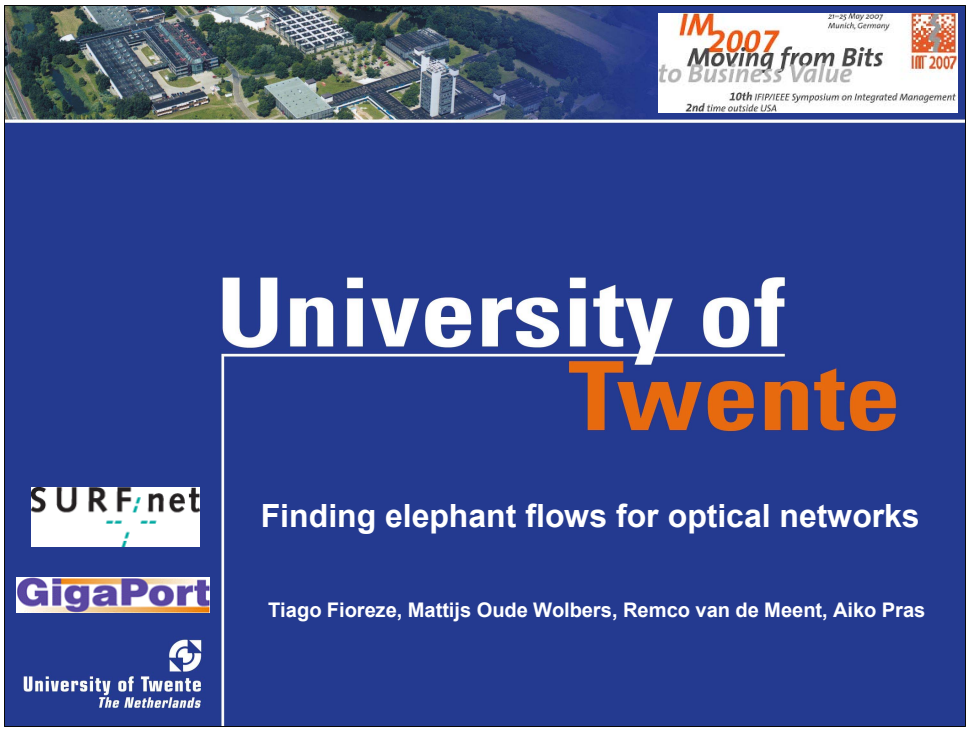
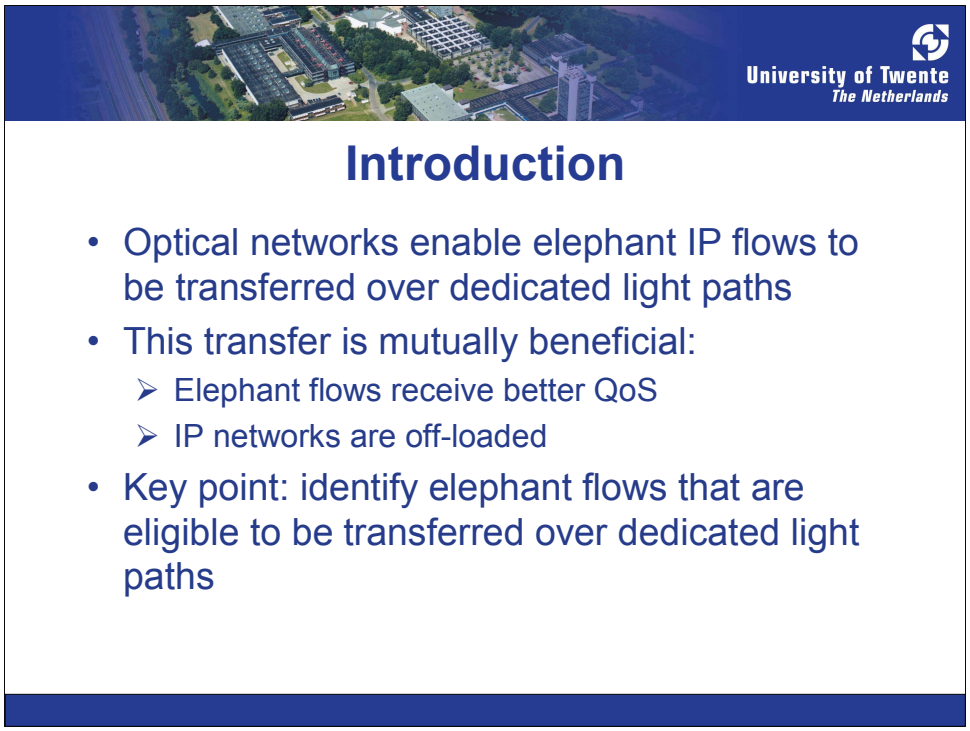


Abstract

Optical networks are fast and reliable networks that enable, amongst others, dedicated light paths to be established for elephant IP flows. Elephant IP flows are characterized by being small in number, but long in time and high in traffic volume. Moving these flows from the general IP network to dedicated light paths can be beneficial for both the elephant flows as well as the general IP network. Elephant flows over light paths would benefit from receiving better Quality of Service (at the optical level there is no jitter and far more bandwidth) and, at the same time, IP networks would be off-loaded and therefore offer better Quality of Service to the remaining, smaller IP flows. Identifying elephant flows in large scale IP networks is therefore an important task in order to effectively manage the network. In practice such flows are generally characterized using 5-tuple flow definition (source/destination address/port and protocol), which may be too restrictive for the purpose of establishing optical light paths. In this paper we evaluate different flow definitions at different levels of granularity. Using measurements at a large national research network, we compare our alternative flow definitions to the traditional 5-tuple definition. We show that the discovery of elephant flows eligible to be transferred over light paths can better be reached using less restrictive flow definitions.

1-4244-0799-0/07/$25.00 ©2007 IEEE

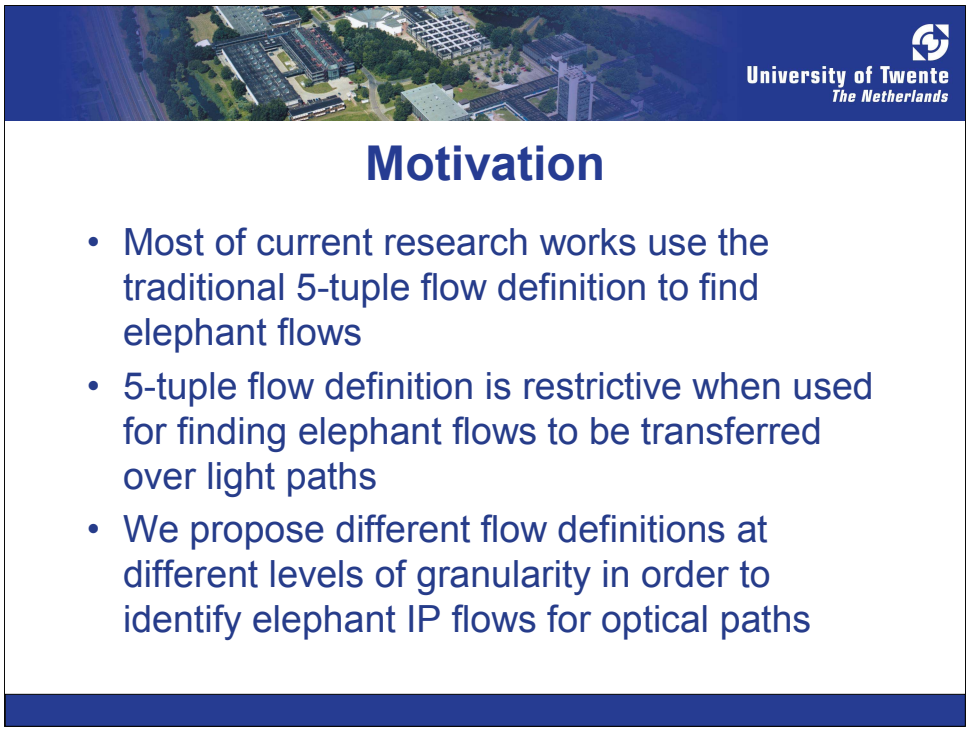


1. Introduction

Optical networks allow huge amounts of data to be transferred over light paths (lambda-connections) in a fast and reliable way through modern multi-service optical switches. Multi-service optical switches are capable of performing data forwarding decisions at different network levels, which enables therefore the transfer of sets of data packets (IP flows) at optical-level (physical layer) instead of at packet-level (network layer).

This transfer may be beneficial when resources in IP networks (e.g., bandwidth) are mostly consumed by a few number of flows, which are known as elephant flows. Studies [1][2] show that few elephant flows contribute to most of the traffic volume in IP networks while most of the other flows (mice flows) are responsible for a small part of the generated traffic. By transferring elephant flows to the optical level, congested IP networks might be therefore off-loaded and offer better services to other flows. At the same time, elephant flows would receive better Quality of Service (no jitter and plenty of bandwidth) by being transferred over dedicated light paths.

The key problem is thus to identify which flows are eligible to be transferred over light paths. Since elephant flows are known to have a high traffic volume and a considerable persistence in time [3], this paper looks therefore for flows that satisfy the requirements long duration and big size.

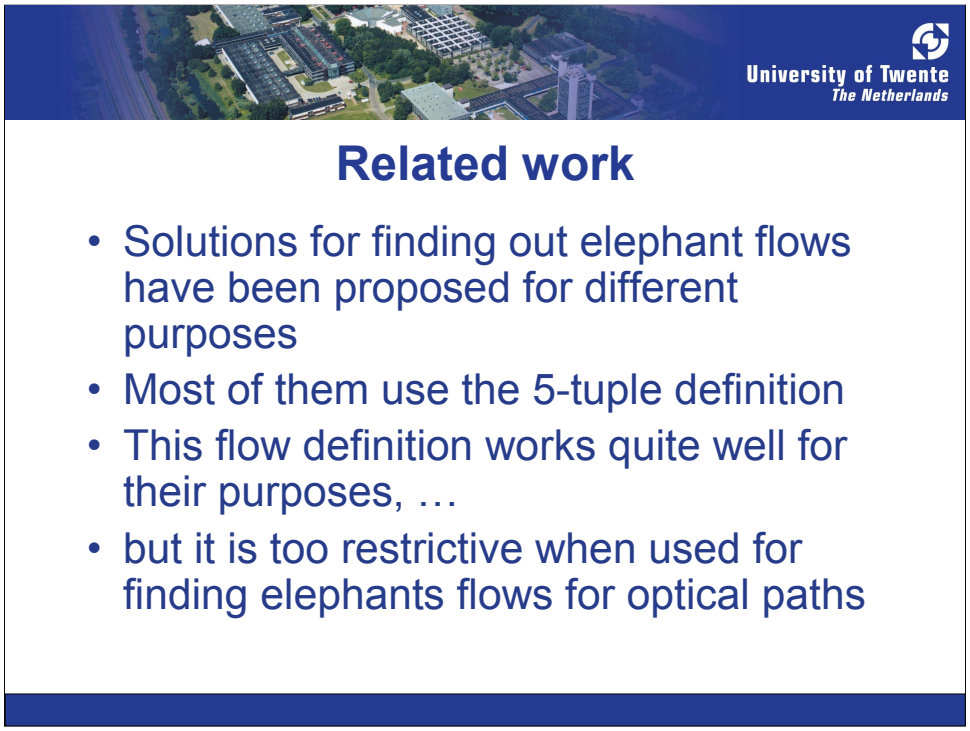


1.1 Motivation

Flows may be defined at different levels of granularity. The higher granular a flow definition is, the more details a flow definition has. An example of flow definition with a high level of granularity is the 5-tuple flow definition. This flow definition consists of characterizing a flow as a set of IP packets that contains the common 5 properties source port, destination port, source address, destination address, and IP protocol.

In most papers the identification of elephant flows is based on the traditional 5-tuple definition [4][5][6]. However, as we are going to show, flow definitions with a high level of granularity are more restrictive when grouping IP packets into flows. As a result of that, the 5-tuple flow definition is considered too restrictive when looking for elephant flows that satisfy our requirement to be transferred over dedicated light paths.

In this paper we present different flow definitions at different levels of granularity to find out elephant flows in optical networks. In addition, we also evaluate our proposed flow definitions with the traditional 5-tuple flow definition by observing the percentage of bytes eligible to be transferred over light paths. The main contribution of this work is to show the quantification of IP traffic to be transferred to the optical level by using flow definitions at different levels of granularity.



2. Related work

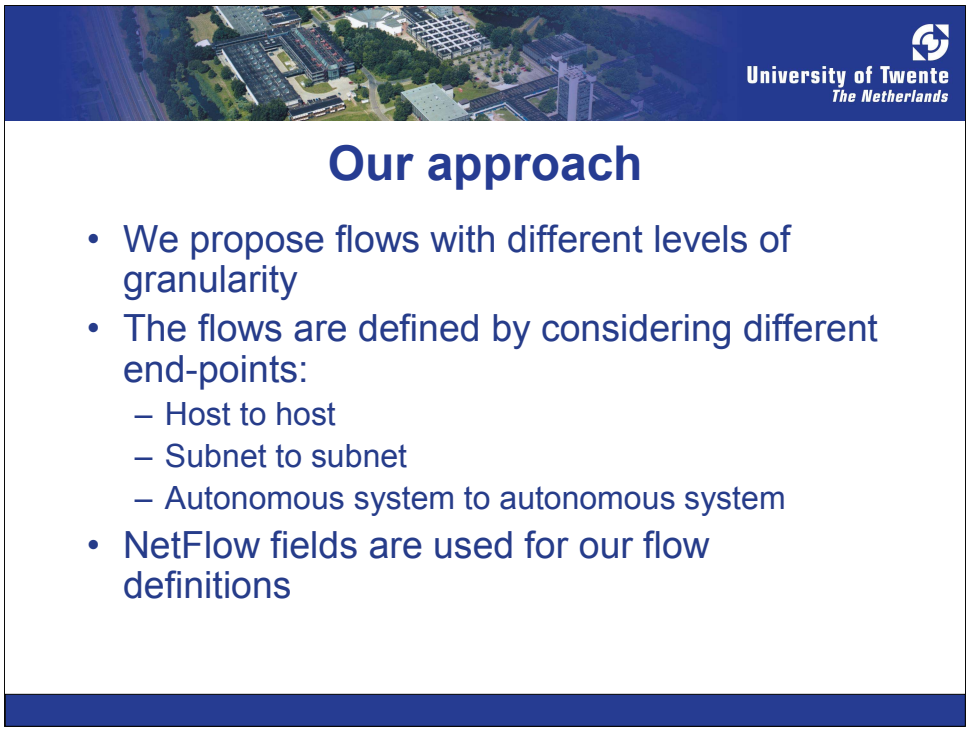
The search for elephants flows in large-scale optical networks has also been addressed in other works.

Mori et al [4] propose to identify elephant flows by defining thresholds of sampled packets for a single flow. The threshold values are obtained based on Bayes' theorem and they define whether a flow is an elephant flow or not based on the number of sampled packets per flow. If the number of packets is greater than the defined threshold, the flow is identified as an elephant flow. Mori et al's approach relies on the premise that elephant flows have a large number of packets. Based on that, their main concern is to find the right threshold in order to find a proper trade-off between misidentified elephant flows and missed elephant flows. In their work the traditional 5-tuple flow definition is used to evaluate the approach.

Wallerich et al [5] focus on the analysis of the persistency properties of elephants flows in order to identify them. The authors use NetFlow data and packet-level traces to perform the study on the elephant flows persistency by considering different time intervals (between 1-10 minutes) and flow definitions (5-tuple and prefix flow definitions). The occurrence of these flow definitions within the time intervals defines how persistent one elephant flow is.

In another related work, Estan and Varghese [6] look for elephant flows by not considering small flows (mice flows) to be stored in cache (SRAM memory). They perform that by using two novel techniques, referred to as *sample-and-hold* and *multistage filters.* Their approach consists of allocating space in cache only for those flows that are bigger than the measured traffic in cache. Different flow definitions are also used in Estan and Varghese work.

One can say that in most related work they do not focus on different flow definitions, but use a 5-tuple flow definition in order to find elephant flows. Although this works well for their purposes, it is too restrictive when looking for elephant flows to be transferred over dedicated light paths, as we are going to show in the coming slides.



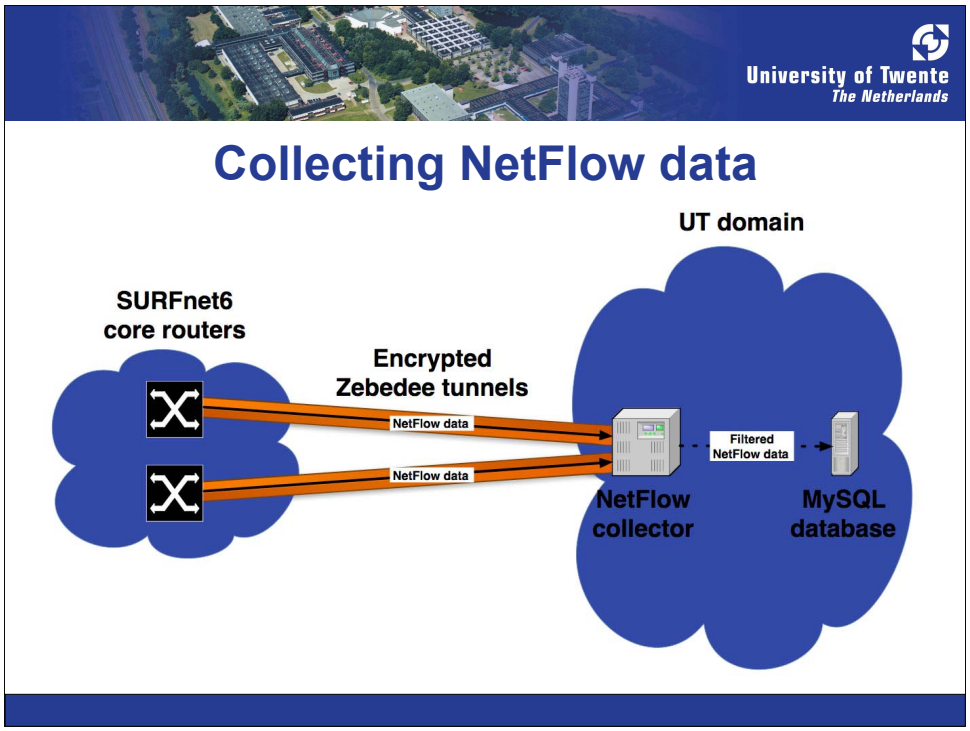
3. Proposed flow definitions

A flow can be defined as a set of packets that match the same properties (e.g., source/destination TCP ports). The amount of properties that a flow has defines its level of granularity. For example, the 5-tuple flow definition is considered as a high granular flow definition because it is defined with a considerable amount of properties. Whereas, the more properties (details) a flow definition has, the more restrictive a flow will be in order to group packets into it.

In our approach, we look for elephant IP flows by using three different flow definitions with different levels of granularity: Host to Host (***Hst2Hst***), Subnet to Subnet (***Sub2Sub***), and Autonomous System to Autonomous System (***AS2AS***). In the case of the ***Sub2Sub*** definition we "simulate" subnets by using different prefix lengths (e.g., /8, /16, and /24). In addition, these flow definitions take into account different flow end-points and they are defined based on the set of fields types provided by NetFlow version 9 [10].

Hst2Hst flow definition consists of grouping a set of packets with the same source and destination IP addresses. ***Sub2Sub*** flow definition groups packets with the same source and destination address prefixes, i.e., with the same number of most significative bits of the source and destination addresses. The last but not the least, the ***AS2AS*** flow definition groups flows with the same source and destination autonomous systems. For the sake of the simplicity, we are going to consider the 5-tuple definition as Application to Application (***App2App***) from now on in this paper.

The next subsections show in more details how network traffic information was collected and stored in our database, as well as how network analysis were performed.



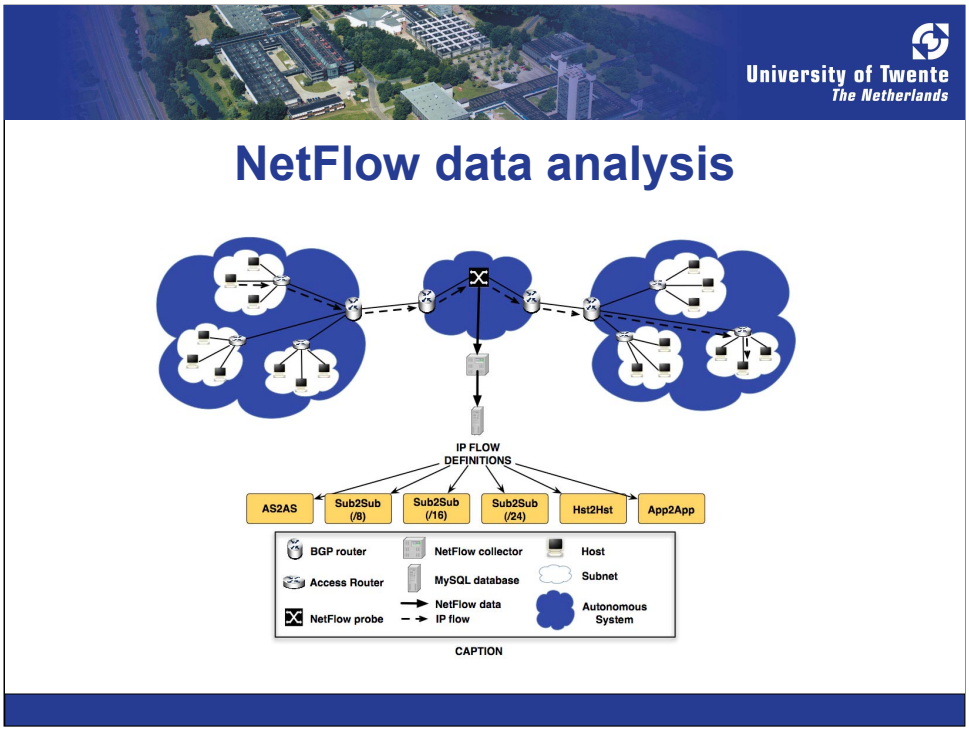
3.1 The collecting of NetFlow data

NetFlow data was collected from the two core routers in SURFnet6 network (our test bed), which is the network built in the RoN GigaPort project [7]. All the SURFnet6 traffic pass through these two core routers.

Since SURFnet6 network is a high-speed network (Gbps links), the collecting of each packet arriving at the SURFnet6 core routers is unfeasible due to high amount of data. In order to solve this issue, a sampling approach provided by Cisco called sampled NetFlow was used. This approach consists of selecting 1 packet out of **n** packets. The sampling used was 1 per 100 packets sampling, which provides NetFlow data with 1 percent of total traffic in SURFnet6.

The sampled network traffic was exported by the core routers to one NetFlow collector located at the University of Twente (UT) domain. Since NetFlow does not have any encryption features, Zebedee [8] was used to provide encryption in order to protect the sensitive exported data.

The data was collected in its "raw" state, i.e., without any aggregation or filtering by using tcpdump [9]. The collecting duration was 1440 minutes (i.e., 24 hours) which gave a total of 11.64 GBytes of NetFlow data. This collected NetFlow data was then filtered and exported into a MySQL database for analysis. A filtering process was necessary to decrease the size of the analysed data and to remove some unnecessary fields (e.g. IP TOS). The following NetFlow fields were considered in our analysis: SOURCE, FIRST_SWITCHED, LAST_SWITCHED, IPV4_SRC_ADDR, IPV4_DST_ADDR, SRC_AS, DST_AS, SRC_MASK, DST_MASK, BYTES and PACKETS.



3.2 The NetFlow data analysis

In order to make the analysis 4 flow definitions were used: ***App2App, Hst2Hst, Sub2Sub,*** and ***AS2AS***. We also considered the usage of the values /8, /16, and /24 for address prefixes in the ***Sub2Sub*** flow definition. All these flow definitions are represented in the picture above.

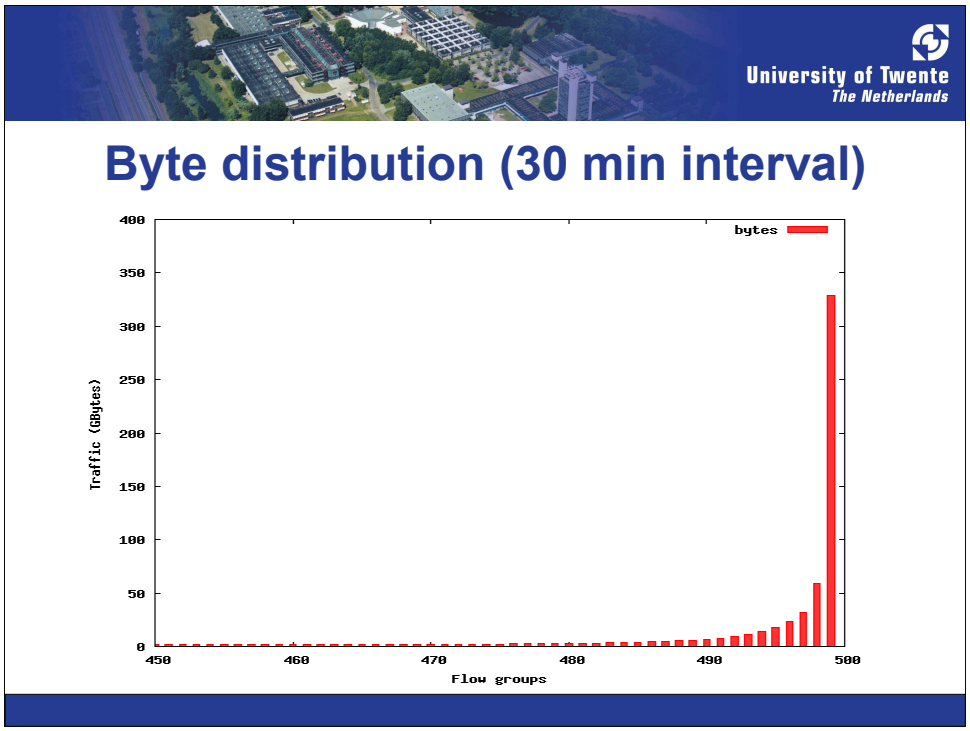
In addition, the total collected traffic was aggregated in 30 minutes (1800 seconds) intervals. As a result of that, there is a loss of information concerning the time resolution of the analyzed flows, which made flow throughput not be considered in our analysis, but total traffic of bytes per flow.

In order to find big and long flows the criteria used in our evaluation was to select flows that are bigger or equal to the minimal unit of transmission in SONET (*Synchronous Optical Networking*) networks (OC-1 = 50.112 Mbps) in a 30 minutes interval.

Based on this criteria, our main goal was to analyze the percentage of bytes to be transferred over light paths by using different flow definitions. To achieve this goal we consider a flow as eligible for a light path whether its total amount of bytes (traffic) is equal or bigger than 11 GBytes (50.112 Mpbs x 30 min).

Since sampled NetFlow data was used, the results that we are going to show are an estimation of the actual traffic volume. As a sampling of 1 out of 100 packets was used, it is reasonably safe to estimate the actual size of the flows by multiplying the reported length by 100.

In the next couple of slides the results of our analysis are going to be presented.



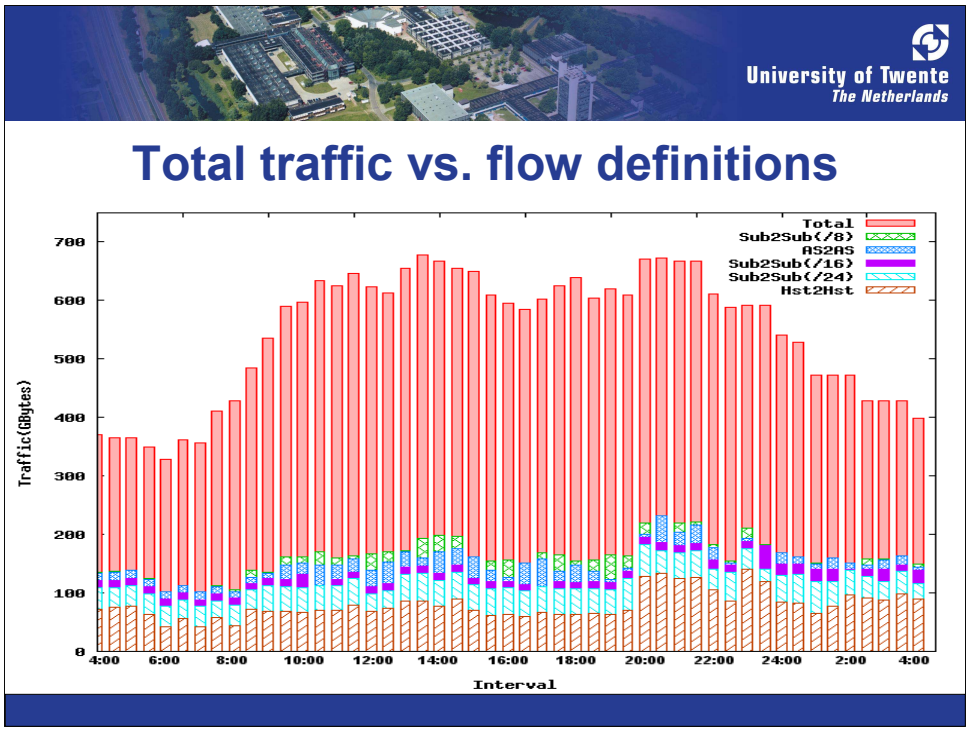
4. The results

This section presents the results of our evaluation. The section starts by showing the byte distribution of our collected data. Then the amount of total data generated by the different flow definitions over 24 hours is presented. Following that, we show how elephant flows are mostly distributed in our 30 minutes intervals. Finally, we show the percentage of bytes that would be transferred over light paths by using the different flow definitions.

4.1 Byte distribution in the collected data

In order to show that the byte distribution of our collected data is similar to the byte distribution found in the literature, we ordered in size (y-axis) 30 minutes of collected Cisco flows and summed them together (x-axis). This resulted in a 500 groups of flows with their sum of bytes they represent.

As the graph shows, the large amount of data is transmitted by few flows, whereas most of the other flows transmit a small amount of data. This proves our hypotheses that improvements made by the transfer of big flows to optical level will have a large impact on the system overall, since they account for a large percentage of the total amount of bytes processed.



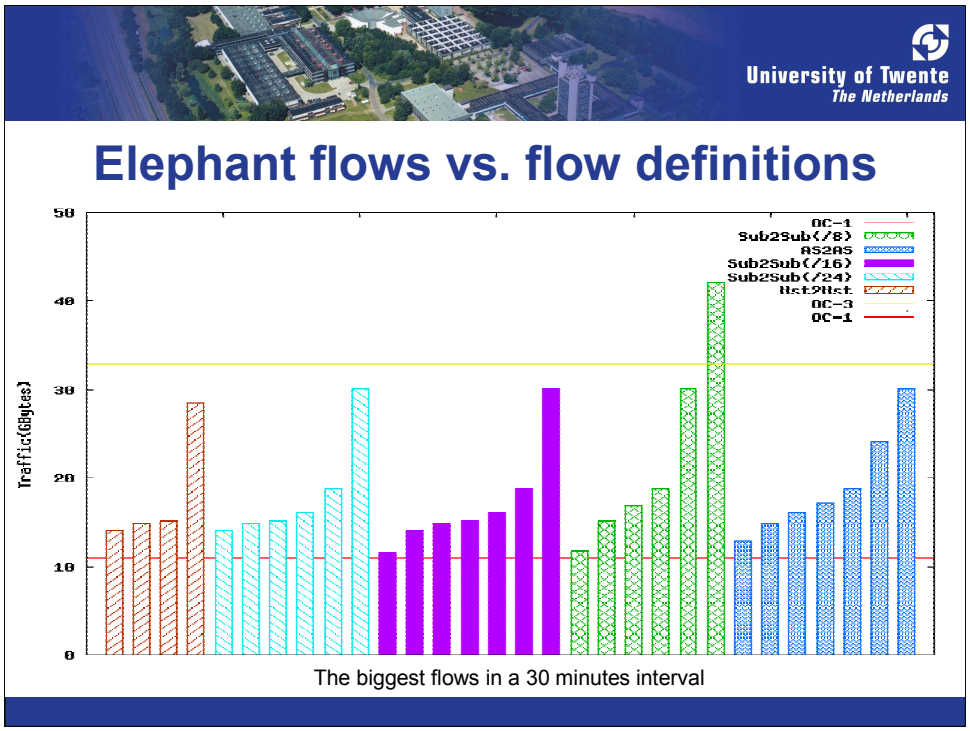
4.2 Total traffic vs. flow definitions

This graph shows the total traffic (y-axis) generated in SURFnet6 in one day (x-axis) of NetFlow collecting data. The traffic was monitored from 04:00 AM till 04:00 AM of the next day and it is divided in 30 minutes time intervals. In addition, it also shows the amount of traffic generated by the different flow definitions in each 30 minutes interval.

In addition, this graph only considers those flows that satisfied our evaluation criteria, i.e., those flows whose traffic volume is equal or greater 11 GBytes. One could say that the graph does not show any ***App2App*** flows. This is correct and the reason is that no one of analyzed ***App2App*** flows satisfied our criteria.

On the other hand, considerable amount of the total traffic can be aggregated in flows by using different flow definitions with different level of granularity. The ***Sub2Sub (/8)*** and the ***AS2AS*** are those that aggregate most of the traffic in all the time intervals. This can be explained by the fact that these flows definitions are the least restrictive when aggregating packets into flows and they also include the traffic volume of flow definitions with higher level of granularity (e.g., ***Sub2Sub (/24)*** or ***Hst2Hst***). As a result of that, flow definitions with lower level of granularity flows tend to accumulate much more bytes than the flow definitions with higher level of granularity.

In a general view, this graph proves that the less restrictive a flow definition is, the most traffic can be aggregated into flows. And as a consequence, the most elephant flows eligible for light paths may be found.

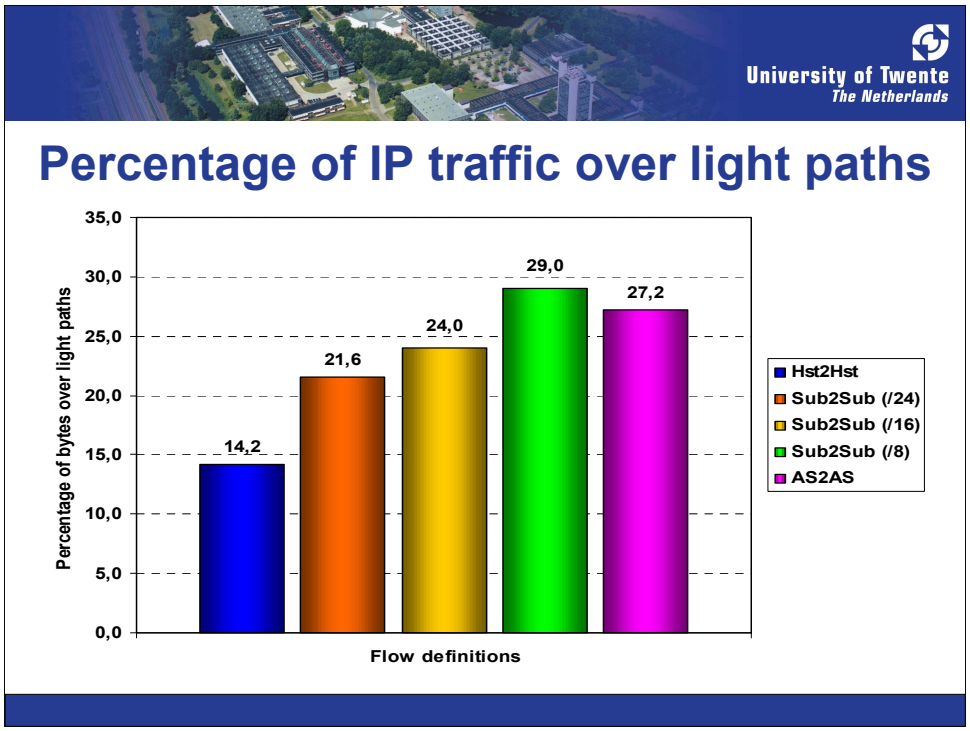


4.2 Elephant flows vs. flow definitions

The previous graph showed the amount of total traffic consumed by different flow definitions in each 30 minutes intervals. The graph above provides a "zoon in" in one of the 30 minutes time interval and it shows the elephant flows distribution within it. We have purposely omitted to show all elephant flows per 30 minutes time interval for a space constraint. However, we can assure that this elephant flow distribution has almost the same pattern for all our 30 minutes intervals.

This graph confirms what it is commonly found in the literature, small number of flows are responsible for a high traffic volume. This can also be seen by the fact that there are no many differences in the number of elephant flows per different flow definition.

As mentioned in the previous slide, there are no ***App2App*** elephant flows in this graph because no one of them reached the minimum requirement. The reason for that is the criteria used in our approach, which aims to find big and long flows. For sake of curiosity, the biggest ***App2App*** flow that we saw in our analyses was 3.63 GB, which is 33% of our minimal requirement for a light path. If different criteria would be used such as to group flows in smaller time intervals (e.g., 5 min), a great number of ***App2App*** flows might be found.



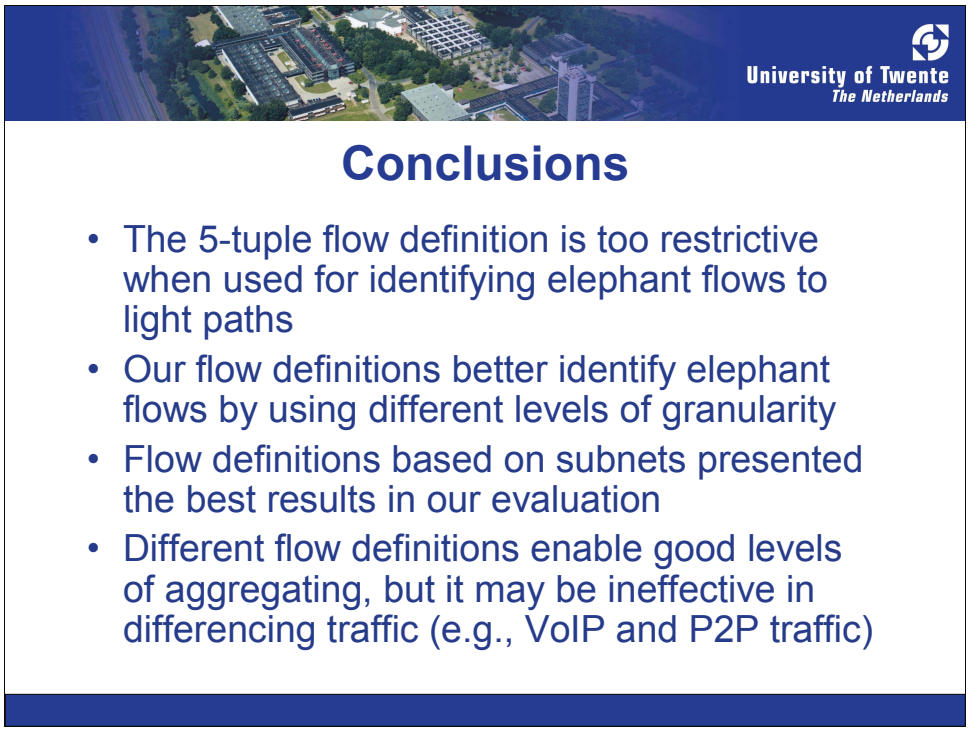
4.3 Percentage of IP traffic over light paths

This graph shows the percentage of IP traffic per different flow definition that could be transferred over light paths. As shown in the previous results, ***App2App2*** does not appear here for the same explanation presented before. As it was already expected, the volume of traffic eligible to be transferred over light paths increases when flow definitions with lower level of granularity are used.

It is also seen that subnets contribute for most of the traffic to be transferred over light paths. This is also true for the case of autonomous system, since they are an aggregation of subnets. Together, the flow definitions based on subnets (***AS2AS*** and ***Sub2Sub*** variants) allow an average of 25% of the total IP traffic to be transferred over light paths.

On the other hand, individual hosts are not responsible for much of traffic generated. This can be seen in the graph above by seeing that the ***Hst2Hst*** bar is about 1,6 times smaller than the others.

A time interval of 24 hours is used to present the percentage in the graph above. One could argue that by using a different time interval, different percentage values can be found, which is true. However, our purpose in presenting this graph is only to demonstrate the fact that by using different flow definitions, this would result in bigger flows and therefore more IP traffic could be transferred over light paths.



5. Conclusions

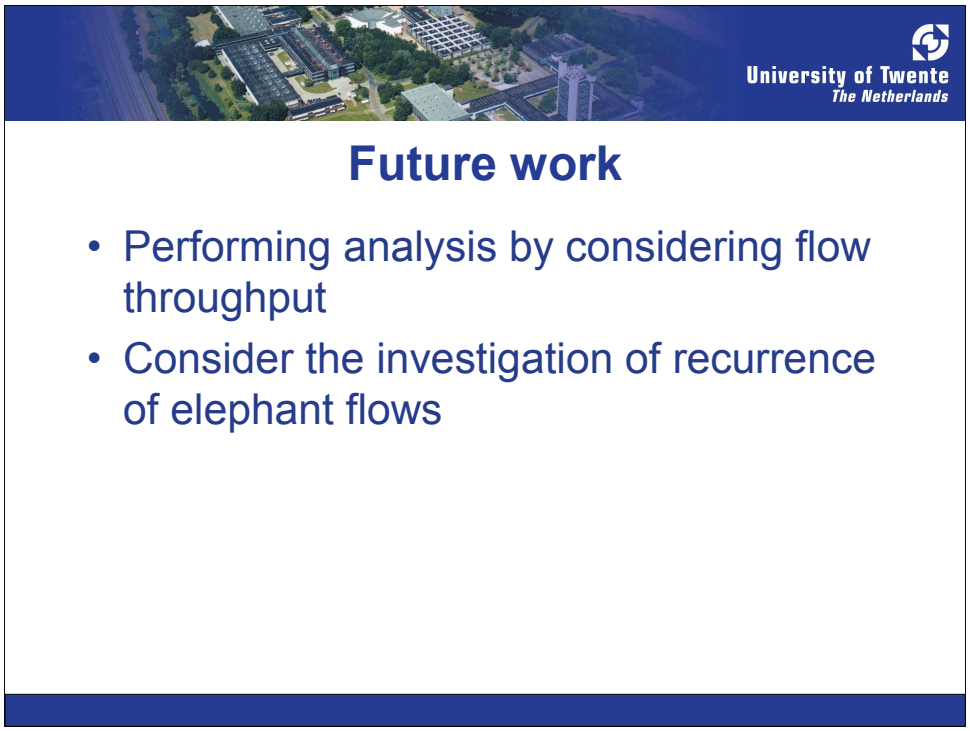
We have presented in this paper different flow definitions in order to find eligible elephant flows to be transferred over light paths. We also presented an evaluation of our proposed flow definitions with the traditional 5-tuple flow definition.

The presented results shows that the 5-tuple definition (***App2App***) is too restrictive for selecting elephant flows to light paths. This is presented in our results, which show that no ***App2App*** flow satisfied our criteria to be considered eligible for a light path. The reason for that is the number of properties that 5-tuple definition has to group packets into it, which results in a few number of packets aggregated per flow, and, as a consequence, the ***App2App*** flow definition tends to be small.

On the other hand, our proposed flow definitions showed that by using less granular flow definitions outcomes in a greater amount of packets that are grouped in much bigger flows (average of 25 GB). With such a size, these flows are easier to be identified and, as a consequence, transferred over light paths.

In addition, the flow definitions based on subnets (***SubSub (/24), SubSub (/16), SubSub (/8),*** and ***AS2AS***) were those that presented the most promising results in our evaluation. Approximately 25% of the total IP traffic analyzed in our work could be transferred over light paths by using these flow definitions.

However, when flow definitions with lower level of granularity are used, different kind of traffic (e.g., VoIP and P2P traffic) may be mixed. This can result in a inefficient method to select flows to light paths if the purpose is to select flows based on traffic differentiation. In a scenario like that the ***App2App*** flow definition is a better choice due to his higher amount of properties to characterize a flow.



5.1 Future work

As future work, we intend to consider the throughput for a more precise evaluation of flows. We used a fixed time interval (30 minutes) for the flows evaluation, which allow us to show that elephant flows could be found in such a interval. However, the aggregation made in the different flow definitions did not allow us to consider flow throughput as a flow requirement for light paths. One solution to overcome this issue would be the usage of smaller time intervals (e.g., 5 minutes time interval).

In addition, it would be interesting to know how often these elephants flows appear in large–scale networks, i.e., to analyze the recurrence of these flows. A better traffic engineering can be achieved by knowing in advance when a specific elephant flows is about to start sending a lot of data.

The last, but not the least, we believe that the main contribution of this work is to show that the probability of finding elephant flows eligible to light paths increases by using different flow definitions. These different flow definitions allow therefore that a greater number of bytes can be transferred over light paths, which could off-loaded congested IP networks.

Thanks for your attention!

- Contact:
  - **Tiago Fioreze**
    (t.fioreze@utwente.nl)
  - **Mattijs Oude Wolbers**
    (m.oudewolbers@utwente.nl)
  - **Remco van de Meent**
    (r.vandemeent@utwente.nl)
  - **Aiko Pras**
    (a.pras@utwente.nl)

Acknowledgements:

We would like to thank SURFnet for allowing us to perform measurements on their network. This paper was supported in part by the EC IST-EMANICS Network of Excellence (#26854).

References

1. N. Larrieu, P. Owezarski, "Measurement based networking approach applied to congestion control in the multi-domain Internet", *9th IFIP/IEEE International Symposium on Integrated Network Management* (IM'2005), Nice, France, pp.485-498, 2005.
2. T. Mori, R. Kawahara, S. Naito, S. Goto, "On the characteristics of Internet traffic variability: spikes and elephants," *Applications and the Internet, 2004. Proceedings. 2004 International Symposium on* , pp. 99-106, 2004.
3. D. Papagiannaki, N. Taft, S. Bhattacharyya, P. Thiran, K. Salamatian, C. Diot, "A pragmatic definition of elephants in internet backbone traffic," *2nd ACM SIGCOMM workshop on Internet measurement* (IMC '02), Marseille, France, pp. 175-176, 2002.
4. T. Mori, M.Uchida, R. Kawahara, J. Pan, S. Goto, "Identifying elephant flows through periodically sampled packets", *4th ACM SIGCOMM conference on Internet measurement* (IMC '04), Sicily, Italy, pp. 115-120, 2004.
5. J. Wallerich, H. Dreger, A. Feldmann, B. Krishnamurthy, W. Willinger, "A methodology for studying persistency aspects of internet flows", *ACM SIGCOMM Computer Communication Review,* 35(2): pp: 23-36, 2005.
6. C. Estan, G. Varghese, "New directions in traffic measurement and accounting: Focusing on the elephants, ignoring the mice", *ACM Transactions on Computer Systems* (TOCS) , 21(3): pp. 270-313, 2003.
7. Gigaport, "GigaPort homepage", 2006, Available in: <http://www.surfnet.nl/info/innovatie/gigaport/>
8. Zebedee, "Zebedee: Secure IP tunnel", 2006, Available in: <http://www.winton.org.uk/zebedee/>.
9. Tcpdump, "TCPDUMP public repository", 2006, Availabe in: <http://www.tcpdump.org/>.
10. B. Claise. "Cisco Systems NetFlow Services Export Version 9 ", Request for Comments 3954, IETF, October 2004.