Dynamic Mobile Cloud Computing: Ad Hoc and Opportunistic Job Sharing

Niroshinie Fernando¹, Seng W. Loke² and Wenny Rahayu³
Department of Computer Science and Computer Engineering, La Trobe University, Australia
{¹tfernando@students., ²S.Loke@, ³W.Rahayu@}@latrobe.edu.au

Abstract—Despite increasing usage of mobile computing, exploiting its full potential is difficult due to problems such as resource sparseness. In this paper, we explore the feasibility of a mobile cloud computing framework to use local resources to solve these problems. The framework aims to determine a priori the usefulness of sharing workload at runtime. The results of experiments conducted in Bluetooth transmission and an initial prototype are also presented. Furthermore, we discuss a preliminary analytical model to determine whether or not a speedup will be possible in offloading.

Keywords—mobile cloud computing; offloading

I. INTRODUCTION

Mobile computing allows the user to be locationally independent by providing a tool when and where it is needed. However, problems arise when trying to support mobility in computing devices: resource sparseness, hazardousness, finite energy source, and low connectivity [18]. Offloading mobile applications to remote clouds such as Amazon EC2 has been discussed [13] where services are provided from the resource rich cloud to the mobile device. But this type of offloading depend on the connection to the remote cloud, and the system fails in low connectivity scenarios [19]. A solution would be to exploit the capabilities of the mobile cloud where a mobile cloud is defined as a ‘cloud’ of local resources utilized to achieve a common goal in a distributed manner. In this paper, we refer to ‘mobile cloud’ in a different view than traditional cloud computing. Typically, most of the local cloud resources would be mobile (mobile phones, PDAs, laptops etc) and would be owned by different individuals. A ‘mobile cloud computing framework’ needs to be dynamic such that it can handle runtime resources and connectivity changes, proactive such that the costs are pre-estimated, opportunistic such that it exploits resources as they are encountered, and cost effective such that its task distribution is based on a cost model with benefits to all participating nodes. Furthermore, such a cloud will not be limited to high end mobile devices, but will be able to cater to low end devices as well. The future computing cloud that surrounds a user will not just be smartphones and PCs, but also computers in shoes, watches, jackets, furniture, cups etc, and such a collection of devices around the user will change as he moves from one environment to another, calling for greater opportunistic behaviour and ad hoc set up, with graceful failing.

Although the concept of offloading to local hardware has been explored in a number of frameworks, these do not satisfy all the requirements of a ‘mobile cloud’ as defined above. In this paper, we discuss the vision towards such a framework and explore the possibilities based on related work and some preliminary experiments. Challenges in such a framework include mobility, job partitioning, job distribution in the cloud, recognizing a potential cloud device, connectivity options, cost estimation and fault tolerance. Ideally, networking would be accomplished via a common existing short-range protocol such as Bluetooth/WiFi etc, rather than special protocols and we experimentally investigate this possibility in this paper.

II. MOTIVATION

Consider the case of Jane traveling on a train in a foreign country. She needs help planning her trip and wishes to ask from the many locals on the train. To solve the language problem, she starts up a speech recognition and synthesis application in her phone. However, such a program would require much computational power and drain her battery. Fortunately, Jane is able to use her mobile phone to initiate setting up a ‘virtual cloud’ of local computational resources comprised of the many mobile devices on board.

The aforementioned scenario is only one example demonstrating the need for a mobile cloud computing framework. In wearable computing, two major challenges are to reduce bulkiness and low battery power [1]. This could be solved by offloading the computational jobs to the local ‘mobile cloud’, while sensors and peripherals facilitate the pervasive experience. In augmented reality, it has been suggested using cloud resources [15] to solve similar problems. Other areas include but are not limited to resource demanding rich media applications, video editing, facial recognition and image search.

We propose comprising the local cloud with other mobile devices, enabling mobility without additional infrastructure. Considering the trends for smart phones, which shows they are getting more powerful each year, a local mobile cloud will be able to provide sufficient resources for intensive mobile apps. Also, since most mobile devices have sensing abilities, a cloud made up of mobile devices will be able to provide the users with context aware services.

In the coming sections, the device that has a job to be completed and initiates the sharing/offloading process will be
referred to as ‘master’ and the other devices doing a share of the job would be referred to as ‘slaves/clients’.

III. PROPOSED ARCHITECTURE

The main components of the framework can be given as Resource handler, Job handler and Cost handler as in Figure 1. The Resource Handler would be responsible for

![Image](image_url)

**Fig. 1.** Main components of a Mobile cloud framework

resource discovery, establishing connections, and exchanging meta data with clients. Meta data includes device details such as CPU, available battery, pricing etc. The Context Manager senses and records contextual information about clients, such as their location, movement, and acceleration which is used by the Resource Monitor to keep track on client nodes. Context information is important to register if new devices are coming in, or if current devices are moving away, resulting in disconnection and fault tolerance mechanisms. Cost handler would need to estimate costs and select suitable client devices based on the device meta data ,user’s specific requirements. A micropayment module should also be included to handle the monetary transactions between client devices and master device. Job handler would be in charge of dynamically partitioning the application, creating a job pool and managing a work distribution mechanism. The basic steps are as follows:

1) Resource discovery: The master device needs to perform a discovery procedure to search for potential clients within range.

2) Calculate costs: A comparison of potential clients, user priorities, requirements and constraints, is needed to estimate the cost of sharing. Clients can be selected based on this calculation.

3) Distribute jobs: Since client devices will most probably differ in their resources and capabilities, a mechanism to decide which job is assigned to whom is needed. Lessons could be drawn from [9].

4) Run the jobs: Once the jobs have been distributed, the clients would proceed to execute their job/s.

5) Collect the results: When the client devices finish their job/s, results are sent back to the master, and reassembled.

6) Cleanup: Once the job is finished, a method to clean up the client devices of data and/or code that were part of the job is needed to ensure data privacy and security.

7) Handle micropayments: Payments to the client nodes could be processed before or after the completion of the job.

IV. COST MODEL, PROTOTYPE IMPLEMENTATION AND EXPERIMENTATION

We conducted experiments on Bluetooth to determine its suitability for connectivity on a mobile cloud and the results of these tests were helpful in the construction of our prototype. Our reasons for conducting the preliminary tests on Bluetooth are due to its advantages such as low radiation, low energy cost[6], and wide spread availability as opposed to other protocols such as WiFi [3] and 3G. Also, according to its specifications, future versions will be faster up to 24Mbps and consume less energy. Therefore, when dealing with low end devices Bluetooth would be a better option, although not the only one. Ideally, the system should be able to switch to other protocols depending on the client capabilities and energy situation.

Two phones: Nokia X6 and Nokia 6500, and a PC were used in experiments. These three devices were used since they represent a range (low end mobile, high end mobile, resource rich PC) of devices. The PC used had Microsoft Windows 7 Enterprise with Intel(R) Core(TM) Duo CPU E7300 @ 2.66 GHz 2.67 GHz as processor, 2 GB RAM and Bluecove 2.1.0 on Winsock as Bluetooth driver. To implement Bluetooth communication for J2SE in the PC, we used the third party library BlueCove. According to Nokia’s device specifications[6], X6 runs Symbian OS v9.4 with 434 MHz while 6500 runs Nokia OS. Although the CPU details for Nokia 6500 are not available there, other sources such as Softpedia[4] gives it to be 170 MHz.

A. Cost Formulae for Work Distribution

As the primary phase in developing a framework for job sharing, a master-slave system consisting of mobile devices on a piconet was designed using Bluetooth and J2ME. The master node initiates contact and seeks other client nodes advertising ‘job sharing’ services. Once the master finds client nodes and forms a piconet, it can proceed with sharing/offloading.

1) Sharing jobs: The time for a master node to complete a job is given in equation 1 where \( T_m \) is the total time to complete the job in master and \( T_{pd} \) is the time to establish the piconet. \( T_d \) is the time to partition and distribute jobs to slaves, \( T_s \) is the time to complete the master’s job and \( T_{wcr} \) is the time the master spends waiting for the slaves. Since computing the master’s job and receiving the results are done in separate threads in parallel, the greater value of \( T_s \) and \( T_{wcr} \) is used. The pseudo code for the master is given in Listing 1.

\[
T_m = T_{pd} + T_d + \max(T_s, T_{wcr}) \tag{1}
\]

The time to complete a job in the slave is given in equation 2 where \( T_d \) is the total time to complete the job, \( T_{pd} \) is the time to connect to the master, \( T_{res} \) is the time to receive the job parameters from the master, \( T_{cp} \) is the time to complete

1https://www.bluetooth.org

2http://www.forum.nokia.com/Devices/Device specifications/X6-00/

\[
\begin{align*}
\text{start master} & \quad //\quad (1) \\
& \quad \text{do device \& service discovery} \\
& \quad \text{until slave services are found} \\
& \quad \text{int numOfDev\_s = numOfServices + 1} \\
& \quad \text{long[]} \text{ readWaitTimes} = \text{ new long[ numOfServices ];} \\
& \quad \text{IF services are found} \\
& \quad \quad \text{FOR each service discovered} \\
& \quad \quad \quad \text{Open a Connection} \\
& \quad \quad \quad \text{Open an outputStream} \\
& \quad \quad \text{END FOR} \\
& \quad \text{ELSE} \\
& \quad \quad \text{Exit app} \\
& \quad \quad \text{END IF} \\
& \quad \text{// (1)} \\
& \quad \text{FOR each slave} \\
& \quad \quad \text{generate \& distribute parameters} \\
& \quad \quad \text{END FOR} \\
& \quad \text{// (3)} \\
& \quad \text{Start thread ownWork for doing own job} \\
& \quad \text{Thread tRead[]} = \text{ new Thread[ numOfServices ];} \\
& \quad \text{FOREACH} \\
& \quad \quad \text{Open inputStream} \\
& \quad \quad \text{Wait for incoming transmissions from slave \_i} \\
& \quad \quad \quad \text{Read result \_i} \\
& \quad \quad \quad // (5) \quad \text{- T\_worker[i]} = (5) - (4) \\
& \quad \text{END FOREACH} \\
& \quad \text{Join all slave threads tRead} \\
& \quad \text{and thread doing own job with main thread} \\
& \quad // (6) \quad T_{\text{cn}} = (6) - (1) \\
& \quad // (7) \quad T_{\text{worker}} = \text{ add all values in readWaitTimes array} \\
& \quad // (8) \quad T_{\text{cn}} = (8) - (1) \\
\end{align*}
\]

Listing 1. Pseudo code for master

slave’s job and \( T_{w} \) is the time to send the completed results back to the master. The pseudo code for the offloaded slave is the same as for the job sharing slave as given in Listing 2.

\[
\begin{align*}
\text{start master} & \quad //\quad (1) \\
& \quad \text{do device \& service discovery until a slave is found} \\
& \quad \text{IF a service is found} \\
& \quad \quad \text{open connection} \\
& \quad \quad \text{open outputStream} \\
& \quad \quad \text{generate \& distribute parameters} \\
& \quad \quad \text{for slave} \\
& \quad \quad \quad // (2) \\
& \quad \quad \quad \text{T_{on}} = (3) - (2) \\
& \quad \quad \quad \text{open InputStream} \\
& \quad \quad \quad \text{read offloaded job results from slave} \\
& \quad \quad \quad \text{close inputStream} \\
& \quad \quad \quad \text{close connection to slave} \\
& \quad \quad \text{END IF} \\
& \quad \text{// (5)} \\
& \quad \text{T_{worker}} = (5) - (4) \quad \text{T_{masterOff} = (5) - (1)} \\
\end{align*}
\]

Listing 3. Pseudo code for master offloading code on slave

\[
T_{\text{masterOff}} = T_{\text{piconetOff}} + T_{\text{on}} + T_{\text{worker}} \quad (3)
\]

\[
T_{\text{off}} = T_{\text{cn}} + T_{\text{rcn}} + T_{\text{cp}} + T_{\text{w}} \quad (4)
\]

3) Analytical model: Offloading to n slaves: We now provide an analytical model generalizing offloading to \( n \) slaves, showing the conditions under which such offloading results in speedups. We assume that a job is uniformly partitioned such that all slaves get an equal sized job so that a \( \text{Job} = \sum_{i=1}^{n} J_{i} \), there is no dependency among each job, and that all \( n \) slaves in the piconet would be dedicated nodes such that their computational power would be predominantly available for the job. Since \( J_1 = J_2 = ... J_i = J_{i+1} ... = J_n \) for \( n \) slaves, \( \text{Job} = n J \). Therefore, if \( T_0 \) is the time to do job sequentially on master, and if the time it takes the master to run a job \( J \) sequentially is \( T_i \), then, \( T_0 = n T_i \) \( (5) \)

If the computational time for one slave \( j_{\text{slave}} \) is \( T_{C_i} \), then \( T_{C_i} \) would be the computational time to do job \( J_i \) to \( \infty \), \( T_{C_i} \in \text{ci} \), \( \forall i \), \( T_i = k_i T_{C_i} \) \( (6) \)

Where \( k_i \) is a constant depending on each slave. Therefore, from (5) and (6), we see that since \( T_i = \frac{J_i}{n} \), \( T_0 = k_i T_{C_i} \) and \( \frac{T_0}{T_i} = n k_i \). If \( n k_i = K_i \) where \( K_i \) is a constant, then \( \frac{T_0}{T_i} = K_i \), \( \forall i \), \( T_0 = k_i T_{C_i} \) \( (7) \)

If the Speedup is \( S \), and \( T_{\text{worker}} \) is time to wait for and receive all results from slaves, then

\[
S = \frac{T_0}{T_{\text{worker}}} = \frac{T_0}{T_{\text{worker}}} \quad (8)
\]

\[
S = \frac{T_m}{T_{\text{worker}}} \quad (9)
\]

For \( n \) devices and a fixed job size of \( J \), \( T_p \) and \( T_{\text{on}} \) would depend on the bluetooth stack/drivers of the device. Hence, for a given device, let us assume that \( T_p + T_{\text{on}} \) would be a constant \( \alpha \). Then, \( S = \frac{n k_i T_{C_i}}{\alpha + T_{\text{worker}}} \) \( (9) \)

If \( t_i \) is the time the \( i^{th} \) thread on master spends waiting and receiving from the \( i^{th} \) slave, then the total wait and receive time is equal to the maximum \( t_i \) value of the \( n \) threads.
This can be represented as: 

\[ T_{wkcr} = \max\{T_1, T_2, \ldots, T_n\} \]

(10) If the communication overhead is \( \delta_i \), \( \delta_i = T_{cr} + T_{sr} + \delta_i \), 
then, considering (13), in 

\[ \alpha \leq \frac{S}{\alpha + \max\{(T_{cr} + T_{sr} + \delta_i)\}} \]

(11)

Since the job size is fixed, \( T_{sr} \) solely depends on the device Bluetooth, i.e, the communication cost \( c_{comm} \). Since \( \delta_i \) is also related to communication overhead, \( T_{cr} + \delta_i \) is proportional to \( c_{comm} \). Therefore, if \( \beta_i = T_{sr} + \delta_i \) then from (IV-A3), \( S = \frac{\alpha \cdot \max\{(T_{cr} + T_{sr} + \delta_i)\}}{\alpha + \max\{(T_{cr} + T_{sr} + \delta_i)\}} \). 

Here, \( \alpha \) depends on the master device, \( K_i \) is the ratio of computational times between the master device and the \( i^{th} \) slave, \( T_{ci} \) is the computational time on \( i^{th} \) slave respectively. So, 

\[ \text{MasterComm}^\text{min} \leq \alpha \leq \text{MasterComm}^\text{max} \]

\[ \text{SlaveComm}^\text{min} \leq \beta_i \leq \text{SlaveComm}^\text{max} \]

\[ \text{SlaveComm}^\text{min} \leq \beta_i \leq \text{SlaveComm}^\text{max} \]

Therefore, the maximum value of \( \frac{\alpha}{T_{ci} \cdot \max\{\alpha\}} \) comes from 

\[ \frac{\text{MasterComm}^\text{max}}{\text{SlaveComm}^\text{max}} \text{ and } \max\{(T_{ci} + \beta_i)\} \]

is related to 

\[ \frac{\text{MasterComm}^\text{max} + \text{SlaveComm}^\text{max}}{\text{SlaveComm}^\text{min}} \]

Therefore, 

\[ K_i > \frac{\text{CPU}_{\text{slave}} \cdot \text{CPU}_{\text{master}}}{\text{CPU}_{\text{master}} + \max\{C_{p} + C_{m}\}} \]

(13)

Where \( C_{p} \) and \( C_{m} \) represent \( \text{SlaveComm} \) and \( \text{SlaveComm} \). From (6) and (13), \( \frac{T_0}{T_{sr}} = K_i \). Since 

\[ T_{sr} \]

ratio is dependent on the computational capabilities of the devices, if \( K_i \propto \frac{\text{CPU}_{\text{slave}}}{\text{CPU}_{\text{master}}} \)
then, considering (13), in order to get a speedup greater than one from offloading on 

\[ \frac{\text{CPU}_{\text{slave}}}{\text{CPU}_{\text{master}}} > \frac{\text{CPU}_{\text{slave}}}{\text{CPU}_{\text{master}}} + \max\{C_{p} + C_{m}\} \]

Based on this derivation, if the ratio between the CPU of the \( i^{th} \) slave and the master device is greater than the sum of the communication cost of master and the maximum sum of computation and communication costs out of slaves 1 to \( n \), there would definitely be a speedup. The right hand side of the inequality gives the maximum possible value and we have an assumption that the CPU clock rates of the \( i^{th} \) slave and master are inversely proportional to their calculation times.

An application of this relation with actual data is given in Section IV-C3.

B. Bluetooth transmission testing

The experiments on Bluetooth were focused on the effect of message length, distance, protocol and buffer size: ‘Buffer size’ refers to the size of a message buffer.

1) Message length and Buffer size: Messages of varying lengths were transmitted from Nokia X6 to 6500 for protocols RFCOMM and OBEX for buffer sizes from 256 to 900 B. The trends for all buffer sizes were almost linear, with 500 B giving the best time.

When comparing these two protocols, it could be seen that RFCOMM was faster.

2) Distance: Our tests showed that OBEX has more range than RFCOMM, and as data lengths increase, the maximum range of transmission decreases and so does the consistency.

3) Computational power and heterogeneity: A PC was used to send and receive data to and from Nokia X6 to investigate the effects of computational power and heterogeneity. Figure 2 shows the trend of time against data lengths up to 2 MB in mobile-to-mobile, mobile-to-PC, and PC-to-mobile. Surprisingly, PC to Nokia X6 has the lowest performance, while Nokia to PC gives the best time suggesting that although the PC is more powerful than X6, transmission also depends on the receiver’s capabilities. It could be that use of third party library to interface with the native Bluetooth stack of the PC (Bluecove) might be adding an overhead. According to the Bluecove speed tests on the web, in some cases mobile to PC is faster, while in other cases, the opposite is true.

C. Initial Experiments with Distributed Mandelbrot Set Generation

As a first step for a framework for job sharing, we implemented a prototype that partitions a predefined problem and distributes the job among mobile and stationary devices. It was developed using Java, and was executed on the same two mobile phones and the PC that was used in the Bluetooth tests. RFCOMM was used since it had proved to be faster than OBEX in previous experiments.

As a sample application, we implemented a distributed Mandelbrot set generation and compared the results with its sequential implementation on the initiating master device. Although it is not a typical candidate for a mobile application, Mandelbrot generation was selected to study the feasibility and issues in the preliminary tests due to its ease of job partitioning.

1) The Mandelbrot set: The Mandelbrot set is a set of points situated in the complex plane and the boundary of these points forms a fractal. Iterating the formula given in Equation (15) gives this set of points.

\[ Z_0 = 0, Z_{n+1} = Z_n^2 + C \]

(15)

Where \( C \) is a constant number on the Complex plane, and \( Z_n \) is the current point. As the number of iterations per point increases, the accuracy of the image increases.

2) Parallelizing the Mandelbrot set generation: For a given point, the Mandelbrot iteration feeds the result of the previous step into the next one. Hence, the simplest method of partitioning this algorithm is to assign each device a
continuous block of the image. An example of how an image of 300 X 350 pixels is split is given in Figure 3.

![Figure 3: Splitting up the Mandelbrot image among three devices](image)

3) Execution results: We tested the application for an image of pixel size 300 X 300 for varying number of iterations from 200 to 5000. The distributed version was tested on two to three devices and the results were compared with the performance of the monolithic version. Figures 4(a,b) shows the results of the Mandelbrot version using Nokia 6500 and X6 as the Master.

![Figure 4: Iterations vs Speedup](image)

Here, Speedup is defined by the following formula in equation (16). Speedup = \( \frac{T_1}{T_m} \)  

(16)

Where \( T_1 \) is the execution time of the sequential algorithm on master device, and \( T_m \) is the distributed version measured on master device. As can be seen from Figure 4(a), sharing/offloading with Nokia 6500 as master gives speedups from 1.29 up to 3.92. However, the results obtained with Nokia X6 as master show that the distributed version failed to produce a speedup except when it was offloaded and shared with the PC, since X6 is more powerful. When work is shared with a device that is substantially weaker (Nokia 6500), the weak device degrades the overall performance, since the total outcome cannot be given until all the devices complete their work.

Figure 5 shows the processing times in local execution and the shared version in, (A)Nokia 6500 as master and (B)Nokia X6 as master. X axis shows the results of executions for varying Mandelbrot iterations from 500, 1000, 2000 and 5000. Here, 1(A) represent the results of execution for 500 iterations in Mandelbrot generation in which Nokia 6500 was the master and X6 was the slave. Here, the computational time measured on 6500 when it shared the workload with the X6 was only 73% that of its monolithic execution. 1 (B) shows the results for 500 iterations when X6 was master and 6500 was slave. In this case, execution time on X6 was 84% that of when the Mandelbrot generation was performed on the device (X6) itself. Likewise for results represented in 2, 3, and 4. In all cases, though only (A) gave a speedup, it is apparent that the processing time has lessened, suggesting an energy saving [12], given the energy cost for Bluetooth communication is also taken into account.

![Figure 5: Processing times Local vs Shared execution](image)

With regards to Section IV-A3, experimental results above, we identify that the CPU ratio for phones X6 and 6500 as 2.553. When we apply average values from an instance of a series of experiments where the master was 6500 and job was offloaded to X6, the value of \( \frac{\text{Master} \times \text{Comm}_{\text{max}} + \text{Master} \times \text{Comm}_{\text{min}}}{\text{CPU}_{\text{X6}}} \) giving the following inequality:

\[
2.553 > 2.1472
\]

thereby satisfying the relation in (IV-A3).

V. RELATED WORK

Two common methods relating to offloading and/or sharing work from mobile devices are, partitioning applications and VM migration.


Enabling mobile clouds with context awareness has been discussed in [14] and [17]. The system proposed in [8] also uses context to identify common goals.

VI. SUMMARY AND FUTURE CHALLENGES

Based on our experiments, we surmise that Bluetooth is one of the viable options as the mode of transmission in a mobile cloud. However, our future experiments will not be confined to Bluetooth and would explore other protocols such as WiFi and 3G as well. Also, our experiments show that Bluetooth transmission time is linear against data length and distance and the buffer size plays a substantial role, and RFCOMM outperforms OBEX in terms of speed, but OBEX has more range and computational power alone is not the deciding factor in transmission speed.

4http://hadoop.apache.org
The results with prototype for distributed Mandelbrot set shows that merely distributing a task regardless of device capabilities will not always give a speedup. In case of equal job sharing, a speedup could only be achieved in cases where the slave node/s were of higher performance than master. Therefore, the device capabilities of the master and potential slaves need to be compared before a job is shared. Tests with offloading show that offloading gives better performance than sharing.

These factors point out the need for a cost model, and we aim to explore ‘work stealing’ [2] as a way of effective load balancing.

REFERENCES


