From Non-Optimal Routing Protocols to Routing on Multiple Optimality Criteria

João Luís Sobrinho, *Senior Member, IEEE, Member, ACM*, and Miguel Alves Ferreira, *Student Member, IEEE*

***Abstract*—At a suitable level of abstraction, all that standard routing protocols do is iterate extension and election operations on path attributes. An extension operation composes the attribute of a path from those of a link and another path, while an election operation produces the most preferred attribute of a set of candidate attributes, given a total order on them. These protocols are guaranteed to compute optimal paths only if the extension operation and the total order are entwined by the algebraic property of isotonicity, which states that the relative preference between two attributes is not inverted when both are extended by a third attribute. We solve the problem of computing and routing on optimal paths with generality by recognizing that every total order contains a partial order for which isotonicity holds. Then, we design a partial-order vectoring protocol where every election operation produces a subset of attributes from a set of candidate attributes, rather than a single attribute as is the case with a standard vectoring protocol; the election operation is derived from the partial order, ensuring that no attribute of the set of candidate attributes is preferred to an attribute of the elected subset. Moreover, we show how partial-order vectoring protocols can be designed to allow routing on optimal paths concurrently for diverse optimality criteria. Our evaluation over publicly available network topologies and attributes, covering both intra- and inter-AS routing, evince that the sizes of elected subsets of attributes are surprisingly small and that the partial-order vectoring protocol converges fast, sometimes faster than a standard vectoring protocol operating in the absence of isotonicity.**

***Index Terms*—Multiple optimality criteria, routing protocol, routing algebra, partial order.**

I. Introduction

THE quickest path to deliver a file [1]; the wireless communication path that drains the least energy from battery-powered devices [2]; or the shortest Autonomous System (AS) path where intermediate ASes profit from transiting traffic [3], are all examples of optimal paths. The definition of optimal path and the computation of optimal paths can be expressed with generality by a routing algebra consisting of: (1) a set of attributes, which represents performance metrics and/or policy choices in context; (2) a binary extension operation on attributes, which allows the computation of the attribute of a path from the attributes of its constituent links; and (3) a total order on attributes, which establishes a relative preference among them [4]–[9]. The total order defines the optimality criterion. A path is optimal if its attribute is preferred to that of any other path from the same source to the same destination. The prototypical routing algebra takes real numbers representing lengths for attributes, the extension operation is addition and the total order is the less-than-or-equal order. An optimal path is a shortest path.

Any shortest path routing protocol can be generalized to a routing algebra by substituting addition and the less-than-or-equal order on real numbers by an arbitrary extension operation and an arbitrary total order on attributes. However, the resulting protocols may fail to compute optimal paths. For example, the Enhanced Interior Gateway Routing Protocol (EIGRP) [10] operates on pairs width-length, preferring those pairs that minimize the latency to deliver a file, but it does not compute or route on quickest paths, in general [5], [11]; the Border Gateway Protocol (BGP) [12] allows configuration of routing policies at each AS of the Internet, but overall it does not compute or route on best policy paths [6]. The generalized protocols are guaranteed to compute optimal paths only if the extension operation is isotone for the total order, meaning that the relative preference between any two attributes is not inverted when both are extended by an arbitrary third attribute [4]–[6], [8].

Besides not computing optimal paths, current routing protocols operate according to a single optimality criterion, whereas different applications sharing a network call for distinct optimality criteria: for instance, a path that is optimal to deliver a file is not necessarily optimal to stream a video. In this paper, we present a general solution to the problem of computing and routing on optimal paths for a single optimality criterion and concurrently for multiple optimality criteria.

Routing on a single optimality criterion. Suppose that we have an optimality criterion for which the extension operation is not isotone. In this case, we re-formulate the concept of routing algebra to preserve isotonicity. Specifically, we trade the totality of the order for isotonicity: we extract a partial order from the total order such that the extension operation is isotone for the partial order [13]. In a partial order, one attribute of a pair of attributes is preferred to the other or the two attributes are incomparable [14]. A partial order defines a dominance criterion. The subset of dominant attributes of a set of attributes consists of those attributes that are not less preferred than any other attribute of the set: it is a plural set

Manuscript received 22 June 2021; revised 3 May 2022; accepted 5 July 2022; approved by IEEE/ACM Transactions on Networking Editor S. Kompella. Date of publication 29 July 2022; date of current version 16 February 2023. This work was supported in part by Portuguese Fundação Para a Ciência e Tecnologia, Portugal, under Grant UIDB/50008/2020. *(Corresponding author: João Luís Sobrinho.)*

João Luís Sobrinho is with Instituto Superior Técnico, University of Lisbon, 1049-001 Lisbon, Portugal and also with the Instituto de Telecomunicações, 3810-193 Aveiro, Portugal (e-mail: joao.sobrinho@lx.it.pt).

Miguel Alves Ferreira is with Instituto Superior Técnico, Instituto de Telecomunicações and Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: maferrei@andrew.cmu.edu).

Digital Object Identifier 10.1109/TNET.2022.3191808

1558-2566 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Tsinghua University. Downloaded on September 12,2025 at 07:14:59 UTC from IEEE Xplore. Restrictions apply.

of pairwise incomparable attributes, in general, which contains the optimal attribute of the original total order.

Next, we design a partial-order vectoring protocol that operates on partial orders to compute the set of dominant path attributes from every source to every destination in a network. The protocol instantiates a separate routing computation per destination, as standard vectoring protocols do, but has every node elect and advertise to neighbors a set of dominant attributes to reach the destination, rather than a single optimal attribute. Every node assigns a locally unique label to each attribute of its elected set that is advertised alongside the attribute [15]. A source of data-packets recognizes the optimal attribute of the original total order among its elected attributes and labels data-packets accordingly, to be forwarded along an optimal path through label-switching.

Routing on multiple optimality criteria. Suppose now that we have a collection of optimality criteria, each defined by a different total order on a common set of attributes. Given any two attributes, either their relative preference differs on at least two criteria of the collection or their relative preference is the same on all criteria. We build a partial order on attributes such that one attribute is preferred to another if the former is preferred to the latter on all criteria. When the extension operation is not isotone for the partial order built this way, we once again extract from it a partial order for which isotonicity holds. Then, the partial-order vectoring protocol computes the set of dominant attributes from every source to every destination in a network. A source of data-packets is able to select the optimal attribute that is most appropriate to each specific data-packet among its elected attributes and labels the data-packet accordingly, to be forwarded along the corresponding optimal path through label-switching.

The routing state maintained by a partial-order vectoring protocol on a network is proportional to the sizes of the sets of dominant attributes from sources to destinations. The practical value of such a protocol depends on these sets being small. We computed the sizes of sets of dominant attributes on the Rocketfuel topologies [16], annotated with metrics related to bandwidth and delay, and on CAIDA topologies [17], annotated with the type of relationship between neighbor ASes. For the Rocketfuel topologies, we found that the average number of dominant attributes from source to destination is below three even for a network with worldwide coverage, hundreds of nodes, and more than a thousand links. For the CAIDA topologies, no more than three dominant attributes ever connect a source to a destination with more than 90% of source-destination pairs being connected by a single dominant attribute.

Another important consideration is the speed with which a partial-order vectoring protocol terminates in a stable state following a network event, such as the announcement of a new destination or the failure of a link. We ran simulations of vectoring protocols on the Rocketfuel topologies. We found that the termination time of a partial-order vectoring protocol is only marginally worse than that of a standard vectoring protocol when both protocols operate under isotonicity and that it is sometimes better than that of a standard vectoring protocol when the latter does not operate under isotonicty.

A. Roadmap

The work herein reported developed from a previous conference paper [18]. The present paper: (1) relies on weaker hypothesis for routing algebras, requiring neither associativity nor commutativity of the extension operation; (2) applies the concepts developed to inter-AS routing; and (3) offers an evaluation of the stable state of a partial-order vectoring protocol on CAIDA topologies. On the other hand, only one type of partial-order vectoring protocols is discussed and the proofs of termination and dominance, which are similar to those presented in [18], are omitted.

Optimal path routing for a single optimality criterion and for multiple optimality criteria is first illustrated in Section II and Section III, in a context solely comprised of performance metrics and in a context involving inter-AS policy decisions, respectively. Section IV develops the general theory of optimal path routing, introducing the concepts of isotonic reduction of an order and intersection of multiple orders. Section V designs a canonical partial-order vectoring protocol that operates on partial orders. Section VI and Section VII present an evaluation of optimal path routing for a single optimality criterion and for multiple optimality criteria on the Rocketfuel topologies and on the CAIDA topologies, respectively. Section VIII reviews related work and Section IX concludes the paper. The appendices contain proofs of two propositions.

II. Routing on Widths and Lengths

We illustrate how vectoring protocols based on partial orders allow optimal path routing for a variety of optimality criteria. In the forthcoming examples, every link and path in a network is characterized by a pair *width-length* belonging to the Cartesian product of the set of positive or infinite widths with the set of nonnegative lengths. Width represents a metric, such as capacity or available bandwidth, that extends along a path with the minimum operator, whereas length represents a metric, such as delay or number of data-packets in queue, that extends along a path with addition. Therefore, the *extension* of width-length (w,l) with width-length (w',l') is width-length $(\min\{w,w'\}, l+l')$.

A *shortest-widest path* is a path of minimum length among those of maximum width from a source to a destination in a network [19].[1] Shortest-widest paths are selected according to the *shortest-widest order* (lexicographic order), which establishes that width-length (w,l) is preferred to width-length (w',l') if its width is greater, $w > w'$, or the widths are equal but its length is smaller, $w = w'$ and $l < l'$.

In the network of Figure 1, each link is annotated with a pair width-length and all nodes want to route data-packets to destination x along shortest-widest paths. By inspection, we readily conclude that the shortest-widest path from v to x is vwx, with width-length $(20,5) = (\min\{20,20\}, 4+1)$, and that the shortest-widest path from u to x is uvx, with width-length $(10,5) = (\min\{10,10\}, 3+2)$.

[1] The mnemonic for shortest-widest path is "shortest of the widest paths." The naming of instances of optimal paths will follow this model throughout the paper.

Authorized licensed use limited to: Tsinghua University. Downloaded on September 12,2025 at 07:14:59 UTC from IEEE Xplore. Restrictions apply.

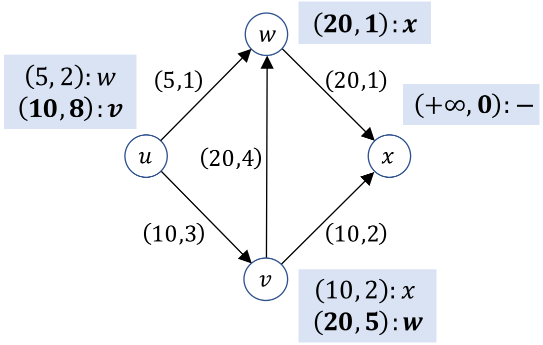


Fig. 1. Stable state of a standard vectoring protocol operating according to the shortest-widest order for destination x. Links are annotated with width-lengths. Elected width-lengths are in bold.

Consider a standard vectoring protocol that operates on the shortest-widest order, with each node electing and advertising to its in-neighbors the most preferred width-length learned from its out-neighbors. The stable state of such a protocol is shown in the figure. Destination x elects $(+\infty, 0)$ and w elects $(20, 1)$. Node v learns $(10, 2)$ from x and $(20, 5)$ from w, which is the extension of $(20, 4)$ of link vw with $(20, 1)$ of the elected width-length at w. It elects $(20, 5)$, learned from w, instead of $(10, 2)$, learned from x, on account of its greater width. Node u learns $(5, 2)$ from w, which is the extension of $(5, 1)$ with $(20, 1)$, and $(10, 8)$ from v, which is the extension of $(10, 3)$ with $(20, 5)$. It elects $(10, 8)$, learned from v, because of its greater width.

Node u forwards data-packets to v, which forwards them to w, which delivers them to x. Thus, data-packets with source at u and destination at x travel along $uvwx$, which is not the shortest-widest path from u to x. The standard vectoring protocol fails to route data-packets along shortest-widest paths; it fails because the extension operation on width-lengths is not isotone for the shortest-widest order. Concretely, $(20, 5)$ is preferred to $(10, 2)$, but $(10, 8)$, which is the extension of $(10, 3)$ with $(20, 5)$, is less preferred than $(10, 5)$, which the extension of $(10, 3)$ with $(10, 2)$.

Note that node u would be able to compute width-length $(10, 5)$ of the shortest-widest path to x if v refrained from making a decision of which of the two width-lengths $(10, 2)$ and $(20, 5)$ is preferred, rather advertising them both to u. In order to pursue this line of inquiry rigorously, consider the *product order on width-lengths* [14], which is such that width-length (w, l) is preferred to width-length (w', l') if it is different from (w', l') and both its width equals or is greater, $w \geq w'$, and its length equals or is smaller, $l \leq l'$, than those of (w', l'). The product order is a partial order. Two width-lengths such that one has greater width but the other has smaller length are *incomparable*, neither of them being preferred to the other. Figure 2 shows the width-length plane where the set of width-lengths that are preferred to (w, l) is shaded in blue and the set of width-lengths that are less preferred than (w, l) is shaded in green. Width-lengths (w, l) and (w', l') are incomparable in the product order on width-lengths, yet (w, l) is preferred to (w', l') in the shortest-widest order owing to its greater width. The extension on width-lengths is clearly isotone for the product order on them.

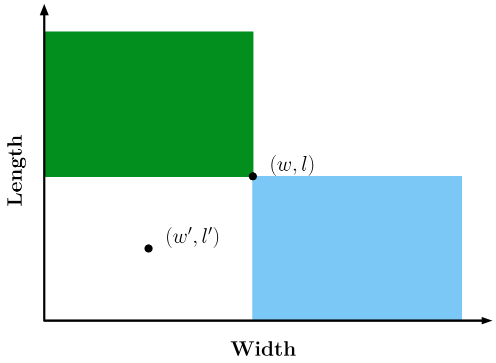


Fig. 2. Product order on width-lengths. The blue area consists of width-lengths that are preferred than (w, l) and the green area of those that are less preferred than (w, l). Width-lengths (w, l) and (w', l') are incomparable on the product order, but (w, l) is preferred to (w', l') on the shortest-widest order.

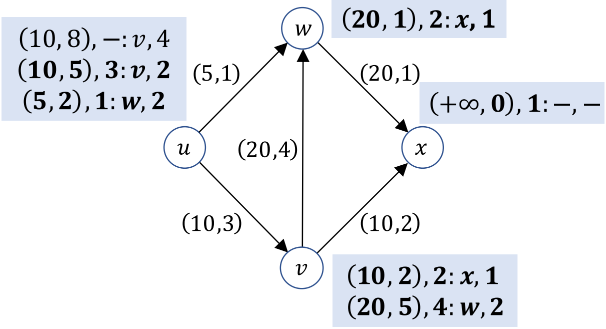


Fig. 3. Stable state of a partial order vectoring protocol operating according to the product order for destination x. Labels guide data-packets along shortest-widest paths or widest-shortest paths as decided by the source.

A width-length in a set of width-lengths is *dominant* if no width-length in the set is preferred to it.[2] A *dominant path* from a source to a destination in a network is one whose width-length is dominant among the width-lengths of all paths from source to destination. Figure 3 shows the same network as Figure 1. The dominant paths from v to x are vx and vwx. Their width-lengths, $(10, 2)$ and $(20, 5)$, respectively, are incomparable. The dominant paths from u to x are uvx and uwx, width-lengths $(10, 5)$ and $(5, 2)$, respectively. Path $uvwx$, the only remaining path from u to x, has width-length $(10, 8)$, which is less preferred than width-length $(10, 5)$ of path uvx.

Consider now a *partial-order vectoring protocol* that operates on the product order on width-lengths, with each node electing and advertising to its in-neighbors the set of dominant width-lengths learned from its out-neighbors. The stable state of such a protocol is shown in Figure 3. Node x elects $(+\infty, 0)$ and w elects $(20, 1)$ as before. Node v learns $(10, 2)$ from x and $(20, 5)$ from w. Since these width-lengths are incomparable, both are elected. Node v differentiates the two width-lengths by assigning them distinct labels [15]. It arbitrarily

[2] In the terminology of order theory, a dominant width-length is a *minimal* width-length.

Authorized licensed use limited to: Tsinghua University. Downloaded on September 12,2025 at 07:14:59 UTC from IEEE Xplore. Restrictions apply.

assigns label 2 to $(10, 2)$ and label 4 to $(20, 5)$. These labels are advertised alongside the associated width-lengths to in-neighbor u in order to enable expedition of data-packets via label-switching.

Therefore, u learns from v both $(10, 5)$ with label 2 and $(10, 8)$ with label 4. Both become candidate width-lengths to reach x via v. From w, u also learns $(5, 2)$ with label 2. The dominant width-lengths of the set $\{(10, 5), (10, 8), (5, 2)\}$ of candidate width-lengths at u are $(10, 5)$ and $(5, 2)$, since these two width-lengths are incomparable, while $(10, 8)$ is less preferred than $(10, 5)$. Node u elects $(10, 5)$ and $(5, 2)$, assigning label 3 to the former and label 1 to the latter.

In Figure 3, an entry at a node of the form

$$(\mathit{width}, \mathit{length})\,,\ \mathit{label}\ :\ \mathit{next.hop}\,,\ \mathit{next.label}$$

reads as follows:

- data-packets generated at the node that need to travel along a path with width-length $(\mathit{width}, \mathit{length})$ are forwarded to out-neighbor $\mathit{next.hop}$ with label $\mathit{next.label}$;
- data-packets arriving at the node from an in-neighbor carrying label label are forwarded to out-neighbor $\mathit{next.hop}$ with the label modified to $\mathit{next.label}$.

With the partial-order vectoring protocol, u is capable of routing data-packets on the shortest-widest path to x. Node u recognizes $(10, 5)$ as the most preferred width-length according to the shortest-widest order among its elected width-lengths. It tags data-packets with label 2 and forwards them to v. At v, incoming label 2 matches the entry pointing to out-neighbor x and outgoing label 1. Consequently, v replaces label 2 with label 1 and forwards the data-packets to x.

A *widest-shortest path* is a path of maximum width among those of minimum length from source to destination in a network [19]. Widest-shortest paths are selected according to the *widest-shortest order* (colexicographic order), which establishes that width-length (w, l) is preferred to width-length (w', l') if its length is smaller, $l < l'$, or the lengths are equal and its width is greater, $l = l'$ and $w > w'$.

The partial-order vectoring protocol allows for routing a flow of data-packets on a shortest-widest path or on a widest-shortest path as is more appropriate for that specific flow. Going back to Figure 3, suppose that u now wants to send data-packets to x along a widest-shortest path. Node u recognizes $(5, 2)$ as the most preferred width-length according to the widest-shortest order among its elected width-lengths. Reading from the entry corresponding to $(5, 2)$, u tags data-packets with label 2 and forwards them to w, which replaces label 2 with label 1 for delivery to x.

In summary, in this section, we: (1) presented an instance of optimal paths that a standard vectoring protocol does not route on and ascribed this shortcoming to the lack of isotonicity; (2) solved the problem of routing on optimal paths with a new kind of vectoring protocol that operates according to a partial order that is included in the original total order defining optimality and satisfies isotonicity; (3) and explored the new kind of vectoring protocol to route concurrently on the optimal paths determined by different optimality criteria. This overall approach is asserted with generality and rigor in Section IV and Section V. But before, in the next section, we discuss its significance to the inter-AS routing context.

III. Inter-AS Routing

Inter-AS routing is mostly decided by the combination of policy choices taken by ASes. Since these choices are ultimately implemented in BGP in terms of ranking of paths and of rules for their importation and exportation, they can be modeled by a routing algebra with its inherent notion of optimality. We illustrate how partial-order vectoring protocols overcome the limitations of BGP in achieving optimal goals for inter-AS routing and how they open up the range of policy choices available to the ASes.

The baseline model for inter-AS routing posits that two neighbor ASes establish either a customer-provider or a peer-peer relationship. A customer AS pays a provider AS to transit its traffic to and from the rest of the internet, while two peer ASes transit each other's and their customer's traffic without mutual compensations [20], [21]. An AS-path is *valid* if, and only if, every intermediate AS is paid to transit traffic, implying that for each such AS at least one of its two neighbors along the path is a customer of the AS. Consequently, valid AS-paths are composed of a sequence of customer-to-provider links (possibly empty) followed by a maximum of one peer-to-peer link followed by a sequence of provider-to-customer links (possibly empty) [22]. Data-packets are forwarded exclusively on valid AS-paths.

The *type* of a valid AS-path is *customer*, *peer*, or *provider*, respectively, if its first link is a provider-to-customer link, a peer-to-peer link, or a customer-to-provider link. Type customer is preferred to type peer, which is preferred to type provider [3].[3] Types customer, peer, and provider are denoted by the letters C, R, and P, respectively. Every valid AS-path in an internet is characterized by a pair *type-length* belonging to the Cartesian product of the set of types with the set of nonnegative lengths, where length represents the number of links (hops) in the AS-path. The extension of type-length $(\alpha, 1)$, corresponding to link uv, with type-length (α', n'), corresponding to valid AS-path P starting at node v, yields type-length $(\alpha, n' + 1)$ if, and only if, $\alpha = \mathrm{P}$ or $\alpha' = \mathrm{C}$, corresponding to valid AS-path uvP: the equality $\alpha = \mathrm{P}$ means that u is a customer of v, while the equality $\alpha' = \mathrm{C}$ means that the second AS along P is a customer of v.

A *shortest-best-type path* is a valid AS-path of minimum length among those of most preferred type from a source to a destination in an internet. Shortest-best-type paths are selected according to the *shortest-best-type order* (lexicographic order), which establishes that type-length (α, n) is preferred to type-length (α', n') if its type α is preferred to type α', or the types are the same but its length is smaller, $\alpha = \alpha'$ and $n < n'$.

Current inter-AS routing has BGP compute paths according to the shortest-best-type order, but fails to route on shortest-best-type paths, in general. This failure is illustrated with the internet of Figure 4. A solid arrow joins a customer and

[3] For simplicity of presentation, but without loss of generality, we assume that type peer is preferred to type provider although this relative preference is not necessary [3].

Authorized licensed use limited to: Tsinghua University. Downloaded on September 12,2025 at 07:14:59 UTC from IEEE Xplore. Restrictions apply.

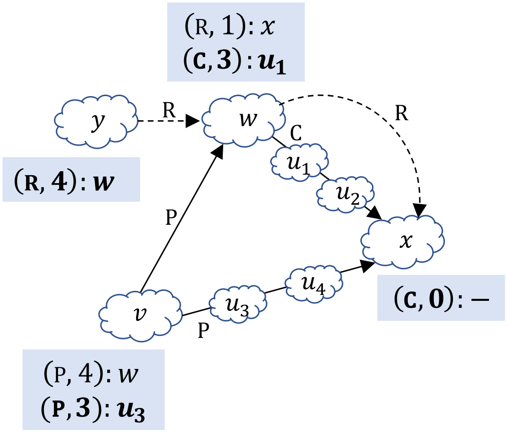


Fig. 4. Stable state of BGP for destination AS x. Links and paths are annotated with their type. Elected type-lengths are in bold.

a provider, with the provider drawn above the customer, and a dashed arrow joins two peers. The arrows point in the direction of the flow of data-packets toward destination AS x. AS w is a peer of AS y. AS x is a peer of AS w and there is a customer path from AS w to AS x through AS u_1 and AS u_2. AS w is a provider of AS v and there is a provider path from AS v to AS x through AS u_3 and AS u_4. All paths from AS v to AS x are provider paths: vwx with length 2; vu_3u_4, with length 3; and vwu_1u_2 with length 4. The shortest-best-type path from AS v to AS x is vwx.

AS w elects type-length $(\mathrm{C}, 3)$, corresponding to customer path wu_1u_2x, to the detriment of type-length $(\mathrm{R}, 1)$, corresponding to the direct peer path wx, because type C is preferred to type R. Consequently, AS v learns $(\mathrm{P}, 4)$ from AS w, which is the extension of $(\mathrm{P}, 1)$ of link vw with $(\mathrm{C}, 3)$ elected at AS w. It also learns $(\mathrm{P}, 3)$ from AS u_3. It elects the latter type-length, since it has the same type but smaller length than that of $(\mathrm{P}, 4)$. Hence, data-packets from AS v to AS x travel along provider path vu_3u_4x, which is not the shortest-best-type path from AS v to AS x. As in the case of shortest-widest paths discussed in the previous section, non-optimality is imputed to the failure of isotonicity. Specifically, $(\mathrm{C}, 3)$ is preferred to $(\mathrm{R}, 1)$, but $(\mathrm{P}, 4)$, which is the extension of $(\mathrm{P}, 1)$ with $(\mathrm{C}, 3)$, is less preferred than $(\mathrm{P}, 2)$, which is the extension of $(\mathrm{P}, 1)$ with $(\mathrm{R}, 1)$. In turn, $(\mathrm{P}, 4)$ is less preferred than $(\mathrm{P}, 3)$, so that AS v forwards data-packets to AS u_3, rather than to AS w, on their way to AS x.

Consider now the *product order on type-lengths*, which is such that type-length (α, n) is preferred to type-length (α', n') if it is different from (α', n') and both its type equals or is preferred to that of (α', n') and its length equals or is smaller than that of (α', n'). The extension on type-lengths is isotone for the product order on them. Figure 5 shows the same internet as Figure 4. There are two dominant paths from AS w to AS x according to the product order on type-lengths: wu_1u_2x and wx. Their type-lengths, respectively, $(\mathrm{C}, 3)$ and $(\mathrm{R}, 1)$ are incomparable. With a partial-order vectoring protocol, AS w elects both $(\mathrm{C}, 3)$ and $(\mathrm{R}, 1)$. AS v learns from AS w both $(\mathrm{P}, 4)$ and $(\mathrm{P}, 2)$, rather than just $(\mathrm{P}, 4)$, as in the case of BGP. In addition, AS v learns $(\mathrm{P}, 3)$ from AS u_3 as before. All three type-lengths are candidates to reach AS x with common type P. The elected type-length is $(\mathrm{P}, 2)$ corresponding to shortest-best-type path vwx.

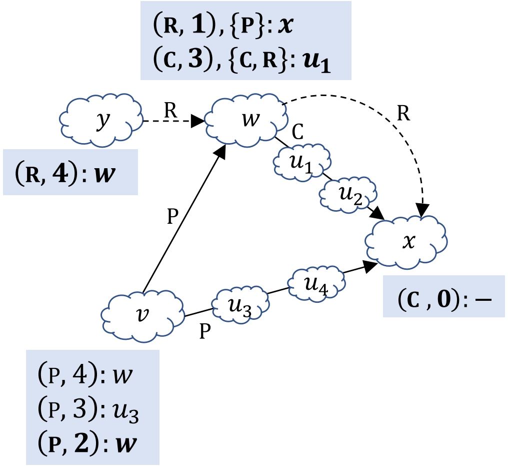


Fig. 5. Stable state of a partial-order vectoring protocol operating according to the product order on type-lengths for destination x. The nature of the relationship between neighbor ASes determines data-packet forwarding.

For the case of the product order on type-lengths, the nature of the relationship between neighbor ASes suffices to guide data-packets along dominant paths, dispensing with labels [23]. Data-packets arriving through provider links (from customers) match the elected type-length of smaller length, while data-packets arriving through customer and peer links (from providers and peers) match the elected type-length of type customer. In the figure, the two type-lengths elected at AS w are discriminated by the types of incoming links using them. Data-packets arriving at AS w from AS v match type-length $(\mathrm{R}, 1)$ and are forwarded directly to AS x, whereas those arriving from AS y are forwarded along customer path wu_1u_2x to AS x.

The partial-order vectoring protocol operating according to the product order on type-lengths provides embedded incentives for uncoordinated adoption by ASes, which is of paramount relevance in a decentralized internet. Stub ASes, which are those without customers and constitute the majority of the ASes of the Internet, can keep using BGP, electing a single type-length per destination. Internet Service Providers (ISPs), which are those ASes with customers, can individually deploy a partial-order vectoring protocol to their own advantage. By offering a shorter valid AS-path option to their customers, ISPs profit from transiting their traffic to a wider set of destinations. In the internet of Figure 5, the adoption of the partial-order vectoring protocol at AS w causes AS v to transit its traffic to destination AS x through AS w rather than through AS u_3.

A *best-type-shortest path* is a valid AS-path of minimum length with ties broken by type. These paths are selected according to the *best-type-shortest order* (colexicographic order), which establishes that type-length (α, n) is preferred to type-length (α', n') if it has smaller length, $n < n'$, or it has the same length and a preferred type, $n = n'$ and α is preferred to α'. Since all paths from AS v to AS x in the internet of Figure 5 are provider paths, the best-type-shortest and the shortest-best-type paths from AS v to AS x coincide in path vwx. Such coincidence in types does not occur between the two paths from AS w to AS x. The best-type-shortest path from AS w to AS x is peer path wx with unit length, whereas the shortest-best-type path from AS w to AS x is customer path wu_1u_2x with length 3. Routing its own traffic on the latter path, AS w is paid by AS u_1. However, the business of ASes is to profit from transit traffic, not from traffic originated locally. AS w may be better off routing the traffic originated locally on the peer link to AS x.

IV. From Optimality to Dominance

We transform the problem of routing on optimal paths into that of routing on dominant paths. Section IV-A reviews routing algebras with an emphasis on isotonicity. Section IV-B introduces the possibility of reducing an optimality criterion that does not satisfy isotonicity to a dominance criterion that satisfies it. Section IV-C consolidates a collection of optimality criteria into a single dominance criterion that respects all criteria of the collection. Section IV-D introduces an algebraic property that guarantees the termination of the partial-order vectoring protocol that we design in Section V.

A. Routing Algebras and Optimality

The definition of optimal path and the computation of optimal paths can be formalized with a routing algebra $(S, \oplus, \preceq)$ consisting of a set S of *attributes*, a *binary extension operation* $\oplus$ on attributes, and a *total order* $\preceq$ on attributes [4], [5], [7]–[9]. Attributes represent arbitrary performance metrics and/or policy choices in context. Every link and path in a network is associated with an attribute. The binary extension operation $\oplus$ allows computation of the attribute of a path from the attributes of its constituent links. Contrary to our previous work [18], we do not assume that the binary extension operation is either associative or commutative. The reason is twofold. On the one hand, associativity and commutativity of the extension operation are never invoked or implied in the computations performed by the vectoring protocols considered. On the other hand, forsaking associativity and commutativity encompasses more optimal path problems. For instance, the extension operation on type-lengths defined in Section III is neither associative nor commutative.

As a convention, in the absence of parenthesis, extension operations are performed from right to left in a sequence of such operations. Thus, for example, $a \oplus b \oplus c$ is meant to represent $a \oplus (b \oplus c)$.[4] Without loss of generality, we assume there is a *neutral attribute* ϵ, which is a right-identity for $\oplus$: $a \oplus \epsilon = a$ for all $a \in S$.

The attribute of link uv is denoted by $a[uv]$. The attribute of path $P = u_0u_1 \cdots u_{n-1}u_n$ is denoted by $a[P]$ and computed from

$$a[P] = a[u_0u_1] \oplus \cdots \oplus a[u_{n-2}u_{n-1}] \oplus a[u_{n-1}u_n].$$

The attribute of a trivial path, consisting of a single node, is ϵ.

The total order $\preceq$ establishes a relative preference among attributes. It is a binary relation on attributes that satisfies antisymmetry, transitivity, and connexity. Connexity means that $a \preceq b$ or $b \preceq a$ for all $a, b \in S$. We write $a \prec b$ for $a \preceq b$ and $a \neq b$, and say that a is *preferred* to b and that b is *less preferred* than a. We assume there is a *null attribute* $\bullet$ that is the least preferred of all attributes and represents the absence of a valid path.

The *optimal attribute* of a set of attributes is the most preferred attribute of the set. The optimal attribute from a source to a destination in a network is the optimal attribute of the set of all attributes of paths from source to destination and an optimal path is a path with such an attribute.

Definition 1: Binary extension operation $\oplus$ is *left-isotone* for total order $\preceq$ if

$$a \preceq b \text{ implies } c \oplus a \preceq c \oplus b \text{ for all } a, b, c \in S;$$

it is *right-isotone* for total order $\preceq$ if

$$a \preceq b \text{ implies } a \oplus c \preceq b \oplus c \text{ for all } a, b, c \in S.$$

Binary extension operation $\oplus$ is *isotone* for total order $\preceq$ if it is both left- and right-isotone for $\preceq$.

Left-isotonicity implies that given any two paths either the relative preference between their attributes is preserved when they are prefixed by an arbitrary common link or their attributes become equal when they are prefixed by the common link [4]–[6], [8]. This key algebraic property determines optimality of standard vectoring protocols. If left-isotonicity holds, then a standard vectoring protocol routes on optimal paths; if it does not hold, then a standard vectoring protocol does not route on optimal paths, in general [5], [6], [11], [24]. Many optimality criteria of practical interest do not satisfy left-isotonicity. The shortest-widest order on width-lengths, Section II, and the shortest-best-type order and the best-type-shortest order on type-lengths, Section III, are just three examples where left-isotonicity does not hold.

We do not have a need for right-isotonicity in this paper, but will use the term isotonicity instead of left-isotonicity when the extension operation is both left- and right-isotone for the total order.

B. Partial Orders and Isotonic Reductions

A *partial order* $\preceq$ on attributes is an antisymmetric, transitive, and reflexive binary relation on attributes. Reflexivity means that $a \preceq a$ for all $a \in S$. Connexity implies reflexivity, so that a total order is a particular case of a partial order. If $a \preceq b$ or $b \preceq a$, then a and b are *comparable*; otherwise they are *incomparable*. We still write $a \prec b$ for $a \preceq b$ and $a \neq b$, and say that a is *preferred* to b and that b is *less preferred* than a. A *dominant attribute* of a set of attributes is defined such that no other attribute of the set is preferred to it. A dominant attribute from a source to a destination in a network is a dominant attribute of the set of all attributes of paths from source to destination and a dominant path is a path with such an attribute. The definitions of left-isotonicity

[4] The reason for this convention is that the sequence of link attributes of a path is written in the direction of the flow of data-packets, from source to destination, while a vectoring protocol performs extensions operations in the opposite direction, from destination to source.

Authorized licensed use limited to: Tsinghua University. Downloaded on September 12,2025 at 07:14:59 UTC from IEEE Xplore. Restrictions apply.

and isotonicity given in the previous section apply to partial orders as well as to total orders.

Definition 2: A *left-isotonic reduction* of a total order $\preceq$ on attributes is a partial order contained in $\preceq$ for which $\oplus$ is left-isotone [13].

Left-isotonic reductions trade connexity for isotonicity. Let $\preceq^R$ be a left-isotonic reduction of total order $\preceq$. Let $a(s,t)$ be the optimal attribute from s to t in a network according to $\preceq$, and $A^R(s,t)$ be the set of dominant attributes from s to t in the same network according to $\preceq^R$. Since $a \preceq^R b$ implies $a \preceq b$, the optimal attribute from s to t belongs to the set of dominant attributes from s to t, $a(s,t) \in A^R(s,t)$. Thus, an idea to obtain optimal attributes is to leverage the left-isotonicity of a left-isotonic reduction to compute dominant attributes with an effective routing protocol (Section V) and then find the optimal attributes among the computed dominant attributes.

The more attributes that are comparable within a partial order, the smaller the sets of dominant attributes and the more efficient the routing protocols. Hence, we aspire to left-isotonic reductions which are as large as possible.

Theorem 1: For every total order on attributes $\preceq$, the binary relation on attributes $\preceq^{R^*}$ defined by $a \preceq^{R^*} b$ if

$$x_n \oplus \cdots \oplus x_1 \oplus a \preceq x_n \oplus \cdots \oplus x_1 \oplus b,$$

for every nonnegative n and every set of attributes $x_1, \ldots, x_n$, is the largest left-isotonic reduction of $\preceq$.

Proof: Binary relation $\preceq^{R^*}$ is a partial order. (1) Reflexivity: $a \preceq^{R^*} a$, because $x_n \oplus \cdots \oplus x_1 \oplus a \preceq x_n \oplus \cdots \oplus x_1 \oplus a$, for every nonnegative n and every set of attributes $x_1, \ldots, x_n$. (2) Antisymmetry: $a \preceq^{R^*} b$ and $b \preceq^{R^*} a$ imply $a = b$, because $a \preceq^{R^*} b$ implies $a \preceq b$, $b \preceq^{R^*} a$ likewise implies $b \preceq a$, and $a \preceq b$ together with $b \preceq a$ imply $a = b$. (3) Transitivity: $a \preceq^{R^*} b$ and $b \preceq^{R^*} c$ imply $a \preceq^{R^*} c$, because $x_n \oplus \cdots \oplus x_1 \oplus a \preceq x_n \oplus \cdots \oplus x_1 \oplus b$ and $x_n \oplus \cdots x_1 \oplus b \preceq x_n \oplus \cdots \oplus x_1 \oplus c$ together imply $x_n \oplus \cdots \oplus x_1 \oplus a \preceq x_n \oplus \cdots \oplus x_1 \oplus c$, for every nonnegative n and every set of attributes $x_1, \ldots, x_n$, which implies $a \preceq^{R^*} c$.

Binary extension operation $\oplus$ is left-isotone for $\preceq^{R^*}$. The inequality $a \preceq^{R^*} b$ implies $y_n \oplus \cdots \oplus y_1 \oplus (c \oplus a) \preceq y_n \oplus \cdots \oplus y_1 \oplus (c \oplus b)$, for every attribute c, every nonnegative n, and every set of attributes $y_1, \ldots, y_n$, which implies $c \oplus a \preceq^{R^*} c \oplus b$ for all $c \in S$.

Partial order $\preceq^{R^*}$ is the largest left-isotonic reduction of $\preceq$ on attributes. In order to arrive at a contradiction, suppose that there is a partial order $\preceq^R$ contained in $\preceq$, but not contained in $\preceq^{R^*}$. Therefore, there are attributes a' and b' such that $a' \preceq^R b'$, while it is not the case that $a' \preceq^{R^*} b'$. Since it is not the case that $a' \preceq^{R^*} b'$, there is a set of attributes $x_1, \ldots, x_n$ for some nonnegative n for which it is not the case that $x_n \oplus \cdots \oplus x_1 \oplus a' \preceq x_n \oplus \cdots \oplus x_1 \oplus b'$. On the other hand, because $\oplus$ is left-isotone for $\preceq^R$, we have $x_n \oplus \cdots \oplus x_1 \oplus a' \preceq^R x_n \oplus \cdots \oplus x_1 \oplus b'$, which implies $x_n \oplus \cdots \oplus x_1 \oplus a' \preceq x_n \oplus \cdots x_1 \oplus b'$: a contradiction was arrived at. ■

The characterization of largest left-isotonic reduction becomes simpler if the extension operation $\oplus$ is associative.

Theorem 2: Suppose that the extension operation $\oplus$ is associative. Then, the largest left-isotonic reduction $\preceq^{R^*}$ of total order $\preceq$ is such that $a \preceq^{R^*} b$ if, and only if, $a \preceq b$ and $x \oplus a \preceq x \oplus b$ for every attribute x.

Proof: Assume that $a \preceq b$ and $x \oplus a \preceq x \oplus b$ for every attribute x. Let $x_1, \ldots, x_n$ be an arbitrary sequence of n attributes, with n positive. Since $\oplus$ is associative, we write $(x_n \oplus \cdots \oplus x_1) \oplus a = x_n \oplus \cdots \oplus x_1 \oplus a$ and $(x_n \oplus \cdots \oplus x_1) \oplus b = x_n \oplus \cdots \oplus x_1 \oplus b$. Thus,

$$(x_n \oplus \cdots \oplus x_1) \oplus a \preceq (x_n \oplus \cdots \oplus x_1) \oplus b,$$

is equivalent to

$$x_n \oplus \cdots \oplus x_1 \oplus a \preceq x_n \oplus \cdots \oplus x_1 \oplus b,$$

which, by Theorem 1, implies $a \preceq^{R^*} b$.

Conversely, assume that $a \preceq^{R^*} b$. By Theorem 1, we have $a \preceq b$ $(n = 0)$ and $x_1 \oplus a \preceq x_1 \oplus b$ for every attribute x_1 $(n = 1)$. ■

Examples of left-isotonic reductions. The largest left-isotonic reduction of the shortest-widest order is the product order on width-lengths, Section II. The largest left-isotonic of the shortest-best-type order is the product order on type-lengths, Section III. The proofs of these facts are easy and are omitted.

We now study a class of optimality criteria on width-lengths for which the largest left-isotonic reduction is strictly larger than the product order on width-lengths. The time required to convey a file of size K along a path with capacity w and propagation delay l is $K/w + l$. A K*-quickest path* is a path that minimizes the time required to convey a file of size K, with ties broken by the largest capacity. K-quickest paths are selected according to the K*-quickest order*, denoted by $\preceq_K$, which establishes that (w, l) is preferred to (w', l') if either $K/w + l < K/w' + l'$, or $K/w + l = K/w' + l'$ and $w > w'$. Incidentally, the operation of EIGRP [10] is based on a K-quickest order.

The binary extension operation on width-lengths is commutative. Thus, we talk of isotonicity instead of left-isotonicity. Except for $K = 0$, the K-quickest order does not satisfy isotonicity. Its largest isotonic reduction is given in the following proposition.

Proposition 1: The largest isotonic reduction of the K-quickest order $\preceq_K$ is such that width-length (w, l) equals or is preferred to width-length (w', l') if, and only if, $(w, l) \preceq_K (w', l')$ and $l \leq l'$.

The proof of Proposition 1 is presented in Appendix A. Figure 6 depicts the largest isotonic reduction of the K-quickest-order in the width-length plane for some value of K. Width-lengths (w, l) and (w', l') dotted in the figure are incomparable because $(w, l) \preceq_K (w', l')$, but the extension of width-length $(w', 1)$ with each of them yields $(w', l + 1)$ and $(w', l' + 1)$, respectively, and it is not the case that $(w', l + 1) \preceq_K (w', l' + 1)$. (Note that $(w', l + 1) \preceq_K (w', l' + 1)$ is equivalent to $l \leq l'$.) Comparing Figure 6 with Figure 2, we readily see that more pairs width-length are comparable in the largest isotonic reduction of the K-quickest order than in the product order on width-lengths.

Authorized licensed use limited to: Tsinghua University. Downloaded on September 12,2025 at 07:14:59 UTC from IEEE Xplore. Restrictions apply.

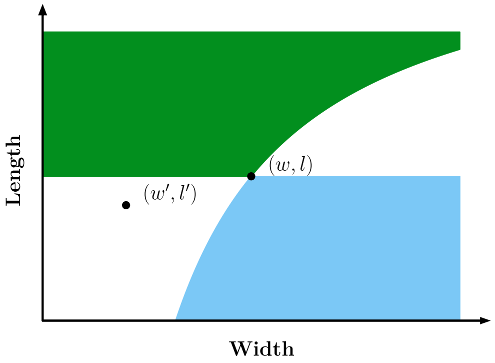


Fig. 6. Largest isotonic reduction of the K-quickest order for some value of K. The blue area consists of width-lengths that are preferred to (w,l) and the green area of those that are less preferred than (w,l). Width-lengths (w,l) and (w',l') are incomparable, although $(w,l) \preceq_K (w',l')$ (compare with Figure 2).

C. Intersection of Multiple Orders

Different total orders on a common set of attributes S with a common extension operation $\oplus$ imply distinct optimality criteria. Let O be a collection of total orders on S.

Definition 3: The *intersection* of total orders $\preceq_i$, $i \in O$, is the binary relation $\preceq^O$ such that $a \preceq^O b$ if $a \preceq_i b$ for all $i \in O$ and all $a, b \in S$.

The binary relation $\preceq^O$ is a partial order which is contained in each of the total orders of the collection. Two attributes are comparable in $\preceq^O$ if, and only if, one of them is preferred to the other for every optimality criteria i. Let $a_i(s,t)$ be the optimal attribute from s to t in a network according to $\preceq_i$ and $A^O(s,t)$ be the set of dominant attributes from s to t according to $\preceq^O$. Since $a \preceq^O b$ implies $a \preceq_i b$, we have that $a_i(s,t) \in A^O(s,t)$ for all $i \in O$.

The concept of largest left-isotonic reduction applies to partial orders as well as to total orders and it is easily shown that the largest left-isotonic reduction of the intersection of partial orders equals the intersection of their left-isotonic reductions. When $\oplus$ is not left-isotone for $\preceq^O$, we take its largest left-isotonic reduction. Therefore, the problem of computing optimal attributes for a collection of optimality criteria is transformed into the problem of computing sets of dominant attributes according to the largest left-isotonic reduction of the intersection of all criteria. The optimal attributes of each criteria can be found among the dominant attributes.

Examples of intersections. The intersection of the shortest-widest order and the widest-shortest order is the product order on width-lengths, Section II, while the intersection of the shortest-best-type order and the best-type-shortest order is the product order on type-lengths, Section III.

The next proposition shows that the intersection of the k-quickest orders on width-lengths for all nonnegative k smaller than or equal to K coincides with the largest isotonic reduction of the K-quickest order.

Proposition 2: The intersection of the k-quickest orders $\preceq_k$ for all k between 0 and K is such that (w,l) equals or is preferred to width-length (w',l') if, and only if, $(w,l) \preceq_K (w',l')$ and $l \leq l'$.

The proof of Proposition 2 is presented in Appendix B. The intersection of the k-quickest orders on width-lengths for all nonnegative k is the product order on width-lengths.

D. Strictly Inflationary Circuits

Left-isotonicity guarantees that vectoring protocols compute optimal or dominant attributes, as the case may be, under the assumption that these protocols terminate in a state devoid of loops. It does not, however, validate this latter assumption. We introduce an algebraic property concerning network circuits that together with left-isotonicity guarantees protocol termination in a state devoid of loops and, thus, in a state where optimal or dominant attributes are computed by the protocol. But first it is convenient to introduce some terminology to cope with the possible non-associativity of the extension operation $\oplus$. Given a path P, we denote by $\hat{a}[P]\oplus$ the function that takes an attribute for argument and gives as result the attribute obtained from a succession of extensions operations with the attributes of the links of P, from right to left, starting with the attribute taken as argument. For instance, if $P = uvw$, then $\hat{a}[P] \oplus b = a[uv] \oplus (a[vw] \oplus b)$ for every attribute b, which we previously agreed to denote simply by $a[uv] \oplus a[vw] \oplus b$. If the extension operation is associative, then $\hat{a}[P] \oplus b = a[P] \oplus b$.

Definition 4: A circuit C is *strictly left-inflationary* if

$$b \prec \hat{a}[C] \oplus b \text{ for all } b \in S - \{\bullet\}.$$

Strict left-inflation of a circuit means that an arbitrary non-null attribute is preferred to its extension around the circuit. An immediate consequence of strict left-inflation is embodied in the next theorem.

Theorem 3: If left-isotonicity holds and all circuits in a network are strictly left-inflationary, then every dominant attribute from a source to a destination in the network is the attribute of a simple path.

Proof: We start by showing that the attribute of any path containing a circuit either equals or is less preferred than the attribute of the path obtained through removal of the circuit. Let PCQ be a path from a source s to a destination t that contains circuit C. Because of strict left-inflation of C, we write

$$a[Q] \prec \hat{a}[C] \oplus a[Q] = a[CQ],$$

and because of left-isotonicity, we obtain

$$a[PQ] = \hat{a}[P] \oplus a[Q] \preceq \hat{a}[P] \oplus a[CQ] = a[PCQ],$$

showing that the attribute of path PCQ either equals or is less preferred than the attribute of path PQ.

Inductively applying the argument above over the circuits contained in a path from s to t, we conclude that the attribute of every non-simple path from s to t equals or is less preferred than the attribute of a simple path from s to t. ■

V. Partial Order Vectoring Protocol

A standard vectoring protocol, such Routing Information Protocol (RIP) [25], EIGRP [10], or BGP [12], instantiates a separate routing computation per destination. The computation is initiated with the destination advertising the neutral attribute ϵ to all its in-neighbors. Every node maintains candidate attributes to reach the destination via each of its out-neighbors and elects the most preferred attribute among the candidates, forwarding data-packets aimed at the destination to the corresponding out-neighbors. Whenever a node receives an advertisement from an out-neighbor, it sets the candidate attribute learned from the out-neighbor to the extension of the attribute of the link to the out-neighbor with the advertised attribute. If, as a consequence, the elected attribute has changed, then it is advertised to all in-neighbors.

A standard vectoring protocol can be generalized to work with a partial order on attributes. Let $\mathcal{D}_{\preceq}(A)$ denote the subset of dominant attributes of set A according to partial order $\preceq$,

$$\mathcal{D}_{\preceq}(A) = \{a \in A \mid \text{there is no } x \in A \text{ such that } x \prec a\}.$$

In the canonical partial-order vectoring protocol, destination t originates singleton $\{\epsilon\}$, which it advertises to all its in-neighbors. Algorithm 1 presents the pseudo-code for when node u, $u \neq t$, receives a set B of attributes advertised by its out-neighbor v pertaining to destination t. Variable $C_u[v,t]$ stores the set of candidate attributes to reach t via out-neighbor v and variable $E_u[t]$ stores the set of elected attributes to reach t.

Algorithm 1 Canonical Partial-Order Vectoring Protocol. Node u Receives Set B of Attributes From Out-Neighbor v to Reach t

```
1: $C_u[v,t] := \{a[uv] \oplus b \mid b \in B\}$
2: $E_u[t] := \mathcal{D}_{\preceq}(\{a \mid a \in C_u[v,t] \text{ with } v \text{ an out-neighbor}\})$
3: if $E_u[t]$ has changed then
4:     for all $r$ an in-neighbor do
5:         send $E_u[t]$ to $r$
```

When u receives set B from v, it first computes the set of attributes learned from v, where each such attribute results from the extension of the attribute of the link to v with an attribute contained in B (line 1). Then, u finds its own new set of elected attributes as the subset of dominant attributes of the union of all sets of candidate attributes learned from each of its out-neighbors (line 2). If there is a change in the set of elected attributes, then u advertises this set to all its in-neighbors (lines 3–5).

Each node assigns a unique label to each of its elected attributes that is advertised to in-neighbors alongside the attribute [15]. Therefore, for a given destination, each node maintains a table with entries of the form

$$attribute\ , label : next.hop\ , next.label.$$

The table is used as follows:

- data-packets generated at the node that need to travel along a path with attribute $attribute$, presumably an optimal path according to some optimality criterion, are forwarded to out-neighbor $next.hop$ with label $next.label$;
- data-packets arriving at the node from an in-neighbor carrying label $label$ are forwarded to out-neighbor $next.hop$ with the label modified to $next.label$.

A node may install multiple entries with a common value of $attribute$. This allows for routing data-packets along multiple dominant paths with a common attribute, a possibility that in standard vectoring protocols is known as ECMP (Equal Cost Multi-Path).

Termination and dominance. A *stable state* is a state without advertisements in transit in any of the links of the network. The partial-order vectoring protocol *terminates* if, in the absence of changes in the network, it reaches a stable state from any initial state. As their standard counterparts, partial-order vectoring protocols are prone to count-to-infinity if the set of all possible paths attributes in the network is infinite [26]. When this condition is not met in a specific routing context, it can be enforced by including a hop-count field in attributes and invalidating paths with hop-count in excess of some prespecified maximum value, as in RIP; or by including a field in attributes that records the path traversed by the sequence of advertisements away from the destination and invalidating looping advertisements, as in BGP.

Theorem 4: If left-isotonicity holds, all circuits in the network are strictly left-inflationary, and the set of all path attributes is finite, then the partial-order vectoring protocol terminates.

We do not prove the theorem here due to space limitations and because the structure and ingredients of the proof can be found in the proof of the cognate theorem that we presented in [18]. However, we mention two differences between Theorem 4 and the theorem from [18]. First, Theorem 4 is premised on the algebra being left-isotone, whereas the theorem from [18] is premised on the algebra being inflationary, as defined in that paper. Second, the present theorem does not assume associativity or commutativity of the extension operation. This does not raise any concerns, since the proof given in [18] does not rely on these two properties.

Theorem 5: If left-isotonicity holds and all circuits in the network are strictly left-inflationary, then the partial-order vectoring protocol elects dominant attributes in stable state.

The proof of the cognate theorem in [18] applies here as well, since it is also not reliant on either associativity or commutativity of the extension operation.

VI. Evaluation of Routing on Widths and Lengths

Our evaluation of routing on widths and lengths intends to answer two main questions. How large are the sets of dominant width-lengths in realistic networks? How does the convergence behavior of a partial-order vectoring protocol compare with that of a standard vectoring protocol? These questions are addressed in Sections VI-A and VI-B, respectively.

The test networks used for evaluation are based on the largest biconnected components of the ISP topologies inferred by the Rocketfuel project [16]. Every link in a topology is annotated with both an Open Shortest Path First (OSPF) weight and a propagation delay. A width was assigned to each link that is equal to the inverse of its weight, since, by default, OSPF weights are set as inverse capacities; a length was assigned to each link that is equal to its propagation delay. We present results for AS 1239, which is the largest of the

Authorized licensed use limited to: Tsinghua University. Downloaded on September 12,2025 at 07:14:59 UTC from IEEE Xplore. Restrictions apply.

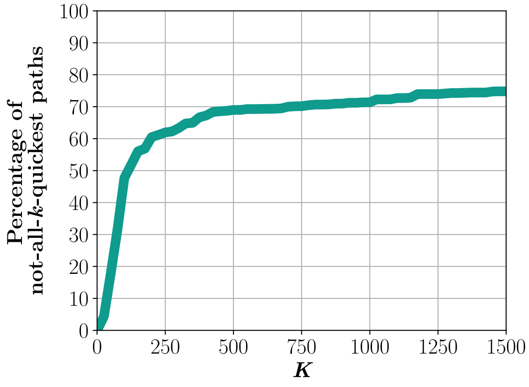


Fig. 7. Percentage of source-destination pairs in AS 1239 for which a standard vectoring protocol operating on the K-quickest order does not compute k-quickest paths for some k, $0 \leq k \leq K$.

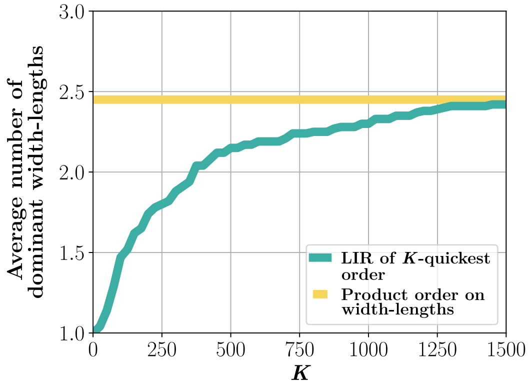


Fig. 8. Average number of dominant width-lengths in AS 1239 for the LIR of the K-quickest order, as a function of K, and for the product order on width-lengths.

Rocketfuel topologies, having worldwide coverage with clusters of nodes in the US, Europe, and Southeast-Asia/Australia. AS 1239 has 284 nodes and 1882 links. Lengths range from 1 to 64 and widths from 7 to 50.

A. Sets of Dominant Width-Lengths

Figure 7 plots the percentage of paths computed with a standard vectoring protocol operating on the K-quickest order (Section IV-B) which are not k-quickest paths for at least one value of k, $0 \leq k \leq K$, over all source-destination pairs in the network. For instance, at $K = 750$ around 70% of the computed paths are not k-quickest paths for some k.

In contrast to the standard vectoring protocol operating on the K-quickest order, the partial-order vectoring protocol operating on the Largest Isotonic Reduction (LIR) of the K-quickest order is able to produce k-quickest paths for all k, $0 \leq k \leq K$, over all source-destination pairs.

Figure 8 plots the average number of dominant width-lengths for the LIR of the K-quickest order as a function of K, over all source-destination pairs in the network. The asymptote belongs to the product order on width-lengths.

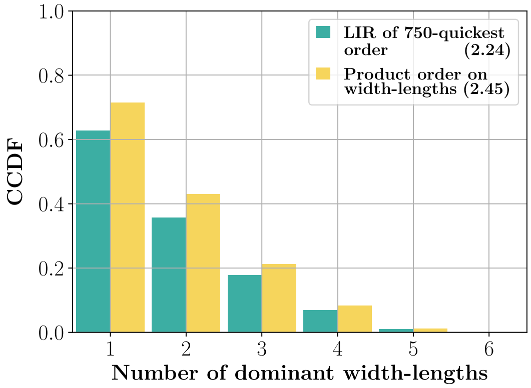


Fig. 9. CCDF of the number of dominant width-lengths in AS 1239 for the LIR of the 750-quickest order and for the product order on width-lengths.

As expected, the average number of dominant width-lengths increases with K. For K greater than 1000, this number is within 5% of the asymptotic value of 2.45. Figure 9 plots the Complementary Cumulative Distribution Function (CCDF) of the number of dominant width-lengths for the LIR of the 750-quickest order and for the product order on width-lengths. The percentages of source-destination pairs connected by more than three dominant width-lengths are 17.8% and 21.2%, respectively, for the LIR of the 750-quickest order and for the product order on width-lengths. For both orders, no source-destination pair is connected by more than seven width-lengths.

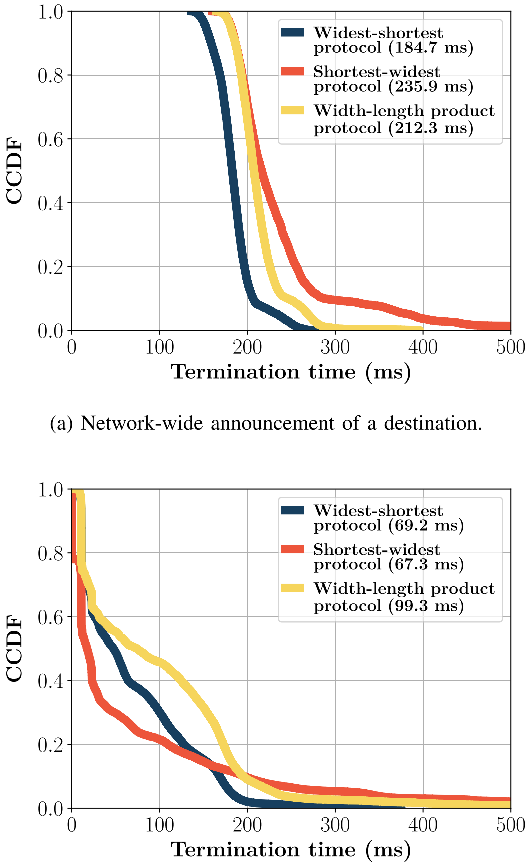
The number of dominant width-lengths from a source to a destination in a given network is upper bounded by the number of distinct widths among the links of the network. AS 1239 exhibits 19 distinct widths. Figures 8 and 9 show that the numbers of dominant width-lengths in AS 1239 are well below the upper bound of 19, corroborating the practicality of routing based on partial orders.

B. Termination Times

The *termination time* of a vectoring protocol is defined as the duration of the interval of time elapsed from the moment a network event occurs until the protocol reaches a stable state, which is a state without advertisements in transit. Two types of network events are considered: the network-wide announcement of a destination and the failure of a link. We built a simulator of vectoring protocols according to which advertisements traverse every link first in, first out subject to the propagation delay of the link plus a random delay taken from a uniform distribution. The propagation delay of a link equals its length in ms and the uniform distribution ranges from 0 to 10 ms. As a measure against count-to-infinity, advertisements that travel more than a prespecified maximum number of hops are invalidated. We set that maximum number to 20. For every network event, we ran 25 independent trials.

Announcement of a destination. Figure 10a plots CCDFs of termination times after an announcement in AS 1239 over all possible destination nodes and all trials. Three protocols are

Authorized licensed use limited to: Tsinghua University. Downloaded on September 12,2025 at 07:14:59 UTC from IEEE Xplore. Restrictions apply.



(a) Network-wide announcement of a destination.

(b) Failure of a link.

Fig. 10. CCDFs of termination times in AS 1239 for the widest-shortest, shortest-widest, and width-length product protocols. Averages of the distributions are given inside parenthesis.

considered: the widest-shortest (standard vectoring) protocol; the shortest-widest (standard vectoring) protocol; and the width-length product (partial-order vectoring) protocol.

The curves for the widest-shortest protocol and for the width-length product protocol have similar behaviors, with average termination times of 184.7 ms and 212.3 ms, respectively. In the presence of isotonicity, attributes elected at each node to reach a destination can only be replaced by more preferred attributes during each trial. Therefore, the termination time equals the time taken to propagate an advertisement all the way up an optimal or dominant path, as the case may be, plus the time to clear this advertisement from the network.

Both the widest-shortest protocol and the width-length product protocol operate according to isotonic orders. The widest-shortest protocol computes, first and foremost, shortest paths, while the width-length product protocol computes sets of paths that include shortest and widest paths, as well as paths whose lengths are in-between those of shortest and widest paths. Since the time to convey an advertisement across a link equals its length plus a random delay, vectoring protocols take longer to compute widest paths than shortest paths, justifying the longer termination times of the width-length product protocol. At the same time, it should be emphasized that the width-length product protocol computes, on average, 2.45 width-length pairs from source to destination, whereas the widest-shortest protocol computes just one. The 2.45 width-length pairs are computed in parallel during execution of the width-length product protocol leading to a modest 15% increase in termination times in relation to the widest-shortest protocol.

The shortest-widest protocol is the slowest to terminate despite electing just one width-length pair from source to destination. Its average termination time is 235.9 ms with 9.5% of the announcements having termination times in excess of 300 ms. The corresponding values for the width-length product protocol are 212.3 ms and 0.6%.

Isotonicity not only guarantees optimality and dominance in stable state, but also promotes short termination times [27]. In the absence of isotonicity, a node may elect an attribute that later on has to be supplanted by a less preferred attribute. Standard vectoring protocols continuously search for the most preferred attribute among candidate attributes learned from their neighbors. Trying to settle on less preferred attributes by always electing the most preferred attributes among candidates learned from neighbors may take many iterations and, hence, lead to long termination times [28].

The shortest-widest protocol does not operate according to an isotonic order, which justifies its longer termination times in comparison with the widest-shortest protocol and the width-length product protocol.

Failure of a link. Figure 10b plots CCDFs of termination times after the failure of a link in AS 1239 over all possible links and all trials. The same vectoring protocols as in the case of an announcement are considered. The average termination times for the widest-shortest, shortest-widest, and width-length product protocols are 69.2 ms, 67.3 ms, and 99.3 ms, respectively. Many link failures have only a local impact on the stable state of the protocol, which explains the shorter average termination times in comparison with the network-wide announcement of a destination. For instance, 21.9% of link failures have no effect at all on the stable state of the shortest-widest protocol. This protocol computes widest paths. The failure of any link of width smaller than those of the widest paths goes by unnoticed by the protocol.

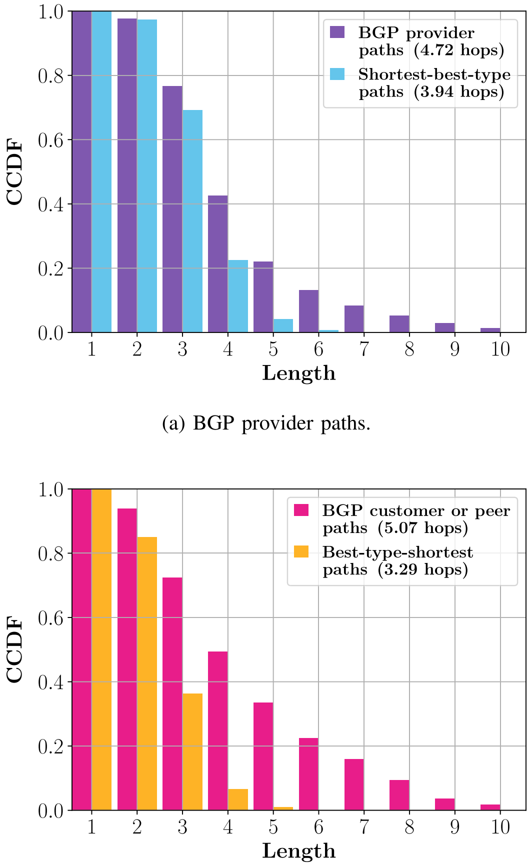
Despite shorter average termination times, the curves pertaining to link failures have a wider variance than the curves pertaining to an announcement. A few link failures lead to long termination times in excess of 300 ms. When a link fails in a network running a vectoring protocol, some nodes will end up electing less preferred attributes than the ones they elected before the failure. As stated previously, the transient process culminating in those elections can be slow.

VII. Evaluation of Inter-AS Routing

Our evaluation of inter-AS routing intends to answer two main questions. How large are the sets of dominant type-lengths in the Internet? How do the lengths of valid AS-paths provided by a partial-order vectoring protocol compare with those computed by BGP?

We used the Internet topologies inferred by CAIDA [29], where every pair of neighbor ASes is described as having

Authorized licensed use limited to: Tsinghua University. Downloaded on September 12,2025 at 07:14:59 UTC from IEEE Xplore. Restrictions apply.



(a) BGP provider paths.

(b) BGP customer or peer paths.

Fig. 11. CCDFs of AS-path lengths for BGP, type-optimal paths, and shortest-valid paths. Averages of the distributions are given inside parenthesis.

either a customer-provider or a peer-peer relationship. The very few ASes that do not have a valid AS-path to every other AS were removed. We present results for the Internet topology of February 2021, characterized by: 59,728 stub ASes and 10,746 ISPs, only 18 of which are tier-1 ISPs, that is, ISPs without providers, for a total of 70,474 ASes; 143,753 customer-provider relationships and 208,386 peer-peer relationships, for a total of 352,139 relationships.

We computed the sets of dominant type-lengths from every AS to every other AS in the Internet. The results are as follows: the percentages of source-destination pairs connected by one, two, and three dominant type-lengths are 93.85%, 6.14%, and 0.01%, respectively. The vast majority of source-destination pairs are connected just by one type-length. These include the 89.47% source-destination pairs that are connected exclusively by provider paths.

Figure 11a plots the CCDF of AS-path length over all source-destination pairs for which BGP has the source elect a provider type-length to reach the destination; and the corresponding CCDF for when the type-length product (partial-order vectoring) protocol is used instead. Recall that the operation of BGP in inter-AS routing obeys a shortest-best-type order, that this order does not satisfy left-isotonicity, and, consequently, that BGP does not route on minimum length provider paths, in general (Section III). Contrastingly, the type-length product protocol satisfies left-isotonicity and can route on minimum length provider paths. Figure 11a shows that the type-length product protocol curtails the distribution of AS-path lengths and reduces its average. The percentages of source-destination pairs connected by more than 5 hops is 22.0% and 4.2% for BGP and for the type-length product protocol, respectively, while the average AS-path lengths are 4.72 and 3.94.

Figure 11b plots the CCDF of AS-path length over all source-destination pairs for which BGP has the source elect a peer or a customer type-length to reach the destination; and the corresponding CCDF of best-type-shortest paths. Routing on best-type-shortest paths is not possible with a standard vectoring protocol, but it is possible with a type-length product protocol (Section III). Figure 11b shows that best-type-shortest paths are significantly shorter than shortest-best-type paths. The percentages of source-destination pairs connected by more than 5 hops is 33.6% and 1.1% for the shortest-best-type and the best-type-shortest orders, respectively, while the average AS-path lengths are 5.07 and 3.29.

The main takeaway from these results is that with only a few cases of ASes electing two or, in rare instances, three type-lengths to reach a destination, AS-path lengths can be significantly shortened in comparison with current inter-AS routing.

VIII. Related Work

Algebraic conceptualization of routing. The algebraic framework proposed in [5], [7] laid the foundations for a unified treatment of routing problems and protocols, abstracting away the specificity of performance metrics, policy choices, and protocol parameters. The framework is premised on a total order on attributes. The present work generalizes the algebraic framework by accepting partial orders on attributes and by devising vectoring protocols that route on the dominant paths determined by those orders. Moreover, it describes a generic procedure that reduces a set of total orders to a common partial order that respects all orders and for which the extension is isotone.

Multi-objective path problems. Multi-objective path problems have been studied by the operations research community [30]–[32]. These problems can be described in concrete algebraic terms. Attributes are tuples of the Cartesian product of elementary metrics, each of which either extends with addition and is ordered by the less-than-or-equal order or extends with the minimum operator and is ordered by the greater-than-or-equal order. Tuples extend term-wise and are partially ordered by the product order of their term-wise total orders. The goal is to find sets of dominant tuples from source to destination in a network and is attained with generalizations of Dijkstra's and Bellman-Ford algorithms [30]–[32].

The setting considered in this work is broader and the problem addressed is different. Attributes are not necessarily tuples of elementary metrics. Even when they are, a partial order on them is derived, rather than assumed a priori, and does not necessarily coincide with the product order. The goal

Authorized licensed use limited to: Tsinghua University. Downloaded on September 12,2025 at 07:14:59 UTC from IEEE Xplore. Restrictions apply.

is to route data-packets on optimal paths, for a single optimality criterion and for multiple optimality criteria. This goal is attained with a partial-order vectoring protocol.

Multipath routing protocols. Since partial-order vectoring protocols typically find multiple paths from source to destination, they can be considered a type of multipath routing protocols. Multipath routing protocols have mostly been proposed as extensions to BGP with one of the following three goals in mind. A first goal is to ensure proper termination both of external BGP [33] and of internal BGP [34], [35]. A second goal is to improve the data-packet delivery capabilities of BGP during convergence of the protocol upon a link failure [36]–[39]. And a third goal is to allow the configuration of more expressive routing policies than is possible with standard BGP [40]. The BGP multipath routing proposal presented in [23] addresses all three goals.

The partial-order vectoring protocol proposed in this work targets a different goal. We seek to route data-packets on optimal paths for a variety of optimality criteria, some of which do not lend themselves to a solution by a standard vectoring protocol. In addition, our partial-order vectoring protocol is formulated with generality rather than being specific to BGP.

IX. Summary and Conclusion

We presented a solution to the problem of routing on optimal paths concurrently for a collection of optimality criteria, which subsumes, as a particular case, the problem of routing on optimal paths for a single, but arbitrary, optimality criterion. A fundamental piece of the solution is the identification of a partial order on attributes that is contained in each of the total orders that define a criterion of the collection and for which the extension operation on attributes is isotone. We designed a partial-order vectoring protocol that operates according to such partial orders to compute dominant attributes from sources to destinations in any given network. By construction, the dominant attributes contain the optimal attributes associated with each of the optimality criteria. Alongside the computation of dominant attributes, partial-order vectoring protocols disseminate the necessary forwarding information to guide data-packets on the diverse types of optimal paths.

The generality of our approach was accompanied with two instantiations. The first concerns optimal path routing based on performance metrics that can be represented as widths and lengths. We concluded that routing on K-quickest paths or on shortest-widest paths is not possible with a standard vectoring protocol, but can be accomplished with a partial-order vectoring protocol that elects just a few width-lengths per destination in stable state and terminates fast. The second instantiation of our approach relates to inter-AS routing. AS-paths are valid if, and only if, all the intermediate ASes profit from transiting traffic. Aside from validity, AS-path length is a major consideration in path selection. We concluded that a partial-order vectoring protocol that elects two or, in rare cases, three type-lengths per destination at a few ASes significantly reduces the lengths of routing paths across the Internet as compared to BGP.

Appendix A
Largest Isotonic Reduction of the K-Quickest Order

Proposition 1: The largest isotonic reduction of the K-quickest order $\preceq_K$ is such that width-length (w, l) equals or is preferred to width-length (w', l') if, and only if, $(w, l) \preceq_K (w', l')$ and $l \leq l'$.

Proof: The extension operation on width-lengths is associative. Therefore, we apply Theorem 2. First, we show that $(w, l) \preceq_K (w', l')$ and $l \leq l'$ together imply $(\min(x, w), m+l) \preceq_K (\min(x, w'), m+l')$ for every width-length (x, m). Two cases are distinguished.

Case 1: $w \geq w'$. We have $\min(x, w) \geq \min(x, w')$. From $l \leq l'$, we write

$$K/\min(x, w) + m + l \leq K/\min(x, w') + m + l'.$$

Consequently, $(\min(x, w), m + l) \preceq_K (\min(x, w'), m + l')$.

Case 2: $w < w'$. From $(w, l) \preceq_K (w', l')$ and $w < w'$, we deduce that $l < l'$. If $x \leq w$, then $x = \min(x, w) = \min(x, w')$, and we write

$$K/\min(x, w) + m + l < K/\min(x, w') + m + l',$$

so that $(\min(x, w), m + l) \preceq_K (\min(x, w'), m + l')$. If $w < x < w'$, then we write

$$\begin{aligned} &K/\min(x, w) + m + l \\ &\quad = K/w + m + l && \text{(from } w < x\text{)} \\ &\quad \leq K/w' + m + l' && \text{(from } (w, l) \preceq_K (w', l')\text{)} \\ &\quad < K/\min(x, w') + m + l'. && \text{(from } x < w'\text{)} \end{aligned}$$

Once again, $(\min(x, w), m+l) \preceq_K (\min(x, w'), m+l')$. Last, if $w' \leq x$, then widths w and w' are not diminished by width x. We obtain $(\min(x, w), m + l) \preceq_K (\min(x, w'), m + l')$ directly from $(w, l) \preceq_K (w', l')$.

Second, we show that if either $(w, l) \preceq_K (w', l')$ does not hold or $l > l'$, then there is width-length (x, m) such that $(\min(x, w), m + l) \preceq_K (\min(x, w'), m + l')$ does not hold. If $(w, l) \preceq_K (w', l')$ does not hold, then we choose $(x, m) = (+\infty, 0)$. Otherwise, if $l > l'$, then we choose $(x, m) = (\min(w, w'), 1)$ to obtain $K/\min(w, w') + l + 1 > K/\min(w, w') + l' + 1$, which implies that $(\min(x, w), m + l) \preceq_K (\min(x, w'), m + l')$ does not hold. ■

Appendix B
Intersection of k-Quickest Orders for $0 \leq k \leq K$

Proposition 2: The intersection of the k-quickest orders $\preceq_k$ for all k between 0 and K is such that (w, l) equals or is preferred to width-length (w', l') if, and only if, $(w, l) \preceq_K (w', l')$ and $l \leq l'$.

Proof: First, we show that $(w, l) \preceq_K (w', l')$ and $l \leq l'$ together imply $(w, l) \preceq_k (w', l')$ for every k, $0 \leq k < K$. Two cases are distinguished.

Case 1: $w \geq w'$. We have $k/w \leq k/w'$. From $l \leq l'$, we write

$$l + k/w \leq l' + k/w'.$$

Consequently, $(w, l) \preceq_k (w', l')$.

Authorized licensed use limited to: Tsinghua University. Downloaded on September 12,2025 at 07:14:59 UTC from IEEE Xplore. Restrictions apply.

Case 2: $w < w'$. We write

$$\begin{aligned} & k/w + l \\ & \quad \le k/w + K(1/w' - 1/w) + l' \\ & \qquad\qquad\qquad\qquad \text{(from } (w,l) \preceq_K (w',l')\text{)} \\ & \quad < k/w + k(1/w' - 1/w) + l' \quad \text{(from } k < K\text{)} \\ & \quad = k/w' + l'. \end{aligned}$$

Once again, $(w,l) \preceq_k (w',l')$.

Second, we show that $(w,l) \preceq_k (w',l')$ for every k, $0 \le k \le K$, implies $(w,l) \preceq_K (w',l')$ and $l \le l'$. Trivially, we take $k = K$ and $k = 0$ in $(w,l) \preceq_k (w',l')$, respectively. ■

References

[1] Y. L. Chen and Y. H. Chin, "The quickest path problem," *Comput. Operations Res.*, vol. 17, no. 2, pp. 153–161, Jan. 1990.

[2] Q. Dong, S. Banerjee, M. Adler, and A. Misra, "Minimum energy reliable paths using unreliable wireless links," in *Proc. 6th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, May 2005, pp. 449–459.

[3] L. Gao and J. Rexford, "Stable internet routing without global coordination," *IEEE/ACM Trans. Netw.*, vol. 9, no. 6, pp. 681–692, Dec. 2001.

[4] B. Carré, *Graphs and Networks*. Oxford, U.K.: Clarendon Press, 1979.

[5] J. L. Sobrinho, "Algebra and algorithms for QoS path computation and hop-by-hop routing in the internet," *IEEE/ACM Trans. Netw.*, vol. 10, no. 4, pp. 541–550, Aug. 2002.

[6] J. L. Sobrinho, "An algebraic theory of dynamic network routing," *IEEE/ACM Trans. Netw.*, vol. 13, no. 5, pp. 1160–1173, Oct. 2005.

[7] T. G. Griffin and J. L. Sobrinho, "Metarouting," in *Proc. ACM SIGCOMM*, 2005, pp. 1–12.

[8] M. Gondran and M. Minoux, *Graphes, Dioides, Semirings*. Cham, Switzerland: Springer, 2008.

[9] J. Baras and G. Theodorakopoulos, *Path Problems in Networks*. San Rafael, CA, USA: Morgan & Claypool Publishers, 2010.

[10] D. Savage, J. Ng, S. Moore, D. Slice, P. Paluch, and R. White, *Cisco's Enhanced Interior Gateway Routing Protocol (EIGRP)*, document RFC 7868, May 2016.

[11] M. G. Gouda and M. Schneider, "Maximizable routing metrics," *IEEE/ACM Trans. Netw.*, vol. 11, no. 4, pp. 663–675, Aug. 2003.

[12] Y. Rekhter, T. Li, and S. Hares, *A Border Gateway Protocol (BGP)*, document RFC 4271, Jan. 2006.

[13] T. Lengauer and D. Theune, "Efficient algorithms for path problems with general cost criteria," in *Proc. Int. Colloq. Automata, Lang. Program.*, 1991, pp. 314–326.

[14] E. Harzheim, *Ordered Sets*. Cham, Switzerland: Springer, 2005.

[15] G. P. Chandranmenon and G. Varghese, "Trading packet headers for packet processing," *IEEE/ACM Trans. Netw.*, vol. 4, no. 2, pp. 141–152, Apr. 1996.

[16] N. Spring, R. Mahajan, D. Wetherall, and T. Anderson, "Measuring ISP topologies with Rocketfuel," *IEEE/ACM Trans. Netw.*, vol. 12, no. 1, pp. 2–16, Feb. 2004.

[17] (Jul. 2008). *AS Relationships Dataset, as-Rel.20080707.a0.01000.Txt*. CAIDA. [Online]. Available: http://www.caida.org/data/active/as-relationships/

[18] J. L. Sobrinho and M. A. Ferreira, "Routing on multiple optimality criteria," in *Proc. ACM SIGCOMM*, 2020, pp. 211–225.

[19] Z. Wang and J. Crowcroft, "Quality-of-service routing for supporting multimedia applications," *IEEE J. Sel. Areas Commun.*, vol. 14, no. 7, pp. 1228–1234, Sep. 1996.

[20] G. Huston, "Interconnection, peering and settlements—Part I," *Internet Protocol J.*, vol. 2, no. 1, pp. 2–16, Mar. 1999.

[21] G. Huston, "Interconnection, peering and settlements—Part II," *Internet Protocol J.*, vol. 2, no. 2, pp. 2–23, Jun. 1999.

[22] L. Gao, "On inferring autonomous system relationships in the internet," *IEEE/ACM Trans. Netw.*, vol. 9, no. 6, pp. 733–745, Dec. 2001.

[23] Y. Wang, M. Schapira, and J. Rexford, "Neighbor-specific BGP: More flexible routing policies while improving global stability," in *Proc. ACM SIGMETRICS*, 2009, pp. 217–228.

[24] J. L. Sobrinho, "Fundamental differences among vectoring routing protocols on non-isotonic metrics," *IEEE Netw. Lett.*, vol. 1, no. 3, pp. 95–98, Sep. 2019.

[25] G. Malkin, *RIP: An Intra-Domain Routing Protocol*. Reading, MA, USA: Addison Wesley, 1999.

[26] D. P. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1991.

[27] M. L. Daggitt and T. G. Griffin, "Rate of convergence of increasing path-vector routing protocols," in *Proc. IEEE ICNP*, 2018, pp. 335–345.

[28] C. Labovitz, A. Ahuja, A. Bose, and F. Jahanian, "Delayed internet routing convergence," *IEEE/ACM Trans. Netw.*, vol. 9, no. 3, pp. 293–306, Jun. 2001.

[29] (Feb. 2021). *The CAIDA AS Relationships Dataset*. [Online]. Available: http://www.caida.org/data/active/as-relationships/

[30] P. Hansen, "Bicriterion path problems," in *Multiple Criteria Decision Making Theory and Application*, G. Fandel and T. Gal, Eds. Cham, Switzerland: Springer Verlag, 1980, pp. 109–127.

[31] E. Q. V. Martins, "On a multicriteria shortest path problem," *Eur. J. Oper. Res.*, vol. 16, no. 2, pp. 236–245, 1984.

[32] J. Brumbaugh-Smith and D. Shier, "An empirical investigation of some bicriterion shortest path algorithms," *Eur. J. Oper. Res.*, vol. 43, no. 2, pp. 216–224, 1989.

[33] R. Agarwal, V. Jalaparti, M. Caesar, and P. B. Godfrey, "Guaranteeing BGP stability with a few extra paths," in *Proc. IEEE 30th Int. Conf. Distrib. Comput. Syst.*, 2010, pp. 221–230.

[34] A. Flavel and M. Roughan, "Stable and flexible iBGP," in *Proc. ACM SIGCOMM*, 2009, pp. 183–194.

[35] V. Van den Schrieck, P. Francois, and O. Bonaventure, "BGP add-paths: The scaling performance tradeoffs," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 8, pp. 1299–1307, Sep. 2010.

[36] N. Kushman, S. Kandula, D. Katabi, and B. M. Maggs, "R-BGP: Staying connected in a connected world," in *Proc. USENIX NSDI*, 2007, pp. 1–14.

[37] I. Ganichev, B. Dai, P. B. Godfrey, and S. Shenker, "YAMR: Yet another multipath routing protocol," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 5, pp. 13–19, Oct. 2010.

[38] F. Wang and L. Gao, "A backup route aware routing protocol—Fast recovery from transient routing failures," in *Proc. IEEE INFOCOM*, 2008, pp. 2333–2341.

[39] Y. Liao, L. Gao, R. Guerin, and Z.-L. Zhang, "Reliable interdomain routing through multiple complementary routing processes," in *Proc. ACM CoNEXT*, 2008, pp. 1–6.

[40] W. Xu and J. Rexford, "MIRO: Multi-path interdomain routing," in *Proc. ACM SIGCOMM*, 2006, pp. 171–182.

João Luís Sobrinho (Senior Member, IEEE) received the Licenciatura and Ph.D. degrees and the title of Agregado in electrical and computer engineering from Instituto Superior Técnico in 1990, 1995, and 2019, respectively. He is an Associate Professor with the Department of Electrical and Computer Engineering, Instituto Superior Técnico, Universidade de Lisboa; and a Senior Researcher with the Instituto de Telecomunicações. Before joining academia in 1997, he was a member of the Technical Staff with Bell Labs. His research interests are in the rigorous analysis and design of networks protocols. He won a 2020 ACM SIGCOMM Best Paper Award, a 2015 Internet Society Applied Networking Research Prize, and the 2006 IEEE Communications Society William R. Bennett Prize. He is a member of ACM.

Miguel Alves Ferreira (Student Member, IEEE) received the M.Sc. degree in electrical and computer engineering from Instituto Superior Técnico in 2020. He is currently pursuing the dual Ph.D. degree in electrical and computer engineering with Carnegie Mellon University and the Instituto Superior Técnico. He is interested in solving routing and congestion control problems through the lens of stochastic control and optimization. He won a 2020 ACM SIGCOMM Best Paper Award.

Authorized licensed use limited to: Tsinghua University. Downloaded on September 12,2025 at 07:14:59 UTC from IEEE Xplore. Restrictions apply.