IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received January 31, 2017, accepted February 27, 2017, date of publication March 16, 2017, date of current version July 31, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2683943

SDN Orchestration for Next Generation Inter-Networking: A Multipath Forwarding Approach

ABDUL BASIT[1], SAAD QAISAR[1], SYED HAMID RASOOL[1], AND MUDASSAR ALI[1,2]
[1]School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Islamabad 44000, Pakistan
[2]University Engineering and Technology, Taxila 47050, Pakistan

Corresponding author: A. Basit (basit.khan@seecs.edu.pk)

ABSTRACT Increasing trend of peering at Internet exchange points (IXPs) provides a topological and network management advantage to the Internet service providers (ISPs) that is otherwise not possible through individual peering arrangements with their neighboring domains. ISPs keep analyzing potential advantages to peer at geographically diversified IXPs. Increasing degree of multihoming in ISPs requires extensive coordination among ISPs in different roles (access and transit), for mutual benefit. We propose a novel approach to build a conducive inter-networking ambiance for future data centric applications. Our approach exploits edge multiplicity of transit ISPs present across different IXPs to leverage a cross layer coordination between network stack at access ISP subscribers and the network infrastructure. Software defined networking provides a useful apparatus to control IXP fabric and also empowers our multipath forwarding strategy for an optimal network resource utilization in the inter-domain configuration. Contributions of this paper include the derivation of abstracted overlay graph from an IXP centric inter-domain connectivity model. This graph is used to provide multipath forwarding for increased reliability and throughput over abstracted graph. Adaptive cross layer coordination increases the efficiency of the proposed framework and provides an online mechanism of forwarding the traffic across the domains. We have observed as high as 54% increase in throughput using proposed scheme as compared with single path routing and forwarding strategy currently employed at the Internet backbone.

INDEX TERMS Internetworking, internet topology, software defined networking, cross layer design, quality of service.

I. INTRODUCTION AND MOTIVATION

Ubiquity of connected devices has lead to a variety of hardware for accessing the Internet ranging from handheld devices to sensors and tablet to home appliances. From an application perspective, the appetite for video content is overtaking every other content type. ISPs are investing heavily on hardware to improve their network capacity and meet ever increasing customer's Quality of Experience (QoE) demands. To enhance QoE, ISPs[1] try to establish and extend peering connections with other ISPs, either on their own or via facilitation of an IXP. Peering at IXP not only simplifies Internet topology but also creates ease in establishing Border Gateway Protocol (BGP) sessions with other Autonomous Systems (ASes). At an IXP, data of Internet subscribers has better path diversity to reach Internet backbone or to Content Delivery Networks (CDNs) that are peering at various IXPs. However, maximum utilization of inter-network capacity is constrained due to under-utilization of path diversity that can gear up throughput. De-facto inter-domain protocol, BGP, uses single path routing and forwarding decision, conforming to routing policies of the domains across the path, to navigate flows belonging to a particular ISP. Despite BGP's limited ability to accommodate innovative experimentation for future networking, ISPs do not have alternate technology with similar scalability and interoperability. We have examined a connectivity model centered around peering at IXPs and a possible abstraction can be made from this model to provide multipath forwarding across the domains, keeping inter-domain routing preference and privacy of the domains intact.

The SDN pins responsibility of control plane from forwarding hardware to a software that manages forwarding hardware centrally through abstractions and operate on a global view of the entire network. Despite Software Defined Networking (SDN) based solutions have greater ability for

[1]Term ISP is used for brevity to represent an Autonomous System or domain, even if consist of more than one domain

abstraction, extensibility and hardware interoperability for uncommon problems [1], SDN is yet to be trialled at network across the domains. Today, routing at inter-domain level is based on BGP's single-path routing strategy. Factors like limited memory of routers, individual peering model, complex inter-domain policies and expensive multipath computations, have all lead protocol designers to opt for single-path as the conventional routing strategy. However, majority of these factors are addressable by employing SDN over IXP centric topological model of inter-domain network [2], [3]. In following subsection, we have examined our motivation to accentuate the problem domain.

A. AVAILABLE PATH DIVERSITY AT IXP CENTRIC INTER-DOMAIN SUBSTRATE HAS GREAT POTENTIAL TO ASSIST EMPLOYING MULTIPATH ROUTING AND FORWARDING STRATEGY FOR THROUGHPUT MAXIMIZATION

It is known that the BGP's single path selection limits the throughput of the inter domain traffic. Capacities of available links in a certain path limit the end to end throughput for a communicating source and destination pair. Oscillation in network configuration due to BGP's churn rate adds another constraint for single-path routing. These two important factors motivate us to investigate utility of employing multipath forwarding strategy in IXP centric inter-domain substrate and suggest an alternate to the single-path routing and forwarding.

B. ADAPTATION OF MULTIPATH ROUTING AND FORWARDING STRATEGIES WITH THE HELP OF SDN IN OTHER AREAS OF NETWORKING ARE ALSO PROMISING

Designers of Datacenters (DC) networking are obtaining fruitful optimizations in network resource utilization, mitigating effects of link failure and maximizing throughput. Unlike other networking solutions, sophisticated DC networking designs essentially employ multipath forwarding mechanisms to increase reliability and robustness. Application requirements over wireless devices drive the advancement in antenna technologies. Eventually, these devices are now able to communicate over heterogeneous networks with the help of enabling technologies like Multipath TCP (MPTCP) [4].

C. MULTIHOMING IN ISPs HAS CONGRUENT BENEFITS AS OF THOSE OBTAINED AT THE END HOSTS

Benefits of multihoming at end hosts are known, however, research community is exploring its significant utilization at ISP level. ISPs with large network resources can offer transit services and get fruits of multihoming. Utility of multihoming in access ISP role is an area necessitating further investigation. We explore whether this can be a value addition in multipath context.

D. MINIMIZING MULTIPATH STRATEGY OVERHEADS IN A VERY LARGE SCALE NETWORK

Joint optimization of shortest path route selection in parallel with congestion minimization is considered a non-convex problem. Investigating all implications of employing multipath over inter-domain network is an interesting challenge. In this study, we focus on the utility of abstracting IXP centric inter-domain connectivity model where multipath forwarding strategy can be employed for reliability and higher throughput.

Following are the key research contributions of the proposed scheme.

- We have derived an online layer-3 specific multipath forwarding strategy over abstracted IXP centric inter domain connectivity model.
- We have investigated utility of multihoming for an access ISP and also researched effect of its degree for the number of multipath selection.
- An adaptive cross layer multipath forwarding framework on top of layer-3 specific multipath forwarding strategy is proposed for an IXP centric inter-domain substrate.

Rest of the paper is organized as follows. In section II, we present related researches, whereas section III and IV have covered discussion related to our proposed design. Implementation details and results have been presented in section V and section VI respectively. In section VII, we have presented important applications of our design to accentuate the utility of proposed framework. We have concluded our discussion in section VIII.

II. RELATED WORK

A. IXP AND ROUTING OUTSOURCING

Establishment of first Internet Exchange Point (IXP) as Metropolitan Area Ethernet (Metro Ethernet) in Washington DC has changed the concept of individual ISP peering. In recent years, rapid increase in IXP numbers and count of peering at every IXP is worth mentioning [14]. Therefore, futuristic approach of solving inter-domain routing problems must be IXP centric contrary to approaches revolved around individual peering relationships amongst ISPs [15]. These relationships are much more complex and routing preferences may differ across prefixes and geographic regions. The vibrant and divergent routing policies of ISPs require crafting solutions on top of IXP interconnection fabric. Recent research proposal for inter-domain routing affirms our argument and build the solution around interconnection of IXPs [3]. 91% of global IPv4 prefixes available in ISPs either directly peer at some IXP or at one hop distance from an IXP. However, this study does not propose any methodology to take advantage of edge multiplicity between different IXPs and greater reach-ability to almost entire spectrum of IPv4 prefixes. Idea of outsourcing routing is seemingly conflicting with its traditional function at inter-domain level [16]. The proposal in [5] has further highlighted need of a neutral entity that could mediate to resolve conflicting routing policies among ISPs. However, peculiarities in outsourcing inter-domain routing are intensively described and addressed by the solution presented in [3].

B. SDN FOR WIDE AREA ROUTING AND TRAFFIC ENGINEERING

Routing and Traffic Engineering (TE) in Wide Area Networks (WANs) is a well investigated topic. However, Software Defined internet eXchange (SDX) [2], along with work presented in [3] provides foundations to our inter-domain multipath forwarding proposal. Unlike TE solutions for IP network, most of the SDN based TE solutions are centralized. Scaling these TE solutions for a very large network, like the Internet, is challenging. DIFANE [6] and DevoFlow [7] are two relevant researches in which mitigation of control plane latency issues due to scalability of the network are discussed. OpenFlow wild card rules were introduced on switches so that they can handle micro flows within the data plane without involving remote controller in a large enterprise network [7]. Control of dealing with elephant flows was kept with the central controller. Whereas in DIFANE [6], incoming traffic, that is not matched with the rules cached in switches (data plane), is forwarded to the designated switch of the area (again data plane) for an appropriate action. Therefore, traffic remains in the data plane and is handled very quickly as compared to the slow control plane decisions. HyperFlow [8] handles issues of the scalability in the enterprise network with a different approach. Localized decision making is kept with local controllers whereas a central view is maintained by the central control to get all benefits of centralization. Hedera [17] presented a dynamic flow scheduling framework that identifies and treats large (elephant) flows separately from the small flows. Basic drawback in this work is heavy traffic overhead of network resource utilization statistics being collected from network to identify elephant flow. Another technique in [18] improved upon the basic deficiency in Hedera. Instead of detecting large flows in the network and wasting network resources for the collection of statistics, it adds shim layer at end host operating system to detect elephant flows. To optimize traffic flow across different datacenters, located across the globe, Google has introduced an industrial TE solution B4 [9]. It is a complex hybrid solution in which SDN operates in parallel with BGP to meet different application requirements running across datacenters and schedules different networking resources for competing flows of these applications. Google achieved up to 70 percent of overall link utilization with the help of B4 approach. However, in B4 solution, SDN controller had small network size with manageable corresponding BGP announcements. Scalability of the TE solution like B4 over a large network, like the Internet, remains questionable.

Global view of a large network with the help of distributed deployment of SDN controllers is proposed in Onix [11]. In this work, scalability issues related to SDN deployment over a large enterprise network is discussed in depth and feasibility for an Application Programming Interface (API) is proposed, to serve the coordination amongst distributed controllers. This API however, does not answer the peculiarities involved in the deployment of controllers at inter-domain level. Whereas, another study, DISCO [12] described features of solely SDN based deployment across inter-domain network. DISCO does not address issues and challenges of fully SDN based inter-domain deployment. Adaptive SDN orchestration on top of optical network spanning across multiple domains is proposed in [19]. The orchestration monitors the real-time network information and controller of respective domain utilizes global information to allocate local network resources in order to achieve system wide optimality. However, the scheme solely focuses on congestion aware service provisioning across domains and does not address policy driven forwarding issues like way-point and negative way-point routing.

C. MULTIPATH SOLUTIONS

Classical material on multipath solutions is usually classified as based on TCP layers or factors like topology, path setup mechanism (centralized and decentralized), computational complexity of objectives (congestion control, load balancing, reliability etc) [20]. In addition to these classifications, Qadir *et al.* [21] recommended to carefully address few fundamental design questions such as control plane complexity for computing multiple paths and data plane complexity to split and route the flows over computed paths.

MIRO, a multipath framework for inter-domain network, presented a solution compatible with legacy BGP routing framework. MIRO proposes creating tunnels to route some of the traffic over these tunnels, in addition to the primary single path route. Policy driven re-advertisement of available routes to the neighbors provides ASes similar control over route propagation as of BGP driven network. However, distributed nature of proposed framework makes it subject to convergence problems. MIRO does not exploit topological simplification offered by IXP centric architectures. Now let us describe few techniques actually applied in the real network. One of these techniques, B4 [9] already examined in this section. Software-driven WAN (SWAN) [10] is a technique that uses global view of the SDN enabled network to maximize inter datacenter (DC) communication. Different priorities are assigned to the network traffic classes (interactive, elastic and background) and network resources are assigned based on these traffic classes. ADCMF [22] is another multipath routing protocol that uses group of algorithms to achieve near optimal solutions for variety of objectives like congestion minimization, scalability, path oscillations etc. ADCMF scheme however, does not propose any methodology to adapt it for an inter-domain network and where transit providers have peak and off-peak traffic conditions. Another multipath routing scheme is proposed in [23] to visualize the efficacy of cross layer coordination. In this scheme, geo-diverse multipath routes are calculated in coordination with MPTCP. This scheme shows significant improvements as compared to single-path TCP in WAN environment. The scheme however, does not address peculiarities involved in the inter-domain deployment. It neither describes how different ISP roles (access, transit, etc) can be helpful to achieve path diversity in IXP centric inter-domain environment nor it states that how

TABLE 1. Summary of related work.

Referred study	Description	Networking paradigm	Relevant TCP/IP Layer	Centralized/ Distributed	Routing domain	Market Availability
Stitching paths[3]	Proposed IXP centric multigraph for Inter-domain routing	SDN	L3	Centralized	Inter-domain	No
Routing outsourcing[5]	Outsourced Inter-domain routing to a neutral agency	Hybrid	L3	Centralized	Inter-domain	No
SDX[2]	Enabling IXP to control switching fabric with SDN	Hybrid	L2	Centralized	Inter-domain	Yes
DIFANE[6]	Addressed controller latency in SDN based networks	SDN	Multiple Layers	Both	Intra-domain	No
DevoFlow[7]	Wildcard rules have been introduced to address control plane latency	SDN	Multiple Layers	Both	Intra-domain	No
Hyperflow[8]	Deals Scalability issues in a large network	Hybrid	L3	logically centralized	Intra-domain	No
B4[9]	Optimized inter-domain link utilization	Hybrid	L3,L2	Centralized	Inter-domain	Proprietary
SWAN[10]	Traffic engineering for inter-datacenter traffic	Hybrid	L3,L2	Centralized	Inter-domain	Proprietary
Onix[11]	Addressed scalability issues in large network	SDN	L3	Both	Intra-domain	Yes
DISCO[12]	Proposed a routing model for inter-domain environment	SDN	L3	Both	Inter-domain	No
RCP[13]	Proposed logically centralized overlay for inter-domain routing	Legacy	L3	Both	Inter-domain	No

inter-domain routing policies (way-point, negative, and valley free) affect multipath route selection.

Table 1 provides comparison and overall summary of the presented work in this section. We propose a multipath forwarding scheme to exploit edge multiplicity that exists between the IXP inter-connect level for better throughput, reliable and secure communication across almost entire IPv4 prefixes. Topological simplification of global peering arrangement in IXP centric inter-domain substrate provides foundation to our proposal.

III. DESIGN CHALLENGES

Increasing trend of establishing IXPs is diversifying peering relationships of the domains over the Internet. ISPs are also increasing their presence over multiple IXPs to facilitate peering interconnections with neighboring ISPs. Exploiting existing infrastructure of such ISPs, with their consent and appropriate accounting, can result in much better path diversity for the inter-domain traffic. However, extracting maximum benefit out of available path diversity is quite challenging. In this section, we discuss some important challenges.

A. ROUTING CONVERGENCE

Routing convergence remains the big problem of de-facto inter-domain routing protocol BGP and routing table convergence takes minutes even in case a single link goes down [5]. Notable contributing factors for inter-domain routing slow convergence include the distributed nature of the protocol, complex individual peering and conflicting inter-domain policies. These factors are being addressed individually by the research community. SDN has addressed distributed control plane factor and provided a centralized platform for protocol development. Whereas, substantial growth in IXP centric peering has simplified individual peering complexities that involved large overheads in terms of hardware as well as BGP peering sessions. Advances in the network policy compilation using SDN [24] has made possible to resolve policy conflicts.

B. CHURN RATE

Addressing routing oscillations, due to high churn rate, is another key challenge. Causes of high churn vary from network to network. Duplicate announcements by different routers and distributed nature of BGP are considered primary factors contributing in routing table oscillations [25]. BGP tries to overcome churn problem by tuning Minimum Route Advertisement Interval (MRAI) timer. The timer prevents removal of an entry from the routing table until its expiry and withhold any specific entry in case it receives quick updates for the entry.

C. SCALABILITY

Routing and forwarding solutions based on multipath strategy are considered resource intensive. In first week of December 2016 [26], number of ASes operating worldwide is approximately 55931 nodes. Selecting combination of source destination pairs out of this number will yield $1.5 \times e^9$ unique pairs. Providing multipath routing and forwarding across this huge number is impractical. According to [3], number of IXPs in the world by year 2014 is about 300. Growth rate of IXPs they have witnessed is also moderate i.e. approximately $\simeq$ 115 IXPs per year. Therefore, abstraction of approximately 650 nodes may be sufficient to reach almost every IPv4 prefix. Degree of multiple paths $k = 3$ in network size of 650 nodes, for every source destination pair is NP-Hard problem.

IV. DESIGN

SDN controller equipped with policy translation engine, similar to pyretic [24], on top of a substrate, built through IXP interconnections, is our design choice for inter-domain routing and forwarding proposal. This design choice is suitably placed to address the routing convergence design challenge mentioned in previous section. In our proposal, multiple paths are available for traffic forwarding between a pair of source and destination. Unavailability of a certain path would not affect all flows of a source destination pair. We have also introduced a timer described later in this section. The timer, on expiry, invokes alternative path search to replace unavailable paths at once, based on available global view of the topology at that moment. This methodology has inherent potential of resisting against churn rate. We provide prefix mapping over an abstracted network of IXP interconnections, in proposed design, for a translation from IPv4 prefix to its owner ISP that peer at some IXP or at one hop neighbor. Prefix mapping and heuristics based approach furnish our proposed design of multipath forwarding with affordable scalability. Another possible tuning to scale our multipath framework for a globe wide network is limiting the number of edge disjoint paths over this abstracted IXP centric substrate. We tune a parameter to control number of multipaths in the system. Value of the parameter depends on historical data, required filtration of paths due to domain specific policies and availability of resources such as computational power as well as number of IXP hops. We employ an adaptive cross layer coordination among design components for better throughput and reliability using AS multihoming information. Before presenting each design component in detail, we start with design assumptions of our proposal.

A. ASSUMPTIONS

We have grouped our assumptions into two distinct categories based on ISP type, access and transit.

1) TRANSIT ISPs

We assume that all ISPs offering transit services can communicate with our framework. We expect to receive peak and off-peak hours for transit service and policy regarding ingress traffic from each ISP over this communication channel. We establish an impartial routing broker Control eXchange Authority (CXA) similar to CXP in [3]. The routing broker CXA, responsible for accounting, charges the transit users on behalf of transit providers.

2) ACCESS ISPs

To perform forwarding, on behalf of access networks, broker CXA is required to have knowledge regarding degree of multihoming with respect to its peering at different IXPs. For example, if an ISP-A peer at two different IXPs, IXP-X and IXP-Y then it must be known to CXA. This topological information is required to determine degree of multipath that can be offered to an ISP user i.e. in order to maximize its Quality of Experience (QoE) or increased reliability. Secondly, we assume that prefixes belonging to an ISP is known to CXA as it is in public domain. This information shall require to efficiently route traffic in abstracted graph that will be constructed from IXP and ISP peering information. The information will also be helpful to carry out cross layer coordination between network users and network management authority, responsible for multipath routing and forwarding. Additionally, every access ISP shall provide lists for way-point and negative way-point routing policies. These policies, in addition to valley free routing, are considered important for inter-domain routing. To start with simple workable solution, we are assuming that ISPs, peering at any given IXP, do not have conflicting way-point and negative lists, for which K-Shortest Path (KSP) algorithm can not find the compliant path.

B. ABSTRACTED IXP CENTRIC INTER-DOMAIN SUBSTRATE

Management of complexity is the classical definition of term abstraction. In other words, it is a concept by which we transform one complex problem into its easier representation without changing its characteristics. As discussed in section II, with fewer number of IXPs and including one hop neighbor of the peering ISPs, we can almost reach out to the entire spectrum of IPv4 prefixes. Consider fig 1 where some ISPs are willing to provide transit services between IXPs. Few other ISPs, in access role, want their traffic to reach the rest of the Internet (prefixes in other ASes). Some other neutral ISPs may also be peering at these IXPs, however, their presence does not affect proposed model. We build an abstracted graph in which IXPs are represented as nodes and transit ISPs are edges between these nodes. We have categorized transit links into two categories. If an ISP provides transit services across exactly two IXPs, then in abstraction, we represent connectivity with solid line and call it dedicated link. Whereas, if ISPs are providing transit across multiple IXPs, then equivalent edges in abstracted graph are represented with dashed line and links are shared. It is important to know why ISPs in access role or one hop neighbors of access ISPs are not included in this abstracted graph. Recall discussion in section III, where we presented our point of view to scale our multipath forwarding scheme by abstracting IXP graph, rather than building graph from ISPs interconnections. IXP centric abstracted multigraph consisting of approximately 650 nodes, will be sufficient to provide routing across almost entire spectrum of the IPv4 prefixes across the Internet. Puzzle of forwarding on top of abstracted graph remains

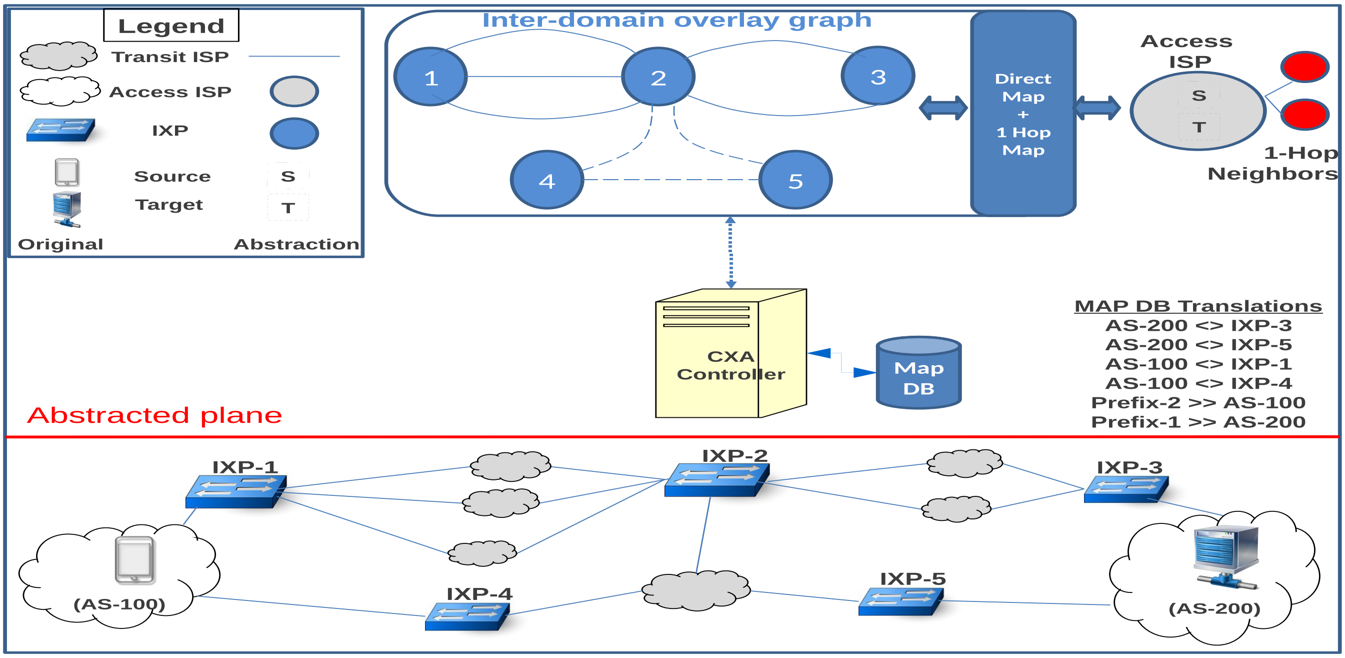


FIGURE 1. **Abstraction of IXP centric multi-domain substrate.**

unsolved if our service doesn't know relationship of prefixes with ISPs and ISP with IXP. Therefore, keeping in view their relationship, we can divide our abstraction into following four distinct processes.

1) BUILDING A MAP DB

In this phase, our framework maps a relationship of IXP nodes and all such ISPs peering at that IXP or one hop away from it. It also registers ownership of prefixes with respect to each ISP. Therefore, a Map database (Map DB) holds relationship between IXP node and all such ISPs peering at that node. It also holds the information of one hop neighbors of peering ISPs. Finally, Map DB knows what prefixes an ISP owns. On query for a prefix, Map DB provides the IXP node where prefix owner domain is directly peering. Otherwise, it provides address of a neighbor from where prefix owner domain is one hop away. This Map DB also keeps track of transit services available for the path diversity: Information like available bandwidth, load, expected delay as well as packet loss rates, peak and off-peak time and negative lists of that transit service.

2) ROUTE DISCOVERY

This phase, deals with finding multipath routes for every source and destination pair. We use K-Shortest Path (KSP) algorithm by Eppstein [27] to find number of paths for a source destination pair because of its low worst case computational complexity. We have kept initial value of $K = 3$ in our design to keep path computational overheads in a moderate range, however, traffic routed on these path is dependent on cost associated with each path explained later in this section. Moreover, to balance fairness and optimality, value of k can oscillate. Provision to discard the selected paths and recomputing of paths can also exercise to mitigate bottleneck links. Another consideration in calculating paths is way point and negative way point lists of the ISPs peering at source IXP node. These lists are assumed to be non conflicting for the scope of this work, so that paths may be discovered those are compliant to the routing policies of ISPs peering at source IXP node.

3) ROUTE RECONCILIATION AND UPDATE STRATEGY

To orchestrate an online multipath forwarding strategy for a network like the Internet, most critical aspect is ensuring reliability and low churn rate. High churn rate causes routing table to oscillate and eventually introduces network instability. Any algorithm or protocol that could result in network to oscillate from one state to another, is simply unacceptable. This is the prime reason to introduce bounds on multipath degree $K = 3$ and adding constraints on such conditions which could invoke path re-evaluation.

We propose scheme that governs route re-evaluation and update process. A flow model of this scheme is represented in fig 2. The scheme starts with initialization of different variables. Variable K is representing number of total paths for each source destination pair, variable $RCC(K)$ is Re-Calculation Counter (RCC) for each path K, Path Re-Calculation Threshold (PRT) and Link Sharing Threshold (LST) are introduced for tuning purposes. After initialization, K paths are generated as discussed in route discovery section. These generated paths are observed and evaluated on two levels. These levels correspond to the mass of the congestion. For link level congestion scenario, if the

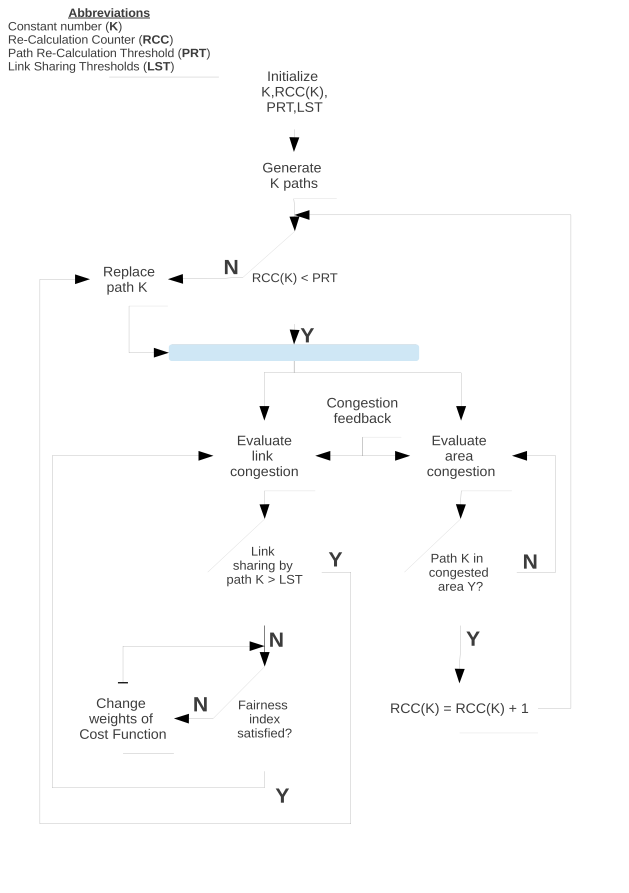


FIGURE 2. Route reconciliation and update methodology at CXA controller.

congestion appears on certain link, the scheme will enforce routing application to tune forwarding of the flows by applying appropriate weights. Tuning the weights in forwarding mechanism, discussed subsequently, can improve the congested situation on that link. However, if the congested link is used by multiple paths and number of such paths exceeds from Link Sharing Threshold (LST) then scheme will invoke path recalculation with edge disjoint settings for this link. In area congestion scenario, directly invoking path recalculation approach is not suitable especially when we have already employed link congestion remedy. Therefore, a counter based technique is adopted which invokes path recalculation on timer expiry. Links involved in area congestion gradually get their timer value high if they persist to play part in congestion. This two level route reconciliation and update strategy has enabled us to keep rationalized computational overhead with low oscillations in routing table.

4) COST FUNCTION BASED FORWARDING

Cost on each route is defined by individual link costs dependent on factors like its capacity, utilization, delay, packet drop rate and time for which transit services are being offered. These factors are defined as factor tuple using a data structure at IXP controller. Transit service number, capacity of service, current utilization, delay expected, and packet loss rate constitute this tuple. Formally, our abstracted graph is defined as, $G = (V, E)$, where $V = \{\text{Set of IXPs}\}$ and $E = \{\text{Set of ISPs offering transit services}\}$. Cost of a path $Cost_p$ is dependent on each intermediate transit edge cost $\mathcal{C}_i$ that is defined as follows.

$$Cost_p = \sum_{i \in p} \mathcal{C}_i \tag{1}$$

Cost of each intermediate transit edge $\mathcal{C}_i$ can be decomposed into three independent functions: Link Congestion State (LCS), Peak Time Constraint (PTC) and Slowly Varying Factors (SVF). LCS and PTC are exponential growth functions and adding a constant $k \in \{1, 2\}$ will result in shifting functions vertically by value k. Shifting vertically will cause horizontal asymptote to shift up by k units.

$$\mathcal{C}_i = SVF_i \cdot (1 + LCS_i) \cdot (2 + PTC_i) \tag{2}$$

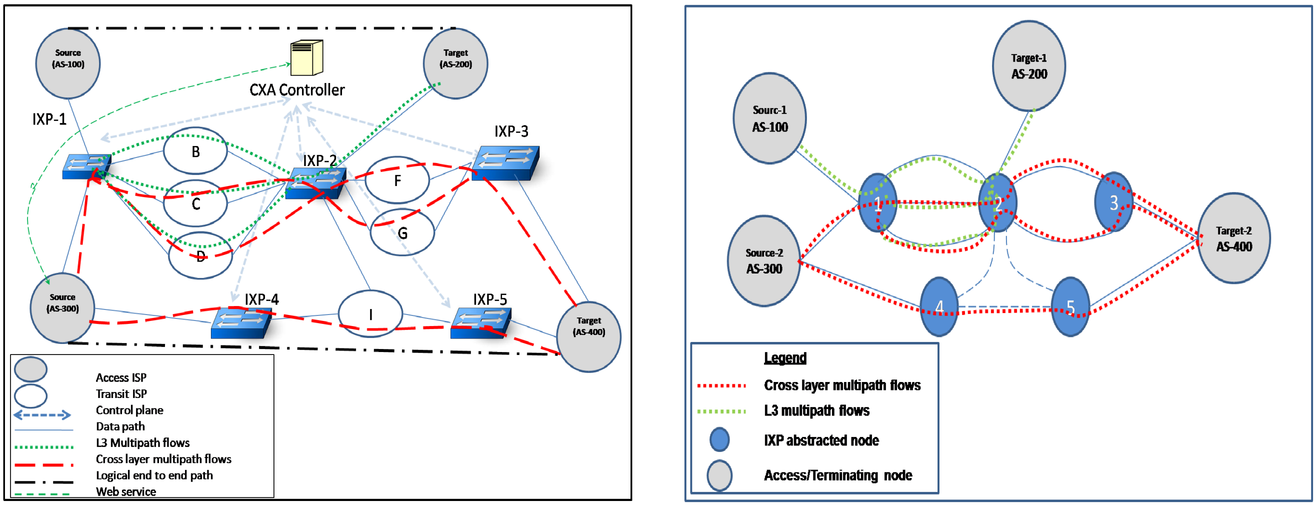
Value of SVF function depends on elements of factor tuple in MapDB for link i, discussed earlier in this section. Sum of weights W_j is equal to 1 for each normalized element X_{ij} in a factor tuple for link i. Means $\sum_{j=1}^{n} W_j = 1$. Whereas, value of X_{ij} is between range $\{0 < X_{ij} < 1\}$ as depicted in equation 3. Large value of some elements in factor tuple convey low cost meaning in our context. For example, if we have more link capacity, the link has low cost for all flows being forwarded through this link. Negative coefficient $-X_{ij}$ is used to sum these factors with corresponding weights w_1 to w_k.

$$SVF_i = \sum_{j=1}^{k} w_j \cdot e^{-X_{ij}} + \sum_{j=k}^{n} w_j \cdot e^{X_{ij}} \tag{3}$$

In function LCS_i of a link i, TP_i is a threshold price and LU_i is an average value of the link utilization history whereas CT_i is a link congestion threshold. Examining carefully, one can identify that power of the exponential function will always yield values greater than 1. Therefore, it is an exponential growth function and base value of this exponential growth function depends on selection of the threshold price for any link in the network. This threshold price varies from $\{0 < TP_i < 1\}$. Link with wider or high bandwidth may contribute with lower base growth function to overall weight calculation. Whereas, links with lesser bandwidth may be assigned TP value that is closer to 1. This will result in exponential growth function with higher base value. The higher base value of exponential function will yield more value to the weight function. This way, non dependent terms bandwidth and congestion are being correlated in this LCS_i price function.

$$LCS_i = \begin{cases} e^{\frac{1}{1-TP_i} - \frac{1}{LU_i - TP_i}} & 0 < LU_i < CT_i \\ \infty & LU_i \geq CT_i \end{cases} \tag{4}$$

In equation 5, w_t is the weight for PTC price function and $T^r = T_i^2 - T_i^1$ is the remaining time left for which this transit ISP is offering transit. If Current time (CT) is in between range time-1 and 85% of the time-2 of transit service offer time, then PTC price function will contribute accordingly to



(a) (b)

FIGURE 3. Multipath Forwarding over IXP centric network. (a) Flows in IXP centric substrate. (b) Flows in abstracted graph.

overall cost, otherwise it will make the link price infinite. Hence, function will yield favourable price for all such links that has more service time as compared to the ones with less remaining service time.

$$PTC_i = \begin{cases} w_t \cdot e^{-\frac{1}{1-T^r}} & T_i^1 < CT < (T_i^2 \cdot 0.85) \\ \infty & \text{Otherwise} \end{cases} \quad (5)$$

C. NETWORK LAYER MULTIPATH INTER-DOMAIN FORWARDING

In this paragraph, we have to discuss, how this abstraction works with proposed forwarding framework that we have presented in this work, with reference to topology in fig 3a. Abstracted overlay graph that is shown in fig 3b have five nodes corresponding each IXP in original topology. Transit ISPs are represented with shared and dedicated edges in abstracted graph. Difference between shared and dedicated edges is discussed earlier in this text. Consider a communicating node inside source AS-100. The node requests a prefix that belongs to one of the Google web servers in AS-200. Once first packet of this web request reaches at IXP-1 switch fabric, it has to follow a match-action SDN rule that is proactively installed at IXP switch. Every IXP switch in the network is installed with rules like this one. These rules are calculated proactively by our framework, based on source IXP and for all destination prefixes in the abstracted graph. As we have described earlier that prefixes are associated with directly peering ISPs or at a distance of one hop from an IXP node. In this case, traffic is entering through abstracted node 1 and its destination prefix is attached to node 2. Effect of source prefix on rout select is discussed later in this section. Node 1 will forward all the traffic to node 2 based on cost function discussed in previous section. If cost allows then three multipath routes are set as forwarding paths. Forwarding decision on multipath routes in this scenario is taken solely based on information available at Layer-3. Only important aspect that remain unattended is forwarding that is compliant to the negative and way-point routing policies. To address this aspect, CXA controller manages lists for way-point and negative way-point routing policies. In discovery phase, these lists are used while finding K shortest paths compliant to the routing policy for every source ISP.

D. CROSS LAYER MULTIPATH-PATH INTER-DOMAIN FORWARDING

Effect of multipath routes in network is generally considered as a source to provide better throughput and reliability. However, TCP is abstained with multipath routes with various delays due to out of order arrival of packets at receiver [28]. In such cases, cross layer coordination between network and end hosts can improve TCP traffic. Multiple TCP connections can be used, each for packets following same route so that amount of delay remains constant for a stream of packets. This was the basic behind the development of MPTCP [4] in which emphasis is to create parallel TCP connections (sub-flows) so that effect of latency on different streams may be handled separately. We propose that in addition to multihoming at the end hosts, ISPs peering diversity at backbone network will outperform to improve end hosts' Quality of Experience (QoE). Effects of using ISPs' multihoming are yet unknown for the better network resource utilization and improving end user QoE. In this study, we propose coordination between ISP end hosts and CXA controller to initiate MPTCP sub-flows based on CXA controller suggestion. CXA has global view of the network and it can suggest better routes based on network resource availability. This coordination can be done by using a web service provided by the CXA controller. Consider fig 3a where end hosts in AS-300 contact CXA controller using a service interface. The service suggests to initiate two MPTCP connections,

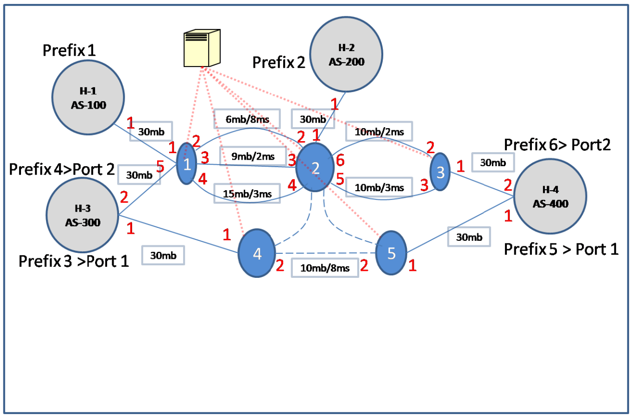


FIGURE 4. Implementation setup.

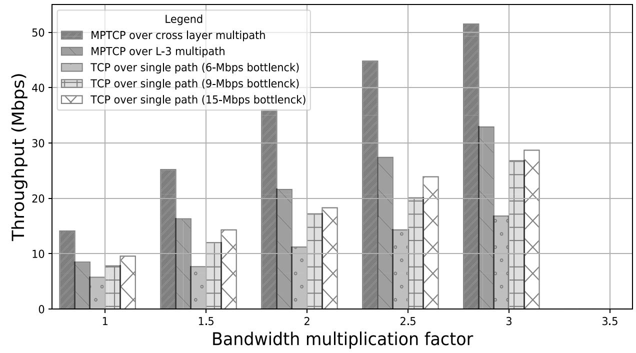


FIGURE 5. Throughput comparison of forwarding methodologies to measure cross layer technique utility.

one from gateway that leads traffic to IXP-1 and second leads traffic to IXP-4 from originating AS-300. For the route selection from IXP-1 to IXP-2, CXA knows that the delay on edge B has different readings from edge C and D. Therefore, for a network based multipath route decision, traffic must be routed via these two links so that it may not affect TCP congestion window. CXA ensures that the minimum traffic should pass through shared links to avoid congestion. Complications due to distance between CXA controller and the IXP switches are known [29] but out of scope for this work.

V. IMPLEMENTATION

In this section, we have presented the implementation details and example scenario where our design provides significant improvements. We have implemented our setup using the Mininet emulated environment [30] with MPCTP stack installed on linux kernel. For the CXA controller, we have used Ryu SDN controller because of its modularity with comparatively simple interface. We have used topology in the fig 4 to develop a testbed setup as given in the abstraction. We considered four ISPs, shown as end hosts in the gray circles peering on different IXP switches. Numbers in red colors are representing port numbers on which these switches and nodes are connected to each other. Two ISPs are having multihoming connections whereas other two ISPs are peering at single IXP. In implementation, our focus is on such cases where data from an ISP is required to traverse across multiple IXPs and proposed abstraction facilitate to let it happen. We use non-uniform traffic pattern between these AS nodes [31].

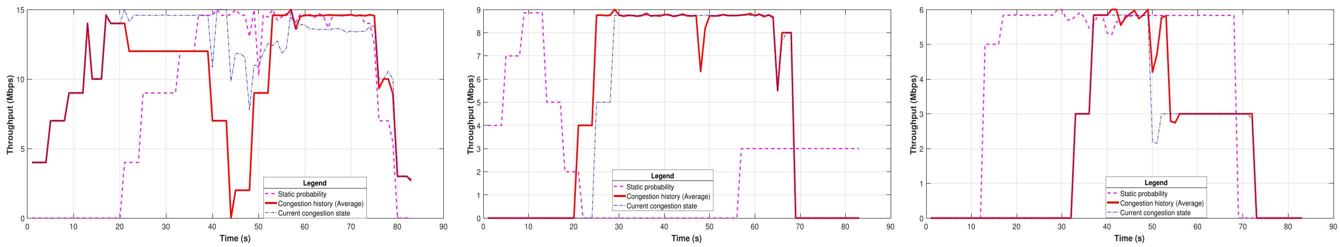
In this experiment, controller is configured to have knowledge of prefixes availability across the abstracted network. Based on this knowledge collected through reporting proxies of each domain, controller installs match action rules in a preemptive manner. Switch-1 is configured to route the traffic that is coming from ISP node-1 and node-3, on three paths with switch-2 that is representing IXP-2. These links are being offered by different transit ISPs. link-1 has low bandwidth with more latency. Whereas, link-2 and link-3 that represent different transit providers offer more bandwidth and less delay as compared to link-1. On the other hand, IXP switch-2 is connected to another IXP switch-3 with two links having similar bandwidth and latency conditions. So this emulated network has many loops in the topology and controller has tackled these loops in calculating end to end paths for the nodes attached to these switching network. It has monitored link statistics on each link across the network. Weight function presented in design section is the core of the CXA controller to calculate weights for the different links and traffic is shaped according to these assigned weights. To employ network augmented MPTCP over ISPs client, a web service is created on controller. Each ISP node coordinates with the controller to get suitable multipath subflow count. The controller calculate this count by observing available path diversity from IXP where this ISP is peering to the desired destination.

VI. RESULTS

Results presented in this section are arranged in subsections. Each subsection addresses different aspect of the design presented earlier in the text.

A. FORWARDING STRATEGIES OVER IXP CENTRIC INTER-DOMAIN SUBSTRATE

We answer basic but prime value question that is utility of multipath forwarding strategy over inter-domain level on abstracted graph. In fig 5, where horizontal axis represents bandwidth multiplication factor and vertical axis depicts throughput, we have analyzed throughput achieved by adopting various multipath forwarding strategies over inter-domain abstracted graph. A path consists of many intermediate links is always constrained by the link with lowest capacity. This fact has same impact over traffic forwarded through inter-domain routing and forwarding strategy. Path selected on various run of the experiment, by the single path routing strategy is constrained due to bottleneck link on that path. Experiment allows single path routing strategy to decide forwarding of the traffic between pair of source and destination in the network. Path selected by the routing strategy was



(a) (b) (c)

FIGURE 6. Comparison of individual path utilization for each multipath forwarding strategy. (a) Path-1 utilization. (b) Path-2 utilization. (c) Path-3 utilization.

constrained by the link with 6Mbps capacity. In other run, flow has taken another single path that was constrained due to link of capacity 9Mbps and lastly, it was constrained due to 15Mbps link. Whereas, in comparison, when we used layer-3 specific multipath forwarding strategy, it introduced an increase in throughput. Although increase in throughput is substantial when multipath strategy controls the forwarding, but the overall performance of the TCP flows are affected due to various delay factors on multiple paths. We observed actual boost in throughput when we controlled traffic forwarding using multipath strategy with cross layer coordination. Multipath forwarding strategy with network path diversity aware MPTCP has outperformed every other strategy. Cross layer coordination has communicated the degree of multiple paths in the network to MPTCP stack and it has initiated sub-flows accordingly. Maximum throughput increase of 66.67% has been observed in this case, that is two third of the TCP with single path routing strategy. Minimum increase in throughput has observed as high as 54.60% in comparison to single path routing strategy. We observed maximum gain in bandwidth is about 18Mbps in our emulated network testbed.

B. COMPARISON OF ONLINE MULTIPATH FORWARDING SCHEMES

After realization that the multipath forwarding strategy at IXP centric inter domain substrate can answer many questions, we have investigated different methodologies and schemes that could lead to near optimum solution.

1) CONGESTION STATE BASED FORWARDING EVALUATION

Results shown in fig 6 depict throughput achieved on individual path by three multipath forwarding schemes. Incoming flows are distributed across three paths for the same source and destination pair based on a static probability corresponding to the capacity of each path in first scheme named as static probability. For example Path-1 has a capacity of 15Mbps and static probability of selecting this path for an incoming flow has been set to 50%. Whereas Path-2 and Path-3 are assigned static probability of 30% and 20% respectively that correspond to their capacities 9Mbps and 6Mbps. In rest of the schemes, switches deployed on each path are periodically queried for the congestion state of their ports at every

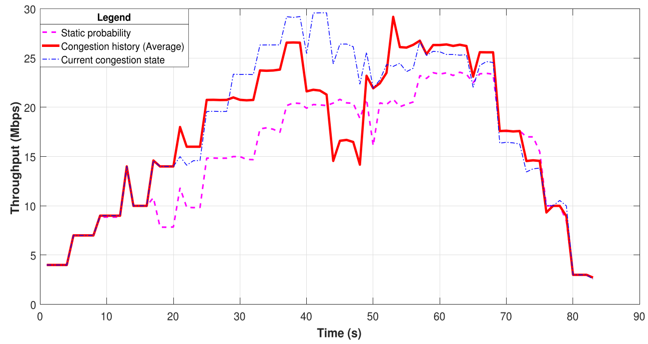


FIGURE 7. Performance of individual online multipath forwarding strategies.

3 seconds interval. This congestion state is used to divert incoming flows to avoid reaching saturation state of the paths. Specifically, in congestion history average scheme, past three samples are averaged to take the decision of path assignment for upcoming new traffic. Whereas, in current congestion state scheme, forwarding of the incoming new flows for the same source and destination pair are distributed across available paths based on current state of the congestion. We can observe individual utilization of each path in fig 6 in which static probability has randomly distributed the load over available multiple paths. Path-1 is fully utilized and Path-2 and 3 have been utilized for a little while. In contrast, congestion state based forwarding scheme has distributed load on all three paths evenly. However, fig 7 shows that congestion current state based scheme has outperformed rest of the schemes. Overall 227.1MB data was transmitted for 100 runs of this experiment. Percentage deviation of each run in this experiment was negligible. We found current congestion state scheme 26% better than the Static probability based multipath strategy. Whereas congestion history (Average) scheme is 16% better than the static probability. Path utilization is also 15% better in current congestion state scheme than the static probability multipath scheme. We can observe a drop in throughput between time 40 to 50 in congestion history (Average) scheme. The scheme has selected less wider path with relatively low congestion average for an incoming larger flow. Whereas, congestion current state scheme has decided based on current state of all paths. For brevity, consult table 2, in which experimental data is organised scheme wise.

TABLE 2. Summary of congestion state based forwarding.

Scheme	Data Transferred (MB)	Percentage of capacity	Overall Avg throughput (Mbps)	Per flow Avg throughput (Mbps)	Path-1 utilization (%)	Path-2 utilization (%)	Path-3 utilization (%)
Static probability	152.1	67.0	14.66	3.40	85.4	57.8	86.7
Congestion history (Average)	176.2	77.6	16.96	3.78	98.8	77.1	67.4
Current congestion state	191.7	84.3	18.40	4.20	100	72.3	67.4

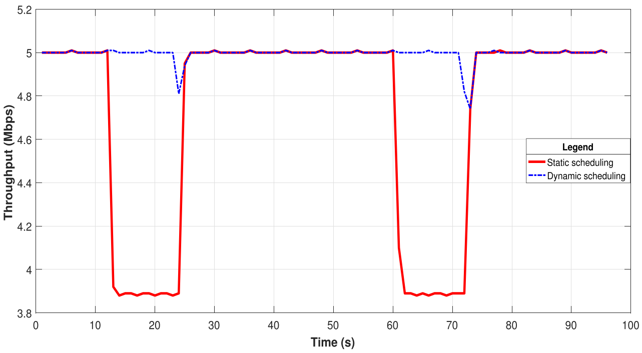


FIGURE 8. Comparison of forwarding methodologies considering peak and off-peak times.

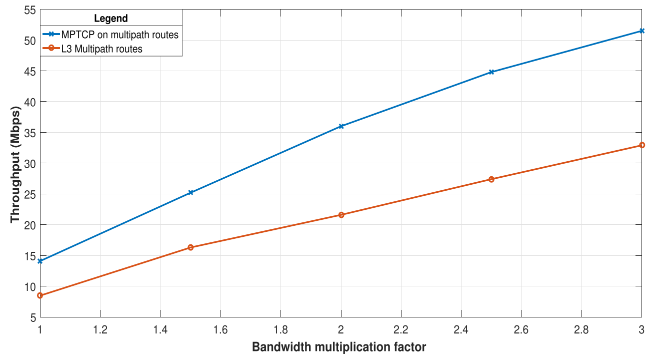


FIGURE 9. Comparison of multipath forwarding strategies.

2) PEAK/OFF-PEAK TIME AWARE SCHEDULING

In section IV, we have introduced a concept of offering network resources for a transit service based on peak and off-peak hours of the ISP. SDN provides means and ways for implementing a strategy to forward traffic based on this concept. Results in fig 8 demonstrate a clear degradation in throughput when a forwarding scheme is not aware of peak and off-peak time scheduling. In our scheme, SDN controller maintains a time record of all paths where degradation in path capacity is expected. Similarly, controller also keeps track of expected up-gradation in path capacity. It reschedules flows different paths keeping in view flow demand with respect to peak and off-peak times of available paths.

3) CROSS-LAYER COORDINATED/UNCOORDINATED MULTIPATH FORWARDING

Lastly, we conducted experiment to identify effect of cross layer coordination on multipath forwarding. Path diversity for a given source destination pair is communicated on MPTCP stack of communication end points to run specific number of MPTCP sub-flows over layer 3 multipath routes. Various throughput are obtained in the experiment by increasing bandwidth of the links and kept the ratio same in the network. In fig 9, results are presented that compare achieved throughput of cross layer coordinated MPTCP flows with non-coordinated multipath flows for a source destination pair. In non-coordinated multipath setup, packets of a single TCP stream are forwarded over multiple available paths where packets were subject to variant delay. Out of order packets were dropped at the destination and overall throughput of TCP stream was degraded consequently. Whereas, in cross layer coordination setup, MPTCP stack at source has initiated sub-flows according to availability of path diversity over the network. Packets of each MPTCP sub-flow has experienced similar delay that resulted in better throughput as compared to single TCP stream case of non-coordinated multipath setup.

VII. USE CASE EXAMPLES

There are numerous use cases of the proposed inter domain forwarding technique. We discuss QoS service provisioning and enhanced network security as two important use cases.

A. QoS SERVICE PROVISIONING

Due to distributed and independent governance of Autonomous Systems (ASes) forming the Internet backbone, provisioning of end to end QoS guaranties are hard to achieve. QoS can be partially ensured by improving routing and forwarding reliability for the traffic passing through this IXP substrate. Adding provision of multipath forwarding over backbone would increase communication reliability. To analyze reliability gains quantitatively, we can measure reliability in terms of flow blocking probability as one of the Key Performance Indicator (KPI). Suppose a network in fig 4 with a constant unit of traffic load (Erlang Unit) over a single source destination path. Erlang B formula [32] says, flow blocking probability over this path is 66.7%. Adding an additional path in the network for the same source destination pair would lower the flow blocking probability by almost 27%. Provisioning of three paths for the same pair would reduce this probability by approximately 46%.

Another KPI to measure proposed framework is delay due to congestion in the network that severely affects QoS provisioning across domains. Our proposal revolves around mitigation of congestion avoidance and selecting less congested links to forward the traffic. Results in fig 5 show cross layer coordination that results in better throughput even in the state of congestion. Although, fairness is not directly linked with the multipath provisioning, however, in multipath network it certainly affect QoS as described in [20] and backed by our result shown in fig 6. In this result, all the flows over respective path are fairly affected in case conditions of congestion arise. In our design, network controller, with global view of the network, having multipath routes available for a source destination pair, can better ensure fairness as compared to a legacy inter domain routing and forwarding protocol like BGP. BGP always has policy driven view of the network with single path provisioning for a source destination pair of domains.

B. ENHANCED NETWORK SECURITY

Internet backbone consists of heterogeneous autonomous domains peer with each other on Internet eXchange Point (IXP). These domains are not only different administratively but also employ atypical objectives. Routing and forwarding of traffic across these domains are carried out by BGP, a de-facto protocol for inter-domain communication. BGP is a plain text protocol and has well known security vulnerabilities [33]. In recent past, attackers have exploited weaknesses of Internet routing to carry out various types of attacks. BGP inherently lacks any mechanism to preserve the integrity of its routing announcements. It also does not provide any means to authenticate the source of these routing announcements. A network node can abuse and announce a forged prefix by impersonating as owner of this prefix. BGP as protocol will simply forward this announcement to neighboring Autonomous Systems (ASes) due to absence of authentication mechanism. Denial of Service (DoS) attacks are very much common over the Internet. In this attack, one can use forged announcement to create black hole over the Internet and achieve its DoS objective by exploiting this vulnerability of BGP. Man-in-Middle attacks are also possible using this forged announcement to divert traffic for eavesdropping. In addition to these vulnerabilities, BGP is also subject to TCP attacks because it is using TCP for all types of communications. Few techniques like S-BGP, BGPSec and RPKI [34] have been proposed in near past but have known limitations due to computational complexities [35] and these techniques offer mitigation of few vulnerabilities only.

BGP based security schemes are subject to slow adoption and costly new hardware deployments. Whereas, our multipath forwarding framework with its inherent security features can be used for improved security of global traffic passing through Internet backbone. Splitting traffic of a target source over multiple paths have natural tendency to resist eavesdropping in case of attack from an adversary. Attacker can not divert all the traffic, following multipath routes, to some AS where it has employed physical means for eavesdropping. Only few among all the paths can be compromised and partial portion of the traffic is vulnerable. Control plane communication in our mechanism is not exposing any global routing information. It is either network statistic's data or SDN forwarding rules for an individual switch. It is contrary in legacy BGP operated network where vital routing information is shared amongst the gateway routers in plain text and is vulnerable for exploitation. In our strategy, global picture of the network in terms of prefix announcements are not present in the network and less vulnerable for security attacks like prefix hijacking etc. To analyze security gains quantitatively, we present statics from recent research [14]. Existing multigraph of 229 IXP nodes with 49,000 edges has approximate edge multiplicity of 4.3 edges. Disjoint path diversity with mean value of 143 and 50% point to point link utilization makes only 3% of global IP traffic vulnerable if it passes through some malicious transit ISP in this multigraph. This estimate of 3% vulnerability will be further reduced in future due to induction of new IXPs. IXP growth rate is above 20% annually mentioned in [3].

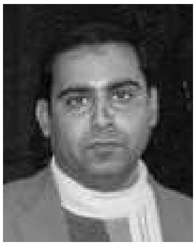
VIII. CONCLUSION

In this work, we have proposed an abstraction over IXP centric inter-domain substrate that enables innovation for future networking applications across the domains. We have employed a cross layer coordinated multipath forwarding scheme glued with SDN constructs to achieve better reliability and throughput. Presence of transit ISPs, peering at multiple IXPs, provides path diversity required to forward traffic over multiple paths. Traffic over these multipath routes is controlled with a function that is subject to parameters like capacity of links over a path, congestion state and peak off-peak hours of the transit providers. Multihoming of access ISPs is extremely useful to generate network augmented MPTCP sub-flows over multiple diversified paths in order to achieve better throughput of TCP streams. We have observed as high as 54.60% increase in throughput with the proposed cross layer coordination. Proposed framework also enables innovative service provisioning at IXP centric environment.

REFERENCES

[1] B. Davie *et al.*, "A database approach to SDN control plane design," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 47, no. 1, pp. 15–26, 2017.
[2] A. Gupta *et al.*, "SDX: A software defined Internet exchange," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 4, pp. 551–562, 2015.
[3] V. Kotronis *et al.*, "Stitching inter-domain paths over IXPS," in *Proc. ACM SOSR*, 2016, p. 17.
[4] A. Ford, C. Raiciu, M. Handley, S. Barre, and J. Iyengar, "Architectural guidelines for multipath TCP development," Internet Eng. Task Force (IETF), Fremont, CA, USA, Tech. Rep. RFC 6182, ISSN: 2070-1721, 2011.
[5] V. Kotronis, X. Dimitropoulos, and B. Ager, "Outsourcing the routing control logic: Better Internet routing based on sdn principles," in *Proc. 11th ACM Workshop Hot Topics Netw.*, 2012, pp. 55–60.
[6] M. Yu, J. Rexford, M. J. Freedman, and J. Wang, "Scalable flow-based networking with DIFANE," *SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 4, pp. 351–362, Oct. 2010.
[7] A. R. Curtis, J. C. Mogul, J. Tourrilhes, P. Yalagandula, P. Sharma, and S. Banerjee, "DevoFlow: Scaling flow management for high-performance networks," *Comput. Commun. Rev.*, vol. 41, no. 4, pp. 254–265, Aug. 2011.
[8] A. Tootoonchian and Y. Ganjali, "Hyperflow: A distributed control plane for openflow," in *Proc. Internet Netw. Manage. Conf. Res. Enterprise*, 2010, p. 3.
[9] 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.

[10] C.-Y. Hong *et al.*, "Achieving high utilization with software-driven wan," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, pp. 15–26, 2013.
[11] T. Koponen *et al.*, "Onix: A distributed control platform for large-scale production networks," in *Proc. OSDI*, vol. 10. 2010, pp. 1–6.
[12] K. Phemius, M. Bouet, and J. Leguay, "Disco: Distributed multi-domain SDN controllers," in *Proc. IEEE Netw. Oper. Manage. Symp. (NOMS)*, May 2014, pp. 1–4.
[13] M. Caesar, D. Caldwell, N. Feamster, J. Rexford, A. Shaikh, and J. van der Merwe, "Design and implementation of a routing control platform," in *Proc. 2nd Conf. Symp. Netw. Syst. Design Implement.*, vol. 2. USENIX Association, 2005, pp. 15–28.
[14] R. Klóti, B. Ager, V. Kotronis, G. Nomikos, and X. Dimitropoulos, "A comparative look into public IXP datasets," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 46, no. 1, pp. 21–29, 2016.
[15] L. Gao and J. Rexford, "Stable Internet routing without global coordination," *IEEE/ACM Trans. Netw.*, vol. 9, no. 6, pp. 681–692, Dec. 2001.
[16] K. Lakshminarayanan, I. Stoica, S. Shenker, and J. Rexford, "Routing as a service," Comput. Sci. Divis., Univ. California Berkeley, Berkeley, CA, USA, 2004.
[17] M. Al-Fares, S. Radhakrishnan, B. Raghavan, N. Huang, and A. Vahdat, "Hedera: Dynamic flow scheduling for data center networks," in *Proc. Netw. Syst. Design Implement. Symp. (NSDI)*, vol. 10. 2010, p. 19.
[18] A. R. Curtis, W. Kim, and P. Yalagandula, "Mahout: Low-overhead datacenter traffic management using end-host-based elephant detection," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 1629–1637.
[19] J. M. Fàbrega *et al.*, "Demonstration of adaptive SDN orchestration: A real-time congestion-aware services provisioning over OFDM-based 400 G OPS and Flexi-WDM OCS," *J. Lightw. Technol.*, vol. 35, no. 3, pp. 506–512, Feb. 1, 2017.
[20] S. K. Singh, T. Das, and A. Jukan, "A survey on Internet multipath routing and provisioning," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2157–2175, 4th Quart., 2015.
[21] J. Qadir, A. Ali, K.-L. A. Yau, A. Sathiaseelan, and J. Crowcroft, "Exploiting the power of multiplicity: A holistic survey of network-layer multipath," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2176–2213, 4th Quart., 2015.
[22] M. Luo, Y. Zeng, J. Li, and W. Chou, "An adaptive multi-path computation framework for centrally controlled networks," *Comput. Netw.*, vol. 83, pp. 30–44, Jun. 2015.
[23] Y. Cheng, M. M. Rahman, S. Gangadhar, M. J. Alenazi, and J. P. Sterbenz, "Cross-layer framework with geodiverse routing in software-defined networking," in *Proc. 11th Int. Conf. Netw. Service Manage. (CNSM)*, Nov. 2015, pp. 348–353.
[24] J. Reich, C. Monsanto, N. Foster, J. Rexford, and D. Walker, "Modular SDN programming with pyretic," USENIX Assoc., Berkeley, CA, USA, Tech. Rep. vol. 38 no. 5, Oct. 2013.
[25] A. Elmokashfi, A. Kvalbein, and C. Dovrolis, "BGP churn evolution: A perspective from the core," *IEEE/ACM Trans. Netw.*, vol. 20, no. 2, pp. 571–584, Feb. 2012.
[26] A. RD. *IPV4 CIDR Report* accessed on Dec. 8, 2016. [Online]. Available: http://www.cidr-report.org/as2.0/
[27] D. Eppstein, "Finding the k shortest paths," *SIAM J. Comput.*, vol. 28, no. 2, pp. 652–673, 1998.
[28] M. Allman, V. Paxson, and E. Blanton, "TCP congestion control," Internet Eng. Task Force (IETF), Fremont, CA, USA, Tech. Rep. RFC 5681, 2009.
[29] U. Javed, A. Iqbal, S. Saleh, J. Kim, J. Alowibdi, and M. Ilyas, "Analytical modeling of end-to-end delay in openflow based networks," *IEEE Access*, to be published.
[30] B. Lantz, B. Heller, and N. McKeown, "A network in a laptop: Rapid prototyping for software-defined networks," in *Proc. 9th ACM SIGCOMM Workshop Hot Topics Netw.*, 2010, p. 19.
[31] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian, "Internet inter-domain traffic," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 40, no. 4, pp. 75–86, 2010.
[32] T. S. Rappaport, *Wireless Communications: Principles and Practice*, vol. 2. Englewood Cliffs, NJ, USA: Prentice-Hall, 1996.
[33] O. Nordstróm and C. Dovrolis, "Beware of BGP attacks," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 34, no. 2, pp. 1–8, 2004.
[34] A. Cohen, Y. Gilad, A. Herzberg, and M. Schapira, "Jumpstarting bgp security with path-end validation," in *Proc. ACM SIGCOM*, 2016, pp. 342–355.
[35] M. Chiesa, G. Di Battista, T. Erlebach, and M. Patrignani, "Computational complexity of traffic hijacking under BGP and S-BGP," *Theor. Comput. Sci.*, vol. 600, pp. 143–154, Oct. 2015.



ABDUL BASIT received the bachelor's degree in computer science from the University of Management and Technology, Lahore, in 2003, and the M.S. degree in software engineering from the National University of Sciences and Technology in 2005, where he is currently pursuing the Ph.D. degree with the Department of Computing, School of Electrical Engineering and Computer Sciences. He served in various positions and in different organizations both in academia and industry for more than seven years. His research interests includes, programmable networking, network modeling and optimization, machine learning, and industrial applications of GIS.

SAAD QAISAR received the master's and Ph.D. degrees in electrical engineering from Michigan State University, USA, in 2005 and 2009, respectively, under the supervision of Dr. H. Radha (Fellow, IEEE). He is currently serving as an Associate Professor with the School of Electrical Engineering and Computer Science, National University of Sciences and Technology (NUST), Pakistan. He is the Lead Researcher and the Founding Director of the CoNNekT Lab: Research Laboratory of Communications, Networks and Multimedia, NUST. As of 2011, he is also the Principal Investigator or a Joint Principal Investigator of seven research projects in amount exceeding 1.4 million USD over next two-three years spanning cyber physical systems, applications of wireless sensor networks, network virtualization, communication and network protocol design, wireless and video communication, multimedia coding, and communication. He has authored over 40 papers at reputed international venues with a vast amount of work in pipeline. He is also serving as the Chair for Mobile-Computing, Sensing and Actuation for Cyber Physical Systems workshop in conjunction with International Conference on Computational Science and Its Applications. Based on his contributions, he was declared Overall Best Researcher for year 2012 by the National University of Sciences and Technology, Pakistan.

SYED HAMID RASOOL received the B.S. degree in electrical engineering from the Pakistan Institute of Engineering and Applied Sciences, Islamabad. He is currently pursuing the M.S. degree in electrical engineering in telecom and computer networks track with the National University of Sciences and Technology (NUST). He is also with the Connekt Research Group, School of Electrical Engineering and Computer Sciences, NUST, under the supervision of Dr. S. Qaisar. His research interests include HTTP adaptive video streaming, software defined networking, and network performance analysis.

MUDASSAR ALI received the B.S. degree in computer engineering and the M.S. degree in telecom engineering from the University of Engineering and Technology, Taxila, Pakistan, in 2006 and 2010, respectively, with major in wireless communication. He is currently pursuing the Ph.D. degree with the School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Pakistan. From 2006 to 2007, he was a Network Performance Engineer with Mobilink, an Orascom Telecom Company. From 2008 to 2012, he was a Senior Engineer Radio Access Network Optimization with Zong, a China Mobile Company. Since 2012, he has been an Assistant Professor with the Telecom Engineering Department, University of Engineering and Technology. His research interests include 5G wireless systems, heterogeneous networks, interference coordination, and energy efficiency in 5G green heterogeneous networks.