

# Power-Law and Stable Distribution of BGP Paths

Julien Clément and Kwan-Wu Chin

**Abstract**—This paper presents two new properties of BGP routes. First, it shows that the frequency of autonomous systems (ASs) appearing in path vectors follow a Power-Law relationship. Secondly, it shows that path lengths can be characterized accurately by stable distributions. The main implication of these properties is that it allows the creation of realistic forwarding tables for simulation studies. Moreover, they extend previous works that have thus far only shown Power-Law relationships exhibited by ASs degree.

**Index Terms**—Internet, Border Gateway Protocol, Power Law, Stable Distributions

## I. INTRODUCTION

The Internet is comprised of thousands of autonomous systems (ASs) that use the Border Gateway Protocol (BGP) [8] to form peering relationships to gain access to the Internet. BGP enables each AS to advertise and receive reachability information from peer ASs, and to learn of routes offered by them. Figure 1 shows an example reachability information where PREFIX describes a reachable network via the path vector indicated in AS\_PATH, i.e., <12956, 22364 10937>. From the path vector, an AS can easily determine the set of ASs its packets will traverse through in order to reach hosts in the network 12.8.184.0/24.

```
TIME: 2007-1-1 11:00:00
TYPE: MSG_TABLE_DUMP/AFL_IP
VIEW: 0
SEQUENCE: 0
PREFIX: 12.8.184.0/24
STATUS: 1
ORIGINATED: Thurs Dec 21 04:56:42 2006
FROM: 194.68.129.179 AS 12956
AS_PATH: 12956 22364 10937
NEXT_HOP: 194.68.129.179
COMMUNITIES: 12956:123 12956:4003 12956:4030 12956:4302
```

Fig. 1. Reachability information.

To investigate the properties of path vectors, we use the publicly available BGP routing information bases (RIBs) from the RouteViews project [2] and RIPE [1]. In particular, those RIBs that were collected at midnight of January 1, 2007. In total, RouteViews has 17.2 million records, and RIPE has 15.9 million records – representing routes seen at fourteen and six locations respectively. In our discussions to follow, we refer to RIBs by their name. RIBs from RouteViews are called, EQIX, OREGON, ISC, OIX, WIDE and LINX, whereas those from RIPE are referred to as RRC0, RRC1, and RRC02 . . . RRC15.

Both authors are with the School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Northfields Ave, NSW, Australia 2522, email: {jyc157, kwanwu}@uow.edu.au  
ICITA2008 ISBN: 978-0-9803267-2-7

Note, we omit RRC08, and RRC09 since there were no data for January 1, 2007.

We first remove all duplicated records. After that we compress all inflated paths. For example, a path vector <10, 20, 30> is compressed to <10, 20, 30>. Finally, we count the number of times or frequency an AS appears in path vectors. We also record the frequency of route lengths.

## II. RESULTS

### A. Distribution of ASs in BGP Routes

Table II-A shows the ten most frequently appearing ASs over all RIBs. 71% of the 33.1 million recorded routes include at least one of these ASs. Also shown is their respective degree. An interesting observation here is that an AS's degree has no bearing on its frequency in BGP routes. We are currently exploring other reasons that better explain the ranking of ASs (in terms of frequency) in BGP routes. Readers are referred to [4] for more details.

TABLE I  
TOP-10 MOST FREQUENTLY APPEARING ASs OVER ALL RIBs.

AS Number	Company/Network Name	Frequency ('000')	Degree
3356	Level 3 Communications	5495	1340
1239	Sprint	4196	1727
3549	Global Crossing	4167	871
1299	Telianet	3558	443
701	MCI	3062	2406
2914	NTT America	3028	551
7018	AT&T	2766	2036
174	Cogent	2280	1628
3527	Tiscali	1899	338
6453	VSNL International	1752	372

Like ASs degree [5], the frequency of ASs in BGP routes for a given RIB also follows a Power-Law relationship. Specifically, if the distribution of ASs is sorted in decreasing order where the AS with the highest frequency is ranked one, followed by the second highest in frequency and so forth, the AS-frequency,  $y$ , is proportional to its rank,  $x$ .

$$y \propto x^a \quad (1)$$

where  $a$  is an exponent that characterizes the Power-Law relationship.

Figure 2 shows the frequency of ASs in routes from the OREGON RIB. The frequency is plotted on a log-log scale, and also shown is the theoretical Power-Law with an exponent of  $-1.19$ . From the figure, we see that the distribution of ASs follows the Power-Law relationship closely. In fact, with a correlation factor of 0.98. The exponent for other RIBs are shown in Table II-A.

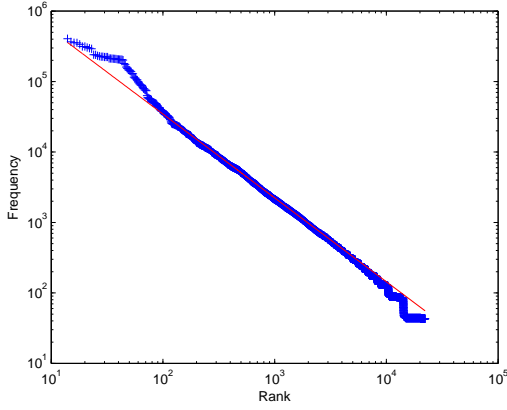


Fig. 2. OREGON RIB: Frequency versus Rank of ASs.

TABLE II  
POWER-LAW EXPONENTS FOR ALL DATASETS.

RIB	Exponent $\alpha$	Correlation Factor
EQIX	-1.22	0.99
OREGON	-1.19	0.98
ISC	-1.21	0.99
OIX	-1.08	0.99
WIDE	-1.20	0.96
LINX	-1.21	0.98
RRC0	-1.20	0.98
RRC1	-1.20	0.99
RRC02	-1.20	0.99
RRC03	-1.20	0.96
RRC04	-1.16	0.96
RRC05	-1.20	0.97
RRC06	-1.18	0.97
RRC07	-1.20	0.96
RRC10	-1.20	0.99
RRC11	-1.15	0.99
RRC12	-1.21	0.98
RRC13	-1.20	0.95
RRC14	-1.20	0.99
RRC15	-1.25	0.99

### B. Route Lengths

The next questions of interest are the length of routes and whether the distribution of routes can be modeled by a probability distribution.

Table II-B shows the path length recorded for each RIB. The longest path is very homogenous between snapshots. The longest path of six snapshots is 12 hops whereas six other snapshots have a longest path length of 13 hops. The average path length is very small. It goes over five in only one snapshot. Nine snapshots have an average path length between four and five whereas 10 of them have an average path length shorter than four. The average path length is 3.87. In other words, on average, it takes a packet less than four hops to reach its destination.

Figure 3 shows the distribution of path lengths for the OREGON RIB. The average path length is 3.84. An interesting question is determining whether there is a probability distribution that adequately describes Figure 3 and also those from other RIBs. Notice that Figure 3 has a heavy tail, and is skewed to the left; indicating that the frequency of path lengths may be characterized by a stable distribution.

According to [7], “stable distributions are a rich class of probability distributions that allow skewness and heavy tails”.

TABLE III  
PATH LENGTHS INFORMATION FOR ROUTEViews RIBs.

RIB	Path Length Range	Average Path Length	Variance
EQIX	1-12	3.63	0.85
OREGON	1-13	3.84	1.08
ISC	1-15	3.19	0.65
OIX	1-8	2.67	0.71
WIDE	1-10	4.08	0.96
LINX	1-12	3.77	0.95
RRC0	1-13	4.34	1.30
RRC1	1-12	3.89	0.94
RRC02	1-6	2.38	0.52
RRC03	1-13	4.00	1.06
RRC04	1-13	4.08	0.82
RRC05	1-12	4.02	0.89
RRC06	1-12	3.94	0.98
RRC07	1-13	4.28	1.28
RRC10	1-10	3.98	0.83
RRC11	1-12	4.01	1.07
RRC12	1-13	4.04	1.13
RRC13	1-14	5.03	1.68
RRC14	1-11	3.59	0.74
RRC15	1-13	4.57	0.96

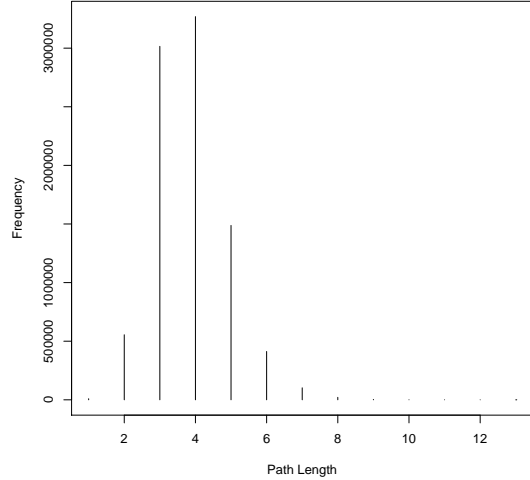


Fig. 3. Path Length versus Frequency.

In particular, they are defined by a characteristic function with four parameters:  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ , which describes their stability, skewness, scale and location respectively. These parameters are restricted to the range  $\alpha \in (0, 2]$ ,  $\beta \in [-1, 1]$ ,  $\gamma \geq 0$ , and  $\delta \in \mathbb{R}$ . Example stable distributions with close form expressions include Gaussian, Cauchy and Lévy. We have tried matching the distribution of path lengths to each of them, but to no avail. Hence, we resorted to finding suitable parameter values for the characteristic function.

The four parameters are estimated using the maximum likelihood estimation capability of STABLE [6]. We first compute the probability density function (PDF) of route lengths, and used it to generate 1000 samples. These samples are then read into STABLE. Note, higher number of samples has a negligible effect on results.

Table II-B shows the resulting parameter values for all RIBs. As an example, Figure 4 shows a plot of a stable distribution (“solid line”) using the estimated parameters for the OREGON

RIB. The corresponding Q-Q plot<sup>1</sup> is shown in Figure 5. Both figures confirm that a stable distribution does indeed characterize the distribution of path lengths.

TABLE IV  
STABLE DISTRIBUTION PARAMETERS.

Dataset	$\alpha$	$\beta$	$\gamma$	$\delta$
EQIX	1.8531	0.9895	0.6180	3.5243
OREGON	1.7998	0.9995	0.6785	3.6781
ISC	1.8530	0.9897	0.6198	3.5459
OIX	1.7301	0.9887	0.5319	2.5161
WIDE	1.8660	0.9884	0.6584	3.9688
LINX	1.8587	0.9899	0.6188	3.6700
RRC0	1.8938	0.9871	0.7933	4.2478
RRC1	1.8657	0.9886	0.6516	3.7921
RRC02	1.9752	0.9842	0.5074	2.3629
RRC03	1.8583	0.9900	0.6931	3.8855
RRC04	1.8591	0.9179	0.6177	3.9948
RRC05	1.8693	0.9858	0.6414	3.9228
RRC06	1.8566	0.9896	0.6555	3.8287
RRC07	1.7252	0.9898	0.7284	4.0622
RRC10	1.8664	0.9211	0.6102	3.8847
RRC11	1.8691	0.9899	0.7023	3.9045
RRC12	1.8620	0.9900	0.7238	3.9321
RRC13	1.9759	-0.1252	0.9256	5.0299
RRC14	1.8521	0.9889	0.5663	3.4957
RRC15	1.8582	0.7933	0.6556	4.4856

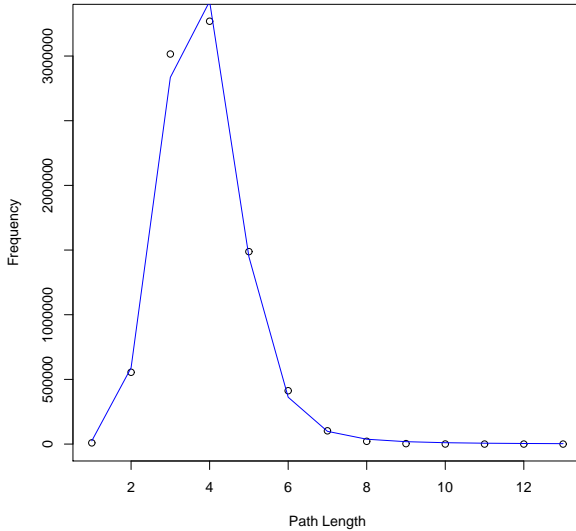


Fig. 4. Fitting a stable distribution with parameters  $\alpha = 1.7998$ ,  $\beta = 0.995$ ,  $\gamma = 0.6785$ ,  $\delta = 3.6781$  onto the distribution of path lengths from the OREGON RIB (shown as circles).

### III. CONCLUSION

This paper presents two interesting contributions. First, unlike existing studies [5][3] that have mainly focused on the Power-Law relationship exhibited by ASs degree, we show that the rank of ASs in path vectors also follow a Power-Law relationship. This means the majority of routes on the Internet involve a handful of ASs. Secondly, we show that stable distributions model the distribution of path lengths

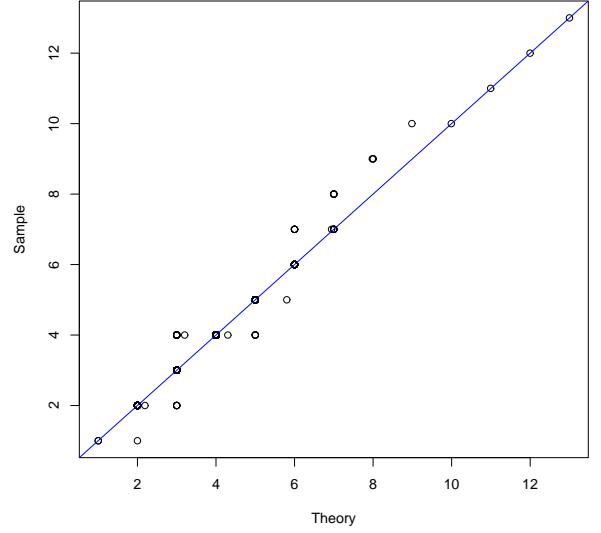


Fig. 5. Q-Q plot of a stable distribution (“Theory”) and the distribution of path lengths (“Sample”) from the OREGON RIB.

accurately. To the best of our knowledge, no other works have provided such model. In closing, both contributions can be used to generate realistic forwarding tables for simulation studies involving routing on the Internet.

### REFERENCES

- [1] Ripe. <http://www.ripe.net/projects/ris/rawdata.html>.
- [2] Routeviews project. <http://www.routeviews.org>.
- [3] H. Chang, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. Towards capturing representative as-level internet topologies. *Elsevier Computer Networks*, 44:737–755, 2004.
- [4] J. Clement and K.-W. Chin. On the characteristics of BGP routes. In *The 7th IEEE International Symposium on Communications and Information Technology (ISCIT'07)*, Sydney, Australia, Oct. 2007.
- [5] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *Proceedings of ACM SIGCOMM*, Sept. 1999.
- [6] J. Nolan. Stable 4.0.xx. Software Documentation, 2006. <http://academic2.american.edu/~jpnolan/stable/stable.html>.
- [7] J. P. Nolan. *Stable Distributions - Models for Heavy Tailed Data*. Birkhäuser, Boston, 2007. In progress, Chapter 1 online at [academic2.american.edu/~jpnolan](http://academic2.american.edu/~jpnolan).
- [8] Y. Rekhter, T. Li, and S. Hares. A border gateway protocol 4 (BGP-4). RFC 4271, Jan. 2006.

<sup>1</sup>Quantile-quantile plot is a graphically technique used to show whether two datasets originate from the same distribution.