|  |
| --- |
|  |
| Megacharge |
| The impact of session rounding on mobile data costs |
| University of Technology Sydney |
| March, 2016 |

****

Megacharge: The impact of session rounding on mobile data costs

Authored by University of Technology Sydney (“UTS”)

Published in 2016

Supported by funding from the Australian Communications Consumer Action Network (ACCAN)

The operation of the Australian Communications Consumer Action Network is made possible by funding provided by the Commonwealth of Australia under section 593 of the *Telecommunications Act 1997*. This funding is recovered from charges on telecommunications carriers.

University of Technology Sydney
ABN: 77 257 686 961
UTS CRICOS Provider Code 00099F
Website: [www.uts.edu.au](http://www.uts.edu.au)
Report prepared by: Brett Hagelstein and Mehran Abolhasan
Faculty of Engineering and Information Technology
E-mail: mehran.abolhasan@uts.edu.au
Telephone: +61 2 9514 2448

Australian Communications Consumer Action Network
Website: [www.accan.org.au](http://www.accan.org.au)
E-mail: research@accan.org.au
Telephone: +61 2 9288 4000
TTY: +61 2 9281 5322

ISBN 978-1-921974-36-6

Cover image: shutterstock.com and Richard Van Der Male



This work is copyright and owned by UTS. It is licensed under the Creative Commons Attribution 4.0 International Public Licence. You are free to cite, copy, communicate and adapt this work, so long as you attribute “University of Technology Sydney (UTS) and the Australian Communications Consumer Action Network (ACCAN)”. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/Iegalcode>

This work can be cited as: University of Technology Sydney 2016, *Megacharge:* *The impact of session rounding on mobile data costs*, Australian Communications Consumer Action Network, Sydney.

Contents

Executive Summary 1

Survey of Over-Plan Data Costs 2

Objective 2

Survey data 2

Discussion 3

Data Use Survey 4

Objective 4

HREC approval 4

Survey design 4

Survey publication 4

Survey distribution 4

Response summary 5

Data Billing Analysis 6

User classification 6

Scaling factor for total download 6

Scaling factor for data per session 7

Scaling factor for minimum session size 9

Discussion and Conclusion 11

Appendix: Bill Processing 12

Method 12

Data storage 12

**Figures**

Figure 1: The cost of going over the data quota can vary significantly between plans 3

Figure 2: The scaling factor is generally larger when the data used is smaller 6

Figure 3: The scaling factor is inversely proportional to the amount of data downloaded per session 8

Figure 4: Data session size distribution is a function of the usage profile 9

Figure 5: A smaller minimum session size results in a significantly smaller scaling factor – rounding to 10KB or 25KB sessions has minimal impact on billed data use 10

# Executive Summary

Mobile data is an important resource in our connected lives, but one that is challenging to estimate when selecting a mobile plan. Our mobile data utilisation can vary significantly from month to month depending on many factors such as connectivity to Wi-Fi when at home or work and use while commuting. Another factor that many do not consider is the impact of data session rounding.

Some data plans record the data session size and round up to the next megabyte (MB). While this may not appear to be significant, particularly if coupled with a low cost plan, it can form a significant proportion of the amount of data billed. For example, this means that a 10 kilobyte (KB) data session for an email push will result in a 1MB usage or charge applied to your bill.

This research studies the impact of megabyte session rounding based on the observation of more than 20000 data sessions. The key observations are:

* The range of scaling factors – the multiple of the billed data to that of the actual data used after session sizes are rounded up – was 1.01 to 5.2, with a median of 1.3.
* *Light* data users (less than 500 MB/month) are the most significantly affected by the use of megabyte session rounding and typically add around 80% of their actual used data in rounding “overhead” in their bill.
* The session rounding overhead is inversely proportional to the average session download, but the session data sizes can vary significantly between users with similar monthly downloads.
* Plans that offer smaller session rounding sizes (such as 10KB, 25KB or even 0.1MB) effectively eliminate the significance of session rounding overhead on the billed data.
* Session rounding overhead tends to have a greater impact on personal data use than business plans.

# Survey of Over-Plan Data Costs

## Objective

Mobile data providers offer a large range of data plans to meet various use profiles and budgets. While the amount of included data is clearly described in most mobile plans, the charge applied once a user exceeds that amount (the over-plan data costs) is less visible and can vary significantly between providers and plans. This survey explores some of the different billing mechanisms used by providers for over-plan data use. A high over-use cost can rapidly become a large bill, especially when coupled with “nearest megabyte” data session rounding.

## Survey data

A brief survey of the over-plan data costs available from major carriers was conducted using the information available on carriers' websites. The results of a survey taken on 26th November 2015 are presented below.

Table 1: Survey of over-plan data costs

|  |  |  |  |
| --- | --- | --- | --- |
| **Provider** | **Type** | **Billing unit** | **Over-quota (up to)** |
| Aldi | Prepaid PAYG | 25KB\* | $0.05/1MB |
| Aldi | Prepaid Pack | 25KB\* | $5/300MB (starting at) |
| Amaysim | Prepaid | KB\* | $0.072/1MB |
| Boost | Prepaid | KB\* | $2/1MB |
| Dodo | Prepaid PAYG | MB | $0.05/1MB |
| GoTalk | Prepaid | MB | $0.05/1MB |
| Hello Mobile | Prepaid | *Not stated* | $0.05/1MB |
| Lebara | Prepaid | KB\* | $0.15/1MB |
| Live Connected | Postpaid | KB | $0.04/1MB |
| Optus | Plan | KB | $10/1GB |
| Optus | Prepaid | KB | $5/500MB |
| SpinTel | Plan | KB | $10/1GB |
| TeleChoice | Plan | 10KB | $10/1GB |
| Telstra | Plan | KB | $10/1GB |
| Telstra | Prepaid | KB\* | $2/1MB |
| TPG | Prepaid | KB | $10/1GB |
| Vaya | Plan | KB\* | $10/1GB |
| Vaya | Data only | KB\* | $0.02/1MB |
| Virgin | Plan | KB | $0.052/1MB |
| Vodafone | Plan | KB | $10/1GB |

\* Indicates advertised changes made between April and November 2015 that reduce the billing units from megabyte session rounding to the value indicated. Still applies to the Telstra Prepaid Simplicity Plan.

## Discussion

Figure 1 shows the actual cost to the user for exceeding the data quota by a given amount for a range of indicative plans. As seen in the table, the cost of over-quota data can be quite significant. The cheapest option for users depends on how much they will go over their quota, but in general, data packs provide a more economical option for users exceeding their quota by more than 150MB. Conversely, while a per-megabyte over-quota charge may appear cheaper, they can cost significantly more than data packs for many practical examples.

Figure 1: The cost of going over the data quota can vary significantly between plans



Furthermore, the higher over-plan rates are somewhat correlated with megabyte data session rounding. In these cases, each data push – such as email and social media notifications – over the data plan limit, can generate separate 1MB (rounded) data sessions costing $0.10 each.

# Data Use Survey

## Objective

A survey of real-world data usage will illustrate how different use-cases can influence the data billing process.

## HREC approval

Human Research Ethics Committee (HREC) approval was required for this survey per UTS research policy.

This project was a shorter duration than most university research projects, generally multi-year higher degree research projects, and required a rapid turnaround time to ensure the project progress goals were met. The survey was designed to comply with UTS HREC’s Nil/Negligible Risk designation which was approved by the Faculty Research Office and did not require a full HREC examination.

## Survey design

The process of collecting detailed data usage information is complex and requires several steps on the part of the participant. The survey was designed to be as informative and direct as possible to maximise the number of participants and the ratio of valid responses.

The most important aspect of the survey was collecting the itemised mobile data information, which is generally available via a web portal or by contacting the provider. This was the single largest hurdle for participants and was a key factor in the survey design. Generating and improving procedures for each provider was an iterative process which incorporated participant feedback.

## Survey publication

The survey was created as an online form using a Google Form. This allowed for a range of question types such as dropdown selections and text boxes with regular expression validation. The response to each question was used as a trigger to load one of a set of custom pages. For example, selecting the provider on one page triggered the appropriate itemised bill collection procedure on the following page.

Participant data received from the survey was stored in a Google Sheet. Like all Sheets, this provided a continuous backup with version control in case someone was able to access it and delete data. A new “user” was signed up to Google to host the survey to reduce exposure to spam and account attacks.

## Survey distribution

The survey was advertised broadly, including in a Whirlpool forum post, and via ACCAN and personal Facebook pages, Twitter accounts, and email lists.

## Response summary

More than 200 responses were received in response to this survey. Of these, 142 responses contained useful itemised billing information – which included more than 20000 data sessions – and was used to generate the following results.

The survey contained a short questionnaire to identify potential causal factors in billing trends. However, there were insufficient valid survey responses to draw conclusions as to the influence of provider, handset hardware, software and application configurations.

# Data Billing Analysis

This section describes the scaling factor results calculated from the received itemised data bills. Details on the process to collect this information are included in the Appendix.

## User classification

Users were classified as either *Personal* or *Business* based on their billing information. *Personal* users were subsequently divided into:

* *Light* data users who downloaded less than 0.5GB per month,
* *Moderate* users who downloaded between 0.5 and 1GB per month, and
* *Heavy* users who downloaded more than 1GB per month

## Scaling factor for total download

The first observation is that *Light* data users are more affected by the scaling factor of megabyte session rounding than *Heavy* users. That is, the rounding factor will form a greater proportion of the billed data when the actual data use is smaller. This is shown in Figure 2, where the blue columns represent the median value and the red error bars show the range of observations.

Figure 2: The scaling factor is generally larger when the data used is smaller



The median *Light* user has a scaling factor of 1.8, that is, for 100MB of actual data used the billed data would be 180MB. The largest scaling factor observed was 5.2 and this was also from a *Light* user. In this case, a user may have a 500MB data limit, and although they only used 100MB of actual data, they would exceed the limit and may have an over-use charge applied to their bill.

The median *Moderate* user has a scaling factor of 1.24. That is, almost 20% of the billed data is simply due to the session rounding. As expected, *Heavy* users who download more data feel a significantly smaller impact from session rounding because any rounding component is a small fraction of a large data session.

*Business* users show a range of scaling factors with a median close to that of a *Heavy* user, although the range of observations is more similar to a *Moderate* user.

Take a hypothetical example of a very careful user on a megabyte session rounding plan with a $0.05/MB over-use charge. The plan is for 500MB of data per month and they set their device to cut the connection when the 500MB is reached. If the user had a scaling factor of 1.8, which is a reasonable value from the observations, this would result in a $20 over-use charge.

There are exceptions to this observation. Take a hypothetical example of a user who only connects to one 400MB data session during the billing cycle and is otherwise not connected – the scaling factor must be less than 1.0025. However, it is expected this use-case accounts for a low proportion of users.

Therefore, in general, users with a lower total data bill are more significantly affected by rounding session downloads up to the nearest megabyte.

## Scaling factor for data per session

The second observation is that the scaling factor will be inversely proportional to the average data session size. That is, smaller data sessions will result in a significantly larger scaling factor because the session rounding is a more significant portion of the data billed.

To show this, the average session data size was calculated for each survey response and plotted against the total scaling factor for that bill. The result is shown in Figure 3 which shows a very consistent inverse relationship between data session size and scaling factor.

Figure 3: The scaling factor is inversely proportional to the amount of data downloaded per session



There is an interesting phenomenon in this figure – the *Light* users always have a lower data download per session than *Heavy* users, but *Moderate* and *Business* users are across the spectrum. Further, a *Moderate* user with the lowest scaling factor of 1.01 used 527MB of data, yet one of the highest scaling factors observed, at 2.24, used 683MB of data which is the reverse of the general pattern. This observation shows that the session handling – which is a complex function of handset hardware, software and applications used, carrier, terrain and usage pattern – plays an important role in the potential scaling factor of megabyte session rounding. That is, even knowing the actual data use, the range of scaling factor values is potentially large and it can be difficult to predict the billed data value when rounding sessions to the next megabyte.

*Business* users present an interesting difference from most private users – Figure 3 suggests they have fewer data sessions and generally download a large amount of data per session. *Business* users may generally connect to a workplace Wi-Fi during the day because this is a cost effective design for the business. It is possible that *Business* users may be required to access large amounts of data remotely, which is why they have the device, and correlates with the high data per session observations.

To expand on this idea, Figure 4 shows the distribution of data sizes per session for each usage category. The first observation is that 48% of *Light* user data sessions are less than 10KB. That is, 48% of *Light* user data sessions have a rounding overhead of more than 99% when using a nearest megabyte session rounding.

All usage categories have a spike of data sessions less than 10KB. This is caused by data pushes, such as email notifications, that generate a data session which is terminated before additional data is downloaded.

Figure 4: Data session size distribution is a function of the usage profile



The second interesting observation is that *Business* users have the highest proportion of data sessions greater than 1MB and the lowest proportion of data sessions less than 0.5MB. This observation agrees with the previous observation that *Business* users generate fewer small data sessions from data pushes, possibly because they are connected to the workplace Wi-Fi. This observation has a greater impact when considering the median monthly *Business* usage is approximately 600MB, which suggests that the business data profile should correlate with the *Moderate* user profile.

## Scaling factor for minimum session size

Recently, some mobile data providers have decreased the session rounding size from “to the nearest MB” to smaller quantiles, such as 0.1MB, 25KB and 10KB. The third observation is that reducing the data session rounding size will greatly reduce the scaling factor. That is, a much smaller amount of data will be lost in session rounding and a larger proportion of the user's data quota is available for actual download.

Figure 5: A smaller minimum session size results in a significantly smaller scaling factor – rounding to 10KB or 25KB sessions has minimal impact on billed data use



Figure 5 shows that reducing the session rounding size even to 0.1MB significantly reduces the scaling factor overhead for most users. The blue columns again represent the median observation and the red error bars show the range. Even the worst-case observation reduces the session rounding overhead from 420% to 38% (scaling factor 5.2 to 1.38).

Further reducing the session rounding size to 25KB or 10KB virtually eliminates this overhead in all observed cases. That is, a data plan that rounds to the next 25KB will mean that the user gets to use almost all of the plan data quota - even *Light* users with small data sessions.

# Discussion and Conclusion

This study collected more than 20000 data sessions from 142 survey responses. It identifies a range of scaling factors of between 1.01 and 5.2, where the amount of billed data can be up to 5.2 times the amount of data used after sessions are rounded up to the nearest megabyte.

The median scaling factor from all personal (non-business) observations was approximately 1.3. That is, a user would be charged for 30% additional data that is lost to session rounding.

It is the light data users who are most heavily affected by session rounding. When users restrict their data usage, this results in smaller data sessions, and a larger proportion of data is wasted when the session size is rounded up.

Business accounts present a different use-case and data session profile than personal users. With a median scaling factor of 1.05, the business connectivity profile shows there is significant opportunity to reduce the overhead of session rounding.

The session handling has a large impact on the session rounding overhead and may have a significant effect on the consumer's data bill. It is a complex function of handset hardware and software configuration, carrier policy, terrain and usage pattern which defines when sessions are terminated. For example, the one device may end a session more rapidly than another in an effort to reduce power consumption. This behaviour is not well understood and provides a significant research opportunity to minimise both consumer data costs and device energy use.

Perhaps not coincidentally, when amaysim changed from megabyte session rounding to kilobyte session measurement, the per megabyte data cost on “As You Go” plans increased from $0.05 to $0.072. This corresponds to a 44% increase, or a multiplication factor of 1.44. It is possible that amaysim modeled a similar value and changed their pricing structure to provide the same value for an “average” user. Importantly, however, the light data users receive a much better deal in this pricing restructure than they previously would have with lower data prices but rounding sessions to the next megabyte.

Consumers with data plans that include session rounding can mitigate the impact on billing by factoring in their own scaling factor in the usage notifications. While the itemised bill after session rounding does not identify the actual data use, consumers can infer the value from the handset's recorded value, and set the notification warning at the appropriate ratio. For example, if the handset recorded a data use of 300MB and the amount billed is 400MB, then the scaling factor was 1.33 and the data use warning should be set at or below 75% of the plan allowance. As the observations in this survey showed, the scaling factor does change due to a variety of factors, and while several samples will help understand individual cases, a more conservative scaling factor should be implemented to avoid over-use penalties.

# Appendix: Bill Processing

The bill processing script reads key information from itemised data bills to calculate the scaling factor and other statistics. This appendix briefly describes the implementation details of this process.

## Method

Survey participants supply information which is stored in the Google Sheet, including the itemised billing data and provider. A Python script is used to parse the billing data and collect statistical values.

Generally, each provider formats their bills slightly differently. The script uses the defined provider information to select the correct custom profile and extract just the data session sizes from the bill. If the survey response is an edited version of the data bill, for example all identifying information is already removed, then the bill is manually checked to confirm that the custom profile still applies. In cases where this is not possible, generally when the manual user editing was not consistent, the response data must be manually prepared and a simple profile is used.

The provider profiles defined which sections of text corresponded to data session sizes while the other values were ignored. In some cases (Vaya, Optus and Telstra), the bill format changed during the survey period or was different between customers. Logic was added to the script to test for each format style to apply the correct profile.

Some providers use PDF encodings which resulted in inconsistent data reading when parsed by the Python PDF module and instead were copied to text file using Adobe Reader and manually sanitised. Similarly, survey responses using Excel spreadsheets were opened and converted to simple text files for parsing. Both processes were manual and required approximately three minutes to convert and validate.

Bill processing is handled using the Python scripting language (version 2.7) and statistics were calculated using the Numerical Python module (NumPy).

## Data storage

Statistics from each response were aggregated in a CSV file for final storage. This allowed for simple importing into spreadsheet or statistical applications to generate figures.

Accidental inclusion of duplicate user responses will skew the results and it is therefore important to detect and remove duplicate entries. However, duplicate entries are harder to detect in non-identifying result tables because there are no dates, phone numbers, names or email addresses to compare. Instead, a hash is calculated on the dataset to detect if two responses contain the same sequence of data session sizes. This hash value is then used as the filename of the session data file and is easily detected if a duplicate user submission is received at a later time.

Other user responses such as handset details can also be collected in the CSV file.