Crowd-Sourcing Wi-Fi Coverage Data to build Self-Mapping Radio Maps

Gareth Ayers
Wireless Network Officer, ISS, Swansea University, Swansea, UK. SA28PP
e-mail: g.j.ayres@swansea.ac.uk
Jason Jones
Tutor/Lecturer, College of Engineering, Swansea University, Swansea, UK. SA28PP
email: j.w.jones@swansea.ac.uk

Paper type
Research paper

Abstract
As Wi-Fi usage continues to grow and become a fundamental technology to facilitate connectivity, services and applications have evolved to fulfill the needs of users. One such service that has evolved to simplify Wi-Fi connectivity and roaming is eduroam. An on-going challenge with eduroam, and enterprise Wi-Fi deployments in general, is keeping up to date with the available locations of coverage on a macro (country/region) and micro (campus/building) scale. Such coverage information (radio maps) could also help facilitate the development and adoption of location based services.

We propose a solution to automatically generate self-mapping radio maps using crowd-sourced Wi-Fi coverage data. We introduce a technique to achieve this using location fingerprinting techniques performed by supplicant configuration tools to produce undirected weighted graphs which can be analysed and interpreted using force-directed graph layout algorithms to produce a self-mapping and self-correcting radio map of Wi-Fi coverage. We will show how we have implemented this at Swansea University using a system we build called LocPriS and the SU1X configuration tool also developed at Swansea. We will also begin to address how multiple institutions could share and combine radio maps using open standardised formats and open data principles.

Keywords—Wireless, eduroam, Visualization, Location-dependent, Graph Theory, Supplicant

1. Introduction

The recent explosion in popularity of mobile devices such as the Apple iPhone and Google Android phones has accelerated the use and deployment of Location Based Services (LBS). LBS's require different levels of granularity of location data with some applications working sufficiently well with low-grained location data such as a town/city while others require much more fine-grained data such as room/building.

In academia a solution to the security and roaming complexities of Wi-Fi users was developed called eduroam. eduroam facilitates the secure roaming of participating institutions users to any other participating institution around the world. Participation of eduroam amongst European institutions is very high, with Asians and American institutions beginning to participate in greater numbers also.

While the adoption of Wi-Fi and eduroam has been significant, the challenge of providing up to date and accurate coverage maps and LBS has been a challenge. Detailed coverage maps would allow for development of LBS as well as facilitate the ease of which roaming users can find Wi-Fi coverage. LBS's depend upon localisation in order to provide a service in context. There has been much research in the area of localisation both indoors and outdoors. The problem of localisation while outdoors has been largely addressed with the use of three technologies, GPS, mobile phone network localisation and Wi-Fi localisation.
Wi-Fi data can usually be obtained from smart phones, and as a result this has begun to emerge as a solution to outdoor localisation. It has become common place for competing mobile OS providers to build their own radio

map of Wi-Fi access points that is used to help assist in the location of smart phones. The Smartphone’s then contribute new information back to the provider to improve and add to the map.

Indoor localisation using Wi-Fi on the other hand is an ongoing problem to which an adequate solution has yet to be adopted by any mainstream smart phone developers. There have been many proposed solutions which will be covered in Section 2.

In this paper we propose a solution to automatically generate self-mapping radio maps using crowd-sourced Wi-Fi coverage data. We introduce a technique to achieve this which is covered in section 4 and uses crowd-sourced location fingerprinting techniques performed by supplicant configuration tools. We will also show how this technique can produce undirected weighted graphs which can be analysed and interpreted using force-directed graph layout algorithms to produce a self-mapping and self-correcting radio map of Wi-Fi coverage that can be used to facilitate LBS's.

We will show how we have implemented this at Swansea University using the SU1X and LocPriS software developed in-house. We will show the level of accuracy and usability currently achievable, and how we can build location based services on top of the radio maps. We will also begin to address how multiple institutions could share and combine radio maps using open standardised formats and open data principles.

At Swansea University we develop and use the SU1X windows supplicant configuration tool to configure our Windows devices for eduroam. We have added new functionality to the SU1X code to allow collection and storage of location fingerprints from devices. We implemented this feature as an opt-out option as part of the SU1X installation. During the period of 17th September 2012 to 12th April 2013 we have collected 282,712 location fingerprints consisting of 3.3 million records. These were collected by 4317 unique devices. We have captured over 4000 eduroam BSSID’s as well as 107 SSIDs, of which only 3 are broadcast by Swansea University’s wireless access points.

Using a supplicant configuration tool such as SU1X or eduroamCAT to capture location fingerprints has some advantages in that users see building the radio map as part of the connecting process and helping improve network performance, and the software used can collect fingerprints and send the data only when the device is successfully connected to the network. There are potential disadvantages also in that the configuration tools may only be run in bursts during arrival periods, and most readings may come from the user’s halls of residence where they first set up eduroam.

It is hoped by presenting a functioning, standardised, open source solution to building radio maps we can gather momentum on the adoption of an eduroam wide solution to building self-mapping radio maps to aid users with coverage maps and other apps, as well as provide the tools and API’s needed to build location based services based on campus Wi-Fi networks.

The remainder of this paper is broken down as such: Section 2 covers related work in indoor localization and location fingerprinting. Section 3 describes the systems used in this paper. Section 4 introduces the technique to use crowd-sourced location fingerprinting techniques to generate Wi-Fi coverage data gathered by the supplicant configuration tool SU1X. Section 5 shows how a radio map can be augmented with anchor points to create an augmented radio map. Section 6 describes some ideas on how to share radio maps. Sections 7 and 8 conclude the paper and highlight some identified areas of further work.

2. Related Work

Self-mapping and indoor Wi-Fi localisation is an active area of research that has its roots in indoor localisation for mobile sensor networks as well as mobile robotics. Both these areas of research have a need to build a self-map of a wireless environment and to locate themselves within it. These areas of research combined with indoor localisation in WLAN environments have produced much work. This section will summarise the work in these areas related to this paper.

2.1 Indoor Localisation

Much work on localisation of mobile devices in an indoor environment has been undertaken in order to provide context-aware and location based services. The first areas of research involved specialist hardware devices that made use of TOA (time of arrival), TDOA (time difference) and AOA (angle of arrival). These specialist setups made use of infrared (Active Badges) (Want, 1992), ultrasonic (Active Bat) (Harter, 1999) and beacon based (Cricket) (Priyantha, 2000) high frequency wireless communications to determine location. While these systems worked well, the requirement for specialist hardware was a hurdle to mainstream adoption.

2.2 Location Fingerprinting

In 2001 Microsoft Research published a paper that detailed a passive system called RADAR (Bahl, 2000) which made use of the 802.11 based Wi-Fi infrastructure for localisation, with only additional software needed and no additional specialist hardware. This paper sparked a drive to improve localisation of indoor devices using the native wireless equipment used by 802.11b networks at the time.

In the RADAR system, the RF received signal strength indicator values (RSS) is used as a measure of distance between an Access Point (AP) and mobile node (MN). An empirical/deterministic algorithm is described (Nearest Neighbour in Signal Space NNSS) which is used to store the RSS of each AP for a specific location in a Vector in an off line phase. This off-line phase is used to create a radio map. Each point that RSS values are added to the vector is called a calibration point (CP). This information is used to compute the 2D position by triangulation, with both an empirical method and a signal propagation modelling method used.

RSS based positioning has developed to generally use three different techniques: Association/Cell ID based methods, triangulation (pattern matching and probabilistic filtering) and fingerprinting. Cell ID based is the simplest, but yields a very low accuracy equal to that of the total range of RSS from an AP. Triangulation through lateration is more accurate with 3 or more APs, but is computationally demanding and requires knowledge of the exact position of AP’s, and does not work well when considering radio propagation / multi-path problems encountered by non line of sight NLOS.

However, for indoor positioning, non-line-of-sight (NLOS) propagation and attenuations caused by walls, other structures, and even people causes significant fluctuations to RSS, which makes the simple path loss models too inaccurate in many real life situations. Location Fingerprinting has become an established area of further research as it copes well with the problem of NLOS in Wi-Fi by modelling the environment and making use of the propagation characteristics by design.

Typically two phases are required for location fingerprinting; an offline training phase to build a radio map of calibration points (CPs) for all locations and an online location determination phase where a radio map is referenced by a device to determine its location.

There are commercial products available to solve indoor localization problems. Most depend upon the labor intensive process uploading floor plans and manually positioning access points. Some also depend upon RFID and similar technologies to work.

3. System Overview

Two systems developed at Swansea University are used to facilitate the work carried out in this paper; SU1X and the LocPriS Framework. Both work on the Swansea University wireless network.

3.1 LocPriS Framework

LocPriS is a modular extensible framework we previously outlined (Ayres, 2010) and are developing that provides tools for the development, analysis, comparison and visualisation of LBS’s that preserve privacy and security. The framework will also assist in the development and testing of Location Based Services through an exposed API. Parts of the code and applications that make up the framework will be made available as open source software.

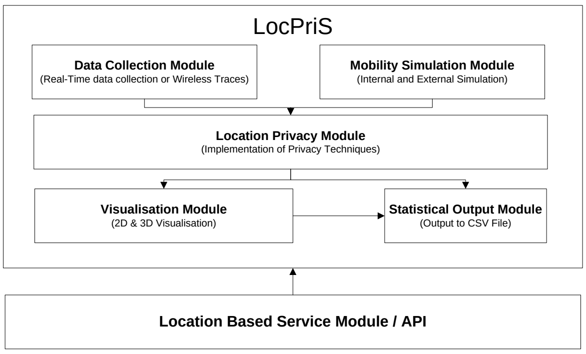

Fig. 1. The modular design of the LocPriS Framework

3.2 SU1X

SU1X is a supplicant configuration tool developed for Microsoft Windows that is used by Swansea University as well as hundreds of other institutions around the world. The main purpose of the application is to configure devices with a suitable profile and other settings to use a 802.1x wired or wireless network. SU1X is available as open source software and the code can be found on GitHub and Sourceforge. More information on SU1X can be found on the SU1X website: http://su1x.swan.ac.uk

3.3 Swansea University's Wi-Fi Network

The system outlined in this paper works interdependently to any particular Wi-Fi equipment manufacturer, but for the purpose of clarity we will briefly describe the setup at Swansea. Cisco WLAN infrastructure is used with 5 WLAN controllers and over 1000 wireless access points providing complete coverage of the university campus. There is a mixture of 802.11g and 802.11an access points. eduroam is the primary Wi-Fi network used by both staff and students. There are roughly 3000 members of staff and over 10,000 students at Swansea. On an average day the Wi-Fi network sees around 8000 unique users and peaks at around 6000 consecutively connected devices.

4. Crowd Sourcing Data using Supplicant Configuration Tools to Build Radio Maps

The location fingerprints are collected by the supplicant configuration tool. At Swansea we use SU1X, so this is used to collect the fingerprints and send them only when the device is connected successfully to the eduroam wireless network.

A fingerprint consists of all the BSSIDs seen by the device at that time, along with a RSS (signal strength) value and the SSID of that network. We also record if it is an 802.11g or 802.11n network, but this is not used. The MAC Address of the device is also captured, but the LocPriS system uses this to generate a salted hash to create a pseudo-anonymous identity.

Location fingerprints are stored in a MySQL database on LocPriS and sent to LocPriS using a HTTPS web GET. The SU1X tool runs when a user configures a profile for the first time, or when the user re-auths after a failed authentication. For March 14th 2013 there were 16314 entries taken from 172 unique devices. Of those 16314 entries we saw 1967 unique BSSID's. We broadcast 3 unique SSIDs at Swansea, but the tool captured 5 this day.

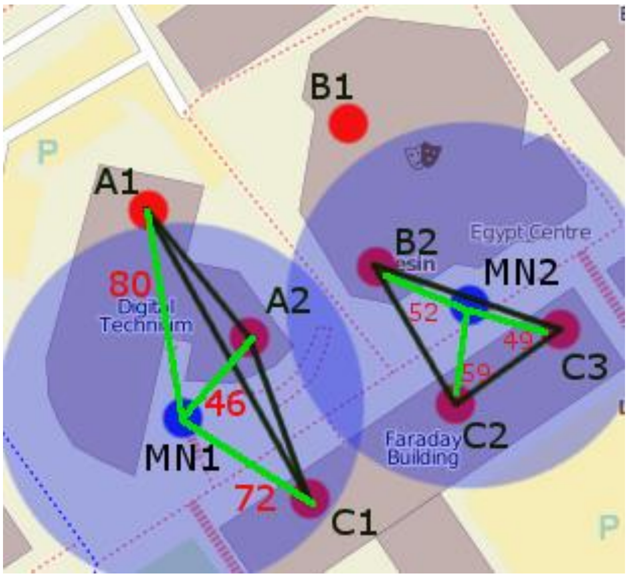

Fig. 2. Location Fingerprint Example with 2 Devices

Figure 2 shows the scenario used to capture fingerprints. Two devices can be seen (blue dots) MN1 and MN2 with the access points shown as red dots and labelled A1 to C3. There are 3 buildings labelled building A, B and C. Two fingerprints are captured in this example. The large transparent blue areas surrounding both devices is the area they can receive WLAN information.

The fingerprints captured by device MN1 can be described as:

[{A1, 80},{A2,46},{C1,72}]

The fingerprints captured by device MN2 can be described as:

[{B2, 52},{C2,59},{C3,49}]

If we say the edge E(Vx,Vy) weight between any two given vertices Vx (access points) is the sum of the RSS values in that fingerprint between the mobile node MNi and that Vx such that:

$$E(Vx,Vy) = \{MNi,Vx\} + \{MNi,Vy\}$$

So for the example in Figure 1 we get the following edge weights:

E(A1,A2)=80+46=126	E(B2,C3)=52+49=101
E(A2,C1)=72+46=118	E(B2,C2)=52+59=111
E(A1,C1)=80+72=152	E(C2,C3)=59+49=108

This produced two graphs, once for each device, each with its own set of Vertex and Edges. These graphs are also undirected and weighted. The graphs are visible in Figure 2 when following the black line between access points. If device MN1 had vertex (access point) also visible by MN2, then this would bring both graphs together into one graph. This simple concept is important when considering the sharing and distribution of graphs/radio maps.

When this process is performed on the data captured by our system, the fingerprints are larger, and the links between fingerprints are quite frequent. This allows the relative positions of access points, and groups of access points such as buildings, to be constructed in a graph when applying layout algorithms.

A force-directed layout algorithm uses the weights between edges to position vertices relative to each other when the links are strong. Often the concept of springs and spring strengths are used to describe this.

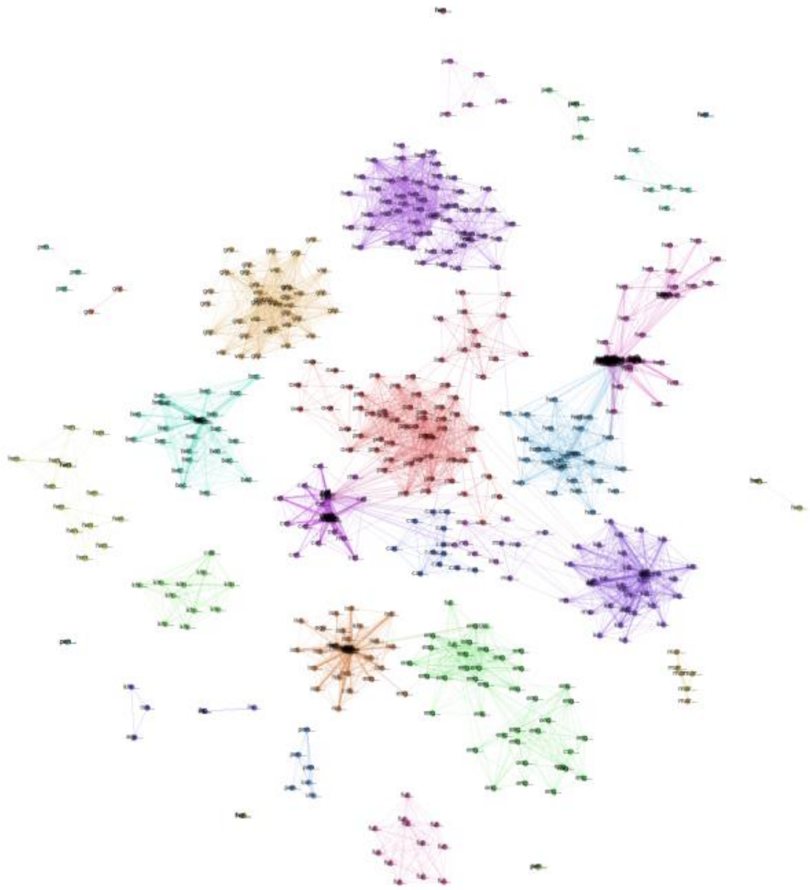

Fig. 3. Radio Map using only one day's data

Figure 3 shows the graph/radio map generated using only one day's data from March 14th 2013. As there are only limited fingerprints captured in this short period of time, the graph shows lots of smaller graphs with only a few larger graphs. Some links between graphs indicate the edges of adjacent buildings to each other.

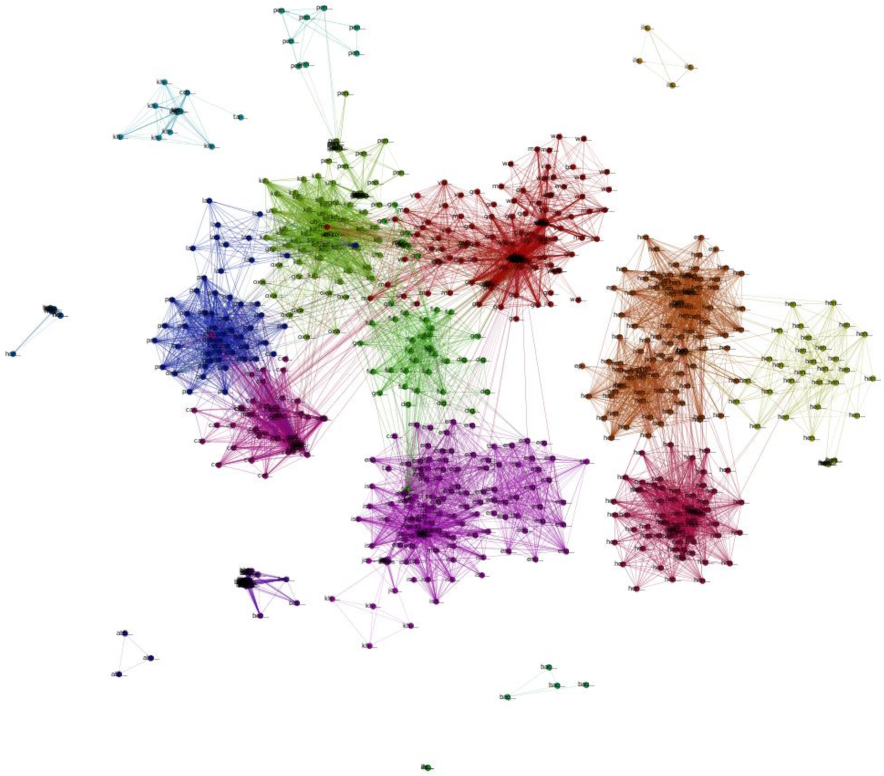

Fig. 4. Radio Map of Swansea University Wi-Fi for 1 week's data

Using more data to build the radio map results in more connected graphs and less unconnected graphs and nodes. Figure 4 shows another radio map generated with more data from 7 days beginning March 11^{th} 2013. This graph is a lot more complete than Figure 3 as it is built from a larger set of location fingerprints. Clustering colouring has been applied to the graph to highlight some artificial clusters. In Figure 3 the colours generally indicate buildings. In figure 4 this is less so the case, with geographic areas becoming coloured the same as a result of the increased links between buildings.

Some abnormalities exist in the position of some access points, but this may be corrected with the addition of more data to the graph. Building labels are added to all graphs artificially to assist in the correctness of the graphs. A numerical process of calculating accuracy and correctness of graphs needs investigation in order to fine-tune the layout algorithm in the future.

5. Augmenting Radio Maps into a Coverage Map

With the automatically generated radio map built, it is possible to augment it with real world information (anchor points from GPS or manual positioning) to produce a radio map that contains enough information on the location of access points to perform indoor and outdoor localisation, or further Location Fingerprinting. We call this an Augmented Radio Map (ARM).

The radio map alone contains information useful to indoor and outdoor localisation, but it lacks orientation and position relative to the real world. By infusing anchor node information into the radio map, it is possible to correctly orientate and position an augmented radio map that contains information useful to indoor and outdoor localisations for use with other location fingerprint techniques.

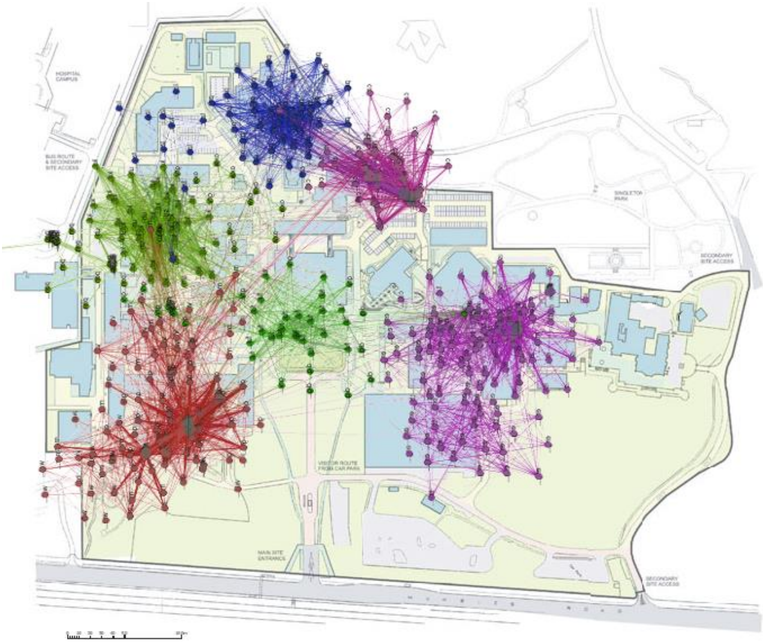

Fig. 5. Augmented Radio Map of Swansea Campus

This augmented radio map is able to locate wireless devices, and will correct itself and improve its accuracy as more fingerprints are fed into the source map. The accuracy of localisation is entirely dependent on the quality of the source map and variety of fingerprints gathered. Figure 5 shows part of the radio map overlaid onto a diagram of Swansea University Singleton campus. This is one of the connected graphs taken from Figure 4. The map is positioned and orientated correctly manually in this example but this could be achieved with a small number of anchor nodes. Anchor nodes could be generated manually or with a smart-phone app such as su1x-droid. While there are some obvious errors with the ARM, the majority of AP's are positioned atop/near the correct building of which they are in. Continuously adding more data to the source maps should improve the ARM accuracy.

6. Sharing and Distribution of Radio Maps

We believe by providing an open and standardised approach to storing location fingerprint data we can build radio maps using multiple sources of data that can be shared and combined to build more complete, accurate and self-corrected coverage maps. A radio map can be stored in a graph file format such as the Graph Exchange XML Format (GEXF) used in this paper. An example of node (access point) and edge (links) is shown below:

```
<node id="0" label="grv-wap-gnd-115"/>
<node id="1" label="grv-wap-2-313"/>
<node id="2" label="grv-wap-gnd-110"/>
<node id="3" label="grv-wap-1-215"/>
<node id="4" label="grv-wap-1-210"/>
```

```
<edge id="0" source="0" target="2" weight="167"/>
<edge id="1" source="0" target="1" weight="106"/>
<edge id="2" source="0" target="4" weight="199"/>
<edge id="3" source="0" target="3" weight="116"/>
<edge id="4" source="0" target="112" weight="89"/>
```

Graphs derived by different institutions could be combined into one master GEXF file that can then be distributed. This file could then be used and have force-layout algorithms applied to it in order to generate large scale graphs of Wi-Fi coverage. This could be achieved by a centralized authority taking ownership of a master map and developing a system to allow contributors to add data to this map, or simply by a number of institutions in a similar geographic region collaborating on a larger map to share.

7. Future Work

The techniques outlined in this paper prove the concept of using crowd-sourced location fingerprinting techniques to produce radio maps for use as coverage maps and to help build LBS's. The techniques used have room for fine-tuning and improvement which would improve the accuracy of the location data in the maps.

The use of supplicant tools to gather location fingerprints could be furthered to gather additional useful data such as signal to noise ratios. Ideally supplicants themselves should perform the fingerprinting. This opens up the possibility of embedding the fingerprints into the EAP traffic itself in order to send it to a centralised server.

Different sources of location fingerprints such a deriving them from the RADIUS logs as people re-authenticate to different access points in fixed time periods could be investigated, along with using smart-phones to generate GPS anchor nodes to contribute to the map.

The layout algorithms used also need additional research and work in order to discover the best ones. There may even be certain layout algorithms for different Wi-Fi environments. Also all radio-maps and graphs in this paper are assumed to be 2D. 3D maps should be considered in the future also.

8. Conclusion

In this paper we introduce a method to generate a Self-Mapping Radio Map to be used with location fingerprinting and indoor localisation that makes use of the LocPriS system and introduces a new method of gathering location fingerprints using the supplicant configuration tool SU1X.

We introduce the novel approach of using supplicant configuration tools to gather location fingerprints with their estimated neighbor locations which then uses undirected weighted graphs and graph layout algorithms such as force-directed graphs to produce a radio map. With the use of anchor nodes, either captured by smart-phone supplicant configuration apps or manual entered, we show how the self-mapping radio map and anchor nodes can be combined to orientate and position the self-mapping radio map into a augmented radio map (ARM).

While this approach has only been implemented at Swansea University we believe with further open standardised work and participation this technique could be used to build larger and more accurate maps that would benefit users and developers of large scale wireless network such as eduroam.

Acknowledgments

This work was carried out jointly by Information Services and Systems and the College of Engineering at Swansea University.

References

Ayres, G. , Mehmood R. LocPriS: a security and privacy preserving location based services development framework, in Proceedings of the 14th international conference on Knowledge-based and intelligent information and engineering systems: Part IV. 2010, Springer-Verlag: Cardiff, UK. p. 566-575

Want, R., et al., The Active Badge Location System. ACM Transaction on Information Systems, 1992(10): p. 91--102.

Harter, A., et al. The Anatomy of a Context-Aware Application. 1999.

Priyantha, A. Chakraborty, and H. Balakrishnan. The Cricket location-support system, in Mobile computing and networking. 2000, ACM: Boston, Massachusetts, United States.

eduroam, http://www.eduroam.org

Chown, Tim and O'Leary, Mark (2012) Towards open eduroam coverage data. In, TERENA Networking Conference 2012, Reykjavík, IS, 21 - 24 May 2012.

Bahl, P. and V.N. Padmanabhan, RADAR: an in-building RF-based user location and tracking system. 2000.

GEXF, http://gexf.net/format/

SU1X, http://su1x.swan.ac.uk

LocPriS, http://locpris.swan.ac.uk

eduroam CAT, https://cat-test.eduroam.org/

SU1X Droid, http://code.google.com/p/su1x-droid

Biographies

Gareth Ayres – Gareth Ayres works as a Wireless Network Officer for Information Services and Systems (ISS) at Swansea University. He previously completed an undergraduate degree at Swansea before working in ISS, and then completed a master's degree in communication systems before starting a PhD with the engineering department part-time. He now works and researches in the area of wireless networking and privacy and enjoys combining the practical side of his work with ISS with the theoretical aspect of his PhD. He is a member of the JANET UK SIG on 802.1X, and developer of the SU1X supplicant configuration tool.

Contact: email: g.j.ayres@swansea.ac.uk Phone: +441792602235

Jason Jones - Dr Jason Jones received a BSc in Computer Science from Swansea University in 1994. He subsequently moved across to the College of Engineering and received his PhD in High Performance Visualisation in 1999. After that he stayed in Swansea as a post-doctoral researcher working on a number of large European projects involving the major Aerospace companies in Europe including BAE Systems, Airbus and EADS. Dr Jones then took up the post of Centre Manager for a new Virtual Reality Centre in the University for 3 years before finally moving back to the College of Engineering as their High Performance Computing Research Manager

Contact: email j.w.jones@swansea.ac.uk