

---

This is an electronic reprint of the original article.  
This reprint may differ from the original in pagination and typographic detail.

Wang, Le; Manner, Jukka

## Energy-efficient mobile web in a bundle

*Published in:*  
Computer Networks

*DOI:*  
[10.1016/j.comnet.2013.08.006](https://doi.org/10.1016/j.comnet.2013.08.006)

Published: 01/01/2013

*Document Version*  
Publisher's PDF, also known as Version of record

*Published under the following license:*  
CC BY-NC-SA

*Please cite the original version:*  
Wang, L., & Manner, J. (2013). Energy-efficient mobile web in a bundle. *Computer Networks*, 57(17), 3581-3600. <https://doi.org/10.1016/j.comnet.2013.08.006>

---

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.



ELSEVIER

Contents lists available at ScienceDirect

# Computer Networks

journal homepage: [www.elsevier.com/locate/comnet](http://www.elsevier.com/locate/comnet)

## Energy-efficient mobile web in a bundle <sup>☆☆</sup>

Le Wang<sup>\*</sup>, Jukka Manner

Aalto University, P.O. Box 13000, FI-00076 Aalto, Finland

### ARTICLE INFO

#### Article history:

Received 2 February 2013  
 Received in revised form 25 July 2013  
 Accepted 2 August 2013  
 Available online 13 August 2013

#### Keywords:

Wireless networks  
 Power consumption  
 Web access  
 Optimisation  
 Bundling

### ABSTRACT

The mobile Internet that was a huge hype ten years ago is finally here. We have a wealth of mobile devices that allow us to enjoy and consume Internet content from any number of web sites and services. With faster processors and radio links, our use of the Internet and the traffic mobile users generate grows at a tremendous pace. In this development, we still have a huge challenge to tackle: energy efficiency. When ten years ago one would recharge his personal mobile device once a week, we now do that daily; Smart phones today are not able to carry enough energy to allow us several days of usage time. The focus on this paper is to present and analyse one solution to help us in our daily lives. We implement and study a scheme where web content, a page, is delivered as a whole to a mobile device, instead of sending each individual object of the page separately. Combined with RRC state based header compression and selective content compression, our proposal allows the radio to keep in low power state for longer durations, and as a consequence brings huge energy savings. Download times also decrease, thus bringing increased Quality of Experience.

© 2013 The Authors. Published by Elsevier B.V. Open access under [CC BY-NC-SA license](https://creativecommons.org/licenses/by-nc-sa/4.0/).

## 1. Introduction

With the advent of almost 6 billion mobile-cellular subscriptions at the end of 2011 [1], the inevitable trend of post-PC era is already clearly visible. In the last decades, rarely have technical innovations changed everyday life as fast and profoundly as the pervasive use of mobile communications. With the advance of wireless communication technology and the emergence of mobile devices, global mobile data traffic grew 2.3-fold in 2011, which was eight times the size of the entire global Internet traffic in 2000, and is expected to increase 18-fold by 2016 [2]. The ever-

increasing demand for mobile connectivity and wireless services has lead to increased energy consumption and short battery life on mobile devices. A major technological challenge for mobile devices is to store a large amount of energy in batteries for increasingly complex mobile devices and yet still deliver reasonable size and weight. So long as batteries continue to be based on electrochemical processes, limitations of power density and limited lifetime will be difficult to overcome, making it hard to cater for mobile devices with power-hungry features and adopt mobile services further [3,4].

The reduction in energy consumption of mobile networks is of great importance to continued adoption of mobile Internet services and sustainability of an acceptable Quality of Experience (QoE) for mobile Internet. This paper focuses on energy saving on mobile web access, which accounts for 10.11% of Internet web access worldwide in May 2012, a growth of 75.8% over last year [5]. The number is even higher in Asia and Africa, where PC penetration is lower on average. Clearly, mobile devices have being rapidly evolving as one of the primary choices of accessing web content. However, mobile web access faces several

<sup>☆☆</sup> This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike License, which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

<sup>\*</sup> Corresponding author. Tel.: +358 440327066; fax: +358 947022474.

E-mail address: [le.wang@aalto.fi](mailto:le.wang@aalto.fi) (L. Wang).

issues regarding energy efficiency. Firstly, the fixed overhead of transmission of radio interfaces is significant once they are in active mode [30]. Thus data should be sent in quick bursts (compared to constant small transfers) to enable a longer battery lifetime. Secondly, unlike wired networks, wireless networks take open air as the transmission medium, which is subject to interferences, fading over the wireless channel and mobility of end devices. As a result, application layer performance in terms of goodput has been significantly reduced due to increased delay and rate variability. Least but not last, the power consumption of transmitting one bit over the air is over 1000 times than that of 32-bit CPU computation [29]. Hypertext Transfer Protocol (HTTP), as a verbose protocol to carry web traffic, introduces additional bytes into its communication resulting in additional energy cost.

Therefore, we undertake a detailed exploration of the effects of power consumption on HTTP traffic, propose and evaluate a proxy-based architecture for energy-efficient web access. In particular, we make the following contributions.

- This work studies the characteristics of transmission in wireless networks, namely Wireless Local Area Network (WLAN) network, and Universal Mobile Telecommunications System (UMTS) network in terms of throughput and energy efficiency performance. The results present the energy consumption characteristics of browsing web content over the wireless networks, and also the potential of improving the power consumption efficiency.
- An energy-efficient proxy is introduced between mobile devices and web servers to improve energy efficiency of web access by optimising the web content delivery through the content bundling, header compression, and selective content compression.
- The energy-efficient proxy was implemented on a commercial smartphone and evaluated through experiments in both WLAN and 3G networks with thorough test cases. We concentrate on the improvement of the proxy-based solution compared to normal web access from several aspects, where we look into how data subscription plan, traffic delay, file content and Radio Resource Control (RRC) inactivity timers affect the power consumption. The results show significant energy consumption reduction of web access.

It is worth to mention that the solution provides energy saving opportunity not only for web browsing. HTML5 is a trending web technology that is being applied to a range of industries with a new method of content delivery. The emerging developments for HTML5 are changing the way to access content, and also boost the potential of energy savings by deploying our solution. For example, Pocket [27] as a mobile application with over 4.5 million users helps people to save an interesting article or web page that users do not have time to view it right away but for reading it later on their mobile devices. Our solution optimises the downloading of the content over wireless networks in an energy-efficient way for the type of services, which take HTTP traffic as a major carrier of interaction between clients and servers.

The remaining sections are organised as follows. We start with reviewing prior work on the field in Section 2 and provide sufficient background information for further discussion in Section 3. Then we present the architecture and design of our system in Section 4. Section 5 evaluates the performance of the energy-efficient proxy and Section 6 discusses related issues regarding the proxy and summarise in Section 7.

## 2. Related work

Energy consumption of network transmission on mobile devices has seen a large number of academic papers. Our work, in particular, focuses on applying a performance enhancing proxy to improve energy efficiency of web access relating to mainly three categories of work: research on analysing and improving TCP performance over wireless links, research on optimising RRC state in UMTS networks, and research on enhancing HTTP performance.

A lot of prior work has been conducted to enhance TCP performance with proxy over wireless links. Some studies focus on TCP throughput improvement and propose that TCP connections can be split using as pivot proxies connected to both wireless and wired links. End-to-end connections are partitioned into two portions so that the impact from the slow and lossy wireless link on the fixed network can be minimised. I-TCP [6] is the very first work in the area, M-TCP [7] uses a zero window ACK scheme to schedule TCP data to provide better preserved end-to-end transmission. Other similar studies can be found in [8–11]. As energy issues have presented a significant barrier to continued adoption of mobile Internet services, many studies have been carried on using a proxy to improve energy efficiency over transmission. Cool-Tether [12] was proposed to minimise energy consumption on mobile devices while they serve as Internet access gateways for other clients by employing gatherer, striper and a reverse-infrastructure mode for WiFi. Network Connectivity Proxy (NCP) [13] proposes a SOCKS-based proxy to preserve existing TCP connections and UDP data flows for a mobile device when it goes to sleep so that the device can enter a low power consumption state and its full network presence is still maintained.

The RRC inactivity timers have also been widely studied in several previous work. The energy inefficiency of the timers was analytically studied in [14,15] to determine the optimal values of the inactivity timers. Then the timer issues have been studied in [36,37,16] on real-measurement basis. Researchers also proposed solutions to deal with the issues. In [47], Tail Optimisation Protocol (TOP) dynamically decides timeouts for the inactivity timers and utilises fast dormancy to eliminate the tail energy. Another solution TailTheft [17] prefetches and delays data transfer to schedule a number of transmissions to the tail time (the period of high power state after the completion of a transmission) of the transmissions.

In comparison, our study utilises the principle of split TCP to optimise HTTP downloading over wireless links, and focuses on leveraging RLC buffer threshold to keep the mobile device in lower power consumption state.

Existing work has studied how to improve HTTP performance. WebExpress [41] was proposed to improve web-browsing performance over wireless links through caching, differencing, and header reduction. GPRSWeb [26] implements CHK-based caching, parse-and-push functions, delta-encoding and utilises its own customised protocol between a client and a proxy to improve web-browsing experience over GPRS links. In [18,19], HTTP is modified to deliver the embedded web objects as a single bundle to a web browser without being explicitly requested. In addition, Opera Mini [23] applied transcoding approaches to reduce the traffic load instead of the power consumption, which requires frequent communications between the proxy and user clients. Thus, it consumes a lot of power. In summary, the solutions were designed for fast page downloading and do not focus on energy-efficiency of web browsing.

The studies of energy-efficient web browsing have also been reported in prior work. Power Aware Web Proxy (PAWP) [20] is an architecture designed to schedule incoming web traffic into intervals of high and no communication so that WLAN interfaces can switch to a low power state after very short idle intervals. Another approach is reported in [43], which proposes an architecture called Virtual-Machine based Proxy (VMP) to shift computing from mobile devices to the proxy in 3G networks. The proxy transcodes dynamic web objects in order to save energy of mobile devices. In comparison of the proxy-based solutions for energy-efficient web browsing, our solution utilises bundling and header compression to cater to the energy consumption characteristics of WLAN and 3G networks. The selective compression applied is lossless compression, which does not alter original web content and still provides significant improvement of energy consumption along with other techniques. In addition, the solution does not require any modification on web browser and web servers, thus it can be deployed incrementally.

### 3. Transmission and energy issues over wireless networks

In the section, we start with TCP performance issues over wireless networks in Section 2.1. We then discuss energy consumption issues of mobile devices and demonstrate power saving mechanism of WLAN and 3G networks in Section 2.2. Furthermore, their effects on HTTP traffic are discussed in Section 2.3.

#### 3.1. TCP Issues over wireless networks

TCP performance issues manifest in various aspects as significant degradation of TCP throughput, under-utilisation of link capacity, excessive interruption of data transmissions, and energy-inefficiency of transmission. There are several reasons resulting in these problems:

*High and variable latency:* generally, wireless links suffer from severe error rates due to external interference, going out-of-range, or blocking of signal. WLAN networks provide relatively smaller coverage area and higher system bandwidth, where the observed transmission and propaga-

tion delay is relatively smaller as well compared to 3G networks. 3G networks exhibit spurious delay and additional jitter not only due to the lower bit rate or wider coverage range, but also because of packets buffering, interleaving, rate adaptation and retransmission in Radio Link Controller (RLC) layer, and allocation and de-allocation of network resources when RRC state transition happens. Based on more than half million of measurements gathered by NetTutka [35] from nation wide user base, we find that the average Round-Trip Time (RTT) is around 83.29 ms in WLAN networks and 150.76 ms in 3G networks (UMTS and HSPA/HSPA+). Fig. 1 shows the cumulative distribution function of latencies of sending packets from mobile clients to our server. 60% of the latencies are below 80 ms in WLAN and 150 ms in 3G networks. However, over 4% of the latencies can go beyond 500 ms in both networks. We also measured the jitter during TCP sessions. There

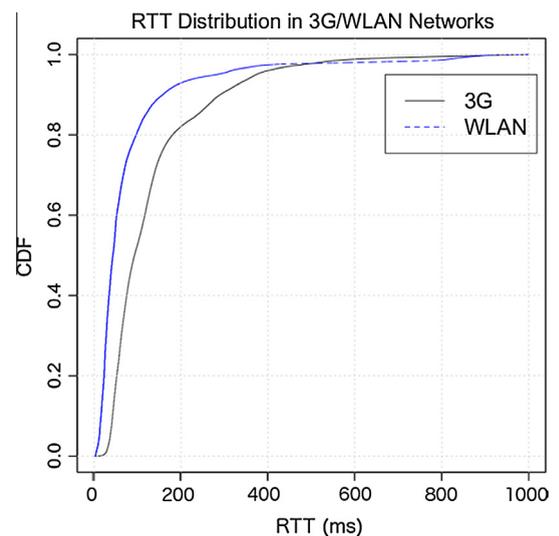


Fig. 1. CDF of network latencies.

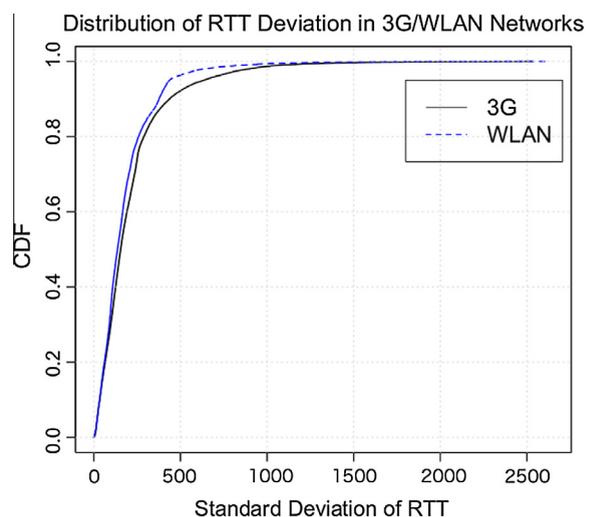


Fig. 2. CDF of network latencies.

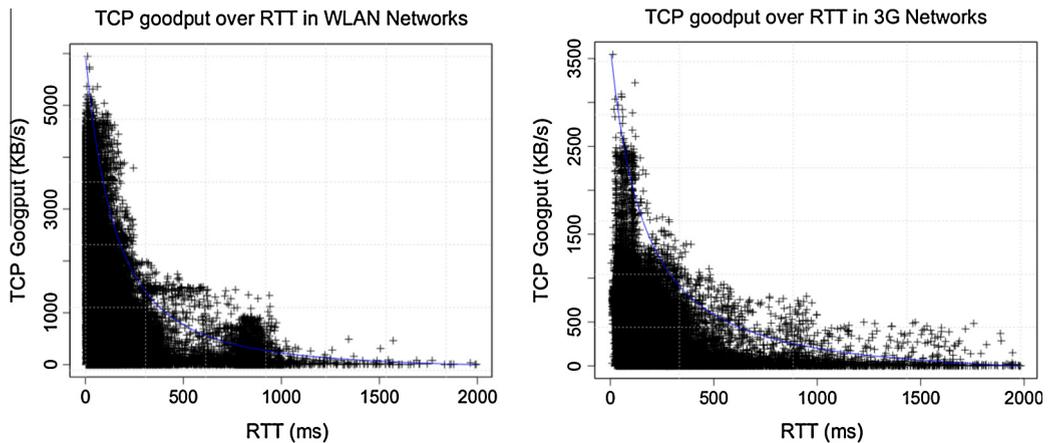


Fig. 3. TCP goodput over RTT.

were nearly 30 thousand TCP sessions measured, during each of which the value of RTT was recorded every 50 ms. During certain TCP sessions, the standard deviations are surprisingly high as seen in Fig. 2, where the cumulative distribution function of the sorted standard deviations of network latencies in ascending order is plotted.

*Fluctuation of TCP throughput:* With the large delay in wireless links, the connection stays longer in slow start state leading to longer TCP start process to fully utilise available link bandwidth. RTT inflation may also lead to high retransmission timer value, potential SYN timeout, higher recovery time, and mistaking wireless losses for congestion, which all impact TCP performance [25,26,28]. Fig. 3 demonstrates how RTT affects TCP goodput in WLAN and 3G networks based on the measurements collected from the Nettiutka database. The increase of RTT results in poor perceived TCP goodput in both networks. The blue solid line in each figure presents a borderline where 95% of measurement dots are within it.

### 3.2. Energy efficiency of transmission

WLAN and 3G radio interfaces of mobile devices present high association energy cost to the networks. We measured the consumed power of one Nokia N9 by using Nokia Energy Profiler Compared to 330 mW operating power of screen with maximum brightness on the Nokia N9, the power of WLAN radio interface is around 310 mW when it stays in active mode without sending/receiving any data, and the operating power is above 1380 mW when sending/receiving data at full-speed rates, which is around 12 Mbps on downlink and 5 Mbps on uplink. The operating power of 3G radio interface in active mode without sending/receiving is around 1085 mW, and the operating power increases to 1470 mW when data are transmitted through the interface.

The two radio technologies present radically different energy consumption characteristics due to the different network resource allocation/de-allocation mechanisms. Specifically, IEEE 802.11 standard defines that the 802.11 WLAN capable devices operate either in Continuously Active Mode (CAM) or Power Saving Mode (PSM). Compared

to CAM, the objective of the 802.11 PSM allows the mobile device to switch from the Active Mode to the Sleep Mode as soon as data transmission is completed. The mobile device synchronises with the infrastructure such as Access Point (AP) by receiving periodical beacon frames every 100 ms or 200 ms. The mobile device suspends its radio activity after the period of inactivity, and then wakes up for buffered frames from the access point. If there is no incoming traffic, the mobile device is able to switch into low power consumption mode right after the beacon interval. In the upper plot of Fig. 4, the power consumption and uplink/downlink rates of downloading a 2.5 MB file are shown. TCP handshaking triggered the N9 into active mode and the power consumption increased dramatically when the downloading started. The operating power then went down due to sleep mode in 200 ms after the end of the traffic.

Due to the scarceness of radio resources in 3G networks, RRC state machine is defined in the Third Generation Partnership Project (3GPP) [38]. It is more dedicated and complex including IDLE state, Cell Paging Channel (Cell\_PCH) state, Cell Forward Access Channel (Cell\_FACH) state and Cell Dedicated Channel (Cell\_DCH) state. Each state presents different requirements of radio resources and level of operating power. There is no radio resource allocated for the mobile device in the IDLE state, even through the device monitors the broadcast messages from the Radio Network Controller (RNC). In the Cell\_PCH state, the device monitors the paging control channel, and yet is still not able to have uplink activity. Packet Data Protocol (PDP) context is maintained so a session could be reconnected rapidly. In the Cell\_FACH state, the device is assigned a common or shared transport channel but not a dedicated channel. Thus, this state is suitable for small or medium amount of data [36,37]. In the Cell\_DCH state, RRC connection is fully established and the device is assigned dedicated transport channels both downlink and uplink. The dedicated resources make the channels more suitable for large traffic volume such as audio/video streaming, file transfer and Web traffic with large objects. Since the resources associated with these states are different, each state also presents different operating power. As seen in

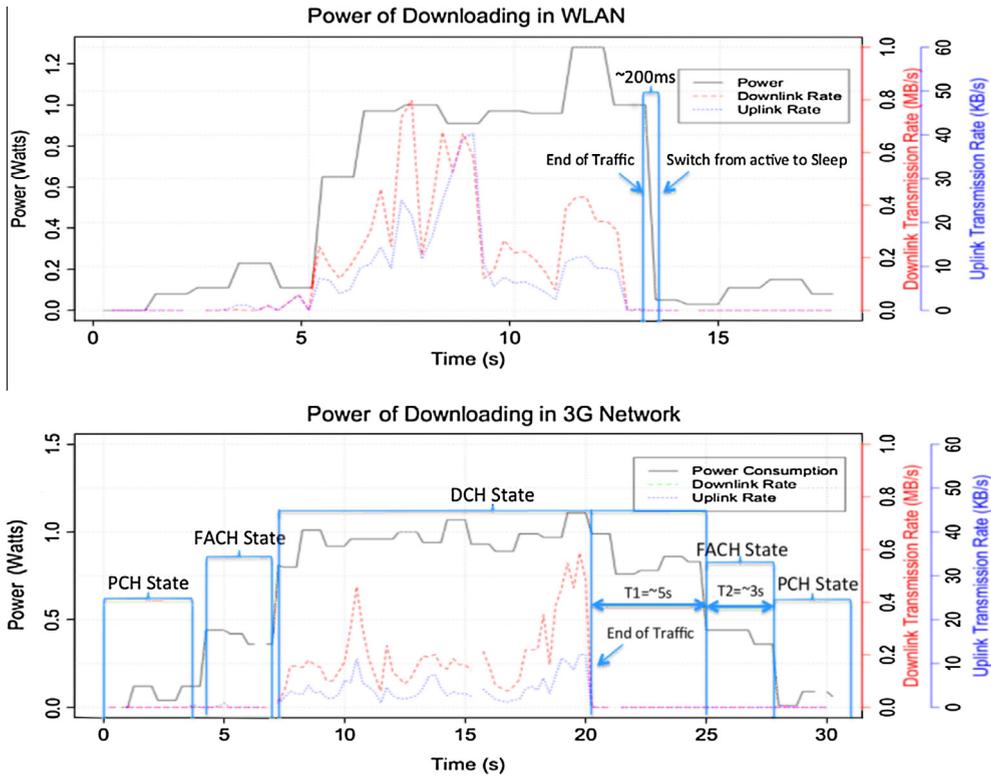


Fig. 4. Power consumption states in WLAN and 3G networks.

the lower plot of Fig. 4, the operating power of the Cell\_FACH state is around 489mW, roughly 50% of that in the Cell\_DCH state, and the Cell\_PCH state only consumes about 1–2% of the operating power of the Cell\_DCH state. After the end of traffic, the 3G link exhibited a residual energy cost. Compared to the operating power characteristic of WLAN, inactivity timers are introduced in UMTS as shown in Fig. 5. Timer T1, T2 and T3 control the durations of staying at the Cell\_DCH, the Cell\_FACH and the Cell\_PCH states, respectively. Mobile device is able to switch from the Cell\_DCH state to the Cell\_FACH state only after T1 times out, and from the Cell\_FACH state to the Cell\_PCH state after T2 times out. When T3 times out, the device falls back from the Cell\_PCH to the IDLE state.

In the RRC state machine, the states promote when switching from lower power consumption states to higher power consumption states, and the states demote when

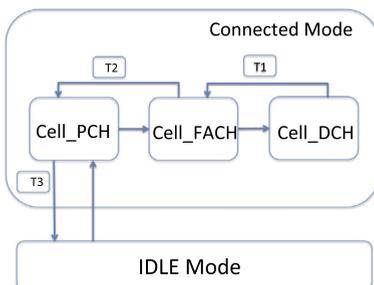


Fig. 5. RRC states transitions.

switching happens in the reverse direction. The state promotion is triggered either from the IDLE/Cell\_PCH state to the Cell\_FACH state once there is any transmission activity or from the Cell\_FACH state to the Cell\_DCH state when the data volume exceeds the RLC buffer threshold, which is controlled by network operators [36]. To explore the threshold value in our measurement environment, we sent different-sized ping packets from the N9 and measured the operating power. In Fig. 6, the average power of sending ping packets including the portion of energy tail, and its 95% confidence interval are shown. The power increases

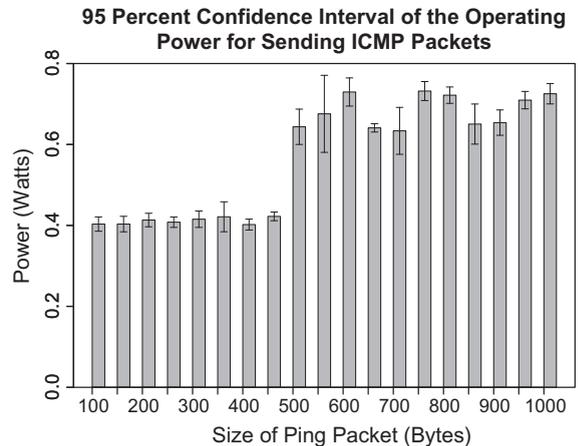


Fig. 6. Operating power of sending ping packets.

significantly from around 400 mW to 700 mW when the size is over 400 bytes. Because the energy tail is included in our calculation, these two values of power are smaller than the operating power of the Cell\_FACH and the Cell\_DCH states. More specifically, the threshold value is reached when IP payload is  $470 \pm 12$  bytes (8 bytes ICMP header + 462 bytes payload) in the network. Before state switching is done, signalling traffic is exchanged between the mobile device and the RNC for resources allocation and de-allocation [39]. Since the allocation is more costly, the promotion needs more control messages yielding a long latency that is around 1 s for the promotion from the IDLE/Cell\_PCH state to the Cell\_FACH state, and 0.5 s for the promotion from the Cell\_FACH state to the Cell\_DCH state. The state demotions are triggered by the inactivity timers, which are also determined by network operators in the RNC.

### 3.3. HTTP traffic over wireless networks

In addition to the limitations of wireless communications in terms of throughput and energy efficiency, the HTTP protocol also presents inefficiencies:

- As a verbose protocol, HTTP is coded in standard ASCII, which increases the number of bytes transmitted over the air.
- Due to the statelessness of HTTP protocol, HTTP relies on cookies or ETags to keep states consistent between web browser and server. According to RFC 2109 [40], the size of one cookie could be up to 4096 bytes and it is normally around 700–800 bytes per HTTP request [43].
- Simultaneous persistent connections in HTTP/1.1 [31] allows the TCP connections to keep open for the following requests without opening new connections. HTTP pipelining offers the ability to pipeline requests for different web objects in a single TCP connection. The technologies eliminate the overhead from a single client [26]. However, modern websites are integrated with the third-party content, such as web analytics tools, social media plugins and embedded advertisements. TCP connections have to be set up between the mobile

device and multi-domains. Thus, associated TCP connection overhead, e.g. SYN, ACK packets, is still high, and TCP three-way handshake delay can be significant while establishing a TCP due to the high latency of wireless links.

Fig. 7 simply depicts the energy consumption performance when downloading a web page under different level of delay. As the delay increases by 200 ms, the energy consumption of loading a webpage increases 60.14% and 33.12% in average in the WLAN and 3G networks respectively.

## 4. System architecture and design

Guided by the findings revealed and the causes of poor energy consumption performance of web access in the wireless environment in Section 2, we propose an energy-efficient proxy system to reduce the operating power of web access. We will first overview the architecture in Section 3.1. Then the design of the key components and its rationales are presented in Sections 3.2, 3.3 and 3.4 respectively.

### 4.1. Overview of components

The overall architecture of the energy-efficient proxy is shown in Fig. 8. The proxy is introduced between the mobile devices and web servers to split HTTP traffic into two portions, one of which is normal HTTP traffic between the proxy and web servers, another operates over our Energy-Efficient Proxy (EEP) protocol to save energy consumed for web content delivery through a number of enhancements. The solution is designed to be generic and transparent between the mobile devices and web servers, and also independent of web browsers. Thus, as part of installation, the web browser only needs to configure its HTTP proxy setting to route all HTTP requests to the device's localhost address, where the client side software is listening. Once any request is received, it is forwarded to the proxy in compact format. The system is designed and implemented to improve energy-efficient performance from the following aspects.

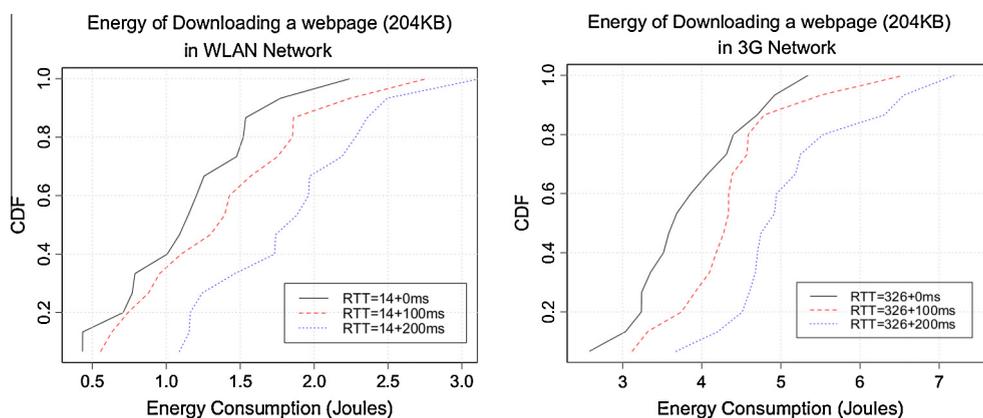


Fig. 7. CDF of energy consumption for downloading a webpage.

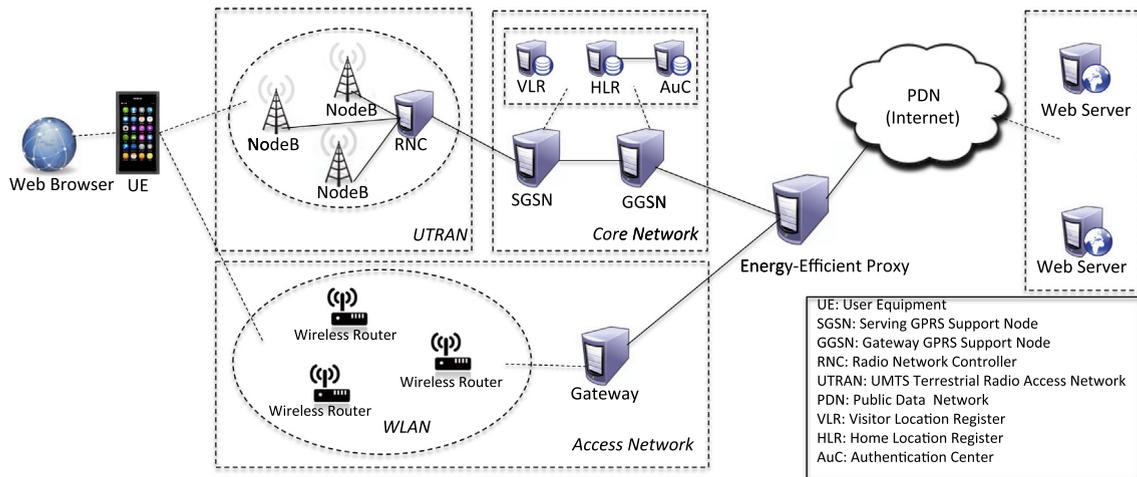


Fig. 8. Architecture of energy-efficient proxy.

- On one hand, we offload entire interactive web fetching from the mobile device to the web proxy. On the other hand, we allow the mobile device to enter low power consumption state till the proxy sends the bundled objects back instead of keeping the radio on until downloading is finished. The solution simplifies the HTTP message exchange procedure between the mobile device and the web servers. Once our running software on the mobile device receives HTTP requests, it encapsulates the requests into EEP packets and forwards them to the proxy. The proxy fetches all the web content and sends all of the web objects in one bundle to the mobile device on behalf of the web browser. Instead of using HTTP, the EEP protocol is applied between the mobile device and the energy-efficient proxy. Furthermore, we offload DNS resolution to the proxy in the wired network leading to faster lookup.
- HTTP headers are compressed not only to fasten transmission rate but also to keep mobile devices in lower power consumption state even when initial HTTP request is sent through 3G networks. Compared to WLAN, UMTS maintains an RRC state machine to manage radio resources for each mobile device. There are three states, each of which allocates different amounts of radio resources, and presents different level of operating power. Promotion from low power consumption state to high power consumption state is decided by the amount of data sent from mobile device to base station. Keeping the size of HTTP request under the promotion threshold is able to maintain the device in relative lower operating power even during the requesting procedure.
- Carefully selecting compression on HTTP payload can provide further energy saving when fulfilling certain conditions, which include considerations of link quality, computation load, file type and compression algorithms [32,33]. The proxy compresses the web objects selectively based on the compression ratio of compressing the objects and operating power of mobile devices required for decompressing during the web fetching.
- The solution firstly utilises a single TCP connection to effectively retrieve multiple web objects instead of multiple TCP connections. Secondly, the TCP connection is split into wireless and wired portions. The improvement of transmission results in lower connection overhead, better utilisation of the wireless network bandwidth, and higher robustness against link variances and agnostic to network heterogeneity.

Embodying the above motivated design principles, we designed and implemented the energy-efficient proxy. The whole system is mainly split into two parts, namely *Local Proxy* residing in the UE and acting as a web proxy for the browser to communicate with the outside, and *Remote Proxy* acting as a proxy between the UE and web servers as shown in Fig. 9, where the detailed protocol stack is illustrated. EEP protocol enables the local-remote communication, where compression algorithms and levels can be specified, the knowledge of transmission medium (3G or WLAN) can be shared, and the payload data can be interpreted by both sides (detailed in Section 4.3).

As seen in Fig. 10, once a new HTTP request is generated by the browser, it is then forwarded to the Local Proxy, where the URL of each HTTP is hashed as a 128-bit index using SHA-1. The hashed indexes are stored in the Local Proxy to map to the corresponding EEP reply, which consists of EEP header, the URL hash and compressed HTTP response. The Local Proxy compresses the request before encapsulating it as EEP payload and sending it over the air.

After being received by the Remote Proxy, the EEP requests are examined and different actions are taken depending on the request types. If the type is for web objects, the requests are then forwarded to a dedicated web engine after decompression. A webpage normally contains a number of web objects not only the HTML page. These eventually generate more than one HTTP request after parsing the HTML document. The web engine is able to build Document Object Model (DOM) tree based on the HTML document, and evaluate JavaScripts, which may generate new requests for web objects. Therefore, all the web

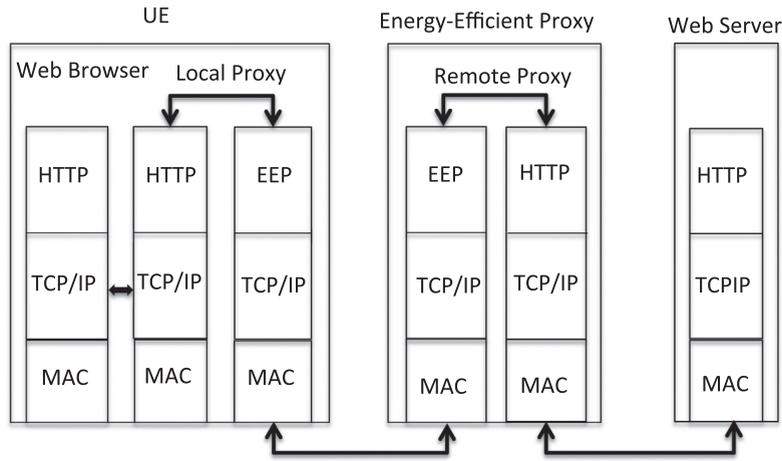


Fig. 9. Protocol stack.

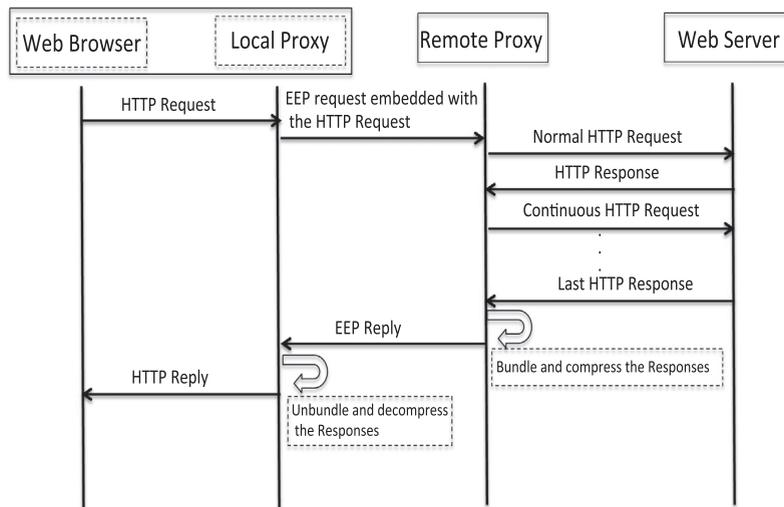


Fig. 10. Flow chart of message exchange between the web browser, local proxy, remote proxy and web server.

objects associated with the request can be fetched from web servers. When the HTTP response is received, the Remote Proxy compresses the response’s header, and the payload selectively (detailed in Section 4.4). Since HTTP is stateless, HTTP cookies and some other header fields are used to maintain certain consistency between the web browser and web servers. Thus, the HTTP response headers are also kept in EEP replies. After all the web objects are downloaded, the Remote Proxy sends them back in sequence as a single bundle to the Local Proxy. Once the Local Proxy receives the EEP reply, it unbundles and decompresses the reply so that the original HTTP responses including headers and payloads are reconstructed. There might be the case that HTTP responses requested by the web browser are missing from the reply due to the differences between the web browser in the UE and web engine in the Remote Proxy, or the requests generated by web engine related JavaScript codes. The web browser

can request for the missing content until the whole webpage is loaded.

#### 4.2. Bundling

As discussed, persistent and pipelining HTTP still suffers from the issue when web objects are kept in multi-domains. By having bundling mechanism, both DNS resolving and HTTP fetching can be entirely offloaded from the mobile device to the Remote Proxy. Besides, bundling splits the TCP connections between the mobile device and web servers to optimise TCP behaviour over congested wireless links. The main benefit from the bundling mechanism is not only to keep the link utilised during the transmission and reduce overhead of maintaining multiple TCP connections, but also to keep the mobile device in lower power consumption state during the period of web object fetching and bundling in the Remote Proxy. In Figs. 11 and 12,

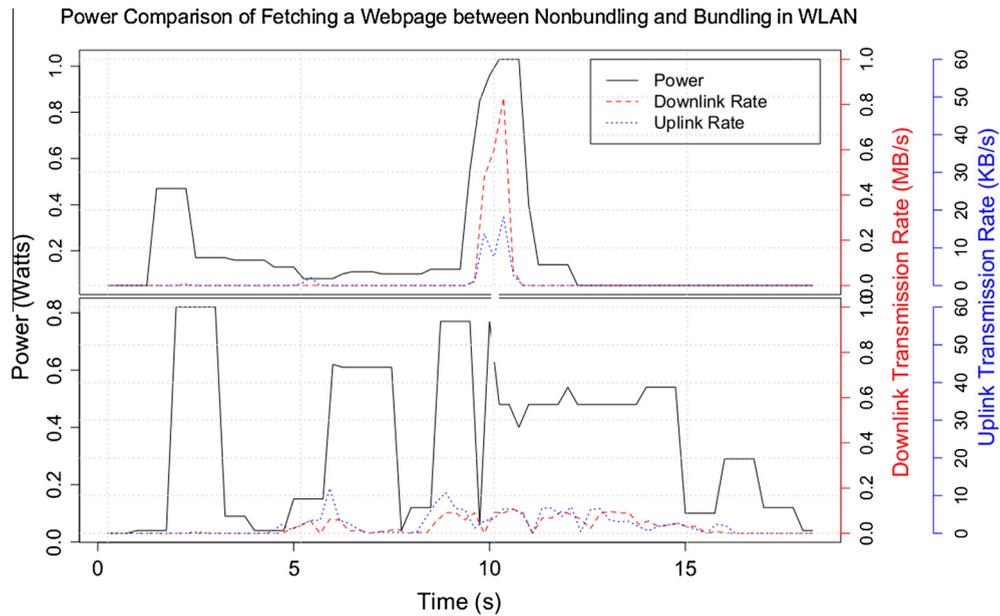


Fig. 11. The operating power of fetching a webpage with/without bundling in WLAN network.

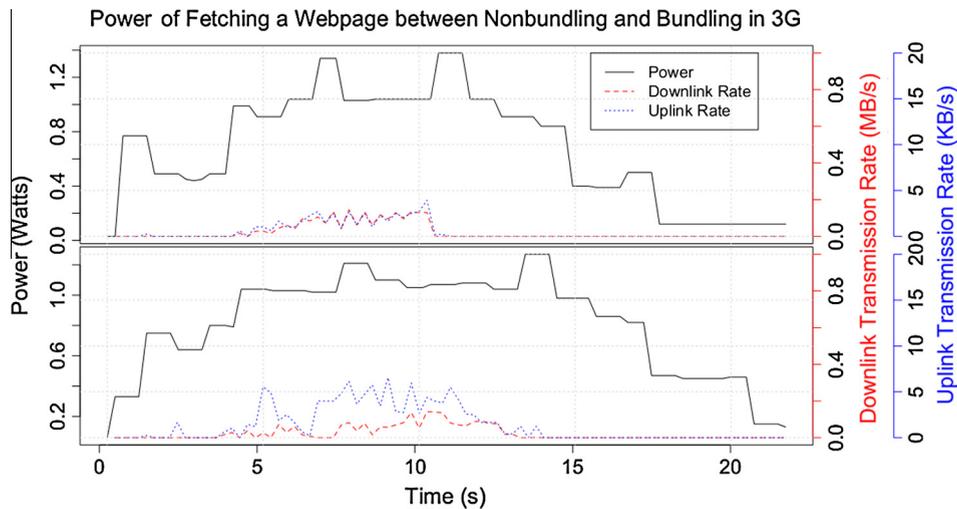


Fig. 12. The operating power of fetching a webpage with/without bundling in 3G network (1 Mbps).

the operating power of fetching a webpage with size of 678 KB using bundling and unbundling mechanisms in WLAN and 3G is depicted. As shown, the download time reduces from 15 s to 9.5 s in WLAN network, and from 20 s to 17 s in 3G networks. After TCP connection setup and the EEP request is sent, the operating power of fetching with bundling in WLAN drops immediately and keeps in lower power consumption state until the bundled reply is back. The total energy consumption reduces from 6.15 Joules to 2.69 Joules. In the case of fetching in 3G network, the difference is that the operating power declines only to the operating power of Cell\_FACH state not Cell\_PCH state, since the value of T2 timer was shorter than the time of

spending on bundling. Furthermore, the energy consumption is able to drop from 17.84 Joules to 14.36 Joules.

It is worth mentioning the fast dormancy here, which was introduced to 3GPP Release 8 [45,46] to offer the possibility for the UE to control the state changes actively rather than only being controlled by RNC. A Signalling Connection Release Indication (SCRI) message is explicitly sent to the network to indicate the desire of state demotion. With the fast dormancy, it is possible for the UE to enter Cell\_FACH or Cell\_PCH immediately after the EEP request is sent. This provides the potential to increase the energy savings offered by the EEP proxy. However, the feasibility of applying the fast dormancy for web traffic requires the

knowledge of the aggregated traffic of all applications to predict a long enough period for the changing to dormant mode. The accuracy of prediction is also often compromised since applications act independently and users inputs introduce randomness. Some research has been carried on studying the problems [47,48].

### 4.3. Header compression

HTTP header compression was already proposed in few studies [26,43] and their main purpose of reducing the size of header is to improve the serialisation latency to send requests, especially in low bandwidth networks, and also reduce the size of replies. Based on the measurements we did against Alexa top 100 websites [43], the average size of all the HTTP requests for the HTML documents during first view (meaning visiting the website for the first time) is 395 bytes, and the average size during repeat view (meaning a returning visit) is 602.17 bytes. The differences between the two views are that some web objects may have been already cached in browser, and HTTP requests normally have cookies associated. After the header compression and EEP header added, the average size of EEP requests without cookies included is 296.84 bytes, and that of EEP requests with cookies included is 432.09 bytes. Compared with non-compressed headers, the average size of EEP headers in the both cases are reduced 28.2% and 24.9% respectively. In Fig. 13, the cumulative distributions of the size of the four types of headers are illustrated.

The most important reason of integrating HTTP header compression in the system is that compressing the requests offers possibility of keeping the mobile device in an even lower power consumption state after sending the new requests. As seen in Fig. 14, a normal HTTP request to “www.wikipedia.org” sent from the N9, and its operating power went up to Cell\_DCH state and then took 24.01 s to switch back to initial power consumption state. When the request was sent as an EEP request, the N9 only

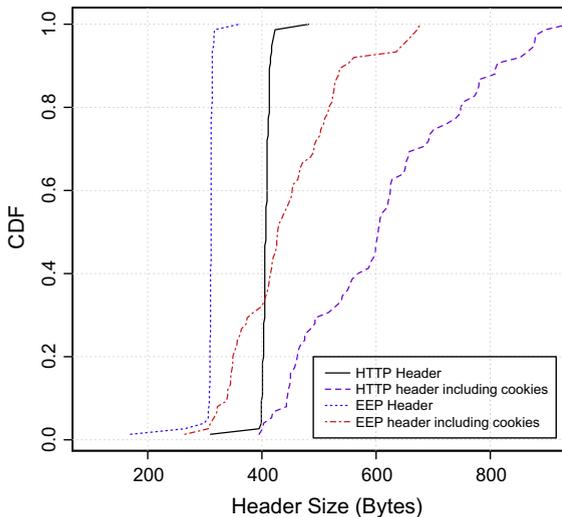


Fig. 13. CDF of the size of HTTP and EEP headers.

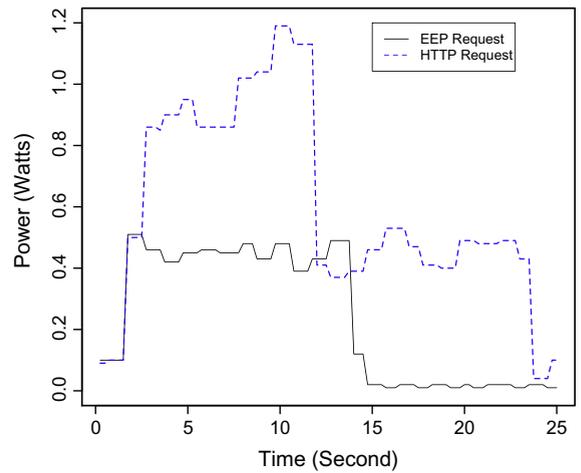


Fig. 14. The operating power of sending HTTP requests with/without HTTP header compression in 3G networks.

switched to Cell\_FACH state and was able to enter Cell\_PCH state in 13.57 s. The size of the HTTP request including its cookie is 451 bytes and it reduces to 337 bytes as a EEP request. The mobile device was kept in Cell\_FACH instead of Cell\_DCH due to the size reduction, which grants the device the chance of staying in a relative lower power consumption state during the period between sending the request and starting to receive the bundled reply.

The triggering to different RRC states is decided by whether the size of requests reaches the buffer threshold. As it is size-dependent, we show how much energy could be saved if header compression is applied over requesting Alexa top 100 websites [44] in Fig. 15. In average, the energy consumption of the requests without cookies included reduces 4.7%, and that of the requests with cookies included reduces 26.7%.

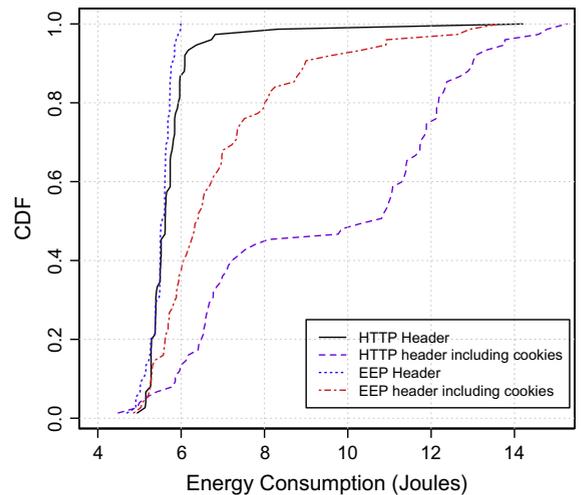


Fig. 15. CDF of energy consumption for sending HTTP requests with/without HTTP header compression in 3G network.

**Table 1**  
Measured webpages.

Wegpage	Type	Size	# of CSS	# of JavaScripts	# of images	Total # of web objects
Copernicus	Small	254 KB	1	0	38	41
Ebay	Medium	678 KB	3	3	37	45
Youtube	Large 1	1.43 MB	2	3	84	92
Tumblr	Large 2	1.2 MB	3	5	18	27

#### 4.4. Selective compression

Nowadays, the size of screen for mobile devices tends to be increasingly bigger, given that the usage penetration of the mobile devices with bigger screen is higher than normal ones in consuming the Internet content and services [24]. Besides, the market share of tablet is expected to reach a new high of 190 million shipment units with year-on-year growth of 48.7%, and will surpass desktop PCs in 2013 [21]. Based on these two facts, it might not be necessary to wrap text and decrease image resolution in original web content. Unlike.mobi [22] and Opera mini [23], our solution applies loss-less compression to reduce the number of bit transmitted over the air, and keep content unchanged when it reaches mobile users.

According to our previous study [32], there is a trade-off between computation and transmission over wireless links when using compression to save energy on mobile devices. Compression squeezes the amount of bytes transmitted over the air to reduce transfer size and improve response time, but increase the energy consumption spending over compression/decompression. Thus, compression-enabled communication should take the type of transmitted data, bandwidth of wireless link and energy consumption of compression/decompression into consideration. The scheme to achieve the energy-efficient communications is called selective compression. The compression condition is determined by two ratios. One is  $R_{threshold}$  defined as the ratio of the energy consumption of receiving one bit to the energy consumption of decompressing one bit. The other is  $R$ , which is defined as the ratio of the number of bits of compressed data to that of deduced data (the delta between compressed and uncompressed data). The derivation can be found in our previous study [42]. The ratio  $R_{threshold}$  is approximately constant and can be obtained from measurements on given specific mobile device. Therefore, selective compression examines every received object. If  $R$  over this object is less than  $R_{threshold}$ , compression is applied. Compared to selective compression, HTTP compression uses public domain compression algorithms to encode text-based files such as HTML, XML, JavaScript and CSS files, and other file formats like ICO and SVG files. Since transmitting one bit over the air is energy-costly, the selective compression is able to improve the energy saving further by carefully checking the web object.

## 5. System evaluation

To evaluate the system, the downloading time of tested webpages, and the energy consumption of fetching the webpages are used as metrics. In this section, we first

introduce the experimental setup and measured performance metrics in Section 5.1, and then the implementation is briefly discussed in Section 5.2. Experimental results are given in Section 5.3.

#### 5.1. Experimental workload, testbed and measured performance metrics

Table 1 shows the characteristics of our selected workload. The solution tends to keep the original content as much as possible, thus the selected webpages are not their mobile versions. There are 4 webpages<sup>1</sup> chosen, namely “ETSI reference web page “Copernicus” [34]”, and the front pages of “www.ebay.com”, “www.youtube.com” and “www.tumblr.com” presenting “Small”, “Medium”, “Large with more web objects” and “Large with less web objects” respectively. All the webpages were kept as original

We set up our experimental testbed as shown in Fig. 16. The Nokia N9 was used as a mobile client, in which we installed the Local Proxy code, a customised web browser based on QWebkit enabling automotive measurements, and the Nokia Energy Profiler to record the operating power and downlink/uplink speed. A DLink DIR-815 wireless router was used for WLAN measurements. The router offers maximum 54 Mbps downlink throughput theoretically and the N9 can achieve around 13 Mbps downlink speed and 5 Mbps uplink speed in practice. The Elisa Finland commercial 3G network was used for initial measurements and a Nokia Siemens Networks Flexi Multiradio BTS with 3GPP Release 6 specifications was used for evaluating the system, with which we were able to control the network parameters such as subscription rates and the values of inactivity timers. The Remote Proxy was installed on a Linux machine with the specification of 2.4 GHz Intel Core i3 CPU and 3.2 GB memory.

In the same machine, two virtual machines were created as middle boxes with netem installed to create delay and packet losses between the N9 and the proxy, and between the proxy and a web server, where tested webpages were served by an HTTP compression enabled Apache HTTP server. Considering the real scenarios in WLAN networks, wireless APs are located close to users’ mobile devices but the devices suffer from the delay and packet loss over the links between the APs and web servers, which are sometimes located far away from each other. Thus, we introduced 150 ms delay in the netem box B during the measurements for the WLAN network, except the measurements of evaluating the performance over RTTs, where the delay varied. As analysed in Section 2.1, the spurious

<sup>1</sup> Captured on April 12, 2012.

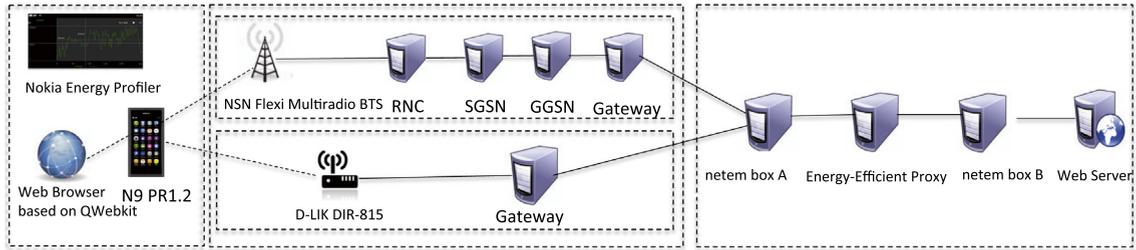


Fig. 16. Measurement setup.

delay in 3G networks is mainly due to allocation/de-allocation of network resources, buffering and bandwidth sharing among users. Since the testbed we had for evaluation in the 3G network did not have traffic load and the end-to-end delay was around 60 ms, we applied 350 ms delay in the netem box A during the measurements for the 3G network, except when the performance over RTTs was evaluated. It is prevalent to delegate static web content to Content Delivery Networks (CDNs) to get the content rapidly and reliably delivered to end-users. Therefore, there was only 15 ms delay introduced in netem box B.

We evaluate the performance of our solution mainly from perspectives of network quality, web content, proxy functionalities and models of mobile devices to explore how these metrics affect QoE of mobile Web access in terms of downloading time and consumed energy. The detailed metrics of measured performance are listed in Table 2.

## 5.2. Implementation

The system was implemented over Qt SDK version 4.8, which is a cross-platform application and UI framework. Qt APIs speed up the development of networking related functionalities and also provide QWebkit library giving the ability to parse web content easily. Since Qt provides high reusability of one code base for multiple platforms, the Local Proxy on the Nokia N9 and Remote Proxy on a Linux machine share part of the source code for the EEP protocol stack, the functionalities of compression/decompression and HTTP connection. We also used Qt Quick, a CSS/JavaScript-like programming language for rapid UI creation, to implement a mobile application to demonstrate the system in term of downloading time and energy saving.

Table 2  
Measured performance metrics.

Network Quality/parameter	Proxy functionalities
<i>Metrics</i>	
RTT (ms)	Bundle
Throughput (Mb/s)	Compression
Packet Loss (%)	Caching
Inactivity timer (s)	
<i>Web content</i>	
Size of page (B)	Device models
# Of web objects	Hardware
	Operating system

## 5.3. Results

In this section, we try to answer the following questions: (1) How much can the proxy improve web access? (2) How much energy the proxy can save? (3) How web content, network delay and link speed affect the results? (4) How the inactivity timers affect the results in 3G network?

### 5.3.1. Network quality

This section presents the results of comparing the performance of downloading the medium-sized test webpage over different RTTs using the normal web browsing or using the proxy with bundling and selective compression enabled. Figs. 17 and 18 show the download time and energy consumption comparison in the WLAN and 3G networks respectively. The average values are presented in bar charts and their 95% confidence intervals are also given to indicate the reliability of the measurements. In order to emphasise the improvements, the saving in percentage for both cases are marked next to the curly brackets.

We evaluated the proxy under four different network latencies, namely 60 ms, 160 ms, 410 ms and 610 ms, which were introduced to netem box B for WLAN and netem box A for 3G. As seen in the figures in general, the savings increase as the delay grows in both the networks. With the normal web browsing, the download time and energy consumption increase dramatically as expected. The time and energy consumption increase 178% and 90% in the WLAN network, and 116% and 109% in the 3G network respectively when the delay increases from 60 ms to 610 ms. Compared to the performance of using the normal browser, the solution of using the proxy is resistant to the network delay resulting in only 17% more spent time and 43% extra energy consumption in the WLAN network. Even through the download time increases considerably when the delay rises in the case of using the proxy, the growth of energy consumption is contained in a small scale since the header compression and bundling give the mobile device the chance to remain in lower power consumption state.

Since the link speed in WLAN is typically not limited in APs in normal scenarios, we mainly focus on how link speed affects the performance of the proxy in the UMTS network as demonstrated in Fig. 19, where the proxy was evaluated under two different subscription plans. As indicated, the solution of using the proxy gains more benefit over slow or congested wireless links. As the link speed in-

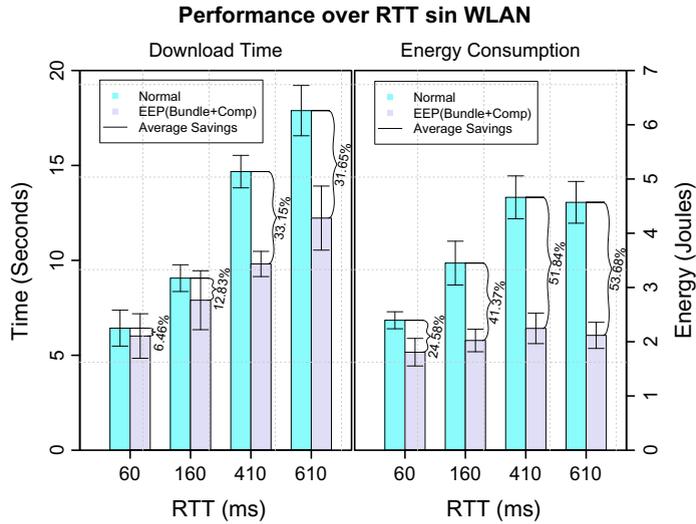


Fig. 17. Download time and energy consumption of Medium-sized webpage over different RTTs in WLAN. There is 0% packet loss rate introduced by the netem boxes.

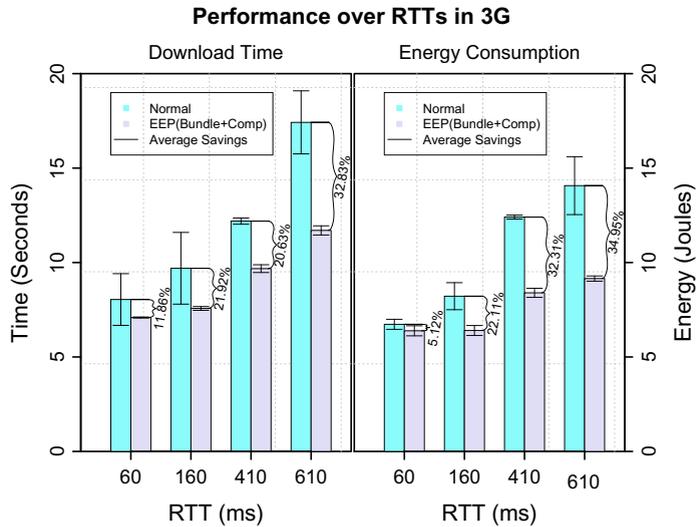


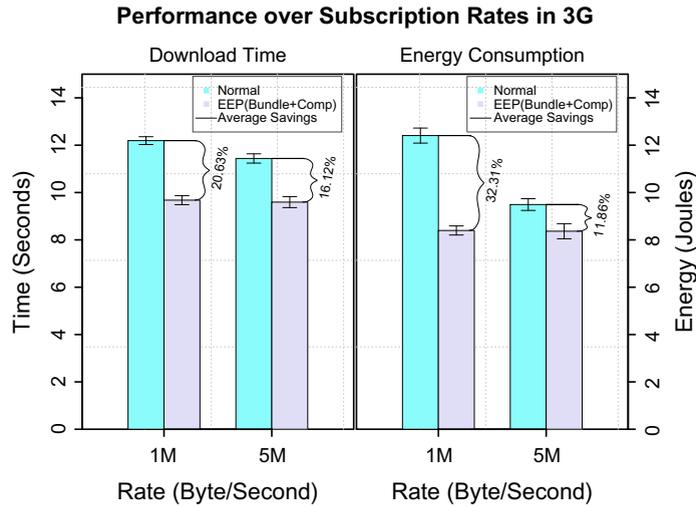
Fig. 18. Download time and energy consumption of Medium-sized webpage over different RTTs in 3G. There is 0% packet loss rate introduced by the netem boxes, and 15 ms delay introduced in netem box B. The subscription rate is 1Mbps, and the timers are 3 s and 2 s for T1 and T2.

increases fivefold in the new subscription plan, the time and energy spent on downloading the bundle decrease resulting in less saving correspondingly. We also evaluated how packet loss rate affect the performance of downloading a webpage in Fig. 20. As the packet loss rate grows from 0% to 2.0%, the time saving of using the proxy increases from 9.12% to nearly 50%. Given the measurement case, the energy can be saved over 58.26% when there is no packet loss, and increases to nearly 70.56% when the packet loss rate grows over 1.5%.

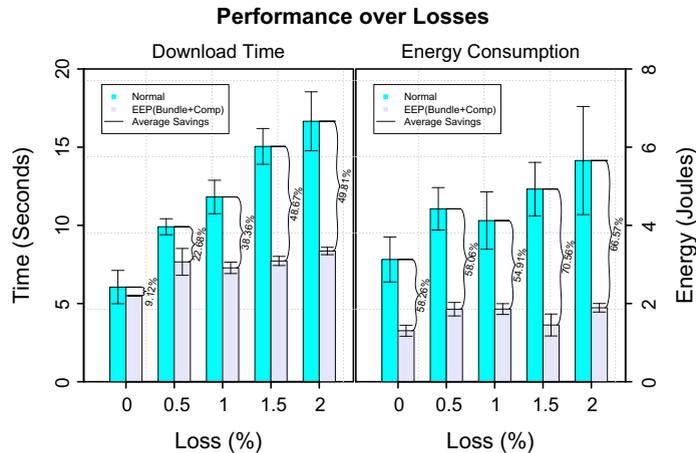
### 5.3.2. Inactivity timers

In order to improve battery life, the mobile device that has been in the higher power consumption states for a period of time without sending or receiving any data can be

moved to the lower power consumption states. The period of time is decided by the inactivity timers, which are configurable in the RNC. We examined three different combinations of T1 and T2 values as shown in Fig. 23. Since the timers do not affect download throughput and the behaviour of the process, download time basically keeps constant when using the proxy thought all the value combinations. The energy savings are slightly decreased as the values of timers are increased. As observed in the measurements, the time spent on fetching and bundling is approximately 3.5 s. Considering the values of the timers, the N9 was able to enter the lower power consumption after sending a request when T1 is less than 3.5 s in this particular test case. Thus, when T1 is 2 s and T2 is 1 s, the N9 can entirely enter Cell\_PCH state but reside there



**Fig. 19.** Download time and energy consumption of Medium-sized webpage over subscription rates in 3G. There is 0% packet loss rate introduced by the netem boxes, 350 ms delay introduced in netem box A and 15 ms delay introduced in netem box B. The timers are 3 s and 2 s for T1 and T2.



**Fig. 20.** Download time and energy consumption of Medium-sized webpage over different packet loss rates in WLAN. There is 150 ms delay introduced in netem box B.

for only half second. This explains why the differences of energy savings are minimal. This type of measurements was done when there was 15 ms delay introduced in netem box B, which is the scenario that the web content is located in CDNs. However, there are still a large fraction websites running without supporting of CDNs. Then the delivery time of web content from these websites to the proxy could be longer and offer the N9 longer residency in the lower power consumption states.

5.3.3. Web objects

Figs. 21 and 22 present the effect of the content of webpages over the performance in the WLAN and 3G networks respectively. The number of bytes of the webpages decides the download time and energy consumption directly as comparing small-sized, medium-sized and large-sized webpages. Since the most of objects in small-sized page are images, the savings over this page is limited but the

proxy still can achieve 31.82% energy saving in the WLAN network and 28.56% energy savings in the 3G network. It is worth mentioning the comparison between Large 1 and Large 2 webpages. Both of the webpages are similar-sized, but the first one outnumbers the second one in terms of the number of web objects. The web browser installed in the N9 was built based QWebKit, which limits the number of requests executed in parallel to 6 for one host/port combination, and no HTTP pipelining enabled. Thus the browser has to request the web objects more times over each connection resulting in longer downloading. As seen in the figures, the download time of the Large 1 webpage is 33% and 30% more than that of Large 2 webpage in WLAN and 3G networks. The correspondent energy consumption is 52% and 51% more, respectively. Using the proxy, we are able to reduce the energy consumption of loading the Large 1 webpage down to 62.38% and 65.37% in each network.

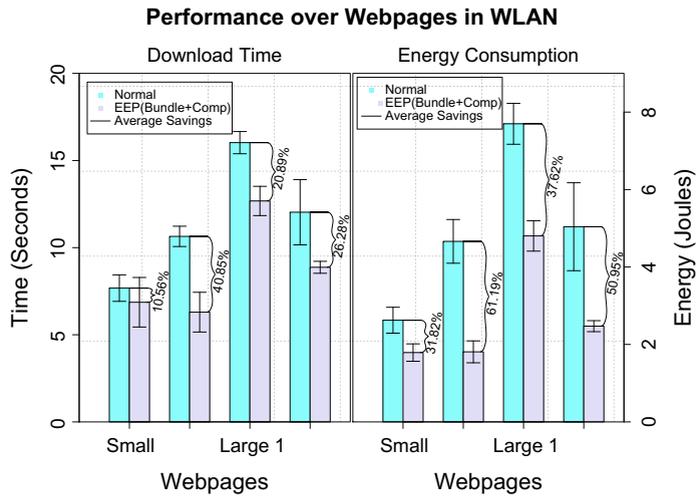


Fig. 21. Download time and energy consumption of Medium-sized webpage over different webpages in WLAN. There is 0% packet loss rate introduced by the netem boxes, and 150 ms delay introduced in netem box B.

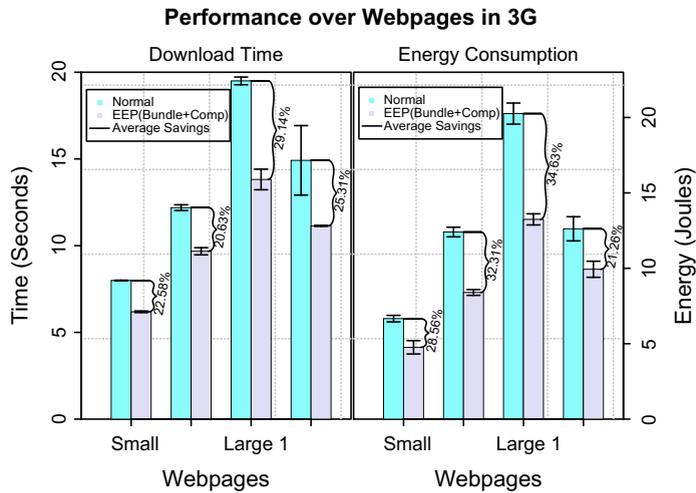


Fig. 22. Download time and energy consumption of Medium-sized webpage over different webpages in 3G. There is 0% packet loss rate introduced by the netem boxes, 350 ms delay introduced in netem box A and 15 ms delay introduced in netem box B. The subscription rate is 1 Mbps, and the timers are 3 s and 2 s for T1 and T2.

5.3.4. Features

Figs. 24 and 25 show the performance over the functionalities provided by the EEP proxy. As seen in the figures, the medium-sized webpage was taken as an example to demonstrate the impacts from each scheme. As more energy saving schemes are applied to the proxy, the savings incrementally increase in general. As indicated in Section 4.3.1, the time saving becomes significant only when the network delay is over 400 ms in the 3G network. Compared to the WLAN network, the energy saving in the 3G network is also less significant when using bundling. Since the energy saving mechanisms in WLAN and 3G networks behave differently, the percentages of the savings brought by bundling are different as well. In the 3G network, the time spent on fetching and bundling of the webpage by proxy is relatively too short to let the inactivity

timers expire resulting in that the N9 does not have the chance to enter Cell\_PCH state or stay there longer enough after the proxy receives the request for the webpage. The selective compression is more meaningful when the webpage is transmitted over high-delayed and congested wireless link, e.g. the 3G network with 350 ms delay and 1Mbps subscription rate. As to the caching, the benefit it brings is trivial since the cache mechanism provided by Qt always asks the web server to send the web object only if it has changed, when the proxy has the web object in its local cache. HTTP requests with a header field “If-Modified-Since” are sent for every cached web object. If the server tells the proxy that the web object has not changed since last fetching time, in this case, the server sends HTTP “304 Not Modified” reply containing only HTTP header without payload. Even though it reduces the amount of

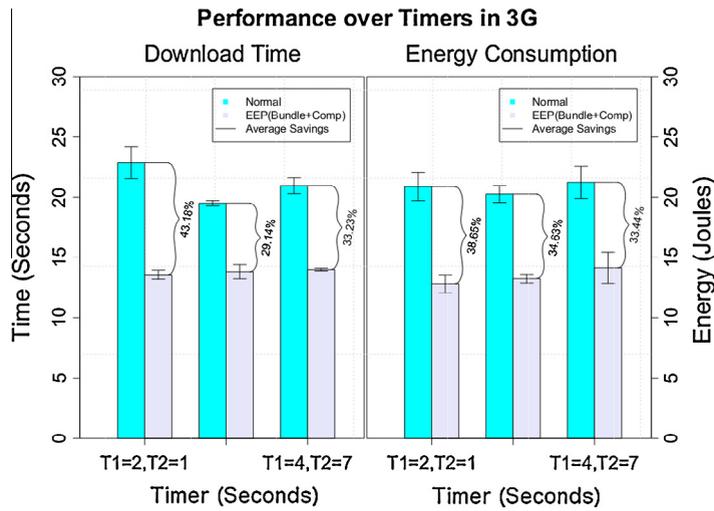


Fig. 23. Download time and energy consumption of Large 1 webpage over inactivity timers in 3G. There is 0% packet loss rate introduced by the netem boxes, 350 ms delay introduced in netem box A and 15 ms delay introduced in netem box B. The subscription rate is 1 Mbps.

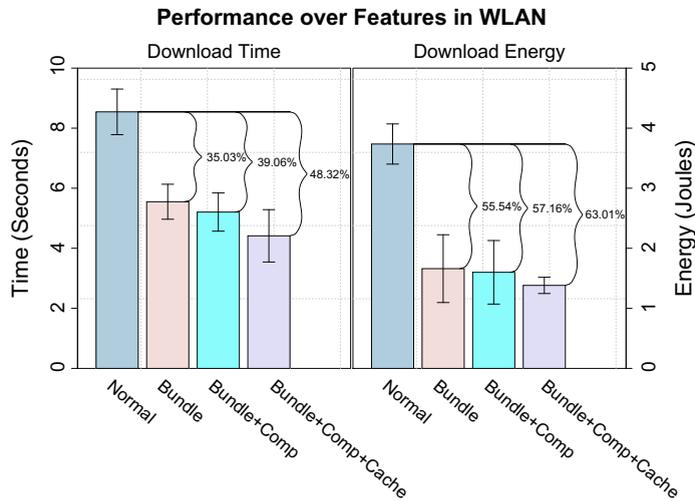


Fig. 24. Download time and energy consumption of Medium-sized webpage over different features in WLAN. There is 0% packet loss rate introduced by the netem boxes, and 150 ms delay introduced in netem box B.

transmitted data by not sending the web object in HTTP reply, it always involves a roundtrip delay between the server and proxy, which compromises the improvement offered by the cache.

5.3.5. Hardware and OS

To evaluate the effect of different hardware and operating system, a more recent popular smart phone, the Samsung Galaxy SIII LTE, was compared to the Nokia N9. The specifications of the two devices are listed in Table 3. Since the original code of our Local Proxy was written in Qt and C++, we ported the code to the Android platform via Necessitas [52], which adds a dynamic library and a Java application wrapper to the original code. To our knowledge, there is no suitable software to provide accurate power measurement results on the Android platform, similar to what we can obtain from the Nokia Energy Profiler on

Meego platform. Thus, we used a power meter MAHEG HM8115-2 with a power socket HZ815 to measure the operating power of the Galaxy SIII. With the power socket, the power measurement of tested devices can be done with micro USB power cable connected to the phone on one side and plugged into the HZ815 on the other side. The comparison used the ETSI reference web page “Copernicus” in real WLAN and 3G networks.

As can be seen in Fig. 26, the performance of the Galaxy SIII is slightly better than that of the N9, which is understandable, given that Galaxy SIII was released nearly one year later. More powerful CPU/GPU leads to faster execution of JavaScript and page rendering. Moreover, a more recent radio chipset can potentially result in less operating power and faster download. In general, the EEP solution provides significant mobile download time and energy savings on both modern mobile devices.

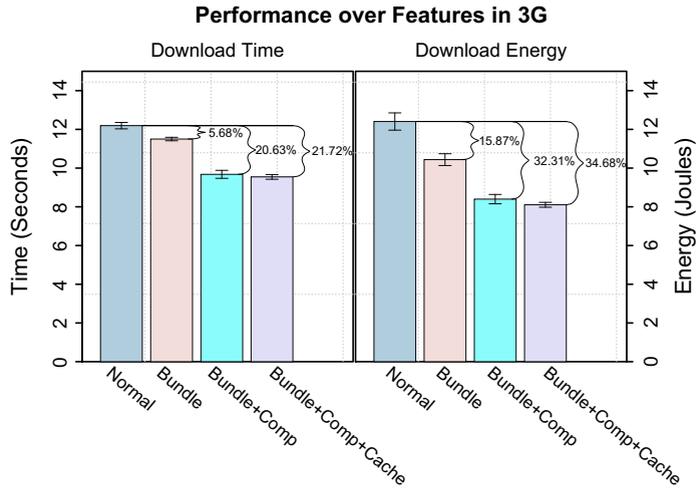


Fig. 25. Download time and energy consumption of Medium-sized webpage over different features in 3G. There is 0% packet loss rate introduced by the netem boxes, 350 ms delay introduced in netem box A and 15 ms delay introduced in netem box B. The subscription rate is 1 Mbps, and the timers are 3 s and 2 s for T1 and T2.

Table 3 Comparison between Nokia N9 and Galaxy SIII.

Model	OS	Processor	Memory	Display	Released date
N9	MeeGo R1.2	1G Cortex-A8	1 GB	AMOLED 3.9"	29.09.2011
Galaxy SIII	Android 4.0.4	1.4G Cortex-A9	1 GB	AMOLED Plus 4.8"	29.05.2012

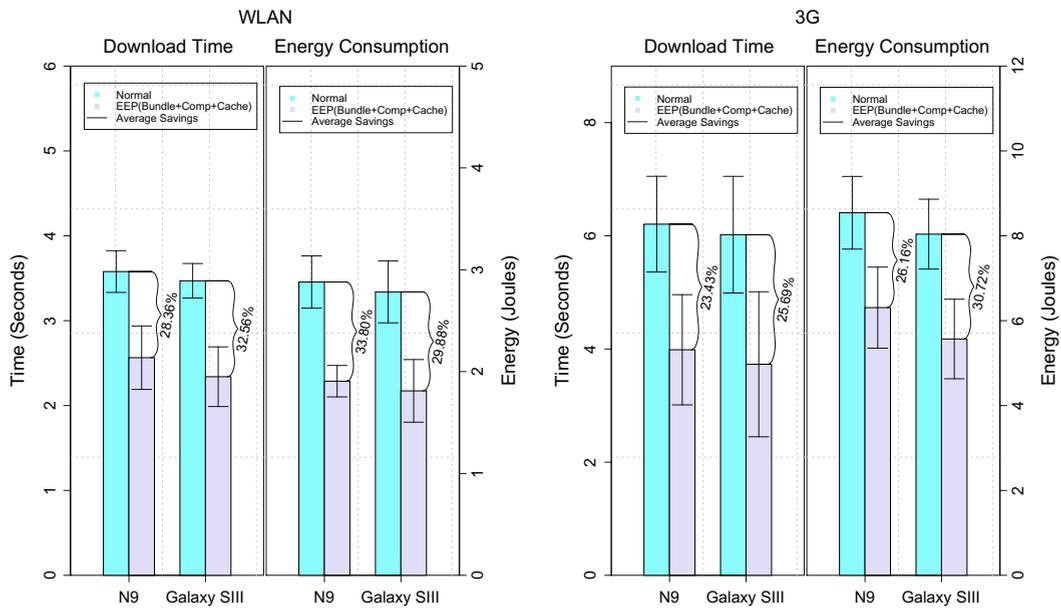


Fig. 26. Download time and energy consumption of Nokia N9 and Samsung Galaxy SIII in WLAN and 3G networks.

6. Discussion

In this section, we present the other observed performance results, and discuss further optimisations and possible improvements for the system.

6.1. CPU and memory

Compared to the normal web browsing, extra resources of the mobile device have to be assigned to the tasks of unbundling and decompression when using the

proxy. As observed, the CPU usage of the normal web browsing is about 15.74% when fetching the medium-sized webpage. In case of having bundling and compression enabled, the CPU usage increases to 19.17%. The memory usage only increases 2.22 MB when the proxy is in use, respectively. The CPU and memory usages vary across different webpages. The CPU usage increases among the four webpages are 21.56%, 7.21%, 17.91% and 9.15% respectively. Moreover, there is no substantial difference in memory usage, the range of which is within 2.8%.

## 6.2. Network traffic

The uplink and downlink network traffic through the mobile device were examined. To determine how much traffic can be reduced, we recorded the number of bytes that went through the 3G interface. The pool of tested webpages was selected from Alexa top 100 global sites and the webpages were determined by pareto distribution based on their ranks. There were 1000 requests for the chosen webpages sent for each case. The EEP approach is extremely efficient and causes 62.7% reduction in terms of the number of bytes sent out from the mobile device, since only one EEP request and TCP acks for downloading EEP replies are needed. For the downlink traffic, the amount reduce 6% due to the compression. Overall, there is 11% traffic reduction for total number of bytes transmitted over the air. The traffic reduction relieves the network burden and especially alleviates the possibility of peak traffic during rush hour. Moreover, considering the current cost of international roaming charges ranges are from \$4.08 in Greece to \$24.06 in Canada per megabyte of data, and the average price of OECD is \$9.27 in 2011 [49], the proxy is also helpful to reduce the expense of users spent on web service when being abroad.

One more thing worth discussing is the statistical properties of web traffic, especially Self-Similarity (SS) and Long-range Dependence (LRD), which are largely verified properties of Internet traffic, severely affecting network queuing performance. SS is the property showing an object's appearance is unchanged regardless of the scale of time. When this property is observed at a wide range of time scales, it can exhibit LRD, meaning values of any instant are positively correlated with values at all future instants. Since the bundling transfers traffic pattern into FTP-alike traffic, which shows a heavy-tailed distribution, the traffic would still maintain self-similar. But the aggregated file transfer can affect the degree of exhibited characteristics of self-similarity [50]. Moreover, the proxy would also change web traffic models, like ON/OFF model, where self-similar traffic can be built by multiplexing a large number of ON/OFF sources that have ON and OFF period. Bundling would change ON times corresponding to the transmission duration of individual web objects, since all the web objects arrive at the same time. However, if a HTTP request matches a bundle, the traffic can still be modelled with ON/OFF model with different ON and OFF periods.

## 6.3. Cookies

One of the reasons to design the system in such way is to keep all necessary information (cookies, ETag, etc.) embedded in HTTP headers delivered from web server to mobile client. The cookies are commonly used as the most important way to keep consistency of web services for remembering browsing histories and user login states. However, the way how the mobile device works with the proxy breaks the consistency of the cookies. After the EEP request is received by the proxy, the mobile device is not able to associate its cookies to the HTTP requests, which are sent from the proxy to the web server. Hence, the proxy needs maintaining databases to make sure that correct cookies for each particular web browser are associated with the HTTP requests. When the proxy receives HTTP responses from the web server, it records all the cookies into the database associated with the browser it sent the HTTP request on the mobile device. When the proxy receives the EEP request, it appends the cookies to each HTTP request sent to the web browser. The databases are created per web browser basis, and which browser an EEP request is sent from can be identified by a hash index, which is the value of a string consisting a Local Proxy ID and browser's name. Since only hashed values are stored in the proxy, the user information is kept anonymously. The Local Proxy also creates a database to ensure the reliability of the consistency. The database is always consistent with the database in the Remote Proxy, since it inherits same design principle. The Local Proxy crosschecks its own cookie database with the web browser's cookie database. If there is any mismatching, two HTTP header fields defined as "CookieToAdd" and "CookieToDelete" are used to notice the Remote Proxy which cookies should be added to the database and which cookies should be removed.

## 6.4. Deployment and scalability

In this paper, the EEP is deployed on a stand-alone proxy to provide energy-efficient and fast web content delivery for end devices. The proxy can be a service offered by an independent third party, or, for example, a telecom operator's serving gateway could integrate the technology to provide the service for their customers. It can also be a solution for customer premises equipment or femtocells to serve home or corporate users. A further deployment scenario would be to integrate our technology directly into the content server; this way the energy efficient delivery of content can be offered by the content provider without a third party in the middle.

In order to provide an overview of the scalability of the solution, we performed a stress test of the EEP. The server machine where the EEP was installed, was equipped with Intel Xeon E5520 4-core, 2.27 GHz CPU, 3 GB memory and 320 GB SATA-II hard disk. We recorded CPU and memory usage of the server while concurrent requests were coming to the proxy. The test clients were initialising multiple connections to fetch the test page "Copernicus" through the proxy. When there were 500 connections and the concurrency of connections was 50, the consumed

memory was around 10 MB and CPU usage was around 12–15%.

When integrating the solution, device or service owners need to take the additional power consumption into account. The power supply for the server is 200 W. However, we need to point out that the operating power highly depends on the use case and the underlying hardware. For example, a home router can serve less users than the server above, and the operating power of such device could be within 10–20 W. The operating power also varies dramatically depending on the architecture of hardware platform. For example, in the study of benchmarking two web server architectures [51], a EnergyCore server was compared with an Intel Core server. At full CPU utilisation, the EnergyCore server consumed only around 5 W and the Intel Xeon platform consumed over 100 W, while the difference in performance was only around 20%.

## 7. Conclusion

Our work has shown that delivery of web content is not optimal for mobile devices. To enable the system and radio interface to save power, we have to transmit data in quick bursts instead of a continuous bursty low bit rate stream. Since we use Internet services more and more from battery-powered devices, the bundling concept we have presented could have a huge impact in people's lives. In the paper, we also analyzed and discussed how radio links and RRC inactivity timers affect energy consumption and content delivery time, demonstrating that significant savings can still be achieved on careful adaption to the radio links and their power saving features and timers.

Our concepts could be well implemented directly into web servers in the future and the client side would be part of web browsers. Current implementation works with most web pages and services on the Internet. In order to cater to emerging web technologies and services, more work is required to make them work properly in our system. In the future, we will put more focus on qualitative analysis of battery lifetime improvements. Furthermore, with the increasing LTE coverage, we need to investigate how this new mobile technology affects the energy-saving potential of our concepts.

## Acknowledgements

The work was partly performed within the ECEWA Project, funded by the Finnish National Technology Agency Tekes and industry. The authors thank Tero Isotalo from Tampere University of Technology for providing valuable help.

## References

- [1] ICT Facts and Figures. <<http://www.itu.int/ITU-D/ict/facts/2011/material/ICTFactsFigures2011.pdf>>.
- [2] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2011–2016. <[http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns7.05/ns827/white\\_paper\\_c11-520862.html](http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns7.05/ns827/white_paper_c11-520862.html)>.
- [3] Batteries for Portable ICT Devices. <[http://www.itu.int/dms\\_pub/itu-t/oth/23/01/T230100000E0001PDFE.pdf](http://www.itu.int/dms_pub/itu-t/oth/23/01/T230100000E0001PDFE.pdf)>.
- [4] M.V.J. Heikkinen, J.K. Nurminen, Consumer attitudes towards energy consumption of mobile phones and services, in: Vehicular Technology Conference (VTC 2010-Fall), 2010.
- [5] StatCounter. <[http://gs.statcounter.com/#mobile\\_vs\\_desktop-ww-monthly-201105-201205](http://gs.statcounter.com/#mobile_vs_desktop-ww-monthly-201105-201205)>.
- [6] A. Bakre, B.R. Badrinath, I-TCP: Indirect TCP for mobile hosts, in: Proc. of the 15th IEEE International Conference on Distributed Computing Systems, Vancouver, BC, 1995, pp. 136–143.
- [7] K. Brown, S. Singh, M-TCP: TCP for Mobile Cellular Networks, ACM Computer Communication Review (1997).
- [8] K. Ratnam, I. Matta, W-TCP: an efficient transmission control protocol for networks with wireless links, in: Proc. IEEE ISCC, 1998.
- [9] Chakravorty, S. Katti, J. Crowcroft, I. Pratt, Flow aggregation for enhanced TCP over wide-area wireless, in Proc. IEEE INFOCOM, 2003.
- [10] M.C. Rosu, D. Rosu, An evaluation of TCP splice benefits in web proxy servers, in Proc. ACM WWW, 2002.
- [11] M. Tariq, A. Zeitoun, V. Valancius, N. Feamster, M. Ammar, Answering What-If Deployment and Configuration Questions with WISE, in Proc. ACM SIGCOMM, 2008.
- [12] A. Misra, V. Navda, R. Ramjee, V.N. Padmanabhan, E.M. Belding, Cool-tether: energy efficient on-the-fly WiFi hot-spots using mobile phones, in Proc. CoNEXT, 2009.
- [13] M. Jimeno, K. Christensen, A network connection proxy to enable hosts to sleep and save energy, in IEEE IPCCC, 2008.
- [14] C.C. Lee, J.H. Yeh, J.C. Chen, Impact of inactivity timer on energy consumption in WCDMA and cdma2000, in Proc. IEEE WTS, 2004.
- [15] J.H. Yeh, J.C. Chen, C.C. Lee, Comparative analysis of energy-saving techniques in 3GPP and 3GPP2 systems, IEEE Transactions on Vehicular Technology 58 (1) (2009).
- [16] N. Balasubramanian, A. Balasubramanian, A. Venkataramani, Energy consumption in mobile phones: a measurement study and implications for network applications, in: Proc. of ACM SIGCOMM Internet Measurement Conference, 2009.
- [17] H. Liu, Y. Zhang, Y. Zhou, TailTheft: leveraging the wasted time for saving energy in cellular communications TailEnder, in: Proc. MobiArch, 2011.
- [18] A. Abhari, A. Serbinski, HTTP modification to reduce client latency, Electrical and Computer Engineering (2008).
- [19] C-H. Chi, H. Wang, W. Ku, Proxy-cache aware object bundling for web access acceleration, Book Web Content Caching and Distribution (2004) 257–273.
- [20] M.C. Rosu, C.M. Olsen, C. Narayanaswami, L. Luo, PAWP: a power aware web proxy for wireless LAN clients, in: 6th IEEE Workshop on Mobile Computing Systems and Applications, 2004.
- [21] Worldwide Smart Connected Device Market Crossed 1 Billion Shipments in 2012, Apple Pulls Near Samsung in Fourth Quarter, According to IDC. <<http://www.idc.com/getdoc.jsp?containerId=prUS24037713#UVNGAldVV8E>>.
- [22] .mobi. <<http://dotmobi.mobi/content/aboutmobi>>.
- [23] Opera Mini. <<http://www.opera.com/mobile/features/>>.
- [24] Shocker: Smartphone Users Like Bigger Screens, Market Share may Respond Accordingly. <<http://www.engadget.com/2012/09/03/shocker-smartphone-users-like-bigger-screens/>>.
- [25] G. Xylomenos, G. Polyzos, P. Mahonen, M. Saaranen, TCP performance issues over wireless links, IEEE Communications Magazine 39 (4) (2001).
- [26] R. Chakravorty, A. Clark, I. Pratt, Optimizing web delivery over wireless links: design, implementation, and experiences, IEEE Journal on Selected Areas in Communications 23 (2) (2006).
- [27] Pocket. <http://getpocket.com/>.
- [28] M.C. Chan, R. Ramjee, TCP/IP performance over 3G wireless links with rate and delay variation, in: Proc. the 8th Annual International Conference on Mobile Computing and Networking, 2002.
- [29] K.C. Barr, K. Asanovic, Energy-aware lossless data compression, ACM Transactions on Computer Systems 24 (3) (2006) 205–291.
- [30] L. Wang, J. Manner, Energy consumption analysis of WLAN, 2G and 3G interfaces, in: Proc. of 2010 IEEE/ACM International Conference on Green Computing and Communications, 2010.
- [31] RFC 2616 – Hypertext Transfer Protocol – HTTP/1.1. <http://www.faqs.org/rfcs/rfc2616.html>.
- [32] L. Wang, J. Manner, Evaluation of data compression for energy-aware communication in mobile networks, in: Proc. of International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, 2009.
- [33] R. Maddah, S. Sharafeddine, Energy-aware adaptive compression scheme for mobile-to-mobile communications, in: Proc. of IEEE 10th International Symposium on Spread Spectrum Techniques and Applications, 2008.

- [34] ETSI Copernicus Reference Web Page. <<http://portal.etsi.org/stqj/WebReferencePage.asp>>.
- [35] Nettiutka Measurement Platform. <<http://www.nettiutka.fi>>.
- [36] P.H.J. Perala, A. BarbuZZi, G. Boggia, Theory and practice of RRC state transitions in UMTS networks, in: Proc. of 5th IEEE Broadband Wireless Access Workshop collocated with IEEE Globecom, BW-WAWS, 2009.
- [37] Q. Feng, Z.G. Wang, A. Gerber, Z.M. Mao, S. Sen, O. Spatscheck, Characterizing radio resource allocation for 3G networks, in: ACM Internet Measurement Conference, 2010.
- [38] 3GPP TR 25.480: Terminal Power Saving Features.
- [39] Signals Research Group, Smartphones and A 3G Network. <[http://www.nokiasiemensnetworks.com/system/files/document/Signals\\_Research\\_-\\_Smartphones\\_and\\_a\\_3G\\_Network\\_May\\_2010.pdf](http://www.nokiasiemensnetworks.com/system/files/document/Signals_Research_-_Smartphones_and_a_3G_Network_May_2010.pdf)>.
- [40] D. Kristol, L. Montulli, RFC 2109:HTTP State Management Mechanism. <<http://www.ietf.org/rfc/rfc2109.txt>>.
- [41] B.C. Housel, D.B. Lindquist, WebExpress: a system for optimizing web browsing in a wireless environment, in: Proc. of the 2nd Annual International Conference on Mobile Computing and Networking, 1996.
- [42] L. Wang, J. Manner, Proxies for energy-efficient web access revisited, in: 2nd International Conference on Energy-Efficient Computing and Networking, 2011.
- [43] B. Zhao, B.C. Tak, G. Cao, Reducing the delay and power consumption of web browsing on smartphones in 3G networks, in: Proc. the 31st International Conference on Distributed Computing System, 2011.
- [44] Alexa Top Site. <<http://s3.amazonaws.com/alexa-static/top-1m.csv.zip>> (09.26.12).
- [45] UE "Fast Dormancy Behavior, 3GPP Discussion and Decision Notes R2-075251, 2007.
- [46] Configuration of Fast Dormancy in Release 8, 3GPP Discussion and Decision Notes RP-090960, 2009.
- [47] F. Qian, Z. Wang, A. Gerber, Z.M. Mao, S. Sen, O. Spatscheck, TOP: tail optimization protocol for cellular radio resource allocation, in: ICNP, 2010.
- [48] Using Heartbeat. <[http://harmattan-dev.nokia.com/docs/library/html/guide/html/Developer\\_Library\\_Best\\_practices\\_for\\_application\\_development\\_Optimising\\_power\\_consumption\\_Using\\_heartbeat.html](http://harmattan-dev.nokia.com/docs/library/html/guide/html/Developer_Library_Best_practices_for_application_development_Optimising_power_consumption_Using_heartbeat.html)>.
- [49] Mobile Roaming Fees Highest in Canada: OECD. <<http://www.cbc.ca/news/technology/story/2011/06/08/technology-mobile-roaming-wireless-oecd.html>>.
- [50] K.J. Christensen, V. Ballingam, Reduction of self-similarity by application-level traffic shaping, in: Proc. the 22nd Annual Conference on Local Computer Networks, 1997.
- [51] Apache Benchmarks for Calxeda's 5-Watt Web Server. <<http://armservers.com/2012/06/18/apache-benchmarks-for-calxedas-5-watt-web-server/>>.
- [52] KDE Necessitas Project. <<http://necessitas.kde.org/>>.



**Le Wang** is a postgraduate student in Aalto University, Finland. He received his MSc. (2009) from Helsinki University of Technology. Since then he has been a full-time researcher at Aalto University, Department of Communications and Networking and focuses on energy efficiency research on mobile devices, and his research interests are networking, mobile computing and energy saving technologies in general.



**Jukka Manner** (born 1972) received his MSc. (1999) and PhD. (2004) degrees in computer science from the University of Helsinki. He is a full professor (tenured) of networking technology at Aalto University, Department of Communications and Networking Comnet). His research and teaching focuses on distributed systems and various networking aspects, most recently on the development of the Internet and its services, particularly in topics related to energy efficient ICT, networking beyond IP, multipath connectivity and transport protocols. He is the Academic Coordinator for the Finnish Future Internet research programme. He is an active peer reviewer and member of various TPCs. He has contributed to standardization of Internet technologies in the IETF for over 10 years, and was the co-chair of the NSIS working group. He has been principal investigator and project manager for over 15 national and international research projects. He has authored over 80 publications, including several IETF RFCs. He is a member of the ACM and the IEEE.