Location Intelligence Application in Digital Data Activity Dimensioning in Smart Cities

Jensen, Michael; Gutierrez Lopez, Jose Manuel; Pedersen, Jens Myrup

Published in:
Procedia Computer Science

DOI (link to publication from Publisher):
10.1016/j.procs.2014.09.015

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 36 (2014) 418 – 424

Procedia Computer Science

Complex Adaptive Systems, Publication 4
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2014- Philadelphia, PA

Location Intelligence Application in Digital Data Activity Dimensioning in Smart Cities

Michael Jensen, Jose Gutierrez*, Jens Pedersen

Department of Electronic Systems, Aalborg University, Fredrik Bajers vej 7, 9220 Aalborg Ø, Denmark

Abstract

Cities around the globe are now following the "smart" trend. Smart cities are complex systems of systems that rely on IT to improve their efficiency in terms of economics or sustainability. Many of the activities involved in this context require data to be transferred via communication networks. However, due to the heterogeneity of the applications involved, different nature traffic patterns and number of connected elements (humans or machines), these data traffic flows become extremely complex to model. This work focuses on the application of location intelligence to simplify the modeling of data networks' activity in such complex systems. It is essential to describe and understand when and where data is generated to effectively design, plan and manage communication networks. We introduce a model based on traffic generation rules and patterns to be applied over Geographic Information Systems data to create "dynamic data activity heat maps". These heat maps provide a spatial-temporal overview of the behavior of network data in cities as a whole. In addition, we illustrate the model's application to a specific geographic area of relevance.

Peer-review under responsibility of scientific committee of Missouri University of Science and Technology
Keywords: Smart cities; GIS; Location Intelligence; Data traffic estimation.

1. Introduction

Now, in the year 2014, around 50% of the population in the world is living in cities, and is expected to increase even more in the near future [1]. In addition, the trend of making these cities “smarter” opens up for a whole new perspective on digitalization when integrating common city infrastructures and services with ICT systems.

There is no consensus regarding the definition of “Smart City”. The points of view range from sociological [2] to Information and Communication Technologies (ICT) [3] passing by urban perspectives [4]. However, it is clear that one of the main points in common is communication networks as a digitalization enabler. Thus, in this work we will adopt the definition presented in [5] as cities that integrate a digital infrastructure with the physical city to improve, among other aspects, environmental impact, quality of life and economic growth. Devices or machines are physically located in cities, people and vehicles move around cities, hence the

**Corresponding author:+45 99408747, E-mail address:jgl@es.aau.dk*

Peer-review under responsibility of scientific committee of Missouri University of Science and Technology
doi:10.1016/j.procs.2014.09.015

relation between data traffic generated by them and their associated physical location is a key feature to take into account when planning and/or managing communication networks in cities.

The types of digital services within the smart city concept are very diverse and the number of devices communicating with each other has been exponentially increasing for the past few years; this is commonly known as Internet of Things [6]. Some devices transmit a few bytes of information at concrete times such as induction loop detectors for vehicle counting [7], while others generate heavy continuous data streams, such as CCTV surveillance [8]. Therefore, smart cities are conceptually understood as systems of systems where the network infrastructure must handle different types of traffic [1], [9], [10]. Consequently, to create a global overview of cities' data activity could become overwhelming due to its volume and heterogeneity.

There is a significant amount of technical research on individual services that could be labelled within the smart city context. However, there is a need for unified urban information models [4] and [9], that can systematically visualize the digital activity in order to manage such a complex system.

Location intelligence may be useful to understand the interaction of data generating elements in relation to their geographical location in cities. Geographical Information Systems (GIS) combined with models abstracting the activity and mobility of data traffic sources as a function of well know city elements (households, roads....) may provide the appropriate spatio-temporal framework to approach the problem. In this way, dynamic data activity "heat maps" can be created to be able to visualize and understand the urban digital ecosystem. These maps consist of a set of overlapping layers, each representing the data traffic activity of a different service or application. The combination of these layers may provide an overview of the data activity to be used for different purposes, ranging from network planning to automated utility management systems.

A key aspect of using GIS data is not only the geographical location of specific traffic activities but also the temporal information associated to them, as the location and activity volume of users/devices may vary depending on the time of the day. Thus, we propose to use a spatial-temporal-activity GIS framework to create dynamic data traffic activity maps.

Summarizing all the concepts described above, the main objective of this paper is to create a framework to be able to quantify and visualize the data traffic activity. The introduced methods within this framework allow us to create heat maps based on basic models and GIS data extrapolation to know where and when digital activity is generated. To the best of our knowledge, such a framework and methods have been never been proposed in the relevant literature. In addition, a case study is carried out on a real geographical area using GIS data combined with simple mobility and service models, to illustrate the application of the methods and framework.

The rest of the document is structured as follows: Section 2 summarizes important definitions and studies in relation to this work. Section 3 presents the framework and methodology proposed and describes the followed models. Section 4 presents an illustrative case study of how the methods may be applied using a real geographic area. In addition, it elaborates on the potential application of the framework. Finally, Section 5 highlights the most relevant conclusions of this work.

2. Background

2.1 Smart Cities and GIS

Smart cities can be defined from many different points of view and their development has been highlighted as beneficial and/or enabler of very diverse services and applications. Examples are the following: Infrastructure monitoring and management in [10], [11], and [12], energy efficiency and operational expenditure (OPEX) savings in power systems in [3] and [11], green sustainability in [5] and [13], urban development [6], citizens' quality of life improvement in [3] and [5], cities' economic growth in [5], health care systems design in [13], or intelligent transport systems (ITS) management in [13].

GIS have been used with very diverse purposes and applications, and play a significant role in urban and infrastructure planning. More concretely in relation to the theme of this work, examples of studies using GIS in relation to smart cities are: A service oriented architecture infrastructure management based on GIS information is proposed in [11]; an illustration of the use of GIS in flood management systems in Rio de Janeiro is shown in [1]; and GIS visualization tools are highlighted to play a key role in decision support and data analysis [7].

2.2 Definitions

Raw GIS data: In this work, it refers to basic GIS data of buildings (as points represented by Cartesian coordinates) and streets (lines that are subdivided into a set of points) . These points are referred to in this document as "Basic Points, *BP*".

Spatio-temporal data: Generically, it is defined as any set of data characterized by both a geographical location and a time instant or time frame. For example, the movement of a person during a day can be described by its position in different reference times. In this work, spatio-temporal data consists of geographical locations where data the volume varies in time.

Location Intelligence: Generically, it is defined as the set of tools and methods that interrelates geographic information to business data, to identify patterns and relationships for decision-making that otherwise may be complex to operate with without a spatial representation [14]. We extrapolate this concept to data networks where geographical information is combined with network models for better understanding the digital activity dynamics in cities. This combination provides a set of methods, tools and data for building a framework to analytically or empirically design, evaluate, and improve cities' communication networks carrying the data traffic generated within the "smart" evolution context.

Data activity: It refers to the digital data transmitted by devices within the smart city applications such as management of critical infrastructure, surveillance, etc. In addition, only internet traffic related to exclusively these applications is considered, in the context that it is extracted and used within the city limits. Common data traffic generated by users when accessing common internet services is out of the scope of this study but can easily be added to the presented framework.

Data point: In this work it refers to the entities that generate digital data. Three types of point are considered:

1. People: Set of points related to the services implying data generated by people such as crowdsourcing.
2. Machines: Set of points related to machines communicating with (usually) other machines. Examples of services used by these entities are Smart Grid metering or road traffic monitoring.
3. Vehicles: Set of points that correspond to both the combination of mobile machines and people or just mobile machines.

Access Network: This work is focused on the data activity generated by devices regardless of the type of access network they are connected to. To avoid having any limitation regarding the network, we assume in the case study that the city's access network is capable of handling all data transactions, in order to create the heat-maps.

3. Methodology

3.1 Method overview

The basic idea is to model the behavior of the aforementioned entities and relate these models to the available GIS data, the Basic Points *BPs*, instead of individually tracking or monitoring them. In this way, keeping and maintaining huge amounts of information passively or actively acquired by individual devices is avoided. The procedure is to abstract the information from these individual entities and the digital services used to be able to relate it to spatial-temporal data activity information and *BPs*. Consequently, temporal heat-maps can be created in order to analyze the behavior of such a complex system.

The idea proposed in this work is based on the quantification of the data volume that is being generated for any time period and at any point in the city, regardless of the entity and application used. The methodology is structured into three basic modelling levels, as illustrated in Fig. 1(a).

Entity Level: Behavior modelling of the entities independently from the services used. It is related to their position in time. Examples of mobility modelling can be found in [15]. As a simplistic example of how to model the location and mobility of entities, a person has a higher probability of being at home at night than during working hours.

Service Level: Modelling of the different services that are accessed by the entities and the use they make of them. An example of a specific service (e-Governance) modelling can be found in [16], however user behavior patterns are very dependent on local factors, and probably these need to be extracted from data gathered locally or regionally. Nevertheless, in a similar way as for the entity level modelling, the use of services (frequency, data volume generated, etc…) may be characterized by time- and/or location-dependent probability distributions.

Geocoding Level: Assignment of the information derived from the combination of entity level and service level modelling to available GIS data (Basic Points, *BPs*). For example, in relation to Smart Grid, if a smart meter sends *X* kbytes every *Y* minutes and there is one meter per household, then it is possible to relate the data activity to a GIS *BP* (household in this case). Consequently, the models can be combined overlapping each other to estimate the data activity for each GIS element involved in the evaluation (in this work households and streets). Fig. 1(b) graphically illustrates the data contribution of different services and entities to *BPs*. In addition, to provide a uniform visualization layout, the area under study can be discretized into grid-cells where the data activity of each cell is the sum of the data volume of all the contained *BPs* at a given time period.

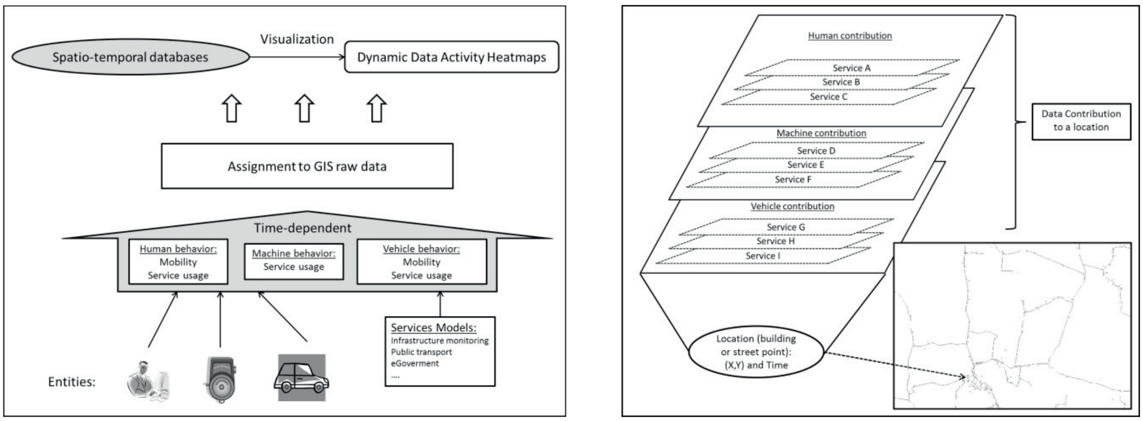


Fig 1. (a) Methodology overview; (b) Data contribution to *BPs*.

3.2 GIS data format

The GIS dataset used is related to and derived from two basic city elements: buildings and streets. Each database used has two parts, raw data which serves as an input to model the behaviour of the involved entities, and resulting data which is the outcome of the combination of the raw data with service models and probability distributions resulting in temporal data activity estimation.

While buildings can be directly abstracted into individual points to work with, streets (lines) must be converted into a collection points representing a certain street segment (e.g. one street segment of 150 m can be divided into 10 points 15 m apart and placed appropriately along the segment line). The format of the databases used are defined as follows:

- Building database:
  - Raw data fields: Identifier (*ID*), coordinates (*X* and *Y*), building type (household, business,...) (*Type*), inhabitants (Living or working) (*Pop*), and population density at location (*Dens*).
  - Resulting data: Time period identifier (*T*), current population (*CPop*), data contribution of service 1 (D_1), data contribution of service 2 (D_2) …, data contribution of service *n* (D_n).

- Street point database:
  - Raw data fields: Identifier (*ID*), coordinates (*X* and *Y*), degree (*Deg*), Street type (Urban, rural,…) (*Type*), and population density at location (*Dens*).
  - Resulting data: Time period identifier (*T*), current population (*CPop*), data contribution of service 1 (D_1), data contribution of service 2 (D_2) …, data contribution of service *n* (D_n).

3.3 Services and Mobility Models

The proposed framework can be indistinctly used for the purpose of evaluating effects of existing services that can be modelled after empirical performance data or future services modelled after design and theoretical analyses. In any of these two cases, essential information about each service is: a) who is involved in the data activity (type of entity), b) from where (Basic Point) and c) data volume in a predefined time period.

In this work, the mobility of isolated entities is not individually modelled. Instead, entities are assigned to *BPs* using probability functions describing how likely it is that entities (vehicles or people) are located at a certain point, and allowing to probabilistically calculate the number of entities at each *BP*.

3.3.1 Time cycle and granularity

The data activity in cities as a whole and individually by independent entities is not constant in time but can be related, to some extent, to repetitive temporal cycles. Usually, data networks activity is represented in weekly or daily cycles. In addition, these cycles can be discretized in time periods which are characterized by the collection of the total data activity within a time frame (e.g one hour). Therefore, each *BP* location has *N* entries in the database, being *N* the ratio (temporal cycle/time period).

3.3.2. Probability distributions

Ideally, real data regarding the use of services and mobility (or location) of entities would be necessary to depict the real data activity of a city. This data is usually obtained empirically, can be modelled after patterns and characterized by probability distributions. However, most of this type of data is not available at this early stage of this work, and the scenario has been set up by defining some basic rules that may illustrate to some extent the dynamics of a city. These unknown parameters are modelled using Poisson distributions characterized by a mean (λ). For example, the number of inhabitants in each individual household is unknown, but the average number of inhabitants per household is known. Consequently, individual inhabitant values can be given to each household following a Poisson distribution based on the known information.

State probability: It is related to the probability of people being on certain state (at work, at home…). For example, it is more likely for people to be at normal households during the evening and in companies during working hours. This probability is also modelled by a Poisson distribution where $\lambda(T)$ is the average percentage of the total number of inhabitants/workers present at the location within a time period (occupancy).

Mobility: In this context, mobile entities can be associated to streets as it is likely that most of them move following some street paths between two locations in the city. Hence, mobile entities can be associated to street *BPs* by probability distributions modelling how likely a *BP* is to have a number of mobile entities when using certain services. The location and usage probabilities can be assigned, for example, as a function of the type of street or time period of the day.

4. Case Study and Application

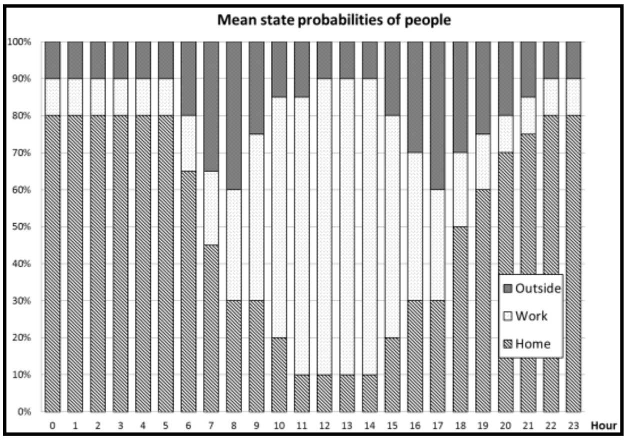
4.1 Case Study

The area under study is the municipality of Aalborg in Denmark, not only covering the city centre but also rural areas and other concentrations of households. As mentioned above, some of the required user behaviour patterns in terms of location, mobility or service usage are not known at this stage of the work. However, numerical values to the necessary parameters (e.g data generated by a service) are given symbolically to be able to illustrate the use of the proposed framework and methods. More accurate models require further investigations on, for example, specific current technologies, their near future evolution and empirical user behaviour pattern studies among others. In addition, a limited number of services are included in the case study which to our understanding represents a wide spectrum of the services offered within the smart city concept.

4.1.1 *Scenario Set up*

The following list summarizes the basic rules followed to set up the scenario for the case study, all based on three types of entities: people, machines and vehicles as introduced in Section 2.

- There are two types of areas: urban (population density higher than 500 inhab. per km^2) and rural.
- Day-time corresponds to the period from 7:00 to 19:00 and night-time from 19:00 to 7:00.
- People have three possible states: At home, at work, or outside. The mean probabilities for humans of being at each state are depicted in Fig 2. (a) and follow a Poisson distribution. For example, if at 10:00 65% of the people is at work, the percentage of workers assigned to each company follows a Poisson distribution of λ=65.
- Average number of inhabitants per household is 1.8 and the number of workers for each individual company is known.
- 50% of the people outside are walking and 50 % is in vehicles, 1.5 people travel per car on average, and the average trip time is 10 min. (these parameters are necessary to calculate number of people and vehicles to be distributed around *BPs*). People only walk in urban areas while vehicles move in rural and urban areas.
- Probability for *BPs* to be associated to people and vehicles is proportional to their own (*BP's*) population density area.
- For simplicity, the data generation of moving entities is assigned to one *BP,* implying that one entity generating some volume of data would do it instantly on a specific place even though it is moving. This simplified modelling is sufficient to illustrate the ideas behind the methods proposed. In more advanced modelling, the data volume could be assigned to a number of consecutive *BPs* depending on the speed and the concrete burst of data generated.
- Fig 2. (b) presents an overview of the *BPs* used for this case study.



Basic Point	Type 1 (# of points)	Type2 (# of points)
Buildings	Business (26.787)	Household (83.085)
Streets	Urban (122728)	Rural (143016)

Fig 2. (a) Mean State Probabilities; (b) Raw data database dimension.

4.1.2 Services

Smart Housing: It represents the collection of data related to monitoring and management of buildings in relation to Smart Grid , domotics, security services, etc. The data generated is assigned to building *BPs* and modelled by constant data per *BP*, variable data as a function of the size (inhabitants or workers), and variable data as a function of occupancy at representative times. Eq. (1) presents $D_i(T)$, the data activity in a period T (of one hour) of a generic building BP_i where $MaxPop_i$ is the number of inhabitants (in households) or workers (in companies), $CurPop_i(T)$ is the occupancy of the building in T, α = 10 if $7{:}00 \leq T \leq 23{:}00$ and α = 1 otherwise for households, and α = 10 for businesses.

$$D_i(T) = 10MB + 5MB \cdot MaxPop_i + \alpha MB \cdot CurPop_i(T) \quad (1)$$

Critical infrastructure management and monitoring: It represents a collection of services related to the basic infrastructure in cities (road, power, sewer…). These services only involve machines at fixed locations. The modelling of these services is summarized in the following list:

- Traffic monitoring video cameras placed at all crossroads *Deg* > 3. Data generated 75 MB/hour daytime and 100 MB/hour nighttime per device.
- Sewer system flow control devices in every crossroad *Deg* > 2 (assumption that sewer system is mapped 1 to 1 to street network). Data generated 5 MB/hour per device.
- Traffic lights and traffic control devices placed at every crossroad *Deg* > 2, only in urban areas. Data generated 20 MB/hour per each 10 % of population outside.

Mobile services, crowdsourcing and ITS: It represents some services accessed by people or vehicles moving around. The data for these services can flow in both directions, from/to the people or vehicle. For this case study, only the data generated by the users is included (as the data to the users is generated just at a limited number of servers).

- Crowdsourcing city information service. People upload (pictures, maps, text…) about the status of the city, traffic congestion, accidents, rare events, etc. 1 % of the walking people and 2 % of the people in vehicles use this service generating 5MB/trip (upstream).
- ITS, vehicles automatically send all sorts of data in relation to the transport infrastructure. 5 % of the vehicles generate 3 MB/trip.

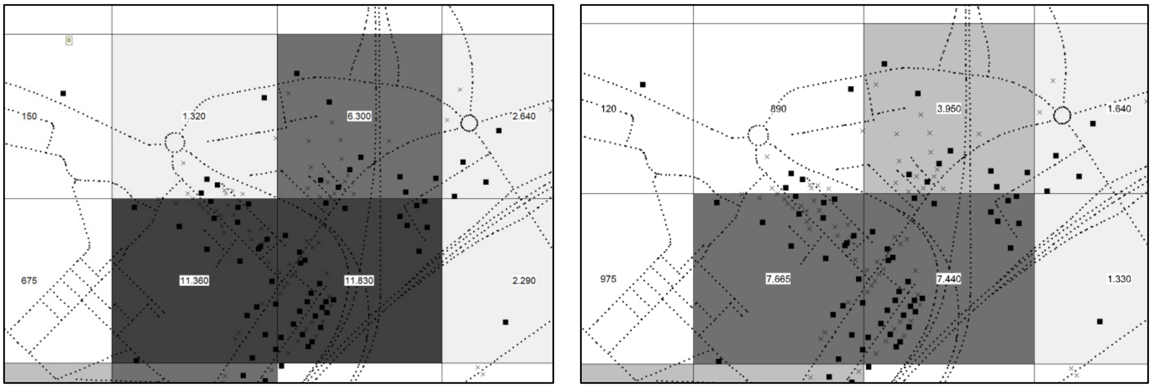


Fig 3. (a) Data activity 11 h.; (b) Data activity 19 h..

Fig. 3 illustrates the resulting heat-maps for the same area on different time periods. This area is a small section of the whole municipality under study. The time cycle is 1 day divided into a granularity of 1 hour periods, and the cell size is 500 meters. The crosses represent households, the squares companies and the dots street points; in addition the value of the data activity is displayed in MB/hour.

4.2 Application

Firstly, the most straight forward application of the proposed framework and methods is to understand how different services (existing or future) of very different nature and running on an enormous variety of devices contribute to the overall picture of digital data activity in smart cities. This understanding does not only come from spatio–temporal GIS data processing but also from its proper visualization. In this way, planning and management decisions (not only limited to data networks) can be taken based on a unified scenario of the city's digital communications, rather than just treating each service as an isolated silo. In the same context, sensitivity analyses can be carried out tuning up parameters of services, changing user behavior patterns, or injecting new traffic to the network to evaluate the effects and variations caused on the results.

In addition, the resulting spatio-temporal GIS data combined with information about existing access technologies infrastructure that carry the data generated may be used to evaluate its readiness, potential weaknesses, bottlenecks, and/or potential upgrade directions.

5. Conclusion

Cities around world are becoming "smarter" and regardless of the context of this evolution, digital data communications is one of the main common factors. In smart cities, a huge number of devices generate data traffic of very diverse nature associated to a large variety of applications. Also, the use of this enormous amount of data is very diverse ranging from infrastructure management to energy efficiency. In this paper, we propose a framework and methodology to get an overview of where and when digital data in relation to smart cities services is generated, based on spatio-temporal Geographic Information Systems (GIS). Within this framework, it is possible to combine and unify the data activity generated by different types of entities (people, vehicles…), using different applications and at different time periods. The method allows quantifying, geographically and dynamically in time, the volume of data generated to be used later for planning, managing, or monitoring purposes, among others. Moreover, the visualization of the resulting data activity may provide a convenient platform to understand the digital data dynamics in cities.

References

1 M. Naphade, G. Banavar, C. Harrison, J. Paraszczak, and R. Morris, “Smarter cities and their innovation challenges,” Computer, vol. 44, pp. 32–39, June 2011.

2. J. V. Winters, “Why are smart cities growing? who moves and who stays,” Journal of Regional Science, vol. 51, no. 2, pp. 253–270, 2011.

3. T. C. G. on behalf of the Global eSustainability Initiative, “The ict behind cities of the future.” SMART 2020, 2010.

4. C. Harrison and I. A. Donnelly, “A theory of smart cities,” in Proceedings of the 55th Annual Meeting of the ISSS, 2011.

5. C. Mulligan and M. Olsson, “Architectural implications of smart city business models: an evolutionary perspective,” Communications Magazine, IEEE, vol. 51, pp. 80–85, June 2013.

6. N. Komnimos, H. Schaffers, M. Pallot “Developing a policy roadmap for smart cities and the future internet,”, IIMC International Information Management Corporation, 2011.

7. V. Kostakos, T. Ojala, and T. Juntunen, “Traffic in the smart city: Exploring citywide sensing for traffic control center augmentation,” Internet Computing, IEEE, vol. 17, pp. 22–29, Nov 2013.

8. M. Valera and S. Velastin, “Intelligent distributed surveillance systems: a review,” Vision, Image and Signal Processing, IEE Proceedings -, vol. 152, pp. 192–204, April 2005.

9. J. M. Hern´andez-Mu˜noz, J. B. Vercher, L. Mu˜noz, J. A. Galache, M. Presser, L. A. H. G´omez, and J. Pettersson, “The future internet,” ch. Smart Cities at the Forefront of the Future Internet, pp. 447–462, Berlin, Heidelberg: Springer-Verlag, 2011.

10. I. Vilajosana, J. Llosa, B. Martinez, M. Domingo-Prieto, A. Angles, and X. Vilajosana, “Bootstrapping smart cities through a self-sustainable model based on big data flows,” Communications Magazine, IEEE, vol. 51, pp. 128–134, June 2013.

11. M. Al-Hader, A. Rodzi, A. Sharif, and N. Ahmad, “Soa of smart city geospatial management,” in Computer Modeling and Simulation, 2009. EMS '09. Third UKSim European Symposium on, pp. 6–10, Nov 2009.

12. T. M. Chen, “Smart grids, smart cities need better networks [editor’s note].,” IEEE Network, vol. 24, no. 2, pp. 2–3, 2010.

13. G. Piro, I. Cianci, L. Grieco, G. Boggia, and P. Camarda, “Information centric services in smart cities,” Journal of Systems and Software, vol. 88, no. 0, pp. 169 – 188, 2014.

14. David Loshin, “Enterprise Location Intelligence. Bringing Location-related Business Insight to Support Better Decision Making and More Profitable Operations”, White paper 2010.

15. Sibren Isaacman, Richard Becker, Ramón Cáceres, Margaret Martonosi, “Human Mobility Modeling at Metropolitan Scales”, 10th International conference on Mobile systems, applications, and services, pp. 239-252 2012.

16. Z. Al-adawi , S. Yousafzai , J. Pallister “Conceptual Model of Citizent Adoption of e-Government”, International Conference on Innovations in Information Technology, 2005.