Content Diffusion Between Wireless Mobile Devices

Bryce Thomas

Supervised by:

{Prof. Ian Atkinson, A/Prof. Raja Jurdak, A/Prof. Bruce Litow}

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Information Technology at James Cook University.

Submitted 21st August 2015

Declaration

This thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institute of tertiary education. Information derived from the published and unpublished work of others has been acknowledged in the text and a list of references is given.

Bryce Thomas

21st August 2015

Statement on the Contribution of Others

This thesis was produced under the supervision of Prof. Ian Atkinson, A/Prof. Raja Jurdak and A/Prof. Bruce Litow, all of whom provided academic and editorial advice. Further editorial support was provided by Patrick Laub and John Cokley.

Primary financial support was provided by an Australian Government Australian Postgraduate Award (APA) scholarship and a Commonwealth Scientific and Industrial Research Organisation (CSIRO) Office of the Chief Executive 'top up' scholarship. Secondary financial support was provided by the Queensland Cyber Infrastructure Foundation (QCIF), the Dromey Foundation and James Cook University. Tertiary financial support was provided by the bull market of 2013.

Data access was facilitated by Timothy O'Donohoe, Niel Bergmann and Thomas King.

Paper co-authors include Raja Jurdak, Ian Atkinson and Kun Zhao whose contributions include academic and editorial advice.

Acknowledgements

I would like to acknowledge good luck. Good luck to be born in a wealthy country. Good luck to have lived on soil free from war. Good luck to have access to clean water, food and shelter. Good luck to have been provided with world-class health care. Good luck to have received an education unparalleled throughout human history. Good luck to be born a gender and race free from discrimination. Good luck to have access to the near sum of human knowledge. And good luck to be in the company of others with similar good fortunes who have reinforced my already abundant opportunities.

For every person whose lot in life has exceeded my own, there are thousands more whose haven't. I therefore acknowledge how fortunate I've been on a grand scale and the critical role this has played in allowing me to comfortably pursue this thesis. As we progress along the arrow of time, I hope the kind of opportunities bestowed upon me are rendered more widely available.

"...some architectural principles inevitably change. Principles that seemed inviolable a few years ago are deprecated today. Principles that seem sacred today will be deprecated tomorrow. The principle of constant change is perhaps the only principle of the Internet that should survive indefinitely."

— RFC 1958: Architectural Principles of the Internet.

Abstract

The architecture of today's Internet was conceived in the 1960s and 1970s. Networking aimed to solve the problem of *resource sharing*—enabling remote access to scarce and expensive devices such as card readers, high-speed tape drives and supercomputers. The resulting communication model was one of a conversation between exactly two hosts—a resource consumer and a resource provider. Over the past 50 years, computers have undergone a metamorphosis from scarce, multimillion-dollar machines occupying large rooms to cheap and ubiquitous pocket-sized devices. The value of the Internet no longer lies in facilitating access to scarce hardware. Rather, the Internet now primarily derives its utility from enabling large-scale electronic content distribution and retrieval.

At odds with the content-centricity of modern network applications is the host-centricity of the prevailing network architecture. The juxtaposition of the two paradigms has been the impetus for new research into alternative Future Internet architectures under the umbrella title of *Information-Centric Networking (ICN)*. Though individual visions for ICN vary substantially in ambition and detail, the central role of content is widely recognized. There is all but consensus that content identity and security ought to be independent of content location, enabling seamless replication throughout the network. This would permit secure content retrieval from a cache geographically and/or topologically close to the end user while balancing load for the content distributor. What separates the current wave of research from the myriad of application-specific distributed caching overlays is

a generalization which intends to render content the primary *network* abstraction.

One proposal emblematic of ICN's core principles is *Content-Centric Networking (CCN)*. CCN envisages an Internet architecture in which Internet Protocol (IP) packets are replaced by comparably sized chunks of named content. Attached to each chunk is a cryptographic signature allowing the validity, relevance and provenance of content to be assessed, irrespective of sender. Like other ICN projects, CCN solves a critical problem on the path towards truly distributed content dissemination—establishing trust in content received from untrusted senders. This opens up compelling new caching strategies in which content may be retrieved from *any* device in the network, including those not typically thought of as content distributors.

This thesis is motivated by one particular application of CCN's universal secure caching—content diffusion (spreading) *directly* between colocated mobile wireless devices. Over the past decade, commercial and private Content Delivery Network (CDN) overlays have proliferated on the Internet as a way of geographically distributing cached copies of content, balancing load for content producers and increasing performance for consumers. These CDNs however are limited in their reach by the granularity of the facilities in which content servers may be installed—typically Internet Service Providers' (ISPs') Points of Presence (PoP). Building a 'CDN' of sorts from the colocated end user devices themselves is a natural progression of the distributed caching architecture and one which we consider the 'final frontier' for CDNs. The meaning of this statement is that there are no content caching locations closer to an end user than another colocated device.

The potential advantages of direct device-to-device content sharing are manifold: (i) higher throughput, (ii) lower latency, (iii) extended network coverage and (iv) reduced load on infrastructure. Despite being conceptually attractive in these respects, relatively little is known about the intrinsic spreading potential of opportunistic encounter (contact) networks. It is to this matter that we direct our attention in this thesis. Drawing upon empirical spatiotemporal traces of device mobility and inferred device encounters, we conduct a simulation-driven exploration of opportunistic content diffusion. We address a number of aspects of spreading potential, which are divided into three separate papers. In the first, we perform simulations parameterized by site, time and number of source devices in order to establish the impact of these variables on content diffusion. We also motivate content diffusion by integrating real-world application usage statistics from a popular campus maps application into our simulations. Our second paper addresses the impact of margins of trace uncertainty on diffusion potential, as well as the inherent variation in diffusion potential as a function of the randomly chosen source device. Our third paper seeks to understand how nodes' spatiotemporal preferences impact on spreading potential. We achieve this through the presentation of a set of novel null models which separately decorrelate the relationships between times, locations and nodes. We also describe how the null models can be generalized to any contact network in which contact events are predicated on node colocation.

As a preamble to the three papers on content diffusion, we present one additional paper, the subject matter of which is the SPDY web protocol. As such, the set of three papers just described are in fact numbered 2, 3 and 4 in the thesis, with the SPDY paper being listed as paper number 1. The SPDY paper is the by-product of a thorough review of the current state of the art in *near-term* protocol innovations which we conducted during the early stages of the research. The SPDY paper offers an interesting point of comparison between the pragmatic strategies behind near-term protocol improvements and the broader long-term research perspectives on computer network architecture.

Contents

1 Introduction **1**

1.1 The Internet's Past, Present and Future 2

1.2 Wireless Peer-to-Peer as The Final Frontier of Content Delivery Networks 4

1.3 Aims, Relevance and Significance of the Thesis 5

1.3.1 Aims 5

1.3.2 Relevance 6

1.3.3 Significance 7

1.4 The Thesis 7

1.4.1 Structure 7

1.4.2 Assumed Knowledge and Intended Audience 9

2 Historical Perspective and Motivation **10**

2.1 The Telephone System (Circuit Switching) 12

2.2 The Internet (Packet Switching) 17

3 Information-centric Networking (Dissemination) 22

3.1 CCN Background 23

3.2 CCN Network Stack 24

3.3 CCN Packet Primitive 25

3.4 CCN Forwarding Engine 25

3.5 CCN as a Framework for High-Speed Peer-to-Peer Content Dissemination 27

3.6 Delay-Tolerant Networking and Mobile Ad hoc NETworks 28

4 Paper 1 – SPDYing Up the Web 31

5 Paper 2 – Content Diffusion in Wireless MANETs: the Impact of Mobility and Demand 44

6 Paper 3 – The Impact of Mobility and Content Demand on Diffusion in Wireless MANETs 54

7 Paper 4 – Diffusion in Colocation Contact Networks: the Impact of Nodal Spatiotemporal Dynamics 71

8 Conclusion 91

8.1 Summary of Contributions and Results 92

8.2 Challenges, Limitations and Future Work 93

8.2.1 Access to Empirical Data 94

8.2.2 External Validity 95

8.2.3 Real-world Applications 95

8.2.4 Interdisciplinary Integration 96

Appendices **97**

A List of Thesis Papers **98**

B Research By-products **99**

B.1 JCUNav 99

B.1.1 JCUNav Versions 100

B.1.2 JCUNav Data 104

B.2 liber80211 104

B.2.1 IEEE 802.11 Wi-Fi Probe Requests 106

B.2.2 Geolocating ESSIDs with WiGLE 107

B.2.3 Technical Overview of liber80211 107

B.2.4 Challenges to Probe Request-based Monitoring 109

B.2.5 Source Code and Related Work 110

9 Acronyms & Glossary 111

Acronyms 111

Glossary 112

1

Introduction

How information can be transmitted quickly and efficiently over long distances has been a topic of human interest for at least several thousand years. Since prehistory, smoke signals have been used to propagate information far beyond the reach of the human voice. Today, we transmit information at terabits per second over optical fiber channels. There is a rich history behind high-speed communication networks, although interestingly, most of the major advancements have been made in the past 200 years alone. This period includes the electric telegraph, telephony and digital packet switching. And in just the past 50 years, the pace of innovation in computer networks has proven so relentless that if today's technology were to be transported back to the mid-20th century, it would seem almost alien. But networks have not only gotten faster, they have also become increasingly reliable, ubiquitous and application-independent. The smoke signals and optical telegraph which preceded the electric telegraph required good weather and a clear line of sight. The electric telegraph solved this problem but was for the most part used only in government and commercial settings. The telephone managed to find a place in everyday households but was limited in its use to human conversation. The Internet ushered in a myriad of new forms of communication but until recently most of its users remained tethered to a desk.

Most recently, compact mobile device form factors and continued innovation in wireless networks have led to increasingly ubiquitous communication networks. In the present day, a user with a mobile device costing less than $100 can be seamlessly connected at home, at work, at a coffee shop and on the go[1].

1.1 The Internet's Past, Present and Future

The combined advances in packet switching and the stored program computer through the 1940s to 1970s ultimately led to the creation of the Internet, the most important application of telegraphy in the 21st century. Interestingly, the engineering principles and architecture of *today's* Internet remain those originally conceived in the 1960s and 1970s. Networking aimed to solve the problem of *resource sharing*—remotely accessing scarce and expensive devices such as card readers, high-speed tape drives or even supercomputers [1]. The resulting communication model was one of a conversation between exactly two hosts—a resource consumer and a resource provider [1]. This is deeply engrained in the Internet Protocol (IP) [2], whose packet header specifies exactly two addresses—a source and a destination. As the applications of the Internet have broadened well beyond those envisaged almost 50 years ago, it has come into question whether a host-centric architecture remains the most appropriate for the Internet moving forward.

The dominant application of the Internet today is content distribution and retrieval. The user, once concerned with the *where* of computer networking (the machine, tape-drive, card reader), is now concerned with the *what* of computer networking (the video, email, news, photographs). At odds with this phenomenon is current networking technology which still only speaks of connections between hosts. The apparent mismatch between the host-centric abstraction of the IP and

[1]and even on the toilet.

the user-centric abstraction of content means that most application layer protocols overlay their own content-centric features on top of the IP. Furthermore, the mismatch has spawned an entirely new category of overlay network—the Content Delivery Network (CDN). Akamai [3] taken as a prominent example is a large commercial CDN whose sole business is quickly and scalably delivering Web content on behalf of site owners. Akamai operates over 160,000 servers in 95 countries inside more than 1,2000 networks [4], with peak traffic in excess of 20 terabits per second and over 50 petabytes [5] of traffic per day. Akamai has built an extensive overlay infrastructure that caches content at locations both geographically and topologically close to end users, accelerating and scaling content delivery beyond what is possible with single host/single location network architectures. Akamai however is only one prominent example that happens to publicly release these statistics. Other large Internet companies such as Google, Amazon, Microsoft and Facebook are all thought to have on the order of hundreds of thousands to millions of servers, a sizeable portion of which likely operate at least in part as distributed content caches.

The content-centricity of modern network applications has increased interest in the past few years around *Information-Centric Networking (ICN)*. ICN broadly encapsulates approaches to computer networking that focus on named content as the primary network abstraction as opposed to host identifiers. The broad motivation of ICN is to identify architectural approaches to computer networking that may be more appropriate for the Internet's dominant use cases at present and into the future. One project emblematic of ICN is *Content-Centric Networking (CCN)*. CCN in its purest form replaces the entire notion of IP addresses with named data. Under CCN, a device wishing to retrieve content from the network requests that content by name without needing to resolve to any specific host identifier. The assumption of CCN is that content does not *live* at any one location, but rather is cached throughout the network. This caching is envisaged to occur not only at dedicated storage locations, but also on other devices not typically thought of as content caches, including routers and even end user devices. An important feature of CCN and one we presuppose in our own research is that all content is cryptographically signed in a way that enables trusted content to be retrieved even from untrusted nodes. This opens up compelling new secure Peer-to-Peer (P2P) content distribution and retrieval possibilities at the very edge of the network.

1.2 Wireless Peer-to-Peer as The Final Frontier of Content Delivery Networks

Throughout this thesis we refer to wireless P2P content sharing as the 'final frontier' for CDNs. This implies that there is no node closer to the end user than another colocated peer device. What makes CDNs useful in general is that content caches are highly geographically distributed so that users, irrespective of their location, are in close proximity to a copy of the cached content. Whereas traditional CDNs are ultimately constrained in deployment granularity to the level at which computational resources can be installed (e.g. inside Internet Service Providers' (ISPs') Points of Presence (PoP)), a P2P 'CDN' operates at the granularity of the consuming devices themselves. The peer devices are in effect the 'leaf' nodes of the network.

Our research is inspired by the idea that it may be possible to make content distribution faster and more efficient if user devices are enabled to seamlessly share previously downloaded content locally, rather than all content retrieval taking place through infrastructure networks. The mobility of wireless devices assists in geographic spreading of content and the physical proximity of device pairs allows one to use low-power short-range transmission, opening up the possibility for greater spectrum reuse in adjacent areas and more power-efficient networks.

Furthermore, content-centric P2P sharing schemes also hold the possibility of supporting content retrieval in environments where broader Internet connectivity is unavailable.

1.3 Aims, Relevance and Significance of the Thesis

1.3.1 Aims

Broadly speaking, the aim of this thesis is to explore the potential of wireless P2P content sharing as a mechanism for disseminating content quickly and efficiently throughout a network. We take special interest in *wireless* networks for two reasons: (i) there is a clear trend in computing towards wireless mobile devices and (ii) wireless networks free devices from having to route on top of artificially constrained wired network topologies. Through a series of papers, we seek to answer questions which include:

- How long does a message propagated via wireless P2P diffusion typically take to reach a given percentage of devices in the network?
- What impact does time of day, day of week, site and the number of diffusing sources have on the rate of information diffusion?
- How is rate of diffusion impacted when modelled in the context of real-world application usage and demand?
- How much uncertainty in content diffusion potential is induced by trace data sets with timing uncertainties?
- How much variation is there in the rate of content diffusion across individual source devices?

- How do the spatiotemporal preferences of nodes accelerate or impede content diffusion?

To address these questions, we draw upon an empirical trace of connections to Access Points(APs) in a large multi-campus university setting. From this trace we are able to infer encounters between pairs of wireless devices based on simultaneous presence at a given AP. This in turn allows us to simulate opportunistic content diffusion enabled by device encounters.

1.3.2 Relevance

The number of Internet-connected devices continues to experience rapid growth, as does the content demands being placed on computer networks. Devices are becoming increasingly mobile and increasingly wireless. In the past five years alone, smartphones and tablets have transitioned from early adopter status to ubiquity. This had led to increasing strain on wireless networks, many of which, it could be argued, are in a persistent state of underprovision. The next generation of devices such as consumer wearables, embedded devices and sensors will further drive content demand, as will the continual improvement of existing applications and device form factors (e.g. 4K video).

One of the most compelling features of a wireless P2P content-centric sharing scheme is that any realised performance or efficiency gains have a multiplier effect on lower level innovations in wireless telegraphy. Essentially this means that rather than competing with the inevitable future advances in physical and link layer technologies, wireless content sharing has the potential to offer persistent complementary performance improvements.

1.3.3 Significance

Though wireless P2P content sharing is a conceptually attractive approach to content distribution, little is known about the intrinsic potential and limits as they pertain to sharing facilitated by mobility and device encounters. Furthermore, little is known about what factors actually *drive* spreading potential (or lack thereof). P2P networks more generally have been studied for some time. However, only in recent times has the proliferation of mobile wireless devices opened up truly compelling possibilities around intermittently connected mesh-based (non-hierarchical) *colocation*-driven content distribution schemes—a specialized area that has received relatively little attention. Understanding the potential of mobile wireless P2P content sharing at the basic level of what mobility alone would facilitate is an important step in understanding the potential applications. Furthermore, by isolating which factors slow or accelerate content diffusion in colocation contact networks, we can better understand in what contexts wireless mobile P2P content distribution is likely to prove most useful and in what contexts it may prove most challenging.

1.4 The Thesis

1.4.1 Structure

This first chapter serves simply to introduce and broadly motivate the topic addressed in this thesis. Chapter 2 provides historical context around how the existing Internet architecture came to be, which motivates our subsequent discussion of ICN in Chapter 3. Chapter 3 also establishes the connection between ICN and mobile wireless P2P content diffusion. By the end of Chapter 3, the author hopes to have sufficiently contextualized and motivated the core of the thesis research which is presented in Chapters 4–7 in the form of one paper per chapter. These papers are introduced independently at the beginning of their respective chapters, however a brief overview follows:

1. *SPDYing Up the Web* [6]: published in the Communications of the ACM (CACM) in December 2012, this paper gives a technical introduction to the SPDY protocol describing the key performance features and reviewing publicly available results on the performance of SPDY versus HTTP 1.x. SPDY has since gone on to form the basis of HTTP 2.0 and will play a critical role in web performance for years to come. This paper is the result of a review of the state of the art in an important category of network protocols—those pertaining to the Web.

2. *Content Diffusion in Wireless MANETs: the Impact of Mobility and Demand* [7]: published in the proceedings of the IEEE International Wireless Communications & Mobile Computing Conference (IWCMC) in August 2014, this paper analyzes the impact of: (i) time of day, (ii) day of week, (iii) site and (iv) number of content sources, on content diffusion potential within university networks. This paper also analyzes content diffusion potential in the context of real-world application demand based on statistics collected from the author-written mobile campus navigation application *JCUNav*.

3. *The Impact of Mobility and Content Demand on Diffusion in Wireless MANETs*: submitted for publication in Elsevier Ad Hoc Networks, this paper is an extension of the earlier IWCMC paper that addresses trace uncertainties and their impact on diffusion potential as well as looking more closely at the impact of variation in diffusion potential across individual devices.

4. *Diffusion in Colocation Contact Networks: the Impact of Nodal Spatiotem-*

poral Dynamics: submitted for publication in PLOS ONE, this paper looks more deeply into the factors impeding and catalyzing spreading in our contact trace. We achieve this through the formulation of a set of 'inducement' shuffling null models which separately destroy correlations between times, locations and nodes. We apply these null models to our trace, allowing us to separate the impact of different correlations on spreading potential.

The four paper chapters are followed by Chapter 8 which concludes the thesis. This chapter consists of a summary of key results and contributions, challenges, limitations and a roadmap for future work. Chapter 9 is a glossary of terms. This is followed by Appendices A and B which list the research paper outputs and research by-products respectively. Finally, a list of references is given. Note that the reference list at the end of this thesis does *not* include the citations made within the four thesis papers, which are instead listed independently within each paper. This allows the papers to be included verbatim in the thesis without introducing ambiguous citation numbering.

1.4.2 Assumed Knowledge and Intended Audience

The author suggests this thesis should be classified as applied computer science with an emphasis on empirical data analysis. It is the author's expectation that a background in computer science or information technology will render this thesis most accessible. Though familiarity with modern computer networks may afford the reader some advantage, it is not a strict prerequisite. Aspects of this thesis also touch on subject matter in broader fields of scientific inquiry including network dynamics and even epidemiology, which may be of interest to a broader audience.

2

Historical Perspective and Motivation

"Mr. Watson—Come here—I want to see you."

— Alexander Graham Bell, 1876.

The road to the present-day Internet has been paved by a most fascinating period of invention over the past 200 years. The discovery of electromagnetism in the early 19th century [8, p. 1], one of the single greatest advances in electrical science, contributed directly to the invention of the electromagnetic telegraph in the 1830s and 1840s. This was followed by the invention of the telephone in 1876 which again revolutionized communications and was the first system to attract widespread consumer appeal. Throughout the 20th century the phone network continued to evolve with increasing levels of automation and connectivity. However, parallel advances were also being made in the stored program computer throughout the 1940s–1970s. These, along with the growing acceptance of the packet-switching network paradigm in the late 1960s and 1970s, laid the groundwork for the *ARPANET*. Far from a ubiquitous platform for high-speed content delivery, the ARPANET was designed as a way for scientists to share scarce resources and basic information between a limited number of machines (four initially). The ARPANET was significant in its own right as a working demonstration of packet-switched technology connecting multiple nodes, though it is just as notable for being the first implementation of the Transmission Control Protocol (TCP)/Internet Protocol (IP) protocols. At the time, the TCP/IP protocol was envisaged as a prototype. Instead, TCP/IP usage grew monumentally to become the standard transport and network protocols of the Internet.

Looking back over 200 years of seemingly disparate communication networks, it is revelatory to see that all of the networks employ the same abstraction of a connection between pairs of nodes. The electromagnetic telegraph made this unmistakably explicit by employing a static network of physical wires. The telephone changed the nature of the transmission but again was based around physical wires. Even as the telephone network advanced to incorporate human operated and later electromechanical and electronic switchboards, connections were simply being wired dynamically rather than statically. The Internet and packet-switching would again change the nature of the transmission but still retained the abstraction of a connection between hosts. Despite the highly dynamic routing capability of packet-switched technology, if anything the connection-oriented nature was made *more* explicit by the IP which encodes all packets with a source and destination address. So pervasive has been the connection-oriented nature of communication networks that is has arguably resulted in an invisible frame of reference, constraining how one can think about higher-level network abstractions. This chapter explains how this came to be, embedding our arguments in the zeitgeist of the respective eras. We look in some detail at both the telephone system and Internet and how both their predecessors and the technology of the time likely influenced their most basic architectural decisions. We crudely classify the telephone and Internet as circuit- [9, p. 282] and packet-switched [10, p. 180] paradigms respectively—a distinction we elucidate throughout this chapter.

The purpose of this chapter is to motivate the subsequent chapter on content dissemination which characterizes the present-day outlook of long-term networks research. This will in turn motivate our own more focused research questions, enumerated earlier in Section 1.3.1.

2.1 The Telephone System (Circuit Switching)

On March 7, 1876, Alexander Graham Bell was granted US patent 174,465—*Improvement in telegraphy* [11]. Bell's patent described a method for converting between acoustic vibrations (e.g. voice) and equivalent electrical undulations. Three days after the patent was granted, Bell had succeeded in translating his idea into the first working telephone[1]. The invention of the telephone came at a most interesting time in history. The electric telegraph (Figure 2.1) along with signalling standards such as Morse code [13] had already been in existence for several decades and yet had not achieved widespread consumer appeal. Furthermore, the telephone had managed to precede the invention of a practical light bulb, without which there was little impetus for electrical cables to be run to households. The challenge therefore of promoting the telephone was not only a matter of perfecting the technology but also laying the physical wire infrastructure to connect individual premises. In an 1876 letter sent from Bell to his father, Bell remarked: *"The day is coming when telegraph wires will be laid on to houses just like water or gas—and friends will converse with each other without leaving home."*. Bell's comments proved prescient as the telephone gained widespread adoption throughout the late 19th and 20th century.

Perhaps all but the very earliest phone networks employed circuit switching as a way to reduce the amount of physical wire needed to connect subscribers. As

[1] The invention of the telephone was hotly contested between Bell and Elisha Gray and remains the subject of much controversy [12].

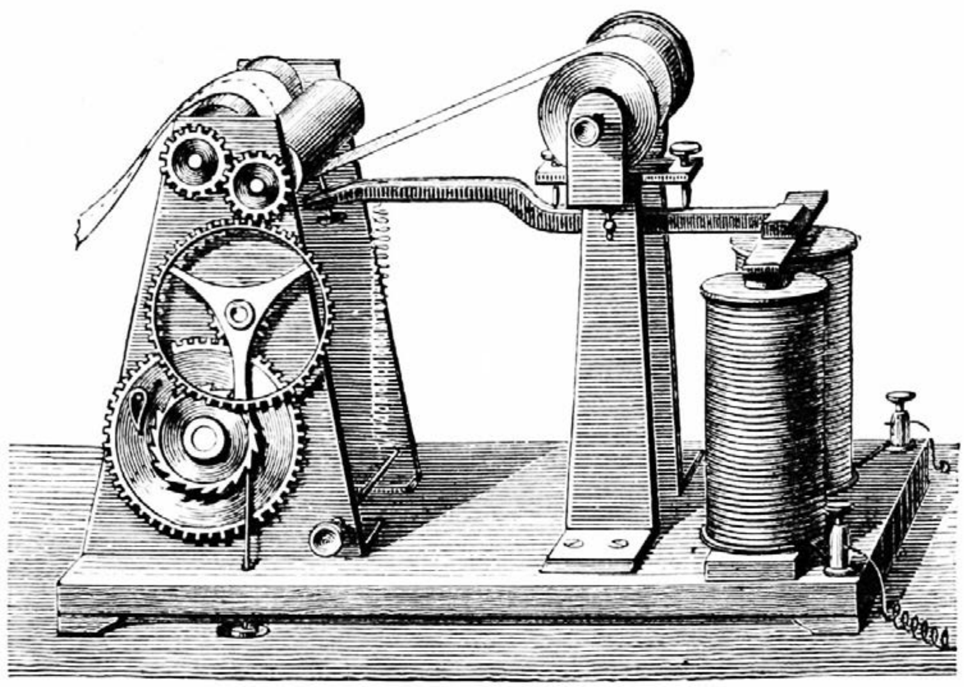

Figure 2.1: Morse telegraph [14].

illustrated in Figure 2.2, in its simplest distillation a circuit-switched network adheres to a star topology allowing connection of n subscribers using only $n \in \Theta(n)$ wires, rather than the well known $\frac{1}{2}n(n-1) \in \Theta(n^2)$ requirement of a pairwise connected network (i.e. a clique). At the central node in a circuit-switched network, wires between any two pairs of subscribers were dynamically interconnected to form a single continuous connection over which a phone call could take place. This central node became known as the *telephone exchange*.

Telephone exchanges in and of themselves have a rich history. The first commercial exchange began operating on January 28, 1878 in a storefront of the Boardman Building in New Haven, Connecticut [15]. At the time, interconnection between caller/callee pairs was a manual process mediated by a human switchboard operator. As illustrated in Figure 2.3, switchboard operators would literally place the right plugs in the right sockets to physically interconnect two subscribers' lines, completing an electrical circuit. Over time the role of the human telephone exchange operator would fall into obsolescence with the creation

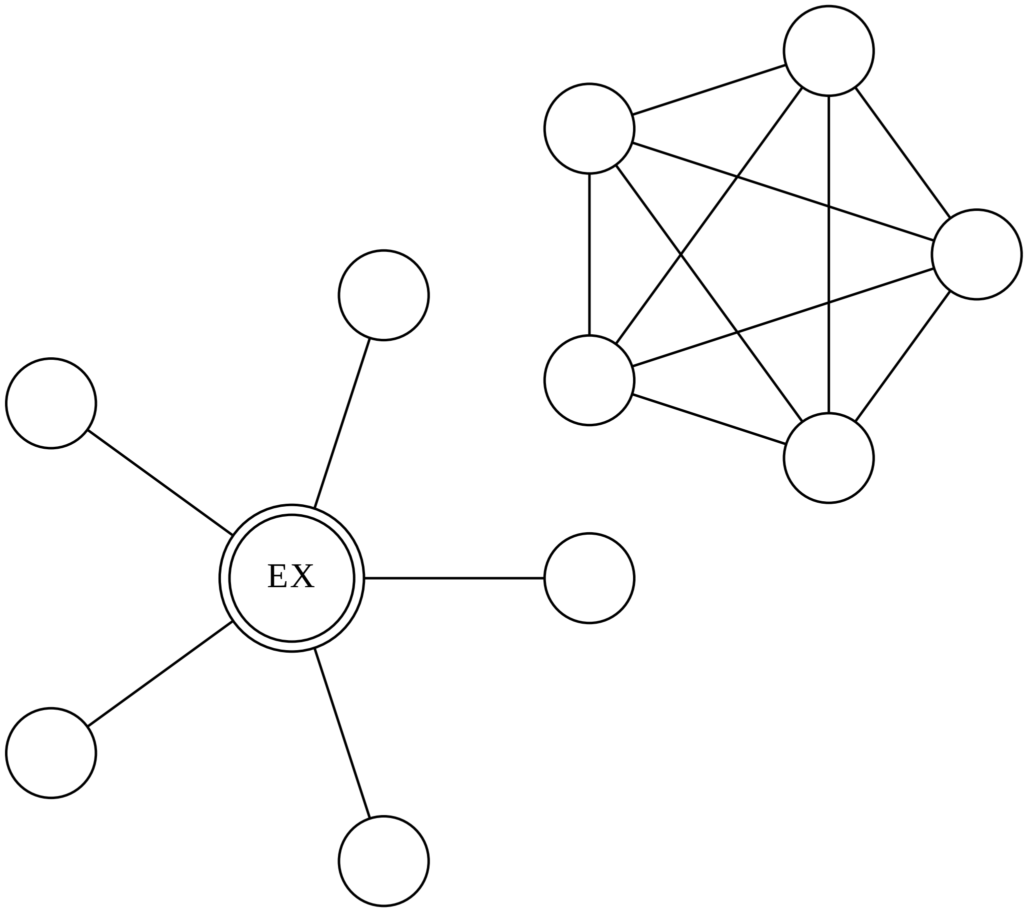

Figure 2.2: Clique (top right) and star (bottom left) network topologies. “EX” marks the telephone exchange in the star topology.

of the automatic telephone exchange.

It was March 10th, 1891 when Almon Strowger was awarded US patent 447,918—*Automatic Telephone Exchange* [17], describing an electromechanical device capable of connecting two subscriber lines without human intervention. Prior to the automatic telephone exchange, telephone handsets were not equipped with dialling pads. Instead, the caller would vocalise to the switchboard operator who they would like to connect to and the operator would fulfill the request by placing a plug in the appropriate jack. Strowger’s patent described how a phone with a dialpad could be used to ‘program’ a number into the telephone system, triggering an electromechanical switching device at the exchange to form a connection with the desired subscriber. Figure 2.4 depicts a Strowger switch which possesses a row-and-column head movement design that bears a striking resemblance to

Figure 2.3: Switchboard operators [16].

modern hard-disk drives (the mechanical platter variety, not solid state drives). With subsequent advances in electronics, electromechanical switches too were ultimately rendered obsolete, this time by digital switching technology still in use today.

What remained constant throughout the telephone's evolution, from manual to electromechanical to digital switching, was a focus on *building a wire* between pairs of subscribers, over which a call could take place. Given the time in history

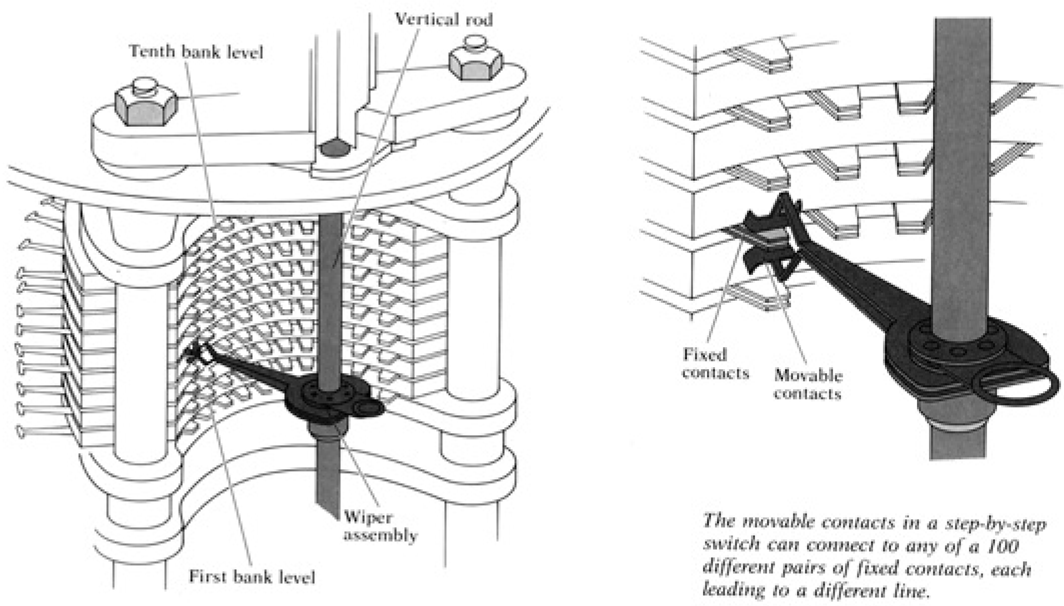

Figure 2.4: The Strowger Switch [18]—an electromechanical switching device.

in which the telephone was invented and the technology of the era, this is in fact largely unsurprising. Forming an end-to-end circuit was the immediate problem which begged to be solved in order to render telephony practical, much as it was with the earlier electric telegraph. The task of rolling out cabling to households was in itself monumental. The key innovation had been discovering a way to convert acoustic vibrations to electrical undulations (and back again) in real-time such that speech could essentially be transmitted over a wire. The notion of digital storage, buffering and flexible digital representations of information was almost certainly not being contemplated around the time of the telephone's invention. It was not until the second half of the 20th century that researchers would begin seriously considering an alternative networking paradigm with a view towards resilience in the face of component failure. This paradigm, *packet switching*, is explored in the following section.

2.2 The Internet (Packet Switching)

In September 1962, RAND Corporation engineer Paul Baran published what would become one of the most seminal papers in the history of the Internet, titled *On Distributed Communications Networks* [19]. Baran's paper introduced the concept of segmenting information into small, explicitly addressed pieces. Baran envisaged that each piece of information could then be routed from the source to the destination independently, potentially over different paths, based on distributed routing decisions. This networking paradigm came to be known simply as *packet switching*.

Baran's publication came during the Cold War and just a month prior to the Cuban missile crisis. A looming concern of the United States at the time was that neither the long-distance telephone plant nor the military's basic command and control network would survive a nuclear attack [20]. Baran's work on packet switching was therefore motivated by the need to build a distributed networking substrate that would be capable of withstanding a nuclear attack, by virtue of being able to route around destroyed links and nodes. Figure 2.5 is an illustration from Baran's paper, depicting three different network topologies Baran considered in his analysis of network resiliency. Under the centralized model (A), the network was highly vulnerable to complete failure, requiring only the destruction of a single node. The decentralized model (B) provided some further protection again complete network disconnection though remained susceptible to heavy partitioning with the destruction of only one or a small number of nodes or links. The distributed model (C) was determined the most resilient, requiring a substantial portion of nodes or links to be destroyed in order to cause even a serious partitioning of the network, let alone complete disconnection. Baran argued that the distributed model came with the virtue of providing a far more resilient computer network with only a modest number of additional links required.

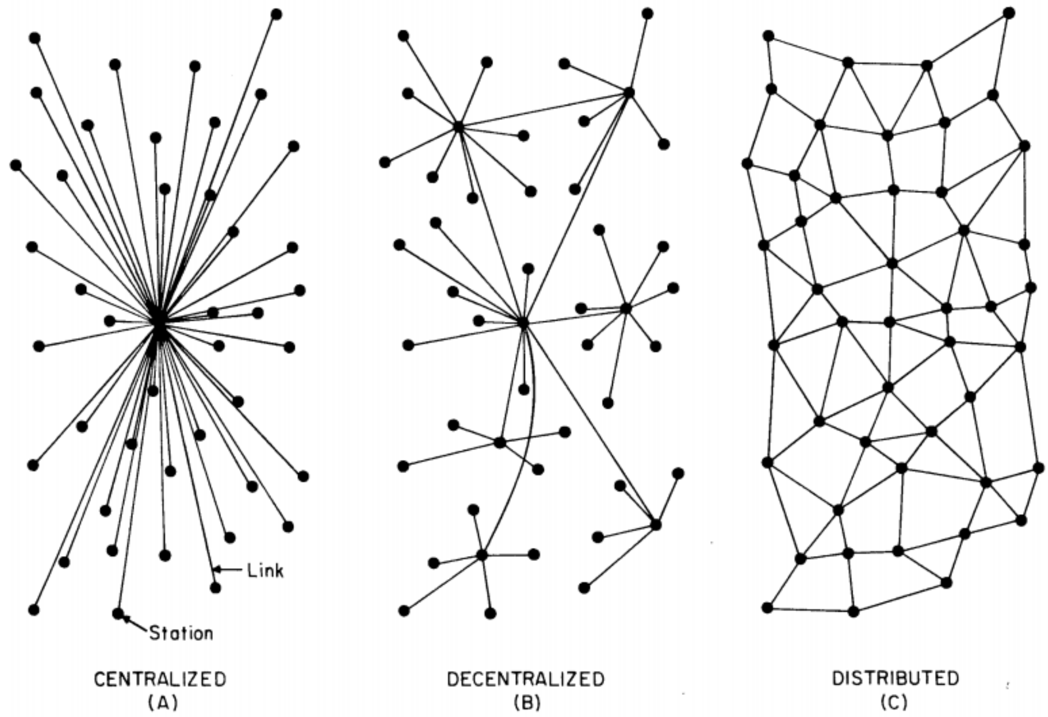

Figure 2.5: Three network topologies from Paul Baran's seminal paper [19].

At a time in which digital computer technology was itself nascent, Baran's work was not well received by the research community. Baran would later remark that *"Many of the things I thought possible would tend to sound like utter nonsense, or impractical, depending on the generosity of spirit in those brought up in an earlier world"* [21]. It would be another seven years before the first experimental implementation of a packet-switched network was created in December 1969—the *ARPANET* [22]. Though the ARPANET integrated the idea of segmenting data for independent routing across the network, it should be noted that the principal motivation was to accommodate inherently unreliable hardware, not to survive a nuclear attack [23].

Though the Internet does not have a well-defined creation date, one working definition is that the Internet came into existence on November 22, 1977 with the first successful demonstration of TCP/IP interconnecting three disparate networks, one of which was the ARPANET [22] (Figure 2.6). This paved the way for the broader interconnectivity of computer systems, with the first commercial

Internet Service Providers(ISPs) emerging in the late 1980s. The continuing unparalleled advancements in information technologies throughout the 1990s and beyond have ultimately led to the Internet of today connecting billions of devices globally.

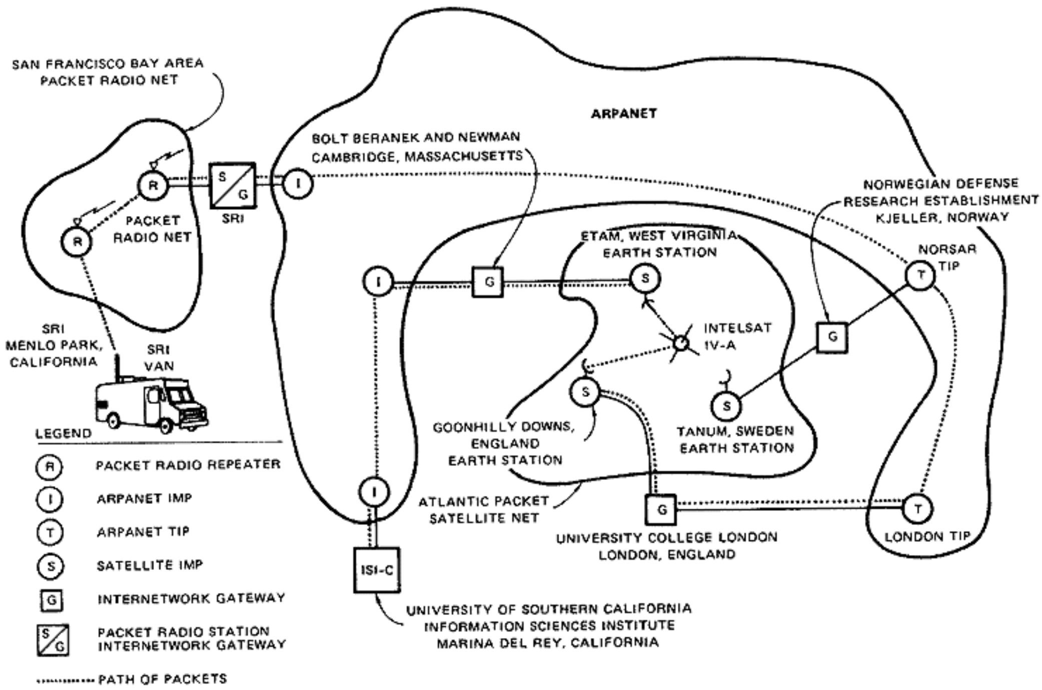

Figure 2.6: First ARPANET multinetwork demonstration [22].

In some respects the Internet is distinctly different from the traditional telephone system, particularly in its use of packet switching rather than circuit switching. Though originally conceived as a means of making network communications more resilient to nuclear attack, packet switching has delivered other advantages. One of the those has been the ability to build highly resilient networks out of redundant low-cost hardware instead of requiring high-cost reliable hardware to offset single points of failure. Another advantage has been the ability to exploit statistical multiplexing [24] to realize a desirable trade-off between variable delay and greater utilization of network capacity. Finally, packet switching can eliminate the need for an initial connection setup which can introduce a particularly large overhead on links with a high Bandwidth-Delay Product (BDP)[2].

[2]However, it is worth noting that in practice reliable transport protocols such as the TCP can negate this advantage.

Irrespective of the different paradigms of circuit and packet switching, what both the telephone and Internet share in common is *an abstraction of a connection between hosts*. In the circuit-switched phone system the connection was implicit in the physical wire link. Although the packet-switched Internet delegated *routing* decisions to the network, the notion of host endpoints prevailed, this time explicitly in the form of IP addresses that travel with each packet. In this way the Internet moved the intelligence of routing into the network itself while retaining pairwise host conversation as the universal communication primitive. As previously discussed, designing the telephone network around circuit switching makes sense when one considers the limited technology of the time and the immediate problem which required solving—the physical interconnection of premises. In much the same way, host-addressed packet switching made sense in the context of the state of the art in technology at the time and the problem computer networking aimed to solve, namely, resource sharing.

Computers in the 1960s and 1970s were expensive, specialized and scarce. Data was minimal and simply having the ability to interconnect machines on disparate networks was viewed as a major achievement. The relatively small number of machines combined with their specialized features rendered host addresses a useful task-oriented abstraction—the user first needed to connect to machine X to perform operation Y. Contrast this to the state of present day computing, with billions of cheap, commoditized, Internet-connected devices. The value of the present day Internet is largely derived from acting as a ubiquitous substrate for massive-scale content distribution and retrieval. This transformation in the Internet's dominant usage and proliferation of commoditized machines has rendered individual machines largely inconsequential to the task of the user. Much of what is distributed and retrieved over the Internet today is massively replicated inside of server farms for load balancing, as well as geographically distributed for availability and performance. The task-oriented user now thinks in terms of names, not ultimately caring from where the associated content is retrieved, provided it is an authentic copy of that which is requested. The way the mismatch between hosts and named content has been handled to date is to build overlay networks running on top of the IP [25]. Perhaps the most conspicuous example of this phenomenon is the World Wide Web in which URLs are used to express the desired content. The Domain Name System (DNS) system maps the identifiers (domain names) to locations (IP addresses) to facilitate content retrieval based on the URL. Over the past few years the research community has increasingly been contemplating whether the dominance of content distribution and retrieval warrants a network layer abstraction which speaks *directly* in terms of named content, rather than relying on separate application-specific overlays. In the next chapter we review the current state of research in this area and establish the connection between this work and mobile wireless Peer-to-Peer (P2P) content dissemination.

3

Information-centric Networking (Dissemination)

The evolution of the Internet from a platform for interconnecting scarce hosts to a substrate for content distribution and retrieval has enlivened work in the research community around Future Internet Architectures(FIAs). FIAs broadly speaking are network architectures, some incremental and some clean slate, designed to meet the computer networking needs of the 21st century. Much of the work around FIAs falls under the umbrella title of Information-Centric Networking (ICN). The term Information-Centric Networking (ICN) encapsulates many competing projects and visions of a future Internet that share the common thread of viewing information (i.e. content) as the primary network abstraction, rather than host addresses. There is all but consensus in the ICN community that content identity and security ought to be independent of content location, enabling seamless content replication throughout the network. This would permit secure content retrieval from a cache geographically and/or topologically close to the end user while balancing load for the content distributor. What separates the current wave of ICN research from the myriad of application-specific distributed caching overlays is a generalization which intends to handle content distribution and retrieval at the network layer directly.

In this chapter we provide a brief introduction to ICN by way of exemplification. That is, we choose one instantiation of ICN known as *Content-Centric Networking (CCN)* [1][1] to describe in some detail. For a broader exposition of ICN we refer the interested reader to [27]. We then briefly relate ICN to Delay-Tolerant Networking (DTN) and Mobile Ad hoc NETwork (MANET)s, areas of computer networking to which our research is naturally aligned, before proceeding to the four core research papers in the subsequent chapters.

3.1 CCN Background

CCN is one proposal for an ICN architecture which treats content names rather than host addresses as the primary network abstraction. The original proposal for CCN can be found in the now well known paper, *Networking Named Content* [1]. CCN in its purest form proposes that named 'content chunks' completely replace Internet Protocol (IP) packets as the 'thin waist' (universal layer) of the network stack, which we describe in more detail in the following section. These packets are cryptographically signed in such a way that the content may be verified as having not been tampered with even when retrieved from an untrusted host. In this thesis we choose to elucidate CCN specifically for three reasons:

1. CCN is emblematic of ICN. Although each ICN project has different implementation and protocol specifics, CCN captures the principal architectural philosophy.

2. CCN is one of the most prominent ICN projects at time of writing, with support from the National Science Foundation (NSF) [26], Palo Alto Re-

[1] Content-Centric Networking (CCN) is also the technical basis of the Named Data Networking (NDN) [26] project. For the purposes of this thesis CCN and Named Data Networking (NDN) can be considered equivalent.

search Center (PARC) [28] and a backing consortium consisting of industry members including Cisco Systems, Alcatel-Lucent and Verisign [29].

3. We frame our own research in the context of the CCN signed packet primitive. This allows us to presume that trustworthy content is retrievable from untrusted hosts, meaning *our* research focus can be concentrated on intrinsic diffusion potential, not the specifics of content security.

In the following sections we describe the core components of the CCN architecture and its applicability to Peer-to-Peer (P2P) content dissemination.

3.2 CCN Network Stack

As illustrated in Figure 3.1, CCN replaces host-addressed IP packets at the 'thin waist' of the network stack with chunks of named content. Devices wishing to retrieve some piece of content request it from the network *directly* by name, without having to perform the translation from *what* the user cares about (named data) to *where* that data is stored (i.e. IP host addresses). Also illustrated in Figure 3.1, above the thin waist CCN does not use a traditional transport layer protocol and instead has a security layer which provides a stronger notion of content integrity (described shortly in Section 3.3). Because CCN names data directly, many of the existing application layer naming schemes are rendered redundant, and so a 'File Stream' (i.e. reconstitution of content chunks into meaningful data objects) is shown as the application layer abstraction. Below the thin waist, the CCN 'Strategy' layer handles the forwarding decisions of interests in content chunks.

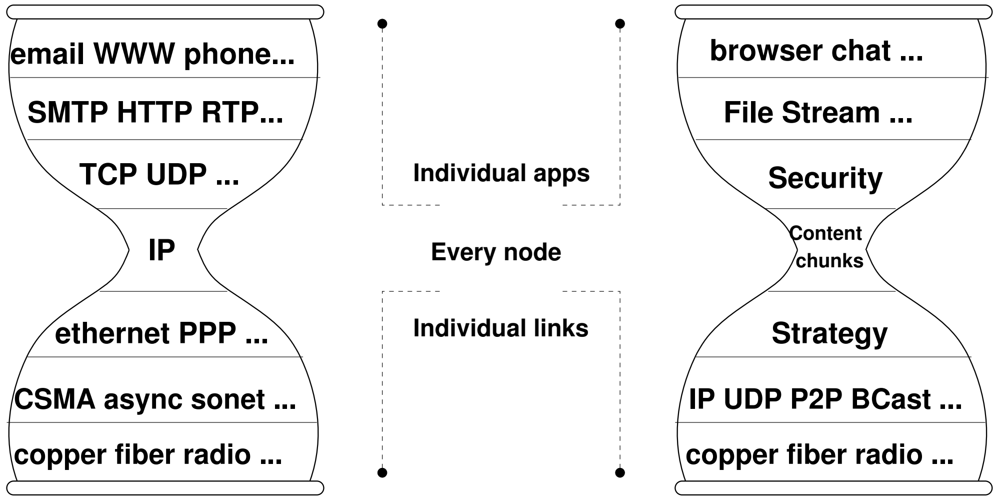

Figure 3.1: CCN replaces host-addressed IP packets, the current universal component of the network stack, with chunks of named content [1].

3.3 CCN Packet Primitive

As illustrated in Figure 3.2, the two CCN packet types are *interests* and *data*. An example content name is shown in Figure 3.3. The desire to receive some content is expressed by transmitting an interest with the appropriate content name, along with some associated metadata. Importantly, the data packet response includes not only the content name and its associated data, but also a cryptographic signature of the *binding* between the name and its data. This signature allows the receiver to verify that the received data is a complete, uncorrupted copy of what the publisher sent (*validity*), that the publisher is one the receiver trusts (*provenance*) and that the content is the answer to the question asked (*relevance*). The interested reader is referred to [30] for an extended discussion of securing named content.

3.4 CCN Forwarding Engine

Figure 3.4 illustrates the key components of a CCN 'router'. Any networked device can act as a router under CCN, *even devices typically not thought of as*

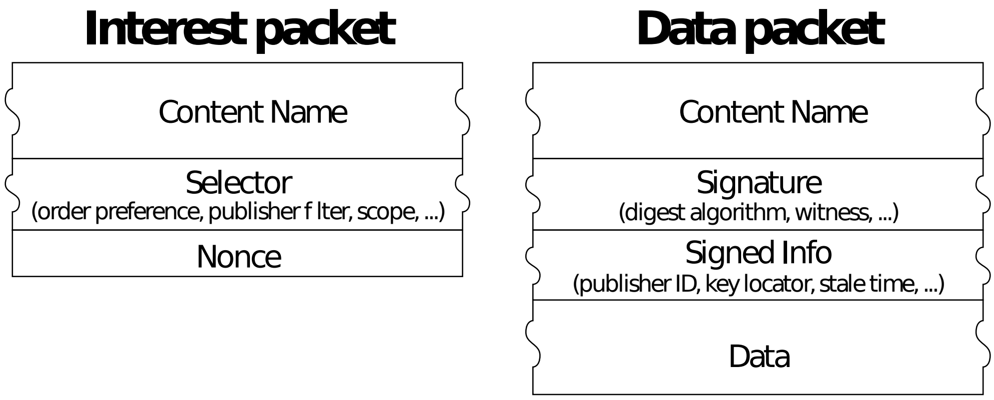

Figure 3.2: CCN packet types [1].

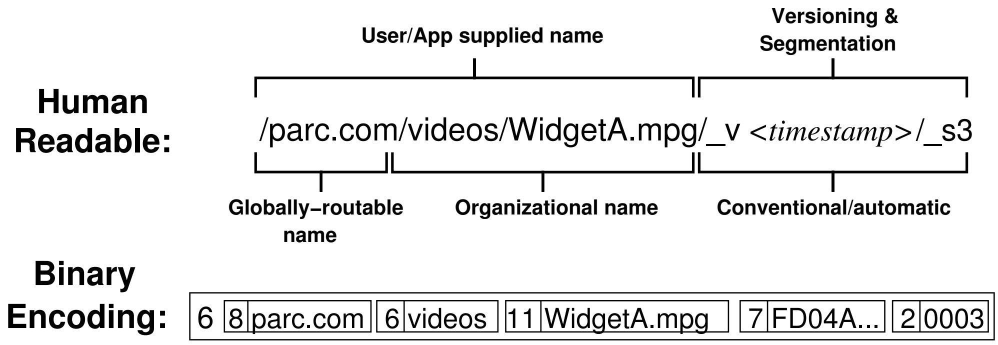

Figure 3.3: Example Content Name [1].

content sources, such as peer wireless mobile devices. The content store is essentially a cache which can store any seen content at a packet granularity. The Pending Interest Table (PIT) holds a record of all requests the router has seen, but has not yet fulfilled. Finally, the Forwarding Information Base (FIB) holds a record of which faces[2] content for a given prefix may be accessible over. Note that unlike IP, CCN is not restricted to forwarding interests over a single interface, allowing seamless multi-source content retrieval.

[2]The term *face* rather than *interface* is used because packets are not only forwarded over hardware network interfaces but also exchanged directly with application processes within a machine.

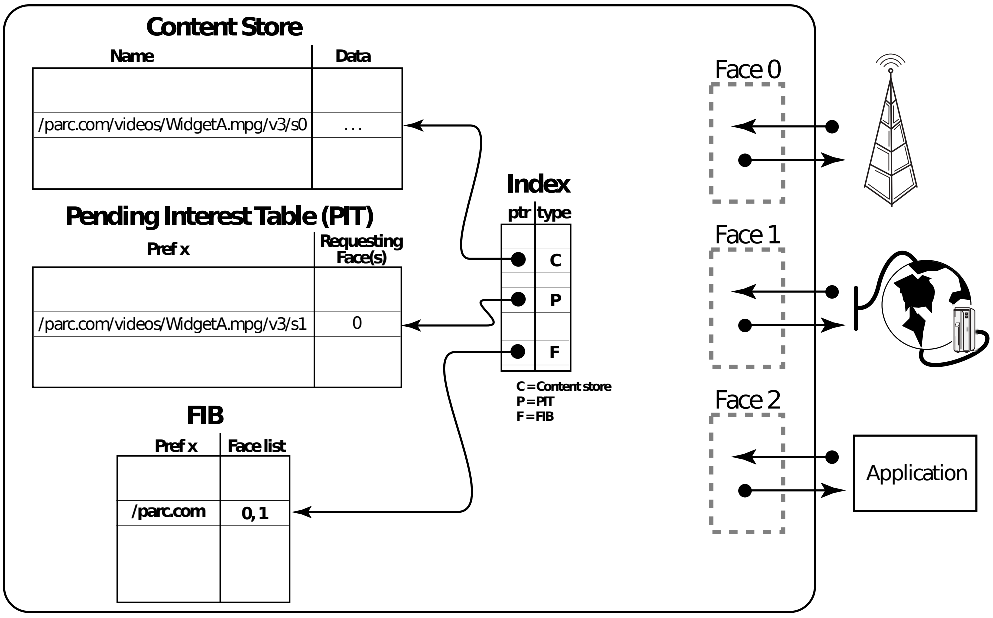

Figure 3.4: CCN forwarding engine model [1].

3.5 CCN as a Framework for High-Speed Peer-to-Peer Content Dissemination

CCN provides a multifaceted image of how a named and secure content primitive may be used to improve the state of content distribution and retrieval. On one hand CCN offers to existing hierarchical network infrastructures a standardized way to cache fine granularity content. On the other hand CCN seeks to open up new possibilities around seamless non-hierarchical multi-source content retrieval, including retrieving content from *peer* devices not typically thought of as potential sources. The secure content primitive of CCN is ideal for such networks as it enables trust in content retrieved from untrusted hosts. Furthermore, the content-centric nature of the network primitive allows P2P content dissemination to be contemplated independently of application-specific protocols. It is for this reason that CCN offers an attractive conceptual framework for thinking about wireless P2P content dissemination.

Studying P2P networks built on top of IP, one inevitably ends up having to consider practical security implications of how to retain content integrity, as well as how application content should be translated down to host-addressed packets. By presuming CCN's secure content chunks as the network primitive, the intrinsic potential of P2P content diffusion can be confidently studied in the knowledge that the prerequisites for secure content dissemination are already in place.

3.6 Delay-Tolerant Networking and Mobile Ad hoc NETworks

Delay-Tolerant Networking (DTN) is the term coined by Kevin Fall in 2003 [31] to describe networks with non-continuous end-to-end connectivity, relying on asynchronous forwarding to propagate messages between nodes. DTN is intrinsically a form of ICN, and CCN is considered one promising protoocol for DTN's widespread realization. The term 'Mobile Ad hoc NETwork (MANET)' is synonomous with the term 'mobile wireless P2P network' and we use the two interchangeably throughout the remainder of this thesis. This will be particularly apparent in the second and third paper. This change in terminology is simply to bring our work in line with the preferred terminology of the respective publication venues. DTN and MANET are closely related, in that MANET is perhaps the most conspicuous example of DTN in the literature. Portions of this thesis' related work have roots in DTN and MANETs and so this final section of the chapter briefly relates the topics to our own research.

A substantial fraction of DTN research has focused on mobility models and routing. The former aims to distill the salient features of real mobility traces into formulas and rules for the purpose of characterization and application to new environments. The latter focuses on the strategy behind message propagation, particularly where the objective is for a specific subset of nodes in the network to receive a message efficiently. We refer the interested reader to [32] and [33] for a survey of mobility models and routing respectively. Our own research is not concerned with modeling mobility or routing per-se. In fact, we draw upon empirical rather than synthetic mobility data and assume content is shared *every* time two devices come into contact. The motivation for our own research is understanding what fundamental limits real-world contact patterns place on the speed at which content is capable of broadly diffusing. Specifically, we seek to answer the research questions enumerated earlier in Section 1.3.1, which have not received the same level of attention in the literature to date.

In addition to mobility modelling and routing, there exists a number of other research areas closely affiliated with DTN and MANETs. These include characterizing the behavioral patterns of individual mobile users [34–37], the classification of mobile users into distinct behavioral groups [38, 39], analysis of encounters between mobile devices [40–43] and studies of the intrinsic content dissemination potential of intermittently connected networks predicated on these encounters [42, 43]. The preponderance of our own research falls into the last category, with the single closest work being that of Hsu & Helmy in *On Nodal Encounters* [42]. In this paper, the authors studied how incremental constraints on participation in content sharing would impact the rate of spreading in a wireless LAN environment. This was quantified in terms of the percentage of nodes in the network yet to receive a message after a given period of time. Our own research tracks a similar reachability metric, but instead quantifies the impact of a different set of variables including the time of day, day of week and number of source devices on content diffusion. Furthermore, we extend our work to a generalized framework for analyzing the impact of nodes' spatiotemporal preferences on spreading phenomenon in spatiotemporal graphs.

By this point in the thesis, the author hopes to have adequately motivated wireless mobile P2P content diffusion and the research questions outlined earlier in Section 1.3.1. The lacuna in the literature the author aims to address has been alluded to briefly in this section and will become more clear in the following four chapters, each of which presents one of the four research papers outlined previously in Section 1.4.

4

Paper 1 – SPDYing Up the Web

The first peer-reviewed article of the thesis appeared in the December 2012 edition of the Communications of the ACM (CACM) and is titled *SPDYing Up the Web*. SPDY is a 'drop-in' replacement for HTTP with a number of key performance improvements. This paper was the first formally published work to (i) provide a consolidated technical summary of the differences between SPDY and HTTP and the underlying performance motivations, (ii) summarize publicly available results on SPDY's empirical performance versus HTTP and (iii) discuss the compelling use cases around SPDY gateways.

At face value, this paper (the "SPDY paper") is an outlier with respect to the remaining three research papers, which all focus directly on content diffusion between mobile wireless devices. To help contextualize the SPDY paper, it is instructive to consider more broadly the motivation behind the author's research; namely, faster and more efficient content delivery. Approaches to networking performance improvements exist along a spectrum from evolutionary to revolutionary. SPDY falls into the evolutionary category, while content diffusion directly between mobile wireless devices assumes more revolutionary change to the way that content is delivered. The SPDY paper offers a complementary perspective on how progress towards faster and more efficient content delivery can be achieved in the short-term without any fundamental changes to the prevailing networking paradigm. From a pragmatic perspective, such incremental improvements offer meaningful value over the short-term while the longer-term solutions mature.

The SPDY paper developed naturally as a consequence of a broader review of the networking performance literature near the beginning of the author's doctorate studies. Two additional factors which served as a catalyst for the SPDY paper were:

1. The thesis author has a long standing interest in Web performance. In 2010 the thesis author published his honours thesis which presented a Peer-to-Peer (P2P) content distribution network focused on fast lookup and retrieval of web resources. The author continued to follow the state of the art in performance improvements for the Web and so was cognizant of the developments around SPDY at the time.

2. The original PhD research plan involved performing Deep Packet Inspection (DPI) in order to identify duplicate content byte strings across mobile devices. Matching these byte strings with time-respecting paths between pairs of peer devices was originally envisaged as a mechanism for calculating what percentage of load might be taken from the infrastructure. Although the content-level dataset ultimately proved unattainable, part of the background research entailed monitoring trends in end-to-end encryption, which, if prevalent enough, might preclude DPI. As explained in the paper, for both pragmatic and philosophical reasons SPDY typically operates over SSL. As such, the increasing popularity of SPDY might have materially affected the broader prevalence of end-to-end encryption and in turn adversely impacted the thesis author's original research plan. Coincidentally, the prevalence of end-to-end encryption has increased substantially in more recent times [44], though this is likely due only in part to SPDY. Other

influencing factors have likely included broader awareness of information privacy and revelations around government mass surveillance of computer networks.

Although the preponderance of the author's research pertains to long-term network architectures, the author believes the SPDY paper offers a valuable point of contrast between short-term and long-term approaches to improving network performance. Since the publication of this paper, SPDY has gone on to become the basis for the version 2.0 specification of the HTTP protocol. SPDY's underlying principles are therefore likely to be in increasingly widespread use on the Web in years to come, albeit under a different name.

The included copy of the manuscript which follows is reprinted with permission. The DOI is `http://dx.doi.org/10.1145/2380656.2380673`.

DOI:10.1145/2380656.2380673

Improved performance and a proven deployment strategy make SPDY a potential successor to HTTP.

BY BRYCE THOMAS, RAJA JURDAK, AND IAN ATKINSON

SPDYing Up the Web

This publication has been removed due to copyright restrictions

5

Paper 2 – Content Diffusion in Wireless MANETs: the Impact of Mobility and Demand

The second peer-reviewed article of the thesis appeared in the Proceedings of the 10th IEEE International Wireless Communications & Mobile Computing Conference, 2014 (IWCMC '14) and is titled *Content Diffusion in Wireless MANETs: the Impact of Mobility and Demand*. The purpose of this paper was twofold:

1. The thesis author had hypothesized that time of day, day of week, site and number of content sources were likely real-world factors that would influence rates of diffusion. Although theoretical rates of epidemic content diffusion between wireless mobile devices had previously been measured, these parameters had been largely overlooked.

2. The thesis author sought to embed content diffusion simulations in the context of realistic application usage.

To address the former, extensive universal content diffusion simulations were performed, varying the parameters of (i) time of day, (ii) day of week, (iii) site and (iv) number of content sources. To address the latter, application-specific diffusion simulations were performed. The application in question was *JCUNav*, a popular campus maps application which the thesis author developed throughout the PhD, with the explicit intent of gathering large-scale empirical data. As is described in the paper, the usage patterns, reflected in terms of distributions of the number of daily users and times of application usage, were gathered from JCUNav and used to parameterize the content diffusion simulations. More information about JCUNav can be found in Appendix B.1.

The included copy of the manuscript which follows is reprinted with permission. The DOI is `http://dx.doi.org/10.1109/IWCMC.2014.6906485`.

Content Diffusion in Wireless MANETs: the Impact of Mobility and Demand

Bryce Thomas, Ian Atkinson
Department of Information Technology
James Cook University
Townsville, Australia
bryce.m.thomas@gmail.com, ian.atkinson@jcu.edu.au

Raja Jurdak
Autonomous Systems
CSIRO
Brisbane, Australia
raja.jurdak@csiro.au

6

Paper 3 – The Impact of Mobility and Content Demand on Diffusion in Wireless MANETs

The third peer-reviewed article of the thesis has been submitted for consideration to Elsevier Ad Hoc Networks and is titled *The Impact of Mobility and Content Demand on Diffusion in Wireless MANETs*. This paper builds directly on the prior paper, addressing two distinct areas not covered previously:

1. *Uncertainties* in session trace timestamps can lead to different rates of content diffusion, depending on the assumptions made. By developing synthetic 'optimistic' and 'pessimistic' traces we are able to illustrate the magnitude of this difference.

2. *Variation* in diffusion performance across individual devices means that the average rate of diffusion is only part of the story. We illustrate this variation across individual diffusion trials.

In addition to covering these two new areas, this paper also provides a generally more complete exposition and background than was possible within the page limits of the prior publication.

The included copy of the manuscript which follows is the authors' preprint.

The Impact of Mobility and Content Demand on Diffusion in Wireless MANETs

Bryce Thomas[a,*], Raja Jurdak[b], Ian Atkinson[c]

[a] *Department of Information Technology, James Cook University, Townsville, Queensland, Australia*
[b] *Autonomous Systems, CSIRO, Brisbane, Queensland, Australia*
[c] *eResearch, James Cook University, Townsville, Queensland, Australia*

Abstract

Opportunistic wireless content sharing via Mobile Ad hoc NETworks (MANETs) can increase throughput, lower latency, extend network coverage and reduce load on infrastructure. While the benefits of content diffusion clearly depend on the underlying movement dynamics and content demand, the impact of these factors on diffusion remains largely unexplored. We analyze content sharing potential based on device encounters inferred from a large multi-site wireless LAN trace. We explore the impact of time, location, and number of sources on diffusion, finding that contexts with higher activity generally promote faster diffusion, while additional content sources improves diffusion mainly in the short-term. We then apply real-world demand patterns from a popular campus maps application to content diffusion simulations. We find that up to 70% of map requests could theoretically be served from the peer network over the first 12 hours. Finally, our analysis of the impact of trace uncertainties and individual device variation on diffusion potential reveals large differences based on the selected assumption and chosen source devices. We discuss these results and their implications for wireless peer-to-peer networks.

Keywords: manets, content-diffusion, contact-networks, spatiotemporal-networks, network-simulation

1. Introduction

Enabling wireless user devices to directly share common-interest content is a conceptually attractive approach to enhancing wireless networks. Each user device caches content retrieved from the infrastructure and makes it transparently available to colocated peers, either pre-emptively or on demand. Devices' content demands are preferentially served from a nearby peer with the infrastructure serving as a fallback when a cached copy is unavailable. The potential benefits of such a scheme include higher throughput, lower latency, greater spectrum reuse, extended network coverage and reduced load on infrastructure.

1.1. Motivating Example

We present a mobile map sharing application as a motivating example. Suppose User A is using their mobile device to navigate a geographic region after having downloaded the region's map from the infrastructure (e.g. a cell tower or wireless access point). Now suppose User B enters the same region and encounters User A. User A proceeds to pre-emptively share the map data with User B. Shortly afterwards, User B would also like to view a map of the region. Rather than having to retrieve the mapping data from the infrastructure, User B already has a local copy available received earlier from User A. We highlight several potential benefits of this peer sharing:

- Being in close geographic proximity allows the devices to transmit at lower power, reducing battery consumption and increasing opportunities for spectrum reuse in adjacent areas.
- User A and B can establish a short-range dedicated connection, increasing throughput. This is particularly important if User B were to retrieve the map on demand, rather than receiving it pre-emptively.
- The devices can communicate with very low latency as a result of the short-range nature of the connection and because the devices are not contending with other devices for access to the infrastructure. Again, this is important for on-demand retrieval.
- If User B is not in range of the infrastructure, User A effectively extends User B's coverage by making otherwise unreachable content available.
- Finally and in many cases most importantly, load has been taken off the fixed wireless infrastructure. Wireless infrastructure and cellular data infrastructure in particular is often viewed as being in a perpetual state of underprovision. Partially offloading content delivery from the infrastructure onto a Mobile Ad hoc NETwork (MANET) may prove a useful strategy for reducing the necessary cost or frequency of infrastructure upgrades.

*Corresponding author
Email addresses: `bryce.m.thomas@gmail.com` (Bryce Thomas), `raja.jurdak@csiro.au` (Raja Jurdak), `ian.atkinson@jcu.edu.au` (Ian Atkinson)

Continuing the maps example, assume that some time later User A transitions to a new geographic region. As a result of A's mobility, maps of the prior region are now available to devices in the new region. This is an example of how content may spread with the aid of device mobility.

We have presented mapping as just one motivating example of MANET-based content sharing via diffusion. The use cases of content diffusion however generalise to any application premised on or enhanced by the ability to move content quickly and efficiently. Content diffusion may prove particularly useful for other applications which like maps exhibit locality of reference [1] in content interests, i.e. content interests tend to be spatially and/or temporally correlated. This includes web content, app content and even personal area networks (PANs) where a single user carries multiple cloud-connected devices synchronizing identical data.

1.2. Contributions

Though wireless peer-to-peer (P2P) content sharing presents as an intellectually attractive approach to improving network efficiency and performance, a lacuna exists in the literature around real-world parameters influencing content diffusion potential. Existing works [2, 3] explore some facets of epidemic content diffusion including the resulting network topologies and diffusion potential under various constraints on participation. Our earlier work in [4] provides a preliminary examination of how site, time of day, day of week, number of content sources and empirical patterns of content demand influence content diffusion potential in wireless LANs. In the present paper we build on our prior work by analyzing the impact of *uncertainty and variation* in trace-driven diffusion simulations. We find diffusion potential to be relatively sensitive to the assumptions chosen to compensate for inherent timing uncertainties in wireless LAN traces. We also find a relatively large amount of variability in diffusion potential between individual content source devices. We discuss currently accepted assumptions of the research community as they pertain to inferring device encounters and highlight why verifying the validity and then perhaps improving these assumptions would be beneficial.

1.3. Paper Structure

The following section covers related work. Section 3 provides background information on the area of content diffusion and formally defines how device encounters are inferred from wireless LAN traces. Our primary wireless LAN trace, the *UQ trace*, is described in Section 4, along with its uncertainties. Our first set of simulations analyze *universal diffusion* on the UQ trace, i.e. how quickly an arbitrary piece of content might spread throughout a network. These simulations are described in Section 5 and the results are presented in Section 6. Our second set of simulations focus on a realistic *application-specific* use-case for content diffusion—diffusing electronic maps. Our application-specific simulations are based off both the UQ trace and empirical real-world "app" usage statistics collected from the separate *JCUNav trace*. The JCUNav trace and our application-specific simulations are described in Section 7. Section 8 provides a discussion of our findings regarding the impact of trace uncertainties and presents avenues for future work. Section 9 concludes the paper.

2. Related Work

Our work fits broadly into the existing body of research around MANET [5] communications and Delay Tolerant Networking (DTN) [6]. Though present-day device and protocol support for seamless device-to-device communication is somewhat deficient, we are particularly motivated in our analysis by promising next generation protocols like Content-Centric Networking (CCN) [7]. The pertinent feature of CCN (and similar protocols) is enabling trustworthy content to be retrieved from untrusted hosts.

Most directly related to our work are empirical studies of device mobility and encounters and the ad hoc content diffusion opportunities these create. Eagle & Pentland [8] recorded 9 months of Bluetooth encounters of 100 mobile devices given to students and faculty at MIT university. Wang *et al.* [9] recorded 3 days of Bluetooth encounters of 41 "iMote" devices given to participants at the 2005 Infocom conference. Su *et al.* [3] recorded device encounters of two groups of students given PDAs, each group being around 20 students in size and the two experiments lasting 2.5 and 8 weeks respectively. Hsu & Helmy [2] analyzed device encounter patterns in traces collected from four university campuses and the Infocom 2005 conference.

Of the aforementioned works, [2] and [3] explicitly analyzed ad hoc multi-hop message dissemination facilitated by device mobility and encounters. Our own work compliments these prior studies by i) analyzing site, time of day, day of week and number of content sources as diffusion parameters and i) providing new findings on application-specific diffusion, trace uncertainties and diffusion variation. Furthermore, we perform our simulations using a late 2012 trace, which compared to traces used in past studies is i) substantially newer (in some instances over a decade so) ii) larger in size and iii) collected with greater temporal and procedural consistency across sites.

A number of other studies [10, 11, 12, 13] have characterized wireless network usage and user behavioural patterns. In addition to these, there have been a multitude of works on mobility models intended to describe the movement of devices in space and time, many of which are reviewed in [14]. Again our work is complementary to these studies, though we focus specifically on information diffusion potential in the context of empirical data, not network characterization or mobility modelling.

3. Background and Definitions

3.1. Opportunistic Mobile Content Diffusion

Opportunistic mobile content diffusion refers to the dissemination of content directly between mobile devices during incidental encounters, i.e. where and when opportunities naturally arise. Content may originate directly from a device or have been downloaded from an infrastructure network at an earlier point in time. For example, a sensor reading may originate from a mobile device, while a cached web page originates from an Internet-connected infrastructure network. Once one or more mobile devices possesses a given piece of content, that content can be shared directly with other mobile devices. These other devices may then further propagate the content causing a (time respecting [15]) transitive spread of content throughout the network. Even a device with no interest in a piece of content may act as a data mule [16] who receives, caches and then further propagates the content during subsequent opportunistic encounters.

3.2. Ideal Diffusion

We define *ideal diffusion* as a special case of opportunistic content diffusion that takes place *every time* an opportunity arises. Essentially this is a form of flooding—each time two devices encounter, they share with one another their respective contents.

3.3. Universal Ideal Diffusion

One of the simplest questions that can be asked about ideal diffusion potential is *what is the maximum percentage of all devices in a network that an arbitrary piece of content might reach after a given amount of time?* Universal ideal diffusion (referred to simply as “ideal diffusion” from hereon forward) can be simulated by firstly selecting a start time and assigning one or more devices as content “sources”. These sources then act as origins of diffusion, sharing content with each encountered device. At each time step where either a device enters the network for the first time or content is shared, the percentage of devices in the network which have received the content is recalculated. Later in Section 5.2, we formally define the *unreachable ratio* which measures the proportion of devices in the network yet to receive the diffusing content.

3.4. Application-specific Diffusion

While universal diffusion gives a broad idea about the intrinsic diffusion potential of a network, it is also possible to analyze diffusion potential in the context of real-world application demand. In this paper we define application-specific diffusion simulations to be those which account for realistic patterns of content demand, both in absolute scale of interested users and the times at which content is desired. Though not considered in this paper, application-specific diffusion simulations may model other factors such as willingness to participate and minimum connection durations required for various content transfers to take place. Later, in Section 7.2, we formally define the *cache miss ratio* as our metric for measuring application-specific diffusion potential. This describes the proportion of *interested* devices in the network which successfully received the desired content from the P2P network, i.e. without having to resort to the infrastructure.

3.5. Wireless LAN Trace-Driven Simulations

In this paper we focus on understanding the content diffusion potential of large Wireless Local Area Networks (WLANs) based on trace-driven simulations. To be of use in diffusion simulations, a wireless LAN trace should for each session that has taken place in the network include a record of i) connection time ii) disconnection time, iii) a unique access point (AP) identifier and iv) a unique user device identifier. From these records it is possible to infer encounters between user devices by identifying concurrent connectivity of devices to a given access point.

3.6. Wireless LAN Encounter Definition

In WLAN traces, mutual transmission range may be approximated by simultaneous connectivity of a and b to a given AP. We follow below with a formal definition of encounters in the context of WLAN traces where encounters are inferred based on concurrent connectivity to a static intermediary (i.e. the AP):

Let $I_{d,p} = \{[j_{d,p,1}, k_{d,p,1}], ..., [j_{d,p,n}, k_{d,p,n}]\}$ be the set of intervals during which device d was connected to access point p, where $k_{d,p,i} < j_{d,p,i+1}$. We then define the encounter set between devices d and e at p as:

$$E_{d,e,p} = \bigcup I_{d,p} \cap \bigcup I_{e,p} \tag{1}$$

As an example, suppose devices d and e were connected to p for intervals $\{[10, 20], [25, 30], [32, 45]\}$ and $\{[18, 22], [41, 60]\}$ respectively. Then:

$$\begin{aligned}
I_{d,p} &= \{[10, 20], [25, 30], [32, 45]\} \\
I_{e,p} &= \{[18, 22], [41, 60]\} \\
E_{d,e,p} &= \bigcup \{[10, 20], [25, 30], [32, 45]\} \cap \bigcup \{[18, 22], [41, 60]\} \\
&= \{10...20, 25...30, 32...45\} \cap \{18...22, 41...60\} \\
&= \{18...20, 41...45\}
\end{aligned}$$

indicating d and e encountered at p during the interval set $\{[18, 20], [41, 45]\}$.

Our encounter definition is equivalent to that used by Hsu & Helmy in [2] and is only an approximation of actual encounters. The first key assumption is transitive reachability, i.e. if devices d and e are in transmission range of AP p, then d and e are in transmission range of each other. The second key assumption is that d and e never encounter at p unless both are simultaneously connected to p. Clearly these assumptions do not capture verbatim real-world encounters—devices connected to the same AP may not be in mutual transmission range, devices connected to different APs may be in transmission range and devices may encounter one another outside of the range of APs. Though imperfect, our encounter definition serves as a useful approximation and is consistent with the earlier work of Hsu & Helmy in [2]. Throughout this paper we will however draw attention to the sensitivity of diffusion results as they pertain to assumptions about *other* sources of uncertainty. In doing so we highlight why encounter definitions and other uncertainties still ought to be validated and improved upon accordingly by the broader research community.

4. The UQ Trace and Uncertainties

4.1. The UQ Trace

The UQ trace is a record of all IEEE 802.11 (Wi-Fi) Access Point (AP) sessions collected from the multi-site University of Queensland (UQ) wireless network between Nov. 27–Dec. 11, 2012. The trace contains 549,002 sessions from 23,931 unique MAC addresses connecting to 3,081 APs across 24 discrete geographic sites. Sites include university campuses, hospitals, research stations and AP installations at other UQ-affiliated locations throughout the state of Queensland, Australia. Each record in the trace corresponds to a single session whose details include i) connecting MAC address, ii) AP name, iii) site name, iv) session start time and v) session end time.

Most of the 24 sites in the UQ trace are relatively small with fewer than 50 APs. As our primary interest in this paper is content diffusion potential at large sites, we limit our analysis to the 5 sites with 50 or more APs. Our analysis excludes one unknown "site" with 337 APs known as "Root Area". The Cisco Network Control System Configuration Guide [17] suggests that Root Area is a default label applied to APs which do not belong to a particular site or at least have not had any site-specific label applied. The session volume over time for each of the 5 selected sites is illustrated in Figure 1 and each site's numeric properties and general characteristics are summarized in Table 1. For convenience, Table 1 includes the derived ratios `MACs:APs`, `sessions:MACs` and `sessions:APs` which we refer to when describing our results in section 6.

4.2. Session Timeframe Uncertainties

A degree of uncertainty exists in the start and end times of sessions in the UQ trace. The first cause of this uncertainty is a trace collection infrastructure which samples and timestamps information about users connected to each access point periodically rather than instantaneously. The second cause of uncertainty arises from the fact that the collection infrastructure times out users after 30 minutes of inactivity, though does not explicitly record in which sessions this timeout has occurred. For content diffusion analyses in Sections 6 and 7 of this paper, we present our findings under both *optimistic* and *pessimistic* session length assumptions which take into account these uncertainties.

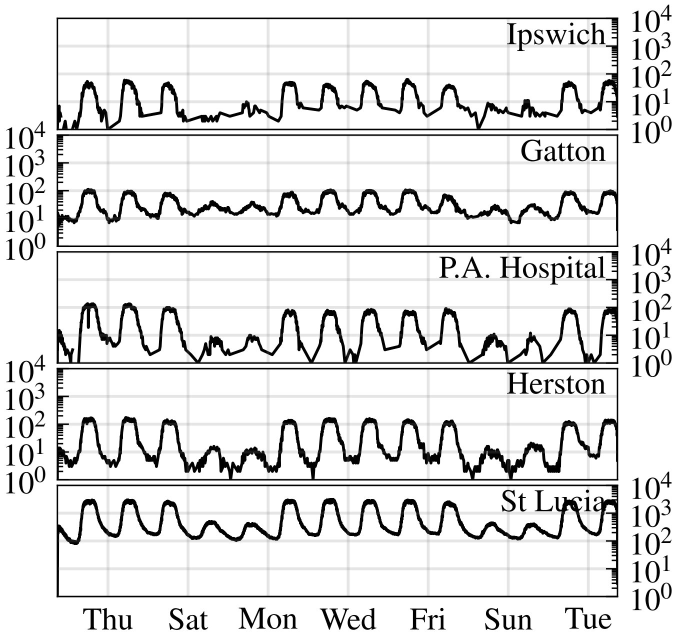

Figure 1: Per-site session volume.

4.2.1. Periodic Timestamping

UQ deploy Cisco APs which are centrally managed by a Cisco Network Control System (NCS) [17]. The NCS periodically polls APs for information about currently connected users. Importantly, the NCS does not use precise timestamp information from APs about the time individual users connect or disconnect. Rather, the NCS applies its own current timestamp at the time the data is recorded. This implies that session start and end timestamps which appear in our trace are greater than or equal to the true time at which the corresponding event occurred. More formally, for a session recorded as spanning the time interval $[u, u']$, the real session time interval is $[v, v']$ such that $v \leq u$ and $v' \leq u'$. Based on empirical observation, we add the further constraint that $u \leq v'$, leaving us with $v \leq u \leq v' \leq u'$. The subtle implication of this constraint is that a session which both starts and ends inside of a single reporting interval never appears in our trace. When analyzing the trace, we noticed that very short sessions never occurred. We conjecture that internally the Cisco NCS compares an AP's connected users across consecutive reporting intervals to infer which users have disconnected during the interim. When a user connects and disconnects during a single reporting interval, neither report witnesses the connection and so the session is never recorded.

The NCS uses a nominal polling interval of 5 minutes. Reporting is a low priority task competing with other tasks for computational resources and so some variation exists around the nominal 5-minute interval. The nature of the

Table 1: Properties of the analyzed sites.

Site Name	MACs	APs	Sessions	MACs:APs	Sessions:MACs	Sessions:APs	Environment
St Lucia	20 339	2 005	448 136	10.14	22.03	223.5	Large university campus
Gatton	731	258	13 867	2.83	18.97	53.75	Medium university campus
Herston	1 323	115	19 066	11.50	14.41	165.79	Medium university campus on hospital grounds
Ipswich	469	167	5 736	2.80	12.23	34.35	Medium university campus
P.A. Hospital	782	92	12 095	8.50	15.47	131.47	Hospital

Table 2: Optimistic and pessimistic session length adjustments.

Adjustment	Source of uncertainty: Periodic Timestamping	Source of uncertainty: Connection Timeouts
Opt. start	-10 minutes	-
Opt. end	-	-
Pess. start	-	-
Pess. end	-10 minutes	-30 minutes iff session > 30 minutes

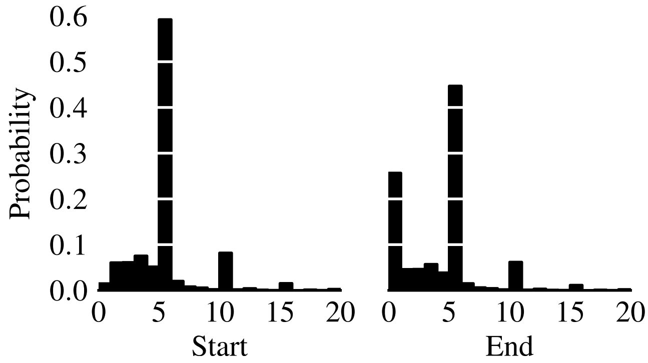

Figure 2: Gap sizes (minutes) between chronologically consecutive start (left) and end (right) timestamps in the source trace. Timestamps with zero gap omitted.

trace makes it impossible to precisely determine the time period between any two consecutive reports. This is because i) no explicit report ID is recorded in the trace and ii) a single report may take on the order of seconds to complete, resulting in sessions with different timestamps even within a single report. Therefore it can be uncertain whether sessions with close but different timestamps belong to the same or different reports. We can however determine the distribution of gap sizes between all chronologically consecutive session start or end timestamps to get an approximate idea of typical reporting intervals. Figure 2 is a histogram of the non-zero gap sizes between chronologically consecutive timestamps in our trace. As can be seen, gap sizes are typically on the order of 5 minutes, with some variation. Gap sizes of 1 minute or less are likely sessions being recorded as a part of a single report, while gap sizes between 1 and 5 minutes may either result from a single slow report or commencement of a new report. We note additional smaller peaks around 10 and 15 minute gap sizes. We suggest such peaks may be caused by low traffic periods during which not a single new user connected or disconnected from the network during a given reporting interval. This would result in one or more empty reports causing the gap size between consecutive timestamps in the trace to widen to approximately some multiple of 5 minutes.

Based on the gap sizes in Figure 2, our first step in deriving pessimistic and optimistic traces from the original trace is to make the following adjustments:

- *pessimistic*: *subtract 10 minutes from reported session end times, leave reported session start times intact.* Subtracting 10 minutes from the reported session end time ensures the derived session will in the majority of cases end at a time prior to when the session truly ended. Leaving the reported session start time as-is ensures that the derived session starts at least as late as the session truly started.
- *optimistic*: *leave reported session end times as is, subtract 10 minutes from reported session start times.* Leaving the session end time as-is ensures the derived session ends at least as late as the real session. Subtracting 10 minutes from the reported session start time ensures the derived session will in the majority of cases start at a time prior to when the session truly started.

4.2.2. Connection Timeouts

The second form of session duration uncertainty is caused by timed out connections—some 802.11 devices will occasionally fail to explicitly disconnect from the network upon leaving. The Cisco hardware from which our trace is derived disconnects such users from the network automatically after a 30-minute window of inactivity. For those users who have timed out, we would like to subtract 30 minutes from the reported session end time. Unfortunately, our trace does not distinguish between users who have explicitly disconnected from the network and those which have timed out. For sessions longer than 30 minutes in duration, there is therefore no way to tell whether the user explicitly disconnected from the network or was subject to the 30-minute timeout. Again, we make session start and end time adjustments to derive pessimistic and optimistic traces:

- *pessimistic*: *for all sessions reported as longer than 30 minutes in duration, subtract 30 minutes from the reported end time.* Subtracting 30 minutes from the end time of all sessions longer than 30 minutes

Algorithm 1 Universal Diffusion Simulations.

```
function RUN_UNIVERSAL()
    sites = {St. Lucia, Gatton, Herston,
        Ipswich, P.A. Hospital}
    times = {Wed 12:06PM Nov 28,
        Thu 04:52AM Nov 29,
        Sat 03:38PM Dec 01, Sun 04:56AM Dec 02}
    sources = {1, 2, 4, 8, 16}
    for ∀⟨s, t, u⟩ ∈ {sites × times × sources} do
        SIMULATE(s, t, u)
    end for
end function

function SIMULATE(site, start, sourceCount)
    for i = 1 to 50 do
        sourceMACs = RANDSOURCES(site, start,
        sourceCount)
        SIMULATEDIFFUSION(site, start, sourceMACs)
    end for
end function
```

ensures that the duration of any timed out session is not overestimated. The side effect is that any session longer than 30 minutes which did *not* timeout also has its duration shortened in the derived trace.

- *optimistic*: *leave all session end times as is.* Leaving session end times as-is ensures the derived sessions end at least as late as the real sessions ended. The side effect is that sessions which did timeout are overestimated in duration by 30 minutes.

We summarize all optimistic and pessimistic session adjustments in Table 2.

5. Simulating Universal Diffusion

5.1. Simulation Overview

We perform multi-site, multi-source simulations for a variable number of source devices, variable diffusion start times and under both pessimistic and optimistic session length assumptions. Our simulation models ideal content diffusion by means of Discrete Event Simulation (DES) implemented as a set of custom Shell, Python and Go scripts. In total we perform 10,000 universal content diffusion simulations. This entails simulating all combinations of 5 sites, 5 quantities of content source devices, 4 diffusion start times and 2 assumptions. For each combination, we perform 50 trials ($5 \times 5 \times 4 \times 2 \times 50 = 10,000$), where each trial elects a random set of devices to act as content sources. The *RUN_UNIVERSAL()* function in Algorithm 1 summarises this procedure which is run over optimistic and pessimistic input traces separately.

The 5 simulated sites are those shown in Table 1. As previously mentioned, these are the five largest sites in the UQ trace. The 4 diffusion start times are chosen to occur during traffic periods corresponding to i) a weekday peak ii) a weekday trough iii) a weekend peak and iv) a weekend trough. These times are summarized in Table 3. Each simulation commences with 1, 2, 4, 8 or 16 selected devices

Table 3: Diffusion start times and traffic level (concurrent sessions) they represent.

Time	Traffic Characteristic
Wed 12:06PM, Nov 28	Weekday Peak
Thu 04:52AM, Nov 29	Weekday Trough
Sat 03:38PM, Dec 01	Weekend Peak
Sun 04:56AM, Dec 02	Weekend Trough

as content sources. Though source devices are selected at random for each of the 50 trials, they are subject to the constraint of having to be present in the network (connected to an AP) at the relevant diffusion start time. This ensures diffusion commences concurrently from all source devices. Note that for any single trial, source devices are sampled without replacement and so each source device is unique. Across multiple trials however, source devices are sampled *with* replacement. Therefore $|F \cap F'| \geq 0$ for source device sets F and F' sampled for two different trials.

An event in our DES is when a device either connects to or disconnects from an AP. When a connection event occurs, we record the device as connected and look for other devices simultaneously connected to the same AP. If the device which has just connected possesses the content being diffused (either because it's a source device or has received it from someone else), it shares the content with all simultaneously connected devices at the same AP. If a device already connected to the AP possesses the content, then that device shares the content with the newly connected device. When a disconnection event occurs, we remove the record of the device being connected to the AP.

5.2. The Unreachable Ratio

The *unreachable ratio*, coined by Hsu & Helmy in [2], is the name of the metric used to describe the percentage of all devices in a network yet to receive a piece of content being diffused. The unreachable ratio is defined as:

$$u = \frac{(|A| - |B|) - (|C| - |B|)}{|A| - |B|} \qquad (2)$$

where A is the set of all devices seen since diffusion began, B is the set of source devices and C is the set of all devices that have received or always possessed a copy of the diffusing content.

The unreachable ratio changes over time and is recalculated whenever a new device enters the network or content is shared with a device. When a device enters the network for the first time, the unreachable ratio increases. When a device receives content, the unreachable ratio decreases. Note that a device's exit from the network does *not* affect the unreachable ratio—the unreachable ratio is calculated over all devices seen so far, not all devices instantaneously connected.

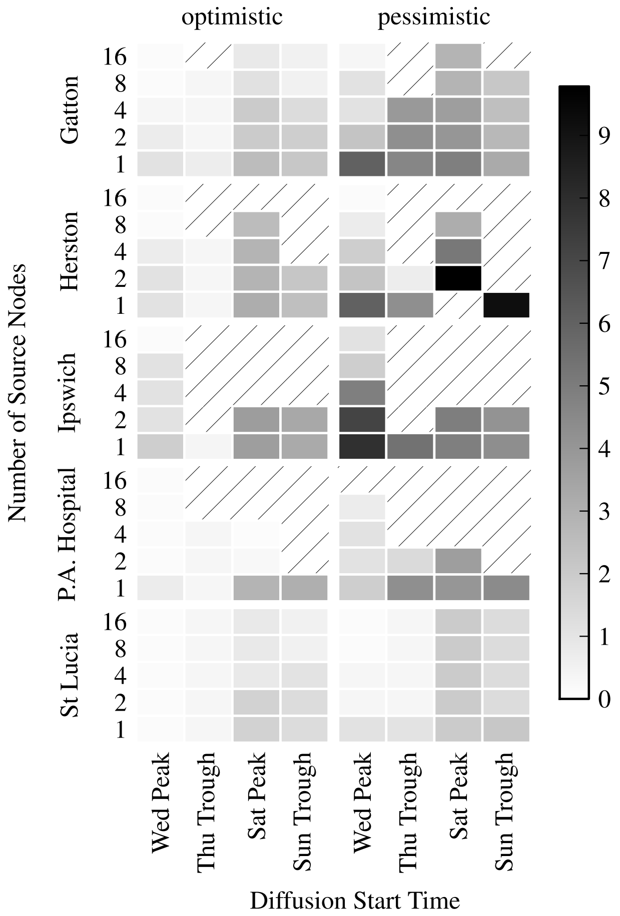

Figure 3: Time taken to reach a 50% unreachable ratio (days). Striped squares indicate insufficient source devices were available to run the simulation, with the exception of Herston with 1 source device/pessimistic assumption/Saturday peak which simply never reached the 50% unreachable ratio.

6. Universal Diffusion Results

6.1. Results Presentation Overview

Throughout this section, we refer to Figures 3, 4, 5 and 6 to illustrate our findings.

Figure 3 is a heatmap of the time taken for the unreachable ratio to drop to 50% under all combinations of the simulated parameters. The purpose of Figure 3 is to provide a coarse summary measure of diffusion performance—the time taken for diffused content to reach half of all devices in the simulated network.

Figure 4 and Figure 5 depict the unreachable ratio over time for each site using different combinations of diffusion start time and number of content sources. The results in Figure 4 and Figure 5 are based on the previously defined pessimistic and optimistic assumptions respectively. Whereas Figure 3 presents a coarse measure of diffusion (time to 50% unreachable ratio), Figure 4 and Figure 5 offer a more detailed view of the progression of information diffusion over the simulated period. The unreachable ratio as presented in each line in Figure 4 and Figure 5 is an average calculated over the 50 trials of information diffusion we perform for each combination of ⟨site, session length assumption, diffusion start time, number of content sources⟩.

Figure 6 is designed to quantify the *variation* in diffusion performance across individual trials. That is, whereas Figure 4 and 5 illustrate the overall expected level of diffusion potential, Figure 6 highlights how some individual devices can be more effective at diffusing content than others. All results in Figure 6 are based on simulations conducted using a single source device starting at the Weekday Peak time (see Table 3).

6.2. Analysis Across Simulated Parameters

6.2.1. Influence of Site

The most obvious finding in Figure 4 and Figure 5 is that the rate of information diffusion is dependent on the site analyzed. Recall that all site traces were collected in a uniform time period, under a single administrative domain, are all from 802.11 Wi-Fi networks and were all processed in the same manner. The difference in rate of diffusion cannot therefore be discounted as caused by heterogeneous trace sources. It is not completely clear what the dominant drivers are behind this variation, though we follow with a preliminary hypothesis.

St Lucia, by far the largest site, has a very strong tendency to outperform other sites in content diffusion under all parameter combinations, with only a small number of exceptions in the first few days of diffusion. St Lucia also has the highest ratio of `sessions:APs` and `sessions:MACs` and the second highest ratio of `MACs:APs`, as seen in Table 1. All else being equal, higher values for these three ratios would increase the rate of information diffusion, as they imply higher levels of network activity. We therefore offer the hypothesis that St Lucia demonstrates superior diffusion capability as a result of either its generally higher rate of campus activity or larger size. Ipswich, the smallest site as measured by both unique MACs and number of sessions, has a relatively strong tendency to underperform other sites in information diffusion with a few exceptions. Ipswich also has the lowest ratios of `MACs:APs`, `sessions:APs` and `sessions:MACs`. Again, all else being equal, these lower ratios would adversely affect diffusion performance. As such, we offer the hypothesis that Ipswich demonstrates inferior diffusion capability either as a result of its generally lower rate of campus activity or smaller campus size. We acknowledge that the size/ratios hypothesis alone is not enough to fully explain the observed behaviour and that further investigation is needed to discover other contributing factors. For example, the relative diffusion performance of P.A. Hospital, Gatton and Herston shows less uniformity across simulation parameters, even though these three sites vary substantially in size and ratios as shown in Table 1.

6.2.2. Influence of Number of Source Devices

Intuitively, increasing the number of devices acting as a content source increases the rate at which content diffuses

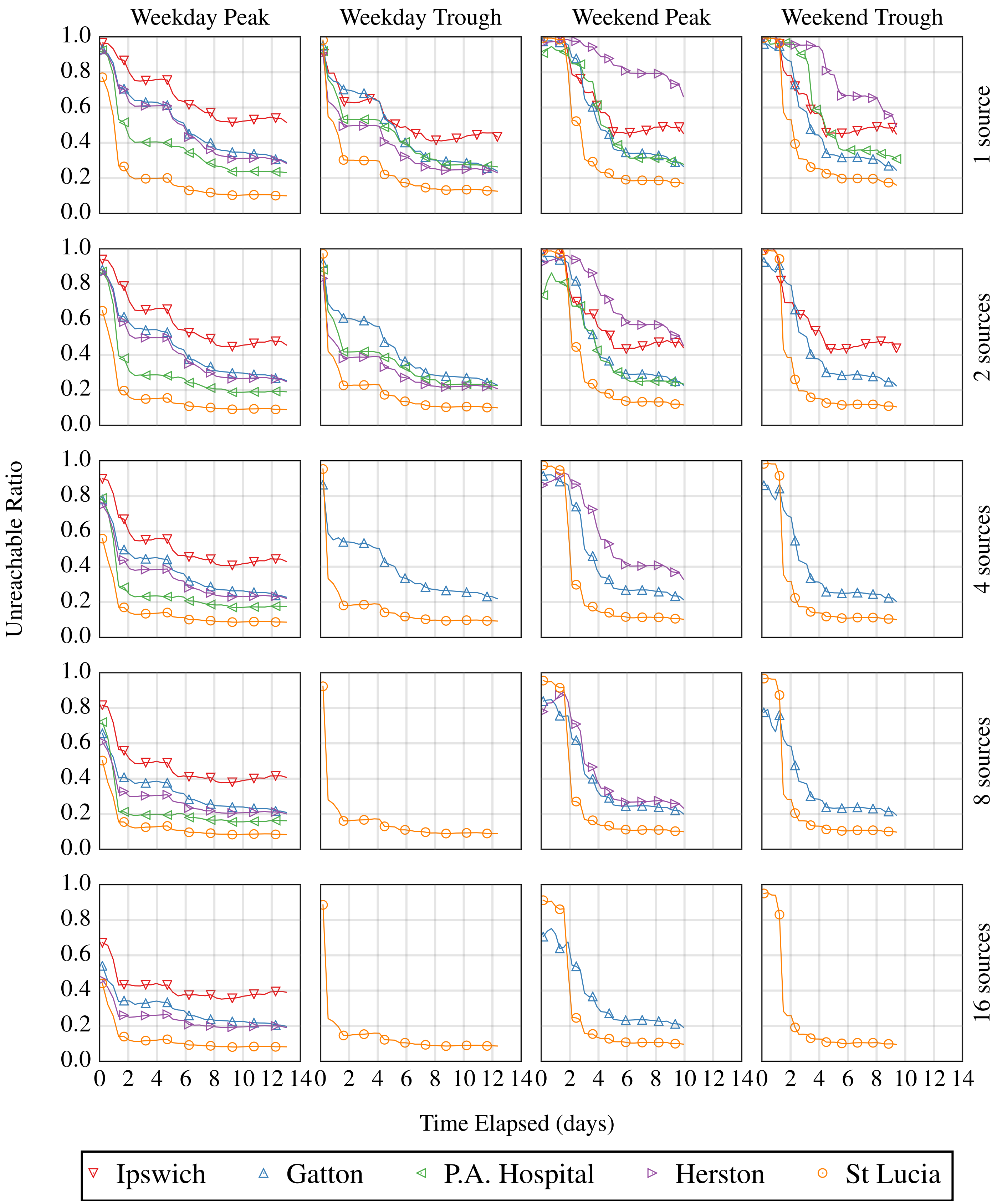

Figure 4: Unreachable ratio based on number of source devices and diffusion start time (pessimistic).

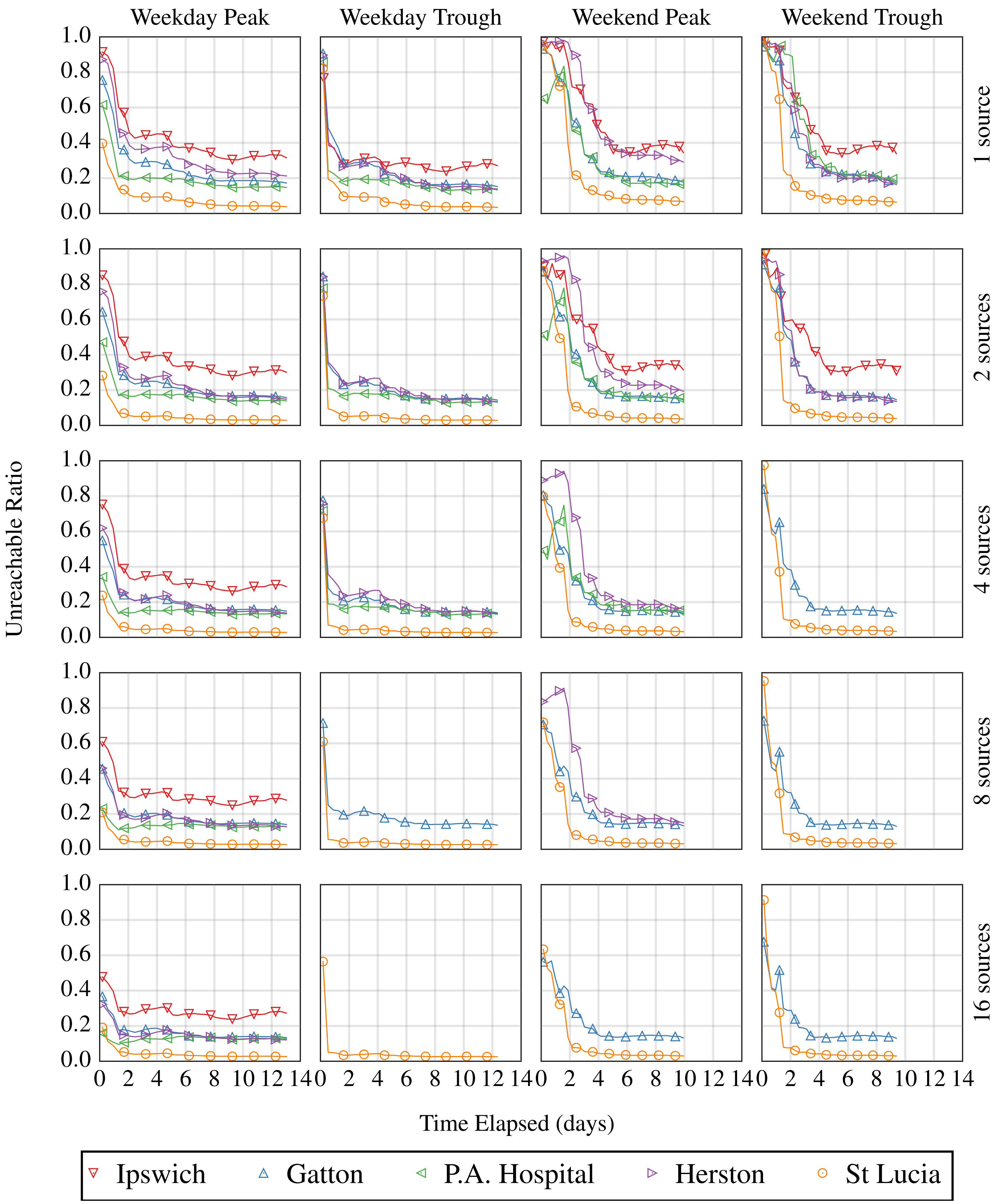

Figure 5: Unreachable ratio based on number of source devices and diffusion start time (optimistic).

throughout the network. In our simulations, the change in rate of information diffusion as a function of using a higher number of source devices is in fact monotonically non-decreasing. This is because the source devices used in a simulation with i source devices are a subset of those used in the otherwise same simulation with j source devices, where $i < j$.

We note that additional source devices often make a marked difference on the rate of diffusion, particularly over the short-term. Over the longer term, we observe that the number of source devices has relatively little influence on diffusion potential and is often negligible by the end of the trace period. This finding suggests that much of the benefit of additional source devices is in the form of content reaching devices sooner, though most of these devices would receive the same content in due course with fewer sources, albeit not as quickly.

6.2.3. Influence of Day and Time

Day and time appears to affect the rate of diffusion differently depending on site and number of source devices. For example, by comparing across individual rows in Figure 3 one can observe that there is no strict ordering of light and dark cells which applies to all rows. One pattern we did observe in Figure 4 and Figure 5 is that when diffusion begins on a weekend there tends to be more activity in the upper left hand corner of the subplot. We conjecture that the lower session volume during the weekend period (see Figure 1) translates to fewer opportunities for content to diffuse and so the rate of diffusion tends to remain low over the weekend. As a rule, diffusion tends to take longer to reach the 50% unreachable ratio (Figure 3) when starting on weekends, though this pattern is not universal.

6.2.4. Influence of Session Length Assumption

The general patterns of content diffusion are comparable between simulations performed over the pessimistic and optimistic traces. Comparing Figure 4 and Figure 5 side by side, we do however note meaningful absolute differences in rates of diffusion, particularly over the short-term. This finding suggests that assumptions around trace uncertainties may not drastically affect the general diffusion behavior, though may meaningfully bias absolute results.

6.3. Diffusion Potential Variation Across Devices

Figure 6 demonstrates substantial variation in rates of diffusion across randomly selected source devices. We note to the reader that due to a flaw in visual perception, humans tend to incorrectly estimate the relative gap sizes between two lines with widely varying slopes [18]. Even in the St Lucia case, where the shaded region appears small due to this phenomenon, the gap size measured vertically is quite large in many regions.

6.3.1. Short-term Variation

Nearly all sites exhibit low to moderate variation in diffusion potential over the short-term (< 1 day), as the unreachable ratio tends to be uniformly high when content is just starting to diffuse. St Lucia is a notable exception, with $P_{95}-P_{05} \approx 0.5$ near the beginning of the trace. Given that St Lucia has already been identified as the site with the greatest content diffusion potential, it is not surprising to find some simulations in which a low unreachable ratio is realized almost instantaneously, increasing variation.

6.3.2. Medium-term Variation

We observe at all sites a moderate to large variation in unreachable ratio at some point over the medium-term (≈ 1 day–9 days). In some cases, $P_{95} - P_{05} \gtrsim 0.5$. Generally speaking, it is medium-term diffusion potential which exhibits the greatest variability.

6.3.3. Long-term Variation

We note that in about half of all cases the variability seems to decrease substantially nearing the end of the trace period, often such that $P_{95} - P_{05} < 0.2$. In other cases, the variability remains much higher even nearing the end of the trace, sometimes with $P_{95} - P_{05} \approx 0.5$. St Lucia is the only site which exhibits low long-term variation under both optimistic and pessimistic session length assumptions. Ipswich, Gatton and P.A. Hospital all exhibit low long-term variation under optimistic assumptions, but higher variation under pessimistic assumptions. Herston reverses this pattern, with relatively low long-term variation under a pessimistic assumption but high variation under an optimistic assumption.

6.3.4. Variation Summary

All sites are susceptible to widely varying diffusion potential across source devices at one point or another throughout our simulations. For those wishing to accurately model content diffusion or design applications where the diffusion potential of individual devices is important, we suggest that the variation in diffusion potential across individual devices is an important consideration.

7. Simulating Application-Specific Diffusion

In this section we examine a concrete use case of information diffusion—sharing electronic maps. Our simulations draw upon both the UQ trace and the JCUNav trace (described next) to model diffusion of maps between wireless devices. From the UQ trace we use the same set of sessions and inferred encounters used earlier in our universal diffusion simulations. We then project the daily and hourly usage patterns from the JCUNav trace (Figure 7 and Figure 8) onto the UQ trace to simulate demand for maps throughout each simulated day and quantify the number of users whose demand could have been served from the MANET.

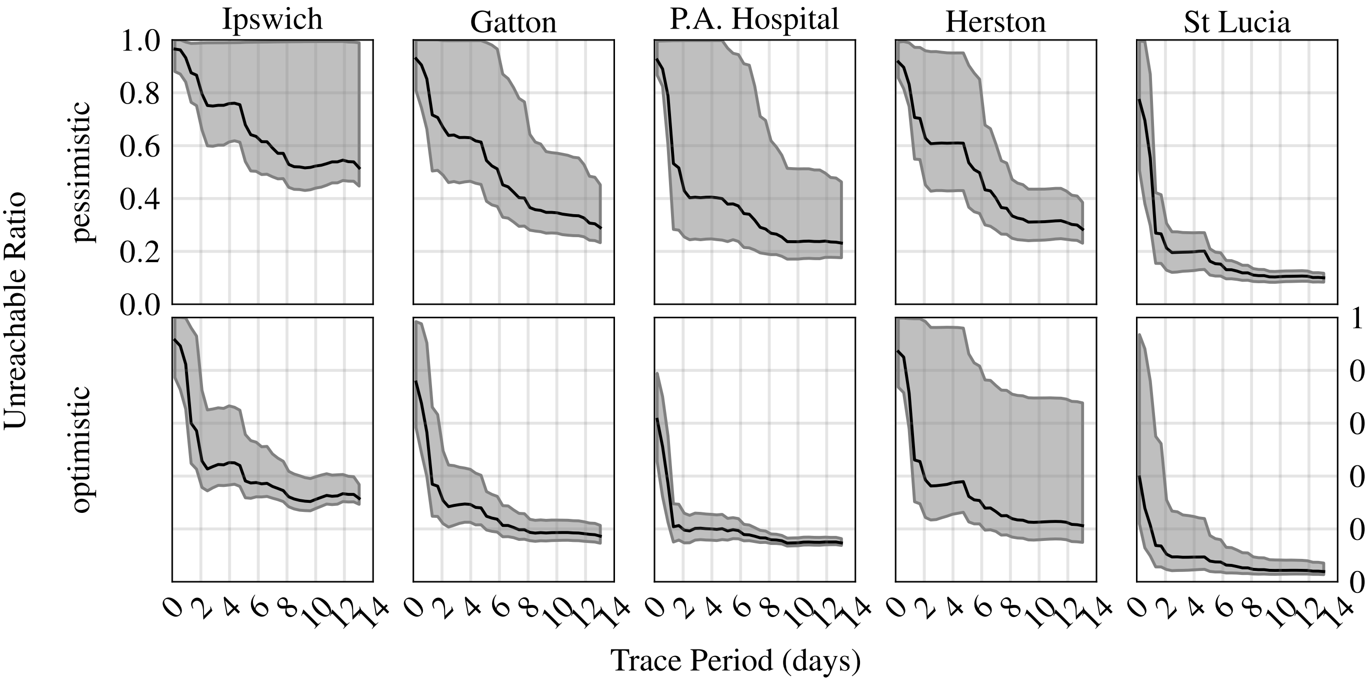

Figure 6: Variation in unreachable ratio across random trials. Black lines depict the average. Shaded regions depict 5th–95th percentile around the average.

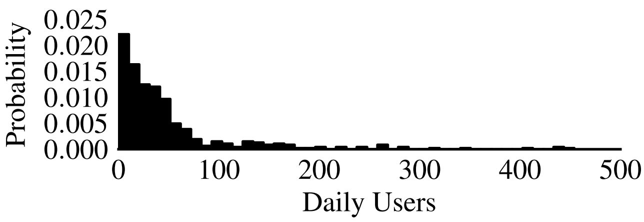

Figure 7: Probability density—number of map users on any given day.

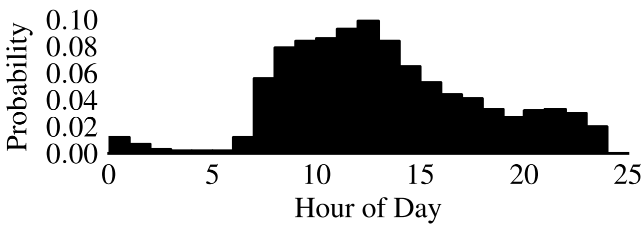

Figure 8: Probability density—number of map users in any given hour of the day.

7.1. The JCUNav Trace

JCUNav [19] is a popular Apple iOS mobile campus navigation application at *James Cook* University (*JC*U, not UQ), written and maintained by the primary paper author. For 450 days spanning 6th September, 2012–29th November, 2013, application usage statistics were gathered from JCUNav using the Flurry Analytics [20] logging framework. Two key pieces of information were extracted from the logged data: i) a count of daily JCUNav users each day over the trace period and ii) a frequency distribution aggregated over the entire trace period describing the daily distribution of application usage delineated into 24 1-hour buckets. Figure 7 illustrates the distribution of number of daily JCUNav users (Freedman-Diaconis binning [21]). Figure 8 illustrates the distribution of application usage throughout the day, retaining the original hourly binning of the JCUNav trace.

7.2. Simulation Overview

Using the first seven days of trace from the St Lucia campus (the largest site in the UQ trace), we simulate ideal maps diffusion over seven discrete time periods, one period for each day of the week. For each day we perform 50 simulation trials and average the results. Our map demand simulations are summarised in Algorithm 2 which is run over optimistic and pessimistic input traces separately. Each key step is described in more detail shortly.

The previously covered UQ trace describes device encounters though does not describe application usage patterns of the studied devices. Conversely, the JCUNav trace describes application usage patterns of a set of studied devices though does not describe device encounters. To simulate diffusion of maps, we therefore project the usage patterns of the JCUNav trace onto the encounter pattern of the UQ trace. We describe our procedure for achieving this in the following steps, which we perform for each individual simulation. We perform 50 simulation trials for each of the 7 days under both pessimistic and optimistic assumptions, for a total of $50 \times 7 \times 2 = 700$ simulations:

- Draw one random sample from the daily users probability distribution in Figure 7 (Algorithm 2, Line 12). This will be the number of users who would like a copy of the map in a given simulation.

- Multiply the random sample by the *UQ scale coefficient* (Line 14). The UQ St Lucia campus population is larger than the JCU Townsville population by around a factor of three and so we must multiply the daily user counts by the UQ scaling coefficient—3. Let the result of this multiplication be called n.

- Multiply n by the *day of week scale coefficient* (Line 15). The level of campus activity at UQ varies depending on the day of the week, particularly between weekdays and weekends. To account for this variability, we apply a scaling factor that is equal to the number of UQ network users on the given simulation day divided by the average number of UQ network users across all simulation days. Table 4 lists the scale factor for each day of the week under both pessimistic and optimistic session length assumptions. Let the result of this multiplication be m.

- Randomly sample m times from the time of day distribution illustrated in Figure 8 (Line 17). The m sampled times become the individual times of day each map requesting user would like to see the map, and we call this vector T. A limitation of the JCUNav trace is that there is no way to discern between users who are on and off campus. As a simplifying assumption, we assume a user to be on campus if the map is requested between 7am-7pm and off campus otherwise. Any time $t \in T$ that falls during an off campus period is discarded from T (Line 18), essentially reducing the number of requesting users for the simulation day to only those who requested the map while on campus.

- For the given simulation day, assign one device DEV_t from the UQ mobility trace to each time $t \in T$. DEV_t must be a device that is online in the UQ trace at time t, as we make the simplifying assumption that a user on campus is always connected to an access point and recall that all of our users in T are considered on campus.

At this stage, we have assigned a randomly chosen set of devices to serve as users interested in the map on a given day, and have defined the time of day each individual user requests the map. We then construct a DES similar to that described earlier in Section 5. This time however, rather than beginning the simulation with a fixed number of content sources, we add "demand" events corresponding to each time of day a device would like to view the map. A demand event can result in one of two outcomes: i) a *cache miss*: the device does not currently possess the map and so must retrieve the map from the infrastructure or ii) a *cache hit*: the device has received the map via diffusion at some time prior to when it would like to view the map, in which case there is no need to resort to the infrastructure. As in the universal content diffusion, the content (in this case the map) diffuses between devices when a device with the content encounters a device without the content. For the map simulation, the first demand event will always result in a cache miss, as nobody in the network possesses the map. This first device is then capable of spreading the content via diffusion. Each subsequent map demand may either result in a cache hit or cache miss, depending on whether the map reached the demanding device via diffusion before being requested.

There are a few additional assumptions worth noting. Firstly, we break the simulations down into individual days, rather than running a single simulation over the entire trace period. Secondly, we assume that the map content is flushed from all user's caches at the end of the day. This has to do with a limitation of the JCUNav trace, which is that there is no way to identify which users are repeat users across multiple days, meaning it is not possible to establish who already does and does not have the map over two or more consecutive days.

Algorithm 2 Maps content demand simulation.

```
 1: function RUN_APP_SPECIFIC()
 2:     days = {Wed Nov 28 7am–7pm, ...,
            Tue Dec 4 7am–7pm}
 3:     site = St Lucia
 4:     for ∀⟨d, s⟩ ∈ {days × site} do
 5:         SIMULATE(d, s)
 6:     end for
 7: end function
 8:
 9: function SIMULATE(day, site)
10:     for i = 1 to 50 do
11:         ▷ returns scalar
12:         numUsers = SAMPLEDAILYUSERS()
13:         uqCoefficient = 3
14:         numUsers *= uqCoefficient
15:         numUsers *= SCALEFACTOR(day)

16:         ▷ returns list of length |numUsers|
17:         dTimes = SAMPLEDEMANDTIMES(numUsers)
18:         dTimes = {d | d ∈ dTimes ∧ d ≥ 7am ∧ d ≤ 7pm}
19:         SIMULATEDIFFUSION(day, site, dTimes)
20:     end for
21: end function
```

Table 4: Day of week scale factors.

Day	Pessimistic	Optimistic
Monday	1.322	1.286
Tuesday	1.337	1.297
Wednesday	1.369	1.330
Thursday	1.359	1.337
Friday	1.175	1.178
Saturday	0.24	0.307
Sunday	0.198	0.265

The measure we are interested in for the map diffusion simulation is the *cache miss ratio*, defined simply as:

$$z = \frac{Cache\,Misses}{Cache\,Hits + Cache\,Misses} \tag{3}$$

The cache miss ratio z reflects the number of times a device which would like the map has to resort to the infrastructure, as opposed to receiving the content ahead of time via diffusion. A lower cache miss ratio therefore implies a more effective MANET.

7.3. Simulation Results

Figure 9 and Figure 10 illustrate the simulation results. We note firstly the pronounced difference in rate of diffusion between weekdays and weekends, with weekdays demonstrating greater diffusion potential. This result is consistent with our earlier findings in universal diffusion. Though particular days clearly demonstrate superior diffusion potential even when controlling for weekdays/weekends, the exact order is not consistent between pessimistic and optiistic simulations. For example, after 12 hours Sunday has more diffusion potential than Saturday in pessimistic simulations, while the pattern is reversed in optimistic simulations. Similar reversals are observable between weekdays also.

Aside from the re-ordering of some day's diffusion potential between optimistic and pessimistic simulations, we draw attention to substantial *absolute* differences in diffusion potential based on the chosen assumption. Under the pessimistic assumption weekends and weekdays exhibit cache miss ratios of around 95–87% and 78–72% respectively. In constrast, under optimistic assumptions these ratios fall to around 69–61% and 40–28%. For weekends this represents a difference of over 25% and for weekdays a difference of as much as 40%. As absolute differences these are non-trivial and again demonstrate the sensitivity of diffusion potential to trace uncertainties.

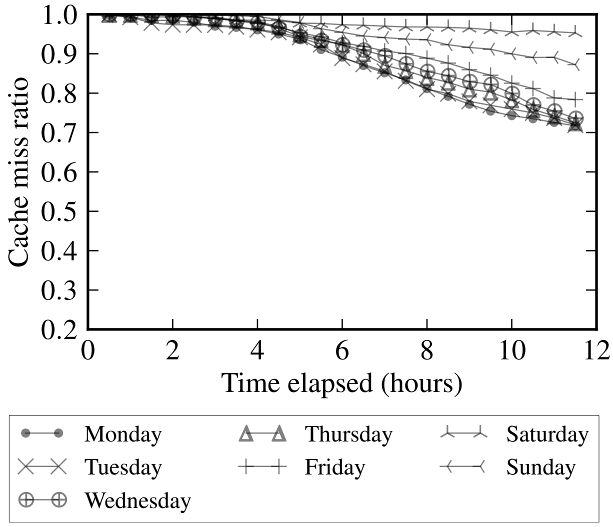

Figure 9: JCUNav diffusion partitioned by day (pessimistic).

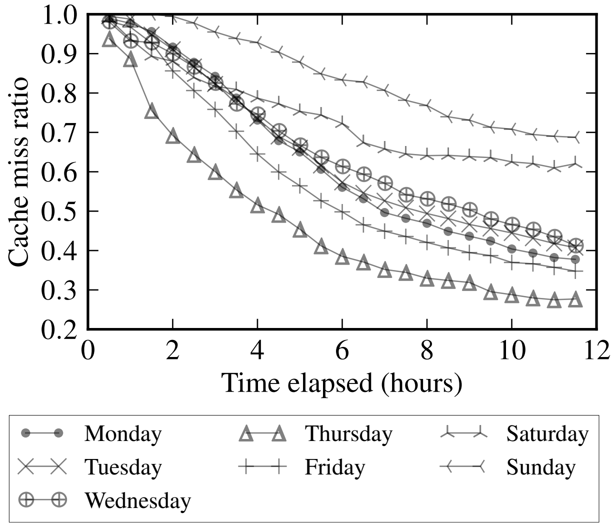

Figure 10: JCUNav diffusion partitioned by day (optimistic).

8. Discussion and Future Work

The results presented in this paper elucidate a number of tangible factors influencing rates of information diffusion. However, our comparison of diffusion potential under optimistic and pessimistic assumptions also highlights diffusion's sensitivity to trace uncertainties. Some traces like the UQ trace embed uncertainties regarding session start and end times which are the result of periodic rather than instananeous sampling of connected devices. Other forms of uncertainty however are more general and intrinsic to wireless traces collected from the view of the wireless infrastructure. Namely:

Disconnection time errors: ideally, associations in wireless networks are explicitly terminated by either the user or infrastructure device. In practice , a user device may simply travel out of range of the infrastructure or otherwise fail to explicitly request a disconnection. In such cases, wireless networks such as 802.11 (Wi-Fi) typically rely on inactivity timeouts to trigger session termination. A Wi-Fi timeout may be on the order of 30 minutes, as is the case in the UQ trace. This creates a session end time uncertainty leaving no way to determine the portion of the timeout period simply spent inactive versus actually absent from the network. Moreover, a device which both exits and re-enters the network inside the timeout window may never be flagged as having been disconnected for the period of absence.

Encounter inference errors: Our own study as well as prior work [2] have made the simplifying assumption that simultaneous connectivity of devices to an access point implies the devices are in transmission range of each other. As described earlier, this assumption inevitably induces both errors of ommission and commission—devices not simultaneously connected to an AP may actually be in transmission range and devices which are simultaneously connected to an AP may not be in transmission range. As with disconnection uncertainties, the magnitude of this error remains unquantified.

Given the differences we observed in diffusion potential between optimistic and pessimistic session length assumptions, we suggest an important area of future work will be addressing the aforementioned spatial and temporal trace uncertainties. We suggest that a valuable contribution in this area would be an encounter trace collected from the device's point of view, rather than the infrastructure. Though examples can be found in the literature of where this has been done, they tend to be susceptible to one or more of the following problems: i) the experiment is contrived [3, 9, 8] (e.g. devices handed out to graduate students) ii) the sample size is small (e.g. 10–300 devices) [3, 9, 22, 8] iii) the instrumented devices are geographically sparse [22] iv) the trace is dated [3, 9, 8]. One avenue for collecting this data within a university or organization may be to instrument one or more site-specific "apps" on smartphones and tablets to gather such data. For example, the majority of students at university X may have the official X app installed, making for a large sample that is geographically dense, less contrived and recent.

Another area for future research is broadening the scope of analyzed trace environments. Also of interest is understanding the way in which the next generation of networked devices and applications intend on harnessing MANET communication to enhance the utility of wireless devices beyond what is possible in infrastructure-only networks. While analysis of device encounters has been seen many times in the literature, there is a lacuna around how these encounters are (if at all) being used today for content dissemination and a need for a less scattered and more systematic review of their proposed uses in future.

9. Conclusion

Our analysis of MANET-based content diffusion reveals several important factors influencing diffusion potential. Firstly, the rate at which content spreads throughout a network is highly site-dependent, even across sites of the same type (university campuses) and even when the trace collection is controlled for both network type and collection period. Secondly, the time at which content is introduced into the MANET greatly influences the success of information diffusion over the short-term. In particular, content introduced into the network over the weekend suffers higher initial delay in reaching other devices than content which is introduced during the working week. This finding is consistent across both universal diffusion and application-specific diffusion simulations. Thirdly, the number of source devices used to diffuse a message can greatly influence the rate of diffusion, particularly over the short-term.

One of the key contributions of this paper is to highlight the impact of the aforementioned parameters on diffusion potential. Another equally important contribution however has been to illustrate that assumptions that are chosen when confronted with trace uncertanities can lead to large *absolute* differences in results. In our simulations of maps diffusion for example, we observed a 25–40% difference in diffusion potential between pessimistic and optimistic assumptions after 12 hours. In addition to trace uncertainties, we have also highlighted in this paper that there exists substantial variation in diffusion potential between devices—a fact easy to overlook when results are presented simply in terms of averages. We expect this aspect of our analysis to motivate the research community towards refining common assumptions and documenting intrinsic variations around averaged results.

Acknowledgements

This work is supported in part by an Australian Government Australian Postgraduate Awards scholarship and Commonwealth Scientific and Industrial Research Organisation Office of the Chief Executive scholarship.

[1] P. J. Denning, The locality principle, Commun. ACM 48 (7) (2005) 19–24.

[2] W. Hsu, A. Helmy, On nodal encounter patterns in wireless lan traces, IEEE Transactions on Mobile Computing 9 (2010) 1563–1577.

[3] J. Su, A. Chin, A. Popivanova, A. Goel, E. de Lara, User mobility for opportunistic ad-hoc networking, in: Proceedings of the Sixth IEEE Workshop on Mobile Computing Systems and Applications, WMCSA '04, IEEE Computer Society, Washington, DC, USA, 2004, pp. 41–50.

[4] B. Thomas, I. Atkinson, R. Jurdak, Content diffusion in wireless MANETs: The impact of mobility and demand, in: Wireless Communications and Mobile Computing Conference (IWCMC), 2014 International, 2014, pp. 959–966.

[5] D. Johnsort, Routing in ad hoc networks of mobile hosts, in: Mobile Computing Systems and Applications, 1994. WMCSA 1994. First Workshop on, 1994, pp. 158–163.

[6] K. Fall, A delay-tolerant network architecture for challenged internets, in: Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications, SIGCOMM '03, ACM, New York, NY, USA, 2003, pp. 27–34.

[7] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, R. L. Braynard, Networking named content, in: Proceedings of the 5th international conference on Emerging networking experiments and technologies, CoNEXT '09, ACM, New York, NY, USA, 2009, pp. 1–12.

[8] N. Eagle, A. (Sandy) Pentland, Reality mining: sensing complex social systems, Personal Ubiquitous Comput. 10 (4) (2006) 255–268.

[9] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, C. Diot, Pocket switched networks and human mobility in conference environments, in: Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking, WDTN '05, ACM, New York, NY, USA, 2005, pp. 244–251.

[10] A. Balachandran, G. M. Voelker, P. Bahl, P. V. Rangan, Characterizing user behavior and network performance in a public wireless lan, SIGMETRICS Perform. Eval. Rev. 30 (1) (2002) 195–205.

[11] D. Kotz, K. Essien, Analysis of a campus-wide wireless network, Wirel. Netw. 11 (1-2) (2005) 115–133.

[12] D. Tang, M. Baker, Analysis of a local-area wireless network, in: Proceedings of the 6th annual international conference on Mobile computing and networking, MobiCom '00, ACM, New York, NY, USA, 2000, pp. 1–10.

[13] D. Tang, M. Baker, Analysis of a metropolitan-area wireless network, Wirel. Netw. 8 (2/3) (2002) 107–120.
[14] N. Aschenbruck, A. Munjal, T. Camp, Trace-based mobility modeling for multi-hop wireless networks, Computer Communications 34 (6) (2011) 704–714.
[15] P. Holme, J. Saramäki, Temporal networks, Physics Reports 519 (3) (2012) 97–125.
[16] R. Shah, S. Roy, S. Jain, W. Brunette, Data mules: modeling a three-tier architecture for sparse sensor networks, in: Sensor Network Protocols and Applications, 2003. Proceedings of the First IEEE. 2003 IEEE International Workshop on, 2003, pp. 30–41.
[17] Cisco Systems Inc., Cisco prime network control system 1.1 data sheet, http://www.cisco.com/en/US/prod/collateral/wireless/ps5755/ps11682/ps11686/ps11688/data_sheet_c78-650051.html, accessed: Jan. 11, 2015.
[18] W. S. Cleveland, R. McGill, Graphical Perception and Graphical Methods for Analyzing Scientific Data, Science 229 (1985) 828–833.
[19] Bryce Thomas, JCUNav—James Cook University navigation application, https://itunes.apple.com/au/app/jcunav/id497618649?mt=8, accessed: Jan. 11, 2015 (2013).
[20] Flurry, Flurry analytics software library, http://www.flurry.com/flurry-analytics.html, accessed: Jan. 11, 2015 (2013).
[21] D. Freedman, P. Diaconis, On the histogram as a density estimator:l 2 theory, Zeitschrift fr Wahrscheinlichkeitstheorie und Verwandte Gebiete 57 (4) (1981) 453–476.
[22] N. Corporation, The mobile data challenge: Big data for mobile computing research, https://research.nokia.com/files/public/MDC2012_Overview_LaurilaGaticaPerezEtAl.pdf, accessed: Jan. 11, 2015 (2012).

Bryce Thomas received the bachelor's degree in Information Technology from Central Queensland University, Australia and honours in Information Technology from James Cook University, Australia. Bryce is currently an Information Technology PhD candidate at James Cook University.

Raja Jurdak is a Principal Research Scientist at CSIRO. He has a PhD in Information and Computer Science at University of California, Irvine in 2005, an MS in Computer Networks and Distributed Computing from the Electrical and Computer Engineering Department at UCI (2001), and a BE in Computer and Communications Engineering from the American University of Beirut (2000). His current research interests focus on mobility in networks. He has over 75 peer-reviewed journal and conference publications, as well as a book published by Springer in 2007 titled Wireless Ad Hoc and Sensor Networks: A Cross-Layer Design Perspective.

Ian Atkinson is a Tropical Leader and Director of the eResearch Centre at James Cook University. As Director of the JCU eResearch Centre he has the privilege of working with researchers to apply new and constantly changing ICT tools and methods to multiple research domains. He has a long-standing interest in eResearch methods, tools, scientific data management and user interfaces for HPC tools. He is also actively involved in researching how new systems and software that connect the physical and virtual worlds, particularly focusing on environmental monitoring with sensor networks.

7

Paper 4 – Diffusion in Colocation Contact Networks: the Impact of Nodal Spatiotemporal Dynamics

The fourth and final peer-reviewed article of the thesis has been submitted for consideration to PLOS ONE and is titled *Diffusion in Colocation Contact Networks: the Impact of Nodal Spatiotemporal Dynamics*. In the prior two papers, the focus was on establishing how various real-world parameters influence rate of spreading. In the final paper, we instead focus on developing a set of generalizable *null models* for shuffling colocation contact networks in a manner which allows one to separately isolate the impact of different spatiotemporal correlations.

The phrase *null models* is purported to have been coined in 1981 by ecologists Robert K. Colwell and David W. Winkler [45]. In theoretical ecology, null models are used to 'randomize away' some ecological process while holding constant other aspects of the data. This is in turn used to reason about the role of non-randomness in the process under consideration in influencing some ecological outcome.

Similar to ecological null models, in the study of mathematical graphs, null models are used to randomize one or more aspects of a graph in order to determine the influence of initially non-random graph features on some observed outcome. Prior graph null models have focused on rewiring *existing* graphs, such that one or more of the graph's properties is randomized while other graph properties are maintained. That is, the null model takes as its input a graph and generates as its output a graph with certain features decorrelated. In the following paper, we instead introduce the notion of *inducement-shuffled* null models, in which the events which led to the graph in the first place are randomized in one or more ways, rather than randomizing the graph itself. The input to an inducement-shuffled null model is not a graph, but rather a set of events on which a graph's structure is predicated. The output of an inducement-shuffled null model is a new set of events, where one or more aspects of those events has been randomized (decorrelated). An inducment-shuffled null model exerts its influence over the structure of a graph not by modifying the graph itself, but rather indirectly by altering the events underlying the calculation of the graph in the first place.

As is elucidated in the paper, inducement-shuffled null models allow us to establish how the spatiotemporal *behavior* of nodes impacts the rate of diffusion, through controlled randomizations of (T)imes, (L)ocations and (N)odes (TLN). In line with the earlier subject matter of the thesis, we apply the inducement-shuffle null models to the same wireless mobile device session trace and measure rates of diffusion. As noted in the paper, we anticipate the inducement-shuffled null models we present may find broader applications throughout the field of complex networks.

The included copy of the manuscript which follows is the authors' preprint.

Diffusion in Colocation Contact Networks: the Impact of Nodal Spatiotemporal Dynamics

Bryce Thomas[1,*], Raja Jurdak[2], Kun Zhao[2], Ian Atkinson[3]
1 Department of Information Technology, James Cook University, Townsville, Queensland, Australia
2 Autonomous Systems, CSIRO, Brisbane, Queensland, Australia
3 eResearch, James Cook University, Townsville, Queensland, Australia
∗ **E-mail: bryce.m.thomas@gmail.com**

Abstract

Temporal contact networks are studied to understand dynamic spreading phenomena such as communicable diseases or information dissemination. To establish how spatiotemporal dynamics of nodes impact spreading potential in colocation contact networks, we propose "inducement-shuffling" null models which break one or more correlations between times, locations and nodes. By reconfiguring the time and/or location of each node's presence in the network, these models induce alternative sets of colocation events giving rise to contact networks with varying spreading potential. This enables second-order causal reasoning about how correlations in nodes' spatiotemporal preferences not only lead to a given contact network but ultimately influence the network's spreading potential. Under each of the presented null models we measure both the number of contacts and infection prevalence as a function of time, with the surprising finding that the two are not strongly related. We find the correlation between nodes and times to be the greatest impediment to spreading while the correlation between times and locations slightly catalyzes spreading, though only in the absence of other correlations.

Introduction

Complex networks [1–3] as a unified framework can describe a wide range of systems including the Internet and World Wide Web [4], biological systems [5–7] and social networks [8–10]. Dynamical processes taking place on networks, such as information diffusion or epidemic spreading, are strongly influenced by the structure and organization of networks [3]. For a long time, studies on the interplay between network structure and dynamical processes have been focused on static networks [11, 12]. Recently, attention has turned to the study of more complicated systems, in which dynamical processes are taking place on temporal networks with time-varying structures [13, 14]. Underlying factors in these systems that drive the evolution of temporal networks can subsequently impact the embedding dynamic phenomenon. Identifying these factors and their intrinsic association with the process is crucial for unravelling the complexity of these systems, as well as modelling, predicting and controlling the dynamic phenomenon.

Contact networks are one important instantiation of temporal networks that model either the logical or physical contacts between individuals. Examples of the former include long-range communications such as phone calls or email. Examples of the latter include colocation-driven short-range communications such as face-to-face human contact [15, 16] or proximity-based direct wireless transmissions [17]. Unlike logical contact networks, physical (i.e. colocation) contact networks are predicated on spatially constrained copresence and depend on mobility for broadscale spreading. The spatiotemporal dynamics of the actors (nodes) in such networks is therefore a critical determinant of spreading potential.

One of the best available sources of empirical data on colocation contact (also known as "encounter") networks comes from electronic mobile wireless device traces [17]. These traces typically describe either directly recorded encounters between devices [18–21] or encounters which are inferred based on mutual presence at a known location [21, 22]. Much work has been devoted to the analysis of these traces, often in terms of their static network-theoretic properties and in some instances also in terms of spreading potential. For example, the authors in [20–22] all explicitly analyzed ad hoc [23] multi-hop message dissemination [24] facilitated by device mobility and encounters, the latter work being our own performed over the same trace analyzed in this paper. While many attempts have been made to model such networks, a lacuna exists insofar as there is no structured framework for establishing how node's spatiotemporal preferences impede or catalyze spreading potential. It is to this matter that we direct our attention in this paper.

One of the core tools for relating observed spreading behavior of a network quantity (e.g. virus, information) to idiosyncratic network features are *null models* [25, 26] which separately destroy correlations in a network's contact structure. By simulating spreading on both the real and null networks one can establish to what extent certain correlations catalyze or impede spreading. Prior null models [25] have focused on *contact* shuffling—given a set of nodes connected by timestamped links, these models shuffle the links in one of a number of ways to retain certain correlations while destroying others. For example, the author's in [25] applied null model contact shufflings to a mobile call network (MCN) in which nodes were mobile subscribers and links were placed between caller/callee pairs and annotated with one or more call timestamps. A limitation of contact shuffled null models as applied to colocation-based contact networks is that they cannot draw any association between spatiotemporal dynamics and spreading. Rather, the models consider the contact network independent of the nature of the events which led to the contacts in the first place.

In this paper we propose an alternative way of generating network null models through a method we refer to as *inducement* shuffling—shuffle the events on which contacts are themselves predicated and so in the process induce a modified set of contact events. Inducement shuffling possesses the attractive property of enabling second-order causal reasoning about spreading behavior. With contact shuffling it is only possible to state how characteristics of the contact network itself influence spreading. With inducement shuffling one can state how the nature of the events that lead to the contact network in the first place influence spreading. We present our inducement-shuffled null models based on the theme of node, location and time which we use to decorrelate spatiotemporal node preferences. We apply the null models to an empirical trace in which (i) nodes are mobile wireless electronic devices such as smartphones, tablets and laptop computers (ii) locations are Access Points (APs) in a large university campus network and (iii) times are the times at which given devices were connected to given APs. Contacts in our empirical trace are predicated on device colocation. That is, two devices simultaneously present at a given location are connected by a timestamped link in the inferred contact network. Through inducement shuffling we destroy node's time and location preferences, i.e. we decorrelate the relationships between nodes, times and locations. This in turn leads to a different set of colocation events and thus a modified contact network is induced. For the original and each induced contact network we simulate diffusion of a quantity starting from a randomly infected node under the Susceptible-Infected (SI) infection model [27]. To the best of the authors' knowledge, the present work is the first to take a null models approach to isolating spreading impediments and catalysts in colocation-driven contact networks. More importantly, we believe the inducement-shuffled null models to be the first to enable reasoning about the second-order causal relationship between the events on which the contact network is predicated and subsequent spreading potential. Though motivated in the context of a wireless mobile device contact trace, we believe the inducement-shuffled null models presented in this paper may find broader applications, some well outside computer networks.

Materials and Methods

Dataset

Our work utilizes a wireless IEEE 802.11 (Wi-Fi) network trace collected by the University of Queensland (UQ) Australia describing device sessions at individual access points (APs) in a large university network over a 14 day period between Tue Nov 27 17:39:12 AEST 2012–Tue Dec 11 17:29:16 AEST 2012 . Each session record in the UQ trace includes (i) the unique MAC address of the connecting mobile device (ii) session start time (iii) session end time (iv) AP name and (v) site. We clean the trace by discarding 1,462 sessions with no end time (active when trace collection ceased) and 1 session with zero duration (start = end). A further 1,279 sessions are discarded as specifying no AP. After cleaning, the UQ trace retains 546,260 sessions from 23,895 devices over 3,079 APs at a total of 24 discrete geographic sites. In this paper we focus only on the largest site in the UQ trace—the *St Lucia* campus. The St Lucia campus is a large university campus accounting for 445,867 sessions from 20,308 devices over 2,004 APs. After filtering on (v) to extract the St Lucia trace from the UQ trace, we then use (i)–(iv) to construct the minimum dataset for our analysis which consists of a set of session 4-tuples, each of the form $\langle N, T_{start}, T_{end}, L\rangle$. These 4-tuples fully describe which (N)ode (MAC address) partakes in a session at what (T)ime (start and end) and at what (L)ocation (AP), i.e. they encode the information about the spatiotemporal patterns of nodes. We refer to these 4-tuples simply as the "session tuples" from this point onwards. Note that nodes and locations just happen to be defined in terms of MAC addresses and APs respectively in our analysis. The notion of nodes and locations may be generalized to accommodate equivalent entities specific to the network under consideration.

Contact Inference

The session tuples extracted from our original dataset and later reconfigured by our inducement shufflings do not explicitly describe contacts between pairs of wireless mobile devices. Rather, as is done in [21] we infer contact from colocation—two devices connected to the same AP at the same time are assumed to be in transmission range and so inferred as making "contact". As noted by the authors in [21], this inference is an approximation. Devices connected to the same AP may not be able to communicate directly, devices connected to different APs may be able to communicate and some contacts take place outside the range of APs. Still, it is believed to be a reasonable approximation in the network under consideration. Contact inference translates the session tuples into a set of 5-tuples of the form $\langle N_i, N_j, T_{start}, T_{end}, L\rangle$, describing the location and duration of a contact between devices N_i and N_j $(i \neq j)$. For our spreading analysis we rely only on the first three fields of the 5-tuples i.e. $\langle N_i, N_j, T_{start}\rangle$, which describes the initiation of contact events. We refer to these 3-tuples simply as the "contact tuples" from this point onwards.

Contact Shuffling

As a preamble to our main inducement shuffling results, we apply the pre-existing contact-shuffled null models presented by the authors in "Small but slow world" [25] to our originally inferred contact network. This allows us to compare and contrast spreading behavior against that observed in prior work. We refer to these earlier null models simply as the "SBSW" or "contact-shuffled" null models. The SBSW models were previously applied to a number of contact networks, most notably a large Mobile Call Network (MCN) with timestamped links between caller/callee pairs. The authors use strings of capital letter abbreviations in naming the contact-shuffled null models, where each letter represents a retained correlation. These correlations are (D)aily pattern i.e. overall event frequency, (C)ommunity structure, (W)eight topology correlation (B)ursty event dynamics on single links and (E)vent-event correlations between links. Below we reproduce verbatim the description of the SBSW null models, more details about which can be found in the original paper:

- DCWB (*equal-weight link-sequence shuffled*): Whole single-link event sequences are randomly exchanged between links having the same number of events. Temporal correlations between links are destroyed.
- DCB (*link-sequence shuffled*): Whole single-link event sequences are randomly exchanged between randomly chosen links. Event-event and weight-topology correlations are destroyed.

– DCW (*time-shuffled*): Time stamps of the whole original event sequence are randomly reshuffled. Temporal correlations are destroyed.

– D (*configuration model*): The original aggregated network is rewired according to the configuration model, where the degree distribution of the nodes and connectedness are maintained but the topology is uncorrelated. Then, original single-link event sequences are randomly placed on the links, and time shuffling as above is performed. All correlations except seasonalities like the daily cycle are destroyed.

Inducement Shuffling

Our main results are based around the inducement-shuffled null models presented below which are framed in terms of (T)imes, (L)ocations and (N)odes. The input to our inducement models are the session tuples. The output is a set of new session tuples of equal length with one or more of the correlations between pairs of T, L and N destroyed. We perform contact inference on the output sessions using the method just described to arrive at a set of new contact tuples. We reiterate that because inducement shuffling results in a different set of colocation events, it *implicitly* induces a modified contact network during contact inference, rather than explicitly reconfiguring the contact network itself.

The listing below describes each of the new inducement null models. Each model's abbreviation is based on paired capital letters which indicate the retained correlation(s) between times, locations and nodes. For example, the null model LN-TN retains (L)ocation/(N)ode and (T)ime/(N)ode correlations. LN retains only (L)ocation/(N)ode correlation. Under this notation, the original session trace could also be referred to as LN-TN-TL as correlations are retained between all pairs of location, node and time. Note that our inducement models only destroy *correlations* between the three variables—they do not alter the independent frequency distribution of times, locations and nodes. Moreover, session *durations* are always retained. That is, even if session times are shuffled, pairs of start and end times always move together. The inducement models are summarized in Table 1 in addition to the detailed listing below. Note that the final column of Table 1 includes spreading prevalence at 1 day, the meaning of which will become clear when we explain spreading dynamics shortly.

– LN-TN (*group by node, shuffle times*): sessions are first grouped by node. For each grouping, the list of time pairs are shuffled (i.e. each start/end pair is exchanged with another pair in the same group). Grouping by node acts to "loosely bind" nodes and times—although the times are shuffled, they are always reallocated to the same node and therefore only the binding between times and locations is destroyed. i.e. LN and TN correlations are retained while TL correlation is destroyed.

– TL-LN (*group by location, shuffle nodes*): sessions are first grouped by location. For each grouping, the list of nodes are shuffled. Again, grouping acts as a loose binding mechanism keeping the same location/node association. i.e. TL and LN correlations are retained while TN correlation is destroyed.

– LN (*shuffle times*): time pairs of the entire trace are shuffled. LN correlation is retained while TN and TL correlations are destroyed.

– TN (*shuffle locations*): locations of the entire trace are shuffled. TN correlation is retained while LN and TL correlations are destroyed.

– TL (*shuffle nodes*): nodes of the entire trace are shuffled. TL correlation is retained while LN and TN correlations are destroyed.

– $\emptyset$ (*shuffle locations, shuffle nodes*): locations of the entire trace are first shuffled and then nodes of the entire trace are shuffled. All correlations between time, location and node are destroyed.

In total there are $2^3 = 8$ potential models (including Original) based on which if any of the correlation pairs chosen from (LN, TN, TL) are retain during shuffling, however we omit one null model TL-TN from our analysis. TL-TN would in theory group by time and shuffle locations in order to destroy only the location/node (LN) correlation. Unlike location and node however, time is not a categorical variable and so grouping would require a discretization of times into an arbitrary number of slots. It is not readily apparent why one number of slots should be chosen over any other and so we leave contemplating TL-TN for future work.

We note that in our trace, locations are inherently discrete—individual APs with known identifiers. The inducement-shuffled null models do however generalize to broader settings, provided locations can be discretized. For example, a trace of mobile wireless bluetooth contacts annotated with known GPS coordinates might be discretized into a grid of lat/lon squares or geographically clustered. On the other hand, a contact network with no known locations (e.g. email) is not amenable to shuffling with the models presented in this paper. Moreover, though we focus our own simulations on a contact trace predicated on colocated wireless devices, the theme of times/locations/nodes is generalizable in the context of contact inference.

Spreading Dynamics

We model spreading atop of all contact networks using the Susceptible-Infected (SI) infection model [27]. Nodes in the SI model are in one of two states: (S)usceptible or (I)nfected. State change is unidirectional with nodes graduating from S to I upon satisfying the infection condition—in our trace, colocation of a susceptible and infected device. Under the SI model, infection prevalence is monotonically nondecreasing as a function of time growing until all devices reachable from initial conditions are infected.

Similar to the work in [25], we start by first infecting a randomly chosen node at a randomly chosen contact event. The chosen event's timestamp is interpreted as the simulation trial start time $t = 0$. We restrict the random sampling of the initial event to the first 4 days of trace to ensure a minimum of 10 days simulation "runway" before reaching the end of the trace. This 4/10 partition provides ample sampling opportunity across peak and trough traffic periods as well as weekdays and weekends while allowing enough simulation runway to observe the prevalence growth pattern. After sampling the initial infection event, we discard those contacts that occur either prior to the event or > 10 days after the event. This ensures all simulations run exactly 10 days, preventing a non-uniform number of samples beyond the 10-day period (e.g. only initial infection events sampled from within the first 12 hours of trace would have known prevalence at 13.5 days). We proceed to calculate the prevalence denominator N as equal to the size of the Largest Connected Component (LCC) of the aggregated 10-day contact network which typically consists of the majority of all devices ($> 95\%$) in the period. Note that on rare occasions the initially sampled device will fall outside of the LCC. In this case we simply resample until the device is a part of the LCC. Starting at the initial infection and stepping chronologically through the contact sequence we then simulate the SI model of ideal diffusion whereby a susceptible device becomes infected upon contact with an already infected device. We denote the set of infected devices at time t as $I(t)$ and the infection prevalence at time t as $P(t) = |I(t)|/N$. We perform a total of 250 random spreading trials (different starting nodes) and average the results. In the supporting material we also provide a comprehensive pairwise comparison of spreading potential for all inducement-shuffled null models along with standard errors around the averaged results.

Results and Discussion

Spreading Under Contact Shuffling

In Figure 1 we present the spreading results under the pre-existing SBSW contact-shuffled null models on the St Lucia contact sequence.

We first note that the time to near full prevalence in our trace is on the order of a few days, in comparison to 100's of days in [25]. Moreover, a clear "wavestep" pattern emerges—relatively flat periods of spreading interleaved with growth periods. In preparing our results we plotted individual trials separately and found the diurnal pattern of session volume lead to an even more pronounced wavestep pattern, with most spreading happening during high traffic periods i.e. during the day and particularly during the working week. What is seen in Figure 1 is merely an attenuated version of this phenomenon—simulation trials implicitly begin more often during busy periods as such events constitute the preponderance of the sampling pool, which tends to temporally align the averaged results "in phase".

Interestingly, the prior work in [25] found that DCWB and Original diffused at nearly identical rates whereas our results show DCWB being substantially faster than Original. This suggests that temporal link correlations alone are a substantial impediment to spreading in our trace unlike in prior work. We find that further destroying link weights (DCB) or bursty single-link event sequences (DCW) leads to little difference in spreading versus DCWB. Again this finding contrasts the prior work which suggested a clear spreading relationship between the three: DCW $>$ DCB $>$ DCWB. We conjecture that link weights and bursty single-link event sequences make little difference in our trace due to the relative homogeneity of link weights—most device contacts occur only once (Table 2). This means destroying weight or bursty event correlations does little to perturb our contact network. By comparison, the MCN in [25] we calculate to have average link weight $\overline{\chi} \approx 34$ (306×10^6 calls on only 9×10^6 links), whereby destruction of link weights or bursty event sequences would lead to substantial perturbation. Finally, consistent with [25] we find D to spread fastest. We note however that the marginal difference is less than observed in the prior work.

Spreading Under Inducement Shuffling

We now present our main result in Figure 2—the subsequent rate of spreading as a function of time after inducement shuffling of the session records. Table 1 also summarizes the prevalence at 1 day.

The original trace is found to spread slower than all of the inducement-shuffled null models, indicating the combined effects of location/node, time/node and time/location correlations result in the slowest spreading. Destroying the time/location correlation with LN-TN only marginally accelerates spreading and further destroying the location/node correlation in TN produces little further acceleration. In contrast, destroying the time/node correlation with TL-LN substantially accelerates spreading. LN alone also spreads relatively quickly suggesting time/node correlation as the common theme of slow spreading. Though it is interesting that some crossover is observed around 2–5 days between TL-LN and LN, there is little statistical strength to the statement that LN is ever truly faster which is made clear by comparing TL-LN and LN in Figure 8 (Supporting Information). TL and $\emptyset$ occupy the top echelon with respect to speed of spreading. This indicates two things: (i) location/node correlations have some impeding effect on spreading (TL being faster than TL-LN) and (ii) as we will show next, relatively fast spreading prevails even in the absence of the contact frequency advantage of time/location correlation (TL produces more contacts than $\emptyset$ but both spread relatively quickly). In summary, time/node correlation is found to be the main impediment to spreading with location/node correlation playing a smaller role.

Contact Frequency and Spreading

To explore the spreading propensity observed under each of the inducement-shuffled null models, we check whether some null models might simply be producing more contact events than others. In this section we plot both the number of contacts as a function of time and the prevalence as a function of contacts in order to establish the strength of the relationship between contact volume and spreading potential. We separate our plots into total and unique contacts. That is, the tally of all contact events including repeats and the tally of unique contact links between node pairs respectively.

Cumulative Contact Frequency

Figures 3 and 4 illustrate the number of contacts as a function of time for each of the inducement models, for total contacts and unique contacts respectively.

We begin by elucidating the total contact counts in Figure 3. All contact tallys exhibit the characteristic diurnal stepping pattern of five weekdays interleaved with two weekend days. Original and TL grow at essentially the same rate (overlapping lines) as they both retain the correlation between times and locations leading to the same number of total colocation events. We say "essentially" rather than "exactly", as all null models which decorrelate time and node (TL-LN, LN, TL, $\emptyset$) can occasionally produce "imaginary" contact events whereby a node is colocated with itself. Such events are discarded during contact inference and so are not accounted for in our tallys. For TL we find these imaginary contacts account for a negligible portion of all contact events ($< 0.03\%$) and so TL appears to overlap Original. TL-LN on the other hand is clearly illustrates the imaginary event phenomenon. In fact, the delta between Original and TL-LN in Figure 3 is a direct measure of the the number of discarded imaginary contacts under TL-LN shuffling which we find typically runs around 3%. It is interesting to note that LN is the only other shuffling with non-negligible imaginary events, again around 3%. The implication is that individual nodes likely have a strong affinity for specific locations (retained by *LN) leading to a (relatively) high number of imaginary events when the timing of node's presence in the network is shuffled. LN-TN produces the fewest total contacts indicating that correlated times and locations are the largest driver of contact volume in the Original trace. Further destroying either the location/node or time/node correlation as is done by TN and LN respectively increases total contacts. We expect that TN, LN and $\emptyset$ all produce the same number of total contacts as they all have the same expected number of nodes at a given location at a given time. The observed disparity of LN having fewer total contacts than TN and $\emptyset$ is simply attributable to the $\approx 3\%$ of imaginary contacts produced and discarded under LN shuffling.

We now turn our focus to the unique contact counts in Figure 4. Again, all contact tallys exhibit the characteristic diurnal stepping pattern being strongly driven by macro-scale periodic activity in the network. Whereas the original trace produces the most total contacts (in a tie with TL and TL-LN before discarding imaginary contacts), it also produces the *least* unique contacts. This suggests the "mixing" effect of the null models which increases unique contacts tends to be stronger than the opposite hindering effect the null models sometimes have on total contact volume. LN-TN and TL-LN produce approximately the same number of unique contacts which implies that destroying either time/location or time/node correlation has about equal effect on unique contacts. Destroying time/location and time/node correlations leaving only location/node correlation (LN) further increases unique contact. TN and $\emptyset$ produce still more unique contacts, possibly because the aforementioned affinity of devices for specific locations which is retained in LN had a stronger impeding effect on new unique contacts than does node's time preferences. TL leads to the most unique contacts, likely because it retains the total contact boost driven by node's preferences to be in the same location at the same time with the added benefit of mixing which nodes partake in the contact event.

As a final point of comparison between Figures 3 and 4 we note how unique contacts are always within a factor of 2.5 of total contacts. Again referring to Table 2, this is likely a by-product of a contact network whose original link "weights" as measured by repeat contact count are quite low versus the MCN analyzed in [25] (average link weight $\overline{\chi} \approx 34$).

Already Figures 3 and 4 appear to suggest that there is a weak relationship between a null model's propensity to produce contact events and its subsequent spreading potential. We now proceed to plot contacts versus prevalence where this will become even more obvious.

Contacts vs. Prevalence

Figures 5 and 6 plot prevalence as a function of the number of total and unique contacts respectively. It is now immediately evident that some inducement-shuffled null models are far more "efficient" at reaching a given prevalence in terms of requiring far fewer contact events. What is less obvious is why?

It is interesting to note that this disparity in spreading efficiency has also been implicitly observed both in this paper and prior work [25] under contact shuffling. This follows from the fact that contact shuffling inherently retains the same total contact frequency as a function of time, even after shuffling. Although unique (non-repeat) contacts may deviate from the original trace throughout the simulation period under D and DCW shufflings, these also equalize to that of the original trace by the end of the simulation. In any case, DCWB and DCB which match Original in both total and unique contacts at each time step present with drastically superior diffusion potential (on a per-contact basis) than the original trace in our analysis. We suggest that future work may wish to further explore the relationship between both the static and temporal structures of contact-shuffled and inducement-shuffled networks. Doing so may reveal common emergent properties such as statistical similiarities that offer a unified explanation to the disparity of spreading efficiency.

Conclusion

This paper has analyzed impediments and catalysts to spreading in a contact network through the introduction of a set of inducement-shuffled null models which separately destroy the correlations between times, locations and nodes. The inducement-shuffled null models have enabled second-order causal reasoning about the observed spreading propensity. That is, how is spreading affected by the idiosyncratic behavior that lead to the observed contact network in the first place, rather than how the observed contact network alone affects spreading. Among our main observations is that (i) spreading is primarily impeded by time/node correlation and (ii) though correlations have a slowing effect in general, retaining time/location correlation alone proves an exception which slightly increases the rate of spreading versus a network with all correlations destroyed. Furthermore we have demonstrated a curious disparity between a null model's ability to produce frequent contact events and its propensity to promote spreading. Finally, we have found that under pre-existing contact-shuffled null models temporal link correlations are the main spreading impediment in our trace, in contrast to earlier reported results.

We conclude by proposing a number of avenues for future work. Firstly, there are several areas already alluded to in this manuscript. These are (i) defining and justifying appropriate grouped shuffling models on non-categorical variables such as time (ii) establishing if and how "imaginary" contact events between a node and itself ought to ever be avoided in shuffling and (iii) clearly articulating the conditions which may lead to large disparities between contact volume and prevalence. More broadly, it would be useful to explore whether the abstraction of times, locations and nodes generalizes well to other types of contact networks, particularly those which do not operate at human scale (e.g. cellular level networks). Complex networks in itself offers a unified framework for contemplating a diverse range of systems and so we expect that some subset of these may well be framed in a similar null model construction to that used in this paper. Moreover, there may exist alternative null model abstractions that will afford the same second-order causal reasoning about the nature of spreading in other types of contact networks. We urge the reader to consider what these abstractions might be.

References

1. Albert R, Barabási AL (2002) Statistical mechanics of complex networks. Rev Mod Phys 74: 47–97.
2. Newman ME (2003) The structure and function of complex networks. SIAM review 45: 167–256.
3. Barrat A, Barthelemy M, Vespignani A (2008) Dynamical processes on complex networks, volume 1. Cambridge University Press Cambridge.
4. Barabási AL, Albert R, Jeong H (2000) Scale-free characteristics of random networks: The topology of the world-wide web. Physica A: Statistical Mechanics and its Applications 281: 69–77.
5. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási AL (2000) The large-scale organization of metabolic networks. Nature 407: 651–654.
6. Mason O, Verwoerd M (2007) Graph theory and networks in biology. Systems Biology, IET 1: 89–119.
7. Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience 10: 186–198.
8. Redner S (1998) How popular is your paper? an empirical study of the citation distribution. The European Physical Journal B-Condensed Matter and Complex Systems 4: 131–134.
9. Newman ME (2004) Coauthorship networks and patterns of scientific collaboration. Proceedings of the National Academy of Sciences 101: 5200–5205.
10. Eagle N, Pentland AS, Lazer D (2009) Inferring friendship network structure by using mobile phone data. Proceedings of the National Academy of Sciences 106: 15274–15278.
11. Pastor-Satorras R, Vespignani A (2001) Epidemic spreading in scale-free networks. Physical review letters 86: 3200.
12. Newman ME (2002) Spread of epidemic disease on networks. Physical review E 66: 016128.
13. Holme P, Saramäki J (2012) Temporal networks. Physics Reports 519: 97–125.
14. Liu SY, Baronchelli A, Perra N (2013) Contagion dynamics in time-varying metapopulation networks. Physical Review E 87: 032805.
15. Zhao K, Stehlé J, Bianconi G, Barrat A (2011) Social network dynamics of face-to-face interactions. Physical Review E 83: 056109.
16. Starnini M, Baronchelli A, Pastor-Satorras R (2013) Modeling human dynamics of face-to-face interaction networks. Physical review letters 110: 168701.
17. Aschenbruck N, Munjal A, Camp T (2011) Trace-based mobility modeling for multi-hop wireless networks. Computer Communications 34: 704–714.
18. Eagle N, (Sandy) Pentland A (2006) Reality mining: Sensing complex social systems. Personal Ubiquitous Comput 10: 255–268.
19. Hui P, Chaintreau A, Scott J, Gass R, Crowcroft J, et al. (2005) Pocket switched networks and human mobility in conference environments. In: Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking. ACM, WDTN ’05, pp. 244–251.

20. Su J, Chin A, Popivanova A, Goel A, de Lara E (2004) User mobility for opportunistic ad-hoc networking. In: Proceedings of the Sixth IEEE Workshop on Mobile Computing Systems and Applications. IEEE Computer Society, WMCSA '04, pp. 41–50.

21. Hsu Wj, Helmy A (2010) On nodal encounter patterns in wireless LAN traces. IEEE Transactions on Mobile Computing 9: 1563–1577.

22. Thomas B, Atkinson I, Jurdak R (2014) Content diffusion in wireless MANETs: The impact of mobility and demand. In: Wireless Communications and Mobile Computing Conference (IWCMC), 2014 International. pp. 959–966.

23. Johnsort D (1994) Routing in ad hoc networks of mobile hosts. In: Mobile Computing Systems and Applications, 1994. WMCSA 1994. First Workshop on. pp. 158–163.

24. Fall K (2003) A delay-tolerant network architecture for challenged internets. In: Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications. ACM, SIGCOMM '03, pp. 27–34.

25. Karsai M, Kivelä M, Pan RK, Kaski K, Kertész J, et al. (2011) Small but slow world: How network topology and burstiness slow down spreading. Phys Rev E 83: 025102.

26. Gauvin L, Panisson A, Cattuto C, Barrat A (2013) Activity clocks: spreading dynamics on temporal networks of human contact. Scientific reports 3.

27. Anderson RM, May RM (1992) Infectious Diseases of Humans Dynamics and Control. Oxford University Press.

Figure Legends

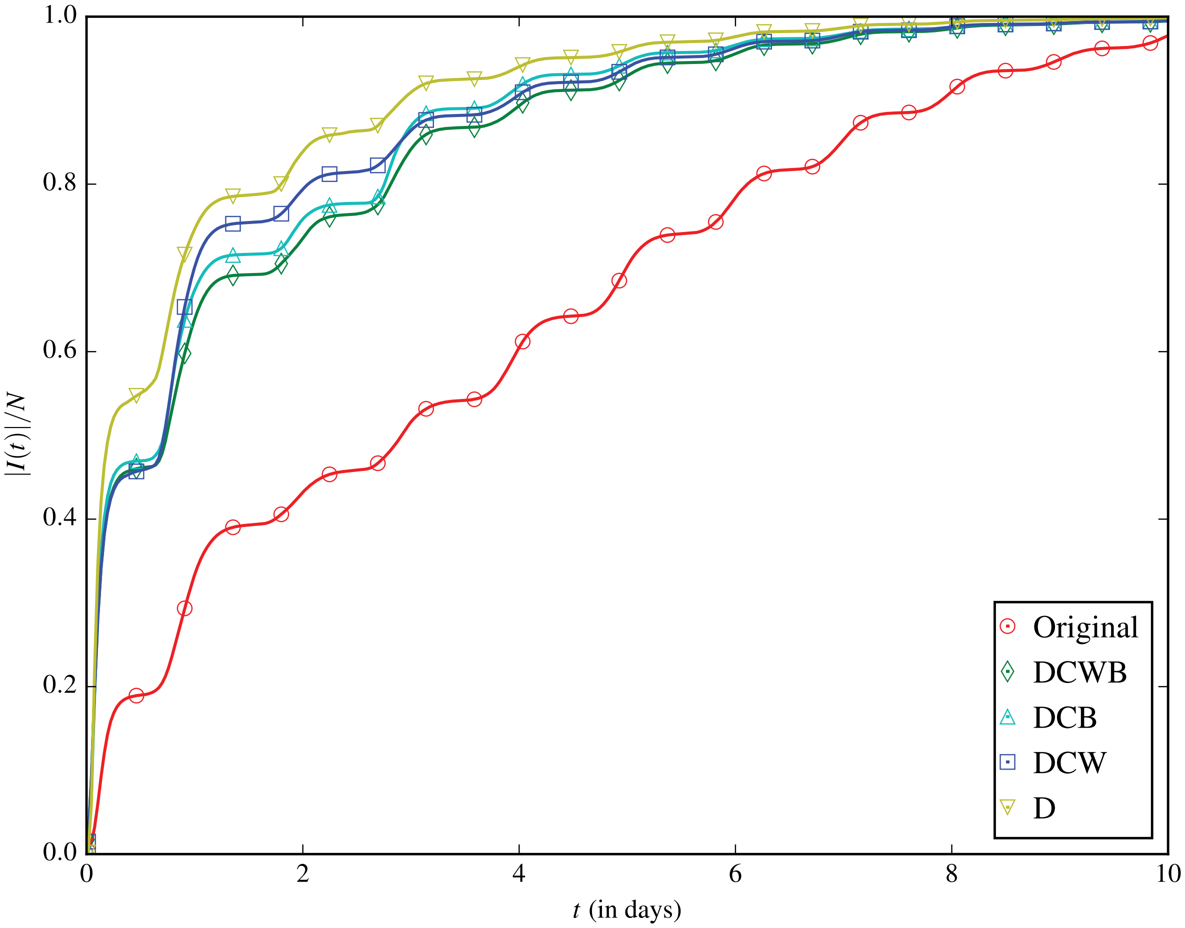

Figure 1. (color online) Fraction of infected devices $|I(t)|/N$ as a function of time, for the original contact trace and "Small But Slow World" null model contact shufflings. Each line is based on a uniformly spaced sampling of prevalence over time. Note that the sampling resolution of lines is higher than their respective markers, as we only plot every n-th marker ($n > 1$) to minimize visual clutter. This (not fitting) is the reason for nonlinearity between plotted points. The same applies for the other plots presented throughout this paper.

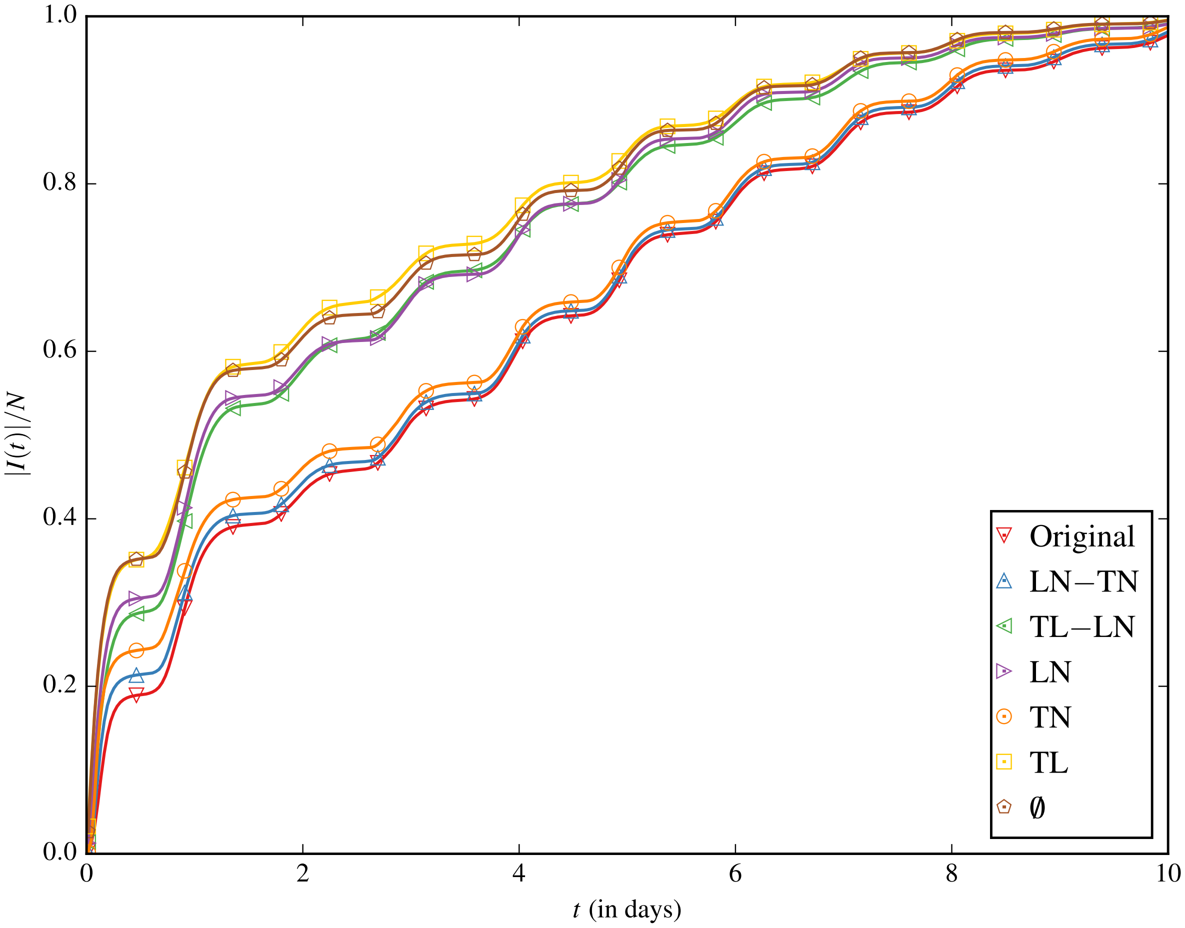

Figure 2. (color online) Fraction of infected devices $|I(t)|/N$ as a function of time, for the original contact trace and inducement-shuffled null models.

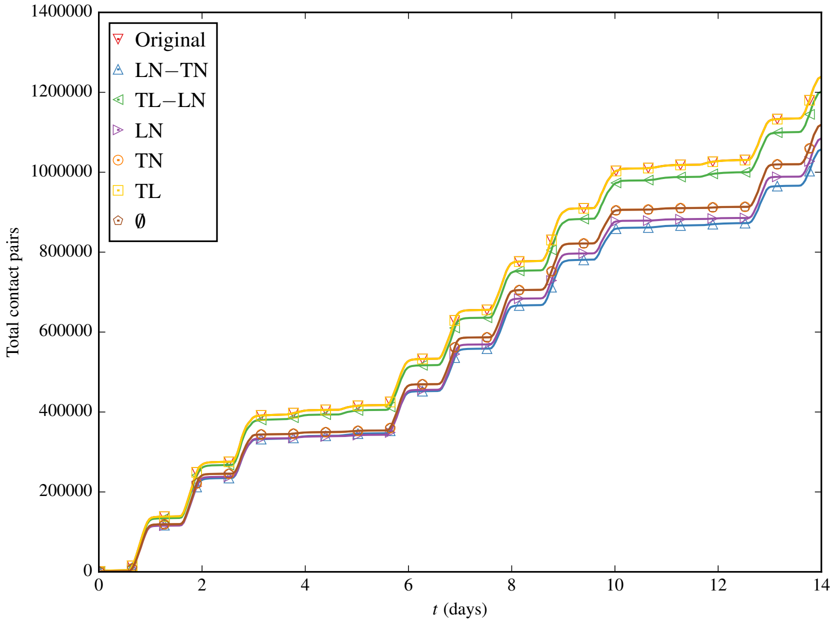

Figure 3. (color online) Total cumulative contacts as a function of time, for the original contact trace and contact traces induced by inducement shuffling.

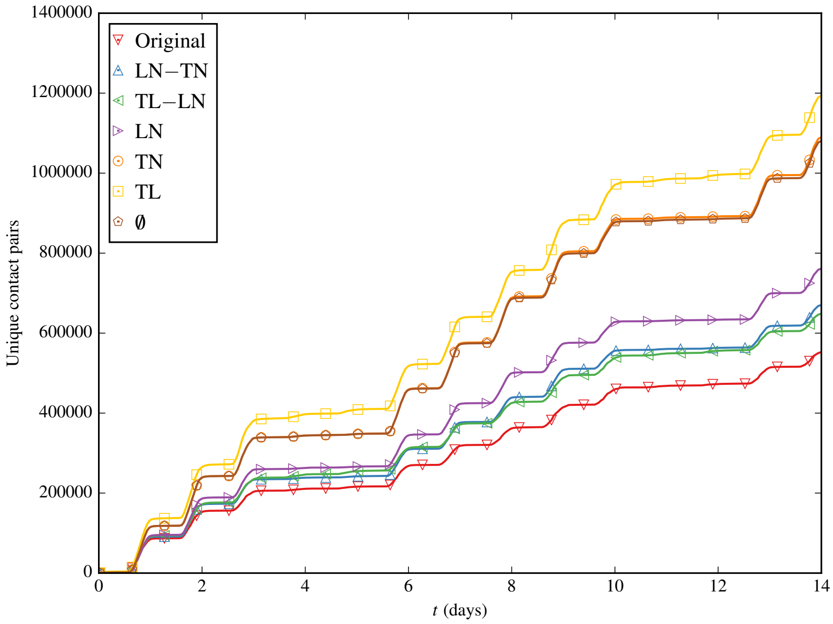

Figure 4. (color online) Unique cumulative contacts as a function of time, for the original contact trace and contact traces induced by inducement shuffling.

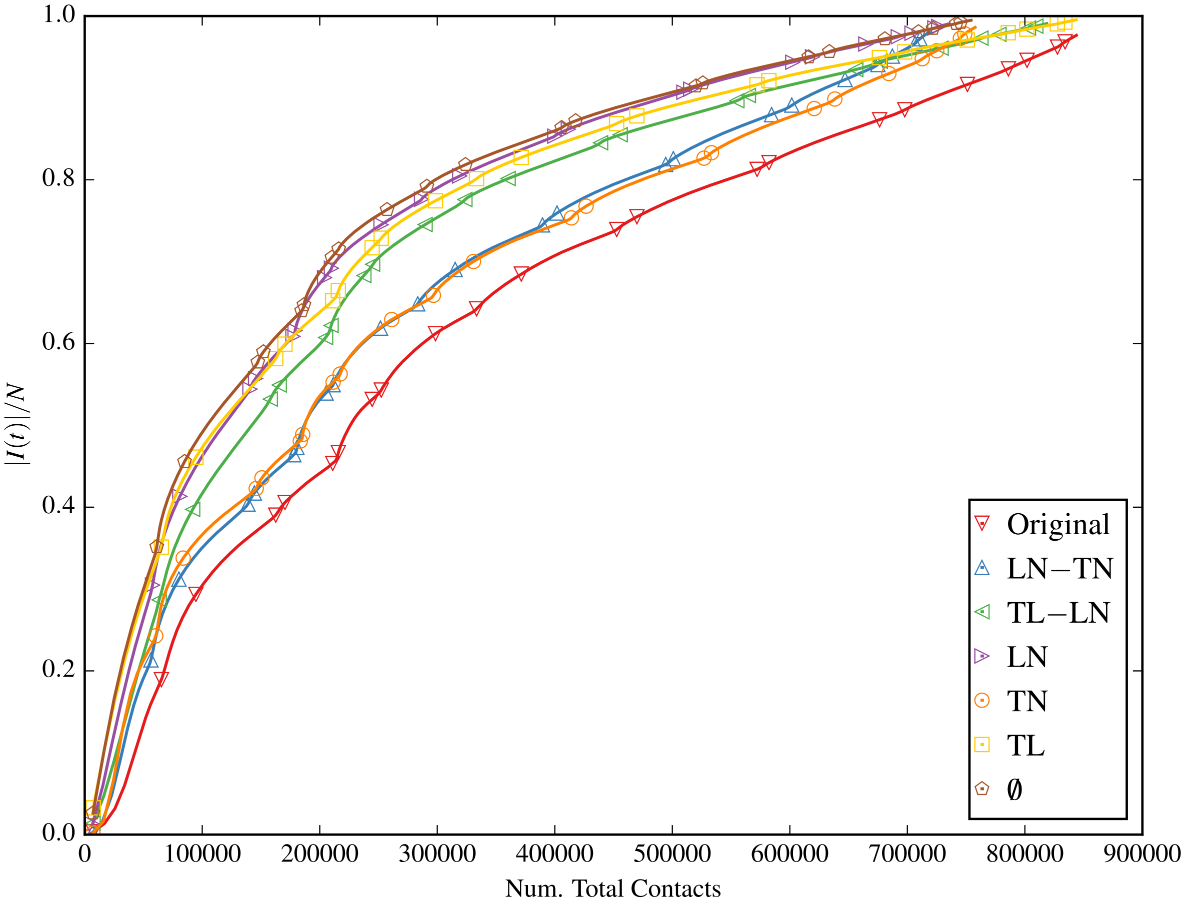

Figure 5. **(color online) Fraction of infected devices $|I(t)|/N$ as a function of the number of *total* contacts, for the original contact trace and inducement-shuffled null models.**

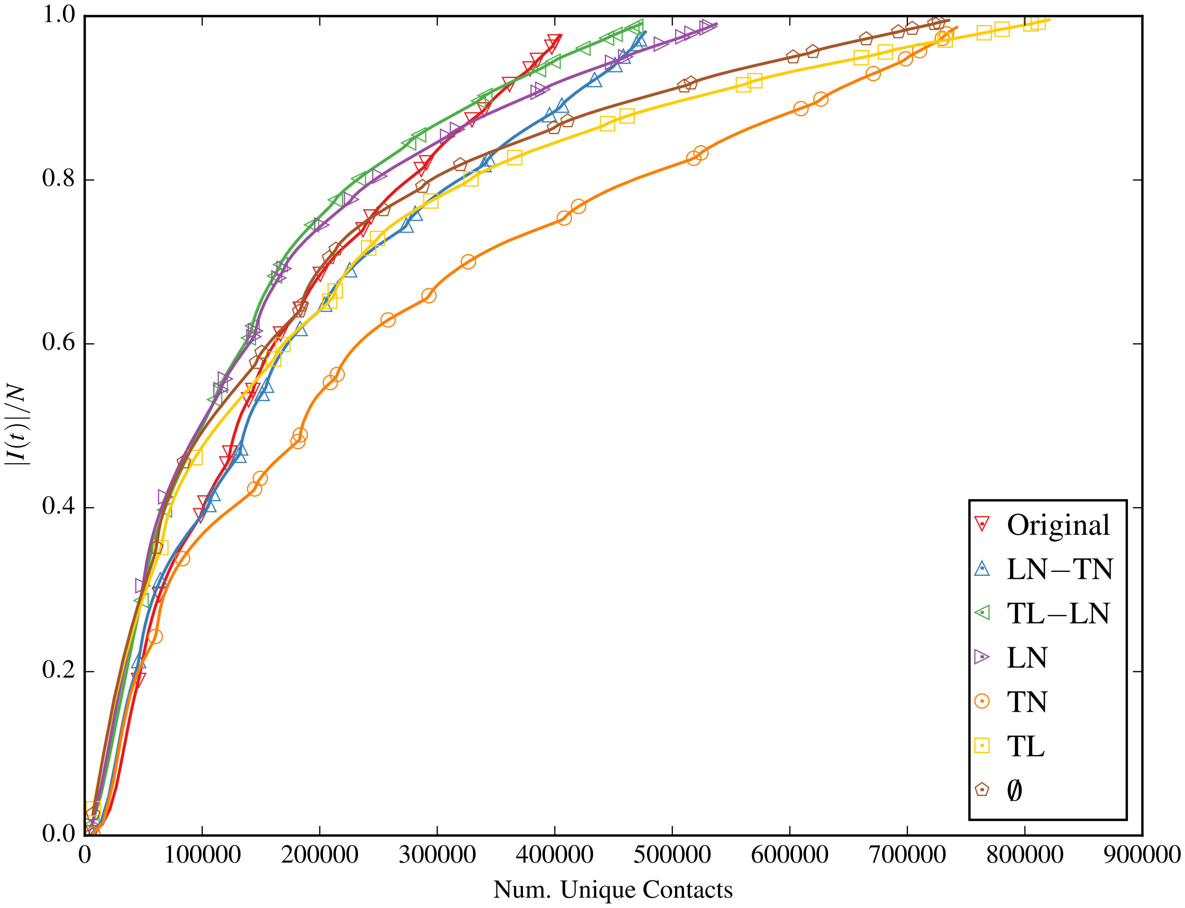

Figure 6. (color online) Fraction of infected devices $|I(t)|/N$ as a function of the number of *unique* contacts, for the original contact trace and inducement-shuffled null models.

Tables

Table 1. Null model summary and prevalence at 1 day ± standard error of the mean.

Shuffling	LN	TN	TL	$\|I(1\ day)\|/N$
Original	✓	✓	✓	33.4%±0.6%
Group node, shuff. time	✓	✓		35.0%±0.6%
Group location, shuff. node	✓		✓	44.6%±0.8%
Shuffle time	✓			46.5%±0.8%
Shuffle location		✓		37.3%±0.6%
Shuffle node			✓	50.5%±0.8%
Shuffle location, shuffle node	-	-	-	50.3%±0.7%

Table 2. Number of repeat contacts

repeats	count
1	385 453
2	70 092
> 2	96 298

Supporting Information Captions

We include Figures 7 and 8 to enable the interested reader to more closely compare pairwise prevalences under the inducement-shuffled null models. Figure 7 is designed primarily for macro-scale comparison of prevalences, i.e. approximately how much faster or slower is one shuffle versus another. By comparing specific row and columns the reader can easily see any large difference in prevalences as a function of time of a specific pair of shufflings as the y-axis scale encompasses the maximum absolute differences. Figure 8 on the other hand is designed for micro-scale comparison of prevalences. The y-axis has been zoomed, which hides differences of large magnitude but allows the reader to more clearly visualize standard errors and overlaps between prevalence pairs. This is designed in particular to allow the reader to gauge whether small but perceivable prevalence differences in Figure 2 are of much statistical significance. For example, the reader might notice the brief crossover between LN and TL-LN in Figure 2 and be curious how significant this result is. By comparing LN and TL-LN in Figure 8 one is then able to ascertain that there exists relatively large standard errors around both LN and TL-LN at the time of this measurement and so it is not likely to hold that TL-LN is ever faster than LN with much statistical significance. On the other hand, the reader may be interested in where there are clear differences, such as between TL and LN where there is a substantial amount of space between the two prevalences (TL being faster).

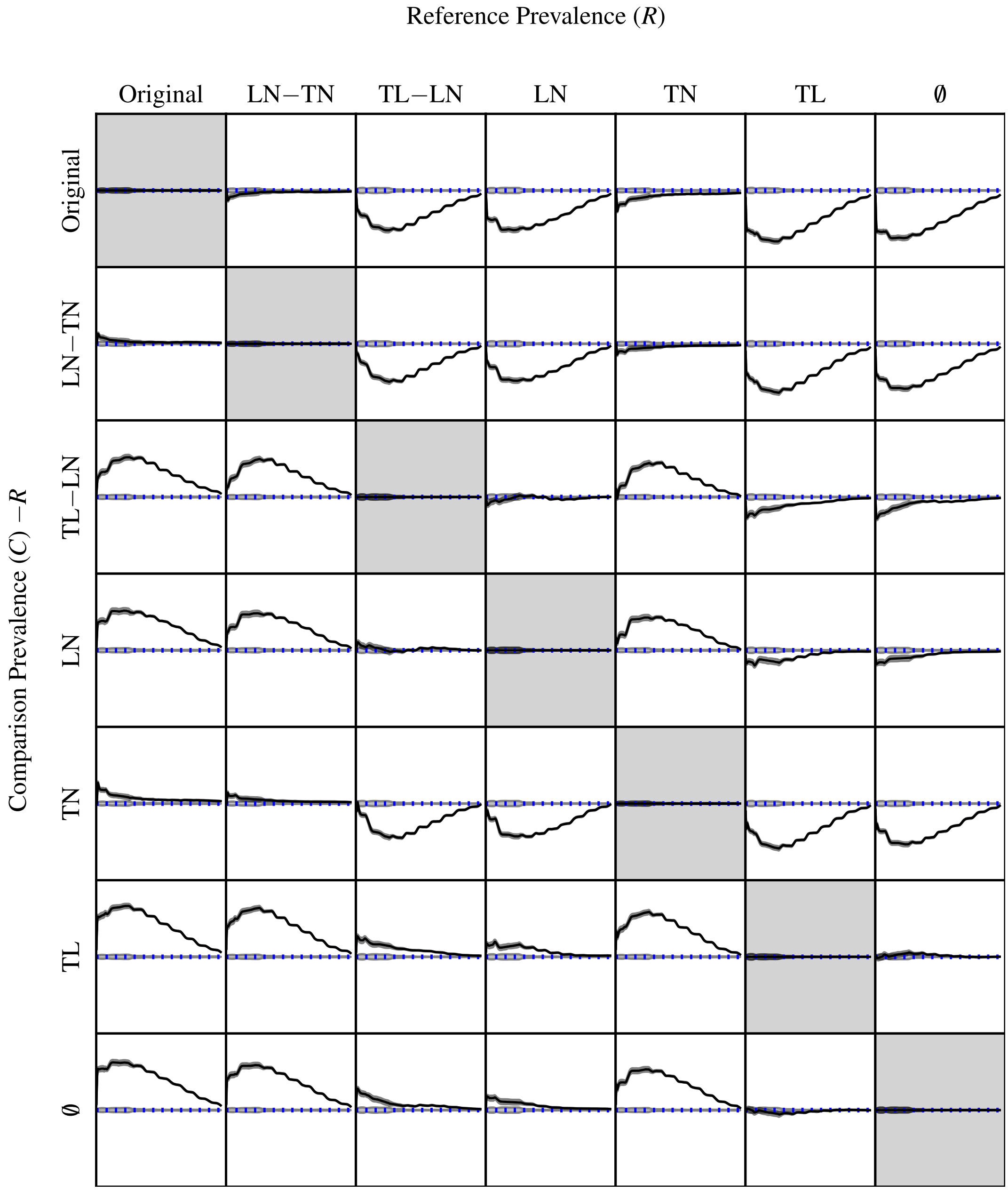

Figure 7. (color online) Macroscale prevalences over time of all shuffling pairs compared. The x-axis of each subplot is time and spans the 10-day simulation interval, domain [0,10]. The y-axis of each subplot is the column label's prevalence subtracted from the row label's prevalence and spans the range interval [-0.3,0.3] i.e. 30% either side of the blue dotted line which is centered at 0.0. The gray shaded region around each black line (very small in most plots but visible at high zoom) is the standard error of the mean of the row label's prevalence. The gray shaded region around the center dotted blue line is the standard error of the mean of the column label's prevalence.

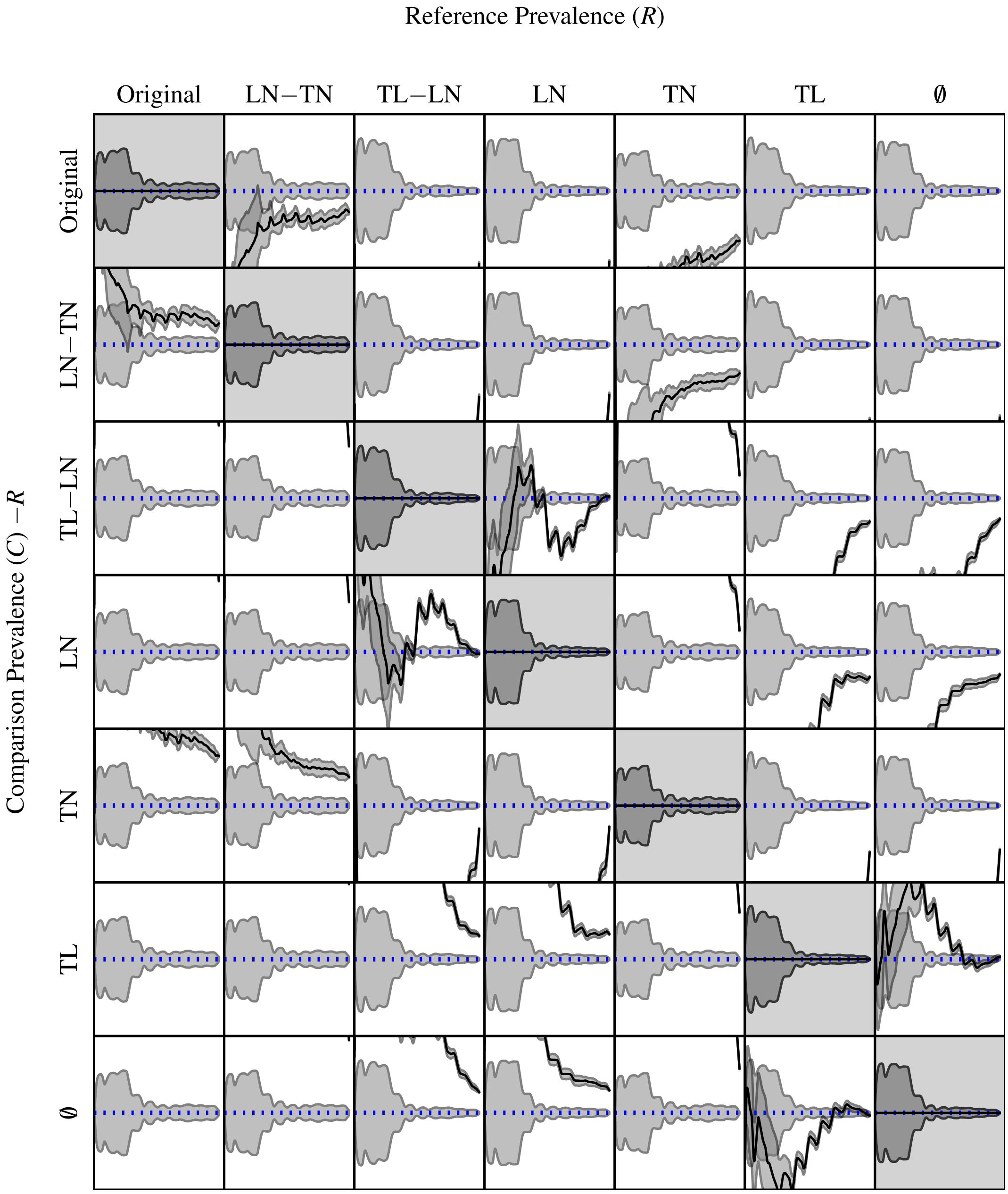

Figure 8. (color online) Microscale prevalences over time of all shuffling pairs compared. This figure is the same as Figure 7, the only difference being that the y-axis range has been zoomed to the interval [-0.015,0.015] i.e. 1.5% either side of the dotted blue line instead of 30% either side. This highlights the standard error of the mean more clearly and allows one to more easily establish regions of overlap between the reference and comparison's standard errors giving a better idea of where differences are of lower statistical significance.

8

Conclusion

50 years ago, few people would have envisioned just how profound an impact computers would come to have on society today. Through an almost unfathomable series of consistent long-term improvements in computer hardware, that which once cost millions of dollars and occupied a large room has been superseded by that which costs less than $100 and fits in your pocket. This ongoing miniaturization of computer electronics has led to devices becoming increasing mobile and increasingly wireless. For the present day Internet dominated by content distribution and retrieval, the proliferation of wireless mobile devices presents both challenges and opportunities around fast and efficient content delivery.

In recent times, huge geographically distributed Content Delivery Networks(CDNs) have proven their utility as a means of improving content distribution and retrieval. Much of the benefit of CDNs is derived from proximity to end users, achieved by situating servers inside Internet Service Providers' (ISPs') Points of Presence (PoP). The logical progression to the 'final frontier' of CDNs is to contemplate distribution networks in which the 'leaf node' client devices themselves act as content caches.

This thesis has analyzed spreading potential predicated on device colocation and mobility using the graph abstraction of a temporal contact network. The thesis has also elucidated complementary near-term practical improvements being built on top of existing network protocols, namely, SPDY as a drop-in replacement for HTTP.

Our initial publication was a technical overview of the SPDY protocol and its performance optimizations. SPDY has since gone on to form the basis of HTTP 2.0 and this paper has served as a point of contrast between near-term network research and the long-term network research which has received most of our attention in this thesis. In our second and third publications we provided empirical results on the diffusion potential of a large campus network parameterized around (i) time of day, (ii) day of week, (iii) site and (iv) number of content sources. In our final publication we proposed a set of general null models which can be used to decorrelate any spatiotemporal contact network predicated on colocation. We applied these models to the same contact trace analyzed in our earlier papers to determine the impact of nodes' spatiotemporal preferences on spreading.

8.1 Summary of Contributions and Results

The main contributions and results of the thesis are summarized as follows:

- We have shown diffusion potential to be highly dependent on site, day of week and the number of content sources. Larger sites with more activity appear more amenable to diffusion and weekdays diffuse more quickly than do weekends. More content sources increases the initial rate of diffusion, though has little substantial impact longer term.
- Trace timestamp uncertainties can have a meaningful impact on rates of diffusion, though mostly in absolute terms. The overall patterns of diffusion hold irrespective of the compensating assumption chosen.

- Diffusion rates can vary widely across individual source devices—some source devices simply have a propensity to result in far faster diffusion than others.
- With no compensating adjustments for trace timestamp uncertainties, up to a third of application requests could theoretically be satisified over a 12-hour period for the mobile maps use case analyzed in the thesis. In the optimistic case this increases to 70% of requests in the same 12-hour period.
- We have introduced a set of null models which can be used to decorrelate the spatiotemporal node preferences of any colocation-driven contact network. This can in turn be used to allow second-order causal reasoning about how the time and location preferences of nodes not only influence the number of contact events but also the subsequent rate of spreading.
- With respect to spatiotemporal node preferences, we find the correlation between nodes and times to be the greatest impediment to spreading. The correlation between times and locations slightly catalyzes spreading, though only in the absence of other correlations.
- Contact frequency and diffusion potential are surprisingly only loosely related. Certain shufflings of spatiotemporal node preferences produce contact networks which are more 'efficient' or 'amenable' to content spreading on a per-contact basis.
- Null models applied to the studied wireless trace suggest the impact of contact correlations has a greater slowing effect on diffusion than suggested by prior works.

8.2 Challenges, Limitations and Future Work

As is perhaps the case with many research projects, the research performed for this thesis has not come without challenges and limitations and there is always room for future work. In this section we highlight several prominent examples.

8.2.1 Access to Empirical Data

One of the greatest hurdles faced by the author was gaining access to empirical data. As mentioned earlier in the preamble to the first paper, the author had originally planned to gain access to complete application layer packet traces describing the content byte strings traversing the network. The objective was to simultaneously identify duplicate byte strings along with time-respecting mobility paths between pairs of devices. This would have served as direct evidence of redundant content which one user could have shared with another. For example, if (i) User A downloaded content X at time t_1, (ii) User B downloaded content X at time t_n and (iii) there was a time-respecting path between User A and User B in the interval $[t_1, t_n]$, then from this we could ascertain that User A could have shared content X with User B. This type of analysis would essentially give a more complete insight into where real content might have been shared without making assumptions about who is and is not interested in an arbitrary piece of content.

During the early stages of the PhD, the thesis author pursued access to full packet traces through James Cook University (JCU). These proved difficult to attain due to technical limitations combined with a lack of timely vendor support. More broadly, issues of privacy are a likely reason why such full packet datasets prove so elusive in public data repositories. Arguably, large private technology companies have better access to rich empirical datasets making this an area where industry likely has a comparative advantage over academia. The thesis author hopes that more organizations will find ways to make rich empirical datasets publicly available while at the same time preserving the privacy of their users. For researchers who are successful in acquiring full packet traces, there are valuable contributions to be made in characterizing content redundancy across wireless mobile devices, as well as its implications on the potential of wireless Peer-to-Peer (P2P) content diffusion.

8.2.2 External Validity

A large portion of the results presented in this thesis are based on a single dataset. Future work should look to either corroborate or contrast our findings in other networked environments. This includes not only different geographic settings such as cities, towns and residential areas, but also different network types including cellular and bluetooth, where nodes may qualify as 'colocated' across either larger or smaller distances than Wi-Fi.

8.2.3 Real-world Applications

Despite Mobile Ad hoc NETworks(MANETs) being an enduring research topic, there are surprisingly few examples of large-scale deployed MANET applications from which practical lessons can be learnt. It is the thesis author's view that the time is now ripe for pragmatists to make valuable contributions in this area. In the present day, smartphones are essentially ubiquitous throughout the developed world. This provides an excellent existing platform upon which MANET-based applications may be built. For example, the JCUNav application referred to throughout this thesis may potentially be extended to *actually* share cached mapping data directly between peer devices. Although the 'appliancized' [46] nature of modern smartphones imposes relatively restrictive Application Programming Interfaces(APIs), there is also much promise in using peripheral devices to overcome these limits. As described later in Appendix B.2, the thesis author succeeded in receiving unfiltered over-the-air Wi-Fi packets on an 'unrooted' Android handset using an external USB Wi-Fi adapter. With some work, this could be extended to packet *transmission*, allowing essentially unrestricted access for experimentation with wireless broadcast and unicast protocols on commodity handset devices. With additional expertise in USB hardware development, it may even prove salutary to develop an affordable bespoke hardware peripheral. This might in turn become a de facto standard instrument in MANET experimentation.

8.2.4 Interdisciplinary Integration

The final paper in this thesis touched on the connection between wireless P2P content diffusion and the broader field of network dynamics. Indeed, the contact network considered in this thesis is just one instantiation of a temporal network—a network abstraction used to describe a wide range of time-variant systems. What the broader field of network dynamics can contribute to the study of MANETs and vice versa is another interesting area for future work.

Appendices

Appendix A

List of Thesis Papers

1. **B. Thomas**, R. Jurdak, I. Atkinson, "SPDYing Up the Web", *Communications of the ACM (CACM)*, vol. 55, no. 12, pp. 64–73, Dec. 2012.

2. **B. Thomas**, I. Atkinson, R. Jurdak, "Content Diffusion in Wireless MANETs: the Impact of Mobility and Demand", in *IEEE International Wireless Communications & Mobile Computing Conference (IWCMC), 2014 International*, Aug 2014, pp. 959–966.

3. **B. Thomas**, R. Jurdak, I. Atkinson, "The Impact of Mobility and Content Demand on Diffusion in Wireless MANETs", *Elsevier Ad Hoc Networks*, Submitted for consideration (not yet accepted).

4. **B. Thomas**, R. Jurdak, K. Zhao, I. Atkinson, "Diffusion in Colocation Contact Networks: the Impact of Nodal Spatiotemporal Dynamics", *PLOS ONE*, Submitted for consideration (not yet accepted).

Appendix B

Research By-products

Aside from the research papers constituting the core of the thesis, a number of noteworthy by-products where produced in the course of the thesis author's studies. In this appendix we highlight two prominent examples—*JCUNav* and *liber80211*.

B.1 JCUNav

In the early planning stages of the author's thesis, substantial effort was invested in identifying appropriate sources of empirical data describing device mobility patterns. During this time, the author made a serendipitous finding—the official James Cook University (JCU) digital map had not been designed to work well on mobile devices. Because the official map did not work well on mobile devices, the author posited that a well designed unofficial application may prove popular. The *JCUNav* application (screenshot in Figure B.1) began with the simple idea that it would be useful to the user to have their current location based on the Global Positioning System (GPS) shown on top of a map of the JCU campus. This would allow the user to more easily find their way around the campus.

The research motivation behind JCUNav was that the application might allow empirical device mobility patterns to be collected which could then be used in the author's content dissemination simulations. Particularly, mobility patterns from the *device's* point of view which may include more continuous data than that gathered from the infrastructure. Moreover, a core objective of JCUNav was to overcome the sampling limitations seen in many research papers where handsets are explicitly provided to a small number of often highly correlated individuals (e.g. 10's of research students in a single building on a university campus). By writing an application which could be promoted to the public and easily installed, JCUNav would offer a larger scale measurement platform with less sampling bias. Furthermore, the application would be run on users' existing handsets and so would avoid the capital expense associated with purchasing dedicated devices.

In the remainder of this section, the major versions of JCUNav are enumerated and the extracted data is discussed. JCUNav is closed-source software—the copyright and intellectual property resides with the thesis author. More information about JCUNav can be found at `jcunav.com`.

B.1.1 JCUNav Versions

From JCUNav's inception to the present day, several major versions of the application have been released for the web, Apple iOS and Google Android platforms. We detail these versions in this section.

JCUNav Beta

JCUNav began as a simple mobile webpage which utilized the HTML5 geolocation Application Programming Interface (API) to show the user's current position on top of a map of the JCU campus. The user was able to select a building from a

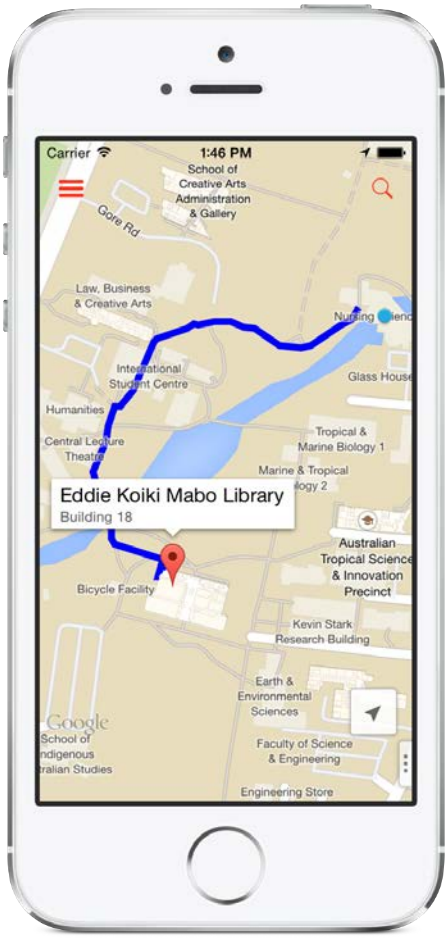

Figure B.1: A screenshot of JCUNav running on the Apple iPhone.

list and a straight line would be overlayed on the map between the user and the destination. This version was promoted in July 2011 through the JCU staff and student bulletin mailing lists.

JCUNav 1.0 for iOS

Subsequent development of JCUNav focused on an installable mobile application. The first version for iOS (1.0) was simply an embedded webpage inside an invisible frame that utilized the native device's compass to show the user's heading in real-time. This version was released in February 2012 on the Apple iOS App Store. The author envisaged that having an installed application would likely provide better access to the device's location for research data collection than would a simple webpage. For the user, the installable application had the advantage of being conveniently located on the home screen, working offline and showing the user's heading. JCUNav's first ever user review on the iOS App Store is pictured in Figure B.2.

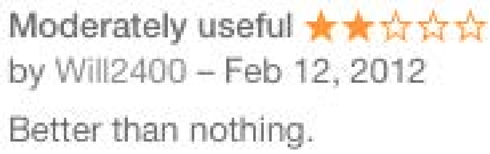

Figure B.2: JCUNav's first ever user review on the iOS App Store.

JCUNav 2.0 for iOS

Version 2.0 of JCUNav was released in March 2012 and was a complete overhaul that replaced the embedded webpage with the native iOS maps framework while retaining the same JCU map overlay. This drastically improved the application's performance and fluidity, particularly when panning and zooming the map. A review of JCUNav version 2.0 is pictured in Figure B.3.

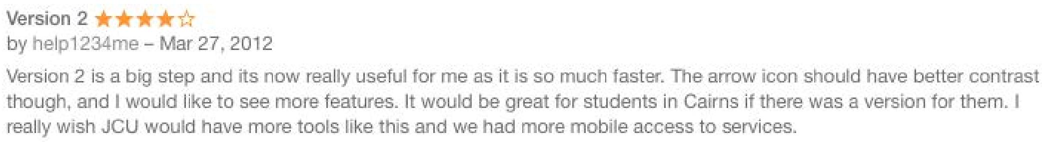

Figure B.3: A user review of JCUNav version 2.0 on the iOS App Store.

JCUNav 1.0 for Android

In October 2012 the first version of JCUNav for the Android platform was released. This version was largely comparable in features to the 2.0 version of JCUNav on iOS though originally lacked building labels as Google Maps was being used alone with no JCU map overlay. Around this time, Google Map Maker [47] became available in Australia which allowed users to contribute building names and outlines. The author subsequently contributed the majority of JCU's building names and outlines to Google which began showing on the Android version of JCUNav within a few days.

JCUNav 3.0 for iOS

Version 3.0 of JCUNav for iOS was released in December 2012. This version added the people search feature which allows users to quickly search for faculty and staff on campus, launching a phone call to the selected person with the press of a button. The people search feature was designed to help make the application increasingly compelling to attract as wide an audience as possible.

JCUNav 2.0 for Android

The 2.0 version of JCUNav for Android was also released in December 2012. At the time, Google had just completed an overhaul of the Google Maps Android library which served as the impetus for this release. The new version of the maps library drastically improved the fluidity of the map as well as offering a more advanced perspective positioning system which made for a more compelling user experience.

JCUNav 4.0 for iOS

Version 4.0 of JCUNav for iOS was released in June 2013 and was the first version to replace the old JCU raster map overlay with Google's own native vector overlay. From the time of release of Google Map Maker in Australia, the author had continually been submitting improved JCU mapping data to Google Maps such that it would appear natively without the need for an overlay. The vector maps reduced the application size dramatically and offered features inherited directly from Google Maps, such as building names and outlines showing only at certain zoom levels. The vector paths added through Google Map Maker also enabled this version of JCUNav to be the first version to integrate detailed directions to destinations using the Google Maps directions API.

Table B.1: JCUNav key statistics.

Installations since inception	> 7000
Peak daily users	> 600
Typical in-term daily users	$\approx$ 20–100

B.1.2 JCUNav Data

Originally the author had envisioned that JCUNav might provide empirical research data on the mobility patterns of devices as they moved around campus. This particular type of data was ultimately not used in the author's research, partly for technical reasons and partly to bound the scope of the research. Rather, the thesis author gathered basic usage statistics describing both the number of users on any given day and the time of day the application was being used. This data was used in both the second and third paper. We conclude our discussion of JCUNav with the key usage statistics summarized in Table B.1.

B.2 liber80211

Early 2012 proved to be a difficult time for gaining access to empirical device mobility and encounter data. Originally it had been planned that in the second half of 2011 JCU would provide a timestamped trace of device connections to individual Access Points(APs) on the JCU campus. Due to the ongoing discovery of technical limitations in JCU's Wi-Fi vendor's software, by 2012 this was looking increasingly unlikely. At the time, JCUNav was still not popular enough to gather a meaningful volume of data, particularly given that it was proving difficult to record device's geolocation without the application running in the foreground. The University of Queensland device connection trace which was ultimately the primary trace used in the PhD had also not yet been established as an option.

The author began contemplating other ways to collect large-scale traces which might describe at least some aspects of device mobility and encounter patterns. While studying the technical details of the Wi-Fi protocol, the thesis author learnt that client devices searching for wireless networks would often broadcast in plain text their identity (MAC address) and in some instances the name of networks to which they had previously connected. This could occur even when the client devices were well outside the range of the prior networks. The author thought that this information, if collected in great enough volume, might serve as a compelling dataset describing certain aspects of spatiotemporal device mobility. Firstly, an observation of a given MAC address at a given monitored location would offer a data point describing both the spatial and temporal presence of a given device. Secondly, an observation of a prior network name (an Extended Service Set IDentifier (ESSID)) might be resolved to a geographic location using a public crowdsourced ESSID database (described shortly). This would offer prior spatial information, albeit without temporal certainty, other than to say the client device had connected to the named network 'sometime in the past'.

One of the challenges the author faced with this thread of research was that collecting the requisite data would require a Wi-Fi adapter that could be put into 'monitor mode'. Furthermore, there would need to be many monitoring devices in order to achieve adequate coverage. Though USB Wi-Fi adapters supporting monitor mode were themselves inexpensive, host machines such as laptops which had sufficient battery life and compact form factor were not. Even cheap computing platforms such as the Raspberry Pi [48] would require the additional expense of a battery and storage, as well as an external GPS if data were to be collected 'on the go'. The thesis author thought that if a USB Wi-Fi adapter were able to be connected to an inexpensive Android mobile handset, the total unit cost could be reduced, making large-scale data collection more affordable. An Android handset would not only come with its own battery and GPS, but would also offer a simple 'plug and play' setup, meaning monitoring units could be easily used by data collection volunteers. The desire to be able to use an inexpensive USB

Wi-Fi adapter in combination with an Android handset is what ultimately led to the development of *liber80211* (pronounced liberate-0211).

Figure B.4: A screenshot of an Android handset running liber80211 interfacing with an external USB Wi-Fi adapter and powered via an external USB battery pack.

B.2.1 IEEE 802.11 Wi-Fi Probe Requests

Probe Requests(PRs) are a special type of IEEE 802.11 management frame which wireless clients use to proactively search for wireless networks. PRs possess two interesting properties in the current discussion: (i) they typically contain the unique MAC address of the transmitting client and (ii) some PRs contain a named ESSID, typically the name of a network the device has previously used. As an example, the client MAC `AA:BB:CC:DD:EE:FF` may probe for the ESSID `Townsville Megamart`. This would be a strong indication that at some point in the past, this client had been connected to a network named 'Townsville Megamart'. Observing a PR in and of itself reveals the presence of a given client at the observation location and observation time. As explained in the next subsection, the ESSID can in some instances also be cross-checked against a geographic database to identify a prior location of the client.

B.2.2 Geolocating ESSIDs with WiGLE

Wi-Fi *networks* typically broadcast their ESSIDs perpetually in *beacon* frames as a way to let clients know of their existence. This is why on most 802.11 Wi-Fi devices a user is able to see the names of networks in range, e.g. 'Starbucks', 'John Smith Wireless', 'Acme Shopping Center'. Crowdsourced mobile applications such as WiGLE [49] allow users to automatically contribute ESSID locations with a simple application running on a commodity handset. The handset's location services (e.g. GPS) are recorded alongside observed ESSIDs which allows a worldwide geographic database of ESSIDs to be constructed. It is these same ESSIDs that *client* devices searching for prior networks often include in PRs. Because of this, one is able to cross-reference observed ESSIDs in PRs to the known location of ESSIDs garnered from independently observed beacon messages. Because PRs that include ESSIDs typically use the ESSIDs of networks to which they have previously connected, an observer is able to use this cross-referencing to identify one of a client device's prior locations.

B.2.3 Technical Overview of liber80211

liber80211 is an open-source Android application that enables IEEE 802.11 Wi-Fi 'monitor mode' on non-rooted Android devices using an external USB Wi-Fi adapter (pictured in Figure B.4). Monitor mode is a special mode of Wi-Fi that enables wireless frames not addressed to the receiving device to be propagated from the operating system's kernel space into user space for processing by non-privileged applications. A regular kernel space Wi-Fi driver (one not operating in monitor mode) drops frames not addressed to the device before user space programs are given the opportunity to process them. PRs are one such type of frame that a regular client Wi-Fi driver drops in kernel space. Having a client Wi-Fi driver operate in monitor mode is therefore a prerequisite for receiving and processing PRs in non-privileged applications.

Since version 3.1 of the Android operating system, Android has supported USB On-The-Go (USB OTG) which allows standard non-privileged Android applications running in user space on a mobile device to interface with arbitrary USB peripherals. The Android operating system is oblivious to the nature of the connected peripheral—it is left to the application programmer to use the Android USB OTG API to programmatically communicate between the handset and the attachment. This communication takes place at a very low level, with the application programmer having to define the I/O operations in terms of byte streams.

The thesis author began work on liber80211 by identifying a high quality and low cost USB Wi-Fi adapter which had the capacity to be put into monitor mode. The adapter in question is the Alfa Network AWUS036H which is based on the Realtek RTL8187 series of Wi-Fi drivers. This Wi-Fi adapter has been popular for several years with Wi-Fi penetration testers and an open source Linux kernel driver is freely available. By carefully inspecting the original Linux driver source code (written in C) and cross checking against thousands of lines of USB I/O on a Linux computer, the author was able to establish enough about the host-to-peripheral messaging to port a portion of the driver to Android user space (written in Java). This allowed 802.11 PR frames to be collected on an Android phone without needing to 'root' the handset.

The AWUS036H is an extremely low-level device in terms of the programming necessary to initialize the card. The card contains an Electrically Erasable Programmable Read-Only Memory (EEPROM) which needs to be initialized manually over the USB connection each time the device is connected. This required thousands of carefully crafted byte stream patterns to be transmitted in the correct order to initialize the card before 802.11 Wi-Fi frames could be received. The Linux source driver was based on the offical Realtek Windows driver. Both the original Windows driver and Linux driver have very little documentation. It was therefore necessary to carefully work through the Linux kernel driver code and ensure the relevant USB messages were faithfully replicated and interpreted in the same way in liber80211.

B.2.4 Challenges to Probe Request-based Monitoring

liber80211 was ultimately not used to produce any research outputs in the author's thesis for a number of reasons:

1. USB OTG, though technically supported by the Android operating system since version 3.1, has been poorly implemented on the majority of Android handsets. This limited handset options to more expensive Android devices which were financially prohibitive for large-scale deployment. The author had thought the price of Android handsets with full USB OTG support might fall faster than it did in practice. Although several compellingly cheap handsets with purported full USB OTG support have since become available, these were not on the market at the time liber80211 was developed.

2. The WiGLE ESSID geo-database has API throttling in place at a level prohibitive to resolving any reasonable portion of all observed ESSIDs. The throttling mechanisms used by WiGLE are opaque and after a large number of experiments well spaced over time, the author found it increasingly difficult to extract more data.

3. A large fraction of PRs included only the broadcast ESSID. This is an

ESSID which reveals nothing about any of a device's prior networks and so cannot be resolved to a prior location.

4. Disambiguation of ESSID locations proved difficult for common ESSIDs. For example, many networks have names such as "netgear" or "dlink" which are common across thousands if not millions of networks worldwide.

PRs have seemingly become an even less reliable tool for mobility analysis since liber80211 was published. In particular, in the iOS 8 operating system, Apple has identified PRs as a potential privacy concern and has begun randomizing the sender's MAC address.

B.2.5 Source Code and Related Work

The source code for liber80211 is publicly available at [50]. The installable Android application is available at [51]. At time of release, liber80211 was the first publicly available implementation of 802.11 monitor mode support on an unrooted Android handset. Since the release of liber80211, another party has released a more feature-rich packet capture tool for Android—*Wi-Fi PCAP Capture* [52].

Around the time liber80211 was being developed, a number of papers were published on the use of PR for inferring other information about devices and their users. This includes features such as language, social structure and handset vendor adoption. We refer the interested reader to the papers in [53] and [54] for more information.

9

Acronyms & Glossary

Acronyms

API Application Programming Interface. 93, 98, 101, 106, 107

AP Access Point. 6, 102

BDP Bandwidth-Delay Product. 19

CCN Content-Centric Networking. vi, 3, 4, 23–27

CDN Content Delivery Network. vi, 3, 4, 89

DNS Domain Name System. 21

DPI Deep Packet Inspection. 31

DTN Delay-Tolerant Networking. 29

EEPROM Electrically Erasable Programmable Read-Only Memory. 106

ESSID Extended Service Set IDentifier. 103–105

FIA Future Internet Architecture. 22

FIB Forwarding Information Base. 26

GPS Global Positioning System. 97, 105

ICN Information-Centric Networking. v, vi, 3, 7, 22, 23

IP Internet Protocol. vi, 2, 3, 11, 18, 20, 21, 23–27

ISP Internet Service Provider. vi, 4, 19, 89

JCU James Cook University. 92, 97, 98, 100, 102

MANET Mobile Ad hoc NETwork. 28, 93, 94

NDN Named Data Networking. 23

NSF National Science Foundation. 23

P2P Peer-to-Peer. 4–7, 21, 24, 27, 28, 30, 93, 94

PARC Palo Alto Research Center. 23

PIT Pending Interest Table. 26

PoP Point of Presence. vi, 4, 89

PR Probe Request. 104–108

TCP Transmission Control Protocol. 11, 18, 19

UQ University of Queensland. 102

USB OTG USB On-The-Go. 106, 107

Glossary

Access Point A wireless infrastructure node to which client devices connect. 6, 102

Application Programming Interface The interface exposed to a programmer against which software is written. 93, 98, 101, 106, 107

Bandwidth-Delay Product The product of a data link's capacity and round-trip latency. 19

circuit switching A method of computer networking in which two parties establish a dedicated communication channel (circuit). 12, 19, 20

Content Delivery Network A geographically and topologically distributed network of servers on the Internet designed to offer high performance and high availability content delivery. vi, 3, 4, 89

Content-Centric Networking A computer networking architecture beginning as a research project at Palo Alto Research Center that treats named content chunks as the primary network abstraction, rather than host IP addresses. vi, 3, 4, 23–27

Deep Packet Inspection A form of packet analysis which considers the packet data in addition to the packet headers. 31

Delay-Tolerant Networking A networking paradigm which emphasizes the eventual delivery of messages in the absence of continuous connectivity between all network nodes. 29

Domain Name System A distributed naming system which, among other tasks, is used to translate domain names to IP addresses. 21

Electrically Erasable Programmable Read-Only Memory A type of non-volatile programmable computer memory. 106

Extended Service Set IDentifier The name of an access point-based wireless infrastructure network, e.g. "Bob's wireless". 103–105

Forwarding Information Base Under the Content-Centric Networking (CCN) paradigm, a table mapping outbound faces to content sources. Longest prefix match lookups are performed against the FIB to determine which face(s) an interest request ought to be forwarded over. 26

Future Internet Architecture Any forward-looking Internet architecture designed to accommodate the networking needs of the 21st century. 22

Information-Centric Networking An umbrella title for computer networking approaches that focuses on named information (content) as the primary network abstraction rather than host addresses. v, vi, 3, 7, 22, 23

Internet Protocol The network ('thin waist') layer protocol of the current Internet. vi, 2, 3, 11, 18, 20, 21, 23–27

James Cook University A tertiary education institution based in Townsville, North Queensland, Australia. 92, 97, 98, 100, 102

JCUNav A mobile campus maps application developed by the thesis author. Usage statistics from JCUNav have been incorporated into the thesis author's research. 8, 43, 93, 97–99, 101, 102

kernel space The restricted area of an operating system's memory where the core functions of the operating system execute. cf. user space. 105, 106

liber80211 A user space application written by the thesis author which runs on the Android mobile operating system in conjunction with an external USB Wi-Fi adapter. liber80211 enables PR frames to be interpreted in user space on the device rather than having the PRs discarded by the device's own kernel space driver. 104, 105

Mobile Ad hoc NETwork An infrastructureless network comprised of wireless mobile nodes which communicate in an ad hoc fashion, often during opportunistic encounters. 28, 93, 94

Named Data Networking A Future Internet Architecture (FIA) with a close resemblance to CCN. 23

National Science Foundation A United States government agency responsible for promoting science and engineering research and education. 23

packet switching A digital networking paradigm that divides transmitted data into explicitly addressed blocks that may be routed over independent paths from a source to a destination. cf. circuit switching. 1, 2, 16, 17, 19, 20

Palo Alto Research Center The research and development company behind the earliest work on CCN. 23

Peer-to-Peer A network structure where client and server devices are not mutually exclusive, i.e. devices can take on both roles.. 4–7, 21, 24, 27, 28, 30, 93, 94

Pending Interest Table Under the CCN paradigm, a table recording content interests which a device is waiting to have fulfilled. 26

Probe Request A type of IEEE 802.11 Wi-Fi frame broadcast by wireless clients in search of wireless networks. 104–108

telegraphy The transmission of symbols/information over long distances without the physical exchange of objects. 2, 6

Transmission Control Protocol The transport layer protocol of the current Internet. 11, 18, 19

USB On-The-Go A USB standard which allows devices to switch between the roles of host and peripheral. For example, a mobile phone may act as the peripheral device when plugged into a desktop computer while acting as the host device when connected to an external USB storage drive. 106, 107

user space A restricted area of memory in which user applications execute without special priviledges. User space is designed to minimize application's ability to adversely effect the stability and security of the operating system. cf. kernel space. 105, 106

References

[1] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, "Networking named content," in *Proceedings of the 5th international conference on Emerging networking experiments and technologies*, ser. CoNEXT '09. New York, NY, USA: ACM, 2009, pp. 1–12.

[2] *RFC 791 Internet Protocol—DARPA Internet Program, Protocol Specification*, Internet Engineering Task Force, Sep. 1981. [Online]. Available: http://tools.ietf.org/html/rfc791

[3] Akamai, "Cloud computing services and Content Delivery Network (CDN) provider," http://www.akamai.com/, accessed: Jan. 11, 2015.

[4] ——, "Facts & figures," http://www.akamai.com/html/about/facts_figures.html, accessed: Jan. 11, 2015.

[5] MIT Technology Review, "Tom Leighton: A culture of innovation," https://www.youtube.com/watch?v=jMOvvQdTBbs#t=400, 2013, accessed: Jan. 11, 2015.

[6] B. Thomas, R. Jurdak, and I. Atkinson, "Spdying up the web," *Commun. ACM*, vol. 55, no. 12, pp. 64–73, Dec. 2012.

[7] B. Thomas, I. Atkinson, and R. Jurdak, "Content diffusion in wireless manets: The impact of mobility and demand," in *Wireless Communications and Mobile Computing Conference (IWCMC), 2014 International*, Aug. 2014, pp. 959–966.

[8] K. B. Miller, *American telephone practice.* McGraw publishing company, 1905. [Online]. Available: http://books.google.com.au/books?id=J6mEAAAAIAAJ

[9] R. R. Panko, *Business Data Networks and Telecommunications*, 6th ed. Pearson, 2007.

[10] B. A. Forouzan, *Tcp/Ip Protocol Suite*, 3rd ed. McGraw-Hill, 2005.

[11] A. G. Bell, "US patent 174,465—improvement in telegraphy," http://www.google.com/patents/US174465, 1876, accessed: Jan. 11, 2015.

[12] B. H. Baker, *The Gray Matter—The Forgotten Story of the Telephone.* Telepress, 2001.

[13] S. F. B. Morse, "US patent 1,647—improvement in the mode of communicating information by signals by the application of electro-magnetism," http://www.google.com/patents/US1647, 1840, accessed: Jan. 11, 2015.

[14] A. P. Deschanel, "Electric telegraphs," *Popular Science Monthly*, vol. 3, pp. 401–419, May 1873.

[15] "National historic landmarks program—site of the first telephone exchange," http://www.cr.nps.gov/nhl/DOE_dedesignations/Telephone.htm, 2011, accessed: Jan. 11, 2015.

[16] Seattle Municipal Archives, "Telephone operators, 1952," http://www.flickr.com/photos/24256351@N04/2680257100/, 1952, accessed: Jan. 11, 2015.

[17] A. Strowger, "US patent 447918—automatic telephone exchange," http://www.google.com/patents?id=PShCAAAAEBAJ&printsec=frontcover&dq=447918&hl=en&ei=hz07ToaXLKremAXi9uX3Ag&sa=X&oi=book_result&ct=result&resnum=4&ved=0CC8Q6AEwAw, 1891, accessed: Jan. 11, 2015.

[18] "The strowger switch," http://people.seas.harvard.edu/~jones/cscie129/nu_lectures/lecture11/switching/strowger/strowger.html, 2004, accessed: Jan. 11, 2015.

[19] P. Baran, "On distributed communications networks," *Communications Systems, IEEE Transactions on*, vol. 12, no. 1, pp. 1–9, Mar. 1964.

[20] RAND Corporation, "Paul baran and the origins of the internet," http://www.rand.org/about/history/baran.html, accessed: Jan. 11, 2015.

[21] K. Hafner and M. Lyon, *Where Wizards Stay up Late: The Origins of the Internet*. Econo-Clad Books, 1998.

[22] V. Cerf, "The internet today," http://www.youtube.com/watch?v=KeAIwLp9YmA, 2009, accessed: Jan. 11, 2015.

[23] B. Leiner, V. Cerf, D. Clark, R. Kahn, L. Kleinrock, D. Lynch, J. Postel, L. Roberts, and S. Wolff, "A brief history of the internet," http://www.isoc.org/internet/history/brief.shtml, 2011, accessed: Jan. 11, 2015.

[24] A. Leon-Garcia and I. Widjaja, *Communication Networks: Fundamental Concepts and Key Architectures*, 2nd ed. McGraw-Hill, 2003.

[25] N. Niebert, S. Baucke, I. El-Khayat, M. Johnsson, B. Ohlman, H. Abramowicz, K. Wuenstel, H. Woesner, J. Quittek, and L. Correia, "The way 4ward to the creation of a future internet," in *Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on*, Sep. 2008, pp. 1–5.

[26] L. Zhang, A. Afanasyev, J. Burke, V. Jacobson, k. claffy, P. Crowley, C. Papadopoulos, L. Wang, and B. Zhang, "Named data networking," *SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 3, pp. 66–73, Jul. 2014.

[27] G. Xylomenos, C. Ververidis, V. Siris, N. Fotiou, C. Tsilopoulos, X. Vasilakos, K. Katsaros, and G. Polyzos, "A survey of information-centric net-

working research," *Communications Surveys Tutorials, IEEE*, vol. 16, no. 2, pp. 1024–1049, Feb. 2014.

[28] Palo Alto Research Center, "PARC—a Xerox company," https://www.parc.com/, accessed: Jan. 11, 2015.

[29] Named Data Networking, "Named data networking consortium—current members," http://named-data.net/consortium-members/, Oct. 2014, accessed: Jan. 11, 2015.

[30] V. Jacobson and D. K. Smetters, "Securing network content," PARC, Tech. Rep., Oct. 2009.

[31] K. Fall, "A delay-tolerant network architecture for challenged internets," in *Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, ser. SIGCOMM '03. New York, NY, USA: ACM, 2003, pp. 27–34.

[32] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless Communications and Mobile Computing*, vol. 2, no. 5, pp. 483–502, 2002.

[33] *RFC 2501 Mobile Ad hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations*, Internet Engineering Task Force, Jan. 1999. [Online]. Available: http://tools.ietf.org/html/rfc2501

[34] M. Balazinska and P. Castro, "Characterizing mobility and network usage in a corporate wireless local-area network," in *Proceedings of the 1st international conference on Mobile systems, applications and services*, ser. MobiSys '03. New York, NY, USA: ACM, 2003, pp. 303–316.

[35] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campus-wide wireless network," in *Proceedings of the 10th annual interna-*

tional conference on Mobile computing and networking, ser. MobiCom '04. New York, NY, USA: ACM, 2004, pp. 187–201.

[36] W.-j. Hsu and A. Helmy, "On important aspects of modeling user associations in wireless lan traces," in *Proceedings of the 2nd International Workshop Wireless Network Measurement*, April 2006.

[37] M. McNett and G. M. Voelker, "Access and mobility of wireless pda users," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 9, pp. 40–55, April 2005.

[38] W.-j. Hsu, D. Dutta, and A. Helmy, "Mining behavioral groups in large wireless lans," in *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*, ser. MobiCom '07. New York, NY, USA: ACM, 2007, pp. 338–341.

[39] M. Kim and D. Kotz, "Periodic properties of user mobility and access-point popularity," *Personal Ubiquitous Comput.*, vol. 11, pp. 465–479, August 2007.

[40] N. Eagle and A. (Sandy) Pentland, "Reality mining: sensing complex social systems," *Personal Ubiquitous Comput.*, vol. 10, pp. 255–268, March 2006.

[41] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot, "Pocket switched networks and human mobility in conference environments," in *Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, ser. WDTN '05. New York, NY, USA: ACM, 2005, pp. 244–251.

[42] W.-j. Hsu and A. Helmy, "On nodal encounter patterns in wireless LAN traces," *IEEE Transactions on Mobile Computing*, vol. 9, pp. 1563–1577, 2010.

[43] J. Su, A. Chin, A. Popivanova, A. Goel, and E. de Lara, "User mobility for opportunistic ad-hoc networking," in *Proceedings of the Sixth IEEE Workshop on Mobile Computing Systems and Applications*. Washington, DC, USA: IEEE Computer Society, 2004, pp. 41–50.

[44] HTTP Archive, “Trends,” http://httparchive.org/trends.php?s=All&minlabel=Dec+15+2011&maxlabel=Jan+1+2015#perHttps, accessed: Jan. 11, 2015.

[45] N. Gotelli and G. R. Graves, *Null Models in Ecology.* Smithsonian Institution Press, 1996.

[46] J. Zittrain, *The Future of the Internet—And How to Stop It.* Yale University Press, 2009.

[47] Google, “Map maker,” http://www.google.com.au/mapmaker, 2015, accessed: Jan. 16, 2015.

[48] Raspberry Pi, “Raspberry pi,” http://www.raspberrypi.org/, 2015, accessed: Jan. 11, 2015.

[49] WiGLE, “Wireless Geographic Logging Engine (WiGLE),” https://wigle.net/, accessed: Jan. 11, 2015.

[50] B. Thomas, “liber80211,” https://github.com/brycethomas/liber80211, Oct. 2012, accessed: Jan. 11, 2015.

[51] ——, “liber80211,” https://play.google.com/store/apps/details?id=com.brycestrosoft.liber80211, Oct. 2012, accessed: Jan. 11, 2015.

[52] M. Kershaw, “Wi-Fi PCAP Capture,” https://www.kismetwireless.net/android-pcap/, Dec. 2012, accessed: Jan. 11, 2015.

[53] M. Cunche, M. Kaafar, and R. Boreli, “I know who you will meet this evening! linking wireless devices using wi-fi probe requests,” in *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2012 IEEE International Symposium on a*, June 2012, pp. 1–9.

[54] M. V. Barbera, A. Epasto, A. Mei, V. C. Perta, and J. Stefa, “Signals from the crowd: Uncovering social relationships through smartphone probes,” in

Proceedings of the 2013 Conference on Internet Measurement Conference, ser. IMC '13. New York, NY, USA: ACM, 2013, pp. 265–276.