Fog-to-Cloud (F2C) Data Management for Smart Cities

Amir Sinaeepourfard, Jordi Garcia, Xavier Masip-Bruin, Eva Marin-Tordera
Advanced Network Architectures Lab (CRAAX),
Universitat Politècnica de Catalunya (UPC BarcelonaTech)
Barcelona, Spain
{amirs, jordig, xmasip, eva}@ac.upc.edu

***Abstract*—Smart cities are the current technological solutions to handle the challenges and complexity of the growing urban density. Traditionally, smart city resources management rely on cloud based solutions where sensors data are collected to provide a centralized and rich set of open data. The advantages of cloud-based frameworks are their ubiquity, as well as an (almost) unlimited resources capacity. However, accessing data from the cloud implies large network traffic, high latencies usually not appropriate for real-time or critical solutions, as well as higher security risks. Alternatively, fog computing emerges as a promising technology to absorb these inconveniences. It proposes the use of devices at the edge to provide closer computing facilities and, therefore, reducing network traffic, reducing latencies drastically while improving security. We have defined a new framework for data management in the context of a smart city through a global fog to cloud resources management architecture. This model has the advantages of both, fog and cloud technologies, as it allows reduced latencies for critical applications while being able to use the high computing capabilities of cloud technology. In this paper, we present the data acquisition block of our framework and discuss the advantages. As a first experiment, we estimate the network traffic in this model during data collection and compare it with a traditional real system.**

Keywords—Smart City; Fog-to-Cloud (F2C) computing; Data Management; Data Lifecycle Model (DLC); Data Aggregation

I. Introduction

Data are the essential fuel for smart cities development because they provide the required information for services to proceed according to the contextual state, and can generate higher value knowledge extracted after some complex data analysis. In fact, smart cities constitute an ideal scenario to generate abundant data from any source type. For instance, data is mainly obtained from the sensors network deployed throughout the city, but also from the increasingly popular participatory sensing (e.g. sensors integrated in citizens' smartphones), from social media or any other third party application, streams of data from surveillance cameras and devices, or any other city resource sensitive to contribute with valuable information.

Managing and organizing efficiently all these diverse sources and tremendous volumes of data in such a context is a critical challenge for an effective smart cities' success. We have estimated that 8 GB of data could be generated every day in the city of Barcelona [1], only considering some basic public sensors' data (for instance, surveillance or traffic control cameras were not considered). However, not many researchers are paying attention to explicit data management strategies in the context of smart cities.

We have defined a novel architecture for efficient data management in the context of a smart city, based on a fog to cloud distributed model for resources management. The advantages of such a model is that it combines the advantages of both the cloud and the fog computing technologies, these are: keeping high performance capabilities for computational intensive applications, reducing communication latencies for real-time or critical services, reducing network data traffic and enhancing fault tolerance and security protection. As a first stage of our research, in this paper we focus on the data acquisition phase, we describe some basic data aggregation optimizations that can be easily implemented in our fog to cloud model, and estimate the effects of such optimizations on the network data traffic reduction.

The rest of this paper is organized as follows. Section 2 discusses some relevant related work about resources and data management models in Smart Cities. Then, Section 3 presents the details of the new architecture for data management in smart cities using the fog to cloud distributed model, and discusses the advantages of this new approach. In Section 4, we describe and implement some basic data aggregation optimizations to illustrate the potential of our proposal. Finally, in Section 5 we conclude this work and present our future research directions.

II. Related Work

Several efforts have traditionally been made to handle data management technologies, generally focused on Relational Database Management (RDBMS) technologies and, more recently, the Extract-Transform-Load (ETL) process for modeling data life stages in the concept of data warehousing environments [2, 3]. Furthermore, the big data paradigm constrained additional challenges to the traditional data management systems in the recent decades [4].

Alternatively, Data LifeCycle (DLC) models represent one great solution towards developing an integral data management framework beyond any specific technology [2, 4]. Several DLC models generated for specific scenarios (such as smart cities [5, 6]), sciences and environments (for instance, big data [2, 7]) have been proposed by several researchers from academia and industries.

With respect to resource management in smart cities, there are two main different trends: i) centralized (cloud) data management, and ii) distributed data management. In the one hand, in the centralized data management model all physical data resources send the sensed data to the cloud data center through a broad area communication network, such as internet. In this context, the cloud environments are the responsible to collect, aggregate, and convert data into meaningful information [6, 8]. On the other hand, the alternative option is the distributed data management model that uses fog technology [9, 10]. Fog computing proposes the use of physical devices at the edge for further processing and preservation. Other authors [11] propose a Fog-to-Cloud (F2C) computing framework that combines the cloud computing (centralized view) with the fog computing (distributed view) models. Although there is few related work about distributed data management [5] in the context of smart cities, it is not yet mature enough how this distributed data management model can manage all data life stages from fog to cloud layers.

In this paper, we argue that data can be organized and managed in any smart city scenario at the fog layers (including data acquisition, data preservation and data processing) while using the all facilities (such as deep computing and unlimited data storage) at the cloud layer. In addition, we show that the F2C distributed data management model provides an excellent opportunity to perform some data optimizations during the data acquisition phase, which provides several advantages, such as reducing latencies for critical applications, reducing network traffic, while being able to use the high computing capabilities of cloud technologies.

III. F2C Data Management

The distributed hierarchical F2C resources management architecture provides an interesting framework for data management in the context of smart cities, according to our Smart City Comprehensive Data LifeCycle (SCC-DLC) model proposal [5]. In this section, we present a novel architecture for efficient fog to cloud data management in smart cities, consisting on the SCC-DLC model mapping onto the smart city F2C resources management architecture. The model is illustrated in Fig. 1. The SCC-DLC consists of three main blocks, named the data acquisition, the data processing, and the data preservation. Data acquisition is mainly performed at fog layer 1, as well as some basic data processing and data preservation actions. The fog layer 2 can enhance the data processing and data preservations capabilities of level 1 by providing higher computing capabilities. And finally, the cloud layer is the responsible of a more complex and more sophisticated data processing over a much broader set of (presumably historical) data, as well as the responsible for permanent data preservation. In the following subsections, the functionalities of each data block are further described, as well as the advantages of the model.

A. Data acquisition

Data acquisition is mainly performed at fog layer 1. In fact, all sensor devices (such as the sensors network deployed throughout the city, but also surveillance cameras or sensor data from smart phones) are part of the fog nodes at this level according to their physical location. So most data is collected at fog layer 1. There can eventually be some additional data collected at cloud level, such as data from web services or third party applications, but these will be smaller compared to the vast volumes of sensor generated data.

As long as data are being collected, the following phases from the data acquisition block can also be performed at fog layer 1, where a reasonable amount of computing resources is available. For instance, the data filtering phase can apply filters to remove redundant data and can apply some additional data aggregation techniques to further reduce the amount of data to be managed. Data quality can also be implemented at this fog layer, assessing and guaranteeing higher data quality. And data description can be performed in order to tag data according to the city business model considered, for instance, timing information (creation, collection, modification, etc.), location positioning (city, country, GPS coordinates), authoring, privacy, and so on.

Data collected at fog layer 1 will be periodically moved upwards to layer 2, and data collected at layer 2 from a set of fog nodes at layer 1 will be combined and periodically moved upwards to the cloud level, which will collect the whole data set from the city and keep it for historical references. Note that data al fog layer 1 can be immediately used at this same level (real-time data), so there is not any need to urgently move these data to higher levels and, therefore, the frequency for the periodical upwards data movements can be strategically decided in order to accommodate it to the network traffic loads.

B. Data storage

Data are generated at fog layer 1, but gradually moved upwards to the fog layer 2, and upwards to the cloud layer, where they will be permanently preserved. So, in fact, the F2C hierarchy acts as a reversed memory hierarchy, where data are created and the lowest cache level (fog layer 1) and moved upwards to “main memory” (cloud layer) instead of being created at the main memory and moved to lower cache levels of the memory hierarchy.

Data generated at fog layer 1 will be temporarily stored at this level, allowing real-time applications an instant access to these data. The smart city business model can decide the amount of temporal data that will be stored at this level, as well as the frequency of updating to upper levels. Similarly, data gathered at fog layer 2, consisting of data received from several fog nodes at layer 1, will be temporarily stored at this level 2. This will make up a set of less recent data (as it has been received after some period of time) but from a broader area, comprising the combination of the respective fog nodes’ areas at layer 1. Finally, data will be moved upwards and permanently preserved at cloud layer, unless any expiry time is defined.

The different management phases included in the data preservation block will be mainly executed at the cloud level, where the permanent storage is performed. Note that these phases are not urgent and, therefore, their execution can be delayed to the time in which data are received to the cloud layer. This is the case of the data classification phase, responsible for classifying and ordering data before storing, and eventually implementing the appropriate techniques for data versioning, data lineage or data provenance. And the data dissemination phase, responsible for providing a user interface for public or private access to stored data, and responsible for implementing any protection, privacy, sharing or security policies according to the city business requirements.

C. Data processing

Data processing can be performed at any layer from the F2C hierarchy, according to the requirements of the application or service. For instance, critical or real-time services will be executed at fog layer 1 in order to have a faster access to the (just generated) real-time data. Note that accessing data locally inside the boundaries of a fog node is much faster than moving the data to a centralized cloud data center and, afterword, reading these same data from the cloud to the local node.

Alternatively, deep computing complex applications will be executed at the cloud layer. Note that i) in the cloud the computing resources are unlimited and, ii) the data set of a high performance computing application will presumably be very large and, therefore, be part of the historical data set stored at the cloud layer. Note that in this case, when computation requires very high capabilities, adding more latency to the first data access will not be significant in the overall performance.

For the other applications, they will be executed at the lowest fog layer that both, provides the required computing capabilities and contains the required data set. As a general rule, the closer the layer, the faster responses times. An additional consideration in this case is when the required data is not present in the current fog node at layer 1, but can be accessed from either a node at a higher layer or a neighbor fog node at the same layer 1. This option may eventually be considered and solved using some sort of cost model to estimate the effects of both cases and proceed according to the lowest cost.

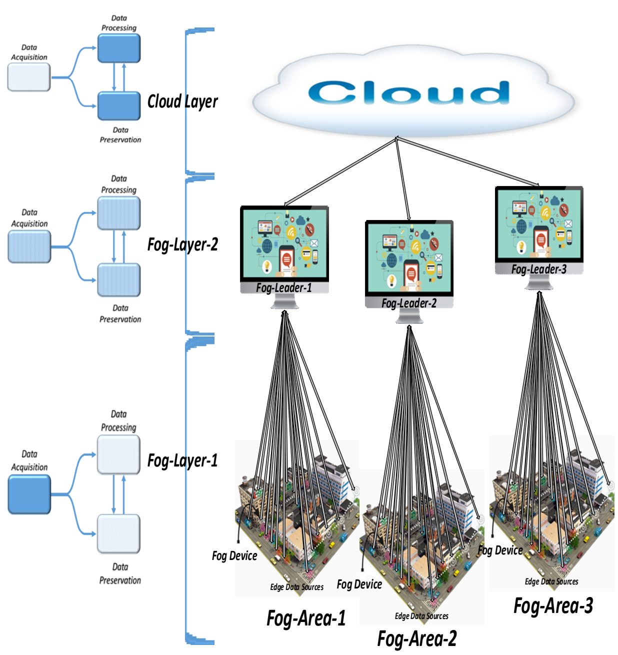


Fig. 1. Mapping of the SCC-DLC model onto the F2C architecture.

D. Advantages of the F2C data management model

The most obvious advantages of this F2C data management model are that it can benefit from the combined advantages of both, the cloud and the fog computing technologies. This is, high computing and storage capabilities from the cloud layer, and reduced network traffic and communication latencies from the fog layers. However, some additional advantages arise from this hierarchical and distributed model, as listed below:

- Real-time data accesses are much faster than in a centralized architecture. This higher speed is not only due to the reduced communication latencies, but due to the fact that accessing data from a centralized system requires the data to be moved first to the cloud, classified and stored there, and then moved back to the edge. So two times the data transfer through the same path must be considered.
- By having the just collected data available at fog layer 1, the global network load is drastically reduced because some applications will be able to access these data locally, avoiding several remote data accesses through the network.
- By having the just collected data available at fog layer 1, the transmission to the cloud can be delayed without any performance loss. This allows additional optimization implementations, such as:
  - Performing some data aggregation techniques to reduce the volume of data to be transmitted upwards, without any computational constraint, as data do no need to be sent immediately.
  - Adjusting the frequency of the data transmission in order to use the network in periods when the traffic load is low.
- Traditional centralized systems define a low frequency policy for data collection from sensors in order to reduce the total amount of data to be transmitted in the network. By having the real-time data available at fog layer 1, the data collection frequency can be increased at this level without overloading network load and, therefore, providing more precision and accuracy from the sensed data at no additional cost.
- By reducing the data transmission length, the security risks and the probability of communication failure are reduced as well and, for this reason, privacy is easily enhanced.

IV. Optimizing Data Collection through Aggregation

In this section, we provide some validation for our distributed data management strategy based on a F2C resources management architecture, by estimating the effects of some basic data aggregation techniques and comparing them with a real centralized cloud system, named Sentilo [18], which manages the municipal open data from the city of Barcelona [1].

A. Data Aggregation

Data Aggregation provides a splendid facility as part of data management to do some kind of processing for gathering, reducing, mixing, or presenting information somehow as a summary [12]. The main objective of data aggregation techniques is reducing the amounts of managed data, and can be obtained through diverse techniques, such as data combination, data redundancy elimination, data compression, bandwidth reduction or power consumption reduction, just to name a few.

Recently, data aggregation has been tailored with the concepts of data and information mining progression, business demands and human analysis techniques, where data must be explored, collected, and presented in a report-based and shortened format in their networks [13]. There are some different view to do data aggregation in theoretical and practical scenarios. Traditional views concentrated to specific network devices and resources such as Wireless Sensor Networks (WSN) to manage data aggregation approaches [14]. The other view extends the previous view to go beyond ubiquitous and distributed scenarios (instead of focusing on specific devices and network) such as big data [13], cloud and distributed computing [13, 15], or real-time systems [16].

In cloud computing environments, cloud computing provides (almost) unlimited, scalable as well as elastic resources. For this reason, cloud computing adopt some data aggregation approaches and techniques to produce high level and sophisticated outcome. In [8], the authors provide a full data model from sensors nodes to cloud computing environments for a smart city scenario. This model has two main layers, which are sensors nodes and cloud computing layers. The sensors nodes collect data from the city and transfer to the cloud computing layer. The cloud layer is responsible to perform data collection and aggregation, data filtering (including classification), and data processing (including preprocessing, processing, and decision making).

With respect to distributed data aggregation, a recent survey presents a taxonomy for distributed data aggregation approaches [15]. They propose two main taxonomies, named communication and computation. The communication taxonomy focuses on the communication aspects (including communication/routing strategy and network topology). It is divided into structured (including hierarchical and ring protocols), unstructured (including flooding/broadcast, random walk, and gossip routing protocols) and hybrid data aggregation approaches. Alternatively, the computation taxonomy encompasses to decomposable functions (including hierarchic, averaging, and sketches basis and principles methods), complex functions (including digests basis and principles methods) and counting (including deterministic and randomized basis and principles methods) data aggregation approaches.

In this work, we will apply some basic aggregation techniques as an example to show the facility and efficiency of our model to use such kind of optimizations. The data aggregation techniques explored are:

- Redundant data elimination: With this technique, we focus on providing a basic yet effective solution to easily reduce the amount of duplicated data collected from the sensors layer. For example, in case of weather measurement, each sensor sends the current temperature measurements, but this type of data is prone to repetitions, so eliminating them may easily reduce such amount of data.
- Compression: As data are collected and transmitted to an upper level delayed, several interesting options arise to accumulate a reasonable amount of data and compute compression, in order to obviously reduce the amount of data transfer.

Many other data aggregation techniques could be easily applied in this architecture; however, these two basic techniques are enough to illustrate the facility and effectiveness of such optimizations in our model.

B. Experimental results

In a previous work [1], we estimated the amounts of data that can be generated (and therefore transmitted through the network to the main cloud data center) in the city of Barcelona, through their data management platform, named Sentilo [18]. In this experimental section, we compare these figures with the estimated data that should be transmitted using a F2C data management model as the one described in the previous section.

The data aggregation and data compression tasks can be performed at the fog device (in fog layer 1), at the fog leader (in fog layer 2), and at the cloud layer as part of the data classification phase, as shown in Fig. 2. According to the current distribution of districts and sections in Barcelona, we estimate that our fog layer 1 should be covered with 73 fog areas, which corresponds to the number of sections in Barcelona. In this case, our fog area covers almost 1 km2, which is a reasonable fog area size. In addition, the fog layer 2 can be defined as 10 main areas (at fog layer 1) which corresponds to the number of districts in Barcelona.

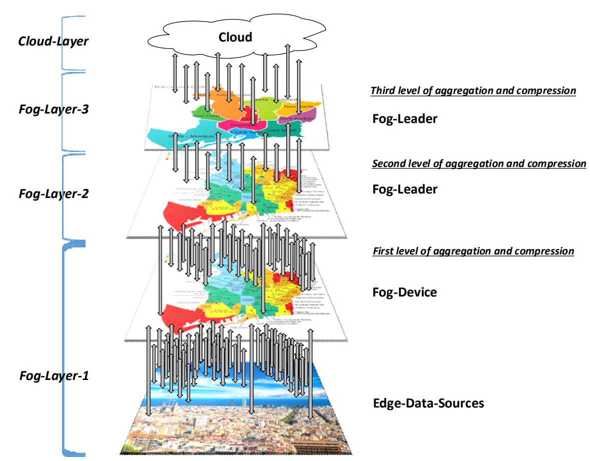


Fig. 2. Representation of the F2C data management in Barcelona.

The data classification phase is the responsible to classify and organize all data collected from the different categories of sensors. In our use case, Sentilo provides five categories of information and services, which are energy, noise, urban, garbage and parking. Each category is divided into different types of information. For instance, the energy category contains electricity meter, external ambient conditions, gas meter, internal ambient conditions, network analyzer, solar thermal installation, and temperature. The noise category includes three different types of information. The urban category encompasses to air quality, bicycle flow, people flow, traffic and weather. The garbage category has container glass, container organic, container paper, container plastic, and container refuse. And finally, the parking category has only one type of information.

The Sentilo platform provides a part of sensors network in Barcelona smart city. In our last paper, we calculated that the Sentilo generated different amount of data per day for different categories of information (including, energy monitoring, noise monitoring, garbage collection, parking spot monitoring, and urban lab monitoring) as shown in below:

- Energy Monitoring category: 2,390,344,704 byte per day.
- Noise Monitoring category: 641,280,000 byte per day.
- Garbage Collection category: 360,000,000 byte per day.
- Parking Spot Monitoring category: 320,000,000 byte per day.
- Urban Lab Monitoring category: 4,723,200,000 byte per day.

We realized that each category of information produced huge amount of the redundant data in every transaction (by different sensors) and per day. Therefore, we monitored a single day of data generation in Sentilo platform. Then, we observed that the redundant data for energy, noise, garbage, parking and urban is almost 50%, 75%, 70%, 40%, and 30% respectively. Therefore, we have an initial thought to absorb this amount of data through the layers of our F2C data management architecture.

As we mentioned in the previous section, we applied two different techniques to reduce the number of data transfer among F2C layers. First, we used data aggregation techniques to eliminate the number of redundant data in the layer. Second, we used data compression techniques (for example, one solution proposed by PKWARE [17]) to compress data size after applying aggregation techniques. Note that data aggregation and data compression techniques can be implemented in each layer of F2C data architecture. Therefore, in the following paragraph we will explain how much data will be reduced at each layer.

The fog device is in the first level of our aggregation model. With respect to number of the redundant observation, we calculated that sensors data would be reduced to 1,194,834,432 bytes for energy monitoring, 160,320,000 bytes for noise monitoring, 108,000,000 bytes for garbage collection, 192,000,000 bytes for parking spot monitoring, and 3,306,240,000 bytes for urban lab as shown in Table I and Fig. 3 (blue lines). Then, the amount of data can be further decreased to smaller sizes through data compression. Therefore, the data size will be 262,863,575 bytes for the energy monitoring, 35,270,400 bytes for the noise monitoring, 23,760,000 bytes for the garbage collection, 42,240,000 bytes for the parking, and 727,372,800 bytes for the urban lab as shown in Table I and Fig.3 (green lines).

Similarly, the fog leaders at Fog-Layer-2 (city sections) play a second level for performing data aggregation and data compression techniques. Therefore, the data volume will be reduced to 597,586,176 bytes for energy monitoring, 39,498,840 bytes for noise monitoring, 32,462,370 bytes for garbage collection, 115,106,400 bytes for parking spot monitoring, and 2,318,823,158 bytes for urban lab as shown in Table I and Fig. 3 (blue lines). Then, the number of data can be shifted to smaller sizes through data compression. Therefore, the data size will go to 131,468,959 bytes for the energy monitoring, 8,689,745 bytes for the noise monitoring, 7,141,721 bytes for garbage collection, 25,323,408 bytes for the parking, and 510,141,095 bytes for the urban lab as shown in Table I and Fig. 3 (green lines).

Next layer are the fog leaders at Fog-Layer-3 (city districts) to handle data compression and data compression techniques. In this layer, data size goes to 298,793,088 bytes for energy monitoring, 9,874,710 bytes for noise monitoring, 9,738,711 bytes for garbage collection, 69,063,840 bytes for parking spot monitoring, and 1,623,176,211 bytes for urban lab after handling data aggregation as shown in Table I and Fig. 3 (blue lines). Then, the number of data can be further reduced through data compression. Therefore, the data size will change to 65,734,479 bytes for the energy monitoring, 2,172,436 bytes for the noise monitoring, 2,142,516 bytes for garbage collection, 15,194,045 bytes for the parking, and 35,098,766 bytes for the urban lab as shown in Table I and Fig. 3 (green lines).

After these computations, we conclude that: i) aggregation efficiency rate at fog devices, fog leaders in Fog-Layer-2 (city sections), and fog leader in Fog-Layer-3 (city districts): is almost 49.98%, 50.01%, and 50.08% efficiency rate for energy monitoring information. Similarly, the noise monitoring efficiency rate is 24.96%, 25.02%, and 25.05%. Then, the garbage collection rate is 29.99%, 30.05%, and 30.08%. In addition, the parking spot monitoring rate is 59.95%, 59.99%, and 60.01%. Indeed, the urban lab-monitoring rate is 68.93%, 69.13%, and 70.01%. ii) compression efficiency rate: is almost 22% for all layers (including Fog device, Fog-Leader (in Fog-Layer-2 layer), and Fog-Leader (in Fog-Layer-3 layer)).

Indeed, the total efficiency rate (including redundant data elimination and compression) at fog devices, fog leaders in Fog-Layer-2 (city sections), and fog leader in Fog-Layer-3 (city districts) is as shown in below:

- Energy Monitoring category: About 71.98%, 72.01%, and 72.08%.
- Noise Monitoring category: Almost 46.96%, 47.02%, and 47.05%.
- Garbage Collection category: Approximately 51.99%, 52.05%, and 52.08%.
- Parking Spot Monitoring category: Almost 81.95%, 81.99%, and 82.01%.
- Urban Lab Monitoring category: About 90.93%, 91.13%, and 92.01%.

V. Conclusion

In this paper, we have presented a novel architecture for hierarchal distributed data management in smart cities based on a distributed hierarchical fog to cloud resources management system. The advantages of this architecture are numerous. The most obvious advantage is that high computing and storage capabilities from the cloud layer can be combined with reduced network traffic and communication latencies from the fog layers, while enhancing fault tolerance and security and privacy protection. However, by providing such a hierarchical and distributed model, some interesting additional advantages arise:

- Real-time data accesses are much faster than in a centralized architecture;
- The network load is drastically reduced because many data can be accessed and used locally;

- Several aggregation techniques can easily be applied to further reduce the volume of data to be transferred through the network;
- The data transmission frequency can be adjusted in order to use the network in periods of low traffic;
- The data collection frequency from sensors can be increased at no additional cost, thus allowing higher precision and accuracy.

We have also explored the effectiveness of this architecture by exploring two basic data aggregation techniques, which are redundant data elimination and data compression, and compared to a real cloud based system from the smart city of Barcelona. We have shown by applying redundant data elimination that, in some cases, the data reduction rate reaches 75%. Additionally, by applying data compression, the data reduction rate increases to up to 78%. Finally, we have explored that the total efficiency rate, by applying both redundant data elimination and data compression, moves to almost 92%, in some cases. Although many other data aggregation techniques could be easily applied in this architecture, these two basic techniques are enough to illustrate the facility and effectiveness of such optimizations in our model.

As part of our future work we will explore more options related to data aggregation, and continue developing other data life cycle phases of our model, including data quality, data processing, data analysis, data storage, and data dissemination.

ACKNOWLEDGMENTS

This work has been supported by the Spanish Ministry of Economy and Competitiveness and by the European Regional Development Fund, under contract TEC2015-66220-R (MINECO/FEDER) and by the Catalan Government under contract 2014SGR371 and FI-DGR scholarship 2015FI_B100186.

REFERENCES

[1] A. Sinaeepourfard, J. Garcia, X. Masip-Bruin, E. Marín-Tordera, J. Cirera, G. Grau, et al., "Estimating Smart City sensors data generation," in The 15th IFIP Annual Mediterranean Ad Hoc Networking Workshop, 2016, pp. 1-8.

[2] H. Hu, Y. Wen, T.-S. Chua, and X. Li, "Toward scalable systems for big data analytics: A technology tutorial," IEEE Access, vol. 2, pp. 652-687, 2014.

[3] S. Henry, S. Hoon, M. Hwang, D. Lee, and M. D. DeVore, "Engineering trade study: extract, transform, load tools for data migration," in IEEE Conference on Design Symposium, Systems and Information Engineering, 2005, pp. 1-8.

[4] A. Levitin and T. Redman, "A model of the data (life) cycles with application to quality," Journal of Information and Software Technology on Elsevier, vol. 35, pp. 217-223, 1993.

[5] A. Sinaeepourfard, J. Garcia, X. Masip-Bruin, E. Marin-Tordera, X. Yin, and C. Wang, "A data lifeCycle model for smart cities," in IEEE International Conference on ICT Convergence (ICTC), 2016, pp. 400-405.

[6] J. Jin, J. Gubbi, S. Marusic, and M. Palaniswami, "An information framework for creating a smart city through internet of things," IEEE Internet of Things Journal, vol. 1, pp. 112-121, 2014.

[7] A. Sinaeepourfard, J. Garcia, X. Masip-Bruin, and E. Marín-Torder, "Towards a comprehensive data lifecycle model for big data environments," in The 3rd IEEE/ACM International Conference on Big Data Computing, Applications and Technologies 2016 (BDCAT), 2016, pp. 100-106.

[8] M. M. Rathore, A. Ahmad, A. Paul, and S. Rho, "Urban planning and building smart cities based on the Internet of Things using Big Data analytics," Journal of Computer Networks on Elsevier, vol. 101, pp. 63-80, 2016.

[9] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the internet of things," in Proceedings of the first edition of the MCC workshop on Mobile cloud computing, 2012, pp. 13-16.

[10] T. V. N. Rao, A. Khan, M. Maschendra, and M. K. Kumar, "A Paradigm Shift from Cloud to Fog Computing," International Journal of Science, Engineering and Computer Technology, vol. 5, p. 385, 2015.

[11] X.Masip, E.Marín, A.Jukan, G.J.Ren, and G.Tashakor, "Foggy clouds and cloudy fogs: a real need for coordinated management of fog-to-cloud (F2C) computing systems," Jounal of IEEE Wireless Communications Magazine, 2016.

[12] P. Patil and U. Kulkarni, "Delay Efficient Distributed Data Aggregation Algorithm in Wireless Sensor Networks," International Journal of Computer Applications, vol. 69, 2013.

[13] N. Karthick and X. A. Kalrani, "A Survey on Data Aggregation in Big Data and Cloud Computing," International Journal of Computer Trends and Technology (IJCTT), vol. 17, pp. 28-32, 2014.

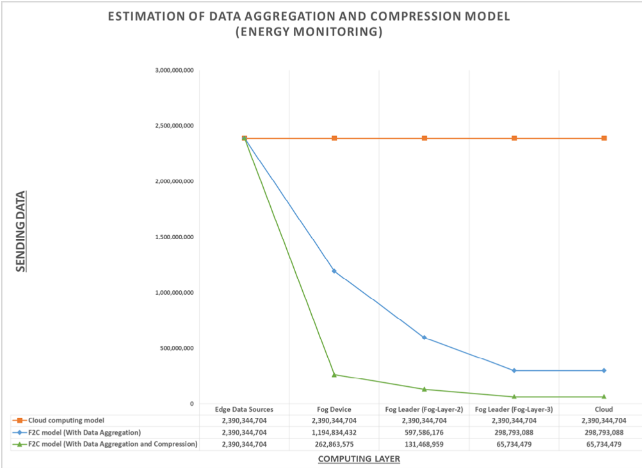
[14] J.-Y. Chen, G. Pandurangan, and D. Xu, "Robust computation of aggregates in wireless sensor networks: distributed randomized algorithms and analysis," IEEE Transactions on Parallel and Distributed Systems, vol. 17, pp. 987-1000, 2006.

[15] P. Jesus, C. Baquero, and P. S. Almeida, "A survey of distributed data aggregation algorithms," IEEE Communications Surveys & Tutorials, vol. 17, pp. 381-404, 2015.

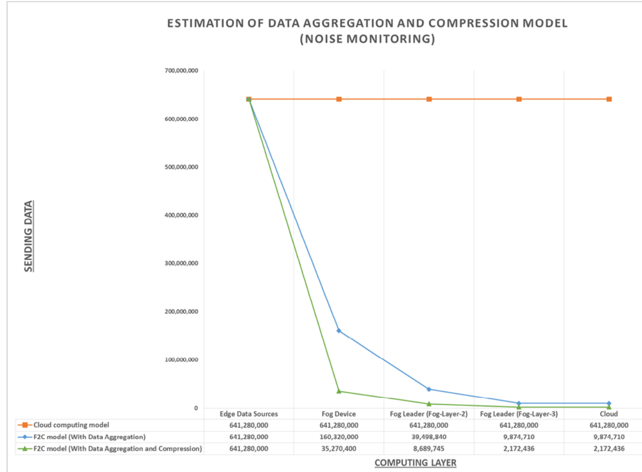
[16] T. He, L. Gu, L. Luo, T. Yan, J. A. Stankovic, and S. H. Son, "An overview of data aggregation architecture for real-time tracking with sensor networks," in 20th IEEE International Parallel & Distributed Processing Symposium, 2006, p. 8 pp.

[17] PKWARE.APPNOTE.
Available: https://support.pkware.com/display/PKZIP/APPNOTE.

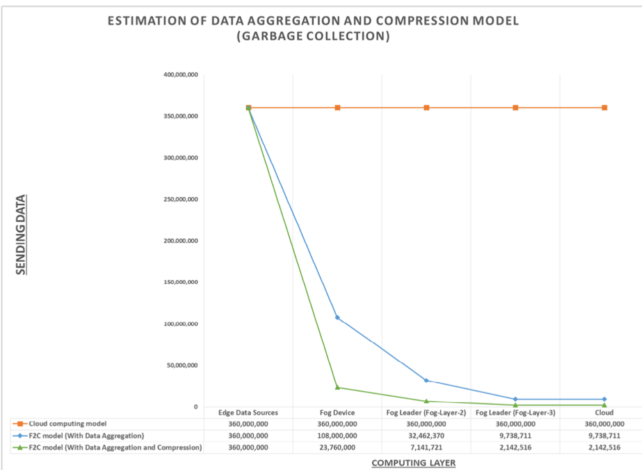
[18] Sentilo Platform.
Available: http://www.sentilo.io/wordpress/..



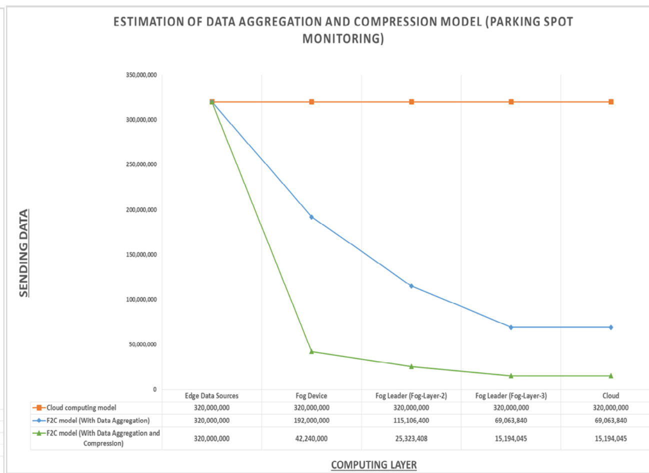
(a) Energy Monitoring category



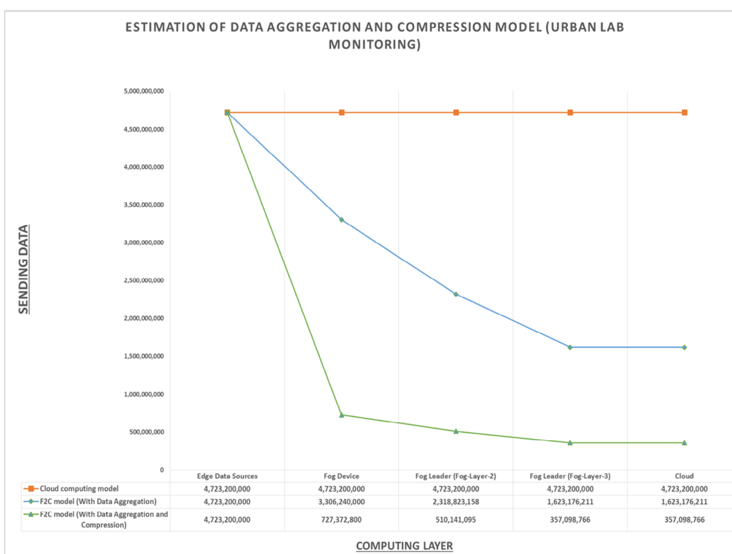
(b) Noise Monitoring category



(c) Garbage Collection category



(d) Parking Spot Monitoring category



(e) Urban Lab Monitoring

Fig. 3. Redundant data aggregation and data compression models in F2C data management architecture.

TABLE I. ESTIMATION OF REDUNDANT DATA AGGREGATION AND DATA COMPRESSION MODELS IN F2C DATA MANAGEMENT ARCHITECTURE

Category of inforamtion	Type of information	Number of sensors	Computing model	Data Filtering techniques	Sections of Barcelona				Districts of Barcelona										
					All				Ciutat Vella	Eixample	Sants-Montjuic	Les Corts	Sarria-Sant Gervasi	Gracia	Horta-Guinardo	Nou Barris	Sant Andreu	Sant Marti	Total Number of Fog Layer 3
					Fog Layer 1			Fog Layer 2	Fog Layer 3										
					Edge Data Sources	Fog Device	Fog-Leader	Fog -Leader	Fog-Leader										
Energy monitoring	Electricity meter	70,717	Cloud Computing	No Action	2,112	2,112	682,176	2,046,528	8,186,112	12,279,168	16,372,224	6,139,584	12,279,168	10,232,640	22,511,808	26,604,864	14,325,696	20,465,280	149,396,544
					149,354,304	149,354,304	149,396,544	149,396,544	8,186,112	12,279,168	16,372,224	6,139,584	12,279,168	10,232,640	22,511,808	26,604,864	14,325,696	20,465,280	149,396,544
			Fog Computing	Data Aggregation	2,112	1,056	170,544	255,816	1,023,264	1,534,896	2,046,528	767,448	1,534,896	1,279,080	2,813,976	3,325,608	1,790,712	2,558,160	18,674,568
					149,354,304	74,677,152	37,349,136	18,674,568	1,023,264	1,534,896	2,046,528	767,448	1,534,896	1,279,080	2,813,976	3,325,608	1,790,712	2,558,160	18,674,568
	External ambient conditions	70,717	Cloud Computing	No Action	2,112	2,112	682,176	2,046,528	8,186,112	12,279,168	16,372,224	6,139,584	12,279,168	10,232,640	22,511,808	26,604,864	14,325,696	20,465,280	149,396,544
					149,354,304	149,354,304	149,396,544	149,396,544	8,186,112	12,279,168	16,372,224	6,139,584	12,279,168	10,232,640	22,511,808	26,604,864	14,325,696	20,465,280	149,396,544
			Fog Computing	Data Aggregation	2,112	1,056	170,544	255,816	1,023,264	1,534,896	2,046,528	767,448	1,534,896	1,279,080	2,813,976	3,325,608	1,790,712	2,558,160	18,674,568
					149,354,304	74,677,152	37,349,136	18,674,568	1,023,264	1,534,896	2,046,528	767,448	1,534,896	1,279,080	2,813,976	3,325,608	1,790,712	2,558,160	18,674,568
	Internal ambient conditions	70,717	Cloud Computing	No Action	2,112	2,112	682,176	2,046,528	8,186,112	12,279,168	16,372,224	6,139,584	12,279,168	10,232,640	22,511,808	26,604,864	14,325,696	20,465,280	149,396,544
					149,354,304	149,354,304	149,396,544	149,396,544	8,186,112	12,279,168	16,372,224	6,139,584	12,279,168	10,232,640	22,511,808	26,604,864	14,325,696	20,465,280	149,396,544
			Fog Computing	Data Aggregation	2,112	1,056	170,544	255,816	1,023,264	1,534,896	2,046,528	767,448	1,534,896	1,279,080	2,813,976	3,325,608	1,790,712	2,558,160	18,674,568
					149,354,304	74,677,152	37,349,136	18,674,568	1,023,264	1,534,896	2,046,528	767,448	1,534,896	1,279,080	2,813,976	3,325,608	1,790,712	2,558,160	18,674,568
	Network analyzer	70,717	Cloud Computing	No Action	23,232	23,232	7,503,936	22,511,808	90,047,232	135,070,848	180,094,464	67,535,424	135,070,848	112,559,040	247,629,888	292,653,504	157,582,656	225,118,080	1,643,361,984
					1,642,897,344	1,642,897,344	1,643,361,984	1,643,361,984	90,047,232	135,070,848	180,094,464	67,535,424	135,070,848	112,559,040	247,629,888	292,653,504	157,582,656	225,118,080	1,643,361,984
			Fog Computing	Data Aggregation	23,232	11,616	1,875,984	2,813,976	11,255,904	16,883,856	22,511,808	8,441,928	16,883,856	14,069,880	30,953,736	36,581,688	19,697,832	28,139,760	205,420,248
					1,642,897,344	821,448,672	410,840,496	205,420,248	11,255,904	16,883,856	22,511,808	8,441,928	16,883,856	14,069,880	30,953,736	36,581,688	19,697,832	28,139,760	205,420,248
	Solar thermal installation	70,717	Cloud Computing	No Action	2,112	2,112	682,176	2,046,528	8,186,112	12,279,168	16,372,224	6,139,584	12,279,168	10,232,640	22,511,808	26,604,864	14,325,696	20,465,280	149,396,544
					149,354,304	149,354,304	149,396,544	149,396,544	8,186,112	12,279,168	16,372,224	6,139,584	12,279,168	10,232,640	22,511,808	26,604,864	14,325,696	20,465,280	149,396,544
			Fog Computing	Data Aggregation	2,112	1,056	170,544	255,816	1,023,264	1,534,896	2,046,528	767,448	1,534,896	1,279,080	2,813,976	3,325,608	1,790,712	2,558,160	18,674,568
					149,354,304	74,677,152	37,349,136	18,674,568	1,023,264	1,534,896	2,046,528	767,448	1,534,896	1,279,080	2,813,976	3,325,608	1,790,712	2,558,160	18,674,568
	Temperature	70,717	Cloud Computing	No Action	2,112	2,112	682,176	2,046,528	8,186,112	12,279,168	16,372,224	6,139,584	12,279,168	10,232,640	22,511,808	26,604,864	14,325,696	20,465,280	149,396,544
					149,354,304	149,354,304	149,396,544	149,396,544	8,186,112	12,279,168	16,372,224	6,139,584	12,279,168	10,232,640	22,511,808	26,604,864	14,325,696	20,465,280	149,396,544
			Fog Computing	Data Aggregation	2,112	1,056	170,544	255,816	1,023,264	1,534,896	2,046,528	767,448	1,534,896	1,279,080	2,813,976	3,325,608	1,790,712	2,558,160	18,674,568
					149,354,304	74,677,152	37,349,136	18,674,568	1,023,264	1,534,896	2,046,528	767,448	1,534,896	1,279,080	2,813,976	3,325,608	1,790,712	2,558,160	18,674,568
	Total number	424,302	Cloud Computing	No Action	2,390,344,704	2,390,344,704	2,390,344,704	2,390,344,704	2,390,344,704										
			Fog Computing	Data Aggregation	2,390,344,704	1,194,834,432	597,586,176	298,793,088	298,793,088										
	Total number	424,302	Cloud Computing	No Action	2,390,344,704	2,390,344,704	2,390,344,704	2,390,344,704	2,390,344,704										
			Fog Computing	Data Aggregation and Data Compression	2,390,344,704	262,863,575	131,468,959	65,734,479	65,734,479										
Noise monitoring	Noise	10,000	Cloud Computing	No Action	768	768	34,560	103,680	414,720	622,080	829,440	311,040	622,080	518,400	1,140,480	1,347,840	725,760	1,036,800	7,680,000
					7,680,000	7,680,000	7,680,000	7,680,000	414,720	622,080	829,440	311,040	622,080	518,400	1,140,480	1,347,840	725,760	1,036,800	7,680,000
			Fog Computing	Data Aggregation	768	192	2,160	1,620	6,480	9,720	12,960	4,860	9,720	8,100	17,820	21,060	11,340	16,200	118,260
					7,680,000	1,920,000	473,040	118,260	6,480	9,720	12,960	4,860	9,720	8,100	17,820	21,060	11,340	16,200	118,260
		10,000	Cloud Computing	No Action	31,680	31,680	1,425,600	4,276,800	17,107,200	25,660,800	34,214,400	12,830,400	25,660,800	21,384,000	47,044,800	55,598,400	29,937,600	42,768,000	316,800,000
					316,800,000	316,800,000	316,800,000	316,800,000	17,107,200	25,660,800	34,214,400	12,830,400	25,660,800	21,384,000	47,044,800	55,598,400	29,937,600	42,768,000	316,800,000
			Fog Computing	Data Aggregation	31,680	7,920	89,100	66,825	267,300	400,950	534,600	200,475	400,950	334,125	735,075	868,725	467,775	668,250	4,878,225
					316,800,000	79,200,000	19,512,900	4,878,225	267,300	400,950	534,600	200,475	400,950	334,125	735,075	868,725	467,775	668,250	4,878,225
		10,000	Cloud Computing	No Action	31,680	31,680	1,425,600	4,276,800	17,107,200	25,660,800	34,214,400	12,830,400	25,660,800	21,384,000	47,044,800	55,598,400	29,937,600	42,768,000	316,800,000
					316,800,000	316,800,000	316,800,000	316,800,000	17,107,200	25,660,800	34,214,400	12,830,400	25,660,800	21,384,000	47,044,800	55,598,400	29,937,600	42,768,000	316,800,000
			Fog Computing	Data Aggregation	31,680	7,920	89,100	66,825	267,300	400,950	534,600	200,475	400,950	334,125	735,075	868,725	467,775	668,250	4,878,225
					316,800,000	79,200,000	19,512,900	4,878,225	267,300	400,950	534,600	200,475	400,950	334,125	735,075	868,725	467,775	668,250	4,878,225
	Total number	424,302	Cloud Computing	No Action	641,280,000	641,280,000	641,280,000	641,280,000	641,280,000										
			Fog Computing	Data Aggregation	641,280,000	160,320,000	39,498,840	9,874,710	9,874,710										
	Total number	424,302	Cloud Computing	No Action	641,280,000	641,280,000	641,280,000	641,280,000	641,280,000										
			Fog Computing	Data Aggregation and Data Compression	641,280,000	35,270,400	8,689,745	2,172,436	2,172,436										
Garbage Collection	Container glass	40,000	Cloud Computing	No Action	1,800	1,800	329,400	988,200	3,952,800	5,929,200	7,905,600	2,964,600	5,929,200	4,941,000	10,870,200	12,846,600	6,917,400	9,882,000	72,000,000
					72,000,000	72,000,000	72,000,000	72,000,000	3,952,800	5,929,200	7,905,600	2,964,600	5,929,200	4,941,000	10,870,200	12,846,600	6,917,400	9,882,000	72,000,000
			Fog Computing	Data Aggregation	1,800	540	29,646	26,681	106,726	160,088	213,451	80,044	160,088	133,407	293,495	346,858	186,770	266,814	1,947,742
					72,000,000	21,600,000	6,492,474	1,947,742	106,726	160,088	213,451	80,044	160,088	133,407	293,495	346,858	186,770	266,814	1,947,742
	Container organic	40,000	Cloud Computing	No Action	1,800	1,800	329,400	988,200	3,952,800	5,929,200	7,905,600	2,964,600	5,929,200	4,941,000	10,870,200	12,846,600	6,917,400	9,882,000	72,000,000
					72,000,000	72,000,000	72,000,000	72,000,000	3,952,800	5,929,200	7,905,600	2,964,600	5,929,200	4,941,000	10,870,200	12,846,600	6,917,400	9,882,000	72,000,000
			Fog Computing	Data Aggregation	1,800	540	29,646	26,681	106,726	160,088	213,451	80,044	160,088	133,407	293,495	346,858	186,770	266,814	1,947,742
					72,000,000	21,600,000	6,492,474	1,947,742	106,726	160,088	213,451	80,044	160,088	133,407	293,495	346,858	186,770	266,814	1,947,742
	Container paper	40,000	Cloud Computing	No Action	1,800	1,800	329,400	988,200	3,952,800	5,929,200	7,905,600	2,964,600	5,929,200	4,941,000	10,870,200	12,846,600	6,917,400	9,882,000	72,000,000
					72,000,000	72,000,000	72,000,000	72,000,000	3,952,800	5,929,200	7,905,600	2,964,600	5,929,200	4,941,000	10,870,200	12,846,600	6,917,400	9,882,000	72,000,000
			Fog Computing	Data Aggregation	1,800	540	29,646	26,681	106,726	160,088	213,451	80,044	160,088	133,407	293,495	346,858	186,770	266,814	1,947,742
					72,000,000	21,600,000	6,492,474	1,947,742	106,726	160,088	213,451	80,044	160,088	133,407	293,495	346,858	186,770	266,814	1,947,742
	Container plastic	40,000	Cloud Computing	No Action	1,800	1,800	329,400	988,200	3,952,800	5,929,200	7,905,600	2,964,600	5,929,200	4,941,000	10,870,200	12,846,600	6,917,400	9,882,000	72,000,000
					72,000,000	72,000,000	72,000,000	72,000,000	3,952,800	5,929,200	7,905,600	2,964,600	5,929,200	4,941,000	10,870,200	12,846,600	6,917,400	9,882,000	72,000,000
			Fog Computing	Data Aggregation	1,800	540	29,646	26,681	106,726	160,088	213,451	80,044	160,088	133,407	293,495	346,858	186,770	266,814	1,947,742
					72,000,000	21,600,000	6,492,474	1,947,742	106,726	160,088	213,451	80,044	160,088	133,407	293,495	346,858	186,770	266,814	1,947,742
	Container refuse	40,000	Cloud Computing	No Action	1,800	1,800	329,400	988,200	3,952,800	5,929,200	7,905,600	2,964,600	5,929,200	4,941,000	10,870,200	12,846,600	6,917,400	9,882,000	72,000,000
					72,000,000	72,000,000	72,000,000	72,000,000	3,952,800	5,929,200	7,905,600	2,964,600	5,929,200	4,941,000	10,870,200	12,846,600	6,917,400	9,882,000	72,000,000
			Fog Computing	Data Aggregation	1,800	540	29,646	26,681	106,726	160,088	213,451	80,044	160,088	133,407	293,495	346,858	186,770	266,814	1,947,742
					72,000,000	21,600,000	6,492,474	1,947,742	106,726	160,088	213,451	80,044	160,088	133,407	293,495	346,858	186,770	266,814	1,947,742
	Total number	424,302	Cloud Computing	No Action	360,000,000	360,000,000	360,000,000	360,000,000	360,000,000										
			Fog Computing	Data Aggregation	360,000,000	108,000,000	32,462,370	9,738,711	9,738,711										
	Total number	200,000	Cloud Computing	No Action	360,000,000	360,000,000	360,000,000	360,000,000	360,000,000										
			Fog Computing	Data Aggregation and Data Compression	360,000,000	23,760,000	7,141,721	2,142,516	2,142,516										
Parking Spot	Parking	80,000	Cloud Computing	No Action	4,000	4,000	1,460,000	4,380,000	17,520,000	26,280,000	35,040,000	13,140,000	26,280,000	21,900,000	48,180,000	56,940,000	30,660,000	43,800,000	320,000,000
					320,000,000	320,000,000	320,000,000	320,000,000	17,520,000	26,280,000	35,040,000	13,140,000	26,280,000	21,900,000	48,180,000	56,940,000	30,660,000	43,800,000	320,000,000
			Fog Computing	Data Aggregation	1,800	2,400	525,600	946,080	3,784,320	5,676,480	7,568,640	2,838,240	5,676,480	4,730,400	10,406,880	12,299,040	6,622,560	9,460,800	69,063,840
					320,000,000	192,000,000	115,106,400	69,063,840	3,784,320	5,676,480	7,568,640	2,838,240	5,676,480	4,730,400	10,406,880	12,299,040	6,622,560	9,460,800	69,063,840
	Total number	200,000	Cloud Computing	No Action	320,000,000	320,000,000	320,000,000	320,000,000	320,000,000										
			Fog Computing	Data Aggregation	320,000,000	192,000,000	115,106,400	69,063,840	69,063,840										
	Total number	80,000	Cloud Computing	No Action	320,000,000	320,000,000	320,000,000	320,000,000	320,000,000										
			Fog Computing	Data Aggregation and Data Compression	320,000,000	42,240,000	25,323,408	15,194,045	15,194,045										
Urban Lab monitoring	Air quality	40,000	Cloud Computing	No Action	13,824	13,824	2,529,792	7,589,376	30,357,504	45,536,256	60,715,008	22,768,128	45,536,256	37,946,880	83,483,136	98,661,888	53,125,632	75,893,760	552,960,000
					552,960,000	552,960,000	552,960,000	552,960,000	30,357,504	45,536,256	60,715,008	22,768,128	45,536,256	37,946,880	83,483,136	98,661,888	53,125,632	75,893,760	552,960,000
			Fog Computing	Data Aggregation	13,824	9,677	1,239,598	2,603,156	10,412,624	15,618,936	20,825,248	7,809,468	15,618,936	13,015,780	28,634,716	33,841,028	18,222,092	26,031,560	190,030,386
					552,960,000	387,072,000	271,471,980	190,030,386	10,412,624	15,618,936	20,825,248	7,809,468	15,618,936	13,015,780	28,634,716	33,841,028	18,222,092	26,031,560	190,030,386
	Bicycle flow	40,000	Cloud Computing	No Action	3,168	3,168	579,744	1,739,232	6,956,928	10,435,392	13,913,856	5,217,696	10,435,392	8,696,160	19,131,552	22,610,016	12,174,624	17,392,320	126,720,000
					126,720,000	126,720,000	126,720,000	126,720,000	6,956,928	10,435,392	13,913,856	5,217,696	10,435,392	8,696,160	19,131,552	22,610,016	12,174,624	17,392,320	126,720,000
			Fog Computing	Data Aggregation	3,168	2,218	284,075	596,557	2,386,226	3,579,339	4,772,453	1,789,670	3,579,339	2,982,783	6,562,122	7,755,235	4,175,896	5,965,566	43,548,630
					126,720,000	88,704,000	62,212,329	43,548,630	2,386,226	3,579,339	4,772,453	1,789,670	3,579,339	2,982,783	6,562,122	7,755,235	4,175,896	5,965,566	43,548,630
	People flow	40,000	Cloud Computing	No Action	3,168	3,168	579,744	1,739,232	6,956,928	10,435,392	13,913,856	5,217,696	10,435,392	8,696,160	19,131,552	22,610,016	12,174,624	17,392,320	126,720,000
					126,720,000	126,720,000	126,720,000	126,720,000	6,956,928	10,435,392	13,913,856	5,217,696	10,435,392	8,696,160	19,131,552	22,610,016	12,174,624	17,392,320	126,720,000
			Fog Computing	Data Aggregation	3,168	2,218	284,075	596,557	2,386,226	3,579,339	4,772,453	1,789,670	3,579,339	2,982,783	6,562,122	7,755,235	4,175,896	5,965,566	43,548,630
					126,720,000	88,704,000	62,212,329	43,548,630	2,386,226	3,579,339	4,772,453	1,789,670	3,579,339	2,982,783	6,562,122	7,755,235	4,175,896	5,965,566	43,548,630
	Traffic	40,000	Cloud Computing	No Action	63,360	63,360	11,594,880	34,784,640	139,138,560	208,707,840	278,277,120	104,353,920	208,707,840	173,923,200	382,631,040	452,200,320	243,492,480	347,846,400	2,534,400,000
					2,534,400,000	2,534,400,000	2,534,400,000	2,534,400,000	139,138,560	208,707,840	278,277,120	104,353,920	208,707,840	173,923,200	382,631,040	452,200,320	243,492,480	347,846,400	2,534,400,000
			Fog Computing	Data Aggregation	63,360	44,352	5,681,491	11,931,132	47,724,526	71,586,789	95,449,052	35,793,395	71,586,789	59,655,658	131,242,447	155,104,710	83,517,921	119,311,315	870,972,601
					2,534,400,000	1,774,080,000	1,244,246,573	870,972,601	47,724,526	71,586,789	95,449,052	35,793,395	71,586,789	59,655,658	131,242,447	155,104,710	83,517,921	119,311,315	870,972,601
	Weather	40,000	Cloud Computing	No Action	34,560	34,560	6,324,480	18,973,440	75,893,760	113,840,640	151,787,520	56,920,320	113,840,640	94,867,200	208,707,840	246,654,720	132,814,080	189,734,400	1,382,400,000
					1,382,400,000	1,382,400,000	1,382,400,000	1,382,400,000	75,893,760	113,840,640	151,787,520	56,920,320	113,840,640	94,867,200	208,707,840	246,654,720	132,814,080	189,734,400	1,382,400,000
			Fog Computing	Data Aggregation	34,560	24,192	3,098,995	6,507,890	26,031,560	39,047,340	52,063,119	19,523,670	39,047,340	32,539,450	71,586,789	84,602,569	45,555,229	65,078,899	475,075,964
					1,382,400,000	967,680,000	678,679,949	475,075,964	26,031,560	39,047,340	52,063,119	19,523,670	39,047,340	32,539,450	71,586,789	84,602,569	45,555,229	65,078,899	475,075,964
	Total number	80,000	Cloud Computing	No Action	4,723,200,000	4,723,200,000	4,723,200,000	4,723,200,000	4,723,200,000										
			Fog Computing	Data Aggregation	4,723,200,000	3,306,240,000	2,318,823,158	1,623,176,211	1,623,176,211										
		200,000	Cloud Computing	No Action	4,723,200,000	4,723,200,000	4,723,200,000	4,723,200,000	4,723,200,000										
			Fog Computing	Data Aggregation and Data Compression	4,723,200,000	727,372,800	510,141,095	357,098,766	357,098,766										
Total number		1,552,906	Cloud Computing	No Action	8,434,824,704	8,434,824,704	8,434,824,704	8,434,824,704	8,434,824,704										
			Fog Computing	Data Aggregation	8,434,824,704	2,622,834,432	1,463,333,735	862,546,313	413,501,909										
		1,328,604	Cloud Computing	No Action	8,434,824,704	8,434,824,704	8,434,824,704	8,434,824,704	8,434,824,704										
			Fog Computing	Data Aggregation and Data Compression	8,434,824,704	1,091,506,775	682,764,928	442,342,243	442,342,243										