CARDAP: A Scalable Energy-Efficient Context Aware Distributed Mobile Data Analytics Platform for the Fog

Prem Prakash Jayaraman[1], João Bártolo Gomes[2], Hai Long Nguyen[2], Zahraa Said Abdallah[1], Shonali Krishnaswamy[2], and Arkady Zaslavsky[1,⋆]

[1] CSIRO Computational Informatics, Canberra, Australia
{prem.jayaraman,arkady.zaslavsky}@csiro.au
[2] Institute for Infocomm Research (I2R), A*STAR, Singapore
{bartologjp,nguyenhl,spkrishna}@i2r.a-star.edu.sg
[3] Monash University, Melbourne, Australia
zahraa.said.abdallah@monash.edu

Abstract. Distributed online data analytics has attracted significant research interest in recent years with the advent of Fog and Cloud computing. The popularity of novel distributed applications such as crowdsourcing and crowdsensing have fostered the need for scalable energy-efficient platforms that can enable distributed data analytics. In this paper, we propose CARDAP, a (C)ontext (A)ware (R)eal-time (D)ata (A)nalytics (P)latform. CARDAP is a generic, flexible and extensible, component-based platform that can be deployed in complex distributed mobile analytics applications e.g. sensing activity of citizens in smart cities. CARDAP incorporates a number of energy efficient data delivery strategies using real-time mobile data stream mining for data reduction and thus less data transmission. Extensive experimental evaluations indicate the CARDAP platform can deliver significant benefits in energy efficiency over naive approaches. Lessons learnt and future work conclude the paper.

1 Introduction

Big Data analytics in the *Fog* has created tremendous opportunities to gain new and exciting value from big data. *Fog computing* or briefly *Fog* is a term recently embraced by Cisco Systems [4]. Fog computing extends Cloud Computing paradigm to the edge of the network. A recent vision paper from Intel [1], clearly highlights the value of the data that resides in the Fog and the need for distributed data analytics and techniques that can work in the Fog (closer to the source of data), where some of the *biggest* big data is generated. Further, this approach is also considered as an alternative to alleviate the current big data challenge of processing massive amounts of data in remote Cloud environments.

Fig. 1 presents the big picture of big data in the Fog. The intelligent systems and sensors (Internet of Things), smart city infrastructure, mobile smart phones

⋆ Prof. Zaslavsky is a visiting Professor at StPetersburg National Research University of IT, Mechanics and Optics, Russia.

Y. Manolopoulos et al. (Eds.): ADBIS 2014, LNCS 8716, pp. 192–206, 2014.

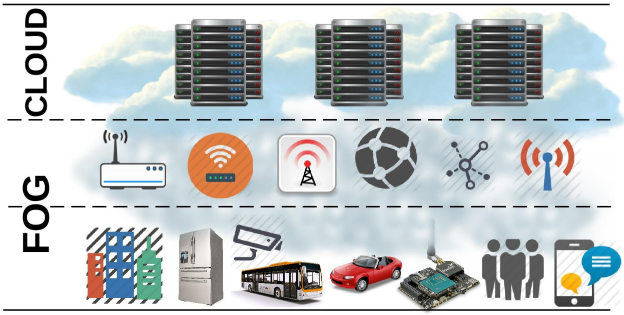


Fig. 1. Smart Devices and Human Entities in the Fog stream data to Cloud

etc., coupled with information from humans are some of the largest volume of streaming complex big data. The data sources as shown in the figure are heterogeneous and distributed. The ubiquitous connectivity available in the Fog (e.g. 2G, 3G, 4G, Internet, Network, WiFi etc) has created new opportunities that can take advantage of these massive data storehouses.

One of the key arguments driving the notion of Fog computing is the emerging wave of Internet deployments, most notably the Internet of Things (IoT) [4]. Although the definition of "Things" has evolved with rapid advancements in technologies, the key notion of enabling "Things" to sense without human intervention has remained the same. IoT is a major generator of live sensor data. The International Data Corporation estimates an installed based of 220 billion "things" by 2020 [2] coupled by the ever increasing growth in mobile smart phone devices. The smart mobile devices can sense the environment and situation around human entities. The enormous increase in data (in petabytes) [1] being generated by the smart devices render the need to move data analytics close to where the data resides/originates. To achieve this, the compute and storage capabilities must be moved to the Fog. To achieve the goal of a distributed analytics infrastructure between the Fog and the Cloud needs addressing of a set of unique challenges including: 1) Smart devices in the Fog need to have a platform to run local analytics in a cost-efficient manner (e.g. energy, visualisation, resources and data transmission); 2) Analytics are very domain dependent e.g. pre-processing of noisy data is different for each domain and 3) Not all application require data in real-time as instant insights are not required. Hence, local data storage and retrieval must be possible.

To this end, in this paper, we propose (C)ontext (A)ware (R)eal-time (D)ata (A)nalytics (P)latform, a context aware distributed mobile data analytics platform for the Fog. The key driving factor behind the development of CARDAP is the ability to efficiently and effectively perform distributed data analytics in Fog/Cloud environments. The proposed CARDAP framework is application domain analytics agnostic i.e. a generic platform that can be extended to suit any application analytics requirement. The CARDAP platform breaks monolithic application silos that are in most cases very difficult and expensive to extend/adapt.

This is achieved by separating the data from application analytics (application logic). CARDAP follows a component based software development model enabling dynamic integration of application-specific analytics. The CARDAP platform reduces the efforts required to develop applications that need distributed mobile analytics. CARDAP addresses the previously identified challenges by providing a unique way for smart devices in the Fog to perform data analytics with features such as component based analytics integration, local storage, query and task processing. We envision the CARDAP approach as a novel way to address the big data challenge by moving analytics closer to the source of data. The following are the key contribution of this paper.

- We propose CARDAP, a context aware distributed mobile data analytics platform. CARDAP enabled efficient data analytics in the Fog by providing a standardised component oriented approach to incorporate the required application-specific analytics. CARDAP also addresses the need for local storage and query processing for application that does not require instant insights.
- We propose a cost model for distributed data analytics in Fog/Cloud environments.
- We conduct experimental evaluations using CARDAP platform to evaluate three distributed data analytics strategies.

The rest of the paper is organised as follows. Section 3 presents the recent work in the area of mobile distributed data analytics platforms for the Fog. Section 2 presents the motivations behind the proposed CARDAP approach. Section 4 provides in-depth details on the proposed CARDAP platform architecture. Section 5 presents a cost model for mobile distributed data analytics. Section 6 provides implementation details of the CARDAP platform followed by discussions on experimental evaluations. Section 7 concludes the paper with remarks on future work.

2 Motivation

A typical example of a IoT big data application in the Fog is a mobile crowdsensing application in smart cities. Mobile crowdsensing popularly called community sensing [9,15] is an autonomous collaborative sensing approach that requires minimal user involvement (e.g. continuous processing of noise level around users' location). Mobile crowdsensing applications takes advantage of a population of individuals to measure large-scale phenomenon that cannot be measured using single individual. In most cases, the population of individuals participating in crowdsensing applications share a common goal.

Let us consider an example scenario *monitoring citizen activity in a smart city*. This example scenario is depicted in Fig. 2. The aim of such an application in a smart city environment is to determine the activity of users in an outdoor park (as highlighted by the red polygon in the Fig. 2). The key requirements to satisfy this scenario includes 1) ability to perform a task on demand, in this

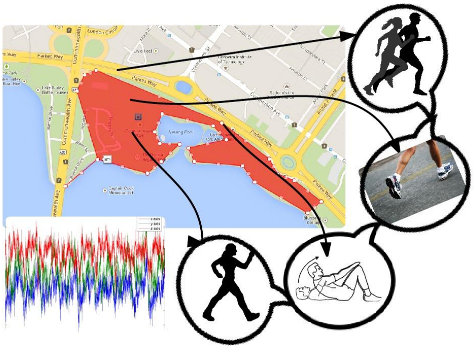


Fig. 2. Scenario: Monitoring Citizen Activity in Smart Cities

example capturing data within a given location from users who are involved in some activity; 2) ability to perform local data analytics, in this example activity recognition on-the-fly [11,10]; 3) ability to store analysed data locally in the Fog that can be queried later by the Cloud (for applications that does not require instantaneous insights) and 4) use an energy-efficient strategy for continuous monitoring and data upload to the Cloud. This scenario can be extended to may such typical smart city example such as monitoring air pollution by mounting sensing component on buses, cars and trams.

As identified by the motivating scenario, there is a need for an efficient mobile data analytics platform that can satisfy the aforementioned requirements to facilitate easy development of IoT applications in the Fog. CSIRO and I2R carry out industry-focussed research and the proposed CARDAP platform has the potential to form the basis for social networking and smart grid applications which have become very popular and have clear commercial advantage.

3 Related Work

Nowadays, with the advent of technology, mobile devices have significantly sensing abilities, computing power, communication and storage resources. They are commonly equipped with various sensors, such as GPS, accelerometer, microphone and camera. Consequently, mobile crowd-sensing (MCS), which analyses crowd behaviour by monitoring large-scale environmental information generated from individual devices, has emerged as a important research topic. Several MCS applications have been successfully developed to discover individual and community trends, such as transportation activities in urban spaces [22], traffic monitoring [23], and collaborative searching [19].

In MCS, mobile devices continuously sense information about the environment and upload the sensed data to the Cloud/remote servers. These processes are obviously energy expensive and may cause battery drain in some cases. Moreover, the raw data generated by physical sensors is usually huge (big data) and can not be used directly as application inputs. Usually, mobile devices need to pre-process the data, perform primitive local processing and only upload intermediate results to the backend servers for further analytics. This approach not only helps to reduce energy and network bandwidth consumption, but also avoid overwhelming of raw data at the backend. Therefore, distributed mobile analytics play an important role to the success of many crowd-sensing applications. A taxonomy of distributed mobile data analytics approaches is presented [21].

In distributed data analytics, a key challenge is to design scalable, ease-to-use frameworks that supports to perform local analytics (mobile device) and global analytics (server) effectively [6]. Mobile devices with limited resources can perform local analytics, such as, converting raw sensed data into application consumable data (e.g. analog-to-digital converter), remove noise, aggregate/summarise sensed data from many sensors and perform light-weight mining tasks [8]. Meanwhile, the aim of global analytics is to discover overall patterns of the environment. Furthermore, the system should understand common information needs of similar procedures in an application or different applications in the same domains to avoid duplicate efforts. Another challenge in distributed data analytics is to understand contexts for proper problem-solving in the right circumstances. For example, mobile devices are configured to upload processed data with 3G connection and raw sensed data with wifi connection. Or, in order to save bandwidth usage, mobile devices only update new data when there are significant changes in the sensed data.

In literature, there are several efforts on designing scalable frameworks for crowd-sensing applications, including high-level abstraction for sensing information [20], task description [16], and a component-based design for quick sensing application deployments [12]. Ye *et al.* proposed MECA [20] (Mobile Edge Capture and Analytics) middle-ware for crowd-sensing. The framework has three logical layers: data, edge, and phenomena/application. After receiving phenomenon specifications from applications, MECA configures edge nodes and devices for corresponding raw data collection and edge analytics processing. Thereby, users focus on implementing their applications' logic without concern with device interaction.

Ravindranath *et al.* [16] introduced CITA, a system that eases the development and running of tasking applications on smart phones. End users create task by writing only on server-side code in form of "condition/action" rules. CITA automatically partitions the code, deals with device coordination, and efficiently executes code on the devices. The framework is currently under developing with their own activity context-aware applications and has not been published yet.

Recently, Jayaraman *et al.* [12] proposed component-based platform for opportunistic sensing applications, named MOSDEN. In this approach, each smartphone occupies a MOSDEN instance in order to run applications with minimal user interaction. Sensors communicate with MOSDEN platform via the concept of plug-in; thereby, a new sensor can be easily added or removed from the system at running time. A conceptual description of the plugin is in XML format. MOSDEN provides a true zero programming middleware, where users do not need to write program code and supports both push and pull data streaming mechanisms.

Context-aware in mobile crowd-sensing has recently attracted a lot of research work. By understanding the context, mobile devices are not simply data collectors, but they can act dynamically according its sensed data and users' needs. Thereby, context-aware crowd-sensing applications can provide more elegant and meaningful solutions, for example, minimise user intervention and optimise the consumption of resources of mobile devices.

The OPPORTUNITY project has been developed to build a ready-to-use middleware and prototypical implementation for evaluating and testing for applications on opportunistic human activity and context recognition [17]. Some preliminary results have been reported, such as, signal conditioning and feature abstraction, autonomous evowhich can lution and adaptation. Carresra *et al.* [5] proposed an adaptive sampling algorithm for user localisation. The key idea is to trade off the accuracy of the location estimation with the battery consumption by varying the type the localisation methods. The application switch to GPS localisation when there is uncertainty on the user location due to the coarse of network localisation.

Sherchan *et al.* [18] developed Context-Aware Real-time Open Mobile Miner (CAROMM) to facilitate data collection from mobile users for crowdsensing applications. CAROMM aims to reduce energy and bandwidth consumption related to continuous sensing and uploading in crowd-sensing applications. Cooperating with resource-aware clustering, CAROMM send only analysed information from each device when it identifies significant changes in the situation. This approach not only reduces the frequency and amount of transferred data, but also guarantee that no important information will be lost. Based on CAROMM framework, Jayaraman *et al.* [13] later demonstrated another context-aware crowd-sensing application that collect sensory data and activity data from a large number of mobile users. It classifies places into different types of contexts as lively, busy and quiet based on light levels, noise levels, crowd intensity, and user activity levels. The application is able to provide real-time reasoning about different situations/ambience of the locations.

4 CARDAP – Distributed Mobile Data Analytics Platform

In this section, we present the system architecture of CARDAP. The proposed CARDAP system is an outcome of our previous works namely Context-aware open mobile miner (CAROMM) [18] and Mobile Sensor Data Engine (MOSDEN) [12]. We first present a big picture view of the proposed CARDAP concept. We then present architectural details of the CARDAP system.

4.1 A Model for Distributed Mobile Data Analytics

As mentioned in Section 3, a key challenge facing distributed data analytics is to design scalable, ease-to-use frameworks that supports local analytics (mobile device) and global analytics (server) effectively. The focus of this paper is on the local analytics area in the Fog. But to provide a big picture, we first present our model for a distributed mobile data analytics system. The big picture illustrated in Fig. 3 captures our vision. In our model, the smart devices in the Fog can include any Internet connected device such as individual to group of smart mobile devices capturing situational context information from the user and his/her environment, micro-sized processing platforms (e.g. raspberry pi mounted on buses).

We model a request from users/other smart devices/data agencies as a set of tasks ta_1,ta_2 ... ta_n. Each task ta_i where $1 < i < n$ has an associated deadline dl_1, dl_2, ... dl_n. Each task is also associated with a set of minimum capabilities (C^{ta}_{min}) c_1,c_2 ... c_m that is required to accomplish the task. Examples of capability include specific sensor requirements, data analytic model requirements, etc. sd_1,sd_2 ... sd_k represents the set of mobile smart devices that are in the Fog. Each mobile smart device sd_j where $1 < j < k$ also has the set of associated capabilities C^{sd}. At any given time t, a mobile smart device can perform a task ta_i if and only if task minimum capability $C^{ta}_{min} \subseteq$ mobile device capabilities C^{sd}.

We note, this paper focuses on CARDAP's architecture and experimental evaluations validating its energy-efficiency. The task assignment functionality of the scheduler as depicted Fig 3 based on capabilities and deadlines is outside the scope of this paper.

4.2 System Architecture

As stated earlier, the CARDAP architecture is developed from our previous works namely CAROMM and MOSDEN. CAROMM is a context-aware open mobile miner platform that is underpinned by continuous mobile data stream

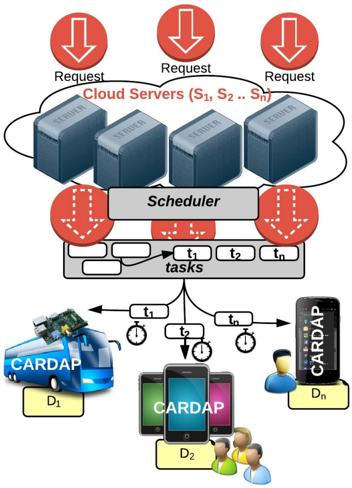


Fig. 3. Distributed Mobile Data Analytics - Big Picture

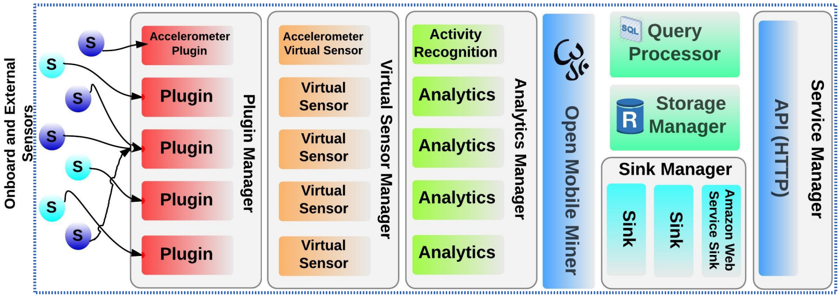


Fig. 4. CARDAP - System Architecture

mining. The mobile data stream mining is used to determine change in contextual information being monitored thus reducing the amount of data being transmitted to the Cloud from the mobile device. The MOSDEN platform is a generic mobile crowdsensing application development platform that breaks the tight coupling between data and application specific processing. MOSDEN framework enables local storage of sensed data. The limitation of CAROMM was its inability to store and query data on-demand and separate analytics and sensing. The limitation of MOSDEN was unavailability of smart processing techniques to reduce bandwidth usage (i.e. data upload). CARDAP was developed to address these limitation of the respective systems. The architecture of CARDAP is presented in Fig. 4.

The five key components of the CARDAP architecture is the data stream capture component, the analytics component, the open mobile miner component, the data sink component and the storage and query processor component.

Data Stream Capture Component: The data stream capture component of CARDAP uses a component-based approach namely plugins and virtual sensors to interface with data sources. The data sources could range from on-board sensors for mobile devices to externally sensors in-case of systems like raspberry pi. The plugin and virtual sensors together enable integration of heterogeneous data sources to the CARDAP platform. The virtual sensor is an abstract representation of a physical/logical sensors in the CARDAP system.

Analytics Component: The analytics component feature allows developers and users to implement application-specific data analytics algorithms. CARDAP, incorporates an mobile activity recognition algorithm namely StreamAR [3]. The StreamAR algorithm is a personalised and adaptive framework for activity recognition that incrementally learns from evolving data stream. The developed framework deals with high speed, multi-dimensional streaming data to learn, model, recognise personalised user's activities. StreamAR system is divided into four phases. A supervised learning phase, where a learning model is built from a set of examples that describe the data domain. An unsupervised learning phase, that employs windowing technique on data stream in order to break down the unlabelled data. A recognition phase handled by an ensemble prediction technique based on a hybrid similarity measures algorithm and finally an incremental and continuous learning phase where the learning model is refined and updated in real time to reflect recent changes.

Open Mobile Miner: The open mobile miner component is the outcome of our CAROMM platform [18]. The CAROMM platform incorporates a data analysis and clustering engine. The engine employs the Light Weight Clustering algorithm [7]. The LWC algorithm uses data adaptation techniques to match high-speed data streams and achieves optimum accuracy based on available resources. CAROMM incorporates a change detect techniques that employs the LWC algorithm to continually monitor significant change in data. This approach has been effectively used by CAROMM for efficient data reduction (reduce number of data transmission) while maintaining a high level of data accuracy [18].

Data Sink Component: The data sink component depicted as Sink in Fig. 4 allows application to push data to any external sink. For e.g. push data to a publish/subscribe bus. This feature is an extension to MOSDEN as MOSDEN only allows local storage and querying of data. Combining the open mobile miner component and the sink, data reduction while transmitting data to a Cloud server can be achieved. The CARDAP platform incorporates sink functionality to upload data to Amazon Web Service (AWS)[1].

Storage and Query Component: The storage and query components of CARDAP perform the functions of storing processed data locally that can be later queried using a RESTful API over HTTP. This feature of CARDAP allows the platform to work independent of a global coordinator (scheduler) thus allowing autonomous task execution. On demand, the global scheduler can perform selective querying of captured and processed data. The query manager uses SQL to resolve incoming requests. The CARDAP platform supports both push and pull approaches to query data from the smart device.

The key feature of the CARDAP platform is its ability to facilitate development of new distributed data analytics applications by wiring the required components using XML configuration files.

5 Cost Model for Mobile Distributed Data Analytics

In this section, we develop cost models for the different distributed mobile data analytic approaches possible in the *CARDAP* . The cost models proposed for the different data collection approaches are:

5.1 Data Transmission Cost Model

The cost of data transmission from n devices to m servers for a considered time period t. The data transmission cost for each device i is defined:

$$Cost_i^{dt} = totaldata(bytes) \tag{1}$$

[1] `http://aws.amazon.com/`

When multiple devices are considered the cost becomes:

$$Cost^{dt}_{n*m} = \sum_{i=1}^{n} Cost^{dt}_{i} * m \tag{2}$$

Let us define the cost of sending all the raw data to the server(s) as $Raw.Cost^{dt}_{n*m}$ and the strategy k (for data collection) data transmission cost as $S_k.Cost^{dt}_{n*m}$. We calculate the bandwidth gain $Gain^{dt}_{S_k}$ of a strategy k in relation to the full raw data sending strategy as:

$$Gain^{dt}_{S_k} = \frac{Raw.Cost^{dt}_{n*m}}{S_k.Cost^{dt}_{n*m}} \tag{3}$$

5.2 Energy Usage Cost Model

The cost of performing data analytics on the local device, in terms of resource consumption (e.g., energy). The energy drainage (%) on the battery of a single device for a time period t is modelled by $Cost^{eu}$. This value can be assessed for the different strategies used in *CARDAP* and can be decomposed into $Cost^{eu}_{s}$ that is the impact that sensing plays on the drain, $Cost^{eu}_{pr}$ that is the energy cost of processing the data and the energy cost of transferring the data $Cost^{eu}_{dt}$.

$$Cost^{eu} = Cost^{eu}_{s} + Cost^{eu}_{pr} + Cost^{eu}_{dt} \tag{4}$$

The $Cost^{eu}_{s} = freq_s * \alpha$ can be described as a function of the sensing frequency $freq_s$, where α is a constant for each device. The $Cost^{eu}_{pr} = CPU_{\%} * \beta$ component can be described as a function of the $CPU_{\%}$ usage and β is another constant defined for each device. The $Cost^{eu}_{dt}$ of transferring the data is a function of the number of (bytes) that need to be transferred. For each different strategy k each of the components of the energy usage cost model will be different. For the particular strategy of sending all the raw data $Raw.Cost^{eu}$ the $Cost^{eu}_{pr} \approx 0$, the $Cost^{eu}_{dt}$ is the maximum possible value and the $Cost^{eu}_{pr}$ takes the value of β since it represents the minimum CPU processing value. We evaluate the energy gain $Gain^{eu}_{S_k}$ of a strategy k $S_k.Cost^{eu}$ in relation to the full raw data sending strategy $Raw.Cost^{eu}$ as:

$$Gain^{eu}_{S_k} = \frac{Raw.Cost^{eu}}{S_k.Cost^{eu}} \tag{5}$$

6 Implementation and Evaluation of CARDAP

6.1 Implementation

In this section, we present implementation details and experimental evaluation outcomes of CARDAP platform. For proof-of-concept implementation purposes, we consider a mobile crowdsensing scenario as described in the motivation section (Section 2). The CARDAP platform has been developed for the Android [2]

[2] `http://www.android.com/`

platform using the Android SDK v4.2.2. For experimentation, we used a Google Nexus 7 tablet (CPU: Quad-core 1.2 GHz Cortex-A9, Memory: 1GB). We implemented interface to sensors as discoverable plugins using the Android interface definition language (AIDL)[3]. This allows CARDAP platform to independently discover interfaces to on-board and external sensors.

The activity recognition engine namely StreamAR was implemented as a data analytics component on CARDAP. To test activity recognition, we wired the accelerometer sensor plugin with the StreamAR component. For experimentation purposes, we used accelerometer dataset from the WISDM lab work on activity recognition [14]. The dataset has 1.1 million data points with activities including walking, jogging, sitting, standing etc.

The LWC algorithm implemented as a part of the CAROMM [18] component in CARDAP detects significant changes in streaming data continuously. Since the activity recognition is state based i.e. either walking or sitting, for evaluation purposes, we use the on-board light sensor to detect significant change in environmental light.

6.2 Evaluation

We evaluate and validate CARDAP's resource and energy efficient performance against the following typical distributed mobile data analytics strategies:

– *Naive approach(baseline/raw data upload)*. All data is collected and sent to the Cloud for further processing (mobile does data collection based on a given time window)
– *Local Analytics (LA)*: Smart mobile device does local analytics and stores data locally which can be queried on demand. For experimentation, we incorporate mobile activity recognition as the local analytics.
– *Local Analytics + Smart data reduction + On-demand sensing (LA-DR-OS)*: Smart mobile device does local analytics, stores data locally and sends data when significant change in the processed data is detected and a pre-defined condition for upload is satisfied. E.g. Record the profile of users who are running within a given location from 5 PM to 8 PM and upload data only when significant change in light value is detected.

We note, the LA and the LA-DR-OS strategies are native features of the proposed CARDAP platform. Depending on application requirements, either of these strategies can be employed for a distributed mobile data analytics application like crowdsensing. The experimental evaluations also validate the performance gain achieved by employing the proposed CARDAP-based strategies namely *Local Analytics - LA* and *Local Analytics + Smart data reduction + On-demand sensing - LA-DR-OS*.

The resource consumption experiment compute the amount of memory and CPU consumed by the CARDAP platform when working under each of the

[3] http://developer.android.com/guide/components/aidl.html

aforementioned strategies. The memory consumption is measured in MB and the CPU consumption in jiffy[4]. The energy consumed by each strategy is computed in milli Watts(mW). To compute the energy and resource consumption, we implemented a modified version of the PowerTutor[5] open source android power monitoring application.

The results of the experimental evaluations are presented in Fig. 5a, 5b, 6a and 6b. We compute the gain for strategies *LA* and *LA-DR-OS* over the naive approach using equation (5). An experimental round consisted of replaying the WISDM dataset for each strategy and computing the average CPU, memory and power consumed by CARDAP platform for a time period of 1 hour. The sampling rate for activity recognition was 1 every minute. This resulted in 60 data uploads when experimenting CARDAP using the naive approach.

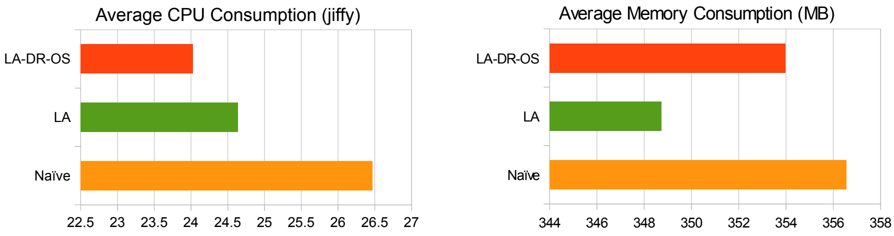
Fig. 5a and Fig. 5b presents the memory consumption of the CARDAP platform when evaluated under each strategy. It is to be noted that the memory and CPU allocation is controlled by Android operating system. Depending on the process workload, android may allocate more memory and CPU cores to maintain system stability. The CPU consumption of LA-DR-OS is lesser than LA as observed in Fig. 5a. Whereas, the memory consumption is vice-versa i.e. LA approach is lesser than LA-DR-OS as noted in Fig 5b. As stated earlier, since memory and CPU allocation is managed by android, the higher memory consumption trend can be attributed to the overheads involved in managing and maintaining the cluster i.e. continuous data stream clustering to monitor significant change in data streams.

A similar trend is observed with the average power consumption experiment presented in Fig. 6a. LA-DR-OS consumes more power when compared to LA approach due to the following factors 1) overhead to maintain and manage clusters and 2) network consumption due to upload of data when significant change in data is detected and satisfies a pre-defined condition. As indicated by the experimental outcomes, the CARADAP native approach namely LA and LA-DR-OS performs significantly better than the baseline approach of uploading raw data. This observation is supported by the energy gain outcome presented in Fig.6b.

Overall, experimental evaluations clearly validate the resource and energy efficiency of the proposed CARDAP platform irrespective of the strategy. Further, CARDAP's native approaches perform significantly better than the baseline approach making CARDAP strategies an efficient and effective technique to realise the development of energy and resource-efficient distributed mobile data analytics applications (e.g. crowdsensing). The proposed CARDAP platform is the first step in the development of a complete distributed mobile data analytics platform as presented in our big picture in Fig. 3.

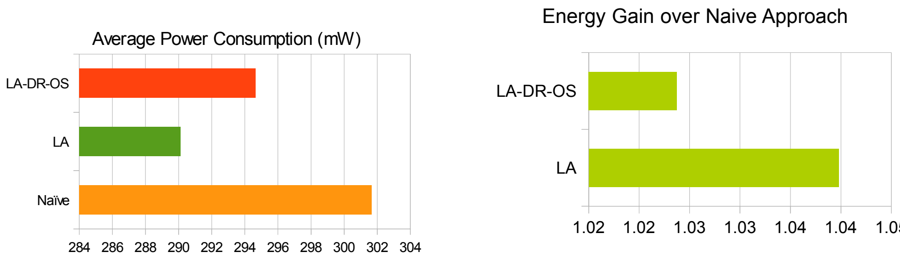
[4] In computing, a jiffy is the duration of one tick of the system timer interrupt. It is not an absolute time interval unit, since its duration depends on the clock interrupt frequency of the particular hardware platform.

[5] `http://powertutor.org/`



(a) CPU (b) Memory

Fig. 5. Average CPU and Memory Consumption



(a) Power (b) Computed Gain

Fig. 6. Power Consumption and Gain

7 Conclusion

In this paper we have presented CARDAP, a context aware real-time data analytics platform for the Fog. CARDAP is a generic, flexible and extensible, component-based platform capable to deploy complex distributed mobile analytics applications such as on-demand distributed mobile crowdsensing. In addition, we discussed different real-world scenarios where using CARDAP can be significantly beneficial. CARDAP incorporates a number of energy efficient data delivery strategies employing real-time mobile data stream mining for data reduction. Our experimental evaluations indicate that the CARDAP platform can deliver significant benefits in terms of CPU, memory and energy efficiency over baseline approaches. In our future work, we aim to investigate the Cloud part of the proposed distributed data analytics model developing cost-efficient task scheduling and smart device selection approaches.

Acknowledgement. Part of this work has been carried out in the scope of the ICT OpenIoT Project which is co-funded by the European Commission under seventh framework program, contract number FP7-ICT-2011-7-287305-OpenIoT. The authors acknowledge help and support from CSIRO Sensors and Sensor Networks Transformational Capability Platform (SSN TCP).

References

1. Vision paper - distributed data mining and big data (August 2012)
2. The internet of things is poised to change everything, says idc (October 03, 2013)
3. Abdallah, Z., Gaber, M., Srinivasan, B., Krishnaswamy, S.: Streamar: Incremental and active learning with evolving sensory data for activity recognition. In: 2012 IEEE 24th International Conference on Tools with Artificial Intelligence (ICTAI), vol. 1, pp. 1163–1170 (November 2012)
4. Bonomi, F., Milito, R., Zhu, J., Addepalli, S.: Fog computing and its role in the internet of things. In: em Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing, MCC 2012, pp. 13–16. ACM, New York (2012)
5. Carreras, I., Miorandi, D., Tamilin, A., Ssebaggala, E.R., Conci, N.: Crowd-sensing: Why context matters. In: 2013 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), pp. 368–371. IEEE (2013)
6. Gaber, M.M., Gama, J., Krishnaswamy, S., Gomes, J.B., Stahl, F.: Data stream mining in ubiquitous environments: state-of-the-art and current directions. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 4(2), 116–138 (2014)
7. Gaber, M.M., Krishnaswamy, S., Zaslavsky, A.: Cost-efficient mining techniques for data streams. In: Proceedings of the Second Workshop on Australasian Information Security, Data Mining and Web Intelligence, and Software Internationalisation, ACSW Frontiers 2004, vol. 32, pp. 109–114. Australian Computer Society, Inc., Darlinghurst (2004)
8. Gaber, M.M., Stahl, F., Gomes, J.B.: Pocket Data Mining: Big Data on Small Devices, vol. 2. Springer (2014)
9. K., R., Ye, F., Ganti, H.L.: Mobile crowdsensing: current state and future challenges. IEEE Communications Magazine 49(11), 32–39 (2011)
10. Gomes, J.B., Krishnaswamy, S., Gaber, M.M., Sousa, P.A., Menasalvas, E.: Mars: a personalised mobile activity recognition system. In: 2012 IEEE 13th International Conference on Mobile Data Management (MDM), pp. 316–319. IEEE (2012)
11. Gomes, J.B., Krishnaswamy, S., Gaber, M.M., Sousa, P.A.C., Menasalvas, E.: Mobile activity recognition using ubiquitous data stream mining. In: Cuzzocrea, A., Dayal, U. (eds.) DaWaK 2012. LNCS, vol. 7448, pp. 130–141. Springer, Heidelberg (2012)
12. Jayaraman, P.P., Perera, C., Georgakopoulos, D., Zaslavsky, A.: Efficient opportunistic sensing using mobile collaborative platform mosden. In: 2013 9th International Conference Conference on Collaborative Computing: Networking, Applications and Worksharing (Collaboratecom), pp. 77–86 (October 2013)
13. Jayaraman, P.P., Sinha, A., Sherchan, W., Krishnaswamy, S., Zaslavsky, A., Haghighi, P.D., Loke, S., Do, M.T.: Here-n-now: A framework for context-aware mobile crowdsensing. In: Proc. of the Tenth International Conference on Pervasive Computing (2012)
14. Kwapisz, J.R., Weiss, G.M., Moore, S.A.: Activity recognition using cell phone accelerometers. SIGKDD Explor. Newsl. 12(2), 74–82 (2011)
15. Le, V.-D., Scholten, H., Havinga, P.: Towards opportunistic data dissemination in mobile phone sensor networks. In: Eleventh International Conference on Networks, ICN 2012, pp. 139–146. International Academy, Research and Industry Association (IARIA), France (2012)

16. Ravindranath, L., Thiagarajan, A., Balakrishnan, H., Madden, S.: Code in the air: simplifying sensing and coordination tasks on smartphones. In: Proceedings of the Twelfth Workshop on Mobile Computing Systems & Applications, p. 4. ACM (2012)
17. Roggen, D., Forster, K., Calatroni, A., Holleczek, T., Fang, Y., Troster, G., Lukowicz, P., Pirkl, G., Bannach, D., Kunze, K.: Opportunity: Towards opportunistic activity and context recognition systems. In: IEEE International Symposium on World of Wireless, Mobile and Multimedia Networks & Workshops, WoWMoM 2009, pp. 1–6. IEEE (2009)
18. Sherchan, W., Jayaraman, P., Krishnaswamy, S., Zaslavsky, A., Loke, S., Sinha, A.: Using on-the-move mining for mobile crowdsensing. In: 2012 IEEE 13th International Conference on Mobile Data Management (MDM), pp. 115–124 (July 2012)
19. Yan, T., Kumar, V., Ganesan, D.: Crowdsearch: exploiting crowds for accurate real-time image search on mobile phones. In: Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services, pp. 77–90. ACM (2010)
20. Ye, F., Ganti, R., Dimaghani, R., Grueneberg, K., Calo, S.: Meca: mobile edge capture and analysis middleware for social sensing applications. In: Proceedings of the 21st International Conference Companion on World Wide Web, pp. 699–702. ACM (2012)
21. Zaslavsky, A., Jayaraman, P.P., Krishnaswamy, S.: Sharelikescrowd: Mobile analytics for participatory sensing and crowd-sourcing applications. In: 2013 IEEE 29th International Conference on Data Engineering Workshops (ICDEW), pp. 128–135. IEEE (2013)
22. Zheng, Y., Zhang, L., Xie, X., Ma, W.-Y.: Mining interesting locations and travel sequences from gps trajectories. In: Proceedings of the 18th International Conference on World Wide Web, pp. 791–800. ACM (2009)
23. Zhou, P., Zheng, Y., Li, M.: How long to wait?: predicting bus arrival time with mobile phone based participatory sensing. In: Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services, pp. 379–392. ACM (2012)