Pervasive Data Sharing as Enabler for Mobile Citizen Sensing Systems

Waldir Moreira and Paulo Mendes
COPELABS, University Lusófona
Lisbon, Portugal
{waldir.junior, paulo.mendes}@ulusofona.pt

Abstract

Today, users can use their personal devices for a wide range of applications and services, such as controlling other devices, monitoring human physiological signals, or accessing information while on the move. Due to the communication and sensing capability of personal devices as well as embedded devices, their pervasive deployment and use may lead to an improvement of social and personal welfare by exploiting novel mobile citizen sensing applications. However, the pervasiveness of such large-scale sensing systems is only possible if devices are able to share sensing data independently of the available communication infrastructure, their location, and applications making use of the collected data. Hence, this paper describes a set of paradigms that should be considered to allow pervasive data sharing for the support mobile citizen sensing systems.

Index Terms

pervasive data sharing; mobile citizen sensing; cooperation incentives; trust; opportunistic communications; data-centric networking;

I. INTRODUCTION

Advancements in computing hardware and communication technologies resulted in personal mobile devices (e.g., smartphones, watches, glasses, gloves, bracelets) encompassing a diversified set of communications (e.g., Bluetooth, Wi-Fi, 3G) and sensing (e.g., proximity, location, human body signals) capabilities. Such devices\(^1\) have a direct impact on the evolution towards an Internet able to accommodate citizen’s needs in real time, through the integration of a large set of mobile nodes.

These powerful nodes may support novel applications for Mobile Citizen Sensing (MCS) systems, which rely on large-scale sensing data produced by the regular users throughout their daily routines, and that may impact different sectors of society (e.g., health care, education, industry, government agencies): over the past decade, the focus of wireless sensor networking has evolved from static networks of specialized nodes deployed to sense the environment, to networks making use of nodes able to exploit people’s mobility to sense large-scale social environments (e.g., MIT SENSEable City Lab, Intel Urban Atmospheres project).

A general purpose MCS system should sense, learn, and share information about its context (e.g., users’ behavior and surroundings). The resulting sensing data (simple or fused) should be made available to any application that is interested in such sensing activity. Moreover, since users (i.e., citizens) are producing and consuming good part of their data while on-the-go, the communication among entities of a MCS system (acting as sensors, actuators or decision makers) must take into account the potential intermittent nature of the networking scenario (e.g., lack of 3G coverage, closed WiFi networks). Therefore, MCS systems should be supported by a Pervasive Data Sharing (PDS) system, which allows the collection, processing, and sharing of any available user sensing data, and that ensures high delivery rates of such data with low latency and cost, even when facing challenged networking scenarios.

\(^1\)This is the author’s preprint version. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotion or for creating new collective works for resale or for redistribution to thirds must be obtained from the copyright owner. The camera-ready version of this work has been accepted for publication in the October IEEE Communications Magazine feature topic “Social Networks Meet Next Generation Mobile Media Internet” and is property of IEEE.”

\(^1\)Note that the words device, node, and peer are used interchangeably throughout this paper.
MCS systems have various applications in real life, such as routine improvement and disaster prevention [1]. In a routine improvement scenario, users produce content (e.g., videos, photos, GPS data), which can be used by other users to improve their daily routines (e.g., avoid clogged roads, or find a room to rent). Regarding a disaster prevention scenario, users in a national park may report smoke sightings (e.g., photos, text) by making such content pervasively available, sharing it with vehicles passing by, and combining it to readings from sensors (e.g., temperature and humidity data) deployed in the park area to produce alerts to the respective authorities (e.g., firefighters).

The wonder of MCS systems is that they can perform on the aforementioned scenarios by relying solely on the capabilities of the devices and information that is exchanged and produced locally among these devices, without depending of expensive networking infrastructure.

The scientific contributions of this paper can be summarized as follows:

- Analysis of the networking characteristics, requirements, and assumptions of pervasive MCS systems (Section II);
- Definition of set of design paradigms to devise Pervasive Data Sharing (PDS) systems (Section III);
- Provision of a general design guideline for the development of a PDS system, based on the proposed set of design paradigms (Section IV);
- Identification of a set of functional building blocks for the instantiation of the proposed PDS design paradigms as well as the resulting PDS node architecture that serves as enabler for MCS systems (Section V).

II. NETWORKING IN MOBILE CITIZEN SENSING SYSTEMS

Sensing data in MCS systems (cf. Fig. 1) is produced by different entities based on specific requirements, and may: i) span and be collected in different areas (i.e., in-building, city wide, forests); ii) be of different types (i.e., physiological, images, location, temperature); iii) be used for a variety of purposes (i.e., connected health, smart small grids, people-centric sensing applications); and iv) be combined with data coming from different sources for a more general purpose (e.g., temperature and humidity combined to identify a fire hazard situation).

Due to the inherent pervasive and opportunistic nature of MCS systems, sensing should be deployed based on a networking system that is able of exploiting any communication opportunity taking place among mobile nodes. As nodes are worn/carried by people and people have social relationships and data interests\(^2\), sensing data may be exchanged considering their social interactions and/or common interests [2].

Despite the increased capabilities (e.g., processing, storage) of personal devices, MCS systems cannot assume that their users are willing to share such capabilities. Thus, egoistic behavior must be overcome by incentive models to encourage users in engaging in the cooperative networking process.

\(^2\)Data interests in the context of this work can refer to the analyzed sensing data, to the node’s contextual data, or to the information required by the sensing process (baseline data and learning models).
Furthermore, mobile communications should only rely on trusted devices that are willing to store and share data by taking advantage of communication opportunities: a malicious user may easily display cooperative behavior with the intention of accessing the users’ sensing data. So, trust mechanisms must be in place to provide users with secure data exchange in dynamic scenarios.

When facing a large number of independent devices, it is necessary to rely on a data-centric networking approach [3], [4] in order to create a robust communication system independent of device location, while being aware of users’ data interests. By focusing on the produced data and not on the device, the networking system is expected to scale while keeping the required robustness.

Considering these characteristics/requirements, next we analyze a set of paradigms that should be followed to design truly deployable PDS solutions to support MCS systems.

III. DESIGN PARADIGMS FOR PERVERSIVE DATA SHARING

Since there are some fundamental questions that have not been answered properly (for instance, what are the general principles that a data sharing system should have in common in order to ensure a pervasive deployment of MCS system?), our goal is to take a step in this direction by proposing an original set of PDS paradigms that serves as enabler for a general purpose MCS system.

PARADIGM #1 - COOPERATION INCENTIVES

There are different ways of encouraging cooperation, i.e., by means of virtual currency or based on reputation [5]. By employing virtual currency mechanisms, the user is rewarded for sharing its sensing, storage, and communication resources. Such rewards can later be exchanged by the user for other resources, such as free Internet access. In the case of reputation-based cooperation, the user has his reputation increased inside the system as long as he cooperates with others. Such increased reputation makes the user very reliable and trustworthy in the system.

This paradigm guarantees the basic operation of the sensing system: by encouraging users to help improving the operation of the network (i.e., carrying and/or relaying data), their own networking experience is improved: for instance, users will always have a way to send and retrieve information of their interest). On the other hand, this paradigm shall penalize egoistic users, for instance, by discouraging other users to carry/relay their data.

PARADIGM #2 - TRUSTWORTHY PEERING

In a pervasive scenario, the PDS system should rely on the possibility to build trust on-the-fly: devices create trust circles (i.e., sets of trusted users) based on the reputation of those with whom they interact, and based on the impact that those devices had on previous interactions. This way, users can establish communication with desirable trust levels and safely share data [6].

Data may be generated by malicious users with the intent of harming the functioning of the PDS system (e.g., a misbehaving user could create fake data causing the triggering of false alarms). This situation can be mitigated by identifying users that are trustworthy.

In a pervasive system of personal devices, this paradigm may be implemented by mechanisms that allow sharing based on the notion of social relationships [7], [8] and shared interest [9] aspects: users may find easier to share data with users with whom they share social relationships and data interests.

PARADIGM #3 - OPPORTUNISTIC COMMUNICATIONS

Robust data dissemination is of great importance while designing a PDS system: users have access to data based on the probability of meeting suitable carriers over any wireless interface (e.g., Bluetooth, Wi-Fi), instead of based on the probability of finding an open hotspot or based on the availability of expensive 3G connectivity. This is what we refer to as right-here-right-now approach, where a carrier can be a device or even an open hotspot configured with a set of data interests that can allow the immediate exchange of data.

Different solutions have emerged to allow this right-here-right-now approach [2]. Since devices are used by people, who happen to have social relationships and diverse data interests, information may be exchanged considering the social interactions existing between users and/or the common interests they share [8], [9], [10].
Nevertheless, of importance to the success of PDS systems is the trade-off between the delivery probability and latency experienced by such opportunistic PDS system, as well as the robustness of the overall sensing system: modern control theory is largely based on the abstraction that data is shared over perfect communication channels, which is not realistic to assume in a large-scale urban sensing systems.

**PARADIGM #4 - DATA-CENTRIC NETWORKING**

The efficiency of a data centric approach relies on a selective management of data to tackle user’s interests and privacy concerns. By understanding the relevancy that data has to the user and to the entities interacting with him/her, network (e.g., bandwidth) and device (e.g., battery) resources can be spared by reducing the number of data transmissions.

Since users are normally concerned about their privacy, the PDS system needs to be able to selectively share data based on privacy concerns, namely when the device does not have the computational power needed to analyze the collected sensing data. Still, a mechanism to ensure data privacy and anonymous sensing operations must not rely on registration authorities or centralized task services [11].

Besides the need to selectively handle data, the robustness and reliability of a PDS system depends upon the availability of useful data: this can be ensured by pre-fetching data based on data interest inference, improving resource utilization and the availability rate of data to the sensing applications.

**IV. HOW TO DESIGN PERVERSIVE DATA SHARING SYSTEMS**

The aforementioned paradigms help building a PDS system that copes with the characteristics and requirements of MCS communications in large-scale, urban pervasive networks.

It is important to note that we are neither claiming that the combination of these paradigms will result in the perfect PDS system, nor interested in the specific approach employed in each paradigm. Instead, our effort is towards a guideline to come up with a realistic and deployable PDS system, able to allow users to exchange collected sensing data through their interactions, supporting different applications and services of different sectors such as health care, education, industry, and government agencies.

Fig. 2 shows our view of how the four paradigms should come together, observing important correlation aspects that must be taken into consideration while building a PDS system.

According to paradigm #1, user willingness to participate in a PDS system can be exploited to overcome the user’s egoistic behavior, thus can aid in identifying cooperative users. One may argue that paradigm #1 is not required given the fact that users are expected (i.e., willing) to cooperate at all times. This assumption
must be dealt with care: although devices have significant resources, their owners will certainly not share all the available resources [6].

Since users are wary of interacting with others who they do not trust, paradigm #2 adds to the PDS system the capability of overcoming this lack of trust. Such trustful property is a baseline to increase the rate at which users participate in a PDS system, as well as avoiding with high probability some hazard operational situations.

Based on paradigms #1 and #2, the PDS system is built based on a set of cooperative and trusted users, who are considered for data exchange. For that, paradigm #3 is employed allowing the PDS system to explore the networking capabilities of trustful devices by implementing the right-here-right-now approach, i.e., making the most out of the different contact opportunities between users to allow the exchange of data.

Note here that paradigm #3 cannot be employed prior to paradigms #1 and #2: we cannot assume that every user, despite of having the potential to be a carrier/disseminator, is willing to cooperate or can be trusted.

While paradigm #3 ensures opportunistic operation over trustworthy pervasive scenarios, paradigm #4 guarantees that the PDS system reduces the network load by selecting data based on users’ interest and privacy considerations, pre-fetching data close to interested users and fusing data based on application requirements.

Generally speaking, there is a direct correlation between all four paradigms: To start with, paradigms #1 and #2 create the conditions for a realistic deployment of MCS systems, which rely on the cooperation among trustful devices. Data flow among trustful devices in pervasive scenarios is only possible by the exploitation of any communication opportunity (paradigm #3), which should ensure a good performance. The efficiency of the MCS system is then ensured by the application of paradigm #4, which allows the system to scale, independently of the location of devices, reaching the desirable level of data robustness and reliability.

V. ENABLING MOBILE CITIZEN SENSING SYSTEMS

We believe that using a named data approach together with a social-based opportunistic forwarding approach may facilitate the development and deployment of MCS systems: since nodes, which are constantly collecting and consuming sensing data (e.g., weather conditions, pollution levels) are carried by users, they can pervasively exchange such data through users’ social interactions and data interests. Moreover, the exchange of data can be done with high probability, low cost and low latency while being agnostic about its location, which is beneficial in dynamic mobile scenarios.

This section starts by overviewing relevant functional blocks that can be used to implement each of the proposed PDS paradigms, as well as their implicit/explicit alignment with one another as to enable MCS systems. Then, a PDS node architecture is presented, followed by an operational example.

A. Functional Blocks for Pervasive Data Sharing

Paradigms #1 and #2 refer to users’ engagement in the PDS cooperation process as trustworthy peers. Within the context of a MCS scenario, cooperation can happen based either on: i) the trust level among users; or ii) rewarding those who engage in cooperation.

A suitable cooperation framework may consider: i) a reciprocity-based incentive mechanism that takes into account users’ reputation (i.e., levels of trust) to allow cooperation in scenarios where nodes know each other; and ii) a reward-based incentive mechanism that encourages nodes to cooperate by allowing the exchange of virtual currency, which overcomes the lack of trust [5], [12].

Such cooperation framework matches the requirements of a MCS scenario: cooperation happens independently of how trusted the environment is. This is a desired feature for a PDS system given the different contexts in which users will find themselves (e.g., their known communities, another country).

Regarding the trust framework, different mechanisms to properly reflect trust associations between users may be used [6]: users are uniquely identified by means of virtual identities based on cryptography to reduce impersonation and non-repudiation issues; users can explicitly set how they trust unknown users (i.e., dispositional trust). Trust computation is given by the different trust associations and may be influenced by local (e.g., user’s reputation) and external (e.g., presence of malicious users in the vicinity) aspects.

Within the MCS context, these trust mechanisms allow the creation of trust circles to allow reliable, trusted communications that may be more efficient than hard-coded security (e.g., public keys) in pervasive dynamic scenarios.
One can clearly see that solutions for paradigms #1 and #2 are closely aligned and intertwined: cooperation among trusted users happens easily. However, such alignment is not so clear for the solutions employed in the remaining paradigms.

In what concerns paradigm #3, there are several opportunistic forwarding solutions that may be exploited in a PDS system, ranging from flooding approaches to more elaborate ones encompassing different social features (i.e., common communities, shared interests, popularity, dynamic social behavior of users) [2].

By considering social features, solutions for paradigm #3 implicitly relate to paradigms #1 and #2 to some extent: users can easily cooperate and are prone to trust others with same interests, or that belong to same community, for instance. Nevertheless, paradigms #1 and #2 are required for real application deployment as users may still be reluctant to exchange data even with those with whom they share interests, communities, or some other level of social affinity. Additionally, solutions for paradigm #3 must be used to allow the system to be aware of the dynamics of users’ social behavior, as this is beneficial for opportunistic communications in urban, dense scenarios [13].

With this in mind, solutions such as Bubble Rap[7], dLife [8] and SCORP [9] fit paradigm #3: these proposals consider how users are socially connected, the communities they belong to, how important they are in the system, and take into account the users’ interests on the data traversing the network. The features of these social-aware and content-based solutions are aligned with the characteristics of nodes operating over a large-scale urban sensing scenario: users are very dynamic, are focused on their interest on sensing data, wish an anytime/anywhere data exchange capability, and do have relevant social interactions.

Finally, there are several frameworks that are centered on the data and could be employed to implement paradigm #4, such as Publish-Subscribe Internet Routing Paradigm (PSIRP), Data-Oriented Network Architecture (DONA), Named Data Networking (NDN), Content-Centric Networking (CCN), and Network of Information (NetInf). Each of these proposals has its own particularities (e.g., employ their own naming scheme) and look into different content-centric aspects (e.g., naming, security, routing) emphasizing few of these aspects according to the application to which they have been devised.

One can observe that solutions for paradigm #4 aim mostly at improving data dissemination in fixed networks (i.e., Internet at large) with few examples supporting ad-hoc, vehicular, and Internet of Things networking. This also makes such solutions not aligned with paradigm #3, since these approaches assume that nodes are always connected. Moreover, features of paradigm #4 such as the capability of increasing the quality and value of the data by means of data aggregation mechanisms, and privacy are still novel research issues when it comes to this paradigm [14], [1].

Thus, concerning a MCS scenario, paradigm #4 could be potentially based on a CCN/NDN-like framework given its decentralized feature. While in the other frameworks nodes rely on specific entities (i.e., rendezvous points in PSIRP, name resolution system in NetInf, resolution handlers in DONA) to handle user interest, queries, and responses, CCN/NDN is more straightforward, requiring nodes to just manifest their interest for retrieving content [15].

This feature of CCN/NDN is interesting as it allows easy integration with paradigm #3 (i.e., nodes easily exchange interest lists and desired content based on such lists), and can cope with the dynamism of user behavior (i.e., users want to send/retrieve content anytime/anywhere).

B. Node Architecture for Pervasive Data Sharing

Following the proposed set of paradigms, design guidelines, and analysis of potential solutions for each paradigm, Fig. 3 presents the proposed node architecture to deploy pervasive data sharing systems.

Generally speaking, in MCS scenarios all nodes may forward data: nodes should not only be able to collect/consume data, but they should also be able to forward such data/interest requests towards the intended destination and/or source of data.

The proposed PDS node architecture comprises two engines, namely trusted cooperation and forwarding. The former is responsible to employ the solutions for paradigms #1 and #2. This informs the current node about which neighboring peers are willing to cooperate (represented by filled circles in the right-hand side graph on Fig. 3) and can be considered part of its trust circles (non-crossed links between them on Fig. 3). As mentioned
in Section V-A, cooperative behavior can be encouraged either with reciprocity-based or reward-based incentives, and trust associations between users are built as they interact in the system.

It is worth noting that a node may assume a cooperative behavior, but it may still not be trusted (filled circle with crossed link on Fig. 3). The PDS system must be able to identify such misbehaving/malicious nodes in order to avoid them.

The forwarding engine is responsible for employing paradigms #3 and #4. Once the current node knows its neighboring peers (trusted, cooperative users), it will start sharing data. From the social-aware and content-based proposals mentioned in Section V-A, opportunistic communication shall be driven by SCORP. This is due to the fact that not only does it consider users’ interests, but also how these users are socially connected. It is worth mentioning that SCORP measures the social weight (SW) of a node towards specific interests [9].

In the proposed node architecture, data is managed based on persistent and temporal data structures, as shown in Fig. 3. Persistent structures are: i) the data cache (a.k.a. content store) which holds data carried by the node, and that is of interest to itself and/or other nodes; and ii) the carried interest table that keeps track of the data interests of the node, as well as its social weights towards interests of other nodes with whom it socially interacts.

Generally speaking, the data cache remains the same as the CCN/NDN content store, but it limits the amount of stored data. This is because nodes in a MCS scenario have different capabilities and users may not be willing to share all of their storage on behalf of others. Thus, stored data can be removed according to the rate of interaction a node has with others who are interested in the carried content. Moreover, since the carried interest table includes the interests of the node and those of others, SW is assigned a value of infinity to identify the interests of the current node, and facilitate content exchange.

It is important to say that data and interest can be generated by local applications (cf. interface 2 on Fig. 3) or received via a networking interface (cf. interfaces 0 and 1). Also note that in any of the two persistent structures, an entry has a defined time-to-live (TTL): data TTL defines the time usefulness of data, as tagged by the source; and interest TTL defines the time period over which some node is interested in a specific type of data. By erasing carried data and interests based on TTL, we ensure that only useful data is transported, which is a desired scalability feature in large-scale urban sensing scenarios.

The temporary structures are: i) the temporary Pending Interest Table (PIT), which lists the interest carried by a neighboring node; and ii) the temporary Carried Data Table (CDT), which replaces the Forwarding Information...
Base found in CCN/NDN. The PIT also has a SW field to describe the social weight between the neighboring node and specific interests it has come across. PIT keeps track of the data missing from data cache, and does not forward an interest request as it would normally happen with content-centric approaches. CDT is populated with information about the data that neighboring nodes are currently carrying. The entries on these tables are temporary, since they are erased as soon as the contact with a neighboring node is broken.

Finally, both PIT and CDT do not map interest/content to interfaces as it would normally happen in content-centric approaches. This is because the PDS node is more concerned in mapping the interest/content to nodes in which it is socially well connected.

One can observe that these changes allow our PDS system to be easily built based on well-known networking paradigms to support mobile citizen sensing in large-scale urban scenarios.

In order to illustrate the operation of the proposed node architecture, let’s exemplify the interaction between two nodes A and B, from the perspective of node A (the same operation is taking place at node B concurrently).

When nodes A and B meet, node B sends two metadata lists concerning i) the data it currently carries (data cache), and ii) its interests as well as its SWs towards the interests of other nodes it has interacted up to this point (carried interest). These lists populate node A’s CDT and PIT, respectively.

Node A then uses the recently updated PIT to quickly determine whether node B is interested in, and/or if node B is socially well connected to other nodes with interest in the data carried by node A.

For every piece of information in its data cache, node A shall forward actual data if it is not in its updated CDT (i.e., missing from node B’s data cache) and:

- node B is interested on it (“/ccn/copelabs/videos” is forwarded since $SW = \infty$, cf. PIT in Fig. 3), or
- node B’s SW towards that specific interest (obtained from the PIT) is greater than the SW of node A towards such interest as specified in its own carried interest list (“/ccn/copelabs/papers/2013” is forwarded since node B’s $SW = 0.9 > node A’s SW = 0.2$, cf. PIT and carried interest in Fig. 3).

It is worth mentioning that the number of interests in a PDS system can be significantly high given its granularity (e.g., one user likes car, while other likes a specific model from a specific manufacturer), which may affect scalability. Thus, the PDS node architecture can employ mechanism to mitigate such effect by creating metadata lists that include only SW to socially relevant interests and with useful TTL.

VI. CONCLUSIONS

Today, we are surrounded by a panoply of devices that produce a massive amount of data and have a diversified set of communications (i.e., Bluetooth, Wi-Fi, 3G), and sensing (i.e., activity, proximity, sound, location, human body signals) capabilities. Based on such properties, mobile devices may support Mobile Citizen Sensing (MCS) applications, which can have an impact in different sectors of society.

Due to its pervasive nature, we cannot assume that the control of MCS systems can be based on a networking system where data is transmitted over perfect communication channels. Instead, the communication between mobile devices should be supported by a data sharing system able to ensure high data exchange probability, low latency and low cost, even when facing challenged networking scenarios.

The creation of MCS systems, based on Pervasive Data Sharing (PDS), requires a constructive engineering approach. This article gives an introduction to networking requirements of MCS systems and proposes a set of design paradigms for PDS systems. These paradigms are derived by looking at different user-centric and data-centric networking approaches and by extracting common features that contribute to the efficiency of pervasive data sharing in large-scale urban sensing systems. We believe that these paradigms represent basic building blocks, and so we suggest a process for the design of pervasive data sharing solutions, and show that MCS systems can be easily devised based on current content-centric and opportunistic networking approaches. We hope that these contributions will stimulate further research to allow the deployment of mobile citizen sensing systems.

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Unions (EU) Horizon 2020 research and innovation programme under grant agreement No 645124 (Action full title: Universal, mobile-centric and opportunistic communications architecture, Action Acronym: UMOBILE). This paper reflects only
the authors views and the Community is not liable for any use that may be made of the information contained therein.

Acknowledgments are also due to CitySense project from COPELABS.

REFERENCES


