This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

I²BGP: A Privacy-Preserving Intra-AS State-Assisted Inter-AS Routing Scheme

Peizhuang Cong, Yuchao Zhang, *Member, IEEE*, Jun Wang, *Student Member, IEEE*, Wendong Wang, *Member, IEEE*, Tong Yang, *Member, IEEE*, Dan Li, *Fellow, IEEE*, and Ke Xu, *Fellow, IEEE*

***Abstract*—BGP is the most widely employed inter-AS routing protocol, connecting millions of ASes worldwide. While it is possible to select the egress for outgoing flows based on administrators' configurations, such schemes are localized due to the privacy of the intra-AS network state. The *AS_Path* field of BGP records all crossed ASes, which can be used to prevent routing loops and select paths, *i.e.*, selecting the minimum AS-hop path among available paths. Although this scheme is simple, effective, and offers a certain degree of global perspective, selecting paths at AS granularity ignores the transmission performance within each intra-AS, which may result in selecting non-optimal routing paths. To enable the use of private intra-AS data for inter-AS routing, we proposed a privacy-preserving intra-AS state-assisted inter-AS routing scheme, which can select optimal inter-AS paths without disclosing specific intra-AS state data. Specifically, we added an additional BGP header field to carry path performance features and designed a three-step data masking mechanism to protect intra-AS state data, enabling the selection of inter-AS paths with intra-AS state awareness. I²BGP has been deployed in the Greater Bay Area Future Network and a large-scale network simulator based on real network topologies. The results show that I²BGP outperforms BGP in terms of specified forwarding hops, delay, and bandwidth metrics.**

***Index Terms*—Inter-AS routing, intra-AS state awareness, private number comparison.**

I. Introduction

SINCE the Internet has connected the world, the across Autonomous Systems (AS) traffic transmission has been deeply studied in the network architecture field [1]. As the variety of network services flourishes, inter-AS traffic transmission becomes especially common and essential [2], [3]. For example, the massive amount of data used to train a Large Language Model (LLM) may originate from different geographic regions or different data centers, requiring large-scale data transmission across ASes, or the LLM may be deployed in specific data centers, making it necessary to efficiently transmit requests from users around the world to the model compute nodes across ASes.

The Border Gateway Protocol (BGP) is the most widely deployed inter-AS routing protocol, linking tens of thousands of ASes to form the World Wide Internet. Related works such as refining scheduling granularity, accelerating route convergence, and anomalous behavior detection have been widely studied [4], [5]. BGP allows routing for inter-AS traffic based on administrators' manual configuration, classic hot potato policies, or using *MULTI_EXIT_DISC* (MED) values notified in BGP messages by neighbor AS, however, which are localized, *i.e.*, only local AS state is taken into account when routing [6]. In addition, the *AS_Path* field of the BGP packet header is designed to record the list of crossed ASes in the route diffusion, which can be used not only to prevent routing loops but also as a reference for routing, *i.e.*, the route with the minimum AS hops can be selected as the inter-AS forwarding path [6]. Although such *AS_Path*-based scheme is simple, effective and present a certain degree of global perspective, selecting forwarding path at the granularity of AS is treating all ASes as undifferentiated black boxes, which ignores the network performance differences among all ASes, resulting in missing the optimal path with most suitable transmission performance that does not consist of minimum AS hops.

As shown in Figure 1, if a BGP *AS_Path*-based routing policy is employed, the border router in client c AS would select Path A for traffic to server s, as it involves two AS hops (with three ASes, s, b, c), compared to four AS hops (with five ASes, s, a_1, a_2, a_3, s) of Path B. Suppose that the transmission costs within each AS, such as delay or forwarding hops, are as indicated by the values, the path with more AS hops offers better overall transmission performance actually. There are some related research studies aimed at enhancing inter-AS routing selection [7], [8], [9], such as the Routing-as-a-Service (RSA) [10], the Software-defined Internetworking (SDI) [11], Multi-Party Computation (MPC) [12], and the Centralized Interdomain Routing (CIRCA) [13]. The key idea of these

Received 10 December 2024; revised 13 May 2025 and 22 July 2025; accepted 17 August 2025; approved by IEEE TRANSACTIONS ON NETWORKING Editor K. Lee. This work was supported in part by the National Natural Science Foundation of China under Grant 62172054, Grant 62372009, and Grant 62502014; in part by the National Natural Science Foundation of China for Distinguished Young Scholars under Grant 62425201; in part by Beijing Nova Program under Grant 2023140; and in part by the National Key Research and Development Program of China under Grant 2024YFB2906900 and Grant 2024YFB2906602. *(Corresponding authors: Yuchao Zhang; Tong Yang.)*

Peizhuang Cong is with Beijing University of Posts and Telecommunications, Beijing 100876, China, and also with the State Key Laboratory of Multimedia Information Processing, School of Computer Science, Peking University, Beijing 100871, China (e-mail: congpeizhuang@pku.edu.cn).

Yuchao Zhang, Jun Wang, and Wendong Wang are with Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: yczhang@bupt.edu.cn).

Tong Yang is with the State Key Laboratory of Multimedia Information Processing, School of Computer Science, Peking University, Beijing 100871, China (e-mail: yangtong@pku.edu.cn).

Dan Li and Ke Xu are with Tsinghua University, Beijing 100190, China.

Digital Object Identifier 10.1109/TON.2025.3606860

2998-4157 © 2025 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

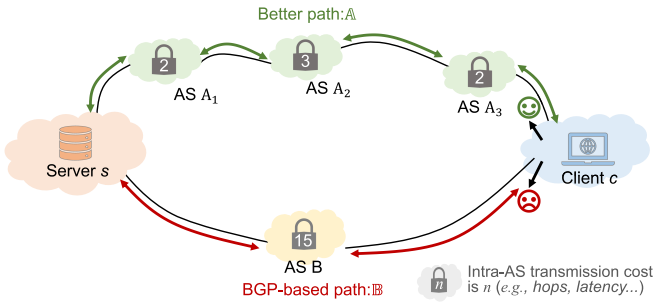


Fig. 1. Example of BGP-based inter-AS routing: the value marked on each AS represents the internal transmission cost, such as hops, latency, *etc*, which remains undisclosed to other ASes due to privacy concerns. However, based on BGP, 𝔹 will be selected due to it crossing fewer ASes.

studies is to employ specialized network controllers or trusted centralized computing systems to schedule traffic centrally. However, such centralized approaches presuppose that each AS will share or report its internal transmission performance information, which may be regarded as confidential by many AS organizers. Moreover, the centralized scheduling way faces scalability, making it challenging to implement across widely distributed internet systems.

Firstly, BGP was initially designed for guaranteed reachability in the network, *i.e.*, it fails to support carrying the performance information of paths in the route diffusion; secondly, the intra-AS network state (links from ingress to egress) is difficult to share directly to the whole network given the security considerations. In this context, we proposed a privacy-preserving intra-AS state-aware inter-AS routing scheme, I^{2}BGP, which can ensure the privacy of intra-AS network state while leveraging it to enhance inter-AS forwarding path selection. It addresses the limitation that link state information cannot be carried and utilized for inter-AS routing of current BGP-based approaches.

Specifically, intra-AS routing rules are independent and principally unaffected by the BGP protocol; it is both sufficient and necessary to only obtain the transmission performance from ingress to egress of each AS for inter-AS routing (herein referred to as intra-AS data, including forwarding hops, latency, bandwidth, *etc*., with hops used as an example for the introduction of I^{2}BGP). Therefore, I^{2}BGP abstracts the entire network into a border router-only topology (where routers within the same AS are interconnected by virtual links, and border routers between different ASes maintain their initial connections), and defines a new field, *Attr*, for the BGP packet header. This field records the quantitative performance values of intra-AS (from ingress to egress) and of the link between two ASes during route propagation, which are maintained and updated by the ASes crossed (similar to the update of *AS_Path* of BGP packet header) (§IV-B1). Additionally, the leakage of intra-AS data, such as hops, could potentially expose the intra-AS topology and subsequently pose network security threats. To protect this intra-AS data from being detected by other ASes, I^{2}BGP proposes two strategies: random number obfuscation (§IV-B3) and private number comparison (§IV-B4). These strategies facilitate the irreversible translation of the *Attr* coordinate system and the pre-calculation of optimal sub-paths, respectively, thereby ensuring that the data within each AS remains undetected. Alternatively, we designed a fast rerouting scheme for I^{2}BGP to enhance the disaster tolerance to cope with link failures.

The prototype of I^{2}BGP has been deployed in the Greater Bay Area Future Network (spanning five ASes across three cities) and the real network topology-based large-scale simulation experiment environment. Experimental results from thousands of flows indicate that I^{2}BGP reduces the average end-to-end forwarding hops by approximately 45% and decreases the Flow Completion Time (FCT) by about 60% compared to existing BGP. Further performance improvements are detailed in §V.

- We revealed that traditional BGP-based inter-AS routing may choose non-optimal paths due to ignoring intra-AS state.
- We designed I^{2}BGP, a privacy-preserving intra-AS state-aware inter-AS routing scheme, and further proposed the corresponding fast reroute scheme to enhance the disaster-tolerance capability of I^{2}BGP.
- We deployed I^{2}BGP in a real cross-city interconnection network and a real network topology based large-scale simulation network environment, experiment results of thousands of flows demonstrated the advantages and improvement of I^{2}BGP over the classic BGP.

In the following of this paper we: review background and motivation (§III), specify the design principle (§IV-A) and detail the methodology of I^{2}BGP (§IV), present experimental results (§V), survey related work (§II), and conclude the paper (§VI).

II. Related Work

A. Enhancements for BGP

BGP has been in operation for decades worldwide, and both industry and academia have expressed great interest in it and proposed enhancements designed to address it in terms of reconvergence, security, configuration, measurement, *etc*. [14], [15], [16], [17], [18]. BGP may undergo re-convergence due to link state updates, which can significantly impact data transmission efficiency. To improve the efficiency of re-convergence, Mattia et al. propose a method to optimize the minimum route advertisement time based on network topology [19]. Additionally, the changes in routing table entries during the re-convergence process can lead to transmission failures. To address this, Juan et al. propose sorting the corresponding entries based on traffic volume, ensuring that entries for large flows are updated first, thus maintaining overall network transmission performance [4].

In view of the success of blockchain technology in information security, many studies use it to augment BGP security [20], [21], [22]. To cope with route hijacking, He et al. stipulate that each AS maintains a globally consistent and tamper-proof ROA repository for verifying the legality of routing sources [23]. There are also some works combine the deep learning with BGP for anomalous detecting [24], [25].

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

B. Inter-as Routing Schemes

Existing inter-AS routing schemes can be mainly grouped into two types based on the employed approach [26], [27], [28], [29].

Third party-assisted schemes [30], [31], [32], [33]: Qiao et al. introduce a software-defined internetworking model, which extends SDN principles to inter-AS routing by enabling programmable route control while preserving network autonomy and develop an efficient Bayesian optimization-based method that leverages inter-AS routing algebra properties to efficiently search for policy-compliant end-to-end routes [11]. Pouryousef et al. propose a logically centralized architecture and system for inter-AS routing, which leverages trusted cloud servers to offload BGP-style route computation while ensuring the confidentiality of proprietary information; these servers, equipped with high-computing power and high-bandwidth, low-latency networks, enable efficient routing computation and policy delivery [13]. As a powerful technique, machine learning, especially reinforcement learning, is widely used in networked systems [34], [35]. Zhao et al. enabled a scalable data-driven inter-AS routing by utilizing multi-agent reinforcement learning and hierarchical network structure [33]. decided However, such solutions are preconditioned on third-party assurances that sensitive data will not be snooped and leaked. And as the scale of the network increases, the efficiency of such centralized approaches leaves much to be discussed and verified.

Tunnel-based overlay schemes [36], [37], [38], [39], [40]: The basic idea behind such approaches is to enable a stub AS to interact with a remote AS to select routes that are different from BGP routes and then establish an AS-to-AS tunnel to adopt the selected routes. Consequently, these systems incur extra overhead for tunnel processing on each data packet. However, since the tunneling information will not be disclosed to other ASes, it may result in undetected violations of data traffic regulations, which makes it difficult to be accepted by network organizations.

Additionally, Google and Meta also develop peering systems for their network that enable edge routers to perform flexible route selection by overriding BGP [41], [42]. However, such methods can only specify the next AS hop for traffic flows.

III. Preliminary and Motivation

A. Preliminaries

1) Border Gateway Protocol: As the de facto most widely used inter-AS routing protocol globally, the Border Gateway Protocol (BGP) connects tens of thousands of ASes. According to BGP policy, each AS is required to appoint at least one router (commonly referred to as a border gateway or border router) to facilitate route exchanges with neighboring ASes and forward incoming and outgoing traffic [6]. BGP operation relies on four kinds of messages: *OPEN*, *UPDATE*, *NOTIFICATION*, and *KEEP ALIVE*. 1) The *OPEN* message initiates the BGP session, exchanges basic parameters, and establishes a peer relationship; 2) The *UPDATE* message transmits new routes or withdraws existing ones, including key attributes for path selection; 3) The *NOTIFICATION* message reports errors or closes sessions, used for handling anomalies and maintaining session stability; 4) The *KEEP ALIVE* message is sent periodically to keep the session active and prevent session timeouts due to lack of data exchange.

Forwarding Path	Des	AS_Path	Priority	State	Hops
[c, b, s] (Path 𝔹)	s	2	high	√	15-> 15
[c, a3, a2, a1,s] (Path 𝔸)	s	4	low	×	7-> 2 3 2
Path detail to s		Route entry			Ideal additional info

Fig. 2. Route entries of client c in Figure 1.

Specifically, the attributes in the *UPDATE* message related to route selection include *Local_Pref*, *MED*, and *AS_Path*: 1) *Local_Pref* is usually manually set by local network administrators to determine the preferred egress router when multiple exit routers are present. In the context of increasingly complex network conditions and the trends in network services, manual configuration poses risks of limited flexibility and configuration errors, as evidenced by the significant economic losses caused by an accidental configuration by an engineer at Meta in 2021 [43]. 2) The *MED* value, announced by neighboring AS, is used to select the exit for traffic to that neighbor, typically a higher *MED* value indicating a lower priority. Both of these attributes lack a global network perspective, showing a limitation in their localized limitation. 3) The *AS_Path* attribute helps track the ASes a route crossed during route diffusion and can be used to prevent routing loops or assist in routing path selection, prioritizing the path with minimum AS hops. However, the AS hops-based routing selection way may lead to a sub-optimal path due to the neglect of performance differences within individual ASes.

2) Homomorphic Encryption: Before discussing the motivation behind our work, it is essential to introduce the homomorphic encryption technology utilized in this study. Homomorphic encryption is a form of encryption that allows for direct computations on encrypted data, producing an encrypted result [44]. This result, when decrypted, is identical to the outcome obtained by performing the same operations on the original data. This technology ensures that computations can be carried out on confidential data without compromising its security, thereby enabling both the protection of sensitive information and the utility of the data in encrypted form. It can be formulated as $D(E(a) \odot E(b)) = a \oplus b$, where $E()$ is the encrypting computation, $D()$ is the decrypting computation, and the symbols $\odot$ and $\oplus$ respectively represent the operations performed in the plaintext and cyphertext domains.

If $\oplus$ signifies addition, then the encryption is classified as additive homomorphic; if $\oplus$ indicates multiplication, then the encryption is identified as multiplicative homomorphic. If an encryption function exhibits both additive and multiplicative homomorphic properties and is capable of supporting an unlimited number of these operations, it is referred to as fully homomorphic encryption.

B. Motivation

The two route entries for client c in Figure 1 to server d are shown in Figure 2. Without considering the manually

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

administrator configuration and notification of *MED* by neighboring ASes (both of which are hard to own a network-wide perspective), according to the *AS_Path*-based path selecting rule of BGP, the entry of PathB will be matched, and the entry of PathA will be discarded and will not continue to notified to the other ASes.

However, if the value in the figure denotes the forwarding hops or the transmission delay, the cumulative overhead of PathA (7) is better than that of PathB (15). Ideally, if the routing entries contain additional information with the cumulative overhead (or quantitative transmission performance evaluation, *etc*.) of the paths, it would enhance the routing selection policy. However, firstly, the BGP lacks fields related to recording the cumulative overhead of the path; and secondly, given network security reasons, ASes would be hesitant to share their internal state data.

Therefore, it is a feasible way that extend BGP fields to record the cumulative overhead of a path while designing a privacy-preserving strategy for the AS internal state that can still assist in routing selection.

To realize the above function, *i.e.*, encrypting the transmission overhead values while ensuring homomorphic additivity during route diffusion, homomorphic encryption algorithms are potentially employed. Briefly, embedding the encrypted transmission overhead of intra-AS to the specified BGP packet header field, which will be cumulatively computed based on the homomorphic encryption cyphertext computability feature during route diffusion. To prevent ASes from providing incorrect intra-AS states deliberately, the inter-AS reputation system can be established based on existing studies [45], [46]. Alternatively, trustworthy third-party organizations and blockchain technology can be employed to periodically verify the authenticity of uploaded information, and ASes that provide incorrect information will be disclosed and penalized accordingly. Therefore, in this paper, it can be assumed that each AS provides accurate information. The details of I^{2}BGP will be presented in the following §IV.

IV. Design Details of I^{2}BGP

In this section, the design principles of I^{2}BGP are first described, followed by the details of I^{2}BGP, including topology abstraction, random number confusion, path performance-embedded BGP notifications, and route diffusion constraint with homomorphic encryption-based private number comparison. Then, the multi-criteria extension and the fast rerouting scheme for I^{2}BGP are presented. Finally, the incremental deployability of I^{2}BGP is discussed.

A. Design Principle

Because inter-AS route selection must be achieved without disclosing intra-AS state, I^{2}BGP is designed to satisfy the following two key criteria.

- **Protection of state privacy**: The intra-AS state, specific value of intra-AS forwarding hops, delay, *etc*., should not be leaked for satisfying the requirement of state-awareness, *i.e.*, other ASes cannot detect the data embedded in the BGP header directly or indirectly.

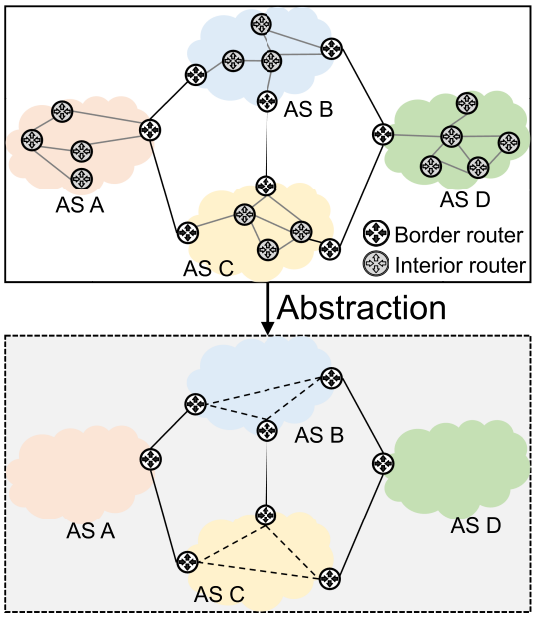


Fig. 3. AS topology abstraction.

- **Consistent state awareness**: In the premise of protecting data privacy, the relative transmission performance of different paths can still be obtained, *i.e.*, the field used to represent the performance of paths is comparable, and the comparison results can accurately reflect the relative merits of the actual performance of different paths.

B. I^{2}BGP Details

To make it possible for BGP to load and diffuse the additional path performance value, which is composed of a combination of the performance of all the links it crossed, in this work, it is advocated to introduce a new additional field *Attr* for the BGP packet header. Alternatively, it is also feasible to reuse existing fields for convenience (*e.g.*, *MED*), which is required to override the functionality of the initial field. For better presentation, I^{2}BGP can be divided into the following main steps and modules.

1) Topology Abstraction: According to the RFC document [6], the ingress border router is responsible for specifying the next hop address for inter-AS traffic, which usually involves only the border router, *i.e.*, the egress router of the AS. Inter-AS routing would not affect or constrain the intra-AS forwarding policy, *i.e.*, there are only ingress and egress hops for an AS from the BGP perspective. In this context, it is reasonable to simplify the intra-AS topology to a topology that only includes the border routers for inter-AS routing, and due to the intra-AS connectivity, the virtual links between each pair of border routers will be maintained, which are the inter-AS sub-paths within AS in I^{2}BGP. The abstraction example is shown in Figure 3.

Most intra-AS protocols or centralized SDN-based network architectures maintain router-to-router transmission performance under intra-AS network state (most common is forwarding hops), which can be directly leveraged by I^{2}BGP. It is also feasible to maintain such states, especially with a

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

complexity of $O(n^2)$, where n is the number of AS border routers. Such an abstraction not only simplifies the scale of the state that needs to be maintained but also fully meets the requirement of I^{2}BGP for intra-AS state information, while also protecting topology details within the AS to a certain degree.

2) Random Number Confusion: The extended BGP header contains the field *Attr* to record the path performance, which can be written directly with the intra-AS state data and spread to other ASes with the route diffusion, and each AS needs to maintain and update the received *Attr* before diffusing it to other ASes until the route converges. Similar to supplementing the current AS index to *AS_Path*, *Attr* needs to be integrated with that value received, the state of the link between the two ASes, and the state of the local AS. For example, for a cumulative type of attributes, *e.g.*, delay, hops, cumulative operations are required. Thus, such cumulative operations during route diffusion can inherently protect the data from being detected. For instance, even obtaining the cumulative sum of the hops of two ASes (y, $y = x_1 + x_2$) is incapable of deterministically identifying the specific hops of these two ASes (x_1, x_2).

However, there exists a case that fails to use this simple and efficient data protection, *i.e.*, the data of the source AS will be directly transmitted to the neighboring AS along with the routing diffusion. As shown in the Figure 3, although the data of ASB, ASC and ASD are not available to ASA, the data of ASD can be directly obtained by ASB and ASC, because the data of ASD is written directly into the *Attr*. By leveraging the inherent protection provided by the accumulation operation, I^{2}BGP stipulates that the data of the source AS should be written into the *Attr* only after an accumulation operation incorporating a random number is performed. This mechanism is referred to as random number confusion in this paper. It specifies the same random number for all route entries announced from the same router, which does not interfere with subsequent *Attr* operations and comparisons. Such an operation can be interpreted as performing a random transformation on the coordinate system in which the *Attr* values are located, without altering the relative relationships among different *Attr* values.

In addition, it is difficult for any AS to infer information from multiple routes received in this case. Specifically, it is assumed that each AS operates as an independent entity. Considering a scenario in which a source router a, located within an AS, announces routes through two border routers, b and c, resulting in route advertisements x_1 and x_2, respectively. And $x_1 = \delta + (a \to b)$ and $x_2 = \delta + (a \to c)$, where δ denotes the random number added for confusion, and $(a \to b)$, $(a \to c)$ represent the internal feature values (*e.g.*, hops) from a to b and c, respectively. Even if a neighboring AS obtains both x_1 and x_2 simultaneously, it remains unable to infer either δ or the actual internal path information $(a \to b)$ and $(a \to c)$. This property extends to cases involving more than two routes, where no meaningful internal information can be inferred.

Up to this point, it is able to protect the intra-AS data by the cumulative operations and random number confusion for route diffusion. In the following, a representative routing diffusing is presented to describe the specific update and maintenance process of *Attr* and path selection.

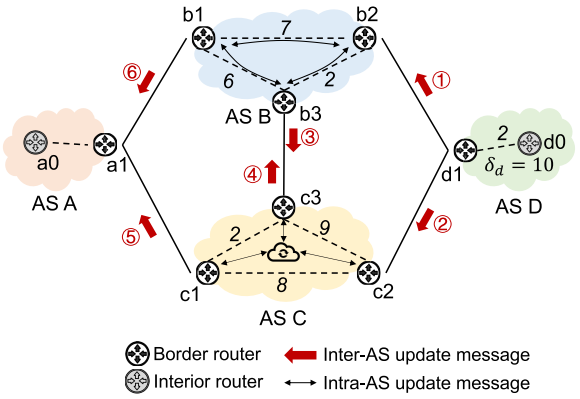


Fig. 4. Diffusion example.

TABLE I
INTRA-AS UPDATE MESSAGES

MSG	S→D	Update MSG details				
		Des	Next-hop	AS-hop	*Attr*	Des-AS
①	d1→b2	d'	d1	1	12	D
②	d1→c2	d'	d1	1	12	D
③	b3→c3	d'	b2	2	15	D
④	c3→b3	d'	c1	2	22	D
⑤	b1→a1	d'	b2	2	20	D
⑥	c1→a1	d'	c3	3	18	D

3) Routing Entry Diffusion: As mentioned above, a new field *Attr* is added to the BGP header of route updates to take the required data with it, and it is optional to redefine the existing field to make a convenient deployment after being authorized. In this way, the corresponding data can be spread with the diffusion of the routes. In addition, whether the AS is configured with traditional routing protocols or SDN-based architectures is feasible for I^{2}BGP, as long as the corresponding network state gets maintained.

As shown in the Figure 4 and the Table I and II, taking forwarding hops as an example, *i.e.*, the number of the link indicates the end-to-end forwarding hops. After ASD has updated the route with destination d', where the corresponding random number is 10, then d1 sends a route update message to the border routers b2 and c2 of the neighboring ASes, where the value of *Attr* is 12. The border router that receives the update from the other AS is responsible for accruing the performance of the link between the two ASes, and the hop is 1 for the link AS to AS link. Each border router decides whether to update the value of the old *Attr* in the local RIB by comparing the new *Attr*, with the principle of retaining routes with small *Attr* in this scenario. Updates within the AS are accomplished through the iBGP protocol, where b1 and b3 receive the update from b2, and c1 and c3 receive the update from c2. Additionally, b3 and c3 send updates to each other. In the case of b3, for example, it receives the *Attr* of 15 from b2 and the *Attr* of 22 from c3, and so it would not update the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE II
LOCAL RIB STATES

Router	Trigger MSG	Action	Local RIB details				
			Des	Next-hop	AS-hop	*Attr*	Des-AS
b2	①	*Attr* ①+1 = *Attr*: *not updata*	d'	d1	2	13	D
c2	②	*Attr* ②+1 = *Attr*: *not updata*	d'	d1	2	13	D
c3	③	*Attr* ③+1 < *Attr*: *update*	d'	c2→b3	2→3	22→16	D
b3	④	*Attr* ④+1 < *Attr*: *not updata*	d'	b2	2	15	D
c1	iBGP update MSG	Intra-AS *update*	d'	c2→c3	2→3	21→18	D
a1	⑤, ⑥	*Attr* ⑥+1 < *Attr* ⑤+1 < *Attr*: *update*	d'	b1→c1	3→4	21→19	D

local route based on an update of c3. Conversely, c3 receives *Attr* of 15 from b3, which is less than the local value of 22, and updates the local route *Attr* to $16 = 15 + 1$, *AS_Path* to 3, and the next hop is updated to b3. Similarly, b1 and c1 will send an update to a1 after completing the update within the AS, where the *Attr* of b1 and c1 are 20 and 18, respectively. The local *Attr* of a1 (21) is greater than the new *Attr* (18) received from c1, then the local *Attr* will be updated to $18 + 1$, the next hop will be c1, and the *AS_Path* will be 4.

Different types of link attributes may involve different cumulative calculations [47], but the overall updating process remains generally consistent with the example above.

4) Private Number Comparison: Due to the nature of cumulative calculations, the exchange of routing information between adjacent ASes may pose a risk of data leakage.

For example, in the Figure 4, b3 can infer the hops from c2 to c3 as 9 after receiving updates from b2 and c3. The specific process is as follows: c3 sends an *Attr* value of 22 to b3, while b3 receives an *Attr* of 15 for the same destination from b2 and can know that b2 received an *Attr* of 12. Combining the *AS_Path* and the AS-level topology, the value from c2 to c3 is calculated as 22-(12 + 1)=9 (since the hops from d1 to c2 equals the hops from d1 to b2). Mathematically, the problem can be formulated as follows: concerning the two equations $x + y = h$ and $x + y + z = l$, the value of z can be obtained by making a difference between the two equations, $z = l - h$, even where x and y are unknown.

From the perspective of topology, such data leakage risk can be occurred in three ASes connected as a triangle, such as the connection of AS B, AS C, and AS D in the Figure 4. This is due to the fact that one of the ASes may receive routes with the same destination from two other ASes, one of which has more overhead generated within this AS than the other, which is an operation that cannot be avoided by I^2BGP. And random number confusion fails to solve this issue, as adding an additional random number to the AS that may leak data will possibly make subsequent comparison results of the *Attr* inconsistent with the actual path performance.

As shown in Figure 4, ASB, ASC, and ASD form a typical triangular connection. To obtain the minimum forwarding hops path, the flows with destination d1 from down stream ASes to ASC will be firstly forwarded to ASB instead of being directly forwarded to ASD. Then, if d1 does not send the corresponding routing update to c1 it still does not change the final routing result, since the path of c1 to d1 is not a subpath

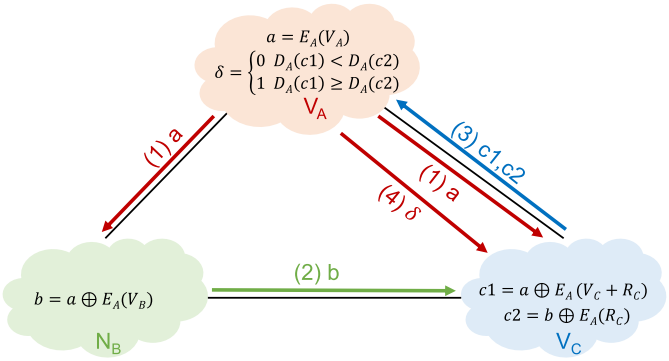


Fig. 5. Comparison illustration.

of the optimal path. Therefore, logically removing the non-optimal subpath in the triangle topology can avoid the data leakage mentioned above.

On this basis, homomorphic encryption can be utilized to facilitate the search for optimal subpaths in triangle connections, *i.e.*, comparing the performance of subpaths in terms of specified metrics while guaranteeing the internal state of the involved ASes will not be leaked, and removing non-optimal subpaths logically according to the comparison results, which can be completed by constraining the routing diffusion. The details and computational steps of the Paillier algorithm employed in this work are not key elements, and they can be referred to in [44]. The details of the above-described scheme will be presented in the following three processes.

- *Maintaining Triangle Connection*. It is feasible to exchange one's own neighbor AS list with adjacent ASes by using the *KEEP ALIVE* message of BGP. The neighbor AS list can be maintained by the *AS_Path* field of the received route update message. Based on the exchanged neighbor information, the triangle connection search can be executed. Since the topology between ASes is relatively stable, the search algorithm for triangle connection will not need to be executed frequently, and the execution can be triggered by the failure of the link between ASes after the initial search. Duplicate triangles can be removed from the returned list if the router-to-router routing path in the network is undirected.
- *Comparing Paths*. The communication and computation process of the comparison is shown in the illustrative example shown in Figure 5, where it is assumed that the intra-AS state values of ASA, ASB, and ASC are V_A, V_B, and V_C,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

respectively. These values can be interpreted as the forwarding hops or other metrics value from the ingress router to the egress/destination router. In the following, all encryption processes employ the public key of the ASA and ASA decrypts with a private key owned only by itself.

First, suppose ASA initiates the comparison process by sending the V_A encryption $E_A(V_A)$ to ASB and ASC. Upon receiving the message from ASA, ASB computes $E_A(V_A) \oplus E_A(V_B)$ and sends it to ASC, where $\oplus$ denotes homomorphic addition. AS C, after receiving the message from ASA, calculates $E_A(V_A) \oplus E_A(V_C + R_C)$, and upon receiving the message from ASB, calculates $E_A(V_A+V_B)\oplus E_A(R_C)$, where R_C is a random number generated by ASC. The results will be sent back to ASA. ASA decrypts the received messages as $D_A(E_A(V_A + V_C + R_C))$ and $D_A(E_A(V_A + V_B + R_C))$, and gets the result of the comparison of these two values, which will also be sent back to ASC. Finally, according to δ, ASC can distinguish the priority between the direct path from ASC to ASA and the bypassing path from ASC via ASB to ASA. Similarly, ASA also knows, which will be leveraged in the next diffusion constraint step. For the convenience of describing the above process, the link between the two ASes is not taken into account, and the AS that receives the update can be assigned to be in charge of calculating the corresponding value. In addition, it is necessary for ASA to send $E_A(V_A)$ to ASB and ASC, because the values of ASC or ASB from ingress to egress in ASA may not be the same, *i.e.*, the V_A that ASA sends to ASB and ASC can be different.

Given that both ASB and ASC lack the private key of ASA, they are not capable of decrypting the encrypted data. The scenarios, such as secret key leakage and malicious probing, are not considered here. And as ASC utilizes random number confusion, ASA also cannot obtain data about ASB and ASC from the two values received from ASC, even ASA owns the private key. Hence, the entire comparison process is secured.

In addition, for operations involving homomorphic encryption, a set of prime numbers can be pre-computed by each border router for timely use. More efficiently, given the inherently public nature of information within an AS, a large pre-computed shared prime pool can be established and periodically updated, enabling all border routers within the AS to obtain prime from it. Such shared prime pool offers two advantages: First, random selection of primes with replacement from a sufficiently large set can mitigate the prime exhaustion issue while preserving the security of the encryption; Second, newly added routers can directly access and utilize the shared primes, eliminating the necessary for independent computations.

- *Constraining Diffusion*. As explained above, in triangle topology, the optimal sub-paths are typically directed or detoured connections. In this context, we propose a route diffusion constraint scheme to remove non-optimal sub-paths logically. When ASA receives an update from another AS, it can specify how this update should be diffused in subsequent ASB and ASC.

In the optimal sub-path is directed connection scenario, as shown in Figure 6(a), ASA directly diffuse the update with an identifier $F = 1$ to inform the downstream ASes, *i.e.*, ASB and ASC, that this update should not be further diffused to another AS within the triangle connection. For the detour situations, as depicted in Figure 6(b), ASA sends the update to ASB with the flag $F = 0$, indicating that ASB needs to forward this update to ASC. ASC, upon receiving the update from ASB, and due to the loop-free feature of the *AS_Path* attribute, will not send this update back to ASA. Alternatively, such constrain can be achieved using the BGP community attribute if it is supported between ASes [48]. Similarly, such an operation can also be applied when the neighboring ASes belong to the same administrative authority to prevent collusion between them to obtain information.

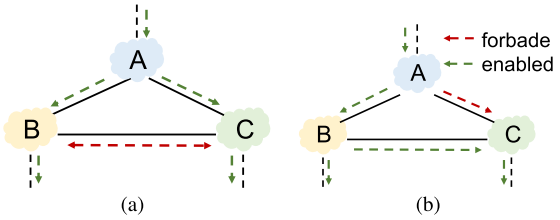


Fig. 6. Two types of diffusion constraint.

By doing so, it neither affects the final selection of the optimal path nor brings the risk of data leakage due to triangle connections.

C. Completeness Analysis of Privacy

To enhance inter-AS routing with intra-AS information, it is crucial to ensure the privacy of each AS's internal data. Specifically, no AS should be able to infer the internal state (*e.g.*, hops) of other ASes from the route entries it receives. During a route convergence, each AS *S* accumulates its maintained local state $cost_S$ on the *Attr* announced by upstream AS, and diffuses it to downstream ASes. As a result, an AS receiving route *i* can obtain a summation value of the equation, $Attr_i = \sum_{J \in AS_Path_i} cost_J$, where each unknown $cost_J$ corresponds to a value maintained by an AS on AS_Path_i. Privacy protection in this context means that an AS must not be able to derive the value of any individual $cost_J$ by collecting and analyzing such information derived from different routes in its local Routing Information Base (RIB).

To formalize this, we frame the privacy guarantee as a system of equations solving problem. The privacy proof of I^{2}BGP can be expressed as demonstrating that the system of equations constructed from all received routes admits no unique solution for the individual $cost_J$ values. In the following, we first mathematically model the problem and then prove the privacy completeness of the I^{2}BGP.

1) Formulation: We define the cumulative cost for AS *S* of being forwarded by its border router *i* to AS *D* as:

$$y_i = COST_S^{i \to D}. \tag{1}$$

Then, based on different *n* border routers, *S* can obtain the set of equations $\mathbb{C}$:

$$y_i = COST_S^{i \to D}, i = 1, 2, \cdots, n. \tag{2}$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

where y_i is the value of *Attr* of each related route *i*. Each $COST_S^{i\mapsto D}$ can be decomposed into the cumulative sum form of *cost* of ASes contained in *AS_Path* of route *i*. So the set $\mathbb{C}$ can be further converted as:

$$y_i = cost^D_{i^{AS_0}} + cost^D_{i^{AS_1}} + \ldots + cost^d_{i^{AS_D}}, i = 1, 2, \cdots, n, \tag{3}$$

where $cost^D_{i^{AS_j}}$ represents the cost of the *j-th* AS of the path forwarded by the border router *i* to AS *D*, $cost^D_i$ is the cost from border router of AS *D* to the destination router *d*. Intra-AS data leakage occurs when any *cost* can be inferred from $\mathbb{C}$.

Suppose there exists a curious AS *S*, which attempts to infer internal information about other ASes based on $\mathbb{C}$ derived from received routing information. The following mathematical derivation formally demonstrates that *S* cannot succeed in this endeavor.

2) Mathematical Analysis: Due to each equation in $\mathbb{C}$ is linearly independent, for any subset $\mathbb{C}' \subseteq \mathbb{C}$, the elements can be inferred if the number of unknown elements equals the number of equations after elimination transformation of $\mathbb{C}'$. According to the Rouche-Capelli theorem [49], which states that a linear system has a unique solution if and only if (**iff**) the rank of the coefficient matrix equals the rank of the augmented matrix and is equal to the number of unknowns.

For a special case, $\mathbb{C}$ only has an equation $y_i = cost^d_{i^{AS_D}}$, *i.e.*, the number of unknowns and the number of equations are both 1. This corresponds to the situation where the curious observer AS *S* is directly connected to the destination AS *D* in the network topology, allowing *S* to directly obtain the value of $cost^D$. However, under the *Random Number Confusion* scheme, an additional random element δ^d is implicitly introduced, transforming the observable value into $cost^D + \delta^d$. This effectively increases the number of unknowns and prevents *S* from accurately inferring the true value of $cost^D$.

For the general case, without loss of generality, assume $\mathbb{C}'$ initially contains an equation, $y_1 = cost_{AS_0} + cost_{AS_1} + \ldots + cost_{AS_n} + cost^d_{AS_D}$, representing the route information from the curious observer AS *S* to a destination router *d* of AS *D*. To get any element of this equation, additional equations must be introduced into $\mathbb{C}'$. If an introduced equation corresponds to a route whose destination router belongs to AS $D'(\neq D)$, this will introduce at least one new unknown element to $\mathbb{C}'$. For instance, simplifying the path cost components, $\mathbb{C}'$ is as follows:

$$\begin{cases} y_1 = cost_{AS_0} + X_D \\ y_2 = cost_{AS_0} + X_{D'} \end{cases}$$

where X_D and $X_{D'}$ are distinct unknown elements associated with AS *D* and D', respectively. However, any element of $\mathbb{C}'$ still lacks a unique solution. Alternatively, if the introduced equation corresponds to a route that with the same AS *D* but a different router $d'(\neq d)$. Similarly, even if the paths of these two routes are identical, the Random Number Confusion scheme ensures a lack of a unique solution for any element. For example, $X^d_D = x_d + r_d$ and $X^{d'}_D = x_{d'} + r_{d'}$, where *x* is the cost from router to border router of AS *D*, *r* is a specified random number, which is independent for different announcements. Therefore, any elements of $\mathbb{C}'$ remain unsolvable for unique values.

Consequently, to maximize the potential for solving unknown elements, it is most constructive to construct $\mathbb{C}'$ with equations that relate routes that share the same destination router. In inter-AS routing scenarios, these route paths adhere to specific constraints, as detailed below.

Theorem 1: Paths Convergence: For AS *S*, suppose there are two distinct routes with paths P_1 and P_2 to the destination *d* of AS *D*. If P_1 and P_2 have a common intermediate AS *I*, then their sub-paths from *I* to *d* are identical.

Proof: Assume that the sub-paths from the common AS *I* to the destination *d* along P_1 and P_2 differ, which implies that AS *I* maintains two distinct inter-AS paths from itself to *d*. This contradicts the standard BGP path selection, which mandates that an AS deterministically maintains the most appropriate path for each destination in the local RIB. Therefore, the assumption is false, and the sub-paths must be identical. □

On the basis of Theorem 1, we represent the cumulative values of ASes before the shared common AS as $COST^{I-d} = \sum_{i=I}^{d} cost_i$. Then, the equations corresponding to the path with intersection can be converted as:

$$y_i = cost^D_{i^{AS_0}} + \cdots + cost^D_{i^{AS_{n_i}}} + COST^D, \ i \in \mathbb{C}'. \tag{4}$$

For equation (4), **iff** $\exists a \in \mathbb{C}' \rightarrow y_a = COST^D$, and $\exists b \in \mathbb{C}' \rightarrow y_b = cost^D_{b^{AS_0}} + COST^D$, it can uniquely infer the value of an unknown quantity. That is, $cost^D_{b^{AS_0}} = y_a - y_b$. This situation corresponds to the aforementioned triangle connection, which can be prevented by the proposed *Constraining Diffusion*. For all other situations, the number of elements is greater than the number of equations; thus, no element can be inferred.

In conclusion, I^{2}BGP eliminates the risk of inferring any link-level information from path-level information of both source and diffusion processes of routes, *i.e.*, I^{2}BGP ensures the link state of each AS remains inaccessible to other ASes while preserving the usability of path information for inter-AS routing.

D. Extension to Multi-Criteria

Current network traffic may no longer be constrained by a single criterion. For instance, some emerging AR/VR services consider both latency and bandwidth. Common criteria of traffic flows can be categorized into two categories. The first category refers to cumulative criteria, such as the hops mentioned above, as well as latency, packet loss rate, *etc*. For these criteria, the path state is influenced by all the links and can be ultimately expressed as the cumulative sum or multiplication of the states of all links. The second category consists of bottleneck criteria, where the state of the path is determined by the extrema (*e.g.*, minimum or maximum) of the attributes across all links on the path. Bandwidth is a typical example of such a criterion; in this case, the path bandwidth is determined by the minimum bandwidth among all the links, *i.e.*, $BW(Path) = min(BW(link_1), BW(link_2), \ldots | link_i \in Path)$.

Before introducing multi-criteria scenarios, operations of bottleneck criteria are briefly presented, which are highly similar to the aforementioned operations of cumulative criteria. First, computations of bottleneck criteria also do not allow

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

for exact inference of input values from the final result, making it inherently practicable for privacy-preserving. For instance, given $x = min(A_1, A_2)$, it is infeasible to determine the exact values of A_1 or A_2 solely from x. This intrinsic indeterminacy ensures a level of confidentiality comparable to that of additive attributes during routing diffusion. Then, in a triangular topology, the processing is even simpler than that of cumulative criteria. By approximating pairwise comparisons and exchanging comparison results, it is possible to derive the overall order of values, which can then be constrained based on the same routing diffusion limitations. Finally, regarding the randomization strategy, a value can be randomly selected between the minimum bandwidth required by the service and the actual available bandwidth. Since this randomized value both satisfies service requirements and introduces uncertainty, neighboring ASes cannot accurately infer the true bandwidth. Moreover, this strategy allows the comparison and selection process to favor paths with higher bandwidth while still meeting the necessary service constraints, thereby improving the overall routing performance. Based on the above operations, the bottleneck criteria can also be computed in a privacy-preserving manner.

I^{2}BGP exhibits scalability from single-criterion to multi-criteria scenarios. Consistent with the single-criterion I^{2}BGP, the multi-criteria I^{2}BGP also consists of three main parts. In topology abstraction, inter-AS links are preserved while intra-AS states are masked. Considering the differences in criteria when calculating the optimal path, different strategies are required for random number confusion. For example, for criteria such as latency and hops where smaller values are preferred, to ensure that the final state value calculation will not be affected by the random numbers added in the *Attr* field of route entry in the source AS, it is possible to randomly select a natural number. However, for criteria like bandwidth, where a larger value is preferable, the minimum value along the entire path is crucial. In this case, a random value greater than the expected value that meets the service needs can be set, ensuring that the specified random number will not result in getting a path that does not meet the requirements. The goal of homomorphic encryption-based private number comparison is to distinguish the magnitudes of two values. Therefore, it is necessary to adjust whether to select larger or smaller values based on the specific criterion. Then, the diffusion constraint strategy can be combined to specify whether to choose the path with larger or smaller *Attr* values.

To be able to carry multiple link attribute states, *Attr* can be decomposed into multiple sub-fields, each responsible for carrying different attribute states. It is possible to appoint the order and bit length of all states to improve the utilization efficiency of the *Attr* field. Additionally, special characters, such as setting subfields to all 0s or all 1s, can indicate forbade diffusion or that the state corresponding to the subfield does not need to be reconverged. The updates to the link state in the RIB are calculated based on all subfields of *Attr* independently.

E. Fast Reroute Scheme

In a triangular topology, I^{2}BGP constrains the route diffusion of links on non-optimal paths, which does not affect the network-wide convergence of routes on the optimal path. However, when a link belonging to the optimal path in the triangle topology fails, and since the other path is ignored in the context of I^{2}BGP, forwarding it back to the ingress will result in routing loops, the traffic flows of the corresponding route need to be forwarded to other AS egresses. If there is no other egress, according to the BGP routing policy, the forwarding will be unreachable, resulting in packet loss. However, in the triangular topology, the three nodes periodically probe the connection states, allowing non-optimal forwarding paths to naturally function as alternative routes. When the link on the optimal path fails, it directly forwards the involved flows to the alternative path by default. At the same time, the new link state is notified by a route update message for network re-convergence.

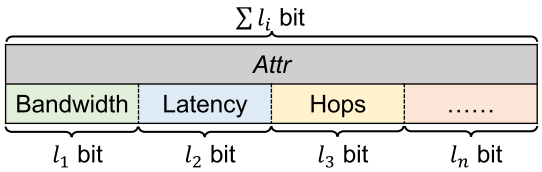


Fig. 7. Multi-criteria *Attr* example.

Specifically, assuming that there are flows that need to be forwarded from ASA to ASC, fast reroute can be categorized into four scenarios as shown in the Figure 8: (a) If the link failure occurs on the initial optimal path between ASA and ASC, ASA forwards the corresponding flows with a failure identifier to ASB, and ASB then forwards these flows to ASC; (b) If the link failure occurs on the initial optimal path between ASA and ASB, ASA can directly forward the corresponding flows to ASC; (c) If the link failure occurs on the initial optimal path between ASB and ASC, and ASA is already informed about the failure, it will directly forward the corresponding flows to ASC; (d) If ASA does not receive the fault identifier, it forwards the traffic to ASB based on I^{2}BGP. If ASB has other egresses, it can forward the flows to other ASes via the available egress, but if not, it forwards the flows back to ASA with the failure identifier, and then ASA will forward the corresponding flows directly to ASC based on the received failure identifier.

Since the original path is the optimal path after convergence, the subsequent path is likely to remain the optimal path to the destination after detouring to the next AS of the original path following a fast reroute. The specific fast rerouting performance will be shown in the evaluation section. Note that the above rerouting is specifically for link failures in the triangular topology of I^{2}BGP, which is orthogonal to other related pre-existing inter-AS fast rerouting or fast re-convergence schemes.

F. Incremental and Fast Deployment Discussion

Since BGP has been operating worldwide for decades, adding new features to BGP on such a large scale is difficult to accomplish at once [50]. For this reason, we discuss the hybrid and rapid deployment of I^{2}BGP. I^{2}BGP aims to make inter-AS routing have a network-wide view by spreading the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

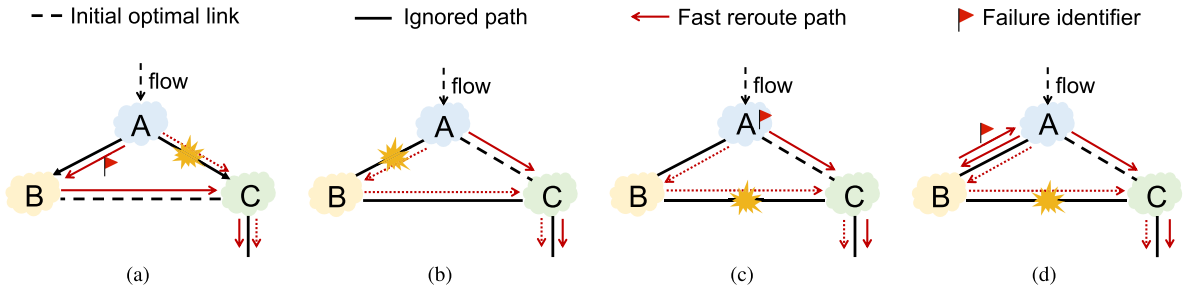


Fig. 8. Fast reroute for link failures in triangular topology.

intra-AS state that is processed by some protection mechanism. When I^{2}BGP and BGP are mixed deployed on the network, the required data can be transmitted via the upper-layer protocols across ASes that are not running I^{2}BGP and then update the forwarding routing table. For path selection, the *AS_Path* attribute can be integrated to jointly prioritize the paths, for example, a path with a higher percentage of ASes running I^{2}BGP has a higher probability of guaranteed performance, so the higher the priority should be; or excluding poorly performing paths based on I^{2}BGP; or specifying the average performance provided by I^{2}BGP as the performance of the ASes that are not running I^{2}BGP along the path, *etc*.

In addition, *MED* is designed to advertise the priority of each ingress to the AS where the incoming traffic comes, and the value of this attribute only works for the ingress routers of the advertised AS. Therefore, it is a feasible way to redefine the *MED* field for rapid and cost-effective deployment. In this case, the update and maintenance rules of the *MED* should be modified to incorporate the cumulative nature required by I^{2}BGP, thereby extending its scope of action to the entire network. This tactic is usually applicable to special inter-AS network scenarios, such as a federated network formed by several companies. Moreover, when all the involved ASes are under the same management organization, it is feasible to apply the route diffusion and update rules of I^{2}BGP only.

V. Evaluation

In this section, we comprehensively evaluate the performance of I^{2}BGP by constructing a prototype system that consists of five ASes across three cities in the Greater Bay Area Future Network and a real network-based large-scale experimental testbed on *NS-3* network simulator.

Explanations of the chosen comparison methods for I^{2}BGP are provided before exhibiting the results. Even disregarding the challenges associated with deploying centralized approaches in large-scale scenarios, their reliance on a third-party to obtain a global network view inherently conflicts with one of the core objectives of this work: preserving intra-AS state privacy. Furthermore, to avoid diverting attention toward debating the relative merits of centralized versus distributed routing paradigms, we focus our comparative experiments on BGP rather than centralized alternatives. This allows us to highlight our primary contribution: enhancing the de facto inter-AS BGP protocol with intra-AS state awareness to enable improved routing decisions.

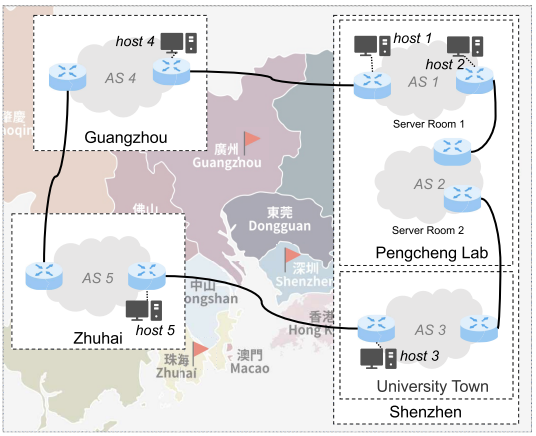
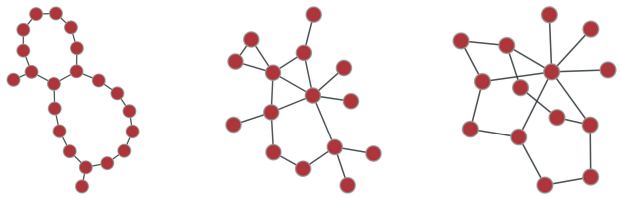
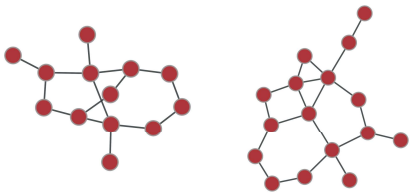


Fig. 9. Implementation topology in Greater Bay Area Future Network.



(a) ATMnet: [21, 22] (b) Claranet: [15, 18] (c) Compuserve: [14, 17]



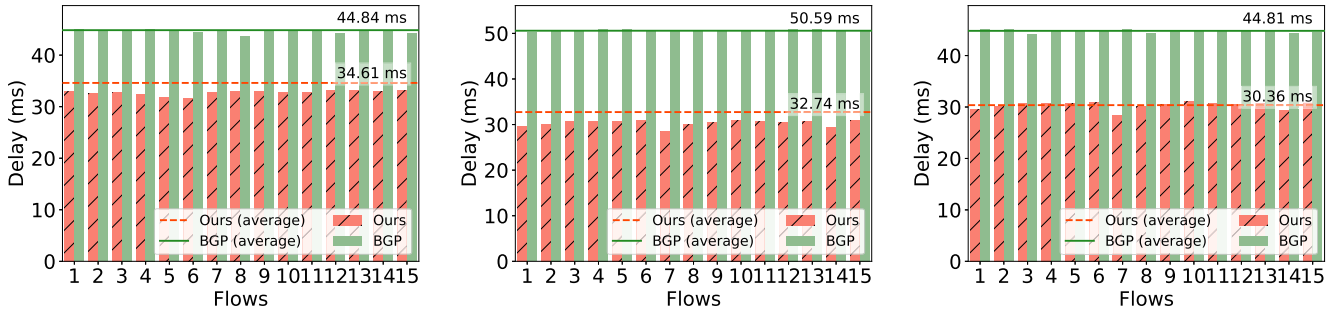
(d) NSFnet: [13, 15] (e) Peer1: [16, 20]

Fig. 10. Topologies sketch. (*Topo name*: [*Num of AS and AS-edge*]).

A. Prototype System Evaluation

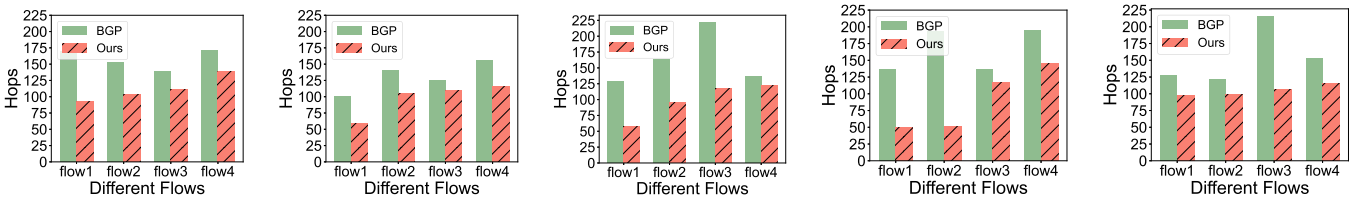
To validate the performance of I^{2}BGP, it was deployed in the Greater Bay Area Future Network, connecting five forwarding nodes across three cities. Each forwarding node is configured as an AS: Pengcheng Laboratory in Shenzhen (AS1 and AS2), University Town of Shenzhen (AS3), Guangzhou (AS4), and Zhuhai (AS5). The two ASes in Pengcheng Laboratory are located in different server rooms. Each AS is configured with two border routers, which are interconnected within the AS

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.



(a) Host 1 to host 5 (b) Host 2 to host 5 (c) Host 3 to host 4

Fig. 11. Delay improvements in the Greater Bay Area.



(a) ATMnet (b) Claranet (c) Compuserve (d) NSFnet (e) Peer1

Fig. 12. Hop performance under 5 topologies.

through multiple virtual machines. Host1 and host2 are set up in AS1, host3 is setup in AS3, and host4 and host5 are set up in AS4 and AS5, respectively. The specific network setup is shown in the Figure 9.

Three sets of flows were configured in the aforementioned network environment: host1 and host2 each send data to host5, and host3 sends data to host4. The packet transmission delay for the three traffic flows based on I^{2}BGP and BGP protocols was recorded, respectively. The transmission delay results of 15 randomly selected packets and the average transmission delay of all packets are shown in Figure 11. The experimental results demonstrate that compared to the existing BGP protocol, the deployment of I^{2}BGP can enhance network transmission performance.

B. Large-Scale Simulation Evaluation

Experiment Setting: The experiments are implemented on the *NS-3* [51] network simulator on the *Ubuntu* operating system. And five real network topologies, as shown in the Figure 10, slATMnet, *Claranet*, *Compuserve*, *NSFnet*, and *Peer1*, were selected from *Topology Zoo* as the experimental topologies [52], [53].

1) Forwarding Hops Improvement: The end-to-end flows between four pairs of routers are randomly generated in each of these five topologies. The source and destination routers for each flow are selected to be far apart within the topology, which ensures that there are multiple available paths between the flows and better shows the difference in routing policy performance between I^{2}BGP and BGP.

The end-to-end forwarding hops of all flows under traditional BGP and I^{2}BGP are recorded, as shown in Figure 12. The results indicate that I^{2}BGP outperforms traditional BGP where there are multiple divergent inter-AS paths.

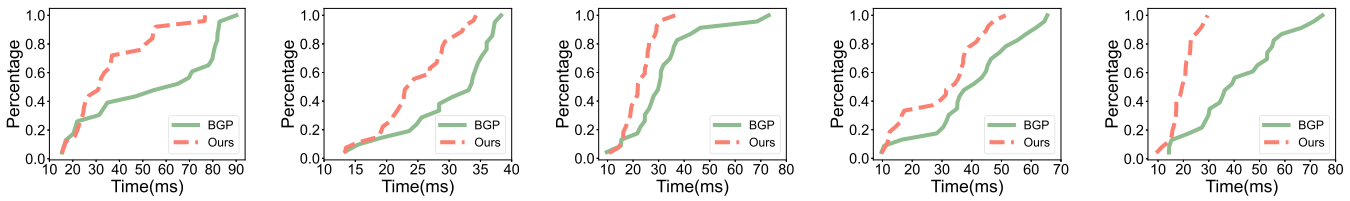
2) Delay Improvement: In the last set of experiments, although I^{2}BGP does not take link latency (*i.e.*, path transmission delay) as the value of *Attr*, the performance improvement in terms of delay can still be achieved by using the forwarding hops as the optimization goal, because a reduction in forwarding hops will reduce the total overhead of the processing at all the passed switches.

For the performance improvement of I^{2}BGP over BGP regarding delay in these five topologies, the Flow Completion Time (FCT) is measured for the same set of large-scale flows. As illustrated in Figure 13, the results exhibit that the FCT of I^{2}BGP is shorter than BGP in all five topologies.

3) Multi-Criteria Improvement: This set of experiments defines two types of services, delay-first flow and bandwidth-first flow, such as online gaming service and offline file transfer service, which prefer the path with the minimum delay and the path with the maximum bandwidth, respectively. The bandwidth of AS links is set to 10Mbps and the bandwidth of links within the AS is randomly set to 2, 4, 6, 8, and 10Mbps. The link latency is kept consistent with the previous settings. I^{2}BGP employs multi-criteria with link latency and bandwidth. Both types of flows are generated between any pair of router nodes in five topologies.

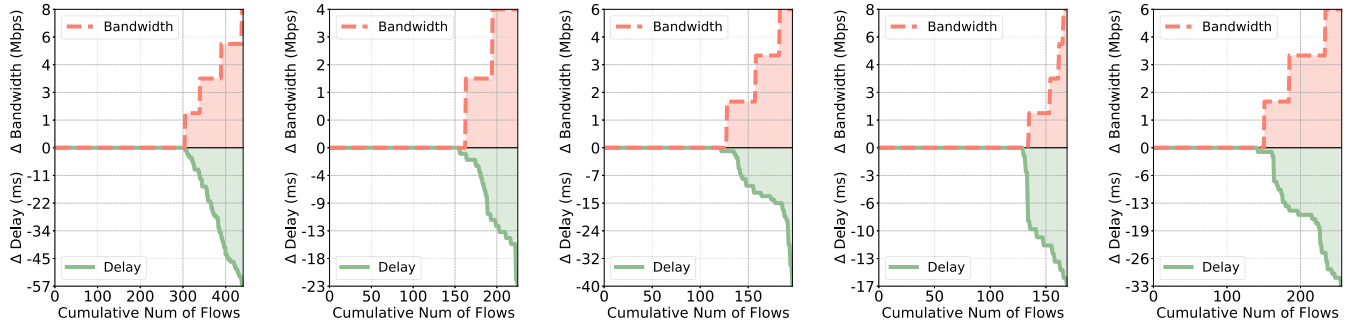
For delay-first flows, the difference in transmission delay between I^{2}BGP and BGP for each flow with the same source and destination is recorded: $Delay_{\text{I}^2BGP}(flow_i) - Delay_{BGP}(flow_i)$. Similarly, for bandwidth-first flows, the difference in transmission path bandwidth for the same flow is recorded: $BW_{\text{I}^2BGP}(flow_i) - BW_{BGP}(flow_i)$. The specific cumulative distributions are shown in Figure 14. It can be observed that I^{2}BGP can at least maintain comparable forwarding performance with BGP for some flows (y=0 part of the figure); for the rest of the flows, I^{2}BGP can flexibly select paths with greater bandwidth for some bandwidth-first flows

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.



(a) ATMnet (b) Claranet (c) Compuserve (d) NSFnet (e) Peer1

Fig. 13. Delay performance under 5 topologies.



(a) ATMnet (b) Claranet (c) Compuserve (d) NSFnet (e) Peer1

Fig. 14. Multi-improvements in both delay and bandwidth criteria.

based on the additional information from *Attr* (shaded area in the upper half-axis), and similarly can select paths with shorter transmission delay for delay-first flows (shaded area in the lower half-axis).

4) Performance in Larger Scale Network: To further validate the performance improvement of I^{2}BGP, we constructed two larger-scale network topology structures, containing 200 and 500 ASes, respectively. Specifically, each AS is randomly connected to 2 to 4 other ASes. Within each AS, in addition to the border routers, 5 to 20 intra-AS routers are randomly generated. The degree of these routers is constrained between 2 and 4, while ensuring the overall connectivity of the network. In these two network topologies, 100 flows are randomly generated, and the hops of the corresponding forwarding paths for these flows are measured. Experimental results demonstrate that, compared with BGP, I^{2}BGP reduces the average path hops by 19.77% and 23.04%, respectively.

5) Full-Set Flows Performance: To evaluate the proportion of end-to-end flows of the network that benefit from I^{2}BGP, we generated approximately 800 inter-AS flows between all pairs of routers in *NSFnet* under the same configuration. We then recorded the transmission delays of all flows under both I^{2}BGP and BGP, respectively.

Note that since I^{2}BGP is theoretically superior to BGP concerning the hops feature, the transmission delay is measured in the experiments to enable the comparison under full-set flow. The transmission delay differences ($Delay_{BGP} - Delay_{I^2BGP}$) are shown in the Figure 15, which indicates that I^{2}BGP performs better than or at least equal to BGP for almost all flows. Due to the network fluctuation uncertainty (*e.g.*, congestion and packet loss), it is reasonable to expect performance gains for all flows. As shown in Figure 15, only 0.5% of the flows show that I^{2}BGP is slightly inferior to BGP. To sum up, I^{2}BGP is superior to BGP on the network wide scope.

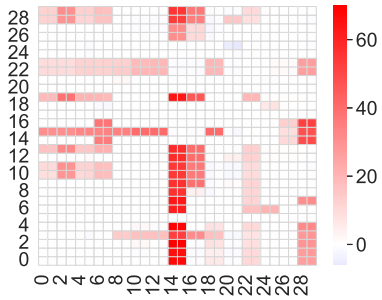
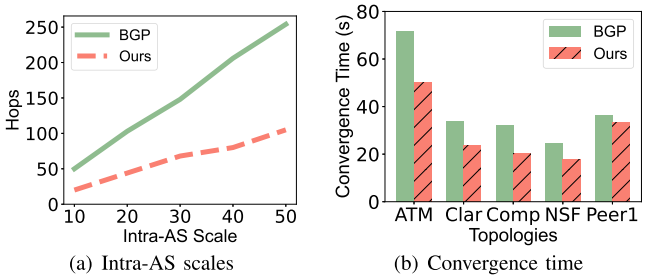


Fig. 15. Delay comparison under full-set flows.



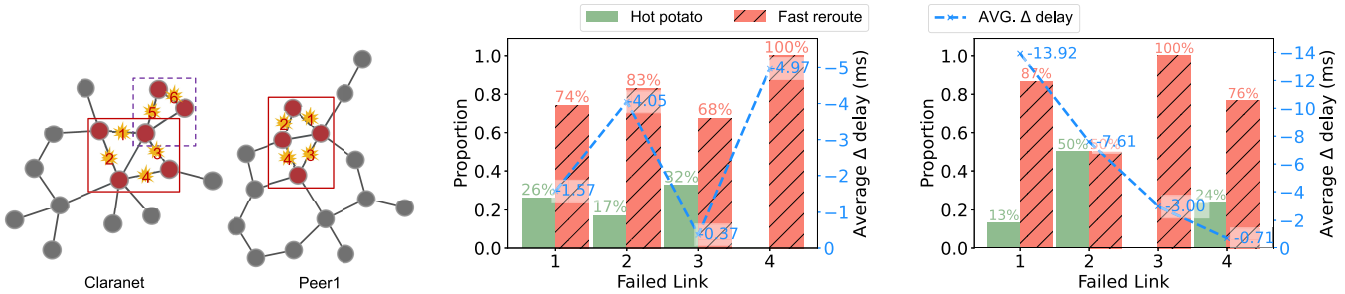
(a) Intra-AS scales (b) Convergence time

Fig. 16. Convergence time and influence of intra-AS scales.

6) Influence of Intra-as Scale: We measured the performance improvement of I^{2}BGP over BGP for different intra-AS scales. The number of forwarding nodes within each AS is respectively set to 10 to 50 for different AS scales experiments, where connections between each pair of nodes are randomly established with an average node degree of 3.

Under the aforementioned experimental setup, the performance of I^{2}BGP and BGP in terms of forwarding hops is shown in Figure 16(a). As can be seen, I^{2}BGP's performance improvement is more pronounced in larger AS scales, with

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.



(a) Failed links setting (b) Claranet (c) Peer1

Fig. 17. Fast reroute performance.

up to a 60% reduction in the end-to-end average forwarding hops. This is because the overall performance of the transmission path between ASes is cumulatively affected by each traversed AS. As the overall scale of the intra-AS increases, the performance disparity between ASes also increases, ultimately leading to greater cumulative impact.

7) Convergence Analysis: The time cost from the initial state to the processing end time of the last UPDATE message packet that updates the route table is defined as the convergence time. With the same scale of forwarding nodes within the AS, the convergence time of I^{2}BGP is shorter than that of BGP, as shown in Figure 16(b). This performance improvement can be attributed to two key factors.

On the one hand, the addition of path attributes during route diffusion involves only simple arithmetic operations, which are comparable to the *AS_Path* update operation in basic BGP. Consequently, the control-plane overhead during route diffusion remains nearly identical to that of BGP. On the other hand, the specific optimal sub-path searching of I^{2}BGP is performed as pre-computation, and the route diffusion constraint policy eliminates some unnecessary *UPDATE* messages. Although this pre-computation is part of the overhead associated with I^{2}BGP's convergence time, the measured overhead of this pre-computation process is only about 10ms, and such computations of the whole network are fully parallelizable, which will not obviously hinder network efficiency.

8) Cryptogram Overhead: The three steps of encryption, homomorphic computation, and decryption are regarded as one operation, and the computational efficiency of private key and public key generation is improved by randomly selecting corresponding numbers from a pre-computed prime number array. Implementing the *Paillier* algorithm in *Python* and the *Number Theory Library* [54], averaging the cumulative consumption time of 100,000 operations, the computational overhead is about 30ms and less than 0.1ms, respectively. Although the comparison phase includes computational and communication overheads, the distributed architecture makes the comparison phase fully parallel and independent of the route diffusion.

To further compare the performance of the *Paillier*, we conducted the same operational overhead measurement on two classical fully homomorphic encryption methods [55], *Brakerski-Gentry-Vaikuntanathan* (BGV) and *Brakerski/Fan-Vercauteren* (BFV). The experimental results show that these two algorithms take an average of approximately 170ms per operation in *Python*-based implementation, and the overhead of implementation based on *PALISADE Library* [56] is about 10ms. This discrepancy arises because fully homomorphic encryption requires support for homomorphic multiplication, which inevitably introduces higher computational overhead.

9) Fast Reroute Performance: *Claranet* and *Peer1* were selected to evaluate the performance of fast reroute. Given that the proposed fast reroute scheme is specific to triangle topology, as shown in Figure 17(a) the failed link-set 1 in *Claranet* includes links 1-4 (circled by the red solid line), and set 2 includes links 5 and 6 (circled by the purple dashed line). In *Peer1*, the failed link-set includes links 1-4 (circled by the red solid line). It is assumed that only one link fails at a time. Two sets of results of flow transmission delay using the fast reroute scheme and the *hot potato* scheme are respectively recorded when the initial path contains the failed link. The *hot potato* rule is to randomly forward the flows to an available egress router, and subsequently select a new transmission path based on the routing table entries within the next AS.

For each pair of routers, flows were generated in the experiment. The results are shown in the Figure 17(b) and Figure 17(c), where the bars represent the proportion of flows that fast reroute or the *hot potato* scheme exhibits better transmission performance over the other scheme when the link fails. The dashed line represents the difference in the average transmission delay of all flows ($Delay_{Fast_reroute}(flow_i) - Delay_{Hot_potato}(flow_i)$). In *Claranet*, packet loss occurred in failed link set 2, as the *hot potato* scheme had no other egress router options apart from the ingress and the unavailable egress under the constraint of I^{2}BGP for the corresponding flows. The results show that more flows benefit from the fast reroute scheme during link failures.

VI. Conclusion

The most widely deployed inter-AS routing protocol, BGP, mainly routes by manual configuration or selects the path with minimum AS hops according to *AS_Path* attribute. However, the AS hops-based routing policy neglects the internal state of each AS. Additionally, each AS keeps the internal network state private due to network security concerns, making it challenging for BGP to provide path-awareness capability. To this end, in this paper, we proposed a privacy-preserving intra-AS state-assisted inter-AS routing scheme, I^{2}BGP, which

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

realizes using intra-AS state to assist in inter-AS routing while ensuring the privacy of intra-AS state data. Moreover, we elaborated on the deployability of I^{2}BGP, as well as designed schemes in terms of multi-criteria extension and fast rerouting. Experimental results obtained from the constructed prototype system and real network topology-based network simulations exhibit that I^{2}BGP outperforms BGP concerning multiple metrics, including forwarding hops, delay, and bandwidth.

ACKNOWLEDGMENT

This article significantly extends [1] by adding the deployment of a prototype system to the Greater Bay Area Future Network and the fast reroute mechanism.

REFERENCES

[1] P. Cong et al., "Break the blackbox! Desensitize intra-domain information for inter-domain routing," in *Proc. IEEE/ACM 30th Int. Symp. Quality Service (IWQoS)*, Jun. 2022, pp. 1–10.

[2] G. Huston. (2021). *BGP Routing Table Analysis Reports*. [Online]. Available: https://bgp.potaroo.net

[3] (2020). *Cisco Annual Internet Report (2018-2023) White Paper*. [Online]. Available: http://www.cisco.com

[4] J. Brenes, A. García-Martínez, M. Bagnulo, A. Lutu, and C. Pelsser, "Power prefixes prioritization for smarter BGP reconvergence," *IEEE/ACM Trans. Netw.*, vol. 28, no. 3, pp. 1074–1087, Jun. 2020.

[5] R. A. K. Fezeu and Z.-L. Zhang, "Anomalous model-driven-telemetry network-stream BGP detection," in *Proc. IEEE 28th Int. Conf. Netw. Protocols (ICNP)*, Oct. 2020, pp. 1–6.

[6] Y. Rekhter, T. Li, and S. Hares, "RFC 4271: A border gateway protocol 4 (BGP-4)," Internet Eng. Task Force, Fremont, CA, USA, RFC 4271 (Proposed Standard), Jan. 2006. [Online]. Available: http://tools.ietf.org/html/rfc4271

[7] S. Jain et al., "B4: Experience with a globally-deployed software defined wan," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, pp. 3–14, 2013.

[8] Z. Yang, Y. Cui, B. Li, Y. Liu, and Y. Xu, "Software-defined wide area network (SD-WAN): Architecture, advances and opportunities," in *Proc. 28th Int. Conf. Comput. Commun. Netw. (ICCCN)*, Jul. 2019, pp. 1–9.

[9] P. Lin et al., "A west-east bridge based SDN inter-domain testbed," *IEEE Commun. Mag.*, vol. 53, no. 2, pp. 190–197, Feb. 2015.

[10] K. Lakshminarayanan, I. Stoica, S. Shenker, and J. Rexford, *Routing As a Service*. Princeton, NJ, USA: Citeseer, 2004.

[11] Q. Xiang et al., "Toward optimal software-defined interdomain routing," in *Proc. IEEE Conf. Comput. Commun.*, Jul. 2020, pp. 1529–1538.

[12] G. Asharov et al., "Privacy-preserving interdomain routing at internet scale," *Proc. Privacy Enhancing Technol.*, vol. 2017, no. 3, pp. 147–167, Jul. 2017.

[13] S. Pouryousef, L. Gao, and A. Venkataramani, "Towards logically centralized interdomain routing," in *Proc. 17th USENIX Symp. Netw. Syst. Design Implement. (NSDI)*, 2020, pp. 739–757.

[14] A. García-Martínez, P. R. Torres, and M. Bagnulo, "BGP convergence in an MRAI-free internet," *Comput. Netw.*, vol. 240, Feb. 2024, Art. no. 110183.

[15] T. Holterbach, T. Alfroy, A. Phokeer, A. Dainotti, and C. Pelsser, "A system to detect forged-origin BGP hijacks," in *Proc. 21st USENIX NSDI*, 2024, pp. 1751–1770.

[16] M. Bahnasy, F. Li, S. Xiao, and X. Cheng, "DeepBGP: A machine learning approach for BGP configuration synthesis," in *Proc. Workshop Netw. Meets AI ML*, Aug. 2020, pp. 48–55.

[17] A. Garcia-Martinez and M. Bagnulo, "Measuring BGP route propagation times," *IEEE Commun. Lett.*, vol. 23, no. 12, pp. 2432–2436, Dec. 2019.

[18] M. Chiesa, A. Kamisinski, J. Rak, G. Rétvári, and S. Schmid, "A survey of fast-recovery mechanisms in packet-switched networks," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 1253–1301, 2nd Quart., 2021.

[19] M. Milani, M. Nesler, M. Segata, L. Baldesi, L. Maccari, and R. L. Cigno, "Improving BGP convergence with Fed4FIRE+ experiments," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Jul. 2020, pp. 816–823.

[20] Q. Yang, L. Ma, S. Tu, S. Ullah, M. Waqas, and H. Alasmary, "Towards blockchain-based secure BGP routing, challenges and future research directions," *Comput., Mater. Continua*, vol. 79, no. 2, pp. 2035–2062, 2024.

[21] M. McBride, X. Fan, and D. Guzman, "BGP blockchain for metaverse—A distributed consensus system for BGP," in *Proc. IEEE Int. Conf. Metaverse Comput., Netw. Appl. (MetaCom)*, Jun. 2023, pp. 769–774.

[22] E. Zeydan, J. Mangues, S. S. Arslan, and Y. Turk, "Blockchain-based self-sovereign identity for routing in inter-domain networks," *IEEE Commun. Mag.*, vol. 62, no. 1, pp. 96–102, Jan. 2024.

[23] G. He, W. Su, S. Gao, and J. Yue, "Securing route origin authorization with blockchain for inter-domain routing," in *Proc. IFIP Netw. Conf. (Netw.)*, Jun. 2020, pp. 504–508.

[24] K. McGlynn, H. B. Acharya, and M. Kwon, "Detecting BGP route anomalies with deep learning," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr. 2019, pp. 1039–1040.

[25] A. Lutu, M. Bagnulo, C. Pelsser, O. Maennel, and J. Cid-Sueiro, "The BGP visibility toolkit: Detecting anomalous internet routing behavior," *IEEE/ACM Trans. Netw.*, vol. 24, no. 2, pp. 1237–1250, Apr. 2016.

[26] A. Gupta et al.., "An industrial-scale software defined internet exchange point," in *Proc. 13th USENIX Symp. Netw. Syst. Design Implement. (NSDI)*, 2016, pp. 1–14.

[27] R. R. Sambasivan, D. Tran-Lam, A. Akella, and P. Steenkiste, "Bootstrapping evolvability for inter-domain routing with D-BGP," in *Proc. Conf. ACM Special Interest Group Data Commun.*, 2017, pp. 474–487.

[28] J. L. Sobrinho, D. Fialho, and P. Mateus, "Stabilizing BGP through distributed elimination of recurrent routing loops," in *Proc. IEEE 25th Int. Conf. Netw. Protocols (ICNP)*, Oct. 2017, pp. 1–10.

[29] T. Holterbach, S. Vissicchio, A. Dainotti, and L. Vanbever, "SWIFT: Predictive fast reroute," in *Proc. Conf. ACM Special Interest Group Data Commun.*, 2017, pp. 460–473.

[30] A. Gupta et al., "SDX: A software defined internet exchange," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 4, pp. 551–562, 2014.

[31] R. Birkner, A. Gupta, N. Feamster, and L. Vanbever, "SDX-based flexibility or internet correctness?: Pick two!," in *Proc. Symp. SDN Res.*, Apr. 2017, pp. 1–7.

[32] A. Dethise, M. Chiesa, and M. Canini, "Prelude: Ensuring inter-domain loop-freedom in SDN-enabled networks," in *Proc. 2nd Asia–Pacific Workshop Netw.*, Aug. 2018, pp. 50–56.

[33] X. Zhao, C. Wu, and F. Le, "Improving inter-domain routing through multi-agent reinforcement learning," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Jul. 2020, pp. 1129–1134.

[34] P. Cong, Y. Zhang, W. Wang, and K. Xu, "SOHO-FL: A fast reconvergent intra-domain routing scheme using federated learning," *IEEE Netw.*, vol. 38, no. 1, pp. 272–279, Jan. 2024.

[35] Y. Zhang et al., "Federated routing scheme for large-scale cross domain network," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Jul. 2020, pp. 1358–1359.

[36] W. Xu and J. Rexford, "MIRO: Multi-path interdomain routing," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 4, pp. 171–182, Aug. 2006.

[37] X. Yang, D. Clark, and A. W. Berger, "NIRA: A new inter-domain routing architecture," *IEEE/ACM Trans. Netw.*, vol. 15, no. 4, pp. 775–788, Aug. 2007.

[38] H. Wang, Y. R. Yang, P. H. Liu, J. Wang, A. Gerber, and A. Greenberg, "Reliability as an interdomain service," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 37, no. 4, pp. 229–240, Oct. 2007.

[39] S. Peter, U. Javed, Q. Zhang, D. Woos, T. Anderson, and A. Krishnamurthy, "One tunnel is (often) enough," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 4, pp. 99–110, Feb. 2015.

[40] Y. Wang, J. Bi, and K. Zhang, "A SDN-based framework for fine-grained inter-domain routing diversity," *Mobile Netw. Appl.*, vol. 22, no. 5, pp. 906–917, Oct. 2017.

[41] B. Schlinker et al., "Engineering egress with edge fabric: Steering oceans of content to the world," in *Proc. Conf. ACM Special Interest Group Data Commun. (SIGCOMM)*, Aug. 2017, pp. 418–431.

[42] K.-K. Yap et al., "Taking the edge off with espresso: Scale, reliability and programmability for global internet peering," in *Proc. Conf. ACM Spec. Interest Group Data Commun.*, 2017, pp. 432–445.

[43] B. Barrett, "Why Facebook, Instagram, and WhatsApp all went down today," WIRED, Oct. 4, 2021. [Online]. Available: https://www.wired.com/story/why-facebook-instagram-whatsapp-went-down-outage/

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[44] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Proc. Int. Conf. Theory Appl. Cryptograph. Techn.*, 1999, pp. 223–238.
[45] S. Zhao, X. Huang, and Z. Zhong, "Research and implementation of reputation-based inter-domain routing selection mechanism," *J. Commun.*, vol. 44, no. 6, pp. 47–56, 2023.
[46] D. Chen, H. Qiu, K. Zhu, Q. Wang, and J. Zhu, "An inter-domain routing reputation model based on autonomous domain collaboration," *Scientia Sinica (Inf.)*, vol. 51, no. 9, pp. 1540–1558, 2021.
[47] P. Cong et al., "A deep reinforcement learning-based multi-optimality routing scheme for dynamic IoT networks," *Comput. Netw.*, vol. 192, Jun. 2021, Art. no. 108057.
[48] Y. Tan, W. Huang, Y. You, S. Su, and H. Lu, "Recognizing BGP communities based on graph neural network," *IEEE Netw.*, vol. 38, no. 6, pp. 282–288, Nov. 2024.
[49] H. Anton and C. Rorres, *Elementary Linear Algebra: Applications Version*. Hoboken, NJ, USA: Wiley, 2013.
[50] D. Gupta et al., "A new approach to interdomain routing based on secure multi-party computation," in *Proc. 11th ACM Workshop Hot Topics Netw.*, 2012, pp. 37–42.
[51] (2024). *NS-3 Network Simulator*. [Online]. Available: https://www.nsnam.org/
[52] Zoo.(2021). *The Internet Topology Zoo*. [Online]. Available: http://www.topology-zoo.org/
[53] Netzschleuder.(2024). *Internet Topology (pop Level)*. [Online]. Available: https://networks.skewed.de
[54] (2022). *NTL: A Library for Doing Number Theory*. [Online]. Available: https://libntl.org/download.html
[55] C. Marcolla, V. Sucasas, M. Manzano, R. Bassoli, F. H. P. Fitzek, and N. Aaraj, "Survey on fully homomorphic encryption, theory, and applications," *Proc. IEEE*, vol. 110, no. 10, pp. 1572–1609, Oct. 2022.
[56] (2025). *Palisade Lattice Cryptography Library*. [Online]. Available: https://palisade-crypto.org



Peizhuang Cong received the Ph.D. degree from the State Key Laboratory of Network and Switching Technology, Beijing University of Posts and Telecommunications, China, in 2023. He is currently a Post-Doctoral Researcher with Peking University. He has published papers in IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, WWW, and CoNEXT. His research interests include next-generation network architecture, data-driven networks, network resource management and QoS, and LLM.



Yuchao Zhang (Member, IEEE) received the B.S. degree in computer science and technology from Jilin University in 2012, and the Ph.D. degree from the Department of Computer Science, Tsinghua University, in 2017. She is currently an Associate Professor with Beijing University of Posts and Telecommunications, Beijing, China; and a Visiting Scholar with the University of Cambridge, where she is also a Research Associate with the Wolfson College. Her research interests include large scale datacenter networks, federated learning, data-driven networks, and edge computing. She is a member of ACM.



Jun Wang (Student Member, IEEE) is currently pursuing the Ph.D. degree with the State Key Laboratory of Network and Switching Technology, Beijing University of Posts and Telecommunications, China. His research interests include next-generation network architecture, congestion control, large-scale data center networks, and WAN.



Wendong Wang (Member, IEEE) received the B.E. and M.E. degrees from Beijing University of Posts and Telecommunications, China, in 1985 and 1991, respectively. He is currently a Full Professor with the State Key Laboratory of Networking and Switching Technology. He has published over 200 papers in various journals and conference proceedings. His current research interests include next generation network architecture, network resources management and QoS, and mobile internet.



Tong Yang (Member, IEEE) received the Ph.D. degree in computer science from Tsinghua University in 2013. He visited the Institute of Computing Technology, Chinese Academy of Sciences (CAS). Currently, he is an Associate Professor with the School of Computer Science, Peking University. He has published papers in SIGCOMM, NSDI, SIGKDD, SIGMOD, VLDB, ICDE, INFOCOM, USENIX ATC, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, IEEE TRANSACTIONS ON COMPUTERS, IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, IEEE/ACM TRANSACTIONS ON NETWORKING, VLDB Journal, and IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. His research interests include network big data, sketches, network measurements, bloom filters, IP lookups, KV stores, and hash tables. He is currently an Associate Editor of *Knowledge and Information Systems*.



Dan Li (Fellow, IEEE) received the Ph.D. degree in computer science from Tsinghua University in 2007. He is currently a Full Professor with the Department of Computer Science, Tsinghua University, and Zhongguancun Laboratory, Beijing, China. His research interests include future internet architecture and data center networking. His research interests also include future internet architecture and data center networking.



Ke Xu (Fellow, IEEE) received the Ph.D. degree from the Department of Computer Science and Technology, Tsinghua University. He is currently a Full Professor with Tsinghua University. His research interests include next generation internet, P2P systems, the Internet of Things, network virtualization, and network economics. He has guest edited several special issues in IEEE and Springer journals.