Inter-domain Link Inference with Confidence Using Naïve Bayes Classifier

Yi Zhao
State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou, China

Yan Liu*
State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou, China

Xiaoyu Guo
State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou, China

Zhonghang Sui
State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou, China

ABSTRACT

Inter-domain link inference is not only important for network security and fault diagnosis, but also helps to conduct research on inter-domain congestion detection and network resilience assessment. Current researches on this issue lack confidence analysis of the inferred results. In this paper, the IP link types (i.e., intra-domain link and inter-domain link) are considered as the latent variable in probability model, while the parameters are probabilities of different link types with particular features. The expectation maximization algorithm is applied to estimate parameters of the model. In each iteration of EM algorithm, Naïve Bayes is used for classification. The final result is determined according to the probability, and the probability is the confidence of the result. The experimental results show that our method can achieve better precision and recall on the validation set than two existing general methods.

CCS CONCEPTS

• **Networks** → Network properties; Network structure.

KEYWORDS

Internet mapping, Inter-domain link inference, AS boundary, Probability model

ACM Reference Format:
Yi Zhao, Yan Liu, Xiaoyu Guo, and Zhonghang Sui. 2021. Inter-domain Link Inference with Confidence Using Naïve Bayes Classifier. In *2021 13th International Conference on Machine Learning and Computing (ICMLC '21), February 26–March 01, 2021, Shenzhen, China.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3457682.3457715

1 INTRODUCTION

An Autonomous System (AS) is a set of routers run by one or more network operators with a single and clearly defined routing policy [1]. Inter-domain link inference is the process of identifying links between different ASes (inter-domain links) in the network topology, distinguishing them from links within an AS (intra-domain links). Inferring inter-domain links is a challenge in Internet topology research, facilitating accurate modeling of Internet topology, diagnosing network faults, analyzing network resilience and robustness, detecting inter-domain congestion, etc.

Intuitively, inter-domain links can be identified simply by using IP-to-AS mapping techniques. For example, if the IP addresses at each end of an IP link belong to different ASes, it is straightforward to determine that this IP link is an inter-domain link. However, the actual Internet infrastructure dictates that the addresses at both ends of a link usually have the same network prefix (i.e., they belong to the same network segment). This can cause the interface address of the border router to potentially originate from the address space of its neighbor AS, leading to the misjudgment of inter-domain links. In addition, when responding to traceroute probes, routers may use different interface addresses, including third-party addresses [2], which may influence the inter-domain link inference.

Existing studies on inter-domain link inference are classified into individual AS and Internet scale depending on the scope of the method to be handled. Luckie *et al.* [3] presented bdrmap which firstly performed targeted traceroutes from various vantage points (VP) inside an AS to its neighbor ASes, and then used knowledge of traceroute idiosyncrasies and topological constraints to correctly identify inter-domain links at the router-level. It can only infer inter-domain links attached to the network hosting VPs and observed by the VPs. Inferring inter-domain links at the Internet scale generally identifies links where the routers at both ends belong to different ASes after inferring the owner of routers. Tangmunarunkit *et al.* [4, 5] chose the most frequently assigned AS in the origin AS set of the router. Chang *et al.* [6] argued that a router is a border router if its interface addresses are assigned from multiple administration domains, and proposed the intersection rule, majority rule, and hole-filling heuristic. Huffaker *et al.* [7] developed five different heuristics to produce an AS-router dual graph. Marder *et al.* [8] created bdrmapIT, which combined bdrmap [3] and their previous algorithm MAP-IT [9] used for iteratively inferring inter-AS links at the interface-level graph, to infer router owners at the Internet scale. Although the current studies have yielded good results, none of them have assigned any confidence to their results.

In this study, in order to assign confidence to the result, we propose a probabilistic model-based method for inferring inter-domain links. First, the IP link classification problem is modeled as a probabilistic model with IP link type as the hidden variable and the probabilities of particular features occur for different link types as the model parameters. The design of features is based on

[*] Corresponding author; E-mail: ms.liuyan@foxmail.com.

ICMLC '21, February 26–March 01, 2021, Shenzhen, China

ACM ISBN 978-1-4503-8931-0/21/02.
https://doi.org/10.1145/3457682.3457715

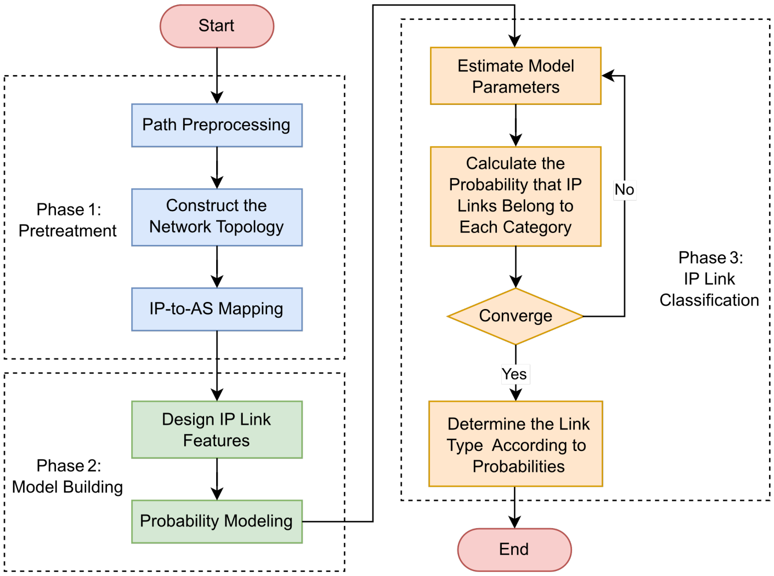
Start
Path Preprocessing
Construct the Network Topology
IP-to-AS Mapping
Phase 1: Pretreatment
Design IP Link Features
Probability Modeling
Phase 2: Model Building
Estimate Model Parameters
Calculate the Probability that IP Links Belong to Each Category
Converge
No
Yes
Determine the Link Type According to Probabilities
Phase 3: IP Link Classification
End

Figure 1: The flow chart of the probabilistic model-based inter-domain link inference method.

domain knowledge, and features that have good differentiation for inter-domain links and intra-domain links are selected. Then, the parameters of the model are estimated using the expectation maximization algorithm. In each iteration of the EM algorithm, Naïve Bayes is used for classification. The algorithm converges and determines the classification result based on final probabilities, which are considered as the confidence of the result. Finally, all IP links are classified into two categories, intra-domain links and inter-domain links, achieving inter-domain link inference with confidence.

This study makes two main contributions as follows.

(1) A probabilistic model-based method for inter-domain link inference is proposed by modeling the IP link classification problem as a probabilistic model with hidden variables, achieving the classification of intra-domain links and inter-domain links.

(2) The inferred results of inter-domain links have confidence. The experiment results show that both the precision and recall on validation sets of our method are better than two existing approaches.

2 METHOD

In order to improve the shortcomings of existing methods, this study considers a probabilistic model-based inter-domain link inference method, whose main idea is to use a probabilistic model to classify IP links into two categories: intra-domain links and inter-domain links, and to make the results with confidence.

The flow chart of the method is shown in Figure 1, which mainly includes three stages: preprocessing, model building and IP link classification.

Below are the main steps of the method:

- **Path preprocessing.** First, to avoid mistaking hosts for routers, when extracting IP addresses from traceroute paths, only addresses that appear as intermediate hops in some paths are included, i.e., the destination addresses are ignored. After that, paths with only one IP address and paths with IP loops are deleted. In addition, delete IP addresses in paths that belong to the BGP route server AS Numbers (ASNs) provided by IXPs for establishing multilateral peering among their member ASes.
- **Construct the network topology.** Using paths processed in the previous step, generate an IP-level network topology map, where IP addresses are nodes, and IP links between adjacent IPs in paths are edges.
- **IP-to-AS mapping.** Annotate the AS to which each IP belongs using IP-to-AS mapping technique.
- **Design IP link features.** Based on domain knowledge and data analysis, summarize and extract features that can distinguish inter-domain links from intra-domain links by considering different perspectives.
- **Build probabilistic models.** The IP link classification problem is modeled as a probabilistic model with IP link type as the hidden variable and probabilities of particular features occur for different link types as parameters.
- **IP link classification.** The parameters of the model are estimated using the expectation maximization algorithm, and the probability of an IP link belonging to each class is calculated using the Naïve Bayes classifier. In each iteration of the EM algorithm, EM can converge and attribute stable parameters to Naïve Bayes. Furthermore, Naïve Bayes can work well even when there are correlations between features [10–12]. Finally, the type is determined according to probabilities after the convergence condition is reached.

In the above steps, IP link features design and IP link classification are keys, which are specified separately below.

2.1 IP Link Features Design

2.1.1 *AS Relation Feature of Adjacent IP Links.*

The link type relates to the business relationship between ASes. Without considering complex hybrid relationship, AS relationships can be categorized as customer-to-provider (c2p), peer-to-peer (p2p) and sibling-to-sibling (s2s) [13]. In Internet practices, when two ASes connect with each other by the point-to-point link, the IP subnet of the link (usually a /30 or /31 in IPv4) is typically from one of the two ASes [14]. It is worth noting that in p2c relationships, the provider usually supplies the address space for the IP subset [15].

In our method, a traceroute path can be represented as a sequence of IP addresses: $p = IP_1, IP_2, ..., IP_n,\ n \geq 2$. A single path contains several IP links: $\langle IP_i, IP_{i+1} \rangle,\ 1 \leq i \leq n-1$. The $IAS(IP_i)$ is expressed as the ASN that IP_i is mapped to. We use $REL(\langle IP_i, IP_{i+1} \rangle)$ to denote the AS relationship between $IAS(IP_i)$ and $IAS(IP_{i+1})$.

Considering AS relationship, with regard to the IP link $\langle IP_i, IP_{i+1} \rangle$, its **forward AS relation feature** is the set $\{(REL(\langle IP_{i-1}, IP_i \rangle), REL(\langle IP_i, IP_{i+1} \rangle))\}$, where the tuple $(REL(\langle IP_{i-1}, IP_i \rangle), REL(\langle IP_i, IP_{i+1} \rangle))$ denotes the relation of AS relationship labels of IP link $\langle IP_i, IP_{i+1} \rangle$ and its forward link $\langle IP_{i-1}, IP_i \rangle$ in a path. The forward AS relation set is obtained after traversing all paths in P. Similarly, the **backward AS relation feature** $\{(REL(\langle IP_i, IP_{i+1} \rangle), REL(\langle IP_{i+1}, IP_{i+2} \rangle))\}$ of IP link $\langle IP_i, IP_{i+1} \rangle$ can also be yielded.

2.1.2 *Fan Feature.*

In general, multiple probe paths passing through different routers in an AS will converge to the border router and cross the inter-domain link to enter the neighbor AS. Similarly, multiple probe packets traversing the inter-domain link to reach an AS will be distributed and forwarded to different routers inside the AS. For an IP link $\langle IP_i, IP_{i+1} \rangle$, $IAS(IP_i) \neq IAS(IP_{i+1})$ means an AS switch occurs. Therefore, for an inter-domain link, it may have more previous or subsequent links with AS switch than an intra-domain link. The fan feature captures the likelihood of a certain number of links where an AS switch occurs before (fan-in feature) or after (fan-out feature) a particular link given its link type.

For each IP link $\langle IP_i, IP_{i+1} \rangle$, traverse all traceroute paths including this link, and count the number of its forward links $\langle IP_{i-1}, IP_i \rangle$ (backward links $\langle IP_{i+1}, IP_{i+2} \rangle$) with an AS switch as the **fan-in (fan-out) feature**.

2.1.3 *IP Vector Distance Feature.*

The IP(v4) address is a 32-bit binary number that is typically split into four bytes. The IP vector distance feature is used to indicate that the two IPs at the ends of an inter-domain link typically have a larger Euclidean distance between them than an intradomain link. Thus, for each IP link $\langle IP_i, IP_{i+1} \rangle$, expressing IP_i and IP_{i+1} as vectors, yields $IP_i = [x_1, x_2, x_3, x_4]$ and $IP_{i+1} = [y_1, y_2, y_3, y_4]$, where x_j and y_j denote the jth byte of IP_i and IP_{i+1}, respectively. The **IP vector distance feature** of the IP link $\langle IP_i, IP_{i+1} \rangle$ is calculated as the following equation 1).

$$Euclidean\,(IP_i, IP_{i+1}) = \sqrt{\sum_{j=1}^{4} (x_j - y_j)^2} \tag{1}$$

2.2 IP Link Classification

After building the probabilistic model, we estimate its parameters using the EM algorithm. Specifically, the conditional probability distributions of each feature for different link types are firstly computed, and then the type of each link is updated by Naïve Bayes classifier, after which the feature distributions are recalculated based on new probability values. The above two steps are repeated until convergence.

The probabilistic model requires initial parameters. Before running the algorithm, the IP-to-AS mapping dataset is used to annotate the initial type of each IP link. Therefore, each IP link has an initial type as input to the algorithm. Moreover, to solve the problem of zero probability, the algorithm uses Laplace (Add-1) Smoothing [16]. The following pseudocode introduces the implementation of the algorithm.

Algorithm 1 IP Link Classification Algorithm

```
Input: IP links $\mathbb{L}$; Initial types of IP links $\mathbb{T}$; Feature vector $\mathbb{F}$ ; $\in$; threshold
Output: Inferred type $T_L$ of each link $L$
while $\mathbb{T}$ – last($\mathbb{T}$) >$\in$, do
    for each feature $f$ in feature vector $\mathbb{F}$, do
        $P(f|inter) = \frac{P(f,inter)}{P(inter)} = \frac{N(f,inter)+\alpha}{N(inter)+\alpha d}$
        $P(f|intra) = \frac{P(f,intra)}{P(intra)} = \frac{N(f,intra)+\alpha}{N(intra)+\alpha d}$
    end for
    for each link $L$, do
        $N(All) \leftarrow N(inter) + N(intra)$
        $P(L = inter) \leftarrow P(inter) = \frac{N(inter)}{N(All)}$
        $P(L = intra) \leftarrow P(intra) = \frac{N(intra)}{N(All)}$
        for each feature $f$ in feature vector $\mathbb{F}$ do
            $P(L = inter) = P(L = inter) \times P(f|inter)$
            $P(L = intra) = P(L = intra) \times P(f|intra)$
        end for
        $P(L = inter) \leftarrow \frac{P(L=inter)}{P(L=inter)+P(L=intra)}$
        $P(L = intra) \leftarrow \frac{P(L=intra)}{P(L=inter)+P(L=intra)}$
        if $\frac{P(L=inter)}{P(L=intra)} > threshold$ then
            $T_L = inter$
        else
            $T_L = intra$
        end if
    end for
end while
```

3 EXPERIMENT

To verify the performance of the probabilistic model-based inter-domain link inference method proposed in this paper, we conducted experiments using publicly available topology probe data.

3.1 Experiment Settings

3.1.1 *Experiment Data.*

- Traceroute Dataset

CAIDA deploys and maintains a globally distributed measurement platform Archipelago (Ark), serving the network research community. In team probing, a set of Ark nodes work together as a team to do large-scale Internet topology measurements, using the measurement tool scamper [17] to perform traceroutes. To validate our algorithm, we use traces collected by team1 of CAIDA's Ark infrastructure for the period December 26, 2018 to January 10, 2019 (cycles 7131 to 7159) [18]. The traceroute data contain a total of 295,408,669 paths.

- IP-to-AS Mapping

For IP-to-AS mapping, we use Prefix-to-AS mapping dataset derived from RouteViews [19] data by CAIDA. We supplement this data by the Team Cymru IP2AS mapping tool [20]. For prefixes still not be mapped, publicly available information from the five Regional Internet Registries (RIRs) is used to match IP prefixes with ASes.

- IXP Prefixes List

IXP prefixes are considered specially. We compile a list of IXP prefixes using data from CAIDA [21], which is derived by combining information from PeeringDB, Hurricane Electric, Packet Clearing House (PCH), Wikipedia, BGP Looking Glass, and GeoNames.

- Route Server ASN List

To facilitate peering connectivity, IXPs can provide BGP route servers to establish multilateral peering among IXP member ASes. Route servers have their own AS Numbers (ASNs). We collect a list of route server ASNs from PeeringDB [22] by extracting the ASN of the network whose type is 'Route Server'. We also add Route Server ASNs in Euro-IX [23] to the list.

- AS Relationships Data

We use the dataset of AS relationships inferred by ProbLink [13], which annotates each AS link with a peer-peer (p2p), customer-provider (c2p), or siblings label. CAIDA's AS relationships dataset [24] is used to augment this dataset.

- Verification Datasets

1) IXP validation set

In IXPs that composed of a layer-2 switch device, IXP member ASes use IP addresses assigned by IXP on their router interfaces attached to the IXP switch. Therefore, if there is an IP address IP_i belonging to an IXP prefix in a traceroute path, the link between this IP and its prior hop IP, i.e., $\langle IP_{i-1}, IP_i \rangle$, can be regarded as an inter-domain link. We accordingly generate a validation set containing 103426 such inter-domain links.

2) TeliaSonera validation set

Analyzing DNS hostnames allows us to obtain the interface information which can help determine many aspects (type, location, and so on) of interfaces and their corresponding routers [25]. Therefore, we also perform verification based on DNS hostnames data in CAIDA ITDK 2019-01 [26]. Specifically, we choose interfaces in traces associated with TeliaSonera (AS1299) to generate the verification dataset manually. The reason why we select these ASes is that their DNS hostnames of inter-AS link often can indicate the name of the connected AS [9]. Our TeliaSonera validation dataset contains a total of 26755 IP links, in which 16180 are inter-domain links, 10575 are intra-domain links. The dataset is accurate enough to be used as approximate ground truth [9].

3.1.2 *Evaluation Metrics.*

In this paper, *Precision*, *Recall* and *F1 Score* are evaluation metrics to assess the performance of the algorithm, reflecting its ability to identify inter-domain links.

The inferred results can be classified into four categories as follows.

- True Positive (TP): both the inferred type and the true type are inter-domain links.
- False Positive (FP): the inferred type is inter-domain link, while the true type is intra-domain link.
- True Negative (TN): both the inferred type and the true type are intra-domain links.
- False Negative (FN): the inferred type is intra-domain link, while the true type is inter-domain link.

After counting the numbers of the above four cases respectively, *Precision* can be interpreted as how many of links predicted to be inter-domain links are true inter-domain links, as shown in equation 2).

$$P = \frac{TP}{TP + FP} \tag{2}$$

Recall means how many of true inter-domain links are correctly inferred, as shown in equation 3).

$$R = \frac{TP}{TP + FN} \tag{3}$$

F1 Score is the weighted average of *Precision* and *Recall*, which is shown in equation 4).

$$F1 = \frac{2 \times P \times R}{P + R} \tag{4}$$

3.2 Analysis of the Confidence

The probability that a link belong to the inferred type in the algorithm is considered as its confidence of the inference. This subsection evaluates the accuracy of inter-domain links with different confidence levels by dividing the TeliaSonera validation set into different subsets according to their confidence levels, and then evaluating the effect on the subsets. The results are shown in Table 1.

The result shows that the higher the confidence level is, the higher accuracies of results related usually are. Overall, the results are significantly better when the confidence is greater than 0.9. Therefore, our method makes the results not simply positive or negative, but with a certain degree of confidence. The results with high confidence can be selected for output as needed.

3.3 Comparison with Existing Methods

Here, we compare our method to two techniques that are commonly considered sufficient for the purpose of inter-domain link inference, which are the simple heuristic and the convention heuristic.

The simple heuristic assumes that the first IP address in a different AS is used for the inter-AS link. While the convention heuristic is similar to the simple approach, but incorporates the conventional wisdom that transit links are typically assigned addresses from the provider's address space [15].

The precision, recall and F1 score of the result on IXP validation set and TeliaSonera set are shown in Table 2 and Table 3, respectively.

Table 1: The numbers of links with different confidence levels and their result evaluations

Confidence Level	# Inter-domain Links	Accuracy
0.9~1	12849	91.7%
0.4~0.9	386	81.6%
0.3~0.4	2108	80.8%
0~0.3	1427	77.1%

Table 2: The IXP verification result

Method	TP	FP	Precision
Simple Heuristic	101138	2288	97.8%
Convention Heuristic	101603	1823	98.2%
Our method	101741	1685	98.4%

Table 3: The TeliaSonera verification result

Method	TP	FP	TN	FN	Precision	Recall	F1 Score
Simple Heuristic	0	102	10473	16180	0	0	
Convention Heuristic	11801	2442	8133	4379	82.9%	72.9%	77.6%
Our method	15112	3018	7557	1068	83.4%	93.4%	88.1%

In the IXP verification, our method outperforms these two techniques. In the TeliaSonera verification, since inter-AS links are assigned addresses from a common prefix, the simple heuristic cannot infer correctly, leading to a precision of 0, and severely hurting the recall. As for the convention heuristic, our precision and recall are both higher than it, and the F1 score is greater. Overall, our method performs better.

4 CONCLUSION AND FUTURE WORK

Faced with the problem in practical applications, this paper proposed a probabilistic model-based IP link classification algorithm to achieve inter-domain link inference with confidence. The experimental results showed that the method in this paper had some improvement for the identification of inter-domain links compared with two existing conventional techniques, and the results had confidence.

However, although this paper achieves better results, there are still two shortcomings: 1) only the connection relationships between IP addresses are considered when constructing the network topology, and the path delay information is not taken into account. If the delay information is fully utilized based on the path information, better experimental results may be obtained; 2) the assignments of routers to their operating ASes are not inferred. In future work, further research will be conducted for the above issues.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (U1804263) and Zhongyuan Science and Technology Innovation Leading Talent Project (No. 214200510019).

REFERENCES

[1] J. Hawkinson, and T. Bates. 1996. Guidelines for creation, selection, and registration of an Autonomous System (AS), RFC1930.

[2] Y. Zhang, R. Oliveira, H. Zhang, and L. Zhang. 2010. Quantifying the pitfalls of traceroute in AS connectivity inference. In Proceedings of the 11th International Conference on Passive and Active Measurement (PAM '10). Springer-Verlag, Berlin, Heidelberg, 91–100. https://doi.org/10.1007/978-3-642-12334-4_10

[3] M. Luckie, A. Dhamdhere, B. Huffaker, D. Clark, and K. Claffy. 2016. Bdrmap: inference of borders between IP networks. In Proceedings of the 2016 Internet Measurement Conference (IMC '16). Association for Computing Machinery, New York, NY, USA, 381–396. https://doi.org/10.1145/2987443.2987467

[4] H. Tangmunarunkit, J. Doyle, R. Govindan, W. Willinger, S. Jamin, and S. Shenker. 2001. Does AS size determine degree in AS topology? SIGCOMM Comput. Commun. Rev., 31, 5 (October 2001), 7–8.

[5] H. Tangmunarunkit, R. Govindan, S. Shenker, and D. Estrin. 2001. The impact of routing policy on Internet paths. In Proceedings of IEEE INFOCOM 2001. Conference on Computer Communications. 20th Annual Joint Conference of the IEEE Computer and Communications Society, 2, 736–742. https://doi.org/10.1109/INFCOM.2001.916262

[6] H. Chang, S. Jamin, and W. Willinger. 2001. Inferring AS-level Internet topology from router-level path traces. In Proceedings of Scalability and Traffic Control in IP Networks, S. Fahmy and K. Park, Eds., 4526, SPIE, 196–207.https://doi.org/10.1117/12.434395

[7] B. Huffaker, A. Dhamdhere, M. Fomenkov, and K. Claffy. 2010. Toward topology dualism: improving the accuracy of AS annotations for routers. In Proceedings of the 11th International Conference on Passive and Active Measurement (PAM '10). Springer-Verlag, Berlin, Heidelberg, 101–110. https://doi.org/10.1007/978-3-642-12334-4_11

[8] A. Marder, M. Luckie, A. Dhamdhere, B. Huffaker, K. Claffy, and J. M. Smith. 2018. Pushing the boundaries with bdrmapIT: mapping router ownership at Internet scale. In Proceedings of the Internet Measurement Conference 2018 (IMC '18). Association for Computing Machinery, New York, NY, USA, 56–69. https://doi.org/10.1145/3278532.3278538

[9] A. Marder and J. M. Smith. 2016. MAP-IT: multipass accurate passive inferences from traceroute. In Proceedings of the 2016 Internet Measurement Conference (IMC '16). Association for Computing Machinery, New York, NY, USA, 397–411. https://doi.org/10.1145/2987443.2987468

[10] K. Chhogyal and A. Nayak. 2016. An empirical study of a simple naive bayes classifier based on ranking functions. In Proceedings of AI 2016: Advances in Artificial Intelligence. B. H. Kang and Q. Bai, Eds. Cham: Springer International Publishing, 324–331. https://doi.org/10.1007/978-3-319-50127-7_27

[11] Q. Wang, G. M. Garrity, J. M. Tiedje, and J. R. Cole. 2007. Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. in

Applied and Environmental Microbiology, 73, 16, 5261–5267. https://doi.org/10.1128/AEM.00062-07

[12] H. Zhang. 2004. The optimality of naive bayes. In Proceedings of the 17th International Florida Artificial Intelligence Research Society Conference. FLAIRS 2004, 2, 562-567.

[13] Y. Jin, C. Scott, A. Dhamdhere, V. Giotsas, A. Krishnamurthy, and S. Shenker. 2019. Stable and practical AS relationship inference with ProbLink. In Proceedings of 16th USENIX Symposium on Networked Systems Design and Implementation (NSDI 19). MA: USENIX Association, Boston, 581–598.

[14] A. Retana, D. R. McPherson, R. White, and V. Fuller. 2000. Using 31-bit prefixes on IPv4 point-to-point links, RFC 3021.

[15] K. Claffy, Y. Hyun, K. Keys, M. Fomenkov, and D. Krioukov. 2009. Internet mapping: from art to science. In Proceedings of 2009 Cybersecurity Applications Technology Conference for Homeland Security. 205–211. https://doi.org/10.1109/CATCH.2009.38

[16] C. D. Manning, P. Raghavan, and H. Schütze. 2008. Introduction to Information Retrieval. Cambridge University Press.

[17] M. Luckie. 2010. Scamper: a scalable and extensible packet prober for active measurement of the Internet. In Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement (IMC '10). Association for Computing Machinery, New York, NY, USA, 239–245. https://doi.org/10.1145/1879141.1879171

[18] The CAIDA UCSD IPv4 Routed /24 Topology Dataset - Dec 26, 2018 to Jan 10, 2019. Retrieved May 27, 2020 from http://www.caida.org/data/active/ipv4_routed_24_topology_dataset.xml

[19] University of Oregon Route Views Project. http://www.routeviews.org/

[20] IP-to-ASN Mapping. http://www.team-cymru.org/IP-ASN-mapping.html

[21] The CAIDA UCSD IXPs Dataset. Jan, 2019. Retrieved May 27, 2020 from https://www.caida.org/data/ixps

[22] PeeringDB. https://beta.peeringdb.com/

[23] Euro-IX Route Servers. https://ixpdb.euro-ix.net/en/ixpdb/route-servers/

[24] The CAIDA AS Relationships Dataset. Jan 1, 2019. Retrieved May 27, 2020 from http://www.caida.org/data/active/as-relationships/

[25] J. Chabarek and P. Barford. What's in a name? decoding router interface names. In Proceedings of the 5th ACM Workshop on HotPlanet (HotPlanet '13). Association for Computing Machinery, New York, NY, USA, 3–8. https://doi.org/10.1145/2491159.2491163

[26] The CAIDA UCSD IPv4 Routed /24 DNS Names Dataset. Jan, 2019. Retrieved May 27, 2020 from http://www.caida.org/data/active/ipv4_dnsnames_dataset.xml.