Towards Bandwidth Estimation
using Flow-level Measurements

Ricardo de O. Schmidt, Anna Sperotto, Ramin Sadre, Aiko Pras

University of Twente
Design and Analysis of Communication Systems (DACS)
Enschede, The Netherlands
{r.schmidt,a.sperotto,r.sadre,a.pras}@utwente.nl

Abstract. Bandwidth estimation is one of the prerequisite for efficient link dimensioning. In the past, several approaches to bandwidth estimation have been proposed, ranging from rules-of-thumb providing over-provisioning guidelines to mathematically backed-up provisioning formulas. The limitation of such approaches, in our eyes, is that they largely rely on packet-based measurements, which are almost unfeasible considering nowadays load and speed (1–10 Gbps). In this context, flow-based measurements seems to be a suitable alternative, addressing both data aggregation as well as scalability issues. However, flows pose a challenge for bandwidth estimation, namely the coarser data granularity compared to packet-based approaches, which can lead to a lower precision in the estimation of the needed bandwidth. In this paper, we investigate what is the impact of flow-based measurements on bandwidth estimation. In particular, we are interested in quantifying the impact of flows on main statistical traffic characteristics, in particular the traffic rate variance. Our approach is validated on real traffic traces captured from 2002 to 2011 at the University of Twente.

1   Introduction

Bandwidth provisioning is one of the key activities of network management toward achieving desired levels of quality of service. Current practices of provisioning by network operators are based on rules-of-thumb and on rough traffic measurements. For example, traffic averages can be obtained using the Simple Network Management Protocol (SNMP) [1], usually at timescales of 5 minutes. The link provisioning is done by adding to the average obtained via SNMP a safety margin (e.g., 30% of the calculated average). However, as shown in [2] and [3], such practices based on traffic measurements at large timescales (i.e., intervals of minutes) completely overlook traffic variations on smaller timescales due to averaging. Consequently, the link usage and the bandwidth need are not correctly estimated, resulting in poor network performance (in the case of underestimation), respectively waste of resources (in the case of overestimation).

   In general, the key challenge in bandwidth provisioning is to create an adequate model of the current and future traffic behavior. For example, the work in

[4,5] is based on the assumption that the traffic rate in fast networks is Normal distributed. Gaussianity is also exploited in [6]. Although these approaches provide accurate results, they require the knowledge of the variance of the traffic rate on small timescales. The variance, as well as other statistics, are usually obtained from packet traces. Similarly, also non-Gaussian models, such as the Batch Markovian processes used in [7], rely on packet traces for their parameterization.

Considering the high traffic rates observed in today's networks, such as Internet backbones, approaches relying on packet measurements can be considered as problematic. Packet monitoring suffers from scalability and requires powerful and often expensive monitoring devices. Researchers try to circumvent this problem. For example, in [4], an inversion procedure was proposed to indirectly estimate traffic statistics by sampling the buffer content.

In this light, *flow* monitoring seems to be promising approach to the bandwidth estimation problem. A flow is defined as a set of packets sharing common properties and passing through an observation point [8]. Flow monitoring is supported by various software tools, as well as by network equipment [9]. Hence, flow-based approaches are considered to be easily deployable into existing high-speed networks.

However, flow monitoring generally results in an inherent loss of information, because individual network packets are aggregated by flow probes into so-called flow records. Such a flow record only contains a summary of the characteristics of the packets inside a flow. This problem has led us to the following research question: *Is it possible to use flow-level measurements instead of packet-level measurements to obtain traffic statistics?* and, if so, *what is the error in flow-level measurements introduced by the averaging of traffic information?*. In order to answer these questions, we will estimate the bandwidth variation in network traffic from flow measurements and compare the results with those obtained from packet measurements. Furthermore, we will study the impact of the parameters of the flow building operation on the accuracy of the results.

We are not the first to study flow-based approaches for provisioning. Most notably, the authors of [10] estimate statistics of the traffic rate from flow characteristics. The underlying model requires that the network links are not congested and an adaptation to the considered traffic is needed. In [11], a link shared by several users is studied, assuming that the traffic can be described by a superposition of flows. However, only elastic data traffic (TCP connections) is considered. Differently to the work above, our approach is more measurement oriented, focusing on the limitations of flow-level measurements and on the information that can be gained from flow monitoring, while using a simple model of the traffic behavior on flow level.

The rest of this paper is organized as follows. A more detailed description of the problem is given in Section 2. In Section 3, we introduce our approach to estimate the bandwidth variation. In Section 4, our experiments are described and results are presented. Finally, conclusions are drawn in Section 5.

(a) Packet transmission from host A to host B.
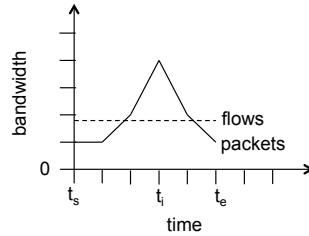
(b) Time series.

Fig. 1: Packet-level vs. flow-level measurements.

2 Problem Definition

Flow-level measurements are a suitable alternative to packet-based measurements in cases in which scalability is a requirement, as it is often the case in high-speed environments with traffic rates of several Gbps. However, scalability does not come for free: since flows provide an aggregated view of streams of packets, we are dealing with information at a coarser granularity compared to packet-based measurements. Particularly and importantly for the case of bandwidth estimation, flow monitoring does not provide us any information about the timing distribution of packets within the flow. Typically, the flow record exported by the flow probe specifies, among others, the total number of packets and bytes, and the observation time of the first and last packet of the flow.

Figure 1 presents an example comparing packet and flow measurements. Host A sends a stream of packets of constant size S to host B over a link from time t_s to time t_e. Around time t_i host A generates a burst of packets, which would be correctly reported by a packet-based measurement. From a flow perspective, the stream of packets from host A to host B is seen as a single flow of n packets with duration $t_e - t_s$. Consequently, we would observe a constant bandwidth usage equal to $nS/(t_e - t_s)$ bytes/sec, and the traffic burst would be canceled out, as shown in the example time series of Figure 1(b).

To overcome this problem, in our approach we assume the bytes to be *uniformly* distributed within the duration of the flow. Such an assumption is critical for the problem of bandwidth estimation, since it can have a direct impact on the measured bandwidth. Under the assumption of uniform distribution, each flow record contributing to the bandwidth calculation also introduces a measurement error compared to the real traffic situation. The aim of this paper is to quantify the impact of such an error in the overall statistical characteristics of the traffic.
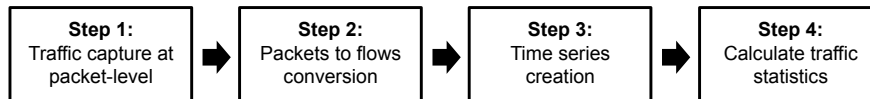
| Step 1:
Traffic capture at
packet-level | Step 2:
Packets to flows
conversion | Step 3:
Time series
creation | Step 4:
Calculate traffic
statistics |

Fig. 2: The steps of the approach.

3 Flow-based Approach

In this section we describe our approach for answering the research questions presented in Section 1. We begin with a general description of the approach and its components, followed by the procedure for creating flow-based time series and extracting the desired statistics from them.

3.1 Approach Structure

As already mentioned, a flow is a set of packets that have common properties and pass through an observation point in the network. Depending on how flow probes are implemented, a single observed flow can be exported as several flow records. The number of exported flow records reporting a single flow will be determined by the exact flow definition used by the flow probe, as further explained in Section 4.2. Usually, flow records contain information such as: start and end time, number of packets and bytes, source and destination IP addresses and ports, and others. The main goal of our approach is to estimate the variance of the traffic rate from flow-level measurements.

Figure 2 illustrates the different steps of our approach. Traffic data captured on packet level in step 1 is first converted into flow records in step 2. In the third step, we create the traffic rate time series. This procedure is detailed in Section 3.2. We calculate the traffic statistics from the resulting data in step 4. Importantly, in the first step we use packet-level captures done with tools like *tcpdump* because our implementation currently operates offline. These captures are stored in *pcap* files and converted into flow records in the second step. For this conversion we use the tool YAF (*Yet Another Flowmeter*) [12].

3.2 Creating Flow-level Time Series

In this section we detail the procedure of creating flow-level time series. For this, we only need the following information from the flow records: the start time s, the end time e and the number of bytes b. Initially, the flow records are placed in a timeline that lasts from s of the first flow record to e of the last flow record in the measurement period.

A flow time series T_F is a sequence of values that represent the link usage (i.e., traffic rate) during the entire measured period. The timeline is defined as $T_F = \{B_{F,i}\}, 1 \leq i \leq m$, where $B_{F,i}$ is the link usage in the time interval $t_i = [i\Delta, (i+1)\Delta[$. The interval size Δ defines the time granularity for the following computations (i.e., timescale).

In order to create T_F, we have to calculate for each time interval t_i the amount of bytes transfered in that interval, under the assumption that the bytes belonging to a flow (record) are uniformly distributed over the flow duration. Let's consider a set of flow records F that contains n records f_j, $1 \leq j \leq n$. The subset of flow records that contribute to the traffic rate in time interval t_i is given by

$$\kappa_i = \{f_j \in F : s_j < (i+1)\Delta \wedge e_j \geqslant i\Delta\}.$$

Then, we calculate for each flow record in κ_i the amount of bytes it contributes to the time interval t_i. The sum over all flow records in κ_i gives the total number of bytes S_i transferred in t_i

$$S_i = \sum_{f_j \in \kappa_i} \frac{\max\left(\min((i+1)\Delta, e_j) - \max(i\Delta, s_j), 1\right)}{\max(e_j - s_j, 1)} \cdot b_j \tag{1}$$

The above formula accounts for the fact that flows do not necessarily start (end) exactly at the beginning (end) of the time interval t_i. It considers the fraction of bytes in relation to the flow record duration within the specific interval. The max function in the divisor avoids a division by zero when the flow record only contains one packet, i.e., the start time is equal to the end time. The flow-based link usage $B_{F,i}$ in t_i is then given by dividing S_i by the size of the chosen timescale

$$B_{F,i} = \frac{S_i}{\Delta} \tag{2}$$

Analogously, we can create a packet timeseries T_P from the packet data. For a time interval t_i, $B_{P,i}$ gives the overall number of bytes of all packets that have been monitoring in that interval, divided by the timescale Δ. In the following, we express $B_{F,i}$ and $B_{P,i}$ in Mbps.

3.3 Metrics of Interest

In this paper, we are interested in assessing the impacts of averaging in flows on the main statistical characteristics of the traffic. Since all packets are accounted in the flow measurements, it will clearly be able to correctly estimate the mean bandwidth usage $\overline{B}_F$. In our experiments we particularly focus our attention on the following metrics of interest.

Variance of the traffic rate: The averaging introduced by the use of flow measurements can have an impact on the variance of the time series T_F. This is due to the fact that short traffic peaks can be evened out by the assumption that the bytes are uniformly distributed over the duration of a flow record. The variance V_F of the flow time series is defined as

$$V_F = \frac{1}{m} \sum_{i=1}^{m} \left(B_{F,i} - \overline{B}_F\right)^2. \tag{3}$$

For the packet data, V_P is defined accordingly.

Variance error: We investigate the error introduced by the flow-based approach by calculating the variance relative error, expressed in percentage

$$V_e rr = \frac{|V_P - V_F|}{V_P} \cdot 100\%. \tag{4}$$

Bandwidth error: In addition to the error of the variance, we are also interested in the error introduced per time interval t_i, i.e. the average bandwidth error calculated by comparing the packet-based and flow-based time series. We define:

$$\overline{B}_{err} = \frac{1}{m} \sum_{i=1}^{m} |B_{P,i} - B_{F,i}|. \tag{5}$$

4 Experiments

This section presents the experimental results obtained when using flow-level traffic time series. First, we describe the measurement dataset used in our analysis. Then, we discuss how different flow definitions may impact on the time series. Finally, we present our analysis methodology and our results.

4.1 Traffic Traces

The dataset used to validate our approach contains hundreds of 5-minute traces measured in three different links. These traces have been collected in 2002, 2003 and 2011. The measurements from 2002 and 2003[1] have already been used for validation in [4,5], and the 2011 traces were collected for these experiments.

Traces from 2002 have been captured on a 300 Mbps Ethernet link that connects a residential network to the core network of a university. At the time of the measurements, the network segment had about 2000 users, each with a 100 Mbps access link, and the link average usage was around 60%. We will refer to this set of traces as *Location 1* (abbr. *loc1*).

Traces from 2003 have been collected on a 1 Gbps link connecting a large college to the Dutch academic and research network. At that time, about 1000 users had access to the link, each with a 100 Mbps access link. The average usage of the measure link was between 10% and 15%. We refer to this set of traces as *Location 2* (abbr. *loc2*).

Finally, traces from 2011 consist of the incoming/outgoing traffic of a university's building, captured on the uplink and downlink connecting the building to the university's main router. Both links allow for 1 Gbps traffic and they have an average usage of around 15%. We will refer to this set of traces as *Location 3* (abbr. *loc3*). Table 1 summarizes our dataset.

---

[1] The anonymized traces are available at `http://traces.simpleweb.org`

Table 2: Flow definitions.

Table 1: Overview of measurements dataset.

abbr.	descr.	period	# hosts	avg. usage
loc1	residential	2002	2000	60%
loc2	college	2003	1000	10–15%
loc3	univ. building	2011	1000	15%

abbr.	timeouts (sec.)	
	active	*inactive*
a2i1	2	1
a5i2	5	2
a15i5	15	5
a30i10	30	10
a60i20	60	20
a120i30	120	30
a300i50	300	50

4.2 Flow Definitions

The flow creation process, which converts packet-based measurements into flows, can influence the precision of the flow-based time series as created by the approach described in Section 3.2. Flow records are usually exported on the basis of timers, namely *active* and *inactive* timeouts. Such timeouts are parameters by which we can obtain shorter or longer flow records. The active timeout controls the maximum duration of a flow record before it is exported. Inactive timeout, on the other hand, is used to decide if a flow can be considered definitely idle, and therefore exported.

In our experiments we generate flow-level time series using different combinations of timeouts, as listed in Table 2. The different timeouts allow us to analyze the loss on traffic information due to averaging at the flow-level and, at the same time, they control the number of flow records we should analyze. When timeouts are considerably long, the traffic averaging may have a significant impact, since the real packet arrival rate is smoothed out for the entire duration of the flow. Consequently, significant traffic fluctuations may completely vanish from the time series. On the other hand, the number of flows to be processed is reduced.

With shorter timeouts, the effect of averaging will be limited because longs flows are divided into multiple flow records. This way, traffic fluctuations may still be present in the flow-level time series. But, due to a larger number of exported flows, there may also be an additional overhead. Our aim is to quantify the error introduced by traffic averaging for different combinations of timeouts.

4.3 Methodology

In this paper, we compare flow-based time series with a *ground truth* consisting of the packet-based time series for the same traffic. The methodology adopted to analyze the traffic traces is structured according to the following two steps:

1. **Qualitative analysis**: the first step consists in the manual comparison of the flow-based and packet-based time series. The aim of this step is to identify situations in which these time series show a different behavior, and to

investigate the causes for such a deviation. Such analysis is conducted for different flow timeouts, as explained above.

2. **Error analysis**: the second step targets the quantitative analysis of the statistical characteristics of the flow-based time series compared to the packet-based one. For this analysis, we compute the main metrics of interest, described in Section 3.3. Also in this case, we investigate the effect of the flow timeouts on the statistical metrics.

4.4 Qualitative Analysis

Due to space issues, we present the qualitative analysis for a single trace from *Location 2* only. We chose this particular trace because it illustrates well all the significant differences between packet-level and flow-level time series. To facilitate explanations and comparisons, the following plot only contain flow-level time series of two flow definitions: *a5i2* and *a120i30*. Time granularity is $\Delta = 1$ second.

Figure 3 illustrates the differences between packet-level and flow-level time series. At the beginning of the time series, one can clearly see the impact of the averaging process. The situation between *12:05:01* and *12:07:01* illustrates the problems of averaging when defining flows with long timeouts. During this period, most traffic was generated by a single flow and, for the definition *a120i30*, it was exported as a single flow record. As one can see from the packet-level time series, the traffic increased significantly at around *12:05:40*. That is, the major portion of bytes was transmitted few seconds after the beginning of the flow. Ignoring the actual distribution and assuming uniformity, our approach averages all the bytes within the entire flow duration.
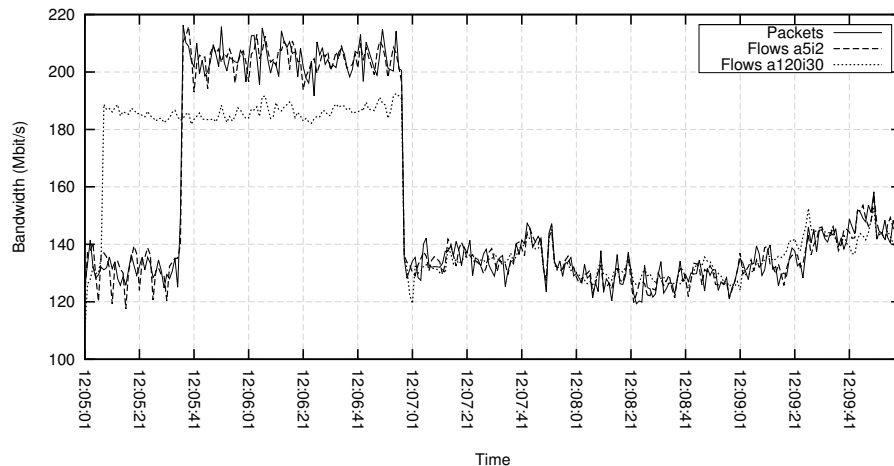

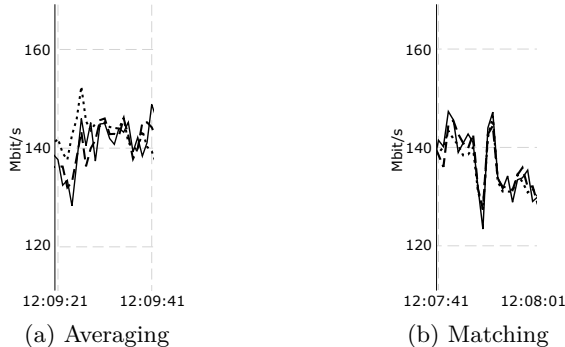Fig. 3: Time series for trace from *Location 2*.

Fig. 4: Intervals of the time series for trace from *Location 2*.

Comparing the two flow time series, we observe that the *a5i2* series closely follows traffic fluctuations. The reason is that, unlike *a120i30*, the *a5i2* definition exports long flows as several consecutive and smaller flow records. It means that the averaging of traffic is limited to intervals of 5 seconds (the active timeout) at maximum.

The problem of averaging in flows is present again at *12:09:21 – 12:09:41*, shown in isolation in Figure 4(a). In this period, there is a sudden drop in the traffic rate for a few seconds. The flow-level time series do not follow this drop because they take into account bytes of flows that started before and ended after this specific time bin. As can be seen in the figure, this effect is stronger for larger timeouts.

A third interesting situation is presented in Figure 4(b), extracted from the period *12:07:41 – 12:08:01* of the trace. In this case, both flow definitions *do* follow remarkably well the traffic fluctuations occurring on packet level. This happens when (i) flows have a constant traffic rate, i.e., the flow records represent well the packet-level dynamics, or when (ii) flows are short (including flows consisting of only one packet) and fit into one time interval.

4.5   Error Analysis

This section presents the traffic statistics calculated from flow-level time series and compares them with those from packet-level time series. Note that we give all results in Mbps.

Table 3 gives the results for the traces from *Location 1* for three different timescales Δ (1, 10 and 30 seconds). For each timescale, we show the average bandwidth error $\overline{B}_{err}$ between the flow and the packet timeseries, the variance V of the traffic rate, and the relative error V_{err} of the variance. One can observe that the best approximation to the packet timeseries is mostly obtained by the flow timeseries with shorter timeouts because they reproduce traffic fluctuations better. Compared to the mean bandwidth of around 277.4 Mbps, the average

Table 3: Statistics summary for the trace from *Location 1*.

source	1s			10s			30s		
	$\overline{B}_{err}$	V	$V_{err}(\%)$	$\overline{B}_{err}$	V	$V_{err}(\%)$	$\overline{B}_{err}$	V	$V_{err}(\%)$
packets	-	137.11	-	-	31.70	-	-	20.11	-
a2i1	3.3402	107.49	21.60	0.4580	31.15	1.74	0.2900	19.52	2.93
a5i2	6.3638	92.78	32.33	0.8587	28.89	8.86	0.6420	17.46	13.18
a15i5	6.5670	83.34	39.22	1.6737	28.23	10.95	0.5070	18.04	10.29
a30i10	6.9092	83.13	39.37	2.9030	31.00	2.21	0.7250	17.10	14.97
a60i20	7.0905	77.54	43.45	3.8253	29.64	6.50	2.7210	12.52	37.74
a120i30	7.0022	73.91	46.09	3.6260	30.77	2.93	2.4570	15.98	20.54
a300i50	6.8888	84.66	38.25	3.9047	49.71	56.81	3.0890	33.51	66.63

bandwidth error is very small, although the variance error is much higher. In addition, we see that the variance and the errors generally decrease with increasing timescale Δ. This is expected because larger timescales smooth short-term fluctuations in the traffic rate.

Table 4 gives the results for *Location 2*. For the first five timeout settings, the variance error is significantly lower than in the first trace, even for small Δ. This also holds for $\overline{B}_{err}$ which is small compared to the mean traffic rate of 153 Mbps. In contrast, we observe large errors for *a120i30* and *a300i50*. This is due to the phenomenon described in Section 4.4.

Finally, Table 5 presents the statistics for *Location 3*. These traces have a much higher variance than the other ones (the mean bandwidth is 440 Mbps) and the variance also remains on larger timescales. Consequently, we observe that the errors quickly increase for longer timeouts. However, for small timeouts, the results are still acceptable.

Table 4: Statistics summary for the trace from *Location 2*.

source	1s			10s			30s		
	$\overline{B}_{err}$	V	$V_{err}(\%)$	$\overline{B}_{err}$	V	$V_{err}(\%)$	$\overline{B}_{err}$	V	$V_{err}(\%)$
packets	-	1042.75	-	-	968.30	-	-	983.27	-
a2i1	2.4249	1037.74	0.48	0.2507	967.95	0.04	0.0590	984.02	0.08
a5i2	3.4849	1036.74	0.58	0.3850	967.18	0.12	0.1450	984.04	0.08
a15i5	3.5848	1039.64	0.30	1.0290	964.88	0.35	0.1470	978.84	0.45
a30i10	3.8320	1053.13	1	1.8563	976.22	0.82	0.5120	975.37	0.80
a60i20	3.6490	1053.08	0.99	1.9730	989.39	2.18	0.9880	999.68	1.67
a120i30	12.6345	656.16	37.07	10.8620	616.19	36.36	9.1510	611.76	37.38
a300i50	12.5833	687.18	34.10	10.9383	652.53	32.61	9.3790	653.35	33.55

Table 5: Statistics summary for the trace from *Location 3.*

source	1s			10s			30s		
	B_{err}	V	$V_{err}(\%)$	B_{err}	V	$V_{err}(\%)$	B_{err}	V	$V_{err}(\%)$
packets	-	35543.35	-	-	30045.00	-	-	29643.97	-
a2i1	26.9391	33835.18	4.81	2.3910	29260.62	2.61	0.9589	27215.77	8.19
a5i2	38.1607	32617.15	8.23	4.0286	28989.83	3.51	1.7722	27188.32	8.28
a15i5	62.0467	29804.32	16.15	30.5776	27130.78	9.70	6.3456	27437.85	7.44
a30i10	74.7265	27710.80	22.04	51.7083	27264.21	9.26	16.4878	27597.72	6.90
a60i20	100.6202	20264.41	42.99	80.7528	19936.01	33.65	74.1400	20093.97	32.22
a120i30	90.0240	20017.42	43.68	69.9117	20192.62	32.79	62.4544	20663.48	30.29
a300i50	95.4344	13331.16	62.49	75.4097	13352.82	55.56	65.5011	13427.02	54.71

5 Conclusions

The research presented in this paper aimed to evaluate whether flow-level time series could be used in place of packet-level time series for extracting traffic statistics. Since flow monitoring generally results in a loss of knowledge of the arrival times of individual packets, we proposed an approach for creating flow-level time series assuming that traffic is uniformly distributed within the flow. We validated the performance of the approach by means of experiments with several empirical traffic traces. In addition, we studied the influence of the parameters (timeouts) of the flow building process on the results.

The results suggest that our approach is promising, that is, it allows to estimate the variance of the traffic bandwidth with low errors compared to packet-level time series. This is especially true when flow timeouts are small. In future work we are going to embed our approach into bandwidth estimation and link dimensioning methods, such as [4,5], and assess whether applying estimated variance from flows into dimensioning formulas will still lead to good results on link dimensioning. In addition, we want to identify whether traffic distribution inside flows changes accordingly to different properties, like transport protocol and flow duration. If significant differences are observed, improvements can be expected by refining the uniformity assumption depending on such properties.

References

1. Schönwalder, J.: Simple Network Management Protocol (SNMP) Context EngineID Discovery. RFC 5343, 2008.
2. Meent, R. van de, Pras, A., Mandjes, M., Berg, H. van den, Nieuwenhuis, L.: Traffic Measurements for Link Dimensioning, A Case of Study. In proceedings of the 14th IFIP/IEEE Workshop on Distributed Systems: Operations and Management (DSOM), pp. 106–117, 2003.
3. Papagiannaki, K., Cruz, R., Diot, C.: Network Performance Monitoring at Small Time Scales. In proceedings of the Internet Management Conference (IMC), 2003.

4. Meent, R. van de: Network Link Dimensioning - a measurement & modeling based approach. PhD thesis, University of Twente, 2006.
5. Pras, A., Nieuwenhuis, L., Meent, R. van de, Mandjes, M.: Dimensioning Network Links: A New Look at Equivalent Bandwidth. IEEE Network, 23(2), pp. 5–10, 2009.
6. Fraleigh, C., Tobagi, F., Diot, C.: Provisioning IP Backbone Networks to Support Latency Sensitive Traffic. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications (INFOCOM), 2003.
7. Klemm, A., Lindemann, C., Lohmann, M.: Modeling IP traffic using the batch Markovian arrival process. Performance Evaluation, 54, pp. 149–173, 2003.
8. Quittek, J., Zseby, T., Claise, B., Zander, S.: Requirements for IP Flow Information Export (IPFIX). RFC 3917, 2004.
9. Cisco Systems Inc.: Cisco IOS Flexible NetFlow Configuration Guide. Release 12.4T, 2008. Available at: http://www.cisco.com/. Accessed in: January 2012.
10. Barakat, C., Thiran, P., Iannaccone, G., Diot, C., Owezarki, P.: Modeling Internet backbone traffic at the flow level. IEEE Transactions on Signal Processing, 51(8), 2003.
11. Bonald, T., Olivier, P., Roberts, J.: Dimensioning high speed IP access networks. In proceedings of the 8th International Teletraffic Congress (ITC), pp. 241–251, 2003.
12. Inacio, C. M., Trammell, B.: YAF: Yet Another Flowmeter. In proceedings of the 24th Large Installation System Administration Conference (LISA), 2010.
13. SimpleWeb. Available at: http://www.simpleweb.org/. Accessed in: January 2012.