Australian Journal of Telecommunications and the Digital Economy

Volume 6 Issue 3

September 2018

Published by Telecommunications Association Inc. ISSN 2203-1693

AJTDE Volume 6, Number 3, September 2018

Table of Contents

Editorial

Digital Economy Focus ii
Mark A Gregory

Public Policy Discussion

Fixed Broadband Penetration and Economic Growth 1
Yudi Purnama Purnama

Framework for Demystifying M2M Spectrum Regulation 17
Arpit Khosla, Anuradha C.B.

Articles

Bitmaps & Bitmasks: Efficient Tools to Compress Deterministic Automata 41
Shiva Shankar Subramanian, PinXing Lin, Andreas Herkersdorf, Thomas Wild

History of Australian Telecommunications

Communications for the America's Cup Challenge 76
Simon Moorhead

Digital Economy Focus

Editorial

Mark A Gregory
RMIT University

Abstract: Papers in the September 2018 issue of the *Journal* provide an interesting mix of public policy debate, technology and the communications for the America's Cup Challenge defence held in Fremantle, Australia. The public policy papers cover fixed broadband adoption and economic growth in ASEAN and a framework to demystify machine-to-machine spectrum regulation. A technical paper on bitmaps and bitmasks provides an insight into the latest tools and techniques. The history of Australian telecommunications paper on the America's Cup communications solution provides an insight into what was a successful and difficult-to-achieve outcome for one of the world's major sporting events. The *Journal* welcomes contributions.

In This Issue

In this issue of the *Journal* papers cover public policy, historical events and new technology solutions. The breadth of papers highlights the exciting changes occurring within our industry both locally and internationally.

Fixed Broadband Take-up and Economic Growth presents an interesting look at the causal relationships between fixed broadband take-up, gross capital accumulation, the degree of urbanisation, and economic growth, using panel data from 9 ASEAN member countries in 2003-2016.

Framework for Demystifying M2M Spectrum Regulation proposes a framework that aims to demystify the spectrum policy regulation in the age of machine-to-machine communications and the Internet of Things.

Bitmaps & Bitmasks: Efficient Tools to Compress Deterministic Automata provides an approach to efficiently compress deterministic automata for deep packet inspection.

Communications for the America's Cup Challenge provides two historic papers from a special issue of the *Journal* in 1986 featuring the communication requirements for the America's Cup 1986/87 challenge defence in Fremantle, Australia.

http://doi.org/10.18080/ajtde.v6n3.163

The drive towards a digital economy

The momentum for Australian business and industry to adopt a proactive approach to becoming a part of the global digital economy continues to grow. As the global economies shift from fossil fuels to a range of modern alternatives, there is an underlying shift in the economic focus of the nations, like Australia, that have been major exporters of fossil fuels.

Australia is currently underperforming in the global digital economy and our ranking amongst the 16 most industrialised nations is middle of the pack. Our ranking for global resilience is falling and similarly for our ranking in the global innovation index.

There are a range of factors affecting the nation's prosperity and one of the factors that will become more important over time is investment in all things digital.

Telecommunications provides the fundamental platform for the global digital economy. To ensure that our participation in the global digital economy increases, a mix of public and private investment should occur over the next decade to enhance the opportunities for business and industry, irrespective of where they're located.

The drive towards a digital economy requires careful planning and consideration of how to meet the demands of regional and remote Australia, the powerhouse of the Australian economy. It will be through innovative ideas that are supported by public and private funding that the nation will enhance participation outside the urban areas.

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

The *Journal*, Looking Forward

Australian telecommunications is moving forward at a rapid rate, and the introduction of 5G next year will speed up the reach and utilisation of telecommunication services. The *Journal* is calling for papers on how 5G will affect Australian telecommunications consumers.

The National Broadband Network (NBN) is soon to enter a new phase and papers on what should be done with the NBN are welcomed.

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 telecommunications market is now, how it got to where it is, and what is going to happen next.

http://doi.org/10.18080/ajtde.v6n3.163

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 the coming year.

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

Fixed Broadband Penetration and Economic Growth:

Evidence from panel data of ASEAN member countries

Yudi Adhi Purnama
Graduate School of Asia-Pacific Studies, Waseda University, Japan

Abstract: This study explores causal relationships between fixed broadband penetration, gross capital accumulation, the degree of urbanisation, and economic growth, using panel data from 9 ASEAN member countries in 2003-2016. Employing multivariate models, we reveal bidirectional causality between fixed broadband penetration and economic growth, with a higher magnitude of effect from GDP per capita to broadband penetration. Furthermore, fixed broadband penetration has bidirectional causality to urbanisation, while a unidirectional link exists from fixed broadband penetration to gross capital accumulation.

Keywords: ASEAN, economic growth, fixed broadband, Granger causality, gross capital accumulation

Introduction

Information and Communication Technology (ICT) has developed rapidly in recent decades and is viewed as an engine of development in the fourth industrial development phase. Telecommunication infrastructure as part of ICT has also evolved and offers better services such as broadband. In 2003, the International Telecommunication Union (ITU) standard defined broadband as access with a speed higher than ISDN, around 1.5 or 2.0 Megabits per second (Mbps). The always-on connection and higher rate of broadband offer more experiences and services, such as video streaming, online gaming, e-commerce and business solutions like cloud computing and the Internet of Things (IoT).

Broadband transforms people's lifestyle through online shopping, distance learning and teleworking. Broadband also facilitates less developed and emerging countries to expand and compete in the global economy. In a rural-urban context, broadband has brought equal opportunity in accessing information, learning and technology dissemination, thereby escalating economic capacity. In summary, broadband supports higher levels of development.

http://doi.org/10.18080/ajtde.v6n3.151

The important role of broadband in economic growth encourages abundant research at the global, regional, national, firm or household level. Despite the enthusiasm, limited studies exist for emerging economies such as those of most countries in the Association of Southeast Asian Nations (ASEAN). Furthermore, the impact of broadband on the economy needs further study because previous work mostly measures the impact of pre-broadband technology such as teledensity or dial-up. Moreover, most of the studies explore the correlation of broadband to economic growth, while causality between factors is not examined.

Another consideration is the integration spirit of ASEAN. The Southeast Asian region is collectively the third largest economy in Asia and the seventh largest in the world, with 622 million people and a US$2.6 trillion market. In 2015, the members established the ASEAN Economic Community (AEC), envisioning economic integration and shared prosperity. As an economic community, AEC also aimed to foster connectivity and cooperation. This vision makes studying ASEAN as a single economic entity plausible.

Therefore, this paper will explore causal relationships between fixed broadband penetration and economic growth in relation to growth of capital accumulation and urbanisation in ASEAN countries in the period 2003-2016. Employing a well-known Granger causality test, our multivariate research presents robust results, since ASEAN consists of countries with different levels of development. Furthermore, the long time span allows for implementing lags, based on the assumption that economic factors require time to have an effect on economic growth.

This paper is structured as follows. Section two gives a brief literature review of broadband and its economic effect, while section three describes broadband development in ASEAN. Section four describes the data and empirical model, and section five offers econometric analysis and empirical results. Section six includes discussion of the test results and a conclusion. Throughout this paper, we use the term "broadband" to mean fixed broadband.

Broadband and the Economy

ICT including broadband plays a substantial role in modern civilisation, especially in the last two decades. The advancement of telecommunication, specifically Internet access from narrowband to broadband, brings transformation, not only in terms of speed but also of varied applications and access quality (Cawley & Preston, 2007; Crandall, Lehr & Litan, 2007; Grubesic, 2004; Lau, Kim & Atkin, 2005). More sophisticated technology, such as multimedia, music and gaming, and later artificial intelligence, IoT and cloud services, have marked the vital role of broadband in every aspect of life.

http://doi.org/10.18080/ajtde.v6n3.151

Broadband access also has a positive impact on the economy, not only in terms of growth but also of income per capita (Bertschek *et al.*, 2015; Van Gaasbeck, 2008). Moreover, broadband development has brought down the price to more affordable levels, increasing user numbers as well as broadband externality. The high impact of broadband on the economy positions it to help emerging economies compete with other countries, including developed ones.

Furthermore, broadband catalyses employment growth (Bouras, Diles & Kokkinos, 2013; Katz *et al.*, 2010; Van Gaasbeck, 2008) and facilitates new business establishments (Kim & Orazem, 2017) such as creative industries and e-commerce. Moreover, broadband also impacts regional development (Bertschek *et al.*, 2015), encourages innovation (Fanfalone, 2015; Ng, Lye & Lim, 2013; Rampersad & Troshani, 2013) and increases business productivity (Czernich *et al.*, 2011; Falch, 2007; Van Gaasbeck, 2008).

In contrast, some studies suggest a weak impact from broadband on the manufacturing sector (Bertschek *et al.*, 2015; Holt & Jamison, 2009; United Nations, 2015) compared to service industries such as tourism, finance and banking. Broadband is positioned as a general-purpose technology that benefits the service sectors more because it affects service mobility, which is challenging for manufacturing industries.

Furthermore, broadband helps human resources development (Kolko, 2012; LaRose *et al.*, 2014). Broadband enables people to access knowledge and technology, giving the same opportunity to developed and emerging countries. Broadband also helps students, teachers and managers to access a broad range of services, content and applications (Tookey, Whalley & Howick, 2006). Moreover, broadband will decrease learning costs by using e-education, enabling knowledge dissemination despite distance, especially in rural areas (Falch & Henten, 2010). Better education will also increase people's capacity for work and finally support growth (Bertschek *et al.*, 2015).

On the other hand, some studies (Bojnec *et al.*, 2012; Grosso, 2006; Lemstra, Voogt & Van Gorp, 2015; OECD, 2002, 2008) found economic growth has affected broadband penetration. Broadband subscribers are increasing in line with rising income per capita as broadband becomes more affordable. In less developed nations, where income per capita is low as a result of low economic growth and productivity, broadband penetration is low. In these countries, broadband subscription services are expensive, since broadband infrastructure requires high investment in comparison to the low economic scale.

In summary, broadband plays a significant role in facilitating economic growth, increasing human resources capacity and closing the development gap between developing and developed countries.

ASEAN and Broadband Development

The formation of ASEAN in Bangkok, Thailand, on 8 August 1967, was marked by the Bangkok Declaration, known as the ASEAN Declaration (Asean.org, 2017). ASEAN was established as a community for economic cooperation and social and cultural development to build a prosperous and peaceful population in Southeast Asia, with mutual respect and equality between members. Furthermore, at the 9th ASEAN Summit in 2007, all members committed to the establishment of the ASEAN community by 2015 (Asean.org, 2017). The community has three pillars: ASEAN Political-Security Community; ASEAN Economic Community; and ASEAN Socio-Cultural Community. The three components are an implementation of ASEAN Vision 2020 to build and manage Southeast Asia as outward looking, living in peace, stability and prosperity, and promoting development through partnership between nations.

The ASEAN initiative for ICT development was marked by the signing of an e-ASEAN framework agreement at the ASEAN Informal Summit in Singapore in November 2000. The agreement focused on six primary actions: establishment of ASEAN information infrastructure; growth of electronic commerce; liberalisation of trade in ICT products and services and of investments; facilitation of trade in ICT products and services; capacity building and e-society; and e-government. Following the e-ASEAN agreement, in 2001, the first ASEAN Telecommunications Ministers Meeting (TELMIN) was held at Kuala Lumpur, Malaysia. At that event, ASEAN members realised the importance of ICT infrastructure and agreed to reduce the digital divide between member countries and develop an ASEAN information infrastructure. The e-ASEAN Integration Roadmap (2004-2010) became part of the ASEAN Framework Agreement for the Integration of Priority Sectors, and ICT was identified as one of 11 key sectors (International Development Research Centre, 2008).

After a series of meetings, ASEAN formulated the ASEAN ICT Masterplan 2015 (AIM2015) in 2011 at the 10th TELMIN in Kuala Lumpur. The AIM2015 aims to establish ICT as an engine of growth for ASEAN Member States; achieve recognition for ASEAN as a global ICT hub; enhance the quality of life for the people of ASEAN; and contribute towards ASEAN integration. The outcome will be achieved by implementing six pillars: economic transformation; people empowerment and engagement; innovation; infrastructure development; human capital development; and bridging the digital divide (Asean.org, 2011). To address the ICT sector dynamics, in November 2015, ASEAN members agreed to expand AIM2015 into the ASEAN ICT Masterplan 2020 (AIM2020), which added several points: ICT in the single market, new media and content and information security and assurance (Asean.org, 2016).

The ASEAN members agreed on infrastructure availability, especially broadband, to support ICT development. In the Bali Declaration 2008, members agreed to forge a partnership to develop high-speed connections to close the digital divide (Asean.org, 2008) and later to implement broadband across ASEAN (Asean.org, 2009).

Broadband development in the ASEAN region varies between countries because their development stages are also different. Different states of broadband are affected mostly by lack of affordable services and inaccessibility to broadband infrastructure (Ruddy & Ozdemir, 2013). For some countries, the unweighted average price of a 1 Mbps annual subscription reaches approximately 30% of GDP per capita, while the lowest price is 0.1% GDP per capita in Singapore, compared to 132.8% of GDP per capita in Myanmar (see Figure 1 for details).

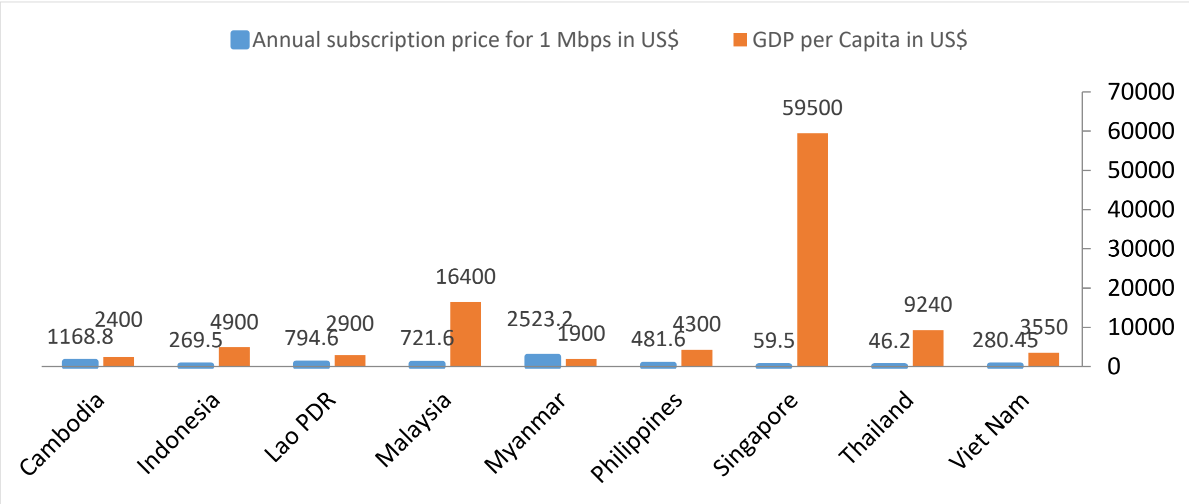


Figure 1. Annual fixed broadband subscription price and GDP per capita in ASEAN member countries in 2012. Data retrieved from Ruddy & Ozdemir (2013).

Moreover, broadband coverage in the ASEAN region is concentrated in city centres since economic scale matters. Geographical obstacles also exist because some ASEAN member countries have extreme land conditions that hinder infrastructure deployment and diminish people's opportunities to access broadband services (see Figure 2). In addition to the backbone, limited last-mile coverage also hinders broadband development because reaching subscribers' premises is difficult. Some member countries tackle this challenge by employing DSL as a transitional technology (Ruddy & Ozdemir, 2013) to lower deployment cost amid a low number of fixed phone subscribers.

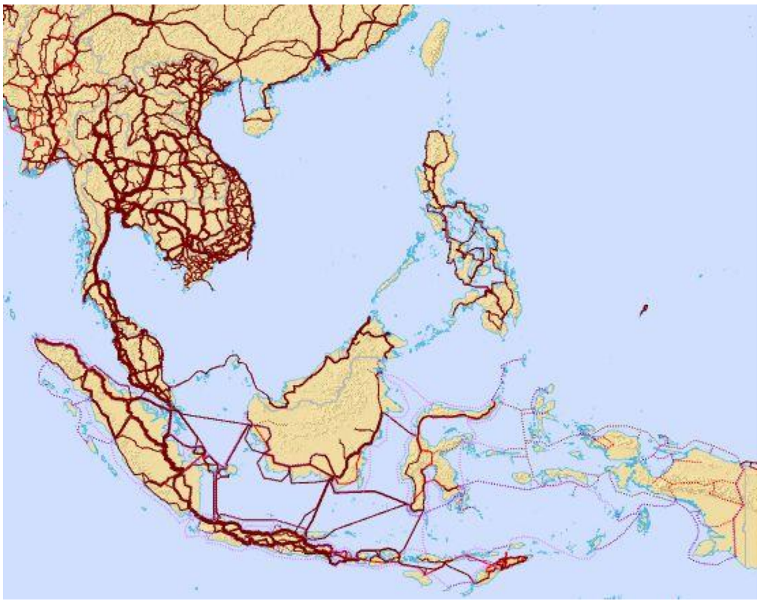


Figure 2. Fibre optic backbone map in Southeast Asia region. Reprinted from ITU interactive terrestrial transmission/ESCAP Asia-Pacific information superhighway maps by ITU, 19 November 2017, retrieved from http://www.itu.int/itu-d/tnd-map-public/ . © Copyright 2017 by ITU.

Data and Empirical Model

Data

This study utilized data of ASEAN member countries between 2003 and 2016 from the World Bank Development Index (WDI). ASEAN consists of 10 countries: Brunei Darussalam, Cambodia, Indonesia, Lao PDR, Malaysia, Myanmar, Philippines, Singapore, Thailand and Viet Nam. This research will omit Myanmar because some data are missing, and the estimation model requires balanced data.

Moreover, the data are treated as panel data since the Granger Causality test requires sufficient observations for each variable. Another consideration is the regional economic integration spirit of ASEAN as a single market, envisaged in the ASEAN Community Vision 2025.

This paper employs multivariate analysis using four variables: GDP per capita with purchasing power parity in international dollars as an economic growth measurement, symbolized by GDPPP; fixed broadband subscriptions per 100 population as a penetration ratio, expressed as BB100; gross capital accumulation percentage of GDP as a form of investment, symbolized by GROSCAP; and percentage of urban population to total population, expressed as URBAN. We employ a natural logarithm form of the variables in the calculation. Table 1 presents summary statistics of the variables in log form.

Table 1. Summary Statistics for the Variables

Variable	Observations	Mean	Std. Dev.	Min	Max
GDPPP	126	9.301858	1.207818	7.219645	11.38345
BB100	126	.0550483	2.397358	-7.717629	3.317656
URBAN	126	3.835072	.485379	2.941118	4.60517
GROSCAP	126	3.194852	.2526225	2.345396	3.678601

Notes: GDPPP is GDP per capita based purchasing power parity current international $; BB100 is fixed broadband subscriptions per 100 people; URBAN is urban population (% of total); and GROSCAP is gross fixed capital formation (% of GDP). Natural logarithms are used in our calculation method.

Panel Unit Root Test

The unit root test is a popular econometric method to analyse time series data. As research has advanced, the time series method has expanded to panel data computation with multiple observations and more prolonged periods, fuelling concern about spurious estimation from non-stationary data (Baltagi, 2013). In this study, we apply a three-panel unit root test: the Levine-Lin-Chu (LLC) test advised by Levin, Lin, and Chu (2002); the Augmented Dickey-Fuller (ADF)-Fisher chi-squared test recommended by Maddala and Wu (1999); and the Phillips Perron-Fisher chi-squared test proposed by Choi (2001). These three tests are recommended in many papers, presented in many econometric books, and included in current statistics software packages.

Panel Causality

A panel causality test was utilised to examine the causal relationship between two variables, because correlation is insufficient to test causality. Furthermore, a causality test can describe the causality direction, whether unidirectional, bidirectional or no causality. The causality calculation method was presented by Clive W. Granger (1969) in his paper titled "Investigating Causal Relations by Econometric Models and Cross-spectral Methods", known as the Granger causality test. Granger causality assumes the future cannot cause the past and, by employing a cross-sectoral model, a relationship between two or more variables in which one factor causes another can be validated (Granger, 1969). The simple causal model by Granger (1969) is:

$$X_t = \sum_{j=1}^{m} a_j X_{t-j} + \sum_{j=1}^{m} b_j Y_{t-j} + \varepsilon_t$$

and

$$Y_t = \sum_{j=1}^{m} c_j X_{t-j} + \sum_{j=1}^{m} d_j Y_{t-j} + \vartheta_t$$

http://doi.org/10.18080/ajtde.v6n3.151

where X_t and Y_t are two stationary variables. Y_t is causing X_t if some b_j is not zero, and, similarly, X_t is causing Y_t if some c_j is not zero. Furthermore, if both conditions occur, then a bidirectional causal relationship exists.

Our research adopts a model from Arvin and Pradhan (2014) to calculate the causal relationship between variables for the 13 years from 2003-2016. As we are focusing on broadband's effect on economic growth in connection to capital and urbanisation, the model is modified as follows:

$$GDPPP_{it} = \mu_{it} + \beta_{1i}BB100_{it} + \beta_{2i}GROSCAP_{it} + \beta_{3i}URBAN_{it} + \varepsilon_{it}$$

where $\mu_{it} = \eta_i + \nu_t$, η_i is an individual (country) effect, ν_t is a time effect, i represents each country in the panel, t refers to year, and ε_{it} is attributed to error, while β_1, $\beta_{2,}$ β_3 are coefficients.

Hypotheses

From the model, we develop two main hypotheses to describe the possibility of a causal relationship between two variables:

H_{01}: BB100 does not Granger-cause GDPPP.

H_{02}: GDPPP does not Granger-cause BB100.

GDPPP is GDP per capita based on purchasing power parity current international $, and BB100 is broadband subscriptions per 100 people.

Econometric Analysis and Empirical Results

Panel Unit Root Test

The Granger causality test requires data to be stationary to avoid spurious regression that is often affected by the existence of a trend in panel data. Table 2 presents the results of a three-unit root test for each variable.

Table 2. Results of Panel Unit Root Test

Statistic	GDPPP		BB100		URBAN		GROSCAP	
	L	1st	L	1st	L	1st	L	1st
LLC	-4.1069***	-4.8439***	-7.9199	-13.0199***	-3.9429***	-3.9280***	-1.3413**	-3.2328***
ADF	17.6577	42.7659***	83.1089***	106.7478***	46.5202***	47.6160***	15.8474	56.6247***
PP	23.1756	45.4471***	278.5666***	117.8568***	443.1846***	72.1020***	22.9062	161.3919***

Notes: L denotes level data, and 1st denotes first differenced data. LLC is the Levine-Lin-Chu test; ADF is Augmented Dickey-Fuller Fisher chi-squared test; and PP is the Philips-Peron Fisher chi-squared test. ***, ** indicate significance at the 1% and 5% levels, respectively, using lags=1.

The unit root test results show that some variables are not stationary in level data. We repeat the computation using first differenced data and achieve the result that all data is stationary at 1% level of significance.

Granger Causality Test

Following the unit root test, we continued with the Granger causality test to detect causal relationships between variables using the models below. The Granger causality test will be employed on first differenced data using lags automatically chosen by the Akaike Information Criterion (AIC) to find optimum lags, as suggested by Akaike (1974).

$$\Delta GDPPP_{it} = \eta_{1j} + \sum_{k=1}^{p} \beta_{1ik} \Delta GDPPP_{it-k} + \sum_{k=1}^{q} \lambda_{1ik} \Delta BB100_{it-k} + \sum_{k=1}^{r} \mu_{1ik} \Delta \mathrm{GROSCAP}_{it-k} + \sum_{k=1}^{s} \pi_{1ik} \Delta URBAN_{it-k} + \varepsilon_{1it} \quad (1)$$

$H_0: \lambda_{1ik} = 0;\ \mu_{1ik} = 0;\ \pi_{1ik} = 0$ for k=*1,2,…, p / q / r / s*

$H_A: \lambda_{1ik} \neq 0;\ \mu_{1ik} \neq 0;\ \pi_{1ik} \neq 0$ for at least one k

where *p, q, r* and *s* are lag lengths for the differenced variables from the equation and can be determined with a Granger causality approach. See, for example, Arvin and Pradhan (2014). Tables 3 and 4 show the results of the panel data Granger causality tests.

Table 3. Panel Granger Causality Test Results

	Independent Variables				
Dependent variables	**GDPPP**	**BB100**	**URBAN**	**GROSCAP**	**Causality Direction**
GDPPP	-	3.9015***	Undefined	3.5121***	BB100 => GDPPP GROSCAP => GDPPP
BB100	10.2105***	-	Undefined	1.3320	GDPPP => BB100
URBAN	Undefined	Undefined	-	Undefined	
GROSCAP	5.3451***	1.6901*	Undefined	-	GDPPP => GROSCAP BB100 => GROSCAP

Notes: The results for first differenced data variable are used here. The Granger non-causality equation (Dumitrescu & Hurlin, 2012) is used. * and *** indicate significance levels at 10% and 1%, respectively. Undefined results appear because the Granger non-causality test gave no output in those cases.

Six tests give undefined output for the URBAN variable, and after carefully checking the data, we found a problem in the urbanisation data for Singapore. Because it is a city-state, Singapore's population all live in the city, making the urban-to-population percentage 100% and thus producing zero value in the first differenced data for the URBAN variable. We omitted data for Singapore and repeated the Granger causality test, achieving the results in Table 4.

http://doi.org/10.18080/ajtde.v6n3.151

Table 4. Panel Granger Causality Test Results with Singapore Data Omitted

	Independent Variables				
Dependent variables	**GDPPP**	**BB100**	**URBAN**	**GROSCAP**	**Causality direction**
GDPPP	-	4.3379*	1.6145	3.1516***	BB100 => GDPPP GROSCAP => GDPPP
BB100	10.8753***	-	1.9873**	1.6436	GDPPP => BB100 URBAN=> BB100
URBAN	7.0669***	7.5635***	-	5.0846***	GDPPP=> URBAN BB100=> URBAN GROSCAP=> URBAN
GROSCAP	5.8368***	1.7195*	- 0.5672	-	GDPPP => GROSCAP BB100 => GROSCAP

Notes: The result for the first differenced data variable is used here without Singapore's data. The Granger non-causality calculation (Dumitrescu & Hurlin, 2012) is employed. *, ** and *** indicate significance levels at 10%, 5% and 1%, respectively.

Based on the calculation of Formula 1, we find a strong Granger causality link variable with the procedure by Dumitrescu and Hurlin (2012). The causality direction is summarised in the right column of Table 4 and described further in the flow diagram in Figure 3. The result presents a wide range of causal relationships between the variables.

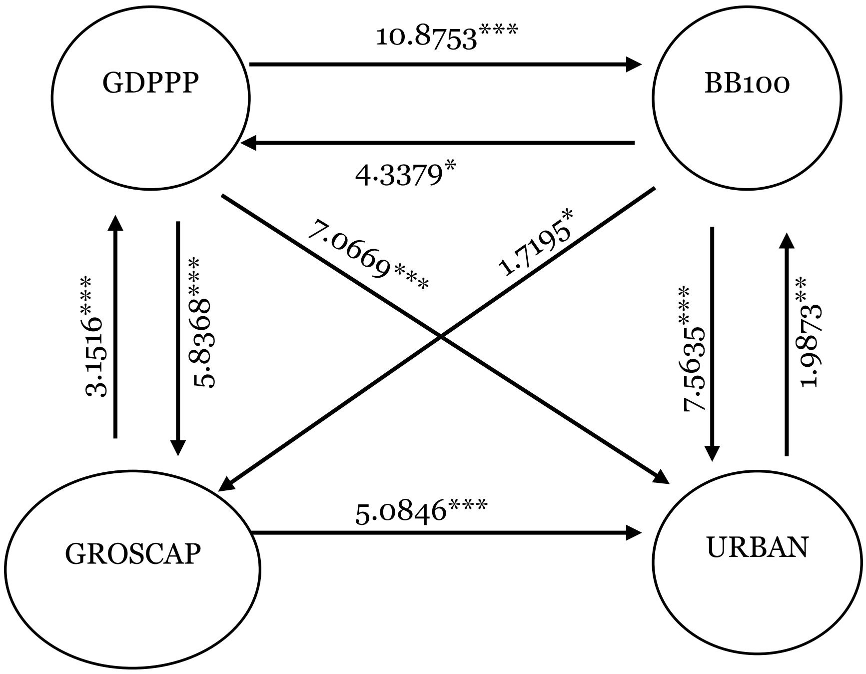


Figure 3. Granger Causality Directions Between Variables

Discussion

Although the result is clear, we intend to focus on the relationships between broadband and GDP growth. Viewing the hypothesis and Granger causality test result, we reject H_{01}: BB100 does not Granger-cause GDPPP and H_{02}: GDPPP does not Granger-cause BB100. We conclude there is a two-way causal link between broadband penetration and economic growth.

The results also suggest several issues in broadband penetration related to economic growth. First, the results show broadband penetration has a causal effect on GDP per capita growth, a view shared with previous studies (e.g. Bertschek *et al.*, 2015; Van Gaasbeck, 2008) that place broadband as an accelerator for economic growth in line with capital and labour. The impact of broadband on economic growth can be in terms of productivity, efficiency and broader market access, as well as new business formulation and employment opportunities. Broadband also affects gross capital accumulation, as business and personal activities benefit from broadband utilisation in terms of increasing productivity and efficiency (Czernich *et al.*, 2011; Falch, 2007; Van Gaasbeck, 2008), as well as decreasing capital requirements and promoting business expansion.

Second, we found the impact of GDP per capita growth on broadband penetration. The effect of GDP on broadband penetration has been presented in many studies (Bojnec *et al.*, 2012; Grosso, 2006; Lemstra, Voogt & Van Gorp, 2015; OECD, 2002, 2008), viewing GDP per capita as the primary supporter for broadband penetration. Higher income per capita will increase purchasing power and make broadband services more affordable, leading to increasing subscriber numbers. Moreover, the result also shows urbanisation has Granger causality for broadband penetration. Urbanisation will facilitate broadband penetration (Lemstra *et al.*, 2015; Wallsten, 2006) or demand (Koutroumpis, 2009), since urban people have higher education, employment status and technology experience, all of which support broadband adoption preferences (Hill, Troshani & Burgan, 2014).

While GDP per capita has a likely effect on broadband penetration, we found gross capital accumulation did not show a causal relationship with broadband penetration development. This can be understood when viewing the history of broadband development in the Southeast Asia region, where Digital Subscriber Line (DSL) became a central technology in the early stage of the broadband rollout (Ruddy & Ozdemir, 2013). DSL as a transitional technology required less capital compared to other technology, such as fibre-optic infrastructure, since DSL services use existing copper cable from plain old telephone service (POTS). In contrast to the operator benefit from utilising its traditional telephone network, the lower teledensity in ASEAN countries impacts broadband availability and limits penetration growth.

Furthermore, the increasing implementation of managed services in the telecommunication sector has decreased capital investment by telecommunication providers, as costs are reduced due to outsourcing efficiency and the transformation of capital expenditure (CAPEX) to operational expenditure (OPEX) (Esonwune, 2011). Managed services, as a consequence of increasing competition and falling telecommunication tariffs, have helped operators to deploy technology with less capital investment (as the assets are owned by the managed services vendor), along with limiting technological risk and the human resources required, as well as speeding deployment time.

Conclusion

Our results in some respects have shown the same view as existing literature, primarily on the impact of broadband penetration on economic growth. However, we also found a higher magnitude of effect from GDP on broadband penetration than from broadband penetration on GDP: this shows the importance of economic growth to promote broadband penetration in the ASEAN region, as affordability issues appear in some ASEAN member countries. Furthermore, broadband penetration benefited from urbanisation as well as economic growth, despite no causality from gross capital accumulation to broadband penetration.

We also affirm that variable and sample selection has a substantial effect on the results, when employing multivariate and panel causality tests.

Acknowledgements

The author gratefully appreciates the financial support from the Indonesian Endowment Fund for Education (LPDP), which gave him the opportunity to conduct this research. He also thanks Professor Hitoshi Mitomo and he acknowledges the useful discussion and helpful comments of his PhD colleagues.

References

Akaike, H. 1974. 'A new look at the statistical model identification'. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Arvin, BMM; Pradhan, RP. 2014. 'Broadband penetration and economic growth nexus: evidence from cross-country panel data'. *Applied Economics*, *46*(35), 4360–4369. https://doi.org/10.1080/00036846.2014.957444

Asean.org. 2008. Bali Declaration in Forging Partnership to Advance High Speed Connection to Bridge Digital Divide in ASEAN Bali, Indonesia, 29 August 2008. Retrieved 14 November 2017 from http://asean.org/?static_post=bali-declaration-in-forging-partnership-to-advance-high-speed-connection-to-bridge-digital-divide-in-asean-bali-indonesia-29-august-2008

Asean.org. 2009. Vientiane Declaration on Promoting the Realisation of Broadband Across ASEAN Vientiane, 15 October 2009. Retrieved 14 November 2017 from http://www.asean.org/wp-content/uploads/images/archive/TELMIN/Vientiane-Declaration-TELMIN-9.pdf

Asean.org. 2011. *ASEAN ICT Masterplan 2015*. Jakarta: ASEAN Secretariat. Retrieved from http://www.scribd.com/doc/111870071/ASEAN-ICT-Masterplan-2015

Asean.org. 2016. *ASEAN ICT Masterplan 2020*. Jakarta: ASEAN Secretariat. Retrieved from http://www.asean.org/storage/images/2015/November/ICT/15b -- AIM 2020_Publication_Final.pdf

Asean.org. 2017. Overview - ASEAN. Retrieved 14 November 2017 from http://asean.org/asean/about-asean/overview/

Baltagi, B. 2013. *Econometric Analysis of Panel Data*. Fifth Edition. John Wiley & Sons. Retrieved from https://books.google.co.jp/books?id=oQdx_70Xmy0C

Bertschek, I; Briglauer, W; Hüschelrath, K; Kauf, B; Niebel, T. 2015. 'The Economic Impacts of Broadband Internet: A survey'. *Review of Network Economics*, *14*(4), 201-227. https://doi.org/10.1515/rne-2016-0032

Bojnec, S; Fertő, I. 2012. 'Broadband availability and economic growth'. *Industrial Management & Data Systems*, *112*(9), 1292–1306. https://doi.org/10.1108/02635571211278938

Bouras, C; Diles, G; Kokkinos, V. 2013. 'Impact of broadband public infrastructures and services on SEE countries' economy'. *4th Global Information Infrastructure and Networking Symposium* (GIIS 2013), Trento, Italy. https://doi:10.1109/GIIS.2013.6684353

Cawley, A; Preston, P. 2007. 'Broadband and digital "content" in the EU-25: Recent trends and challenges'. *Telematics and Informatics*, *24*(4), 259–271. https://doi.org/10.1016/j.tele.2007.01.015

Choi, I. 2001. 'Unit root tests for panel data'. *Journal of International Money and Finance*, *20*(2), 249–272. https://doi.org/10.1016/S0261-5606(00)00048-6

Crandall, R; Lehr, W; Litan, R. 2007. 'The Effects of Broadband Deployment on Output and Employment: A Cross-sectional Analysis of U.S. Data'. *Issues in Economic Policy*, *6*, July 2007, 1–34.

Czernich, N; Falck, O; Kretschmer, T; Woessmann, L. 2011. 'Broadband Infrastructure and Economic Growth'. *The Economic Journal*, *121*, 505–532. https://doi.org/10.1111/j.1468-0297.2011.02420.x.

Dumitrescu, E.-I; Hurlin, C. 2012. 'Testing for Granger non-causality in heterogeneous panels'. *Economic Modelling*, *29*(4), 1450–1460. https://doi.org/10.1016/j.econmod.2012.02.014

Esonwune, C. 2011. *Telecommunications Managed Services Model in sub-Saharan Africa –*

The Pros and Cons : The Case of Nigeria. Blekinge Institue of Technology. Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:bth-5578%0A

Falch, M. 2007. 'Penetration of broadband services – The role of policies'. *Telematics and Informatics*, *24*(4), 246–258. https://doi.org/10.1016/j.tele.2007.01.008

Falch, M; Henten, A. 2010. 'Public private partnerships as a tool for stimulating investments in broadband'. *Telecommunications Policy*, *34*(9), 496–504. https://doi.org/10.1016/j.telpol.2010.07.010

Fanfalone, AG. 2015. *Post-2015 Development Agenda- Mexico Perspectives: ICT Infrastructure*. Retrieved from https://www.copenhagenconsensus.com/sites/default/files/mexico_ict_resouces.pdf

Granger, CWJ. 1969. 'Investigating Causal Relations by Econometric Models and Cross-spectral Methods'. *Econometrica*, *37*(3), 424-438. https://doi.org/10.2307/1912791

Grosso, M. 2006. 'Determinants of Broadband Penetration in OECD Nations'. *Communications*, 1–31.

Grubesic, TH. 2004. 'The geodemographic correlates of broadband access and availability in the United States'. *Telematics and Informatics*, *21*(4), 335–358. https://doi.org/10.1016/j.tele.2004.02.003

Hill, SR; Troshani, I; Burgan, B. 2014. 'Broadband adoption in regional and urban households'. *Journal of Computer Information Systems*, *54*(3), 57-66.

Holt, L; Jamison, M. 2009. 'Broadband and contributions to economic growth: Lessons from the US experience'. *Telecommunications Policy*, *33*(10–11), 575–581. https://doi.org/10.1016/j.telpol.2009.08.008

International Development Research Centre. 2008. *Digital Review of Asia Pacific 2007/2008*. Singapore: SAGE Publications. Retrieved from https://idl-bnc-idrc.dspacedirect.org/bitstream/handle/10625/34958/IDL-34958.pdf

Katz, RL; Vaterlaus, S; Zenhäusern, P; Suter, S. 2010. 'The impact of broadband on jobs and the German economy'. *Intereconomics*, *45*(1). https://doi.org/10.1007/s10272-010-0322-y

Kim, Y; Orazem, PFPF. 2017. 'Broadband Internet and new firm location decisions in rural areas'. *American Journal of Agricultural Economics*, *99*(1), 285–302. https://doi.org/10.1093/ajae/aaw082

Kolko, J. 2012. 'Broadband and local growth'. *Journal of Urban Economics*, *71*(1), 100–113. https://doi.org/10.1016/j.jue.2011.07.004

Koutroumpis, P. 2009. 'The economic impact of broadband on growth: A simultaneous approach'. *Telecommunications Policy*, *33*(9), 471–485. https://doi.org/10.1016/j.telpol.2009.07.004

LaRose, R; Bauer, JM; DeMaagd, K; Chew, HE; Ma, W; Jung, Y. 2014. 'Public broadband investment priorities in the United States: an analysis of the broadband technology opportunities program'. *Government Information Quarterly*, *31*(1), 53–64. https://doi.org/10.1016/j.giq.2012.11.004

Lau, TY; Kim, SW; Atkin, D. 2005. 'An examination of factors contributing to South Korea's global leadership in broadband adoption'. *Telematics and Informatics*, *22*(4), 349–359. https://doi.org/10.1016/j.tele.2004.11.004

Lemstra, W; Voogt, B; Van Gorp, N. 2015. 'Measuring broadband in Europe: Development of a market model and performance index using structural equations modelling'. *Telecommunications Policy*, *39*(3–4), 253–268. https://doi.org/10.1016/j.telpol.2015.01.006

Levin, A; Lin, C-F; Chu, C-SJ. 2002. 'Unit root tests in panel data: asymptotic and finite-sample properties'. *Journal of Econometrics*, *108*(1), 1–24. https://doi.org/10.1016/S0304-4076(01)00098-7

Maddala, GS; Wu, S. 1999. 'A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test'. *Oxford Bulletin of Economics and Statistics*, *61*(s1), 631–652. https://doi.org/10.1111/1468-0084.61.s1.13

Ng, TH; Lye, CT; Lim, Y. S. 2013. 'Broadband penetration and economic growth in ASEAN countries: a generalized method of moments approach'. *Applied Economics Letters*, *20*(9), 857–862. https://doi.org/10.1080/13504851.2012.754538

OECD. 2002. *The Development of Broadband Access in OECD Countries*. OECD Publishing. https://doi.org/10.1787/9789264034754-en

OECD. 2008. *Broadband Growth and Policies in OECD Countries*. OECD Publishing. https://doi.org/10.1787/9789264046764-en

Rampersad, G; Troshani, I. 2013. 'Innovation in the digital economy: The impact of high-speed broadband on innovating SMEs'. In *Proceedings of the 24th Australasian Conference on Information Systems*, 4-6December 2013, Melbourne.

Ruddy, M; Ozdemir, E. 2013. *An In-Depth Study of Broadband Infrastructure in the ASEAN Region*. Bangkok: UN ESCAP.

Tookey, A; Whalley, J; Howick, S. 2006. 'Broadband diffusion in remote and rural Scotland'. *Telecommunications Policy*, *30*(8–9), 481–495. https://doi.org/10.1016/j.telpol.2006.

06.001

United Nations. 2015. 'Information Economy Report 2015: Unlocking the Potential of E-commerce for Developing Countries'. In *United Nations Conference on Trade and Development*, 19, 1–28. Geneva. https://doi.org/10.1093/ijlit/eaq011

Van Gaasbeck, KA. 2008. 'A rising tide: Measuring the economic effects of broadband use across California'. *The Social Science Journal*, *45*(4), 691–699. https://doi.org/10.1016/j.soscij.2008.09.017

Wallsten, S. 2006. 'Broadband and Unbundling Regulations in OECD Countries'. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.906865

Framework for Demystifying M2M Spectrum Regulation

Anuradha C. Bhakthavatsalam
Senior Consultant, Ericsson India Global Services Pvt. Ltd, India

Arpit Khosla
Senior Consultant, Ericsson India Global Services Pvt. Ltd, India

Abstract: The evolving M2M landscape not only cuts across diverse verticals but also embraces a range of networks and devices. This diversity leads to varied and dynamic requirements, which make regulatory policy formulation a daunting challenge, especially the regulation of spectrum. Through this paper, we would like to share our perspective on the changed way in which a regulator should look at demand, supply, and utilization of spectrum. The proposed framework encapsulates various factors which should be considered by regulators. This study also illustrates how some of the proposed approaches (e.g. understanding spectrum demand) have been implemented by certain regulators. We also augment our view with specific data points from certain vertical industries. Finally, our study details the policy levers (e.g. spectrum fees, sharing regulations, License Authorization Model) which can help a regulator in reaching the desired policy posture.

Overall this framework attempts to demystify the spectrum policy regulation in the age of M2M/IoT. Additionally, our paper can also serve as a reference to new M2M/IOT players, who look forward to the information about the impact of spectrum regulation on their solution or their vertical.

Keywords: M2M, IoT, Regulator Policies, Spectrum, Framework

Introduction

Post the liberalization era of telecom services, there has been a clamour to deregulate the sector and take a true market-based approach towards telecom services. However, telecom services (in most cases) come within the category of essential public services and hence attract significant control and audit by government. Hence, telecom regulations tend to take a hybrid path between regulation and de-regulation. The key objectives of a regulator — promoting universal access, fostering competitive markets, preventing abuse of market power, and optimizing the use of scarce resources (e.g. radio spectrum, rights of way, identification numbers etc.) — continue to be the same with the advent of Machine-to-Machine (M2M) communications and the Internet of Things (IoT).

Implementation of such a hybrid (regulated free market) approach in an M2M-driven, fast-changing environment is expected to continue to keep the regulators on high alert. To be responsive with an effective, timely and optimal policy regulation, regulators will need to assiduously keep track of spectrum demand movements and utilization trends.

Until the recent past, voice, SMS and web browsing (using broadband) were the only applications that primarily influenced the traffic demand projections. Additionally, other factors such as the number of devices/subscriber combination profiles were also limited: hence demand forecasts were not the biggest of all challenges. However, with the increasing adoption of M2M/IoT, the traffic demands are expected to undergo a sea change. This directly implies that the approach and decisions on spectrum policy (including analysis of demand and supply) would need to adapt as well.

At this juncture, a comprehensive framework or a structured catalogue of factors, which are essential for M2M policy regulation, could immensely help regulators in demystifying the impact of this fast-changing environment on policy regulation.

The regulators would generally consider multiple dimensions while developing such a regulatory model. These dimensions typically consist of Licensing and Authorization of Spectrum, Roaming-related Regulations, numbering plan, BMAQ (Billing, Mobility, Authentication and Quality of Service – QoS), security and privacy postures. While each of these dimensions throws an intricate challenge to the regulator, we believe that the spectrum regulation is one of the most daunting and primary challenges of all.

The typical dilemmas that regulators face in the domain of spectrum can be boiled down to questions like the following. Is the unlicensed band sufficient to maintain scalability with QoS for various use cases? Do the demand and other requirements warrant a separate licensed band for M2M? As expected, there is no panacea for spectrum access regulation. This can be attributed to the fact that the technical requirements (for example, data throughput, reliability, range, and output power) vary dramatically across the gamut of use cases. To exacerbate the confusion in sizing up the varied and fast-changing technical requirements, a regulator is also expected to align its spectrum access policies to regulations and standards of other verticals. The advent of new cognitive radio, Dynamic Spectrum Access (DSA) and other technologies, definitely do not make the task of the regulator any simpler.

http://doi.org/10.18080/ajtde.v6n3.139

Here we present our perspective on the evolved approach (Framework of factors) that could help the regulator in deciphering demand, supply and utilization of spectrum in the age of M2M/IoT.

The Framework

Our proposed framework of factors, when considered, will help regulators look at this new world with the same old dimensions of supply, demand and utilization, but with an evolved lens. An in-depth study and analysis of the three dimensions would help in deciphering an optimal target M2M spectrum policy posture. We also plan to use selective sample analysis of some regulators to illustrate how some aspects of our proposed approaches can be applied. Additionally, in explanation of certain sections of the framework, we also plan to augment our view with specific data points from certain vertical industries.

The target spectrum posture in a specific spectrum band is defined in terms of the application of various policy levers in the specific band of consideration. Hence, the last part of our study plan shall delve into earmarking and detailing the policy levers (e.g. spectrum fees, sharing regulations, License Authorization Model) which can help the regulator in reaching the desired policy posture.

In the next sections, we elaborate how the analysis of spectrum, demand, supply, utilization and policy orchestration has evolved with the influence of M2M/IoT.

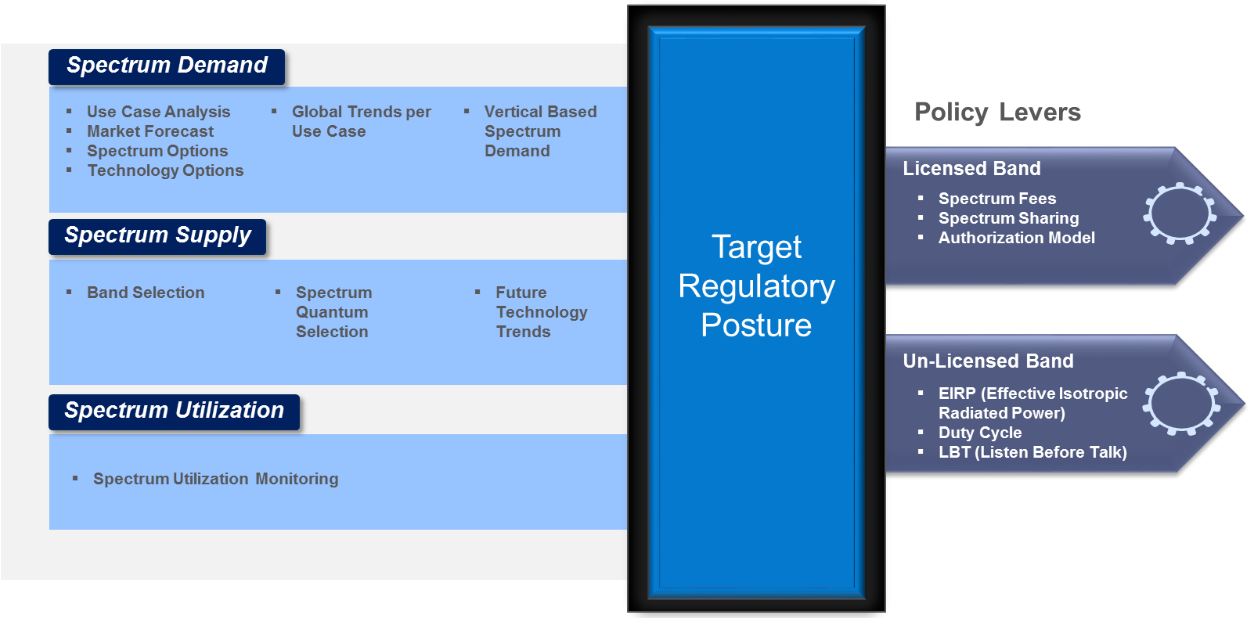


Figure 1. Framework for M2M Spectrum Regulation

Factors Shaping Target Spectrum Policy Posture

In this section, we offer a detailed explanation of the factors enumerated in the framework.

http://doi.org/10.18080/ajtde.v6n3.139

Spectrum Demand

This section is dedicated to detail and illustrate the new-age (influenced by M2M/IoT) factors that the typical regulator might look into for sizing and deconstructing spectrum demand and thereupon how the regulator might choose to classify this demand in terms of spectrum requirements. Local demand insights, the impact of global demand and the perspective of vertical specific demands are three principal areas of demand analysis. In the subsequent sections, we shall detail each one of these.

Local Demand Insights

Regulators would typically resolve the answers to the following four questions sequentially to deconstruct local market demand insights.

1. Describe and categorize use cases that are forecasted to grow in short/medium and long term.
2. For various use case categories, evaluate the options in terms of:
   a. Band of Operation;
   b. Available Access Technology.
3. Decipher implications of the above on spectrum demand.

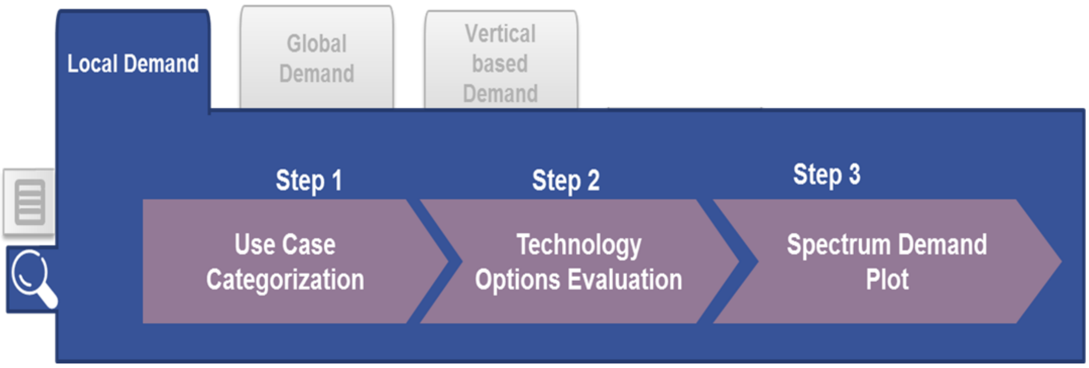


Figure 2. Local Demand Analysis

As a first step, the regulator must identify the use cases that are expected to grow within the geography of its influence: it is essential for the regulator to do an independent and in-depth analysis of each identified use case. Typically, the regulator should enlist all the parameters of identified use cases shown in Figure 3.

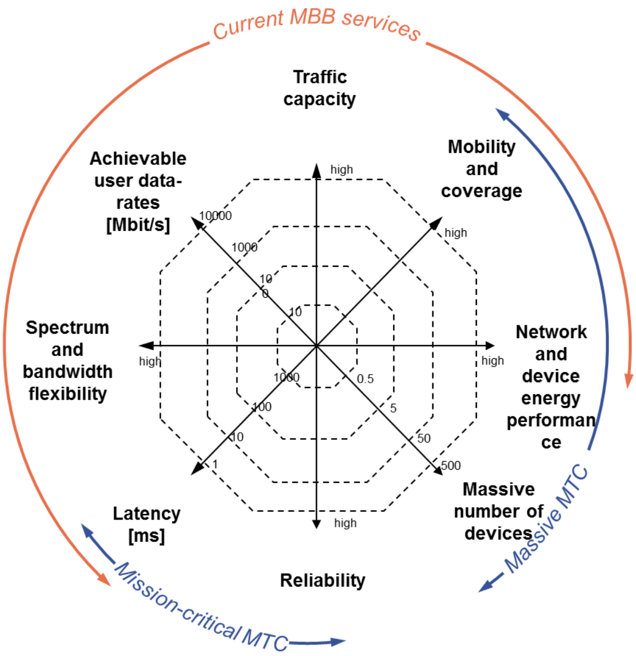


Figure 3. M2M Use Case Analysis Dimensions (Source: Murara, 2017, p. 15)

When a regulator delves into all 8 dimensions of use cases, they shall be able to categorize use cases and model the requirements. Below is an example of such a categorization exercise that was undertaken by Hattachi & Erfanian (2015).

Table 1. Example M2M Use Cases Categorization (Source: Hattachi & Erfanian, 2015, p.27)

Use Case Category	User Experience Data Rate	Data Latency	Mobility
Broadband access in dense areas	DL: 300 Mbps UL: 50 Mbps	10 ms	On demand, 0-100 km/h
Indoor ultra-high broadband access	DL: 1 Gbps UL: 500 Mbps	10 ms	Pedestrian
Broadband access in a crowd	DL: 25 Mbps UL: 50 Mbps	10 ms	Pedestrian
50+ Mbps everywhere	DL: 50 Mbps UL: 25 Mbps	10 ms	0-120 km/h
Ultra-low cost broadband access for low ARPU areas	DL: 10 Mbps UL: 10 Mbps	50 ms	On demand, 0-50 km/h
Mobile broadband (MBB) in vehicles (cars, trains)	DL: 50 Mbps UL: 25 Mbps	10 ms	On demand, up to 500 km/h
Airplanes connectivity	DL:15 Mbps per user UL: 7.5 Mbps per user	10 ms	Up to 1000 km/h
Massive low-cost/long-range/low-power MTC (Machine type Communications)	Low (typically 1-100 kbps)	Seconds to hours	On demand, 0-500 km/h
Broadband MTC	See the requirements for Broadband access in dense areas and 50+Mbps everywhere categories		
Ultra-low latency	DL: 50 Mbps UL: 25 Mbps	<1 ms	Pedestrian
Resilience and traffic surge	DL: 0.1-1 Mbps UL: 0.1-1 Mbps	Regular communication: not critical	0-120 km/h

Use Case Category	User Experience Data Rate	Data Latency	Mobility
Ultra-high reliability & Ultra-low latency	DL: From 50 kbps to 10 Mbps UL: From few bps to 10 Mbps	1 ms	On demand: 0-500 km/h
Ultra-high availability & reliability	DL: 10 Mbps UL: 10 Mbps	10 ms	On demand: 0-500 km/h
Broadcast like services	DL: Up to 200 Mbps UL: Modest (e.g. 500 kbps)	<100 ms	On demand: 0- 500 km/h

A regulator would also need to consider the topographic distribution of use cases in the above categories.

Subsequently, in step 2 (Technology Options Evaluation – see Figure 2), the regulator typically superimposes the use case requirements categories in step 1 onto the spectrum band and available technology options. For identifying the appropriate spectrum band, a regulator can come up with feasible options for each category of use cases by understanding the behaviour of these bands. Multiple studies have analyzed the factors to be considered for selecting a suitable operational band for a specific use case. Typically, these studies articulate that the selection of a carrier frequency band is a trade-off between radio-frequency (RF) propagation characteristics, noise floor, and terminal antenna size.

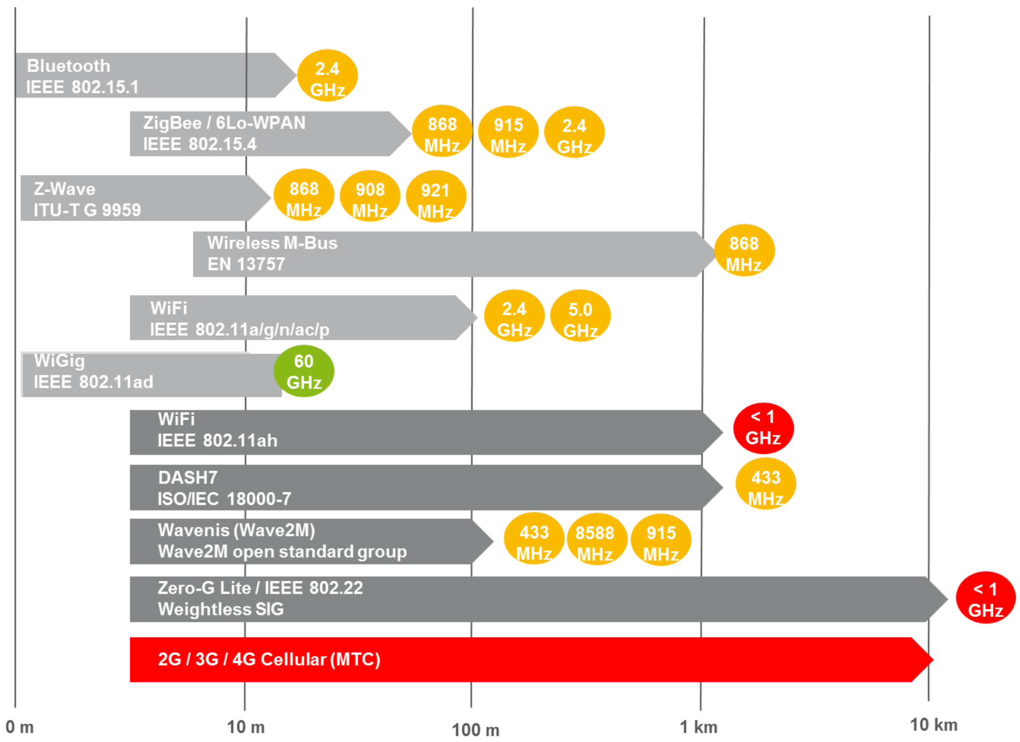


Figure 4. Example M2M Technologies and Spectrum Bands (Gupta, 2017, p. 14)

The next key step that the regulator must look at is the availability of technology options for those use cases in identified bands. We illustrate one such analysis (undertaken by the Institution of Engineering and Technology) in Figure 4.

Lastly as step 3, based on all the above factors, including all 8 dimensions of use cases, the preferable band of operation and technology availability, a regulator can come up with a local spectrum demand forecast. In order to do this, a regulator can classify requirements into categories like spectrum requirements for wide area coverage and short-range coverage. We illustrate this below with a similar thought process adopted by RSPG (2016b).

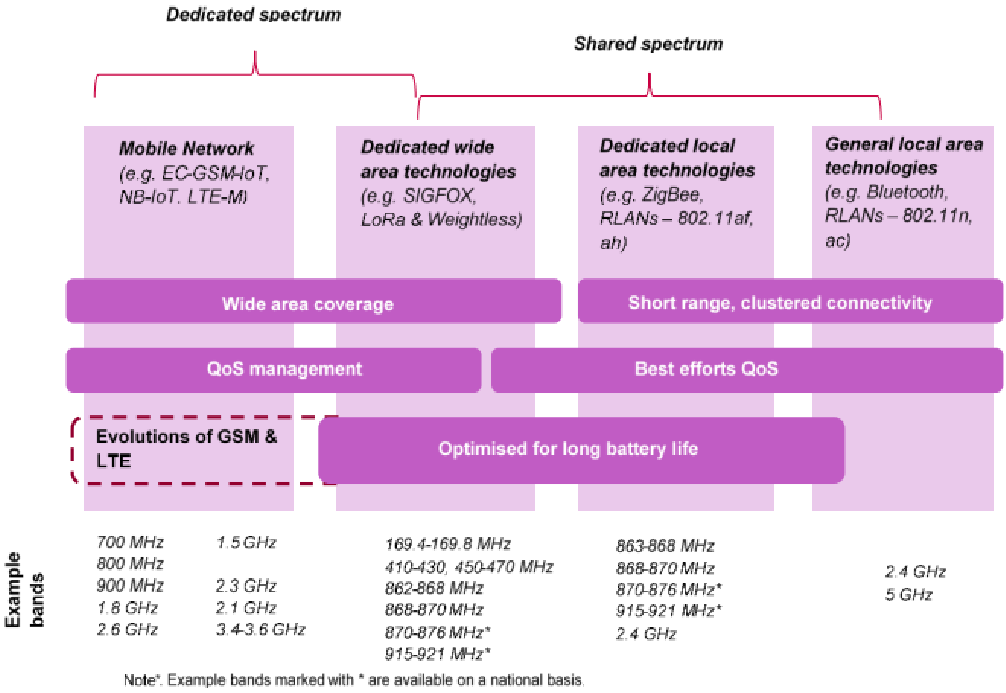


Figure 5. Example M2M Demand Plot (Source: RSPG, 2016b).

With such a classification, the regulator can ensure that it has considered demand in different bands and bandwidths for use cases that require dedicated as well as shared spectrum.

Global Demand Insights

As M2M is inherently a global business, over and above the local analysis it is imperative that regulatory policy caters to the impact of global M2M use cases. The policy must recognize as well as facilitate cross-border data flow, amongst many other requirements. A regulator must consider the scenarios wherein a significant number of roaming M2M SIMs or devices can be expected to enter the regulator's region of authority. For instance, if BMW cars come with pre-fitted global SIMs and are operating on LTE-M, then, while calculating demand, this factor must be accounted for as well. Conclusively, on top of local demand, for some sectors, the regulator might need to add a sliver of demand generated due to the global nature of the business.

Vertical-Based Point of View:

To date, regulators have never had to give specific attention to technical or regulatory requirements of different verticals. But, based on the scale that M2M is expected to take and the impact it is anticipated to have on all verticals, telecom regulators have to work not only with vertical stakeholders but also with vertical regulators to understand the requirements and forecast spectrum demand. Hence, joint consideration is a key underpinning of any equitable and just policy framework. As an illustration, we highlight the points of view of select verticals in the following sub-sections.

Utility Vertical

One of the biggest applications of M2M is smart metering. Smart meters are increasingly being deployed by utility companies to automate meter reading and provide opportunities to save energy, e.g. by providing real-time information on energy use. The deployment of smart meters is forecasted to connect 53 million meters by 2020 (Machina Research, 2014).

The existing and future communication requirements of utilities can be met in various ways. Differences in the grid layout, grid density, geographic grid coverage, amounts and types of renewable energy to be integrated, as well as varying demands in terms of resilience, security, latency, longevity, security, data rates, availability and criticality of communication, determine which communication solution is optimal in technical terms.

To meet these requirements, utilities need access to radio spectrum in a range of frequency bands. Utility-sector-based M2M solution providers are currently focusing on the 450-470 MHz band as a preferred band to meet current and future needs. In addition to this, in future, spectrum above 1 GHz (i.e. 1500 MHz range) may also be needed to support data-intensive applications. These bands offer an ideal compromise between coverage and the limited bandwidth requirements of the critically important utility applications. A harmonized spectrum allocation for utilities will facilitate synergies between utility companies, bring industrial benefits, facilitate cross-border coordination, increase the security of supply and lower energy costs to consumers.

To bring the perspective of a vertical based regulator, let us consider UTCC, Utilities Technology Council of Canada. UTCC supports access to additional spectrum by utilities and other critical infrastructure industries. UTCC has opined that it is critically important for public policies to support access to additional spectrum for utilities because everything in modern society depends on electricity, heat, and water (UTCC, 2017).

The key takeaways for the regulator from the above analysis could be summarized in the following 3 points:

http://doi.org/10.18080/ajtde.v6n3.139

a) How many of those 53 million meters are forecasted in the regulator's territory?
b) What is the license policy and utilization status of the preferred band (450-470 MHz)?
c) Is the technology ecosystem needed to support this thriving in our economy?

It would be prudent to give these issues their due importance when policy regulation is being formulated.

Automotive Vertical

In the automotive vertical, spectrum is essential in providing a host of valuable services, such as navigation services, concierge services, emergency calling and road-side assistance, door unlock, stolen vehicle tracking, crash notifications, and hands-free voice calling.

The automotive M2M sector comprises two main elements, namely factory-fitted "Vehicle Platforms", which support multiple applications, and aftermarket devices typically designed for a single function, such as stolen vehicle recovery or usage-based telematics.

It is anticipated that by 2022 80% or more of new vehicles will be connected with factory-fitted embedded connectivity, which is equivalent to 41% of all vehicles on the road in 2022 (Machina Research, 2014).

The after-market devices come in many different types, including usage-based insurance, stolen vehicle recovery, and navigation. This is driven by vehicle replacement rates and the adoption of usage-based insurance (UBI). From around 1% today, UBI is forecasted to become the norm by 2022 (Machina Research, 2014). These high-demand automotive applications will be significantly affected by constraints on the supply of spectrum. The automotive sector requires a large number of radio frequencies, as given in the below table.

Table 2. M2M Automotive Spectrum Requirements (Source: Bhattacharya, 2015)

Spectrum Band	Automotive Requirement
24.05-81 GHz	Blind-spot detection, lane departure prevention system, collision avoidance system, adaptive cruise control
22-29 GHz	Automatic start/stop technique
5.9 GHz	Car-to-car, car-to-infrastructure communication
1.602 GHz	Global navigation satellite system
1.575 GHz	Global positioning system
868.10 to 868.40 MHz	Short-range communication, keyless on/off

Spectrum Band	Automotive Requirement
433 to 434.79 MHz	Remote keyless entry, tyre pressure monitoring system (TPMS), immobilizers, keyless go
314 to 315 MHz	Remote keyless go, TPMS

Regulators see the need for interference-free access to spectrum in order for that technology to work as expected. We also notice steps that ITU is taking to enable a harmonized approach towards addressing such sector-specific requirements. For instance, ITU has recently allocated 79 GHz frequency spectrum for the operation of short-range high-resolution automotive radar (ITU, 2015). The allocation of the 79 GHz frequency band provides a globally harmonized regulatory frequency for automotive radar to prevent collisions, which will improve vehicular safety and reduce traffic accidents. Based on the criticality of the autonomous driving use case, it is expected that an automotive regulator might have its own regulations in place. Hence, over and above the three key takeaways that we mentioned in the utility sector, in Automotive it is imperative that telecom regulators take the perspective of automotive standards and regulations into account.

Overall, it is important that in this M2M-driven changed environment a telecom regulator looks at the above vertical-based market forecasts, technical requirements and vertical regulators' perspective for a policy that is optimally tuned to stimulate and harness the potential of all key verticals.

Over and above local demand, global demand and vertical-based points of view, there are two key stakeholders, whose interests are one of the primary drivers of telecom regulation – Mobile Network Operators (MNOs) and Mobile Virtual Network Operators (MVNOs).

MNO

Since AT&T is a force to reckon with in the MNO landscape, we choose to take AT&T's perspective on spectrum regulation in M2M/IoT as representative of MNOs. AT&T's documented response to a regulator's (TRAI) call for consultation reads:

> [T]here is no need for governments to allocate dedicated spectrum specifically for IoT or IoT segments. The government should continue efforts to find and reallocate spectrum for commercial mobile broadband use. If sufficient licensed spectrum is allocated for mobile broadband use, there is no reason to expect that dedicated spectrum to support IoT devices should be needed: it should be left up to Spectrum licensees to manage and employ their spectrum in an optimized fashion for the mix of traffic types that may be simultaneously using licensed bands. (AT&T, 2016)

In fact, not only AT&T but most telcos appear at odds with the sector regulator's call to delicense 1 MHz of spectrum at 867-868 MHz and another six MHz from the 915-935 MHz range for machine-to-machine (M2M) communication services. They also see no immediate need to delicense the 'V band' (in the range 57-64 GHz).

Overall it is evident that unlicensed spectrum and associated authorization policy might encourage new players to enter the M2M/IoT space, which telcos want to discourage. We expect global responses from the other telcos to be on the same lines as well.

MVNO/M2M Service Provider

We do not expect the MVNO point of view to be aligned with the MNO on this subject. Dedicated license-free spectrum along with an MVNO's core network would reduce the MVNO's dependence on an MNO for radio spectrum requirements. Though the forecasts suggest an exponential increase in the number of M2M/IoT devices and connections (the majority expected to operate in an unlicensed band), the spectrum allocation continues to be lopsided in favour of a licensed band. For instance, in most developed countries the licensed bands are to the tune of 400-700 MHz, whereas the unlicensed remains to be less than 26 MHz. Additionally, most of this 26 MHz is also shared with other ISM applications. Hence, we expect MVNO's to clamour for a dedicated unlicensed band.

Though it is important for regulators to be aware of the views of all stakeholders in terms of the spectrum bands, quantum of bandwidth and associated licensing model, the decision should be more driven by offering an actual use-case-based demand-driven analysis and conclusions.

Once a regulator has considered all the above factors of spectrum demand, despite the advent of M2M/IoT and changed ecosystem, we believe regulator would have not only an understanding of spectrum demand in each band but also the mapping of the demand to the vertical and use case driving that demand.

From a hypothetical instance perspective, which illustrates the entire spectrum demand analysis thought process, we believe a run-down of demand analysis activity might not be unlike the following description.

Illustration (Australia example): From an Australian perspective, multiple secondary research firms suggest that Smart Agriculture would be the sector with the most number of devices in Australia by 2026. To be precise, one study (Mackenzie, 2018) suggests nearly 16 million smart agricultural devices in Australia by 2026. Once the Australian agriculture regulator validates the plausibility of these forecasts and shares a typical expected geographic or demographic distribution, the telecom regulator's work begins. Though an Australian regulator would

http://doi.org/10.18080/ajtde.v6n3.139

formulate its own classification table, for our illustration sake, we assume that the regulator is following the classification suggested in Table 1. This would imply that Smart Agriculture can be a sub-category within the Massive low-cost/long-range/low-power MTC category. For smart agriculture we do not see a variety of difference in technical requirements from different use cases. However, had that not been the case, the regulator could have defined an additional sliver of subcategories within a Low-power MTC–Smart Agriculture category. A simple call for consultation would easily make the regulator aware of all the technologies (with proposed bands) available to address the use cases within this category. For illustration, let us assume the call for consultation returns LoRa (838 MHz) band (Semtech, 2018) and NBIoT (GSM bands) (Flynn, 2015) as the most appropriate technologies. The regulator, at this stage, might have its own spectrum demand picture like the one illustrated in Figure 5. Thereupon, as the next step, the regulator would need to build a model to forecast the carrier bandwidth required for addressing the 16 million prospective users with similar device and application profiles. Additionally, the regulator would take into account any global impact. (From the perspective of use cases that we are considering, for illustration, we do not foresee any global impact as such.) Lastly, the regulator would hold a round of discussion with the vertical regulator and stakeholders to cater to their perspective on the future in terms of technology and use cases (vertical specific). Hence, from a demand posture perspective, a regulator can have a clear goal for 2026; for example: "We need to provide for 16 million smart agricultural device types in the 868 MHz band". A similar thought process can be applied to all other relevant verticals.

Spectrum Supply

The two key aspects that regulators typically consider from the perspective of spectrum supply are: a) supply should be able to cater to demand forecasts; b) regulations should encourage technologies that have higher spectral efficiency. In the wake of M2M/IoT, it is essential to consider the above two factors in terms of both local and global perspectives for licensed and unlicensed bands. Overall, regulators must look at spectrum supply across bands and apply the above principles across the following three key steps.

Step 1: Band Selection

The objective of this step is to identify the spectrum bands that are relevant from the demand forecast perspective. To identify the appropriate bands, a regulator's own version of Figure 4 can be considered to be a good starting point. For each category of requirements defined in Table 1, the regulator needs to check the bands in which currently available technologies work. For instance, for a category "LPWA (low power wireless access) networks with no QoS requirements", LoRa is a potential candidate technology. LoRa can operate in both licensed and license-free ISM bands, like 335 MHz, 433 MHz, 868 MHz, and 915 MHz etc. An alternative technology Sigfox (n.d.) radio link uses unlicensed ISM radio bands as well, the frequencies of which are according to national regulation: in US, 915 MHz is used, while in Europe the 868 MHz band is used. By choosing any of the above options, the regulator can ensure that the spectrum supply policy enables one or a set of technologies to address the category of "LPWA networks with no QoS requirements". Just to illustrate the importance of the above perspective, we would like to take the example of an Indian regulator, which missed taking timely action on this front. In India, 865 MHz to 867 MHz was part of an unlicensed band until recently but 868 MHz was not part of this license-free band. It was only in 2017 that TRAI (Indian telecom regulator) recommended to include 868 MHz in this license-free bloc as well (TRAI, 2017). In such scenarios, wherein a regulator has to retrospectively reallocate an existing band for IoT, we anticipate two key impacts: firstly, it delays the adoption of technology in the country; secondly, there are chances of impact on existing users as well.

Step 2: Spectrum Quantum Selection

As a second step, a regulator must look at supply in selected bands from the perspective of the capability to support the prospective scale of demand. Each selected band in step 1 will typically service a particular set of device types. For spectrum supply determination, we believe the regulator can start by categorizing different regions based on device density expected. Then, based on application(s) being served, the regulator can discern peak and average throughput expected per device and optimal duty-cycle times for applications to run reliably. Thereupon, based on the technologies available in the band under consideration, regulators can estimate the supported spectral efficiency and typical coverage range of a cell. Eventually, this will help in the calculation of the number of cells per Km^2 and finally the amount of spectrum required. The significant difference between the M2M/IoT environment and the traditional environment is heterogeneity. The profile types of device/application combinations, available technologies and bands under consideration throw up a wide variety of possibilities. Additionally, variables such as some applications requiring indoor coverage, validity of current propagation models (such as the Hata Model (Hata, 1980)) and vulnerability to interference at higher order carrier bands make the whole estimation process quite complex. As licensed band technologies typically operate in a regulated environment, modelling interference and propagation might not be considered as the most challenging aspect of policy formulation. However, the regulator will have to handle the complexity rendered by a host of new technological paradigms, such as cloud-based radio access network (RAN), dynamic spectrum access, cognitive radio, models of spectrum sharing and possibilities via white spaces. Lastly, multiple technologies operating in the same band might make it difficult to predict the latency that the spectrum quantum can deliver.

Having said that, it is not debatable that the most challenging area in defining spectrum quantum will be the unlicensed spectrum. From a regulator's perspective, the M2M/IoT world has brought about a sea change in the world of unlicensed bands. The license-exempt bands with limitations on EIRP (Equivalent Isotropically Radiated Power), duty cycle etc., were defined based on fundamentals of short-range usage. However, for massive IoT, license-exempt bands are being used for long-range communications. Additionally, retrospectively, it was the downlink which was usually the constraint, but in massive IoT the uplink throughput and number of messages are more critical. These factors will directly impact the interference and performance issues; hence, they need to be modelled in unlicensed band quantum calculations. (IoTA, 2016, p. 10)

Lastly, while defining the spectrum quantum, regulators could also take leanings from other countries and their ratio of licensed to unlicensed spectrum. To illustrate the kind of learning that a regulator can draw from relative comparisons, we would take an example. Example: Australia has a concept of a class licence, which authorises open and shared access to segments of spectrum for designated services.

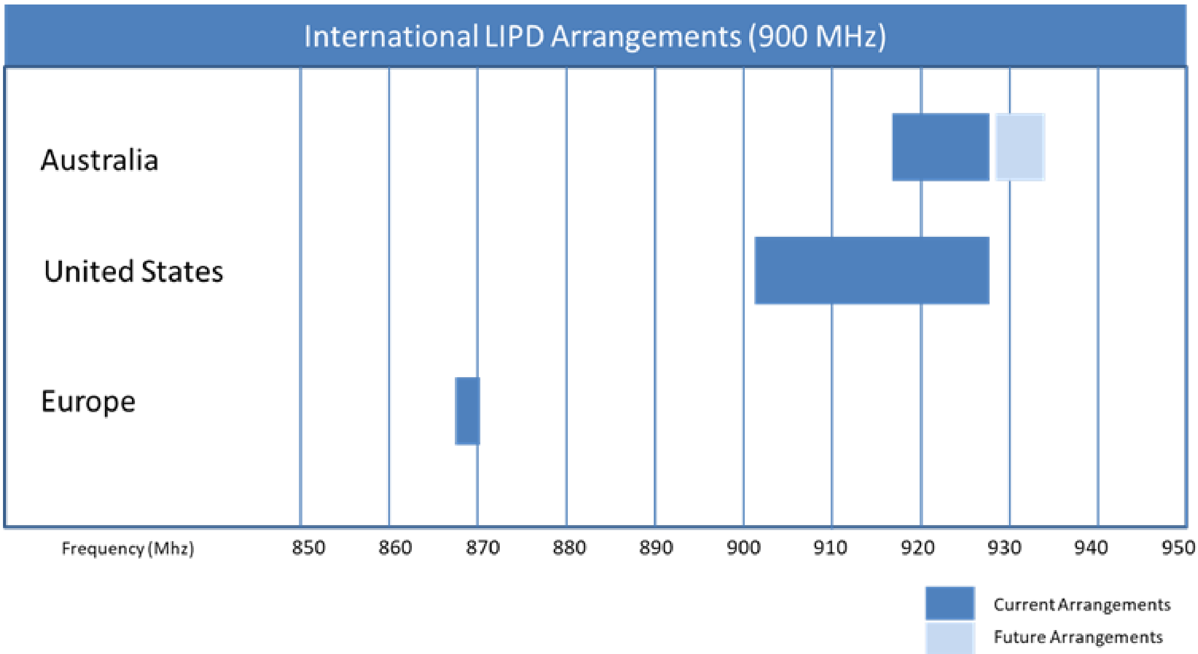


Figure 6. Comparison of Class Licenses (LIPD: Low Interference Potential Devices) in 800-950 MHz Bands (IoTA, 2016, p.4)

The above diagram illustrates the differences in class license allocation in the 800-950 MHz band across Australia, United States and Europe. It is evident that some IoT solutions in specific US bands (900-915 MHz) might not work out of the box in Australia. Another example could be the 868 MHz band, which is globally quite popular for LPWA deployments (with low duty cycle) but which is already allocated to land mobile services in Australia.

Similarly, a regulator can also compare the total quantum of unlicensed band to licensed band locally or compare unlicensed bands in local and global contexts. Such comparisons can help a regulator in selective and strategic harmonization of spectrum.

Step 3: Future Technology Trends

Lastly, it is also important to shape the current policy keeping expected short-term and long-term ecosystem changes in view. As far as some of the current trends are concerned, MNOs are typically using NBIoT, LTE-M and extended-coverage GSM technologies for massive IoT use cases. Some players are augmenting these technologies with other ISM band technologies, like LoRa and Sigfox. For critical IoT use cases, LTE and License Assisted Access seem to be the technologies of choice. A lot of operators are also choosing to switch off their 2G network, which may make available the 800 MHz and 1800 MHz bands.

As a future trend, we foresee a significant number of operators considering TV white spaces, spectrum sharing, cognitive radio and cloud RAN in the very near term.

Finally, on almost every operator's roadmap radar is 5G. Though the final recommendation on official spectrum bands of 5G will be formulated by the World Radiocommunication Conference in 2019, many telcos are launching pre-standard and pre-commercial 5G solutions. Regulators would need to keep a close watch on global 5G movements.

Each of the above future trends can have a profound impact on the spectrum supply posture of a regulator.

Overall, from a spectrum supply perspective, we believe a regulator's spectrum-supply-related decision-making model needs to evolve. In the wake of M2M/IoT-based changes in the ecosystem, if a regulator chooses to consider all factors mentioned in the above section, it will certainly help the regulator in arriving at an optimal spectrum supply policy

Spectrum Utilization

The third step of the analysis is to determine the present utilization of allocated spectrum in all the bands under consideration. The objective of a regulator is to balance the cost of interference against the spectrum utilization gains. The regulator achieves this objective by spectrum monitoring. The core aspects that are monitored include place, frequency and time (van der Vorst, Veldman & van Rees, 2016, p. 85). For a given interval of time, spectrum monitoring verifies if the earmarked spectrum for radio communication is being used appropriately and, as well, that the geographic location aligns with the applicable spectrum regulations. There are various instruments that can be used for monitoring. Typically, each monitoring node is equipped with high-quality software-defined radios (SDRs) that can be configured to collect data on specific frequency bands in specific time frames (Anker, 2010, p.

5). From a broad perspective, spectrum monitoring of diverse applications with different requirements is expected to be much more complex than the spectrum monitoring for technologies supporting traditional voice, SMS and broadband service only.

In licensed bands, the typical application of monitoring spectrum utilization is to ensure that stakeholders that acquire the licenses deploy legitimate bearer services with allowed technology variants, in the shortest time from the date of acquisition of licenses. The other application of monitoring the spectrum in a licensed band is so that the regulator can ensure that the quality of service experienced by users meets the defined benchmarks for urban and rural areas. The tools and practices for licensed band monitoring are also quite mature. The only change that we anticipate is in terms of spectrum monitoring in higher order frequencies (mm-wave) and for new RF modulations, which might be released as a part of 5G specifications. Conclusively maintaining effectiveness and efficiency in licensed band use cases might not be a challenge in general. However, for application-based monitoring, especially for use cases with requirements such as 1-10 ms latency, the regulator might have to adapt the incumbent tools and practices.

While the tools for monitoring spectral efficiency in licensed bands are available and known, unlicensed bands so far were primarily monitored for EIRP. In an M2M/IoT environment, we anticipate the following challenges in spectrum monitoring:

1. An unlicensed band might be carrying data for both short-range and long-range applications;
2. An unlicensed band might be carrying different long-range technologies in the same band.

Here, we would like to give an example of a change that might be required from the spectrum monitoring perspective (van der Vorst, Veldman & van Rees, 2016). Typically, the current monitoring network provides the data 24/7 at a 5 kHz and 1 minute resolution. However, if the actual monitoring of IoT networks is to be implemented, then the occupation of 125 kHz channels at a resolution of 50 ms (LoRa) and the occupation of 100 Hz channels at a time resolution of 1 second (Sigfox) should be made measurable (van der Vorst, Veldman & van Rees, 2016, p. 88).

Formulating the Spectrum Policy Posture

Based on all steps listed above, including a detailed demand analysis, supply analysis, utilization levels and considering the vertical stakeholder views, the regulator can choose its own spectrum target posture. This target posture basically signifies all the spectrum bands the regulator chooses to regulate heavily and the bands that are chosen to be handled with a light touch. But to actually regulate and reach the target posture, the regulator uses the levers which are to his disposal. In the next section, we identify and delve into these policy levers

Levers for Shaping the Spectrum Policy

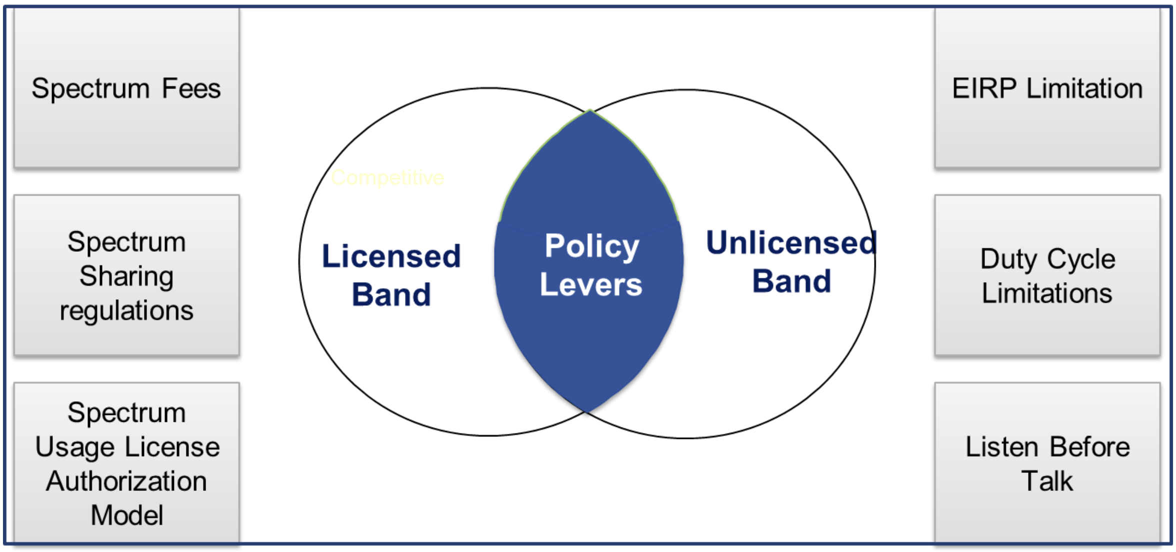


Figure 7. M2M Spectrum Regulatory Levers

A regulator has different levers to identify and regulate the licensed and unlicensed spectrum.

For licensed spectrum, a regulator looks at orchestrating:

- Spectrum fees
- Spectrum sharing regulations;
- Spectrum Usage License Authorization Model.

For unlicensed, a regulator typically looks at

- Technical specification based regulation.

Spectrum fees typically have two key components: spectrum acquisition fees and revenue-based spectrum usage charges.

Regulators have multiple options/approaches to value acquisition fees (Bakker, 2016). One of the methods is to make the acquisition fee equal to the cost of management of that spectrum. This method is, however, not tied to the value of spectrum used and hence may not encourage spectrum efficiency. Another method is to base the prices on market simulation-based shadow prices (based on willingness to pay). There are multiple detailed papers written on the pros and cons of auction-based spectrum acquisition pricing and the impact of a lack of an aftermarket on auction-based pricing, which is beyond the scope of this paper.

The second component of spectrum fees is revenue-based spectrum usage charges, which, though easy to administer, indirectly disincentivizes the efficient use of spectrum. Some experts in the field (Mazar & Azzarelli, 2016) have come up with formulas for such fee determinations, based on bandwidth allocated, position within RF band, allowed area for usage and more. However, this is just one way of calculating revenue-based fees; based on scenarios, many such permutations can be developed. Overall spectrum fees in specific bands should be determined by the usage in that band and with a vision to encourage free market-based pricing and efficient use of spectrum. Since this is a new evolving market, we urge regulators to initially take annual spectrum management fees (for massive IoT spectrum, if allocated separately in a licensed band) and subsequently, once the sector matures, the renewal can be based on willingness to pay. For critical IoT use cases, for instance, which are 5G based and expected to accrue attractive revenues to telcos, regulators can consider auction-based allocation, e.g. in higher-order mm wave bands near 60 GHz.

The next aspect is that of spectrum-sharing regulations, which until very recently encompassed only static spectrum-sharing regulations. However, the impending crunch in demand versus. supply is propelling ideas like dynamic spectrum sharing. With discussion of software-defined radio and cognitive radio, Dynamic Spectrum Access (DSA) is no longer only an academic theory. From a regulator's perspective, just like other unresolved questions, the debate about dynamic vs static spectrum management can be expected to be centre stage in future. We understand that no regulator can immediately adopt a dynamic spectrum management approach. However, from an implementation perspective, we agree with the paper by Anker (2010, p. 2), where it recommends a stepwise approach for implementing DSA. The stepwise approach encompasses introduction of cognitive radio within the current regime, creation of a spectrum commons and realization of a more fluid market for spectrum property rights. Interestingly, at the time of writing, there is no international regulation in place that prohibits the use of cognitive radio. However, any cognitive radio has to meet the interference, sharing and other conditions for the service that it is delivering in the respective band. Additionally, regulators will have a choice to exercise the interweave, underlay or overlay models for allowing DSA. We do not have a generic opinion as to which model is better. In fact, it is quite possible that different models are optimal in different circumstances (Song *et al.*, 2012).

Additionally, one other form that spectrum sharing acquires is spectrum leasing. Spectrum leasing can be of multiple types: for instance, a) where a licensee decides to allow other licensees to use its spectrum, but it retains the ownership; b) whereby owners of spectrum usage rights can sell or lease all or part of the rights associated with the license; and c) whereby owners of licenses can have their spectrum usage rights changed if they meet conditions defined by the regulator. Spectrum leasing is quite a subject on its own; hence we do not plan to delve deeper into it here (Srivastava, 2017).

Thirdly, a License or Authorization Model is by far one of the most powerful tools of a regulator. Using this privilege, a player can be authorized to use a specific span of channels in identified band:

- In a Specific Circle/Area or Pan-national;
- For a Service or Service agnostic;
- Via a particular technology or technology agnostic (liberalized).

Additionally, regulators can bind a licensee to deliver services not only in profitable areas, but also in rural circles/areas. From the perspective of geography, we anticipate most M2M licenses (if a separate licensed band is allocated) to be national and service agnostic. From the standpoint of technology, liberalization is still an open question. Technology Liberalization policy mostly refers to technology lockins within a particular spectrum band. For instance, in a liberalized 900 MHz band, a licensee would be free to choose any technology that fits the purpose. Considering the fact that M2M technologies are evolving, we believe that a liberal stand would be more reasonable. However, it is very important that, at this juncture, we bring out the contrary view that we encountered in the research. Some stakeholders believe that it is important that M2M bands are licensed and not technology neutral, as this would encourage ecosystem development around the technology of choice. Lastly, it would need assessment of local considerations, such as incumbent technologies. For instance, in Australia both LoRa and Sigfox networks have entered and hence it would be wise to take a liberal approach, whereas, in an underdeveloped economy, where there is less incentive for players to go in, a non-liberal environment might be more appropriate.

For the unlicensed band, typically to allow fair usage, regulators put a limit on EIRP (radiated power) and duty cycles, and mandate listen-before-talk kind of protocols. For massive M2M, when a technology is using an unlicensed band, these limitations allow a more efficient use of spectrum, and hence are very well suited. We do not expect critical IoT use cases to utilize unlicensed spectrum; hence, these regulations will typically not be applicable to critical IoT use cases.

Overall, regulators use the above levers to create an environment not only to allow fair and equitable access to the national resource of spectrum, but also to monitor and manage the players such that spectrum is efficiently and effectively used for the good of society at large.

Conclusion

The challenges that M2M services pose for a regulator cannot be understated. The diversity of verticals, use cases and stakeholders make spectrum regulation a very daunting challenge. The objective of this paper is to indicate major elements to focus on when formulating radio spectrum policy for new applications and services in the realm of M2M/IoT.

The traditional spectrum demand analysis was primarily dependent on the forecast of voice, SMS and broadband services. Via our analysis, we uncovered the perspective of looking at demand on a per-application basis. Our analysis implied that the same device profile can expect very different spectrum requirements based on the application being invoked. Our framework on demand also tried to bring to attention the importance and impact of global demand on local demand. Lastly, our research clearly propounds the relevance of collaboration between regulators across verticals for demand analysis, which was unheard of until the advent of M2M/IoT.

Our research on spectrum supply has pointed to a clear change of path in terms of both licensed and license-exempt bands. We highlighted how the complexity rendered by a host of new technological paradigms, such as cloud RAN, dynamic spectrum access, cognitive radio, models of spectrum sharing and possibilities via white spaces, is significantly changing the licensed band spectrum supply considerations. In the unlicensed bands arena, we strongly recommend the regulator consider the challenges of long-range technologies operating in license-exempt bands, which were essentially designed for short range. We urge the regulator to look at the impact of uplink messages in the unlicensed bands, which was never the case previously.

Based on our understanding, especially in the area of massive IoT, we have recommended the regulator apply annual spectrum charges in the initial years. Subsequently, once the sector matures, we have advised that fees based on willingness to pay might be a much more suitable approach for spectrum pricing.

Though we did not come across a significant change in the objectives of spectrum monitoring, we did highlight the elevated impact of QoS monitoring in critical IoT use cases and interference monitoring in license-exempt LPWA bands.

Last but not the least, our study points to the reasons that have brought look-before-talk protocols back into regulatory discussions.

Overall, our proposed framework offers not only a comprehensive list of design factors to consider for defining an optimal target M2M spectrum policy posture, but also indicates the policy levers available to achieve the same.

References

Aegis Systems. 2014. 'M2M application characteristics and their implications for spectrum'. Retrieved from https://www.ofcom.org.uk/__data/assets/pdf_file/0040/68989/m2m_finalreportapril2014.pdf

Anker, P. 2010. 'Cognitive Radio, the Market and the Regulator'. *IEEE Symposium on New Frontiers in Dynamic Spectrum (DySPAN)*, Singapore. Retrieved from: http://www.telecomabc.com/specials/dsa/resources/dyspan2010-cr-market-regulator.pdf

AT&T India. 2016. 'AT&T Comments on Telecom Regulatory Authority of India Consultant Paper'. Retrieved from http://trai.gov.in/sites/default/files/AT%26T_India_AGNSI_CP_18102016.pdf

Bakker, H. 2016. 'Spectrum and licensing in the mobile telecommunications market'. ITU Regional Workshop on "Competition in Telecommunications Market", Khartoum, Sudan. Retrieved from ITU website: https://www.itu.int/en/ITU-D/Regional-Presence/ArabStates/Documents/events/2016/CT/Final%20Documents/Session9/Spectrum%20and%20licensing.pdf

Bhattacharya, R. 2015. 'Auto industry asks govt. to de-license several spectrum bands for telematics tech'. *Business Standard*. Retrieved from http://www.business-standard.com/article/companies/auto-industry-asks-govt-to-de-license-several-spectrum-bands-for-telematics-tech-115021700826_1.html

Bhawan, MD; Marg, JLN. 2017. 'Recommendations on "Spectrum, Roaming and QoS related requirements in Machine-to-Machine (M2M) Communications"'. Retrieved from Telecom Regulatory Authority of India website: http://www.trai.gov.in/sites/default/files/Recommendations_M2M_05092017.pdf

BMVIT [*Bundesministerium für Verkehr, Innovation und Technologie*, Austria]. 2012. 'Wireless Local Area Networks (WAS, WLAN, RLAN)'. Information of the Austrian Telecommunications Authority. Retrieved from https://www.bmvit.gv.at/ofb/publikationen/infoblaetter/downloads/201005en.pdf

Chang, H; Chen, K. 2011. 'Cooperative spectrum sharing economy for heterogeneous wireless networks'. In *GLOBECOM Workshops (GC Wkshps), 2011 IEEE*. Retrieved from http://ieeexplore.ieee.org/abstract/document/6162491/

FCC [Federal Communications Commission, US]. 2015. 'Comments Sought on M2m Spectrum Networks'. *Public Notice*. Retrieved from FCC website: https://www.fcc.gov/document/comments-sought-m2m-spectrum-networks

Flynn, K. 2015. NarrowBand IOT. Retrieved from http://www.3gpp.org/news-events/3gpp-news/1733-nio

GSMA. 2016. '3GPP Low Power Wide Area Technologies'. *GSMA White Paper*. Retrieved from GSMA website: https://www.gsma.com/IoT/wp-content/uploads/2016/10/3GPP-Low-Power-Wide-Area-Technologies-GSMA-White-Paper.pdf

Gupta, M. 2017. *Indian Telecom Industry getting ready for M2M/IoT*. Retrieved from Institution of Engineering and Technology website: http://www.broadbandindiaforum.com/pdf/New-Technologies/Indian-telecom-getting-ready-for-M2M-IoT_%20whitepaper_final_OCT11.pdf

Hata, M. 1980. 'Empirical Formula for Propagation Loss in Land Mobile Radio Services'. *IEEE Transactions on Vehicular Technology*, VT-29 (3) pp. 317–325. See also https://en.wikipedia.org/wiki/Hata_model

Hattachi, R; Erfanian, J. 2015. *NGMN 5G White Paper*. Retrieved from NGMN Alliance website: https://www.ngmn.org/fileadmin/ngmn/content/downloads/Technical/2015/NGMN_5G_White_Paper_V1_0.pdf

Intven, H. 2000. *Telecommunication Regulation Handbook Module 5 Competition*. Retrieved from https://www.itu.int/ITU-D/treg/Documentation/Infodev_handbook/5_Competition.pdf

IoTA [IoT Alliance]. 2016. 'Spectrum available for IoT'. Retrieved from http://www.IoT.org.au/wp/wp-content/uploads/2016/12/IoTSpectrumFactSheet.pdf

ITU. 2015. 'ITU World Radio communication Conference approves spectrum for automotive radar'. *Press release*. Retrieved from http://www.itu.int/net/pressoffice/press_releases/2015/52.aspx#.WuDURy5ubIV

Kim, G. 2015. 'What is spectrum sharing, and why does it matter?' *Spectrum Futures*. Retrieved from https://spectrumfutures.org/what-is-spectrum-sharing-and-why-does-it-matter/

Loy, M; Karingattil, R; Williams, L (eds). 2005. 'ISM-Band and Short Range Device Regulatory Compliance Overview'. Retrieved from http://www.ti.com/lit/an/swra048/swra048.pdf

Machina Research. 2014. 'M2M communications for policy makers'. Retrieved from http://www.iicom.org/images/iic/events/annual-conference/2014/Workshop_Presentations/m2m2014-Matt-Hatton-Machina-Research.pdf

Mazar, MH; Azzarelli, T. 2016. 'Spectrum Fees'. In *Radio Spectrum Management: Policies, Regulations and Techniques*. Somerset: Wiley.

Mackenzie, M. 2018. 'Australia: IoT market report 2018'. *Analysys Mason*. Retrieved from http://www.analysysmason.com/Research/Content/Country-reports/australia-IoT-report-rdme0

McBride, S. 2017. 'Regulating Permanent Roaming for M2M and IoT Devices'. *Ovum*. Retrieved from https://ovum.informa.com/resources/product-content/te0007-001165

Murara, B. 2017. *IMT-2020 Network High Level Requirements, How African Countries can cope*. Retrieved from ITU website: https://www.itu.int/en/ITU-T/Workshops-and-Seminars/standardization/20170402/Documents/S2_4.%20Presentation_IMT%202020%20Requirements-how%20developing%20countries%20can%20cope.pdf

NTIA [National Telecommunications and Information Administration]. n.d. 'US Spectrum Management Policy: An Agenda for the Future'. *NTIA Special Publication* 91-23. Retrieved from https://books.google.co.in/books?id=3gHKZZ7Alh0C&printsec=frontcover#v=onepage&q&f=true

Parbat, K. 2017. 'Telcos oppose Trai view of delicensing airwaves for M2M'. *The Economic Times*. Retrieved from https://economictimes.indiatimes.com/news/economy/policy/telcos-oppose-trai-view-of-delicensing-airwaves-for-m2m/articleshow/60400046.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst

Pillai, R; Kumar, A; Thukral, H. 2015. *M2M enablement in Power Sector*. TEC-TR-S&D-M2M-002-02. Retrieved from Telecommunication Engineering Centre website: http://tec.gov.in/pdf/M2M/Spectrum%20requirements%20%20for%20PLC%20and%20Low%20power%20RF%20communications.pdf

RSPG [Radio Spectrum Policy Group]. 2016a. '[Draft] Opinion for public consultation on A Spectrum Roadmap for IoT: Draft Opinion on the Spectrum Aspects of the Internet-of-things (IoT) including M2M'. European Commission. Retrieved from https://circabc.europa.eu/sd/a/50d58028-eb1d-487f-87d5-5efb9c76b140/RSPG16-045-Draft_Opinion_IoT.pdf

RSPG [Radio Spectrum Policy Group]. 2016b. 'A Spectrum Roadmap for IoT: Opinion on the Spectrum Aspects of the Internet-of-things (IoT) including M2M'. European Commission. Retrieved from https://circabc.europa.eu/sd/a/a0faa1a5-ca41-42c3-83d5-561b197419b0/RSPG17-006-Final_IoT_Opinion.pdf

Semtech. 2018. 'What is LoRa®?' Retrieved from https://www.semtech.com/lora/what-is-lora

Sigfox. n.d. 'Sigfox Technology Overview'. Retrieved from https://www.sigfox.com/en/sigfox-iot-technology-overview

Song, M; Xin, C; Zhao, Y; Cheng, X. 2012. 'Dynamic spectrum access: from cognitive radio to network radio'. *IEEE Wireless Communications*, 19 (1), February. Retrieved from https://ieeexplore.ieee.org/document/6155873

Srivastava, UK. 2017. 'Liberalised Spectrum'. *ITU-TRAI International Training Program*, New Delhi. Retrieved from https://trai.gov.in/sites/default/files/presentations_&_cv/Day-1_23Aug2017/Session-2_Conv%20Lic%20Regime/Spectrum%20Liberalisation-%20UK%20Srivastava.pdf

TRAI [Telecom Regulatory Authority of India]. 2017. 'TRAI releases recommendations on "Spectrum, Roaming and QoS related requirements in Machine-to-Machine (M2M) Communications"'. Press Release No. 71/2017. Retrieved from TRAI website: http://trai.gov.in/sites/default/files/PR_No_71of2017.pdf

TSDM [Telecoms Systems Design & Manufacturers Association, India]. 2017. 'Comments on TRAI Consultation Paper on "Spectrum, Roaming and QoS Related Requirements in M2M communication"'. Retrieved from Telecom System Design and Manufacturer's Association website: http://www.trai.gov.in/sites/default/files/TSDM%20M2M%20CP.pdf

UTCC [Utilities Technology Council of Canada]. 2017. 'Why Do Utilities Need Access To Spectrum', October. Retrieved from https://utc.org/wp-content/uploads/2018/02/Why_Do_Utilities_Need_Access_To_SpectrumOCT-2017FINAL.pdf

Van der Vorst, T; Veldman, J; van Rees, J. 2016. 'The wireless Internet of Things: Spectrum utilisation and monitoring'. Retrieved from http://www.IoTjournaal.nl/wp-content/uploads/2017/02/dialogic_-_internet_of_things_spectrum_utilisation_and_monitoring_v2.3.pdf

Bitmaps and Bitmasks: Efficient Tools to Compress Deterministic Automata

Subramanian Shiva Shankar, Lin PinXing
Intel Technology Asia Pte Ltd

Andreas Herkersdorf, Thomas Wild
Technical University of Munich

Abstract: Accelerating the signature matching function is essential to perform Deep Packet Inspection (DPI) at line rates. The conversion of the signatures into the Deterministic Finite Automaton (DFA) enables performance of this function at linear time. However, since the DFA is extremely storage inefficient, it is compressed before being stored in the memory. Although state-of-the-art bitmap-based compression algorithms can perform line rate signature matching, they only achieve transition compression of ~90-95%. Addressing the storage inefficiency, two bitmap-based transition compression algorithms were proposed by Subramanian *et al.* in 2016 to achieve transition compression of over 98%. A theoretical relationship is established in this article between the achievable signature matching throughput and the number of pipeline stages required to perform the decompression through the hardware accelerator based on the proposed techniques. Additional optimizations are proposed and evaluated to improve the per-stream signature matching throughput through the proposed decompression engines. The experimental evaluation of the optimizations shows that the per-stream signature matching throughput can be improved by a factor of 1.2–1.4x. A software model of the proposed decompression engines was designed and evaluated across a multitude of payload byte streams to verify the functional correctness of the proposed compression methods.

Keywords: Deep Packet Inspection; Network Security; Regular Expressions; Finite Automata; Transition Compression;

1. Introduction

The Residential Gateway Router (RGR) has traditionally been used to connect the Local Area Network (LAN) in the home with the Wide Area Network (WAN) of the service providers. Services such as Internet Protocol Television (IPTV), media sharing, home security and other value added services are driving the RGR from its traditional function of simple packet

http://doi.org/10.18080/ajtde.v6n3.125

forwarding applications. The home network has gained prominence in the networking ecosystem due to the emergence of the smart home and an increased presence of consumer Internet of Things (IoT) devices in the home network. Moreover, researchers have found various security vulnerabilities in the RGR and have identified that it is easy prey for network attacks (Holcomb, 2016; Team Cymru Threat Intelligence Group, 2016). Value added services, such as content-based Quality of Service (QoS), and security applications, such as intrusion prevention, drive the need for Content Aware Networking (CAN) in the home network. Moreover, improving broadband speeds (Sappington, 2016) necessitate the need for multi-gigabit line-rate packet processing, not only for packet forwarding applications in the RGR but also for content aware networking applications.

Deep Packet Inspection (DPI) is the process of inspecting the network packet payloads with a predefined database of signatures to perform CAN. DPI has various functions, which include network packet normalization (Handley, Paxson & Kreibich, 2001), flow based packet reordering, signature subset allocation based on header analysis (Song *et al.*, 2005), signature matching and packet post processing (Xu, 2016). However, signature matching is the most time critical function in DPI and defines the rate at which DPI can be performed in modern network processors (Nourani & Katta, 2007). Considering the processing and memory capabilities of the RGR, an analysis in Subramanian, Lin & Herkersdorf (2014) identified that hardware acceleration of signature matching is essential to support line-rate DPI in the RGR.

The signatures which are used for DPI applications comprise strings and regular expressions and are used to identify the patterns in the network traffic. Since automata-based methodologies support both string and regular expression signatures, the signature set is converted into an automaton (state table) with which the network payload is compared (Becchi & Crowley, 2007). The signatures can either be converted into a Non-Deterministic Finite Automaton (NFA) or a Deterministic Finite Automaton (DFA) to perform the signature matching against the network traffic. The NFA is time inefficient and space efficient, while the DFA is time efficient and space inefficient (Yu *et al.*, 2006). DFA-based architectures are preferred over NFA-based architectures, as they allow the signature matching to be done in linear time (Becchi & Crowley, 2007). A set of signatures can be converted into a DFA based on the algorithms described in John, Rajeev & Ullman (2007).

The DFA represents the signatures in the form of a finite state machine with states and directed edges between the states to represent the state transitions (Xu, 2016). Since the internet traffic is primarily constructed over the extended ASCII character set, each state in the DFA has 256 outgoing transitions, while the total number of states in the DFA depends on the characteristics of the signatures which it represents (Becchi & Crowley, 2007b). Various compression algorithms have been proposed in the literature to address the storage inefficiency associated with the DFA and can be further classified into transition and state compression algorithms (Xu, 2016). The transition compression algorithms focus on compressing the redundant state transitions in the DFA, while the state compression algorithms focus on reducing the number of states in the DFA. However, the transition and the state compression algorithms are orthogonal to each other (Xu, 2016).

Certain regular expressions contain terms such as the Kleene closure operator which, when it interacts with other signatures, results in an exponential growth in the number of states generated in the DFA. This is referred to as the state explosion problem (Becchi & Crowley, 2007b). Many modified automata such as the Hybrid Finite Automaton (HFA) (Becchi & Crowley, 2007b), eXtended Finite Automaton (XFA) (Smith *et al.*, 2008), Tunable Finite Automaton (TFA) (Yang *et al.*, 2014), Jump Finite Automaton (JFA) (Yu, Lin & Becchi, 2014) and DFA with Extended Character Set (DFA/EC) (Cong *et al.*, 2014) have been proposed in the literature to counter the state explosion problem.

The transition compression algorithms proposed in the literature can be classified into hardware-oriented and software-oriented algorithms (Subramanian *et al.*, 2016). The software oriented transition compression techniques such as the D²FA (Kumar *et al.*, 2006), A-DFA (Becchi & Crowley, 2013), δFA (Ficara *et al.*, 2008) and RCDFA (Rafael *et al.*, 2015) compress the redundant state transitions in the DFA at the cost of additional memory bandwidth to fetch the compressed state transition (Xu, 2016). Moreover, multiple terabits of on-chip memory bandwidth will be required, when the signature matching has to be performed at line rates after performing the compression through these techniques (Qi *et al.*, 2011). Moreover, designing memories with such huge memory bandwidth is not practically achievable. To address the problem of efficient transition compression and performing signature matching at line rates, various bitmap-based transition compression techniques (Qi *et al.*, 2011; Wang *et al.*, 2011) have been proposed in the literature. However, the usage of bitmaps to compress the redundant state transitions also requires the bitmap to be stored in the memory along with the compressed state transitions. So, the existing bitmap-based methods compromise on the transition compression rates and only achieve compression rates of the order of 90-95%, primarily to minimize the number of bitmaps stored in the memory. However, this approach increases the memory footprint of the compressed DFA and results in inefficient usage of on-chip memories (Subramanian *et al.*, 2016). Addressing this problem, the Member State Bitmask Technique and the Leader State Compression Technique were proposed, in which an additional layer of indexing called bitmask was introduced to improve the achievable transition compression rates to over 98% (Subramanian *et al.*, 2016). The improved transition compression resulted in a 50% reduction in the memory footprint of the compressed DFA in comparison to the existing methods. Moreover, in addition to the transition compression methods, Subramanian *et al.* (2016) also proposed a corresponding hardware accelerator to perform line-rate signature matching. Extending the idea proposed in Subramanian *et al.* (2016), this paper further discusses the structural hardware building blocks which are required to perform the transition decompression at line rates through a hardware accelerator. A software model of the transition decompression engines was designed to verify the compression algorithms and was evaluated with synthetic traffic workloads spanning across different levels of maliciousness. The signature matching results extracted from the software model were identical to that of the DFA-based signature matching engine, further validating the correctness of the compression methods. Moreover, this paper discusses and evaluates additional optimization steps in the transition decompression process to further improve the per-stream signature matching throughput that can be achieved through the hardware accelerator. Experimental evaluations further show that the proposed optimizations result in 1.25–1.4x improvement in the per-stream signature matching throughput in comparison to the per-stream throughput without the optimizations.

2. Prior Art and Key Contributions

2.1 Introduction to DFA

If S' represents the set of 'N' DFA states and C' represents the set of characters, each entry in the DFA stores the state transition $\delta(s, c) = t$, where $s, t \in S'$ represents the current state and the next state, respectively; $c \in C'$ represents the character; and δ represents the state transition function. The total number of state transitions generated in the DFA is a product of the total number of states generated and the number of characters for which the transitions are represented in a state as shown in (1).

$$\#\ \text{Total State Transitions} = \mathrm{N}\ \times \mathrm{C}' \quad (1)$$

If the DFA is directly stored in the memory, the amount of memory required to store the DFA corresponds to the memory required to store all the state transitions as shown in (1). However, only a very small portion of the transitions in a DFA lead to a successful signature, while the majority of others do not. These transitions which do not lead to a signature match are called the failure transitions. Many of these failure transitions are redundant and compressing these transitions enables efficient storage of the DFA. As mentioned previously, the transition compression techniques which are used to compress these redundant transitions can be broadly classified into software- and hardware-oriented techniques. Though the aim of both these techniques is to compress the redundant transitions in a DFA, the algorithmic approach towards transition compression in hardware-oriented techniques enables the decompression to be performed in a dedicated hardware accelerator, which is not possible in the case of software-oriented solutions (Qi *et al.*, 2011).

2.2. Software-Oriented Transition Compression Techniques

Kumar *et al.* (2006) identified equivalent transitions between states in a DFA, which are compressed by introducing a default transition to create a D^2FA. However, the default transition paths can be very long, which in turn results in multiple memory lookups to fetch a compressed transition. To avoid the long default path, Becchi and Crowley (2007a, 2013), proposed the Amortized time-bandwidth overhead DFA (A-DFA) with directional default path selection. This approach reduced the length of the default paths, but also reduced the transition compression achieved in certain cases. However, in both D^2FA and A-DFA, the compressed transitions in a state have to be sequentially searched, which further leads to additional memory lookups. Analysis in Qi *et al.* (2011) identified that the average number of transitions traversed to identify the next state is about 100 per input character using the D^2FA and A-DFA compression algorithms. To avoid this problem, δFA (Ficara *et al.*, 2008) was proposed, which only stored those transitions in a state that were different from its parent state. δFA also defined how to identify the parent state for a state whose transitions are compressed. Transitions corresponding to the character set are periodically updated in a local cache as the states are traversed through. Even though the local cache memory is used to store the transitions, the cache has to be frequently updated and the frequency depends on whether a transition is fetched from the parent state or not. The average transition compression rates achieved by D^2FA, A-DFA and δFA are typically of the order of about 90% (Xu, 2016). Rafael *et al.* (2015) identified that a large set of transitions in a state are directed to a same next state and can be compressed by representing them as a ranged transition representation and defined the Ranged Compressed Deterministic Finite Automata (RCDFA). The RCDFA achieved transition compression rates of the order of 97% and, in turn, resulted in fewer transitions stored per state in comparison to the other proposals discussed above. The signature matching throughput that could be achieved by the various methods described above was compared in Rafael *et al.* (2015)[i]. The signature matching throughput achieved when the payload was compared against the uncompressed DFA representation was used as the benchmark for comparison purposes. The average signature matching throughput achieved by the DFA-based (uncompressed) signature matching engine was ~400 Mbps. On the other hand, the average signature matching throughput achieved through the A-DFA and the RCDFA implementations was of the order of 10 Mbps and 200 Mbps, respectively. It was pointed out in the evaluation that the reduction in the signature matching throughput was due to the sequential fetch of additional state transitions as part of the decompression process.

http://doi.org/10.18080/ajtde.v6n3.125

Moreover, Qi *et al.* (2011) pointed out that multiple terabits of on-chip memory bandwidth are required to support line-rate signature matching in methods such as D²FA and A-DFA.

2.3. Hardware-Oriented Transition Compression Techniques

Hardware-oriented transition compression techniques can be classified into hash-based and bitmap-based techniques (Subramanian *et al.*, 2016). Hash-based solutions identify and store the non-redundant transitions in a DFA in a hash table by using the current state and character as hash keys. A hash-based technique was proposed in Lunteren & Guanella (2012) where the DFA transitions are converted into rules which are stored in the memory. There can be multiple rules associated with every unique transition and all of them have to be stored in on-chip memory. As the number of signatures increases, the number of customized rules also increases, which are eventually stored in the off-chip DRAM. The latency associated with a rule fetch from the DRAM reduces the signature matching throughput affecting the scalability of the solution. In worst case scenarios, the throughput achieved by the solution dropped to 2 Gbps, while the maximum throughput that can be achieved by the solution is 73.6 Gbps.

On the other hand, bitmap-based techniques compress adjacent transitions that are identical to each other in a DFA and a bitmap is used to identify the transition indices which have been compressed. For example, a K-bit bitmap is used to identify if a transition is compressed or not in a sequence of K state transitions. An index 'i' in the K-bit bitmap has a '0' stored, if the transition corresponding to that index is compressed. On the other hand, the index has a '1', to identify a transition that is not compressed. The transitions that are not compressed are stored in a unique transition list with a unique transition index to identify their location. The unique transition index corresponding to a transition can be found by calculating the number of 1's in the bitmap before the index of interest until the least significant bit position. Figure 1 shows a transition sequence with 8 transitions. The transitions corresponding to indices 1, 4, 5 and 7 represented in blue are compressed, while the ones represented in green are uncompressed. The bitmap corresponding to each character is shown, along with the unique transition list, in Figure 1.

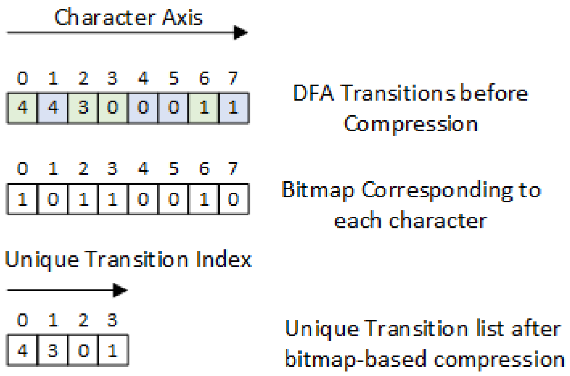


Figure 1. Example of a bitmap and a unique transition list.

Reorganized and Compact DFA (RCDFA) (Wang *et al.*, 2011) is a bitmap-based transition compression technique which performs bitmap-based compression along the state axis. The RCDFA originally achieves transition compression rates of the order of 97-98%. However, in order to reduce the number of bitmaps stored in the memory, RCDFA stores additional redundant state transitions, which reduces the compression rates to 95%, resulting in an increased overall memory footprint of the compressed DFA.

Another bitmap-based transition compression technique is proposed in Qi *et al.* (2011), which achieves transition compression of the order of about 90% by grouping states and is called Front-End Acceleration for Content-Aware Network processing (FEACAN). This technique observes both intra-state as well as inter-state transition redundancy in a DFA. The intra-state redundancy is removed by compressing the transitions using bitmap along the character axis. For example (see Figure 2(a)), the state transitions corresponding to the index in character 7 is compressed in states 0, 2, 3, 4 and 7 since it is identical to that of character 6. The inter-state redundancy is removed by grouping the states into subsets and by comparing the compressed transitions within the state groups. After state grouping, one of the states in the group is referred to as the leader state and all other states in the group are called the member states. After the state grouping, a comparison is made at each unique transition index between the leader transition and all the member transitions. Member transitions corresponding to each unique transition index are compressed, if and only if all of them are identical to the leader transition. This step removes inter-state redundancy. Figure 2(a) shows a DFA with 8 states and 8 characters and Figure 2(b) shows the DFA after the bitmap-based intra-state compression step, while Figure 2(c) shows the compressed DFA after inter-state compression. The transitions shown in red in Figure 2(c) are the redundant ones which are stored in memory, even though they are the same as the leader transition. It can be seen from Figure 2(c) that only 33 out of 64 transitions are compressed using the above-mentioned compression algorithm. Both the compression algorithms discussed above propose a hardware-based decompression engine to achieve signature matching at multi gigabit rates.

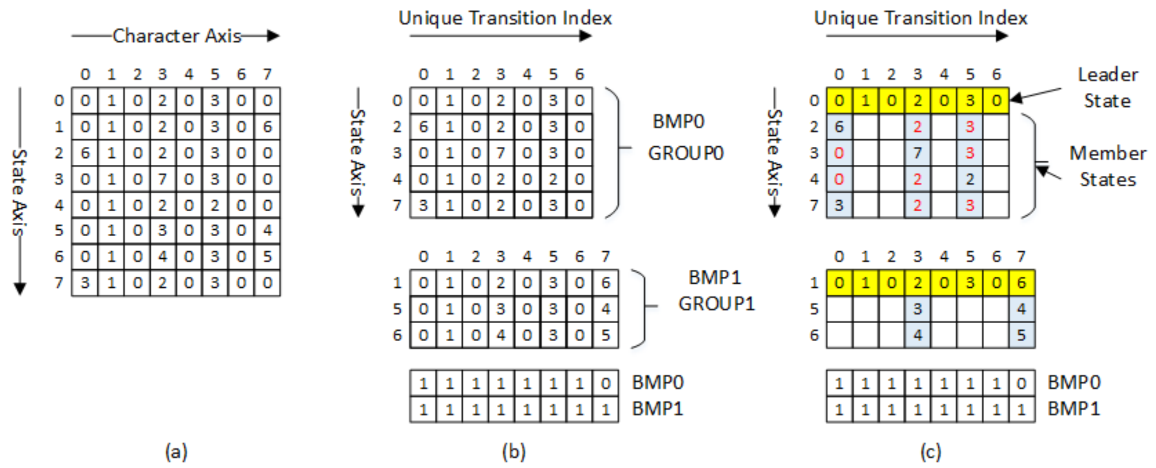


Figure 2. (a) An Example of a DFA with 8 states and 8 characters; (b) Compressed DFA after bitmap-based compression in the character axis and state grouping; (c) Compressed DFA after intra-state and inter-state compression.

To summarize, the memory components in bitmap-based techniques can be split into control and transition memories. The transition compression rates achieved through the bitmap-based techniques are around 90% in the case of FEACAN and 95% in the case of RCDFA. Thus, the resulting transition compression rates result in inefficient storage of the compressed DFA in the on-chip memories.

2.4. Key Contributions

The state-of-the-art bitmap-based compression techniques do not result in efficient transition compression, leading to inefficient usage of on-chip memories to store the compressed transitions. This can either be due to the algorithmic limitations as in the case of Qi *et al.* (2011) or the redundant transitions being stored to reduce the number of unique bitmaps stored in memory, as in the case of Wang *et al.* (2011). Addressing these weaknesses, two bitmap-based transition compression techniques, the Member State Bitmask Technique (MSBT) and the Leader State Compression Technique (LSCT) were proposed in Subramanian *et al.* (2016). The key idea behind these two techniques is an additional level of indexing with the introduction of bitmasks, which efficiently index the non-redundant transitions after bitmap-based transition compression. The additional indexing not only results in a reduced transition memory usage, but also reduces the overall memory usage to store the compressed transitions. This paper describes an extension of the MSBT and LSCT ideas with the following key contributions:

- This paper proposes the key hardware building blocks for the decompression system using the proposed MSBT and the LSCT compression methodologies. Additionally, this paper proposes and evaluates the theoretical relationship between the best- and the worst-case throughput that can be achieved by the MSBT and LSCT based hardware decompression engines and the number of pipeline stages associated with the hardware engines. The experimental evaluation of the optimizations introduced in the hardware accelerator shows further increases in the per-stream signature matching throughput by a factor of 1.2 to 1.4x.

- This paper details the validation of the software model of the proposed transition decompression engines. The software model of the decompression engines is validated by injecting 1 MB of synthetic traffic of various maliciousness levels. The identical signature matching results between the proposed models and the DFA further validate the functional correctness of the proposed methods.

3. Member State Bitmask Technique

3.1. Masking the Redundant Member Transitions – Member Transition Bitmask

MSBT is a two-dimensional transition compression technique that performs intra-state and inter-state transition compression. The intra-state redundant transitions are compressed along the character axis through bitmaps and the compressed states are grouped into subsets of states. The state grouping algorithm proposed in Qi *et al.* (2011) is used for state grouping. States which share the same bitmap and a certain percentage of identical transitions, defined by the transition threshold, are clustered into a group. After grouping the states into subsets of states, a leader state is identified for each group, while the rest of the states are called the member states. As part of the inter-state compression, the transitions in a member state that are identical to the transitions in a leader state at each unique transition index are compressed. The member transition which is not identical to that of the leader transition in a member state can be identified using a Member Transition Bitmask (MTB) for each of the member states. The MTB is composed of a sequence of mask bits, where each bit corresponds to a unique transition index and represents whether a member transition at an index is identical or different in comparison to the leader transition at the same index. If the member and leader transitions are identical at the unique transition index, then the bitmask bit corresponding to the index is marked '0' in the MTB. If not, the bitmask bit for the index is marked '1' in the MTB.

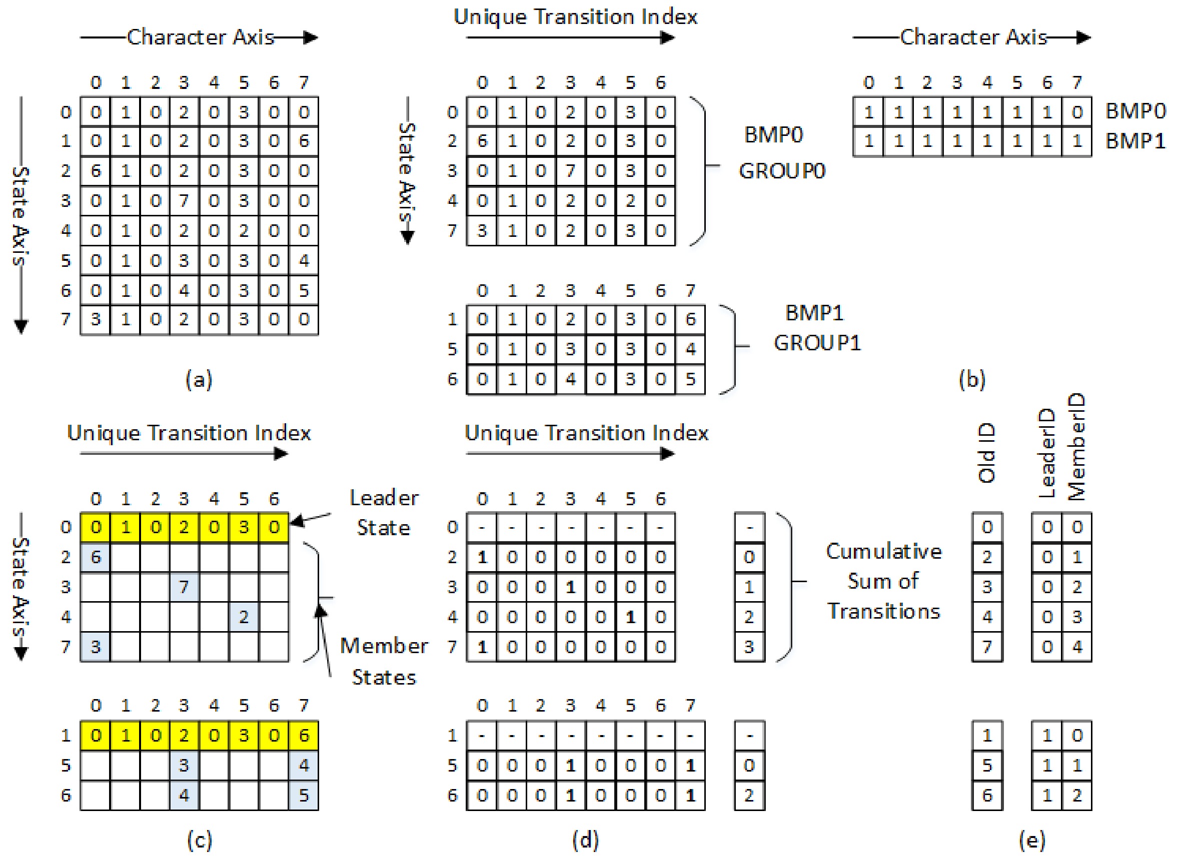


Figure 3. (a) Original DFA before compression; (b) Compressed DFA after bitmap-based intra-state compression and state grouping; (c) Compressed DFA after MSBT; (d) Member Transition Bitmask for each member state; (e) Encoded State representation after MSBT.

Figure 3(a) shows the original DFA before transition compression and Figure 3(b) shows the DFA after the intra-state compression and the state grouping step. Figure 3(d) shows the MTB for each member state in a group. Figure 3(c) shows the compressed transitions after the inter-state compression step. For example, the bitmask bit at index ‘0’ for state ‘2’ has a ‘1’, representing that the member transition at the index is different from the leader transition at the same index. On the other hand, the bitmask bit at index ‘3’ for the state ‘2’ has a ‘0’, representing that the member transition at the index is the same as the leader transition at the same index. It can be seen from Figure 3(c) that 41[ii] transitions are compressed using the MSBT compression in comparison to 33 transitions that were compressed as shown in Figure 2(c) using the same reference DFA. The transitions which remain uncompressed in the leader state are shown in yellow and the transitions which remain uncompressed in the member states are shown in blue in Figure 3(c). These are the transitions which are stored in memory after implementing the MSBT-based transition compression.

If a DFA is compressed using the MSBT, a next state transition can be decompressed in the following way. If the current state is a leader state, then the transition at the unique transition index corresponding to the incoming character is directly assigned as the next state. If the current state is a member state, the bitmask bit corresponding to the unique transition index decides the next state. If the bitmask bit at the unique transition index is a '1', the member transition which remains uncompressed corresponding to the unique transition index is assigned as the next state. If the bitmask bit corresponding to the unique transition index is a '0', the leader transition corresponding to the unique transition index is chosen as the next state.

The cost paid for the additional compression is the memory used to store the MTB. The maximum width of the unique transition index for a group defines the length of the MTB. For example, the maximum width of the unique transition index corresponding to group '0' is 7, resulting in a 7-bit MTB for each member state in the group. A cumulative sum of all member transitions, which remain uncompressed until each member state in a group is stored in the memory along with the MTBs, is shown in Figure 3(d). For example, a cumulative sum of '3', corresponding to state 7 represents that 3 member transitions are stored in memory before the first uncompressed member transition belonging to state 7 is stored.

In the MSBT, the states are encoded and represented as a combination of leaderID and memberID similar to the state encoding technique used in Qi *et al.* (2011). Figure 3(e) shows the state encoding between the two representations. The leaderID identifies the group to which a state belongs and the memberID identifies the member representation within a group of states. The memberID for the leader state is always kept '0' to easily differentiate between a leader state and other member states.

3.2. Functional Description of the Hardware Decompression Engine (MSBT)

3.2.1. Components of the Hardware Decompression Engine

A hardware acceleration engine is proposed to decompress the transitions that are compressed using the MSBT. Figure 4 shows the functional architecture of the signature matching engine. The engine is split across three processing stages, to include the Address Lookup Stage (ALS), the Leader Transition and Bitmask Fetch Stage (LTBFS) and the Member Fetch Stage (MFS). There are four lookup tables across which the compressed transitions and the control information are split. The Leader Transition Table (LTT) belongs to the second stage and stores the transitions which remain uncompressed after the bitmap-based compression among the leader states. The Member Bitmask Table (MBT) also belongs to the second stage and stores the MTB for each member state along with the cumulative sum of transitions. The Member Transition Table (MTT) belongs to the third stage and stores the member transitions which remain uncompressed after the intra- and inter-state compression. The Address Mapping Table (AMT) belongs to the first stage and stores the address location of the first transition in LTT, MTT and the first MTB for each group, which are referred to as LTT base address, MTT base address and MTB base address, respectively. The AMT also stores the bitmap for each group.

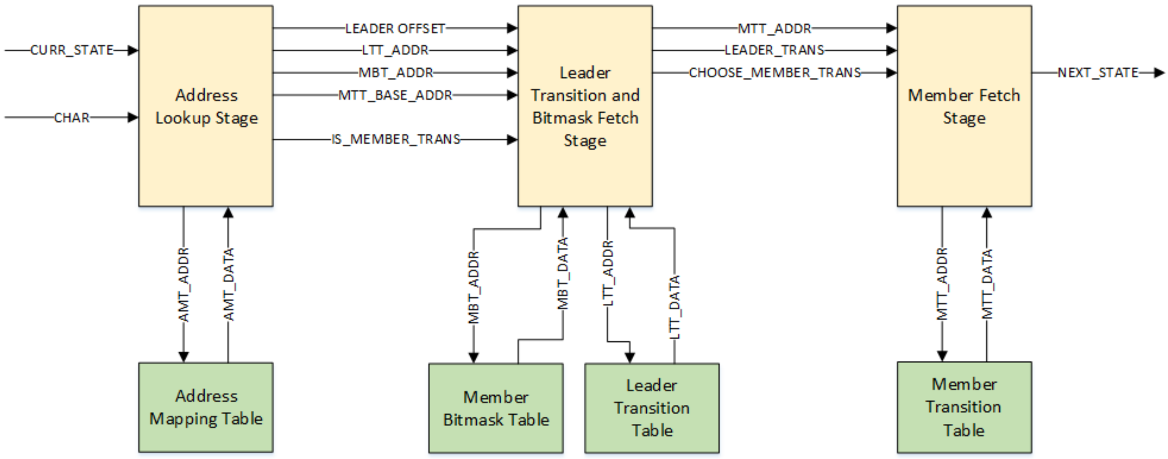


Figure 4. Functional description of the hardware-based decompression architecture for MSBT.

3.2.2. How to Fetch a Compressed Transition

The current state and the incoming character are passed as inputs to the first stage to compute the leader offset, the address location for LTT and MBT based on the data fetched from the AMT. As mentioned earlier, the DFA state is represented as a combination of the leaderID and the memberID after compression. The leaderID is used as the address to fetch the information from the AMT. The leader offset identifies the unique transition index corresponding to the character. The leader offset can be calculated by computing the total number of 1's in the bitmap from the least significant bit position to the position before the character. The LTT base address fetched from the AMT is added together with the leader offset to generate the address location for the LTT. The MTB base address fetched from AMT is added together with 'memberID-1' to generate the address location for the MBT.

The second stage uses the addresses calculated in the first stage to fetch the data from the LTT and the MBT. The bitmask bit corresponding to the leader offset position in the MTB will identify if the member state transition is compressed. If the current state is a member state and the bitmask bit corresponding to the leader offset is '1', a member offset is calculated. The member offset is calculated by computing the total 1's in the MTB from the least significant bit position to the position before the leader offset. The member offset identifies the relative position of a transition in the member state with respect to its first transition that remains uncompressed. The member offset information when added together with the cumulative sum provides the relative position of a transition that remains uncompressed in the group. This information is added together with the MTT base address to generate the address location for the transition in the MTT. The third stage takes the generated leader transition and the MTT address location as inputs. The next state assignment is multiplexed between the transition fetched from the LTT and the transition fetched from the MTT depending on the bitmask bit corresponding to the leader offset and the current state, as discussed previously.

All memories used in the MSBT implementation are single port memories and can be categorized into control and transition memories. AMT and MBT belong to the control memories, as they store the control information such as the base addresses, the bitmaps and the bitmasks which are used to compute the location of a compressed transition. On the other hand, the LTT and the MTT are transition memories as they actually store the compressed state transitions. The basic idea of the proposal is to store more information in the control memory in comparison to the state-of-the-art implementations to improve the transition compression, resulting in an overall reduction in the memory usage.

3.2.3. Example of a Transition Fetch

Figure 5 illustrates the various tables explained above with respect to the compressed DFA shown in Figure 3(c). To maintain simplicity with the representation of the transitions in the transition memories, the encoded representation is not shown in Figure 5. The addresses and the state ID's are represented as decimal numbers. Moreover, the bitmap and the MTB have been represented in the binary format with the least significant bit on the right, which denotes the index with the least value.

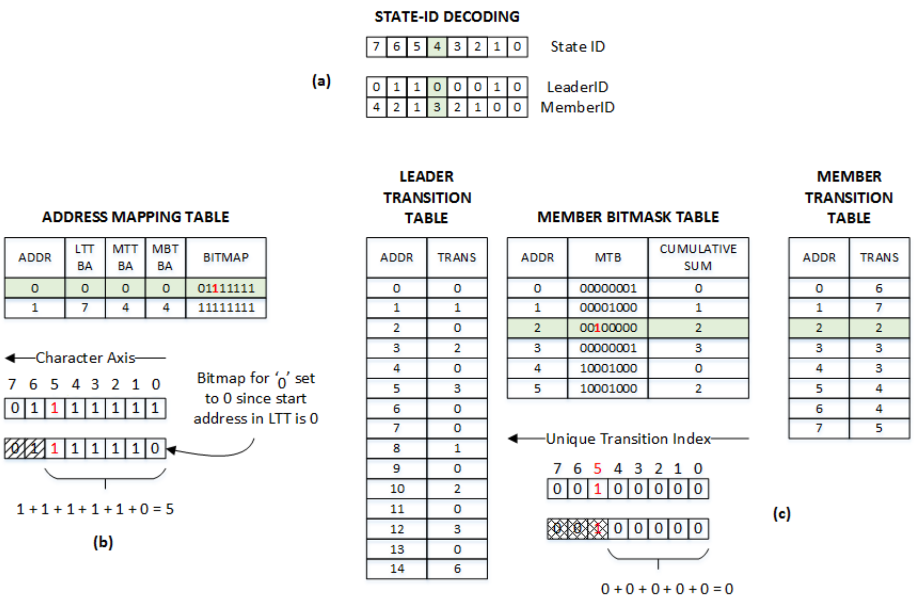


Figure 5. Representation of the compressed DFA with respect to MSBT decompression system.

The transition fetch for a state character combination of '4' and '5' is used as an example to explain the decompression process. The information fetched from various memories is highlighted in green, while the bits of interest in the bitmap and the bitmask are highlighted in red in Figure 5. The leaderID corresponding to state 4 is '0' and is used as the address to fetch the data from the AMT. The bitmap bit corresponding to character '5' is '1' and, in turn, the computed leader offset is '5'. The LTT base address value of '0' is fetched from the AMT which, when added to the leader offset, generates '5' as the LTT address location. Since the state '4' is a member state in group 0, the bitmask bit at the leader offset position is checked to identify whether the next state is assigned from the LTT or the MTT. Since the bitmask bit at the leader offset position is '1', the transition fetched from the MTT is assigned as the next state. In this case, since both the MTT base address and the member offset are '0' and the cumulative sum of transitions is 2, the computed MTT address location is '2'. The transition fetched from MTT address location '2' is '2', which exactly matches the transition corresponding to the state character combination seen in Figure 3(a).

4. Leader State Compression Technique

4.1. Compressing the Most Repeated Leader Transitions – Leader Transition Bitmask

Bitmap-based intra-state transition compression can effectively compress transitions which are identical and at the same time adjacent to each other. Even after bitmap-based compression, there are certain transitions in the unique transition list which have the same transition entry, but cannot be compressed, just because they are not contiguous. During the MSBT-based compression, if these transitions belong to a member state, there is a chance of them being compressed during the inter-state compression step. On the other hand, if these transitions belong to a leader state, they cannot be compressed at all. Specifically, this scenario occurs in the case of failure transitions belonging to a leader state. These failure transitions can be spread across the character axis and can be blocked by those transitions which lead to a forward match in a signature. For example, after the MSBT, the leader state '0' corresponding to group '0' shown in Figure 3(c) has transition '0' at unique transition indices 0, 2, 4 and 6. Even though these transitions are exactly the same, they are not compressed as they are not adjacent to each other. Based on our observation, on average about 40% of the transitions in the unique transition list among the leader states converge to a same next state and are not compressed because of the reasons explained above.

In order to reduce the redundancy in the leader state unique transition list, as discussed above, an efficient method is proposed to index a single most repeated transition through the Leader Transition Bitmask (LTB). The single most repeated transition can be identified by sorting all the transitions in the unique transition list based on the frequency of their occurrence using sorting algorithms such as the quick sort (Cormen *et al.*, 2009). After identifying the single most repeated transition, a single bit at each unique transition index position is used to distinguish the indices with the most repeated transition in the leader state. An index position with the most repeated transition is represented with a '0'; and a '1' if not. The most repeated transition is stored only once, while the transitions that differ remain uncompressed and are stored in the memory. The proposed LTB can be combined with the MSBT. The LTB is introduced after the inter-state compression step in the MSBT and brings additional compression to a DFA. This methodology of compressing the repeated leader transitions together with MSBT is called the Leader State Compression Technique (LSCT).

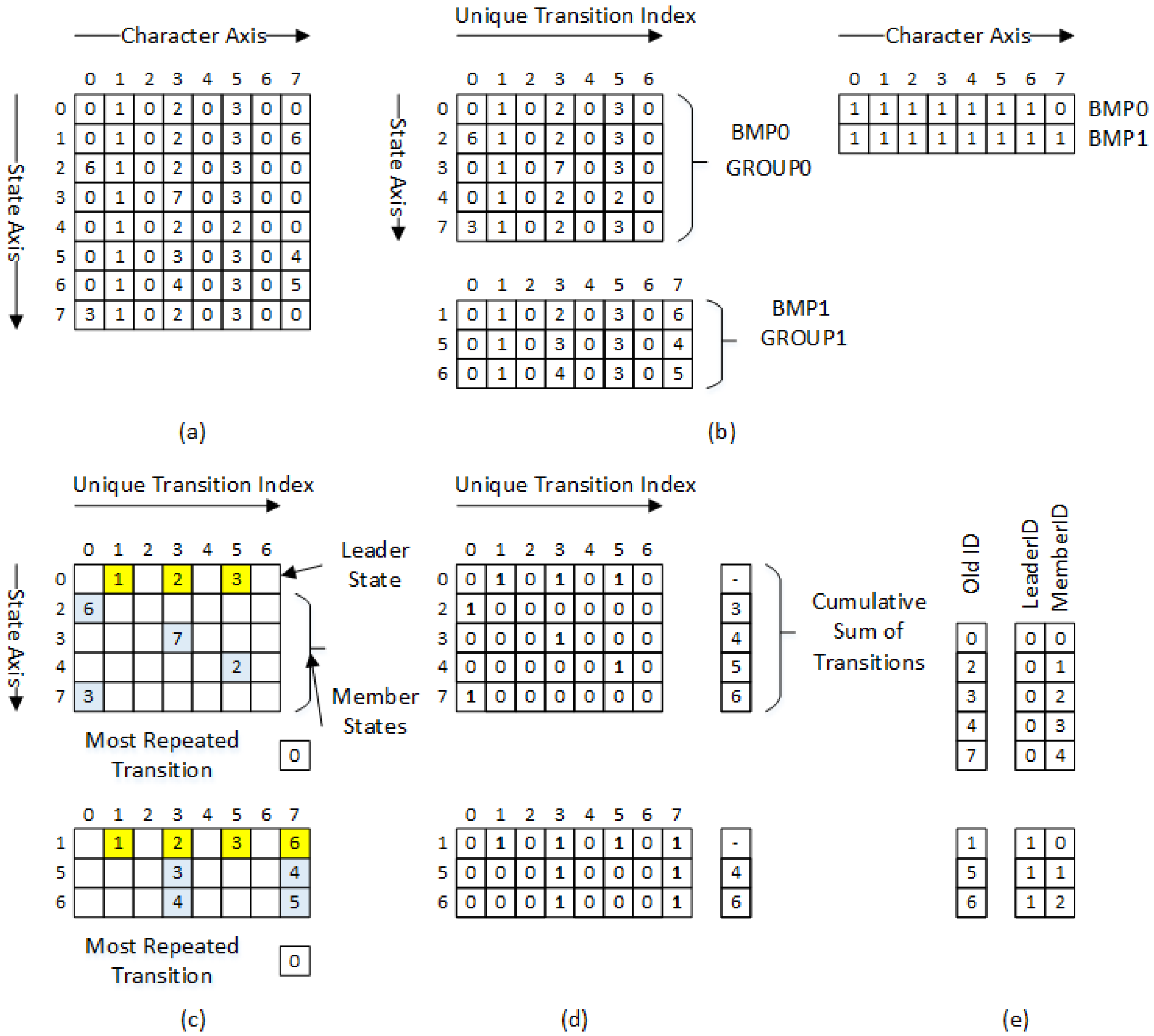


Figure 6. (a) Original DFA before compression; (b) Compressed DFA after bitmap-based intra-state compression and state grouping; (c) Compressed DFA after LSCT; (d) Member Transition Bitmask and Leader Transition Bitmask for each state; (e) Encoded State representation after MSBT.

Figure 6(a) shows the original DFA and Figure 6(b) shows the DFA after the intra-state compression by using bitmaps and state grouping. Figure 6(c) shows the compressed DFA after the intra-state and inter-state compression based on the LSCT. It can be seen that a total of 47[iii] transitions are compressed using LSCT, while only 41 transitions were compressed using MSBT, as shown in Figure 3(c). Transitions represented in yellow are those leader transitions which remain uncompressed, while the transitions represented in blue are those member transitions which remain uncompressed after the LSCT. The LTB and the MTBs corresponding to the leader and member states are shown in Figure 6(d). For example, the entry corresponding to the unique transition index ‘2’ in group ‘0’ has a ‘0’ which represents that the leader transition corresponding to the entry is the most repeated transition. On the other hand, the entry corresponding to the unique transition index ‘3’ in group ‘0’ has a ‘1’ representing that the leader transition corresponding to the entry is not the most repeated transition.

If a DFA is compressed using the LSCT, the next state transition can be decompressed in the following way. If the current state is a leader state, then the bitmask bit corresponding to the unique transition index in the LTB is calculated. If the bitmask bit is ‘0’, then the most repeated transition is assigned as the next state. If the bitmask bit is ‘1’, then the leader transition which remains uncompressed corresponding to the unique transition index is assigned as the next state. If the current state is a member state, the bitmask bit corresponding to the unique transition index in the MTB is identified. If the bitmask bit is ‘0’, then the same procedure is followed as in the case of the leader state. If the bitmask bit is ‘1’, then the member transition which remains uncompressed corresponding to the unique transition index is assigned as the next state.

4.2. Functional Description of the Hardware Decompression Engine (LSCT)

4.2.1. Components of the Hardware Decompression Engine

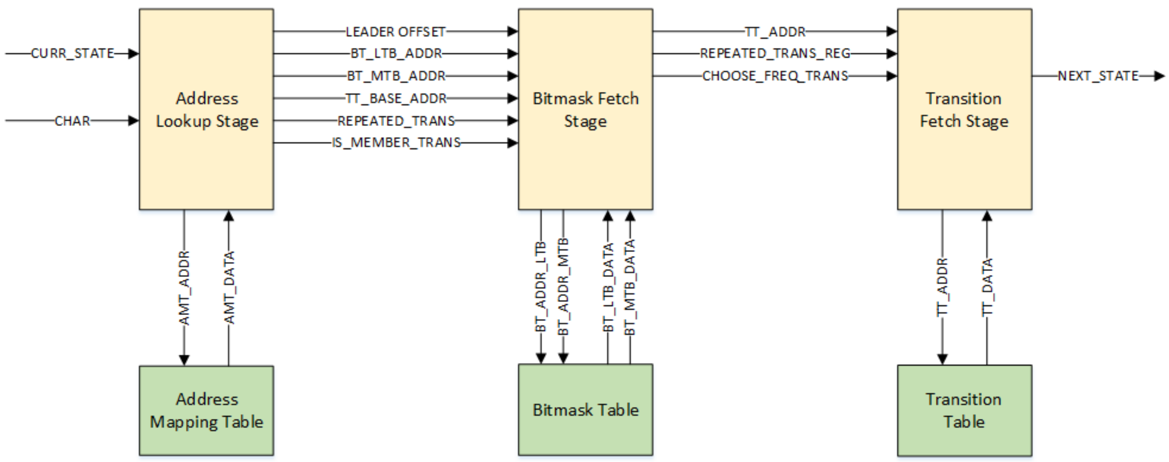


Figure 7. Functional description of the hardware-based decompression architecture for LSCT.

A hardware acceleration architecture is proposed to decompress the transitions which are compressed using the LSCT. Figure 7 shows the functional architecture of the signature matching engine. The engine is split across three processing stages, to include the Address Lookup Stage (ALS), the Bitmask Fetch Stage (BFS) and the Transition Fetch Stage (TFS), respectively. The transitions which remain uncompressed and the control information are split across three lookup tables. The Bitmask Table (BT) belongs to the second stage and stores the LTB in the case of the leader state and the MTB in the case of a member state, along with the cumulative sum of transitions for each state. Unlike the MSBT, the BT in LSCT is implemented using a dual port memory, as both the MTB and the LTB have to be fetched simultaneously. The Transition Table (TT) belongs to the third stage and stores the leader and the member transitions that remain uncompressed after the intra- and inter-state compression. The Address Mapping Table (AMT) belongs to the first stage and stores the base address of the first transition that remains uncompressed and the first bitmask for each group. These are referred to as TT base address and BT base address, respectively. The AMT also stores the bitmap for each of the groups and the most repeated transition in the leader state. Similar to the MSBT, the AMT and BT are the control memories and TT is the transition memory.

4.2.2. How to Fetch a Compressed Transition

The current state and the incoming character are passed as inputs to the first stage to compute the leader offset and the address location to fetch the LTB and the MTB (BT_LTB_ADDR & BT_MTB_ADDR). The leaderID (which is part of the state encoding) corresponding to the current state is used as the address to fetch the data from the AMT. The leader offset is calculated analogous to how the leader offset was computed in the case of MSBT. The BT base address represents the address from which the LTB is fetched. The memberID when added to the BT base address provides the address from which the MTB and the cumulative sum of transitions are fetched. The first stage also provides the TT base address and the most repeated transition as inputs to the second stage.

The addresses generated by the first stage are used to fetch the LTB and the MTB from the BT simultaneously. The bitmask bit corresponding to the leader offset is checked in both the LTB and MTB; these bits are denoted as the leader bitmask bit and the member bitmask bit. If the current state is a leader state and the leader bitmask bit is '1' or if the current state is a member state and the member bitmask bit is '0', a transition offset is calculated similar to the member offset calculation in the case of MSBT. The transition offset is then added together with the TT base address to generate the TT address location. On the other hand, if the current state is a member state and the member bitmask bit is '1', the transition offset which is calculated is added together with the cumulative sum of transitions and the TT base address to generate the TT address location. If the current state is a member state and both the leader bitmask bit and the member bitmask bit are '0' or if the current state is a leader state and the leader bitmask is '0', then there is no need for any address computation as the most repeated transition can be assigned as the next state. The second stage provides the information on when the generated address can be used for TT to the third stage based on the above combinations.

The third stage takes the generated TT address, the most repeated transition and assigns the compressed state transitions depending on the bitmasks. The transition is fetched from the generated TT address and the next state is multiplexed between the transition fetched from TT and the most repeated transition based on the combinations discussed above.

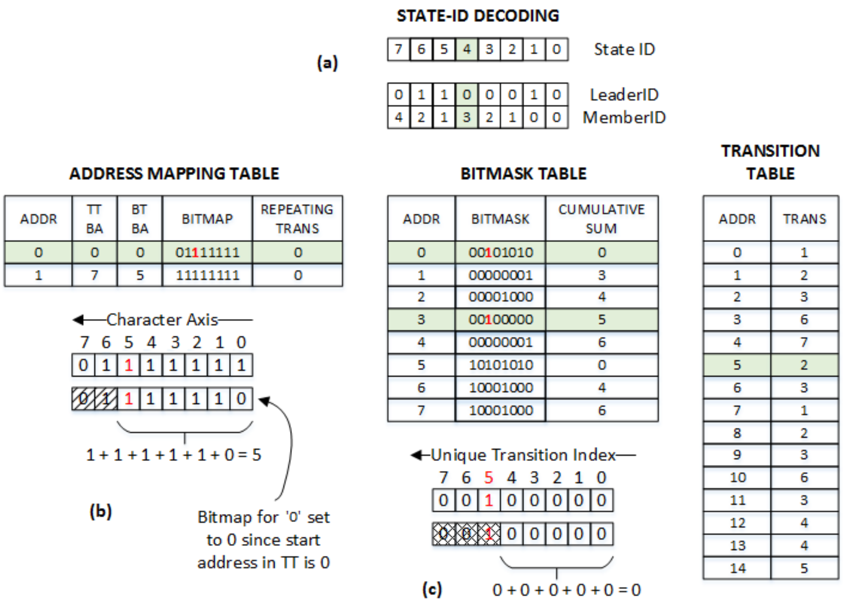


Figure 8. Representation of the compressed DFA with respect to MSBT decompression system.

4.2.3. Example of a Transition Fetch

Figure 8 illustrates the various tables explained above with respect to the compressed DFA shown in Figure 6(c). Similar to the example shown for MSBT, the decimal numbers are used to represent the addresses and transitions. The bitmap and the MTB have been represented in binary format with the least significant bit on the right, which denotes the index with the least value. The transition fetch for a state character combination of ‘4’ and ‘5’ respectively is shown as an example to explain the decompression based on the LSCT compression. The information fetched from various memories is highlighted in green, while the bits of interest in the bitmap and the bitmask are highlighted in red in Figure 8.

The leaderID corresponding to state 4 is ‘0’ and is used as the address to fetch the data from the AMT. The bitmap bit corresponding to the character is ‘1’ as calculated from the bitmap, which results in the leader offset of ‘5’. Since state 4 is a member state in group ‘0’, the LTB and the MTB are fetched from address locations ‘0’ and ‘3’, respectively. Both the leader bitmask bit and the member bitmask bit corresponding to leader offset ‘5’ are calculated to be ‘1’. Since the current state is a member state, the member bitmask bit takes priority, which represents that the transition corresponding to the state character combination is not compressed. The TT address from which the compressed transition is fetched is calculated to be ‘5’, since both the transition offset and the TT base address are ‘0’. The transition ‘2’, which is fetched from TT address location ‘5’, is the same as seen in the uncompressed DFA shown in Figure 6(a).

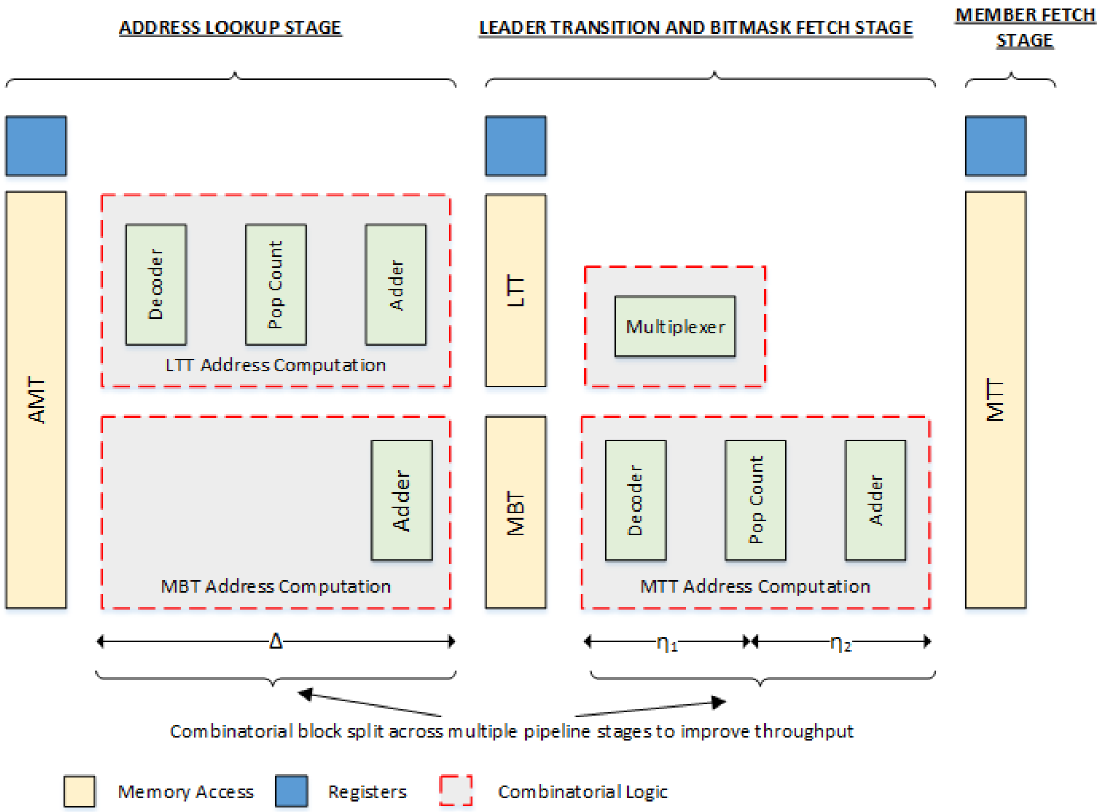


Figure 9. Hardware building blocks for MSBT-based decompression system.

5. Considerations for Hardware Implementation

5.1. Hardware Building Blocks Description

Figures 9 and 10 show the hardware building blocks for the MSBT and the LSCT decompression engines, respectively. The memory accesses and the associated combinatorial blocks have been shown in yellow and grey blocks, respectively. For example, the address lookup stage in both Figures 9 and 10 consists of a single memory access for the AMT, followed by a combinatorial logic block which processes the data fetched from the memory to generate the addresses for the next stage. The same applies to all other functional blocks. Assuming that all the memories are clocked at the same frequency, it would take 3 clock cycles to decompress a transition based on the MSBT or the LSCT. Since there are 3 memory accesses that are part of the transition decompression, it would take a minimum of 3 clock cycles to determine the next state.

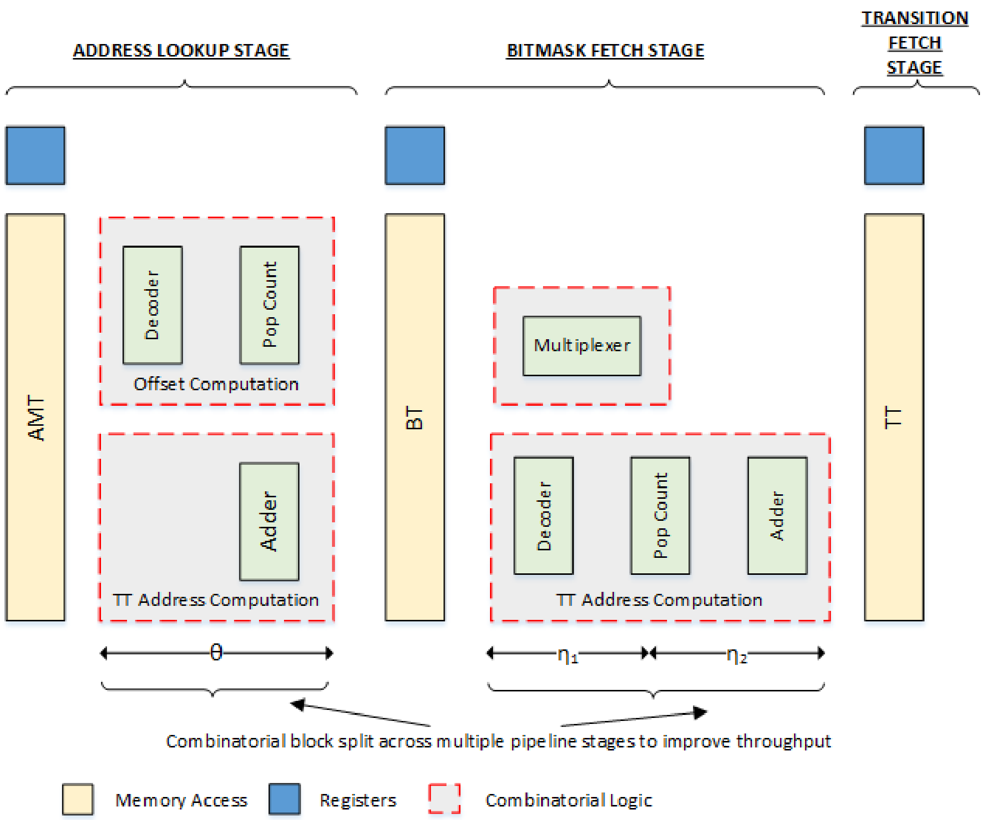


Figure 10. Hardware building blocks for LSCT-based decompression system.

The key combinatorial building blocks required for the hardware realization of the MSBT and the LSCT decompression can be classified into the decoder, the population count (Pop Count), the adder and the multiplexer blocks. The Pop Count block is used to compute the offsets from the bitmap or the bitmask by computing the total number of set bits (1's) in an input vector. The pop-count block can be realized similar to the 64-bit Wallace tree structure proposed in Rajaraman *et al.* (2008). MSBT and the LSCT implementations require the pop-count to be performed on a 256-bit input vector to support the ASCII character set. So, the 64-bit structure proposed in the literature can be extended to a 256-bit structure. The decoder block masks the bits in the bitmap or the bitmask which have to be excluded from the pop-count computation based on the character or the leader offset position in the 256-bit vector. These are the bits from the position of interest, i.e. the character index in the case of the bitmap and the leader offset index in the case of the bitmask until the most significant bit position. Adder circuits such as the ones proposed in Zeydel, Baran & Oklobdzija (2010) can be used to achieve high speed and low power implementations in the address generation blocks. The multiplexer block chooses one of the 256 bits in the bitmap or the bitmask to identify the bitmap or the bitmask bit associated with an index. The multiplexer block can be constructed using multiple smaller multiplexers, for example 2 to 1 multiplexers in a logarithmic fashion using the binary tree topology.

5.2. Pipelining vs Throughput

As mentioned in the previous section, each functional stage is a combination of a single memory lookup followed by a combinatorial function which processes the data from the memory. So, it would take 3 clock cycles for a transition to be decompressed from the memory. For example, if the character is consumed in the first clock cycle as an input, the subsequent character can only be consumed in the 4th clock cycle, as it takes 3 clock cycles to process the character and identify the next state corresponding to it. So, in order to keep the pipeline busy and to fully utilize the hardware resources, characters from multiple data streams can be interleaved as proposed in Basu & Narlikar (2005). Since there are 3 pipeline stages (corresponding to 3 memory accesses), characters from 3 different data streams are passed to the pipeline in an interleaved manner once every 3 clock cycles.

To generalize, if 'P' is assumed to be the total number of pipeline stages to process a character from one data stream, characters from 'P' different data streams have to be interleaved to extract the best from the hardware resources. Since each character corresponds to 8 bits and the system can consume one character every clock cycle, the throughput that can be achieved is a product of the frequency and the character width. Assuming F to be the frequency at which the system is clocked, the throughput T achieved by the decompression system can be generalized as shown in (2).

$$T = F \times 8 \text{ bps} \tag{2}$$

Since characters from multiple streams are input to effectively utilize the hardware resources, the maximum throughput that is achieved by one of the interleaved streams is inversely proportional to the number of pipeline stages. Equation (3) shows the maximum achievable throughput $T_{stream\_max\_pipeline}$ for a single stream among the interleaved streams during pipelined operation.

$$T_{stream\_max\_pipeline} = \frac{F \times 8}{P} \text{ bps} \tag{3}$$

Based on (1), the throughput, T, can either be increased by increasing the clock frequency or increasing the character width, i.e., by sending multiple characters per stream per clock cycle. Processing multiple characters per stream requires the conversion of the DFA into a multi-stride DFA, which results in an exponential memory growth and is not a scalable approach (Becchi & Crowley, 2013). Thus, the only way in which the throughput can be improved is by increasing the operating frequency of the transition decompression. The frequency that can be achieved depends on two factors: the latency associated with the SRAM fetch; and the latency associated with the combinatorial processing path in the pipeline. The latency associated with the SRAM fetch can be optimized by effectively configuring the memories as part of the ASIC design flow. Similarly, the combinatorial processing logic associated with various functional stages can be broken down into multiple pipeline stages by introducing additional registers to increase the frequency of operation. On the contrary, splitting the combinatorial path into pipelines will also reduce the maximum throughput that can be achieved by a single stream, since they are inversely proportional to each other, as shown in (3).

In the case of the MSBT and the LSCT, the number of pipeline stages required to process an incoming character can be broken down into a fixed and a variable count. The former is the number of pipeline stages which is a bare minimum requirement due to the associated memory fetches. In both the MSBT and the LSCT, the fixed pipeline stage count is 3, as there are 3 memory fetches as part of the transition decompression. The latter is a variable component, which results from the combinatorial block in each stage being split into multiple smaller stages to improve the clock frequency. In the case of the MSBT and the LSCT, the variable pipeline stage counts for the first stage are defined as Δ and θ, respectively. For example, if the combinatorial block in the ALS is split into 3 (Δ) stages, it would take 4 pipeline stages (3 (Δ) + 1 stage for AMT lookup) in the MSBT to finish the processing associated with the ALS. On the other hand, in the MSBT and the LSCT, the variable pipeline count associated with the second stage is defined by η, as the processing associated with these stages is exactly the same. Equations (4) and (5), respectively, define the total pipeline stage count P_{MSBT} and P_{LSCT} for MSBT and LSCT after design pipelining.

$$P_{MSBT} = 3 + \Delta + \eta \tag{4}$$

$$P_{LSCT} = 3 + \theta + \eta \tag{5}$$

Providing multiple streams to the system is a best-case scenario to completely utilize the hardware resources. In the worst-case scenario, when only one stream is available to the pipeline, the maximum throughput that can be achieved is the same as described in (3). In this scenario, the throughput of the signature matching engine can be increased by directly assigning the next state from the transition fetched from the second functional stage whenever possible. This can happen in the case of MSBT, when the next state transition can be assigned from the transition fetched from the LTT. Similarly, in the case of LSCT, the most repeated leader state transition can be assigned as the next state based on the bitmasks that are fetched in the second stage. However, it should be noted that this is highly dependent on the sequence of characters which are inspected as part of the data stream. Additional circuitry can be added in the hardware to consider this special scenario to achieve a higher throughput when compared with $T_{stream\_max\_pipeline}$. The number of variable pipeline stages η in the second stage is split into η_1 and η_2, where η_1 represents the number of pipeline stages required to decide if the next state can be assigned in the second stage, while η_2 represents the number pipeline stages required for address calculation for the third stage. So, the minimum number of pipeline stages for the MSBT and the LSCT are represented as P_{MSBT_min} and P_{LSCT_min}, as shown in (6) and (7), respectively. The fixed pipeline component in (6) and (7) is represented as 2, as there are only 2 memory accesses required to identify the next state.

$$P_{MSBT\_min} = 2 + \Delta + \eta_1 \tag{6}$$

$$P_{LSCT\_min} = 2 + \theta + \eta_1 \tag{7}$$

$P(c)_{MSBT}$[iv] is defined as the probability of a transition fetch from the third functional stage assigned as the next state, when a sequence of M bytes is inspected by the MSBT decompression system. Equation (8) shows the throughput $T_{stream\_max\_MSBT}$ that can be achieved by a single stream in the worst-case scenario in relation to the probability of the transition fetch. Since P_{MSBT_min} is smaller than P_{MSBT}, a higher throughput can be achieved when $P(c)_{MSBT}$ is smaller. Equation (9) shows a similar relationship between $T_{stream\_max\_LSCT}$ and the probability of a transition fetched from the third functional stage $P(c)_{LSCT}$ being assigned as the next state. Since the transitions which differ from the most repeated leader transitions are also stored in the TT, which belongs to the third stage, $P(c)_{LSCT}$ will be higher than $P(c)_{MSBT}$ for the same character sequence. Similar to MSBT, a higher throughput will be achieved in the case of LSCT when $P(c)_{LSCT}$ is small.

$$T_{stream\_max\_MSBT} = \frac{F \times 8}{\left(P_{MSBT\_min} \times (1 - P(c)_{MSBT})\right) + (P_{MSBT} \times P(c)_{MSBT})} \text{ bps} \tag{8}$$

$$T_{stream\_max\_LSCT} = \frac{F \times 8}{\left(P_{LSCT\_min} \times (1 - P(c)_{LSCT})\right) + (P_{LSCT} \times P(c)_{LSCT})} \text{ bps} \tag{9}$$

6. Results and Discussion

This section analyses the compressed DFA resulting from the MSBT and the LSCT techniques in comparison to various other techniques discussed in the literature. The results section is split into two sub-sections to discuss the effectiveness of the transition compression techniques. The first section describes the experimental evaluation of the proposed techniques. The transition compression achieved using the proposed techniques is compared with A-DFA and FEACAN. The A-DFA and FEACAN are chosen as they represent the state-of-the-art software and the hardware-oriented compression techniques, respectively. This section also compares the memory used to store the compressed DFA in comparison to FEACAN. On the other hand, the second section describes the software simulation setup used to verify the functional correctness of the compression methods. This section further analyses the throughput of the decompression system based on the traffic generated with different levels of maliciousness.

6.1. Experimental Evaluation

A compiler was developed which takes the DFA as an input to generate the compressed transitions along with the control information such as the bitmaps and bitmasks as outputs. The maximum number of states in a group was restricted to 256 states and the transition threshold used for inter-state compression was set to 80%. The same set of states is used as inputs for all the three bitmap-based compression techniques. The compression algorithm was implemented in a Xeon server machine running at 4.4 GHz with 500 GB of main memory.

As proof of concept, the compression scheme was evaluated on DFAs generated across 5 different rule-sets listed in Table 1. The rule-sets were carefully identified to contain both strings and regular expression signatures. Exact match is a group of 500 string signatures synthetically generated from the tool developed by Becchi (2016). The other four rule-sets are extracted from Snort (Roesch, 1999) and Bro (Paxson, 1999) intrusion detection systems, respectively, and are a combination of simple strings and complex regular expressions. The signature sets were converted into the DFA using the regex tool (Becchi, 2016) and the custom compiler performs the MSBT and the LSCT on the generated DFAs. Column 3 in Table 1 presents the total DFA states generated, while columns 5 and 6 present the total leader states and member states after state grouping. Column 4 presents the total number of uncompressed transitions in the DFA.

Table 1. Rulesets Used in Simulations

Signature Set	#Signatures	#DFA States	#Transitions	#Leader States	#Member States
Snort34	34	13834	3541504	575	13259
Snort31	31	19522	4997632	584	18938
Snort24	24	13882	3553792	538	13344
Exact Match	500	15149	3878144	298	14851
Bro	217	6533	1672448	173	6360

6.1.1. Transition Compression Ratio

The transition compression ratio is the ratio of the number of transitions that remain uncompressed to the number of transitions in the original DFA. Figure 11 compares the transition compression achieved by MSBT and LSCT in comparison to the theoretical maximum compression limit. Along with these, the transition compression results are compared across A-DFA and FEACAN as well. The parameters used for the A-DFA algorithm were tuned to achieve the best compression results.

The MSBT clearly shows an improvement of the order of 4-5% in the transition compression when compared with FEACAN. The improvement is a direct consequence of using MTBs to index the position of the redundant member transitions. The MTBs introduced as part of the

MSBT index the redundant transitions in the member state which are not stored in the memory. Table 2 compares the average number of transitions that remain uncompressed in a member state after compressing the DFA using FEACAN and the MSBT. The information in Table 3 is extracted from the compilation results after performing the MSBT and the FEACAN compression on the signature sets described in Table 1. It can be seen that about 50 to 80% of the member transitions stored in FEACAN are redundant and can be compressed efficiently through the introduction of MTBs in the MSBT.

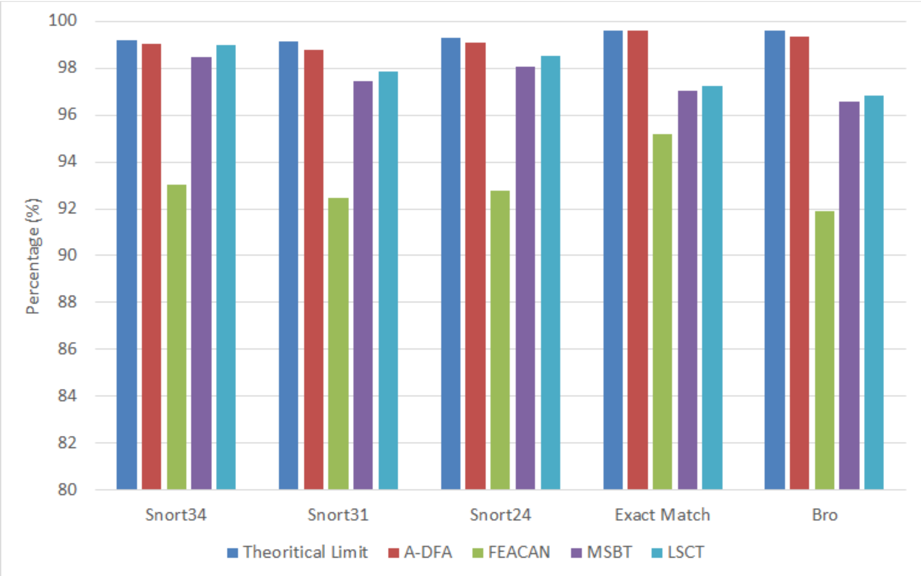


Figure 11. Comparison of transition compression across various techniques.

Table 2. Comparison of average number of transitions per member state after compression

	FEACAN	MSBT
Snort34	17.12	2.16
Snort31	18.38	5.18
Snort24	17.49	3.14
Exact Match	10.62	5.84
Bro	19.27	6.74

Table 3. Comparison of average number of transitions per leader state after compression

	MSBT	LSCT
Snort34	43.53	13.98
Snort31	52.66	14.63
Snort24	51.07	19.13
Exact Match	104.96	82.42
Bro	79.66	56.80

A slight improvement in the transition compression is seen in the case of LSCT in comparison to MSBT as the most repeated leader transition for each leader state is compressed using the LTB. Table 3 shows the average number of leader transitions that remain uncompressed after the LSCT in comparison to the MSBT. The average number of transitions generated in the leader state is directly extracted from the compiler results after performing the MSBT and the

http://doi.org/10.18080/ajtde.v6n3.125

LSCT on the described signature sets. It can be observed from Table 3 that 30-60% of the transitions in the leader are the single most repeated transitions and can be compressed effectively, which results in an increase in the transition compression.

The A-DFA achieves the best transition compression results when compared with all the other bitmap-based compression techniques and is also close to the theoretical limit. Table 4 shows a comparison of the average number of transitions which have to be fetched from the memory in each of the compression methods, before identifying the compressed state transition corresponding to a state character combination. The data in Table 4 was generated by directly analysing the compressed DFA generated through each of the techniques. A closer look at the results from Table 4 will detail the cost paid for the transition compression in the case of A-DFA. Columns 3, 4, 5 and 6 present the average number of transitions that have to be fetched from the memory to identify the compressed transition for A-DFA, FEACAN, MSBT and LSCT, respectively. Column 2 presents the maximum number of transitions that have to be fetched from the memory in the case of A-DFA before identifying the compressed transition. Columns 2 and 3 clearly show that multiple tens of transitions have to be fetched from memory before identifying the compressed transition corresponding to the state-character combination in the case of A-DFA. The sequential search associated with the software-oriented techniques such as A-DFA result in low signature matching throughput. The increased memory bandwidth is the cost paid by software-oriented solutions, such as A-DFA, to compress the redundant transitions. In the case of bitmap-based techniques, a maximum of 2 memory accesses is only required to fetch the compressed transition. The best-case scenario to identify the compressed transition is a single memory fetch in all cases. The ability to quickly fetch the compressed transition from the memory is a differentiating factor in bitmap-based transition compression techniques to achieve high throughput signature matching in comparison to software-oriented techniques. The numbers shown in columns 3 to 6 are average numbers to fetch a compressed transition extracted from the compressed automaton. These average numbers will vary based on access patterns in an automaton.

Table 4. Average number of transitions fetched before fetching the compressed transition

Signature Set	A-DFA		FEACAN	MSBT	LSCT
	Max.	Avg.	Avg.	Avg.	Avg.
Snort34	1036	9.8	1.06	1.01	1.01
Snort31	288	12.84	1.07	1.02	1.02
Snort24	1109	21.01	1.07	1.01	1.01
Exact Match	26	6.65	1.04	1.02	1.03
Bro	44	10.15	1.07	1.03	1.03

Table 5. Parameters Used for Memory Estimation

Parameter	Bit-width (in bits)		
	FEACAN	MSBT	LSCT
LTT Base Address	16	16	-
MTT Base Address	16	16	-
TT Base Address	-	-	18
MBT Base Address	-	16	-
BT Base Address	-	-	16
Bitmask	-	Variable	
Bitmap	256	256	256
Transition	20	20	20

6.1.2. Memory Usage

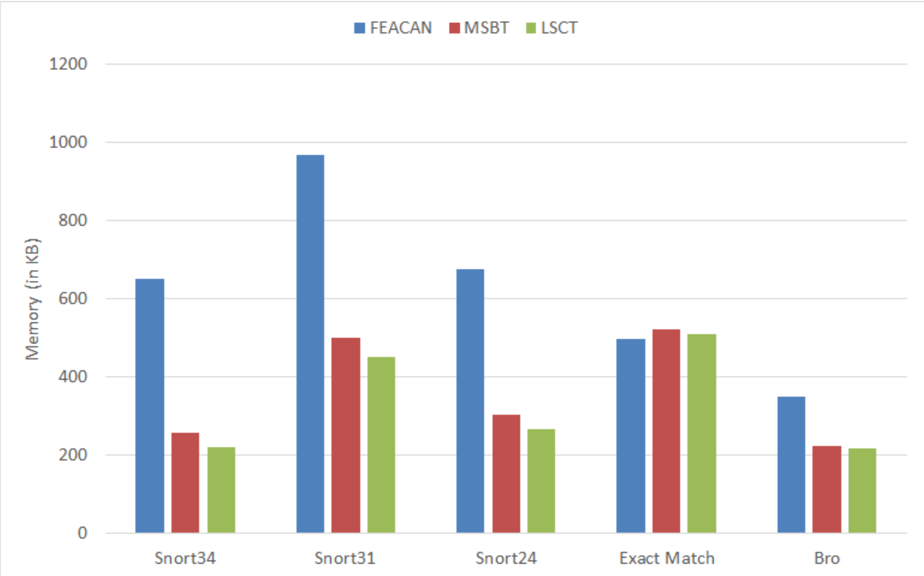


Figure 12. Comparison of memory usage across various techniques.

Memory used to store the compressed automata is calculated by computing the sum of the memory used to store the compressed transitions and the memory used to store the control information, such as the base addresses, bitmaps and bitmasks. Table 5 lists the width of various information that is stored in memory as part of various techniques.

Figure 12 shows a comparison of the memory usage across different bitmap-based compression techniques. A small improvement of 4-5% in the transition compression ratio seen in Figure 11, in the case of MSBT, translates into an overall reduction in memory by 50% in most of the signature sets. Similarly, in the case of LSCT, even a very minute increase in the transition compression results in a significant 5–10% reduction in memory usage when compared with MSBT. Figure 13 shows a comparison of the transition and control memory usage across various techniques. As part of the control memory, FEACAN only stores the base addresses for the transition memories and the bitmap for certain groups of states. This represents a very negligible portion of the overall memory usage, while a major portion of the memory is used to store the compressed state transitions, as shown in Figure 13. On the other hand, in the case of MSBT and LSCT, a considerable portion of the overall memory is needed for the control memory to store the bitmasks, which in turn reduces the memory used by the transition memories. Even though storing the bitmasks for states increases the control portion of the memories, there is a huge reduction in the transition memory usage and also an overall reduction in the memory usage to store the compressed automata. As seen in Figures 12 and 13, Exact Match ruleset is an exception where the average length of the bitmask is 112 bits which, when stored per member state, needed a substantial portion of memory in comparison to other signature sets. The reduction in the transition memory which results by eliminating the redundant transitions in a member state is smaller in comparison to the memory used to store the bitmasks in case of the Exact Match ruleset. The average number of transitions per state after the intra-state compression directly corresponds to the length of the bitmask stored for a state. The memory used to store the compressed transitions in the case of software-based techniques such as A-DFA depends on the chosen memory layout (Chen *et al.*, 2016) and is not compared with the proposed techniques.

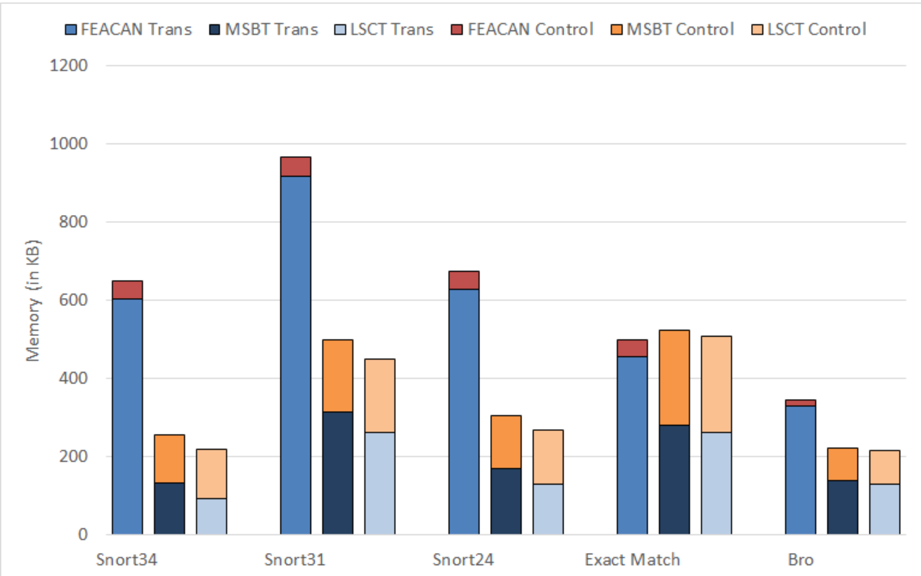


Figure 13. Comparison of transition and control memory usage across various techniques.

6.2. Simulation Setup

A software-based simulation was performed to verify the decompression system based on the MSBT and the LSCT compression. Figure 14 shows the overview of the software simulation setup. The signature sets were compiled into an NFA and a DFA using the regex tool (Becchi, 2016). The synthetic traffic generation module described by Michela, Mark & Patrick (2008), implemented in the regex tool, was used to generate the sequence of characters. The synthetic traffic generation module takes the NFA as an input and generates a sequence of characters, where each character in the sequence is decided based on a maliciousness probability P_M by traversing the NFA. The P_M value is the probability with which a forward transition is made for each character leading to a signature match. A lower P_M value indicates a lower probability of a successful signature match than a higher P_M value. Four different character sequences of 1 MB were generated based on the values chosen for P_M, which were 0.35, 0.55, 0.75 and 0.95. The DFA generated by the regex tool is used as an input to perform the transition compression using the MSBT and the LSCT compilers. A software model of the decompression system was developed for both the MSBT and the LSCT, which was used to perform signature matching on the compressed signatures. Table 6 shows a summary of the signature matching results obtained from the software model of the decompression systems and compares the results with a DFA-based signature matching engine across different signature sets and across the various P_M values. The identical signature matching results seen in Table 6 show the functional correctness of the proposed compression methods. In addition to tracking the total number of signature matches, as shown in Table 6, the next state transitions generated for each of the state character combinations were also monitored and the results were identical in all three systems across all the signatures sets across all the P_M values (not shown in the results section).

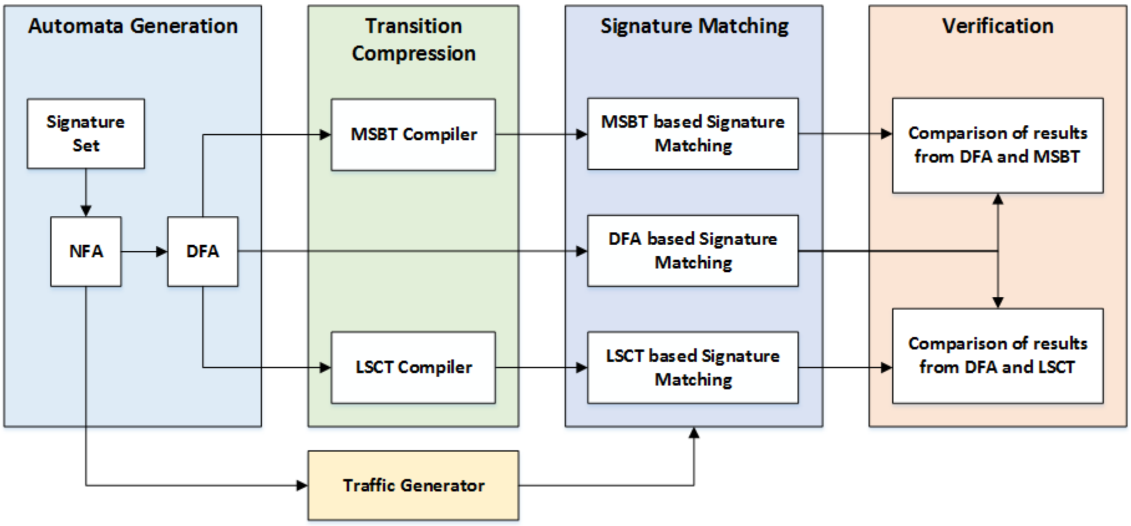


Figure 14. Overview of the software-based simulation environment to verify the decompression system.

6.2.1. Transition Fetch — Dynamic

Figure 15(a) shows the statistics of the number of transitions fetched from the third stage in the case of MSBT as a percentage of total transition fetches, considering traffic traces with various levels of maliciousness. As the probability of maliciousness increases with different character traces, it can be clearly seen that more transitions are fetched from the third functional stage. This can be attributed to two reasons. Firstly, as the maliciousness level in the traces increases, the states which are at higher depths are visited, where the depth of a state refers to the number of positive character matches in a signature. The states which are visited in these cases are traversed if and only if a sequence of positive character matches leads to it. Secondly, only a very small portion of the states are the leader states, while the majority of states are member states after the state grouping process. In the case of the traces with higher maliciousness levels, the probability that the states traversed are member states is very high. The transitions which lead to the state at a higher depth are generally distinct and cannot be compressed. Thus, these transitions will belong to the member states which remain uncompressed, resulting in more transitions fetched from the third stage. Figure 16(a) shows a similar comparison in the case of LSCT and, even at the lowest maliciousness level, which is 0.35, the percentage of transitions fetched from the third functional stage is much higher when compared with MSBT. The compressed transition is fetched from the second stage if and only if the transition is the most repeated transition in the group. Those leader transitions which are different from the most repeated transition are also fetched from the third stage on top of all the member transitions which are fetched. These transitions are directly responsible for the difference in the access patterns in comparison to MSBT.

Table 6. Summary of results after injecting 1MB of byte sequence into the MSBT, the LSCT and the DFA

P_M	Method	Signature Sets				
		Snort34	Snort31	Snort24	Exact_Match	Bro217
0.35	DFA	5	1	29	46	11631
	MSBT	5	1	29	46	11631
	LSCT	5	1	29	46	11631
0.55	DFA	7	2	45	117	6347
	MSBT	7	2	45	117	6347
	LSCT	7	2	45	117	6347
0.75	DFA	8	1	22	331	28701
	MSBT	8	1	22	331	28701
	LSCT	8	1	22	331	28701
0.95	DFA	14591	1286	22818	7841	69976
	MSBT	14591	1286	22818	7841	69976
	LSCT	14591	1286	22818	7841	69976

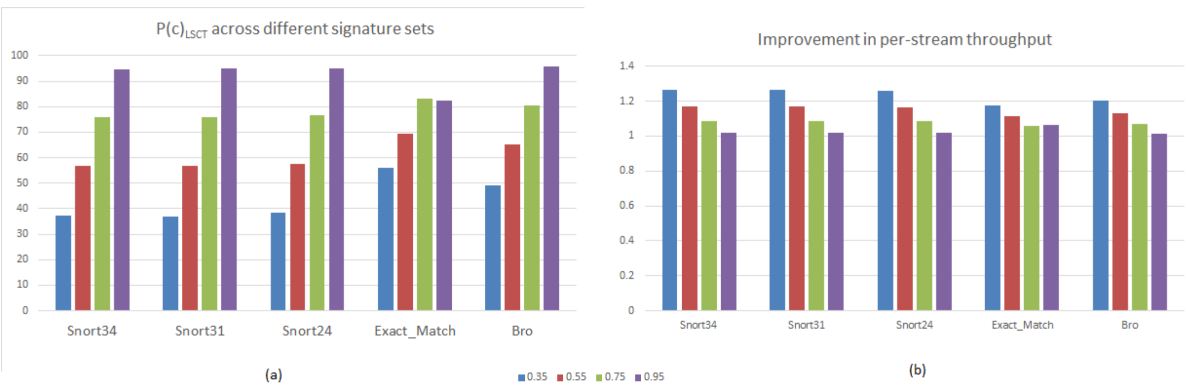


Figure 15. Statistics of transition fetch from the third functional stage in MSBT decompression system and the effective improvement in throughput.

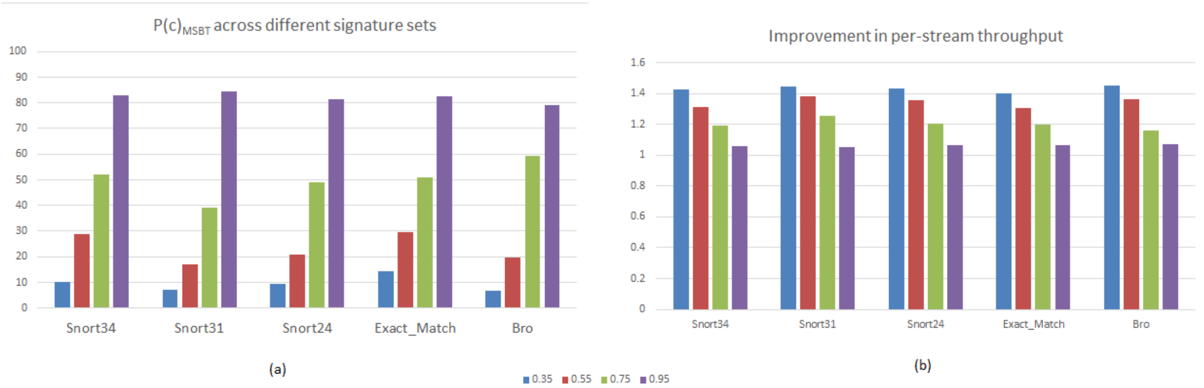


Figure 16. Statistics of transition fetch from the third functional stage in LSCT-based decompression system and the effective improvement in throughput.

Figure 15(b) and Figure 16(b), respectively, shows the improvement in the per-stream throughput that can be achieved in the worst-case scenarios based on the relationship established in the previous sections. It is assumed that it takes two clock cycles to fetch the compressed state transition from the second stage, while it takes three clock cycles to fetch the compressed transition from the third stage. Based on this assumption, the improvement in the per-stream throughput is calculated based on (8) and (9), as shown in Figure 15(b) and Figure 16(b). As the level of maliciousness increases in the traffic, more transitions are fetched from the third functional stage, which reduces the per-stream throughput in the worst-case scenarios. In the case of lower levels of maliciousness, the per-stream throughput that can be achieved in the case of LSCT is less than for MSBT, as the probability of a transition fetch from the third functional stage is higher in LSCT. The difference in the throughput gradually reduces as the levels of maliciousness increase, due to the distribution of the compressed transitions. According to Michela *et al.* (2008), traffic traces with the highest maliciousness levels are not a common occurrence in network traffic traces. So, with low levels of maliciousness, MSBT can be used to achieve better signature matching throughput than LSCT. By assigning the next state directly from the second stage, the per-stream signature matching throughput can be increased by a factor of 1.2 to 1.4 times in the case of the MSBT and the LSCT, as shown in Figure 15(b) and Figure 16(b), respectively.

7. Conclusion

Hardware acceleration of signature matching is a key requirement to perform deep packet inspection at line rates. The signatures which are used for DPI are converted into the DFA to perform signature matching in linear time, while the storage inefficiency associated with the automata is addressed through various transition and state compression algorithms. Bitmap-based transition compression techniques enable hardware acceleration of transition decompression, in turn allowing the signature matching function to be performed in a dedicated hardware accelerator. State-of-the-art bitmap-based transition compression

http://doi.org/10.18080/ajtde.v6n3.125

techniques do not efficiently compress the DFA. Addressing this problem, Subramanian *et al.* (2016) proposed two bitmap-based transition compression techniques to achieve transition compression rates of the order of over 98%. The transition decompression through the proposed techniques is performed in a hardware accelerator so that the signature matching can be performed at line rates. The fundamental building blocks of the hardware accelerator performing the transition decompression corresponding to these techniques were first proposed in this article. Furthermore, a software model corresponding to the decompression engines was designed and verified to validate the proposed compression methods. The functionality of the decompression engines was verified by injecting multiple 1 MB streams of bytes of different levels of maliciousness, across different signature sets, into the software models and the DFA. The identical signature matching results further validated the functional correctness of the proposed compression methods. Furthermore, a theoretical relationship was established between the signature matching throughput achieved through these systems and the number of pipeline stages required by the hardware accelerator to perform the transition decompression. Based on this analysis, further optimization methods were proposed to improve the per-stream signature matching throughput. Experimental evaluations further showed that the proposed optimizations improve the per-stream signature matching throughput by a factor of 1.2x to 1.4x in comparison to the throughput that is achieved without the optimizations.

Acknowledgement

The authors wish to thank Dharmesh Shah for his valuable feedback on this manuscript.

References

Basu, A; Narlikar, G. 2005. 'Fast incremental updates for pipelined forwarding engines'. *IEEE/ACM Transactions on Networking*, 690-703.

Becchi, M. 2016. Regular Expression Processor. Retrieved from http://regex.wustl.edu/

Becchi, M; Crowley, P. 2007a. 'An improved algorithm to accelerate regular expression evaluation'. 3rd ACM/IEEE Symposium on Architecture for networking and communications systems (pp. 145-154). Florida: ACM.

Becchi, M; Crowley, P. 2007b. 'A hybrid finite automaton for practical deep packet inspection'. CoNEXT '07, Proceedings of the 2007 ACM CoNEXT conference. New York.

Becchi, M; Crowley, P. 2013. 'A-DFA: A Time- and Space-Efficient DFA Compression Algorithm'. *ACM Transactions on Architecture and Code Optimization*, 4:1-4:26.

Chen, X; Jones, B; Becchi, M; Wolf, T. 2016. 'Picking Pesky Parameters: Optimizing Regular Expression Matching in Practice', *IEEE Transactions on Parallel and Distributed Systems*, 27 (5), May, pp. 1430-1442.

Cong, L; Yan, P; Ai, C; Jie, W. 2014. 'A DFA with Extended Character-Set for Fast Deep Packet Inspection'. *IEEE Transactions on Computers*, 1925-1936.

Cormen, TH; Leiserson, CE; Rivest, RL; Stein, C. 2009. *Introduction to Algorithms*, Third Edition. MIT Press.

Ficara, D; Giordano, S; Procissi, G; Vitucci, F; Antichi, G; Pietro, AD. 2008. 'An improved DFA for fast regular expression matching'. *ACM SIGCOMM Computer Communication Review*, 38(5), 29-40.

Handley, M; Paxson, V; Kreibich, C. 2001. 'Network Intrusion Detection: Evasion, Traffic Normalization, and End-to-End Protocol Semantics'. Proceedings of the 10th USENIX Symposium. Washington D.C.

Holcomb, J. 2016. securityevaluators.com. Retrieved 26 March 2016 from https://security evaluators.com/knowledge/presentations/soho_defcon21.pdf

John, EH; Rajeev, M; Ullman, JD. 2007. *Introduction to Automata Theory, Languages, and Computation*, 3rd Edition. Pearson.

Kumar, S; Dharmapurikar, S; Yu, F; Crowley, P; Turner, J. 2006. 'Algorithms to accelerate multiple regular expressions matching for deep packet inspection'. ACM SIGCOMM. New York.

Lunteren, JV; Guanella, A. 2012. 'Hardware-accelerated regular expression matching at multiple tens of Gb/s'. INFOCOM, 2012 Proceedings IEEE. Orlando.

Michela, B; Mark, F; Patrick, C. 2008. 'A workload for evaluating deep packet inspection architectures'. IEEE International Symposium on Workload Characterization, 2008. IISWC 2008 (pp. 79-89).

Nourani, M; P.Katta. 2007. 'Bloom Filter Accelerator for String Matching'. Proc. ICCCN (pp. 185-190). Honolulu.

Paxson, V. 1999. 'Bro: a system for detecting network intruders in real-time'. *Computer Networks: The International Journal of Computer and Telecommunications Networking*, 2435-2463. Retrieved from http://www.bro-ids.org/

Qi, Y; Wang, K; Fong, J; Xue, Y; Li, J; Jiang, W; Prasanna, V. 2011. 'FEACAN: Front-end acceleration for content-aware network processing'. INFOCOM, 2011 Proceedings IEEE. Shanghai.

Rafael, A; Stenio, F; Djamel, S; Judith, K; Géza, S. 2012. 'Deterministic Finite Automaton for scalable traffic identification: The power of compressing by range'. Network Operations and Management Symposium. IEEE.

Rafael, A; Stenio, F; Djamel, S; Judith, K; Géza, S. 2015. 'Design and optimizations for efficient regular expression matching in DPI systems'. *Computer Communications*, 103-120.

Rajaraman, R; Sanu, M; Vasantha, E; Ram, K; Shay, G. 2008. 'A 2.1 GHz 6.5 mW 64-bit Unified PopCount/BitScan Datapath Unit for 65 nm High-Performance Microprocessor Execution Cores'. 21st International Conference on VLSI Design (VLSID 2008). IEEE.

Ramanarayanan, R; Mathew, SK; Krishnamurthy, RK; Gueron, S; Erraguntla, VK. 2012. US Patent No. 8214414: Combined set bit count and detector logic.

Roesch, M. 1999. 'Snort - Lightweight Intrusion Detection for Networks'. Proceedings of the 13th USENIX conference on System Administration (pp. 229-238). Washington. Retrieved from http://www.snort.org

Sappington, B. 2016. Parks Associates. Retrieved from http://www.jungo.com/wp-content/uploads/Jungo%20WP_CiscoUpdate3_3-11-13.pdf

Smith, R; Estan, C; Jha, S; Kong, S. 2008. 'Deflating the big bang: fast and scalable deep packet inspection with extended finite automata'. Proceedings of the ACM SIGCOMM 2008 conference on Data communication. New York.

Song, H; Sproull, T; Attig, M; Lockwood, J. 2005. 'Snort Offloader: A Reconfigurable Hardware NIDS Filter'. Proceedings of Field Programmable Logic and Applications.

Subramanian, SS; Lin, P; Herkersdorf, A. 2014. 'Deep packet inspection in residential gateways and routers Issues and challenges'. 14th International Symposium on Integrated Circuits (ISIC). Singapore.

Subramanian, SS; Lin, P; Herkersdorf, A; Wild, T. 2016. 'Hardware acceleration of signature matching through multi-layer transition bit masking'. International Telecommunication Networks and Applications Conference. Dunedin: IEEE.

Team Cymru Threat Intelligence Group. 2016. 'SOHO Pharming'. Retrieved 26 March 2016, from https://www.doc.ic.ac.uk/~maffeis/331/TeamCymruSOHOPharming.pdf

Tuck, N; Sherwood, T; Calder, B; Varghese, G. 2004. 'Deterministic memory-efficient string matching algorithms for intrusion detection'. INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies. Hong Kong.

Wang, K; Qi, Y; Xue, Y; Li, J. 2011. 'Reorganized and Compact DFA for Efficient Regular Expression Matching'. 2011 IEEE International Conference on Communications. Kyoto.

Wesley, M; Stenio, F; Rafael, A; Djamel, S; Judith, K; Szabó, G. 2012. 'Benchmarking of compressed DFAs for traffic identification: Decoupling data structures from models'. IEEE Global Communications Conference. IEEE.

Xu, C; Chen, S; Su, J; Yiu, SM; Hui, LC. 2016. 'A Survey on Regular Expression Matching for Deep Packet Inspection: Applications, Algorithms, and Hardware Platforms'. *IEEE Communications Surveys & Tutorials*, 18(4).

Yang, X; Junchen, J; Rihua, W; Yang, S; H. Jonathan, C. 2014. 'TFA: A Tunable Finite Automaton for Pattern Matching in Network Intrusion Detection Systems'. *IEEE Journal on Selected Areas in Communications*, 1810-1821.

Yu, F; Chen, Z; Diao, Y; Lakshman, T; Katz, R. H. 2006. 'Fast and memory-efficient regular expression matching for deep packet inspection'. Proceedings of the 2006 ACM/IEEE symposium on Architecture for networking and communications systems. New York.

Yu, X; Lin, B; Becchi, M. 2014. 'Revisiting State Blow-Up: Automatically Building Augmented-FA While Preserving Functional Equivalence'. *IEEE Journal on Selected Areas in Communications*, 32(10), 1822-1833.

Zeydel, BR; Baran, D; Oklobdzija, VG. 2010. 'Energy-Efficient Design Methodologies: High-Performance VLSI Adders'. *IEEE Journal of Solid-State Circuits*, 1220-1233.

Endnotes

[i] The signature matching engines were implemented as software programs and the evaluation was performed on an Intel core i7-2600 running at 3.4 GHz with 8 GB of memory.

[ii] The original uncompressed DFA consists of 64 state transitions. However, after MSBT, the compressed DFA only consists of 23 state transitions, so a total of 41 state transitions were compressed through MSBT.

[iii] The original uncompressed DFA consisted of 64 state transitions. On the other hand, the compressed DFA after the LSCT consists of 17 compressed state transitions (7 leader transitions + 8 member transitions + 2 most repeated transitions). This results in a total of 47 state transitions compressed as part of the LSCT.

[iv] P(c) is used to define the probability of the next state being fetched from the third stage, corresponding to a character c among the sequence of M bytes in the payload.

Communications for the America's Cup Challenge

Simon Moorhead
Ericsson Australia and New Zealand

Abstract: Two historic papers from a special issue of the *Journal* in 1986 featuring the communication requirements for the America's Cup 1986/87 challenge in Fremantle.

Keywords: Telecommunications, History, America's Cup, Fremantle

Introduction

In 1986, the Society produced a special issue of the *Telecommunication Journal of Australia* (Volume 36, Number 2) featuring ten papers on the communication needs for the America's Cup challenge scheduled for 1986/87. Two of these papers have been selected for this historic review.

In 1983, *Australia II* from the Royal Perth Yacht Club in Fremantle, Western Australia, was successful in winning the America's Cup from the New York Yacht Club, ending over 130 years of possession and 24 previously unsuccessful challenges. The race history is well known, with *Australia II* coming from 1/3 down to win 4/3, and the world yachting spotlight focussed on Fremantle for the Cup's defence in 1986/87.

The first paper (Hume, 1986) provides a summary of the America's Cup history and an overview of the infrastructure and telecommunication requirements in Western Australia. Logistically, seven syndicates challenged for the cup in the US in 1983. Initially, twenty-six syndicates indicated they were serious about challenging in 1986/87, which later was reduced to thirteen by the time of the elimination races. The paper details the necessary upgrades to dock facilities at Fremantle, and provides an explanation of the "12 metre" boat dimension rule and the race program for this first Australian defence.

The second paper (Herring, 1986) discusses the infrastructure establishment, service provision, television operations and navigational equipment. The number of additional visitors to Perth was estimated at well over one million people, of which twelve percent would be overseas visitors. This influx forced major capital works to be undertaken and helped the introduction of new technology, such as debit-card public phones. All spare telecommunications capacity in the surrounding area was also brought on-line for the event.

http://doi.org/10.18080/ajtde.v6n3.160

The other papers in this special issue of the *Journal* on the communications for the America's Cup 1986/87 challenge are highly recommended for further reading and cover the following topics;

- Video and Sound Programme Network for America's Cup;
- Cabling for the Cup – A District Prospective;
- America's Cup Impact on Perth Interstate and International Telephony Traffic;
- 12 Metre Yacht Telemetry System;
- Frequency Management for the America's Cup;
- America's Cup Media Complex;
- The Official America's Cup Directory.

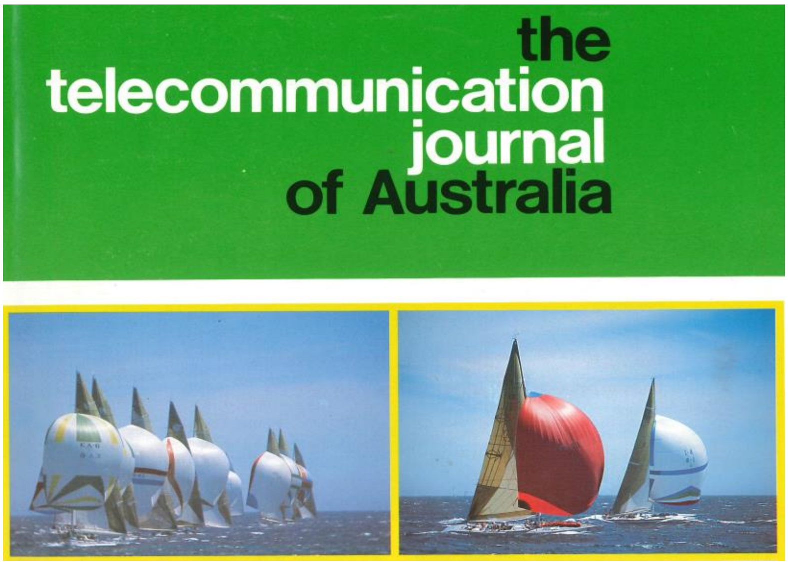
Dennis Conner skippered the *Stars and Stripes 87* yacht to a 4/0 victory over *Kookaburra III* at Fremantle in early 1987. The Cup subsequently returned to San Diego Yacht Club in America. This was the last time twelve metre yachts were raced for the America's Cup, as the rules were changed to allow for different boat designs such as multi-hull catamarans.

References

Hume, B. R. 1986. 'America's Cup History and Telecommunications Needs', *Telecommunication Journal of Australia,* Vol. 36, No. 2, pp. 3-10.

Herring, C. N. 1986. 'Communications Infrastructure and Commercial Aspects – America's Cup', *Telecommunication Journal of Australia,* Vol. 36, No. 2, pp. 11-14.

The Historic Papers



America's Cup History and Telecommunications Needs

B. R. HUME B.E. (Hons), M.I.E.Aust.

This paper briefly outlines the saga of the America's Cup and how the Twenty Sixth Challenge came to be held at Fremantle, Western Australia. The situation is considerably different to Newport, Rhode Island, both geographically and telecommunications-wise, and this latest challenge has many more challengers than any previous event. Telecom has had to develop the telecommunications facilities to meet all needs of syndicates, media and visitors.

THE ORIGIN OF THE AMERICA'S CUP

The saga of the America's Cup began in 1851, when a syndicate of members of the New York Yacht Club got together to build an advanced racing yacht called the America. They wanted to show off this yacht in England so they sailed across the Atlantic to Cowes, the home of the Royal Yacht Squadron. One thing led to another and a challenge was mounted between the America and the best yachts that the Royal Yacht Squadron could enter in the field. The race took place against fifteen English yachts, around the Isle of Wight, and the America was of such a superior construction that it beat the English yachts with ease. Thus the America's Cup was presented to the New York Yacht Club in recognition of this great event.

The letter of presentation which was sent from the Royal Yacht Squadron subsequently became what is known as the "Deed of Gift." The letter stated that "any organised yacht club of any foreign country shall always be entitled, through any one or more of its members, to claim the right of sailing a match for this cup." Over the past 130 years there have been 24 unsuccessful challenges to try and regain the cup from the New York Yacht Club.

OUTLINE OF SUBSEQUENT CHALLENGES

The first challenge was held in 1870, when the Englishmen set out to try and regain their cup, which they had lost to the America. In those days the challenging yacht had to sail across the Atlantic Ocean to challenge the New York Yacht Club at New York. These challenges were then held at irregular intervals up until 1980 and in total England mounted 14 challenges, Scotland 1, Canada 2 and Australia had 6 attempts to wrest the cup from the New York Yacht Club.

Many changes were made to the "Deed of Gift" from time to time as conditions varied and yachting designs advanced. The major change was that it was not necessary for the yacht to sail itself to New York.

Some highlights of the Challenges:

In 1930 the 14th challenge was held and the location of the races was shifted to Newport because of the congestion of shipping in the vicinity of New York.

In 1958, the 17th challenge, the class of yacht was changed to the 12 metre class for economic reasons. Over the previous challenges the yachts had become bigger and bigger and much more expensive and it was decided to rationalize to the 12 metre class.

In 1962 the 18th challenge was held and this was the first one where Australia challenged with the yacht Gretel.

In 1970, for the 21st challenge, trials had to be arranged between the foreign challengers to select the one to meet New York Yacht Club's defender. This was the first time that more than one challenger had made a simultaneous challenge and this time the Australian Yacht, Gretel II, beat the French entry for the right to make the challenge.

THE 1983 SUCCESSFUL CHALLENGE

The 25th challenge, held in 1983, was notable for the increased number of challengers. In fact, there were seven yachts entered from five nations so that an extensive pre-selection series of races was needed. The Royal Sydney Yacht Syndicate was selected as the Challenger of Record and many races were conducted over a 3 month period at Newport.

The yacht chosen to challenge the New York Yacht Club was Australia II, entered by the Royal Perth Yacht Club and owned by a syndicate headed up by Alan Bond.

There were four yachts entered to become defender of the cup for the New York Yacht Club. Over the same 3 month period trials were conducted by the New York Yacht Club to pick the best defender and finally Liberty, skippered by Dennis Connor, was selected.

The history of the challenge is well known, with Australia II being down 1/3 after the first 4 races and then coming back to level the score at 3 all. In the tradition of America's Cup there was plenty of controversy during these races. The final race was then won by Australia II to make it a 4/3 victory and the public and media attention was aroused as never before. It was realised that the New York Yacht Club had finally lost the cup.

REACTION IN WESTERN AUSTRALIA

The Royal Perth Yacht Club suddenly found themselves to be the possessor of the America's Cup and immediately set about preparing for a new challenge series. An avalanche of challenges was received by the Club and, at the time of paying the first deposit, there were 26 syndicates that said they were serious about challenging for the cup.

The W.A. State Government and the Royal Perth Yacht Club quickly decided that there was insufficient accommodation to cater for such a large number of challengers and a new marina was put on the drawing-board. With the large interest being generated over the America's Cup, it appeared that Perth and Fremantle were in for a tourist invasion and careful preparation was needed to ensure that the 26th Challenge would be a great success.

It was realised very early by Telecom that extensive telecommunication facilities would be needed to support the 26th Challenge for the America's Cup. Early discussions were held with people who had been at Newport at the time of the successful challenge to assess the communication provisions and shortages at that time. These discussions showed the need to prepare a very extensive media centre to house a large number of media personnel who could be expected to cover the event. Similarly, television could be expected to place heavy demands on the infrastructure. To assess the situation at first hand the author visited Newport late in 1984.

THE SITUATION IN NEWPORT, RHODE ISLAND

Newport is about 200 km from New York, 50 km from Providence, the capital of Rhode Island, and in summer is the centre of a very popular yachting season. Several oceanic events start from Newport, including the famous annual Newport-to-Bermuda race. Newport is located on Narragansett bay, which houses 25 individual yacht clubs. The America's Cup challenge is just one of a series of racing carnivals to be held on Narragansett Bay during the summer.

Newport is historically orientated as many of its buildings date back to the 17th Century. The main street, Thames Street, feeds the shipyards, where the 12 metre yachts were docked and the area is so compact that it is ony 100 m from the pens to the main street. The media centre was located in Thames Street among the shipyards so that it was a short walk between any of the points of the waterfront. As the crowds built up Thames Street had to be closed to make it available for pedestrians.

Newport has hosted several challenges, which have all been very one-sided affairs. As the event has grown, Newport has been able to accommodate it with little effort. The seven challengers in 1983 were located within the existing harbour facilities. **Fig 1** shows the typical berthing arrangements which were temporarily converted for syndicate use. The large restaurant "Christies" was a favourite haunt for Australian supporters.

The local telephone company provides telephones at Newport. For the 25th Challenge they installed additional lines and telephones in the harbour area and in the media building. Telex services are provided by Western Union International using cable pairs provided by the local telephone company. Apart from these additional lines and augmentation of the circuits from Newport back to

FIG. 1. Waterfront at Newport near Australia II berth.

Providence, the capital of Rhode Island, there was little work needed by the telephone company. At that time there was no International dialling and all calls had to be placed through an operator. There was no paging or mobile telephone service available.

The major television networks in the United States largely ignored the America's Cup series. The local television station provided outside broadcast links to a boat, helicopter and the blimp and a video link from Newport to Providence. This station provided live coverage during the final series.

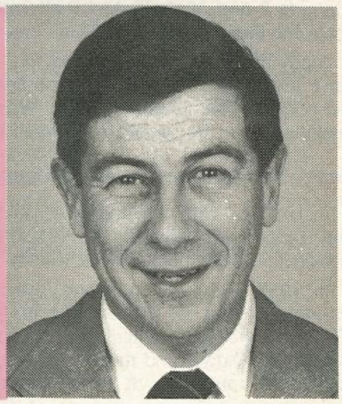
The only radio broadcast coverage was provided by the local Newport Radio Station. They used a reporter on a boat following the race and provided frequent updates on the contestants' positions during their normal programme.

The arrangements outlined above worked well during the elimination series of races and the first few races of the final seven. However, when Australia II won its second and then its third race there was a totally unprecedented increased interest in following the event. Suddenly Newport was deluged with media representatives who wanted to cover the final races and television stations were demanding live coverage of the event. Only the NBC network was able to get the live coverage because it has affiliations with the local television station in Providence. The local radio company was also able to sell its coverage of the America's Cup extensively throughout the United States, using dial-up STD circuits.

The overall impression of Newport was that they were not prepared for the huge increase in interest during the latter races. To avoid this situation at Fremantle,

BRIAN HUME has worked in PMG/Telecom since he graduated from the University of Western Australia with a Bachelor of Engineering, majoring in Communications in 1957. He has worked in a variety of positions including Country operations, Trunk Service, Exchange Installation and Planning.

In 1983, when Australia won the cup, he was Superintending Engineer of the Planning Branch, Western Australia. Here he controlled the planning of telecommunications facilities for the twenty-sixth defence to provide the basic switching, cabling and transmission infrastructure. In December 1985 he was promoted to the position of Regional Engineer, Metro Region where he now also controls the installation of the engineering plant to support America's Cup telecommunications.



telecommunications facilities would have to be augmented considerably.

COMPARATIVE SITUATION AT FREMANTLE

Fremantle is 16 km from Perth at the mouth of the Swan River and is the major port for Western Australia. It is also an historical site, being the point where the first European settlers landed here in 1829. The aim was to develop America's Cup facilities in a compact location to allow people to walk between the various facilities.

It became apparent that the Fishing Boat Harbour would be able to provide some facilities for the visiting syndicates. In fact two syndicates, Bond's Australia III and the New York Yacht Club, established themselves in this harbour at the end of 1983. Not all of the visitors could be accommodated in this area, so an additional marina called Challenger Harbour was constructed. **Fig 2** shows a panoramic view of the harbours during preparations for the elimination series.

The most suitable existing building for the media complex proved to be a hall owned by the Fremantle Port Authority **(Ref 1).** This is located about 400 m from the new Challenger Harbour towards the main Fremantle Harbour. **Fig. 3** shows the layout of the area and the location of the significant points together with berthing arrangements for the syndicates. The theatrette in the Fremantle Port Authority building was also designated as the media interview room.

The area selected is not as compact as Newport. This is to be expected as the Fremantle challenge will be a much larger event involving many more syndicates. There are several advantages at the Fremantle location as compared to Newport.

- the site is free of main streets and traffic will not be a problem;
- there is a railway skirting the area to provide access for large numbers of people without the necessity of providing parking space;
- the access from the marina to the America's Cup course area is very direct for the competing yachts;
- the main deep water port in Fremantle for overseas ships is within walking distance;
- the area is more open and will allow larger groups to be accommodated.

Telecom Australia is in a much better situation to provide facilities for this event as it provides all types of communication facilities and is able to plan the complete facilities for telephones, telex, facsimiles, data, TV bearers, radio programme lines, mobile telephones and radio paging. The early selection of the basic plan enabled Telecom to move into the construction phase of the new facilities.

ASSESSMENT OF TELECOMMUNICATION REQUIREMENTS

Because of the long lead times to purchase telecommunication equipment, early advice of the requirements was needed. This proved impossible to obtain as those involved had not reached that stage of their planning. From discussions it appeared that the syndicates and the media complex would be the main focus for telecommunication efforts. The author's overseas visit included other locations which were related because they were all sporting events which aroused worldwide interest. The experiences of overseas organisations had then to be translated into what would be required in the Fremantle situation.

All the early planning work by Telecom was done by the Engineering Department as an extension of normal network development activities. The increased activity at Fremantle will require additional telephone services, causing greater traffic loads on connecting junctions. An additional 1000 lines of subscribers capacity was provided, together with augmented cable to the marina area and the media complex. Junction relief was a more complex problem requiring the earlier installation of an optic fibre cable, containing 10 fibres, from Perth to Fremantle. This cable is now providing circuits between the Wellington and Fremantle exchanges on a 140 Mbit/s bearer. Additional switches have been provided at coastal

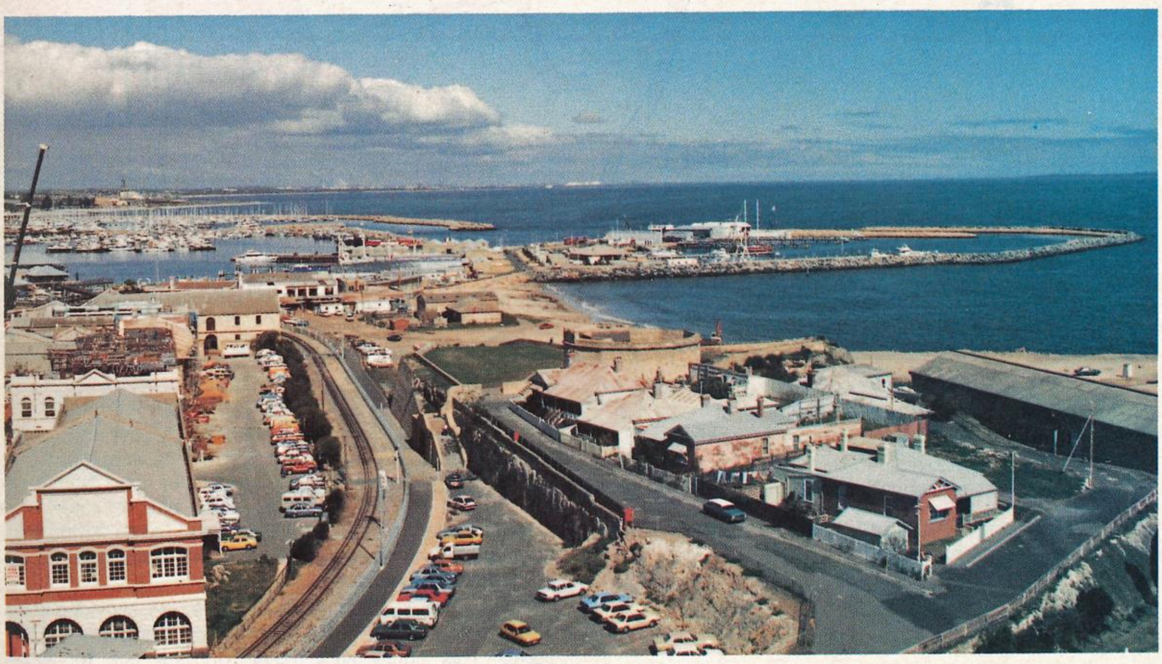


FIG. 2. View from Fremantle Port Authority Office of the harbours for the America's Cup.

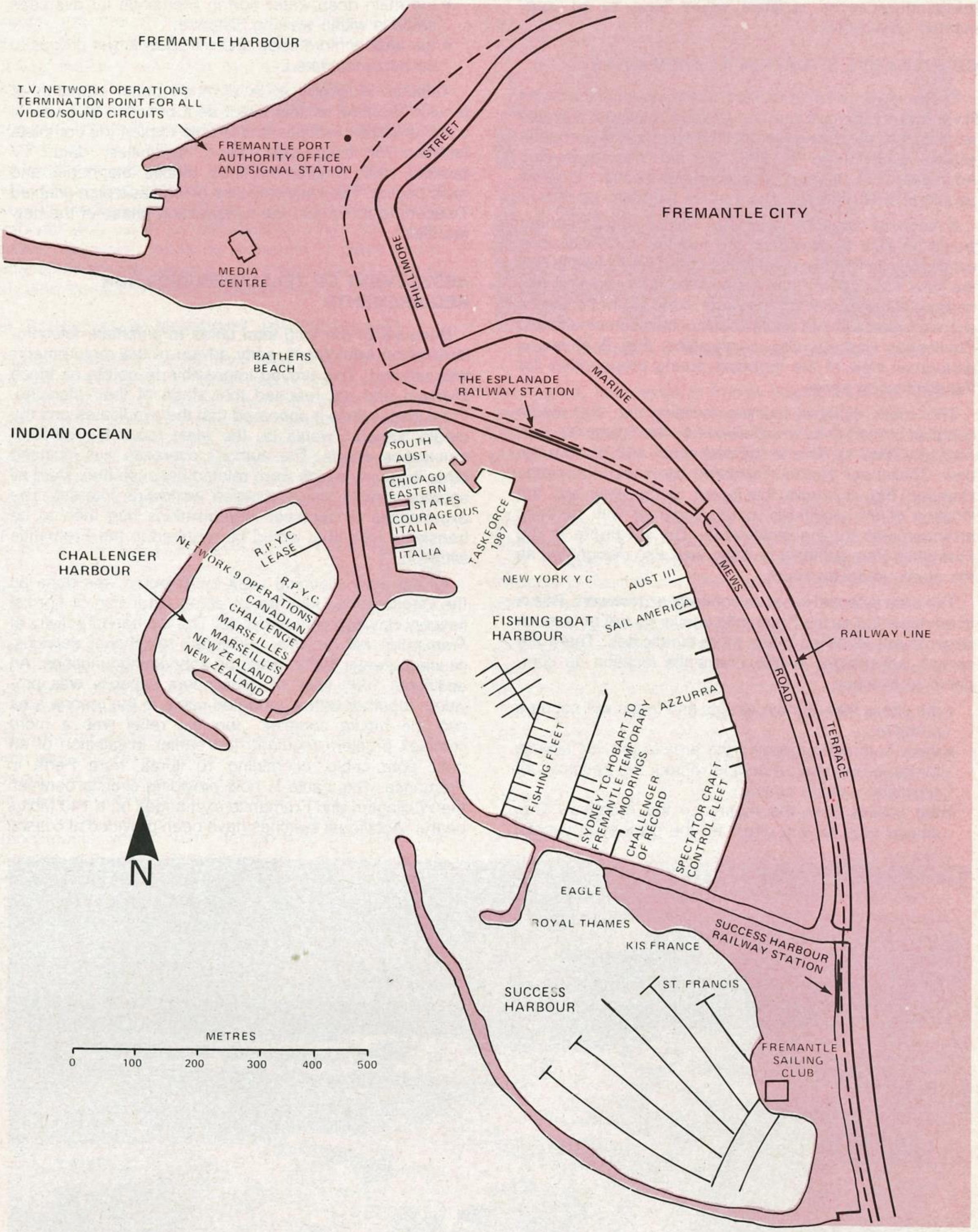


FIG. 3. General Layout at Fremantle, showing America's Cup Syndicate Sites.

telephone exchanges north of Fremantle to cater for the expected tourist influx.

The design and provision of facilities in the media complex was a special project and Reference 1 in this issue gives a description of this work.

The additional traffic load is expected to have a heavy interstate and international bias. We have planned to augment these routes by installing the 1987 normal circuit requirements by October 1986.

To meet the need for television coverage, six video links from Fremantle to Perth were planned using spare optic fibres. This figure was based on one link for each Australian network, with two for itinerant overseas programmes. **Ref 2** shows how this early estimate has been augmented to meet later requirements.

Radio broadcast circuits were planned to be provided over 2 Mbit/s links from Fremantle to Perth. Again, the early requirements have had to be augmented as in **Ref 2.** This increased circuit loading has considerably exceeded the capacity of the Television and Radio Operating Centre which has consequently been re-arranged.

To coincide with the Cup, the Commonwealth Government brought forward the completion date of the new Perth International Airport building to October 1986. This required another optic fibre solution as the new building is remote from any telephone exchange. **Ref 3** gives more details.

Other increases have been made to the Mobile Telephone Service (500 customers), public telephones (coin and card), paging and telephony links to cruise ships which are expected to visit during the Cup period.

When the first syndicate arrived there was a necessity to set up a full-time Commercial Manager position to establish and maintain contact with all parties involved in the challenge. The occupant is Mr C. Herring, who has ensured that all Telecom groups coordinate their efforts to meet the needs as they arise **(Ref 4)**.

NUMBER OF SYNDICATES

As mentioned earlier there were initially 26 challenge syndicates. Our advice was that these would probably reduce to about 16. This has occurred and the withdrawals so far have brought the number down to 14 challenging syndicates. The Challenger of Record is the Costa Smeralda Yacht Club of Sardinia and they will be organising the elimination series to determine the best yacht to be the challenger. The large number of races required will need two or three race courses to be used simultaneously. These elimination series will provide plenty of match sailing experience for the challengers.

A similar story was experienced with the number of defenders which started off about 10 and has now been reduced to four. The Royal Perth Yacht Club will be conducting an elimination series to pick the best Australian boat to be defender. This series will be conducted at the same time as the challenger series. Whichever Australian boat is selected as the defender will be representing the Royal Perth Yacht Club and will seek to retain the Cup for that Club.

CONCLUSION

The 26th America's Cup Challenge is by far the largest in its 136 year history. It has grown to be a very prestigious event and the large entry will attract visitors from all around the world. The media coverage must be of a very high standard to satisfy the interests of people in many countries.

This article outlines Telecom Australia's preparations for communications for the America's Cup series. The advanced planning work has allowed ample infrastructure to meet the communications needs. The other papers in this edition of the Journal give more details of the facilities provided.

REFERENCES

1. Herring, C. (1986), **"America's Cup Media Complex,"** Telecommunication Journal of Australia, Vol 36, No 2, 1986.
2. Norman M. J. (1986), **"Video and Sound Programme Network for the America's Cup,"** The Telecommunication Journal of Australia, Vol 36, No 2 1986.
3. Norman M. J. (1986), **"Perth International Airport — Telecom Planning and Installation Aspects,"** The Telecommunication Journal of Australia, Vol 36, No 2 1986.
4. Herring, C. (1986), **"Communications Infrastructure and Commercial Aspects — America's Cup.** Telecommunication Journal of Australia, Vol 36, No 2, 1986.

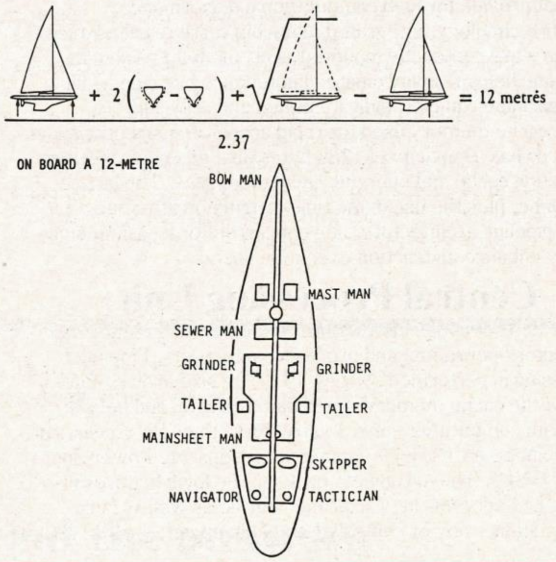
The 12 Metre Yacht

WHAT IS A 12-METRE?

It is a yacht which complies in every respect with the requirements regarding construction and equipment contained in the Deed of Gift and the Interpreting Resolutions applying to national origin of design and construction.

Bilges shall be kept as reasonably dry as possible while racing. No device shall be fitted or employed which would permit the tilting of the mast athwartship.

This class of racing sailboat is based on a mathematical formula which takes account of hull length, skin and chain girth, sail area and freeboard. When formula values are summed and divided by a mathematical constant the resultant rating (12 metres or 39.37 feet) should result.



FIRST AUSTRALIAN DEFENCE THE AMERICA'S CUP 1986-87 WESTERN AUSTRALIA

RACE PROGRAMME

13 CHALLENGERS – 6 COUNTRIES

LOUIS VUITTON CUP

ROUND ROBINS

ROUND ROBIN 1 OCT 5-20
ROUND ROBIN 2 NOV 2-19
ROUND ROBIN 3 DEC 2-19
Each yacht will meet the others once.

LOUIS VUITTON CUP SEMI-FINALS

DEC 28-JAN 7, 87
The four top-scoring yachts will enter the semifinals.
The first yacht to win four races.

12 Louis Vuitton Cup

LOUYIS VUITTON CUP FINALS

JAN 13-23, 87
The winner of the Louis Vuitton Cup is the one who first wins four races. He is selected to fight against the best defender in the final match.

4 DEFENDERS – AUSTRALIA

DEFENDER'S CUP

SERIES

SERIES A OCT 18-29
SERIES B NOV 10-23
SERIES C DEC 06-21
Each yacht will meet the others twice.

SERIES D
JAN 10-16, 87

DEFENDER'S CUP FINALS

JAN 18-25, 87
The two top-scoring yachts enter the finals. The winner defends t4he America's Cup against the Louis Vuitton Cup's winner.

AMERICA'S CUP

CHALLENGER ⟷ DEFENDER
JAN 31-FEB 15, 1987
The first yacht to win four races.

Communications Infrastructure and Commercial Aspects — America's Cup

C. N. HERRING

An event of the magnitude of the 1987 America's Cup Defence places great reliance on a wide range of quality communications services. Their operation and efficiency of connection will be closely appraised by many people and to, an extent, Australia will be judged on the outcome.

But much of the America's Cup story is about money and the returns on investment which other Country participants would anticipate through a successful association with an individual challenger or the event itself. Never before has Western Australia had such an opportunity to show itself to the world. The organisation arrangements for the event, facilities, tourist attractions and other activities all play a part in determining the long term benefits to both the State and Australia.

COUNTRY

The America's Cup is now a major international event. If it were on the world market for competitive bids in the same manner as the Commonwealth or Olympic Games, cities and nations would spend millions of dollars to try to attract it and hundreds of millions to service it. Beginning life as a leisurely yachting regatta 135 years ago, the Defence of the Cup has now become a contest between Nations generating intense interest in sport, travel and technology.

An event of this magnitude demands and places great reliance on a wide range of communications services. In this regard, Telecom, the National carrier and major provider and manager of telecommunications within Australia, has established an infrastructure to meet the resultant communications demand, a demand which will come not only from yachting syndicates but also from the many sponsors and organisers, the hospitality industry, entrepreneurs, visitors to Western Australia and the media.

INFRASTRUCTURE ESTABLISHMENT

In planning for the Cup events, Telecom actions have been guided by adoption of the following principles —

- Any decision to undertake major works was strongly influenced by their potential for post-Cup community use.
- The application of standard charges to the provision of telecommunications services for yachting syndicates, sponsor operations, the media, etc.
- The charges for provision and extension of underground cables to locations beyond those normally serviced by Telecom should be at commercial rates.
- Non-standard work undertaken on a discretionary basis at a customer's request should be charged at commercial rates.
- The opportunity should be taken to promote and advertise Telecom's role in support of this major event.

Towards the end of 1983 (and soon after Australia II's success at Newport) funds were requested by Telecom WA for inclusion in future years capital works programmes, especially for the Challenge and related Cup events. At the same time, a letter to the State Government offered Telecom's support and expertise in assisting with the staging of the 1987 Defence. Late 1984 saw the first major expenditure of funds as work to upgrade telephone exchanges was started and external plant crews commenced laying ducts and cables to service 12 metre syndicate operations in the Fremantle Fishing Boat Harbour, Challenger Harbour and the Fremantle Sailing Club. Attention was also directed to the needs of the media at this time and planning, in conjunction with the Fremantle Port Authority, to upgrade facilities along Victoria Quay (which alone will become home for some 10,000 people living in luxury liners) was undertaken.

At the beginning of 1985, Telecom devoted considerable attention to supporting both the City of Fremantle and business operations in their plans to upgrade facilities, accommodation and other developments in the Fremantle area. The City commenced an extensive programme of footpath replacement and construction, together with roadworks and drainage projects. Business groups became involved with housing redevelopments, hotel upgrading and other activities. In almost every case Telecom underground plant was affected. These early works, and others, were completed at Telecom expense to provide a significant part of the initial establishment of an America's Cup communications infrastructure.

SERVICE PROVISION

As overseas yachting syndicates arrived in Fremantle with their plans for headquarters and dockside facilities, commercial negotiations were quickly completed to ensure almost immediate provision of telephone, telex and data services. In many cases, cable extensions into transportable accommodation were required and special facilities such as locks and private metering equipment fitted to telephones to control usage.

Having upgraded telecommunications including all internal cabling at the Royal Perth Yacht Club (RPYC) in Perth, attention was then directed to the needs of the newly established RPYC annexe located in the midst of activities in Challenger Harbour.

This venue, being the Race Central Headquarters, requested a flexible communications plan to cater for not only the yacht racing, but also for activities in up to 22 hospitality areas and to service 3 associated jetties with 60 telephone points.

In conjunction with the Club management and architects, cabling was extended from the clubhouse annexe to locations within the area of lease to service the 'Wimbledon style' marquee accommodation for corporate bodies, and to the three jetties for provision of a range of immediate communications options to boats on both a temporary and semi-permanent basis. Other locations serviced included the jury, and measurers and harbour masters transportables, these being positioned some distance from the clubhouse building to ensure a degree of visual separation.

The annexe communications required the inclusion of a small business system, telex, facsimile, data lines, sound circuits for the host radio broadcaster, links to the Media Centre PABX Challenger of Record Operations and the 17 competing yachting syndicates. Seiko timing signals and printouts of progress, and IBM history/results information is also available at the annexe.

Estimates have indicated that, over the five month period October-February, the Cup would be expected to attract to Perth an additional 119,000 WA Country visitors (over and above the normal 512,000 for the period), an additional 304,000 interstate visitors (2.2 times the normal 133,000) and an additional 70,000 international visitors (over and above the normal 76,000 international visitors for the period).

In total, the number of people expected to visit Perth during the America's Cup 5 month period is 1,216,000 of which some 12% will be visitors from overseas **(ref. 1)**.

Plans to cater for this 'visitor market' included:—

- Special arrangements through real estate and travel agents to ensure the provision of a telephone service on the day of arrival, dedicated to visitors renting private accommodation.
- The installation of additional pay phones at locations attractive to tourists, including the provision of more services to cater for an anticipated significant increase in the South Australia to Western Australia Eyre Highway vehicular traffic.
- A field trial of a magnetic strip debit and credit 'Card Phone' timed to coincide with the Cup racing period in WA as part of a National technical/market evaluation of this product.
- An increase in the number of interstate circuits for STD/ISD connections and the retention by OTC of additional international satellite circuits, provisioned for Christmas traffic, until after the final Challenge match in February.

TELEVISION OPERATIONS

Elsewhere in this journal, attention has been directed to the facilities and operation of the Media Centre which largely caters for the press (including photographers) and radio broadcasters **(ref 2).** In this regard, most requests for service accompanied accreditation documents or were negotiated in conjunction with media centre management.

For television operations, commercial negotiations were conducted separately from other media activities and with a number of distinct groups namely —

- A consortium comprising representatives of four Australian networks whose role is to provide a host signal initially of limited duration increasing progressively to full time coverage of the final, best of seven, 1987 Defence races.
- The International Management Group (a division of the McCormack organisation) appointed by the RPYC to market the Cup.
- Australian networks who, as individuals, established their own operations and entered into arrangements with overseas networks.
- Overseas networks during their investigations and negotiations with production companies and Australian networks in their quest for local support for their operations.
- The Overseas Telecommunication Commission in the establishment of the Perth International Terminal at Gnangara.

These negotiations carried with them an element of commercial risk as, due to the need to forward order video equipment and special cable, and complete external plant work at an early stage for telephone services, decisions on television circuit requirements were often anticipated in advance.

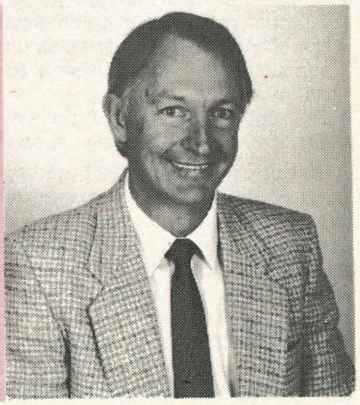
Nevertheless through the utilisation of spare optic fibre cable capacity between Fremantle and Perth and the efficient use of existing underground ducts coupled with the interstate loan and reuse after the Cup of video radio links and terminating equipment, Telecom was able to effectively satisfy the majority of network requests. (Ref 3).

To ensure full availablity of bearers in allowing for conditions of yacht racing which could see events delayed, postponed or require increased coverage through moments of vivid interest, local video links were arranged under a special 19 week full time lease agreement. This arrangement also allowed individual network operation staff to carry out line patching at Fremantle, much of which will be carried out 'after hours,' as required.

OTHER ACTIVITIES

The defence of the Cup will attract to Perth the biggest collection ever of the world's leaders in commerce and industry to assemble in Australia. Most sponsors of yachting syndicates are sending high-powered teams to Western Australia for at least part of the Cup period.

COLIN HERRING is Manager — America's Cup (Commercial) Telecom Western Australia. He joined the then P.M.G.'s Department in 1952 and has previously been involved with Sub Professional Training and the Common User Data Network (C.U.D.N.) project. More recently he has held various positions in the Operations Department organisation. Since commencing on the America's Cup project in December 1984 Colin has been closely involved with the Western Australian State Government, Royal Perth Yacht Club and other organisations providing communications consultancy and support initially for the staging of the World 12 Metre Championships in February 1986 and presently for the first Australian defence of the America's Cup to be sailed off Fremantle early in 1987.



Debit/Credit card telephones which will be trialled in Western Australia during the period of the America's Cup are expected to be popular amongst visitors and tourists.

In addition to requiring personal accommodation, many businesses are seeking locations to display and market their respective products and services and to promote these through media liaison. Since early 1985 there has been a constant demand for Telecom to provide communications advice and service to these groups.

The large numbers of business and professional conventions being held during the Cup period have added to Telecom's workload. Two concurrent events that will highlight Western Australia's economic importance will be the Perth America's Cup International (PACI) Exposition to be held in Fremantle in November 1986 and PacRim '86, an international symposium of finance, trade and investment interests centred on the Pacific Rim region to be held in Perth November 17 to 19. Each of these events demands international communications of world standard.

Additionally, the America's Cup Festival of Sport, September 19 to February 20, which includes a world boxing contest at the new $280M Burswood Island Casino, a $1,000,000 horse race and other entrepreneurial activities have required forward planning to ensure provision of adequate communications facilities.

The mini-America's Cup, usually conducted at Cowes on the Isle of Wight during the Admirals Cup, and sailed in scaled down 5-metre repiicas of 12-metre yachts will, in 1987, be sailed off Bathers Beach, Fremantle between January 27 to 29. The Beach will be an ideal amphitheatre for the mini-Cup because it has vantage spots along the South Harbour and Challenger Harbour. It will also be popular for spectators wishing to view the departure and return of the 12-metre boats on both race and lay days.

Anchorages and moorings between Rockingham, 20 nautical miles south of the America's Cup course and Hillarys marina at Ocean Reef, just north of the race area will 'park' an armada of spectator vessels ranging from 10-metre sportsfishing launches to 70 metre floating palaces, all visiting to see the competition. For the larger vessels private charter companies are providing victualling, fuelling, sewage disposal and commuter transport services with computers being used to co-ordinate the many needs.

Overhead a 59 metre helium-filled airship will be employed as a television camera platform and for airborne advertising. Other airspace will cater for up to 40 helicopters hovering over the courses with fixed wing aircraft operating above.

THE BOTTOM LINE

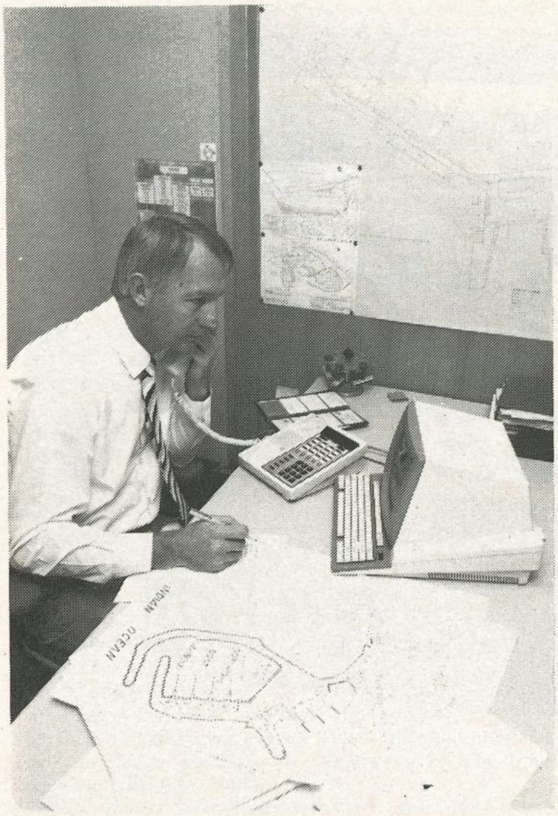
It will cost the Royal Perth Yacht Club (through its trust company ADAC — Australia's Defence America's Cup 1987 Pty Ltd) some $4M to stage the first Australian defence of the Cup.

The majority of this money will be provided by major sponsors and companies who have signed agreements to use sponsors' and licencees' symbols, along with the sale of television and radio broadcast rights and 'Official Supporters Hospitality Packages.'

The Federal Government gave a direct grant of $30M and the capital cost to the WA State Government for preparations is expected to be about $54M.

Much of the Government expenditure, however, can be classified as payment for improvements in the social infrastructure of the region with a useful life well beyond

http://doi.org/10.18080/ajtde.v6n3.160



Colin Herring, Telecom's Manager, America's Cup is shown at work handling the commercial aspects of the event.

the Cup period. Only a little can be identified as America's Cup specific.

Similarly, Telecom Australia with an overall Cup budget of some $8.3M has constructed World class communications services which will provide long term benefits to the community.

By the standards of other 'hallmark events' the level of public sector commitment to stage the Cup Defence is relatively modest. Compared to Olympic and Commonwealth Games and to Expositions, which require very significant commitment of Government resources for an infrastructure which is arguably not well utilized after the event, the cost of $7M to construct the major Cup venue (Challenger Harbour) is minimal.

The previously mentioned 'America's Cup — Economic Impact' study report says the defence could inject at least $600M into the Western Australian economy and up to $1.117 billion. The spending could create some 14,400 jobs over 12 months, half of these resulting from visitor spending.

Finally, even if the America's Cup is not retained, the 1987 Series will have a substantial positive impact on the economy through promotional and trade opportunities flowing from the staging of the event.

References

1. 'America's Cup — Economic Impact' report to Western Australian Government Centre for Applied and Business Research, the University of Western Australia.
2. Herring, C. 'America's Cup Media Complex,' Telecommunication Journal of Australia, Vol. 36, No. 3, 1986.
3. Norman, M. 'Video and Sound Program Network for the America's Cup,' Telecommunication Journal of Australia, Vol. 36, No. 3, 1986.

Information Transfer News

Navigation Accuracy for the America's Cup

The Royal Perth Yacht Club are taking no risks with the placing of the navigational marks for the America's Cup races. They have appointed a Perth firm, Seatronics Pty. Ltd., as official suppliers of navigational equipment, and all mark laying will be done with the aid of a Syledis navigational system.

The difference between Syledis and other navigation systems is that the others tend to work on a range to range mode, whereas Syledis works on hyperbolic mode as well. To put that into simpler terms, with a range to range navigation system, the unit on the boat sends out a signal to a series of fixed beacons and times how long they take to come back, then by a simple triangulation works out the position. The Syledis has what is called a passive receiver on the vessel that just receives signals from other beacons and that is done by a synchronised set of signals that come from a master beacon. By calculating the time dfference in a hyperbolic fashion, the unit can work out precisely where it is in relation to those beacons either on a grid or geographically. Because Syledis uses a passive receiver, it has unlimited user capability, whereas most others can only handle about four users at once.

Syledis, a French made system, is used all over the world. There are chains of beacons in the Middle East, all around the coast of Britain and in Australia. Since putting up the chain off Fremantle for the America's Cup, it has been found that other people besides the yachtsmen have a use for it. The Marine and Harbours Department have a definite need for very precise positioning, if for example they are called out in the middle of the night to rescue someone, it is very comforting for them to know exactly where they are to within a metre or so.

The navigators on the yachts will not only be able to tell exactly where they are, in latitude and longitude, there will be a graphic display on the screen that will give the yacht's position relative to the next mark, the distances up to thirty marks and the yacht's speed. The system has been tried and tested over the past months, both by the Royal Perth Yacht Club and the yachts that are here training. The system was also used in the World Championships in February, 1986.

Australian Journal of Telecommunications and the Digital Economy, ISSN 2203-1693, Volume 6 Number 3 September 2018

http://doi.org/10.18080/ajtde.v6n3.160