

## An Alias Resolution Method Based on Delay Sequence Analysis

Yang Tao<sup>1</sup>, Gang Hu<sup>1</sup>, Bingnan Hou<sup>1</sup>, Zhiping Cai<sup>1,\*</sup>, Jing Xia<sup>1</sup> and  
Cheang Chak Fong<sup>2</sup>

**Abstract:** Alias resolution, mapping IP addresses to routers, is a critical step in obtaining a network topology. The latest work on alias resolution is based on special fields in the packet, such as IP ID, port number, etc. However, for security reasons, most network devices block packets for setting options, and some related fields exist only in IPv4, so these methods cannot be used for alias resolution of IPv6. In order to solve the above problems, we propose an alias analysis method based on delay sequence analysis. In this article, we present a new model to describe the distribution of Internet delays and give a mathematical proof. After experimental measurements using the Macroscopic Internet Topology Data Kit (ITDK) and Ark IPv6 Topology Dataset, it was found that the statistical differences in most alias delay models were very small. The statistical differences in the non-alias delay models are spread over a wide range. Using the wavelet decomposition in delay sequence, it was found that the approximate components and the detail components of the delay sequence of aliases were the same after filtering out the noise, which provided a theoretical explanation for the experimental results. This technology is applicable to both IPv4 and IPv6.

**Keywords:** Alias resolution, IPv6, Internet topology, measurement methodology.

### 1 Introduction

Ping tool is widely used to measure Internet delay, based on ICMP, UDP or TCP protocol. The tool shows the reachability of a host on Internet and round-trip time from source host to destination host. The program reports errors, packet loss, time to live (TTL) and a statistical summary of the results, typically including the minimum, maximum, the mean round-trip times, and standard deviation of the mean [Muuss (1983)]. Because of its cross-platform (windows, mac and Linux), and its cross-version (IPv4 and IPv6), ping tool is selected to obtain reliable experimental data.

Alias resolution is the process of identifying which interface IP addresses belong to the same routers and is required to convert the abstract IP-level topology discovered by

---

<sup>1</sup> College of Computer, National University of Defense Technology, Changsha, 410073, China.

<sup>2</sup> Macau University of Science and Technology, 999078, Macau.

\*Corresponding Author: Zhiping Cai. Email: zpcai@nudt.edu.cn.

Received: 22 January 2020; Accepted: 02 March 2020.

traceroute to a more concrete router-level topology [Pansiot and Grad (1998); Gunes and Sarac (2007)]. Accurate and efficient identification of alias IPs is important for obtaining real router-level network topologies. The existing alias resolution methods can be divided into two types based on active probe and statistical analysis. The method based on the active probe obtains the response packet by detecting the IP address of the interface, and performs alias resolution based on the source address, the IP identifier, and the optional field of the response packet. The method based on the statistical analysis is based on the router host naming rules, IP address assignment conventions and network composition, and statistical analysis results such as network graph structure, for alias resolution.

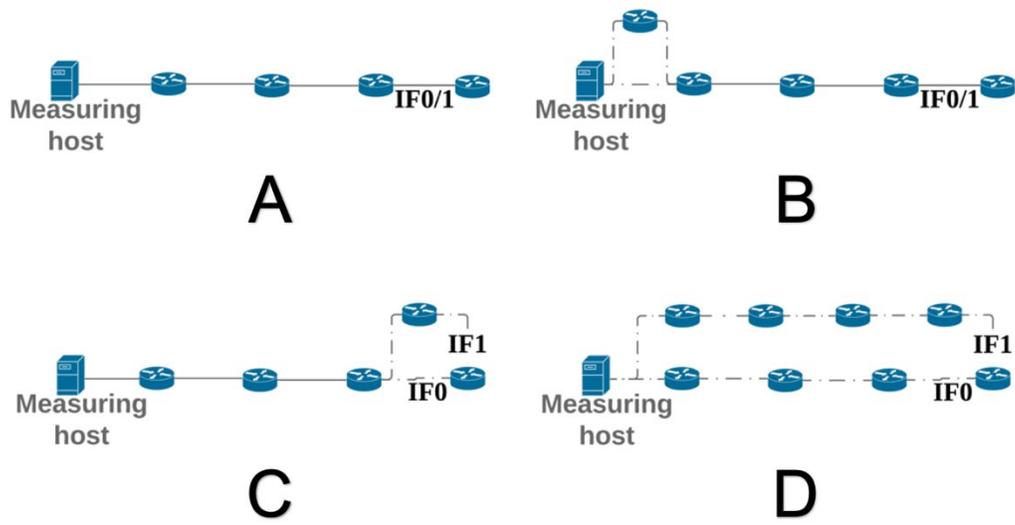
Typical alias resolution methods based active probe are as follows: 1) alias resolution according to responding to the content of the packets, including header and data. Multiple interfaces of the router usually share a counter. When a packet is generated, the counter value will be the IP identification of the packets, and the IP-ID series is in linear or nonlinear relationship. the IP-ID values of aliases in response packets are ordered and adjacent [Spring, Mahajan and Wetherall (2002)]. Radar Gun [Bender, Sherwood and Spring (2008)] considers that the IP-ID series in multiple response packets of aliases are similar. MIDAR [Keys, Hyun, Luckie et al. (2013)] considers IP-ID series of alias have monotonic trend in common. However, most routers or hosts were found to have a constant IPID (39%) or a local counter (34%), and global counters account for approximately 18%, counters using random number account for approximately 9% [Salutari, Cicalese and Rossi (2018)]. Options in ICMP is used to alias resolution such as Sidecar [Sherwood and Spring (2006)], RIPAPT [Sherry, Katz-Bassett, Pimenova et al. (2010)], Pythia [Marchetta, Persico and Pescapè (2013)], etc. In 2014, the IETF (The Internet Engineering Task Force) recommended that network devices should discard the packets with option settings for security. This strategy leads to the inability to measure based on options. 2) alias resolution according to inductive detection. Mercator, iffinder [Govindan and Tangmunarunkit (2000)] consider that when using the UDP high port probe for the router interface IP, the source address of the response packet header may be the other IP in the same router, which is used to determine alias by comparing the IP address of the probe packet with the IP address of the source address in the response packet. But only few IPs will response UDP high port probe and the fewer will response from the other source [Govindan and Tangmunarunkit (2000)]. Robert [Beverly, Brinkmeyer, Luckie et al. (2013)] take packets of a specific size to induce alias information This packet exceeds the 1280 B limit of the IPv6 protocol MTU and is less than the 1500 B limit of the MTU of the IPv4 protocol. The target splits the message into 2 packets for transmission and joins the segment ID. It can be used for resolution. But this measurement mechanism causes huge network load.

The alias resolution theory based on statistical analysis is as follows: IPs with the same host name or similar naming rules [Gunes and Sarac (2006)], in the same subnet [Augustin, Cuvellier, Orgogozo et al. (2006)] is aliases or two addresses that directly precede a common successor are aliases, assuming point-to-point links are used [Spring, Dontcheva, Rodrig et al. (2004)]. The practicality of these theories is not satisfactory. Router host name acquisition is difficult, and host naming rules are not in same standard. There are a large number of unknown interfaces on the router, and there is no stable connection between the interfaces.

With the development of technology, machine learning is used in alias resolution. using neural networks to improve the accuracy and universality of analysis. Grailet et al. [Grailet and Donnet (2017)] proposed an alias resolution algorithm by combining various measurement methods. With machine learning, a large number of aliases clustering is realistic. However, interpretability of neural networks used for alias resolution is still unknown. This paper proposes an alias analysis technique using delay series and also explains why neural network usage delays for alias resolution. Based on the previous work, we proposed the mathematical form of the delay model of the Internet link and gave proof. This model was used to calculate the difference in delay models. And we compared two methods of measuring the difference in delay. This paper introduces wavelet decomposition to distinguish and analysis aliases and discovers delay series can be used for alias resolution.

**2 Theory of link delay distribution**

Round trip times is the elapsed time between the instant a packet is released by the source to the instant the corresponding ack is received by the source [Shakkottai, Brownlee and Broido (2004)]. The distribution of RTTs can dynamically display network link status, congestion and bandwidth in interval moments. Many theories show that network delays are regular, and a large number of researches have explained the network delay distribution with accurate models. There are queuing delays present in the RTTs which are accentuated by distance in geography [Shakkottai, Brownlee and Broido (2004)]. Aliases in routers are in same location, and the queuing delays of these are almost equal.



**Figure 1:** There are the circumstances that traceroute from the measure host to the target alias. The alias is bound to interface 0 and interface 1. The solid line represents the public route and the dotted line represents the difference route

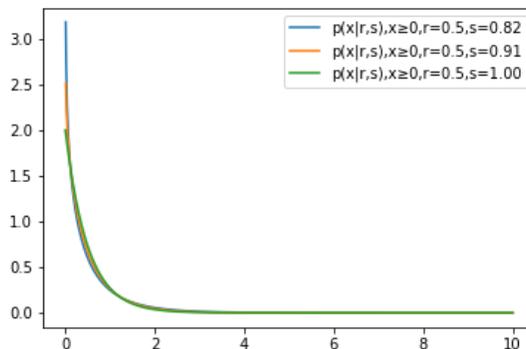
From Fig. 1A, 1B and 1C have many public routes, and the path similarity is at a high level. The routing paths in D has almost no public route, and this circumstance is not considered in alias resolution.

Aliases have many characteristics in common [Keys, Hyun, Luckie et al. (2013)]. Using data sets from 2000000 IPs from CAIDA, experiments show that about 96.7% of the measuring hosts to aliases routes belong to circumstances A, B and C, and rarely belong to circumstance D. There is usually no public route for non-aliases in traceroute, about 16.7% of which have only a few hop (less than 3) public routes.

The theory is based on the fact that the same datagrams of the same link at the same interval has the same delay [Hu, Xiang, Wu et al. (2019)]. Due to the dynamics of the Internet, a small number of delay measurements are not enough to reflect the characteristics of the entire link, there needs a model of the delay distribution. It was found that the round-trip delay distribution could be well approximated by a truncated normal distribution [Shakkottai, Brownlee and Broido (2004); Elteto and Molnar (1999)]. And Konstantina Papagiannaki proves that single-hop queueing delay from an operational backbone network is in Weibull distribution [Papagiannaki, Moon, Fraleigh et al. (2002)], the probability density function (PDF) of that is given by:

$$p(x|r, s) = \frac{sx^{s-1}}{r^s} e^{-\left(\frac{x}{r}\right)^s} + C, x \geq 0, r, s \geq 0 \quad (1)$$

C is a constant representing the processing delay, transmission delay and propagation delay. Considering a fixed link, the transmission delay and propagation delay are the same for the same packets. And the processing delay is negligible because routers usually process the packets within 1 ms. For ease of calculation, single-hop model of queueing delay will degenerate into an Exponential distribution within the error tolerance due to  $s = 0.82$  [Papagiannaki, Moon, Fraleigh et al. (2003)]. From Fig. 2, the Weibull distribution is very similar to the exponential distribution when the shape parameter is close to 1.



**Figure 2:** there are difference between Weibull distribution and Exponential distribution in same scale parameter  $r = 0.5$  and similar shape parameter  $s \leq 1$

Therefore, the probability density function can be written as:

$$p(x|r) = \frac{1}{r} e^{-\left(\frac{x}{r}\right)}, x \geq 0, r \geq 0 \quad (2)$$

We prove that the entire link delay is consistent with the Gamma distribution in appendix:

$$p(x|r, s) = \frac{x^{s-1}}{\Gamma(s)r^s} e^{-\frac{x}{r}} + C, x \geq 0, r, s \geq 0 \quad (3)$$

Obviously, Gamma distribution is equivalent to Weibull distribution only if  $s = 1$ . We consider the queue event is “bad” thing in transactions and there is not queue delay in perfect circumstances because it is the most efficient for packet in transaction without queuing. The shape parameter is that power plus one, and so this parameter can be interpreted that  $s = 1$  indicates that the “bad” rate is constant over time. This might suggest random external events are causing “bad” events and correspond to the generation mechanism of queuing delay.

The network delay conforms a certain distribution and model. In fact, this model shows that the network delay distribution is affected by the entire link, including TTL, the congestion of per hop, the forwarding performance of each router, etc. The difference of delay model can also display the difference of entire link characteristics. We introduce the energy distance [Székely and Rizzo (2013)] and Wasserstein distance used to measure the difference in the network delay distribution model, and the differential values of aliases is typically less than no-aliases.

### **3 Data sources**

Because the dynamics of the internet, there is some mistake in the dataset. IP in alias group may be not in same router now and no-aliases may be assigned in one router. We drop those that are significantly different from other aliases due to high loss rate.

#### **3.1 IPv4**

Macroscopic Internet Topology Data Kit (ITDK) is a classical dataset of Internet measure, including 3.01 million addresses extracted from the IPv4 Routed/24 Topology Dataset. The ITDK contains data about connectivity and routing gathered from a large cross-section of the global Internet. This dataset is useful for studying the topology of the Internet at the router-level, among other uses. The nodes file of ITDK lists the set of interfaces that were inferred to be on each router.

#### **3.2 IPv6**

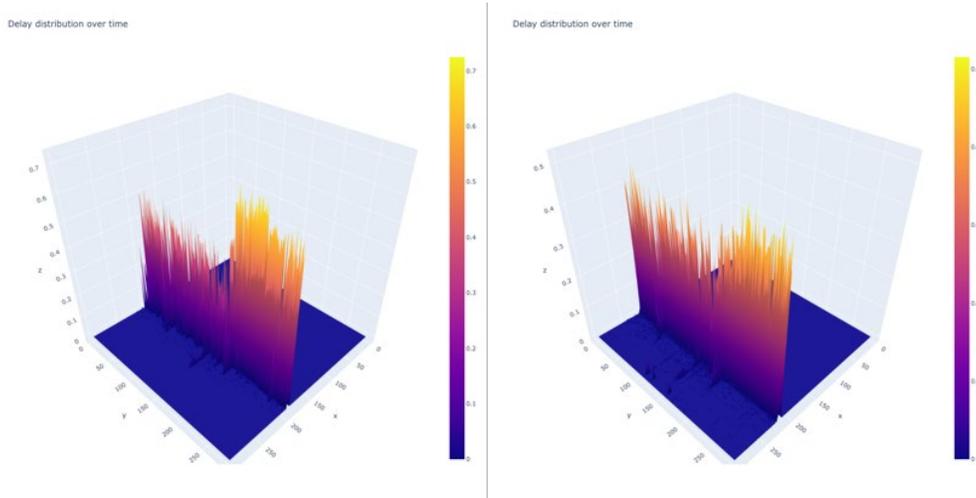
The Ark IPv6 Topology Dataset contains information useful for studying the IP-topology of the IPv6 Internet. It uses scamper to perform ICMP-based traceroutes using the Paris traceroute technique. For each probed path, this dataset includes the IP address, RTT, reply TTL, and ICMP responses for all hops. This dataset verifies the validity of the delay series theory in IPv6.

### **4 Methodology and data analysis**

We analyze the delay distribution of a large number of aliases in 24 hours, whose granularity is 5 minutes. Obvious and stratified results are obtained, Internet delay dynamically change in a small range. Experiments also show difference of no-aliases by energy distance. We believe these conclusions will help later research.

#### 4.1 Delay distribution over time

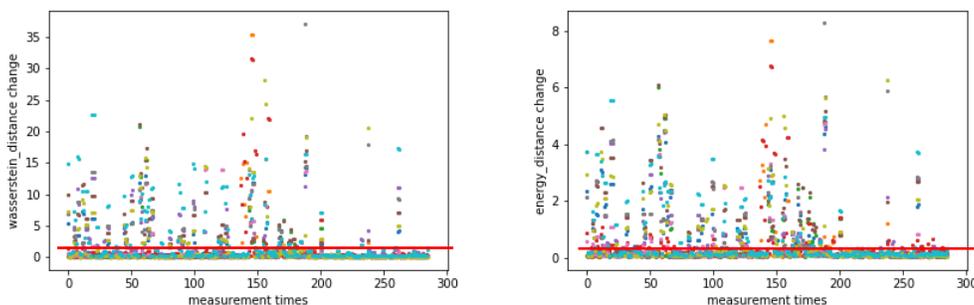
We send measuring probes to alias group include about 100 aliases in 5 minutes granularity. Measurement lasted for 24 hours and 288 times. The characteristics of the delay distribution clearly demonstrated by the measurement results in a chronological 3D diagram.



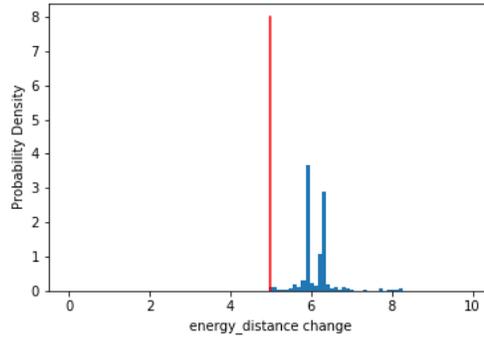
**Figure 3:** The X-axis represents the delay in units of millisecond, and Y-axis represents the number of measurements

From Fig. 3, The pattern of the delay distribution slightly changes over time but the position of the maximum value of the probability density is fixed. In fact, all the experiments show that delay distribution over time of the same IP changes less.

For an IP, there is stability in the latency distribution state in Fig. 4. Using Wasserstein distance as a metric to calculate, 88.2% of IP's continuous change over time does not exceed 2.0. When using energy distance as measure method, IPs of less than 0.5 accounts for 92.7% of the total aliases. It means that delay distribution is stable and experimentally verifiable in proper interval. But a large interval between the two measurement times of an IP results in large errors, even more than no-aliases. The accumulation of small errors in a short period of time can make a huge difference. Therefore, measurement duration should be within a small range.



**Figure 4:** Continuous changes in the latency distribution of one IP



**Figure 5:** Energy distance of different timestamps

From Fig. 5, delay in different timestamps have strong fluctuation. It results in that energy distance is in high level, about 5.0 and there is obvious distinction between delay distribution in an interval and in different timestamps.

Experiments show that delay change in interval is in low level and can be considered as a stable model. But the distribution models have high level distinction with each other on a long-term scale, for example one hour. Measurement of delay distribution is consisting of a large number of values and can maintain timeliness over measurement time.

#### **4.2 Difference of aliases and no-aliases in delay**

We deploy three hosts in Los Angeles and Seattle and performed extensive latency measurements of 400,000 IPs belonging to approximately 6,000 routers from ITDK. After the analysis of delay distribution, it is found that aliases have great similarity with others which reflected in the low level of difference in the energy distance, Wasserstein distance and the wavelet analysis of delay series. On the contrary, no-aliases are in high level.

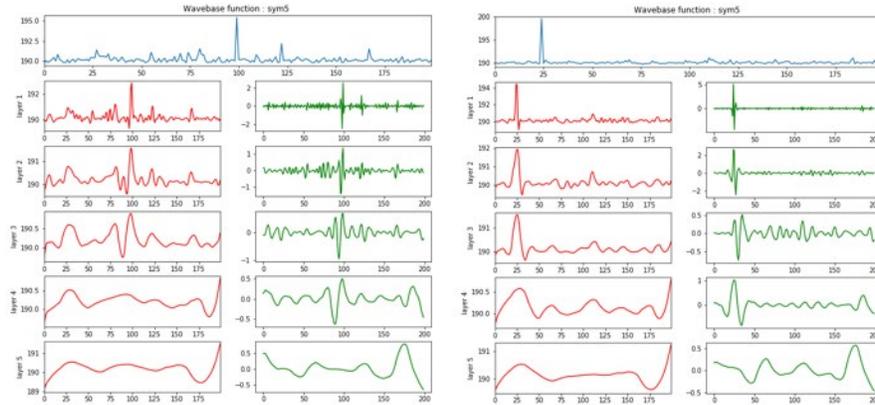
To distinguish between non-aliases in latency distribution, we randomly sampled 10,000 IP pairs from 4 million IP pairs, of which 5,000 were alias pairs and 5,000 non-alias pairs. And thresholds are set to 5.0 and 2.0 in the calculation of energy distance and Wasserstein distance after delay distribution regularization. There is a clear difference between the alias and the non-alias.

**Table 1:** Energy and Wasserstein distance in aliases

	Thresholds	Alias pairs	No-alias pairs
Energy distance	<3.0	72.6%	4.9%
	>5.0	5.0%	80.5%
Wasserstein distance	<100.0	81.9%	3.1%
	>120.0	9.5%	93.9%

A series of delay cannot be only summed up as a probability distribution model, but the sequential information of these values itself should also be used. Network delay has multiple sources: signal transmission in the link, router processing information,

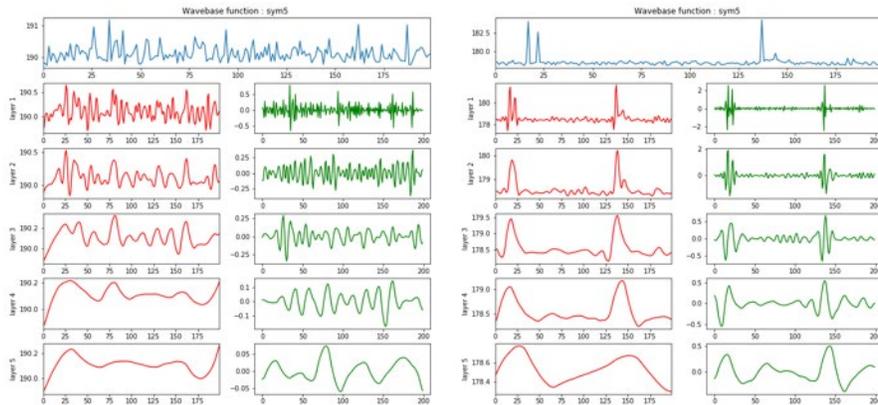
congestion control and traffic control, etc. We use wavelet decomposition to get important information for alias resolution in Figs. 6 and 7.



**Figure 6:** A wavelet analysis of the delay series of aliases was and there are 5 layers of decomposition

The delay series waveform of alias IP is similar in approximate component and local component. We use sym5 as wave-base function for decomposition due to its excellent performance in eliminating noise.

Despite the deviation in the time domain and Subtle noise wave, the main components of the waveforms are almost identical, especially in the fifth layer of decomposition. For no-aliases, there are typical difference between wavelet analysis results. After filtering out the noise, the approximate component gradient direction and values are still different.



**Figure 7:** A wavelet analysis of the delay series of no-aliases was and there are 5 layers of decomposition

Because there are a large number of public routes for alias pairs, the link status and the running status of the routers are the same at the same time, so the primary component of the delay is the same. The error of alias pairs is due to the fact that the moment is not fully simultaneous in the actual measurement.

## 5 Conclusion

This paper verifies and illustrates the mathematical expression of the overall model of network delay. Results show that Gamma distribution corresponds to the network delay distribution very well. This paper proposes an alias resolution method for the difference of delay probability distribution. Wavelet analysis was used to decompose the delay series and it was found that the aliases not only have similarity in the latency distribution, but the approximate components of the delay series without noise are almost the same in alias pairs, which used sequential information on the time domain. This idea is closer to reality, because the running status of the router do not always present a perfect periodicity, and these can partly be recorded in the delay series.

Extracting the main frequency and time domain information and filtering out the unrelated noise, which are main task for alias resolution using delay series. Based on the wavelet analysis results, it was shown that the delay is more than just an isolated number for discrimination, distribution and order information of which can also assist the alias identification. Delay series is feasible for alias resolution.

**Funding Statement:** This work is supported by The National Key Research and Development Program of China (2018YFB1800202, 2018YFB0204301, 2016YFB1000302, SQ2019ZD090149).

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- Augustin, B.; Cuvellier, X.; Orgogozo, B.; Viger, F.; Friedman, T. et al.** (2006): Avoiding traceroute anomalies with Paris traceroute. *Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement*, pp. 153-158.
- Bender, A.; Sherwood, R.; Spring, N.** (2008): Fixing ally's growing pains with velocity modeling. *Proceedings of the 8th ACM Sigcomm Conference on Internet Measurement*, pp. 337-342.
- Beverly, R.; Brinkmeyer, W.; Luckie, M.; Rohrer, J. P.** (2013): IPv6 alias resolution via induced fragmentation. *International Conference on Passive and Active Network Measurement*, pp. 155-165.
- Elteto, T.; Molnar, S.** (1999): On the distribution of round-trip delays in TCP/IP networks. *Proceedings 24th Conference on Local Computer Networks*, pp. 172-181.
- Govindan, R.; Tangmunarunkit, H.** (2000): Heuristics for internet map discovery. *Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 3, pp. 1371-1380.
- Graillet, J. F.; Donnet, B.** (2017): Towards a renewed alias resolution with space search reduction and IP fingerprinting. *Network Traffic Measurement and Analysis Conference*, pp. 1-9.

- Gunes, M. H.; Sarac, K.** (2006): Analytical IP alias resolution. *IEEE International Conference on Communications*, vol. 1, pp. 459-464.
- Gunes, M. H.; Sarac, K.** (2007): Importance of IP alias resolution in sampling Internet topologies. *IEEE Global Internet Symposium*, pp. 19-24.
- Hu, B.; Xiang, F.; Wu, F.; Liu, J.; Sun, Z.** (2019): Research on time synchronization method under arbitrary network delay in wireless sensor networks. *Computers, Materials & Continua*, vol. 61, no. 3, pp. 1323-1344.
- Keys, K.; Hyun, Y.; Luckie, M.; Claffy, K.** (2013): Internet-scale IPv4 alias resolution with MIDAR. *IEEE/ACM Transactions on Networking*, pp. 383-399.
- Liu, F.; Tang, G. M.; Li, Y. H.; Cai, Z. P.; Zhang, X. Z. et al.** (2019): A survey on edge computing systems and tools. *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1537-1562.
- Liu, F.; Guo, Y. T.; Cai, Z. P.; Xiao, N.; Zhao, Z. M.** (2019): Edge-enabled disaster rescue: a case study of searching for missing people. *ACM Transactions on Intelligent Systems and Technology*, vol. 10.
- Marchetta, P.; Persico, V.; Pescapè, A.** (2013): Pythia: yet another active probing technique for alias resolution. *Proceedings of the Ninth ACM Conference on Emerging Networking Experiments and Technologies*, pp. 229-234.
- Muuss, M.** (1983): The story of the PING program. *Retrieved March*.
- Pansiot, J. J.; Grad, D.** (1998): On routes and multicast trees in the Internet. *ACM SIGCOMM Computer Communication Review*, pp. 41-50.
- Papagiannaki, K.; Moon, S.; Fraleigh, C.; Thiran, P.; Diot, C.** (2003): Measurement and analysis of single-hop delay on an IP backbone network. *IEEE Journal on Selected Areas in Communications*, pp. 908-921.
- Papagiannaki, K.; Moon, S.; Fraleigh, C.; Thiran, P.; Tobagi, F. et al.** (2002): Analysis of measured single-hop delay from an operational backbone network. *Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 2, pp. 535-544.
- Salutari, F.; Cicalese, D.; Rossi, D. J.** (2018): A closer look at IP-ID behavior in the wild. *International Conference on Passive and Active Network Measurement*, pp. 243-254.
- Shakkottai, S.; Brownlee, N.; Broido, A.** (2004): The RTT distribution of TCP flows on the Internet and its impact on TCP based flow control. *Cooperative Association for Internet Data Analysis*.
- Sherry, J.; Katz-Bassett, E.; Pimenova, M.; Madhyastha, H. V.; Anderson, T. et al.** (2010): Resolving IP aliases with prespecified timestamps. *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, pp. 172-178.
- Sherwood, R.; Spring, N.** (2006): Touring the Internet in a TCP sidecar. *Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement*, pp. 339-344.
- Spring, N.; Mahajan, R.; Wetherall, D.** (2002): Measuring ISP topologies with Rocketfuel. *ACM SIGCOMM Computer Communication Review*, vol. 32, no. 4, pp. 133-145.
- Spring, N.; Dontcheva, M.; Rodrig, M.; Wetherall, D.** (2004): How to resolve IP Aliases. *Technical Report, University of Washington*.

**Székely, G. J.; Rizzo, M. L.** (2013): Energy statistics: a class of statistics based on distances. *Journal of Statistical Planning and Inference*, pp. 1249-1272.

**Wu, H. J.; Zhang, J.; Cai, Z. P.; Liu, F.; Li, Y. Y. et al.** (2020): Towards energy-aware caching for intelligent connected vehicles. *IEEE Internet of Things Journal*.

**Yu, G.; Cai, Z. P.; Wang, S. Q.; Chen, H. W.; Liu, F. et al.** (2019): Unsupervised online anomaly detection with parameter adaptation for KPI abrupt changes. *IEEE Transactions on Network and Service Management*.

### Appendix A. Proof of the entire link model

The probability density function of single-hop delay depredates into exponential distribution, because  $s \approx 1$ :

$$p(x|r) = \frac{1}{r} e^{-\left(\frac{x}{r}\right)}, x \geq 0, r \geq 0 \quad (4)$$

At the macro level, the delay from different hops is independent and in exponential distribution for packet switching. The problem is translated into that if random variables for single-hop delay:

$$X_i \sim \frac{1}{r} e^{-\left(\frac{x}{r}\right)}, x \geq 0, r \geq 0, i = 1, 2, \dots, n \quad (5)$$

then random variables for entire link:

$$X = X_1 + X_2 + \dots + X_n \quad (6)$$

Assume the probability density function of entire link  $p(x|s, r) = \frac{x^{s-1}}{\Gamma(s)r^s} e^{-\left(\frac{x}{r}\right)} + C, x \geq 0, r, s \geq 0$ , let  $f(x) = p(x|s, r) - C$ , Obviously, while  $M = 1, s = 1$ :

$$f(x) = \frac{1}{r} e^{-\left(\frac{x}{r}\right)}, x \geq 0, r \geq 0 \quad (7)$$

Hypothesis  $M \leq n$ :

$$f(x) = \frac{x^{s-1}}{\Gamma(s)r^s} e^{-\left(\frac{x}{r}\right)}, x \geq 0, r, s \geq 0 \quad (8)$$

While  $M = n + 1$ , let  $y = x - x_{n+1}$ :

$$f(x) = \int_0^x f(y)f(x-y) dy \quad (9)$$

Then:

$$f(x) = \frac{e^{-\frac{x}{r_{n+1}}}}{\Gamma(s)r^s r_{n+1}} \frac{r^s r_{n+1}^s \Gamma\left(s, \frac{(r_{n+1}-r)x}{r_{n+1}r}\right)}{(r_{n+1}-r)^s} \quad (10)$$

Because  $\forall r_i \approx \text{contant}$ ,  $r_{n+1} - r \approx 0$ ,

$$f(x) = \frac{x^s}{\Gamma(s+1)r^{s+1}} e^{-\left(\frac{x}{r}\right)}, x \geq 0, r, s \geq 0 \quad (11)$$

we can conclude that the delay of entire link follow gamma distribution.