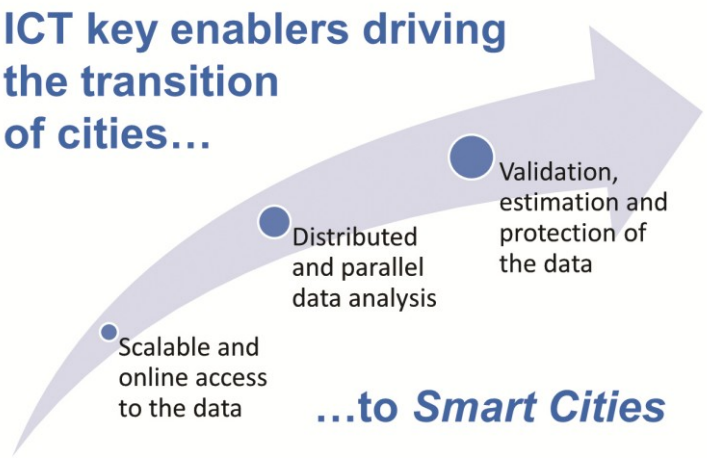
When Smart Cities meet Big Data

by Vincenzo Gulisano, Magnus Almgren and Marina Papatriantafilou[1]

Sharing information is a key enabler in the transition of a city becoming *smart*. Information, generated by the ICT backbone of a city, and maintained by distinct public and private entities, comes with processing challenges that must be addressed in order to increase citizens' quality of life and make their cities sustainable. In CRISALIS and SysSec, we investigate such challenges from a security perspective in order to protect and enhance smart cities' sensitive infrastructures.

The possibilities enabled by Information and Communication Technologies (ICTs) are driving the evolution and transition of cities to Smart Cities. The ultimate goal is to increase the awareness of citizens', companies' and authorities' and improve their quality of life while also making it sustainable. A considerable number of research directions embrace Smart Cities: users' privacy protection [1], detection of malicious actions and misuses and users' awareness through social media. More research efforts are dedicated to specific features of a Smart City. As an example, the energy forecast techniques used to predict consumption and allow the usage of alternative energy resources (e.g., solar or wind power) to be scheduled. What all these research fields have in common is their dependency on the (possibly sensitive) data produced by the devices forming the Internet of Things (IoT) of a city. The possibilities enabled by Smart Cities demand for novel data processing paradigms to form the expertise of public and private companies. Based on our experience with both academic and industrial partners, in this article we discuss some of the challenges associated with data processing in Smart Cities.



Scalable and online access to the data

In a Smart City, millions of messages will be exchanged on a daily basis by hundreds of thousands of devices (e.g., mobile phones, electrical meters, weather stations, etc.). For example, more than 1.2 million messages are exchanged on a daily basis within an AMI infrastructure (owned by one of our industrial partners) that covers a metropolitan area with roughly 600,000 inhabitants [2]. The information generated by such devices could be matched and joined to enhance the management of Smart Cities. For example, energy or water losses caused by faulty devices could be reduced by matching the consumption measured by users' meters with the one measured by other utilities' systems. To this end, on-the-fly processing of data becomes all the more important while traditional

[1] Published in ERCIM News #98, http://ercim-news.ercim.eu/en98/special/when-smart-cities-meet-big-data

store-then-process approaches in which each company retrieves its data and stores it in order to access it sometime in the future might be no longer appropriate.

Think in a distributed and parallel fashion.
Smart Cities will be composed of several independent networks (even within the same stakeholder). Hence, no centralized application will embrace the information carried by the messages exchanged by the devices. At the same time, the huge volume of information shared by ICT devices will make parallel processing a necessity [3]. To this end, pushing the analysis closer to the sources of information would be a natural way of analyzing the messages exchanged by them and leverage the information they carry. Challenging aspects in this context will be imposed by the constrained resources of such devices.

Validate, estimate and protect the data.
Cheap, resource-constrained devices are largely employed to build the networks that will form the IoT of a Smart City. Unfortunately, the data measured and reported by such devices (e.g., energy consumption readings) is usually noisy and lossy. Reasons of this are not limited uniquely to the devices themselves (e.g., faulty or badly calibrated devices, lossy or overloaded communication channels) but can also be caused by (possibly malicious) citizens. As an example, incorrect consumption readings could be manipulated by malicious users aiming to lower their bills. To this end, validation schemes, estimation schemes and security countermeasures must be adopted in order to ensure that who leverages the information is not mislead by incorrect, partial or malicious data.

The shift from cities to Smart Cities depends on the efficiency with which information is shared among citizens and private and public companies. This information brings challenges, and, following the big data revolution, novel processing schemes must be adopted to enable the possibilities that exist of this domain. All the possibilities enabled by smart cities, like improved quality of life or energy efficiency, shall build on top of efficient data processing and users' privacy protection schemes.

CRISALIS may be contacted at contact@crisalis-project.eu. SysSec may be contacted at the corresponding contact@syssec-project.eu, followed in twitter (twitter:syssecproject) and Facebook (http://www.facebook.com/SysSec).

References:

[1] V. Tudor, M. Almgre y M. Papatriantafilou, «Analysis of the Impact of Data Granularity on Privacy for the Smart Grid,» de *WPES '13 Proceedings of the 12th ACM workshop on Workshop on privacy in the electronic society*, 2013.

[2] Z. Fu, O. Landsiedel, M. Almgren and M. Papatriantafilou, "Managing your Trees: Insights from a Metropolitan-Scale Low-Power Wireless Network," in *CCSES'14: Proceedings of the 3rd Workshop on Communications and Control for Smart Energy Systems held in conjunction with the 33rd IEEE International Conference on Computer Communications (INFOCOM)*, 2014.

[3] V. Gulisano, M. Almgren and M. Papatriantafilou, "METIS: a Two-Tier Intrusion Detection System for Advanced Metering Infrastructures," in *The fifth International Conference on Future Energy Systems (ACM e-Energy)*, 2014.

Please contact:

Vincenzo Gulisano
vinmas@chalmers.se