Wireless Location

Ali H. Sayed and Nabil R. Yousef

Abstract—This article provides an overview of wireless location technologies, methods and applications.

I. Definition

Wireless location refers to obtaining the position information of a mobile subscriber in a cellular environment. Such position information is usually given in terms of geographic coordinates of the mobile subscriber with respect to a reference point. Wireless location is also commonly termed mobile-positioning, radiolocation, and geolocation.

II. Applications

Wireless location is an important public safety feature of future cellular systems since it can add a number of important services to the capabilities of such systems. Among these services and applications of wireless location are:

1. E-911. A high percentage of emergency 911 (E-911) calls nowadays come from mobile phones [1], [2]. However, these wireless E-911 calls do not get the same quality of emergency assistance that fixed-network E-911 calls enjoy. This is due to the unknown location of the wireless E-911 caller. To face this problem, the Federal Communications Commission (FCC) issued an order on July 12, 1996 [1], which required all wireless service providers to report accurate mobile station (MS) location to the E-911 operator at the public safety answering point (PSAP). According to the FCC order, it is mandated that within five years from the effective date of the order, October 1, 1996, wireless service providers must convey to the PSAP the location of the MS within 100 meters of its actual location for at least 67 percent of all wireless E-911 calls. It is also expected that the FCC will further tighten the required location accuracy level in the recent future [3]. This FCC mandate has motivated research efforts towards developing accurate wireless location algorithms and in fact has led to significant enhancements to the wireless location technology (see, e.g., [4]–[11]).

2. Location sensitive billing. Using accurate location information of wireless users, wireless service providers can offer variable-rate call plans that are based on the caller location. For example, the cell-phone call rate might vary according to whether the call was made at home, in the office, or on the road. This will enable wireless service providers to offer competitive rate packages to those of wire-line phone companies.

3. Fraud protection. Cellular phone fraud has attained a notorious level, which serves to increase the usage and operation costs of cellular networks. This cost increase is directly passed to the consumer in the form of higher service rates. Furthermore, cellular fraud weakens the consumer confidence in wireless services. Wireless location technology can be effective in combating cellular fraud since it can enable pinpointing perpetrators.

4. Person/asset tracking. Wireless location technology can provide advanced public safety applications including locating and retrieving lost children, Alzheimer patients, or even pets. It could also be used to track valuable assets such as vehicles or laptops that might be lost or stolen. Furthermore, wireless location systems could be used to monitor and record the location of dangerous criminals.

5. Fleet management. Many fleet operators, such as police force, emergency vehicles, and other services including shuttle and taxi companies, can make use of the wireless location technology to track and operate their vehicles in an efficient way in order to minimize response times.

6. Intelligent transportation systems. A large number of drivers on road or highways carry cellular phones while driving. The wireless location technology can serve to track these phones, thus transforming them into sources of real-time traffic information that can be used to enhance transportation safety.

7. Cellular system design and management. Using information gathered from wireless location systems, cellular network planners could improve the cell planning of the wireless network based on call/location statistics. Improved channel allocation could be based on the location of active users [9], [10].

8. Mobile yellow pages. According to the available location information, a mobile user could obtain road information of the nearest resource that the user might need such as a gas station or a hospital. Thus, a cellular phone will act as smart handy mobile yellow pages on demand. Cellular users could obtain real-time traffic information according to their location.
III. Wireless Location Technologies

Wireless location technologies fall into two main categories: mobile-based and network-based techniques. In mobile-based location systems, the mobile station determines its own location by measuring signal parameters of an external system, which can be the signals of cellular base stations or satellite signals of the Global Positioning System (GPS). On the other hand, network-based location systems determine the position of the mobile station by measuring its signal parameters when received at the network cellular base stations. Thus, in the later type of wireless location systems, the mobile station plays a minimal or no role in the location process.

A. Mobile-based wireless location

GPS mobile-based location systems. In GPS-based location systems, the MS receives and measures the signal parameters of at least four different satellites of a currently existing network of 24 satellites that circle the globe at an altitude of 20,000 km and which constitute the Global Positioning System. Each GPS satellite transmits a binary code, which greatly resembles a code division multiple access (CDMA) code. This code is multiplied by a 50 Hz unknown binary signal to form the transmitted satellite signal. Each GPS satellite periodically transmits its location and the corresponding time-stamp, which it obtains from a highly accurate clock that each satellite carries. The satellite signal parameter, which the MS measures for each satellite, is the time the satellite signal takes until it reaches the MS. Cellular handsets usually carry a less accurate clock than the satellite clock. To avoid any errors resulting from this clock inaccuracy, the MS time-stamp is often added to the set of unknowns that need to be calculated, thus making the number of unknowns equal to four (three MS position coordinates plus time-stamp). This is why four satellite signal parameters have to be measured by the MS. Further information on the GPS systems is available in [12]–[13].

After measuring the satellite signal parameters, the MS can proceed in one of two manners. The first is to calculate its own position and then broadcast this position to the cellular network. Processing the measured signal parameter to obtain a position estimate is known as data fusion. In the other scenario, the MS broadcasts the unprocessed satellite signal parameters to another node (or server) in which the data fusion process is performed to obtain an estimate of the MS position. The later systems are known as server-aided GPS systems, while the first are known as pure GPS systems [14], [15].

A general scheme for server-aided GPS systems is shown in Figure 1. The server-aided GPS approach is successful in a microcell cellular environment, where the diameter of cellular cells is relatively small (a few hundred meters to a few kilometers). This environment is common in urban areas. On the other hand, in macrocell environments, which are common in suburban or rural areas, base stations, and thus servers, are widely spread out. This increases the average distance between the MS and the server leading to ineffective correction information. This is why, in many mobile-based GPS location system designs, handsets have to support both server-aided GPS and pure GPS location modes of operation (see, e.g., [14]).

GPS-based mobile location systems have the following advantages. GPS receivers usually have a relatively high degree of accuracy, which can reach less than 10 meters with differential GPS server-aided systems [16]. Moreover, the GPS satellite signals are available all over the globe, thus providing global location information. Finally, GPS technology has been studied and enhanced for a relatively long time and for various applications, and is a rather mature technology. Despite these advantages, wireless service providers may be unwilling to embrace GPS fully as the principal location technology due to the following disadvantages of GPS-based location systems:

1. Embedding a GPS receiver in the mobile handset directly leads to increased cost, size, and battery consumption of the mobile handset.
2. The need to replace hundreds of millions of handsets that are already in the market with new GPS-aided handsets. This will directly impact the rates the wireless carriers offer their users and can cause considerable inconvenience to both users and carriers during the replacement period.
3. The degraded accuracy of GPS measurements in urban environments, when one or more satellites are obscured by buildings, or when the mobile antenna is located inside a vehicle.
4. The need for handsets to support both server-aided and pure GPS modes of operation, which increases the average cost, complexity and power consumption of the mobile handset. Furthermore, the power consumption of the handset can increase dramatically when used in the pure GPS mode. Moreover, the need to deploy GPS aiding servers in wireless base stations adds up to the total cost of GPS-aided location systems.
5. GPS-based location systems face a political issue raised by the fact that the GPS satellite network is controlled by the U.S. government, which reserves the right to shut GPS signals off to any given region worldwide. This might make some wireless service providers outside the United States unwilling to rely solely on this technology.

Cellular mobile-based location systems. Cellular mobile-based wireless location technology is similar to GPS based location technology, in the sense that the MS uses external signals to determine its own location. However, in this type
of location systems, the MS relies on wireless signals originating from cellular base stations. These signals could be actual traffic cellular signals or special purpose probing signals, which are specifically broadcast for location purposes. Although this approach, which is also known as forward link wireless location, avoids the need for GPS technology, it has the same disadvantages that GPS location systems have, which is the need to modify existing handsets, and may even have increased handset power consumption over that of the GPS solution. In addition, this solution leads to lower location accuracy than that of the GPS solution. This makes cellular mobile-based location systems less favorable to use by wireless service providers.

**B. Network-based wireless location**

Network-based location technology depends on using the current cellular network to obtain wireless users location information. In these systems, the base stations (BSs) measure the signals transmitted from the MS and relay them to a central site for processing and calculating the MS location. The central processing site then relays the MS location information to the associated PSAP, as shown in Figure 2. Such a technique is also known as reverse link wireless location. Reverse link wireless location has the main advantage of not requiring any modifications or specialized equipment in the MS handset, thus accommodating a large cluster of handsets already in use in existing cellular networks. The main disadvantage of network-based wireless location is its relatively lower accuracy, when compared to GPS-based location methods [3].

Network-based wireless location techniques have the significant advantage that the MS is not involved in the location-finding process, thus these systems do not require any modifications to existing handsets. Moreover, they do not require the use of GPS components, thus avoiding any political issue that may arise due to their use. However, unlike GPS location systems, many aspects of network-based location are not fully studied yet. This is due to the relatively recent introduction of this technology. In most of the rest of this article, we will focus on network-based wireless location. First, we will review the MS signal parameters that need to be estimated by the cellular base stations and how these signals are combined to obtain a MS location estimate in data fusion. We will also discuss the sources of error that limit the accuracy of network-based location. Finally, we study different MS signal parameter estimation techniques along with some hardware implementation issues. Here, we may add that although many of the studied aspects apply to both GPS based location and forward link location, we will focus on reverse link network-based location. From this point on until the end of the article, we will refer to network-based wireless location simply as wireless location.

**IV. DATA FUSION METHODS**

Data fusion for wireless location refers to combining signal parameter estimates obtained from different base stations to obtain an estimate of the MS location. We will study the conventional data fusion methods. The MS location coordinates in a Cartesian coordinate system are denoted by \((x_0^i, y_0^i)\), with the superscript ‘o’ used to denote quantities that are unknown and which we wish to estimate. These coordinates can be estimated from measured MS signal parameters, when measured at three or more base stations (BSs). The coordinates of the nearest three BSs to the MS, denoted by BS\(_1\), BS\(_2\), and BS\(_3\), are \((x_1, y_1), (x_2, y_2)\), and \((x_3, y_3)\), respectively. Without loss of generality, the origin of the Cartesian coordinate system is set to those of BS\(_1\), i.e.,

\[ (x_1, y_1) = (0, 0). \]

We will denote the time instant at which the MS starts transmission as time instant \(t_0\). This MS signal reaches the three BSs involved in the MS location process at instants \(t_1^o\), \(t_2^o\), and \(t_3^o\), respectively. The amplitudes of arrival of the MS signal at the main and adjacent sectors of BS\(_i\) are respectively denoted by \(A_{i1}^o\) and \(A_{i2}^o\), for \(i = 1, 2, 3\). Data fusion methods obtain estimates for the MS coordinates, say \((x_o, y_o)\), by combining the MS signals through

\[ (x_0, y_0) = g(t_0, t_i, A_{i1}, A_{i2}), \]

\[ i = 1, 2, 3. \]

In cellular systems, a sectored antenna structure is very common. Each BS usually contains three different antennas, with the main lobe of each antenna facing a different direction, and with an angle of 120 degrees between each of the directions. The sector whose antenna faces a specific MS is termed the main sector serving this MS. The sector next to the main sector from the MS side is termed the adjacent sector.
Another performance index for data fusion methods is the errors of variance of the time. In other words, it is the value of the error, \( e \), which depends on the data fusion method. The resulting location error from the data fusion operation is thus given by

\[
e = \sqrt{(x_0 - x_0^*)^2 + (y_0 - y_0^*)^2}
\]  

(2)

One performance index, which is used to compare the accuracy of data fusion methods, is the location mean-square error (MSE), defined by

\[
MSE = Ee^2 = E \left[ (x_0 - x_0^*)^2 + (y_0 - y_0^*)^2 \right].
\]  

(3)

Another performance index for data fusion methods is the value below which the error magnitude, \( |e| \), lies for 67% of the time. In other words, it is the value of the error, \( e_{67\%} \), at which the error cumulative density function (CDF) is equal to 0.67. The 67% error limit is the performance index that is used by the FCC to set the required location accuracy. Here we may add that for zero-mean Gaussian errors of variance \( \sigma_e^2 \), we have

\[
e_{67\%} = \sigma_e = \sqrt{MSE}.
\]  

(4)

Several wireless location data fusion techniques have been introduced since the late 1990s, all of which are based on combining estimates of the time and/or amplitude of arrival of the mobile station (MS) signal when received at various base stations (BSs). These methods fall into the following categories:

- Time of arrival (TOA).
- Time difference of arrival (TDOA).
- Angle of arrival (AOA).
- Hybrid techniques.

A. Time of arrival (TOA)

The time of arrival (TOA) data fusion method is based on combining estimates of the time of arrival of the MS signal, when arriving at three different BSs. Since the wireless signal travels at the speed of light (C), thus the actual distance between the MS and BS, \( r_i \), is given by

\[
r_i^2 = (t_i^0 - t_0^0)C,
\]  

(5)

where \( t_i^0 \) is the actual time instant at which the MS starts transmission and \( t_0^0 \) is the actual time of arrival of the MS signal at \( BS_i \). Each TOA estimate, \( t_i \), serves to form an estimate of the distance between the MS and the corresponding BS as

\[
r_i = (t_i - t_0) C.
\]  

(6)

These estimated distances between the MS and each of the three BSs are then used to obtain \( (x_0, y_0) \) by solving the following set of equations

\[
\begin{align*}
r_1^2 &= x_0^2 + y_0^2, \\
r_2^2 &= (x_2 - x_0)^2 + (y_2 - y_0)^2, \\
r_3^2 &= (x_3 - x_0)^2 + (y_3 - y_0)^2.
\end{align*}
\]  

(7-9)

Without loss of generality, it can be assumed that \( r_1 < r_2 < r_3 \).

Now, a conventional way of solving this overdetermined nonlinear system of equations is as follows. First, equations (7) and (8) are solved for the two unknowns \((x_0, y_0)\) to yield two solutions. As shown in Figure 3(a), each equation defines a locus on which the MS must lie. Second, the distance between each of the two solutions and the circle, whose equation is given by (9) is calculated. Finally, the solution that results in the shortest distance from the circle (9) is chosen to be an estimate of the MS location coordinates \([4]\).

Although this method will help resolve the ambiguity between the two solutions resulting from solving equations (7) and (8), it does not combine the third measurement \( r_3 \) in an optimal way. Furthermore, it is not possible to combine more TOA measurements from BSs more than three.

This can be solved by combining all the available set of measurements using a least-squares approach into a more accurate estimate. This approach can be summarized as follows. Subtracting (7) from (8), we obtain

\[
r_2^2 - r_1^2 = x_2^2 - 2x_2x_0 + y_2^2 - 2y_2y_0.
\]

Similarly, subtracting (7) from (9), we obtain

\[
r_3^2 - r_1^2 = x_3^2 - 2x_3x_0 + y_3^2 - 2y_3y_0.
\]

Rearranging terms, the previous two equations can be written in matrix form as

\[
\begin{bmatrix}
x_2 & y_2 \\
x_3 & y_3
\end{bmatrix}
\begin{bmatrix}
x_0 \\
y_0
\end{bmatrix} = \frac{1}{2} \begin{bmatrix}
K_2^2 - r_2^2 + r_1^2 \\
K_3^2 - r_3^2 + r_1^2
\end{bmatrix},
\]

(10)

where

\[
K_i^2 = x_i^2 + y_i^2.
\]

(11)

Equation (10) can be rewritten as

\[
Hx = b.
\]

(12)

where

\[
H = \begin{bmatrix}
x_2 & y_2 \\
x_3 & y_3 \\
x_4 & y_4
\end{bmatrix}, \quad x = \begin{bmatrix}
x_0 \\
y_0
\end{bmatrix}, \quad b = \frac{1}{2} \begin{bmatrix}
K_2^2 - r_2^2 + r_1^2 \\
K_3^2 - r_3^2 + r_1^2 \\
K_4^2 - r_4^2 + r_1^2
\end{bmatrix}.
\]

The solution of (12) is given by

\[
x = (H^T H)^{-1} H^T b.
\]

(13)

If more than three TOA measurements are available, it can be verified that (12) still holds, with

\[
H = \begin{bmatrix}
x_2 & y_2 \\
x_3 & y_3 \\
x_4 & y_4 \\
\vdots & \vdots
\end{bmatrix}, \quad b = \frac{1}{2} \begin{bmatrix}
K_2^2 - r_2^2 + r_1^2 \\
K_3^2 - r_3^2 + r_1^2 \\
K_4^2 - r_4^2 + r_1^2 \\
\vdots & \vdots
\end{bmatrix}.
\]

In this case, the least-squares solution of (12) is given by

\[
x = (H^T H)^{-1} H^T b.
\]
The TOA method requires accurate synchronization between the BSs and MS clocks. Many of the current wireless system standards only mandate tight timing synchronization among BSs (see, e.g., [17]). However, the MS clock might have a drift that can reach a few micro seconds. This drift directly reflects into an error in the location estimate of the TOA method.

B. Time difference of arrival (TDOA)

Another widely-used technique that avoids the need for MS clock synchronization is based on time difference of arrival (TDOA) of the MS signal at two BSs. Each TDOA measurement forms a hyperbolic locus for the MS. Combining two or more TDOA measurements results in a MS location estimate that avoids MS clock synchronization errors (e.g., [18]-[21]).

We now illustrate how a closed form location solution can be obtained from TDOA measurements in the case of three BSs involved in the MS location. The TDOA measurement between BS$_i$ and BS$_1$ is defined by

$$r_{i,1} \triangleq r_i - r_1 = (t_i - t_0)C - (t_1 - t_0)C = (t_i - t_1)C.$$ \hspace{1cm} (14)

Note that TDOA measurements are not affected by errors in the MS clock time ($t_0$) as it cancels out when subtracting two TOA measurements. Equation (8) can be rewritten, in terms of the TDOA measurement $r_{2,1}$, as

$$(r_{2,1} + r_1)^2 = K_2^2 - 2x_2x_0 - 2y_2y_0 + r_1^2.$$ 

Expanding and rearranging terms, we get

$$-x_2x_0 - y_2y_0 = r_{2,1}r_1 + \frac{1}{2}(r_{2,1}^2 - K_2^2).$$

Similarly, we can write

$$-x_3x_0 - y_3y_0 = r_{3,1}r_1 + \frac{1}{2}(r_{3,1}^2 - K_3^2).$$

Rewriting these equations in matrix form we get

$$Hx = cr_1 + d,$$ \hspace{1cm} (15)

where

$$c = \begin{bmatrix} -r_{2,1} \\ -r_{3,1} \end{bmatrix}, \quad d = \frac{1}{2} \begin{bmatrix} K_2^2 - r_{2,1}^2 \\ K_3^2 - r_{3,1}^2 \end{bmatrix}.$$ 

This equations can be used to solve for $x$, in terms of the unknown $r_1$, to get

$$x = H^{-1}cr_1 + H^{-1}d.$$ 

Substituting this intermediate result into (7), we obtain a quadratic equation in $r_1$. Substituting the positive root back into the above equation yields the final solution for $x$.

If more than 3 BSs are involved in the MS location, equation (15) still holds with

$$H = \begin{bmatrix} x_2 & y_2 \\ x_3 & y_3 \\ x_4 & y_4 \\ \vdots & \vdots \end{bmatrix}, \quad c = \begin{bmatrix} -r_{2,1} \\ -r_{3,1} \\ -r_{4,1} \\ \vdots \end{bmatrix}, \quad d = \frac{1}{2} \begin{bmatrix} K_2^2 - r_{2,1}^2 \\ K_3^2 - r_{3,1}^2 \\ K_4^2 - r_{4,1}^2 \\ \vdots \end{bmatrix}$$

which yields the following least-squares intermediate solution

$$x = \left( H^T H \right)^{-1} H^T (cr_1 + d).$$ \hspace{1cm} (16)

Combining this intermediate result with (7), the final estimate for $x$ is obtained. A more accurate solution can be obtained as in [19] if the second order statistics of the TDOA measurement errors are known.
C. Angle of Arrival (AOA)

In cellular systems, AOA estimates can be obtained by using antenna arrays. The direction of arrival of the MS signal can be calculated by measuring the phase difference between the antenna array elements or by measuring the power spectral density across the antenna array in what is known as beamforming (see, e.g., [22] and the references therein). Combining the AOA estimates of two BSs, an estimate of the MS position can be obtained (see Figure 3(b)). Thus the number of BSs needed for the location process is less than that of TOA and TDOA methods by one. Another advantage of AOA location methods is that they do not need any BS clock synchronization. However, one disadvantage of using antenna array based location methods is that antenna array structures do not currently exist in second generation (2G) cellular systems. Deploying antenna arrays in all existing BSs may lead to high cost burdens on wireless service providers. The use of antenna arrays is planned in some third generation (3G) cellular systems, such as UMTS networks (see, e.g., [23], [24]), which will use antenna arrays to provide directional transmission in order to improve the network capacity.

AOA estimates can also be obtained using sectored multi-beam antennas, which already exist in current cellular systems, using the technique described in [25]. In this technique, an estimate of the AOA (θ — see Fig. 4) is obtained based on the difference between the measured signal amplitude of arrival (AmpOA) at the main beam (beam 1) and the corresponding AmpOA measured at the adjacent beam (beam 2)\(^3\). This difference is denoted by \(A_1 - A_2\) in Fig. 5, where \(A_1\) and \(A_2\) are the measured amplitude levels in dB. The measured AmpOA at the third beam may be used to resolve any ambiguity that might result from antenna side lobes. One main challenge facing this technique is the relatively low signal-to-noise ratio (SNR) of the received MS signal at the adjacent beam, especially in cases where the AOA is close to a null in the adjacent beam field pattern (e.g., θ close to 0 degrees in Figures 4 and 5). This significantly limits the AmpOA estimation accuracy at the adjacent beam.

\[\text{Fig. 4. Sectored-antenna field pattern.}\]

\(\text{Fig. 5. Measured AmpOA level patterns (in dB) for a three-beam antenna vs. the AOA (θ).}\)

\(\text{D. Hybrid techniques}\)

In TOA, TDOA, and AOA methods, two or more BSs are involved in the MS location process. In situations where the MS is much closer to one BS (serving site) than the other BSs, the accuracy of these methods is significantly degraded because of the relatively low SNR of the received MS signal at one or more BSs. Such accuracy is further reduced due to the use of power control, which requires the MS to reduce its transmitted power when it approaches a BS, causing what is known as the hearability problem [26]. Such problems will be discussed in the next section. In these cases, an alternate location procedure is to obtain an angle of arrival estimate (AOA) from the serving site and combine it with a TOA estimate of the serving site (see, e.g., [27]). Combining TOA and AOA estimates from one BS leads to one well-defined MS position estimate, which corresponds to the intersection of a circle and a straight line that starts at the center of the circle. The precision of this hybrid technique is limited by the accuracy of the TOA measurement, which is dictated by the accuracy of the MS clock. Many other hybrid location data fusion techniques can be used, such as combining TDOA and AOA measurements (see, e.g., [28]).

V. Signal Parameter Estimation

From the previous discussion, we can see that the wireless location methods depend on combining estimates of the TOA and/or AOA of the received signal at/from different BSs. Although estimating the time and amplitude of arrival of wireless signals has been studied in many works over the past decade as it is needed in many cellular systems for online signal decoding purposes (see, e.g., [29]), parameter estimation for wireless location is actually a different estimation problem in many respects. This makes the success of using conventional estimation algorithms very limited in wireless location problems. In this section, we will illustrate the differences between signal parameter estimation for conventional signal decoding and wireless location. We will then discuss some particular system issues that makes
signal estimation for wireless location different from one cellular system to the other (e.g., GSM, 2G and 3G CDMA systems).

Signal parameter estimation for wireless location purposes is different than that for online signal decoding in the following aspects:

Lower signal-to-noise ratios. Cellular systems usually suffer from high multiple access interference levels that degrade the SNR of the received signal, thus degrading the signal parameter estimation accuracy in general. Moreover, for network-based wireless location, the ability to detect the MS signal at multiple base stations is limited by the use of power control algorithms, which require the MS to decrease its transmitted power when it approaches the serving BS. This significantly decreases the received MS signal power level, when received at other BSs involved in the location process.

This scenario is shown in Figure 6, where the received SNRs at BS1 and BS2 are significantly reduced as the target MS approaches BS3. In a typical CDMA IS-95 cellular environment, the received SNR of the serving BS is in the order of -15 dB. Conventional signal estimation algorithms are usually designed to work at this SNR level. However, the received SNR at BSs other than the main serving BS can be as low as -40 dB, which poses a challenge for wireless location in such environments.

Almost-perfect knowledge of transmitted signals. In conventional signal parameter estimation for online signal decoding, the transmitted MS bits are unknown. This forces signal estimation algorithms to perform a squaring operation to remove any bit ambiguity. The squaring operation limits the period over which coherent signal integration (averaging) is possible to the bit period. Further signal integration is only possible in a noncoherent manner, i.e., averaging after squaring.

In wireless location applications, signal estimation algorithms can have almost-perfect knowledge of the MS signal in many cases. For example, at the serving site, the MS signal is decoded with reasonably high accuracy (within a one percent frame error rate). The decoded bits become ready for use after a delay that is equal to the decoded frame period used in the cellular system (20 ms for IS-95 systems). Because of the nature of wireless location applications, such a delay is not critical. Thus, the received MS signal can be buffered or delayed until the decoded bits become available through the conventional decoding process. Moreover, in many cellular systems, a cyclic redundancy check (CRC) feature is used. This enables the decoder to point out the erroneous frames after the decoding process. These erroneous frames can be ignored in the signal estimation process. The decoded bit information, obtained from the main sector of the serving site, can also be used by other adjacent sectors of the same site. Furthermore, this bit information can be transmitted through the network infrastructure to other BSs involved in locating the MS. This is known as tape recording of the MS signal.

Another technique that avoids the tape recording process is known as the power-up function (PUF), which requires the MS in emergency situations to override the power control commands and raise its transmitted power level above the conventional level. Moreover, the MS transmits known probing bit sequences instead of its regular unknown bit sequence for a part or all of the transmission period. Although this solution overcomes many of the difficulties encountered at far BSs, it requires modifying the existing handsets or at least the used power control algorithms. Furthermore, it can cause a decrease in the overall network capacity [26].

Channel fading. Channel fading is considered constant during the relatively short estimation period of conventional signal parameter algorithms for online signal decoding, and is thus ignored in the design of such algorithms. This assumption cannot be made for wireless location applications where the estimation period could be considerably longer (might reach a few seconds). Furthermore, coherent integration periods are no longer limited by the bit duration, much longer coherent averaging periods could be achieved in wireless location applications (see, e.g., [30], [31]). In this case, the coherent integration period is limited by the received signal phase rotation. Thus, unlike the case of online channel estimators, channel fading plays an important role in any successful design of signal parameter estimators for wireless location. In many cases, the system parameters have to be adapted to the available knowledge of the channel fading characteristics.

Need to resolve overlapping multipath. Multipath propaga-
ation is often encountered in wireless channels (see, e.g., [32] and the references therein). In wireless location systems, the accurate estimation of the time and amplitude of arrival of the first arriving ray of the multipath channel is vital. In general, the first arriving (prompt) ray is assumed to correspond to the most direct path between the MS and BS. However, in many wireless propagation scenarios, the prompt ray is succeeded by a multipath component that arrives at the receiver within a short delay from the prompt ray. If this delay is shorter than the duration of the pulse-shape used in the wireless system, these two rays overlap causing significant errors in the prompt ray time and amplitude of arrival estimation. Resolving these overlapping multipath components becomes rather difficult in low SNR and rapid channel fading situations. On the other hand, resolving these overlapping components is not vital for signal decoding applications as it does not significantly affect the performance of the signal decoding operation, for which coarse estimates for the channel time delays and amplitudes are sufficient.

Fig. 7 shows an example for the combined impulse response of a two ray channel and a conventional pulse-shape, for a conventional CDMA IS-95 system, in two cases (a,b). In case (a), the delay between the two channel rays is equal to twice the chip duration ($2T_c$). It is clear that the peaks of both rays are resolvable, by a simple peak-picking procedure, thus allowing for relatively accurate estimation of the prompt ray time and the amplitude of arrival. However, in case (b), both multipath components overlap and are nonresolvable via peak-picking. This can lead to significant errors in the prompt ray time and amplitude of arrival estimation. These errors cannot be tolerated for wireless location applications, especially in the case of a relatively wide pulse-shaping waveform.

A. Parameter estimation schemes

We now elaborate on some schemes that are used to estimate the wireless signal time and amplitude of arrival. The aim of such schemes is to estimate an unknown constant discrete-time delay, $\tau^o$, of a known real-valued sequence $\{s(n)\}$. The signal is transmitted over a single path time varying channel, and the designer has access to a measured sequence $\{r(n)\}_{n=1}^K$ that relates to $\{s(n)\}$ via

$$r(n) = A x^o(n) s(n - \tau^o) + v(n),$$

where $v(n)$ is additive white Gaussian noise, and $\{x^o(n)\}$ accounts for the time-varying nature of the fading channel gain over which the sequence $\{s(n)\}$ is transmitted, while $A$ is a constant unknown received signal amplitude that accounts for both the gain of the static channel if fading were not present and the antenna beam gain. Multipath issues are considered later in this section.

A conventional estimation scheme for $\tau^o$ for online bit decoding purposes is shown in Figure 8. In this scheme, the received sequence, $r(n)$, is correlated with replicas of $\{s(n - \tau)\}$ over a grid of $\tau$ values, say $\{\tau_1, \tau_2, \ldots, \tau_F\}$. The coherent averaging period, $N$, is set to the bit interval. The outputs of the correlation process are squared to remove any bit ambiguity and then non-coherently averaged over the rest of the available estimation period.

Figure 9 shows a block diagram of a wireless location TOA/AOA estimation scheme [30]. In this scheme, the received sequence $\{r(n)\}$ is also multiplied by a replica of the transmitted sequence $\{s(n - \tau)\}$ for different values of $\tau$. The resulting sequence is then averaged coherently over an interval of $N$ samples, and further averaged non-coherently for $M$ samples to build a power delay profile, $J(\tau)$. The averaging intervals $N$ and $M$ are positive integers that satisfy $K = NM$, and the value of $N$ is picked adaptively in an optimal manner by using an estimate of the maximum Doppler frequency of the fading channel ($f_D$), which can be estimated using some suggested techniques (e.g., [33]).

The searcher picks the maximum of $J(\tau)$, which is given by

$$J(\tau) = \frac{1}{M} \sum_{m=1}^{M} \left| \sum_{n=m}^{mN} r(n) s(n - \tau) \right|^2,$$

and assigns its index to the TOA estimate, according to

$$\hat{\tau}^o = \arg \max_{\tau} J(\tau).$$

The optimal value of the coherent averaging period ($N_{opt}$) is obtained by maximizing the SNR gain at the output of the estimation scheme with respect to $N$, which leads to [30]:

$$\sum_{i=1}^{N_{opt}-1} i R_x(i) = 0,$$  

where $R_x(i)$ is the autocorrelation function of the sequence $\{x(n)\}$. For a Rayleigh fading channel, $R_x(i)$ is given by

$$R_x(|i|) = J_o(2\pi f_D T_s |i|),$$
arg max \tau \left( s(n - \tau) \right)

\arg \max_{\tau} \left( \frac{1}{N} \sum_{i=1}^{N} |s(n - \tau)|^2 \right)

\arg \max_{\tau} \left( \frac{1}{M} \sum_{i=1}^{M} |s(n - \tau)|^2 \right)

Noise Bias Estimate

Fading Bias Estimate

Doppler Estimator

where $J_0(\cdot)$ is the first order Bessel function, $T_s$ is the sampling period of the received sequence \{r(n)\}, and $f_D$ is the maximum Doppler frequency of the Rayleigh fading channel. Equation (20) shows that the coherent averaging interval $N$ should be adapted according to the channel autocorrelation function.

It has been shown in [30], [31] that when coherent/non-coherent averaging estimation schemes are used for wireless location applications, where an extended coherent averaging interval is used, two biases arise at the output of the estimation scheme. Both biases affect the accuracy of the amplitude estimate significantly. The first bias is an additive noise bias that increases with the noise variance and is given by

$$B_n = \frac{\sigma_n^2}{N}. \quad (21)$$

The second bias is a multiplicative fading bias that depends on the autocorrelation function and is given by

$$B_f = \frac{R_x(0)}{N} + \frac{N^{-1}}{N^2} \sum_{i=1}^{N} 2(N - i)R_x(i). \quad (22)$$

It is clear that $B_f$ is less than or equal to unity (it is unity for static channels, which explains why previous conventional designs ignored this bias as fading was not considered in these designs — see [29]. The value of $B_f$ is also unity for $N = 1$.)

To correct for these biases, the searcher equalizes the peak value of $J(\tau)$ by subtracting two fading and noise biases, which are estimated by means of the upper and lower branches of the scheme of Figure 9. The output of this correction procedure is taken as an estimate for the amplitude-of-arrival, which is given by

$$A = \sqrt{C_f [J(\tau^o) - B_n]}.$$  

The value of $C_f$ (the fading correction factor) is $C_f = 1/B_f$, i.e.,

$$C_f = \left[ \frac{R_x(0)}{N} + \sum_{i=1}^{N-1} \frac{2(N - i)R_x(i)}{N^2} \right]^{-1}. \quad (23)$$

For a Rayleigh fading channel, this correction factor increases with the maximum Doppler frequency of the fading channel. When $f_D$ is estimated, we actually end up with an estimate for $C_f$. For the case of CDMA systems,
the quantity $B_n$ can be estimated as follows. Note first that
the noise variance $\sigma^2_v$ can be estimated directly from
the received sequence $\{r(n)\}$ since, for CDMA signals, the
SNR is typically very low. In other words, we can get an
estimate for $\sigma^2_v$ as follows:

$$\hat{\sigma}^2_v = \frac{1}{K} \sum_{i=1}^{K} |r(i)|^2.$$  

Then, an estimate for $B_n$ is given, from (21), by

$$\hat{B}_n = \frac{\sigma^2_v}{N} = \frac{1}{NK} \sum_{i=1}^{K} |r(i)|^2.$$  

(24)

With $\{\hat{B}_n, C_f\}$ so computed, we obtain an estimate for $A$
via the expression

$$\hat{A} = \sqrt{C_f \left(J(\hat{\sigma}^2_v) - \hat{B}_n\right)}.$$  

(25)

More details on this scheme and simulation results can be
found in [30], [31], [34].

B. Overlapping multipath resolving

As mentioned before, wireless propagation usually suffers
from severe multipath conditions. In situations where
the prompt ray overlaps with a successive ray, a significant
error in both the time and amplitude of arrival estimation is
encountered.

Overlapping multipath components can be modeled by
considering the relation

$$r(n) = c(n) * p(n) * h(n) + v(n),$$  

(26)

where $\{r(n)\}$ continues to denote the received sequence,
$c(n)$ is a known binary sequence, $\{p(n)\}$ is a known pulse
shape impulse response sequence, $v(n)$ is additive white
Gaussian noise of variance $\sigma^2_v$, and $h(n)$ now refers to a
multipath channel that is described by

$$h(n) = \sum_{l=1}^{L} \alpha_l x_l(n) \delta(n - \tau^p_l).$$  

(27)

Here $\alpha_l$, $\{x_l(n)\}$, and $\tau^p_l$ are respectively the unknown
gain, the normalized amplitude sequence, and the time of arrival
of the $l^{th}$ multipath component (ray). The above model
assumes that there is a multipath component at each delay
with corresponding amplitude $\alpha_l$. In practice, most of these
amplitudes will be zero or insignificant. For this reason,
a common procedure is to estimate the amplitudes at all
delays and to compare them to a threshold value that is
proportional to the noise variance. If the amplitude $\alpha_l$, at
a specific delay $\tau^p_l$, is larger than the threshold, then it
is declared to correspond to a multipath component. The
time and amplitude of arrival are then taken as the time and
amplitude of the earliest ray higher than this threshold.

In this regard, the required estimation problem is one of
estimating the vector of amplitudes at all possible delays,
which is given by

$$h \triangleq \text{col}[\alpha_1, \alpha_2, ..., \alpha_L].$$

Several least-squares-type methods have been suggested for
this purpose (e.g., [35], [36], [37]). These methods exploit
the known transmitted pulse-shape to resolve overlapping
rays. For example, it is shown in [37] that, under some
reasonable assumptions, the vector $h$ can be estimated by
means of the following procedure. The received sequence
is multiplied by delayed replica of the known transmitted
sequence, $\{s(n - \tau)\}$. Each $N$ samples of the resulting
sequence are coherently averaged and the resulting averages
at all delays are collected into a vector, say $\mathbf{r}$. An estimate
of $h$ is then obtained from $\mathbf{r}$ by solving a least-squares
problem, which leads to

$$\hat{h} = (A^T A)^{-1} A^T \mathbf{r},$$  

(28)

where $A$ denotes a convolution matrix that is constructed
from the pulse-shaping waveform. A general block diagram
for such least-squares based techniques is shown in Figure
10. Alternative so-called super-resolution techniques are
also available that are based on methods known as ESPRIT
and MUSIC (see [22] and the references therein).

Least-squares multipath resolving techniques, however,
suffer from noise boosting, which is usually caused by the
ill-conditioning of the matrices involved in the LS operation.
This ill-conditioning magnifies the noise at the output of
the LS stage. For wireless-location finding applications,
where the received signal-to-noise ratio (SNR) is relatively
low, noise magnification leads to significant errors in the
time and amplitude of arrival estimates, which in turn result
in low location precision. Other modified LS techniques
that attempt to avoid matrix ill conditioning, such as
regularized least squares, total least squares and singular
value decomposition methods, lack the required fidelity to
resolve overlapping multipath components. Furthermore,
applying least-squares methods may produce unnecessary
errors in the case of single-path propagation.

An adaptive filtering technique for multipath resolving
that avoids the aforementioned difficulties of is discussed in
[38]. Although adaptive filters do not suffer from noise
amplification, they can still suffer from slow convergence
and also divergence in some cases. These problems can
be addressed by using knowledge about the channel auto-
correlation and the fact that each channel ray fades at a
different Doppler frequency.

VI. HARDWARE IMPLEMENTATION ISSUES

It is clear from the previous considerations that signal
parameter estimation for wireless location purposes often
requires performing an extensive search over a dense grid of
the estimated parameter (e.g., TOA estimation). The hardware implementation of these search schemes requires special attention as they might introduce a dramatic increase in the overall system hardware complexity and power consumption. In this section we review two hardware architectures for implementing TOA estimation schemes. The first scheme depends on combing the hardware of both channel and location searchers, while the second involves an FFT-based estimation scheme. Both architectures aim at reducing the overall hardware complexity.

A. Combined channel/location searchers

A main hardware block in CDMA receivers is the conventional RAKE receiver, which consists of a dedicated channel searcher and a minimum of three RAKE fingers. Channel searchers obtain coarse estimates of the time and amplitude of arrival of the strongest multipath components of the mobile station signal. This information is then used by the receiver RAKE fingers and delay lock loops (DLLs) to lock onto the strongest channel multipath components, which are combined and used in bit decoding. Estimates of the time and amplitude of arrival of the strongest rays are continuously fed from the channel searcher to the RAKE fingers.

Although the location searcher and RAKE receiver differ in purpose, structure, and estimation period, several basic building hardware blocks used in each of them are common. This fact can be exploited to combine both searchers into a single architecture that serves to save hardware blocks with added design flexibility.

Figure 11 shows the scheme, proposed in [39], for the combined searcher architecture. The scheme is formed from $L_t$ data branches. Each data branch starts with a correlator over $N_1$ samples (despreader), where $N_1$ is the number of chips per symbol multiplied by the number of data samples per chip (4, 8, or 16). The output of the correlator is then multiplexed between two paths, marked ‘1’ and ‘2’ in Figure 11. In path 1, which corresponds to data path of the channel searcher or RAKE finger branches, the despread signal is squared and noncoherently averaged over $M_1$ samples, where $M_1$ is optionally adapted to an estimate of the maximum Doppler frequency of the fading channel. For path 2, which is needed for the location parameter estimation, the despread sequence is delayed for a frame period, multiplied by an estimate of the transmitted bit sequence $b(n)$, and continuously averaged over $N_2$ samples. Every $N_2$ symbols, the $N_2$ register is reset and its output is squared using the shared squaring circuits and averaged over $M_2$ samples. After the total location estimation period ($N_2 \times M_2$ symbols), the time and amplitude of arrival of the prompt ray are equalized for fading and noise biases and used to extract needed location parameters.

This architecture has the following advantages:

1. Saving a large number of hardware building blocks via multiplexing basic hardware blocks between $3L_t + L_c$ location searcher and RAKE receiver branches.

2. Improving the performance of the RAKE receiver by continuously adapting the estimation period $M_1$ to an estimate of the maximum Doppler frequency. This period is conventionally adjusted to track a fading channel in the worst (fastest) case, which restricts this period to a small value (around 6 symbols for IS-95 systems). Adapting the estimation period of the channel searcher has two advantages. First, it will increase the accuracy of the delay and amplitude estimates. Second, it will help save power as it will reduce the number of times the RAKE fingers need to change their lock point, especially for low maximum Doppler frequency cases.

3. Reducing the hardware complexity significantly by eliminating the need to use DLLs for fine tracking in the cases where the accuracy of the used combined architecture is $T_c/8$ or higher (which is typical for location applications). In such cases the accuracy of the RAKE receiver will be adequate for online bit decoding without the use of DLLs. Hardware implementation of DLLs is extremely complex, especially with regard to the analog front end [40].

Further details of the operation of this architecture is given in [39], along with performance simulation results.

B. FFT-based searchers

In mobile-based wireless location systems, a maximum likelihood searcher is embedded in the MS handset. It is very common for such searchers to involve multiple correlation operations of the received signals, from cellular BSs or GPS satellites, with local delayed replica of the transmitted signals. Performing these extensive correlations in the time-domain may be a burden for the MS hardware. Often these correlations are performed using a general purpose DSP processor, which is embedded in the MS to perform many other tasks including the correlation process. DSP processors have many advantages, such as low-cost, versatility, and design flexibility. An efficient way of imple-
Evaluating this sum requires \( N \) multiplication operations for every value of the delay \( \tau \). Thus, computing \( y(\tau) \) in the time-domain needs \( N^2 \) multiplications. On the other hand, the correlation operation can be viewed as a multiplication of two sequences in the frequency domain, which has the form

\[
y(\tau) = \sum_{n=1}^{N} r(n) s(n - \tau)
\]

Evaluating this sum requires \( N \) multiplication operations for every value of the delay \( \tau \). Thus, computing \( y(\tau) \) in the time-domain needs \( N^2 \) multiplications. On the other hand, the correlation operation can be viewed as a multiplication of two sequences in the frequency domain, which has the form

\[
Y(\omega) = R(\omega) \cdot S(\omega)
\]

where \( Y(\omega) \), \( R(\omega) \), and \( S(\omega) \) are the Fourier transforms of \( y(\tau) \), \( r(n) \), and \( s(n) \), respectively. Thus, an alternate way of computing \( y(\tau) \), is via

\[
y(\tau) = \mathcal{F}^{-1} [ R(\omega) \cdot S(\omega) ]
\]

Thus, the total number of multiplications needed to perform this operation is the number of multiplications needed to obtain the FFT of the sequences \( r(n) \) and \( s(n) \), the product of \( R(\omega) \) and \( S(\omega) \), and finally the inverse-FFT of this product. Notice that if the sequence \( s(n) \) is perfectly known, its FFT can also be known and stored instead of storing the signal \( s(n) \) itself. Thus the total number of multiplications needed is given by \( (N + N \log_2 N) \). When \( N \) is relatively large, this number of multiplications can be significantly less than \( N^2 \). For example, correlating GPS signals of length 1024 using the FFT approach is faster than performing the correlation in the time-domain by a factor of 64 [15], [41]. This directly reflects to a huge saving in the complexity and power consumption of the MS handset. We may also note that this procedure works only for sequences whose length is a power of two. The general case can still be efficiently treated using the chirp-z transform (CZT), which can handle sequences whose length is not a power of two (see [15], [41] for more details).

VII. Concluding Remarks

As can be inferred from the discussions in the body of the article, and from the extended list of references below, wireless location is an active field of investigation with many open issues and with a variety of possible approaches and techniques. The final word is yet to come, which opens the road to much further work and, ultimately, to tremendous benefits.

Acknowledgement

The authors would like to thank Dr. L. M. A. Jalloul for his input, insights, and collaboration in this area of research.

References


