

A Hybrid Approach for Data Analytics for Internet of Things

Badraddin Alturki
University of Leicester
Leicester
United Kingdom
baba1@leicester.ac.uk

Stephan Reiff-Marganiec
University of Leicester
Leicester
United Kingdom
srm13@leicester.ac.uk

Charith Perera
Newcastle University
Newcastle upon Tyne
United Kingdom
charith.perera@ncl.ac.uk

ABSTRACT

The vision of the Internet of Things is to allow currently unconnected physical objects to be connected to the internet. There will be an extremely large number of internet connected devices that will be much more than the number of human being in the world all producing data. These data will be collected and delivered to the cloud for processing, especially with a view of finding meaningful information to then take action. However, ideally the data needs to be analysed locally to increase privacy, give quick responses to people and to reduce use of network and storage resources. To tackle these problems, distributed data analytics can be proposed to collect and analyse the data either in the edge or fog devices. In this paper, we explore a hybrid approach which means that both in-network level and cloud level processing should work together to build effective IoT data analytics in order to overcome their respective weaknesses and use their specific strengths. Specifically, we collected raw data locally and extracted features by applying data fusion techniques on the data on resource constrained devices to reduce the data and then send the extracted features to the cloud for processing. We evaluated the accuracy and data consumption over network and thus show that it is feasible to increase privacy and maintain accuracy while reducing data communication demands.

ACM Classification Keywords

D.2.11. Software Engineering: Software Architectures — Data Abstraction; H.3.4. Information Systems: Systems and Software — Distributed Systems

Author Keywords

Internet of Things; Cloud; Data Analytics; Fog Computing; Edge Computing; Distributed Data Analytics

INTRODUCTION

The Internet of Things (IoT) has become one of the most active areas in computer science and beyond, both for researchers and companies and it is interpreted by communities in a variety of ways. The cluster of European research projects defined the

IoT as allowing "people and things to be connected Anytime, Anyplace, with Anything and Anyone, ideally using Any network and Any service [18]". This is important for applications to monitor, track, and communicate (amongst others) with things for various purposes remotely. Sensors embedded in many everyday physical objects around us play a key role in IoT. These embedded sensors will include vast sensing capabilities [15] and can then send the data through the network to decision points – typically the cloud. After collecting the data, there is a need to analyse them to gain insights to help, automate and speed up decision making [16]. In a data driven economy, the data and insights can be considered as the main goods [14]. According to [17] the number of devices connected to the internet will be more than 50 billion devices in the very near future. However, the greater awareness promised by so many smart things will produce an ever greater volume of data at increasing rates of delivery.

Big data is not a new term in computer science [24], it has been created by big technological companies like Yahoo, Microsoft and Google. Big data as researched has three key characteristics: volume, variety and velocity [24]. As result of improvements in electronics, the cost of equipment has decreased dramatically and sensors have become more affordable and are already embedded into many electronic devices. A great number of data has been generated by these sensors and companies have started storing it – in fact the predominant business model at the moment seems to be around storing and owning data for possible later analytics: most fitness trackers or smart watches will send their recorded data to the cloud of the maker. Part of this desire to store comes from the value of data, part of the fact that data analytics is a notable challenge to which solutions are still being explored. However, the IoT moves the game to an entirely new level by increasing the scale of deployed devices dramatically – posing new challenges to gathering, processing, transporting, storing and analysing data.

There are five steps (Collection, Collation, Evaluation, Decide, and Act) in the so called IoT monitoring cycle [19]. Considering the overall IoT systems, we always have devices at the edge as well as in the network and the cloud in a central position. Data processing can be at the in-network (so edge and devices in the network) and the cloud level [3, 19] – and these levels play a role in the various stages.

Most of the research and existing work in the field of big data focuses on cloud computing because of the offered power in terms of processing and storage. The common way to process

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

IoT 2017, October 22–25, 2017, Linz, Austria

© 2017 ACM. ISBN 978-1-4503-5318-2/17/10...\$15.00

DOI: <https://doi.org/10.1145/3131542.3131558>

the data is to send all data to the cloud and return results after analysis. In addition to the significant power available, processing in the cloud also means that as complete a collection of data is available to analysis as can be obtained. However, processing all streaming raw data in the cloud negatively affects several aspects, such as increased network traffic, latency (to get actions back to the user), energy consumption and privacy. As the IoT grows the need to tackle these issues grows.

We suggest that in the longer term there is an opportunity to move the computation as much as possible off the cloud to the fog or edge device side. This means that data analytics should be handled in the device or fog before sending the data to cloud (possibly going as far as avoiding the cloud altogether for processing of operational data). The cloud would still have a role in longer term backup and also in helping to compute models to guide analytics, in fact there is no doubt that the cloud has an important role in enabling the IoT since it provides high power processing and storage [25]. A key problem to be tackled is to understand how accuracy of analytic results is effected if computations and decisions on raw data are made elsewhere in the processing chain and infrastructure.

So, in this paper we propose a hybrid approach that moves some processing off the cloud and allows us to study the savings in data transfer and changes to accuracy. In the proposed work, we are fusing and filtering data close to the source and then send meaningful higher level data rather than raw data to the cloud. As less details are being transmitted some privacy protection (not every little move is known, only the general picture) is already taking place – however further work in studying the privacy angle needs to be undertaken.

This data fusion technique will directly influence the collation and evaluation steps and hence the crucial question arising is: how can we fuse sensors data locally without harming the accuracy of the overall decision? A secondary question is considering the feasibility of distributing the processing considering that many network and edge devices have less processing power.

The novel contributions of this paper are:

- We propose a hybrid approach, which moves the computation as much as possible to fog/ edge side of the network.
- We extensively evaluate the approach using the WISDM dataset [13] and five of the most popular data analytics techniques.
- We explore the feasibility of applying these data aggregation techniques via resource constrained device, particularly a Raspberry Pi 3 Model B.

The rest of this paper is organised as follows: section 2 describes the solution space while section 3 explores our solution. We then evaluate, consider related work and draw conclusions.

SOLUTION SPACE: DATA AGGREGATION

CERP-IoT [17] provides the following characteristics of the IoT: Autonomous, Intelligence, connectivity, sensing, energy, dynamism, interoperability, privacy and security. The characteristics of data in the IoT are heterogeneity, redundancy,

dynamism and variety. Considering that "data fusion and mining present an efficient way to manipulate, integrate, manage and preserve mass data collected from various things" [23]. Processing IoT data means to add value to the raw data by extracting important aspects and creating meaningful information – an essential element of the IoT [22], [4] identifies five steps to follow when processing IoT data, namely data collection, data pre-processing, transformation of data, mining and evaluation. In this paper we are specifically interested in data fusion, which fits into the area of data pre-processing and transformation and allows to reduce the volume data but increase its value. Data fusion is referred to by other 'synonyms' such as information fusion, decision fusion, data combination, multi-sensor data fusion, sensor fusion and data aggregation. While there is no general agreement on these terms, there are some differences that can be observed: in some cases data fusion is applied on raw sensor data while information fusion is used to determine analysed data, meaning that the latter has a higher semantic grade than data fusion[6]. Similarly, data fusion techniques are used to integrate data from a variety of sources to produce more meaningful and effective inferences and associations, whereas data aggregation can be considered as subcomponent of data fusion which summarises the sensor data to remove data redundancy [1]. The most common definitions by researchers are as follows:

- data fusion is defined by the Joint Directors of Laboratories (JDL) workshop [20] as "a multi-level process dealing with the association, correlation, combination of data and information from single and multiple sources to achieve refined position, identify estimates and complete and timely assessments of situations, threats and their significance."
- Hall and Llinas [11] say that "data fusion techniques combine data from multiple sensors and related information from associated databases to achieve improved accuracy and more specific inferences than could be achieved by the use of a single sensor alone."

Data fusion can be classified depending on a variety of attributes as shown in figure 1 [7]. These attributes are discussed in detail in [1] and generally capture the idea that there are different dimensions such as the abstraction level or the relation between the data items from one or multiple sensors.

Dasarathy's data fusion classification system formalises the attributes just discussed and can be considered as one of the most common approaches [9]. Dasarathy's classification focuses on details of input and output based on the abstraction level. The classification contains five classes as follows [6]):

Data In-Data Out (DAI-DAO) is the primary method of data fusion in the classification model. It processes the raw data that are collected directly from sensors resulting in more accurate data. In addition, image and signal processing algorithms can be used at this stage.

Data In-Feature Out (DAI-FEO) processes the raw data to produce features which can depict a structure about the environment.

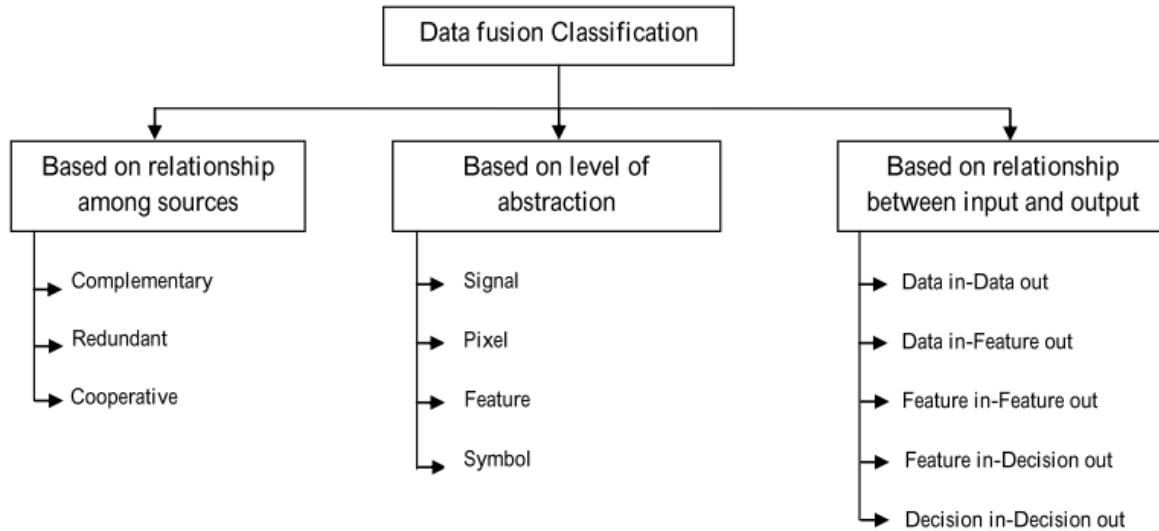


Figure 1. Data Fusion Classification

Feature In-Feature Out (FEI-FEO) processes a collection of features to get more effective feature results.

Feature In-Decision out (FEI-DEO) processes the features to acquire a collection of decisions.

Decision In-Decision Out (DEI-DEO) processes the decisions to extract more efficient decisions.

Features are defined as the single measurements that are used to create the training model. In other words, they are the columns of data that are created for the training set [12]. In addition, data fusion can provide the required knowledge that is essential in a decision-making process, therefore, the amount of the available knowledge / data can effect the final decision at any stage. Many techniques use symbolic information and the data fusion process to determine the uncertainties and restrictions that are part of / effect the decision-making process [6]. In other word, the decision can be captured depending on the knowledge of the events that are collected from variety sources by fusing them.

Informally, our working definition of data fusion can be that it aggregates and integrates all sensor data to allow obtaining accurate and meaningful data while eliminating unneeded and useless data.

Understanding what data fusion can achieve, one also needs to consider the architectural aspect of where data fusion is applied. Options include a centralised, decentralised or distributed architectures as follows [6]:

- **Centralized architecture:** all the collected data from sensors will be sent to the cloud for processing which means that everything is held in one single server. It is known that the cloud is capable to process very large amounts of data effectively. However, in real time scenarios data consumption over the network will be high, which will make the cloud not sufficient for effective fusion of the data. This architecture is also very problematic if the data consists of images

such as earth observation imagery. The reason is that there will be more delays in terms of data arrival time and this will impact badly on the output of data. Additionally, privacy will be one of the main issues because this architecture receives all the raw data without applying any reduction or aggregation previously. Finally, energy consumption has been important in IoT because transferring raw data all time from devices using any network such 3G and WiFi will consume significant amounts of energy.

- **Decentralized architecture:** there are several nodes in the network and each of them has their specific computation capabilities, so there is no single server like centralised system. Every node applies data aggregation autonomously on its local data and data received from peers. One of the major limitations of this architecture is the high communication cost between peers. In this case, if we increase the number of nodes, then there might be a lack of scalability.
- **Distributed architecture:** sensor readings are processed at the source level before applying data aggregation in a specific node that is capable of data fusion. This can overcome various issues of the centralised architecture and can reduce communication costs over the decentralized architecture.
- **Hierarchical architecture:** The data fusion step is performed at a variety of levels in the hierarchy and it can be considered as a combination of both distributed and decentralised architecture.

It is true that it is not possible to say that one of these architecture is the best, as it often depends on specific requirements and technology. Both decentralised and distributed architectures are quite similar to each other in many ways. However, they differ in terms of the place for pre-processing the data. In decentralised architectures the whole data aggregation happens in every node which produces comprehensive output. Whereas, in distributed architectures the raw data is firstly pre-processed at source to extract features, and then these features

are fused. The main advantages of the distributed architecture over the centralised one are reducing the processing and communication costs because it pre-processes the data in a distributed manner before fusing data [6].

It is generally accepted that increasing accuracy and reducing energy usage are major aspects of data fusion [7], so any architecture that is presented needs to consider these aspects. While accuracy is self explanatory, reducing energy is more difficult as the energy used is a combination of costs for storage, transport and processing with transport being very expensive on wireless transmissions technologies.

As there are obvious trade-offs between the different architectures it seems desirable to formulate solutions which combine the different ideas in ways that reduce the disadvantages and benefit from the advantages of each. Our method presented below attempts to achieve this.

PROPOSED SOLUTION: ADAPTIVE DATA AGGREGATION

Overview

We propose a hybrid approach that moves the computation as much as possible from the cloud to the fog/edge level. The overview of this approach is demonstrated in Figure 2. We begin with applying data fusion techniques on sensors data to minimise the number of data points and extract features in IoT devices. Then, we extract features from this data and send it to edge/fog node. This step is important because it is widely accepted that raw time-series data cannot be efficiently analysed by ordinary algorithms for classification. After that, the features will be sent to the cloud for training purposes and creating inferences.

Architecture

The architecture of our proposed solution is divided into two main parts: First, the cloud level has the responsibility of data training and creating inferences. Second, the in-network level aggregates the sensor data to reduce data transmission cost over network. The aim is to save energy, reduce decision times and to increase privacy by analysing and processing data locally while maintaining accuracy as much as possible, reduce data transmission cost over network and save energy by analysing and processing data locally.

The communications between the nodes or peers can be undertaken in different ways as is typical (such as WiFi, 3G and any other solutions). Figure 2 shows the architecture of the system. The distributed processing architecture contains three types of node including IoT Devices, fog, and the cloud as follows:

- A **sensor node** is at the lowest level of the system and is typically embedded in physical objects. Sensor nodes are small and cheap in terms of price to make the process of deploying sensors to objects easy and inexpensive. It senses real-world inputs such as motion detection, temperature and so on. These sensor nodes are connected to Fog nodes via wireless or wired communication.
- A **fog node** resides next to the sensors or along the communication path to the cloud and collects sensors data (or data received from a 'downstream' fog node) and applies data

fusion techniques to extract features. For our architecture they form the main component. Obviously a fog node has less power and a less global data view than a cloud node and hence it can apply less sophisticated data aggregation algorithms. It sends the transformed and fused data to the cloud for further processing and storage if required (ideally the fog node can make the ultimate decision). In an investigation into energy limitation, the authors in [5] found that these devices have restricted energy for particular tasks.

- **Cloud nodes** reside in the cloud and provide the final processing mechanism, obtaining the transformed data from fog nodes. They mainly apply machine learning algorithms and store the data. It is clear that the processing power and storage capability of the cloud is high. This power can be used even more effectively by using the presented approach. According to [5], there is no energy limitation in the devices which are in cloud.

Activity Recognition Using Accelerometer Traces

To validate our architecture we have used the WISDM [13] data set which is a set of accelerometer data on mobiles (particularly Android based) from 36 users who are doing 6 activities (walking, jogging, climbing upstairs, descending downstairs, sitting and standing). These users carried their mobiles while they were performing these activities for a fixed time.

We divided the data into 10 seconds chunks. In addition, 43 features are created depending on 200 readings within the specified chunks. The transformed data contains 5418 accelerometer traces from the 36 users, with in average 150.50 traces per user and a standard deviation of 44.73.

We conducted 3 sets of experiments: Firstly, we apply analytical algorithms on the transformed data in the cloud to calculate the accuracy of each algorithms and the execution time as a baseline. Secondly, we apply analytical algorithms on the transformed data in a fog gateway to calculate the execution time and to check the feasibility of the resource constraint devices while processing the data. Finally, we apply data aggregation algorithms on the raw data to extract features. The final approach minimises the data as much possible in the fog then sends the transformed data to the cloud for analysis. We measure the accuracy and execution time as well as the data amount send to the cloud.

Our hope was that a similar accuracy can be achieved with the third approach without increasing processing time and with significantly reducing network data transmissions.

EVALUATION AND DISCUSSION

Experimental Set up

As mentioned earlier it is not possible to apply classification algorithms on raw data which is time series data. Therefore, there is a need to transform raw data into features [13]. In our experiment, we used a Raspberry Pi 3 model as an example of a low power fog gateway. The used Raspberry Pi has 1GB RAM and runs Raspbian Jessie with Pixel installed as operating system. In addition, to simulate the cloud device we used a 16GB RAM Linux System. We used the weka tool on both sides and we adjusted the heap size in both cloud

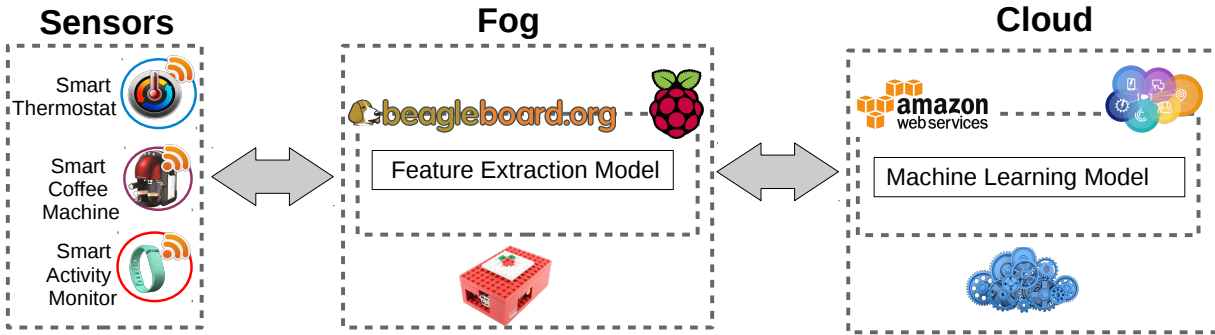


Figure 2. Distributed Processing in Internet of Things

and fog. In the fog the heap size was 650MB, in the cloud we allowed 8GB RAM for our experiment. Moreover, we used the same data aggregation methods that were used to extract features in [13] to allow for comparability. We run data said aggregation methods in the Raspberry Pi to generate meaningful features depending on 200 readings where each has x,y and z acceleration information.

When we used the statistical measurements that are used in [13] we created 43 features including the average of each axis, standard deviation of each axis, average absolute difference of each axis, average resultant acceleration for all axis, time between peaks of each axis and binned distribution for every axis (10 equal sized bins and totally 30 bins).

After the data is prepared we applied five classification methods from the Weka data mining and machine learning tools. The methods include decision tree (J48), logistic regression, multilayer perceptron, and naïve Bayesian. Throughout our experiment we have used 10 fold cross validation.

Results

Figure 3.(a) shows the accuracy results of the 5 analysis algorithms that we applied on the transformed data. It is clear that from results that the multilayer perceptron has the highest accuracy percentage (100% is the best result from the in Cloud analysis on raw data).

Figure 3.(b) shows the data communication time over network from fog (Raspberry Pi) to cloud. There are two bars visible: one for raw data and the other for transformed data. While applying this experiment the upload speed of the internet was 1Mbps. It is clear that the fog only device has no data communication cost because the processing happened in the device and no communication to the cloud took place. However, in the cloud approach the raw data communication over network from fog to cloud is extremely high, whereas in the hybrid approach the transformed data communication over the network from fog to cloud is low. This not surprising result confirms that we can save significantly on data communications by aggregating and pre-processing data early in the chain.

Figure 3.(c) illustrates the execution time of the 5 analytics algorithms in both the cloud and fog device. The results show

us that two algorithms (logistic regression and multilayer perceptron) have significant differences between the two sides. Obviously, the IoT device takes more time than the cloud to execute analytics algorithms because of its resource constraints.

Figure 3.(d) demonstrates the total processing time for the three architectures. There are three measurements for each architecture including the execution time of analytics (ML) algorithms, the execution time of the data transformation process and the data communication time between local device and cloud. This graph needs a bit more explanation as the results are more interesting, so the details are as follows:

- **Fog (Raspberry PI):** The data transformation process is conducted locally and it is clear that the processing time is higher than in cloud. The analytics algorithms have been processed locally and they took much more time than cloud because of the processing power. However, data communication (the time to send data to the cloud) is very low as only aggregated data is being sent for storage. So, overall processing time is in the middle of the measured approaches.
- **Cloud:** Data Communication is the time that the raw data takes from IoT device to cloud, which is clearly high as all raw data is being transmitted. The data transformation process was done in the cloud and due to the available resource runs quickly. Also, the analytics algorithms have been processed in the cloud and they took much less time than the fog because of the processing power. However, overall due to the significant amount of transmission time the cloud is the slowest approach in the given setting.
- **Hybrid:** Here the data transformation process is done locally (in the fog) on the Raspberry Pi, with the usual observation. Data Communication is the time the transformed data takes from IoT device to cloud which as before is low. Finally the analytics algorithms have been processed in the cloud on the transformed data they took much less time than locally because of the processing power. Overall by combining the various strength this leads to a very good execution time.

Both fog and hybrid approaches looks similar to each other in most cases. However, they differ in terms of the place for applying the machine learning algorithms on the transformed data. In the fog approach the whole processing (data fusion and

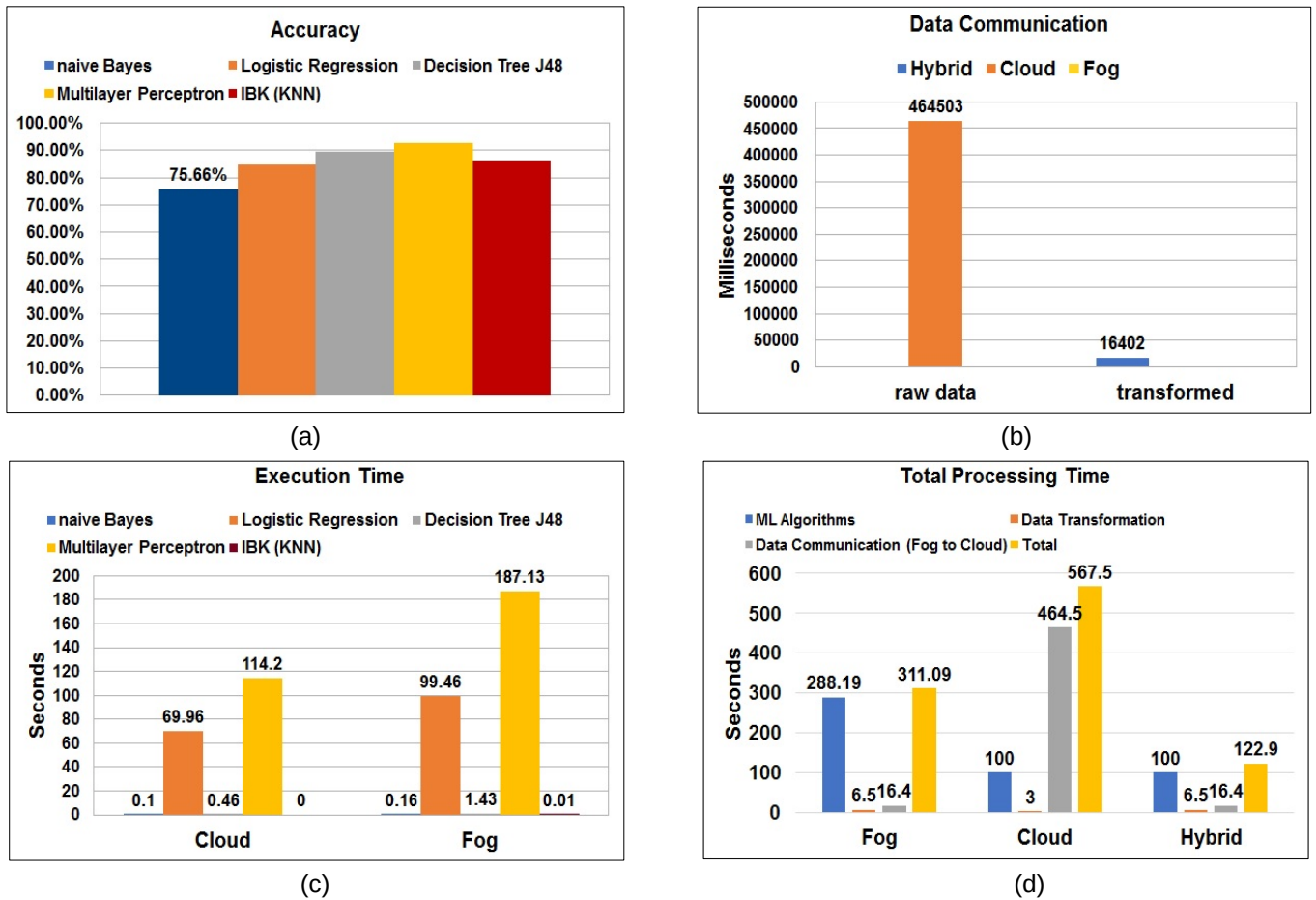


Figure 3. Data Processing Results (a - d)

machine learning algorithms) happens in the node itself which can be considered as decentralised architecture. Whereas, in the hybrid approach the raw data is firstly fused in the fog node to extract features, and then these features are sent to the cloud for applying the machine learning algorithms. The major benefit of the hybrid approach over the fog one is using the power of the cloud for applying machine learning algorithms which need more processing power. Therefore, this step helps in reducing the processing time as a contributor to the overall data processing time.

These initial results show us that the proposed hybrid approach is good enough for the chosen dataset and analytical methods. It is clear from the results that data communication is efficient and provides significant gains.

Observation 1: Data consumption over network. It is accepted that when the size of data is large the data consumption will be more expensive. The raw data was around 1 million rows which is equal to approximately 50 MB. However, after aggregating data into features by using data aggregation algorithms, the number of rows became 5418 rows and the size became 1.2 MB. This means that very significant savings to data transmission and storage can be made by early aggregation. This observation will gradually gain in importance

as the number and quality of sensors increases rapidly and thus the rate and resolution at which data will be delivered grows quickly. By fusing the data locally before sending it to the cloud, we are not only reducing the data we are also determining which data is meaningful and only send that. This will reduce the energy consumption of fog and sensor devices which typically gain internet connectivity through 3, 4 or 5G, thus batteries in the devices will last longer.

Observation 2: Accuracy. Aggregated data leads to less accuracy in the results compared to working with raw data. All presented approaches are effected in the same way. Overall, the loss of accuracy is not drastic: the lowest bar is 75% with the highest being around 93%. Clearly the used analysis method has an impact with trade-offs such as the used local processing power as well as methods optimised for this localised setting being factors that can influence the accuracy. The right balance will in terms of privacy, accuracy, resource cost, energy consumption and data transmission will need to be identified and our future work will further this area.

Out of these experiments and observations we can conclude that one important aspect for future work is a development that combines the ideas of distributed data aggregation with analysis methods that can also be distributed effectively and be

run in low power environments. The fact that they can operate on smaller data sets will help, but somehow the methods need to contain a core part based on global understanding.

Aside. For completeness we like to note that we initially attempted to conduct the experiments with a Raspberry Pi model b with 512MB RAM, however this was not capable to apply some of the weka toolbox analysis algorithms because of the RAM constraint. Hence we used a slightly more powerful version as reported above.

RELATED WORK

In recent years, the main cloud providers have been promising new IoT services with various functionalities and advantages. One of the main cloud providers is Microsoft with its Azure Stack [21] which offers a hybrid cloud that allows companies to transfer benefits from their servers while keeping the management of servers for new types of cloud (hybrid cloud). In addition, they provide gateway devices in the cloud and data analytics. Similarly, IBM has an online web analytics system with IBM Digital Analytics. This service provides tracking and analysing of behaviours from visitors. The data analytics uses high power servers inside IBM. The IBM PureData system promises fast data analytics and warehouse that combine warehouse, data centres and analytics [8]. Although the above systems are promising powerful data analytics approaches, they do not support a fog gateway concept which resides between IoT devices and the cloud. As we can see from our results that uploading high volumes of raw data consumes time and energy. Therefore, a fog gateway concept is important for real time services to save time, energy and resource cost.

Data fusion is an active area in research and business particularly with a view to optimised data analytics. There are several data fusion techniques that focus on reducing the consumption of energy in [10, 2]. They have used a variety of methods including fuzzy set theory and neural networks. They succeeded in terms of removing redundancies while fusing the data. However, they did not focus on the resource constraints of devices that embed the sensors. In contrast, they assume that these devices work efficiently without a need to pay attention to their limitations. More importantly, these mechanisms send all the data to centralised computation systems, which affects the data communication cost, privacy and energy as well. As we could see in our experiments sending the raw data to the cloud is not efficient in terms of data communication over the network.

CONCLUSIONS AND FUTURE WORK

This paper presents a hybrid approach in which data is fused in the fog before being send to the cloud to reduce data communication over the network. The results show that this architecture is successful in terms of reducing data communication cost over network without significantly reducing accuracy of later decision making. We presented the proposed approach and its relevant methods. In addition, we used the WISDM dataset [13] to validate our architecture.

On the basis of the promising findings presented in this paper, future work will involve creating different features with dif-

ferent algorithms for better data aggregation further reducing data communication while attempting to increase (or at least maintain) accuracy. One particular key piece of work has been alluded to in the results section: the development of analysis methods that can be distributed and executed in efficient ways on low power devices. A key strategy here will be the exploration of the balance of analysis with limited (local) data sets vs the availability of a global view. In addition, evaluation of energy consumption and consideration of the positive impact on privacy will be aspects of future work. Furthermore, additional datasets to test and evaluate our hybrid approach further will be investigated.

ACKNOWLEDGMENTS

Dr. Charith Perera's work is funded by EPSRC award number DERC EP/M023001/1 (Digital Economy Research Centre). Badraddin Alturki's research is funded by Saudi Arabian Cultural bureau in London and his scholarship is granted by King Abdul Aziz University.

REFERENCES

1. Ahmed Abdelgawad and Magdy Bayoumi. 2012a. Data fusion in WSN. In *Resource-aware data fusion algorithms for wireless sensor networks*. Springer, 17–35.
2. Ahmed Abdelgawad and Magdy Bayoumi. 2012b. *Resource-Aware data fusion algorithms for wireless sensor networks*. Vol. 118. Springer Science & Business Media.
3. Mervat Abu-Elkheir, Mohammad Hayajneh, and Najah Abu Ali. 2013. Data management for the internet of things: Design primitives and solution. *Sensors* 13, 11 (2013), 15582–15612.
4. Henrique C. M. Andrade, BuÄşra Gedik, and Deepak S. Turaga. 2014. *Fundamentals of Stream Processing: Application Design, Systems, and Analytics*. Cambridge University Press. DOI: <http://dx.doi.org/10.1017/CBO9781139058940>
5. Carsten Bormann, Mehmet Ersue, and A Keranen. 2014. *Terminology for constrained-node networks*. Technical Report.
6. Federico Castanedo. 2013. A review of data fusion techniques. *The Scientific World Journal* 2013 (2013).
7. Sakshi Chhabra and Dinesh Singh. 2015. Data Fusion and Data Aggregation/Summarization Techniques in WSNs: A Review. *International Journal of Computer Applications* 121, 19 (2015).
8. Larry Coyne, Joe Dain, Phil Gilmer, Patrizia Guitani, Ian Hancock, Antoine Maille, Tony Pearson, Brian Sherman, Christopher Vollmar, and others. 2017. *Ibm private, public, and hybrid cloud storage solutions*. IBM Redbooks.
9. Belur V Dasarathy. 1997. Sensor fusion potential exploitation-innovative architectures and illustrative applications. *Proc. IEEE* 85, 1 (1997), 24–38.

10. Hevin Rajesh Dhasian and Paramasivan Balasubramanian. 2013. Survey of data aggregation techniques using soft computing in wireless sensor networks. *IET Information Security* 7, 4 (2013), 336–342.
11. David L Hall and James Llinas. 1997. An introduction to multisensor data fusion. *Proc. IEEE* 85, 1 (1997), 6–23.
12. Peter Harrington. 2012. *Machine learning in action*. Vol. 5. Manning Greenwich, CT.
13. Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. 2011. Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter* 12, 2 (2011), 74–82.
14. IC Ng. 2014. Engineering a Market for Personal Data: The Hub-of-all-Things (HAT), A Briefing Paper. *WMG Service Systems Research Group Working Paper Series* (2014).
15. C. Perera, C. H. Liu, and S. Jayawardena. 2015. The Emerging Internet of Things Marketplace From an Industrial Perspective: A Survey. *IEEE Transactions on Emerging Topics in Computing* 3, 4 (Dec 2015), 585–598. DOI : <http://dx.doi.org/10.1109/TETC.2015.2390034>
16. Jurgo Preden, Jaanus Kaugerand, Erki Suurjaak, Sergei Astapov, Leo Motus, and Raido Pahtma. 2015. Data to decision: pushing situational information needs to the edge of the network. In *Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2015 IEEE International Inter-Disciplinary Conference on*. IEEE, 158–164.
17. Harald Sundmaeker, Patrick Guillemin, Peter Friess, and Sylvie Woelfflé. 2010. Vision and challenges for realising the Internet of Things. *Cluster of European Research Projects on the Internet of Things, European Commission* (2010).
18. Ovidiu Vermesan, Peter Friess, Patrick Guillemin, Sergio Gusmeroli, Harald Sundmaeker, Alessandro Bassi, Ignacio Soler Jubert, Margaretha Mazura, Mark Harrison, Markus Eisenhauer, Pat Doody, Friess Peter, Guillemin Patrick, Gusmeroli Sergio, Bassi Harald, Sundmaeker Alessandro, Jubert Ignacio Soler, Mazura Margaretha, Harrison Mark, Eisenhauer Markus, and Doody Pat. 2009. Internet of Things Strategic Research Roadmap. *Internet of Things Strategic Research Roadmap* (2009), 9–52. DOI : http://dx.doi.org/pdf/IoT_Cluster_Strategic_Research_Agenda_2011.pdf
19. Meisong Wang, Charith Perera, Prem Prakash Jayaraman, Miranda Zhang, Peter Strazdins, and Rajiv Ranjan. 2015. City data fusion: Sensor data fusion in the internet of things. *arXiv preprint arXiv:1506.09118* (2015).
20. Franklin E White. 1991. *Data fusion lexicon*. Technical Report. DTIC Document.
21. J. Woolsey. 2016. Powering the Next Generation Cloud with Azure Stack. (2016).
22. Fatos Xhafa and Leonard Barolli. 2014. Semantics, intelligent processing and services for big data. (2014).
23. Zheng Yan, Jun Liu, Athanasios V Vasilakos, and Laurence T Yang. 2015. Trustworthy data fusion and mining in Internet of Things. *Future Generation Computer Systems* 49, C (2015), 45–46.
24. Arkady Zaslavsky, Charith Perera, and Dimitrios Georgakopoulos. 2013a. Sensing as a service and big data. *arXiv preprint arXiv:1301.0159* (2013).
25. Arkady B. Zaslavsky, Charith Perera, and Dimitrios Georgakopoulos. 2013b. Sensing as a Service and Big Data. *CoRR* abs/1301.0159 (2013). <http://arxiv.org/abs/1301.0159>