Journal of Telecommunications and the Digital Economy

Volume 12, Number 4
December 2024

Published by
Telecommunications Association Inc.

ISSN 2203-1693

The *Journal of Telecommunications and the Digital Economy* is published by TelSoc four times a year, in March, June, September and December.

Journal of Telecommunications and the Digital Economy

Volume 12, Number 4
December 2024

Table of Contents

History of Telecommunications

Editorial Team

The *Journal* is published by the Telecommunications Association (TelSoc), a not-for-profit society registered as an incorporated association. It is the Australian telecommunication industry's oldest learned society. The *Journal* has been published (with various titles) since 1935.

Editorial

Social Media and Identifiers

Leith H. Campbell
Managing Editor

Abstract: This editorial offers some observations about recent moves by the Australian Government to restrict access to social media by young people. It also outlines the other content of this December issue and notes some personnel changes in 2025.

Keywords: Editorial, Social Media, Age verification, Identity management

Social Media and Age Restrictions

The Australian Government plans to restrict access to social media for children under the age of 16 – a proposal that has garnered international media attention. The relevant law has been passed, but the details of how these new restrictions will operate are yet to be published. Nor have the relevant social media platforms been specified, but the basic definition is of those services whose "sole purpose, or a significant purpose, of the service is to enable online social interaction between 2 or more end-users" (Australian Government, 2024b, clause 63C, (1)(a)(i)).

What is the purpose of such a restriction? The Australian Human Rights Commission, while expressing "serious reservations" about the new law, suggested four reasons for restricting access by children: "Protection from Harm"; "Promoting Healthy Development"; "Addressing Online Privacy Concerns"; and "Supporting Parents" (AHRC, 2024). The Prime Minister has been quoted as saying that there is a "clear, causal link between the rise of social media and the harm [to] the mental health of young Australians" (Sullivan, 2024).

There has been some reporting that this is an Australian "world first". However, this is not so, as the world now knows after the CEO of TikTok, Chew Shou Zi, a resident of Singapore, appeared before a US Congressional Committee ("5 key moments", 2024). Singapore restricts access for the under-13s. The Singaporean experience is instructive (Chia, 2024). In

https://doi.org/10.18080/jtde.v12n4.1191

Singapore, TikTok bars users under 13 but does not ask for proof of age. TikTok, Instagram and Snapchat have parental-control features for child accounts that accept parental supervision. Facebook, Twitter, Reddit and Discord have no direct parental controls but are, in theory, only available to users over 13 years. One may rightly be sceptical about the effectiveness of these controls.

The Australian Government's approach to age verification and restriction is not yet known, but it has awarded a tender for a trial:

> The trial will examine age verification, age estimation, age inference, parental certification or controls, technology stack deployments and technology readiness assessments in the Australian context. It will invite Australians to participate in testing these different age assurance solutions in a live environment (Australian Government, 2024a).

The trial is due to be completed by the middle of 2025, in order for regulations to be in place by the end of the year.

The Government and most commentators have focussed on what social media platforms will or should do. Given (2024), for example, says that "tech companies" should have a "digital duty of care". She says that "social media platforms should be safe spaces for all users" and that the "onus is now on the tech companies to restrict access for youth under 16".

While the social media platforms should, can, and probably will implement some age-restricted access methods, they cannot provide the whole solution. Truly verified age belongs to the individual and to the government. In the long run, if age restrictions are to be fully effective, age will need to be verified by a trusted third party.

In the immediate term – and perhaps in the long term – there are potential solutions outside the social media platforms. The trial of age-verification methods includes "technology stack deployments" (Australian Government, 2024a, quoted above), whatever they are, but it could provide an opportunity to look more widely at the "technology stacks" used to access the platforms.

The telecommunications companies could be the trusted third party: they actually control access to the Internet and the World Wide Web. Each access point has an "owner", the person or organisation that is billed for the access. This owner could be given controls to determine who could use the access; the telco would be responsible for verifying the credentials of each user. Once admitted, the user would have access to all resources appropriate to their age and/or other characteristics. In essence, a mobile phone locked to a single user already provides the basis on which the required verification could be built. On a fixed access where there are potentially multiple users, more personalised control would be required.

https://doi.org/10.18080/jtde.v12n4.1191

Importantly, each access should retain a "guest" or "anonymous" option to open the Internet and Web to an unverified user. This would permit normal web browsing (but the search engines may restrict content) and access to sites that do not require verification. In addition, for added security, a bank, for example, may not accept a login to its website from an anonymous user, but only permit verified network users to login (with the usual two-step verification) to its website.

In addition, verified age is of value to all social media platforms, whether age-restricted or not, and to the advertisers that are their paying customers. This opens up a possible new revenue stream from the platforms to the telcos to defray the cost of implementing and maintaining strict access controls.

Telecommunications companies are used to dealing with regulators and there would be nothing new, in broad terms, in handling regulation of verified access. It would involve new processes and protocols, all of which would take some time to be discussed, trialled and agreed. This would not necessarily meet the government's preferred timetable.

Whatever the outcome for the present case of age restriction, it is important for all stakeholders to recognise that the Internet and Web are not working as originally envisaged and that verified identity in some form (Campbell, 2024) will be necessary. A consistent, systemic solution will be found not from the social media platforms alone, but will include the telcos and, potentially, other stakeholders.

In This Issue

As usual, we cover a wide range of topics in this issue.

We publish a Special Interest paper – the Charles Todd Oration 2024 presented by Will Irving; and the question-and-answer session that followed – entitled *If At First You Don't Succeed, …* This will be particularly relevant to readers interested in the history leading to Australia's National Broadband Network and the future of universal service.

In the Public Policy section, we publish one paper, *The Impact of Fixed and Mobile Broadband Adoption on Economic Growth*, which uses panel data from 34 OECD and G20 countries.

In the Digital Economy & Society section, there are three papers. *Decoding Consumer Behaviour in Indonesian E-Commerce: A Stimulus-Organism-Response Analysis* considers the barriers to acceptance of e-commerce by consumers and looks at possible solutions. *Sharing Business Data Securely: Insights from the European Gaia-X Project on Technical and Economic Roles Enabling Federated Data Spaces* provides results from a major European research project. *Autonomous Robot Navigation System Workflow for Monitoring*

https://doi.org/10.18080/jtde.v12n4.1191

and Maintenance in Industry 4.0 Applications studies autonomous navigation combining computer vision with path-planning algorithms.

In the Telecommunications section, there are two papers. *Review and Enhancement of VoIP Security: Identifying Vulnerabilities and Proposing Integrated Solutions* describes identified vulnerabilities in Voice over Internet Protocol and proposes a system for enhanced security. *Dynamic Touchstroke Analysis with Explainable Artificial Intelligence Tree-Based Learners* describes an explainable artificial intelligence method for touch-stroke analysis to be used for identity authentication.

We also publish two book reviews: *Data Rules: Reinventing the Market Economy by Cristina Alaimo and Jannis Kallinikos*; and *William Webb's Contrarian Thesis: A Book Review of "The End of Telecoms History"*.

The History of Telecommunications section has two papers. The first, continuing the e-commerce topic, is a reprint from the *Telecommunication Journal of Australia* in 2000, introduced as *E-Commerce Security Revisited*. The proposed publication of this reprint also triggered a further contribution, *E-Commerce Security Issues, Then and Now: Thoughts Stimulated by an Historic Paper by Dez Blanchfield (2000) on E-Commerce Security*.

As always, we encourage you to consider submitting articles to the *Journal* and we welcome comments and suggestions on which topics or special issues would be of interest. Feedback on the current issue would be welcome.

Editorial Personnel Changes

As announced in September, I am retiring as Managing Editor after this issue. The new Managing Editor is Dr Michael de Percy, Senior Lecturer in Political Science at the University of Canberra's School of Politics, Economics and Society and a long-standing Section Editor of this *Journal*.

A second notable change is the retirement of Professor Peter Gerrand from the Editorial Advisory Board after 30 years of service to this *Journal* and its predecessor. He was Editor-in-Chief of the *Telecommunication Journal of Australia* from June 1994 until its last issue in 2013; and then the first Managing Editor of this journal, published initially as the *Australian Journal of Telecommunications and the Digital Economy*, serving until the end of 2014. The length of Professor Gerrand's service is unlikely to be surpassed. More about Professor Gerrand's contributions and the wider history of the *Journal* will be published in 2025, the 90th year since publication of this *Journal* and its predecessors began.

https://doi.org/10.18080/jtde.v12n4.1191

References

Australian Government. (2024a, November 15). Tender awarded for Australian Government's age assurance trial. Department of Infrastructure, Transport, Regional Development, Communications and the Arts. Available at https://www.infrastructure.gov.au/department/media/news/tender-awarded-australian-governments-age-assurance-trial#:%7E:text=ACCS%20is%20an%20independent%20accredited,Government%20by%20mid%2Dnext%20year.

Australian Government. (2024b). *Online Safety Amendment (Social Media Minimum Age) Act 2024*. Available at https://parlinfo.aph.gov.au/parlInfo/download/legislation/bills/r7284_aspassed/toc_pdf/24150b01.pdf;fileType=application%2Fpdf

AHRC [Australian Human Rights Commission]. (2024, November 21). Proposed Social Media Ban for Under-16s in Australia. Available at https://humanrights.gov.au/about/news/proposed-social-media-ban-under-16s-australia

Campbell, L. H. (2024). Editorial: Identity is Coming. *Journal of Telecommunications and the Digital Economy*, *12*(2), iii-vi. https://doi.org/10.18080/jtde.v12n2.994

Chia, O. (2024, November 22). Amid TikTok scrutiny, how effective are parental curbs on social media? *The Straits Times*. https://www.straitstimes.com/tech/a-look-at-how-effective-parental-controls-are-on-social-media-as-tiktok-faces-scrutiny-in-the-us

Given, L. M. (2024, November 28). Australia's social media ban for kids under 16 just became law. How it will work remains a mystery. *The Conversation*. https://theconversation.com/australias-social-media-ban-for-kids-under-16-just-became-law-how-it-will-work-remains-a-mystery-244736

Sullivan, H. (2024, November 28). Australia passes world-first law banning under-16s from social media despite safety concerns. *The Guardian*, Australian edition. https://www.theguardian.com/media/2024/nov/28/australia-passes-world-first-law-banning-under-16s-from-social-media-despite-safety-concerns

"5 key moments from TikTok CEO Singaporean Chew Shou Zi's combative hearing in US Congress". (2024, November 22). *The Straights Times*. https://www.straitstimes.com/world/united-states/5-key-moments-from-tiktok-ceo-s-combative-hearing-in-us-congress

If At First You Don't Succeed, …

Charles Todd Oration 2024

Will Irving
Chief Strategy & Transformation Officer, nbn Co Ltd

Abstract: This is the Charles Todd Oration 2024, delivered by Will Irving on 17 October 2024 in Sydney. It outlines, from a Telstra-internal viewpoint, the various proposed network enhancements that eventually led to Australia's National Broadband Network. It then draws lessons for the future from this history and the current state of telecommunications in Australia. The question-and-answer session after the Oration is also summarised.

Keywords: nbn, National Broadband Network, History of Australian Telecommunications, Charles Todd Oration

Introduction

This is a near-verbatim record of the Oration. It has been edited for clarity.

Well, good afternoon. Thank you, Rob, for your very kind introduction and for acknowledging the traditional owners of the country on which we are meeting. Just as Sir Charles Todd's era has had books written about it, so **nbn**'s Chair, Kate McKenzie, has sometimes mused: "Will, you should write a book about this". But, to borrow one of Clint Eastwood's lines: "A man's got to know his limitations". So, you will be relieved to know there will be no book.

Speaking of limitations, I should stress that I am not here in any official capacity. This story and these reflections are mine, and mine alone — to aid informed debate, based on imperfect memory — and are inherently subjective. Some of the quotes you will hear I remember verbatim. Others are 'words to that effect', so I hope those whom I misquote or misunderstood will forgive me.

Today, I will take you to a few pivotal moments in the evolution of the National Broadband Network (NBN) — to help explain why things panned out the way they did — and then I will draw out some lessons for Universal Service reform.

https://doi.org/10.18080/jtde.v12n4.1184

Speaking of lessons, as a teenager, I was given a cartoon of Garfield the Cat saying: "If you're supposed to learn from your mistakes, then I should be a genius by now". In other words, it is easy to make mistakes, but often hard to learn from them. On that basis, the path to an NBN created a few geniuses — me included!

There were four failed attempts at a National Broadband Network — an 'NBN' — before today's **nbn** Co Limited came into being. Hence today's subject:

> "If at first you don't succeed, try, try again".

This phrase was popularised in Edward Hickson's *Moral Song*, republished in 1857. This was the year after Charles Todd — at age 30, South Australia's Observer and Superintendent of the Electric Telegraph — and his Victorian counterpart had jointly proposed constructing a single Overland Telegraph line from Darwin to Adelaide (and hence to Melbourne and Sydney), with Darwin connected by sub-sea cable to Singapore and the Empire.

This transformative project brought Australia a step-change in connectedness. Completed in 1872, it cost £480,000. Using comparative average weekly earnings, that is $950 million today. When you consider that Australia's European population back then was 1.8 million — 1/15th of today — that is the equivalent of $14 billion in relative national income terms. It was indeed the **nbn** of its day!

The NBN and Attempt Number 1

Sol Trujillo had had 17 different jobs at US West — the US telco most like Telstra — before becoming Telstra's CEO on Friday, 1 July 2005. He had started as a graduate and overcame significant prejudice along the way. He had also been CEO at Orange Mobile in Europe and on boards including PepsiCo and Bank of America. He remained on the board of US retailer Target.

Welcome to the Telstra Boardroom on Wednesday evening, 6 July 2005. Phil Burgess, Sol's incoming head of Public Policy & Communications, has come straight from the airport. He has been reading up on Australia. Phil reads a lot.

Sol's Transformation plans — to be announced in November — are Big. Really Big. We are already working round the clock on them.

I am Telstra's General Counsel. I had briefed Sol on regulation two days prior. He is keenly interested in Access Regulation: and incensed that Telstra's merits appeal rights have been removed.

We turn to Access Regulation. Sol looks over at me:

> "So, Will, there's no Bill of Rights in Australia, is there?". "No, Sol", I reply.

https://doi.org/10.18080/jtde.v12n4.1184

> "So that's how come they could take our appeal rights away."
>
> "Well Sol, there isn't a Bill of Rights, but we have a constitution modelled on the US and it contains some similar protections but is less prescriptive than the Bill of Rights."

Phil jumps in:

> "Sol, ya gotta understand — the Minister here is a dictator — she can do what she likes." [The Minister at the time was Senator Helen Coonan.]

Sue Cato — also a new face to me that evening and there to help Sol navigate Australian media — bristles, as do I.

> "Well Sol, it's not quite that simple. The Minister sits in parliament and has Ministerial accountability. Unlike the President's cabinet, she is elected. Phil is right that the government can do a lot — especially as this government has control of the Senate — but the will of the people in a democracy means it's fair enough that elected officials can govern."

I get a glare from Sol.

> "And there are rules about due process and implied constitutional protections too and there's a free press, so governments get scrutinised."

Sol is not convinced:

> "But, Will, the Minister can issue licence conditions and the ACCC has driven wholesale pricing below cost, and it seems like there's nothing we can do about it."

I reply:

> "Kind of, but not entirely — we have a takings clause in the Constitution, just like the US, and it's famous here as 'the vibe' clause, thanks to a movie called *The Castle*. But it's not that useful given the access regime is legislated and in theory we choose to hold our licence.
>
> "Also, telecoms is a really big political issue in a country as big as the US with a fraction of the population. Telecoms is the fifth Federal Power — it's right up the front between funding government and defence."

"Go Will!", says Sue.

Sol concludes:

> "Well, Will, we have to find a way to change things. I'm going to need your plan to help save this company."

So, some observations:

https://doi.org/10.18080/jtde.v12n4.1184

1. Telstra subsequently ran and lost a 'vibe' case in the High Court. It was later cited as one reason the 2009 NBN fibre plan bypassed Telstra.
2. I had to chuckle about Phil's "dictator" comment in 2012 when Communications Minister Conroy publicly noted that his powers meant: "If I say… 'you better wear red underpants on your head'…[then] you'll be wearing them on your head"! (ABC News, 2012).
3. Sol's depth of thought – in this case to go to the foundations of Australia's democracy — and his passion for a 'fair go' were under-appreciated, but core to him. To this day, he is a champion for diversity and inclusion — based always on hard work, ability and results.

Next, to late July 2005: Sol is four weeks into his tenure. Sol and Greg Winn — Telstra's new Chief Operations Officer — go to Blackall in Queensland and visit a small school at the start of the day. One student turns on the computer, connects the modem, and then the class goes outside because their first lesson download takes 45 minutes.

Greg, who had fought in Vietnam and put himself through higher education, was shocked. After telling the Telstra Leadership Team what he had seen, he added: "We're giving rural kids a third-world education with barely working dial-up internet. We have to fix this."

Scroll forward two weeks to 11 August 2005. Sol and Donald McGauchie, Telstra's Chair, head to Canberra to meet with the Prime Minister, Treasurer, Communications Minister and Leader of the Nationals to propose the first NBN: a 6 Mbps network (100x faster than dial-up); $2.6B from the government for rural Australia; and $3.1 billion from Telstra. The catch: Telstra wants an access holiday to avoid Unconditioned Local Loop pricing. There is also a $4.7B option that we will return to later.

The Government quickly says no, but suggests talking to the Australian Competition and Consumer Commission (ACCC).

Forward to November's Investor Day. Telstra's 5-year Transformation is announced: a national 3G mobile network — to be the world's fastest; an MPLS core network; Marketing and IT Transformations; and Bigpond will start movie downloads. But no NBN. In December 2005, Telstra holds a more detailed briefing to urge regulatory reform.

NBN Attempt Number 2

In March 2006, Phil Burgess and ACCC chair, Graeme Samuel, start discussing an access deal. Much progress is made but pricing is the problem, especially the rural cross-subsidy required. Discussions end in early August (*Herald Sun*, 2006). There is disappointment, but no real surprise.

https://doi.org/10.18080/jtde.v12n4.1184

In parallel, preparations for T3 [the third tranche of Telstra shares; ANAO, 2008] are in full swing, but they are very tense. Negotiating how to accurately portray the regulatory risks to investors — when the vendor shareholder controls those risks and also has to sign off — is a delicate balancing act. There is external pressure on some of us to be 'good Australians'. We resist. Finally, in late August, T3 is announced.

Then, in October 2006, way ahead of what anyone outside Telstra thought possible, the world's fastest *national* 3G network, NextG, is launched. It is a measure of Telstra's heightened competitive instincts that, when sprinklers malfunction inside Sydney's Overseas Passenger Terminal Auditorium during the launch, it is only half-jokingly suggested that Optus might have been behind our drenching.

Into 2007 and Attempt Number 3

Telstra hopes a new version of an NBN might be a signature win ahead of the upcoming election. So too does Communications Minister Helen Coonan and her 'we can get things done' Chief of Staff, Peta Credlin. We go to work on a Telstra-Government deal. At one point, Sol and Helen meet to nut out remaining details. The advisers wait outside. Out they come and Helen's first words are: "It's a boy".

This will be superfast FTTN/VDSL, up to 50Mbps; $5 billion plus – all Telstra's money; the major cities; and, in Sol's back pocket, Telstra will fund 100 regional towns as the deal sweetener.

All it needs is an access price that guarantees a risk-based return. An NBN is risky. Will enough people buy broadband? Remember that the iPhone has not been launched. Nokia has half the mobile market. Blockbuster DVD rental shops are in every suburb.

By ANZAC Day 2007 [25 April 2007], we have a Heads of Agreement almost done, just awaiting a final price. Multiple rounds with the ACCC have us close. It is lower than Sol wants, but Telstra can live with it. There is a narrow legislative drafting window with the Office of Parliamentary Counsel for necessary enabling legislation, but they need instructions quickly.

But the ACCC just keeps asking for lower pricing. Maybe the Prime Minister will save the day and step in? Hours, days tick by and then: no deal. After yet another request to cut the price, Sol has had more than enough and decides they are not serious (Sainsbury, 2007; O'Sullivan *et al.*, 2007).

Things then go from bad to worse for Telstra in Canberra.

By the time of the Merrill Lynch Australia Investor Conference in New York in September 2007, Sol does not even dodge the election question when asked. He says:

https://doi.org/10.18080/jtde.v12n4.1184

> "Well, as an American, I don't get to vote, but if I did and I was voting on the basis of broadband policy, then I'd vote for Kevin Rudd" ("Trujillo praises Labor", 2007).

November 2007's 'Rudd-slide' Occurs and Ushers in Attempt Number 4

With the new government implementing its election policy, Telstra's August 2005, $4.7 billion option is opened to all-comers. As new Communications Minister Conroy wryly notes: "Telstra's never underquoted on the cost of anything before"! (Australian Senate, 2008, p. 202). The Request-For-Proposals (RFP) process begins in April 2008, with a November closing date.

A month later, the Telstra 'Brains Trust' goes to Bowral to war-game what could happen. (The GFC [Global Financial Crisis] is rumbling. A prescient Telstra Chief Financial Officer, John Stanhope, has just modelled a major US bank going broke.) Our conclusion: it will be either Telstra or the Government building the NBN. Everyone else will fall by the wayside.

Now if we are right, this presents a real problem. Buried in the RFP are two troublesome requirements. Firstly, to win, proponents will likely need to provide detailed engineering plans. Secondly, the Intellectual Property in the bid becomes government property.

Whilst Telstra prepares a full bid, those requirements mean that Telstra is unlikely to lodge it. Telstra handing over crucial intellectual property for free? No fiduciary could sensibly volunteer for that.

Then there are the political rumblings through 2008 about forcibly separating Telstra. This further reduces Telstra's confidence that its intellectual property would be safe. Telstra boycotts the final RFP briefing in protest.

By the time bids are due in November, there is a bigger issue than the RFP: the GFC. A former bank CEO tells us: "If we don't end up in a great depression, it will be the closest we come to it in our lifetimes".

On the day of submission, all 5000 pages of bid materials are put in his car boot and driven to Canberra by Telstra's Bid team head. After much debate, Telstra lodges only a summary proposal and, to show good faith, puts the full bid into escrow at a law firm.

Telstra is hoping, but not expecting, that the government might be flexible. It is not to be. Telstra is quickly kicked out of the RFP process, but, having boycotted the final RFP briefing, is surprised by the reason given (Stevens, 2008).

It makes little difference to the outcome.

https://doi.org/10.18080/jtde.v12n4.1184

I have mentioned the threats of separation. In parallel to the RFP, Telstra, like Telecom New Zealand, is considering demerging its regulated and retail businesses. The regulated business would then have clear, wholesale-only incentives and could negotiate to build an NBN. Or, if extreme regulation continues, the regulated company might go broke and the regulatory failure would be obvious. Telecom NZ did choose demerger — and, whilst a regulated Chorus had a very bumpy regulatory patch a few years later, New Zealand now provides an insightful counter-factual to Australia.

In Telstra's case, the proposed demerger companies have the working names of 'Copper Loop Co' and 'EEFRA' — Everything Except the Fixed Regulated Assets (aka 'Telstra Retail').

At the Consumer Electronics Show in Las Vegas in January 2009, the Telstra Board seriously considers a demerger. Now, you will appreciate that 'What happens in Vegas, stays in Vegas'. So, whilst Telstra never proceeds, that is not for want of a full and frank discussion.

Sol's departure is announced in February 2009, with a June exit. Meanwhile, as Telstra's war game had predicted, it seems like there are no other bankable bids. Should Telstra go back to the Government with a fresh proposal? Offer to demerge? Some of us think so.

Sol is still in charge:

> "Will, there's a time to hold 'em, a time to fold 'em and a time to run away. Right now, they need the copper. We own it, and we're holding."

Then, at the start of April, there are rumours that the government is going to build an all-fibre network, so they would not need Telstra's copper. But a fibre network will cost at least $40 billion. Telstra's market capitalisation is well lower than that. With that kind of money, they could nationalise Telstra if they wanted!

You can imagine sitting in Sol's office watching *Sky News* on 7 April 2009 as the Prime Minister and Senator Conroy announce a $43 billion plan for an NBN co. to bypass Telstra — along with the accompanying "gun to Telstra's head to encourage co-operation" (Rodgers, 2009). You may think Sol's reaction would have been very predictable. It is more one of lost opportunity than I expect.

My current employer, nbn Co Limited, is incorporated two days later.

Solomon Trujillo and Donald McGauchie

Sol and Donald McGauchie both leave Telstra in early May. Before I move on, this Oration is the appropriate place to note two people and a partnership that has had a profound influence on our industry: Telstra Chair, Donald McGauchie, and Solomon Trujillo.

https://doi.org/10.18080/jtde.v12n4.1184

Sol's energy, intellect and force of will have arguably shaped our industry in Australia more than anyone in recent times. Before Sol arrived, Telstra had two-and-a-half sub-scale mobile networks, Australia was not really on the global telco map, and an NBN was not a political issue.

Telstra's prior CEO, Ziggy Switkowski, had fought hard too. He had even faced down a notional billion dollars in ACCC fines for creating a rational broadband market. But, when Sol arrived, there was a feeling at Telstra that only so much was possible.

Whilst the importance of communications was understood, what was possible was not understood. When Telstra launched NextG in 2006, it ran ads with a school classroom saying a mobile phone was a map, a phone book, a TV, even a bank. People laughed.

Today, Australia has three of the world's best mobile networks. Led by Telstra, the others invested to compete. We have a generation of telco leaders who learned how to work and compete with an intensity they had not needed before. And we have an **nbn** that, in terms of national availability and customer choice, is world leading.

I mention Donald McGauchie — and it took the Board too, of course — but the CEO-Chair relationship was the key. Sol could not have persevered without Donald's staunch backing and willingness to put his reputation and old friendships at risk in pursuit of a dramatic improvement in Australia's digital infrastructure.

Lessons to be Learnt

So, 20 years on, what are the lessons from that era? In my view, the mistakes to be avoided are:

1. Poor communication, made far worse when the media is used to convey messages.
2. Misunderstood motivations and suspicion.
3. Lack of technical understanding and/or great engineering advice.
4. Lack of economic understanding: companies cannot act expecting to hurt shareholders.
5. Populist thinking, but equally, ignoring political reality.
6. A desire to overly shift risk, or to overreach when regulating.
7. Rigid processes that do not allow iteration or learning.

To borrow the primary message of the Harvard Negotiating School: Focus on interests and not positions, and work for a deal that lets the other side have it your way (Shonk, 2024).

https://doi.org/10.18080/jtde.v12n4.1184

So, to the Future

As we think about Universal Service Obligation (USO) reform, we need to start with the world today, not the legacy regulation we are used to.

Historically, we have separated voice and data regulation. But today, voice — and I include Triple Zero [emergency service number] — is just another app needing high reliability. Treating voice separately to data is an anachronism.

In terms of volume, the average household uses 500 GB a month compared to 15 GB for a mobile. Capacity really matters.

Consistent speed is also important. Waiting hours to download a new game on a 50 Mbps service does not make for a great family Christmas. Likewise, speed is valuable if you learn or work on video, in the cloud, with digital media or, increasingly, with AI.

We should not be put off by some views that we are reaching peak demand for data. Whilst video watching will peak, screen sizes and definitions, localised AI models, 3D virtualisation, co-bots (robots working with humans) and pretty much every other technology under development right now is connected and needs data. I think in 20 years' time, by 2044, we will look back at 50 Mbps in 2024 the same way that we now look back 20 years at the standard speed in Australia of 256 Kbps in 2004 — it is hard to imagine that the standard speed in Australian homes was so slow.

So, for reliability, consistency, capacity, latency and speed, here are the primary 'go-to' on-premises technologies we should prioritise, based on population density:

1. FTTP — or multi-Gigabit-capable precursors like HFC — to around 91%.
2. Terrestrial dedicated Fixed Wireless for capacity — for the 91-98% areas.
3. Low Earth Orbit Satellites (LEOs) for the last 2%.

The **nbn** today carries close to 85% of all data in Australia, so these three networks need to carry that load.

Mobile capacity is, of course, very important too and is continually growing, but so too is mobile usage. In terms of future capacity, it seems wise to keep mobile capacity for mobile uses. For example, if ever car cameras are networked for self-driving, mobile capacity will be fully used. Conversely, all the better to keep dedicated stationary networks for high-capacity fixed uses.

Then, rather than overlapping retail voice and wholesale data obligations, to me, an option worthy of consideration is a single Statutory Communications Provider (SCP) in each location where there is locally owned high-capacity fixed services infrastructure.

https://doi.org/10.18080/jtde.v12n4.1184

This would mean **nbn** or its vibrant greenfields competitors are the SCPs. Today, such an SCP footprint would cover all of Australia, given **nbn** currently has two geostationary satellites (which are running better than ever, by the way). However, those GEOs will be decommissioned within the next decade, so **nbn** will not have infrastructure for the last 2% forever. This is where LEOs come to the fore.

Starlink covers everyone today and it is likely that at least Amazon's LEO service will too in coming years, with others possible as well. Starlink is expensive, but that will likely not always be so — at least not everywhere.

You then add in Telstra's 99.6% mobile population coverage, Optus at close to 99%, and TPG expanding too. Layer in an **nbn** with huge capacity, by comparison, out to 98%+, maybe more, and the redundancy issue is not the hurdle it may seem.

Sitting underneath all that capacity is Telstra's rural copper. If it did not exist today, there is no way we would build it, let alone spend $270 m+ per year maintaining it. It is even more outdated now that the dramatic speed and capacity upgrade to **nbn**'s Fixed Wireless network is almost complete.

So, the only logical question to me is: How fast can we exit USO copper and the disproportionately expensive service obligations attached to it, and save a chunk of that $270+ million annual spend in the long term?

Remote twisted-pair copper is one of the least reliable technologies in my opinion — especially in a hot and wet climate. And, when remote copper lines are washed away, they can take months to repair.

Exchange-based power is the only reason I can see left for keeping copper, but mobile phone battery recharging packs are now cheap. Solar-powered LEOs do not run out of sunlight in orbit.

So, in the 2% or so beyond **nbn**'s terrestrial networks: mobile coverage mostly, plus LEOs everywhere, will be the answer for emergency voice.

So then, provided LEOs are available and priced competitively, do we need to regulate more?

Telstra is no longer required to deliver telephone directories to every doorstep in Australia, thanks to global-scale players like Google, Facebook and LinkedIn. So too, in the case of LEOs, we should seek the benefits of global scale and competition, and be careful not to regulate simply because we are used to 'having a throat to choke'.

Having said that, for professional installs and service, or to enable more retail choices, there may be roles for a wholesaler to play. The key then is that the reasons for adding regulation

https://doi.org/10.18080/jtde.v12n4.1184

and funding are clear, well-articulated, fill gaps not otherwise filled, and are done in the most efficient way possible.

As an aside, I do not see LEOs displacing modern terrestrial networks. They will compete — and vigorously in less dense areas — but, given short lifecycles of 5-7 years and associated costs, plus the amount of actual overhead capacity growth versus demand growth, it is hard to see major substitution. If anything, genuinely 'always on' 100% coverage may encourage digitisation and more usage of all networks.

How to pay for it

Finally, support of rural and remote communities is important for Australia's wellbeing. Funding — for past investments, for network augmentation, for service quality, and for customer choice — is rightly all our responsibility.

Firstly, we need to reduce costs where we can. Exiting copper particularly, and **nbn**'s GEOs when prudent, will save money in the long term.

Today's technology and functionally-specific levies create distortions. The challenge with continuing current approaches is that we will have fixed, fixed wireless, handheld mobile and fixed LEO, portable LEO, handheld LEO, plus non-**nbn** GEO services too. To me, there is a simple solution, but, to borrow another phrase I first heard from Phil Burgess, that topic is "above my pay grade".

I do think Telstra will need to give up its TUSOPA [Telstra Universal Service Obligation Performance Agreement] entitlements, but equally, it avoids worsening copper economics and various down-side scenarios. Hopefully, it has learned a lesson about not over-playing its hand.

Conclusion

So, in conclusion:

1. Listening — especially to customers, this industry, stakeholder groups and all decision makers — is vital. Clear, consistent, balanced but suitably optimistic communication to assist change is then essential. There is much to be optimistic about. To quote a former Telstra colleague: "Sell hope, but paint reality".
2. Every day of delay means wasted money and people missing out on better outcomes. We in this industry need to show what is possible and engage widely.

In words Sol might find familiar and Sir Charles would also have understood: 'You can't fight city hall, but sometimes they can be persuaded with a very good idea'.

https://doi.org/10.18080/jtde.v12n4.1184

Question-and-Answer Session

This section was taken from the video of the Oration. It has been edited, mainly for brevity and to remove repetition.

Q: In the industry in 2007, there were a large number of ISPs [Internet Service Providers], who were later consolidated into larger entities. Could the NBN have come about from this, despite Telstra?

An interesting question. Before Sol, with Ziggy back in 2004, we looked at broadband pricing and the state of competition. There were more than 600 ISPs in Australia. That was more globally than in any other country, by a mile, even though we have a fraction of the population of much larger countries. And, from Telstra's perspective, it was just driving a whole lot of inefficiency. It was suggestive of regulatory arbitrage, not businesses with capital.

To give one example: After complaints about one ISP, which had gone offline, the Telstra Wholesale service people called to ask: "Why are you offline?" And the lady on the line said: "I told him to clean his room and I turned off the box [the ISP server], and it's not going on until he does". I am not making that up — that was an ISP.

The focus at Telstra was to create a rational broadband market. Without consolidation, we were not going to see, in my view, the quality of service that was needed to give customers the confidence to move, in those days, from dial-up to broadband. Remember there was a discussion at the time with Graeme Samuel about "fraudband": whether 128 kbps was broadband or not? We needed to get industrial scale and quality to get to facilities-based competition, which was the holy grail of the regime in the 90s. That wasn't going to happen with 600 ISPs.

Then you think about content. In November 2005, Bigpond spent a huge amount of Telstra money on launching downloads of content (i.e., movies), but that also triggered the rest of the industry to invest. For example, think about Optus's strategy in the last five years with the English Premier League. Think about the arrival of Fetch (and amazing that that has ended up as a Telstra product given that for a long time it was a prime competitor to Telstra), but that's again global scale and everybody (Netflix, Disney etc.) coming in.

So back to your question. No, we had to go through that consolidation in those days. Then you think about the next wave of competition now with a lot of the infrastructure build and you look at this week's announcements around Vocus. If you thought back to 2005/6, all those names that ultimately went into TPG — into iiNet — and then into Amcom and M2 and Vocus and Commander and so on — we've now got industrial scale and we've got a measure of facilities competition that was undreamed of in 2004. So, I think we had to go through the

https://doi.org/10.18080/jtde.v12n4.1184

consolidation because, effectively, if one of the goals of regulation is efficient competition, then you have to create the grounds for efficiency.

Q: Please comment on the 121 Points of Interconnect (PoIs) for **nbn**. They create a problem of scale for a small RSP [Retail Service Provider] to reach all of them. Is this necessary?

Lots in that question! It is fair to say that, if you go back originally, there were two things driving the number of PoIs. One was the concept that you would need data caching close to customers for a bunch of reasons, so 121 PoIs put you close to customers. Secondly, back to consolidation, it was also going to drive, at the core network level, a degree of consolidation — think about Telstra, Optus, TPG, Vocus and Aussie Broadband — and it took us pretty much until 2017/18 before you got those core 121 PoIs all connected to those kind of larger wholesale infrastructure players.

There is still a question again: if you think about most parts of the Australian economy, what kind of number of material suppliers do you get to, and have a level of efficiency with, given 7.2 million square kilometres and 27 million people? It's not my place at **nbn** to comment right now, because the 121 is set. I think it is fair to say that if you were to arrive from Mars today, you would do a lot less than that, but, again, like so many things, we are in the world we are in. It is that old Irish joke that you may not want to start from here to get to Dublin — but this is where we've got to start from.

Q: Please comment on serving the last 2% of customers by LEOs vs GEOs, noting that satellites are expensive.

Again, **nbn** has quite publicly run a request for information from LEO providers. As I mentioned, we are now more than halfway through the life of our geostationary satellites, so it's a very clear and present issue; **nbn** has a statutory obligation to provide service to 100% of Australians, so, as we don't have LEOs right now, we should be working on what other alternatives there are, in an economically rational and sensible way, to provide that service. As I said in the speech, there's a question in fact whether that will be needed.

Today, there is only the one LEO provider, Starlink. If you look in the Australian market, Telstra resells that service, and you can also buy it from Starlink — but it's expensive, about $125-odd at retail for the Telstra service, about $140 if you buy the full product direct from Starlink. You compare that to the offers from Sky Muster Plus that are below $100 in an unconstrained data context today. But that has a limited lifetime because of the lifetime of the GEO satellites.

The technology is definitely there. In New Zealand, Starlink has a product it is selling for NZ$79 (about AUD 70). It's not available everywhere and it's slower speed, but it's perfectly good for many people if you read the Commerce Commission reports on it. Amazon — not a

https://doi.org/10.18080/jtde.v12n4.1184

small company, very well resourced, very technically capable — is, I think, and I think we all hope, going to get into the LEO business in the not-too-distant future. And, at that point, I think we will see pricing broadly comparable to what we see today in Australia for non Starlink services.

Remember today, if you buy Sky Muster, you don't pay for the installation of the dish; if you go to Starlink, you've got to do it yourself; and Telstra has blended an offer in between. There's going to be a lot more competition. If you look at the other businesses owned by Elon Musk, Space X's ultimate owner, you see a lot of competitive activity, price differentiation, and in different markets depending on different contexts, so, in that sense, I am an optimist.

Why? Firstly: global scale — as with the directories business — will end up replacing the need for local initiatives. Note though that when you think about the billions spent on **nbn** and a billion plus spent on the geostationary satellites, there was a phenomenal expenditure and we are paying for it today in the context of the regional broadband levy. We will need to pay that off — and that's a legacy that's sitting there.

Secondly, as I kind of mentioned, where the USO funding goes in the future should logically enable, even if its underlying price is higher than today's GEO-based offers, people in the bush to receive something at a broadly similar price to today. The service conditions, though, may be a bit different.

Q: What would have happened to **nbn** if the (previous) government had not mandated a multi-technology mix?

The analogy I have often used is that it is like a young couple, they're thinking about buying a house, and they've sort of got two choices. They're planning a family: do they buy a bigger house now? It is going to cost more, but it's all there. They don't need to move later. And it's going to be a lot more convenient. Or do they buy the smaller house, but with some land with room at the back, and then later on do the extension. They'll spend a little bit less now. They're going to spend maybe a bit more later. It's going to be more disruptive.

The first version of the NBN was to buy the big house early. The second version (multi-technology-mix) was not to buy the big house and to add the extension later. Nobody at the time had the benefit of hindsight, but with the knowledge now of what happened with COVID, I actually think that what was done was the right plan for Australia to get through a pandemic, because the NBN was, you know, declared built and fully operational during the first year of the pandemic. Certainly, when I was sitting at Telstra in 2010/11 negotiating the NBN deal, had someone suggested the idea that the NBN — and it was pure fibre at that point — would be finished by 2020, well, we would have fallen around laughing. In fact, a bunch of guarantees

https://doi.org/10.18080/jtde.v12n4.1184

were slated to run to 2027, which gives you an indication that Telstra was thinking that it might well take the NBN until 2027 to get finished.

It is a purely hypothetical question. You can look at other markets. I mentioned New Zealand, where things played out very differently. The key point is now, as I said, let's make the best of where we are from here. Sure, learn lessons, but, ultimately, the world ends up with multi-Gigabit service in my opinion. Now, is there an ideology around fibre? No, there's no ideology around what is going to deliver a reliable service at the capacity that is needed.

Q: Please comment on William Webb's thesis, in his recent book, *The End of Telecoms History*, that we have reached the speeds ultimately required for broadband.

I didn't mention the issue of speed in the speech. I don't think humans are going to watch a lot more TV; if the core thesis says we're not going to watch a lot more TV, well, I frankly agree, and having at least one teenager left in the house, I kind of hope that's true. But in the world of what we used to call, 10 or 20 years ago, machine-to-machine communications, the amount of data you need to download and continue to update things is large — like a large language model for what I'll call onboard AI.

Why does **nbn** carry close to 90% of the traffic in the country? Because most of the devices you are carrying will wait until they can get to a fixed network to do any communicating at scale. And it's not about: can you do it over however many hours it takes? It's about: can I do it now?

Now, to return to that example I was using earlier. You get on the freeway and, all right, it has a speed limit of 100 kph —and at 2 am or 4 am or even 6 am, you can travel at 100 kph. But, get on there at 5:45 pm at night, you're crawling along at 20 kph. The average speed is probably 90 kph — sounds great — but, when you need it, it's 20. And that's when that high speed — i.e., capacity at a point in time —becomes so important. And that's why I *do* think, you know, we will use higher capacity.

Finally, if you look at all the technology companies, you think about what at **nbn** we call "cobots" — this is robots working with humans which suddenly need LIDAR and spatial understanding. The robot in a factory doesn't need to know that, for example, my microphone is here [waving to touch the microphone]. A robot I'm working with needs to know where I am. And that then brings LIDAR; it brings multi-Gigabit needs.

We saw on the news last night that Australia is well below replacement rate for our population. China is getting to that point; Japan is already well at that point. Some of us when we're a lot older are going to likely have more than the robot vacuum cleaner floating around — and they're going to need, and we are going to need, a level of connectedness and data capability that we can only dream of. I meant it when I said 256 kbps 20 years ago: some of us wondered:

https://doi.org/10.18080/jtde.v12n4.1184

Yeah, why would you need more? Only fifteen percent of Telstra's customers were on Telstra's top speed of 1.5 Mbps back then. Why would you spend the money on 1.5 Mbps?

If you look at all of the investment in new technology globally, it is hard to think of a technology which does not rely on connected data to deliver it. Back in 2004, a guy called Reed Hastings just started dropping DVDs in satchels with the word 'Netflix' written on the side. When Netflix arrived in Australia, data use per Netflix customer went up 80% in one month. I don't know what the next Netflix will be, but I'm reasonably convinced — after 30 years this year since I first started working with Telstra —that data growth is not going to suddenly disappear.

Sell a lot of books, make a great debate for sure, but I'm not in the Webb camp.

Q: If Telstra had built the 2007 FTTN network, would we have been in a better position or, given the aging of the copper, would we effectively be in a worse position?

It's always interesting with any new technology. If you go early, you have that leading edge, early adopter problem; if you go late, you get the benefit of seeing what others have done, but you miss out for a period on whatever gains there are. I think what would have happened was essentially what has kind of happened in the context of the multi-technology mix. Telstra would've started on ADSL and then FTTN. With its $5 billion, by about 2010/11, it would have got the cities done. And — back to our earlier conversation — we might well be sitting here today and you're saying: "you know what, 50 Mbps is not really enough".

We would have seen, I think, the HFC as it is today, but, rather than having it as, you know, either/or HFC or fibre, we would've seen a competitive layer in between the service and the HFC. At some point, Telstra would have had to separate some parts of its business. It might have been the HFC network. In North America, you know, that was the separation point. So, I think we would've ended up with those split up, rather than **nbn** controlling a single network.

We certainly would've got things done a lot quicker. The fact is that Chorus started with all of the incumbent assets of Telecom New Zealand and was finished earlier at a much, much smaller cost by comparison. I should note that Chorus received some government money as interest-free loans; it didn't receive the kind of capital of **nbn**.

But, like all these things, again, we are where we are. The job now for all of us, whether we are part of the **nbn** ecosystem or competing against the NBN, is: what do customers need? How can we most efficiently deliver that? If this industry focuses on customers, it will be OK. If we focus on ourselves, we'll continue to go around in circles.

Q: What is the impediment to USO reform?

It's a complex one, and some of it is probably — to borrow a phrase — "above my pay grade", but there is always a resistance to change. And, if you think about copper services, a lot of

https://doi.org/10.18080/jtde.v12n4.1184

people feel, for good reason, a measure of safety. They know it's powered from the exchange and they think if the power goes down — and, you know, in the middle of a natural disaster power is often the first thing that goes — well, I've still got my landline. And, yes, today we have power cubes; and the mobile operators, compared to a few years ago, have done a huge amount more on redundancy power at base stations and so on. But it is far from perfect. So, there's been a reluctance to say, well, I'm going to take something you've relied on for a long time — it has a measure of safety, it has one attribute that is different to its alternatives — and for the customers to say, gee, I don't want to let that go.

I didn't get into it in the speech, but I mentioned things going from bad to worse. In 2007, Telstra was pushing to switch off its 2G CDMA rural network to take the 850 MHz spectrum and refarm it to NextG. It proposed that, it had blue-ticked phones and a whole bunch of things, and there was a huge amount of resistance, so that the Minister of the day said no. Telstra, for a bunch of reasons that probably no telco would do today, decided to sue the Minister. The change of government happened. The new Minister made a different decision.

The day before that network got switched off — and, again, I am not particularly exaggerating — Telstra was getting some pretty serious threats about very nasty things being done to some of its staff, if it went ahead with the switch-off, because we were 'going to kill people'. And the day after, people were going: "Wow, this is so good. It's so much faster, it's so much better".

And, early on, there were some gaps in coverage. The old network had gone to a few places that the 3G network did not. But, over time, that network went far further. We are now out to 99.6%-odd of the population, compared to give-or-take 99% at that point. So, sometimes, yes, there is a short period of loss, but it's always that resistance to change.

That, to me, is the big issue. That then feeds into: well, do I want to be, as somebody who represents people, the person who is going to tell people I'm taking a "definite" away with a "maybe" to follow? Back to my point about the listening, and then the understanding, and then the explaining: maybe there's a case for doing it in different places first. We already do that internationally, but whether we do that domestically is a question that I think will get asked.

Q: What about sovereign risk with regard to foreign-owned LEO satellites?

It doesn't just apply in a LEO context, there are very many things where Australia, as less than 1% of the global economy, is dependent on others. There are kind of two layers to it, I think. One is: what potentially happens in a competitive market? Is there a dysfunctional market that therefore would cause a rational player to do something that otherwise might be irrational? I'm going to exit the market because something's happened here and there's sort of, you know, a risk that I choose to do so. So, I'm having a bad day and throwing a tantrum because X or Y

https://doi.org/10.18080/jtde.v12n4.1184

or Z happened and I just feel like having a go at someone. What's going on in Brazil at the moment, where you've got a bit of a power struggle going on, might be an example.

In my experience of these things, people ultimately are rational if you give them a chance to be rational. If you publicly box people into a corner — and this can be businesses trying to box governments into a corner or, you know, business to business or whatever it is — people pushed into a corner will come out with something you don't expect. It happened with the $43 billion. Telstra did not expect it. The government was boxed into a corner and it came out, absolutely, in a way that, you know, triggered things.

I do think we will end up in a competitive market. I don't think it will be in any LEO operator's interest to decide that they want to depart. It is five-eyes based, which is a huge advantage. So I am relatively — you'll note my words, relatively — cautiously confident. There's a whole political sovereign risk: we watch Canada and India right now and see that things can happen totally outside what people expect. None of us can know what may happen. But you'd say: what's the relative cost for relative benefit? I would suggest that actually being confident enough about a future market for LEOs, in the context that we're in, makes sense.

And then, as I said, what's the worst case in the short term? There's about a half percent of the population who would not have coverage at that point. The question then becomes: what quickly would happen to fill in that coverage? Cells on wheels and all sorts of things. And, even if the LEOs go, there are still GEOs and there are a bunch of global providers of GEOs. At **nbn**, looking from the outside, the GEOs are going to finish, but there are plenty of people — Singtel Optus is a great example — who have a lot of GEO capacity and you can buy from them. So, the idea that all the satellites will, you know, proverbially fall out of orbit, I don't think so. I think there will be options for that last piece of coverage if the LEO market doesn't evolve.

References

ABC News. (2012). Conroy plays down 'red underpants' comments. September 28, 2012. Available at https://www.abc.net.au/news/2012-09-28/conroy-plays-down-red-underpants-comments/4286562

ANAO [Australian National Audit Office]. (2008). Third Tranche Sale of Telstra Shares. Available at https://www.anao.gov.au/work/performance-audit/third-tranche-sale-telstra-shares

Australian Senate. (2008, February 2008). Estimates. Official Committee Hansard. Senate Standing Committee on Environment, Communications and the Arts. Canberra. https://www.aph.gov.au/~/media/Estimates/Live/commttee/S10631.ashx

Herald Sun. (2006, August 8). Pulling the Plug: Hard road to the node. [chart] *Herald Sun*.

O'Sullivan, M., Murray, L., & Barker, G. (2007, May 3). Telstra, Coonan at odds over high-speed broadband deal. *The Age*.

https://doi.org/10.18080/jtde.v12n4.1184

Rodgers, E. (2009, September 15). Opposition slams proposed Telstra reforms. ABC News. https://www.abc.net.au/news/2009-09-15/opposition-slams-proposed-telstra-reforms/1429922

Sainsbury, M. (2007, May 3). Secrecy cloaks fast net failure. *The Australian*.

Shonk, K. (2024, October 24). Principled Negotiation: Focus on Interests to Create Value. *Daily Blog*, Program on Negotiation, Harvard Law School. https://www.pon.harvard.edu/daily/negotiation-skills-daily/principled-negotiation-focus-interests-create-value/

Stevens, M. (2008, December 15). Telco's crossed lines to Canberra. *The Australian*.

"Trujillo praises Labor; Coonan not happy". (2007, September 14). *SmartCompany*. https://www.smartcompany.com.au/technology/trujillo-praises-labor-coonan-not-happy/

https://doi.org/10.18080/jtde.v12n4.1184

The Impact of Fixed and Mobile Broadband Adoption on Economic Growth

A Panel Data Analysis

James Endres
SR Economics

Mariana Steula
SR Economics

Abstract: This study analyses the impact of fixed and mobile broadband take-up on economic output for 34 OECD and G20 countries over the period 2008 to 2022. The results suggest that for the period both fixed and mobile broadband made positive contributions to a nation's GDP per persons employed, with mobile broadband having the larger impact. The results also suggest that during the pandemic period (2020–2022) both fixed and mobile broadband were positively associated with GDP output. That said, many countries in the sample experienced notable declines in the take-up of mobile broadband during the pandemic period, resulting in a drag on GDP. This was offset, however, by increased take-up of fixed-line broadband services, which resulted in higher levels of economic output.

Keywords: Broadband, Economic Growth, Fixed, Wireless

Introduction

The extent to which economic and social welfare is enhanced by investments in communications technologies has been a focus of policy makers, industry and academics for several decades. This focus is evidenced by a large and diverse body of literature which examines the nature of, and extent to which, economic impacts arise from investments in, and use of, communications technologies (Deloitte, 2021).

Over the past 50 years ongoing innovation has transformed communications technologies and their uses. Hence, as communications technologies have evolved, so too has the focus of the relevant literature. For example, early studies were focused on the role of the telephone in economic development (Hardy, 1980; Shapiro, 1976). By contrast, recent studies attempt to estimate the extent to which future investments in fibre broadband networks — to enable the

https://doi.org/10.18080/jtde.v12n4.1128

delivery of ultra-fast broadband services — will further promote economic growth (Campbell, *et al.*, 2021; Briglauer *et al.*, 2021). Despite this shifting focus over time, the relevant literature provides robust and consistent empirical evidence that investments in communications networks and associated technologies have material impacts on firm-level productivity (Grimes *et al.*, 2009) and macroeconomic outcomes, including higher rates of economic growth, improved employment outcomes and enhanced productivity (Crandall *et al.*, 2007; Allen Consulting Group, 2002; Koutroumpis, 2019).

The relevant literature not only reports the findings of the empirical analysis, but it also describes the mechanisms for how communications technologies, including broadband, improve economic outcomes over time. Investments in communications technologies impact how and where economic activity takes place, resulting in increased economies of scale and improved resource allocation (Hardy, 1980). This means that, in addition to the direct economic impacts that flow from the construction of communications networks and the manufacture of associated communications equipment, there are important indirect effects that impact both firms and households and are often measurable at the firm, industry, regional and national levels (Allen Consulting Group, 2002).

This paper seeks to add to the existing body of literature by investigating: (1) the extent to which investments in, and use of, mobile broadband networks drive economic growth and improvements in productivity; (2) how these economic impacts compare to those arising from investment in, and use of, fixed-line broadband; and (3) the extent to which both fixed-line and mobile broadband contributed to economic output during the COVID-19 pandemic period (i.e., for the period 2020 to 2022). To do this, a static fixed effects and a random effects regression model are estimated using a panel dataset for 34 OECD and G20 countries,[i] for the period 2008 to 2022. Data is sourced from both the OECD and the World Bank.

The estimates indicate that, over the period 2008–2022, there has been a strong causal relationship between mobile broadband and economic growth and that, relative to fixed-line technologies, mobile broadband has a larger impact on GDP. Despite this, the results also indicate that, during the period of the COVID-19 pandemic (i.e., 2020 to 2022), several countries experienced large reductions in the number of mobile broadband subscribers (as measured by number of mobile subscribers per 100 inhabitants), which for some countries had the effect of reducing GDP in one or more years of the pandemic. Finally, the estimated coefficients suggest that mobile broadband may serve as a substitute for fixed-line broadband. These findings may have important policy implications, particularly for the design of future policy frameworks with respect to competition regulation, as well as achieving universal access to high-speed broadband services.

https://doi.org/10.18080/jtde.v12n4.1128

The remainder of this paper is organised as follows. The second section briefly summarises the existing literature dealing with the relationship between broadband and economic growth. The third section introduces the empirical framework and the data. The fourth section presents the empirical results of the estimation methods. The fifth section discusses potential policy implications for Australia and identifies issues that warrant further empirical assessment and consideration. Finally, the last section identifies some important limitations of the study.

Literature on the Economic Impacts of Fixed and Mobile Broadband

The Organisation for Economic Cooperation and Development (OECD, 2007) considers broadband to be a General-Purpose Technology (GPT) that, when combined with other information communications technologies (ICTs), can fundamentally change how and where economic activity is organised. This change arises from both direct and indirect impacts. Direct impacts stem from direct investments in the infrastructure itself as well as associated ICT equipment and services. Indirect impacts flow from all aspects of economic activity which are impacted by broadband, and which drive changes in economic output (Collins *et al.*, 2007).

The impact of broadband on firm-level productivity

Several studies indicate that the adoption and use of broadband by businesses has had a material impact on firm-level productivity by enabling businesses to adopt new business models, processes and innovations, which in turn enables businesses to reduce costs, increase productivity and improve overall firm efficiency.

An influential Australian study by the Allen Consulting Group (2002) reported that Australian businesses experienced cost savings of around 6.3 per cent from the use of broadband Internet compared to 1.5 per cent from the use of dial-up Internet. The study claimed that the reported average cost savings would result in an overall productivity gain of around 0.32 per cent for Australian businesses.

Similarly, a study by Grimes *et al.* (2009) used panel data of 6060 New Zealand firms, sourced from two surveys conducted by Statistics New Zealand, to estimate the impact of broadband adoption on firm-level productivity. The authors employed two estimation approaches: propensity score matching (PSM) and an instrumental variables (IV) estimator. Both methods indicate that New Zealand firms that adopted a broadband connection experienced a material productivity boost relative to those firms that relied on a narrowband Internet connection or had no Internet connection at all. Specifically, results from the PSM estimation indicated that broadband connectivity increased firm-level productivity by 6.9 to 9.7 per cent. This increase

https://doi.org/10.18080/jtde.v12n4.1128

was consistent across different types of firms with no significant differences between an urban versus rural split or between high versus low knowledge industries. The IV estimation results indicated even higher productivity impacts of between 21 and 25 per cent.

The macroeconomic impacts of basic broadband

In addition to those studies that look at the impact of broadband on firm-level productivity, there is a large body of literature that has examined the association between broadband and macroeconomic-level indicators, such as economic growth, employment, wage levels and productivity growth (see Atif *et al.* (2012) and Wieck & Vidal (2011) for a comprehensive overview of this literature). A key theme of this literature is that for both developed and developing nations there is a strong causal relationship between broadband and growth in economic output, and that countries with higher rates of broadband adoption experience higher rates of economic growth. For example, Qiang *et al.* (2009), using data for 120 developed and developing countries, employed an endogenous growth model to evaluate the impact of broadband adoption on the average growth rate of GDP per capita between 1980 and 2006. The results suggest that, for a developed country, all else being equal, a 10 per cent increase in broadband adoption would yield a 1.21 per cent increase in the growth rate of GDP per capita. Similarly, for developing countries, all else being equal, a 10 per cent increase in broadband adoption would yield a 1.38 per cent increase in the growth rate of GDP per capita.

Koutroumpis (2019) provides an in-depth analysis of the relationship between broadband infrastructure and economic growth across 35 OECD countries over the period 2002 to 2016. The study finds a consistent positive effect of broadband adoption on national economic output with a 10 per cent increase in broadband adoption being associated with a 0.3 per cent increase in GDP per annum on average across the OECD countries. The study also finds that the impact of broadband on economic growth exhibits diminishing returns to scale, meaning that the marginal benefits decrease as broadband adoption increases.

A more recent study by Katz & Jung (2022) examined the role of broadband in mitigating the economic losses resulting from COVID-19 in the United States. Using a Cobb-Douglas production function, estimated within a structural multi-equation model, the study reveals that US states with higher broadband adoption were less economically impacted in 2020 by the COVID-19 pandemic compared to states with lower broadband adoption. Extrapolating these results to a national level, the study indicates that, if the US national broadband adoption rate was equal to that of the most connected state, Delaware,[ii] then national GDP would have contracted by only 1.0 per cent in 2020, compared to the actual contraction of 2.2 percent. Conversely, if the US national broadband adoption rate had been equal to the least connected state, Arkansas,[iii] then GDP would have contracted by about 2.7 per cent. The authors argue

https://doi.org/10.18080/jtde.v12n4.1128

that the COVID-19 pandemic highlights the critical need to close the digital divide and to ensure universal adoption of broadband connectivity in the United States.

Economic impacts of higher speed broadband network

In addition to those studies analysing the economic benefits of broadband availability and its initial take-up, several more recent studies have explored whether incremental economic benefits arise from users moving up the broadband speed curve. A paper by Bai (2017) examines the impact of different broadband speed levels using US county-level data from 2011 to 2014. Employing a first-differenced regression analysis, a positive relationship between broadband availability and county-level employment is identified. However, compared to lower-speed broadband, faster broadband did not significantly impact employment.

Briglauer & Gugler (2018) investigate the economic impacts of fast and ultra-fast broadband technologies in 27 European Union (EU) member states from 2003 to 2015. The authors define fast and ultra-fast broadband as typically either hybrid-fibre ("fast") or wholly fibre ("ultra-fast") technologies that typically provide data rates from at least 30 Mbps up to several Gbps. The authors conclude that the availability of ultra-fast broadband results in a small but significant incremental benefit on GDP. Importantly, using a cost-benefit framework they conclude that economic welfare would not be maximised by government policies targeting 100 per cent coverage of ultra-fast broadband; rather, economic welfare would be maximised by a combination of 50 per cent fast and ultra-fast broadband, with the remainder being basic broadband.

Using data for 35 OECD countries for the period 2002 to 2016, Koutroumpis (2019) finds that broadband speed acts as a moderator of economic impacts and that there is a speed threshold beyond which further increases in broadband quality do not yield significant economic benefits. Importantly, this speed threshold increases over time as more advanced applications and skills become available, suggesting that policymakers should encourage continuous investment to ensure that broadband infrastructure keeps pace with increasing demands for higher speeds.

Hasbi (2022) estimates the impact of high-speed broadband on local economic growth utilising data on almost 5,000 French municipalities for the period 2010 to 2014. The findings indicate that municipalities with a very high-speed broadband network tend to be more attractive for companies operating in non-primary sectors and that these municipalities exhibit higher rates of business creation.

Finally, an influential study by Briglauer *et al.* (2021) estimated the economic benefits of high-speed broadband within and across neighbouring counties in Germany, using a panel dataset of 401 counties for the period 2010 to 2015. The key finding of this study is that a 1 Mbps

https://doi.org/10.18080/jtde.v12n4.1128

increase in average broadband download speeds induces a rise in regional GDP of 0.18 per cent and that when regional externalities[iv] are considered the rise in GDP is further increased to 0.31 per cent.

The economic impacts of mobile broadband

As noted by Eisenach & Kulick (2020) there are comparatively few studies examining the impacts of mobile broadband on economic growth and productivity, and there are even fewer which directly compare the economic impacts of mobile broadband with those arising from fixed-line broadband. Nevertheless, a few recent studies provide robust evidence that investment in, and use of, mobile broadband is positively associated with improved macroeconomic outcomes. This is particularly the case in rural areas and developing countries where fixed-line networks are expensive to roll out and therefore are either not widely available or non-existent.

An influential report prepared on behalf of Vodafone by Waverman *et al.* (2005) looked at the impact of mobile telephony on economic growth in both developed and developing nations. This study employed both an annual production function (APF) approach and an endogenous technical change (ETC) approach to estimate the impact of mobile phone networks on economic growth using data from 92 countries (both high and low income) for the period 1980 to 2003. The empirical results indicate that mobile telephony has a positive and significant impact on economic growth, and this impact may be twice as large in developing countries as compared to developed countries. Specifically, a developing country that had an average of 10 or more mobile phones per 100 persons between 1996 and 2003 would have enjoyed per capita GDP growth that was 0.59 per cent higher than an otherwise identical country. While the study is focused on mobile telephony, it is nonetheless instructive given that, in addition to reporting the results of their empirical analysis, the authors made several important observations about the role that mobile telephones are playing in less developed countries. These observations may be relevant for mobile broadband. Specifically, they observed that:

- In less developed countries, mobile phones are playing the same crucial role that fixed telephony played in the richer economies in the 1970s and 1980s;
- Mobile phones are a substitute for fixed-line telephones in poorer countries but are a complement to fixed-line phones in rich countries;
- Many countries with an under-developed fixed-line network have achieved rapid mobile telephony growth with much less investment than would be required to roll out a fixed-line network.

A study prepared for the GSM Association by Deloitte (2012) employed an endogenous growth model to test whether the transition of users from 2G to 3G mobile networks has a positive

https://doi.org/10.18080/jtde.v12n4.1128

impact on growth of GDP per capita. The study used a panel dataset for a sample of 96 countries over a 4-year period from 2008 to 2011. The study finds that a 10 per cent higher 3G adoption increased annual GDP growth per capita by 0.15 percentage points. Additionally, the study estimated the impact of mobile data usage across 14 countries, finding that a doubling of mobile data consumption increased GDP by 0.5 percentage points.

Eisenach & Kulick (2020) modelled the economic effects of 4G mobile technology adoption in the United States on employment and economic growth and, based on those results, projected the economic benefits of 5G adoption under different counterfactual scenarios. Using panel vector autoregression techniques and state-level data on 4G adoption over the period Q1 2010 to Q4 2014 (i.e., 18 quarters), the authors find strong evidence of a direct positive relationship between the pace of 4G adoption and growth in both employment and economic output. Specifically, a sustained one percentage point increase in the adoption of 4G mobile broadband over a period of eight quarters would increase employment at the end of Q4 2014 by approximately 147,000 jobs and increase annualised GDP by $121 billion. Based on these findings, the authors project that, if 5G adoption in the US follows a similar path to the adoption of 4G, then, at its peak, 5G will contribute approximately 3 million jobs and $635 billion in GDP to the US economy.

A 2022 study by Leo & Clement (2022) employed structural equations and generalised method of moments (GMM) to analyse the effects of mobile broadband on economic growth and employment in Nigeria between 2010 and 2020. The study finds that mobile broadband positively influences economic growth and reduces unemployment in Nigeria. The findings prompted the authors to recommend that the Nigerian government should pursue policies which enhance the availability and affordability of mobile broadband in Nigeria.

Methodology

To evaluate the impact of fixed and mobile broadband take-up on economic output per capita, a structural regression model is estimated, based on a Cobb-Douglas production function in which a country's economic output (GDP per persons employed, *GDP_PW*)[v] is assumed to have a relationship to that country's level of capital (*K*), labour (*L*) and human capital (*EDU*). As a general-purpose technology, both fixed-line and mobile broadband are assumed to effect economic output through their impact on total factor productivity (TFP). Thus, the effect of a 1 per cent increase in the take-up of fixed-line broadband (*FL_BB*) and mobile broadband (*Mob_BB*) on GDP per worker can be directly interpreted from the coefficients $\beta 3$ and $\beta 4$, respectively, in the following equation:

$$\log(GDP\_PW_{it}) = \beta_1 \log(K_{it-1}) + \beta_2 log(L_{it}) + \beta_3 \log(FL\_BB_{it}) + \beta_4 \log(Mob\_BB_{it}) + \beta_4 \log(EDU_{it}) + \beta_6 \log(Pop_{it}) + \beta_7 \log(R\&D_{it}) + \beta_8 \log(Exp_{it}) \quad (1)$$

https://doi.org/10.18080/jtde.v12n4.1128

The remaining variables are *Pop*: total population, *R&D*: total expenditure on Research & development as a percent of GDP, and *Exp*: the total value of exports as a percent of GDP.

Specifically, three different models were estimated:

- Model 1 – this model is focused on the impacts on GDP per person employed arising from the use of fixed-line broadband services in the absence of mobile broadband technology. This model omits the Mob_BB_{it} variable from equation (1).
- Model 2 - this model estimates the economic contribution of mobile broadband in the absence of fixed-line broadband. This model omits the term FL_BB_{it} from equation (1).
- Model 3 – this model estimates the economic impact of both fixed-line and mobile broadband, as per equation (1), on the basis that both technologies co-exist, albeit with different adoption rates and geographic availabilities.

The above models were estimated using panel data, collected from the World Bank for the period 2008 to 2022 across 34 countries. All countries are member states of either the OECD or the G20 that provided sufficient data for the model period of analyses.

A Hausman test of the data indicated the presence of unobserved heterogeneity. To control for this, the fixed effects estimation method was used. All variables were transformed into their logarithmic form to correct for data skewness. Consequently, models 1, 2, and 3 were specified as log-log functions. A descriptive summary of the variables is provided at Appendix B.

The panel dataset covers the period 2020 to 2022, in which the global economy was impacted by the COVID-19 pandemic and the associated lockdowns and economic disruption. This provides an opportunity to test claims by many Governments and industry advocates that the adoption of broadband connectivity allowed nations to mitigate the damaging economic effects of the COVID-19 pandemic.

To do this, a random effects model[vi] that reflects the coexistence of both fixed-line and mobile broadband is estimated using the same data for each of the 34 OECD and G20 countries in the sample. A dummy variable was used to account for the impacts of the COVID-19 pandemic, including the pandemic itself, the imposition of lockdowns, travel restrictions and other emergency measures by national governments, as well as disruptions to international supply chains and the world economy. Noting that it was not until May 2023 that the World Health Organization (WHO) declared an end to the global health emergency caused by the COVID-19 pandemic (WHO, 2023) and that throughout 2022 the pandemic continued to disrupt the global economy[vii] (see World Economic Forum, 2022), this study treats each of the years 2020, 2021 and 2022 as being impacted by the COVID-19 pandemic.

https://doi.org/10.18080/jtde.v12n4.1128

Additionally, for each country we calculate the extent to which annual changes in fixed-line and mobile broadband adoption rates during the period 2020 to 2022 contributed to economic output. To do this, we first calculate the extent to which both mobile and fixed-line broadband adoption changed in each of the years 2020-2022.

Results

Table 1 reports the results for each of the three fixed effects models.

Table 1. Estimation results for Fixed Effects Models

GDP	Coefficient	Std Err	*t*	P>|*t*|
Model 1: Standalone fixed-line broadband				
Constant***	6.73	.258	26.1	0.00
K***	.330	.0152	21.7	0.00
EDU***	.161	.032	5.02	0.00
R&D***	.056	.0135	4.19	0.00
L***	-.121	.0211	-5.74	0.00
EXP***	.157	.018	8.90	0.00
POP***	-.203	.025	-8.21	0.00
FL_BB***	.111	.019	5.94	0.00
Model 2: Standalone mobile broadband				
Constant***	5.26	.345	15.3	0.00
Kl***	.370	.014	26.5	0.00
EDU***	.275	.026	10.6	0.00
EXP***	.178	.018	9.91	0.00
R&D***	.074	.013	5.54	0.00
L***	-.149	.021	-7.05	0.00
POP***	-.20	. 025	-8.32	0.00
MOB_BB***	.166	.039	4.24	0.00
Model 3: Coexistence of fixed-line and mobile broadband				
Constant***	5.86	.353	16.6	0.00
K***	.334	.016	22.1	0.00
EDU***	.170	.032	5.35	0.00
R&D***	.057	.013	4.27	0.00
L***	-.127	.021	-6.09	0.00
FL_BB***	.102	.019	5.46	0.00
MOB_BB***	.137	.038	3.56	0.00
POP***	-.192	.025	-7.79	0.00
EXP***	.165	.018	9.39	0.00

*Significant at 10% level, **significant at 5% level, ***significant at 1% level
Abbreviations: Std Err, Standard Error; *t*, t-statistic; P>|t|, probability of a value greater than absolute value of *t*.

All three models confirm a positive association between broadband and GDP per capita. In particular:

https://doi.org/10.18080/jtde.v12n4.1128

- Model 1 confirms that, on a standalone basis, increased adoption of fixed-line broadband increases GDP per person employed. The value of the estimated coefficients suggests that a 1 per cent increase in fixed broadband subscriptions per 100 people results in an average increase of 0.111 per cent in GDP per person employed. This finding is consistent with the findings of several previous studies (for example, Atif *et al.*, 2012).
- Consistent with expectations and previous empirical studies, Model 2 reveals a strong association between the increased take-up of mobile broadband and GDP per employee. Specifically, the estimated coefficients suggest that over the period a 1 per cent increase in the total number of mobile broadband subscriptions per 100 people resulted in an average increase of 0.166 per cent in GDP per person employed. This finding is consistent with those of Deloitte (2012) and Eisenach & Kulick (2020).
- Model 3 confirms that, in the presence of fixed-line broadband, increased adoption of mobile broadband is associated with higher economic output. Likewise, in the presence of mobile broadband, increased take-up of fixed-line broadband has a positive impact on GDP per employee. In this regard, the estimated coefficients suggest that a 1 per cent increase in the take up of fixed-line broadband increases GDP per employee by 0.102 per cent, compared to a 1 per cent increase in mobile broadband, which increases GDP per employee by over 0.137 per cent.

A holistic review of the model results reveals several interesting findings. First, a comparison of Models 1 and 2 suggests that, over the period, mobile broadband had a greater impact on economic output than fixed-line broadband. This might be explained by several reasons, including diminishing returns from fixed-line broadband given that it is more established than mobile broadband, or perhaps that either the lower costs to deploy mobile broadband infrastructure or the existence of a 'mobility premium' results in a greater productivity impact and therefore a greater impact on economic output. Unfortunately, the extent to which these factors explain why, over the period, mobile broadband had a 50 per cent greater impact on economic growth than fixed-line broadband is beyond the scope of this study.

A second issue to consider is that the extent to which increased mobile broadband positively impacts economic output is significantly greater on a stand-alone basis (as per Model 2) relative to when fixed-line broadband is available (as per Model 3). In particular, the economic impact of mobile broadband is more than 20 per cent greater when fixed-line broadband is excluded from the model. While this may arise because of omitted variable bias, the high R^2 statistic and F-statistic for both Model 2 and Model 3 suggests otherwise.[viii] Another explanation could be that there is a degree of substitutability between mobile and fixed-line broadband services.

https://doi.org/10.18080/jtde.v12n4.1128

Empirical evidence of substitutability between mobile and fixed broadband raises two important questions: (1) is substitutability supported by observed behaviour in the market? and (2) what is the nature and extent of any substitutability? In respect of the first question, there are many online applications and use cases in which a mobile broadband connection could substitute for a fixed-line connection. Indeed, given that mobile broadband provides high-speed access to the Internet, there is scope for it to serve as a substitute for fixed-line broadband in areas that are under-served or not served by fixed broadband. Such a scenario would not be dissimilar to those observations by Waverman *et al.* (2005) in respect of the role that mobile telephony played in developing nations without a fixed-line telephone network. In respect of the second question, the estimated value of the co-efficient for the fixed-line broadband variable in Model 1 versus Model 3 suggests that the contribution of fixed-line broadband to economic output is similar with, or without, the presence of mobile broadband. This suggests that, while mobile broadband may serve as a substitute for a fixed-line broadband connection, fixed-line broadband is not a substitute for mobile broadband.

Evidence that mobile broadband is a substitute for fixed-line broadband, but that the reverse does not hold, may also lend weight to the concept of a *mobility dividend*. This concept implies that mobile broadband offers unique advantages over fixed-line broadband, including the ability to access the Internet on the go. Related to this is the fact that, because mobile networks can be deployed more quickly and at a lower cost compared to fixed-line networks, they may have a bigger impact on economic output and productivity growth.

Table 2. Estimation results for Random Effects model

GDP	Coefficient	Std Err	z	P>|z|
Model 4: Inclusion of COVID-19 dummy variable				
Constant***	5.99	.341	17.6	0.00
K***	.331	.015	22.6	0.00
EDU***	.171	.032	5.40	0.00
R&D***	.059	.013	4.45	0.00
L***	-.127	.0218	-6.11	0.00
FL_BB***	.101	.018	5.58	0.00
MOB_BB***	.124	.037	3.37	0.00
EXP***	. 163	.017	9.43	0.00
POP***	-.191	.024	-7.89	0.00
Dummy**	-.036	.017	-2.10	0.04

*Significant at 10% level, **significant at 5% level, ***significant at 1% level
Abbreviations: Std Err, Standard Error; *z*, z-statistic (used to test the null hypothesis that a coefficient is equal to zero); P>|z|, probability of a value greater than absolute value of *z*.

Table 2 reports the results of the estimated random effects model to test the extent to which the availability of fixed-line and mobile broadband allowed nations to mitigate the harmful economic effects of the COVID-19 pandemic. As anticipated, the estimated coefficient for the

https://doi.org/10.18080/jtde.v12n4.1128

dummy variable is both statistically significant at the 5 per cent level and exhibits the expected sign, suggesting that the COVID-19 pandemic had a negative impact on GDP per worker.

The results also suggest that, when controlling for the effects of the COVID-19 pandemic, both fixed-line and mobile broadband continue to have a positive impact on economic output per worker. Specifically, a 1.0 per cent increase in mobile broadband would lift GDP per worker by 0.124 percentage points, and a 1.0 per cent increase in fixed broadband increases GDP per worker by 0.101 percentage points. These results are consistent with the estimated coefficients calculated using the Fixed Effects method (see Model 3 above) and the findings of Katz & Jung (2022).

Despite the estimated coefficients confirming the positive association between the adoption of both fixed-line and mobile broadband and economic output per worker during the period of the COVID-19 pandemic, they do not say anything about the actual extent to which changes in the adoption of fixed-line and mobile broadband impacted GDP throughout this period. This is because, for either fixed or mobile broadband to have had a positive impact on GDP per worker during the period, there needs to have been an increase in either fixed-line and/or mobile broadband take-up.

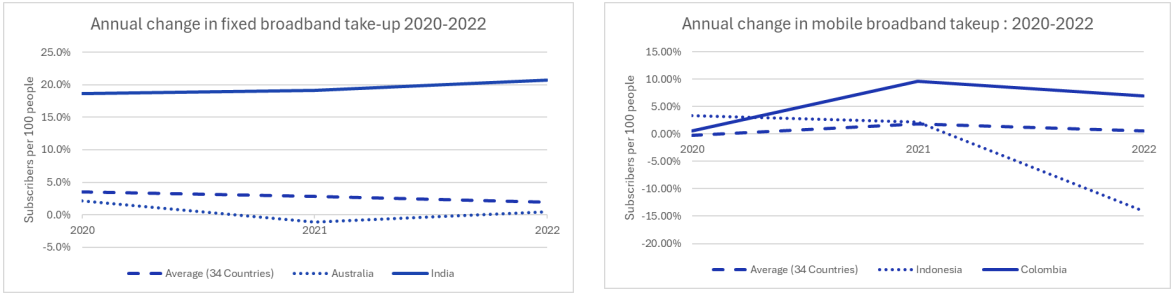


Figure 1. Annual growth in take-up rates for fixed and mobile broadband 2020 to 2022

Figure 1 shows the annual average growth in the take-up of fixed-line and mobile broadband for all 34 countries for the period 2020 to 2022; it also details the growth rate of the country that experienced the highest increase in fixed-line and mobile broadband take-up, as well as the country that experienced the lowest. Of note, all countries witnessed increased take-up of fixed-line broadband over the period, with cumulative increased take-up ranging from 1.4 per cent for Australia to 70.7 per cent for India; the average for all countries in the sample was a cumulative increase of 8.5 per cent over the period 2020 to 2022.

By contrast, 14 countries in the sample witnessed a reduction in take-up rates for mobile broadband over the period. Furthermore, relative to the average increase in take-up of fixed broadband, the average cumulative increase for mobile broadband across all countries was significantly less at 2.2 per cent; Indonesia experienced a 9.2 per cent reduction in mobile subscriptions over the period, while Columbia experienced an 18 per cent increase.

https://doi.org/10.18080/jtde.v12n4.1128

This data allows for the calculation of the annual percentage point change in GDP for each of the years 2020 to 2022 arising from changes in the take-up of fixed-line and mobile broadband for each of the 34 countries in the sample, as well as on average for all countries in the sample. It also allows for the calculation of the cumulative total percentage point impact on GDP over the COVID-19 pandemic period (i.e., 2020–2022). Appendix A details the results of these calculations.

These results suggest that for most countries, but not all, increased adoption of broadband during the period 2020 to 2022 resulted in higher levels of economic output (as measured by GDP per worker), both on an annual basis and for the period. This was mostly due to increased take-up of fixed-line broadband during the pandemic years, which allowed users to work from home and continue to engage in economic activity despite the restrictions imposed by many national governments and the disruption to supply chains and economic activity. For some countries, increased adoption of mobile broadband also contributed to higher levels of economic output, which was in addition to the economic benefits generated by fixed-line broadband.

Of the 34 countries included in the sample, four countries experienced lower output per worker over the period due to changes in the take-up of fixed-line and mobile broadband services; these countries are Switzerland, Australia, New Zealand and the Netherlands. All four of these countries experienced an increased take-up of fixed-line broadband services over the period, which resulted in higher economic output; however, these benefits were offset by reduced economic output arising from a reduction in the take-up of mobile broadband. Several other countries in the sample also experienced an overall reduction in the take-up of mobile broadband, which in all cases was a drag on GDP per employee over the period but was insufficient to offset the increased GDP arising from higher take-up of fixed broadband.

The reduction in the take-up of mobile broadband during the pandemic merits further consideration. Nine countries (more than a quarter of the countries in the sample) experienced an overall net reduction in mobile subscribers per 100 people over the period. By comparison, a total of 25 countries (almost 75 per cent of the sample) experienced at least one annual reduction in the take-up of mobile broadband during the pandemic period; of note, 20 countries (approximately 60 per cent of countries) witnessed a reduction in the take-up of mobile broadband in the first year of the pandemic, with the average subscriber rate across all countries contracting by approximately 27 per cent in 2020.

Explanations for these large and widespread reductions in the take-up of mobile broadband include the following:

https://doi.org/10.18080/jtde.v12n4.1128

- COVID restrictions eroded the mobile premium, prompting users to cancel their mobile subscription: The imposition of travel restrictions, social distancing rules and work-from-home arrangements may have eroded the value of the so called "mobility premium" that some subscribers ascribe to mobile broadband. Notably, in 2020, over 90 countries-imposed lockdown measures that impacted nearly half of the global population. The year 2020 was also when the most significant declines in mobile subscriber rates occurred, with many countries experiencing a yearly drop in mobile broadband take-up rates. This explanation is consistent with the fact that, in many countries, as the initial lockdowns were lifted and people spent more time away from their homes, mobile take-up rates began to recover as consumers once again placed value on the "mobility premium" associated with a mobile broadband connection.
- There was a shift in user preference in favour of fixed-line broadband: An alternative explanation might be that, to support remote working arrangements, broadband users preferenced fixed-line connections over mobile broadband during the COVID-19 period. This change in preferences may reflect higher bandwidth requirements to support bandwidth intensive applications, such as remote learning, remote working and greater demand for online entertainment services during the pandemic. While this shift in consumer preferences goes some way to explain the increased take-up in fixed-line broadband during the 2020 to 2022 period, it does not explain why the reductions in mobile subscriptions, particularly in the first year of the pandemic, were so pronounced (i.e. in 2020 across all 34 countries, the average number of mobile subscribers per 100 inhabitants declined by almost 27 per cent compared to an average increase in the number of fixed broadband connections of 3.5 per cent).
- A third explanation could be that, in the face of economic uncertainty triggered by the pandemic, consumers and businesses reduced spending on communications goods and services, resulting in fewer mobile subscribers per 100 people. This could have resulted in a combination of households and businesses switching from a mobile broadband connection to a lower cost fixed-line broadband subscription as well as a reduction in the number of active mobile subscriptions. Unfortunately, reliable data which supports this proposition does not appear to be available.

Potential Policy Implications and Further Work

There are several important policy implications arising from this study. While the focus of this section is on the policy implications for Australia, it is noted that many advanced and developing nations face similar policy challenges to Australia, especially in respect of policy

https://doi.org/10.18080/jtde.v12n4.1128

interventions intended to promote the take-up of broadband services and increase digital inclusion.

While not controversial, an important finding of this study is confirmation that investments in both fixed-line and mobile broadband infrastructure have strong positive impacts on economic output. These impacts are both direct and indirect in nature; the more significant being that the adoption and use of broadband allows firms and consumers to better organise how they undertake economic activities, resulting in a more efficient allocation of resources. These indirect impacts are not once-off in nature but are ongoing. Hence, greater availability of higher speed and more reliable broadband services to support new online applications and services will continue to drive improved economic and social outcomes into the future. This provides a strong policy rationale in support of universal service policies and minimum coverage obligations.

The results support the notion that mobile broadband services provide users with a "mobility premium" which is not a feature of fixed-line broadband. This has important policy implications. First, to the extent that fixed-line and mobile broadband may be substitutes, this is limited to mobile broadband being a substitute for fixed broadband. To the extent that consumers value mobility, it is unlikely that they will perceive fixed-line broadband as an effective substitute for mobile broadband.

Second, given that mobile broadband has a bigger impact on GDP relative to fixed-line networks, government policies which aim to provide affordable broadband access in locations that are uneconomic to serve should favour mobile broadband over fixed-line broadband. This is because:

- All things being equal, increased take-up of mobile broadband will deliver a greater increase in GDP compared to fixed-line broadband. This incremental increase in GDP will offset, either partially or fully, the cost to government of subsidising broadband availability in otherwise uneconomic-to-serve locations.
- Because mobile networks can be deployed at lower cost than fixed-line networks, preferencing mobile infrastructure to serve uneconomic regions will maximise both productive and allocative efficiency. Productive efficiency will be maximised because the services will be delivered at least cost, while allocative efficiency will be improved due to a reduction in the deadweight loss of any consequential funding arrangements, such as higher taxes or an industry levy.
- Mobile networks can be deployed more quickly than fixed-line networks. Hence, the use of mobile networks to serve households and business in uneconomic-to-serve regions will allow those users to capture the benefits of improved broadband more

https://doi.org/10.18080/jtde.v12n4.1128

quickly. This will also bring forward the indirect economic impacts (i.e., increased economic output, increase employment and wages and improved productivity) that an increase in the adoption broadband will generate. It may also bring forward the realisation of quality-of-life benefits for the region, such as increased population, higher property values and improved social inclusion.

Third, further empirical work should be undertaken to understand the extent to which fixed-line and mobile broadband are effective substitutes. Evidence of effective substitutability, or otherwise, should not only inform future decisions in respect of economic regulation but also in respect of potential reform of the Universal Service Obligation (USO) and the Regional Broadband Scheme (RBS), as well as the design of consumer safeguards to support low-income consumers. It is important to note that further consideration of this issue should in no way call into question the wisdom of the Commonwealth's Government's investment in NBN Co. To the contrary, a better understanding of substitutability between fixed-line and mobile broadband will allow policymakers, investors and regulators to make more informed decisions about how to efficiently allocate resources across alternative technologies. Furthermore, the empirical results presented in the previous section provide robust evidence that, in the presence of mobile broadband, the take-up and use of fixed broadband services provides significant economic benefits in addition to those delivered by the mobile networks.

Limitations of this Study

It is worth noting two important limitations of this study. First, this study has relied on *the number of fixed-line broadband subscribers per 100 people* and *the number of mobile subscriptions per 100 people* as a measure of the take-up and use of fixed-line and mobile broadband, respectively. From a policy perspective this is an important limitation because it says nothing about how intensively subscribers use these services and for what purpose. Given that the indirect economic benefits of broadband arise not solely from its adoption but from its use, it follows that how households and businesses use fixed and mobile broadband is a key driver of economic benefits. Unfortunately, due to limitations in the available data, our study is unable to shed any light on how the different uses of broadband applications and services, by different groups of users, impacts economic outcomes. Further work is needed in this regard to inform current and emerging questions about the ongoing importance of fixed and mobile broadband in the presence of ongoing technological change and increased digitisation. For example, will the expected growth in the use of Artificial Intelligence (AI) by households and businesses increase the flow-on economic benefits from broadband networks?

A second limitation of this study, which also arises from the lack of available data, is that it offers no insight as to whether increased availability and adoption of higher-speed broadband

https://doi.org/10.18080/jtde.v12n4.1128

services – capable of data rates of more than 100 Mbps (often referred to as ultra-fast broadband) – will deliver incremental economic benefits. This is an important issue from a policy perspective, especially given that in many countries network operators will likely seek government support to fund the deployment of the required infrastructure and network upgrades. From an empirical perspective, it is also important as there are challenges to ensure that the benefits of sunk broadband investments are not double counted. Any double counting of the economic and social benefits that have arisen from sunk broadband investments to justify future investments in higher-speed broadband networks risks prematurely allocating resources to investments that may undermine the inherent economic benefits of broadband technology.

References

Alimi, I. A., Patel, R. K., Muga, N. J., & Pinto, A. (2023). Towards enhanced mobile broadband communications: A tutorial on enabling technologies, design considerations, and prospects of 5G and beyond fixed wireless access networks. *Applied Sciences*, *11*(21), 10427. https://doi.org/10.3390/app112110427

Allen Consulting Group. (2002). Built for business II: Beyond basic connectivity, the Internet and Australian business. Sydney. https://webarchive.nla.gov.au/awa/20021221050701/http://www.allenconsult.com.au/publications_research.php

Atif, S. M., Macdonald, J., & Endres, J. (2012). Broadband infrastructure and economic growth: A panel data analysis of OECD countries. Sydney. https://ssrn.com/abstract=2166167

Australian Industry Group. (2008). National CEO Survey — High Speed to Broadband: Measuring industry demand for a world class service. https://webarchive.nla.gov.au/awa/20110303164403/http://pandora.nla.gov.au/pan/125552/20110304-0001/www.aigroup.com.au/portal/binary/com.epicentric.contentmanagement.servlet.ContentDeliveryServlet/LIVE_CONTENT/Publications/Reports/2008/7122_CEO_Broadband_web.pdf

Bai, Y. (2017). The faster, the better? The impact of internet speed on employment. Information *Economics and Policy*, *40*, 21–25. https://doi.org/10.1016/j.infoecopol.2017.06.004

Briglauer, W., & Gugler, K. (2018). Go for gigabit? First evidence on economic benefits of (ultra-)fast broadband technologies in Europe. ZEW [Centre for European Economic Research], Discussion Paper No. 18-020. Available at https://www.econstor.eu/bitstream/10419/177828/1/1019312734.pdf

Briglauer, W., Durr, N., & Gugler, K. (2021). A retrospective study on the regional benefits and spillover effects of high-speed broadband networks: Evidence from German counties. *International Journal of Industrial Organization*, *74*, 102677. https://doi.org/10.1016/j.ijindorg.2020.102677

Campbell, S., Ruiz Castro, J., & Wessel, D. (2021, August 18). The benefits and costs of broadband expansion. Brookings. https://www.brookings.edu/articles/the-benefits-and-costs-of-broadband-expansion/

Carew, D., Martin, N., Blumenthal, M., Armour, P., & Lastunen, J. (2018). The potential economic value of unlicensed spectrum in the 5.9 GHz frequency band: Insights for allocation policy. RAND Corporation. https://www.rand.org/pubs/research_reports/RR2

Collins, P., Day, D., & Williams, C. (2007). *The economic effects of broadband: An Australian perspective*. Canberra: Research Statistics and Technology Branch, Department of Communications, Information Technology and the Arts. Accessed 24 May 2010 from http://www.oecd.org/dataoecd/29/9/38698062.pdf

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*. https://d1bcsfjk95uj19.cloudfront.net/files/Brookings-BroadbandDeployment.pdf

Czernich, N., Falck, O., Kretschmer, T., & Woessmann, L. (2011). Broadband Infrastructure and Economic Growth. *The Economic Journal*, *121*(552), 505–532. https://doi.org/10.1111/j.1468-0297.2011.02420.x

Deloitte. (2012). What is the impact of mobile telephony on economic growth? ictData.org. GSM Association. https://www.gsma.com/solutions-and-impact/connectivity-for-good/public-policy/wp-content/uploads/2012/11/gsma-deloitte-impact-mobile-telephony-econo

Deloitte. (2021). Broadband for all: charting a path to economic growth. https://www2.deloitte.com/content/dam/Deloitte/us/Documents/process-and-operations/us-charting-a-path-to-economic-growth.pdf

Eisenach, L., & Kulick, R. (2020). Economic impacts of mobile broadband innovation: Evidence from the Transition to 4G. American Enterprise Institute, Economics Working Paper 2020-05. http://dx.doi.org/10.2139/ssrn.3607196

Grimes, A., Ren, C., & Stevens, P. (2009). The need for speed: impacts of internet. Motu Economic and Public Policy Research. https://motu-www.motu.org.nz/wpapers/09_15.pdf

Hardy, A. P. (1980). The role of the telephone in economic development. *Telecommunications Policy*, *4*(4), 278–296. https://doi.org/10.1016/0308-5961(80)90044-0

Hasbi, M. (2022). Impact of very high-speed broadband on company creation and entrepreneurship: Empirical Evidence. *Telecommunications Policy*, *44*(3), 101873. https://doi.org/10.1016/j.telpol.2019.101873

Katz, R., & Jung, J. (2022). The role of broadband infrastructure in building economic resiliency in the United States during the COVID-19 pandemic. *Mathematics*, *10*(16), 2988. https://doi.org/10.3390/math10162988

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

Koutroumpis, P. (2019). The economic impact of broadband: Evidence from OECD countries. *Technological Forecasting and Social Change*, 148, 119719. https://doi.org/10.1016/j.techfore.2019.119719

Leo, J. G., & Clement, A. C. (2022). The effects of mobile broadband on economic growth in Nigeria. *Journal of Economics and Allied Research*, *7*(4), 71–85. https://jearecons.com/index.php/jearecons/article/view/257

OECD. (2007). Broadband and the Economy. Paris. https://one.oecd.org/document/DSTI/ICCP/IE(2007)3/FINAL/en/pdf

Pollock, A. J. (2016). GDP per-worker vs. GDP per-capita. rstreet.org. https://www.rstreet.org/commentary/gdp-per-worker-vs-gdp-per-capita/

Qiang, C. Z.-W., & Rossotto, C. M., & Kimura, K. (2009). Economic Impacts of Broadband. In The World Bank, *Information and Communications for Development 2009: Extending Reach and Increasing Impact*. Washington D.C.: The World Bank. https://api.semanticscholar.org/CorpusID:14700576

Shapiro, P. D. (1976). Telecommunications and Industrial Development. *IEEE Transactions on Communications*, *24*, 305–311. https://doi.org/10.1109/TCOM.1976.1093286

Shapiro, R. J., & Hassett, K. A. (2012). The Employment Effects of Advances in Internet and Wireless Technology: Evaluating the Transitions from 2G to 3G and from 3G to 4G. Washington D.C.: The New Policy Institute. https://www.sonecon.com/docs/studies/Wir

Waverman, L., Meschi, M., & Fuss, M. (2005). The impact of telecoms on economic growth in developing countries. The Vodafone Policy Paper Series, *2*(03), 10–24. Available from https://www.assignmentpoint.com/wp-content/uploads/2012/04/L_Waverman_Telecoms_Growth_in_Dev_Countries.pdf

Wieck, R., & Vidal, M. (2011). Investment in telecommunications infrastructure, growth and employment – recent research. *International Journal of Management and Network Economics (IJMNE)*, *2*(2). https://doi.org/10.1504/IJMNE.2011.043351

World Economic Forum. (2022, January 23). These 5 charts reveal the global economic outlook for 2022. Retrieved from https://www.weforum.org/stories/2022/01/global-economic-outlook-5-charts-world-bank/

World Health Organization. (2023). WHO Director-General's opening remarks at the media briefing – 5 May 2023. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing---5-may-2023

Appendix A: Calculated Changes in GDP from Broadband Take-up

Table A.1. Percentage point change in GDP due to changes in the take-up of fixed and mobile broadband 2020–2022

Country	2020	2021	2022	Total
Austria				
Fixed	0.32	-0.07	0.09	0.34
Mobile	-0.05	0.17	0.15	0.27
Total	0.27	0.10	0.24	0.61
Australia				
Fixed	0.21	-0.11	0.04	0.14
Mobile	-0.57	-0.08	0.30	-0.36
Total	-0.36	-0.20	0.34	-0.21
Belgium				
Fixed	0.27	0.35	0.27	0.89
Mobile	-0.03	0.17	0.09	0.23
Total	0.24	0.53	0.36	1.13
Brazil				
Fixed	0.98	1.42	0.80	3.20
Mobile	0.15	0.76	-0.44	0.48
Total	1.13	2.18	0.36	3.68
Canada				
Fixed	0.19	0.24	0.24	0.67
Mobile	-0.84	0.39	0.44	-0.01
Total	-0.65	0.63	0.68	0.66
Chile				
Fixed	0.84	1.28	0.34	2.46
Mobile	-0.16	0.61	-0.14	0.31
Total	0.68	1.90	0.20	2.77
Columbia				
Fixed	1.02	0.75	0.36	2.13
Mobile	0.08	1.20	0.86	2.13
Total	1.10	1.94	1.22	4.27
Costa Rica				
Fixed	0.90	0.61	0.39	1.90
Mobile	0.25	0.45	0.00	0.70
Total	1.15	1.06	0.39	2.60
Czechia				
Fixed	0.29	0.29	0.16	0.74
Mobile	-0.09	0.30	0.20	0.41
Total	0.20	0.59	0.35	1.14
Denmark				
Fixed	0.16	0.05	0.03	0.25
Mobile	-0.05	0.13	0.08	0.16
Total	0.11	0.18	0.11	0.40

https://doi.org/10.18080/jtde.v12n4.1128

Country	2020	2021	2022	Total
Finland				
Fixed	0.26	0.09	0.07	0.42
Mobile	-0.07	0.04	-0.05	-0.08
Total	0.19	0.12	0.03	0.34
France				
Fixed	0.28	0.27	0.13	0.68
Mobile	0.11	0.43	0.23	0.76
Total	0.39	0.70	0.36	1.45
Germany				
Fixed	0.27	0.18	0.17	0.62
Mobile	0.00	-0.13	-0.23	-0.36
Total	0.27	0.05	-0.05	0.26
Greece				
Fixed	0.42	0.49	0.09	1.00
Mobile	-0.42	0.17	-0.11	-0.36
Total	0.00	0.66	-0.02	0.63
India				
Fixed	1.89	1.93	2.10	5.92
Mobile	-0.09	-0.09	-0.20	-0.39
Total	1.79	1.84	1.89	5.52
Indonesia				
Fixed	1.32	0.53	0.76	2.61
Mobile	0.41	0.27	-1.74	-1.06
Total	1.73	0.80	-0.98	1.55
Ireland				
Fixed	0.27	0.32	0.15	0.73
Mobile	0.05	0.23	0.63	0.91
Total	0.32	0.55	0.78	1.64
Israel				
Fixed	0.31	0.05	-0.16	0.20
Mobile	0.38	0.03	1.04	1.45
Total	0.69	0.08	0.88	1.65
Italy				
Fixed	0.42	0.36	0.74	1.52
Mobile	-0.25	0.14	0.10	-0.01
Total	0.17	0.50	0.84	1.51
Japan				
Fixed	0.40	-0.03	0.27	0.64
Mobile	0.63	0.41	0.51	1.55
Total	1.03	0.38	0.79	2.19
Lithuania				
Fixed	0.19	0.19	0.21	0.59
Mobile	0.02	0.34	0.50	0.86
Total	0.20	0.53	0.71	1.45
Mexico				
Fixed	1.56	0.87	0.05	2.48
Mobile	0.00	0.29	0.06	0.35
Total	1.56	1.16	0.11	2.82

https://doi.org/10.18080/jtde.v12n4.1128

Country	2020	2021	2022	Total
Netherlands				
Fixed	0.05	0.08	0.22	0.35
Mobile	-0.25	0.23	-0.69	-0.72
Total	-0.20	0.31	-0.47	-0.36
New Zealand				
Fixed	0.19	0.07	0.23	0.49
Mobile	0.20	-0.93	0.08	-0.65
Total	0.39	-0.86	0.31	-0.15
Norway				
Fixed	0.51	0.14	0.18	0.82
Mobile	0.03	0.16	0.11	0.31
Total	0.54	0.30	0.29	1.13
Poland				
Fixed	0.70	0.41	0.14	1.25
Mobile	0.27	0.35	-0.01	0.61
Total	0.97	0.76	0.13	1.86
Portugal				
Fixed	0.48	0.38	0.39	1.26
Mobile	-0.07	0.63	0.37	0.93
Total	0.41	1.01	0.76	2.19
Slovak Republic				
Fixed	0.70	0.45	0.12	1.27
Mobile	-0.19	0.14	-0.30	-0.34
Total	0.51	0.60	-0.17	0.93
South Africa				
Fixed	0.30	2.91	1.41	4.62
Mobile	-0.28	0.44	-0.11	0.04
Total	0.02	3.34	1.30	4.66
Spain				
Fixed	0.32	0.28	0.14	0.73
Mobile	0.00	0.23	0.46	0.69
Total	0.32	0.50	0.60	1.43
Sweden				
Fixed	0.25	0.08	-0.06	0.27
Mobile	-0.22	0.10	0.07	-0.05
Total	0.03	0.18	0.01	0.23
Switzerland				
Fixed	-0.09	0.33	0.32	0.56
Mobile	0.04	-0.39	-0.39	-0.73
Total	-0.05	-0.05	-0.07	-0.17
United Kingdom				
Fixed	0.12	0.14	0.03	0.30
Mobile	-0.31	0.08	0.24	0.00
Total	-0.19	0.22	0.27	0.30
United States				
Fixed	0.56	0.36	0.06	0.98
Mobile	-0.17	0.28	0.33	0.44
Total	0.39	0.64	0.39	1.42

https://doi.org/10.18080/jtde.v12n4.1128

Country	2020	2021	2022	Total
		Average (34 Countries)		
Fixed	0.358	0.281	0.196	0.835
Mobile	-0.033	0.226	0.073	0.266
Total	0.325	0.507	0.269	1.101

Appendix B: Summary of Variables and Data Plots

Table B.1. Summary of the variables

Variable	Description	Obs	Mean	Std Dev.	Min	Max
GDP_PW	GDP per persons employed (constant 2021 PPP $)*	510	96170.04	39807.35	13054.71	246149.10
K^{ix}	Gross fixed capital formation (current US$) lagged by one year**	510	22.13552	4.404687	10.68743	54.27
L	Total labour force	510	3.85e+07	8.82e+07	1461249	5.54e+08
EDU	Gross percentage of school enrolment in tertiary education	510	67.97	22.49	15.81	150.20
MOB_BB	Mobile cellular subscriptions (per 100 people)	510	118.21	21.46	28.74	177.02
FL_BB	Fixed broadband subscriptions (per 100 people)	510	26.56	12.40	.41	49.55
POP	Total population	510	8.89e+07	2.27e+08	2794137	1.42e+09
R&D	Research and development expenditure (per cent of GDP)	510	1.77	1.10	.08	5.71
EXP	Exports (per cent of GDP)	510	42.32	24.16	10.08	137.09

* GDP is quantified using purchasing power parity (PPP) dollars, denoted as PPP $.
** Gross Fixed Capital Stock is quantified in current US dollars, representing a nominal measure of each country's gross fixed capital stock over time. This measure does not account for the effects of inflation or exchange rate fluctuations

https://doi.org/10.18080/jtde.v12n4.1128

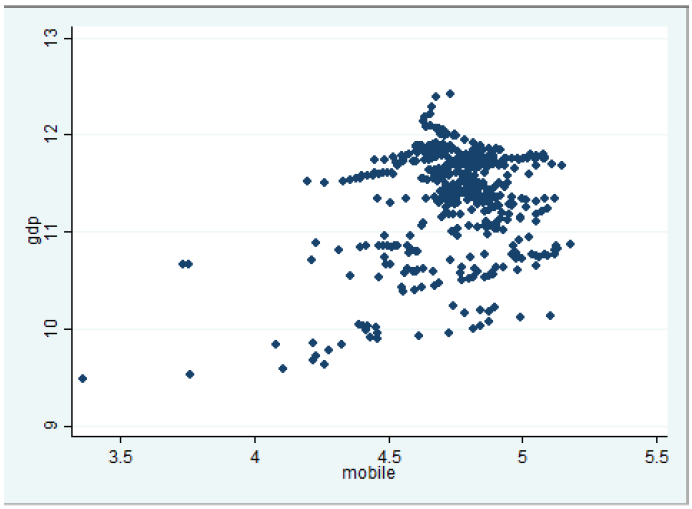


Figure B.1 Scatterplot: Mobile BB vs GDP

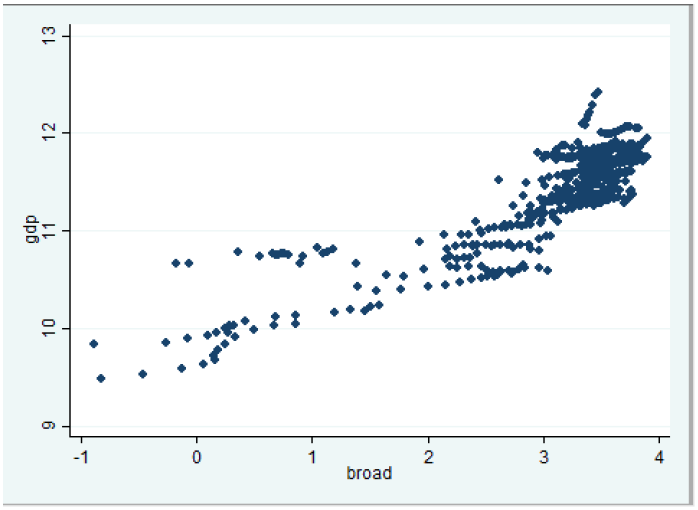


Figure B.2: Scatterplot: Fixed BB vs GDP

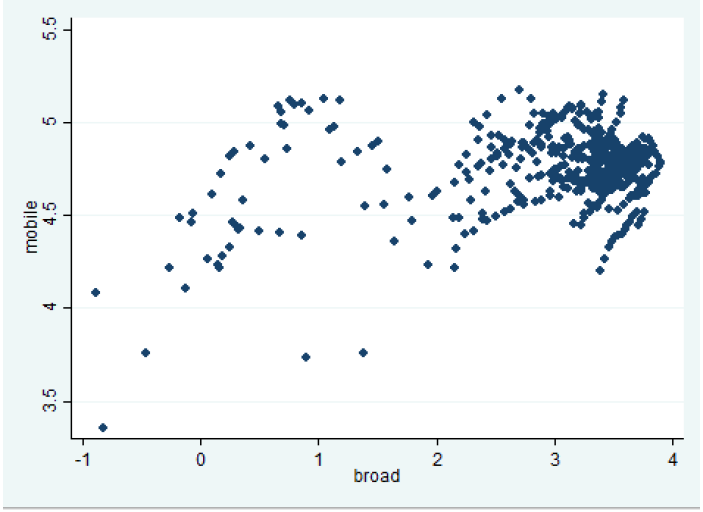


Figure B.3. Scatterplot: Mobile BB versus Fixed BB

https://doi.org/10.18080/jtde.v12n4.1128

Endnotes

[i] Countries included in the panel dataset are Austria, Australia, Belgium, Brazil, Canada, Chile, Colombia, Costa Rica, Czechia, Denmark, Finland, France, Greece, Germany, Indonesia, India, Ireland, Israel, Italy, Japan, Lithuania, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, South Africa, Spain, Sweden, Switzerland, United Kingdom and United States.

[ii] In 2020, Delaware had a broadband adoption rate of 91.4 percent compared to the US national adoption rate of 70.5 per cent.

[iii] In 2020 Arkansas had a broadband adoption rate of 70.5 per cent.

[iv] Regional externalities occur when the impact of a policy, investment, or economic activity in one region extends beyond its borders, influencing neighbouring regions. These externalities can be positive or negative and may include environmental impacts (i.e., increased air or water pollution), economic impacts (i.e., increased employment, higher property values or reduced wages) and social impacts (i.e., increased knowledge, increased regional migration).

[v] GDP per persons employed is used as it is a measure of labour productivity as opposed to overall economic wellbeing and adjusts for shifts in the composition of the population among those who are employed. See Pollock (2016).

[vi] The Random Effects estimation method was used to address the problem of perfect multicollinearity which was a consequence of including a dummy variable for the pandemic years. A Hausman test was performed to confirm that random effects is the appropriate estimation method for the data.

[vii] In late 2021 and early 2022 several countries imposed either partial or full lockdown measures due to the emergence of the Omicron variant.

[viii] The overall R^2 Statistics for Model-1, Model-2 and Model-3 are 0.919, 0.916 and 0.928, respectively. The F-Statistics for Model-1, Model-2 and Model-3 are all statistically significant at the 1% level.

[ix] The choice to use the lagged Gross Fixed Capital Formation (current US$) stems from the high correlation between gross fixed capital formation and GDP (constant 2021 PPP $) per person employed. Adopting a lagged version of Gross Fixed Capital Formation (using data from 2007–2021) mitigates the issue of contemporaneous correlation between these variables and helps address potential endogeneity, enabling the capture of long-term effects.

https://doi.org/10.18080/jtde.v12n4.1128

Decoding Consumer Behaviour in Indonesian E-Commerce: A Stimulus-Organism-Response Analysis

Kin Meng Cheng
Game Studies Department, Faculty of Creative Industries, Universiti Tunku Abdul Rahman, Malaysia

Lianna Wijaya
Management Department, BINUS Online Learning, Bina Nusantara University, Indonesia

Kah Choon Ng
Geography Department, Faculty of Arts and Social Sciences, University of Malaya, Kuala Lumpur, Malaysia

Lilian Anthonysamy
Faculty of Management, Multimedia University, Cyberjaya, Malaysia

Abstract: Indonesia holds a significant position in Southeast Asian e-commerce, witnessing notable shifts in consumer shopping patterns owing to the rapid expansion of online markets. This study concentrates on understanding e-commerce dynamics within Indonesian suburban areas. This research aims to unravel the hindering factors in Indonesian e-commerce, employing the Stimulus-Organism-Response framework to decode the intricate relationships between stimuli, organisms, and responses. This study endeavours to comprehend the intricate perceptions and behavioural tendencies exhibited by a cohort of 280 individuals by implementing a targeted quantitative questionnaire. The investigation seeks to scrutinize the impact of security, privacy, perceived risk, and cognitive trust on the attitudes and behavioural intentions of suburban consumers engaging in e-commerce activities. Employing SmartPLS as the analytical framework and utilizing Partial Least Squares Structural Equation Modelling (PLS-SEM), the research aims to investigate the causal relationships among key factors in the Indonesian suburban e-commerce industry. With a focus on the suburban population, this empirical study will offer comprehensive insights into the intricate dynamics of e-commerce. The findings are expected to contribute meaningful insights to the growing domain of e-commerce research and offer practical recommendations to foster the development of a resilient and thriving e-commerce ecosystem in Indonesia.

Keywords: E-commerce, Hindering Factors, Stimulus-Organism-Response, Perceived Risk, Cognitive Trust

https://doi.org/10.18080/jtde.v12n4.1009

Introduction

During the period of digital transformation, e-commerce has become a fundamental aspect of contemporary business, altering traditional retail models and impacting how consumers engage with online platforms (Dahbi & Benmoussa, 2019). The digital era has brought about significant changes in how commerce is performed, with e-commerce playing a crucial role in influencing consumer behaviours and market dynamics (Kim, 2019). Indonesia, known for its extensive group of islands and rapidly growing online population, is leading the way in the e-commerce revolution in Southeast Asia (Ma'Mun *et al.*, 2023; Sensuse *et al.*, 2020). The growth of online business has been driven by the growing popularity of smartphones, increased Internet accessibility, and the rise of the middle class over the past decade (Ariansyah *et al.*, 2021). Nevertheless, within this prosperous sector, there exists an intricate interplay of elements that impede the smooth advance of e-commerce (Kurniawan *et al.*, 2021). Within the dynamic and varied Indonesian market, the path of e-commerce expansion is characterised by exceptional prospects and obstacles. This study seeks to explore the complexities of the digital environment, with a specific emphasis on the obstacles that impact customer participation in e-commerce.

Indonesia is a significant participant in the rapidly changing Southeast Asian market, experiencing a notable increase in online commercial activity due to reasons (Lynn *et al.*, 2020). The e-commerce industry has been undergoing substantial expansion in the past decades (Kim, 2019). Due to the extensive adoption of smartphones, mobile commerce is of utmost importance in the e-commerce industry of Indonesia. A significant number of consumers utilise mobile applications to access e-commerce platforms (Ma'Mun *et al.*, 2023). The key e-commerce giants in Indonesia comprise Tokopedia, Bukalapak, Shopee, and Lazada (Chong & Ali, 2022; Lynn *et al.*, 2020). These platforms provide a diverse selection of products, encompassing electronics, fashion, beauty products, and other items. Key challenges in the Indonesian e-commerce market are platform competition, infrastructure development requirements, and catering to the varied needs of the people dispersed across the islands (Ariansyah *et al.*, 2021).

Indonesia has one of the highest rates of e-commerce penetration in the world in 2020. Up to 90% of Internet users aged between 16 and 64 purchase online. In the same year, Indonesia's gross market value (GMV) for e-commerce was above USD 30 billion, and it is projected to reach USD 83 billion in 2025 due to shifting consumer behaviour among tech-savvy consumers who are prepared to pay more for convenience (International Trade Administration, 2023). In 2024, Indonesia as the eighth-largest e-commerce market, behind only India and ahead of Canada, with a predicted revenue of USD 94.85 million, a market

https://doi.org/10.18080/jtde.v12n4.1009

growth of 32%, and a contribution to the worldwide growth rate of 29.8% (ECDB, 2024). Much of the promise in Indonesia's e-commerce sector has not been realised despite growing competition, particularly in places outside of Java Island. The Indonesian government has recognised the importance of the digital economy and e-commerce. Regulatory frameworks were developed to address issues such as data privacy, consumer protection, and taxation, providing a more stable environment for e-commerce growth.

The growth of e-commerce in Indonesia has been influenced by a range of factors including financial, technological, cultural, organisational, and economic considerations (Dahbi & Benmoussa, 2019). The process is complex and driven by various factors, such as technical breakthroughs, evolving consumer behaviour, legal frameworks, and the entrepreneurial drive of both local and foreign participants. The landscape is constantly changing, positioning itself as a significant participant in the e-commerce sector of Southeast Asia (Kurniawan *et al.*, 2021). The e-commerce sector in Indonesia is being shaped by ongoing technology breakthroughs, such as enhancements in logistics and payment systems. These advances offer potential for additional growth and innovation (Ariansyah *et al.*, 2021).

Although experiencing significant expansion, the Indonesian e-commerce industry encounters obstacles that hinder its smooth progress (Sensuse *et al.*, 2020). This study aims to investigate the elements that impede consumer engagement behaviour in the e-commerce industry, using the Stimulus-Organism-Response paradigm for a thorough analysis (Primadewi & Fitriasari, 2022). Previous research has thoroughly examined the aspects that contribute to the success of e-commerce. However, there is a significant lack of knowledge regarding the factors that obstruct optimal consumer engagement in the Indonesian e-commerce ecosystem. This study aims to fill this void by using the Stimulus-Organism-Response (S-O-R) framework, a widely recognised psychological model, to methodically examine the complex dynamics involved (Kurniawan *et al.*, 2021).

This empirical research seeks to achieve the following objectives: (1) Identify and categorise hindering factors in Indonesian e-commerce; (2) Understand how the organism (perceived risk and cognitive trust) is impacted by the stimulus (data security and privacy) and how this influences the response (attitude and behaviour intention); and (3) Provide actionable insights and recommendations for stakeholders in the e-commerce ecosystem. To achieve these objectives, a literature review is necessary to examine and comprehend the complex web of stimuli, individual characteristics, and responses in the context of e-commerce. In the following sections, the literature surrounding the variables of impact of security, privacy, perceived risk, and cognitive trust on the attitudes and behavioural intentions of suburban consumers engaging in e-commerce activities will be examined.

https://doi.org/10.18080/jtde.v12n4.1009

Theoretical Background and Hypothesis Development

Previous research on Stimulus-Organism-Response (S-O-R) framework

The S-O-R (Stimulus-Organism-Response) Framework, introduced by Mehrabian and Russell in 1974, outlines a triadic model encompassing an independent stimulus, a mediating organism, and a dependent response (Mehrabian & Russell, 1974; Vieira, 2013). This conceptual structure illustrates how the generated environment (Stimulus) influences the emotional state of a customer (Organism), subsequently prompting a behavioural response (Response) (Goi *et al.*, 2014; Mehrabian & Russell, 1974). The model serves to portray the intricate interplay between external stimuli, individual perception, and resulting behaviours, comprehensively capturing the dynamics of these elements.

Jiménez-Jiménez & Sanz-Valle (2011) asserted that the behaviour change process is a learning process. It involves: 1) Presentation of a stimulus, either accepted or rejected, influencing attention; 2) Accepted stimuli lead to understanding; 3) The organism processes the stimulus, fostering a willingness to act; 4) Environmental support influences the individual's action. Studies applying the model to retail reveal that environmental stimuli impact consumers' internal states, influencing behaviour in-store or online. In online retail, stimuli encompass design features (Alotebi *et al.*, 2018). Internal states include emotions and cognitive evaluations. Responses manifest themselves as various consumer behaviours (Alghizzawi, 2019). The S-O-R model is apt for this study due to its widespread application in online consumer behaviour research. Priporas *et al.* (2017) applied the model, examining the effects of task and mood on cognitive and emotional experiences and subsequent online buying behaviour.

Previous research on e-commerce success factors

Previous research employed Partial Least Squares Structural Equation Modelling (PLS-SEM) to analyse a survey of 611 Indonesian consumers influenced by YouTube reviews, introducing a model from the stimulus-organism-response paradigm. It detailed how sensory marketing and information adoption impact parasocial interaction, trust and information usefulness, correlating with consumer responses (Huang *et al.*, 2022). Fikri *et al.* (2019) studied 331 Indonesian respondents on fresh produce repurchase intentions, finding a notable connection between emotion and perceived risk influencing repurchase intentions. Surprisingly, the e-retailer reputation did not impact perceived risk. The study emphasised perceived risk and emotion in Indonesian e-shoppers' fresh produce repurchase willingness.

https://doi.org/10.18080/jtde.v12n4.1009

Hindering factors in global e-commerce studies

E-commerce adoption is influenced by its advantages over traditional commerce, and alignment with individual needs and values. Essential for expansion are digital skills and widespread infrastructure. Additionally, exposure to harmful content hinders adoption. A literature review on global e-commerce hindering factors of the e-commerce industry pinpointed technology, environment and customer trust as primary challenges, while benefits include the environment, organisation, and performance. Interconnected challenges and benefits emphasise the importance of managerial perception and understanding for successful e-commerce adoption (Hendricks & Mwapwele, 2024). Privacy concerns impede the e-commerce industry. A United Arab Emirates consumer study revealed that factors influencing online transaction intentions include perceptions of Internet safety, acceptance of e-commerce, privacy concerns and personal interests (Akour *et al.*, 2022).

Specifics of Indonesian e-commerce hindering factors

The influence of online platform security on consumers' perception of risk in e-commerce is significant. Research conducted by Tsai & Yeh (2010) and Bojjagani *et al.* (2023) demonstrated that implementing strong security measures, such as encryption, secure payment gateways, and authentication systems, significantly reduces consumers' worries over the safety of their transactions. These methods instil confidence in users regarding the security of their data and financial information, hence mitigating perceived risks associated with potential threats like data breaches and fraud. The enhancement of security significantly impacts consumers' confidence and trust, affecting their preference for online transactions. As a result, enhanced security not only reduces perceived dangers but also greatly facilitates the adoption of e-commerce.

The security procedures of online shopping platforms have a significant impact on consumers' cognitive trust. These platforms ensure reliability and integrity by protecting user data, assuring secure transactions, and implementing authentication methods. By using encryption and robust login protocols, users' confidence is bolstered while their personal and financial information is safeguarded. Effective communication regarding security protocols, trust seals, and timely responses to security concerns enhance cognitive trust. The trust necessary to maintain consumer loyalty arises from the perception of the platform's competency and commitment to safeguarding users in the e-commerce industry (Tran & Nguyen, 2022). The strong security measures implemented by online purchasing platforms have a profound impact on consumers' reasonable perceptions of dependability, hence cultivating trust and assurance in their engagements with these platforms.

https://doi.org/10.18080/jtde.v12n4.1009

The importance of ensuring the security of online shopping platforms to the development of cognitive trust in e-commerce is emphasised (Lăzăroiu *et al.*, 2020). The use of stringent security protocols, such as encryption, robust authentication, and defense mechanisms against cyber threats, significantly enhances the level of trust that customers place in these platforms. The platform's ability to protect personal information and guarantee secure transactions instils a sense of reliability. Open and clear communication regarding security measures and effective handling of security issues enhance cognitive trust. In the end, a secure online environment is crucial in influencing consumers' cognitive impressions and building a solid foundation of trust in e-commerce platforms.

The Indonesian government has implemented measures to strengthen the security and supervision of online shopping platforms through the enactment of laws and regulations. The Electronic Information and Transactions Law (UU ITE) is a crucial framework that provides legal grounds for addressing cyber threats and safeguarding data (Lubis *et al.*, 2024). Partnerships with industry stakeholders facilitate the establishment and enforcement of standards for secure online transactions. The National Cyber and Crypto Agency (BSSN) places high importance on cybersecurity, making a substantial contribution to the establishment of a more secure digital environment (Rahayu, 2018). The government's focused endeavours underscore its commitment to protecting consumers and fostering a safe digital marketplace in Indonesia.

Gaps and research questions

With the literature above regarding different aspects of shopping styles and behaviour in Indonesia, four research questions and 10 hypotheses were developed through the lens of S-O-R theory.

Research Questions:

RQ1: To what extent does online platform security influence the perceived risk of the consumers?

RQ2: To what extent does online shopping platform security impact the cognitive trust of consumers?

RQ3: To what extent does online platform security influence the cognitive trust of consumers?

RQ4: To what extent does the Indonesian government's enforcement of laws and regulations enhance the security of e-commerce platforms?

Hypotheses:

H1: Security influences Perceived Risk

H2: Privacy influences Perceived Risk

H3: Security influences Cognitive Trust

https://doi.org/10.18080/jtde.v12n4.1009

H4: Privacy influences Cognitive Trust

H5: Perceived Risk influences Attitude

H6: Cognitive Trust influences Attitude

H7: Perceived Risk influences Behaviour Intention

H8: Cognitive Trust influences Behaviour Intention

H9: Cognitive Trust influences Perceived Risk

H10: Attitude influences Behaviour Intention

Based on the above research questions and hypotheses, a conceptual framework, as shown in Figure 1, was developed. The framework consists of multiple interconnected hypotheses that investigate the interactions between crucial aspects within the setting of e-commerce. The analysis commences by investigating the distinct effects of security and privacy on the perception of risk (H1, H2). Furthermore, it explores the impact of security and privacy on cognitive trust, as indicated by hypotheses H3 and H4. The paradigm examines how perceived risk and cognitive trust influence attitude (H5, H6) and subsequently affect behavioural intentions (H7, H8). Moreover, it examines the reciprocal connection between cognitive trust and perceived risk (H9) and the influence of attitude on behavioural intentions (H10). This paradigm seeks to clarify the complex relationships between security, privacy, perceived risk, cognitive trust, attitude, and action intention in the field of e-commerce.

Research Methodology

This study utilised the S-O-R (Stimulus-Organism-Response) paradigm to examine the intricate dynamics of e-commerce in suburban areas of Indonesia, which were Tangerang, Bekasi, Karawang, Depok, and Bogor. The study aimed to investigate the effects of different factors (such as security, privacy, and perceived risk) on the attitudes, perceptions, and behaviours of suburban consumers who engage in e-commerce activities. This was done by distributing questionnaires to 280 individuals who were selected using cluster sampling. The paradigm enabled the examination of how external stimuli in the digital marketplace impact internal responses and affect behavioural intentions within this group. This methodology facilitated a thorough examination, investigating the cause-and-effect connections among important components while offering a systematic framework to analyse the intricate interconnections between stimuli, individual perceptions, and behavioural consequences within the specific context of Indonesian suburban e-commerce.

Population and sample

The study focused on persons involved in e-commerce operations, specifically in suburban regions of Indonesia. The sampling approach used was cluster sampling, which involved selecting specific geographic clusters or subdivisions within the suburban regions.

https://doi.org/10.18080/jtde.v12n4.1009

The researchers did not individually choose respondents inside these clusters; instead, they selected and sampled entire groups or clusters representing certain geographic locations. For example, the researchers could have selected particular suburban communities, districts, or zones recognised for their e-commerce activities. This method ensured the inclusion of a varied and inclusive sample from various pockets of suburban regions in Indonesia.

Suppose the researchers partitioned the suburban areas into several groups, such as communities or districts that are recognised for their involvement in online commerce. The researchers employed a random selection process from each cluster to include all persons who met the study's requirements, such as those actively engaged in e-commerce activities. The objective of this strategy was to encompass a wide range of experiences and viewpoints in the Indonesian suburban e-commerce industry, to gain a more thorough comprehension of the behaviours and preferences of this particular group of people.

Questionnaire design, measures, and data collection

The measurement items have been developed and established using previous research, presented in Table 1. Initially, the surveyed audience will be requested to provide their details, including gender, age, occupation, the most frequently used e-commerce, and personal reasons for using the apps. The measurement items for privacy, security, and cognitive trust consisted of 4 items each and were adapted to the e-commerce apps context from previous literature (Riquelme & Román, 2014). The measurement items of perceived risk consist of 4 items that were adapted in the e-commerce app context (Forsythe & Shi, 2003). For measurement items of Attitude and Behaviour Intention were adapted from previous literature (Ajzen, 1991).

Table 1. Measurement Items

Construct/Item	Measurement Items
Attitude (ATT)	
ATT1	I liked shopping via e-commerce apps.
ATT2	Using e-commerce apps for online shopping is a good idea.
ATT3	Using e-commerce apps for online shopping is a wise decision.
Behaviour Intention (BI)	
BI1	If I am satisfied and have a positive encounter with the e-commerce app, then I will use it in the future.
BI2	Near the future, I have the intention to make purchases on the e-commerce apps.
BI3	I have plans to shop on the e-commerce platforms in the future
Privacy (PR)	
PR1	When using E-commerce apps, you can be assured that your private data is securely protected.
PR2	The e-commerce apps used adequately protect information about payments and banks.
PR3	Only the relevant personal details that are required to finalize the transaction need to be provided.
PR4	The e-commerce app demonstrates its adherence to the rules and regulations that oversee the protection of data online.

https://doi.org/10.18080/jtde.v12n4.1009

Construct/Item	Measurement Items
Perceived Risk (RS)	
RS1	I feel risk the security of my credit card number will be ensured.
RS2	I feel risk the assurance that the confidentiality of my data shall be preserved.
RS3	Fears regarding the security of credit card information are a concern when entering it on E-commerce platforms.
RS4	The security of personal information is a concern for users.
Security (SC)	
SC1	The e-commerce apps have adequate online privacy.
SC2	The e-commerce apps provide secure options for making payments.
SC3	The e-commerce apps possess sufficient security measures.
SC4	Before making the payment in e-commerce apps, you have the option to verify the specifics of the transaction.
Cognitive Trust (TR)	
TR1	I am confident in the reliability of this e-commerce app's performance.
TR2	This e-commerce app is dependable when it comes to purchasing items online.
TR3	This e-commerce app is trustworthy.
TR4	The e-commerce app always ensures that customers are provided with a service policy when conducting transactions.

Data Analysis

Demographics of respondents

Table 2. Respondents' Demographics

Respondents' Profile	Frequency (N=280)	Percentage (%)
Gender		
Male	157	56%
Female	123	44%
Age		
Below 21 years old	136	48.6%
21–30 years old	95	33.9%
31–40 years old	24	8.6%
41–50 years old	14	5%
Above 50 years old	11	3.9%
Employment Status		
Full-Time Government	10	3.6%
Full-Time Private	65	23.2%
Students	102	36.4%
Entrepreneurs	27	9.6%
Retired	4	1.4%
Unemployed	72	25.8%
The most frequently used E-commerce		
Shopee	205	73.1%
Tokopedia	24	8.6%
Lazada	14	5%
Others	37	13.3%
Reason for Using E-commerce Apps		
Competitive Price	34	12.1%
Responsible E-commerce Operators	33	11.8%
Responsive Sellers	8	2.9%
Complete Products	98	35%
Promotion Offered	107	38.2%

https://doi.org/10.18080/jtde.v12n4.1009

In the constantly changing environment of Indonesian e-commerce in suburban areas, a vibrant combination of digital interaction emerges. The demographic composition as shown in Table 2 presents a dynamic image: an increasing presence of young individuals, with almost half of the population being under the age of 21, engaging actively with digital devices. Students make up more than one-third of the group, while the unemployed account for one-quarter of the participants. Amid this vibrant gathering, the gender distribution somewhat favours the daring nature of the male group, creating a dynamic blend of diversity and passion for technology.

However, amidst all these activities, one platform stands out as the primary organizer of this e-commerce event: Shopee. With significant recognition in the digital landscape, it has attracted over 70% of participants. This notable engagement may be attributed to effective marketing strategies that communicate appealing offers through various digital channels. These advertisements highlight the extensive selection of high-quality products available, emphasizing both convenience and variety.

In the online business environment, the themes of youth, opportunity, and convenience resonate strongly, creating a powerful message that spreads throughout the suburban areas of Indonesia. Engaging digital displays attract individuals who utilise their devices to participate in modern commerce and digital communication. The results of the survey based on the participants' selection of the items and scale are presented in Table 3. The survey results indicate that participants primarily possess favourable views and intentions towards the topic matter. A significant proportion agreed or strongly agreed with the statements regarding the Attitude (ATT) construct: 72.5% (n=203) for ATT1; 72.8% (n=204) for ATT2; and 64.3% (n=180) for ATT3. In the Behaviour Intention (BI) construct, positive intents are apparent, with 76.5% (n=214) agreeing or strongly agreeing with BI1, 65.3% (n=183) with BI2, and 69.3% (n=194) with BI3. Privacy (PR) issues elicited diverse responses; still, a clear majority expressed agreement, particularly with PR4, where 76.1% (n=213) agreed or strongly agreed. The replies on Perceived Risk (RS) were predominantly neutral; nevertheless, a substantial proportion expressed agreement or strong agreement—43.6% (n=122) for RS2, 56.8% (n=159) for RS3, and 56.8% (n=159) for RS4. The Security (SC) construct reflects favourable perceptions, with more than 60% of respondents agreeing or strongly agreeing on all items, particularly 75.4% (n=211) for SC4. Finally, Cognitive Trust (TR) ratings are elevated, with 73.6% (n=206) either agreeing or strongly agreeing with TR2, and comparable levels for other items, indicating a robust level of trust among participants.

https://doi.org/10.18080/jtde.v12n4.1009

Table 3. Summary of Survey Results (N= 280)

Item	Strongly Disagree n (%)	Disagree n (%)	Neutral n (%)	Agree n (%)	Strongly Agree n (%)
Attitude (ATT)					
ATT1	8 (2.9%)	9 (3.2%)	60 (21.4%)	100 (35.7%)	103 (36.8%)
ATT2	8 (2.9%)	9 (3.2%)	59 (21%)	100 (35.7%)	104 (37.1%)
ATT3	6 (2.1%)	12 (4.3%)	82 (29.3%)	112 (40%)	68 (24.3%
Behaviour Intention (BI)					
BI1	5 (1.8%)	6 (2.1%)	55 (19.6%)	97 (34.6%)	117 (41.9%)
BI2	4 (1.4%)	15 (5.4%)	78 (27.9%)	90 (32.1%)	93 (33.2%)
BI3	7 (2.5%)	10 (3.6%)	69 (24.6%)	94 (33.6%)	100 (35.7%)
Privacy (PR)					
PR1	10 (3.6%)	22 (7.9%)	80 (28.6%)	93 (33.2%)	75 (26.7%)
PR2	7 (2.5%)	13 (4.6%)	66 (23.6%)	109 (38.9%)	85 (30.4%)
PR3	14 (5%)	18 (6.4%)	73 (26.1%)	105 (37.5%)	70 (25%)
PR4	7 (2.5%)	13 (4.6%)	47 (16.8%)	110 (39.3%)	103 (36.8%)
Perceived Risk (RS)					
RS1	23 (8.2%)	36 (12.9%)	99 (35.3%)	70 (25%)	52 (18.6%)
RS2	12 (4.3%)	27 (9.6%)	84 (30%)	92 (32.9%)	65 (23.2%)
RS3	11 (3.9%)	29 (10.4%)	81 (28.9%)	88 (31.4%)	71 (25.4%)
RS4	11 (3.9%)	28 (10%)	82 (29.3%)	91 (32.5%)	68 (24.3%)
Security (SC)					
SC1	23 (8.2%)	23 (8.2%)	64 (22.9%)	63 (22.5%)	107 (38.2%)
SC2	6 (2.1%)	19 (6.8%)	76 (27.1%)	92 (32.9%)	87 (31.1%)
SC3	10 (3.6%)	8 (2.9%)	67 (23.9%)	114 (40.7%)	81 (28.9%)
SC4	9 (3.2%)	7 (2.5%)	53 (18.9%)	100 (35.7%)	111 (39.7%)
Cognitive Trust (TR)					
TR1	2 (0.7%)	12 (4.3%)	58 (20.7%)	111 (39.7%)	97 (34.6%)
TR2	3 (1.1%)	11 (3.9%)	60 (21.4%)	120 (42.9%)	86 (30.7%)
TR3	7 (2.5%)	11 (3.9%)	57 (20.4%)	118 (42.1%)	87 (31.1%)
TR4	7 (2.5%)	9 (3.2%)	61 (21.8%)	111 (39.7%)	92 (32.8%)

Measurement model

The evaluation of the measurement model reveals the strength and reliability of the concepts used in the study, providing insights into the connections between stimuli, perceptions, and behavioural responses in the context of e-commerce in Indonesian suburban areas. Table 4 presents the loading factors for connected items in critical domains, such as Attitude, Behaviour Intention, Privacy, Perceived Risk, Security, and Cognitive Trust, which show significant relationships, with values ranging from 0.639 to 0.906. These results indicate robust correlations between the items and their corresponding components, highlighting the significant impact these factors have on customer behaviour. The substantial composite reliability scores, which range from 0.855 to 0.925, emphasise the internal consistency within each construct, guaranteeing dependable measurements (Sarstedt *et al.*, 2019). Furthermore, the average variance extracted (AVE) metrics, as shown in Table 4, which range from 0.597 to 0.763, indicate

https://doi.org/10.18080/jtde.v12n4.1009

that a substantial amount of variance, approximately 59.7–76.3%, is accurately represented by the items within each construct (Fornell & Larcker, 1981). In this data, the VIF (Variance Inflation Factor) values for each item range from 1.260 to 2.716, indicating no significant multicollinearity, as all values are below the common threshold of 5 (O'Brien, 2007). Table 5 presents the discriminant validity of this study. This highlights the validity of the constructs in accurately depicting their intended meaning.

Table 4. Measurement Model Evaluation

Construct/Item	Loading Factor	Composite Reliability	AVE	VIF
Attitude (ATT)		0.867	0.685	
ATT1	0.803			1.573
ATT2	0.871			1.975
ATT3	0.806			1.507
Behaviour Intention (BI)		0.906	0.763	
BI1	0.879			1.955
BI2	0.854			2.023
BI3	0.888			2.144
Privacy (PR)		0.925	0.756	
PR1	0.772			1.702
PR2	0.846			1.954
PR3	0.771			1.605
PR4	0.818			1.645
Perceived Risk (RS)		0.899	0.692	
RS1	0.656			1.439
RS2	0.891			2.626
RS3	0.906			2.618
RS4	0.850			2.147
Security (SC)		0.878	0.644	
SC1	0.639			1.260
SC2	0.777			1.557
SC3	0.855			2.028
SC4	0.805			1.638
Cognitive Trust (TR)		0.855	0.597	
TR1	0.848			2.211
TR2	0.890			2.716
TR3	0.880			2.683
TR4	0.859			2.477

Table 5. Fornell-Larcker Criterion for Discriminant Validity Evaluation

Construct	ATT	BI	TR	RS	PR	SC
ATT	0.827					
BI	0.770	0.874				
TR	0.639	0.673	0.869			
RS	0.408	0.435	0.345	0.832		
PR	0.589	0.571	0.613	0.288	0.802	
SC	0.618	0.570	0.567	0.244	0.632	0.773

This review confirms the accuracy and effectiveness of the measuring methodology used to study the complex dynamics of e-commerce in suburban areas of Indonesia. The strong loading factors, high composite reliability scores, and significant average variance extracted across different constructs collectively validate the model's capacity

https://doi.org/10.18080/jtde.v12n4.1009

to capture and understand the interaction between stimuli, perceptions, and behavioural intentions in this distinct consumer environment. The measurement model path diagram is as shown in Figure 1. These findings confirm the constructs' impact on consumer decisions in e-commerce and establish a solid basis for comprehending and forecasting consumer behaviours and preferences in Indonesian suburban regions in the digital marketplace.

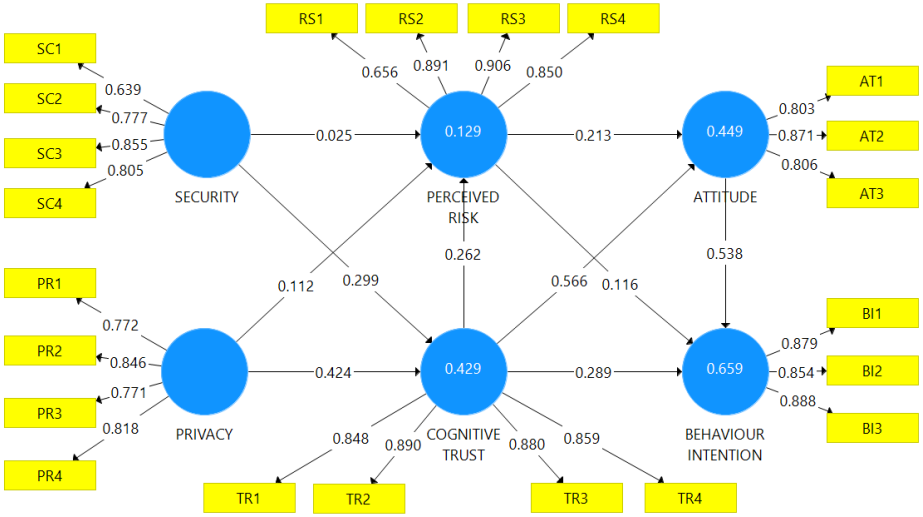


Figure 1. Measurement Model Path Diagram

Structural model

The evaluation of the structural model provides important insights into the connections between different elements within the e-commerce system in Indonesian suburbs. This evaluation uses parameters such as effect size (f^2), explained variance (R^2), predictive relevance (Q^2), and model fit (SRMR), presented in Table 6. Significant effect sizes are observed, specifically demonstrated by the strong impact of Cognitive Trust (TR) on Attitude (ATT) with a f^2 value of 0.512, confirming the influential role of trust in determining consumer attitudes. The R^2 values, particularly the 0.655 correlation between Cognitive Trust and Behavioural Intention (BI), show that Cognitive Trust explains about 65.5% of the variation in consumers' behavioural intentions in e-commerce. This suggests that while Cognitive Trust is a substantial factor, other influences likely play a role in the remaining 34.5% of cases. Furthermore, the positive Q^2 values indicate a correlation of 0.489 between Cognitive Trust and action Intention, suggesting a noteworthy relationship that may help predict customer behaviour in this context. The model's goodness-of-fit is demonstrated by the SRMR values, which are mostly below 0.08. For example, the SRMR value of 0.066 between Security (SC) and Perceived Risk (RS) indicates a strong alignment between the observed and anticipated covariances. This reinforces the validity and suitability of the structural model.

https://doi.org/10.18080/jtde.v12n4.1009

Table 6. Structural Model Evaluation

Construct	f^2	R^2	Q^2	SRMR
SC → RS	0.000	0.119	0.078	0.066
PR → RS	0.007			
SC → TR	0.094	0.425	0.317	
PR → TR	0.189			
RS → ATT	0.073	0.445	0.298	
TR → ATT	0.512			
RS → BI	0.032	0.655	0.489	
TR → BI	0.143			
TR → RS	0.045			
ATT → BI	0.468			

The effect sizes, explained variances, predictive relevance, and model fit observed all emphasise the significant influence of variables such as Cognitive Trust on shaping attitudes and behavioural intentions. Supported by established scholarly principles set forth by Chin (1998) and Hair *et al.* (2021), these findings affirm the model's robustness in capturing and elucidating the underlying relationships between stimuli, perceptions, and subsequent behavioural responses within the complex and evolving landscape of e-commerce in Indonesian suburban areas.

Hypothesis testing

Hypothesis testing in the context of structural equation modelling offers crucial insights into the validity and importance of interactions between constructs. The results presented in Table 7 show the evaluation of path coefficients, together with corresponding statistical metrics including standard deviation, T-values, and P-values. Additionally, Figure 2 shows the Structural Model path diagram; each path is labelled with a coefficient, representing the strength of the relationship between the variables.

Various hypotheses were examined to determine the connections between components in the context of e-commerce in Indonesian suburbs. Hypotheses 3, 4, 5, 6, 7, 8, 9, and 10 were supported, as indicated by substantial T-values (ranging from 2.518 to 9.250) and matching P-values of 0.000, confirming the statistical importance of these correlations. Hypothesis 6 (TR →ATT) shows a significant path coefficient of 0.566 with a small standard deviation of 0.061. This results in a high T-value of 9.250 and a statistically significant P-value of 0.000, demonstrating a strong and significant link between Cognitive Trust and Attitude. Hypothesis 10 (ATT →BI) shows a strong path coefficient of 0.538, with a low standard deviation and a high T-value of 9.132. The P-value of 0.000 indicates statistical significance, proving a strong connection between Attitude and Behavioural Intention in the e-commerce scenario.

Nevertheless, Hypotheses 1 and 2 did not receive support, as indicated by the non-significant T-values (0.275 and 1.289, respectively) and higher P-values (0.783 and 0.197, respectively)

https://doi.org/10.18080/jtde.v12n4.1009

that were above the conventional significance threshold of 0.05. The results suggest that there is not enough evidence to establish substantial connections between Security (SC) leading to Perceived Risk (RS) and Privacy (PR) leading to Perceived Risk (RS) in the e-commerce scenario being studied. These findings offer useful insights into the links between major dimensions, both supported and unsupported. They enhance our understanding of the processes that influence customer behaviour in Indonesian suburban e-commerce.

Table 7. Hypothesis Testing

Construct	Path Coefficient	StDev	T-value	P-value	Supported
H1: SC → RS	0.025	0.090	0.275	0.783	No
H2: PR → RS	0.112	0.087	1.289	0.197	No
H3: SC → TR	0.299	0.073	4.090	0.000	Yes
H4: PR → TR	0.424	0.079	5.387	0.000	Yes
H5: RS → ATT	0.213	0.060	3.560	0.000	Yes
H6: TR → ATT	0.566	0.061	9.250	0.000	Yes
H7: RS → BI	0.116	0.046	2.518	0.012	Yes
H8: TR → BI	0.289	0.058	4.949	0.000	Yes
H9: TR → RS	0.262	0.093	2.832	0.004	Yes
H10: ATT → BI	0.538	0.059	9.132	0.000	Yes

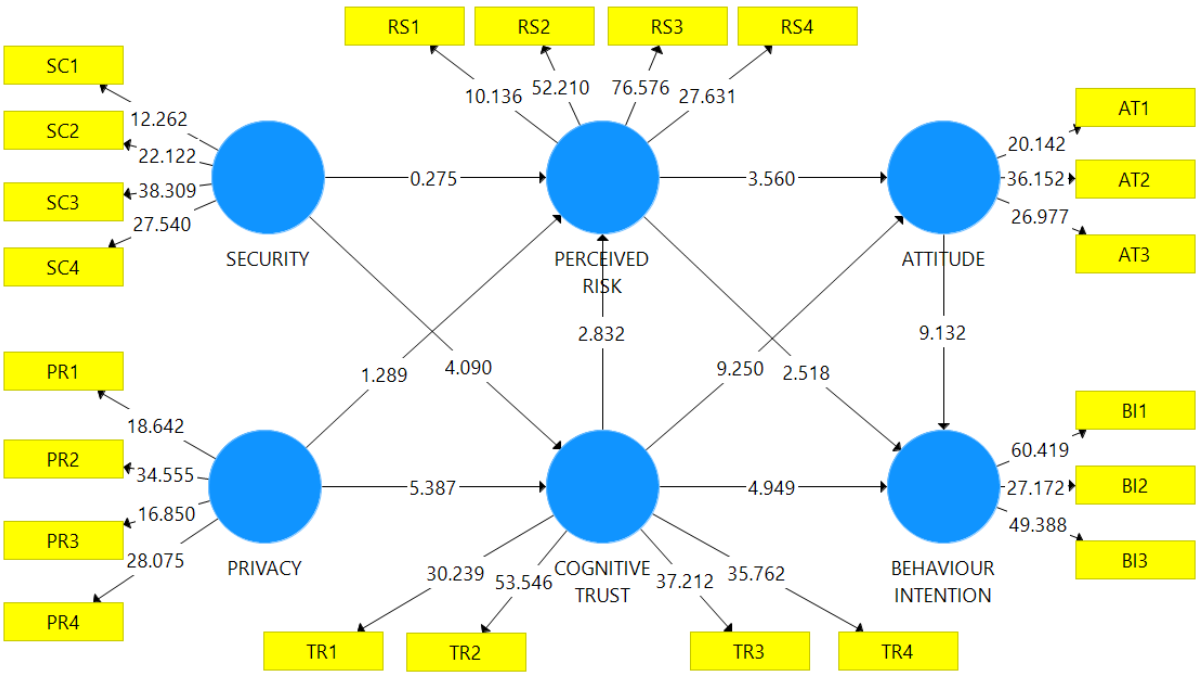


Figure 2. Structural Model Path Diagram

Discussion

Theoretical implications

The results of this study have important theoretical implications that are in line with the stated aims. They contribute to a better understanding of the obstacles in Indonesian e-commerce and the complex relationship between stimuli, the organism, and customer responses.

https://doi.org/10.18080/jtde.v12n4.1009

Firstly, the process of identifying and categorising limiting variables, particularly the non-supported linkages between Security (SC) →Perceived Risk (RS) and Privacy (PR) →Perceived Risk (RS), contributes to the theoretical understanding by emphasising the complex nature of these relationships. This highlights the necessity for a thorough examination of the various factors and intricacies related to data security and privacy issues concerning anticipated hazards in the Indonesian e-commerce industry.

Furthermore, the links that are supported, specifically those connecting Cognitive Trust (TR) and Perceived Risk (RS) to Attitude (ATT) and Behaviour Intention (BI), provide a theoretical understanding of the substantial influence of trust and perceived risk on influencing consumer attitudes and behavioural intentions. Within the context of e-commerce, customers' assessment of risk is shaped by various aspects that extend beyond mere security and privacy concerns. It is interesting to note that security (SC) and privacy (PR) concerns have no significant effect of perceived risk (RS), for which further insights require more investigation. The finding runs in contrast to the popular view that security and privacy concerns drive the perception of risk in e-commerce. A possible explanation is that Indonesian customers might consider other factors, such as platform familiarity, brand trust, and transaction benefits, to be more important than specific security and privacy concerns when assessing risk. This finding sets up a possible cultural or contextual variance in which trust and familiarity may mean more than the expected security and privacy problems. By challenging these theoretical models of privacy and security as primary contributors of risk, this study opens ways for further investigations into the contextual factors which shape perceived risk in e-commerce environments, such as cultural attitudes toward data handling or levels of digital literacy.

These considerations encompass trust, familiarity with the platform, prior experiences, and the perceived advantages associated with the transaction (Carstens *et al.*, 2019). These findings enhance theoretical models by highlighting the crucial roles of trust and perceived risk as factors that influence consumer responses in the e-commerce field.

Finally, to achieve the goal of offering practical insights and recommendations for stakeholders, the theoretical implications involve clarifying the key components that influence consumer behaviour. The established connections between constructs, such as Cognitive Trust (TR) impacting Attitude (ATT) and Behaviour Intention (BI), provide a theoretical basis for stakeholders to promote trust-building efforts and methods to reduce risks (Al-Hujran *et al.*, 2015). In addition, the interactions that are not supported provide insight into areas that need more research and intervention. This helps stakeholders to adopt more focused approaches in addressing data security and privacy problems, thereby creating a favourable e-commerce environment.

https://doi.org/10.18080/jtde.v12n4.1009

Essentially, these theoretical implications enhance our understanding of factors that hinder Indonesian e-commerce and the stimulus-organism-response framework. They also provide a theoretical foundation for stakeholders to strategically navigate and improve the e-commerce ecosystem by gaining a nuanced understanding of consumer behaviour and perceptions.

Practical implications

The Indonesian e-commerce industry faces substantial challenges in a fast-changing digital landscape. Some of them deal with a wide variety of issues, including user trust and confidence and broader concerns over market dynamics and regulatory frameworks. Although conventional wisdom and extant literature emphasize security as a key facilitator in building trust (Li & Wang, 2020), findings in this study suggest otherwise. Security and privacy issues are not viewed as major contributors to perceived risk in this context. Instead, other factors like cognitive trust and familiarity with the platform seem to hold greater influence in shaping consumer perceptions of risk.

Considering these findings, the practitioners must devise a strategy that will help them build consumer confidence through the development of cognitive trust and familiarity. The practitioners may emphasize clear communication regarding transaction benefits, designing user experience for familiarity, and ensuring consistent and reliable service quality. While security is key to guarding data integrity and meeting regulatory requirements, e-commerce platforms should be aware that Indonesian users have other priorities besides pure security features. Resource allocation needs, therefore, to balance both, investing in activities that build trust in the platform; this may call for a revisit of priorities with a view to matching local user perceptions. The most important thing it would underpin is the definite need for more focused approaches to enhancing consumer confidence, so that Indonesian e-commerce could apply more flexibility and consumer orientation to its strategies.

Limitations and future direction

The research faces limitations, with a modest sample size of 280 participants that may restrict the broad applicability of results. The cross-sectional design impedes the examination of long-term trends, and the reliance on self-reported data introduces potential response bias. While the Stimulus-Organism-Response framework is robust, it might not cover all pertinent factors. The study may not fully represent the cultural and regional diversity in Indonesia, and external variables influencing e-commerce were not extensively explored. The recommendations may not comprehensively address all hindering factors. Acknowledging these constraints emphasises the need for caution in generalising findings and points toward potential avenues for future research improvement.

https://doi.org/10.18080/jtde.v12n4.1009

Conclusion

Data obtained through hypotheses testing give relevant insights from some important relationships in the study. The accepted hypotheses indicated that TR and RS have significantly influenced ATT and BI in the e-commerce context. Such results bring into focus that a need exists to evoke trust and reduce the perceptions of risk as critical pre-requisites for any attempts to influence customer responses and build positive attitudes towards e-businesses.

However, some assumptions were not well supported, such as the respective hypothesized links of Security (SC) and Privacy (PR) with Perceived Risk (RS). Although security and privacy measures may not be as important in driving perceived risk for Indonesian e-commerce, insofar as consumers would prefer to rely upon other trust factors associated with familiarity and reliability. Given the above, building trust may be far more efficient and reach further than sophisticated security investments, thus making companies capable of coping better with the expectations set by local consumers.

This advocates for the efficiency of targeted trust-building strategies in shaping positive consumer responses and improving engagement in Indonesian e-commerce. The succeeding sections detail the results of hypothesis testing and variable decisions that firmly establish the interlinked relationship between trust, perceived risk, attitude, and behavioural intentions in the Indonesian e-commerce setting. These insights provide hands-on guidance to stakeholders that building trust and thereby understanding consumer preference will be rather efficient than investing heavily in sophisticated security measures. This would eventually lead to a positive and customer-oriented e-commerce environment in Indonesia.

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Akour, I., Alnazzawi, N., Alshurideh, M., Almaiah, M. A., Al Kurdi, B., Alfaisal, R. M., & Salloum, S. (2022). A conceptual model for investigating the effect of privacy concerns on E-commerce adoption: a study on United Arab Emirates consumers. *Electronics*, *11*(22), 3648. https://doi.org/10.3390/electronics11223648

Al-Hujran, O., Al-Debei, M. M., Chatfield, A., & Migdadi, M. (2015). The imperative of influencing citizen attitude toward e-government adoption and use. *Computers in Human Behavior*, *53*, 189–203. https://doi.org/10.1016/j.chb.2015.06.025

Alghizzawi, M. (2019). The role of digital marketing in consumer behavior: A survey. *International Journal of Information Technology Language Studies*, *3*(1), 24–31. https://journals.sfu.ca/ijitls/index.php/ijitls/article/view/67

Alotebi, H., Alharbi, O., & Masmali, A. (2018). Effective leadership in virtual learning

environments. *International Journal of Information and Education Technology*, *8*(2), 156–160. https://doi.org/10.18178/ijiet.2018.8.2.1026

Ariansyah, K., Sirait, E. R. E., Nugroho, B. A., & Suryanegara, M. (2021). Drivers of and barriers to e-commerce adoption in Indonesia: Individuals' perspectives and the implications. *Telecommunications Policy*, *45*(8), 102219. https://doi.org/10.1016/j.telpol.2021.102219

Bojjagani, S., Sastry, V. N., Chen, C.-M., Kumari, S., & Khan, M. K. (2023). Systematic survey of mobile payments, protocols, and security infrastructure. *Journal of Ambient Intelligence and Humanized Computing*, *14*(1), 609–654. https://doi.org/10.1007/s12652-021-03316-4

Carstens, M., Ungerer, M., & Human, G. (2019). Perceived risk, trust and familiarity of online multisided pure-play platforms selling physical offerings in an emerging market. *Southern African Business Review*, *23*(1). 5594-5625. https://doi.org/10.25159/1998-8125/5594

Chin, W. W. (1998). The partial least squares approach to structural equation modeling. *Modern Methods for Business Research*, *295*(2), 295–336. https://psycnet.apa.org/record/1998-07269-010

Chong, D., & Ali, H. (2022). Literature review: Competitive strategy, competitive advantages, and marketing performance on e-commerce Shopee Indonesia. *Dinasti International Journal of Digital Business Management*, *3*(2), 299–309. https://doi.org/10.38035/jkmt.v1i1

Dahbi, S., & Benmoussa, C. (2019). What Hinder SMEs from Adopting E-commerce? A Multiple Case Analysis. *Procedia Computer Science*, *158*, 811–818. https://doi.org/10.1016/j.procs.2019.09.118

ECDB. (2024). *eCommerce market in Indonesia*. https://Ecommercedb.Com/Markets/Id/All

Fikri, A., Nurmalina, R., Najib, M., & Simanjuntak, M. (2019). The Effect of reputation on online repurchase intention of fruits/vegetables in Indonesia with emotional and perceived risk as antecedent: based on the stimulus-organism-response model. *Jurnal Manajemen & Agribisnis*, *16*(2), 111–122.https://doi.org/10.17358/jma.16.2.111

Fornell, C. and Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50. https://doi.org/10.2307/3151312

Forsythe, S. M., & Shi, B. (2003). Consumer patronage and risk perceptions in Internet shopping. *Journal of Business Research*, *56*(11), 867–875. https://doi.org/10.1016/S0148-2963(01)00273-9

Goi, M. T., Kalidas, V., & Zeeshan, M. (2014). Comparison of stimulus-organism-response framework between international and local retailer. *Procedia-Social and Behavioral Sciences*, *130*, 461–468. https://doi.org/10.1016/j.sbspro.2014.04.054

Hair, J. F., Jr, Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2021). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (3rd ed.). SAGE Publications.

Hendricks, S., & Mwapwele, S. D. (2024). A systematic literature review on the factors influencing e-commerce adoption in developing countries. *Data and Information*

Management, *8*(1), 100045. https://doi.org/10.1016/j.dim.2023.100045

Huang, T.-Y., Chen, W.-K., Chen, C.-W., & Silalahi, A. D. K. (2022). Understanding How Product Reviews on YouTube Affect Consumers' Purchase Behaviors in Indonesia: An Exploration Using the Stimulus-Organism-Response Paradigm. *Human Behavior and Emerging Technologies*, *2022*(1), 4976980. https://doi.org/10.1155/2022/4976980

International Trade Administration. (2023). *eCommerce*. https://www.trade.gov/country-commercial-guides/indonesia-ecommerce

Jiménez-Jiménez, D., & Sanz-Valle, R. (2011). Innovation, organizational learning, and performance. *Journal of Business Research*, *64*(4), 408–417. https://doi.org/10.1016/j.jbusres.2010.09.010

Kim, H. (2019). Globalization and regulatory change: The interplay of laws and technologies in E-commerce in Southeast Asia. *Computer Law & Security Review*, *35*(5), https://doi.org/105315.10.1016/j.clsr.2019.03.009

Kurniawan, A., Wibowo, L., Rahayu, A., Yulianti, C., Annisa, T., & Riswanto, A. (2021). Online brand community strategy in achieving e-loyalty in the Indonesian e-commerce industry. *International Journal of Data and Network Science*, *5*(4), 785–790. https://doi.org/10.5267/j.ijdns.2021.7.002

Lăzăroiu, G., Neguriţă, O., Grecu, I., Grecu, G., & Mitran, P. C. (2020). Consumers' decision-making process on social commerce platforms: Online trust, perceived risk, and purchase intentions. *Frontiers in Psychology*, *11*, 890. https://doi.org/10.3389/fpsyg.2020.00890

Li, L., & Wang, W. (2020). The Effects of Online Trust-Building Mechanisms on Trust in the Sharing Economy: The Perspective of Providers. *Sustainability*, *12*(5), 1717. https://doi.org/10.3390/su12051717

Lubis, M., Safitra, M. F., Fakhrurroja, H., & Putri, D. P. (2024). Navigating Online Privacy: Insights from Cybersecurity Expert. *Procedia Computer Science*, *234*, 1388–1395. https://doi.org/10.1016/j.procs.2024.03.137

Lynn, N. D., Sourav, A. I., & Setyohadi, D. B. (2020). Increasing user satisfaction of mobile commerce using usability. *International Journal of Advanced Computer Science and Applications*, *11*(8), 300–308. https://doi.org/10.14569/IJACSA.2020.0110839

Ma'Mun, D. K., Aprillia, R., Riko, R., & Wijaya, L. (2023). Antecedents of the Online Shopping Behavior during the Covid-19 Pandemic. *Proceedings of the 2023 6th International Conference on Computers in Management and Business*, 86–92. https://doi.org/10.1145/3584816.3584844

Mehrabian, A., & Russell, J. A. (1974). The basic emotional impact of environments. *Perceptual and Motor Skills*, *38*(1), 283–301. https://doi.org/10.2466/pms.1974.38.1.283

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, *41*, 673–690. https://doi.org/10.1007/s11135-006-9018-6

Primadewi, S., & Fitriasari, W. (2022). Analisis Faktor Yang Mempengaruhi Pembelian Impulse Pada Live Streaming E-Commerce Berdasarkan SOR (Stimulus Organism

Response) Framework. *Jurnal Sosial Teknologi*, *2*(10), 846–856. https://doi.org/10.59188/jurnalsostech.v2i10.427

Priporas, C.-V., Stylos, N., & Fotiadis, A. K. (2017). Generation Z consumers' expectations of interactions in smart retailing: A future agenda. *Computers in Human Behavior*, *77*, 374–381. https://doi.org/10.1016/j.chb.2017.01.058

Rahayu, D. (2018). Indonesia national cybersecurity review: Before and after establishment national cyber and crypto agency (BSSN). *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, 1–6. https://doi.org/10.1109/CITSM.2018.8674265

Riquelme, I. P., & Román, S. (2014). The Influence of Consumers' Cognitive and Psychographic Traits on Perceived Deception: A Comparison Between Online and Offline Retailing Contexts. *Journal of Business Ethics*, *119*(3), 405–422. https://doi.org/10.1007/s10551-013-1628-z

Sarstedt, M., Hair, J. F., Cheah, J. H., Becker, J. M., & Ringle, C. M. (2019). How to specify, estimate, and validate higher-order constructs in PLS-SEM. *Australasian Marketing Journal*, *27*(3), 197–211. https://doi.org/10.1016/j.ausmj.2019.05.003

Sensuse, D. I., Sipahutar, R. J., Jamra, R. K., Suryono, R. R., & Kautsarina. (2020). Challenges and recommended solutions for change management in Indonesian e-commerce. *2020 International Conference on Information Technology Systems and Innovation, ICITSI 2020 - Proceedings*, 250–255. https://doi.org/10.1109/ICITSI50517.2020.9264950

Tran, V. D., & Nguyen, T. D. (2022). The impact of security, individuality, reputation, and consumer attitudes on purchase intention of online shopping: The evidence in Vietnam. *Cogent Psychology*, *9*(1). https://doi.org/10.1080/23311908.2022.2035530

Tsai, Y. C., & Yeh, J. C. (2010). Perceived risk of information security and privacy in online shopping: A study of environmentally sustainable products. *African Journal of Business Management*, *4*(18), 4057. https://doi.org/10.5897/AJBM.9000457

Vieira, V. A. (2013). Stimuli–organism-response framework: A meta-analytic review in the store environment. *Journal of Business Research*, *66*(9), 1420–1426. https://doi.org/10.1016/j.jbusres.2012.05.009

Sharing Business Data Securely

Insights from the European Gaia-X Project on Technical and Economic Roles Enabling Federated Data Spaces

Christoph Heinbach
FH Münster University of Applied Sciences, Münster School of Business (MSB), Münster, Germany

Jens Gessler
Zeppelin University, Friedrichshafen, Germany

Hanspeter Rychlik
Zeppelin University, Friedrichshafen, Germany

Christine Stecenko
HTW Saar, University of Applied Sciences, Saarbrücken, Germany

Horst Wieker
HTW Saar, University of Applied Sciences, Saarbrücken, Germany

Wolfgang H. Schulz
Zeppelin University, Friedrichshafen, Germany

Abstract: Federated data spaces (FDSs) represent an innovative approach to foster sovereign and interoperable data sharing across various business domains, promising substantial opportunities for value creation. The European Gaia-X initiative has emerged as a key driver in promoting FDS developments, particularly through its emphasis on data sovereignty and collaborative innovation. Effective design and operation of FDSs require a wide array of skills, expertise, services and technological components, creating a complex landscape for participating organisations. In this paper, we explore the technical and economic roles necessary for the successful implementation of FDSs, focusing on insights derived from two mobility use cases. Through interviews with experts engaged in a Gaia-X project, we identify 39 distinct roles, which we further abstract into eight meta-roles. These roles illustrate the structure and dynamics of inter-organisational collaboration in FDSs. Our analysis contributes to existing knowledge by illuminating the inter-organisational networks in FDSs, with a specific focus on the roles that support technical integration and economic value generation.

Keywords: Data sharing, federated data space, organisational role, Gaia-X, digital federation

https://doi.org/10.18080/jtde.v12n4.1037

Introduction

In our digitised economy, the use of data has become an asset for driving and facilitating service innovations and business models driven by the application of digital technologies (Yoo *et al.*, 2010). Data is consequently recognised as the enabler for digital transformation and disrupts the mechanisms of value creation in all business domains (Parker *et al.*, 2017). This situation has shaped the emergence of platform business and data ecosystems even further, allowing the sharing of data between independent actors and enabling data-driven innovations (Cennamo, 2021; de Reuver *et al.*, 2018). However, platform-based data ecosystems arise in the digital economy through the concept of data spaces entailing federated characteristics with sovereign data opportunities facilitating digital competition, specifically in an industrial context (Hutterer & Krumay, 2022; Möller *et al.*, 2024). Within an organisational context: "Data sharing is the domain-independent process of giving third parties access to the data sets of others [...] to develop new applications and services." (Jussen *et al.*, 2023, p. 4). This understanding has emerged in recent years and extends beyond a technical view of data exchange encompassing the broader transformational data process (Golshanara & Chomicki, 2020).

Today, data spaces are considered as enablers to foster federated data sharing among organisations (Otto & Jarke, 2019), while trust and data sovereignty remain key concerns in digital business (Hummel *et al.*, 2021; Senyo *et al.*, 2019). These issues result particularly from the dominance of hyperscalers focusing on data as a value-adding resource (Mosch *et al.*, 2023). Platform ecosystems, therefore, emerge within unregulated and non-transparent digital markets (Lee *et al.*, 2019). Establishing data spaces, characterised as function-oriented data-sharing systems (Hutterer *et al.*, 2023), consequently comes with a multitude of challenges from legal, technological, market and organisational perspectives (Curry *et al.*, 2022). To cope with this situation, the European data space initiative Gaia-X was launched in 2019. Gaia-X represents a European initiative and aims to create a scalable federated data infrastructure yielding digital sovereignty, trust and secure data sharing between ecosystem participants (BMWi, 2019). The Gaia-X approach supports the development of data ecosystems following the provisions of transparency, independence and interoperability (Braud *et al.*, 2021). In this context, "data spaces are a central means for the implementation of the European strategy on data." (Otto *et al.*, 2022, p. viii). Advanced data space concepts such as Gaia-X build on a decentralised data architecture (Eggers *et al.*, 2020) and use federated services yielding FDS to allow sovereign and interoperable data sharing between participants (Otto, 2022a).

https://doi.org/10.18080/jtde.v12n4.1037

Against this backdrop, developing federated data spaces (FDSs) facilitates data-driven innovations by promising data value opportunities (e.g., trust, sovereignty, interoperability), and maintaining the competition of data economies (Möller *et al.*, 2024). Currently, FDSs prevail to be developed focusing on technological implementations to provide normative software components for application in different FDS projects (Atzori *et al.*, 2024; Solmaz *et al.*, 2022). Nevertheless, the successful introduction and use of FDSs are associated with an economic benefit for the participants to implement digital sovereignty operationally. Without economic consideration, using FDSs with Gaia-X in practice is prone to remaining as an isolated technical development (Autolitano & Pawlowska, 2021). Addressing technical and economic perspectives is therefore considered an important approach to ensure that the development and use of FDSs in business domains can be achieved (Otto & Jarke, 2019). Following these thoughts, understanding the inter-organisational network contributing to data value functions for FDSs developed with Gaia-X provides a fruitful avenue for researchers. The corresponding interacting organisational roles represent relevant actors for the management of the network (Knight & Harland, 2005). Investigating the associated technical and economic roles with Gaia-X is therefore beneficial for discovering the interrelationships between organisations that make data a value-creating asset in FDSs. For this reason, the following research question (RQ) guides this paper:

RQ: *Which technical and economic roles can be identified that contribute to FDSs with Gaia-X to achieve sovereign and interoperable data sharing?*

To answer the RQ, the remainder of this paper is organised as follows: In the following section, we provide theoretical background information about organisational roles, data spaces and the investigated Gaia-X consortium research project. Thereafter, we present details of our research methodology applied to a qualitative study that follows grounded theory based on explorative expert interviews involved in the introduced Gaia-X project and uses an institutional concept to identify roles in FDSs. This is followed by a presentation of the results from our analysis and proposes a set of conceptualised technical roles, economic roles and the grouped meta-roles. Subsequently, we discuss our findings and draw limitations of our qualitative study. Finally, the paper concludes and provides an outlook for future research in the realm of data ecosystems.

Theoretical Background

Data space concept and Gaia-X

The data space concept has emerged exponentially and was coined by Franklin *et al.* (2005). Since then, the benefits of data spaces arise from "the possibility of exchanging heterogeneous

https://doi.org/10.18080/jtde.v12n4.1037

data within or outside of the organisation." (Hutterer & Krumay, 2024, p. 4506). Over time, data spaces have emergently been described as a federated concept comprising technical, legal and economic concerns of interacting organisations with the aim of inter-organisational data sharing (Beverungen *et al.*, 2022; Kagermann *et al.*, 2021). In this way, data spaces arise as the basis for platform-based alliances comprising federated functions (Otto & Jarke, 2019) established in data ecosystems (Hutterer & Krumay, 2022; Möller *et al.*, 2024). While the data management concept behind data spaces incorporates trust, security and compliance as fundamental attributes yielding federated capabilities, it also results in abstraction related to their practical usage in business domains (Solmaz *et al.*, 2022). Given the promising problem-solving opportunities in practice, using data spaces unfolds opportunities for organisations to capture data value. Data value is manifested in the sovereign provision and consumption of data or services in a federated and interoperable manner (Pettenpohl *et al.*, 2022). To achieve this, two data ecosystem initiatives were initialised in Europe to drive data space developments toward sovereign and interoperable data sharing (and exchange): Gaia-X and the International Data Spaces Association (Braud *et al.*, 2021, Möller *et al.*, 2024).

Gaia-X is attracting attention particularly across Europe through the provision of a decentralised data infrastructure that utilises data or services, enabling participants to innovate data-based business models (Gaia-X European Association for Data and Cloud AISBL, 2024). FDSs developed with Gaia-X therefore follow an advanced data space approach defined as: "A federated, open infrastructure for sovereign data sharing, based on common policies, rules and standards" (Reiberg *et al.*, 2022, p. 11). Although the technical concept of Gaia-X details relevant roles and responsibilities (e.g., Gaia-X European Association for Data and Cloud AISBL, 2022), an economic perspective toward operational usability in practice remains thus far challenging due to the nascent level of mature data spaces (Otto & Jarke, 2019, Möller *et al.*, 2024). To promote potential use cases for FDS projects with Gaia-X in business domains, Solmaz *et al.* (2022) suggest three key layers for consideration: (1) data connectors and infrastructure, (2) data interoperability, and (3) data value. Addressing these layers to operate use cases with data spaces such as Gaia-X requires organisational roles that realise value-based offerings between data providers and consumers (Möller *et al.*, 2024) or the provision of Gaia-X functions by software components such as an Identity Provider (Solmaz *et al.*, 2022).

Organisational roles

In any organisation the management of roles helps to align expectations and reduce conflicts in business activities (Tubre & Collins, 2000). Structuring and defining roles ensure that responsibilities and associated functions are addressed to facilitate efficient processes (Biddle,

https://doi.org/10.18080/jtde.v12n4.1037

1979; Mintzberg, 1973). Each role requires a delineation to avoid ambiguities that can lead to inefficiencies or failures in the organisational system (Lindman *et al.*, 2016). However, the dynamic nature of roles requires continuous adaptation to meet the evolving business needs, essentially affected by the digital transformation of inter-organisational networks. Given the various interrelationships between organisations existing within emerging FDSs, managing organisational roles is likewise critical to allow seamless interactions among different stakeholders yielding data-sharing value capabilities. In this context, relevant responsibilities addressed by organisational roles are found in, for instance, data governance, technical infrastructure or economic mechanisms (Otto & Jarke, 2019). The underlying interactions between the corresponding organisational roles in FDS are crucial for achieving integrity and effectiveness (Schleimer *et al.*, 2023), while at the same time fostering trust and interoperability among the organisations (Huber *et al.*, 2022). In this light, the successful operation of such digital alliances requires organisational roles in both technical and economic areas. For instance, roles responsible for system coordination and data integrity require an alignment with monetisation and compliance. Nevertheless, to accommodate the FDS concept in practice, organisational roles follow developments regarding the functional scope of data space activities. Integrating the relevant organisational roles into the framework is therefore important to meet the FDS objectives.

Overall, aligning organisational roles with the technological structure of FDSs facilitates the complexity of capturing promising (federated) value propositions (e.g., data sovereignty, interoperability and security) beyond scalability (Nagel & Lycklama, 2022; Otto, 2022a). Furthermore, data-sharing capabilities are leveraged using a decentralised data infrastructure approach applied by the interacting organisations and based on concisely defined responsibilities and tasks to be fulfilled. In essence, the exploration of organisational roles that contribute to the successful operation of FDSs allows for creating robust and flexible digital networks in the sphere of advanced data ecosystems within the digital economy.

Consortium research project Gaia-X 4 ROMS

Gaia-X can be recognised as a "technology-in-practice" (Orlikowski, 2000, p. 407) and supports a structured use case approach based on inter-organisational interactions within business domains (Maghazei *et al.*, 2022). This approach supports the transfer of the data space concept into an operational environment found in various research projects, such as the German consortium research project Gaia-X 4 ROMS. The project incorporates mobility use cases and is part of the Gaia-X 4 Future Mobility project family comprising six projects funded by the German Federal Ministry of Economics and Climate Action (DiMOS Operations GmbH, 2024). To be more precise, the Gaia-X 4 ROMS focuses on two use cases to achieve sovereign

https://doi.org/10.18080/jtde.v12n4.1037

and interoperable data sharing supported by 17 participating organisations. The first use case is about a public transportation system, while the second use case implements an autonomous freight transportation system to demonstrate the federated and decentralised capabilities of Gaia-X in practice (Heinbach *et al.* 2023). In both use cases, the key benefits of the FDS are derived from a federated catalogue to be implemented for the provision and consumption of data or services (Eggers *et al.*, 2020). Each use case involves several roles related to the fleet operation of vehicles (e.g., remote operation, routing) in which multiple organisations collaborate to operate an FDS based on federated functions (i.e., data sharing and data exchange). To this end, the project is organised in an interdisciplinary manner along various project stages to facilitate the implementation of the use cases with Gaia-X by the participants (e.g., research institutions, software providers, vehicle manufacturers). Associated technical and economic roles consequently concern the data space layers suggested by Solmaz *et al.* (2022) to achieve function-oriented FDSs (e.g., sovereign data sharing), rendering a role-based investigation of the FDS network as a promising approach.

Research Methodology

To identify the technical and economic roles in FDSs and answer our RQ, we follow the grounded theory methodology (Glaser & Strauss, 1967) and perform a qualitative study based on explorative expert interviews. We chose this methodology since the grounded theory is suggested to analyse the complex phenomenon of unknown social interactions (Glaser & Strauss, 1967; Strauss & Corbin, 1990). Generally, grounded theory is well established in qualitative platform research (e.g., Heinbach *et al.*, 2022; Papert & Pflaum, 2017), implying exploratory opportunities with data spaces (e.g., Hutterer & Krumay, 2024). For this reason, the objective of our exploration is to classify unknown objects (i.e., technical and economic roles) in FDS research based on data, providing us greater flexibility with which to add knowledge to an existing phenomenon.

Data collection

The data collection unit of analysis in our study involves the introduced consortium research project Gaia-X 4 ROMS (cf. Consortium research project Gaia-X 4 ROMS) due to the authors' involvement and access to it. Therefore, we opted to focus on data sharing and exchange functions represented by data and service offerings originating from the use cases. Due to the novelty of Gaia-X, we aimed to identify roles with sufficient information from a user's point of view to gain relevant data and a broad perspective for FDS operations (Huber *et al.*, 2022). Pursuing the qualitative approach supports the identification of the technical and economic roles in the Gaia-X data space and the different actors' perspectives. To collect qualitative data, we interviewed 16 experts involved in the Gaia-X project. The created database followed the

https://doi.org/10.18080/jtde.v12n4.1037

principles of grounded theory (Strauss & Corbin, 1990) by selecting experts based on their project activities (e.g., technical implementation, conceptual work), organisational character (e.g., company, institution), and use case involvement to achieve the intended development goals (e.g., user, developer). In Table 1, we provide an overview of the qualified experts including the organisation type, expert position, work experience and organisation size per EU definitions (European Commission, 2024). The interviews, with an average duration of 43 minutes, were conducted in the year 2022 between September and December via online video calls which were recorded, anonymised and transcribed.

Table 1. Overview of interviewed experts involved in the Gaia-X project

ID	Organisation type	Expert position	Experience	Organisation size
01	IT Services & Digitalisation	Specialist Mobility Services	6 yrs	Medium
02	IT Services & Digitalisation	Senior Engineer	11 yrs	Large
03	Digital Business Consultancy	Mobility & Digitalisation Manager	5 yrs	Large
04	Digital Business Consultancy	Technical Lead Architect	7 yrs	Large
05	Digital Business Consultancy	Digital Transformation Consultant	14 yrs	Small
06	Mobility System Provider	Digital Business Manager	18 yrs	Large
07	Commercial Vehicle OEM	Digital Service Manager	8 yrs	Large
08	Commercial Vehicle OEM	Product Engineer	5 yrs	Large
09	Commercial Vehicle OEM	Business Strategy & Innovation	10 yrs	Large
10	Commercial Vehicle OEM	Senior Account Manager	25 yrs	Large
11	Digital Automotive Solutions	Data Ecosystem Architect	7 yrs	Large
12	Digital Automotive Solutions	Head of Mobility Services	12 yrs	Large
13	Mobility Research Institute	Research Assistant	2 yrs	Medium
14	Mobility Research Institute	Researcher	5 yrs	Medium
15	AI Research Institute	Researcher	4 yrs	Large
16	AI Research Institute	Senior Researcher	14 yrs	Large

The interviews were guided by consideration of the Institutional Role Model (IRM) to identify the technical and economic roles in the FDS with Gaia-X (Schulz & Franck, 2022). The IRM is a conceptual approach to reduce complexity and fosters trust within inter-organisational cooperations that are relevant for FDSs. At the same time, the IRM enables reduction of the complexity of stakeholder relationships and clearly defines the roles necessary for operating FDSs. By focusing on a non-discriminatory, role-based collaboration, the IRM enables effective coordination of the institutions involved by assigning specific tasks and responsibilities (Schulz *et al.*, 2021). The primary goal of the IRM is to identify institutions, roles and market phases required to establish emerging technologies in digital markets (i.e., FDSs with Gaia-X). In addition, the successful application of the IRM for structuring complex relationships and promoting trust among participants in technology-based organisational networks has already been demonstrated within the mobility domain in practice (e.g., Schulz

https://doi.org/10.18080/jtde.v12n4.1037

et al., 2019). The IRM is likewise suited for the exploration of inter-organisational relationships and related roles based on use cases as demonstrated by our exploratory study.

We applied the IRM in our study in two steps. First, we categorised tasks derived from the project description into conceptual roles focusing on the Gaia-X implementation (e.g., governance, identity) and the use case implementation (e.g., fleet operations, order management). Second, we conducted semi-structured interviews with experts and presented the conceptual roles, while proposing questions comprising three sections: (1) relevance of the conceptual roles' contribution to the Gaia-X and use case implementations, (2) identification of technical or economic roles to support the implementation and usage of Gaia-X in the market, and (3) self-assignment of the organisation to the identified roles supporting the Gaia-X project objectives. The conversations with the experts led to an open and interactive discussion to comprehensively understand the organisational context. We also left room for the individual thoughts and ideas of the respondents based on their experiences within the scope of emerging FDSs.

Data analysis

To analyse the collected data, we applied open and axial coding techniques to ground the intended roles (Strauss & Corbin, 1990). At least one author and one scientific assistant analysed the data and identified key categories from the data (e.g., "software agents enable automated order booking", ID01) comprising key task responsibilities in the Gaia-X-enabled FDS (open coding). The identified categories were specified based on a composition of similar codes. For instance, the code "software agents enable automated order booking" (ID01) was associated with "fleet management is based on autonomous systems using artificial intelligence" (ID16) resulting in the category "Software Agent Provider." In line with this process, we grouped and summarised categories by examining their relationship based on their technical or economic contribution and data space link. For instance, the categories "Software Agent Provider" and "Product System Service Provider" were assigned as technical and grouped as "Data Tethering". This challenging step was performed intuitively by the authors focusing on exploring technical and economic roles grouped into conceptual meta-roles. The identified roles were verified by available Gaia-X documents and publications to ensure consistency and clarity of the findings.

Results

Based on the collected data and the conducted analysis, we explored the data space of the Gaia-X 4 ROMS project and revealed 39 roles in a platform-based network shaping a digital "cloud-federation" (Autolitano & Pawlowska, 2021, p. 12). We further derived 20 technical roles and

https://doi.org/10.18080/jtde.v12n4.1037

19 economic roles grouped into eight conceptualised meta-roles. Our results based on the Gaia-X project are elaborated in more detail as follows.

Technical roles contributing to FDSs with Gaia-X

The technical roles explored address the technical backbone of the data space to realise FDS capabilities following the trust framework provided by the technical committee of Gaia-X (Gaia-X European Association for Data and Cloud AISBL, 2022). In this way, the role **System Federator** is responsible in the decentralised inter-organisational network to for setting (data) governance. The role initiates coordination and decision-making processes within the digital federation and ensures that federation rules are achieved by the provision of federation services (e.g., identity and trust). The role **System Participant** likewise describes natural persons, legal entities or technical machines aiming to participate in the FDS based on verifiable credentials issued by the federator. To achieve this, a **Trusted Third Party** supports the creation of verifiable credentials in a machine-readable format for the validation of services based on self-descriptions. Applying the federation services by participants consequently requires an enabling role represented by **Portal and Integration**. The role offers services such as a graphical user interface for interacting functions between organisations. To make data and services consumable in a trustworthy and verifiable manner, the role **Service Catalogue** sustains the search and selection of offerings between providers and consumers by providing a Gaia-X repository. This role complies with the technical concept of the "federated catalogue" in the Gaia-X initiative based on self-descriptions from participants (Gaia-X European Association for Data and Cloud AISBL, 2021, p. 6). In this context, **Data Sovereignty** is identified as a role, enabling transparency and control based on specific capabilities such as the data contract transaction (e.g., the scope of data usage) and the data-sharing logging (e.g., creation of an auditable transaction log). The role **Identity and Trust** ensures the provision of Gaia-X-compliant identity solutions. It introduces compliance within the FDS, for instance, by providing federated data services to use verifiable identities. To give access to the data and services according to the roles and rights concept defined in the digital federation, the role **Conformity and Onboarding** was identified. Relevant activities include the realisation of a validation process for participants, data services and data offerings to ensure that the federation access promotes secure and trustworthy interactions between providers and consumers. A **Data Space Connector Provider** offers software components to enable peer-to-peer data sharing or exchange between networking organisations (e.g., Möller et al., 2024). Data space connectors yield a data connection based on a plane architecture (i.e., control plane and data plane) to facilitate interoperable data sharing between FDSs. In particular, the Eclipse Data Space Connector was specified in our survey for the implementation of the Gaia-X project (Eclipse Foundation AISBL, 2024).

https://doi.org/10.18080/jtde.v12n4.1037

The roles of **Data Service Prosumer** and **Data Prosumer** provide and/or use/consume data or services as key participants within the Gaia-X data space. They are reflected by freight fleet and public transport operators providing and consuming data based on data or service offerings. The role **Software Agent Provider** specifically enables the optimisation and automation of freight deliveries in intermodal transport chains using multi-agent system software (Maecker *et al.*, 2023). The software incorporates a system of various agents, each representing digital entities of physical actors leveraging automated transport functions. The role **Gaia-X Data Space Integrator** connects existing Gaia-X data spaces on a federated level to realise data interoperability. Likewise, aligning the existing governance rules to ensure sovereign data sharing is included. In contrast, a **Non-Gaia-X Data Space Integrator** was identified that assures the data transfer and manages interfaces between existing data spaces, cloud services and platforms hosted outside the Gaia-X federation. While the connections are supported, the range of data and service offerings in the catalogue can be expanded in addition to the standardised integration processes (e.g., onboarding). We identify the role **Product Service System Provider** as supporting data usage from applied IoT technologies (i.e., telematics systems, on board unit) from commercial freight vehicles based on the data backend of a fleet management system. In addition, a **Hosting Infrastructure Provider** is responsible for decentralised Gaia-X cloud infrastructure and hosted as-a-service components, for instance, to run Gaia-X federation services or individual service offerings in an indexed repository (i.e., federated catalogue). Moreover, the role **Communication Infrastructure Provider** supports communication between Edge and Cloud services and the provision of communication technologies or transmission between network operators. The role **Depot Provider** was found within the scope of use case implementation for the handling of goods in the freight transportation systems. Depots represent specific warehouses where operators perform the loading and unloading of goods. Furthermore, a **Maintenance Provider** role was explored providing physical workshop resources for vehicle operations (e.g., repair, maintenance). Finally, the role **Traffic Control Centre Provider** is responsible for the traffic control system operating in public transportation (e.g., traffic light control, traffic guidance).

Economic roles contributing to FDSs with Gaia-X

Our Gaia-X study has uncovered several economic roles for capturing data value opportunities to drive market competition. We identified the role **Governance** as a federating actor to set up data-sharing rules (e.g., terms and conditions) and specify design processes (e.g., data/service offering) for the FDS participants. This role is crucial for the creation of applied data policies in the data-sharing process. To identify and reduce transaction costs using the federation services, the role **Transaction Management** supports ecosystem participants

https://doi.org/10.18080/jtde.v12n4.1037

and enables the initiation and execution of transactions within the data space (e.g., streamlining operations and interactions between stakeholders). Thereby, the **Compliance** role is required to apply legal obligations and assert data standards of the business domain (e.g., General Data Protection Regulation). The **Marketing and Sales** role identified during the analysis is responsible for activities to foster positive market awareness and advertising activities of the Gaia-X data space. The role helps to attract potential new participants or partners to expand offerings or functional scope within the FDS. Likewise, a **Finance** role is relevant to monitor the payment transactions by the data value functions within the Gaia-X project. Finance activities refer to subsequent data space initiatives (e.g., financial funding, sponsorship). Moreover, the role **Communication** coordinates federation tasks between participants and external partners through stakeholder events (e.g., meetings, workshops) to drive the federation activities forward. Likewise, this includes the communication of interests with relevant stakeholders outside the FDS (e.g., authorities). The role **Legal** supports the compliance aspects within the Gaia-X project and the provision of relevant services (e.g., contracts). Legal advice (e.g., data protection) is therefore provided to participants. A legal framework structure within the FDS is important for the participants to increase trust and protect organisations against risks. Through the role **Portfolio Management**, the management and control of the Gaia-X federation services for application is enabled. Based on the provided portfolio and data value opportunities from the data space, participants receive strategic support to manage their data or service offerings. The roles **Data Owner** and **Service Owner** were identified as legal entities providing data sets and services to consumers. Owners encompass the authorisation management of the data sets (e.g., access to anonymised data) and the maintenance of the services provided (e.g., software updates). The role **Consulting** was explored for analysation, for instance, existing business resources of participants or potential business models. To make the FDS a profitable effort for participants, the role **Monetisation Management** focuses on developing suitable monetisation models for the services, ensuring economic and competitive returns respond to customer needs. To address the customer markets, the role **Service Channel Management** is responsible for exploring new customer segments and providing suitable market channels for the offerings of the Gaia-X data space. Managing the customer channels enables organisations to increase revenue by offering data (e.g., GPS vehicle data is provided to municipalities for advanced traffic planning) or services for new customers and use cases. In addition, we identified the role **Freight Fleet Provider** which reflects vehicle operations delivering transportation services based on a transport order. These providers ensure the availability and reliability of the physical fleet assets in use (e.g., buses, freight trailers, delivery robots) associated with the public and freight fleet operations. In particular, the role **Remote Transport Service Provider** provides services to the markets addressing remote or teleoperated fleets for local

https://doi.org/10.18080/jtde.v12n4.1037

passenger transport. As both aforementioned use cases are based on the processing of transport orders, the role **Transport Management** is responsible for ensuring that the order process is implemented across the fleet vehicles in operation. Likewise, to address the technical operation tasks related to the fleets, the role **Fleet Management** entails the technical provision of vehicles to assure qualitative transport services and fulfillment of vehicle safety requirements in the mobility markets. The maintenance tasks for the vehicles are addressed by the role **Maintenance Management** and coordinate the necessary activities for regular maintenance (e.g., inspections) and repairs (e.g., damages). Finally, for freight transportation activities, the role **Depot Management** is related to the handling of freight for the transshipment of parcels between different modes of transport or loading units.

Meta-roles contributing to FDSs with Gaia-X

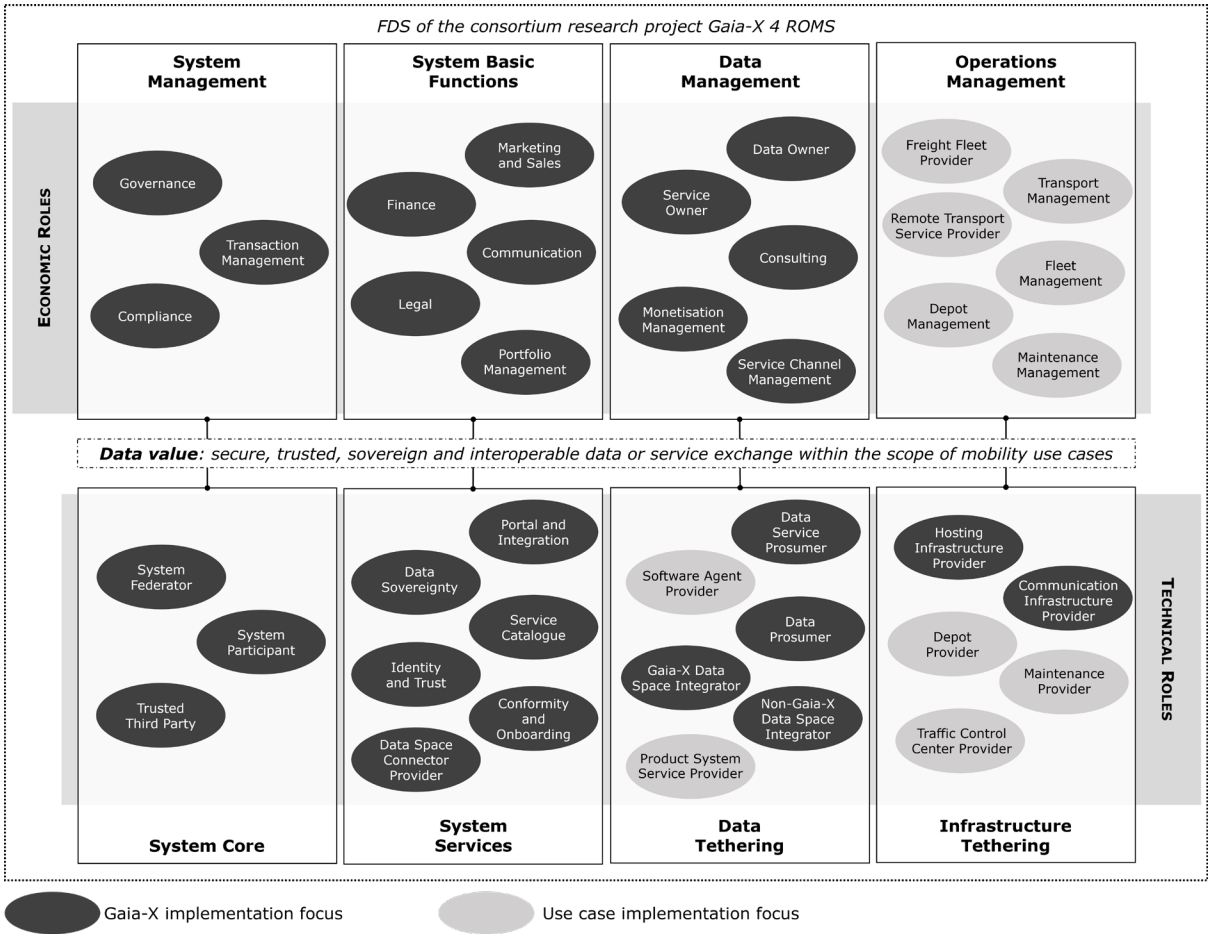


Figure 1. Proposed set of technical roles, economic roles and meta-roles in the investigated FDS

Based on the findings, we grouped the roles into eight conceptual meta-roles. The identified technical and economic roles and assigned meta-roles are illustrated in Figure 1. The proposed set incorporates the role-based functionalities aligned to the tasks and network interrelationships in the investigated Gaia-X project. In the figure, data value is captured by secure, trusted, sovereign and interoperable data sharing between the organisations in the use

https://doi.org/10.18080/jtde.v12n4.1037

cases. A differentiation between roles and superordinate meta-roles focusing on use case or Gaia-X implementation is made according to the insights obtained from our survey.

Discussion and Limitations

The presented study is one of the first, to the best of the authors' knowledge, to explore organisational roles in FDSs with Gaia-X to achieve sovereign and interoperable data sharing and exchange. The investigated network is grounded on a decentralised cloud-based infrastructure to deploy federation services, forming an inter-organisational alliance characterised as a "cloud-federation" (Autolitano & Pawlowska, 2021, p. 12). In this vein, data spaces provide the opportunity to orchestrate digital capabilities between organisations for achieving common objectives (i.e., secure, trusted, sovereign, and interoperable data sharing and exchange) based on a boundary that we call a digital federation. From the 16 qualitative interviews conducted within the consortium research project Gaia-X 4 ROMS, we identified 20 technical and 19 economic roles required to establish a digital federation incorporating a dedicated data space for two mobility use cases. The theoretical contribution of our work provides a set of 39 roles found in the examined Gaia-X project, which were summarised into eight conceptualised meta-roles (Figure 1).

In answering our RQ through presenting the identified roles (cf. Results), we have revealed that FDSs embrace more than technological layers (e.g., infrastructure, data interoperability). The investigated FDSs provided data value opportunities (Solmaz *et al.*, 2022) by sharing data in a trusted manner manifested by two use cases in the mobility domain. Our survey revealed that capturing data value is conducive to using FDS in practice, centring on four economic meta-roles (i.e., system management, system basic functions, data management and operations management). These roles foster an economic business understanding of applying FDSs and set an entry point for use case owners beyond the existing socio-technical system design focus (Otto & Jarke, 2019). We further supported organisational understanding of FDS drivers by assigning trust, interoperability and cost control to specific roles in the inter-organisational network (Hutterer & Krumay, 2024). In particular, the role of Monetisation Management has been identified as the prime economic concern for organisations to recognise FDS as a competitive data technology in digital markets. However, it eventuated during the conducted study that the maturity of technical components and the flexible integration of federation services into the use case is a prerequisite for the application of the FDS concept in an operational context. Combining technical and economic perspectives in emerging FDSs is therefore subject to the operational requirements to be considered within sector-specific business domains. This situation renders current FDS projects an individual venture and corresponds with the current challenges of interoperability between data spaces discussed in

https://doi.org/10.18080/jtde.v12n4.1037

academia (e.g., Möller *et al.*, 2024). Our work, however, contributes to the theoretical concept of data spaces by providing insights into an organisation's set of actors for a specific data space, which calls for more analysis about the FDS configurations in general across business domains.

In addition to contributing to a theoretical body of knowledge, our study also offers managerial implications for decision-makers. First, the conceptualised roles may serve as a blueprint that guides other FDS research projects in terms of the institutional structure for implementation activities and responsibilities, fostering a transition from data infrastructure to data value. Second, our findings suggest adapting the set of roles when applying them to use cases in other business domains. Third, we see the opportunity for practitioners to use our results as a boundary role concept, allowing interacting organisations to define the tasks and obligations in data space initiatives to achieve data-sharing objectives collaboratively. The identified roles help to span an initial bridge between organisations for operationalising data sharing following a decentralised data infrastructure approach and provide an entry point for building trusted data spaces based on use cases in practice.

Notwithstanding these insights, the explorative character of this study is subject to some limitations. First, the conceptualisation of technical and economic roles is grounded in data collected from 16 interviewed experts in the mobility domain. Although we discovered use cases with diverse service settings, our results can neither be considered as comprehensive nor representative of general FDS projects. Second, the interviewed experts are related to two use cases implemented with limited Gaia-X functionality, which may have introduced a bias in our results. Investigating market-ready FDSs (e.g., DRM Datenraum Mobilität GmbH, 2024) could have revealed other insights. Third, the proposed organisational roles are subject to our subjective interpretation of the qualitative data collected from the interviews. While the authors are experienced with the topic of Gaia-X, mobility and organisational roles, the findings underlie inherent limitations that should be addressed by conducting further quantitative research to arrive at a generalised set of organisational roles applicable to dedicated business domains in the sphere of data spaces.

Conclusion

In this paper, we identified the organisational roles of digital federations which are emerging exponentially in the context of data spaces currently being built. Our exploratory survey is motivated by the promising data value opportunities of Gaia-X, positioned as a technical infrastructure concept in Europe, and the associated sovereign data-sharing capabilities for organisations. To achieve our research objective and answer the RQ, we conceptualised 39 roles summarised into eight overarching meta-roles contributing to FDS with Gaia-X. The

https://doi.org/10.18080/jtde.v12n4.1037

roles were derived according to the tasks, organisational interrelationships and activities based on interviews conducted with 16 Gaia-X experts in the mobility domain. Given the current data ecosystem initiatives and FDS activities, a more comprehensive and domain-specific understanding of their use to drive digital innovations is required. Future research opportunities arise in validating our findings in other FDS projects and domains. In this vein, more scholarly efforts are necessary to explore for organisations the granularity of data value functions (e.g., data sharing and data exchange) and their economic mechanisms (e.g., payment and pricing) within FDSs that are currently being built. It is hoped that the insights from this study span the gap between theory and practice on a contemporary issue and, in doing so, promote the expansion of sovereign digital forces in digital economies on a global scale.

Acknowledgements

This article was written as part of the Gaia-X 4 ROMS project – Support and Remote Operation of Automated and Networked Mobility Services (Action Number: 19S21005P). The project is funded by the German Federal Ministry of Economics and Climate Action (BMWK). The authors are responsible for the content of this article.

References

Atzori, M., Ciaramella, A., Diamantini, C., Martino, B., Distefano, S., Facchinetti, T., Montecchiani, F., Nocera, A., Ruffo, G., Trasarti, R. (2024). Dataspaces: Concepts, architectures and initiatives. *CEUR workshop proceedings*, *3606*. CEUR-WS.org. https://hdl.handle.net/11584/389724

Autolitano, S., & Pawlowska, A. (2021). *Europe's quest for digital sovereignty: GAIA-X as a case study*. Istituto Affari Internazionali (IAI). [PDF] https://www.iai.it/sites/default/files/iaip2114.pdf

Beverungen, D., Hess, T., Köster, A., & Lehrer, C. (2022). From private digital platforms to public data spaces: Implications for the digital transformation. *Electronic Markets*, *32*(2), 493–501. https://doi.org/10.1007/s12525-022-00553-z

Biddle, B. J. (1979). *Role theory: Expectations, identities, and behaviors*. Academic Press. https://doi.org/10.1177/000169938202500214

BMWi (2019). Project GAIA-X – *A Federated Data Infrastructure as the Cradle of a Vibrant European Ecosystem. Federal Ministry for Economic Affairs and Energy (BMWi).* https://www.bmwk.de/Redaktion/EN/Publikationen/Digitale-Welt/project-gaia-x.pdf?__blob=publicationFile&v=1 (Retrieved on 13 November 2024)

Braud, A., Fromentoux, G., Radier, B., & Le Grand, O. (2021). The road to European digital sovereignty with Gaia-X and IDSA. *IEEE Network*, *35*(2), 4–5. https://doi.org/10.1109/MNET.2021.9387709

Catena-X Automotive Network e.V. (2024). R*adical change calls for even more radical collaboration*. https://catena-x.net/en/ (Retrieved on 11 November 2024)

Cennamo, C. (2021). Competing in digital markets: A platform-based perspective. *Academy of Management Perspectives*, *35*(2), 265–291. https://doi.org/10.5465/amp.2016.0048

Curry, E., Tuikka, T., Metzger, A., Zillner, S., Bertels, N., Ducuing, C., Dalle Carbonare, D., Gusmeroli, S., Scerri, S., López de Vallejo, I., & García Robles, A. (2022). Data sharing spaces: The BDVA perspective. In B. Otto, M. Ten Hompel, & S. Wrobel (Eds.), *Designing Data Spaces* (pp. 365–382). Springer International Publishing. https://doi.org/10.1007/978-3-030-93975-5_22

de Reuver, M., Sørensen, C., & Basole, R. C. (2018). The digital platform: A research agenda. *Journal of Information Technology*, *33*(2), 124–135. https://doi.org/10.1057/s41265-016-0033-3

DiMOS Operations GmbH (2024). *Gaia-X 4 ROMS: Support and Remote-Operation of Automated and Networked Mobility Services*. https://www.gaia-x4futuremobility.de/en/projects/roms (Retrieved on 13 November 2024)

DRM Datenraum Mobilität GmbH (2024). *Mobility Data Space: The data space for future mobility*. https://mobility-dataspace.eu/ (Retrieved on 13 November 2024)

Eclipse Foundation AISBL (2024). *Dataspaces and the Eclipse Dataspace Components (EDC)*. https://projects.eclipse.org/projects/technology.edc (Retrieved on 13 November 2024)

Eggers, G., Fondermann, B., Maier, B., Ottradovetz, K., Pfrommer, J., Reinhardt, R., Rollin, H., Schmieg, A., Steinbuß, S., Trinius, P., Weiss, A., Weiss, C., & Wilfling, S. (2020). *GAIA-X: Technical Architecture*. Federal Ministry for Economic Affairs and Energy (BMWi). https://www.bmwk.de/Redaktion/EN/Publikationen/gaia-x-technical-architecture.pdf?__blob=publicationFile&v=1 (Retrieved on 13 November 2024)

European Commission (2024). *SME Definition*. https://single-market-economy.ec.europa.eu/smes/sme-fundamentals/sme-definition_en (Retrieved on 13 November 2024)

Franklin, M., Halevy, A., & Maier, D. (2005). From databases to dataspaces: A new abstraction for information management. *ACM SIGMOD Record*, *34*(4), 27–33. https://doi.org/10.1145/1107499.1107502

Gaia-X European Association for Data and Cloud AISBL (2021). *Gaia-X Federation Services (GXFS)*. https://gaia-x.eu/wp-content/uploads/files/2022-01/Gaia-X_Federation_Services_White_Paper_1_December_2021.pdf (Retrieved on 11 November 2024)

Gaia-X European Association for Data and Cloud AISBL (2022). *Gaia-X – Architecture Document – 22.04 Release*. https://gaia-x.eu/wp-content/uploads/2022/06/Gaia-x-Architecture-Document-22.04-Release.pdf (Retrieved on 11 November 2024)

Gaia-X European Association for Data and Cloud AISBL (2024). *Gaia-X: European Association for Data and Cloud AISBL*. https://gaia-x.eu/wp-content/uploads/2024/01/Gaia-X-Brochure_2024_Online_Spread.pdf (Retrieved on 11 November 2024)

Glaser, B., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine Transaction.

Golshanara, L., & Chomicki, J. (2020). Temporal data exchange. *Information Systems, 87*, 101414. https://doi.org/10.1016/j.is.2019.07.004

Heinbach, C., Beinke, J., Kammler, F., & Thomas, O. (2022). Data-driven forwarding: A typology of digital platforms for road freight transport management. *Electronic Markets*, *32*(2), 807–828. https://doi.org/10.1007/s12525-022-00540-4

Heinbach, C., Gösling, H., Meier, P. & Thomas, O. (2023). Smart managed freight fleet: Ein automatisiertes und vernetztes Flottenmanagement in einem föderierten Datenökosystem. *HMD Praxis der Wirtschaftsinformatik*, *60*(1), 193–213. https://doi.org/10.1365/s40702-022-00887-4

Huber, M., Wessel, S., Brost, G., & Menz, N. (2022). Building trust in data spaces. In B. Otto, M. Ten Hompel, & S. Wrobel (Eds.), *Designing Data Spaces* (pp. 147–164). Springer International Publishing. https://doi.org/10.1007/978-3-030-93975-5_9

Hummel, P., Braun, M., Tretter, M., & Dabrock, P. (2021). Data sovereignty: A review. *Big Data & Society*, *8*(1). https://doi.org/10.1177/2053951720982012

Hutterer, A., & Krumay, B. (2022). Integrating Heterogeneous data in dataspaces – a systematic mapping study. *Proceedings of Pacific Asia Conference on Information Systems (PACIS) 2022*. 222. https://aisel.aisnet.org/pacis2022/222

Hutterer, A., & Krumay, B. (2024). The adoption of data spaces: Drivers toward federated data sharing. *Proceedings of the 57th Hawaii International Conference on System Sciences (HICSS) 2024* (pp. 4506–4515). https://hdl.handle.net/10125/106926

Hutterer, A., Krumay, B. & Mühlburger, M. (2023). What constitutes a dataspace? Conceptual clarity beyond technical aspects. In *Proceedings of 29th Americas Conference on Information Systems (AMCIS) 2023*. 5. https://aisel.aisnet.org/amcis2023/eco_systems/eco_systems/5

Jussen, I., Schweihoff, J., Dahms, V., Möller, F., & Otto, B. (2023). Data sharing fundamentals: Definition and characteristics. *Proceedings of the 56th Hawaii International Conference on System Sciences (HICSS) 2023 (pp. 3685–3694)*. https://hdl.handle.net/10125/103083

Kagermann, H., Streibich, K.-H., & Suder, K. (2021). *Digital sovereignty. Status quo and perspectives*. acatech IMPULSE– National Academy of Science and Engineering. https://en.acatech.de/publication/digital-sovereignty/ (Retrieved on 13 November 2024)

Knight, L., & Harland, C. (2005). Managing supply networks: Organizational roles in network management. *European Management Journal*, *23*(3), 281–292. https://doi.org/10.1016/j.emj.2005.04.006

Lee, S. U., Zhu, L., & Jeffery, R. (2019). Data governance decisions for platform cosystems. *Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS) 2019* (pp. 6377–6386). https://hdl.handle.net/10125/60072

Lindman, J., Kinnari, T., & Rossi, M. (2016). Business roles in the emerging open-data ecosystem. *IEEE Software*, *33*(5), 54–59. https://doi.org/10.1109/MS.2015.25

Maecker, D., Gösling, H., Heinbach, C., & Kammler, F. (2023). Exploring multi-agent systems for intermodal freight fleets: Literature-based justification of a new concept. *Proceedings of the 18 Internationale Tagung Wirtschaftsinformatik (WI 2023)* (pp. 1–17). https://aisel.aisnet.org/wi2023/97

Maghazei, O., Lewis, M. A., & Netland, T. H. (2022). Emerging technologies and the use case: A multi-year study of drone adoption. *Journal of Operations Management*, *68*(6–7), 560–591. https://doi.org/10.1002/joom.1196

Mintzberg, H. (1973). *The nature of managerial work*. Harper & Row.

Mosch, P., Majocco, P., & Obermaier, R. (2023). Contrasting value creation strategies of industrial-IoT-platforms – a multiple case study. *International Journal of Production Economics*, *263*, 1–14. https://doi.org/10.1016/j.ijpe.2023.108937

Möller, F., Jussen, I., Springer, V., Gieß, A., Schweihoff, J. C., Gelhaar, J., Guggenberger, T., & Otto, B. (2024). Industrial data ecosystems and data spaces. *Electronic Markets*, *34*(1), 1–17. https://doi.org/10.1007/s12525-024-00724-0

Nagel, L., & Lycklama, D. (2022). How to build, run, and govern data spaces. In B. Otto, M. Ten Hompel, & S. Wrobel (Eds.), *Designing Data Spaces* (p. 17–28). Springer International Publishing. https://doi.org/10.1007/978-3-030-93975-5_2

Orlikowski, W. J. (2000). Using technology and constituting structures: A practice lens for studying technology in organizations. *Organization Science*, *11*(4), 404–428. https://doi.org/10.1287/orsc.11.4.404.14600

Otto, B. (2022a). A federated infrastructure for European data spaces. *Communications of the ACM*, *65*(4), 44–45. https://doi.org/10.1145/3512341

Otto, B. (2022b). The evolution of data spaces. In B. Otto, M. Ten Hompel, & S. Wrobel (Eds.), *Designing Data Spaces*. (p. 3–15). Springer International Publishing. https://doi.org/10.1007/978-3-030-93975-5_1

Otto, B., & Jarke, M. (2019). Designing a multi-sided data platform: Findings from the international data spaces case. *Electronic Markets*, *29*(4), 561–580. https://doi.org/10.1007/s12525-019-00362-x

Otto, B., Ten Hompel, M., & Wrobel, S. (2022). Preface. In B. Otto, M. Ten Hompel, & S. Wrobel (Eds.), *Designing Data Spaces*. (pp. vi–viii). Springer International Publishing. https://doi.org/10.1007/978-3-030-93975-5

Papert, M., & Pflaum, A. (2017). Development of an ecosystem model for the realization of Internet of Things (IoT) services in supply chain management. *Electronic Markets*, *27*(2), 175–189. https://doi.org/10.1007/s12525-017-0251-8

Parker, G., Van Alstyne, M., & Jiang, X. (2017). Platform cosystems: How developers invert the firm. *MIS Quarterly*, *41*(1), 255–266. https://doi.org/10.25300/MISQ/2017/41.1.13

Pettenpohl, H., Spiekermann, M., & Both, J. R. (2022). International data spaces in a nutshell. In B. Otto, M. Ten Hompel, & S. Wrobel (Eds.), *Designing Data Spaces* (pp. 29–40). Springer International Publishing. https://doi.org/10.1007/978-3-030-93975-5_3

Reiberg, A., Niebel, C., & Kraemer, P. (2022). *What Is a Data Space?* [White paper]. Gaia-X Hub Germany. https://gaia-x-hub.de/wp-content/uploads/2023/11/GX-White-Paper-Data-Space.pdf (Retrieved on 15 November 2024)

Schleimer, A. M., Jahnke, N., & Otto, B. (2023). Architecture design options for federated data spaces. *Proceedings of the 56th Hawaii International Conference on System Sciences (HICSS) 2023 (pp. 3643–3652)*. https://hdl.handle.net/10125/103078

Schulz, W. H. & Franck, O. (2022). The institutional role model: A system-dynamic approach to reduce complexity. *International Journal of Sustainable Development and Planning*, *17*(2), 351–361. https://doi.org/10.18280/ijsdp.170201

Schulz, W. H., Franck, O., & Smolka, S. (2021). Die Theorie der institutionellen Rollenmodelle – der Restrukturierungsansatz für Unternehmen zur Bewältigung der COVID-19 Krise. In W. H. Schulz, N. Joisten, & C. F. Edye (Eds.), *Mobilität nach COVID-19* (pp. 1–32). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-33308-9_1

Schulz, W. H., Joisten, N., & Arnegger, B. (2019). Development of the institutional role model as a contribution to the implementation of co-operative transport systems. *SSRN Electronic Journal*, 1–31. http://dx.doi.org/10.2139/ssrn.3421107

Senyo, P. K., Liu, K., & Effah, J. (2019). Digital business ecosystem: Literature review and a framework for future research. *International Journal of Information Management*, *47*, 52–64. https://doi.org/10.1016/j.ijinfomgt.2019.01.002

Solmaz, G., Cirillo, F., Fürst, J., Jacobs, T., Bauer, M., Kovacs, E., Santana, J. R., & Sánchez, L. (2022). Enabling data spaces: Existing developments and challenges. *Proceedings of the 1st International Workshop on Data Economy 2022* (pp. 42–48). https://doi.org/10.1145/3565011.3569058

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Sage publications.

Tubre, T. C., & Collins, J. M. (2000). Jackson and Schuler (1985) revisited: A meta-analysis of the relationships between role ambiguity, role conflict, and job performance. *Journal of Management*, *26*(1), 155–169. https://doi.org/10.1177/014920630002600104

Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). Research commentary – The new organizing logic of digital innovation: An agenda for information systems research. *Information Systems Research*, *21*(4), 724–735. https://doi.org/10.1287/isre.1100.0322

Autonomous Robot Navigation System Workflow for Monitoring and Maintenance in Industry 4.0 Applications

Simon Peter Cornelius
Tunku Abdul Rahman University of Management and Technology, Malaysia

Jia Jan Ong
Tunku Abdul Rahman University of Management and Technology, Malaysia

Tsung Heng Chiew
Tunku Abdul Rahman University of Management and Technology, Malaysia

Kai Ming Chang
Tunku Abdul Rahman University of Management and Technology, Malaysia

Yoon Ket Lee
Tunku Abdul Rahman University of Management and Technology, Malaysia

Abstract: Autonomous navigation in factories faces a different challenge with a lack of GPS, frequently changing environment, and human interference. Current methods employed include autonomous guided vehicles, which require extensive setup and lack flexibility, making it unsuitable for frequently changing environments. Prohibiting the adoption of an autonomous mobile robot is the slow mapping time and cost. A method of autonomous navigation combining computer vision with path-planning algorithms is presented. This method uses cameras attached to the environment for navigation and not on the robot to leverage security cameras commonly available. Out of the four aspects of navigation, only three were successful, namely perception, localisation, and cognition. Mapping via grayscale thresholding is found faster than simultaneous localisation and mapping but is less accurate because it is dependent on lighting. Faster region-based convolutional neural network or You Only Look Once version 4 (YOLOv4) found no difference because the travel time is significantly longer than the processing time of both. The proposed method using A* path planning with Euclidean heuristic successfully reaches the goal for at least 30 repeats. Navigational abilities are still limited in real-world settings because of accumulation of error from odometry and inertial measurement unit sensors and lack of localisation feedback.

Keywords: Computer vision, autonomous navigation, autonomous mobile robots, automated guided vehicles, robot operating system.

https://doi.org/10.18080/jtde.v12n4.1017

Introduction

An important aspect in using robots in an Industry 4.0 setting is the need for these robots to navigate without human intervention in a dynamic environment. In a factory environment, we cannot rely on the global positioning system (GPS) for localisation because it may not be available; in addition, the environment is frequently changing, with objects constantly moving around the factory floor. This paper seeks to develop a method of autonomous navigation that also adapts to changes in the environmental conditions. Computer vision and navigation algorithms can be combined to produce a more efficient method to meet the needs of the robot.

Autonomous guided vehicles (AGV) and autonomous mobile robots (AMR) are commonly used for material transportation. Typically, AGV tends to refer to autonomous vehicles that require external path guidance like magnetic tape (Zaman *et al.*, 2021) or quick-response (QR) stickers (Zhang *et al.*, 2023) but is unable to plot a free path outside its predefined path (Ahmadi *et al.*, 2023). Alternatively, AMR addresses the problems faced by AGV using the simultaneous localisation and mapping (SLAM) algorithm. SLAM allows robots to create maps and localise its map position in an unknown environment using a myriad of sensors (Choi *et al.*, 2023; Zhao & Chidambareswaran, 2023). Once the map of the environment is made, a path-planning and obstacle-avoidance algorithm is used for navigation and achieving its goals.

The shortcomings of AGVs are that they are inflexible while requiring extensive setup, and AMRs are expensive given the sensors used (Merzlyakov & Macenski, 2021; Neto & Tonidandel, 2022) and it is difficult getting reliably accurate localisation in dynamic environments (Leong & Ahmad, 2024). We intend to address these issues by using cameras attached to the environment for navigation unlike AMRs, which commonly only rely on its onboard sensors for navigation. Hence, achieving a successful mobile robot navigation requires four main aspects (Alatise & Hancke, 2020; Roland *et al.*, 2011):

- Perception: the robot must extract useful data from its sensors.
- Localisation: the robot must know its location in an environment.
- Cognition or path planning: decide which action best taken to complete its task.
- Motion control or navigation: control motor outputs to achieve the trajectory with zero error.

Out of the four aspects, our proposed method will swap out the perception of mapping the environment and localisation of the robot's pose. Reintegration to the full workflow of navigation is done to determine the viability of the proposed method for autonomous navigation. The proposed method aims to improve navigational efficiency by using commonly available equipment. Currently, AMRs will have to encounter obstacles before knowing the

https://doi.org/10.18080/jtde.v12n4.1017

path is blocked, whereas the proposed method captures the entire map and identifies blocked paths, allowing better global path planning. The proposed method also reduces the sensors needed on the robot, allowing cheaper processors while improving battery life.

A common method for mapping and localisation is the SLAM algorithm. To compare AMR using SLAM, TurtleBot3 is used, which has four officially supported SLAM algorithms: Gmapping, Cartographer, Hector SLAM, and Karto SLAM. From previous studies, it is found that Cartographer is the best in terms of accuracy followed by Gmapping. Hector SLAM accuracy decreases with more turns as no odometry data is used, whereas Karto SLAM accuracy decreases if it cannot close the loop. In terms of mapping speed, Cartographer is the fastest at 16.4 minutes, whereas Gmapping is 19.7 minutes, which is the reason Cartographer is chosen as the SLAM comparison (Chen *et al.*, 2021; Li *et al.*, 2021). Cartographer is a two-dimensional (2D) light detection and ranging (LiDAR) SLAM that uses scan-to-submap matching and pose graph optimisation to reduce the accumulation of errors (Hess *et al.*, 2016; Yang *et al.*, 2022).

Continuous improvements on Cartographer SLAM are made by reducing the odometer and inertial measurement unit (IMU) sensor reading error by fusing it with extended Kalman filter (Zheng *et al.*, 2024). Another paper introduces map-to-map feedback for close loop detection in environments with fewer features, improving its mapping accuracy (Yang *et al.*, 2022). Applications of LiDAR SLAM is seen by improving its use cases by attaching a robotic arm to an AMR to pick and place objects (Chen *et al.*, 2023; Kusashio *et al.*, 2024).

Replacing the 2D LiDAR with other cheaper sensors like cameras has led to the development of ORB-SLAM3 (Campos *et al.*, 2021) and Object-Visual SLAM (Adkins *et al.*, 2024). Its larger uncertainties in its depth estimation without a LiDAR sensor causes a distorted map to be created (Kok *et al.*, 2023). In addition, image recognition techniques tend to require heavy processing power, which is not always available on each individual AMR (Ahmed *et al.*, 2023).

To achieve perception, an RGB camera is used with an object recognition algorithm. Object recognition uses convolutional neural network (CNN), which is trained to identify objects (Anoopa *et al.*, 2023). Comparing image recognition between region-based CNN (R-CNN), Faster R-CNN, You Only Look Once version 4 (YOLOv4) and Single Shot Multibox Detector (SSD), Kim *et al.* found that YOLOv4 performed the best at object identification among the four methods (Kim *et al.*, 2020). Several other studies found that Faster R-CNN performed the best at object identification but at a slower speed (Krishna *et al.*, 2021; Tan *et al.*, 2021). Another problem is YOLOv4 has difficulties recognising small objects compared with Faster R-CNN, whereas for SSD, its accuracy is lower than Faster R-CNN (Niranjan *et al.*, 2021).

https://doi.org/10.18080/jtde.v12n4.1017

There are several techniques to solve localisation problems, such as beacon-based localisation using beacons like Bluetooth, Wi-Fi, or radio frequency signal to determine the location by measuring different parameters of the signal emitted. The average error of location is found to be above one metre (Kaewpinjai *et al.*, 2020; Mavilia *et al.*, 2023). These techniques only provide localisation unlike SLAM that does both localisation and mapping. Another approach is marker-based localisation using radio frequency identification tags, QR code, or barcode to determine location. The markers are placed throughout the environment at strategic locations, and a scanner on the robot will read it and know its exact location on the map. The drawback of this technique is that the reader has to be close to the tag to read it (Gunawan *et al.*, 2022).

Path planning is used to achieve cognition, which are split into global and local path planning, in which global path planning uses static known information, like Google maps. Local path planning is real-time dynamic path planning based on information available around the robot, like avoiding other vehicles and barriers on the road (Shen *et al.*, 2021; Wang & Mao, 2019). Among global path-planning methods, A* have the shortest path, although the disadvantage is the time taken for the algorithm to compute (Qiong *et al.*, 2020). Research on improving A* is still ongoing, indicating it is still relevant and applicable as a global path-planning algorithm (Yi & Hongtu, 2019). The path planning and localisation can be simulated in programs such as Robot Operating System and Gazebo.

Motion control is to modulate its actuators to achieve the desired input; depending on the type of robot used, the motion control will be different due to the kinematic constraints of the robot (Roland *et al.*, 2011). The TurtleBot3 uses a proportional integral (PI) controller for velocity control and PI derivative controller for position. Because the TurtleBot3 has its own motion control algorithms, this aspect of navigation was not further investigated.

Methodology

Computer vision with automated algorithms for navigation is chosen for this paper as it allows the mapping to be completed in a shorter time, which is important because the map needs to be updated frequently to account for the dynamic environment. Furthermore, this method bypasses the need for comprehensive and highly accurate onboard sensors on the robot, allowing the method to be used despite a factory having robots with a variety of makes and sizes that may not be equipped with onboard sensors. Furthermore, such environments tend to have security cameras placed at strategic locations to maximise coverage. Leveraging this, images or videos from cameras mounted on walls and ceilings can be used to create a map and navigate a robot to its destination.

The navigation system uses the top camera to map and localise the environment and then feed the information to the TurtleBot3 via Wi-Fi using Samba protocol. TurtleBot3 then uses A*

https://doi.org/10.18080/jtde.v12n4.1017

path-planning algorithm to plot the shortest route while using LiDAR for obstacle avoidance. The top camera is then used to feedback to the TurtleBot3 whether it has reached its destination. Figure 1 shows the flowchart of the proposed method.

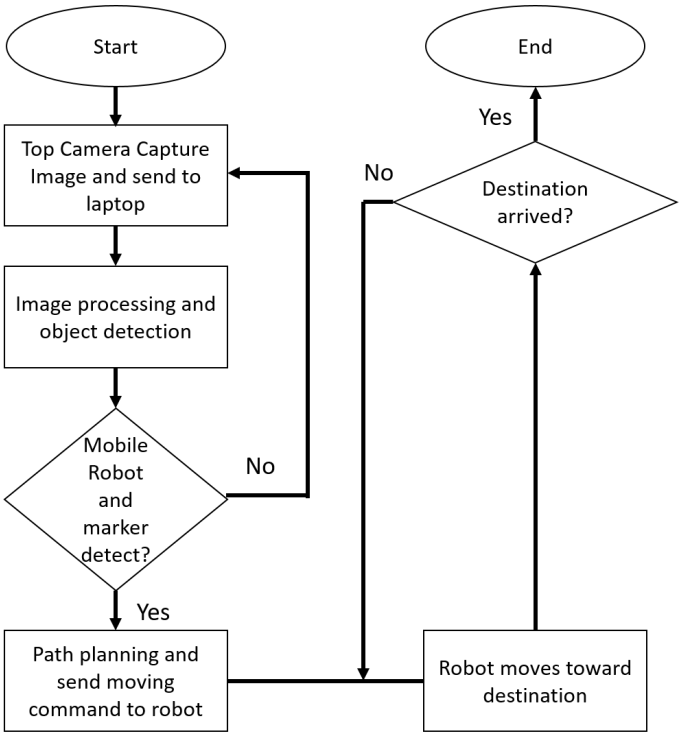


Figure 1. Flowchart of the proposed system architecture.

Physical setup and initial image capture

The TurtleBot3 Burger was chosen as the robot because it is a small and versatile platform to aid development without extensive modifications. This robot is configured to use the standard onboard LiDAR sensor to assess the performance of this method in more practical scenarios in which state-of-the-art equipment may not be available. The onboard Raspberry Pi 3 Model B is connected by secure shell protocol through Wi-Fi to control the robot's physical movement and read onboard sensors. A combination of odometry (measuring the rotation of the wheels) and the LiDAR sensor is used to aid the navigation of the robot. The detailed specifications are shown in Table 1. The minimal distance of 10 mm to trigger the obstacle detection is used as an arbitrary set value for the LiDAR sensor because the sensor reading drops to zero for obstacles less than 9.9 mm from the middle of the sensor. The robot has a rotation radius of 80 mm (100 mm from the middle of the sensor to the edge of the circle of rotation), so this distance is small enough for the rotation without triggering false positives.

The computing hardware used is a central "server" laptop computer (Asus S451LN laptop with 4 GB RAM and GPU of NVIDIA GeForce 840M) that can be used for the whole process. The cameras are used in a distributed fashion by connecting each to a Raspberry Pi 4 Model B to send the images to the central server.

https://doi.org/10.18080/jtde.v12n4.1017

Table 1. Specifications for Turtlebot3 Burger.

Specification	Description
Maximum velocity	0.22 m/s
Maximum payload	15 kg
Threshold of climbing	10 mm or lower
Battery	Lithium polymer 11.1 V 1800 mAh, expected to last 2 h, 30 min
Single board computers	Raspberry Pi 3 Model B and B+
MCU	OpenCR1.0, 32-bit ARM Cortex-M7 with FPU (216 MHz, 462 DMIPS)
Light distance sensor	360 laser distance sensor LDS-01

Because an RGB camera can be used to identify objects, it has been decided that the same sensor can be used for localisation. This equipment is also common enough that it is possible to perform the assessments without the need to extensively retrofit the environment if scaled up. A Raspberry Pi 4 with 8-megapixel camera is mounted on the ceiling as shown in Figure 2.

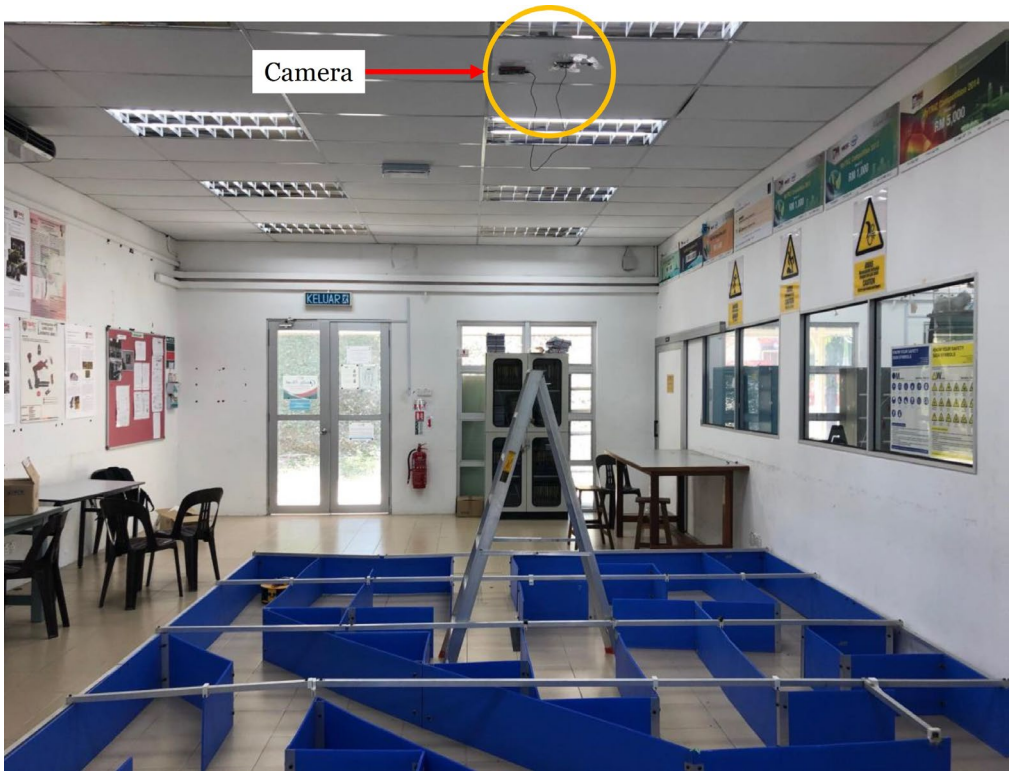
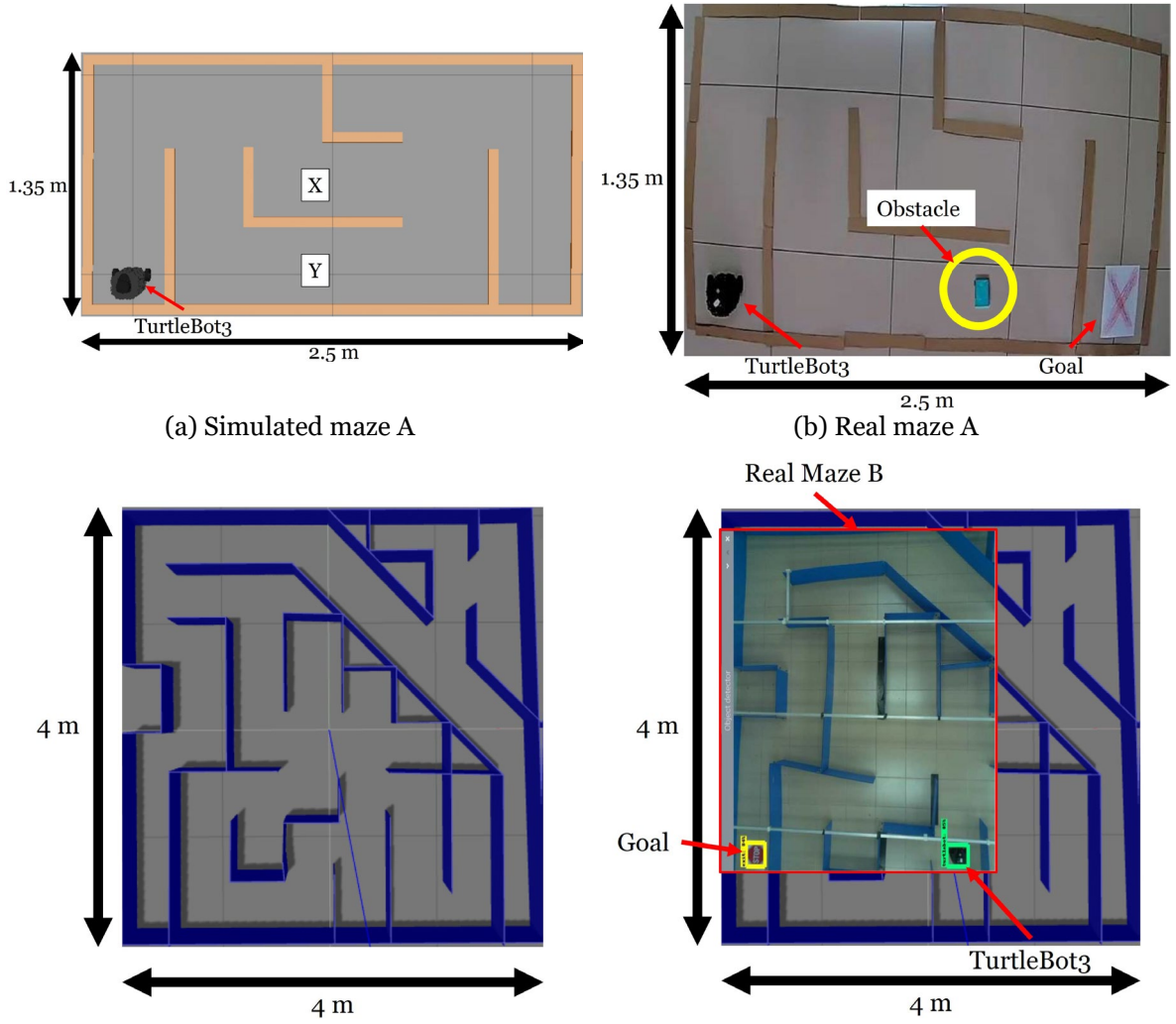


Figure 2. Overall experiment setup for maze B.

A simpler 2.5 m × 1.35 m maze is constructed in Gazebo simulation as shown in Figure 3 (a) and real life as shown in Figure 3 (b). This maze is referred to as "Maze A". This maze has nine straight segments with two paths labelled X and Y. The walls of Maze A are constructed using flat cardboard pieces to reduce the complexity from the robot colliding with the wall and make modifications to the layout simpler. Because this maze is constructed with flat pieces of

https://doi.org/10.18080/jtde.v12n4.1017

cardboard, the LiDAR sensor cannot detect them, and so the camera image processing is the only method of navigating to avoid collisions.



(a) Simulated maze A (b) Real maze A

(c) Simulated maze B (d) Real maze B overlaid on simulated maze B

Figure 3. Simulated and real top-down view of mazes A and B.

A larger and more complex 4 m × 4 m maze is also tested and referred to as Maze B. This maze is constructed both in Gazebo simulation as in Figure 3 (c) and in real life as in Figure 3 (d) with taller walls to facilitate testing with a longer path to the goal to make it more challenging for both algorithms to map and to navigate. Due to limited ceiling height and Raspberry Pi camera viewing angle without additional lens, the real Maze B only captures a section of the entire maze measuring approximately 3.6 m × 2.7 m. This maze has 27 straight segments, although only 17 are used for this real-life experiment. In this case, the maze has tall 200-mm cardboard walls; so the LiDAR can detect these walls, and the robot can collide with the maze and lose its odometry localisation.

An obstacle is then introduced to Maze A in random locations, for example as annotated in yellow in Figure 3 (b). In this case, the obstacle chosen is a light blue whiteboard duster as it

https://doi.org/10.18080/jtde.v12n4.1017

does not completely block the path and is simpler to determine its edges. The goal used is a printed red STOP sign and a white paper with a written "X" on it. Both are to evaluate if arbitrary objects of different complexity and contrast can be detected and reconfigured as navigation markers. These mazes are tested both in idealised simulations as well as using the whole process of converting the physical maze into digital coordinates to then be used for physically navigating the robot to the goal.

Object detection, image processing, and path planning

Computer vision is used to detect the initial position of the TurtleBot3 and the target goal. Any obstacles can also be detected the same way. These can be threshold, and the positions marked as solid square shapes to act as impassable areas in the black and white threshold maze image. This marks the coordinates for the path planning using A* algorithm. Control signals are then sent to physically move the robot. Figure 4 shows the flowchart for image processing.

First, the lighting is kept consistent as the experiment is ran indoors with overhead lighting from the ceiling lights, although some small changes can be observed throughout the day from the sunlight shining through the windows and the resulting reflections. Objects and people moving around also affect the shadows and reflections, but this is minimised because the area is kept clear of additional moving objects and people during the experiment. These lighting conditions only need to be kept consistent for the initial image taken for the localisation and path-planning process, and during the movement of the robot, this is no longer necessary because there are no images taken to repeat the localisation after this initial step.

The image captured only at the start of the process by the top camera is converted into grayscale in which the brightness of each pixel is between 0 and 255. Thresholding is then applied, in which each pixel is set to either 0 or 255, turning darker parts into black and brighter parts into white. The image then undergoes median filtering to remove noise by taking the values around a pixel and sets it to the same median value of the other pixels around it within a set range. This range is adjusted manually at the start of each real-life experiment by increasing or decreasing the setting and observing if too many features are removed from the final image or if too much noise remains. It can then be reused for subsequent experiments viewing the same environment. This removes small features such as grooves between tiles, reflections, and noisy details of the obstacles to simplify their shapes for detection.

The threshold value setting is manually selected in this case by visually determining the threshold value most suitable for isolating the free path. The boundary to select as pure white (255) or pure black (0) pixels is increased and decreased manually by inputting the value into the program with a sample photo taken just before the experiment has started. White pixels

https://doi.org/10.18080/jtde.v12n4.1017

indicate free path, whereas black pixels indicate obstacles. Once the threshold value is calibrated for a given lighting condition, it can be reused, and this step can be skipped.

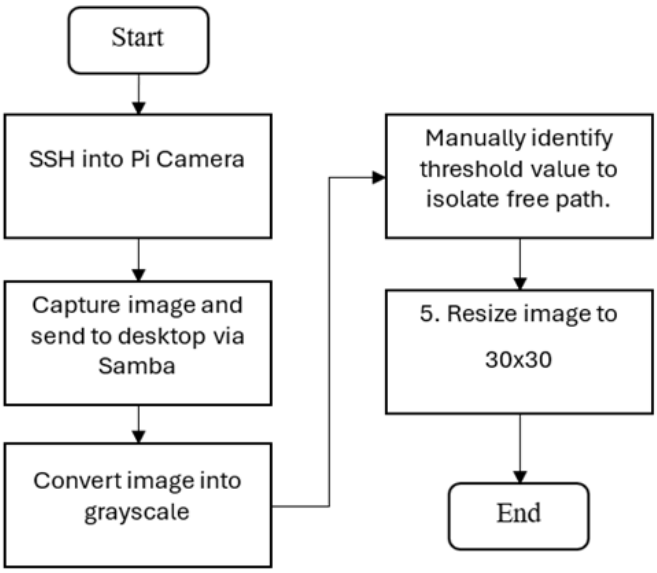
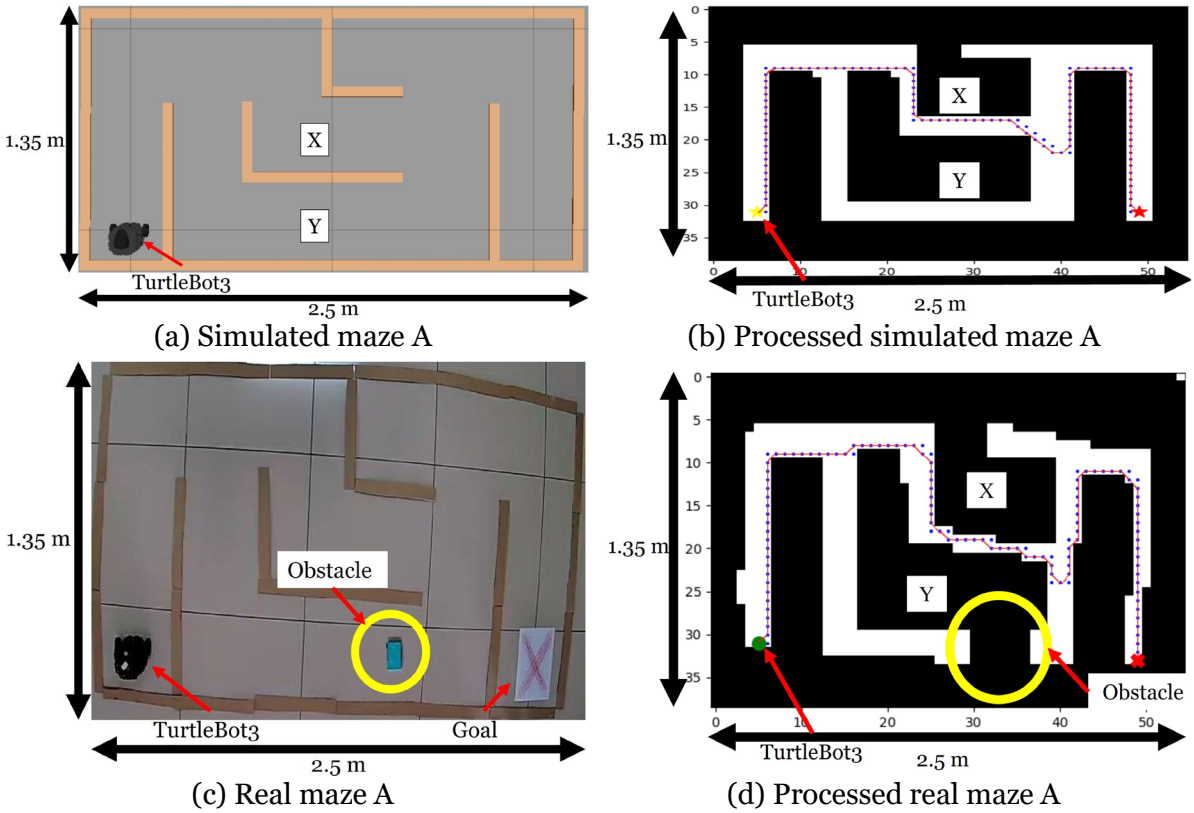


Figure 4. Flowchart of image processing for map creation.



(a) Simulated maze A (b) Processed simulated maze A

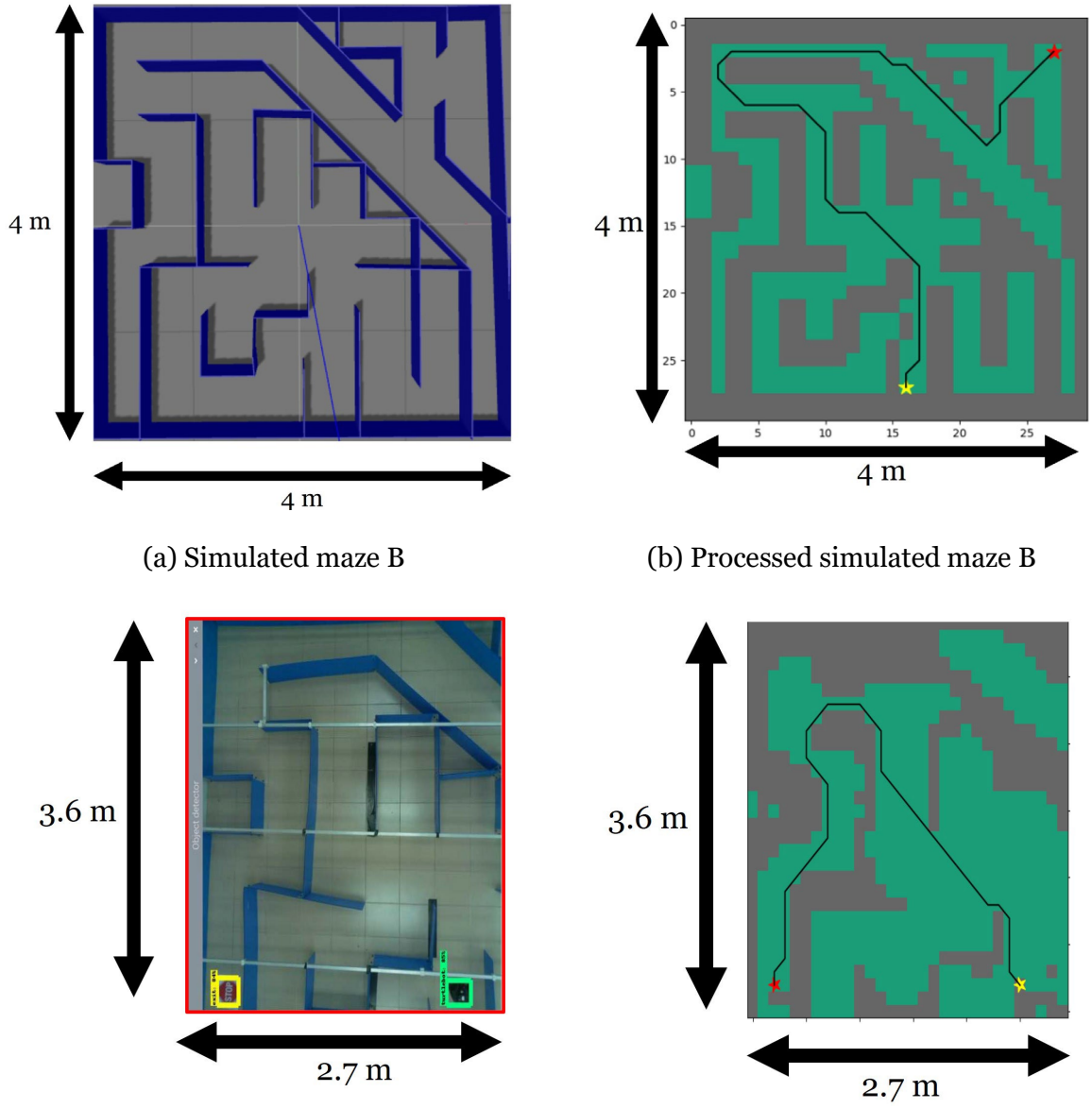
(c) Real maze A (d) Processed real maze A

Figure 5. Processed images of maze A in simulation and real life.

This method works if there is a distinguishable grayscale colour between the free path and obstacles. This converts the walls of the maze into a black and white image, which can be fed into the path-planning algorithm as seen in the corresponding processed images of Figure 5 and Figure 6 (b) and (d). However, the limitation is that reflections and shadows can disrupt

https://doi.org/10.18080/jtde.v12n4.1017

the thresholding such that these darker shadow areas can be mistaken as obstacles, and reflective obstacles may appear as open paths by this thresholding. This effect is most noticeable in real Maze B processed image where some walls are missing and with the top path completely mistaken as obstacle. The darker-coloured grooves between the tiles in Maze A causes this issue without median filtering.



(a) Simulated maze B

(b) Processed simulated maze B

(c) Real maze B

(d) Processed real maze B

Figure 6. Processed images of maze B in simulation and real life.

To compensate for the fisheye lens effect on the camera images as seen in real Maze A shown in Figure 5 (a), known dimensions such as the edges of the maze and other features are used to calibrate different sections of the distorted image by scaling the image up or down at those locations according to the ratio found. This is done manually for the first time of each setup

https://doi.org/10.18080/jtde.v12n4.1017

and maze and then reused for subsequent repeats; for example, for Maze A, the scale found is 8.1 for the left, 8.4 on the middle, and 7.4 on the right side.

YOLOv4 is chosen because it has faster processing while having adequate accuracy, which will be crucial when scaling up with multiple cameras and robots to identify. Faster R-CNN with Inception Resnet V2 is also chosen for further testing as it has a higher accuracy but also slower processing, so it is more effective when used in a more powerful central server computer. A compromise between accuracy and speed must be made because the TurtleBot3 moves rather slowly; Faster R-CNN is the better choice for image recognition as the obstacles faced by TurtleBot3 can be small. However, YOLOv4 is also tested to show that the lightweight but less accurate method can still successfully be used. The obstacles detected with this method are then coloured as black pixels as well.

With the map and coordinates of the TurtleBot3 and goal, the path-planning algorithm A* can begin to find the shortest path. This whole process of mapping using image processing and finding the shortest path is referred to as the "proposed method".

Evaluation experiments

These experiments are first carried out in Gazebo simulation to validate its consistency and its average time in perfect simulated conditions. This is to isolate effects such as the differences in the perfect maze of the simulation shown in Figure 5 and Figure 6 (a) because the map processed from real life will end up with paths that are no longer straight due to optical inaccuracies such as the camera's fisheye effect and angled perspectives because of small misalignments in the camera.

The experiment is then attempted in real life to validate the algorithm when faced with physical inaccuracies, such as the TurtleBot3's wheels having a slightly different size, so even when rotated at precise angles, the distance travelled is slightly different, resulting in straight movements becoming slightly angled. The slippery floor and gaps between tiles also contribute to the wheel movement becoming inaccurate.

For Maze A, the accuracy of movement is prioritised over minimising the travel time, so the robot is instructed to totally stop, then rotate, and then resume movement for each node (step) of the path. This results in more accurate movement at the nodes to avoid collisions but a longer travel time because they need to stop before rotating. The robot is also instructed to move at 0.08 m/s moving forward, which is approximately 36% of its full speed of 0.22 m/s. This makes the speed more stable as wear and tear reduces the motors' ability to meet this maximum rated speed, and higher speeds also can cause the wheels to slip on the slippery tiles.

https://doi.org/10.18080/jtde.v12n4.1017

In comparison, for Maze B, the robot is instructed to continue moving forward at 0.1 m/s speed while turning. It is also instructed to move at the highest possible speed by setting it to move at 0.25 m/s when travelling straight. This is to make the movement smoother and challenge the system to react to obstacles for the worst case of traveling quickly to the goal.

To test the viability of the basic algorithm, the proposed method is first tested using YOLOv4 to detect objects. For navigation, it uses the "enhanced Euclidean heuristic", which is Euclidean heuristic with minimal safe distance obstacle avoidance using the simpler Maze A without obstacles. This is to test the viability of the algorithm to detect the TurtleBot3, goal, and obstacles with a less accurate object recognition algorithm. Because Maze A walls are short, flat pieces of cardboard, the LiDAR sensor is not able to detect these obstacles, and so the camera mapping is the only viable method to avoid collision. The TurtleBot3 is placed on the bottom left of Maze A while the goal is also on the bottom right. The timer is started when the workflow is started and stopped once the TurtleBot3 reaches the goal. The maze is processed into 55 × 39 pixels for the map.

This experiment is then repeated by introducing an obstacle (in this case, a blue whiteboard duster) into the maze as seen in Figure 5 (c) and that diverts the TurtleBot3 from its shortest path—in this case, the bottom straight section Path Y. This experiment is repeated 30 times, and the average time and number of collisions are measured.

Once it is validated that the algorithm is stable, it can be compared with the more conventional SLAM navigation method using LiDAR sensors for localisation. The more complex Maze B is then used to test the viability when using a slower object recognition method with heavier load in a more complex environment. Additionally, the "proposed method" is used with Faster R-CNN for the recognition of the TurtleBot3 and the goal (printed red stop sign). This is to further increase the computational load in the more complex maze to observe if the calculations can fail with the higher load.

Maze B also has taller walls and so can be detected using the LiDAR sensor for SLAM navigation. These taller walls also introduce an additional issue that can disrupt the localisation using odometry because collisions will result in the wheel spinning without the expected robot movement.

Similar to the previous experiments, the proposed method with enhanced Euclidean heuristic is compared with SLAM in a Gazebo simulation first and then in real-life tests to validate the results. The criteria of comparison are time taken, path cost, and collision. In simulation, Maze B is generated as in Figure 6 (a) and processed from its original pixel size to a 30 × 30 pixel size shown in Figure 6 (b). A path-planning algorithm is then executed to plot the path shown as a black line.

https://doi.org/10.18080/jtde.v12n4.1017

Additional tests are performed in Maze B to compare different heuristic methods. The enhanced Euclidean method is compared with the navigation when using only Euclidean heuristic and then using only Manhattan heuristic to determine its reliability and performance in real life. This is to validate that the enhanced Euclidean heuristic is more reliable because even in the complex environment it can navigate diagonally compared with the Manhattan heuristic. It is also able to prevent and correct errors from colliding with the walls because it can detect a wall using LiDAR and perform obstacle avoidance before the collision disrupts the localisation. For the real-life experiment of Maze B, because of camera viewing angle limitations, only a partial section of Maze B is used for the validation, as shown in Figure 6 (c) with its corresponding processed image shown in Figure 6 (d).

These experiments can validate the algorithm for use in situations with a low computational load but lower accuracy, higher computational load, in simulation and real life, and in comparison, with different navigational heuristic methods. This also shows the versatility of the algorithm because it can be modified with different components and evaluates the resulting changes in efficiency.

Results and Discussion

The experiments validated the performance of the algorithms and compared them when used in different situations. This is done in Maze A and B in simulation and real life and by varying the object recognition used and the type of navigation heuristic.

Evaluation in Maze A with obstacles

The proposed method is tested in the smaller Maze A in simulation and real life, which is repeated 30 times to validate performance and reliability. Table 2 shows the time taken in simulation averaging 2 minutes 9 seconds with a maximum and minimum of 2 minutes 10 seconds and 2 minutes 7 seconds, respectively. This timing in the simulation is consistent with no collisions. The path chosen is through Path X as the shortest path. The total nodes (steps) of the path generated is 101, and the robot rotated 765° (17 turns).

From Table 3, it is seen that the time taken to complete the maze without obstacles in real life is consistent, averaging 2 minutes 21 seconds with a maximum and minimum of 2 minutes 24 seconds and 2 minutes 18 seconds, respectively. Only a single collision was recorded in round 5, where Path Y is the path chosen in this case. The total nodes (steps) of the path generated is 120, and the robot rotated 630° (14 turns).

As shown above in Figure 5 (b) and (d), the planned shortest path changed due to inaccuracies in the real-life processed map, changing the straight path of the maze into an angled one. As the resolution and movement used does not allow for angles in increments less than 45°, this

https://doi.org/10.18080/jtde.v12n4.1017

results in a zig-zag motion that takes much longer due to additional turns and more steps. This is besides the difference from the warping from the construction of the maze.

When an obstacle is introduced, Path Y was blocked, and so it was observed that a new path through the previously longer zig-zag Path X above was chosen. The time taken to complete the maze increases as shown in Table 4, with the average time taken to complete the maze is 2 minutes 51 seconds with a maximum and minimum of 2 minutes 53 seconds and 2 minutes 48 seconds, respectively, without any collisions.

Table 2. Time taken and number of collisions for TurtleBot3 to complete Maze A in simulation. This experiment uses Path X.

Round	Time taken (min:s)	Collision	Round	Time taken (min:s)	Collision
1	2:07	0	16	2:07	0
2	2:09	0	17	2:08	0
3	2:10	0	18	2:10	0
4	2:10	0	19	2:10	0
5	2:08	0	20	2:09	0
6	2:09	0	21	2:10	0
7	2:08	0	22	2:08	0
8	2:09	0	23	2:10	0
9	2:10	0	24	2:10	0
10	2:10	0	25	2:09	0
11	2:08	0	26	2:10	0
12	2:10	0	27	2:07	0
13	2:10	0	28	2:09	0
14	2:09	0	29	2:10	0
15	2:10	0	30	2:10	0
AVERAGE				2:09	Total: 0

Table 3. Time taken and number of collisions for TurtleBot3 to complete Maze A in real life. This experiment uses Path Y.

Round	Time taken (min:s)	Collision	Round	Time taken (min:s)	Collision
1	2:20	0	16	2:23	0
2	2:21	0	17	2:22	0
3	2:18	0	18	2:20	0
4	2:19	0	19	2:20	0
5	2:24	1	20	2:22	0
6	2:22	0	21	2:22	0
7	2:20	0	22	2:21	0
8	2:21	0	23	2:18	0
9	2:21	0	24	2:19	0
10	2:23	0	25	2:19	0
11	2:18	0	26	2:21	0
12	2:20	0	27	2:21	0
13	2:18	0	28	2:20	0
14	2:19	0	29	2:22	0
15	2:21	0	30	2:23	0
AVERAGE				2:20	Total: 1

https://doi.org/10.18080/jtde.v12n4.1017

Despite the same number of nodes for Path X and Path Y that the robot needs to pass through, the extra rotation angles resulted in more steps of movement and so a longer time to navigate. So, the algorithm chose Path Y as shorter. This shows the algorithm is able to choose a new shortest path automatically when obstacles are introduced to block the original path.

To summarise the findings of these experiments, a comparison between the number of nodes and rotations is shown in Table 5. For all these cases, the computing time is negligible compared with the movement of the robot. This compares the difference between the simulated and real-life Path X that became jagged due to the image processing and warped maze to produce more rotations and steps to navigate. It also compares that Path Y became chosen as the shortest path because the algorithm considered the number of rotations needed and not just the number of nodes. Finally, it shows the algorithm can automatically select the new shortest path (Path X) again if an obstacle is blocking the previously determined shortest path (Path Y).

This shows the method is consistent while being used in real life and not only in simulations. It also showed that the TurtleBot3 was able to navigate despite small errors from the localisation inaccuracies stated in Methodology and Evaluation experiments.

Table 4. Time taken and number of collisions for TurtleBot3 Burger to complete the Maze A with the presence of an obstacle. This experiment uses Path X.

Round	Time taken (min:s)	Collision	Round	Time taken (min:s)	Collision
1	2:51	0	16	2:53	0
2	2:53	0	17	2:49	0
3	2:48	0	18	2:49	0
4	2:53	0	19	2:53	0
5	2:50	0	20	2:50	0
6	2:51	0	21	2:53	0
7	2:52	0	22	2:52	0
8	2:51	0	23	2:53	0
9	2:48	0	24	2:50	0
10	2:49	0	25	2:50	0
11	2:48	0	26	2:51	0
12	2:53	0	27	2:52	0
13	2:52	0	28	2:50	0
14	2:51	0	29	2:51	0
15	2:52	0	30	2:51	0
AVERAGE				2:50	Total: 0

Table 5. Comparison between Path X and Y and the number of nodes and rotations taken between real-life and simulated paths. This is summarising Table 2, Table 3 and Table 4.

Navigation type	Path Type	Nodes (Steps)	No. of 45° rotations	Average Time (min:s)
Simulation	X	101	17	2:09
Real life	Y	120	14	2:20
Real life with obstacle	X	120	24	2:50

https://doi.org/10.18080/jtde.v12n4.1017

Evaluation in Maze B with SLAM

It is found that despite the higher computational load of using Faster R-CNN and the larger, more complex maze, the TurtleBot3 and goal were still successfully recognised and detected without a significant increase in time, in that the increased total time taken for the enhanced Euclidean heuristic is still mostly due to the travel time, whereas the increased processing time is negligible within the range of milliseconds to process the captured maze. As compared in the literature review, the difference is 82.1 frames per second (FPS) for YOLOv4 and 36.32 FPS for Faster R-CNN (Kim *et al.*, 2020).

In this case with the larger, more complex maze, the proposed method was plagued with cumulative inaccuracies from the motor causing poor motion control. Due to the uneven surface caused by the tiles, the TurtleBot3 got stuck in between the grooves, causing it to fail. The SLAM method is observed to fail significantly more often than the proposed method, although SLAM takes a significantly longer time due to the extra time needed for mapping.

Another issue faced with the proposed method is in the map creation. Because grayscale thresholding method is used, it is observed that highly reflective obstacles are mistaken as an open path. Furthermore, if the free path and obstacle have no distinguishable colour differences, the method will not be able to isolate free paths from obstacles. However, the light blue duster shown in Figure 5 (c) could still be detected, despite it having a relatively low contrast. These errors are reduced as consistent lighting is used during the experiments by performing the experiments indoors with overhead lighting from indoor bulbs.

First, the path planning is validated as functioning using a screenshot of the simulated maze as shown in Figure 6 (a). This is only to confirm that the method can function, so for direct comparison only, the setup as in Figure 6 (c) is presented. It is then validated using camera images of the maze with different start and end positions. It is observed that the time taken for the processing is still negligible compared with the travel time of the robot, that is within seconds and with almost no difference for different variations. It is also negligible compared with the process for SLAM.

From simulation in Gazebo, the path cost is calculated from a 30 × 30 grid, as is shown in Table 6. The smaller path cost is from the diagonal motion of the Euclidean heuristic, which translates to significantly shorter time taken as shown when compared with the Manhattan heuristic.

https://doi.org/10.18080/jtde.v12n4.1017

Table 6. Movement taken and path cost in a fixed path for Maze B.

Type of heuristic	Generate path on 30 × 30 pixel array	Path cost
Manhattan	(16,27), (16,26), (16,25), (16,24), (16,23), (16,22), (16,21), (15,21), (14,21), (13,21), (12,21), (11,21), (11,20), (11,19), (11,18), (11,17), (11,16), (11,15), (11,14), (10,14), (10,14), (9,14), (8,14), (7,14), (7,13), (6,12), (6,11), (7,11), (8,11), (9,11), (9,11), (10,11), (11,11), (12,11), (9,1), (9,2), (10,2), (10,3), (11,3), (11,4), (12,4), (12,5), (13,5), (13,6), (14,6), (14,7), (15,7), (16,7), (16,6), (16,5), (16,4), (16,3), (17,3), (18,3), (18,2), (18,1), (27,2)	57
Euclidean	(16,27), (16,26), (16,25), (16,24), (16,23), (16,22), (15,11), (9,11), (8,11), (7,11), (6,10), (6,9), (6,8), (6,7), (6,6), (6,5), (5,4), (4,4), (3,4), (2,4), (1,3), (1,2), (2,1), (3,1), (4,1), (5,1), (6,1), (7,1), (8,1), (9,1), (10,2), (11,3), (12,4), (13,5), (14,6), (15,7), (16,6), (16,5), (16,4), (17,3), (18,2), (18,1), (27,2)	43

Table 7 shows the time taken for different heuristics used and SLAM in Gazebo simulation in Maze B. Collisions with the walls or inability to move was recorded as failure. It is seen from this that using the Euclidean heuristic, which allows for diagonal movement, is faster than Manhattan heuristic averaging 3 minutes 30 seconds compared with 4 minutes 46 seconds for Manhattan but a success rate of 40%. The diagonal movement causes the TurtleBot3 to clip the wall; hence, by using the LiDAR sensor readjustments, it is made to maintain a minimal distance to the wall. The LiDAR sensor could be replaced with a cheaper range sensor. Pairing this with Euclidean heuristic, the enhanced Euclidean heuristic maintains the same average time as the Euclidean heuristic while having no collisions. The performance of SLAM is not reliable, having only succeeded 40% while taking a significantly longer time, averaging 20 minutes and 38 seconds.

Table 7. Time taken for TurtleBot3 to complete simulated maze B for Euclidean heuristic, Manhattan heuristic, enhanced Euclidean heuristic, and SLAM

Round (n)	Euclidean heuristic	Manhattan heuristic	Enhanced Euclidean heuristic	SLAM
	Time taken (min:s)	Time taken (min:s)	Time taken (min:s)	Time taken (min:s)
1	3:31	4:48	3:31	15:37
2	Failed	4:45	3:30	Failed
3	Failed	4:45	3:29	Failed
4	3:30	4:46	3:30	25:39
5	Failed	4:47	3:30	Failed
Average	3:30.5	4:46.2	3:30	20:38
Success (%)	40	100	100	40

https://doi.org/10.18080/jtde.v12n4.1017

Table 8 shows the comparison of SLAM and the proposed method with enhanced Euclidean heuristics in real life with a section of Maze B. The success rate of the proposed method was 40% compared with SLAM at 60%. The issue is due to the accumulation of errors in localisation via odometry due to the slippery conditions of the floor and variations of wheel size and conditions. This issue can be improved by continuously providing feedback of the location through the camera instead of only once at the beginning. The average time taken was 7 minutes 14 seconds compared with SLAM's 19 minutes 59.7 seconds because as with the simulation results, the proposed method is faster. SLAM requires travel to map the environment, whereas the proposed method obtains it through the camera overviewing the environment.

Table 8. Time taken for TurtleBot3 navigation in Maze B using SLAM and proposed method with enhanced Euclidean heuristic in real life.

Round (n)	Proposed method with enhanced Euclidean heuristic	SLAM
	Time taken (min:s)	Time taken (min:s)
1	6:57	Failed
2	Failed	15:18
3	Failed	Failed
4	Failed	22:48
5	7:32	21:53
Average	7:14	19:59.7
Success rate (%)	40	60

Limitations observed

From these experiments, it is observed that the system is able to automatically determine the shortest path through Maze A and Maze B, so it can handle different path complexities and the presence of obstacles blocking their normal shortest paths.

However, during navigation, it has some reliability issues because it can lose its localisation and collide more frequently, although this is similarly experienced by the conventional SLAM navigation. However, compared with conventional SLAM methods, the proposed method is still significantly faster to localise and navigate, improving the time from approximately 20 minutes to 7 minutes to perform the whole process.

The four aspects of navigation, perception, localisation, and cognition can be deemed successful in real life, whereas motion control is not as reliable due to improper feedback system and cumulative motor inaccuracies when navigating a larger maze. However, with further development on the feedback by increasing the computer vision update frequency and

https://doi.org/10.18080/jtde.v12n4.1017

processing, this should be adequate to provide navigation in a constantly changing environment while maintaining the shortest route.

The first and main limitation of this proposed method is that uneven surfaces of the maze caused by the tiles results in the TurtleBot3 getting stuck in between the grooves, resulting in inaccuracy and errors. These grooves introduce some extra distance detected by the odometer, whereas slippery tiles can make the wheels lose grip and move a shorter distance than expected. The TurtleBot3's IMU can also lose accuracy if used for a long distance without recalibration. However, these sensor issues can be minimised in future applications by doing calibrations again in the middle of the navigation operation. Localisation and path planning can be done periodically or when an unexpected obstacle is detected instead of only at the start.

The second limitation is mapping the environment using image processing methods such as thresholding, which can fail to differentiate path and wall due to a shadow mistaking it as a wall or illumination by light bulbs on reflective material mistaking it as a path.

The third limitation will be the TurtleBot3 and exit detection. As the objects get further away, more features are lost and can no longer be recognised by the neural network. The recognition can also fail if the angle is isometric from certain angles that were not trained in the pretrained model, although this is reduced because the camera angle chosen is directly from the top. This model can still function when certain markers such as a coloured identification sticker is attached on top of the TurtleBot3, so this opens possibilities for future work.

The last limitation will be the coverage area. Because the coverage area that is the maze is not constantly fixed in place, the wall of the maze can be moved by accident. This causes frequent tuning in scale for the TurtleBot3 because every time the wall moved, tuning will be required to ensure the TurtleBot3 will not collide with the wall while using the same scale and path. This suggests further work on real-time calibration can be considered in cases in which the camera or walls can experience changes in maze features over time. If there are no collisions or changes to the path from the navigation maze image at the start of the process, this is not required.

Besides these limitations with the image processing workflow, some hardware limitations were also present. The camera viewing area is limited by the viewing angle and picture quality of the Raspberry Pi camera. When adding a fisheye lens, the current rudimentary method of dealing with the image distortion by rescaling different sections of the picture is no longer enough to deal with the larger 2D distortion, so this method was not presented. It is possible to concatenate the images of two or more cameras to stitch together the image of a larger area, but this is not yet explored. Image concatenation can introduce some complexity, such as the

https://doi.org/10.18080/jtde.v12n4.1017

issue of overlapping images at the seams and the slight distortions from the different camera angle of the cameras.

It is possible to run it using smaller computers such as Raspberry Pi and Jetson Nano if the Gazebo simulation is not needed because the image processing and path planning are relatively lightweight in comparison. However, because the simulations and visualisations are needed for the comparison, this was not explored in detail. Instead, this paper presents the analysed time as from the start of the process until the TurtleBot3 reaches its destination because it is a fairer comparison with existing methods such as SLAM that need to also include the mapping time into the process. The time taken for the image processing and path-planning calculations on the desktop is negligible (within seconds) compared with the actual movement of the robot and so was not focused on in this paper. Also, the processing requirements increase when using a maze size larger than 30 × 30 pixels, but this is not significantly explored; instead, a resolution in which one pixel is approximately the footprint of the TurtleBot3 is used for all experiments.

Overall, these limitations open opportunities for further work to improve the algorithm to be more reliable with dynamic environments. With the current experimental setup, it is capable of planning paths with obstacles that appear at the start of the navigation and can react with the help of onboard sensors such as the LiDAR detecting new obstacles and repeating the localisation and path-planning process.

Conclusion

In conclusion, a navigation method has been developed using TurtleBot3 and a camera mounted to the ceiling. Through simulation and experiment, it is shown that the proposed method can map and navigate through a simple maze faster than SLAM because of the time taken for SLAM to map the environment. In more complex mazes, it is unreliable due to inaccuracies from odometry and IMU data, wheel slip or getting stuck, and lack of localisation feedback. When obstacles are introduced, unlike SLAM, the proposed method can detect the obstacle and plot a path to avoid from the start.

Mapping via the camera using grayscale thresholding is seen to have more inaccuracies than SLAM due to being highly dependent on lighting conditions. Here, we simplified the mapping by using a top-down view of the maze, which will need to be improved to accommodate camera views at an angle as well as with differing lenses.

Localisation using Faster R-CNN and YOLOv4 in the framework showed no difference because the travel time took significantly longer than the processing time. Larger maps and multiple

https://doi.org/10.18080/jtde.v12n4.1017

robots may yield some differences. More work is needed for localisation of multiple robots of the same type.

Out of the four aspects of navigation, we conclude that perception, localisation, and path planning of the proposed method are successful but motion control is inadequate. The error accumulation from the sensors can be addressed by continuously localising the robot's pose via the camera and not from the odometry and IMU sensor during navigation. The proposed method will add on to existing navigational methods, providing greater global path planning by continuously updating the map.

References

Anoopa, S., Salim, A., & Nadera Beevi, S. (2023, 19-21 April 2023). Comparison of Faster RCNN and YOLO V3 for Video Anomaly Localization. 2023 International Conference on Power, Instrumentation, Control and Computing (PICC). https://doi.org/10.1109/PICC57976.2023.10142815

Adkins, A., Chen, T., & Biswas, J. (2024). ObVi-SLAM: Long-Term Object-Visual SLAM. *IEEE Robotics and Automation Letters*, *9*(3), 2909–2916. https://doi.org/10.1109/LRA.2024.3363534

Ahmadi, B., Vanany, I., & Dewi, R. S. (2023, 18-21 Dec. 2023). Smart Automated Guided Vehicles and Autonomous Mobile Robots in Warehouse Operations: A Bibliometric Analysis. 2023 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM). https://doi.org/10.1109/IEEM58616.2023.10406697

Ahmed, A. J. O., Babiker, A., Elhag, A., & Drar, M. (2023, 28-30 Nov. 2023). Real-Time Agricultural Monitoring with Agrobot: A Raspberry Pi and YOLO Based Solution. 2023 International Conference on Computer and Applications (ICCA). https://doi.org/10.1109/ICCA59364.2023.10401368

Alatise, M. B., & Hancke, G. P. (2020). A Review on Challenges of Autonomous Mobile Robot and Sensor Fusion Methods. *IEEE Access*, *8*, 39830–39846. https://doi.org/10.1109/ACCESS.2020.2975643

Campos, C., Elvira, R., Rodríguez, J. J. G., Montiel, J. M. M., & Tardós, J. D. (2021). ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM. *IEEE Transactions on Robotics*, *37*(6), 1874–1890. https://doi.org/10.1109/TRO.2021.3075644

Chen, C. W., Lin, C. L., Hsu, J. J., Tseng, S. P., & Wang, J. F. (2021, 16-17 Dec. 2021). Design and Implementation of AMR Robot Based on RGBD, VSLAM and SLAM. 2021 9th International Conference on Orange Technology (ICOT). https://doi.org/10.1109/ICOT54518.2021.9680621

Chen, L. B., Huang, X. R., & Chen, W. H. (2023). Design and Implementation of an Artificial Intelligence of Things-Based Autonomous Mobile Robot System for Pitaya Harvesting. *IEEE Sensors Journal*, *23*(12), 13220–13235. https://doi.org/10.1109/JSEN.2023.3270844

Choi, J. H., Bae, S. H., An, Y. C., & Kuc, T. Y. (2023, 17-20 Oct. 2023). Development of an Advanced Navigation System for Autonomous Mobile Robots for Logistics Environments. 2023 23rd International Conference on Control, Automation and Systems (ICCAS). https://doi.org/10.23919/ICCAS59377.2023.10316962

Gunawan, R., Chandra, P. D., Kusmadi, Azwar, A. G., Nurwathi, & Risnanto, S. (2022, 13-14 Oct. 2022). Autonomous Vehicle Guided with RFID Position Detection for Warehouse Management System. 2022 16th International Conference on Telecommunication Systems, Services, and Applications (TSSA). https://doi.org/10.1109/TSSA56819.2022.10063925

Hess, W., Kohler, D., Rapp, H., & Andor, D. (2016, 16-21 May 2016). Real-time loop closure in 2D LIDAR SLAM. 2016 IEEE International Conference on Robotics and Automation (ICRA). https://doi.org/10.1109/ICRA.2016.7487258

Kaewpinjai, R., Chuaubon, T., & Apavatjrut, A. (2020, 24-27 June 2020). On Improving Indoor Navigation Accuracy Using Bluetooth Beacons. 2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON). https://doi.org/10.1109/ECTI-CON49241.2020.9157901

Kim, J.-a., Sung, J.-Y., & Park, S.-h. (2020, 1-3 Nov. 2020). Comparison of Faster-RCNN, YOLO, and SSD for Real-Time Vehicle Type Recognition. 2020 IEEE International Conference on Consumer Electronics — Asia (ICCE-Asia). https://doi.org/10.1109/ICCE-Asia49877.2020.9277040

Kok, S. Y., Yu, W. W., & Ng, D. W. K. (2023, 21-23 April 2023). A Study on the Application of ORBSLAM3 on a Mobile Differential Drive Robot. 2023 8th International Conference on Control and Robotics Engineering (ICCRE). https://doi.org/10.1109/ICCRE57112.2023.10155617

Krishna, N. M., Reddy, R. Y., Reddy, M. S. C., Madhav, K. P., & Sudham, G. (2021, 2-4 Sept. 2021). Object Detection and Tracking Using Yolo. 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA). https://doi.org/10.1109/ICIRCA51532.2021.9544598

Kusashio, K., Yakoh, T., & Kakinuma, Y. (2024, 25-27 March 2024). Task Planning Optimization in Assisting Multiple NC Machine Tools Using AMR. 2024 IEEE International Conference on Industrial Technology (ICIT). https://doi.org/10.1109/ICIT58233.2024.10540943

Leong, P. Y., & Ahmad, N. S. (2024). Exploring Autonomous Load-Carrying Mobile Robots in Indoor Settings: A Comprehensive Review. *IEEE Access*, *12*, 131395–131417. https://doi.org/10.1109/ACCESS.2024.3435689

Li, Z. X., Cui, G. H., Li, C. L., & Zhang, Z. S. (2021, 9-11 Oct. 2021). Comparative Study of Slam Algorithms for Mobile Robots in Complex Environment. 2021 6th International Conference on Control, Robotics and Cybernetics (CRC). https://doi.org/10.1109/CRC52766.2021.9620122

Mavilia, F., Barsocchi, P., Furfari, F., & Girolami, M. (2023, 30 Jan.-1 Feb. 2023). Evaluating the Impact of Anchors Deployment for an AoA-based Indoor Localization System. 2023 18th Wireless On-Demand Network Systems and Services Conference (WONS). https://doi.org/10.23919/WONS57325.2023.10061949

Merzlyakov, A., & Macenski, S. (2021, 27 Sept.-1 Oct. 2021). A Comparison of Modern General-Purpose Visual SLAM Approaches. 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). https://doi.org/10.1109/IROS51168.2021.9636615

Neto, A. P., & Tonidandel, F. (2022, 29-30 April 2022). Analysis of WiFi localization techniques for kidnapped robot problem. 2022 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC). https://doi.org/10.1109/ICARSC55462.2022.9784792

Niranjan, D., VinayKarthik, B., & Mohana. (2021, 15-17 Sept. 2021). Performance Analysis of SSD and Faster RCNN Multi-class Object Detection Model for Autonomous Driving Vehicle Research Using CARLA Simulator. 2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT). https://doi.org/10.1109/ICECCT52121.2021.9616712

Qiong, L., Xudong, C., Jizhuang, H., & Ruihao, M. (2020, 14-16 Aug. 2020). Research on Robot Path Planning Method Based on Tangent Intersection Method. 2020 International Conference on Information Science, Parallel and Distributed Systems (ISPDS). https://doi.org/10.1109/ISPDS51347.2020.00063

Roland, S., Illah Reza, N., & Davide, S. (2011). Mobile Robot Localization. In *Introduction to Autonomous Mobile Robots* (pp. 265–367). MIT Press. https://doi.org/10.1017/s0269888905210354

Shen, Y., Liu, J., & Luo, Y. (2021, 17-19 Dec. 2021). Review of Path Planning Algorithms for Unmanned Vehicles. 2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA). https://doi.org/10.1109/ICIBA52610.2021.9688064

Tan, L., Huangfu, T., Wu, L., & Chen, W. (2021). Comparison of yolo v3, faster r-cnn, and ssd for real-time pill identification. https://doi.org/10.21203/rs.3.rs-668895/v1

Wang, C., & Mao, J. (2019, 18-20 Oct. 2019). Summary of AGV Path Planning. 2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE). https://doi.org/10.1109/EITCE47263.2019.9094825

Yang, S., Geng, C., Li, M., Liu, J., & Cui, F. (2022, 28-30 Oct. 2022). Improved Cartographer Algorithm Based on Map-to-Map Loopback Detection. 2022 6th CAA International Conference on Vehicular Control and Intelligence (CVCI). https://doi.org/10.1109/CVCI56766.2022.9964502

Yi, C., & Hongtu, X. (2019, 3-5 June 2019). Global Dynamic Path Planning Based on Fusion of Improved A* Algorithm and Morphin Algorithm. 2019 Chinese Control And Decision Conference (CCDC). https://doi.org/10.1109/CCDC.2019.8833360

Zaman, U. K. U., Aqeel, A. B., Naveed, K., Asad, U., Nawaz, H., & Gufran, M. (2021, 26-27 Oct. 2021). Development of Automated Guided Vehicle for Warehouse Automation of a Textile Factory. 2021 International Conference on Robotics and Automation in Industry (ICRAI). https://doi.org/10.1109/ICRAI54018.2021.9651360

Zhang, S., Liang, W., Yuan, X., An, H., Zheng, M., & Zhang, Y. (2023, 20-22 May 2023). A Robust Surveillance System for AGV Using Forward-Downward Visual Measurements.

2023 35th Chinese Control and Decision Conference (CCDC). https://doi.org/10.1109/CCDC58219.2023.10326630

Zhao, X., & Chidambareswaran, T. (2023, 26-30 Aug. 2023). Autonomous Mobile Robots in Manufacturing Operations. 2023 IEEE 19th International Conference on Automation Science and Engineering (CASE). https://doi.org/10.1109/CASE56687.2023.10260631

Zheng, Z., Xie, M., Jiang, C., Bian, H., & Wang, W. (2024, 23-23 Aug. 2024). Indoor Mobile Robot Map Construction Based on Improved Cartographer Algorithm. 2024 WRC Symposium on Advanced Robotics and Automation (WRC SARA). https://doi.org/10.1109/WRCSARA64167.2024.10685767

Review and Enhancement of VoIP Security: Identifying Vulnerabilities and Proposing Integrated Solutions

Athirah Mohd Ramly
School of Architecture, Computing and Engineering, University of East London, United Kingdom

Zi Wei Ng
Department of Computing and Information Systems, Sunway University, Malaysia

Yahya Khamayseh
Department of Computing and Information Systems, Sunway University, Malaysia

Charis Shwu Chen Kwan
Department of Computing and Information Systems, Sunway University, Malaysia

Angela Amphawan
Smart Photonics Research Laboratory, School of Engineering and Technology, Sunway University, Malaysia

Tse-Kian Neo
CAMELOT, Faculty of Creative Multimedia, Multimedia University, Malaysia

Abstract: The prolific use of Voice over Internet Protocol (VoIP) causes this system's vulnerabilities to become a major concern for individuals or organizations. Based on this concern, this paper aims to provide a comprehensive analysis of the cyber threats and attacks in VoIP systems. VoIP offers several benefits and advantages; however, it poses significant security risks which can cause serious consequences. At the beginning of this paper, an overview of VoIP, including its architecture, components, protocols and advantages are discussed. Following this, attack vectors and component vulnerabilities are examined, and types of security threats are categorized into different sections for clarity. The types of attacks under discussion involved Denial-of-Service (DoS) and Man-in-the-Middle (MITM) attacks, spoofing threats, registration and call hijacking, Spam over Internet Telephony (SPIT), vishing, malware and toll fraud. Subsequently, existing security solutions for VoIP systems are reviewed, highlighting their strengths, weaknesses and applicability. While some solutions are effective in protecting VoIP, a lack of robustness still exists. Hence, this paper proposes an enhanced security method: the combination of Virtual Private Networks (VPNs) and firewalls within VoIP

https://doi.org/10.18080/jtde.v12n4.1022

systems. This combination can reduce potential cyber risk by enhancing data privacy and protection, and overall system security.

Keywords: VoIP, DoS, VPN, SPIT, Malware.

Introduction

The rapid development of Voice over Internet Protocol (VoIP), a digital communication system, has influenced the communication of individuals and enterprises. According to Future Market Insights Inc. (2023), a Compound Annual Growth Rate of 9.8% and a determined value of US$118.86 billion is assumed to be achieved in the VoIP market by 2033, based on the valuation in 2023 of US$43.92 billion. Traditional telephony systems offer limited flexibility and scalability as compared to VoIP due to their operation on dedicated physical circuits for voice transmissions (Luhach *et al.*, 2019). This causes VoIP, which operates by converting voice signals into digital packets and transmitting over IP networks, to have more effective transmission and consequently lead the communication trend (Suthar & Rughani, 2020; Chakraborty *et al.*, 2019a; Jimoh & Al-Juboori, 2024).

Furthermore, the capability to transmit voice and multimedia content over IP networks has made it an integral portion of various sectors comprising business, healthcare, education and even government services (Kumar & Roy, 2021; Chakraborty & Telgote, 2019; Oproiu *et al.*, 2020; Munusamy & Khodadi, 2023; Tay, Ooi, Pang, Gan & Lew, 2023).

However, cyber threats and attacks in VoIP have become major safety risks due to their deployment on IP data networks (Jingi, 2017). These threats encompass but are not restricted to denial-of-service (DoS), eavesdropping, identity spoofing, malware and toll fraud. They are disruptive as they have the potential to interrupt communication services, and breach data confidentiality, integrity and availability (Kumar & Roy, 2021; Abdulazeez *et al.*, 2020; Rathore *et al.*, 2021; Mrewa, Ramly, Amphawan, & Neo, 2024).

To address the issues effectively, this paper begins with an overview of VoIP, including its protocols, component design and benefits, to help understand the weaknesses that attackers take advantage of. The next section of the study focuses on classifying different kinds of cyberattacks that specifically target VoIP systems and highlights how architectural deficiencies in VoIP pose serious threats to users' confidentiality and security.

Furthermore, the research paper suggests an enhanced security approach that combines VPNs and firewalls with VoIP systems. This integration is aimed to improve data security, privacy and system safety in general by utilizing strong encryption for data transfer and implementing strict access control measures.

https://doi.org/10.18080/jtde.v12n4.1022

The wide adoption of VoIP systems has exposed significant cyber threats. According to IBM's *2024 Cost of a Data Breach Report*, the global average cost of a data breach has reached an all-time high of US$4.88 million; a 10% increase from the previous year (IBM, 2024). Notably, 40% of these breaches involved data stored across multiple environments, including public clouds, which incurred the highest average breach cost at US$5.17 million (IBM, 2024). This underscores the growing risk to VoIP systems, as they often operate across diverse environments, making them susceptible to breaches that can compromise confidentiality, integrity and availability.

According to Liu (2024), fraud losses from Internet Protocol Private Branch eXchange (IP-PBX) hacking totalled $1.82 billion; a 28% increase over 2019. This statistic underscores the urgent need for robust and comprehensive security solutions that can effectively mitigate this risk. Nevertheless, current mitigation techniques fail to effectively protect the VoIP system which leads to the pressing need for enhanced solutions in VoIP security.

This research aims to address these issues by providing an in-depth analysis of cyber threats and attacks in VoIP, evaluating the shortcomings of current security measures, and proposing new strategies for enhancing VoIP security. The goal is to develop a robust, adaptable and scalable security solution that can safeguard VoIP systems against evolving cyber threats and ensure reliable and secure communication.

The objectives of this research are to:

1. Understand VoIP: Develop a comprehensive understanding of VoIP systems including its protocols, components and advantages. Determine the potential attack vectors of VoIP systems and identify the types of cyber threats that could exploit these systems.
2. Evaluate existing security measures: Assess the effectiveness and limitations of current security measures implemented in VoIP systems.
3. Propose enhanced security strategies: Develop and propose an enhanced security countermeasure to mitigate the identified vulnerabilities in VoIP systems. This involves the integration of technologies such as VPNs and firewalls.

This project aims to improve the understanding of VoIP systems and focus on their security aspects to strengthen their defences. Its primary focus explores VoIP's underlying structure, functionalities and related protocols such as Session Initiation Protocol (SIP), Media Gateway Control Protocol (MGCP) and Real-time Transport Protocol (RTP).

Furthermore, a significant part of this research analyses various cyber threats and attacks on VoIP systems. This includes but is not limited to DoS, distributed DoS, Man-in-the-Middle (MITM) attacks, spoofing threats, registration and call hijacking, eavesdropping, vishing, Spam over Internet Telephony (SPIT), malware, viruses, and toll fraud.

https://doi.org/10.18080/jtde.v12n4.1022

In terms of countermeasures, this study assesses existing security solutions implemented in VoIP systems. This includes specifically addressing DoS, DDoS and SPIT as they are among the popular ongoing threats in the field of VoIP. While the investigation also consists of other solutions, it is important to note that it may not cover all of the potential security solutions available for VoIP systems.

The research then moves forward to propose a methodology for enhancing VoIP security. This methodology involves the utilization of Asterisk, a VPN, a firewall and other tools. In addition, the project will cover security parameters used in the VPN and firewall, followed by the design of the architecture and the setting up of the experiment. Security testing procedures are established and a comparative analysis is conducted.

The research does not delve into the non-security aspects of VoIP technology such as cost-effectiveness, communication efficiency and other operational benefits. Additionally, although the research covers a broad range of cyber threats, it may not exhaustively cover all possible threats due to the evolving nature of cyber threats.

Literature Review

VoIP is a modern digital technology that allows individuals to make phone calls using an internet connection rather than a conventional analogue telephone line (Sadiwala, 2018). The methods and principles required in initiating VoIP calls are like those involved in traditional digital telephony, including signalling, channel setup, digitization of analogue speech signals and encoding. Despite this, instead of being sent across a circuit-switched network, digital information is packetized and transmitted as IP packets via a packet-switched network (Fasiku, 2018). VoIP systems use signalling and transporting protocols to control call signalling, setup and termination. For example, various well-known network applications that make use of VoIP technology such as Skype and Zoom handle millions of active accounts every day for video conferencing and two-party as well as multi-party calls.

VoIP communication can be classified into four types (Kumar & Roy, 2021):

1. Computer-to-computer
2. Computer-to-phone
3. Phone-to-computer
4. Phone-to-phone.

Indeed, VoIP is becoming more popular because of its low cost, enhanced network administration and ability to integrate data and voice traffic over existing networking infrastructure (Kumar & Roy, 2021). By implementing VoIP, large corporations have saved millions of dollars by replacing old telephone systems for long-distance calls with IP network-

https://doi.org/10.18080/jtde.v12n4.1022

based alternatives. In addition, network consolidation in VoIP allows the transmission of data, phone, and video information across a single network, reducing initial setup and maintenance expenses (Mentsiev & Supaeva, 2019).

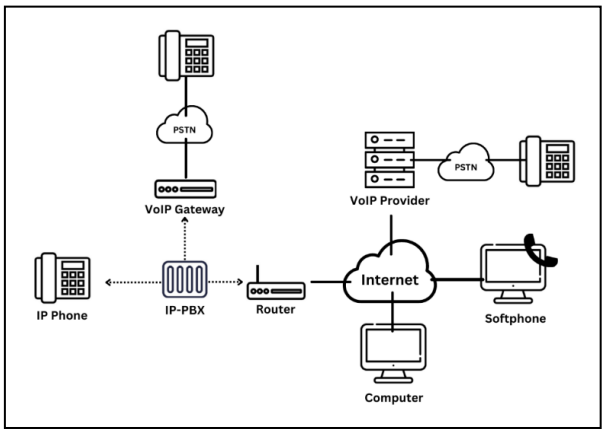


Figure 1. Basic architecture of VoIP system

Based on Figure 1, VoIP architecture comprises a set of interconnected components to facilitate seamless voice data transmission. These components include endpoints, codecs, signalling protocols, VoIP gateways and IP networks. Each plays a vital role in ensuring the efficient and reliable delivery of voice packets from the sender to the receiver.

This system is adaptable and can accommodate a wide range of devices, from traditional phones linked via gateways to current softphones on computers or mobile devices. Moreover, its core is an IP-based exchange, which is identical to traditional telephone networks but has been converted for digital VoIP communication.

The following sections will cover each component of the VoIP system in detail, describing how they work together.

VoIP components

Following are the descriptions for each key element in the system.

1. End-user equipment: Refers to the device used to access and connect with other endpoints through the VoIP system. It is responsible for initiating and maintaining the necessary signalling processes during the communication establishment, as well as enabling users to connect, make calls and communicate (Nazih *et al.*, 2019). IP phones, softphones (software-based phones), hard phones, computers, tablets and other mobile devices are examples of end-user equipment in VoIP systems. Additionally, every device is assigned a unique IP address to exchange information within the VoIP system.
2. IP-PBX: Is also known as a soft PBX and is an advanced version of traditional PBX. It utilizes internet protocol to handle signalling, data and voice traffic (Nazih *et al.*, 2019).

https://doi.org/10.18080/jtde.v12n4.1022

It is the central component that manages call routing, call control and other telephony features within an organization. The IP-PBX system consists of a VoIP gateway, an IP-PBX server and VoIP phones, while the IP-PBX server acts as a proxy server. Moreover, SIP clients such as hardware or softphones, register with the IP-PBX server and request the IP-PBX when establishing a connection. The IP-PBX also stores a directory of all users and their matching SIP addresses to allow it to connect a call internally or to route a third-party call to the desired destination via a VoIP gateway or a VoIP service provider (Almutairi, 2018).

3. Codecs: Also known as coders and decoders, play a crucial role in packet-switched networks for transferring audio and video data (Kumar & Roy, 2021). Analog voice signals are transformed into digital signals, compressed and finally encoded into a specified format such as G.711, G.729, G.723.1a, etc (Sadiwala, 2018).
4. VoIP gateway: The gateway functions as an interface between the VoIP network and traditional telephone networks (PSTNs), receiving incoming calls from the caller and establishing a connection with the recipient (Mentsiev & Dzhangarov, 2019). It converts voice signals between packet-switched IP format and circuit-switched format that can be transmitted over an IP network and telephone network (Mentsiev & Dzhangarov, 2019). In short, it enables communication between VoIP phones and PSTN-based telephones (Chakraborty *et al.*, 2019a).
5. IP network: Enables connectivity among all terminals, regardless of whether they are private, part of an intranet or connected to the internet.

VoIP signalling and transport protocols

VoIP systems have two main functions: signalling and media transmission function (Nazih *et al.*, 2019). These functions work together to enable optimal and consistent communication over IP networks.

Signalling functions in a VoIP system involve the exchange of control information between endpoints to establish, modify and terminate a communication session (Chakraborty et al., 2019b). Several protocols are used in signalling, including H.323, SIP and MGCP. H.323 was the first protocol introduced in VoIP; however, the emergence of SIP caused the reduced demand for H.323. As a result, Cisco has announced the discontinuation of H.323 call control functionality in Cisco IOS XE software. Alternatively, SIP is a more recent and lightweight signalling protocol that manages multimedia conference signalling and control (Chakraborty *et al.*, 2019a). Therefore, this paper will only focus on the use of SIP in VoIP systems. Moreover, the MGCP protocol is used to exchange control information between VoIP network components.

https://doi.org/10.18080/jtde.v12n4.1022

Once the signalling phase is completed and the call is established, the media transmission function takes over. This function is responsible for sending the actual voice or multimedia data between the devices involved in the call (Abualhaj *et al.*, 2020). The media transmission relies on protocols such as RTP, which handles the packetization, sequencing and timestamping of the media data (Abualhaj *et al.*, 2020). It collaborates with real-time protocols to transport data across the networks.

How VoIP works

VoIP is a system that uses an IP network to send analogue voice signals. It is performed through encoding, compressing, packetizing and other operations. At the receiving end, the reverse techniques are used to reassemble the voice data once it has been sent to the destination over the network to allow it to be received. According to Figures 2 and 3, there are a few steps involved in the process.

First, the process begins with the transformation of analogue data into digital data, which is known as the digitization of the analogue voice signal (Sadiwala, 2018). The input analogue signal is sampled at least 8,000 times per second before converting into the digital form (Chakraborty *et al.*, 2019b). The digital data is a bit stream typically represented by discrete values such as 0s and 1s. This process is carried out using the analogue-to-digital (A/D) converter shown in Figure 2. It is important to note that the coding algorithm used throughout the digitalization process must be the same to ensure the digital data can be converted back to the responding analogue data (Sadiwala, 2018).

After the analogue voice signal is digitalized into a bit stream, the compression process will compress the digitized data into smaller sizes by using a complex algorithm (Sadiwala, 2018). Although the compression algorithm will reduce the size of the data, it guarantees the quality of the signal.

Next, the digitized data will be grouped into data packets that are suitable for transmission. For example, if the encoder handles 10 ms of data at a time, the first 40 ms of digitized data will be divided into four different chunks and coded in sequence. Each chunk of the data will then be encapsulated into an IP packet with a header that consists of the necessary information for routing. This process is known as packetization (Sadiwala, 2018); following which the IP packets will be transmitted to the entire network for transmission.

During the transmission, the IP packets are carried by IP to the entire network. Every intermediate node in the network examines the header information in the IP data such as source and destination address and then forwards it to the next node according to the information provided (Sadiwala, 2018).

https://doi.org/10.18080/jtde.v12n4.1022

When the packets finally reach the receiving end, the IP packets undergo decapsulation. The compressed digitized data is extracted from the IP packets and placed in order based on information in the packet headers. During the decapsulation process, the address and other control data are removed as they are not necessarily needed during this process (Sadiwala, 2018).

Then the compressed digitized data is decompressed back into the digitized data, which will then be decoded into an analogue voice signal that is audible using the digital-to-analogue (D/A) converter.

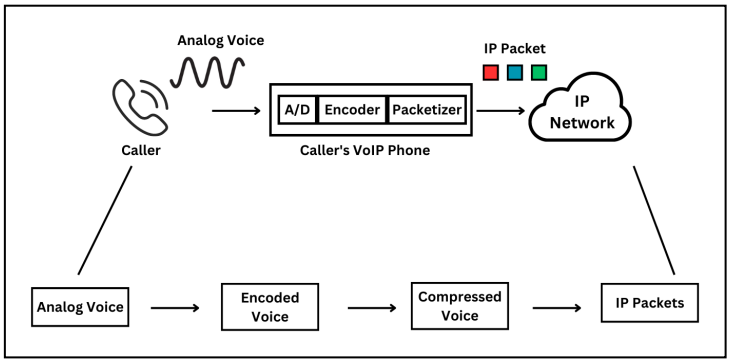


Figure 2. Working mechanism of a VoIP communication system

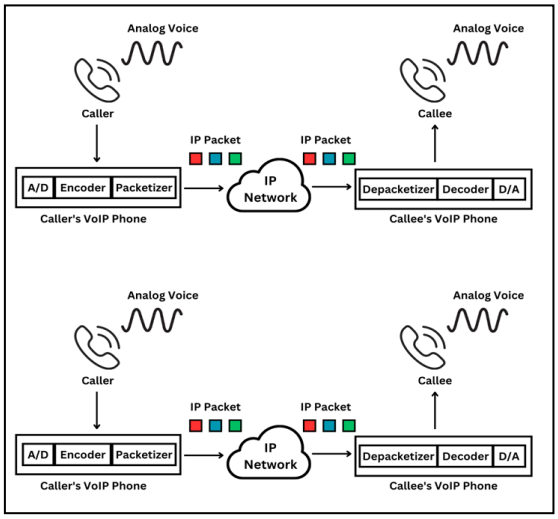


Figure 3. VoIP communication between the caller and the callee

Performance metrics

The Quality of Service provided to the end user is of upmost importance and one of the main issues for the implementation of a VoIP system: if the conversation is unintelligible then there is no point in providing the service. The main factors that affect the quality of service are latency, jitter, packet loss and Mean Opinion Score (MOS) (Fasiku, 2018).

Latency is referring to the time gap between when someone speaks and when the listener hears it. It is crucial for good voice quality in VoIP. A significant issue it can create is speech overlap. In a normal conversation, when one person finishes speaking, they wait for the other person

https://doi.org/10.18080/jtde.v12n4.1022

to respond. If the response takes too long, the speaker might start talking again and this can lead to their speech colliding with the delayed response. The International Telecommunication Union suggests that, for optimal quality, the total delay from source to destination should be around 150 ms (IBM, 2024). Delays between 150 and 300 ms are still acceptable. However, if the delay exceeds 300 ms, efforts should be made to reduce it.

Meanwhile, jitter refers to the packets of data arriving at different times because of network congestion and poor queuing. This happens when data is split into pieces (packets) and sent from one place to another through various routes, causing the arrival times to vary due to changes in network traffic. For optimal voice quality, jitter should be between 20 and 50 ms. The voice quality will be poor if the jitter value is higher than 50 ms.

Packet loss occurs when the network experiences congestion due to excessive traffic or overwhelmed bandwidth, as well as when network quality is weak, such as unstable network components or insufficiently powered equipment. These causes are applicable to both voice and data packets. However, voice packets face an additional challenge: they can be discarded if they arrive at their destination too late to be of any use, exceeding the time limit set by the jitter buffer. Packet loss does not pose a problem for data packets that use the Transmission Control Protocol (TCP). TCP ensures reliability by identifying and retransmitting lost packets. Therefore, VoIP cannot apply on TCP's retransmission mechanism due to reliance on real-time packet transmission. This is because, for a VoIP application, a delayed packet is equivalent to a lost one, and TCP introduces unacceptable delays for voice packets.

MOS is a measurement of voice quality for assessing how people perceive call quality. In VoIP networks, it is extensively employed to guarantee clear voice communication, identify quality problems, and offer a standard for evaluating voice performance and deterioration. As VoIP services become more popular, MOS scores play a crucial role in ensuring customer satisfaction and supporting the expansion of these networks.

VoIP advantages

VoIP is an emerging technology that aims to improve traditional voice communication through also utilising digital communication. VoIP enables the simultaneous transmission of data, pictures and videos, and provides several benefits including:

1. Cost efficiency: VoIP reduces communication costs for users by enabling communication over private or internet data network lines, rather than relying on expensive commercial telecommunications lines (Sadiwala, 2018).
2. Mobility: Users can enjoy consistent services such as call features, call logs, voicemail, service policies, security features and more wherever they go (Fasiku, 2018).

https://doi.org/10.18080/jtde.v12n4.1022

3. Integration with other services: Individual users can integrate VoIP services with their personal email, calendar or other applications, enhancing their productivity and efficiency (Fasiku, 2018).
4. Scalability: Additional phone lines and users can conveniently be added online by purchasing a new phone number, without the need for complex setup or wiring (Drew, 2023). For instance, add-on features such as additional communication channels, extra storage, professional voicemail greetings and more can be easily obtained by upgrading to a higher-tiered plan.
5. Easy to use: Users do not have to deal with complicated and bulky equipment. Only VoIP configurations and software are required if there is an internet connection (Suthar & Rughani, 2020).

VoIP disadvantages

While VoIP technology provides numerous advantages, it also introduces security concerns. The increasing popularity of VoIP services has made them targets for cybercriminals, emphasizing the need for strong security measures. Adopting VoIP involves not only embracing its benefits but also addressing significant security challenges to mitigate technological risks. This section primarily focuses on reviewing existing solutions related to VPNs and firewalls to set the foundation for the methodology proposed in the subsequent section.

As an early attempt to integrate VPN technology within VoIP systems, Surasak & Huang (2019) developed an Android application that combines VoIP with VPN technology, leading to an increase in the overall security of VoIP communications by generating anonymous packets. However, the system's adaptability to other platforms remained untested, which indicates scope for further research in this area. It may also not fully represent realistic network conditions as their performance analysis was conducted in an environment devoid of an internet connection. This limitation raises questions about the real-world applicability of their findings as an Internet connection is essential for the functionality and security of a network in everyday situations. This method offers a deeper understanding of the use of VPN but also reveals gaps that necessitate a better solution. Despite being limited to an offline environment, their work is significant for its early exploration of mobile VoIP VPN applications, which paved the way for further research on network adaptability.

Abdulazeez *et al.* (2020) conducted an in-depth comparison of popular VPN protocols, including WireGuard, Internet Protocol security (IPsec) and generic routing encapsulating (GRE), highlighting the crucial role of VPN protocols in securing network communication. Their study revealed that WireGuard stands out due to its strengths in various characteristics.

https://doi.org/10.18080/jtde.v12n4.1022

These findings align with Pudelko *et al.*'s (2020) performance analysis of VPN gateways, which emphasized the limitations of IPsec and OpenVPN when compared to newer solutions like WireGuard. They suggested the future use of WireGuard for its improved performance and potential, supporting the notion that WireGuard is a secure and promising VPN solution. Pudelko *et al.*'s (2020) serve as a foundation for the choice of WireGuard in the methodology. Recent research by Jumakhan & Mirzaeinia (2024) shows that WireGuard is a lightweight and efficient VPN protocol, outperforming older options like OpenVPN and IPsec in speed and security. Its low latency and strong encryption make it ideal for modern VoIP systems that need fast, secure communication without high resource demands.

While these studies underscore the significance of VPN protocols like WireGuard, He (2021) shifted focus to the role of firewall technology as a vital defense mechanism in computer networks. He concluded that firewall technology effectively prevents data breaches and reduces cyberattack risks, hence giving valuable insights about using a firewall to achieve secure VoIP communication. However, it is important to note that while He (2021)discusses the theoretical importance of firewall technology in securing VoIP communication, it does not provide practical implementation or empirical results to validate these theoretical claims. Although theoretical, his study offers a crucial framework for firewall application in VoIP, highlighting its importance in conceptual security models despite the lack of practical testing.

To address this gap between theory and application, Tuleun (2024) conducted an in-depth analysis of VoIP security in an Asterisk-based system, focusing on integrating IPsec VPN with firewall technologies to protect SIP-based communication. The study demonstrated that combining VPNs with firewall configurations effectively secures VoIP traffic, blocking unauthorized access and ensuring data integrity through packet encapsulation and encryption. However, Tuleun's approach lacked advanced firewall features such as stateful inspection and rate limiting, which are essential for handling VoIP-specific traffic patterns and mitigating threats like DoS attacks. This gap highlights the need for enhanced security measures, including port filtering, stateful inspection and rate limiting. Building on Tuleun's foundation, Arpaci & Şentürk (2024) further explored the performance impact of VPN and firewall configurations on real-time applications, offering insights into balancing security and communication quality in VoIP.

In a recent study, Arpaci & Şentürk (2024) examined the performance impact of firewall and VPN usage in real-time applications, specifically focusing on video conferencing as a proxy for VoIP environments. Their research highlights how firewalls and VPNs contribute to security in high-traffic, real-time communications, with a particular focus on performance metrics critical to VoIP quality, such as end-to-end delay, packet delay variation (jitter), and packet loss. Using the OPNET simulation tool, the study tested various configurations, including

https://doi.org/10.18080/jtde.v12n4.1022

firewall-only, VPN-only, and combined setups. The findings revealed that, while firewalls and VPNs both add latency due to packet inspection and encryption processes, the delay remained within acceptable ranges for quality communication. Notably, VPN usage was found to increase end-to-end delay slightly more than firewall-only configurations, though both provided effective security without significant quality degradation. Arpaci & Şentürk (2024) introduced empirical data underscores the balance between security and performance in VoIP-related applications, offering insights into optimal configurations that minimize performance impacts. These findings support the methodology of combining VPNs with firewall-based security in VoIP to ensure robust security without compromising real-time communication quality.

Table 1. Functional comparison of existing VPN and firewall solutions for VoIP security

Author	Year	VPN protocol	Firewall integration	Security features	Limitations
Surasak & Huang	2019	Not specified	Not integrated	Generates anonymous packets to enhance security	Limited to Android platform; no Internet connection in testing, which raises concerns about real-world applicability
Abdulazeez *et al.*	2020	WireGuard, IPsec, GRE	Not discussed	Highlights WireGuard's speed and security advantages in securing network communication	Comparative study without a VoIP focus, lacks application-specific firewall integration
Pudelko *et al.*	2020	WireGuard, IPsec, OpenVPN	Not discussed	Analyzes performance, noting WireGuard's strengths over IPsec and OpenVPN	Limited to VPN gateway analysis, without specific firewall or VoIP integration
Jumakhan & Mirzaeinia	2024	WireGuard	Not discussed	Demonstrates WireGuard's low latency and strong encryption, suited for VoIP	Focuses on VPN protocol performance, lacking details on firewall integration or VoIP-specific requirements
He	2021	Not specified	Theoretical framework	Describes firewall importance in preventing data breaches and cyberattack risks	Conceptual; lacks empirical testing and practical firewall implementation details for VoIP security
Tuleun	2024	IPsec	Basic firewall in Asterisk system	Enhances security with packet encapsulation and encryption for SIP communication	Lacks advanced firewall features (e.g. stateful inspection, rate limiting), which are essential for VoIP-specific threat management
Arpaci & Şentürk	2024	Various (not specified)	Firewall and VPN tested separately and combined	Balances security with acceptable performance for real-time VoIP applications	Shows minor latency with VPN, but lacks discussion of advanced firewall mechanisms such as stateful inspection

In recent years, significant research has been conducted on securing VoIP systems, focusing

https://doi.org/10.18080/jtde.v12n4.1022

on improving both security and performance. Arpaci & Şentürk (2024) and Tuleun (2024) provided insights into the latest VPN and firewall integration techniques for VoIP, highlighting minimal impact on latency and packet loss. While recent research is central, certain earlier works are included for foundational concepts in VoIP traffic analysis and protocol handling, which continue to inform current practices.

To conclude, the existing protective measures appear inadequate to provide robustness. Therefore, an enhanced technique that combines VPN and firewall technologies appears to be an appropriate solution with which to address the problem. With WireGuard's performance as a VPN solution for VoIP security and the integration of adaptive firewall configurations, this combined approach meets the evolving needs of secure VoIP communication.

Methodology

In this study, the proposal is to combine VPNs and firewalls to enhance VoIP system security. Existing solutions lack a complete integration of advanced VPN technology and advanced firewall mechanisms, resulting in gaps in security and system efficiency for VoIP communication. The objective is to imitate specific attacks on the SIP server and associated SIP phones to assess the effectiveness of current security measures. The study specifically examines SIP scanning, eavesdropping, DoS and brute-force attacks. Performance data is collected and analysed in three scenarios: one without any security measures, another with a VPN in place, and a third with both a VPN and an active firewall. As shown in Figure 4, it illustrates the proposed security architecture in detail, which includes three core functions.

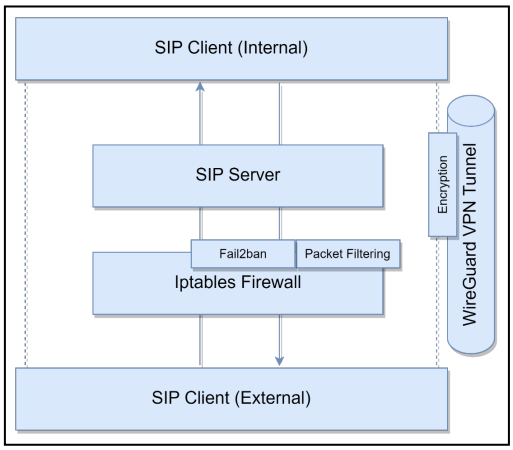


Figure 4. Proposed architecture

Encryption

WireGuard is utilized to offer encryption, ensuring the confidentiality of user agents when initiating or receiving VoIP calls. WireGuard is a new, open-source VPN application that uses advanced cryptography. It is designed to be faster and easier to use than earlier VPN solutions

https://doi.org/10.18080/jtde.v12n4.1022

like IPsec and OpenVPN, and has its own unique protocol. Traditional VPN protocols, such as IPsec, often rely on complex and resource-intensive encryption algorithms like AES (Advanced Encryption Standard) or 3DES (Triple Data Encryption Standard). These older algorithms, while secure, can be computationally demanding, especially in real-time communication scenarios like VoIP. WireGuard, in contrast, focuses on simplicity and efficiency in its design. Its authentication mechanism relies on straightforward public and private keys for authentication. Similar to Secure Shell (SSH), WireGuard's authentication process involves the exchange of public and private keys between the server and client before a connection is established. Once the connection is established using public keys, WireGuard takes care of key exchanges, connections, reconnects and disconnects automatically, streamlining the process for users.

WireGuard takes a different approach compared to widely used ciphers and protocols like AES, 3DES, IPsec, and various SSL VPNs. Instead, WireGuard adopts newer, secure alternatives that have undergone rigorous analysis. It employs Noise for the initial key exchange, computes encryption keys using the Elliptic Curve Cryptography (ECC) with Curve25519 function, and utilizes ChaCha20 with Poly1305 for authenticated symmetric encryption. Additionally, WireGuard uses BLAKE2 for hashing.

In this implementation, WireGuard VPN is chosen over earlier protocols like IPsec and OpenVPN due to its lower latency and simpler configuration, which is particularly beneficial for real-time VoIP traffic. While older protocols provide robust security, WireGuard's modern encryption algorithms allow for faster handshake times and reduced packet overhead. This selection enables high-security levels without compromising performance; a critical aspect given VoIP's sensitivity to latency and jitter.

The streamlined design of WireGuard focuses on minimalism, reducing the number of steps and avoiding extra features that can slow down VPN connections. This design choice makes WireGuard faster because it needs fewer resources to encrypt and decrypt data, allowing for high-speed, low-latency connections. By using efficient algorithms like ChaCha20 and Curve25519, WireGuard provides strong security without compromising performance, even on devices with limited processing power. WireGuard breaks the usual trade-off between security and performance by using modern cryptographic methods that are both efficient and secure. This combination allows WireGuard to improve security while still delivering high performance, making it a great option for applications like VoIP that need low latency.

The operational flow of WireGuard VPN depicted in Figure 5 is a representation of how encrypted traffic is routed between two SIP clients (7002 and 7001) over a VPN, with a router functioning as the VPN server in the middle. Following is the step-by-step process:

https://doi.org/10.18080/jtde.v12n4.1022

1. Step 1: SIP Client 7002 initiates a call to SIP Client 7001: The SIP Client 7002 wants to initiate a call to SIP Client 7001. Both clients are on a VPN network, which allows them to communicate securely over possibly insecure networks like the internet.
2. Step 2 – Encryption on SIP Client 7002: SIP Client 7002 uses its WireGuard interface (wg0) to encrypt the packet. The packet is a UDP packet addressed to SIP Client 7001's endpoint, which includes the VPN IP address of SIP Client 7001 (192.168.133.6).
3. Step 3 – Packet transmission to the router (VPN server): The encrypted packet is sent out via SIP Client 7002's network interface enp0s3.
4. Step 4 – Router (VPN server) processing: The router, functioning as a VPN server, receives the packet on its enp0s3 interface. It uses its own wg0 interface to decrypt the packet to determine the destination, which is SIP Client 7001. The router's VPN IP is 192.168.133.10. After decrypting the packet and reading the destination address, the router forwards the packet to SIP Client 7001 via its enps03 interface (not shown in the diagram but indicated by the directional flow).
5. Step 5 – Packet reception by SIP Client 7001: SIP Client 7001 receives the packet on its enp0s3 interface. The wg0 interface on SIP Client 7001 then decrypts the packet so that the SIP client can process the call initiation from SIP Client 7002.
6. Step 6 – Call connection: Once decrypted, SIP Client 7001 has the necessary information to establish a call with SIP Client 7002, and the SIP protocol takes care of setting up and managing the call session.

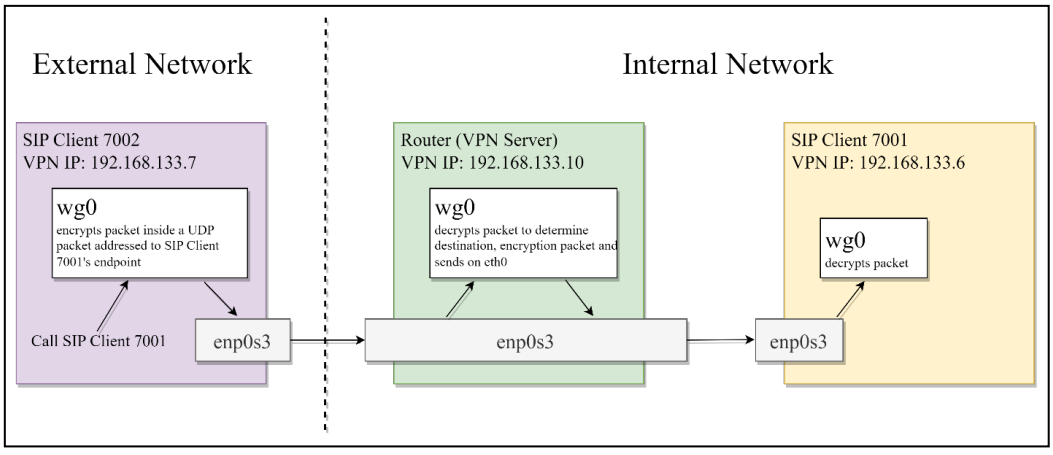


Figure 5. Operational flow of WireGuard VPN when 7001 calls 7002

This flow ensures that the SIP traffic between the two clients is encrypted across the network, protecting the communication from eavesdropping or tampering. WireGuard's role in this process is critical, as it provides the encryption and decryption mechanism on both ends of the communication and on the VPN server.

https://doi.org/10.18080/jtde.v12n4.1022

Packet filtering

The proposed architecture focuses on safeguarding the network perimeter from external threats and attacks. This security layer is achieved through the implementation of an iptables-based firewall responsible for filtering network traffic and permitting only authorized traffic according to predefined rules. The firewall serves as the initial line of defence against potential attacks, positioned between the private network and the public internet. It acts as a crucial chokepoint for network traffic, ensuring that all data entering or exiting the LAN must pass through it when appropriately configured. Various types of firewalls exist, including packet filter firewalls, proxy firewalls and stateful event monitors. In this methodology, a packet filter firewall is employed. A packet filter firewall, categorized as a network-level firewall, operates as a router device that not only forwards packets between networks but also assesses and determines whether to allow or deny the flow of these packets based on established filtering rules.

Since VoIP relies on specific signalling protocols like SIP and H.323, each with unique message structures, packet filtering must account for these differences. For instance, SIP uses SDP (Session Description Protocol) within its messages to negotiate media parameters like RTP and RTCP (Real-time Transport Control Protocol) stream details, which are essential for establishing VoIP calls. Firewall rules can be configured to monitor specific ports and IPs related to these RTP/RTCP streams, filtering out unauthorized or unexpected packets, which mitigates risks such as spoofing or MITM attacks. Additionally, H.323's reliance on channels like RAS (Registration, Admission, and Status) and H.245 for call setup and media control require separate rules. These rules ensure only authorized H.323 traffic is allowed, and they can limit exposure by only permitting traffic from known, trusted sources on specific ports. This protocol-specific filtering significantly reduces the likelihood of unauthorized access and improves the security of VoIP communication by handling each signalling protocol's unique requirements.

In this architecture, the firewall will implement several firewall rules such as rate limiting, stateful packet inspection, whitelisting, customize rules and logging to mitigate proposed attack effectively, proving that the adoption of a firewall can effectively protect the VoIP system from a wide range of attack types.

The operational flow for a packet filter firewall, as depicted in Figure 6, begins when a network packet arrives and triggers the firewall's processing sequence. Initially, the firewall reads the header of the packet, which contains vital information such as source and destination IP addresses, port numbers and protocol type. With this information, the firewall proceeds to

https://doi.org/10.18080/jtde.v12n4.1022

compare the packet against its set of predefined filtering rules. It checks for a match between the packet's header information and the criteria defined in the rules.

If the packet's header matches a rule, the firewall then logs the details of the packet for auditing and monitoring purposes. Following the logging, the firewall must decide whether to accept or drop the packet based on the action specified in the matching rule. If the rule requires the packet to be dropped, it is discarded and no further action is taken. Conversely, if the rule allows the packet, it is accepted and forwarded to its destination.

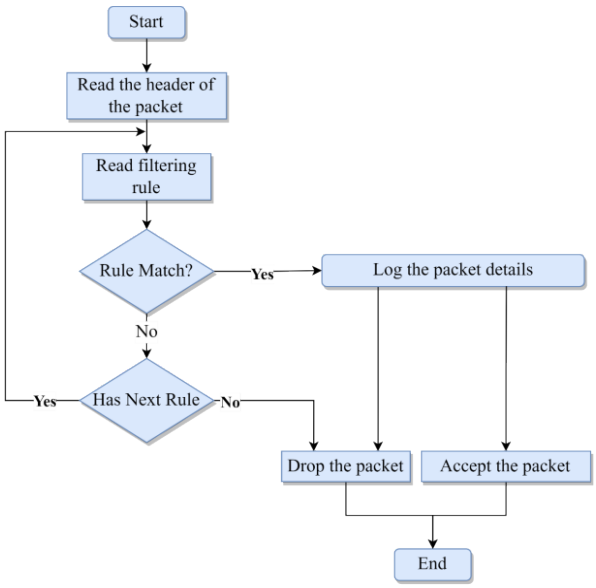


Figure 6. Operational flow of packet filter firewall

In cases where the packet does not match the current rule, the firewall looks to see if there are additional rules to consider. If more rules are available, the firewall reads the next rule and re-evaluates the packet against this new set of criteria. This loop continues until a matching rule is found. If no matching rule is found after all rules have been evaluated, the firewall defaults to dropping the packet to ensure the security of the network. This process marks the end of the firewall's decision-making for the individual packet.

Fail2Ban

Fail2Ban analyzes log files to identify potential malicious activities, such as repeated unsuccessful authentication attempts. Typically, this tool is utilized to modify firewall settings, in this case, iptables, to block or reject IP addresses for a set duration. Additionally, it can be configured to perform various other actions, including sending email alerts detailing the detected suspicious activities.

In this architecture (Figure 7), Fail2Ban is integrated into the firewall protecting the SIP server, establishing an additional security measure for the communication system. The firewall, equipped with Fail2Ban, actively monitors and filters traffic, allowing it to intercept

https://doi.org/10.18080/jtde.v12n4.1022

potentially harmful packets before they reach the SIP server. Fail2Ban scrutinizes the traffic for patterns of malicious activity, such as numerous failed logins attempts or irregular request patterns.

Furthermore, Fail2Ban is configured to issue email alerts upon the implementation of an IP ban, providing immediate notification of the security response. This strategy provides a proactive layer of security, ensuring immediate action and awareness of potential threats. Upon detecting such activities, Fail2Ban triggers an IP ban on the firewall. This proactive approach ensures that the suspicious activity is stopped at the network's edge, preventing any impact on the SIP server itself. By integrating Fail2Ban with the firewall's capabilities, such as iptables, an enhanced security layer is formed. Iptables act as the first barrier by setting the traffic rules, while Fail2Ban offers a dynamic and responsive security layer, adapting to potential threats in real time. This strategy protects SIP servers against unauthorized access and potential compromises of client information, utilizing a comprehensive security model that includes a VPN, a firewall and Fail2Ban. The efficacy of this multi-layered defensive strategy is thoroughly evaluated through simulated cyberattack scenarios to ensure robust protection.

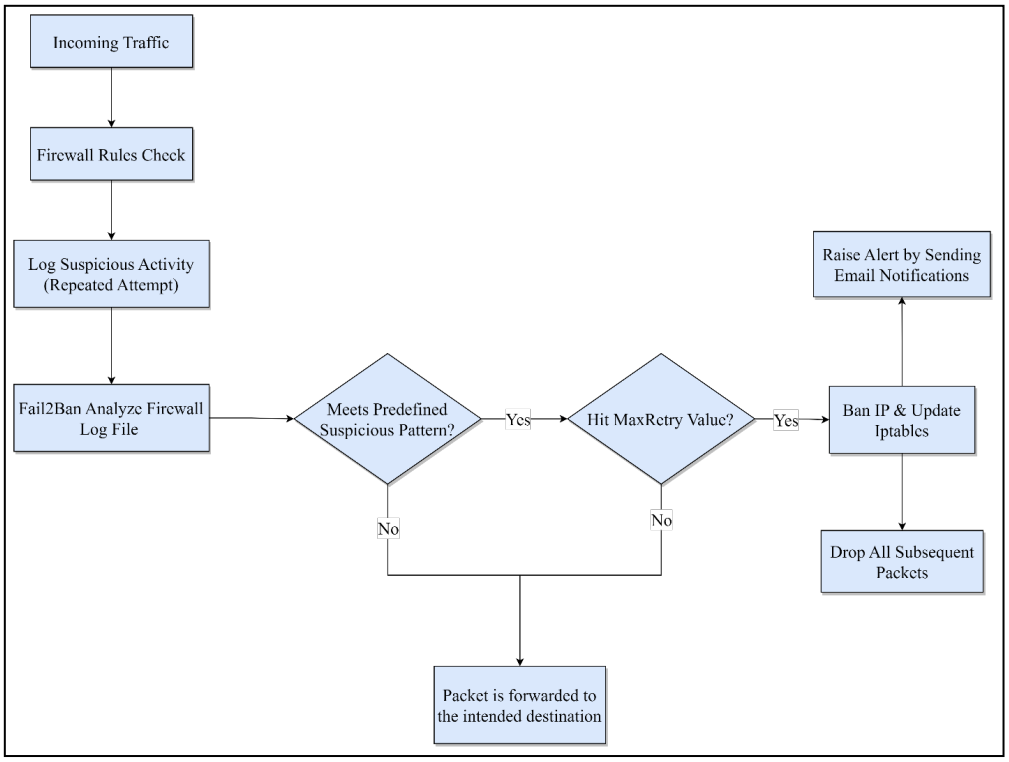


Figure 7. Operational flow of Fail2Ban

This methodology consists of four key sections, each essential for understanding the security and performance aspects of a controlled VoIP network.

https://doi.org/10.18080/jtde.v12n4.1022

1. Laboratory setup: This section describes the controlled environment where the VoIP infrastructure is replicated, including the configuration of virtual machines, network settings and security parameters.
2. Attack simulations: In this part, the approach to simulating various attack vectors, such as SIP scanning, eavesdropping, DoS and brute-force attacks, is outlined. Each attack's methodology is explained, considering the tools and techniques utilized to challenge the VoIP system's defences.
3. Performance analysis: This segment delves into the methods used to assess the VoIP system's performance, considering the delay, bandwidth, jitter and packet loss under different setup scenarios.
4. Test scenarios for security effectiveness and performance evaluation: The final section presents specific test scenarios to gauge the effectiveness of the proposed security solutions. It investigates the results of implementing VPN and firewall measures, examining how they influence the system's resilience to the simulated attacks.

Implementation

In our VoIP setup, we utilize a SIP server configured as both the proxy server and PBX, managing call signalling, media sessions and routing for SIP phones 7001 and 7002. Security is enhanced through the deployment of a WireGuard VPN and a firewall, creating a secure environment to mitigate risks such as unauthorized access and data breaches.

To address VoIP vulnerabilities**,** WireGuard VPN is employed for its low-latency, high-security performance, ensuring encrypted communication channels for all VoIP traffic. This approach is aligned with findings by Arpaci and Şentürk (2024), who demonstrated that combining VPNs with firewalls provides effective security without significantly impacting real-time performance metrics such as latency and jitter.

Additionally, a firewall was configured with iptables implements packet filtering and rate limiting to prevent common VoIP threats, including DoS attacks. This setup mirrors the security integration described by Tuleun (2024), who successfully applied IPsec VPN and firewall technology within an Asterisk-based VoIP system to enhance SIP communication security.

Simulation setup

The testbed operates on Ubuntu 20.04, with each virtual machine configured with 8GB RAM and dual-core processors to ensure stable performance for VoIP simulations and effective handling of firewall and VPN processes. The setup includes several key components: SIP phones, acting as user agents for call initiation; a SIP server configured as both a proxy server

https://doi.org/10.18080/jtde.v12n4.1022

and PBX, managing signalling and call routing; a router that handles network segmentation; and an attacker, which simulates external threats. Each SIP phone is assigned a static IP address and unique credentials to connect with the SIP proxy for authentication and call management over the default SIP port, 5060. These components work within a secure environment enhanced by VPN encryption and firewall protections.

The attacker is positioned within the 'External Network' segment, representing a public, non-trusted environment where potential eavesdropping or compromise of SIP phones or servers could occur. This setup allows simulation of attacks on the internal network to test the firewall and VPN's security efficacy, as the attacker shares the network segment with the external SIP phone, connected via a public or unsecured network.

Table 2. Virtual machine specifications

VM	Name	Software	Public IP	VPN IP	OS	Network adapter	Adapter type
1	SIP phone 7001	Linphone/3.12.0 (belle-sip/1.6.3)	10.0.5.5	192.168.133.6	Ubuntu 20.04.4 LTS	(Network Address Translation) NAT network	Intel PRO/1000 MT desktop
2	SIP server	Asterisk PBX 16.2.1	10.0.5.4	192.168.133.15			
3	Router	–	10.0.2.11 10.0.5.6	192.168.133.10			
4	SIP phone 7002	Linphone/3.12.0 (belle-sip/1.6.3)	10.0.2.7	192.168.133.7			
5	Attacker	–	10.0.2.9	–	Kali Linux 2021.1		

Note: The IP addresses labelled as 'Public IP' are used within the simulation to represent internet-exposed IP addresses but are not actual public IPs.

With the VPN enabled, the network utilizes three distinct IP ranges: 10.0.2.0/24 (external), 10.0.5.0/24 (internal), and 192.168.133.0/24 (VPN), providing logical separation between network elements and securing SIP client communication through encrypted VPN routing. The WireGuard VPN tunnel employs Curve25519 for key exchange and ChaCha20 for encryption, balancing low latency with strong security for real-time VoIP traffic. Public–private key pairs ensure mutual authentication between endpoints.

The firewall configuration further reinforces security, with rate limiting, IP whitelisting and stateful packet inspection. Fail2Ban dynamically blocks IPs exhibiting suspicious behaviour, such as repeated failed login attempts. This segmented IP structure not only separates internal and external traffic but also routes SIP client communication exclusively through the VPN, shielding it from direct exposure to potential attacks.

Overall, this configuration ensures that each component of the testbed is protected against common VoIP vulnerabilities, with robust firewall rules, VPN encryption and dynamic IP blocking to secure VoIP communication without compromising performance.

https://doi.org/10.18080/jtde.v12n4.1022

Results and Discussion

Attack mitigation results

While our firewall configuration lacks full layer 7 next-generation firewall capabilities, such as application-level SSL decryption, it provides essential stateful inspection to monitor and control VoIP traffic. Combined with WireGuard VPN encryption, the firewall helps secure data transmission, protecting against eavesdropping while maintaining efficiency and cost-effectiveness. By applying rate limiting, IP whitelisting and packet inspection at network layers 3–4, this setup manages traffic effectively without the higher costs or processing demands of more advanced layer 7 firewalls, which may be more than needed for basic VoIP security.

The attack simulations revealed the effectiveness of this security setup in handling four common VoIP attacks: SIP scanning, eavesdropping, DoS attacks, and SIP account brute-force attacks. Without security measures, the VoIP network showed high vulnerability. For example, SIP scanning enabled attackers to find open ports and identify valid SIP accounts, exposing network weaknesses. The eavesdropping simulation demonstrated that unencrypted SIP and RTP traffic could be intercepted, allowing unauthorized access to sensitive data.

After enabling the VPN, encryption prevented data interception, significantly reducing the risk of eavesdropping and improving overall security. However, during a DoS attack simulated through SIP INVITE flooding, system resources were still overloaded, reaching 100% traffic without filtering. By adding the firewall and Fail2Ban, most of the excess DoS traffic was successfully blocked, reducing captured INVITE requests to just 0.002%. This demonstrates that the tools effectively blocked attack patterns, preserving service quality. Similarly, Fail2Ban limited brute-force attacks on SIP accounts by restricting failed login attempts, allowing only three tries before blocking the attacker's IP. This layered defence mitigated brute-force vulnerabilities without reducing VoIP performance.

This study demonstrates that integrating VPN, a firewall and Fail2Ban effectively enhances VoIP security by addressing multiple threat vectors. By implementing these measures, the VoIP system maintained a secure environment with minimal performance impact, proving viable for real-world applications where consistent service quality is critical.

Performance analysis results

The results of the VoIP performance analysis are shown below. As mentioned in the previous section, a comparison of various security measures has been implemented.

VoIP and unified communication traffic rely on UDP, making them particularly sensitive to delay, jitter and packet loss. While adding a security layer, such as encryption via VPN or

https://doi.org/10.18080/jtde.v12n4.1022

packet filtering with firewalls, there are minimal performance impacts that can be managed to avoid disruption. Acceptable metrics for uninterrupted voice and video calls generally include a latency of under 150 ms, jitter within 20–50 ms, and packet loss close to 0%, ensuring minimal impact on MOS. Effective security implementations should balance these factors, maintaining the required quality of service while enhancing protection.

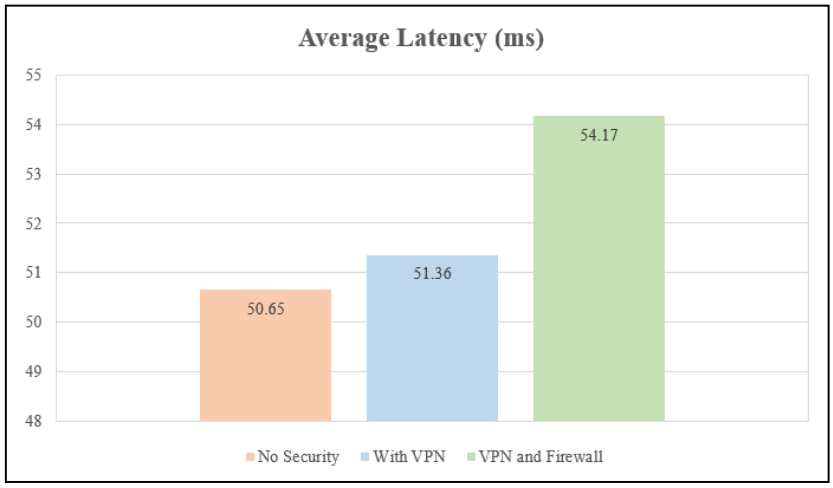


Figure 8. Average latency

As shown in Figure 8, latency is lowest with no security measures at roughly 50.65 ms, indicating the most efficient communication. Introducing a VPN increases the latency slightly to about 51.36 ms, likely due to the encryption and rerouting processes inherent to VPNs. The latency peaks at approximately 54.17 ms when both a VPN and a firewall are in use. The presence of a firewall, which inspects and filters traffic, combined with the VPN, contributes to this additional delay. This pattern underscores the impact of security protocols on VoIP performance, where each layer of security tends to add some degree of latency to voice data transmission. Indeed, these insights are invaluable for organizations that need to strike a balance between ensuring the security of their communication systems and minimizing any delays in voice exchanges.

The 'No Security' scenario in Figure 9 has the lowest average jitter at approximately 6.71 ms. This suggests that VoIP communications experience the least variation in packet arrival time when no security measures are implemented. When a VPN is added, the average jitter slightly increases to around 6.8 ms. This increment could be due to the VPN's encryption and tunnelling processes, which might introduce variability in packet transmission times. The highest average jitter is observed with the 'VPN and Firewall' configuration, at roughly 6.84 ms. The increase in packet arrival time variation can be attributed to the additional checks and balances carried out by the firewall, coupled with the VPN operations.

These security measures introduce extra processing steps and verification procedures that packets must go through before reaching their destination. As a result, the time it takes for

https://doi.org/10.18080/jtde.v12n4.1022

packets to travel this extended path can vary more significantly compared to a more direct, unsecured communication route. While these security measures are essential for protecting data and ensuring network security, they can introduce latency and variations in packet arrival times, which organizations need to consider when balancing security and network performance.

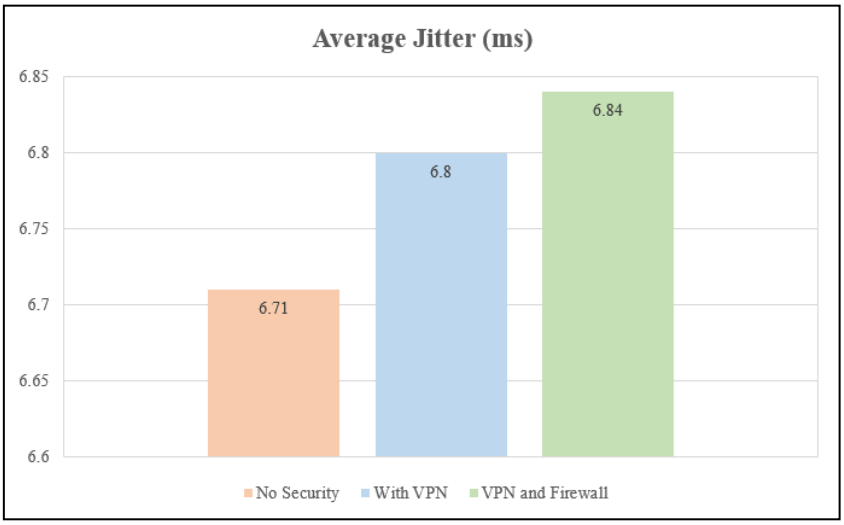


Figure 9. Average jitter

These observations highlight the impact of security measures on jitter in VoIP communications, with each layer of security potentially increasing the variability of packet delivery. This variability is an important aspect to consider for maintaining quality in VoIP communications, as high jitter can lead to poor audio quality.

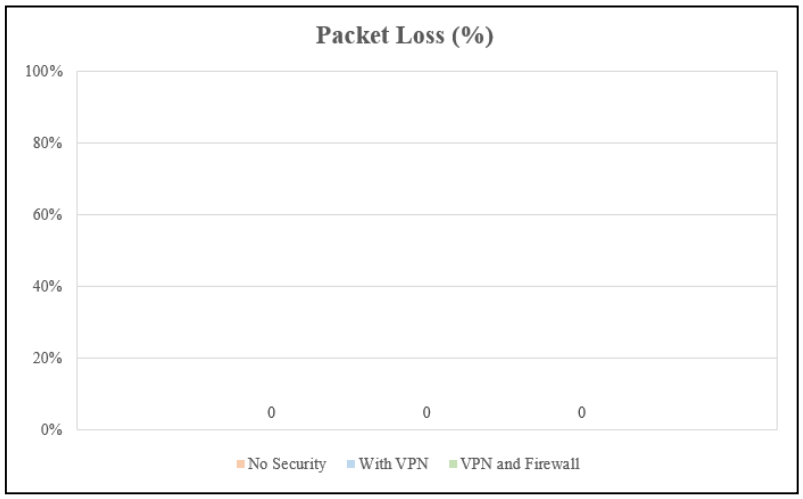


Figure 10. Packet loss

All three conditions in Figure 10 display a packet loss of 0%, indicating that there is no loss of data packets during transmission in any of the scenarios. This could imply that implementing a VPN and a firewall as security measures do not negatively impact the integrity of data transmission in terms of packet loss. It is also a positive indicator of network reliability and efficiency in the context of VoIP communication for these security settings.

https://doi.org/10.18080/jtde.v12n4.1022

The 'No Security' configuration in Figure 11 shows an MOS of approximately 4.3706, suggesting users perceive the quality of VoIP communication as very good when no security measures are implemented. This suggests that in a network environment without security measures, the quality of the VoIP calls is highly satisfactory. The introduction of a VPN shows a virtually identical MOS of about 4.3701, indicating that the use of a VPN has no significant impact on the perceived quality of the VoIP communication. This is a positive outcome, as it suggests that users can enjoy secure communications without sacrificing call quality. Finally, the 'VPN and Firewall' configuration has an MOS of around 4.3685, which is slightly lower than the other two configurations but still within a range indicating good quality. This configuration provides an additional layer of security through the firewall while maintaining reasonable call quality.

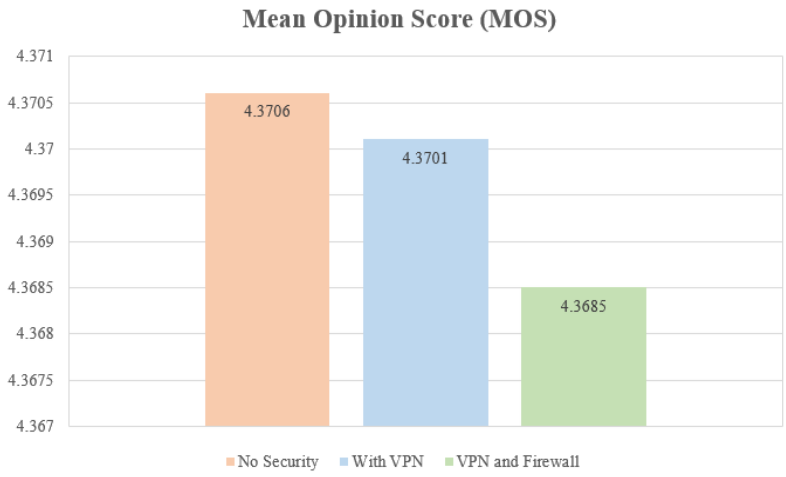


Figure 11. MOS

These slight variations in the MOS across different security configurations suggest that while security measures like VPNs and firewalls are in place, they do not substantially alter the perceived quality of VoIP communications according to the users' evaluations. This is beneficial for environments requiring secure communication without compromising the quality of voice transmission.

In short, the performance analysis across different security configurations for VoIP demonstrates that the implementation of VPNs and firewalls has a minimal impact on the quality and reliability of VoIP communication. Latency and jitter metrics show only slight increases when these security protocols are introduced, indicating that while there may be a slight performance trade-off, it does not significantly detract from the overall user experience. Furthermore, packet loss is unaffected by the addition of these security measures, which is critical for maintaining the integrity of voice communication. The MOS serves as a valuable and straightforward measure of how users perceive the quality of a particular experience. This measure consistently supports our findings by displaying similar scores in different scenarios.

https://doi.org/10.18080/jtde.v12n4.1022

Therefore, the takeaway from these results is that the proposed enhanced security solutions can improve the VoIP security without concern about a significant drop in voice quality.

Conclusions and Recommendations

The study analyzes VoIP systems, identifying vulnerabilities and cyber threats, and assesses current security measures. It proposes a methodology that combines VPNs, firewalls and Fail2Ban techniques to enhance security. The analysis confirms that this integrated approach significantly boosts protection against threats, with data supporting the effectiveness of this strategy.

Initial analysis revealed that unencrypted SIP and RTP traffic made the VoIP system vulnerable to eavesdropping. Implementing a VPN encrypted this traffic, while using a VPN alongside a firewall in a simulated DoS attack reduced attack traffic from 100% to just 0.002%. The firewall blocked malicious traffic and reduced server load. Additionally, integrating Fail2Ban with the firewall effectively thwarted brute-force attacks by limiting REGISTER attempt packets to three, based on Fail2Ban's settings. This enhanced the system's resilience, with a minor increase in service response and call setup times from 0.23 ms to 0.33 ms.

The performance analysis showed that while security measures increased latency and jitter by no more than 5 ms, and packet loss remained at 0%, the MOS slightly decreased from 4.37 to 4.36, still indicating high user-perceived quality. Despite minor performance impacts, the security enhancements effectively mitigated all simulated attacks, demonstrating the effectiveness of the proposed architecture.

Overall, the integration of firewall and VPN solutions significantly improved VoIP security while maintaining service quality, proving that security enhancements can be made without compromising performance or user experience.

Acknowledgements

We would like to acknowledge the support of TM R&D grant project (RDTC/231106; MMUE/230053) from TM R&D Sdn. Bhd. MMU and thank the members for their contribution to and collaboration with this research publication.

References

Abdulazeez, A. M., Salim B. W., Zeebaree D. Q., & Doghramachi D. (2020). Comparison of VPN protocols at network layer focusing on wire guard protocol. *International Journal of Interactive Mobile Technologies*. *14*(18), 157. https://doi.org/10.3991/ijim.v14i18.16507

Abualhaj, M. M., Al-Khatib, S. N., Kolhar, M., Munther, A., & Alraba'nah, Y. (2020). Effective voice frame pruning method to increase VoIP call capacity. *TEM Journal*, *9*(1). https://doi.org/10.18421/TEM91-08

Almutairi, A. A. (2018). Toll-fraud protection, detection and prevention. *International Journal of Intelligent Computing Research*, *9*(3). 939–943. https://doi.org/10.20533/ijicr.2042.4655.2018.0113

Arpaci, S., & Şentürk, A. (2024). Performance analysis of firewall and virtual private network (VPN) usage in video conferencing applications. *Duzce University Journal of Science and Technology*, *12*(4), 1879–1894. https://doi.org/10.29130/dubited.1462133

Chakraborty, P., & Telgote, A. M. (2019). Performance analysis of LAN, MAN, WAN, and WLAN topologies for VoIP services using OPNET modeler. In Iyer, B., Nalbalwar, S., Pathak, N. (Eds.), *Computing, Communication and Signal Processing, (*p. 810*).* Springer. https://doi.org/10.1007/978-981-13-1513-8_20

Chakraborty, T., Misra, I. S. & Prasad, R. (2019a). Overview of VoIP technology. In *VoIP Technology: Applications and Challenges*. Springer. https://doi.org/10.1007/978-3-319-95594-0_1

Chakraborty, T., Misra, I. S. & Prasad, R. (2019b). VoIP protocol fundamentals. In *VoIP Technology: Applications and Challenges*. Springer. https://doi.org/10.1007/978-3-319-95594-0_2

Drew, R. (2023). *VoIP advantages & disadvantages: All you need to know*. getvoip.com. https://getvoip.com/blog/voip-advantages-and-disadvantages/#easily-scalable Accessed July 24, 2023.

Fasiku, A. I. (2018). A review of voice over internet protocol. (2018). *Journal of Scientific and Engineering Research*, *5*(7), 96–10.

Future Market Insights Inc. (2023). *Voice over internet protocol (VoIP) market.* https://www.futuremarketinsights.com/reports/voice-over-internet-protocol-market Accessed July 21, 2023.

He, X. (2021). Research on computer network security based on firewall technology. *Journal of Physics: Conference Series*, *1744*(4), 042037. https://doi.org/10.1088/1742-6596/1744/4/042037

IBM. (2024). *Cost of a Data Breach Report 2024*. https://www.ibm.com/reports/data-breach. Accessed 25/11/2024.

Jingi, A. M. (2017). VoIP security: Common attacks and their countermeasures. *International Journal of Computer Science and Information Security*, *15*(3).

Jumakhan, H., & Mirzaeinia, A. (2024). WireGuard: An efficient solution for securing IoT device connectivity. arXiv. https://export.arxiv.org/abs/2402.02093

Jimoh, S. T., & S Al-Juboori, S. (2024). Cyber-Securing Medical Devices Using Machine Learning: A Case Study of Pacemaker. *Journal of Informatics and Web Engineering*, *3*(3), 271–289. https://doi.org/10.33093/jiwe.2024.3.3.17

Kumar, V., & Roy, O. P. (2021). Security and challenges in voice over internet protocols: a survey. *Materials Science and Engineering A*, *1020*(1), 012020. https://doi.org/10.1088/1757-899X/1020/1/012020

Liu N., (2024). *VoIP security: Vulnerabilities & best practices*. Yeastar. https://www.yeastar.com/blog/voip-secuirty-best-practices/ Accessed July 24, 2023.

Luhach, R., Jha, C. K., & Luhach, A. K. (2019). Research and analysis for adaptive IFIR Filters for voice quality enhancement in wireless VoIP. *Journal of Discrete Mathematical Sciences and Cryptography*, *22*(4), 581–590, https://doi.org/10.1080/09720529.2019.1645392

Mentsiev, A. U., & Supaeva, Kh. (2019). VoIP techniques. *Engineering Bulletin of the Don 1*(52), 65.

Mentsiev, A. U., & Dzhangarov, A. I. (2019). VoIP security threats. Инженерный вестник Дона [Engineering Bulletin of the Don], *1*(52), 75. Accessed July 24, 2023. https://cyberleninka.ru/article/n/voip-security-threats

Mrewa, N., Mohd Ramly, A., Amphawan, A., & Neo, T. K. (2024). Optimizing Medical IoT Disaster Management with Data Compression. *Journal of Informatics and Web Engineering*, *3*(1), 55–66. https://doi.org/10.33093/jiwe.2024.3.1.4

Munusamy, T., & Khodadi, T. (2023). Building Cyber Resilience: Key Factors for Enhancing Organizational Cyber Security. *Journal of Informatics and Web Engineering*, *2*(2), 59–71. https://doi.org/10.33093/jiwe.2023.2.2.5

Nazih, W., Hifny, Y., Elkilani, W., Abdelkader, T., & Faheem, H. (2019). Efficient detection of attacks in SIP based VoIP networks using linear L1-SVM classifier. *International Journal of Computers Communications & Control*, *14*(4), 518–529. https://doi.org/10.15837/ijccc.2019.4.3563

Oproiu, M., Muşuroi, C., & Volmer, M. (2020). Low cost and integrable healthcare services using VoIP for remote patient monitoring. *2020 International Conference on e-Health and Bioengineering*, 1–4. IEEE Xplore. https://doi.org/10.1109/EHB50910.2020.9280206

Pudelko, M., Emmerich, P., Gallenmüller, S., & Carle, G. (2020). Performance analysis of VPN gateways. *2020 IFIP Networking Conference (Networking)*, *France*, 325–333.

Rathore, V. S., Dey, N., Piuri, V., Babo, R., Polkowski, Z., & Tavares, J. M. R. S. (2021). *Rising threats in expert applications and solutions*. Springer Singapore.

Sadiwala, D. R. (2018). Analysis of security threats of VoIP systems. *RKDF University Journal of Science and Engineering*, *01*(02), 34.

Surasak, T., & Huang, S. C-H. (2019). Enhancing VoIP security and efficiency using VPN. *2019 International Conference on Computing, Networking and Communications (ICNC)*, 180–184. IEEE Xplore. https://doi.org/10.1109/ICCNC.2019.8685553

Suthar, D., & Rughani, P. H. (2020). A comprehensive study of VoIP security. *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, 812–817. IEEE Xplore. https://doi.org/10.1109/ICACCCN51052.2020.9362943

Tay, Y. H., Ooi, S. Y., Pang, Y. H., Gan, Y. H., & Lew, S. L. (2023). Ensuring Privacy and Security on Banking Websites in Malaysia: A Cookies Scanner Solution. *Journal of Informatics and Web Engineering*, *2*(2), 153–167. https://doi.org/10.33093/jiwe.2023.2.2.12

Tuleun, W. (2024). Design of an asterisk based VoIP system and the implementation of security solution across the VoIP network. *World Journal of Advanced Research and Reviews*, *23*(1), 10–30574. https://doi.org/10.30574/wjarr.2024.23.1.2048

Dynamic Touchstroke Analysis with Explainable Artificial Intelligence Tree-Based Learners

Wun Puo Lim
Faculty of Information Science and Technology (FIST), Multimedia University, Melaka, Malaysia

Shih Yin Ooi
Faculty of Information Science and Technology (FIST), Multimedia University, Melaka, Malaysia

Ying Han Pang
Faculty of Information Science and Technology (FIST), Multimedia University, Melaka, Malaysia

Soodamani Ramalingam
School of Physics, Engineering & Computer Science, Department of Engineering and Technology, University of Hertfordshire, United Kingdom

Yee Jian Chew
Faculty of Information Science and Technology (FIST), Multimedia University, Melaka, Malaysia

Abstract: As mobile devices become integral to daily life, robust authentication methods are essential for ensuring security. Traditional methods like personal identification numbers and swipe patterns remain vulnerable to social engineering attacks. To address these risks, this study investigates behavioural biometrics, specifically touch-stroke dynamics, as a transparent and secure alternative. By leveraging unique user interaction patterns, such as touch speed and pressure, this approach provides a distinctive means of authentication. Although various machine learning techniques are available for touch-stroke analysis, the interpretability of classification decisions is vital. This paper implements explainable artificial intelligence with tree-based learners, specifically decision trees and random forests, to enhance the transparency and effectiveness of touch-stroke dynamic authentication. Performance evaluations show that random forests achieve equal error rates (EER) between 0.03% and 0.05%, and decision trees yield EERs between 0.02% and 0.07%, demonstrating a balance between security and interpretability for mobile authentication.

Keywords: Behavioural Biometrics, Touch-stroke dynamic, Explainable Artificial Intelligence (XAI), Random forest (RF), Decision tree (DT)

https://doi.org/10.18080/jtde.v12n4.1023

Introduction

Mobile devices have enhanced human convenience in this modern generation (Wang *et al.*, 2020; Seek *et al.*, 2023). Despite their convenience, traditional authentication methods such as PINs and swipe patterns are increasingly vulnerable to security threats. Personal identification numbers (PINs), for instance, are often simple to guess, especially when users rely on common or easily deducible combinations. Similarly, although appearing more complex, swipe patterns can be easily observed and replicated by onlookers. To address the shortcomings of traditional authentication methods, biometric authentication has emerged as a promising alternative.

Several researchers (Miraoui & El-Etriby, 2019; Spaling & Singh Uppal, n.d.; Voege *et al.*, 2022; Voege & Ouda, 2022) suggested that authentication can be based on four factors: "Something you know", "Something you have", "Something you are", and "Something you do". In this paper, biometric authentication applies two of these factors: "Something you are" and "Something you do". "Something you are" refers to unique personal characteristics, such as fingerprints, facial features, or irises, that distinguish one individual from another. "Something you do" refers to biometric behaviours, which are based on an individual's habits or actions, such as keystroke dynamics, mouse movement patterns, touch-stroke dynamics, and gait recognition. Together, "Something you are" and "Something you do" provide a higher level of security, as each person's physical attributes and behavioural patterns are distinct.

Biometric authentication can be separated into two categories. The first is physiological, and the second is behavioural (Alwahaishi & Zdralek, 2020). Examples of physiological biometrics modalities include fingerprint, facial, iris, hand geometry, retina, etc. On the other hand, behavioural biometrics include voice recognition, signature verification, keystroke dynamic, gesture biometrics, touch-stroke dynamic, etc. Touch-stroke analysis is one of the most suitable implementations and is comparable to typing PINs or swipes on mobile devices. Compared with others, this modal is relatively new, and its process can be completely transparent to the users. The user still types the PIN and swipes as usual. However, when it comes to authentication, the factors to be considered now are not only the right match of PIN (a string of digits) or the swiping patterns because touch-stroke analysis also considers the behavioural patterns of how the users key in the PIN and how they perform the swipe, such as the dynamic speed and pressure of touching the screens (Ooi & Teoh, 2019).

Beyond classification performance, it is essential to grasp the critical factors driving a machine learning algorithm's decisions because this knowledge is pivotal for defending against social engineering and replay attacks. Thus, in this paper, two tree-based learners are adopted to test their classification ability on three benchmark touch-stroke datasets: (1) Frank dataset, (2) E-

https://doi.org/10.18080/jtde.v12n4.1023

BioDigit database, and (3) MobileTouchDB database. On top of this, explainable artificial intelligence (XAI) frameworks are adopted to visualise the factors and rules governing touch-stroke classification. We postulated that this visualization will aid security professionals in countering replay attacks by identifying which factors or attributes should be encrypted and understanding what information may be vulnerable to tampering through sniffing attacks.

Literature Review

Touch-stroke dynamics is a relatively new and developing biometric behavioural authentication method that analyses how individuals interact with touch screens. It involves analysing the distinctive manner in which an individual engages with a touch screen, considering factors such as pressure, speed, and finger movement (Ooi & Teoh, 2019). However, one of the main challenges to adopting it in real-world applications would be collecting and processing larger data on mobile devices, which requires efficient algorithms and collection capabilities to derive verified user identities accurately.

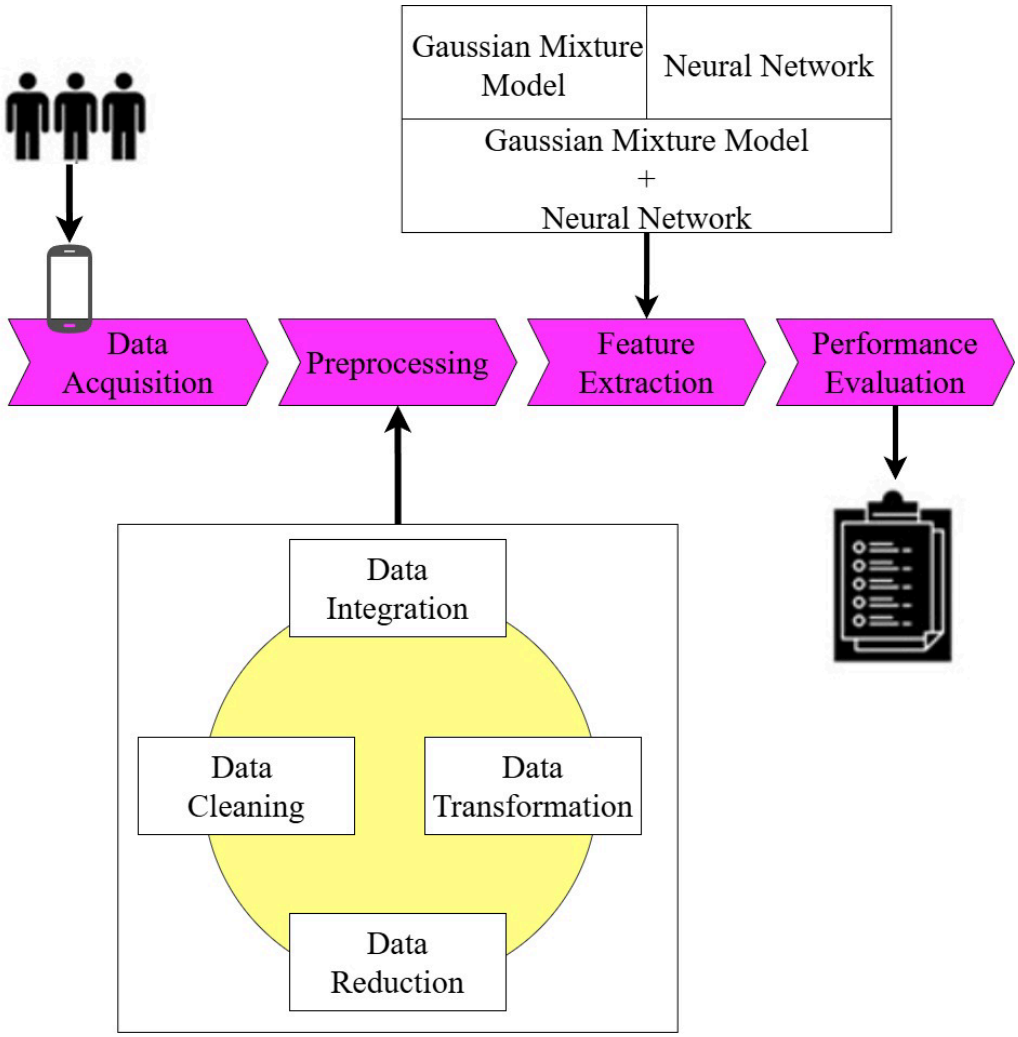


Figure 1. Workflow of touch-stroke dynamic dataset training process.

https://doi.org/10.18080/jtde.v12n4.1023

The benchmark methodologies for most biometric authentication systems, including touch-stroke dynamic, are (1) data acquisition, (2) data preprocessing, (3) feature extraction, and (4) performance evaluation (Gavisiddappa *et al.*, 2020; Liu *et al.*, 2022; Z. Wang *et al.*, 2021), as shown in Figure 1. Data acquisition for touch-stroke dynamic is used to collect data on touch pressure, speed, acceleration, finger angle preference, contact area, and position data by using sensitive touch screens and built-in sensors on the smartphone. The next process is data processing, which aims to reduce the interference in collected data to minimise the error message and capture valuable data. The third process is feature extraction, which derives the touch dynamic–related parameters collected. The popular methods adopted include the Gaussian mixture model (GMM), support vector machine (SVM), neural network (NN), label encoder, and one hot encoder.

Lastly, a classifier is an algorithm that can automatically categorise data into classes (*Classifier Definition | DeepAI*, 2023). There are many types of classifiers, such as *k*-nearest neighbour, random forest (RF), dynamic time warping, multilayer perceptron (MLP), decision tree, etc. Some of them are discussed further in the following sections of this paper.

Random Forest (RF)

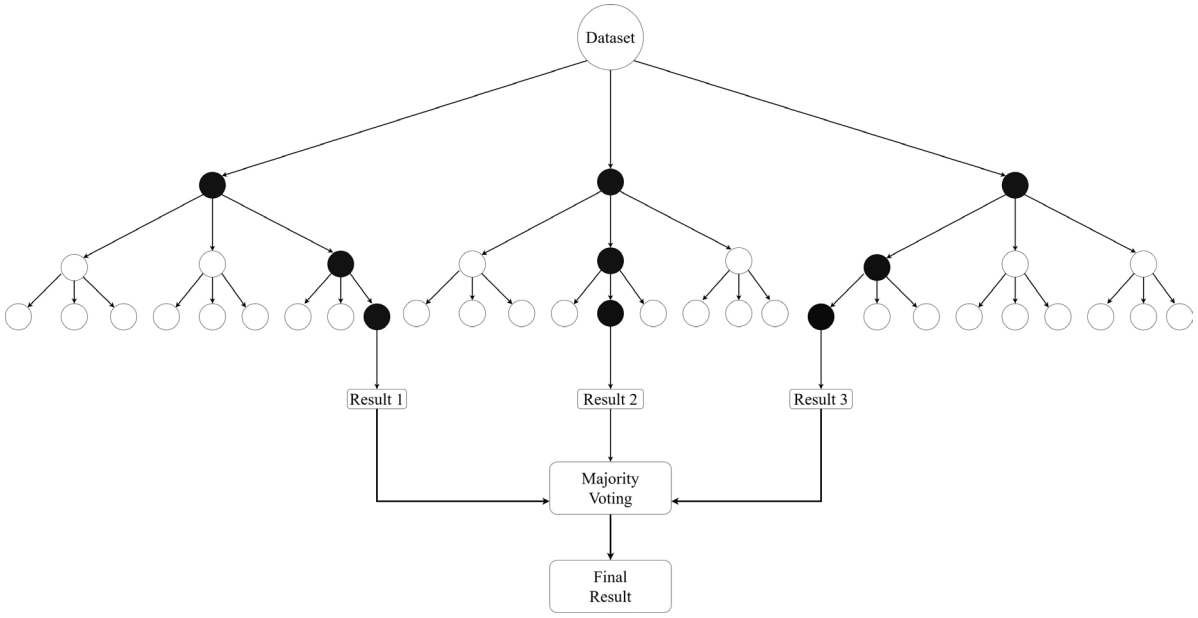


Figure 2. Random forest ensemble model diagram (modified from *Random Forest Algorithm in Machine Learning | Great Learning, 2024*).

The RF is a machine-learning algorithm introduced by Bello *et al.* (2020). The final decision is formed based on the votes given in each decision tree within the forest (*Random Forest Algorithm for Absolute Beginners in Data Science*, 2024). With its ensemble learning mechanism, it is believed to tolerate more biases than single decision trees. The RF algorithm is a tree classifier using bootstrapped samples and selected attributes to generate training data. The majority vote sum in an RF can then be used to classify data samples. RF develops several

https://doi.org/10.18080/jtde.v12n4.1023

different decision trees and classifies or predicts the outcome for each tree during training, as shown in Figure 2.

After training the dataset (Equation (1)), a majority vote will be determined x (Equation (2)).

$$X = x_1, \ldots, x_n$$
$$Y = y_1, \ldots, y_n \quad (1)$$
$$f = \frac{1}{B}\sum_{b=1}^{B} f_b(x) \quad (2)$$

In addition, regression forest is a specific variation of the RF algorithm tailored to perform purely regression-based functions (Ooi & Teoh, 2019). RF performs regression or classification tasks, whereas regression forest handles regression operations alone. Ooi & Teoh (2019) proposed a new temporal regression forest architecture that emerges by juxtaposing temporal frames and classical regression forests. This integration allows classical regression forests to deal with the order of touch-stroke feature vectors. The new splitter selection randomisation approach will guide tree growth in temporal regression forests for temporal ordering purposes, and this procedure can be employed in SVM, MLP, or naïve Bayes to illuminate the relevance of the incorporation of temporality.

RF has proven to reduce overfitting in decision trees, thereby enhancing accuracy. Additionally, it is a robust ensemble learning technique that effectively addresses both classification and regression tasks, demonstrating flexibility even when working with discrete or continuous input data attributes. Furthermore, RF is highly automated, handling missing values seamlessly during training (Random Forest Algorithm in Machine Learning | Great Learning, 2024). Notably, it operates effectively without the need for data normalisation due to its rule-based approach, making it a practical and user-friendly solution for a wide range of machine learning applications.

Support Vector Machine (SVM)

SVM is a binary classifier that uses points in space to represent data samples. Furthermore, SVM aims to differentiate various groupings by a considerable margin. It uses the nearest data points, also called support vectors, to discover the hyperplane that contains the maximising between two objects, as shown in Figure 3.

Bello *et al.* (2020) proposed a rudimentary SVM calculation. The training set T of data in SVM can be written as in Equation (3). In addition, the training set T comprises n-dimensional vectors X_i and corresponding classes Y_i (1 or −1). The "w" mean weight vector for the "b" is the error rate calculated in the training process, as depicted in Equation (4). Finally, to obtain good classification results in a training set, it is required that $F(.)$ (w & b) to return a positive

https://doi.org/10.18080/jtde.v12n4.1023

value if the point X_i is classified positively and negatively otherwise (refer to Equation (5) and Figure 4).

$$T = \{(X_1, X_1), (X_2, X_2), (X_3, X_3), \dots, (X_m, X_m)\} \quad (3)$$

$$F(x) = w \cdot x - b \quad (4)$$

$$w \cdot x_i - b > 0 \;\; if \; y_i = 1 \; \& \; w \cdot x_i - b > 0 \;\; if \; y_i = -1 \quad (5)$$

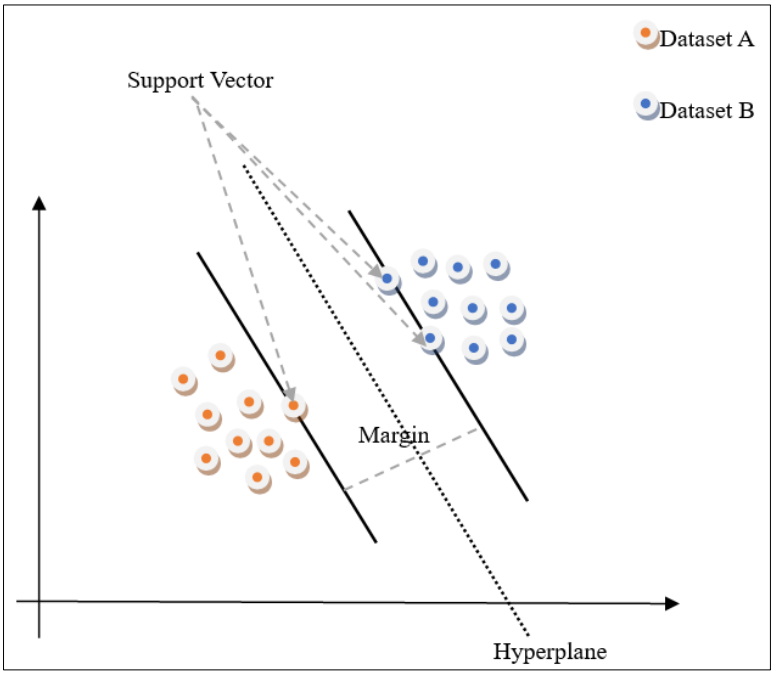


Figure 3. Support vector machine ensemble model diagram (modified from Chapter 7. Learning (II): SVM & Ensemble Learning | Data Analytics: A Small Data Approach, n.d.)

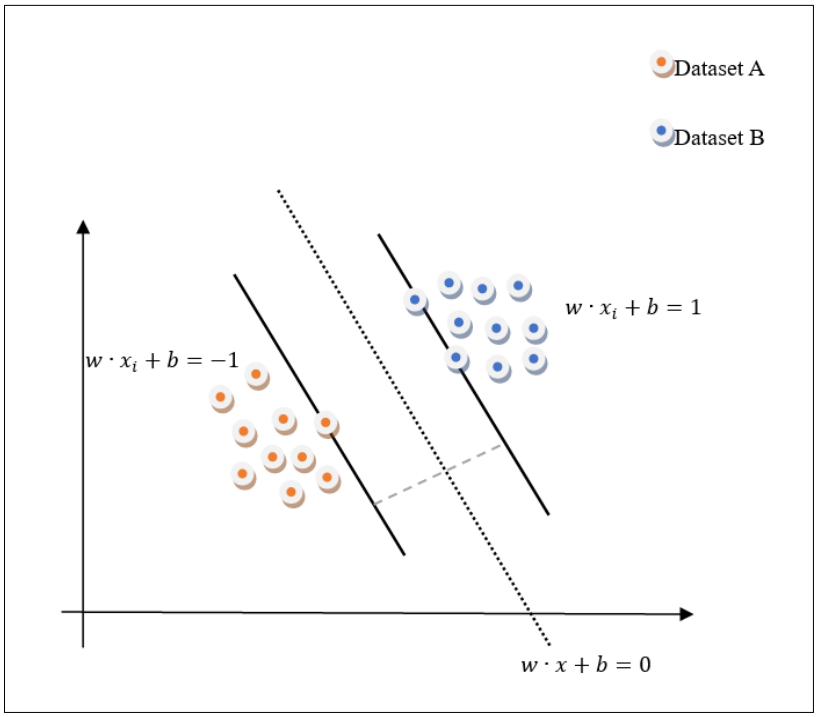


Figure 4. Support vector machine ensemble model diagram (modified from Chapter 7. Learning (II): SVM & Ensemble Learning | Data Analytics: A Small Data Approach, n.d.).

https://doi.org/10.18080/jtde.v12n4.1023

There are a few reasons for choosing SVM as a classifier. Firstly, the SVM margin of separation between classes is clear; it can work very well. It also works well in multidimensional domains in which the number of dimensions exceeds the number of samples (Top 4 Advantages and Disadvantages of Support Vector Machine or SVM | by Dhiraj K | Medium, 2024). However, SVM also has weaknesses, such as complex algorithms, multiclassification problems, and weaknesses when performing on imbalanced datasets. The SVM algorithm becomes complex due to the large training dataset. Besides that, SVM also involves solving quadratic programming (QP) problems. So, it needs to take a lot of processing effort and memory. SVM also faces the challenge of dealing with unbalanced data, in which the imbalanced dataset can cause the model to be biased and perform poorly in several classes.

To address the limitations of SVM, Cervantes *et al.* (2020) explored various methods to handle the complexity of SVM algorithms. Simple random sampling is a data selection technique that reduces dataset size while preserving the hyperplane's separation properties. Distance-based approaches, using measures like Euclidean or Mahalanobis distance, are effective, as are methods such as condensed nearest neighbour and hierarchical clustering, which enhance SVM's performance on large datasets. Decomposition techniques, including chunking and sequential minimal optimisation, focus on active constraints to reduce training time, although they can be time-consuming if not carefully managed. Alternative formulations, like least-squares SVM and parallelising SVM, allow faster training, albeit sometimes at the expense of accuracy. Parallel implementations, such as approximating the kernel matrix with a block diagonal matrix, and geometric approaches, which seek optimal hyperplanes through convex hulls, further optimise SVM. Heuristic methods, like alpha seeding, are also used to generate initial solutions for QP problems.

Neural Network (NN)

An NN is a computational model that mimics the way nerve cells function in the human brain (Bello *et al.*, 2020). It belongs to a class of machine learning algorithms and is fundamental in various applications, including pattern recognition, classification, regression analysis, and decision-making. The basic building blocks of an NN are artificial neurons, or perceptrons, which are organised into three layers: the input layer, hidden layers, and the output layer, as illustrated in Figure 5. There is only one layer for the input and output layer, but the hidden layer can be more than one (Figure 6). Each connection between neurons has an associated weight. NNs learn by adjusting these weights based on input data and desired output.

https://doi.org/10.18080/jtde.v12n4.1023

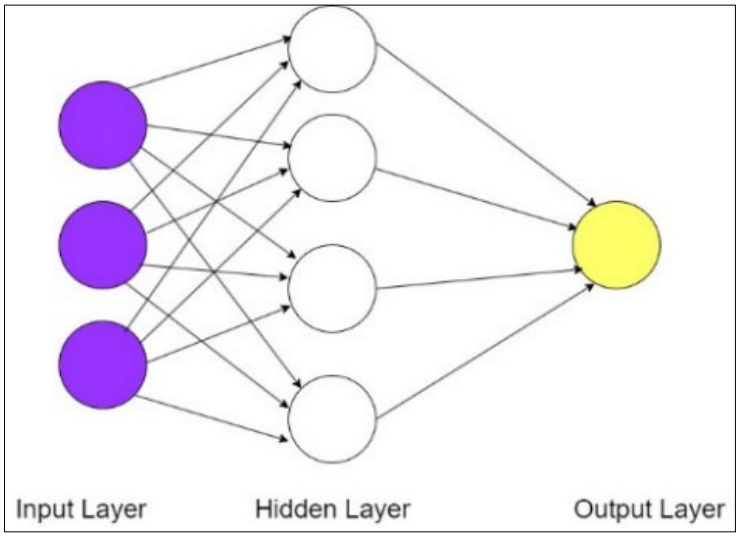


Figure 5. Neural network ensemble model diagram with one hidden layer diagram (modified from *Deep Neural Network Architecture | Download Scientific Diagram*, n.d.).

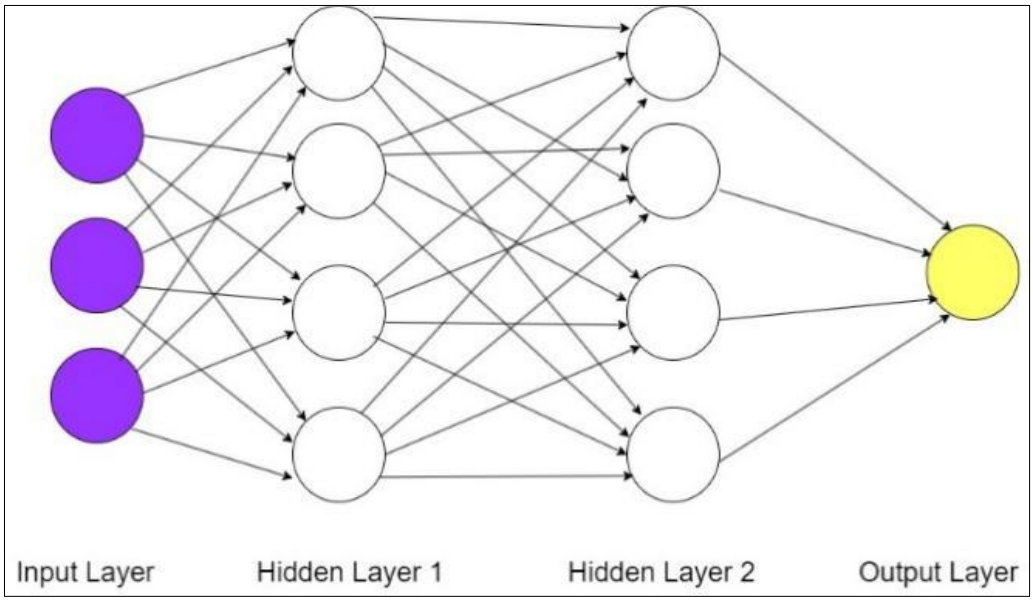


Figure 6. Neural network ensemble model diagram with more than one hidden layer diagram (modified from *Deep Neural Network Architecture | Download Scientific Diagram*, n.d.).

In the learning process, an NN is trained on a dataset. The input data are fed into the network and compared with the desired output. The difference between the expected output and the actual one is used to readjust weights in the network using a process called backpropagation. The iterative training process allows the NN to improve at making accurate predictions or classifications (Bello *et al.*, 2020). The input layer is designated "d", the hidden layer is "M", and finally, the output layer is "c". As shown in Equation (6), the input features, x_i, are combined with weights, $w_{ji}^{(1)}$, and added together for each neuron in the hidden layer. This is essentially to calculate a weighted sum for each hidden neuron. Then, an activation function, g, is applied to introduce non-linearity and allow the model to learn more complex patterns. Next, the outputs from the hidden layer are combined again, this time using another set of

https://doi.org/10.18080/jtde.v12n4.1023

weights, $w_{kj}^{(2)}$, to compute a weighted sum for each output neuron. Finally, the activation function, δ, is used to produce the final output, y_k. To summarise, Equation (6) shows the process from the input layer through the hidden layer to the output layer using weight and activation functions at each layer.

$$y_k = \delta\left(\sum_{j=1}^{M} w_{kj}^{(2)} g\left(\sum_{i=1}^{d} w_{ji}^{(1)} x_i\right)\right) \quad (6)$$

NNs are especially advantageous because they solve regression, classification, and numerical problems. Besides that, they enable us to take approximation functions of different forms. For nonlinear data or jobs with many features like image processing, NNs are extremely effective. For complicated problems of the probability of classification, they also excel in splitting large and complex ones into a collection integrated set of hierarchical trees made up of simple pieces. They can be tailored to any input and number of layers. The greater the dataset size, the better their performance will be. However, neural nets are black boxes. One cannot know what effect independent variables have on dependent variables. They also are computationally expensive and time-consuming to train (mostly on conventional central processing units). It also suffers from overfitting or interpolating the model, making it perform badly (Neural Networks with R, 2024). Popular variations of NNs include artificial NNs, convolutional NNs, recurrent NNs, and MLP.

Explainable Artificial Intelligence (XAI)

There are two types of AI algorithms: inherently explainable and inherently opaque. When inherently explainable and designed for humans to understand how the algorithms (classifiers) make decisions, it can also be called a “glass box”. Few methods, such as decision trees, regression algorithms, Bayesian classifiers, and SVMs, are inherently explainable. On the other hand, inherently opaque systems, known as “black box” models, include algorithms like NNs, which although highly effective, are difficult to interpret and understand.

To address the opacity of these models, XAI aims to interpret and elucidate the results of model classifications, thereby enhancing transparency in the decision-making process, fostering a rapid increase in its adoption (Qin *et al.*, 2024). XAI refers to the process of enunciating the inner workings of an AI model, such as its expected results and potentially biased outcomes (Barredo Arrieta *et al.*, 2020; *Explainable AI | Royal Society*, 2024). It ensures that AI-backed decision-making systems are exact, accurate, and reliable. Besides that, XAI integration is required for organisations because it builds trust and confidence in AI models when deployed in real-time.

https://doi.org/10.18080/jtde.v12n4.1023

XAI can assist developers in checking whether the system is working as desired, complying with the relevant regulations, or enabling those affected by a decision to engage or make enough changes to reverse that decision (Barredo Arrieta *et al.*, 2020; Explainable AI | Royal Society, 2024). However, XAI still faces some challenges, such as the transferability of post hoc explainability methods, the lack of globally accepted standards and measures for AI and/or machine learning system explainability, the trade-off between explainability and performance, and the difficulty of explaining deep learning models (Rawal *et al.*, 2022). There are several models and techniques within XAI, including decision trees, Shapley additive explanations, local interpretable model-agnostic explanations, counterfactual explanations, permutation feature importance, etc (Types of Explainable AI, 2024).

Decision Trees

Decision trees are one of the most straightforward and transparent models, making them easy for humans to understand and interpret (Rawal *et al.*, 2022; Types of Explainable AI, 2024). A decision tree is structured with different nodes and branches, starting with a root node that represents the initial point of decision-making. From the root, branches emerge representing different decision paths, each leading to further nodes that represent subsequent decisions or outcomes (Decision Trees - Graphical Explainable AI • Aggregata, 2024). This branching structure allows users to follow the entire sequence of decisions, building the tree structure until terminal nodes are reached, which signifies the final classification or regression result.

Decision trees are very flexible, suitable for classification and regression, and easy to understand the true or false logic of the decision tree (Decision Trees - Graphical Explainable AI • Aggregata, 2024) and thus have been customized in difference variations in tackling different datasets (Ooi et al, 2019). However, it has high variance estimators. When the data have small changes, it results in completely different decision trees. Figure 7 shows an example of the decision tree.

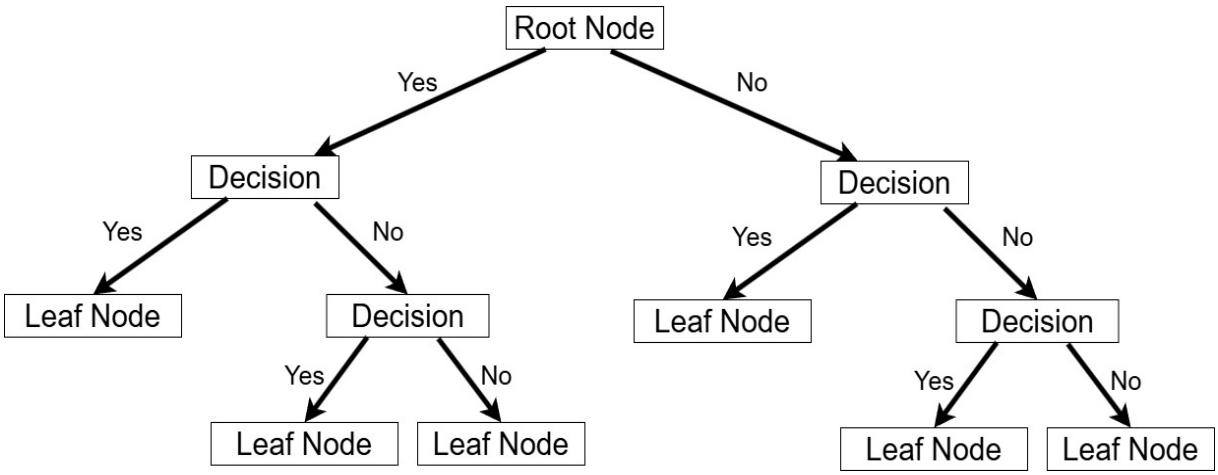


Figure 7. Decision Tree ensemble model diagram (modified from *Decision Tree Tutorials & Notes | Machine Learning | HackerEarth*, n.d.).

https://doi.org/10.18080/jtde.v12n4.1023

Benchmark and Public Touch-Stroke Dataset/Database

There are several public datasets available for touch-stroke analysis and research purposes. In this paper, three datasets are adopted for the empirical study. Each of them is further discussed in the following section.

Frank Dataset

This Frank dataset has been widely used in research, as evidenced by its implementation in various studies (Bajaber *et al.*, 2022; Delgado-Santos *et al.*, 2024; Montgomery *et al.*, 2019; Ooi & Teoh, 2019; Keykhaie & Pierre, 2020). It is a comprehensive collection of swiping data from 41 users gathered during two successive sessions conducted one week apart on Android devices, primarily for the purpose of image matching and text reading applications (iRAPs). The dataset captures several parameters, including the x and y coordinates of the finger touch point, timestamps, finger area (in pixels), pressure, device orientation, and finger orientation. Data were collected using several devices with varying sampling frequencies, but all subjects provided portrait-oriented data, with downward strokes being significantly more common than upward ones. Although downward strokes dominate the dataset, landscape orientation strokes are relatively rare. The Frank dataset is publicly available and can be downloaded from the following link: https://www.mariofrank.net/touchalytics/.

E-BioDigit Database

Tolosana *et al.* (2020) designed the E-BioDigit Database specifically for experimental use with online handwritten digits ranging from zero to nine. The data were captured using a Samsung Galaxy Note tab with a 10.1-inch display and LCD. Data were collected over two sessions spaced at least three weeks apart to reduce intersession variability. Participants had to write out four numerical sequences for each session, namely zero to nine, using only their fingers. It resulted in a total of eight samples per digit and user.

To reduce user variability, a rectangular space was chosen for writing with the same proportions as modern 5-inch screens. Participants could assess the quality of their input using the "OK" and "Cancel" buttons. The database includes data from 93 users, with an age range of 17 to 27 years. Out of these 93 participants, 89.2% were right-handed, and the male-to-female ratio was 66.7% to 33.3%. This dataset is available for download by official request from https://github.com/BiDAlab/eBioDigitDB.

MobileTouch Database

Tolosana *et al.* (2020) also developed the MobileTouch database, a comprehensive collection of touch biometric handwritten character data. This database includes over 64,000 online character samples from 217 participants. Data acquisition was facilitated by an Android-based

https://doi.org/10.18080/jtde.v12n4.1023

mobile app featuring interfaces for "drawing characters" and "password" input, with the "OK" and "Cancel" buttons. An unsupervised mobile scenario simulation was developed, allowing users to generate touch screen data from any position on the screen using an app available on the Play Store. Signed data captured spatial coordinates, finger area, time stamp, accelerometer, and gyroscope signals. Nevertheless, some information concerning older acquisition devices was missing. Six sessions were taken apart into intervals (with varying delays), and each session consisted of eight snapshots in a new sequence. The tasks were designed to mitigate shoulder surfing attacks, incorporating both visible and invisible modes. There was a notable decrease in participant numbers across sessions, with a diverse range of ages and a higher prevalence of right-handed users. This dataset can be freely downloaded upon official request from https://github.com/BiDAlab/MobileTouchDB.

Table 1 summarises all the studies that have used these datasets, sorted by the year of publication to illustrate the research trends.

Table 1. Summary of findings.

Study	Classifiers	Performance	Dataset	Limitations
Montgomery *et al.*, 2019	Logistic regression SVM K-nearest neighbour Gaussian naïve Bayes VGG network DNN	Acc: 67%–95%	Frank dataset TouchTrack	Given the limited number of users, the high classification accuracy achieved with decision trees and DNN classifiers may indicate potential overlearning or overfitting.
Ooi & Teoh, 2019	T-RF	EER: 2.5%–4.0%	Frank database Serwadda database	The LSTM and GRU architectures may have encountered difficulties, possibly due to sample size constraints that hinder generalization in high-dimensional temporal models.
Tolosana, Vera-Rodriguez, & Fierrez, 2020	DTW RNN	EER: 3.80%	e-BioDigit Database	The system can only extract numerical digits (0–9) for authentication, which limits its effectiveness when passwords contain both uppercase and lowercase letters.

https://doi.org/10.18080/jtde.v12n4.1023

Study	Classifiers	Performance	Dataset	Limitations
Tolosana, Vera-Rodriguez, Fierrez *et al.*, 2020	DTW SW-DTW RNN TA-RNN	EER: 20%–29%	e-BioDigit database MobileTouch database	The MobileTouch database gathers data on finger contact area, accelerometer signals, and gyroscopic signals; however, the discriminative potential of these recordings was not fully used in this study.
Keykhaie & Pierre, 2020	NN L-SVM	EER: 11%–19%	Frank database Serwadda database	The resource constraints of smart cards necessitate model quantization and simplification, which may impact accuracy and limit scalability.
Acien *et al.*, 2021	SVM k-NN RBF	Acc: 80%–100%	HuMIdb database	Agnostic classification faces declining accuracy when classifiers trained on one type of synthetic sample are tested on a different type.
Bajaber *et al.*, n.d.	LSTM CNN GRU RNN	EER: 0%–4%	Bioident dataset Touchalytics dataset	Limited dataset availability and the lack of secure template storage limit versatility and data security.
Delgado-Santos *et al.*, 2024	SVM GMM	EER: 3.6%–11.30%	In-house database Frank dataset HuMIdb database	High computational requirements are associated with practical implementation and the need for additional testing to enhance stability.

Abbreviations: Acc, accuracy; CNN, convolutional NN; DNN, deep NN; DTW, dynamic time warping; EER, equal error rate; GMM, Gaussian mixture model; GRU, gated recurrent unit; k-NN, K-nearest neighbour; LSTM, long short-term memory; L-SVM, least-squares SVM; NN, neural network; RBF, radial basis function; RNN, recurrent NN; SVM, support vector machine; SW-DTW, subsequence DTW; TA-RNN, temporal attention–based RNN; T-RF, temporal random forest; VGG, Visual Geometry Group.

Feature Extraction

As previously discussed, feature extraction involves transforming raw data into numerical features that can be processed by machine learning algorithms while retaining the essential information from the original dataset (Feature Extraction Explained - MATLAB & Simulink, 2024; Khoh *et al.*, 2023). This process typically yields better outcomes than applying machine

https://doi.org/10.18080/jtde.v12n4.1023

learning directly to raw data. This paper applies two feature extraction techniques, the GMM and NN, to all three datasets.

Gaussian Mixture Model (GMM)

The GMM is a statistical model that suppresses a probability distribution as a sum of a few Gaussian functions with weighting (Ma *et al.*, 2021). This model is well-suited for tasks such as clustering and density estimation. Moreover, GMM also can be used for feature extraction. It can extract and visualise complex features from datasets. For example, the mean and covariance of the Gaussian component can be used as characteristics representing the statistics of the data distribution. These characteristics can be used for subsequent classification, clustering, or dimensionality reduction tasks. For instance, in visualisation tasks, tools like the Lasso tool allow users to mark complex target features across multiple slices of volumetric data, after which GMM can automatically extract and visualise these features in a three-dimensional view, aiding in deeper data exploration and analysis (Ma *et al.*, 2021).

Neural Network (NN)

NNs have also proven to be valuable tools for feature extraction in data preprocessing, as supported by various studies (Devi *et al.*, 2022; Garg, 2019). An NN typically consists of three layers: the input layer, the hidden layer, and the output layer. In the feature extraction process, raw data are first fed into the input layer. The hidden layer then performs a series of mathematical operations, adjusting internal weights based on the data. Through this learning process, the network identifies and extracts features that capture the underlying structure of the data. Finally, these features are passed through the output layer, producing a refined set of characteristics that can be used in further analysis or decision-making tasks. The ability of NNs to automatically learn and extract relevant features makes them particularly useful in complex data environments (Ooi *et al.*, 2017).

Proposed Methodology

This paper explores the XAI decision trees model on three touchstroke datasets: Frank, eBioDigit, and MobileTouch databases. Moreover, this paper has compared the equal error rate (EER) for the RF classifier and decision tree classifier. Before the data training process, feature extraction is conducted to prepare the datasets for analysis. The previous section discusses the feature extraction process in detail, including the techniques employed. In this experiment, we implemented GMM feature extraction on the Frank dataset. For the eBioDigit database, we used an NN for feature extraction, whereas the MobileTouch database used a combination of GMM and NN techniques. The following subsection explains the functioning

https://doi.org/10.18080/jtde.v12n4.1023

of the decision tree in this experiment and provides insights into the implementation of feature extraction in the coding process.

Explainable artificial intelligence (XAI)

The literature review section of this paper has already discussed the definition of XAI and its benefits. Building on this foundation, this study implements an XAI framework using decision tree models within the data training process, specifically employing RF and decision tree classifiers. Decision trees are simple and understandable models that make decisions based on rules (Ooi *et al.*, 2017). Each internal node reflects a judgment based on a single attribute, whereas each leaf node represents a classification label. Decision trees give a clear path for decision-making and are intrinsically explainable (*Explainable AI (XAI) with a Decision Tree | by Idit Cohen | Towards Data Science*, 2024). Figure 8 shows the sample design of XAI showing the decision tree.

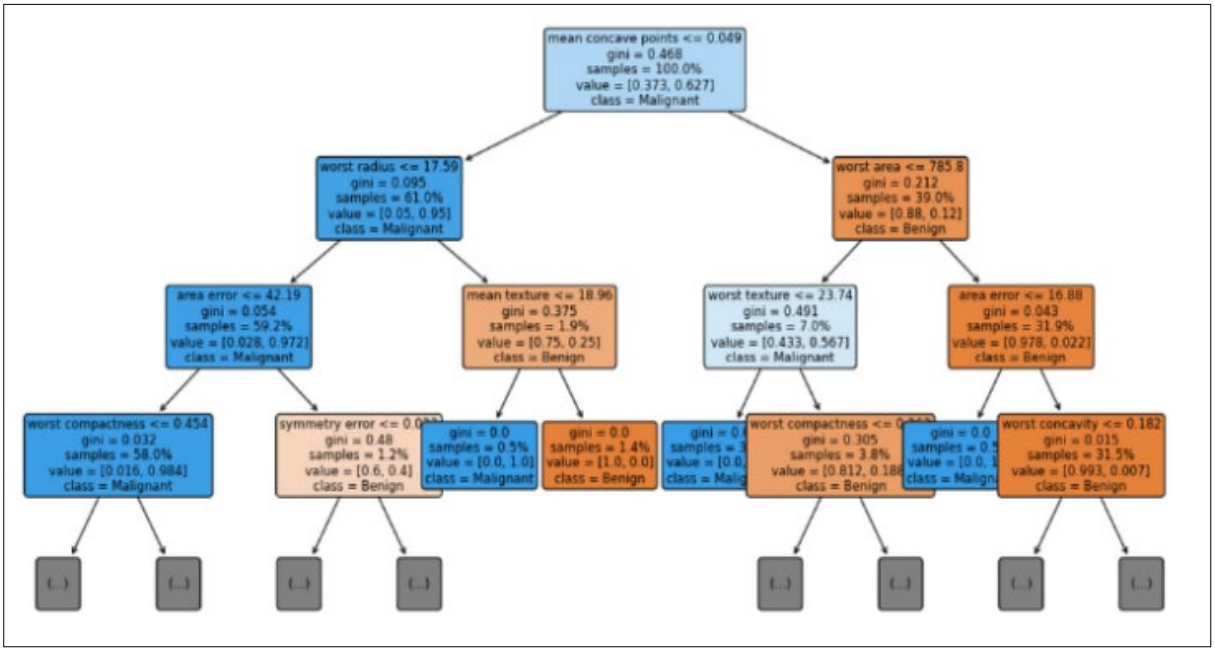


Figure 8. Sample XAI showing decision tree (Explainable AI (XAI) with a Decision Tree | by Idit Cohen | Towards Data Science, 2024).

Feature Extraction Implementation

GMM Feature Extraction

The GMM is implemented as the feature extraction technique for the Frank dataset. Figure 9 shows the code to apply a GMM through the Frank dataset and stored in a data frame (df). '$n\_components = 5$' is to fit the data with five Gaussian distributions (or clusters). The number of $n\_components$ could be filled out as your desired number. Besides that, '$gmm.fit(df)$' is used to fit the GMM to the data in the data frame. This step estimates the parameters of the

https://doi.org/10.18080/jtde.v12n4.1023

Gaussian distributions (mean and covariance) and their mixing coefficients using an expectation-maximisation algorithm.

Additionally, this code '$gmm\_features = gmm.predict\_proba(df)$' is used to predict the probability of each data point belonging to the Gaussian components. The last two-line codes loop over each Gaussian component and then extract the probability scores for the 'i' Gaussian component for all data points. These scores are then stored in columns named '$gmm\_feature\_i$'.

```
# Fit Gaussian Mixture Model
gmm = GaussianMixture(n_components=5, random_state=42)
gmm.fit(df)

# Add GMM features to dataframe
gmm_features = gmm.predict_proba(df)
for i in range(gmm.n_components):
    df[f'gmm_feature_{i}'] = gmm_features[:, i]
```

Figure 9. GMM feature extraction example coding implemented in Frank dataset.

NN Feature Extraction

An NN is used for feature extraction for the E-BioDigit database. The first part of the code is to create a feedforward NN model using Keras's 'Sequential' application programming interface, which consists of four dense layers with 64, 32, 16, and 8 neurons (Goh et al, 2024). Each dense layer is followed by the ReLU activation function, except for the last layer.

The second part of the code creates a new sequential model called '$feature\_extractor$' by removing the last layer from the original '$model$'. This last layer is removed because it does not have an activation function. The modified network is then used to generate the features, as shown in Figure 10.

```
# Define a neural network model
model = Sequential([
    Dense(64, activation='relu', input_shape=(X.shape[1],)),
    Dense(32, activation='relu'),
    Dense(16, activation='relu'),
    Dense(8, activation='relu')
])

# Extract features using the model
feature_extractor = Sequential(model.layers[:-1])
dnn_features = feature_extractor.predict(X)
```

Figure 10. Neural network feature extraction example coding implemented in E-BioDigit database.

https://doi.org/10.18080/jtde.v12n4.1023

Experimental Evaluation

Dataset Settings

This experiment aims to identify or classify users based on their interactions, such as touch patterns. In this case, the user identifier had to be our target variable, as shown in Figure 11.

```
# Extract features and target variable
X = data.drop(columns=['user ID'])
y = data['user ID']
```

Figure 11. User identifier as target variable.

Moreover, Figure 12 shows the code to split the dataset into a training set and testing set, and this code is applied consistently across all classifier training models. Additionally, this split can be generated every time the code is run. After the split, the variables obtained are as follows:

- *X_train* (features for training)
- *X_test* (features for testing)
- *y_train* (target variable for training)
- *y_train* (target variable for testing)

```
# Split the dataset into the training set and test set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Figure 12. Code to split the dataset into training and testing.

Experiment Result

Frank dataset, e-BioDigit database, and MobileTouch database were used in this paper. In this case, Table 2 shows the EER results for each classifier based on the three datasets. The EER for an RF classifier based on the three datasets is around 0.03%—0.05%. For a decision tree classifier, the EER is around 0.02%—0.07%.

Table 2. Table comparison result between RF and DT.

Dataset	Classifier	EER	FAR	FRR
Frank dataset, 12 features, GMM feature extraction	RF	0.05%	0.00%	0.10%
	DT	0.07%	0.00%	0.13%
E-BioDigit database, 5 features, NN feature extraction	RF	0.05%	0.00%	0.09%
	DT	0.02%	0.00%	0.05%
MobileTouch database, 21 features, GMM + NN feature extraction	RF	0.03%	0.00%	0.05%
	DT	0.02%	0.00%	0.05%

Abbreviations: DT, decision tree; EER, equal error rate; FAR, false acceptance rate; FRR, false rejection rate; GMM, Gaussian mixture model; NN, neural network; RF, random forest.

https://doi.org/10.18080/jtde.v12n4.1023

XAI Decision Tree Model Output

This section discusses the XAI framework and visualises the results for both the RF and decision tree classifiers. However, due to the large dataset size, it is not feasible to display the full decision trees. The detailed results and explanation for each dataset are discussed in the subsequent subsection.

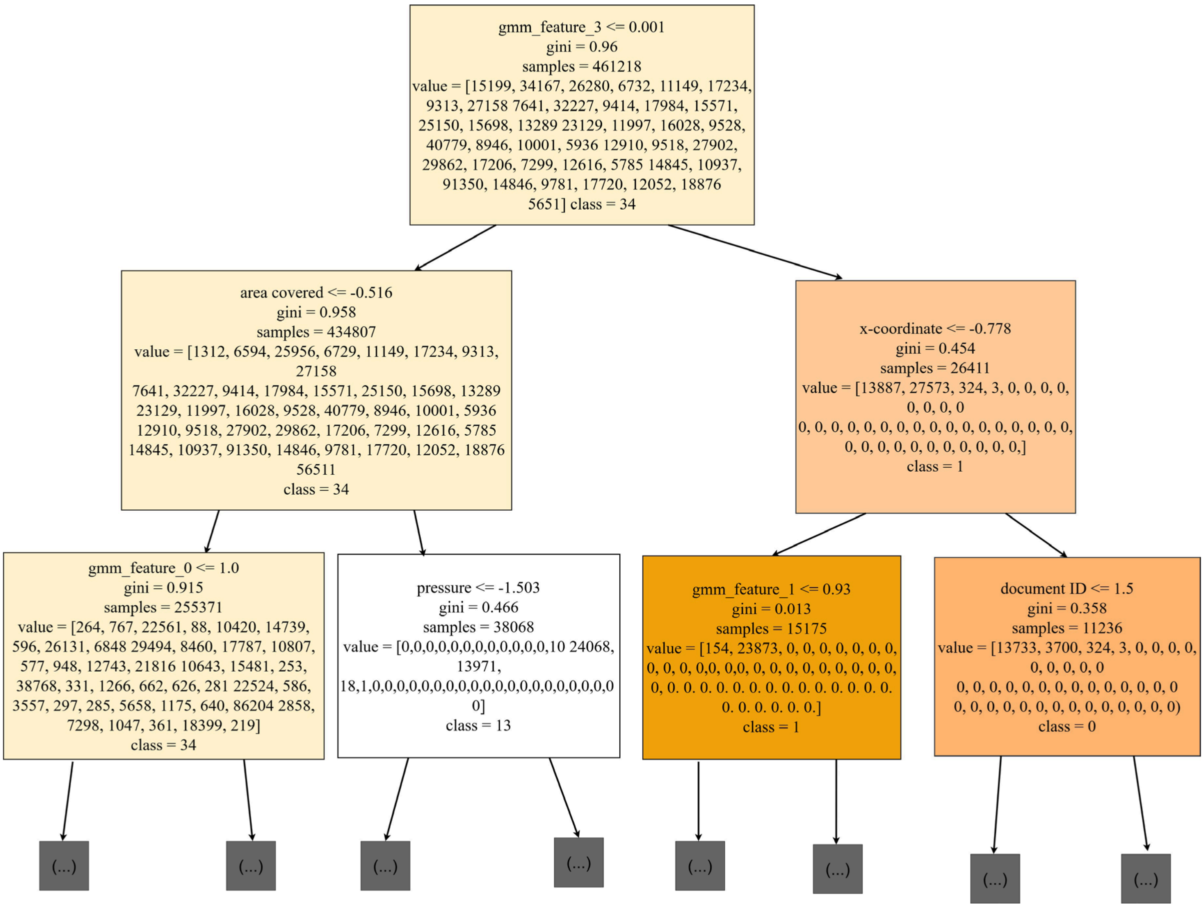


Figure 13. XAI decision tree based on the Frank dataset by implementing random forest classifier.

Frank Dataset XAI Decision Tree

Figure 13 and Figure 14 display the XAI decision tree output for the Frank dataset. However, both are implemented with different classifiers; Figure 13 presents the tree generated by an RF classifier, whereas Figure 14 shows the tree from a decision tree classifier. Both of them have the same decision tree output. In both figures, we can observe that the root node submits a judgment concerning whether “$gmm\_feature\_3$” is less than or equal to 0.001. Besides that, the root node has a Gini index of 0.96, and the class is 34 according to the value array, providing the distribution of samples from different classes.

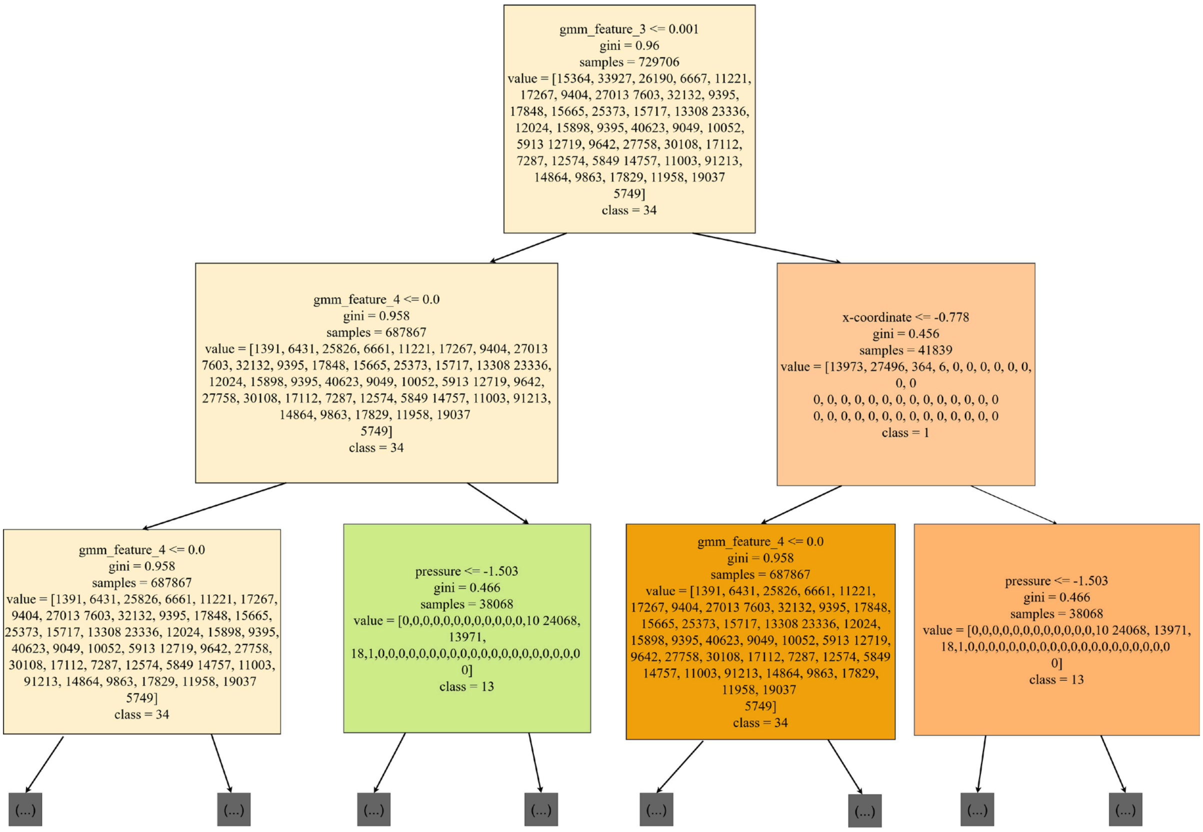


Figure 14. XAI decision tree based on the Frank dataset by implementing decision tree classifier.

E-BioDigit Database XAI Decision Tree

Figure 15 and Figure 16 illustrate the XAI decision tree output for the E-BioDigit database, each generated using different classifiers. Figure 15 shows the tree created by an RF classifier, whereas Figure 16 displays the tree generated by a decision tree classifier. From Figure 15, we can see that it starts with a condition on the variable y and splits based on whether y is less than or equal to 0.653. Besides that, it also shows the Gini index, the number of samples, and the distribution of values for each class at that node. On the other hand, Figure 16 starts with a different variable, x. After progressing to the next leaf, the variable time is evaluated.

MobileTouch Database XAI Decision Tree

Figure 17 and Figure 18 show the XAI decision tree output based on the MobileTouch database. Similar to the previous two datasets, Figure 17 shows the tree created by an RF classifier, whereas Figure 18 presents the tree generated by a decision tree classifier. Comparing both figures, it can be observed that both models have started with different features at the root node. Figure 17 starts with the "$dnn\_feature\_2$" feature, and Figure 18 starts with the "$area\_covered$" feature.

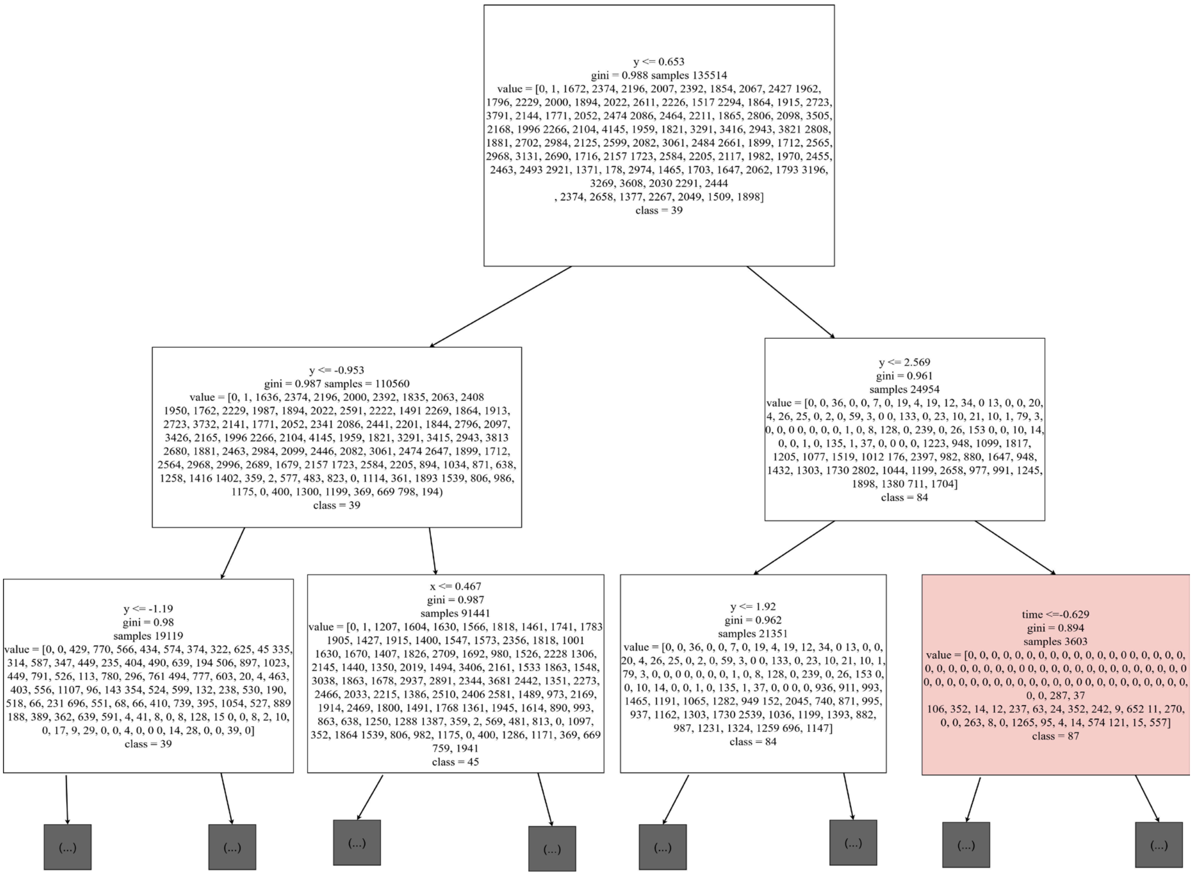


Figure 15. XAI decision tree based on E-BioDigit database by implementing random forest classifier.

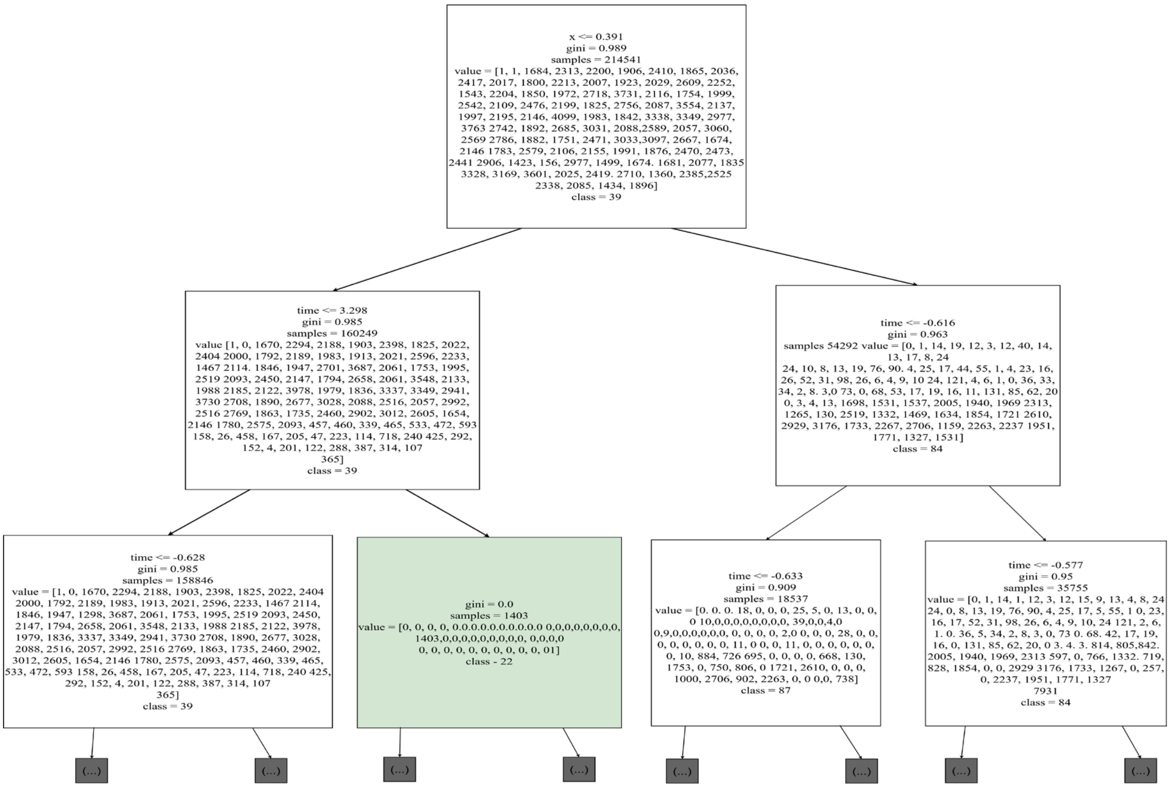


Figure 16. XAI decision tree based on E-BioDigit database by implementing decision tree classifier.

https://doi.org/10.18080/jtde.v12n4.1023

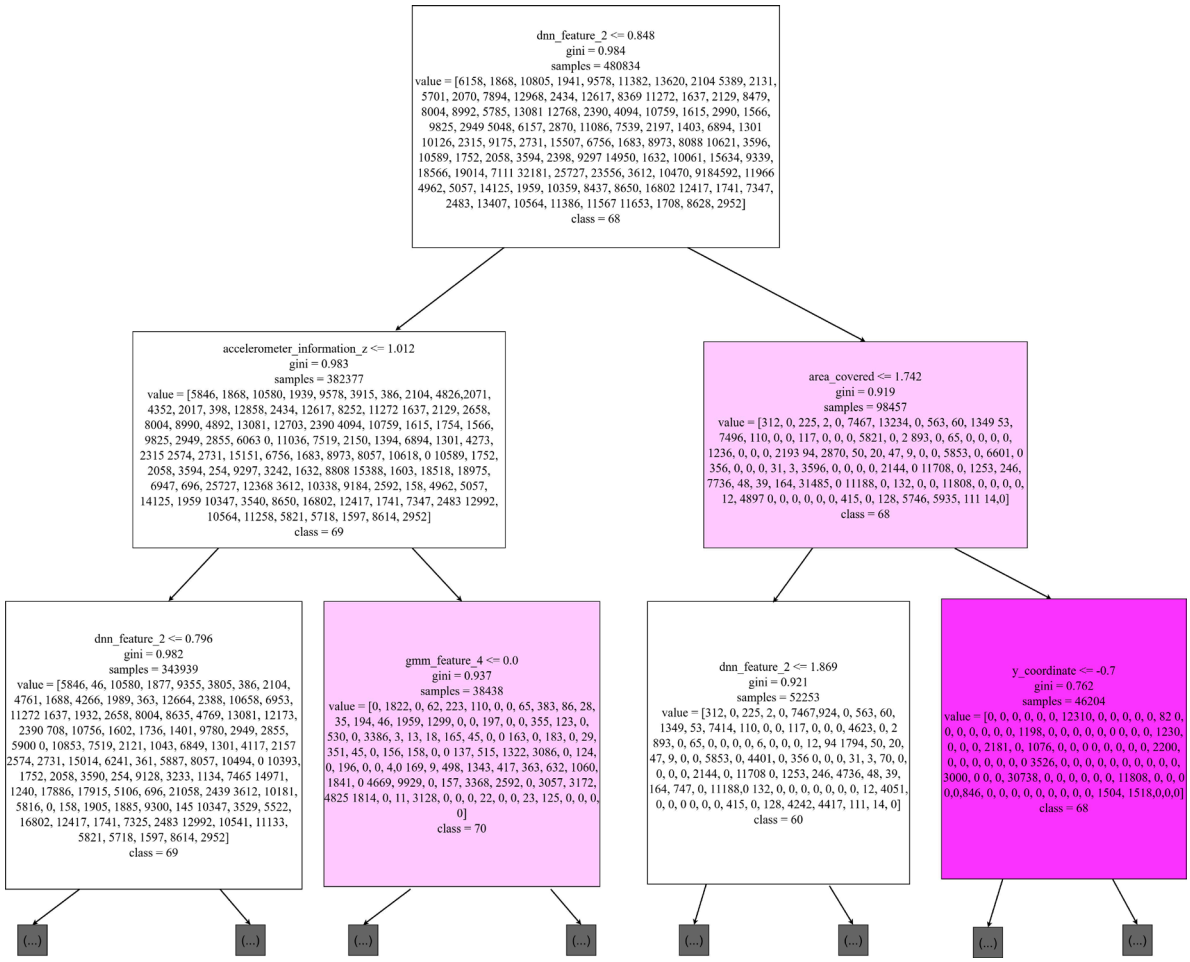


Figure 17. XAI decision tree based on MobileTouch database by implementing random forest classifier.

Conclusions

In conclusion, this paper implements two feature extraction techniques, GMM and NN, in the data preprocessing stage. The performance of two classifiers, RF and decision tree, is compared in terms of EER on the Frank dataset, the E-BioDigit database, and the MobileTouch database. Experimental results demonstrate that both classifiers achieve excellent empirical outcomes, with EERs ranging from 0.02% to 0.07%. Additionally, XAI technology is used to provide in-depth insights into the classifiers' decision-making processes, enhancing interpretability and enabling users to better understand how decisions are made. This approach substantiated the importance of both classification performance and interpretability in developing robust biometric authentication systems.

Acknowledgments

This research was funded by a Matching Grant from Multimedia University and the University of Hertfordshire (MMUI/230078).

https://doi.org/10.18080/jtde.v12n4.1023

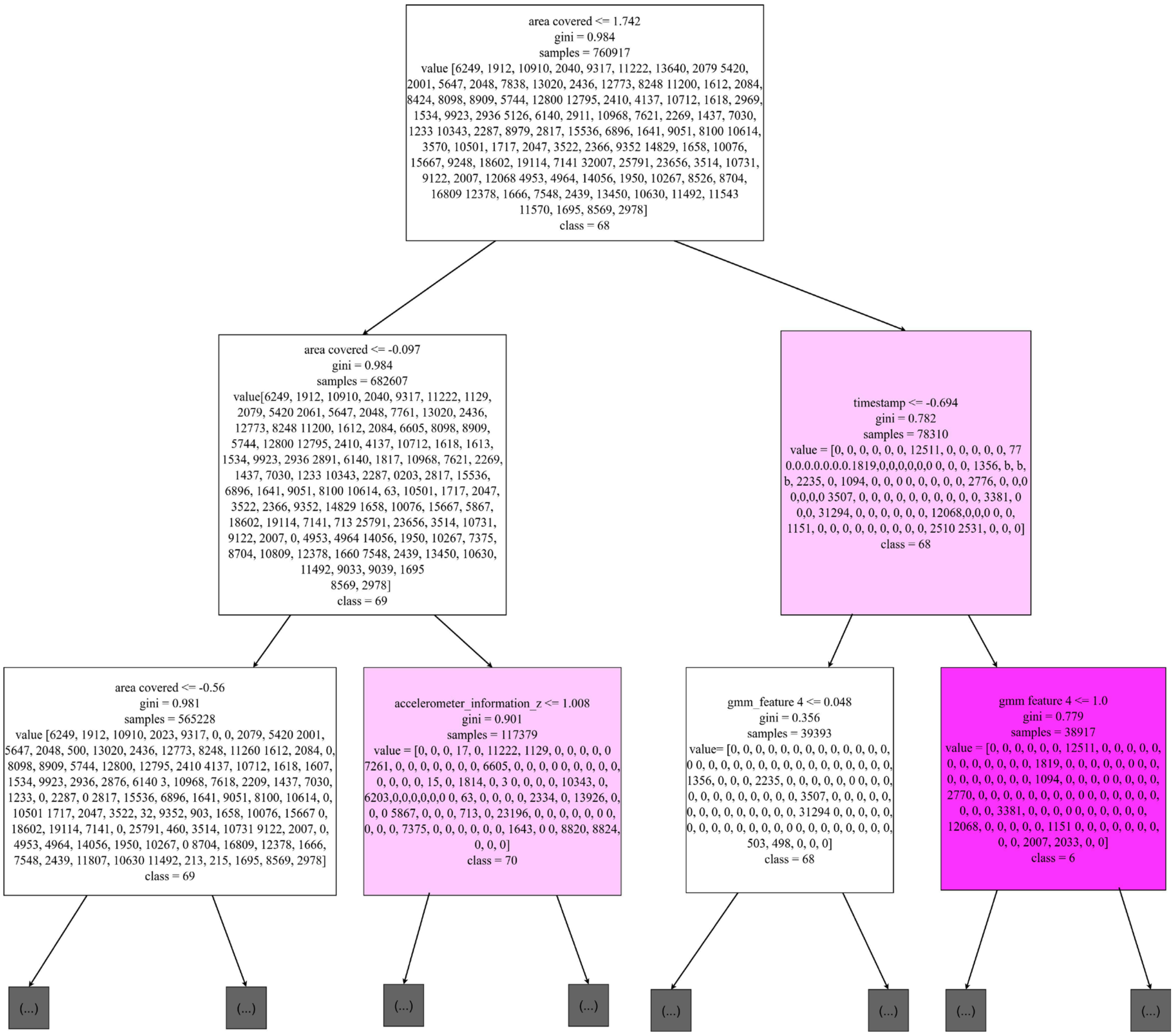


Figure 18. XAI decision tree based on MobileTouch database by implementing decision tree classifier.

References

Acien, A., Morales, A., Fierrez, J., Vera-Rodriguez, R., & Delgado-Mohatar, O. (2021). BeCAPTCHA: Behavioral bot detection using touchscreen and mobile sensors benchmarked on HuMIdb. *Engineering Applications of Artificial Intelligence*, *98*, 104058. https://doi.org/10.1016/J.ENGAPPAI.2020.104058

Alwahaishi, S., & Zdralek, J. (2020). Biometric Authentication Security: An Overview. *Proceedings - 2020 IEEE International Conference on Cloud Computing in Emerging Markets, CCEM 2020*, 87–91. https://doi.org/10.1109/CCEM50674.2020.00027

Bajaber, A., Fadel, M., & Elrefaei, L. (n.d.). *Evaluation of Deep Learning Models for Person Authentication Based on Touch Gesture.* https://doi.org/10.32604/csse.2022.022003

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Bello, A. A., Chiroma, H., Gital, A. Y., Gabralla, L. A., Abdulhamid, S. M., & Shuib, L. (2020). Machine learning algorithms for improving security on touch screen devices: A survey, challenges and new perspectives. *Neural Computing & Applications*, *32*(17), 13651–13678. https://doi.org/10.1007/S00521-020-04775-0/FIGURES/9

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, *408*, 189–215. https://doi.org/10.1016/J.NEUCOM.2019.10.118

Chapter 7. Learning (II): SVM & Ensemble Learning | Data Analytics: A Small Data Approach. (n.d.). Retrieved 1 November 2024, from https://dataanalyticsbook.info/chapter-7.-learning-ii-svm-ensemble-learning.html

Classifier Definition | DeepAI. (n.d.). Retrieved 28 November 2023, from https://deepai.org/machine-learning-glossary-and-terms/classifier

Decision Tree Tutorials & Notes | Machine Learning | HackerEarth. (n.d.). Retrieved 1 November 2024, from https://www.hackerearth.com/practice/machine-learning/machine-learning-algorithms/ml-decision-tree/tutorial/

Decision Trees - Graphical Explainable AI • Aggregata. (n.d.). Retrieved 19 April 2024, from https://aggregata.de/en/blog/supervised-learning/decision-trees/

Deep neural network architecture | Download Scientific Diagram. (n.d.). Retrieved 1 November 2024, from https://www.researchgate.net/figure/Deep-neural-network-architecture_fig2_347813247

Delgado-Santos, P., Tolosana, R., Guest, R., Lamb, P., Khmelnitsky, A., Coughlan, C., & Fierrez, J. (2024). SwipeFormer: Transformers for mobile touchscreen biometrics. *Expert Systems with Applications*, *237*, 121537. https://doi.org/10.1016/J.ESWA.2023.121537

Devi, N. B., Kavida, A. C., & Murugan, R. (2022). Feature Extraction and Object Detection Using Fast-Convolutional Neural Network for Remote Sensing Satellite Image. *Journal of the Indian Society of Remote Sensing*, *50*(6), 961–973. https://doi.org/10.1007/S12524-022-01506-X/TABLES/2

Explainable AI | Royal Society. (n.d.). Retrieved 22 March 2024, from https://royalsociety.org/news-resources/projects/explainable-ai/

Explainable AI (XAI) with a Decision Tree | by Idit Cohen | Towards Data Science. (n.d.). Retrieved 3 April 2024, from https://towardsdatascience.com/explainable-ai-xai-with-a-decision-tree-960d60b240bd

Feature Extraction Explained - MATLAB & Simulink. (n.d.). Retrieved 26 March 2024, from https://www.mathworks.com/discovery/feature-extraction.html

Garg, S. (2019). *Face Recognition System: A Review*. Proceedings of DHE Sponsored 1 Day National Seminar on Recent Advancement in IT & E-Commerce: Present Scenario & Future Prospects RAITECOM-2019

Gavisiddappa, G., Mahadevappa, S., & Mohan Patil, C. (2020). Multimodal Biometric Authentication System Using Modified ReliefF Feature Selection and Multi Support

Vector Machine. *International Journal of Intelligent Engineering and Systems*, *13*(1). https://doi.org/10.22266/ijies2020.0229.01

Goh, T.-J., Chong, L.-Y., Chong, S.-C., & Goh, P.-Y. (2024). A Campus-based Chatbot System using Natural Language Processing and Neural Network. *Journal of Informatics and Web Engineering*, *3*(1), 96–116. https://doi.org/10.33093/jiwe.2024.3.1.7

Keykhaie, S., & Pierre, S. (2020). Mobile Match on Card Active Authentication Using Touchscreen Biometric. *IEEE Transactions on Consumer Electronics*, *66*(4), 376–385. https://doi.org/10.1109/TCE.2020.3029955

Khoh, W. H., Pang, Y. H., Ooi, S. Y., Wang, L.-Y.-K., & Poh, Q. W. (2023). Predictive Churn Modeling for Sustainable Business In The Telecommunication Industry: Optimized Weighted Ensemble Machine Learning. *Sustainability*, *15*(11), 8631. https://doi.org/10.3390/su15118631

Liu, S., Shao, W., Li, T., Xu, W., & Song, L. (2022). Recent advances in biometrics-based user authentication for wearable devices: A contemporary survey. *Digital Signal Processing*, *125*, 103120. https://doi.org/10.1016/J.DSP.2021.103120

Ma, J., Chen, J., Chen, L., Zhou, X., Qin, X., Tang, Y., Sun, G., & Chen, J. (2021). Gaussian mixture model-based target feature extraction and visualization. *Journal of Visualization*, *24*(3), 545–563. https://doi.org/10.1007/S12650-020-00724-0/TABLES/4

Miraoui, M., & El-Etriby, S. (2019). A context-aware authentication approach for smartphones. *2019 International Conference on Computer and Information Sciences, ICCIS 2019*. https://doi.org/10.1109/ICCISCI.2019.8716453

Montgomery, M., Chatterjee, P., Jenkins, J., & Roy, K. (2019). Touch Analysis: An Empirical Evaluation of Machine Learning Classification Algorithms on Touch Data. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11611 LNCS*, 147–156. https://doi.org/10.1007/978-3-030-24907-6_12/FIGURES/6

Neural Networks with R. (n.d.). Retrieved 24 March 2024, from https://subscription.packtpub.com/book/data/9781788397872/1/ch01lvl1sec27/pros-and-cons-of-neural-networks

Ooi, S. Y., Tan, S. C. & Cheah, W. P. (2017) Temporal Sampling Forest (TS-F): An Ensemble Temporal Learner. *Soft Computing*, *21*, 7039–7052. https://doi.org/10.1007/s00500-016-2242-7

Ooi, S. Y., Tan, S. C., & Cheah, W. P. (2018). Temporal Sleuth Machine with Decision Tree for Temporal Classification. *Soft Computing*, *22*(24), 8077–8095. https://doi.org/10.1007/s00500-017-2747-8

Ooi, S. Y., & Teoh, A. B. J. (2019). Touch-Stroke Dynamics Authentication Using Temporal Regression Forest. *IEEE Signal Processing Letters*, *26*(7), 1001–1005. https://doi.org/10.1109/LSP.2019.2916420

Qin, D., Amariucai, G.T., Qiao, D., & Guan, Y. (2024). Improving behavior based authentication against adversarial attack using XAI. ArXiv. https://arxiv.org/abs/2402.16430 .

Random Forest Algorithm for Absolute Beginners in Data Science. (n.d.). Retrieved 24 March 2024, from https://www.analyticsvidhya.com/blog/2021/10/an-introduction-to-random-forest-algorithm-for-beginners/

Random forest Algorithm in Machine learning | Great Learning. (n.d.). Retrieved 24 March 2024, from https://www.mygreatlearning.com/blog/random-forest-algorithm/

Random Forest in Machine Learning. (n.d.). Retrieved 1 November 2024, from https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/

Rawal, A., McCoy, J., Rawat, D. B., Sadler, B. M., & Amant, R. S. (2022). Recent Advances in Trustworthy Explainable Artificial Intelligence: Status, Challenges, and Perspectives. *IEEE Transactions on Artificial Intelligence*, *3*(6), 852–866. https://doi.org/10.1109/TAI.2021.3133846

Seek, C. Y., Ooi, S. Y., Pang, Y. H., Lew, S. L., & Heng, X. Y. (2023). Elderly and Smartphone Apps: Case Study with Lightweight MySejahtera. *Journal of Informatics and Web Engineering*, *2*(1), 13–24. https://doi.org/10.33093/jiwe.2023.2.1.2

Spaling, M. M., & Singh Uppal, A. (2021). *Multi-factor authentication in network security* [Master's thesis, University of Alberta]. https://doi.org/10.7939/r3-ftat-7h78 .

Tolosana, R., Vera-Rodriguez, R., & Fierrez, J. (2020). BioTouchPass: Handwritten Passwords for Touchscreen Biometrics. *IEEE Transactions on Mobile Computing*, *19*(7), 1532–1543. https://doi.org/10.1109/TMC.2019.2911506

Tolosana, R., Vera-Rodriguez, R., Fierrez, J., & Ortega-Garcia, J. (2020). BioTouchPass2: Touchscreen Password Biometrics Using Time-Aligned Recurrent Neural Networks. *IEEE Transactions on Information Forensics and Security*, *15*, 2616–2628. https://doi.org/10.1109/TIFS.2020.2973832

Top 4 advantages and disadvantages of Support Vector Machine or SVM | by Dhiraj K | Medium. (n.d.). Retrieved 23 March 2024, from https://dhirajkumarblog.medium.com/top-4-advantages-and-disadvantages-of-support-vector-machine-or-svm-a3c06a2b107

Types of explainable AI. (n.d.). Retrieved 19 April 2024, from https://courses.minnalearn.com/en/courses/trustworthy-ai/preview/explainability/types-of-explainable-ai/

Voege, P., Abu Sulayman, I. I. M., & Ouda, A. (2022). Smart Chatbot for User Authentication. *Electronics 2022, 11*(23), 4016. https://doi.org/10.3390/ELECTRONICS11234016

Voege, P., & Ouda, A. (2022). An Innovative Multi-Factor Authentication Approach. *2022 International Symposium on Networks, Computers and Communications, ISNCC 2022*. https://doi.org/10.1109/ISNCC55209.2022.9851710

Wang, C., Wang, Y., Chen, Y., Liu, H., & Liu, J. (2020). User authentication on mobile devices: Approaches, threats and trends. *Computer Networks*, *170*, 107118. https://doi.org/10.1016/j.comnet.2020.107118

Wang, Z., Chen, F., Zhou, N., Ma, M., Li, X., Guo, Y., & Chen, D. (2021). Identity authentication based on dynamic touch behavior on smartphone. *2021 6th International Conference on Image, Vision and Computing, ICIVC 2021*, 469–474. https://doi.org/10.1109/ICIVC52351.2021.9527023

Book Review

Data Rules: Reinventing the Market Economy

by Cristina Alaimo and Jannis Kallinikos

Rob Nicholls
University of Sydney and UTS Law

Abstract: *Data Rules: Reinventing the Market Economy* by Cristina Alaimo and Jannis Kallinikos offers an exploration of the societal, cultural, and epistemological dimensions of data, challenging conventional views of data as neutral or merely technical. By introducing original concepts such as 'data rules' and 'data complementarities', the authors provide a unique perspective on how data reshapes markets, social relationships, and innovation ecosystems.

The book balances historical context, theoretical insight, and contemporary examples, making it a useful resource for interdisciplinary readers. The book leaves gaps in addressing actionable regulatory frameworks or the political dimensions of platform capitalism. The volume represents a useful 'rounding out' of the literature in this specialist field.

Keywords: cultural data, data complementarities, data rules, epistemological dimensions of data

Introduction

Data Rules presents an exploration of the social, cultural, and epistemological dimensions of data, moving beyond the purely technical framing often seen in discussions of the data revolution. The book challenges readers to see data not merely as neutral tools or passive resources but as powerful forces reshaping social and economic relationships.

Book details: Data Rules: Reinventing the Market Economy, by Cristina Alaimo and Jannis Kallinikos. MIT Press. 238 pp. ISBN 978-0262547932. Published 9 July 2024. Available open access at https://direct.mit.edu/books/oa-monograph/5792/Data-Rules Reinventing-the-Market-Economy. Paperback: $US45 (about $82 online).

https://doi.org/10.18080/jtde.v12n4.1160

The Book

Chapter 1

Chapter 1 is called “Introduction: Data and Socioeconomic Transformations” and introduces the book’s main objective. This is to analyse the forces that place data at the heart of modern life and their impact on the economy and society. The authors claim that this objective departs from typical discussions of the data revolution, which often focus on data as a purely technical phenomenon. Instead, the chapter, building on ideas from our conversation history, emphasises the social, cultural, and epistemological dimensions of data and their impact on various aspects of contemporary life.

The authors acknowledge that digital data, due to its inherent qualities of reproducibility and malleability, has expanded the scope and impact of data practices that have long existed in society. However, the chapter cautions against viewing digital data solely as a technological artifact. It argues that understanding the social and economic implications of digital data requires recognising its complex relationship with knowledge, communication, and existing social institutions.

The authors suggest that data rules, alongside market and design rules, are fundamentally reshaping the economy by influencing the creation of new products and services. The emergence of digital platforms and ecosystems is presented as a key example of this transformation, with platforms acting as vehicles for challenging the traditional separation between transactions and social relationships. The chapter argues that understanding data rules is crucial for understanding the broader socioeconomic transformations brought about by the increasing prevalence of digital data.

The chapter concludes by outlining the structure of the book.

Chapter 2

Chapter 2 is entitled “The Epistemic Foundations of Data” and traces the history of data and how it has been linked to cognition and social action. The chapter explores the evolution of data from ancient record-keeping practices to the digital age, examining the multiple ties data have maintained with social practices and institutions.

One important historical example the chapter discusses is the use of clay tokens in ancient Mesopotamia. These served as a form of record-keeping and accounting, demonstrating the long-standing link between data and social organisation. The chapter also highlights the rise of modern statistical practices in the 19th century. This period saw an increased emphasis on measurement and the use of data to understand social phenomena.

https://doi.org/10.18080/jtde.v12n4.1160

The chapter argues that the advent of digital data tends to obscure the historical functions and practices associated with data. The chapter concludes by arguing that understanding the 'politics of data', or how data are used to shape reality and influence social action, is essential for understanding the contemporary data-driven world.

Chapter 3

Chapter 3, "The Digital Data Revolution", focuses on how mechanisation and digitisation have transformed the very nature and function of data. The chapter argues that while data have always been social and cultural artefacts, the advent of digital technologies has led to data being increasingly perceived as purely technical items, obscuring their rich history and complex social dimensions.

The authors argue that mechanical tabulating machines marked a shift towards the formalisation and de-contextualisation of data processing and that this accelerated with the rise of digital computers. Digital data became detached from specific contexts and could be easily manipulated and transmitted across various domains. This process, while enabling new possibilities for data analysis and use, also led to a tendency to view data as raw, objective, and context-free.

The chapter argues that this narrow perception of digital data overlooks their continued role as knowledge artefacts. Instead, data are shaped by social practices, institutional contexts, and the choices made during their production and interpretation. The chapter concludes by advocating for a more nuanced understanding of digital data that recognises both the continuities and discontinuities introduced by digital technologies.

Chapter 4

The fourth chapter has the title of "The Data Life Cycle". It examines the process by which data are created, standardised and aggregated into objects that take on novel social and economic lives. This chapter analyses the decisions made in data production, arguing that these choices shape the meaning and use of data.

The authors emphasise that data are not simply 'out there' but are actively 'made' through a series of choices and classifications. Data production involves selecting and defining what counts as a relevant event or property, and this selection process reflects the social and organisational contexts in which the data is produced; examples are provided.

The authors introduce the concept of 'data objects', which they describe as recurring arrangements of data that serve specific knowledge-making and use purposes, and they

https://doi.org/10.18080/jtde.v12n4.1160

provide examples. These data objects enable new knowledge and organisational processes around virtual representations of machines, such as digital twins.

The chapter concludes by arguing that understanding the data life cycle and the choices made in data production is crucial for understanding the social and economic implications of data. The authors suggest that the value of data is not inherent, but rather emerges through processes of negotiation, legitimisation, and embedding in market mechanisms.

Chapter 5

Chapter 5 is called "Technologies of Difference: Excursus on Surveillance" and examines the debate surrounding data regulation and concerns about surveillance in the digital age. The chapter argues that simply regulating data as technical elements is insufficient and overlooks the complex roles data play in society. Instead, the authors advocate for understanding how data intertwine with innovation, value creation, and societal wellbeing to develop more effective regulation strategies.

The authors argue that data are not neutral representations of reality but active forces shaping personal and institutional relationships. This perspective challenges the notion of a clear separation between data tracking and the broader social practices that generate data.

The authors argue against viewing data as purely technical or simply as commodities to be traded. They emphasise the need to recognise data's diverse functions, including their role in forming new types of markets and mediating social and economic interactions on platforms. In contrast to Zuboff (2018), the authors suggest that the focus on surveillance as a primary concern often obscures the broader socioeconomic implications of data-driven platforms and ecosystems.

The chapter concludes by calling for a more nuanced approach to data regulation that recognises the multifaceted nature of data and their role in shaping social and economic relations. It suggests that effective regulation requires a deeper understanding of how data are implicated in processes of innovation, knowledge creation, and value generation within the digital economy.

Chapter 6

Chapter 6 is "Decentering Organizations: Data, Knowledge, and Institutional Change". It examines how the increasing prevalence and use of data is transforming the nature of organisations and challenging traditional theories of the firm. The chapter argues that the influx of external data, coupled with the ability to repurpose data across domains, is leading

https://doi.org/10.18080/jtde.v12n4.1160

to a “decentering” of organisations. This means that organisations are more reliant on data and knowledge from external sources.

The chapter examines the implications of these developments for organisational knowledge and capability development. The authors argue that the ability to repurpose data across domains is leading to a more fluid and dynamic understanding of knowledge.

The authors argue that organisations are increasingly engaging in more open and collaborative forms of innovation and value creation, relying on external data and expertise to complement their internal capabilities.

The chapter concludes by discussing the implications of these developments for the future of organisations. The authors suggest that the decentring of organisations is likely to lead to the emergence of new organisational forms, such as platforms and ecosystems, which are better suited to managing the complexities of a data-driven world.

Chapter 7

Chapter 7 is called “Platforms and Ecosystems” and examines the growing research on digital platforms and digital business ecosystems.

The authors discuss the management and innovation literature, which understands platforms as product configurations with a stable core and variable peripheral components. Next, the chapter explores the concept of ‘multisided platforms’, which are exchange systems marked by the dynamics of network effects and the ability to manage large user populations. These platforms often blur the boundaries between traditional market mechanisms and organisational structures. The chapter argues that these platforms heavily rely on data to manage user participation, create value, and coordinate interactions. Finally, the chapter examines research on digital business ecosystems, which are cross-industry networks of loosely coupled organisations. It contrasts the emphasis on ‘complementarities’ in ecosystem literature with the focus on ‘network effects’ in platform literature. The chapter suggests that the distinction between these concepts becomes blurred in the context of data-driven platforms and ecosystems, where value creation arises from the complex interplay of user interactions, data flows, and technological capabilities.

The chapter concludes by highlighting the lack of attention to data and data technologies in mainstream research on platforms and ecosystems.

Chapter 8

Chapter 8, “Data and Ecosystems”, extends the arguments made in previous chapters, claiming that the dynamics of digital business ecosystems are significantly shaped by what the

https://doi.org/10.18080/jtde.v12n4.1160

authors call ‘data rules’. The chapter posits that data rules, alongside market rules and design rules, play a pivotal role in shaping the structure and behaviour of ecosystems.

The authors emphasise that understanding the relationships between actors in an ecosystem requires moving beyond traditional economic notions of complementarity. The authors introduce the concept of ‘data complementarities’, arguing that data can create value-reinforcing synergies that go beyond the traditional distinctions between generic and specific resources. This concept builds upon the previous discussions about the malleability and repurposability of data objects, highlighting how data can be used in unforeseen ways to connect actors and resources across different domains.

The authors provide several examples to illustrate the concept of data complementarities. The chapter also discusses the role of data objects in shaping ecosystem dynamics. The authors argue that data objects serve as crucial building blocks for ecosystems. The minimalist and adaptable nature of data objects allows them to represent a wide range of entities and facilitate interactions within complex and constantly evolving ecosystems.

The authors conclude by advocating for a shift in the regulatory approach to platforms and ecosystems. They argue that current regulatory frameworks often focus too narrowly on algorithmic systems and fail to recognise the intricate relationships between data, algorithms, and ecosystem dynamics.

Chapter 9

The book ends with a chapter called “Epilogue”. This summarises the book’s key arguments and explores their implications for understanding the transformative impact of data on the market economy and society. The chapter echoes a consistent theme: data are not simply neutral tools or passive resources; they are powerful forces that are reshaping social and economic relations in profound ways.

The authors critique the limited attention given to data in existing research on platforms and ecosystems. They argue that much of this research focuses too narrowly on business models and economic outcomes, failing to grasp the broader societal implications of data-driven technologies. They call for a ‘social science of data’ that moves beyond the technical aspects of data science and examines how data are implicated in processes of knowledge production, social control, and cultural change.

https://doi.org/10.18080/jtde.v12n4.1160

Contributions and Readership

Contributions

From the perspective of readers of the *Journal*, the key contribution of *Data Rules* is its exploration of the social, cultural, and epistemological dimensions of data, moving beyond the purely technical framing often seen in discussions of the data revolution. The book challenges readers to see data not merely as neutral tools or passive resources but as powerful forces reshaping social and economic relationships.

However, in doing so, *Data Rules* effectively forecloses addressing some of the issues associated with the regulation of platforms and businesses with network effects. The call for a 'social science of data', which addresses social, cultural, and political implications of data-driven technologies, seems to miss the political implications of platforms and networks.

Target readers

The book challenges readers to move beyond simplistic understandings of data as purely technical entities and engage with the broader social, cultural, and epistemological implications of the data revolution. As such, it may justify a place on a bookshelf. However, I think that I would want Zuboff (2018) and the following books nearby:

- *The Tech Coup: How to Save Democracy from Silicon Valley* by Marietje Schaake (Schaake, 2024);
- *Feeding the Machine: The Hidden Human Labour Powering AI* by James Muldoon, Mark Graham & Callum Cant (Muldoon *et al.*, 2024);
- *Technofeudalism: What Killed Capitalism* by Yanis Varoufakis (Varoufakis, 2024);
- *Broken Code: Inside Facebook and the Fight to Expose Its Toxic Secrets* by Jeff Horwitz (Horwitz, 2023);
- *How Big-Tech Barons Smash Innovation—and How to Strike Back* by Ariel Ezrachi & Maurice E. Stucke (Ezrachi & Stucke, 2022);
- *Big Tech and the Digital Economy: The Moligopoly Scenario* by Nicolas Petit (Petit, 2020).

Bibliography

Ezrachi, A., & Stucke, M. E. (2022). *How Big-Tech Barons Smash Innovation—And How To Strike Back* (Standard Edition). HarperCollins US.

Horwitz, J. (2023). *Broken Code: Inside Facebook and the fight to expose its toxic secrets*. Transworld Digital.

Muldoon, J., Graham, M., & Cant, C. (2024). *Feeding the Machine: The Hidden Human Labour Powering AI* (Export edition). Canongate Trade.

https://doi.org/10.18080/jtde.v12n4.1160

Petit, N. (2020). *Big Tech and the Digital Economy The Moligopoly Scenario*. Oxford University Press UK.

Schaake, M. (2024). *The Tech Coup: How to Save Democracy from Silicon Valley*. Princeton University Press.

Varoufakis, Y. (2024). *Technofeudalism: What Killed Capitalism*. Vintage.

Zuboff, S. (2018). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (First edition). PublicAffairs.

https://doi.org/10.18080/jtde.v12n4.1160

William Webb's Contrarian Thesis

A Book Review of "The End of Telecoms History"

Jim Holmes
Vice President, TelSoc and Director, Incyte Consulting

Abstract: William Webb's new book, *The End of Telecom's History*, argues that 5G has failed to deliver on its promises and, further, that this is largely because increased data rates and data capacity are approaching sufficiency in the developed world. Thus, the basis of the 5G business case has not materialised. In this book, Web draws an analogy with the end of history generally, meaning the resolution of challenges that have defined an epoch. The implications of his thesis, in terms of industry investment and structure, and the impacts on all stakeholders, are set out. Webb's conclusions are contentious, and there are many in the industry who challenge his basic arguments. Some of the responses are mentioned in this review.

Keywords: William Webb, broadband, 5G, data rates, telecommunications industry

Introduction

Professor William Webb is an independent consultant based in the United Kingdom, who provides advice "across all telecommunications matters" (Webb, 2024, cover bio). He is also a prolific author. According to the biographical material on the back cover of *The End of Telecoms History,* he has published 18 books, 100 papers and 18 patents. With experience in academia, Ofcom, the industry and learned societies, he makes a very serious contribution to public discussion on where the telecommunications industry in general, and the mobile segment in particular, are headed.

Structure of the Book

The End of Telecoms History is a slender paperback of 106 pages, formatted in 1.5 line spacing, with many internal cross references. Consequently, there is an appreciable level of repetition. At $A24, the book appears to be highly priced – for what it is. The list of abbreviations is neither complete nor consistently applied, but these flaws are minor. The book has an index

https://doi.org/10.18080/jtde.v12n4.1176

of three pages, although the size and structure of the book make this less important than in some lengthier works.

The book is structured as seven chapters dealing with the notion of the end of telecoms history; an abridged telecoms history; user requirements; growth prospects; the implications of the thesis (that our needs in terms of data speeds and capacity have been substantially met) for the sector and more generally; delivering ubiquity; and conclusions.

The structure of the book is logical. I find it difficult to shake the thought that it is a write-up of a presentation that the author has given in PowerPoint, or has planned to give. That is not a criticism as such, but raises the question of an appropriate price for an evolved slide pack.

The End of Telecoms History

The concept of the end of history is derived, with appropriate acknowledgements, from Francis Fukuyama, an American academic and writer, whose book, *The End of History and the Last Man* (Fukuyama, 1992), was published in 1992, following the collapse of the Soviet Union and the end of the Cold War. Fukuyama argued that "not just … the passing of a particular period of post-war history, but the end of history as such: That is, the endpoint of mankind's ideological evolution and the universalization of Western liberal democracy as the final form of human government". In other words, civil society had reached its goal. In my view, Fukuyama was asserting, at minimum, that a major chapter in human history had concluded, and that whatever comes after would be heavily impacted by that point of major inflection: somehow, with the prevailing of democracy, the story of the human race must be viewed and told differently. Whatever Fukuyama had in mind, subsequent events did not support his forecast. Democracy has been under threat on many fronts, including even on the home front, where adherence to democratic norms have long been assumed to be strong and abiding (for example, *Washington Post*, 2024). One thing that Fukuyama succeeded in doing was attracting attention well beyond the levels that his book might otherwise have attracted.

I suspect that Webb has been taken with the catchy notion of "the end of something" and the sense of achievement of a major goal or target, rather than with anything else about Fukuyama's book. Getting attention on preferred issues is important. I doubt that Webb thought that the unravelling of Fukuyama's predictions in less than a generation, and arguably within little more than a decade, is something to aspire to.

Webb's version of the end of history, as applied to telecommunications, is that we now have the data rates that we need for most current and foreseeable broadband applications; and also that we have the data capacity that we need or are likely to need. Therefore, we do not need to continue investing in telecommunications systems to massively increase average rates or

https://doi.org/10.18080/jtde.v12n4.1176

capacity. These challenges, and the challenges that have driven the development of the industry for the last 150 years, have been met in developed Western countries.

Webb provides an impressive range of evidence in support of this argument.

In relation to 5G, he refers to the thesis of his earlier book, *The 5G Myth* (Webb, 2016), in which he predicted that the features being offered through 5G deployments would be of little interest and that the investment would not generate commercial returns. In the 2024 book, Webb notes that his 2016 forecast has been realised, and that three of the main promises of 5G – the increased data rates over 4G; massive machine connectivity known as the Internet of Things (IoT); and anticipated applications for ultra-low latency communications – had turned out to be of little interest to consumers in general (Webb, 2024, p. 7). Although each of these promised features of 5G has materialised to some extent, they have not resulted in the substantial take-up of substantially higher speed services, or in continuing high average growth of data capacity consumed.

The stock markets in Europe and elsewhere reflect the declining value of mobile network operator (MNO) shares compared to overall share indices over the past 15 years. MNO share prices in Europe have declined by around 50%, compared to an overall share price index increase of around 65% since 2008 (Webb, 2024, p. 10).

Data Rates

Webb discusses why optimum data rates have already been achieved and concludes: "Beyond a certain speed there is no benefit in going faster. The user experience will not change. The speed appears to be around 10 Mbit/s for mobile connections and around 20 Mbit/s for fixed connections. … [T]here may be multiple fixed users in a house. So 50 Mbit/s per house may be a safer upper limit" (Webb, 2024, p. 34).

Data Capacity

In terms of capacity, Webb notes that "fixed capacity is, in essence, unlimited. Most broadband subscribers have a dedicated connection and the amount they use it makes little difference, other than requiring relatively low-cost core network upgrades" (Webb, 2024, p. 34). As he notes, the situation is completely different with mobile networks.

There have been major stimulants for mobile data usage, such as the launch of the iPhone in 2007. However, Webb treats this change as essentially one-off, with nothing of similar impact being foreseeable in 2024. The annual growth rate in mobile data use has declined from 2016 and is continuing to decline (Webb, 2024, p. 36).

https://doi.org/10.18080/jtde.v12n4.1176

Webb points out that video applications drive mobile data consumption, and cites Ericsson reports to show that video accounted for around 70% in 2022 and an anticipated 80% of mobile data consumption in 2028 (Webb, 2024, pp. 10–11; Ericsson, 2022). However, no new source of increased video usage demand is in sight. What we are seeing, according to the information Webb has assembled, is a continued growth in mobile data consumption but at an ever-declining rate. The annual growth has declined from 80% in 2013 to around 20% at present. Webb's extrapolation suggests that there will be no growth, a plateauing, by 2027, (Webb, 2024, p. 43). This contrasts with Nokia's assessment of current growth at around 30% per annum and continuing at that rate (Nokia, 2023).

New Services and Applications

Chapter 4 of *The End of Telecoms History* is entitled "Is there anything that could reignite growth?" Webb considers a range of candidate applications and services that might generate renewed data capacity growth. Fifteen candidates, including high resolution person-to-person calling, smart wearables, remote surgery, sensor networks and automated driving (all of which Webb describes as "5G hopefuls"), are considered, and all are found wanting because they either have had no traction in the market, and are unlikely to in future, or because they are insufficiently data intensive to support a return to 30% annual growth levels (Webb, 2024, pp. 50–52).

But what about 6G? Will that be justified by meeting demands for new high-capacity services? Webb refers to the lists put forward by Ericsson and by the European Union (Webb, 2024, p. 53). The Ericsson list includes e-health for all, precision health care, smart agriculture, earth monitoring, digital twins, collaborative robots (or 'cobots') and robot navigation. The EU list includes 'truly, immersive extended reality (XR), high fidelity mobile holograms and digital twins of real-world objects'. Webb reviews the main items and dismisses their likely impact on data capacity requirements. This analysis is quite valuable for the general reader who might be unaware of what these applications are or entail. Whether they will be further developed and gain traction in the market is unclear. Webb makes the case that, whether or not they are taken up, they do not individually or collectively appear to be the magic bullet needed to reverse the decline in data growth. He concludes that 6G, like 5G, is a solution in search of a problem (Webb, 2024, p. 65).

Implications

Webb is careful to exclude coverage from the list of requirements that have been met, especially in developing economies and high-cost rural areas. However, extending coverage is costly and not a priority for a mobile sector that is struggling to achieve commercial returns

https://doi.org/10.18080/jtde.v12n4.1176

on its recent investments (Webb, 2024, p. 92). He notes that coverage extensions have been subsidised and will continue to require subsidies. Chapter 6 ("Delivering ubiquity") is devoted to ways to improve coverage and connectivity for all.

Webb sets out the implications of his thesis for each category of stakeholder in the industry, including manufacturers and operators. He sees operators as needing to recognise that they are utilities, albeit important ones, and they "should ideally restructure and cut costs to adjust to this new reality" (Webb, 2024, p. 71). Readers will be aware of staffing and other cost cutting measures announced with inevitable regularity in Australia (for example, Telstra – see ABC News, 2024), although without the recognition of the utility-reality mentioned by Webb.

Amongst the cost cutting options suggested for consideration, Webb mentions transformation to a project management entity, through massive outsourcing and becoming purely online operations without shops or physical presence (Webb, 2024, pp. 70–71). Unfortunately, that is the emerging reality now for many operators and for their under-appreciated customers. It may address the problem of declining profitability in the sector.

He also advocates substantial infrastructure sharing, especially in high-cost areas, and the use of HAPs (high altitude platforms, such as tethered balloons) and satellites, especially for rural coverage (Webb, 2024, p. 73).

Reactions

If Webb was seeking attention and encouraging reaction, he has been successful. It is not as though his ideas are new or that the issues of continued growth and the justification for major industry investments have not been raised by Webb and many others in the past. Webb anticipates much of the reaction in his book and seeks to address it there, with references to the views of major vendors, such as Ericsson and Nokia, and to the EU.

However, the book and its attendant publicity has elicited a new wave of responses. I have selected some of those responses, mainly from Australian sources.

One argument, in relation to data rates, is that certain applications will need greater data speeds some of the time. Rob Joyce, NBN Co's Executive General Manager, Customer Strategy and Innovation, was quoted in *Communications Day* as saying that increased speed was critical for some customers: "'Whether you're downloading a large presentation when you're working from home, or whether it's your kids downloading a Fortnite or Call of Duty update, in these moments that matter, speed is essential and fast is never enough.' He [Joyce] noted Webb's suggestion that most households do not require more than 50 Mbps. Joyce countered by observing that the upcoming Sony PlayStation 5 Pro console being released later this year was not equipped with a disk drive and relied on a network to download games, with many

https://doi.org/10.18080/jtde.v12n4.1176

titles in excess of 100 GB. He contrasted the seven hours it would take to download a title over 50 Mbps to the 20 minutes on a gigabit connection" (*Communications Day*, 17 October 2024, p. 3).

It seems to me that the question here is whether those who demand speed for certain applications are prepared to pay for it, or whether the costs associated with additional investment are going to be cross-subsidised by the majority of users, many of whom might have ongoing requirements that are met by at most 50 Mbps. It is possible that NBN Co has specific problems in convincing its customers to pay for data speeds above 100 Mbps. The latest ACCC *NBN Wholesale Market Indicators Report* suggests that around 5.5% of services have data speeds above 100 Mbps, and around 71% of services have data speeds of up to 50 Mbps (ACCC, 2024). This compares unfavourably with New Zealand (Canstar Blue, 2021).

Another line of criticism was that Webb relied too much on averages and that many countries, including the United States, India and Finland, had monthly data usage in excess of the 20 GB per month that Webb suggested was sufficient for users (Zehle, 2024). Without wishing to defend Webb in relation to the specifics of this argument, it seems to me that Webb is making a broader argument that the rate of data usage growth that the industry was accustomed to in the recent past is slowing and there is no basis on which to assume that the reduction will be reversed; and, hence, investment in greater capacity and higher data rates needs to be seriously reviewed. When viewed this way, the fact that some countries have average data usage levels above the 20 GB per month that Webb calculates is needed (possibly with the UK in mind) is not fatal to his overall argument.

A study by Robert Kenny of Communications Chambers, entitled *Patterns of Fixed Traffic Growth, 2024*, is consistent with Webb's overall thesis (Kenny, 2024). The study suggests that fixed traffic growth per line of around 10% per year (it was 11% for the year ending June 2024 for the aggregate of the countries covered by Kenny) is the "new normal", well down from the 40% growth rate 10 years ago and during the COVID years (Kenny, 2024, p. 1). Kenny attributes the decline in part to the reduced rate of growth of streamed video, and suggests that demographic usage might cause the decline to accelerate in future (Kenny, 2024, pp. 3–4).

Conclusion

It is not the purpose of this review to attempt to offer a final view on the arguments that Webb (or his commentators and critics) have made about the sufficiency of broadband data speeds and capacity, or about the forecasts that he has made about future growth-rate declines and industry impacts. However, I do think that Webb has made a sufficiently compelling case that the growth and service assumptions that underpin current sector investment strategies and

https://doi.org/10.18080/jtde.v12n4.1176

industry business plans need to be reviewed. Many stakeholders have a vested interest, as Webb notes, in the narrative of continued high growth (Webb, 2024, p. 43). Webb himself has a vested interest in his role and reputation as a serious contrarian and forecaster. Nevertheless, there appears to be a serious ongoing public dialogue which might not have emerged in the way it has without the book. Inevitably the dynamics of the industry will play out and plans will be adjusted accordingly, so the influence of one particular book might need to be assessed in a broader context.

I found the book interesting and thought-provoking. Others who want to see the evidence assembled and the detailed arguments made by Webb might think the same.

References

ABC News. (22 May 2024). https://www.abc.net.au/news/2024-05-22/telstra-mass-sacking-a-worrying-sign-of-things-to-come/103876130 referring to Telstra's plan to reduce its workforce by 2,800 or around 10%.

ACCC [Australian Competition and Consumer Commission]. (2024). NBN Wholesale Market Indicators Report, September 2024. https://www.accc.gov.au/by-industry/telecommunications-and-internet/national-broadband-network-nbn-access-regulation/nbn-wholesale-market-indicators-report/september-quarter-2024-report

CanstarBlue. (2021). https://www.canstarblue.com.au/internet/accc-australia-new-zealand-broadband/#:~:text=While%20Australia's%20results%20are%20based,(HFC)%2C%20and%20in%20some

Communications Day. (17 October 2024). Sydney, New South Wales. https://www.commsday.com

Ericsson. (2022). Mobility Report. https://www.ericsson.com/en/reports-and-papers/mobility-report/reports

Fukuyama, F. (1992). *The End of History and the Last Man*. Free Press, New York City, New York.

Kenny, R. (2024). *Patterns of Fixed Traffic Growth, 2024*. Communications Chambers (November 2024). https://www.commcham.com/pubs/2024/11/13/patterns-of-fixed-traffic-growth-2024.html

Nokia Corporation. (2022). https://nokia.com/about-us/news/releases/2023/10/31/nokia-technology-strategy-2030-emerging-technology-trends-and-their-impact-on-networks/ . "In the Global Network Traffic 2030 report, Nokia projects that end-user data traffic demand will increase at a compounded annual growth rate (CAGR) of 22% to 25% from 2022 through 2030. Global network traffic demand is expected to reach between 2,443 to 3,109 exabytes (EB) per month in 2030. If there is a higher adoption rate of cloud gaming and XR in the second half of this decade, Nokia projects a CAGR that reaches as high as 32%."

https://doi.org/10.18080/jtde.v12n4.1176

Washington Post. (2024). Trump's victory threatens democracy, experts say. https://www.washingtonpost.com/politics/2024/11/06/trump-victory-threatens-democracy/, 6 November 2024.

Webb, W. (2016). *The 5G Myth: When Vision Decoupled from Reality*. Amazon, 2016.

Webb, W. (2024). *The End of Telecoms History: How we reached the point of having all that we need*. Amazon.com.au, Sydney, New South Wales, 2024.

Zehle, S. (2024). Letter to the Editor of *Communications Day*, published 1 November 2024, at pp. 12–13, under the heading *The End of Telecoms History? Not really*.

https://doi.org/10.18080/jtde.v12n4.1176

E-Commerce Security Revisited

Simon Moorhead
Telecommunications Manager

Abstract: The *Journal* revisits an historic paper from 2000 flagging the potential security risks in e-commerce systems.

Keywords: History of Australian Telecommunications, e-Commerce, Security

Introduction

It is appropriate to reprise the historic paper (Blanchfield, 2000) on e-commerce security, given the significant number of online scams and fraudulent transactions occurring today. The paper was written over twenty years ago and warns of the implications of assuming security is someone else's problem, to be fitted separately outside the usual e-commerce software build cycles.

The paper recommends a pro-active rather than a reactive response to security. Often a standard design that is considered safe can contain one or more Achilles' heels that can be exploited by outsiders. Designers need to do more to ensure the security and testing of "our new and shiny house of cards".

The paper gives examples of contemporary security breaches in both government (Australian Taxation Office) and commercial (Sanity Entertainment) e-commerce systems that were supported by evidence in external links (which, unfortunately, are no longer working, given the age of the original paper).

However, a search today reveals an SBS news article (Webster, 2024) from August 2024 describing that the Australian Ombudsman has reported that "hackers were exploiting Medicare and Centrelink accounts through the myGov platform by linking them to bogus myGov accounts and making bogus tax claims worth thousands of dollars, or falsely claiming support payments" (Webster, 2024, second paragraph).

The historic paper also provides some sobering scenarios related to credit card fraud and the hijacking of websites to perpetuate credit card fraud – the takeaway being: it is much easier to

https://doi.org/10.18080/jtde.v12n4.1167

secure your e-commerce platform up front, rather than suffering the human resources cost of dealing with angry customers and the related fraud investigation and rectification.

It was difficult for the author to cover the necessary e-commerce security at a technical level, given the confines of the paper. However, he hoped to do the next best thing and "change your view on the effects of not taking into account the broader security aspects of eCommerce when considering planning, designing, and building your current or next exciting project" (Blanchfield, 2000, p. 18).

References

Blanchfield, D. (2000). e-Commerce Security – It's not an oxymoron!! *Telecommunication Journal of Austr*alia, *50*(4), 13–18.

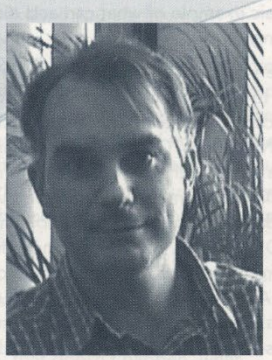
Webster, M. (2024, August 9). Revealed: How fraudsters steal from Australians through a myGov's 'side entrance'. SBS News. Available at https://www.sbs.com.au/news/article/revealed-how-fraudsters-steal-from-australians-through-a-mygov-side-entrance/8bqh9jx2c

https://doi.org/10.18080/jtde.v12n4.1167

The Historic Paper

e-Commerce Security — It's not an oxymoron !!

Dez Blanchfield



Dez Blanchfield

It is necessary to consider the possibilities of the less technical aspects of eCommerce security, and some of the peripheral aspects, which are certainly just as relevant.
Many projects are well into the development or build cycles before the security issues are even considered.
Reactive rather than pro-active response to security is not recommended. All too easily, in what might have been considered a safe, secure design, we find that perhaps we might not have done all we could, or perhaps should, to ensure the security of this shiny new house of cards.
Some illustrative examples and scenarios are described to help understand the unpleasant possibilities.

PERCEPTIONS OF SECURITY

I recently overheard a friend replace their all time favourite oxymoron one line joke 'military intelligence' with 'eCommerce security'. Upon hearing this, I knew all was not well with the state of eCommerce 'public perception'. Consider the fact that this friend doesn't actually know what the 'e' in electronic commerce actually means, and you get the idea that we might have a problem, if only perception. ..It's still an issue which requires our attention.

It's time we let our minds wander and consider the possibilities of the less technical aspects of eCommerce security, and a few of the peripheral aspects, although less technical and nerd driven, certainly just as relevant.

All too often, when talking about eCommerce, invariably the term 'security' is thrown in for good measure, almost like that dash of salt or pepper in any good pasta sauce. But when it comes to dealing with the topic in the real world of delivering solutions, even the best of us are all too often far from what might be considered ideal.

The difficulty we all face when considering what security implications a given project might have, is that all too often the term security is placed off to one side, as a component to be fitted in somewhere. Many projects find themselves well into the development or build cycles before the security issues are even considered.

How many projects throw the term security into project plans and design documents, to ensure that we can demonstrate that they have covered the issue, that they have complied with some often unwritten set of standards pertaining to security?

'Security — that's somebody else's problem. I just have to get this thing working, then I'll deal with security..'. Sound familiar?

It isn't only when some hard case, often from the anonymous 'other' side of a computer screen, manages to weed out some previously unconsidered Achilles heel, we get the point.

Reactive rather than pro-active response to security is not recommended. All too easily, in what might have been considered a safe, and secure, design, we find that perhaps we might not have done all we could have, or perhaps should have, to ensure the security of our shiny new house of cards.

Sure, I can hear you scoffing already. 'Give me a break', I hear you say. 'Not my project, we know what we're doing, we wrote the book'. Is that what you're thinking? That zillion dollar project you just completed, it's leading the world in its brilliance, the press releases said so, right, but is it secure? Are

Journal of Telecommunications and the Digital Economy, ISSN 2203-1693, Volume 12 Number 4 December 2024

https://doi.org/10.18080/jtde.v12n4.1167

you sure, are you really sure? Even if you are, check again, please!

SOME EXAMPLES

It happens to the best of us. Let me give you a couple of recent examples, right here in our own back yard.

Example #1:

One of the highest profile World Wide Web sites we've seen in years, the Australian Tax Office (ATO) GST Assist program online is a classic example of what can and did go wrong.

When building an online eCommerce web site to better manage the process of allocating Australian Business Numbers (ABNs), the ATO unfortunately managed to miss some of the basics. It took a moment of fiddling for someone to stumble over a fundamental, but disastrous hole in the ATO's web site. Subtracting the value of one (1) from the account ID allocated to you by the ATO's web site, would give you the full details registered by the person who subscribed one session before you!

A bright young fellow by the name of Kelly managed to do just that. ..A self confessed 'non-hacker' (no hacking was required), Kelly, by merely adjusting his ID in the web page URL, was able to gain access to the confidential details of at least 17,000 taxpaying entities registered via the ATO site.

Kelly was able to gain access to business phone and fax numbers, postal addresses, all bank account details, and email addresses.

Kelly sent emails to most of the 17,000 people and companies whose details he had accessed, to let them know that they too had a problem. ..Imagine the nightmare the ATO had on their hands! A world leading solution to one of the government's more difficult challenges in recent years, and what should have been text book stuff for all the right reasons, became text book stuff for all the wrong reasons.

Before you choke from scoffing, review what the Sydney Morning Herald had to say about it at
http://www.smh.com.au/news/0006/29/update/news2.html
and get the full story.

Example #2:

Internet users swamped the brand new Sanity.com music web site registering for the chance of winning a free music CD – unfortunately the folks at Sanity apparently shipped quite a few free CD's before realising they had a problem. It seems the ordering process kindly permitted some customers to order CDs without having to provide their credit card details as part of the purchase process. ..Sure you can turn this into a publicity stunt, but at what cost?

Did I hear you scoff once more? Surely not, well ok, take a moment to go visit
http://www.internetnews.com/intl-news/article/0,,6_245621,00.html
and then come back and tell me that you're not thinking about eCommerce security.

Now I hate to say this, but I actually could go on like this for hours and hours. But filling this short article with war stories from the bleeding edge of business via the World Wide Web under the guise of eCommerce, isn't going to hold your interest for long.

Almost every article I've read about doing business online, and in particular about security issues facing eCommerce development, tends to be littered with the IT industry's staple diet of three letter acronyms (TLAs) and jargon such as SSL, SET, digital signatures, digital certificates, smart cards, private key encryption, etc.

I think we've all had enough of the jargon and TLAs, of nicely detailed colour graphics showing the flows and processes of how Netscape's Secure Sockets Layer (SSL) protocol works and why you would need one. If you haven't had enough of it, go visit Netscape's home page, they are dying to tell you all about it.

I've found far too few quality articles about the fundamentals, the actual 'low tech' behind the scenes, as it's more often than not the little things that trip you up. Sure, you may have super computer power running your web server but the absolute basics, such as your registered domain name, or the software you are using on your web server, the naming of your sites on that server, how people login to that server and such, are critically important.

Issues like 'plain text log files'; 'public networks being used for messaging; local area network vulnerabilities; open ports on Ethernet switches; unlocked communications racks in data centres; unshredded printouts of consignment orders; and shipping details being dropped in trash cans.

False 'mirror' web sites under similar domain names to yours, of email addresses like sales4yourcompany@yahoo.com and the like – are these things taken care of?

I'm of the opinion that what is needed currently, is a great deal more open forum style discussion of the fundamental issues we face in eCommerce today.

Journal of Telecommunications and the Digital Economy, ISSN 2203-1693, Volume 12 Number 4 December 2024

https://doi.org/10.18080/jtde.v12n4.1167

The more the fundamental issues are debated openly, the sooner we can deal with them and hopefully re-educate those of us who too easily throw in a quick patch up, rather than solve a serious issue right up front. ..It's the fundamentals that often trip us up; technology is making it harder to see the delimiting lines. Those lines are getting more and more grey by the day, and the guys in the black hats look more and more like you and me.

Certainly pressures such as economics, time and politics in even small firms can force us to grey the lines a little, but that's often when we should be on our toes more than ever.

Perhaps it's primarily a case of education on both sides of the browser? Let's talk through an example of some of my current favourites, you might be surprised by this simple little story.

Recent research conducted by Citibank indicates that 60 percent of Australians still do not trust the Internet with their credit card details.

That same 60 percent said they would gladly provide their credit card details over the phone to pay for a purchase made on an Internet world wide web site. Oh no!

Most folk I tell this to simply nod their heads in agreement, even the die-hard geeks. Almost all say that they don't trust the systems currently being used online to take payments.

Even when an SSL enabled server is handling the transaction, and their World Wide Web browser reports that the link is secure by displaying a safely 'locked' session with a nice little icon on the screen too! Gosh!

I choke at this and ask how they feel about some young kid, let's call him little Johnny, on the other end of the phone writing down their own personal copy of your details as they enter them into a computer database to record the purchase, over a computer network.

A CAUTIONARY TALE

What happens if that credit card is used, let's say by our same little Johnny, to make a purchase over the phone, of some shiny new music CD's. Being a cunning little fellow, he makes arrangements for them to be delivered to a false address, where little Johnny might be sitting on the steps of a stranger's house waiting to sign for the delivery. He then gets on his razor scooter and rides off into the sunset, chuckling all the way listening to his new CD's.

OK, so you get your credit card statement in the mail. If you're diligent, you pick it up and get on the phone with the bank, they make you go down to your branch, fill out a declaration stating you did not make that purchase. Some 24 hours later the money gets put back into your account by the bank, and is taken out of the merchant's account.

So if you are the owner of the credit card, you don't lose, the merchant does, which is fine, that is unless you're the merchant.

More than likely, if you are like most of today's modern lifestyle society, you won't notice a $29.95 purchase on your statement under 'Sanity Online', as you won't have paid your account with the first notice from the bank. When the second 'late' notice comes through without a detailed statement, you panic like the rest of us, and hurriedly pay your monthly bill just hoping they don't take your card away.

Either way, little Johnny got his free CD and is now enjoying it with his pals, with the volume turned all the way up to eleven.

'Yeah, sure, that doesn't happen, what rubbish' I hear you say.

Well consider this scenario:

Scenario #1:

Picture an Internet Service Provider (ISP) who has an eCommerce-enabled subscription form. Our little Johnny uses your stolen credit card details to pay this same ISP for an Internet dialup access account. He connects to the ISP's home page, clicks on 'join', chooses the flat fee monthly account at $29.95 for unlimited access. The ISP web page lets Johnny enter the basic details the ISP needs for login, password, and payment details, including your credit card number and expiry date. (These are the same details you provided when you purchased that pizza, remember).

Because Johnny is a bright kid, he makes sure he clicks on 'receive invoices by email', so now the ISP will email the monthly accounts to 'gotcha' via the ISP's email server rather than sending them to a postal address. A postal invoice would be a problem as it could be traced or would be sent 'Return To Sender' possibly if it went to a false address, but our Johnny isn't that silly.

Johnny signs up with the login 'gotcha' and off he goes, one month's free Internet access.

It doesn't sound like a big deal to most of us — that one little $29.95 a month Internet account, but if you end up with even ten

Journal of Telecommunications and the Digital Economy, ISSN 2203-1693, Volume 12 Number 4 December 2024

https://doi.org/10.18080/jtde.v12n4.1167

people doing this each month for let's say three months each, that's around $89.95 per person, and a total of $899.50 per month. Multiply that by three months and you get a total of $2,698.50 per quarter.

Now factor in the human resource time to deal with it:

- a receptionist taking the angry phone calls as they come in
- accounts admin staff handling the enquiries
- technical support and admin staff providing answers to the accounts admin staff
- finance staff sorting out the billing issues as well as the on-flowing effect to cash flow
- financial management dealing with the bank who took money from the company account
- company management answering bank requests for an explanation to what the bank sees as ongoing wrongful billing
- marketing and sales management
- marketing communications staff team dealing with any press which might arise

and it goes on. Certainly I've stretched it a little to illustrate the point, but here's the moral of this little story:

A single phone call to the card holder's phone number provided with the order online to perhaps confirm the purchase, or even better, to provide the login name and password as well as confirm the order, would have saved that ISP almost immeasurable cost and resource.

I consider this to be just a tiny part of what can only be considered a step in an eCommerce security procedure. Surely one, which at $0.40c per phone call, would be an enormous saving to that business.

While running an ISP myself, I regularly had a pile of uncollectable invoices placed on my desk each month. This was a reminder that we had ongoing problems like this – last tally I recall the total of uncollectable or bank-recovered credit card billing to be around AUD$114,000 over 12 months, and I am of the opinion that we ran a very tight ship. I should note that the $114,000 was the sum of the invoices — it did not take into account the peripheral costs of human and technical resources incurred to reach that cost.

The problems faced when considering eCommerce security are not just those found behind the black boxes selling you goods —, consider the issues for the unwary traveller on the internet for a moment and you will quickly see yet another mine field.

Let's consider some of the less considered traps a new web venture might fall prey to.

Consider this scenario (the company name has been changed so I don't get sued!):

Scenario #2:
To register that all important DOT COM domain name for your Internet start-up you just need a valid credit card and email address.

So, little Johnny pops into the scene once more. Johnny jumps on the Internet with his dad's home computer. He again chooses to use the credit card details you gave him 'securely' over the phone while ordering those pizzas for the recent late night with the sales team planning your next eCommerce project.

Johnny pops on over to the web site http://www.networksolutions.com/ where it takes him around five minutes to register MelodyMusic.COM which Melody Music Company Pty Limited hasn't yet registered – they are happy with Melody.COM which they are currently using.

To activate his newly-registered domain name Johnny jumps onto a free web site network like GeoCities, which is sponsored by Yahoo! He signs up for a free web page, with his own domain as the address, and agrees to have Yahoo! place banner advertising on his free web site to let the nice folk over at Yahoo! recover their running costs in providing this wonderful service of free web sites.

A little web page editing later, Johnny now has a free web page online. It's actually a copy of Melody.COM – Johnny is a smart kid, he just copied the original HTML source code right off the home page from Melody.COM.

Five minutes' worth of cutting and pasting from his browser, which has this neat feature of allowing visitors to 'View Source' from your page, and Johnny has uploaded Melody.COM's home page to his MelodyMusic.COM Web site. Oh yes, Johnny had to make allowances for that silly free pop-up banner ad in the upload, but he's a HTML guru as a result of a TAFE computer course his parents recently paid for, so a few layout changes to the site took another five minutes, no big deal.

Now this is where you enter the scene – it would be no fun if you didn't play a part, so I've included you – just for fun.

You get on the web; you've just finished

Journal of Telecommunications and the Digital Economy, ISSN 2203-1693, Volume 12 Number 4 December 2024

https://doi.org/10.18080/jtde.v12n4.1167

reading about the Melody Music Company and their wonderful new eCommerce enabled web site in the Financial Review, and you figure you might go and buy that Kylie CD you've heard so much about.

You can't recall the address or URL of the site, and Mary down the hall has pinched the Financial Review and taken it to lunch to read, so you pop on over to your favourite internet search engine http://www.WebSearch.com.au and go hunting for the Melody music web site.

What you don't know at this stage is that our little Johnny has had the foresight to pop on over to http://www.jimtools.com and use their neat free Internet site registration utility. He just typed in his Internet World Wide Web page address (or URL), and through the magic of computers, MelodyMusic.COM gets submitted to over 8,500 Internet search engines and directories like WebSearch.COM.AU and Yahoo! where it's sure to be found.

Melody Music Company Pty Limited and their Melody.COM have not submitted their site to any online directories or search engines, and so it's not likely that their site will be found by anyone searching for them in the likes of the directories provided by LookSmart, or Yahoo! or search engines like WebSearch.COM.AU, Altavista or similar.

You, meanwhile, unwittingly search for Melody Music on good old http://www.WebSearch.COM.AU and hey presto, they find MelodyMusic.com which looks like the site you wanted, so you click on the link and off you go.

WebSearch.COM.AU is one of the 8,500 search engines and directories that are capable of accepting automatic site submissions from places like Jimtools.COM, so within 24 hours of little Johnny using Jimtools.COM the bogus Melody Music web site will appear in the WebSearch.COM.AU search engine.

So, now you are connected to the MelodyMusic.COM World Wide Web site, and it looks great. Right there on the front page, they have lots of copies of that new Kylie CD you wanted, so you immediately click on ADD TO SHOPPING CART, then you click on CHECKOUT, and start punching in payment and delivery details.

Johnny again has had the wisdom to use the free secure payment services (SSL). Johnny chooses to use a service provided by a great bunch of folks he found on an adult pornography site. A payment gateway service company called EasyPaymentGatewayServices Inc.

EasyPaymentGatewayServices Inc will let Johnny use their payment gateway and SSL servers for a small fee of just 15% of any transaction made through his web site MelodyMusic.COM. What a fantastic service. No set-up fees. Just sign up on line and you're ready in 5 minutes!

There is a further benefit of using EasyPaymentGatewayServices Inc. It turns out that EasyPaymentGatewayServices Inc likes their 'affiliates' to spend their earnings online.

So the folk over at EasyPaymentGatewayServices Inc will let Johnny redeem his payments online through their network of other affiliated web sites.

So Johnny now can spend his earnings from online CD sales through his bogus copy of the Melody Music Company web site, by spending it online with World Wide Web sites affiliated with EasyPaymentGatewayServices Inc.

The great thing about this whole set-up of course is that at no stage does our cunning little Johnny have to transfer his fraudulently earned income to a bank account in his or any other traceable name. Everything is online, it's virtual, and practically impossible to back step to an actual human.

Oh, guess who just happens to be an affiliate site to EasyPaymentGatewayServices Inc. Yes, you guessed it. None other than our friends over at the Melody Music Company and their site Melody.COM!

Now this is really getting interesting.

After a few days, Johnny has loads of credit, as a result of his bogus sales to you and thousands of others which he sent SPAM email to, as part of launch of his bogus web site.

So once again, our little Johnny is sitting on yet another set of random house steps, where the owners are away. He has taken the day off school, he has to sit and wait for the courier to deliver his box of 50 CD's he ordered from Melody.COM – why go to school, he's too smart to waste the day at school anyway – there's some serious eCommerce to be done!

God help that US firm taking the payments online for Johnny, and for Melody Music Company who can't work out why whole boxes of CD's get delivered and signed for by Mr Gotcha Smith, time and time again. As for the fulfilment house, they are still trying to read the signature, seems it looks a little like 'Gotcha Again' but it's not easy to read.

Journal of Telecommunications and the Digital Economy, ISSN 2203-1693, Volume 12 Number 4 December 2024

https://doi.org/10.18080/jtde.v12n4.1167

And as for the residents at #1 Gotcha Street where little Johnny had the CD's delivered — well they are still filling out the statutory declaration forms for the police, and everybody's getting a little flustered!

Now there's the matter of you not getting your CD's some weeks after you placed your order. You still don't realise of course that you've been interacting with the false site.

Of course you placed an order with little Johnny's copy of the Melody Music site, not the real one, but you won't realise this until you get weird responses from the 1800-MELODY-HELP phone support service. They of course have no record of your purchase.

In fact, due to the fact that 1800-MELODY-HELP is outsourced to a call centre, they are invariably not Internet savvy. Because MelodyMusic.COM sounds like what they have on their 'script' to follow when trying to help you, they don't pick up the significance of the different URL while talking to you on the phone.

As far as 1800-MELODY-HELP are concerned, you're some crank on the phone screaming at them because you think you purchased a CD online. They can't see you in the purchasing system; they simply tell you that it's impossible that you made any such order, and apologetically they advise you to check that you visited the correct site, and hang up, click!

Now if that second scenario doesn't give you something to think about, perhaps you shouldn't be considering eCommerce.

Unfortunately, it would be difficult for me to cover eCommerce security at a technical level, even in brief detail within the confines of this article. I'm still working on the book. So I've hoped to do the next best thing, that is, change your view on the effects of not taking into account the broader security aspects of eCommerce when considering, planning, designing and building your current or next exciting project.

If I can have achieved one thing here, I hope it is to have you re-consider what you previously understood to be the scope of eCommerce security at the very least, and hopefully given you reason to open your browser and go hunting for further reading on the topic.

THE AUTHOR

Des Blanchfield has 16 years of experience in IT, in most aspects of software, systems, and telecommunications from development to implementation, from hands-on to management. Currently Dez is a consultant to a select few eCommerce and Telecommunications firms in Australia, New Zealand and Hong Kong. Key areas of interest over the past few years have seen Dez develop technologies in Clustering, Network Redundancy, Systems, Network and eCommerce Data Security, Firewalls, Load Balancing, Content Servers, Search Engines, Ad Serving Engines, and Remote Monitoring. He may be contacted at:

dez@blanchfield.com.au

https://doi.org/10.18080/jtde.v12n4.1167

E-Commerce Security Issues, Then and Now

Thoughts Stimulated by an Historic Paper by Dez Blanchfield (2000) on E-Commerce Security

Graham Shepherd
Life Member, TelSoc

Abstract: Simon Moorhead's recent historical reprint (December 2024) revisiting Dez Blanchfield's December 2000 paper in *TJA* on e-commerce security has stimulated this author to make additional comments on the weaknesses of major websites and of email systems today in permitting fraud and deception, in comparison to 24 years ago.

Keywords: Internet security, e-commerce, phishing, spam

Introduction

Simon Moorhead's paper 'e-Commerce Security Revisited' (Moorhead, 2024) has usefully drawn our attention to the astute warnings by Dez Blanchfield (Blanchfield, 2000) in this *Journal's* predecessor, the *Telecommunication Journal of Australia*, concerning the dangers of designing online systems without sufficient regard to security.

Moorhead's introduction to the historic paper, and Blanchfield's original, have stimulated this reviewer to make some additional observations on the hazards of some current practices in the operation of both consumer websites and email. The History section editor of the *Journal*, Peter Gerrand, suggested that these observations deserved publication as a companion paper, despite its slimness.

The Dangers of Outsourcing Security

Practically no e-commerce sites today, except the very big ones like Amazon and Google, use their own internal security systems for transactions. Most sites, even big ones, pass the buyer off to a renowned gateway, such as PayPal, Stripe, Square, or one of the big banks. In the process, credit card numbers never get transmitted through or stored on the e-commerce site. To a large extent, the e-commerce sites are passing off this part of the risk to the gateway for

https://doi.org/10.18080/jtde.v12n4.1182

a small fee per transaction. The big risk remaining for these sites is securing personal details, particularly names, email and physical addresses, and phone numbers. The website software alone must protect this information or avoid saving it. This is still a bane for website administrators, who have to choose from a small number of highly secure Content Management Systems to do that hard work and to keep up-to-date. Weekly updates would be the norm, indicating that the author of the original paper was very perceptive about the growing scale of the problem.

Vulnerability to Phishing and Spamming

Today, this is where phishing and spam emails come in. They rely on huge email address lists stolen by specialist hackers. This way only a few consumers need to be hooked to make a big payoff. Companies, on the other hand, pay the price in ransom demands and reputation. The scale of this problem in 2022 was that an estimated 3.4 billion spam emails were sent every day, that is, over 48% of all emails. Of this number Google blocks only 100 million per day, barely scratching the surface (https://aag-it.com/the-latest-phishing-statistics/). In 2022 the number of ransomware attacks in the United States alone amounted to around 217.5 million (https://www.statista.com/statistics/1377918/ransomware-target-countries/). Australia has not been free of the problem: e.g., Optus, Medicare, Canva, ANU etc. (https://www.upguard.com/blog/biggest-data-breaches-australia/).

Poor Verification of Email Sources

Whilst most Internet protocols are subject to constant upgrading to stay ahead of the game, email remains a very weak link.

The email protocols SMTP (client) and IMAP and POP (server) have been around for a very long time and have only been patched in simplistic ways to try and verify that an email comes from whom it purports to be: for example, DMARC (Domain-based Message Authentication Reporting & Conformance), SPF (Sender Policy Framework) and DKIM (DomainKeys Identified Mail) records are (often, but not always) attached to the domain name server (https://dmarcian.com/). Most spam still gets through. The telecommunications operators who once invested heavily in standards have largely abandoned the field to the vendors, who have less interest in keeping the traffic on the Internet valid, perhaps the opposite. The email service providers, like Google, Apple Microsoft and Yahoo, have a poor record and little interest in collaboration and so the patches have come from smaller players with smaller impact.

https://doi.org/10.18080/jtde.v12n4.1182

Shifting Risk onto the Customer

At about the time the original paper was being written, banks were closing branches and shifting as many transactions as possible online. The risk for them was loss of confidence if their security systems were breached, particularly by hijacking a credit card number. In the early days they bore that risk and usually refunded the buyer (not the vendor because they needed the warning to sort their systems out). However, the banks have largely stopped accepting this risk, so the old saying, "buyer beware", is still very pertinent.

Recommendations

The huge growth in both the good and malicious uses of the Internet has focused attention on how critical cybersecurity is for our ordinary everyday lives. The Internet Engineering Task Force (IETF) is the premier standards development organisation for the Internet, but by design it is entirely voluntary, has no membership, nor does it enforce compliance (https://www.ietf.org/). Email is amongst the technologies it addresses, including DMARC mentioned above, and a new protocol, JMAP (JSON Meta Application Protocol), to replace IMAP (one day, if sufficient service providers take it up). However, security and trustworthiness of email for consumers and businesses do not appear to be considered matters of great urgency at the moment. The Internet Society also has a mission addressing "the development of the Internet as a global technical infrastructure, a resource to enrich people's lives, and a force for good in society" (https://www.internetsociety.org/).

Another approach to trust is blockchain technology, which creates direct "trustless" (as opposed to "untrustworthy") contracts directly between users. But blockchain is a long way from achieving a central place in e-commerce and has its own problems, such as enormous energy usage, its avoidance of government regulation and its attraction to speculators and money launderers.

Fixing email remains the most attractive path for solving the problem.

The ICT (information and communication technology) industry would seem to be the best lobby group to initiate the requirements definition and development (within the IETF) of a much more robust email protocol to replace the aging IMAP and its associated protocols. Australia has a number of active ICT industry bodies which collaborate on major issues, including TelSoc, the Australian Computer Society (ACS), The Pearcey Foundation and the Australian Information Industry Association (AIIA). Sponsorship for a series of forums on this subject could be sought from the telcos, banks, Google, Microsoft and others, with the aim of developing a new and evolving set of protocols to address this challenge.

https://doi.org/10.18080/jtde.v12n4.1182

References

Blanchfield, D. (2000). e-Commerce Security – It's not an oxymoron!! *Telecommunication Journal of Austr*alia, *50*(4), 13–18.

Moorhead, S. (2024). E-Commerce Security Revisited, *Journal of Telecommunications and the Digital Economy, 12*(4), 178–185. https://doi.org.10.18080/jtde.v12n4.1167