



Technische Universität München

Chair of Media Technology

Prof. Dr.-Ing. Eckehard Steinbach

Engineer Practice

Priority Based Buffer Flushing With the Click Modular
Software Router

Author:	Arturo Buitrago Méndez
Matriculation Number:	03646899
Address:	Türkenstr. 58a 80799 München
Advisor:	Christoph Bachhuber
Begin:	Datum des Arbeitsbeginns
End:	Datum des Vortrags

With my signature below, I assert that the work in this thesis has been composed by myself independently and no source materials or aids other than those mentioned in the thesis have been used.

München, August 7, 2016

Place, Date

Signature

This work is licensed under the Creative Commons Attribution 3.0 Germany License. To view a copy of the license, visit <http://creativecommons.org/licenses/by/3.0/de>

Or

Send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA.

München, August 7, 2016

Place, Date

Signature

Contents

Contents	i
1 Introduction	1
2 Click Modular Router	2
2.1 Why Click?	2
3 Buffer Flusher	4
3.1 Click Element	4
3.1.1 Example	5
3.1.2 Kernel integration	6
3.2 Packet creation	6
4 Testing	7
4.1 Test Results	9
4.1.1 No cross-traffic	9
4.1.2 Random cross-traffic	9
4.1.3 Sample pattern of cross-traffic	9
5 Zusammenfassung	11
A Packet structure	12
List of Figures	13
List of Tables	14
Bibliography	15

Chapter 1

Introduction

The aim of this Engineer Practice is the implementation of a buffer flusher mechanism that works as described in the paper "Low Delay Video Transmission Strategies" [1]. Its main goal is to give certain video frames, called "*keyframes*", a higher transmission priority, by "*flushing*" the buffer containing the outgoing packets of any belonging to non-keyframes, thus allowing the more important packets to be transmitted immediately. This would ideally result in a lower end-to-end transmission time in real systems. The router devised worked as intended at the user level but kernel integration proved to be too laborious for the scope of this work.

The buffer flusher itself was implemented through the Click Modular Router, a virtual router creation software with Linux kernel integration. The corresponding Click Element files are included in the repository of the project, as well as the scripts used to create and send the packages. Tests were then run in a variety of different transmission conditions in order to gauge the effectiveness of the application.

The rest of this report includes a comment on why Click was chosen and a more in-depth explanation of the software and the buffer flusher specifically. It is then followed by a discussion of the prospect of kernel integration. The testing environment is introduced and the test results are presented and discussed.

Chapter 2

Click Modular Router

The **Click Modular Router** is a project that was born out of a paper by Eddie Kohler and other collaborators in 2000 [2]. It is still being given very sporadic, but important, updates by Kohler himself in his personal github repository. It is designed to be a software architecture for building configurable routers. These routers are composed of simple packet routing modules, called *elements*, that perform relatively straightforward functions. Click works by concatenating these elements (with possibly multiple inputs and outputs), in a meaningful way. This property makes it particularly easy to expand the software with user-defined elements, which was the case in this project.

Click configuration files, which specify the flow of packets by listing the elements and their respective connections, are written in their own, simple syntax. Element files in themselves use C++, with its own implementation of the C++ Standard Library, which contains equivalents to most of the STL data structures.

The software is capable of running the router either at the user level or as a Linux kernel thread. Be advised that not all elements included in the software distribution can be executed at a kernel level. One of the original goals of this project was kernel integration of the resulting application, but unforeseen errors lead to debugging at a kernel level being out of the time scope devised for it. For further details, please read the appropriate chapter

Click has a significant learning curve, especially due to lack of documentation and online resources, so perseverance is advised.

2.1 Why Click?

The Click Modular Router software had already been identified as the option with most potential from before the start of the practice in itself by my supervisor, Mr. Bachhuber.

However, part of the literature research needed included evaluating other options for the implementation of the application.

Alternatives seemed to be relatively scarce and deplorably out of date. The main ones identified were:

- X-kernel
- Scout
- OpenVZ

X-kernel is a project that was mostly maintained by researchers from the University of Arizona. Its aim is to be an Object-based framework for implementing network protocols. Most of its functionality was aimed at the TCP/IP protocol, and was deemed to lack the flexibility needed in order to process packets that did not expressly follow the protocol. I was also unable to find a reference or paper from this millennium, which further made it seem nonviable for the realization of the flusher.

Scout was devised as a spin-off of X-kernel, also with support from University of Arizona personnel. It was described as being a communication-oriented operating system targeted at network appliances. Despite being operating system, which would probably grant some greater measure of speed or expediency, it was also seen as not fitting the specifications. This was in no small part due to the woeful lack of updates and the sorry state of its host website and documentation.

OpenVZ, as discussed more thoroughly in a research paper by Doriguzzi et al [3], is a modified Linux kernel tree supporting virtualization, isolation and resource management; and it is mostly used for virtualization of Linux containers. It was found by the authors to be too limiting compared to Click in different aspects for their simulation work on multi-hop wireless networks.

Adding to these considerations, the paper published by Bianco et al [4], which aims to prove the feasibility of a high performance IP router, which was more in line with our expectations of the software. It compared Click favorably in terms of reception and transmission (bit and packet) rate, among other measures, in a series of evaluations against the standard Linux network stack implementation. It especially attributes the differences to elements such as better buffer management in Click.

Regarding buffer flushing or other forms of frame dropping in general, despite being a relatively common occurrence in video delivery for bandwidth-constrained applications, as in [5] and [6], the actual action is glossed over quickly and given little attention in the available literature.

Chapter 3

Buffer Flusher

3.1 Click Element

The buffer flusher was ultimately implemented as a single element in the Click software. Like all Click elements, it consists of a header file, in which the overall specifications of the element are given, and an element file, which contains the actual code. The header file contains the information that Click needs in order for it to know in what contexts the element makes sense, how many ports it has, how it reacts to packets being pulled or pushed, etc. In this file, the flusher has a very similar profile to that of a queue, with one input and one output port, that can handle any number of pushed packets and returns the appropriate packet when pulled, if there are any present.

The element file contains the actual implementation of the buffer flusher. It was modeled from the file that belonged to the default QuickQueue implementation. The definition for the pull function is left unchanged, with the element returning the first packet in the queue when any are present and returning a Null pointer when empty. Buffer flushes can only occur when a new packet is pushed into the queue.

The push function starts by setting a pointer to the byte in the packet header which indicates whether the packet in question belongs to a *keyframe*, a video frame which was deemed to be of higher relevance and thus marked for faster forwarding. Note that the buffer flusher has no functionality to actually tell whether a frame is a *keyframe* or not, that distinction should be made by a different algorithm in another part of the process altogether, before splitting the frame itself into packets. The position of the *keyframe* identifier was set beforehand (see the section dealing with packet creation), but the position of the pointer can be easily changed in the element file code.

If the frame is not a *keyframe*, the frame gets pushed into the awaiting queue and the number of *non-keyframes* since a flush is incremented by one. If, on the contrary, the packet does identify itself as having high transmit priority, another pointer into the packet

checks whether the packet is the start of a frame or not. This makes the flushing function happen only for the start of frames, assuming of course that the whole frame has the same priority, saving some CPU power. Should that not be an issue, then the second pointer can be altogether ignored.

At this point, a flush of the buffer is called. This leads to two possible cases for flushing:

1. Queue is full of *non-keyframes*
2. Queue has a mixture of *keyframes* and *non-keyframes*

In the first instance, the whole queue is simply purged. All incoming *keyframes* are pushed to the start of the queue and can be pulled immediately. In the second case, the algorithm drops a number of packets at the back of the queue equal to the number of *non-keyframe* packets since the last flush. Since all *keyframes* cause a buffer flush unless the queue is empty, this should lead the prioritized packets to be pushed to the first slot not occupied by the same kind of parcels. Thus we have an element that implements the buffer flushing according to specifications.

3.1.1 Example

Given the following packet flow, with 0 denoting a *non-keyframe* and 1 a *keyframe*, with the first packet pushed beginning on the left:

$$0 \leftarrow 0 \leftarrow 1 \leftarrow 0 \leftarrow 1 \leftarrow 0 \leftarrow 0 \leftarrow 0$$

The buffer flusher would execute the following with each packet, assuming no pull requests reach the buffer flusher:

1. Push *non-keyframe* packet, 1 packet since last flush.
2. Push *non-keyframe* packet, 2 packets since last flush.
3. Buffer flush (Case 1, whole queue discarded), push *keyframe* packet. No packets since last flush.
4. Push *non-keyframe* packet, 1 packet since last flush.
5. Buffer flush (Case 2, 1 packet flushed), push *keyframe* packet. No packets since last flush.
6. Push *non-keyframe* packet, 1 packet since last flush.
7. Push *non-keyframe* packet, 2 packets since last flush.
8. Push *non-keyframe* packet, 3 packets since last flush.

This leaves the queue in the following order, with queue front at the right:

$$0 \rightarrow 0 \rightarrow 0 \rightarrow 1 \rightarrow 1$$

3.1.2 Kernel integration

As previously mentioned, one of the goals of the project was the integration of the application into the Linux kernel. The relative ease with which Click allows this was also one of the reasons why it was chosen as an appropriate software. The development and early testing of the buffer flusher occurred at the user level. After the correct functionality of the application was tested at this level, I tried to run it as a kernel thread, but ran into all sorts of trouble.

Running the router configuration in the kernel level causes Linux to intercept the thread and lock the CPU in which it is running for a certain amount of time. This loops after the timeout expires. I managed to isolate the problem to the buffer flusher element in itself, but could get no further without a deep understanding of the kernel and how it works with Click.

The painstaking process of kernel thread debugging was deemed too laborious for the scope of this engineer's practice, as well as too complex for my current skill set. My supervisor and I decided jointly to run the tests at the user level and leave any kernel integration for possible future projects.

3.2 Packet creation

Part of the scope of the project included the creation of the packets from the original trace files provided by my supervisor. To preserve the integrity of test results and the functionality of the application, I sought to build the packages in a way that was unobtrusive and maintained both the size specified in the trace and all the information indicated in the files.

I decided on building the packets with a simple script with the help of "*Scapy*", a Python-based packet manipulation program and corresponding module with especially accessible forging of parcels. The information was codified in simple fields of four Bytes and then filled with dummy data to achieve the correct size. These would then be sent from one machine to another through a UDP socket in order to crudely simulate the traffic coming in from the camera. This also significantly facilitates the packet capture by Click itself. The machine receiving the packages would then run the flushable buffer with the incoming information. For more details on packet structure, please see Appendix A.

Despite the fact that UDP does not represent a realistic implementation of the video file transfer, seeing as the files being worked on are filled with dummy data, the relative convenience overruled any other minor concerns.

Chapter 4

Testing

The testing was centered around evaluating the way the buffer flusher deals with different volumes and patterns of incoming packet traffic. This meant simulating different situations of channel saturation for the path between the creation of the packages and the buffer flusher. Evaluation of the forwarding of the packages at the output of the buffer flusher was deemed unimportant and as such was not as thoroughly investigated. The test environment was as follows:

Packet Creator → Buffer Flusher → Packet Capture

Path A is a channel emulated with the *Netropy Network Emulator* in the institute Robotics lab, under a series of bandwidths and loads. **Path B** is a virtual channel emulated with a certain bandwidth that is included in the router configuration file for Click. It routes the output packages of the buffer flusher into a packet capture software (e.g. *Wireshark*) and stores them as a *.pcap* dump file.

The following bandwidths (among others) were investigated in certain relevant combinations for both paths:

- 1 Mbps
- 4 Mbps
- 16 Mbps

For readability and legibility, only certain combinations are presented and analyzed in this report. The rest of the data is contained in the *test_results* repository.

These bandwidth combinations were chosen for illustrative purposes and do not necessarily represent maximum or minimum values for correct function of the buffer flusher.

- **Path A** : 1 Mbps, **Path B** : 4 Mbps

This configuration is meant to test the case in which the outgoing packet rate to the flusher exceeds the income rate, which should not be the norm for most. There should be little need for flushing due to very short queues in the flusher itself.

- Path A : 16 Mbps, Path B : 1 Mbps

This configuration tests a high incoming packet rate with a low outgoing rate, a more realistic example. This environment should trigger the flushing mechanism at fairly regular intervals and evidence a much higher percentage of *keyframe* packages after the flusher.

- Path A : 16 Mbps, Path B : 16 Mbps

Lastly, both income and outcome rate are roughly equal. A similar result to the first configuration should be observed in terms of package count, but with much higher expediency.

Additionally, the following channel loads or cross-traffic patterns were investigated.

- No channel load
- Random cross-traffic
 - 25% channel load
 - 50% channel load
 - 75% channel load
- Sample pattern of cross-traffic

The random cross-traffic patterns produce Poisson-distributed bursts of packets of 1500 Bytes of length, around the size of the average packet length.

The sample pattern of cross-traffic was a

Two video traces were used and had the following properties:

Table 4.1: Video traces used for testing

Video traces			
Channel type	Total packages	<i>Keyframe</i> packages	<i>Keyframe</i> percentage
Slow Channel	1465	687	46.9%
Medium Channel	15677	632	4.0%

4.1 Test Results

4.1.1 No cross-traffic

The tests run with no cross-traffic serve primarily to set a benchmark to compare other testing conditions with, as well as corroborate the function of the application. Note that the chosen bandwidths are meant to have an illustrative purpose for the

Table 4.2: No cross-traffic

BW A (Mbps)	BW B (Mbps)	Total pack- ages	Keyframe pack- ages	%	Non- Keyframe Packages	Time elapsed (sec)
Slow Channel						
1	4	1465	687	46.9%	778	17.35
16	1	697	687	98.6%	10	1.08
16	16	1465	687	46.9%	778	1.08
Medium Channel						
1	4	15677	632	4.0%	15045	186.26
16	1	1163	632	54.3%	531	11.57
16	16	15677	633	4.0%	15046	11.64

4.1.2 Random cross-traffic

4.1.3 Sample pattern of cross-traffic

Lorem Ipsum

Table 4.3: Random cross-traffic, 50% channel occupation

BW A (Mbps)	BW B (Mbps)	Total pack- ages	Keyframe pack- ages	%	Non- Keyframe Packages	Time elapsed (sec)
Slow Channel						
1	4	1465	687	46.9%	778	35.6
16	1	710	687	96.8%	23	2.3
16	16	1465	687	46.9%	778	2.2
Medium Channel						
1	4	15677	632	4.0%	15045	371.8
16	1	2195	632	28.8%	1563	23.8
16	16	15677	633	4.0%	15046	23.7

Chapter 5

Zusammenfassung

Am Schluß werden noch einmal alle wesentlichen Ergebnisse zusammengefaßt. Hier können auch gemachte Erfahrungen beschrieben werden. Am Ende der Zusammenfassung kann auch ein Ausblick folgen, der die zukünftige Entwicklung der behandelten Thematik aus der Sicht des Autors darstellt.

Appendix A

Packet structure

Table A.1: Breakdown of packet structure by byte

Packet Structure		
Content	Bytes	Description
Ethernet Header	0 - 13	Header due to packaging.
IPv4 Header	14 - 33	Header due to packaging.
UDP Header	34 - 41	Header due to packaging.
Frame Identifier	42 - 45	Binary variable coded as integer, either <i>keyframe</i> or not.
Frame Number	46 - 49	Used to monitor that order of packages is not scrambled. Coded as integer.
Segment Number	50 - 53	Can be kept track of to detect any drops, coded as an integer.
Packet Size	54 - 57	For most packets it equals 1538 bites. Coded as long.
Event Occurred	58 - 61	Uninteresting for our purposes. Coded as float.
Sampling time	62 - 65	Same as above. Coded as a float.
Dummy Data	66 onward	Necessary filler to achieve specified length. Filled with zeroes.

List of Figures

List of Tables

4.1	Video traces used for testing	8
4.2	No cross-traffic	9
4.3	Random cross-traffic, 50% channel occupation	10
A.1	Breakdown of packet structure by byte	12

Bibliography

- [1] Christoph Bachhuber, Eckehard Steinbach, Technische Universit, and Martin Reisslein. Low Delay Video Transmission Strategies Chair for Media Technology.
- [2] Eddie Kohler, Robert Morris, Benjie Chen, John Jannotti, and M. Frans Kaashoek. The click modular router. *ACM Transactions on Computer Systems*, 18(212):263–297, 2000.
- [3] Roberto Doriguzzi Corin, Roberto Riggio, Daniele Miorandi, and Elio Salvadori. Airo-LAB: Leveraging on virtualization to introduce controlled experimentation in operational multi-hop wireless networks. *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, 46:331–344, 2011.
- [4] A. Bianco, R. Birke, D. Bolognesi, J.M. Finochietto, G. Galante, M. Mellia, M.L.N.P.P. Prashant, and F. Neri. Click vs. Linux: two efficient open-source IP network stacks for software routers. *HPSR. 2005 Workshop on High Performance Switching and Routing, 2005.*, pages 18–23, 2005.
- [5] Tiecheng Liu and Chekuri Choudary. Real-time content analysis and adaptive transmission of lecture videos for mobile applications. *ACM Multimedia 2004 - proceedings of the 12th ACM International Conference on Multimedia*, pages 400–403, 2004.
- [6] Zhi-Li Zhang, Srihari Nelakuditi, Rahul Aggarwal, and Rose P Tsang. Efficient Selective Frame Discard Algorithms For Stored Video Delivery Across Resource Constrained Networks. *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, pages 472–479, 1999.