ICT-216372

Trilogy

Trilogy: Re-Architecting the Internet.
An Hourglass Control Architecture for the Internet, Supporting Extremes of Commercial, Social and Technical Control

Large Scale Integrating Project

FP7 ICT Objective 1.1 – The Network of the Future

D16 Public record of 'industry workshop' held sometime during Year 3

Due date of deliverable: Dec 2010

Actual submission date: Dec 2010

Start date of project: 1 January 2008 Duration: 39 months

Lead contractor for this deliverable: DTAG

Version 1

Confidentiality status: "Public"

Abstract

This document contains a report on the workshop organized by the Trilogy project during the third year. In particular, it contains the details of the Implementers workshop on Multipath TCP, which was held in conjunction with the Maastricht IETF.

Target audience

General public.

Disclaimer

Impressum

Full project title: Trilogy: Re-Architecting the Internet. An Hourglass control Architecture for the Internet, Supporting Extremes of Commercial, Social and Technical Control

Title of the workpackage: <WP-5>

Document title: D16 Public record of ‘industry workshop’ held sometime during Year 3

Editor: Philip Eardley, BT

Workpackage Leader: Andrea Soppera, BT

Project Co-ordinator: Andrea Soppera, BT

Technical Manager: Philip Eardley, BT

This project is co-funded by the European Union through the ICT programme under FP7.

Copyright notice

Executive Summary

This document describes Trilogy’s industrial workshop, which was organised as an “Implementers Workshop” for Multipath TCP. Interactive discussions covered use cases for MPTCP and barriers to deployment. It was held in Maastricht on Saturday 24th July, which was co-located with, and immediately before, the IETF. The workshop was a good success with very interesting discussions amongst approximately 25 participants.

List of Authors

Philip Eardley – BT

Alan Ford - RMR

Table of contents

1 Introduction

The Trilogy project promised to realise an industrial workshop during the third year of the project. We didn't want the workshop to be a simple dissemination exercise of talks, lunch, more talks, go home. We wanted something that would help Trilogy ideas actually happen, ie increases the chances that Trilogy stuff moves from research to real deployment - a stepping stone on the path. After careful consideration we decided it would be more effective to focus the workshop on one topic. Multipath TCP was selected, since it is one of the main topics of Trilogy and probably the one nearest to realisation. The workshop was a success and the details are included below.

2 Venue and Organisation

We decided to organize the workshop co-located with the IETF, because several of the key people we hoped would participate would be going to the IETF anyway. Also, the July 2010 IETF was in Maastricht and thus more convenient for Europeans than a normal IETF. After much discussion it was decided to hold the workshop on the Saturday before the IETF: people would be too busy during the IETF, whilst the Sunday was impossible for several key players who had to be at other meetings like the IESG and IAB.

The workshop was organised under the auspices of the IETF's MPTCP Working Group. The organising team was led by Philip Eardley (WG Chair, Trilogy Technical Manager), with much help from other Trilogists, as well as support from the IETF Secretariat and technical folks. We prepared a list of external experts whom we would have liked to participate in the workshop. We sent direct invitations to them, as well as broadcasting the "advert" below on various IETF mailing lists.

3 Scope

The scope of the workshop was described in the following "advert":-

Title: Multipath TCP Implementors workshop, Saturday 24th July

The IETF's Multipath TCP (MPTCP) working group is organising a workshop with the OS and applications communities on the Saturday afternoon immediately before the IETF meeting in Maastricht (i.e. on 24th July) (http://www.ietf.org/meeting/78/index.html)

If you would like to come, please register to mptcp-chairs at tools.ietf.org. Due to room size, numbers are limited and registration is required.

Multipath TCP enables a single TCP connection to use multiple network interfaces and paths simultaneously for one TCP connection. From an applications perspective, this increases resilience and enables a basic form of connection mobility, whilst from an OS perspective it requires a change to the host TCP stack.

IETF working group: http://www.ietf.org/dyn/wg/charter/mptcp-charter

Supplemental information: http://nrg.cs.ucl.ac.uk/mptcp/

The objective of the workshop is to help make MPTCP real, i.e., to get it implemented in many operating systems and to get it used by key applications. This will be an interactive workshop with application designers (who would use MPTCP), OS implementors (who would implement and ship MPTCP) and MPTCP working group people (who are designing and standardising MPTCP). Note that this is not a passive "dissemination" workshop! The aim is to have an active discussion with the OS

and applications communities about their requirements and needs, in order to influence and improve the MPTCP protocol design.

The workshop will be highly interactive and focus on two topics:

- On the one hand, it is to understand the process through which Multipath TCP could make its way into OSes. The OS implementers and active community members can explain their real world requirements and constraints on the various platforms. What are the potential stumbling blocks for MPTCP's implementation and how could the protocol designers lower the deployment barriers?
- On the other hand, the WG would like to discuss the use cases for Multipath TCP with the applications community. What are the gotchas and how can the protocol designers increase the usefulness of MPTCP?

Where & when?

The workshop will be held on July 24, 2010 in Maastricht at the IETF meeting venue, "MECC", starting at 2pm (room to be confirmed). This is the Saturday afternoon before the start of the IETF week.

We hope that remote participation will also be possible, via Webex.

We would be pleased to hear feedback on the scope and purpose of the workshop.

The workshop is organised under the auspices of the MPTCP working group, but is not a formal WG meeting - for instance, WG consensus calls will not be made.

Best wishes,

Philip Eardley

Yoshifumi Nishida

mptcp-chairs at tools.ietf.org

4 Agenda

The detailed agenda was:

- Intro, background, assumptions - Philip Eardley
- Demo of implementation - Sébastien Barré / Costin Raicu
- Use cases discussion
    - Mobility - discussion led by Lars Eggert
    - Data centres - discussion led by Costin Raicu
- OS implementation
    - Linux implementation - discussion led by Sébastien Barré and Costin Raicu
    - Offloading - Robert Watson

The slides are available from the website, along with the jabber log. Streamed audio was also available for remote participants, although not available for download due to unexpected technical issues. http://trac.tools.ietf.org/wg/mptcp/trac/wiki/Maastricht_workshop

5 Participants

The following people were registered for the workshop, there were about 25 participants. Most people attended in person, whilst there were also some remote participants.

Alan Ford	Roke Manor Research
Ananth Ramaiah	Cisco
Arnd Hannemann	RWTH Aachen University
Christian Vogt	Ericsson
Christoph Paasch	Université catholique de Louvain
Costin Raiciu	UCL, London
Dave Harrington	HuaweiSymantec
Deepak Kumar-Gupta	Nokia
Denis Collange	Orange FT
Ehsan Elahi	M. A. Jinnah University, Pakistan
Erik Nordmark	Oracle
Hui Deng	China Mobile
Iljitsch van Beijnum	Universidad Carlos III de Madrid
Javier Ubillos	Swedish Institute of Computer Science
Jinmei Tatuya	Internet Systems Consortium, Inc.
Julien Laganier	Qualcomm
Lars Eggert	Nokia
Lawrence Stewart	Swinburne University of Technology, Australia
Luigi Iannone	Technische Universität Berlin
Marcelo Bagnulo Braun	Universidad Carlos III de Madrid
Michael Scharf	Alcatel-Lucent
Mike Silby Silbersack	FreeBSD
Nandita Dukkipati	Google
Olivier Bonaventure	Université catholique de Louvain
Philip Eardley	BT
Richard Scheffenegger	NetApp
Robert N. M. Watson	University of Cambridge & FreeBSD
Rolf Winter	NEC
Ronald van der Pol	SARA (Dutch National High Performance Computing and e-Science Support Center)
Salvatore Loreto	Ericsson
Sébastien Barré	Université catholique de Louvain
Tim Shepard	Independent consultant and researcher

Yoshifumi Nishida WIDE Project
Zhen Cao China Mobile

6 Conclusions

6.1 Use cases

One of the most obvious use cases for MPTCP is to provide a form of mobility for mobile terminals – especially as terminals now have multiple interfaces. But the discussion suggested that there are alternative lower layer techniques for providing resilience. However, mobile terminals are highly power-constrained – so an interesting idea is to use MPTCP to shift traffic more aggressively away from the more power-hungry interface(s). There is already some research going on around this idea (for example, at Cambridge University – Ovidiu Popa had made a Guest Presentation during our Trilogy Plenary meeting a fortnight earlier).

In terms of the work of the WG, this may set a requirement on the "advanced API". Although unaltered applications can use MPTCP (ie the API is the same as for TCP), enhanced applications could use an advanced API with extra features – for example with information about the energy usage of different interfaces.

There was considerable interest in the use of MPTCP in data centres, for load balancing to improve performance. Throughput and resilience are both important in data centres, and proprietary solutions are often used. MPTCP would provide a standard solution to these issues, potentially allowing a wider range of standards-compliant equipment to be used. Also, early simulations (made within the Trilogy project) show there are significant gains in typical topologies, perhaps increasing the throughput from 40% to 80% of the theoretical maximum.

The discussion during the workshop suggests that this use case is particularly interesting because everything is under the control of the data centre operator. Also, deployment in controlled environments such as this would potentially permit the use of a cut-down MPTCP without the need of middlebox-traversing features, and with minimal security. However, the general feeling was that a lightweight variant of MPTCP is not worthwhile. Firstly, it creates interoperability issues, such as negotiating between versions and avoiding potential bidding-down attacks; and secondly, offloading engines on data centre line cards are effectively a form of middlebox. It is clear there are plenty of research issues about MPTCP in data centres, for example to detail the performance gains and to research how it changes the optimum topology.

The workshop concentrated on the above two use cases. Other potential ones include: for a campus network with two providers (for resilience), inside a network (for fast traffic engineering), and with a proxy server for deployment to MPTCP-unaware hosts or hosts on single-homed links.

6.2 OS Implementation

The important outcome of the workshop was the insight that that the design of MPTCP must support the *offloading* of MPTCP to NICs. Today, many TCP/IP features (such as segmentation (TSO), reassembly (TRO), and checksumming) are already offloaded to NICs, and (especially for high-performance environments) it is essential that MPTCP too can be offloaded, otherwise the performance gains will be offset by the need to implement on the CPU. This discussion has had significant impact on MPTCP design decisions, not least the decision to implement using TCP options, which can be kept separate from the data stream.

There are other considerations, such as packets are likely received out of order (and reassembled) more often. This implies an increased buffer size and delay. Further study is needed.

We were very pleased to get active participation from the OS community, including those who are “gatekeepers” for the FreeBSD kernel. We also had some participation from commercial OS vendors. The impression seemed to be that commercial vendors would wait and make an assessment from free versions. Specialist vendors (e.g. data centres) may take the free version, when proven, and adapt it to their needs.

6.3 Organisational

The participants were enthusiastic about the workshop and felt it had been useful. People believe it would be worthwhile to have another workshop, perhaps in the Bay Area and perhaps when the implementations are more advanced.

The remote participants did not find the workshop easy. During the IETF meetings the audio for remote participants (normally) works well. However, on the Saturday before the IETF (when we held the workshop) the audio was in the process of being set-up. This led to a few glitches, such as a change of audio stream and unfortunately no recording. There was also a need for more dedicated discussion time, and for the WG Chair to remember to poll remote participants! Two remote participants kindly made some detailed suggestions, which will be taken into account in a future workshop.