Report-based topology discovery schemes for centrally-managed Wi-Fi deployments

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

***Abstract*—The density of IEEE 802.11 Wireless Local Area Networks, especially in metropolitan areas, has resulted in almost ubiquitous wireless presence in such environments. Apart from the benefits of increased wireless coverage, though, uncontrolled WLAN deployment has brought to attention interference issues among neighbor wireless networks. Combating interference has become one of the keys to successful wireless infrastructure management. To this end, network topology information, such as information about overlapping cell coverage and client presence is necessary input. We wish to exploit the spectrum sensing capabilities and inherent mobility of clients in order to gather a more up-to-date view of the network topology and a user-perceived view of interference conditions. Thus, we rely on trusted clients to scan for Wi-Fi coverage and report to a centralized entity. When no trusted client reports are available, we resort to AP-based ones. We evaluate the efficiency of the above reporting scheme in accurately discovering AP topology via analysis and simulation for various urban settings with different client and AP densities and compare it with pure client-based and pure AP-based approaches.**

***Index Terms*—Wi-Fi, Network management and control, Interference, Spectrum management**

I. Introduction

In recent years, we have witnessed a clear trend towards open wireless access, expressed, to a significant extent, with the proliferation of technologies operating in unlicensed spectrum bands, such as IEEE 802.11 or Bluetooth. Advances in the area of Cognitive Radio Networking [1] support this trend and ensure that open wireless systems will be a critical component of the future Internet. The plethora of devices sharing spectrum, especially as we are approaching the *Internet of Things* [2] era, though, raises the need for advanced spectrum management.

Already, with Wi-Fi pervading modern metropolitan areas, unlicensed spectrum scarcity is becoming a reality. While in densely populated urban areas Wi-Fi coverage is no more an issue, continuous unplanned and anarchic Wi-Fi network deployment raises interference issues; for IEEE 802.11b/g there are only 3 non-overlapping frequency bands (channels) on which a Wi-Fi cell can operate, and in the scenarios we study, the probability of coexistence of more than 3 WLANs at the same spot is high.

Therefore, optimizing network operation 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.

The focus of this work is on studying schemes for accurate Wi-Fi topology discovery. The necessary information can be reported by the wireless infrastructure (Access Points) or the clients themselves. There are significant advantages in involving clients in this process, as opposed to leaving this task to the APs. 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. We target centrally managed Wi-Fi networks, such as corporate or campus WLANs or Wi-Fi hotspot-based networks deployed in public spaces by Wireless Internet Service Providers (WISPs). Deployment of such networks is usually planned and their operating parameters can be centrally tuned. This is contrary to the mass of residential (or other autonomous) Wi-Fi networks, which are set up in an uncontrolled manner and their typically novice administrators often leave them operating in default settings.

Most frequency selection, power control or load balancing schemes imply knowledge of network topology and client presence, among others. The main contribution of this work lies in the analytical and simulation-based study of a Wi-Fi topology reporting architecture, which can provide accurate input to such sophisticated spectrum management schemes so as to limit interference. Our architecture exploits reports from trusted clients along with measurements from trusted APs (which are under the operator's full control) when necessary.

The remainder of this paper is structured as follows. In Section II we present our architecture and system model and in Section III we analytically evaluate its performance. Numerical results for realistic urban settings are shown in Section IV. In Section V we review the state-of-the-art in related research areas and in Section VI we discuss various relevant issues and indicate future research directions. We conclude the paper in Section VII.

II. Architecture and system model

We focus on managed Wi-Fi deployments, whose configuration is centrally controlled. For example, imagine a WISP who has set up a number of Wi-Fi hotspots around a city. Users registered with this WISP attach to the WISP's APs, but can also roam around foreign APs which may be providing Internet access.

The purpose of the operator is to collect information from the radio environment around its own APs in order to optimize their operation by tuning parameters such as the transmission power and frequency (channel). To this end, registered users, which are assumed trusted, periodically (or upon request) scan the medium for AP presence and report the results to a central *collector*, responsible for building a Wi-Fi coverage map. Each report contains details about cell operation (channel number, received signal strength, etc.). There are, thus, two classes of clients in our architecture, *trusted* and *foreigners*. Trusted clients are affiliated with the provider and always submit truthful reports. We assume that trusted users are certified by the operator and their reports are authenticated. Note that a managed AP may server both trusted clients and foreigners.

There are cases when no trusted users are associated with a managed AP, though. In such cases, the AP needs to perform a site survey by itself and report to the collector. This improves discovery accuracy in the presence of few trusted clients and implies that the lower bound to the performance of our architecture is that of a pure AP-centric scheme (where all reports are provided by APs).

It should be noticed that, although the APs belonging to the operator are under its full control and their operation can be carefully planned, there are a lot of other APs which may be interfering with the operation of the managed ones. Our reporting scheme aims at revealing such cases of cell overlap, too.

Our architecture is shown in Fig. 1. Only trusted entities (e.g. APs A and C and clients in white) participate in the reporting process. Our system does not require any reporting-related communication between APs and clients; a trusted client, whether attached to a trusted AP or not, submits a report directly to the collector over the Internet. Also, reports that do not involve any managed AP are ignored. When there are no trusted clients associated with a managed AP (see AP C), it is *activated* and submits a report by itself.

We model Wi-Fi topology as a weighted undirected *Coverage Graph (CG)*, where vertices represent APs and edges represent coverage overlap between neighbor Wi-Fi cells. As shown in Fig. 2, there are two cases of overlap. In the first case (*Type-1* edges), two APs are within range of each other. Even if there are no reports about 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 can capture 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]. Based on reports, the aim of our system is to expose as many CG edges as possible.

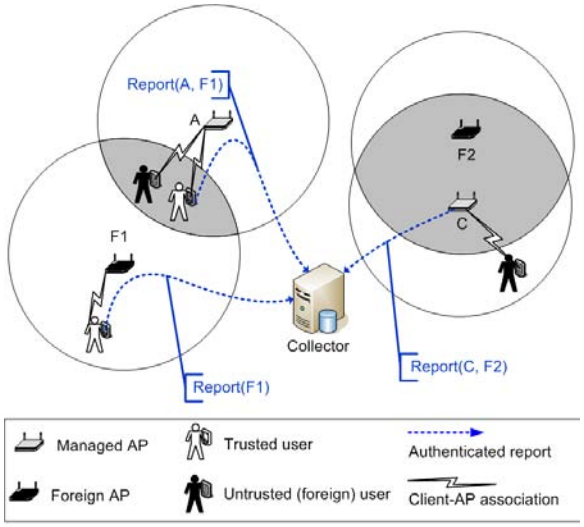


Fig. 1. The proposed reporting architecture. Trusted clients, be they attached to trusted (managed) APs or not, perform a site survey and submit an authenticated report to the collector over the Internet. On the other hand, if a managed AP does not have any trusted clients associated with it (e.g., Access Point C), it is activated and performs a site survey itself and submits an authenticated report to the collector. Foreign clients and APs do not participate in the reporting process.

Our system only considers edges which connect two managed APs or a managed AP and a foreign one. Edges between foreign APs are irrelevant for the operator, since both APs are outside its control and he is unable to resolve such conflict (e.g. by controlling their transmission power).

Fig. 2 shows an instance of the CG where each report contributes a unit to an edge's weight. It should be noted that edges may not always be detected, due to lack of reports. For example, if all APs in Fig. 2 are managed, the A - B edge will not be detected if neither of the two APs is activated (there are no trusted clients in the overlap area to report it).

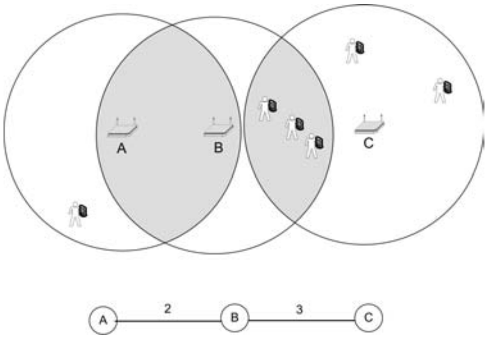


Fig. 2. The Coverage Graph. A-B is a Type-1 edge and B-C is a Type-2 edge.

III. Performance analysis

In this section, we analytically evaluate the topology discovery accuracy of our scheme. Our methodology involves comparing the actual network topology to the detected one and our evaluation metric is the percentage of discovered cases of overlap/interference (percentage of detected CG edges). It should be noted that we only consider edges between managed APs or between a managed AP and a foreign one. The CG does not contain edges between foreign APs, since the latter cannot be controlled and are thus considered irrelevant. 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 cells (intersection of two circles) is given by the following formula:

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

where $d < 2R$ 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

$$P(n,d) = e^{-\lambda_c A(d)} \frac{(\lambda_c A(d))^n}{n!} \tag{2}$$

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 conflicting APs is d, the probability that the respective edge exists in the CG is given by:

$$P_{edge}(d) = \begin{cases} 1 & \text{if } 0 \leq d \leq R \\ 1 - e^{-(\lambda_c + \lambda_{AP})A(d)} & \text{if } R < d \leq 2R \end{cases} \tag{3}$$

B. Neighbor distance distribution

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

$$F(x) = P(X \leq x) = \frac{\pi x^2}{4\pi R^2} = \frac{x^2}{4R^2},\ 0 \leq x \leq 2R \tag{4}$$

and, thus, its PDF is given by

$$f(x) = F'(x) = \frac{x}{2R^2},\ 0 \leq x \leq 2R \tag{5}$$

C. Number of CG edges

From Eq. (3) and (5), it follows that from the N_{pe} potential edges (i.e. cases of cell overlap where at least one of the two APs is managed), the number of actual CG edges is:

$$N_e = \int_0^R N_{pe} f(x)dx + \int_R^{2R} N_{pe} f(x) P_{edge}(x)dx \tag{6}$$

The first integral in Eq. (6) corresponds to Type-1 edges (the respective edge exists in the real CG, even if no clients or APs are located in the overlap area). The number of d-distance such edges is $N_{pe}f(d)$.

The second integral refers to Type-2 edges, where, in order for the 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. Access Point activation probability

Whenever an AP does not have any trusted clients associated with it, it performs a Wi-Fi scan and reports its findings to the collector.

APs and trusted clients are PPP-distributed with intensities λ_{AP} and $\lambda_{tc} = \lambda_c P_{tc}$ respectively, where P_{tc} is the constant probability that a client is trusted. In order to calculate the probability that a trusted AP is activated, we will first derive the distribution of associated trusted clients.

On average, each client can choose among $\lambda_{AP}\pi R^2$ APs which are in range to associate with and we assume that he picks one of them uniformly at random. We consider scenarios where a client has more than one APs available on average, i.e. $\lambda_{AP}\pi R^2 > 1$. Therefore, the intensity of the distribution of associated trusted clients is $\lambda_a = \frac{\lambda_{tc}}{\lambda_{AP}\pi R^2}$. The probability that, of the number of clients within an AP's range, n trusted clients are associated with it is given by

$$\begin{aligned} P_{assoc}(n) &= Pr\{n \text{ trusted clients associated with an AP}\} \\ &= e^{-\lambda_a \pi R^2} \frac{(\lambda_a \pi R^2)^n}{n!} \end{aligned} \tag{7}$$

A managed AP is activated when there are no trusted clients associated with it. From Eq. (7) it follows that the AP activation probability is

$$P_{act} = P_{assoc}(0) = e^{-\lambda_a \pi R^2} = e^{-\frac{\lambda_c P_{tc}}{\lambda_{AP}}} \tag{8}$$

E. Edge discovery probability

1) Mixed reporting scheme: Consider a CG edge corresponding to two APs with distance d. In order for this edge to be discovered, an authenticated report including it should be submitted to the collector. In the following, we calculate the edge detection probability for Type-1 and Type-2 edges in the mixed reporting scheme which we have proposed (reports by trusted clients and activated APs).

a) Type-1 edges: For a Type-1 edge to be discovered, at least one of the following conditions should be met:

- At least one of the two APs is activated
- There is at least one trusted client located in the overlapping region
- At least one other (managed) activated AP is located in the overlapping region

Therefore, the Type-1 edge discovery probability is given by

$$P_d^{(1)} = 1 - (1 - P_{act})^2 e^{-(\lambda_{tc} + \lambda_{act})A(d)} \tag{9}$$

where $\lambda_{tc} = \lambda_c P_{tc}$ is the intensity of the PPP according to which trusted clients are spatially distributed. Also, $\lambda_{act} = \lambda_{AP} P_{tAP} P_{act}$ is the intensity of the PPP corresponding to activated APs (APs are PPP distributed and we split this process using P_{tAP}, the ratio of managed APs over the total AP population, and the probability that an AP is activated, which is given by Eq. (8)).

b) Type-2 edges: As for a Type-2 edge, one of the following should be true:

- At least one (managed) activated AP is located in the overlapping region, or
- at least one trusted client is located there.

Then, the probability that a Type-2 d-distance edge is discovered is given by

$$P_d^{(2)} = 1 - e^{-(\lambda_{tc}+\lambda_{act})A(d)} \tag{10}$$

2) Pure AP-centric scheme: We wish to compare the performance of our mixed reporting scheme with a pure AP-centric one. In such a system, only the APs which are centrally managed scan the medium and provide reports. Using similar arguments, we will derive the probability that an edge is discovered.

a) Type-1 edges: Type-1 edges are always discovered, since at least one of the two APs is managed and all managed APs submit reports. Therefore,

$$P_d^{(1)}(d) = 1, \forall d \in [0, R]. \tag{11}$$

b) Type-2 edges: In a pure AP-based reporting scheme, in order for a Type-2 edge to be discovered, at least one managed AP should be located in the overlapping region. Therefore, the discovery probability for a d-distance Type-2 edge is given by

$$P_d^{(2)}(d) = 1 - e^{-\lambda_{tAP}A(d)} \tag{12}$$

where $\lambda_{tAP} = \lambda_{AP}P_{tAP}$ is the intensity of the PPP for distributing managed APs.

3) Pure client-centric scheme: To complete our analysis, we consider the other extreme, i.e. a pure client-centric scheme, where only trusted clients submit reports and APs do not participate in the reporting process. Again, we derive the probability that a d-distance Type-1 or Type-2 edge is discovered.

Since APs do not contribute reports, an edge (be it Type-1 or Type-2) is discovered if at least one *trusted* client is located in the overlapping region. Thus, the edge discovery probability is

$$P_d^{(1)}(d) = P_d^{(2)}(d) = 1 - e^{-\lambda_{tc}A(d)} \tag{13}$$

where λ_{tc} is the intensity of the PPP according to which trusted clients are distributed.

F. Percentage of detected edges

Our performance metric is the percentage of detected CG edges. There are cases when an edge is missed. This happens when there are no trusted clients or activated APs located in the overlapping region. In Section III-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 $N_{pe}f(d)P_d^{(1)}(d)$. Also, of the $N_{pe}f(d)P_{edge}(d)$ d-distance Type-2 edges, the number of discovered ones is $N_{pe}f(d)P_{edge}(d)P_d^{(2)}(d)$. In total, the number of discovered edges (N_d) is given by

$$N_d = \int_0^R N_{pe}f(x)P_d^{(1)}(x)dx + \int_R^{2R} N_{pe}f(x)P_{edge}(x)P_d^{(2)}(x)dx \tag{14}$$

The performance of our mechanism is thus given by

$$\begin{aligned} R &= \frac{N_d}{N_e} \\ &= \frac{\int_0^R f(x)P_d^{(1)}(x)dx + \int_R^{2R} f(x)P_{edge}(x)P_d^{(2)}(x)dx}{\int_0^R f(x)dx + \int_R^{2R} f(x)P_{edge}(x)dx} \end{aligned} \tag{15}$$

IV. Numerical results

In this section we present the results of the performance analysis of our system. We have used AP density information from a 2007 study [4] and population density data from the 2000 US census for various US metropolitan areas. In each case, we assume that a percentage of the total AP population is centrally managed (P_{tAP}; see Section III) and measure the efficiency of our system (the R metric; see Section III-F), i.e. the percentage of CG edges that involve these APs which are discovered under varying numbers of trusted users (P_{tc}; see Section III). Table I summarizes client and AP densities for our target scenarios. We assume that APs have a fixed transmission range of $100m$.

TABLE I
AP and client densities

	AP density	Client density
Manhattan	1854/km^2	27490/km^2
Boston	729/km^2	4947/km^2
San Francisco	326/km^2	6688/km^2
Seattle	395/km^2	2755/km^2
Atlanta	142/km^2	1552/km^2
Las Vegas	109/km^2	1604/km^2

A. Performance of a mixed reporting scheme

Fig. 3 shows the results of our analysis. In each subfigure, we plot 4 curves, each one demonstrating the efficiency (the R metric) of our system for a different percentage of managed APs (the P_{tAP} parameter in our analysis) over the total AP population and as the percentage of trusted clients (P_{tc}) increases.

As expected, the higher the number of trusted users, the better the discovery accuracy. When both the density of managed APs (see the 10% or 40% curves) and that of trusted clients is low, the performance of our scheme is decreased. However, even with few trusted clients, in most cases (and especially in dense Wi-Fi deployments, such as that of the city of Manhattan) our system manages well in discovering network topology.

Notice that as P_{tc} increases, the 4 curves converge. This happens because with high P_{tc} values, the AP activation

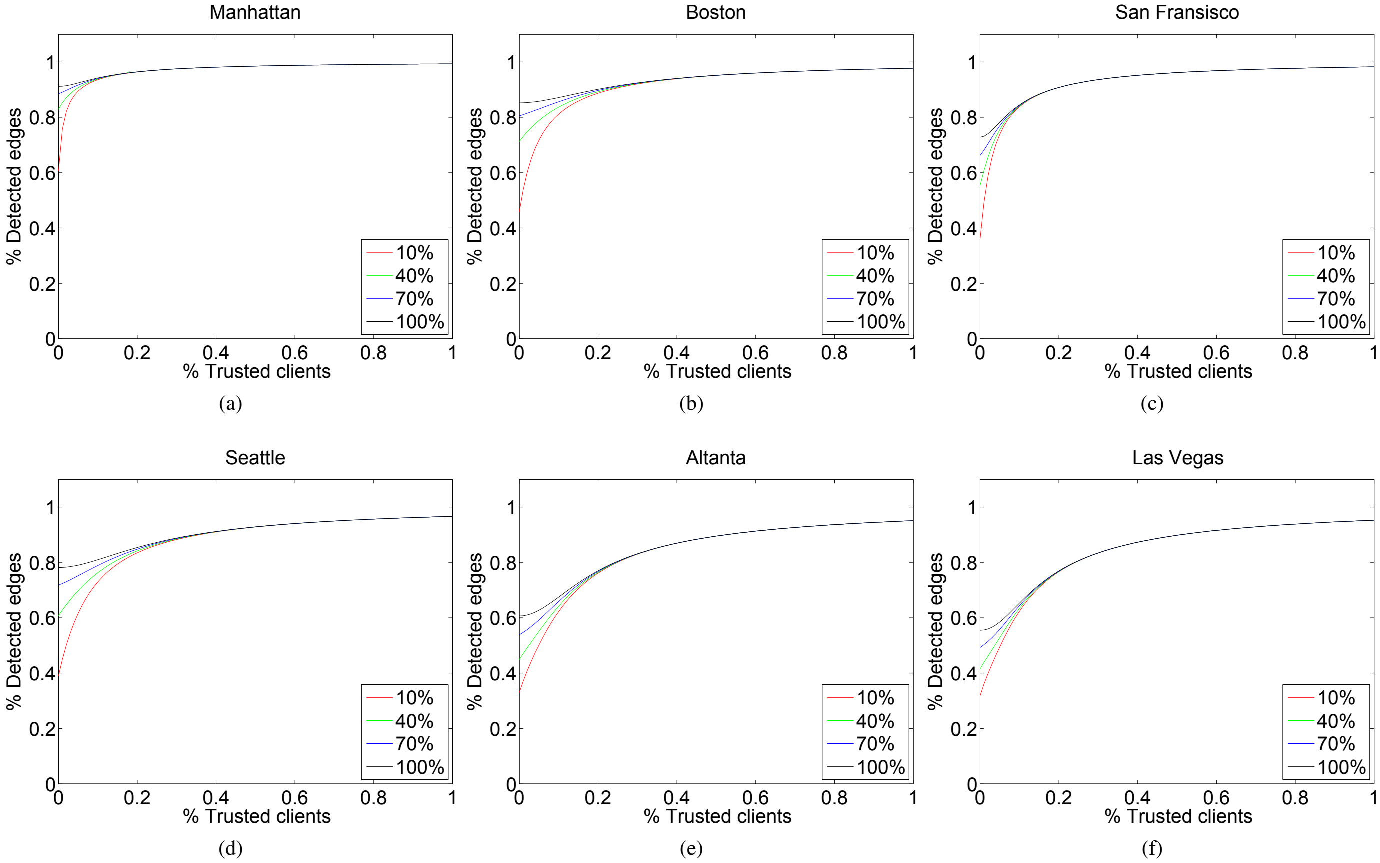


Fig. 3. Topology discovery accuracy in the presence of increasing numbers of trusted clients. The more the trusted clients, the fewer the APs activated. In each subfigure, we plot the efficiency of our scheme (y axis) for increasing percentages of trusted clients (x axis) for 4 different percentages of managed APs over the total AP population.

probability of Eq. (8) is minimized (when trusted client density is high, the probability that no trusted clients are attached to a managed AP is small). This, in turn, means that and the edge detection probability, as shown in Eq. (9) and Eq. (10), only depends on the density of trusted clients, since $(1-P_{act})^2 \rightarrow 1$ and $\lambda_{act} \rightarrow 0$.

B. Performance of pure AP-centric and client-centric schemes

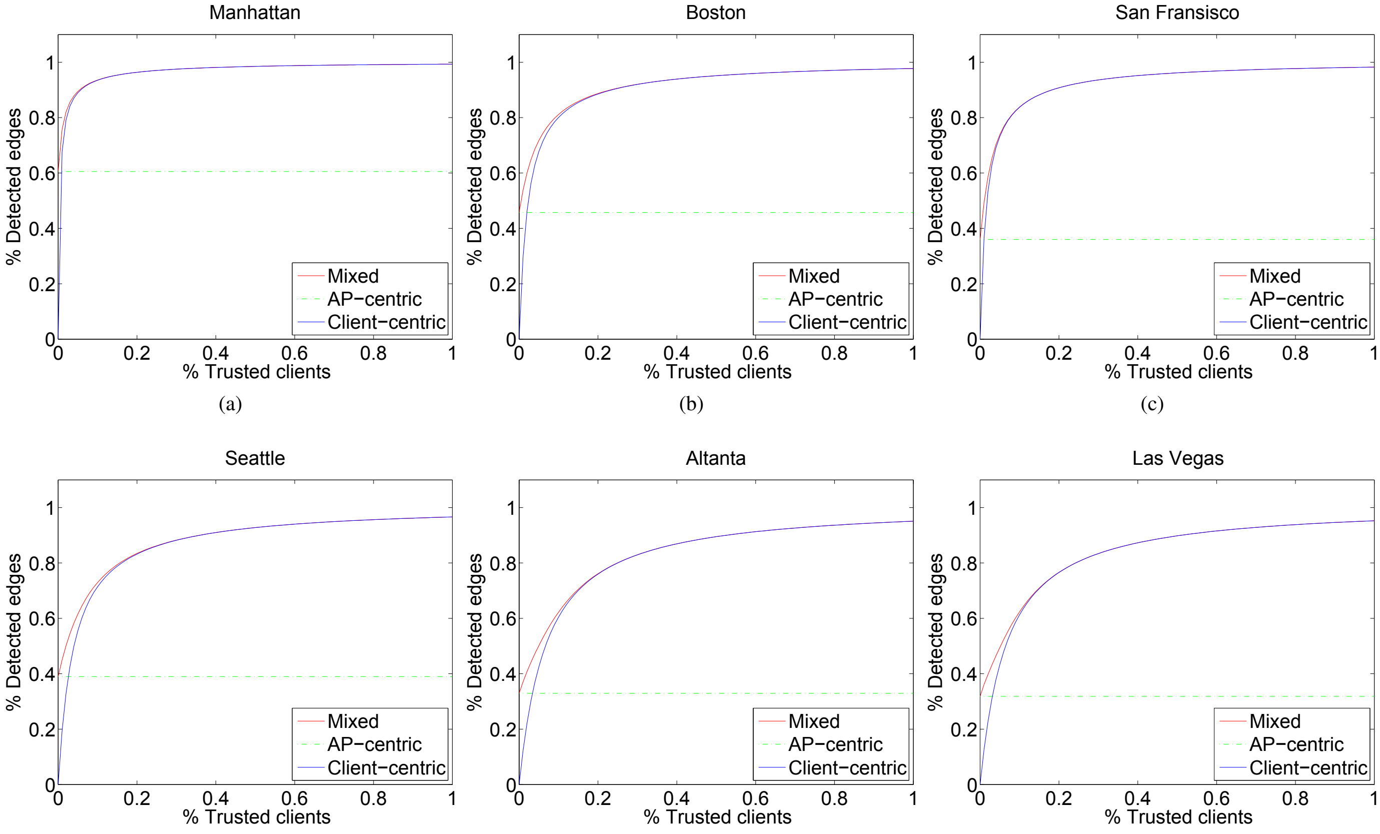
In this section we quantify the advantages of a mixed reporting scheme over pure client-based and, more importantly, pure AP-based ones. In a client-based scheme, all trusted clients submit reports, but APs do not participate in this process. In Fig. 4, we present the efficiency (the R metric) of our scheme for the 6 urban settings we study, fixing the ratio of managed APs over the total AP population to $P_{tAP} = 0.1$.

As our results indicate, the performance benefits of a mixed scheme are more evident when (trusted) client density is low (below 20%, in these cases), since measurements from APs are useful when few trusted clients are in place. However, when client density is high, the performance of a pure client-centric scheme converges to that of a mixed scheme. Again, this happens because AP activation probability, shown in Eq. (8), is minimized, thus the mixed scheme behaves just like the pure client-centric one.

On the other hand, in a pure AP-centric scheme, topology discovery accuracy only depends on AP density. The higher the number of managed APs, the more the CG edges that are discovered. It should be noticed that the percentage of detected edges does not depend on the number of trusted clients, since the latter do not participate in the reporting process. The percentage of edges discovered by an AP-centric scheme is a lower bound to the performance of the mixed one (when there are no trusted clients, thus all managed APs are activated). This is demonstrated by the horizontal lines in Fig. 4.

C. Simulation-based evaluation

To demonstrate the accuracy of our analysis, we have also performed a set of simulations. For reasons of scalability (especially for dense deployments), we have programmed a custom simulator. As in our analysis, we have simulated 6 urban settings with different AP and client densities and with varying percentages of managed APs and trusted clients. Clients and APs are PPP-distributed on a $1km^2$ terrain and their densities are shown in Table I. Again, we assume a fixed $100m$ cell radius. Our simulation results, presented in Fig. 5 (Manhattan area) and Fig. 6 (Las Vegas), are confined to a few representative cases for reasons of brevity. In each figure, the percentage of managed APs is $P_{tAP} = 0.1$. In all the



(a) (b) (c)

(d) (e) (f)

Fig. 4. Comparison of three reporting schemes. In each figure, the topology discovery accuracy in the presence of increasing numbers of trusted clients and assuming that 10% of the AP population are managed is shown. The horizontal curves represent the performance of a pure AP-centric scheme, which also constitutes a lower bound on the performance of the mixed scheme. As it seems, the mixed scheme offers benefits over a pure client-centric one when the number of trusted clients is small (less than 20% over the total client population).

scenarios we have studied, simulation results closely match the analytical ones. Each data point is the mean of 10 iterations, presented with 99% confidence intervals.

V. Related work

A. Wi-Fi topology models

To represent Wi-Fi topology, we have adopted a weighted graph model similar to the one introduced by Mishra et al. [3]. They solve the channel assignment problem as a weighted vertex coloring one. The input graph is composed of vertices corresponding to APs and edges denoting interfering neighbor APs. Edge weights are a function of the number of clients associated with the corresponding APs and affected by interference.

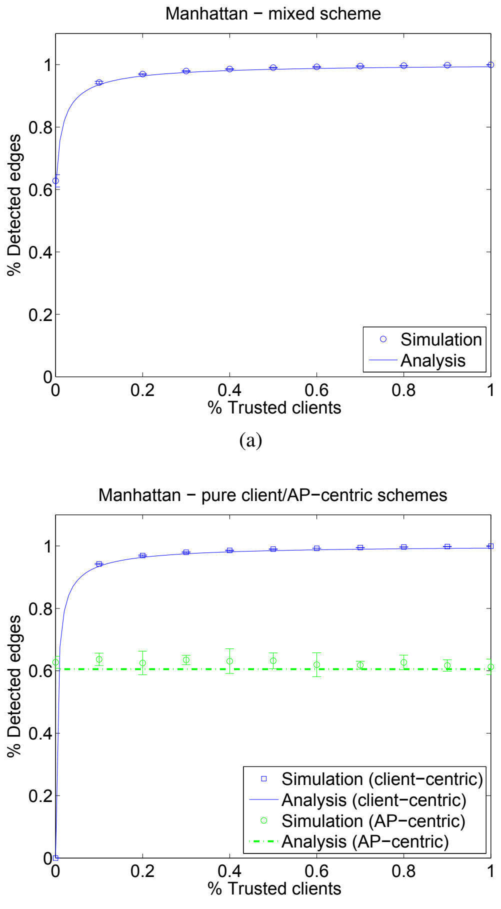
There are many variants of this model. Ahmed and Keshav [5], for instance, to address interference asymmetry between APs and to be able to capture client and AP load, necessary for performing power control, use an *annotated conflict graph*, which includes additional client vertices, undirected client-AP *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 possible representation of interference conditions is by modeling a link between two nodes as a graph vertex and place an edge between two vertices if the respective links are conflicting [7].

B. Wi-Fi reconfiguration schemes

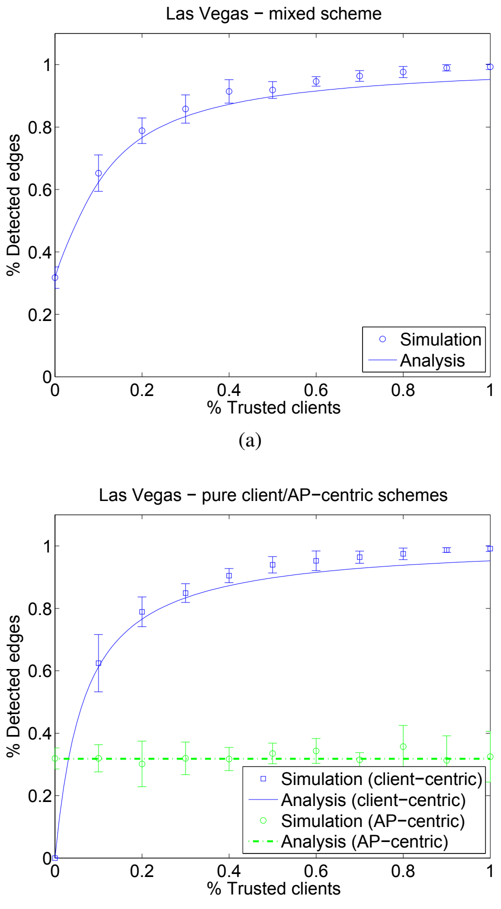
Numerous approaches aim at adding reconfiguration features to Wi-Fi networks for more efficient spectrum utilization. All these schemes assume that spectrum usage information is available. The next step is to apply sophisticated reconfiguration mechanisms by means of frequency selection [3], [6], [8], power control [5], [9], rate adaptation [10], adaptation of the carrier sensing threshold [9], [11], or their combinations. In a similar spirit with our work, Murty et al. [12], [13] focus on enterprise WLAN environments where most wireless management decisions are pushed to the infrastructure and measurements from clients and APs are necessary to perform them. Our work focuses on studying the accuracy of the information collection process and its ability to capture network topology under various AP and client densities in



(a)

(b)

Fig. 5. Simulation results for the three schemes for a setting simulating the Manhattan area, assuming that 10% of the total AP population are managed.



(a)

(b)

Fig. 6. Simulation results for the three schemes for a setting simulating the Las Vegas area, assuming that 10% of the total AP population are managed.

urban settings.

C. *Involving users in information collection*

Most of the above schemes [3], [6], [10], [11], [12], [13] require client participation for the collection of input for the respective spectrum sharing mechanisms. In a different context, Pang et al. [14] present a collaborative service offering information about AP capabilities, which can be used for improved AP selection. This information is built by reports from users themselves. Importantly, the authors propose reporting protocols that preserve user privacy, at the same time limiting fake reporting.

D. *Spectrum sensing in Cognitive Radio Networks*

Our work is related to the process of spectrum sensing in Cognitive Radio Networks (CRN) [15]. In a typical CRN scenario, *secondary* (i.e., unlicensed) users, independently or collectively, monitor spectrum usage to detect the presence or absence of *primary* (i.e., licensed) ones. Recent standardization efforts within the IEEE 802.22 working group [16] also focus on spectrum sensing issues.

VI. Discussion and future work

A. *Implementation considerations*

The information collection system we have described is straightforward to implement with current off-the-shelf wireless equipment. Considering the operation of reporting entities, there is no need to develop a communication protocol between clients and APs, since the former directly communicate their measurements to the collector. If we considered a different system model, though, with a more distributed implementation in mind where each AP controls the reporting process itself and requests/collects client-based measurements, the recently standardized IEEE 802.11k [17] protocol could be used.

One of the controllable parameters in client operation is the frequency of scanning and reporting to the central entity. In prior work [18] we have shown via testbed experiments that frequent scanning for Wi-Fi presence may negatively affect the *Quality of Experience* of real-time applications, such as VoIP. However, realistic scanning/reporting frequencies (in the order of one scan per minute or more) should not have significant performance effects.

B. *Authentication vs privacy*

We rely on trustworthy reporters to reveal network topology. We thus need to ensure that reports from untrusted sources are excluded from our system, so report authentication is necessary. To this end, standard cryptographic techniques relying on Public Key Infrastructures could be used, and the fact that our system is centralized facilitates this from an architectural point of view. However, report authentication may contradict with user privacy requirements; reporting to the collector reveals the user's approximate location. It is thus important to study report anonymization techniques which do not compromise the system's robustness against potential attacks. This tradeoff and relevant cryptographic techniques to address it have been studied by Pang et al. [14].

C. Exploiting untrusted reports and the need for security measures

As evident from our performance analysis, topology discovery accuracy in our architecture increases as the density of *trusted* users increases. In low-density scenarios the performance of our scheme could be improved if we exploited reports from non-trusted users, too. In this case, though, we will have to tackle cases of fraudulent reporting. Therefore, effective countermeasures should be developed in order to filter fake reports. In our prior work [19] we addressed the issue of attacks to the reporting process and proposed simple filtering rules for some particular adversarial scenarios, albeit under a different system model where only clients participated in the reporting process and there were different trust assumptions.

D. Addressing user mobility

We have yet only considered stationary users. The implications of user mobility on practical aspects of our system, such as when to perform reports, how to tackle with the dynamics of user movement and how to represent interference conditions incorporating the time dimension remain open issues.

E. An integrated architecture and its extended performance evaluation

We envisage an integrated architecture where advanced spectrum management decisions and proper infrastructure reconfiguration will be taking place based on the information from the underlying reporting layer. It is therefore significant to measure the efficiency of the reporting process in terms of network and application oriented metrics, such as application throughput or user Quality of Experience, when the aforementioned mechanisms are in operation. To this end, applying Wi-Fi reconfiguration schemes (such as channel selection algorithms) on top of the proposed reporting layer is part of our work in progress.

VII. Conclusion

In this work we studied the performance of reporting-based Wi-Fi topology discovery schemes, focusing on centrally managed Wi-Fi deployments, where the infrastructure (APs) and some clients can be assumed trustworthy. Our evaluation reveals that it is beneficial to involve trusted clients in the reporting process, since, apart from the fact that, this way, a user-perceived view of interference conditions is acquired, Wi-Fi coverage can be more accurately discovered. We have shown via analysis and simulation that a mixed reporting scheme where AP-based measurements are also used when no trusted client-based reports are available is more efficient than pure AP-centric or pure client-centric schemes, especially when the population of trusted users is low and the Wi-Fi deployment is sparser. Our findings also lead us to believe that, under such circumstances, reports from non-trusted users would also help, given that appropriate security measures are in place to counter fraudulent reporting.

References

[1] J. Mitola and G. Q. Maguire, "Cognitive radio: making software radios more personal," *Personal Communications, IEEE*, vol. 6, no. 4, pp. 13–18, 1999.

[2] *ITU Internet Reports 2005: The Internet of Things*, ITU, 2005.

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

[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] R. Murty, J. Padhye, A. Wolman, and M. Welsh, "DYSON: An architecture for extensible wireless LANs," in *Proc. ACM SIGCOMM 2009 Poster Session*, August 2009.

[14] 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.

[15] T. Yücek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Communications Surveys and Tutorials*, vol. 11, no. 1, 2009.

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

[17] 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.

[18] P. A. Frangoudis and G. C. Polyzos, "Coupling QoS provision with interference reporting in WLAN sharing communities," in *Proc. IEEE PIMRC 2008 SocialNets Workshop*, September 2008, pp. 1–5.

[19] P. A. Frangoudis, D. I. Zografos, and G. C. Polyzos, "Secure interference reporting for dense Wi-Fi deployments," in *Proc. ACM CoNEXT 2009 Student Workshop*, December 2009.