Structure and Evolution of a Large-Scale Wireless Community Network

Fotios A. Elianos, Georgia Plakia, Pantelis A. Frangoudis and George C. Polyzos
Mobile Multimedia Laboratory
Department of Informatics
Athens University of Economics and Business
{ailianos,plakia,pfrag,polyzos}@aueb.gr

Abstract

In recent years, we have witnessed a trend towards open wireless access, mainly driven by the low cost of IEEE 802.11-capable equipment and its operation in unlicensed spectrum. This trend has many faces; municipalities set up free Wi-Fi hotspots for Internet connectivity in public spaces, while Wi-Fi access is provided as an extra service to customers in other venues, such as restaurants or hotels. Also, in many metropolitan areas worldwide, community-initiated wireless mesh networks have emerged. Their members use inexpensive technologies to build multi-hop wireless networks and communicate autonomously. In this work, we document the structure and evolution of one of the largest community wireless mesh networks worldwide, the Athens Wireless Metropolitan Network (AWMN). We focus on how the network has grown in size, complexity and service offerings in the last few years. To be able to dynamically discover the structure of such a network, we have developed a suitable topology discovery methodology.

1. Introduction

The popularity of IEEE 802.11-based technologies for local area wireless connectivity has resulted in increased wireless coverage, especially in densely-populated metropolitan areas. This is largely attributable to their low cost, ease of installation, configuration and maintenance, and, importantly, to their operation in unlicensed spectrum. The proliferation of this technology has given rise to new service architectures, business models and applications. Commercial operators, academic institutions, or even plain radio communications enthusiasts are building public infrastructures for Internet access based on Wi-Fi. The trend towards open and low-cost wireless access is, thus, evident.

An offspring of this trend is the emergence of *Wireless Community Networks (WCNs)*; in parallel with the deployment of residential, corporate, municipal and campus WLANs, WCNs have appeared as grassroots movements of WLAN enthusiasts, who use inexpensive networking equipment for free interconnection. Thus, they create all-wireless autonomous mesh networks, or open their Wi-Fi access points (APs) for public Internet access, on an altruistic basis or with the anticipation that their service will be reciprocated by other community members. Local factors have also contributed to their success, such as the degree of penetration of fixed broadband access services in the area. People have often resorted to WCNs as an alternative to more expensive broadband access solutions, as is the case for Athens, Greece, where one of the largest WCNs worldwide, the *Athens Wireless Metropolitan Network (AWMN)* operates.

We believe that the properties of such networks are worth studying. To this end, we have focused on AWMN as a large-scale example of a WCN and present a detailed study on its structure and evolution. Our goals and main contributions of this work are the following:

1. To document a large-scale community wireless mesh network and observe its evolution over the years.
2. To devise and apply a topology discovery methodology suitable for community-based wireless mesh networks, after studying relevant Internet-oriented tools.

The remainder of this paper is structured as follows. Section 2 provides background information on WCNs worldwide, as well as on relevant topology discovery approaches. In Section 3 we study the characteristics and evolution of AWMN. In Section 4 we propose a suitable topology discovery methodology for WCNs and apply it to AWMN, before we conclude the paper in Section 5.

2. Background

2.1. Wireless Community Networks

Wireless Community Networks are public wireless access schemes, driven by community, commercial or munic-

978-1-4244-4439-7/09/$25.00 ©2009 IEEE

ipal initiatives. In the first case, which is also the focus of this work, the development of the network is a result of collective efforts of individual volunteers and function on a not-for-profit basis.

Following the above trend, commercial players such as FON [3] have entered the scene, offering mediation services for the development of wireless communities and trying to make a profit out of this service.

In many cases, municipality-controlled wireless APs are set up in public places, offering Internet access to citizens.

As to their architecture, there are two basic alternatives for WCNs. One the one hand, in *wireless mesh* architectures [8, 1, 2, 4, 13, 5] users build a wireless backhaul using Wi-Fi technologies, with nodes having multiple interfaces and potentially directional point-to-point links with one another. Their members enjoy community-wide services, such as VoIP, online games, FTP, Web access and many more. Contrary to mesh-like WCNs, which aim at providing autonomous wireless interconnection among their members, *hotspot-based* [10, 7, 3] community networks typically target mobile users who use wireless hotspots to get access to the Internet. These hotspots are Wi-Fi access points usually attached to fixed broadband lines.

2.2. Topology discovery

Knowing the topology of a network can assist in network troubleshooting and can be used for predicting and controlling the network's spreading rate. From a potential attacker's point of view, knowing the topology of a target network can increase the probability that an attack is successful. There are both active and passive topology discovery methods.

2.2.1 Active discovery

Active topology discovery methods involve probing the network and observing its behavior. Such methodologies and tools will be presented in the following section.

Distributed tracerouting Traceroute reveals the path from a single source to a target host by sending sequences of ICMP packets to the latter, incrementing the TTL value for each successive packet, starting by value of one. This way, at each round, traceroute receives a "time-exceeded" ICMP error message from the next hop en route to the destination.

Traceroute is an excellent tool for topology discovery, but as network complexity increases each packet does not always follow a unique path. Initiating a traceroute procedure from a number of different nodes scattered in the network will reveal a more realistic and complete network view. The question is, what is the ideal number of traceroute sources that must be used? Barford et al. [11] have shown that the marginal utility of an additional tracerouting host is very small. For example one source could discover 4500 different nodes out of 122200 but 8 hosts only managed to increase the discovered nodes by 6000. On the other hand the marginal utility of increasing the number of destination nodes is much higher; the number of discovered hosts increases linearly as destination nodes are added.

Alias resolution Aliases are different IP addresses of a single host. Each of these aliases corresponds to a physical or logical interface. There are three ways to resolve aliases [15]: (i) DNS records, (ii) Routing tables and (iii) IP-ID counters and TTLs.

From DNS records one can make reasonable assumptions on the topology of a network. For example if name1.domain.com and name2.domain.com belong to the same traceroute hop then there it is great probability that those two addresses belong to the same host. As to routing tables, given two router addresses, if we can access the routing tables of each address, then those addresses are aliases if those tables are identical. Finally, a powerful and flexible tool for alias resolution is the IP-ID value of IP datagrams [14]. Originally used as an identifier for assembling fragmented packets, the IP-ID field is nowadays used more as a packet counter. If two packets are sent to two different addresses that are aliases, then those packets should have a difference in IP-ID values by less than 1000. Also they will have equal TTL values. The implementation of IP-ID differs in each operating system so, usually, in addition to IP-ID and TTL, the DNS name is also taken into account. An example of an alias resolution tool is RocketFuel [16], which relies on both DNS records and IP-ID counters.

2.2.2 Passive discovery

Passive discovery methods involve processing existing routing information. There exist registers with published routing data over the Internet, called *looking glasses*. Various organizations such as CAIDA and Route Views, of the University of Oregon, have published data about the Internet topology. Usually passive and active discovery methods are combined for increased accuracy.

3. The Athens Wireless Metropolitan Network

The main purpose of this work is to document AWMN, a large community wireless mesh network. We begin our study looking back to the days of its conception, attempting to reason about its emergence and popularity. We present its architecture and operation and observe its evolution over the last 4 years. We pay attention to the services offered to members of the community, as well as the underlying technologies. Quantitative data presented in this section are based on information available from AWMN's central registrar (WiND database[1]). This registrar is managed by a group of volunteers, who are also responsible for controlling the admission of new members to the network.

3.1. History and evolution

The idea for the development of a community wireless mesh network in Athens was conceived in February 2002, influenced by similar movements worldwide, such as Seattle Wireless [8]. At the time, broadband penetration in Greece was very limited. The creation of a wireless mesh network with a community-oriented character was expected to offer wider broadband coverage and higher data rates.

The first AWMN links were set up in October 2002. Since then the network has kept growing. In the beginning, it was composed of isolated "islets". Key point for the evolution of AWMN was surpassing some physical obstacles (hills and mountains) in the Athens metropolitan area, linking these islets and creating a unified network.

AWMN's growth rate kept increasing until 2006. The network continued to expand, albeit at a declining rate, reaching 2022 nodes, as of mid 2008. This is easily explained by the fact that what was originally one of the highlights of AWMN, i.e. inexpensive broadband connectivity, has now generally became a commodity, with the drop in DSL prices. Still, the self-organizing spirit of WCNs, the opportunity to experiment with wireless technologies and the content and services available to community members keep attracting new members.

Figure 1 depicts the growth rate of the network. It reveals the great dimensions that it has taken, but also the descending number of new nodes during the last couple of years.

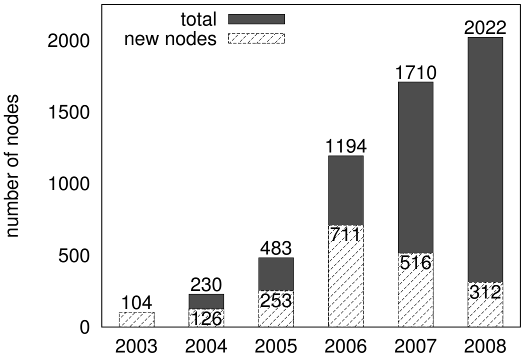


Figure 1. The total number of AWMN nodes and the number of nodes joining each year.

Table 1. Number of links per IEEE 802.11 protocol variant (a/b/g)

Protocol	Number of links
IEEE 802.11a	1486
IEEE 802.11b	2660
IEEE 802.11g	84

3.2. Architecture and technologies

3.2.1 Node types

There are two types of nodes in AWMN. *Backbone nodes* are those upon which the backhaul of the network is built. They are considered more stable and reliable, forming the core of the network. Due to their reliability, they run routing software and provide services to the other nodes. They maintain two or more interfaces and they are interconnected with directional point-to-point links. At the same time, they may also function as *access points* providing connectivity to the rest of the nodes, i.e. the *clients*. Clients do not contribute to the routing process, being the "leaves" of the network. As of mid 2008, there were 515 active backbone and 1504 client nodes. It should be noticed that client connections are typically not ephemeral; clients are usually registered AWMN nodes and their links to APs are fixed.

The distribution of the clients to the backbone nodes is shown in Figure 2. Each subfigure represents a snapshot of the network taken every year, from September 30, 2006. In every period of time, there is a significant number of backbone nodes that do not support any clients. This means that they either do not operate an access point or that, simply, they do not have any clients attached. The remaining usually serve 1 to 5 clients. There are also particular instances of backbone nodes that support 21 to 25 clients.

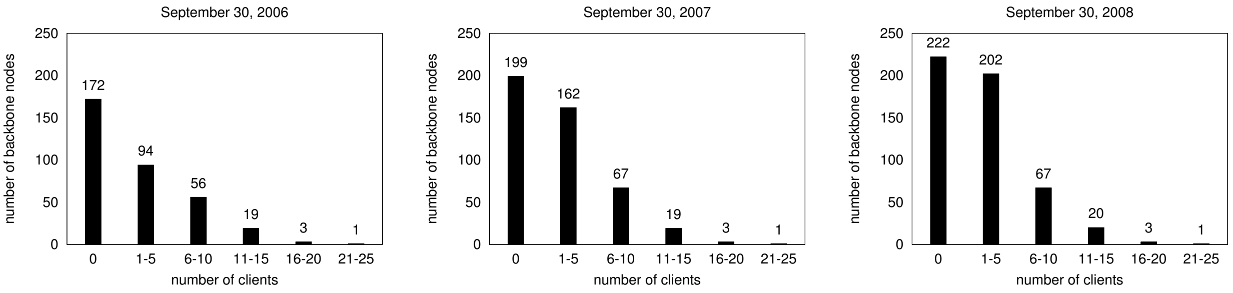
3.2.2 Link types

Backbone nodes are connected with directional point-to-point links, using high-gain directional antennas (attached to Wi-Fi network interfaces), to achieve long-distance links. Today there are 1732 active backbone links. Most of the backbone nodes maintain less than 5 backbone links. In Figure 3 we can observe how many such backbone links each backbone node maintains. Again, we present data for three snapshots of the network, staring from September 30, 2006.

Typically backbone nodes also operate access points, mainly using omni-directional antennas. There are 1432 active access points in the AWMN as of October 2008.

As to link distances, we have found that the shortest wireless link registered is 8 meters and the longest reaches

[1] http://wind.awmn.net



(a) (b) (c)

Figure 2. Distribution of clients to backbone nodes

the distance of 124 kilometers [2]. However, small distance links of about 1 kilometer are preferred. In Figure 4 a distribution of the links as to their distance is shown.

Backhaul links are usually implemented using IEEE 802.11a, which offers high data rates (up to 54Mbps) and less interference than IEEE 802.11b/g. Table 1 shows the number of links using each of the 3 standards. Although there are 2660 links using 802.11b, only 99 of them are backbone ones.

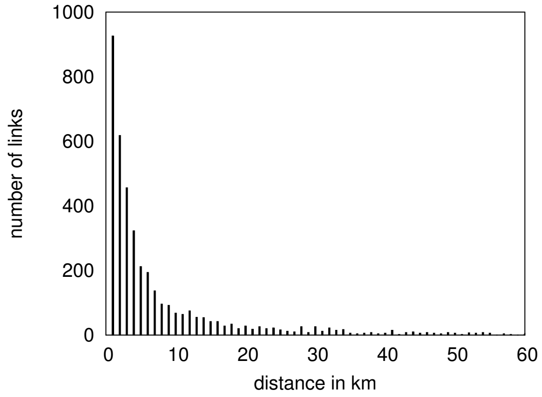


Figure 4. Distribution of link distances

3.2.3 Addressing and routing

Each AWMN node is assigned a private IP address range. Routing is based on BGP (Border Gateway Protocol), with each backbone node and its clients forming a single Autonomous System (AS).

Table 2. Services offered by backbone nodes

Service	Number of nodes
Web site	90
FTP server	84
DNS hosting	64
Video streaming	42
Game server	18
VoIP	15
Time server	12

3.2.4 Services

File sharing (via FTP or Bittorrent) tops the list of the most popular services among AWMN users. VoIP services, video streaming, game servers, websites, and web hosting are offered as well. Importantly, on some occasions, members share their fixed broadband connections with the community, so that Internet access is achieved through WCN-to-Internet proxies. Table 2 shows the number of backbone nodes offering various popular services.

4. Discovering AWMN's topology

4.1. Our goals and challenges

Open wireless communities such as AWMN often grow in an unplanned manner, with frequent topological changes, and given that they are sometimes operated by non-expert users, routing problems (e.g. unreachable nodes) can appear. Their open nature also makes them more vulnerable to malicious actions. Motivated by the above, we carried out a topology discovery process in order to find out whether it is possible to record AWMN's topology by performing sim-

[2] We have calculated link distances from node location information registered by node operators in the WiND database, thus minor inaccuracies may appear.

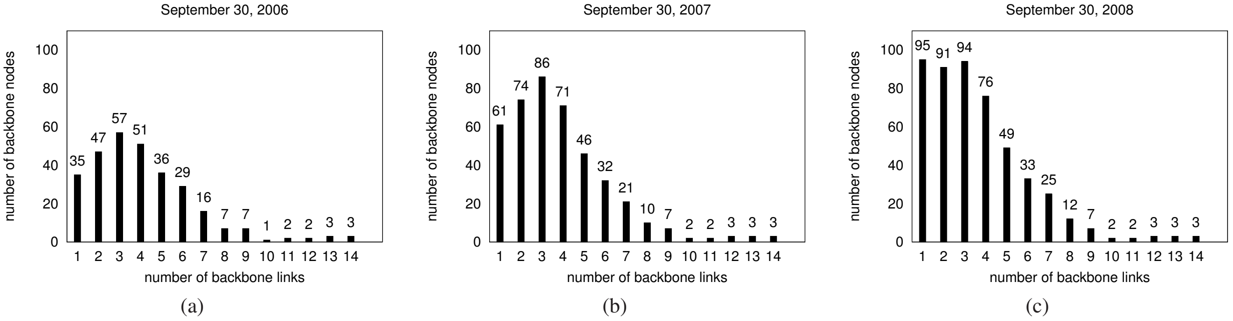


Figure 3. Distribution of backbone links per (backbone) node

ple scans from a single workstation (like a potential attacker would do).

Although performing topology discovery scans in WCN may at first seem similar to methods that have been applied to the Internet, there are some significant differences.

1. In the AWMN there are no official registers, such as *looking glass* servers or published traceroute dumps.
2. There is no consistent DNS naming.

In the Internet, consistent DNS naming can reveal a lot about a network's structure. In AWMN there is a spoken rule that a backbone link between nodes A and B is named `gw-B.A.awmn`. This information can reveal a lot about the topology, but our experience shows that there are lots of cases where this information was inconsistent or missing, as a result of, e.g., outdated DNS servers or typos in DNS records. In most cases, though, a DNS server had a default configuration not allowing reverse DNS lookups, which are very useful to link an IP address to a domain.

4.2. Discovery of active hosts

We used the *nmap* [6] security scanner to find all active hosts in AWMN's IP address range (10.0.0.0 – 10.99.255.255). We discovered 4436 active hosts at most. The scans were conducted during a period of a week by 12 nmap threads started in a round-robin fashion, in 2-hour intervals (to cover a whole 24 hour period). A lot of workstations were online for a very small period of time. Because we were interested in hosts that are routers (backbone nodes), we also scanned AWMN probing for an open 179 TCP port, which is the port the BGP routing daemon listens to. We discovered approximately 2627 router IP addresses.

4.3. AS-level topology discovery

In the AWMN there are no looking glass servers offering public information about routing tables. On the other hand we discovered that many AWMN routers, running software such as the *Quagga* and *Zebra* routing suites, offered Command Line Interfaces (CLI) for remote administration which operated on default username/password settings. We programmed a crawler which connects to these CLIs and fetches the routing tables of each router by sending the appropriate command (`show ip bgp`). From the 2627 router addresses we discovered that 1434 of them had an open port for CLI (usually port 2601 or 2605) and we managed to successfully login with the default username/password to 784 of them. BGP routing tables have the following format:

```
TargetNetwork#NextHop#various_info#AS_path
```

For example:

```
10.2.8.0/24#10.2.79.241#0,2113,72,280
```

means that in order to reach an address in the 10.2.8.0/24 subnet a packet must be routed from 10.2.79.241 of AS 2113. Also the 10.2.8.0/24 subnet belongs to the AS 280 and will follow the path 2113 → 72 → 280. An AWMN AS number is the same as the node's ID. Therefore, backbone node #66 will be advertised at the BGP level as "AS 66". By parsing this information we were able to (i) discover AWMN nodes, and (ii) match IP addresses with AS numbers. We have detected 586 Autonomous Systems and 1579 links among them.

Given that we did not have access to every BGP routing table we needed to verify whether our information is complete. We compared our list of discovered ASs, and thus, AWMN backbone nodes, with the backbone node list of the WiND database and verified that we have discovered approximately 96% of WiND's registered nodes. The rest 4% where offline or unreachable from our spot. Given that we managed to discover almost all ASs, we can safely assume that BGP can't "go wrong" and can offer us all potential paths between those nodes.

4.4. IP-level topology discovery

Our goal here is to find whether an IP address of a router is an alias of an already discovered router or a new one. Tools such as RocketFuel [16] work outstandingly well over the Internet, but in the AWMN they are almost useless because they are also based on DNS records. As mentioned, there are inconsistencies in AWMN DNS records and not all DNS servers support reverse lookup. We discovered that only 1196 out of 2627 IP addresses could be looked up. Given that our data sources are IP addresses, not being able to resolve them to DNS names and solely based on IP-IDs and TTLs, we cannot positively decide if an IP address is an alias or not.

We created a machine-learning-based tool that can be trained to decide whether an IP address is an alias of an already discovered host. First, we selected a consistent set of resolvable IP addresses as training data (approximately half of all discovered hosts). Our tool first sends ICMP packets to the discovered hosts and collects the IP-ID and TTL of each packet. Also, it resolves the DNS name of each IP address and, additionally, groups data with the same DNS domain name and TTL. Each group contains different IP-IDs. We included the TTL as a grouping criterion because there can be routers that connect via intra-domain routing protocols. Those routers can be identified by having same domain name thus belonging to the same AS but also very close TTL values due to the intra-domain hops.

The question to our classifier is whether two groups of our collected data belong to the same router. Our tool can also give an idea of how close IP-ID values should be to assume that those packets belong to the same router. For the Internet this value is less than 1000, but how about a WCN with much less traffic?

We used a data mining method called *record linkage* [12]. Given two *records* (i.e. groups, as described above) we can determine if their distance is small enough to be characterized identical.

In our training data, each pair of records is marked as a "MATCH", if they represent aliases of the same router (typically having the same DNS and TTL and small IP-ID difference), otherwise marked as "NOTMATCHED". We used the J48 algorithm implementation of the *weka* [9] library to classify our data, with approximately 99% accuracy. Thus, we could easily classify data sets for which we had no DNS information, based on the behaviour of our training data.

To verify our results, we connected once more to the BGP CLI of each IP address (wherever we had access). IP addresses which are aliases of a router should have identical routing tables. Our results have shown that our methodology accurately resolves aliases.

5. Conclusion

We attempted to document the structure, characteristics, operation and growth of the Athens Wireless Metropolitan Network, a large-scale WCN. We observed the network's evolution from its inception, back in 2002, to date and studied various of its technical aspects. In line with this work, we developed a topology discovery methodology and tools applicable to similar WCNs. Applying the proposed methodology to AWMN, we managed to discover all existing and reachable ASs and their interconnections, as well as the IP level network topology. Our study gave us valuable insight on the reasons for the emergence and success of the AWMN (but also for relevant actions worldwide), helped us observe trends as to its usage and growth, and can thus assist in making predictions for its future.

References

[1] Athens Wireless Metropolitan Network. `http://www.awmn.net`.

[2] CUWiN - Community Wireless. `http://www.cuwireless.net/`.

[3] FON. `http://en.fon.com`.

[4] Freifunk Berlin. `http://berlin.freifunk.net`.

[5] NetEquality. `http://www.netequality.org`.

[6] Nmap free security scanner. `http://nmap.org`.

[7] NYCwireless. `http://www.nycwireless.net`.

[8] Seattle Wireless. `http://www.seattlewireless.net`.

[9] Weka 3: Data Mining with Open Source Machine Learning Software in Java. `http://www.cs.waikato.ac.nz/ml/weka/`.

[10] Wireless Philadelphia Executive Committee. `http://www.phila.gov/wireless`.

[11] P. Barford, A. Bestavros, J. Byers, and M. Crovella. On the marginal utility of network topology measurements. In *Proc. 1st ACM SIGCOMM Workshop on Internet Measurement (IMW'01)*, pages 5–17, 2001.

[12] K. Goiser and P. Christen. Towards automated record linkage. In *Proc. Fifth Australasian Conference on Data Mining and Analystics (AusDM '06)*, pages 23–31, Darlinghurst, Australia, Australia, 2006. Australian Computer Society, Inc.

[13] R. D. J. Kramer, A. Lopez, and A. M. J. Koonen. Municipal broadband access networks in the Netherlands – three successful cases, and how New Europe may benefit. In *Proc. AccessNets '06*, Athens, Greece, September 2006.

[14] R. Sherwood and N. Spring. Touring the internet in a TCP sidecar. In *Proc. 6th ACM SIGCOMM Conference on Internet Measurement (IMC '06)*, pages 339–344, New York, NY, USA, 2006. ACM.

[15] N. Spring, M. Dontcheva, M. Rodrig, and D. Wetherall. How to resolve IP aliases. Technical Report 04-05-04, University of Washington Dept. CSE, 2004.

[16] N. Spring, R. Mahajan, D. Wetherall, and T. Anderson. Measuring ISP topologies with Rocketfuel. *IEEE/ACM Trans. Netw.*, 12(1):2–16, 2004.