Evaluating Location Fingerprinting Methods for Underground GSM Networks deployed over Leaky Feeder

Fernando Pereira[1,2], Chris Theis[1], Adriano Moreira[3] , Manuel Ricardo[2],

[1] CERN/Radiation Protection, Geneva, Switzerland. Email: [fernando.pereira, christian.theis]@cern.ch
[2] INESC Porto, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal. Email: mricardo@inescporto.pt
[3] University of Minho / Dept. of Information Systems, Guimarães, Portugal. Email: adriano@dsi.uminho.pt

Abstract—**Accurate localization techniques have long been of major importance for safety systems and a lot of research has been conducted in the distributed computing field regarding its functionality and reliability. In the specific scenario of long yet narrow tunnels existing at CERN, localization methods will enable a number of applications and processes to substantially reduce human intervention.**
In this paper we evaluate the use of Fingerprinting techniques with GSM signal available throughout the LHC tunnel via a radiating cable and compare some methods to estimate the location. The existing GSM infrastructure and tunnel conditions seem to be favorable to the adoption of these Fingerprinting methods. Nevertheless significant variations in the signal have been observed which might be traced back to different operational states of accelerator equipment. These effects and their sources will be analyzed in more detail in order to improve the applied techniques and their accuracy under such challenging conditions.

Keywords—**Fingerprinting; GSM; Leaky Feeder**

I. INTRODUCTION

Location estimation has received considerable attention in the latest years after the success of the GPS system and the innumerable applications that emerged thereafter. Close to 1 billion GPS receivers are used worldwide, both for private, military and industrial uses, including process control and logistics management. GPS is nevertheless known to be very limited in indoors and dense urban environments and its accuracy lies in the range of 10 to 100 m, depending on the receiver and the atmospheric conditions [1].

In the context of CERN's activities, GPS is used at the surface being an integral component of general safety plans. However, in the vast underground areas, automatic localization techniques would also be highly advantageous for many applications from the various technical departments at CERN. Besides the fact that GPS is not applicable underground, achieving higher accuracy would be desirable as well. Of particular interest are the frequent radiation surveys carried throughout the entire accelerator complex by the radiation protection group. It involves radiation measurements in thousands of points around the accelerator facilities, for which an accurate position tag is required. In this context, the correct auto-determination of the position would allow for a much faster or even unmanned processing – a remarkable advantage in terms of efficiency, reliability and as a consequence also personnel safety.

Among indoor location techniques, those based in the Received Signal Strength (RSS) are of particular interest since they require neither the installation of extra infrastructure hardware nor allocation of extra spectrum [2] [3]. This is particularly important for the case of an accelerator tunnel where the risk of hardware damage due to radiation or interference with pre-existing components is high.

In this paper we show some preliminary results on the appropriateness of fingerprinting localization methods. We started by collecting several GSM fingerprints in a limited sector of the tunnel and, subsequently, we developed a first approach to estimate location. In the end we compare the performance of some variants of this method and give an outlook of further research that will be carried out.

II. PROBLEM CHARACTERIZATION

The LHC [4] tunnel at CERN is located 100 m below the surface; it is divided into 8 sections and measures nearly 27 km in perimeter. The tunnel walls are made of concrete forming an arched layout with a radius of 2.2 m. Besides the LHC machine, a number of auxiliary support systems exist in the tunnel, including the cryogenics system and a large amount of cabling and electronics.

GSM network coverage is available all along the tunnel's length via a set of *leaky-feeder* cables installed at nearly 2 m from the ground. For each tunnel section, two radio signals with distinct frequencies are injected at each cable's ends (see Figure 1), therefore creating two GSM cells.
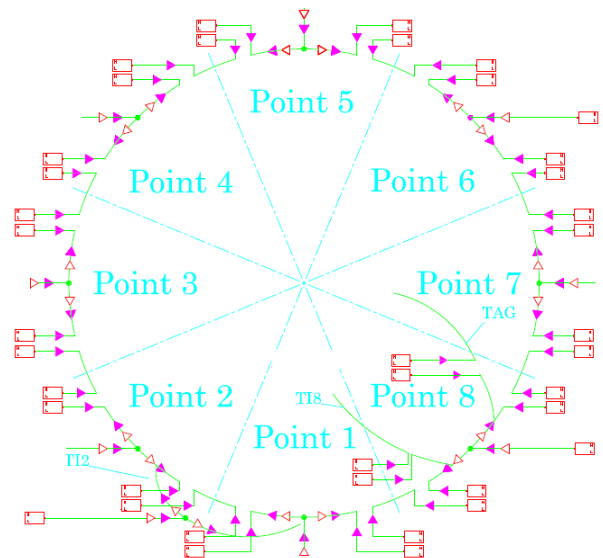

Figure 1 - GSM frequencies along the tunnel.
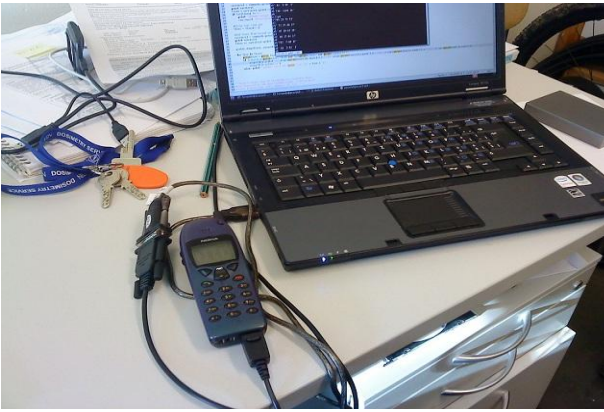The pink arrows represent the injection points of GSM frequencies

Figure 2 - Equipment used for data collection

According to the vendor specifications [5], the cable is specially designed for tunnels, providing low coupling loss variations. It propagates electromagnetic waves of up to 1950 MHz and it exhibits a longitudinal loss of 3.16 dB/km at 900 MHz. With this configuration, as one goes along the tunnel, one of the radio channels gets stronger while the other attenuates. Our objective is to exploit this radio environment for positioning inside the tunnel, thus avoiding the installation of further equipment.

III. METHODOLOGY

A. Data Collection

Several studies [6] [7] show that the use of the neighbor cells' information can definitely increase the performance of fingerprinting methods when compared to cases where only the data about the currently associated cell is used.

Except for specific vendor solutions, information about the neighbor cells is not directly accessible through the terminals' APIs. In our tests we used a Nokia 6150 mobile phone in which we enabled the NetMonitor menus (see Figure 2 for a picture of the test equipment) providing access to the signal strength of the 6 strongest cells. All fingerprints were automatically recorded to a laptop running a capture and parsing script taking advantage of the Gammu utility [8].

Measurements were taken in a tunnel sector comprising both straight and bending sections and fall into four groups:

1) *Comprehensive coarse-grained measurements* with fingerprints taken every 200 m throughout an entire LHC section and 10 samples per fingerprint. It accounts for major network channel changes.

2) *Detailed measurements* with fingerprints taken every 40 meters in a section of 600 m and 10 samples per fingerprint. Four of such sessions were carried out.

3) *Fixed location measurements I,* where 150 samples of the signal strength were collected at the same position to acocunt for variations depending on the measurement conditions.

4) *Fixed location measurements II*, where 200 samples of the signal strength were taken in identical measurement conditions, at the same position, to account for time dependent network variations.

The collected data – one file per measurement session – was then parsed and analyzed using Python scripts. For this purpose a data query library for Python which mimics the SQL standard was implemented. Aided by these software utilities several time and statistical plots of the radio map were generated.

B. A First approach for location estimation

In a first attempt to estimate location a modified weighted *K-nearest neighbor* (KNN) approach, in 16 variants, was explored. In this method raw data are being preprocessed so that for each point of the radio-map the average, the maximum and the standard deviation of the RSS values are calculated. The parameter *K* simply controls how many nearest points are returned to the user, this number being limited by the radio-map's size (N).

Since we are using RSS information from a number of GSM channels, and given that it is difficult to predict which norm will provide the best performance, we tested both Manhattan and Euclidian norms. In case of a missing channel, either in the sample or in the *radio-map* point, the device's sensitivity threshold -115dB is assumed. Each difference is additionally weighted by one out of four methods, trying to take advantage of the standard deviation.

To summarize, the variants of the method result from the combination of the following parameters:

1) *RSS grouping function:*
   a. Average
   b. Maximum
2) *Distance measurement norm:*
   a. Manhattan (1-norm)
   b. Euclidean (2-norm)
3) *Weighting method:*
   a. No weighting
   b. Weighted by Cumulative Distribution Function (CDF) of the normally-approximated RSS values in each point.
   c. Weighted by squared standard deviation
   d. Weighted by standard deviation

For a given RSS sample *i*, a score is computed, depending on the method's variant, against the *radio-map*. In the end, the probability of a point *k* in the *radio-map* being the correct one is calculated as the inverse proportion of its score against the sum of all points' score, as:

$$P_{i,k} = \frac{1/score_{i,k}}{\sum_{j=1}^{N} 1/score_{i,j}} \qquad (1)$$

Supposing a scenario where point A was given score 4 and four other points were given score 8, this point will be given probability of 1/3 while the others will be given 1/6.

C. Evaluation of the different method variants

In order to compare the method's variants and evaluate which one provides the best results, a scoring method was applied to data collected in the detailed measurements. It takes every signal sample and processes them against all radio-maps except the one they belong to. This technique is similar to that used by [7]. The *a-priory* knowledge of the

actual position of the sample allows to automatically assess the effectiveness of the method by calculating a score according to the guess order of the point k, the calculated probability P and the distance Δ to the actual position, in the following way:

$$Rank_{Method} = \sum_{i=0}^{n_{samples}} \sum_{k_i=0}^{5} \frac{P_{i,k}}{2^{\Delta_i} * 2^{k_i}} \qquad (2)$$

This method accepts contributions until the 5th most probable guess to allow the best score to be achieved when the correct point and the four closest neighbors are given the highest probability in the right order. By applying equation (2), the ideal method with ideal data would be given 1 for every sample and therefore the method's score is bounded effectively by the number of samples.

IV. RESULTS EVALUATION

A. RSS according to position

As we go along the tunnel we notice that the received signal strength can change significantly. In Figure 3 we present the mean and standard deviation of the RSS values for each detailed measurement as a function of the distance. We can clearly see two dominant GSM channels and traces of several channels from nearby tunnel sections. In the plot we can also identify two distinct areas, specifically before and after measurement location 6. In fact, the *leaky-feeder* is injected a new GSM channel at this location. Until then the new frequency experiences high attenuation as it propagates through air. After that, the propagation occurs normally in the cable and therefore, the attenuation is much lower. In this region, approximating the RSS evolution to a linear function, we obtained attenuation factors of 3.9 and 4.5 dB/km, slightly larger than the attenuation of the cable.

Despite the fact that the cable's longitudinal attenuation introduces measurable signal changes, the variations in the RSS values tend to obfuscate them, dramatically reducing the accuracy of the estimation. These effects are generally caused by the volatility of the measurement conditions but also suffer the influence of *multipath* propagation in the tunnel and *multicoupling* in the *leaky-feeder* itself [9].
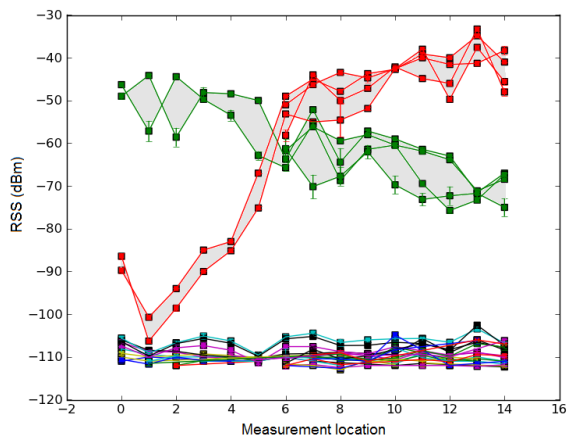

Figure 3 - RSS for consecutive measurement points.
Each line color represents a different frequency, error bars represent standard deviation and shaded areas are delimited by RSS values of the same frequency obtained in different measurements.

B. RSS dependence upon measurement conditions

Although the four detailed measurement sessions were performed in rather similar conditions, the observed RSS values at each point exhibit divergent behaviors. It was curious how one can have almost null variance in two consecutive measurements while their average differs by more than 10 dB (e.g.: measurement point 11 in Figure 3).

This fact motivated the stationary measurements in where we tested three slightly different conditions:

1) Optimal: We ensured no one was close to the equipment during the measurement process, by at least 30 meters.

2) Sub-optimal: At least one person was standing besides the equipment during the measurement

3) Realistic: One person was holding the equipment and slightly moving it during the measurement.

Figure 4 shows the RSS evolution for the measurements under the different conditions. The more adverse conditions the higher the variations and the averaged RSS values tend to drop. Such observations are clearly confirmed by their respective histograms (Figure 5), where one clearly notices a larger spread of the distribution as well as a small shift to the left.

The results of this test provide a very strong argument supporting the hypothesis that the environment characteristics of this underground tunnel, under optimal conditions, are static enough to cause very little signal fluctuations, which is quite favorable for fingerprinting methods. However, for scenarios requiring some manual handling of the equipment, as it will be the case for radiation surveys, we must account for some signal variations which will inevitably affect the accuracy.
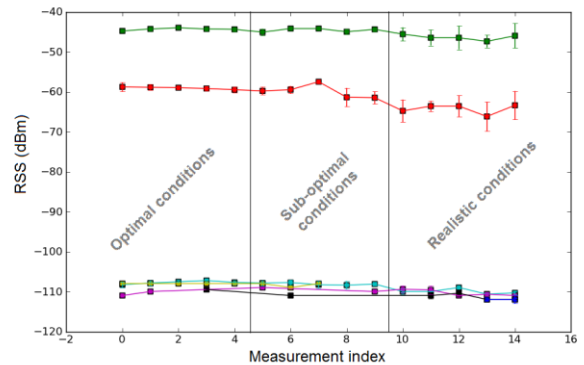

Figure 4 - RSS dependence on measurement conditions.
Each group was taken with slightly different conditions, being
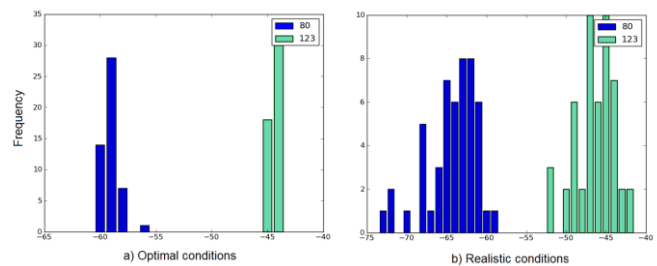a) Optimal, b) Sub-optimal or c) Realistic

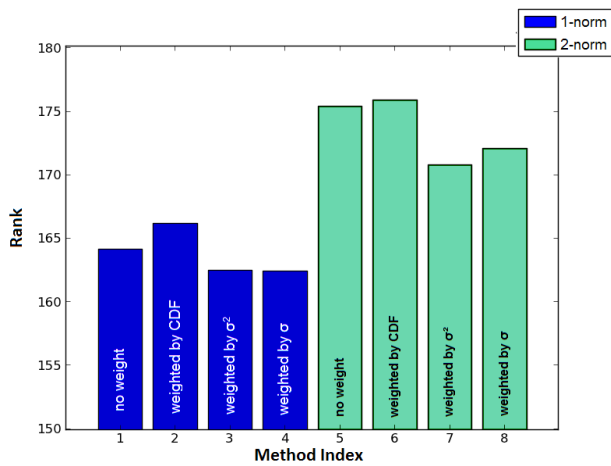Figure 5 - RSS histograms for two distinct measurement conditions

Figure 6 - Rank of the different estimation methods

Figure 7 – Accuracy of the method's first guess

Similar results were obtained in tests performed for WLAN networks [3]. Even though the RSS is more stable using *Leaky-Feeder* cables, significant signal variations occur due to *fast fading* and eventual presence of obstacles. To minimize these effects, it is important that calibration samples are all taken in the same exact locations and *line-of-sight (LoS)* propagation is available and dominant.

C. Method's comparison

The 16 variants of the method were tested according to the procedure described in section II-B. We tested all samples taken out of the calibration sets from the detailed measurement, as seen in Figure 3, which represents a realistic scenario where considerable signal variations exist.

In these tests, all the variants using the average as the grouping function for the calibration map outperformed those using the maximum by a factor between 0.5% and 8%.

The impact of the different distance measurement norms and weighting methods, as captured by the ranking algorithm, is presented in Figure 6. We conclude that the Euclidean norm is better suited to this case in calculating the distance. Within this group the best results were obtained by weighting the difference by the cumulative distribution function (method b.), followed by method a. then method d. and finally method c. To note that the weighting method a. performed reasonably well although it didn't make any use of the standard deviation. In turn, methods b. and c. clearly didn't take the best advantage of it, which indicates that a smaller standard deviation doesn't necessarily represent a better fingerprint in the *radio-map*.

The best method has the following characteristics: 1) it uses a *radio-map* whose calibration points were averaged; 2) the error distances are measured by the Euclidean norm; 3) the error distances are weighted by the CDF of the normally approximated power distribution. With these parameters the method has yielded an accuracy better than 100 m in 64% of the cases for the best guess - see Figure 7 .

V. CONCLUSIONS AND FUTURE WORK

In this paper we presented the preliminary studies on using the pre-existing GSM infrastructure in the LHC tunnel for localization using fingerprinting techniques. We realized that, al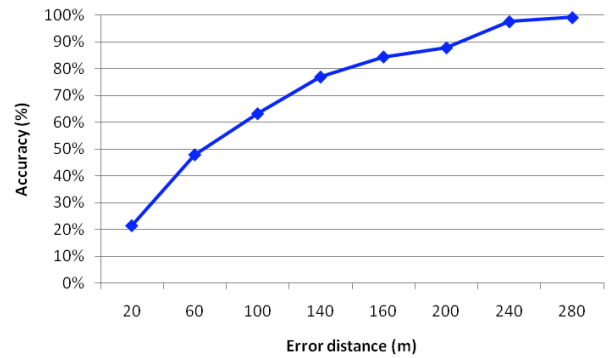though the RSS remained impressively constant under optimal measurement conditions, significant differences arise among different measurement sessions. This could be due to the fact that the equipment located in the accelerator tunnel changes their operation modes (e.g. magnets being powered) between different measurement sessions, or that the tunnel configuration promotes *fast-fading* phenomena. Despite this fact, it was shown that the RSS of different channels could be successfully used to determine the position within an acceptable range requiring neither additional hardware nor sophisticated algorithms.

Given the simplicity of the method and the absence of any kind of raw data filtering, we are convinced that the results are still very conservative. We expect therefore to achieve much higher accuracy by increasing the resolution of the calibration map, improving the signal measurement process and by exploring more advanced location estimation approaches taking advantage of the RSS statistical properties and filtering techniques.

REFERENCES

[1] R Bajaj, S L Ranaweera, and D P Agrawal, "GPS: location-tracking technology," vol. 35, pp. 92 -94, apr 2002.

[2] Alan Bensky, "Received Signal Strength," in *Wireless Positioning Technologies and Applications*.: Artech house, 2008.

[3] M Weber, U Birkel, and R Collmann, "Indoor RF Fingerprinting using leaky feeder cable considering environmental changes," 2009, pp. 1--6.

[4] CERN - The Large Hadron Collider. [Online]. http://public.web.cern.ch/public/en/LHC/LHC-en.html

[5] 1-1/4" RADIAFLEX® RLKW Cable, A-series. [Online]. http://www.rfsworld.com/dataxpress/Datasheets/?q=RLKW114-50JFLA

[6] B Denby, Y Oussar, I Ahriz, and G Dreyfus, "High-Performance Indoor Localization with Full-Band GSM Fingerprints," in *IEEE International Conference on Communications Workshops*, Dresden, 2009, pp. 1 -5.

[7] V Otsason, A Varshavsky, A LaMarca, and E de Lara, "Accurate GSM indoor localization," in *UbiComp*, LNCS, 2005, pp. 141–158.

[8] Michal Čihař. Gammu. [Online]. http://wammu.eu/gammu/

[9] M Weber, U Birkel, R Collmann, and J Engelbrecht, "Comparison of various Methods for Inddor RF Fingerprinting using Leaky feeder Cable," , 2010, pp. 291 -298.