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).

Figure 7 presents the results (in log scale). The worst case relates to a scenario where only one transaction is stored in each block (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.

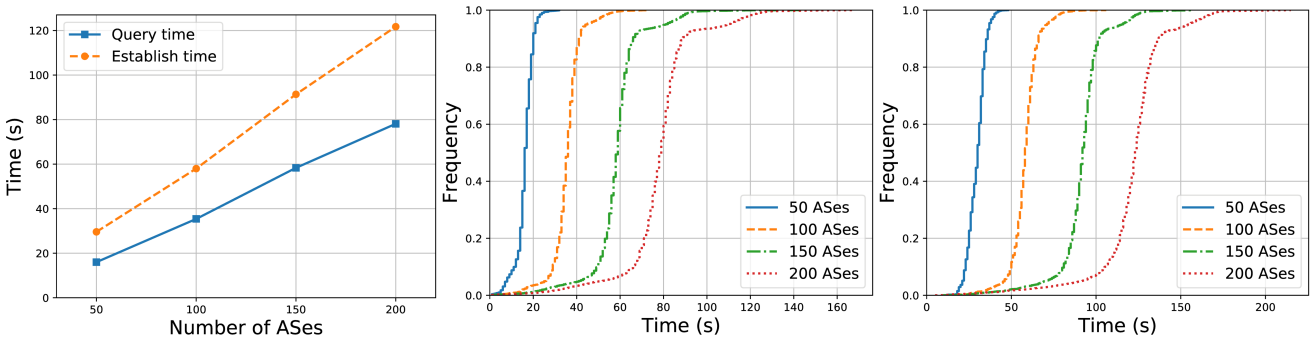
We evaluate the blockchain growth for three different block creation timeouts (15, 30, and 60 seconds) and two transaction rates, 1 (the minimum number reported in [26]) and 4 per second. The number of cases is the combination of timeouts and rates (e.g., with 4 transactions per second and a 30s timeout we have 120 transactions per block - TPB). As expected, the size of the blockchain grows more slowly with longer timeouts. Specifically, with a timeout of 60 seconds, the size of the ledger after 10k transactions is between 21.32 MB and 31.44 MB, for 1 and 4 transactions per second, respectively.

To further illustrate, consider the same scenario with 4 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 an IXP with 1500 members (size of the largest one in terms of members [10]) and a period of 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.

6.3 What are the bandwidth requirements?

Dynam-IX is designed to operate alongside the infrastructure of an IXP. This 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 aforementioned experiments) as reported by Amazon CloudWatch [15]. 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.

The traffic related to the regular ASes is approximately the same in all experiments, with individual peaks of 4.8 Mbps (ingress traffic) and 0.8 Mbps (egress traffic). The AS receiving all the requests



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

Figure 6: Response time for different number of ASes.

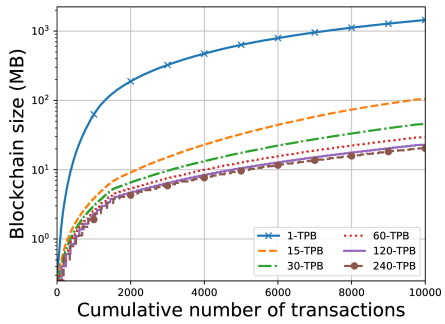


Figure 7: Ledger size for different configurations.

presents egress peaks between 9.3 Mbps and 12.4 Mbps (for experiments with 50 and 200 ASes respectively), and ingress peaks between 8.1 Mbps and 9.4 Mbps likewise. 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. 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 IXP 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 IXPs such as DE-CIX [8] and AMS-IX [6] carry an aggregated traffic of more than 5 Tbps, the demands of Dynam-IX from the IXP 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 IXPs. We note that the scenario of the throughput experiment describes an extreme case and in regular conditions the bandwidth requirements tend to be much smaller than 7 Gbps of traffic for 1500 ASes.

7 RELATED WORK

To innovate the interconnection ecosystem, related work mainly seeks to improve inter-domain routing, the interconnection negotiation and set up process, or the interconnection location (i.e., IXPs). We survey them below.

Evolving 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 [52], BGP extended [64], the use of large communities [49] to carry (more) meta-information, and BGP session culling [48], 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 [66], researchers, and engineers try to overcome the limitations with external systems. Edge Fabric [68] and Espresso [74] 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 IXP.

Reduced interconnection agreement setup time. More related to our work, MINT [70], ChoiceNet [71], and RouteBazaar [24] are previous efforts from the research community. They discuss the concept of marketplaces aiming to provide alternatives for reducing the interconnection negotiation and setup time. They highlight potential benefits in composing end-to-end paths to accommodate their inter-domain traffic. MINT presents a high-level system design where through a centralized entity ASes can advertise path segments and query for end-to-end paths, and the centralized entity is responsible for matching offers and demands. ChoiceNet follows the same principle (using a centralized entity) but differs on the matching that is performed by the ASes. Both proposals insert a new entity in the process of establishing interconnection agreements and expose the business policies of the participant networks to the other members and the marketplace operator. Such approaches also do not offer a method to build trust among the operators. RouteBazaar discusses (does not implement) the use of a public blockchain (public ledger) through which ASes can advertise pathlets [42] and use them to compose end-to-end paths. While this approach removes the centralized entity, it still exposes the

interconnection policies of the ASes. Differently from MINT and ChoiceNet, RouteBazaar discusses alternatives to provide information to help ASes decide whether or not to interconnect with another AS. Such method, however, leaks information about the interconnection agreement properties. Dynam-IX instead offers the necessary components to improve wide-area traffic delivery performance in a privacy-preserving manner.

Interconnection companies such as Megaport [58], Packet Fabric [62], Epsilon [34], and ConsoleConnect [29] offer on-demand connectivity to cloud providers (e.g., Amazon AWS, Google GCC, MS Azure) for networks connected to their Points of Presence (PoPs). The use of marketplaces or brokers has been considered as well in the context of CDNs [60]. While these approaches could easily be extended to allow any two networks to establish on-demand agreements, they fail to guarantee business policy confidentiality and to offer a mechanism to build trust – requirements that Dynam-IX respects. In fact, Dynam-IX is not a competitor of such companies, but a way to offer more flexible connectivity for their customers without learning about their interconnection policies.

IXPs as service enablers. Due to their nature as convergence points of hundreds of ASes, IXPs have been advocated as places to spur innovation and promote new services for network operators. Control eXchange Point [53] 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) [47] and a number of refinements and extensions [18, 27, 46] 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 [32] and allow the use of economic aspects in the policy configuration at IXPs [45]. These proposals can work together with Dynam-IX and enable network operators to optimize route configuration after the establishment of the interconnection agreement.

8 CONCLUSIONS

Dynam-IX is a step towards a more dynamic interconnection ecosystem. By providing 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 IXPs. Among the potential benefits are more responsiveness to the traffic dynamics, increased utilization of peering ports, and new economic opportunities. We 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 IXP infrastructure. Currently, we are working with a large international interconnection facility to offer Dynam-IX for its connected ASes. Through this deployment, we plan to investigate how the ASes will benefit from Dynam-IX and its impact on the traffic patterns and spare capacity at the IXP. We also plan to improve the way ASes build trust by replacing subjective scores with objective information that can be verified. We envision that blockchain *smart contracts* [72] can be combined with forwarding performance verification algorithms [19] to produce objective and verifiable scores for ASes.

ACKNOWLEDGEMENTS

We thank the anonymous reviewers and our shepherd, Cristel Pelsser, for their valuable feedback on our paper. We are also thankful to Leandro Bertholdo, Raul Sejas, Philippe Duguet, Eric Loos, Ankit Singla, Ignacio Castro, Josh Bailey, and Nikolaos Laoutaris for their excellent feedback and discussions that helped to improve our work. We are also thankful to all network operators and peering coordinators for taking part in our survey. This research is (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.

REFERENCES

[1] ios 5 update causes massive internet traffic spike - to users' frustration, 2011. Available at https://www.theguardian.com/technology/2011/oct/13/ios-5-update-internet-traffic-spike.
[2] ios 7 downloads consumed 20 percent of an isp's traffic on release day, 2013. Available at https://arstechnica.com/information-technology/2013/11/ios-7-downloads-consumed-20-percent-of-an-isps-traffic-on-release-day/.
[3] 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.
[4] 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.
[5] Hopus - the routed exchange, 2017. Available at http://hopus.net.
[6] AMS-IX, 2018. Available at https://ams-ix.net.
[7] Caida as-rank, 2018. Available at http://as-rank.caida.org.
[8] DE-CIX, 2018. Available at https://www.de-cix.net.
[9] Internet Exchange Map, 2018. Available at https://www.internetexchangemap.com/.
[10] São Paulo IXP members list, 2018. Available at http://ix.br/particip/sp.
[11] I. 6connect. IRR Power Tools – A utility for managing Internet Routing Registry (IRR) filters, 2018. Available at https://github.com/6connect/irrpt.
[12] B. Ager, N. Chatzis, A. Feldmann, N. Sarrar, S. Uhlig, and W. Willinger. Anatomy of a large european ixp. In *SIGCOMM '12*, 2012.
[13] A. Ahmed, Z. Shafiq, H. Bedi, and A. Khakpour. Peering vs. Transit: Performance Comparison of Peering and Transit Interconnections. In *IEEE ICNP*, 2017.
[14] Y. Alowayed, M. Canini, P. Marcos, M. Chiesa, and M. Barcellos. Picking a partner: A fair blockchain based scoring protocol for autonomous systems. In *Proceedings of the Applied Networking Research Workshop*, ANRW '18, pages 33–39, New York, NY, USA, 2018. ACM.
[15] Amazon. Amazon cloudwatch, 2018. Available at https://aws.amazon.com/cloudwatch.
[16] Amazon. Amazon ec2, 2018. Available at https://aws.amazon.com/ec2/.
[17] E. Androulaki et al. Hyperledger fabric: A distributed operating system for permissioned blockchains. In *Proceedings of the Thirteenth EuroSys Conference*, EuroSys '18, 2018.
[18] G. Antichi, I. Castro, M. Chiesa, E. L. Fernandes, R. Lapeyrade, D. Kopp, J. H. Han, M. Bruyere, C. Dietzel, M. Gusat, et al. Endeavour: A scalable sdn architecture for real-world ixps. *IEEE Journal on Selected Areas in Communications*, 2017.
[19] K. Argyraki, P. Maniatis, and A. Singla. Verifiable network-performance measurements. In *CoNEXT '10*, 2010.
[20] S. Bafna, A. Pandey, and K. Verma. Anatomy of the internet peering disputes. *CoRR*, abs/1409.6526, 2014.
[21] N. Bakker, E. Jasinska, R. Raszuk, and N. Hilliard. Internet Exchange BGP Route Server (RFC7947). 2016. Available at https://tools.ietf.org/html/rfc7947.
[22] M. Blum. Coin Flipping by Telephone a Protocol for Solving Impossible Problems. *ACM SIGACT News*, 15(1), 1983.
[23] F. Bornstaedt. New levels of cooperation between eyeball isps and ott/cdns, 2017. Available at https://ripe75.ripe.net/archives/video/126/ – starting at 7min50s.
[24] I. Castro, A. Panda, B. Raghavan, S. Shenker, and S. Gorinsky. Route bazaar: Automatic interdomain contract negotiation. In *USENIX HotOS 2015*.
[25] M. Chiesa, D. Demmler, M. Canini, M. Schapira, and T. Schneider. Internet routing privacy survey, 2017. Available at https://six-pack.bitbucket.io/media/privacy-survey-2017.pdf.
[26] M. Chiesa, D. Demmler, M. Canini, M. Schapira, and T. Schneider. Sixpack: Securing internet exchange points against curious onlookers. In *CoNEXT '17*,

2017.
[27] M. Chiesa, C. Dietzel, G. Antichi, M. Bruyere, I. Castro, M. Gusat, T. King, A. W. Moore, T. D. Nguyen, P. Owezarski, S. Uhlig, and M. Canini. Inter-domain networking innovation on steroids: empowering ixps with SDN capabilities. *IEEE Communications Magazine*, 2016.
[28] Y.-C. Chiu, B. Schlinker, A. B. Radhakrishnan, E. Katz-Bassett, and R. Govindan. Are We One Hop Away from a Better Internet? In *IMC 2015*.
[29] Console. Console - the cloud connection company, 2017. Available at https://www.consoleconnect.com/.
[30] V. Costan and S. Devadas. Intel SGX explained. Cryptology ePrint Archive, Report 2016/086, 2016. http://ia.cr/2016/086.
[31] A. Dhamdhere and C. Dovrolis. The internet is flat: Modeling the transition from a transit hierarchy to a peering mesh. In *CoNEXT 2010*.
[32] C. Dietzel, G. Antichi, I. Castro, E. L. Fernandes, M. Chiesa, and D. Kopp. Sdn-enabled traffic engineering and advanced blackholing at ixps. In *ACM SOSR'17*, 2017.
[33] X. Dimitropoulos, P. Hurley, A. Kind, and M. P. Stoecklin. On the 95-percentile billing method. In *International Conference on Passive and Active Network Measurement*, pages 207–216. Springer, 2009.
[34] Epsilon. Epsilon telecommunications limited – connectivity made simple, 2017. Available at www.epsilontel.com/.
[35] R. Fanou, F. Valera, and A. Dhamdhere. Investigating the causes of congestion on the african ixp substrate. In *IMC '17*, 2017.
[36] N. Feamster. Revealing utilization at internet interconnection points. 2016. Available at http://arxiv.org/abs/1603.03656.
[37] P. Ferreira, J. Mindel, and L. McKnight. Why bandwidth trading markets have not matured? analysis of technological and market issues. *International Journal of Technology, Management and Policy*, 2, 2004.
[38] C. Fraleigh, F. A. Tobagi, and C. Diot. Provisioning IP backbone networks to support latency sensitive traffic. In *INFOCOM 2003*, pages 375–385, 2003.
[39] P. C. Fusaro and R. M. Miller. *What went wrong at Enron: Everyone's guide to the largest bankruptcy in US history*. John Wiley & Sons, 2002.
[40] V. Giotsas, C. Dietzel, G. Smaragdakis, A. Feldmann, A. Berger, and E. Aben. Detecting peering infrastructure outages in the wild. In *ACM SIGCOMM'17*, 2017.
[41] E. Giovannetti and C. A. Ristuccia. Estimating market power in the Internet backbone. Using the IP transit Band-X database. *Telecommunications Policy*, 2005.
[42] P. B. Godfrey, I. Ganichev, S. Shenker, and I. Stoica. Pathlet routing. In *ACM SIGCOMM '09*, 2009.
[43] O. Goldreich, S. Micali, and A. Wigderson. How to Play any Mental Game or A Completeness Theorem for Protocols with Honest Majority. In *STOC'87*, pages 218–229, 1987.
[44] K. Goncharoff. Bandwidth's new bargaineers, 1998. Available at https://www.technologyreview.com/s/400275/bandwidths-new-bargaineers/.
[45] J. Griffioen, T. Wolf, and K. L. Calvert. A Coin-Operated Software-Defined Exchange. In *ICCCN 2016*.
[46] A. Gupta, R. MacDavid, R. Birkner, M. Canini, N. Feamster, J. Rexford, and L. Vanbever. An industrial-scale software defined internet exchange point. In *NSDI'16*.
[47] A. Gupta, L. Vanbever, M. Shahbaz, S. P. Donovan, B. Schlinker, N. Feamster, J. Rexford, S. Shenker, R. J. Clark, and E. Katz-Bassett. Sdx: a software defined internet exchange. In *SIGCOMM*, 2014.
[48] W. Hargrave, M. Griswold, J. Snijders, and N. Hilliard. 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.
[49] J. Heitz, J. Snijders, K. Patel, I. Bagdonas, and N. Hilliard. Bgp large communities attribute (rfc8092). 2017. Available athttps://tools.ietf.org/html/rfc8092.
[50] T. Holterbach, S. Vissicchio, A. Dainotti, and L. Vanbever. Swift: Predictive fast reroute. In *ACM SIGCOMM '17*, 2017.
[51] Hopus. Pricing scheme, 2018. Available at http://hopus.net/price.
[52] D. Katz, R. Chandra, Y. Rekhter, and T. Bates. Multiprotocol extensions for bgp-4 (rfc4760, updated by 7606). 2007.
[53] V. Kotronis, R. Kloti, M. Rost, P. Georgopoulos, B. Ager, S. Schmid, and X. Dimitropoulos. Stitching inter-domain paths over ixps. In *Proceedings of the Symposium on SDN Research*, 2016.
[54] A. Lodhi, N. Laoutaris, A. Dhamdhere, and C. Dovrolis. Complexities in internet peering: Understanding the "black" in the "black art". In *INFOCOM 2015*.
[55] A. Lodhi, N. Larson, A. Dhamdhere, C. Dovrolis, and k. claffy. Using peeringdb to understand the peering ecosystem. *SIGCOMM Comput. Commun. Rev.*, 44(2):20–27, Apr. 2014.
[56] M. Luckie, B. Huffaker, A. Dhamdhere, V. Giotsas, and k. claffy. As relationships, customer cones, and validation. In *ACM IMC '13*, 2013.
[57] P. Marcos, M. Chiesa, L. Muller, P. Kathiravelu, C. Dietzel, M. Canini, and M. Barcellos. Internet interconnection ecosystem survey, 2018. Available at https://dynam-ix.github.io/docs/internet_interconnection_ecosystem_survey.pdf.
[58] Megaport. Megaport - a better way to connect, 2017. Available at http://megaport.com/.
[59] C. Morales. Netscout arbor confirms 1.7 tbps ddos attack; the terabit attack era is upon us, 2018. https://asert.arbornetworks.com/netscout-arbor-confirms-1-7-tbps-ddos-attack-terabit-attack-era-upon-us/.
[60] M. K. Mukerjee, I. N. Bozkurt, D. Ray, B. M. Maggs, S. Seshan, and H. Zhang. Redesigning cdn-broker interactions for improved content delivery. In *CoNEXT*, 2017.
[61] L. H. Newman. Github survived the biggest ddos attack ever recorded, 2018. https://www.wired.com/story/github-ddos-memcached/.
[62] PacketFabric. Packetfabric, 2017. Available at https://www.packetfabric.com/.
[63] Peeringdb. Peeringdb, 2018. Available at https://www.peeringdb.com.
[64] Y. Rekhter and S. R. Sangli. Bgp extended communities attribute (rfc4360, updated by 7153, 7606). 2006. Available athttps://tools.ietf.org/html/rfc4360.
[65] P. Richter, G. Smaragdakis, A. Feldmann, N. Chatzis, J. Boettger, and W. Willinger. Peering at peerings: On the role of ixp route servers. In *ACM IMC '14*, 2014.
[66] R. R. Sambasivan, D. Tran-Lam, A. Akella, and P. Steenkiste. Bootstrapping evolvability for inter-domain routing with d-bgp. In *ACM SIGCOMM '17*, 2017.
[67] Sandvine. FIFA 16 - The Beautiful Game?, 2015. Available at http://www.internetphenomena.com/2015/09/fifa-16-the-beautiful-game/.
[68] B. Schlinker, H. Kim, T. Cui, E. Katz-Bassett, H. V. Madhyastha, I. Cunha, J. Quinn, S. Hasan, P. Lapukhov, and H. Zeng. Engineering egress with edge fabric: Steering oceans of content to the world. In *ACM SIGCOMM '17*, 2017.
[69] J. Sousa, A. Bessani, and M. Vukolic. A byzantine fault-tolerant ordering service for the hyperledger fabric blockchain platform. *CoRR*, abs/1709.06921, 2017. Available at http://arxiv.org/abs/1709.06921.
[70] V. Valancius, N. Feamster, R. Johari, and V. Vazirani. MINT: A Market for INternet Transit. In *ReArch'08*.
[71] T. Wolf, J. Griffioen, K. L. Calvert, R. Dutta, G. N. Rouskas, I. Baldin, and A. Nagurney. Choicenet: Toward an economy plane for the internet. *SIGCOMM Comput. Commun. Rev.*, 2014.
[72] G. Wood. Ethereum: A secure decentralised generalised transaction ledger. *Ethereum Project Yellow Paper*, 2014.
[73] B. Woodcock and M. Frigino. 2016 survey of internet carrier interconnection agreements. *Packet Clearing House, November*, 2016.
[74] K.-K. Yap et al. Taking the edge off with espresso: Scale, reliability and programmability for global internet peering. In *ACM SIGCOMM '17*, 2017.