Robust Client-Based Wi-Fi Topology Discovery

Pantelis A. Frangoudis, Dimitrios I. Zografos, and George C. Polyzos
Department of Informatics
Athens University of Economics and Business
Email: {pfrag,zwgrafos,polyzos}@aueb.gr

***Abstract*—The low cost and ease of installation of Wi-Fi equipment operating in unlicensed spectrum have made dense Wi-Fi deployments a reality in most modern urban areas. With the lack of many non-overlapping frequencies to operate on, interference among neighbor Wi-Fi cells can cause significant performance degradation. Here we study the problem of topology discovery in such dense deployments, which is necessary in order to combat interference. To this end, we apply a client-driven scheme, where client devices sense the spectrum and report overlapping cells. However, reporting entities cannot always be assumed trustworthy. We therefore study cases where reporters attack the discovery process by submitting fake information and propose simple countermeasures to tackle some attacker strategies. We show analytically and via simulations that, in realistic urban scenarios, our mechanisms are effective, even in the presence of a large number of attackers.**

I. Introduction

With the proliferation of IEEE 802.11-based WLAN equipment, Wi-Fi pervades modern metropolitan areas. Residential users, municipalities, university campuses and Wireless ISPs, among others, set up Access Points (APs) for public or private access. The low cost and ease of installation of Wi-Fi equipment, as well as its operation in unlicensed spectrum are the main reasons for its popularity. While in densely populated urban areas wireless coverage is no more an issue, unplanned and anarchic deployment of Wi-Fi networks comes with the cost of interference. For IEEE 802.11b/g there are only 3 non-overlapping frequency bands (channels) on which a Wi-Fi cell can operate. In the scenarios we study, the probability of coexistence of more than 3 WLANs at the same spot is high.

Combating interference in chaotic WLAN deployments necessitates sophisticated interference mitigation strategies by means of transmission power control or frequency selection, among others. Information on the topology of the network is vital input to such schemes. Discovering the topology of Wi-Fi deployments requires detecting overlapping Wi-Fi cells sharing common spectrum, but also collecting information about the number of clients affected by interference.

Such information can be reported by the wireless infrastructure (APs) or the clients themselves. There are significant advantages in involving clients in this process. First, reports by clients offer a user-perceived view of interference conditions, which an AP-centric scheme might fail to capture. Second, client density is typically higher than that of APs, thus a client-centric topology discovery scheme offers greater coverage, also exploiting user mobility.

In a different context, the density of wireless APs in metropolitan areas has made it possible to build Wi-Fi-based positioning systems based on recorded AP beacons [1]. This requires extensive site surveying to correlate AP beacons with locations. Delegating the task of AP mapping to roaming clients could offer similar coverage advantages.

In any case, it should be noted that in order for such schemes to be successful, reporting entities should be trustworthy. Otherwise, effective countermeasures need to be in place to filter fake information (or erroneous feedback due to equipment failures). There may be disincentives to contribute truthful feedback, such as the overhead of spectrum monitoring [1]. Also, in a competitive environment where clients subscribed with different Wi-Fi service providers visit one another's hotspots, one may be tempted to submit fraudulent reports to manipulate the spectrum sharing mechanism to his affiliated provider's advantage, in return for better service or other benefits. Fake reports pollute the system's view of interference conditions and, consequently, affect interference mitigation mechanisms.

In this paper, we study reporting schemes for discovering Wi-Fi topology that involve client participation. Our particular focus is on their security and robustness. Our contribution is the study of specific classes of attacks and the development of simple countermeasures based on majority rules. For realistic client and AP densities in urban settings [2], we show analytically and via simulations that, even when there is a large percentage of attackers, our mechanisms perform well in discovering network topology.

II. System model

We model our system as a weighted undirected graph. Vertices of the *Coverage Graph (CG)* represent APs and edges represent coverage overlap between neighbor Wi-Fi cells. As shown in Fig. 1, there are two cases of overlap. In the first case (*Type-1* edges), two APs are within range of each other. Even if no clients are there to report it, the operation of both cells will be affected. In the second case (*Type-2* edges), two APs are not within range of each other, but clients or other APs are located in the overlap area. The weight of an edge is a function of the number of reports about it and captures user-perceived interference. High-weight edges should be more carefully considered while assigning channels or adapting the transmission power of the respective APs, since they affect more users. Our model is very similar to the one proposed by Mishra et al. [3].

[1] An IEEE 802.11 active scan may take more than 250ms, during which time the client station cannot transmit/receive application data. More advanced spectrum usage measurements may be more time consuming.

978-1-4244-8788-2/11/$26.00 ©2011 IEEE

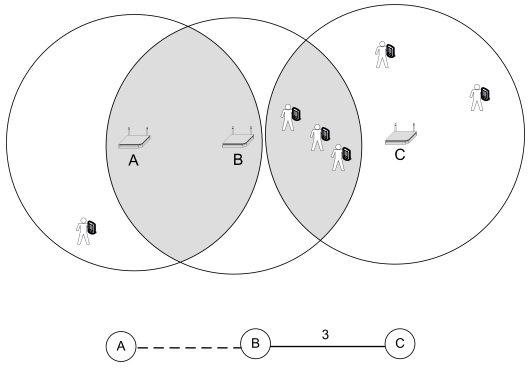


Fig. 1. The Coverage Graph. A-B is a Type-1 edge. If we assume a system where only clients submit reports, this edge will not be reported. B-C is a Type-2 edge.

The basic components of our system are the *reporting entities* and the *collector*. Reporting entities can either be APs or clients, although our system focuses on the latter. Based on reports, the aim of our system is to expose as many edges as possible; clients situated in areas where neighbor cells overlap report this fact, together with details about cell operation (channel number, received signal strength, etc.) to the APs they are attached to, and the latter forward reports to the collector, which is a centralized entity maintaining the reported CG. The recently standardized IEEE 802.11k [4] protocol could be used by clients and APs in the reporting process. Fig. 1 shows an instance of the CG where each report contributes a unit to an edge's weight, but this may not always be the case, as we will explain in Section III.

Reporting to the AP a client is attached to is not a strict requirement, since the client could directly communicate its measurements to the collector. This would make the system easier to deploy, since it would not require any modifications to the AP functionality. In the scenarios we study (see Section III), though, knowledge about which AP a client reports to may be necessary. In some centrally-managed deployments, as is the case in corporate WLANs, this information can be readily available to the controller. For simplicity and presentation clarity, we will assume, as well, that this information is available to the collector.

It should be noted that edges may not always be detected, due to lack of reporters located there (see for example the A - B edge in Fig. 1). Also, since reporters may misbehave, the reported CG does not necessarily encode the actual network topology. Fraudulent reports contribute fake CG edges. Our mechanisms aim at eliminating these edges and accurately representing true coverage and interference.

We do not assume that reports are trustworthy, but we assume that they are authenticated and that appropriate mechanisms are in place to ensure that a single user cannot send multiple reports in small timescales to deliberately increase the weight of particular edges. Authentication can limit *Sybil attacks*, but does not exclude collusion among reporters. The exact implementation of such protection mechanisms will be the topic of future work. Also, in this work, we assume that APs are always trustworthy.

III. Attacks and filtering mechanisms

We study two attack scenarios, for which there are different underlying assumptions as to user-AP affiliations. For each scenario, we devise simple filtering rules to effectively exclude fake edges from the CG. The intuition behind our filtering mechanisms is that for an edge to get accounted for, there should be sufficiently many reports about it. This consensus-based scheme comes with the cost of filtering edges which are not reported by many clients. However, bearing in mind that we model user-perceived interference, few reports about an edge indicate few affected users and are, thus, less important. Our analysis shows that for realistic urban Wi-Fi deployments where client density is high, our mechanisms perform well.

A. Scenario 1: independent attackers

In the first scenario, we do not assume any user-provider affiliations or other trust relationship. Clients act independently and their reports are considered of equal weight. For each interfering AP pair reported, a unit is added to the respective edge's weight. We assume no cooperation among attackers, each of whom submits reports containing a number of random fake AP identifiers. If we assume that the probability that two or more attackers report the same fake edge is negligible, each fake report contributes unit-weight edges to the CG, as well as fake vertices for each fake AP. These edges connect a real vertex (corresponding to the AP the reporter is associated with) to fake ones. Also, fake edges among fake vertices are added. The above lead us to the following observation.

Observation 1. In the first scenario, pruning all unit-weight edges eliminates the probability that a fake edge appears in the CG.

Thus, to combat this attack, we simply remove unit-weight edges from the reported CG.

B. Scenario 2: colluding attackers

In the second scenario, APs belong to a number of competing Wireless Internet Service Providers (WISPs) and each user is affiliated with one of them. There are two classes of users: *roamers*, i.e., those attached to APs belonging to "foreign" WISPs and *non-roamers*. Non-roamers always submit truthful reports, while roamers are not always honest and may form *colluding groups* as follows: Dishonest roamers affiliated with provider A currently attached to a single AP of provider B, agree to report the same fake set of random APs. The filtering scheme of the first scenario is useless here; these fake reports would contribute edges with higher weight to the CG, which the filtering mechanism would fail to detect.

To counter this attack, each AP values more reports that originate from *trusted* clients, i.e., clients affiliated with the same provider. Roamer reports are *discounted* so that their cumulative weight per AP does not exceed that of a single trusted report. Based on the number of roamers associated with it, each AP independently calculates the weight w assigned to the reports of each of these n roamers so that

$$w = \frac{1}{n} - e, \quad 0 < e \ll 1. \tag{1}$$

After discounting, a roamer's report is forwarded to the collecting entity. Now, for each pair of APs contained in it, the weight of the respective CG edge is incremented by w. The filtering mechanism prunes all CG edges which have weight less than 1. The following conditions are sufficient for an edge to be detected: (1) At least 1 non-roamer reports the edge, or (2) there are sufficiently many roamers reporting it, so that the sum of the weights of their reports is at least 1.

One should bear in mind that not all roamers reporting an edge need be assigned the same weight, since they may be attached to different APs, which may in turn have a different number of roaming users attached to them. The above process strictly bounds the weight of an edge reported by a colluding group below 1, which leads us to the following observation:

Observation 2. In the second scenario, pruning all edges with weight less than 1 eliminates the probability that a fake edge appears in the CG.

In both scenarios, the filtering mechanism's efficiency is only limited by potential *false negatives*, namely, real edges which fail to reach the unit-weight threshold (to the eyes of the collecting entity, such a case is equivalent to an attack).

IV. Performance analysis

We analytically determine the detection accuracy of our scheme in the presence of varying numbers of attackers, for the two scenarios discussed in Section III. Our methodology involves comparing the actual network topology to the discovered one and our evaluation metric is the percentage of detected CG edges. We assume idealized conditions, where AP coverage area is a disk of radius R.

A. Probability that an edge exists

We assume that clients and APs are spatially distributed following homogeneous Poisson Point Processes (PPP) with intensities λ_c and λ_{AP} respectively. The area of the overlap region between two APs is given by the following formula:

$$A(d) = 2R^2 cos^{-1}(\frac{d}{2R}) - \frac{d}{2}\sqrt{4R^2 - d^2}, \tag{2}$$

where $d \leq 2R$ is the distance between the two APs and R the cell radius, which we assume constant. Therefore, the probability that n clients are located in such a region is

$$P(n, d) = e^{-\lambda_c A(d)} \frac{(\lambda_c A(d))^n}{n!}. \tag{3}$$

The probability that n APs are located in such a region is calculated in a similar fashion. A CG edge exists if the respective APs are within range of each other (i.e., $d \leq R$) or, otherwise, there is at least one client or one AP in the overlap area $A(d)$. Thus, given that the distance between two APs is d, the probability that the respective edge exists in the CG is given by:

$$P_{edge}(d) = \begin{cases} 1 & \text{if } 0 \leq d \leq R \\ 1 - e^{-(\lambda_c + \lambda_{AP})A(d)} & \text{if } R < d \leq 2R \end{cases} \tag{4}$$

B. Neighbor distance distribution

We assume that APs are PPP-distributed. Let X be the random variable representing the distance between an AP A and a random neighbor AP B picked from a $2R$-radius disk centered at A. The CDF of X is given by

$$F(x) = P(X \leq x) = \frac{\pi x^2}{4\pi R^2} = \frac{x^2}{4R^2} \tag{5}$$

and its PDF is given by

$$f(x) = \frac{x}{2R^2}\ . \tag{6}$$

C. Number of CG edges

From (4) and (6), it follows that from the N_{pe} cases of cell overlap (potential edges), the number of actual CG edges is:

$$N_e = \int_0^R N_{pe} f(x) dx + \int_R^{2R} N_{pe} f(x) P_{edge}(x) dx. \tag{7}$$

The first integral in (7) corresponds to neighbor Wi-Fi cells where the APs are within range of each other and, therefore, the respective edge exists in the real CG, even if no clients are located in the overlap area (Type-1 edges). The number of d-distance such edges is $N_{pe} f(d)$.

The second integral refers to Type-2 edges, where, in order for an edge to be part of the CG, at least one client or AP needs to be located in the overlap area; otherwise, no nodes are affected and the respective edge is ignored. The number of d-distance Type-2 edges is $N_{pe} f(d) P_{edge}(d)$.

D. Detection probability

We derive the probability that an edge gets detected when dealing with the attack scenarios that we have described in Section III and applying the respective filtering mechanisms. Edges finally accounted for in the CG are those reported by a sufficient number of clients. The edge detection probability is also a function of the distance d between the respective APs.

1) Attack scenario 1: An edge exists in the *filtered* CG if it is reported by at least 2 clients. Detection probability depends on the ratio of truthful reporters, since attacker reports are by definition filtered out (see Observation 1). Each reporter is truthful with a fixed probability P_t, thus the intensity of the distribution of *truthful reporters* is $\lambda_c P_t$ and the probability that more than 1 truthful reporters are located in the overlap region $A(d)$ between two APs is

$$\begin{aligned} P_d(d) &= Pr\{\text{more than 1 truthful reporters in area A(d)}\} \\ &= 1 - e^{-\lambda_c P_t A(d)} \lambda_c P_t A(d) - e^{-\lambda_c P_t A(d)}. \end{aligned} \tag{8}$$

2) Attack scenario 2: We consider two Poisson Point Processes to distribute different types of clients (the main PPP with intensity λ_c is split); non-roamers are distributed with intensity $\lambda_{nr} = (1 - P_{roam})\lambda_c$ and truthful roamers are distributed with intensity $\lambda_r = P_{roam} P_t \lambda_c$, where P_{roam} is the (constant) probability that a client is a roamer and P_t the probability that a *roamer* is truthful. All non-roamers are assumed truthful.

The mean number of roamers per AP determines the average roamer report weight. We assume dense AP deployments where there is full wireless coverage. Thus, every client is in range of at least one AP. On average, there are $\overline{N_r} = \frac{N_c}{N_{AP}} P_{roam}$ roamers associated with an AP, where N_c is the total number of clients and N_{AP} the total number of APs.

To calculate the probability that an edge is detected, let X and Y denote the random variables representing the number of non-roamers and the number of roamers reporting an edge respectively. Then, this probability is given by:

$$\begin{aligned} P_d(d) &= Pr\{X > 0\} + Pr\{Y > \lfloor \overline{N_r} \rfloor\} Pr\{X = 0\} \\ &= 1 - e^{-\lambda_{nr} A(d)} + (1 - \sum_{i=0}^{\lfloor \overline{N_r} \rfloor} e^{-\lambda_r A(d)} \frac{(\lambda_r A(d))^i}{i!}) e^{-\lambda_{nr} A(d)} \\ &= 1 - \sum_{i=0}^{\lfloor \overline{N_r} \rfloor} e^{-(\lambda_r + \lambda_{nr}) A(d)} \frac{(\lambda_r A(d))^i}{i!}. \end{aligned} \tag{9}$$

E. Percentage of detected edges

In the absence of *false positives*, i.e., fake edges in the reported/filtered CG (see Observations 1 and 2), our performance metric is the percentage of true edges that are discovered. In Section IV-C, we calculated N_e, i.e., the total number of CG edges. Using a similar analysis, we can calculate the total number of detected ones. Of the $N_{pe} f(d)$ d-distance Type-1 CG edges, the number of discovered ones is $N_{pe} f(d) P_d(d)$. Also, of the $N_{pe} f(d) P_{edge}(d)$ d-distance Type-2 edges, the number of discovered ones is $N_{pe} f(d) P_{edge}(d) P_d(d)$. In total, the number of discovered edges (N_d) is given by

$$N_d = \int_0^R N_{pe} f(x) P_d(x) dx + \int_R^{2R} N_{pe} f(x) P_{edge}(x) P_d(x) dx. \tag{10}$$

The performance of our mechanism is thus given by

$$\begin{aligned} R &= \frac{N_d}{N_e} \\ &= \frac{\int_0^R f(x) P_d(x) dx + \int_R^{2R} f(x) P_{edge}(x) P_d(x) dx}{\int_0^R f(x) dx + \int_R^{2R} f(x) P_{edge}(x) dx}. \end{aligned} \tag{11}$$

V. Numerical results

In this section we present the results of the performance analysis of our system. Using AP density information from a 2007 study [2] and population density data from the 2000 US census, we simulated Wi-Fi deployments corresponding to the Manhattan and Boston metropolitan areas and measured the efficiency of our filtering schemes for varying numbers of attackers for each attack scenario.

For reasons of scalability, we opted to develop our own custom simulator. APs and clients are PPP-distributed on a $1km \times 1km$ terrain, and AP transmission range is fixed to $100m$. Each client submits a report (which may be fake) about the APs within range and the reported CG is built, filtered and compared to the actual CG. It should be noted that we do not address user mobility; clients are assumed stationary. Simulator settings are summarized in Table I.

TABLE I
Simulator settings

	Manhattan	Boston
AP density	1854/km^2	729/km^2
Client density	27490/km^2	4947/km^2
Cell radius	100m	
Terrain size	$1km^2$	

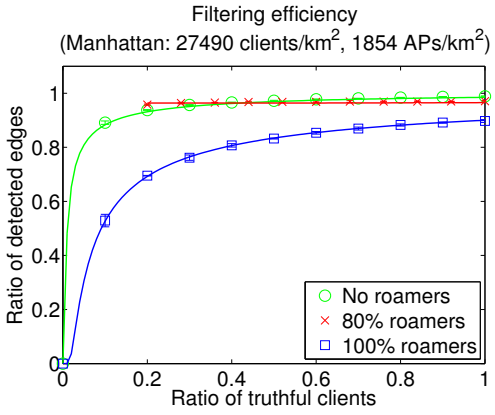
For each of the following experiments, we plot simulation results (points) and the results from our analysis (curves). Each data point is the mean of 5 iterations (i.e., simulations of different random topologies with the same characteristics as to client/AP densities and percentages of roamers and attackers). We have calculated 99% confidence intervals, which are, however, too narrow to be easily discernible.

We plot 3 curves for 3 different experiments for each setting; the "no roamers" curve represents the first attack scenario, where adversaries are independent. The other 2 curves represent the second scenario, with different roamer percentages each. All attacking roamers attached to an AP form a single colluding group. Note that in the experiments involving 80% roamers, the remaining non-roamers (20% of the clients) are truthful and trusted and account for the very high performance.

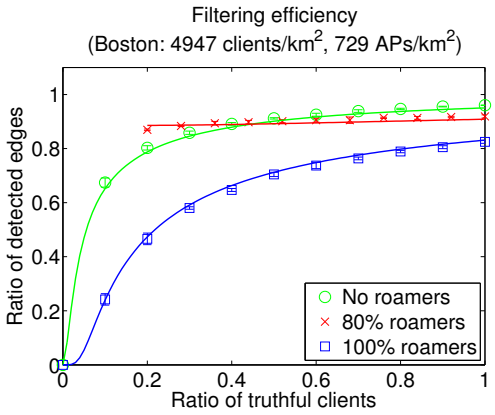
Fig. 2a shows the percentage of detected CG edges in a setting corresponding to the Manhattan area, characterized by very high AP and client density. Fig. 2b depicts a sparser setting (City of Boston). Even with very large attacker ratios, our simple mechanisms manage well in discovering network topology, at the same time filtering *all* fake reports. One should notice that the drop in client density is followed by a drop in performance. This could become more obvious in sparser (as far as clients are concerned) deployments and especially in the presence of many non-trusted users (even if they are truthful). Although our system performs well in the scenarios we target, this observation leads us to believe that in order for it to be effective when client density is low, less strict and more adaptive report evaluation mechanisms would be necessary.

VI. Related work

Our model is very similar to the one introduced by Mishra et al. [3] for solving the channel assignment problem. To address interference asymmetry between APs and to capture client and AP load, necessary for performing power control, Ahmed and Keshav [5] use an *annotated conflict graph* with additional client vertices, undirected client-AP *association* edges and directed *interference* edges. Another approach [6] is to apply a *conflict set coloring* formulation to the problem of jointly performing channel assignment and load balancing, where, for each client, there is a *range* set (APs in range) and an *interference* set (APs not in range, but with interfering clients associated to them) and the objective is to minimize interference suffered by each client. A alternative representation of interference is by modeling a link between two nodes as a graph vertex and placing an edge between two vertices if the respective links are conflicting [7].



(a)



(b)

Fig. 2. Filtering efficiency in a very high-density random topology simulating the Manhattan area (a) and in a lower-density area simulating the City of Boston (b).

Numerous approaches aim at adding reconfiguration features to Wi-Fi networks. Their common denominator is the need to collect information from the wireless environment. The next step is to apply sophisticated reconfiguration mechanisms by means of frequency selection [3], [6], [8], power control [5], [9], rate adaptation [10], adaptation of the carrier sensing threshold [9], [11], or their combinations. Murty et al. [12] focus on enterprise WLANs where most wireless management decisions are pushed to the infrastructure. Again, they need measurements from clients and APs to perform them. Our work serves in improving the robustness of information collection and providing valid input to the above mechanisms.

Most of the above schemes [3], [6], [10], [11], [12] require client participation for the collection of input for the respective spectrum sharing mechanisms. In a different context, Pang et al. [13] present a collaborative service, offering information about AP capabilities, which can be used for improved AP selection. This information is built by user-provided reports. Importantly, they propose reporting protocols which preserve user privacy while limiting fake reporting. Some of their mechanisms are applicable to our system and can limit practical attacks like multiple reports from the same attacker.

Our work is related to the process of distributed spectrum sensing in Cognitive Radio Networks (CRN). In a typical CRN scenario, *secondary* (i.e., unlicensed) users collectively monitor spectrum usage to detect the presence of *primary* (i.e., licensed) ones. Recent standardization efforts within the IEEE 802.22 working group [14] also focus on spectrum sensing. In this context, Chen et al. [15] study two potential attacks, namely Incumbent Emulation, where an adversary's CR transmits signals that emulate the characteristics of a primary user's transmissions, and Spectrum Sensing Data Falsification. In the latter, which is similar in spirit with the attacks we address, adversaries submit fake sensing data to the collecting entity to tamper with the sensing decision.

VII. Conclusion

Wi-Fi topology discovery is an important first step for sophisticated interference mitigation schemes, since it provides the input to processes such as frequency selection or power control. Our goal is to exploit the inherent benefits of delegating the task of carrying out measurements and reporting on network topology to end-users, at the same time dealing with potential fraudulent reporting. In this paper, we have shown that in today's high density urban wireless deployments and given that client density is relatively high, it is possible to combat such attacks with simple countermeasures.

References

[1] "Skyhook wireless," http://www.skyhookwireless.com.

[2] K. Jones and L. Liu, "What Where Wi: An analysis of millions of Wi-Fi access points," in *Proc. IEEE PORTABLE 2007*, May 2007.

[3] A. Mishra, S. Banerjee, and W. Arbaugh, "Weighted coloring based channel assignment for WLANs," *ACM SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 9, no. 3, pp. 19–31, 2005.

[4] IEEE 802.11 WG, *IEEE Standard for information technology - Telecommunications and information exchange between systems - Local and metropolitan area networks - Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 1: Radio Resource Measurement of Wireless LANs, IEEE 802.11k-2008*, The Institute of Electrical and Electronics Engineers, Inc., New York, USA, June 2008.

[5] N. Ahmed and S. Keshav, "Smarta: a self-managing architecture for thin access points," in *Proc. ACM CoNEXT '06*, December 2006, pp. 1–12.

[6] A. Mishra, V. Brik, S. Banerjee, A. Srinivasan, and W. A. Arbaugh, "A client-driven approach for channel management in wireless LANs," in *Proc. IEEE INFOCOM 2006*, Barcelona, Spain, April 2006.

[7] K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless network performance," *Wirel. Netw.*, vol. 11, no. 4, pp. 471–487, 2005.

[8] B. Kauffmann, F. Baccelli, A. Chaintreau, V. Mhatre, K. Papagiannaki, and C. Diot, "Measurement-based self organization of interfering 802.11 wireless access networks," in *Proc. IEEE INFOCOM 2007*, May 2007, pp. 1451–1459.

[9] V. Mhatre, K. Papagiannaki, and F. Baccelli, "Interference mitigation through power control in high density 802.11 WLANs," in *Proc. IEEE INFOCOM 2007*, May 2007, pp. 535–543.

[10] G. Judd, X. Wang, and P. Steenkiste, "Efficient channel-aware rate adaptation in dynamic environments," in *Proc. ACM MobiSys 2008*, June 2008, pp. 118–131.

[11] A. Vasan, R. Ramjee, and T. Y. C. Woo, "ECHOS - enhanced capacity 802.11 hotspots," in *Proc. IEEE INFOCOM 2005*, March 2005, pp. 1562–1572.

[12] R. Murty, A. Wolman, J. Padhye, and M. Welsh, "An architecture for extensible wireless LANs," in *Proc. HotNets VII*, 2008.

[13] J. Pang, B. Greenstein, M. Kaminsky, D. McCoy, and S. Seshan, "Wifi-reports: improving wireless network selection with collaboration," in *Proc. ACM MobiSys 2009*, June 2009, pp. 123–136.

[14] IEEE 802.22 Working Group on Wireless Regional Area Networks, http://www.ieee802.org/22/.

[15] R. Chen, J.-M. Park, Y. T. Hou, and J. H. Reed, "Toward secure distributed spectrum sensing in cognitive radio networks," *IEEE Communications*, April 2008.