Accurate Active Inter-domain Measurements

Otto J Wittner and Arne Øslebø
UNINETT AS

Extended Abstract[1]

The Internet is a significant part of today's communication infrastructure. It is commonly recognized that such an infrastructure should provide transport services with at least an availability in the order of five nines, i.e., the service should be available at least 99.999% of the time. With the objective to measure how close the availability of the global Internet is to this criteria, and to understand how such a level can be reached, an international lab, illustrated in Figure 1, has been setup. The lab enables accurate inter-domain measurements at low cost. This abstract presents the lab setup and results from accuracy measurements of the lab.

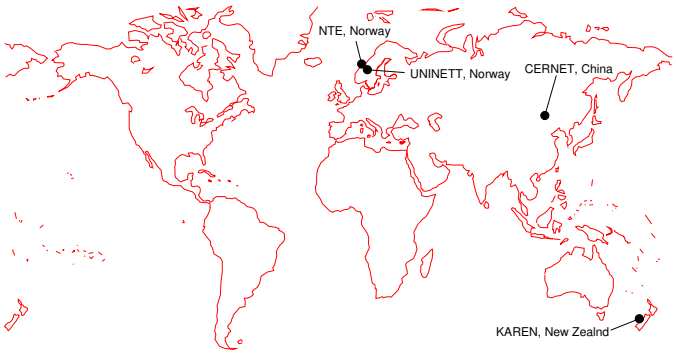


Figure 1: International measurement lab

To capture impairments down on a sub-second timescale as well as to ensure that detected failure events are related to relevant networks, the use of strategically placed and lightly loaded servers were found necessary. Hence currently available global lab-environments, e.g. [1], [2], are not (yet) suitable.

Table I presents hardware and software operative in our lab. Standard "off-the-shelf" hardware is applied together with an open source OS and rather well known open source software measurement tools. As indicated in Table I passive monitoring hardware is available in relation to two of our lab servers. Such cards provide time-stamp accuracy on a nanosecond timescale with GPS synchronized clocks. However they are (still) somewhat costly (i.e. not "off-the-shelf") and complicated to install due to GPS antenna requirements.

Real-time UDP Data Emitter (*rude*) and Collector for RUDE (*crude*) [3] are lightweight traffic generator and sink respectively. Constant bit-rate traffic at high rates may be generated. Sequence number, TX and RX timestamps are added to packets. Time-stamp accuracy is on a millisecond timescale. Measurement results confirming this accuracy are presented below.

Traceroute [4] utilizes ICMP packets and TTL values to actively probe all routers along a paths to a specific address. Hence an indication of the paths traversed by traffic between to servers is output. To map a path *traceroute* applies at least as many packets as there are hops in a path. There is no guarantee that all packets flow along the same path. To avoid being overloaded by ICMP packet processing routers may choose to ignore ICMP packets. Hence traceroute has medium/low accuracy, and may only be applied infrequently.

Tcpdump [5] dumps packets arriving on or leaving an interface. All headers and content of the packets can be made available in the dumps. Hence *tcpdump* supplements *curde* by making IP header information available for inspection (e.g. TTL).

Network time protocol daemons (*ntpd)* provide synchronization of system clocks by applying the NTP protocol [6]. Accuracy is reasonably good since clock adjustments are done in very small steps (<< 1ms) causing only low frequency noise in measurement data.

To verify the accuracy of the the timestamps added to the packets by *rude* and *crude* processes on the servers, a locale lab setup was arranged. A sending and receiving server with Intel Xenon 3.2Ghz and Dual Intel Xeon 514 2.33Ghz CPUs respectively where connected back-to-back through a DAG passive measurement card. Sending and receiving interface hardware were Broadcom NetXtreme BCM5721 and Broadcom NetXtreme II BCM5708 Gigabit ethernet cards respectively. Both server were running Linux kernel 2.6.26. On the sending server *rude* was configured to send 100 packets per second while receiving server applied *crude* to receive packets. *vmstat* was applied on both servers to monitor CPU load.

Figure 2 presents results form a 40 minute test-run. The results clearly show good performance of the sender process. *rude* transmits packets with an average deviation of only 0.00067ms from the configured 10ms packet interval. Maximum deviation was at 7.64ms and standard deviation from the average at 0.035ms. 99.988% of all packet had a deviation < 1ms, and only 28 > 1ms. Time-stamp accuracy is also acceptable with only a single gap above 1ms (1.2ms) when comparing *rude* (TX) timestamps and DAG generated timestamps. The period of worse performance between packet number 150000 and 180000 i Figure 2a is due to an attempt to stress the system by transferring a large file concurrently with the active measurements.

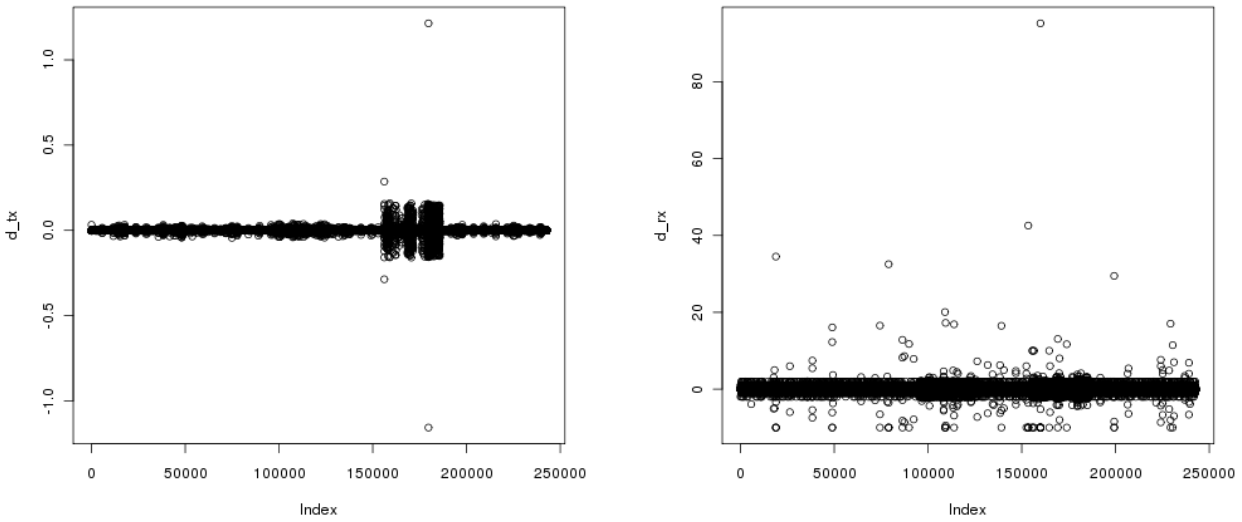
The receiving server running *crude* also performs acceptably, however somewhat less accurate than the sender. *crude*

UNINETT AS, the operator of the Norwegian educational and research IP network, http://uninett.no.

[1]This extended abstract was originally submitted to TNC2011. It is however now published as an technical abstract only.

Table I: Measurement server configurations

	UNINETT, Norway	*NTE, Norway*	*CESNET, China*	*KAREN, New Zealand*
Network domain	NREN	Commercial	NREN	NREN
CPU	Intel Xenon 4x4 core 2Ghz	Intel Pentium 2x1 core 3 Ghz	AMD Opteron 2 core 1.8Ghz	Intel Pentium 2 core 3Ghz
Memory	3 GB	2GB	2GB	1GB
OS	Linux 2.6.26	Linux 2.6.26	Linux 2.6.18	Linux 2.4.21
Network interface	Intel 82572EI Gigabit Ethernet	3Com 3c985 Gigabit Ethernet	Broadcom BCM5704 Gigabit Ethernet	Intel 82546GB Gigabit Ethernet*
Passive measurement card	Napatech NT20E	-	-	Endace DAG
Traffic source/sink	rude/crude v. 0.7	rude/crude v. 0.7	rude/crude v. 0.7	rude/crude v. 0.7
Path trace	traceroute v. 2.0.11	traceroute v. 2.0.12	traceroute v. 1.4a12	traceroute 1.4a12
Traffic dump	tcpdump v. 3.9.8	tcpdump v. 4.0	tcpdump v. 3.9.5	tcpdump 3.7.2
Clock synch	ntpd v. 4.2.4p4	ntpd v. 4.2.4p6	ntpd v. 4.0.98d	ntpd 4.1.1b



(a) *Index* is packet number and *d_tx* is deviation of *rude* (TX) time-stamp from configured value.

(b) *Index* is packet number and *d_rx* is deviation of *crude* (RX) timestamps from *rude* (TX) timestamps.

Figure 2: *rude/crude* accuracy measurements.

(RX) timestamps deviates on average 0.0468ms from *rude* TX timestamps, with maximum deviation at 95.19ms and standard deviation from the average at 0.38ms. Compared to DAG card timestamps, there are 101 gaps > 11ms and 174 gaps < 9ms during the measurement period. Some outliers exist, 1859 gaps > 1 ms, with maximum gap of 102.3ms.

During the tests the server CPUs were loaded heavily by applying the *CPUburn* [7] tool, however practically no correlation where found between CPU load and sender/receiver process behavior.

Further investigations and attempts in adjusting system parameters concluded in that the network interface cards including their OS drivers with high probability are responsible for most of the inaccuracies found during the local test setup.

The international lab has already collected valuable measurements. Two papers have been publish [8], [9] and more are in the pipeline, all analyzing the collected data. Future work on the lab setup include adding more measurement servers as well as ensuring each server is configured as close to the optimal setup with respect to accuracy as possible, e.g. by ensuring desirable network cards are being applied and/or by making passive measurement cards available.

REFERENCES

[1] L. Peterson, A. Bavier, M. Fiuczynski, and S. Muir, "Experiences building planetlab," in *Proceedings of the Seventh Symposium on Operating System Design and Implementation (OSDI)*, (Seattle, WA, USA), November 2006.

[2] C. Dovrolis, K. Gummadi, A. Kuzmanovic, and S. D. Meinrath, "Measurement lab: Overview and an invitation to the research community." CCR online, June 2006.

[3] J. Laine, S. Saaristo, and R. Prior, "Rude/crude real-time udp data emitter/collector." http://rude.sourceforge.net/, 2010.

[4] D. Butskoy, "traceroute for linux." Avaiable via http://traceroute.sourceforge.net, 2010.

[5] V. Jacobson, C. Leres, and S. McCanne, "tcpdump - dump traffic on a network." Available via http://www.tcpdump.org, 2010.

[6] D. Mills, U. Delaware, J. Martin, J. Burbank, and W. Kasch, "RFC5906 - network time protocol version 4: Protocol and algorithms specification." IETF, June 2010.

[7] R. Redelmeier, "CPUburn - CPU testing utility." Available at http://pages.sbcglobal.net/redelm/, 2010.

[8] E. Myakotnykh, B. Helvik, O. J. Wittner, O. Kvittem, J. K. Hellan, T. Skjesol, and A. Øslebø, "An empirical analysis of dependability characteristics on a global route," in *Proceedings of IEEE International Workshop on Quality of Services (IWQoS 2010)*, (Beijing, China), IEEE, June 2010.

[9] E. Myakotnykh, B. E. Helvik, O. J. Wittner, O. Kvittem, J. K. Hellan, T. Skjesol, and A. ÃŸslebÃÿ, "Analyzing causes of failures in the global research network using active measurements," in *Proceedings of the Second IEEE International Workshop on Reliable Network Design and Modeling (RNDM 2010)*, (Moscow, Russia), IEEE Com Soc, Octber 2010.