Cell-Level Modeling of IEEE 802.11 WLANs

Manoj Panda[a,*], Anurag Kumar[b]

[a]*Swinburne University of Technology, John Street, Hawthorn, 3122 Victoria, Australia.*
[b]*ECE Department, IISc Bangalore – 560012.*

Abstract

We develop a scalable *cell-level* analytical model for multi-cell infrastructure IEEE 802.11 WLANs under a so-called Pairwise Binary Dependence (PBD) condition. The PBD condition is a geometric property under which the relative locations of the nodes inside a cell do not matter and the network is free of *hidden nodes*. For a given number of cells, the computational complexity of our cell-level model remains constant even if the number of nodes per cell increases. For the cases of saturated nodes and TCP-controlled long-file downloads, we provide accurate predictions of cell throughputs. Similar to Bonald et al (Sigmetrics, 2008), we model a multi-cell WLAN under short-file downloads as "a network of processor-sharing queues with state-dependent service rates." Whereas the state-dependent service rates proposed by Bonald et al are based only on the *number* of contending neighbors, we employ state-dependent service rates that incorporate the impact of the overall *topology* of the network. We propose an *effective service rate approximation* technique and obtain good approximations for the *mean flow transfer delay* in each cell. For TCP-controlled downloads where the Access Points (APs) transmit for a much larger fraction of time than the stations (STAs), we consider the case when the APs can sense all the nodes in the neighboring cell, but $\approx 50\%$ of the STAs in each cell can sense only a subset of STAs in the other cell. Our cell-level model can predict the throughputs quite accurately in this case as well even though the PBD condition does not strictly hold.

Keywords: IEEE 802.11; multi-cell WLAN; analytical modeling; throughput and delay; fixed point

1. Introduction

With widespread deployment of WiFi networks (or, more formally, IEEE 802.11 networks) in office buildings, university campuses, homes, hotels, airports and other public places, it has become very important to understand the performance of Wireless Local Area Networks (WLANs) that are based on the IEEE 802.11 standard, and also to know how to effectively design, deploy and manage them.

The IEEE 802.11 standard [2] provides two modes of operation, namely, the *ad hoc* mode and the *infrastructure* mode. Commercial and enterprise WLANs usually operate in the infrastructure mode. An infrastructure WLAN contains one or more Access Points (APs) which provide service to a set of users or client stations (STAs). Every STA in the WLAN associates iself with exactly one AP. Each AP, along with its associated STAs, constitutes a so-called *cell*. Each cell operates on a specific channel. Cells that operate on the same channel are called *co-channel*. We call a WLAN containing multiple APs a *multi-cell* WLAN.

In a multi-cell infrastructure WLAN, the APs are usually connected among themselves and to the Internet by a high-speed *wireline* Local Area Network (LAN), e.g., a Gigabit Ethernet. The STAs, on the other hand, access the Internet *only through their respective APs*. Thus, the 802.11 MAC protocol is employed only for *single-hop intra-cell* frame exchanges between an AP and its associated STAs

---

☆This paper is an extended version of our earlier work [1]. In this paper, we extend our analytical model in [1] to TCP-controlled short-file downloads, characterize the associated service process, and provide approximations for the mean flow-transfer delays. We also demonstrate the applicability of our analytical model when the "PBD condition" does not strictly hold.

☆☆Much of the research work for this paper was carried out when the first author was affiliated with the Indian Institute of Science Bangalore.

*Corresponding Author

Email addresses: mpanda@swin.edu.au (Manoj Panda), anurag@ece.iisc.ernet.in (Anurag Kumar)

Figure 1: A multi-cell infrastructure WLAN: The 802.11 MAC protocol is employed only for single-hop intra-cell frame exchanges within the cells (shown by dashed ovals). A high-speed wireline LAN connects the APs and provides access to the Internet.
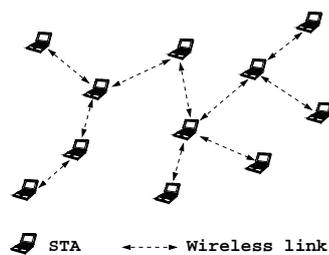
Figure 2: A multi-hop ad hoc WLAN.

(and not among the APs and/or STAs belonging to different cells). Figure 1 depicts an example of such a multi-cell WLAN.

This paper is concerned with analytical modeling of infrastructure WLANs (such as in Figure 1) that are based on the Distributed Coordination Function (DCF) Medium Access Control (MAC) protocol as defined in the IEEE 802.11 standard [2]. Analytical modeling can provide important insights to effectively design, deploy and manage the WLANs. However, accurate analytical modeling of multi-cell WLANs is a challenging problem. Nodes (i.e. AP or STA) in two closely located co-channel cells can suppress each other's transmissions via carrier sensing and interfere with each other's receptions causing packet losses. Thus, the activities of the nodes in proximal co-channel cells are essentially coupled, which makes the analytical modeling difficult.

1.1. Literature Survey

The seminal analytical model for *single cell* WLANs was developed in [3], and later general-

ized in [4]. In a single cell, nodes can sense and decode each other's transmissions. Thus, nodes in a single cell have the same *global* view of the activities on the common medium. Nodes in a multi-cell WLAN, however, can have different *local* views of the network activity around themselves, and their own activity is determined by this local view.

In the context of multi-hop ad hoc networks (see Figure 2), *node-* and *link-level* models have been proposed to capture the *local* characteristics of network activities [5, 6, 7]. However, for networks of realistic size, a node- or link-level model is intractable, since its complexity increases exponentially with the number of nodes or links [8].

Since a node- or link-level model is often intractable and provides little insight, several simplifying assumptions have been made in order to gain insights. A common simplification is to ignore collisions due to *hidden nodes* by assuming that hidden nodes are suppressed either via RTS/CTS handshaking [9, 10], or via a so-called "hidden-node-free design" [11, 12, 13, 14]. Another simplification is to assume an infinite number of nodes placed according to either some regular topology [9] or a regular point process [15]. The performance analysis of a single-AP WLAN in presence of hidden nodes has been reported in [16, 17, 18].

TCP-controlled data transfers are usually classified into two types: (i) *long-lived* flows (e.g., file transfer), and (ii) *short-lived* flows (e.g., web browsing). The case of long-lived flows in a single isolated cell has been analyzed in [19, 20]. A single isolated cell with short-lived flows was modeled as "a processor-sharing queue with state-dependent service rates" in [21, 22]. The seminal papers by Bonald et al. provide the most comprehensive treatment of multi-cell networks with short-lived flows, both for cellular data networks [23] as well as for 802.11-based WLANs [10]. In [10] the authors model a multi-cell infrastructure WLAN under short-lived downloads as a "network of *multi-class* processor-sharing queues with state-dependent service rates."

1.2. Our Contributions

To accurately model a multi-cell infrastructure WLAN, one needs to account for the node- and link-level interactions, as in the case of multi-hop ad hoc networks. However, comparing Figures 1 and 2, we observe the following key differences:

- The ad hoc WLAN does not possess any specific stucture (topology). However, the infrastructure WLAN has a nice *cellular* structure. Let R denote the *cell radius* (i.e., the maximum distance between an AP and its associated STAs) and D denote the distance between two nearest co-channel APs. For communication at high Physical layer (PHY) rate to be possible between an AP and its associated STAs, R should be small. However, with multiple non-overlapping channels, D could be much larger than R. Thus, the APs and STAs that operate on the same channel look like a set of clusters (or cells) centered at the APs and well separated from each other (for example, see Figures 3(a)-3(d)).

- If hidden nodes are due to physical distances and not due to obstacles, then a WLAN can be designed to be hidden-node-free by tuning the carrier sensing range [11, 12, 14]. The hidden-node-free condition in a cellular infrastructure WLAN, however, is different from that in an ad-hoc network. A hidden node scenario depends on the relative positions of an interferer w.r.t. a transmitter-receiver pair, and in a cellular infrastructure WLAN, a transmitter-receiver pair must belong to the same cell. If the carrier sensing range is at least twice the cell radius, then a hidden node must belong to another cell.

The above observations motivate us to take a different perspective on the analytical modeling of multi-cell infrastructure WLANs. Specifically, we exploit the cellular structure of multi-cell infrastructure WLANs by treating a cell as a single entity and develop a *cell-level* analytical model.

We make the following novel contributions:

(1) We identify a geometric property, which we call the *Pairwise Binary Dependence* (PBD) condition (see A1 in Section 2), under which we could develop a cell-level model. Our cell-level model is scalable since its computational complexity increases with the number of cells rather than nodes or links. In fact, its complexity remains constant as the number of nodes per cell increases (see Section 3.3).

(2) We develop analytical models with saturated nodes as well as for TCP-controlled long- and short-file downloads (Sections 3 and 4). For the case of saturated nodes and TCP-controlled long-lived downloads (resp. short-lived downloads), we predict throughputs (resp. flow transfer delays) and illustrate the accuracy of our analysis (Section 5). The equations for collision probabilities, throughputs and mean flow transfer delays that we derive are new.

(3) The PBD condition, if holds, essentially implies that the multi-cell infrastructure WLAN is free of hidden nodes (Section 2). We develop our analytical model in a hidden-node-free setting as in [9, 10, 13]. However, we demonstrate that our model can be applied also when the APs can sense all the nodes in the neighboring cell, but ≈50% of the STAs in each cell can sense only a subset of nodes in the other cell so that the PBD condition does not strictly hold (Section 5.5).

(4) In [6, 9, 13], valuable insights have been obtained by taking the *access intensity*[1] to infinity. One such insight is:

I_1 As the access intensity goes to infinity, only those nodes that belong to some *maximum independent set*[2] obtain non-zero throughputs [6, 9, 13].

We confirm Insight (I_1) at the cell-level. Another insight, which is a simple consequence of Insight (I_1), is:

I_2 As the access intensity goes to infinity, the aggregate network throughput is maximized.

Insight (I_2) has been recently established by [9] in the context of an infinite linear chain of saturated nodes. We establish it, under the PBD condition, for arbitrary cell topologies for saturated nodes and also for TCP-controlled long-lived downloads (see Theorem 3.1). Moreover, we provide a third insight:

I_3 Realistic access intensities with default backoff parameter settings and sufficiently large packet payloads (e.g., ≥ 500 bytes) are indeed *large* so that the values of the cell throughputs in such situations are very close to their values with infinite access intensities (Section 3.4).

---

[1] This term was coined by [9] and is defined in (5).

[2] A maximum independent set of a graph G is an independent set of G with maximum cardinality. In contrast, a *maximal* independent set of G is an independent set of G that is not a proper subset of any other independent set of G. Every maximum set is also a maximal set, but the converse is not true.

(5) For TCP-controlled short-file downloads, we improve the service model of [10]. The service model in [10] is based on the assumption of *"synchronous and slotted"* evolution of network activities with the service rates depending only on the *number* of contending neighbors. However, inter-cell blocking due to carrier sensing is "asynchronous" (see Section 3.1). The service rates in our model capture this asynchrony and the service rates in our model depend on the overall *topology* of the contending APs. Taking an example of a three-cell network, we show that (i) the service rates proposed in [10] can lead to inaccurate prediction of delays (Figure 20), but (ii) significantly more accurate results can be obtained with our improved service model (Figure 21).

1.3. Outline of the paper

The remainder of the paper is structured as follows. In Section 2, we provide our network model and discuss our assumptions. In Section 3, we first develop our analytical model with saturated nodes, and then extend to the case of TCP-controlled long-file downloads (Section 3.2). In Section 4, we apply the insights obtained in Section 3.4 to extend to the case of TCP-controlled short-file downloads. We illustrate the accuracy of our analytical models in Section 5. The paper concludes in Section 6. A derivation of the equation for collision probability has been provided in the Appendix.

2. Network Model and Assumptions

Let R_{cs} denote the *carrier sensing range*, i.e., R_{cs} is the maximum distance up to which a transmitter can cause a "medium busy" condition at idle receivers [24, 12].

Definition 2.1. Two nodes are said to be *dependent* if the distance between them is less than R_{cs} **and** they operate on the same channel; otherwise, the two nodes are said to be *independent*. Two cells are said to be independent if every node in a cell is independent w.r.t. every node in the other cell; otherwise, the two cells are said to be dependent. Two dependent cells are said to be *completely dependent* if every node in a cell is dependent w.r.t. every node in the other cell. □

In the above definitions, we implicitly assume that (i) only non-overlapping channels are used, (ii) nodes can neither sense nor interfere (and be interfered) with transmissions on a different non-overlapping channel, and (iii) the carrier sensing range R_{cs} is a *sharp* boundary within which nodes sense and interfere (and be interfered) with each other's transmissions, and do not interact outside of it. We also make the following explicit assumptions about the network:

A0. Nodes in the same cell are dependent.

A1. **Pairwise Binary Dependence (PBD):** Any pair of cells is either independent or completely dependent.

A2. Simultaneous transmissions by independent transmitters are successfully received at their respective receivers.

A3. Simultaneous transmissions by dependent transmitters are always lost due to collision.

A4. Each cell contains a *fixed* number of *identical* STAs.

A5. Packet losses due to channel errors are negligible.

By A0, nodes in the same cell can sense and interfere (and be interfered) with each other's transmission. By the PBD condition (A1), nodes in the same cell sense and interfere (and be interfered) with the same set of nodes in the other cells. Since communication is always between an AP and one of its associated STAs (in the same cell), any transmitter-receiver pair can sense the same set of interferers. Thus, the PBD condition implies that the network is free of *hidden nodes*. As stated earlier in Section 1.1, assuming a hidden-node-free setting is a common simplification and applied in [9, 10, 13]. In Section 5.5 we demonstrate that our cell-level model, which we develop assuming the PBD condition, can also be applied in certain scenarios where hidden nodes are indeed present and the PBD condition does not strictly hold.

Simultaneous transmissions by nodes that are located within R_{cs} of each other (i.e., nodes that sense each other) lead to *synchronous* collisions, in which case, the colliding transmissions arrive at the receiver within a backoff slot duration [24]. Simultaneous transmissions by nodes that cannot sense each other's transmissions lead to *asynchronous* (or hidden node) collisions at a receiver. In our network model, there cannot be packet losses due to

asynchronous hidden node collisions (A2). However, there can be packet losses due to synchronous collisions, e.g., A0 and A3 imply that simultaneous transmissions in the same cell are always lost due to (synchronous) collisions.

3. An Analytical Model for Multi-Cell WLANs with Arbitrary Cell Topology

In this section, we develop a cell-level model for WLANs that satisfy A0-A5. We index the cells by positive integers $1, 2, \ldots, N$, where N denotes the number of cells. Let $\mathcal{N} = \{1, 2, \ldots, N\}$ denote the set of cells. Let M denote the number of non-overlapping channels, and let $\mathcal{C} = \{1, 2, \ldots, M\}$ denote the set of channels. Let $\boldsymbol{c} = (c_1, c_2, \ldots, c_N)$ denote the *channel assignment* where, $\forall i \in \mathcal{N}$, $c_i \in \mathcal{C}$ denotes the channel assigned to Cell-i.

We construct a *contention graph* $\mathcal{G}(\boldsymbol{c})$ as follows. Every cell in the network is represented by a vertex in $\mathcal{G}(\boldsymbol{c})$. There exists an edge between two vertices in $\mathcal{G}(\boldsymbol{c})$ if and only if the corresponding cells are completely dependent. Owing to the PBD condition, any two cells are either independent or completely dependent. Moreover, two cells are completely dependent if the distance between the respective APs is less than R_{cs} and the two cells operate on the same channel. Thus, given the locations of the APs and a channel assignment $\boldsymbol{c}$, $\mathcal{G}(\boldsymbol{c})$ can be constructed, and our model applies to any $\mathcal{G}(\boldsymbol{c})$. To simplify notations, we assume the channel assignment $\boldsymbol{c}$ to be given and fixed, and henceforth, we denote the contention graph $\mathcal{G}(\boldsymbol{c})$ (for the given channel assignment $\boldsymbol{c}$) by $\mathcal{G}$. Two cells whose corresponding vertices in the contention graph $\mathcal{G}$ are joined by an edge will be called *neighbors*. Let $\mathcal{N}_i$ ($\subset \mathcal{N}$) denote the set of neighboring cells of Cell-i ($i \in \mathcal{N}$). By definition, $i \notin \mathcal{N}_i$.

We first develop our model for saturated nodes (Section 3.1), and then extend to TCP-controlled long-lived downloads (Section 3.2). The case of TCP-controlled short-lived downloads is covered in Section 4. We develop a generic model and demonstrate the accuracy of our model by comparing with NS-2 simulations of a few example networks given in Figure 3(a)-3(d). Note that in Figure 3(a)-3(d), by a *"contention domain"* we mean a *maximal clique* of $\mathcal{G}$.
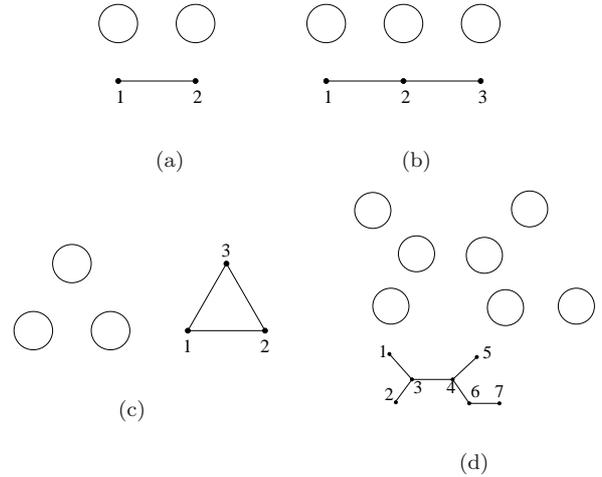
Figure 3: Examples of multi-cell systems: (a) two co-channel cells with a single contention domain, (b) three linearly placed co-channel cells with two contention domains, (c) three co-channel cells with a single contention domain, and (d) seven co-channel cells with an arbitrary cell topology and seven contention domains. The contention graphs have also been shown.

3.1. Modeling with Saturated Nodes and UDP Traffic

Consider the scenario where the nodes (i.e., the APs and the STAs) are *saturated* or infinitely backlogged, and are transferring packets to one or more nodes in the same cell over UDP. Let n_i, $n_i \geq 2$, denote the number of nodes in Cell-i. Thus, Cell-i consists of a saturated AP, AP-i, and $n_i - 1$ saturated STAs. The APs and their associated STAs exchange fixed size packets over UDP.

In general, nodes belonging to different cells can have different views of the network activity. Due to the PBD condition, however, nodes in the same cell have an identical view of the rest of the network. When one node senses the medium idle, so do the other nodes in the same cell, and we say that *the cell is sensing the medium idle*. Since the nodes are saturated, whenever a cell senses the medium idle, all the nodes in the cell decrement their backoff counters per idle backoff slot that elapses in the *local* medium of the cell, and we say that *the cell is in backoff*. If the nodes were not saturated, a node with an empty transmission queue would not have been counting down during the "medium idle" periods and the number of *contending* nodes would have been time-varying. With saturated nodes, however, nodes contend whenever they sense the medium idle, and the number of contending nodes

in each cell remains constant.

We say that *a cell is transmitting when one or more nodes in the cell are transmitting*. When a cell starts transmitting, its neighbors (i.e., neighboring cells) sense the transmission within one backoff slot and they defer medium access. We then say that the neighbors are *blocked* due to carrier sensing. When a cell is blocked, the backoff counters of all the nodes in the cell are *frozen*. Thus, a cell can be in one of three states: (a) **Transmitting** (when at least one node in the cell is transmitting), (b) **Backoff** (when every node in the cell is decrementing its backoff counter per idle backoff slot), or (c) **Blocked** (when the backoff counters of all the nodes in the cell are frozen due to transmissions in some neighboring cell(s)).

We emphasize that *the evolution of the network is partly synchronous and partly asynchronous.* Nodes in the same cell have an identical view of the rest of the network and they are synchronized. When two or more nodes in the same cell start transmitting together, a synchronous *intra-cell* collision occurs. Two neighboring cells can start transmitting together (i.e., within a backoff slot) before they could sense each other's transmissions, resulting in synchronous *inter-cell* collisions. For example, consider Figure 3(b) and suppose that Cell-1 and Cell-2 are counting down together. With positive probability, there would be simultaneous transmissions in both Cell-1 and Cell-2 leading to synchronous inter-cell collisions. Recall that, due to Assumption (A2), there cannot be asynchronous hidden node collisions in the network. However, blocking due to carrier sensing is, in general, asynchronous in nature. For example, consider Figure 3(b) and suppose that Cell-1 starts transmitting first and blocks Cell-2 (no later than a backoff slot). Since Cell-3 is independent of Cell-1, Cell-3 can start transmitting at any instant during the transmission in Cell-1. Thus, transmissions in Cell-1 and Cell-3 would overlap at the nodes in Cell-2 in an asynchronous manner keeping Cell-2 blocked.

To capture both the synchrony and the asynchrony, we extend the *two-stage* approach of [7] from node-level to cell-level. In the first stage, we ignore inter-cell collisions and assume that inter-cell blocking due to carrier sensing is *immediate*. We develop a continuous time model by extending the independent-set approach of [5] from node-level to cell-level, and obtain the fraction of *time* for which each cell is transmitting/blocked/in backoff. This requires careful choice of cell-level transition rates,

which we obtain in (2) and (3). In the second stage, we obtain the fraction of *backoff slots* in which various subsets of neighboring cells can start transmitting together, and compute the collision probabilities in (9) by accounting for synchronous inter-cell collisions. We combine the above by a fixed-point equation and compute the throughputs using the solution of the fixed-point equation. We define the following notation:

$\beta_i :=$ (transmission) attempt probability (over the backoff slots) of the nodes in Cell-i

$\gamma_i :=$ conditional collision probability of the nodes in Cell-i (conditioned on an attempt being made)

Thus, at every backoff slot boundary, every node in Cell-i starts transmitting with probability β_i, i.e., the attempt process of each node is a Bernoulli process over the backoff slots. As in [4], β_i is related to γ_i by

$$\beta_i = G(\gamma_i) := \frac{1 + \gamma_i \cdots + \gamma_i^K}{b_0 + \gamma_i b_1 \cdots + \gamma_i^k b_k + \cdots + \gamma_i^K b_K}, \quad (1)$$

where K denotes the *retransmit limit* and b_k, $0 \le k \le K$, denotes the mean backoff sampled after the kth collision (for the same packet). The *backoff parameters* K and b_k, $0 \le k \le K$, are fixed in the DCF and their values depend on the PHY layer being used.

3.1.1. The First Stage

When Cell-i and some (or all) of its neighboring cells are in backoff, they contend until one of the cells, say, Cell-j, $j \in \mathcal{N}_i \cup \{i\}$, starts transmitting. Since we ignore inter-cell collisions in the first stage, the possibility of two or more neighboring cells starting transmission together is ruled out. When Cell-i starts transmitting, it gains control over its local medium by *immediately* blocking its neighbors that are not yet blocked. We assume that the time until Cell-i goes from the backoff state to the transmitting state is exponentially distributed with mean $1/\lambda_i$.

The *"activation rate"* λ_i of Cell-i is given by

$$\lambda_i = \frac{1 - (1 - \beta_i)^{n_i}}{\sigma}, \quad (2)$$

where we recall that n_i denotes the number of nodes in Cell-i, σ denotes the duration of a backoff slot (in seconds), and $1 - (1 - \beta_i)^{n_i}$ is the probability that there is an attempt in Cell-i per backoff slot.

Remarks 3.1. In (2) we have converted the aggregate transmission attempt probability in a cell per (discrete) backoff *slot* to an attempt rate over (continuous) backoff *time*. Our assumption of "exponential time until transition from the backoff state to the transmitting state" is the continuous time analogue of the assumption of "geometric number of backoff slots until transmission attempt" in the discrete time models of [3] and [4]. □

Remarks 3.2. The authors in [7] use an unconditional activation rate λ over all times as well as a conditional activation rate g over the backoff times. We use a single activation rate λ which is conditional on being in the backoff state. Our modified approach with a single conditional activation rate can also be applied to simplify the node- and link-level model of [7]. □

When Cell-i transmits, its neighbor cells remain blocked (due to Cell-i) until Cell-i's transmission finishes and an idle DIFS period elapses. If the transmission is a success (resp. an intra-cell collision) in Cell-i, then Cell-i's neighbors remain blocked for a *success time T_s* (resp. a *collision time T_c*). In the Basic Access (resp. RTS/CTS) mechanism, T_s corresponds to the time DATA-SIFS-ACK-DIFS (resp. RTS-SIFS-CTS-SIFS-DATA-SIFS-ACK-DIFS) and T_c corresponds to the time DATA-EIFS (resp. RTS-EIFS). T_s and T_c can be computed using the protocol parameters and mean DATA payload size L. We denote the mean duration for which Cell-i's neighbors remain blocked due to a transmission in Cell-i by $1/\mu_i$, where

$$\frac{1}{\mu_i} = p_{succ,i} \cdot T_s + (1 - p_{succ,i}) \cdot T_c, \qquad (3)$$

where $p_{succ,i} = \frac{n_i \beta_i (1-\beta_i)^{n_i-1}}{1-(1-\beta_i)^{n_i}}$ is the probability that the transmission in Cell-i is successful given that there is a transmission in Cell-i. We take the activity periods of Cell-i to be i.i.d. exponential random variables with mean $1/\mu_i$.

Our approximation of exponential activity periods is justified by a well-known insensitivity result (see [25, Theorem 1] and [5]) and the detailed balance equations satisfied by our model (see Equation (4)).

Due to carrier sensing, at any point in time, only a set $\mathcal{A}$ ($\subset \mathcal{N}$) of mutually independent *cells* can be transmitting together, i.e., $\mathcal{A}$ must be an *independent set* (of vertices) of $\mathcal{G}$. From $\mathcal{G}$, we can determine the set of cells $\mathcal{B}_\mathcal{A}$ that get blocked due to $\mathcal{A}$, and the set of cells $\mathcal{U}_\mathcal{A}$ that remain in backoff, i.e., the set of cells in which nodes can continue to decrement their backoff counters. Note that $\mathcal{A}$, $\mathcal{B}_\mathcal{A}$ and $\mathcal{U}_\mathcal{A}$ form a partition of $\mathcal{N}$.

We take $\mathcal{A}(t)$, i.e., the set $\mathcal{A}$ of cells that are transmitting at time t, as the state of the multi-cell system at time t. Due to the exponential approximations of the distributions of the time to activation and the activity periods, at any time t, the rate of transition to the next state is completely determined by the current state $\mathcal{A}(t)$. For example, Cell-j, $j \in \mathcal{U}_\mathcal{A}$, joins the set $\mathcal{A}$ (and its neighboring cells that are also in $\mathcal{U}_\mathcal{A}$ join the set $\mathcal{B}_\mathcal{A}$) at a rate λ_j. Similarly, Cell-i, $i \in \mathcal{A}$, leaves the set $\mathcal{A}$ (and its neighboring cells that are blocked only due to Cell-i leave the set $\mathcal{B}_\mathcal{A}$) to join the set $\mathcal{U}_\mathcal{A}$ at a rate μ_i. We conclude that, *the process $\{\mathcal{A}(t), t \geq 0\}$ has the structure of a Continuous Time Markov Chain* (CTMC).

The CTMC $\{\mathcal{A}(t), t \geq 0\}$ has a finite number of states, and it is irreducible as long as the λ_i's and the μ_i's are non-zero, which is the case when the contention windows and the packet payloads are finite. Then, the CTMC $\{\mathcal{A}(t), t \geq 0\}$ is *stationary* and *ergodic*. The set of all possible independent sets of $\mathcal{G}$ which constitutes the state space of the CTMC $\{\mathcal{A}(t), t \geq 0\}$ is denoted by $\boldsymbol{\mathcal{A}}$. The state when none of the cells is transmitting will be denoted by Φ, where Φ denotes the *empty set*. For a given contention graph $\mathcal{G}$, $\boldsymbol{\mathcal{A}}$ can be determined. For the topology given in Figure 3(c) we have $\boldsymbol{\mathcal{A}} = \{\Phi, \{1\}, \{2\}, \{3\}\}$, and for the topology given in Figure 3(b), we have $\boldsymbol{\mathcal{A}} = \{\Phi, \{1\}, \{2\}, \{3\}, \{1,3\}\}$. The corresponding CTMCs are given in Figures 4(a) and 4(b).

It can be checked that the transition structure of the CTMC $\{\mathcal{A}(t), t \geq 0\}$ satisfies the "Kolmogorov Criterion" for reversibility [26]. Hence, the stationary probability distribution $\pi(\mathcal{A}), \mathcal{A} \in \boldsymbol{\mathcal{A}}$, satisfies the *detailed balance equations*, $\forall i \in \mathcal{U}_\mathcal{A}$,

$$\pi(\mathcal{A})\lambda_i = \pi(\mathcal{A} \cup \{i\})\mu_i. \qquad (4)$$

Define the *"access intensity"* ρ_i of Cell-i by

$$\rho_i := \frac{\lambda_i}{\mu_i}. \qquad (5)$$

Due to reversibility, the stationary distribution of $\{\mathcal{A}(t), t \geq 0\}$ has the following product-form, $\forall \mathcal{A} \in$

(a)

(b)
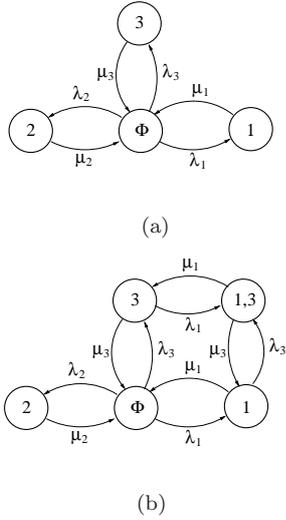
Figure 4: The CTMCs describing the cell-level contention for example networks in (a) Figure 3(c), and (b) Figure 3(b).

$\mathcal{A}$,

$$\pi(\mathcal{A}) = \left(\prod_{i \in \mathcal{A}} \rho_i \right) \pi(\Phi), \qquad (6)$$

where $\pi(\Phi)$ denotes the stationary probability that none of the cells is transmitting. $\pi(\Phi)$ can be obtained by applying the normalization equation

$$\sum_{\mathcal{A} \in \boldsymbol{\mathcal{A}}} \pi(\mathcal{A}) = 1. \qquad (7)$$

Combining (6) and (7), we obtain, $\forall \mathcal{A} \in \boldsymbol{\mathcal{A}}$, the stationary probability $\pi(\mathcal{A})$ that the particular set $\mathcal{A}$ of cells is transmitting by

$$\pi(\mathcal{A}) = \frac{\left(\prod_{i \in \mathcal{A}} \rho_i \right)}{\sum_{\mathcal{A} \in \boldsymbol{\mathcal{A}}} \left(\prod_{j \in \mathcal{A}} \rho_j \right)}, \qquad (8)$$

where we recall that a product $\prod$ over an empty index set is taken to be equal to 1.

3.1.2. The Second Stage

In the first stage analysis in Section 3.1.1, we had ignored inter-cell collisions. We now compute the collision probabilities γ_i's accounting for inter-cell collisions. Note that γ_i is conditional on an attempt being made by a node in Cell-i. Hence, to compute γ_i, we focus only on those states in which Cell-i can attempt. Clearly, Cell-i can attempt in State-$\mathcal{A}$ iff it is in back-off in State-$\mathcal{A}$, i.e., iff $i \in \mathcal{U}_\mathcal{A}$. In all such states a node in Cell-i can incur intra-cell

collisions due the other nodes in Cell-i. Furthermore, some (or all) of Cell-i's neighbors might also be in back-off in State-$\mathcal{A}$. If a neighboring cell, say, Cell-j, $j \in \mathcal{N}_i$, is also in back-off in State-$\mathcal{A}$, i.e., if $j \in \mathcal{U}_\mathcal{A}$, then a node in Cell-$i$ can incur inter-cell collisions due to the nodes in Cell-j. The collision probability γ_i is then given by, $\forall i \in \mathcal{N}$,

$$\gamma_i = \frac{\sum_{\mathcal{A} \in \boldsymbol{\mathcal{A}} \,:\, i \in \mathcal{U}_\mathcal{A}} \pi(\mathcal{A}) \gamma_i(\mathcal{A})}{\sum_{\mathcal{A} \in \boldsymbol{\mathcal{A}} \,:\, i \in \mathcal{U}_\mathcal{A}} \pi(\mathcal{A})}, \qquad (9)$$

where

$$\gamma_i(\mathcal{A}) = 1 - (1 - \beta_i)^{n_i - 1} \prod_{j \in \mathcal{N}_i \,:\, j \in \mathcal{U}_\mathcal{A}} (1 - \beta_j)^{n_j}$$

represents the (conditional) collision probability of the nodes in Cell-i in State-$\mathcal{A}$. A derivation of (9) can be found in the Appendix at the end of the paper. We emphasize that (9) is different from the equation for synchronous collisions, Equation (10), of [7].

3.1.3. Fixed Point Formulation

The γ_i's can be written as functions of the β_i's by applying (2)-(9). Let, $\forall i \in \mathcal{N}$,

$$\gamma_i = \Gamma_i(\beta_1, \beta_2, \ldots, \beta_N)$$

denote this functional dependence. Also, $\forall i \in \mathcal{N}$, (1) provides the functional dependence of β_i on γ_i as $\beta_i = G(\gamma_i)$. Using the notation $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_N)$, we obtain an N-dimensional fixed point equation in terms of $\boldsymbol{\beta}$ given by

$$\boldsymbol{\beta} = (G(\Gamma_1(\boldsymbol{\beta})), G(\Gamma_2(\boldsymbol{\beta})), \ldots, G(\Gamma_N(\boldsymbol{\beta}))), \quad (10)$$

where we recall that N is the total number of cells. This N-dimensional fixed point equation can be numerically solved to obtain the attempt probabilities β_i's. Once the β_i's are known, the stationary state probabilities $\pi(\mathcal{A})$'s and the collision probabilities γ_i's can also be obtained by applying (2)-(9). In all of the cases that we have considered, the fixed point iterations were observed to converge to the same solutions irrespective of starting points.

3.1.4. Computation of Saturation Throughputs

Let $x_i(\mathcal{G})$ denote the fraction of time for which Cell-i is *not* blocked in a multi-cell network with contention graph $\mathcal{G}$. Thus, $x_i(\mathcal{G})$ is the fraction of time when Cell-i is either transmitting ($i \in \mathcal{A}$) or

in backoff ($i \in \mathcal{U}_\mathcal{A}$), and hence, $\forall i \in \mathcal{N}$,

$$x_i(\mathcal{G}) = \sum_{\mathcal{A} \in \boldsymbol{\mathcal{A}} \,:\, i \in \mathcal{A} \cup \mathcal{U}_\mathcal{A}} \pi(\mathcal{A}), \qquad (11)$$

where we recall that $\boldsymbol{\mathcal{A}}$ depends on $\mathcal{G}$.

Let $\Theta_i^{sat}(\mathcal{G})$ denote the aggregate throughput in packets/sec of (the n_i saturated nodes in) Cell-i in the multi-cell network with contention graph $\mathcal{G}$. Let $\Theta_{n,singlecell}^{sat}$ denote the aggregate throughput in packets/sec of a single *isolated* cell containing n saturated nodes. Note that $\Theta_{n,singlecell}^{sat}$ can be computed as in [3] or [4]. We approximate $\Theta_i^{sat}(\mathcal{G})$ by

$$\Theta_i^{sat}(\mathcal{G}) \approx x_i(\mathcal{G}) \cdot \Theta_{n_i,singlecell}^{sat} \qquad (12)$$

and $\Theta_i^{sat}(\mathcal{G})$ divided by n_i gives the per node throughput $\theta_i^{sat}(\mathcal{G})$ in Cell-i, i.e., $\theta_i^{sat}(\mathcal{G}) = \Theta_i^{sat}(\mathcal{G})/n_i$.

If Cell-i is indeed an isolated cell, then we have $x_i(\mathcal{G}) = 1$ and $\Theta_i^{sat}(\mathcal{G}) = \Theta_{n_i,singlecell}^{sat}$. However, in general, Cell-i remains blocked (by other cells' activity) for a fraction of time $1 - x_i(\mathcal{G})$, and hence, $x_i(\mathcal{G})$ appears as the reduction factor in (12). The approximation in (12) is explained as follows. If we could ignore the time wasted in inter-cell collisions, then the times during which Cell-i is not blocked would consist of only of the backoff slots and the activities of Cell-i by itself. However, in general, a fraction of the medium time is wasted in inter-cell collisions. Thus, the single cell throughput $\Theta_{n_i,singlecell}^{sat}$ of Cell-i is only an approximation of the aggregate throughput of Cell-i, over the times during which it is not blocked, and $\Theta_{n_i,singlecell}^{sat}$ multiplied by $x_i(\mathcal{G})$ provides an approximation of the aggregate throughput $\Theta_i^{sat}(\mathcal{G})$ of Cell-i in the multi-cell network. In Section 5 we show that our analytical results (that are based on (12)) are quite accurate when compared with simulations.

3.2. Extension to TCP-controlled Long-lived Downloads

We now extend the analysis of Section 3.1 to the case when STAs download long files via their respective APs from a *local* server using TCP connections. Our extension to TCP-controlled long-lived flows is based on the analytical model for multi-cell WLANs with saturated nodes, developed in Section 3.1, and the *equivalent saturated model* of [20] for TCP-WLAN interaction in a single cell. The model proposed by [20] has been shown to be quite accurate under the following assumptions:

Aa. The TCP source agents reside in a *local server* which is connected with the AP by a relatively fast wireline LAN (which is our network setting) such that the AP in the WLAN is the bottleneck for every TCP connection.

Ab. Every STA has a *single* long-lived TCP connection.

Ac. There are no packet losses due to buffer overflow.

Ad. The TCP timeouts are set large enough to avoid timeout expiration due to Round Trip Time (RTT) fluctuations.

Ae. The delayed ACK mechanism has been disabled.

Af. The TCP connections have equal maximum windows.

We assume Aa-Af in this section. Under Aa-Af, [20] propose to model a single cell having an AP and an *arbitrary* number of STAs (with long-lived TCP connections) by an *equivalent saturated network* which consists of a saturated AP and a *single* saturated STA. "This equivalent saturated model greatly simplifies the modeling problem since the TCP flow control mechanisms are now implicitly hidden and the aggregate throughput can be computed using the saturation analysis [20]."

Using the equivalent saturated model of [20], the analysis of Section 3.1 can be directly applied to TCP-controlled long file downloads by taking $n_i = 2, \forall i \in \mathcal{N}$, i.e., *by assuming two saturated nodes per cell regardless of how many STAs are actually present in the cells*. Of the two saturated nodes, one represents a saturated AP and the other represents an equivalent saturated STA.

Remarks 3.3. In TCP-controlled downloads, APs send TCP DATA packets to their associated STAs and STAs send TCP ACK packets to their respective APs. Note that, both TCP DATA and TCP ACK packets are DATA units from the point of view of the MAC. Thus, to properly account for TCP DATA and TCP ACK traffic, we take the size of the MAC data units of one of the saturated nodes to be $L_{TCP-DATA}$ and that of the other saturated node to be $L_{TCP-ACK}$, where $L_{TCP-DATA}$ (resp. $L_{TCP-ACK}$) denotes the size of a TCP DATA (resp. TCP ACK) packet plus the IP header. $\square$

3.2.1. AP Throughputs with Long-lived Flows

The throughput of an AP, by definition, is equal to the aggregate throughput obtained by all the users/STAs that are served by the AP. Let $\theta_{singlecell}^{AP-TCP}$ denote the throughput in packets/sec of the AP in a single isolated cell under long-lived downloads. We obtain $\theta_{singlecell}^{AP-TCP}$ by applying the equivalent saturated model of [20]. Clearly, half the successful transmissions must belong to the AP (no-delayed-ACKs). Thus, we must have

$$\theta_{singlecell}^{AP-TCP} = \Theta_{n=2,singlecell}^{sat,L_{TCP,avg}}/2,$$

where $\Theta_{n=2,singlecell}^{sat,L_{TCP,avg}}$ is computed by taking $n = 2$ saturated nodes and an average payload size of

$$L_{TCP,avg} = (L_{TCP-DATA} + L_{TCP-ACK})/2.$$

The throughput, $\theta_i^{AP-TCP}(\mathcal{G})$, of the AP in Cell-$i$ in a multi-cell network with contention graph $\mathcal{G}$, is given by

$$\theta_i^{AP-TCP}(\mathcal{G}) \approx x_i(\mathcal{G}) \cdot \theta_{singlecell}^{AP-TCP}, \qquad (13)$$

where $x_i(\mathcal{G})$ is computed by the analysis of Section 3.1, taking $n_i = 2, \forall i \in \mathcal{N}$.

3.3. Complexity of the Model

The computational benefits of our cell-level model as compared to a node-level model is due to the fact that its complexity grows with the number of cells rather than nodes. The computational complexity lies in: (i) finding the state space $\mathcal{A}$ by searching for all possible independent sets $\mathcal{A}$, and (ii) finding the fixed point given the state space $\mathcal{A}$.

(i) The complexity of finding all the independent sets is akin to the *maximum independent set problem*, which is known to be NP-hard [27]. In theory, the complexity of finding the state space $\mathcal{A}$ can grow exponentially with the number of vertices in the contention graph [8]. However, for realistic topologies, where connectivity in the contention graph is related to physical distances between the nodes, [8] proposed an efficient algorithm whose complexity, in practice, grows only polynomially.

(ii) Finding the fixed point falls into the class of FIXP [28]. In our fixed point routine, one does not have to solve the Markov chain's balance equations to obtain the stationary distribution. Given the state space $\mathcal{A}$, computation of the stationary distribution is immediate by (8). However, the num-

ber of equations inside the fixed point routine increases linearly with the number of vetices in the contention graph. The class FIXP is believed to lie between P and NP [28], and the complexity of finding the fixed point can grow exponentially with the number of vertices in the contention graph.

In a node-level model, every node or link creates one vertex in the contention graph, and its complexity grows with the number of nodes. In our cell-level model, however, a vertex in the contention graph corresponds to a cell rather than a single wireless node. Thus, the complexity of our cell-level model grows with the number of cells rather than nodes. In particular, given the number of cells, the complexity of our cell-level model remains constant as the number of nodes per cell increases.

Consider, for example, the network scenario in Figure 3(d) which consists of 7 cells. For a node-level model, say, with 10 nodes per cell, one would have to search through 2^{70} possible subsets to determine the state space. However, in our cell-level model, we had to search through only $2^7 = 128$ possible subsets to determine the state space. Similarly, the number of equations in the fixed point routine in a node-level model would be 10 times that in our cell-level model.

3.4. Large ρ Regime

For fast computation of the cell throughputs, we extend (from node-level to cell-level) an approach originally proposed by [6] and later adopted by [9] and [13]. As in [6], we take $\rho_i = \rho, \forall i \in \mathcal{N}$, and then let $\rho \longrightarrow \infty$. We call this *the infinite ρ approximation*, and justify its applicability to our WLAN setting later in this subsection (see Example 3.2).

Let $\eta(\mathcal{G})$ denote the number of Maximum Independent Sets (MISs) of vertices of $\mathcal{G}$. Let $\eta_i(\mathcal{G})$ denote the number of MISs of vertices of $\mathcal{G}$ to which Cell-i belongs. Let $\alpha(\mathcal{G})$ denote the cardinality of an MIS of vertices of $\mathcal{G}$. For any graph G, $\alpha(G)$ is called the *independence number* of G.

Example 3.1. For the network in Figure 3(b) we have $\mathcal{A} = \{\Phi, \{1\}, \{2\}, \{3\}, \{1, 3\}\}$. There is only one MIS given by $\{1, 3\}$, and so $\eta = 1$. This MIS consists of two vertices, and so the cardinality of the MIS is $\alpha = 2$. Cell-1 and Cell-3 belong to this MIS, and so $\eta_1 = \eta_3 = 1$. Cell-2 does not belong to this MIS, and so $\eta_2 = 0$. For the network in Figure 3(d), there are three MISs, namely, $\{1, 2, 4, 7\}$, $\{1, 2, 5, 6\}$ and $\{1, 2, 5, 7\}$, and so $\eta = 3$. The cardinality of an MIS in this case is $\alpha = 4$. Cell-1 and Cell-2

belong to all three MISs, and so $\eta_1 = \eta_2 = 3$. Cell-5 and Cell-7 belong to two of the three MISs, and so $\eta_5 = \eta_7 = 2$. Cell-4 and Cell-6 belong to one of the three MISs, and so $\eta_4 = \eta_6 = 1$. Cell-3 does not belong to any of the three MISs, and so $\eta_3 = 0$. □

Applying the infinite ρ approximation to (8), we observe that, as $\rho \longrightarrow \infty$, we have

$$
\pi(\mathcal{A}) \longrightarrow \begin{cases} \dfrac{1}{\eta(\mathcal{G})} & \text{if } \mathcal{A} \text{ is an MIS of } \mathcal{G}, \\ 0 & \text{otherwise.} \end{cases}
$$

Thus, as $\rho \longrightarrow \infty$, only an MIS of cells can be transmitting at any point in time. Note that when an MIS of cells is transmitting, all other cells must be blocked. This implies that, as $\rho \longrightarrow \infty$, every cell is either transmitting or is blocked at all times, i.e., the probability that a cell is in backoff goes to 0. This essentially provides a physical interpretation for infinite ρ approximation, namely, *"the medium time wasted in contention via backoffs is negligible as compared to the time spent in transmissions."*

Applying the infinite ρ approximation to (11), we observe that, as $\rho \longrightarrow \infty$, we have

$$
x_i(\mathcal{G}) \longrightarrow \frac{\eta_i(\mathcal{G})}{\eta(\mathcal{G})}. \tag{14}
$$

The quantity

$$
x_i(\mathcal{G}) = \frac{\Theta_i}{\Theta_{singlecell}}
$$

can be interpreted as the throughput of Cell-i, normalized with respect to the single cell throughput (recall (12) and (13)). Defining the *normalized network throughput* $\bar{\Theta}(\mathcal{G})$ by

$$
\bar{\Theta}(\mathcal{G}) := \sum_{i=1}^{N} x_i(\mathcal{G}), \tag{15}
$$

we prove the following theorem.

Theorem 3.1. *As $\rho_i = \rho \longrightarrow \infty$ for all $i \in \mathcal{N}$, we have*

$$
\bar{\Theta}(\mathcal{G}) \longrightarrow \alpha(\mathcal{G}).
$$

Proof: By definition, we have $\bar{\Theta}(\mathcal{G}) = \sum_{i=1}^{N} x_i(\mathcal{G})$. Also, as $\rho_i = \rho \longrightarrow \infty$ for all $i \in \mathcal{N}$, we have $x_i(\mathcal{G}) \longrightarrow \frac{\eta_i(\mathcal{G})}{\eta(\mathcal{G})}$. Summing over all $i \in \mathcal{N}$, as $\rho_i = \rho \longrightarrow \infty$ for all $i \in \mathcal{N}$, we have $\sum_{i=1}^{N} x_i(\mathcal{G}) \longrightarrow \sum_{i=1}^{N} \frac{\eta_i(\mathcal{G})}{\eta(\mathcal{G})}$. Next we prove that

$\sum_{i=1}^{N} \frac{\eta_i(\mathcal{G})}{\eta(\mathcal{G})} = \alpha(\mathcal{G})$. Consider the equivalent equation

$$
\sum_{i=1}^{N} \eta_i(\mathcal{G}) = \alpha(\mathcal{G})\, \eta(\mathcal{G}),
$$

which can be proved by a counting argument as follows. Imagine the $\eta(\mathcal{G})$ MISs are written as $\eta(\mathcal{G})$ rows with each row listing the cell indices in the corresponding MIS. Since each row contains $\alpha(\mathcal{G})$ elements and there are $\eta(\mathcal{G})$ rows, the total number of elements is equal to $\alpha(\mathcal{G})\, \eta(\mathcal{G})$ which is given by the right hand side. The total number of elements can also be counted as follows. First count how many times Cell-i, $i = 1, 2, \ldots, N$, appears. Clearly, Cell-i appears exactly $\eta_i(\mathcal{G})$ times. Now summing up the counts $\eta_i(\mathcal{G})$'s we get the total number of elements which is given by the left hand side. Hence, Theorem 3.1 is proved. □

Remarks 3.4. Recall that, $\alpha(\mathcal{G})$ is the cardinality of an MIS of $\mathcal{G}$. Thus, the maximum possible value which the normalized network throughput $\bar{\Theta}(\mathcal{G})$ can take is $\alpha(\mathcal{G})$. Theorem 3.1 says that, as $\rho_i = \rho \longrightarrow \infty$ for all $i \in \mathcal{N}$, the normalized network throughput $\bar{\Theta}(\mathcal{G})$ is automatically maximized in a distributed manner without any additional control. This property of the CSMA/CA protocol has been recently established by [9] in the context of an infinite linear chain of saturated nodes. In Theorem 3.1 we prove it for multi-cell WLANs satisfying the PBD condition with arbitrary contention graph $\mathcal{G}$ and with saturated nodes as well as for TCP-controlled long-lived downloads. □

Next, we provide an example to argue that *realistic access intensities are indeed "large" in the sense that if the PBD condition holds, then the values of the normalized cell throughputs with realistic access intensities are very close to their values with $\rho \longrightarrow \infty$.*

Example 3.2. Consider the scenario in Figure 3(d) with 802.11b default backoff parameter settings, 11 Mbps data and control rates, and 10 saturated nodes in each cell. Since, it is difficult to control ρ directly, we examine the variation of the x_i's with increase in ρ in two steps. First, notice in Figure 5 that the access intensities increase monotonically as the MAC payload size increases. Thus, it is reasonable to infer the variation of the x_i's with ρ by examining the variation of the x_i's with the MAC payload size. Second, notice in Figure 6 that, above a payload size, say, of 500 bytes, the x_i's are very close to their values with $\rho \to \infty$, i.e., they are very

11
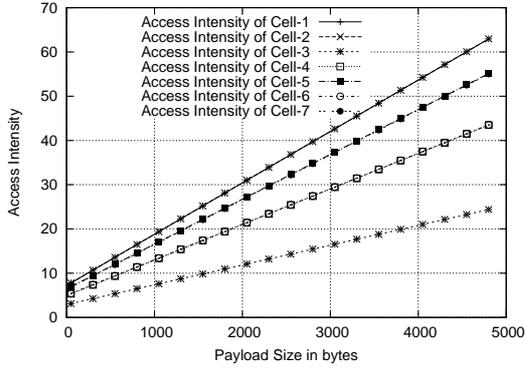
Figure 5: Variation of access intensities with MAC payload size: for the scenario in Figure 3(d) with 802.11b default parameter settings, 11 Mbps data and control rate, and 10 saturated nodes in each cell.
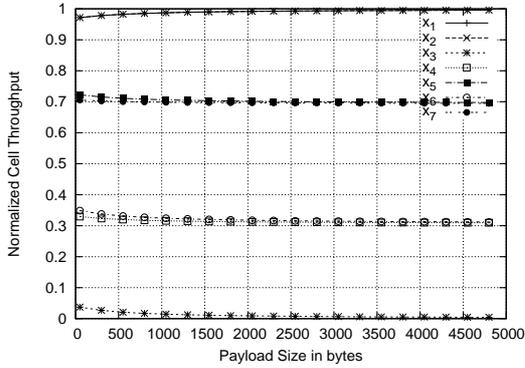
Figure 6: Variation of normalized cell throughputs with MAC payload size: for the scenario in Figure 3(d) with 802.11b default parameter settings, 11 Mbps rates, and 10 saturated nodes in each cell.

close to the values $x_1 = x_2 = 1, x_3 = 0, x_4 = x_6 = 1/3, x_5 = x_7 = 2/3$, which are predicted by (14). Figures 5 and 6 indicate that for access intensities as small as 5-15 (which correspond to the payload size of 500 bytes) the infinite ρ approximation is already a very good approximation. □

4. Short-lived Downloads

In Section 3.4 we showed that the infinite ρ approximation can indeed be a very good approximation to predict cell throughputs at realistic access intensities. In this section, we apply the infinite ρ approximation to develop a simple and accurate model under short-lived downloads, and to obtain predictions for mean flow transfer delays in each cell. We keep A0-A5 of Section 2, and also Aa-Af of

Section 3.2, except Ab. As in [10], applying the so-called *time-scale separation* assumption, we model a multi-cell infrastructure WLAN under short-lived downloads as a network of state-dependent processor sharing queues. Our network model under A0-A5, Aa, and Ac-Af is a special case of the network model in [10] and our traffic model in this case is the same as that in [10]. The key difference lies in the service models.

4.1. Traffic Model

TCP-controlled short file downloads by users in Cell-i and served by AP-i are referred to as class-i flows. Flows of class-i arrive at AP-i according to a Poisson process of rate ν_i (flows/sec). The flow arrival processes of different classes are independent. Flow sizes are i.i.d. exponentially distributed random variables with mean $\mathsf{E}(V)$ (bits/flow).[3] Flow sizes are independent of the arrival processes. Note that, flow sizes correspond to *application-level* data.

4.2. Network State

Let z_i denote the number of ongoing flows at AP-i in Cell-i. We take

$$\boldsymbol{z} = (z_1, z_2, \ldots, z_N)$$

as the state of the network, where we recall that N denotes the number of APs in the WLAN. Thus, the state space is $\mathbb{Z}_+^N$ where $\mathbb{Z}_+$ denotes the set of all non-negative integers.

4.3. Service Model

We now develop a model for the service process according to which the APs serve the ongoing flows. Let $L_{APP-DATA}$ denote the size (in bits) of application-level data per TCP DATA packet. Let Θ denote the rate (in bits/sec) at which a single isolated AP transfers application-level data to its STAs under long-lived downloads. Then, we have

$$\Theta = \theta_{singlecell}^{AP-TCP} \times L_{APP-DATA},$$

where we recall that $\theta_{singlecell}^{AP-TCP}$ denotes the throughput in packets/sec of the AP in a single isolated cell under long-lived downloads. It was shown by [19] and [20] that $\theta_{singlecell}^{AP-TCP}$ is largely insensitive to

---

[3]It is possible to generalize to class-dependent flow size distributions, but we consider the same flow size distribution for all classes.

the number of STAs in the cell (which is why in Section 3.2 we could replace the arbitrary number of STAs by a single saturated STA). Hence, Θ is also insensitive to the number of STAs in the cell. In our service model, Θ represents the service rate in bits/sec of a single isolated (processor sharing) queue, which could represent a single isolated cell under short-lived downloads.

4.3.1. The Service Model of [10]

In [10], different STAs belonging to the same cell might block and be blocked by different subsets of APs and STAs in the other cells, and thus, the users in the same cell could be divided into multiple classes accordingly (see [10] for details). In our network setting, due to the PBD condition (A1), nodes in the same cell have an identical set of interferers, and they belong to the same class. Thus, our network model is a special case of the more general network model in [10].

Let $\mathbf{1}_{\{\cdot\}}$ denote the indicator function. Let $\varphi_i(\boldsymbol{z})$ denote the rate in bits/sec at which application-level data is transferred by AP-i to its STAs in state $\boldsymbol{z}$, according to the service model in [10]. It turns out that, the service rate $\varphi_i(\boldsymbol{z})$ in bits/sec of the ith processor-sharing queue in state $\boldsymbol{z}$, $z_i > 0$, as applied to our network model (of single user class per cell), is given by

$$\varphi_i(\boldsymbol{z}) = \frac{\Theta}{1 + \sum_{j \in \mathcal{N}_i} \mathbf{1}_{\{z_j > 0\}}}. \tag{16}$$

Thus, for our network model, the service model in [10] says that *the service rate of an AP in any state is equal to the service rate of an isolated AP divided by "1 plus the number of neighboring APs that are non-empty in that state."* Clearly, the service model in [10] is based on the assumption that, on the average, the single cell capacity is equally shared among the contending neighbors, i.e., among the neighboring APs that have non-empty queues.

We shall refer to the service model in [10], restricted to our network model and given by (16), as *Model-1*. In Section 5 we provide an example to show that Model-1, which is based only on the *number* of contending neighbors and does not capture the impact of the overall *topology* of non-empty APs, can lead to inaccurate results.

The fundamental assumption which leads to "an equal sharing of capacity among contending neighbors" is the assumption of "synchronous and slotted evolution" [10] which is valid if the network consists of a single contention domain, i.e., if $\mathcal{G}$ is *complete* (or, more generally, if $\mathcal{G}$ is a union of complete subgraphs). (Recall that every maximal clique of $\mathcal{G}$ corresponds to a contention domain.) However, as discussed in Section 3.1, the evolution of activities in networks with multiple overlapping contention domains is **not** synchronous owing to asynchronous inter-cell blocking. Thus, Model-1 leads to inaccurate results in such networks.

Figure 3(b) provides the simplest multi-cell network for which Model-1 leads to inaccurate results (see Example 4.1 in the next subsubsection).

4.3.2. Our Improved Service Model

We now propose our improved service model with short-lived downloads as follows. Let

$$S_{\boldsymbol{z}} := \{i \in \mathcal{N} \ : \ z_i > 0\}$$

denote the set of cells with non-empty AP queues in state $\boldsymbol{z}$. Let $\mathcal{G}_{\boldsymbol{z}}$ denote the subgraph of $\mathcal{G}$ restricted to $S_{\boldsymbol{z}}$. The service rate $\phi_i(\boldsymbol{z})$ in bits/sec of AP-i in state $\boldsymbol{z}$, $z_i > 0$, in our improved service model is given by

$$\phi_i(\boldsymbol{z}) = x_i(\mathcal{G}_{\boldsymbol{z}}) \cdot \Theta \tag{17}$$

where $x_i(\mathcal{G}_{\boldsymbol{z}})$ can be approximated by the infinite ρ approximation as

$$x_i(\mathcal{G}_{\boldsymbol{z}}) \approx \frac{\eta_i(\mathcal{G}_{\boldsymbol{z}})}{\eta(\mathcal{G}_{\boldsymbol{z}})}. \tag{18}$$

Thus, in any state $\boldsymbol{z}$, we (i) obtain the subset $S_{\boldsymbol{z}}$ of cells with non-empty AP queues, (ii) obtain the restriction $\mathcal{G}_{\boldsymbol{z}}$ of the contention graph $\mathcal{G}$ by restricting $\mathcal{G}$ to the cells in $S_{\boldsymbol{z}}$, (iii) obtain the normalized throughputs of the APs in $S_{\boldsymbol{z}}$ w.r.t. the contention graph $\mathcal{G}_{\boldsymbol{z}}$ by invoking the infinite ρ approximation, and (iv) obtain the state-dependent service rates in state $\boldsymbol{z}$ by taking the product of normalized throughputs and the application-level single cell throughput Θ.

We shall refer to our service model, given by (17), as *Model-2*, and demonstrate its accuracy in Section 5. We emphasize that *Model-2 reduces to Model-1 for networks with a single contention domain.* However, for networks with multiple overlapping contention domains, Model-2 can capture the impact of the overall topology of the network which Model-1 cannot.

Example 4.1. Consider Figure 3(b) and suppose that the APs in all the three cells have non-empty

queues, i.e., $z_1, z_2, z_3 > 0$. According to Model-1, the ratio of their service rates should be $\frac{1}{2} : \frac{1}{3} : \frac{1}{2}$. However, according to Model-2, the ratio of their service rates should be $1 : 0 : 1$. In NS-2 simulations, if we saturate the AP queues so as to keep them non-empty all the time, the ratio of throughputs is observed to be $0.96 : 0.04 : 0.96$ which supports Model-2 rather than Model-1. Thus, as shown in Section 5, Model-1 provides inaccurate results, but Model-2 is quite accurate. $\qquad\square$

4.4. Approximate Delay Analysis

Let $\hat{x}_i$ denote the long-term average normalized throughput of AP-i, i.e., $\hat{x}_i$ denotes the long-term average fraction of time for which the nodes in Cell-i are not blocked. The *"effective service rate"* $\hat{\phi}_i$ in bits/sec of the ith processor-sharing queue is then given by $\hat{\phi}_i = \hat{x}_i\Theta$.

We decouple the queues by assuming that each queue i evolves as an independent processor-sharing queue with arrival rate ν_i flows/sec, i.i.d. exponentially distributed flow sizes with mean $\mathsf{E}(V)$ bits and service rate $\hat{\phi}_i$ bits/sec. The probability that queue i is empty (resp. non-empty) is then given by $\max\{0, 1 - \frac{\nu_i\mathsf{E}(V)}{\hat{x}_i\Theta}\}$ (resp. $\min\{1, \frac{\nu_i\mathsf{E}(V)}{\hat{x}_i\Theta}\}$).

We model the inter-dependence among the queues by requiring that, $\forall i \in \mathcal{N}$, the $\hat{x}_i$'s satisfy Equation (19) (placed at the top of the next page). Equation (19) is explained as follows. We consider the queues, other than queue i, and obtain the probability $p(S)$ that the subset $S \subseteq \mathcal{N} \setminus \{i\}$ of queues is non-empty, assuming the queues to be independent given the effective service rates $\hat{\phi}_j$'s. We then multiply $p(S)$ with the normalized throughput of AP-i w.r.t. the graph $\mathcal{G}[S \cup \{i\}]$, where $\mathcal{G}[S \cup \{i\}]$ denotes the restriction of $\mathcal{G}$ to the subset $S \cup \{i\}$ of vertices. Finally, we sum over all possible subsets $S \subseteq \mathcal{N} \setminus \{i\}$ and obtain (19).

The set of equations given by Equation (19) constitute a multi-dimensional fixed-point equation which can be solved to obtain the $\hat{x}_i$'s. Once the $\hat{x}_i$'s are obtained, we approximate the *mean flow transfer delay* $\mathsf{E}(D)_i$ in Cell-i by

$$\mathsf{E}(D)_i = \frac{\frac{\mathsf{E}(V)}{\hat{x}_i\Theta}}{1 - \frac{\nu_i\mathsf{E}(V)}{\hat{x}_i\Theta}}$$

which resembles the standard expression for mean flow transfer delay in a processor-sharing queue with mean arrival rate ν_i flows/sec, mean flow size $\mathsf{E}(V)$ bits, and service rate $\hat{x}_i\Theta$ bits/sec. Results

obtained from our approximate delay analysis are provided in Section 5.6.

5. Numerical and Simulation Results

In this section we validate our analytical model by comparing with the results obtained from NS-2 simulations [29]. We created the example topologies in Figures 3(a)-3(d). We chose the cell radii and the carrier sensing range to ensure that the PBD condition is satisfied. Under the PBD condition, the collision model in NS-2 is exactly as given by Assumptions (A2)-(A3). The channel was configured to be error-free. Nodes were randomly placed within the cells. The AP buffer size was set large enough to avoid buffer losses. The delayed ACK mechanism was disabled. Each case was simulated 20 times, each run for 200 sec of "network time". (The short-file download case was simulated for 1000 flows per cell per run.) We report the results for "Basic Access". We took 11 Mbps data and control rates and data payload size of 1000 Bytes. We report only the mean values of simulation results. In almost all cases, the 99% confidence intervals about the mean values were observed to be within 5-10% of the mean values.

All parameters are as per 802.11b PHY layer settings, except in Sec. 5.4 where we provide some results with 802.11g PHY layer settings.

The analytical values of single cell throughputs per-node for the saturated case were obtained by applying the model of [4]. The analytical values of single cell throughputs of the APs for the TCP case were obtained by applying the model of [20]. The analytical single cell throughputs per-node (resp. of the AP) multiplied with the x_i's obtained from our multi-cell model provide the analytical throughputs per-node (resp. of the AP) in a multi-cell network (see (12)).

5.1. Saturated Nodes with UDP Traffic

Figures 7 and 8 compare the collision probability γ and the throughput per node θ, respectively, for the example network in Figure 3(d) with saturated nodes and UDP traffic. For comparison, in Figures 7 and 8 we also show the corresponding single cell results, i.e., the results one would expect had the seven cells been mutually independent. Referring to Figures 7 and 8, we make the following observations:

14

$$\hat{x}_i = \sum_{S \subseteq \mathcal{N} \setminus \{i\}} \left(\prod_{j \in S} \min\left\{1, \frac{\nu_j \mathsf{E}(V)}{\hat{x}_j \Theta}\right\} \right) \times \left(\prod_{k \in \mathcal{N} \setminus (S \cup \{i\})} \max\left\{0, 1 - \frac{\nu_k \mathsf{E}(V)}{\hat{x}_k \Theta}\right\} \right) \times \left(\frac{\eta_i\big(\mathcal{G}[S \cup \{i\}]\big)}{\eta\big(\mathcal{G}[S \cup \{i\}]\big)} \right) \tag{19}$$
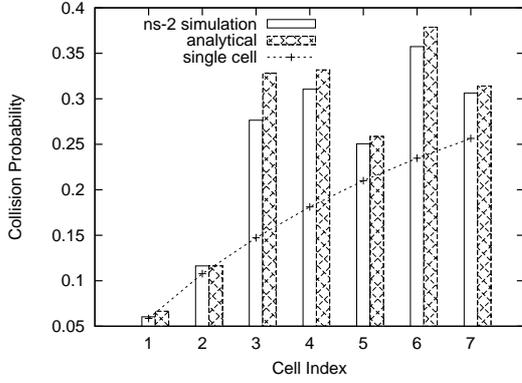
Figure 7: Collision probability γ_i for each cell i, $1 \leq i \leq 7$, for the scenario in Figure 3(d) with UDP traffic when Cell-i contains $n_i = i + 1$ saturated nodes.
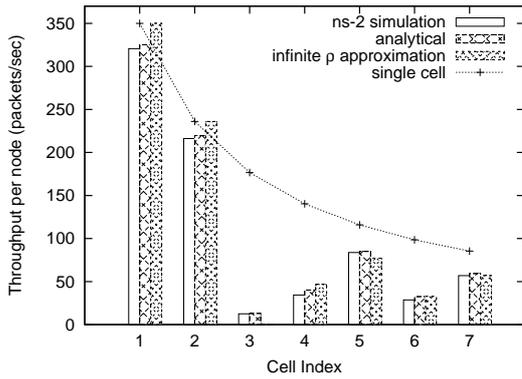
Figure 8: Per-node throughput θ_i^{sat} for each cell i, $1 \leq i \leq 7$, for the scenario in Figure 3(d) with UDP traffic when Cell-i contains $n_i = i + 1$ saturated nodes.

O_1 Our fixed-point analysis provides quite accurate predictions of collision probabilities and throughputs (less than 10% error in most cases). The slight over-estimation of throughputs is due to the approximation in (12). The infinite ρ approximation (which does not require solving a fixed-point equation) provides a simple and efficient method to quickly compute the throughputs fairly accurately. Nevertheless, our fixed point analysis provides more accurate predictions for throughputs and also provides accurate predictions for collision probabilities which cannot be provided by the infinite ρ approximation. Had we not ac-

counted for the inter-cell collisions in the second stage, our analytical collision probabilities would have been equal to the corresponding single cell collision probabilities (see Figure 7).

O_2 The throughput of a cell cannot be accurately determined based only on the *number* of neighboring cells. In Figure 3(d), Cell-3 and Cell-4 each have three neighbors but their per node throughputs θ are quite different (see Figure 8). In particular, $\theta_4 > \theta_3$ even though $n_3 = 4 < n_4 = 5$. This is due to Cell-7 which blocks Cell-6 for certain fraction of time during which Cell-4 gets opportunity to transmit whereas Cell-1 and Cell-2 are almost never blocked and Cell-3 is almost always blocked due to Cell-1 and Cell-2. Thus, *topology of the entire network plays the key role.*

5.2. Long-lived Downloads

Figures 9 and 10 compare the collision probability γ and the throughput of the AP, respectively, for the example network in Figure 3(d). Our simulation results in this case correspond to $n = 5$ STAs in each cell. Note, however, that *the analytical values in Figures 9 and 10 were obtained by taking $n_i = 2$ saturated nodes in every cell.* Referring to Figures 9 and 10, we conclude that the foregoing observations (O_1) and (O_2) for the saturated case carry over to TCP-controlled long file downloads as well.

Referring to Figure 10, we can compare the state-dependent service rates as per Model-1 and Model-2. Model-1, the service model of [10], would predict the throughput ratios as $\frac{1}{2} : \frac{1}{2} : \frac{1}{4} : \frac{1}{4} : \frac{1}{2} : \frac{1}{3} : \frac{1}{2}$. The observed throughput ratios in Figure 10 are much closer to Model-2's prediction given by $1 : 1 : 0 : \frac{1}{3} : \frac{2}{3} : \frac{1}{3} : \frac{2}{3}$.

5.3. Varying the Number of Nodes Per Cell

Figure 11 depicts the variation of cell throughput with the number of nodes per cell for the example scenario in Figure 3(d) with $n = 10, 15, 20, 25, 30$ saturated nodes per cell and UDP traffic. Figure 12 depicts the variation of AP throughput with the number of STAs per cell for the example scenario in Figure 3(d) with $n = 10, 15, 20, 25, 30$ STAs and

15
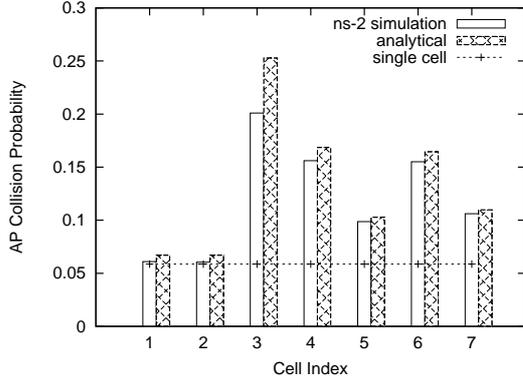
Figure 9: Collision probability of the AP γ_i^{AP} for each cell i, $1 \leq i \leq 7$, for the example scenario in Figure 3(d) when each cell contains an AP and $n = 5$ STAs. STAs are downloading files through their respective APs using long-lived TCP connections.

Figure 11: Cell throughput Θ_i^{sat} for each cell i, $1 \leq i \leq 7$, with $n = 5, 10, 15, 20, 25, 30$ saturated nodes per cell for the example scenario in Figure 3(d).

Figure 10: Throughput of the AP θ_i^{AP-TCP} for each cell i, $1 \leq i \leq 7$, for the example scenario in Figure 3(d) when each cell contains an AP and $n = 5$ STAs. STAs are downloading files through their respective APs using long-lived TCP connections.
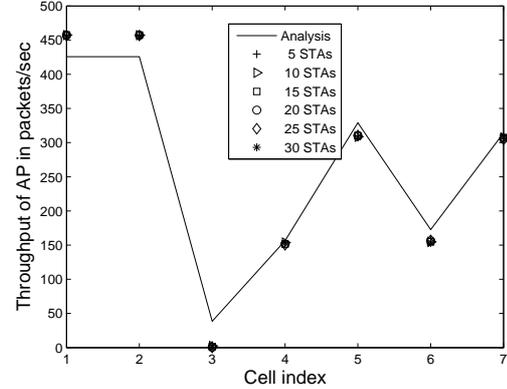
Figure 12: Throughput of the AP θ_i^{AP-TCP} for each cell i, $1 \leq i \leq 7$, with $n = 5, 10, 15, 20, 25, 30$ STAs for the example scenario in Figure 3(d). STAs are downloading files through their respective APs using long-lived TCP connections.

TCP-controlled long-file downloads. Referring to Figures 11 and 12, we observe that

O_3 For TCP-controlled long-file downloads (Figure 12), increasing the number of STAs per cell has no effect on the AP throughput because, effectively, only two nodes are contending at any point in time irrespective of how many STAs are actually there. For saturated nodes and UDP traffic (Figure 11), the cell throughput slowly decreases with the number of nodes per cell due to increased collisions. In both cases, our cell-level model can predict the throughputs quite acurately (maximum 12% error, except in Figure 12 for Cell-3, which was found to be severely blocked in simulations).

5.4. Using 802.11g PHY Layer

Figure 13 depicts the throughput per node with 802.11g PHY layer settings for the example scenario in Figure 3(d) when each cell contains $n = 15$ saturated nodes. We have used the PHY layer parameter settings pertaining to the 802.11g-only mode (and not the 802.11b-compatibility mode) with the highest possible data rate of 54 Mbps. Due to a higher data rate and a shorter slot duration, we observed improvements in the throughputs (approx. four-fold) with 802.11g PHY layer settings.

Since our analytical model can be applied with any given values of timing parameters and contention windows (given the PBD condition holds), we observe good agreement between analysis and simulation with 802.11g PHY layer settings as well

16
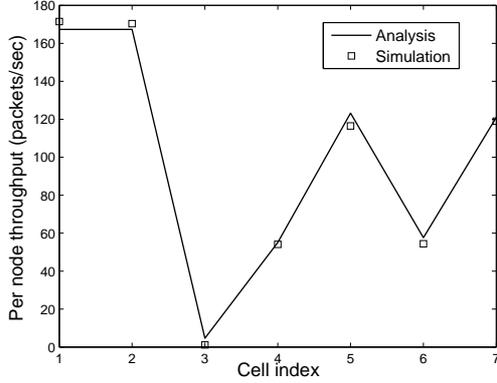
Figure 13: Throughput per node θ_i^{sat} for each cell i, $1 \leq i \leq 7$, with 802.11g PHY layer settings for the example scenario in Figure 3(d) when each cell contains $n = 15$ saturated nodes per cell.

(Figure 13). It can be observed that the ratios of cell throughputs are close to Model-2's prediction given by $1 : 1 : 0 : \frac{1}{3} : \frac{2}{3} : \frac{1}{3} : \frac{2}{3}$. This indicates that the access intensities are large enough even at the high data rate of 54 Mbps.

5.5. Relaxing the PBD Condition in Simulations

Till now we have demonstrated that our analytical model is able to predict the collision probabilities and throughputs quite accurately if the PBD condition holds. In this subsection we report simulation results for a few scenarios where the PBD condition does not hold, i.e., when two dependent cells are not necessarily completely dependent (recall Definition 2.1) such that only a subset of nodes in one cell are dependent w.r.t. a subset of nodes in the other cell.

The impact of relaxing the PBD condition is studied by varying the carrier sensing range R_{cs}. We chose the example scenario in Figure 3(d) with cell radius $R = 50$ for each cell. The coordinates of the cell centres for Cell-i, $i = 1, \ldots, 7$, are given by $(0, 800)$, $(0, 0)$, $(300, 400)$, $(800, 400)$, $(1100, 800)$, $(1100, 0)$ and $(1600, 0)$, respectively. Hence, the distance between the cell centres in neighbouring cells is 500. R_{cs} is varied from 400 to 600 with increments of 50. $R_{cs} = 600$ corresponds to complete dependence and $R_{cs} = 400$ corresponds to independence (both cases with no hidden nodes). The cases $R_{cs} = 550, 500, 450$ correspond to different percentages of hidden nodes as shown in Figures 14-16, respectively, and we call them *Relaxation R1, R2, and R3* of the PBD condition, respectively.
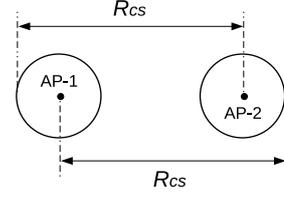
Figure 14: Relaxation (R1) of the PBD condition: APs can sense all the nodes in the other cell, but $\approx 50\%$ STA(s) can sense only a subset of STAs in the other cell.
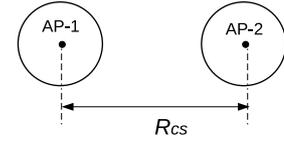
Figure 15: Relaxation (R2) of the PBD condition: APs can sense each other, but the APs cannot sense $\approx 50\%$ STA(s) in the other cell.
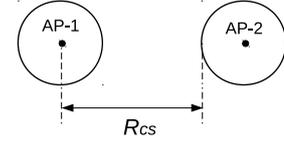
Figure 16: Relaxation (R3) of the PBD condition: APs can barely sense the nodes located on the boundary of the other cell, but STA(s) nearer to the neighboring cells can sense and interfere with each other.

We consider TCP-controlled long-lived downloads where the APs transmit for a much larger fraction of time than the STAs. Figure 17 depicts the variation of AP throughput with $R_{cs} = 400, 450, 500, 550, 600$ and 10 STAs per cell undergoing TCP-controlled long-file downloads for the example scenario in Figure 3(d). Referring to Figure 17, we observe that

O4 For $R_{cs} = 400$, the cells are independent and the AP throughputs are equal to the single cell throughput, which can be predicted by a single cell analysis. For $R_{cs} = 600$, the dependent cells are completely dependent, the PBD condition holds and the AP throughputs are quite accurately predicted by our cell-level model.

O5 For Relaxation R1 or $R_{cs} = 550$ (Figure 14), our cell-level model quite accurately predicts the AP throughputs for $R_{cs} = 550$ even though the dependent cells are not completely dependent and there are hidden nodes. In fact, for Relaxation R1, the TCP DATA traffic from the AP to the STAs is hidden-node-free, but the
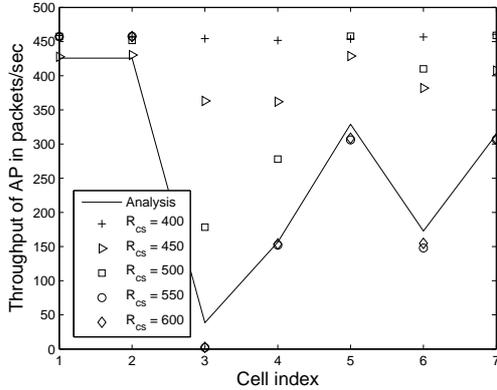
Figure 17: Variation of long-lived TCP throughput of the AP θ_i^{AP-TCP} for each cell i, $1 \leq i \leq 7$, with respect to the carrier sensing range $R_{cs} = 400, 450, 500, 550, 600$, with 10 STAs per cell for the example scenario in Figure 3(d).

TCP ACK traffic from the STAs to the AP may suffer from hidden node collisions. However, since TCP ACKs are cumulative, the loss of one ACK may be compensated by subsequent ACKs and the cell-level model works.

O_6 For Relaxation R2 or $R_{cs} = 500$ (Figure 15), even though the number of interfering APs is equal for Cell-3 and Cell-4, Cell-3 is blocked for a larger fraction of time than Cell-4, and our analytical model captures this fact. Accurate prediction of throughputs for Relaxation R2 and R3 requires the modeling of the impact of hidden nodes.

5.6. Short-lived Downloads

To compare Model-1 and Model-2, we developed a customized event-driven queueing simulator which can simulate networks of single-class processor-sharing queues with Poisson flow arrivals, i.i.d. exponentially distributed flow sizes and the two service models, namely, Model-1 and Model-2. Results from the model-based flow-level queueing simulations are compared against the results obtained with the NS-2 simulator [29] which performs protocol-based packet-level detailed simulations with $n = 5$ STAs per cell. We also compare the above simulation results with our approximate delay analysis.

5.6.1. Single Contention Domain

Figure 18 compares the mean flow transfer delays obtained from (i) NS-2 simulations, (ii) our cus-
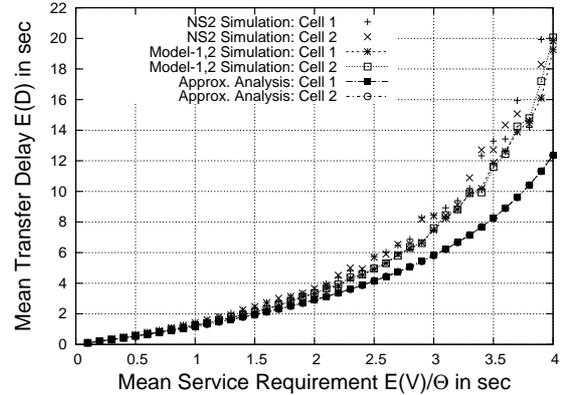
Figure 18: Comparing the mean flow transfer delays for the two-cell network in Figure 3(a) with $\nu_1 = \nu_2 = 0.1 \ sec^{-1}$.
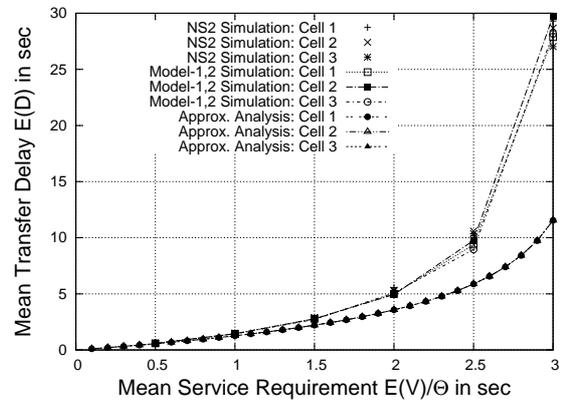
Figure 19: Comparing the mean flow transfer delays for the three-cell network in Figure 3(c) with $\nu_1 = \nu_2 = \nu_3 = 0.1 \ sec^{-1}$.

tomized queueing simulations according to Model-1 and Model-2, and (iii) our approximate delay analysis, for the two-cell network in Figure 3(a) with $\nu_1 = \nu_2 = 0.1 \ sec^{-1}$. Figure 19 shows a similar comparison for the three-cell network in Figure 3(c). The x-axis depicts the *mean service requirement*, $\frac{E(V)}{\Theta}$, of flows in seconds. For example, $\frac{E(V)}{\Theta} = 3 \ sec$, corresponds to exponentially distributed flow sizes of mean $E(V)$ bits where the value of $E(V)$ is such that, a server which works at a constant rate of Θ bits/sec has to dedicate $3 \ sec$ to completely serve a flow. Since the two-cell (resp. three-cell) network in Figure 3(a) (resp. Figure 3(c)) consists of a single contention domain, Model-1 and Model-2 are identical for these cases. We observe that:

18
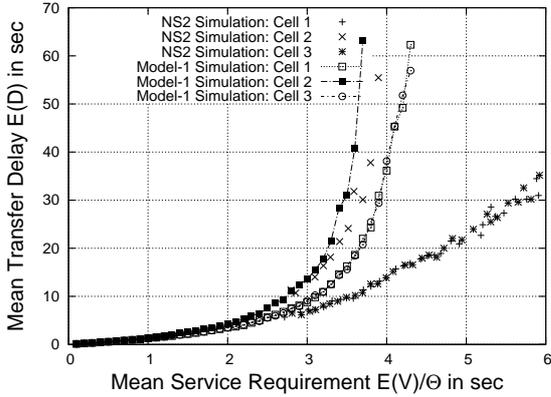
Figure 20: Comparing the mean flow transfer delays from NS-2 simulations with the delays corresponding to Model-1 for the three-cell network in Figure 3(b) with $\nu_1 = \nu_2 = \nu_3 = 0.1 \; sec^{-1}$.
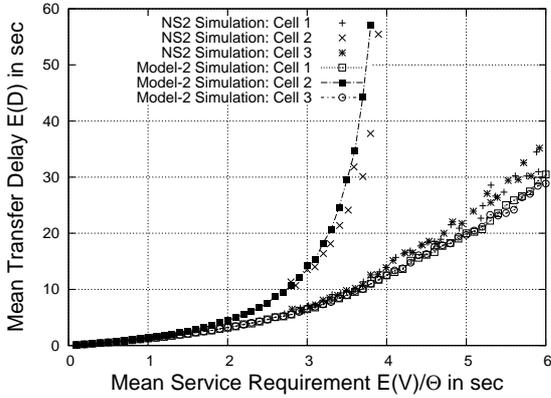
Figure 21: Comparing the mean flow transfer delays from NS-2 simulations with the delays corresponding to Model-2 for the three-cell network in Figure 3(b) with $\nu_1 = \nu_2 = \nu_3 = 0.1 \; sec^{-1}$.
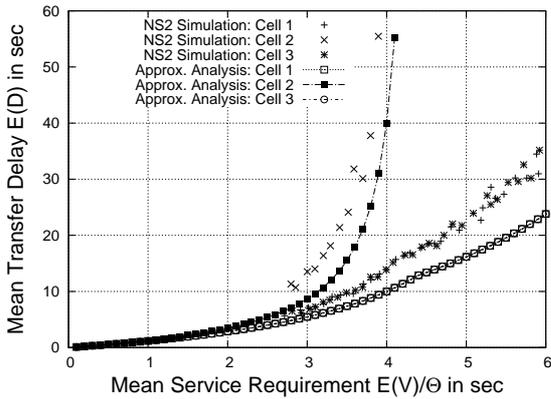
Figure 22: Comparing the mean flow transfer delays from NS-2 simulations with the delays from our approximate delay analysis for the three-cell network in Figure 3(b) with $\nu_1 = \nu_2 = \nu_3 = 0.1 \; sec^{-1}$. Note that the points corresponding to our approximate delay analysis for Cell-1 and Cell-3 are overlapping.

O_7 A close match between our customized queueing simulations and NS-2 simulations validates Model-1 (also Model-2) for networks with a single contention domain.

5.6.2. Multiple Overlapping Contention Domains

Figures 20, 21, and 22 compare the mean flow transfer delays from NS-2 simulations with the mean flow transfer delays corresponding to Model-1, Model-2, and our approximate delay analysis, respectively, for the three-cell network in Figure 3(b) with $\nu_i = 0.1 \; sec^{-1}$. We observe that:

O_8 Model-1 leads to inaccurate results, since it predicts delays for Cell-1 and Cell-3 that are much higher than that obtained from NS-2 simulations (see Figure 20). Apart from the mismatch, Model-1 does not capture the way the network behaves with variation in load. However, Model-2 matches with NS-2 simulations extremely well (see Figure 21).

O_9 Our approximate delay analysis provides good predictions for the mean transfer delays (see Figure 22).

6. Conclusion

In this paper, we identified a Pairwise Binary Dependence (PBD) condition which enables scalable cell-level modeling of multi-cell infrastructure WLANs. Under a unified framework, we developed accurate analytical models for multi-cell infrastructure WLANs with saturated nodes, and for TCP-controlled long- and short-file downloads. An important property of CSMA, namely, "maximization of the network throughput in a distributed manner" has been established in the recent work by [9] in the context of an infinite linear chain of saturated nodes. We established it for multi-cell infrastructure WLANs with arbitrary cell topologies under the assumption of the PBD condition. We also provided analytical and simulation results in support of the so-called "infinite ρ approximation" which exhibit the throughput maximization property of CSMA. We characterized the service process in multi-cell infrastructure WLANs with TCP-controlled short-lived downloads. Such characterization is essential for accurately computing the mean flow transfer delays. We improved the service model of [10] by incorporating the impact of the *topology* of the entire network and demonstrated

19

its accuracy. By applying an "effective service rate approximation" technique we also obtained good approximations for the mean flow transfer delay in each cell.

Appendix A. Derivation of Equation (9)

Consider a tagged node in Cell-i. Let

$A_i(T)$: number of transmission attempts made by the tagged node up to time T

$C_i(T)$: number of collisions as seen by the tagged node up to time T

Then, the (conditional) collision probability γ_i of a tagged node in Cell-i, i.e., the probability that a transmission attempt in Cell-i collides, is given by

$$\gamma_i := \lim_{T \longrightarrow \infty} \frac{C_i(T)}{A_i(T)}. \qquad (A.1)$$

Consider any state $\mathcal{A}$ such that $i \in \mathcal{U}_\mathcal{A}$. Let

$A_i^\mathcal{A}(T)$: number of transmission attempts made by the tagged node in State-$\mathcal{A}$, up to time T

$C_i^\mathcal{A}(T)$: number of collisions as seen by the tagged node, up to time T, given that the transmission attempts were made in State-$\mathcal{A}$

$B_i^\mathcal{A}(T)$: number of backoff slots elapsed in the local medium of Cell-i in State-$\mathcal{A}$, up to time T

Since the CTMC $\{\mathcal{A}(t), t \geq 0\}$ is stationary and ergodic, the (long-term) fraction of time for which the network remains in State-$\mathcal{A}$ is equal to the stationary probability $\pi(\mathcal{A})$. Consider now a sufficiently long time T. Then, the time spent in State-$\mathcal{A}$ up to time T is approximately equal to $\pi(\mathcal{A})T$, and $B_i^\mathcal{A}(T) \approx \dfrac{\pi(\mathcal{A})T}{\sigma}$ where recall that σ is the duration of a backoff slot. Thus, we have

$$\lim_{T \longrightarrow \infty} \frac{B_i^\mathcal{A}(T)}{T} = \frac{\pi(\mathcal{A})}{\sigma}. \qquad (A.2)$$

Since, the nodes in Cell-i attempt with probability β_i in every backoff slot irrespective of whether the other cells can attempt or not, we have $A_i^\mathcal{A}(T) \approx \beta_i B_i^\mathcal{A}(T)$. Thus,

$$\lim_{T \longrightarrow \infty} \frac{A_i^\mathcal{A}(T)}{B_i^\mathcal{A}(T)} = \beta_i. \qquad (A.3)$$

By definition (see (A.1)),

$$\gamma_i = \lim_{T \longrightarrow \infty} \frac{C_i(T)}{A_i(T)} = \frac{\lim_{T \longrightarrow \infty} \frac{1}{T} C_i(T)}{\lim_{T \longrightarrow \infty} \frac{1}{T} A_i(T)}$$

$$= \frac{\lim_{T \longrightarrow \infty} \frac{1}{T} \sum_{\mathcal{A} \in \mathbf{\mathcal{A}} \,:\, i \in \mathcal{U}_\mathcal{A}} C_i^\mathcal{A}(T)}{\lim_{T \longrightarrow \infty} \frac{1}{T} \sum_{\mathcal{A} \in \mathbf{\mathcal{A}} \,:\, i \in \mathcal{U}_\mathcal{A}} A_i^\mathcal{A}(T)}$$

$$= \frac{\lim_{T \longrightarrow \infty} \frac{1}{T} \sum_{\mathcal{A} \in \mathbf{\mathcal{A}} \,:\, i \in \mathcal{U}_\mathcal{A}} B_i^\mathcal{A}(T) \times \frac{C_i^\mathcal{A}(T)}{B_i^\mathcal{A}(T)}}{\lim_{T \longrightarrow \infty} \frac{1}{T} \sum_{\mathcal{A} \in \mathbf{\mathcal{A}} \,:\, i \in \mathcal{U}_\mathcal{A}} B_i^\mathcal{A}(T) \times \frac{A_i^\mathcal{A}(T)}{B_i^\mathcal{A}(T)}}$$

$$= \frac{\sum_{\mathcal{A} \in \mathbf{\mathcal{A}} \,:\, i \in \mathcal{U}_\mathcal{A}} \left(\lim_{T \longrightarrow \infty} \frac{B_i^\mathcal{A}(T)}{T} \right) \times \left(\lim_{T \longrightarrow \infty} \frac{C_i^\mathcal{A}(T)}{B_i^\mathcal{A}(T)} \right)}{\sum_{\mathcal{A} \in \mathbf{\mathcal{A}} \,:\, i \in \mathcal{U}_\mathcal{A}} \left(\lim_{T \longrightarrow \infty} \frac{B_i^\mathcal{A}(T)}{T} \right) \times \left(\lim_{T \longrightarrow \infty} \frac{A_i^\mathcal{A}(T)}{B_i^\mathcal{A}(T)} \right)}$$

$$(A.4)$$

We can write

$$\lim_{T \longrightarrow \infty} \frac{C_i^\mathcal{A}(T)}{B_i^\mathcal{A}(T)} = \lim_{T \longrightarrow \infty} \frac{A_i^\mathcal{A}(T)}{B_i^\mathcal{A}(T)} \times \lim_{T \longrightarrow \infty} \frac{C_i^\mathcal{A}(T)}{A_i^\mathcal{A}(T)} \qquad (A.5)$$

where $\lim_{T \longrightarrow \infty} \frac{C_i^\mathcal{A}(T)}{A_i^\mathcal{A}(T)}$ can be obtained by

$$\lim_{T \longrightarrow \infty} \frac{C_i^\mathcal{A}(T)}{A_i^\mathcal{A}(T)} = 1 - (1-\beta_i)^{n_i-1} \prod_{j \in \mathcal{N}_i \,:\, j \in \mathcal{U}_\mathcal{A}} (1-\beta_j)^{n_j}. \qquad (A.6)$$

Equation (A.6) can be explained as follows. Given that a node in Cell-i attempts a transmission, the transmission is successful if: (i) the $n_i - 1$ other nodes in Cell-i do not attempt transmissions in the same backoff slot, which happens with probability $(1 - \beta_i)^{n_i-1}$, and (ii) for every neighboring cell, Cell-j, that is also in backoff in the state $\mathcal{A}$, the n_j nodes in Cell-j do not attempt transmissions in the same slot, which happens with probability $\prod_{j \in \mathcal{N}_i \,:\, j \in \mathcal{U}_\mathcal{A}} (1 - \beta_j)^{n_j}$. Every transmission attempt by a node in Cell-i in State-$\mathcal{A}$ collides with the complementary probability $p_{coll,\mathcal{A}}$ given by

$$p_{coll,\mathcal{A}} := 1 - (1 - \beta_i)^{n_i-1} \prod_{j \in \mathcal{N}_i \,:\, j \in \mathcal{U}_\mathcal{A}} (1 - \beta_j)^{n_j}.$$

20

Thus, for T sufficiently long, we have

$$C_i^{\mathcal{A}}(T) \approx p_{coll,\mathcal{A}} A_i^{\mathcal{A}}(T).$$

Equation (A.6) is obtained by taking the limit $T \longrightarrow \infty$.

Combining (A.2), (A.3), (A.4), (A.5), (A.6), we obtain (9).

[1] M. K. Panda, A. Kumar, Modeling multi-cell IEEE 802.11 WLANs with application to channel assignment, in: Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, 2009. WiOPT 2009. 7th International Symposium on, IEEE, pp. 1–10.

[2] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE Std 802.11-2007, 2007.

[3] G. Bianchi, Performance analysis of the IEEE 802.11 distributed coordination function, Selected Areas in Communications, IEEE Journal on 18 (2000) 535–547.

[4] A. Kumar, E. Altman, D. Miorandi, M. Goyal, New insights from a fixed-point analysis of single cell IEEE 802.11 WLANs, Networking, IEEE/ACM Transactions on 15 (2007) 588–601.

[5] R. R. Boorstyn, A. Kershenbaum, B. Maglaris, V. Sahin, Throughput analysis in multihop CSMA packet radio networks, Communications, IEEE Transactions on 35 (1987) 267–274.

[6] X. Wang, K. Kar, Throughput modelling and fairness issues in CSMA/CA based ad-hoc networks, in: INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE, volume 1, IEEE, pp. 23–34.

[7] M. Garetto, T. Salonidis, E. W. Knightly, Modeling per-flow throughput and capturing starvation in csma multi-hop wireless networks, IEEE/ACM Transactions on Networking (TON) 16 (2008) 864–877.

[8] A. Kershenbaum, R. R. Boorstyn, M.-S. Chen, An algorithm for evaluation of throughput in multihop packet radio networks with complex topologies, Selected Areas in Communications, IEEE Journal on 5 (1987) 1003–1012.

[9] M. Durvy, O. Dousse, P. Thiran, Self-organization properties of CSMA/CA systems and their consequences on fairness, Information Theory, IEEE Transactions on 55 (2009) 931–943.

[10] T. Bonald, A. Ibrahim, J. Roberts, Traffic capacity of multi-cell WLANs, ACM SIGMETRICS Performance Evaluation Review 36 (2008) 419–430.

[11] L. B. Jiang, S. C. Liew, Improving throughput and fairness by reducing exposed and hidden nodes in 802.11 networks, Mobile Computing, IEEE Transactions on 7 (2008) 34–49.

[12] C.-K. Chau, M. Chen, S. C. Liew, Capacity of large-scale CSMA wireless networks, in: Proceedings of the 15th annual international conference on Mobile computing and networking, ACM, pp. 97–108. Detailed technical report available at http://arxiv.org/pdf/0909.3356v4.pdf.

[13] S. C. Liew, C. H. Kai, H. C. Leung, P. Wong, Back-of-the-envelope computation of throughput distributions in CSMA wireless networks, Mobile Computing, IEEE Transactions on 9 (2010) 1319–1331.

[14] L. Fu, S. C. Liew, J. Huang, Effective carrier sensing in CSMA networks under cumulative interference, Mobile Computing, IEEE Transactions on 12 (2013) 748–760.

[15] H. Q. Nguyen, F. Baccelli, D. Kofman, A stochastic geometry analysis of dense IEEE 802.11 networks, in: INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE, IEEE, pp. 1199–1207.

[16] B. Jang, M. L. Sichitiu, IEEE 802.11 saturation throughput analysis in the presence of hidden terminals, IEEE/ACM Transactions on Networking (TON) 20 (2012) 557–570.

[17] F.-Y. Hung, I. Marsic, Performance analysis of the IEEE 802.11 DCF in the presence of the hidden stations, Computer Networks 54 (2010) 2674–2687.

[18] K.-L. Hung, B. Bensaou, Throughput analysis and bandwidth allocation for IEEE 802.11 WLAN with hidden terminals, Journal of Parallel and Distributed Computing 71 (2011) 1201–1214.

[19] G. Kuriakose, S. Harsha, A. Kumar, V. Sharma, Analytical models for capacity estimation of IEEE 802.11 WLANs using DCF for Internet applications, Wireless Networks 15 (2009) 259–277.

[20] R. Bruno, M. Conti, E. Gregori, An accurate closed-form formula for the throughput of long-lived TCP connections in IEEE 802.11 WLANs, Computer Networks 52 (2008) 199–212.

[21] R. Litjens, F. Roijers, J. van den Berg, R. Boucherie, M. Fleuren, Performance analysis of wireless LANs: an integrated packet/flow level approach, Teletraffic Science and Engineering 5 (2003) 931–940.

[22] D. Miorandi, A. A. Kherani, E. Altman, A queueing model for HTTP traffic over IEEE 802.11 WLANs, Computer networks 50 (2006) 63–79.

[23] T. Bonald, S. Borst, N. Hegde, A. Proutière, Wireless data performance in multi-cell scenarios, ACM SIGMETRICS Performance Evaluation Review 32 (2004) 378–380.

[24] H. Ma, R. Vijayakumar, S. Roy, J. Zhu, Optimizing 802.11 wireless mesh networks based on physical carrier sensing, Networking, IEEE/ACM Transactions on 17 (2009) 1550–1563.

[25] P. Whittle, Partial balance and insensitivity, Journal of Applied Probability (1985) 168–176.

[26] F. P. Kelly, Reversibility and stochastic networks, Cambridge University Press, 2011.

[27] G. J. Woeginger, Exact algorithms for NP-hard problems: A survey, in: Combinatorial OptimizationEureka, You Shrink!, Springer, 2003, pp. 185–207.

[28] K. Etessami, M. Yannakakis, On the complexity of Nash equilibria and other fixed points, SIAM Journal on Computing 39 (2010) 2531–2597.

[29] S. McCanne, S. Floyd, The ns Network Simulator (v2.33), 2008. http://www.isi.edu/nsnam/ns/.