# A Message Ferrying Approach to Low-Cost Backhaul in Cellular Networks

Sayed Amir Hoseini<sup>1,2</sup>, Azade Fotouhi<sup>1,2</sup>, Mahbub Hassan<sup>1,2</sup>, Chun Tung Chou<sup>1</sup>,

Mostafa H. Ammar<sup>3<sup>7</sup></sup>

<sup>1</sup>University of New South Wales (UNSW), Sydney 2052, NSW, Australia <sup>2</sup>National ICT Australia (NICTA) <sup>3</sup>Georgia Institute of Technology, Atlanta, GA, USA

{amirhoseini, azadef, mahbub, ctchou}@cse.unsw.edu.au, ammar@cc.gatech.edu

## ABSTRACT

Cellular operators are considering large-scale small cell deployment in urban traffic hot spots to combat the looming capacity crisis in their networks. However, connecting a large number of small cells to the network core using current backhaul technology is costly. In this paper, we propose message ferrying as a low-cost backhaul solution that uses mobile phones of vehicle occupants as an army of ferries to transfer network data between small cell base stations and nearby switching centres. There are a number of challenges in the design and deployment of such a system. Use of phone memory for network data transfer without interfering with user applications is one such challenge that we study in this paper. Although the memory capacity in mobile phones has increased significantly over the years, it is considered a volatile ferrying resource as it is claimed and released dvnamically by user applications. If message ferrying is to be transparent to the user, base stations must be able to predict memory availability over a sufficiently long horizon to ensure that data carried in the phone do not get wiped out during transit due to dynamic memory claim by user applications. We experiment with real phones by logging memory availability for extended hours under different usage scenarios. By applying autoregressive models on memory usage traces, we are able to predict the minimum memory available over a 7.5-minute horizon with 94% accuracy.

#### **Categories and Subject Descriptors**

C.2.1 [Ccomputer-Communication Networks]: Network Architecture and Design—Store and forward networks, Network topology, Wireless communication; D.4.2 [Software]: Storage Management—Main memory

CHANTS'14, September 7, 2014, Maui, Hawaii, USA.

Copyright 2014 ACM 978-1-4503-3071-8/14/09 ...\$15.00.

http://dx.doi.org/10.1145/2645672.2645677.

#### **Keywords**

Message Ferrying; Delay Tolerant Networking; Cellular Backhaul; Small Cell; Memory Usage Prediction

#### 1. INTRODUCTION

Cellular operators are constantly looking for new ways to increase their network capacity. Although acquisition and more efficient use of additional spectrum would increase the capacity of existing macro cells, it is now widely believed [3] that the spectrum reuse through small cells deployed at specific locations (*hot spots*) where mobile traffic is concentrated is the most scalable way to meet the capacity crisis in mobile networks. However, connecting the outdoor small cells, also known as *metro* cells, to the core network with existing backhaul solutions, e.g., fibre, microwave, satellite, and leased lines, is expensive and can easily make up 80% of metro cell total cost of ownership [1]. The small cell evolution is inevitably shifting the mobile bottleneck from the radio access to the backhaul [3], demanding novel approaches to cellular backhaul.

In this paper, we propose to employ the well-known concept of message ferrying [16] as a low-cost data transportation option for small cell backhaul. In the proposed architecture, mobile devices of vehicle occupants act as an army of ferries to continuously transfer data back and forth between small cells and nearby switching centres, creating a virtual ferry-based backhaul for small cells (see Figure 1). The basic ingredients required to realise this approach are already in place. For example, current mobile phones are equipped with large memory capacity providing ample storage to ferry data. Typical small cells are deployed on street furniture, such as lamp posts, making it practical for the small base stations to easily reach the mobile devices inside passing vehicles despite having short communication ranges. Since mobile phones are designed to detect and connect to available base stations automatically and quickly, ferry discovery and association is not an issue. Finally, over the years, we have gathered considerable knowledge and experience with fundamental mechanisms of delay and disruption tolerant networks [11, 2, 4, 14], which serves as a strong foundation for developing the proposed ferry-based cellular backhaul.

Despite the feasibility, there are a number of challenges in the design and deployment of such a system. Because personal user data is carried in third party devices, privacy and security of such message ferrying would have to

<sup>&</sup>lt;sup>\*</sup>Mostafa Ammar's work is funded in part by NSF grant CNS 1161879.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions @acm.org.

be addressed satisfactorily. Advanced protocol mechanisms must be developed to ensure that ferry-based backhaul can meet the end-to-end quality of service requirements of cellular networks, especially for video traffic. Finally, if message ferrying using unused storage space in mobile phones is to be transparent to the user, base stations must be able to predict memory availability over a sufficiently long horizon to ensure that data carried in the phone do not get wiped out during transit due to dynamic memory claim by user applications. While all these challenges are real and must be addressed systematically, in this paper, we carry out a detailed study of the issue of phone memory predictability and leave the other issues as future work.

We experiment with real phones by logging memory availability for extended hours under different usage scenarios. By applying autoregressive models on memory usage traces, we are able to predict the minimum memory available over a 7.5-minute horizon with 94% accuracy. The memory availability question is going to be relevant to any application where message ferrying using mobile phones is proposed. As such the analysis in the paper is of wider applicability although it is done with the focus on the backhaul application.

The remaining of the paper is structured as follows. We present the proposed message ferrying architecture for small cell backhaul in Section 2. Details of our phone memory experiments and the results are provided, respectively, in Sections 3 and 4. We discuss related work in Section 5 followed by our conclusion in Section 6.

### 2. ARCHITECTURE

Small cells are deployed at traffic hot spots as well as next to every switching centre enabling a centre to act as a message ferrying hub for all nearby small cells (Figure 1). As soon as the mobile phones inside cars come within the wireless communication range of the small cells, they automatically connect with the cell base stations using existing cellular signalling. The base stations employ appropriate scheduling algorithms to upload or download data to and from the phone storage. For example, when a phone travels from hot spot to a switching centre, the hot spot base station downloads data items to be ferried into the phone storage, which is then uploaded to the base station at the switching centre completing the ferrying task. Thus, the continuous flow of mobile phones at both directions creates a bi-directional backhaul link for the cellular network. For a given road segment, the capacity of such ferry-based backhaul can be roughly estimated by multiplying the vehicular volume with the amount of available memory space that could be expected in a phone assuming sufficient local radio capacity exists in the small cells. Figure 2 shows daily traffic volumes in some of the Sydney roads. For these statistics, we find [5] that with only a 5% use of the phone memory, message ferrying could deliver giga bits per second capacity and transfer tens of peta bytes per week between two points on these roads.

While most data can be transported over the ferry-based backhaul, the operators may lease a low-speed low-cost backhaul link connecting the hot spot small cell to the switching centre directly to carry urgent interactive traffic not tolerant to the scale of delay characteristic of message ferrying. The leased link can also serve as a backup for unexpected disruptions in the message ferrying link or as a low-delay 'control



Figure 1: Proposed message ferrying architecture for realizing cellular backhaul.

channel' to improve the quality of service of applications supported by the higher-delay message ferrying link. The combination is particularly useful for delivering high quality video as a few initial video segments could be transported quickly over the leased link to avoid starting delay and the rest can be carried over the message ferrying link.

There are a number of challenges in the design and deployment of such a system:

- Local spectrum: We would require an abundance of short-range high capacity local spectrum within small cells to ensure that small cell radio capacity does not become a bottleneck for message ferrying. Early small cell deployments are using the same low frequency (below GHz) spectrum used for the macro cells, which is costly and offers limited data rates. Spectrum authorities, however, are planning to release significant new high frequency (above GHz) spectrum, which is currently underutilised, for small cell use at an exceptionally low licensing cost [12]. At the same time, there have been some recent breakthroughs that will soon allow mobile devices to transmit and receive data at an extraordinarily high rate for a short distance using frequency bands in the range of hundreds of GHz to terahertz [9]. Finally, there is a move to develop a new base station technology [12] that can efficiently use the existing unlicensed bands, such as the ones used by WiFi and radars, which are totally free of cost. These developments will help building higher capacity backhaul using message ferrying concept.
- Ferry trajectory and mobility: Unlike some previously studied applications of message ferrying [7], there is absolutely no control over the trajectory and mobility of the ferries. Cars can suddenly leave the road to the switching centre or get stuck in an unexpected traffic jam causing serious disruptions to end-to-end QoS.
- Privacy and security: Given that personal user data would be temporarily stored in a third-party device, privacy and security may be a concern in choosing the type of storage. Table 1 shows the storage capacity



Figure 2: Hourly traffic volume for (a) Paramatta Road Eastbound (Station 02.273.E), (b) Cleveland Street Eastbound (Station 02.038.E), and (c) Enmore Road Northbound (Station 02.062.N) collected during the week commencing 15 August 2005 (Source: Roads and Maritime Services, Sydney, Australia [13]).

available in smartphones from different manufacturers as of 2014. Although there is an enormous amount of *internal* and *removable* storage to boost the capacity of message ferrying, they are 'easily visible' to the user. We therefore consider RAM, referred to as 'memory' in this paper, as the most suitable storage type for the proposed architecture. Note that RAM provides a better obscurity compared to the other storage types, but encryption may be employed for ultimate security.

• Storage volatility: One major issue with memory is that it is claimed and released dynamically by user applications. If message ferrying is to be transparent to the user, base stations must be able to predict memory availability over a sufficiently long horizon to ensure that data carried in the phone do not get wiped out during transit due to dynamic memory claim by user applications. However, to the best of our knowledge, no memory prediction studies have been reported in the literature so far. In the following section, we describe our experiments to study the predictability of mobile phone memory.

Table 1: Smartphone storage capacity in GB.

Smartphone	RAM	Internal	Removable
Samsung S4 4G	2	16	64
Samsung Galaxy Note 3	3	32	64
iPhone 5S	1	64	None
HTC One 4G 801S	2	64	None
HTC Desire 600	1	8	64
Sony Xperia Z 4G	2	16	32
Sony Xperia Z 4G	2	16	64

#### 3. PREDICTION OF AVAILABLE MEMORY

We collected memory usage data on smartphones and used time series analysis to predict memory usage.

#### **3.1 Data Collection**

We used two Samsung Galaxy smartphones, which we will refer to as SP1 and SP2, for data collection. The phone SP1 (resp. SP2) is of model S3 (S4) with 1.8 GB (2 GB) RAM. Both of them use Android 4.3. Three traces (or time series) of memory usage were collected from these two phones using "MKSysMon" available from Google Play. One memory usage trace was collected from SP1 with a sampling period of 3s; the trace contains 9819 samples, which is more than 8 hours. The user of SP1 was asked to use the smartphone in the usual manner during data collection. The trace SP1 covers stand by mode and commonly used applications for messaging, calling, working with emails, watching video, web browsing and so on. For smartphone SP2, two traces were collected. The sampling period was 5s and each trace contains 1180 samples, which is just over 1.5 hours. The trace SP2-N (where N denotes normal) was collected while the user used the phone normally like the SP1. In order to understand the memory usage pattern for memory intensive applications, the trace SP2-H (where H denotes high) was collected as the user played the graphic intensive games, Asphalt8 and Fifa2014, while running the same apps as SP2-N in the background. A summary of the three traces is in Table 2.

Trace	SP1	SP2-N	SP2-H
Phone Model	S3	S4	S4
Total RAM	$1.8~\mathrm{GB}$	2  GB	2  GB
Sampling period	3s	5s	5s
# samples	9819	1180	1180

Table 2: Three memory usage traces collected.

#### 3.2 Methodology

For message ferrying, we would like to be able to predict the amount of memory that will be available in the data ferries within a certain time frame. Therefore, our aim is to investigate whether we can use past data on available memory to predict the amount of available memory in the future. While there are many time series models available for this purpose, we deploy a basic autoregressive (AR) model in this paper. Finding the best prediction model will be pursued in future works.

We consider each collected trace as a time series  $\{y_t\}$ where t = 1, ..., N where N is the number of samples in the time series and  $y_t$  denotes the amount of available memory at the *t*-th sampling instance. We use AR model of order *p* (where *p* is a parameter to be studied later on) for modelling. This model has the following form:

$$y_t = \mu + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \dots + \rho_p y_{t-p} + \varepsilon_t \tag{1}$$

where  $\mu$  and  $\rho_j$  are the coefficients of the AR model, and  $\varepsilon_t$  is zero mean Gaussian white noise. For the estimation of the AR model coefficients at sampling time n, we assume that we have all the data at or before time instance n. This is inevitably not the most efficient method for model estimation because all past data has to be stored. Since our aim is to study the feasibility of predicting memory usage, we will leave efficient estimation methods, such as recursive AR modelling, as future work. We use the ar function from the MATLAB (System Identification Toolbox) to estimate the AR model. Our intention is to use the identified AR model to predict the available memory in the time window [n+1, ..., n+h], i.e. a prediction horizon of h steps where h is a parameter to be studied later on. Let  $\hat{\mu}, \hat{\rho}_1, .., \hat{\rho}_q$  denote the estimated coefficient, then the predicted  $y_{n+1}$  (or 1-step ahead prediction) is given by:

$$\hat{y}_{n+1} = \hat{\mu} + \hat{\rho}_1 y_n + \hat{\rho}_2 y_{n-1} + \dots + \hat{\rho}_p y_{n-(p-1)}$$
(2)

The 2-step ahead prediction is :

$$\hat{y}_{n+2} = \hat{\mu} + \hat{\rho}_1 \hat{y}_{n+1} + \hat{\rho}_2 y_n + \dots + \hat{\rho}_p y_{n-(p-2)}$$
(3)

and so on. When h > p, the *h*-step prediction follows the recursive model:

$$\hat{y}_{n+h} = \hat{\mu} + \hat{\rho}_1 \hat{y}_{n+h-1} + \hat{\rho}_2 \hat{y}_{n+h-2} + \dots + \hat{\rho}_p \hat{y}_{n-(p-h)} \quad (4)$$

Such predictions can be computed by using the MATLAB **forecast** function. Since we need a certain amount of data to obtain an AR model, we can only use n in the range [q, N - h] where q is the minimum amount of data for estimating AR model; the upper limit of N - h is so that we can predict the full h steps ahead and then compare with collected data. The value of q should be at least 2p.

We consider a prediction horizon of up to 450 seconds (= 7.5 minutes). Therefore, for the trace SP1, h ranges from 1 to 150; while for SP2-N and SP2-H, h is between 1 and 90. We use the following metrics to measure the prediction performance. Note that  $\Delta = N - h - q + 1$ .

• Mean Absolute Error (MAE)

$$MAE = \frac{1}{\Delta} \sum_{n=q}^{N-h} MAE_n$$
 where  $MAE_n = \frac{1}{h} \sum_{k=1}^{h} |\hat{y}_{n+k} - y_{n+k}|$ 

• Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{1}{\Delta} \sum_{n=q}^{N-h} MAPE_n$$
 where  $MAPE_n = \frac{1}{h} \sum_{k=1}^{h} |\frac{\hat{y}_{n+k} - y_{n+k}}{y_{n+k}}|$ 

• Mean Percentage Error of Predicting Minimum Available Memory (MMPE)

$$MMPE = \frac{1}{\Delta} \sum_{n=q}^{N-h} \tilde{m}_n \text{ where}$$
$$\tilde{m}_n = \frac{|\min_{k=1,\dots,h} \hat{y}_{n+k} - \min_{k=1,\dots,h} y_{n+k}|}{\min_{k=1,\dots,h} y_{n+k}}$$



Figure 3: Available memory for SP1.



Figure 4: Available memory for SP2-N and SP2-H.

#### 4. **RESULTS**

Figure 3 plots the amount of available memory (in MB) for trace SP1, and Figure 4 shows the available memory for traces SP2-N and SP2-H. It can be seen that sharp rises and drops took place in all three traces. A rise occurred when a new application or services was started, and a drop occurred when some applications were closed or were sent to the background. Note that the amount of available memory for SP2-H is lower than that of SP2-N because memory intensive applications were ran during the collection of SP2-H.

Figure 5 shows the empirical cumulative distribution function (CDF) of available memory for the three traces. We see from Figures 3 and 4 that the amount of available memory never falls below 200 MB for SP1, 500 MB for SP2-N and 230 MB for SP2-H. Thus, the CDF is zero for values lower than these thresholds. The empirical CDF can therefore be used to estimate the 'storage failure probability' for a given choice of memory use for ferrying. For example for SP1, storage failure probability is negligible if 200 MB or less is used, but it can be as high as 50% for 250 MB. Figure 6 presents the CDFs normalised to the mean available memory observed for the entire trace. An important observation from Figure 6 is that there is a sharp cliff when the normalised available memory is around 0.8. If we use A (in MB) to denote the amount of available memory at this point, then there is a high probability that A MB or less is available but there is a very small probability that more than A MB is available.

To investigate whether the AR model can accurately predict future available memory, we use SP2-N and SP2-H and vary the prediction horizon h from 1 to 90, which means up



Figure 5: Empirical CDFs of available memory.



Figure 6: CDF normalised to mean available memory.

to 450s. We use four values of p: 5, 15, 25 and 29. Figures 8 and 7 shows the MAPE for, respectively, SP2-N and SP2-H. It is not surprising that lower MAPE is obtained with larger p and lower h. For a prediction horizon h of 90 and p = 29, the MAPE for SP2-N and SP2-H are, respectively, 1.2% ( $\approx$  7.7 MB) and 5.2% ( $\approx$  19.6 MB). Generally, the MAPE for SP2-H is higher than SP2-N for the same values of p and h. This can be explained by the higher volatility of SP2-H.

Figure 9 compares MMPE with MAPE. MMPE is prediction error of the minimum available memory and is more related to our target. However, they are approximately the same for SP2-N and MMPE is higher than MAPE for SP2-H because of its high volatility. It can be also concluded that in the worst case, SP2-H, the available memory has been predicted by 6% error or 94% accuracy.

AR model estimation is computationally intensive, so we want to investigate whether some simpler prediction methods can give similar level of performance. We use a simple predictor where we use the mean available memory in the past p sample instances as the predicted available memory for the next h steps. We will call this method mean predictor. For this investigation, we use the trace SP1, and choose p = 20 and vary h from 1 to 150, which means up to 450 seconds for SP1.

Figure 10 compares the MAE for the AR predictor and the mean predictor for SP1. It can be seen that the AR predictor outperforms the mean predictor for all values of h. Note that for this comparison, the MAE is calculated over the 'legitimate' section (i.e.  $n \in [q, N - h]$ ) of test data. We now



Figure 7: MAPE for SP2-N.



Figure 8: MAPE for SP2-H.

investigate what happens if we limit the test data to 9-64th minutes of the trace SP1 where the volatility of available memory is high. The results are shown in Figure 11. It shows that for  $h \leq 71$ , the AR predictor has lower MAE but for h > 71, the mean predictor is better. This suggests that the AR predictor, which assumes stationarity, cannot deal with high level of fluctuations. We will investigate the choice of prediction methods, taking into account both accuracy and complexity, in future work.

#### 5. RELATED WORK

Over the last decade, researchers have studied message ferrying or delay tolerant networking (DTN) for a diverse range



Figure 9: MAPE and MMPE for SP2-N and SP2-H



Figure 10: MAEs for AR and mean predictors for legitimate section of trace SP1.



Figure 11: MAEs for AR and mean predictors for 9-64th minutes of trace SP1.

of applications, including e-mail delivery to disconnected villages [11], restoring connectivity between partitioned mobile nodes [16, 7], collecting data from a sensor network [6], and providing roadside-to-roadside (r2r) [8] data transfer service using vehicles as mobile ferries. Among them, r2r is the closest to the proposed ferry-based cellular backhaul service considered in this paper. However, r2r message ferrying studied in [8] considered WiFi-based devices, whereas we consider cellular networks and mobile phones. As such, we do not encounter the issues related to long scanning and association times characteristic of WiFi networks. However, because we consider users' mobile phones as ferries to carry network data, we face new challenges such as privacy and transparency of phone memory usage by two different entities, the user and the network. In this paper, we studied the predictability of unused phone memory to analyse the feasibility of transparent deployment of the proposed ferrybased cellular backhaul. Although there are prior works on predicting application [15] or energy [10] usage in a mobile phone, to the best of our knowledge, prediction of minimum available memory over a given horizon has not been studied before.

#### CONCLUSION 6.

In this paper, we proposed the idea of using mobile phones of vehicle occupants as message ferries to realise low-cost backhaul service in cellular networks. We discussed the feasibility as well as key challenges facing the ultimate design

and deployment of such a service. Predictability of unused phone memory is identified as one of the issues if the proposed ferry-based backhaul is to remain transparent to the users. We experimented with real phones and found that standard autoregressive models can predict the minimum memory available over a 7.5-minute horizon with 94% accuracy using the time series of memory usage. Given that the memory availability is going to be relevant to any message ferrying application that uses mobile phones, the analysis in the paper has applicability beyond the proposed backhaul application. Finally, we acknowledge that understanding memory availability is necessary but not sufficient for the viability of the proposed message ferrying system. Investigations of other issues are the subject of future work.

- 7. **REFERENCES**[1] Alcatel-Lucent. Deploying metro cells means thinking differently. http://www.smallcellsamericas.com/files/ alu\_metro3\_info\_gfk\_v008\_p1.pdf, 2013. Accessed: 06 Mar, 2014.
- [2] A. Ali, M. Panda, T. Chahed, and E. Altman. Improving the transport performance in delay tolerant networks by random linear network coding and global acknowledgments. Ad Hoc Networks, 11(8), Nov 2013.
- Jeffrey G. Andrews. Seven ways that hetnets are a cellular [3] paradigm shift. IEEE Communications Magazine, 51(3):136-144, 2013.
- [4] Internet Research Task Force. Delay tolerant networking research group. http://dtnrg.org/. Accessed: 03 Mar, 2014.
- M. Hassan and CT. Chou. Could message ferrying be a [5]viable technology for small cell backhaul? Technical Report 201407, School of Computer Science and Engineering, UNSW, 2014.
- I. Jawhar, M. Ammar, S. Zhang, J. Wu, and N. Mohamed. [6] Ferry-based linear wireless sensor networks. In IEEE Globecom, Atlanta, USA, 2013.
- [7]C.H. Liu, Ting He, Kang-Won Lee, K.K. Leung, and A. Swami. Dynamic control of data ferries under partial observations. In WCNC, Sydney, Australia, 2010.
- [8] A. Mansy, M. Ammar, and E. Zegura. Reliable roadside-to-roadside data transfer using vehicular traffic. In MASS, 2007.
- [9] L. Newman, A new record for terahertz transmission, *IEEE* Spectrum, 50(12):15–16, December 2013.
- [10] E.A. Oliver and S. Keshav. An empirical approach to smartphone energy level prediction. In International conference on Ubiquitous computing, 2011.
- [11] Alex (Sandy) Pentland, Richard Fletcher, and Amir Hasson. DakNet: Rethinking connectivity in developing nations. COMPUTER, 37(1):78-83, January 2004.
- [12] Qualcomm. 1000x more spectrum especially for small cells. http://www.gualcomm.com/media/documents/ 1000x-more-spectrum-especially-small-cells November 2013. Accessed: 15 Feb, 2014.
- [13] Roads and Maritime. Traffic volume data. http://www.rms. nsw.gov.au/publicationsstatisticsforms/aadtdata/, 2005. Accessed: 25 Feb, 2014.
- [14]K. Scott and S. Burleigh. Bundle protocol specification, RFC5050, November 2007. Internet Engineering Task Force.
- [15] Chang Tan, Qi Liu, Enhong Chen, and Hui Xiong. Prediction for mobile application usage patterns. In Nokia MDC Workshop, volume 12, 2012.
- [16]W. Zhao, M. Ammar, and E. Zegura. A message ferrying approach for data delivery in sparse mobile ad hoc networks. In MOBIHOC, Tokyo, Japan, 2004.