UNIVERSITY OF TECHNOLOGY SYDNEY
Faculty of Engineering and Information Technology

Data Classification and Transportation in Rail Networks

by

Mahdi Saki

Principle Supervisor

A/Prof. Mehran Abolhasan

Co-Supervisor

A/Prof. Justin Lipman

THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

Sydney, Australia

2020

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Mahdi Saki, declare that this thesis is submitted in fulfilment of the requirements for the award of PhD, in the School of Electrical and Data Engineering/Faculty of Engineering and IT at the University of Technology Sydney. This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature: Signature removed prior to publication.

Date: 18/01/2021

Abstract

IoT is a revolutionary technology in the digital world, with a diverse range of services being created and deployed. One of the major challenges involved in efficiently implementing IoT is the management and transportation of large volumes of data that this solution generates. Modern approaches for IoT completely rely on cellular networks. As the demand for such networks is massively growing, in this thesis, we explore other communication methods as alternatives for management and delivery of IoT data in rail networks. Particularly, the focus will be on developing strategies that utilize existing trains and the rail network as a mode of data transportation. Furthermore, the thesis will combine physical delivery of IoT data by trains to strategic collection points in rail networks with cellular infrastructure to minimize costs and increase communication scalability and efficiency. Therefore, in this thesis, we introduce a new framework into future data-driven rail networks. For this purpose, we propose an edge processing unit that includes two main parts. The first part is a data classification model that classifies IoT data into maintenance-critical data (MCD) and maintenance-non-critical data (MnCD). The second part is a data transmission unit that based on the class of data, employs appropriate communication methods to transmit data to strategic collection points. The MCD is immediately forwarded through real-time communication methods such as cellular networks. However, for the transmission of MnCD, we propose three travel pattern methods including train-to-station (T2S), train-to-train (T2T) and train-to-wayside (T2W) communications that employ trains as data carriers. We validate the

classification model and all the transmission methods through extensive experiments. The simulation results show the effectiveness of our models as follows. The data classification model was validated under different operating conditions with over 98% accuracy. For the T2S model, we showed that over 5 GB data can be offloaded through T2S communications. Additionally, our proposed mobility model for T2T communications was tested with real GPS data and showed over 98% accuracy. Furthermore, for the T2W communications, we showed that the proposed AP placement approach could improve the efficiency of data offloading up to 165%. Finally, we proved that we can offload over 250 Gigabits through T2W communications over WiFi networks.

To my wife, Atefeh,

and my sons, Kian and Ryan

Acknowledgment

This thesis would not have been fulfilled without the supervision and guidance of many individuals who contributed and extended their valuable assistance in the preparation and completion of this study.

At first, I offer my sincerest gratitude to my principal supervisor, A./Prof. Mehran Abolhasan, who has supported me throughout my thesis with his patience and knowledge. I attribute the level of my Ph.D. degree to his encouragement and effort, and without his support, this thesis would not have been completed.

I am grateful to my co-supervisor, A./Prof. Justin Lipman, for his unconditional support and valuable guidance during this thesis. It is an honour for me to work with him during my studies. He has not been only a great advisor, but also an encouraging and motivating friend.

My sincere thanks goes to Prof. Abbas Jamalipour, from the University of Sydney, for his contribution in two of my papers that both were published in the prestigus journals of IEEE Transactions. I really learned a lot from his professional comments and wish to use his constructive advice in the other works in the future.

I greatly appreciate the financial support from the Rail Manufacturing Cooperative Research Centre (funded jointly by participating rail organisations and the Australian

Federal Government's Business Cooperative Research Centre Program) through Project R3.7.1 – Data Classification and Transportation in Rail Networks.

I would like to express my main thanks to my wife, Mrs. Atefeh Pourmohammadi, who really supported me with her patience, efforts and encouragements. I really belive that she did the main and hard job by caring and growing our two sons, Kian and Ryan. Without her support, it was really impossible for me to progress and complete my PhD.

Last but not least, I would like to thank all my friends specially, Dr. Farzad Tofigh, who was one of my first motivation to start my PhD; Dr. Ali Baraytee, a great friend, who learned me the alphabets of machine learning and data science and Dr. Majid Azadi, for his great friendship and all the nice times we spent during my PhD.

Contents

List of Figures

List of Tables

List of Algorithms

Technical Terms and Acronyms

2-D two-dimensional

3-D three-dimensional

A-GPS Assisted GPS

AARF Adaptive Auto Rate Fallback

AP access point

AP access point

BDA Big Data Analytics

CBM condition-based monitoring

DFT Discrete Fourier Transform

ECA environment class arrangement

ECR environment class ratio

EDP equally distributed placement

EOP energy optimization problem

FE feature extraction

FSPL free space path-loss

FT Fourier Transform

HP hybrid placement

iDMM IoT data management module

IoT internet of things

ISM Industrial, Scientific, and Medical

ITS Intelligent transportation system

LoS line of sight

LTE-R Long Term Evolution-Railway

MBP Measurement-Based Placement

MCD maintenance-critical data

MCS modulation and coding scheme

ML machine learning

MnCD maintenance-non-critical data

OP optimal placement

OSU on-board storage unit

PC principle component

PCA Principle Component Analysis

PL path-loss

PSD Power Spectral Density

RBAR Receiver Based Auto Rate

RCM railway condition monitoring

RMSE Root Mean Squared Error

RSS received signal strength

RTS Rail transportation systems

SNR signal-to-noise ratio

SVM Support Vector Machine

T2S train-to-station

T2T train-to-train

T2W train-to-wayside

T2W train-to-wayside

TCM train contact model

TMM train mobility model

UWB ultra wide-band

WLAN wireless local area network

Nomenclature and Notation

Chapter 3

Symbol	Description
x	data sample
N	number of samples
X	matrix of data
i	bearing number
j	segment number
c	channel number
n	number of channels
m	number of sensors
F	matrix of features
$\hat{F}$	matrix of effective features
l	number of effective features
w	window length
d_{tot}	total delay
d_{col}	duration of data collection
d_{alg}	algorithm processing time

Chapter 4

Symbol	Description
t_r	WiFi resilience time
t_{en}	entering time
t_{dw}	dwelling time
t_{lv}	leaving time
stp	stopping station
ps	passing station
d	displacement
a	acceleration
v	velocity
P_{ref}	received power at reference distance
P_t	transmitter power
λ	wavelength
c	light speed
f	radio carrier frequency
n^{dBm}	noise in dBm
γ	path-loss exponent
bw	bandwidth
N_{ss}	number of spatial streams
A	offloading capacity
th	throughput
N_{stp}	number of stopping stations
N_{ps}	number of passing stations

Chapter 5

Symbol	Description
v	velocity
d	distance
a_1	traction acceleration
a_4	breaking deceleration
a_3	coasting deceleration
a_f	acceleration due to friction
T	total trip time
D	total trip distance
$\overline{x}$	position vector
$\overline{v}$	velocity vector
$\overline{X_c}$	vector of contact positions
$\overline{T_c}$	vector of contact duration
N_c	number of contacts

Chapter 6

Symbol	Description
p_{Tx}	transmitter power
p_{Rx}	receiver power
PL	path-loss
i	index of access points
j	index of train position
PL_0	path-loss at reference distance
G_a	antenna gain
$C_{\pm}$	clearance between round lines
C_{AP}	clearance between APs and line
l	length of track line
α	ratio of environment class
ECR	environment class ratio
ECA	environment class arrangement
eff	energy efficiency
D	data capacity
E	consumed energy
ρ	MAC efficiency factor
P_N	noise power
PL_{av}	average path-loss