Bootstrapping Evolvability for Inter-Domain Routing with D-BGP

Raja R. Sambasivan[*], **David Tran-Lam**[†], **Aditya Akella**[†], **Peter Steenkiste**[*]

June 2016
CMU-CS-16-117

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

[*]School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA
[†]Department of Computer Sciences, University of Wisconsin-Madison, Madison, WI, USA

This research was funded in part by NSF under award numbers CNS-1345305 and CNS-1345249.

Keywords: BGP, control plane, evolvability

Abstract

It is extremely difficult to utilize new routing protocols in today's Internet. As a result, the Internet's baseline inter-domain protocol for connectivity (BGP) has remained largely unchanged, despite known significant flaws. The difficulty of using new protocols has also depressed opportunities for (currently commoditized) transit providers to provide value-added routing services. To help, this paper proposes Darwin's BGP (D-BGP), a modified version of BGP that can support evolvability to new protocols. D-BGP modifies BGP's advertisements and advertisement processing based on requirements imposed by key evolvability scenarios, which we identified via analyses of recently-proposed routing protocols.

1 Introduction

The Internet's inter-domain routing protocol, BGP, is critical to the Internet's architecture. All of the services and content we hold dear are accessible because of the routing paths that it computes. But, this important protocol is plagued with problems. For example, it does not provide domains (stubs or transit providers) sufficient influence to limit incoming traffic [12]; its paths are slow to converge and prone to oscillations [32]; it indiscriminately chooses a single best-effort path per router, robbing other domains of paths they may prefer more [50]; and it is prone to numerous attacks, including prefix hijacking, traffic interception, and black-holing [38]. Worst of all, BGP is architecturally rigid and cannot facilitate the introduction of new protocols [14].

This architectural rigidity has made rolling out the plethora of critical fixes and improvements to BGP proposed by the research and the operator community painfully difficult. Examples include adding secure path announcements via S-BGP [25] to prevent prefix hijacking, adding awareness of path costs to limit incoming traffic to domains [31], and adding backup paths to reduce convergence times [27].

BGP's rigidity has also prevented more sophisticated protocols that address BGP's increasing unsuitability for today's Internet from being deployed (e.g., path-based or multi-path routing [48, 50] to offer source domains more control over path selection and multi-hop routing to allow for rich policies [13, 15]). As such, BGP, flaws and all, has remained virtually unchanged for almost 20 years. Without efforts to understand how to deploy new inter-domain routing protocols, BGP will likely remain unchanged for years to come.

To help, this paper presents Darwin's BGP (D-BGP), a modified version of BGP that transforms it from a rigid protocol to one that can *bootstrap evolvability*—i.e., help new protocols gain traction and seamlessly deprecate itself in favor of them. As such, D-BGP allows operators to rapidly deploy fixes to BGP—either across all or a subset of domains—whenever new use cases bring critical deficiencies to the fore. In the extreme, D-BGP can help the Internet transition from an old routing protocol to one that uses a fundamentally different paradigm (e.g., move from hop-based to path-based forwarding). It can also facilitate the simultaneous co-existence of multiple disparate protocols, improving the richness of the Internet architecture as a whole.

D-BGP's modifications are based on requirements for enabling evolvability for three key scenarios, which we identified via an analysis of recently-proposed routing protocols. We find that three modifications are necessary. First is the ability to pass-through information about protocol a domain does not support to other domains. Second is multi-protocol support within BGP's advertisements to inform routers of what protocols are used on routing paths and how to use them to send data packets. Third is multi-protocol support within BGP's advertisement processing, allowing protocol-specific information to be forwarded to the relevant protocol's path-selection algorithm.

Our analysis of D-BGP's control-plane costs reveal that it can facilitate a rich Internet composed of 10s to 100s of BGP fixes and sophisticated replacements with only modest overheads compared to a single protocol deployment. This result is largely due to the fact that many of BGP's fixes can share most of their protocol-specific information with BGP. Experiments run using D-BGP show that it incentivizes adoption of new protocols by increasing the rate at which early adopters see the benefits of a new protocol. D-BGP itself can be incrementally deployed across contiguous domains, allowing participants to enjoy the rich and evolvable Internet it can enable.

This paper builds on previous work on evolvability for inter-domain routing [39] and presents the following contributions:

1) Based on an analysis of 13 recently proposed inter-domain routing protocols, we identify key evolvability requirements any mechanism that aims to facilitate evolution to new inter-domain protocols must satisfy.

2) We identify modifications needed to BGP to satisfy these requirements and describe the design of D-BGP, a version of BGP that incorporates them.

3) We show that D-BGP's control plane overheads are reasonable even when supporting large numbers of inter-domain routing protocols. We show that D-BGP incentivizes adoption for important types of new protocols by accelerating the rate at which adopters see the benefits of using them.

The rest of this paper is organized as follows. Section 2 describes the evolvability scenarios we identified and the requirements they impose for routing evolvability. Section 3 identifies the modifications needed to BGP to satisfy these requirements and illustrates how they are incorporated within D-BGP. Section 4 presents a case study of the evolvable Internet D-BGP could enable. Section 5 evaluates D-BGP's overheads and its ability to increase the rate at which adopters see the benefits of a new protocol. Section 6 presents related work and Section 7 concludes.

2 Evolvability scenarios

To identity what requirements any mechanisms for routing evolvability must provide, we analyzed recently proposed protocols from the research and operator communities [4, 11, 13, 15, 25, 27, 31, 34, 41, 46–48, 50]. We identified three key evolvability scenarios: modifying some baseline protocol with a critical fix, running a custom protocol side-by-side with the baseline, and replacing the baseline with a fundamentally different protocol. Protocols suited to these scenarios differ in three ways. First, they differ in their goals. Second, they differ in the type of data-plane support they require in an Internet that is running multiple inter-domain protocols (see Section2.1). Third, they differ in operators' incentives for deploying them. Some protocols are suited to multiple scenarios. This section describes these scenarios and the requirements they impose. We start with a discussion of the data-plane issues that can arise when deploying multiple protocols.

Assumptions: At the beginning of time, we assume that all domains, ASes for short, are using a baseline routing protocol for inter-connectivity that is BGP-like. It is a hop-based protocol that uses path-vector-style loop detection. Its advertisements carry connectivity information from traffic sinks (destinations) to traffic sources. Most importantly, each advertisement identifies a single best-effort path for routing data packets to a sink. Routers only advertise a single path to each neighbor.

Data packets flow downstream from sources to sinks and are guaranteed to traverse only the first hop of advertised paths (partially because advertised paths may change while data packets are in flight). Our discussion below is agnostic to whether ASes use distributed control (i.e., routers choose paths) or centralized control (e.g., SDNs [18, 22]) and whether ASes support different sets of protocols on different routers.

Terminology: *Islands* refer to a cluster of one or more contiguous ASes that support the same set of routing protocols. Neighbors of islands run a different set of protocols. *Baseline ASes / Islands* refer to those that run the baseline protocol (e.g., BGP). *Upgraded ASes / Islands* refer to those that support the new protocol being discussed. We refer to the set of ASes separating two upgraded islands as *gulfs*.

2.1 Routing consistency

The data plane or *network protocol* is responsible for enforcing routing protocols' path choices. When multiple routing protocols are deployed concurrently, consistency of routing decisions becomes an issue. If care is not taken to consistently enforce the same routing protocol's path choices for a destination address at every location (e.g., router or AS), the resulting end-to-end path may not be the result of any single protocol's choices. Such hybrid paths may violate the goals of a given protocol (e.g., one that aims to avoid a congested AS). Also, paths chosen by one protocol at one location may prevent data packets from using better paths selected by a more preferred protocol at other locations. These issues can severely curtail new protocols' benefits. Whether or not a protocol needs its routing decisions to be consistently enforced informs the evolvability scenario to which it is suited.

Enforcing consistency requires different mechanisms within islands and across islands. Our discussions assume an IP prefix as the destination address, but are equally valid for other types (e.g., content names [49]). To ensure consistency within islands, protocols must be careful not to install conflicting entries at different points in the path. This requires assigning different protocols different addresses that name the same physical destinations.

Additionally ensuring consistency for routing decisions across islands requires the relevant protocol's path choices to be enforced at locations that do not support it (i.e., within gulfs). Doing so requires data packets to be encapsulated and tunneled, thus hiding their within-island addresses from other protocols and islands.

Note that if routing protocols use different network protocols or use a network protocol that supports multiple address types (e.g., XIA [17]), consistency issues can be avoided.

2.2 Baseline → Baseline with critical fix

In this evolvability scenario, the goal is to deploy a modified version of the baseline that incorporates some critical fix. These critical fixes usually extend the baseline by disseminating extra control information to improve path selection or the protocol itself. Examples of critical fixes to today's baseline, BGP, include Wiser [31], for fixing BGP's lack of support for limiting ingress traffic at ASes [12], S-BGP [25], for fixing BGP's susceptibility to route hijacking [38], R-BGP [27], for providing backup paths in case of failures, and LISP [11], for supporting mobility.

Data-plane support: Routing consistency is not needed. Protocols suited to this evolvability scenario can use each other's path-selection choices. This is because these protocols are simply different versions of each other that are backward compatible (i.e., they use the same routing paradigm). ASes that use critical fixes that require participation by all hops on a routing path can verify end-to-end participation transitively. For example, to guarantee that ASes on an advertised path are not being spoofed, S-BGP [25] disseminates extra route attestations with advertisements, which describe the

path and the authenticity of the ASes that constitute it. The attestation is signed by each AS that advertises a S-BGP path. Neighbors that receive S-BGP advertisements verify that the latest signer is the AS that it received the advertisement from.

Example: Figure 1 shows a scenario in which the white ASes start to deploy Wiser [31] as an update to BGP. Wiser fixes BGP's inability to help ASes limit ingress traffic by disseminating an extra *path cost* in advertisements, which influences path selection. ASes independently add their internal costs of routing traffic along advertised paths before selecting the one with the lowest cost. Costs are normalized between neighbors to allow ASes to choose their own cost metric (e.g., price, congestion) and to prevent sensitive information from being leaked.

In the Figure, the two ASes at the edge of the large Wiser island, E1 and E2, must use BGP to advertise paths to their neighbors in the BGP gulf. Lines show paths advertised and arrows show the direction of the advertisement. This creates two problems . First, the source, which supports Wiser, must use BGP to select paths because it cannot see path costs. As such, it will choose the shortest path (due to BGP's decision criteria), which has the highest path cost.

Second, E1 and E2 are at a disadvantage because their path-cost contribution will not be taken into account when the source or any ASes in the gulf select best paths. Yet, they must honor the path costs they receive from downstream neighbors when selecting best paths themselves. This may dis-incentivize them from supporting Wiser, especially if this requirement increases their payments to providers or peers.

Requirements: As the above example shows, today, non-contiguous islands or ASes that deploy updated baseline protocols cannot quickly leverage the improvements afforded by them. This is because updated baselines' extra control information cannot be disseminated across BGP gulfs. Thus, we end up with this requirement:

CF-R1 *Disseminate new protocols' control information across gulfs.*

Also, the critical fix must succeed the baseline eventually. Thus, we end up with:

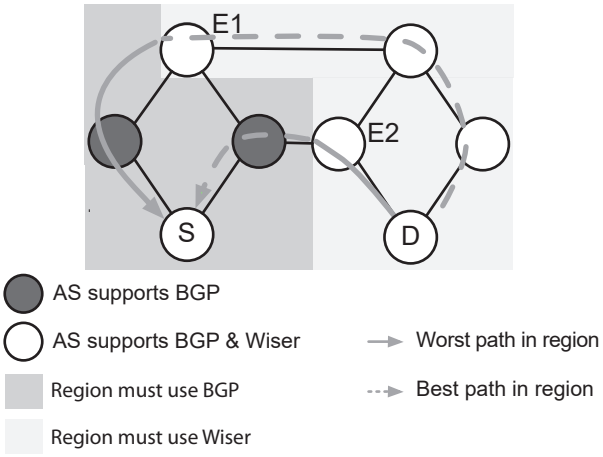

Figure 1: S cannot see path costs, so it will choose the highest-cost one.

CF-R2 *Disseminate new protocols' control information side-by-side or in-band with BGP's advertisements .*

Incentives for deployment are incremental benefits: Operators will be incentivized to deploy a critical fix if the benefits afforded by that protocol can be realized quickly. Such benefits will increase incrementally as a function of the protocol and the number of ASes that can route among each other using it.

Constraints: This evolvability scenario applies only to a very restricted set of protocol improvements that can directly use each other's path-selection choices. This scenario is also restricted to path-vector-style protocols.

2.3 Baseline → Baseline // custom protocol

The goal of this evolvability scenario is to allow ASes or islands to deploy a new protocol in parallel (//) with the baseline. The new protocol is used to route select traffic, while BGP is used for the rest. It is most apt for protocols that provide value-added services, which operators sell to other customers (e.g., other AS operators or end users). Examples include providing alternate paths from BGP's single path [27, 34, 47] and providing extra functionality on existing paths [34] (e.g., higher intra-domain or intra-island QoS).

This scenario also describes situations in which different islands aim to use some non-baseline protocol to deliver select traffic to each other. These non-baseline protocols may include ones that use different routing paradigms than the baseline. Examples include multi-hop routing protocols [13, 15], such as Pathlets [15], which allow ASes to construct end-to-end paths out of advertised path fragments, and path-based routing protocols [48], such as SCION [50], which advertise multiple paths per router and let sources pick which ones they want to use.

Data-plane issues: Islands will run multiple inter-domain routing protocols concurrently (e.g., the baseline and new protocol). The new protocol's routing decisions must be enforced consistently, both within islands, and across gulfs. Assuming all routing protocols use the same network protocol and address types, separate address ranges must be assigned to custom protocols within islands. Packets must be encapsulated and tunneled across gulfs. Otherwise, the baseline protocol may divert packets from ever reaching an upgraded island.

Example: Figure 2 describes a scenario in which a transit AS (marked T) wishes to avoid the single poorly performing path advertised by BGP (the dashed path). An AS that supports MIRO [47] offers alternate paths for payment (the rightmost one). However, transit AS T cannot discover the MIRO-enabled AS because BGP does not allow discovery of ASes' custom services or the extra coordination required to use them. This lack of discovery mechanism limits the MIRO AS's potential customers, perhaps only to its direct neighbors. It could use bespoke approaches for discovery (e.g., a web site), but these may go unnoticed.

Requirements: As the above example shows, ASes or islands supporting the new protocol must be able to both discover each other and how to coordinate out-of-band in order to exchange relevant control information, including protocol-specific information (e.g., alternate paths) and the type of encapsulation method that will be used to route packets across gulfs. Thus, we require:

CP-R3 *Facilitate off-path discovery of custom protocols.*

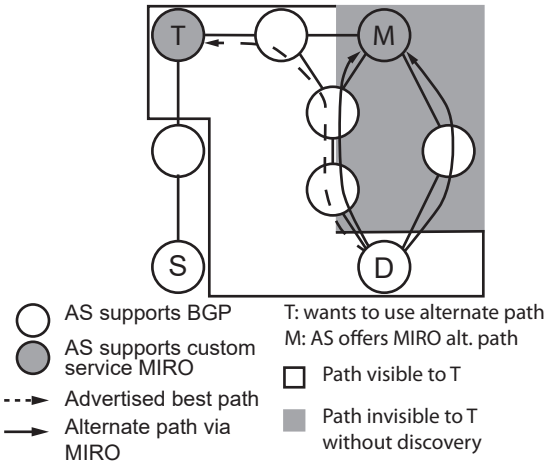

Figure 2: T cannot discover M's alternate path.

Incentives for deployment: ASes or islands that deploy custom protocols do so in hopes of selling value-added services or because of an explicit contractual relationship with direct or indirect neighbors. Other ASes are incentivized to facilitate discovery of ASes that use custom protocols because their customers—who might desire the value-added services offered or may have a contractual relationship with some indirect neighbor—might seek alternate providers otherwise.

Constraints: This scenario is not apt for protocols that aim to replace the baseline because the coordination required to use custom protocols leverages the baseline's paths. It will be difficult to deploy a large number of custom-routing protocols that use the same network protocol because each will have to be assigned increasingly smaller pools of addresses. This limitation is avoided if custom protocols use different network protocols (or different address types within an existing protocol).

2.4 Baseline → Replacement protocol

The goal of this evolvability scenario is to allow new protocols to completely replace the baseline within islands. So, the key difference between this and the previous scenario is that the protocol is used for *all traffic* in these upgraded islands. Doing so is a very aggressive model and likely to be only attractive if there are strong incentives or requirements that are impossible to meet with the baseline (e.g., high QoE for all traffic or specific economic relationships). As such, it is likely to be useful only within single islands. However, this scenario can also be used to gradually introduce protocols that are radically different than the baseline and aim to succeed it. In this case, multiple islands would use the same protocol and route traffic among each other using it.

Protocols apt for this scenario include ones that use very different routing paradigms than the baseline, such as HLP [41], which is a hybrid path-vector-based/link-state protocol, or path-based ones [15, 48, 50].

Data-plane issues: Within islands, consistency is not an issue because only a single routing protocol is used. Across islands, paths to destinations will be jointly controlled by the replacement and the baseline. Across gulfs, whether routing consistency is needed depends on individual proto-

cols' goals. Routing consistency across islands is not possible in this scenario because the custom coordination needed to identify how to tunnel traffic is not feasible (see below). Instead, paths will be jointly established by upgraded islands and ASes in gulfs. Many protocols that use different routing paradigms than the baseline require data packets to incorporate protocol-specific headers. As such, packets must be encapsulated using both the baseline's header and that needed by the new protocol.

Example: Figure 3 illustrates a scenario in which BGP is being replaced by SCION [50], a path-based protocol. Sources can be advertised multiple path options (exposed at the granularity of border routers). They pick which one they want to use by encoding the path in packet headers, which routers key on to forward traffic. In this case, the rightmost SCION region in the diagram exposes two paths to the destination.

The scenario illustrates two key problems: the SCION source in the diagram cannot discover other SCION islands or route traffic to them. Unlike the previous model, out-of-band coordination and tunneling cannot be used to address this problem as it will not scale to handle all traffic (i.e., from the entire Internet). Also, ASes within BGP gulfs cannot route to destinations within SCION islands. Both problems can be addressed by re-distributing SCION routes into BGP [28]. But, BGP can only advertise one path per router, so one of the SCION paths would be lost in this example.

Requirements: Solving the above problems requires the ability to disseminate new protocols' control information in-band with the baseline protocol. Doing so sidesteps scalability issues and avoids redistribution issues that may result in loss of important information. Also, such control information must be capable of crossing gulfs. This is the same requirements as that for critical fixes (**CF-R1** and **CF-R2**).

Incentives for deployment are incremental benefits: Similar to incentives for critical fixes, AS operators will be inclined to deploy a replacement protocol if its benefits can be realized quickly. Also, stringent contractual obligations among neighbors might require use of a replacement protocol (e.g., one that provides high QoS).

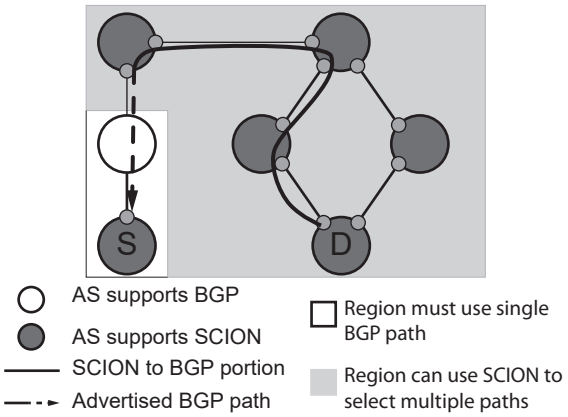

Figure 3: S cannot be advertised both paths to D.

3 Design of D-BGP

This section describes the design of Darwin's BGP (D-BGP), a version of BGP that can bootstrap evolvability to new protocols. D-BGP incorporates two key building blocks that satisfy the evolvability requirements identified in the previous section. *Pass-through support* enables routers or ASes to pass-through control information for protocols they do not support to adjacent ones. This allows new protocols' control information to cross gulfs ((CF-R1). It also facilitates off-path discovery (*CP-R3*) as we shall see later in this section. The second building block is *multi-protocol support* in advertisements and routers, which provides the expressiveness necessary to encode multiple protocols' information side-by-side ((CF-R2) and the coordination necessary to disseminate and use paths in an Internet composed of many inter-domain routing protocols. D-BGP does require some data-plane support, similar that used for MPLS [2] and Arrow [34].

The rest of this section describes how D-BGP modifies BGP's advertisements and advertisement processing to provide the needed building blocks. It also describes how D-BGP addresses the scenarios presented in the previous section. D-BGP is agnostic as to whether ASes use centralized control (e.g., SDNs) or distributed control (i.e., traditional routers). We assume the latter in this section.

3.1 Assumptions

We assume that all inter-domain routing protocols will be assigned unique protocol IDs by some governing body (e.g., the IETF [21]). We assume critical fixes and replacement protocols that aim to succeed BGP (the current baseline) will also vetted by a governing body and assigned some global ordering (e.g, critical fix version 5 > critical fix version 4). Vetting prevents broken protocols or those that interact poorly with current ones from being deployed. Examples include those that cause transient oscillations [43] or those that maximize some metric that an existing protocol aims to minimize. Ordering allows for eventual convergence toward some new baseline. Custom protocols and replacement protocols that do not aim to succeed BGP are not subject to such scrutiny as their modifications are local to individual islands.

We assume that all inter-domain routing protocols will be path-vector based, as scalability issues will likely prevent other protocol types (e.g., link state) from being deployed. Note that individual islands may use non-path-vector protocols internally. For example, they could use a link-state protocol for intra-island communication and path-vector for inter-island communication [41].

Finally, we assume that all routers support both BGPv4 as the baseline routing protocol and IPv4 as the baseline network address format. These provide a common denominator on which to create end-to-end paths and name destinations.

3.2 Integrated advertisements

In D-BGP, integrated advertisements (IAs) extend BGP's advertisements. Each IA compactly describes a path that can be used to reach a destination address, named using the baseline address format (i.e., an IPv4 prefix). The path represents a *best path* choice toward the destination of the router or AS that created the advertisement. Figure 4 shows an example IA. It has been populated with

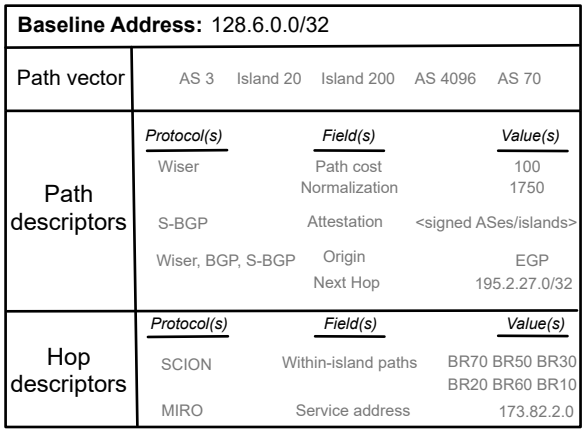

Figure 4: A basic integrated advertisement and example fields.

several protocols' control information to illustrate how it could be used. Only fields relevant to multi-protocol support are shown, so standard BGP fields, such as withdrawn prefixes and non-transitive community attributes have been omitted.

IAs provide multi-protocol support by including three key fields: a *path vector*, per-protocol *path descriptors*, and *hop descriptors*. The first field states the path and is used to avoid routing loops. It is similar to the path-vector field in BGP advertisements today, except it is expanded to list all ASes or islands involved in routing on a path, regardless of protocol used. Island IDs can be new values issued by some governing body (e.g., the ARIN [1]) or simply a concatenated list of islands' border ASes. The second field bears resemblance to BGP's community attributes, but is always transitive and is explicitly structured for multi-protocol support. To our knowledge, BGP does not have an explicit analog to the third field. The rest of this section describes the latter two fields in detail.

Path descriptors describe per-protocol attributes of the entire path. Critical fixes can use them to encode their bespoke control information. The example shown in Figure 4 includes Wiser' path cost [31] and S-BGP's route attestations [25, 51]. Other potential path descriptors include Xiao et al's and EQ-BGP's QoS metrics [4, 46].

Hop descriptors describe attributes of individual routing hops (i.e., islands or ASes) on the path that support enhanced functionalities not offered by the baseline or its critical fixes. The information listed by a given hop descriptor includes the functionality(ies) offered by the corresponding hop and any control information needed to use it. Hop descriptors are listed in the order in which data packets will traverse them, allowing sources to layer headers in the right sequence when encapsulating traffic to use the desired functionalities.

Custom and replacement protocols can use hop descriptors to include bespoke information relevant to their island. For example, a MIRO or Arrow island could originate an advertisement listing the service it offers and the IP address of a portal that potential customers could contact to identify alternate paths offered and coordination necessary to use them. This enables discovery. Alternatively, a SCION island could include its extra intra-island paths that other SCION islands can use to route packets to the destination. These paths are specified at the level of border routers (e.g., BR 50).

We note that hop descriptors can also be used to translate between the baseline address format and other address formats, such as content names [49] or IPv6 (not shown in Figure 4). For example, an island that routes traffic based on content names in addition to BGP can list the content name associated with an IP address in a hop descriptor, allowing other ASes that use content names to route traffic using it.

To reduce IA sizes, protocols listed within IAs share control information that is identical across them. Such sharing can drastically reduce overhead for critical fixes, which share most of their control information with BGP (e.g., S-BGP uses only one extra field). In Figure 4, BGP, S-BGP, and Wiser all share control information. Custom and replacement protocols' contribution to IA size will be small because they do not need to disseminate much bespoke control information outside their islands. For example, a SCION island that disseminates 5 within-island paths, each consisting of 5 intra-island hops, will only need to disseminate about 200 bytes of control information to describe them (this assumes 4-byte border router IDs). IAs can be compressed to further reduce their size.

3.3 Advertisement processing

D-BGP modifies BGP's advertisement processing to support IAs, provide multi-protocol support, and provide pass-through functionality. A router with D-BGP's advertisement processing is shown in Figure 5. In addition to standard import/export filters, which allow operators to enforce polices, such as valley-free routing [35], D-BGP's processing includes the following novel components.

First, it includes multiple *decision modules*, corresponding to BGP and the critical fixes and replacement protocols it supports. Each decision module encapsulates the data structures (e.g., RIBs) and algorithms a given protocol uses to choose best paths. Figure 5 shows a decision module for the baseline BGP and for some critical fix.

Second, it includes an *IA factory*, which replaces similar functionality for BGP's advertisements. The IA factory is responsible for receiving IAs, communicating per-protocol information contained in IAs to relevant decision modules, and creating new advertisements for the best path selected. It also adds pass-through information associated with the best-path chosen.

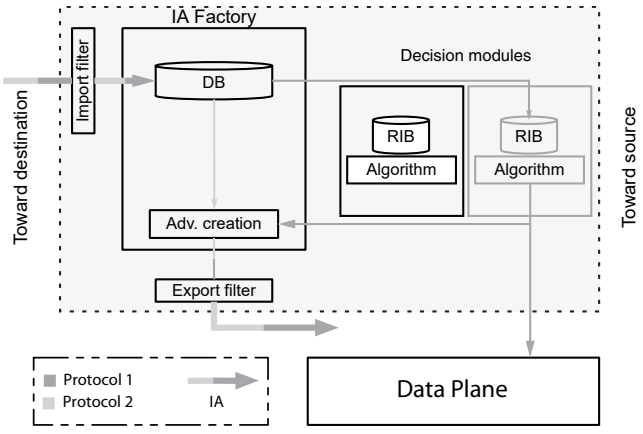

Figure 5: A router's advertisement processing.

When a D-BGP router receives a new IA, the following steps occur to choose a new best path for the prefix it names. First, the IA factory decodes the message and identifies control information for the most recent protocol version supported by the router. It also stores the incoming IA in a database for later use. Second, this control information is forwarded to the relevant decision module, which selects a new best path. It does so by comparing the control information it has just received to that from previously received IAs for the same prefix, which are stored in its RIB. It outputs both the new best path and modified control information (e.g., a modified path cost that reflects the router's or relevant AS's current load or a new route attestation). Third, the path choice is installed in the router's forwarding table.

Fourth, the IA factory builds a new IA for the selected best path. To do so, it inserts modified control information for the protocol used to select the best path in the new IA. It also adds pass-through information by indexing into its database of stored IAs to retrieve the incoming IA for the best-path chosen. It opaquely copies over control information for all protocols that were not used for best-path selection from this IA into the new advertisement, effectively tunneling this information across the selected path.

3.4 Example usage

Evolvability for critical fixes: Figure 6 illustrates the result if the ASes in the scenario from Section 2.2 supported D-BGP. E1 and E2 include Wiser's path costs in IAs they advertise to ASes in the gulfs. These path costs are passed through so that the source AS (S) is able to see them and use them to select the lower cost, longer path. E1 and E2 are still at the mercy of ASes that run only BGP, but their situation incrementally improves as more non-contiguous ASes or islands adopt Wiser.

Discovery for custom protocols: D-BGP allows the transit AS in the example from Section 2.3 to discover and use the MIRO AS's alternate path as follows. First, the MIRO AS (M) uses IAs to advertise a path to a service portal it provides. A hop descriptor included within the IA includes the ID of the service offered and the custom coordination required to use this service portal (e.g., a

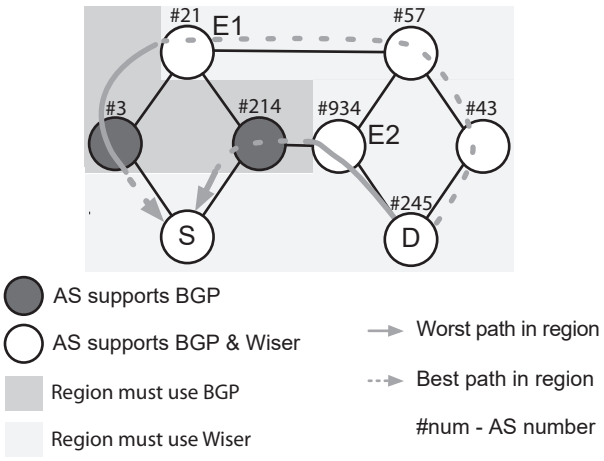

Figure 6: S sees path costs in IAs, so it chooses the lowest-cost one.

specific protocol). Second, the transit AS (T) contacts the service portal to negotiate the alternate path to the destination and the data-plane encapsulation technique (e.g., an additional prefix) that will be used to cross gulfs and selectively route the transit (T)'s traffic. Third, the transit (T) uses the necessary encapsulation technique to tunnel its traffic destined for the destination AS. As an optimization, the initial advertisement could include a list of the most popular alternate paths the MIRO-enabled AS provides (e.g., alternate paths to Google).

Evolvability for replacement protocols: Figure 7 illustrates how D-BGP enables evolvability for the SCION scenario discussed in Section 2.4. The edge AS's border router in the large SCION island creates an IA for the prefix advertised that includes control information for a SCION path that has been redistributed into BGP (or some critical fix). It also includes other paths other SCION islands can use to route traffic to the description within a hop descriptor. When the SCION source (S) receives the IA, it extracts the SCION-specific control information, chooses a within-island path, and encodes it in a SCION header, which it attaches to data packets. It also encapsulates the packet with an IPv4 header so that the packet can cross gulfs. When receiving packets, SCION border routers in the rightmost island de-encapsulate packets to probe for a SCION header. If it exists, it routes packets using the specified path choice.

3.5 Supporting IA aggregation

To further reduce message sizes, we allow IAs to describe a group of paths that can be used to reach a set of contiguous addresses (i.e., a broad prefix). Such aggregation affords protocols the opportunity to summarize control information across all the paths listed in the IA. For example, Wiser [31] could list a single average path cost for all of the paths summarized in an aggregated IA instead of including separate ones in individual IAs. Additional savings in message sizes and processing overhead results from having to send fewer aggregated IAs than individual ones. IA aggregation is similar to the aggregation that occurs in BGP today [37].

Figure 8 shows the revised IA structure that supports aggregation. Note that it is associated with a broad prefix that covers many destination addresses (i.e., a /16, instead of the /32 used in Figure 4). The *path vector* field now allows for an OR ([]) relationship, allowing the IA to compactly state all the

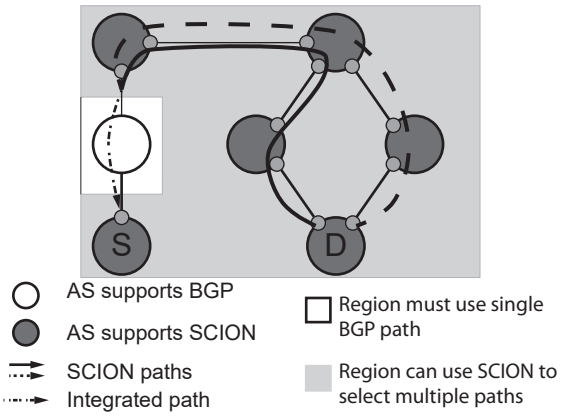

Figure 7: S sees both paths in the integrated advertisement.

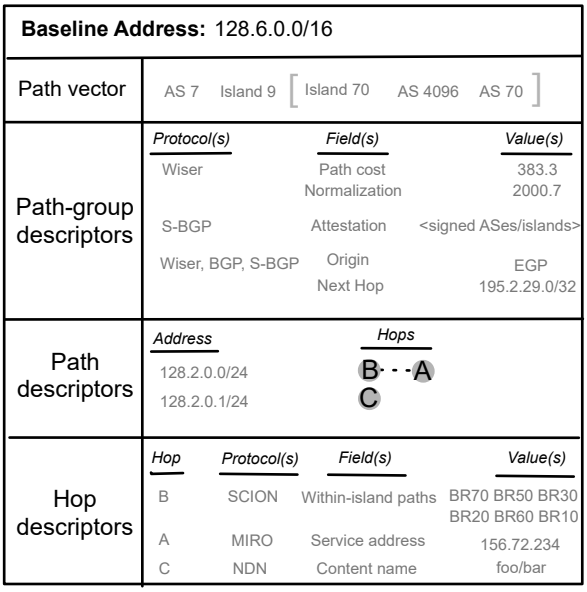

Figure 8: An integrated advertisement that supports aggregation.

possible ASes or islands that could be used to reach destinations covered by the broad prefix. It also contains a new *path-group descriptor* field, which allows individual protocols to summarize control information across all the paths described by the IA. Figure 8 shows an average path cost for Wiser and an aggregated path attestation for S-BGP [51].

Path descriptors now describe individual paths contained in an IA whose hop(s) offer enhanced functionality(ies). They are needed because these functionalities are specific to the relevant path and hop and so cannot be summarized in path-group descriptors. Hops are listed in the order that data packets will traverse them so that sources can encapsulate data packets accordingly. Corresponding hop descriptors are identical to the basic IA case described earlier and list bespoke control information. Figure 8 shows control information for SCION [50], NDN [49], and MIRO [47].

A router that wishes to aggregate a set of IAs that it receives must support all of the protocols listed in their path-group descriptor fields. Also, all of those protocols must agree to summarize control information for the same range of addresses (e.g., Wiser may wish not to summarize control information across paths with very different costs). These requirements are necessary to avoid potential for the outgoing aggregated IA to be larger than the sum of the incoming ones. This can happen when different protocols' control information is summarized across different (potentially overlapping) subranges of the outgoing IA's broad prefix. Since stub ASes' routers will likely support the same set of protocols, we expect them to be capable of disseminating aggregated IAs for the majority of addresses they own. However, we expect potential for aggregation to be limited at transits and tier-1 ASes.

3.6 Limitations

D-BGP's method of enabling evolvability is subject to the limitations and policies of the protocols used within gulfs. As such, indiscriminate path choices within gulfs may reduce the benefits afforded by a new protocol (e.g., alternate paths or higher-quality paths) when routing across them. D-BGP requires sources to layer data packets with the headers needed to use custom or replacement protocols. Such layering makes it difficult for D-BGP to support protocols, such as Pathlet routing [15] that modify the header at locations other than stub ASes/islands.

4 A rich & evolvable Internet

To illustrate the utility of D-BGP, Figure 9 illustrates the type of rich and evolvable Internet with many routing options it could enable. This rich Internet is comprised of several different types of protocols, including BGP, different critical fixes to it, different types of replacement protocols (path-based, multi-hop), and custom protocols. Our example uses protocols already discussed in this paper (or extensions to them), but others could also be used. Examples include other types of critical fixes [4, 46], other types of replacement protocols (multi-hop [13] or path-based [48]), and other custom protocols [34]. The rest of this section further describes our example. We also illustrate IAs at interesting points in the topology to show how they enable this rich world and how they are constructed.

In our rich Internet example, a governing body (e.g., the IETF) has decided on an ordering to critical fixes. WS-BGP is preferred over Wiser [31], which is preferred over BGP. WS-BGP is similar to S-BGP [25]. But, because it is a later critical fix than Wiser, it also understands and disseminates Wiser's path cost metric and normalization factor. Wiser shares all of its control information with BGP, except for a path cost and normalization factor. WS-BGP shares all of its control information with Wiser except for an extra route attestation. These critical fixes reflect the governing body's efforts to prevent route hijacking (e.g., ISPs in China posing as popular web sites, such as cnn.com [38]) by adding secure paths and their efforts to fix BGP's broken ability to help ASes limit ingress traffic [12, 31] by adding path costs.

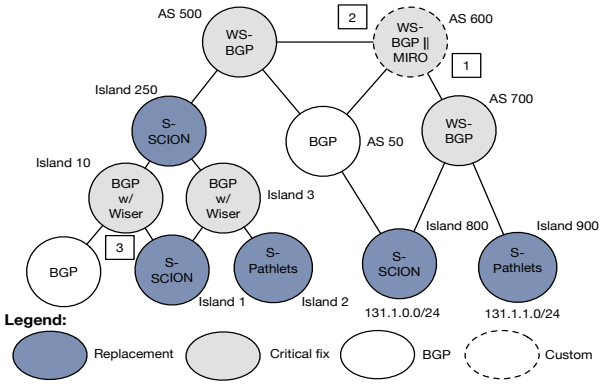

Figure 9: A rich & evolvable Internet that D-BGP could facilitate.

We assume that the governing body has decided that a version of SCION [50] that supports route attestations (S-SCION) will eventually be the replacement baseline protocol for the Internet because it gives sophisticated stub networks a richer set of path options to choose from. Because a secure version of Pathlet routing [15] (S-Pathlets) allows islands to use rich policies, some islands use it internally. But, we assume that it has not been ratified as a protocol that will replace BGP and its critical fixes. We assume S-SCION and S-Pathlets redistribute their routes (and route attestions) into WS-BGP. This allows them to leverage Wiser's path costs when routing among Wiser-enabled islands and maintain secure routes and path costs when routing via WS-BGP-enabled islands.

In our example, all ASes may eventually converge to using S-SCION because D-BGP incentivizes adoption by increasing the speed at which AS operators can enjoy the benefits of a new protocol. Even if this does not happen, D-BGP will still enable non-contiguous ASes that support the same set of protocols to observe some of the benefits those protocols afford. Security benefits cannot be achieved until all ASes on a routing path support a secure protocol (i.e., S-SCION, S-Pathlets, or WS-BGP) [30].

To illustrate what the contents of an aggregated IA might look like, Figure 10 shows the aggregated IA received by AS 600 from AS 700 (point *[1]* in the topology). It describes paths to reach the 131.1.0.0/24 and 131.1.1.0/24 prefixes, which were originated by Islands 800 and 900 respectively. We assume islands 800 and 900 disseminated aggregated advertisements for the broad /24 prefix themselves, but they could have sent non-aggregated ones as well. AS 700 was able to aggregate the IAs it received from both islands because they form a larger contiguous address block and because it supports the same protocols' as those specified in the incoming IAs' path-group descriptors.

The IA in Figure 10 contains five path-group descriptors. The first three correspond to those used by WS-BGP: a path cost, a normalization factor, and a route attestation. The rest are shared between BGP and WS-BGP (Note that many shared fields have been omitted in Figure 10). All of the path-group descriptors have been summarized over all of paths described in the IA.

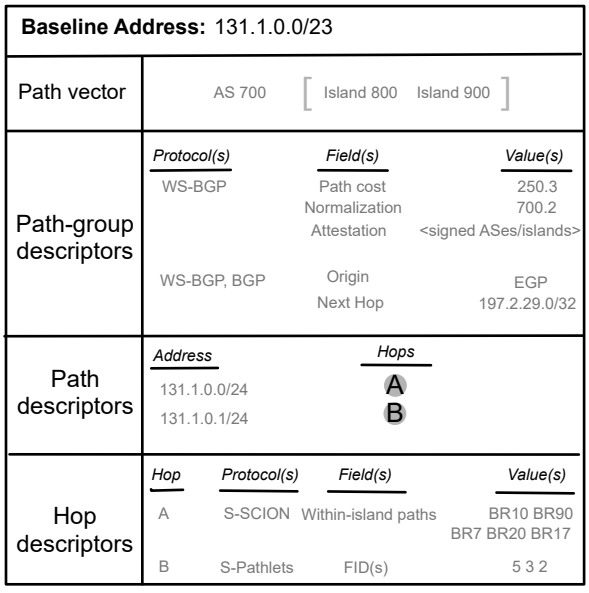

Figure 10: IA that would be received at point 1 of our rich world.

The advertisement from AS 700 to 600 also includes two path descriptors, associated with the more specific prefixes originated by Islands 800 and 900. For simplicity, we only list path descriptors associated with the two /24 addresses advertised by islands 800 and 900, but, additional path descriptors for more specific prefixes could be included as well. Islands running replacement and custom protocols will perform some form of aggregation internally, so we do not expect them to disseminate prefixes for individual destinations (i.e., /32s).

Both path descriptors contain one hop descriptor. The one associated with S-SCION contains the paths within island 800 that other S-SCION islands can use to reach destinations covered by prefix 131.1.0.0/24. The one associated with S-Pathlets contains a sequence of forwarding IDs used by Pathlet routing to reach destinations covered by prefix 131.1.1.0/24. These IDs correspond to path fragments that can be assembled into larger path fragments by other Pathlet islands.

We illustrate how islands running custom protocols could use IAs to facilitate their discovery at point *[2]* in Figure 9. At this point in the topology, AS 600 disseminates regular IAs for best paths to prefixes advertised to it and one additional IA, which it originates, for its MIRO service. This IA for the MIRO service describes a path to a specific prefix (i.e., a /32) corresponding to the service portal interested customers should contact to identify the alternate paths AS 600 offers and coordinate their use. In our example, AS 600 chooses the incoming advertisement from AS 700 as the best path to reach 131.1.0.0/24. The MIRO service can offer the path through AS 50 as an alternate, for example, if AS 700 proves to be unreliable.

The IA received by the S-SCION Island 1 from Island 10 for the aggregate prefix 131.1.0.0/23 is shown in Figure 11 (labeled as *[3]* in Figure 9). We illustrate it to show how a traffic source might use an IA that contains a rich set of protocols to select paths and forward traffic. After receiving the IA, the S-SCION island can learn that the paths described by it are not secure by examining the route attestation to find that it was not signed by its upstream neighbor. The path cost contained in the

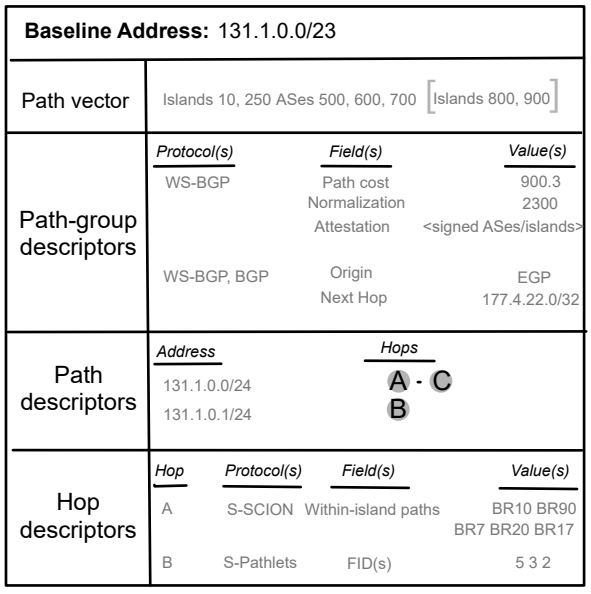

Figure 11: IA received at point 3 of our rich-world topology.

IA represents the cost to all Wiser and WS-BGP islands of using the advertised path. The S-SCION island can use the path cost contained in this advertisement and the one it receives from Island 3 to decide which path to choose as its best path. It could also choose best paths based on which ones contain custom or replacement protocols it wants to use.

Individual path descriptors in Figure 11 contain a sequence of two routing hops for prefix 131.1.0.0/24. These identify the paths that can be used within the two downstream SCION islands on the path (Islands 250 and 800). To use these within-island paths, the S-SCION Island 1 must encapsulate its choices between two IPv4 headers. The path descriptor for prefix 131.1.1.0/24 lists the forwarding IDs advertised by the other Pathlet island in our topology (Island 900).

5 Evaluation

In this section, we evaluate the control-plane overhead of D-BGP and its ability to incentivize adoption of critical fixes and replacement protocols. We seek to answer the following two questions. First, what is the control-plane overhead of using D-BGP to facilitate an Internet that runs many inter-domain routing protocols? (See Sections 5.1 and 5.2.) Second, how does D-BGP's ability to allow non-contiguous islands to route traffic amongst each other using a new protocol accelerate the incremental benefits afforded to them as more ASes/islands adopt the same protocol? (See Sections 5.3 and 5.4).

5.1 Control-plane overhead methodology

We evaluate control-plane overhead by estimating properties of IAs that would be received at a tier-1 AS in an Internet that is using D-BGP to run multiple inter-domain routing protocols. We analyze three types of overhead: the size of individual IAs that are received (indicative of per-IA serialization cost at the tier-1 AS), the number of IAs that are received (reflective of CPU cost), and aggregate size of all IAs received (reflective of total overhead and the amount of state that must kept at the tier-1 AS). Tier-1 ASes reside at the top of the Internet hierarchy, so they will see the highest overheads.

To derive estimated IAs sizes and their number, we use characterizations of the Internet topology and protocols' expected control-message sizes culled from recent research and RFCs [5, 10, 20, 29, 37] Table 1 lists key parameters, the ranges of values we consider, and our reasoning behind our choices for these ranges. Whenever possible, we choose ranges based on estimates in the literature. For parameters whose values are more uncertain (e.g., number of critical fixes), we consider a broad range of possible values to allow for future protocols' as-yet undetermined needs.

5.2 Control-plane overhead results

Table 2 shows our results. IA sizes are further broken down into contribution by protocol type (critical fix or custom/replacement). For each overhead type estimated, we list a range of minimum and maximum values, derived from the equations, parameters, and values discussed in Table 1.

We find that a basic analysis that assumes individual IAs received at a tier-1 will contain information for all protocols yields very large aggregate overheads. + *Avg. path lengths* improves

Parameter	Variable	Ranges considered	Rationale
		General Internet topology	
# of stubs == # of prefixes	P	10,000 - 100,000	Assume single aggregated prefix / stub AS; 10,000s of stubs [10, 20, 29]
Avg. BGP path length	PL	3 - 5	Analysis of routing tables [5]
Avg. # of stubs / provider	SpP	2 - 5	Range around node degree in Internet topology [20]
		Critical fixes (CFs)	
# of critical fixes	CFs	10s - 100s	Assume governing body will limit total number
Critical fixes / path	$CFs\ /\ path$	3-5	Assume one critical fix per hop on path
Control info / critical fix	$CI\ /\ CF$	8 KB - 256 KB	8 KB is max size for BGP [37]; up to 256KB for future protocols
Unique control info / critical fix	CFu	0.05 - 0.25	Most critical fixes share majority of control info w/each other
		Custom or replacement protocols (CRs)	
# of custom or replacements	CRs	10 - 1000s	Many possible because large fraction need not be regulated
Custom or replacements / path	$CR\ /\ path$	3 - 5	Assume one custom/replacement per hop on path
Control info / custom or replacement	$CI\ /\ CR$	100 B - 10 KB	Not much info needs to be disseminated outside islands
		Aggregation potential at stubs' providers	
w/Multiple protocols	agg_m	0.1 - 0.2	Small because provider must support same protocols as stub
w/Single protocol	agg_s	0.25 - 0.75	Some prefixes will not be aggregated due to traffic engineering

Table 1: Parameters and ranges considered for analyzing D-BGP's control-plane overhead.

this analysis by accounting for the fact that IA size is a function of the number of protocols on a routing path, not the total number. This reduces our estimate of maximum aggregate overhead by an order of magnitude. + *Sharing* improves our analysis by accounting for the fact that many critical fixes can share the majority of their control information (e.g., S-BGP disseminates only an additional route attestation compared to BGP). This yields significant savings and reduces both our minimum and maximum estimates by an additional order of magnitude. + *IA aggregation* allows stub ASes' providers to aggregate incoming IAs if they support all of the same protocols for which path-descriptor information is listed in an IA (e.g., they support the same critical fix). In our analysis, we assume that this requirement means that only a small fraction of IAs can be aggregated, yielding only small improvements.

We also compare D-BGP's overheads with multiple protocols to the case where only a single protocol is running, which should be similar to the overheads seen today with BGP. The single protocol case allows for a factor of 2-4x increase in aggregation levels. Despite this and our assumption of 3-5 critical fixes and 3-5 custom/replacement protocols on routing paths, we find that D-BGP's aggregate overhead to be within a factor of 3 of the single protocol case. This is largely a result of the savings due to sharing of critical fixes' control information.

5.3 Incremental benefits methodology

We analyzed D-BGP's ability to accelerate incremental benefits by simulating protocols' path-selection choices on an AS-level topology in which ASes are running both BGP and a given new protocol. Our AS-level topology is generated by BRITE, which is configured to generate 500 ASes using use a Waxman model with (α = 0.15 and β = 0.25) [7, 9, 19]. Both new protocols' and BGP's path selection choices are simulated using a modified version of the BGP simulator used by by John et al. and Peter et al. [23, 34]. Path choices always are valley free [35] and policy compliant.

For each new protocol, we measure the benefits afforded to upgraded ASes as a function of

	Contribution to IA size by....			
Name	**CFs**	**CRs**	**# of advertisements**	**Total overhead**
Basic	$CFs \cdot \frac{CI}{CF}$	$CRs \cdot \frac{CI}{CR}$	P	
	80 KB - 26 MB	1 KB - 10 MB	10,000 - 100,000	0.8 GB - 3515 GB
+ Avg. path lengths	$\frac{CFs}{path} \cdot \frac{CI}{CF}$	$\frac{CRs}{path} \cdot \frac{CI}{CR}$	"	
	24 KB - 1.3 MB	0 KB - 50 KB	"	0.2 GB - 132 GB
+ Sharing	$\frac{CFs}{path} \cdot \frac{CI}{CF} \cdot (CFu) + \frac{CI}{CF} \cdot (1 - CFu)$	"	"	
	9 KB - 0.5 MB	"	"	0.09 GB - 54 GB
+ IA aggregation	"	$\sim \frac{CRs}{path} \cdot \frac{CI}{CR}$	$\frac{P}{SpP} + (PpS - 1) \cdot (1 - agg_m) \cdot \frac{P}{SpP}$	
	"	$\sim 0KB - \sim 50KB$	8,400 - 95,000	.07 GB - 51 GB
Single protocol	$\frac{CI}{CF}$	0	$\frac{P}{(SpP)} + (PpS - 1) \cdot (1 - agg_s) \cdot \frac{P}{SpP}$	
	8 KB - 256 KB	0	4,000 - 75,000	0.03 GB - 18 GB

Table 2: Control-plane overhead of D-BGP. This table shows estimated IA sizes and number of IAs that would be received at a tier-1 AS as a function of various parameters. Equations or values identical to the previous corresponding entry are marked with a ".

the percentage of upgraded ASes in the topology (i.e., the adoption level). The slope of this value corresponds to incremental benefits. We consider two cases: an Internet that is running D-BGP, which allows non-contiguous islands running the same protocol to choose paths and route traffic to each other using it, and an Internet w/o D-BGP, which does not facilitate such functionality. As such, the new protocol can only be used within ASes or islands and not across them. Comparing the these two cases illustrates D-BGP's ability to accelerate incremental benefits. In both cases, ASes that have not been upgraded (i.e., within gulfs) choose paths with the lowest path length [6].

We examine how incremental benefits differ for different types of protocols by considering two fundamentally different archetypes. The first corresponds to a type of critical fix and the second one is representative of a replacement protocol. The archetypes are summarized below.

A bottleneck metric critical fix: With this archetype, a metric is disseminated with advertisements and is used to influence path selection. If an AS's contribution to the metric is less than the value in incoming advertisements, it replaces the incoming value with its value before selecting best paths. As such, incremental benefits and the amount of benefits observed at different adoption levels depends heavily on which ASes have been upgraded. Examples of critical fix that fit this archetype include the one described by Xiao et al. [46] and ones that aim to choose paths with high residual or bottleneck bandwidth.

Note that the objective of our critical fix archetype is a function of all ASes on a routing path, including ones in gulfs. As such, benefits afforded by a protocol may decrease compared to the 0% adoption case. This will occur when only a small percentage of ASes have upgraded and the incomplete information on which they choose best paths causes them to choose worse ones than what BGP was providing.

A generic replacement protocol: With this archetype, a small amount of control information is disseminated outside islands. With D-BGP, whether this control information is delivered to other upgraded islands depends on the best-path choices of ASes in gulfs. Without D-BGP, this

information can only be leveraged by contiguous deployments. Examples that fit this archetype include Pathlets [15], SCION [50], and NIRA [48]. For simplicity, we assume this information is extra within-island paths to destinations in this section. We cap the number of paths that can be disseminated by an AS 10 to more accurately reflect real-world scenarios.

For the bottleneck metric experiment, we assign values on links uniformly between a range of 10 and 1024 (the range doesn't affect results much). Upgraded ASes are chosen randomly, reflecting the ideal case of providing ASes the flexibility to deploy a new protocol at their convenience. Each of our results reflect the average of 10 trials, each with different random seeds. Benefits are plotted at increments of 10% adoption of the new protocol.

We also analyzed incremental benefits for an **additive sum critical fix** archetype, in which ASes independently add their contribution to the metric being disseminated. Compared to the bottleneck metric critical fix, this archetype is much more sensitive to individual ASes' contributions. We found that D-BGP always provided higher benefits than the w/o D-BGP case. However, we do not include results for it because we found that the exact values for our results were highly dependent on the distribution of link values we chose. Protocol corresponding to this archetype is Wiser [31] and Q-BGP [4]. We also do not evaluate benefits for protocols that require all ASes on routing paths to be upgraded ones because D-BGP does not provide extra value for them.

5.4 Incremental benefits results

Figure 12 shows the benefit afforded to upgraded ASes by the *bottleneck metric critical fix* archetype as a function of the percentage of them that have upgraded. We measure benefit as the average bottleneck metric value associated with the best path chosen at each AS. As more ASes upgrade, this value increases. The status quo line represents the average bottleneck value associated with the best path chosen at each AS with 0% adoption. The best case line shows the value when all ASes have upgraded.

We see that with D-BGP, the incremental benefit (the slope of the D-BGP line) is almost always greater than the w/o D-BGP case. The crossover point is at 80% adoption, when large upgraded

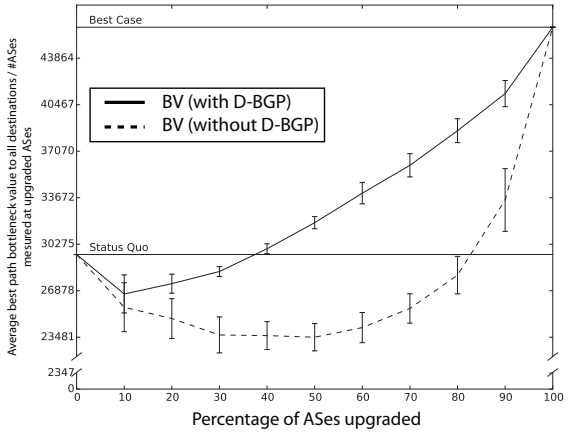

Figure 12: Incremental benefits for bottleneck-metric archetype

islands in the w/o D-BGP case start to connect and very quickly see large benefits as a result of doing so. The fact the D-BGP accelerates incremental benefits at lower adoption levels means that the total benefits facilitated by D-BGP always outperform the w/o D-BGP case. The maximum percentage difference in the bottleneck metric for both cases is 37% at 80% adoption.

With D-BGP, benefits compared to the status quo decline initially (until 30% adoption) because upgraded ASes make best-path decisions using very incomplete information. W/o D-BGP, benefits decline until 80% adoption. This difference arises because at equivalent adoption levels, non-contiguous ASes in the D-BGP case always have more complete information on which to base their best-path choices.

Figure 13 shows the benefit afforded by the *generic replacement protocol* to upgraded ASes as a function of the percentage of them that have upgraded. We measure benefit as the number of extra paths available to destinations at upgraded stubs, since these are the entities that would be able to use these paths. We see that with D-BGP, the incremental benefit (the slope of the D-BGP line) is always greater than the w/o D-BGP case until 70% of ASes are upgraded. Once again, the fact that D-BGP accelerates incremental benefits at lower adoption levels means that the total benefits facilitated by D-BGP is always greater than the w/o D-BGP case. The maximum percentage difference in number of paths to destinations is 89% at 60% adoption.

6 Related work

Several research efforts focus on evolvability for the data plane [17, 42, 44, 45, 49]. Our research complements them by focusing on the control plane. Other research has focused on requirements for general network evolvability [8, 14, 36]. Those listed by Ratnasamy et al. [36] are compatible with our requirements, but we extend them to inter-domain routing.

BGP's community attributes [37] could be leveraged to implement the path descriptors used by our IAs, however they are not usually passed through to indirect neighbors [24] (we also verified this via our own experiments on PEERING [40]). Multi-protocol extensions to BGP [3] allow direct

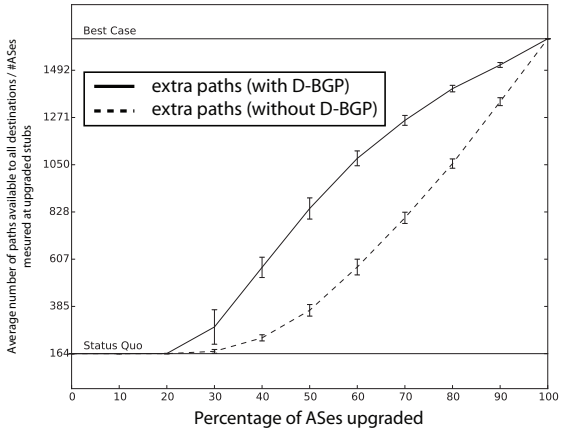

Figure 13: Incremental benefits for replacement archetype

neighbors to name the same physical destination using different address formats (e.g., IPv4 and IPv6). This enables evolvability for addressing within ASes or islands, but not across them.

Koponen et al. [26] propose using Pathlet routing [15] to enable evolvability—i.e., as the new baseline—because of its ability to emulate many routing protocols. Nikkah et al. [33] explore what protocol-specific factors lead to a new protocol's' successful adoption (e.g., ability to improve performance or backward compatibility). Our work complements these efforts by providing mechanisms to make deployment of promising protocols easier. Software-defined exchanges [16] present a promising point at which to jump start deployment of new inter-domain routing protocols, however D-BGP would still be needed to allow non-contiguous neighbors to enjoy a new protocol's benefits and to enable Internet-wide discovery of custom protocols.

7 Summary

BGP cannot easily be evolved. This prevents new protocols from being widely deployed. Based on requirements identified by an analysis of key evolvability scenarios, we identified key building blocks needed to support evolvability and modified BGP to include them. Our modified version of D-BGP can support evolution to a wide range of critical fixes to BGP, sophisticated BGP replacements, and protocols that run in parallel with BGP to provide functionality it doesn't.

References

[1] American registry of internet numbers.

[2] D. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, and G. Swallow. RSVP-TE: Extensions to RSVP for LSP Tunnels. RFC 3209, IETF, December 2001.

[3] T. Bates, R. Chandra, D. Katz, and Y. Rekhter. Multiprotocol extensions for bgp-4. RFC 4760, IETF, January 2007. http://www.rfc-editor.org/rfc/rfc4760.txt.

[4] A. Beben. EQ-BGP: an efficient inter-domain QoS routing protocol. ... *Networking and Applications*, 2:5 pp., 2006.

[5] Bgp routing table analysis reports. http://bgp.potaroo.net/as6447/.

[6] M. Caesar and J. Rexford. BGP routing policies in ISP networks. *Network, IEEE*, 19(6):5–11, 2005.

[7] W. Chen, C. Sommer, S.-H. Teng, and Y. Wang. A compact routing scheme and approximate distance oracle for power-law graphs. *ACM Trans. Algorithms*, 9(1):4:1–4:26, Dec. 2012.

[8] D. D. Clark, J. Wroclawski, K. R. Sollins, and R. Braden. Tussle in Cyberspace: Defining Tomorrow's Internet. *IEEE/ACM Transactions on Networking*, 13(3):462–475, June 2005.

[9] Q. Duan, E. Al-Shaer, and H. Jafarian. Efficient random route mutation considering flow and network constraints. In *Proc. IEEE Conference on Communications and Network Security (CNS)*, Oct 2013.

[10] B. Edwards, S. Hofmeyr, G. Stelle, and S. Forrest. Internet Topology over Time. *arXiv.org*, Feb. 2012.

[11] D. Farinacci, V. Fuller, D. Meyer, and D. Lewis. The Locator/ID Separation Protocol (LISP). RFC 6830, IETF, January 2013. http://www.rfc-editor.org/rfc/rfc6830.txt.

[12] N. Feamster, H. Balakrishnan, and J. Rexford. Some foundational problems in interdomain routing. In *Proc. HotNets*, November 2004.

[13] I. Ganichev, B. Dai, P. B. Godfrey, and S. Shenker. YAMR: yet another multipath routing protocol. *ACM SIGCOMM Computer Communication Review*, 40(5):13–19, Oct. 2010.

[14] A. Ghodsi, S. Shenker, T. Koponen, A. Singla, B. Raghavan, and J. Wilcox. Intelligent Design Enables Architectural Evolution. In *Proc. HotNets*, Nov. 2011.

[15] P. B. Godfrey, I. Ganichev, S. Shenker, and I. Stoica. Pathlet Routing. In *Proc. SIGCOMM*, Aug. 2009.

[16] A. Gupta, L. Vanbever, M. Shahbaz, S. P. Donovan, B. Schlinker, N. Feamster, J. Rexford, S. Shenker, R. Clark, and E. Katz-Bassett. SDX: a software defined internet exchange. In *Proc. SIGCOMM*, Aug. 2014.

[17] D. Han, A. Anand, F. Dogar, B. Li, H. Lim, M. Machado, A. Mukundan, W. Wu, A. Akella, D. G. Andersen, J. W. Byers, S. Seshan, and P. Steenkiste. XIA: efficient support for evolvable internetworking. In *Proc. NSDI*, Apr. 2012.

[18] C.-Y. Hong, S. Kandula, R. Mahaan, M. Zhang, V. Gill, M. Nanduri, and R. Wattenhofer. Achieving High Utilization with Software-Driven WAN. In *Proc. SIGCOMM*, June 2013.

[19] Y.-Y. Huang, M.-W. Lee, T.-Y. Fan-Chiang, X. Huang, and C.-H. Hsu. Minimizing flow initialization latency in software defined networks. In *Proc. Network Operations and Management Symposium (APNOMS)*, Aug 2015.

[20] B. Huffaker, M. Fomenkov, and k. claffy. Internet Topology Data Comparison. Technical report, Cooperative Association for Internet Data Analysis (CAIDA), May 2012.

[21] Internet engineering task force. http://www.ietf.org.

[22] S. Jain, A. Kumar, S. Mandal, J. Ong, L. Poutievski, A. Singh, S. Venkata, J. Wanderer, J. Zhou, M. Zhu, J. Zolla, U. Hölzle, S. stuart, and A. Vahdat. B4: Experience with a Globally-Deployed Software Defined WAN. In *Proc. SIGCOMM*, June 2013.

[23] J. P. John, E. Katz-Bassett, A. Krishnamurthy, T. Anderson, and A. Venkataramani. Consensus routing: The internet as a distributed system. In *Proc. NSDI*, 2008.

[24] E. Katz-Bassett, C. Scott, D. R. Choffnes, I. Cunha, V. Valancius, N. Feamster, H. V. Madhyastha, T. Anderson, and A. Krishnamurthy. LIFEGUARD: practical repair of persistent route failures. In *Proc. SIGCOMM*, Aug. 2012.

[25] S. Kent, C. Lynn, and K. Seo. Secure Border Gateway Protocol (S-BGP). *IEE Journal on Selected Areas in Communications*, 18(4):582–592, 2000.

[26] T. Koponen, S. Shenker, H. Balakrishnan, N. Feamster, I. Ganichev, A. Ghodsi, P. B. Godfrey, N. McKeown, G. Parulkar, B. Raghavan, J. Rexford, S. Arianfar, and D. Kuptsov. Architecting for Innovation. *SIGCOMM Computer Communication Review*, 41(3):24–36, July 2011.

[27] N. Kushman, S. Kandula, D. Katabi, and B. M. Maggs. R-BGP: staying connected In a connected world. In *Proc. NSDI*, Apr. 2007.

[28] F. Le, G. G. Xie, and H. Zhang. Understanding Route Redistribution. In *Proc. ICNP*, 2007.

[29] K. Liu, B. Jabbari, and S. Secci. Understanding transit-edge routing separation: Analysis and characterization. In *Proc. Network of the Future (NOF)*, Nov 2011.

[30] R. Lychev, S. Goldberg, M. Schapira, and R. Lychev. BGP security in partial deployment: is the juice worth the squeeze? *ACM SIGCOMM Computer Communication Review*, 43(4):171–182, Aug. 2013.

[31] R. Mahajan, D. Wetherall, and T. Anderson. Mutually controlled routing with independent ISPs. In *Proc. NSDI*, Apr. 2007.

[32] Z. M. Mao, R. Govindan, G. Varghese, and R. H. Katz. Route flap damping exacerbates internet routing convergence. In *Proc. SIGCOMM*, Aug. 2002.

[33] M. Nikkhah, C. Dovrolis, and R. Guérin. Why didn't my (great!) protocol get adopted? In *Proc. HotNets*, 2015.

[34] S. Peter, U. Javed, Q. Zhang, D. Woos, T. Anderson, and A. Krishnamurthy. One tunnel is (often) enough. In *Proc. SIGCOMM*, Aug. 2014.

[35] S. Y. Qiu, P. D. McDaniel, and F. Monrose. Toward Valley-Free Inter-domain Routing. In *Proc. IEE ICC*, 2007.

[36] S. Ratnasamy, S. Shenker, and S. McCanne. Towards an Evolvable Internet Architecture. *SIGCOMM Computer Communication Review*, 35(4):313–324, Aug. 2005.

[37] Y. Rekhter and T. Li. A Border Gateway Protocol 4 (BGP-4). RFC 4271, IETF, Jan 2006. http://www.rfc-editor.org/rfc/rfc1771.txt.

[38] Chinese ISP hijacks internet. http://www.bgpmon.net/chinese-isp-hijacked-10-of-the-internet/.

[39] R. R. Sambasivan, D. Tran-Lam, A. Akella, and P. Steenkiste. Bootstrapping Evolvability for Inter-Domain Routing. In *Proc. HotNets*, Nov. 2015.

[40] B. Schlinker, K. Zarifis, I. Cunha, N. Feamster, and E. Katz-Bassett. Peering: An as for us. In *Proc. HotNets*, 2014.

[41] L. Subramanian, M. Caesar, C. T. Ee, M. Handley, M. Mao, S. Shenker, and I. Stoica. HLP: a next generation inter-domain routing protocol. In *Proc. SIGCOMM*, Aug. 2005.

[42] D. L. Tennenhouse and D. J. Wetherall. Towards an Active Network Architecture. *SIGCOMM Computer Communication Review*, 26(2):5–17, Apr. 1996.

[43] K. Varadhan, R. Govindan, and D. Estrin. Persistent route oscillations in inter-domain routing. *Computer Networks*, 32(1):1–16, Jan. 2000.

[44] A. Venkataramani, J. F. Kurose, D. Raychaudhuri, K. Nagaraja, M. Mao, and S. Banerjee. MobilityFirst: a mobility-centric and trustworthy internet architecture. *SIGCOMM Computer Communication Review*, 44(3), July 2014.

[45] T. Wolf, J. Griffioen, K. L. Calvert, R. Dutta, G. N. Rouskas, I. Baldin, and A. Nagurney. ChoiceNet: toward an economy plane for the internet. *SIGCOMM Computer Communication Review*, 44(3), July 2014.

[46] L. Xiao, J. Wang, K.-S. Lui, and K. Nahrstedt. Advertising interdomain QoS routing information. *IEEE Journal on Selected Areas in Communications*, 22(10):1949–1964, 2004.

[47] W. Xu and J. Rexford. MIRO: multi-path interdomain routing. In *Proc. SIGCOMM*, Aug. 2006.

[48] X. Yang, D. Clark, and A. W. Berger. NIRA: A New Inter-Domain Routing Architecture. *IEEE/ACM Transactions on Networking*, 15(4):775–788, 2007.

[49] L. Zhang, A. Afanasyev, J. Burke, V. Jacobson, k. claffy, P. Crowley, C. Papadopoulos, L. Wang, and B. Zhang. Named data networking. *SIGCOMM Computer Communication Review*, 44(3), July 2014.

[50] X. Zhang, H.-C. Hsiao, G. Hasker, H. Chan, A. Perrig, and D. G. Andersen. SCION: Scalability, Control, and Isolation on Next-Generation Networks. *IEEE Symposium on Security and Privacy (SP)*, pages 212–227, 2011.

[51] M. Zhao, S. W. Smith, and D. M. Nicol. Aggregated Path Authentication for Efficient BGP Security. In *Proc. CCS*, 2005.