Survivability strategies for emerging wireless networks with data mining techniques: a case study with NetLogo and RapidMiner

Iván García-Magariño, Geraldine Gray, Raquel Lacuesta, *Senior Member, IEEE* and Jaime Lloret, *Senior Member, IEEE*

***Abstract*—Emerging wireless networks have brought Internet and communications to more users and areas. Some of the most relevant emerging wireless technologies are WiMAX, LTE-A, ad-hoc and mesh networks. An open challenge is to ensure the reliability and robustness of these networks when individual components fail. The survivability and performance of these networks can be especially relevant when emergencies arise in rural areas, for example supporting communications during a medical emergency. This can be done by anticipating failures and finding alternative solutions. This paper proposes using big data analytics techniques such as decision trees for detecting nodes that are likely to fail, and so avoid them when routing traffic. This can improve the survivability and performance of networks. The current approach is illustrated with an agent-based simulator of wireless networks developed with NetLogo and data mining processes designed with RapidMiner.**

***Index Terms*—agent-based-simulation, big data, multi-agent system, wireless network**

I. Introduction

Information and Communication Technology (ICT) enterprises usually avoid low population areas, since the installation cost would be too high per user. In this context, WiMAX (Worldwide Interoperability for Microwave Access) networks offer a more economical communications solution for these areas. In addition, other protocols such as LTE-A (LTE Advanced) facilitate rapid communications in mobile devices by increasing the number of antennas. In wireless ad-hoc networks and mesh networks, links can be changed dynamically, facilitating real-time rerouting.

To increase the reliability of these emerging technologies, self-organized networks can adapt routes based on an analysis of the current state of the network. This means that self-organized wireless networks can improve communication reliability even when some network nodes malfunction. In particular, [1] concluded that the transmission reliability increased with their approach, which used network topologies represented as binary trees. In addition, [2] analyzed several routing protocols for mobile ad-hoc network (MANET). The network topology was dynamically changed. They analyzed the bandwidth utilization, delay, media access delay, traffic transmitted and lost, sent and received data. Moreover, [3] proposed a secure protocol for establishing spontaneous ad-hoc networks bearing in mind resource, processing and energy limitations of ad hoc services. Furthermore, a secure spontaneous mobile ad-hoc cloud computing network approach [4] provided communication infrastructure with mobile devices with their limited storage and processing resources, guaranteeing secure communications by encryption methods.

A confirmation message can determine whether a communication has succeeded. If it is not confirmed, the message can be sent again with the same route or a different one [5]. However, resending messages increases the energy consumption (against the green computing principles [6]) and decreases the performance of the network (in terms of transmission time). In this context, the prediction and anticipation of failures can allow wireless base stations to better plan the route of communications. Obtaining a diagnosis of wireless sensor network faults is a challenging problem [7].

The main goal of the current work is to analyze possible strategies for predicting failures, and so improve network performance and survivability. The proposed approach is based on data mining techniques and their incorporation to a simulation model. We use agent-based simulation (ABS) for assessing survivability strategies, since autonomous and reactive agents can simulate autonomous and reactive network nodes. In the current proposal, RapidMiner [8] has facilitated an exploration of different big data analytics techniques to inform smarter routing strategies by anticipating failures in the network nodes. NetLogo has been selected as the ABS environment due to its support for simulating networks [9].

The remainder of this article is organized as follows. The next section introduces the background related work to show the context and relevance of the current approach. Section III presents a technique for rapidly prototyping and testing data mining techniques for augmenting survivability and performance of emerging wireless networks. Section IV presents the experimentation done in the current approach, in which we tested several data mining techniques and selected one of them to improve an agent-based wireless survivability strategy. Concluding remarks are in Section V including possible future research lines.

I. García-Magariño and R. Lacuesta are with the Department of Computer Science and Engineering of Systems of the Universtiy of Zaragoza, Teruel, 44003, and with the *Instituto de Investigación Sanitaria Aragón*, Zaragoza, Spain. e-mails: ivangmg@unizar.es, lacuesta@unizar.es

G. Gray is with the Department of Informatics of the Institute of Technology Blanchardstown, Dublin, Ireland. e-mail: geraldine.gray@itb.ie

J. Lloret is with the Instituto de Investigacion para la Gestión Integrada de Zonas Costeras, Universitat Politecnica de Valencia, Valencia, Spain. e-mail: jlloret@dcom.upv.es

Manuscript received April 19, 2005; revised August 26, 2015.

II. Related work

A. Emerging wireless networks

The emerging wireless networks WiMAX use radio waves (normally between 2.5 and 5.8 GHz) and can reach up to 70 km away. This technology is useful for reaching rural areas in which wired communications would be too expensive per user. For example, [10] compared the spectral efficiency of the two WiMAX types called WHT-WiMAX and FFT-WiMAX concluding that the former one obtained better results.

LTE-A is a mobile communication standard that extends Long Term Evolution (LTE). Its main improvement is to transmit information in parellel from several antennas in the transmitter to several ones in the receiver (e.g. a cell phone). In this way, LTE-A can speed up the wireless communications. For instance, [11] explored the similarity between LTE and LTE-A to determine whether network traffic prediction resulted in an improvement in network performance. They demonstrated the existence of self-similarity in real world data traffic in both networks, concluding that self-similarity was higher in LTE-A. Self-similarity could facilitate forecasting traffic workload.

Wireless ad hoc networks (WANETs) are networks that don't have a predefined structure like the one in wired communications with routers. Instead, each node dynamically routes the information based on the network connectivity. One of the most relevant challenges about WANETs is to minimize their energy consumption. For example, [12] proposed a hybrid algorithm based on particle swarm optimization and local search. Their approach reconstructed the broadcast network efficiently.

Mesh networks is a local network topology in which the nodes connect dynamically and cooperate among each other. These networks can self-organize considering changing aspects such as current workloads. One of the problems of mesh networks is the instability caused by the link quality fluctuations. In this context, [13] proposed a stability-based routing protocol based on an entropy function. They improved the network performance compared to other alternatives.

Therefore, survivability strategies can be useful for improving the reliability, performance and energy consumption of emerging wireless networks such as WiMAX, WANETS and Mesh networks.

Moreover, [9] indicated that traditional network simulators are mainly focused in low level simulation, and they explicitly proposed multi-agent system (MAS) programming as a solution for high-level prototyping and testing MANETs. In particular, they proposed NetLogo for this purpose, probably because of its in-built support for modeling networks among other reasons. In general, MASs can be an appropriate mechanism for prototyping high-level simulations, as for example when using Ingenias for simulation [14], which used model-driven development (MDD).

B. Big data analytics applied to networks

RapidMiner is a tool for performing predictive analytics and data mining [8]. It allows one to apply several algorithms of machine learning, including decision trees and deep learning among others. Several extensions incorporate additional functionality. For example, the Linked Open Data extension of RapidMiner allows users to extract some relevant information from web pages and their embedded hyperlinks [15]. This extension was successfully tested for extracting statistical data about the World Bank from scientific publications. The accuracy of RapidMiner is comparable with the other similar tools such as WEKA, Tanagra, Orange and Knime [16]. RapidMiner has been used in the context of a wireless mesh network [17]. In particular, this work analyzed the performance for detecting the presence of humans through WiFi sensors. However, none of these works applied big data analytics for improving the survivability and performance of wireless networks.

Wireless multimedia sensor networks have been analyzed from a big data viewpoint. In particular, [18] proposed a big data simulation model, in which big data was generated simulating the multimedia detected from cameras of multiple devices. They compared the graph-based NoSQL databases Neo4j and OrientDB, and the relational database MySQL. This study determined which database management systems may be more useful for managing multimedia data from sensors of Internet of Things (IoT). Nevertheless, this work was more focused on storing and managing big data rather than analyzing it.

Big data analytics have been applied to detect intruders from wireless sensor networks. More concretely, [19] proposed a big data analytic architecture to reconstruct images from a low number of camera sensors with low quality to analyze the behavior of the intruder. In this manner, their approach was able to store less amount of data with the same relevance for this purpose and process it more efficiently. Nonetheless, they did not focus on possible wireless network problems.

Several artificial intelligence techniques supported the development of intrusion detection systems (IDSs) that analyzed the transmitted big data. More specifically, [20] reviewed IDSs with big data and artificial intelligence techniques inspired in biological patterns, including genetic algorithms, artificial immune and artificial neural networks. In addition, [21] proposed an IDS that analyzed big data from the Internet of Things (IoT) for detecting malicious activity. They used a variational autoencoder for classifying the communication patterns, outperforming other similar unsupervised methods. Their approach also provided a novel algorithm for performing feature recovery. However, these works were focused on detecting malware patterns in the network activity rather than guaranteeing its survivability based on the prediction of connection loses.

It is worth mentioning that data-driven approaches are becoming integrated in ABSs. For instance, [22] proposed a technique for injecting data into ABSs. They focused on feeding the ABS with data in the simulation, and corroborating the similarity of the outputted data with the real data. This article was one of the works that motivate the use of data and its analysis in ABSs, either for simulating networks or other contexts.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2825954, IEEE Access

C. Agent-based simulation applied to networks

ABSs have been applied to analyze different aspects of networks. For instance, ABS-TrustSDN [23] is another ABS that simulates trust strategies for allowing the controller of a software-defined network (SDN) to detect the network nodes with malware or unreliable. More specifically, one of its strategies was based on the history of each network node. In this manner, the SDN controller could select the appropriate routes. These ABSs were designed for improving performance of networks, for improving either the time response of urgent services or the isolation protocols of network nodes with malware. However, this ABSs did not use big data analytics for improving the survivability of emergent wireless technologies.

In a broader context, ABSs have been used to model different kinds of networks. In particular, [24] proposed an ABS that simulated a network of friendship relations. They used a fuzzy-logic approach for modeling and simulating the evolution of these friendships. In addition, [25] used ABSs for microscopic road network simulation. More concretely, the focused of efficiently using graphics processing units (GPUs) for modeling over a half million vehicles in road networks with a high performance. Nevertheless, these works did not solve the survivability of the network for ensuring communications,

Furthermore, [26] analyzed the communications among agents through the network with MASs. In particular, this work proposed a set of metrics for guiding the design of communications in MASs, and they assessed their approach with an ABS about the management of crisis situations in cities. This work used a different approach to the current one, since it focused mostly in the measurement through metrics instead of applying data mining techniques. The next section introduces the current approach that addresses a gap in the literature regarding improving survivability strategies of emerging wireless networks by applying data mining techniques.

III. A TECHNIQUE FOR EXPLORING SURVIVABILITY STRATEGIES WITH DATA MINING TECHNIQUES

Figure 1 introduces the current approach for guaranteeing survivability of emerging wireless technologies. In particular, this approach is based on the prediction of network problems based on big data analytics of network information. In the current approach, the history of each network node is stored locally, considering problems such as software attacks, hardware component failures and vandalism acts. This information will be periodically shared with the neighbor nodes (e.g. once per day to not overload the network information). In this way, each network node can perform data mining in its information and information from neighboring nodes, to predict (a) which are the more reliable nodes, and (b) which are more likely to have problems that require maintenance or repair. In this way, the nodes can determine links that are more reliable for making the network survive with a high performance (avoiding resending data). This approach can also identify nodes requiring maintenance.

The current work proposes a technique with the following steps for improving the survivability and performance of the emerging wireless networks with data mining techniques:

- *Creation of a simulation model*: ABS-SurviveWireless provides a framework for simulating different strategies of survivability of wireless networks. These strategies should manage and export data for subsequent analysis.
- *Analysis of the results with data mining*: In this phase, we propose to use RapidMiner as it allows rapid prototyping and evaluation of different data mining techniques applied to the ABS data.
- *Improvement of survivability strategies*: The designer considers the results of the data mining analyses and improves some aspects of the survivability strategies.

ABS-SurviveWireless is the novel ABS that allows simulating survivability strategies in wireless networks. It has been developed as a simulation model for NetLogo. It allows testing different strategies regarding the establishment and removal of frequent links among nodes. One can define the features of the nodes. It allows designers to find the shortest path considering either the number of links or some weights of these.

This work has used an ABS for reproducing realistic network information, and has applied big data analytics techniques to assess the viability of the current approach.

A. ABS-SurviveWireless simulation model: an ABS for simulating survivability strategies in wireless networks

ABS-SurviveWireless was developed with NetLogo, as this directly supports the creation, the update and some basic operations over the network [27]. We developed this ABS following PEABS (a Process for developing Efficient ABSs) [28], so that this ABS was efficient in terms of simulation responses. From a high-level design viewpoint, the three main components were (a) initialized the model creating all the agents, (b) evolved the simulation by simulating communications in a wireless network and updating its links for reliable routing, and (c) managed the visualization and storing of the different kinds of outputted data.

In the NetLogo simulation engine, simulation models are usually defined separating (1) the setup methods commonly triggered by a "Setup" button" and (2) the evolution methods normally triggered by a "Go" button. In the setup of the proposed simulation model, the nodes are distributed in random locations in the simulated area. The simulator establishes the features of the network nodes considering the probabilities of certain internal parameters. This assignment of features is performed following the principle of non-deterministic decisions of TABSAOND (a technique for developing agent-based simulation apps and online tools with nondeterministic decisions) [29]. The features of each node are represented as three boolean values that determine (a) whether the node is outdoors, (b) whether it has vulnerable software, and (c) whether it has components that are more than five years old.

Initially, some links are created among the nodes considering the restriction of the maximum possible distance among nodes indicated by the user. TABSAOND was applied for selecting which links are created from the ones that satisfy the distance restriction. Another internal parameter determined the probability of these selections.

In wireless networks, it is not necessary to install a wire for raising a new link, and consequently the simulator creates new

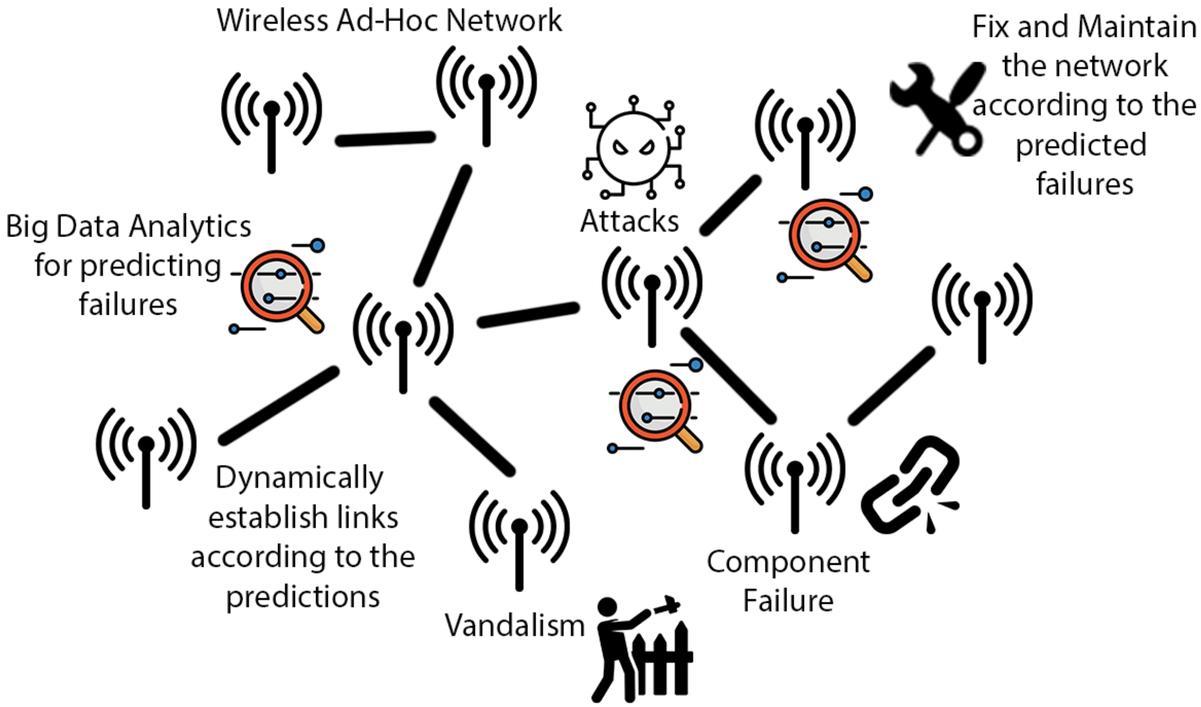

Wireless Ad-Hoc Network
Fix and Maintain the network according to the predicted failures
Attacks
Big Data Analytics for predicting failures
Dynamically establish links according to the predictions
Vandalism
Component Failure

Fig. 1. Overview of the current approach for guaranteeing survivability of emerging wireless technologies

```
; It simulates a communication. It increases the energy consumption
; as the number of links to reach the target
to communicate
  ; Get source and target
  let source random count turtles
  let target random count turtles
  ask turtle source[
    let dist nw:distance-to turtle target
    ifelse dist = false
      [set num-failures num-failures + 1]
      [  set energy-consumption energy-consumption + dist
         if success = -1
           [set success 0]
         set success success + 1
         foreach nw:turtles-on-path-to turtle target
           [transmit-msg]
      ]
    set communications communications + 1
  ]
end
```

Fig. 2. Programming code for simulating a communication with NetLogo language

```
; In a node (i.e. turtle), it simulates the transmission of a message
; throug the node
to transmit-msg
  let hasSuccess true
  if ((outdoor and (random-float 100 < chance-vandalism)) or
      (not outdoor and (random-float 100 < chance-default)))
    [set vandalism vandalism + 1
     set hasSuccess false]
  if ((vulnerable and (random-float 100 < chance-attacks)) or
      (not vulnerable and (random-float 100 < chance-default)))
    [set attacks attacks + 1
     set hasSuccess false]
  if ((old and (random-float 100 < chance-component-failures)) or
      (not old and (random-float 100 < chance-default)))
    [set component-failures component-failures + 1
     set hasSuccess false]
  set msgs msgs + 1
  if hasSuccess
    [set success success + 1]
end
```

Fig. 3. Programming code for simulation the the forwarding of a message through a network node

links dynamically. The proposed simulation model simulates the network failures in the simulation evolution.

In every iteration, the tool simulates a certain number of communications. Each communication departs from a source node, and aims at reaching a target node. Both nodes are selected randomly from the total set of nodes, guaranteeing that these nodes are different from each other.

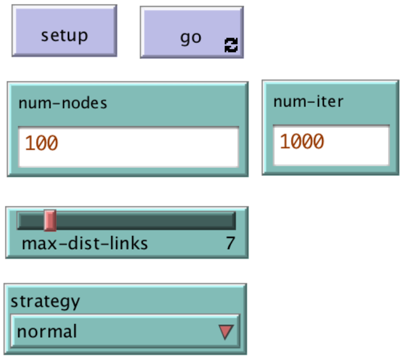
A simple strategy calculates the shortest path in terms of number of links. Then, the message transmission is simulated from the source node to target one, going through all the nodes on the path. Figure 2 shows the NetLogo code that simulates these actions.

Each node of the path simulates the possibilities of (1) having suffered vandalism, (2) having been damaged by a software attack, and (3) having a component failure. Each problem is simulated taking the features of the nodes and certain probabilities into account, following the TABSAOND approach. Figure 3 shows the corresponding NetLogo programming code. In particular: outdoor nodes can suffer from vandalism more easily; nodes with vulnerable software usually suffer from software attacks more frequently; and old nodes are more likely to have failures.

B. User interface of ABS-SurviveWireless

The user interface (UI) of ABS-SurviveWireless allows users (a) to enter some input parameters, (b) to see the evolution of simulation in a map of the network, and (c) to track evolution of some features in certain charts.

Users can enter the input parameters in the UI fields shown in Figure 4. For example, Users can establish the number



num-nodes 100
num-iter 1000
max-dist-links 7
strategy normal

Fig. 4. Input parameters of the user interface

of network nodes with a numeric input field. They can also establish the maximum distance between nodes for setting links with a slider bar. Although the simulator can run forever and be stopped when users decide so, users can set a number of iterations to determine when all the relevant simulation data will be stored in a CSV (comma separated values) file. The Setup button initializes the network. The Go button starts running the simulation, performing communications among pairs of random nodes and updating the links between nodes. A dropdown list allows the designer to choose the survivability strategy. For example, the simulator incorporates a simple strategy called "Normal". A smart strategy will be added when incorporating certain knowledge after testing some data mining techniques.

ABS-SurviveWireless shows the evolution of the network considering the different kinds of nodes and the links among these, as one can observe in Figure 5. The different kinds of nodes are presented with colors. Blue nodes represent outdoor nodes. Red ones represent nodes with vulnerable software (e.g. outdated). Gray ones represent nodes with old components (i.e. above 5 years old). The remaining nodes are represented with white.

As presented in Figure 6, the UI has a chart that shows the average results of nodes considering nodes that (a) have suffered from vandalism, (b) that have been compromised with attacks, and (c) have had failures in some of its components. In addition, a chart of the UI represents the global energy consumption. Another chart shows the failures of connectivity, the cases in which there was not any path among the two corresponding nodes.

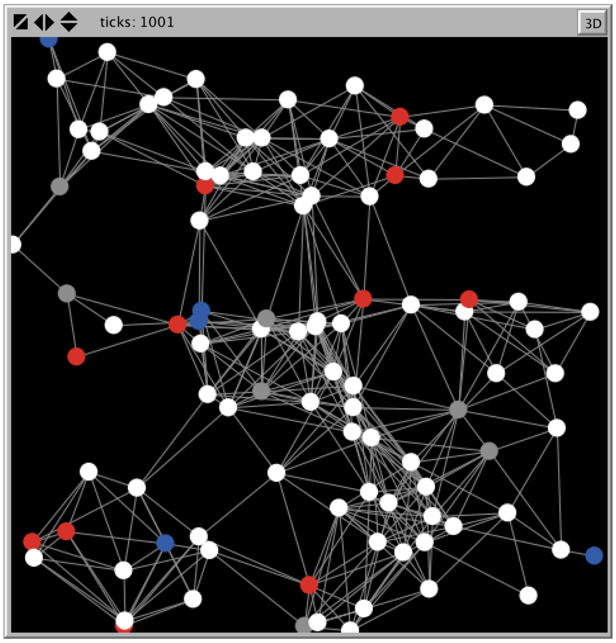


Fig. 5. Simulated network of the user interface

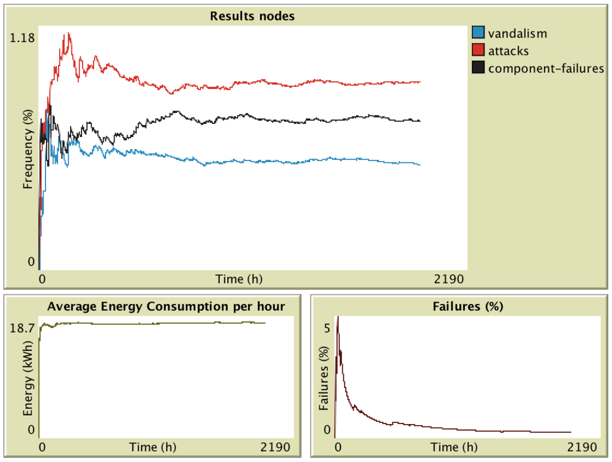


Fig. 6. Charts of the user interface

C. Data mining for survivability of wireless sensor networks

We propose data mining as an effective mechanism to inform decisions about selecting the best routes for communications.

The current approach proposes that network nodes collect data about previous problems such as vandalism, attacks and component failures. In particular, this approach mainly considers the percentage of times a message was sent to a node and this did not forward it because of each of these problems. This data can be analyzed to detect feature value combinations that are predictive of node failure. In this manner, the survivability strategy can determine reliable paths by avoiding nodes that are predicted to fail.

RapidMiner was used to test different big data analytics approaches. The data was initially modelled using a Decision Tree. Figure 7 shows the process used. The process receives input from the CSV file generated by ABS-SurviveWireless. "Set Role" defines the class label, the feature to be predicted. This was a boolean value indicating if the node successfully transmitted messages above a certain threshold range (e.g. we used a threshold of 99% in the current experimentation). Section IV will show an example of a learned decision tree, as part of the experimentation results.

A correlation matrix was used to identify correlated variables. This can be useful in determining replacement variables if some variables are missing. Similarly, it can be used to identifying redundant features, where two or more features

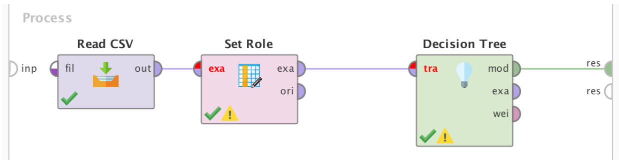


Fig. 7. Process for obtaining a decision tree of success of network nodes with RapidMiner

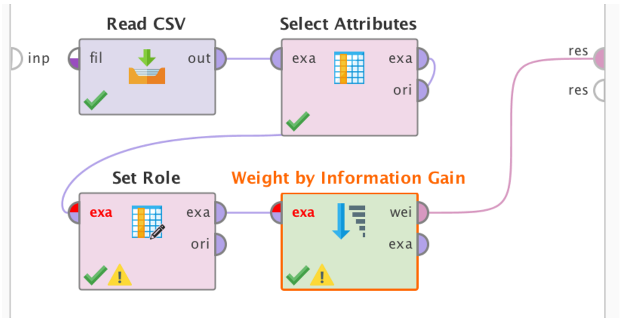


Fig. 8. Process for obtaining the weights by information gain of success of network nodes with RapidMiner

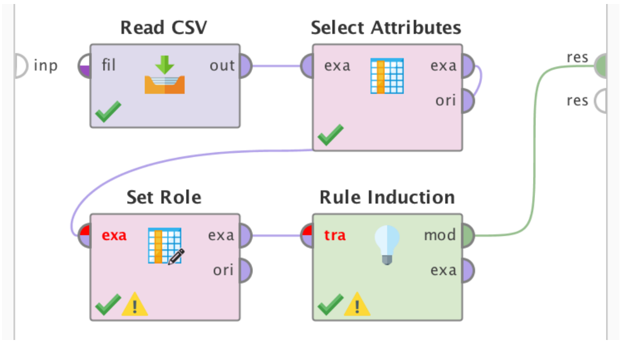


Fig. 9. Process for rule induction with RapidMiner

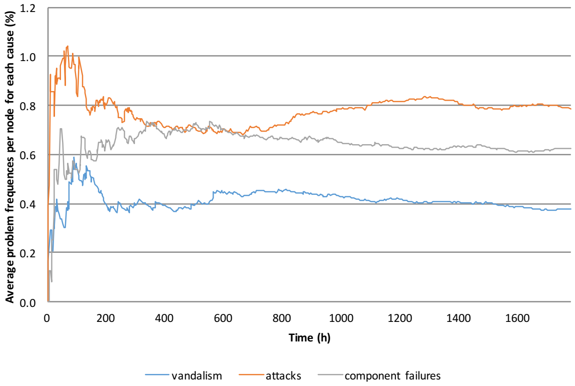


Fig. 10. Average frequencies of each problem kind per network node

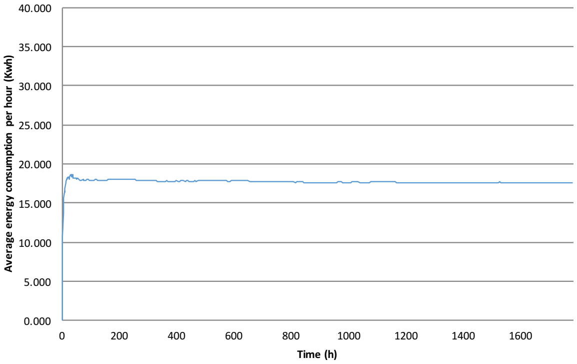


Fig. 11. Evolution of average energy consumption per hour

represent the same information.

Rapidminer implements a range of methods for feature weighting. This allows designers to rank network features by their relevance in predicting a class label. For example, “Weight by Information Gain” ranks features in the range [0,1] based on their information gain, a measure of the decrease in entropy if the dataset was split by feature value. The process is illustrated in Figure 8. Two alternatives were also considered: “Weight by Uncertainty” which ranks features based on symmetric uncertainty; and “Weight by Relief” which gives a higher rank to features that have the same value amongst nearest neighbours from the same class but a different value amongst nearest neighbours that are in different classes.

A rule induction classifier was trained on the features as illustrated in Figure 9. This process learns a set of rules determined from available feature values that maximize the correct classification of the nodes, in this case between reliable and error-prone based on the comparison of success rate with a threshold.

Results from these data analytics processes can inform changes to the survivability strategy implemented in the ABS model. Smart routing strategies can use patterns predictive of node failure to identify problematic nodes.

IV. Experimentation

We executed several simulations of 1780 iterations in the ABS-Survive Wireless. Table I indicates the values of both UI input parameters and internal ones used in this experimentation.

Figure 10 presents a chart of the simulation evolution regarding the frequencies of each of the analyzed kinds of network node problems. This chart presents the average of failed transmissions due to each problem per node. In the beginning of the simulation, the average varies frequently as the sample is low. The more iterations the simulation has gone through, the more stable the average frequencies are, since the samples are higher. At the end of the simulation, the most severe problem was attacks with a frequency of 0.78%. The component failures had a frequency of 0.62%. The least frequent problem was vandalism with a frequency of 0.38%.

In addition, Figure 11 shows the evolution of global average energy consumption per hour for the whole simulated network. The average was 17.7 kWh. It is worth noting the energy consumption per hour is quite stable (SD = 0.38 kWh) since this simulation considers a fixed number of communications per hour. The variations can be due to the distance from the sender and receiver (both of them selected randomly), and whether there is some connectivity failure.

TABLE I
UI INPUT PARAMETERS AND INTERNAL VALUES OF ABS-SURVIVEWIRELESS

Name	Description	Value
num-nodes	The number of network nodes	100
max-dist-links	Maximum distance between nodes	7 km
num-communications-per-iter	Number of communications per each iteration	5
chance-link-init	The percentage probability (chance) that two nearby nodes connect	30%
chance-link-die	The percentage change of losing a link by a software attack	0.8%
chance-link-create	The percentage change of creating a link in each iteration	1%
chance-outdoor	Percentage probability of a node being outdoor	10%
chance-vulnerable	Percentage probability that a node has vulnerable software	10%
chance-old	Percentage probability of a node being old	10%
chance-vandalism	Probability of suffering vandalism when a node is outdoor	3%
chance-attacks	Probability of suffering software attacks when a node has vulnerable software	10%
chance-component-failures	Probability that a node has a component failure when it has components with more than 5 years old	7%
chance-default	Chance any of the above problems when correspondingly not outdoor, not vulnerable or not old	0.1%
threshold-success	Threshold for converting the success percentage into a nominal value for data mining analyses.	99%

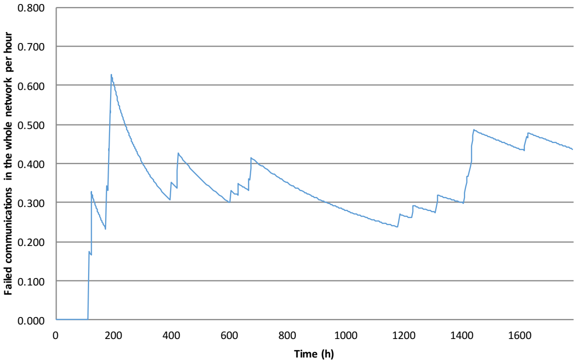


Fig. 12. Failed communications in the network per hour

Figure 12 shows the evolution of the failed communications per hour. These failures can be due to lack of connectivity between the sender and receiver, or due to failure of any of the nodes of the path. The average of the whole simulation was 0.336, and the SD was 0.118.

Figure 13 presents the decision tree model of the data. According to this analysis, the features recording component failures, vandalism, and attacks are more predictive of node failures than the features of the nodes themselves (i.e. outdoor, vulnerable software, and old components). However, the correlation matrix in Figure 14 shows strong correlations between features included in the tree, and features not included. Vandalism is correlated with network nodes being outdoor ($r = 0.742$); software attacks are correlated with vulnerable software ($r = 0.948$); and component failures are correlated to node components older than five years ($r = -0.928$). These correlations suggest models could use features of the network nodes themselves when history information is non-existent, for example in early simulation iterations.

Figures 15, 16 and 17 illustrate the results of Weight By Information Gain, Weight by Relief and Weight By Uncertainty respectively. There is some disagreement in the results, as they use different strategies to weight attributes. Weight By Information Gain favours the numeric attributes based on node history, namely vandalism, component failure and attacks. As expected, this concurs with the decision tree model. Weight by relief on the other hand favours the intrinsic, binary attributes of outdoor, old and vulnerable, indicating these attributes would be preferable if using a nearest neighbour model. As decisions made are local, they may be less useful in informing code changes to a global ABS model. Finally, weight by uncertainty gives similar weightings to correlated features. Old and component failure have the highest ranking, while vandalism and outdoor have the lowest ranking. Figure 18 shows these results in a bar chart.

Figure 19 presents the results of applying rule induction. As one can observe, it obtained a rule in which the result was correct in 90,094 out of 96,069 cases. Thus a single rule represents 93.8% successful predictions.

The output of rule induction was used to illustrate how results from data analytics can enhance the ABS model. This was chosen both because of its relatively good prediction accuracy, and the ease with which a single rule can be added to the programming code of the ABS model. In particular, this rule was included to create virtual links between nodes identified by the rule model as being reliable. Figure 20 shows the programming code in NetLogo that implemented this recommendation in ABS-SurviveWireless.

We analyzed the results of the ABS model enhanced with the corresponding data mining technique. From this point forward, this new strategy will be referred as the smart one. Figure 21 shows the average frequencies of the different kind of problems from a simulation with the smart strategy. In particular, the average percentage reduction of attacks was 17.7%. It also reduced the average components failures by

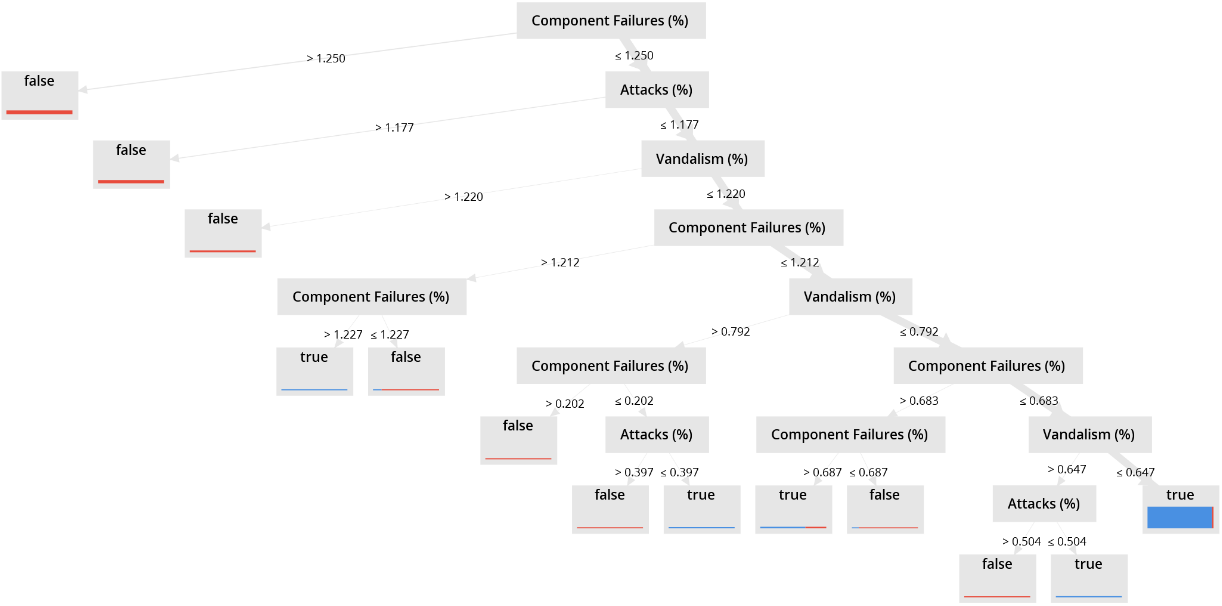
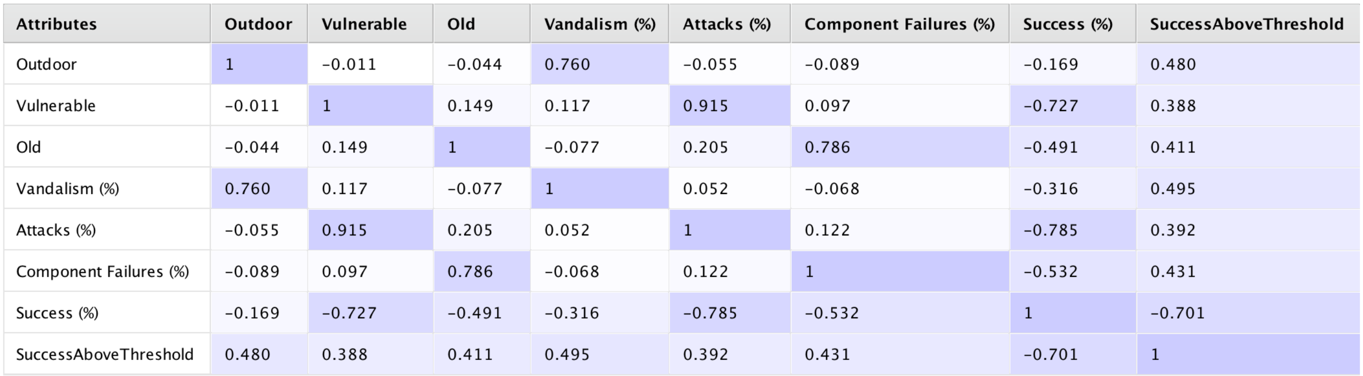
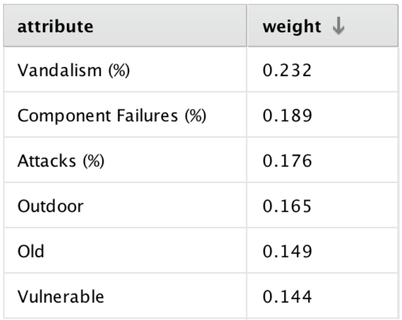


Fig. 13. Decision tree of success of network nodes considering their features



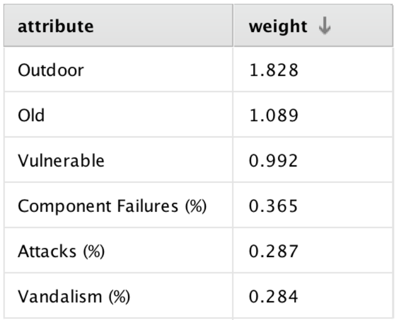
Attributes	Outdoor	Vulnerable	Old	Vandalism (%)	Attacks (%)	Component Failures (%)	Success (%)	SuccessAboveThreshold
Outdoor	1	−0.011	−0.044	0.760	−0.055	−0.089	−0.169	0.480
Vulnerable	−0.011	1	0.149	0.117	0.915	0.097	−0.727	0.388
Old	−0.044	0.149	1	−0.077	0.205	0.786	−0.491	0.411
Vandalism (%)	0.760	0.117	−0.077	1	0.052	−0.068	−0.316	0.495
Attacks (%)	−0.055	0.915	0.205	0.052	1	0.122	−0.785	0.392
Component Failures (%)	−0.089	0.097	0.786	−0.068	0.122	1	−0.532	0.431
Success (%)	−0.169	−0.727	−0.491	−0.316	−0.785	−0.532	1	−0.701
SuccessAboveThreshold	0.480	0.388	0.411	0.495	0.392	0.431	−0.701	1

Fig. 14. Correlation matrix among the different variables of wireless network nodes



attribute	weight ↓
Vandalism (%)	0.232
Component Failures (%)	0.189
Attacks (%)	0.176
Outdoor	0.165
Old	0.149
Vulnerable	0.144

Fig. 15. Results of the analysis of Weight By Information Gain



attribute	weight ↓
Outdoor	1.828
Old	1.089
Vulnerable	0.992
Component Failures (%)	0.365
Attacks (%)	0.287
Vandalism (%)	0.284

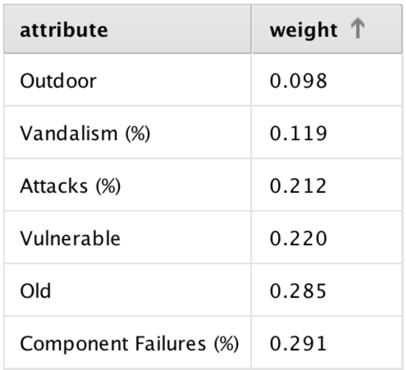
Fig. 16. Results of the analysis of Weight By Relief

9.1%.

In addition, Figure 22 shows the energy evolution in the ABS with the smart strategy. The average energy reduction was 3.30% in comparison to the previous strategy without data mining.

Figure 23 shows the evolution of failures with the smart strategy. The percentage reduction of communication failures was 51.6%. It is worth noting that the small reduction of problems in nodes had a great impact in the reduction of global communication failures since each communication depended on the failures of several nodes in each path.

Figure 24 shows an example of the network when using the



attribute	weight ↑
Outdoor	0.098
Vandalism (%)	0.119
Attacks (%)	0.212
Vulnerable	0.220
Old	0.285
Component Failures (%)	0.291

Fig. 17. Results of the analysis of Weight By Uncertainty

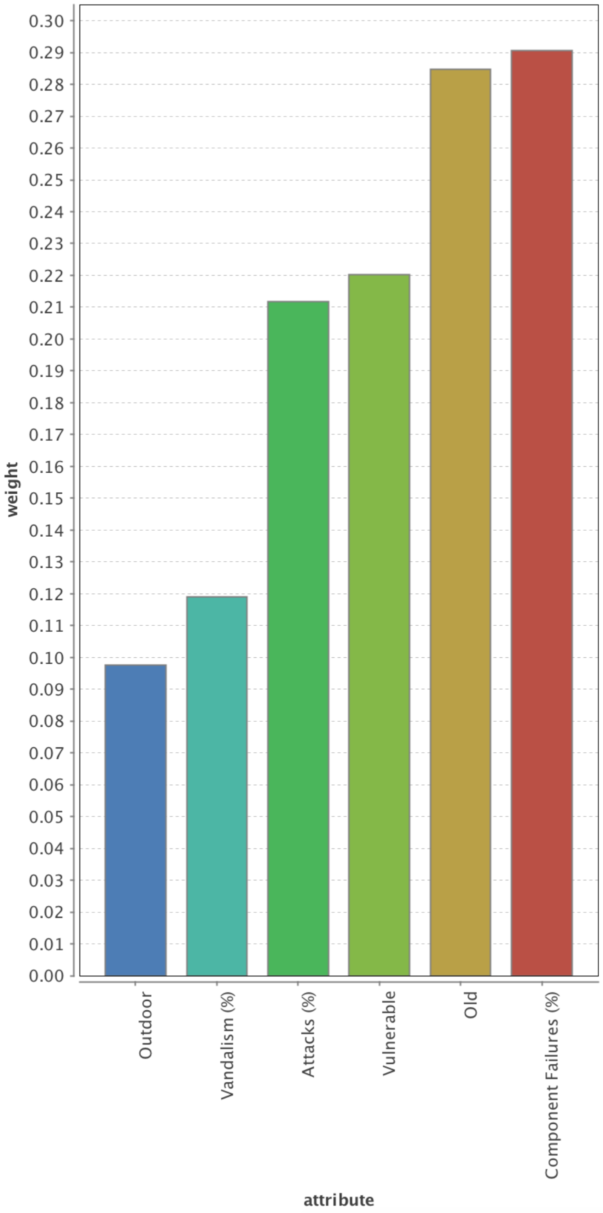
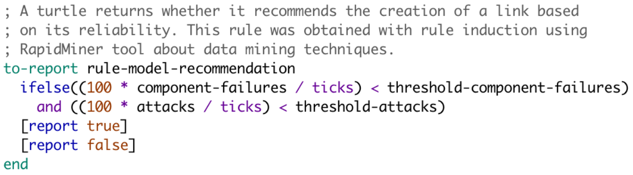


Fig. 18. Charts of the analysis of Weight By Uncertainty

Rule Model
if Component Failures (%) $\leq$ 0.597 **and** Attacks (%) $\leq$ 1.184
then true (70926 / 5632)
else false (343 / 19168)
correct: 90094 out of 96069 training examples.

Fig. 19. Results of Rule Induction analysis



```
; A turtle returns whether it recommends the creation of a link based
; on its reliability. This rule was obtained with rule induction using
; RapidMiner tool about data mining techniques.
to-report rule-model-recommendation
  ifelse((100 * component-failures / ticks) < threshold-component-failures)
    and ((100 * attacks / ticks) < threshold-attacks)
   [report true]
   [report false]
end
```

Fig. 20. NetLogo programming code for the recommendation about creating virtual links based on the result of rule induction

smart strategy. As one can observe, the software-vulnerable and hardware-old components had slightly less links in the network. The reason is probably that software attacks and component failures are respectively related with these intrinsic features (as indicated by the correlation matrix analysis), and these problem frequencies were considered in the incorporated rule model.

V. Conclusions and Future Work

The current work explores the possibility of applying data mining techniques for improving the performance and survivability of wireless ad-hoc networks. This can be useful for different purposes such as attending critical health situations in rural areas. To illustrate the current approach, this work presents a novel ABS of wireless networks called ABS-SurviveWireless. We have also used RapidMiner to apply different data mining techniques to explore their utility in the current approach. Some data mining results can be incorporated to ABS model. In particular, the experimentation case study incorporated a rule obtained by rule induction in a smart strategy, which improved the results of similar strategy but without using this inducted rule.

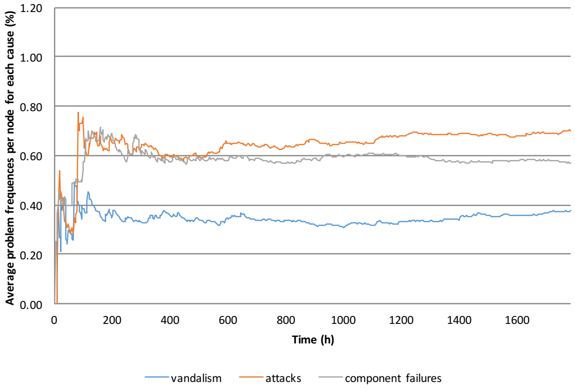


Fig. 21. Average frequency of problems per node for each cause in the smart strategy

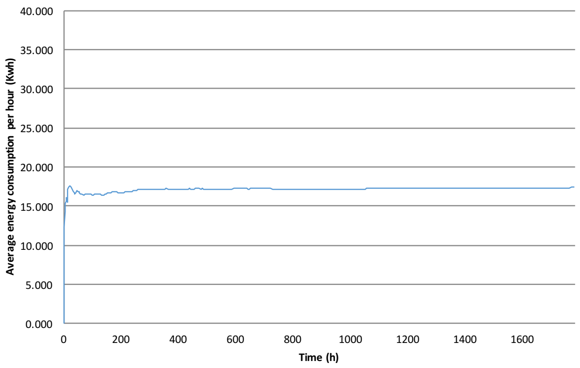


Fig. 22. Evolution of average energy consumption per hour in the smart strategy

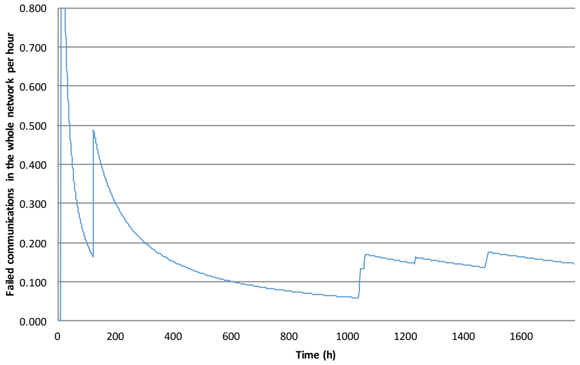


Fig. 23. Failed communications in the whole network per hour in the smart strategy

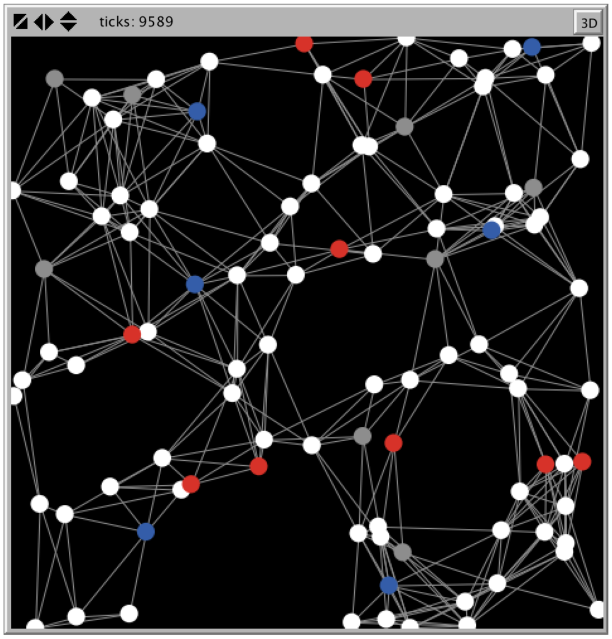


Fig. 24. An example of the network when using the smart strategy

The current work will be extended by deploying these data mining techniques in several kinds of emerging wireless technologies such as WiMAX, LTE-A and Mesh networks. In this manner, we will be able to assess and compare the utility of the current approach in each of these emerging wireless technologies. In addition, we will improve our techniques by a more rigorous evaluation of each data mining model(s) and the resulting changes to wireless network strategies.

Acknowledgment

This work acknowledges the research project "Construcción de un framework para agilizar el desarrollo de aplicaciones móviles en el ámbito de la salud" funded by University of Zaragoza and Foundation Ibercaja with grant reference JIUZ-2017-TEC-03. This work has been supported by the program "Estancias de movilidad en el extranjero José Castillejo para jóvenes doctores" funded by the Spanish Ministry of Education, Culture and Sport with reference CAS17/00005. We also acknowledge support from "Universidad de Zaragoza", "Fundación Bancaria Ibercaja" and "Fundación CAI" in the "Programa Ibercaja-CAI de Estancias de Investigación" with reference IT1/18. We acknowledge the research project "Desarrollo Colaborativo de Soluciones AAL" with reference TIN2014-57028-R funded by the Spanish Ministry of Economy and Competitiveness. It has also been supported by "Organismo Autónomo Programas Educativos Europeos" with reference 2013-1-CZ1-GRU06-14277. This work was also supported in part by the Ministerio de Economía y Competitividad in the Programa Estatal de Fomento de la Investigación Científica y Técnica de Excelencia, Subprograma Estatal de Generación de Conocimiento within the project with reference TIN2017-84802-C2-1-P.

References

[1] J.-D. Decotignie, "Propagation delays in self-organized wireless sensor networks," *IFAC Proceedings Volumes*, vol. 36, no. 13, pp. 53–57, 2003.

[2] D. Ali, M. Yohanna, and W. Silikwa, "Routing protocols source of self-similarity on a wireless network," *Alexandria Engineering Journal*, 2017.

[3] R. Lacuesta, J. Lloret, M. Garcia, and L. Penalver, "A secure protocol for spontaneous wireless ad hoc networks creation," *IEEE transactions on parallel and distributed systems*, vol. 24, no. 4, pp. 629–641, 2013.

[4] S. Sendra, R. Lacuesta, J. Lloret, and E. Macias-López, "A secure spontaneous mobile ad hoc cloud computing network," *Journal of Internet Technology*, vol. 18, no. 7, pp. 1485–1498, 2017.

[5] M. Gilesh and R. Hansdah, "An adaptive reliable transport protocol based on automatic resend request (asq) technique for wireless sensor networks," in *Advanced Information Networking and Applications (WAINA), 2011 IEEE Workshops of International Conference on*. IEEE, 2011, pp. 409–416.

[6] A. Karmokar and A. Anpalagan, "Green computing and communication techniques for future wireless systems and networks," *IEEE Potentials*, vol. 32, no. 4, pp. 38–42, 2013.

[7] R. R. Swain, P. M. Khilar, and S. K. Bhoi, "Heterogeneous fault diagnosis for wireless sensor networks," *Ad Hoc Networks*, vol. 69, pp. 15–37, 2018.

[8] V. Kotu and B. Deshpande, *Predictive analytics and data mining: concepts and practice with RapidMiner*. Morgan Kaufmann, Elsevier: Waltham, MA, USA, 2014.

[9] M. Babiš and P. Magula, "NetLogoAn alternative way of simulating mobile ad hoc networks," in *Wireless and Mobile Networking Conference (WMNC), 2012 5th Joint IFIP*. IEEE, 2012, pp. 122–125.

[10] L. Kansal, V. Sharma, and J. Singh, "Performance evaluation of FFT-WiMAX against WHT-WiMAX over Rayleigh fading channel," *Optik - International Journal for Light and Electron Optics*, vol. 127, no. 10, pp. 4514 – 4519, 2016.

[11] R. K. Polaganga and Q. Liang, "Self-Similarity and modeling of LTE/LTE-A data traffic," *Measurement*, vol. 75, pp. 218 – 229, 2015.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2018.2825954, IEEE Access

[12] P.-C. Hsiao, T.-C. Chiang, and L.-C. Fu, "Static and dynamic minimum energy broadcast problem in wireless ad-hoc networks: A pso-based approach and analysis," *Applied Soft Computing*, vol. 13, no. 12, pp. 4786 – 4801, 2013.

[13] M. Boushaba, A. Hafid, and M. Gendreau, "Node stability-based routing in wireless mesh networks," *Journal of Network and Computer Applications*, vol. 93, pp. 1 – 12, 2017.

[14] J. J. Gómez-Sanz, C. R. Fernández, and J. Arroyo, "Model driven development and simulations with the ingenias agent framework," *Simulation Modelling Practice and Theory*, vol. 18, no. 10, pp. 1468–1482, 2010.

[15] P. Ristoski, C. Bizer, and H. Paulheim, "Mining the web of linked data with rapidminer," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 35, pp. 142–151, 2015.

[16] A. Naik and L. Samant, "Correlation review of classification algorithm using data mining tool: WEKA, Rapidminer, Tanagra, Orange and Knime," *Procedia Computer Science*, vol. 85, pp. 662–668, 2016.

[17] M. M. Noor and M. Norulazmi, "Cell-based intrusion detection using wireless mesh network." *International Journal of Academic Research*, vol. 5, no. 5, pp. 94–99, 2013.

[18] C. Küçükkeçeci *et al.*, "Big data model simulation on a graph database for surveillance in wireless multimedia sensor networks," *Big Data Research*, p. https://doi.org/10.1016/j.bdr.2017.09.003, 2017.

[19] S. K. Mohapatra, P. K. Sahoo, and S.-L. Wu, "Big data analytic architecture for intruder detection in heterogeneous wireless sensor networks," *Journal of Network and Computer Applications*, vol. 66, pp. 236–249, 2016.

[20] N. A. Alrajeh and J. Lloret, "Intrusion detection systems based on artificial intelligence techniques in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 9, no. 10, p. 351047, 2013.

[21] M. Lopez-Martin, B. Carro, A. Sanchez-Esguevillas, and J. Lloret, "Conditional Variational Autoencoder for Prediction and Feature Recovery Applied to Intrusion Detection in IoT," *Sensors*, vol. 17, no. 9, p. 1967, 2017.

[22] S. Hassan, J. Pavón, L. Antunes, and N. Gilbert, "Injecting data into agent-based simulation," in *Simulating Interacting Agents and Social Phenomena*, ser. Agent-Based Social Systems. Springer, 2010, vol. 7, pp. 177–191.

[23] I. García-Magariño and R. Lacuesta, "ABS-TrustSDN: An agent-based simulator of trust strategies in software-defined networks," *Security and Communication Networks*, vol. 2017, pp. Article ID 8 575 842, 9 pages, https://doi.org/10.1155/2017/8 575 842, 2017.

[24] S. Hassan, M. Salgado, and J. Pavón, "Friendship dynamics: modelling social relationships through a fuzzy agent-based simulation," *Discrete Dynamics in Nature and Society*, vol. 2011, pp. Article ID 765 640, 19 pages, http://dx.doi.org/10.1155/2011/765 640, 2011.

[25] P. Heywood, S. Maddock, J. Casas, D. Garcia, M. Brackstone, and P. Richmond, "Data-parallel agent-based microscopic road network simulation using graphics processing units," *Simulation Modelling Practice and Theory*, p. https://doi.org/10.1016/j.simpat.2017.11.002, 2017.

[26] C. Gutierrez and I. Garcia-Magariño, "A metrics suite for the communication of multi-agent systems," *Journal of Physical Agents*, vol. 3, no. 2, pp. 7–14, 2009.

[27] S. Balev, A. Dutot, and D. Olivier, "Networking, networks and dynamic graphs," in *Agent-based Spatial Simulation with NetLogo, Volume 2*, A. Banos, C. Lang, and N. Marilleau, Eds. London, UK: ISTE Press Ltd., Elsevier:, 2017, pp. 85–116.

[28] I. García-Magariño, A. Gómez-Rodríguez, J. C. González-Moreno, and G. Palacios-Navarro, "PEABS: a process for developing efficient agent-based simulators," *Engineering Applications of Artificial Intelligence*, vol. 46, pp. 104–112, 2015.

[29] I. García-Magariño, G. Palacios-Navarro, and R. Lacuesta, "TABSAOND: A technique for developing agent-based simulation apps and online tools with nondeterministic decisions," *Simulation Modelling Practice and Theory*, vol. 77, pp. 84–107, 2017.