Understanding Directional Load Balancing using Per Call Measurement Data

Anu Vaidyanathan
Electrical and Computer Engg.
HIT Lab
Univ. of Canterbury
Christchurch
anuradha.vaidyanathan@canterbury.ac.nz

Wilfred Wong
Alcatel-Lucent
Wellington
wong@alcatel-lucent.com

Mark Billinghurst
HIT Lab
Univ. of Canterbury
Christchurch
mark.billinghurst@hitlabnz.org

Harsha Sirisena
Electrical and Computer Engg.
Univ. of Canterbury
Christchurch
harsha.sirisena@ canterbury.ac.nz

Abstract— **Several load balancing techniques have been proposed and studied for CDMA systems using a variety of simulation-based data-sets. In this study, we present actual call data captured at Wellington, New Zealand, from their leading telephone services provider, in order to understand how calls in the active-set are shared amongst cell-sites. Two groups of cell-sites are analysed to understand this phenomenon including (a) The busiest cell-sites, which experience the highest call volume and (b) The cell-sites with the poorest coverage or reference pilot signal strength. We obtain a day long trace on the date of an All Blacks Rugby game in Wellington, New Zealand and further delineate the rush-hour data on this day. Rush hour is defined as the hour at which the most calls are placed during the day, which co-incided with the time at which the Rugby game started. We observe that the busiest sites are located in downtown Wellington and calls in the active set migrate in the direction of caller traffic, which was moving towards the Westpac stadium, where the Rugby game was held. Calls are not picked up by the nearest neighbors (to the busiest sites) and are absorbed by the neighbors that are proximal to the stadium, suggesting directional load balancing. In the case of cell-sites with poor coverage, the load is distributed amongst the nearest neighbors, when the coverage of a particular cell-site reduces, over time.**

Index Terms—**Load balancing, CDMA Networks, Per Call Measurement Data**

I. Introduction

Per call measurement data (PCMD) is collected at three switches within the network of New Zealand's leading telephone services provider, in order to characterize user and traffic patterns [7] within three major cities. Understanding such characteristics as call-traffic, user-mobility, data-traffic, performance metrics for enhancing the network and points of interest on special occasions or holidays with PCMD, provides a powerful medium to both performance engineers on the service-provider's end, and the customers or end-users, by enabling a host of services including safety monitoring, crisis-intervention, crisis-prevention and a host of Location Based Services. Particularly, by mapping cell-phone activity, as users move across the city and providing a visualization interface, the performance engineers can avoid such things as bottlenecks and other ineffective load balancing problems during rush-hour [1]. Further, the proliferation and ubiquity of mobile phones make them powerful sensors for measuring the pulse of an urban centre and enabling other end-uses.

In this study, we extend the utility of PCMD data to understand the effects of directional load balancing by isolating the busiest cell-sites and comparing call traffic from certain distances against reference pilot signal strength or coverage values of those cell sites, in order to understand how load balancing is carried out for the calls in the active set within the immediate neighbours. In this study, data is collected from Wellington, New Zealand on the date of an All Blacks Rugby game,June7th,2008 [6]

The rest of this paper is organized as follows: section 2 details the related work in this area and section 3 talks about the methodology used for analysis and characterization. Section 4 presents the effects of load balancing by migration of calls in the active set to a different originating cell site on the phone call data for New Zealand and details the effects of load balancing as measured on the busiest cells within Wellington City on the date of an All Blacks game for the busiest traffic hour and the whole day. Section 5 outlines the future work and conclusions are included in section 6.

II. Related Work

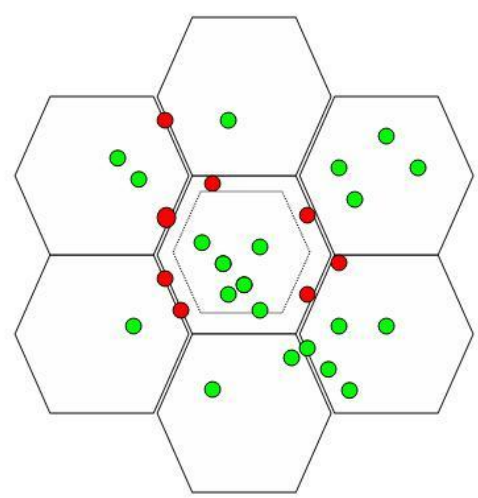


Fig. 1. Cells in the network under normal load where the GREEN calls belong to the respective cells and the RED calls are placed on the periphery of the cell.

Fig. 1 shows the normal load within seven cells in the network. The Green Calls belong to individual cells while the red calls are placed in a geographic area that is on the periphery of adjacent cells.

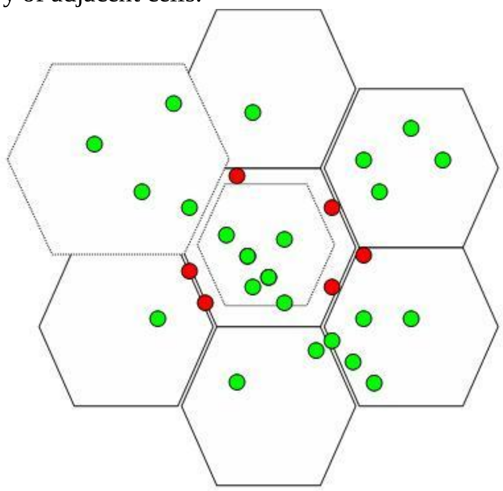


Fig. 2. The central cell shrinks due to an overload of calls and the cell on the top left-hand corner expands to make up for this, leading to a re-distribution of call ownership amongst cells.

Fig. 2 shows the central cell shrinking under a heavy load of calls. This causes the cell on the top left-hand corner to expand its coverage in order to maintain QoS guarantees and not drop the calls that belong to the central cell. Therefore, a call that was originally in the periphery of these cells is now owned by the cell at the top left-hand corner.

Mathematical Optimization models have been proposed in order to investigate load-balancing problems in CDMA systems [2] for non-uniform traffic distribution. Genetic algorithms have been proposed to solve the problem of traffic hot spots and unbalanced call distributions [3]. Adjustment of pilot power has been proposed [3] in order to assist in better load balancing. A comparison of load-balancing metrics has been presented in [5] wherein a novel vertical handoff algorithm is proposed, in order to optimise performance across WLAN and CDMA networks.

The data-sets in previous studies rely heavily on models and do not reflect actual calls made on a telecom services provider. In this paper, we utilize Per Call Measurement Data (PCMD) collected at various switches, within New Zealand from a leading telecommunications services provider which are records of actual calls placed on the network. We study load balancing as it actually occurs by profiling those cell sites that have the highest call-volume and those that have the poorest coverage or signal pilot strength, in order to compare and contrast directional effects.

III. METHODOLOGY

In this section, we present the methods employed to collect and analyse the data. These are important to acknowledge, as changing the analysis or collection methods can impact the nature of the results.

A. Data Collection

Traces were collected from a leading telephone service provider's CDMA network, where a Per Call Measurement Data (PCMD) feature is present. PCMD provides access to key network information for every 3G1x (voice, SMS or data) call that is placed via the network. The data recorded contains several aspects such as the identifier pertaining to individual phones, service type, number dialled, call length, signal quality, timing from pilots, sector in which the call was placed, the call result, the reason why the call ended etc. PCMD records provide an unprecedented, unobtrusive view of customer behaviour and end-user performance, collected live from ubiquitous and pervasive mobile sensors, cell phones. In order to obtain a geographical view of information pertaining to the network, a geo-location algorithm was used to extract location information from PCMD. CDMA has a unique timing and hand-off system, which make the data very fine-grained. By utilizing maximum likelihood methods in the geo-location algorithm, triangulation estimates are refined to the highest probability location. Field calibration has a per-call median error of ~140meters, averaged over all call locations wherein increased accuracy results from the user placing the call when closest to cell sites. The main data-set used in this study was collected from the Wellington on the date of an All-Blacks rugby game [6] to understand how traffic varies on important or event-oriented days of the year. Further, data was collected on a Friday and a Sunday, wherein the number of calls on a Friday is orders of magnitude larger than those placed on Sunday on the dates of March 27th, 2009 and March 29th, 2009, in Wellington.

B. Data Collection

Analysis of the traces was carried out by incorporating the data sets into the system shown in Fig. 3.

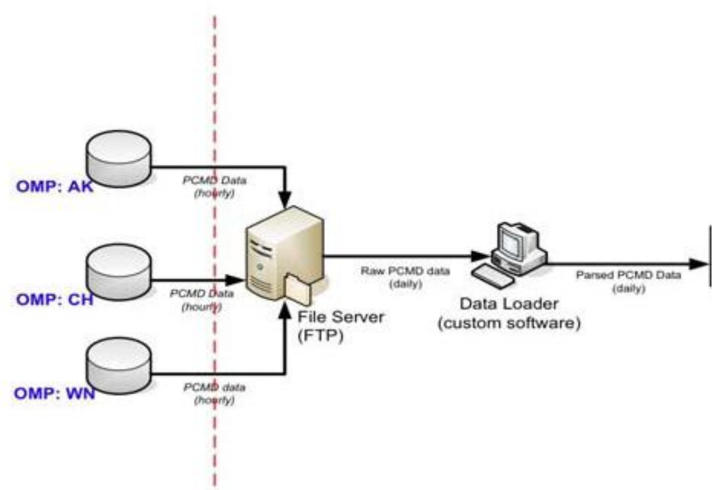


Fig. 3. Data being collected by the telecom services provider at three locations, Auckland, Christchurch and Wellington.

PCMD traces are collected from three switches in Auckland (AK), Christchurch (CH) and Wellington (WN), in order to analyse user patterns in these big urban centres. The data is accumulated at a file server and transferred there via FTP. This is then sent to a data-loader with our custom software, which parses the raw PCMD to scan for characteristics of interest. The data-loader contains means to load the data into custom databases, run queries on the data and means to interact with the data and feed it to a visualization interface. Standard maps of New Zealand in general and the three urban centres in particular are also stored in the data-loader, in order to map the location of calls and track them through the lifetime of the call. Fig. 3 shows how the data is collected from three different switches and sent to an FTP server, where it is temporarily stored before being passed onto a Data Loader and processed for later use.

TABLE I TRACES SETS USED IN THE STUDY (FROM WELLINGTON)

Trace	Date	Event	Sample Size
All day data	June 7th, 2008	All Blacks Rugby Game	1667914
Rush Hour	June 7th, 2008	All Blacks Rugby Game	127184
Friday Data	March 27th, 2009	Regular weekday	48632
Sunday Data	March 29th, 2009	Regular weekend	24676

Four traces (or datasets) are used in this study, including a rush hour trace and a whole day trace, collected on the day of the Rugby game, at Wellington. The data was collected in hour-long blocks and rush hour was identified as the hour in which the most calls were placed. Two additional traces were collected on March 27th (a Friday) and March 29th (a Sunday), 2009, in order to compare and contrast a cell-sites pilot strength with calls made at a particular distance from the cell. Table I summarizes the traces used in this study.

C. *Data Characteristics*

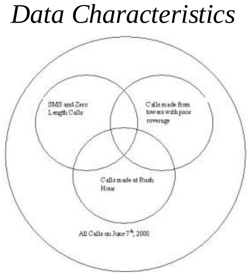


Fig. 4. Segregation of calls in the trace depending on the type of call being made.

The data used in this study was collected on June 7th, 2008 and segregated into three main categories of calls: calls made during rush hour, SMS and Zero length calls and calls made on cell towers with poor coverage, as measured by signal pilot strength. This is shown in Fig. 4.

Rush hour is defined as the hour in which the most calls are placed, during the day. As the trace contains information about the time at which the call was placed, the trace can be processed to obtain an hourly breakdown of calls.

The base station pilot signal's strength relative to the initial sequence offset is represented mathematically as Ec/Io where Ec (energy per chip) is the relative strength of the signal relevant to the communication and Io is the power of the entire transmission signal within the bandwidth. The unit of Ec/Io measurements is in decibels. Zero length and SMS calls are self-explanatory. Every record in the PCMD data-set pertains to one call and has several attributes associated with it. One of these attributes describes the kind of call placed from the phone and one other attribute is the length of the call and these are profiled to filter out SMS and zero-length calls. Table II details the size of the various data-sets used.

TABLE II EXPERIMENTS PERFORMED

Time Period	Zero Length Calls	SMS Calls	Low pilot strength or coverage
All day data for June 7th, 2008	No	No	Yes
	No	Yes	Yes
	No	Yes	No
	No	Yes (ONLY SMS)	No
	No	Yes (ONLY SMS)	Yes
	No	No	No
Rush Hour Data for June 7th, 2008	No	No	Yes
	No	Yes	Yes
	No	Yes	No
	No	Yes (ONLY SMS)	No
	No	Yes (ONLY SMS)	Yes
	No	No	No

IV. LOAD BALANCING EFFECTS

In this section we seek to characterize load balancing effects, as observed over the geographical area of Wellington, New Zealand. Cell sites are monitored based on a unique numeric identifier, associated with every cell site. The number of calls placed at a cell site is collected and stored for analysis. This information coupled with other metrics pertaining to the pilot strength of the reference pilot, and the distances at which calls in the active set are placed are taken into consideration, in order to characterize the data properly.

A. Cell Loading statistics based on call volume and distance

In this section, we discuss the cell loading statistics based on number of calls made at a cell site, the pilot strength of that particular site and the distance from which the calls in the active set are placed.

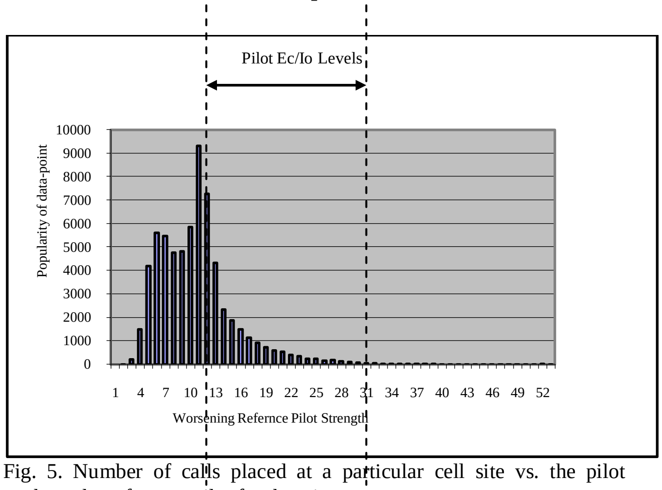


Fig. 5. Number of calls placed at a particular cell site vs. the pilot strength on the reference pilot for that site.

Fig. 5 shows the number of calls made across all cell sites in a data-set that excluded zero-length calls over the day-long trace. The x-axis shows the pilot strength (or normalized Ec/Io values) of the reference pilot as it worsens from left to right and the y-axis shows the number of calls placed. It is observed that on pilots with poor coverage, the total number of calls made drops off. This is because although a call may be placed at a particular site, if the coverage is too poor, the call fails. As the coverage of a site deteriorates, the number of calls made drops because the site is unable to allow the call to go through.

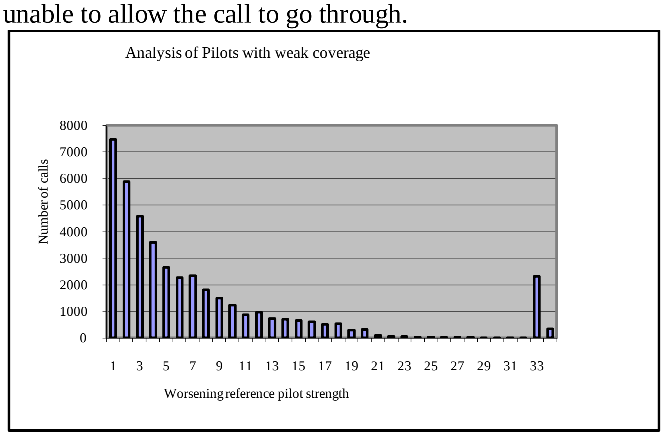


Fig. 6. Number of calls placed at a cell sites with poor coverage.

Fig. 6 places in focus the calls made on cell sites with poor coverage in a data-set that excluded zero-length calls over the day-long trace. The x-axis shows the pilot strength (or normalized Ec/Io values) of the reference pilot as it worsens from left to right and the y-axis shows the number of calls placed. It is seen in this figure that there are a few very cell sites with poor Ec/Io values, wherein the number of calls placed is still high. This could be explained as reflecting the calls that possibly failed to originate or those that terminated abnormally. There is an anomalously high number of calls at the right-hand side of the graph at a very poor coverage value and we point out that these are probably calls that terminated abnormally.

Since the data was collected on the date of an All Blacks rugby game, two candidate sets of cell sites can be scrutinized to understand how calls in the active set migrate amongst cells from which they originate and their immediate neighbors: (a) Monitoring cell towers in downtown Wellington, close to the stadium where the rugby game was being played and (b) Monitoring cell sites with extremely poor coverage. As downtown experiences a vast spike in call traffic, as the hour of the game approaches we are able to study the effects of how calls in the active set are shared and infer whether it is the closest neighbors that pick up the calls from congested sites or neighbors more proximal to where majority of the traffic is moving.

The distance from which calls are made is defined as the distance at which the caller is located from the base of the cell site and this is calculated according to equation 1, by using the round-trip delay time, which is calculated on a per-call basis.

```
RTD(chip): [CDMA_RT_DELAY]/8

Access Dist (km): [CDMA_RT_DELAY]/16*0.243
```

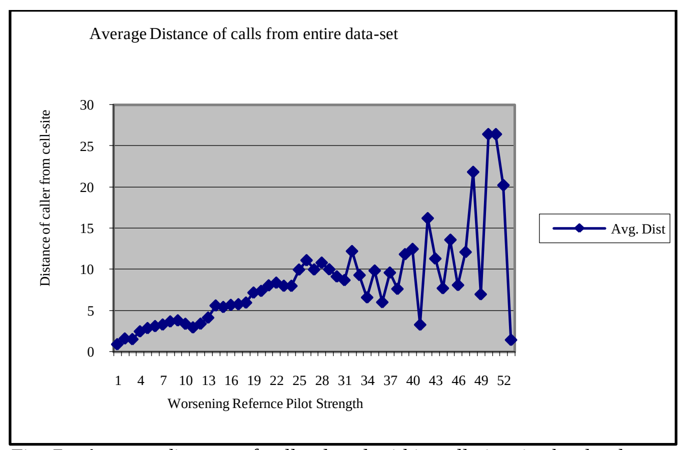


Fig. 7. Average distance of calls placed within cell sites in the day-long trace, for all cell-sites

Fig. 7 shows the average distance at which calls were placed on all cell towers in the day-long trace. The x-axis shows the pilot strength (or normalized Ec/Io values) and the y-axis shows the distance from which the calls are placed. We see that with poorer coverage, the average distance from which the calls are placed seems to increase.

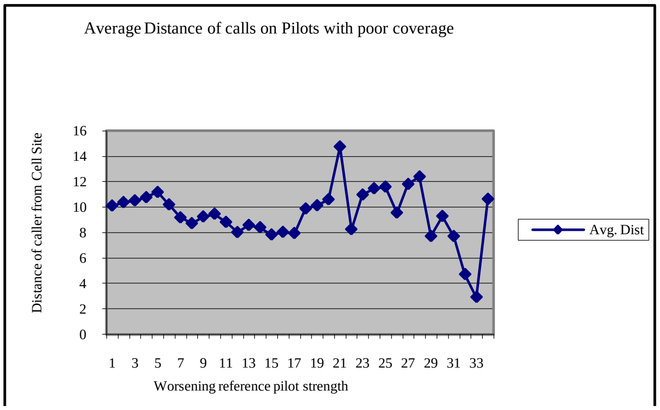


Fig. 8. Average distance of calls placed within cell sites whose pilot strength is worsening.

Fig. 8 focuses on those cell-sites with coverage worse than the threshold and shows the average distance at which calls were placed on cell towers, with increasingly poor Ec/Io values, in a data-set that excluded zero-length calls over the day-long trace. The x-axis shows the pilot strength (or normalized Ec/Io values) and the y-axis shows the average distance from which calls were placed. At the worst pilot strength values, the distance at which calls placed drops suddenly and this anomaly is once again explained as calls that failed to originate or were terminated abnormally.

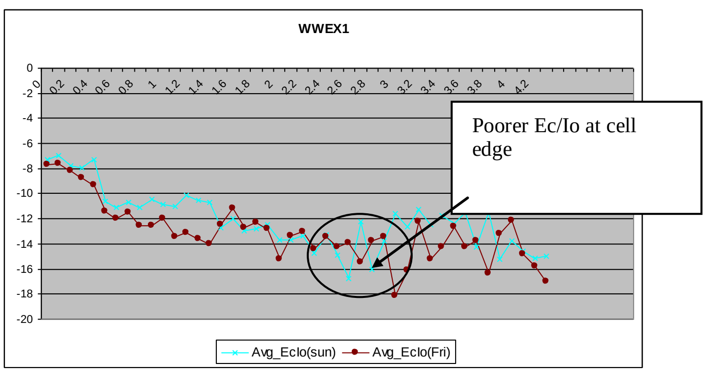


Fig. 9 Average distance of calls placed within one particular cell-site on a Friday and a Sunday to compare Ec/Io values at similar distances

Fig. 9 plots the average distance of the calls placed on a Friday (March 27th) as opposed to a Sunday (March 29th) when the call volume is about 50% that of the call volume on Friday (Table 1). The plot shows the Avg Ec/Io versus distance. We observe that the Ec/Io values beyond 3km appear to be weaker (when compared between Sunday and Friday). This could possibly due the users at the edge of the cells experiencing poorer Ec/Io due to the increased number of calls made on the cell. This would create a good candidate set of cells that can be analysed to understand how calls are shared.

B. Busy Cell Sites demonstrating load balancing

Table 3 shows the number of calls placed from the busiest cell sites, which have unique numerical identifiers associated with them. The ten busiest sites are collected in the all day trace and the rush hour trace in order to compare and contrast the load on the sites. These are later juxtaposed with the Ec/Io values on these cell sites in order to understand whether the cells drop calls or shed their load to their neighbors. Neighboring cells are often located at various radii from the originating cell and absorb calls as the coverage or signal strength of the originating cell site fluctuates.

TABLE III NUMBER OF CALLS PLACED FROM UNIQUE CELL SITES

Cell Descriptor	Cell-Site Number	Number of calls placed
All Day Data	219	36762
	348	36565
	222	35141
	99	34097
	66	33313
	100	32883
	168	32637
	186	32365
	40	31487
	12	30819
Rush-Hour	186	2768
	66	1916
	222	1820
	219	1787
	100	1738
	348	1701
	99	1634
	96	1618
	12	1571
	228	1564

As seen in this table, cell site 186 is one of the busiest sites during rush hour, which we know to be the time before the Rugby game, as the traffic congregates at downtown Wellington. Picking this cell site is a valid choice as even during the day, this cell site receives a large number of calls. The traces used to identify the busiest sites were compared across those in which the zero-length calls and the SMS data was included and excluded and 186 is established to be the busiest site.

Fig. 10 shows the relative placement of cell-site 186 and its immediate neighbors that have also been observed to handle a high volume of calls, in order to understand **load** balancing effects with calls in the active set. The distance of cell site 186 from its neighbors has also been indicated in the figure. The nearest neighbor is 208 at 0.7 KM followed by site 60 at 0.8KM, site 178 at 2.5KM, site 298 at 2.6KM and site 184 at 2.7K.

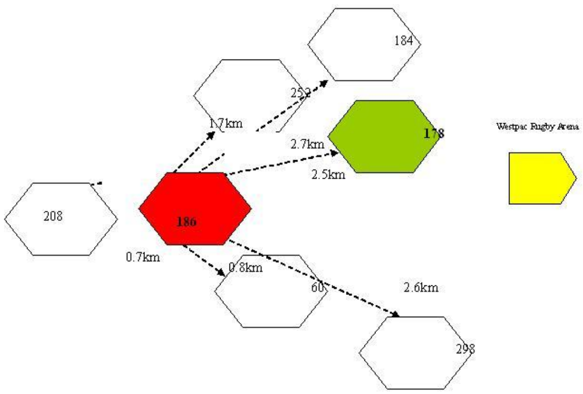


Fig. 10. Distance between cell-site 186 and its immediate neighbors.

We notice the relative distances between the busiest site and its immediate neighbors as we expect that when cell site 186 sheds its load (or shrinks) it will do so to its immediate neighbors.

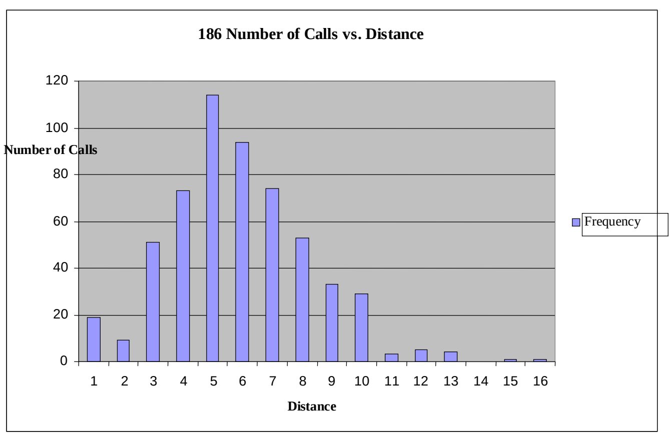
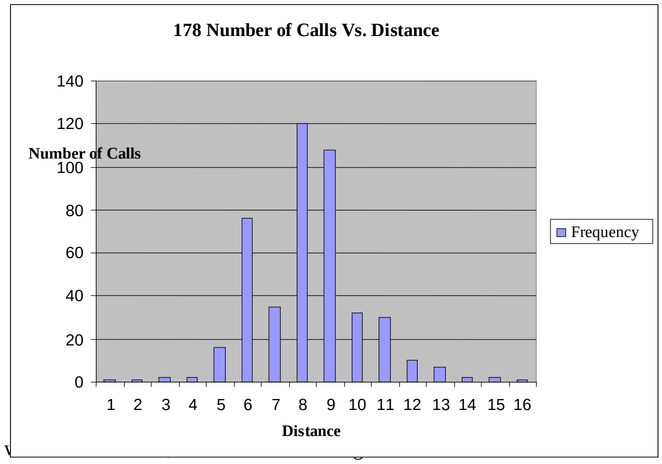


Fig. 11. Average number of calls placed vs. distance of calls placed within cell site 186, the busiest site during rush-hour.

Fig. 11 shows the number of calls made to cell site 186, vs. the distance from which the calls are placed. The x-axis shows the distance from cell-site 186 and the y-axis shows the number of calls that were placed from that distance. We see that the average distance from which calls are placed are 5km-6km. These calls are in the active set of the most loaded site in the trace as this is not an arbitrary cell site that absorbs calls from its periphery or from its neighbors. The calls placed in this site are genuinely originating from this site.

Fig. 12 shows the number of calls made to cell site 178, (which has the highest number of calls after site 186, compared to the rest of the neighbors and compared to site 252 which is the nearest neighbor by distance but is not as busy as site 178), vs. the distance from which the calls are placed.



The x-axis shows the distance from cell-site 178 and the y-axis shows the number of calls that were placed from that distance. We see that the average distance from which calls are placed are 8km-9km. This implies that site 178 is picking up the load from the periphery of site 186 in order to avoid calls being dropped.

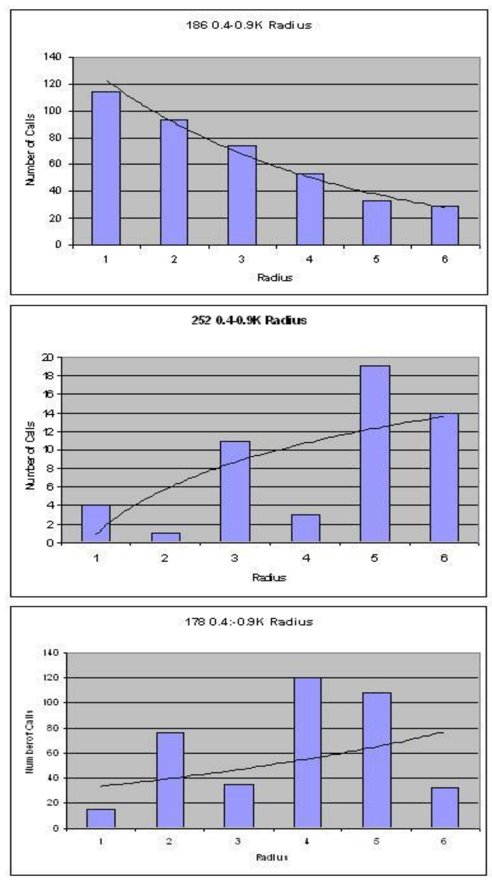


Fig. 13. Absolute and exponential rendition of the number of calls observed at a 0.4km-0.9km radius around cell site 186 and its immediate neighbors at rush hour.

Fig. 13 shows the exponential and absolute number of calls made from a radius of 0.4km-0.9km on the busiest cell site 186 and its two neighbours, 252 and 178. We see that 178, which is closer to Westpac Arena where the Rugby match is held picks up a larger portion of the load shed by cell site 186.

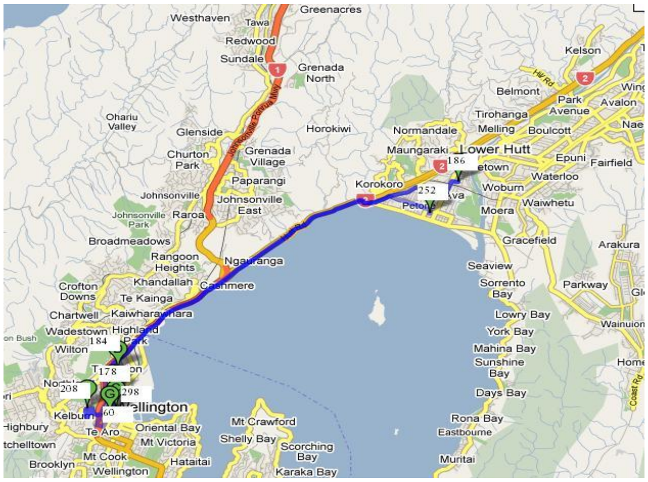


Fig. 14. Geographic location of the seven cell sites (186 and neighbors) on a map of Wellington

Fig. 14 shows the geographic location of the seven cell sites (186 and neighbors) on a map of Wellington. Table IV shows the location of the cell-site alongside the street address of the site and the distance of the neighbors from site 186. The Westpac arena, where the Rugby match was held on June 7th, 2008, whose rush hour call traffic is being analyzed in this section, is located at Waterloo Quay in Wellington. This location is 2.5km from cell-site 186 (Wellington Central), 0.4km from cell-site 178 (Lambton Quay) and 0.8km from cell-site 252 (Bolton Quay). The proximity of the cell-site to the stadium explains why cell site 178 picks up more of the load shed from 186 (even though site 252 is CLOSER to site 186, when compared to 178). As the traffic moves towards the stadium, calls get picked up by the sites en-route.

TABLE IV STREET LOCATION OF CELL-SITE 186 AND NEIGHBORS

Cell-Site Number	**Geographic Location**	**Distance from 186 in km**
186	Central Wellington	0
252	Bolton Quay	1.7
178	Lambton Quay	2.5
298	Victoria Villeston	2.6
184	Thorndon	2.7
208	Kelburn	0.7
60	Majestic Center	0.8

C. . Cell Sites with poor-coverage demonstrating load balancing

In this section, we briefly analyze the load balancing amongst the cells with the poorest coverage (or worst Ec/Io values) in our day-long trace collected at the Wellington switch. Since this switch collects data for a broader geographic area than just Wellington, the cells with poor coverage was identified to be cell-site 96, Blagdon Hill, whose six immediate neighbors were cell sites 99, 93, 261, 343, 262 and 260. The location of the site with the poorest coverage and its immediate neighbors and distance form site 96 is presented in Table IV.

TABLE IV STREET LOCATION OF CELL-SITE 96 AND NEIGHBORS

Cell-Site Number	**Geographic Location**	**Distance from 186 in km**
96	Blagdon Hill	0
99	New Plymouth City	3.7
93	Brooklands	5.7
261	Fitzroy	6.9
343	Mangorei	9.1
262	Bell Block	11
200	Oakura	13

The absolute distances are seen to be a lot farther amongst neighbours in this area as this is not a bustling city. Fig. 15 shows the location of the three neighbours, with respect to each other.

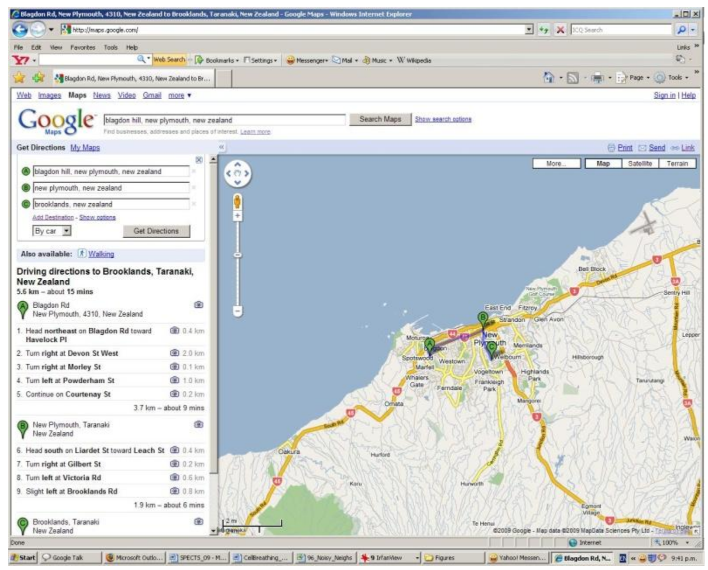


Fig. 15. Location of the tower with poorest coverage and it's two immediate neighbors.

Fig. 16 shows how the Ec/Io levels vary between site 96 (Blagdon hill) and its immediate neighbors 93 (Brooklands) and 99 (New Plymouth City). As seen in this figure, the closest cell site absorbs most of the load when tower 96 starts to drop its calls. The absorption of calls by the neighbors of cell site 96 does not have a directional component to it, as was the case with the busiest site 186 where the calls were being absorbed in the direction of movement of caller traffic. As these traces were captured with time-stamps on each of these calls, the x-axis shows the time of the call on each site and the y-axis shows the number of calls made at that instant in time, on that particular site for the three sites.

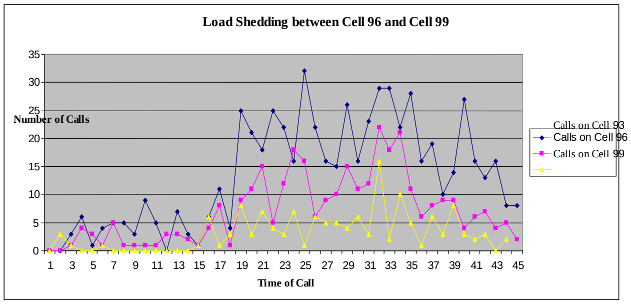


Fig. 16. Load balancing amongst cell sites with poor coverage.

V. CONCLUSION

In this study, we present Per Call Measurement Data in two traces collected on June 7th, 2008 at the Wellington switch, which co-incided with the date on which the All Blacks played an important Rugby match in Wellington New Zealand. The data relates to information about calls placed on the country's leading telephone networks provider and is collected over the course of 24 hours with rush-hour being identified as the hour when the maximum number of calls are placed on the network. We present data pertaining to the general aspects of the number of calls placed on various cell sites during the course of the day. We further present data pertaining to the general aspects of average caller distance from the cell sites, during the course of the day. We then go on to make two hypotheses about the nature of call traffic and load balancing thereof: Load balancing can be observed on two candidate sets of cell sites (a) Busy cell-sites that have too many calls and end up shedding calls in the active set such that the originating site changes, during the call (b) Cell-sites with poor coverage that are also shedding calls.

In the case of the busiest cell sites, (that are identified to be located in downtown Wellington), observed at rush-hour, wherein the load balancing effects are studied and presented in terms of number of calls made on the busiest site and its immediate neighbors, the radius from the sites wherein the call traffic is observed in order to understand which of the neighbors are absorbing the call traffic and the proximity of the sites shedding and absorbing call traffic to the Westpac Stadium, which is where the Rugby game took place. We observe that load balancing is directional, in that the cell-site that is closer to the stadium absorbed more calls from the busiest site, rather than the cell site that was the nearest neighbor to the busiest site.

In the case of the sites with poor coverage, we observe that the coverage levels of the neighboring sites complement each other as the trace is studied alongside the time-stamp on the call that was placed on the sites in question. This suggests that when a cell-site has reduced coverage, in order to handle its calls, in the same temporal proximity, its immediate neighbors absorb calls as indicated by complementary Ec/Io levels.

Future work would include characterizing the movement of users in the area of the noisy cell sites and further understand the impact of geographical location, interference and terrain on call volume and load balancing in the underlying network.

ACKNOWLEDGMENT

We would like to acknowledge the work of the team at Alcatel-Lucent including Vinu, Aaron and Andrew McKay.

REFERENCES

[1]AnuVaidyanathan, Mark Billinghurst, NZCSRC"Characterizing user-mobility in wireless telephonic networks", April 2008

[2] Kuo-Chung Chu, Frank Yeong-Sung Lin"Adaptive Load Balancing of Cellular CDMA Systems Considering Non-uniform Traffic Distributions ", Lecture Notes in Computer Science, 2005.

[3]H. Zhu, T. Buot, R. Nagaike, and S. Harmen, "Load Balancing in WCDMA systems by adjusting pilot power," Wireless Personal Multimedia Communications. The 5th International Symposium, Vol. 3, pp. 936 – 940, 27-30 Oct. 2002.

[4] Chae Y. Lee and Hyon G. Kang, "A Dynamic Load Balancing in Wireless CDMA Cellular Networks", White Paper, Korea Advanced Institute of Science and Technology.

[5] S. K. Das, S. K. Sen, and R. Jayaram, "Handoff Triggering and Network Selection Algorithms for Load-Balancing Handoff in CDMA-WLAN Integrated Networks", Proceedings of EURASIP Journal on Wireless Communications and Networking Volume 2008 (2008), Article ID 136939, 14 pages

[6]http://www.allblacks.co.nz/index.cfm?layout=2008results