Location Verification Systems Based on Received Signal Strength with Unknown Transmit Powers

Author 1, Author 2, Author 3, and Author 4

Abstract—In the context of location verification systems (LVs), this work proves that knowledge on a legitimate user’s transmit power has no effect on the optimal performance of a RSS-based LVS. Specifically, we prove that the detection performance of a generalized likelihood ratio test (GLRT), where the unknown transmit power is estimated, is identical to that of a differential likelihood ratio test (D-LRT), where the impact of the unknown transmit power is removed by differencing. Our analysis also proves the asymptotic optimality of D-LRT for an RSS-based LVS with unknown transmit power. These results are important for real-world deployments since D-LRT incurs a significantly lower implementation cost relative to GLRT.

Index Terms—Physical layer security, location verification, generalized likelihood ratio test, received signal strength.

I. INTRODUCTION

Location-based technologies and services (e.g., geographic routing protocols, location-based access control protocols, and location-based key generation) are becoming widely used in emerging wireless networks [1–5]. Meanwhile, current mainstream positioning systems, such as the now ubiquitous WiFi positioning systems and GPS, are highly vulnerable to location-spoofing attacks due to their openness and wide public availability. Against this background, the deployment of a location verification system (LVS), which provides methods to guarantee the reliability of the location information (e.g., [6–10]), is of growing importance. The main purpose of an LVS is to verify whether the claimed location of a user is consistent with his true position based on signal observations.

The LVS based on received signal strength (RSS) is of particular interest due to the simplicity in acquiring RSS observations [11]. However, a challenging issue to address in an RSS-based LVS is that the transmit power of a legitimate user (who reports his true location) may be unknown for a range of reasons - automatic power-saving functionality at battery exhaustion being one example [12]. In addition, from the perspective of designing an RSS-based LVS, whether we should set the legitimate user’s actual transmit power to be a known (i.e. public) or unknown variable is another challenging issue. This is due to the fact that if transmit power is an unknown variable, a malicious user who spoofs his claimed location will not have to meet any specific signal value at the receiving base station (BS). This, in turn, begs the question as to whether, in such a case, there remains any benefit to an RSS-based LVS. These two issues, which are the focus of this work, turn out to have a rather surprising answer - the known or unknown legitimate user’s transmit power does not matter.

Considering the case in which the legitimate user’s power is unknown, both the null and alternative hypothesis are composite. That is, the likelihood functions depend on the unknown transmit powers of the legitimate and malicious users. In this first case, the location verification is a composite binary detection problem with unknown transmit powers at all elements of the observation vector, for which the generalized likelihood ratio test (GLRT) is known to be asymptotically optimal (e.g., [13]). In the GLRT, the transmit powers have to be estimated first, which means that the complexity of the GLRT is high (from a signal processing perspective) [14]. However, we note that the above composite binary detection problem can also be solved by the likelihood ratio test (LRT) based on differential observations (D-LRT). In the D-LRT, the transmit powers are removed by differencing, and thus D-LRT is of lower complexity relative to the GLRT. In the second case, where the legitimate user’s power is publicly known, the likelihood functions are completely determined in the null hypothesis and alternative hypothesis. This is due to the fact that the malicious user will optimize his transmit power accordingly (otherwise he becomes easier to be detected as shown in [9]) and this optimized transmit power can be determined by the BS as well. For the second case, the binary detection problem can be solved by LRT, which is the uniformly most powerful test.

Our main contribution in this work is to formally prove that D-LRT is exactly equivalent (exact same detection performance) to GLRT without and with arbitrary localization errors on the true location of the legitimate or malicious user. It is well known that GLRT is asymptotically optimal (e.g., [13]) for the composite binary detection problem. As such, our proof indirectly demonstrates the asymptotic optimality of D-LRT. This rather counter-intuitive result provides useful guidelines for real world RSS-based LVs, due to the lower complexity of the D-LRT solution. With the aid of [9], this work also proves that keeping the legitimate user’s transmit power publicly known or unknown does not affect the detection performance of RSS-based LVs.

II. SYSTEM MODEL

We now outline the system model and state the assumptions adopted in this work. We denote the null hypothesis (i.e., the user is legitimate) and the alternative hypothesis (i.e., the user is malicious) by \( \mathcal{H}_0 \) and \( \mathcal{H}_1 \), respectively. The composite lognormal RSS observation model is given by [11, 15]

\[
\begin{align*}
\mathcal{H}_0 &: \ y = \theta_0 \mathbf{1}_N + \mathbf{u} + \mathbf{w}, \\
\mathcal{H}_1 &: \ y = \theta_1 \mathbf{1}_N + \mathbf{v} + \mathbf{w},
\end{align*}
\]

where \( y \) is the \( N \times 1 \) original RSS observation vector, \( \theta_0 \) presents the unknown transmit power of the legitimate user, \( \theta_1 \) presents the unknown transmit power of the malicious user,
and $1_N$ is the $N \times 1$ vector with all elements set to unity. In (1), each element of $u$ is given by $u_i = p - 10 \gamma \log_{10} \left( \frac{d_i^2}{\sigma} \right)$, $i = 1, 2, \ldots, N$, where $p$ is a reference received power corresponding to a reference distance $d$, $\gamma$ is the path loss exponent, $d_i^2$ is the Euclidean distance from the $i$-th BS to the legitimate user’s claimed location (also his true location). Each element of $v$ is given by $v_i = p - 10 \gamma \log_{10} \left( \frac{d_i^2}{\sigma} \right)$, where $d_i^2$ is the Euclidean distance from the $i$-th BS to the malicious user’s true location.

We first note that in practice the malicious user’s true location cannot be known, which leads to the fact that the vector $v$ may not be known for hypothesis testing. However, following the methodology adopted in [9], we also note that the best location (in terms of leading to the minimum detection errors) for the malicious user to launch location spoofing attacks can be determined under some practical constraints (e.g., the malicious user should be on a specific road section). In this work we consider the worst-case scenario, where the malicious user is actually at the optimal location. Such a circumstance leads to the vector $v$ being known due to the fact that the BS can determine this location. In the log-normal RSS observation model, the measurement is in dB and the noise $w$ in (1) is widely assumed to be a normal random variable with zero mean and covariance matrix $R$ [11, 15]. As such, $y$ under $\mathcal{H}_0$ conditional on $\theta_0$ follows a multivariate normal distribution, which is given by

$$f(y|\theta_0, \mathcal{H}_0) = \mathcal{N}(\theta_0 1_N + u, R).$$  \hspace{1cm} (2)

Then, $y$ under $\mathcal{H}_1$ also follows a multivariate normal distribution, i.e., $f(y|\theta_1, \mathcal{H}_1) = \mathcal{N}(\theta_1 1_N + v, R)$.

III. GENERALIZED LIKELIHOOD RATIO TEST (GLRT)

BASED ON ORIGINAL RSS OBSERVATIONS

When the transmit powers in the observation model are unknown, the binary detection problem in the RSS-based LVS becomes a composite hypothesis test, for which the GLRT is asymptotically optimal [13]. As such, in this section we first consider GLRT.

The binary decision rule embedded in the GLRT based on the original observations obtained from (1) is given by

$$\Lambda(y) \triangleq \frac{f(y|\hat{\theta}_1, \mathcal{H}_1)}{f(y|\hat{\theta}_0, \mathcal{H}_0)} \begin{cases} \lambda_R & \text{for } y \in D_1, \\ \lambda_0 & \text{for } y \in D_0, \end{cases}$$  \hspace{1cm} (3)

where $\Lambda(y)$ is the likelihood ratio of $y$, $\lambda_R$ is the threshold corresponding to $\Lambda(y)$, $\hat{\theta}_0$ and $\hat{\theta}_1$ are the maximum-likelihood estimations of $\theta_0$ and $\theta_1$, respectively, and $D_0$ and $D_1$ are the binary decisions that infer whether $y$ is from $\mathcal{H}_0$ or $\mathcal{H}_1$, respectively. We note that the specific value of $\lambda_R$ can be set by using different strategies or optimization frameworks, e.g., through predetermining a false positive rate [13], to minimize the Bayesian average cost (e.g., the total error rate that is the sum of the false positive rate and miss detection rate) [13], or to maximize the mutual information between the system input and output [16]. Following (3), we present the variant of the decision rule embedded in the GLRT based on $y$ in Lemma 1.

Lemma 1: The binary decision rule of the GLRT based on $y$ is given by

$$\mathcal{T}(y) \begin{cases} \mathcal{D}_1 & \text{for } \mathcal{T}(y) > \lambda_R, \\ \mathcal{D}_0 & \text{for } \mathcal{T}(y) \leq \lambda_0. \end{cases}$$  \hspace{1cm} (4)

where $\mathcal{T}(y)$ is the test statistic given by

$$\mathcal{T}(y) \equiv c^T R^{-1} \left( y - \frac{y^T R^{-1} 1_N}{1_N^T R^{-1} 1_N} 1_N \right).$$  \hspace{1cm} (5)

$\lambda_R$ is the threshold corresponding to $\mathcal{T}(y)$ given by

$$\lambda_R \triangleq \ln c + \frac{1}{2} c^T R^{-1} e,$$  \hspace{1cm} (6)

and the definitions of $c$ and $e$ are given by

$$c = \frac{v - u}{1_N^T R^{-1} 1_N},$$  \hspace{1cm} (7)

$$e = \frac{v + u}{1_N^T R^{-1} 1_N}.$$  \hspace{1cm} (8)

Proof: We first derive the closed-form expressions for $\hat{\theta}_0$ and $\hat{\theta}_1$. Based on (2), the log likelihood function of $y$ conditioned on $\theta_0$ under $\mathcal{H}_0$ is

$$\ln f(y|\theta_0, \mathcal{H}_0) = -\frac{1}{2} \ln |R| - \frac{N}{2} \ln(2\pi) - \frac{1}{2} (y - \theta_0 1_N - u)^T R^{-1} (y - \theta_0 1_N - u).$$

Then, the first derivative of $\ln f(y|\theta_0, \mathcal{H}_0)$ with respect to $\theta_0$ can be derived as

$$\frac{\partial \ln f(y|\theta_0, \mathcal{H}_0)}{\partial \theta_0} = \frac{\partial \ln f(y|\theta_0, \mathcal{H}_0)}{\partial (\theta_0 1_N)} \frac{\partial (\theta_0 1_N)}{\partial \theta_0} = -\theta_0 1_N^T R^{-1} 1_N + (y - u)^T R^{-1} 1_N.$$  \hspace{1cm} (9)

The second derivative of $\ln f(y|\theta_0, \mathcal{H}_0)$ with respect to $\theta_0$ is derived as $-1_N^T R^{-1} 1_N$, which is less than zero due to the fact that $R$ is a positive-definite symmetric matrix. As such, $\theta_0$ is derived in a closed-form expression, which is given by

$$\hat{\theta}_0 = \frac{y - u^T R^{-1} 1_N}{1_N^T R^{-1} 1_N}. \hspace{1cm} (10)$$

Following a similar procedure, we also derive $\hat{\theta}_1$ as

$$\hat{\theta}_1 = \frac{y - v^T R^{-1} 1_N}{1_N^T R^{-1} 1_N}. \hspace{1cm} (11)$$

We then obtain the likelihood ratio conditioned on $\hat{\theta}_0$ and $\hat{\theta}_1$ in log domain as

$$\ln \Lambda(y) = \left( \hat{\theta}_1 - \hat{\theta}_0 \right) 1_N + v - u \right)^T R^{-1} y - \frac{1}{2} \left( \hat{\theta}_1 - \hat{\theta}_0 \right) 1_N + v - u \right)^T R^{-1} \left( \hat{\theta}_1 + \hat{\theta}_0 \right) 1_N + v + u \right).$$  \hspace{1cm} (12)

Substituting (10) and (11) into (12), we obtain

$$\ln \Lambda(y) = c^T R^{-1} \left( y - \frac{y^T R^{-1} 1_N}{1_N^T R^{-1} 1_N} 1_N \right) - \frac{1}{2} c^T R^{-1} e.$$  \hspace{1cm} (13)

Following (13), we obtain the desired result in Lemma 1 after some algebraic manipulations.
The results provided in Lemma 1 can be used to derive the false positive rate, $\alpha_R$, and detection rate, $\beta_R$, of GLRT, which are presented in Theorem 1 as below.

**Theorem 1:** The false positive and detection rates of GLRT based on $y$ are given by

$$
\alpha_R = Q \left( \frac{\ln \lambda_R + \frac{1}{2} c^T R^{-1} c}{\sqrt{c^T R^{-1} c}} \right),
$$

$$
\beta_R = Q \left( \frac{\ln \lambda_R - \frac{1}{2} c^T R^{-1} c}{\sqrt{c^T R^{-1} c}} \right),
$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-t^2/2)dt$.

**Proof:** In order to obtain the false positive and detection rates, we have to derive the distributions of the test statistic $T(y)$ under both $H_0$ and $H_1$. We first derive the covariance matrix of $T(y)$, which is the same for $H_0$ and $H_1$ due to the fact that the covariance matrix of $y$ under $H_0$ and $H_1$ is the same. To this end, we first define $z$ as

$$
z = y^T R^{-1} 1_N 1_N^T R^{-1} y.
$$

The variance of $z$ is derived as

$$
\sigma_z^2 = \frac{1}{(1_N^T R^{-1} 1_N)^2} = \frac{1}{1_N^T R^{-1} 1_N}.
$$

Denoting the covariance of $(y - 1N)_j$ as $G$, the $(i, j)$-th $(i, j = 1, 2, \ldots, N)$ element of $G$ is given by

$$G_{ij} = \sigma_z^2 + R_{ij} - \text{Cov}(y_i, z) - \text{Cov}(y_j, z),$$

where $R_{ij}$ is the $(i, j)$-th element of $R$. In (18), $\text{Cov}(y_k, z)$ is the covariance of $y_k$ and $z$, which is defined as $k = i$ or $k = j$.

$$\text{Cov}(y_k, z) = \frac{1}{N} (R^{-1})^T R(:, k),$$

where $R(:, k)$ denotes the $k$-th column of $R$. Since $R$ is a positive-definite symmetric matrix, we have $(R^{-1})^T R = R^{-1} R = I_N$ ($I_N$ is the $N \times N$ identity matrix), which results in $1_N^T (R^{-1})^T R(:, k) = 1$. As such, defining $R_{kk} = \sigma_z^2$, we have $\text{Cov}(y_k, z) = 1/\sigma_z^2$, which is not dependent on $k$. Therefore, we obtain $G = R + \xi I_N$, where $\xi$ is

$$\xi = \frac{1}{1_N^T R^{-1} 1_N} = \frac{2}{\sigma_z^2} \sqrt{1_N^T R^{-1} 1_N},$$

and $1_N \times N$ is the $N \times N$ matrix with all elements set to unity. With the definition of $\xi$, following (5) the test statistic can be rewritten as $T(y) = c^T R^{-1} (y - 1N)$, and therefore the covariance matrix of $T(y)$ is given by

$$c^T R^{-1} G (c^T R^{-1})^T = c^T R^{-1} c + \xi c^T R^{-1} 1_N 1_N^T (c^T R^{-1}) = (\xi c^T R^{-1} 1_N 1_N^T (c^T R^{-1})^T. \hspace{1cm} (21)$$

As per the definition of $c$ given in (7), we have

$$c^T R^{-1} 1_N = \left( (v - u)^T - \frac{(v - u)^T R^{-1} 1_N 1_N^T R^{-1} 1_N}{R^{-1} 1_N 1_N^T R^{-1}} \right) R^{-1} 1_N = 0.$$

As such, with regard to the second term in (21) we have

$$\xi c^T R^{-1} 1_N 1_N^T (c^T R^{-1})^T = \xi \left( c^T R^{-1} 1_N 1_N^T (c^T R^{-1}) \right) = 0_{N \times N}. \hspace{1cm} (22)$$

where $0_{N \times N}$ is the $N \times N$ matrix with all elements equal zero. Substituting (22) into (21), we obtain the final covariance matrix of the test statistic $T(y)$ as

$$c^T R^{-1} G (c^T R^{-1})^T = c^T R^{-1} c. \hspace{1cm} (23)$$

The means of $y$ under $H_0$ and $H_1$ are $(\hat{\theta}_0 1_N + u)$ and $(\hat{\theta}_1 1_N + v)$, respectively. As such, the distributions of $T(y)$ under $H_0$ and $H_1$ are given by

$$T(y) | H_0 \sim N \left( c^T R^{-1} (u - u^T R^{-1} 1_N 1_N^T R^{-1} 1_N), c^T R^{-1} c \right), \hspace{1cm} (24)$$

$$T(y) | H_1 \sim N \left( c^T R^{-1} (v - v^T R^{-1} 1_N 1_N^T R^{-1} 1_N), c^T R^{-1} c \right). \hspace{1cm} (25)$$

As per the decision rule in (4) and the definitions of the false positive and detection rates, we obtain the results in (14) and (15) after some algebraic manipulations.

**IV. COMPARISON BETWEEN GLRT AND D-LRT IN RSS-BASED LVSs**

For the specific observation model given in (1), the composite binary detection problem in the RSS-based LVS can also be solved by the D-LRT [9], where the unknown transmit powers, $\theta_0$ and $\theta_1$, are removed by differencing. For convenience, we represent the detection performance of the D-LRT in the following lemma, which is Theorem 2 in [9].

**Lemma 2:** The false positive and detection rates of D-LRT are given by

$$\alpha_D = Q \left( \frac{\ln \lambda_D + \frac{1}{2} (\Delta v - \Delta u)^T D^{-1} (\Delta v - \Delta u)}{\sqrt{(\Delta v - \Delta u)^T D^{-1} (\Delta v - \Delta u)}} \right) \hspace{1cm} (26)$$

$$\beta_D = Q \left( \frac{\ln \lambda_D - \frac{1}{2} (\Delta v - \Delta u)^T D^{-1} (\Delta v - \Delta u)}{\sqrt{(\Delta v - \Delta u)^T D^{-1} (\Delta v - \Delta u)}} \right) \hspace{1cm} (27)$$

where $\lambda_D$ is the threshold corresponding to the likelihood ratio of $\Delta y$, $\Delta u_m = u_m - u_N$, $\Delta v_m = v_m - v_N$, $D_{mn} = R_{NN} + R_{mm} - R_{nm} - R_{mn}$, $m = 1, \ldots, N - 1$, and $n = 1, \ldots, N - 1$. We have the following proposition with regard to the performance comparison between GLRT and D-LRT.

**Proposition 1:** We have $\alpha_R = \alpha_D$ and $\beta_R = \beta_D$ for $\lambda_R = \lambda_D$. That is, the performance of the D-LRT is equivalent to the performance of the GLRT based on $y$.

**Proof:** From (14), (15), (26), and (27), we can see that $\alpha_R$, $\beta_R$, $\alpha_D$, and $\beta_D$ are all in the form of the $Q$ function. We denote $\alpha_R = Q(\xi R)$, $\beta_R = Q(\eta R)$, $\alpha_D = Q(\xi D)$, and $\beta_D = Q(\eta D)$. In order to prove $\alpha_R = \alpha_D$ and $\beta_R = \beta_D$ for $\lambda_R = \lambda_D$, we only need to prove $\xi R = \eta R = \xi D = \eta D$. As per (14), (15), (26), and (27), in order to prove $\xi R = \eta R = \xi D = \eta D$, it suffices to prove the following equation

$$c^T R^{-1} c = (\Delta v - \Delta u)^T D^{-1} (\Delta v - \Delta u). \hspace{1cm} (28)$$

This equation is the same as (55) in [9], which has been proved, and thus the proof of Proposition 1 follows.

Based on Proposition 1, the composite hypothesis testing problem in the RSS-based LVS can be solved through the more efficient D-LRT, in which the detection performance
known and the malicious user optimizes his transmit power.

**Scenario 2**: The legitimate user’s transmit power is unknown to BSs or the malicious user.

**Proof**: The proof follows from our Theorem 1, Lemma 2 and the Theorem 1 in [9].

The intuitive explanation of Proposition 2 is that the benefits of knowing the legitimate user’s transmit power by the BS are counteracted by the malicious user by optimizing his transmit power accordingly. As such, we can conclude that setting the legitimate user’s transmit power to be known or unknown has no effect on the RSS-based LVSs.

V. Conclusion

In this work, we first proved that the unknown transmit power in an RSS-based LVS, which can be estimated or removed by differencing, has no effect on its detection performance. In addition, we analytically showed that setting the legitimate user’s transmit power to be known or unknown has no effect on the detection performance of the RSS-based LVSs.

### References


