

Active Goodput Measurements from a Public 3G/UMTS Network

Kostas Pentikousis, Marko Palola, Marko Jurvensuu, and Pekka Perälä

Abstract—Although third generation cellular networks are being deployed in many countries, rigorous measurement studies from public networks have yet to be published. After extensive experimentation, we collect measurements from a commercial WCDMA 3G/UMTS network and observe that the goodput of the first of a series of back-to-back transfers is consistently below par. We compare these results with those obtained from an all-wired network and show that 3G signaling plays an important, albeit detrimental role in end-user goodput. Finally, we illustrate that near-nominal goodput is realizable for large payloads only.

Index Terms—Wireless networks, WCDMA 3G/UMTS, protocols for mobile networks, network traffic measurement.

I. INTRODUCTION

THIRD generation wide-area mobile networks, such as the WCDMA universal mobile telecommunications system (3G/UMTS), are expected to “deliver wideband information as well as voice, data, and multimedia” [1]. Besides delivering higher peak data rates, for example 384 kb/s downlink, the Universal Terrestrial Radio Access (UTRA) transport control function plays a central role in WCDMA 3G/UMTS providing operators many degrees of freedom in network implementation. The network can be parameterized using both dynamic and semi-static factors, yielding a large number of data rate combinations and the ability to change transmission rates frequently (even once every 10 ms) [1]. This way, the network can seamlessly transport different kinds of traffic and support four classes of quality of service [2]. However, flexibility and parameterization comes at a cost: WCDMA 3G/UMTS is a complex system, which may not always perform adequately.

Backwards compatibility and coexistence with second generation mobile networks (2G) allows for a gradual deployment and introduction of new types of services, such as streaming media [3]. Although there are several studies for 2G (see [4]–[6] and the references therein), to the best of our knowledge, measurements from *public* 3G networks from the user perspective have not been published so far. We fill this void and identify the strengths and weaknesses of an actually deployed 3G network using our mobile service testing software (MOSET), designed for traffic measurement, application-level, performance evaluation studies. MOSET can quantify content retrieval goodput through the entire service

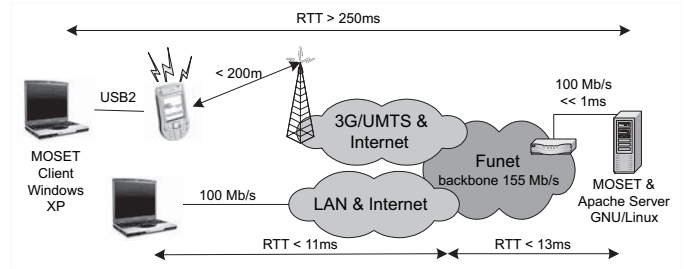


Fig. 1. The experimental topology used in our measurement study.

chain, from the content provider to the user equipment (UE), over different networks [7].

II. METHODOLOGY

Our experimental setup (Fig. 1) includes a laptop running the MOSET client on Windows XP and a GNU/Linux host running the Apache web server. A typical experiment has the client initiating n back-to-back downloads of the same object. We will refer to the object size in bytes as the *MOSET payload*. The client opens a new TCP [8] connection and issues an HTTP GET [9] request for each download. Once the MOSET payload is received, the TCP connection is torn down and a *new* TCP connection is initiated.

The next section presents results from two configurations. First, the laptop is connected via USB 2.0 to a Nokia 6630 mobile phone which, in turn, is connected to a public WCDMA 3G/UMTS network. The server is connected via Fast Ethernet and is located on a separate network. In the second setup, the laptop is connected via a Fast Ethernet LAN. As illustrated in Fig. 1, on average, the end-to-end round-trip time (RTT) is greater than 250 ms over the 3G network and less than 25 ms when the laptop is connected via the LAN. The two end-to-end paths share a common route within Funet, the Finnish University and Research Network, which contributes approximately 13 ms in each case. In all experiments, end-system load is negligible, and the TCP peers are configured to use a maximum segment size (MSS) of 1460 bytes [8] and selective acknowledgments (SACK) [10]; timestamps are not used. The receiver's (initial) advertised window ($rwnd$) is 32 KB, well above the estimated upper bound of the end-to-end bandwidth-delay product ($384 \text{ kb/s} \times 0.25 \text{ s} \approx 11.7 \text{ KB}$).

Unfortunately, the 3G network operator did not disclose any details about the radio access network (RAN), the core network configuration, or the load of the mobile network. We dealt with this information shortage by making extensive measurements with both setups. In addition to the data gathered

Manuscript received March 7, 2005. The associate editor coordinating the review of this letter and approving it for publication was Prof. Samuel Pierre. This work was carried out during the tenure of an ERCIM Fellowship, and was supported in part by TEKES, the National Technology Agency of Finland, under the EUREKA/ITEA Easy Wireless project (ITEA 03008).

The authors are with VTT, Technical Research Centre of Finland, Oulu, Finland (e-mail: pentikousis@ieee.org).

Digital Object Identifier 10.1109/LCOMM.2005.09021.

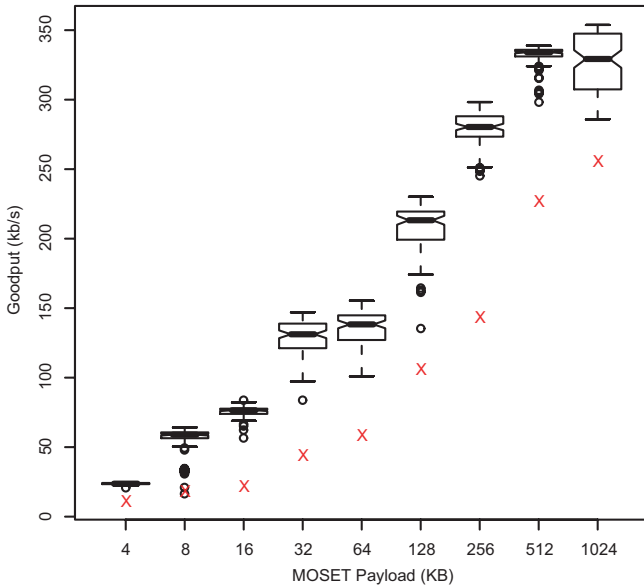


Fig. 2. MOSET measurements over a 3G/UMTS network reveal that the first connection goodput (“X”) is significantly below par.

by the MOSET measurement server, we collected traffic traces using *Ethereal* while each MOSET experiment is run, and employ *traceroute* and *ping* between successive sets of tests to capture any singularities in the end-to-end path.

III. RESULTS

We conducted measurements during different days and periods within the same day using MOSET payloads ranging from 64 B to 1 MB. Due to space limitations, we present results from measurements on February 18 and 22, 2005 only. Our metric of choice is end-user goodput, $g = D/T$, where D is the MOSET payload, and T is the total connection time. We are interested in the end-user perspective and, thus, take T to start from the moment the GET request is issued at the client and end when D is completely received by the client.

Fig. 2 presents a composite box-and-whisker and scatter plot of the measured goodput over the 3G network for nine values of D ranging from 4 KB to 1 MB, $n = 101$. The scattered “X” points correspond to the first connection goodput (g_1). The box plots illustrate the goodput distribution of the 100 subsequent back-to-back transfers ($\{g_i\}, i = 2 \dots n$). The box contains the middle 50% of the data, the line in the middle represents the median, \tilde{g} , and the “hinges” the $Q1$ and $Q3$ quartiles [11]. The width of the notch in the middle corresponds to the 95% confidence interval (CI) of the distribution median. Values outside the whiskers, shown as circles, are considered outliers.

Overall, goodput increases with D . This is attributed mainly to TCP: as the protocol goes through its connection establishment and slow start phases, probing the network for available capacity, the sender cannot fully utilize the allocated bandwidth. The large RTT prolongs the time spent in slow start, putting a considerable toll on goodput. During this phase, any dedicated WCDMA 3G/UMTS channel allocated to the transfer is, to a large extent, underutilized. As explained

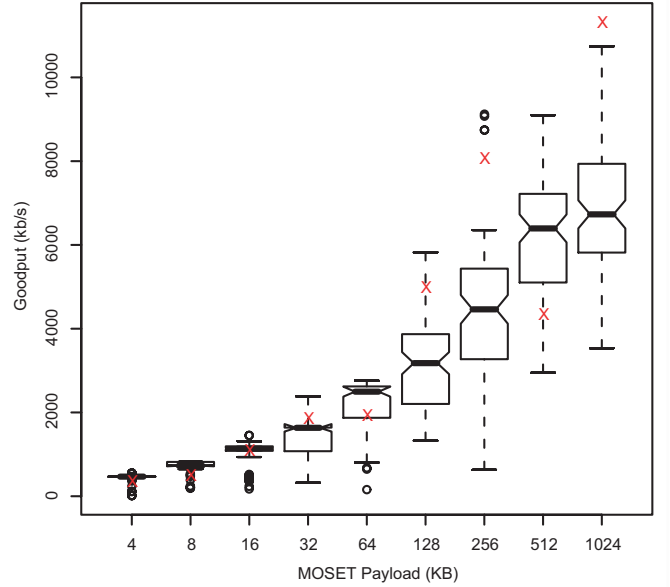


Fig. 3. MOSET measurements over a LAN show that goodput-wise the first connection (“X”) is equipotential with any of the subsequent downloads.

in [12], short flows suffer more than large ones. For D large, the TCP connection establishment and slow start costs are amortized over time and goodput can exceed 300 kb/s. However, the figure hints that this is unlikely for $D \leq 256$ KB. Note also that the box plot notches indicate that, at the 95% CI, the median goodput plateaus for $D = 512$ KB and 1 MB.

More remarkably, perhaps, Fig. 2 accentuates that

$$g_1 < \tilde{g}$$

In fact, g_1 is the worst of the entire lot. If we consider the mean goodput, \bar{g} , for $\{g_i\}, i = 2 \dots n$, we find that g_1 ranges from 29% ($D = 16$ KB) to 78% ($D = 1$ MB) of this average. Measurements from other days (not reported here due to space limitations) confirm that g_1 is always worse than both \bar{g} and \tilde{g} . Sporadically, one of the subsequent transfers progressed slower than the first one. Traffic traces showed that these exceptions were due to network congestion during the experiment. We could not resolve whether the segment losses occurred in the wired or wireless part of the network.

The underperformance of the first transfer is not observed in the second set of tests that involve a wired LAN (Fig. 3). Not surprisingly, goodput in these tests is at least an order of magnitude larger than in the previous set of tests and, as before, increases with the MOSET payload. However, g_1 is never the worst. In fact, in several cases $g_1 \geq \bar{g}$, and in the case of $D = 1$ MB, $g_1 > \max_i g_i, i = 2 \dots n$.

One might assume that domain name system (DNS) lookups could explain the poor first connection goodput [6]. However, in our experimental setup, the MOSET server has a fixed IP address, and the results of previous DNS lookups are cached locally. The *Ethereal* traces revealed no DNS lookups during the course of our experiments. As such, our experiments highlight a finding which, to the best of our knowledge, has not been reported in the literature so far: the first data transfer over a 3G/UMTS network has a notable handicap, not related with or explained by application or session layer shortcomings.

IV. DISCUSSION

From the end user perspective, the first connection goodput is the most relevant. Users typically download an object only once and, therefore, are bound to experience a goodput closer to g_1 than \tilde{g} or \bar{g} . Poor goodput in wireless networks is often attributed to TCP, due to well-known issues [12], but this is not manifested in our experiments. Network congestion or false reaction to corruption-based losses do not explain our finding. As a matter of fact, one could argue that WCDMA 3G/UMTS is quite successful in shielding TCP from packet loss. Furthermore, we point out that trace analysis showed that no SYNs were retransmitted in any of the experiments reported here. This means that all TCP connection establishments were always successfully completed in less than 3 s, the default initial TCP retransmission timeout (RTO). Earlier work by Ludwig et al. [4] showed that TCP performs “mostly ideal” over 2G networks. Our results show that stock TCP performs quite well on WCDMA 3G/UMTS too, in particular for large D . Nevertheless, end users should not expect to regularly enjoy goodputs close to the advertised data rates.

The reason for the below-par g_1 figures lies in the way UTRA allocates resources for packet-switched transfers. 3G/UMTS UE/hosts are considered “always on” but before they can transmit and receive data, they have to activate a Packet Data Protocol (PDP) context. Even if a PDP context is active, the UE still has to acquire access to a transport channel before it can, say, issue the GET request in our experiments. For reliable transfers, such as those handled by TCP, the UE will preferably request a dedicated channel (DCH) [1], [2]. This is the differentiating factor in our 3G measurements: the first transfer has to undergo the process of acquiring the channel while the subsequent ones do not have to do so.

More formally, when calculating $g_1 = D/T_1$, T_1 includes the time needed to acquire a transport channel and, possibly, the time to activate a PDP context. In contrast, $\tilde{g} = D/\tilde{T}$, where \tilde{T} is the median of $\{T_i\}$, $i = 2 \dots n$, is calculated with $\tilde{T} < T_1$, because \tilde{T} does not include a component corresponding to the time required for channel acquisition. For transfers over the LAN, T_i , $i = 1 \dots n$ does not have such a component either. Of course, in both sets of experiments, and for all n transfers, T_i includes the TCP connection establishment, slow start and, for D large, congestion avoidance phases, but not the connection teardown.

After allocating the bearer channel, UTRA allows both the UE and the network to cancel the PDP context and/or reallocate resources. A 3G/UMTS network is expected to reallocate resources if they are left unused. Once the resources have been reallocated, the UE will need to request them once more before any further transmission can commence. Taking the typical web usage pattern into consideration—“download a page, read it, then continue with a new download,” and so on—it seems that the goodput level that most users enjoy is close to g_1 .

V. CONCLUSION

Quantifying goodput and variation in performance is essential in understanding the real value of 3G services. After

extensive experimentation and measurements with a public WCDMA 3G/UMTS network, we can conclude that stock TCP performs quite well, especially when transferring large payloads. Our results indicate that the first transfer is particularly affected by the 3G/UMTS signaling required in establishing a channel before data transfers can commence. Furthermore, the “first connection goodput phenomenon” is likely to be observed in any transfer that follows an inactivity period exceeding the channel release threshold, typically in the order of seconds. In many scenarios, end users will experience goodputs closer to g_1 than to the significantly higher goodput \tilde{g} or \bar{g} of the n back-to-back subsequent transfers.

Our results are of great interest to researchers evaluating 3G/UMTS performance, especially those using off-the-shelf, generic network simulators only. Mobile application developers also need to pay attention to the effect of 3G signaling in end-to-end performance and keep in mind that high goodput rates are only achieved for large payloads. We are currently studying the effect of 3G signaling more closely, quantifying the unutilized network capacity and investigating ways to alleviate it. Along with more detailed performance analysis, we are also interested in scenarios which, for example, include long-lasting TCP connections with periods of inactivity as well as scenarios where several connections originating from the same mobile host compete for network capacity.

ACKNOWLEDGMENT

We thank Jarmo Prokkola for his comments & suggestions.

REFERENCES

- [1] R. Prasad, W. Mohr, and W. Konhäuser, Eds., *Third Generation Mobile Communication Systems*. Boston, MA: Artech House, 2000.
- [2] R. Lloyd-Evans, *QoS in integrated 3G networks*. Norwood, MA: Artech House, 2002.
- [3] A. J. Huer and J. F. Huber, *UMTS and Mobile Computing*. Boston, MA: Artech House, 2002.
- [4] R. Ludwig, B. Rathonyi, A. Konrad, K. Oden, and A. Joseph, “Multi-layer tracing of TCP over a reliable wireless link,” *SIGMETRICS Perform. Eval. Rev.*, vol. 27, pp. 144–154, June 1999.
- [5] P. Benko, G. Maliesko, and A. Veres, “A large-scale, passive analysis of end-to-end TCP performance over GPRS,” in *Proc. IEEE INFOCOM 2004*, vol. 23, Mar. 2004, pp. 1883–1893.
- [6] R. Chakravorty, J. Chesterfield, P. Rodriguez, and S. Banerjee, “Measurement approaches to evaluate performance optimizations for wide-area wireless networks,” *Lecture Notes in Computer Science (Proc. of PAM 2004)*, vol. 3015, Apr. 2004, pp. 257–266.
- [7] M. Palola, M. Juvansuu, and J. Korva, “Breaking down the mobile service response time,” in *Proc. IEEE ICON 2004*, vol. 1, Nov. 2004, pp. 31–34.
- [8] J. Postel, “Transmission Control Protocol - DARPA Internet Program Protocol Specification,” STD 7, RFC 793, Sept. 1981.
- [9] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, and T. Berners-Lee, “Hypertext Transfer Protocol - HTTP/1.1,” RFC 2616, June 1999.
- [10] M. Mathis, J. Mahdavi, S. Floyd, and A. Romanow, “TCP Selective Acknowledgement Options,” RFC 1818, Apr. 1996.
- [11] R Development Core Team, *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2005.
- [12] K. Pentikousis, “TCP in wired-cum-wireless environments,” *IEEE Commun. Surveys and Tutorials*, vol. 3, pp. 2–14, Fourth Quarter 2000.