Probability-Based Routing Symmetry Metrics

Qin Wang[1], Fang Dong[2(✉)], Xin-Li Yang[3], and Rui Yin[2]

[1] China University Program, Texas Instruments Semiconductor Technologies (Shanghai), Shanghai, China
qin-wang@ti.com

[2] College of Information and Electronic Engineering, Zhejiang University City College, Hangzhou, China
{dongf, yinrui}@zucc.edu.cn

[3] Port Management Office of Haiyan County, Jiaxing, Zhejiang Province, China
hyhangyun@163.com

Abstract. In communication networks, if streams between two endpoints follow the same physical paths for both forward and reverse direction, they are symmetric. Routing asymmetry affects several protocols, and impacts part of traffic analysis techniques. We propose two routing symmetry metrics to express different meanings when talking about routing symmetry, namely, (1) the forward and reverse flows coming from one node to another are exactly the same, and (2) one single node is visited by both flows. The two metrics are termed as identity symmetry and cross symmetry, respectively. Then, we build a model to link the macroscopic symmetry with the microscopic routing behavior, and present some analysis results, thus make it possible to design a routing algorithm with some desired symmetry. The simulation and dataset study show that routing algorithms that generate next hop randomly will lead to a symmetric network, but it is not the case for Internet. Because the paths of Internet are heavily dominated by a small number of prevalent routes, Internet is highly asymmetry.

Keywords: Routing symmetry · Routing behavior model · Statistical process

1 Introduction

In communication networks, if streams between two endpoints follow the same physical paths for both forward and reverse direction, they are symmetric [1]. Routing asymmetry affects several protocols and impacts traffic analysis techniques. Knowing to which degree the routings are symmetric is helpful in protocol design and traffic analysis.

In practice, the one-way propagation time is commonly estimated to be half of the round-trip time (RTT) between nodes, e.g., the NTP (Network Time Protocol) of Internet [2, 3]. However, this estimate will be inappropriate if routes are asymmetric.

The paper is supported by Basic Public Welfare Research Projects (No. LGF18F010007) of Zhejiang Province and National Natural Science Foundation Program of China (No. 61771429).

L. Meng and Y. Zhang (Eds.): MLICOM 2018, LNICST 251, pp. 370–380, 2018.
https://doi.org/10.1007/978-3-030-00557-3_37

Today's communication protocols rely heavily on the estimate of link condition to exploit available communication resources effectively, which is especially the case in wireless communications. The estimate is usually based on measurement of the statistical parameters of incoming packet, which will not infer the real condition of the outgoing link in situations of routing asymmetry.

Some traffic analysis techniques [4–6] are embedded in an assumption that routings are symmetric, i.e., all the packets of a session on both directions can be monitored by a sniffer located on a specific link. But it is not the case in practice [7, 8]. Routing asymmetry has a significant impact of these techniques [9, 10].

A common cause of routing asymmetry is that routing is selected independently for each flow, and at each node, taking many factors into account, including load-balancing and congestion controlling, which varied among nodes. This cause is especially prominent in the case of multipath routing. Another commonly mentioned cause is the "hot-potato routing", which is a business practice of passing traffic off to another autonomous system (AS) as soon as possible. By autonomous system, we mean a domain in which the routers and hosts are unified by a single administrative authority, and a set of interior gateway protocols [11].

2 Related Works

Literatures contain many studies of routing protocols, but considerably few studies of routing behavior [2, 3]. But recently there is a growing research of macroscopic properties of the network routing, including routing asymmetry, by studying datasets or modeling network behavior.

It is more than ten years since Paxson revealed that about 50% of the time an Internet path includes a major asymmetry [12]. In the past few years, it became clearer that this phenomenon has a significant impact on network measuring, modeling, and managing. [9] studied impact that asymmetric routing can have on statistical traffic classifiers. [8] pointed out that over 60% of AS-level paths are asymmetric, and path asymmetry will increasingly spread in the future. [12] studied the path stability and symmetry in 6 levels of granularity: router, point of presence (PoP), address prefix (AP), autonomous systems (AS), city and country. [1] used passively captured network data to estimate the amount of routing symmetry on a specific link, and [5, 13, 14] made an impractical assumption of traffic symmetry in tools and analysis.

Most work quantified the asymmetry with the number of different nodes between the forward and reverse paths, or classified a path as either asymmetric or symmetric, without considering quantifying the degree of symmetry [2, 3]. [7] proposed an approach to quantify the magnitude of routing asymmetry, measuring the dissimilarity between a pair of routes by aligning the two routes together and counting the minimal total cost incurred in aligning them. [15] defined the similarity coefficient as the number of similar nodes divided by the total number of distinct nodes in the two paths. [12, 16] quantified the difference between two routes (at any level) by calculating their Edit Distance [17] value. The metric defined in [16, 17] are different because the only operation considered in the former was aligning, while the latter considered inserting, deleting, and modifying.

These metrics of symmetry/asymmetry are not suited for modeling the behavior of routing theoretically, since they can only be calculated by algorithm or program. Some metrics [15, 16] does not meet the reality, because routing symmetry means different things under different contexts: (1) when estimating path condition, it means whether or not the forward and reverse path coming from a specific node to another specific node are exactly the same; (2) when talking a sniffer can or cannot monitor flows from both directions, it means whether or not a specific node is visited by both flows. So, accordingly, we will define 2 metrics, which will be called identity routing symmetry and cross routing symmetry.

3 Modeling and Statistical Analysis

3.1 Routing Symmetry Metrics

In this paper, when $\boldsymbol{A}$ and $\boldsymbol{B}$ be $m \times n$ matrices, the element-wise product of $\boldsymbol{A}$ and $\boldsymbol{B}$ is defined by $[\boldsymbol{A} \circ \boldsymbol{B}]_{i,j} = [\boldsymbol{A}]_{i,j}[\boldsymbol{B}]_{i,j}$, for all $1 \le i \le m, 1 \le j \le n$.

In a connected network, there are an infinite number of paths from any node s to any other node d, i.e. $path_1, path_2, path_3, \ldots$ when counting circles. Suppose the frequency that the source node s selects these paths to route packets to be $p_1, p_2, p_3, \ldots$. Similarly, there are also an infinite number of paths from node d to $s, path_1^{-1}$, $path_2^{-1}, path_3^{-1}, \ldots$, with $path_i^{-1}$ we mean the reverse path of $path_i$, and the corresponding selecting frequency $q_1, q_2, q_3, \ldots$. Informally, we will use $\boldsymbol{p}$ to denote the vector $[p_1, p_2, p_3, \ldots]$ and $\boldsymbol{q}$ to denote $[q_1, q_2, q_3, \ldots]$.

A. Identity routing symmetry

We use the normalized inner product of $\boldsymbol{p}$ and $\boldsymbol{q}$ to define identity routing symmetry:

$$\rho^{id}(s,d) = \frac{(\boldsymbol{p},\boldsymbol{q})}{\sqrt{(\boldsymbol{p},\boldsymbol{p})}\sqrt{(\boldsymbol{q},\boldsymbol{q})}}. \tag{1}$$

In algebra, ρ^{id} is also viewed as the cosine of the angle between $\boldsymbol{p}$ and $\boldsymbol{q}$. It is varied in the range [0, 1]. When ρ^{id} is close to 0, $\boldsymbol{p}$ and $\boldsymbol{q}$ are orthogonal to each other. So, if for some $path_i$, the frequency that it is selected as the forward path, namely p_i, is large, then the frequency that $path_i^{-1}$ is selected as the reverse path, q_i, must be small, otherwise ρ^{id} will not be close to 0. Conversely, when ρ^{id} is close to 1, $\boldsymbol{p}$ and $\boldsymbol{q}$ are parallel to one another, so for each i, p_i and q_i are both large or both small. This makes ρ^{id} a good choice for defining our identity routing symmetry. Note that the defined ρ^{id} is not a linear function of the angle between $\boldsymbol{p}$ and $\boldsymbol{q}$, so we may use $1 - arccos(\rho^{id})$ to calculate the identity symmetry. As this is an increasing function of ρ^{id}, they are essentially the same metric.

In practical networks, usually a small number of paths are used to transfer packet stream. In such cases, $\boldsymbol{p}$ and $\boldsymbol{q}$ are sparse vectors.

B. Cross routing symmetry

Cross routing symmetry is defined by the probability that a specific node is visited by the forward flow and the reverse flow:

$$\rho^{cross}(v, s, d) = \mathrm{P}\{v \in path_i \text{ and } v \in path_j^{-1}, \text{for any } i, j\}, \tag{2}$$

where v is different from s and d.

3.2 Modeling

The model links the macroscopic symmetry with the microscopic routing behavior, thus make it possible to design a routing algorithm with a desired symmetry.

The routing selection is modeled as a Markov Chain. The *routing probability* from node i to j is the probability that node i select a neighboring node j as the next hop to route data packets. Figure 1 gives a further demonstration of routing probability. The probability of data packet routed from node i to j in one hop is denoted by p_{ij}. The routing probability matrix, or *routing matrix* for short, is given by using p_{ij} as the i-th row and j-th column element. The assumptions are:

(a) Routing probability is time-invariant;
(b) Routing probability is independent of the source node, but depends on the destination node.

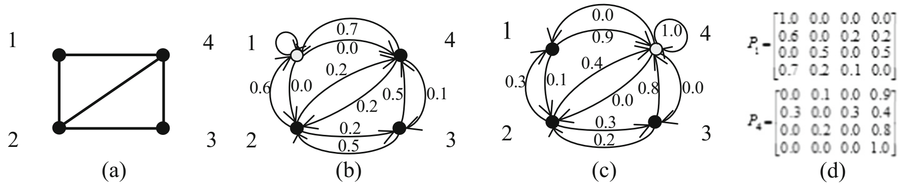

Fig. 1. Explanation of routing matrix (a) A network with 4 nodes; (b) Packets are destined to node 1. Each routing probability is labeled on a directed link from a source node to a destination node; (c) When data packets are destined to node 4, the routing probabilities are different from (b); (d) The routing matrices of (b) and (c).

Without the second assumption, the model is identical to the *random walk* model, which is a well-known routing behavior model. Actually, this model is an extension of random walk, so assumption (2) is not a restriction but rather a generalization. The introduction of assumption (2) make the model more realistic, as in communication networks, many routing protocols are designed to behave destination dependent. Thus the subscript to specify the destination node is used as $\boldsymbol{P}_1$ and $\boldsymbol{P}_4$ in Fig. 1(d). Some notes of $\boldsymbol{P}_d$ are:

(a) Node d is a destination node, and thus d is an absorbing state in Markov Chain, so $[\boldsymbol{P}_d]_{d,d}= 1$, and $[\boldsymbol{P}_d]_{d,j}= 0, j \neq d$;
(b) For any i, $\sum_j [\boldsymbol{P}_d]_{i,j}= 1$.

A. Identity routing symmetry

Let nodes s and d be any two different nodes of a connected network. There is only one possible path from s to d in the network with length 1, which is s,d. This path will be chosen to transfer packets with probability $[\boldsymbol{P}_d]_{s,d}$ (if path s,d does not exist, $[\boldsymbol{P}_d]_{s,d}$ will be zero). There are (N-2) possible paths in the network with length 2, chosen with probability $[\boldsymbol{P}_d]_{s,h_1}[\boldsymbol{P}_d]_{h_1,d}, h_1 = 1, 2, \ldots, N$ and $h_1 \neq s, d$. There are (N-2)^2 paths in the network with length 2, chosen with probability $[\boldsymbol{P}_d]_{s,h_1}[\boldsymbol{P}_d]_{h_1,h_2}[\boldsymbol{P}_d]_{h_2,d}, h_1, h_2 = 1, 2, \ldots, N$ and $h_1, h_2 \neq s, d$. And the rest can be deduced by analogy. So the inner product of previously mentioned vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ is

$$\begin{aligned}(\boldsymbol{p}, \boldsymbol{q}) = [\boldsymbol{P}_d]_{s,d} \cdot [\boldsymbol{P}_s]_{d,s} + \sum_{h_1 \neq s,d} [\boldsymbol{P}_d]_{s,h_1}[\boldsymbol{P}_d]_{h_1,d} \cdot [\boldsymbol{P}_s]_{d,h_1}[\boldsymbol{P}_s]_{h_1,s} + \\ \sum_{h_1,h_2 \neq s,d} [\boldsymbol{P}_d]_{s,h_1}[\boldsymbol{P}_d]_{h_1,h_2}[\boldsymbol{P}_d]_{h_2,d} \cdot [\boldsymbol{P}_s]_{d,h_2}[\boldsymbol{P}_s]_{h_2,h_1}[\boldsymbol{P}_s]_{h_1,s} + \cdots .\end{aligned} \tag{3}$$

All p_i's are not larger than 1 and the sum of all q_i's is 1, so the partial sum of the RHS (right-hand side) of (3) is less than 1, and the sequence of partial sums are incremental. So the RHS of (3) is convergence.

Use the definition of element-wise product operator '$\circ$', Eq. (3) can be written in another form:

$$(\boldsymbol{p}, \boldsymbol{q}) = [\sum_{i=1}^{\infty} (\boldsymbol{P}_d^{(s)} \circ (\boldsymbol{P}_s^{(d)})^T)^i]_{s,d}. \tag{4}$$

Matrix $\boldsymbol{P}_d^{(s)}$ is $\boldsymbol{P}_d$ with s-th column replaced by a zero vector, and similarly, $\boldsymbol{P}_s^{(d)}$ is $\boldsymbol{P}_s$ with d-th column replaced by zero vector.

[10] in appendix makes it possible to write (4) with a closed form. The only requirement is $\boldsymbol{I} - \boldsymbol{P}_d^{(s)} \circ \boldsymbol{P}_s^{(d)}$ to be invertible. Suppose this requirement is satisfied, then

$$(\boldsymbol{p}, \boldsymbol{q}) = [\lim_{n \to \infty}((\boldsymbol{I} - \boldsymbol{P}_d^{(s)} \circ (\boldsymbol{P}_s^{(d)})^T)^{-1}(\boldsymbol{P}_d^{(s)} \circ (\boldsymbol{P}_s^{(d)})^T - (\boldsymbol{P}_d^{(s)} \circ (\boldsymbol{P}_s^{(d)})^T)^{n+1}))]_{s,d}. \tag{5}$$

Because the d-th column and d-th row of $(\boldsymbol{P}_d^{(s)} \circ (\boldsymbol{P}_s^{(d)})^T)$ are all zeros, so for any n, the term $(\boldsymbol{P}_d^{(s)} \circ (\boldsymbol{P}_s^{(d)})^T)^{n+1}$ will have d-th column be zeros. So,

$$(\boldsymbol{p}, \boldsymbol{q}) = [(\boldsymbol{I} - \boldsymbol{P}_d^{(s)} \circ (\boldsymbol{P}_s^{(d)})^T)^{-1}(\boldsymbol{P}_d^{(s)} \circ (\boldsymbol{P}_s^{(d)})^T)]_{s,d}. \tag{6}$$

With some similar but simpler steps, we get

$$(\boldsymbol{p},\boldsymbol{p}) = [\sum_{n=0}^{\infty}(\boldsymbol{P}_d^{(d)} \circ \boldsymbol{P}_d^{(d)})^n(\boldsymbol{P}_d \circ \boldsymbol{P}_d)]_{s,d} = [(\boldsymbol{I} - \boldsymbol{P}_d^{(d)} \circ \boldsymbol{P}_d^{(d)})^{-1}(\boldsymbol{P}_d \circ \boldsymbol{P}_d)]_{s,d}, \quad (7)$$

$$(\boldsymbol{q},\boldsymbol{q}) = [\sum_{n=0}^{\infty}(\boldsymbol{P}_s^{(s)} \circ \boldsymbol{P}_s^{(s)})^n(\boldsymbol{P}_s \circ \boldsymbol{P}_s)]_{d,s} = [(\boldsymbol{I} - \boldsymbol{P}_s^{(s)} \circ \boldsymbol{P}_s^{(s)})^{-1}(\boldsymbol{P}_s \circ \boldsymbol{P}_s)]_{d,s}. \quad (8)$$

Finally,

$$\rho^{id} = \frac{[(\boldsymbol{I} - \boldsymbol{P}_d^{(s)} \circ (\boldsymbol{P}_s^{(d)})^T)^{-1}(\boldsymbol{P}_d^{(s)} \circ (\boldsymbol{P}_s^{(d)})^T)]_{s,d}}{\sqrt{[(\boldsymbol{I} - \boldsymbol{P}_d^{(d)} \circ \boldsymbol{P}_d^{(d)})^{-1}(\boldsymbol{P}_d \circ \boldsymbol{P}_d)]_{s,d}}\sqrt{[(\boldsymbol{I} - \boldsymbol{P}_s^{(s)} \circ \boldsymbol{P}_s^{(s)})^{-1}(\boldsymbol{P}_s \circ \boldsymbol{P}_s)]_{d,s}}}. \quad (9)$$

B. Cross routing symmetry

With the aforementioned assumptions, the forward routing process is independent of the reverse routing process. So, the cross routing symmetry is

$$\begin{aligned}\rho^{cross}(v,s,d) &= \mathrm{P}\{v \in path_i, \text{for any } i\}\mathrm{P}\{v \in path_j^{-1}, \text{for any } j\} \\ &= (1 - \mathrm{P}\{v \notin path_i, \text{for all } i\})(1 - \mathrm{P}\{v \notin path_j^{-1}, \text{for all } j\}).\end{aligned} \quad (10)$$

Similar to the derivation of previous section, there is only one possible path in the network with length 1, which is s,d, chosen with probability $[\boldsymbol{P}_d]_{s,d}$ (if path s,d does not exist, $[\boldsymbol{P}_d]_{s,d}$ will be zero). Flow that follows this path definitely will not visit node v (which will be called "miss v" in the following). There are N-2 possible paths in the network with length 2. Flow that follows these paths will miss node v with a probability $[\boldsymbol{P}_d]_{s,h_1}[\boldsymbol{P}_d]_{h_1,d}$ respectively, $h_1 = 1,2,\ldots,N$ and $h_1 \neq s,v,d$. There are $(N\text{-}2)^2$ possible paths in the network with length 2. Flow that follows these paths will miss node v with a probability $[\boldsymbol{P}_d]_{s,h_1}[\boldsymbol{P}_d]_{h_1,h_2}[\boldsymbol{P}_d]_{h_2,d}$ respectively, $h_1,h_2 = 1,2,\ldots,N$ and $h_1,h_2 \neq s,v,d$. And the rest can be deduced by analogy. So ρ^{cross} is

$$\begin{aligned}\rho^{cross}(v,s,d) &= (1 - \mathrm{P}\{v \notin path_i, \text{for all } i\})(1 - \mathrm{P}\{v \notin path_j^{-1}, \text{for all } j\}) \\ &= (1 - [\boldsymbol{P}_d]_{s,d} - \sum_{h_1 \neq s,v,d}[\boldsymbol{P}_d]_{s,h_1}[\boldsymbol{P}_d]_{h_1,d} - \sum_{h_1,h_2 \neq s,v,d}[\boldsymbol{P}_d]_{s,h_1}[\mathbf{P}_d]_{h_1,h_2}[\boldsymbol{P}_d]_{h_2,d} - \cdots)\cdot \\ &\quad (1 - [\boldsymbol{P}_s]_{d,s} - \sum_{h_1 \neq s,v,d}[\boldsymbol{P}_s]_{d,h_1}[\boldsymbol{P}_s]_{h_1,s} - \sum_{h_1,h_2 \neq s,v,d}[\boldsymbol{P}_s]_{d,h_1}[\boldsymbol{P}_s]_{h_1,h_2}[\boldsymbol{P}_s]_{h_2,s} - \cdots) \\ &= (1 - [\sum_{n=0}^{\infty}(\boldsymbol{P}_d^{(s,v,d)})^n\boldsymbol{P}_d]_{s,d})(1 - [\sum_{n=0}^{\infty}(\boldsymbol{P}_s^{(s,v,d)})^n\boldsymbol{P}_s]_{d,s}).\end{aligned} \quad (11)$$

Matrix $\boldsymbol{P}_d^{(s,v,d)}$ is $\boldsymbol{P}_d$ with s-th column, v-th column and d-th column replaced by zero vectors, while matrix $\boldsymbol{P}_s^{(s,v,d)}$ is $\boldsymbol{P}_s$ with s-th column, v-th column and d-th column replaced by zero vectors. Suppose matrix $\left(\boldsymbol{I}-\boldsymbol{P}_d^{(s,v,d)}\right)$ and $\left(\boldsymbol{I}-\boldsymbol{P}_s^{(s,v,d)}\right)$ to be invertible,

$$\begin{aligned}\rho^{cross}(v,s,d) &= (1-[\lim_{n\to\infty}(\boldsymbol{I}-\boldsymbol{P}_d^{(s,v,d)})^{-1}(\boldsymbol{I}-(\boldsymbol{P}_d^{(s,v,d)})^n)\boldsymbol{P}_d]_{s,d})\cdot \\ &(1-[\lim_{n\to\infty}(\boldsymbol{I}-\boldsymbol{P}_s^{(s,v,d)})^{-1}(\boldsymbol{I}-(\boldsymbol{P}_s^{(s,v,d)})^n)\boldsymbol{P}_s]_{d,s})\end{aligned} \tag{12}$$

Finally, we get

$$\rho^{cross}(v,s,d) = (1-[(\boldsymbol{I}-\boldsymbol{P}_d^{(s,v,d)})^{-1}\boldsymbol{P}_d]_{s,d})(1-[(\boldsymbol{I}-\boldsymbol{P}_s^{(s,v,d)})^{-1}\boldsymbol{P}_s]_{d,s}). \tag{13}$$

4 Evaluation and Analysis

4.1 Evaluation of Random Walk Based Routing

According to their topology, networks are usually classified into random networks, regular networks, small world networks and scale free networks. To avoid bias introduced by topology, three typical networks (random network, WS [18] network and BA [19] network) are considered. In the simulation, each of these 3 networks is composed of 128 nodes, thus there will be (128 × 127)/2 = 8128 different pairs of nodes to be considered when evaluating identity symmetry. Cross symmetry of all intermediate nodes of two fixed nodes is also evaluated. Three different random walk-based routing algorithms are evaluated. There are some literatures focusing on random walk-based routings in practical networks [20, 21].

The routing probability from node i to j of these routing algorithms is

$$p_{ij} = \frac{d_j^\alpha}{\sum\limits_{k\in N(i)} d_k^\alpha}, \tag{14}$$

but taking different values of parameter α, namely, −1, 0, and 1, respectively. Notation d_j is the degree of node j, and $N(i)$ the set of all neighboring nodes of node i.

Results are shown in Figs. 2 and 3.

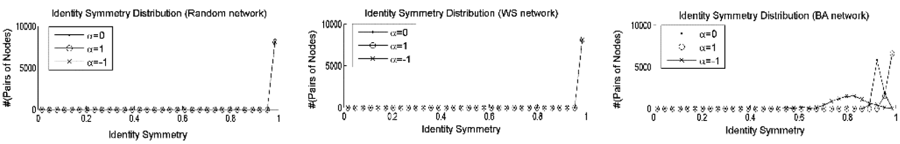

Fig. 2. Identity symmetry distribution

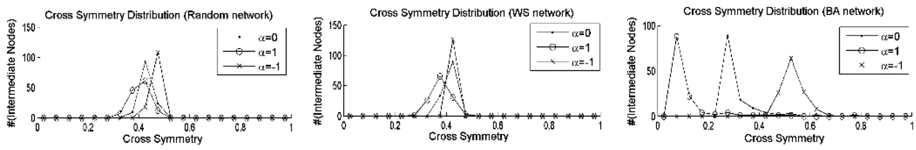

Fig. 3. Cross symmetry distribution

In Fig. 2, most node pairs' identity symmetry metrics are close to 1 in random network and WS network, regardless of which routing algorithm is applied. This result is consistent with assumptions that networks are symmetric. The reason for the first two routings are symmetric is explained by an example. Consider a path s, i, j, d, the probability that it is selected as a forward route is

$$p_f = \frac{d_i^\alpha}{\sum\limits_{k \in N(s)} d_k^\alpha} \frac{d_j^\alpha}{\sum\limits_{k \in N(i)} d_k^\alpha} \frac{d_d^\alpha}{\sum\limits_{k \in N(j)} d_k^\alpha}, \tag{15}$$

and correspondingly, the probability that path d, j, i, s is selected as a reverse route is

$$p_r = \frac{d_j^\alpha}{\sum\limits_{k \in N(d)} d_k^\alpha} \frac{d_i^\alpha}{\sum\limits_{k \in N(j)} d_k^\alpha} \frac{d_s^\alpha}{\sum\limits_{k \in N(i)} d_k^\alpha}. \tag{16}$$

With a little more effort we can calculate the ratio of p_f to p_r. In random networks or small world networks, nodes are of similar degrees, so the ratio will be approximately 1. Thus, we can see the angle between vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ will be very small, thus the identity symmetry is close to 1. But in scale-free network, degrees of nodes are varied significantly, thus the identity symmetry metrics are scattered.

Figure 3 shows that no matter which topology is, the routing algorithm with $\alpha = 1$ will be of lowest cross symmetry, the difference of the three routing algorithms is especially significant in scale free networks. According to the motivation that this metric is present, a conclusion can be drawn that routing algorithm with $\alpha = 1$ will be the safest of the three, especially in a scale-free network.

Many researchers have found that a large number of networks, including Internet, have scale free property [22]. Figures 2 and 3 shows that scale free networks are not symmetric under either definition of symmetry.

4.2 Dataset Evaluation

We do not have sufficient evidence to conclude that some networks are asymmetric without studying some widely deployed networks. In the rest of this section, a dataset study of Internet is presented. The data is measured by [2, 3] (only the second set of measurements, which is termed D2 there, is suitable and thus used in this paper for symmetry analysis), using a computer program called *trace route*, which can display the route path. The measurement is conducted on Internet, including nodes (computers) from one hundred or so cities of different countries. To reduce complexity and make it tractable for symmetry metric calculation, we choose the method used by [2, 3], abstracting computers from the same city with a single node, thus constructing a new network with a relatively small number of nodes. The constructed network contains 101 nodes, with 27 nodes have route to them. The identity symmetry metrics are evaluated of any possible pairs of these 27 nodes.

To show the difference of the measured data and the data generated by random walk, we use an algorithm, similar to the algorithm of random redistribution of link weights [22], to randomize the originally data. First, for each destination d, by analyzing the data, we get the routing probability matrix $\boldsymbol{P}_d$. Then, for each row of $\boldsymbol{P}_d$, the non-zero entries are divided into a smaller unit Δ. Each unit is extracted randomly with probability p, unless it is the last unit of this entry. Lastly, we equiprobably lay back each extracted unit to all non-zero entries in the same row. The parameter p controls the degree to which the routing probability is randomized, without changing the topology of the network.

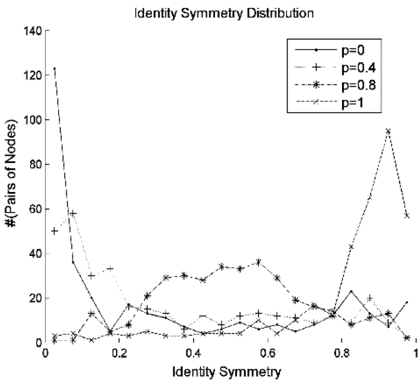

Fig. 4. Identity symmetry distribution of internet

The result is shown in Fig. 4. The curve corresponds to the original data ($p = 0$) shows that routing in Internet is highly asymmetric, with almost all the pairs' identity symmetry metrics centered in the range (0, 0.2). While p increases, the metrics are gradually moves to 1. When $p = 1$, similar to the result of the previous simulation (Fig. 2), most node pairs' identity symmetry metrics are close to 1. This implies that routing algorithms that generate next hop randomly will lead to a symmetric network, and that Internet does not work in this way. The routing is not random but rather specialized, consistent with [2], which shows that Internet paths are heavily dominated by a small number of prevalent routes.

5 Conclusions

In this work, we propose two routing symmetry metrics to express different meanings when talking about routing symmetry, namely, (1) the forward and reverse flows coming from one node to another are exactly the same, and (2) one node is visited by both flows. Then, we build a model to link the macroscopic symmetry with the microscopic routing behavior, thus make it possible to design a routing algorithm with a desired symmetry. The simulation and dataset study shows that routing algorithms that generate next hop randomly will lead to a symmetric network, but Internet does not work in this way, because the paths of which are heavily dominated by a small number of prevalent routes, it is highly asymmetry.

References

1. John, W., Dusi, M., Claffy, K.: Estimating routing symmetry on single links by passive flow measurements, pp. 473–478. ACM (2010)
2. Paxson, V.: End-to-end routing behavior in the internet. IEEE/ACM Trans. Netw. **5**(5), 601–615 (1997). https://doi.org/10.1109/90.649563
3. Alderson, D., Chang, H., Roughan, M., Uhlig, S., Willinger, W.: The many facets of internet topology and traffic. Netw. Heterog. Media **1**(4), 569–600 (2017)
4. Nguyen, T., Armitage, G.: A survey of techniques for internet traffic classification using machine learning. IEEE Commun. Surv. Tutor. **10**(4), 56–76 (2009)
5. McGregor, A., Hall, M., Lorier, P., Brunskill, J.: Flow clustering using machine learning techniques. In: Barakat, C., Pratt, I. (eds.) PAM 2004. LNCS, vol. 3015, pp. 205–214. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-24668-8_21
6. Fahad, A., Alshatri, N., Tari, Z., Alamri, A.: A survey of clustering algorithms for big data: taxonomy and empirical analysis. IEEE Trans. Emerg. Top. Comput. **2**(3), 267–279 (2014)
7. He, Y., Faloutsos, M., Krishnamurthy, S.: Quantifying routing asymmetry in the internet at the as level, pp. 1474–1479. IEEE(2004)
8. Mao, Z.M., Qiu, L., Wang, J., Zhang, Y.: On as-level path inference, pp. 339–349. ACM (2005)
9. Crotti, M., Gringoli, F., Salgarelli, L.: Impact of asymmetric routing on statistical traffic classification, pp. 1–8. IEEE (2009)
10. Dong, F., Liu, J., Dai, S.: Identity routing symmetry metrics for routing behavior. In: Atlantis Conference, pp. 1853–1856 (2016)
11. Tozal, M.: Autonomous system ranking by topological characteristics: a comparative study. In: Systems Conference, pp. 1–8 (2017)
12. Weinsberg, U., Shavitt, Y., Schwartz, Y.: Stability and symmetry of internet routing, pp. 1–2. IEEE (2009)
13. Keralapura, R., Mellia, M., Grimaudo, L.: Self-learning classifier for internet traffic. US 8694630 B1. IEEE (2014)
14. Zhang, J., Chen, X., Xiang, Y., Zhou, W., Wu, J.: Robust network traffic classification. IEEE/ACM Trans. Netw. **23**(4), 1257–1270 (2015)
15. Pucha, H., Zhang, Y., Mao, Z.M., Hu, Y.C.: Understanding network delay changes caused by routing events, pp. 73–84. ACM (2007)
16. Schwartz, Y., Shavitt, Y., Weinsberg, U.: On the diversity, stability and symmetry of end-to-end internet routes, pp. 1–6. IEEE (2010)

17. Lcvenshtcin, V.: Binary coors capable or ‘correcting deletions, insertions, and reversals (1966)
18. Watts, D.J., Strogatz, S.H.: Collective dynamics of ‘small-world’ networks. Nature **393** (6684), 440–442 (1998)
19. Barabási, A.L., Albert, R.: Emergence of scaling in random networks. Science **286**(5439), 509–512 (1999)
20. Servetto, S.D., Barrenechea, G.: Constrained random walks on random graphs: routing algorithms for large scale wireless sensor networks. In: Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications, pp. 12–21. ACM (2002)
21. Blondel, O., Hilario, M.R., Santos, R.S.D., Sidoravicius, V., Teixeira, A.: Random walk on random walks: low densities. Mathematics (2017)
22. Li, M., et al.: Effects of weight on structure and dynamics in complex networks. arXiv preprint cond-mat/0601495 (2006)