Journal of Telecommunications and the Digital Economy

Volume 7 Issue 2

June 2019

JTDE Volume 7, Number 2, June 2019
Table of Contents

A Forum on the Future of Australia's NBN

Editorial

Mark A. Gregory
RMIT University

Abstract: The future of the $51 billion Australian National Broadband Network (NBN) remains unknown, with the Government still to commit to a course of action after the current build phase. Industry representatives have recently voiced their concerns about a potential future sale of the NBN and how this would occur. In response, the Telecommunications Association is hosting a public forum on the future of the NBN on 31 July 2019 at RMIT University. Papers in the June 2019 issue of the *Journal* include discussion on consumer interest in 5G in New Zealand, the history of Australian mail handling and technical papers covering a range of interesting topics. This month we include a paper titled *Measuring Digital Inequality in Australia: the Australian Digital Inclusion Index* that provides an important insight into digital inclusion. The *Journal* welcomes further contributions on telecommunications and the digital economy.

A Forum on the Future of Australia's NBN

The Telecommunications Association is hosting a forum on the future of the NBN at RMIT University in Melbourne on 31 July 2019. More details can be found on the *telsoc.org* website.

The National Broadband Network (NBN) is Australia's largest Government-funded infrastructure project, which is anticipated to cost the taxpayer about $51 billion by the time it is built and fully operational in 2022.

The NBN is about the future, vital infrastructure that will underpin the nation's future participation in the global digital economy. For the younger generations, the NBN is a key Government contribution towards successful careers, innovation and wealth creation.

The NBN has been controversial, in its inception and with the change by the Coalition Government to the multi-technology mix approach in September 2013. For the telecommunications industry, the NBN has been an important step towards a fair, open and competitive telecommunications market through the provision of uniform national wholesale bitstream products.

Last year, the *Journal* published a paper titled *Australian Wholesale Telecommunications Reforms* ([Gregory, 2018](#)) that takes a look at the future ownership options for the NBN and how there are a number of interlinked policies and regulations that would need to be reviewed and updated, depending on the approach taken by the Government to proceed with retaining or selling the NBN, either as a single entity or by separating access technologies.

Now that the NBN is approaching the end of the rollout, attention is turning to what happens next. Recently, the new Minister for Communications, Cyber Safety and the Arts, Paul Fletcher stated that the NBN would not be sold to Telstra or a company linked to Telstra. Further clarification was provided by the Government that this was a preliminary view and that a decision on what would be done with the NBN would be taken at some point in the future. Broadly, there remains no support in the telecommunications industry for Telstra or a wholly owned subsidiary company, e.g. a restructured InfraCo, to become a future owner of the NBN.

What should be done with the NBN? The Telecommunications Association is hosting a public forum on the future of the NBN to provide a venue for debate on the different approaches that might be taken by the Government after the 2022 Federal election. It is anticipated that there will be a range of views, some of which are at the opposite ends of the spectrum.

If you have an interest in the future of the telecommunications market, then this forum is a must attend.

In This Issue

In this issue of the *Journal* papers cover public policy, improvements in telecommunications technology, digital inclusion, 5G in New Zealand and a historical look at mail handling.

The Australian Mail Handling Scene presents two historic papers from 1966 detailing the mail handling scene in Australia and the development of the state-of-the-art Sydney Mail Exchange.

A Cluster-Driven Energy Routing Protocol for Optimal Network Lifetime in Ad Hoc Networks proposes PEGADyn – a hybrid version of Power-Efficient Gathering in Sensor Information Systems and Dynamic State algorithm for a new energy-efficient routing protocol in ad hoc networks.

Implementation of RF Band-Pass Filter on UMTS Systems to Improve Quality of Service proposes that an RF band-pass filter device be implemented on UMTS systems to reduce high interference and reports on a trial installation in Indonesia.

Energy-Efficient Network Protocols for Domestic IoT Application Design investigates the gathering of data from household IoT-enabled devices, the associated energy cost and the impact of different communications technologies and protocols on that cost.

Dynamic Vehicular Traffic Load: Definition and Quantification proposes a method that quantifies the amount of traffic over time by the help of a cloud calculation service and vehicular communication.

New Zealand Consumer Interest Growing for 5G Mobile summarizes results from a survey of the New Zealand mobile market by Venture Insights in October 2018. The survey had over 1,000 respondents, all of whom were responsible for making their mobile purchasing decisions, with a representative spread across New Zealand, all adult age groups, and customers from the three major mobile service providers.

Measuring Digital Inequality in Australia: the Australian Digital Inclusion Index describes the development of the Australian Digital Inclusion Index, its architecture and the dataset used to populate it.

The *Journal,* Looking Forward

The *Journal* welcomes papers on telecommunications and the digital economy, including theory, public policy and case studies.

Technological change is happening at a rapid rate and consumers anticipate that governments and industry keep pace to ensure that the benefits can be fully utilised. The *Journal* is calling for papers on how new technologies – especially 5G –will affect Australian telecommunications consumers.

The topics of *International Telecommunications Legislation and Regulations* and *International Mobile Cellular Regulation and Competition* are set to continue for some time, as the opportunity to attract papers from around the globe continues. We encourage papers that reflect on where the global telecommunications market is now, how it got to where it is, and what is going to happen next.

Papers are invited for upcoming issues. With your contributions, the *Journal* will continue to provide readers with exciting and informative papers covering a range of local and international topics. The Editorial Advisory Board also values input from our readership, so please let us know what themes you would like to see in future issues.

All papers related to telecommunications and the digital economy are welcome and will be considered for publication after the double-blind peer-review process.

Mark A. Gregory

Reference

Gregory, M. A. (2018). Australian Wholesale Telecommunications Reforms. *Journal of Telecommunications and the Digital Economy*, 6(2), 1-34. https://doi.org/10.18080/jtde.v6n2.155

The Australian Mail Handling Scene

Simon Moorhead

Ericsson Australia and New Zealand

Abstract: Two historic papers from 1966 detailing the mail handling scene in Australia and the development of the state-of-the-art Sydney Mail Exchange.

Keywords: Telecommunications, History, Australia Post, Mail Handling.

Introduction

The two historic papers were published in the *Telecommunication Journal of Australia* in October 1966. Mail handling and telecommunications were the primary focus of the Postmaster General's Department (PMG) from Federation in 1901 until Australia Post and Telecom Australia were separately established in 1975.

The first paper (Page, 1966) details the mail handling scene in Australia in 1966 and highlights the challenges of delivering an increasing volume of mail to a geographically dispersed customer base. At the time, 2.5 billion articles were handled per annum and this increased continuously until 2008 when they peaked at 5.6 billion (Australia Post, 2008).

In 1966, the PMG was well aware of the growing volumes and the implications for standardisation, automation, staffing levels and delivery times. The construction of the state-of-the-art Sydney Mail Exchange in 1966 and the introduction of the current four-digit post code system in 1967 were no coincidence. This was a deliberate strategy by the PMG to cost-effectively manage the increasing postal volumes.

The second paper (Magnusson, 1966) summarises the development of mechanised mail handling around the world and details how the PMG used the new four-digit post code system to significantly simplify the design of the Sydney Mail Exchange. Any reader familiar with the British six-digit, alphanumeric post code system would appreciate the simplicity of the Australian four-digit numeric system, at a time where machine learning was very crude.

Both papers make fascinating reading, given Australia Post still delivers over three billion articles a year but the mix is shifting dramatically to parcels as the community embraces eCommerce and new forms of digital communication (Australia Post, 2018).

References

Australia Post (2008). Connecting People Every Day, Annual Report 2007/08. Available at https://auspost.com.au/content/dam/auspost_corp/media/documents/auspost-annrpt-2008.pdf

Australia Post (2018). Everyone Matters, Annual Report 2018. Available at https://auspost.com.au/content/dam/auspost_corp/media/documents/annual-report-2018.pdf

Magnusson, V. St. G. (1966). The Development of the A.P.O. Mechanised Mail Handling Concept and Overseas Trends, *Telecommunication Journal of Australia, 16* (3), October, 207-214.

Page, R. J. (1966). The Australian Mail Handling Scene. *Telecommunication Journal of Australia, 16* (3), October, 202-206.

The Historic Papers

R. J. PAGE

R. J. PAGE, author of the article "The Australian Mail Handling Scene", joined the Postmaster-General's Department in 1937 as a Telegraph Messenger. After completing the Cadet Engineers' Course in 1941, he worked as an Engineer and Divisional Engineer in the Telephone Equipment and Lines Sections, Victoria, before being seconded on special duty to Central Office in 1951. After completing a Post Graduate Scholarship awarded by the Public Service Board in 1952, during which time he attended the Administrative Staff College, Henley, United Kingdom, he returned to the Engineering Division, Central Office. In 1955, he was promoted to the Postal Services Division as Controller, Transport Branch, and later Controller, Mail Exchange Branch. He was appointed Assistant Director-General, Postal Services Division in 1961, and is now First Assistant Director-General.

Mr. Page graduated Bachelor of Science in 1945, completed the Diploma of Public Administration in 1948, and graduated Bachelor of Commerce in 1950. He is an Associate Member of the Institution of Engineers, Australia.

V. ST. G. MAGNUSSON

V. ST. G. MAGNUSSON, author of the article, "The Development of the A.P.O. Mechanised Mail Handling Concept and Overseas Trends", commenced service in the Sydney Office of the Department where he qualified in 1929 for appointment as engineer. He was transferred to the Central Staff in 1935 where he took up duty in the Telegraphs and General Works Section. During the war years he was seconded to the Department of Munitions to initiate and oversight the production in Australia of communication equipment for the armed services in the Pacific Area.

Soon after his return to the Department he was appointed to the Central Office position of Supervising Engineer, Buildings Branch. It was in this position that he developed the present philosophy of mail flow and the modern concept of the machine systems for the handling of large volumes of mail. Later he took control of the Planning and Development Branch on its establishment in the Postal Services Division where he now holds the position of Deputy Assistant Director-General.

He has taken a leading part in the initiation and launching of the Sydney

Mail Exchange Project, which ranks as one of the largest of its kind in the world. Mr. Magnusson has represented the Department on several overseas missions and is regarded as one of the leaders in his particular field. He is a Senior Member of the Institution of Radio and Electronic Engineers (Australia).

THE AUSTRALIAN MAIL HANDLING SCENE

R. J. PAGE, B.Sc., B.Com., Dip.P.A., A.M.I.E.Aust.*

INTRODUCTION

Australia presents a number of unique problems in mail handling. With a total area of just under three million square miles, 60% of the population live in the six State capital cities, with over 40% in the metropolitan areas of Sydney and Melbourne. The capital cities are separated by distances ranging from 400 miles to almost 3,000 miles. Outside these capital cities, the population is scattered widely; many of the communities being very small and very remote.

The Australian postal service has been linked with the development of each country centre. The existence of a Post Office in each town, catering for a wide range of community needs

----
* Mr. Page is First Assistant Director-General, Postal Services Division.

and acting as a focal point for Government contact on a widely decentralised basis, is characteristic of our development.

The distribution of mail along the established transport routes, combined with the uneven concentration of population, leads to unusually heavy volumes of mail circulating through the capital cities and to equipment and cost problems in its handling. The need for delivery of letters, packets and parcels to outback areas has produced rural road mail services, using every known means of land transport and covering all types of country over distances up to 1,200 miles.

In addition to the capital city G.P.O.'s, there are approximately 8,500 Post Offices scattered throughout Australia, from the smallest non-official Post Office operated in conjunction with a store in the outback,

to a Post Office of the size of Newcastle, having a staff complement of 67.

Mail (and there were over 2,500 million articles processed in the year ended 30th June, 1966, equivalent to 79 articles per second) must be distributed to all these offices and, in most cases, delivered by postman or mail contractor, to over three million delivery points.

Although departmental transport and the railways are used widely, it is not surprising that the aeroplane is used extensively. Since 1959, the normal means of conveyance for the ordinary letter is by air, where it is quicker to do so. Since that date, the amount of domestic mail conveyed by air has quadrupled, and today one in every five articles handled is conveyed by air during some part of its journey.

The normal pattern of mail flow is illustrated in Fig. 1. Mail flows

MAIL CIRCULATION PATTERN IN AUSTRALIA

Fig. 1.
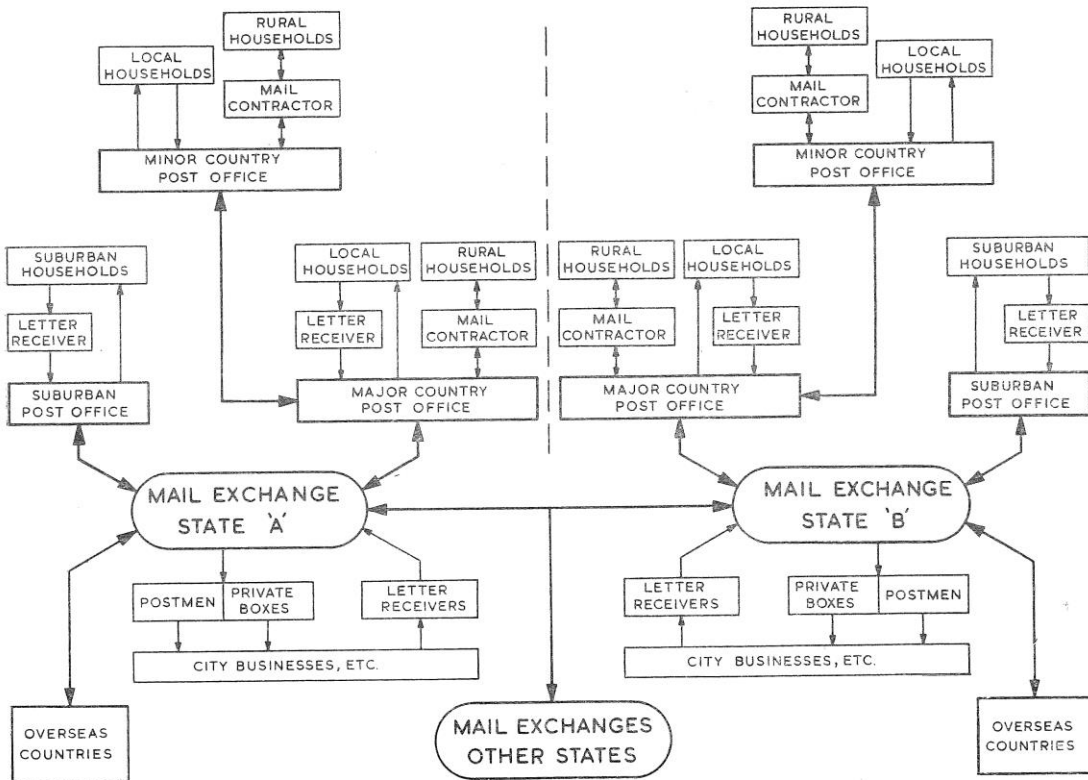
INCIDENCE OF LETTER RECEIPTS
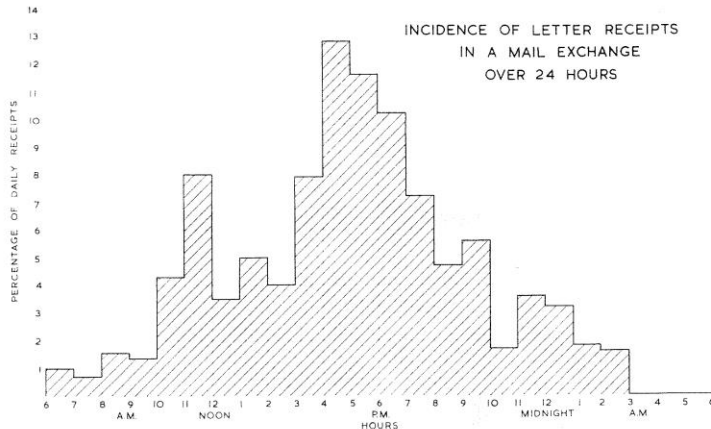IN A MAIL EXCHANGE
OVER 24 HOURS

Fig. 2.

from Post Offices to the central mail handling centre, is processed and then despatched according to its particular destination. It is of interest to note that of mail processed —

35% is for suburban destinations
25% is for city destinations
22% for country destinations
14% is for interstate destinations
4% is for overseas destinations

The interstate component, generally speaking, is conveyed to the mail exchange in the State of destination where arrangements are made for it to be included in the despatch to the particular suburb, town or city concerned.

STANDARDS OF SERVICE

To maintain an efficient postal service, standards of service must be determined. This operates from the time the letter is cleared from the letter receiver, conveyed to the Post Office or Mail Exchange, processed through the various operations at that centre, transported to the Post Office of destination and finally delivered by the postman.

Speaking generally, within the whole capital city area, which in the cases of Sydney and Melbourne each cover about 650 square miles, same day delivery is provided for letters posted in the morning, and next morning delivery for letters posted in the afternoon and evening. Three deliveries a day are provided in the inner city area and two deliveries in most suburbs. Next morning delivery of letters to interstate destinations is provided in most instances. Similarly, standards have been determined for other categories of mail such as newspapers, packets and parcels.

Scientific sampling techniques are used to check the grade of service actually given and information obtained from such checks is used to highlight and remedy weaknesses.

The postal service must meet our customers' demand but, in so doing, we must provide a service which is

reasonable, from both the cost and service standard points of view, and reliable. In some cases in the past, we have provided a service of a standard which is higher than the customer really wants. To enable us to know what the customer does want, and is prepared to pay for, surveys are now undertaken.

Standards of service must be reviewed from time to time. Requirements and habits of the public change and it is essential that the Post Office be aware of the current needs of the customers.

SOME PROBLEMS IN MAIL HANDLING

Inherent in mail handling are the problems of peak traffic so well known to communication engineers. These seasonal, monthly and daily peaks cause many difficulties, which increase rapidly with increase in volume. Fig. 2 gives a general indication of the spread of traffic through the hours of a day in a typical mail exchange.

The "cut off" time in a mail exchange for afternoon suburban delivery is approximately 11.30 a.m. Despatch time for this delivery is approximately 12.45 p.m. This means that there is $1\frac{1}{4}$ hours for the processing of this mail. If the volume of mail for this suburban delivery doubles and if the "cut off" and despatch times remain constant (and any variation in these times has a marked effect on the standard of service), twice the volume of mail must be handled within the same given time. Bearing in mind the peak nature of the traffic input, if the staff is increased to handle this increase in mail receipts, the amount of ineffective time over the stretch of the shift can be increased substantially.

In addition, the spread of the metropolitan areas means that the mail "pick up" runs are attenuated and transit times between the mail exchange and the suburban post offices are becoming greater.

Management has, of course, over the years, introduced innovations, modified procedures, and resorted to a number of expedients to retain the

RELATIONSHIP BETWEEN STAFF AND TRAFFIC
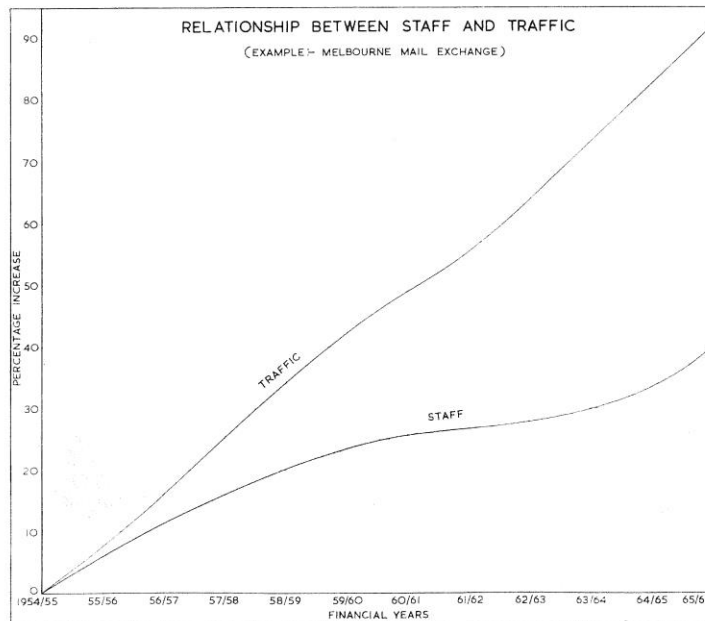(EXAMPLE :- MELBOURNE MAIL EXCHANGE)

Fig. 3.

desired standard of service, having in mind at the same time the economy of operation. Such efforts have been aided by a continuous programme of mechanisation designed to speed the processing of mail. Our efforts have been fruitful and the reward has been the handling of greater volumes of mail without commensurate increase in staff. The graph illustrated in Fig. 3 indicates the relationship of growth in volume of mail handled as against increase in staff numbers, both expressed in percentage. It also indicates that, regardless of the degree of efficiency exercised by management, a point could be reached where staff increases would tend to follow more directly that of volume. Indeed, if volumes continue to increase, a convergence of the two growth curves could eventually be expected. The postal service has always been a big user of labour. At the present time, about 70% of postal costs are labour costs.

It has become obvious that there could be a limit to the improvements that can be brought about with the use of existing handling facilities. Every consideration must therefore be given to the exploitation of major break-through areas or finally face the dilemma of service standards versus operating costs.

There is, too, an overriding motivating force which bears considerable influence on our problem. That is the need in our rapidly developing country to conserve labour to the utmost, so that it may be utilised on various essential capital works of a national character or to strengthen the industrial capacity of the country. It is important then that the Department plays its part in all phases of its activity and utilises modern techniques to best advantage. Conversely, the demand for labour in other directions makes it difficult to obtain operatives for repetitive tasks such as mail sorting.

INTRODUCTION OF AUTOMATIC MAIL HANDLING EQUIPMENT

At this juncture, it would be as well to observe that the Department has not, over the years, been idle in its efforts to apply the principles of mechanisation and automatisation to the handling of mails. There is every reason to believe that we are as advanced as any overseas country in the effective introduction of machinery in mail handling centres. It is most likely the Australian Post Office was as early in the field of mechanical handling as any other country. The first machinery for the handling of "other articles" (packets and newspapers) was installed in the Sydney G.P.O. prior to 1930. The design was superseded some years ago by a later design which is now in general use.

In all mail exchanges, we have had in operation for many years machine systems for the sorting of other articles, machines for the sorting of parcels, bag handling systems, and other

machine aids, such as large letter handling machines. In addition, upwards of 60 letter handling machines have been installed singly as as part of a system, as an aid to the sorting and distribution processes. It is the processing of letter-form articles where the need for advanced machine development is urgent. The number of letter-form articles processed daily

exceeds by far that for all other articles, as may be seen by reference to Fig. 5.

During the financial year ending June, 1966, the total number of letters posted annually in the Commonwealth and including those received from overseas was about 2,100 million, whilst the corresponding number of "other articles" was 400 million. It is

Fig. 4.——Developmental Stages in Letter Processing. Top: Until the year 1960, all letters were primarily sorted on upright presses in all mail exchanges. Middle: Machines designed to aid the sorting and distribution processes and installed to replace upright primaries. Bottom: One of the five suites of coding positions installed in the Sydney Mail Exchange. Letters, when coded, can be processed automatically.

estimated that these figures will double over the next 20 year period.

It is the rapid advance in the field of electronics that has provided the tools for a major break-through in our processing concept, particularly for the treatment of letters. The installation of the letter coding system is the culmination of the first stage of plans to mechanise to the maximum extent possible the processing of letter-form articles. Details of this system have been dealt with in subsequent pages of this Journal. It is, however, relevant to note by reference to Fig. 4, the progress since 1950 when upright sorting presses were still in use throughout the Commonwealth. Their use for primary and secondary sorting was replaced by letter handling machines which are now in turn gradually being replaced by coding and decoding machine systems.

Our problems do not end with the placing of a highly modernised letter handling system into service. Management must develop new skills and recognise too that operations now require changed disciplines. It must foster a closer relationship with the postal user, particularly large business concerns, to ensure that, wherever possible, postal articles passing through the post are suitable for machine or automatic processing. Already new sections of customer education and correction are operating in an endeavour to achieve these ends.

We will be obliged to develop user specifications and modify our rules and regulations as may be necessary to obtain maximum operating efficiency from our equipment. A committee has already been convened by the Standards Association of Australia to prepare an Australian standard for envelopes. When issued this will be an extension of the requirements now being expressed by the Universal Postal Union. The issue of further Australian Standard Specifications will be necessary to eliminate the use of coloured invoices and other envelope insertions on which the address must be read through a panel. We must also endeavour to bring some influence to bear on the size of cheques and documents, etc., designed for transmission through the post.

NEED FOR A MEASURE OF STANDARDISATION.

The necessity for the introduction of a measure of standardisation of mails, particularly letter mail, can be illustrated by reference to the process of "facing up" of letters. Letters from letter receivers arrive at the mail exchange for processing. In the past, before this mail could be sorted, staff had to arrange the letters so that all the stamps were in the top right hand corner. The letters were then fed through a stamp cancelling machine.

This "facing up" process was very costly, and is one of the processes which have been mechanised. Letters, as you can well imagine, are of various types, sizes and shapes. To "face-up" a square envelope automatically is very much more costly than to "face-up" a rectangular envelope because, on a square envelope to be "faced up", a stamp can be in any one of eight positions, whilst on a rectangular envelope, it can only be in one of four positions. (See Fig. 6.)

In addition, if the length of an envelope is considerably greater than its width, the problems of conveying that letter automatically through the various channels are reduced. It is for these reasons that the Universal Postal Union has standardised certain basic requirements for envelopes; one of which is that the ratio of length to breadth should be at least $\sqrt{2}:1$. In any high speed automatic system, some measure of standardisation is inevitable.

DEVELOPMENT OF THE AUSTRALIAN SYSTEM

Although the system used for the coding of letters is explained in detail in later articles, I would like to briefly indicate the type of problems that were encountered in its development. The manual system consists of a series of operations where a Mail Officer reads the address and sorts the letter into a particular slot or pigeon-hole, depending on its destination. The letter, generally, is conveyed by belt to the next sorting stage where a similar manual operation takes place. After a letter is sorted on the final divisions, it is despatched to its destination. Most of the letters are handled two or three times.
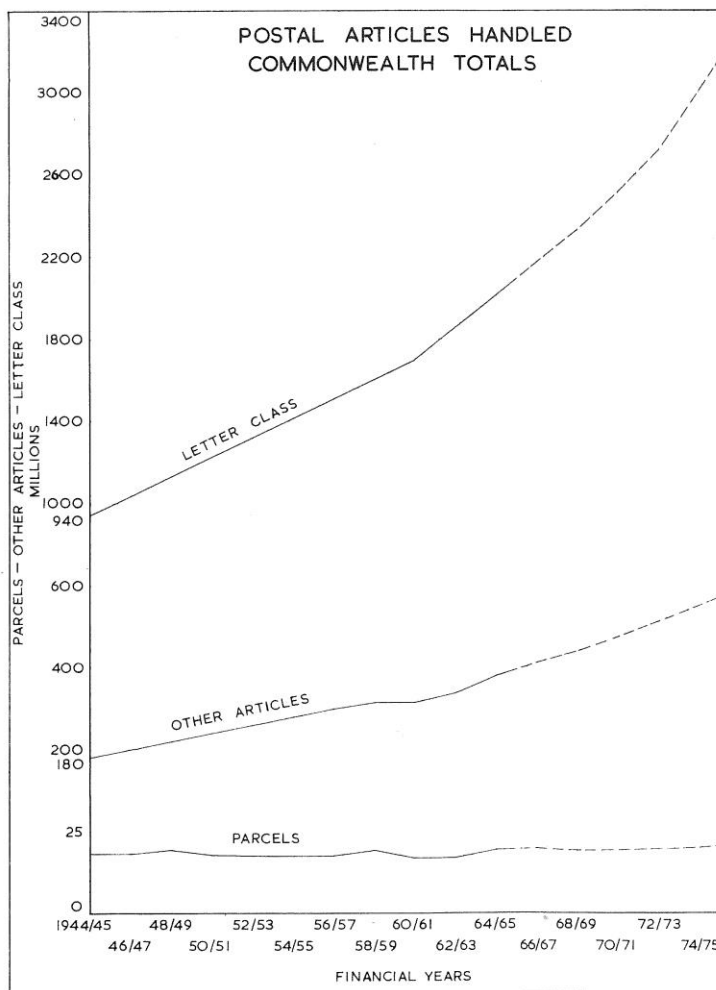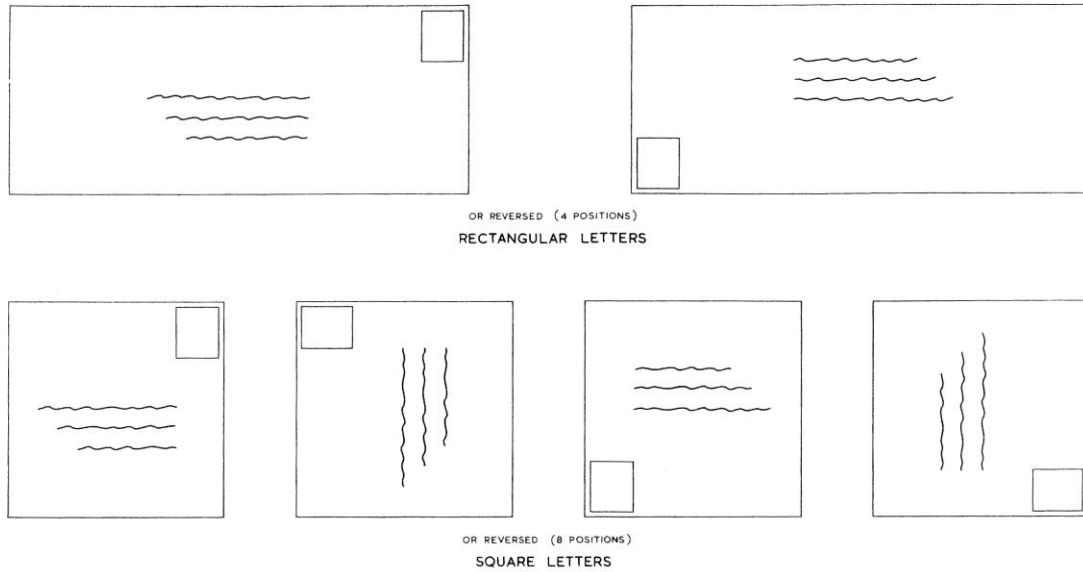

Fig. 5.

OR REVERSED (4 POSITIONS)
RECTANGULAR LETTERS

OR REVERSED (8 POSITIONS)
SQUARE LETTERS

Fig. 6.

In developing a system of electronic coding of letters, the problem of the place name is, of course, fundamentally important. Some overseas countries are using a system of numeric coding — where the public simplifies the problem by including in the address the four or five digit numeric code appropriate for the particular place of destination. This is a very good scheme and a numeric code of four digits has been developed for all Australian place names and a national publicity campaign has been planned to sell this to the Australian public and business community.

It was felt, however, that this would take an appreciable time to be included in the majority of addresses. So, in addition, a system of alpha extraction was developed. Under this plan, a number of the letters of the place name is extracted by the operator and coded on the keyboard. The problem here was how many elements should there be in the code; how many letters should be extracted from the place name. The number of letters extracted should be as low as possible consistent with an acceptable level of ambiguity. After an analysis of the names of all Post Offices in Australia, it was decided that a five element code consisting of the first, second, fourth and last two letters of the name gave an acceptable low level of ambiguity.

The memory used in the Australian system will therefore recognise either the appropriate numeric code of a place name or the alpha extraction of the place name. For example, if Tamworth in N.S.W. has the numeric code 2340, the code marks for 2340 and TAWTH are the same and therefore the letter will be directed to the same final despatch stacker.

CONCLUSION

There is no doubt that the system of mail handling installed in the Sydney Mail Exchange is one of the largest and most modern in the world. It provides management with a facility having considerable flexibility in the handling of traffic in that there are a number of differing or additional demands that can be readily satisfied to meet changes in traffic requirement. Naturally, it is only the first stage in the coding of mail and from its design and development, experience has been gained which will enable us to further advance towards our goal of providing the Australian public with a postal service of which all can be proud.

THE DEVELOPMENT OF THE A.P.O. MECHANISED MAIL HANDLING CONCEPT AND OVERSEAS TRENDS

V. St.G. MAGNUSSON, S.M.I.R.E.E.(Aust.)

The Australian Post Office concept for the design and application of machines for the processing of mail has been based on observations made over many years on methods of handling mail, the physical and other characteristics of articles and the labour components required in the various work areas. The most important factor influencing present day thinking is the rapidity with which mail processing costs are increasing.

Basically, the conditions under which mechanisation in the postal service becomes most efficient are when the desired standard of service to the community is provided with the greatest economy of operation. This broad statement assumes that full consideration has been given to such factors as staff comfort and the desire to eliminate burdensome tasks.

Insofar as the standard of service to the community is concerned, there could be two conditions under which an Administration accepts the principles of mechanisation. It could be that the present standard of service is satisfactory and that the reason for mechanisation is to reduce the costs of providing service; or it may be that mechanisation is necessary to upgrade the standards of service. There could conceivably be a further condition where it becomes necesary to re-adjust service in a downward direction. This would apply only in various parts of the total area over which the service is given and would only mean a reduction in service standards for certain groups which at present enjoy a higher standard than others.

This would suggest that service standards should be assessed and determined as an independent factor regardless of the method of handling mail. This in itself is a complex problem, having many facets, some of which are the determination of boundaries for country and metropolitan delivery; the form of inter-office transport employed; and the time of the day when the first letter delivery is to be effected. Once the standard of service to be provided is determined as a fundamental fixed condition to be met for a period of time, then all other calculations or comparative economic studies should be developed to meet this condition, whether the methods of handling remain manual or are mechanised. It is, of course, realised that variations to any so-called fixed service standard will occur from time to time, but that the major variations which are within the control of Administrations should, if possible, be planned to provide for a definite period of time.

The processing of mail in itself could be regarded as a relatively simple problem, but the operation tends to become complex when large volumes of mail are to be treated. For convenience of handling, articles of mail must be divided into several characteristic groups each requiring its own unique machine system. Processing must also be governed by the time elements which must be observed to provide service. There is also the problem of documentation and secondary activities associated with the handling of large volumes of mail.

It becomes obvious in any analysis of the ratio of staff to traffic load that a proportion of the operatives must be occupied on activities other than the prime process of sorting. It seems essential, therefore, in the attempt to reduce the cost of the sorting process, that we should also consider the labour expended in other areas. This then highlights the fundamental need for the co-ordination of the output of operatives engaged on the same sorting process and the automatic transfer of articles from process to process, particularly in large mail exchanges. It also points to the basic need to think in terms of machine systems rather than of machines. It is most unlikely that machines designed to process low volume flow can be adapted for use where peak hour volumes are considerable. However, the converse, that design should be based on the evolution of machine systems the elements of which could be utilised at centres of low volume peaks, could be true.

The second important conclusion is that design efforts should not contemplate the use of machines or systems which in themselves separate articles into many directions. If the natural evolution is through mechanisation towards automation, some balance should be made between the number of primary separations as against those required subsequently.

With few exceptions, the trend until recently has been to develop letter processing machines with a single, or at most, a limited number of associated operating positions. These may enable letters to be divided into a considerable number of directions but where two or more machines would be required to meet peak load conditions, there is no ready means for automatic co-ordination of outputs. (See Figs. 1, 2 and 3.)

It seems evident that the solution is to develop systems in which the coding desks are self-contained and detached from the sorting or decoding operation. This is the basis for development in West Germany where there is a system utilising 14 coding desks which feed to several decoders.

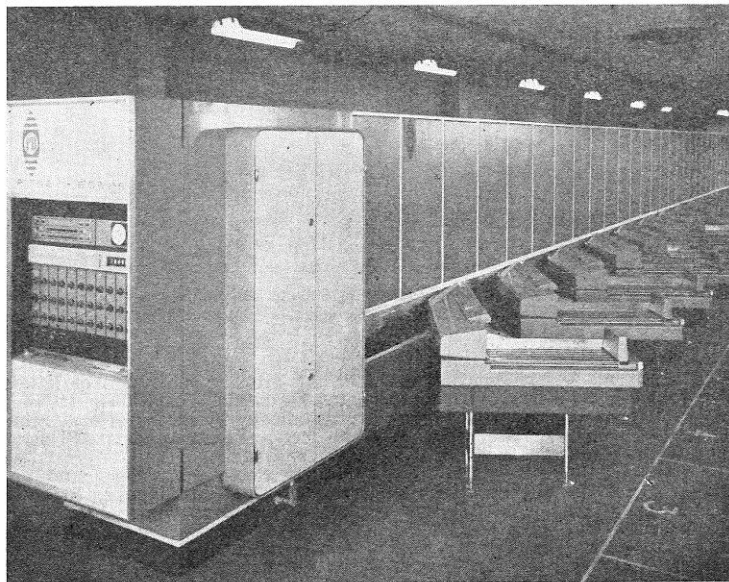* Mr. Magnusson is Deputy Assistant Director-General, Postal Services Division.

Fig. 1 — Sorting Machine of a Type Used by the U.S. Post Office. 12 positions — 300 separations. Manual feed to operators — manual clearing.
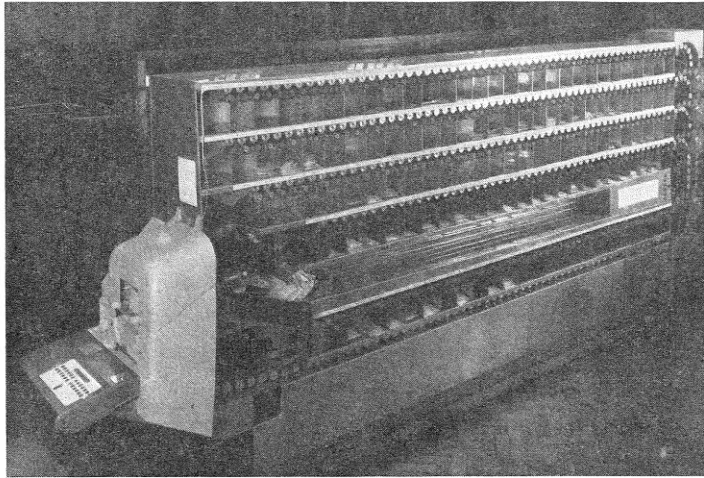
Fig. 2 — Sorting Machine of a Type Used by the B.P.O. Single position — 144 separations. Manual feed to operator — manual clearing.

Fig. 3 — Transorma Machine Manufactured in Holland. 5 positions — 300 or more separations. Manual feed to operators — manual clearing.

tion for a particular address is set up at the operator's keyboard. The translator transmits back to a printer the equivalent in machine language of the code which has been keyed. The printer impresses in bar form with phosphorescent ink the translated code on the back of the letter. The letter then passes through the appropriate primary channel to the decoder.

The Australian design for electronic reading and sorting machines (referred to as decoders) does not follow those of overseas countries, due to a distinctive approach to the overall problem. Whereas the tendency in other administrations is to develop machines which will divide letters into 100 or more directions in one pass, the A.P.O. only provides for 30, but, two or more subsequent runs may be made if required. Fig. 7 shows a decoder designed for the B.P.O. to give 144 separations.

The need for a relatively large number of separations at the decoding units becomes unnecessary when an automatic primary is associated with the coding operation. If the number is limited to 30, then designs can be simplified and high speed operation— a feature of most machines — can be avoided. Additional passes may be run through the machines installed in the Sydney Mail Exchange. However, 95% of letters can be separated into final despatches during the first pass, one further pass only being required on each decoder.

The A.P.O. concept has influenced the keying arrangement and code information printed on each letter. These tend primarily to specify the route a coded letter is to follow to reach its final destination point. Once a letter has been coded, further reference to a central memory is unnecessary even where the final electronic sorting operation takes place in any other State.

This, together with the specialised keyboard, gives a high degree of flexibility to management in that routing arrangements may be altered at will. Other groupings, although not at present envisaged, will surely be brought into being after traffic experience has been gained with the system. It is probable that within the near future, numeric codes will be developed for sorting to delivery rounds in capital city areas.

The design of machine systems for the sorting of non letter-form articles which consist mainly of rolled newspapers and packets referred to generally as other articles (O/A.) is also based on the use of a distribution network. The first O/A. machines were placed in service in the Sydney Mail Exchange prior to the year 1930. These have since been replaced by machines to the present design. The system installed in the Redfern building consists of a primary made up of three machines, the outputs of which are co-ordinated and automatically conveyed to four secondary machines

(See Figs. 4 and 5.) The British Post Office proposes to isolate the coding desks but has interposed an intermediate electronic separator between the desk suite and the decoders.

The A.P.O. letter sorting system is designed to accommodate 150 keyboard operators for the processing of peak traffic volumes in the vicinity of 300,000 letters per hour. The installation is capable of extension or modification to handle loads of greater magnitude. As a by-product of the coding operation, the system provides

for the primary separation of letters into 30 channels and their transfer automatically by means of a distribution network to decoders. Without further handling the letters pass through electronic readers and are sorted to final destinations. Refer to Fig. 6 which shows the primary separations effected at the primary coding positions.

Access to a central translator with capacity for the storage of over 12,000 place names is possible from any coding position when a code combina-
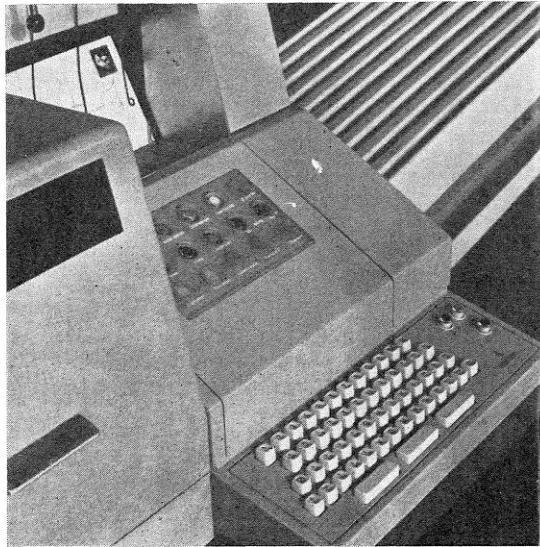
Fig 4 — Coding Desk. One of 14 used in conjunction with Decoders shown in Fig. 5. Automatic infeed and clearance.
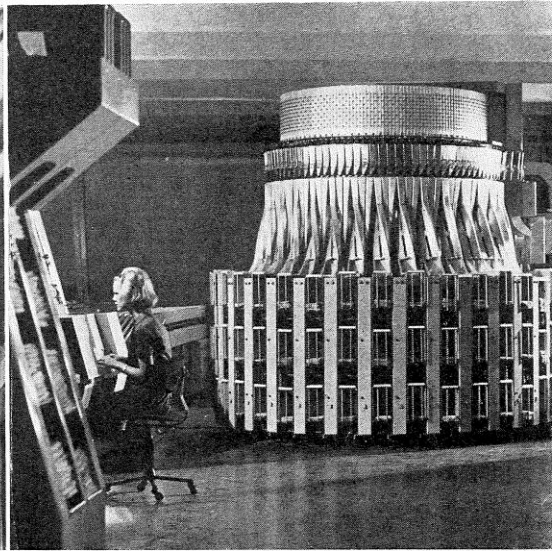
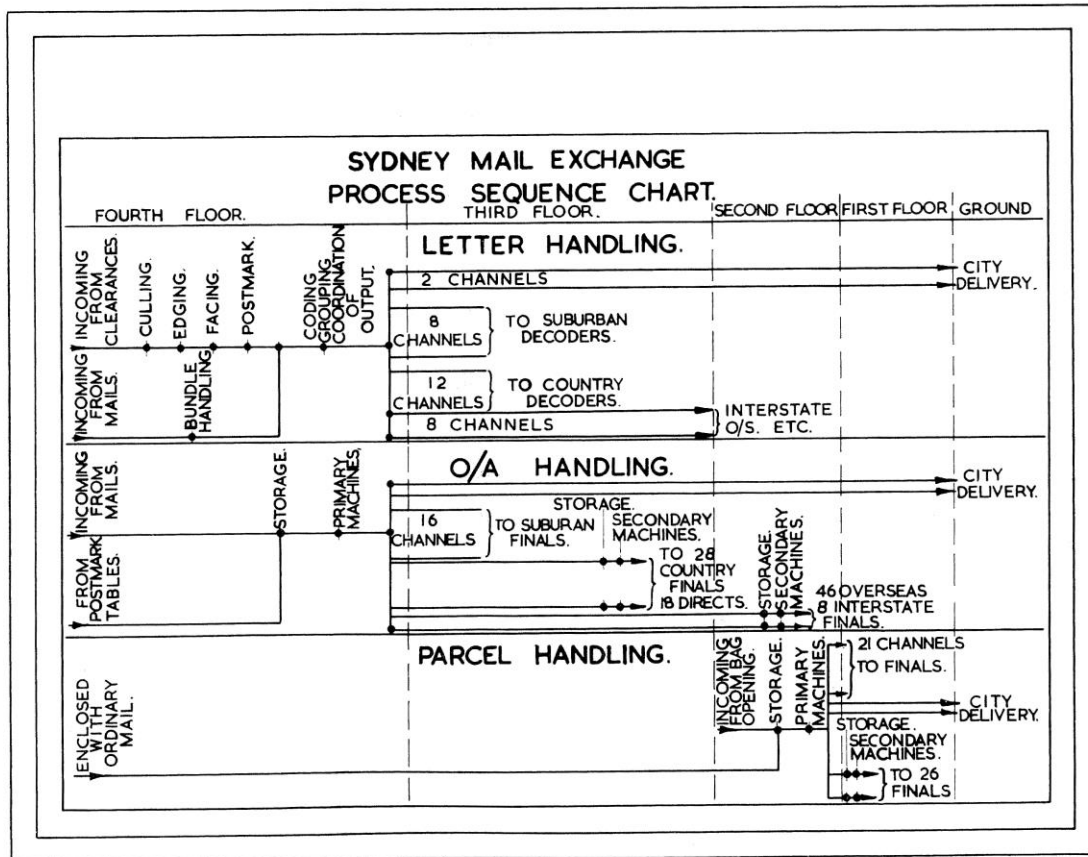Fig. 5 — Decoder Used in Conjunction with Coding Units shown in Fig. 4. Automatic infeed — manual clearance.
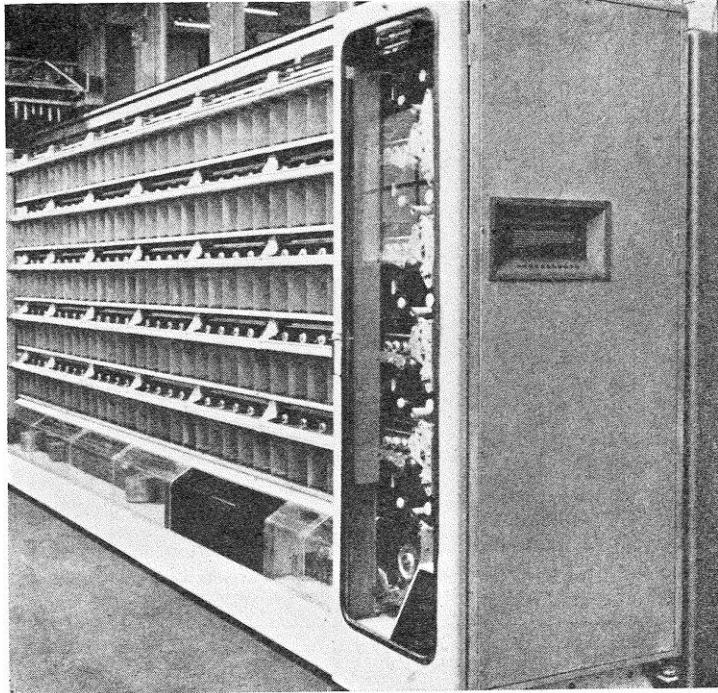

Fig. 6.

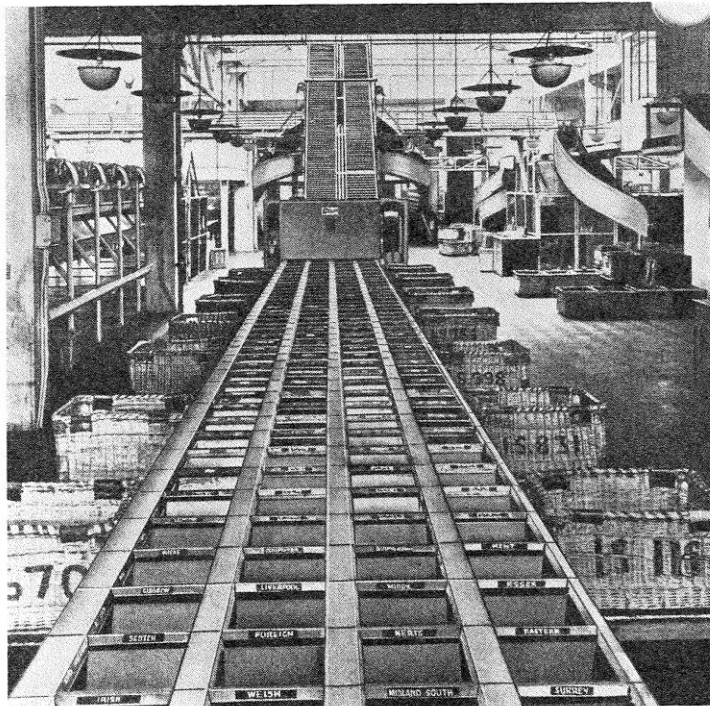Fig. 7 — B.P.O. type Machine Modified for Use as a Decoder.

or to storage hoppers. Articles are sorted manually at each stage. The only other machine of this type known to be in operation is that installed in the Mount Pleasant Sorting Office, London. (See Fig. 8.)

There are several types of machines designed for the sorting of parcels in use in other countries. Machines to the A.P.O. design, the basis of which has been adopted by the B.P.O. and the New Zealand Administration, have been in use for some years. The design of most machines developed in recent years is based on keyboard operation. This common approach has been influenced by the size and weight of the articles for which they are designed to sort. The Australian design is based on the sloping belt and drop panel method of diversion. The machine is keyboard operated and an electro-magnetic memory unit is used for the storage of pulses necessary for the sorting of parcels. There are several other methods whereby keyboard selections may be held in store to perform subsequent operations. The Australian designed machine can be manufactured at low cost and is simple to operate with an extremely low fault incidence. For an appreciation of machines in operation overseas, see Figs 9, 10 and 11.

It is thought necessary to refer also to one other machine system, which has been installed for the handling of bagged mail. It is usual for overseas administrations to use chain conveyors for the purpose. Whilst this method may be satisfactory for small or medium size sorting offices, the method is not suitable for a mail handling centre the size of that in Sydney, where considerable distance must be traversed with limited time available. A system utilising twin band conveyors for the elevation of bags and spiral chutes for descending bags has been installed. The outgoing bags are automatically distributed to 64 or more vehicular dock points.

The mail handling machine systems have only been referred to briefly as each has been dealt with in some detail later in this issue. Reference to equipment installed for the pre-treatment of letters and other articles will also be included with its appropriate handling system. There are, however, certain basic design aspects which should be discussed in a paper such as this, as they have been observed throughout the development of of the new systems, the outstanding fundamental being the co-ordination of output and the routing of mail.

The layout of the various working areas and of the equipment in each demanded considerable thought. Having regard to the total floor area in the Sydney Mail Exchange, the factors uppermost were:

(i) the transit time for mail passing between the various processes; and

(ii) the need to transfer operatives to other work areas to meet demands.

Fig. 8 — Machine Installed at the Mount Pleasant Sorting Office, London, for the Sorting of "Other Articles".
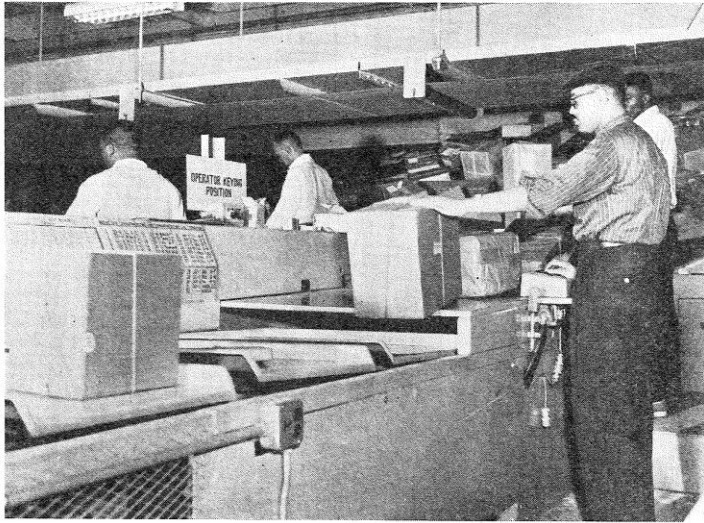
Fig. 9 — Parcel Sorting Machine — Using Tilted Tray Method for Parcel Diversion — U.S. Post Office.

Fig. 10 — Parcel Sorting Machine Using Sweeps for Parcel Diversion — U.S. Post Office.

It is essential that the total transit time for each processing line including the conveyance of bagged mail from and to the dock area be kept at a minimum. It was for this reason that the bagged mail handling system described in this issue was designed, in order that the average transit time for outgoing bags may be kept at an average of 90 seconds. It follows, therefore, that sequential operations should be adjacent or, where possible, in a vertical line between floors. Obviously, the length of time required to convey mail from point to point has a bearing on peak staff requirements. This factor must also be taken into consideration when determining equipment provision and the capital expenditure involved.

It should be mentioned, however, that peak period volumes are influenced by several other factors, some of which are under the control of management. In comparing peak demand, it must be borne in mind that, unlike automatic telephone exchange traffic, mail can often be readily held in store for varying periods without affecting service standards.

It is, for instance, important that machine designs be such that equipment may be readily installed in most buildings particularly existing sorting offices. This condition requires that equipment must be unit constructed and light in weight. Heavy construction is unnecessary, and costly, having regard to lightness of the loads the equipment is to carry. Furthermore, equipment designs should take into account the need for economy in the use of floor space. Considerable attention has been given to the appearance of machines and generally form has followed function. Surface finish has been given due importance, working surfaces are of plastic materials, and pleasing colour treatments have been used.

It is important, too, that designs take into account future developments not only to meet expanding traffic requirements but changes that may be dictated by progress in technology. Obviously a modern letter handling and distribution system should be designed so that letters, if electronically scanned and coded, may be inducted into the main flow. This is of particular significance with the A.P.O. concept where coding identifies the route as well as the Post Town.

Design characteristics must, of necessity, be such that maximum work outputs are possible. The study of human engineering problems is therefore important. The effort required to complete each unit of work should not contain avoidable expenditure of energy. For example, where the sorting operation is manual such as the sorting of O/A.'s, the optimum number of primary separations is 24. To go beyond this number would reduce the hourly rate of sorting and increase the possibility of mis-sorting.

1. INCOMING PARCELS FOR SORTING.

2. PARCEL STORAGE SPACES.

3. FEEDER BELT. (OPERATOR CONTROLLED).

4. SELECTIVE BUCKET CONVEYOR. (ONE PER SORTING POSITION).

5. BELT CONVEYORS TO SECONDARY SORTING OR BAGGING OFF. (EACH COLLECTING FROM ALL BUCKET CONVEYORS).
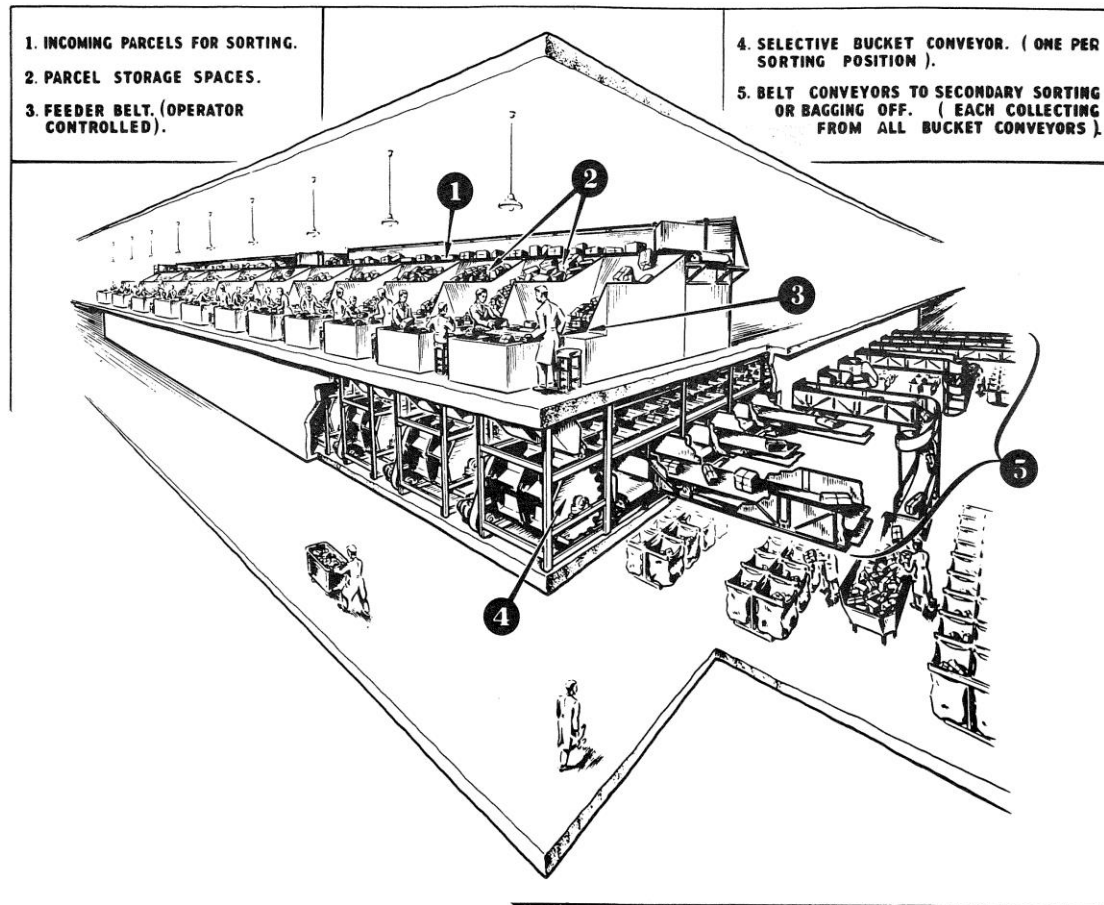
Fig. 11 — Artist's Conception of a Large Machine Installation for the Sorting of Parcels. A number of these Machines single or as a system have been installed by the B.P.O.

To reduce the number could require additional handling at a subsequent process.

Without doubt, the most difficult of design problems arises from the un-standard nature of articles passing through the mail exchange. Even the method of sealing letters or the tying of packets, etc., causes unpredictable vagaries in transit through machine systems. Perhaps the most important Australian development has been the sloping belt. By its use articles may be diverted from conveyors without the possibility of causing interruption to the normal mail flow. The sloping belt feature forms part of every machine system installed in the Mail Exchange. Before this development, sweeps and other methods were used for the diversion of mail flow. In other countries, even jets of compress-ed air have been used for the purpose.

The manner in which the sloping belt is used in each machine system will be described in the associated articles. The technical aspects of the sloping belt will, however, be given in the one describing the O/A. system, together with a reference to other methods used elsewhere.

An important fundamental feature of design which is of particular sig-nificance when entering into a rela-tively unknown field of development has been the avoidance of over sophistication. Where several tech-niques or methods have been devised to perform a function, the simple has been preferred, wherever possible, to the complex. This is confirmed by re-ference, for example to the sloping belt technique or the magnetic mem-ory device used with the parcel mach-ine system.

However, the need for machine systems to perform efficiently and with low fault incidence cannot be denied; circuitry and mechanical com-ponents must be designed with this in mind. They must take into considera-tion, too, the need for adequate con-trol and guard facilities. The provis-ion of these facilities must naturally have high essentiality for a system such as that installed in the Sydney Mail Exchange for the code sorting of letters, where the output capacity is dependent on links which may be common to a particular coding suite.

There are a number of refinements which could have been included with the first stage of the project to assist management. Whilst local super-visory control facilities have been in-cluded with each machine system, the ultimate will be the provision of a centralised supervisory control system so that traffic flow and loadings may be oversighted.

Future refinements must also envi-sage a centralised statistics centre

where volumes and numbers of articles passing through the Mail Exchange may be automatically recorded for examination or analysis. Access to reliable statistical information is essential to officers concerned with traffic management and staffing arangements.

Provision has been made in the circuitry of each machine system for the tapping of leaks for both supervisory control and the centralised recording of statistics. Electronically actuated units which can be fitted to conveyor runs have already been made for the weighing of mail matter in flow.

As the letter handling system required new methods of work processing, it was necessary to extend study into other technical fields. One interesting example is the application of the science of ergonomics to determine the location of the keyboard at coding positions. The keyboard is located at approximately waist level and is adjustable in both the horizontal and vertical directions. The operator's seat is fitted with a rest for the left arm only.

Many keyboard layout arrangements are possible. Some similar to typewriters fitted with up to 50 keys are in use overseas. A keyboard having only five keys has also been devised and used to provide for coding systems with limited application. A further variation is the use of two sets of five keys, one for each hand. This arrangement required a key in each set to be pressed simultaneously to register a character.

The development of a standard keyboard arrangement for A.P.O. use, was a most difficult problem, involving finally, tests to determine mental reflex actions of keyboard operators. These tests did not favour the use of keyboards where the operation required a finger of each hand to depress keys simultaneously, as even slight variations in timing caused errors. Such errors would increase where the keyboard is to be used for numeric as well as alpha codes.

The keyboard finally adopted as standard is based on one developed for use with tabulators and computers. On this particular keyboard, alpha characters are registered by depressing two keys simultaneously with the thumb and index fingure of the one hand, i.e., a one-handed keyer. To ensure efficient keying operation, the circuitry is not operative until the pressure is removed from the keys. By this means, tension on the operator has been removed. To ensure that touch coding commences early in the training period, the keys have not been individually designated. However, certain key groupings have been given a distinctive colouring for ready identification and as an aid to quick memorising of key locations. Fig. 12 shows the keyboard adopted by the A.P.O. Fig. 1 on Page 216 gives details of layout and usage.
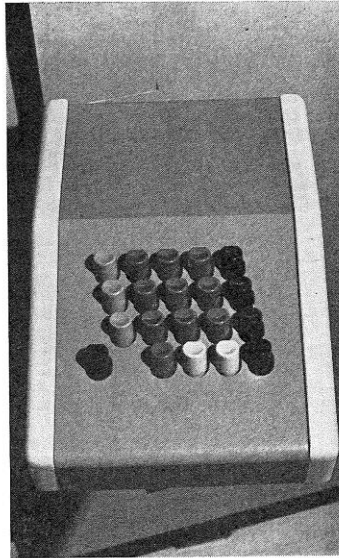

Fig. 12 — Keyboard Adopted by the A.P.O. for the Coding of Letters.

Again to ensure that there is no unnecessary mental strain or undue demand on operators, the development of machines which require the operator to work in rhythm with its movement has been avoided. Whilst earlier machines developed overseas, embodied this method, it has been generally discarded in later machine designs.

It is appropriate to make reference in this article to the code systems devised for use by the A.P.O., as it has had a direct bearing on the type of keyboard design. The development of a code is, of course, essential if electronic equipment is to be employed for automatic sorting. It provides a means whereby the keyboard operator may cause each envelope to be correctly code marked for subsequent sorting by an automatic process.

There are at least three methods of coding which may be devised for the purpose, i.e.:

(i) Extraction coding

(ii) Public Coding

(iii) Keyboard coding.

These methods, each of which is made use of by the A.P.O., are discussed in the following paragraphs.

Extraction coding is based on the selection of alpha characters from the destination name included with the address. The determination of the optimum code pattern is a difficult problem, as one of the disadvantages of extraction coding is the number of characters which must be taken from the place name if the number of ambiguities is to be kept within

reasonable limits. Ambiguity is a condition where the extracted characters and their sequence are the same for two or more localities. In Australia, it has been found necessary to extract five characters with the particular order of selection being the first, second, fourth and last two.

Several simple rules are necessary when using the code as some place names may not contain five characters, whilst others may consist of two words. Other rules are also necessary to avoid ambiguity. In total, a keyboard operator is required to memorise eight rules when using the extraction method for coding letters.

Public coding is a system which can be devised for use by the public when addressing mail. The value of public coding however depends on the degree of co-operation obtained from the users of the postal service. The A.P.O. has allotted a unique four-digit combination to each Post Town in the Commonwealth, the first digit representing the State in which the place name is located, i.e., 2 for N.S.W., 3 for Victoria etc. There are sufficient spare number combinations to meet development in New South Wales and Victoria for at least 20 years and for an indefinite period in other States. It is possible to extend the code to 5 digits should future development demand such a course. This could be done with no appreciable inconvenience to the user.

Numbers from the bracket (001-999) have been distributed to localities on a broad geographical basis. In New South Wales, for example, the following areas are observed:

Area	All N.S.W. number Combinations pre-fixed by "2"
Suburban	001 - 249
Country North	250 - 419
East	420 - 559
South	560 - 739
West	740 - 899
Spares	900 - 999

The adoption of the numeric code has provided an additional aid to the sorting process as digits can be keyed with one finger more readily than alpha characters. Futhermore, only the last three digits need be fed into the system when coding for intrastate Post Towns, the first digit only being required when coding for interstate destinations. A numeric code also provides an aid for the manual sorting of non-letter form articles.

Obviously, a public code could be made up of alpha characters or a combination of alpha and numeric digits. Codes of this nature would not meet the requirements of future planning, particularly with respect to the eventual introduction of automatic reading of addresses by scanning.

Keyboard coding requires the operator to memorise a code for each place name. This method is also used to a limited degree mainly to overcome the problems arising from ambiguities between place names where they occur. It could also be used for the numeric keying of letters to Post Towns receiving considerable volumes of mail. In such cases, the keyboard operator is required to memorise the allotted numeric code for each.

It will be seen that the A.P.O. has made use of the three recognised methods of coding, each for its particular purpose. Having in mind the need to ensure that maximum efficiency is gained from the letter coding and automatic sorting systems, it is essential that the alpha extraction method be used as a base. On this has been superimposed the simpler and more desirable numeric code method which depends for efficient use on the co-operation of the public. In its absence, extraction coding must be resorted to. The keyboard operators could, of course, be given the formidable task of memorising the allotted numeric code for every Post Town in the Commonwealth.

The Post Office recognises in the Commonwealth approximately 9,000

place names which may be included with addresses. All of these must be recorded in the memory system. In addition, all Post Towns of which there are 2,750 receiving direct despatches have been given a numeric code combination. These must also be included in the memory system.

It was the intention, when undertaking this exercise, to give, as far as possible, emphasis to the philosophy of approach to the problems of providing machine aids for the handling of mail. This philosophy cannot be expressed in few words as there are many associated aspects and fundamentals which must be observed. The ultimate in provision arises from efforts to meet the needs woven into the foregoing paragraphs. All these must be weighed and given their proper value in the final summation of effort, so that the result is the moulding of machine systems and traffic flow, capable of best meeting the needs of management. If this is done, the elements must live and extend to other sorting centres in the Commonwealth.

There is, of course, much more to be done and, in this era of technological development, it is obvious that we shall never be able to say "finish". It is essential however, that each step

we take is in an orderly progression towards a logical goal. Even during the course of the project, design features have been inbuilt to better meet that which is regarded as the eventual concept. For, in common with other projects of similar nature and magnitude, each step taken crystallises the way to the next.

REFERENCES

1.   G. P. Copping and H. J. Langdon "Sorting Letters by Machine"; Post Office Electrical Engineers' Journal, Jan., 1959.

2. R. S. Phillips and G. N. Davison, "Mechanising the Postal Service"; Journal of Inst. Elect. Engrs., Vol. 177, No. 8.

3. Otto Steiner, "Conveying Operations in Siemens Automatic Mail Handling System"; Siemens Review, Vol. XXX, No. 8.

4. V.St.G. Magnusson, "Planning a Modern Mail Handling Centre"; Union Postale Journal, Issue 11, November, 1965.

5. A. B. Corbett, "The Mechanised Handling of Mails"; The Australian Engineer, Dec. 1936.

The Sydney Mail Exchange Building showing the Cleveland Street frontage (right) and the Chalmers Street frontage.

A Cluster-Driven Energy Routing Protocol for Optimal Network Lifetime in Ad Hoc Networks

David Airehrour
Unitec Institute of Technology, Auckland, New Zealand

Marianne Cherrington
Otago Polytechnic, Auckland, New Zealand

Samaneh Madanian
Auckland University of Technology, Auckland, New Zealand

Abstract: Mobile Ad hoc Network (MANET) is a group of networked mobile devices working in a cooperative manner. Due to usage, these devices quickly run out of battery power in critical situations and consequently fail in packet transmission. Power-Efficient Gathering in Sensor Information Systems (PEGASIS) and Dynamic State algorithm are two research works with unique energy efficiency concepts that, if harmonized and refined, will deliver a better optimized energy-efficient routing protocol for MANETs. This study, therefore, proposes PEGADyn – a hybrid version of PEGASIS and Dynamic State algorithm for a new energy-efficient routing protocol in ad hoc networks. PEGADyn creates a virtual grid classification of nodes based on current location, followed by a cluster formation of nodes in each virtual grid created. In each cluster and virtual grid formed, cluster heads (CHs) and designated cluster heads (dCH) are selected based on their energy levels. CHs and dCHs are used for communication between clusters and virtual grids. The use of CHs and dCHs limits communication overheads among nodes, reducing the energy expended and increasing the network lifetime. A simulation comparison of PEGADyn with PEGASIS and Dynamic State shows PEGADyn to be better in extending network lifetime and maintaining network throughput.

Keywords: Ad Hoc, Cluster Head, Energy, PEGASIS, Self-Organising.

Introduction

A Mobile Ad hoc NETwork (MANET) is a wireless communication network system that is comprised of mobile devices like laptops, mobile phones, and drones, amongst others. These mobile devices are commonly referred to as nodes and they have the ability to communicate with other nodes without requiring any intermediate network infrastructure; hence, they are considered self-configuring or self-organizing networks (SON), which are easily deployable,

that are capable of self-directed operations. Generally, wireless networks can be classified into four categories: Mobile Ad hoc Network (MANET), Wireless Sensor Network (WSN), Wireless Mesh Network (WMN), and Vehicular Ad hoc Network (VANET). In MANETs, nodes work together to deliver network access and resources without centralized management (Ilyas, 2003). Nodes in MANET are usually connected within a defined, but limited distance coverage area and they can freely join an ad hoc network, move within the network and leave the ad hoc network. A unique feature in MANETs is the ability to incorporate MANET nodes into cellular wireless networks. Incorporating MANET nodes into a cellular wireless network allows nodes to communicate directly with other nodes without going through base stations.

MANETs find their usefulness in areas where nodes are mobile and there is no pre-existing communication or network infrastructure, like military battlefield networks, collaborative robotic mobile networks and vehicular networks (VANET). Other application areas include where pre-existing communication infrastructure has been destroyed, as in disaster relief networks and sensor networks; or areas where it is not feasible for either financial reasons, physical constraints or mobility to use wired networks. A typical application scenario of MANET can be seen during the rescue of 12 teenage Thai footballers and their coach in a cave. This was an operation that involved about 10,000 workers working in an ad hoc manner (Smith, 2018). For proper communication and coordination of activities, a MANET network infrastructure becomes necessary. In another circumstance, the Carr wildfire of California (Morris, 2018) involved more than 27,000 fire fighters and logistic workers, some of whom were from other nations. For effective communication and coordination of activities, MANET network infrastructure also becomes vital.

For any MANET to be effective, a routing protocol is required for nodes in a network to communicate and create routes for data transmission. These paths are logically structured so that the data packets can travel over multiple hops to get to the destination node (Pattanayak et al., 2011). Routing is the process of creating a path in a network by using some predefined metrics with the aim of transmitting data from source to destination. The routing process is an important layer in the Open System Interconnect (OSI) reference model, and it is the nucleus to the proper functioning of any multi-hop network system like MANET. However, the design of an efficient routing protocol is confronted by several challenges due to the constrained resources of MANET devices, like limited memory, processing power, battery life and bandwidth (Pattanayak et al., 2011). Some MANET devices (nodes) die out during operation due to limited battery life, and getting a charging system in an ad hoc situation could prove challenging. It becomes imperative to design an energy-efficient MANET routing protocol that could provide a longer lifetime for MANET nodes. Therefore, this paper addresses this research gap.

The rest of the paper is organized as follows. A review of related energy-efficient protocols for ad hoc networks is discussed with a focus on the Power-Efficient Gathering in Sensor Information Systems (PEGASIS) and Energy Efficiency Dynamic State (EEDS) algorithm. Further, we propose and present our hybrid model, PEGADyn. Through NS2 simulation we demonstrate that our protocol (PEGADyn) provides better network lifetime, throughput ratio and packet drop rate compared with PEGASIS and EEDS.

Energy-based Routing Protocols in MANET

In MANETs, different metrics for finding routes from source to destination have been proposed (Boukerche *et al.*, 2011; Aziz & Al-Akaidi, 2007; Son *et al.*, 2014). MANET algorithmic routing methods include link-state and distance-vector algorithms. However, in this study, an exploration is undertaken of power-efficient protocols that could improve the battery life of nodes while providing good network throughput.

Low Energy Adaptive Clustering Hierarchy (LEACH)

Heinzelman, Chandrakasan & Balakrishnan (2000) presented a Low Energy Adaptive Clustering Hierarchy protocol, also referred to as the LEACH protocol. LEACH is a self-configuring and self-organizing hierarchical routing protocol. It uses a clustering system which uses a random-based mechanism for the selection of local cluster-heads that uniformly apportion the energy cost among network nodes. The LEACH algorithm process is divided into *rounds* (time steps performed during packet transmission) of configuration and each round consists of two phases. The first phase is the Cluster setup, while the second phase is the Steady-state phase. A node sends its data to a nearby cluster head, which in turn compresses the data and forwards the data to a base station. A fundamental assumption in LEACH is that the radio strength of each node is enough to reach any base station or at least to a nearby cluster head (CH). In the LEACH algorithm, a voting process occurs for the selection of a cluster-head at a specific time period. This is necessary to maintain a state of energy equilibrium among the CHs in the network. The voting process in LEACH is a weakness, as this causes additional energy and network overhead to be spent in the process.

Power-Aware Ad hoc On-Demand Distance Vector Routing (PAAODV)

Pattanayak *et al.* (2011) proposed a Power-Aware AODV (PAAODV) protocol, which is a power consumption improvement to the AODV routing protocol. The concept behind PAAODV is the regulation of the transmission power of nodes in the network to extend network lifetime and sustain network connectivity. PAAODV uses a node's transmission power level to regulate the number of packets transmitted. The researchers submitted that regulating the transmit power of a node conserves a node's energy and thus extends its network lifetime. They only compared

the performance of PAAODV with AODV and other non-energy-based protocols. However, a true justification for the authors' claim of improved network lifetime will be a benchmark against similar energy-based routing protocols, which was not provided.

Energy Efficiency Dynamic State (EEDS)

According to Gautam *et al.* (2016), the lifetime or energy of nodes in a network can be improved and maintained using the concept known as Energy Efficiency Dynamic State (EEDS). This concept uses three states, namely: the active state; the night state; and the discovery state. Nodes under EEDS are required to be in one of these states, while only nodes in the active (and ready) state can be considered for routing. Nodes in the night state are either depleted or recouping their energy. In EEDS, nodes are grouped into a predefined number of virtual grids based on their current positions. Nodes in the active state are involved in routing functions. Nodes in a grid can often transfer their tasks to other nodes in the active state within their own virtual smart grid once they observe their energy levels to be critical or depleted to 10%. Nodes with critical energy level then begin to transit into the night state where they are not involved in routing or any other tasks, in order to recoup their energy. Eventually, these nodes recoup their energy and thus move to the discovery state, where they are now ready to participate in routing tasks.

Table 1 provides a summary of some cluster-based protocol schema used in ad hoc networks. The table shows parameters used for CH selection based on CH identity (ID), permissibility of node to overlap between clusters, the density of nodes per cluster and the frequency of CH change.

Table 1. Energy-aware and Cluster-based Protocols. Adapted from Udaykumar & Thirugnanam (2015).

Clustering Schema	Based on	CH's Selection	Overlap Cluster	Number of Nodes	CH Change
Linked Cluster Algorithm (LCA)	Neighbour Node	Lowest ID	Possible	High	Low
Heuristic Based Algorithm	Neighbour Node	Node ID	No	Low	Low
High Connectivity Cluster	Neighbour Node	Lowest ID	High	Very Low	Very High
Power-Efficient Gathering in Sensor Information System (PEGASIS)	Energy Node or Threshold Node	Node ID	Possible	Moderate	Low
α – Stability Structure Clustering	Neighbour Node	Highest ID	Low	Moderate	Low
Novel Cluster Algorithm	Mobility Node	Lowest ID	Low	Moderate	Very Low

Power Efficient Gathering in Sensor Information System (PEGASIS)

To provide a significant refinement to the LEACH protocol, Lindsey & Raghavendra (2002) proposed a Power-Efficient GAthering in Sensor Information Systems (PEGASIS). The focus of PEGASIS is to organise the nodes so that each node takes its turn to receive and transmit packets to some nearby neighbour nodes. The collected data passes from one node to another, and ultimately a designated node transmits the collected data to a base station. In PEGASIS, nodes take their turn to transmit packets to the base station with the aim of minimizing the average energy spent by each node in transmitting from source to destination and back again (round trip time). In contrast to LEACH, the PEGASIS algorithm shows inherent benefits. First, it performs a localized node gathering where the distance covered to transmit data to the destination is much shorter in comparison to the LEACH protocol. Secondly, the number of rounds (time steps performed during packet transmission) made per node during packet transmission to the base station is kept minimal. Thirdly, since the selection of CH is made once, the probability of node overlap among cluster groups is eliminated.

There are, however, variants of the PEGASIS protocol and these are discussed below.

1. Improved PEGASIS Routing Protocol Based on Neural Network and Ant Colony Algorithm (ACON-PEG). ACON-PEG embeds the ant colony algorithm and neural network instead of using the greedy algorithm to build the network routes. ACON-PEG subdivides the route coverage area into multiple, but equal parts and creates a chain that makes the route path evenly spread while reducing the square of the total transmission radius. The protocol utilises the concept of a neural network algorithm to choose the chain head and applies the ant colony algorithm to discover the optimal path for sending data to the base station. Li *et al*. (2015) claim that, in comparison with PEGASIS by Lindsey & Raghavendra (2002), ACON-PEG achieves better route optimization.

2. Improved Energy-Efficient PEGASIS-Based Protocol. The Improved Energy-Efficient PEGASIS-based protocol (IEEPB) was proposed to address the problems of Energy-Efficient PEGASIS-Based protocol (EEPB), which was challenged with the sub-optimal election of leader node and long links of chain nodes. Sen, Bing & Liangrui (2011) use a weighting scheme to select the leader node by assigning every node a weight that demonstrates the node's suitability to become a head/leader. The weight is based on a node's properties, including its residual battery power and the distance between the node and the base station. The paper claims that IEEPB has higher energy efficiency and, hence, longer network lifetime.

PEGADyn: Hybridizing PEGASIS and Dynamic State Routing Algorithms

Our study into the PEGASIS algorithm shows it has some features that make it unique, although there are some aspects wherein energy is being wasted. Similarly, the virtual classification of nodes into grids in the Dynamic State (EEDS) algorithm provides an efficient topological structure for an optimized energy efficiency implementation. This study, therefore, explores and amalgamates the unique features of PEGASIS and EEDS to develop a new hybrid optimized energy-efficient algorithm called PEGADyn.

In PEGADyn, a virtual grid classification of nodes based on their current location is generated, and this is followed by a cluster formation of nodes in each virtual grid created. A virtual grid has a default radius of 50 metres, while a default cluster radius is 25 metres. In each cluster formed, a cluster head (CH) is selected based on the node with the highest energy level. The node with the second highest energy level is selected as the associate cluster head (aCH). Where there is a tie in energy level, an arbitrary selection is made by the algorithm.

Furthermore, to aid inter-virtual-grid communication, designated CHs (dCH) are selected from among CHs in each virtual grid. The selection of a dCH is a combination of minimal mobility of a CH node, and the CH with the highest energy level in a virtual grid. Thus, a dCH also doubles as the CH for its own cluster group. A dCH transmits packets for the cluster groups in its virtual grid. Inter-virtual-grid communication ensures that CHs communicate only with clusters in their own virtual grid. For communication with clusters or nodes outside their virtual grid, a node sends its packet to a CH it is associated with, and the CH in turn sends the packet to a dCH that routes the packet to another dCH (in a different virtual grid) that can fulfil the route request.
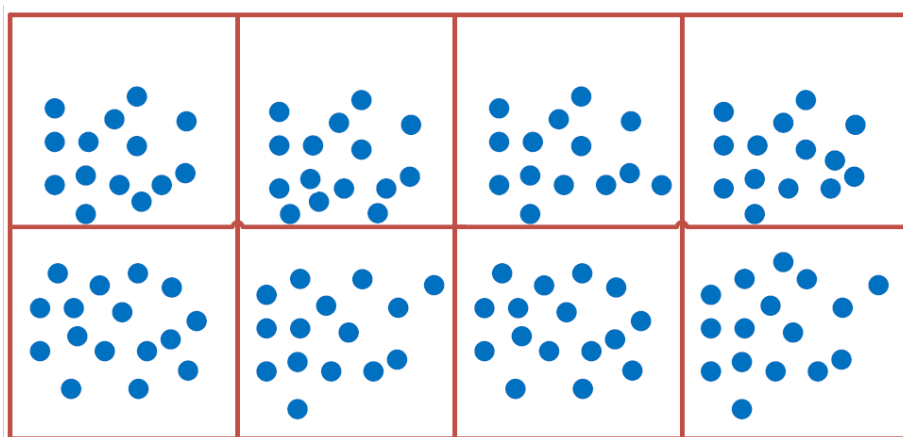

Figure 1. Virtual grid classification of nodes based on proximity.

All CHs advertise their route details to their cluster members and each node member aligns and sends its packet via the CH. Every CH also advertises its route details to other CHs. This

way, a source-to-destination path is formed from one end node to another. Figure 1 shows the virtual grid classification based on proximity of nodes. Figure 2 shows the final cluster and network formation of nodes and inter-virtual-grid routing communication.
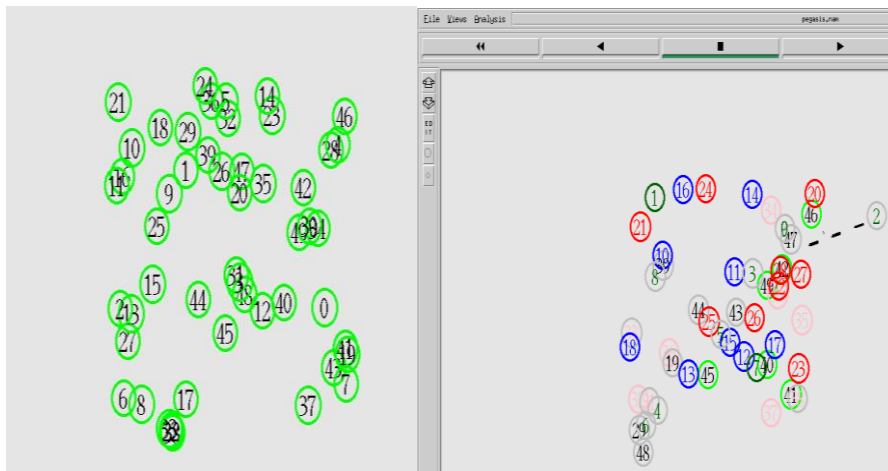

Figure 2. (a) Node deployment; and (b) Cluster formation and CH selection.

Since nodes are mobile within the network, when a node's proximity threshold is reached (default is 50 metres), it switches its grid and cluster to join a new virtual grid and cluster. In this respect, a node identifies with a cluster group in order to identify a CH that it could use to route its traffic. A summary of the procedure for CH, aCH and dCH selection is given in Figure 3. When a CH is depleted in energy or moves away to another cluster, it becomes an ordinary node in the new cluster it joins and aligns itself with the CH in that cluster for packet routing. However, if there is no CH in that cluster, the node assumes itself as the CH and advertises its details to proximity nodes to form a cluster. The classification of the virtual grid is to create a proximity boundary, so that only two CHs communicate across the virtual grids. While CHs within a virtual grid can communicate as they want, inter-virtual-grid communication can only be performed by designated CHs (dCHs).
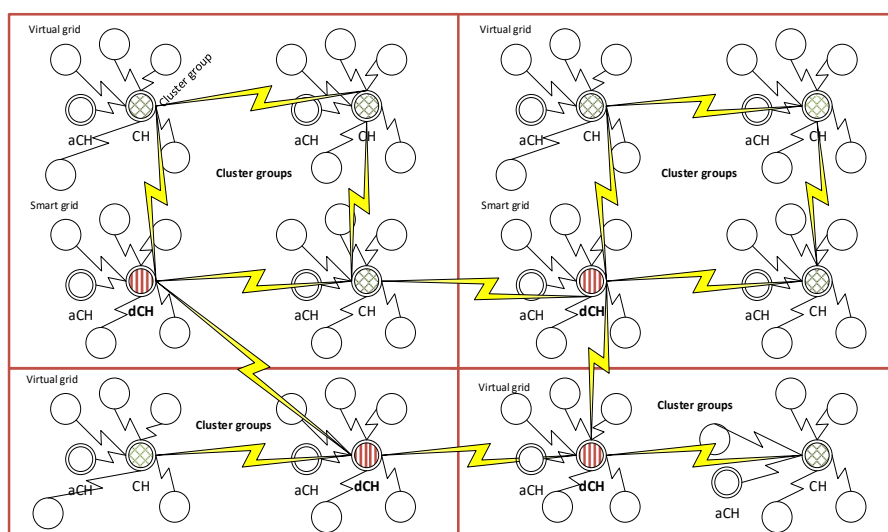

Figure 3. Virtual grid, cluster formation and CH selection in PEGADyn.
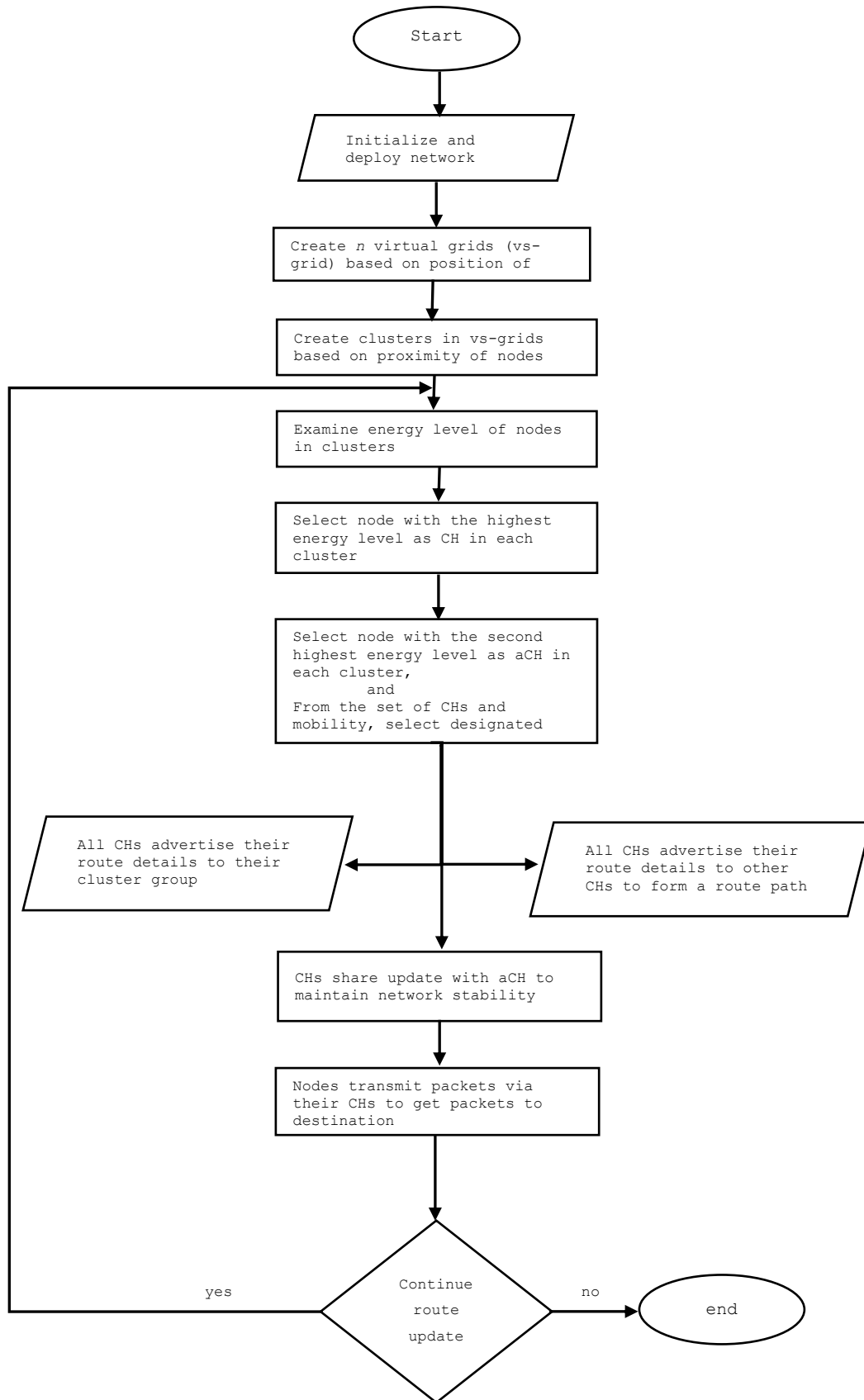
Figure 4. Flowchart for the creation of virtual grids and clusters; and selection of CHs, aCHs and dCHs selection.

The deployment of nodes and the creation of virtual grids and clusters is shown in Figure 4. At the start of the flowchart, the deployed nodes are initialised. This is followed by the

formation of virtual grids (vs-grid) based on the positions of the deployed nodes. The virtual grids aid inter-virtual-grid communication.

Further to that, cluster formations are created within each virtual grid. This is to facilitate inter-nodal communication between CHs and nodes within each cluster. This phase forms the initial network topology formation.

Next, nodes within each cluster examine their energy levels to elect a CH and an aCH. The node with the highest energy level within a cluster is elected the CH, while the node with the second highest energy level is selected as the aCH. Now, within every virtual grid (vs-grid) a dCH is selected, which is a CH with the highest energy level. The dCH facilitates communication between virtual grids in the network. All CHs advertise their route details to the nodes within their cluster groups and to other CHs to form route paths to other virtual grids. To maintain network stability and to avoid becoming depleted of energy during the routing process, every CH shares network updates (synchronises) with its aCH. Once network convergence is achieved, nodes are able to transmit their packets via their CHs to other nodes within the cluster. However, packets destined to other virtual grids are sent from a CH to a dCH, which transmits the packet to a counterpart dCH that delivers the packet to the destined node within its cluster or forwards it to another dCH for onward delivery.

Energy efficiency plays an important role in ad hoc networks hence, the core of PEGADyn is to minimize the power consumption of nodes during network transmission. It achieves this by reducing inter-node communication by assigning the responsibilities to designated nodes at cluster and virtual grid levels. By this communication system, nodes take turns to become CHs, aCHs and/or dCHs to balance the energy requirement among all nodes. This way, the network lifetime is extended, and throughput is not compromised.

Our proposed system was compared through simulation with PEGASIS and the Dynamic State algorithm. The comparison was based on energy consumption, residual energy and network throughput.

Simulation and Results

The simulation setup was performed using NS2 version 2.34. No route overhead has been considered in the simulation because of the link stability and the route lifetime. The total coverage area was set at 500m x 500m, which is a realistic space to cover for an emergency or a disaster scenario. The NS2 simulator provides support for the multicast routing protocols and the IP protocols, such as TCP and UDP. A total of 50 nodes were deployed and each node has a transmission range of 100 metres; a mobility speed of 5 metres per second was assumed.

Multiple simulation runs were performed, and the average results were computed and presented. A summary of network simulation parameter settings is presented in Table 2.

Table 2. Network Simulation Parameters

Simulation Settings	
Simulation platform	NS2 v2.34
Coverage Area	500m x 500m
Antenna Type	Omnidirectional
Node Transmission range	100 metres
Node Speed	5 m/s
Number of Nodes	50
Traffic Type	UDP/CBR
Routing Protocols	EEDS, PEGASIS and PEGADyn
Simulation Time	500 seconds

Figure 2(a) shows the initial deployment of network nodes while Figure 2(b) shows cluster formation and the selection of CHs. CHs have been selected based on the energy of the nodes in the network. The node with the highest energy is selected as the CH for every cluster. Nodes of the same cluster group have the same colour (see Figure 2(b)). Hence, nodes in yellow belong to the same cluster, while nodes in red are in the same cluster group.

Power Consumption

Power consumption is the total amount of energy consumed by a node during packet transmission and reception. Figure 5 shows the graph of energy consumed (Joules) by PEGADyn is less compared to PEGASIS and Dynamic State. This supports our point that our proposed PEGADyn protocol has better energy efficiency in comparison to PEGASIS and Dynamic State algorithms.
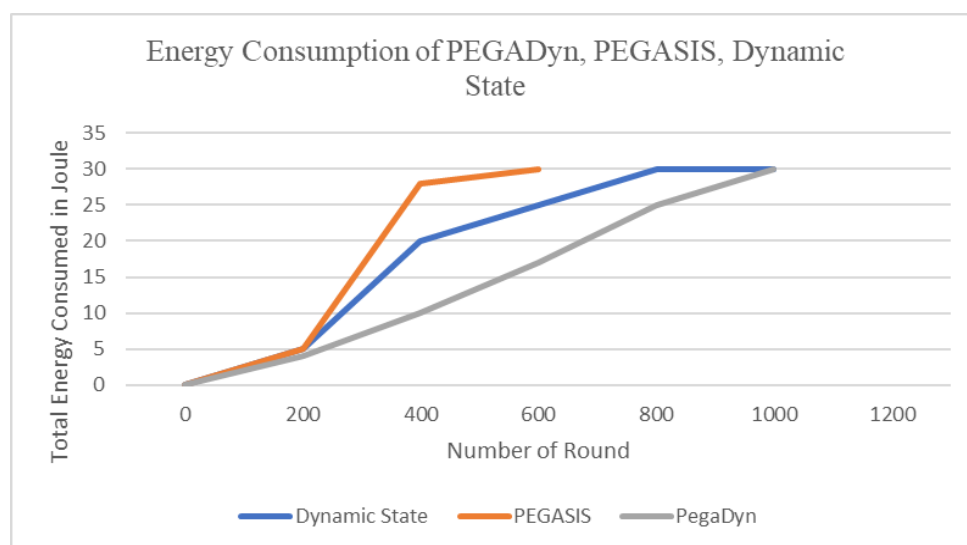

Figure 5. Energy consumption comparison.

The virtual grid classification of nodes was a feature in Dynamic State algorithm, which was incorporated into PEGADyn and contributed to the improved energy efficiency of PEGADyn.

The virtual classification provided an efficient structure of how each node could transmit its packets in the network, thus limiting excessive node communication and transmission within the network. As can be observed in Figure 5, the Dynamic State algorithm also had better energy savings over PEGASIS.

Minimum Residual Energy

The minimum residual energy indicates the overall lifetime extension of a node in an active state (Valikannu, George & Srivatsa, 2015). From our simulation results and as shown in Figure 6, PEGADyn had a higher amount of residual energy than PEGASIS and Dynamic State, an indication that nodes under PEGADyn experienced longer network lifetime. This is attributed to the classification of nodes into virtual grid and cluster groups, and the use of dCHs for inter-virtual-grid communication among cluster groups to minimize energy consumption. In addition, because every node takes its turn to become either a CH or dCH, this evenly distributes the energy consumption across all nodes in the network.
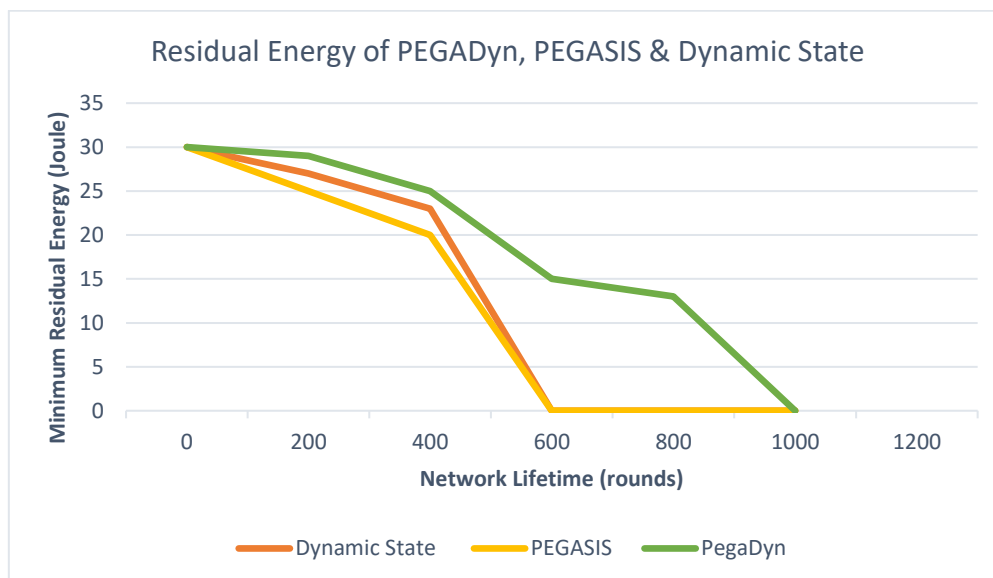
Figure 6. Residual energy comparison among protocols.

Throughput Ratio

The throughput ratio is the amount of data received at a node's destination during a time period, which comprises host overhead and contention on the media link. Figure 7 shows the ratio of packets delivered by all three protocols. PEGADyn had consistently higher network throughput ratio than PEGASIS and Dynamic State protocols throughout the simulation period.
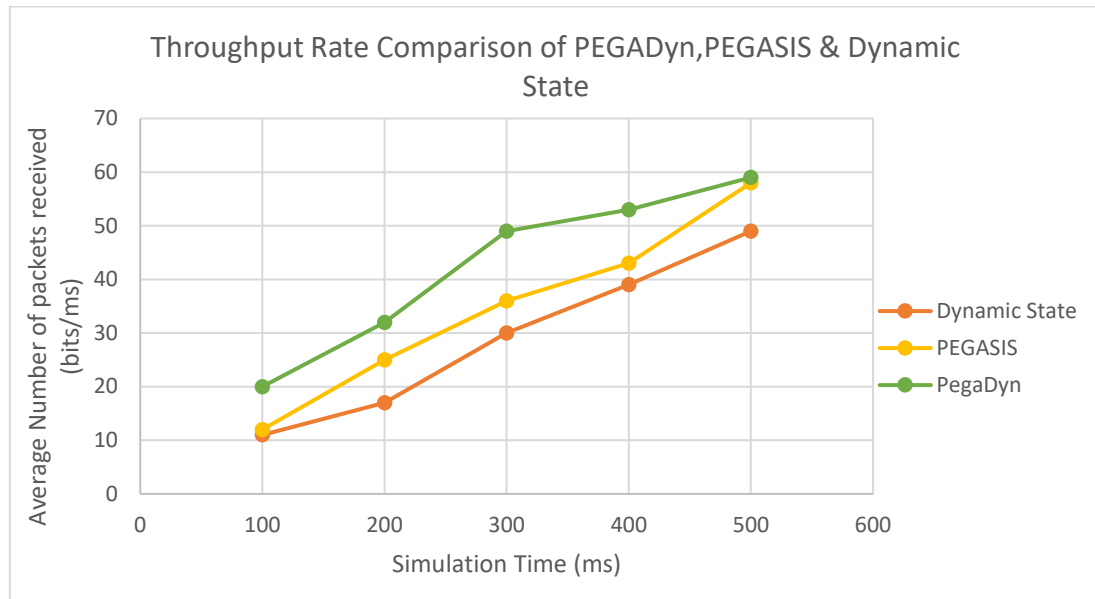
Figure 7. Throughput rates of PEGADyn, PEGASIS & Dynamic State.

Packet Drop Ratio

The Packet Drop Ratio shows the rate of dropped/lost packets to the total amount of packets sent. Every transmitted packet has a time-to-live period (TTL) in order to get to its destination (Das & Tripathi, 2018). Upon expiry of this time period and the packet's inability to get to its destination (due to delay or collision etc.) the packet is dropped. Figure 8 shows PEGADyn having a better (lesser) packet drop rate than PEGASIS and EEDS. Over the simulation period, PEGADyn had drop rate that peaked at 42%, while PEGASIS's drop rate peaked at 60%. However, EEDS had a higher drop rate of packets peaking at 72%. This clearly demonstrates PEGADyn as a better protocol in comparison to PEGASIS and EEDS.

Figure 8. Packet Drop Ratio with Time.

Simulations Study Analysis

To establish the accuracy of the simulation experiments performed in this research, the Confidence Level (CL), Standard Deviation, Sample Variance and Standard Error of measurements for statistical means were applied to the mean of generated simulation data. The CL presented in th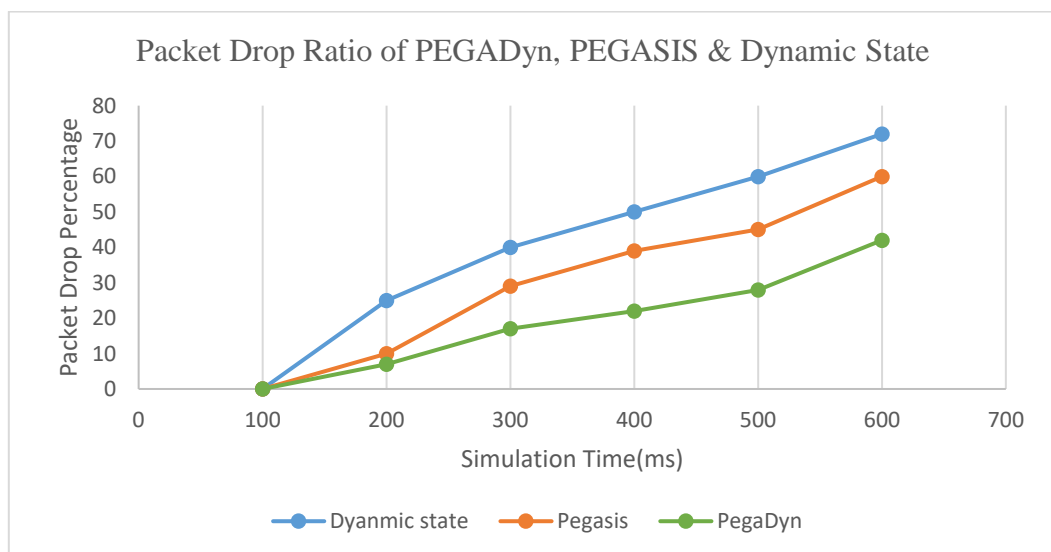is study is based on the mean of the data set, given the sample's size of 200 simulation runs. The simulation result presumes a Normal distribution from the sample data set analysed, and a 95% confidence level. A 95% confidence level indicates that 95% of our sampled data set contained the population mean with just 5% of our data set possibly outside the population sample.

Table 3 presents the statistical summary of the residual energy per round for all three protocols. The CLs for PEGADyn, PEGASIS and EEDS, which are 0.063, 0.062 and 0.065, respectively, reflect good confidence levels of generated data from simulations performed. In addition, the standard error values for PEGADyn (0.028), PEGASIS (0.027) and EEDS (0.029) and the standard deviations for PEGADyn (0.089), PEGASIS (0.087) and EEDS (0.091) show negligible sampling error.

Table 3. Statistical Summary of Residual Energy Per Round

Description	PEGADyn	PEGASIS	EEDS
Mean (residual energy per round)	0.070	0.055	0.058
Standard Error	0.028	0.027	0.029
Standard Deviation	0.089	0.087	0.091
Sample Variance	0.008	0.008	0.008
Confidence Level (95%)	0.063	0.062	0.065

Furthermore, Table 4 shows the statistical summary of the throughput ratios of PEGADyn, PEGASIS and EEDS. The standard deviation and standard error of measurement of sampled data for all three protocols showed insignificant error margins – PEGADyn (SD=0.089, SE=0.028), PEGASIS (SD=0.087, SE=0.027), EEDS (SD=0.091, SE=0.029) – an indication of consistency of data sampled across all three protocols. In addition, the CLs for all three protocols, including PEGADyn (0.020), PEGASIS (0.005) and EEDS (0.005), further attest to the high confidence of our sampled data.

Table 4. Statistical Summary of Throughput Ratio

Throughput Rate	PEGADyn	PEGASIS	EEDS
Mean (throughput ratio)	0.154	0.118	0.098
Standard Error	0.009	0.002	0.002
Standard Deviation	0.027	0.006	0.006
Sample Variance	0.001	0.000	0.005
Confidence Level (95.0%)	0.020	0.005	0.005

Finally, the statistical analysis of the packet drop ratio in Table 5 shows that the standard deviation and standard error of measurement of sampled data for all three protocols also show insignificant error margins – PEGADyn (SD=0.019, SE=0.006), PEGASIS (SD=0.031, SE=0.012), EEDS (SD=0.039, SE=0.012) – an indication of consistency of simulation data across all three protocols. In addition, the CLs for all three protocols, including PEGADyn (0.013), PEGASIS (0.021) and EEDS (0.026), further attest to the high confidence of our sampled data.

Table 5. Statistical Summary of Packet Drop Ratio

Description	PEGADyn	PEGASIS	EEDS
Mean (throughput ratio)	0.046	0.073	0.106
Standard Error	0.006	0.009	0.012
Standard Deviation	0.019	0.031	0.039
Sample Variance	0.000	0.001	0.002
Confidence Level (95.0%)	0.013	0.021	0.026

Conclusion

In conclusion, this study has presented a hybrid version of PEGASIS and Dynamic State routing protocols to form a new energy efficient routing protocol, PEGADyn. PEGADyn minimizes energy consumption among nodes by creating virtual grids and cluster heads in order to form a hierarchical route path that is energy efficient and balances the energy requirements among nodes in the network. Results from our simulation using NS2 have shown that PEGADyn offers a more promising energy efficient routing protocol for ad hoc networks, especially in extending the lifetime of nodes within the network.

References

Aziz, M. & Al-Akaidi, M. (2007). Security Issues in Wireless Ad Hoc Networks and the Application to the Telecare Project, in *Digital Signal Processing, 15th International Conference on Digital Signal Processing*, Cardiff, pp. 491-494.

Boukerche, A., Turgut, B., Aydin, N., Ahmad, M. Z., Bölöni, L. & Turgut, D. (2011). Routing protocols in ad hoc networks: A survey, *Computer Networks*, *55* (13), 3032-3080, September.

Das, S. K. & Tripathi, S. (2018). Intelligent energy-aware efficient routing for MANET, *Wireless Networks*, *24* (4), 1139-1159. DOI: 10.1007/s11276-016-1388-7

Gautam, J., Fathima, B. L., Sangeetha, K. & Muzammil, P. M. (2016). Energy Resource Optimization in Wireless Ad-hoc Network Using Dynamic States, *International Conference on Engineering and Technology Systems*, *13* (2), 57-61.

Heinzelman, W. R., Chandrakasan, A. & Balakrishnan, H. (2000). Energy efficient Communication Protocol for Wireless Microsensor Networks, *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, Maui, HI, USA.

Ilyas, M. (2003). *The Handbook of Ad Hoc Wireless Networks*. Boca Raton: CRC Press.

Li, T., Ruan, F., Fan, Z., Wang, J. & Kim, J. (2015). An Improved PEGASIS Routing Protocol Based on Neural Network and Ant Colony Algorithm, *International Journal of Future Generation Communication and Networking*, *8* (6), 149-160.

Lindsey, S. & Raghavendra, C. S. (2002). PEGASIS: Power-efficient gathering in sensor information systems, *Proceedings in IEEE Aerospace Conference*, Big Sky, MT, USA.

Morris, C. (2018). California's Mendocino Complex Fires Devoured 7,000 Acres While You Slept, *Fortune Inc*, USA, 7 August. [Online]. Available: http://fortune.com/2018/08/07/california-mendocino-complex-fires-latest-numbers/. [Accessed 10 August 2018].

Pattanayak, B. K., Mishra M. K.,Jagadev, A. K., & Nayak A. K. (2011). Power Aware Ad Hoc On-demand Distance Vector (PAAODV) Routing for MANETS, *Journal of Convergence Information Technology*, *6* (6), 212-220.

Sen, F., Bing, Q. & Liangrui, T. (2011). An improved Energy-Efficient PEGASIS-Based protocol in Wireless Sensor Networks, *Eighth International Conference on Fuzzy Systems*, Shanghai, China. July. DOI: 10.1109/FSKD.2011.6020058

Sen, S., Clark, J. A. & Tapiador, J. E. (2011). Security threats in mobile ad hoc networks, pp. 128-136 in *Security of Self-Organising networks: MANET WSN WMN VANET*, A. K. Pathan (Ed.), Boca Raton: Auerbach Publications.

Smith, N. (2018). Timeline: The full story of the Thai cave rescue, Stuff Limited, July. [Online]. Available: https://www.stuff.co.nz/world/105495849/timeline-the-full-story-of-the-thai-cave-rescue [Accessed 10 August 2018].

Son, T.T., Minh, H. L., Sexton, G. & Aslam, N. A. (2014). Novel Encounter-Based Metric For Mobile Ad-Hoc Networks Routing, *Ad Hoc Networks* (Elsevier), *14*, March, 2-14.

Udaykumar K. & Thirugnanam, T. (2015). Analysis of Various Clustering Algorithms in Wireless Sensor Network. *International Journal of Computer Science and Information Technologies*, *6* (2), 1685-1691.

Valikannu, R., George, A., & Srivatsa, S. K. (2015). A novel energy consumption model using Residual Energy Based Mobile Agent selection scheme (REMA) in MANETs, *2nd International Conference on Signal Processing and Integrated Networks*, Noida, India. DOI: 10.1109/SPIN.2015.7095410

Implementation of RF Band-Pass Filter on UMTS Systems to Improve Quality of Service

Nyoman Gunantara
Electrical Engineering Department, Universitas Udayana, Indonesia

Made Adi Surya Antara
STIKOM, Indonesia

NMAE Dewi Wirastuti
Electrical Engineering Department, Universitas Udayana, Indonesia

Abstract: A large number of cellular operators in an area causes high interference effects. One indicator is the problem with received total wideband power (RTWP) found in a universal mobile telecommunication system (UMTS). High RTWP is one of the causes of a decrease in the quality of service (QoS) in cellular communication networks. To solve this problem, an RF band-pass filter device can be implemented on a UMTS system, where RF band-pass filters have been designed and tested in previous studies. In this research, the RF band-pass filter was installed between the BTS node and bottom jumper. After implementing it, QoS measurements with the Huawei M2000 application and Drive Test were carried out. The QoS analysis included RTWP, Circuit-Switched access rate, Packet-Switched access rate, and Received Signal Code Power (RSCP) value. The results show that the RTWP value decreases, while the CS access rate, PS access rate, and RSCP value increase. These results indicate that QoS on the UMTS increases.

Keywords: RTWP, RF band-pass filter, UMTS quality of services.

Introduction

The problems that occur in mobile communication are very complex, such as noise, fading, and interference. These problems cause Quality of Service (QoS) in mobile communication to decrease (Pratama, Endroyono & Suwadi, 2014), with a large number of cellular operators causing high levels of interference. One indicator is the problem with received total wideband power (RTWP) in the Universal Mobile Telecommunications Service (UMTS). This problem is the main cause of the decline in QoS in mobile communication (Nader, 2006).

Radio interference is a key factor that influences the quality of mobile communication. In general, radio interference is classified into external interferences and internal interferences. External interferences are those caused by a repeater, radar, analog base transceiver station (BTS) or walkie-talkie. Internal interferences are caused by the system – often resulting from an illegal terminal, parameter setting problems or BTS fault equipment – and also in-band and out-of-band interference. The results of radio interference affect several aspects of the system, such as decreased sensitivity of a BTS, decreased system capacity, increased call drop rate, decreased access success rate, call quality, call drop, handoff, conversation quality, network coverage, and capacity (Nosiri *et al.*, 2014a).

Measuring the QoS in mobile services is done by cellular operators to know and maintain the stability of the BTS infrastructure, to provide suitable QoS, quality of experience (QoE) and customer satisfaction from the mobile communication services. According to International Telecommunication Union (ITU) recommendation P.800, the standard of evaluation of mobile communication is classified into three key performance indicators (KPI), which are accessibility, retainability, and integrity (Melvi & Fernando, 2014).

QoS analysis of UMTS using Qualnet simulation has been performed by changing the value of the precedence bit in the constant bit rate (CBR) application: measured performance parameters were throughput, average jitter, and average end-to-end delay (Kaur, Rakheja & Kaur, 2010; Thethi & Sawhney, 2010). To improve the QoS of UMTS, several handover techniques have been simulated using OPNET: the measured QoS parameters were sent and received traffic, end-to-end delay variation, upload response time and download response time (Ali, Saleem & Tareen, 2012). In addition, the soft handover process in UMTS has been analyzed with OPNET simulation by adjusting the speed of the user equipment: the QoS parameters analyzed were delay, packet loss, throughput, jitter, and bit error rate (BER) (Derar & Mustafa, 2014a).

Voice over IP (VOIP) in UMTS has been studied for a variety of voice encoders using OPNET simulation. The QoS parameters analyzed were end-to-end packet loss, throughput, and end-to-end delay (Derar & Mustafa, 2014b; Werner & Vary, 2005). The integration of WLAN and UMTS networks for video conferencing and FTP applications has been studied with OPNET, measuring performance by throughput and delay (Verma & Baghla, 2014).

Besides the methods discussed above, QoS performance degradation due to interference effects can be improved using filters. The total interference in a network, with measurements of RTWP and outdoor interference, has been evaluated using Monte Carlo simulation (Tchao *et al.*, 2013). The effect of interference and the noise arising from spurious emissions in the

case of co-located channel bandwidth spectrum among operators have been described (Nosiri *et al.*, 2014b), with various industry solutions proffered to improve coverage performance and system capacity. A solution to co-channel interference and adjacent channel interference in a WCDMA system is to add an RF filter to each BTS (Pratama, Endroyono & Suwadi, 2014). This can show improvement in cellular BTS performance parameters such as call completion success rate-packet switched (CCSR-PS), call completion success rate-circuit switched (CCSR-CS), and dropped call rate (DCR) (Melvi & Fernando, 2014). The effect of interference can be overcome using the Chebyshev filter approach, with a filter that is a dielectric resonator filter with 9 poles. Such a filter can result in high selectivity and low insertion loss (Knack, Mazierska & Piel, 2008). RF band-pass filters have been designed and tested for performance improvement in UMTS through the decrease of RTWP value. The design and testing of the RF band-pass filter use the image parameter method, with components, namely L inductor and C capacitor, arranged in series and parallel. Determination of LC values uses a circuit of two half sections to get 2 poles (Antara, Gunantara & Wirastuti, 2016).

The problem of RTWP in UMTS in big cities is unavoidable, so the selection of research sites is very important (Martin-Escalona & Barcelo, 2006). The city used as the research site in the present study is Denpasar City, Bali, Indonesia. The UMTS in the Denpasar area has RTWP problems caused by high levels of radio interference because Denpasar has many cellular operators and the BTSs of different cellular operators are close together. Meanwhile, in other areas in Bali excluding Denpasar, there are more likely fewer cellular operators.

To overcome the problem of RTWP in UMTS systems in Denpasar, RF band-pass filters have been installed. The RF band pass filter implemented in Denpasar's UMTS has been designed and tested, as described in a previous paper (Antara, Gunantara & Wirastuti, 2016). After the RF band pass filter was installed, measurements were collected using the M2000 application (Huawei, 2013) and Drive Test from August 1st to September 29th, 2015. This paper summarizes these measurements. Measurements were made in two stages: before the installation of the RF band-pass filter; and after the installation using the dedicated lock mode. The QoS analysis uses measurements of RTWP, Circuit-Switched (CS) access rate, Packet-Switched (PS) access rate, and Received Signal Code Power (RSCP).

RF Band-Pass Filter and UMTS Performance

A. The UMTS

The UMTS network architecture consists of mutually supportive devices, namely user equipment (UE), UMTS terrestrial radio access network (UTRAN) and a core network (CN),

all of which can be seen in Figure 1 (Kreher, 2006). In the UMTS, the access network, UTRAN, provides the connection between the UE and CN. UTRAN consists of a radio network controller (RNC) and one or more Node-Bs (Kreher, 2006; Wardhana, 2011).
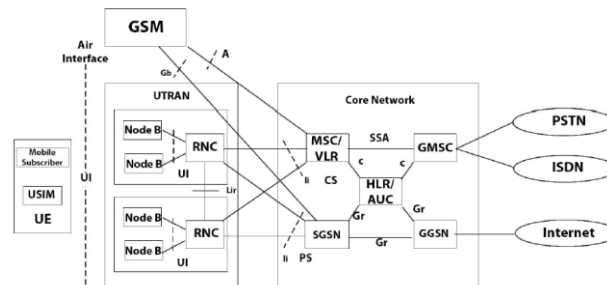

Figure 1. UMTS Network Architecture

1) RNC (Radio Network Controller)

The RNC is responsible for controlling the radio resources on the UTRAN. It oversees several Node-Bs, connecting the CN to the UE, and acts as the end of the radio resource control (RRC) protocol.

2) Node B

Node-B is the same as the base transceiver station (BTS) in a GSM network. It is a transmitting and receiving device that provides radio services to the UE. The main function of a Node-B is to process layer 1, including channel coding, interleaving, spreading, de-spreading, modulation, demodulation, and other processes. Node-B also performs several radio resource management (RRM) operations, such as handover and power control. The Node-B used in this research is designated node-B 3612998G. It has poor performance or a low QoS, with an RTWP value of -90 dBm, indicating a QoS problem. Node-B 3612998G is one of the many Node-Bs owned by cellular operators in Denpasar.

3) CN (Core Network)

The CN combines the functions of intelligence and transport. It supports signalling and transporting information from traffic, including traffic load relief. The intelligent functions are logic, control facilities of service through a clearly defined interface, and mobility settings. UMTS is also connected with other telecommunication networks, so communication is possible not only between UMTS mobile users but also with other networks.

The UMTS implemented in Bali is governed by the Minister of Communication and Informatics (Indonesian: *Menteri Komunikasi dan Informatika*, or Menkominfo) with an uplink frequency in the range 1920-1980 MHz and a downlink frequency range of 2110-2170 MHz. This study discusses the UMTS uplink frequency with a bandwidth of 5 MHz (Menkominfo, 2013).

B. RF Band Pass Filter

The three types of RF filter are band pass, low pass, and high pass. This research uses a band-pass type of RF filter to adjust the centre frequency that can be passed in the RF filter circuit (Lacanette, 1991; Sayre, 2008). The RF filter application in this research has been done through design and choosing a two half-section band-pass filter with the best image parameter (Antara, Gunantara & Wirastuti, 2016). This section will describe the design briefly. The band pass filter design uses the image parameter method. The image parameter method is a framework for analyzing the value calculations of series and parallel components in a passive filter (Sayre, 2008). The band-pass filter design with the image parameter method is similar to the RF low-pass and high-pass designs; however, the complexity is greater due to the components and the two cut-off frequencies. Similar to the filter design with low pass and high pass, this research uses two half sections, as shown in Figure 2.

According to Figure 2, in the band pass filter circuit, each LC pair is a single pole, so the half section consists of two inductors and two capacitors. The two half-section band-pass filter circuits can be connected to create a filter with more poles. The *Ro* element value is the value of comparison between input and output impedance value (Z_{in}/Z_{out}) or matching impedance with a maximum value of 50 ohms (Sayre, 2008).

To design the band pass filter with the image parameter method, it is necessary to calculate the LC element value for the half-section circuit. The equation element value L_s describes the series inductance value, C_s is the series capacitance value, L_p is the parallel inductance value, and C_p describes the parallel capacitance value. For the two half-section band-pass filter circuits as shown in Figure 2, the value of the LC element can be calculated using the following equations:

$$L_s = \frac{(\frac{R_0}{(f_{2c} - f_{1c})\pi})}{2} \tag{1}$$

$$C_s = 2\left(\frac{f_{2c} - f_{1c}}{R_0(f_{2c}.f_{1c})4\pi}\right) \tag{2}$$

$$L_p = 2\left(\frac{R_0(f_{2c} - f_{1c})}{(f_{2c}.f_{1c})4\pi}\right) \tag{3}$$

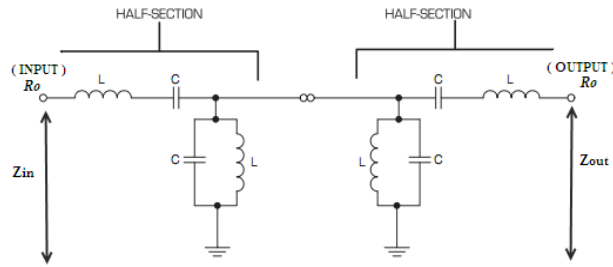$$C_p = \frac{(\frac{1}{R_0(f_{2c} - f_{1c})\pi})}{2} \tag{4}$$

Figure 2. The composition of two half-section band-pass filter circuit

The two half-section band-pass filter circuit consists of several LC pairs or two poles, with inductance and capacitance elements, L_s, C_s, L_p, and C_p, in series and parallel. The circuit can be seen in Figure 3 (Sayre, 2008).
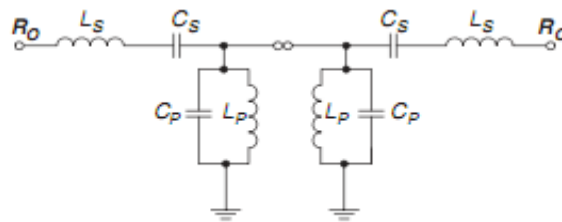

Figure 3. Elements L_s, C_s, L_p, and C_p in the two half-section band-pass filter

From the design and simulation results that have been carried out in the study by Antara, Gunantara & Wirastuti (2016), the parameters of the RF band-pass filter can be seen in Table 1.

Table 1. Parameters of RF Band-Pass Filter

Parameter	Value
f_{Low}	1955 MHz
f_{High}	1980 MHz
f_{center}	1967.5 MHz
B	25 MHz
Ro	50 Ohm
L_s	318 nH
L_p	0.0514 nH
C_s	0.0206 pF
C_p	0.1273 nF

The RF band pass filter simulation results can be seen in Table 2 (Antara, Gunantara & Wirastuti, 2016).

Table 2. Simulation Results for Parameters from RF Band-Pass Filter

Parameter	Specification		Simulation Result	
Frequency Range	UMTS UL: 1955-1980 MHz UMTS DL: 2145-2170 MHz		Up	Down
UMTS UL (Up Link)				
Insertion Loss	Typical	0.7 dB max	0.7 dB	0.7 dB
	Edge	1.0 dB max	1.0 dB	1.0 dB
Return Loss	18 dB min		20.08	21.28
Rejection	1983-1990 MHz	65 dB Min	66.26	66.31

Parameter	Specification		Simulation Result	
UMTS DL (Down Link)				
Insertion Loss	Typical	0.5 dB Max	0.18	0.15
Return Loss	18 dB min		20.75	23.60
Rejection	1983-1990 MHz	65 dB Min	66.14	66.25
DC/AISG (DC-3.0 MHz)				
Insertion Loss	1.0 dB Max		0.21	0.22
Return Loss	12 dB Min		25.54	25.61

C. The UMTS System Performance

The UMTS performance parameters written about in this paper are as follows.

1) RTWP

RTWP is the total power of the entire received signal in the uplink frequency at the antenna. The received signal is from user equipment or external interference or noise in the UMTS. RTWP values can be calculated through the following equation (Junjie and Qiong, n.d.):

$$RTWP = KTB + NF \qquad (5)$$

where:

K : Boltzmann constant

T : temperature (290 K)

B : bandwidth (Hz)

NF : noise factor

Noise factor (NF) is also specified as the total equivalent input noise power divided by the input noise power due to the source only.

RTWP can also be referred to as the total power received in the antenna at an uplink frequency, as can be seen in Table 3 (Kreher, 2006).

Table 3. Parameter Values of RTWP

Parameter	Threshold	Adjective
RTWP	< -105 dBm	Excellent
	-105 dBm to -92 dBm	Good
	92 dBm to -85 dBm	Fair
	> -85 dBm	Bad

2) CS Access Rate

CS Access Rate is used to measure the level of network availability in providing services in the form of voice services and SMS.

$$CS\ access\ rate = \frac{\sum call\ establish - \sum block\ call}{\sum call\ establish}\ x\ 100\% \qquad (6)$$

- Call establish is a successful process when an MS (mobile station) contacts another MS.

- Block call is a process when an MS fails to contact another MS.

3) PS Access Rate

PS Access Rate is used to measure the level of network availability in providing services in the form of internet data packet traffic service in the UMTS network.

$$PS\ access\ rate\ = \frac{\sum packet\ establish - \sum block\ packet}{\sum packet\ establish}\ x\ 100\% \qquad (7)$$

4) CPICH RSCP (Common Pilot Indicator Channel Received Signal Code Power)

CPICH RSCP is the level of signal strength of the UMTS network received by the user equipment. The RSCP can be used to analyze coverage. Areas with an RSCP under -97 dBm are considered to have low coverage.

Implementation and Measurement

This section begins with the placement of the RF band-pass filter on the UMTS with RTWP problems, followed by measurements using the M2000 data interface and Drive Test.
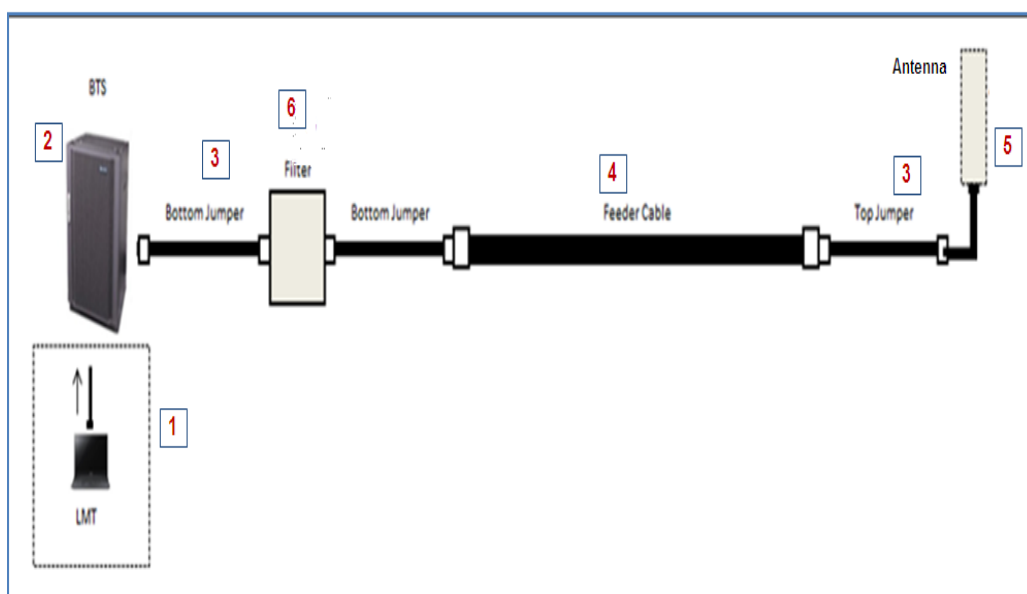
A. RF Band-Pass Filter Placement

Figure 4. Node-B block diagram arrangement prior to and after the RF band-pass filter installation (*Legend:* 1. Local Maintenance Terminal; 2. BTS; 3. Bottom and Top Jumpers; 4. Feeder Cable; 5. RF Antenna; 6. RF band-pass filter)

A Node-B, as shown in Figure 4, generally consists of a local maintenance terminal (LMT), BTS, bottom jumper, feeder cable, top jumper, and RF antenna. The RF band-pass filter that has been designed and tested is placed between the bottom jumper and BTS in order to reduce interference at the radio frequency unit installed on the BTS. In Figure 4, the arrangement of building-blocks 1 to 5 is prior to the installation of the RF band-pass filter. Once it is installed, the building-block number 6 is added.

From Figure 4, it can be seen also that the LMT measuring instrument is connected to the Node-B (BTS) interface to monitor the status of Radio Base Station, so the internal node condition towards the installed RF filter can be known (Huawei, 2011).

B. M2000

Data is collected through the M2000 application (Huawei, 2013), which previously had shown that the RTWP value is high (poor). Statistics on the RTWP value, PS access rate and CS access rate have been collected from the node implemented with the RF band-pass filter. The measurement process using M2000 can be seen in Figure 5.
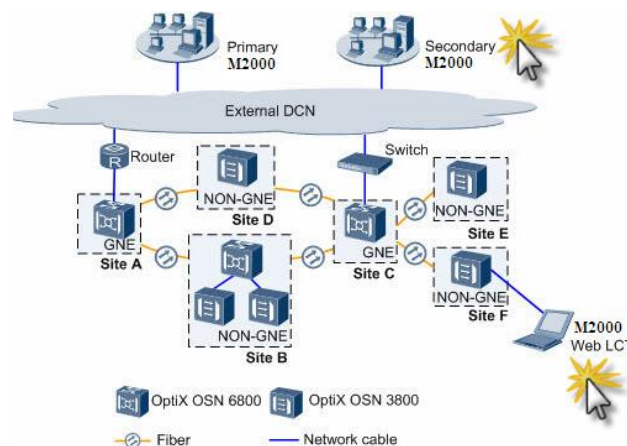
Figure 5. Measurements using M2000

In Figure 5, the arrows in yellow indicate the position of the M2000 measuring instrument. The data processing performance is obtained from monitoring the UMTS network system using the M2000 application. This performance data is collected by every UMTS node in serving every activity performed by the user, including voice, access data, or even a video call. The performance data collected by the UMTS node is then sent hourly by the node to the RNC, from which it is available. The performance data is later processed by the system so that it can be accessed through the M2000 application (Huawei, 2013).

C. Drive Test

The Drive Test is used to measure the signal quality perceived by the user and to optimize the existing UMTS system. There are several tools that can be used in a drive test, including

TEMS Investigation (Ericsson), NEMO (Nokia) and GENEX Probe (Huawei). The Drive Test in this paper uses the TEMS Investigation device. The parameter whose performance is to be measured is the CPICH RSCP (Wardhana, 2011). The measurement process using the Drive Test can be seen in Figure 6.
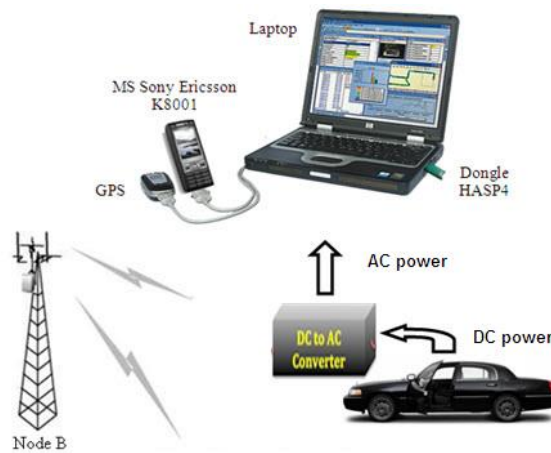

Figure 6. Measurements using Drive Test

There are two types of measurement in this Drive Test: idle mode and dedicated mode. Idle mode is when a mobile station (MS) is on but does not have a connection to the radio network, also referred to as a standby condition. In this condition, the MS can still access the radio network system and the radio network can also reach the MS. This mode is usually done only to know the signal strength of an area where a low signal or no service is indicated. Dedicated mode is the measurement of signal quality followed by a canal occupation (long call or short call to a certain destination number) to measure and identify the quality of voice and data.

Data collection on the Drive Test is done using two methods: single site verification (SSV) and cluster method. SSV is a commonly used Drive Test on a new site with close coverage area to check functionalities such as voice call, video call, PS download, HSDPA download, power emitted according to coverage, and ability to swap feeders. Meanwhile, the Cluster method is a Drive Test conducted on an area or cluster consisting of several sites, in order to observe the network of a mobile carrier (Kreher, 2006).

The equipment selected for the Drive Test includes a laptop, TEMS, dongle HASP4, TEMS cellular phone, data cable, GPS tracker, and accessories (Kreher, 2006).

The Drive Test route is determined according to the proximity of the node position with other nodes, while considering the coverage at farther distances and also operational issues. These will affect the results of the Drive Test. With the positions of the nodes far away from each other, a Drive Test route is developed in order to obtain maximum results so that the

UMTS node performance indicator after the implementation of RF band-pass filter can be acquired.

The SSV Drive Test routes and cluster Drive Test routes to see the signal quality provided in the RF band-pass filter implementation can be seen in Figure 7.
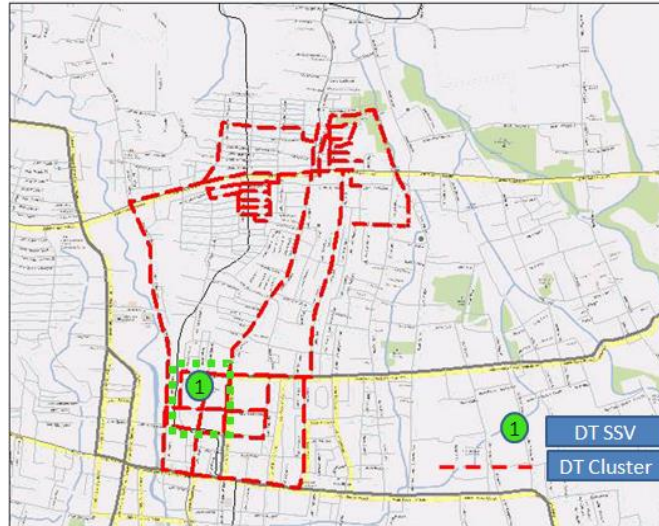
Figure 7. SSV Drive Test route and cluster Drive Test route after the implementation of RF band-pass filter node in UMTS Denpasar area

As shown in Figure 7, from the existing UMTS network area routes, a measurement using the Drive Test is done using the SSV and cluster method. The green lines show the SSV Drive Test route, while the red lines indicate the cluster Drive Test route. A vehicle with in-built Drive Test tools will pass through the route to obtain the signal quality data from the measured network. The results of each UMTS network performance measurement are then processed and analyzed.

Measurement Results and QoS Analysis

Since the RF band-pass filter is installed in the UMTS, a hardware radio frequency unit (RFU) scan is needed to find out the internal node condition towards the installed RF band-pass filter. The scanning results of the RFU hardware with the LMT show that there are no active alarms on the nodes, and only the door open alarm is on, which indicates that there is ongoing operational and maintenance activity in the UMTS. The scanning results can be seen in Figure 8.
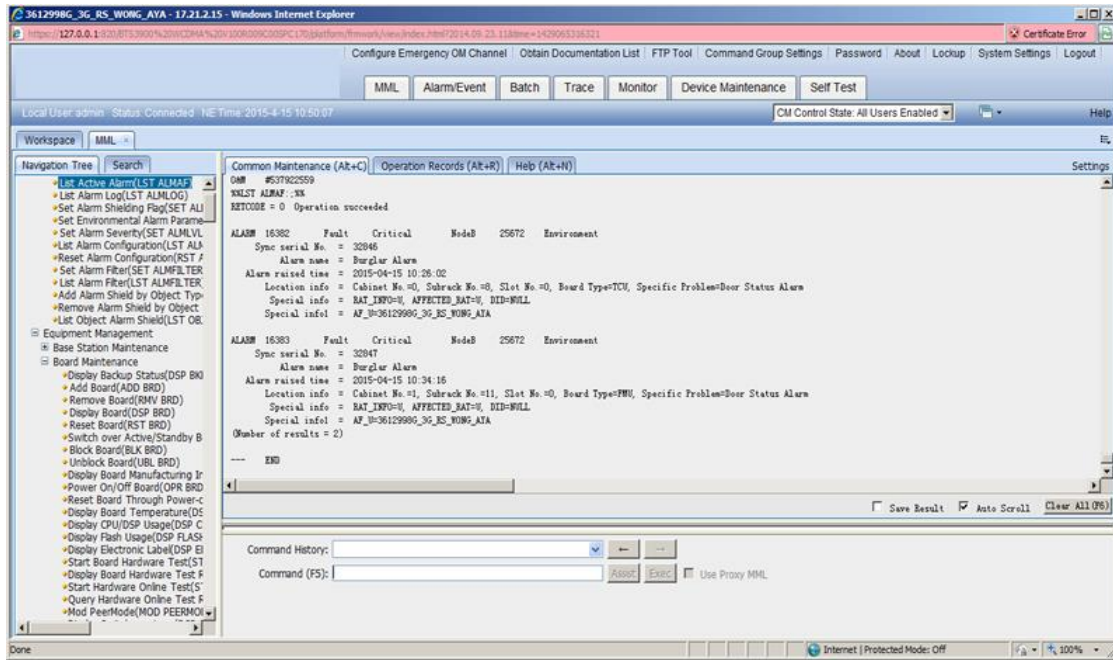
Figure 8. Result of Local Maintenance Terminal Scan of RF Unit

Based on the RFU hardware scanning results in Figure 8, the internal node condition is good, in terms of cable connection, board configuration, transmission line, and functionality board. Further analysis leads to external interference at the node where the RTWP issues exist.

A. M2000 Measurement Results

Statistics of the mean RTWP at node 3612998G, which were processed using the M2000 application conducted from August 1st to September 29th, 2015 can be seen in Figure 9.

From Figure 9, we can see the fluctuations of parameters F1 and F2. F1 is the first carrier frequency of the UMTS system of the mobile carrier whose designation is for voice and video traffic, whereas F2 is the second carrier frequency of the UMTS system of the mobile carrier whose designation is for data and HSDPA traffic.

Figure 9. Mean RTWP at node 3612998G

From Figure 9, the RTWP mean value decreased after the implementation of RF band-pass filter on September 10th, 2015. Before that date, the mean RTWP at node 3612998G was still between -90 dBm and -95 dBm. This serves as an indicator of interference on the uplink.

With the implementation of the RF band-pass filter on September 10th, the mean RTWP value steadily improves to a value of between -102 dBm and -104 dBm. This is caused by the implementation of the RF band-pass filter. The noise factor value decreases, causing the RTWP value to decrease.
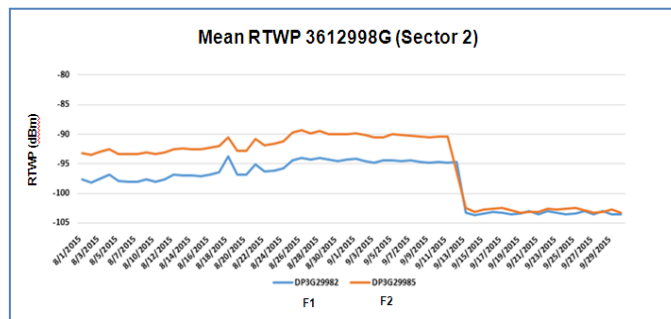
The CS access rate at node 3612998G, processed by the M2000 application in the period August 1st to September 29th, 2015, can be seen in Figure 10.
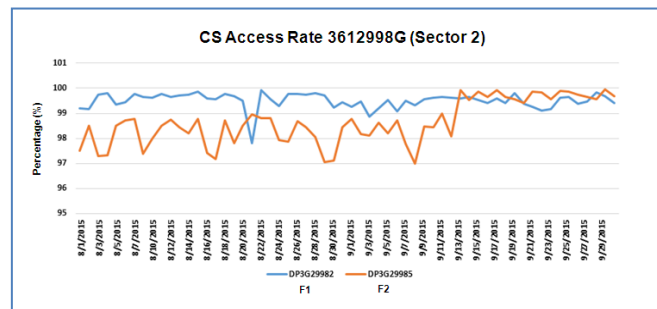
Figure 10. CS access rate at node 3612998G

From Figure 10, the value of the CS access rate has improved after the RF band-pass filter implementation on September 10th, 2015. Before the implementation, the CS access rate at node 3612998G on its second carrier frequency (F2) showed extremely high fluctuations, with values between 97% and 98%. After the implementation, the value only fluctuates slightly, with an increase that reaches almost 100%.

The PS access rate at node 3612998G from the M2000 application for the period August 1st to September 29th, 2015 can be seen in Figure 11.
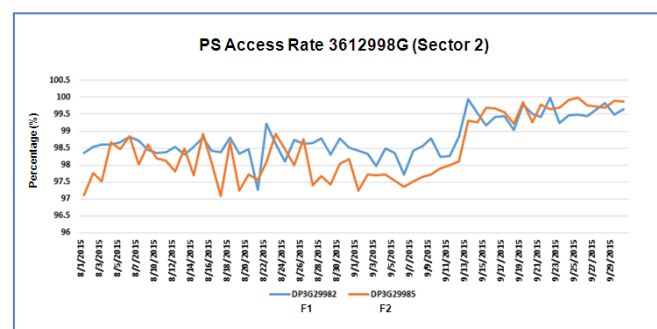
Figure 11. PS access rate at node 3612998G

The data given in Figure 11 shows that the PS access rate improved after the implementation of the RF band-pass filter on September 10th, 2015. Before the implementation, the PS access rate at node 3612998G fluctuates extremely highly, between 97% and 98.5% for the second carrier frequency. After the RF band-pass filter implementation, the PS access rate steadily increases to almost 99.5%.

B. Drive Test Measurement Results

The results of the Drive Test show symbols coloured red, yellow, light green, and dark green, depending on the values of the measured parameters. Here, we show the results for CPICH RSCP. The colours have the following meanings: Red means the RSCP level is at -120 to -95 dBm; yellow shows an RSCP level of -95 to -85 dBm; light green indicates an RSCP level of -85 to -75 dBm; and dark green indicates an RSCP level of -75 to 0 dBm.

The SSV Drive Test measurement process is done in two stages: before the RF band-pass filter implementation and after the RF band-pass filter implementation with a dedicated lock mode. The results of the SSV Drive Test measurements before the RF band pass filter is installed on the UMTS can be seen in Figure 12.
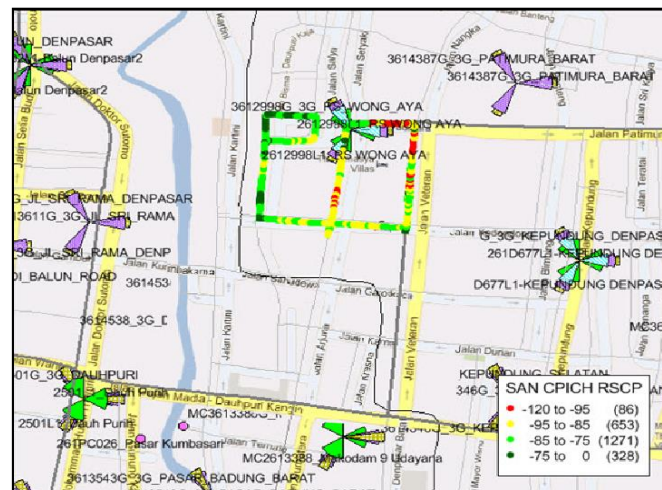
Figure 12. Measurement of SSV Drive Test prior to RF band-pass filter implementation with a dedicated lock mode

From the results shown in Figure 12, the sample data for each indicator is obtained using the following equation:

$$\frac{\text{Drive Test category } (bad, \; fair, \; good, \; excellent)}{\text{Number of samples from Drive Test}} \, x \, 100 \, \% \qquad (8)$$

The percentage and indicators of each measurement are:

RSCP value -120 to -95 dBm: 86/2338 x 100% = 3.68% (bad)

RSCP value -95 to -85 dBm: 653/2338 x 100% = 27.93% (fair)

RSCP value -85 to -75 dBm: 1271/2338 x 100% = 54.36% (good)

RSCP value -75 to 0 dBm: 328/2338 x 100% = 14.03% (excellent)

The lowest RSCP value, coloured red, -120 to -95 dBm, with a sample number of 86 or 3.68%, is categorized as *bad*; the colour yellow, at -95 to -85 dBm with a sample of 653 or

27.93%, is considered *fair*; and the dark green colour, at -75 to 0 dBm with a sample of 328 or 14.03%, is categorized as *excellent*.

After the RF band-pass filter is installed, the RSCP is measured using the SSV drive test done on the UMTS using dedicated lock mode. The results can be seen in Figure 13.

Figure 13. Measurement of SSV Drive Test after the implementation of RF band-pass filter with a dedicated lock mode

From the results that appear in Figure 13, the sample data from each indicator is obtained through equation (8). The percentage and indicators of each measurement are:

RSCP value -120 to -95 dBm: 0/2338 x 100% = 0% (bad)

RSCP value -95 to -85 dBm: 0/2338 x 100% = 0% (fair)

RSCP value -85 to -75 dBm: 407/2338 x 100% = 17.41% (good)

RSCP value -75 to 0 dBm: 1931/2338 x 100% = 82.59% (excellent)

The RSCP values coloured red, at -120 to -95 dBm, and yellow, at -95 to -85 dBm, show 0 samples. QoS analysis on RSCP with SSV drive test, after the RF band-pass filter implementation, shows that the level of RSCP improves or increases.

The next measurement is done using the cluster method. This process is done in two stages: before the RF band-pass filter implementation; and after the RF band-pass filter implementation with a dedicated lock mode. The measurements before the implementation can be seen in Figure 14. From the measurements in Figure 14, the data sample is obtained for each indicator through equation (8). The percentage and indicators of each measurement are:

RSCP value -120 to -95 dBm: 3659/221735 x 100%=1.65% (bad)

RSCP value -95 to -85 dBm: 44382/221735 x 100%=20.02% (fair)

RSCP value -85 to -75 dBm: 78332/221735 x 100%=35.33% (good)

RSCP value-75 to 0 dBm: 95362/221735 x100%=43.00% (excellent)

The RSCP value coloured red, with a range of -120 to -95 dBm and a sample of 3659 or 1.65%, is considered *bad*; the colour yellow, with -95 to -85 dBm and a sample of 44382 or 20.02%, is categorized as *fair*; and dark green, at -75 to 0 dBm with a sample of 95362 or 43.00%, is categorized as *excellent*.

The measurement results from the cluster Drive Test after implementing the RF band-pass filter using the dedicated lock mode can be seen in Figure 15.
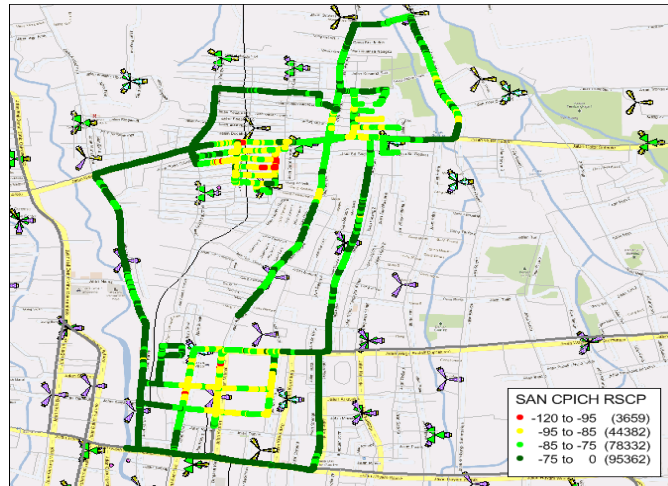
Figure 14. Measurement of Cluster Drive Test prior to RF band-pass filter implementation with a dedicated lock mode

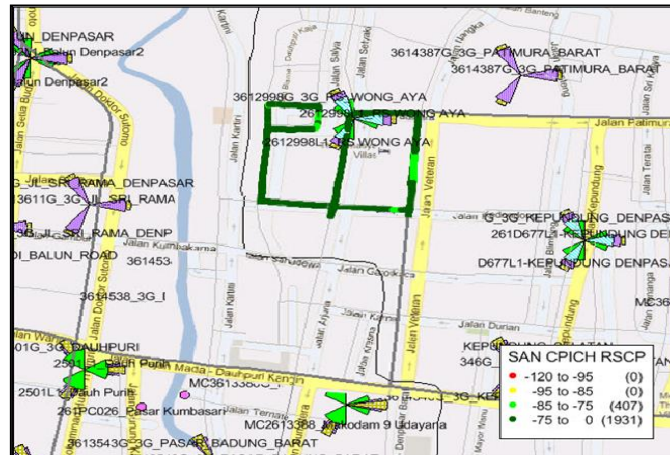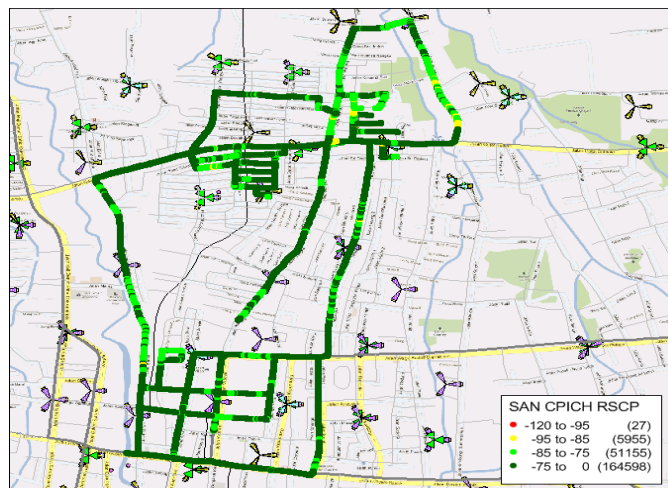**Figure 15. Measurement of Cluster Drive Test after the implementation of RF band-pass filter with a dedicated lock mode**

From the measurement results in Figure 15 the sample data for each indicator is obtained through equation (8). The percentage and indicators of each measurement are:

RSCP value -120 to -95 dBm: 27/221735 x 100% = 0.012% (*bad*)

RSCP value -95 to -85 dBm: 5955/221735 x 100% = 2.69% (fair)

RSCP value -85 to -75 dBm: 51155/221735 x 100% = 23.07% (good)

RSCP value -75 to 0 dBm: 164598/221735 x 100% = 74.23% (excellent)

The RSCP value coloured red, at -120 to -95 dBm with a sample of 27 or 0.012%, is categorized as *bad*; the value coloured yellow, at -95 to -85 dBm with a sample of 5955 or 2.69%, is considered *fair*; and the dark green indicator, at -75 to 0 dBm with a sample of 164598 or 74.23%, is categorized as *excellent*.

In the cluster drive test, after the RF band-pass filter implementation, there is a difference in colour: there is no red and only a little yellow. This gives an indication of improvement in the RSCP value of the UMTS system.

Discussion and Conclusion

QoS measurements have been carried out on UMTS systems with the M2000 and Drive Test. QoS measurement has been done in two stages. In the first stage, QoS measurements are carried out without the RF band-pass filter. In the second stage, QoS measurements are carried out with the RF band-pass filter in place.

Through the M2000, the QoS measurement results are in the form of RTWP, CS access rate, and PS access rate. Before the RF band-pass filter is installed, the resulting QoS is a high RTWP value and the values of the access rate fluctuate very highly. After the RF band-pass filter is installed, the RTWP value decreases because the value of the noise factor decreases. For the performance of the access rates, the results obtained are more stable and increase.

With the Drive Test (both SSV and Cluster), the QoS measurement results are RSCP values. Before the RF band-pass filter is installed, the SSV and Cluster Drive Tests show very small RSCP values. This is indicated by the symbols being red or yellow. After installing the RF band-pass filter, the RSCP value increases. This is indicated by a reduction in the number of symbols of red and yellow colour, and an increase in dark green colour.

The QoS measurement results before installing the RF band-pass filter show poor performance due to the signals in the 1955-1980 MHz band being mixed with noise and interference from the 1983-1990 MHz band. After installing the RF band-pass filter, the QoS measurement results are better. This is because the RF band-pass filter stops the noise and interference from the 1983-1990 MHz band, so that only the desired frequency can pass to the UMTS system.

After implementing the RF band-pass filter, which is located between the BTS node and bottom jumper, QoS performance has been measured. From the results of the QoS analysis, it can be concluded first that the average value of RTWP decreases. This is indicated by the RTWP value: from a range of -90 dBm to -95 dBm to a range of -102 dBm to -104 dBm. Second, CS access rate value only fluctuates slightly, with an increase that reaches almost

100%. Third, the PS access rate steadily increases from 97% to almost 99.5%. Fourth, the RSCP value from the SSV drive test improves or increases, with no red or yellow symbols. Lastly, the RSCP value from the cluster drive test indicates improvement of the RSCP value, shown by no red symbols and only a little yellow colour.

Acknowledgements

References

Ali, S., Saleem, N., & Tareen, T. (2012). Measuring the Performance of Handover Mechanisms in UMTS for Diverse Traffic Services Classes to Improve QoS. *International Journal of Computer Applications, 55* (11). DOI:10.5120/8799-3021

Antara, M.A.S., Gunantara, N., & Wirastuti, N.M.A.E.D. (2016). RF Band Pass Filter Method for Performance Improvement Received Total Wideband Power on UMTS System. *Majalah Ilmiah Teknologi Elektro*, *15* (1), Januari-Juli [Indonesian Journal].

Derar, L.B.A., & Mustafa, A.B.A. (2014a). Improvement Performance of Soft Hand Over on UMTS Network. *Journal of Scientific and Engineering Research*, *1* (2).

Derar, L.B.A., & Mustafa, A.B.A. (2014b). Quality Of Service In UMTS Network And Improvement VOIP Performance. *International Journal of Technology Enhancements and Emerging Engineering Research*, *2* (10).

Huawei. (2011). *NodeB V200R013 Technical description*. Huawei Technologies Co. Ltd.

Huawei Technologies Co. Ltd. (2013). *BTS3900A Monitoring System* (Manual Guidebook), China: Huawei.

Junjie, Z., & Qiong, L. (n.d.). *RTWP Troubleshooting Guide*. Huawei Technologies Co. Ltd.

Kaur, D., Rakheja, P., and Kaur, A. (2010). Analysing Quality of Service in UMTS. *International Journal of Computer Applications, 11* (12).

Knack, A., Mazierska, J., & Piel, H. (2008). Dielectric Resonator Filters for UMTS Systems. *IEEE MTT-S International Microwave Workshop Series on Art of Miniaturizing RF and Microwave Passive Components*. Chengdu, China, December.

Kreher, R. (2006). *UMTS Performance Measurement, A Practical Guide to KPIs for the UTRAN Environment*. Chichester: John Wiley & Sons Ltd. DOI: 10.1002/9780470034866

Lacanette, K. (1991). *A Basic Introduction to Filters – Active, Passive, and Switched-Capacitor,* National Semiconductor Application Note 779. Available at http://www.ti.com/lit/an/snoa224a/snoa224a.pdf

Martin-Escalona, I., & Barcelo, F. (2006). Middleware-Controlled Resource consumption for Location Traffic in Cellular Networks. *Journal of Communications Software and Systems*, *2* (4), December.

Melvi, A.U., & Fernando, R. (2014). Mobile Based Transceiver Station Performance Analysis for the Stability of Infrastructure of Cellular Communication System. *Jurnal Rekayasa Elektrika*, *11* (1), 1-7, April [Indonesian Journal].

Menkominfo. (2013). News Bulletin SDPPI, Media Information and Communication, Directorate General Resources and Equipment Post and Informatics, Ministry of Communications and Informatics. The fifth edition [Indonesian Bulletin].

Nader, G. (2006). *Ultra Wideband Interference on Third-Generation Wireless Networks*. Dissertation, Virginia Polytechnic Institute and State University.

Nosiri, O.C., Idigo, V.E., Ohaneme, C.O., & Akpado, K.A. (2014a). Coverage and Capacity Performance Degradation on a Co-Located Network Involving CDMA2000 and WCDMA @1.9GH, *International Journal of Computer, Information, Systems and Control Engineering*, *8* (3).

Nosiri, O.C., Onoh, G.N., Idigo, V.E., & Ohaneme, C.O. (2014b). Mechanisms and Industry Solutions for RF Interference in a Co-located Network. *Journal of Communications Engineering and Networks*, *2* (3), 92-102, July.

Pratama, W., Endroyono & Suwadi. (2014). Solutions for Suppressing Co-Channel and Adjacent Channel Interference on Multi-Operator WCDMA Cellular System. *Jurnal Teknik POMITS*, *1* (1), 1-6 [Indonesian Journal].

Sayre, C.W. (2008). *Complete Wireless Design*. The McGraw-Hill Companies, Inc.

Tchao, E.T., Ofosu, W.K., Diawuo, K., Affum, E., & Agyekum, K. (2013). Interference Simulation and Measurements for deployed 4G-WiMAX Network In an Urban Sub-Saharan African Environment. *International Journal of Computer Applications*, *71* (14), June. DOI: [10.5120/12424-8973](10.5120/12424-8973)

Thethi, A. S. & Sawhney, R.S. (2010). Performance Evaluation of QoS parameters in UMTS Network Using Qualnet. *International Journal of Distributed and Parallel Systems (IJDPS)*, *1* (2).

Verma, V. & Baghla, S. (2014). Performance Evaluation of QoS in WLAN-UMTS Network Using OPNET Modeller. *International Journal of Science and Research*, *3* (8).

Wardhana, L. (2011). *2G/3G RF Planning and Optimization for Consultant*. Jakarta: Nulis Buku.

Werner, M. & Vary, P. (2005). Quality Control for UMTS-AMR Speech Channels. *9th European Conference on Speech Communication and Technology*, Lisbon, Portugal, September 4-8.

Energy-Efficient Network Protocols for Domestic IoT Application Design

Chrispin Gray

Melbourne School of Engineering, The University of Melbourne

Leith Campbell

Melbourne School of Engineering, The University of Melbourne

Robert Ayre

Melbourne School of Engineering, The University of Melbourne

Kerry Hinton

Melbourne School of Engineering, The University of Melbourne

Abstract: In the future Internet of Things, many common household devices will have communications interfaces added. The gathering of data from these household IoT-enabled devices will incur an energy cost and, in this paper, we investigate the impact of different communications technologies and protocols on that cost. As a first step towards energy-efficient design, we have measured the power consumption of several popular wireless interfaces – Bluetooth (Classic and Low Energy), ZigBee, Wi-Fi and 433 MHz module (RF433). We then combine these measurements through the example of a simple domestic stock control application and we show how an energy-efficient communications paradigm can be designed in each case. In general, both the communications paradigm and the amount of traffic need to be considered for an energy-efficient design. This is a contribution to design guidelines for energy-efficient communication in the Internet of Things as it expands to encompass all consumer devices.

Introduction

The Internet of Things (IoT) (Atzori, Iera & Morabito, 2010) is not just about new sensors and actuators but will include, in the domestic realm, many consumer devices that are not currently networked and do not draw electricity when they are not in use. Bringing these devices into the IoT ecosystem will require the addition of communications interfaces, which will mostly be wireless. A naive approach to adding communications interfaces (e.g. a Wi-Fi

module on each device) could lead to substantial additional energy consumption without any corresponding benefit in functionality, if an inappropriate design option is chosen. While this effect may be marginal for a single device, energy efficiency in communications design will be important for individual households and in aggregate may have considerable effect (for example, see chapter 14 of Weldon (2016)).

A variety of wireless network communications interfaces have been considered as enablers of the IoT, with a wide range of characteristics (e.g. bit rate, topology and energy consumption) (Atzori, Iera & Morabito, 2010; James, 2014). They include Bluetooth (Classic and Low-Energy), 802.15.4/ZigBee, 6LowPAN, Z-Wave, Thread, DECT-ULE, EnOcean, Wi-Fi, NFC, ANT/ANT+, WirelessHART and a number of proprietary protocols. Here, we compare the energy consumption characteristics of five common commercial off-the-shelf (COTS) wireless interfaces, which represent today's dominant COTS interfaces for IoT application and are expected to see significant growth in the near future (OnWorld, 2017; OnWorld, 2018). They are Bluetooth Classic (BT), Bluetooth Low Energy (BLE), ZigBee, Wi-Fi and 433 MHz module (referred to as RF433 in this study). Power consumption of the COTS interface options was measured for their various states of operation and combined to determine their energy use in different scenarios.

To examine the options for energy-efficient communication using these COTS devices, we employ a simple stock control application involving a domestic toaster communicating with a gateway hosting an inventory system. This application is rich enough to encompass the full range of communication options while being simple enough to enable a straightforward analysis of energy consumption. The application choice is in line with developing interest in 'smart kitchens' (Pal Amutha, Sethukkarasi & Pitchiah, 2012) and automated stock control for online reordering of basic grocery items (Hsu *et al.*, 2006; Yang *et al.*, 2014) as part of assisted living.

While the example given here may be somewhat contrived, it illustrates the point that careful design for energy usage and efficiency in the IoT will be important. The type and volume of data, the frequency of data transmission, and the type of communications interface should all be considered.

In this paper, we first give a brief description of the five wireless network communication protocols and describe the measurements obtained through experiments. We then combine these measurements using the example IoT application to determine the energy consumption of several communications paradigms. We use these results to describe energy-efficient design options for future domestic IoT applications.

Wireless Network Protocols

There are several short-range wireless network communications protocols that are considered as part of the enabling technologies of the in-home IoT. In this section, we present a brief description of the five commonly employed wireless network protocols for IoT applications: Bluetooth Classic; Bluetooth Low Energy; Wi-Fi; ZigBee; and Radio Frequency 433 MHz (RF433).

Bluetooth Classic

Bluetooth technology, based on the IEEE 802.15.1 standard, is designed to support short-range, ad-hoc connectivity amongst devices. Traditional or Bluetooth Classic (BT) operates in the 2.4 GHz Industrial Scientific and Medical (ISM) band using adaptive Frequency Hopping Spread Spectrum (FHSS) channel access (1600 hops/sec), with 79 defined channels of 1 MHz channel width, 32 of which are used for device discovery (Ferro & Potorti, 2005). BT devices can operate either as a master or slave, with one master interconnecting up to seven active slave devices – hence a star topology. BT is capable of data rates up to 2 Mb/s. However, BT has some major disadvantages, which limit its widespread implementation in IoT applications. These disadvantages include a relatively high power consumption, large data packets with huge overhead, a complex protocol stack with large memory demand, and long connection time (Mackensen, Lai & Wendt, 2012). Additionally, while BT has some low power modes (e.g. sniff mode), it lacks an effective sleep-mode regime, which is critical for minimising energy consumption in the case of battery-operated IoT devices.

Bluetooth Low Energy

Bluetooth Low Energy (BLE) is defined in the Bluetooth Specification 4.0 (Bluetooth-SIG, 2010) as the low energy version of BT, designed for IoT-like applications – low cost and complexity, low duty cycle and infrequent transmission. Like BT, BLE (or *Bluetooth Smart*) is also based on the 2.4 GHz ISM band and uses FHSS but with 40 channels (including 3 advertisement channels for device discovery) of 2 MHz channel width, and a data rate of 1 Mb/s. BLE can either operate as a central (master) or peripheral (slave) device, with a slave able to connect to only one master device at a time (i.e. star topology). An important feature of BLE is its sleep-mode capability, which, in combination with a low duty-cycle application, significantly reduces the device energy consumption. A slave may broadcast a limited amount of data in an advertisement packet or, alternatively, a communications link between a master device and a slave device may be established for dedicated or higher volume traffic applications. The latest version of BLE is defined in Bluetooth Specification 5.1 (Bluetooth-SIG, 2019).

Wi-Fi

Based on the IEEE 802.11 series of standards for wireless local area network (WLAN), Wi-Fi operates on multiple frequency bands (e.g. 2.4 GHz, 5 GHz, 60 GHz) and is a widely used short-range wireless protocol. The Wi-Fi protocol has undergone several revisions (802.11a/b/g/n/ac) with 802.11ac capable of achieving broadband speeds in excess of 500 Mb/s (Perahia & Stacey, 2013). Wi-Fi supports two operating topologies: Peer-to-Peer (Ad-hoc) network topology and Star (Infrastructure) network topology. While Wi-Fi was originally designed for high-speed, short-range (up to 100 m) communication, its ubiquity in homes and places of work has resulted in an increasing use in a number of IoT applications (e.g. Smart Light Bulbs or Smart Appliances).

ZigBee

Based on the IEEE 802.15.4 standard, which defines the PHY and MAC layers, ZigBee has an established set of specifications for low power, low data-rate and short-range applications. The ZigBee Alliance defines the Network and Application Support layers (ZigBee, 2012). ZigBee operates on the class-licensed 868 MHz, 915 MHz and 2.4 GHz bands with data rates of 20 kb/s, 40 kb/s and 250 kb/s, respectively. ZigBee defines three types of devices: (a) ZigBee Coordinator (ZC) – acts as a gateway/hub and delivers the greatest capability; (b) ZigBee Router (ZR) – acts as an intermediate router in addition to application functions; (c) ZigBee End-Device (ZED) – performs application functions and is the least capable device. ZC and ZR are classified as fully functional devices, while ZED is a reduced-function device. The ZigBee protocol was designed for low complexity and low energy usage with a sleep-mode capability. In addition to its star and cluster-tree networking capabilities, the ZigBee protocol can be configured for mesh networking, which offers long reach by means of data relay. Through its mesh networking functions, ZigBee offers a "self-healing" capability, which is the ability to create alternate paths when one node fails or a connection is lost.

RF433

Based on the ISM 433 MHz band, RF433 can be used for wireless data transmission, including in wireless sensors and home automation systems. Its frequency range is 433.05-434.79 MHz. RF433 is not a standardised protocol like Bluetooth, ZigBee and Wi-Fi but is a widely employed, inexpensive wireless communication option for many IoT devices and applications. Based on this background and the general belief that RF433 will be a common hardware selection in future IoT applications, it is included in this study.

Related Studies

There are a number of comparative studies in the literature, many of which feature two or more wireless network communication protocols. Some of these studies have focused on the energy utilization of the wireless protocols irrespective of any specific application domain. For example, while using representative device datasheets, one study compared, among other metrics, the power consumption of four wireless protocol standards (BT, ZigBee, Wi-Fi and Ultra-Wideband) (Jin-Shyan, Yu-Wei & Chung-Chou, 2007). They showed that, while BT and ZigBee consume remarkably less power, they are far less energy-efficient than Wi-Fi and UWB (Ultra-Wide Band) for higher data rates. A more comprehensive survey of the characteristics of BT and Wi-Fi, using an experimental study of similar example chipsets programmed with the same data rate for a single connection scenario showed that Wi-Fi consumes five times more power than BT (Ferro & Potorti, 2005). Applying datasheet values in a simulation study of raw data transmission for BLE, ZigBee and Wi-Fi modules, Shahzad & Oelmann (2014) showed ZigBee being more energy efficient for low data payloads, while Wi-Fi was least energy efficient; but the reverse for high data payloads. The difference is mainly attributed to the shorter connection time of ZigBee as compared to Wi-Fi, but the larger frame size of Wi-Fi enabled more data transmission per Tx/Rx transaction for high data payloads. From the measurement-based comparisons, Mikhaylov, Plevritakis & Tervonen (2013) examined the energy consumption of BLE, ZigBee and SimpliciTI (proprietary) transceivers in different modes of operation. Siekkinen *et al.* (2012), on the other hand, compared only BLE and ZigBee transceivers, with their result showing the energy efficiency of BLE being 2.5 times better than that of ZigBee. Dementyev *et al.* (2013) compared and reported power consumption of BLE, ZigBee and ANT+ protocols in varying cyclic sleep intervals. They showed that a device sleep current draw is not always indicative of its long-term energy consumption, highlighting the importance of shorter start-up connection time.

While these studies aim to show one protocol as being more energy efficient than the other, it should be noted that energy consumption ultimately depends on the device hardware implementation and this may vary from one manufacturer to the other. With some IoT services now mainstream (e.g. Home Automation), and an increasing use of wireless protocols in such applications, there is a need for application-specific studies. IoT applications vary in requirements of bit rate, latency, packet sizes and QoS; and these must be taken into account while assessing the energy-efficiency of wireless network protocols.

Measuring Energy Consumption

We measured the power consumption (energy consumption) characteristics of each of the above communication technologies, in order to be able to compare their performance in typical applications. A description of the experiment and measurement setup is given in this section. It is assumed that, for each deployment scenario, each COTS wireless interface communicates with an always-on gateway device (i.e. master, ZC, etc.).

Measurement Setup

The components used in the measurements include two COTS short-range RF modules for each communication protocol, an Arduino Duemilanove (Due) board ([Arduino, 2009](#)) equipped with Atmel's Atmega 328P-PU microcontroller (MCU), a dc power supply and power meter unit, a test computer (PC) and smartphone. The Arduino acts as a data source. A block diagram of the experiment setup is shown in Figure 1.
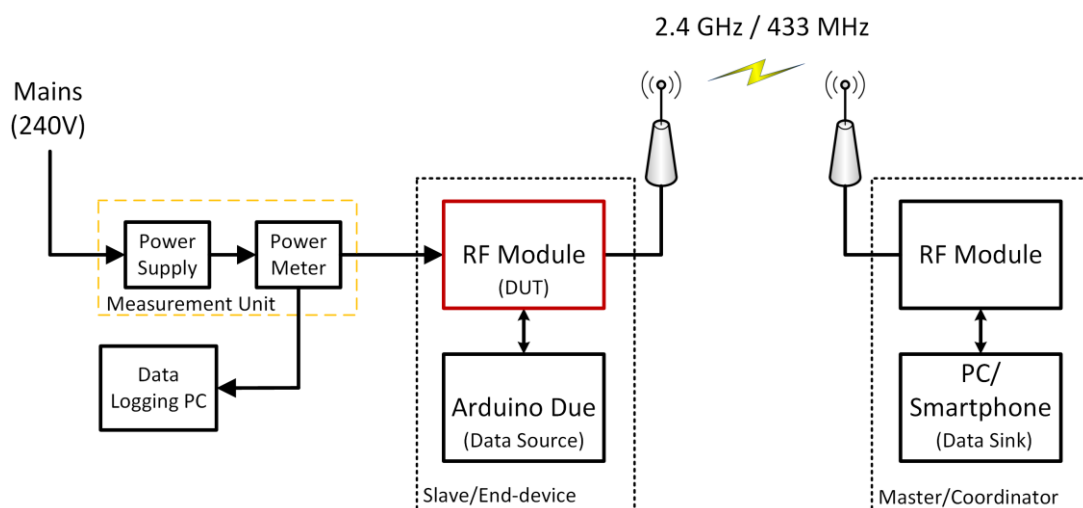
Figure 1. Block diagram of experiment setup.

Earlier RF modules might include separated transceiver circuitry, antenna, MCU and serial interface, but modern RF modules today are most commonly designed as a System-on-Chip (SoC) with integrated transceiver module and MCU. Hence, the RF modules in this study are measured as one SoC unit, given that in practice they may be deployed as such.

The power meter records power consumption, current draw (mA) and voltage (V) of the device under test (DUT)/end-device, and the measurements are logged on a PC for post-processing. The Arduino Due generates a pre-set 10-byte (10 B) test payload and is independently powered (not measured). Measurements are recorded at an average of 1 ms intervals with an accuracy of 10 µA and 5 mV voltage. With the exception of the RF433 modules, which are either transmit-only or receive-only, each RF module has 4 leads connected: +ve power input (VCC), Ground (GND), transmit (Tx) and receive (Rx). Figure 2 displays an image showing this setup

with an XBee ZigBee RF module as an example. A similar measurement format was used for modules with a USB connection; the power line within the USB cable was interrupted and voltage and current measured, while the data lines were passed through uninterrupted. For accurate measurements of the DUT power consumption, the VCC and GND leads are directly connected to the dc power supply via a power meter, while the Tx and Rx leads are either connected to the transmit and receive ports on the Arduino board or a USB port. A steady voltage of 5V (or 3.3V where applicable) is supplied to the modules during the tests.

The master/coordinator (gateway device) modules are powered either from the power rails on an Arduino board, a USB interface of a PC or the smartphone battery but these are not measured. Packets received from the DUT are recorded with timestamps. Measurements are collected over 5 minutes. In one case, a digital oscilloscope (e.g. Digilent Analog Discovery USB Oscilloscope) in a setup with a 10 Ω current metering shunt resistor was used for the measurement.
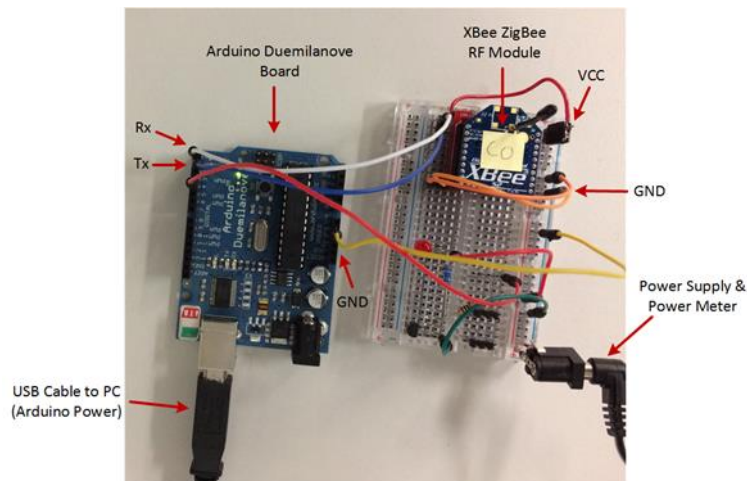
Figure 2. Measurement setup for an XBee ZigBee module.

In the experiment, the data source for the DUT is programmed to send the pre-set 10 B data payload every 5 seconds while the master/coordinator receives the transmitted data, which is displayed on a serial monitor using a terminal emulator such as Tera Term or Bluetooth Smart Data app.

RF Module Power Measurement Setup

Table 1 gives details of the RF modules used in the experiments. For BT measurements, a JY-MCU HC-06 module, compliant with Bluetooth v2.1 standard, was used. The serial interface of the BT DUT (slave) was configured at 9600 Baud. Bluetooth v2.0 standard lacks the very low power sleep-mode; therefore, it is assumed that the module, when not operated continuously, is turned off after every Tx and Rx operation of the DUT.

Table 1. RF modules used in experiments.

Standard	Device Name	RF Module Manufacturer	RF Module Number	Firmware Version	Operating Voltage (V)
Bluetooth	JY-MCU	Linvor	HC-06	1.8	5
Bluetooth LE	BLE Nano v2	Nordic	nRF52832	-	3.3
ZigBee	XBee	Digi	XBee S2	28A7	3.3
Wi-Fi	ESP-12F	Ai-thinker	ESP8266EX	-	3.3
RF433	N/A	Exportise	FS1000A/ZRW-01 (transmitter/receiver)	-	5

For the measurement of BLE, the Redbear BLE Nano v2 (Redbear, 2018) was used. The BLE Nano is a breakout board that simplifies prototyping or small-scale production of new IoT products, based on the Nordic nRF52832, which is a SoC including the RF module and an ARM Cortex-M4F processor at its core. The BLE Nano comes with Bluetooth v4.2 protocol stack (although Bluetooth 5 ready) and can be configured as a central or peripheral device. As the DUT, the BLE Nano was configured as a peripheral device (see Figure 1).

For ZigBee measurements, the Digi International's XBee Series 2 radio modules (Digi, 2011) were used. Using Digi's XCTU software, one of the modules was configured as an end-device and the other as a coordinator, both with their respective firmware. Transparent mode communication was set in both XBees (AT mode) for an asynchronous serial interface of 9600 Baud. Since ZigBee modules have sleep-mode capabilities, the end-device RF module was configured in cyclic sleep mode, waking up at regular intervals to transmit its data payload before returning to sleep. The power consumption and duration for the different stages were recorded.

The ESP-12F Wi-Fi SoC module was used as a representative Wi-Fi RF module. The ESP-12F is based on the ESP8266EX chipset and is IEEE 802.11b/g/n compliant. To avoid significant variation in power due to dynamic transmit power control, the Wi-Fi module (end-device) and Access Point (AP), i.e. master, were configured in Infrastructure mode while maintaining a separation distance between 0.5 and 1 metre.

For the measurement of an RF433 module, the FS1000A transmitter and ZRW-01 receiver radio modules were used. They both use ASK modulation or On-off keying and permit bit rates up to 9.6 kb/s. Since the transmitter and receiver are independent modules, two sets of experimental configuration were needed – one for the transmitter and the other for the receiver. Our examination of a range of commercial wireless sensor devices that use RF433 radio modules showed that they are mostly transmit-only devices, designed to re-send each packet multiple times to compensate for the lack of acknowledgement (ACK) packets from the receivers. This redundancy mechanism is implemented in our application case-study described below, assuming each message is sent three times.

Power Consumption Measurement

This section describes the measurements, power consumption plots and values for examples of each type of wireless technology considered in the experiments. In addition to the main communication task (i.e. Tx and Rx), we monitor and report on other housekeeping functions (e.g., listening, wake-up, sleep) with each technology and their associated energy consumption. Power consumption values generally depend on the device hardware implementation, and may vary slightly between examples of each module type.

Table 2 lists the power draw of the RF modules during their main operational phases, while Table 3 lists their respective duration (in milliseconds) for transmitting a 10-Byte packet. Some of the values reported in this paper were previously published in Gray & Campbell (2016). For simplicity, the duration of the wake-up, pre-processing and synchronization phases was concatenated into one, i.e. wake-up, as given in Table 3. A small amount of energy is incurred to turn off or shut down the modules but these are insignificant in comparison with the other phases.

Table 2. Power consumption of RF modules in different operational phases.

RF Module	Power Consumption (mW)				
	Sleep	Listening/ Standby	Wake-up	Transmit (Tx)	Receive (Rx)
Bluetooth	-	16	96	199	185
Bluetooth LE	0.03	2.2	21	41	33
ZigBee	0.17	30	91	139	129
Wi-Fi	0.03	62	412	227	227
RF433	-	29	12	45	33

Table 3. Duration of operation in different operational phases.

RF Module	Phase Duration (ms)		
	Wake-up	Transmit (Tx)	Receive (Rx)
Bluetooth	1836	1	1
Bluetooth LE	4	1	1
ZigBee	5	2.5	3.5
Wi-Fi	5.5	50	50
RF433	1	126	406

While the individual power consumption values are discussed in the following section, an observation of Table 2 reveals not so subtle differences between the DUT modules. For the Tx/Rx power, we see magnitudes between 3 and 6 times the lowest recorded value (i.e. BLE), while a more significant difference (28 times) can be seen in the standby power consumption. The latter is important in this study as some applications require modules to remain in the standby state, which could result in much higher disparity in total energy consumption. We

acknowledge that the above comparisons are for different wireless technologies and also that the same technology from different manufacturers may show different results. For similar modules like BT and BLE, however, we recorded a notable difference (~5 times) indicating some energy efficiency improvement trend, while improving communication range and other metrics.

BT Measurements

A Bluetooth link controller has a number of connection states including standby, inquiry (scan/advertise), paging, connected/active, sniff, park and hold states (Ferro & Potorti, 2005). Figure 3(a) shows the power consumption of the BT module in a non-connected standby state depicting the advertisement and scanning phases. The device sends advertisements on different channels, each round lasting about 5 ms. During advertisement, the BT module draws an average of 79 mW, consuming about 0.4 mJ each round. During scanning with nearly 100% duty cycle (duration $\approx$ 320 ms), the module draws about 208 mW consuming 67 mJ, which is over 50 times more than its advertisement phase. Such a huge disparity in energy usage explains the reasoning for slave/end-devices (which are often energy constrained) being limited to sending advertisements, while master devices (often with more resources, e.g. PC/Smartphone) engage in scanning during the discovery process.

(a)   (b)
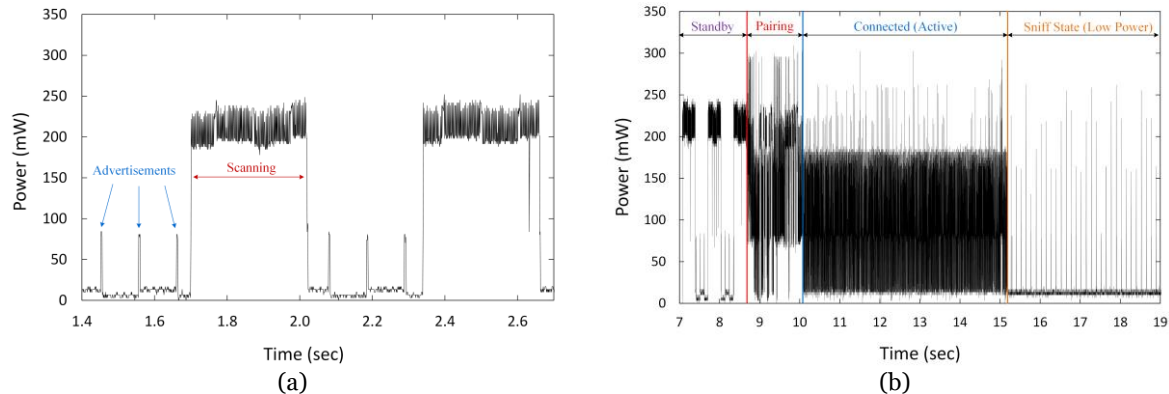
Figure 3. Power trace of BT module: (a) in standby state showing the scanning and advertising phases; (b) showing the state transitions from standby to pairing, connected and sniff states.

Figure 3(b) gives a much wider view of the major stages a BT device undergoes. The figure shows the standby state, when scanning and advertisements are performed; the pairing stage, when the paging and synchronisation process occurs; the connected stage, which includes data Tx, Rx and ACK; and lastly the low-power sniff state (i.e. listening), when the module is less-active but listens for transmissions at set intervals (125 ms in this case), depicted by the spikes on the right-hand-side of the figure. The average power consumption in the low-power sniff mode is about 16 mW. For BT, a complete data frame is transmitted per timeslot (1 timeslot = duration in 1 frequency state) in 0.625 ms, with an ACK packet received in a subsequent

timeslot. Since the power meter can only resolve variations down to 1 ms, we assume a lowest duration to be 1 ms.

BLE Measurement

The measurement was conducted for the BLE non-connected and connected/active states. Figure 4(a) shows the power consumption trace of a BLE peripheral device advertising at 200 ms intervals, and in connected mode, with connection events depicted on the right-hand-side. Each advertisement payload was 30 B long and was transmitted to three advertisement channels. For one advertisement round, the peripheral device consumed about 0.08 mJ, lasting about 2 ms (0.625 ms per timeslot). Figure 4(b) shows the power consumption trace of a non-connected BLE central device scanning three advertisement channels with a 20% duty cycle (i.e. scan window = 200 ms, scan interval = 1s). The steps in Figure 4(b) can be attributed to its post-processing activities, drawing ≈ 23 mW. For scanning, the central device consumes 14.4 mJ per cycle.
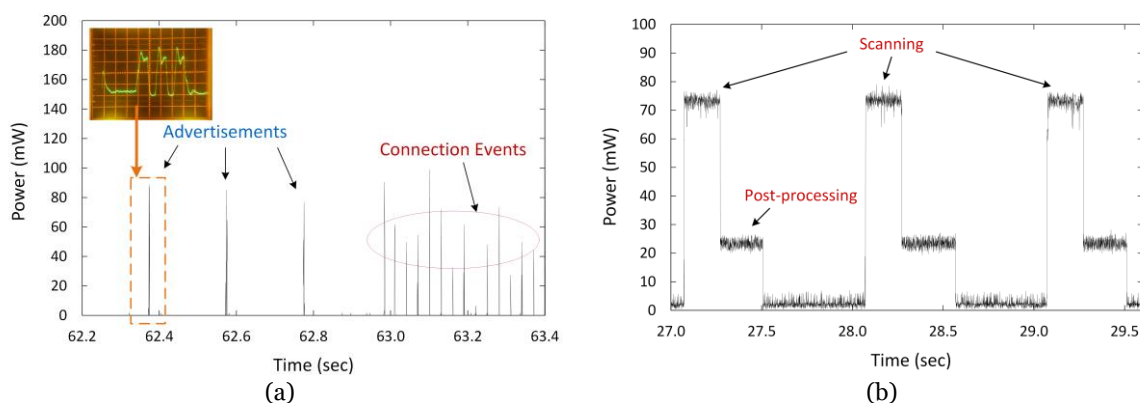
Figure 4. Power consumption trace of BLE module configured as (a) peripheral (slave) device showing advertisements and connection events; (b) central (master) device in scanning mode.

The energy consumption of a BLE module in its non-connected states ultimately depends on the choice of a number of parameters (e.g. connection interval), which would be selected based on an application usage scenario. The measured power draw and duration of BLE connected phases are given in Table 2 and Table 3.

ZigBee Measurement

With ZigBee, the measurement was carried out on a configured XBee ZigBee end-device as the DUT, which connects to an XBee ZigBee coordinator (i.e. gateway). The DUT was configured for cyclic sleep mode with reverse polling. In reverse polling, the end-device (in sleep mode) wakes up at regular intervals (default = 100 ms but can be modified in firmware) and polls its coordinator to request any data sent to its address while sleeping. The end-device sends a poll once after its wake-up sequence and again before going to sleep. In sleep-mode, the XBee draws a maximum of 50 µA at 3.3V (Digi, 2011).
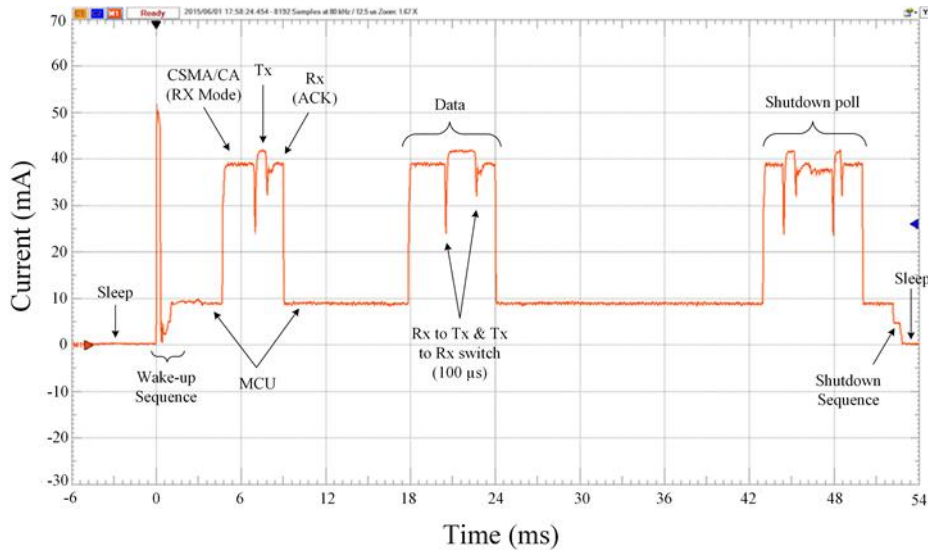
Figure 5. Current draw of the XBee ZigBee end-device module during a connection event.

Figure 5 shows a detailed plot of a ZigBee end-device module current draw for one connection event. The plot shows the wake-up sequence, which includes MCU wake-up, pre-processing and synchronisation. The initial poll is shown in the next phase, starting with the channel access process, which is the contention-based CSMA/CA as defined by the 802.15.4 MAC specification. The duration of the CSMA/CA process may vary depending on channel availability. The power consumption and duration values are reported in Table 2 and Table 3 for an operating voltage of 3.3V.

Wi-Fi Measurement

In measuring the Wi-Fi module, an infrastructure operating mode was used. The Wi-Fi module and the AP were configured for the IEEE 802.11g standard, with a maximum theoretical data rate of 54 Mb/s (the highest of all the RF modules considered). At start-up the Wi-Fi module scans the channels for periodic beacons from the AP before a connection process is initiated. Regular beacons are received from the AP at 100 ms intervals and a block of data is transmitted to the gateway every 3 sec. Figure 6 shows a power consumption plot of the ESP-12F module depicting the beacons and data transmit phase. (There is some variation in the magnitude of the initial spike, due to the brevity of the spike duration compared with the power meter sampling interval.) As with every TCP/IP process, Tx packets are followed by ACK. However, the Rx process is indistinguishable from that of the Tx in the figure. The Wi-Fi module power draw in different states is reported in Table 2 and their duration in Table 3.
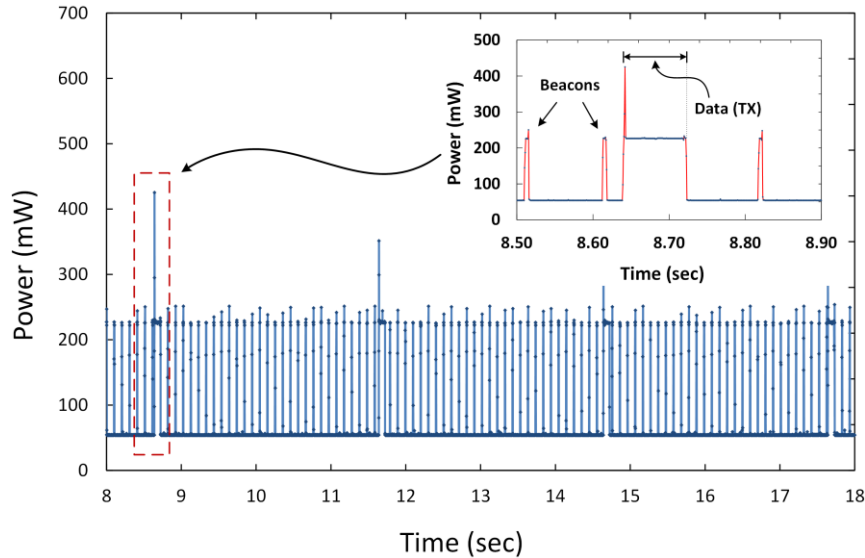
Figure 6. Power consumption trace of ESP-12F Wi-Fi module.

RF433 Measurement

The RF433 transmitter and receiver modules were measured separately. Figure 7(a) shows the power trace of the 433 MHz Tx module sending 10 Bytes of data and Figure 7(b) shows the Rx module receiving the same amount of data at 5 sec intervals.

(a)   (b)
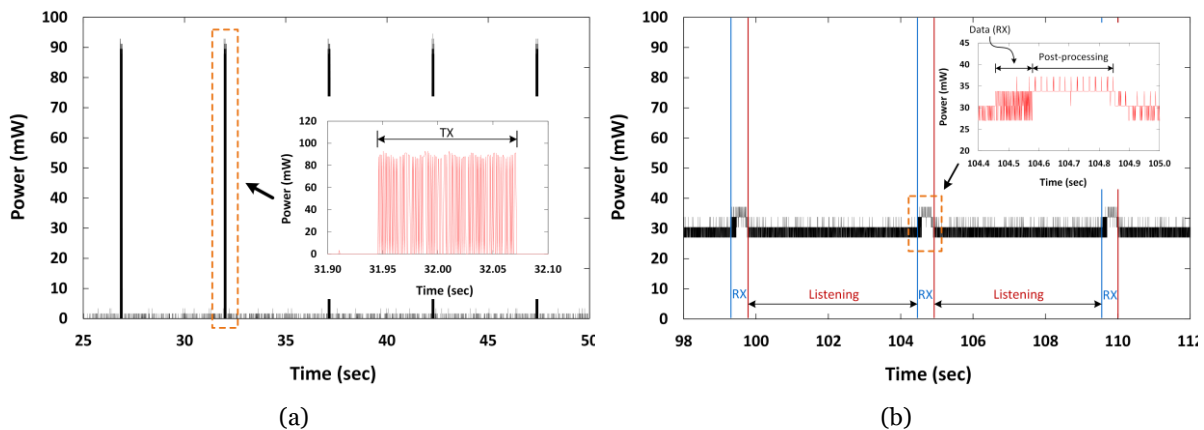
Figure 7. Power trace of an RF433: (a) transmitter module sending a 10 B message; (b) receiver module receiving the same amount of data.

The transmitter module consumes a small amount of power ($\approx$ 0.03 mW) when in standby and about 45 mW on average when transmitting. This is depicted by the spikes in Figure 7(a), indicative of its on-off keying modulation. Table 2 and Table 3 list the power draw and duration values of the RF433 modules. Because the Tx and Rx modules lack an integrated MCU, there is no substantial wake-up time for the transmitter and very little for the receiver (about 3 ms). The receiver module, however, maintains a steady but relatively higher power consumption level ($\approx$29 mW) when in standby (i.e. listening mode), as can be seen from Figure 7(b). The RF433 module has the longest over-the-air transmission time due to its low data transfer rate (9.6 kb/s max), which is several orders of magnitude less than that of the other RF modules. Furthermore, while the Tx and Rx duration are similar for the same data transfer,

the receiver remained at a higher power level for an additional duration of about 280 ms for post-processing or data verification.

Domestic Stock Control IoT Application – A Case Study

Application Architecture

The basic application architecture of the IoT is considered to include a large number of end devices (sensors and domestic appliances), communicating via a short-range wireless medium to a gateway device (a "home gateway") and being controlled by that gateway. In some cases, the home gateway will communicate through an external network with cloud-based applications. While the popular wireless interface options discussed in the preceding sections have propagation or performance characteristics that may be important in particular applications, the focus of this case study is on the energy efficiency of each option in the context of a domestic operational paradigm, with short messages communicated over a range of a few metres.

As an example, we consider a simple domestic stock control application, in which a toaster reports to the home gateway the number of bread slices it has toasted up to a particular time. Coupled with reporting of the bread quantity when it first enters the household (and other events), this could enable automated stock control and replenishment of bread supply by a commercial supplier. This application is simple enough for a straightforward analysis of energy consumption while encompassing all the major design features. Whilst it may seem to be an unusual choice, the toaster is considered the first connected domestic 'thing' in the IoT (Oweis *et al.*, 2016).

To support this application, a wireless communications interface must be added to the toaster, assuming that the home gateway is already equipped with suitable interfaces. The toaster interface will consume power in addition to the normal operation of the toaster and will, in some circumstances, add a few percentage points to the energy consumption over time. For this application, it is assumed that the gateway is within the 'smart kitchen' so that the communication range is a few metres at most. The toaster need only report the number of slices of toast in a given period, so the payload is only a few bytes. All the measured wireless options are capable of providing the necessary range and throughput. The energy consumption of the application over a period of one week is considered.

Communication Paradigms

There are three basic communications paradigms, which have different energy implications. These are:

(i)   Broadcast;

(ii)   Polling;

(iii)   Event-driven.

In each instance, we assume that the communications interface is in sleep or standby mode until it wakes up (or is woken up by a polling request), transmits its data, receives an acknowledgement (if possible), and goes back to sleep or standby mode.

Which of these communications paradigms uses the least energy for a specific wireless interface depends both on the relative energy usage of the interface in its various states and on the frequency of use, that is, on the traffic at the toaster. No one communications paradigm is optimal over the full traffic range.

Broadcast Mode

In Broadcast mode, the toaster reports its status (number of bread slices toasted) at regular intervals (e.g. every hour). The communications interface of the toaster will be active periodically, perhaps frequently. Because the application will require reliable communications between the toaster and the gateway, the communications interface will remain active after wake-up from its sleep or off state until an acknowledgement has been successfully received (if possible). It can then return to an inactive 'sleep' state until the next communications cycle.

In Broadcast mode, the natural cycle adopted for this application is once per hour: i.e. the stock control algorithm determines when more bread will be required and schedules a delivery at some future hour.

Polling Mode

In Polling mode, the gateway asks the toaster at particular times to report its status and the toaster does so. The communications interface should be in a listening state at all times and should become fully active when a polling request is received. When responding to a polling request, the communications interface is fully active until an acknowledgement has been successfully received. It can then return to a listening state.

The Polling mode is assumed to be governed by a standard stock-control application in which the economic order quantity (EoQ) is proportional to the square root of the demand, i.e. $EoQ = k\sqrt{D}$, where D is the demand and k is a constant (see, for example, Corke (1977)). Figure 8 shows a plot of polling instances against demand. The red curve is a scaled square root of demand and the blue curve is the average number of polling instances. Given the assumption of EoQ, the next polling event is set at one-half of the estimated time to exhaustion of bread supply, with the restriction that polling should be no more frequent than once per hour and no less frequent than once per day. (These constraints give rise to the steps in the figure.)
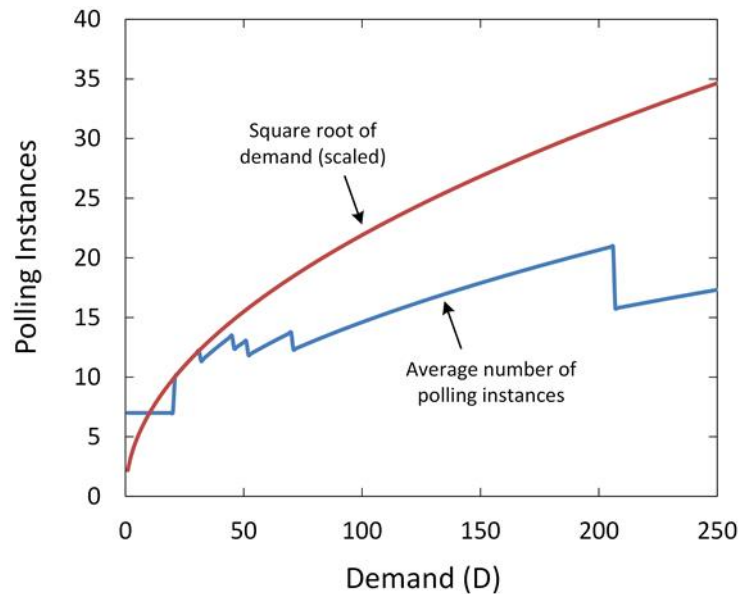
Figure 8. Plot of polling instances against demand.

This is a minimal polling design: a 'simpler' design (e.g. polling once per hour) could generate many more polling events. In this design, the toaster is polled only once per day for low numbers of toasting events. The number of polling events increases only as the square root of demand and, with a suitable value of k, can be kept below 25 polling events per week even for large numbers of toasting events.

Event-driven Mode

In Event-driven mode, the toaster reports only when a toasting operation begins. The toaster interface becomes fully active when a new slice of bread is introduced to the toaster. It remains active until an acknowledgement has been successfully received (or communication is deemed successful) before returning to inactive state. The number of events is usage driven.

Energy Consumption Model

The main operational modes of the wireless interfaces (RF modules) for which power consumption measurements were conducted are:

- **Inactive or sleep mode** – in cases without explicit sleep mode, this is when the interface is turned off;
- **Wake-up mode** – Transitions between sleep and active and from active to inactive;
- **Listening or standby mode** – when the interface can detect and receive a message but may not be able to transmit;
- **Transmit (Tx) mode** – when the interface is transmitting data;
- **Receive (Rx) mode** – when the interface is receiving data.

The total energy consumption, E_{total}, of the RF module in a given period (in this case, one week) in which there are N communication events is the sum of the energy consumed in each event – the wake-up (E_{wu}), transmit (E_{tx}) and receive (E_{rx}) energies – plus the energy consumed in the sleep or standby states (E_{sby}). This can be expressed as:

$$E_{\text{total}} = E_{\text{sby}} + N\left(E_{\text{wu}} + E_{\text{tx}} + E_{\text{rx}}\right) \tag{1}$$

or
$$E_{\text{total}} = P_{\text{sby}}D_{\text{sby}} + N\left(P_{\text{wu}}D_{\text{wu}} + P_{\text{tx}}D_{\text{tx}} + P_{\text{rx}}D_{\text{rx}}\right) \tag{2}$$

where P_{sby}, P_{wu}, P_{tx} and P_{rx} are the power consumption of the RF module in the standby, wake-up, Tx and Rx modes and D_{sby}, D_{wu}, D_{tx} and D_{rx} are their respective durations. N is the number of communications: i.e. number of events, broadcasts or polling requests.

Energy Consumption Calculation

We use the power draw and duration measurements given in Table 2 and Table 3 for each RF module. The calculated energy is the product of the power draw and duration.

With polling communication (home gateway to end-device), the RF module remains in standby state in order to detect incoming packets. Therefore, BT (in sniff mode) and RF433 with no sleep-mode capability will stay in the listening mode when not communicating. In a weekly cycle, the time taken in listening mode is the total weekly period (604,800 seconds) less the time for all transmissions. BLE, Wi-Fi and ZigBee do have sleep-mode options. However, only ZigBee with its reverse polling technique can utilize its sleep-mode (end-device to home gateway), sending requests to its coordinator/master at regular intervals after wake-up. BLE and Wi-Fi, at minimum, require their receivers to remain active, which prevents an effective sleep-mode.

When designed for broadcast-only communication, a 10 B payload is sent by the end-device (toaster) once every hour. Consequently, there will be 168 transmissions ($N = 168$) per week irrespective of the frequency of toast slices made.

In an event-driven communication, however, the energy consumed by the modules scales linearly with the number N of toast events. The energy consumption of a BT module is greatly influenced by the time required (see Table 3) to power on and re-establish connection with the master/coordinator. An average of 640 ms was measured for a complete inquiry scan and advertisement (stages of BT protocol), assuming that the interface is completely turned off at the end of each transmission. It takes an average of 1836 ms for the BT wake-up phase (including turn on, inquiry scan, paging and synchronization) during which the HC-06 consumes an average of 96 mW. Similarly, the Wi-Fi module spends about ≈ 5.5 ms in the wake-up phase but with much higher average power consumption (412 mW).

Processing and Storage Energy

A simple IoT hardware application like the toaster example requires some processing and storage of data bits. Although the main focus of this work is on the measurement and assessment of the communication energy consumption, an IoT toaster needs some memory to store toast counts, which can be incremented step-wise until its reset point. It is therefore necessary to consider its potential processing and storage energy usage.

For the same MCU and clock crystals on a circuit board, the processing energy usage may be small but similar for the three communication methods. With regard to storage, while an event-driven method may not necessarily require storage, the broadcast and polling methods do. Recent RF modules are commonly designed as a SoC, with a few kilobytes or megabytes of static random access memory (SRAM) or NAND flash memory. For example, the toaster could be designed with a 32-bit ARM Cortex M4F MCU (as in Nordic (2017)), which has 64 KB SRAM and 512 KB flash memory. The M4F consumes 1.2 μA (2V operating voltage) in a low-power state with full SRAM retention. Hence, a toaster may consume about 1.5 J per week to retain the toast count data of less than 10 B, which is far smaller than the capacity of the SRAM. Utilizing flash memory, on the other hand, does not require power to retain data (non-volatile). A study by Lee & Chang (2003) shows that a NAND flash memory consumes 4.72 μJ and 38.04 μJ to read and write data, respectively, to a 2 KB memory page within a 128 KB memory block. For 250 read/write operations, a flash memory may consume as much as 10.7 mJ of energy. The choice of memory is dependent on the type of application, the frequency of data requests and amount of data per operational duty cycle.

We can, however, conclude that the energy use for running the processor and storing or retaining the data bits is small and can be ignored if flash memory is utilized for this application.

Comparison of Energy Consumption

The plots in Figure 9 show the weekly energy usage comparison of BT, BLE, ZigBee, Wi-Fi and RF433 for the three communication paradigms considered in this study. For event-driven and broadcast modes, it is assumed that the interface is turned off between events. For the event-driven and broadcast modes, the BT module is the most energy hungry, while Wi-Fi is most energy demanding for Polling mode. BLE is the most energy-efficient for all but one (i.e. Polling) mode. ZigBee is more energy-efficient than BLE for polling due to its reverse polling capability. This allows the ZigBee module to sleep for longer periods while BLE periodically listens to the home gateway for polling requests. This result demonstrates an incentive for the communication to be driven by the end-device.
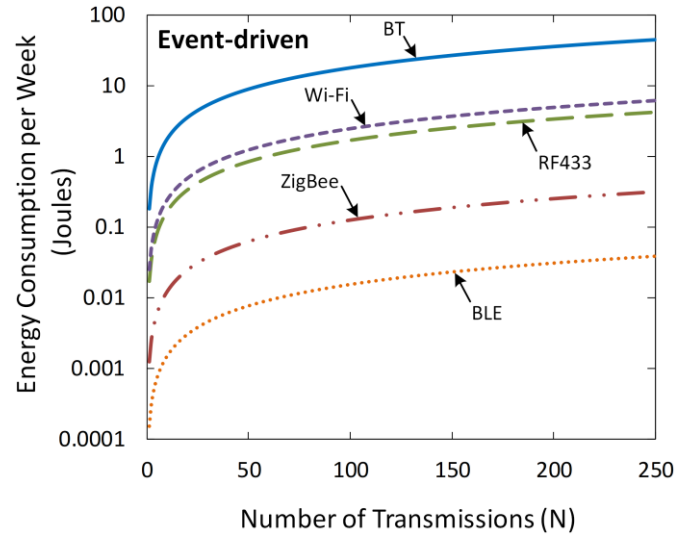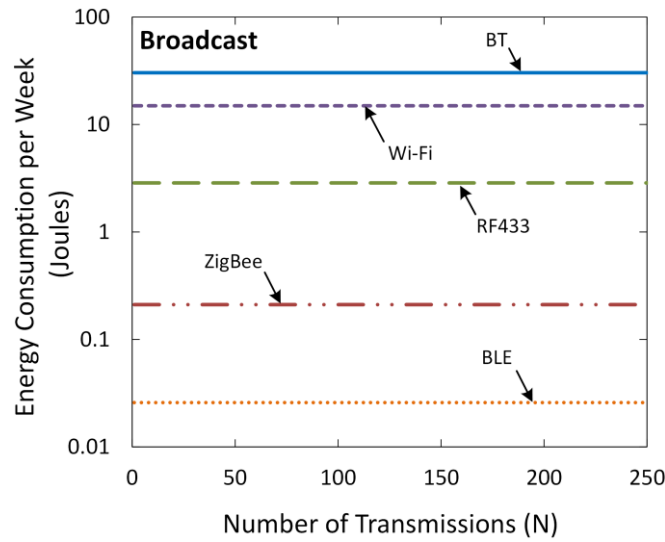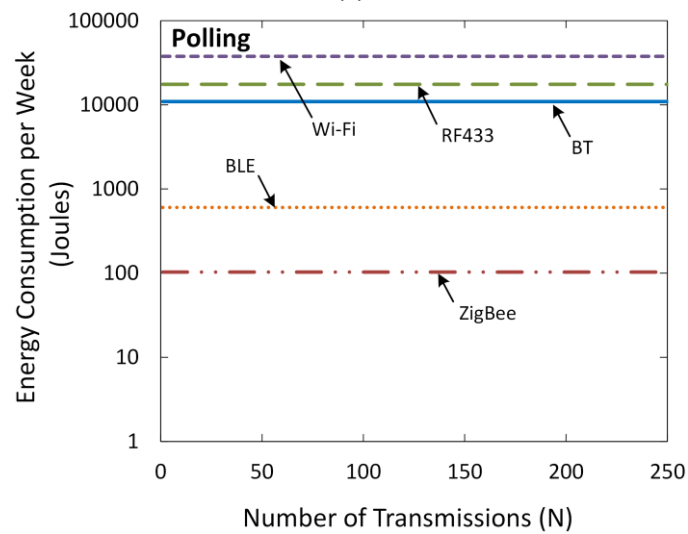
Figure 9. Energy consumption per week for BT, BLE, ZigBee, Wi-Fi and RF433 using communications paradigms (a) Event-driven, (b) Broadcast, and (c) Polling.

Whilst it is energy-wise (visibly from Figure 9(a)) to choose event-driven mode for lower usage rates, the broadcast mode is more energy-efficient beyond 168 servings per week. The energy usage of polling mode, however, is several orders of magnitude higher than both broadcast and event-driven modes, since the RF modules must stay in standby mode to receive incoming polling requests from the home gateway. Note that the variation of the polling energy numbers is masked on the plot due to the use of a logarithmic scale.

As an example, if toasting 20 slices of bread in a week, the BT interface, configured in polling, broadcast or event-driven mode, will consume 10.9 kJ, 30.3 J or 3.6 J, respectively. Hence polling mode uses just over 3000 times more energy (~10000 times more for Wi-Fi at 38 kJ) than event-driven mode. To put this in perspective, an 800 W toaster will consume 72 kJ of energy in one toasting operation (1.5 minutes average toasting time). Hence, the weekly energy consumption of an always-on Wi-Fi interface added to a toaster, is equivalent to about 5% of the total energy consumed for toasting 20 slices of bread for a week, assuming 2 slices per toasting event.

Energy-Efficient Design

This section draws some conclusions about energy-efficient design using the application architecture from the above section and the power measurements reported in Table 2. In particular, we show that the most energy-efficient communications paradigm for each wireless interface depends on the number of events reported (or amount of traffic) over the interface.

Interfaces Always On

The comparisons in the preceding section indicated that polling is not an energy-efficient option. This assumed, however, that the interfaces would be turned off between events in event-driven and broadcast modes. A naive design would just add a standard wireless interface to an end-device and assume that the interface is always on. Figure 10 shows the total energy per week used by an RF433 interface with the three communications paradigms if the RF module stays on the entire time. In this case (and in general, if the interfaces are always on), event-driven communication is preferred in the case where the number of transmission events is very small. Once the number of events exceeds a threshold (about 7 for this example), polling mode is more energy efficient. Broadcast mode will eventually become the most energy efficient mode when the number of servings is very large (not shown in the figure).

Although this is a somewhat contrived example, it shows that the energy-efficient design of the local wireless communications for the Internet of Things can depend on the choice of wireless interface, the number of communications events and the amount of traffic (or application) to be carried.
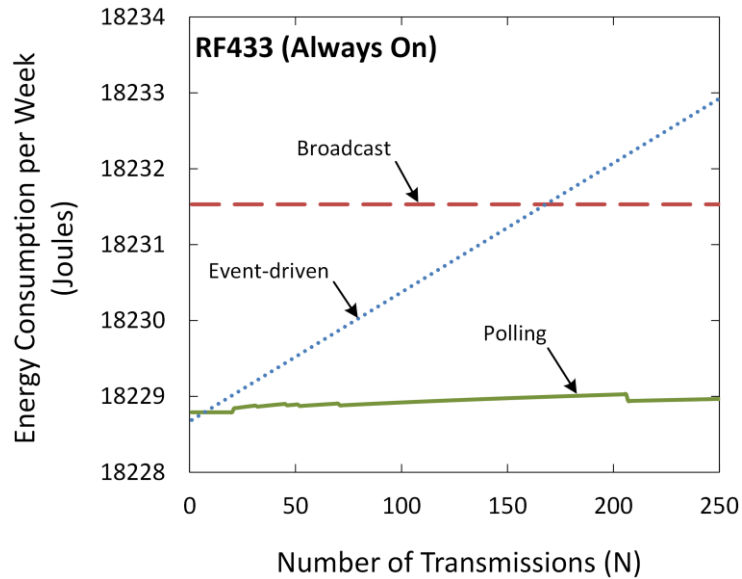
Figure 10. Energy consumption per week of an RF433 interface (Always On).

Interfaces Powered Down When Not In Use

For most traffic levels, sleep mode is important for energy-efficient communications. For very high traffic levels, always-on polling will eventually be preferable to event-driven communication but, in this case, broadcast will be the most energy-efficient paradigm. Figure 11 shows the total weekly energy used by a BT interface when toggling on/off in the broadcast and event-driven modes. (Note the logarithmic scale in order to accommodate all three communication modes.) In this case, the event-driven mode is preferred until the amount of traffic exceeds that of the constant hourly broadcast, after which broadcast mode is preferred.
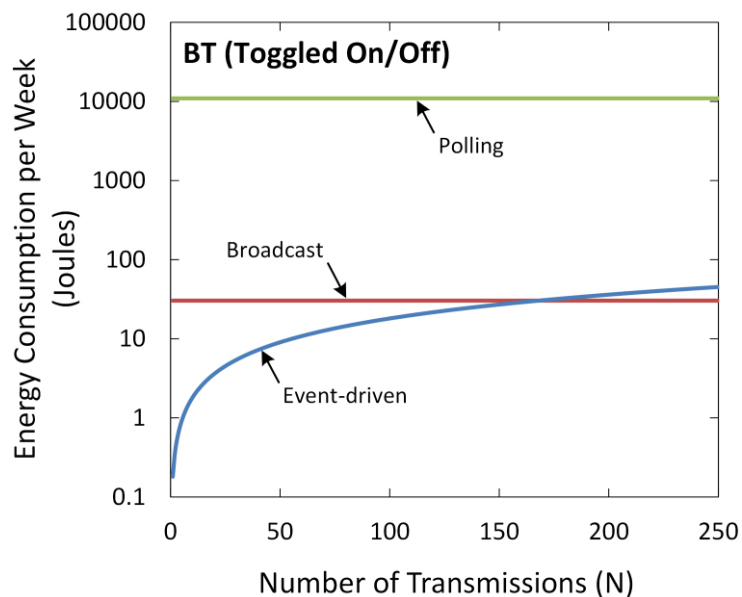
Figure 11. Energy consumption per week of the BT interface (toggled on/off).

Conclusions

In this paper we have measured the energy consumption of representative samples of five popular wireless interfaces when they are in their inactive, listening, transmitting and receiving states. Using these measured values, we have calculated the total energy usage for three different communications paradigms – broadcast, polling and event-driven, in one (somewhat undemanding) application. In each case, it has been shown that the most energy-efficient communications paradigm depends both on the relative energy usage by the interface in its various modes and on the frequency and volume of traffic transmitted over the interface.

For any particular IoT application, the designer has a range of choices in terms of the communications paradigm and of the communications hardware to employ. This paper has shown that it is important to consider the interplay between these and the design of the application itself, in order to achieve an energy-efficient solution.

Our results suggest that, for an IoT application developer, careful consideration of the applications to be run over the wireless interfaces will be required to determine an energy-efficient design. In many cases, the level of traffic will be uncertain (or will change over time). The results suggest that applications should be designed to adapt to traffic levels by selecting a different communications paradigm when it becomes more energy efficient to do so. It would be beneficial if IoT communications protocols were easily adaptable for changes in communications paradigm or, equivalently, traffic levels.

Note also that a least-capital-cost solution may not be the most energy efficient (and hence the least cost over the lifetime of the interface). Bluetooth and Wi-Fi interfaces, for example, may be lowest cost and most readily available because of their volume production, but may consume more energy than alternatives. Their many communications features may also be of no benefit for the specific application for which the communications interface has been added. A design process that takes into account total cost of ownership (including both initial and operating costs) will be preferable.

In terms of the technology choices themselves, it is clear from the values in Table 2 and Table 3 that BLE and ZigBee provide significant energy savings over classic Bluetooth and Wi-Fi. In our measurements, BLE outperforms ZigBee in all instances, suggesting that BLE should always be preferred for the type of single-hop, short-range communication considered here. RF433 remains a competitive low-energy option if the interface can be in Sleep mode most of the time (as in the case, for example, of a temperature sensor). Wi-Fi and Bluetooth, on the other hand, should only be used if their range or other characteristics are required for the application under consideration.

Whilst the actual devices measured are representative of what is on the market today, these consumption values will most likely change over time. That notwithstanding, the principles outlined here can be employed in designing efficient interfaces for IoT appliances.

Acknowledgement

References

Arduino (2009). Arduino Duemilanove. Available from: https://www.arduino.cc/en/Main/arduinoBoardDuemilanove

Atzori, L., Iera, A. & Morabito, G. (2010). The Internet of Things: A survey. *Computer Networks, 54*(15), 2787-2805.

Bluetooth-SIG. (2010). Bluetooth Core Specification 4.0 - Bluetooth Low Energy. Available from https://www.bluetooth.com/specifications/archived-specifications .

Bluetooth-SIG. (2019). Bluetooth Core Specification 5.1. Available from https://www.bluetooth.com/specifications/bluetooth-core-specification .

Corke, D. K. (1977). *Production control in engineering*. London: Edward Arnold.

Dementyev, A., Hodges, S., Taylor, S. & Smith, J (2013). Power consumption analysis of Bluetooth Low Energy, ZigBee and ANT sensor nodes in a cyclic sleep scenario. *2013 IEEE International Wireless Symposium (IWS)*, April, Beijing, China.

Digi (2011). XBee & XBee-PRO ZB**:** ZigBee Embedded RF Module Family for OEMs. Available from: https://www.digi.com/hottag?ht=/pdf/ds_xbee_zigbee.pdf

Ferro, E. & Potorti, F. (2005). Bluetooth and Wi-Fi wireless protocols: a survey and a comparison. *Wireless Communications, IEEE, 12*(1): 12-26.

Gray, C. & Campbell, L. (2016). Should my toaster be polled? Towards an energy-efficient Internet of Things. *2016 26th International Telecommunication Networks and Applications Conference (ITNAC)*, Dunedin, New Zealand.

Hsu, C. F., Liao, H. Y. M., Hsiu, P. C., Lin, Y. S., Shih, C. S., Kuo, T. W., & Liu, J. W. (2006). Smart pantries for homes. *2006 IEEE International Conference on Systems, Man and Cybernetics*, *5*, 4276-4283, October.

James, R. (2014). The Internet of Things: A Study in Hype, Reality, Disruption and Growth. Available from: http://sitic.org/wp-content/uploads/The-Internet-of-Things-A-Studyin-Hype-Reality-Disruption-and-Growth.pdf .

Jin-Shyan, L., Yu-Wei, S. & Chung-Chou, S. (2007). A Comparative Study of Wireless Protocols: Bluetooth, UWB, ZigBee, and Wi-Fi. *IECON 2007. 33rd Annual Conference of the IEEE Industrial Electronics Society, 2007.*

Lee, H. G. & Chang, N. (2003). Energy-aware memory allocation in heterogeneous non-volatile memory systems. *Proceedings of the 2003 International Symposium on Low Power Electronics and Design*, 420-423. ACM, August.

Mackensen, E., Lai, M. & Wendt, T. M. (2012). Performance analysis of a Bluetooth Low Energy sensor system. *2012 IEEE 1st International Symposium on Wireless Systems (IDAACS-SWS).*

Mikhaylov, K., Plevritakis, N. & Tervonen, J. (2013). Performance analysis and comparison of Bluetooth Low Energy with IEEE 802.15. 4 and SimpliciTI. *Journal of Sensor and Actuator Networks, 2*(3), 589-613.

Nordic (2017). Nordic nRF52832 Specification. Available from: http://infocenter.nordicsemi.com/pdf/nRF52832_PS_v1.4.pdf

OnWorld. (2017). Bluetooth Low Energy IoT: A Market Dynamics Report. Available from https://www.researchandmarkets.com/research/45cbbg/bluetooth_low .

OnWorld. (2018). 802.15.4 IoT Markets: Zigbee, Thread, 6LoWPAN, Wi-SUN and Others. Available from https://onworld.com/research/zigbeealliance/vip/ .

Oweis, N. E., Aracenay, C., George, W., Oweis, M., Soori, H., & Snasel, V. (2016). Internet of Things: Overview, Sources, Applications and Challenges. *Proceedings of the Second International Afro-European Conference for Industrial Advancement AECIA 2015*, 57-67. Springer, Cham.

Pal Amutha, K., Sethukkarasi, C. & Pitchiah, R. (2012). Smart Kitchen Cabinet for Aware Home. *SMART 2012, The First International Conference on Smart Systems, Devices and Technologies*, Stuttgart, Germany.

Perahia, E. & Stacey, R. (2013). *Next Generation Wireless LANS: 802.11 n and 802.11 ac.* Cambridge: Cambridge University Press.

Redbear. (2018). Redbear Nano v2. Available from: https://redbear.cc/product/ble/ble-nano-2.html

Shahzad, K. & Oelmann, B. (2014). A comparative study of in-sensor processing vs. raw data transmission using ZigBee, BLE and Wi-Fi for data intensive monitoring applications. *2014 11th International Symposium on Wireless Communications Systems (ISWCS).*

Siekkinen, M., Hiienkari, M., Nurminen, J. K. & Nieminen, J. (2012). How low energy is bluetooth low energy? Comparative measurements with ZigBee/802.15.4. *Wireless Communications and Networking Conference Workshops (WCNCW)*, IEEE.

Weldon, M. K. (2016). *The Future X Network: a Bell Labs Perspective*. CRC press.

Yang, H., Sawhney, R., Zhang, G., Marella, L. & Han, Z. (2014). Using Smart Kitchen for grocery purchase prediction. *Proceedings of the 2014 Industrial and Systems Engineering Research Conference.*

ZigBee Alliance. (2012). ZigBee Specification (IEEE 802.15.4). Available from: https://www.zigbee.org/download/standards-zigbee-specification/#

Dynamic Vehicular Traffic Load: Definition and Quantification

Gerald Ostermayer
University of Applied Sciences Upper Austria

Christian Backfrieder
University of Applied Sciences Upper Austria

Manuel Lindorfer
University of Applied Sciences Upper Austria

Abstract: In this paper, we introduce a method that quantifies the amount of traffic over time by the help of a cloud calculation service and vehicular communication. Furthermore, the approach is applicable also in vehicular traffic simulations, which are widely used in research to demonstrate the effects of proposed solutions to traffic problems. As unused road segments strongly influence the overall traffic load (i.e. used vs full road capacity), we propose a methodology that dynamically calculates the load over time and considers whether specific parts of the road network are used. We introduce two possibilities to filter out distortion of the created statistics due to variation in usage over time. Our novel approach is both simple but widely configurable to fit individual needs. The approach is proven by simulations and application of the load calculation in combination with an intelligent route optimization approach by comparing the optimization gain with the calculated traffic load.

Keywords: vehicular traffic load quantification, cloud service, vehicular communication, vehicular traffic.

Introduction

Investigations in connection with vehicular traffic are an enormously wide field in current and past research. The necessity of analysing various kinds of challenges in this field is ever increasing. These challenges are multifaceted, ranging from optimization of traffic flow, investigations regarding connected vehicles and vehicular communication, analysis of environmental effects resulting from traffic, case studies before road works and many other applications. Out of these, a huge portion depends on traffic flow and has strong relations to traffic density, as elucidated in the following.

Firstly, congestion is one of the biggest economical, societal and environmental problems related to transportation, both in industrialized and lower developed countries (Hogendoorn & Bovy, 2001). Nowadays, traffic jams are a major problem throughout the world, especially in urban and metropolitan areas, which increases the demand for emphasizing research in this field. Mostly, the proposed approaches are evaluated and proven by simulations, and their performance depends on the traffic density (Boskovich & Barth, 2013; Mandziuk & Swiechowski, 2016; Horvitz *et al.*, 2009; Chan, Dillon & Chang, 2013; Nafi *et al.*, 2014; Backfrieder, Ostermayer & Mecklenbräuker, 2017). Secondly, research regarding vehicular communication, such as Vehicle-to-Vehicle (V2V) or Vehicle-to-Infrastructure (V2I) communication, requires simulations in order to evaluate the developed methodology (Elbery *et al.*, 2015; Kumar & Chhabra, 2016; Zhang *et al.*, 2017). The reliability of communication systems is strongly dependent on channel quality, load, number of participants and their distances between each other (Blazek *et al.*, 2017); the number of participants and the distances between them can again be summarized as vehicle density. Thirdly, the investigation of environmental effects resulting from intensive vehicular traffic is in some respects related to traffic optimization. In order to determine or estimate local concentrations of air pollutants such as CO_2 or NO_x as major human health concerns (An *et al.*, 2013), measurements in combination with simulations can provide a means to achieve this goal (Jing *et al.*, 2016). The knowledge gained can help to assess the need for regulation in terms of speed limits or driving bans in order to diminish the environmental burden. Emissions depend on multiple influences, including the current weather conditions, average vehicle speed, the composition of vehicles, characteristics of the road and traffic volume, as well as traffic density (Zhou *et al.*, 2015; Stevanovic *et al.*, 2009; Cappiello *et al.*, 2002). Fourthly, studies to investigate effects of upcoming road works in terms of capacity expansion or temporary facilities, like road or lane closures, are common in transportation planning (Chen, Subprasom & Ji, 2006; *Szeto et al.*, 2015) and, among other things, also depend on traffic density (Abadi, Rajabioun & Ioannou, 2015). Finally, of course, combinations of the above are also required: e.g., making use of vehicular communication in order to optimize traffic or to broadcast important safety information.

These mentioned research applications regarding vehicular traffic in the widest possible sense mostly have in common that the research goal is dependent on traffic volume or density, which is, of course, not limited to the topics above. The number of involved vehicles and the local concentration of those is of high interest. Of course, this is true for both simulation environments and real-world scenarios. As elucidated above, a huge number of investigations apply traffic simulations due to the infeasibility or impossibility of real-world experiments. The simulation results demonstrate the applicability or performance of the presented

approach or methodology and try to represent the reality in an appropriate way, precisely enough and, at the same time, as simply as possible to fulfil the research goal. This further means that a certain number of involved vehicles is modelled within a defined road network extract, and movement of those vehicles is simulated, or observed in the case of a real-world experiment. This is not only true for static situations but also most commonly for dynamic (in the sense of time) situations. In other words, a defined time interval is investigated in order to answer the research question.

This strengthens the need for a versatile methodology to quantify traffic density over time, with special focus on the applicability in the real world but also within microscopic vehicular traffic simulations. We introduce a novel approach to fulfil this goal, which is presented and evaluated by microscopic vehicular traffic simulations. We particularly emphasize continuous calculation, so that a traffic density classification number is available at any time during operation. The required information, such as the current vehicle positions, are reported continuously to the introduced cloud service by applying V2I communication. In addition, a single classification number can be calculated for an arbitrary time interval.

The remainder of this paper is organized as follows. In the next section, "Related Work", we analyse literature, present some related work and demonstrate how the introduced approach differs from existing work. Section "Methodology" describes the methodology of load calculation and how the area of interest (i.e. the relevant road network extract) is adapted to the usage over time. In section "Evaluation by Microscopic Traffic Simulations", we evaluate the approach by comparing the calculated load quantification to the achieved improvements of an existing traffic optimization approach. The last section, "Summary and Conclusions", summarizes the findings and concludes the paper.

Related Work

Without any doubt, research on the subject of traffic density is not a new topic in the literature. It started more than 60 years ago, when Lighthill and Whitham (1955) presented an analogy between vehicles in a traffic flow and particles in a fluid. Also, mathematical specification of the terminology was made very early (Gazis & Edie, 1968). Traffic on the road is often described to be comparable to liquids from the perspective of transportation engineers (Aw & Rascle, 2000). Based on that, the description of traffic density ρ (or also often denoted as k) can be estimated using the hydrodynamic relation between flow (or volume) in vehicles per time and average speed, according to equation (1) (Treiber & Kesting, 2013):

$$\rho = \frac{q}{v} = \frac{\text{flow}}{\text{speed}} \qquad (1)$$

However, as the density is defined as a spatial quantity, the time-based averaging from detectors on the road induces errors. The reason is that vehicles with higher speeds are seen more frequently than slower ones, leading to an unwanted bias for the estimated density if speeds are divergent (Treiber & Kesting, 2013, especially pp. 13-24). This further means that the relationship in equation (1) will hold only for homogeneous traffic flows with low speed and dimension variation among the vehicles. For this reason, the application of the basic thermodynamic relation is unfavourable. Other studies represent the density as the number of vehicles per kilometre of road related to their speeds (Greenshields, 1935) or, as a refinement of that concept, a dimensionless quantity describing the percentage of time during which a certain point on the road is covered by any vehicle (Athol, 1965). An even more precise method is based on overall space occupied by vehicles on the road considering vehicle dimensions. Furthermore, this quantification of road network occupancy is independent of the vehicle speeds, which leads to a better estimate and therefore is a widely used method of measuring this quantity (Treiber & Kesting, 2013; Cassidy & Coifman, 1997; Ni, 2015).

Studies in the field of transportation and vehicular systems do not only require the definition of a certain traffic density at a certain point in time or at a specific location, but can also be influenced by the amount and density of involved vehicles over a longer time span and a greater area. Applications cover simulations of Vehicular Ad-hoc Networks (VANETs), where packet loss due to the varying locations and distances between vehicles plays a major role (Blazek *et al.*, 2017; Nebbou, Fouchal & Lehsaini, 2017; Abbas *et al.*, 2015). A different major field of application of network occupancy over time is traffic optimization and intelligent routing (Backfrieder, Ostermayer & Mecklenbräuker, 2017; Pan *et al.*, 2012) or modelling of driver behaviour (Lindorfer, Mecklenbräuker & Ostermayer, 2017). Naturally, the possible improvements and potential of the approach depend on the severity of emerging congestion, which further is a direct consequence of the amount of and local concentration of affected vehicles.

Existing approaches for traffic optimization obviously focus on the applied methodology itself and, however, point out the relation to traffic density, but rather in a way that describes it as the number of considered vehicles within a certain time interval. In other words, the traffic density is described statically rather than in a dynamic manner. Nevertheless, conditions on the road can vary over time. Likewise, the currently used part of the road network changes, which complicates the description of a globally valid measure of the density or, more precisely, the occupancy or load factor of the network. However, exactly this understanding of the observed traffic needs to be of high reliability and quality. The goal of this work is to introduce a general and configurable method that provides a measure for the network load in concrete terms for the network part of interest continuously, and is applicable for various kinds of

microscopic vehicular traffic simulations, independent of the specific investigation target. In addition, the methodology is applicable in real environments, if vehicles are equipped with communication devices that transmit the required data to the cloud, where a load classification number can be calculated, as described in the next section.

Methodology

Basically, the position and geometric properties (especially length) as well as space headway to the preceding vehicle must be available in the desired temporal resolution of the load quantification. Built-in Global Positioning System (GPS) and displacement sensors are used to sense the required information. With the help of vehicular communication (commonly abbreviated as Vehicle-to-Everything (V2X)), the data is transmitted anonymised in regular intervals by each individual vehicle to a cloud service, which is aware of the road network geometry and characteristics. For the scope of the proposed methodology, we do not model the communication separately, but assume that a working infrastructure such as Long Term Evolution (LTE) can be used. The communication delay is not very critical for applications of this kind, as we already analysed in Backfrieder *et al.* (2018).

In microscopic simulation environments, each vehicle is represented by a separately modelled entity, as is also the case for real environments. Time and space headways are available precisely and at any arbitrary time, which makes it easier to provide the required information. In principle, the source of vehicle information for the service which calculates the desired traffic load quantification number is irrelevant. Hence, it can run anywhere in the cloud, provided that the required data is available there, be it from sensors on the vehicles or just taken from a traffic simulation. Especially in real conditions, vehicles naturally have different characteristics in terms of dimensions and performance features, in contrast to homogeneous traffic in macroscopic modelling. However, as described earlier in section "Related Work", in such heterogeneous environments some simplifications regarding density are not allowed in order to enable precise calculation of the network load. Thus, only microscopic simulations provide the necessary detail to allow a more precise definition of a description for road network occupancy based on used road space.

The relevant road networks need to be modelled as an omnidirectional graph, consisting of edges (road segments) and vertices (intersections), which is the case in common microscopic traffic simulators (Backfrieder, Ostermayer & Mecklenbräuker, 2014; Behrisch *et al.*, 2011). In the case of real-world applications, the network structure can be extracted from free and open services like OpenStreetMap (OSM), if not already available to the operator of the cloud service. Due to the fact that it is a common use case for investigations in the field of road traffic to constrain the area to a certain region of interest, the proposed methodology allows the

calculation of the occupancy for a defined extract of the road network on a road segment basis, as follows.

Load calculation

The cloud service calculates the load L for a single road segment R at time t as defined in equation (2), considering the number of vehicles on this segment (n_v), the specific length s_i and space headway h_i of vehicle i. A road segment is basically defined as the part of a network between two intersections, although such a connection can be split up into more than one edge if a finer spatial resolution of the load number is required.

$$L(R,t) = \frac{\sum_{i=1}^{n_v}(s_i + h_i)}{\text{length}(R)} \tag{2}$$

This represents the load for one single segment R by the share of occupied space at time t. The term length(R) includes all lanes on the referenced road segment, so, for example, if there are two lanes on this segment, the segment length needs to be doubled, because consequently also the space is doubled.

The result for $L(R,t)$ is within the interval [0, 1], where the lower limit represents a situation with a completely free segment. The upper limit means total standstill, since this is the only case where the minimum space headway can be reached in common longitudinal models (Treiber & Kesting, 2013). Figure 1 depicts an example of a road segment where four vehicles $v1$ to $v4$ are currently driving through. The coloured areas mark the lengths required by each vehicle, including the length s of the vehicle and the specific minimum headway h, separated by a dashed line (highlighted for vehicle $v2$), as defined in the respective longitudinal model.
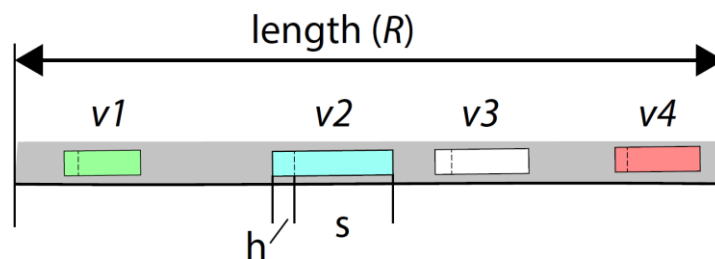
Figure 1. Single road segment containing vehicles with specific length *s* and headway *h*

As the parameter t in equation (2) indicates, the calculated load is a snapshot at a specific time. To a greater degree, the goal is to provide statistics for a longer time period $\Delta t = t_2 - t_1$, which is calculated by an averaging function over time, according to equation (3), representing a simple moving average (SMA), or equation (4), representing an exponential moving average (EMA) function. The latter provides higher weights for more recent values and is also referred to as the exponentially weighted moving average, where m represents the weighting multiplier, index j denotes the current time index, and EMA$_{\Delta t}$ represents the current average value for the whole observation interval Δt (consisting of S values). S denotes the number of time steps

under observation. The resolution r can be defined arbitrarily in real-world observations or for time-continuous simulators. In simulation environments, however, time-discrete simulators in a microscopic manner are more popular, which simplifies the selection of r, since the time step size of the simulation can be chosen.

$$L(R, \Delta t)^{[\text{SMA}]} = \frac{\sum_{t=t_1}^{t_2} L(R,t)}{S} \tag{3}$$

or
$$L(R, t_j)^{[\text{EMA}]} = \left(L(R, t_{j-1}) - \text{EMA}_{\Delta t}\right) \cdot m + \text{EMA}_{\Delta t} \tag{4}$$

where
$$S = \frac{t_2 - t_1}{r} \tag{5}$$

and
$$m = \frac{2}{S+1} \tag{6}$$

With equation (3), we have the load statistics over the time interval Δt for a specific road segment R. When observing or simulating vehicular road traffic, a common use case is to have a larger road network region of interest, naturally consisting of multiple road segments and intersections. The simplest solution to calculate a load $L(N, \Delta t)$ for this network part (denoted as N) is again to average the gained load values $L(R, \Delta t)$ over the total number of road segments n, weighted by their length, as equation (7) elucidates:

$$L(N, \Delta t) = \frac{\sum_{i=1}^{n} L(R_i, \Delta t) \cdot \text{length}(R_i)}{\sum_{i=1}^{n} \text{length}(R_i)} \tag{7}$$

By that means, a full load for a defined extract of a road network can be calculated for a specific time interval. For some applications this may be sufficient, and thus can be applied without hesitation. The problem is that $L(N, \Delta t)$ tends to become more and more meaningless to a certain degree when investigating the effects of a specific action between individual observations. One could consider the variations between several days as several observations in the case of real environments, or likewise multiple simulation runs in the case of simulation environments. In each observation, the traffic density or load may be distinct. This is especially the case if the extract of the road network to consider for this calculation differs between single observations: i.e., some of the road segments are not used by any driver for a longer time in one observation, while this does not happen in another one. In other words, this happens especially when the observed properties are subject to influence by the network load and the routes of the vehicles are variable, which is true for all kinds of investigations concerning intelligent routing, dynamic route calculation and other intelligent traffic management strategies. One solution could be to select only those road segments which are used by any vehicle at any time. But, then, a single vehicle using a specific route in one observation that does not do so in the next observation would influence the network load at each time step, and hence of course also the overall result. For this reason, road segments which are unused at

least during part of the observation time should be considered only temporarily (as described in the next subsection) when calculating the network load.

Time-dynamic selection of area of interest

In order to diminish the problems connected with consideration of a static selection of road segments, the segments which are included in the calculation are selected based upon their utilization. A road segment is only considered for calculation of the network load $L(N, \Delta t)$ if the inequality $L(R, t) > 0$ is satisfied. If a specific road segment is not considered, i.e. becomes inactive, of course its length is also not considered any more in equation (7). Mathematically expressed, a road segment is not considered if the number of vehicles on it is equal to 0 for at least the time period Δt, consisting of S intervals (see equations (3) to (5)).
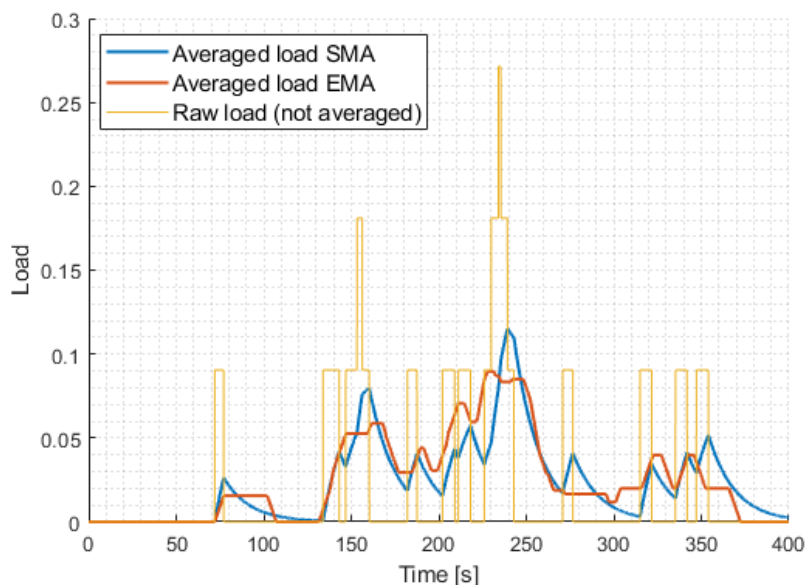
Figure 2. Load by application of simple and exponential moving averages for a single road segment

Figure 2 reveals the difference between SMA and EMA averaging for a single road segment and a total moving average length of 30 seconds. The different weighting of events become obvious therein, e.g. when focusing at second 72. The first vehicle enters the Road Segment (RS) and leaves at second 77, which causes a greater gradient for the EMA curve, which exponentially decreases in contrast to the SMA method, where the resulting average stays at a constant level for the time of the averaging interval. What is essential here is the fact that a single road segment disappears from the overall average as its particular load function result drops down to zero. Thus, it is irrelevant for the overall load quantification value for the unused periods in time.

Subsequently, this process is applied to all road segments and further leads to a single quantification value of the network load at each time interval and, according to equation (7), also for the whole observed network region. Depending on the needs of the application, the

model self-evidently can be parameterized with different averaging lengths, and of course also with an arbitrary averaging method.

Evaluation by Microscopic Traffic Simulations

Undoubtedly, real-world experiments in the field of vehicular traffic with numerous involved vehicles are very costly, if not impossible, and definitely would exceed the budget of our research project. As already mentioned earlier, the cloud service which calculates the traffic load is independent of the source of information. In other words, it is irrelevant whether the data is provided by sensors of vehicles that are currently on the road and transmitted via V2X communication or by a microscopic traffic simulator. Therefore, we selected an example dealing with traffic optimization by intelligent vehicle rerouting in a microscopic simulation environment in order to demonstrate the practical applicability of the introduced methodology. The applied traffic optimization algorithm is introduced in one of our previous publications (Backfrieder, Ostermayer & Mecklenbräuker, 2017) and presents an approach to relieve congestion by traffic jam predictions, called Predictive Congestion Minimization in combination with an A*-based Router (PCMA*). Affected vehicles can then be rerouted by assigning alternative routes with free space in a timely manner and thus bypass congestion, which is proven by simulations with the microscopic traffic simulator, TraffSim (Backfrieder, Ostermayer & Mecklenbräuker, 2014). The performance of PCMA* is strongly dependent on the traffic density as well. Currently, it quantifies the traffic density by a combination of the overall number of simulated vehicles and the vehicle Inter-Arrival Interval (IAI), i.e. the static time interval between the trip start of two consecutively starting vehicles. The performance of PCMA* is measured by average travel time per vehicle, assuming all vehicles have defined immutable origin and destination locations. The system has the freedom of dynamically assigning the routes. This, of course, leads to a different utilization of the road network when the algorithm is enabled, compared to a situation with no rerouting; and also with variations depending on the selected inter-arrival interval and vehicle volume. Some road segments may be used in one configuration while they are unused in a different one, due to the dynamic route assignment. In the following analysis, we evaluate to what degree the presented methodology is representative of the network load, in place of the combination of total vehicle volume and IAI, which is more circumstantial and difficult to relate to each other.

The first hypothesis under investigation is that the network load is assumed to decrease after enabling the intelligent routing mechanism. While this seems obvious, it is also relevant to what degree the load decreases and its correlation with the figure of merit in Backfrieder, Ostermayer & Mecklenbräuker (2017), which evaluates (among other things) the performance in travel time. Subsequently, the second hypothesis claims that the network load needs to drop

in a similar manner to the observed average travel time of each vehicle, since the only possibility for improvements for the PCMA* routing approach is to distribute vehicles' routes differently compared to the original routes. We admit that this correlation is a strong simplification, but it still turns out to be fundamentally correct. Figure 6 indicates the correlation between traffic load (as a single value for a whole and highly dynamic traffic simulation) and average vehicle travel time.

We therefore also use the traffic simulator TraffSim, configured as defined in Backfrieder, Ostermayer & Mecklenbräuker (2017) in order to gain meaningful and comparable perceptions. Two different road networks (scenarios) are under consideration. One is an artificially designed network (Figure 3 left) and the other one is a real-world scenario from the city of Linz (Figure 3 right). All vehicles start within the green ellipse and travel towards the blue destination area, which is intended to somehow represent commuters travelling to their workplace within the city. Furthermore, in the artificial scenario, the low-level roads are rarely used, whereas, in the Linz scenario, the western part of the map is rarely used or even unused. These facts elicit the need for a dynamic consideration of the road infrastructure, as explained previously in the related work section.

Figure 3. Investigated road network scenarios (left: artificial street network, right: City of Linz, Austria)

Evaluation of achieved average travel time vs traffic load

Initially, we compare the achieved average travel time with the calculated traffic load of the very same simulation run. As explained earlier, the average vehicle travel time depends on both the total number of simulated vehicles and the IAI. The evaluation is done for two configurations, which represent a non-optimized and an optimized setup. We use the SMA method with averaging length 30 seconds. Note that we only show results for these parameters, since the behaviour is very similar when setting the parameters to different values.

Setup with disabled traffic optimization

First, the traffic load is calculated in a situation where traffic optimization is disabled, i.e. no route advice is provided for any vehicle and all of them choose the shortest route, independent of any emerging traffic jams. Figures 4a and 5a show the achieved average travel time vs IAI and overall vehicle amount, while Figures 4c and 5c depict the resulting load as defined in equation (7). What becomes obvious here is that the figure pairs 4a/4c and 5a/5c reveal a very similar shape. Both have their peak in the back upper corner, where arrival intervals are short and the number of vehicles is high. However, what also can be extracted from these figure pairs is the fact that there exist different combinations of IAI and vehicle volume that lead to the same load and very similar performance in terms of average travel time. This is one big benefit of quantifying the load in a general manner instead of the dependence on vehicle volume and arrival interval.

(a) average (∅) vehicle travel time

(b) average (∅) vehicle travel time

(c) network load without traffic optimisation

(d) network load with traffic optimisation

Figure 4. Average vehicle travel time and load without (a, c) and with (b, d) intelligent traffic optimization (artificial scenario)

Despite the general similarity of the surfaces, in the Linz scenario (Figure 5) there is still a slight difference visible when comparing the two shapes. Figure 5a has a more linear gradient

when looking backwards (i.e. lower IAI and higher vehicle amount), while Figure 5c is more belly-shaped with medium x- and y-values. This can be explained by long traffic jams that occur due to the heavy overload of the road network, which disappear very slowly and dramatically increase the travel time of the vehicles. In this scenario a large share of the entire road network is congested. The travel time therefore comparatively is much higher, while the load still does not increase any more if all used roads are jammed. In the artificial scenario this behaviour does not occur, since the chosen combinations of the parameters "number of vehicles" and "IAI" fit better to the capacity of the road network, meaning that a large part of the network is not congested. This leads to the fact that the shapes of the surfaces in Figures 4a and 4c are very similar.

(a) average (∅) vehicle travel time

(b) average (∅) vehicle travel time

(c) network load without traffic optimisation

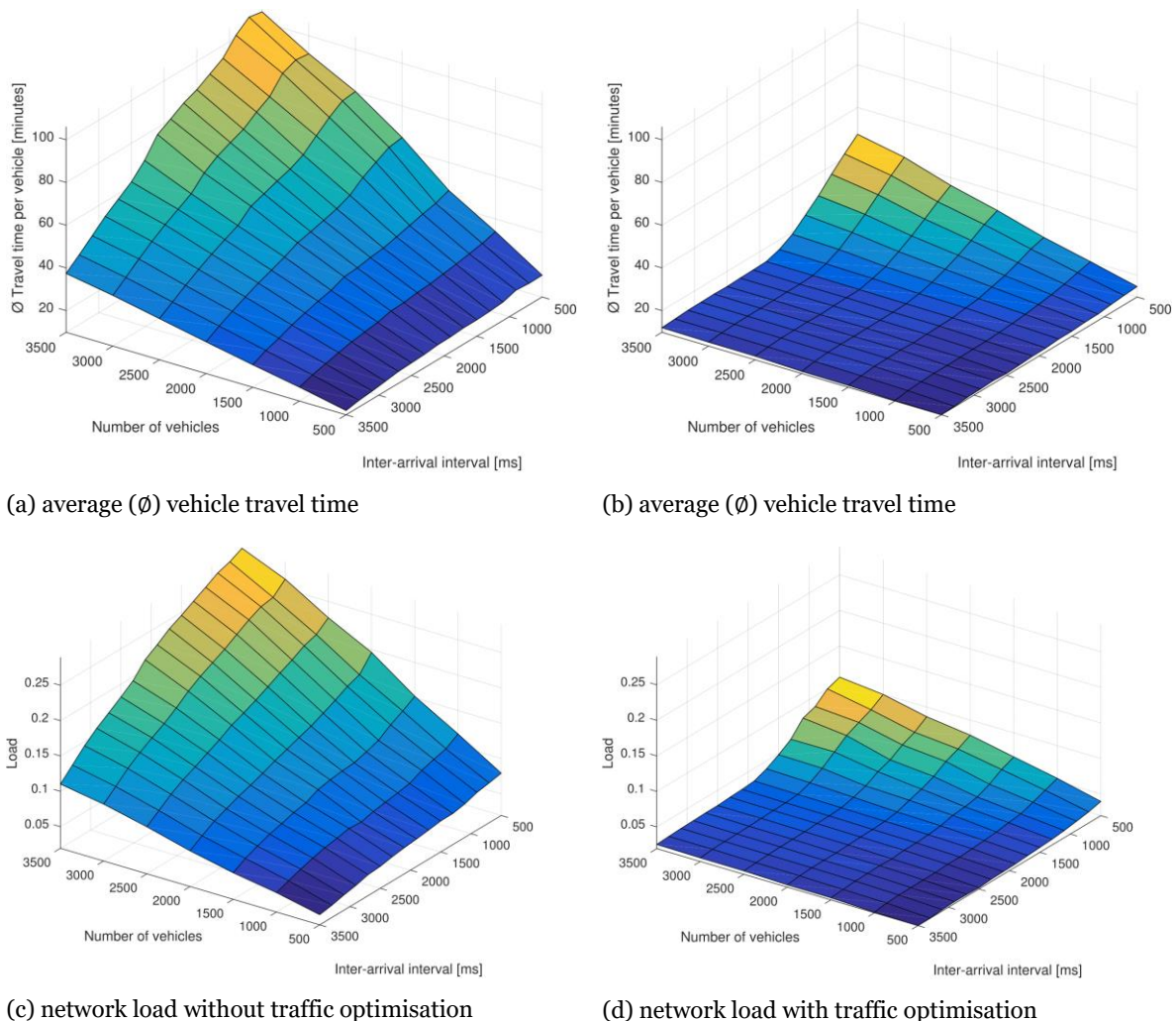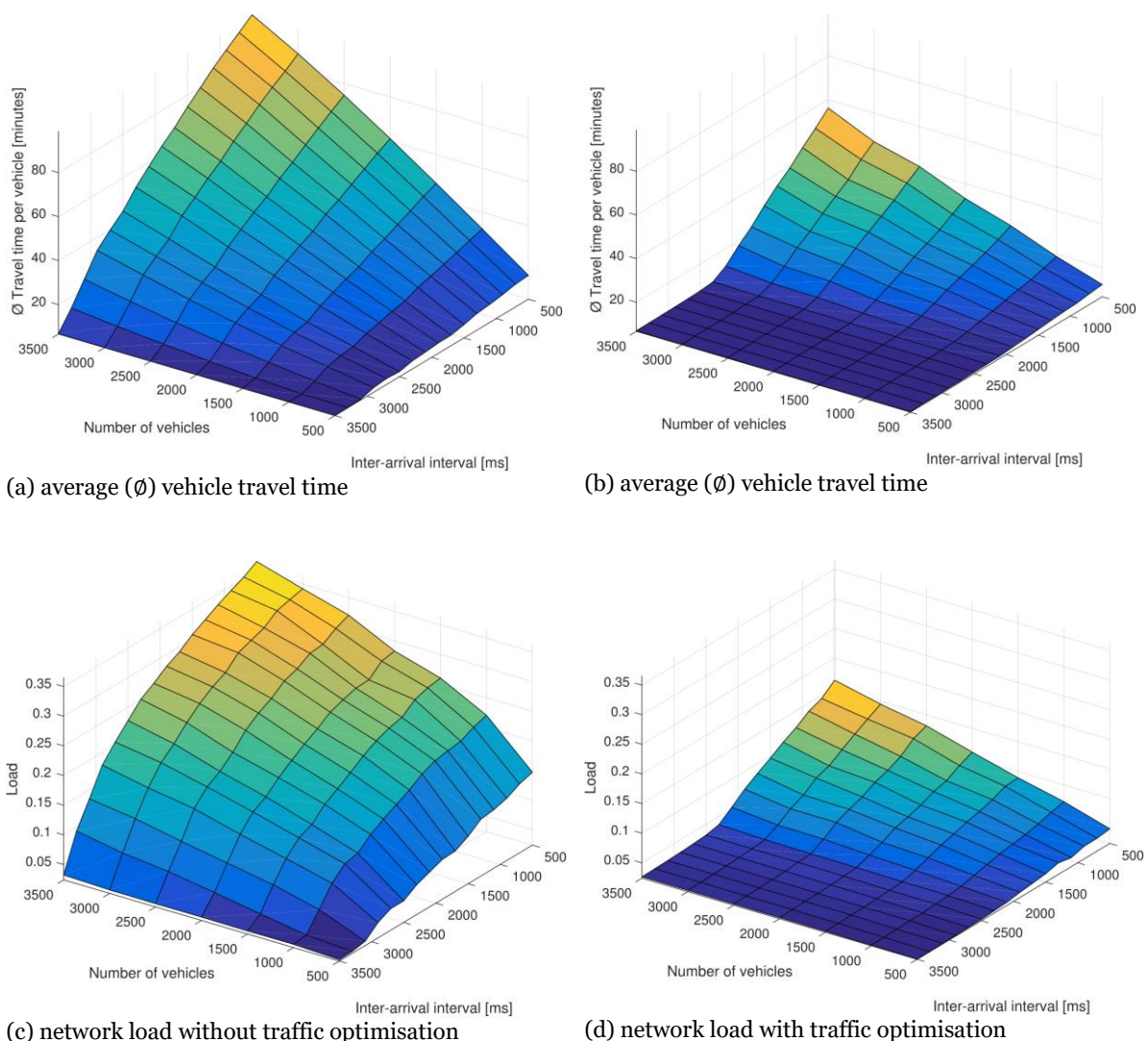(d) network load with traffic optimisation

Figure 5. Average vehicle travel time and load without (a, c) and with (b, d) intelligent traffic optimization (Linz scenario)

Setup with enabled traffic optimization

In the second setup, the best performing setup of PCMA* is chosen (according to Backfrieder, Ostermayer & Mecklenbräuker (2017)) in order to evaluate how the calculated traffic load

correlates with the average vehicle travel time if dynamic rerouting is enabled. Figures 4 (artificial scenario) and 5 (Linz scenario) show the achieved average vehicle travel time (Figures 4b and 5b, respectively) in contrast to the calculated load (Figures 4d and 5d, respectively). In the artificial scenario the surfaces are still very close to each other, whereas in the Linz scenario they are much closer to each other than before, which reveals the fact that the optimized average vehicle travel time is directly proportional to the configured traffic load for a wide range of parameter combinations of IAI and number of vehicles. The next subsection clarifies the relation between the two surfaces in more detail.

Comparison of normalized traffic load and travel time

In order to point out the correlation between the calculated load and the achieved average vehicle travel time of the optimization, we calculate the fraction of the average vehicle travel time $TT(\Delta t)$ and the respective calculated load $L(N, \Delta t)$ as the proportionality factor $k_{\text{TT,L}}$:

$$k_{\text{TT,L}} = \frac{TT(\Delta t)}{L(N,\Delta t)}. \tag{8}$$

This value is normalized to its mean value and results in the normalized proportionality factor $k_{\text{TT,L,norm}}$. Figure 6 shows $k_{\text{TT,L,norm}}$ for different parameter pairs IAI and number of vehicles and Δt as the time for a whole simulation run of 3000 vehicles.

Basically, a surface parallel to the IAI/number-of-vehicles plane means a constant proportionality factor $k_{\text{TT,L,norm}}$ for the entire parameter pair range. This means that the mapping from load to average vehicle travel time can be done for the entire parameter range with the knowledge of $k_{\text{TT,L}}$ for a single parameter pair. Usually there is a dependency on the parameters IAI and number of vehicles (#v) leading to $k_{\text{TT,L}}$(IAI, #v) and $k_{\text{TT,L,norm}}$(IAI, #v), respectively.

One can see that enabling the traffic optimization algorithm (Figure 6 right) has an impact on the proportionality factor. This is because the optimization algorithm helps to distribute the traffic over a higher number of road segments, leading to fewer traffic jams. A quite high dependency of $k_{\text{TT,L,norm}}$ on IAI and #v can be observed in Figure 6c. The reason for that was already explained in the two previous sections. Introducing the traffic optimization algorithm reduces the dependency of $k_{\text{TT,L,norm}}$ on IAI and #v.

What we can see is that there is a dependency on both road network and application of traffic optimization algorithms. Note that the proportionality factor $k_{\text{TT,L}}$ is almost constant in case of not too heavy traffic jams, whereas, in the other case, the proportionality between load and average vehicle travel time gets more dependent on the selected parameter pair and a single proportionality factor for the entire parameter pair range loses validity. Comparing Figures 6c and 6d shows that the application of the traffic optimization algorithm flattens the surface in

the range of heavy traffic (small IAI and high number of vehicles) due to its ability to distribute the traffic over more street segments, avoiding heavy traffic jams. But there still remains an increase of $k_{\text{TT,L,norm}}$ in the case of low traffic (high IAI and low number of vehicles). We expect that low traffic density leads to higher traffic flow, causing higher average speeds of the vehicles. In that case, the headway of the vehicles increases for safe driving. Since in our metric this headway is constant, the load is underestimated, leading to an increase of the proportionality factor. A consideration of this issue is a task for future work.
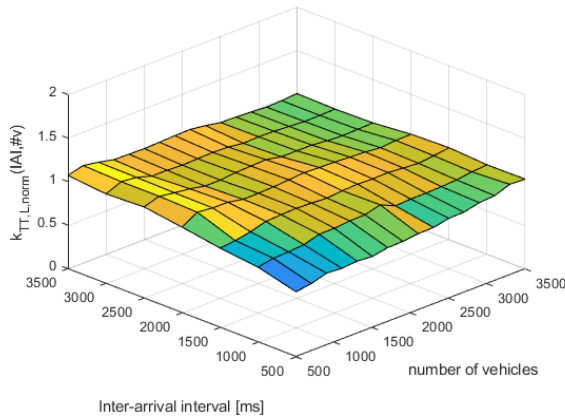
(a) artificial street network, optimization disabled
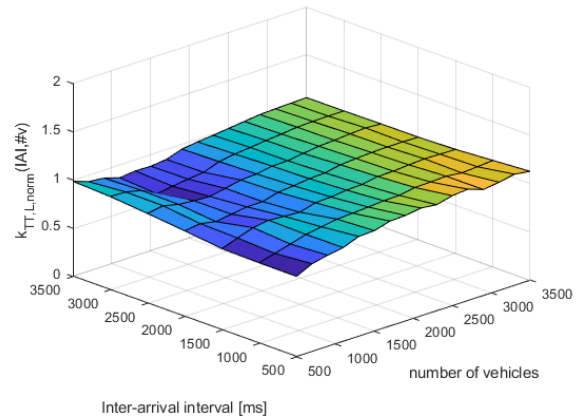
(b) artificial street network, optimization enabled
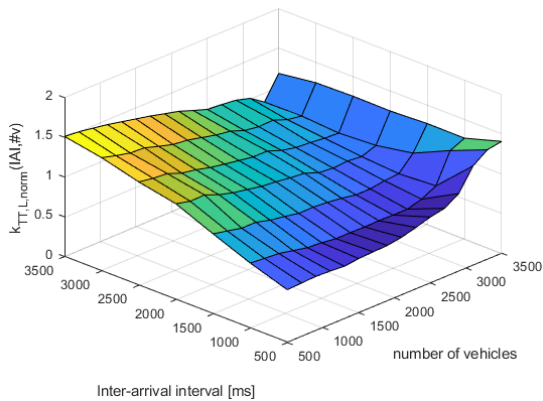
(c) Linz scenario, optimization disabled
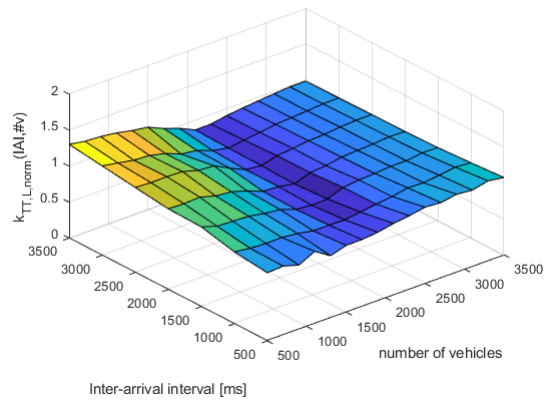
(d) Linz scenario, optimization enabled

Figure 6. Normalized proportionality factor $k_{\text{TT,L,norm}}(N,\Delta t)$ between load and average vehicle travel time without (a, c) and with (b, d) intelligent traffic optimization in the artificial road network scenario (top) and the Linz scenario (bottom)

Summary and Conclusions

We have proposed a novel methodology to quantify the traffic load over time, with special focus on scenarios where the usage of road space is dynamic and varies over time. Our

approach is able to determine the traffic load of an extract of the road network during a specific time period, but also to calculate a single load quantification number. For the latter, it is necessary to focus on the parts of the road that are actually used, instead of always taking into account a predefined part of the road network as done by related research in this field. In order to achieve this, we introduce two possibilities to eliminate distortions of the traffic load.

The proposed methodology is implemented as a cloud service, which requires location information in short time intervals to work precisely. The current positions of the vehicles are transmitted to the cloud at regular time intervals by V2I communication. However, this implies that all vehicles are somehow able to report the required data via wireless communication; otherwise the calculated numbers will underestimate the actual conditions on the road.

We demonstrate the applicability of the presented methodology by simulations, which compare the performance of PCMA*, a predictive rerouting and traffic optimization approach from the literature, to the calculated traffic load. As the only possibility for the dynamic re-router is to make use of additional, previously unused road space, the traffic load is assumed to decrease and correlates well with the achieved optimization performance (average vehicle travel time). The results of this comparison show very good proportionality between the calculated traffic load quantification and the achieved optimized average travel time in the artificial road network case and an acceptable proportionality for the Linz scenario.

In the end, this paper introduces a novel methodology to quantify the traffic load statistic for a defined road network extract and over a defined time period by a single number between 0 and 1, which is also a (very) good indicator of the achievable average vehicle travel time. It is much easier to apply than comparable approaches, and still adjustable to fulfil an individual application's needs.

Acknowledgements

References

Abadi, A., Rajabioun, T., Ioannou, P. A. (2015). Traffic Flow Prediction for Road Transportation Networks with Limited Traffic Data. *IEEE Transactions on Intelligent Transportation Systems, 16(2)*, 653–662.

Abbas, T., Sjöberg, K., Karedal, J., Tufvesson, F. (2015). A Measurement Based Shadow Fading Model for Vehicle-to-Vehicle Network Simulations. *International Journal of Antennas and Propagation, 2015*, 1-12. DOI: 10.1155/2015/190607

An, X., Hou, Q., Li, N., Zhai, S. (2013). Assessment of Human Exposure Level to PM10 in China. *Atmospheric Environment, 70*, 376–386.

Athol, P. (1965). Interdependence of Certain Operational Characteristics within a Moving Traffic Stream. *Highway Research Record, 72*, 58–87.

Aw, A., Rascle, M. (2000). Resurrection of "Second Order" Models of Traffic Flow. *SIAM Journal on Applied Mathematics, 60*(3), 916–938.

Backfrieder, C., Ostermayer, G., Mecklenbräuker, C. F. (2014). TraffSim – A Traffic Simulator for Investigations of Congestion Minimization through Dynamic Vehicle Rerouting. *International Journal of Simulation – Systems, Science and Technology, 15*(4), 38–47.

Backfrieder, C., Ostermayer, G., Mecklenbräuker, C. F. (2017). Increased Traffic Flow through Node-based Bottleneck Prediction and V2X Communication. *IEEE Transactions on Intelligent Transportation Systems, 18*(2), 349–363.

Backfrieder, C., Lindorfer, M., Mecklenbräuker, C. F., Ostermayer, G. (2018). Robustness of Intelligent Vehicular Rerouting towards Non-ideal Communication Delay. *Proceedings of the Future of Information and Communication Conference*, 143-164, Singapore.

Behrisch, M., Bieker, L., Erdmann, J., Krajzewicz, D. (2011). SUMO - Simulation of Urban MObility - an Overview. *Proceedings of the 3rd International Conference on Advances in System Simulation*, 55–60, Barcelona, Spain.

Blazek, T., Backfrieder, C., Mecklenbräuker, C. F., Ostermayer, G. (2017). Improving Communication Reliability in Intelligent Transport Systems through Cooperative Driving. *Proceedings of the 10th IFIP Wireless and Mobile Networking Conference*, 1-6, Valencia, Spain.

Boskovich, S., Barth, M. (2013). Vehicular network Rerouting Autonomy with a V2V, I2V, and V2I Communication Matrix Classification. *Proceedings of the 16th International IEEE Conference on Intelligent Transportation Systems*, 172-177, The Hague, Netherlands.

Cassidy, M., Coifman, B. (1997). Relation among Average Speed, Flow, and Density and Analogous Relation between Density and Occupancy. *Transportation Research Record: Journal of the Transportation Research Board, 1591*, 1–6.

Cappiello, A., Chabini, I., Nam, E. K., Lue, A., Zeid, M. A. (2002). A Statistical Model of Vehicle Emissions and Fuel Consumption. *Proceedings of the 5th IEEE International Conference on Intelligent Transportation Systems*, 801–809, Singapore.

Chan, K. Y., Dillon, T., Chang, E. (2013). An Intelligent Particle Swarm Optimization for Short-term Traffic Flow Forecasting Using on-road Sensor Systems. *IEEE Transactions on Industrial Electronics, 60* (10), 4714–4725.

Chen, A., Subprasom, K., Ji, Z. (2006). A Simulation-based Multi-objective Genetic Algorithm (SMOGA) Procedure for BOT Network Design Problem. *Optimization and Engineering, 7*(3), 225–247.

Elbery, A., Rakha, H., Elnainay, M., Drira, W., Filali, F. (2015). Eco-Routing Using V2I Communication: System Evaluation. *Proceedings of the 18th IEEE International Conference on Intelligent Transportation Systems*, 71–76, Las Palmas, Spain.

Gazis, D. C., Edie, L. C. (1968). Traffic Flow Theory. *Proceedings of the IEEE, 56*(4), 458–471.

Greenshields, B. D. (1935). A Study of Traffic Capacity. *Highway Research Board, 14*, 448-477.

Hoogendoorn, S. P., Bovy, P. H. (2001). State-of-the-art of vehicular traffic flow modelling. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering, 215,* 283-303.

Horvitz, E. J., Apacible, J., Sarin, R, Liao, L. (2009). Prediction, Expectation, and Surprise: Methods, Designs, and Study of a Deployed Traffic Forecasting Service. *arXiv:1207.1352 [physics], 2012*. [Online]. Available: http://arxiv.org/abs/1207.1352

Jing, B., Wu, L., Mao, H., Gong, S., He, J., Zou, C., Song, G., Li, X., Wu, Z. (2016). Development of a Vehicle Emission Inventory with High Temporal Spatial Resolution based on NRT Traffic Data and its Impact on Air Pollution in Beijing Part 1: Development and Evaluation of Vehicle Emission Inventory. *Atmospheric Chemistry and Physics, 16*(5), 3161–3170.

Kumar, R., Chhabra, S. (2016). Efficient Routing in Vehicular Ad-hoc Networks Using Firefly Optimization. *Proceedings of the 2016 International Conference on Inventive Computation Technologies*, 1–6, Coimbatore, India.

Lighthill, M. J., Whitham, G. B. (1955). On Kinematic Waves. I. Flood Movement in Long Rivers. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 281–316.

Lindorfer, M., Mecklenbräuker, C. F., Ostermayer, G. (2017). Modeling the Imperfect Driver: Incorporating Human Factors in a Microscopic Traffic Model. *IEEE Transactions on Intelligent Transportation Systems, 19*(9), 2856–2870.

Mandziuk, J., Swiechovski, M. (2016). Simulation-based approach to Vehicle Routing Problem with Traffic Jams. *Proceedings of IEEE Symposium Series on Computational Intelligence*, 1-8, Athens, Greece.

Nafi, N., Khan, R., Khan, J., Gregory M. (2014). A Predictive Road Traffic Management System based on Vehicular Ad-hoc Network. *Proceedings of the International Telecommunication Networks and Applications Conference*, 135-140, Melbourne, Australia.

Nebbou, T., Fouchal, H., Lehsaini, M. (2017). Metrics for Vehicle Density in Urban Environment. *Proceedings of the 2017 IEEE Global Communications Conference*, 1–6, Singapore.

Ni, D. (2015). *Traffic Flow Theory: Characteristics, Experimental Methods, and Numerical Techniques*. Butterworth-Heinemann.

Pan, J., Khan, M., Sandu Popa, I., Zeitouni, K., Borcea, C. (2012). Proactive Vehicle Re-routing Strategies for Congestion Avoidance. *Proceedings of the 8th IEEE International Conference on Distributed Computing in Sensor Systems*, 265–272, Hangzhou, China.

Stevanovic, A., Stevanovic, J., Zhang, K., Batterman, S. (2009). Optimizing Traffic Control to Reduce Fuel Consumption and Vehicular Emissions: Integrated Approach with VISSIM, CMEM, and VISGAOST. *Transportation Research Record: Journal of the Transportation Research Board, 2128*, 105–113.

Szeto, W. Y., Jiang, Y., Wang, D. Z. W., Sumalee, A. (2015). A Sustainable Road Network Design Problem with Land Use Transportation Interaction Over Time. *Networks and Spatial Economics, 15(3)*, 791–822.

Treiber, M., Kesting, A. (2013). *Traffic Flow Dynamics*. Berlin, Heidelberg. Springer.

Zhang, B., Ma, M., Liu, C., Shu, Y. (2017). Delay Guaranteed MDP Scheduling Scheme for HCCA based on 802.11p Protocol in V2R Environments. *International Journal of Communication Systems, 30(15)*, 1-12.

Zhou, X., Tanvir, S., Lei, H., Taylor, J., Liu, B., Rouphail, N. M., Frey, H. C. (2015). Integrating a Simplified Emission Estimation Model and Mesoscopic Dynamic Traffic Simulator to Efficiently Evaluate Emission Impacts of Traffic Management Strategies. *Transportation Research Part D: Transport and Environment, 37*, 123–136.

New Zealand Consumer Interest Growing for 5G Mobile

Nigel Pugh

Managing Director, Venture Insights

Abstract: This paper summarizes results from a survey of the New Zealand mobile market by Venture Insights in October 2018. The survey had over 1,000 respondents, all of whom were responsible for making their mobile purchasing decisions, with a representative spread across New Zealand, all adult age groups, and customers from the three major mobile service providers. The survey identified a strong consumer interest in and awareness of 5G, with 31% of respondents willing to consider a move to 5G within two years of launch. Mobile video will be a strong driver, with 69% of respondents having viewed video on their mobile and, again, a willingness to move to 5G within two years. Fixed wireless broadband service was also strongly supported, with 18% of those with a household fixed broadband service indicating a possible transition within two years to a 5G fixed wireless service. However, the Ultra-Fast Broadband fixed service is likely to be a strong competitor to fixed wireless broadband.

Keywords: 5G, Fixed wireless, Wireless broadband, New Zealand market

Introduction

In October 2018, Venture Insights conducted a market survey in order to better understand consumer interest in the evolving New Zealand mobile market and the early adoption of 5G services and platforms. This paper, after some background on the respondents, focusses on four key aspects of the survey:

- Consumer 5G awareness and competition for the early adopter market;

- Consumer usage of mobile video and streaming services;

- 5G use case for mobile and telco media bundling;

- 5G use case for fixed wireless broadband.

New Zealand Consumer Survey Background

The Venture Insights New Zealand mobile consumer survey was conducted in October 2018 and included 1,007 completed responses spread across New Zealand – see Figure 1.
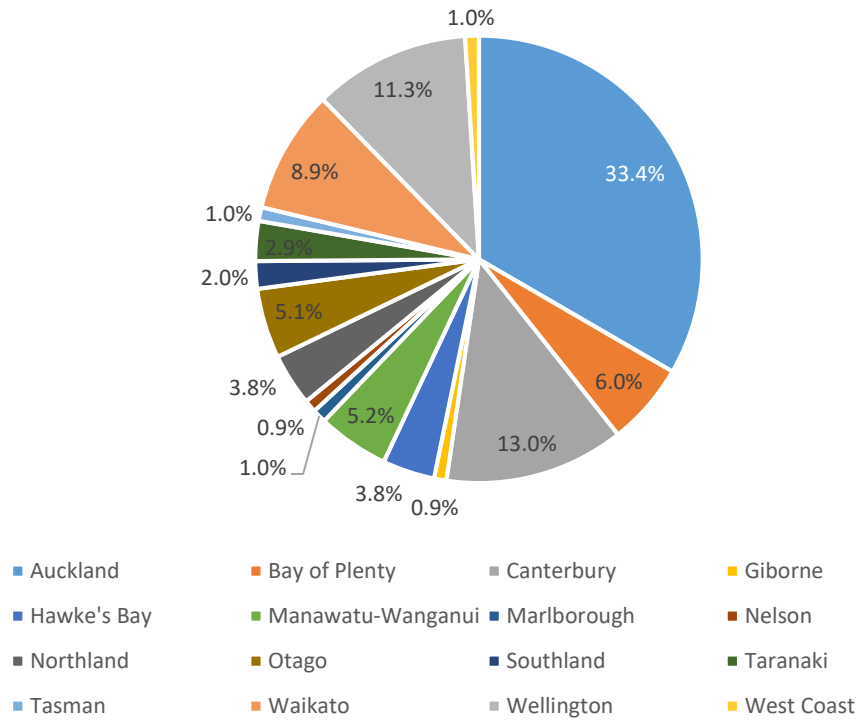
Figure 1. New Zealand Survey - Geographic Spread
(Source: Venture Insights NZ Consumer Mobile Survey, October 2018, n=1,007)
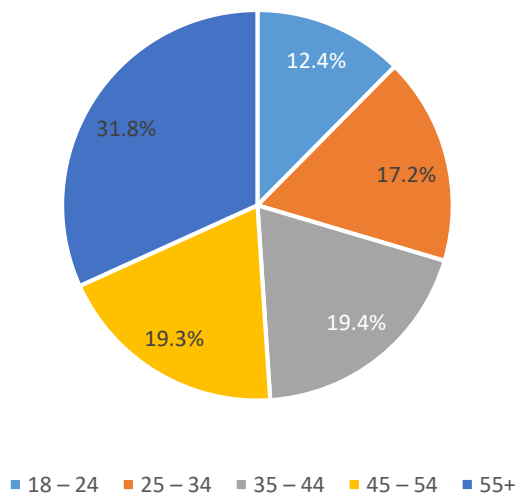
Figure 2. New Zealand Survey – Age Bands
(Source: Venture Insights NZ Consumer Mobile Survey, October 2018, n=1,007)

The survey included a gating question to ensure that all respondents were responsible for purchasing the service. In addition to regional coverage, the survey also represented a fair spread of age and income bands – see Figure 2.

When asked which mobile operator was their primary provider (Figure 3), Vodafone recorded 35.4%, Spark 31.2% and 2Degrees 23.2%. Given the competition for market share in New Zealand, we believe the 5G launch phase will be highly competitive as the mobile operators seek to attract the early 5G adopter market and push forward with their 5G network rollouts.
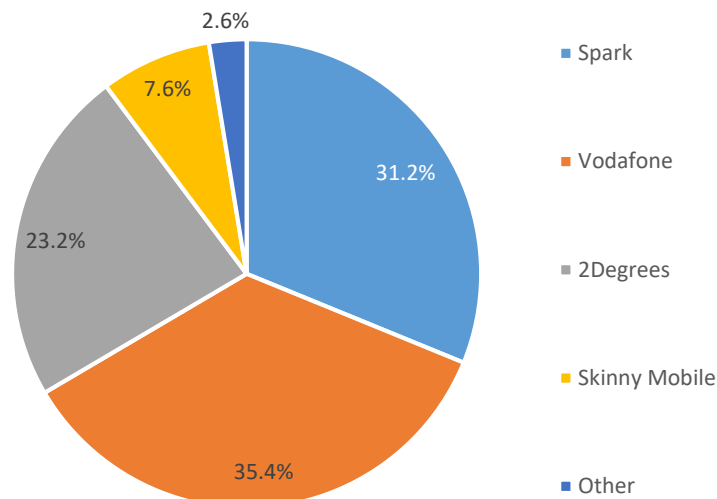
Figure 3. New Zealand Survey – Primary Mobile Account of Respondents
(Source: Venture Insights NZ Consumer Mobile Survey, October 2018, n=1,007)

Consumer Interest in 5G

Similar to Venture Insight's Australian mobile survey (Pugh, 2019), the New Zealand survey tested consumer interest in 5G and respondents' willingness to take up a 5G service.

The survey showed that 30% of respondents were aware of plans to deploy 5G networks, with the majority of these being existing Vodafone and Spark customers. This suggests that recent 5G trials and press coverage – see, for example, Spark's 5G briefing paper (Spark, 2018a) – are getting the attention of consumers.

However, unlike the Australian telcos – who have flagged 2019 launch timeframes, the New Zealand market is still discussing launch timeframes of 2020. At the time of the survey (October 2018), there was uncertainty about the New Zealand spectrum auction planning process, with the NZ Government then yet to respond to 5G spectrum planning submissions, which had closed on 30 April 2018. Subsequently, in February 2019, the Minister has announced plans (Faafoi, 2019) to auction spectrum in the 3.5 GHz band "early in 2020", with full rights to the spectrum "from November 2022". As such, the mobile operators are still

waiting for greater clarity on when additional 5G spectrum will be available and in which bands.

In terms of willingness to upgrade to 5G (Figure 4), we found that 6% of all respondents would consider migrating at launch, with a further 25% willing to migrate within 24 months. Interestingly, the majority of these early adopters were existing Vodafone customers.

Figure 4. New Zealand Survey – Consumer Willingness to Move to 5G
(Source: Venture Insights NZ Consumer Mobile Survey, October 2018, n=1,007)

The survey also showed that 11% of respondents would be willing to pay more for a 5G service, indicating demand for superior networks could be matched by innovative pricing for some segments of the market. Similar results were found for willingness to pay more for a 5G handset.

Given the strong growth in smartphone usage across a wide range of applications, the survey results confirm our view that consumers care about technology upgrades and are willing to shift (and in some cases) pay more to gain an improved mobile experience.

Mobile Video Usage

Mobile video usage was strong, with 69% of respondents indicating they had viewed video on their mobile and 43% reporting that they watched video streaming services regularly (i.e. on either a daily or weekly basis). In addition, 73% of respondents who watched video on their mobile indicated that they watched more video content now than a year ago.

Time-of-day viewing was reasonably spread, with the highest viewing times being late afternoon and evening between 5 pm and 8 pm, reflecting a mix of travelling from work and evening viewing. Indeed, the majority (78%) of respondents mainly watched video streaming content at home (71%) or when travelling (7%). The relative high percentage of 'at home'

viewing was also supported by 76% of respondents reporting watching mobile video content only when connected to a Wi-Fi connection.

On a daily basis ▪ On a weekly basis ▪ On a monthly basis ▪ Only sometimes ▪ Never
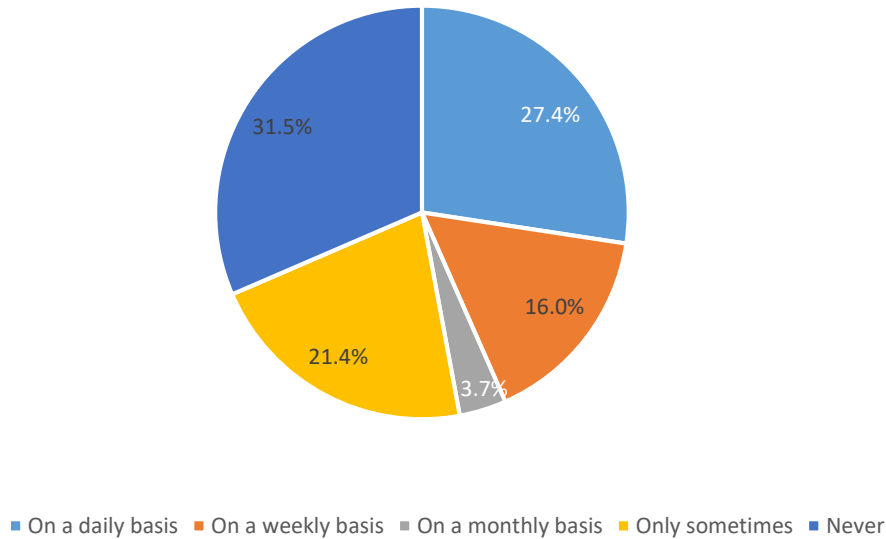
Figure 5. Mobile Video Viewing
(Source: Venture Insights NZ Consumer Mobile Survey, October 2018, n=1,007)

Of the 69% of respondents who watched video content on their mobile, the top three most watched streaming services were YouTube, Netflix and TVNZ OnDemand

Home ▪ Work ▪ Travelling ▪ External venue ▪ At a friend's house ▪ Other
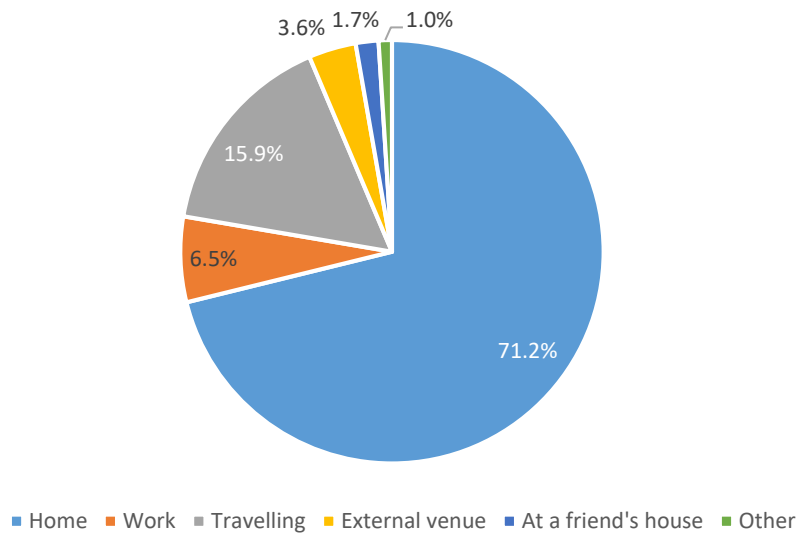
Figure 6. Main Locations for Viewing Video Content on Mobile
(Source: Venture Insights NZ Consumer Mobile Survey, October 2018, n=1,007)

When asked to rate the factors that limited video viewing, the top three limiting factors were mobile data limits (24%), the screen being too small (30%) and battery life (17%). We note that all three of these factors are trending upwards, with consistently larger mobile data limits every year and mobile phone screen and battery sizes typically increasing with each new generation.
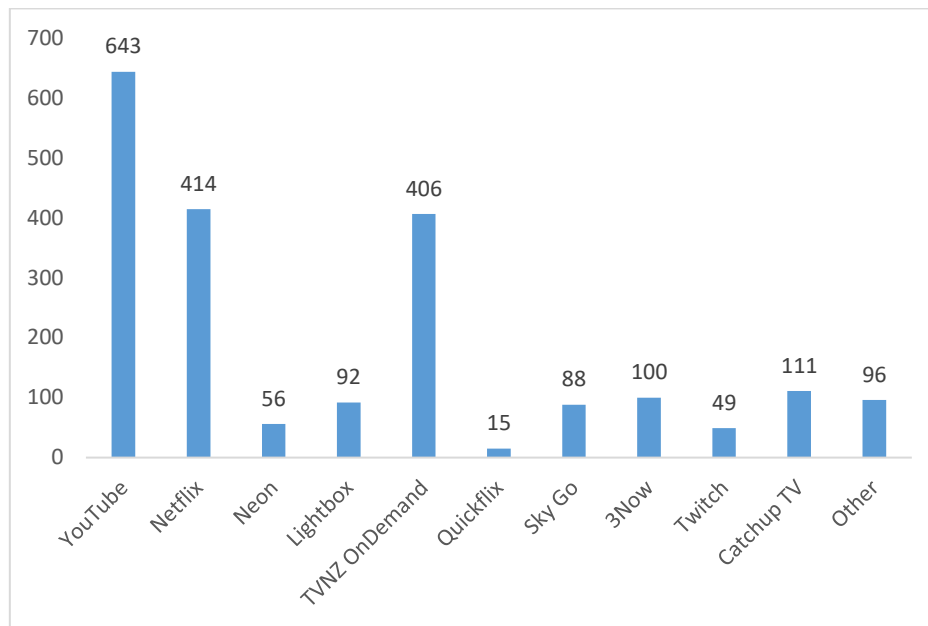
Figure 7. Video Streaming Services Watched the Most: top 3 video services per respondent (Source: Venture Insights NZ Consumer Mobile Survey, October 2018, n=1,007)

5G Video Use Case and Telco Media Bundling

Of the respondents who indicated they were willing to move to 5G within 2 years, 73% watched video content on their mobile either daily, weekly or monthly. This suggests that video viewing will be a strong use case for consumer 5G – especially when marketed with the benefits of 5G technology such as faster download speeds and lower latency.

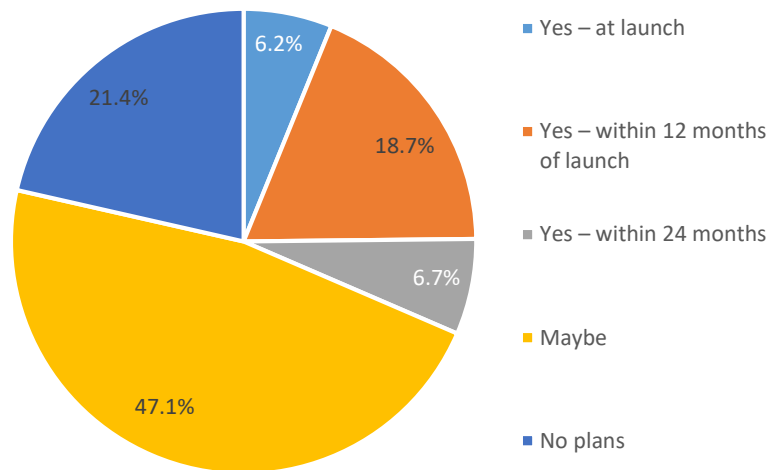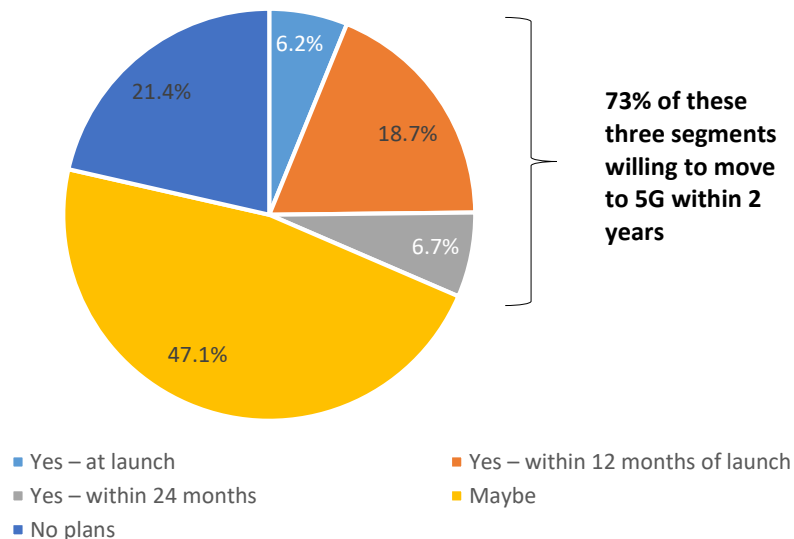**Figure 8. Consumer Willingness to Move to 5G (Source: Venture Insights NZ Consumer Mobile Survey, October 2018, n=1,007)**

The NZ consumer mobile survey also asked respondents to rate six factors when purchasing their next mobile phone service. Unsurprisingly price rated at number one, with 90% of respondents rating this as important/very important, closely followed by network

performance/speed rated at 81%. However, the importance of bundled media with mobile plans was rated at just 24%. Whilst this indicates there is still a healthy segment of the NZ consumer market interested in bundled media, operators should clearly emphasise network performance in their 5G video use case to maximise uptake.

Table 1. Importance of Mobile Factors at Respondent's Next Mobile Service Purchase

Mobile Factor	Consumer Ranking
Price	90%
Network Performance/Speed	81%
Data Allowance	68%
Regional Network Coverage	79%
Handset included	30%
Bundled media/video services	24%

Source: Venture Insights NZ Consumer Mobile Survey, October 2018, n=1,007, Ranking = Important/Very Important

The media-bundling strategies of the major New Zealand telcos are summarized in Table 2.

Table 2. New Zealand Major Telco Media Strategies

Company	Strategy	Monetisation
Spark nz	Mixed Strategy: • Partnership with Netflix offering a free subscription for a year • A deal with Spotify	Bundling Strategy: • Netflix, Spotify and Lightbox are only available to customers who purchase mobile or broadband plans
vodafone	Mixed Strategy: • Vodafone TV is a new standalone set top box that integrates content such as Netflix, Sky and YouTube • Partnership with Sky NZ to show Sky basic on Vodafone TV • Vodafone also offers deals with partnered cinemas across the country for their customers	Bundling Strategy: • Vodafone TV is offered in a bundle with broadband plans that use Fibre or FibreX
2	Partnership Strategy: • Partnership with Google Play to allow access through 2Degrees account	Bundling Strategy: • Google Play access is included in mobile plans

Source: Venture Insights

5G Fixed Wireless Broadband Use Case

Another strong use case for 5G is fixed wireless broadband. In Australia we have seen announcements from both Optus and Telstra with plans to launch a fixed wireless 5G service (in metro areas). Optus reported three live sites at 31 January 2019, with plans for 1,200 5G sites by March 2020 (Optus, 2019). Telstra has reported that it has rolled out "200 5G-capable

base stations across Australia by the end of 2018" (Turner, 2019). From late May 2019, Telstra has been offering an HTC 5G Hub for home broadband (Duckett, 2019).

When we asked the New Zealand survey respondents with fixed broadband household services (such as fibre or copper) if they would consider switching to a wireless broadband service, 18% indicated they would consider switching within the next two years. Interestingly, this is a lower result when compared with our Australian broadband survey (Pugh, 2019) – which indicated 30% of existing fixed broadband respondents would consider migrating within the next 2 years. We believe the larger result in Australia was attributable to consumer dissatisfaction with the Australian NBN (where some respondents indicated they had concerns with the NBN).

■ Yes – within 1 year ■ Yes – within 2 years ■ No ■ Unsure
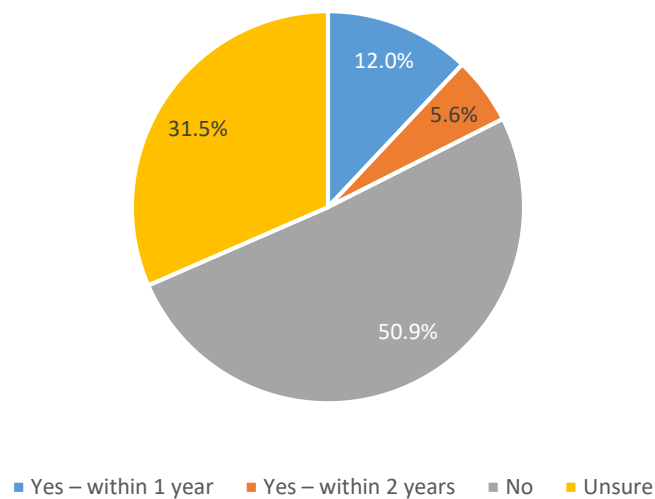
Figure 9. Fixed Broadband Respondents Considering Shift to Wireless Broadband (Source: Venture Insights NZ Consumer Mobile Survey, October 2018, n=649)

Over the last two years, the New Zealand mobile market has seen the launch of several wireless broadband services, first by Spark-owned Skinny Mobile, which was followed by Spark's own wireless broadband service in April 2016. Vodafone launched its wireless broadband service in December 2016.

Spark has been marketing its wireless broadband as an upgrade for 1) areas that are serviced by old copper lines, 2) areas that may not receive the UFB (Ultra-Fast Broadband fixed-line service) and 3) users with low to moderate data usage. It has been fairly aggressive in its approach and has signed up 116,000 customers by August 2018 – which Spark claims has resulted in operating cost savings and a 6.5% uplift in broadband gross margin (Spark, 2018b).

Clearly, wireless broadband represents a good option for NZ mobile operators to bypass the UFB (for some customer segments) and protect their margins. With the upcoming launch of 5G, there is clear scope for the 5G wireless broadband plans to improve and place further

pressure on Chorus. However, the UFB is a strong competitor, with unlimited 100 Mbps plans and 1 Gbps plans already in service, trial 10 Gbps services announced (Chorus, 2019a), and over 529,000 fibre connections in place at 31 December 2018 (Chorus, 2019b). As such, although the consumer survey shows there is demand for wireless broadband services, actual take-up may be lower than the survey results, given the strong positioning of the UFB services and the relative delay in launching 5G services.

Conclusion

Venture Insights' New Zealand mobile survey has indicated strong consumer interest in 5G, with 30% of respondents aware of operator plans to deploy 5G networks and 31% willing to migrate to 5G within two years. These results are lower than those from the corresponding Australian survey, which we believe reflects the NZ operators' reluctance to push 5G rollouts while uncertainties remain about when additional 5G spectrum will be available and in which bands.

The survey indicated that 11% of NZ respondents would be willing to pay more for a 5G service, which suggests that the demand for superior networks could be matched by innovative pricing for some segments. Video will be a strong consumer use case for 5G: 43% of respondents regularly view video on their mobile, with close to 50% of those users wanting to move to 5G within two years.

In the NZ survey, 18% of fixed broadband respondents indicated that they would consider switching to a fixed wireless service within two years. This could build on Spark's existing push for fixed wireless services, especially when 5G services are launched. The result on fixed-wireless migration is lower than in the equivalent Australian survey, but the larger Australian result is probably attributable to consumer dissatisfaction with the Australian NBN. Migration to 5G fixed wireless services may also be less likely, due to the strong position of the UFB in New Zealand.

Acknowledgements

References

Chorus. (2019a). Supercharging New Zealand's broadband with a 10Gbps trial. *Broadband Basics*, 4 February. Available at https://www.chorus.co.nz/blog/supercharging-new-zealands-broadband-10gbps-trial

Chorus. (2019b). Half year result. *Media Release*, 25 February. Available at https://company.chorus.co.nz/half-year-result

Duckett, C. (2019). Telstra releases HTC 5G Hub online. *ZDNet*, 22 May. Available at https://www.zdnet.com/article/telstra-releases-htc-5g-hub-online/

Faafoi, K. (2019). 5G on track for 2020: 3.5 GHz spectrum the first available. *Ministerial press release*, 28 February. Available at https://www.beehive.govt.nz/release/5g-track-2020-35-ghz-spectrum-first-available

Optus. (2019). Optus 5G Whiz: Optus announces plans for 1,200 5G sites in first wave, $70 per month unlimited 5G broadband plans, live 5G sites and the chance to be amongst the first Australians to experience 5G. *Optus Media Release*, 31 January. Available at https://www.optus.com.au/about/media-centre/media-releases/2019/01/Optus-5G-Whiz

Pugh, N. (2019). The Wireless Threat to Fixed Broadband Services. *Journal of Telecommunications and the Digital Economy, 7*(1), 7-19. doi:10.18080/jtde.v7n1.178

Spark. (2018a). 5G: The evolution towards a revolution. *Spark NZ Briefing Paper*, August. Available at https://www.sparknz.co.nz/content/dam/SparkNZ/pdf-documents/5G%20Briefing%20document.pdf

Spark. (2018b). Spark New Zealand FY18 results show strong performance in growth areas while transforming business for the future. *Spark NZ Investor Paper*, August. Available at https://www.sparknz.co.nz/news/Spark_FY18_results/

Turner, A. (2019). Telstra set to start offering 5G smartphones in the first half of 2019. *Sydney Morning Herald*, 10 January. Available at https://www.smh.com.au/technology/telstra-set-to-start-offering-5g-smartphones-in-the-first-half-of-2019-20190110-p50qm6.html

Measuring Digital Inequality in Australia: the Australian Digital Inclusion Index

Chris K Wilson
Swinburne University

Julian Thomas
RMIT University

Jo Barraket
Swinburne University

Abstract: In the past two decades digital inequality has come to be understood as a complex, evolving and critical issue in Australia, as it has elsewhere. This conceptual shift has generated demand for more complex measurement tools that can capture and combine multiple and graduated indicators of digital inequality. The Australian Digital Inclusion Index (ADII), developed in 2015 and now including annual data covering the period 2014 to 2018, is a composite index that addresses this demand. This paper describes the development of the ADII, its architecture and the dataset used to populate it. It also provides an overview of the findings of the 2018 edition of the index. The 2018 index reveals that, although on aggregate digital inclusion is improving in Australia, it continues to follow distinct geographic, social and socio-economic contours. In general, rural and regional Australians, older Australians and Australians with low levels of income, employment, and education are less digitally included than their compatriots. For some of these groups the inclusion gap is widening.

Keywords: digital inclusion, digital divide, Australian Digital Inclusion Index

Introduction

In recent years digital inclusion has emerged as a significant social challenge, and thus a critical area of investigation and policy development in Australia, as it has elsewhere. Digital technologies have become progressively more deeply embedded in work, education, government, healthcare services and other aspects of everyday experience. Indeed, digital communication is now widely seen as an essential service, comparable to basic utilities in its contribution to welfare.

Our recent work through the Australian Digital Inclusion Index (ADII) reveals that, although digital inclusion is improving in Australia, levels of digital access, affordability and abilities (the three fundamental elements of digital inclusion captured by the ADII) continue to follow

distinct social and socio-economic contours (Thomas *et al.*, 2018). Indeed, despite the overall lift in digital inclusion in Australia, the 'inclusion gap' has expanded for some groups. For instance, the gap in the level of digital inclusion between those in the lowest and highest income quintiles is greater in 2018 than it has been over any of the years since the ADII was first calculated in 2014.

In this paper we provide an overview of the changes to digital inclusion in Australia. Before we present this analysis, we briefly discuss the emergence of increasingly nuanced and complex conceptualisations of digital inequality and the consequent development of multi-dimensional indexes in a bid to quantify and track relative levels of inclusion across populations and over time. We outline the emergence of the Australian Digital Inclusion Index as one such tool, describe its architecture and the Roy Morgan Research Single Source dataset that is used to populate it.

Conceptualising and Measuring Digital Inclusion

Problematising the social impact of differential access and use of communication technologies has a history that extends beyond the emergence of the public internet (Sassi, 2005; Andreasson, 2015). But the issue gained significant traction as the World Wide Web began to achieve mass diffusion in a number of developed countries by the late 1990s (van Dijk, 2012, p. 2). The term 'digital divide', which entered circulation in the United States in the mid-to-late 1990s, before its broad international adoption, was the primary moniker under which digital inequality was described in these early years of the internet (Gunkel, 2003). Although use of the term varied, its framing in the public policy of the Clinton-Gore Administration ensured it came to primarily denote "a form of socioeconomic inequality demarcated by the level of *access* that one has to IT" (Gunkel, 2003, p. 503). Conceptualising the divide as one centred on material access to technology informed a substantial volume of research (see literature reviews by DiMaggio *et al.* (2004), Dewan & Riggins (2005), van Dijk (2006) and Mubarak (2015)). Particularly important were studies conducted by the U.S. Department of Commerce National Telecommunications and Information Administration (NTIA) between 1995 and 1999. These NTIA studies measured and benchmarked the divide in terms of personal access to telephone, computer and internet connections (NTIA, 1995; NTIA, 1998; NTIA, 1999).

Reducing socio-technological differences to material access to communication technologies and applying a binary logic that divided those *with* and *without* such access would increasingly be criticised for failing to capture the complexity of the personal digital experience and the need for more targeted policy interventions to address inequality (van Dijk, 2006, p. 222; Gunkel, 2003, pp. 505-509; Strover, 2003, p. 275). One response has been to classify the

digital divide into a hierarchical taxonomy of digital divide levels. Hargittai (2002) initiated this process. Arguing that to understand online inequality it was "increasingly important to look at not only who uses the Internet, but also to distinguish varying levels of online skills among individuals", Hargittai labelled access as the first-level digital divide and online capabilities the second-level divide (Hargittai, 2002). A third level was later added to the taxonomy. It seeks to capture how internet use translates into offline outcomes (Kwok-Kee *et al.*, 2011; Dewan & Riggins, 2005; Helsper & van Deursen, 2015), thus suggesting integration between online and offline experiences of advantage and disadvantage.

While the concept of the digital divide has retained its scholarly and political currency, particularly in its renovated multi-level form, there has also been moves to conceptualise inequality in new ways that more clearly emphasise its multiplicity, continuity and social constitution. DiMaggio & Hargittai (2001, p. 8) suggested using the term digital inequality as a means of moving "beyond the digital divide" and its binary logics. Hargittai (2003, p. 2) defined digital inequality as a "refined understanding of the 'digital divide' that emphasizes a spectrum of inequality across segments of the population depending on differences along several dimensions of technology access and use". More recently, and consistent with assets-based and people-centred logics emerging in other disciplinary domains, the term digital inclusion has emerged as a more positive and human, rather than technology-centred, term to conceptualise digital inequality and its impact in undermining the broader goal of social inclusion (Walton *et al.*, 2013, p. 9.4; Tsatsou, 2011; Helsper, 2008; Crandall & Fisher, 2009, p. 5).

As the conceptualisation of digital inequality has become more nuanced to reveal its multifaceted and graduated nature, the demand for more complex methods for measuring the nature and extent of such inequality arose (Selwyn, 2004). In 2006, Barzilai-Nahon bemoaned that, despite the conceptual shift in discourse to digital inequality, most analyses continued to be atomic and monotopical, focusing on a single factor rather than taking a holistic and comprehensive view through "integrative frameworks and measurements" (Barzilai-Nahon, 2006, p. 269). Over time, multi-factor digital inclusion analyses have become more prevalent, with increasing levels of sophistication around identifying the interaction between factors (cf. van Deursen *et al.*, 2017). Composite indexes (Nardo *et al.*, 2008), which are capable of systematically quantifying digital inclusion as a complex phenomenon based on a suite of specific indicators, have also been developed and refined.

Digital inclusion indexes first emerged at the global analytical level (Dewan & Riggins, 2005). These global level indexes focus on quantifying digital inclusion at the national level to enable international comparison (Barzilai-Nahon, 2006, p. 271). The International Telecommun-ications Union (ITU) has been a pivotal player in the development of such indexes, beginning

with the Digital Access Index in 2003 (ITU, 2009, Chapter 3; Bruno *et al.*, 2011). Its latest index, the ICT Development Index, combines data on communication service subscriptions, home computer and internet access, internet usage, and skills proxy indicators (mean years of schooling, gross secondary enrolment, and gross tertiary enrolment) to generate three sub-indexes: access, use and skills. Since 2017, The Economist Intelligence Unit (on behalf of Facebook) has collated an annual Inclusive Internet Index. The index combines personal, institutional and infrastructural indicators divided into four domains (availability, affordability, relevance and readiness) to generate a holistic view of a country's level of internet inclusion (EIU, 2019, p. 13).

Over time, there has been increasing interest in developing indexes that seek to compose a comprehensive picture of digital inclusion for individuals that would enable demographic and intra-national geographic analysis. One of the first such indexes was developed in South Korea, where the government had been quick to engage with the issue of digital inclusion, even passing a Digital Divide Act in 2001 (Park & Jae Kim, 2014, p. 75). South Korea's Digital Divide Index (DDI) was first compiled in 2004. It incorporates indicators across "three dimensions of digital divide – access, skills and utilisation" and measures relative digital inequality between a number of socio-economically disadvantaged groups and the general population at a single point and over time (Park & Jae Kim, 2014, p. 76). Initially, four groups were identified: "the disabled, the elderly, low-income persons and agriculture/fishery workers" (Park & Jae Kim, 2014, p. 76). From 2011, new groups were added, including "North Korean refugees and transnational marriage migrants" (Park & Jae Kim, 2014, p. 76). A more recent development is the Lloyds Bank UK Consumer Digital Index. It has been compiled annually since 2016. Reflecting an increasing use of data analytics to make best use of latent information, this index aggregates data from multiple surveys and bank transaction records to generate financial and digital capability scores (Lloyds Bank, 2018). This data has been further used as an input into the Digital Exclusion Heatmap created by Tech Partnership, a UK digital skills development alliance (The Tech Partnership, 2017). The Digital Exclusion Heatmap is a predictive geographic digital inclusion index which combines data on digital access, infrastructure and a range of social metrics.

The ADII: History, Dataset and Architecture

In 2015, Telstra, Australia's largest telecommunications company, initiated a project with Swinburne University to develop an Australian digital inclusion index. Substantial consultation with commercial, community and government stakeholders was undertaken, centring on the *Australian Digital Inclusion Index Discussion Paper* (SISR *et al.*, 2015). The consultation confirmed the desirability of a complex digital inclusion measure that integrated

indicators across three dimensions: access, affordability and digital capabilities. The consultation also highlighted the need for the index to be populated with data that would enable relatively refined geographic and socio-economic analysis and that had the potential to support benchmarking the progress of digital inclusion initiatives. The research team identified a range of government, community and market research datasets that contained indicators which might be integrated using statistical modelling techniques. These included those compiled by/for the Australian Bureau of Statistics, the Australian Communications and Media Authority, Roy Morgan Research and The Smith Family (ABS, 2011; ABS, 2014; ACMA, 2015; Roy Morgan Research Ltd, 2015; The Smith Family, 2013). During this scoping process, the team determined that the Roy Morgan Single Source survey included a sufficient range of indictors to populate an index, enabling the generation of a digital inclusion score for each survey respondent and thereby obviating the need for any form of multi-source statistical overlay. The Australian Digital Inclusion Index (ADII) that emerged from the consultation process is exclusively populated with Roy Morgan Single Source data.

The Single Source, the world's largest ongoing single source survey, is compiled by Roy Morgan Research (Roy Morgan Research Ltd, 2018). It is administered every weekend as a face-to-face interview combined with a supplementary leave-behind self-completion booklet to a representative sample of households spread over more than 500 sampling areas across Australia (Roy Morgan Research Ltd, 2017). Over the course of 12 months, the survey yields around 15,000 returns suitable for use in the ADII. The large sample size and geographic coverage often permits geographic and demographic reporting with low margins of error at a high level of confidence (95%+). As an ongoing dataset, it has been possible to generate annual index results for five years to date (2014-2018). The annual datasets represent a 12-month sample to March of that year (e.g. 2014 contains data collected between April 2013 and March 2014). Although the dataset is robust, as with any social research instrument it does have limitations. For instance, the survey is administered in English, which may restrict coverage of those with limited English and/or literacy. Around 800,000 people whose second language is English do not speak it well or at all, while 1.5 million English speakers have low levels of literacy, which may cause issues in relation to the leave-behind booklet (ABS, 2016; ABS, 2013).

Overall, data from more than 100 questions in the Single Source relating to internet products, services, activities, attitudes and income are used to generate an individual's digital inclusion index score (ADII score). The computation of the ADII score is based on the equally weighted aggregation of three sub-indexes conforming to the three explanatory dimensions identified in the consultation phase: Access, Affordability and Digital Ability. These sub-indexes are also

the result of hierarchical component and variable aggregation, an example of which is displayed in Figure 1.
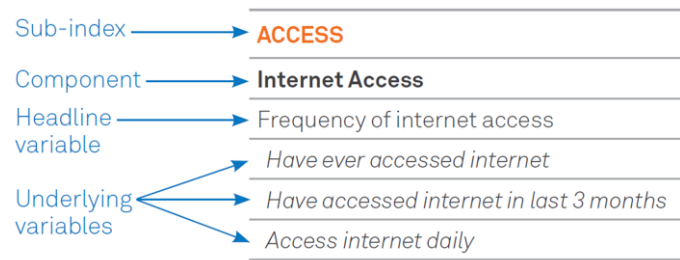
Figure 1. Example of ADII sub-index structure

Each of the sub-indexes is made up of a number of components, which have themselves been calculated as the simple average of numerous headline variables. Headline variables are thematic composites of underlying variables (generally individual survey questions) and are usually calculated as simple averages. In the example provided as Figure 1, the underlying variable 'Have ever accessed internet' feeds into the headline variable 'Frequency of internet access', itself the average of this and two other underlying variables. This headline variable then contributes in equal weight to two other headline variables (Places of internet access, and Number of internet products) to the Internet Access component. This is clear from Table 1 which provides a basic overview of the structure of the ADII. A more refined outline of the entire ADII architecture is provided as an appendix to the annual ADII reports (cf. Thomas *et al.*, 2018, pp. 46-47).

Respondents receive a score between 0 and 100 for each variable, with the upper limit indicating a higher level of inclusion. Where the variable is binary, the score received is 0 or 100. For example (from Figure 1), respondents who have not ever accessed the internet receive 0, those that have score 100. For non-binary variables (e.g. relative expenditure, which equates to percentage of household income spent on internet access) responses are ranged and recoded into a figure between 0 and 100. In this scoring system, a 'perfectly included' respondent will receive an overall ADII score of 100. Given the simple averaging architecture of the index, 100 is the 'perfectly-included' benchmark for all variables, headline variables, components and sub-indexes. Examination of the spread of results across the dataset have enabled us to generate 'inclusion' bands for the index and three sub-indexes that provide a handy comparative frame (see Table 2).

Table 1. ADII sub-index structure (components, headline variables and select underlying variables)

Access		
Internet Access	**Internet Technology**	**Internet Data Allowance**
Frequency of internet access: - Have ever accessed internet - Have accessed internet in last 3 months - Access internet daily Places of internet access: - Have accessed internet from home - Have accessed internet away from home Number of internet products: - One or more internet products - Two or more internet products	Computer technology: - Have personal computer or tablet computer in household Mobile internet technology: - Own or use mobile phone - Have mobile internet Fixed internet technology: - Have fixed broadband - Have cable or NBN fixed Broadband	Mobile internet data: - Have mobile internet - Have mobile internet data allowance over 1GB - Mobile internet data allowance relative to benchmark Fixed internet data: - Have fixed broadband - Have Fixed Broadband data allowance over 10GB - Fixed Broadband data allowance relative to benchmark
Affordability		
Relative Expenditure	**Value of Expenditure**	
Share of household income spent on internet access relative to benchmark (see Notes for Table 1)	Total internet data allowance per dollar of expenditure on internet access, relative to benchmark (see Notes for Table 1)	
Digital Ability		
Attitudes	**Basic Skills**	**Activities**
Computers and technology give me more control over my life I am interested in being able to access the Internet wherever I am I go out of my way to learn everything I can about new technology I find technology is changing so fast, it's difficult to keep up with it (negative) I keep my computer up to date with security software	General internet skills Mobile phone skills Internet banking skills Internet shopping skills Internet community skills Internet information skills	Streamed, played, or downloaded content online AV communication via the internet Internet transaction or payment Purchased or sold a product online Created or managed a site or blog Searched for advanced information

Source: Thomas et al., 2018.

Notes for Table 1: *Relative Expenditure Benchmark*: A percentage of household income expended on internet connections was derived for all respondents with internet connections in the 2016 dataset. Respondents were ranked using this percentage and divided into five equal groups to determine benchmark ranges. Respondents receive an index score based on the range they fall within as follows: 0.01–73% (100); 0.74–1.13% (75); 1.14–1.65% (50); 1.66–2.75% (25); 2.75% or more (0). Changes in affordability over time are measured against the base year of 2016. *Value of Expenditure Benchmark*: A data allowance per dollar of expenditure was derived for all respondents with internet connections in the 2016 (April 2015-March 2016) dataset. Respondents were ranked using this value and divided into five equal groups to determine benchmark ranges. Respondents receive an index score based on the range they fall within as follows: 0.01–0.1 GB/$ (0); 0.11–0.7 GB/$ (25); 0.71–2.6 GB/$ (50); 2.61–6.8 GB/$ (75); 6.81 GB/$ or more (100). Changes in affordability over time are measured against the base year of 2016.

Table 2. ADII and sub-index comparative ranges

	Low	Medium	High
ACCESS	<60	65-75	>80
AFFORDABILITY	<45	50-60	>65
DIGITAL ABILITY	<40	45-55	>60
ADII SCORE	<50	55-65	>70

Source: Thomas et al., 2018.

In the next section, we present findings from the 2018 ADII dataset. We will also present some findings based on a supplementary survey developed by the ADII research team and administered in the remote Indigenous community of Ali Curung (a community of approximately 500 people located 380 km north of Alice Springs) and with some members of the deaf and hard of hearing (DHH) community. The ADII Supplementary Survey consists of specific questions from the Roy Morgan Single Source survey used to compile the index. The majority of these questions are directly transposed. Some have minor modifications to ensure they work in the online tool used to administer the survey. The tool can be used for self-completion (as it was for the DHH community) or as part of a face-to-face method (as it was in Ali Curung). In-field pilot testing, using a national representative online panel, confirms that the composition of the ADII Supplementary Survey generates comparable scores to the Single Source.

Digital Inclusion in Australia in 2018

At a national level, the ADII reveals that digital inclusion has been steadily improving in Australia since 2014 (see Table 3). In 2018, the national ADII score was 60.2, some 6.2 points higher than in 2014 (54.0). Furthermore, in the previous 12 months the national score rose 2.2 points, the largest annual improvement since the ADII data series began in 2014. There have been marked improvements in the Access and Digital Ability dimension of digital inclusion, but progress in relation to Affordability has been poor.

Access has improved steadily over the four years since 2014, rising from 63.9 in 2014 to 73.4 in 2018. All three components of the Access sub-index have improved steadily since 2014. The Internet Access component was already relatively high at 82.7 in 2014 and has made marginal annual improvements since then to hit 87.1 in 2018. The Internet Technology and Internet Data Allowance component scores both started from a lower base and have steadily improved over the four years to 2018. The Internet Technology score rose from 68.2 in 2014 to 78.7 in 2018, while the Internet Data Allowance score rose from 40.8 in 2014 to 54.4 in 2018. These improvements reflect several developments, including the proliferation of connected consumer devices, especially smart phones, and the growing demand for data as Australians

spend more time – and do more things – online. It also reflects improvements to mobile and fixed network infrastructure, including the National Broadband Network (NBN).

There is emerging evidence that the rollout of NBN infrastructure is linked to improvements in the Internet Technology and Data Allowance components of digital inclusion. It is the 2018 ADII results for Tasmania where this link is currently most discernible. Tasmania was one of the initial NBN rollout locations and, by 2018, more than 95% of the state rollout was complete (NBN Co, 2017, p. 38). Between the 2017 and 2018 ADII periods, the percentage of Tasmanians with NBN connections more than doubled, from 29% to 60% and this development seems to have driven fixed broadband uptake, access to higher quality of fixed broadband connections, and larger data allowances (Thomas & Wilson, 2018).

At the national level, affordability has improved only marginally since 2014 and was in decline between 2014 and 2016 before a slight recovery since. The limited improvement in Affordability does not reflect rising costs. In fact, internet services are becoming less expensive as indicated by the Value of Expenditure component (a measure of gigabytes accessible per dollar spent). This affordability component increased from 51.6 points in 2014 to 60.9 points in 2018. However, while cost per gigabyte of accessible data continues to fall, Australians are spending more money on internet services, as demands and opportunities for digital participation increase. Expenditure on internet services has increased faster than household income, resulting in a consistent annual decline in the Relative Expenditure component (which reflects the share of household income spent on internet services). Overall, the proportion of household income devoted to internet services has risen from 1.00% in 2014 to 1.17% in 2018. Relative expenditure issues have had a very negative impact on Australians with lower incomes because they have less discretionary income to spend. As a result, many of Australia's more digitally excluded groups did not report any overall improvement in Affordability between 2014 and 2018. These groups include people from low income households, people with disability, people not in the labour force (NILF), and people who did not complete secondary school.

All three components of the Digital Ability sub-index have improved at the national level since 2014. The Attitudes component rose 5.1 points, Basic Skills 10.1 points, and Activities 6.9 points. However, all three have risen from a low base. The data shows that, while Australians report increasing interest in having continuous internet access, they struggle to keep up with new technologies, and relatively few users engage in more advanced activities. Furthermore, less than half of the population consider digital technologies to be empowering. Digital Ability should remain an important area for attention for policy makers, business, education, and community groups interested in improving digital inclusion.

Table 3. Australian ADII scores over time (ADII 2014-2018)

	ACCESS				AFFORDAB.			DIGITAL ABILITY				ADII
	Internet Access	Internet Technology	Internet Data Allowance	*Access Score*	Relative Expenditure	Value of Expenditure	*Affordability Score*	Attitudes	Basic Skills	Activities	*Dig Ability Score*	ADII Score
2014	82.7	68.2	40.8	**63.9**	60.3	51.6	**56.0**	45.9	46.6	34.1	**42.2**	**54.0**
2015	83.3	69.1	41.5	**64.6**	58.8	49.9	**54.3**	47.3	49.7	36.2	**44.4**	**54.4**
2016	84.4	73.0	45.7	**67.7**	55.0	52.9	**54.0**	49.2	51.7	37.2	**46.0**	**55.9**
2017	85.4	75.7	51.2	**70.8**	54.9	56.9	**55.9**	50.1	53.3	38.4	**47.3**	**58.0**
2018	87.1	78.7	54.4	**73.4**	54.3	60.9	**57.6**	51.0	56.7	41.0	**49.5**	**60.2**

Source: Thomas et al., 2018

While the ADII reveals a steady improvement in digital inclusion at the national level, it also shows that deep divisions exist between different geographic areas and demographic, economic and social groups; and, in some cases, these have increased over time (see Table 4 and Table 5). In general, rural, regional and remote Australians, older Australians and Australians with low levels of income, employment and education are significantly less digitally included.

Geography plays a critical role in digital inclusion in Australia. The ADII reveals significant differences between rural and urban areas. This 'Capital–Country gap' is evident across all three sub-indexes – Access, Affordability, and Digital Ability. In 2018, digital inclusion is 8.5 points higher in capital cities (62.4) than in country areas (53.9). The overall 'Capital–Country gap' has narrowed slightly over the past three years, from 9.5 (2015) to 8.5 (2018), but remains at the same level as it was in 2014.

Older Australians, people aged 65 years and over, record significantly lower levels of digital inclusion than younger people. A closer look at the population within this 65+ age group also reveals a pattern of diminishing digital inclusion as age increases. The ADII score for older Australians is 46.0, some 14.2 points below the national average and 20.5 points below the highest scoring age group (those aged 25-34 years). Figure 2 indicates the increasing gap between older Australians and those in the 25-34-year-old age range over the ADII data collation period (2014-2018). A key issue faced by older Australians – as with other groups reporting relatively low incomes – is the rising proportion of income spent on network access. As a result, digital communication affordability declined for older Australians between 2014 and 2018.
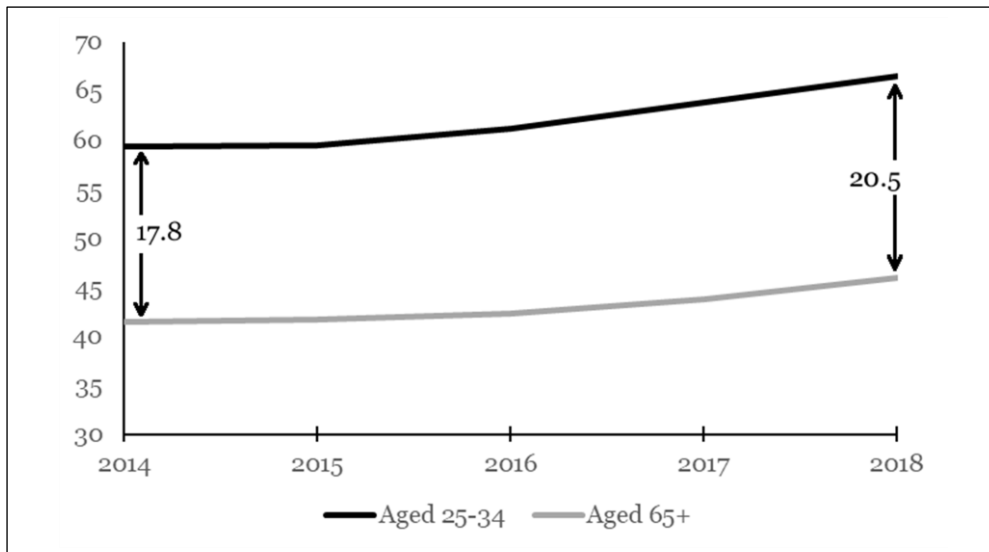
Figure 2. The age gap in digital inclusion, 2014-2018
Source: Thomas et al., 2018
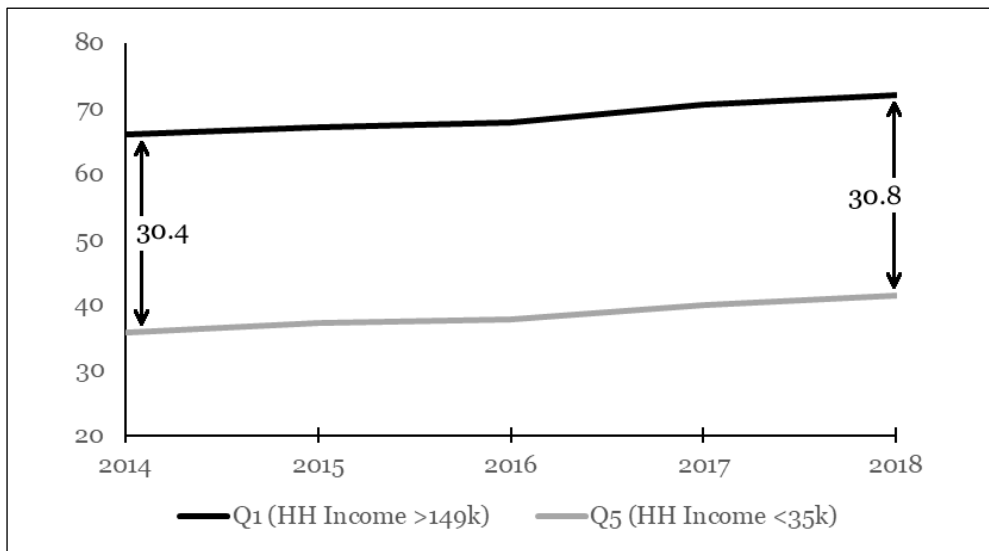
Figure 3. The income gap in digital inclusion, 2014-2018
Source: Thomas et al., 2018

The ADII illustrates a clear digital divide between richer and poorer Australians. In 2018, individuals in households with an annual income of less than $35,000 (Q5) recorded an ADII score of 41.3. This is 30.8 points lower than those living in households with an income over $150,000 (Q1) and 18.9 points below the national average score. As might be expected, it is in the area of Affordability where the digital divide between low and high-income households is greatest. People in the lowest income bracket spent approximately 3.6% of their income on network access, which translated into a Relative Expenditure score of 12.0. This is in sharp contrast to those in the highest household income bracket, who spent less than 1% of household income on network access for a Relative Expenditure score of 86.0. As Figure 3 shows, the overall digital inclusion gap between people in low and high income households has also widened since 2014. In the four years since 2014, those with the highest household

income (Q1) recorded the largest ADII gain (6.0 points) of all income quintiles. In contrast, those in the lowest income bracket (Q5) recorded a slightly smaller increase of 5.6 points.

There is also a clear 'employment gap' in digital inclusion. In 2018, the ADII score for people not in the labour force is 52.0 (8.2 points below the national average), while those who are employed have an ADII score of 65.0 (4.8 above the national average). The digital inclusion gap between those not in the labour force and employed groups has also widened since 2014, largely because of differences in the Affordability sub-index score.

The 'education gap' also remains significant. People who did not complete secondary school scored 47.4 (12.8 points below the national average). Those with a secondary education scored 58.3 (1.9 points below the national average), while tertiary-educated people scored 65.0 (4.8 points above the national average).

Some particular groups identified through the ADII as having distinct digital inclusion patterns are Indigenous Australians, Australians with disability, single parents and the mobile only (those with a mobile phone or mobile broadband device with a data allowance but no fixed connection). (See Table 4 and Table 5.)

Indigenous Australians living in urban and regional areas record low levels of digital inclusion. The ADII score for this population group in 2018 is 54.4—some 5.8 points below the national average. Indigenous Australians record scores below the national average on each of the three sub-indexes. The largest gap is in Affordability, where the score for Indigenous Australians (49.7) is 7.9 points below the national average (57.6). The prevalence of mobile-only connectivity amongst Indigenous Australians, which carries higher costs per gigabyte than fixed connections, contributes to this Affordability result. Results from the ADII Supplementary Survey conducted in the remote Indigenous community of Ali Curung suggest that remoteness further diminishes digital inclusion for Indigenous Australians, particularly with regards to Access and Affordability. The digital inclusion score for the Ali Curung community (42.9) is 17.3 points lower than the Australian average (60.2) and 11.5 points lower than that recorded by Indigenous Australians in urban and regional areas.

Australians with disability (defined in the ADII as those receiving either the disability support pension or disability pension) report low levels of digital inclusion. In 2018, the ADII score for this group was 49.2 (11.0 points below the national average). Between 2014 and 2017, the gap between people with disability and the national average narrowed, largely due to gains by this group in Access and Digital Ability. However, the gap in Affordability expanded in this period. As a consequence of a lack of improvement in Affordability and Digital Ability over 2017-2018, the overall digital inclusion gap between people with disability and other Australians has

widened again. These results represent outcomes for a distinct subset of the wider community of Australians with disability.

Table 4. ADII and sub-index comparative ranges

2018	ACCESS				AFFORDAB.			DIGITAL ABILITY				ADII
	Internet Access	Internet Technology	Internet Data Allowance	*Access Score*	Relative Expenditure	Value of Expenditure	*Affordability Score*	Attitudes	Basic Skills	Activities	*Dig Ability Score*	*ADII Score*
Australia	87.1	78.7	54.4	**73.4**	54.3	60.9	**57.6**	51.0	56.7	41.0	**49.5**	**60.2**
Capital Cities	88.8	79.9	56.5	**75.1**	56.8	63.3	**60.0**	53.1	59.3	43.8	**52.1**	**62.4**
Rural	82.5	74.8	47.9	**68.4**	47.2	53.5	**50.4**	45.3	49.6	33.6	**42.9**	**53.9**
Aged 14-24	91.7	80.6	57.2	**76.5**	60.8	63.9	**62.4**	64.3	56.3	43.5	**54.7**	**64.5**
Aged 25-34	92.3	83.4	65.7	**80.5**	52.6	64.3	**58.4**	60.2	68.5	52.8	**60.5**	**66.5**
Aged 35-49	93.0	83.0	62.2	**79.4**	56.9	63.7	**60.3**	53.3	67.2	48.7	**56.4**	**65.4**
Aged 50-64	86.1	77.9	50.8	**71.6**	55.3	59.9	**57.6**	44.3	54.5	36.4	**45.1**	**58.1**
Aged 65+	72.4	68.2	36.1	**58.9**	44.2	51.1	**47.7**	34.3	36.2	23.9	**31.5**	**46.0**
Household Inc >149k (Q1)	94.7	84.6	63.9	**81.0**	86.0	64.6	**75.3**	58.2	69.8	51.5	**59.8**	**72.1**
Household Inc <35k (Q5)	72.0	66.2	38.8	**59.0**	12.0	50.0	**31.0**	36.6	38.1	26.7	**33.8**	**41.3**
Employed	92.2	82.6	60.8	**78.5**	58.6	62.9	**60.8**	55.3	64.9	47.1	**55.8**	**65.0**
Unemployed	88.1	79.0	56.6	**74.6**	48.3	61.7	**55.0**	56.7	57.3	45.7	**53.2**	**60.9**
Not in Labour Force	78.9	72.3	43.5	**64.9**	47.7	57.0	**52.3**	43.1	43.3	30.4	**38.9**	**52.0**
Below Secondary Ed.	73.4	68.3	40.7	**60.8**	46.4	51.7	**49.0**	34.7	37.5	24.6	**32.2**	**47.4**
Completed Secondary Ed.	86.9	78.6	54.4	**73.3**	50.0	60.2	**55.1**	48.1	54.8	36.7	**46.5**	**58.3**
Completed Tertiary Ed.	91.6	82.5	59.3	**77.8**	57.7	63.8	**60.8**	54.2	66.1	49.0	**56.4**	**65.0**
Disability	75.1	70.9	47.8	**64.6**	38.6	53.2	**45.9**	42.6	40.0	28.8	**37.1**	**49.2**
Indigenous Australians	82.4	73.5	49.6	**68.5**	48.1	51.3	**49.7**	53.4	47.2	34.3	**45.0**	**54.4**
Mobile Only	74.8	60.0	29.7	**54.8**	55.3	10.7	**33.0**	43.2	46.1	31.5	**40.3**	**42.7**
Two Parent Families	92.6	83.8	63.4	**79.9**	59.7	66.3	**63.0**	53.6	67.1	49.0	**56.6**	**66.5**
Single Parent Families	89.2	79.2	58.3	**75.5**	32.2	58.4	**45.3**	45.9	58.1	41.8	**48.6**	**56.5**
DHH Survey Resp. (n: 115)	97.4	89.9	83.6	**90.3**	32.2	68.9	**50.5**	77.6	89.8	80.4	**82.6**	**74.5**
Ali Curung Resp. (n: 112)	64.3	40.5	37.2	**47.3**	39.6	12.1	**25.8**	47.7	64.5	44.8	**52.3**	**42.9**

Source: Thomas et al., 2018

Notes for Table 4 and Table 5: Single Parent Families is based on survey responses from parents living in a single-parent household structure with children aged under 18 years; Two-Parent Families is based on survey responses from parents living in a two-parent household structure with children aged under 18 years; Household Income categories based on division of population into five equal quintile bands; Disability: Respondents receiving the disability support pension or disability pension; Mobile only: respondents with a mobile phone or mobile broadband device with a data allowance, but no fixed connection.

Table 5. ADII and sub-index comparative ranges

Change 2014 - 2018	ACCESS				AFFORDAB.			DIGITAL ABILITY				ADII
	Internet Access	Internet Technology	Internet Data Allowance	*Access Score*	Relative Expenditure	Value of Expenditure	*Affordability Score*	Attitudes	Basic Skills	Activities	*Dig Ability Score*	*ADII Score*
Australia	+4.4	+10.5	+13.6	**+9.5**	-6.0	+9.3	**+1.6**	+5.1	+10.1	+6.9	**+7.3**	**+6.2**
Capital Cities	+4.0	+9.0	+12.4	**+8.5**	-4.9	+7.4	**+1.2**	+5.1	+10.6	+7.7	**+7.8**	**+5.9**
Rural	+5.0	+12.3	+14.4	**+10.6**	-9.1	+10.6	**+0.8**	+4.8	+8.9	+5.1	**+6.3**	**+5.9**
Aged 14-24	+1.9	+8.2	+13.9	**+8.0**	-1.7	+11.6	**+5.0**	+2.1	+7.4	+5.8	**+5.1**	**+6.0**
Aged 25-34	+1.7	+9.4	+16.6	**+9.3**	-4.2	+17.0	**+6.4**	+3.9	+6.7	+6.4	**+5.7**	**+7.2**
Aged 35-49	+3.0	+8.8	+14.1	**+8.6**	-3.4	+12.7	**+4.7**	+6.7	+10.5	+8.9	**+8.7**	**+7.4**
Aged 50-64	+5.3	+11.5	+13.2	**+10.0**	-8.7	+7.8	**-0.5**	+5.5	+13.2	+7.1	**+8.7**	**+6.1**
Aged 65+	+12.6	+15.8	+12.0	**+13.5**	-12.1	-6.7	**-9.4**	+6.5	+14.4	+7.5	**+9.5**	**+4.5**
Household Inc >149k (Q1)	+1.4	+7.1	+11.6	**+6.7**	-0.7	+11.9	**+5.6**	+2.4	+6.7	+7.4	**+5.5**	**+6.0**
Household Inc <35k (Q5)	+9.1	+12.6	+13.0	**+11.6**	-6.0	+1.9	**-2.1**	+4.4	+11.1	+6.5	**+7.3**	**+5.6**
Employed	+1.8	+8.9	+13.4	**+8.0**	-4.7	+12.0	**+3.7**	+4.5	+8.8	+6.8	**+6.7**	**+6.1**
Unemployed	+5.8	+11.2	+18.7	**+11.9**	-3.0	+14.3	**+5.7**	+4.5	+11.4	+7.6	**+7.8**	**+8.4**
Not in Labour Force	+8.7	+12.9	+12.7	**+11.5**	-8.4	+2.9	**-2.8**	+6.4	+12.1	+7.1	**+8.5**	**+5.7**
Below Secondary Ed.	+8.6	+12.1	+12.7	**+11.2**	-7.1	+3.3	**-2.0**	+4.9	+10.6	+7.1	**+7.5**	**+5.6**
Completed Secondary Ed.	+5.0	+10.9	+14.7	**+10.2**	-6.7	+10.1	**+1.7**	+5.9	+11.1	+5.4	**+7.4**	**+6.5**
Completed Tertiary Ed.	+1.6	+8.9	+11.9	**+7.5**	-7.2	+11.5	**+2.2**	+3.0	+8.0	+6.2	**+5.7**	**+5.1**
Disability	+6.7	+11.8	+17.1	**+11.8**	-5.7	+2.9	**-1.4**	+8.1	+11.2	+6.9	**+8.7**	**+6.4**
Indigenous Australians	+6.9	+11.9	+17.7	**+12.2**	-7.3	+15.1	**+3.9**	+11.2	+11.3	+11.2	**+11.3**	**+9.1**
Mobile Only	-9.6	-3.8	+2.2	**-3.8**	-5.7	+2.0	**-1.8**	-0.8	-0.3	-0.8	**-0.6**	**-2.1**
Two Parent Families	+1.1	+7.9	+12.8	**+7.2**	-2.2	+14.4	**+6.1**	+4.2	+9.2	+8.8	**+7.4**	**+6.9**
Single Parent Families	+4.6	+9.1	+16.8	**+10.1**	-4.5	+15.5	**+5.5**	+2.6	+11.0	+7.3	**+7.0**	**+7.6**
DHH Survey Resp. (n: 115)	N/A	N/A	N/A	**N/A**	N/A	N/A	**N/A**	N/A	N/A	N/A	**N/A**	**N/A**
Ali Curung Resp. (n: 112)	N/A	N/A	N/A	**N/A**	N/A	N/A	**N/A**	N/A	N/A	N/A	**N/A**	**N/A**

Source: Thomas et al., 2018

In 2018, the ADII research team conducted a supplementary survey of members of the deaf and hard of hearing (DHH) community. Results from this survey reveal very high levels of digital access and digital ability compared to the national average, but these are tempered by a lower level of affordability. On average, respondents spent a high proportion of their household income on internet access (captured by the relative expenditure component). The result for this component recorded by the DHH community was 32.2. Although a similar result

(36.3) is posted by the ADII Disability cohort, there is a different dynamic at play here. The ADII Disability cohort rely on disability pensions, translating moderate expenditure on internet access into a poor Relative Expenditure result. By contrast, four in five DHH survey respondents were employed and it was a high internet spend (43% above average) that resulted in poor Relative Expenditure.

The socio-economic disadvantage of single parent families with dependent children is well documented (AIHW, 2017). Single parents have low rates of employment and many rely entirely on government benefits, which results in very low household income – more than 20% live below the poverty line (Wilkins, 2017). The ADII reveals this socio-economic disadvantage translates into digital disadvantage. Overall, single parent families have an ADII score of 56.5, 3.7 points lower than the national average and 10.0 points lower than two-parent families. Although their Access sub-index score is higher than the national average, single parent families are less likely to invest in fixed broadband access than other Australians (67.0% versus 72.9% national average). A greater dependence on rental housing and the higher levels of uncertainty and mobility this entails is one barrier to fixed broadband investment. A greater reliance on mobile-only access translates into lower levels of engagement in higher-bandwidth streaming and communication activities by single parent families. Instead, single parents are more likely to engage in functional online activities, such as financial transactions and government interactions. Affordability is the key barrier to greater digital inclusion for single parents. The impact of internet access on single-parent family budgets is substantial – it accounts for 2% of their household income compared to the national average of 1.17%. This results in a Relative Expenditure score of 32.2 – 22.1 points lower than the national average. With greater reliance on mobile connections, single parents, on average, get poorer value for money than other Australians – their Value of Expenditure score is 58.4 compared to the national average of 60.9.

More than four million Australians access the internet solely through a mobile connection – this means they have a mobile phone or mobile broadband device with a data allowance, but no fixed connection (Thomas *et al.*, 2018, p. 16). In 2018, mobile-only users have an ADII score of 42.7, some 17.5 points below the national average (60.2). Being mobile-only not only diminishes access, but also impacts on the affordability and digital ability aspects of digital inclusion. Mobile-only use is clearly linked with a number of socio-economic factors, with people in the lowest household income quintile (29.6%), those with low levels of education (27.2%) and the unemployed (27.0%) more likely to be mobile-only. Single parents (28.8%) and Australians with disability (30.6) are also more likely to be mobile only, as are Indigenous Australians (34.7%) — indeed, all of the respondents with internet connections in the Ali Curung Supplementary Survey sample were mobile only. The position of mobile-only users

that emerges from the ADII research of course reflects the current, somewhat transitional, state of the mobile internet and the resource-constrained consumer hardware and software that support it. We make no assumptions about the affordances of future mobile systems, and we note the appearance of 5G, fast Wi-Fi, and increasingly capable mobile apps and devices. Improved mobile networks and new, AI-enabled, voice-based interfaces are just two of the developments in this area that have the potential to enhance substantially the future capabilities for mobile-only users.

Conclusion

In the past two decades, digital inequality has come to be understood as a complex, evolving and critical issue in Australia, as it has elsewhere. The conceptualization of digital inequality has shifted away from a binary logic that divided those with and without internet access towards a recognition of the multifaceted and graduated nature of online experiences and outcomes, and the increasing interrelationships between online and offline experiences of inclusion/exclusion. Such a shift has generated demand from policy makers, consumer advocates and researchers for more complex tools to measure digital inequality. One approach that has emerged is the formation of aggregate measures that systematically combine a suite of digital inequality indicators. The Australian Digital Inclusion Index (ADII), developed in 2015 and now including annual data covering the period 2014 to 2018, is one such composite index.

The 2018 ADII data reveals that digital inclusion is improving in Australia, with positive changes in relation to the nature and extent of personal internet access and digital abilities since 2014. Affordability of internet access remains a substantial issue. Although internet services are becoming less expensive, the proportion of household income spent on these services continues to rise, as Australians spend more time and do more things online. For those on low incomes, the pressure of greater spending on network access, which is now an essential rather than discretionary service, is of great concern. Indeed, affordability is one of the key fault lines of digital inclusion in Australia.

The ADII reveals that digital access, affordability and abilities continue to follow distinct geographic, social and socio-economic contours. In general, rural and regional Australians, older Australians and Australians with low levels of income, employment, and education are less digitally included than their compatriots. Despite the promise of better network infrastructure through the NBN, for some of these groups the inclusion gap is widening. As the internet and its associated technologies become increasingly embedded in Australians' economic, social and cultural lives, the risks of entrenching an increasingly unequal digital economy are also growing. Over a long period, Australian governments have supported

inclusive and accessible national communications; that goal has played a significant part in the horizontal equity objective underpinning the NBN. The ADII adds to the growing body of evidence that a more systematic approach to policy beyond the improvement of network infrastructure is now necessary. The realization of an inclusive Australian internet will demand further interventions, especially in the areas of affordability and skills.

Acknowledgement

References

ABS. 2011. *Census of Population and Housing 2011,* Australian Bureau of Statistics, Canberra.

ABS. 2013. *Programme for the International Assessment of Adult Competencies, Australia 2011-12*, Cat 4228.0, Australian Bureau of Statistics, Canberra.

ABS. 2014. *Household Use of Information Technology, Australia, 2012-13*, Cat 8146.0, Australian Bureau of Statistics, Canberra.

ABS. 2016. *Census of Population and Housing 2016: Table Builder,* Australian Bureau of Statistics, Canberra.

ACMA. 2015. *Communications Report 2014-15*, Australian Communications and Media Authority, Melbourne.

AIHW. 2017. *Australia's Welfare 2017*, Australia's welfare series no. 13, Australian Institute of Health and Welfare, Canberra.

Andreasson, KJ. 2015. 'Introduction', in KJ Andreasson (ed) 2015, *Digital divides: the new challenges and opportunities of e-inclusion*, Boca Raton: CRC Press, pp. xxi-xxviii.

Barzilai-Nahon, K. 2006. 'Gaps and Bits: Conceptualizing Measurements for Digital Divide/s', *The Information Society,* vol. 22**,** no. 5, pp. 269-278.

Bruno, G; Esposito, E; Genovese, A; Gwebu, KL. 2011. 'A Critical Analysis of Current Indexes for Digital Divide Measurement', *The Information Society,* vol. 27**,** no. 1, pp. 16-28.

Crandall, M; Fisher, KE (eds). 2009. *Digital Inclusion: Measuring the Impact of Information and Community Technology,* New Jersey: Information Today Inc.

Dewan, S; Riggins, F, J. 2005. 'The Digital Divide: Current and Future Research Directions', *Journal of the Association for Information Systems,* vol. 6**,** no. 12, pp. 298-336.

DiMaggio, P; Hargittai, E. 2001. *From the 'Digital Divide' to 'Digital Inequality': Studying Internet Use as Penetration Increases*, Working Paper-Center for Arts and Cultural Policy Studies, New Jersey.

DiMaggio, P; Hargittai, E; Celeste, C; Shafer, S. 2004. 'From Unequal Access to Differentiated Use: A Literature Review and Agenda for Research on Digital Inequality', in KM

Neckerman (ed) 2004, *Social inequality*, New York: Russell Sage Foundation, pp. 355-400.

EIU. 2019. *The Inclusive Internet Index 2019: Executive Summary*, The Economist Intelligence Unit for Facebook, London.

Gunkel, DJ. 2003. 'Second Thoughts: Toward a Critique of the Digital Divide', *New Media & Society,* vol. 5**,** no. 4, pp. 499-522.

Hargittai, E. 2002. 'Second-Level Digital Divide: Differences in People's Online Skills', *First Monday*, vol. 7, no. 4 (online).

Hargittai, E. 2003. 'The Digital Divide and What To Do About It', in DC Jones (ed) 2003, *New economy handbook*, San Diego: Academic Press, pp. 821-839.

Helsper, EJ. 2008. *Digital inclusion: An analysis of social disadvantage and the information society*, Department for Communities and Local Government, London.

Helsper, EJ; van Deursen, AJAM. 2015. 'The Third-Level Digital Divide: Who Benefits Most from Being Online?', in L Robinson; SR Cotton; J Schulz; TM Hale; A Williams (eds) 2015, *Communication and Information Technologies Annual*, Bingley: Emerald Group Publishing Limited, pp. 29-52.

ITU. 2009. *Measuring the Information Society: The ICT Development Index*, International Telecommunication Union, Geneva.

Kwok-Kee, W; Hock-Hai, T; Hock Chuan, C; Tan, BCY. 2011. 'Conceptualizing and Testing a Social Cognitive Model of the Digital Divide', *Information Systems Research,* vol. 22**,** no. 1, pp. 170-187.

Lloyds Bank. 2018. *UK Consumer Digital Index 2018*, Lloyds Bank, London.

Mubarak, F. 2015. 'Towards a renewed understanding of the complex nerves of the digital divide', *Journal of Social Inclusion,* vol. 6**,** no. 1, pp. 71-103.

Nardo, M; Saisana, M; Saltelli, A; Tarantola, S; Hoffman, A; Giovannini, E. 2008. *Handbook on Constructing Composite Indicators and User Guide*, OECD, Paris.

NBN Co. 2017. *NBN Co Corporate Plan 2018-2021*, NBN Co, Sydney.

NTIA. 1995. *Falling through the net: A Survey of the "Have Nots" in Rural and Urban America.*, U.S. Dept of Commerce National Telecommunications and Information Administration, Washington, D.C.

NTIA. 1998. *Falling through the net II: new data on the digital divide*, U.S. Dept of Commerce National Telecommunications and Information Administration, Washington, D.C.

NTIA. 1999. *Falling through the net: defining the digital divide*, U.S. Dept of Commerce National Telecommunications and Information Administration, Washington, D.C.

Park, S; Jae Kim, G. 2014. 'Lessons from South Korea's Digital Divide Index', *Info: The Journal of Policy, Regulation and Strategy for Telecommunications, Information and Media,* vol. 16**,** no. 3, pp. 72-84.

Roy Morgan Research Ltd. 2015. *Single Source Survey*, Roy Morgan Research Ltd, Melbourne.

Roy Morgan Research Ltd. 2017. *How we collect and process Single Source data in Australia*, Roy Morgan Research Ltd, Melbourne.

Roy Morgan Research Ltd. 2018. *How We Score Our Customer Satisfaction*, Melbourne, Available: https://www.customersatisfactionawards.com/how-we-score [Accessed 26 Feb 2019].

Sassi, S. 2005. 'Cultural differentiation or social segregation? Four approaches to the digital divide', *New Media & Society,* vol. 7**,** no. 5, pp. 684-700.

Selwyn, N. 2004. 'Reconsidering Political and Popular Understandings of the Digital Divide', *New Media & Society,* vol. 6**,** no. 3, pp. 341-362.

SISR; CSI; Telstra. 2015. *Australian Digital Inclusion Index Discussion Paper*, Swinburne Institute for Social Research, Centre for Social Impact, Telstra Corporation Ltd, Melbourne.

Strover, S. 2003. 'Remapping the Digital Divide', *The Information Society,* vol. 19**,** no. 4, pp. 275-277.

The Smith Family. 2013. *Sport, culture and the internet: Are Australian children participating?*, The Smith Family, Melbourne.

The Tech Partnership. 2017. *Get Digital Heatmap*, The Tech Partnership, London, http://heatmap.thetechpartnership.com/.

Thomas, J; Barraket, J; Wilson, CK; Cook, K; Louie, Y; Holcombe-James, I; Ewing, S; MacDonald, T. 2018. *Measuring Australia's Digital Divide: The Australian Digital Inclusion Index 2018*, RMIT University for Telstra, Melbourne.

Thomas, J; Wilson, CK. 2018. 'Digital inclusion in Tasmania has improved in line with NBN rollout – will the other states follow? (Commentary)', *The Conversation*, 29 August.

Tsatsou, P. 2011. 'Digital divides revisited: what is new about divides and their research?', *Media, Culture & Society,* vol. 33**,** no. 2, pp. 317-331.

van Deursen, A; Helsper, E; Eynon, R; van Dijk, J. 2017. 'The Compoundness and Sequentiality of Digital Inequality', *International Journal of Communication;* vol. 11, pp. 452-473.

van Dijk, J. 2006. 'Digital divide research, achievements and shortcomings', *Poetics,* vol. 34**,** no. 4, pp. 221-235.

van Dijk, J. 2005. *The Deepening Divide: Inequality in the Information Society,* Thousand Oaks: SAGE Publications.

Walton, P; Kop, T; Spriggs, D; Fitzgerald, B. 2013. 'A digital inclusion: Empowering all Australians', *Australian Journal of Telecommunications and the Digital Economy,* vol. 1**,** no. 1, pp. 2-17.

Wilkins, R. 2017. *The Household, Income and Labour Dynamics in Australia Survey: Selected Findings from Waves 1 to 15*, Melbourne Institute of Applied Economic and Social Research, Melbourne.