



Mobile collaborative sensing: framework and algorithm design

Yuanfang Chen

► To cite this version:

Yuanfang Chen. Mobile collaborative sensing: framework and algorithm design. Networking and Internet Architecture [cs.NI]. Institut National des Télécommunications, 2017. English. NNT: 2017TELE0016 . tel-01597064

HAL Id: tel-01597064

<https://theses.hal.science/tel-01597064>

Submitted on 28 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Doctor of Philosophy (Ph.D.) Thesis
Université Pierre & Marie Curie - Sorbonne Universités and
Télécom SudParis

Specialization

COMPUTER SCIENCE

presented by

Yuanfang Chen

Submitted for the partial requirement of

Doctor of Philosophy
from
Université Pierre & Marie Curie (UPMC) - Sorbonne Universités and Télécom
SudParis

**Mobile Collaborative Sensing:
Framework and Algorithm Design**

12th July, 2017

Committee:

Noel Crespi	Director of Thesis	Professor, Institut Mines-Télécom, Télécom SudParis - France
Francoise Soulie-Fogelman	Reporter	Professor, Tianjin University - China
Pierluigi Siano	Reporter	Professor, Università degli Studi di Salerno - Italy
Zhangbing Zhou	Examiner	Professor, China University of Geosciences (Beijing) - China
Bruno DEFUDE	Examiner	Professor, Institut Mines-Télécom, Télécom SudParis - France
Sebastien Tixieul	Examiner	Professor, UPMC - Sorbonne Universités – Paris - France

nr. 2017TELE0016

**Thèse de Doctorat (PhD) de
l'Université Pierre & Marie Curie - Sorbonne Universités and
Télécom SudParis**

Spécialité

INFORMATIQUE

présentée par

Yuanfang Chen

pour obtenir le grade de

**Docteur de l'Université Pierre & Marie Curie (UPMC) - Sorbonne Universités
and Télécom SudParis**

Framework et Algorithmes pour la Conception d'Applications Collaboratives de Capteurs

12th July, 2017

Jury composé de :

Noel Crespi	Directeur de thèse	Professor, Institut Mines-Télécom, Télécom SudParis - France
Francoise Soulie-Fogelman	Rapporteur	Professor, Tianjin University - China
Pierluigi Siano	Rapporteur	Professor, Università degli Studi di Salerno - Italy
Zhangbing Zhou	Examineur	Professor, China University of Geosciences (Beijing) - China
Bruno DEFUDE	Examineur	Professor, Institut Mines-Télécom, Télécom SudParis - France
Sebastien Tixieul	Examineur	Professor, UPMC - Sorbonne Universités – Paris - France

Dedication

To my Dearest Mother

Acknowledgments

To whom it may concern, Many thanks for paying attention to my thesis. Doing a PhD is a long and wonderful journey of mine, where I moved to France and left my dad and mom, where I met many new friends from different countries, where I gradually turned myself from an engineer to a researcher. All of these impact my life.

I gratefully thank following persons who helped me during my PhD, my supervisor and thesis director Prof. Noel CRESPI, the dean of our department Prof. Djamal ZEGHLACHE, the secretary of our department Valerie MATEUS, and my colleagues including Mr. Antonio M. ORTIZ, Prof. Gyu Myoung LEE, and Miss Xiao HAN. All help from them is valuable and should be highly appreciated. Moreover, I would like to appreciate the collaborators of my research including Prof. Lei SHU from University of Lincoln in the UK, Prof. Zhangbing ZHOU from University of Geosciences in China, Prof. Pierluigi SIANO from University of Salerno in Italy, and Prof. Quang Trung DUONG from Queen's University Belfast in the UK.

I would like to appreciate the jury members of my PhD defense committee including Prof. Bruno DEFUDE, Prof. Françoise Soulie FOGELMAN, who allocate me a slice of their busy hours so as to help me validate my PhD research.

I am pretty sure that I miss some important people I should thank. Thus, I would like to say “Thank You!” to all people surrounding me or having contacted me. It must be the LORD who leads me here or there to know you and get help from you. Please accept my most sincere thanks!

Yours faithfully,

Yuanfang CHEN

Abstract

Nowadays, there is an increasing demand to provide real-time information from the environment, e.g., the infection status of infectious diseases, signal strength, traffic conditions, and air quality, to citizens in urban areas for various purposes. The proliferation of sensor-equipped devices and the mobility of people are making Mobile Collaborative Sensing an effective way to sense and collect information at a low deployment cost. In Mobile Collaborative Sensing, instead of just deploying static sensors in an interested area, people with mobile devices play the role of mobile sensors to sense the information of their surroundings, and the communication network (3G, WiFi, etc.) is used to transfer data for Mobile Collaborative Sensing applications.

Typically, a Mobile Collaborative Sensing application not only requires each participant's mobile device to possess the capability of performing sensing, and returning sensed results to a central server, but also requires to collaborate with other mobile and static devices.

In order to make sensed results well represent the physical information of a target region, and well be suitable to a certain application, what kind of data can be used for different applications, and what kind of information needs to be included into the collected sensing data? Spatio-temporal data can be used by different applications to well represent the target region. In different applications, location and time information is two kinds of common information, and by using such information, the target region of an application is under comprehensive monitoring from the view of time and space. Different applications require different information to achieve different sensing purposes. For example, in this thesis: (i) MCS-Locating application. Signal strength information needs to be included into the sensed data by mobile devices from signal transmitters; (ii) MCS-Prediction application. The relationship between infecting and infected cases needs to be included into the sensed data by mobile devices from disease outbreak areas; (iii) MCS-Routing application. Real-time traffic and road information from different traffic roads, e.g., traffic velocity and road gradient, needs to be included into the sensed data by road-embedded and vehicle-mounted devices.

With *sensing the physical information of a target region*, and *making mobile and static devices collaborate with each other* in mind, in this thesis three sensing based optimization applications are studied, and following four research works are conducted:

- **Mobile Collaborative Sensing Framework.** In this work, a Mobile Collaborative Sensing framework is designed. This framework facilitates the cooperativity of data collection, sharing, and analysis among different devices. Data is collected from different sources and time points. For deploying the framework into applications, relevant key challenges and open issues are discussed.
- **MCS-Locating.** In this work, algorithm LiCS (Locating in Collaborative Sensing based Data Space) is proposed to achieve target locating. LiCS uses Received Signal Strength (RSS) that exists in any wireless devices as location fingerprints to differentiate different locations, so LiCS can be directly supported by off-the-shelf wireless infrastructure. This algorithm uses trace data from individuals' mobile devices, and a location estimation model. LiCS trains the location estimation model by using the trace data to achieve collaborative target locating. Such collaboration between different devices is data-level, and model-supported.
- **MCS-Prediction.** In this work, a recognition model is designed to dynamically acquire the

structure knowledge of the relevant RCN during disease spread. On the basis of this model, a prediction algorithm is proposed to predict the parameter R . R is the reproductive number which is used to quantify the disease dynamics during disease spread.

- **MCS-Routing.** In this work, an eco-friendly navigation algorithm, *eRouting*, is designed by combining real-time traffic information and a representative factor based energy/emission model. Based on the off-the-shelf infrastructure of an intelligent traffic system, the traffic information is collected.

Résumé

De nos jours, il y a une demande croissante pour fournir de l'information en temps réel à partir de l'environnement, par exemple, l'état infectieux des maladies infectieuses, la force du signal, les conditions de circulation et la qualité de l'air, aux citoyens dans les zones urbaines à diverses fins. La prolifération des dispositifs équipés de capteurs et la mobilité des personnes font de la Mobile Collaborative Sensing un moyen efficace de détecter et de collecter de l'information à un faible coût de déploiement. Dans Mobile Collaborative Sensing, au lieu de simplement déployer des capteurs statiques dans une zone intéressée, les personnes disposant d'appareils mobiles jouent le rôle de capteurs mobiles pour détecter l'information de leur environnement et le réseau de communication (3G, WiFi, etc.) pour les applications mobiles de détection collaborative.

En général, une application mobile de détection collaborative exige non seulement que l'appareil mobile de chaque participant ait la capacité d'effectuer la détection et retourne les résultats détectés à un serveur central, mais nécessite également de collaborer avec d'autres dispositifs mobiles et statiques.

Pour que les résultats détectés puissent bien représenter l'information physique d'une région cible et bien convenir à une certaine application, quel type de données peut être utilisé pour différentes applications et quel type d'information doit être inclus dans les données de détection collectées ? Les données spatio-temporelles peuvent être utilisées par différentes applications pour bien représenter la région cible. Dans des applications différentes, l'information de localisation et de temps sont deux types d'information communes, et en utilisant cette information, la région cible d'une application est sous surveillance complète de la vue du temps et de l'espace. Différentes applications nécessitent de l'information différente pour atteindre des objectifs de détection différents. Par exemple, dans cette thèse : (i) MCS-Locating application. L'information de résistance du signal doit être incluse dans les données détectées par des dispositifs mobiles à partir d'émetteurs de signaux ; (ii) MCS-Prédiction application. La relation entre les cas d'infection et les cas infectés doit être incluse dans les données détectées par les dispositifs mobiles provenant des zones de flambée de la maladie ; (iii) MCS-Routing application. L'information routière et routière en temps réel provenant de différentes routes de circulation, par exemple la vitesse du trafic et le gradient de la route, doit être incluse dans les données détectées par des dispositifs embarqués et montés sur véhicule.

Avec la détection de l'information physique d'une région cible, et la mise en interaction des dispositifs mobiles et statiques à l'esprit, trois thèmes d'optimisation basés sur la détection sont étudiés et quatre travaux de recherche sont menés :

- **Mobile Collaboratif Détection Cadre.** Dans ce travail, un cadre mobile de détection collaborative est conçu. Ce cadre facilite la coopérativité de la collecte, du partage et de l'analyse des données entre différents dispositifs. Les données sont collectées à partir de sources et de points temporels différents. Pour le déploiement du cadre dans les applications, les défis clés pertinents et les problèmes ouverts sont discutés.
- **MCS-Locating.** Dans ce travail, l'algorithme LiCS (localisation dans l'espace de données basé sur la détection collaborative) est proposé pour atteindre la localisation de la cible. LiCS utilise la puissance du signal reçu qui existe dans tous les périphériques sans fil comme empreintes digitales de localisation pour différencier les différents emplacements, de sorte LiCS peut être

directement pris en charge par l'infrastructure sans fil standard. Cet algorithme utilise des données de trace provenant d'appareils mobiles d'individus, et un modèle d'estimation d'emplacement. LiCS forme le modèle d'estimation de localisation en utilisant les données de trace pour atteindre la localisation de la cible collaborative. Cette collaboration entre différents périphériques est au niveau des données et est supportée par un modèle.

- **MCS-Prédiction.** Dans ce travail, un modèle de reconnaissance est conçu pour acquérir dynamiquement la connaissance de structure de la RCN pertinente pendant la propagation de la maladie. Sur la base de ce modèle, un algorithme de prédiction est proposé pour prédire le paramètre R . R est le nombre de reproduction qui est utilisé pour quantifier la dynamique de la maladie pendant la propagation de la maladie.
- **MCS-Routing.** Dans ce travail, un algorithme de navigation écologique, *eRouting*, est conçu en combinant de l'information de trafic en temps réel et un modèle d'énergie/émission basé sur des facteurs représentatifs. Sur la base de l'infrastructure standard d'un système de trafic intelligent, l'information sur le trafic sont collectées.

Table of contents

1	Introduction	19
1.1	Background	19
1.2	Research Motivation	22
1.3	Contributions	25
1.4	Organization of this Thesis	26
2	State of the Art Review	27
2.1	Mobile Collaborative Sensing	27
2.2	Mobile Crowd Sensing	29
2.3	Heterogeneous Devices and Communication Standards	30
2.4	Participatory Sensing	32
3	Mobile Collaborative Sensing Framework	33
3.1	Introduction	34
3.2	Definitions and Advances	34
3.2.1	What is Collaborative Intelligence?	34
3.2.2	What is Sensing Intelligence?	35
3.2.3	Advances	36
3.2.3.1	Collaborative Intelligence	36
3.2.3.2	Sensing Intelligence	38
3.3	Mobile Collaborative Sensing Framework	40
3.3.1	Key Components of the Mobile Collaborative Sensing Framework	41
3.3.2	On-Going Efforts	43
3.3.2.1	Dynamic Detection of Toxic Gases	43
3.3.2.2	Citizen Sensing of La Poste	44
3.4	Key Challenges and Open Issues	47
3.4.1	Key Challenges	47
3.4.1.1	Data	47
3.4.1.2	Functionality	48
3.4.2	Open Issues	49

3.5	Conclusion	50
4	MCS-Locating	51
4.1	Introduction	52
4.2	Related Work	53
4.2.1	Wireless Indoor Target Locating	53
4.3	System Model	54
4.4	Locating in Mobile Collaborative Sensing based Data Space	55
4.5	Evaluation	58
4.5.1	Experimental Setup	58
4.5.2	Performance Evaluation	59
4.6	Conclusion	61
5	MCS-Prediction	63
5.1	Introduction	64
5.2	Related Work	65
5.3	Network Model	66
5.4	Recognition Model	68
5.5	Prediction Algorithm	70
5.6	Evaluation	72
5.6.1	Experimental Setup	72
5.6.2	Experimental Results and Discussion	73
5.7	Conclusion	76
6	MCS-Routing	77
6.1	Introduction	78
6.2	Related Work	80
6.2.1	Eco-Friendly Navigation	80
6.2.2	Energy/Emission Estimation Models	81
6.3	System Model	81
6.4	Problem Formulation	82
6.5	Energy/Emission Model	83
6.6	Energy/Emission-related Cost Function	84
6.7	Algorithm Design	85
6.8	Evaluation and Results	86
6.8.1	Experimental Setup	86
6.8.2	Results and Discussion	87
6.9	Conclusion	89
7	Conclusion	91
7.1	Summary	91
7.1.1	Summary of Mobile Collaborative Sensing Framework	92
7.1.2	Summary of MCS-Locating	92
7.1.3	Summary of MCS-Prediction	92

<i>TABLE OF CONTENTS</i>	17
7.1.4 Summary of MCS-Routing	93
References	95

Introduction

Contents

1.1	Background	19
1.2	Research Motivation	22
1.3	Contributions	25
1.4	Organization of this Thesis	26

1.1 Background

Mobile Collaborative Sensing [1,2] has been proposed to intelligently capture physical information, as the number of mobile devices equipped with sensors ¹ shows dramatic growth. It allows heterogeneous mobile devices to seamlessly work together, and even work with the crowd wisdom of humans. Figure 1.1 illustrates potential impacts of Mobile Collaborative Sensing for the society and economy, and shows partial potential application domains.

Applications of Mobile Collaborative Sensing. Facilitated by the widespread adoption of sensor-equipped mobile devices, Mobile Collaborative Sensing has been successfully adopted to enable an ever-increasing number of sensing applications in different domains, e.g., congestion detection of traffic [3], and industrial toxic gas detection [4]. The study of Mobile Collaborative Sensing is concerned with enabling to use distributed wireless devices to achieve applications that do not just rely on one type of dedicated wireless network infrastructure, e.g., wireless sensor networks, and cellular networks. The collaboration between different devices achieves the seamless information integration and sharing, for example, the toxic gas detection of a petrochemical plant, continuous information in space and time is necessary for real-time tracking. The objective of each Mobile Collaborative Sensing application is

1. Such mobile devices include phones, tablets, mobile robots, and unmanned aerial vehicles.

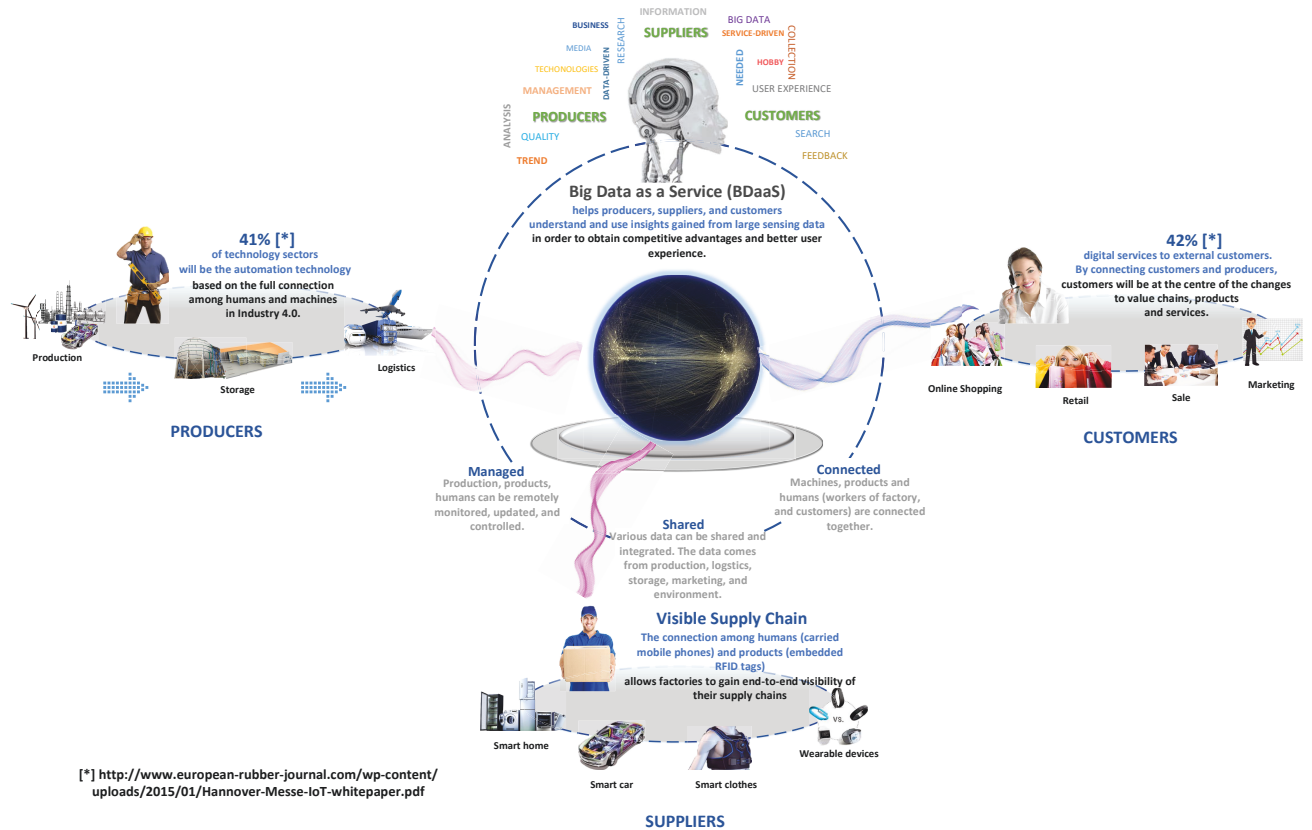


Figure 1.1: Potential impacts of Mobile Collaborative Sensing for the society and economy, and partial potential application domains, for example, Big Data as a Service (BDaaS), it helps the persons with different social roles (e.g., producers, suppliers, and customers) to understand and use insights gained from large sensing data in order to obtain competitive advantages and a better user experience. Mobile Collaborative Sensing fully integrates the information from different persons and devices, and enables collaborative sensing between persons and between persons and devices.

quite different with others, considering the specific requirements for sensing. However, these different applications have the common requirement on the collaboration between different wireless devices.

Process of Mobile Collaborative Sensing. For each Mobile Collaborative Sensing application, the target sensing area can be divided into a set of subareas. The sensing cycles also need to be specified for each subarea, for example, each cycle lasts for an hour in subarea i , and each cycle lasts for two hours in subarea j . The sensing devices of different subareas collaborate with each other to reduce their efforts on sensing, e.g., reducing energy consumption. For example, by the duty-cycled mechanism of sensing devices between different subareas, energy can be saved during a sensing process. Considering each sensing cycle and subarea, the objective of each participant is to collect certain environmental information in respective subarea in each sensing cycle, with the goal of

achieving the full coverage for the target sensing area to support specific Mobile Collaborative Sensing applications. Taking an industrial Mobile Collaborative Sensing task as an example, toxic gas detection of petrochemical plants: (i) considering the different levels of importance, the target sensing area can be divided into several subareas, e.g., production area, purifying area, storage area, logistics management area, and administration area, (ii) in each area, according to the real situation, the specific sensing cycle can be decided. In the most dangerous production area, the sensing cycle is longer compared with other areas. The sensing devices of different subareas have a duty-cycled mechanism to collaborate with each other for energy saving, and (iii) the application aims at collecting sensed gas concentration from each subarea in each sensing cycle. According to the gas concentration from different areas and different cycles, concentration change can be tracked among these different areas as time goes on.

While the objectives of Mobile Collaborative Sensing might be different due to the different targets and requirements of applications in collecting data, the design of Mobile Collaborative Sensing applications usually follows a similar paradigm. In general, the life cycle of an Mobile Collaborative Sensing process consists of four phases: dividing the target sensing area, deciding the specific sensing cycle in each divided subarea, and uploading and integrating the sensed data from different subareas and cycles.

The functionality of each phase is described as follows in detail:

- *Dividing*: Any Mobile Collaborative Sensing application contains two elements: a target sensing area, and sensing participants. The requirement of Mobile Collaborative Sensing to the target area is to achieve full coverage, and to minimize the overlap of different subareas. The target sensing area can be divided according to: (i) different levels of importance, and (ii) the activity area of each participant, for example, in a target sensing area, there are 100 participants, and they are active in their respective areas. For minimizing the overlap, the target sensing area needs to be divided by considering the coverage of different participants, and the full coverage needs to be met. Such division helps to achieve this target: collecting the overall information of the target area, and avoiding overlapping information.
- *Deciding*: The specific sensing cycle needs to be decided in each divided subarea. Because of the importance, each subarea has different sensing cycles. Finding an appropriate sensing cycle in each subarea can help save sensing efforts of participants, e.g., the energy consumption of sensing devices. There is a collaborative process between the devices of different subareas. For example, in a gas leakage event, to track the gas concentration, with the diffusion of the gas, the sensing cycle needs to be dynamically adjusted in each subarea, and for saving energy, the sensing devices of different subareas collaborate with each other to achieve duty-cycled sensing.
- *Uploading and Integrating*: Each participant uploads the sensed data that is collected in specific cycles and the corresponding subareas. Then, this phase takes the data collected from all the participants as input, analyzes the data correlation between different areas and cycles, and provides the integration of different sensing data from different subareas and cycles.

1.2 Research Motivation

Considering the aforementioned background and the difference between Mobile Crowd Sensing and Mobile Collaborative Sensing (Table 1.1), the research of this thesis is based on the following evolution trend of sensing: the collaboration between heterogeneous devices is necessary. These devices can be from Internet of Things (IoT), Internet of Vehicles (IoV), and even mobile networks. By considering the particularity of Mobile Collaborative Sensing in usage scenarios and devices, its definition is provided in Definition 1.

Table 1.1: Differences between Mobile Crowd Sensing and Mobile Collaborative Sensing

	Mobile Crowd Sensing [5]	Mobile Collaborative Sensing [4]
Participant Objects	Humans carried mobile devices	Humans carried mobile devices + other devices (e.g., static sensor nodes, sensor-embedded mobile robots, and sensor-embedded unmanned aerial vehicles)
Communication Mode	Device to Centre (D2C): mobile devices communicate with a data processing centre.	Device to Device (D2D) and Device to Centre (D2C): all devices (mobile and other devices) communicate with each other, and communicate with a data processing centre.
Collaboration: Between Different Devices	No	Yes
Crowd Wisdom	Yes	Yes
Participatory Sensing	Yes	Yes
Mobility	Yes	Yes

Definition 1 *Mobile Collaborative Sensing is a sensing paradigm that allows heterogeneous devices to participate in sensing and sharing the information around participants. In such sensing, although there are various types of devices involved, mobile devices play the major role. As an advantage, mobility makes such sensing be better in flexibility and extendibility. During the sensing, by the communication and cooperation between different devices, these devices collaborate with each other to complete a sensing task.*

Figure 1.2 is an instance of the Mobile Collaborative Sensing system to achieve the collaboration of an industrial ecosystem.

In this industrial sensing and monitoring scenario, there are a large number of: (i) static wireless devices embedded in industrial equipment, e.g., sensor nodes, and smart meters, (ii) mobile wireless devices, e.g., smart phones, smart helmets, and wearable sensors. These static and mobile devices, and workers have to work together seamlessly, and these devices and workers are not just independent collectors of surrounding physical information (crowd sensing). On this basis, mobile-device-participant collaborative sensing is proposed.

From Definition 1, these three important aspects of Mobile Collaborative Sensing can be observed to complete a sensing task. By proposing a suitable framework and algorithms to solve the problems

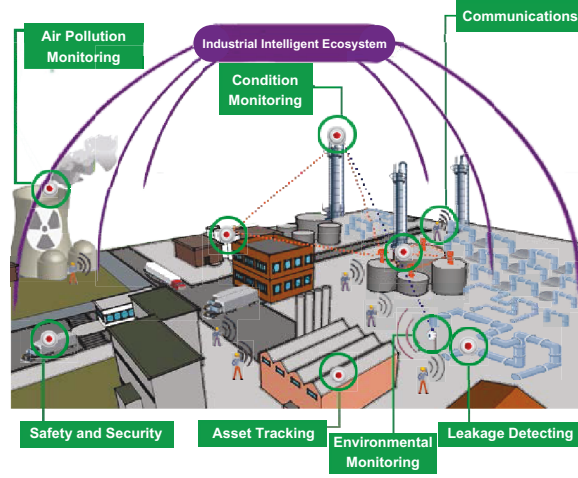


Figure 1.2: Collaborative Industrial Ecosystem. Different types of devices collaborate with each other to sense and monitor a large industrial ecosystem. These heterogeneous devices communicate and share their respective information with each other, and the workers can share their knowledge and experience in this ecosystem.

of these three aspects, it is possible to take the full advantage of the ability of Mobile Collaborative Sensing.

- Participation of heterogeneous devices (device collaboration). How to make these devices collaboratively and seamlessly work together? During a sensing task, as a specific advantage, in Mobile Collaborative Sensing, there are many different types of devices to sense and cover the physical information of an interested area. These devices work together under a uniform framework, and this framework enables the number of and the types of devices to be flexible and extendible.
- Location information acquisition and optimization. As a specific advantage of Mobile Collaborative Sensing, mobility can help us to improve the efficiency of sensing, for example, in a large interested area, only a certain number of devices are needed to cover this area by the flexible move of mobile devices. During the process of moving, location information is important to make such mobility efficient, for example, if there is no location information, the physical information collected by devices cannot be used to reflect where what is happening.
- Dynamic sensing. Timeliness and spatio-temporal continuity is the important aspect in the Mobile Collaborative Sensing based applications. For example, to track the disease propagation, the network topology is different at different time points. How to track such dynamics is a key to enable the disease propagation.

As an important example, in the industrial space, Mobile Collaborative Sensing is necessary and effective to make the different industrial sectors seamlessly work together by sharing respective information. It brings these benefits for industry:

- Mobile Collaborative Sensing enables the industrial value chain to be more efficient, and it fully integrates suppliers, producers, and customers. Moreover, it enables the collaborative sensing between humans, and between humans and machines. As a typical example of this benefit, Ford’s PowerShift dual clutch transmission, which is launched in 2010, meets a big quality problem. Because there is no any direct connection between the Ford’s production line (producers) and end-users (customers), they didn’t explain to consumers that PowerShift is based on manual transmission. This blunder makes users deem that the PowerShift would automatically change gears. So, when the PowerShift didn’t behave in a manual way, Ford’s quality scores took a big hit.
- Based on the above-mentioned full integration of suppliers, producers, and customers, Big Data as a Service (BDaaS) can be achieved to help producers, suppliers, and customers understand and use insights learned from large sensing data, in order to obtain competitive advantages and better user experience.
- Mobile Collaborative Sensing based BDaaS can be achieved to: (i) manage production, products, and humans, (ii) share various data from different sectors, e.g., production, logistics, storage, and marketing, and (iii) connect machines, products and humans.

On this basis, then, why Mobile Collaborative Sensing rather than Mobile Crowd Sensing?

- To satisfy the integration requirements from heterogeneous sensing devices. With the evolution of sensing and wireless networking techniques, various sensing devices are used in our daily lives and industry. These devices have different functions, and use different communication standards, so, as a further step, Mobile Collaborative Sensing is needed to achieve the integration requirements from these different sensing devices. Mobile and static sensing devices are all the objects of collaborating: in a sensing task, mobile devices play a major role, and static devices are auxiliaries.
- To achieve the collaboration between different sensing devices, not just data collection for completing a sensing task. In Mobile Collaborative Sensing, the devices can collaborate with each other to share information for meeting specific requirements. For example, to save the energy of sensing in industrial monitoring, the sensing devices of different industrial sectors need a duty-cycled sleeping mechanism. By the collaboration between different devices, for example, if the monitoring ranges of nodes s_i and s_j cover the same area, when s_i is awake, s_j is able to sleep; the optimized sleeping scheduling is obtained. By using this scheduling, both sides can be taken into account: the saving of energy, and the continuity of monitoring. In Mobile Crowd Sensing, each device is used to sense the physical information around them, and the sensed data is respectively submitted by each device, to complete a sensing task.

1.3 Contributions

Based on the aforementioned research motivation, the Mobile Collaborative Sensing framework is proposed to address the collaborative problem among devices. This framework facilitates the collaboration among different devices by integrating and serializing the massive spatio-temporal data from different devices and time points. In addition to designing the framework, this thesis contains the other three contributions:

1. **MCS-Locating** - In this contribution, a localization algorithm is proposed, LiCS (Locating in Collaborative Sensing based Data Space). For LiCS, the Received Signal Strength (RSS) from different devices is used to train a location estimation model: crowd-sourced data is used to train the model, and the model is installed in distributed servers, so the collaboration between different servers is necessary to provide a locating service for an individual. Any individual can use the trained model to estimate his/her location, and then his/her RSS can be used to train the model again. With the crowd-sourced RSS, the model is trained to make its parameters accordant with the practical situations. The prototype system of LiCS is implemented. Experimental results show that LiCS achieves comparable localization accuracy to previous approaches even without any special hardware. As another important aspect of this contribution, the spatial distribution pattern of RSS is investigated. It is observed that RSS is extremely sensitive to environment, for example, even in the same position, the fluctuation of RSS is very big.
2. **MCS-Prediction** - In this contribution, a prediction algorithm is proposed for disease dynamics. As an important aspect of propagation dynamics, predicting disease dynamics during an epidemic is important in e-Health applications. In such prediction, Realistic Contact Networks (RCNs) have been widely used to characterize disease dynamics. The structure of such networks is dynamically changed during an epidemic. Capturing such kind of dynamic structure is the basis of prediction. With the popularity of mobile devices, it is possible to capture the dynamic change of the network structure. By the collaboration between different devices, a propagation network can be constructed to model and reflect the RCN. For example, two nodes a and b locate at L_a and L_b at time point t , and the edge between them denotes the relationship between these two nodes. On the basis of such a propagation network, in this contribution, the impact of the network structure on disease dynamics is evaluated. These devices are carried by the volunteers of Ebola outbreak areas. Based on the results of this evaluation, a model is designed to recognize the dynamic structure of RCNs. On the basis of this model, the prediction algorithm is proposed.
3. **MCS-Routing** - In this contribution, an eco-friendly navigation algorithm is proposed, eRouting, to save energy and reduce CO₂ emission. The important research issue of traffic information industry, eco-friendly navigation has been widely studied. As an improvement, in this contribution, combining real-time traffic information and a representative factor based energy/emission model, a calculated route is dynamically adjusted during the travel of a vehicle. The traffic velocity v

is used as the real-time traffic information. It is estimated from the driving velocity of vehicles for the corresponding traffic segment. In such an estimate, various wireless devices are used, for example, Automatic Number Plate Recognition (ANPR) cameras, in-vehicle and GPS-enabled wireless devices, and inductive loops built into road surfaces. Based on various devices, massive real-time data can be collected, regarding traffic information from different traffic segments. The collaboration between these different devices is necessary to estimate the traffic velocity of a traffic segment. For example, vehicle a crosses a recognition device at time t_a and driving velocity v_a , vehicle b crosses the recognition device at time t_b and driving velocity v_b , and the traffic velocity in this time period $t_b - t_a$ is $\frac{v_a + v_b}{2}$.

1.4 Organization of this Thesis

The rest of this thesis is organized as:

- **Chapter2** gives a comprehensive survey on the state of the art of Mobile Collaborative Sensing, including the related work of (a) recent applications of Mobile Collaborative Sensing, (b) heterogeneous devices and corresponding communication standards, which are used in Mobile Collaborative Sensing, and (c) two closely related sensing paradigms: Participatory Sensing and Mobile Crowd Sensing. They are the foundation of Mobile Collaborative Sensing, and Mobile Collaborative Sensing is the extension of these two sensing paradigms. Because of the evolution of requirements, in many scenarios, a more open sensing paradigm is needed to integrate various devices, and let them collaborate with each other to work together under a unified framework.
- **Chapter3, Chapter4, Chapter5, Chapter6** present our work, Mobile Collaborative Sensing Framework, MCS-Locating, MCS-Prediction and MCS-Routing respectively, where these are introduced: (a) the motivation example and typical applications of Mobile Collaborative Sensing, open issues and research challenges, (b) research objectives/problem formulation for algorithm design, (c) detailed framework/algorithm design, and (d) evaluation results by using the real world data sets/prototype system.
- **Chapter7** summarizes the designed framework and three algorithms. These three algorithms are for applications of different three domains.

State of the Art Review

Contents

2.1	Mobile Collaborative Sensing	27
2.2	Mobile Crowd Sensing	29
2.3	Heterogeneous Devices and Communication Standards	30
2.4	Participatory Sensing	32

2.1 Mobile Collaborative Sensing

Collaborative sensing has been well studied in mobile sensor networks. There has been much recent research leading to the development of many different applications in mobile collaborative sensing [6, 7], for example, environmental monitoring [8], social networking [9], healthcare [10], and transportation [11]. These applications can come from different domains. Table 2.1 classifies the typical applications of mobile collaborative sensing, and provides some instances.

Table 2.1 shows six types of applications classified by domains, and in these applications various sensors, mobile phones, and sensor-embedded devices are largely used.

The applications in the urban transportation domain mainly focus on these three aspects: parking, traffic, and trajectories. Although the wide availability of sensors offers very interesting opportunities for mobile collaborative sensing in the urban transportation, there are special challenges that need to be tackled in this domain: (i) exploitation of the data from different types of sensors in dynamic transportation; (ii) on-line analysis of large-scale data in this dynamic transportation. How to effectively exploit and integrate all the information that sensors can provide is still an unsolved issue; (iii) management of trust and privacy. Collaborative sensing for transportation implies the collection and storage of partial data which is about the daily trajectories of private vehicles. The applications in the social event

Table 2.1: Applications of Mobile Collaborative Sensing

Applications	Domains	Used Devices
Parking, traffic, and trajectories [12]	Urban transportation	Physical sensors, virtual sensors, social sensors, and human sensors
MoVi [13]	Social event detection	Mobile phones
VUPoints [14]	Social event detection	Mobile phones
SignalGuru [11]	Traffic signal schedule	Mobile phones
Locale [15]	Localization estimation	Mobile sensors
E-Gesture [16]	Gesture recognition	Hand-worn sensor devices and mobile phones
Pickle [17]	Privacy-preserving collaborative learning	Mobile phones

detection domain have several limitations in its current form: (i) the number of triggers is limited and may not be sufficient to capture all the socially interesting moments that arise. Improved information processing is necessary to identify complex patterns that are together indicative of a prospective event; (ii) even if most events are captured, some important moments may be missed; (iii) energy and privacy concerns with systems are certainly open questions. Continuous sensing on multiple sensors, as well as periodic communication to data centres, is likely to drain phones' batteries. However, such continuous sensing is necessary to continuously capture social events; (iv) a realistic social function may pose greater challenges in grouping and trigger detection; however, the collaboration between more phones may greatly improve the efficacy of zone demarcation and trigger detection. For the applications in the traffic signal schedule domain, uncontrolled environment composition and false detection are two main challenges in this special domain. Windshield-mounted smartphones capture the real world while moving. As a result, there is no control over the composition of the content captured by their video cameras. The applications in the localization estimation domain use sensor readings to locate and track targets, and then rely on collaboration to merge observations. The efficacy of the merge is the important aspect to improve the localization accuracy. An important challenge in the applications of the gesture recognition domain is to support gesture-based functions in energy-limited mobile and sensor devices. In some applications of the collaborative learning domain, some privacy information has to be used to conduct collaboration.

To support the above-mentioned applications, many relevant algorithms, systems/platforms and

frameworks have been proposed [13, 14, 18]. For example, [13] envisions a social application where mobile phones collaboratively sense their ambience, and recognize socially interesting events; [14] proposes VUPoints, a collaborative sensing and video-recording system that takes advantage of the convergent ability of mobile phones¹. In VUPoints, mobile phones must sense their surroundings and collaboratively detect events that qualify for recording; [18] presents a collaborative mobile sensing framework namely Mobile Sensor Data EngiNe (MOSDEN) that can operate on smart phones capturing and sharing sensed data between multiple distributed applications and users. MOSDEN has been implemented on Android-based mobile platforms to achieve the collaborative operation in mobile opportunistic sensing applications.

2.2 Mobile Crowd Sensing

As a similar concept, there are many differences and similarities between Mobile Crowd Sensing and Mobile Collaborative Sensing. It means that these two sensing paradigms will have some of the same specialties. On this basis, Mobile Collaborative Sensing can learn a few things from Mobile Crowd Sensing, e.g., solutions and corresponding algorithms for a specific sensing problem.

Successful large-scale urban and industrial management relies on efficiently sensing and acquiring the physical information of surroundings for decision and policy making. To achieve this, traditional sensing techniques usually leverage distributed sensors to acquire real-world conditions [19]. However, the spatial coverage of the currently deployed sensor networks in the real world is far from enough, and the scalability and mobility of such networks is insufficient [20]. Along with the evolution of requirements on sensing, Mobile Crowd Sensing is proposed as a kind of potentially effective sensing paradigm. This sensing paradigm has four impressive specialties: cost effectiveness, coverability, scalability and mobility. These four specialties make Mobile Crowd Sensing effective and suitable to be widely used in various applications. There is a real and typical story about how the feedback of customers impacts the industrial quality tracking. By Mobile Crowd Sensing, the feedback from customers can be easily and quickly got to help the industrial quality tracking. Ford's PowerShift dual clutch transmission, which is launched in 2010, meets a big quality problem. Because there is no any direct connection between the Ford's production line and end-users, they didn't explain to consumers that PowerShift is based on manual transmission. This blunder makes users deem that the PowerShift would automatically change gears. When the PowerShift didn't behave in a manual way, Ford's quality scores took a big hit.

Mobile Crowd Sensing is a large-scale sensing paradigm by making good use of the power of human-accompanying mobile devices [5]. It enables a large number of humans to share surrounding information and their experiential knowledge by sensor-enabled mobile devices. It is a kind of sensing paradigm to be able to ensure the spatial coverage and cost effectiveness, because of these four inherent properties of Mobile Crowd Sensing: (i) mobile devices can be moved to any place by users to get enough coverage,

1. Mobile phones are becoming into a convergent platform for sensing, computation, and communication.

and (ii) if the number of mobile devices is not enough to sense the entire target area, more mobile devices can be added anytime and without any additional efforts.

A broad range of Mobile Crowd Sensing based applications are thus enabled, and most of relevant studies put their attention to urban space, including urban environment monitoring [21], mobile social recommendation [22], public safety [23], traffic control and planning [24], geospatial information gathering [25], and so on. Mobile Crowd Sensing is also friendly to solve the problems in large-scale industrial environment.

A formal definition of Mobile Crowd Sensing is described in Definition 2.

Definition 2 (Mobile Crowd Sensing) *Mobile Crowd Sensing is a sensing paradigm that allows crowd participation to sense and share the information around participants by mobile devices. Such sensing is without having to be connected to a fixed physical link, thus it is better in flexibility and extendibility. Meanwhile, it is without special devices and communication protocols. It means that the participants of such sensing are easy in networking, thus it is better in cost effectiveness.*

2.3 Heterogeneous Devices and Communication Standards

For Mobile Collaborative Sensing, these objects are potential participants: (i) the humans carried mobile devices, and (ii) various sensing devices, e.g., static sensor nodes, sensor-embedded mobile robots and unmanned aerial vehicles, and sensor-mounted vehicles. These heterogeneous devices have different functions. For achieving the communications between these heterogeneous devices, these communication standards are able to be used in Mobile Collaborative Sensing. The corresponding standards are listed in Table 2.2.

Bluetooth is a wireless technology standard for exchanging data over short distances from fixed and mobile devices, and building personal area networks (PANs). Bluetooth Low Energy (BLE) is a wireless personal area network technology designed by the Bluetooth Special Interest Group aimed at novel applications in healthcare, fitness, beacons, security, and home entertainment industries. It is a version of Bluetooth designed for lower-powered devices. WiFi is a technology for wireless local area networking with devices which are based on IEEE 802.11 standards. Three standards are for the 2G data transfer of mobile phones. GSM (Global System for Mobile Communications) is a standard which is developed to describe the protocols for second-generation (2G) digital cellular networks used by mobile phones. General Packet Radio Service (GPRS) is a packet-oriented mobile data service on the 2G and 3G cellular systems for mobile communications (GSM). Enhanced Data rates for GSM Evolution (EDGE) is a digital mobile phone technology that allows improved data transmission rates as a backward-compatible extension of GSM. For the 3G data transfer of mobile phones, Universal Mobile Telecommunications Service (UMTS) is a 3G broadband, packet-based transmission service at data rates up to 2 megabits per second (Mbps). UMTS is based on the GSM communication standard. High Speed Packet Access (HSPA) is an amalgamation of two mobile protocols, High Speed Downlink

Table 2.2: Typical Communication Standards

Typical Communication Standards	Descriptions
Bluetooth: IEEE 802.15.1 [26]	Bluetooth is a wireless technology standard for exchanging data over short distances.
Bluetooth Low Energy (BLE): IEEE 802.15.1 [27, 28]	BLE is a version of Bluetooth designed for lower-powered devices.
WiFi: IEEE 802.11 [29–31]	WiFi is a technology that allows devices to connect to a Wireless Local Area Network (WLAN).
2G: GSM (Global System for Mobile Communications)/GPRS (General Packet Radio Service)/EDGE (Enhanced Data Rate for GSM Evolution) [32], 3G: UMTS (Universal Mobile Telecommunications Service)/HSPA (High Speed Packet Access) [33, 34], 4G: LTE (Long Term Evolution) [35]	Different standards of data transfer for mobile phones.
ZigBee: IEEE 802.15.4 [36]	ZigBee is an IEEE 802.15.4 based specification for a suite of high-level communication protocols used to create personal area networks with small, low-power digital radios.
Z-Wave [37]	Z-Wave is a wireless communication protocol for home automation.

Packet Access (HSDPA) and High Speed Uplink Packet Access (HSUPA), that extends and improves the performance of existing 3G mobile telecommunication networks using the WCDMA (Wideband Code Division Multiple Access) protocols. Long-Term Evolution (LTE) is a standard for the 4G data transfer of mobile phones, and it is based on the GSM/EDGE and UMTS/HSPA technologies. For small and low-power devices, Zigbee is an IEEE 802.15.4-based specification for a suite of high-level communication protocols used to create personal area networks. Z-Wave is a wireless communication protocol used primarily for home automation. A Z-Wave automation system can be controlled via the Internet, with a Z-Wave gateway or central control device serving as both the Z-Wave hub controller and portal to the outside.

2.4 Participatory Sensing

Participatory sensing is the process which uses individuals and communities to collect and analyze systematic data [38]. This sensing paradigm uses sensing-capable mobile phones and cloud services to complete a sensing task.

Participatory sensing consists of two important technical components, ubiquitous data capture and data processing [39].

Ubiquitous data capture. For collecting the empirical data from our daily lives, the mobile phones carried by humans can be used as a kind of unprecedented tool for this data collection job. Mobile phones have become a kind of mobile computing, sensing, and communication platform to capture image, audio, video, motion, proximity, and location data, and to share such data by data communications. They are capable of being programmed for manual, automatic, and context-aware data capture and sharing, because of the ubiquity of mobile phones and associated communication infrastructure. Moreover, it is possible to include people of all backgrounds nearly everywhere in the world. Because mobile phones travel with us, they can help us make sustainable observations on an personal level, and they provide unmatched coverage in space and time.

Ubiquitous data processing. In some cases, the data collected by a mobile device is enough to reveal an interesting pattern on its own. However, for inferring the complex phenomena of individuals and groups, a series of interaction data among users is necessary. So how to analyze such ubiquitous data, and ensure the robustness of such analysis, are both central requirements for participatory sensing.

Mobile Collaborative Sensing Framework

Contents

3.1	Introduction	34
3.2	Definitions and Advances	34
3.2.1	What is Collaborative Intelligence?	34
3.2.2	What is Sensing Intelligence?	35
3.2.3	Advances	36
3.3	Mobile Collaborative Sensing Framework	40
3.3.1	Key Components of the Mobile Collaborative Sensing Framework	41
3.3.2	On-Going Efforts	43
3.4	Key Challenges and Open Issues	47
3.4.1	Key Challenges	47
3.4.2	Open Issues	49
3.5	Conclusion	50

A collaborative sensing framework enables different sensing-capable devices to work together. Under this framework, different sensing participants constitute an intelligent ecosystem to collaboratively collect, share, and analyze data. From the perspective of data, collaboration is a kind of spatio-temporal data collection, sharing, and analysis to cooperatively complete a sensing task. As a typical example, for monitoring a large-scale petrochemical plant, a large number of sensor nodes are deployed in different areas, and these nodes construct a monitoring network. In this network, keeping all of the nodes in the waking state is not necessary, and it causes unnecessary energy consumption, for example, there is a potential dangerous source in area A, and area B is the neighbor of A; the sensor nodes in area B do not need to wake up until the sensor nodes in area A detect danger. On the basis of this

example, it is clear to understand: the collaboration between different devices is necessary, and such collaboration can improve the performance of a monitoring network.

3.1 Introduction

Various sensors and sensing-capable devices have been widely deployed in the real physical world [40, 41]. Massive spatio-temporal data is being collected from these sensors and devices on a daily basis. The kind of solutions by collaboratively analyzing the spatio-temporal data are flexible and low-cost, because only data and corresponding data analysis algorithms are necessary. How to analyze the spatio-temporal data to improve the effectiveness of solutions is a valuable research issue: for a real problem¹, firstly, massive data can be collected from the corresponding domain, and such data has a strong relationship with the real problem. Then a solution for this real problem can be designed by analyzing the massive data. The performance of the solution is decided by the quality of the collected data, and the performance of the corresponding data analysis algorithm.

This work designs a Mobile Collaborative Sensing framework. This framework facilitates the cooperativity of data collection, sharing, and analysis. The data is collected from different sources and time points. For the real deployment of the framework, the key challenges and open issues are discussed.

The scientific contributions of this work are listed as follows.

- This work clearly answers why and how to design a Mobile Collaborative Sensing framework. The key components of this framework are described in detail.
- Two on-going efforts on developing the framework are introduced and discussed. This framework aims to achieve the dynamic collaboration between different objects, and such collaboration is based on massive spatio-temporal data.
- The challenges and open issues on developing and realizing the framework are analyzed and listed.

3.2 Definitions and Advances

As the basis of a Mobile Collaborative Sensing framework, Collaborative Intelligence (CI) and Sensing Intelligence (SI) are clearly defined, and their advances are discussed in this section.

3.2.1 What is Collaborative Intelligence?

By answering these two questions, the term CI can be clearly defined: (i) what is and why do we need intelligence? and (ii) what is and why do we need collaboration from the point of view of realizing intelligence?

1. The problem exists in the real physical world. Solving this kind of problem can help to improve the performance of a real system.

Intelligence is developed to minimize the manual intervention of a series of complex and dangerous operations. It is different from automation [42]. Intelligence can be defined as the ability to acquire information or knowledge from humans and machines, and apply the acquired information or knowledge to construct a deliverable problem-solving network.

In a complex and dangerous operating environment, there are a large number of humans and heterogeneous devices. To acquire information or knowledge, and realize intelligence, the collaboration among the humans and heterogeneous devices is important and necessary. That is, the intelligence on a series of complex and dangerous operations is a battery of collaborations among the relevant humans and heterogeneous devices.

The definition of CI is described in Definition 3.

Definition 3 (Collaborative Intelligence) *Collaborative Intelligence is the ability to acquire information or knowledge from a large number of humans and heterogeneous devices, to construct a problem-solving network². By using this network, the purpose of collaborative intelligence is to realize the automation of complex and dangerous operations, or to improve the performance of the automation.*

3.2.2 What is Sensing Intelligence?

Considering fully mining the potential of sensors and sensing-capable devices, SI is defined, and the corresponding definition is described in Definition 4.

Definition 4 *Through dynamically mining and analyzing the spatio-temporal sensing data collected by sensors and sensing-capable devices, useful information or knowledge can be acquired to improve the ability of automation.*

This definition considers the spatio-temporal continuity of the sensing process. Such consideration reflects the dynamic nature of sensing. From the data perspective, continuous spatio-temporal data is collected from relevant sensing-capable objects.

Definition 4 has taken into account these two main aspects:

- Mining and analyzing spatio-temporal sensing data. There are various sensors and wireless devices to sense surroundings and to collect the data from different data sources and time points. Then the collected data is mined and analyzed according to certain logic.
- Acquiring useful information or knowledge. This is the important step to achieve intelligence. Automation is the first step of realizing intelligence. By integrating the acquired useful information or knowledge into automation, the intelligence can be realized.

From the definition of SI, it is obvious that SI consists of physical sensing, data mining and analysis, as well as information/knowledge acquirement and utilization.

2. A problem-solving network is proposed to exploit the potential of the collaboration among humans and heterogeneous devices, and transfer information-intensive organizations to a network society. It is set for solving problems rather than building relationships.

3.2.3 Advances

3.2.3.1 Collaborative Intelligence

CI is able to utilize extensive information or knowledge to construct a problem-solving network. On the basis of the network, collaborative intelligent systems are constructed to realize the automation of complex and dangerous operations.

CI involves extensive collaboration among different members to be as an efficient team for problem solving. Such collaboration possesses great potential on problem solving under challenging environments [43], because it obviously can provide more information or knowledge to design improved solutions than any single member could. It achieves the flexibility of how members are deployed. It provides a non-stop real-time learning opportunity for a team. For example, different members of collaboration to achieve the flexibility: they do not have to in the same position, and can do communications by heterogeneous devices; and these members can conduct collaborative sleep scheduling: by this kind of scheduling, a team can achieve non-stop workings. Moreover, such collaboration has the potential of integrating diverse contributions into a platform to produce a creative solution for successfully solving a problem [44]: different members contribute different information or knowledge, skill and experience to problem solving.

Based on the above advantages, CI has been widely studied. As an important existing platform of CI, HUB-CI (HUB with CI) [45] is developed at Purdue University. Based on this platform, Prabhu Devadasan *et al.* have designed the model Collaborative Intelligence Measure of KBS³ (CIMK) that measures CI by the multi-objective optimization on the parameters of collaboration, and suggests the optimal operating points for various clients, with great flexibility.

The advance of CI is briefly discussed. Relevant studies are classified in Table 3.1, and some typical literature is listed for each classification as examples. Additionally, several studies are discussed in detail to make the meaning of each classification easy to be understood.

Table 3.1: Classification of the Studies on Collaborative Intelligence (CI)

Classification	Typical Application	Typical Recent Literature
Human-based CI	Smart search and recommendation in social networks	[46–52]
IoT-based CI	Optimizing the performance of intelligent systems	[53–55]

In Table 3.1, the relevant studies can be classified into two classes, human-based CI and IoT-based

CI, depending on the difference of participants.

Human-based CI. As the typical applications of human-based CI, the smart search and recommendation of social networks have been widely studied.

In the literature [46,47], Vincent W. Zheng *et al.* have developed a mobile recommendation system to answer two popular location-related queries in our daily life: (i) if we want to do sightseeing or dining in a large city like Beijing, where should we go? (ii) if we want to visit a place such as the Bird's Nest in Beijing Olympic park, what can we do there? This system includes three important algorithms that are based on collaborative filtering to address the data sparsity problem⁴. The first algorithm uses a collective matrix factorization model to provide a recommendation, based on the merged data from all of the users. The second algorithm uses a collective tensor and matrix factorization model to provide a personalized recommendation. The third algorithm further improves the previous two algorithms by using a ranking-based collective tensor and matrix factorization model.

As the important supportive work of the above-mentioned achievement from Vincent W. Zheng *et al.*, in the literature [48], they have presented user-centred collaborative location and activity filtering (UCLAF) to merge the data from different users together, and have applied the collaborative filtering to find like-minded users and like-patterned activities at different locations.

IoT-based CI. As the typical application of such CI, optimizing the performance of intelligent systems has attracted attention.

In the studies of IoT and intelligent systems, the intelligent transportation system is an important aspect. In the literature [53], a collaborative framework is proposed for the real-time traffic information collection, fusion and sharing. The real-time traffic information is reported by various front-end devices of intelligent transportation systems, for example a vehicle-mounted GPS receiver. The framework integrates real-time traffic information from different data sources to be able to improve the performance of the intelligent transportation system, for example enabling the high-accuracy prediction for real-time traffic status.

As another important intelligent system, the intelligent healthcare service system, Byung Mun Lee *et al.* have introduced a collaboration protocol to share health information among IoT personal health devices [54]. By such information sharing, the quality of the healthcare service can be improved.

On the other side, the collaboration between different members perhaps results in serious mistakes. If a collaboration is not efficient and even incongruous, a minor mistake in this collaboration will fall into a syndrome known as “groupthink” [56], and the syndrome causes the mistake to be amplified, which results in a fiasco [57]. How to make a collaboration efficient is an important and difficult problem. The book [58] presents an approach. Its premise is that preliminary work is performed by professionals of intelligent community: mining information/discovering knowledge from the target work and members of a collaborative team. The effectiveness and correctness about this premise have been verified in the

4. Data comes from each user, and is used to make recommendation, but the data of each user is limited, which reduces the accuracy of recommendation.

research achievement [59].

3.2.3.2 Sensing Intelligence

With the development of IoT technology, sensing intelligence has drawn wide attention, on account of these advantages: (i) with the help of sensing intelligence, efficient monitoring can be achieved, and such monitoring is cost-effective, and (ii) with the help of sensors and wireless devices embedded in various machines and systems, the maintenance of these machines and systems is controllable and automatable; especially, these machines and systems are deployed in remote and hard-to-reach areas. Sensing intelligence has been successfully applied to many applications, such as monitoring, controlling, maintenance and security [60]. Typical industrial applications of sensing intelligence are introduced as follows.

Factory automation. A factory is a highly dynamic ecosystem, in such ecosystems, automation is indispensable. Traditional actuators combined with control units have been used in factory automation. With the development of wireless and sensor technologies, the adoption of Wireless Sensor Networks (WSNs) and Radio Frequency Identification (RFID) on the actuators and control units for factory automation has experienced impressive growth over the past decade [61, 62]. This is SI based factory automation.

In the manufacturing environment of a factory, two main activities are included, manufacturing operations and equipment maintenance [63]. In recent years, based on these two main activities, the studies on factory automation pay much attention to these four aspects [64, 65]: (i) the monitoring and controlling for manufacturing processes, (ii) the safety and maintenance for machines, (iii) the resource tracking for manufacturing workshops, and (iv) high-level logistics and supply chain management.

A SI based factory automation system consists of various devices, e.g., sensors, controllers, and heterogeneous machines, and these devices can be combined together through the communications between each other. The communication component is the most important part of factory automation. Table 3.2 lists the communication protocols that can be used in SI based factory automation.

Table 3.2: Relevant Protocols for Sensing Intelligence based Factory Automation

Wireless Communication Protocol	Relevant Standard	Maximum Data Rate (Mbit/s)	Maximum Data Payload (Bytes)
Bluetooth	IEEE 802.15.1	1	339
Ultra-WideBand (UWB)	IEEE 802.15.3	110	2044
ZigBee	IEEE 802.15.4	0.25	102
WiFi	IEEE 802.11a/b/g	54/11/54	2312

By using SI based factory automation, the ability of factory automation can be enhanced to achieve safe, efficient and eco-friendly factory production.

Energy industry. As another important application of SI, the application environment of the energy industry and factory automation is different. In the energy industry, SI is mainly applied to inaccessible environments to monitor and control industrial systems. In factory automation, SI is mainly applied to highly dynamic and large-scale environments.

With the development of sensing technology and the extensive deployment of sensors, the sensing intelligence supportive renewable energy industry (e.g., solar, tidal and geothermal energy) has become a new and important study aspect. The equipment for accessing renewable energy is often located in remote areas, such as mountains, seas and volcanoes. Despite this, real-time control is necessary for the units of energy harvesting, for example, for a wind turbine, based on the data from wind-direction sensors, a yaw-drive motor turns the nacelle to face the wind. Moreover, the sophisticated units that are embedded in equipment require frequent maintenance [66]. SI is proposed for both purposes, real-time control and maintenance, in the renewable energy industry [67].

- Real-time control. Based on the development of SI in real-time control, first, the real-time data of environmental conditions includes wind speed, temperature, humidity, rainfall and geothermal activity, and it can be collected by the spatially distributed sensors and wireless devices. These sensors and wireless devices are embedded in energy-harvesting systems. Then, by using the collected environmental data, the relationship between generated energy and different seasons can be analyzed. With the analyzed results, the optimal parameter configuration can be acquired and used to control the equipment that is the main component of the energy-harvesting system. In a word, by using SI, the process of energy harvesting is highly efficient and automatic [68]. Moreover, such real-time intelligent control has been used in smart home services as well [69].
- Maintenance. The sensors that are embedded in various units of equipment interact with the equipment to take a number of measures, such as the scheduling of maintenance [70], the reconfiguration of certain operations [71] and the emergency shutdown of equipment [72]. With the SI in maintenance, unnecessary downtime can be prevented, and equipment failure costs can be reduced.

In recent years, as the important part of the energy industry, the smart grid has attracted great attention of researchers. The smart grid represents a vision of the future electricity grid, and it is radically different from current electricity grids that have been deployed. It is an electricity grid that uses analogue or digital communication technology to collect information, and take action for automatically improving the efficiency, reliability, economic benefit and sustainability of the production and distribution of electricity [73]. In the literature [74], Ramchurn *et al.* have presented the delivery of the decentralized, autonomous and intelligent system, smart grid, as a great challenge for computer science and artificial intelligence research. As a typical case that is tightly related to SI in the smart grid, optimizing the electricity usage of electric vehicles is worth studying. For example, with analyzing the spatio-temporal trajectory data from an intelligent transportation system, the routing pattern of electric vehicles can be acquired, and then, a national electric supply company can make time- and area-

divisory electricity prices to control the usage of electricity and, therefore, to improve the efficiency of the smart grid.

3.3 Mobile Collaborative Sensing Framework

Why do we design the Mobile Collaborative Sensing framework? This framework improves the efficiency of a heterogeneous device platform. As an example, in industrial production/service, the internal logical processes are intricate and precise [75]. A large amount of different equipment is involved in these logical processes. For achieving the high efficiency of industrial production/service, effective collaboration is necessary among different equipment and different logical processes. The Mobile Collaborative Sensing framework can organize multi-sourced data, and make different data sources collaborative with each other based on the data. The multi-sourced data is collected from the different equipment and different logical processes of industrial production/service based on the IoT.

Effective collaboration is possible with the help of massive data. First, with the application of IoT technology, massive data can be collected by widely distributed sensors and wireless devices [41]. Then, as the natural advantage of data, different data is easily processed, and even merged together [76]. Finally, the effective collaboration among different equipment and processes can be achieved with processing and merging different data which comes from multiple sources.

Moreover, as an important example, the data based collaboration can cost-effectively develop the intelligence of industrial production/service [77]. For example, in the chemical industry, different equipment is used in different production stages, and different types of data are collected. For improving the ability of acquiring information or knowledge, and applying the acquired information or knowledge to realize the automation of production, the different equipment collaborating based on the data is an effective and low-cost method.

How do we design the Mobile Collaborative Sensing framework? Various sensors and wireless devices have been widely deployed to different equipment, and massive data is collected by these sensors and wireless devices. On this basis, the Mobile Collaborative Sensing framework is designed. Figure 3.1 illustrates the work flow of the Mobile Collaborative Sensing framework.

There are three main challenges to achieve this framework: (i) how to integrate the data from different sources, (ii) how to filter out noise to find the data we need, and (iii) how to get the data into the right hands to discover useful information/knowledge. CI empowers systems to intelligently transform vast amounts of operational data into actionable information/knowledge that is accessible and available anytime, anywhere.

Based on the available data from heterogeneous devices, how to construct a problem-solving network is an important and difficult problem, and constructing the problem-solving network is the main target and contribution of the Mobile Collaborative Sensing framework. As the common and important features of the heterogeneous data, time and location, can be used as collaborative parameters to

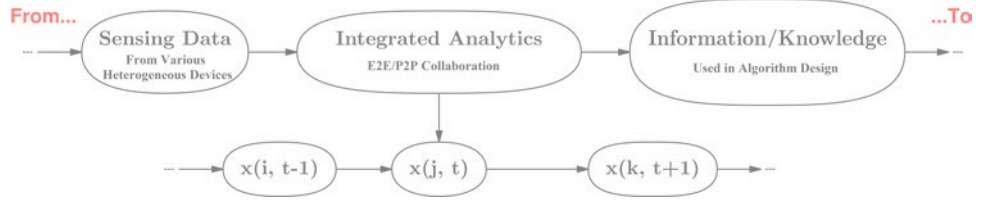


Figure 3.1: Work flow of the Mobile Collaborative Sensing framework. Sensing data is collected from various heterogeneous devices. Information and knowledge is obtained by analyzing the massive heterogeneous data, which can be used in algorithm design to solve the real-world problems. $[..., x(i, t-1), x(j, t), x(k, t+1), ...]$ is the spatio-temporal state sequence from different location's devices, which is used to achieve the collaborative analysis among heterogeneous devices. E2E denotes equipment-to-equipment collaboration, and P2P is for person-to-person. From sensing data to information/knowledge, heterogeneous data is integrated under a uniform framework.

integrate the heterogeneous data. A time or location sequence can be considered as a Markov chain. With the change of time or location, the state of a problem which needs to be solved undergoes transitions from one state to another in a state space, and the state space includes various current states from different relevant devices. With the help of the feature parameters of data, the data can be integrated to achieve the collaboration of heterogeneous devices, and the integrated data can be used to mine and discover useful and actionable information/knowledge. Such information/knowledge is used to algorithm design to solve corresponding problems. On this basis, the problem-solving network can be constructed.

3.3.1 Key Components of the Mobile Collaborative Sensing Framework

The Mobile Collaborative Sensing framework consists of three components (Figure 3.1): (i) sensing data collection, (ii) integrated analytics, and (iii) information mining and knowledge discovery.

Sensing data collection. A large number of heterogeneous sensing devices have been widely deployed to develop various sensing applications. Through these sensing applications, massive heterogeneous data has been collected. This component is the basis of integrated analytics, so collecting enough spatio-temporal data is important and necessary to get accurate analysis results for a sensing application. For example, before analyzing sparse data, data fitting is necessary, and the fitting data may cause the inaccuracy of data analysis compared with directly analyzing the real physical data.

Integrated analytics. Effective integration of heterogeneous data is an important and basic premise to mine/discover useful and actionable information/knowledge. How to make different objects collaborate with each other is an interesting research issue to realize the integrated analytics.

Any sensing application includes a series of processes and actions, and these processes and actions are location- and time-aware. A spatio-temporal Markov chain can be used to process the relationships between these processes and actions. Based on such processing, the collaboration among different

objects is achieved.

The detailed design and description of a spatio-temporal Markov chain [78, 79] is shown as follows. A series of processes and actions of a sensing application produces a series of different states, ..., $x(i, t - 1)$, $x(j, t)$, $x(k, t + 1)$, ..., where $x(., .)$ is the function of the parameters “location” and “time”. These states meet the Markov property that is described in Definition 5. The state transitions of sensing processes can be denoted by a spatio-temporal Markov chain, and the state transitions are based on the state space of the sensing application (an example is illustrated in Figure 3.1).

Definition 5 *A stochastic process has the Markov property, if the conditional probability distribution of future states of the process depends only on the current state, not on a series of preceding states. Therefore, the Markov property can be formulated as: let $\{X(t), t \geq 0\}$ be a time continuous stochastic process, which is assumed to be the set of non-negative integers, and then for every $n \geq 0$, time points $0 \leq t_0 < t_1 < \dots < t_n$, and states $x_0, x_1, \dots, x_n$, the process holds that $P(X(t_n) = x_n \mid X(t_{n-1}) = x_{n-1}, X(t_{n-2}) = x_{n-2}, \dots, X(t_0) = x_0) = P(X(t_n) = x_n \mid X(t_{n-1}) = x_{n-1})$.*

This definition shows that only the current state provides information to the future behaviour of the process. Historical states of the process do not add any new information.

Figure 3.2 provides an example to explain how to do data processing by the spatio-temporal Markov chain.

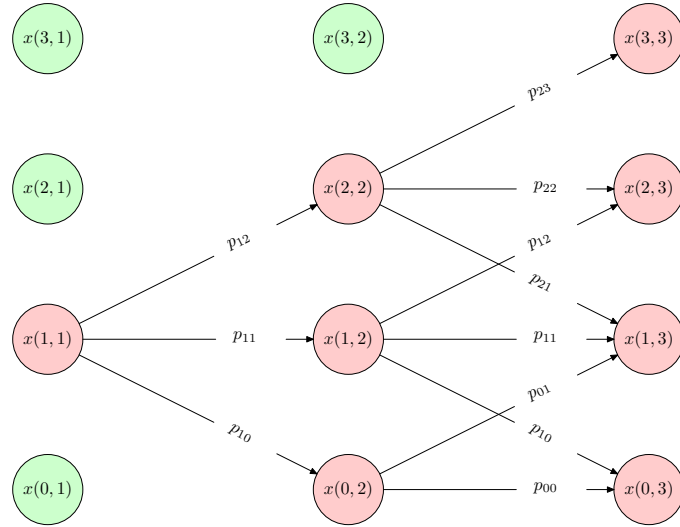


Figure 3.2: A spatio-temporal Markov chain on the processes of a sensing application. $P = \{p_{10}, \dots, p_{ij}, \dots\}$ ($i, j \in \{0, 1, 2, 3\}$) is the set of processes; $x(i, t)$ ($i \in \{0, 1, 2, 3\}, t \in \{1, 2, 3\}$) denotes the state space at the time t , and i is the location number of the device that has the state $x(i, t)$.

The spatio-temporal data of this example is a series of states, $x(i, t)$, and the states at different time points are linked by a set of processes (p_{ij}). As the most important information that can be used to link

two different states, the location and time stamp are included in each state. In this example, there are four states in the state space of the time point $t = 1$, $x(0, 1), x(1, 1), x(2, 1), x(3, 1)$. The state $x(1, 1)$ transfers to $x(0, 2), x(1, 2), x(2, 2)$, with corresponding processes p_{10}, p_{11}, p_{12} , and these transitions are based on certain probabilities. As time goes on, step by step, the Markov chain of these specific sensing processes can be achieved. Such a Markov chain enables the collaboration among different things and time points, based on the massive spatio-temporal data.

Information mining and knowledge discovery. By the integrated analytics, and considering: (i) the representative parameters of a special sensing application, and (ii) the spatio-temporal Markov chain that is based on the representative parameters, the rules about the processes of the application can be learned, and then, these rules form useful and actionable information/knowledge according to a particular logical sequence. Based on the mined information and the discovered knowledge, designing a targeted intelligent algorithm is feasible to solve a real-world problem.

3.3.2 On-Going Efforts

The Mobile Collaborative Sensing framework simplifies the integrated analytics among different data sources, and integrates these data sources with their individual semantics. By introducing and analyzing two on-going efforts, the details of developing the framework in different sensing applications are clearly presented.

3.3.2.1 Dynamic Detection of Toxic Gases

As an important part of industry, in large-scale petrochemical plants, the leakage of toxic gases is a serious threat to the surrounding citizens and environment [80]. It is necessary to develop an intelligent leakage detection solution for timely rescue and control.

However, in most existing large-scale petrochemical plants of China, wireless sensor nodes are deployed to detect toxic gases. These sensor nodes are independent of each other to report the concentration of toxic gases in their individual ranges. There are four disadvantageous aspects about the detection scheme which is based on independent sensor nodes to provide scattered reports:

- It is difficult to locate the leakage source of a toxic gas without tracking the change of concentration of the toxic gas. The concentration of a toxic gas is constantly changing as location shifts and time goes by. In such a dynamic environment, only using independent static sensor nodes, the change of the concentration cannot be tracked without the collaboration among different sensor nodes.
- It is difficult to track and monitor the active workers in a large-scale petrochemical plant. In a petrochemical plant, it is vitally important to identify the geographical locations of workers and to monitor the life signs (e.g., heart rate) of these workers when the leakage of toxic gases happens. The collaboration is necessary among different active workers to locate a worker and to

- estimate/predict the impact of the production environment on the health of the worker.
- For a certain sensor, it only can detect a kind of toxic gas, and in a detecting system, different sensors are needed to detect different toxic gases. In the complex environment of a petrochemical plant, it is hard to make an optimal decision about what certain types of sensors are needed in a certain location to detect certain toxic gases. In addition, a petrochemical plant is an uncertain environment, and under this environment, a chemical reaction is possible by mixing different toxic gases. This reaction produces new toxic gases that cannot be detected by the deployed sensor nodes. Moreover, embedding all possible sensors into a detecting system is not cost-effective and is impractical.
 - It is difficult to set the optimal threshold for the sensed reading of toxic gas concentration. For example, in a carbon monoxide sensor, the predefined threshold is x , and in an accident, the leaking source of the carbon monoxide gas is far away from this sensor. When the sensed reading of this sensor is larger than the predefined threshold x , the carbon monoxide gas has been widely diffused, and has already gotten out of control.

Based on the characteristics of industrial problems, the Mobile Collaborative Sensing framework is designed and used to solve existing problems in industrial systems. It is based on analysing massive spatio-temporal data from various devices of Industrial IoT environments.

Figure 3.3 illustrates an on-going effort, a Mobile Collaborative Sensing based system, which improves the capability of detecting toxic gases in a large-scale petrochemical plant.

As the important two components of this on-going effort, Figure 3.4 provides the details of sensor-embedded wearable wireless devices and static wireless sensor nodes.

In this on-going effort, first, with the daily walking of workers in a petrochemical plant, massive spatio-temporal data is collected by smart helmets, and the smart helmets collaborate with static sensor nodes via communication-enabled wrist watches. Then, the collected data by smart helmets and static nodes is submitted to a remote monitoring centre. Finally, the massive spatio-temporal data is analyzed by the Mobile Collaborative Sensing framework. Such an analysis enables the collaborative working among different wireless devices to construct a problem-solving network.

For the special problem, the leakage of toxic gases in large-scale petrochemical plants, because of the wide deployment of wireless devices, massive data is collected from these heterogeneous devices. The collected data includes different information from different locations and time points. By using the spatio-temporal data analyzing based framework, the widest detecting can be achieved as the efficient and cost-effective solution of the leakage problem.

3.3.2.2 Citizen Sensing of La Poste

Figure 3.5 provides an on-going effort: integrating two different data sources to improve the performance of services or solutions for mail delivery. This effort is based on citizen sensing and machine sensing. Based on sensing and communication operations, sensors share their data, which provides

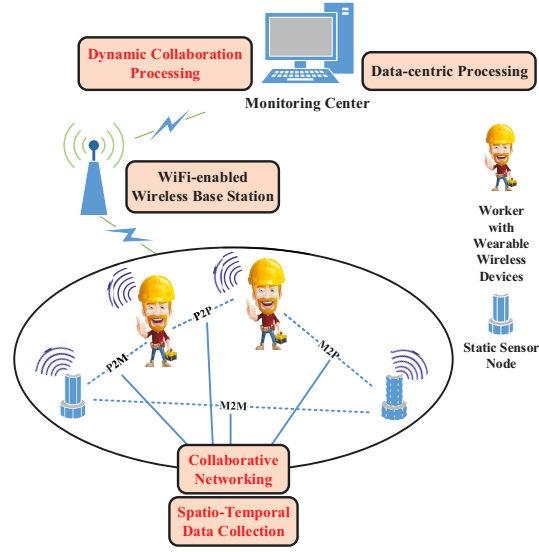
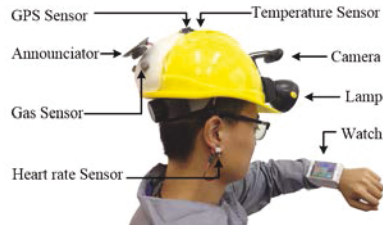


Figure 3.3: An application scenario of the Mobile Collaborative Sensing framework to improve the capability of detecting toxic gases in a large-scale petrochemical plant. There are four components in this application: sensor-embedded wearable wireless devices, static wireless sensor nodes, WiFi-enabled wireless base stations, and a remote monitoring centre. The wearable wireless devices are worn by workers and collaborate with static wireless sensor nodes to sense the surrounding environment and collect spatio-temporal data. The collected data is sent to the remote monitoring centre via WiFi-enabled wireless base stations. In the monitoring centre, by data-centric dynamic collaboration, the collaborative networking among different wireless devices can be achieved. Such networking constructs a problem-solving network to detect the leakage of toxic gases.



(a) Sensor-embedded wearable wireless devices: smart helmet and wrist watch



(b) Static wireless sensor node

Figure 3.4: The smart helmet is sensor-embedded, and it works with the wrist watch to detect toxic gases. The static node is supported by solar energy to persistently measure the concentration of gases, e.g., CO , SO_2 and CH_4 , and collect other environmental information, e.g., wind speed, humidity and temperature.

enhanced situational awareness that any system cannot offer alone.

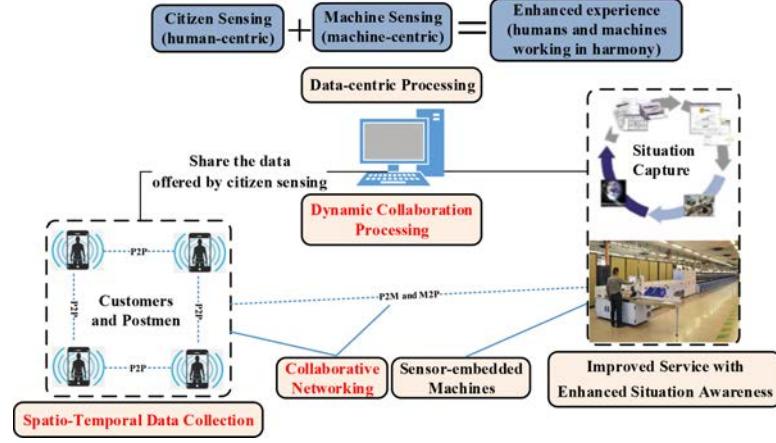


Figure 3.5: An effort of networking two different data sources to improve the quality of services (a use case from La Poste). Integrating the spatio-temporal data from citizen sensing with the data from machine sensing provides enhanced experience and situational awareness. Such integration forms more complete information than either form of sensing can provide alone. Additionally, it enables collaborative networking among different wireless devices.

The collaboration of different data sources provides the enhanced performance of services or solutions by integrating the data from different data sources, and the integration process is based on a certain logical sequence of these different data sources.

Moreover, observing: (i) the advances of CI and SI, and (ii) the on-going efforts on the Mobile Collaborative Sensing framework, a new trend is clear: realizing the interaction between the crowd wisdom of humans and sensing intelligence to solve complex real-world problems.

Sensing intelligence interacting with the crowd wisdom of humans.

- Participatory sensing. Burke *et al.* assert: participatory sensing will make deployed devices interactive, and participatory sensor networks enable heterogeneous sensor-embedded machines to collect, analyze and mine data, and then to discover and share individual knowledge [81]. In the era of big data, participatory sensing is the process where individuals and communities use devices or modules to collect and analyze systematic data to learn and discover knowledge [39].
- Crowd wisdom of humans. For example, as of March 2014, Twitter receives 500 million tweets per day, so mining the wisdom of crowds based on this type of big data has been made possible. To strengthen the decision-making ability of sensing systems, as an effective strategy, interacting with the crowd wisdom of humans has attracted the attention of researchers [82], and the strategy has the prospect of improving the ability of sensing intelligence [83].

In summary, the production/service of industry consists of a series of complex processes. High safety, efficiency and eco-friendliness are required during such production/service. However, how to

make industrial environments and machines be safe, and how to improve the efficiency of industrial production/service, are long-term challenges. Meanwhile, the industrial production/service needs to ensure the friendly interaction with surrounding environment. The data-centric collaboration uses comprehensive sensors and big data analytics to provide an efficient and cost-effective solution for a complex industrial problem.

3.4 Key Challenges and Open Issues

3.4.1 Key Challenges

There are two aspects of challenges to achieve the Mobile Collaborative Sensing framework: data and functionality.

3.4.1.1 Data

- Data analytics [84, 85]. The Mobile Collaborative Sensing framework analyzes spatio-temporal heterogeneous data. Such data is collected by heterogeneous devices, and collected from different locations and different time points. Data from different sources has different semantics, formats, sizes, and contexts. Data analytics consider how to analyze this heterogeneous data to let heterogeneous devices work together under a unified framework.
- Structuring data: transforming unstructured data into a unified structured format. As the basis of the Mobile Collaborative Sensing framework, heterogeneous spatio-temporal data from different devices is not natively structured, e.g., billions of log messages from large fleets of medical equipment [86], such unstructured data is typically text-heavy, and contains important log information, such as dates, running parameters of equipment, and values of these running parameters.
- Data privacy and knowledge access authorization [87, 88]. They are important for data owners. However, in the Mobile Collaborative Sensing framework, between data owners and data consumers, sharing data and knowledge is needed and important to achieve good collaboration. For example, two different industrial systems, they are data sources, and they belong to different departments. Because of the high correlation of industrial processes, what level is just enough, and how to define the level of privacy and access authorization between these two different industrial systems, are two challenges that are worth studying.
- Generic data model [89]. For making the spatio-temporal data from heterogeneous devices be able to be used in knowledge discovery, in the Mobile Collaborative Sensing framework, a generic data model is used to format and unify the heterogeneous data. Such heterogeneous data from different devices has different formats, contexts, semantics, complexity and privacy requirements. The design of the generic data model is a challenge.

3.4.1.2 Functionality

- Knowledge discovery [90]. In the Mobile Collaborative Sensing framework, data-driven knowledge discovery is an important component. Such discovery is by mining and analyzing massive spatio-temporal heterogeneous data. It is still a challenge in mining and analyzing data correlations, data contexts and data semantics to discover knowledge.
- Effective and high-efficiency knowledge utilization [91]. With widely deploying and using sensors and wireless devices, massive data has been produced by device holders and machines at an unprecedented rate. Exploring the possibility of data-driven innovation has become an important research topic. The major challenge of such exploring in the Mobile Collaborative Sensing framework, is how to release and use the knowledge that is mined from the massive data.
- Supporting particular applications. In a particular application, specific data mining and training is required to perform knowledge discovery. For example, for detecting the leakage of toxic gases, real-time data mining algorithms are needed to mine the massive dynamic data⁵ collected from static and mobile sensor-embedded wireless devices. These devices are used to monitor the dynamic environment of a petrochemical plant. The Mobile Collaborative Sensing framework is required to have the ability to support special requirements and to make data owners and data consumers be able to communicate with each other for effective data mining and knowledge discovery.
- Real-time processing/controlling [92]. For example, in an industrial production environment, because of the dynamic nature of industrial production, real-time processing/controlling is necessary. However, due to the complexity of industrial production, real-time processing/controlling is hard to achieve, e.g., in petrochemical production, there are a series of fast chemical reaction processes; the real-time controlling of these processes is not so easy to achieve.
- Interfaces between internal modules. The interfaces between different internal modules play an important role in affecting the performance of workflow. However, how to design effective interfaces is a challenge in Mobile Collaborative Sensing framework. First, the inside of each internal module needs to be made clear enough, and then, each internal module needs to provide individual parameters to design the corresponding interface. The difficulty of this design is: which parameters of each internal module affect workflow performance, and how they affect it.
- Development of a security model [93]. A security model is capable of providing privacy and authority management. In the Mobile Collaborative Sensing framework, there are numerous roles and various corresponding parameters, e.g., data owners and data consumers. Therefore, how to design an appropriate and moderate security model is a challenge for achieving a safe and resource shared framework.

5. In a toxic gas detection application, dynamic data is the sensing records with time stamps and location tags.

3.4.2 Open Issues

Considering the aforementioned challenges, the open research issues are listed as follows.

- Data integration [94]. Data is the basis of the Mobile Collaborative Sensing framework, and to achieve the collaborative capability among different data sources, data integration is an important research issue. Data integration is used to combine different data sources into a unified framework, and these different data sources are controlled by different owners. In the book [95], AnHai Doan *et al.* have provided and discussed: (i) the typical examples of data integration applications from different domains such as business, science and government; (ii) the goal of data integration and why it is a hard problem; and (iii) a data integration architecture.
- Data mining algorithms [96]. Adequate data mining is an important research issue for the Mobile Collaborative Sensing framework. Such mining can be used in large-scale, complex and dynamic sensing environment. For example, by mining the big monitoring data from a large-scale petrochemical plant, the potential leakage sources of toxic gases can be predicted, and by using such prediction, the safety of large-scale industrial environment can be improved. The study in this topic is still very limited, due to the limitation of technology on big data analytics.
- Collaborative knowledge discovery algorithms [97]. In the Mobile Collaborative Sensing framework, designing algorithms to enable the collaboration between crowd wisdom and industrial sensing intelligence for discovering useful knowledge is a valuable research issue. However, due to the limitation of technology on the big data analytics and data processing in a large-scale, complex and dynamic sensing environment, as well as the problem of data integration, the study in collaborative knowledge discovery is still limited.
- Real-time algorithms [98]. The real-time algorithms on data processing are necessary in an intelligent framework to improve the timeliness of processing dynamic processes. Shen Yin *et al.* [99] have proposed two real-time schemes for the fault-tolerant architecture proposed in [100]. This architecture is designed for the fault-tolerant control of industrial systems. One is a gradient-based iterative tuning scheme for the real-time optimization of system performance. The other is an adaptive residual generator scheme for the real-time identification of the abnormal change of system parameters. Other than this fault-tolerant control, in other aspects, real-time algorithms are very important, as well, for example detecting toxic gases in a highly dynamic production environment. However, there are no achievements in these “other aspects”.
- Trusted and privacy-protected model design [101]. The privacy of data and knowledge is important for data owners and data consumers in a collaborative framework. In the Mobile Collaborative Sensing framework, it is indispensable to study and design a trusted and privacy-protected (i) data model for data processing and analysis and (ii) knowledge model for knowledge discovery and utilization. Such models are an important part of the collaborative framework. However, the model design needs to consider the different requirements from data owners and data consumers

for different applications. There is no a unified standard for such design.

3.5 Conclusion

Facing the growing demands of industrial production/service on improving the safety, efficiency and eco-friendliness, as well as meeting the cost-effective objectives, based on the Industrial Internet of Things (IIoT) and the characteristics of industrial problems, this work proposes a mobile collaborative sensing framework. This sensing- and collaboration-based intelligence framework has the potential to improve the performance of industrial systems by providing better awareness and control to dynamic industrial environments and correlated production/service processes, with analysing and integrating massive spatio-temporal data. Moreover, because the spatio-temporal data is collected from things and humans, this mobile collaborative sensing framework can achieve improved automated decision making collaborating with the crowd wisdom of humans. In addition, the challenges and open issues for developing the framework have been explored and discussed. The aim is to identify innovative research issues for industrial intelligence and to deploy the framework to practical industrial applications.

MCS-Locating: Locating in Mobile Collaborative Sensing based Data Space

Contents

4.1	Introduction	52
4.2	Related Work	53
4.2.1	Wireless Indoor Target Locating	53
4.3	System Model	54
4.4	Locating in Mobile Collaborative Sensing based Data Space	55
4.5	Evaluation	58
4.5.1	Experimental Setup	58
4.5.2	Performance Evaluation	59
4.6	Conclusion	61

Locating a target in an indoor social environment by a Mobile Network is important and difficult for location-based applications such as targeted advertisements, geosocial networking and emergency services [102]. A number of radio based solutions have been proposed [103–106]. However, these solutions, more or less, require special infrastructure or extensive pre-training of a site survey. Since people habitually carry their mobile devices in their daily lives, it enables to locate a target by collaborating these mobile devices. This work proposes a locating algorithm, LiCS, which is based on sensing and analyzing the wireless information from the space where the mobile devices are inside. The prototype system of LiCS is developed. Experimental results show that LiCS achieves comparable locating accuracy with previous approaches, even without any special hardware.

4.1 Introduction

Mobile indoor locating is an important research topic in many applications [107,108], and it has been widely and successfully used in a great number of domains. However, it is still a research challenge. A number of existing locating approaches require special infrastructure (e.g., indoor beacons), or extensive pre-training before performing target locating, for example, locating a target in an indoor environment by WiFi signal fingerprinting: signal fingerprint survey in this environment is necessary, and pre-training these signal fingerprints is important to the accuracy of locating.

With the dramatic increase in the number of and the potential functionality of mobile devices, massive sensing data has been produced in our day-to-day social lives. By such sensing data, it is possible to achieve “special infrastructure free”¹ and even “pre-training free” indoor locating.

In this work, LiCS is proposed. LiCS uses Received Signal Strength (RSS) that exists in any wireless devices as location fingerprints to differentiate different locations, so LiCS can be directly supported by off-the-shelf wireless infrastructure. This algorithm uses trace data from individuals’ mobile devices, and a location estimation model. The trace data includes the following information: MAC addresses of devices, MAC addresses of signal transmitters, and corresponding RSSIs (Received Signal Strength Indication)². Mobile devices periodically report their trace data to a Data Analysis Center (DAC). The DAC runs a machine learning algorithm that accepts the wireless trace data as features of user mobility patterns, and periodically estimates the locations of mobile devices in real time. Since wireless information obtained from the social environment around us is used, LiCS can achieve fine-grained target locating.

LiCS exploits the advantage of model-based technique (versatility and conciseness), and avoids its drawback (accuracy loss for target locating) by training a location estimation model using real-time trace data of individuals. Figure 4.1 provides an example for the motivation of applying LiCS, and the system architecture of LiCS.

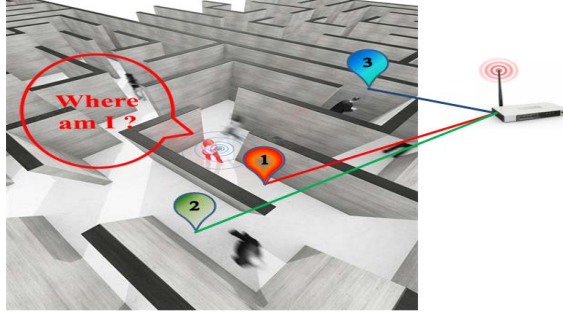
To validate the effectiveness of this algorithm, a prototype system is developed to conduct long-term experiments in two research laboratories and a corridor of a middle-size academic building covering over $39,725m^2$. Experimental results show that LiCS achieves comparable locating accuracy with previous approaches even without a site survey. LiCS provides a room-level target locating service without using special devices and extensive pre-training of a site survey.

The scientific contributions of this work are summarized as follows:

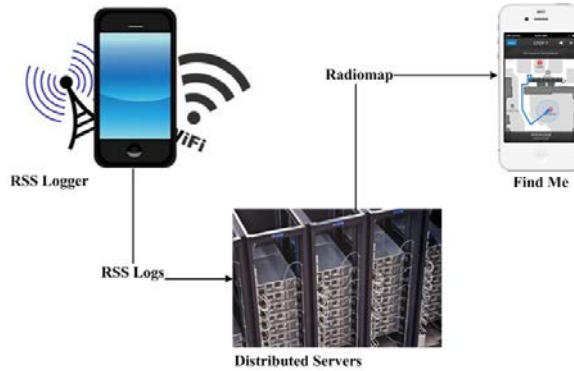
1. A locating algorithm LiCS is proposed. It is an indoor target locating algorithm for social environments. LiCS utilizes automatic self training to target trace data without any specific configuration on mobile devices. The trace data includes general information of wireless devices; there is no any special information from special devices.

1. Special infrastructure denotes that the infrastructure consists of customized devices.

2. RSSI is used to indicate the value of RSS.



(a) Motivation example. The location estimation ability of signal based indoor target locating heavily depends on the signal features and strength in different locations. For LiCS, any individual can use the trained model to estimate his/her location, and then his/her RSS is used to train the model again. By such training, the model is trained to make its parameters be accordant with the practical signal features.



(b) System architecture. As an improvement, LiCS exploits the advantage of the model-based technique (versatility and conciseness), and avoids its drawback, with training a location estimation model: the RSS data from users is used to train the model, and the model is installed in distributed servers. These servers collaborate with each other to achieve wide indoor target locating.

Figure 4.1: Motivation example and system architecture.

2. A prototype system of LiCS is developed on Android devices, and an extensive set of experiments is performed on this system.

4.2 Related Work

In this work, the related work is mobile device based indoor target locating: by mobile devices and corresponding wireless sensing data, the current achievements on indoor target locating.

4.2.1 Wireless Indoor Target Locating

In mobile computing, a user carries a mobile device (e.g., smart phone) to take a random movement or to move within an area deployed static sensor nodes [109]. In either case, the location information

of the user is useful in many mobile based applications, e.g., Location-Based Services (LBS) [110]. Outdoor target locating is well solved by GPS, but indoor target locating is still a challenge in many cases. For the indoor target locating issue, a number of algorithms have been proposed in the past two decades. These algorithms fall into two categories:

1. Sensor nodes or beacon nodes are installed to locate a target [111–113]. In this kind of locating algorithms, a lot of special hardware is required.
2. Wireless signals are widely used to distinguish different locations. It implies that in different locations the signal feature is different. By knowing the unique feature of each location, locating a target is only to detect these features. Two techniques are widely used in such signal feature based algorithms: fingerprinting, and geometrical and statistical model based modelling.
 - (a) A large number of indoor target locating algorithms use fingerprint matching as the basic scheme of location determination. The main idea of this technique is to collect the fingerprints of different locations in an interested area, and then builds a fingerprint database. The location of a user is then estimated by matching the new measured fingerprint from the user with the records of the fingerprint database. Many kinds of wireless features have been used as the fingerprint such as WiFi signals [114] (e.g., LiFS algorithm [103]), Radio Frequency (RF) signals [115, 116], and Frequency Modulation (FM) radio signals [117]. However, in addition to the inflexibility on dynamic environment, fingerprinting based algorithms are suffering in considerable manual costs and efforts.
 - (b) Using geometrical and statistical models is helpful in reducing manual costs and efforts comparing with searching for best-fit fingerprints from pre-built fingerprint databases [118]. For instance, the prevalent Log-Distance Path Loss (LDPL) model [119] provides a semi-statistical function between RSS values and RF propagation distances. However, these model based algorithms trade the measurement efforts with good locating accuracy.

4.3 System Model

The prototype system is developed on Android smart phones, and follows mobile-based network-assisted architecture (Fig. 4.1b). In this prototype system, there are N fixed signal transmitters $T = \{t_1, t_2, \dots, t_i, \dots, t_N\}$, and M mobilizable signal receivers $R = \{r_1, r_2, \dots, r_i, \dots, r_M\}$. The parameter $p(t)$ denotes the estimation of a mobile terminal location at time t . The parameter $p'(t)$ is used to denote the real location of a mobile terminal at time t . Moreover, the parameter $d(t)$ is the measured distance between the estimated location $p(t)$ and the real location $p'(t)$. The “step” is used as the unit to measure the distance between $p(t)$ and $p'(t)$. Such distance is set as the locating error of a mobile terminal at time t in evaluation experiments (Section 5.6). The problem of mobile sensing based location estimation can be defined as an identification procedure. The matched fingerprints can be

identified from the model-assisted fingerprint database. The model is trained in real time by learning the wireless information around the mobile terminals.

4.4 Locating in Mobile Collaborative Sensing based Data Space

As an effective measurement, RSS is easily available from various wireless signals, e.g., WiFi or Bluetooth signals. A large number of RSS based indoor target locating algorithms have been proposed. However, setting an RSS fingerprint database to support indoor target locating, it is time-consuming and labor-intensive. Especially, extensive experiments have shown that the RSS is vulnerable to environment (an example is shown in Fig. 4.2a). How to avoid these weaknesses to improve the performance of RSS based indoor target locating? It is worth noting that the trend of RSS change is obvious in different locations (Figure 4.2b).

To try to avoid RSS weaknesses, LiCS is proposed, and the details are shown as follows.

1. **Input:** Signal triples³ from individuals.
2. First, at time $t + 1$, the value $p(t + 1)$ can be estimated by the “observed locations” $p(t), p(t - 1), p(t - 2), \dots, p(t - k + 1)$. Moreover, the relationship between the output $p(t + 1)$ and the input $p(t), p(t - 1), p(t - 2), \dots, p(t - k + 1)$ has the following mathematical representation (Eq.(4.1)):

$$p(t + 1) = \alpha_0 + \sum_{j=1}^q \alpha_j g(\beta_{0j} + \sum_{i=1}^k \beta_{ij} p(t - i + 1)) + \epsilon, \quad (4.1)$$

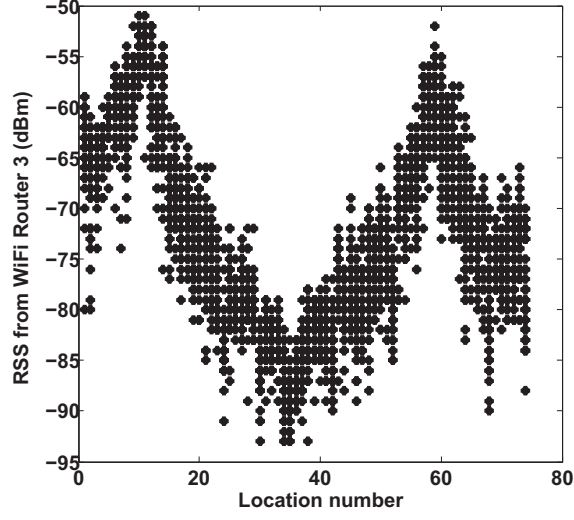
where $\alpha_j (j = 0, 1, 2, \dots, q)$ and $\beta_{ij} (i = 0, 1, 2, \dots, k; j = 1, 2, 3, \dots, q)$ are the connection weights between time series, k is the number of “observed locations”, q is the number of nodes of the hidden layer⁴, and ϵ is noise of the estimation. The logistic function $g(x) = \frac{1}{1+e^{-x}}$ is used as a hidden-layer transfer function. In this work, an optimal location estimation model is built by training the model with the wireless data collected from the real physical space around us (the collected data can be denoted as signal triples). The training steps are shown as follows:

Step 1: Cluster the signal triples. Partition all triples into several clusters using an Expectation-Maximization (EM) clustering algorithm. Moreover a cluster center can be obtained for each cluster. Each cluster is given a unique number as its location.

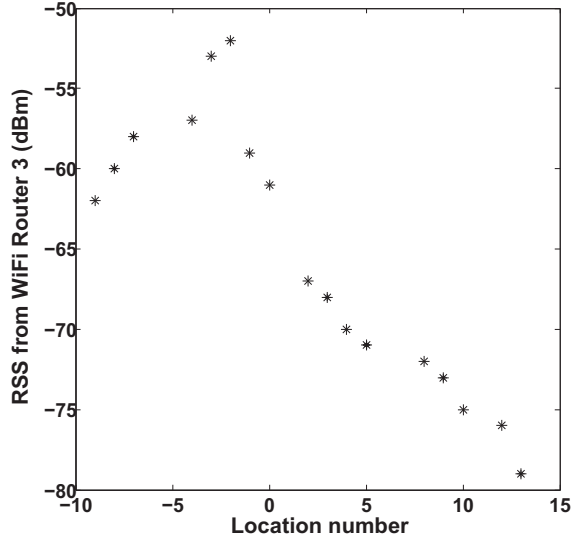
Step 2: Input some selected time-serial signal triples with corresponding locations of clusters into Eq.(4.1) for learning the optimal configuration of parameters, α_j, β_{ij} and ϵ . A location estimation model with optimal parameter configuration can be obtained. Moreover, the growth

3. This content is included in a signal triple, and the triple can be denoted as: $[RSS, MAC_T, MAC_R]$. For the specific RSS of a location, MAC_T is the MAC address of corresponding signal transmitter, and MAC_R is the MAC address of corresponding signal receiver.

4. In machine learning, using the hidden layer enables greater processing power and system flexibility. The nodes in the hidden layer are named as hidden nodes. Hidden nodes are the nodes that are neither in the input layer nor the output layer of a learning network. These nodes are essentially hidden from view, and their numbers and organization can typically be treated as a black box to people who are interfacing with the network.



(a) An example: the fluctuation range of RSS in different locations. For example, in the location 60, the RSS is not an exact value, and it is a value range. So the fingerprint of a location cannot be denoted by an exact RSS value.



(b) Changing trend of RSS. Location number= 0 is the location of a signal transmitter, and with a signal receiver moves away from the transmitter, the RSS is gradually weakened.

Figure 4.2: Instability of RSS, and changing trend in different locations.

of logistic function $g(\cdot)$ satisfies: the initial stage of growth is approximately exponential; then, as saturation begins, the growth slows, and at maturity, the growth stops. So if the training time is long enough, the parameter configuration of Eq.(4.1) will gradually converge to the optimal.

3. Then, a target can be located with the optimal location estimation model. (i) Give the target a start location $p(0)$. Calculate the Euclidean distances between a received new signal triple and all cluster centers (the new signal triple is from the target). If the shortest Euclidean distance is relative to the cluster k , $p(0) = k$. (ii) Calculate the location at time $t + 1$. Using trained Eq.(4.1), from the start location $p(0)$, time-serial locations can be obtained. Based on “observed locations”, $[p(t), p(t - 1), \dots, p(0)]$, $p(t + 1)$ can be calculated.
4. **Output:** The real-time location (cluster number) of a target. For some special applications, if the physical coordinates of clusters are available, the physical location of the target is obtained.

In addition to obtaining the real-time location of a target, in this algorithm, the optimal location estimation model is periodically trained by new signal triples. If more signal transmitters can be detected by receivers, it will help to distinguish different locations more effectively to achieve higher accuracy on target locating.

Note that Eq.(4.1) is the core of LiCS. From above descriptions of the algorithm LiCS, these parameters affect the accuracy of LiCS: the connection weights between time series, α_j and β_{ij} . Parameter α_j reflects the importance of hidden-layer transfer for the location estimation of time $t + 1$. In other words, it denotes the degree of the correlation between different time series. Parameter β_{ij} reflects the importance of the j^{th} node in the hidden layer, when the hidden layer transfers the influence of observed location $p(t - i + 1)$ to $p(t + 1)$. Appropriate values of these parameters at time $t + 1$ are helpful to improve the accuracy of target locating.

The variables that are used in LiCS are summarized in Tab. 4.1.

Table 4.1: Variables and Explanations

Variable	Explanation
$[RSS, MAC_T, MAC_R]$	A triple consists of three elements, where RSS is Received Signal Strength received by a mobile terminal, MAC_T is the MAC address of corresponding signal transmitter, and MAC_R is the MAC address of corresponding signal receiver (the mobile terminal). The triple is used to train the location estimation model, and construct a fingerprint database.
$p(t)$	It denotes the location of a mobile terminal at time t .
α_j	It is the connection weight between time series, and it reflects the importance of hidden-layer transfer in the location estimation of time $t + 1$.
β_{ij}	It is the connection weight between time series, and it reflects the importance of j^{th} node in the hidden layer, when the hidden layer transfers the influence of observed location $p(t - i + 1)$ to $p(t + 1)$.

4.5 Evaluation

The prototype system of LiCS is developed on the increasingly popular Android OS which supports WiFi and Bluetooth. Long-term experiments are conducted in two laboratories ($84m^2$ and $53m^2$, respectively) and a corridor (Fig. 4.3) of a middle-size academic building where a number of WiFi routers without location information have been installed. Moreover, in each experimental site, three Bluetooth transmitters are installed (laptop-embedded Bluetooth transmitters are used in the experiment, so they are not special devices. The signal of Bluetooth is full-coverage for each experimental site). The experiment lasted one month using 9 volunteers. WiFi signals and Bluetooth signals are measured.

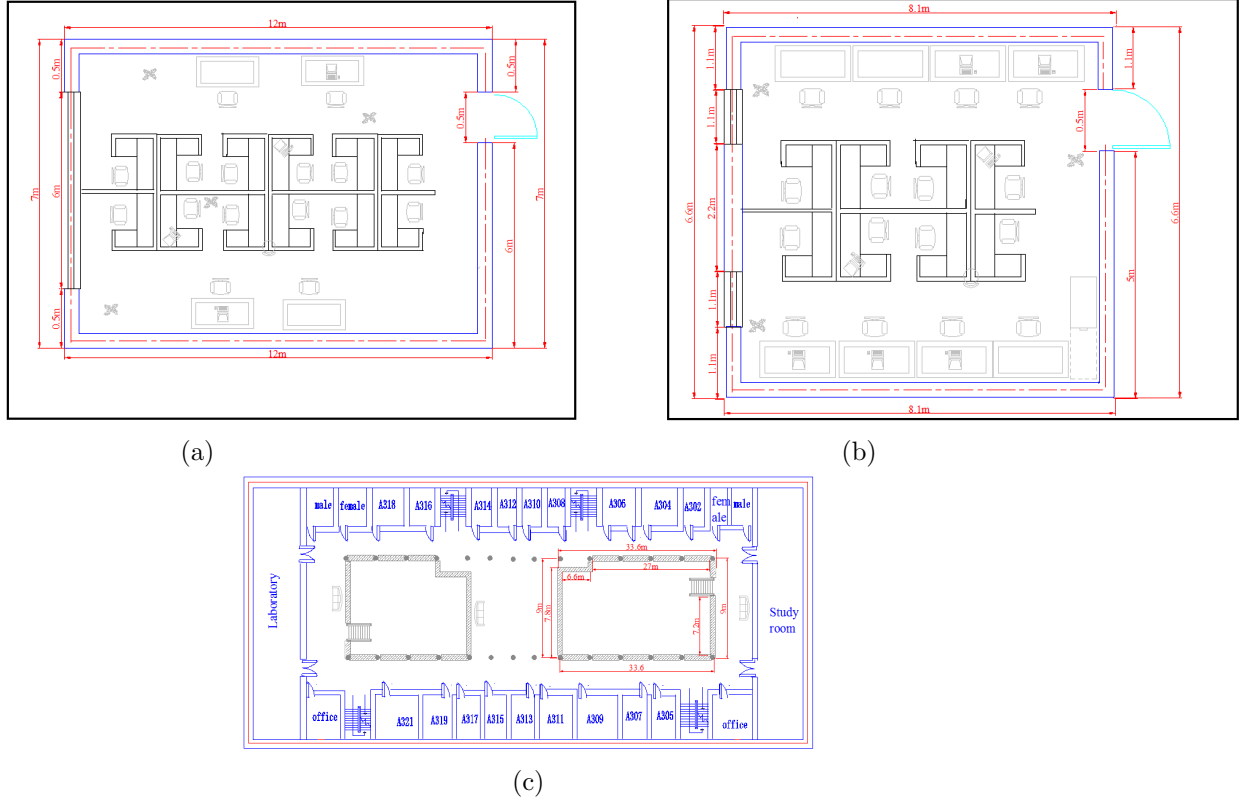


Figure 4.3: Floor plans of experimental sites. (a) Laboratory covering over $84m^2$. (b) Laboratory covering over $53m^2$. (c) Corridor covering over $302m^2$.

4.5.1 Experimental Setup

Each volunteer carries a mobile phone, and can take any activity in any area of experimental sites. The trace data of each volunteer is recorded every 30 seconds during working hours (from 9:00 a.m. to 10:00 p.m.). Moreover, the trace data from volunteers covers most of the areas of experimental

sites. WiFi and Bluetooth signals are used in experiment. Bluetooth is a wireless technology standard for exchanging data over short distances, so Bluetooth signals attenuate more rapidly with distance compared with WiFi signals.

LiFS [103] is used to conduct comparison under the same experimental conditions. LiFS is an RSS based indoor target locating algorithm (using WiFi signals). The key idea behind LiFS is that human motion can be applied to connect previously independent radio fingerprints under certain semantics. In LiFS, exact values of RSS are used to establish a fingerprint database. When a user sends a location query with his/her current RSS fingerprint, LiFS retrieves the fingerprint database, and returns the matched fingerprints as well as the corresponding locations.

4.5.2 Performance Evaluation

Comparative results of LiFS and LiCS are shown and described in this section. This comparative evaluation estimates 248 location queries, and cumulates all of the locating errors of these queries by Cumulative Distribution Function (CDF)⁵ for both algorithms, respectively. The results are illustrated in Fig. 4.4 and Fig. 4.5.

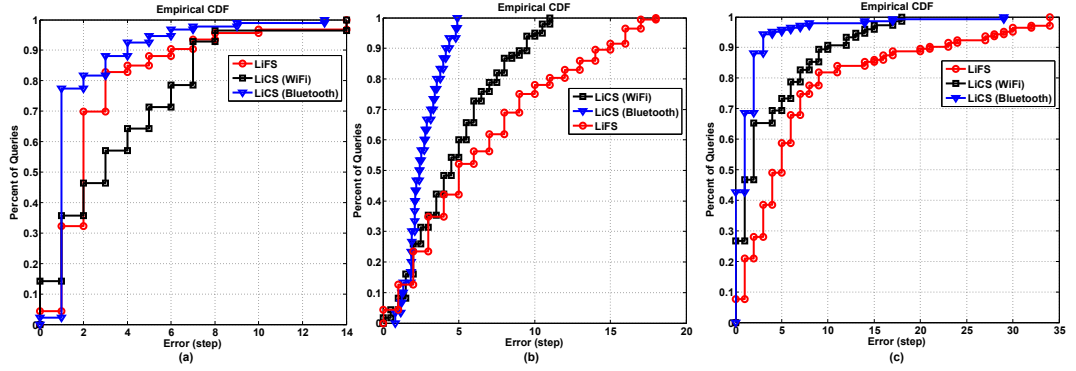


Figure 4.4: CDFs of target locating errors for both algorithms in three different experimental sites. (a) CDF of target locating errors in the laboratory covering over $84m^2$. (b) CDF of target locating errors in the laboratory covering over $53m^2$. (c) CDF of target locating errors in the corridor covering over $302m^2$.

The unit of the estimated error for target locating is $step \approx 1.2m$. The average locating error of LiCS with Bluetooth signals is $2.6575m$, LiCS with WiFi signals is $4.35m$, and LiFS is $5.95m$. The maximum locating error of LiCS with Bluetooth signals is $33.6m$, LiCS with WiFi signals is $34.8m$, and LiFS is $40.8m$.

For comparing the overall performance of three experimental sites for both algorithms, the average results of three experimental sites are calculated for LiCS (Bluetooth), LiCS (WiFi), and LiFS. From

5. Cumulative Distribution Function describes the probability that a real-valued random variable X with a given probability distribution is found at a value less than or equal to x . It can be formulated as $F_X(x) = P(X \leq x)$.

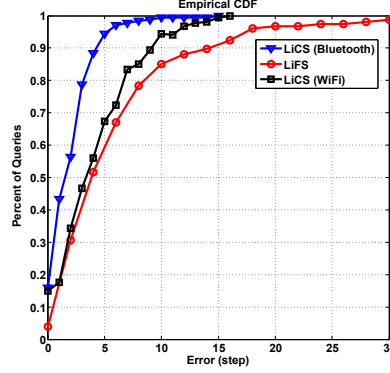


Figure 4.5: CDF of target locating errors for the average of three experimental sites.

Fig. 4.5, it is observed that LiCS is better than LiFS for the average of three experimental sites. For example, the locating errors of 95% queries are less than $6m$ for LiCS (Bluetooth), 69% queries are less than $6m$ for LiCS (WiFi), and 60% for LiFS. LiCS uses the model training with real-time data, so the locating accuracy for location queries is improved compared with LiFS. Moreover, why the locating accuracy of LiCS (Bluetooth) is higher than LiCS (WiFi)? Considering the attenuation characteristics of Bluetooth signals and WiFi signals, the difference of RSS in different locations for Bluetooth is greater than the difference for WiFi. So using the RSS of Bluetooth to distinguish different locations is more accurate than using the RSS of WiFi. If different locations can be distinguished more clearly, obtaining higher locating accuracy is possible by using an RSS based fingerprint database.

Furthermore, from the experimental results of Fig. 4.4, these conclusions are obtained: (i) on average, for locating accuracy, LiCS is better than LiFS in the three different experimental sites. For example, the locating errors of 80% queries are less than $2.4m$ for LiCS (Bluetooth), and 70% for LiFS, in the laboratory covering over $84m^2$; (ii) the locating accuracy of LiCS (Bluetooth) is better than LiCS (WiFi) and LiFS. For example, the locating errors of 50% queries are under $2.4m$ for LiCS (Bluetooth), while about 30% for LiCS (WiFi) and about 25% for LiFS, in the laboratory covering over $53m^2$. Bluetooth improves the average locating error up to 39% compared with LiCS (WiFi), and up to 55% compared with LiFS. Because the signal strength of Bluetooth is changed sharply in different locations, which makes the distinction of signal strength in different locations more remarkable (the “remarkable” is conducive to improving the accuracy of locating).

Moreover, LiFS is based on a priori database (some human intervention is necessary in the build phase of a database). LiCS uses wireless sensing and model training, so only wireless information is required, which is received by users in their daily lives. The locating process of LiCS is automatic and priori-database-free. LiCS is based on WiFi and Bluetooth information which is readily available. The above-mentioned features of LiCS make the rapid deployment of a target locating system possible.

4.6 Conclusion

Mobile Collaborative Sensing is a distributed problem-solving scheme that has emerged in recent years. It exploits the potential and wisdom of collaboration to support various applications and to improve the performance of various algorithms in a cost-effective fashion. In this work, with sensed and collected WiFi and Bluetooth information from surroundings, a time-serial location estimation algorithm LiCS is proposed, and a prototype system is developed to verify the validity of LiCS. It is an indoor target localization algorithm based on: (i) mobile devices carried by individuals, and (ii) a location estimation model which is trained by the collected data from individuals about WiFi and Bluetooth information. Experimental results show that LiCS achieves competitive location accuracy without any special infrastructure. This work sets up a novel perspective to Mobile Collaborative Sensing based indoor localization algorithms.

MCS-Prediction: Predicting Disease Dynamics by Mobile Collaborative Sensing

Contents

5.1	Introduction	64
5.2	Related Work	65
5.3	Network Model	66
5.4	Recognition Model	68
5.5	Prediction Algorithm	70
5.6	Evaluation	72
5.6.1	Experimental Setup	72
5.6.2	Experimental Results and Discussion	73
5.7	Conclusion	76

Predicting disease dynamics during the spread of infectious diseases is an important aspect of e-Health applications. In such prediction, Realistic Contact Networks (RCNs) have been widely used to characterize disease dynamics. The structure of such networks is dynamically changed during the spread. Capturing such kind of dynamic structure is the basis of prediction. With the popularity of mobile devices, it is possible to capture the dynamic change of network structure by the collaboration of mobile devices. On this basis, this work designs a model to recognize the dynamic structure of RCNs. On the basis of this model, a prediction algorithm is proposed for disease dynamics. By comparison experiments, it is observed that the algorithm improves the prediction accuracy compared with the Spatial Risk Model (SRM) based prediction.

The Spatial Risk Model (SRM) [120] is used in this work to evaluate the impact of network structure

on disease dynamics. It is a statistical model used in communicable diseases to estimate or predict the presence or incidence of infected cases within a particular geographical area. As the important parameters of SRM, β and γ , (i) β is a non-negative scalar, the infection probability of an individual. This individual is susceptible and has a single infected neighbor. The infection probability of a susceptible individual with n infected neighbors is $n*\beta$, and (ii) γ is a positive scalar, the recovery probability of an infected individual, and if this individual recovers from a disease, this individual will be disconnected with other individuals, in a contact network. Based on the above explanations for β and γ , this can be observed: these two parameters β and γ are all degree-related, so β and γ can be used to reflect the structure knowledge of a contact network in the SRM model.

5.1 Introduction

As an important aspect of e-Health [121–123], quantifying and even predicting disease dynamics during the spread of infectious diseases [124–127] is very important to effectively allocate resources and to quickly make a response in a public health event. For the public health, underestimating the impact of a disease may lead to an inadequate response, while overestimating it, can lead to the misallocation on the limited resources.

The reproductive number R^1 is used to quantify the disease dynamics during the disease spread, and a wide range of methods have been proposed to estimate or predict R by mining surveillance data [128–132]. However, these existing methods, almost all of them, are based on the networks which are assumed to have special network structure, for example, the networks with exponential degree distributions. During disease spread, the network structure of the relevant Realistic Contact Network (RCN) is dynamically changed along with the spread of a disease. The definition of the RCN is described in Definition 6.

Definition 6 *In the real physical world, a Realistic Contact Network consists of a group of people who can get in touch with each other. In this network, nodes represent people, and edges represent the direct contact between two nodes. If there is an edge between two nodes, it means that there is physical contact between two individuals corresponding to the two nodes.*

During disease spread, capturing the dynamic change of the relevant RCN is helpful to improve the prediction accuracy of disease dynamics. With the popularity of mobile devices in public health [133], it is possible to acquire the dynamic change of network structure. An example is illustrated in Fig. 5.1.

In this work, a recognition model is designed to dynamically acquire the structure knowledge of the relevant RCN during disease spread. On the basis of this model, a prediction algorithm is proposed to predict the parameter R .

1. The number of cases generates in an infectious period, in an uninfected population.

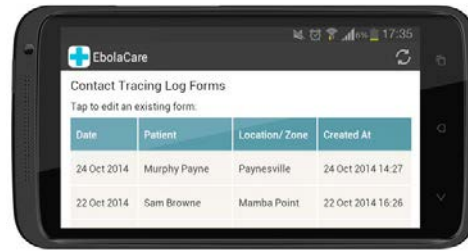


Figure 5.1: A mobile device installed Contact Tracing Application. This application is used to track the Ebola outbreak of West Africa [134]. It can track everyone who directly contacts with a sick Ebola patient. Such devices are carried by the volunteers of Ebola outbreak areas. The data collected by this application is shared with the WHO (World Health Organization), who is using information from hundreds of aid organizations to make big strategic decisions.

The scientific contributions of this work are shown as follows:

- A recognition model is designed to acquire the structure knowledge of an RCN. Three structure properties are used to design the model.
- A prediction algorithm is proposed. This algorithm uses the acquired structure knowledge by the recognition model.

The proposed prediction algorithm uses the recognized structure knowledge by the recognition model as an important aspect to improve the prediction accuracy. Such a network-structure-based prediction algorithm improves the prediction accuracy, even when the network structure is dynamically changed.

5.2 Related Work

This section provides a brief overview on disease dynamics, from the perspective of widely used models and prediction methods.

Epidemic models describe the spread of a communicable disease in a population. In these models, the individuals of a population are taken and placed into one of these three states: Susceptible, Infectious or Recovered (SIR). Modelling the transitions among these states generates the SIR model. This simple SIR model has been extended in a multitude of ways, e.g., by adding/deleting states, or by considering a special pattern of a transition. For example: (i) SIS model [135]. For a disease with no immunity, infected susceptible individuals return to the susceptible state after recovering. (ii) Non-equilibrium transitions [136]. Replacing the homogeneous mixing hypothesis that any individual can contact with any other, non-equilibrium transitions assume that each individual has a certain number of contacts, which is reflected as the node's degree k in a contact network. The degrees of nodes in a network can be denoted as the degree distribution of the network. Degree distribution is the important structural property of a network. Different degree distributions bring different impacts on the transitions among different states of disease dynamics. The relevant studies on such impacts enable disease dynamics to be

associated with a network, and to explore the impact of spatial contact patterns [137–141]. For example, in a dynamic contact network with an arbitrary degree distribution, the real transition threshold of three states during an epidemic is: $\lambda_{th} = \frac{k}{k^2}$, where k is the average degree of the network, and k^2 is the average degree at the next moment of the dynamic network.

As another important aspect of the overview for disease dynamics, previous studies on prediction methods are briefly classified and summarized in Table 5.1.

Through the brief overview, it is feasible to improve the prediction accuracy by acquiring the realistic structure knowledge of a dynamic contact network, during disease spread. Three structure properties are used to represent the structure knowledge in this work: clustering coefficient, degree distribution and degree correlation.

5.3 Network Model

The RCN studied in this work is constructed by processing surveillance data from mobile devices which are carried by the volunteers of Ebola outbreak areas.

As a recent outbreak of Ebola, from 27th March, 2014, Ebola Virus Disease (it is commonly known as “Ebola”) has killed 11,323 individuals, and the total number of cases has reached 28,646 [151]. Researchers generally believe that from a 2-year-old boy of Guinea to his mother, sister and grandmother (this is a contact network) Ebola rapidly spreads in West Africa since March 2014. A series of time-aware Ebola cases is collected by the WHO as well as the Ministries of Health of epidemic countries.

In this work, three groups of data from three typical outbreak countries are selected, Guinea, Nigeria and Liberia. Guinea is the source of this outbreak and has relatively high quantity of confirmed cases, Nigeria is far away from the source of the outbreak and has relatively low quantity of confirmed cases, and Liberia is close to the source of the outbreak and has high quantity of confirmed cases.

By using the above outbreak data, a contact network is constructed. There are 941 nodes corresponding to different cases in this network (Fig. 5.2 illustrates a slice of the network). With the spread of a disease, the contact network is gradually constructed, and by the order of time stamps of cases, the spread process of the disease is very clear. Such a contact network can be modelled as a dynamic graph G_t . In this network, there are four parameters: (i) Case ID. A unique number to indicate a case; (ii) Source ID. It indicates the ID of the infection source for a case; (iii) Date. The reported date of a case; (iv) Location. The coordinates (longitude and latitude) of a case.

The dynamic graph G_t is described in detail as follows. An undirected weighted graph $G_t = (V_t, E_t, W_t)$, where V_t is a set of n_t vertices to indicate cases, E_t is the set of edges, and W_t is the set of weights. If there is an edge between vertex i and vertex j , e_{ij} , it indicates that there is contact between the corresponding individuals of i and j . The weight w_{ij} is the transmission probability (p_{ij}) of a disease from vertex i to vertex j (in the corresponding edge e_{ij}).

The graph G_t is dynamically changed, so it is with a sequence of online updates: (i) Delete(e_{ij}). It

Table 5.1: A Qualitative Overview of the Prediction Methods on Disease Dynamics

Prediction Method	Addressed Problem	Main Idea	Open Issue
Prediction based on online social networks [142, 143]	Predicting the health of real-world populations in real time, to understand actual threats and ongoing disease outbreaks.	This kind of method uses the geo-tagged status updates of traveling Twitter users to infer the properties of individual flow between different cities.	(i) Use of Twitter data. First, Tweets are typed by users who are not experts: the correctness of their judgment for their suffered diseases cannot be guaranteed. Moreover, for example, influenza can be described by different words (even such description does not contain the keywords “flu” or “influenza”). Second, the observed data comes from a small fraction of travellers, and it is only partial data of these travellers. (ii) Based on our experience in Twitter data acquisition by the Twitter Search API, Tweets sent from mobile devices do not always have accurate GPS coordinates. Even if two Tweets are located in the same location (with a certain locating error), it cannot assure that the relevant physical individuals contact with each other and are within the communicable distance of a disease.
Prediction based on RCNs [131, 144–146]	This kind of study has proposed that the structure of contact networks impacts disease dynamics. Such impacts are different for different network structures. For the RCNs, there are three typical network structures that are widely used in studies. They are exponential, power-law [147] and random [148, 149].	Constructing RCNs is by analyzing the actual surveillance data from disease outbreaks.	Exponential and power-law degree distributions have been found in many real-world complex networks [145, 150]. However, the structure of a contact network is dynamically changed during the spread of a disease, and structure knowledge can be used to characterize the dynamics of the spread process on this contact network. Further work is necessary to improve the prediction accuracy of disease dynamics by mining the realistic structure knowledge of the dynamic contact network in real time.

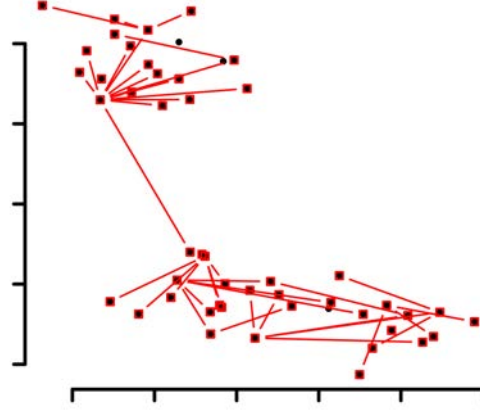


Figure 5.2: A slice of a dynamic contact network. This slice displays 50 cases and their relationships (contact) from three typical countries of the Ebola outbreak since 2014. The three considered countries are: Guinea, Nigeria and Liberia. The black dots are cases (suspected and confirmed), and if there is an edge between two dots, it means that there is contact between two corresponding individuals of two cases. Only a slice of a contact network is displayed in this example, so there are some isolated nodes (black dots without connections).

deletes the edge e_{ij} from E_t , and the corresponding vertices i and j from V_t ; (ii) $\text{Insert}(e_{ij})$. It inserts the edge e_{ij} into E_t , and the corresponding vertices i and j into V_t ; (iii) $\text{Update}(w_{ij})$. It updates the weight w_{ij} that is related to the edge e_{ij} . On the basis of above (i), (ii) and (iii), the graph G_t is updated from $G_t = (V_t, E_t, W_t)$ to $G_{t+1} = (V_{t+1}, E_{t+1}, W_{t+1})$. It means that at different time points, with the spread of a disease, the active subnetworks are different. Such a contact network is time-varying.

5.4 Recognition Model

The structure of a dynamic network is time-varying. It means that in a dynamic network the network structure is different at different time points. RCN is a kind of typical dynamic network. In such a network, the dynamic recognition of network structure is necessary to study how the network structure impacts disease dynamics, and even predict the disease dynamics. In this section, a recognition model is designed to recognize the dynamic structure of an RCN.

There are two components in this recognition model: (i) measure. This component calculates the values of three structure properties, and (ii) knowledge acquisition. The values measured from the first component need to be combined together to reflect and quantify the structure knowledge of a network, and then the structure knowledge can be used into the prediction of disease dynamics (measuring and predicting the value of R).

This model can be formulated and described as follows:

- Measure. As the important structure properties to reflect the structure of a network, in this work,

three structure properties are used, and they are: clustering coefficient (C_i^w), degree distribution ($E[X]$), and degree correlation (r). The formulae to measure the values of these structure properties are shown in Eq.(5.1), Eq.(5.2) and Eq.(5.3), respectively. Using these three formulae, three values are calculated, and these values reflect and quantify the structure knowledge of a network, and by this quantification, the structure knowledge can be used into the design of the prediction algorithm for disease dynamics.

$$C_i^w = \frac{1}{s_i(k_i - 1)} \sum_{j,h} \frac{w_{ij} + w_{ih}}{2} a_{ij} a_{ih} a_{jh}, \quad (5.1)$$

where k_i is the degree of node i , s_i is the strength (summing up the edge weights of the adjacent edges of node i), a_{ij} , a_{ih} and a_{jh} are elements of the adjacent matrix of the network, and w_{ij} and w_{ih} are the weights of corresponding edges.

$$E[X] = x_1 p_1 + x_2 p_2 + \dots + x_k p_k, \quad (5.2)$$

where $X = \{x_1, x_2, \dots, x_i, \dots, x_k\}$ is the set of degrees, and $P = \{p_1, p_2, \dots, p_i, \dots, p_{k'}\}$ ($p_1 + p_2 + \dots + p_{k'} = 1$ and $k' \leq k$) represents the probabilities of the degree values $\{x_1, x_2, \dots, x_i, \dots, x_k\}$, for example, if the set of degrees is $X = \{1, 1, 1, 2\}$, then the set of probabilities for this X is $P = \{\frac{3}{4}, \frac{1}{4}\}$, and $E[X] = 1 * \frac{1}{4} + 1 * \frac{1}{4} + 1 * \frac{1}{4} + 2 * \frac{1}{4} = \frac{5}{4}$.

$$r = \frac{1}{\sigma_q^2} \sum_{j,k} jk (\mathcal{E}_{jk} - q_j q_k), \quad (5.3)$$

where (i) $\sigma_q^2 = \sum_k k^2 q_k - [\sum_k k q_k]^2$, (ii) $\mathcal{E}_{jk}$ is the joint probability distribution of the degrees of the two nodes (j and k) at either end of a randomly chosen edge e_{jk} . This quantity is symmetric in an undirected network, $\mathcal{E}_{jk} = \mathcal{E}_{kj}$, and obeys these rules: $\sum_{j,k} \mathcal{E}_{jk} = 1$ and $\sum_j \mathcal{E}_{jk} = q_k$, (iii) $q_k = \frac{p_{k+1}}{\sum_{j \geq 1} p_j}$, where p_{k+1} is the degree of node $k+1$, and (iv) the value of “ k ” in Eq.(5.3) is set as the degree of the node k .

- Knowledge acquisition. Eq.(5.4) is used to combine the values of three structure properties.

$$M(C_i^w, E[X], r) = E[k_i] + E[X] + E[\mathcal{E}_{jk}], \quad (5.4)$$

where these three parts, $E[k_i]$, $E[X]$ and $E[\mathcal{E}_{jk}]$ have been introduced in Eq.(5.5), Eq.(5.2) and Eq.(5.6), and Eq.(5.5) and Eq.(5.6) are deduced from Eq.(5.1) and Eq.(5.3) by converting and unifying the different measurement parameters (C_i^w and r) to degree-relevant parameters (k_i and $\mathcal{E}_{jk}$).

$$E[k_i] = E\left[\frac{\sum_{j,h} a_{ij} a_{ih} a_{jh}}{C_i^w s_i} + 1\right]. \quad (5.5)$$

An undirected weighted network is used in this work, without loss of generality, assuming $w_{ij} = 1$ and $w_{ih} = 1$, when there is an edge between i and j , and between i and h . Considering Eq.(5.1), this result can be deduced: $k_i = \frac{\sum_{j,h} a_{ij}a_{ih}a_{jh}}{C_i^w s_i} + 1$, and as the limit cases, when $C_i^w = 0$, $k_i = \infty$ and when $C_i^w = \infty$, $k_i = 1$. It means that there is a like-mirrored relationship between “clustering coefficient” and “degree distribution”.

$$E[\mathcal{E}_{jk}] = \sum_{j,k} jk\mathcal{E}_{jk} = r\sigma_q^2 + \sum_{j,k} jkq_jq_k, \quad (5.6)$$

5.5 Prediction Algorithm

On the basis of the above recognition model, a prediction algorithm is proposed for disease dynamics. The algorithm consists of two parts: (i) acquiring structure knowledge by the recognition model, and (ii) estimating the parameter R by SRM.

1: **Begin** Prediction Algorithm:

2: **First part: Acquiring structure knowledge**

3: Input: a contact network of a disease outbreak, G_t

4: $C_i^w = \text{transitivity}(G_t, \text{type} = c(\text{local}))$ $\triangleright$ The function $\text{transitivity}(\cdot)$ is used to calculate the clustering coefficient of each node $i \in V_t$.

5: $S_i = \text{graph.strength}(G_t)$ $\triangleright$ The function $\text{graph.strength}(\cdot)$ is used to calculate the strength of each node $i \in V_t$.

6: =====

7: $\triangleright$ This for-loop structure is used to calculate the degree of each node $i \in V_t$, k_i .

8: **for** (i in $1 : |V_t|$) **do**

9: $\quad tmp = \frac{\sum_{j,h} a_{ij}a_{ih}a_{jh}}{C_i^w s_i} + 1$

$\triangleright tmp$ is a temporary variable.

10: $\quad K = \text{combine}(K, tmp)$

$\triangleright K$ is the set of k_i ($i \in V_t$).

11: **end for**

12: =====

13: $\triangleright$ This for-loop structure is used to calculate $E[k_i]$ that is the expectation of K .

14: **for** (i in $1 : |K|$) **do**

15: $\quad E[k_i] = E[k_i] + K[i] * K[i] / \text{sum}(K)$

16: **end for**

17: =====

18: $\triangleright$ This for-loop structure is used to calculate $E[X]$ that is the expectation of degree distribution.

19: **for** (i in $1 : |dg|$) **do**

$\triangleright dg = \text{degree}(G_t)$ is the degree distribution of network G_t .

20: $\quad E[X] = E[X] + dg[i] * dg[i] / \text{sum}(dg)$

$\triangleright dg[\cdot]$ is used to indicate the degree of a node.

```

21: end for
22: =====
23:  $r = assortativity.degree(G_t, directed = FALSE)$   $\triangleright$  The function assortativity.degree(.) is used
    to calculate the degree correlation of each node  $i \in V_t$ .
24: =====
25:  $E_t = get.edges(G_t)$   $\triangleright$  The function get.edges(.) is used to get all edges of network  $G_t$ .
26: =====  $\triangleright$  This
    for-loop structure is used to calculate the two parts of  $E[\mathcal{E}_{jk}]$ : (i)  $\mathcal{E}_1: \sigma_q^2$ , and (ii)  $\mathcal{E}_2: \sum_{j,k} jkq_jq_k$ .
27: for ( $i$  in  $1 : |E_t|$ ) do
28:      $\triangleright$  The function which(.) is used to get the numbers of the two nodes jointed by an edge.
29:      $j = which(V_t == E_t[i, 1])$ 
30:      $k = which(V_t == E_t[i, 2])$ 
31:      $q_j = \frac{p_{j+1}}{\sum_{k \geq 1} p_k}$   $\triangleright$  Calculate  $q_j$ .
32:      $q_k = \frac{p_{k+1}}{\sum_{j \geq 1} p_j}$   $\triangleright$  Calculate  $q_k$ .
33:      $\triangleright$  Calculate the two parts of  $\sigma_q^2$ : (i)  $\sigma_1: \sum_k k^2 q_k$ , and (ii)  $\sigma_2: \sum_k k q_k$ .
34:      $\sigma_1 = \sigma_1 + dg[k] * dg[k] * q_k$ 
35:      $\sigma_2 = \sigma_2 + dg[k] * q_k$ 
36:      $\mathcal{E}_2 = \mathcal{E}_2 + dg[j] * dg[k] * q_j * q_k$   $\triangleright$  Calculate  $\sum_{j,k} jkq_jq_k$ .
37: end for
38:  $\sigma_q^2 = \sigma_1 - [\sigma_2]^2$ 
39:  $E[\mathcal{E}_{jk}] = r * \sigma_q^2 + \mathcal{E}_2$ 
40: =====
41:  $M = \frac{E[k_i] + E[X] + E[\mathcal{E}_{jk}]}{3}$   $\triangleright$  Combine these three expectations:  $E[k_i]$ ,  $E[X]$  and  $E[\mathcal{E}_{jk}]$ .
42: =====
43: Second part: Estimating the parameter  $R$ 
44:  $SRM(\beta = M, \gamma = M, no.iteration)$   $\triangleright$  “no.iteration” is used to indicate the number of iterations.
    SRM: Spatial Risk Model
45: =====
46: Output: a predicted value of the parameter  $R$ , related to input network  $G_t$ 
47: End

```

5.6 Evaluation

5.6.1 Experimental Setup

A series of time-aware Ebola cases² (941 nodes) and their relationships (938 edges) are collected and used to evaluate the performance of the proposed prediction algorithm. These cases and relationships come from the outbreak of Ebola in West Africa from March 2014. The network structure of the corresponding RCN is dynamically changed during this outbreak, with the spread of the Ebola virus disease. On this basis, nine networks are constructed according to the time stamps of cases. These networks are $G_t = \{G_1, G_2, \dots, G_9\}$, where $t = 1, 2, \dots, 9$ is corresponding to nine weeks of August (four weeks), September (four weeks) and October (one week) 2014. The detailed information of these nine networks is shown in Tab. 5.2. Graphx of Spark [152] is used to process these networks.

Table 5.2: Detailed Information of Nine Dynamic Networks, G_t ($t = 1, 2, \dots, 9$).

Number of Nodes	Number of Edges	λ (Standard Error)	Time Period
333	330	0.5045455 (0.02764892)	from 1st August to 7th August
340	337	0.504451 (0.0273577)	from 8th August to 14th August
340	337	0.504451 (0.0273577)	from 15th August to 21st August
531	528	0.5028409 (0.02182144)	from 22nd August to 31st August
647	644	0.5023292 (0.0197486)	from 1st September to 7th September
707	704	0.5021307 (0.01888457)	from 8th September to 14th September
779	776	0.501933 (0.01798362)	from 15th September to 21st September
916	913	0.5016429 (0.01657475)	from 22nd September to 30th September
941	938	0.5015991 (0.01635166)	from 1st October to 7th October

In Tab. 5.2, this process is used to obtain the values of λ (the exponent of degree distribution) and the corresponding standard errors for each network: a maximum-likelihood fitting is conducted to fit the degree distribution of each network into an exponential distribution, and then the corresponding values of λ and standard errors are obtained on the fitted exponential distribution, for each network. The maximum-likelihood fitting uses a maximum-likelihood estimation [153] to estimate the values of

2. Using the wireless devices carried by volunteers, the cases and relationships (contact) can be tracked and recorded. These devices are GPS-enabled, and the reported records include time stamps.

distribution parameters. Moreover, the sizes of the nine networks are increasing gradually: from 333 nodes (the first week of August) to 941 nodes (the first week of October), with the spread of Ebola virus disease.

For comparison, the original SRM based prediction algorithm that does not integrate a recognition model is used to conduct comparison experiments. It has recently been demonstrated that empirical human contact networks are best described as having exponential degree distributions [131, 154]. The prediction algorithm that does not integrate a recognition model is based on the networks which have exponential degree distributions. Moreover, the predicted results of two algorithms compare with the real values of the reproductive number R . These real values are counted from the collected outbreak data. As a standard, compared with real data, the performance of prediction algorithms can be measured, e.g., how far is it from the predicted results to the real data.

5.6.2 Experimental Results and Discussion

Figure 5.3 illustrates experimental results. Comparison is conducted in the predicted accuracy of the proposed prediction algorithm and the prediction algorithm that does not integrate a recognition model. The predicted results of two algorithms are compared with the real values of R . Extensive experiments are conducted for the dynamic network of each time period under the parameter of SRM $no.iteration = 100$. It means that the two algorithms are iterated 100 times to get the average of predicted values of R in each time period.

By analyzing the comparative results illustrated in Fig. 5.3, it is obvious that these two observations can be obtained. Moreover, following discusses how the proposed algorithm improves the accuracy of prediction as well.

- In terms of the predictive performance for the parameter R , the proposed prediction algorithm performs better than the prediction algorithm that does not integrate a recognition model. The relevant standard deviations³ of R 's real values and R 's predicted values are 100.133, 8.271831 and 3.146532, so the variation of the predicted values calculated by the proposed algorithm is closer to the variation of real values. The values corresponding to Fig. 5.3 and relevant three standard deviations are listed in Tab. 5.3.

The prediction performed by the proposed algorithm is based on acquiring the structure knowledge of a contact network. By the acquired structure knowledge, the more reasonable values of β and γ can be set in SRM. On the basis of such a parameter setting, it can help us to get better prediction results.

- In each time period, the deviation between $\mathfrak{R}$ and $\mathfrak{I}$ is different from the deviation between $\mathfrak{R}$ and $\mathfrak{U}$. $\mathfrak{R}$ denotes the real value, $\mathfrak{I}$ is the predicted value calculated by the proposed algorithm, and $\mathfrak{U}$ indicates the predicted value calculated by the algorithm that is used to compare with the

3. In statistics, the Standard Deviation (SD) is a measure that is used to quantify the amount of variation or dispersion of a set of data values.

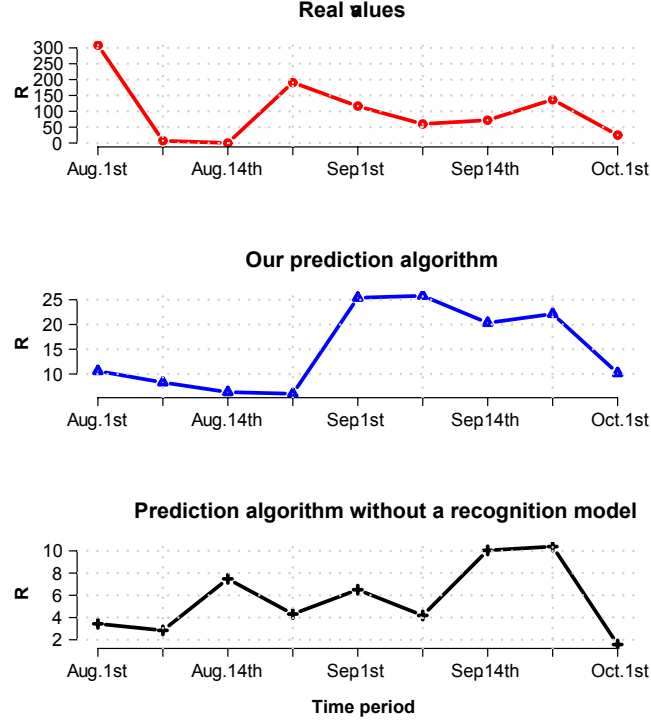


Figure 5.3: Comparative experiment results. Nine time periods from August to October are selected to construct nine dynamic networks. In each time period, based on the corresponding dynamic network, the predicted results from two algorithms are illustrated in the second and third sub-figures, respectively. The first sub-figure provides the real values of R , which are counted from the collected outbreak data.

proposed algorithm. The deviations between the real value and the predicted values (from two algorithms), for each time period, are listed in Tab. 5.4.

As shown in Tab. 5.4, the proposed algorithm has a smaller predicted deviation compared with the algorithm used for comparison. The proposed algorithm is based on the structure recognition of the RCN during the spread of a disease, and the infectious disease spreads through this RCN.

By recognizing the structure of a real disease transmission network, and using the calculated values of structure properties, setting the values of parameters β and γ to SRM can better conform to the actual situation. On this basis, compared with the algorithm used for comparison, the proposed prediction algorithm achieves a smaller predicted deviation to the real value for each time period. It also means that the proposed algorithm improves prediction accuracy.

Moreover, by analyzing the deviations shown in Tab. 5.4, it is observed that there are big deviations between real values and predicted values, for example, 298.40895 and 305.5625. Why some deviation values are so big? During the outbreak of an epidemic, only a subnetwork of the complete RCN can be obtained and used to carry out prediction, because of the limited number of mobile devices carried

Table 5.3: Values Corresponding to Fig. 5.3 and Relevant Three Standard Deviations

	Real Value	Proposed Prediction Algorithm	Prediction Algorithm that Does Not Integrate a Recognition Model
	309	10.59105	3.4375
	7	8.299401	2.842105
	0	6.337838	7.482759
	191	6.009524	4.315315
	116	25.39712	6.503185
	60	25.79491	4.186813
	72	20.32645	10.04628
	137	22.13499	10.3887
	25	10.17526	1.571429
Standard Deviation	100.133	8.271831	3.146532

Table 5.4: Deviations between $\mathfrak{R}$ and $\mathfrak{I}/\mathfrak{U}$.

Deviation between $\mathfrak{R}$ and $\mathfrak{I}$	Deviation between $\mathfrak{R}$ and $\mathfrak{U}$	Time Period
298.40895	305.5625	from 1st August to 7th August
1.299401	4.157895	from 8th August to 14th August
6.337838	7.482759	from 15th August to 21st August
184.990476	186.684685	from 22nd August to 31st August
90.60288	109.496815	from 1st September to 7th September
34.20509	55.813187	from 8th September to 14th September
51.67355	61.95372	from 15th September to 21st September
114.86501	126.6113	from 22nd September to 30th September
14.82474	23.428571	from 1st October to 7th October

by volunteers. It makes the number of individuals in the subnetwork be less than the total number of individuals (there is contact among these individuals during the spread of a disease). On the basis of this reason, the predicted values calculated by the two algorithms are much less than the real values.

If a complete RCN can be acquired during the disease spread, the prediction accuracy of the proposed algorithm can be improved. However, in addition to the absolute value of R , the changing trend is very important to reflect the prediction accuracy of an algorithm as well. From Fig. 5.3, the prediction results of the proposed algorithm follow the changing trend of real data better.

5.7 Conclusion

This work designs a recognition model to recognize the dynamic structure of the contact network during the spread of disease, and based on the structure knowledge mined by the recognition model, a prediction algorithm is proposed. By using the model, the prediction accuracy for disease dynamics is improved. By evaluating and comparing the accuracy of prediction for the time-varying disease dynamics parameter R , the performance of the proposed prediction algorithm has been verified, which can improve the prediction accuracy by using realistic structure knowledge that is mined by the recognition model. Moreover, in the comparison, the predicted results for R by two algorithms (proposed algorithm and the algorithm used to compare with) have been compared with the real values of R . These real values are counted from the collected surveillance data.

MCS-Routing: Eco-Friendly Vehicle Navigation by Mobile Collaborative Sensing

Contents

6.1	Introduction	78
6.2	Related Work	80
6.2.1	Eco-Friendly Navigation	80
6.2.2	Energy/Emission Estimation Models	81
6.3	System Model	81
6.4	Problem Formulation	82
6.5	Energy/Emission Model	83
6.6	Energy/Emission-related Cost Function	84
6.7	Algorithm Design	85
6.8	Evaluation and Results	86
6.8.1	Experimental Setup	86
6.8.2	Results and Discussion	87
6.9	Conclusion	89

This work proposes an eco-friendly navigation algorithm, *eRouting*, to save energy and reduce CO₂ emission. The important research issue of traffic information industry, eco-friendly navigation has been widely studied. As an improvement, in this work, combining real-time traffic information and a representative factor based energy/emission model, a calculated route is dynamically adjusted during the travel of a vehicle. *eRouting* is a centralized algorithm. It profits from the following aspects to achieve improved performance: (i) a representative factor based energy/emission model, (ii) real-time

traffic information based dynamic adjustment, and (iii) an objective function to control the optimization direction to optimize the final energy consumption of vehicle's travel. As a peculiarity of this work, the design of the representative factor based model mines the impact of road-level parameters on energy consumption and CO₂ emission. Such mining is helpful to improve the pertinence of a model by formulating the key influence factors into the model. Experimental results are presented to prove the validity of *eRouting*. In addition, by contrast experiments, *eRouting* shows improved performance compared with an eco-friendly navigation algorithm and three traditional navigation algorithms.

6.1 Introduction

In a total of 471 U.S. urban areas of 2014 [155], the extra energy cost on congestion was \$160 billion (3.1 billion gallons of fuel), and the additional carbon dioxide (CO₂) greenhouse gas was released into the atmosphere during congested conditions. Moreover, because of the excess exhaust emission during periods of congestion, the public health risk from the congestion is becoming a significant issue around the world [156]. Therefore, the issues of road-level energy consumption and exhaust emission need to be addressed, not only by improving vehicle efficiency and developing alternative fuel, [157], but also by making road travel more efficient as well. Designing an eco-friendly navigation algorithm¹ for the road travel of a vehicle is an effective solution. The eco-friendly navigation of traffic information industry [158] has attracted wide attention of researchers in recent decades [159–161], as an important part of constructing ecological and harmonious intelligent environments [162–164].

In a traffic system, during the road travel of a vehicle, a significant portion of energy consumption and exhaust emission is related with: (i) traffic segment features, and (ii) real-time traffic conditions. For example, on a congested traffic segment (road congestion [165,166]), the driving velocity of a vehicle changes frequently, and such frequent change results in extra energy consumption and exhaust emission (Figure 6.1).

The concept of eco-friendly navigation has been developed [167]. However, almost all of the existing eco-friendly navigation algorithms only consider static environments, and ignore the different impacts of different factors on energy consumption and exhaust emission.

In this work, an eco-friendly navigation algorithm, *eRouting*, is designed by combining real-time traffic information and a representative factor based energy/emission model. Based on the off-the-shelf infrastructure of an intelligent traffic system [168], the traffic information is collected. A scenario of the traffic information collection is illustrated in Figure 6.2. The collected information comes from different traffic segments, and it is: (i) the feature information of traffic segments, road grade G_r ², and (ii) real-time traffic information, traffic velocity v (it is estimated from the driving velocity of vehicles).

1. In this work, it is energy-saving and emission-reduced navigation, during the road travel of a vehicle. Such navigation enables minimal or no harm upon ecosystems or the environment.

2. The road grade value of a traffic segment is: the tangent of the angle of the road surface to the horizontal times 100%. The road grade is signed as a percentage.

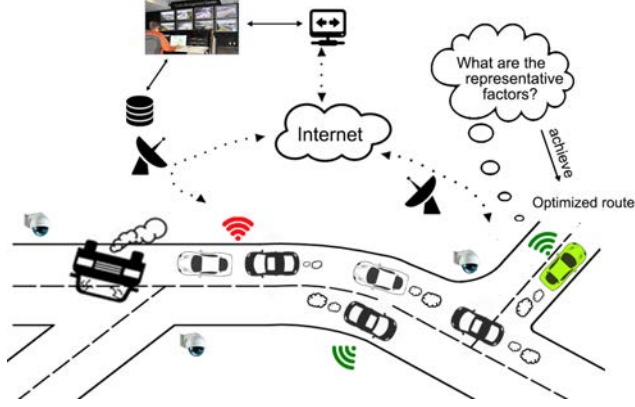


Figure 6.1: Problem description. In this scenario, there is a crash on a road. Such a sudden crash causes the change of vehicle's driving velocity. For adapting to this change, extra energy consumption is needed to control the driving velocity. How to avoid such energy consumption to make an energy-optimized route decision? Global real-time traffic information is necessary, and combining a representative factor based energy/emission model can help improve the pertinence of such route optimization for navigation.



Figure 6.2: A scenario of traffic information collection. The information is collected from an intelligent traffic system. In this traffic system, there are Automatic Number Plate Recognition (ANPR) cameras, in-vehicle and GPS-enabled wireless devices, and inductive loops built into road surfaces. Based on various devices, massive real-time data can be collected, regarding traffic information from different traffic segments.

As the points that are different from traditional algorithms to solve the eco-friendly navigation problem, in *eRouting*: (i) real-time traffic information is used to achieve dynamic navigation, and (ii) representative factors are investigated and mined to be used into the model design, for improving the efficiency and pertinence of a model.

This work investigates the impact of two representative factors (G_r and Δv) on energy consumption³

3. In this work, fuel consumption is used to measure and estimate the energy consumption.

and exhaust emission, with the help of massive data from an intelligent traffic system. Then, based on this investigation, a representative factor based energy/emission model is designed. This model can be used to estimate the costs of relevant traffic segments. Finally, with the calculated costs, *eRouting* is proposed to achieve dynamic navigation. The evaluation of *eRouting* performance is conducted based on the real data from the England traffic system, and the results of this evaluation are analyzed in detail.

Scientific contributions of this research work are listed as follows:

- G_r and Δv are selected as representative factors, and the reason is analyzed in detail.
- *eRouting* improves the performance of eco-friendly navigation by combining real-time traffic information and a representative factor based energy/emission model.
- The representative factor based energy/emission model mines the impact of road-level key influence factors on the vehicle's energy consumption and CO₂ emission. Such mining is helpful to improve the efficiency and pertinence of a model by formulating the key influence factors into the model.
- The pertinence of a model can help to avoid the negative effects of multi-information on the route decision. For example, there is a cumulative effect in the information error: more information sources will cause a higher error with a higher probability.
- An objective function is designed to control, step by step, the optimization direction in order to optimize the total energy consumption of vehicle's travel.

6.2 Related Work

6.2.1 Eco-Friendly Navigation

Eco-friendly navigation systems are used to find a route that saves energy and/or reduces emission [169–171], compared with the traditional navigation systems that are not aware of the “eco-friendliness”. These studies have shown that selecting different travel routes between a start-destination pair can result in significant differences in the amount of consumed energy and produced exhaust. In [169], Eva Ericsson *et al.* found that 46% of the trips were not made on the most energy-efficient route, and these trips can save, on average, an 8.2% of energy by using an energy-optimized navigation system. Therefore, there is much to be done in terms of finding energy-efficient and emission-optimized routes.

These previous studies use all kinds of traffic information from different sources to improve the performance of eco-friendly navigation, for example, historical and real-time traffic information from multiple data sources. However, such information is statically used by existing eco-friendly navigation algorithms, and these algorithms ignore the different impacts of different factors on energy consumption and exhaust emission. A dynamic algorithm is effective to improve the performance of the navigation that is used in a time-varying environment. Traffic information is dynamically changed along with the

travel of a vehicle⁴, so a real-time adjustment is necessary to obtain efficient navigation. That is to say, the first route that is calculated at the start of navigation, needs to be adjusted in real time. Moreover, designing a representative factor based energy/emission model can help to improve the pertinence of optimization, and avoid the negative effects of weak correlation factors. For example, when a vehicle makes a route decision, huge parameters can be used and considered into such a decision. In these parameters, the impacts are different for different parameters on vehicle's energy consumption, and if these impacts cannot be clearly distinguished, and even setting these parameters has the same impacts, the performance of eco-friendly navigation will be seriously affected.

6.2.2 Energy/Emission Estimation Models

As an important part of *eRouting*, a cost function is used to measure the cost of each traffic road. For designing the cost function, the model of estimating energy consumption and exhaust emission is necessary. This section introduces the existing models of such estimation.

For estimating the energy/emission of a vehicle, there are two kinds of models which can be used, microscopic models and mesoscopic models. As the typical microscopic models, Comprehensive Modal Emissions Model (CMEM) [172] and VT-Micro Model [173] are used to accurately estimate the energy/emission of a vehicle. However, such microscopic models require extensive input data from the vehicle themselves, e.g., second-by-second vehicle velocity information, and such second-by-second data processing produces a high computational cost. On this basis, such models are not suitable for real-time applications such as road navigation. As an alternative, the mesoscopic models seem to be a more viable and practical option. Such models estimate energy/emission as a function of a set of road-level variables. Based on the mesoscopic models, a high-accuracy instantaneous model is proposed in [174]. This model uses the real measurement for the relationships between parameters to estimate energy consumption, and the parameters are measured at any instant during the travel of a vehicle. The instantaneous model is formulated as a function of the tractive power required by the vehicle. However, using mesoscopic models to provide a road-level energy/emission estimation that is used in road navigation, requires significant energy/emission measurement efforts. The proposed energy/emission model in this work extends the instantaneous model, and requires less measurement efforts.

6.3 System Model

A time-aware weighted directed graph $G = (S, E, C, T)$ is used as the system model of this work, where S is the set of k vertices related to the intersections of a traffic system, E is the set of n directed edges connecting the vertices (these edges are corresponding to traffic segments), C is the set of h costs associated with the directed edges, and T is a time stamp. For example, the cost of the edge

4. For example, (i) road grade: different selectable traffic roads have different road grades, which impact the energy consumption and exhaust emission during vehicle's travel, and (ii) traffic velocity: it changes with time.

$e_{i,j} = (s_i, s_j) \in E$ is denoted as $c_{i,j}$ at the time point $T = x$. Moreover, a route p includes a series of edges, and the cost of the route p is denoted as c_p , so $c_p = \sum_{e_{i,j} \in p} c_{i,j}$, and the corresponding time series is $T = 1, T = 2, \dots, T = m$ (m is the number of intersections included in the route p). The navigation for an object m is represented as a selection process of edges for constituting the route p_m . Thus, a binary decision variable is defined as $x_{ijm} \in \{0, 1\}$ with $x_{ijm} = 1$, if in the route p_m the vertex s_j is visited after the vertex s_i , where $s_i \in S_{r_1}$, $s_j \in S_{r_2}$, $r_1 \neq r_2$ and $r_1, r_2 \in \{1, \dots, R\}$ ⁵. For the edge $e_{i,j} = (s_i, s_j) \in E$ with the associated decision variable x_{ijm} , the edge-traversal cost $c_{i,j}$ is defined as a function of energy consumption and exhaust emission (it is shown in Eq.(6.14)).

Table 6.1 lists the signs which will be used in this work.

Table 6.1: Signs and Description

Sign	Description	Unit
G_r	Road grade	%
v	Traffic velocity	m/s
P_T	Total tractive power of a vehicle	kw
P_C	Cruise component of total tractive power	kw
P_I	Inertia component of total tractive power	kw
P_{G_r}	Grade component of total tractive power	kw
P_{max}	Maximum tractive power	kw
$c_{i,j}$	Energy/emission-related cost of the traffic segment $e_{i,j}$	mL

6.4 Problem Formulation

The objective of navigation is to seek a route for an object m based on real-time traffic information, and this route optimizes the total cost of the travel on the route. It can be formulated as follows.

$$\min \sum_{s_i \in S} \sum_{s_j \in S} c_{i,j} x_{ijm}, \quad x_{ijm} = \{0, 1\}, \quad (6.1)$$

subject to,

$$\sum_{s_j \in S_r} \sum_{s_i \in S} x_{ijm} = 1, \forall r \in \{1, \dots, R\}, \quad (6.2)$$

5. $|R|$ is the number of intersections in an actual traffic system. Correspondingly, in the graph G , $|R|$ is the number of vertex partitions.

$$\sum_{s_i \in S_r} \sum_{s_j \in S} x_{ijm} = 1, \forall r \in \{1, \dots, R\}, \quad (6.3)$$

$$\sum_{s_i, s_j \in S} x_{ikm} - x_{kjm} = 0, \forall s_k \in S. \quad (6.4)$$

The objective function that is shown in Eq.(6.1) seeks to optimize the energy consumption and exhaust emission with optimizing the travelling route of object m . Equations 6.2 and 6.3 are the constraints of the objective function, and these constraints guarantee that each vertex partition (intersection) along a travelling route is visited just once. This means that any cycle is not existent on the route. The constraint Eq.(6.4) checks that the in-degree and the out-degree of each vertex are equal for a route.

6.5 Energy/Emission Model

In this work, a mesoscopic model is proposed. This model estimates the rate of energy consumption (mL/s) as a value per unit time, as a function of the tractive power required by a vehicle, and as a function of the driving velocity of the vehicle. The energy/emission model extends from an instantaneous model introduced in [174], and uses two groups of parameters: (i) vehicle parameters, and (ii) traffic and road parameters. As a kind of traffic parameter, traffic velocity is estimated by the data collected from: (i) GPS-enabled buses and taxis and, (ii) the cameras and inductive loops that are mounted on roads. With fixed routes and schedules, the buses shuttle among different traffic roads. By continuously monitoring and analyzing GPS traces and the data from the mounted cameras and inductive loops, the traffic velocities of different traffic roads can be estimated. Detailed modeling procedure is given as follows.

When the tractive power P_T is equal to or less than zero, the rate of energy consumption f_t is equal to a typically small constant value. Otherwise, the rate of energy consumption is mainly dependent on: instantaneous acceleration a , driving velocity v_d , and tractive power P_T .

$$f_t = \begin{cases} \alpha + \beta_1 P_T + [\beta_2 a P_I]_{a>0}, & \text{for } P_T > 0, \\ \alpha, & \text{for } P_T \leq 0, \end{cases} \quad (6.5)$$

$$P_T = \min(P_{max}, P_C + P_I + P_{G_r}), \quad (6.6)$$

$$P_C = b_1 v_d + b_2 v_d^3, \quad (6.7)$$

$$P_I = M_v a v_d, \quad (6.8)$$

$$P_{G_r} = 9.81 M_v (\sin G_r) v_d, \quad (6.9)$$

$$\alpha = f_i / 3600, \quad (6.10)$$

where,

- f_t is the rate of energy consumption (mL/s).
- β_1 is the efficiency parameter of consumed energy, and it is the energy consumption per unit of work.
- β_2 is the efficiency parameter of consumed energy during positive acceleration, and the positive acceleration is used to produce the acceleration rate and inertia power of a vehicle.
- P_{max} is the maximum tractive power (kw).
- α is the constant idle energy consumption rate of a vehicle (mL/s), which applies during all driving modes (as an estimate of energy used to maintain engine in operation).
- f_i is the constant idle energy consumption rate in mL/h .

Moreover, the cruise energy consumption rate of a vehicle ($a = 0, P_I = 0$) on a flat road ($G_r = 0, P_{G_r} = 0$) is calculated by:

$$f_{ct} = \alpha + \beta_1 P_C = \alpha + \beta_1 (b_1 v_d + b_2 v_d^3). \quad (6.11)$$

Rate of Carbon Dioxide (CO_2) emission (g/s , as a value per unit time) is estimated directly from the rate of energy consumption:

$$f_t(CO_2) = f_{CO_2} f_t(energy), \quad (6.12)$$

where $f_t(energy)$ is the rate of energy consumption in mL/s , and f_{CO_2} is the correlation coefficient between the rate of CO_2 emission and the rate of energy consumption, in grams per millilitre (g/mL) or kg per litre (kg/L). For different types of vehicles, the correlation coefficient is different, e.g., three kinds of typical vehicles that are used in the experiments of this work: the correlation coefficient is 2.5 for the “passenger car” and “light vehicle”, and is 2.6 for the “heavy vehicle”.

6.6 Energy/Emission-related Cost Function

In this work, the energy/emission-related cost of each traffic road, $c_{i,j}$, is calculated or updated by using available traffic information. Based on the energy/emission model constructed in Section 6.5, the corresponding cost function that is used to calculate the cost is designed and formulated as:

$$C = f_t(energy) * \frac{D_E}{V_E}, \quad (6.13)$$

where C is the cost set of traffic roads, D_E is the length set of the traffic roads, and V_E is the traffic velocity set.

By the formulation of Eq.(6.13), the cost function of each traffic road can be described as:

$$c_{i,j} = f_{ijt}(energy) * \frac{d_{e_{i,j}}}{v_{e_{i,j}}}, \quad (6.14)$$

where $d_{e_{i,j}}$ denotes the length of traffic road $e_{i,j}$, and $v_{e_{i,j}}$ is the traffic velocity on road $e_{i,j}$. Using $\frac{d_{e_{i,j}}}{v_{e_{i,j}}}$, the elapsed time on road $e_{i,j}$ can be calculated.

This work is to optimize the energy consumption during the travel on traffic roads E^6 by the calculated results of Eq.(6.13) and Eq.(6.14). Moreover, because: (i) there is a proportional relationship between energy consumption and exhaust emission, and (ii) the emission of exhaust for a vehicle is because of energy consumption, the exhaust emission during the travel is optimized as well, with optimizing the energy consumption.

6.7 Algorithm Design

Detailed steps of algorithm *eRouting* are described in Algorithm 1.

Algorithm 1 *eRouting* $\triangleright$ Each vehicle runs the same steps at each intersection along the calculated route

```

1: Input: Weighted directed graph  $G$  with vertex partitions (intersections) and real-time traffic in-
   information from relevant traffic roads.  $\triangleright G$  has vertices  $s_1, \dots, s_k$  and weight  $c_{i,j}$  corresponding to
   edge  $e_{i,j}$ , where  $c_{i,j} = \infty$  if  $e_{i,j} = (s_i, s_j)$  is not an edge of  $G$ .
2: Output: A feasible route for vehicle  $m$ .  $\triangleright$  Using this route, the total cost of vehicle travel can be
   optimized.
3: Start
4:  $c_p := 0$   $\triangleright c_p$  is the cost from start vertex  $s_s$  to destination vertex  $s_d$ .
5: for each vertex  $s_i$  in  $G$  do  $\triangleright$  Initialization
6:   if  $s_i \neq \text{destination } (s_d)$  then
7:      $c_{s,i} := \infty$   $\triangleright$  Unknown cost from the start vertex  $s_s$  to the vertex  $s_i$ .
8:      $previous[s_i] := \text{undefined}$   $\triangleright$  A set of previous vertices. It starts from the start vertex  $s_s$ .
9:   end if
10:  Add  $s_i$  to  $P$ .  $\triangleright P = \{s_s\}$  in the initial state.
11: end for
12: while  $P$  is not empty do  $\triangleright$  Main loop.
13:   $s_u :=$  vertex in  $P$  with the optimized cost from  $s_s$  to  $s_u$ .
14:  Remove  $s_u$  from  $P$ .
15:  for each neighbor  $s_n$  of  $s_u$  do  $\triangleright s_n$  has not yet been removed from  $P$ .
16:     $alt := c_{s,u} + c_{u,n}$ 
17:    if  $alt < c_{s,n}$  then  $\triangleright$  A route to  $s_n$  has been found.
18:       $c_{s,n} := alt$ 
19:       $previous[s_n] := s_u$ 
20:    end if
21:  end for
22: end while
23: return  $c_{s,d}, previous[.]$ 
24: End

```

6. $E = \{e_{i,j}; i \neq j, i, j = 0, 1, \dots\}$, where i and j are two different unique numbers. They indicate two different ports of two different intersections. These two ports are linked by a traffic road.

eRouting is a dynamic algorithm. It adjusts the navigable route of a vehicle at each intersection. With this dynamic adjustment, *eRouting* achieves the target of optimizing the cost during road travel. Even if *eRouting* is a dynamic algorithm, its costs of communications and calculation are lower than traditional GPS-based navigation algorithms. A vehicle only receives navigation signals from the control center of a traffic system, and the real-time traffic information of the traffic system is submitted by the sensing devices mounted on traffic roads, buses and taxis (they are all public transport facilities). Each target vehicle receives navigation signals at “intersections”. Conversely, traditional GPS-based navigation algorithms, frequently receive signals for an uninterrupted route adjustment.

eRouting is based on the Dijkstra algorithm [175] with the energy/emission-related cost function and the objective function defined in Eq.(6.1), which aims at seeking a route that optimizes the total cost of the travel on the route. Moreover, by using the objective function, the optimized direction is under control.

Time complexity of *eRouting* is the number of selectable next intersections ($S' \subseteq S$) times ($\times$) the number of the hops from the current intersection to the intersection of the destination ($E' \subseteq E$), and it can be denoted as: $O(|S'| \times |E'|)$.

eRouting improves the performance of navigation by enabling the algorithm to be suitable for a time-varying environment, and by using a representative factor based energy/emission model to improve the pertinence of optimization.

On the basis of the above description concerning the dynamic of *eRouting*, it is necessary to clearly answer an important question for the “dynamic” property of the *eRouting* algorithm. The answer helps to clarify and highlight the special features of *eRouting*. What are the advantages of the “dynamic” property in comparison with other published eco-routing algorithms, in particular, the algorithm proposed in [176]? An objective function is designed in *eRouting* (Eq.(6.1 6.2 6.3 6.4)), and this objective function includes some constraints. These constraints guarantee that each intersection along a travelling route is visited just once. This means that any loop is not existent on the currently calculated route, which can help to optimize the total energy consumption and exhaust emission of vehicle travel. On this basis, at each intersection, new route is calculated to dynamically adjust the preceding calculated route, and there is no any loop on all of these calculated routes, and step by step, the final route is optimized and the optimization objective (optimizing the total energy consumption and exhaust emission of the travel on the route) can be achieved.

6.8 Evaluation and Results

6.8.1 Experimental Setup

Table 6.2 lists the experimental parameters and corresponding value ranges of *eRouting*, where $\bar{a}$ is the average acceleration calculated for each relevant traffic road. Moreover, each parameter is described in the description column of the table.

Table 6.2: Parameters related to the Energy Consumption and Exhaust Emission of a Vehicle, and Corresponding Value Ranges

Parameter	Unit	Value	Description
α	mL/s	0.361	Idle energy consumption rate
β_1	mL/kJ	0.0900	Efficiency parameter
β_2	mL/(kJ·m/s ²)	0.0300	Energy-acceleration efficiency parameter
a ($\bar{a}$)	m/s ²	$\bar{a} = \frac{\Delta v}{\Delta t}$ (it is replaced with average acceleration, in this work)	Instantaneous acceleration
b_1	kJ/m	0.2222	Drag energy consumption parameter related to rolling resistance
b_2	kJ/m(m/s) ²	0.00072	Drag energy consumption parameter related to aerodynamic drag
v_d	m/s or (km/h)/3.6	current v_d	Driving velocity
M_v	kg	1250	Vehicle mass including occupants and any other loads
G_r	%	(-15, 15)	Road grade (negative if downhill)
P_{max}	kw	80	Maximum power
f_{CO_2}	g/mL	2.50	Correlation coefficient

6.8.2 Results and Discussion

This section presents experimental results to prove the validity of *eRouting*, and from the results of contrast experiments, *eRouting* displays the improved performance on energy consumption and exhaust emission. Meanwhile, the reasons behind these experimental outcomes are analyzed.

Moreover, in experiments, the route is calculated with real-time traffic information and the parameters of road and vehicle, e.g., v (the traffic velocity of a traffic segment), G_r (the grade of a road) and M_v (the mass of a vehicle).

The results are listed in Table 6.3. Five algorithms are implemented in the same way (based on real-time traffic information), and experiments are conducted under the same parameter configuration, to verify: (i) the performance (energy consumption and exhaust emission) of *eRouting* is really better than that of the other four algorithms, and (ii) the “dynamic” property (“real-time traffic information based recalculation” + “objective function based control in the optimization direction”) of *eRouting* is really effective to improve the navigation performance. Moreover, the results of the experiments are the average at different time points and in different locations. Twelve different start points and destinations are selected, to obtain the average energy consumption and exhaust emission for different

algorithms, in the 24 hours of a day⁷.

Table 6.3: Average energy consumption and exhaust emission for different algorithms during 24 hours (five algorithms are developed in the same way with using real-time traffic information)

	Energy Consumption (mL)	Exhaust Emission (g)
<i>eRouting</i>	355.563	888.913
Eco-friendly navigation algorithm [176]	381.069	952.679
Shortest distance navigation algorithm	376.086	940.229
Shortest time navigation algorithm	1767.495	4418.738
Random navigation algorithm	490.033	1225.09

In Table 6.3, compared with the other four algorithms, *eRouting* improves the performance from the aspects of energy consumption and exhaust emission. Based on these results, why *eRouting* can make such improvement, by comparing with other four algorithms.

- Eco-friendly navigation algorithm [176]. The routing process of this algorithm is based on the Dijkstra algorithm. On this basis, if recalculation is conducted at each intersection using real-time traffic information, without any constraint, the recalculated route has the risk of reusing some traffic segments, without treating the recalculation at each intersection as a continuous optimization process. Based on such reusing, the travel of vehicle on this final calculated route consumes more energy and emits more exhaust.
- Shortest distance navigation algorithm. The purpose of this algorithm focuses on calculating a route covering the shortest distance between a start point and a destination. The length of each traffic segment is a fixed value. In a real traffic system, under the time-varying traffic conditions, the route with the shortest distance does not mean that the route has the optimized energy consumption and exhaust emission.
- Shortest time navigation algorithm. The target of this algorithm is to find a route on which a vehicle can use the shortest time to complete the travel. Based on real-time traffic information and recalculation at each intersection, without any constraint, the reuse of routing segments is existent. Moreover, without considering the impact of road grade G_r , the energy consumption and exhaust emission cannot be optimized from the road grade aspect.
- Random navigation algorithm. This algorithm has no constraints. A vehicle can select any available route to reach a destination based on the driver's experience. There is a big risk for this algorithm on the energy consumption and exhaust emission of travel on the calculated route.

7. Through long-term observation, in the aspect of velocity of traffic segments, there is no a great difference between a normal weekday and a normal weekend.

Rerouting is an important issue that deserves discussion. It is proposed for using real-time data to improve the performance of navigation: the traffic conditions of each route segment are time-varying; a route is calculated from a starting point to a destination: $1 \rightarrow 2 \rightarrow 3$, but when a vehicle travels through the route segment 1, the traffic conditions of the traffic segment 2 and other selectable segments have changed, so at this moment, maybe there is another route segment which is a better choice than the segment 2; if we want to improve the performance of navigation, the recalculation is needed to get this better choice. This is a universal problem, and rerouting is a kind of feasible way to improve the performance of navigation by using real-time traffic information.

Reusing traffic segment problem is another important aspect which needs to be discussed. However reusing traffic segments is based on the traffic conditions. For example, if there is a traffic accident on the traffic segment e_x , it causes a vehicle which is driving on e_x to reroute back to the only optional traffic segment e_y . It means that the traffic segment e_y is reused; in a separate route calculation, segment reusing needs to be avoided from a starting point to a destination. Other methods using Dijkstra have chosen to avoid segment reusing, but in some conditions, it is necessary to reuse traffic segments. To improve the performance of traffic navigation, it is necessary to be able to distinguish in what conditions to reuse traffic segments and in what conditions to avoid segment reusing. *eRouting* can achieve this. The objective function can avoid the segment reusing in traffic navigation, and the recalculation of *eRouting* can make a vehicle reuse traffic segments under some special traffic conditions, e.g., a traffic accident caused a vehicle U turn after rerouting.

6.9 Conclusion

This work presents *eRouting*, a centralized navigation algorithm in traffic information industry. This algorithm can dynamically calculate an eco-friendly route for vehicle's travel, by using real-time traffic information, with optimizing energy consumption and exhaust emission. Along with the process of travel, the real-time traffic information includes: the traffic velocity v and road grade G_r , on different traffic segments. Moreover, the eco-friendly centralized navigation of a vehicle is formulated as a real-time route decision problem in a weighted directed graph with vertex partitions (intersections), and the problem is subject to the constraints about energy consumption and exhaust emission. Based on extensive experiments, the validity of *eRouting* is verified, and *eRouting* shows improved performance on energy consumption and exhaust emission.

Conclusion

Contents

7.1 Summary	91
7.1.1 Summary of Mobile Collaborative Sensing Framework	92
7.1.2 Summary of MCS-Locating	92
7.1.3 Summary of MCS-Prediction	92
7.1.4 Summary of MCS-Routing	93

7.1 Summary

This thesis studied and clarified the fundamental question:

How can a mobile collaborative sensing application be designed, in order to use sensor data to energy-efficiently and cost-effectively solve complex real-world problems?

Mobile Collaborative Sensing opens up a kind of cost-effective sensing paradigm to solve various complex problems by information sensing and information collaboration among different sensing participants.

This thesis presented a framework of Mobile Collaborative Sensing to facilitate the cooperativity of data collection, sharing, and analysis. The data comes from various sensors and sensing-capable devices which have been widely deployed in the real physical world. Massive spatio-temporal data is being collected from these sensors and devices on a daily basis.

As four key contributions of this thesis, following sections briefly summarize the Mobile Collaborative Sensing Framework, and three Mobile Collaborative Sensing based applications presented in this thesis.

7.1.1 Summary of Mobile Collaborative Sensing Framework

Heterogeneous sensing devices have been widely deployed in the real physical world. To improve the performance of sensing applications by integrating the physical information from heterogeneous sensing devices, the Mobile Collaborative Sensing framework is designed. In this framework, although there are various types of devices involved, mobile devices play the major role to participate in a sensing task. This sensing- and collaboration-based intelligent framework has the potential to improve the performance of sensing systems by providing better awareness to physical environments. Moreover, this framework analyzes the spatio-temporal data which is collected from device holders and sensor embedded devices, so it can achieve improved automated decision making by integrating the crowd wisdom of holders into machine intelligence. In addition, the challenges and open issues to develop the framework have been explored and discussed.

7.1.2 Summary of MCS-Locating

Mobile Collaborative Sensing is a kind of emerging sensing paradigm, which is able to be used into distributed problem resolution. With sensing, collecting and analyzing WiFi and Bluetooth signals from the social surroundings around us, a time-serial location estimation model has been designed and trained. Using this model, LiCS has been proposed in this work. It is an indoor target locating algorithm based on two aspects, (i) mobile devices carried by individuals, and (ii) a location estimation model which is trained by the collected data from individuals about WiFi and Bluetooth signals. The prototype system of LiCS has been developed, and long-term experiment has been conducted in actual physical environment. Experimental results have shown that LiCS achieves competitive locating accuracy without any special infrastructure. This work sets up a novel perspective to sensing based indoor target locating algorithms.

7.1.3 Summary of MCS-Prediction

In this work, a prediction algorithm has been proposed for disease dynamics. This algorithm considers the impact of network structure on disease dynamics. For recognizing the structure of RCNs during disease spread, a recognition model has been designed. By evaluating and comparing the accuracy of prediction for the time-varying reproductive number R , it is verified that the proposed prediction algorithm improves the prediction accuracy by considering realistic structure knowledge that is mined by the proposed recognition model. Moreover, in the comparison, the predicted results of R by two algorithms (the proposed algorithm and the algorithm used for comparison) have been compared with the real values of R . Such real values are counted from the collected surveillance data.

7.1.4 Summary of MCS-Routing

This work presents *eRouting*, a centralized navigation algorithm in traffic information industry. This algorithm can dynamically calculate an eco-friendly route for vehicle's travel, by using real-time traffic information, with optimizing energy consumption and exhaust emission. The real-time traffic information includes: the traffic velocity v and road grade G_r , on different traffic roads. Moreover, the eco-friendly centralized navigation for a vehicle is formulated as a real-time route decision problem in a weighted directed graph with vertex partitions (intersections), and the problem is subject to the constraints about energy consumption and exhaust emission. Based on extensive experiments, the validity of *eRouting* is verified, and *eRouting* shows improved performance on energy consumption and exhaust emission.

References

- [1] Z. Feng, Y. Zhu, Q. Zhang, L. M. Ni, and A. V. Vasilakos, "TRAC: Truthful auction for location-aware collaborative sensing in mobile crowdsourcing," in *IEEE International Conference on Computer Communications (INFOCOM)*. IEEE, 2014, pp. 1231–1239.
- [2] X. Sheng, J. Tang, and W. Zhang, "Energy-efficient collaborative sensing with mobile phones," in *INFOCOM, 2012 Proceedings IEEE*. IEEE, 2012, pp. 1916–1924.
- [3] R. Bauza, J. Gozalvez, and J. Sanchez-Soriano, "Road traffic congestion detection through cooperative vehicle-to-vehicle communications," in *IEEE Conference on Local Computer Networks (LCN)*. IEEE, 2010, pp. 606–612.
- [4] Y. Chen, G. M. Lee, L. Shu, and N. Crespi, "Industrial Internet of things-based collaborative sensing intelligence: framework and research challenges," *Sensors*, vol. 16, no. 2, p. 215, 2016.
- [5] R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: current state and future challenges," *IEEE Communications Magazine*, vol. 49, no. 11, pp. 32–39, 2011.
- [6] W. Z. Khan, Y. Xiang, M. Y. Aalsalem, and Q. Arshad, "Mobile phone sensing systems: A survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 402–427, 2013.
- [7] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell, "A survey of mobile phone sensing," *IEEE Communications Magazine*, vol. 48, no. 9, pp. 140–150, 2010.
- [8] M. Duchon, K. Wiesner, A. Müller, and C. Linnhoff-Popien, "Collaborative sensing platform for eco routing and environmental monitoring," in *International Conference on Sensor Systems and Software*. Springer, 2012, pp. 89–104.
- [9] E. Miluzzo, N. D. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. B. Eisenman, X. Zheng, and A. T. Campbell, "Sensing meets mobile social networks: the design, implementation and evaluation of the cenceme application," in *Proceedings of the 6th ACM conference on Embedded network sensor systems*. ACM, 2008, pp. 337–350.
- [10] R. Istepanian, S. Laxminarayan, and C. S. Pattichis, *M-health*. Springer, 2006.
- [11] E. Koukoumidis, L.-S. Peh, and M. R. Martonosi, "Signalguru: leveraging mobile phones for collaborative traffic signal schedule advisory," in *Proceedings of the 9th international conference on Mobile systems, applications, and services*. ACM, 2011, pp. 127–140.
- [12] S. Ilarri, O. Wolfson, and T. Delot, "Collaborative sensing for urban transportation," *IEEE Data Engineering Bulletin*, vol. 37, no. 4, pp. 3–14, 2014.
- [13] X. Bao and R. Roy Choudhury, "Movi: mobile phone based video highlights via collaborative sensing," in *Proceedings of the 8th international conference on Mobile systems, applications, and services*. ACM, 2010, pp. 357–370.
- [14] X. Bao and R. R. Choudhury, "Vupoints: collaborative sensing and video recording through mobile phones," *ACM SIGCOMM Computer Communication Review*, vol. 40, no. 1, pp. 100–105, 2010.
- [15] P. Zhang and M. Martonosi, "Locale: Collaborative localization estimation for sparse mobile sensor networks," in *Information Processing in Sensor Networks, 2008. IPSN'08. International Conference on*. IEEE, 2008, pp. 195–206.
- [16] T. Park, J. Lee, I. Hwang, C. Yoo, L. Nachman, and J. Song, "E-gesture: a collaborative architecture for energy-efficient gesture recognition with hand-worn sensor and mobile devices," in *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2011, pp. 260–273.
- [17] B. Liu, Y. Jiang, F. Sha, and R. Govindan, "Cloud-enabled privacy-preserving collaborative learning for mobile sensing," in *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*. ACM, 2012, pp. 57–70.
- [18] P. P. Jayaraman, C. Perera, D. Georgakopoulos, and A. Zaslavsky, "Efficient opportunistic sensing using mobile collaborative platform mosden," in *Collaborative Computing: Networking, Applications and Worksharing (Collaboratecom), 2013 9th International Conference*. IEEE, 2013, pp. 77–86.

- [19] S. S. Iyengar and R. R. Brooks, *Distributed sensor networks: sensor networking and applications*. CRC press, 2016.
- [20] Y. Liu, Y. He, M. Li, J. Wang, K. Liu, and X. Li, "Does wireless sensor network scale? a measurement study on greenorbs," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 10, pp. 1983–1993, 2013.
- [21] P. Dutta, P. M. Aoki, N. Kumar, A. Mainwaring, C. Myers, W. Willett, and A. Woodruff, "Common sense: participatory urban sensing using a network of handheld air quality monitors," in *Proceedings of the 7th ACM conference on embedded networked sensor systems*. ACM, 2009, pp. 349–350.
- [22] X. Hu, X. Li, E. C.-H. Ngai, V. C. Leung, and P. Kruchten, "Multidimensional context-aware social network architecture for mobile crowdsensing," *IEEE Communications Magazine*, vol. 52, no. 6, pp. 78–87, 2014.
- [23] H. Roitman, J. Mamou, S. Mehta, A. Satt, and L. Subramaniam, "Harnessing the crowds for smart city sensing," in *Proceedings of the 1st international workshop on Multimodal crowd sensing*. ACM, 2012, pp. 17–18.
- [24] B. Pan, Y. Zheng, D. Wilkie, and C. Shahabi, "Crowd sensing of traffic anomalies based on human mobility and social media," in *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2013, pp. 344–353.
- [25] V. Freschi, S. Delpriori, L. C. Klopfenstein, E. Lattanzi, G. Luchetti, and A. Bogliolo, "Geospatial data aggregation and reduction in vehicular sensing applications: the case of road surface monitoring," in *2014 International Conference on Connected Vehicles and Expo (ICCVE)*. IEEE, 2014, pp. 711–716.
- [26] S. Bluetooth, "Bluetooth specification," 2003.
- [27] C. Gomez, J. Oller, and J. Paradells, "Overview and evaluation of bluetooth low energy: An emerging low-power wireless technology," *Sensors*, vol. 12, no. 9, pp. 11 734–11 753, 2012.
- [28] R. Heydon, *Bluetooth low energy: the developer's handbook*. Prentice Hall, 2013.
- [29] T. Adame, A. Bel, B. Bellalta, J. Barcelo, and M. Oliver, "Ieee 802.11 ah: the wifi approach for m2m communications," *IEEE Wireless Communications*, vol. 21, no. 6, pp. 144–152, 2014.
- [30] D. Jiang and L. Delgrossi, "Ieee 802.11 p: Towards an international standard for wireless access in vehicular environments," in *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*. IEEE, 2008, pp. 2036–2040.
- [31] C. Cordeiro, D. Akhmetov, and M. Park, "Ieee 802.11 ad: Introduction and performance evaluation of the first multi-gbps wifi technology," in *Proceedings of the 2010 ACM international workshop on mmWave communications: from circuits to networks*. ACM, 2010, pp. 3–8.
- [32] T. Halonen, J. Romero, and J. Melero, *GSM, GPRS and EDGE performance: evolution towards 3G/UMTS*. John Wiley & Sons, 2004.
- [33] H. Holma and A. Toskala, *HSDPA/HSUPA for UMTS: high speed radio access for mobile communications*. John Wiley & Sons, 2007.
- [34] E. Dahlman, S. Parkvall, J. Skold, and P. Beming, *3G evolution: HSPA and LTE for mobile broadband*. Academic press, 2010.
- [35] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-advanced for mobile broadband*. Academic press, 2013.
- [36] P. Baronti, P. Pillai, V. W. Chook, S. Chessa, A. Gotta, and Y. F. Hu, "Wireless sensor networks: A survey on the state of the art and the 802.15. 4 and zigbee standards," *Computer communications*, vol. 30, no. 7, pp. 1655–1695, 2007.
- [37] M. Knight, "How safe is z-wave?[wireless standards]," *Computing and Control Engineering*, vol. 17, no. 6, pp. 18–23, 2006.
- [38] I. Koutsopoulos, "Optimal incentive-driven design of participatory sensing systems," in *IEEE International Conference on Computer Communications (INFOCOM)*. IEEE, 2013, pp. 1402–1410.
- [39] D. Estrin, K. M. Chandy, R. M. Young, L. Smarr, A. Odlyzko, D. Clark, V. Reding, T. Ishida, S. Sharma, V. G. Cerf, *et al.*, "Participatory sensing: applications and architecture [internet predictions]," *IEEE Internet Computing*, vol. 14, no. 1, pp. 12–42, 2010.
- [40] Q. Chi, H. Yan, C. Zhang, Z. Pang, and L. Da Xu, "A reconfigurable smart sensor interface for industrial wsn in iot environment," *Industrial Informatics, IEEE Transactions on*, vol. 10, no. 2, pp. 1417–1425, 2014.
- [41] L. Da Xu, W. He, and S. Li, "Internet of things in industries: A survey," *Industrial Informatics, IEEE Transactions on*, vol. 10, no. 4, pp. 2233–2243, Nov 2014.
- [42] V. Vyatkin, "Software engineering in industrial automation: State-of-the-art review," *Industrial Informatics, IEEE Transactions on*, vol. 9, no. 3, pp. 1234–1249, 2013.
- [43] M. R. Haas, "Knowledge gathering, team capabilities, and project performance in challenging work environments," *Management Science*, vol. 52, no. 8, pp. 1170–1184, 2006.
- [44] K. Albrecht, *Social Intelligence: The new science of success*. John Wiley & Sons, 2006.
- [45] P. Devadasan, H. Zhong, and S. Y. Nof, "Collaborative intelligence in knowledge based service planning," *Expert Systems with Applications*, vol. 40, no. 17, pp. 6778–6787, 2013.
- [46] V. Zheng, Y. Zheng, X. Xie, and Q. Yang, "Towards mobile intelligence: Learning from gps history data for collaborative recommendation," *Artificial Intelligence*, vol. 184, pp. 17–37, 2012.

- [47] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang, "Collaborative location and activity recommendations with gps history data," in *Proceedings of the 19th international conference on World wide web*. ACM, 2010, pp. 1029–1038.
- [48] V. Zheng, B. Cao, Y. Zheng, X. Xie, and Q. Yang, "Collaborative filtering meets mobile recommendation: A user-centered approach," in *AAAI*, vol. 10, 2010, pp. 236–241.
- [49] K. W.-T. Leung, D. L. Lee, and W.-C. Lee, "Clr: a collaborative location recommendation framework based on co-clustering," in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM, 2011, pp. 305–314.
- [50] C. Cheng, H. Yang, M. R. Lyu, and I. King, "Where you like to go next: Successive point-of-interest recommendation," in *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*. AAAI Press, 2013, pp. 2605–2611.
- [51] J. Bao, Y. Zheng, D. Wilkie, and M. F. Mokbel, "A survey on recommendations in location-based social networks," *ACM Transaction on Intelligent Systems and Technology*, 2013.
- [52] M. Sattari, M. Manguoglu, I. H. Toroslu, P. Symeonidis, P. Senkul, and Y. Manolopoulos, "Geo-activity recommendations by using improved feature combination," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 2012, pp. 996–1003.
- [53] W.-H. Lee, S.-S. Tseng, and W.-Y. Shieh, "Collaborative real-time traffic information generation and sharing framework for the intelligent transportation system," *Information Sciences*, vol. 180, no. 1, pp. 62–70, 2010.
- [54] B. M. Lee and J. Ouyang, "Intelligent healthcare service by using collaborations between iot personal health devices," *International journal of bio-science and bio-technology*, vol. 6, no. 1, pp. 155–164, 2014.
- [55] J. Pitt, A. Bourazeri, A. Nowak, M. Roszczynska-Kurasinska, A. Rychwalska, I. R. Santiago, M. L. Sanchez, M. Florea, and M. Sanduleac, "Transforming big data into collective awareness," *Computer*, vol. 46, no. 6, pp. 40–45, 2013.
- [56] R. S. Baron, "So right it's wrong: Groupthink and the ubiquitous nature of polarized group decision making," *Advances in experimental social psychology*, vol. 37, pp. 219–253, 2005.
- [57] M. Gardner and J. Bieker, "Data mining solves tough semiconductor manufacturing problems," in *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2000, pp. 376–383.
- [58] J. R. Hackman, *Collaborative intelligence: Using teams to solve hard problems*. Berrett-Koehler Publishers, 2011.
- [59] L. Argote, D. Gruenfeld, and C. Naquin, "Group learning in organizations," *Groups at work: Theory and research*, pp. 369–411, 2001.
- [60] N. Cai, M. Gholami, L. Yang, and R. W. Brennan, "Application-oriented intelligent middleware for distributed sensing and control," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 42, no. 6, pp. 947–956, 2012.
- [61] F. De Pellegrini, D. Miorandi, S. Vitturi, and A. Zanella, "On the use of wireless networks at low level of factory automation systems," *IEEE Transactions on Industrial Informatics*, vol. 2, no. 2, pp. 129–143, 2006.
- [62] X. Cao, J. Chen, Y. Xiao, and Y. Sun, "Building-environment control with wireless sensor and actuator networks: centralized versus distributed," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 11, pp. 3596–3605, 2010.
- [63] S. C. Lee, T. G. Jeon, H.-S. Hwang, and C.-S. Kim, "Design and implementation of wireless sensor based-monitoring system for smart factory," *Lecture Notes in Computer Science*, pp. 584–592, 2007.
- [64] V. C. Gungor and G. P. Hancke, "Industrial wireless sensor networks: Challenges, design principles, and technical approaches," *Industrial Electronics, IEEE Transactions on*, vol. 56, no. 10, pp. 4258–4265, 2009.
- [65] H.-J. Korber, H. Wattar, and G. Scholl, "Modular wireless real-time sensor/actuator network for factory automation applications," *Industrial Informatics, IEEE Transactions on*, vol. 3, no. 2, pp. 111–119, 2007.
- [66] P. Tavner, J. Xiang, and F. Spinato, "Reliability analysis for wind turbines," *Wind Energy*, vol. 10, no. 1, pp. 1–18, 2007.
- [67] C. Cecati, F. Guinjoan, P. Siano, and G. Spagnuolo, "Introduction to the special section on smart devices for renewable energy systems," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 3, pp. 1119–1121, 2013.
- [68] G. Zhabelova and V. Vyatkin, "Multiagent smart grid automation architecture based on iec 61850/61499 intelligent logical nodes," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 5, pp. 2351–2362, 2012.
- [69] J. Byun, B. Jeon, J. Noh, Y. Kim, and S. Park, "An intelligent self-adjusting sensor for smart home services based on zigbee communications," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 3, pp. 794–802, 2012.
- [70] T. Xia, L. Xi, X. Zhou, and J. Lee, "Condition-based maintenance for intelligent monitored series system with independent machine failure modes," *International Journal of Production Research*, vol. 51, no. 15, pp. 4585–4596, 2013.
- [71] J. Cecilio, P. Martins, J. Costa, and P. Furtado, "A configurable middleware for processing in heterogeneous industrial intelligent sensors," in *Intelligent Engineering Systems (INES), 2012 IEEE 16th International Conference on*. IEEE, 2012, pp. 145–150.
- [72] R. Wisniewski, M. Svenstrup, A. S. Pedersen, and C. S. Steiniche, "Certificate for safe emergency shutdown of wind turbines," in *American Control Conference (ACC), 2013*. IEEE, 2013, pp. 3667–3672.

- [73] V. C. Gungor, B. Lu, and G. P. Hancke, "Opportunities and challenges of wireless sensor networks in smart grid," *Industrial Electronics, IEEE Transactions on*, vol. 57, no. 10, pp. 3557–3564, 2010.
- [74] S. D. Ramchurn, P. Vytelingum, A. Rogers, and N. R. Jennings, "Putting the 'smarts' into the smart grid: a grand challenge for artificial intelligence," *Communications of the ACM*, vol. 55, no. 4, pp. 86–97, 2012.
- [75] M. Castells *et al.*, *Technopoles of the world: The making of 21st century industrial complexes*. Routledge, 2014.
- [76] K.-H. Lee, Y.-J. Lee, H. Choi, Y. D. Chung, and B. Moon, "Parallel data processing with mapreduce: a survey," *AcM SIGMOD Record*, vol. 40, no. 4, pp. 11–20, 2012.
- [77] F. J. Martínez-López and J. Casillas, "Artificial intelligence-based systems applied in industrial marketing: An historical overview, current and future insights," *Industrial Marketing Management*, vol. 42, no. 4, pp. 489–495, 2013.
- [78] Q. Yu, G. Medioni, and I. Cohen, "Multiple target tracking using spatio-temporal markov chain monte carlo data association," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.
- [79] W. R. Gilks, *Markov chain monte carlo*. Wiley Online Library, 2005.
- [80] K. Wang, H. Lu, L. Shu, and J. J. Rodrigues, "A context-aware system architecture for leak point detection in the large-scale petrochemical industry," *Communications Magazine, IEEE*, vol. 52, no. 6, pp. 62–69, 2014.
- [81] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. Srivastava, "Participatory sensing," in *Workshop on World-Sensor-Web (WSW) at SenSys*, 2006, pp. 1–5.
- [82] B. Guo, Z. Yu, D. Zhang, and X. Zhou, "From participatory sensing to mobile crowd sensing," *CoRR*, vol. abs/1401.3090, pp. 1–6, 2014.
- [83] B. Guo, D. Zhang, Z. Yu, Y. Liang, Z. Wang, and X. Zhou, "From the internet of things to embedded intelligence," *World Wide Web*, vol. 16, no. 4, pp. 399–420, 2013.
- [84] H. Chen, R. H. Chiang, and V. C. Storey, "Business intelligence and analytics: From big data to big impact," *MIS quarterly*, vol. 36, no. 4, pp. 1165–1188, 2012.
- [85] T. Bernecker, F. Graf, H.-P. Kriegel, N. Seiler, C. Türmer, and D. Dill, "Knowing: A generic data analysis application," in *Proceedings of the 15th International Conference on Extending Database Technology*. ACM, 2012, pp. 630–633.
- [86] R. Sipos, D. Fradkin, F. Moerchen, and Z. Wang, "Log-based predictive maintenance," in *Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 2014, pp. 1867–1876.
- [87] M. Li, W. Lou, and K. Ren, "Data security and privacy in wireless body area networks," *Wireless Communications, IEEE*, vol. 17, no. 1, pp. 51–58, 2010.
- [88] Q. Liu, X. Zhang, X. Chen, and L. Wang, "The resource access authorization route problem in a collaborative manufacturing system," *Journal of Intelligent Manufacturing*, vol. 25, no. 3, pp. 413–425, 2014.
- [89] L. I. Gómez, S. A. Gómez, and A. A. Vaisman, "A generic data model and query language for spatiotemporal olap cube analysis," in *Proceedings of the 15th International Conference on Extending Database Technology*. ACM, 2012, pp. 300–311.
- [90] W. Van Der Aalst, "Process mining: making knowledge discovery process centric," *ACM SIGKDD Explorations Newsletter*, vol. 13, no. 2, pp. 45–49, 2012.
- [91] L. Dilling and M. C. Lemos, "Creating usable science: Opportunities and constraints for climate knowledge use and their implications for science policy," *Global Environmental Change*, vol. 21, no. 2, pp. 680–689, 2011.
- [92] H. Kopetz, *Real-time systems: design principles for distributed embedded applications*. Springer, 2011.
- [93] M. Chau, S.-H. Li, S. Urs, S. Srinivasa, and G. A. Wang, *Intelligence and Security Informatics*. Springer, 2010.
- [94] M. Lenzerini, "Data integration: A theoretical perspective," in *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. ACM, 2002, pp. 233–246.
- [95] A. Doan, A. Halevy, and Z. Ives, *Principles of data integration*. Elsevier, 2012.
- [96] D. T. Larose, *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons, 2014.
- [97] M. Atkinson, R. Baxter, P. Brezany, O. Corcho, M. Galea, M. Parsons, D. Snelling, and J. van Hemert, *The data bonanza: improving knowledge discovery in science, engineering, and business*. John Wiley & Sons, 2013, vol. 90.
- [98] D. Hatley and I. Pirbhai, *Strategies for real-time system specification*. Addison-Wesley, 2013.
- [99] S. Yin, H. Luo, and S. X. Ding, "Real-time implementation of fault-tolerant control systems with performance optimization," *Industrial Electronics, IEEE Transactions on*, vol. 61, no. 5, pp. 2402–2411, 2014.
- [100] S. X. Ding, Y. Wang, S. Yin, P. Zhang, Y. Yang, and E. Ding, "Data-driven design of fault-tolerant control systems," in *Fault Detection, Supervision and Safety of Technical Processes*, vol. 8, no. 1, 2012, pp. 1323–1328.
- [101] S. Bajaj and R. Sion, "Trusteddb: A trusted hardware-based database with privacy and data confidentiality," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 3, pp. 752–765, 2014.
- [102] Y. Chen, L. Shu, A. M. Ortiz, N. Crespi, and L. Lv, "Locating in crowdsourcing-based dataspace: Wireless indoor localization without special devices," *Mobile Networks and Applications*, vol. 19, no. 4, pp. 534–542, 2014.

- [103] Z. Yang, C. Wu, and Y. Liu, "Locating in fingerprint space: wireless indoor localization with little human intervention," in *Proceedings of the 18th annual international conference on Mobile computing and networking*. ACM, 2012, pp. 269–280.
- [104] A. S. Paul and E. A. Wan, "Rssi-based indoor localization and tracking using sigma-point kalman smoothers," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 5, pp. 860–873, 2009.
- [105] H.-S. Ahn and W. Yu, "Environmental-adaptive rssi-based indoor localization," *IEEE Transactions on Automation Science and Engineering*, vol. 6, no. 4, pp. 626–633, 2009.
- [106] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 6, pp. 1067–1080, 2007.
- [107] Y. Gu, A. Lo, and I. Niemegeers, "A survey of indoor positioning systems for wireless personal networks," *IEEE Communications Surveys & Tutorials*, vol. 11, no. 1, pp. 13–32, 2009.
- [108] U. Varshney and R. Vetter, "Mobile commerce: framework, applications and networking support," *Mobile Networks and Applications*, vol. 7, no. 3, pp. 185–198, 2002.
- [109] M. Chen, S. Gonzalez, A. Vasilakos, H. Cao, and V. C. Leung, "Body area networks: A survey," *Mobile Networks and Applications*, vol. 16, no. 2, pp. 171–193, 2011.
- [110] A. Dey, J. Hightower, E. de Lara, and N. Davies, "Location-based services," *Pervasive Computing*, vol. 9, no. 1, pp. 11–12, 2010.
- [111] A. Savvides, C.-C. Han, and M. B. Strivastava, "Dynamic fine-grained localization in ad-hoc networks of sensors," in *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking (MOBICOM)*. ACM, 2001, pp. 166–179.
- [112] J. Scott and B. Dragovic, "Audio location: Accurate low-cost location sensing," *Pervasive Computing*, vol. 3468, pp. 1–18, 2005.
- [113] G. Borriello, A. Liu, T. Offer, C. Palistrant, and R. Sharp, "Walrus: wireless acoustic location with room-level resolution using ultrasound," in *Proceedings of the 3rd International Conference on Mobile systems, Applications, and Services (MobiSys)*. ACM, 2005, pp. 191–203.
- [114] K. Wu, J. Xiao, Y. Yi, M. Gao, and L. M. Ni, "Fila: fine-grained indoor localization," in *Proceedings of the 31st Annual International Conference on Computer Communications (INFOCOM)*. IEEE, 2012, pp. 2210–2218.
- [115] P. Bahl and V. N. Padmanabhan, "Radar: An in-building rf-based user location and tracking system," in *Proceedings of the 19th Annual International Conference on Computer Communications (INFOCOM)*. IEEE, 2000, pp. 775–784.
- [116] M. Youssef and A. Agrawala, "The horus wlan location determination system," in *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services (MobiSys)*. ACM, 2005, pp. 205–218.
- [117] Y. Chen, D. Lymberopoulos, J. Liu, and B. Priyantha, "Fm-based indoor localization," in *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services (MobiSys)*. ACM, 2012, pp. 169–182.
- [118] T. Roos, P. Myllymaki, and H. Tirri, "A statistical modeling approach to location estimation," *IEEE Transactions on Mobile Computing*, vol. 1, no. 1, pp. 59–69, 2002.
- [119] S. Jung, C.-o. Lee, and D. Han, "Wi-fi fingerprint-based approaches following log-distance path loss model for indoor positioning," in *Proceedings of the 1st MTT-S International Microwave Workshop Series on Intelligent Radio for Future Personal Terminals*. IEEE, 2011, pp. 1–2.
- [120] L. Eisen and R. J. Eisen, "Using geographic information systems and decision support systems for the prediction, prevention, and control of vector-borne diseases," *Annual review of entomology*, vol. 56, pp. 41–61, 2011.
- [121] A. Kumar and G. Hancke, "A zigbee-based animal health monitoring system," *IEEE Sensors Journal*, vol. 15, no. 1, pp. 610–617, Jan 2015.
- [122] L. Li, R.-L. Ge, S.-M. Zhou, and R. Valerdi, "Guest editorial: Integrated healthcare information systems," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 4, pp. 515–517, 2012.
- [123] D. A. Luke and K. A. Stamatakis, "Systems science methods in public health: dynamics, networks, and agents," *Annual Review of Public Health*, vol. 33, pp. 357–376, 2012.
- [124] H. Heesterbeek, R. M. Anderson, V. Andreasen, S. Bansal, D. De Angelis, C. Dye, K. T. Eames, W. J. Edmunds, S. D. Frost, S. Funk, *et al.*, "Modeling infectious disease dynamics in the complex landscape of global health," *Science*, vol. 347, no. 6227, pp. aaa4339–1–aaa4339–10, 2015.
- [125] W. Yang, M. Lipsitch, and J. Shaman, "Inference of seasonal and pandemic influenza transmission dynamics," *Proceedings of the National Academy of Sciences*, vol. 112, no. 9, pp. 2723–2728, 2015.
- [126] G. M. Vazquez-Prokopec, D. Bisanzio, S. T. Stoddard, V. Paz-Soldan, A. C. Morrison, J. P. Elder, J. Ramirez-Paredes, E. S. Halsey, T. J. Kochel, T. W. Scott, *et al.*, "Using gps technology to quantify human mobility, dynamic contacts and infectious disease dynamics in a resource-poor urban environment," *PloS ONE*, vol. 8, no. 4, p. e58802, 2013.
- [127] M. Woolhouse, "How to make predictions about future infectious disease risks," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 366, no. 1573, pp. 2045–2054, 2011.
- [128] W. E. R. Team *et al.*, "Ebola virus disease in west africa, the first 9 months of the epidemic and forward projections," *N Engl J Med*, vol. 371, no. 16, pp. 1481–1495, 2014.

- [129] T. Stadler, R. Kouyos, V. von Wyl, S. Yerly, J. Böni, P. Bürgisser, T. Klimkait, B. Joos, P. Rieder, D. Xie, *et al.*, “Estimating the basic reproductive number from viral sequence data,” *Molecular Biology and Evolution*, vol. 29, no. 1, pp. 347–357, 2012.
- [130] C. Groendyke, D. Welch, and D. R. Hunter, “A network-based analysis of the 1861 hagelloch measles data,” *Biometrics*, vol. 68, no. 3, pp. 755–765, 2012.
- [131] G. M. Ames, D. B. George, C. P. Hampson, A. R. Kanarek, C. D. McBee, D. R. Lockwood, J. D. Achter, and C. T. Webb, “Using network properties to predict disease dynamics on human contact networks,” *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 278, no. 1724, pp. 3544–3550, 2011.
- [132] Z. Mukandavire, S. Liao, J. Wang, H. Gaff, D. L. Smith, and J. G. Morris, “Estimating the reproductive numbers for the 2008–2009 cholera outbreaks in zimbabwe,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 21, pp. 8767–8772, 2011.
- [133] Y. J. Fan, Y. H. Yin, L. Da Xu, Y. Zeng, and F. Wu, “Iot-based smart rehabilitation system,” *IEEE Transactions on Industrial Informatics*, vol. 2, no. 10, pp. 1568–1577, 2014.
- [134] “Contact tracing app for the ebola outbreak in west africa,” 2014, <http://www.appsagainstebola.org/#the-app>.
- [135] J. Juang and Y.-H. Liang, “Analysis of a general sis model with infective vectors on the complex networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 437, pp. 382–395, 2015.
- [136] J. Marro and R. Dickman, *Nonequilibrium phase transitions in lattice models*, 2nd ed. Cambridge University Press, 2005.
- [137] M. E. Craft, “Infectious disease transmission and contact networks in wildlife and livestock,” *Phil. Trans. R. Soc. B*, vol. 370, no. 1669, p. 20140107, 2015.
- [138] G. Q. Sun, A. Chakraborty, Q.-X. Liu, Z. Jin, K. E. Anderson, and B.-L. Li, “Influence of time delay and nonlinear diffusion on herbivore outbreak,” *Communications in Nonlinear Science and Numerical Simulation*, vol. 19, no. 5, pp. 1507–1518, 2014.
- [139] A. Cardillo, C. Reyes-Suárez, F. Naranjo, and J. Gómez-Gardeñes, “Evolutionary vaccination dilemma in complex networks,” *Physical Review E*, vol. 88, no. 3, p. 032803, 2013.
- [140] C.-y. Xia, Z. Wang, J. Sanz, S. Meloni, and Y. Moreno, “Effects of delayed recovery and nonuniform transmission on the spreading of diseases in complex networks,” *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 7, pp. 1577–1585, 2013.
- [141] G.-Q. Sun, Q.-X. Liu, Z. Jin, A. Chakraborty, and B.-L. Li, “Influence of infection rate and migration on extinction of disease in spatial epidemics,” *Journal of Theoretical Biology*, vol. 264, no. 1, pp. 95–103, 2010.
- [142] L. E. Charles-Smith, T. L. Reynolds, M. A. Cameron, M. Conway, E. H. Lau, J. M. Olsen, J. A. Pavlin, M. Shigematsu, L. C. Streichert, K. J. Suda, *et al.*, “Using social media for actionable disease surveillance and outbreak management: A systematic literature review,” *PloS ONE*, vol. 10, no. 10, p. e0139701, 2015.
- [143] S. Brennan, A. Sadilek, and H. Kautz, “Towards understanding global spread of disease from everyday interpersonal interactions,” in *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*. AAAI Press, 2013, pp. 2783–2789.
- [144] M. Newman and C. R. Ferrario, “Interacting epidemics and coinfection on contact networks,” *PloS ONE*, vol. 8, no. 8, p. e71321, 2013.
- [145] A. Saumell-Mendiola, M. Á. Serrano, and M. Boguñá, “Epidemic spreading on interconnected networks,” *Physical Review E*, vol. 86, no. 2, p. 026106, 2012.
- [146] S. Eubank, H. Guclu, V. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, “Modelling disease outbreaks in realistic urban social networks,” *Nature*, vol. 429, no. 6988, pp. 180–184, 2004.
- [147] S. Meyer, L. Held, *et al.*, “Power-law models for infectious disease spread,” *The Annals of Applied Statistics*, vol. 8, no. 3, pp. 1612–1639, 2014.
- [148] Y. Shang, “Mixed si(r) epidemic dynamics in random graphs with general degree distributions,” *Applied Mathematics and Computation*, vol. 219, no. 10, pp. 5042–5048, 2013.
- [149] J. Bartlett and M. J. Plank, “Epidemic dynamics on random and scale-free networks,” *The ANZIAM Journal*, vol. 54, no. 1-2, pp. 3–22, 2012.
- [150] G. F. De Arruda, A. L. Barbieri, P. M. Rodríguez, F. A. Rodrigues, Y. Moreno, and L. da Fontoura Costa, “Role of centrality for the identification of influential spreaders in complex networks,” *Physical Review E*, vol. 90, no. 3, p. 032812, 2014.
- [151] “Ebola situation reports,” 2016, <http://apps.who.int/ebola/ebola-situation-reports>.
- [152] “GraphX of Spark,” Spark 1.6.2, <http://spark.apache.org/docs/latest/graphx-programming-guide.html#overview>.
- [153] I. J. Myung, “Tutorial on maximum likelihood estimation,” *Journal of Mathematical Psychology*, vol. 47, no. 1, pp. 90–100, 2003.
- [154] S. Bansal, B. T. Grenfell, and L. A. Meyers, “When individual behaviour matters: homogeneous and network models in epidemiology,” *Journal of the Royal Society Interface*, vol. 4, no. 16, pp. 879–891, 2007.
- [155] T. A. T. Institute, “Annual urban mobility report.” [Online]. Available: <http://mobility.tamu.edu/>

- [156] J. I. Levy, J. J. Buonocore, and K. Von Stackelberg, "Evaluation of the public health impacts of traffic congestion: a health risk assessment," *Environmental Health*, vol. 9, no. 65, pp. 1–12, 2010.
- [157] S. Chopra and P. Bauer, "Driving range extension of ev with on-road contactless power transfer—case study," *Industrial Electronics, IEEE Transactions on*, vol. 60, no. 1, pp. 329–338, 2013.
- [158] W. Su, H. Eichl, W. Zeng, and M.-Y. Chow, "A survey on the electrification of transportation in a smart grid environment," *Industrial Informatics, IEEE Transactions on*, vol. 8, no. 1, pp. 1–10, 2012.
- [159] W. He, G. Yan, and L. D. Xu, "Developing vehicular data cloud services in the iot environment," *Industrial Informatics, IEEE Transactions on*, vol. 10, no. 2, pp. 1587–1595, 2014.
- [160] J. Zheng and L. M. Ni, "Modeling heterogeneous routing decisions in trajectories for driving experience learning," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2014, pp. 951–961.
- [161] R. Wang, L. Zhang, R. Sun, J. Gong, and L. Cui, "Easitia: A pervasive traffic information acquisition system based on wireless sensor networks," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 12, no. 2, pp. 615–621, 2011.
- [162] P. X. Gao, A. R. Curtis, B. Wong, and S. Keshav, "It's not easy being green," *ACM SIGCOMM Computer Communication Review*, vol. 42, no. 4, pp. 211–222, 2012.
- [163] W. W. Lin, J. H. K. Wong, and A. K. Wong, "A novel real-time traffic sensing model to improve the performance of web-based industrial ecosystems," *Industrial Electronics, IEEE Transactions on*, vol. 58, no. 6, pp. 2147–2153, 2011.
- [164] V. Vyatkin, "Iec 61499 as enabler of distributed and intelligent automation: State-of-the-art review," *Industrial Informatics, IEEE Transactions on*, vol. 7, no. 4, pp. 768–781, 2011.
- [165] A. Janeczek, K. A. Hummel, D. Valerio, F. Ricciato, and H. Hlavacs, "Cellular data meet vehicular traffic theory: location area updates and cell transitions for travel time estimation," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 2012, pp. 361–370.
- [166] D. Schrank, B. Eisele, and T. Lomax, "Tti's 2012 urban mobility report," *Proceedings of the 2012 annual urban mobility report. Texas A&M Transportation Institute, Texas, USA*, 2012.
- [167] M. Hoy, A. S. Matveev, and A. V. Savkin, "Collision free cooperative navigation of multiple wheeled robots in unknown cluttered environments," *Robotics and Autonomous Systems*, vol. 60, no. 10, pp. 1253–1266, 2012.
- [168] M. A. Kafi, Y. Challal, D. Djenouri, M. Doudou, A. Bouabdallah, and N. Badache, "A study of wireless sensor networks for urban traffic monitoring: applications and architectures," *Procedia Computer Science*, vol. 19, pp. 617–626, 2013.
- [169] E. Ericsson, H. Larsson, and K. Brundell-Freij, "Optimizing route choice for lowest fuel consumption—potential effects of a new driver support tool," *Transportation Research Part C: Emerging Technologies*, vol. 14, no. 6, pp. 369–383, 2006.
- [170] M. Barth, K. Boriboonsomsin, and A. Vu, "Environmentally-friendly navigation," in *Intelligent Transportation Systems Conference (ITSC), 2007*. IEEE, 2007, pp. 684–689.
- [171] C. Guo, B. Yang, O. Andersen, C. S. Jensen, and K. Torp, "Ecosky: reducing vehicular environmental impact through eco-routing," in *Data Engineering (ICDE), 2015 IEEE 31st International Conference on*. IEEE, 2015, pp. 1412–1415.
- [172] M. Barth, F. An, T. Younglove, C. Levine, G. Scora, M. Ross, and T. Wenzel, *Development of a comprehensive modal emissions model*. National Cooperative Highway Research Program, Transportation Research Board of the National Academies, 2000.
- [173] H. Rakha, K. Ahn, and A. Trani, "Development of vt-micro model for estimating hot stabilized light duty vehicle and truck emissions," *Transportation Research Part D: Transport and Environment*, vol. 9, no. 1, pp. 49–74, 2004.
- [174] R. Akçelik, R. Smit, and M. Besley, "Recalibration of a vehicle power model for fuel and emission estimation and its effect on assessment of alternative intersection treatments," in *Proceedings of the TRB 4th International Conference on Roundabouts, Seattle, DC, USA*, 2014, pp. 1–16.
- [175] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische mathematik*, vol. 1, no. 1, pp. 269–271, 1959.
- [176] K. Boriboonsomsin, M. J. Barth, W. Zhu, and A. Vu, "Eco-routing navigation system based on multisource historical and real-time traffic information," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 13, no. 4, pp. 1694–1704, 2012.