Re-sequencing Buffer Occupancy of a Concurrent Multipath Transmission Mechanism for Transport System Virtualization

K. Tutschku[1], T. Zinner[2], A. Nakao[3,4] und P. Tran-Gia[2]

[1]Universität Wien, Professur "Future Communication",
Universitätsstrasse 10, 1090 Wien, Austria.

[2]Universität Würzburg, Lehrstuhl für Informatik III,
Am Hubland, 97074 Würzburg, Germany.

[3]University of Tokyo, Graduate School of Interdisciplinary Information Studies,
7-3-1, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.

[4]also with the NICT (National Institute for Communication and Information Technology), 1-33-16, Hakusan, Bunkyo-ku, Tokyo, 113-0001, Japan.

Abstract From the viewpoint of a methodology, the concept of *Network Virtualization (NV)* extends beyond pure operational issues. In this paper, we will first transfer the concept of location independence of resources, as known in operating systems, to the area of data transport in communication networks. This idea is denoted as *transport system virtualization (TSV)*. We will outline an example for TSV which uses *concurrent multipath (CMP)* transmission and discuss an important performance issue of CMP transmission, the *re-sequencing buffer occupancy probability distribution*. The investigation of this probability distribution gives insights in how to select the set of paths when using CMP transmission for transport system virtualization.

1 Introduction

For practitioners, the concept of *Network Virtualization (NV)* enables the consolidation of multiple networks or overlays into a single physical system. Thus, NV may lower the required amount of hardware (capital expenditures, CAPEX) and of operational expenditures (OPEX) of multiple network structures since less systems have to be configured. As a result, NV is considered mainly as an operational technique.

From the viewpoint of a methodology, the concept of NV extends beyond the above outlined asset and operational issues. One of the main benefit of virtualization techniques is the *abstraction of computer resources*. An operating system, for example, may provide virtual memory to an application program. This memory gives the program the feeling that it can use a contiguous working memory, while in fact it may be physically fragmented and may even overflow on to disk storage. The actual location of the data in the physically memory doesn't matter and is hidden. Thus, this virtualization technique makes the resource "memory" independent from its physical "location". In this paper, we transfer the concept of location independence of resources to the area of data transport in communication networks. This idea will be denoted here as *transport system virtualization (TSV)*.

The short paper is structured as follows. First, we will outline briefly the idea of TSV. Then, we detail an example for TSV using a *concurrent multipath (CMP)* transmission mechanism. Finally, we discuss an important performance issue in CMP transmission, which is the *re-sequencing buffer occupancy probability distribution*. The investigation of this probability distribution gives insights in how to select the set of paths used in a TSV mechanism using CMP transfer.

2 Transport System Virtualization

The idea of *tranport system virtualization* is motivated partly by the abstraction introduced in P2P content distribution networks (CDNs). Advanced P2P CDNs, such as eDeonkey or BitTorrent, apply the concept of *multi-source download*. Here, a peer downloads multiple parts of a file in parallel from different peers. As a result, the downloading peer doesn't rely any more on a single peer which provides the data, and the reliability is increased. The providing peers are typically selected such that the throughput is optimized. The actual physical location of the file doesn't matter. Thus, these P2P CDNs can be viewed as an abstract and almost infinite storage for the data files. Thus, an abstraction of a storage resource, similar to the example of virtual memory in the introduction, is achieved.

The above outlined abstraction of a storage resource is transferred to the area of data transport. *Transport System Virtualization (TSV)* can be viewed as an abstraction concept for data transport resources. Hereby, the physical location of the transport resource doesn't matter as long as this resource is accessible. In TSV an abstract data transport resource can be combined from one or more physical or overlay data transport resources. Such a resource can be, e.g., a leased line, a wave length path, an overlay link, or an IP forwarding capability to a certain destination. These resources can be used preclusive or concurrently and can be located in even different physical networks or administrative domains. Thus, an abstract transport resource exhibits again the feature of location independence.

3 Implementing TSV using Concurrent Multipath Transfer in Advanced Routing Overlays

A scalable approach for routing overlays is the concept of *one-hop source routing*, [1,2]. Hereby, the user data is forwarded to a specific intermediate node, denoted as One-hop Source Router (OSR), which then relays the traffic to its destination using ordinary IP routing. The details of this architecture can be found in [2] and in accompanying paper to this short paper [3].

The TSV in the considered OSR architecture is achieved by a *Concurrent Multipath (CMP)* transfer mechanism. The mechanism combines multiple overlay paths (even from different overlays) into a single virtual high-capacity pipe. The combined paths are used in parallel by sending data packets concurrently on different overlay paths. This principle is also know as *stripping*. The paths

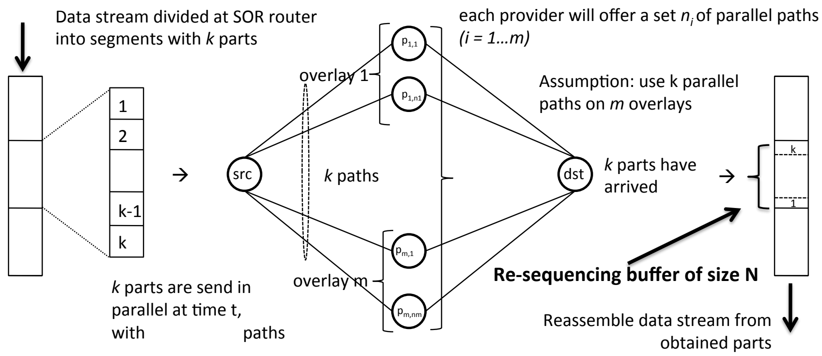

Figure 1: Transmission Mechanism

which form the virtual pipe are chosen by the ingress router out of large number of potential paths, cf. [2] and [3]. The *path oracle* discovers the available paths, i.e. the components of the abstract data transport resource. Current discussions suggest that a path oracle can be provided by the network operator or by other institutions [4]. Altogether, the CMP mechanism combined with the path oracle facilitates an abstraction of a data transport resource. Instead using a single fixed data transport resource, the system relies now on location independent, multiple and varying resources.

Figure 1 shows a detailed model of the stripping mechanism. The data stream is divided into segments which are split into k smaller parts. The k parts are transmitted by the set of paths, i.e. in parallel on k different overlay links. The receiving router reassembles these parts. Unfortunately, the parts can arrive at the receiving router after different time intervals since they experience stochastically varying delay. Therefore, it is possible that parts arrive "out of order".[1]

Part or packet re-ordering may have a severe impact on the application performance. In order to level this behavior, the receiving router maintains a finite re-sequencing buffer. However, when the re-sequencing buffer is filled and the receiving router is still waiting for parts, part loss can still occur. This loss of parts is harmful for the application and should be minimized. Therefore, an important objective in operation of the system is to minimize the re-sequencing buffer occupancy. This can be achieved by a selection of paths with appropriate delay characteristics. The influence of path delay and its distribution on the buffer size is discussed next.

4 Re-sequencing Buffer Occupancy in Concurrent Multipath Transport

The performance of the CMP mechanisms is investigated by time discrete, event-based simulations. The simulation model assumes a continuous data stream. The stream is divided into parts which are send in parallel on either two or three

[1] Please note that part re-ordering can only happen between different paths. The order of parts on a path is maintained since packets typically can not overtake each other.

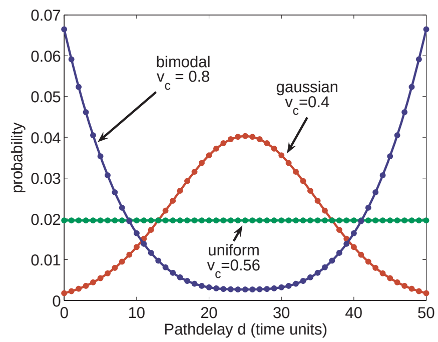

Figure 2: Used distributions

paths. The delay on the paths is modeled by discrete delay distributions with a resolution of one time unit. A packet is transmitted every time unit on a path.

4.1 Impact of Type of Delay Distribution

In a first simulation study, the impact of the type of delay distribution on the buffer occupancy is investigated. Therefore, three different path delay distributions are considered for the paths: a truncated gaussian (label *gaus*), a uniform (*uni*) and a bimodal distribution (*bi*). The probability density functions (PDFs) of the distributions are depicted in Figure 2. The mean delay value for each the distributions is $E[d] = 25$, the minimum delay is $d_{min} = 0$ and the maximum delay $d_{max} = 50$. The coefficient of variation c_v varies between $v_c = 0.4$ for the gaussian distribution to $v_c = 0.8$ for the bimodal distribution. We decided to investigate these distributions in order to evaluate the system behavior under highly different condition, e.g. gaussian vs. bimodal delay.

We start with the investigation of two concurrent paths. The buffer occupancies for different delay combinations are depicted in Figure 3. The y-axis denotes the probability of the packets stored in the re-sequencing buffer, assigned on the x-axis. For the sake of clarity we plotted only the *bi,bi* buffer occupancy distribution with confidence intervals for a confidence level of 99%.[2]

For the case of two gaussian delay distributions, the buffer occupancy is left leaning and higher buffer occupancies are not very likely. However, for two bimodal delay distributions a large fraction of the probability mass covers a buffer occupancy bigger than 30 packets. It should be noted that the maximum buffer occupancy in the investigated scenario $o_{max} = 50$. As we see, the type and the variation of the path delay have a major influence on the buffer occupancy of the receiver. We can conclude that the buffer occupancy is not invariant to the delay distribution of the used paths.

In Figure 4 the buffer occupancies for three concurrent paths and different delay distributions are shown. It should be noticed, that the maximum buffer occupancy in this scenario $o_{max} = 100$. That denotes the worst case which can

[2] The size of the confidence intervals for the other curves is similar.

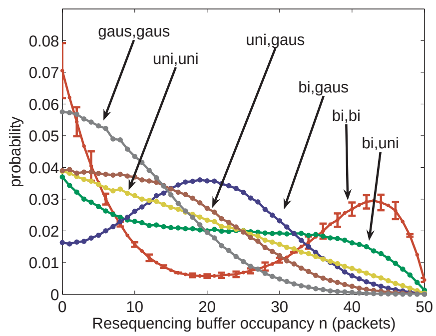

Figure 3: Buffer occupancy for two paths

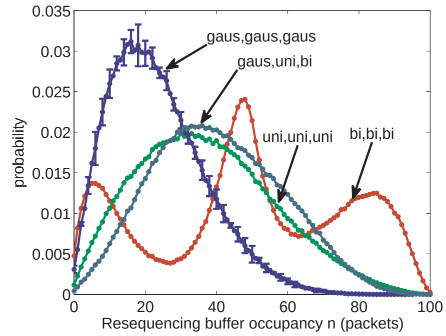

Figure 4: Buffer occupancy for three paths

occur if a packet over one path has the maximum delay. In this case up to 100 packets might be transmitted over the other paths and have to be stored in the buffer until the next packet in sequence arrives.

Furthermore it can also be seen, that the buffer occupancy for three gaussian distributions is the smallest. For three times uniform and for one delay distribution of each type the occupancy is rather the same and higher than in the only-gaussian case. For three times bimodal the buffer occupation is the highest of the investigated scenarios, and a noticeable part of the of the probability mass covers an occupancy bigger than 60 packets.

4.2 Path Selection Criteria

In a second simulation study we investigates the criteria for path selection. Intuitively, a path selection algorithm should select those paths, which provide the shortest delay. This is of typical interest when the traffic comes from an interactive application which requires realtime responsiveness, but is leveled when non-interactive traffic is considered. For the investigation, we considered two truncated Poisson-like delay distributions with mean packet delay of $E[d] = 17$ and $E[d] = 34$ and with the same standard deviation of 4.1, cf. Fig.5.

Figure 6 shows actually four buffer occupancy distributions: a) all paths have the same low mean delay, b) all paths have the same high mean delay, c) two paths have low and one path a high mean delay, and d) two paths have high and one path a low mean delay. A close look at Figure 6 shows that the case a) and b) have the same buffer occupancy distribution. This indicates that the pure delay has no impact on the buffer occupancy. For the cases c) and d) it is outstanding that case d) with two high delay paths has a better performance in terms of occupancy than case c). This is very remarkable since intuitively more paths with lower delay should yield better performance. However, the coefficient of variation is lower for the high delay path and therefore the two high delay paths are more immune against delay variation. In turn, this yields a low buffer

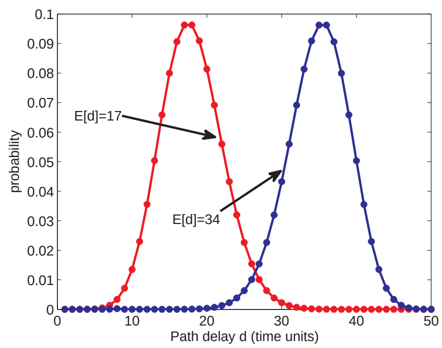

Figure 5: Poisson-like Delay Distributions

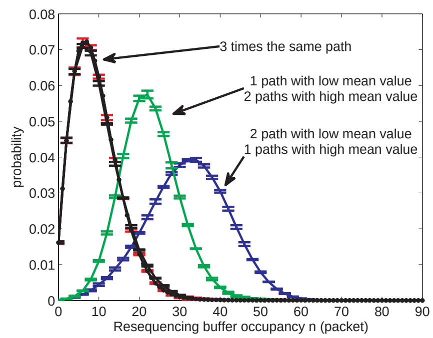

Figure 6: Buffer occupancy considering Poisson-like Delay Distributions

occupany. These simple investigation shows that also the second moment of the path delay has to be considered in a real world path selection algorithms.

5 Conclusion

In this paper we have introduced the idea of *transport system virtualization (TSV)*, which provides a location independent abstraction for data transport resources. We have outlined the TSV concept by the example of *concurrent multipath (CMP)* transmission in one-hop source routing overlays and discussed an important performance issue of CMP transmission, the *re-sequencing buffer occupancy probability distribution* under the influence of the delay distribution on used the paths in the CMP mechanism. The performance evaluation shows that the delay and its distribution type has significant influence on the buffer occupancy. However, the intuitive approach of considering only mean delay is not sufficient. The second moment of the delay, which determines the coefficient of variation of the delay, influences strongly the buffer occupancy. More path selection criteria will be investigated in the future.

6 Literature

References

1. K. Gummadi, H. Madhyastha, S. Gribble, H. Levy, and D. Wetherall, "Improving the reliability of internet paths with one-hop source routing," in *Proc. OSDI'04*, San Francisco, Ca., USA, Dec. 2004.
2. S. Khor and A. Nakao, "AI-RON-E: Prophecy of One-hop Source Routers," in *Proc. of Globecom08*, New Orleans, LA., Nov./Dec. 2008.
3. K. Tutschku, T. Zinner, A. Nakao, and P. Tran-Gia, "Network Virtualization: Implementation Steps Towards the Future Internet," in *Proc. of the Workshop on Overlay and Network Virtualization at KiVS 2009*, Kasel, Germany, Mar. 2009.
4. V. Aggarwal, A. Feldmann, and C. Scheideler, "Can isps and p2p systems co-operate for improved performance?.," *ACM SIGCOMM CCR*, vol. 37, no. 3, Jul. 2007.