Wireless Netw (2019) 25:1567–1584
DOI 10.1007/s11276-017-1612-0

# Predicting throughput in IEEE 802.11 based wireless networks using directional antenna

**Saravanan Kandasamy[1] · Ricardo Morla[1] · Patrícia Ramos[2] · Manuel Ricardo[1]**

Published online: 9 November 2017

**Abstract** In IEEE 802.11 based wireless networks interference increases as more access points are added. A metric helping to quantize this interference seems to be of high interest. In this paper we study the relationship between the *improved attacking case* metric, which captures interference, and throughput for IEEE 802.11 based network using directional antenna. The $y^{1/3} = a + b\ (\ln \mathrm{x})^3$ model was found to best represent the relationship between the interference metric and the network throughput. We use this model to predict the performance of similar networks and decide the best configuration a network operator could use for planning his network.

**Keywords** Directional antenna · Prediction · Regression analysis · IEEE 802.11 · Wireless networks

✉ Saravanan Kandasamy
kandasamy@inesctec.pt

Ricardo Morla
ricardo.morla@inesctec.pt

Patrícia Ramos
pramos@inesctec.pt

Manuel Ricardo
mricardo@inesctec.pt

[1] Centre for Telecommunications and Multimedia (CTM), INESC TEC - Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

[2] Instituto Politécnico do Porto (ISCAP) and Centre for Enterprise Systems Engineering (CESE), INESC TEC - Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

## 1 Introduction

In recent years there has been keen interest in the IEEE 802.11 based Wireless Local Area Network (WLAN) technologies. Equipments that are inexpensive, easily available and that could be operated without a license are the key contributing factors for the technology to gain fame, spurring rapid deployment. Due to the ever increasing user and application demands, WLANs are expected to provide good network performance characteristics such as throughput. Unfortunately it is difficult to meet this expectation because providing coverage for areas such as shopping complexes, universities or metropolitan cities require a high number of Access Points (APs). Moreover, multiple overlapping WLANs arise due to different entities setting up networks unplanned in the same geographical area. As a consequence, the networks saturate due to sensing and interference, and the capacity of network is reached fastly [1–4]. Adding more APs to the network is a common solution to address this problem but it may not increase the network's capacity beyond a certain limit. In fact, the performance of the network could degrade even further if this is not done carefully due to the inherent hidden and exposed nodes problems.

Omnidirectional antenna (OA) is the only antenna supported by the IEEE 802.11 standard [5] but there are many directional antenna (DA) based IEEE 802.11 networks that have been deployed [6–10]. The advantage of using DA [11, 12] include the following: (1) a node could send signals to desired directions allowing the receiver node to avoid interference that comes from unwanted directions; this increases the signal to interference plus noise ratio (SINR); (2) the higher spatial reuse factor of DA, when compared to OA, could allow for more users to utilize a network simultaneously; (3) a source node could potentially reach its destination node in a lesser hop count in multihop scenario, due to the increased transmission range obtained from the higher gain of antenna. As such, DA may be more appealing than OA in some wireless network scenarios.

This paper aims to study the relationship between the *improved attacking case* metric which we proposed in [13] and throughput for WLAN with nodes using DA. Through regression analysis the relationship between the two variables is studied. This relationship is then used to predict the throughput of a similar network utilizing the calculated *improved attacking case* metric as its input parameter.

As in [13], the wireless video surveillance network (refer Fig. 1) is considered the basic scenario of our study. An IEEE 802.11 based station (STA) with an attached video surveillance camera is randomly placed in a network. The STA connects to its closest AP stationed at a fixed location and transmits its video traffic towards the AP. The APs have access to Internet via a wired connection. As we aim to study various degrees of interference in WLAN, the network is evaluated using various channels and power control strategies employing the Basic Access scheme of Distributed Coordinated Function also known as Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) of the IEEE 802.11 MAC protocol. Each STA always has traffic to send as it is fitted with a video surveillance camera and it continuously competes for accessing the medium.

The major contribution of this paper is a model for estimating the aggregated throughput of a IEEE 802.11 based wireless network using the *improved attacking case* metric. This model could be used: (a) to predict similar network's aggregated throughput upon calculating its *improved attacking case* which is obtained from the network topology characteristics; (b) to decide the best configuration for a network based on aggregated throughput requirement among the options available such as node positions, antenna type, number of channels, and scheme used to control transmission power of nodes. These contributions can be particularly useful for network planners to design their wireless network.

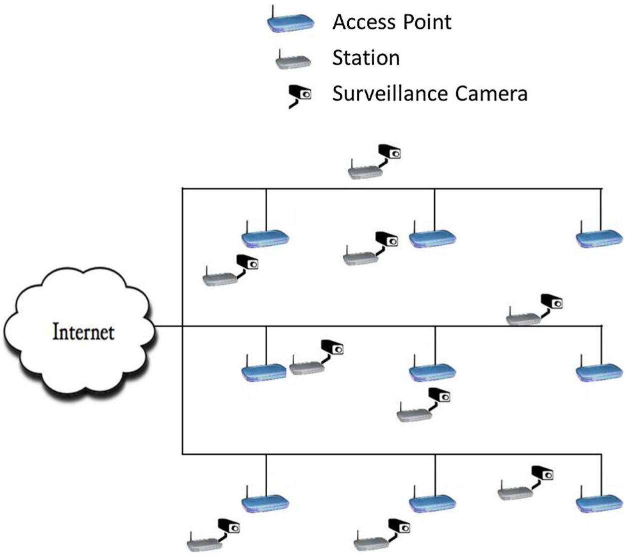


**Fig. 1** The wireless videos surveillance network deployed as a basic scenario

The rest of the paper is organized as follows. In Sect. 2 we present related work and discuss the research space our work fills. In Sect. 3 we provide an overview of the *improved attacking case* metric which characterizes the interference in IEEE 802.11 networks. In Sect. 4 we describe the simulation carried out and the performance results obtained. In Sect. 5, the regression analysis of the simulation results are presented and a regression model chosen to predict throughput from the *improved attacking case*. In Sect. 6, the applications of the chosen regression model to a network planner are presented. Finally, in Sect. 7 we draw the conclusions and indicate topics for future work.

## 2 Related work

In this section we present relevant related works and review the literature from two perspectives: (a) quantizing interference and (b) throughput prediction.

### 2.1 Quantizing interference

Interference is a major issue in WLANs where a node's radio frequency transmission disturbs the node(s) within its radio range. Transmitting using high power increases the number of nodes being interfered. Interference quantization is useful to explain the intensity of the interference in a network. It involves characterizing the interference and represent a metric to explain the severity of it. Parameters such as throughput and packet error ratio do not directly explain the interference that exists in a network. Vlavianos et. al [14] has studied SINR. It is perhaps the closest way to quantize interference but it is not only a local metric, measuring the exact SINR value is also hard.

To the best knowledge of the authors there is only one work done to quantize the severity of interference for a IEEE 802.11 based wireless network in aggregated form. S.C.Liew [15] proposed the *attacking case*, a metric that considers the interference caused by protocol dependent and independent constraints which are captured in graphs form. Information such as transmission power, nodes position, signal to interference ratio and radio propagation model are captured by the *attacking case* metric and used to identify the instances where simultaneous transmissions are not allowed or, if allowed, one or both of the transmission(s) would fail. Although very good, the approach is not suitable for nodes using DA.

We have extended the *attacking case* in [13] to cater for nodes using DA. The *improved attacking case* metric introduced may be used in nodes using OA or DA. A high value of the *improved attacking case* metric indicates severe interference in the network. As such this metric is helpful to understand the performance of a wireless network.

### 2.2 Throughput prediction

Throughput prediction has an important role in network planning, measurement and management processes. Papadopouli et. al [16] have characterized the traffic load of an IEEE 802.11 network and proposed traffic prediction algorithms based on traffic models. Chen et. al [17] used the ARIMA model to predict short-term traffic in IEEE 802.11 networks. Yu et. al [18] found that the more accurate the traffic prediction is the higher efficiency and utilization ratio of network bandwidth can be guaranteed. They proposed a Minimax Probability Machine Regression model to predict the wireless network traffic in IEEE 802.11 networks. The works by Papadopouli, Chen and Yu are short term prediction models, where the validity of prediction is rather short lived as it depends on the data behavior that changes in time. We use the *improved attacking case* metric which is cross sectional data rather than time-series data to predict the network throughput. In this way the model could be applied in any type of similar network regardless the instantaneous behavior changes of data.

Chen et. al [19] proposed throughput prediction model using the signal to noise ratio (SNR) information from the traffic statistics of a wireless network. They considered piecewise and exponential model to do the curve fitting. The correlation coefficient R is used as one of its performance metric. Curve fitting with only two models is inadequate to determine the best model to represent the data. Further, R is not a good metric to sufficiently explain non-linear models. In our work we consider 1936 models when doing the curve fitting and used the coefficient of determination $R^2$ which is more stringent than R as one of the metric to explain the prediction model.

Nghia and Robert [20] predicted throughput using the contention window value. They found that a large contention window size can lead to larger throughput. In [21] Bruno estimated of the throughput obtained by persistent TCP flows. Dely et. al [22] estimated the load of the wireless channel using channel busy fraction as an indicator of fraction of time in which the wireless channel is sensed busy due to successful or unsuccessful transmissions. They showed the channel busy fraction allows an accurate prediction of the available bandwidth with small error. Tang and Wang [23] introduced $q(n)$, a parameter that describes the network load condition which is used as an input parameter to predict the throughput. Although these works [20–23] could be used to predict the throughput of networks tested, the authors have not shown how the proposed approaches could be utilized in planning a similar network. In our work, we will not just predict the throughput of a network but also show how to use the model to improve the CSMA/CA based wireless network.

## 3 *Improved attacking case*

This section provides an overview of the *improved attacking case* metric that is proposed in [13]. The metric is used to quantize the severity of interference in IEEE 802.11 (CSMA/CA) based wireless network. Interested readers are suggested to read [13] for further details on this metric. This metric addresses the IEEE 802.11 based wireless networks with nodes using both OAs and DAs.

### 3.1 Constraints

The *improved attacking case* is characterized by the Link-Interference Graph, Transmitter-side Protocol Collision Prevention Graph, and Receiver-side Protocol Collision Prevention Graph which are built using the Physical Collision Constraints and Protocol Collision Prevention Constraints.

#### *3.1.1 Physical collision constraints*

The Physical Collision Constraints are modeled using the pair-wise interference model. Consider two data links, Link $i$ and Link $j$, communicating using the Basic Access Scheme of IEEE 802.11 MAC protocol (DATA and ACK). Let $T_i$ to represent the position of the transmitter and $R_i$ of the receiver of Link $i$. For simplicity $T_i$ and $R_i$ are also used to refer to the nodes.

$T_i$ transmits DATA and receives ACK while $R_i$ receives DATA and transmits ACK. Four different possible combination of simultaneous transmissions by Link $i$ and Link $j$ may occur: $\mathrm{DATA}_i - \mathrm{DATA}_j$, $\mathrm{DATA}_i - \mathrm{ACK}_j$, $\mathrm{ACK}_i - \mathrm{DATA}_j$, and $\mathrm{ACK}_i - \mathrm{ACK}_j$. Four combinations of the following Physical Collision Constraints can be derived. The transmission of Link $i$ will be interfering with the transmission of Link $j$ if,

$$P(T_j, \theta_{R_j}, R_j) < KP(T_i, \theta_{R_j}, R_j) \quad (\mathrm{DATA}_i - \mathrm{DATA}_j) \tag{1}$$

$$P(R_j, \theta_{T_j}, T_j) < KP(T_i, \theta_{T_j}, T_j) \quad (\mathrm{DATA}_i - \mathrm{ACK}_j) \tag{2}$$

$$P(T_j, \theta_{R_j}, R_j) < KP(R_i, \theta_{R_j}, R_j) \quad (\mathrm{ACK}_i - \mathrm{DATA}_j) \tag{3}$$

$$P(R_j, \theta_{T_j}, T_j) < KP(R_i, \theta_{T_j}, T_j) \quad (\mathrm{ACK}_i - \mathrm{ACK}_j) \tag{4}$$

where $P(a, \theta_b, b)$ is the power received by node $b$ from angle $\theta_b$ of node $a$. $K$ is the Signal to Interference Ratio (SIR) threshold for a packet to be successfully decoded by the IEEE 802.11 protocol (e.g 10 dB).

### 3.1.2 Protocol collision prevention constraints

The Protocol Collision Prevention Constraints of IEEE 802.11 consider the effect of carrier sensing with the goal of preventing simultaneous transmissions within the same radio region. The prevention of a transmission can be induced at the transmitter's side, at the receiver's side, or at both sides.

*Transmitter side*—A transmitter would refrain from transmitting a DATA packet if it can sense the transmission of another ongoing transmission. The transmission of Link $i$ will interfere with the transmission of Link $j$ if,

$$|T_j - T_i| < CSRange\left(P_{T_i}^{\theta_{T_j}}\right) \quad (\mathrm{DATA}_i - \mathrm{DATA}_j) \tag{5}$$

$$|T_j - R_i| < CSRange\left(P_{R_i}^{\theta_{T_j}}\right) \quad (\mathrm{ACK}_i - \mathrm{DATA}_j) \tag{6}$$

$$|T_j - T_i| < TXRange\left(P_{T_i}^{\theta_{T_j}}\right) \quad (\mathrm{DATA}_i - \mathrm{DATA}_j) \tag{7}$$

where CSRange is the carrier sensing range and TXRange is the transmission range of a transmitting node.

*Receiver side*—In IEEE 802.11 commercial products, when $T_i$ is already transmitting, $T_j$ can still transmit if Eqs. 5, 6, and 7 is not true. However, $R_j$ will ignore the DATA packet and not return an ACK packet, causing $T_j$ to interpret that as a collision triggering for a backoff and a retransmission [15, 24]. The transmission of Link $i$ will interfere with the transmission of Link $j$ if,

$$|R_j - T_i| < CSRange\left(P_{T_i}^{\theta_{R_j}}\right) \quad (\mathrm{DATA}_i - \mathrm{ACK}_j) \tag{8}$$

$$|R_j - R_i| < CSRange\left(P_{R_i}^{\theta_{R_j}}\right) \quad (\mathrm{ACK}_i - \mathrm{ACK}_j) \tag{9}$$

$$|R_j - T_i| < TXRange\left(P_{T_i}^{\theta_{R_j}}\right) \quad (\mathrm{DATA}_i - \mathrm{ACK}_j) \tag{10}$$

## 3.2 Graph models

Three weighted directed graphs are modeled using the Physical Collision Constraints and the Protocol Collision Prevention Constraints: the Link-Interference Graph; the Transmitter-side Protocol Collision Prevention Graph; and the Receiver-side Protocol Collision Prevention Graph. These three graphs are used to construct the *improved attacking case* metric. A general graph $G$ is defined as a collection of vertices $V$ and unidirectional edges $E$ that connect pairs of vertices with weights $w$.

$$G = (V, E, w) \tag{11}$$

For any unidirectional edge $e_{ij} \in E$ where $i, j \in V$, vertex $i$ represents Link $i$ consisting of $\mathrm{T}_i$ and $\mathrm{R}_i$ nodes, while $e_{ij}$ represents a relationship between Link $i$ and Link $j$. The weight is a function of $e_{ij}$ where $w(e_{ij}) \in \mathbb{N}$. The value of $w(e_{ij})$ depends on the type graph being modeled.

### 3.2.1 Link-interference graph (i-graph)

The Physical Collision Constraints can be represented by a Link-Interference Graph. The graph captures the SIR effects among links and represented as follows:

$$G_I = (V_I, E_I, w_I) \tag{12}$$

If any of the constraints in Eqs. 1, 2, 3 or 4 is satisfied, there is an edge from vertex $i$ to vertex $j$ to signify that Link $i$ is interfering with Link $j$ with a weight $w_I(e_{ij})$ characterized as follows:

$$\begin{aligned} w_I(e_{ij}) = {} & \mathbb{1}_{\left[P_{T_j}^{\theta_{R_j}} |T_i - R_j|^\alpha < KP_{T_i}^{\theta_{R_j}} |T_j - R_j|^\alpha\right]} \\ & + \mathbb{1}_{\left[P_{R_j}^{\theta_{T_j}} |T_i - T_j|^\alpha < KP_{T_i}^{\theta_{T_j}} |T_j - R_j|^\alpha\right]} \\ & + \mathbb{1}_{\left[P_{T_j}^{\theta_{R_j}} |R_i - R_j|^\alpha < KP_{R_i}^{\theta_{R_j}} |T_j - R_j|^\alpha\right]} \\ & + \mathbb{1}_{\left[P_{R_j}^{\theta_{T_j}} |R_i - T_j|^\alpha < KP_{R_i}^{\theta_{T_j}} |T_j - R_j|^\alpha\right]} \end{aligned} \tag{13}$$

where Eq. 13 is built using components of indicator function as defined in Eq. 14 and $\alpha$ is the path-loss exponent. Since $w_I(e_{ij})$ exists only when there is an $e_{ij}$, $w_I(e_{ij}) \in \{1, 2, 3, 4\}$ for i-graph.

$$\mathbb{1}_{[C]} = \begin{cases} 1, & \textit{if } C = \textit{TRUE} \\ 0, & \textit{if } C = \textit{FALSE} \end{cases} \tag{14}$$

### 3.2.2 Transmitter-side protocol collision prevention graph (tc-graph)

The effect of carrier sensing by the transmitters is modeled by tc-graph and it is represented as follows:

$$G_{TC} = (V_{TC}, E_{TC}, w_{TC}) \tag{15}$$

Formally, if any of the Eqs. 5, 6 or 7 holds true then there is a tc-edge from vertex $i$ to vertex $j$. The weight $w_{TC}(e_{ij})$ of tc-edge is characterized as follows:

$$w_{TC}(e_{ij}) = \mathbb{1}_{\left[\left(|T_j - T_i| < CSRange\left(P_{T_i}^{\theta_{T_j}}\right)\right) \vee \left(|T_j - T_i| < TXRange\left(P_{T_i}^{\theta_{T_j}}\right)\right)\right]} + \mathbb{1}_{\left[|T_j - R_i| < CSRange\left(P_{R_i}^{\theta_{T_j}}\right)\right]} \tag{16}$$

Since $w_{TC}(e_{ij})$ exists only when there is an $e_{ij}$, $w_{TC}(e_{ij}) \in \{1, 2\}$ for tc-graph. As the tc-graph models the effect of carrier sensing purely from the transmitter point of view, it does not consider tc-edges created due to the $DATA_1 - ACK_2$ and $ACK_1 - ACK_2$ pairs of transmission from vertex 1 to vertex 2 and $DATA_2 - ACK_1$ and $ACK_2 - ACK_1$ pairs of transmission from vertex 2 to vertex 1 due to its effect solely at the receiver.

### 3.2.3 Receiver-side protocol collision prevention graph (rc-graph)

The effect of carrier sensing by the receivers is modeled by rc-graph and it is represented as follows:

$$G_{RC} = (V_{RC}, E_{RC}, w_{RC}) \tag{17}$$

If any of Eqs. 8, 9 or 10 is true, there is an rc-edge from vertex $i$ to vertex $j$ with a weight $w_{RC}(e_{ij})$ characterized as follows:

$$w_{RC}(e_{ij}) = \mathbb{1}_{\left[\left(|R_j - T_i| < CSRange\left(P_{T_i}^{\theta_{R_j}}\right)\right) \vee \left(|R_j - T_i| < TXRange\left(P_{T_i}^{\theta_{R_j}}\right)\right)\right]} + \mathbb{1}_{\left[|R_j - R_i| < CSRange\left(P_{R_i}^{\theta_{R_j}}\right)\right]} \tag{18}$$

Since $w_{RC}(e_{ij})$ exist only when there is an $e_{ij}$, $w_{RC}(e_{ij}) \in \{1, 2\}$ for rc-graph. Since rc-graph models the effect of carrier sensing purely from the receiver point of view, it does not consider rc-edges created due to the $ACK_1 - DATA_2$ and $DATA_1 - DATA_2$ pairs of transmission from vertex 1 to vertex 2, and $ACK_2 - DATA_1$ and $DATA_2 - DATA_1$ pairs of transmission from vertex 2 to vertex 1.

For i-graph, tc-graph and rc-graph all the vertices are the same, where $V = V_I = V_{TC} = V_{RC}$.

## 3.3 Formulation of improved attacking case metric

The *improved attacking case* ($AC_{Imp}$) corresponds to the number of cases where simultaneous transmissions are either not allowed or if allowed will not be successful. It is formulated as following: (1) if $e_{i,j}$ is an i-edge then twice the i-edge's weight is added to the *improved attacking case* else; (2) if $e_{i,j}$ is a tc-edge then the tc-edge's weight is added to the improved *attacking case*, and (3) if $e_{i,j}$ is a rc-edge then the rc-edge's weight is added to the improved *attacking case* for all *i,j* where $i \neq j$ as shown in Eq. 19.

$$AC_{Imp} = \sum_{\substack{i,j \in V \\ i \neq j}} \left[ 2 \times w_I(e_{i,j}) \times \mathbb{1}_{[e_{i,j} \in E_I]} + w_{TC}(e_{i,j}) \times \mathbb{1}_{[e_{i,j} \in E_{TC} \wedge e_{i,j} \notin E_I]} + w_{RC}(e_{i,j}) \times \mathbb{1}_{[e_{i,j} \in E_{RC} \wedge e_{i,j} \notin E_I]} \right] \tag{19}$$

## 4 Evaluation setup and data acquisition

In this section, the *improved attacking case* metric is used to quantize the severity of interference of WLAN with nodes using DA and its throughput performance is assessed by means of Network Simulator 2 (NS2) simulation. The performance of network with nodes utilizing OA is also evaluated for benchmarking purpose as the *improved attacking case* is also able to cater for OA. The *improved attacking case* metric and throughput data pairs are used to study the relationship between them and a throughput prediction model is built.

### 4.1 NS2 supporting directional antenna

NS2 was extended to support nodes with DA. Each node supporting DA is assumed to have 4 interfaces where each of its interface is associated with a 90° passive DA of gain 2. This allow a node to able to listen, transmit and receive packets in 360° direction as OA. Each interface has directional MAC, directional NAV, its own IFQ, and maintains its own ARP table as shown by the stack in Table 1. The DA in interfaces 0, 1, 2 and 3 are pointed to angle 0°, 90°, 180°, 270° respectively. The work presented in this paper consider this model for DA.

### 4.2 Scenario setup

A 3 x 3 grid topology is defined as the basic scenario with nodes separated by 250 m which act as APs. STAs are represented by the additional nodes placed randomly in the network. These STAs connect to the closest AP. Traffic is sent from the STAs towards the APs following the video surveillance network scenario as in Fig. 1. Being a single hop wireless network, routing was not considered. In this paper, we have investigated IEEE 802.11b with 11 Mbps data rate as a representative release of IEEE 802.11 protocol. The number of random STAs in the network was gradually varied from 9 to 18, 27, and 36, aiming to increase the amount of interference in the network. The number of channels utilized by the network varied from single channel (SC), to two channels (TC) where neighbour APs use different channels, and nine channels (NC) where

**Table 1** Stack of directional antenna for a node in NS2

| RTR | | | |
|---|---|---|---|
| LL_0 + ARP_0 | LL_1 + ARP_1 | LL_2 + ARP_2 | LL_3 + ARP_3 |
| IFQ_0 | IFQ_1 | IFQ_2 | IFQ_3 |
| MAC_0 | MAC_1 | MAC_2 | MAC_3 |
| NetIF_0 | NetIF_1 | NetIF_2 | NetIF_3 |

each AP uses a unique channel, to study the channel's effect on the network's interference.

Other than the default transmission power, the network is also evaluated using the minimal transmission power algorithm presented in Algorithm 1. The transmission power is enough for a transmitter node to get its transmitted packets decoded by its receiving (closest in physical distance) node defined by $\mathrm{RX}_{th}$. $\mathrm{RX}_{th}$ is the received signal strength threshold required to decode a packet. The equations in Line 2 and Line 3 of Algorithm 1 ensure that the reduced powers satisfy the minimum received power threshold required to maintain an arbitrary Link $i$'s connectivity.

**Algorithm 1** Minimum Transmit Power

```
Require: $ActiveLinks \subset RadioLinks, n \in Nodes, int \in n$
Ensure: $MP\_NT = \{\}, MP\_ND = \{\}, MP\_IN = \{\}$
 1: for all $(T_i, \theta_{R_i}, R_i) \in ActiveLinks$ do
 2:     $\left(P_{T_i}^{\theta_{R_i}}\right)_{new} = \frac{P_{T_i}^{\theta_{R_i}}}{P(T_i, \theta_{R_i}, R_i)} \times RX_{th}$
 3:     $\left(P_{R_i}^{\theta_{T_i}}\right)_{new} = \frac{P_{R_i}^{\theta_{T_i}}}{P(R_i, \theta_{T_i}, T_i)} \times RX_{th}$
 4: end for
 5:
 6: for all $n \in Nodes$ do
 7:     for all $\theta \in n$ do
 8:         $MP\_IN.add(P_n^{\theta})$
 9:     end for
10:     $MP\_ND.add(max(MP\_IN))$
11: end for
12: $MP\_NT.add(max(MP\_ND))$
13:
14: return $MP\_IN, MP\_ND, MP\_NT$
```

The minimum transmit power algorithm is implemented in 3 strategies:

(a) the minimum power per network (MP-NT) – in this approach the interfaces in nodes are allowed to reduce its transmit power, but all the interfaces in the network must use the same transmit power (Line 12 of Algorithm 1). OA and DA use this.

(b) the minimum power per node (MP-ND) – in this approach, as the above, the interfaces are allowed to reduce its transmit power. Each node is allowed to have its own transmit power but all the interfaces of a node must use the same transmit power (Line 10 of Algorithm 1). OA and DA use this.

(c) the minimum power per interface (MP-IN) – in this approach, each interface is allowed to reduce and use its own transmit power (Line 8 of Algorithm 1). Only DA use this.

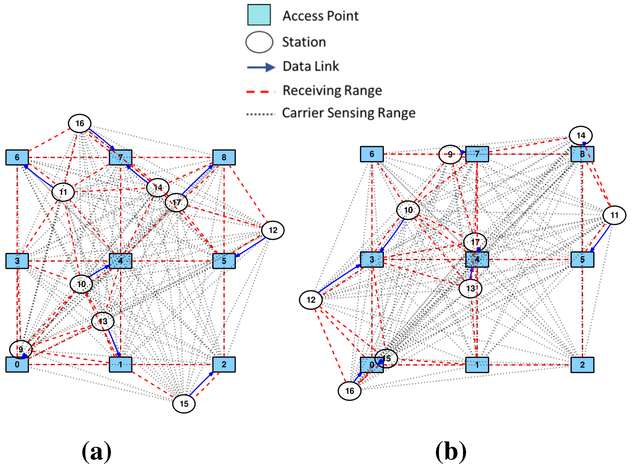


**Fig. 2** Example of random topologies for OA; 9 APs and 9 STAs deployed. **a** Random topology 1. **b** Random topology 2

The minimum transmit power strategies defined above create different severity of interference in the wireless network. For each ratio of STA and AP, 40 random topologies were generated. Some examples of the random topologies used in the simulation are presented in Fig. 2 when the nodes use NS2's default transmission power. The solid lines represent data links, the dashed lines represent nodes within receiving threshold, and the dotted lines represent nodes within carrier sensing threshold.

### 4.3 Dataset

A total of 160 random topologies were generated for all the STA:AP ratio. The throughput performance of the topologies is evaluated using OA/DA, SC/TC/NC channel configuration, MP-NT/MP-ND/MP-IN transmit power strategies and the default transmission power. The rest of the simulation parameters used in NS2 are shown in Table 2.

Throughput is measured as the total number of packets successfully received at the destinations times the packet

**Table 2** Parameters setting

| Parameter | Setting |
|---|---|
| Access mode | Basic (DATA, ACK) |
| Data rate; basic rate | 11; 1 Mbps |
| MAC | IEEE 802.11b |
| Traffic generation model | UDP, Poisson process, 1818.181$\mu$s mean arrival |
| Offered load per STA | 55 pkt/s |
| Packet size | 1500 bytes |
| Interface queue (IFQ) length | 50 packets |
| SIR | 10 dB |
| Propagation model | 2-ray ground reflection |
| NS2's default transmission power | 281.8 mW |
| RXThresh; CSThresh | 36.5 nW; 156 nW |
| Simulation time | 120 s |
| Directional antenna angles | 0°, 90°, 180°, 270° |
| Number of DAs/node | 4, 90° beamwidth each |
| Antenna gain | OA:1, DA: 2 *(in ref. to isotropic antenna)* |
| Number of simulations/scenario | 40 |
| Number of APs | 9 |
| Number of STAs | 9, 18, 27, 36 |

size over the duration of the flows. Formally, the throughput is calculated using Eq. 20.

$$Tput(Mbps) = \frac{\left(\sum_{i=1}^{n} RcvdPkt_i\right) \times Packet_{size}}{T_D} \tag{20}$$

where $n$ is the number of flows, $i$ is the flow number and $T_D$ is the simulation time. The *improved attacking case* is calculated using Eq. 19 for the 160 random topologies. There are a total of 3360 data generated in the dataset contributed by various combination of topologies, antenna, channel, power strategies and default transmission power, each consisting a pair of the *improved attacking case* and throughput information.

## 5 Regression analysis and discussion

In this section we present and discuss the relationship between the *improved attacking case* and the aggregated throughput observed in NS2 simulations based on the results obtained in Sect. 4.3. A model that best fits the dataset is proposed by using simple linear regression (SLR) which is defined by a relationship between one dependent variable and one independent variable [25] in the following general form:

$$y = a + bx + \epsilon \tag{21}$$

where $y$ is the dependent variable that is being predicted, $x$ is the independent variable that is used to predict, $\epsilon$ is the error term while $a$ and $b$ are the coefficients in the regression equation. In our case, the *improved attacking case* and throughput are the independent ($x$) and dependent ($y$) variables respectively. This model will be used to predict the throughput of a similar WLAN. R language [26] which is a statistical computing software is used in investigating the SLR.

### 5.1 Exploratory analysis

Scatter plots are most useful for studying the relationship between the independent and dependent variables through visual method. Often unusual observations are detected

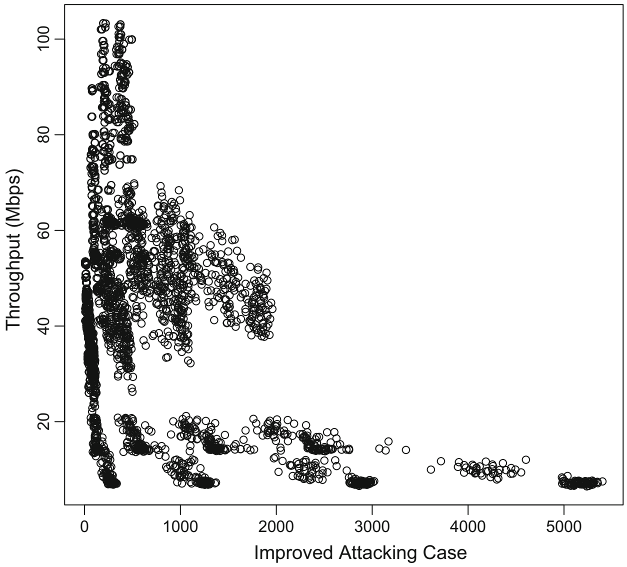


**Fig. 3** Scatter plot of the throughput versus the *improved attacking case*

through this plot [27]. Fig. 3 shows the scatter plot of the *improved attacking case* versus throughput. In the figure it can be observed that 1) multiple curves exist; and 2) throughput has a non-linear relationship with the *improved attacking case*.

SLR is not suitable to be applied to this dataset due to the heterogeneity of the data. In effort to rectify this problem, the dataset is segregated by the antenna type (i.e. OA, DA) and STA (i.e. 09, 18, 27, 36) to form eight smaller groups of data (setups). The scatter plots for these setups are shown in Fig. 4. It is observed that the dataset in each of the setup is now more homogenous. A SLR can be carried out on each of these groups of data to find one model that fits all the eight datasets.

We also observe in Fig. 4 that as the number of STAs increase, the minimum and the maximum values of the *improved attacking case* for both OA and DA setups increase. This affirms that higher number of STAs create more interference in the network as captured by the *improved attacking case* metric. The maximum value of throughput is similar for OA at approximately 60 Mbps when STAs were increased from 9 to 18, 27 and 36. This means that the WLAN is operating in a saturated region. Having more STAs does not help to increase the throughput but only increases the severity of interference in the network. DA has higher throughput and lower *improved attacking case* than OA as it is able to reduce interference on unwanted directions.

In Fig. 4, the data form clusters as the number of STAs increases. The clustering is more evident in the OA setups. Three clusters are evident, each representing the SC, TC and NC channel configuration used in the evaluation. The cluster of SC has the highest *improved attacking case* and the lowest throughput while the cluster of NC has the lowest *improved attacking case* and the highest throughput. We may conclude that the usage of more channels reduces the improved *attacking case* and increases the throughput.

In summary, as the *improved attacking case* increases the throughput decreases and the scatter plots in Fig. 4 suggest that there is relationship between them. This relationship implies that a model for throughput must include the *improved attacking case* as a predictor variable in the SLR equation.

## 5.2 Transformation to achieve linearity

As the relationship of the *improved attacking case* and the throughput in Fig. 4 is non-linear, it need to be linearized in order for SLR techniques can be applied. Table 3 presents the common methods for transforming variables to achieve linearity by applying natural logarithm (ln) to the independent variable $x$ and/or the dependent variable $y$ when a curve is observed in the scatter plot of $y$ against $x$ [28]. For benchmarking purpose the untransformed form (lin) of the $x$ and $y$ variables are maintained in this study and also shown in Table 3 as linear model.

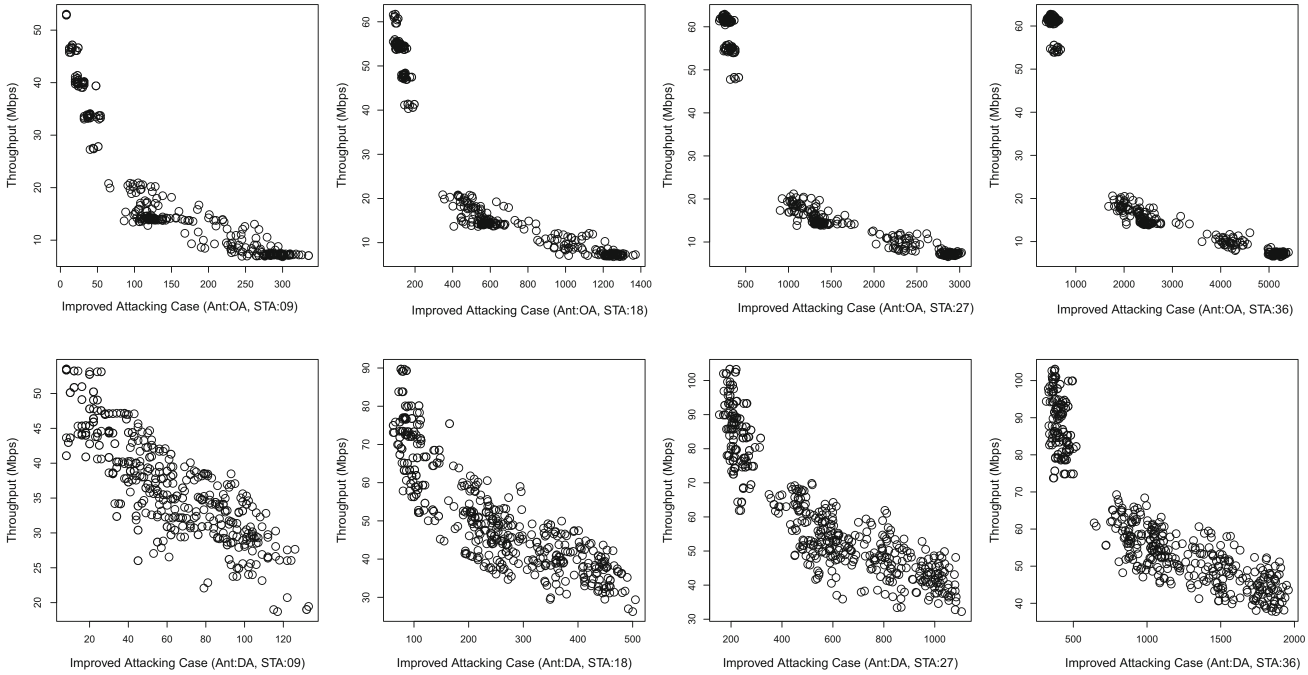


**Fig. 4** Scatter plots of the throughput versus the *improved attacking case* for OA (upper panel) and DA (lower panel) with 9, 18, 27, 36 STAs (left to right).

**Table 3** Variables transformation to achieve linearity

| Method | Regression function | Transformation | Linearized form | Model representation |
|---|---|---|---|---|
| Linear model | $y = a + bx$ | none | $y = a + bx$ | $lin \sim lin$ |
| Logarithmic model | $y = a + b \ln x$ | $x' = \ln x$ | $y = a + bx'$ | $lin \sim \ln$ |
| Exponential model | $y = ab^x$ | $y' = \ln y$ | $y' = \ln a + (\ln b)x$ | $\ln \sim lin$ |
| Power model | $y = ax^b$ | $x' = \ln x$, $y' = \ln y$ | $y' = \ln a + bx'$ | $\ln \sim \ln$ |

The first column in Table 3 shows the four methods of regression model used in our analysis, i.e., linear, logarithmic, exponential and power models. The second and third columns represent the original regression function and the specific transformation that is applied to the *x* and/or *y* variables respectively. The fourth column represents the transformed form of the regression function to achieve linearity which will be used in our regression analysis. The final column presents the model representation which is utilized in our paper to denote these linearized regression functions. To investigate the adequacy of these regression functions to our dataset, we decided to fit all the linearized models in Table 3 to the eight datasets in Fig. 4 where the transformations enable us to use SLR techniques to decide which model fits our dataset the best.

Fig. 5 shows the scatter plots of *improved attacking case* versus throughput and the fitted linear models in the transformed scales from Table 3. The fitted regression lines were estimated by using Ordinary Least Squares (OLS) method of SLR. The goal of OLS is to closely fit a model with the dataset by minimizing the sum of squared errors from the data. It can be noted that the non-linear data now appears to be linear in few of the transformed model in Fig. 5 when the wireless network use OA and DA. Some of the fitted lines in the transformed model also appear to represent the data better compared with other fitted lines.

## 5.3 Evaluation of linear regression model's residual errors

Two of the assumptions of linear regression are that the residual errors must be: 1) homoscedastic, i.e., the variance of the errors is the same across all values of the independent variable; 2) normally distributed. Violation of these

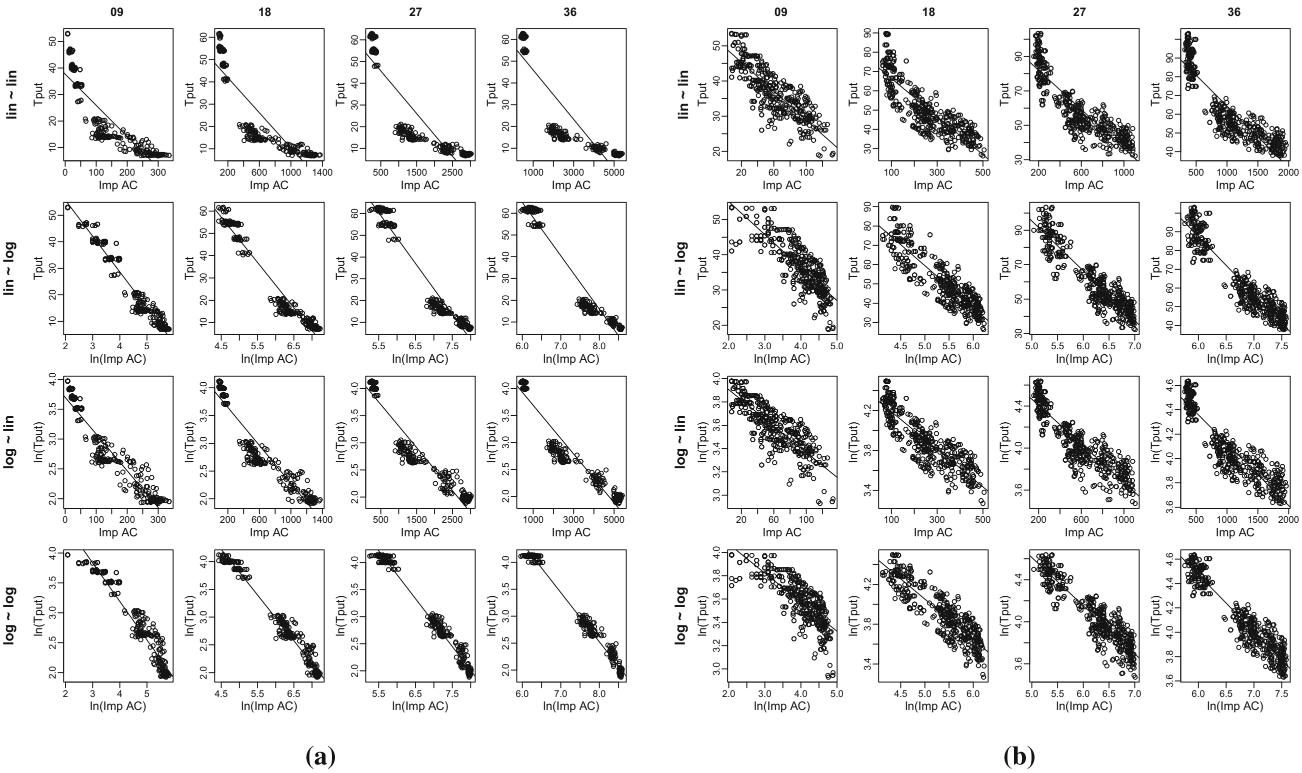


**Fig. 5** Scatter plots of the throughput versus the *improved attacking case* and fitted linear models in the transformed scales. **a** Omnidirectional antenna, **b** directional antenna

assumptions compromise the estimation of coefficients, accuracy of prediction and would give a false sense of trustworthiness. In our work, these two assumptions are validated [29, 30] and model(s) that fulfill these assumptions is short listed as the candidate model(s) to represent the data.

Breusch–Pagan (BP) and Jarque–Bera (JB) tests are used respectively to check for homoscedasticity and normality of the residual errors of the models presented in Fig. 5 [29, 30] with a significance level of 5%. If the probability values (p-values) associated with the computed BP and JB tests using R Language are above 0.05 the tests are passed and the residual errors are considered homoscedastic and normally distributed.

Unfortunately, the presence of influential points such as outliers could fail the BP and JB tests especially if it has large effect on a regression model. An outlier is defined as the data which is far detached from the general population of data set. Outliers may exist in our case since the *improved attacking case* value is a non-controlled outcome of a network depending on parameters such as the nodes position, transmission power and type of antenna used. If a BP or JB test fails when evaluating the residual errors, Cook's distance D is used as a statistical filter to identify the influential points as shown in Eq. 22.

$$D_i = \frac{\sum_{j=1}^{n} \left(\hat{y}_j - \hat{y}_{j(i)}\right)^2}{k \times MSE} \tag{22}$$

where $D_i$ is the Cook's distance for observation $i$, $\hat{y}_j$ is the j-th response value of $y$ of the fitted line, $\hat{y}_j(i)$ is the j-th response value of $y$ of the fitted line where the fit did not include observation $i$ and MSE is the mean squared error. Cook's distance is an index measure where all the points $i$ in the dataset are evaluated and compared to a critical value defined by $4/(n-k-1)$, where $n$ is the number of samples in the dataset and $k$ is the number of independent variables [31].

Once identified, the influential point with the maximum Cook's distance value is removed from the dataset. The BP and JB tests are redone to this reduced dataset. If either of these tests fails again, the elimination process is repeated. Usually up to 5% of the total data count are removed [31]. If the BP or JB tests continue to fail, the model should be removed and not considered as a candidate.

Table 4 shows the $p$ values for BP and JB tests for all the antenna type and STA. It can be concluded as none of the lin ~ lin, lin ~ ln, ln ~ lin and ln ~ ln models passed the BP and JB tests simultaneously for all the antenna type and STA.

### 5.4 Transformation for model respecification

As the residual errors in Table 4 are not homoscedastic and normally distributed for all the 8 datasets, the models warrants for respecification using a suitable transformation. This transformation is aimed to stabilize the error variance by making it constant for all the observations. Tukey's ladder of transformation [32] is used to redefine the variables in orderly manner using a power transformation. The following relationship is explored when transforming the variables:

$$Y^{\lambda_1} = a + bX^{\lambda_1} \tag{23}$$

where

$$\lambda_1, \lambda_2 \in \{-6, -5, -4, -3, -2, -1, -3/4, -2/3, -1/2, -1/3, -1/4, 1/4, 1/3, 1/2, 2/3, 3/4, 1, 2, 3, 4, 5, 6\}$$

are the exponent parameters. $Y$ and $X$ are variables either in original form or in the natural logarithm transformed form depending on lin ~ lin, lin ~ ln, ln ~ lin and ln ~ ln models. A total of 1936 models are evaluated as a product of the transformations to achieve linearity and the transformations for model respecification to determine which model(s) is(are) good to represent the 8 datasets.

**Table 4** *P* value results from the Breusch–Pagan and Jarque–Bera test using R language evaluating the assumption of homoscedasticity and normality for the residual errors

| Antenna | STA | *lin ~ lin* | *lin ~ ln* | *ln ~ lin* | *ln ~ ln* |
|---|---|---|---|---|---|
| OA | 09 | 1.06E−08/2.95E−04 | **7.28E−02/5.22E−01** | **5.14E−02**/3.00E−10 | **5.69E−01**/8.27E−11 |
| | 18 | 6.33E−04/1.57E−08 | **5.96E−01**/1.56E−05 | **6.18E−02**/2.44E−12 | **5.55E−01**/7.06E−12 |
| | 27 | 5.91E−03/2.60E−10 | **2.19E−01**/2.85E−07 | **5.94E−02**/4.31E−13 | **8.85E−01**/1.46E−12 |
| | 36 | 9.93E−03/1.67E−11 | 4.86E−02/2.04E−08 | 2.47E−02/2.08E−13 | **8.85E−01**/5.07E−13 |
| DA | 09 | **2.07E−01**/5.71E−03 | **9.37E−01/5.16E−02** | **8.69E−01**/3.87E−04 | **7.29E−02**/2.50E−04 |
| | 18 | 6.16E−06/2.43E−02 | 5.33E−05/1.01E−02 | **2.52E−01**/5.40E−08 | **9.64E−01**/2.39E−08 |
| | 27 | 2.40E−04/2.48E−02 | 1.23E−02/**6.12E−02** | **5.26E−01**/8.86E−09 | **1.06E−01**/1.22E−09 |
| | 36 | 1.70E−03/1.07E−02 | 1.13E−02/**1.43E−01** | **6.41E−01**/1.43E−10 | **1.57E−01**/3.50E−11 |

Presented in <*p* value Breusch–Pagan/*p* value Jarque–Bera> format. The bold values represents the presence of homoscedasticity or normality

The evaluation of the residual errors as described in Sect. 5.3 is explored again for all these 1936 models. Four candidate models, one from lin $\sim$ lin and three from lin $\sim$ ln, as shown in Eq. 24, have passed both the BP and JB tests simultaneously for all the antenna type and STAs. During the evaluation of the residual errors of these four model only around 1% of the data were discarded due to being influential points.

$$\begin{aligned} y^{1/3} &= a + b\, x^{1/3} \\ y^{1/4} &= a + b\ (\ln \mathrm{x})^4 \\ y^{1/3} &= a + b\ (\ln \mathrm{x})^3 \\ y^{1/2} &= a + b\ (\ln \mathrm{x})^2 \end{aligned} \tag{24}$$

### 5.5 Model selection from the candidate models

In this subsection the best model that represents the eight datasets is selected from the four candidate models presented in Eq. 24. The mean absolute percentage error (MAPE) is used as the selection criteria using the leave one out cross validation (LOOCV) method [33, 34]. Cross validation is a statistical method for evaluating and comparing models by dividing data into two segments: one used to train a model and the other used to validate the model. In LOOCV, the training and validation data are crossed over in successive rounds such that each data point has a chance of being validated against the remaining data that are used for training. At each iteration $i$, the data point which consists of a pair of independent value $x_i$ and measured value $y_i$ is removed from the dataset. The remaining data which acts as a training set is used to create a linear model $\hat{y} = a + bx$ via SLR. Using $x_i$ from the validation data point, a predicted value $\hat{y}_i$ is obtained. This predicted value $\hat{y}_i$ is compared against the measured value $y_i$ and the error of prediction $e_i$ is calculated using Eq. 25.

$$e_i = y_i - \hat{y}_i \quad \forall\ i = 1, 2, \ldots, n \tag{25}$$

The model's MAPE can be calculated using Eq. 26. Since it is a scale-independent matric MAPE can be more easily used to compare the performance of various models that are not in the same scale, as is the case here. The MAPE for the candidate models are presented in Table 5.

$$MAPE = mean\left(\left|\frac{100 \times e_i}{y_i}\right|\right) \quad \forall\ i = 1, 2, \ldots, n \tag{26}$$

As the objective is to find one common model that best fits the dataset for both OA and DA irrespective of STAs, the model with the smallest average MAPE value is selected as shown by the bold row in Table 5. It is found that $y^{1/3} = a + b\ (\ln \mathrm{x})^3$ model is the most attractive as it has the smallest average MAPE value at 11.50%. The scatter plot for the selected model is shown in Fig. 6 with fitting for both OA and DA and all the STAs.

### 5.6 Inference test for the selected model

In this subsection the $y^{1/3} = a + b\ (\ln \mathrm{x})^3$ model which has been selected to represent the network irrespective on the number of STAs and antenna type is evaluated.

#### 5.6.1 *Coefficient of determination, $R^2$*

$R^2$, the coefficient of determination, is a goodness of fit measure indicating how much of the data conform to the hypothesized model. It indicates if the selected model is an acceptable description of the data and the data are not completely random. The $R^2$ has values in [0,1]. In general, the higher the $R^2$, the better the model fits the data. The average value of $R^2$ for the selected model is 0.8883 calculated using R Language. This means that 88.8% of the variability of the throughput around its mean is explained by the variability of the *improved attacking case*. Thus we can conclude that $y^{1/3} = a + b\ (\ln \mathrm{x})^3$ is a good model to represent our network.

#### 5.6.2 *t-test*

The values of the $a$ and $b$ coefficients for the selected model are presented in Table 6. T-test is used to evaluate the $b$ coefficient with a significant level of 5% using R Language. If the $p$ value is below 0.05 it indicates that the independent variable is useful to predict the dependent

**Table 5** Criteria for choosing a model: mean absolute percentage error (MAPE)

| Model | OA | | | | DA | | | | Average |
|---|---|---|---|---|---|---|---|---|---|
| | 09 | 18 | 27 | 36 | 09 | 18 | 27 | 36 | |
| $y^{1/3} = a + b\, x^{1/3}$ | 10.87 | 13.98 | 15.69 | 16.22 | 8.37 | 10.30 | 9.32 | 8.12 | 11.61 |
| $y^{1/4} = a + b\ (\ln \mathrm{x})^4$ | 15.39 | 15.78 | 15.95 | 15.61 | 7.85 | 10.57 | 9.49 | 8.26 | 12.36 |
| $\boldsymbol{y^{1/3} = a + b(\ln x)^3}$ | **12.92** | **13.90** | **14.68** | **14.61** | **8.24** | **10.33** | **9.31** | **8.04** | **11.50** |
| $y^{1/2} = a + b\ (\ln \mathrm{x})^2$ | 11.31 | 13.94 | 15.48 | 15.91 | 8.40 | 10.11 | 9.18 | 7.90 | 11.53 |

The bold row represents the model selected based on the lowest average MAPE

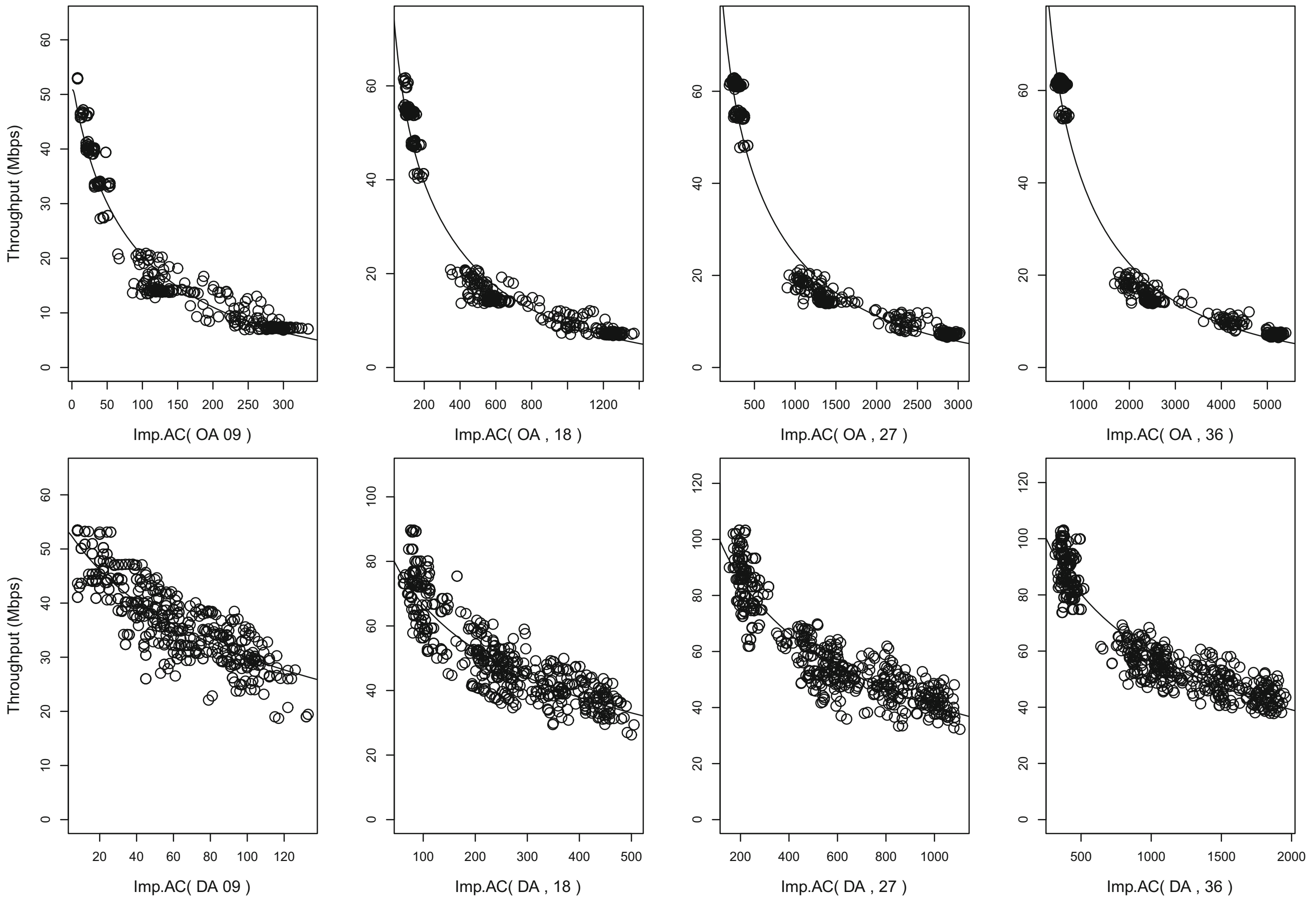


**Fig. 6** Scatter plot for the selected model $y^{1/3} = a + b\ (\ln\ x)^3$ with fitting for OA and DA, STA:[09, 18, 27, 36]

**Table 6** Coefficient values for the selected model, $y^{1/3} = a + b\ (\ln\ x)^3$

| Coefficient | OA | | | | DA | | | |
|---|---|---|---|---|---|---|---|---|
| | 09 | 18 | 27 | 36 | 09 | 18 | 27 | 36 |
| *a* | 3.7040 | 4.4872 | 4.9262 | 5.1761 | 3.7679 | 4.6302 | 5.2126 | 5.4215 |
| *b* | − 0.0099 | − 0.0073 | − 0.0061 | − 0.0054 | − 0.0068 | − 0.0059 | − 0.0054 | − 0.0046 |

variable. In our case the average p-values are below 0.001 for all the antenna type and STAs.

### 5.6.3 Root mean square error

The Root Mean Square Error (RMSE) for the selected model is calculated using Eq. 27 and the results are presented in Table 7. The average RMSE for the selected model is 4.64 Mbps.

$$RMSE = \sqrt{mean(e_i^2)} \tag{27}$$

We have shown in this section that our selected model $y^{1/3} = a + b\ (\ln\ x)^3$, is a good fit and a statistically strong relationship present between the dependent and independent variables.

## 6 Predicting using $y^{1/3} = a + b\ (\ln\ x)^3$ and discussion

In this section the $y^{1/3} = a + b\ (\ln\ x)^3$ model is used: (a) to predict the throughput of a similar random network; (b) in deciding the best configuration to use for a specific topology to achieve the maximum throughput.

**Table 7** RMSE values for the selected model, $y^{1/3} = a + b\ (\ln x)^3$

| OA | | | | DA | | | | Average |
|---|---|---|---|---|---|---|---|---|
| 09 | 18 | 27 | 36 | 09 | 18 | 27 | 36 | |
| 2.9458 | 3.3897 | 3.6559 | 3.6190 | 3.7181 | 6.4623 | 7.0482 | 6.2728 | 4.64 |

### 6.1 Case: predicting a random network

A network operator planning to setup a similar network can use the selected $y^{1/3} = a + b\ (\ln x)^3$ model to predict the throughput of his network knowing the *improved attacking case* value. To validate this, 20 random topologies are used with the random configuration from the various number of STAs, type of antenna, channel configuration, and transmit power strategies options presented in Sects. 4.2 and 4.3. These replicate the scenario where 20 different network operators using the selected model to predict the throughput of their planned network that comes with various configuration. For each of the topology the *improved attacking case*, its predicted throughput using the selected model, and measured throughput is shown in Table 8.

The predicted and measured throughputs are shown on the scatter plot in Fig. 7 for all the 20 random topologies. The predicted throughput is labeled using a diamond symbol which lie on the predicted line of the selected model. It's 95% confidence interval is presented in vertical bar and the measured throughput value is shown using a square symbol. Out of the 20 random topologies, 19 of them are within the 95% confidence (i.e. 95% of the cases).

The error of prediction square, $e_i^2$ verses the predicted throughput chart as shown in Fig. 8 is plotted for the random network not within the 95% confidence interval. It is found that the $e_i^2$ for the measured throughput out of the 95% confidence interval, shown using square symbol, is far away from the cloud of points showing to be an outlier. The MAPE for all the 20 random topologies, calculated using Eq. 26, is 12.56%, but when the one random network lying outside the 95% confidence interval is removed, the MAPE is improved to 11.80%, being both on the same scale of the model's MAPE which was at 11.5%. The RMSE for the 20 random topologies is 5.24 Mbps, and 4.14 Mbps when the random network is removed, both within the same scale of the prediction model's RMSE.

**Table 8** Prediction results for 20 random topology with random configuration using $y^{1/3} = a + b\ (\ln x)^3$ Model

| Random topology | Antenna | STA | Transmit power | Channel | Improved Attacking case | Measured Throughput (Mbps) | Predicted Throughput (Mbps) |
|---|---|---|---|---|---|---|---|
| 1 | OA | 9 | Default | SC | 308 | 7.21 | 6.19 |
| 2 | OA | 9 | Default | TC | 126 | 14.01 | 17.19 |
| 3 | OA | 9 | MP-NT | SC | 302 | 7.13 | 6.38 |
| 4 | OA | 18 | Default | SC | 1296 | 7.43 | 5.94 |
| 5 | OA | 18 | MP-ND | NC | 112 | 60.72 | 51.64 |
| 6 | OA | 27 | Default | TC | 1362 | 14.00 | 18.09 |
| 7 | OA | 27 | MP-NT | SC | 2810 | 7.67 | 6.43 |
| 8 | OA | 27 | MP-ND | TC | 1106 | 19.58 | 22.39 |
| 9 | OA | 36 | MP-NT | NC | 503 | 62.10 | 58.68 |
| 10 | DA | 9 | Default | SC | 104 | 31.73 | 29.48 |
| 11 | DA | 9 | MP-ND | TC | 41 | 36.47 | 40.04 |
| 12 | DA | 18 | MP-ND | NC | 82 | 55.22 | 70.15 |
| 13 | DA | 18 | MP-IN | SC | 353 | 44.99 | 40.55 |
| 14 | DA | 18 | MP-ND | NC | 74 | 78.01 | 71.93 |
| 15 | DA | 27 | Default | TC | 564 | 57.47 | 56.58 |
| 16 | DA | 27 | MP-IN | SC | 818 | 50.22 | 45.98 |
| 17 | DA | 36 | MP-ND | TC | 931 | 57.19 | 61.49 |
| 18 | DA | 36 | Default | SC | 1853 | 47.37 | 41.24 |
| 19 | DA | 36 | MP-IN | SC | 1488 | 45.38 | 47.52 |
| 20 | DA | 36 | MP-NT | NC | 454 | 89.29 | 83.16 |

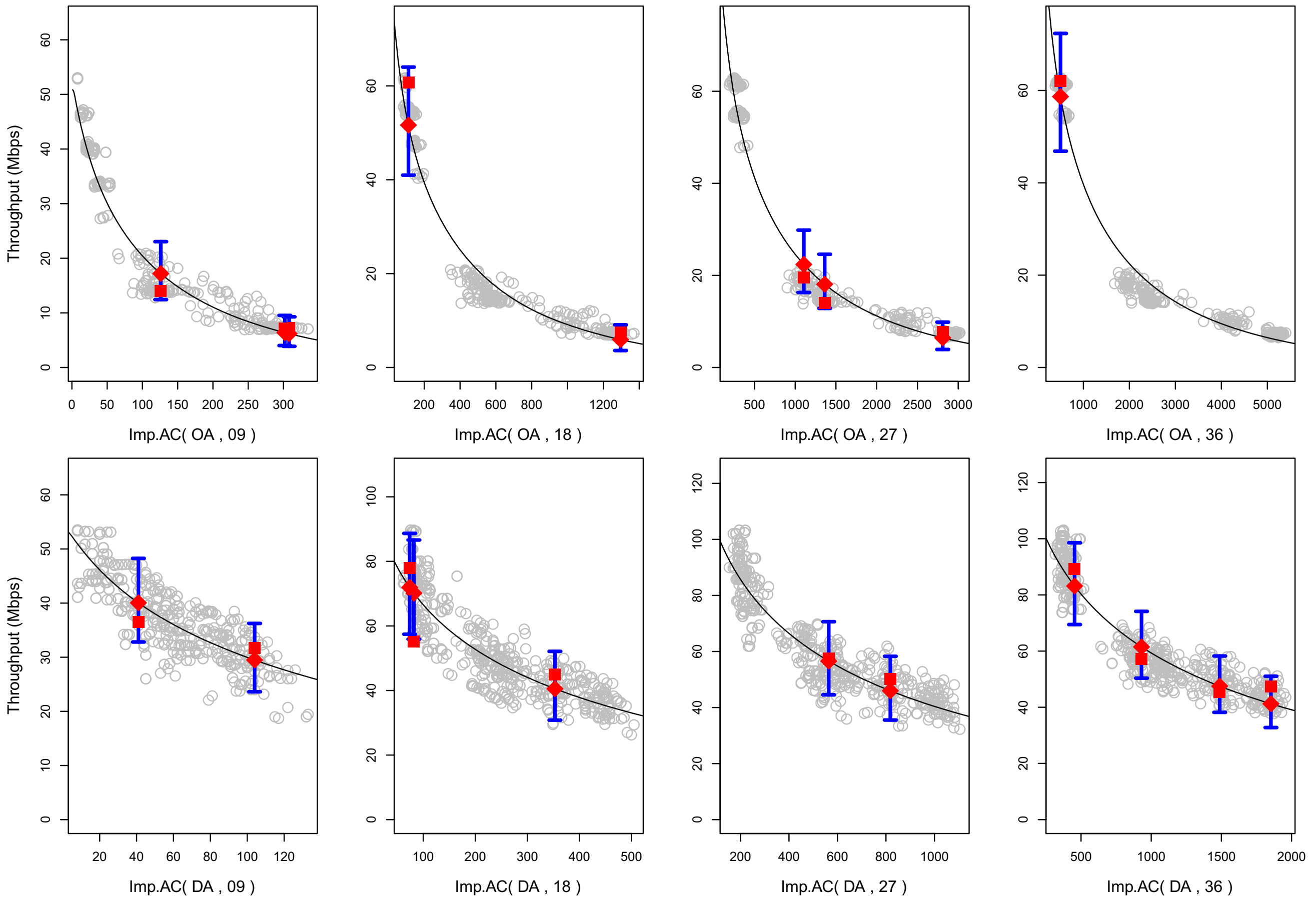


**Fig. 7** Predicting throughput for 20 random topologies using the selected model

## 6.2 Case: best configuration to use for a specific topology

A network operator may have difficulty in deciding the type of antenna to use, whether or not to employ a power control algorithm and if yes if he should be configuring the transmission power per network, per node or per interface in order to get the best performance in terms of throughput for his network. The need to get the highest throughput is limited by other external parameters such as financial constraints, or the availability of the number of interference free channel. The problem now became complex and multi-pronged. In this section, the selected model is used to assist a network operator to decide which configuration is best to be used for his network.

To validate this, a network operator is assumed to be intending to deploy a video surveillance network with 36 STAs as shown in Fig. 9. The position of the AP and STAs depends very much on the region the network operator aims to cover, obstacles that exist in the region (not shown on the figure), and availability of infrastructures such as buildings, lamp post, sign post to install the camera. For simplicity only the position of the AP, STAs and the flow of data are shown in Fig. 9. The network operator has several options to consider and now it has to decide which configuration would be optimal among a set of alternatives, considering the following options: (a) OA, SC and Default Power; (b) DA, SC, and Default Power; (c) OA, TC, and Default Power and (d) DA, SC and MP-IN.

Setup a) is the default setup for a typical WLAN. Having many STAs in the network, as shown in Fig. 9, induces high interference. The network operator may have doubt if the network can indeed support video surveillance. Using DA as in Setup b) would be a solution, but additional investment would be needed considering there are 45 nodes (36 STA + 9 AP) in total in the network with each nodes using 4 DA antennas. The capital expenditure is higher than in Setup a). Another option is to keep the OA but explore using 2 channels as in Setup c). While this could be a cheaper solution, scarcity of additional interference free channel might be a problem in some locations. If the network operator had indeed decided to invest in DA, he might as well take into advantage to use minimum transmit power algorithm to increase the throughput. It would be good for future planning in case there are needs to add

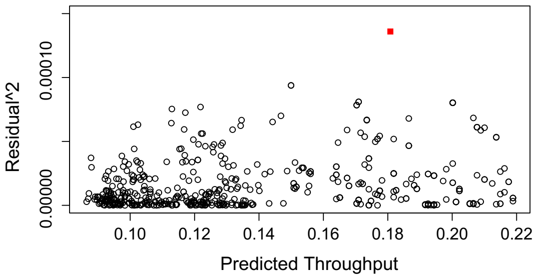


**Fig. 8** Outlier cases DA 18 111

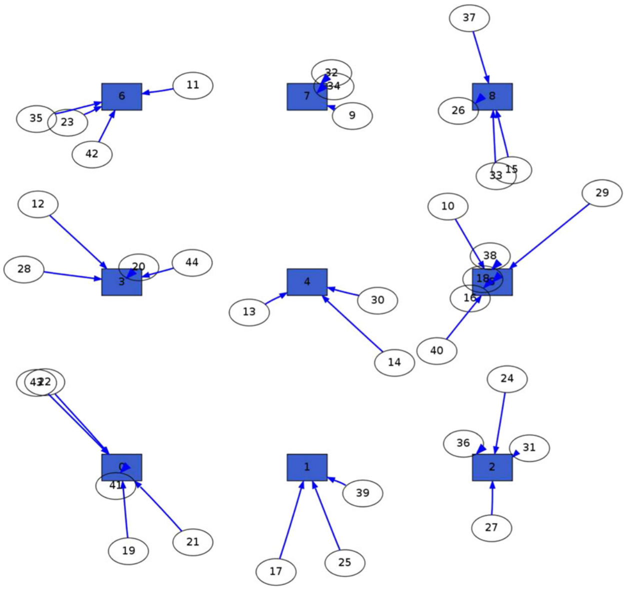


**Fig. 9** The topology to decide the best configuration for maximum throughput

more cameras to the network. These are possible dilemma of a network operator.

For each setup the *improved attacking case*, its predicted throughput using the proposed model, and the throughput measured by simulation is shown in Table 9.

Referring to Table 9, Setup a) has the highest interference in the network with a value of 5319 for the *improved attacking case* metric followed by Setup c) and b) at 2413 and 1816 respectively. Setup d) has the least amount of interference as the *improved attacking case* is only 1413. Using the selected model, Setup d) is predicted to have the best throughput, i.e., approximately 49 Mbps. This is later confirmed by the measured result where the setup was having around 53 Mbps of throughput. The network operator could choose Setup d) that proposes to use DA with MP-IN in case he has interference free channel constraints. This setup was predicted to offer 8.5 times more throughput than Setup a). The next best option is Setup b) where no power control is implemented with around 42 Mbps of throughout predicted. In case constraints due to finance is much higher than the availability of interference free channel then the network operator could opt for Setup c) as no additional investment needed for DA. Though the predicted throughout for this setup is around 18 Mbps, it is predicted to be 3.2 times much higher than Setup a). This might be sufficient to support a video surveillance network with 36 STAs.

## 7 Conclusion

In this paper we have studied the relationship between the *improved attacking case* metric and throughput for IEEE 802.11 based network using DA. After considering 1936 models, it is found that the $y^{1/3} = a + b\ (\ln x)^3$ model best fits our dataset. The model also fulfills the homoscedasticity and normality assumptions of the linear regression's residual errors. We used this relationship: (a) to predict the performance of similar network and found the MAPE is 12.56% for the prediction of 20 random topologies; (b) to decide the best configuration a network operator could use to plan their network.

As the *improved attacking case* is proven to have a strong relationship with throughput, that is when the *improved attacking case* reduces the throughput increases, our future work would involve to find algorithms aimed to reduce transmit power or the interference in network to maximize the throughput. In this work, the number of STA is rather finite, i.e., limited to 9, 18, 27, 36. The *a* and *b* coefficients presented in Table 6 shows increasing in value for OA and DA when the number of STA's increases.

**Table 9** Prediction results for setups a–d using $y^{1/3} = a + b\ (\ln x)^3$ Model

| Setup | Antenna | Transmit power | Channel | Improved Attacking case | Measured Throughput (Mbps) | Predicted Throughput (Mbps) |
|---|---|---|---|---|---|---|
| a | OA | Default | SC | 5319 | 7.58 | 5.74 |
| b | DA | Default | SC | 1816 | 46.87 | 41.81 |
| c | OA | Default | TC | 2413 | 14.56 | 18.46 |
| d | DA | MP-IN | SC | 1413 | 52.99 | 49.03 |

This relation could be studied and used to predict the throughput for STAs not covered in this work.

We have considered wireless network with homogenous: (a) nodes, where all of them using either OA or DA; (b) technology, where IEEE 802.11 is assumed to be the only technology operating in the 2.4GHz ISM band; (c) IEEE 802.11 release, where IEEE 802.11b with 11 Mbps data rate is used as a representative release of the IEEE 802.11 protocol. There could be networks that are heterogeneous where it consists of a combination of nodes that use OA and DA, different technology other than IEE 802.11 protocol co-existing in the same 2.4 GHz ISM band and accommodate different releases of IEEE 802.11 protocol as APs/STAs are usually backwards compatible. Additionally, our work can be further extended to consider the effect of co-channel interference and also routing which is useful for Wireless Mesh Network scenarios. Considering these setups, the relationship between *improved attacking case*, which may be improved further, and throughput can be studied. We have considered a line-of-sight scenario to ease the graph modeling aspect. Our work can be extended to non line-of-sight scenario by building the graphs using the received power information at the nodes. The information may be shared among the nodes either by taking advantage of the information in the packets exchanged in the network or by creating additional control packets specific to carry this information. However, these remains as our future work.

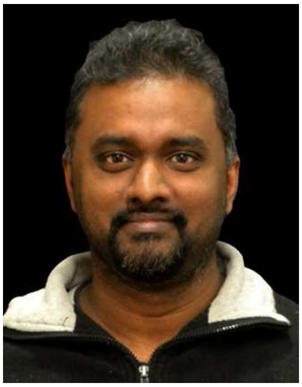
We aim to build a network planning software which can aid to predict the throughput of a similar IEEE 802.11 based wireless network. A user is expected to input the network topology into the software, then choose the configuration from the pool of antenna type, power control strategy, and number of channels. The software can automatically calculate the *improved attacking case* and predict its throughput. The software can also assist a network operator to choose the optimal configuration for his planned network based on constraints such as finance, the availability of interference free channels and physical obstructions.

**Acknowledgements** The authors would like to thank the Fundação para a Ciência e a Tecnologia (FCT) of Ministério da Ciência, Tecnologia e Ensino Superior (MCTES), Portugal for supporting this work through grant SFRH/BD/43744/2008 and PTDC/EEA-TEL/120176/2010.

## References

1. Liu, Q., Jiang, X., & Qiu, X. (2017). The effects of topology on throughput capacity of large scale wireless networks. *Journal of Information*,. https://doi.org/10.3390/info8010032.
2. Xu, Y., Liu, J., Shen, Y., Li, X., & Jiang, X. (2017). On throughput capacity of large-scale ad hoc networks with realistic buffer constraint. *Wireless Networks*, *23*(1), 193–204. https://doi.org/10.1007/s11276-015-1146-2.
3. Hua, Y., Huang, Y., & Garcia-Luna-Aceves, J. J. (2006). Maximizing the throughput of large ad hoc wireless networks. *IEEE Signal Processing Magazine*, *23*(5), 84–94. https://doi.org/10.1109/MSP.2006.1708415.
4. Haenggi, M., & Ganti, R. K. (2009). Interference in large wireless networks. *Foundations and Trends in Networking*, *3*(2), 127–248. https://doi.org/10.1561/1300000015.
5. IEEE Standard for Info. Technology–Telecommunications and Info. Exchange between Systems LAN and MAN–Specific requirements Part 11: Wireless LAN MAC and Physical Layer (PHY) Specifications, IEEE Std 802.11-2012 (Revision of IEEE Std 802.11-2007), 2012 (pp. 1–2793). https://doi.org/10.1109/IEEESTD.2012.6178212.
6. Adeyeye, M., & Gardner-Stephen, P. (2011). The Village Telco project: A reliable and practical wireless mesh telephony infra. *EURASIP Journal on Wireless Communication and Network*, *1*, 78. https://doi.org/10.1186/1687-1499-2011-78.
7. Irwin, D., Sharma, N., Shenoy, P., & Zink, M. (2011). Towards a virtualized sensing environment. In *Testbeds and Research Infrastructures. Development of NTs and Communities*, Springer, Berlin (Vol. 46, pp. 133-142). https://doi.org/10.1007/978-3-642-17851-1_10.
8. Anderson, E., Phillips, C., Yee, G., Sicker, D., & Grunwald, D. (2011). Challenges in deploying steerable wireless testbeds. In *Testbeds and Research Infrastructures. Development of NTs and Communities*, Springer, Berlin (Vol. 46, pp. 231–240). https://doi.org/10.1007/978-3-642-17851-1_19.
9. Kapnadak, V., Senel, M., & Coyle, E. J. (2011). Low-complexity, distributed characterization of interferers in wireless networks. *International Journal of Distributed Sensor Networks*, *2011*, 17. https://doi.org/10.1155/2011/980953.
10. Karrer, R., Pescape, A., & Huehn, T. (2008). Challenges in second-generation wireless mesh networks. *EURASIP Journal on Wireless Communications and Networking*, *1*, 2008. https://doi.org/10.1155/2008/274790.
11. Kandasamy, S., Campos, R., Morla, R., & Ricardo, M. (2010). Using directional antennas on stub WMN: Impact on throughput, delay, and fairness. In *Proceeding of the 19th International Conference on Computer Communications and NTs (ICCCN)* (pp. 1–6). https://doi.org/10.1109/ICCCN.2010.5560027.
12. Kandasamy, S., Campos, R., Morla, R., & Ricardo, M. (2009). Improving the performance of IEEE 802.11s NTs using directional antennas over multi-radio/multi-channel implementation - the research challenges. In *Proceeding of the 4th Doctoral Symposium on Infomatics Engineering (DSIE)* (pp. 1–12).
13. Kandasamy, S., Morla, R., & Ricardo, M. (2016). Power interference modeling for CSMA/CA based networks using directional antennas. *Elsevier's Journal of Computer Communications*, *86*, 86–98. https://doi.org/10.1016/j.comcom.2016.01.012.
14. Vlavianos, A., Law, L., Broustis, I., Krishnamurthy, S., & Faloutsos, M. (2008). Assessing link quality in IEEE 802.11 wireless NTs: Which is the right metric?. In *Personal, Indoor and Mobile Radio Communications 2008. IEEE 19th International Symposium on* (pp. 1–6). https://doi.org/10.1109/PIMRC.2008.4699837.
15. Ho, I.-H., & Liew, S. C. (2007). Impact of power control on performance of IEEE 802.11 wireless networks. *IEEE Transactions on Mobile Computing*, *6*(11), 1245–1258. https://doi.org/10.1109/TMC.2007.1045.
16. Papadopouli, M., Raftopoulos, E., & Shen, H. (2006). Evaluation of short-term traffic forecasting algorithms in wireless network. In *Next Generation Internet Design and Engineering 2nd Conference on* (pp. 8). https://doi.org/10.1109/NGI.2006.1678229.

17. Chen, C., Pei, Q., & Ning, L. (2009). Forecasting 802.11 traffic using seasonal ARIMA model. In *Computer Science-Technology and Applications, 2009. IFCSTA '09. International Forum on*, (Vol. 2, pp. 347–350). https://doi.org/10.1109/IFCSTA.2009.207.
18. Kong, Y., Liu, X.-W., & Zhang, S. (2009). Minimax probability machine regression for wireless traffic short term forecasting. In *Cognitive Wireless System., 1st UK-India International Workshop on* (pp. 1–5). https://doi.org/10.1109/UKIWCWS.2009.5749407.
19. Na, C., Chen, J., & Rappaport, T. (2006). Measured traffic statistics and throughput of IEEE 802.11b public WLAN hotspots with three different applications. *IEEE Transactions on Wireless Communications*, *5*(11), 3296–3305. https://doi.org/10.1109/TWC.2006.05043.
20. Dao, N., & Malaney, R. (2007). Throughput performance of saturated 802.11g NTs. In *Wireless Broadband and Ultra Wideband Communication, 2007. AusWireless 2007. The 2nd International Conferance on* (pp. 31–31). https://doi.org/10.1109/AUSWIRELESS.2007.82.
21. Bruno, R., Conti, M., & Gregori, E. (2009). Average-value analysis of 802.11 WLANs with persistent TCP flows. *IEEE Communications Letters*, *13*(4), 218–220. https://doi.org/10.1109/LCOMM.2009.080653.
22. Dely, P., Kassler, A., & Sivchenko, D. (2010). Theoretical and experimental analysis of the channel busy fraction in IEEE 802.11. In *Future Network and Mobile Summit* (pp. 1–9).
23. Liangrui, T., & Wenjin, W. (2011). An Improved Algorithm based on NT Load Prediction for 802.11 DCF. In *Natural Computation (ICNC), 2011 7th International Conferance on* (Vol. 3, pp. 1466–1469). https://doi.org/10.1109/ICNC.2011.6022299.
24. Jiang, L. B., & Liew, S. C. (2008). Improving throughput and fairness by reducing exposed and hidden nodes in 802.11 networks. *IEEE Transactions on Mobile Computing*, *7*(1), 34–49. https://doi.org/10.1109/TMC.2007.1070.
25. de Carvalho, C., Gomes, D., de Souza, J., & Agoulmine, N. (2011). Multiple linear regression to improve prediction accuracy in WSN data reduction. In *Network Operations and Management Symposium (LANOMS), 2011, 7th Latin American, 2011* (pp. 1–8). https://doi.org/10.1109/LANOMS.2011.6102268
26. R Core Team, R. (2013). A Language and Environment for Stats. Comp., R Foundation for Statistical Comp., Vienna, Austria, ISBN 3-900051-07-0.
27. Montgomery, D., Jennings, C., & Kulahci, M. (2011). *Introduction to time series analysis and forecasting, Wiley series in probability and statistics*. Hoboken: Wiley. ISBN 9-781118211-50-2.
28. Gujarati, D. N. (2004). *Basic econometrics*. New York: The McGraw-Hill Companies. ISBN 9-780070597-93-8.
29. Ariza, C., Rugeles, L., Saavedra, D., & Guaitero, B. (2013). Measuring innovation in agricultural firms: A methodological approach. *Electronic Journal of Knowledge Management*, *11*(3), 185–198.
30. Stonewall, A. J., & Bragg, H. M. (2012). Suspended-sediment characteristics for the Johnson Creek basin, Oregon, water years 2007–10. *Scientific Investigations Report*, *2012*, 1–32.
31. R. Kabacoff, R in Action: Data Analysis and Graphics with R, Manning Pubs Co Series, Manning, ISBN 9-781935182-39-9 (2011).
32. Tukey, J. W. (1977). *Exploratory data analysis*. Boston: Addison-Wesley.
33. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data mining: Practical machine learning tools and techniques* (4th ed.). Burlington: Morgan Kaufmann Publication. ISBN: 0128043571.
34. Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistical Survey*, *4*, 40–79. https://doi.org/10.1214/09-SS054.



**Saravanan Kandasamy** received the B.Eng. (2000) in Electronics majoring in Computer from Multimedia University, Malaysia, M.Sc. (2004) in Communications and Network Engineering from University Putra Malaysia and Ph.D. (Telecommunications Engineering) from University of Porto, Portugal. He is currently a post doctoral fellow in the Centre for Telecommunications and Multimedia of INESC TEC (www.inesctec.pt). His research interest include directional antenna, radio resource management, transmission power control and statistical modeling for IEEE 802.11 based wireless networks.

**Ricardo Morla** is an Assistant Professor with the Electrical and Computer Engineering Department, and a principal investigator with INESC Porto, at the Faculty of Engineering of the University of Porto. His research interests are in the field of modeling and management of IT systems with an emphasis on probabilistic and machine learning approaches applied to networks and ambient intelligence. Ricardo graduated from U.Porto in Electrical and Computer Engineering and holds a Ph.D. in Computing from Lancaster University. He was a lecturer and post-doc at UC Irvine in 2007, and a visiting faculty at Carnegie Mellon University in 2010 under the CMU-Portugal program.

**Patrícia Ramos** obtained her licenciate degree in Applied Mathematics Field of Computer Science from Faculty of Sciences of University of Porto in 1993, Master degree in Electrical and Computer Engineering-Field of Systems from Faculty of Engineer of University of Porto (FEUP) in 1996 and Ph.D. in Engineering Sciences from FEUP in 2005. She is Assistant Professor at the Department of Mathematics of Institute of Accountancy and Administration of Porto, Polytechnic School of Porto. Her research activity in forecasting and predictive analytics is carried out in the Centre for Enterprise Systems Engineering at INESC Porto, in the areas of supply chain forecasting for retailing and manufacturing, computer-intensive forecasting methods, forecasting model selection, model evaluation, econometrics and data mining.

**Manuel Ricardo** received a Licenciatura (1988), M.Sc. (1992), and Ph.D. (2000) degrees in Electrical and Computer Engineering from Porto University. Currently, he is an associate professor at the Faculty of Engineering of University of Porto, where he gives courses in mobile communications and computer networks. He also coordinates Centre for Telecommunications and Multimedia of INESC TEC (www.inesctec.pt).