Review Report for the Ph.D thesis
submitted to Czech Technical University in Prague
Faculty of Information Technology

Candidate: Ladislava Smítková Janů
Title: Improvement of the Routing in Opportunistic Networks by the Application of Unsupervised and Supervised Machine Learning Techniques
Supervisor: Dr. Ing. Kateřina Hyniová

Up-to-dateness of the dissertation.

The objective of the thesis is to investigate the use of machine learning techniques for enhancing the data communication in opportunistic networks. To this end, the candidate proposes four routing protocols for opportunistic networks where various levels of network knowledge and machine learning techniques are combined. Specifically, she focuses on the use of unsupervised and supervised machine learning techniques, also combined with active nodes behavior. Although opportunistic networks represent a quite consolidated networking paradigm, however they continue to generate interest and new research opportunities exist. The topic of the dissertation is current and relevant in such context, because it aims to explore the benefits that the integration and exploitation of heterogeneous learning techniques can bring to opportunistic routing protocols.

Formal structure and organization of the dissertation.

The dissertation thesis is composed of 172 pages and consists of an abstract, five chapters and a rich bibliography. The first chapter introduces the reader to the dissertation topic by explaining the motivations, depicting the problem under investigation, and providing a high-level overview of the proposed solutions for solving the problem. The second chapter first outlines the network context studied (i.e., mobile ad hoc networks and their evolutions), then - focusing on opportunistic networks - it classifies the routing protocols describing the most important ones, and finally it discusses the most significant mobility models used in opportunistic networks. Chapter 3 is the core of the research part of the dissertation where the four proposed routing solutions for opportunistic networks are presented after a section where similar context-aware approaches are described.
Chapter 4 focuses on the performance evaluation based on a simulation comparison with already existing solutions, and discusses the obtained results. Chapter 5 summarizes the main contributions and findings of the dissertation, outlining also some possible future research directions for further improving the proposed algorithms. Finally, the scientific activity of the candidate is appended at the end of the dissertation.

Completion of the dissertation objectives.

The objectives of the dissertation are the exploitation of heterogeneous machine learning techniques and the development of efficient routing protocols in order to enhance the data communication in opportunistic networks. They are depicted in the first chapter and supported by a description of the current state of the art in the field of opportunistic networks in Chapter 2 and 3. They are clearly identified and explained, and they are fulfilled by an appropriate research activity.

Assessment of the methods used in the dissertation.

The methods used in the dissertation, which are briefly outlined in Chapter 1, are carefully described in Chapter 3. Specifically, the work hypothesis, a classification of similar context-aware approaches, and the fundamentals of machine learning techniques and classifiers are described in this chapter. Then an in-depth characterization of the four proposed routing protocols is provided. Specifically, two proposed solutions combine the knowledge of geographical data with unsupervised machine learning techniques. The differences among them rely in the way the detected geographic region is represented and exploited in the routing decisions. The third solution exploits the Support Vector Machine classifier, a specific type of supervised machine learning techniques, to guide the forwarding of messages when two nodes encounter. The fourth solution uses the statistical knowledge of node mobility and the active behavior of nodes to improve message forwarding. Block diagrams as well as pseudo-codes of the proposed routing protocols enrich this chapter. The results of the research and its interpretation, and conclusions are reported in Chapters 4 and 5, respectively.

The research subject is relevant, and the way the candidate approaches the problem is interesting. The candidate represents the ideas with sufficient theoretical background, and the presented results are suitable and offer new research hints.

Evaluation of the results and contributions to the dissertation.

The performance evaluation is based on a simulation analysis and it is composed of five parts, each focusing on a different scenario. Each proposed routing protocol is compared and contrast against four well-known routing protocols for opportunistic networks, i.e., Epidemic, PROPHET, First Contact and Bubble Rap. Overall, the analysis presented provides a general picture of the way the proposed protocols behave as it covers a sufficient number of scenarios under different parameters settings as well as studying the behavior of different performance indexes. It has also been enriched with a sensitivity analysis of a set of influence parameters and the change of the network conditions (such as the network size, the traffic load), which allows to investigate also some scalability properties.

The scientific merits of the dissertation rely on the exploitation of various level of network knowledge combined with machine learning techniques to enhance data communication in opportunistic networks. The result is the proposal of four different routing protocols based on unsupervised learning techniques, supervised
learning techniques, and active node behaviors. While the first three proposed protocols are standard solutions, the idea of using active nodes that move on nodes’ density basis exploited by the last solution is original.

**Remarks, objectives, notes, and questions for the defense.**

The dissertation has significantly improved from its previous version. It is well structured and readable, and all the parts are presented now in a more rigorous form.

The following remarks occurred to me while reading:

- At pg 2 – Problem 1: it is not clear which type of “data” is used to improve the routing algorithm. It should refer to geographical information but there is no mention of it in the paragraph.
- At pg 4 (and later in the thesis): the concept of “motifs” is introduced without any explanation.
- At pg 30: the paragraph related to the attention made by various scientific disciplines to analyze communities is not linked with the rest of the section. A better collocation might be found.
- Maybe a brief description of the steps implemented by HRC in Section 3.5.5.3 would further increase the presentation of the protocol.
- The ANMA performance in experiment 5 is discussed only in the text and no graphical representations are shown. Figures related to such performance evaluation might be included for completeness.
- Since simulations involve random variables, multiple runs should be performed for each experiment, to show that the results are consistent and that values present a small standard deviation. This will provide more confidence in the results and conclusions. However, there is no mention of the number of runs performed neither the confidence intervals are shown in the figures. The term “avg” suggests the presence of multiple runs, however this should be clearly mentioned.
- A careful proofreading is required as there are many typos that need correction, not only in the text but also in the labels of some figures (specifically those related to experiment 3).

Moreover, I have the following questions for the defense:

- Which type of MAC protocol has been simulated?
- Is the transmission bandwidth during node contacts infinite or not? The sentence “cause network congestion” appears several times in the text, which suggest a finite bandwidth, but there is no mention of it.
- Linked with the previous question, in case of working hypothesis finite bandwidth, how has the packet loss been simulated?
- The explanations for the presence of high peaks experienced by some metrics are a bit weak. For example, related to the overhead in figure 4.3, why only does the 300 node configuration strongly affect the overhead performance, while the 400 node configuration no? Both configurations should lead to large communities. Am I missing some implementation details? Maybe is the pick due to a single or low number of simulation runs?
- What do you mean when stating that Epidemic and First Contact outperform the other methods with respect to the delivery delay because “these methods use simple routing rules”? This sentence seems to imply that more complex routing rules always lead to low delay performance, statement that is not true in general. Epidemic reaches the minimum delay as, under infinite bandwidth and unlimited node cache size hypotheses, it explores all the possible paths due its flood nature and it finds the optimal path in a quickest way. Perhaps the candidate wanted to provide a different explanation here.
- Why does Prophet show such a strange behavior in Figure 4.26?
- In ANMA, after having deviated from its original position, does the node go back to the planned route or continue from the last position? In its performance evaluation, which is the maximum number of deviations allowed in Experiment 5?
- Even though there is no a direct comparison among the four proposed routing protocols, but basing exclusively on the collected results, has the candidate an idea which is the most performant protocol among them?
The overall evaluation of the dissertation.

Summarizing, the dissertation addresses an interesting and relevant topic in the area of opportunistic networks. It proposes new solutions that approach the problem from different perspectives and use different methodologies. The way the problem has been critically evaluated is adequate and sufficient. The dissertation thesis results now well organized and balanced in all its parts.

The author of the dissertation proved her ability to conduct research and achieve results. In accordance with par 47, letter (4) of the Law Nr. 111/1998 (The Higher Education Act) I do recommend the thesis for the presentation and defense with the aim of receiving the Ph.D. degree.

Pisa, May 7, 2019

Dr. Eleonora Borgia, Ph.D.

Signed ---

Up-to-dateness of the dissertation

In general, the dissertation thesis is a nice piece of work in which the candidate has provided her ability to undertake scientific research, to clearly set and meet the goals, to synthesize obtained results in a complete work. Based on the current state of the art, we can see that ad-hoc networks, as well as delay tolerant network or opportunistic networks, are going to the forefront of scientific research. Dissertation thesis is dealing with the actual problems of opportunistic networks. Dissertation thesis is focused on the applicability of Hierarchical Routing with Clustering, SVM-based routing, routing scheme combining GMRF (Gaussian Random Fields) and ANMA (Active Node Movement Algorithm) in order to the enhancement of routing algorithms and the active node behavior. The theses are high current at this time.

Completion of the dissertation objectives

Based on the description of the state of the art in the area, the author clearly identified the solved problems and the works how to solve them. A description of the current situation points to the fact that the author of the dissertation is well acquainted with the given issue. In Chapter 1.1 and Chapter 1.2, the author summarized why we need to take attention to the opportunistic networks. The goals of the dissertation works (Page 4 – Chapter 1.4) I consider to be high current, original and beneficial to the research community.

Formal structure and organization of the dissertation

A dissertation thesis has 174 pages. The content of the dissertation is divided into 5 chapters. There are List of figures (62 figures), List of tables (3 tables), List of used abbreviations (15 abbreviations), Bibliography (231 publications), Reviewed publications of the author relevant to the thesis (38 publications), Remaining publications of the author relevant to the thesis (10 publications). An overview of motivation, problem statement and goals of the dissertation thesis are given in Chapter 1. Chapter 2 gives an overview of the current state in the fields of ad-hoc communication, routing of opportunistic networks and human mobility models.

Assessment of the methods used in the dissertation

The description of the proposed method Hierarchical Routing with Clustering 1 and 2, SVM-based routing, routing scheme combining GMRF (Gaussian Random Fields) and ANMA (Active Node Movement Algorithm) are described in Chapter 3. Chapter 4 is focused on the experimental verification of the proposed methods. In Chapter 5 there are described the
main contributions of the dissertation thesis and there are outlined the future steps of the research.

The above mentioned parts of the thesis are processed at a high professional and formal level. Selected methods and procedures that have emerged from the analysis of the current state of the matter I consider it appropriate. The results obtained are very interesting and motivating for the next research activities.

The evaluation of the results and contributions of the dissertation

The presented dissertation thesis summarize theoretical and practical knowledge in the area of the opportunistic networks with regard to enhancement of the routing techniques with hierarchical routing with clustering, SVM-based routing, routing scheme combining GMRF (Gaussian Random Fields) and ANMA (Active Node Movement Algorithm) in order to the enhancement of routing algorithms and the active node behavior. From my point of view, the original benefits of scientific research work are the following:

- A designed a new method of the combining utilization of geographical data and unsupervised machine learning (cluster analysis) called Hierarchical Routing with Clustering (HCR) based on detected geographic regions. The regions are represented by the geographic structure and graph representation.
- A designed new protocol based on the Support Vector Machines (SVM) that enables decisions about routing in an opportunistic network with regular node mobility patterns based on dividing the area of the OPN into cells and SVN classifier.
- A designed new routing method to make decisions about message forwarding and active node deviation from its planned route based on GMRF (Gaussian Random Fields) and ANMA (Active Node Movement Algorithm).
- Using geographic information about the network for cluster analysis in routing algorithm with the knowledge obtained from the contact graph.
- Implementation of the proposed protocol to simulator ONE and a thorough analysis of the results for 5 different scenarios.
- Simulation testbed and comparative study with other opportunistic protocols (First Contact, Epidemic routing, PRoPHET, BUBBLE-Rap).

Remarks, objection, notes for defense

In general, the dissertation thesis of Mrs. Ladislava Šmitková Janků is well readable, however, during my reading, I have made a few notes, or let's say critical comments and/or suggestions. In the following part of my review, I will list and stress out some remarks.

The Ph.D. thesis presents formal errors and stylistic inaccuracies that do not reduce the overall content level of work. In the work, the term routing is used to mark the forwarding process, and in the case of OPN, I think the better is term forward/forwarding. The performance analyses for 5 scenarios are described in Chapter 4. Some graphs are no easy readable (Figure 4.43 and 4.44), for these reasons I recommend use the table form of the value's representation. There is described transmission range, but there is missing information about wireless technologies used during a simulation in ONE. I do not have any other fundamental comments on the dissertation thesis. The assessed dissertation formally fulfills the general criteria that apply to this type of qualification work.
Questions for the defense

I have the following questions about the dissertation thesis:

- I think, that using the term routing in DTN/OPN, when the store-carry-forward model is used is not adequate, much better is term forwarding. Can you explain what is the main difference between routing and forwarding mechanisms?
- On page 127, there is written: „The routing metrics have been designed in order to select the most optimal nodes which have the highest probability to be a part of the paths of successfully delivered messages with respect to maximization of message delivery ratio and the minimization of message delivery delay.“ Explain the term optimal nodes in OPN and how we can find these nodes by the proposed algorithms.
- How can the number of the nodes affect the training process during clustering in the sense of the message delivery ratio and the minimization of message delivery delay?
- How can the number of detected geographic regions affect the routing mechanisms?
- In your work, the value $\delta$ is equal to 0.05, e.g. at least 5% of all node positions must be inside the geographical sector. Why only this value has been selected for the simulations? What will happen with the given algorithms. If the value $\delta$ is lower and higher than 0.05.
- How can the speed of moving nodes affect the message delivery ratio?
- How the social relations between nodes should affect the contact graph, social ties in k-clique communities from the probability of the packet delivery point of view?
- What do you think about the using of the opportunistic network with implemented algorithms from the GDPR policy point of view?

The overall evaluation of the dissertation

In general, I think the candidate successfully fulfilled what she set up. Despite several formal and stylistic inaccuracies within the text, the main aims of the thesis have been fulfilled; the outputs are summarized at the end of the thesis in the chapter: 5.2 Contributions of the Dissertation Thesis. Dissertation thesis of Mrs. Ladislava Smítková Janků brought much valuable and original information for the research in the field of opportunistic networks.

By conducting and completing these aims, Mrs. Ladislava Smítková Janků has demonstrated that she has become a professional possessing broad knowledge of the discipline as well as simulation toolkit allowing addressing and resolving complex scientific questions. The author of the dissertation proved the ability to conduct research and achieve scientific results. I do recommend the thesis for presentation and defense with the aim of receiving the degree

„philosophiae doctor (Ph. D.)“

Sincerely,

[Signature]

Košice, 11.04.2019

doc. Ing. Ján Papaj, PhD.
Prof. Floriano De Rango, Ph.D
DIMES Dept., University of Calabria
Via P. Bucci, 87036 Rende (CS)
e.mail: derango@dimes.unical.it

Review Report for the Ph. D thesis
Submitted to Czech Technical University in Prague
Faculty of Information Technology

Candidate: Ladislava Smitkova Janku

Title: Improvement of the Routing in Opportunistic Networks by the Application of Unsupervised and Supervised Machine Learning Techniques

Supervisor: Dr. Ing. Katerina Hyniova

Up-to-dateness of the dissertation

The main objective of the thesis is the application of artificial intelligence and supervised and unsupervised machine learning to improve the data dissemination and forwarding in opportunistic networks. Also if opportunistic networks is a topic well studied and consolidated in the past years, the application of machine learning can be interesting and novel in this research area. Moreover, the application of different learning techniques and a joint application of these techniques to the data forwarding case can be worth of investigation.

Formal structure and organization of the dissertation

The thesis is divided in 5 chapters. The first one explains the main motivations of the work under study and the general approach to improve and face some research issues.
related to the data forwarding in opportunistic networks. The second chapter presents a classification of the routing and data dissemination techniques applied to the Opportunistic Networks. Chapter 3 is the core of the thesis and there 4 routing techniques that are presented. Performance evaluations are presented in section 4 and finally the main contributions with possible future research on the research topic are presented in section 5.

**Completion of the dissertation objectives**

Author completed the established objective summarized in the first chapter and provided enough simulation results and explanation of the fulfilled objectives.

**Assessment of the methods used in the dissertation**

Author described properly different methods applied in her research specifying two data dissemination techniques based on geographical data combined with unsupervised machine learning, one techniques based on Support Vector Machine classifier and supervised machine learning and the last solution based on the statistical knowledge of node mobility. All these techniques have been described in chapter 3.

**Evaluation of the results and contributions to the dissertation**

The author presented many simulation results considering 5 main experiments: the first one considered a Geographically structured OPN with a hierarchical clustering and the same routing algorithm has been applied also to the other 2 experiments (Experiment 3 and Experiment 4). In all these experiments, different parameters affecting the performance have been considered such as Influence of a period of message generation, Influence of TTL, Influence of Buffer size, Influence of Geographical Structure of Network, Influence of Time dependent connectivity between two OPN networks. The performance evaluation metrics considered have the classical metrics applied in DTN and Opportunistic Networks such as Data Delivery Ratio, average latency and overhead on the basis of the duplicated/transmitted messages over the first generated messages at sources. In the second experiment (Experiment 2) the considered routing strategy has been SVM-based routing (svm) and the parameters and performance metrics are the same of the other experiments. The last experiment has been the simulation scenario 5 where the ANMA (Active Node Movement Algorithm) method has been evaluated in comparison with Prophet, Epidemic and First Contact (FC) method. Final results show as the proposed methods are more effective in the data delivery and dissemination in
comparison with the introduced overhead in comparison with classical data diffusion techniques such as Epidemic, First Contact (FC) and Prophet.

Remarks, objectives, notes and questions for the defense

The overall thesis work has been considered good in the evaluation. Author considered and analyzed many scenarios to test their proposal and some insights in the understanding of data diffusion applying geographical knowledge and supervised and unsupervised learning has been led out. The prefixed objectives listed in the first chapter have been achieved and the methodology applied in the explanation and validation of the proposal is satisfying. Just some small notes and questions are presented below to better clarify some points of the Ph.D dissertation:

1. It is not clear as the mobility models can affect on the performance evaluation. Author considered synthetic mobility model and it seems that not real mobility model has been considered in the thesis. Could you please explain if a potential realistic mobility model can change the results presented in the thesis?
2. No reference has been provided to the statistical validity of the results. How many simulations have been led out to enter in the confidence interval? Are the plotted values given by an average of values obtained by independent simulations? If not, which methodology has been applied?
3. In the last scenario, the author did not consider the Bubble Rap in the comparison with other data dissemination techniques. Why? It was not necessary?
4. The physical layer in the communication seems to be not considered. It is well known that ONE simulator does not provide valid physical layer implementation and there is abstraction for all issues related to the wireless channel, packet error rate etc. How this physical and mac layers missing can affect the performance of the network? How the proposed solutions could be affected? The overall trend of the compared algorithms can completely change?
5. Concerning the learning phase of the algorithm, which are the critical elements to consider in the training phase? How the number of nodes or their mobility can affect the training phase? How long time can get the training phase?

The overall evaluation of the dissertation

The overall evaluation of the thesis is good and satisfying. Author addressed the reviewer suggestions improving significantly the work through the addition of formal
description of the routing/forwarding strategies and supervised/unsupervised machine learning techniques. Different methodologies have been presented and this enriched the understanding and the state of the art in the data dissemination techniques applied to opportunistic networks.

The author of the dissertation proved her ability to conduct research and achieve results. I recommend the thesis for the final presentation and defense with the aim of receiving the Ph.D. degree.

Rende, 06 Sept. 2019.

Prof. Floriano De Rango, Ph.D