Regular Article

Defining Quantiles for Estimating IP Packet Delay Variation in NGN Core Networks

Dao Ngoc Lam, Le Nhat Thang, Le Huu Lap

Posts and Telecommunications Institute of Technology, Hanoi, Vietnam

Correspondence: Le Nhat Thang, thangln@ptit.edu.vn
Manuscript communication: received 4 October 2011, revised 16 November 2011

Abstract—Traffic delay is one of the important metrics used for evaluating network performance. Delay and delay variation characteristics of IP packets transferred over multi-section networks can be derived, estimated or composed from component distributions of IP package delay in each network section. Approximate methods are needed in the cases of unknown or complicated delay distribution functions, which are unavailable or unusable in practice. The ITU-T has proposed a method for estimating IP packet delay variation. One of noticeable factors affecting the estimation accuracy is the packet delay population quantile which has not been adequately considered. The objective of this paper is to examine the optimal range of quantiles used for estimating the IP packet delay variation in the NGN (Next Generation Network) core networks. The paper is composed from the following ideas. Firstly, several concepts and mathematical formulas related to delay metrics based on probability and statistics theory are defined. The approximate method of ITU-T for estimating the IP packet delay variation in a multi-section network is revised. Then, another method based on convolution for composing the empirical IPTD distribution functions is proposed for the same target as the first one. Secondly, a number of test cases are implemented to measure the IP packet delay on several sections of an NGN core network. Sample data are used for computing and estimating the IP packet delay variation for multi-section networks by two methods with certain hypotheses. Finally, these methods are compared and evaluated both theoretically and empirically in regards to the estimation accuracy versus quantiles of the IP packet transfer delay. The best range of quantiles is determined to ensure the accuracy of the estimation method applied for the NGN core network.

Keywords—Delay distribution, delay variation, quantile, multi-section network, NGN core network.

1 Introduction

An NGN core network can be partitioned into different network sections, which is a set of hosts together with all of their interconnecting links that provide a part of the whole network and are under a single jurisdictional responsibility of a service provider or operator. Performance parameters characterizing packet delay in an IP network are represented by two key metrics—the IP Packet Transfer Delay (IPTD) and the IP Packet Delay Variation (IPDV)—, which are defined in the ITU Y.1540 recommendation [1]. IPTD is a random variable measuring time interval from when the first bit of a packet crosses an ingress measurement point to when the last bit of the same packet crosses an egress measurement point. IPDV is defined as the difference between upper and lower quantiles of IPTD population of interest:

\[
\text{IPDV} = \text{IPTD}_{p_2} - \text{IPTD}_{p_1}, \quad 0 \leq p_i \leq 1, \quad i = \{1, 2\} \quad (1)
\]

in which IPTD_{p_i} or p_i-quantile of IPTD is the T value such that \(P(T < \text{IPTD}_{p_i}) \leq p_i\) and \(P(T \leq \text{IPTD}_{p_i}) \geq p_i\), where \(P\) is a probability operator. \(T\) can be derived from the inverse of the cumulative distribution function.

Performance metrics are normally monitored or measured independently in each separated network section but not over multiple ones whereas detailed measurement data will seldom be shared among operators. Moreover, fewer segment measurement results used in calculating thousands of concatenated network sections will help minimize much cost and overhead. For this reason, a provisional method based on mathematical theory for estimating a value of IPDV over a multi-section network is recommended by the ITU [2].

In spite of low computational complexity and overhead, the ITU method encounters unavoidable data error and method error. The first one is due to empirical actions and may be improved by repeating or correcting actions. The second one is due to approximate estimation of the own method. This kind of error has been pointed out in a research paper of [3]. However, the paper approach was only based on pure mathematical calculations with a hypothesis of ideal simple distributions, which are not the true for the real network. Besides, there is almost no research for examining the accuracy of the method yet. In addition, the factors influencing errors and selection criteria for quantiles have not been considered adequately so far. Whereas quantiles have a considerable effect on the accuracy of resulted IPDV. How the quantiles affect the method accuracy and which optimal value ranges of quantiles for estimating IPDV in a multi-section NGN core network are essential questions under investigation in this paper.

The remainder consists of the following sections. Two methods for determining IPDV over multi-section...
networks are mentioned in Section 2. The first one is an approximate estimation method revised from an ITU recommendation. The second one is a new method proposed based on convolution of empirical IPTD distribution function (convolution method for short). After that, the accuracy of both methods will be analyzed, compared and evaluated versus quantiles in Section 3. Mathematical reasoning and experimental results will be used to prove and evaluate errors of each method. Finally, the best ranges of quantiles will be determined to ensure the accuracy of the method applied for the NGN core networks.

2 Methods for Determining IPDV over Multi-Section Networks

2.1 The ITU-T Method for Estimating a Value of IPDV over Multi-Section Networks

A provisional way to estimate IPDV across a network consisting of N sections is suggested in the ITU-T Y.1541 recommendation [2, Section 8.2.4]. It is assumed that the IPTDs on each section are mutually independent. Approximate estimation is used to derive a value of IPDV from IPTD samples in the case of no explicit or simple distribution function available. This method is based on a normal power approximation (NPA) which transforms approximately a random variable of \( T \) based on a normal power approximation (NPA) which

\[
\text{erf}^{-1}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt
\]

where \( x \) is the probability of interest. \( \text{erf}^{-1} \) is the inverse error function whose values can be found from standard probability texts. It can be deduced from (3) to have

\[
\gamma \approx 6 \times \frac{x_p - \frac{T_p - \mu}{\sigma}}{1 - x_p^2}
\]

where \( x_p = \Phi^{-1}(p) = \sqrt{2} \text{erf}^{-1}(2p - 1) \).

In (4), \( \text{erf}^{-1} \) is the inverse error function whose values can be found from standard probability texts. It can be deduced from (3) to have

\[
\gamma \approx 6 \times \frac{x_p - \frac{T_p - \mu}{\sigma}}{1 - x_p^2}
\]

which yields

\[
T_p \approx \mu + \sigma \left[ x_p - \frac{\gamma}{6} (1 - x_p^2) \right].
\]

The ITU method based on the following mathematical relations includes three steps:

1. Compute the means and variances of IPTD:

\[
\mu_k = \frac{1}{M} \sum_{i=1}^{M} D_i,
\]

\[
\sigma_k^2 = \frac{1}{M} \sum_{i=1}^{M} (D_i - \mu_k)^2,
\]

\[
\mu = \frac{1}{N} \sum_{k=1}^{N} \mu_k,
\]

\[
\sigma^2 = \frac{1}{N} \sum_{k=1}^{N} \sigma_k^2,
\]

where \( M \) is the IPTD sample size, \( D_i \) is the \( i \)-th measurement of IPTD, \( N \) is the number of network sections, \( \mu_k \) and \( \sigma_k \) are, respectively, the mean and the standard deviation of the IPTD on the \( k \)-th network section, \( \mu \) and \( \sigma \) are, respectively, the mean and the standard deviation of the IPTD distribution on the whole network.

2. Estimate skewness of the IPTD distribution:

\[
\gamma_k \approx 6 \times \frac{x_p - \frac{T_k - \mu_k}{\sigma_k}}{1 - x_p^2}
\]

\[
\omega_k = \gamma_k \times \sigma_k^3
\]

\[
\omega = \sum_{k=1}^{N} \omega_k
\]

\[
\gamma = \frac{\omega}{\sigma^3}
\]

where \( p \) is the probability of interest, \( t_k \) is the \( k \)-quantile of IPTD in the \( k \)-th section, \( x_p \) is the \( p \)-quantile of \( \mathcal{N}(0,1) \), \( \gamma_k \) and \( \omega_k \) are, respectively, the skewness and the third moment of the IPTD distribution on the \( k \)-th section, \( \gamma \) and \( \omega \) are, respectively, the skewness and the third moment of the IPTD distribution on the whole network.

3. Estimate IPDV over the whole network:

An upper quantile of IPTD over the whole network is estimated from (5) as

\[
T_p \approx \mu + \sigma \left[ x_p - \frac{\gamma}{6} (1 - x_p^2) \right].
\]

The ITU method based on the following mathematical relations includes three steps:
They are rarely approximate to some well-known function as well. For these reasons, IPTD samples need to be processed to build an empirical discrete distribution function, which is a transformation of the original continuous distribution function. The following steps are used in the convolution method to compute a value of IPDV over multiple network sections from empirical data samples of IPTD over each network section.

Step 1 – Determine the empirical probability distribution of the IPTD over each network section: A range of IPTD values $[0, \max(D)]$ is determined from the measured samples $D = \{D_i\}_{i=1,\ldots,M}$ with the size of $M$. Data samples are ordered and mapped according to their numeric values into $K$ bins of the same size of $L = \max(D)/K$. The $k$th bin $(r_k)$ is specified by the size of $[(k-1) \times L + 1, k \times L]$ for $k = 1,\ldots,K$. A bin can be considered as a discrete random variable $R$ with the value range or state space of $r = \{r_k\}$, where $r_k = k \times L - L/2$. The number of elements within $r_k$ is counted as $c_k$ which is used to compute values of the empirical probability mass function of $R$, given by $f_k(r_k) = c_k/M$. These values also represent the probability density of IPTD at the corresponding values within $r_k$.

Step 2 – Determine the empirical probability distribution of IPTD over the whole network: The empirical probability mass function of each network section will be used to compose probability mass function of the whole network as follows:

$$f_{\Sigma_i} (r) = f_{R_1} \ast f_{R_2} \cdots \ast f_{R_N} (r),$$

where $N$ is number of network sections, $R_i$ is a random variable of $R$ defined for the $i$-th section, $f_{R_i}, i=1,\ldots,N$ is the probability mass function of $i$-th network section for random variable $R_i$. The empirical cumulative distribution function can be inferred from (8) as follows:

$$F_{\Sigma_i} (r) = P \left( \sum_{i=1}^N R_i \leq r \right) = \sum_{r_k \leq r} f_{\Sigma_i} (r_k).$$

Step 3 – Determine IPDV over the whole network: A $p$-quantile of IPTD can be found from (9) as follows:

$$T_p = F_{\Sigma_i}^{-1} (p),$$

where $F^{-1}$ is the inverse of $F$. Finally, a value of IPDV can be obtained from (1) as $\text{IPDV} = T_p - \min(\text{IPTD})$.

3 Effects of Quantiles on Accuracy

3.1 Theoretical Analysis and Accuracy Evaluation

According to the ITU method (see Section 2.1), it is expected to have approximately equal raw moments from (2):

$$E \left[ \left( \frac{T - \mu}{\sigma} \right)^k \right] \approx E \left[ \left( Y + \frac{T}{6} \times (Y^2 - 1) \right)^k \right],$$

where $E$ is the expectation operator and $E(X^k)$ is the $k$-th raw moment of $X$. However, with the definitions from (1) and mathematical relations for moments, it can be found that

$$E \left[ \left( \frac{T - \mu}{\sigma} \right)^k \right] = \begin{cases} 0, & \text{if } k = 1 \\ 1, & \text{if } k = 2 \\ \gamma^k, & \text{if } k = 3 \end{cases}$$

while

$$E \left[ \left( Y + \frac{T}{6} \times (Y^2 - 1) \right)^k \right] = \begin{cases} 0, & \text{if } k = 1 \\ 1 + \frac{\gamma^2}{12}, & \text{if } k = 2 \\ \gamma^k, & \text{if } k = 3 \end{cases}$$

The percent error ($\epsilon$) by the approximation in (11), as compared to the exact forms of (12) and (13), is given by

$$\epsilon = 100 \cdot \frac{E \left[ \left( \frac{T - \mu}{\sigma} \right)^k \right] - E \left[ \left( Y + \frac{T}{6} \times (Y^2 - 1) \right)^k \right]}{E \left[ \left( \frac{T - \mu}{\sigma} \right)^k \right]} = \begin{cases} 0, & \text{if } k = 1 \\ \frac{50}{\gamma^2}, & \text{if } k = 2 \\ \frac{100}{\gamma^2}, & \text{if } k = 3 \end{cases}$$

Therefore, the accuracy of the approximation in (2) and (5) depends on the value of skewness $\gamma$, which is a measure of asymmetry of the IPTD probability distribution and can be computed from the measured IPTD sample as follows:

$$\gamma = E \left[ \left( \frac{\text{IPTD} - \mu}{\sigma} \right)^3 \right] = \frac{1}{M} \sum_{i=1}^M \left( \frac{D_i - \mu}{\sigma} \right)^3,$$
that the convolution method is reliable in the aspect of methodology and its results can be used as a reference for evaluating approximation errors of the ITU method. The percent error ($\epsilon$) of the ITU method in comparison with the convolution method is evaluated as follows:

$$\epsilon = \frac{|IPDV_{\text{Conv}} - IPDV_{\text{ITU}}|}{IPDV_{\text{Conv}}} \times 100\%,$$  

(16)

where $IPDV_{\text{Conv}}$ is a value of IPDV computed by the convolution method and $IPDV_{\text{ITU}}$ is a value of IPDV estimated approximately by the ITU method.

3.2 Experimental Analysis and Accuracy Evaluation

The network configuration used for collecting IPTD data is described in Figure 1. In a homogeneous core network domain, it can be assumed that all network sections are identical so that IPTD distributions on the three sections are the same. The samples of IPTD over each section with the size of 100000 are measured during an interval of 24 hours for each test case. This sample size and measurement interval is sufficiently large to include enough packets of the different traffic classes, to reflect a period of typical usage and to cover different delay distribution characteristics depending on different traffic conditions. This also helps the IPTD empirical distribution function to be converted to a true distribution function. A range of IP packet sizes in octets is selected according to IETF RFC 2544 for testing [7]; that is, Test case 1: 64, Test case 2: 128, Test case 3: 256, Test case 4: 512, Test case 5: 1024, Test case 6: 1280, Test case 7: 1518. IPTD data are used to compute, compose and estimate a value of IPDV over a 3-section network by both methods mentioned in Section 2. The granularity of 1 µs is chosen for IPTD bin size (Step 1 in Section 2.2) in order to ensure that errors of discretization process mentioned in Section 3.1 for the convolution method are negligible (on order of microsecond).

Table I shows skewness and moment deviation resulted from experimental data by (14) and (15) respectively.

It can be observed that the skewness of IPTD over the 3-section network is smaller than the skewness of IPTD over each section but both of them are rather large. This causes great errors of distribution moments when NPA is applied, as shown in the column number 4 to 7 in Table I. However, the errors in (14) are calculated for expected values whereas approximation errors of the distribution function may vary according to quantiles. Therefore, the adequate quantiles need to be examined in order to minimize errors of the estimation method.

In this paper, the empirical quantiles in the range of $p \in (0.9, 1.0)$ are used for surveying. The thresholds of 1% to 5% as in Figure 6 and Table II are selected as the reference levels for evaluating the percent error from (16) and they are also used as criteria for selecting the best p-range of quantiles for the method.

Several qualitative and quantitative remarks on experimental results of both methods can be inferred from Figure 2 to Figure 10 and Table II.

Figure 2 shows a statistic of the original IPTD sample data measured from each section of the NGN core network deployed by the VNPT (Vietnam Posts and Telecommunications) as described in Figure 1.

Figure 3 illustrates an empirical probability distribution of the IPTD over each section and another one over the whole 3-section network. These are the results of Step 1 and (8) in Step 2 of the convolution method (see Section 2.2). It can be observed that the IPTD distribution of the whole network is three times as wide as each network section and they do not conform to any well-known function as mentioned in Section 2.2.

Figure 4 describes an empirical cumulative distribution of IPTD over 1-section network and another one over the whole 3-section network. This is resulted from (9). It can be observed that the IPTD cumulative distribution of each section approaches the value of 1 faster than that of the whole network.

Figure 5 shows both values of IPDV over the whole network versus p-quantiles. One of them is computed by the ITU method (Section 2.1) and the other by the convolution method (Section 2.2). It can be observed that there are 3 distinguished segments of quantiles. The first one includes quantiles lower than 0.96, the
second one between 0.96 and 0.999, and the third one higher than 0.999. The results of IPDV are approximately converged in the second segments and diverged in the first and the third segments where the ITU method gives smaller IPDV values than that by the convolution method.

Figure 6 illustrates results of (16) which are the IPDV errors of the ITU method, as compared to the convolution method, versus quantiles. It can be observed that there are also the same segments of quantile as in Figure 5. With the growth of quantiles, the errors gradually reduce quasi-linearly with \( p \) in the first segment, then fluctuate within 5% in the second one, and tend to suddenly increase over 5% with \( p \) in the third segment. The quantiles with \( p < 0.96 \) cause significant errors (e.g., 45% for 0.9-quantile and 12% for 0.95-quantile).

Table II gives us the selected ranges of IPTD quantiles guaranteeing the accuracy of estimation. Broken ranges satisfied the same criteria are separated by semicolons in the table. It can be observed that different \( p \)-values of IPTD quantiles cause different IPDV error levels. In most cases, the range of quantiles ensuring errors within 5% is \( 0.958 < p < 0.9995 \), within 4% is \( 0.96 < p < 0.9994 \), within 3% is \( 0.963 < p < 0.9994 \), within 2% is \( 0.96 < p < 0.974, 0.981 < p < 0.996, \) and \( 0.998 < p < 0.9992 \), and within 1% is \( 0.97 < p < 0.972, 0.983 < p < 0.986 \) and \( 0.999 < p < 0.9991 \). As a
result, the most stable and safe range of IPTD quantiles maintaining negligible errors is $p \in (0.983, 0.986)$ where the central point is located at $p \approx 0.9845$. The results of other test cases are illustrated in Figures 7, 8, 9, 10, 11, and 12.

From the above empirical results, analyses, and remarks, it can be summarized in the following key points:

1) The distribution of IPTD over the NGN core network is not closed to any well-known distribution function.
2) Large skewness value of the distribution of IPTD over the NGN core network causes great error of the NPA.
3) The result of IPDV is considerably dependent on the quantile of the IPTD distribution.
4) A specific rule exists for the relation between the error of the IPDV estimation method and the quantiles of IPTD over the NGN core network.

### Table II

<table>
<thead>
<tr>
<th>Test case</th>
<th>Thresholds for percent error from (16)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5 (%)</td>
</tr>
<tr>
<td>1</td>
<td>0.958-0.9995</td>
</tr>
<tr>
<td>2</td>
<td>0.958-0.9996</td>
</tr>
<tr>
<td>3</td>
<td>0.956-0.9996</td>
</tr>
<tr>
<td>4</td>
<td>0.956-0.9998</td>
</tr>
<tr>
<td>5</td>
<td>0.956-0.9996</td>
</tr>
<tr>
<td>6</td>
<td>0.958-0.9996</td>
</tr>
<tr>
<td>7</td>
<td>0.958-0.9995</td>
</tr>
</tbody>
</table>

Figure 6. Errors of the ITU method against the convolution method.

Figure 7. Results of Test case 2.

Figure 8. Results of Test case 3.
Figure 9. Results of Test case 4.

Figure 10. Results of Test case 5.

Figure 11. Results of Test case 6.

Figure 12. Results of Test case 7.
5) A value of quantile should be selected in the ranges of \( \{ (0.97, 0.972); (0.983, 0.986); (0.999, 0.9991) \} \) for minimizing the estimation error.

In fact, the result of IPDV is dependent on the \( p \)-quantile level of the IPTD distribution. If \( p \) is too low, many significant IPTD samples are not taken into account. Inversely, if it is too high, false samples, which may be generated due to network congestion, abnormal states or data errors, etc., may be involved in the estimation thus causing errors.

4 Conclusion

The convolution method for estimating IPDV over the concatenated networks from empirical data of each network section has been proposed in this paper in order to evaluate the accuracy of the ITU method for the same purpose. Theoretical analyses and experiments are combined together to prove, quantify, and evaluate the errors of the methods, which are dependent on the quantiles of IPTD selected for calculation. An insight view of IPTD quantiles and its effect on accuracy of estimated IPDV have been examined by both qualitative and quantitative analyses. The ranges of quantiles that ensure acceptable accuracy for the method to be applied in the NGN core network are examined and recommended, as a new finding. Although the IPDV estimation method of the ITU has been specified on a provisional basis, and this clause may change in the future based on new findings or real operational experience, the proposed method can still be used for composing and testing different kinds of empirical distributions, which are not possible to be well represented in mathematics. The research in this paper may be extended to other empirical cases such as large-scale network with more than three sections, multicast traffic, different services and traffic classes, heterogeneous network sections, etc.

Acknowledgment

This work has been partly presented at the International Workshop on Communication Systems at Vietnam National University in October 2011. The authors would like to thank the conference organization staff and reviewers for their constructive and helpful comments.

References