A Protocol-Ignorance Perspective on Incremental Deployability of Routing Protocols

Vadim Kirilin
IMDEA Networks Institute and UC3M, Spain

Sergey Gorinsky
IMDEA Networks Institute, Spain

Abstract*—New protocols for Internet inter-domain routing struggle to get widely adopted. Because the Internet consists of more than 50,000 autonomous systems (ASes), deployment of a new routing protocol has to be incremental. In this work, we study such incremental deployment. We first formulate the routing problem in regard to a metric of routing cost. Then, the paper proposes and rigorously defines a statistical notion of protocol ignorance that quantifies the inability of a routing protocol to accurately determine routing prices with respect to the metric of interest. The proposed protocol-ignorance model of a routing protocol is fairly generic and can be applied to routing in both inter-domain and intra-domain settings, as well as to transportation and other types of networks. Our model of protocol deployment makes our study specific to Internet inter-domain routing. Through a combination of mathematical analysis and simulation, we demonstrate that the benefits from adopting a new inter-domain protocol accumulate smoothly during its incremental deployment. In particular, the simulation shows that decreasing the routing price by 25% requires between 43% and 53% of all nodes to adopt the new protocol. Our findings elucidate the deployment struggle of new inter-domain routing protocols and indicate that wide deployment of such a protocol necessitates involving a large number of relevant ASes into a coordinated effort to adopt the new protocol.

I. Introduction

For a long time, Border Gateway Protocol (BGP) [1] has remained the only prominent protocol in inter-domain routing practice. Based on information propagated by neighboring nodes, a BGP node decides which paths it uses and which routing information it shares with neighbors. The local selection and filtering of path announcements by each node provides the node with means to realize its political, economic, and security policies.

On the flip side, BGP suffers from serious problems inherent in its design concept. The local filtering results in unintended information hiding, which artificially reduces the diversity of usable paths. BGP is vulnerable to hijacking [2]. Other problems include slow convergence and lack of support for multipath routing or end-to-end Quality of Service (QoS). While refinements of BGP mitigate some of its weaknesses [3]–[5], a thorough solution to the BGP problems requires changes in the conceptual design.

Radically different designs for inter-domain routing have been proposed and yet not deployed widely [6]–[8]. For example, Route Bazaar is a blockchain-inspired approach that uses a public ledger to announce, select, and verify end-to-end QoS-aware routing in a privacy-preserving manner [8]. Such solutions empower an autonomous system (AS) to not only enforce its security, political, and economic policies but also obtain flexible secure routing based on global information. Despite the promise of significant improvements, the new inter-domain protocols fail to get widely deployed.

ISBN 978-3-903176-08-9

This paper studies the problem of deploying a new inter-domain routing protocol. With the Internet composed by more than 50,000 ASes, replacement of BGP with a new protocol has to be incremental because it is virtually impossible for all the ASes to agree on simultaneously adopting the new protocol on a flag day. Moreover, the benefits of partially deploying the new protocol have to be significant compared to the extra hardware, staff training, and other expenses incurred by the protocol adopters. We develop a rigorous mathematical approach to understand how the extent of partial deployment affects the amount of the benefits realized by the deployment. The sought understanding is of practical importance due to its potential to both explain the deployment struggle of new inter-domain routing protocols and guide a successful deployment for such a protocol.

The cornerstone of our approach is a novel model of a routing protocol, which is equally applicable to inter-domain and intra-domain protocols. The model centers on the ability of a routing protocol to solve the routing problem with respect to a metric of routing cost. The metric of interest can be end-to-end path latency, monetary cost of traffic transit through other nodes, consumed network capacity, or a hybrid of multiple simple metrics. Given a global set of paths constrained by the political, economic, and security policies (if any) of the network nodes, the protocol determines prices of all network links with respect to the routing-cost metric and then constructs routing to deliver a global traffic-demand matrix along the available paths. The considered model of a routing protocol is fairly generic and can be applied to not only computer networking but also transportation and other kinds of networks.

The key innovation in our routing-protocol model is a statistical notion of *protocol ignorance* that quantifies the inability of a routing protocol to accurately determine the price of a network link with respect to the routing-cost metric. This inability arises due to various reasons:

1) A protocol is designed to operate with a different metric than the metric of interest. For example, if the metric of interest is latency, the routing prices of links are in general determined inaccurately by BGP, Routing

Information Protocol (RIP) [9], and the other routing protocols that use the hop count as a proxy for latency.
2) Even when the security, political, and economic policies of any node do not prohibit routing along a link, a protocol artificially excludes the link from the constructed paths, which effectively renders the link price infinite. For instance, this happens in BGP due to its local filtering and single-path routing.
3) Because the routing-cost metric changes its value over time, a protocol measures the dynamic value inaccurately. For example, when latency is the metric of interest, Open Shortest Path First (OSPF) [10] sets the link price to mean latency and ignores higher moments of latency, which still reduces the protocol ignorance of OSPF compared to RIP.

For simplicity, we model a link price as a random variable, rather than a stochastic process.

What makes our study specific to inter-domain routing protocols is our model of protocol deployment. With a *deployment trajectory* referring to a sequence of network nodes that adopt a new inter-domain routing protocol, we assume that each deployment trajectory for the same number of adopting nodes is equally likely in inter-domain settings because ASes act as independent players. This feature distinguishes our model of incrementally deploying an inter-domain routing protocol from intra-domain settings, where the operator of a domain can hand-pick adopting nodes and deployment trajectories to maximize the amount of benefits realized by incremental deployment of a new intra-domain routing protocol.

Via theoretical analysis and packet-level simulation, we evaluate how the routing cost changes with incremental deployment of a new inter-domain routing protocol. We analytically characterize the dependence of the routing cost on a change in routing. In its turn, the simulation examines how protocol ignorance affects routing when the new protocol is incrementally deployed. Combining the two dependencies reveals that the routing cost changes smoothly during incremental deployment. The main contributions of our paper are as follows:

- We propose and rigorously define a statistical notion of protocol ignorance that quantifies the inability of a routing protocol to accurately determine link prices with respect to a routing-cost metric. Our protocol-ignorance model of a routing protocol is equally applicable to inter-domain and intra-domain protocols.
- Based on the notion of a deployment trajectory, we model incremental deployment of a new inter-domain routing protocol.
- Our analysis and simulation show that the routing cost changes smoothly during incremental deployment of a new inter-domain protocol. This explains the struggle of new inter-domain routing protocols to get widely deployed and indicates that their successful deployment necessitates a coordinated adoption effort by a large number of relevant ASes.

Notation	Semantics
$G = (V, E)$	Network topology with node set V and edge set E
$g = \|V\|$	Number of nodes in topology G
$n = \|E\|$	Number of edges in topology G
Z	Set of source-destination pairs
$m = \|Z\|$	Number of source-destination pairs
$z = (s_z, d_z)$	Node pair with source s_z and destination d_z
$R = (r_z)^T$	Traffic demands of all source-destination pairs z
P_z	Set of available paths for source-destination pair z
$l_z = \|P_z\|$	Number of available paths for source-destination pair z
P	Set of all available paths in the topology
$l = \|P\| = \sum_{z=1}^{m} l_z$	Total number of available paths in the topology
p	Path
e	Edge
$B = (b_{ep})$	Edge composition of all available paths
$W_z = (w_{zp})^T$	Traffic-demand split for source-destination pair z
W	Traffic-demand splits for all source-destination pairs
U	All-ones vector
a_e	Aggregate traffic demand on edge e
$F(\cdot) = (f_e(\cdot))^T$	Routing-price functions on all edges
C_e	Routing cost on edge e
$C = \sum_{e \in E} C_e$	Total routing cost

TABLE I: Notation in our model of the routing problem.

The paper has the following structure. Section II presents our model. Sections III and IV evaluate the model via analysis and simulations respectively. Section V discusses related work. Finally, section VI sums up the paper and its contributions.

II. Modeling

A. Routing problem

While our model of a routing protocol focuses on its ability to solve a routing problem, we first formalize the routing problem. Table I sums up relevant notation. We model the network topology as a directed graph $G = (V, E)$ with $g = |V|$ nodes and $n = |E|$ edges. Set Z of size m contains all source-destination pairs $z = (s_z, d_z)$, which have traffic demands $R = (r_z)^T$. Set P_z of size l_z contains all paths available for source-destination pair z. Then, $P = \bigcup_{\forall z \in Z} P_z$ of size $l = \sum_{z=1}^{m} l_z$ constitutes the set of all available paths in the topology. Matrix $B = (b_{ep})$ of size $n \times l$ expresses the edge composition of all available paths. Bit b_{ep} is 1 for path p containing edge e and equals 0 otherwise.

The considered routing problem is a problem of splitting all traffic demands R among available paths. Our model is for multipath routing and includes single-path routing as its special case. With w_{zp} denoting the fraction of traffic demand r_z routed along path p, we express the split of this traffic demand as vector $W_z = (w_{zp})^T$. Constraint $W_z^T U = 1$ ensures routing for the entire demand of source-destination pair z, where U is an all-ones vector. To represent the traffic-demand splits of all source-destination pairs z, we compose block matrix W of size $l \times m$ by forming its diagonal from

Notation	Semantics
Ω	Traffic demand as a random variable
Λ	Edge price as a random variable
X	Ω or Λ
$\phi_X(t)$	Characteristic function of X
$\Phi_X(x)$	Cumulative distribution function of X
k	Moment order
μ_k	Estimate for the k-th lowest moment of X by a real protocol
q	Number of X's lowest moments estimated by a real protocol
$\rho_X(t)$	Estimated characteristic function of X
$\psi_e(\cdot)$	Estimated routing-price function on edge e
$\Psi(\cdot)$	Vector $(\psi_e(\cdot))^T$ of the estimated routing-price functions
α or β	Real routing protocol
$\mathfrak{i}_e^{\alpha}$	Protocol ignorance of protocol α on edge e
$\mathfrak{i}_e^{\alpha\beta}$	Relative protocol ignorance of protocols α and β on edge e
$\mathfrak{I}^{\alpha\beta}$	Relative protocol ignorance of protocols α and β

TABLE II: Notation in our model of a routing protocol.

vectors W_z and setting all its other elements to zero. Each row of matrix W corresponds to the same path as in the respective column of matrix B. Given the global traffic-demand splits, we add up the traffic demands on edge e to compute aggregate traffic demand a_e on each edge e.

We define the routing problem with respect to a metric of routing cost. Following the approach by Roughgarden and Tardos [11], our model determines routing cost C_e on edge e as the product of its traffic demand and routing price: $C_e = a_e f_e(a_e)$ where routing-price functions $F(\cdot) = (f_e(\cdot))^T$ on all edges are non-negative and monotonic. For example, the edge price can be a monetary price of transiting one Mbps of traffic along the edge, latency experienced by the traffic on the edge, or a combination of multiple simple metrics. The routing-price functions can represent such effects as limited edge capacities and congestion, e.g., account for congestion-induced latency when the price is latency. With $C = \sum_{e \in E} C_e$ denoting the total routing cost for the entire network, we formulate the routing problem as a minimization of this total cost:

$$\begin{aligned}
&\text{minimize } C = F(BWR)^T BWR \\
&\text{under constraints } W_z^T U = 1 \ \forall z \in Z, \\
&\qquad\qquad w_{zp} \geq 0 \ \forall z \in Z \ \forall p \in P_z, \\
&\text{with inputs } G, Z, P, R, F(\cdot), \text{ and} \\
&\text{with outputs } W \text{ and } C.
\end{aligned}$$

B. Routing protocol

While section II-A formalizes the routing problem, we now present our model of a routing protocol, with Table II reporting respective additional notation. Our protocol model abstracts away operational details of the protocol, such as the format of its control messages, events that trigger them, etc. Instead, we focus on the inability of a routing protocol to optimally solve the routing problem with respect to the metric of interest due to *protocol ignorance*, which refers to the inability of the protocol to accurately determine the routing prices of network links. This section introduces and rigorously defines the stochastic notion of protocol ignorance. Our protocol-ignorance model of a routing protocol is fairly general and applicable to not only inter-domain but also intra-domain routing, as well as to transportation and other types of networks.

This inability of a routing protocol to know the routing prices exactly arises due to a variety of reasons. First, the protocol might be designed to operate with a different metric than the metric of current interest. For instance, while BGP and RIP use the hop count as the metric of routing cost, the metric of current interest might be latency, and the prominent hop-based protocols determine the routing prices of network links in regard to the latter metric imprecisely. Furthermore, the hop count is increasingly becoming a less representative proxy for path latency due to massive emergence of tunneling techniques that make some hops invisible to the routing protocol, e.g., because of remote peering in Internet inter-domain routing [12].

Second, the design of a routing protocol might unnecessarily exclude a link from routing some traffic, which effectively renders the link price infinite for the purposes of routing this traffic. For example, such link exclusion occurs in BGP due to local filtering of a path by an AS even when routing along the excluded link does not violate any economic, political, or security policy of any AS. Also, single-path routing in BGP unnecessarily prevents routing of some traffic along some links, which similarly undermines the ability of BGP to solve the routing problem optimally. Note that although single-path routing and local filtering in BGP simplify the protocol design and improve its scalability, these design choices are not fundamental for Internet inter-domain routing. For instance, Route Bazaar is an alternative inter-domain routing approach that supports multipath routing and uses a decentralized global public ledger for enabling each AS to make local routing decisions and enforcing the security, political, and economic policies of all ASes in a privacy-preserving manner.

Third, even when a protocol is designed for the same routing metric of interest, the protocol might be unable to exactly measure the dynamic values of the metric. For example, the values of path latency continuously change due to packet queuing in network nodes.

Regardless of the reasons why a particular protocol does not know the exact routing prices, the statistical notion of protocol ignorance quantifies this inability. Below, we refer to a protocol with imperfect knowledge of the routing prices as a *real protocol*. An *optimal protocol* measures the routing prices exactly.

1) Representation of an optimal protocol: For each edge e, we view its aggregate traffic demand and routing price as random variables Ω and Λ respectively and refer to either of them as X for exposition brevity. The characteristic function of X is $\phi_X(t) = \mathbb{E}[e^{itX}]$ where i is the imaginary unit, and $t \in \mathbb{R}$. In our model, an optimal protocol knows exactly all moments $\mathbb{E}[X^k]$ of X, where $k = 1, 2, \ldots, \infty$. According to the Hausdorff moment problem [13], the collection of all the moments uniquely determines the probability density function

(PDF) of X. Specifically, assuming that $\phi_X(t)$ is an analytic function, the optimal protocol expands it into a Taylor series:

$$\phi_X(t) = 1 + \sum_{k=1}^{\infty} \frac{(it)^k}{k!} \mathbb{E}[X^k] \tag{1}$$

and recovers the PDF of X from $\phi_X(t)$ through the inverse Fourier transform as $\frac{1}{2\pi} \int_{\mathbb{R}} \phi_X(t) e^{itx} dt$. By integrating the obtained PDFs of Ω and Λ, the optimal protocol obtains the cumulative distribution function (CDF) for each of these two random variables, $\Phi_\Omega(x)$ and $\Phi_\Lambda(x)$ respectively. Because routing-price function $f_e(\cdot)$ is monotonic, the optimal protocol computes it as $f_e(\cdot) = \Phi_\Omega^{-1}(\Phi_\Lambda(x))$ for each edge e and solves the routing problem of section II-A optimally.

2) Representation of a real protocol: On the other hand, a real protocol observes only samples drawn from the probability distribution of variable X and uses them to compute estimates μ_k for the q lowest moments of X, i.e., for $k = 1, \dots, q$. The real protocol computes an estimated characteristic function $\rho_X(t)$, an estimate of $\phi_X(t)$, as:

$$\rho_X(t) = 1 + \sum_{k=1}^{q} \frac{(it)^k}{k!} \mu_k. \tag{2}$$

By applying the inverse Fourier transform to $\rho_X(t)$ and then integrating the obtained PDF, the real protocol computes estimated routing-price functions $\Psi(\cdot) = (\psi_e(\cdot))^T$ and uses them instead of functions $F(\cdot) = (f_e(\cdot))^T$ when solving the routing problem of section II-A. Because $\Psi(\cdot)$ are only estimates of $F(\cdot)$, the real protocol computes routing W and its total cost C suboptimally in general.

3) Relevance to prominent existing protocols: Whereas existing routing protocols do not actually perform inverse Fourier transforms, integration, or other complicated operations described above, we now show that our model of a routing protocol realistically represents the handling of routing prices by prominent existing protocols.

Hop-based protocols. This kind of routing protocols uses the hop count as the metric of routing cost. BGP and RIP are prominent representatives of such protocols in the inter-domain and intra-domain settings respectively. In our model, a hop-based protocol does not measure any moments of X, i.e., $\mu_k = 0$ for $k = 1, \dots, \infty$, even when the routing metric of interest has dynamic values, e.g., when the metric of interest is latency. Thus, the respective estimated characteristic function is $\rho_X(t) = 1$. The inverse Fourier transform produces the Dirac delta function as the PDF of X, implying that X is a constant and that the edge cost is the same for all the edges, i.e., the model realistically represents the link pricing in a hop-based protocol.

Mean-measuring protocols. A mean-measuring protocol measures only the first moment, i.e., mean μ_1, of routing price Λ. OSPF is a prominent mean-measuring intra-domain protocol when it is configured to measure the routing price as mean latency, e.g., by using a sliding window estimation. While the hop-based BGP constitutes the only prominent existing protocol for inter-domain routing, Route Bazaar is an alternative Internet connectivity approach where mean-measuring protocols can be used for inter-domain routing. In our model of a mean-measuring protocol, the corresponding estimated characteristic function is $\rho_\Lambda(t) = 1 + (it)^k \mu_1$. The inverse Fourier transform yields $H(x) - \mu_1 \delta'(x)$ as the PDF of Λ, where $H(x)$ denotes the Heaviside step function, and $\delta'(x)$ is the derivative of the Dirac delta function. The integration of this function leads to estimating each edge cost as the mean of the metric, i.e., the model realistically represents the handling of routing prices by a mean-measuring protocol.

4) Mathematical definition of protocol ignorance: To model how accurately a real protocol α estimates edge price Λ in comparison to an optimal protocol, we define *protocol ignorance* $\mathfrak{i}_e^\alpha$ of protocol α on edge e as:

$$\mathfrak{i}_e^\alpha = \int_0^c \Big| \rho_\Lambda(t) - \phi_\Lambda(t) \Big| dt \tag{3}$$

where c is a constant ensuring existence of the integral. Based on equations 1 and 2, we express this protocol ignorance as:

$$\mathfrak{i}_e^\alpha = \int_0^c \Big| \sum_{k=0}^{\infty} (\frac{(it)^k}{k!} (\mathbb{E}[\Lambda^k] - \mu_k) \Big| dt \tag{4}$$

where $\mu_k = 0$ for $k > q$. The protocol ignorance of an optimal protocol equals 0. For a real protocol α, we have $\mathfrak{i}_e^\alpha > 0$. As the real protocol estimates more moments of Λ and measures each moment more accurately, $\mathfrak{i}_e^\alpha$ decreases toward 0, and the smaller protocol ignorance enables real protocol α to estimate the routing-price function on edge e more accurately.

The notion of protocol ignorance forms a basis for comparing two real routing protocols α and β. We define *relative protocol ignorance* of protocols α and β on edge e as:

$$\mathfrak{i}_e^{\alpha\beta} = \lim_{c \to \infty} \frac{\mathfrak{i}_e^\alpha - \mathfrak{i}_e^\beta}{c^{1+\max\{q^\alpha, q^\beta\}}} \tag{5}$$

which no longer depends on the choice of constant c. Here, q^α and q^β denote the number of moments estimated for edge price Λ by protocols α and β respectively. Vector of $\mathfrak{i}_e^{\alpha\beta}$ for all edges e in E provides a topology-wide perspective on the relative protocol ignorance. We define *relative protocol ignorance* of protocols α and β as a norm of this vector:

$$\mathfrak{I}^{\alpha\beta} = \sqrt{\sum_{e \in E} \left(\mathfrak{i}_e^{\alpha\beta} \right)^2}. \tag{6}$$

Example 1. Let α and β refer respectively to hop-based and mean-measuring protocols. For edge e, protocol β observes the following five samples of edge latency Λ, which has an exponential distribution: 0.81, 0.63, 2.10, 1.02, and 0.66. Using the samples, protocol β computes $\mu_1 = 1.044$ as an estimate of moment $\mathbb{E}[\Lambda]$, and $\rho_\Lambda(t) = \frac{1}{1-it}$ as an estimate of characteristic function $\phi_\Lambda(t)$. Then, the protocol ignorance of protocol β on edge e is $\mathfrak{i}_e^\beta = |\ln(1 - ic) - c - 0.522ic^2|$. Protocol α, which does not measure the edge latency at all, has a larger protocol ignorance $\mathfrak{i}_e^\alpha = |\ln(1 - ic)|$. Thus, the relative protocol ignorance of protocols α and β on edge e is $\mathfrak{i}_e^{\alpha\beta} = 0.022$, confirming the better awareness of protocol β

Notation	Semantics
h	Number of nodes that adopt the new protocol
j	Deployment trajectory
$\Psi_{hj}(\cdot)$	Routing-price functions for deployment trajectory j of h nodes
W_{hj}	Routing for deployment trajectory j of h nodes
C_{hj}	Routing cost for deployment trajectory j of h nodes
C_h	Average routing cost C_h for all deployments of h nodes
Θ	Edge-sharing matrix
γ^{ab}	Bilinear form
L_e	Lipschitz constant of estimated price function $\psi_e(\cdot)$ on edge e
L	Maximum L_e among all edges e
Υ_z^a	Auxiliary block matrix in the proof of theorem 2

TABLE III: Notation in our model of protocol deployment.

about the edge latency. If protocol β estimated the both lowest moments of edge latency Λ, its protocol ignorance on edge e would change to $\text{i}_e^{\beta} = |ln(1-ic)-c-0.522ic^2+0.463c^3|$, and the relative protocol ignorance of protocols α and β on this edge would increase to $\text{i}_e^{\alpha\beta} = 0.13$, representing the increased advantage of protocol β over protocol α in knowing the latency distribution on edge e. △

C. Incremental deployment of a new inter-domain protocol

While the model of a routing protocol in section II-B is equally applicable to inter-domain and intra-domain protocols, we now present our model for incremental protocol deployment specific to inter-domain routing protocols. Table III reports corresponding extra notation.

Suppose that all nodes in topology G support an incumbent inter-domain routing protocol α. Adopting a new inter-domain routing protocol β can reduce the routing cost in topology G because protocol β measures more accurately the routing prices on those edges where the new protocol is used. Some nodes deploy protocol β. When a node deploys protocol β, this protocol is used on all outcoming edges of this node. Protocol β is backward compatible with protocol α and runs on top of the incumbent protocol, e.g., by using Generic Route Encapsulation (GRE) tunnels [14] or another tunneling technique.

The routing and its cost depend on not only how many nodes adopt the new protocol but also which specific nodes are the adopters. Hence, we define a *deployment trajectory* of h nodes as a sequence of the first h adopting nodes in the order of their deployment of protocol β. Because ASes in the practice of inter-domain routing act as independent players, we assume that every deployment trajectory of h nodes is equally likely. The equal likelihood of deployment trajectories is the main feature distinguishing our inter-domain deployment model from intra-domain settings, where the domain operator can cherry-pick h adopting nodes to maximize the reduction in the routing cost. For the inter-domain settings, we express average routing cost C_h for all deployments of h nodes as:

$$C_h = \frac{1}{P(g,h)} \sum_{j=1}^{P(g,h)} C_{hj} \tag{7}$$

where g is the number of nodes in the topology, C_{hj} refers to the routing cost for the deployment of h nodes that has the j-th h-permutation of g as its trajectory, and $P(|V|,h)$ is the total number of such h-permutations of g.

Let us examine a full deployment of g nodes with trajectory j. Estimated routing-price functions $\Psi_{hj}(\cdot)$, routing W_{hj}, and routing cost C_{hj} for a deployment of h nodes might all change at each stage h along this trajectory, where $h = 1,\ldots,g$. As h increases, cost C_{hj} changes due to two conflated effects: (a) changes in routing W_{hj} and (b) changes in estimates $\Psi_{hj}(\cdot)$ of routing-price functions $F(\cdot)$. To segregate the two effects, we can track the value of $\tilde{C}_{hj} - C_{gj}$ at each stage h, where C_{gj} is the routing cost with the full deployment of protocol β, and $\tilde{C}_{hj}$ denotes the cost of routing W_{hj} computed with full-deployment routing-price functions $\Psi_{gj}(\cdot)$.

III. Analysis

The salient outcomes of our extensive modeling effort in section II include the formulation of the routing problem with respect to a metric of routing cost, statistical notion of protocol ignorance that quantitatively characterizes the inability of a protocol to measure routing prices accurately, and model for incremental deployment of a new inter-domain routing protocol. This section analyzes such incremental deployment. Specifically, we assess how much a change in routing affects the routing cost. The analysis is the first step towards understanding why BGP remains the only prominent inter-domain routing protocol and what fraction of the Internet ASes need to adopt a new inter-domain routing protocol to substantially benefit from the adoption.

A. Routing for one source-destination pair

For ease of exposition, we start the analysis by considering the simple scenario where the routing problem needs to be solved for only one source-destination pair z_1, i.e., $Z = \{z_1\}$. The set of available paths for the pair is P_1. Without loss of generality, we normalize the traffic demand of pair z_1 to $r_1 = 1$. Protocol β computes estimates $\Psi(\cdot)$ of routing-price functions $F(\cdot)$ as described in section II-B. With this, protocol β solves the following instance of the routing problem from section II-A:

minimize $C = \Psi(BW_1)^T BW_1$
under constraints $W_1^T U = 1$,
$w_{1p} \geq 0 \ \forall p \in P_1$,
with inputs $G, \{z_1\}, P_1, r_1 = 1, \Psi(\cdot)$, and
with outputs W_1 and C.

Consider two routings W_1^x and W_1^y that have costs C^x and C^y respectively. Vector $W_1^\epsilon = W_1^x - W_1^y$ of size $l_1 = |P_1|$ represents the difference between these routings. Let $\Theta = (\theta_{pu})$ of size $l_1 \times l_1$ denote an edge-sharing matrix $B^T B$ (where matrix B expresses the edge composition of all available paths), and θ_{pu} represents the number of edges shared by paths p and u, implying that θ_{pu} is at most the diameter of topology G. Then, we represent bilinear form $(W_1^a)^T \Theta W_1^b$

as γ^{ab}. By construction, estimated routing-price functions $\Psi(\cdot)$ have Lipschitz continuity. We define constant $L = \max_{e \in E}\{L_e\}$ where L_e is the Lipschitz constant of estimated routing-price function $\psi_e(\cdot)$ on edge e.

Theorem 1. *In routing for one source-destination pair, a change in the total routing cost is bounded from above as follows:*

$$|C^x - C^y| \leq L(|\gamma^{\epsilon\epsilon}| + 2|\gamma^{\epsilon y}|). \tag{8}$$

Proof. Let C^{ab} denote $\Psi(BW_1^a)^T BW_1^b$. Then, we express cost C^x as $\Psi(BW_1^x)^T B(W_1^y + (W_1^x - W_1^y)) = \Psi(BW_1^x)^T BW_1^y + \Psi(BW_1^x)^T W_1^\epsilon = C^{xy} + C^{x\epsilon}$. Similarly, we express cost C^y as $\Psi(BW_1^y)^T B(W_1^x - (W_1^x - W_1^y)) = \Psi(BW_1^y)^T BW_1^x - \Psi(BW_1^y)^T W_1^\epsilon = C^{yx} - C^{y\epsilon}$. Thus, we have:

$$C^x = C^{xy} + C^{x\epsilon} \text{ and } C^y = C^{yx} - C^{y\epsilon} \tag{9}$$

and express the sum of these two costs as:

$$C^x + C^y = C^{xy} + C^{yx} + C^{x\epsilon} - C^{y\epsilon}. \tag{10}$$

Using equation 10, we express the difference of the two costs as:

$$C^x - C^y = (C^{xy} - C^y) + (C^{yx} - C^y) + (C^{x\epsilon} - C^{y\epsilon}). \tag{11}$$

The three terms on the right-hand side of equation 11 have the following upper bounds: $|C^{xy} - C^y| = |(\Psi(BW_1^x)^T - \Psi(BW_1^y)^T)BW_1^y| \leq L|\epsilon^T BW_1^y| = L|\gamma^{\epsilon y}|$, $|C^{yx} - C^y| = |\Psi(BW_1^y)^T BW_1^\epsilon| = |(\Psi(BW_1^y)^T - \Psi(BO)^T)BW_1^\epsilon| \leq L|\gamma^{y\epsilon}|$, and $|C^{x\epsilon} - C^{y\epsilon}| = |(\Psi(BW_1^x)^T - \Psi(BW_1^y)^T)BW_1^\epsilon| \leq L|\gamma^{\epsilon\epsilon}|$ where O is a zero vector. Because the symmetry of matrix Θ implies $\gamma^{y\epsilon} = \gamma^{\epsilon y}$, we combine the above three bounds to derive equation 8. □

B. Routing for an arbitrary set of source-destination pairs

Now, we extend the result of theorem 1 for the general formulation of the routing problem in section II-A, i.e., when set Z of source-destination pairs and traffic demands R are arbitrary. The extension is fairly straightforward and largely related to generalizing the notation from vectors to matrices. In particular, matrix W^ϵ denotes the difference between two routings W^x and W^y, where W_z^ϵ equals $W_z^x - W_z^y$. Also, we define $\gamma_z^{\epsilon\epsilon} = (W_z^\epsilon)^T \Theta_z (W_z^\epsilon)$ where Θ_z equals $B_z^T B_z$.

Theorem 2. *In the general routing problem, a change in the total routing cost is bounded from above as follows:*

$$|C^x - C^y| \leq L(|\sum_{z=1}^m r_z^2 \gamma_z^{\epsilon\epsilon}| + 2|\sum_{z=1}^m r_z^2 \gamma_z^{\epsilon y}|). \tag{12}$$

Proof. By substituting W_1^x and W_1^y with $W^x R$ and $W^y R$ respectively, we follow the reasoning pattern in the proof of theorem 1 to show that

$$|C^x - C^y| \leq L(|(W^\epsilon R)^T \Theta(W^\epsilon R)| + 2|(W^\epsilon R)^T \Theta(W^y R)|).$$

Let Υ_z^a denote an auxiliary block matrix that has the same size as W. Its z-th block is W_z^a, and all the other elements equal zero. Then, we express W^a as $\sum_{z=1}^m \Upsilon_z^a$. Because $(\Upsilon_z^a R)^T \Theta(\Upsilon_j^b R)$ is zero for $j \neq z$, we represent $(W^a R)^T \Theta(W^b R)$ as $\sum_{z=1}^m (\Upsilon_z^a R)^T \Theta(\Upsilon_z^b R)$. By expressing $(\Upsilon_z^a R)^T \Theta(\Upsilon_z^b R)$ as $r_z (W_z^a)^T \Theta_z (W_z^b)$, we derive:

$$(W^a R)^T \Theta(W^b R) = \sum_{z=1}^m r_z (W_z^a)^T \Theta_z (W_z^b). \tag{13}$$

By applying equation 13 to both terms on the right-hand side of the bound earlier in the proof, we establish equation 12. □

The above results demonstrate that a change in the routing affects the routing cost smoothly, i.e., a small change in routing do not cause a large change in the routing cost.

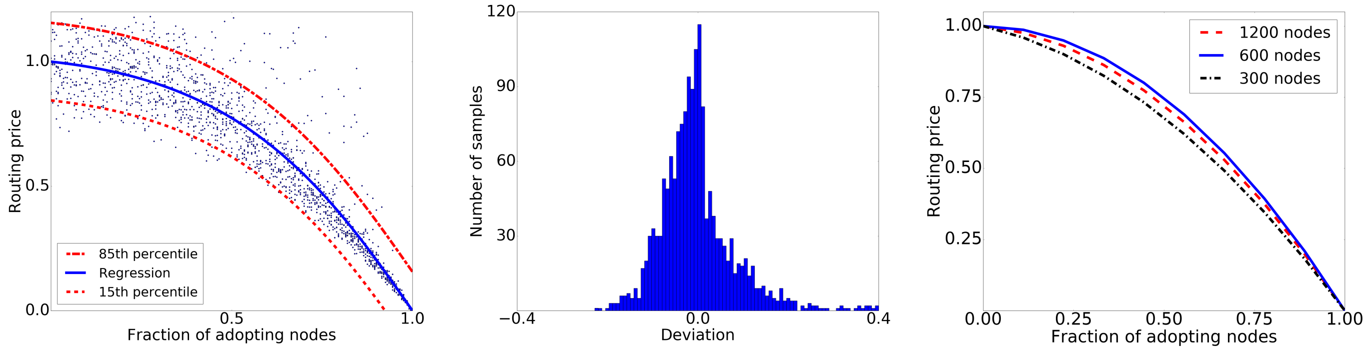
IV. Simulation

The analytic results in section III tell only half the story. They show how a change in routing affects the routing cost. To complete the story, we now examine how the lower ignorance of an incrementally deployed inter-domain protocol affects routing.

A. Methodology

We use real network data to conduct packet-level simulation. Simulating the global Internet faces two steep challenges: scale and fidelity. Because the Internet consists of more than 50,000 ASes connected by around a million inter-domain links, simulation of the entire topology would require enormous computational resources. Furthermore, neither the Internet topology nor its traffic-demand matrix is known with high precision for such simulation to produce highly accurate quantitative answers. In dealing with these challenges, we openly admit necessary limitations of the simulated model (such as using a single node to represent an AS), avoid a focus on exact quantitative results, and instead strive to expose the qualitative dependence of routing on protocol ignorance.

1) Topology: To tackle the challenge of topology scale, we characterize statistical properties of the global Internet topology and generate a family of smaller topologies with the same statistical properties. Specifically, we reconstruct snapshots of the AS-level Internet topology from the CAIDA dataset [15] based on traceroute [16] measurements. Our characterization of the snapshots confirms the observation that the AS-level Internet topology is a scale-free graph. Route Views [17] and other prominent sources of Internet connectivity data can be used to make the same observation. Based on the scale-free characterization, we use NetworkX [18], [19] to generate synthetic topologies that preserve the statistical properties of the global Internet topology and range in their size from 100 to 1,200 nodes, with the default size of 500 nodes. The probability to add an edge to a node during the topology generation varies between 0 and 50%, with 10% being the default value. The minimum number of edges adjacent to each node equals 3 by default and changes from 1 to 50.



(a) 2,000 runs with 10 runs per setting (b) Deviations of the 2,000 samples (c) Regression for different topology sizes

Fig. 1: Impact of protocol ignorance on the routing price during incremental deployment of the new inter-domain protocol.

2) Traffic matrix: While the Transmission Control Protocol (TCP) [20] transmits packets in bursts, the basic level of traffic in the simulation is a packet burst. Each source-destination pair communicates around 10,000 packet bursts. We size every packet to 64 KB. The number of packets in a burst varies from 1 to 32 and is distributed binomially with 32 independent experiments and success probability 0.5. Time between subsequent packet bursts has a random distribution with a mean of 1 ms. We consider three such distributions: exponential, uniform, and Weibull, with the exponential distribution being used in the default settings. We include the Weibull distribution due to prior measurement studies [21], [22] and independently validate it on another CAIDA dataset [23].

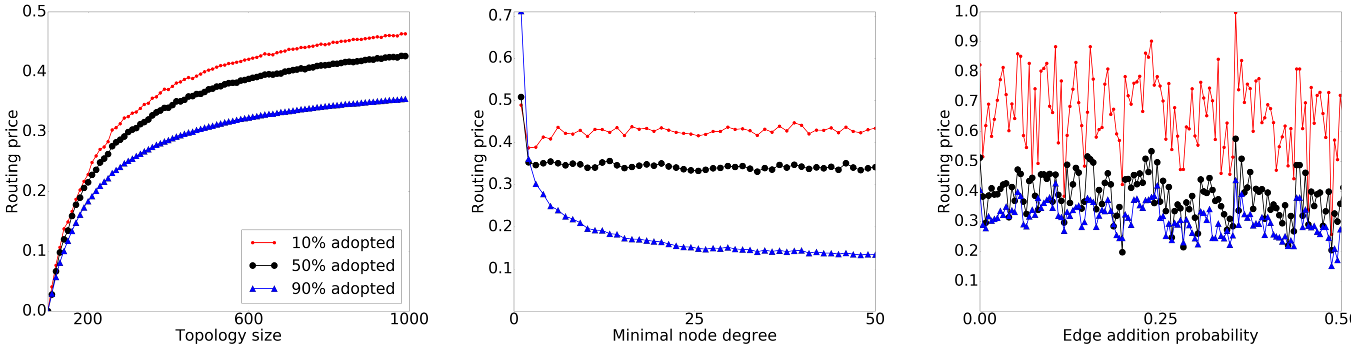
3) Routing protocols: The routing metric of interest in the simulation is latency. For the incumbent and new inter-domain routing protocols, we respectively consider the hop-based and mean-measuring protocols described in section II-B.

4) Simulator: To improve scalability of the simulation, we develop and utilize our own simulator. Unlike ns-3 [24] and other generic simulators that support many features at the price of significant overhead, our tool is customized for the problem in hand to scalably simulate traffic generation, routing, and delivery for each source-destination pair in the network topology. The tool is a discrete-time event simulator that represents each AS as a single node. In addition to generating packet bursts, every node also forwards packet bursts from other sources. The node forwards all packets of a burst together as a whole. The forwarded burst experiences transmission latency determined by dividing the burst size by the internal capacity of the AS; this internal capacity of the node is drawn from a truncated normal distribution. Additionally, the forwarded packet burst experiences queuing latency drawn from the exponential distribution with a mean of 0.05 ms. The simulator keeps the average network utilization at 50% by: (1) setting the capacity of each edge according to closeness centrality of both nodes incident to this edge and (2) then characterizing each edge with extra latency drawn from the same exponential distribution for all edges, with the rate parameter of this distribution being determined experimentally. For every simulated setting, we conduct 10 runs and, for each run, measure the routing price as the average end-to-end latency in the network. The code of our simulator is available in [25].

B. Simulation results

1) Impact of protocol ignorance: Figure 1a depicts how the routing price changes when the fraction of nodes adopting the new protocol increases with a step of 0.5% from 0 to 100%, i.e., from no deployment to full deployment. For each of the 10 runs in every simulated setting, we plot the routing price as a point. Figure 1a also plots a polynomial regression and its 15th and 85th percentiles for the results, with the elbow method consistently identifying a quadratic regression as the best fit. We normalize the plotted results by linearly scaling them to map interval $[f,n]$ into [0,1] where f and n refer to the regression values in the full-deployment and no-deployment settings respectively. Figure 1b reports a histogram of the deviations of the individual results from the corresponding regression values. Figure 1c plots the regression for three other sizes of the network topology. The dependence of the routing price on the deployment extent exhibits the same qualitative profile and only minor quantitative differences. The routing price undergoes a smooth quadratic decline over the entire range of incremental deployment. For the four considered sizes of the topology, between 43% and 53% of all nodes have to adopt the new protocol to decrease the routing price by 25%.

Combining the simulation insights with the analytical results from section III, we conclude that the benefits from adopting the new inter-domain protocol accumulate smoothly during incremental deployment and that protocol deployment by natural early adopters [26], [27] is insufficient to incentivize other ASes to deploy the protocol later. Our findings explain the struggle of new Internet inter-domain routing protocols to get widely deployed. Our results also indicate that widespread deployment of a new inter-domain protocol necessitates involving a large number of relevant ASes into a coordinated effort to adopt the protocol.



(a) Topology size (b) Minimal connectivity (c) Edge addition probability

Fig. 2: Sensitivity of the routing price to parameters of the topology generation algorithm.

2) Parameter sensitivity: Figure 1c already shows that the topology size makes a small quantitative impact on the quadratic dependence between the deployment extent and routing price. We also evaluate sensitivity of this dependence to the distribution of inter-burst time (exponential, uniform, or Weibull) and parameters of the topology generation algorithm. Compared to the results in figure 1c, these sensitivity studies unveil the same quadratic qualitative profile and even smaller quantitative differences for the dependence of the routing price on the fraction of adopting nodes. Due to such similarity and space constraints, we do not report the respective graphs here.

Figure 2 explores sensitivity of the routing price to three parameters of the topology generation algorithm when the fraction of adopting nodes is fixed at 10%, 50%, or 90%. We normalize the plotted results by linearly scaling them to map interval $[l,h]$ into [0,1] where l and h refer respectively to the minimum and maximum routing price across all settings in these three parameter sensitivity studies.

Figure 2a shows that the routing price grows sublinearly as the topology size increases from 100 to 1,000 nodes. The result corroborates the intuition that the topology diameter grows slower than the topology size. For scale-free graphs, [28] analytically shows that the diameter grows on average with rate $\frac{log(g)}{log(log(g))}$ where g is the number of nodes. Because the diameter is determined by the longest shortest path, and the end-to-end routing cost grows on average linearly with the path length, the dependency depicted in figure 2a matches the theoretical expectation.

Figure 2b exhibits dependence of the routing price on the minimal node degree in the topology. With a larger fraction of adopting nodes, the routing price falls steeper as the minimal node degree increases from 1 to 50. This happens due to dependency between the node degree and number of paths in the topology. Because the incumbent protocol uses the number of hops as a proxy metric for latency and thus estimates actual routing prices inaccurately, the decrease in the routing price is more pronounced for larger deployments of the new protocol.

Figure 2c reports on varying the probability of adding a random edge during the topology construction. The topology generation algorithm keeps the number of added edges proportional to the topology size. Whereas the increase of the minimal node degree in our previous sensitivity study weakens the scale-free property of the topology, the addition of random edges strengthens this property without increasing the number of paths exponentially. Figure 2c demonstrates low sensitivity of the routing price to the edge addition probability.

Overall, among all conducted sensitivity studies, the routing price is most sensitive to the topology size and fraction of adopting nodes.

V. Related work

While prior work on inter-domain routing is extensive, its main focuses are not on the problem of incrementally transitioning from BGP to a new protocol. Even those papers that explicitly consider incremental migration to the new protocol tend to deal with technical issues of the transition [6] and do not provide clear answers on economic incentives for adopting ASes, especially for early adopters [26], [27]. Whereas [29] proposes a method for service composition that can be used to combine different routing protocols, the paper develops the marketplace support without studying incremental adoption of the marketplaces. [30] argues that it is possible to select a relatively small set of routing brokers, about 7% of the Internet ASes, to enable QoS-aware and, in particular, latency-aware routing for most of the Internet; while [30] hand-picks the routing brokers among strategically positioned ASes, our work makes a more realistic assumption that ASes adopt the new routing protocol voluntarily and randomly.

Incremental deployability attracts more direct attention in other problem domains of computer networking. In the context of Internet addressing, [31] studies incremental migration from IPv4 to IPv6 and estimates its costs. [32] analyzes a potential way to incrementally deploy a secure version of BGP. Our paper differs from these previous efforts in not only tackling a different problem domain but also using a new method based on protocol ignorance. The distinguishing trait of our work is its model that captures the inability of a protocol to estimate costs accurately.

Our work leverages various prominent previous efforts. The protocols designed in [6], [8], [33]–[35] inspire us to develop the concept of protocol ignorance. Our analysis extends the classical theoretical work by Roughgarden and Tardos on the price of anarchy in different types of networks [11]. Their research paves the way to formalize the routing problem and characterize dependence of the routing cost on routing. [11], [36] analyze different scenarios of network behaviour in its dependency on node behaviour. Our simulations rely on realistic network topologies [37] and use real traffic traces collected by CAIDA [15], [23]. The approaches in [21], [22], [38] guide our modeling work.

[39], [40] report on mathematical modeling of Internet protocols. Our work belongs to the same type of research. We develop a novel abstract model for inter-domain routing that ties together a routing protocol, routing constructed by the protocol, and cost of the constructed routing.

VI. Conclusion

In this paper, we studied incremental deployment of a new inter-domain routing protocol in the Internet. The paper formalized the routing problem in terms of minimizing a metric of routing cost. Then, we introduced and rigorously defined a statistical notion of protocol ignorance that quantifies the inability of a routing protocol to accurately determine routing prices with respect to the metric of interest. Our protocol-ignorance model of a routing protocol is fairly generic and applicable to not only inter-domain but also intra-domain routing, as well as to transportation and other kinds of networks. The considered model of protocol deployment made our study specific to Internet inter-domain routing. Using theoretical analysis and simulation, we showed that the benefits from adopting the new inter-domain protocol accumulate smoothly during incremental deployment. In the simulated topologies, between 43% and 53% of all nodes had to adopt the new protocol to decrease the routing price by 25%. Our results explained the lack of widespread adoption for new inter-domain routing protocols and indicated that their successful deployment necessitated a coordinated adoption effort by a large number of relevant ASes.

VII. Acknowledgments

This research was financially supported in part by the Regional Government of Madrid on Cloud4BigData grant S2013/ICE-2894.

References

[1] Y. Rekhter, T. Li, and S. Hares, "A Border Gateway Protocol 4 (BGP-4)," RFC 4271, 2006.
[2] P. Bangera and S. Gorinsky, "Impact of Prefix Hijacking on Payments of Providers." COMSNETS, 2011.
[3] J. Karlin, S. Forrest, and J. Rexford, "Pretty Good BGP: Improving BGP by Cautiously Adopting Routes." ICNP, 2006.
[4] M. Caesar and J. Rexford, "BPG Routing Policies in ISP Networks," *IEEE Network, 19(6)*, 2005.
[5] W. Sun, Z. Mao, and K. Shin, "Differentiated BGP Update Processing for Improved Routing Convergence." ICNP, 2006.
[6] P. B. Godfrey, I. Ganichev, S. Shenker, and I. Stoica, "Pathlet Routing." SIGCOMM, 2009.
[7] V. Valancius, N. Feamster, R. Johari, and V. Vazirani, "MINT: A Market for INternet Transit." ReArch, 2008.
[8] I. Castro, A. Panda, B. Raghavan, S. Shenker, and S. Gorinsky, "Route Bazaar: Automatic Interdomain Contract Negotiation." HotOS, 2015.
[9] C. Hedrick, "Routing Information Protocol," RFC 1058, 1998.
[10] J. Moy, "OSPF Version 2," RFC 2328, 1998.
[11] T. Roughgarden and E. Tardos, "How Bad is Selfish Routing?" *Journal of the ACM, 49(2)*, 2002.
[12] I. Castro, J. C. Cardona, S. Gorinsky, and P. Francois, "Remote Peering: More Peering Without Internet Flattening." CoNEXT, 2014.
[13] J. Shohat and J. Tamarkin, *The Problem of Moments*. American Mathematical Society, 1943.
[14] D. Farinacci, S. Hanks, and P. Traina, "Generic Routing Encapsulation (GRE)," RFC 1701, 1994.
[15] CAIDA, "Anonymized Internet Traces 2012," 2016. [Online]. Available: http://www.caida.org/data/request_user_info_forms/ark.xml
[16] V. Jacobson, "Traceroute," 1989.
[17] University of Oregon, "Routeviews Prefix to AS Mappings Dataset (pfx2as) for IPv4 and IPv6," 2005. [Online]. Available: http://www.routeviews.org/routeviews/
[18] A. Hagberg, D. Schult, and P. Swar, "Exploring Network Structure, Dynamics, and Function using NetworkX." SciPy, 2008.
[19] P. Holme and B. Kim, "Growing Scale-Free Networks with Tunable Clustering," *Physical review E, 65(3)*, 2002.
[20] Information Sciences Institute University of Southern California, "Transmission Control Protocol," RFC 793, 1981.
[21] N. Hariri, B. Hariri, and S. Shirmohammadi, "A Distributed Measurement Scheme for Internet Latency Estimation," *IEEE Transactions on Instrumentation and Measurement, 60(5)*, 2011.
[22] K. P. Gummadi, S. Saroiu, and S. D. Gribble, "King : Estimating Latency between Arbitrary Internet End Hosts." IMW, 2002.
[23] CAIDA, "IPv4 Routed /24 Topology Dataset," 2016. [Online]. Available: http://www.caida.org/data/passive/passive_dataset_request.xml
[24] NS-3 Consortium, "ns-3," 2017. [Online]. Available: https://www.nsnam.org/
[25] V. Kirilin and S. Gorinsky, "Simulator Source Code," 2017. [Online]. Available: https://github.com/WVadim/SimulatorC-Large
[26] C. Catalini and C. Tucker, "When Early Adopters Don't Adopt," *Science, 365(6)*, 2017.
[27] J. R. Douceur and T. Moscibroda, "Lottery Trees: Motivational Deployment of Networked Systems." SIGCOMM, 2007.
[28] B. Bollobás and O. Riordan, "The Diameter of a Scale-Free Random Graph," *Combinatorica, 24(1)*, 2004.
[29] S. Bhat, R. Udechukwu, R. Dutta, and G. N. Rouskas, "On Service Composition Algorithm for Open Marketplaces of Network Services." EuCNC, 2017.
[30] D. Lin, D. Hui, W. Wu, T. Liu, Y. Yang, Y. Wang, J. C. Lui, G. Zhang, and Y. Li, "On the Feasibility of Inter-Domain Routing via a Small Broker Set." ICDCS, 2017.
[31] A. Durand, "Deploying IPv6," *IEEE Internet Computing, 5(1)*, 2001.
[32] H. Chan, D. Dash, A. Perrig, and H. Zhang, "Modeling Adoptability of Secure BGP Protocol." SIGCOMM, 2006.
[33] V. Fuller, T. Li, J. Yu, and K. Varadhan, "Classless Inter-Domain Routing (CIDR): An Address Assignment and Aggregation Strategy," RFC 4632, 2006.
[34] V. Kotronis, X. Dimitropoulos, R. Klöti, B. Ager, P. Georgopoulos, and S. Schmid, "Control Exchange Points: Providing QoS-enabled End-to-End Services via SDN-based Inter-domain Routing Orchestration." ONS, 2014.
[35] D. Farinacci, D. Lewis, D. Meyer, and V. Fuller, "The Locator/ID Separation Protocol (LISP)," RFC 6830, 2013.
[36] J. R. Correa, A. S. Schulz, and N. E. Stier-Moses, "Selfish Routing in Capacitated Networks," *Mathematics of Operations Research, 29(4)*, 2004.
[37] S.-H. Yook, H. Jeong, and A.-L. Barabasi, "Modeling the Internet's Large-Scale Topology," *Proceedings of the National Academy of Sciences, 99(21)*, 2002.
[38] V. Ramasubramanian, D. Malkhi, F. Kuhn, M. Balakrishnan, A. Gupta, and A. Akella, "On the Treeness of Internet Latency and Bandwidth." SIGMETRICS, 2009.
[39] F. Kelly, *Mathematical Modelling of the Internet*. Springer-Verlag, 2001.
[40] C. Papadimitriou, "Algorithms, Games, and the Internet." STOC, 2001.