Optimal valid path prediction method for inter-domain networks considering commercial relationships*

Jingye Zhang*, Jun Zhang, Xinli Xiong, Yongjie Wang*

College of Electronic Engineering, National University of Defense Technology, Hefei 230037, China

Anhui Province Key Laboratory of Cyberspace Security Situation Awareness and Evaluation, Hefei 230037, China

azjy@nudt.edu.cn, emlzjy@139.com

Abstract. In the security research of large-scale inter-domain networks, it is often necessary to consider network state changes under the condition of failure or attack. In this case, the connection between ASes often switches between disconnection and recovery frequently, requiring frequent prediction of the AS-level path. Previous AS-level path prediction methods generally require detailed network information data and are complicated to calculate, which is difficult to meet the needs of such large-scale dynamic computation. In this paper, an optimal valid path prediction method for inter-domain networks considering commercial relations is proposed. This method predicts the optimal valid path according to the commercial relationships between ASes, which requires less network information data and low computational complexity. Through experiments, it is found that compared with the shortest path method commonly used in the security simulation of large-scale inter-domain networks, the prediction results are closer to the actual path of the network and can better reflect the actual path situation of the inter-domain network with lower computational complexity.

1. Introduction

In the security research of large-scale inter-domain networks, it is often necessary to consider network state changes under the condition of failure or attack. It is necessary to continuously predict the end-to-end paths at AS-level in order to reasonably present and analyze the dynamic changes of the networks. In this case, the connection between ASes often switches between disconnection and recovery frequently, requiring frequent prediction of AS-level paths.

Most inter-domain routing path prediction methods[1-4] are oriented to network diagnosis, performance optimization, and reliability enhancement. These methods usually need to be based on detailed network information data and have high computational complexity. They are not suitable for dynamic path prediction under large-scale and frequent changes in network connection states. In the security research of large-scale inter-domain networks[5-7], the shortest path method is mostly used to calculate the end-to-end path at AS-level. However, the path calculated by this method is often quite different from the actual path, which is difficult to reflect the essential characteristics of policy routing in inter-domain routing systems. Yang[8] proposes that the calculation of end-to-end paths at AS-level should be close to the actual network path in the security simulation of large-scale inter-domain networks, but no specific path prediction method is given.

In this paper, we propose an optimal valid path prediction method for inter-domain network. This method mainly predicts AS-level path according to the commercial relationships between ASes.
Through experiments, it is found that compared with the shortest path method commonly used in the security simulation of large-scale inter-domain networks, the prediction results are closer to the actual path of the network and can better reflect the actual path situation of the inter-domain routing system with lower computational complexity.

This paper is organized as follows. In Section II, we analyse the commercial relationship in the inter-domain routing system and discuss the valid path and invalid path based on this commercial relationship. In Section III, we propose the method to calculate the optimal valid path. In Section IV, We compare the optimal valid path with the shortest path through experiments. Section V summarizes the critical point of this paper.

2. Valid AS-level path considering commercial relationships

2.1. Commercial relationships in inter-domain routing system

Internet is composed of many service providers’ networks. Internet service providers route traffics on the Internet through formal and informal relationships. These relationships are usually realized in the form of commercial agreements. Gao[9] generalizes four kinds of commercial relationships between autonomous systems: customer-to-provider(C2P), provider-to-customer(P2C), peer-to-peer(P2P) and sibling-to-sibling(S2S). An S2S link connects two autonomous systems with common management boundaries. This connection is usually the result of merging or occurs in some network scenarios. Due to the particular circumstances, this paper does not consider it from the perspective of research universality. The commercial relationships between autonomous systems are shown in Fig.1.

Due to the commercial relationships between autonomous systems, customers need to pay providers to access Internet, while peers are equal and exchanges network traffic free of charge. Not all paths in the network generated by simply taking ASes as nodes and connections between ASes as edges can transmit traffic. The real transmission paths in the inter-domain network are very complex. Each autonomous system has its own routing strategy, which is usually unknown to the outside. However, in general, it can be considered that the import and export rules of BGP are as follows.

(1) Export Rules
   1) Routes learned from customers are declared to all neighbours;
   2) Routes learned from peers and providers are declared to customers only.

(2) Import Rules
   1) It is preferred that the local preference is high, usually in the order of customer > peer > provider from high to low;
   2) Secondly, the length of AS-Path is considered. The smaller the length, the more priority.

2.2. Valid and invalid AS-level path

According to the commercial relationship between ASes, valid AS-level paths can be distinguished from invalid AS-level paths. In the valid path from the source AS to the destination AS, for each
transit AS, there is a customer next to it. In the invalid AS path, at least one transit AS has customer who do not pay it. Fig.2 shows several common AS path modes.

Fig.2 Valid path and invalid path

Fig.2(a) and Fig.2(b) show valid paths. Fig.2(c) and Fig.2(d) show invalid paths. The arrows in Fig.2 only represent the commercial relationships between autonomous systems, and the bold lines represent the paths. In Fig.2(a), A, B and C are transit ASes, and their customers are B and C, D, and E respectively. Each transit AS has customers who pay for it, and the path is valid. Fig.2 (b) differs from Fig.2 (a) in that B to C are p2p links and also valid path. B and C are transit ASes, and their customers are D and E respectively. In Fig.2 (c), B is a transit AS, which forwards traffic for A and F, and there are no customers paying B in this path. In Fig.2 (d), B and C are transit ASes, C has a customer E who pays it, while B does not have a customer who pays it, so the path is invalid.

3. Optimal valid path prediction algorithm

3.1. Algorithm description

Based on the characteristics of the valid path, the optimal valid path needs to meet the following conditions:

1. There is at most one P2P connection in a valid AS path.
2. A P2C link in a valid AS path cannot be followed by a C2P link.
3. A P2C link in a valid AS path cannot be followed by a P2P link.
4. A P2P link in a valid AS path cannot be followed by a C2P link.
5. The optimal valid AS path is the shortest of the currently known valid paths.

According to the above conditions, the optimal valid path prediction algorithm is designed. The algorithm flow is shown in Fig.3.
The algorithm is described as follows:

**Algorithm** Optimal valid path prediction

**Input** Network G, AS A,B

**Output** Path\_A,B

Initialize the Child, Ancestor and Peer collections of each AS node. //Child set is all nodes that it can reach only through p2c link. Ancestor set is all nodes that it can reach only through c2p link. Peer set is all nodes that it can directly reach only through p2p link.

if B in A.Child or B in A.Ancestor or B in A.Peer, Path\_A,B = BFS(G,A,B); //Find the shortest path from A to B through breadth first search

else if B in A.Ancestor.Child, Path\_A,B = BFS(G,A,A.Ancestor) | BFS(G, A.Ancestor,B);

else if B in A.Ancestor.Peer, Path\_A,B = BFS(G,A,A.Ancestor) | A.Ancestor.Peer | BFS(G, A.Ancestor.Peer,B);

else, Path\_A,B = None; //There is no optimal valid path from A to B

The algorithm mainly includes two parts: determining the category of the path between A and B and breadth-first search. In calculating the shortest path between two points in an undirected unweighted graph and the data structure uses an adjacency list, the time complexity of the breadth-first search is $O(N)$. $N$ is the number of ASes in the network. If the hash table is used to determine whether node B is in A.Child, A.Ancestor, and A.Peer, the time complexity is $O(1)$. If node B is in the A.Ancestor.Peer, A.Ancestor.Child or A.Ancestor.Peer.Child set, the time complexity of the judgment is $O(N)$. Therefore, the overall time complexity of the algorithm is $O(N^2)$.

3.2. Evaluation metrics

Longest Common Sub-sequence (LCS) and Length Difference (LD) are used in this paper to evaluate the similarity between two paths. Length Difference (LD) is a simple indicator. In this paper, it refers to the difference in the number of links between two AS-level paths.

Longest Common Sub-sequence (LCS) can be described as: a sequence S, if it is a sub-sequence of two or more known sequences, and it is the longest among all sequences that meet this condition, then S is called the longest common subsequence of the sequence. Let two sequences be defined as follows: $X = (x_1 x_2 \cdots x_m)$ and $Y = (y_1 y_2 \cdots y_n)$. The prefixes of X are $X_{1,2,\cdots,m}$; the prefixes of Y are $Y_{1,2,\cdots,n}$. Let LCS($X_i, Y_j$) represent the set of longest common subsequence of prefixes $X_i$ and $Y_j$. This set of sequences is given by the following.

$$\text{LCS}(X_i, Y_j) = \begin{cases} \emptyset & \text{if } i = 0 \text{ or } j = 0 \\ \text{LCS}(X_{i-1}, Y_{j-1}) \cup x_i & \text{if } i, j > 0 \text{ and } x_i = y_i \\ \max \{\text{LCS}(X_i, Y_{j-1}), \text{LCS}(X_{i-1}, Y_j)\} & \text{if } i, j > 0 \text{ and } x_i \neq y_i \end{cases}$$

In this paper, it is considered that the two paths with longer LCS and shorter LD have better similarity.

4. Experiment and analysis

We use the as-relationships dataset[10] from CAIDA in December 2019 as the basic AS-level network topology, and calculate the optimal valid path based on the commercial relationships in the dataset.
The shortest path is used for comparison with the optimal valid path. The shortest path algorithm is a common method for calculating AS-level paths in many large-scale inter-domain network security research. The BGP routing information database in the Route Views project[11] at the same time was used to calculate the actual network path. Longest Common Sub-sequence (LCS) and Length Difference (LD) are used to measure the closeness of the optimal valid path and the shortest path to the actual path. Hereinafter, \(LCS(ovp, rv)\) represents the longest common sub-sequence length of the optimal valid path and the actual path. \(LCS(s, rv)\) represents the longest common sub-sequence length of the shortest path and the actual path. \(LD(ovp, rv)\) represents the length difference between the optimal valid path and the actual path. \(LD(s, rv)\) represents the length difference between the shortest path and the actual path. The experimental comparison method is shown in Fig.4.

![Fig.4 Comparison of the paths](image)

Each time, 50, 100, 500, 1000, 2500 pairs of AS nodes are randomly selected in the as-relationships dataset[10] to calculate the optimal valid path and the shortest path. Then the \(LCS(ovp, rv)\), \(LCS(s, rv)\), \(LD(ovp, rv)\), \(LD(s, rv)\) of the paths are calculated and compared.

The experiment was repeated 10 times, and the average value of the comparison results is shown in Table 1. Column 1 in Table 1 represents the number of randomly selected ASes pairs, that is, the number of paths to compare. Column 2 represents the average number of paths that satisfy \(LCS(ovp, rv) >= LCS(s, rv)\). Column 3 represents the average number of paths satisfying \(LD(ovp, rv) <= LD(s, rv)\).

| Number of ASes pairs | LCS(ovp,rv) >= LCS(s,rv) (avg) | LD(ovp,rv) <= LD(s,rv) (avg) |
|----------------------|-------------------------------|-------------------------------|
| 50                   | 44                            | 32                            |
| 100                  | 82                            | 62                            |
| 500                  | 409                           | 320                           |
| 1000                 | 815                           | 624                           |
| 2500                 | 2045                          | 1619                          |

As can be seen from the results in Table 1, the optimal valid path is better than the shortest path in terms of both the longest common sub-sequence and the length difference, which shows that the optimal valid path is closer to the actual routing path.
Figure 5 shows the comparison results for each of the 10 repeated experiments. Figure 5 (a), (b), (c) show the comparison results when 100, 500, and 1000 AS node pairs are randomly selected. It can be seen that in different rounds of experiments, the comparison results of the optimal valid path relative to the shortest path are stable, and there will be no obvious fluctuation due to the random selection of ASes in different network locations. The algorithm has good applicability for path prediction between ASes in various network locations.

5. Conclusion
In the security research of large-scale inter-domain networks, simulation is often used to present the state of the network when it fails or is attacked. It is necessary to frequently predict the paths at AS level in order to show and analyze the dynamic changes of the network reasonably. Previous AS-level path prediction methods generally require detailed network information data and are complicated to calculate, which is difficult to meet the needs of such large-scale dynamic calculations. This paper proposes an optimal valid path prediction method for inter-domain network considering commercial relations. This method can predict the valid path at AS-level according to ASes connection data with commercial relationships. As it needs less network information data and is low computational complexity, it can be applied to AS-level path prediction under large-scale and high dynamic demand. Through experiments, it is found that compared with the shortest path method commonly used in security simulation of large-scale inter-domain networks, the prediction result is closer to the actual path of the network. Besides, the experimental results show that selecting ASes at different positions...
of the network as the starting and ending points for prediction has little influence on the results. The algorithm has good applicability for path prediction between ASes in various network locations.

References

[1] Su S, Fang B. Towards route dynamics in as-level path prediction[J]. IEICE Transactions on Communications, 2016, E99B(2): 412–421.

[2] Mühlbauer W, Feldmann A, Maennel O, et al. Building an AS-topology model that captures route diversity[J]. Computer Communication Review, 2006, 36(4): 195–206.

[3] Mu X Y, Chen Y X, Deng Y H, et al. A novel algorithm for AS path inference based on BGP routing tables[J]. Proceedings of 2013 3rd International Conference on Computer Science and Network Technology, ICCSNT 2013, IEEE, 2014: 196–199.

[4] Li X, Cai Z, Hou B, et al. ProblInfer: Probability-based AS path inference from multigraph perspective[J]. Computer Networks, Elsevier B.V., 2020, 180(February): 107377.

[5] Guo Y, Wang Z, Luo S, et al. A cascading failure model for interdomain routing system[J]. International Journal of Communication Systems, 2012, 25(8): 1068–1076.

[6] Schuchard M, Mohaisen A, Foo Kune D, et al. Losing control of the Internet: Using the data plane to attack the control plane[A]. Proceedings of the ACM Conference on Computer and Communications Security[C]. 2010: 726–728.

[7] Liu Y, Peng W, Su J, et al. Assessing survivability of inter-domain routing system under cascading failures[A]. Communications in Computer and Information Science[C]. Zhangjiajie, China: 2013, 401: 97–108.

[8] Yang B, Zhang Y, Lu Y. A new methods for cascading failures analysis in inter-domain routing system[J]. Proceedings - 5th International Conference on Instrumentation and Measurement, Computer, Communication, and Control, IMCCC 2015, IEEE, 2016: 382–385.

[9] Gao L, Rexford J. Stable Internet routing without global coordination[J]. IEEE/ACM Transactions on Networking, IEEE, 2001, 9(6): 681–692.

[10] As-relationships datasets[EB/OL]. [2020–07–31]. http://data.caida.org/datasets/as-relationships/.

[11] Route Views Archive Project Page[EB/OL]. [2020–07–29]. http://routeviews.org/.