Peer Me Maybe?: A Data-Centric Approach to ISP Peer Selection

Shahzeb Mustafa, Prasun Kanti Dey*, and Murat Yuksel
University of Central Florida, Orlando, USA
Shahzeb.Mustafa@knights.ucf.edu, Prasun@knights.ucf.edu, Murat.Yuksel@ucf.edu

***Abstract*—The Internet landscape is progressively transitioning towards a flat hierarchical model to prune multiple Internet Service Provider (ISP) tiers. At the core of this transition is settlement free peering, which plays a critical role in mediating traffic exchange among ISPs. It is pertinent to take a closer look at peering and accurately emulate their operative model into a computation model that enables a concrete characterization without losing generality. In this paper, we utilize publicly available data-sets to identify the importance of several factors that play role in the peering process. We conduct a detailed analysis on the relationship of ISPs and their motivation behind selecting a peer ISP and use these findings to develop a Machine Learning (ML) based model that identifies feasible peering relationships. Preliminary results show a high correlation to the ground truth.**

***Index Terms*—Network Economics; ISP Peering; Inter-ISP routing; Network Measurement; Machine Learning**

I. Introduction

ISPs around the world rely on each other through transit and peering relationships for global connectivity. Transit ISPs often possess an enormous network which other ISPs can use to gain paid access to the Internet. The peering model, on the other hand, operates in a more collaborative manner, often leading to a settlement-free "exchange of service" type of contract. Different kinds of interconnections have their pros and cons, and ISPs may choose one over the other depending on their requirements and expectations.

"Peering is like dating" [39]: Two ISPs meet and assess if they should be peering, and if they do decide to peer but it does not work, they *de-peer*. While this brief description fails to capture the entire peering model, multiple surveys reveal that it is in fact not that far from the truth [50], [51], [38]. Many different channels, such as forums, email, and PeeringDB, can be used to contact other potential peer ISPs; however, selecting a viable candidate and then setting up an acceptable peering agreement involves surprising amount of human involvement, even today. The entire process can take months, which means that peering agreements are not dynamic enough, and can stay sub-optimal due to the large number of steps and bulky amount of paperwork needed to set them up. The need for a faster and more efficient peer selection model becomes more obvious as we look into the benefits of peering and how it is changing the Internet architecture over time [16], [25], [32], [40], [6], [20], [49]. The focus of this paper is not to discuss the importance of peering but to present a new research direction for its automation. Multiple large scale data-sets (CAIDA [11], PeeringDB [42], RIPE [36], RouteViews [44], etc.) with measurements from over a decade offer a great opportunity to take a data-centric approach in making peering decisions.

Though there is a recent realization of the need to better facilitate peering relationships [15], [35], most peering-related studies stayed in the analysis stage. Game-theoretic approaches focus mostly on economic analysis by considering both routing and congestion cost [46] to study the capacity and pricing decisions made by service providers [47]. Earlier works focused on formulating an optimal peering problem to determine the maximum peering points along with their strategic placement or a negotiation-based platform for ISPs to jointly determine routing path for traffic exchange [28], [34]. As steps towards understanding Internet-wide negotiation mechanisms, the goal of these studies was minimizing the interconnection cost without any loss of service quality.

The Internet traffic is highly volatile. An ISP admin can make estimations about internal traffic flows but external behavior is unpredictable. This implies that peering relations are often established based on speculations and trust among ISPs. To avoid future disputes, ISPs typically undergo a temporary "trial peering" period of several weeks to determine the exchanged traffic amount and patterns before provisioning the long-term peering session [10]. Adding to the already long process of peering setup, this does not leave much room for peering-based dynamic traffic engineering, which is taking place in the order of hours or minutes.

Operators are reporting an increase in periodic traffic surges as a result of, for example, software and game releases [9], [27]. In case of such surges or link failures, the ability to dynamically form short-term peering relations can help ISPs in traffic engineering. Available data can reveal temporal and spatial peering trends and help network administrators in choosing the right peer, at the right locations, and the right time, with the correct specifications. The extent to which this can be achieved needs more exploration. Towards such dynamic and automated peering, we present an analysis and the use of PeeringDB in suggesting peers such that these recommendations align with the industry expectations.

This work was supported in part by U.S. NSF award 1814086.
*Dr. Prasun Dey is currently affiliated with MathWorks, Inc.

978-1-6654-0601-7/22/$31.00

A. Contributions

Selecting the right candidates for peering requires a multi-level investigation in terms of compatibility and feasibility. Even in this age of automation, this decision is often made using personal connections, which can raise questions on its reliability and optimality. Formulating the peering decision problem is complicated and the best solution may not be as simple and straightforward. To find *good* peers, we first need a contextual and formal definition of *good*, which may vary regionally or within ISPs. We try to unfold the different aspects attached to this problem, and explore how the vast amount of available data from decades of Internet measurement research can help us reach a universal solution. To that end, we present an ML-based model that uses multiple data-sets to construct a comprehensive feature set and generate peering recommendations. Its performance reassures our confidence in the use of ML in network optimization tasks.

Our key contributions can be summarized as follows:

- Explore the potential of a data-driven approach towards making an informed peer selection.
- In-depth analysis of PeeringDB data to observe peering trends and identify common expectations that ISPs have while peering. This is an important step to make sure that the model we design is inherently in line with the industry practices and the overall ISP business model.
- Development and evaluation of an ML-based classifier that recommends if two Autonomous Systems (ASes) should establish a peering relation. Our modeling approach achieves 85% accuracy when validated using CAIDA AS-Relationships data-set.

The rest of the paper is organized as follows: Section II discusses several relevant research and data-sets on ISP peering and its management, Section III highlights some of the key observations from an analysis of the PeeringDB data, Section IV presents a ML model that identifies potential peering candidates for an AS, and finally, Section V showcases the potential of such a model while listing some of the limitations.

II. Related Work

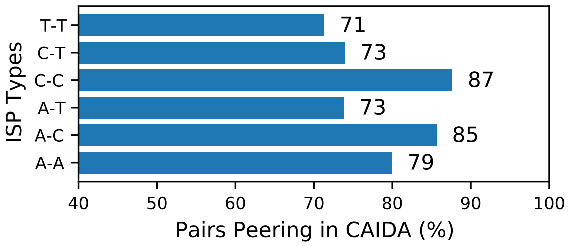
There has been a sizable amount of work on facilitating and reducing the setup costs of inter-ISP peering relations. Route Bazaar [15] and Dynam-IX [35] provide a multi-layered platform to Internet eXchange Point (IXP) members for a more efficient inter-ISP communication and connection establishment. "Picking a Partner" provides a blockchain-based AS scoring that can help in *reliable* peer selection [3]. GENIX is a framework that uses a public networking test-bed (GENI) to emulate IXPs [37]. This can be particularly helpful in recreating internal traffic scenarios and testing the performance of different automation tools.

A large number of papers exist on inter-AS routing and peering measurement. In particular, the measurement studies focusing on IXPs are the most relevant to our work. R. Klöti et al. presents the first comparative analysis of three IXP databases (PeeringDB, Euro-IX, and PCH) [29] and highlights their key characteristics. In-depth analysis of PeeringDB and what it reveals in terms of the peering ecosystem has also been explored [32]. Many researches focus on the internal activities of IXPs, their evolution, and their impact on the overall Internet architecture. Some of these conduct elaborate studies, clear misunderstandings, and reveal surprising facts that were previously unknown [14], [1], [2], [16], [24]. Similarly, "Try Before You Buy" provides a network test-bed that is designed to experiment with inter-AS relations using Software-Defined Networking (SDN) [45]. Cardigan [48] is a distributed router that uses 'routing as a service' abstraction for reducing operational complexity. Endeavour and a Software-Defined IXP (SDX) present and evaluate an SDN-based IXP architecture that can give administrators more control in traffic engineering [26], [4]. These prior efforts focused on the management of peering relationships once a peering decision has been made by an ISP, while our work primarily focuses on automating the peering process by providing tools to help ISPs before their peering decisions. Meta-Peering, our prior work, takes a step in the same direction and presents a tool that can help network administrators in selecting feasible peering locations [18] by formulating it as an optimization problem. In this paper, we use an ML-based data-driven approach to model which factors influence the peering decisions.

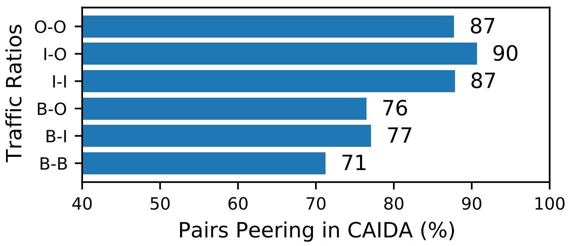
While the majority of the researches study the AS network and present an in-depth analysis of their evolution over time [30], [20], [23], [19], some studies focus on the types of disputes, issues, and complexities that exist in the peering market among different stakeholders [43], [21], [31], [5], [52]. Packet Clearing House (PCH) and CAIDA have conducted large scale surveys with network administrators in an effort to understand the peering trends and the AS relationships. Some of the well-maintained and high dimension data-sets, and network tools available to the public include PeeringDB, CAIDA (multiple data-sets), PCH [41], Euro-IX [22], Route-Views, BGPStream, Route Atlas. In this paper, we primarily rely on PeeringDB historical data dumps and two data-sets from CAIDA (AS Relationships and AS Rank).

III. Understanding Peering Trends

In order to understand what matters in a peering deal from an administrator's point of view, we construct an AS profile directory using PeeringDB and CAIDA AS-Rank [12], and identify the AS pairs that are peering according to CAIDA AS-Relationships data [13]. Although PeeringDB does not guarantee complete accuracy, its usage in understanding peering trends is justified because of the fact that this information is provided by ISPs themselves. For example, an ISP advertising only a subset of its Point-of-Presence (PoP) locations is probably willing to peer at those locations only. For the purpose of this paper, we focus only on ISPs with at least three PoPs in the United States since peering trends that we are studying can be relative to their respective regions.



(a) For different AS types (Access/Transit/Content)



(b) For different Traffic Ratios (Inbound/Outbound/Balanced)

Fig. 1: Percentage of AS pairs peering according to CAIDA AS Relationship data

TABLE I: Frequently asked requirements by ASes

Requirements	*% of ASes*
24*7*365 Support	37
No static route/ default route	11
Accurate peeringDB entry	14
IPv6 Peering Required	48
Multilateral/Bilateral Preference	8
Do not announce third party routes. (Only self customer cone)	40
Minimum geographic presence (peering at least in PoP count)	28
Interconnection speed at each point	18
Provide security; handle DDoS and abuse	48
Traffic ratio (in-bound: out-bound)	8
Routes registered in IRR, RIPE, ARIN	20

Some ISPs post their peering expectations and requirements on peeringDB. In some cases these are posted in the optional “notes” section, and sometimes as a web-page link or an email address which can be contacted for information. Out of 1,295 ASes in the US, we found that only 262 have publicly posted their requirements clearly. While there is no set standard for such reporting, we observe that most ISPs share the same set of requirements. We noticed that a large number of ISPs are concerned about IPv6 peering, round-the-clock support, and substantial security. Table I shows some more detail about these requirements. The percentage values reported are corresponding to 262 ASes that have publicly posted their requirements.

A. *ISP Types and Peering*

We analysed peering trends among different kinds of ISPs and found that the highest “peering rate” is among Content-Content (C-C) and Access-Content (A-C) pairs at 87% and 85%, respectively. Here we refer to “peering rate” as the percentage of pairs in the CAIDA-inferred AS Relationships data that are peering. In other words, 87% of C-C pairs in CAIDA AS Relationships data are peering, as illustrated in Fig. 1a. Dey et. al. present an interesting analysis of A-C peering and how this *vertical* integration is changing the Internet architecture and economics [19]. Figure 2 shows that traffic ratio (Inbound/Outbound/Balanced) and ISP type (Content/Transit/Access) are strongly correlated. Content ASes tend to be more outbound as they are providing content to consumers while access ASes tend to be mostly inbound. As expected, transit ASes are mostly balanced. Therefore, observing the peering rates for different traffic ratios revealed very similar trends when compared to AS type distribution in Figure 1b. ISP pairs in CAIDA AS Relationships data with Inbound-Outbound (I-O) and Inbound-Inbound (I-I) traffic ratios categories showed the highest peering rates at 90% and 87% respectively.

ISPs can also advertise their peering policy (Open/Selective/Restrictive) for each of their AS on peeringDB to present their willingness to accept new requests. A very small number of ASes, 10%, were closed to peering. We examined the ‘notes’ section for such ASes and found that most of them were either in the process of integrating with another AS or had another AS (as part of the same institution) designated for peering. In some cases, ASes were only interested in private peering. According to these notes and advertised policies, we did not find any access, content or transit AS that was absolutely against peering. We further observed that more than 64% of the content ASes are *Open* to peering and only about 6% are closed to it (Figure 2).

The Internet traffic volume has grown exponentially in the last few years and the effect can also be seen in peeringDB. Figure 3 reflects the increase in network wide traffic volume with the increase in AS port capacities which is now moving into PetaBytes. The recent developments in high resolution media (4k, 8k) and the increase in consumption of video content with the introduction of several new streaming services have resulted in a significant increase in content ISPs’ port capacities. Comparing this to 2016 when most ASes had less than a TeraByte of capacity, the number of ASes with coverage in the US has more than doubled from around 500 to 1,300 in the last 5 years. The number of peers has also increased from 5,000 to more than 11,000.

B. *AS Path Comparison using BGP Dumps*

In order to inspect the impact of peering on inter-AS paths, we analyzed BGP path advertisements in BGP dumps from 11 Route-Views collectors [44] in the U.S. For this purpose, we selected 51 AS pairs. For each ISP pair we searched for advertised BGP paths between both ISPs, using their

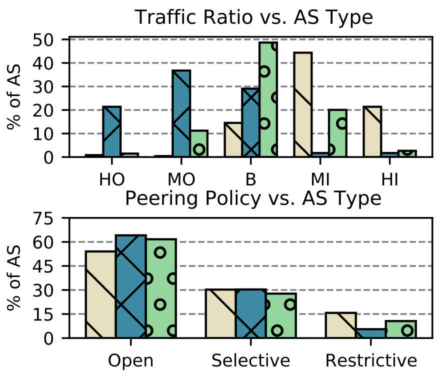


Fig. 2: AS traffic ratio and peering policy distribution w.r.t. AS types.

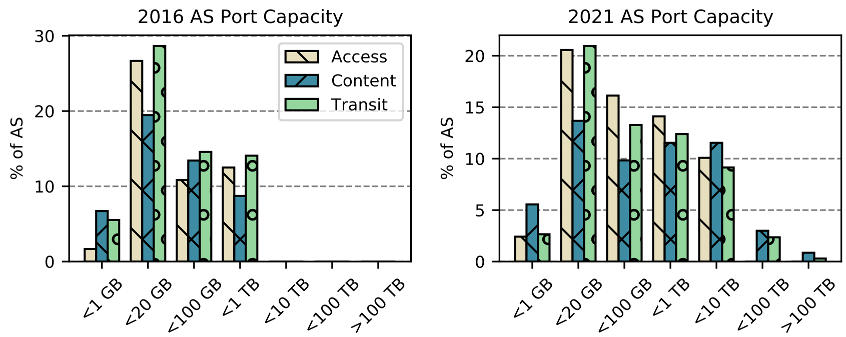
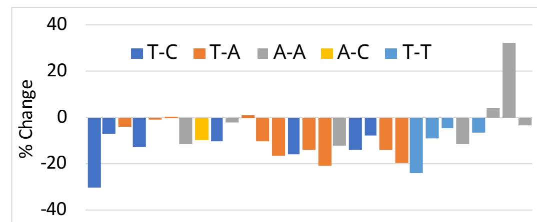


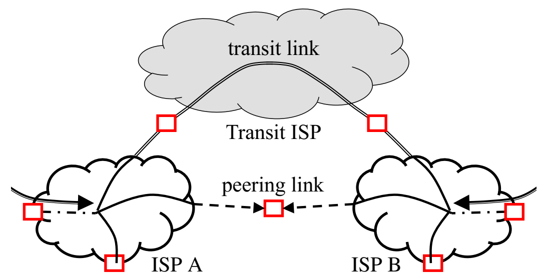
Fig. 3: AS Port Capacities have increased significantly in response to an exponential growth in network traffic volume.

AS numbers AS_A and AS_C. Among these, 22 had direct BGP paths between them since they had a customer-provider relationship, according to CAIDA. For the remaining 29 pairs, we selected the shortest of all advertised AS paths between ASes and were able to identify at least one transit ISP (with AS number as AS_T) in each case. These ASes were not peering and were therefore using a transit route for communication. While this route was costing them an extra AS hop ($AS_C \rightarrow AS_T \rightarrow AS_A$), we wanted to evaluate this extra cost in terms of geographical distance. In other words, was routing traffic through AS_T resulting in a longer than necessary route?

We gathered corresponding PoP location information from PeeringDB for the three ASes (AS_C, AS_T, AS_A) and used PoP IDs to identify common PoPs among them, which are the possible points of traffic exchange. Restricting the flow of traffic through AS_T, we calculated the geo-distance for routes between each connected PoP. Using this value as edge weight, we then calculated shortest path between each of AS_A and AS_C's PoPs, storing the total route weight/cost in a distance matrix. Similarly, to simulate a peering relationship, we populated another distance matrix representing only direct routes from AS_A and AS_C. Note that these are not straight line distances from source to destination, but length of the shortest route from source to destination, which may include multiple router hops. Figure 4a shows that, apart from only four cases, we observed a reduction in average route geographical distance between PoP locations. Interestingly enough, in many cases, AS_A and AS_C shared the same PoP location with AS_T and therefore showed no change in distance because of peering. In such cases, ISPs can potentially save on transit costs by using the same route using settlement-free peering.



(a) Change in average path distance because of peering



(b) Transit vs. Peering

Fig. 4: (a) Most ISPs ASes show a reduction in average route length because of peering, (b) Routing through a transit AS may result in a longer path in certain cases.

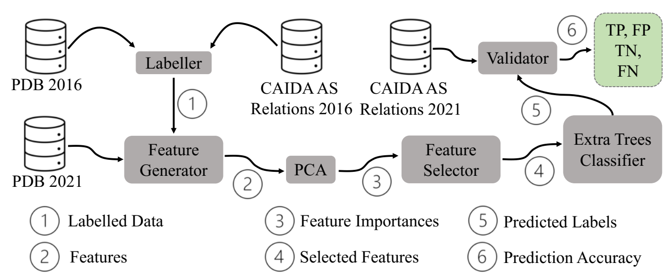


Fig. 5: Peering predictor framework

IV. Peer Selection Model

Extending our analysis of the collected data, we design an Extra Trees Classifier [8] that makes use of a feature set derived from the above observations to predict whether two ASes should be peering. Figure 5 illustrates a detailed overview of the peering predictor framework. We clean and process PeeringDB (PDB) data from 2016 and 2021. The reason we chose the year 2016 was because that is when PeeringDB changed how it reports its data, and the same format is being used currently. Therefore, we used the earliest possible data in the current format, that is 2016, and the latest data available at the time of testing, that is 2021. The *Labeller* then uses the 2016 CAIDA AS Relations to output labelled data with information on which of the pairs are peering. Next, the *Feature Generator* creates different features as discussed later in this section. These features are then fed into the *PCA* module, which calculates the importance of each feature in predicting whether or not two ASes are peering and feeds these importance values to the *Feature Selector*. Selected features are fed into the *Extra Trees Classifier*. The labelled 2016 data is used to train the classifier, and the unlabelled 2021 data is used to test the model and generate predicted labels which are fed to the *Validator*. Here, the predicted labels are matched with CAIDA 2021 AS Relations data, and the final accuracy of the predictive model is calculated in terms of True Positive (TP), False Positive (FP), True Negative (TN), False Negative (FN). We use the labelled data to train and experiment with three different classifiers. First, we test a Linear Regression based classifier which gave 83% accuracy. Next, we tested two perturb-and-combine [7] techniques (Extra Trees and Random Forests) based on randomized decision trees [17] Of the two, Random Forests gave an accuracy of 84% and Extra-Trees gave an accuracy of 87% so we chose that as our primary classifier.

The trained model predicts whether an AS pair should peer. We use the 2021 PeeringDB processed data from step 2 for labeling each pair as peering/not-peering, which is then validated against 2021 CAIDA AS relationship data.

We observed that some of the features held very similar importance and also represented the same aspect of an AS. For example, the number of advertised AS numbers (ASNs), IP prefixes, and IP addresses all represent the cone size of an AS. We also realize that, in such cases, the features have a very strong direct relation to each other, i.e., if one increases, the other increases too. To reduce the computational cost of the model and to enhance its performance by reducing noise, we removed some of the irrelevant features and combined some of the remaining ones as described later in Sec. IV-A. Figure 6 shows the final list of features that we use and their respective importance.

Figure 7 shows the correlation of these features to the probability of peering calculated using Pearsons method. A negative value indicates an inverse relation, and vice versa. As the difference in the cone size of two ASes increases, their willingness to peer is expected to drop. Similarly, the probability of peering increases as peering policy between two ASes becomes more *Open*. While recommending peering, we optimize the probability threshold by minimizing the misclassification rate in the training stage. For the results posted in this paper, 0.63 was used as the optimal threshold. In other words, the model will recommend peering only if the predicted probability for a pair is more than the set threshold.

A. Feature Set

We utilized the set of measurements collected by CAIDA as well as peering policy parameters posted by ISPs at PeeringDB.

For each AS pair, we derived 24 features. Among these features, we observed that multiple measurements relate to the same aspect of the AS. For example, the difference in the number of PoPs, the number for common PoPs, and the number of non-common PoPs, all relate to *PoP Affinity* as explained later in this section. Similarly, the number of providers, customers and peers for an AS refer to how well-connected the AS is with other ASes. We therefore take an average of these three values to represent connectivity, and for an AS pair, derive *Connectivity Difference*, later discussed in this section. Lastly, the number of advertised IP prefixes, addresses, and ASNs all relate to the AS cone size. For each AS pair, we combine these values into *Cone Size Difference*. In this section we have discussed in detail some features that we derived like *PoP Affinity* and *Cone Overlap* and also some features where we converted raw values according to a custom scale like *Traffic Level Difference* and *Traffic Ratio Difference*. A full list of features and their descriptions can be found in Table II.

Cone Size Difference: Ideal ISP peers should have similar sizes in terms of the number of customers they serve. Typical way to quantify the size of an AS (belonging to an ISP) is to measure its customer cone size. In the literature, in order to represent the cone size for an AS, various studies considered the number of advertised IP prefixes, addresses, and ASNs. We use the average of these numbers (collected from CAIDA AS Rank) to derive a single metric that relates to AS cone size. To give same weight to all three numbers, we normalize them before taking the average. For an AS pair, we then use the difference of their cone sizes as a feature. Figure 7 shows that the cone size difference between the two ASes is inversely related to the probability of peering. Intuitively this makes sense, a large AS is less likely to peer with a smaller AS in most cases because of asymmetric traffic loads.

Cone Overlap: Using the CAIDA AS Relationships inference, we constructed customer cones for each AS, i.e., the customer cone of an AS includes all the ASes that are customer to that AS or within the cone of the customer ASes. Usually peering among two ASes does not allow traffic transit between their *indirect* customers. An ASes indirect customers refer to the customers of its customers. This means that if AS A and AS C are peering, indirect customer ASes of A will not be able to reach indirect customer ASes of C using the peering link as a transit route. If a provider AS A starts peering with one of its customers AS C, its customer cone size will reduce (see Figure 8). Therefore, for each AS pair, we calculate the number of ASes that are present in the customer cones of both ASes. We refer to this as the *Cone Overlap* and expect it to play a significant role in peering decisions of ASes.

Traffic Ratio Difference: Many ASes have advertised their average traffic ratios and levels on PeeringDB (Balanced/Inbound/Outbound). Content ASes (like Facebook, Amazon, and Netflix) tend to be Heavily Outbound since they have to provide data to users. On the other hand, Access

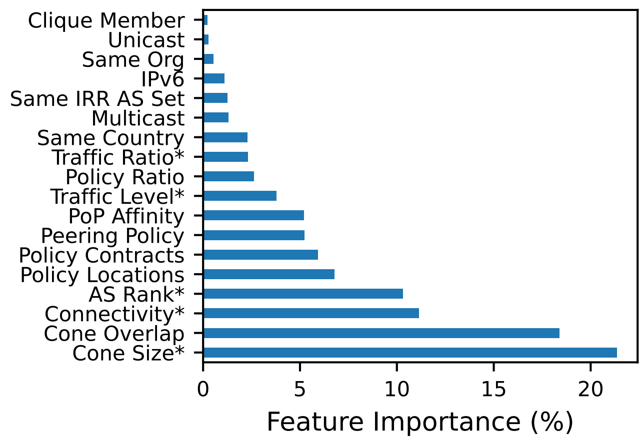


Fig. 6: Principal Component Analysis (PCA) shows that some features are more important for peer suggestion.
**: Difference between two ASes*

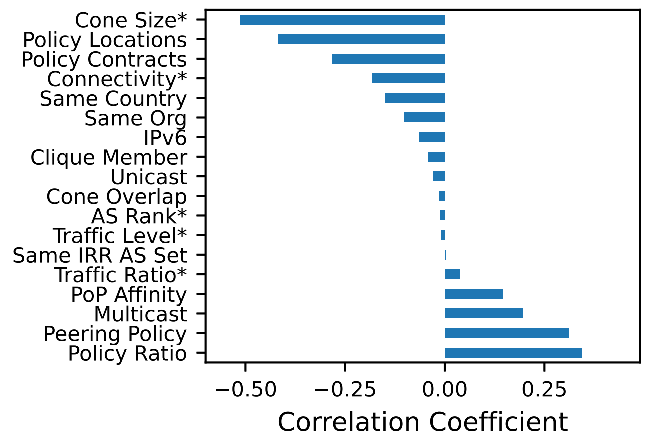


Fig. 7: Correlation of features to peering probability. Positive value implies direct relation, negative implies inverse.

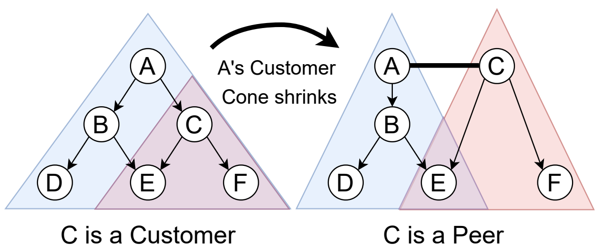


Fig. 8: Because of a *Cone Overlap*, AS A loses some ASes from its customer cone as a result of peering. Here A and C have an overlap of 2.

ASes tend to be Heavily Inbound as their users are content consumers. For each AS, we convert its advertised traffic ratio to an integer value according to the scale below.

```
"Heavy Inbound": -2,
"Mostly Inbound": -1,
"Balanced": 0,
"Mostly Outbound": 1,
"Heavy Outbound": 2
```

Then, for each AS pair, we use the sum of their converted traffic ratio values as a metric. For example, the traffic ratio sum for two Heavy Inbound (-2) ASes will be $-2+(-2) = -4$. Similarly, the traffic ratio sum for a Heavy Inbound and Heavy Outbound AS will be $-2+2=0$.

Traffic Level Difference: In addition to their traffic ratios, many ASes have also advertised their *traffic levels* reported in terms of throughput ranges, e.g., 0–20 Mbps. The advertised traffic levels tend express the size of an AS. For each AS, we convert these ranges into integer values using the scale below.

```
"0-20 Mbps": 0,
"20-100 Mbps": 1,
...
"500-1,000 Gbps": 11,
"1 Tbps+": 12,
...
"50-100 Tbps": 17,
"100+ Tbps": 18
```

Then, for each AS pair, we use the difference in their converted traffic levels as a feature in our classifier model for determining whether they will be compatible for peering. As an example, an AS with reported traffic level of 500-100 Gbps will have a converted score of 11, and an AS with a reported traffic level of 50-100 Tbps will have a converted score of 17. The difference between their traffic levels will, then, be $17-11=6$.

Peering Policy: Many ASes have advertised their peering policy on PeeringDB (Open, Selective, Restrictive). ASes willing to peer advertise an *Open* peering policy while some also advertise *Selective*, implying that they have specific requirements for a peering agreement. We use a simple scoring method to represent the likelihood of two ASes peering by using the intuition that ASes with an *Open* peering policy are more likely to peer compared to ASes with a *Restrictive* policy:

```
"Open": 2,
"Selective": 1,
"Restrictive/No": 0
```

For each pair, the total score is the sum of the individual policy scores shown above. This way we are able to assign the highest score (4) to Open-Open pairs and the lowest score to Restrictive-Restrictive (0) pairs.

Connectivity Difference: CAIDA analyzes BGP dump data to predict the relationship among ASes and estimates, for each AS, the number of providers, customers, and peers. We

Name	Type	Description
Clique Member	Boolean	Whether both ASes are clique members.
Unicast	Boolean	Whether both Ases require multicast.
Same IRR AS Set	Boolean	Whether the two Ases are in the same Internet Routing Registry (IRR) AS Set.
Multicast	Boolean	Whether both ASes require multicast.
IPv6	Integer	Difference in recommended number of IPv6 routes/prefixes to be configured on peering sessions.
Same Country	Boolean	Whether the two ASes are based in the same country
Policy Ratio	Boolean	Whether both ASes have a peering ratio requirement.
Policy Locations	Boolean	Whether both ASes require peering at multiple locations.
Traffic Ratio Difference	String	How different the traffic ratio for the two ASes is.
Same Org	Boolean	Whether the two ASes belong to the same Organization.
Traffic Level Difference	String	How different are the advertised traffic for the two ASes.
Peering Policy	String	How different is the peering openness for both ASes.
Policy Contracts	Boolean	Whether both ASes require peering contract
Pop Count Difference	Integer	Difference in the number of PoPs for both ASes.
Common Pop Count	Integer	Number of common PoPs for both ASes.
Provider Count Difference	Integer	Difference in the number of provider ASes for both ASes.
Customer Count Difference	Integer	Difference in the number of provider ASes for both ASes.
Non-Common Pops Count	Integer	Number of non-common PoPs for both ASes.
Rank Difference	Integer	Difference in CAIDA AS ranks for both Ases.
ASN Count Difference	Integer	Difference in the number of ASNs addresses for both ASes.
Peers Count Difference	Integer	Difference in the number of peer ASes for both ASes.
Num Addresses Difference	Integer	Difference in the number of advertised addresses for both ASes.
Num Prefixes Difference	Integer	Difference in the number of prefixes addresses for both ASes.
Cone Overlap	Integer	The number of ASes that are in both AS's customer cone.

TABLE II: Feature Descriptions

use these counts to define the connectivity of an AS to other ASes. We take the average of these three counts to derive a connectivity metric. Then, for each AS pair, we calculate the difference in connectivity and use it as a feature for our ML-based model. The intuition here is that connectivity of an AS represents what 'tier' it sits within the inter-AS mesh and expresses how it is positioned with respect to the other ASes. Hence, for peering ASes, the difference between their connectivity levels should be small, i.e., it shows an inverse relation with the probability of peering.

AS Rank Difference: CAIDA uses its relationship inference algorithm to assign a rank to each AS, which represents its cone size relative to others. An AS's rank is one greater than number of ASes with larger customer cone sizes [12]. ASes with similar customer cone sizes are likely to have similar ranks. Similar to cone size difference, we expect AS rank difference to play a notable role in expressing the peering potential of an AS pair, and hence use it as a feature. PCA shows that the difference in AS rank between two ASes is highly important in choosing a peer.

PoP Affinity: An AS will be interested in peering if the relationship would expand its coverage area; otherwise, there may not be enough incentive to peer with an AS that is covering the same locations or has a smaller coverage area. For a given AS pair, we define this interest as *PoP Affinity*:

$$\alpha_A = \frac{P_C - P_o}{P_A \cup P_C} = \frac{P_C - P_o}{(P_A - P_o) + (P_C - P_o) + P_o} \tag{1}$$

$$\alpha_C = \frac{P_A - P_o}{(P_A - P_o) + (P_C - P_o) + P_o} \tag{2}$$

where P_A and P_C are the number of PoPs for ASes A and C respectively, and P_o is the number of their common PoPs. Since the number of PoPs can be used as a measurement of the size and expanse of an AS, this metric helps us gauge the benefit of peering in terms of increased coverage. We use geometric mean to calculate the combined PoP Affinity:

$$\alpha_{AC} = \sqrt{\alpha_A * \alpha_C}. \tag{3}$$

The geometric mean assures that both ASes A and C will increase their coverage if they peer. A situation where only A or C increases its coverage from peering would not be desirable since only one AS would benefit from peering in that case. As expected, PoP affinity α_{AC} has a positive impact on the probability of peering.

B. Validation

We validate peering recommendation model results on 630 access ISPs, 472 content ISPs, and 987 transit ISPs using 2021 CAIDA AS Relations data. Figure 9 shows that 89.3% of the peering suggestions (TP + FP) made were found to be in fact peering (TP). Among the actually peering pairs (TP + FN), our model predicted 95.1% of them to be peering (TP). On the other hand, 81.4% of the NOT peering suggestions (TN + FN) made were found to be in fact not peering (TN). Among the not peering pairs (TN + FP), our model correctly predicted 65.3% of them (TN). Table III provides a detailed explanation of what each of the four graphs are referring to.

This shows that our model aligns very well with the real world peering trends. It performs particularly well when suggesting peering, but as seen it needs improvement when suggesting that two ASes shouldn't peer. These predictions were made based on a number of different factors as discussed earlier. While each ISP may have a different motivation for peering, they still share some key interests and that justifies the use of the same feature set across all ASes.

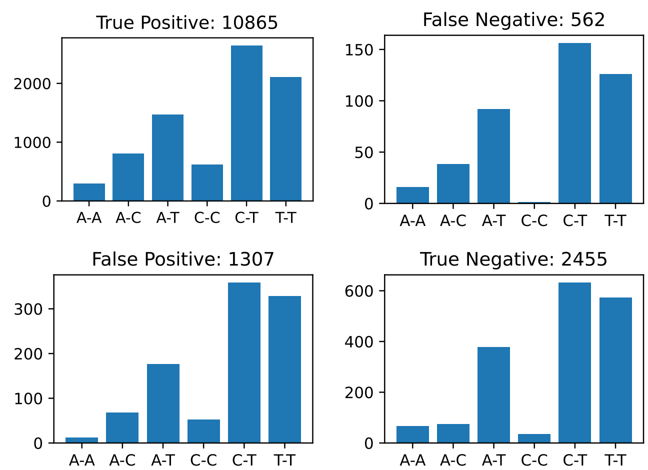


Fig. 9: CAIDA validation for peering recommendation.

The classifier assigns weights to different features and constructs a criteria that defines a potential peering opportunity. In accordance with this criteria, the model makes 1,633 new pair suggestions that are not peering and also suggested that 400 of the peering pairs should not be peering. This brings our overall accuracy to 87.7%. It would be interesting to see if the pairs we suggested to peer end up actually peering in a few years.

Result	Descriptions	
	Prediction	Ground Truth
True Positive (TP)	PEERING	PEERING
False Positive (FP)	PEERING	NOT PEERING
True Negative (TN)	NOT PEERING	NOT PEERING
False Negative (FN)	NOT PEERING	PEERING

TABLE III: Explanation about what each results category means with reference to CAIDA AS Relations.

Figure 9 shows a detailed distribution of our results in comparison to CAIDA. In these predictions, we did not find our classifier to be biased towards a specific AS type. The higher values for T-T and C-T categories in every graphs do not necessarily mean that the model is more successful in classifying transit ISP's peering relationship, it merely reflects the presence of more transit ISPs in our analysis.

V. Summary and Discussion

We studied AS-level measurements from CAIDA datasets and peering policy posts of ISPs at PeeringDB to devise a model that predicts the probability of peering between two ASes. Understanding the dynamics of peering decisions will be key to automating the peering relation establishment. Recent studies show an increased interest in this direction as inter-AS peering decisions can play significant role in critical inter-domain problems such as traffic engineering and path security. We believe that the Internet architecture can greatly benefit from a data-centric approach towards peer selection by utilizing several years of measurement data from different layers. We defined a feature set for modeling the ISP peer selection. These features have notable importance in the peering decision-making process in the following order: Cone Size Difference, Cone Overlap, Peering Policy, AS Rank Difference, Connectivity Difference, PoP Affinity, Traffic Level Difference, and Traffic Ratio Difference. The classifier we developed is the first of its kind and its simplicity and performance demonstrate its great potential.

A. Limitations

What we presented is an initial step in making use of network measurement data towards automating the ISP peering process. This initial work has the following limitations:

- Training and testing labels (CAIDA AS Relations) are based on an inference algorithm, the accuracy of which is not completely validated. In fact, only 34.6% of the inferred relationships are verified by network admins [33]. Still, this is the single largest verified data available.
- The ML model focuses more on the holistic relationship among ISPs, rather than specific AS paths and link health. This particular model cannot work in a dynamic environment, which is one of the key motivations of automated peer selection. Since not all features can be converted to columns, as in our case, a more complex deep learning model will be needed to take full advantage of the available data, that can also keep track of changes in the network links.
- Our primary source of data is PeeringDB which is highly unstructured and incomplete in several aspects. We had to manually go through the advertised requirements and expectations of each AS. In many cases, they only provided an email address that can be used for requesting peering information. Such limitations of PeeringDB hinder large scale analyses of the peering ecosystem.

B. Future Prospects

We believe that there is a large potential for future work in this direction of research:

- A chronological study of the AS graph can reveal several other features, such as latency and path stretch, that can help further optimize peer selection and also bring this model one step closer to experimental deployment.
- The classifier can be integrated with other tools discussed in Sec. II, particularly the ones that are more accepted in the industry. This will be useful in evaluating its performance and impact in the wild.
- Peering among different AS types may stem from varying motives. For example, C-A peering helps improve latency for content delivery to end users, while T-T peering allows inter-regional connectivity. Extending this model to make AS type aware recommendations would make it more versatile and accurate.

REFERENCES

[1] B. Ager, N. Chatzis, A. Feldmann, N. Sarrar, S. Uhlig, and W. Willinger. Anatomy of a large european IXP. In *Proceedings of the ACM SIGCOMM 2012 conference on Applications, technologies, architectures, and protocols for computer communication*, pages 163–174, 2012.

[2] M. Z. Ahmad and R. Guha. A tale of nine internet exchange points: Studying path latencies through major regional IXPs. In *37th Annual IEEE Conference on Local Computer Networks*, pages 618–625. IEEE, 2012.

[3] Y. Alowayed, M. Canini, P. Marcos, M. Chiesa, and M. Barcellos. Picking a partner: a fair blockchain based scoring protocol for autonomous systems. In *Proceedings of the Applied Networking Research Workshop*, pages 33–39, 2018.

[4] G. Antichi, I. Castro, M. Chiesa, E. L. Fernandes, R. Lapeyrade, D. Kopp, J. H. Han, M. Bruyere, C. Dietzel, M. Gusat, et al. Endeavour: A scalable SDN architecture for real-world IXPs. *IEEE Journal on Selected Areas in Communications*, 35(11):2553–2562, 2017.

[5] S. Bafna, A. Pandey, and K. Verma. Anatomy of the internet peering disputes. *arXiv preprint arXiv:1409.6526*, 2014.

[6] T. Böttger, G. Antichi, E. L. Fernandes, R. di Lallo, M. Bruyere, S. Uhlig, and I. Castro. The elusive internet flattening: 10 years of IXP growth. *CoRR*, abs/1810.10963, 2018.

[7] L. Breiman. Rejoinder: Arcing classifiers. *The Annals of Statistics*, 26(3):841–849, 1998.

[8] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[9] J. Brodkin. iOS 7 downloads consumed 20 percent of an ISP's traffic on release day, Nov 2013.

[10] T. W. Cable. IPv4 and IPv6 settlement-free peering policy.

[11] "CAIDA". Center for Applied Internet Data Analysis (CAIDA). Technical report, "CAIDA", 2016.

[12] CAIDA UCSD - AS Rank. www.api.asrank.caida.org/v2/docs.

[13] CAIDA UCSD - AS Relationships. www.caida.org/catalog/datasets/as-relationships/, 2013.

[14] J. C. Cardona Restrepo and R. Stanojevic. A history of an internet exchange point. *ACM SIGCOMM Computer Communication Review*, 42(2):58–64, 2012.

[15] I. Castro, A. Panda, B. Raghavan, S. Shenker, and S. Gorinsky. Route bazaar: Automatic interdomain contract negotiation. In *15th Workshop on Hot Topics in Operating Systems (HotOS {XV})*, 2015.

[16] N. Chatzis, G. Smaragdakis, A. Feldmann, and W. Willinger. There is more to IXPs than meets the eye. *ACM SIGCOMM Computer Communication Review*, 43(5):19–28, 2013.

[17] Scikit Learn - Decision Trees. https://scikit-learn.org/stable/modules/tree.html.

[18] P. K. Dey, S. Mustafa, and M. Yuksel. Meta-peering: towards automated isp peer selection. In *Proceedings of the Applied Networking Research Workshop*, pages 8–14, 2021.

[19] P. K. Dey and M. Yuksel. Peering among content-dominated vertical ISPs. *IEEE Networking Letters*, 1(3):132–135, 2019.

[20] A. Dhamdhere and C. Dovrolis. The internet is flat: modeling the transition from a transit hierarchy to a peering mesh. In *Proceedings of the 6th International COnference*, pages 1–12, 2010.

[21] A. D'Ignazio and E. Giovannetti. Asymmetry and discrimination in internet peering: evidence from the LINX. *International Journal of Industrial Organization*, 27(3):441–448, 2009.

[22] Euro-IX. https://www.euro-ix.net/, 2014.

[23] V. Giotsas, M. Luckie, B. Huffaker, and K. Claffy. Inferring complex AS relationships. In *Proceedings of the 2014 Conference on Internet Measurement Conference*, pages 23–30, 2014.

[24] E. Gregori, A. Improta, L. Lenzini, and C. Orsini. The impact of IXPs on the AS-level topology structure of the Internet. *Computer Communications*, 34(1):68–82, 2011.

[25] A. Gupta, M. Calder, N. Feamster, M. Chetty, E. Calandro, and E. Katz-Bassett. Peering at the internet's frontier: A first look at ISP interconnectivity in Africa. In *International Conference on Passive and Active Network Measurement*, pages 204–213. Springer, 2014.

[26] A. Gupta, R. MacDavid, R. Birkner, M. Canini, N. Feamster, J. Rexford, and L. Vanbever. An industrial-scale software defined internet exchange point. In *Proceedings of USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, pages 1–14, Santa Clara, CA, March 2016.

[27] M. Jackson. ISP BT sees 40% increase in internet traffic from fortnite season 10, Aug 2019.

[28] R. Johari and J. N. Tsitsiklis. Routing and peering in a competitive Internet. In *2004 43rd IEEE CDC (IEEE Cat. No. 04CH37601)*, volume 2, pages 1556–1561. IEEE, 2004.

[29] R. Klöti, B. Ager, V. Kotronis, G. Nomikos, and X. Dimitropoulos. A comparative look into public IXP datasets. *ACM SIGCOMM Computer Communication Review*, 46(1):21–29, 2016.

[30] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian. Internet inter-domain traffic. *ACM SIGCOMM Computer Communication Review*, 40(4):75–86, 2010.

[31] A. Lodhi, N. Laoutaris, A. Dhamdhere, and C. Dovrolis. Complexities in internet peering: Understanding the "black" in the "black art". In *2015 IEEE Conference on Computer Communications (INFOCOM)*, pages 1778–1786. IEEE, 2015.

[32] A. Lodhi, N. Larson, A. Dhamdhere, C. Dovrolis, et al. Using peeringDB to understand the peering ecosystem. *ACM SIGCOMM Computer Communication Review*, 44(2):20–27, 2014.

[33] M. Luckie, B. Huffaker, A. Dhamdhere, V. Giotsas, and K. Claffy. As relationships, customer cones, and validation. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 243–256, 2013.

[34] R. Mahajan, D. Wetherall, and T. Anderson. Negotiation-based routing between neighboring ISPs. In *Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation-Volume 2*, pages 29–42. USENIX Association, 2005.

[35] P. Marcos, M. Chiesa, L. Müller, P. Kathiravelu, C. Dietzel, M. Canini, and M. Barcellos. Dynam-IX: A dynamic interconnection exchange. In *Proceedings of the 14th International Conference on emerging Networking EXperiments and Technologies*, pages 228–240, 2018.

[36] T. McGregor, S. Alcock, and D. Karrenberg. The ripe ncc internet measurement data repository. In *International Conference on Passive and Active Network Measurement*, pages 111–120. Springer, 2010.

[37] S. Mustafa, P. K. Dey, and M. Yuksel. GENIX: A GENI-based IXP emulation. In *IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 1091–1092. IEEE, 2020.

[38] W. B. Norton. Internet service providers and peering. In *Proceedings of NANOG*, volume 19, pages 1–17, 2001.

[39] W. B. Norton. *The Internet peering playbook: connecting to the core of the Internet*. DrPeering Press, 2011.

[40] W. B. Norton. The history of internet peering. DrPeering.net, 2014.

[41] Packet clearing house. https://www.pch.net.

[42] PeeringDB. https://www.peeringdb.com, 2004.

[43] D. Reed, D. Warbritton, and D. Sicker. Current trends and controversies in internet peering and transit: Implications for the future evolution of the internet. In *2014 TPRC Conference Paper*, 2014.

[44] Route-views: Collectors. Routeviews.

[45] B. Schlinker, K. Zarifis, I. Cunha, N. Feamster, E. Katz-Bassett, and M. Yu. Try before you buy:{SDN} emulation with (real) interdomain routing. In *Open Networking Summit 2014 ({ONS} 2014)*, 2014.

[46] S. Secci, J.-L. Rougier, A. Pattavina, F. Patrone, and G. Maier. Peering equilibrium multipath routing: a game theory framework for internet peering settlements. *IEEE/ACM Transactions on Networking (TON)*, 19(2):419–432, 2011.

[47] N. Shetty, G. Schwartz, and J. Walrand. Internet QoS and regulations. *IEEE/ACM Transactions on Networking (TON)*, 18(6):1725–1737, 2010.

[48] J. Stringer, D. Pemberton, Q. Fu, C. Lorier, R. Nelson, J. Bailey, C. N. Corrêa, and C. E. Rothenberg. Cardigan: SDN distributed routing fabric going live at an Internet exchange. In *2014 IEEE Symposium on Computers and Communications (ISCC)*, pages 1–7. IEEE, 2014.

[49] The 'Donut Peering' Model: Optimizing IP transit for online video. MZIMA White paper, Sept. 2009.

[50] B. Woodcock and V. Adhikari. Survey of characteristics of Internet carrier interconnection agreements. *Packet Clearing House*, 2011.

[51] B. Woodcock, M. Frigino, and P. C. House. 2016 survey of Internet carrier interconnection agreements, 2016.

[52] M. Yannuzzi, X. Masip-Bruin, and O. Bonaventure. Open issues in interdomain routing: a survey. *IEEE network*, 19(6):49–56, 2005.