Multi-Path TCP Performance Evaluation in Dual-Homed (Wired/Wireless) Devices

Juan Antonio Cordero[a,*]

[a]*ICTEAM, Université catholique de Louvain. Place Sainte Barbe, 2. B-1348 Louvain-la-Neuve (Belgium)*

Abstract

Multipath TCP is a major extension of TCP, designed for leveraging the increasing availability of multiple interfaces in end hosts, on one side, and the existence of diverse Internet paths between hosts, on the other. This paper proposes a measurement methodology and provides a first evaluation, based on real Internet experiments, of the user benefit of using MPTCP instead of TCP in devices with multiple wireless/wired networking interfaces. We focus on bandwidth utilization and file transfer delays. Our experiments, on a testbed with two disjoint paths connecting a server and a dual-homed probe, indicate that MPTCP is able, in most cases, to take advantage of additional bandwidth with limited cost in terms of delay, but also show that the MPTCP bandwidth benefit substantially degrades when the interfaces have very different bandwidth capacities.

Keywords: TCP; measurements; transport; wireless; MPTCP; multipath; bandwidth; throughput; delay; experimental

1. Introduction

In the TCP/IP architecture, a transport connection is a single path between two host interfaces. This was adapted to the original situation in the Internet, in which end devices were mostly connected to the Internet via a single networking interface. This has however changed in the last decades: more and more end devices are multi-homed hosts, meaning that they have multiple networking interfaces. Popular examples include smartphones (3G/4G and WiFi interfaces), laptops (Ethernet/WiFi interfaces) and multi-homed servers in datacenters. Even for single-interface hosts, transport throughput is not usually limited by the host interfaces, and therefore throughput could be increased by leveraging path diversity in the Internet, *e.g.* by way of ECMP inside ISPs. In

*Corresponding author. E-mail: cscordero@comp.polyu.edu.hk
Present address: *Department of Computing, The Kong Kong Polytechnic University. Mong Man Wai Building. Hung Hom, Kowloon (Hong Kong SAR, China)*

this context, the use of a single transport path often entails an underutilization of network resources.

Multipath TCP (MPTCP) [5] is a major extension of TCP that leverages at the transport layer the existence of multiple Internet paths (typically available over multiple networking interfaces) between end devices. MPTCP has been standardized at the IETF (RFCs 6182 [16], 6824 [9]) and, unlike other multipath transport proposals, *e.g.* SCTP, it is transparent for applications and backwards-compatible with regular TCP. This eases its wide adoption by users and its incremental deployment in the Internet. MPTCP has been adopted, for instance, by Apple in iOS7-based devices. The growth of multipath opportunities in the Internet and the rise of MPTCP as a feasible and increasingly deployed extension to single-path TCP has attracted considerable attention in the research community to the performance evaluation of Multipath TCP in multiple scenarios.

1.1. Contributions

This paper focuses on the study of the potential benefit of using Multipath TCP instead of regular TCP, from a final user perspective, for devices with both wired and wireless (WiFi) networking interfaces. While several works have recently focused on multipath transport evaluation, the WiFi/Ethernet case in the Internet edge has not been properly explored.

The contribution of this paper is thus three-fold. First, it discusses and proposes basic user-experience metrics for evaluating transport performance and specifies two new bandwidth and delay aggregation benefit metrics for MPTCP. Second, it derives a consistent measurement methodology from these two metrics; this measurement methodology is implemented and publicly available, and can be used or extended for more demanding purposes. Third, the paper explores the benefits of using MPTCP instead of TCP in a real Internet networking testbed with multi-homed (wireless/wired) devices.

This is the first paper that addresses, in a real networking scenario, the performance of multipath transport for multi-homed devices connected to (partly) separate paths via WiFi and Ethernet interfaces. This is a situation that can be found in laptops and is also a feasible scenario for IoT-based sensor deployments in which sensoring devices support wired and wireless connectivity. Based on the result of 3000 experiments distributed along several weeks, the paper investigates how MPTCP leverages the existence of multiple (wireless/wired) Internet paths between endpoints of a transport connection and explores the interaction between multipath transport and the most relevant congestion control mechanisms. The analysis of performed experiments shows that the use MPTCP can be beneficial, but it degrades substantially when available paths have very different bandwidth capacities. Also, differences in path latency may lead MPTCP to perform worse than TCP, if transmission duration is not sufficient to take advantage to utilize several paths.

1.2. Paper Outline

The remainder of the paper is organized as follows. Section 2 reviews and discusses existing literature related to MPTCP performance analysis. Section 3 presents the aspects of MPTCP performance that are studied in the paper, specifies the considered metrics, the procedure and the experimental scenarios in which measurements are performed. Section 4 describes the main observations from the performed experiments. Finally, section 5 concludes the paper.

2. Related Work

Several tools have been proposed in the literature for bandwidth and network performance estimation purposes [21]. Prasad *et al.* (2003) [20] provide an extensive survey on metrics and standard available tools for bandwidth estimation. The measurement tool proposed and used in this paper focuses on the estimation of end-to-end capacity and transport delay. It is partly inspired by *Iperf* [25] and *NetPerfMeter* [23], but its design is simplified and adapted to the specificities of MPTCP/TCP measurements and comparison.

We examine the performance of Multipath TCP by way of two user metrics: the Bandwidth Aggregation Benefit and the Delay Benefit. The first metric relies on a definition originally proposed by Kaspar (2011) [17] and later adopted by Paasch *et al.* (2013) [8]. We adapt it in this paper to support non-disjoint transport paths. The second metric is also inspired by the same intuition of Kaspar. In both cases, the objective is to provide a bounded estimation of the relative benefit (or penalty) when using MPTCP instead of TCP.

Evaluation and optimization of TCP have been extensively addressed in the last decades, but the interest for Multipath TCP is more recent. MPTCP performance evaluation has only attracted attention in the last years. Raiciu *et al.* (2012) [14] describe the main design choices on MPTCP implementation in Linux kernel, and provide a first evaluation of their impact, both in simulated environments and with real dual-homed wireless/wireless (3G/WiFi) scenarios.

The wireless/wireless scenario (with 3G or LTE and WiFi) has been widely explored in other works. Results obtained in this scenario, however, cannot be mechanically extrapolated to other scenarios involving wired paths, partly due to the specific characteristics of cellular and wired networks [11, 13]. Raiciu, Niculescu *et al.* (2011) [19] provide the first evaluation, both by way of simulations and by indoor mobility experiments, of the potential bandwidth benefits of MPTCP in 3G/WiFi scenarios. Paasch *et al.* (2012) use the kernel implementation to examine, through real experiments, the ability of MPTCP to handle handovers between WiFi and 3G networks. Chen *et al.* (2013, 2014) [7] [6] analyze the performance of MPTCP in real mobile wireless/wireless scenario, with dual-homed devices (smartphones) with 3G/4G and WiFi interfaces. Their work focuses on the user benefits in terms of bandwidth and delay in these scenarios [7], and the modeling and understanding of some of the main issues that arise, in particular the delayed startup of the second subflow and bufferbloat (that is, the excessive variation of delays due to overdimensioned buffers, typically observed in cellular and WiFi scenarios), and their impact on MPTCP performance [6]. They observe that MPTCP achieves similar latencies to those obtained by the best available path alone, can outperform it for sufficiently large downloads and is able to reduce latency variability. More recently, Deng *et al.* (2014) compare the performance of WiFi and LTE paths and examine the potential gains of MPTCP in this case.

Raiciu, Barré *et al.* (2011) [18] explore the performance of MPTCP in datacenters and show that the use of MPTCP can leverage efficiently the existence of multiple redundant paths in this situation, and allows further optimizations of datacenters topology. From a wider evaluation perspective, Paasch (2013) [8] introduces the notion of "experimental design" and applies it to perform an extensive MPTCP performance evaluation in a broad range of emulated environments, with dual-homed devices, via Mininet.

This paper complements the previous literature on MPTCP performance, but introduces a novel measurement methodology/metric approach and fills a gap in the space of MPTCP measurement studies. To the best of our knowledge, it is the first to investigate user benefits, behavior and limiting factors of MPTCP with real experiments in the edge of the Internet, over scenarios with separate wireless (WiFi) / wired paths, both local and in the Internet. Aside from specific contributions, presented results confirm previous observations in simulation-based experiments and are consistent with real testbed observations in other scenarios explored by the literature – in particular, 3G/WiFi scenarios.

3. Measurement Methodology

Experiments are performed by way of a specific software, designed to define and launch sequences of experiments between probes (Measurement Agents, MAs) and a centralized server (Measurement Server, MS). Source code is publicly available [26]. The MA determines the experiments setup, scheduling and performs metric computation; measurements are reported to the MS. This section details the main focus of the MPTCP performance analysis (section 3.1), presents the measurement architecture and experimental procedures (section 3.2), and describes the testbed and additional scenarios in which measurements have been performed (section 3.3).

3.1. Metrics

Multipath TCP performance is studied from the perspective of the final user. The evaluation compares basic performance indicators such as bandwidth (transport goodput) and file transfer delay in MPTCP and standard TCP. Two metrics are identified: a slight variation of the Bandwidth Aggregation Benefit metric, proposed by Kaspar (2011) [17] and already used in MPTCP analysis by Paasch (2013) [8], and the Delay Aggregation Benefit metric, derived from the same principle. Implementation of these metrics is detailed in section 3.2.

3.1.1. Transport Disjointivity

Given a set of n Internet paths $S = \{p_1, p_2, ..., p_n\}$, the *transport disjointivity index* (TDI) of S is defined as the ratio between the maximum transport throughput that can be achieved when using simultaneously all paths in S, denoted by C_{total}, and the addition of transport throughput achievable (separately) over each path i, C_i:

$$\text{TDI} = \frac{C_{\text{total}}}{\sum_{i=1}^{n} C_i} \in [0, 1] \tag{1}$$

According to this definition, the TDI reaches the maximum value (1) when $C_{\text{total}} = \sum_{i=1}^{n} C_i$, and approaches zero when the combined multipath transport throughput becomes negligible with respect to the addition of separate single-path transport throughputs ($C_{\text{total}} \ll \sum_{i=1}^{n} C_i$). Note that this is not equivalent to the topological notion of path disjointivity: topologically-disjoint paths are necessarily transport-disjoint paths, but transport-disjoint paths do not need to be topologically disjoint (see example in Fig. 1).

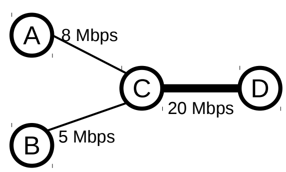

Figure 1: Example of transport disjointivity of paths not topologically disjoints. *Paths $\{A, C, D\}$ and $\{B, C, D\}$ are topologically non-disjoints (they share link (C, D) but are transport-disjoint, since their joint capacity equals the addition of separate capacities).*

In the general case, when a set S of paths is not transport-disjoint, C_{total} needs to be explicitly measured to estimate the maximum theoretical throughput, as it cannot be deduced from the sum of individual throughputs C_i, for each interface i.

3.1.2. Bandwidth Aggregation Benefit

In this context, *bandwidth* between two endpoints is defined as the average goodput achievable during a certain amount of time via a reliable transport protocol (TCP or MPTCP).

Let n be the number of networking interfaces available at the Measurement Agent (MA). Then, MPTCP will try to set up n different transport subflows (leading to potentially disjoint transport paths) between the MA and the Measurement Server (MS). For each MA's interface i, let C_i be the bandwidth capacity for the path traversed by the subflow associated to interface i and $C_{\max} = \max_{i \leq n}\{C_i\}$. Note that, if the corresponding path is compound by links $\{l_1, l_2, ..., l_m\}$, and $c(.)$ denotes the bandwidth capacity of a link, $C_i = \min_{i \leq m}\{c(l_i)\}$. Let C_{total} be the joint capacity of paths from MA's interfaces and the MS. Note that $C_{total} \leq \sum_{i=1}^{n} C_i$, equality holding when the n paths are completely disjoint (from a transport perspective, see section 3.1.1).

This latter case of fully disjoint paths ($C_{total} = \sum_{i=1}^{n} C_i$) corresponds to the strict definition of Kaspar (2011) [17].

The Bandwidth Aggregation Benefit when using MPTCP instead of regular TCP over a set S of Internet paths, and achieving a goodput x, $BW_{\text{benefit}}(S, x)$, is defined as follows:

$$BW_{\text{benefit}}(S,x) = \begin{cases} \frac{x - C_{\max}}{C_{\text{total}} - C_{\max}}, & x \geq C_{\max} \\ \frac{x - C_{\max}}{C_{\max}}, & x < C_{\max} \end{cases} \in [-1, 1] \tag{2}$$

$BW = 1$ corresponds to the maximum possible benefit (*i.e.*, MPTCP achieves all the available bandwidth C_{total}), $BW = -1$ corresponds to the worst case (MPTCP gets 0 Mbps), and $BW = 0$ implies no benefit, *i.e.*, MPTCP behaves just as good as TCP.

3.1.3. Delay Aggregation Benefit

For a file to be transmitted between two endpoints by way of a reliable transport protocol (TCP or Multipath TCP), the term *delay* is defined as the amount of time between the transmission of the first bit by the transmitting endpoint and the reception of the acknowledgement for the last bit of the file at the transmitting endpoint. This definition excludes the time required for the establishment of a TCP/MPTCP connection, which is considered separately in this work.

Let D_i be the delay when using a TCP connection over MA's interface i. $D_{\min} = \min_{i=1}^{n}\{D_i\}$, $D_{\max} = \max_{i=1}^{n}\{D_i\}$ and D^*_{th} defines the theoretical delay that could be ideally achieved if all available paths were used with perfect scheduling at the same time. In general, $D^*_{\text{th}} \leq D_{\min}$. Following the philosophy of Kaspar's aggregation benefit metrics, the Delay Aggregation Benefit for a set S of Internet paths between the MA and the MS, when using MPTCP instead of regular TCP, and measuring a delay x when transmitting a file of B bytes, is denoted by $D_{\text{benefit}}(S, B, x) \in [-1, 1]$, and is defined:

- Positive if the delay with MPTCP is better (smaller) than the delay achieved by TCP over the fastest interface; negative otherwise.
- Maximum (=1) when the MPTCP delay corresponds to (or outperforms) the best possible delay D^*_{th}.
- Minimum (=0) when MPTCP achieves a delay equal or larger than the TCP delay over the slowest interface, $D_{\max}$.

Or, equivalently:

$$D_{\text{benefit}}(S,B,x) = \begin{cases} 1 & , x \leq D^*_{\text{th}} \\ \frac{D_{\min} - x}{D_{\min} - D^*_{\text{th}}} & , D^*_{\text{th}} < x \leq D_{\min} \\ \frac{D_{\min} - x}{D_{\max} - D_{\min}} & , D_{\min} < x \leq D_{\max} \\ -1 & , x > D_{\max} \end{cases} \in [-1, 1] \tag{3}$$

Particular case: two subflows ($n = 2$). This paper explores the performance of Multipath TCP in a two-subflow scenario. In this particular case, the general expression for the Delay Benefit metric (Eq. (3)) can be reformulated as follows.

Let $k = \frac{D_{\text{max}}}{D_{\text{min}}}$ be a compression parameter. It is immediate to observe that the smallest theoretical delay that MPTCP could reach in a two-subflow scenario is $\frac{k}{k+1}$ times the minimum delay obtained in a single path.

Therefore, in the considered scenario with dual-homed devices ($n = 2$), it can be shown that D^*_{th} can be computed as $D^*_{\text{th}} = \frac{k}{k+1} D_{\text{min}}$, and Eq. (3) becomes:

$$D_{\text{benefit}}(S, B, x)|_{n=2} = \begin{cases} 1 & , x \leq \frac{k}{k+1} D_{\text{min}} \\ (k+1)\frac{D_{\text{min}}-x}{D_{\text{min}}} & , \frac{k}{k+1} D_{\text{min}} < x \leq D_{\text{min}} \\ \frac{D_{\text{min}}-x}{(k-1)D_{\text{min}}} & , D_{\text{min}} < x \leq k D_{\text{min}} \\ -1 & , x > k D_{\text{min}} \end{cases} \tag{4}$$

This characterization of $k = \frac{D_{\text{max}}}{D_{\text{min}}}$ allows to obtain a configuration-independent expression for the MPTCP delay benefit.

3.2. Measurement Architecture and Procedures

The architecture of the measurement tool and the interactions between MAs and MS are displayed in Fig. 2. For each individual experiment launched by a MA, a control communication channel is set up between the MA and the MS. Different measurements are collected over exchanges performed through one or several dedicated data communication channels.

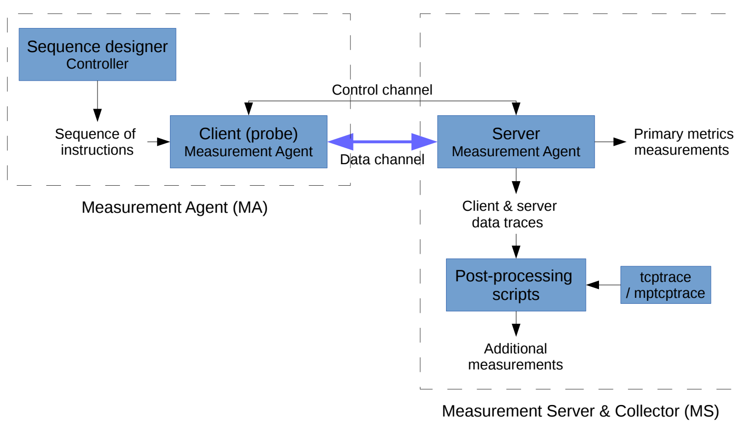

Figure 2: Measurement tool architecture.

The tool supports bandwidth and delay experiments, with two transmission modes emulating two standard patterns on user data transport: oneway transmission in MA $\longrightarrow$ MS direction (`oneway` mode), and request/response exchange in MA $\longleftarrow$ MS direction (`reqresp` mode). In each case, the delay is measured between the transmission of MA's first bit until the connection closing (by the MA after receiving the acknowledgement of the last bit from the MS, by the MS after receiving the acknowledgement of the last bit from the MA).

Experiment	*Input*	*Mode*	*Main output*
Bandwidth	Transmission time ($\sim$ 60 sec)	`output|reqresp`	BW Aggregation Benefit $\in [-1, 1]$
Delay	Number of bytes $\in [10^3, 10^7]$ B	`output|reqresp`	Delay Aggregation Benefit $\in [-1, 1]$

Table 1: Main parameters of experiments.

3.2.1. Procedure for an Individual Experiment

A typical experiment involves three main steps:

1. Hosts negotiate measurement parameters, over the control channel MA-MS.
2. The MA performs different active measurements, specific to the type of experiment, over one or several data channels MA$\longleftrightarrow$MS.
3. MA reports measurements to the MS, over the control channel.

MA reports involve the metrics computation, and can also include TCP and MPTCP connection traces. Collection of these traces allow further and finer analysis by way of specific tools, such as *tcptrace* [22] and *mptcptrace* [4].

3.2.2. Types of Experiments

In bandwidth experiments, the MPTCP Bandwidth Aggregation Benefit is computed by measuring the amount of user data that can be sent over the available flows (TCP or MPTCP, see below) during a fixed amount of time (typically, 60 sec). In delay experiments, the MPTCP Delay Aggregation Benefit is computed by measuring the *delay* (in the sense of section 3.1.3) to transmit a fixed amount of bytes over the available flows (TCP or MPTCP, see below). Table 1 summarizes each experiment's main characteristics.

During an experiment, measurements on the network state are assumed to be valid for the duration of the experiment. That is, the tool implicitly assumes that the state of the network *does not change* during a single experiment, which can last several minutes.

A bandwidth experiment consists of the following active measurements:

1. For each available interface i, TCP connection and transmission over the path (data channel) associated to MA's interface i (for n MA's interfaces, n TCP MA $\longleftrightarrow$ MS connections).
2. Simultaneous TCP connections (data channels) over all paths associated to MA's available interfaces, to estimate the total available bandwidth for reliable transport and the TCP throughput over all interfaces (for n MA's interfaces, n TCP MA $\longleftrightarrow$ MS connections). Note that this step is only redundant with step 1 if paths associated with different MA interfaces are transport-disjoint (*i.e.*, the total bandwidth is the addition of the bandwidth available over each available interface, separately).

3. Transmission using MPTCP over a (multipath) data channel using all MA's available interfaces (1 MPTCP MA $\longleftrightarrow$ MS connection).

Comparing the total available TCP throughput obtained in steps 1 and 2 allows to estimate the transport disjointivity index (TDI) of paths between MA and MS, defined as in Eq. (1).

For a transmission time t, the expected time for the completion of a bandwidth experiment in the data plane is $(n+2) \times t$ (exchanges on the control plane have negligible duration with respect to active measurements in the data plane).

For a delay experiment, active measurements consist of:

1. For each available interface i, TCP connection and transmission over the path (data channel) associated to MA's interface i (for n MA's interfaces, n TCP MA $\longleftrightarrow$ MS connections).

2. Transmission using MPTCP over a (multipath) data channel using all MA's available interfaces (1 MPTCP MA $\longleftrightarrow$ MS connection).

Given b bytes to be transferred, the expected time for the completion of a delay experiment in the data plane is thus $b \times \left(\sum_{i=1}^{n} \frac{1}{BW_{\text{TCP}(i)}} + \frac{1}{BW_{\text{MPTCP}}} \right)$.

Fig. 3 displays the scheme of the steps performed for MPTCP bandwidth and delay tests over a dual-homed MA. Scheme (c) is only performed in the case of bandwidth tests (step 2), as it does not bring any meaningful information in terms of delay. The order of these steps in each test is randomized at each experiment, to avoid possible statistical bias.

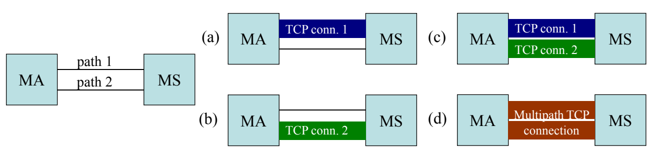

Figure 3: Diagram with steps performed in MPTCP bandwidth and delay measurement tests.

3.2.3. Sequences of Experiments

In the measurements campaigns described in this paper, MAs execute sequences of independent experiments. These sequences are also randomized and stored in the MAs, for reproducibility purposes. Time between two consecutive experiments from the same MA can be constant (thus, leading to periodic MA measurements) or random following an exponential distribution (thus, emulating a Poisson event distribution on the MS), depending on the setup. See section 3.3 for more details on the experimental setup.

	Server (MS)	*UCL Probe (MA1)*	*Aalto Probe (MA2)*
MPTCP version	0.88.8	0.88.x	0.89.2
Congestion control	LIA	OLIA, LIA, Cubic	Cubic
Receiving window	4096, 87380, 3752192	4096, 87380, 1009856	4096, 87380, 6291456
Sending window	4096, 16384, 3752192	4096, 16384, 1009856	4096, 16384, 4194304
Networking interfaces	`eth0`	`eth0.2`, `wlan1`	`eth0`, `eth1`, `wlan0`

Table 2: Main characteristics of the MS and the MAs at UCL (MA1) and Aalto (MA2). *Values for receiving and sending windows correspond to minimum, initial, maximum.*

3.3. Testbed Characteristics

A centralized Measurement Server (MS) is deployed in an Internet-reachable machine at UCL; Measurement Agents are expected to run experiments against this MS. All involved machines use the MPTCP kernel implementation [24]. Main characteristics of the MS are detailed in Table 2 and additional details on the setup are available in further documentation [27]. We have deployed a local MA at the *Université catholique de Louvain* (UCL, Belgium) and a remote MA at Aalto University (Finland). Fig. 4 shows schematically the topology connecting the involved devices and indicating the observed order of magnitude of path RTTs; the following subsections describe more in detail the two experimental setups.

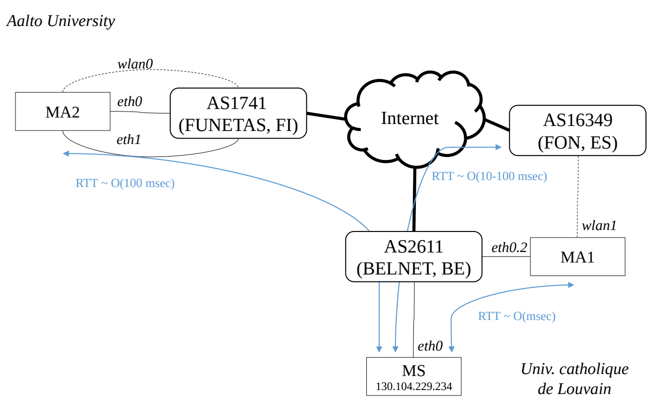

Figure 4: Topology of local and remote measurement setup. *Dotted lines represent wireless interfaces, solid lines represent wired interfaces. Values in the arrows indicate orders of magnitude of RTTs of paths between MAs and MS.*

3.3.1. Local Testbed (UCL)

For the local testbed experiments, the MA has been deployed on a router TL-WDR4900 with a MPTCP-capable kernel (see Table 2). This MA (MA1) is connected to the Internet by way of two networking interfaces: an Ethernet interface attached to the UCL network (connected to AS2611, BELNET), and a WiFi interface to a FON/Belgacom WiFi hotspot (AS16349), able to reach the MS through the Internet via a commercial ADSL connection. The two available paths between MA and MS are both topologically- and transport-disjoint; one is internal in the UCL network (130.104.0.0/16 BGP prefix, inside BELNET AS) and the other is an inter-domain path (see Fig. 4). The intra-domain path available through the MA wired interface shows a RTT in the order of milliseconds, with peaks at 40 msec. The inter-domain path available through the MA wireless interface is significantly slower (between tens and hundreds of milliseconds, with peaks of seconds) and more variant.

The measurement sequences are computed and randomized at the MA, and the corresponding measurements are performed periodically, with a 5 min (300 sec) interval between the start of two consecutive experiments. For bandwidth experiments, each transport connection lasts 60 sec. For delay experiments, the amount of transmitted bytes is selected randomly. The duration of each performed experiment is typically smaller than 60sec – in most cases substantially smaller.

Several measurement sequences have been performed in this scenario. Bandwidth configurations are parametrized by way of the *Bandwidth Ratio* (BWR), defined as the quotient between the goodput obtained on the MA's wired interface and the goodput on the MA's wireless (WiFi) interface. Scenarios with BWR= $\{1...15, 20\}$ are tested. Different BWRs are emulated by shaping traffic in the MA wired (`eth0.2`) interface, by way of the `tc` tool. The uploading rate on the wireless interface being $\sim$ 1 Mbps, the traffic rate is shaped in interface `eth0.2` at k Mbps, for $BWR = k$. Since traffic shaping only affects outgoing traffic (from the MA), only `oneway` experiments are performed. For each BWR configuration, bandwidth and delay measurements are performed. In the latter ones, the amount of bytes to be transmitted for each experiment is determined randomly in each sequence following a LogUniform distribution within intervals $[10^2, 10^6]$ bytes and $[10^3, 10^7]$ bytes.

Measurements are mostly performed by using the two main congestion mechanisms for MPTCP: the Linked-Increases Algorithm (LIA) [15] and the Opportunistic Linked-Increases Algorithm (OLIA) [2]. In LIA, the increase of a subflow congestion window in the additive increase phase is related to the state of the global (MPTCP-level) congestion window, and it is upper-bounded by the increase that would be performed by standard TCP. OLIA adopts a more complex strategy, more aggressive in high capacity paths and less in low capacity paths, thus more efficient in moving traffic away from congested paths [10].

In total, 3064 experiments have been performed at disjoint time intervals; 1024 with OLIA configuration, 979 with LIA configuration and the rest with the MA default congestion control (Cubic). BWR configurations are distributed uniformly among the total number of experiments.

3.3.2. Remote MA configuration

The remote MA (MA2) runs a sequence of `oneway` and `reqresp` experiments. Time between two consecutive experiments is computed randomly, following an exponential distribution with $\lambda = 0.0023\frac{\text{exp}}{\text{sec}}$, *i.e.*, 200 daily experiments in average. The MA has three networking interfaces (see Table 2). From our observations, paths between MA2 and MS in both directions are transport-disjoint, with estimated capacities (average) 4, 93 and 220 Mbps (MA2 $\longrightarrow$ MS) and 2, 22 and 50 Mbps (MS $\longrightarrow$ MA2). Path capacity estimations vary however significantly along the time. Note that, according to Fig. 4, observed paths are not topologically disjoint in the Internet, as the three interfaces are connected to the same Autonomous System (AS1741) and share the same BGP prefix (195.148.0.0/16). No traffic shaping is performed in this MA. In total, 132 experiments have been performed on this probe, half on each direction of communication. Path RTTs are in the order of magnitude of hundreds of milliseconds, with high variance.

4. Results and Discussion

4.1. MPTCP Goodput

Fig. 5 compares the BW benefit for different configurations (control congestion mechanism and BWR). It can be observed that the benefit is higher, in ideal conditions (completely disjoint paths), when bandwidth capacities of available interfaces are similar (*i.e.*, benefit tends to 1 when BWR $\sim$ 1). The curve may change depending on the congestion control or other factors, as it can be observed by comparing OLIA [2] to LIA [15] but the trend is common. When the difference (in terms of goodput) between interfaces grows bigger, MPTCP's ability to achieve available bandwidth degrades, regardless of the congestion control mechanism; MPTCP may even achieve less goodput than TCP over the fastest interface.

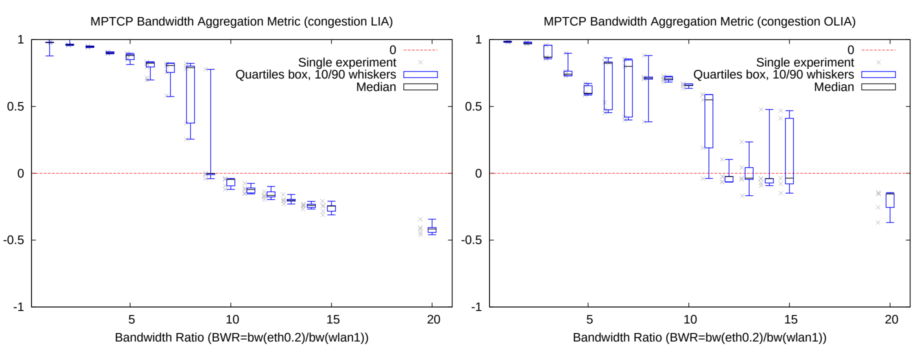

Figure 5: BW Aggregation Benefit, for LIA and OLIA, and different BWRs.

Fig. 6 displays the absolute value of the MPTCP goodput, with OLIA [2] and LIA [15] congestion control schemes. The dashed line indicates the upper bound for MPTCP goodput, corresponding to the joint capacity of available interfaces. Degradation (with respect to the maximum possible goodput) grows in both cases as the BWR increases. Experiments indicate that OLIA improves substantially, compared to LIA, the ability of MPTCP to utilize a larger fraction of the available bandwidth, especially in scenarios in which interfaces are significantly different. This is consistent with the more aggressive heuristic of OLIA (with respect to LIA) for increasing the congestion window during the congestion avoidance phase, already documented in the literature [3] [10].

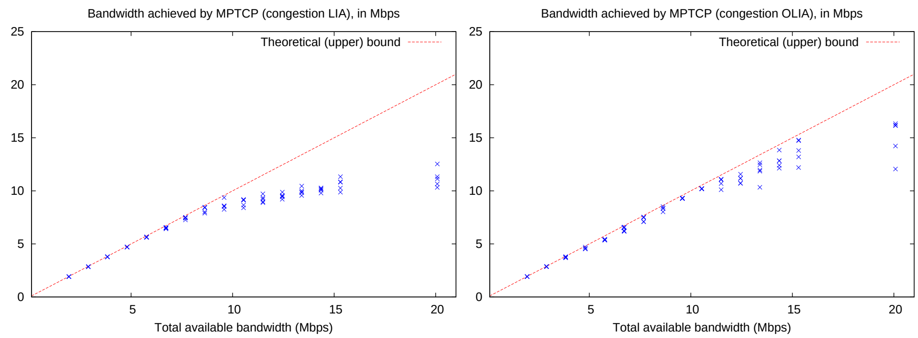

Figure 6: Available goodput achieved by Multipath TCP, for LIA and OLIA congestion control mechanisms, for different available transport capacities ($\sim 1 + \text{BWR}$ Mbps).

4.2. File Transfer Delay

Fig. 7 displays the cumulative distribution function (CDF) of the transmissions for which the delay benefit was positive and negative, both with respect to the file size (Fig. 7(a)) and the MPTCP delay (Fig. 7(b)). It can be observed that the statistical distribution of the experiments for which the delay benefit is negative is close to the overall distribution of experiments (see section 3.3.1 for a detailed description of the experimental setup). Experiments with a positive MPTCP delay benefit, in contrast, show a clearly different statistical distribution: cases in which the use of MPTCP is beneficial in terms of delay correspond to relatively long (in time) / large (in number of bytes) transmissions.

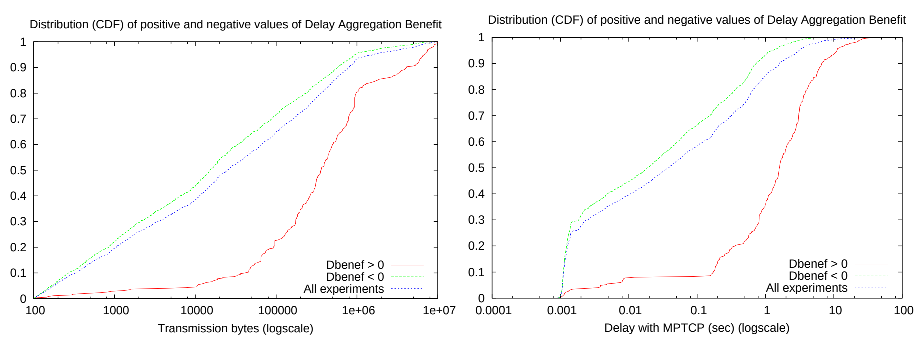

Figure 7: CDF of positive and negative values of Delay Aggregation Benefit, with respect to **(a)** the file size, in bytes, and **(b)** the MPTCP delay, in sec.

This can be observed in more detail in Fig. 8, which shows the point cloud for delay benefit over all performed experiments. For better understanding, a box-and-whiskers diagram is included, showing the median, the 25% and 75% quartiles (box bounds) and the 10%-90% deciles (whiskers). Displayed quantiles are computed over the experiments involving transmissions of bytes within intervals $[100 \times 10^k, 200 \times 10^k)$, $[200 \times 10^k, 500 \times 10^k)$ and $[500 \times 10^k, 1000 \times 10^k)$, $\forall 0 \leq k \leq 5$.

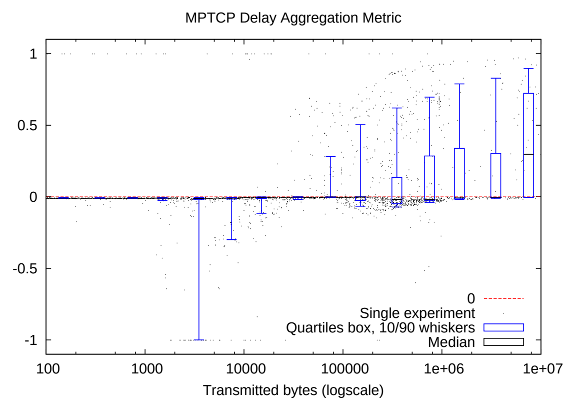

Figure 8: Box-and-whiskers diagram of MPTCP Delay Aggregation Benefit.

For all file sizes, the MPTCP delay benefit is close to 0 in most of the experiments, meaning that, in practice, the MPTCP delay is similar to the delay achieved by TCP over the fastest interface. Delays with MPTCP approaching the TCP delay over the slowest interface remain rare; the use of MPTCP seems to not harm transport performance in terms of delay.

Two regions can be observed in Fig. 8. For very short/small files (up to 200000 bytes), MPTCP incurs in (small) additional delay with respect to TCP over the fastest interface. The MPTCP delay benefit is thus usually negative but almost negligible (close to zero), and has a low variance. For longer/larger transmissions, the variance in the benefit grows bigger and, more in particular, the fraction of experiments with a positive aggregation benefit (that is, in which MPTCP outperforms the fastest TCP transmission) increases substantially. For very large transmissions (in the order of 10 MB), MPTCP achieves a smaller delay than TCP over the best interface in most of the performed experiments. The observed benefit of MPTCP for relatively long/large transmissions is consistent with previous observations in other (3G/WiFi) scenarios [7].

4.2.1. Multipath Utilization

This different behavior is related to the MPTCP utilization of the available paths: for experiments in the first region (small/short transmissions), MPTCP is only able to effectively use one subflow: the transmission terminates before the second subflow has been successfully established and is operational. Fig. 9(a) shows the utilization of the most loaded path; it is 1 when only one path is used in the transmission (*i.e.*, MPTCP behaves as TCP). The separation between both regions thus depends on the transmission time, as expected and shown in Fig. 9(b), in which different transitions can be observed for different BWR configurations. From our observations, the time between the establishment of two consecutive subflows (which determines the threshold from which multiple paths can be effectively used) is in the order of hundreds of msec, with minimum value in 35 msec: file transfers with shorter duration in this scenario are not able to take benefit of multiple paths.

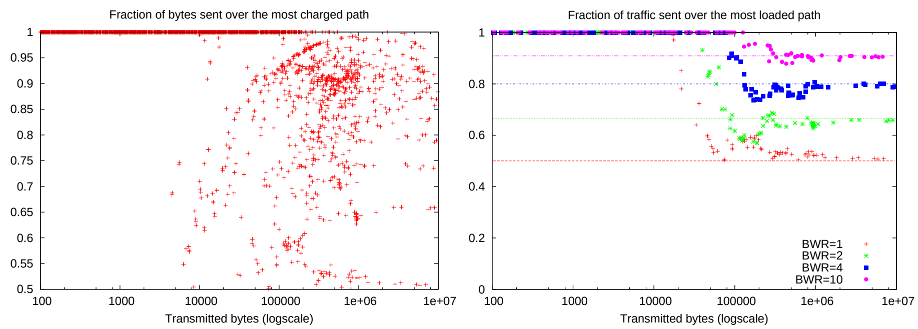

Figure 9: Fraction of bytes sent over the most loaded path, **(a)** for all BWRs confound, and **(b)** for BWR$= 1, 2, 4, 10$.

Note that, for sufficiently long/large transmissions, load distribution among the available paths converges to its bandwidth distribution; in the studied 2-flow scenario, the portion of traffic over the best path roughly corresponds to $\frac{\text{BWR}}{1+\text{BWR}}$.

4.2.2. MPTCP Connection Establishment Time

Performed experiments allow to compare the connection establishment (time immediately before and after the `connect()` call towards the data socket, measured in the client) in MPTCP and in regular TCP, not included in previous delay comparisons (see section 3.1.3). It can be observed that the connection establishment in MPTCP is slightly slower than the fastest TCP connection establishment. This is expected (as the 3-way handshake in MPTCP involves the exchange of more data and computing keys at both endpoints), but the difference is small, in the order of hundreds of microseconds, as it can be observed in Fig. 10.

4.3. Remote Probe Observations and General Discussion

Experiments on the remote probe MA2 confirm the main patterns observed in the local UCL testbed; results show a more significant variance, as expected due to the longer length and higher delays of studied Internet paths, with respect to those from the local testbed. Bandwidth experiments show the same trend: the metric degrades, in average, as the quotient between highest and lowest path goodput estimation (BWR) increases. BWR being in the order of 10 for

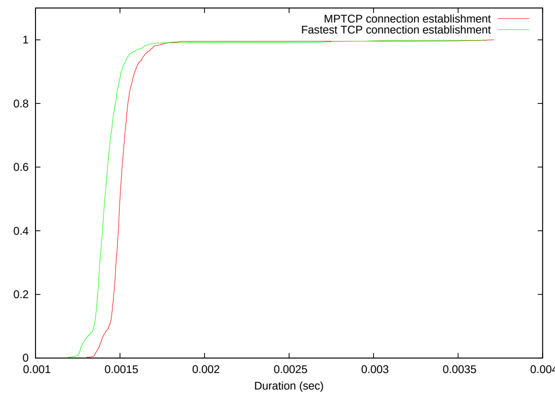

Figure 10: CDF of the MPTCP and fastest TCP connection establishment time.

most of the performed experiments (and in the order of 100 for many `oneway` experiments), the BW benefit is typically negative.

Similarly, for delay experiments, it can also be observed that the probability of positive delay benefit increases with the transmission size. In particular, in `oneway` experiments, transmissions larger than 60000 bytes are sent typically faster with MPTCP than with TCP over any interface, as MPTCP is able to leverage the presence of multiple available paths.

Although local and remote observations are not sufficiently general to draw general conclusions on the impact of topology of MPTCP performance, some aspects can be stressed from the performed experiments. Transport performance is naturally affected by the characteristics of traversed paths, mostly their capacity (bandwidth) and latency. In the case of multipath transport, remote observations confirm empirically that MPTCP performance relates to transport-disjointivity of involved paths, and can leverage the presence of several (transport-disjoint) paths even if these are not topologically disjoint, as in the case of MA2 (see section 3.1.1). Experiments from both probes also show that benefits from multiple transport-disjoint paths can be severely damaged, and even completely counterbalanced, by an excessive difference between bandwidth capacities of different paths – which corresponds in the two-subflow scenario of UCL testbed to a high value of the BWR parameter. Observations in the two-subflow scenario also suggest that relative latency (*i.e.*, difference between the two smaller latencies of the available paths) has a direct effect on MPTCP performance (see section 4.2.1), as it determines the minimum duration of a transmission to allow MPTCP to leverage the existence of a second path, below which MPTCP behaves *at most* as well as single-path TCP.

5. Conclusion

We have proposed, implemented and evaluated a methodology for estimating the user benefits of using MPTCP instead of TCP, concentrating on bandwidth utilization and file transmission delay. To our knowledge, this paper is the first paper that explores the performance of MPTCP in a multi-homed wired/wireless scenario (WiFi/Ethernet), with experiments on a real networking testbed, both with a local, relatively controlled setup and with a remote setup. Wired/wireless host multi-homing is a relevant scenario (*e.g.* with laptops) which could become more important as the Internet of Things paradigm becomes more widespread. Our experiments indicate that MPTCP benefits with respect to TCP can be significantly affected by the difference between the device interfaces capacities: bandwidth utilization improves as interfaces have more similar capacity. Results also show that MPTCP benefits strongly depend on the relationship between interfaces' capacities (BWR): goodput improves when interfaces have similar throughput (BWR ~ 1), but the use of MPTCP may reveal counterproductive when capacities of interfaces have different order of magnitude (BWR $\gtrsim 10$). In terms of delay, our measurements show that MPTCP typically achieves, in the worst case, a similar delay to TCP over the fastest interface; the probability of outperforming TCP increases as the connection is larger or longer. From its very principle of operation, very short transmissions are unable to leverage path diversity if they terminate before MPTCP has established a second subflow. Internet measurements provide additional support to these trends.

Future work includes the generalization of this methodology, both in terms of geographical diversity (measurements from more probes would allow to understand the behavior of MPTCP in a more representative set of multipaths) and analytical depth (additional aspects could be addressed, such as the emulation of streaming or other applications potentially running on top of MPTCP).

Acknowledgements

This work has been mainly supported by the EC's 7$^{\text{th}}$ Framework Programme Grant No. 317647 (Leone), and by the Belgian IAP BESTCOM program. It was supported in part by Hong Kong RGC (GRF PolyU-521312), Hong Kong Polytechnic University (4-BCB6, G-YBJU), and National Natural Science Foundation of China (No. 61272463). I am thankful to Pr. Olivier Bonaventure for his useful comments, and to Saba Ahsan, from Aalto University, for her support in the setup of remote experiments.

References

[1] S. Deng, R. Netravali, A. Sivaraman, H. Balakrishnan: WiFi, LTE or Both? Measuring Multi-Homed Wireless Internet Performance. Proc. IMC'14. Vancouver, Canada. Nov. 2014.

[2] R. Khalili, N. Gast, M. Popovic, J.-Y. Le Boudec: Opportunistic Link-Increases Congestion Control Algorithm for MPTCP. draft-khalili-mptcp-congestion-control-05, IETF I-D, July 2014. *(Work in progress)*

[3] R. Khalili, N. Gast, M. Popovic, J.-Y. Le Boudec: Performance Issues with MPTCP. draft-khalili-mptcp-performance-issues-06, IETF I-D, July 2014.

[4] B. Hesmans, O. Bonaventure: Tracing Multipath TCP Connections. Proc. ACM SIGCOMM'14, Chicago, USA, 2014.

[5] C. Paasch, O. Bonaventure: Multipath TCP. *Communications of the ACM*, Vol. 57, No. 4, pp. 51-57. April 2014.

[6] Y.-C. Chen, D. Townsley: On Bufferbloat and Delay Analysis of Multipath TCP in Wireless Networks. Proc. IFIP Networking 2014. Trondheim, Norway, June 2014.

[7] Y.-C. Chen, Y.-S. Lim, R. J. Gibbens, E. M. Nahum, R. Khalili, D. Townsley: A Measurement-based Study of MultiPath TCP Performance over Wireless Networks. Proc. IMC'13, Barcelona, Spain, Oct. 2013.

[8] C. Paasch, R. Khalili, O. Bonaventure: On the Benefits of Applying Experimental Design to Improve Multipath TCP. Proc. CoNEXT'13, Santa Bárbara, USA, Dec. 2013.

[9] A. Ford, C. Raiciu, M. Handley, O. Bonaventure: TCP Extensions for Multipath Operation with Multiple Addresses. RFC 6824, IETF, January 2013.

[10] R. Khalili, N. Gast, M. Popovic, J.Y. Le Boudec: MPTCP is not Pareto-Optimal: Performance Issues and a Possible Solution. Proc. CoNEXT'12, Nice, France, Dec. 2012.

[11] J. Sommers, P. Barford: Cell vs. WiFi: On the Performance of Metro Area Mobile Connections. Proc. IMC'2012, Boston, MA, USA, November 2012.

[12] C. Paasch, G. Detal, F. Duchêne, C. Raiciu, O. Bonaventure: Exploring Mobile/WiFi Handover with Multipath TCP. Proc. CellNet'12, Helsinki, Finland, Aug. 2012.

[13] Y. Zhang, A. Arvidsson: Understanding the Characteristics of Cellular Data Traffic. Proc. CellNet'12, Helsinki, Finland, Aug. 2012.

[14] C. Raiciu, C. Paasch, S. Barré, A. Ford, M. Honda, F. Duchêne, O. Bonaventure, M. Handley: How Hard Can It Be? Designing and Implementing a Deployable Multipath TCP. Proc. NSDI'2012. San José, CA, USA, April 2012.

[15] C. Raiciu, M. Handley, D. Wischik: Coupled Congestion Control for Multipath Transport Protocols. RFC 6356, IETF, October 2011.

[16] A. Ford, C. Raiciu, M. Handley, S. Barré, J. Iyengar: Architectural Guidelines for Multipath TCP Development. RFC 6182, IETF, March 2011.

[17] D. Kaspar: Multipath Aggregation of Heterogeneous Access Networks. PhD. 2011.

[18] C. Raiciu, S. Barré, C. Pluntke, A. Greenhalgh, D. Wischik, M. Handley: Improving datacenter performance and robustness with Multipath TCP. Proc. ACM SIGCOMM'11. New York, USA, 2011.

[19] C. Raiciu, D. Niculescu, M. Bagnulo, M. Handley: Opportunistic Mobility with Multipath TCP. Proc. ACM MobiArch 2011. 2011.

[20] R. Prasad, C. Dovrolis, M. Murray, kc Claffy: Bandwidth Estimation: Metrics, Measurement Techniques and Tools. *IEEE Network*, November/December 2003.

[21] M. Jain, C. Dovrolis: End-to-End Available Bandwidth: Measurement Methodology, Dynamics and Relation with TCP Throughput. *IEEE/ACM ToN*, Aug. 2003.

[22] S. Ostermann: TCPTRACE, available from http://www.tcptrace.org.

[23] T. Dreibholz, M. Becke, H. Adhari, E. P. Rathgeb: Evaluation of A New Multipath Congestion Control Scheme using the NetPerfMeter Tool-Chain. Proc. SoftCOM'2011. Hvar, Croatia, Sept. 2011.

[24] C. Paasch, S. Barré *et al.*: Multipath TCP in the Linux Kernel, available from http://www.multipath-tcp.org.

[25] NLANR/DAST: Iperf - The TCP/UDP Bandwidth Measurement Tool, available from https://github.com/esnet/iperf.

[26] J. A. Cordero: MPTCP BW/Delay Benefit Measurements, available from http://inl.info.ucl.ac.be/softwares/mptcp-bwdelay-benefit-measurements.

[27] J. A. Cordero: Measuring Multi-Path TCP Performance in Dual-Homed Devices. Technical Report, UCL. Available from http://hdl.handle.net/2078.1/156662.