Ad Hoc Networks xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Ad Hoc Networks

journal homepage: www.elsevier.com/locate/adhoc

Opportunistic content diffusion in mobile ad hoc networks

Q1 Bryce Thomas[a,*], Raja Jurdak[b], Ian Atkinson[c]

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

A R T I C L E I N F O

Article history:
Received 17 January 2015
Revised 20 February 2016
Accepted 29 February 2016
Available online xxx

Keywords:
Manets
Content-diffusion
Contact-networks

A B S T R A C T

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 improve 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 h. 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 content-diffusion in MANETs.

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 collocated 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.

Q2 * Correspondingx author. Tel.: +61 12064553104.
E-mail addresses: bryce.m.thomas@gmail.com (B. Thomas), raja.jurdak@csiro.au (R. Jurdak), ian.atkinson@jcu.edu.au (I. Atkinson).

http://dx.doi.org/10.1016/j.adhoc.2016.02.022
1570-8705/© 2016 Published by Elsevier B.V.

- 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.

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 generalize 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 is 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 from a large university campus is described in Section 4, along with its associated uncertainties in session timestamps. Our first set of simulations analyze *universal diffusion* on the empirical 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. We then focus on a realistic *application-specific* use-case for content diffusion in Section 7—diffusing electronic maps based on the LAN trace and on empirical usage statistics from a university navigation app. 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 ii) 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 substantially newer (in some instances over a decade),

Please cite this article as: B. Thomas et al., Opportunistic content diffusion in mobile ad hoc networks, Ad Hoc Networks (2016), http://dx.doi.org/10.1016/j.adhoc.2016.02.022

larger in size, and is collected with greater temporal and procedural consistency across sites.

A number of other studies [10–13] have characterized wireless network usage and user behavioral 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 modeling.

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] that 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 on a real-world mobility trace 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 $I_{d,p} = \{[j_{d,p,1}, k_{d,p,1}], \ldots, [j_{d,p,n}, k_{d,p,n}]\}$ 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:

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

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

Please cite this article as: B. Thomas et al., Opportunistic content diffusion in mobile ad hoc networks, Ad Hoc Networks (2016), http://dx.doi.org/10.1016/j.adhoc.2016.02.022

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

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 precisely capture 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.

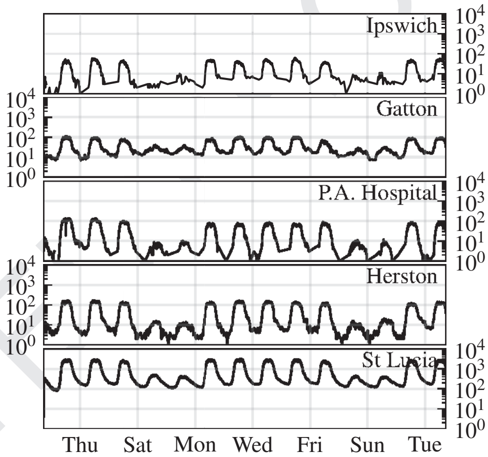

Fig. 1. Per-site session volume.

4. Uncertainties in trace-driven simulations

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 3081 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 Fig. 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 min 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.

4.2.1. Periodic timestamping

UQ deploys 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 $v \le u$ and $v' \le u'$.

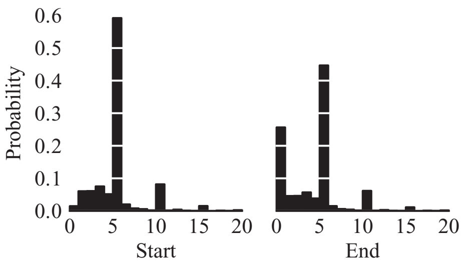

Fig. 2. Gap sizes (minutes) between chronologically consecutive start (left) and end (right) timestamps in the source trace. Timestamps with zero gap omitted.

Based on empirical observation, we add the further constraint that $u \le v'$, leaving us with $v \le u \le v' \le u'$. 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 min. Reporting is a low priority task competing with other tasks for computational resources and so some variation exists around the nominal 5-min interval. The nature of the 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. Fig. 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 min, with some variation. Gap sizes of 1 min or less are likely sessions being recorded as a part of a single report, while gap sizes between 1 and 5 min may either result from a single slow report or commencement of a new report. We note additional smaller peaks around 10 and 15 min 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 min.

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

- *pessimistic*: *subtract 10 min from reported session end times, leave reported session start times intact*. Subtracting 10 min 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 min 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 min 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.

Table 2
Optimistic and pessimistic session length adjustments.

	Source of uncertainty	
Adjustment	Periodic Timestamping	Connection Timeouts
Opt. start	−10 min	–
Opt. end	–	–
Pess. start	–	–
Pess. end	−10 min	−30 min iff session > 30 min

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-min window of inactivity. For those users who have timed out, we would like to subtract 30 min 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 min in duration, there is therefore no way to tell whether the user explicitly disconnected from the network or was subject to the 30-min 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 min in duration, subtract 30 min from the reported end time*. Subtracting 30 min from the end time of all sessions longer than 30 min ensures that the duration of any timed out session is not overestimated. The side effect is that any session longer than 30 minwhich 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 min.

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

5. Simulating universal diffusion

5.1. *Simulation overview*

Using our empirical traces, we perform multi-site, multi-source simulations for a variable number of source

Please cite this article as: B. Thomas et al., Opportunistic content diffusion in mobile ad hoc networks, Ad Hoc Networks (2016), http://dx.doi.org/10.1016/j.adhoc.2016.02.022

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

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 summarizes this procedure

Algorithm 1 Universal Diffusion Simulations.

```
1:  function RUN_UNIVERSAL()
2:      sites = {St. Lucia, Gatton, Herston,
            Ipswich, P.A. Hospital}
3:      times = {Wed 12:06PM Nov 28,
            Thu 04:52AM Nov 29,
            Sat 03:38PM Dec 01, Sun 04:56AM Dec
    02}
4:      sources = {1, 2, 4, 8, 16}
5:      for ∀⟨s, t, u⟩ ∈ {sites × times × sources} do
6:          SIMULATE(s, t, u)
7:      end for
8:  end function
9:
10: function SIMULATE(site, start, sourceCount)
11:     for i = 1 to 50 do
12:         sourceMACs = RANDSOURCES(site, start,
            sourceCount)
13:         SIMULATEDIFFUSION(site, start, sourceMACs)
14:     end for
15: end function
```

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 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:

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

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 simulation tracks the set A by simply maintaining a list of unique device IDs that appear since the start of the simulation. The number of source devices B is selected from the list {1, 2, 4, 8, 16}, with individual source devices varying for each simulation run (line 4 in Algorithm 1). Finally, the simulation maintains a flag for each device to indicate whether it has received the content. Source devices for a specific simulation run are considered to be in possession of the content for the duration of that run. For every simulation step, the size of set C is simply determined as the total number of devices in possession of the content.

As a result, 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.

6. Universal diffusion results

6.1. Results presentation overview

Throughout this section, we refer to Figs. 3–6 to illustrate our findings.

Fig. 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 Fig. 3 is to provide a coarse summary measure of diffusion performance—the

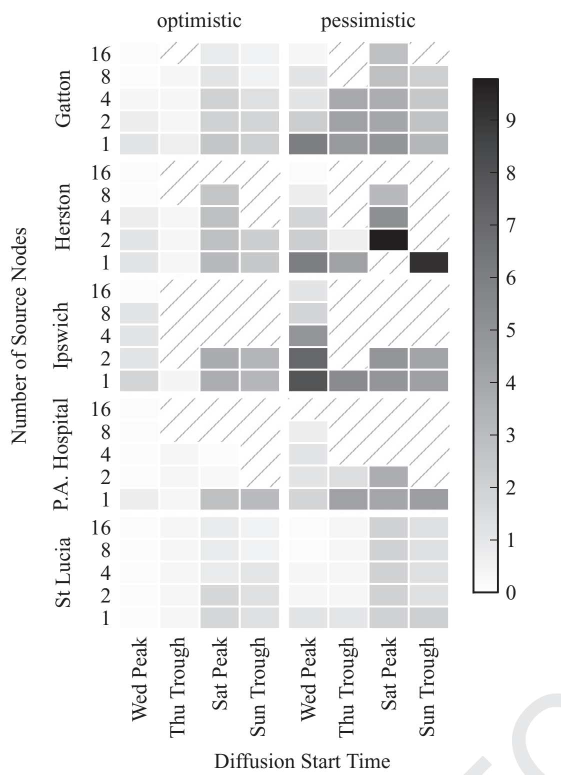

Fig. 3. Time taken to reach a 50% unreachable ratio (days). Striped squares indicate insufficient source devices were available to run the simulation, with the exception of Herston with 1 source device/pessimistic assumption/Saturday peak which simply never reached the 50% unreachable ratio.

time taken for diffused content to reach half of all devices in the simulated network.

Figs. 4 and 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 Figs. 4 and 5 are based on the previously defined pessimistic and optimistic assumptions respectively. Whereas Fig. 3 presents a coarse measure of diffusion (time to 50% unreachable ratio), Figs. 4 and 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 Figs. 4 and 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⟩.

Fig. 6 is designed to quantify the *variation* in diffusion performance across individual trials. That is, whereas Figs. 4 and 5 illustrate the overall expected level of diffusion potential, Fig. 6 highlights how some individual devices can be more effective at diffusing content than others. All results in Fig. 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 Figs. 4 and 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 behavior 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 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.

Please cite this article as: B. Thomas et al., Opportunistic content diffusion in mobile ad hoc networks, Ad Hoc Networks (2016), http://dx.doi.org/10.1016/j.adhoc.2016.02.022

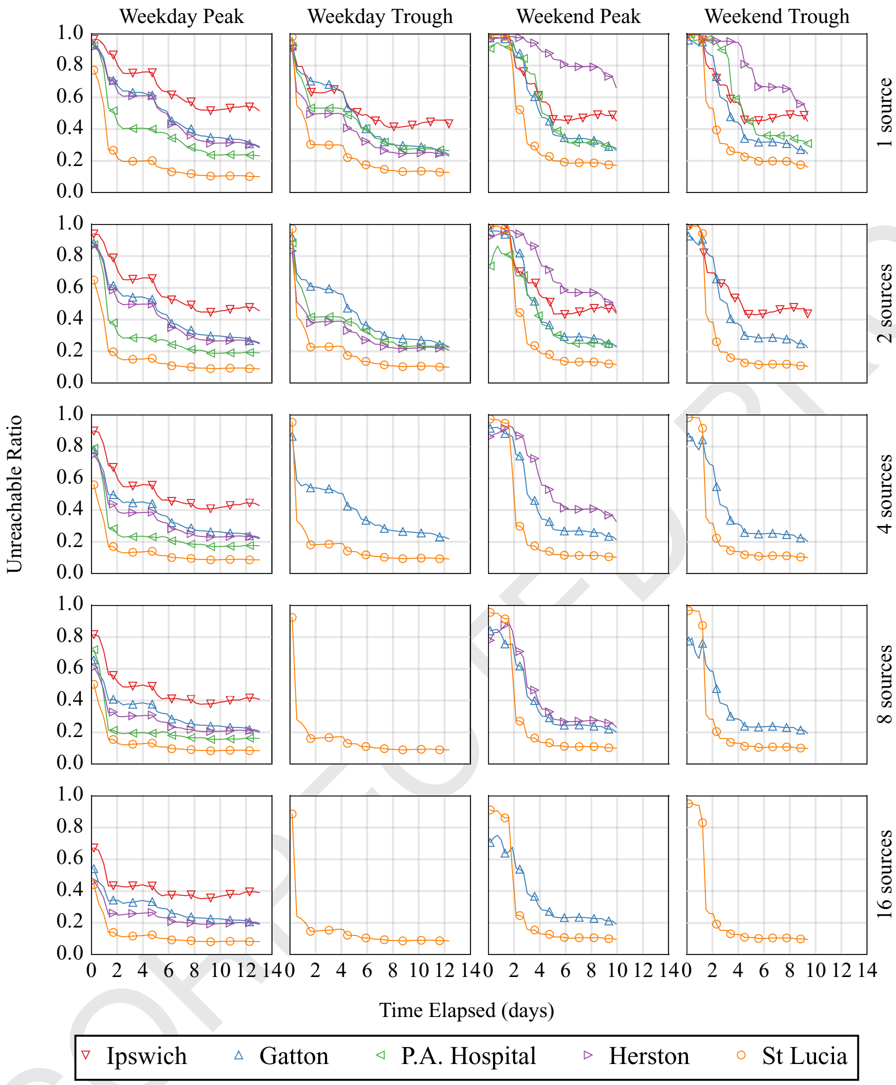

Fig. 4. Unreachable ratio based on number of source devices and diffusion start time (pessimistic).

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 Fig. 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 Figs. 4 and 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 Fig. 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 (Fig. 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 Figs. 4 and 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

Fig. 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.

Please cite this article as: B. Thomas et al., Opportunistic content diffusion in mobile ad hoc networks, Ad Hoc Networks (2016), http://dx.doi.org/10.1016/j.adhoc.2016.02.022

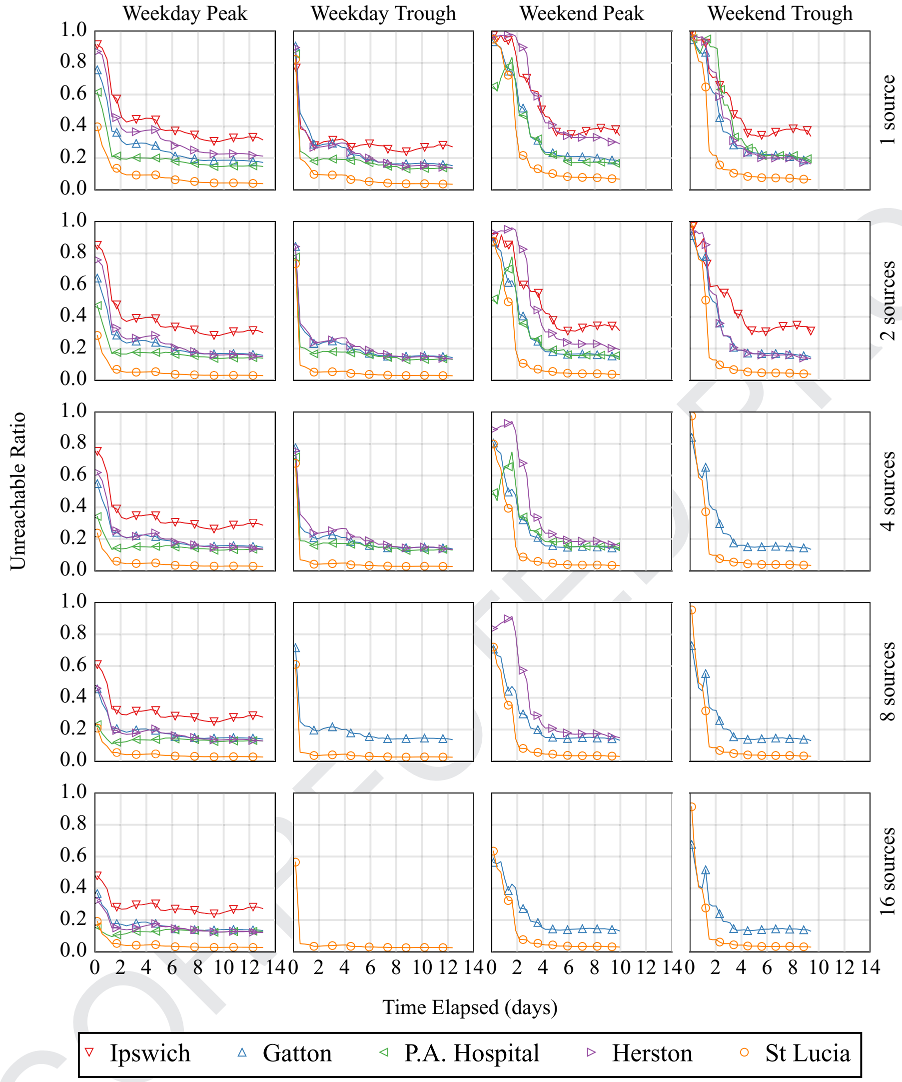

Fig. 5. Unreachable ratio based on number of source devices and diffusion start time (optimistic).

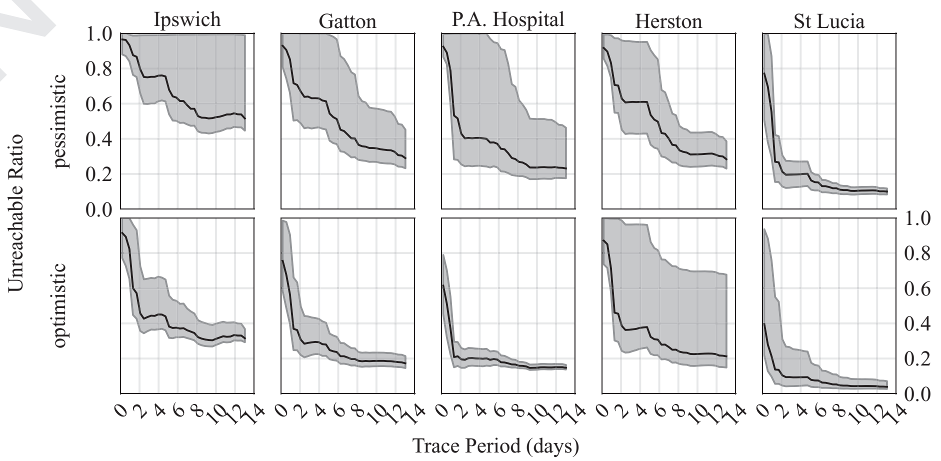

Fig. 6. Variation in unreachable ratio across random trials. Black lines depict the average. Shaded regions depict 5th–95th percentile around the average.

Please cite this article as: B. Thomas et al., Opportunistic content diffusion in mobile ad hoc networks, Ad Hoc Networks (2016), http://dx.doi.org/10.1016/j.adhoc.2016.02.022

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.

Overall, 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.

6.4. Summary of universal diffusion results

We observe that weekday starts to the process lead to faster diffusion, as do more content sources. We also find that for the largest site (St. Lucia) exhibits the fastest content diffusion rate, as expected. More interestingly, the diffusion rate is comparable for both optimistic and pessimistic assumptions in this larger site, suggesting that the large population size of the site dominates its diffusion rate regardless of session connection times. A similar trend is evident for the number of source devices for this large site that appear to have minor effect on diffusion potential, suggesting that the certain underlying correlations in space, time, and between nodes are governing the diffusion, rather than the number of copies of content in the network. Finally, for this larger site, we note the difference between weekday and weekend diffusion speed having higher and lower rates respectively. However, the time of day at which diffusion starts on a weekday or weekend does not appear to make a major difference to the diffusion speed.

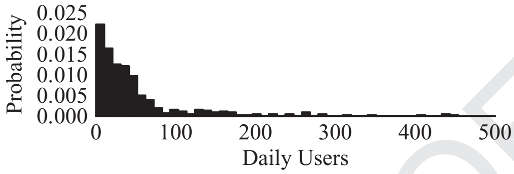

Fig. 7. Probability density—number of map users on any given day.

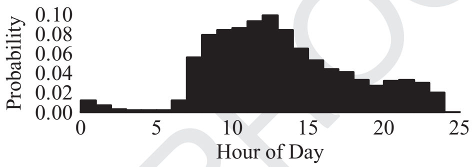

Fig. 8. Probability density—number of map users in any given hour of the day.

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 (Figs. 7 and 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.

7.1. The JCUNav trace

JCUNav [19] is a popular Apple iOS mobile campus navigation application at *James Cook* University (*JCU*, 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-h buckets. Fig. 7 illustrates the distribution of number of daily JCUNav users (Freedman-Diaconis binning [21]). Fig. 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, based on the statistics from Figs. 7 and 8. For each day we perform 50 simulation trials and average the results. Our map demand simulations are summarized in Algorithm 2 which is run over optimistic

Please cite this article as: B. Thomas et al., Opportunistic content diffusion in mobile ad hoc networks, Ad Hoc Networks (2016), http://dx.doi.org/10.1016/j.adhoc.2016.02.022

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

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 Fig. 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 Fig. 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.

Please cite this article as: B. Thomas et al., Opportunistic content diffusion in mobile ad hoc networks, Ad Hoc Networks (2016), http://dx.doi.org/10.1016/j.adhoc.2016.02.022

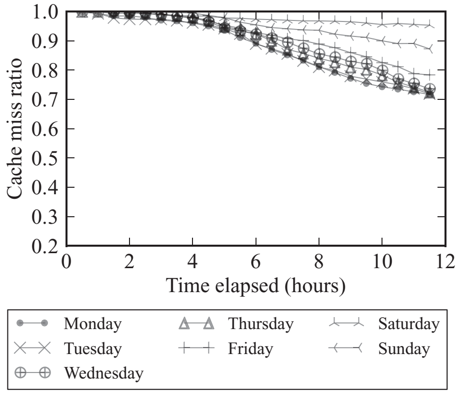

Fig. 9. JCUNav diffusion partitioned by day (pessimistic).

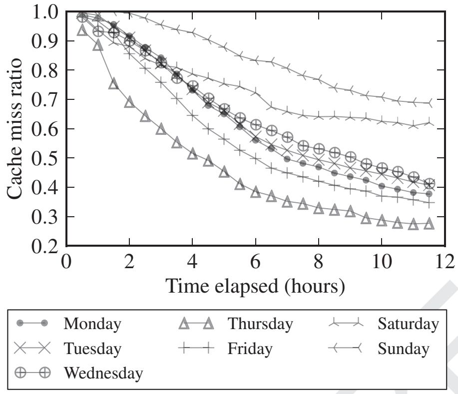

Fig. 10. JCUNav diffusion partitioned by day (optimistic).

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.

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

Figs. 9 and 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 optimistic simulations. For example, after 12 h 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 contrast, 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.

In summary, our results have highlighted that, for trace-based simulations of content diffusion in MANETS, uncertainties arising from the timestamps in traces can contribute up to 40% difference to the observed diffusion potential. This effect is higher for busier times, such as weekdays, where more encounters happen and therefore the cumulative effect of uncertainties results in a larger overall difference in diffusion rate.

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 instantaneous 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 min, 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 omission and commission—devices not simultaneously connected to an AP may actually be in transmission range and devices which are simultaneously

Please cite this article as: B. Thomas et al., Opportunistic content diffusion in mobile ad hoc networks, Ad Hoc Networks (2016), http://dx.doi.org/10.1016/j.adhoc.2016.02.022

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 [3,8,9] (e.g. devices handed out to graduate students) ii) the sample size is small (e.g. 10–300 devices) [3,8,9,22] iii) the instrumented devices are geographically sparse [22] iv) the trace is dated [3,8,9]. 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.

While our analysis has studied the impact of content demand and mobility context on diffusion dynamics, there remain practical considerations for implementing such a peer-to-peer content sharing architecture. Ensuring fairness and cooperative behavior among peer devices will be a primary requirement. Ultimately, users of electronic devices have their own applications running, and will need assurance that their participation in a peer-to-peer content sharing network will not quickly deplete their battery or slow down their device. Such objectives can be met through setting limits on the portion of battery energy or CPU time allocated for content relating. Most importantly, the users need to perceive the value of participating in content sharing, supported by mechanisms for ensuring that their peers are cooperating openly for content sharing. For instance, reputation-based mechanisms can provide users that share content more often higher priority for when these users demand content from the network. An interesting direction for future work is to design and test such mechanisms.

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 uncertainties 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 h. 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.

Acknowledgments

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

References

[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 Trans. Mob. Comput. 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, in: 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, in: 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, in: CoNEXT '09, ACM, New York, NY, USA, 2009, pp. 1–12.
[8] N. Eagle, A. (Sandy) Pentland, Reality mining: sensing complex social systems, Pers. 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, in: 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, in: 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, Comput. Commun. 34 (6) (2011) 704–714.
[15] P. Holme, J. Saramäki, Temporal networks, Phys. Rep. 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, Z. Wahrscheinlichkeitstheorie Verwandte Geb. 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 honors 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, where he leads the Distributed Sensing Systems Group. 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 energy-efficiency and mobility in networks. He has over 100 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.

Please cite this article as: B. Thomas et al., Opportunistic content diffusion in mobile ad hoc networks, Ad Hoc Networks (2016), http://dx.doi.org/10.1016/j.adhoc.2016.02.022