

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 prestigious 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 believe 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

Certificate of Authorship/Originality	i
Abstract	ii
Dedication	iv
Acknowledgement	v
List of Figures	xi
List of Tables	xiv
List of Algorithms	xvi
Technical Terms and Acronyms	xvii
Nomenclature and Notation	xxi
1 Introduction	1
1.1 Background	1
1.2 Research Questions	3
1.3 Research Objectives and Contributions	3
1.4 Thesis Structure	4
1.5 List of Publications	4

2	Literature Review	6
2.1	Data Classification and Transportation in Rail Networks	6
2.2	Train-to-Station Communication Method	8
2.3	Train-to-Train Data Communication Method	10
2.4	Train-to-Wayside Communication Method	12
2.5	Conclusion	15
3	A Comprehensive Scheme for Data Classification and Transportation in Rail Networks	17
3.1	INTRODUCTION	17
3.2	PROPOSED ARCHITECTURE	19
3.3	DATA CLASSIFICATION	21
3.3.1	Feature Extraction	22
3.3.2	Classification Algorithm	26
3.3.2.1	SVM Theory	26
3.3.2.2	Algorithm Explanation	27
3.4	DATA TRANSPORTATION	31
3.4.1	STORAGE AND OFFLOADING	31
3.4.2	REAL-TIME TRANSMISSION	33
3.5	EXPERIMENTAL VERIFICATION	37
3.6	CONCLUSION	43
4	Train-to-Station Communication Method	44
4.1	Introduction	44
4.2	The Proposed Offloading Scheme	46
4.3	The Analytical Offloading Model	48
4.3.1	Offloading Model for Stopping Stations	49
4.3.2	Model for Passing Stations	52

Contents

4.3.3	Total Model for Offloading in a Rail Network	53
4.4	Simulation	53
4.5	Conclusion	57
5	Train-to-Train Data Communication Method	58
5.1	Introduction	58
5.2	system models	61
5.2.1	Train mobility Model (TMM)	62
5.2.2	T2T contact model (TCM)	69
5.2.3	T2T communication approach	70
5.3	Simulation and Results	72
5.3.1	Mobility Model Simulations	73
5.3.1.1	Discussion about timetable changes	77
5.3.2	T2T Contact Simulation	80
5.3.3	T2T offloading simulation	84
5.4	Conclusion and Future Work	88
6	Train-to-Wayside Communication Method	90
6.1	Introduction	90
6.2	System Models and Problem Formulation	92
6.2.1	Problem Formulation	93
6.2.2	Initial Placement Approaches	98
6.2.3	Equally Distributed Placement (EDP) algorithm	99
6.2.4	Optimal Placement (OP) Algorithm	101
6.2.5	Hybrid Placement (HP) Algorithm	102
6.2.6	Railway Environment Model (REM) for Wireless Communications	103
6.3	Simulation and Results	105
6.3.1	REM Selection Strategy in Simulations	106

Contents

6.3.2	Efficiency	108
6.3.3	EDP Algorithm	111
6.3.4	OP Algorithm	114
6.3.5	HP Algorithm	116
6.3.6	The Effect of Different Scenarios on AP Placement	118
6.3.7	Comparison with Measurement-Based Method	122
6.3.8	Discussion about Doppler Effect	123
6.4	Conclusion and Future Works	125
7	Conclusion and Future Works	127
7.1	Data Classification	128
7.2	T2S Communication	129
7.3	T2T Communication	130
7.4	T2W Communication	131
7.5	Future Work	131
	Bibliography	133
	Appendix	149

List of Figures

1.1	Different kinds of communications used in the thesis	3
1.2	Thesis technical overview	5
3.1	Proposed architecture	20
3.2	Functional modes of data management for trains	21
3.3	0.1 second snapshot of a raw vibration signal in time-domain	23
3.4	PSD vs. FFT	25
3.5	Number of features for approximation of the whole data set	25
3.6	Classification algorithm flowchart	29
3.7	A single wagon with four bearings for our experiments	35
3.8	Run-to-failure data	35
3.9	Data transmission costs with and without classification algorithm	36
3.10	Results of PCA for each dataset	38
3.11	Scatter diagram of the obtained principle components for DS1	39
3.12	Classifier confusion matrix	42
4.1	Overall diagram of the proposed station-based offloading scenario	46
4.2	Timing diagrams of the offloading model for a) stopping stations, b) passing stations	49
4.3	SNR vs. time for different values of a) PLE and b) transmitter power	54

List of Figures

4.4	Throughput vs. time for different values of a) PLE and b) transmitter power	55
4.5	The theoretical lower and upper bounds of offloaded data estimated by the proposed model for different environments at $P_t = 30mw$: a) for stopping stations, b) for passing stations	56
5.1	System overview	62
5.2	Train mobility model	63
5.3	Train traffic model	64
5.4	Optimal train guidance trajectory [1]	66
5.5	T2T communication approach	71
5.6	A section of Sydney Trains network as our case study [2]	73
5.7	Simulation results and real data for line T6, between Clyde and Carlingford	75
5.8	Simulation results and real data for line T1, between Penrith and Parramatta	78
5.9	Distance errors caused by unrecognized changes in timetables for lines T1 and T6	80
5.10	Distance vs. time during contacts for two trains in line T6	81
5.11	Results of TCM in line T7 for two-trains scenario	82
5.12	Results of TCM between trains in line T7 for all-trains scenario	83
5.13	The case study used for simulation of T2T offloading	85
5.14	T2T Offloading diagrams: a) throughput, b) end-to-end delay	89
6.1	Three different short-range communication methods in rail networks	91
6.2	Problem overview	93
6.3	EDP algorithm	101
6.4	OP algorithm	102
6.5	HP algorithm	103

List of Figures

6.6	The proposed REM for a given rail path	105
6.7	The proposed REM for evaluation of the proposed placement algorithms	108
6.8	APs required number for every scenario obtained by EDP algorithm . . .	113
6.9	Statistics of mean PL for a) scenarios 1, and b) scenario 9.	115
6.10	Efficiencies obtained by EDP and OP and the related improvement for $n_{AP} = 6$ and communication mode 1, for scenarios 1 and 9.	115
6.11	Efficiencies obtained by HP algorithm compared to EDP algorithm and the related improvements for scenarios 1 and 9	119
6.12	Low and high bounds of DEC for scenario 9	119
6.13	AP Placement by EDP and HP algorithms for all the scenarios assuming a fixed PL threshold	120
6.14	Results obtained by OP and EDP algorithms for a fixed number of APs:a) AP placement for every scenario, b) average PL at every scenario	122

List of Tables

3.1	Specifications of selected data sets	34
3.2	The results obtained from our data classifier algorithm. The numbers in red color and bold format are related to the faulty bearings.	36
3.3	System specifications for simulation process	43
5.1	Comparison between accuracy of our proposed method and Lomnosoff-based method	74
5.2	Input data for simulation at line T6	76
5.3	Input data for simulation at line T1: between Penrith-Parramatta	77
5.4	The maximum errors obtained due to the possible change of the accelerations	77
5.5	Settings used for simulation of the T2T offloading network	85
5.6	The amount of FSPL with and without DS	87
6.1	PLE and STD for different environments in rail networks	104
6.2	Simulation settings	107
6.3	The scenarios proposed for evaluation of the placement algorithms	109
6.4	Maximum PL obtained by EDP algorithm at every scenario	114
6.5	Improvement of efficiency obtained OP over EDP for $n_{AP} = 3 - 6$ and two different communication modes at every scenario	116
6.6	Results obtained by HP algorithm for every scenario	118

List of Tables

6.7	Comparison between our proposed algorithms with MBP for scenarios 1 and 9 and $n_{AP} = 4$	124
6.8	Impact of Doppler Effect on PL, for different frequencies and various speeds	124

List of Algorithms

5.1	MOS determiner	65
5.2	TCM algorithm	70

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

Technical Terms and Acronyms

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

Technical Terms and Acronyms

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

Technical Terms and Acronyms

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
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	braking deceleration
a_3	coasting deceleration
a_f	acceleration due to friction
T	total trip time
D	total trip distance
\bar{x}	position vector
\bar{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