Improving resiliency and throughput of transport networks with OpenFlow and Multipath TCP

Demonstration of results over the Géant OpenFlow testbed

Felicián Németh, Balázs Sonkoly, András Gulyás, Levente Csikor,
János Tapolcai, Péter Babarczi and Gábor Rétvári
Budapest University of Technology and Economics, TMIT, HSNLab
MTA-BME Future Internet Research Group 1117, Budapest Magyar tudósok krt. 2. Hungary
E-mail: {nemethf,sonkoly,gulyas,csikor,tapolcai,babarczi,retvari}@tmit.bme.hu

Motivation: Currently, each networking layer redundantly has its own recovery mechanism resulting in more expensive networking equipment and higher operational costs. Can we get rid of all these mechanisms below the transport layer and use multipath transport protocol to provide the required resiliency?

Background: As a recent study showed, more than half of the network failures causing service degradation to end-users remain totally invisible below the network layer. Clearly, faults classified as power outage, fiber cut, circuit/carrier problem, hardware problem, interface down are visible at these layers, however they are just 46% of the total faults. The rest of the faults such as routing problems, congestion, sluggish, and also the faults classified as unreachable, miscellaneous can be totally invisible though. As a consequence, each networking layer redundantly implements its own recovery mechanism.

Multipath TCP (MTCP) is an experimental technology with non-negligible future potential and remarkable recent past, currently on the standards track in IETF. MPTCP enables the simultaneous use of several network interfaces of a single host bundling them to a single TCP interface towards applications, while in fact spreading data across several subflows traversing different paths. This mechanism can be the basis of a novel approach to network resiliency.

In order to provide disjoint paths for MPTCP in a dynamic and flexible way, OpenFlow is a perfect candidate, and the Géant OpenFlow facility has provided a suitable environment for conducting proof-of-concept measurements validating the concept.

Objectives:

- Implement a control framework for resiliency measurements
    - running on a shared facility
    - supporting different recovery mechanisms and
    - transport protocols
- Understand MPTCP in this environment
    - reveal the basic characteristics based on defined metrics
    - identify current shortcomings and open research issues

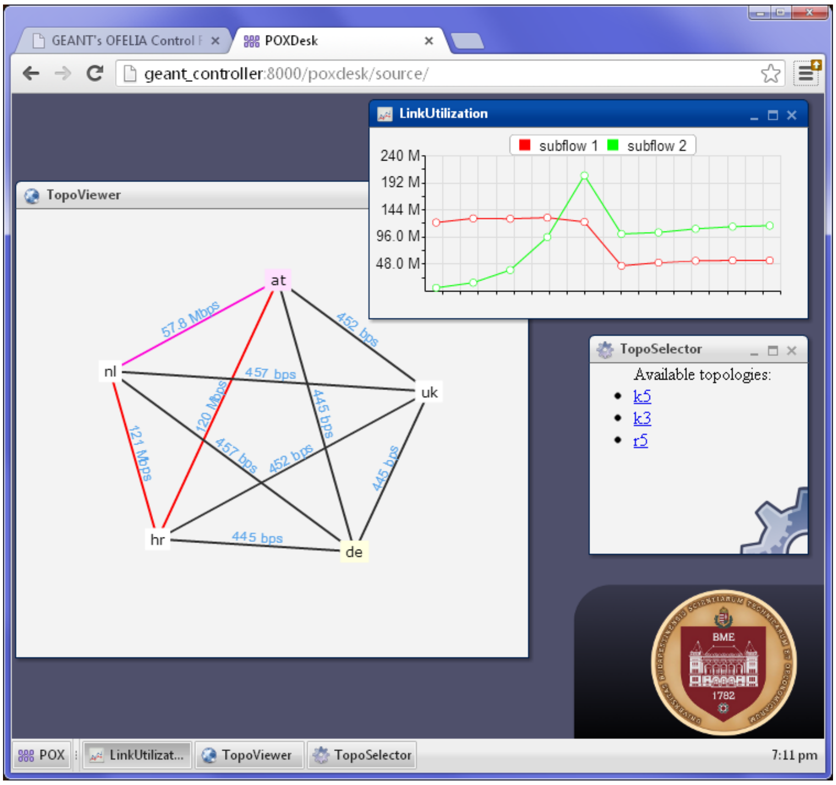
Control Framework:

Our extensible control framework provides automated orchestration for measuring the characteristics of different transport protocols under several error models and recovery strategies. It is able to emulate pre-scheduled link down events at different time scales (from 50ms to several seconds) in a shared OpenFlow testbed environment even when, e.g., FlowVisor filters out port configuration commands. The following restoration scenarios are considered and their impact on the end-to-end performance can be evaluated:

- *Optical recovery* – At the optical layer, a 50 ms failure restoration is the well-accepted time limit while the network has to respond to any optical failure and restore the disrupted connection before any upper layer protocol would detect and react to these failures.
- *IP recovery* – This mechanism contains three steps: (i) detecting the failure at the OpenFlow switch, (ii) setting up a new route flow entry from the controller (or using another pre-installed one) and (iii) forwarding the packets using the new flow entry.
- *Manual recovery/No recovery* – This scenario simulates the lack of an automated restoration technique in the network (e.g., infinite recovery time).

The framework is capable of running single flow measurements with pre-configured background traffic and more complex ones with configurable traffic mixes between different source-destination pairs. Several parameters of TCPs can also be tuned (buffer sizes, initial behavior, etc.) and during an experiment, individual and overall throughput performances and congestion windows can be shown on-the-fly.

Furthermore, the extended POXDesk component visualizes network topology with current link loads, plots per-flow throughput performance and provides network configuration options (see the screenshot below).



Selected Measurements/Results:

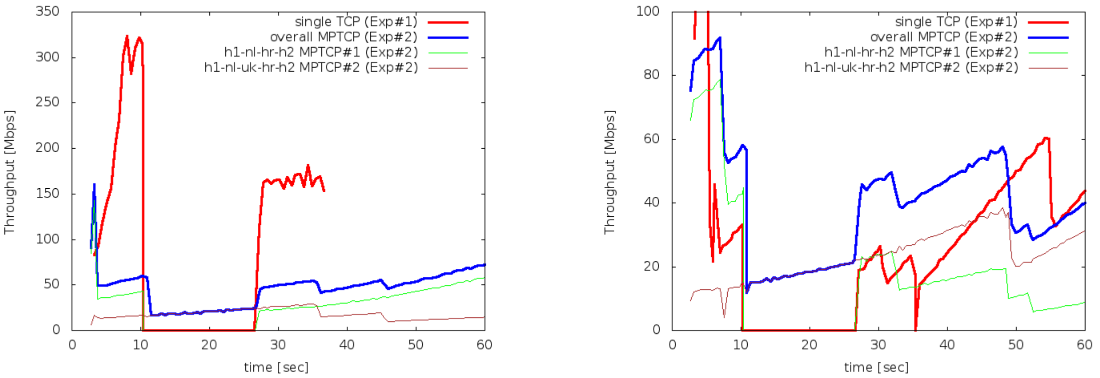


Fig.1 Comparing throughput performance of single TCP and MPTCP with link failure from 10s to 20s when there is no background traffic (left) and with 200Mbps UDP background (right). In case of large Bandwidth-Delay Product (BDP) (left), MPTCP is not responsive enough due to its conservative additive increase mechanism and shows longer flow completion times. When lower bottleneck capacity is available (right), MPTCP outperforms single TCP in the same environment. (Exp#1 and Exp#2 are different experiments in both cases.)

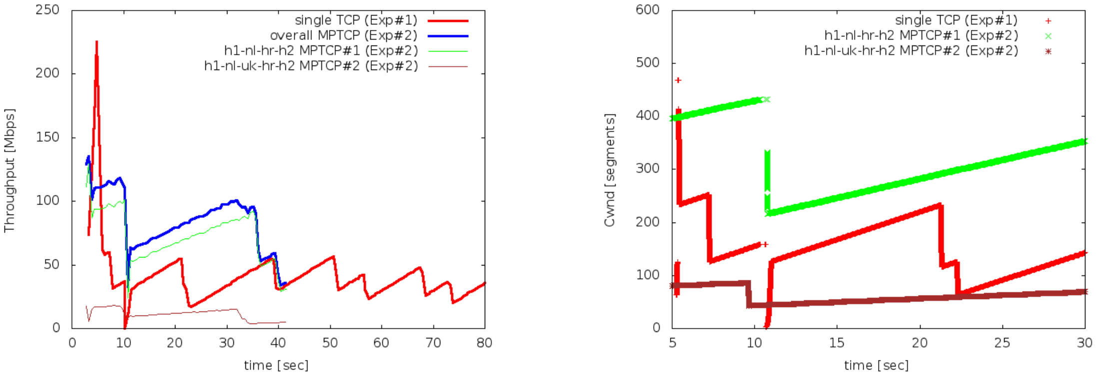


Fig.2 Comparing performance of a single TCP flow and MPTCP in the case of a 100ms link failure at 10s with reduced BDP. Throughput plots (left) indicate the much better performance of MPTCP while the congestion windows (right) help to explain the phenomenon around the link failure event. Here, MPTCP is able to recover from losses with fast recovery whereas single-flow TCP is forced back to slow start.

Powered by the Géant OpenFlow testbed GÉANT