Neural Network Assisted Identification of the Absence of Direct Path in Indoor Localization

Mohammad Heidari, Ferit Ozan Akgül, Nayef Ali Alsindi, Kaveh Pahlavan
Center for Wireless Information Network Studies, Worcester Polytechnic Institute
100 Institute Rd.
Worcester, MA, 01609, USA
{mheidari, ferit, nalsindi, kaveh}@wpi.edu

Abstract—Time of Arrival (TOA) based indoor positioning systems are considered to be the high precision alternatives to other positioning systems employing received signal strength (RSS) or Angle of Arrival (AOA). However, such systems suffer from the blockage of the direct path (DP) and occurrence of undetected direct path (UDP) condition and their performance degrades drastically in such conditions. Erroneous detection of the other multipath components (MPCs) as DP, which is the indicator of the true distance between the transmitter and the receiver, will introduce substantial ranging and localization errors into the system. Therefore, identification of the occurrence of large ranging errors and absence of DP from the received radio signal is our subsequent concern. After identification, the second step is to remedy the ranging errors in such UDP conditions. In this paper we present a methodology, based on an application of artificial neural network (ANN) design, to identify the UDP conditions and mitigate the ranging error using statistics extracted from wideband frequency domain indoor measurements conducted in a typical office building. The system bandwidth used for the frequency domain measurement was 500 MHz centered around 1 GHz.

I. INTRODUCTION

Localization using radio signals has attracted increasing attention in the field of tracking and navigation. The first researches to address this difficult problem on a worldwide scale resulted in launching a series of satellites for the global positioning systems (GPS) [1]. Although widely used today for personal and commercial outdoor applications, it is recognized that GPS does not perform satisfactorily in indoor environments which sparked interest among researchers and practitioners to focus their efforts on indoor localization systems.

With the advancements in signal processing and channel measurements, the indoor localization concept has paved the way for a variety of applications in commercial, health care, public safety, and military domains. In commercial applications, localization has been instrumental in supply-chain and asset management in warehouses. In the health care domain, there are important uses for tracking/locating patients, medical instruments in hospitals, as well as tracking people with special needs. In the public safety and military domains, precise localization is of utmost importance to assist firefighters and military personnel in accomplishing their missions [2].

Estimating the location of an individual or an object in an indoor environment can be a difficult task, often producing ambiguous results, due to the harsh wireless propagation environment in most such areas. The indoor radio propagation channel is characterized as site-specific, exhibiting severe multipath and low probability of line of sight (LOS) signal propagation between the transmitter and receiver [3], making accurate indoor localization very challenging and necessitates novel approaches in their respective model design.

Respective channel models addressing the behavior of TOA metric for telecommunication purposes have been developed and are available in the literature. However, these models are aimed primarily at telecommunication applications, while localization applications call for different approaches. In telecommunication applications we are mainly interested in modeling the behavior of the multipath spread of the channel [4] while in TOA-based indoor localization applications we focus on the behavior of the direct path (DP) between the transmitter and the receiver antennas, which identifies the ranging error between the transmitter and receiver. The difference between the observed TOA of the first path and actual TOA of the DP is referred to as ranging error. The existence of the DP in the received channel profile categorizes the receiver location into two main classes: detected direct path (DDP) and undetected direct path (UDP) [5].

Previous studies for identification and modeling of the ranging error associated with each of these conditions have been carried out in [6, 7]. Here we propose an identification method of the UDP conditions, based on an application of artificial neural network (ANN) design and propagation parameters obtained from wideband frequency domain measurements conducted in a typical office building. Previously, ANNs are exploited in the field of localization and tracking. Power measurements from different access points can be used to form an ANN for location estimation [8], while [9] utilizes variety of propagation parameters to form and train the ANN for location purposes. In this paper, the propagation parameters of the radio signal is initially used to form the likelihood functions, therefore, to construct and train the ANN.

The paper is organized as follows; section II introduces the fundamentals of TOA-based localization and describes the metrics being used for UDP identification. Section III describes the results of the binary likelihood ratio tests and ANN for UDP identification while section IV presents the
simulation setup and results for a sample indoor environment. Finally, section V concludes the paper.

II. PROBLEM STATEMENT

In this section a brief introduction to the fundamentals of localization utilizing TOA metric is presented.

A. General Description

The rich multipath indoor channel environment is characterized as an impulse response

\[ h_\infty(t, \theta) = \sum_{i=1}^{L_p} \alpha_k \delta(t - \tau_k, \theta - \theta_k) \]  

(1)

where \( L_p \) is the number of MPCs, \( \alpha_k = |\alpha_k|e^{j\phi_k} \) represents the amplitude and phase of the \( k^{th} \) path, \( \tau_k \) represents the propagation delay of the \( k^{th} \) path and \( \theta_k \) is the AOA of the \( k^{th} \) path [10].

The ideal channel impulse response (CIR) is usually referred to as the infinite-bandwidth channel profile since with infinite bandwidth the receiver could theoretically acquire every detectable path. In practice, however, the bandwidth of the system is limited. Filtering the CIR with a limited bandwidth filter results in paths with pulse shapes which can be represented as

\[ h(t, \theta) = \sum_{i=1}^{L_p} \alpha_k p(t - \tau_k, \theta - \theta_k) \]  

(2)

where \( p \) represents the time-domain pulse shape of the filter.

In TOA-based localization, TOA of the DP of the received signal is used to determine the time of flight, \( \tau_{DP} \), between the transmitter and the receiver and consequently the distance between the transmitter and the receiver [11]

\[ d_{DP} = \tau_{DP} \times c \]  

(3)

where \( d_{DP} \) represents the distance of the antenna pair, \( c \) represents the speed of electromagnetic waves in vacuum, and \( \tau_{DP} \) represents the TOA of the DP.

Applying a peak detection algorithm to the filtered channel results in detecting the first detected path (FDP) and its respective TOA. The TOA of FDP, \( \tau_{FDP} \), is then used to approximate the distance of the antenna pair

\[ d_{FDP} = \tau_{FDP} \times c \]  

(4)

where \( d_{FDP} \) is the estimate of the distance of the antenna pair and \( \tau_{FDP} \) represents the estimate of the \( \tau_{DP} \). The erroneous detection of the DP component results in ranging error, \( \varepsilon_d \), which can then be defined as

\[ \varepsilon_d = d_{FDP} - d_{DP} \]  

(5)

In the absence of multipath and presence of LOS condition, the estimate of \( \tau_{FDP} \) is very close to the true value, \( \tau_{DP} \), therefore, error is insignificant. However, in practice there exist three main sources of ranging errors in indoor localization systems. The first source of error is the multipath error which is the shift of FDP from DP due to a combination of bandwidth limitation and presence of rich multipath phenomenon in indoor environments. The ranging error caused by multipath is inversely proportional to the order of the bandwidth of the measurement system. Indeed in the presence of multipath and of course LOS conditions accurate UWB TOA estimates of the distance are feasible due to their high time domain resolution [5] which allows the TOA-based localization systems to efficiently and properly function under such conditions. However, in the absence of LOS conditions it is possible for UWB systems to observe large errors depending on the availability of DP. It is worth mentioning that the absence of LOS condition does not necessarily imply that ranging error is bound to be large as discussed below.

The performance of TOA-based localization systems also depends on the availability of DP. Indeed, as mentioned before, in the presence of DP, even in rich multipath environment, accurate UWB ranging and localization is feasible. However, in non-LOS (NLOS) conditions accurate estimates of the distance of the antenna pair are difficult to obtain. The absence of DP (also referred to as UDP) due to blockage is the dominant source of error. The third source of error is associated with propagation delay and difference of the speed of the radio waves in different media.

Previous research relegates the receiver locations of a sample indoor environment into four main categories [5, 12, 13] which are best described in Fig. 1, in which main contributors of ranging error of each class of receiver location are identified.

Detecting the DP component results in accurate ranging estimate of the distance of the antenna pair. On the other hand, the erroneous estimation of the distance of the antenna pair results in large ranging error observed by the localization system and drastically degrades the performance of such systems. Therefore, we face two hypotheses

\[
\begin{align*}
H_0 : & \quad DDP \mid d_{FDP} \approx d_{DP}, \varepsilon_d \approx 0 \\
H_1 : & \quad UDP \mid d_{FDP} \not\approx d_{DP}, \varepsilon_d \not\approx 0
\end{align*}
\]  

(6)

where \( H_0 \) denotes the DDP hypothesis, which indicates that the channel profile can effectively be used for localization,
and $H_1$ denotes the UDP hypothesis, which indicates that the channel profile is not appropriate for being used for localization purposes. $d_{DP}$, $d_{FDP}$, and $\varepsilon_d$ are defined in (3) through (5).

In this paper, we propose a methodology to distinguish between the DDP and UDP conditions by investigating the propagation parameters of the channel profile and an application of ANN. For the purpose of simulation we conducted wideband frequency domain measurements on the grid of receiver locations on the third floor of the Atwater Kent building at Worcester Polytechnic Institute. The measurement system consisted of vector network analyzer (VNA), monopole quarter wave antenna, low-noise amplifier and power amplifier. Measurements were taken from 750 MHz to 1.25 GHz, centered around 1 GHz, to provide enough bandwidth to resolve the MPCs. The results of the frequency domain measurements were then post-processed in Matlab® to obtain the time domain response using chirp-z method along with raised-cosine filter. Peak detection algorithm were then used to extract the MPCs, $\tau_{FDP}$, and other desired metrics. $\tau_{FDP}$ is then used to approximate the spacing of the antenna pair.

Amongst the important parameters of a radio signal, time delay propagation parameters and power propagation parameters have been used in the field of telecommunication. Here, we examine these radio propagation characteristics in order to identify the receiver locations with large ranging error. A hybrid metric consisted of both time delay and power characteristics can also be used for UDP identification.

### B. Time Metrics

The time characteristics of channel profiles have been used in literature for variety of applications [14–16]. RMS delay spread and mean excess delay are being used to determine the data-rate in the communication systems. Here, we utilize the time characteristic to identify the UDP condition.

**RMS delay spread:** Amongst all of the delay metrics the RMS delay spread of the channel profile is one of the easiest to find and perhaps the most effective metric, relatively, to efficiently identify the UDP conditions. RMS delay spread is defined as the

$$\tau_{rms}^2 = \frac{\sum_{i=1}^{L_{dp}} (\hat{\tau}_i - \tau_m)^2 |\alpha_i|^2}{\sum_{i=1}^{L_{dp}} |\alpha_i|^2}$$  \hspace{1cm} (7)

where $\hat{\tau}_i$ and $\alpha_i$ represent the TOA and complex amplitude of the $i^{th}$ detected peak, respectively, $L_{dp}$ represents the number of detected peaks, and $\tau_m$ is the mean excess delay of the channel profile defined as

$$\tau_m = \frac{\sum_{i=1}^{L_{dp}} \hat{\tau}_i |\alpha_i|^2}{\sum_{i=1}^{L_{dp}} |\alpha_i|^2}$$  \hspace{1cm} (8)

Conceptually, it can be observed that profiles with higher RMS delay spread are more likely to be UDP conditions as it is illustrated in Fig. 2 in which the $\tau_{rms}$ values are converted to distances. The probability plots of the distribution of DDP and UDP clearly indicates that they can be best modeled with normal distribution and their separation indicates that their normal distribution parameters are distinct.

In order to qualitatively determine the goodness-of-fit of the data to the normal distribution we apply the Kolmogorov-Smirnov ($K - S$) and $\chi^2$ hypothesis tests. The results of the normal distribution parameters, $K - S$ test and $\chi^2$ test are summarized in Table I.

### C. Power Metrics

The other class of metrics that can be extracted from the channel profile are power characteristics. Amongst the useful power metrics, total power and FDP power are the most popular ones.

**Total Power:** RSS is a simple metric that can be measured easily and it is measured and reported by most wireless devices. For example, the MAC layer of IEEE 802.11 WLAN standard provides RSS information from all active access points (APs) in a quasi-periodic beacon signal that can be used as a metric for localization [11].
The curves illustrate the difference of the plots with their weibull fits are sketched. The separation of the curves illustrates the difference of the $P_{tot}$ behavior for different DDP/UDP conditions.

Again, in order to qualitatively determine the goodness-of-fit of the weibull distribution to the data we apply the Kolmogorov-Smirnov $K - S$ and $\chi^2$ hypothesis tests. The results of the weibull distribution parameters, $K - S$ test and $\chi^2$ test are summarized in Table II.

![Figure 3. Weibull distribution modeling of total power](image)

![Figure 4. Weibull distribution modeling of hybrid metric](image)

It can be observed that weibull distribution passes the $K - S$ hypothesis test but fails the $\chi^2$ hypothesis test. As it can be concluded from Fig. 3 the assumption of the weibull distribution passes the goodness-of-fit of the weibull distribution to the data we apply the Kolmogorov-Smirnov $K - S$ and $\chi^2$ hypothesis tests. As it can be observed that weibull distribution passes the $K - S$ hypothesis test but fails the $\chi^2$ hypothesis test. As it can be concluded from Fig. 3 the assumption of the weibull distribution passes the goodness-of-fit of the weibull distribution to the data we apply the Kolmogorov-Smirnov $K - S$ and $\chi^2$ hypothesis tests.

\[ P_{tot} = P_r = 10 \log_{10}(\sum_{i=1}^{L_p} |\alpha_i|^2) \]  

(9)

For identification, we used $-P_{tot}$ which is referred to as power loss. It can be observed that profiles with higher power loss are more likely to be UDP conditions. This is best illustrated in Fig. 3 in which their respective probability plots with their weibull fits are sketched. The separation of the curves illustrates the difference of the $P_{tot}$ behavior for different DDP/UDP conditions.

Again, in order to qualitatively determine the goodness-of-fit of the weibull distribution to the data we apply the Kolmogorov-Smirnov $K - S$ and $\chi^2$ hypothesis tests. The results of the weibull distribution parameters, $K - S$ test and $\chi^2$ test are summarized in Table II.

### Table II

<table>
<thead>
<tr>
<th>Channel Profile</th>
<th>$a$</th>
<th>$b$</th>
<th>$K - S$</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDP</td>
<td>54.23</td>
<td>7.59</td>
<td>89.57%</td>
<td>28.72%</td>
</tr>
<tr>
<td>UDP</td>
<td>70.45</td>
<td>9.63</td>
<td>88.55%</td>
<td>30.63%</td>
</tr>
</tbody>
</table>

It can be observed that weibull distribution passes the $K - S$ hypothesis test but fails the $\chi^2$ hypothesis test. As it can be concluded from Fig. 3 the assumption of the weibull distribution only fits the middle part of the probability plot of the both curves which is the reason behind failing the $\chi^2$ hypothesis test.

### D. Hybrid Time/Power metric

Although, each time or power metric can be used individually to identify the class of receiver locations, but one can form a hybrid metric to achieve better results in identification of the UDP conditions. Here, we propose to use a hybrid metric consists of TOA of DP component and its respective power as the metric to identify the UDP conditions. Mathematically

\[ \xi_{hyb} = -P_{FDP} \times \tau_{FDP} \]  

(10)

where $\xi_{hyb}$ represents the metric being extracted. It can be shown that the desired metric can be best modeled with weibull distribution. Figure 4 represents the separation of the fits and

proves that, indeed, the proposed metric can efficiently be used in UDP condition identification.

The results of $K - S$ and $\chi^2$ tests for goodness-of-fit show close agreement for the assumption of the weibull distribution. The results of the weibull distribution parameters, $K - S$ test and $\chi^2$ test are summarized in Table III.

### Table III

<table>
<thead>
<tr>
<th>Channel Profile</th>
<th>$a_m$</th>
<th>$b_m$</th>
<th>$K - S_m$</th>
<th>$\chi^2_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDP</td>
<td>995.05</td>
<td>1.90</td>
<td>91.00%</td>
<td>52.00%</td>
</tr>
<tr>
<td>UDP</td>
<td>2219.09</td>
<td>3.03</td>
<td>94.92%</td>
<td>87.25%</td>
</tr>
</tbody>
</table>

III. Binary Hypothesis Testing and Artificial Neural Network for UDP Identification

Knowledge of the statistics of $\tau_{rms}$, $P_{tot}$, and $\xi_{hyb}$ enables us to identify the UDP conditions. In order to do so binary likelihood ratio tests can be performed to select the most probable hypothesis. For this purpose, we picked a random profile and extracted their respective metrics. The likelihood function of $\tau_{rms}$ can then be determined as

\[ L(H_0|\tau_{rms}) = \sup\{f_{DDP}(\tau_{rms})\} \]  

similarly we can define the likelihood functions for $P_{tot}$ and $\xi_{hyb}$ as

\[ L(H_0|P_{tot}) = \sup\{f_{DDP}(P_{tot})\} \]  

\[ L(H_0|\xi_{hyb}) = \sup\{f_{DDP}(\xi_{hyb})\} \]

The defined likelihood functions are the simplified Bayesian alternative to the traditional hypothesis testing. The outcome of the likelihood functions can be compared to a certain threshold to make a decision, i.e. $L(H_0|\tau_{rms}) \geq H_1$.

The outcome of the likelihood ratio test being greater than unity indicates that the receiver location is more likely to be a DDP condition and can appropriately be used in localization algorithm while the outcome less than unity indicates that the
profile is, indeed, UDP class of receive location; hence, the estimated $\tau_{FDP}$ has to be mitigated before being used in the localization algorithm.

For more effectively using the likelihood functions, we can combine the outcomes of the likelihood functions and form a linear simple ANN as illustrated in Fig. 5, where three outputs of the likelihood functions are the inputs of the ANN, the hidden layer of the ANN consist of 10 neurons and the output of the ANN, $\delta$, represents the UDP identification flag.

Once the ANN is trained, we can simulate the unknown channel profiles by feeding their extracted $\tau_{rms}$, $P_{tot}$, and $\xi_{hyb}$ to the ANN to identify the UDP condition. For the purpose of training and simulation we used 400 training channel profiles and 100 channel profiles for simulation, the results of accuracy of the likelihood ratio tests and ANN is summarized in Table IV.

### TABLE IV

**ACCURACY OF THE LIKELIHOOD HYPOTHESIS TEST**

<table>
<thead>
<tr>
<th>Likelihood Ratio</th>
<th>Correct Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_{rms}$</td>
<td>72.40%</td>
</tr>
<tr>
<td>$P_{tot}$</td>
<td>78.30%</td>
</tr>
<tr>
<td>$\xi_{hyb}$</td>
<td>85.48%</td>
</tr>
<tr>
<td>$\delta$</td>
<td>92.00%</td>
</tr>
</tbody>
</table>

It can be observed that amongst individual metrics, the hybrid metric performs superior, as expected, while the other two yield reasonable identification pointers. The overall performance of ANN, however, is finer than individual metrics, as the ANN has adapted itself to the classification problem and the weights are adjusted accordingly.

### IV. SIMULATION AND RESULTS OF LOCALIZATION USING TOA MEASUREMENT

In two-dimensional localization, knowledge of three accurate distance measurements from three known reference points (RPs) will be sufficient to accurately locate the mobile terminal with the help of triangulation. Increasing the number of RPs will enhance the accuracy of the system as it increases the likelihood of observing three or more precise ranging measurements from RPs. In such scenarios, one approach to mitigate the UDP problem is to simply discard the profiles suspicious to be UDP. In practice, however, those scenarios are rare and positioning systems often have access to only three or less RPs. In these cases, we have to mitigate the ranging estimate of the TOA measurement system. In practice, harsh indoor wireless environment introduces erroneous estimates of distance which degrades the performance of the localization algorithm.

One approach is to use least-squares (LS) method to solve the triangulation equations of the form

$$
\begin{align*}
(x_r - x_1)^2 + (y_r - y_1)^2 &= d_{DP}^2, \\
(x_r - x_2)^2 + (y_r - y_2)^2 &= d_{DP}^2, \\
(x_r - x_3)^2 + (y_r - y_3)^2 &= d_{DP}^2,
\end{align*}
$$

in which $d_{DP}$ represents the distance of the transmitter and the receiver. However, erroneous detection of DP and approximating $d_{DP}$ with $d_{FDP}$ degrades the performance of the localization algorithm. The particular instance of the LS algorithm that has been used for our evaluations is the one by Davidon [17], which attempts to minimize the objective function

$$
f(x) = \sum_{k=1}^{N} \left( d_k - \sqrt{(x - X_k)^2 + (y - Y_k)^2} \right)^2
$$

in an iterative manner using the following relation

$$
x_{k+1} = x_k - H_k g(x_k)
$$

where $H_k$ represents an approximation to the inverse of the Hessian of $f(x)$, $G(x)$, which is defined as

$$
G(x) = \begin{pmatrix}
\frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\
\frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2}
\end{pmatrix}
$$

and $g(x)$ is the gradient of $f(x)$, defined as

$$
g(x) = \nabla f(x)
$$

The following relation defines when the computations will be terminated

$$
\rho_k = (g(x_{k+1}))^T H_k (g(x_{k+1}))
$$

so that the iterations will stop when $\rho_k \leq \epsilon$, where $\epsilon$ is a small tolerance value.

To emphasize on the importance and significance of the ANN assisted identification of UDP condition, Fig. 6 describes localization with only two known RPs in the presence of multipath and blockage (dashed line). The RMSE value of localization error is found to be 8.42 meters.

Next we examine the UDP identification and mitigation the error associated with it. Identifying the UDP scenarios with the aid of ANN gives us an edge to mitigate the ranging associated with the profile. Since the exact value of the error is not known, the statistics of the ranging error [6] in such conditions can be used to remedy the distance estimate. A correction value can be subtracted from the observed estimated ranging reading

$$
\tau_{DP} \simeq \tau_{FDP} - \varepsilon_r \\
d_{DP} \simeq d_{FDP} - \varepsilon_d
$$

![Fig. 5. Weibull distribution modeling of hybrid metric](image)
Mitigating the distance measurement of the antenna pair and employing the idea on the previous scenario results in enhanced localization error observed by each receiver location and greatly improves the accuracy of the localization system. Figure 6 illustrates the results (solid line). It can be observed that identifying the UDP condition and mitigating the ranging error associated with such receiver locations, greatly improves the localization error observed in UDP receiver locations.

The RMSE value of the ranging error has also improved by 60% and has dropped to 3.53 meters. Although, employing two RPs limits the efficiency of the localization algorithm, as the minimum number of RPs for accurate 2D positioning is three, but it can be observed that the ANN assisted identification of UDP conditions has improved the accuracy of the localization system. However, the identification is not certain for all the receiver locations and occasionally identifies some DDP profiles as UDP and unintentionally subtract a correction value from their estimated TOA which will result in accuracy degradation at some receiver locations. Nevertheless, as Fig. 6 illustrates, the identification/mitigation algorithm can enhance the accuracy of the localization algorithm majority of the time. Figure 7 compares the CDF of localization error associated with each scenario. Again, it can be observed that, although, localization error has increased at certain receiver locations, on average it has helped the algorithm to identify the UDP conditions and mitigate their ranging measurement to a more accurate location estimation.

V. Conclusion

In this paper we have introduced a methodology to identify the UDP condition and mitigate the large ranging errors associated with it. We have proposed the use of power and time metrics from the received channel profile to obtain the likelihood functions for binary hypothesis testing. Comparing the outcome of the hypothesis tests to a certain threshold determines if a receiver is in DDP or UDP condition. In addition to the binary hypothesis tests, the aforementioned metrics can be used to train a ANN for the classification problem. Once trained, the ANN can accurately identify the UDP conditions. Once such condition with large ranging error is identified, the ranging error associated with it can be mitigated according to the statistics of the ranging error observed in such conditions and the accuracy of the localization algorithm can be reasonably improved.

REFERENCES