

Geolocalización de Nodos Inalámbricos

Matías Parodi, Guillermo Reisch

Abstract—The abstract goes here.

Index Terms—Algoritmos Evolutivos, triangulación, WiFi

I. INTRODUCCIÓN

EL presente informe consiste en la evaluación del comportamiento de un algoritmo evolutivo aplicado a la triangulación de dispositivos móviles en relación a la posición de al menos tres routers inalámbricos, cuyas posiciones absolutas o relativas a un centro arbitrario sí se conocen.

Se asumirá que los routers presentes en el entorno comparten características similares (e.g. alcance máximo), lo cual resulta razonable asumir para un gran número de routers WiFi existentes en el mercado. Esto será considerado al momento de la inicialización de la población inicial.

Por otro lado, resulta inapropiado asumir que las señales no se ven afectadas por obstáculos encontrados en el entorno, lo cual obligará a estimar, además de la posición, cuál es la caída o atenuación de la señal hacia cada router y para esto se estudiarán varias alternativas de modelo de propagación de ondas.

19 Julio, 2013

II. PROBLEMA

Debido a la explosión de las redes WiFi en todo el mundo [?], en los últimos años han surgido servicios de geolocalización por triangulación WiFi [?] que ofrecen APIs [?] y SKDs [?] para el desarrollo de aplicaciones sensibles a la posición del usuario.

Desafortunadamente, ninguna de las grandes empresas en estas áreas han publicado especificaciones de sus implementaciones, por lo que se desconocen muchos detalles del funcionamiento interno y en particular de los algoritmos que utilizan.

En el ámbito científico se exploraron varias alternativas. Quizás el enfoque más sencillo consiste en la triangulación de nodos a partir simplemente de las intensidades de las señales a cada router inalámbrico, como se analiza en [?]. Este enfoque presenta varias desventajas ya que no considera muchos aspectos de la realidad como la caída de señal por ruido electromagnético u obstáculos en el ambiente, condiciones climatológicas que afectan la señal, diversidad de antenas, etc.

Otro grupo de investigadores han explorado el problema desde el área de machine learning, consiguiendo buenos resultados utilizando dos enfoques distintos. Por un lado,

se han implementado satisfactoriamente sistemas basados en *fuzzy logic* [?], mientras que otros han intentado la integración de más de un sistema de triangulación, con el fin de tomar los beneficios de cada uno de ellos, combinando los resultados a partir de constantes de peso aprendidas y que se van ajustando a medida que se ejecuta el algoritmo [?].

Dado el tamaño del espacio de búsqueda, resulta interesante atacar el problema mediante un algoritmo evolutivo con el objetivo de lograr aproximar la posición con una precisión comparable a los servicios existentes pero para eso habrá que restringir el problema de lo contrario sería imposible abordarlo dada la cantidad de variables libres presentes.

III. DIFICULTADES

El nivel de señal recibido por el dispositivo que se intenta triangular varía con el tiempo, aunque no hayan cambiados sus posiciones relativas. Es importante entender qué ocasiona dichas fluctuaciones, ya que tener una buena estimación de la señal será esencial para la correcta evaluación del individuo.

Por mayores detalles respecto a cuales son los factores más importantes que afectan a la señal recibida y como estos influyen en la misma referirse a [?] y [?].

A. Tecnología

Los cuatro factores más importantes respecto a los dispositivos con los que se trabaja son: la frecuencia de transmisión, el tipo de antena, la ganancia de la antena y la potencia de transmisión.

A pesar de que existen muchas marcas y modelos existentes en el mercado, los fabricantes de routers modernos respetan estándares internacionales para garantizar el correcto funcionamiento de redes WiFi con cualquier dispositivo inalámbrico.

Mediante un estudio de mercado se encontró que los routers domésticos son bastante homogéneos en cuanto a sus características, lo cual nos permite asumir algunas cosas que ayudarán a simplificar el el problema de manera de hacerlo abordable.

La frecuencia de transmisión depende del canal en el que se encuentre transmitiendo el router, sin embargo el rango en el que pueden estar transmitiendo no es demasiado grande, pudiendo variar entre 2,412GHz a 2,472GHz. Para simplificar el problema se asumirá entonces que el router opera en una

frecuencia similar al promedio, 2,442GHz, correspondiente al canal 7.

En cuanto al tipo de antena, se encontró que todos los routers modernos tienen entre una y dos antenas isotrópicas de ganancia 5dBm, que emiten uniformemente hacia todos lados, por lo que la energía se disipará en el ambiente proporcionalmente al cuadrado de la distancia.

En cuanto a la potencia de transmisión, muchos routers permiten configurarla, pudiendo variar de 1dBm (1mW) a 20dBm (100mW), sin embargo en la práctica la mayoría de routers usan la potencia máxima con el fin de hacer llegar la señal más lejos.

B. Entorno

El entorno puede afectar significativamente la calidad e intensidad de la señal, en particular el ruido electromagnético, condiciones climáticas y obstáculos presentes ocasionan una pérdida significativa de la misma.

Dado que es imposible construir un modelo de propagación para cada entorno, ya que este afecta de manera diferente a cada dispositivo, resulta vital considerar la caída de la señal percibida por el dispositivo móvil, en lugar de intentar modelar cada variable por separado.

Como se verá más adelante, cada router tendrá asociada una caída de señal, la cual el algoritmo intentará aproximar de manera de encontrar la posición más probable del individuo en relación a ellos.

Se tendrá entonces que la señal real será igual a la señal emitida por el router, menos la pérdida por disipación (dependiente de la distancia), menos la pérdida por obstáculos (dependiente del entorno).

C. Muestreo

Una tarjeta WiFi estándar solo puede escuchar un canal por vez, por lo que para obtener todas las redes visibles en un entorno es necesario iterar en todos los canales en busca del anuncio de los routers inalámbricos.

Esto ocasiona que la información obtenida sobre una red en un determinado momento esté desactualizada por algunos segundos. Esto, sumado a los factores antes mencionados, ocasiona que mediciones consecutivas pueden dar valores de intensidad de señal relativamente distintos.

Para evitar este problema se tomará como intensidad de la señal hacia un router el promedio de las cinco últimas lecturas.

D. Cantidad de Nodos

La cantidad de nodos afectará la precisión con la que se consiga triangular al dispositivo móvil pero serán necesarios al menos 3 routers para un correcto posicionamiento del mismo.

IV. MODELOS EXISTENTES

El modelo de propagación de ondas permitirá la traducción de una intensidad de señal a una distancia y viceversa. Se usará fuertemente este modelo para evaluar al individuo y encontrar el error entre la estimación de la posición y la distancia teórica real.

Existen muchas aproximaciones que fueron analizadas en [?], [?] y [?], aquí se describirán 4 de ellas, partiendo de la más simple a la más compleja.

A. Empirical Model

Este modelo de propagación es quizás el más preciso de todos a pesar de ser muy simple. La principal desventaja, que lo hace impracticable en muchos casos, es que requiere del estudio del ambiente en cuestión, pero como vimos antes esto depende fuertemente de las condiciones en un instante dado, por lo que su utilidad es limitada a medida que avanza el tiempo.

$$d(rs) = |rs - ts| \times k$$

En donde d es la distancia entre el nodo transmisor y el nodo receptor, rs es la intensidad de señal recibida, ts es la intensidad de señal transmitida por el router y k es un valor empírico que corresponde a la distancia cubierta por 1dBm.

B. Wall Attenuation Factor Propagation Model

Este modelo es una adaptación del *Floor Attenuation Factor Propagation Model* sugerido en [?], [?].

En lugar de considerar la atenuación de la señal debido al factor del piso del edificio, considera el efecto causado por las paredes entre el transmisor y receptor.

La ecuación que relaciona la potencia en función de la distancia es:

$$P(d) = P(d_0) - 10 \times n \times \log\left(\frac{d}{d_0}\right) - \begin{cases} n \times W \times WAF & nW < C \\ C \times WAF & nW \geq C \end{cases}$$

En donde n indica el ratio de incremento de la atenuación debido a la distancia, $P(d_0)$ es la potencia de la señal en la distancia de referencia d_0 , d es la distancia entre el transmisor y el emisor, C es la cantidad de paredes hasta la que el factor de atenuación se ve afectado y WAF es el factor de atenuación.

Como se puede apreciar este modelo es bastante más complejo y sigue requiriendo un análisis empírico para determinar el factor de atenuación.

C. Land Propagation Model

Un tercer modelo teórico bastante más complejo considera la pérdida de potencia en la señal debido a tres factores principales:

- *Path loss* (L_p): se refiere a la perdida promedio debido a la propagación de las ondas en el espacio
- *Slow fading* (L_s): se refiere a perdidas de señal causadas en distancias largas debido a variaciones lentas en el espacio tiempo [1 de generic model]
- *Fast fading* (L_f), se refiere a perdidas de señal por variaciones bruscas en el espacio. [2, 9 y 11 de generic model]

De esta forma la ecuación que relaciona potencia con distancia está dada por:

$$P_r = \frac{G_r \times G_t}{L} \times P_t$$

$$L = L_p \times L_s \times L_f$$

En donde G_r y G_t son las ganancias de las antenas de recepción y transmisión respectivamente y P_t es la potencia con la que transmite el router.

Calcular L es extremadamente complejo y requiere de mediciones del ambiente en donde se va a desenvolver el algoritmo, lo que lo hace poco práctico en este contexto.

D. Free Space Propagation Model

Este modelo de propagación está inspirado en la ecuación de propagación de Friis [?] y es quizás el más útil de todos ya que ofrece un compromiso entre simplicidad y utilidad.

$$P_r = \frac{G_t \times G_r \times \lambda^2}{(4\pi \times d)^2} \times P_t$$

En donde P_r es la potencia recibida, G_t y G_r son la ganancia de las antenas de transmisión y recepción respectivamente, λ es la longitud de onda usada y d la distancia entre el transmisor y receptor.

A partir de esta ecuación se puede despejar la distancia teórica entre ambos nodos.

$$d(L_{dB}) = \sqrt{\frac{G_t \times G_r \times \lambda^2}{4\pi \times P_r}} \times P_t$$

λ se puede obtener a partir de la frecuencia (F), utilizando la velocidad de la luz (C).

$$\lambda = \frac{C}{F}$$

Por otro lado, usando las fórmulas de decibel se puede obtener la potencia (P_1) a partir de la intensidad (L_{dB}) en relación a una potencia de referencia (P_0) que en el caso de las redes inalámbricas es 1mW.

$$P_1 = 10^{\frac{L_{dB}}{10}} \times P_0$$

Finalmente, obtenemos la distancia teórica entre los nodos a partir de la siguiente ecuación.

$$d(L_{dB}) = \sqrt{\frac{G_t \times G_r \times (\frac{C}{F})^2}{4\pi \times 10^{\frac{L_{dB}}{10}} \times P_0}} \times P_t$$

En donde simplificándola y usando los valores antes mencionados se tiene:

$$d(L_{dB}) = \sqrt{\frac{5 \times 5 \times (\frac{299.792.458}{2.442.000.000})^2}{4\pi \times 10^{\frac{L_{dB}}{10}} \times 0,001}} \times P_t$$

$$= \sqrt{\frac{29,98336}{10^{\frac{L_{dB}}{10}}}} \times P_t$$

V. ALGORITMO

A. Tecnologías usadas

Los lenguajes de programación usados serán C/C++ [?] debido a su sencillez, bajo consumo de memoria, velocidad de ejecución y herramientas disponibles [?].

Se implementará el algoritmo usando el framework *Malva* [?], el cual es un fork de *Mallba* [?]. Se estudiaron otras alternativas pero finalmente se optó por este debido a recomendaciones hechas en el curso y al hecho de que contábamos con el soporte del foro de discusión del mismo.

B. Representación interna

Cada individuo será una n-úpla de la forma:

$$I = (x, y, d_1, \dots, d_n)$$

$$x, y \in (-\infty, +\infty)$$

$$d_i \in [0, 1] \forall i \in (1, \dots, n)$$

En donde (x, y) se corresponde con las coordenadas estimadas del individuo en el plano y las constantes d_i representan la atenuación estimada de la señal entre dicho nodo y el router i .

Por ejemplo, consideremos el siguiente individuo:

$$I = (10, -5, 0.15, 0, 0.1)$$

De acuerdo a la representación descripta, este individuo se estima que está en la posición (10, -5) en el plano (con respecto al origen de coordenadas arbitrario usado para inicializar los routers) y que tiene un decaimiento estimado de señal del 15%, 0% y 10% con el router R_1 , R_2 y R_3 respectivamente.

En otras palabras, si la intensidad de la señal que percibe de, por ejemplo, el router R_1 es de 82%, para el cálculo del fitness se asumirá que la misma debió ser 15% mayor, por lo cual se usará una intensidad teórica de $82 \times 1.15 = 94\%$.

C. Función de fitness

La función de fitness a minimizar es la suma de errores entre la posición estimada por el algoritmo evolutivo y la distancia teórica obtenida mediante el modelo de propagación *free space propagation model* descrito anteriormente.

$$f(x, y, d_1, \dots, d_n) = \sum_{i=1}^n |\sqrt{(x - x_o)^2 + (y - y_o)^2} - d(s_i \times (1 + d_i))|$$

En donde s_i es la intensidad recibida del router R_i .

D. Inicialización

Como se ha dicho antes, se asume que los routers comparten características similares, en particular, que el alcance máximo para todos ellos es de A metros.

La población será inicializada eligiendo valores aleatorios para (x, y) de manera que todos los individuos comiencen en la intersección de los círculos de centro (R_{i_x}, R_{i_y}) y radio A , en donde (R_{i_x}, R_{i_y}) es la posición del router R_i .

De esta manera, durante la inicialización de la población, los valores estarán limitados por:

$$x_{max} = \text{rand}(\max_{i=1} R_{i_x} - A, \min_{i=1} R_{i_x} + A)$$

$$y_{max} = \text{rand}(\max_{i=1} R_{i_y} - A, \min_{i=1} R_{i_y} + A)$$

En donde $\text{rand}(\min, \max)$ es una función que devuelve un número en punto flotante aleatorio entre \min y \max .

E. Operadores

Dado que el espacio de búsqueda es inmenso, resulta esencial mantener la diversidad hasta haber muestreado una amplia área del plano, ya que en principio se puede asumir la existencia de muchos óptimos locales. Dicho esto, los operadores que serán utilizados se describen a continuación.

1) *Selección*: Se usará selección estocástica universal con punteros equiespaciados.

2) *Cruzamiento*: El operador de cruzamiento promediará los valores de ambos padres.

$$x_3 = \frac{x_1 + x_2}{2}, y_3 = \frac{y_1 + y_2}{2}, d_{3_i} = \frac{d_{1_i} + d_{2_i}}{2}$$

3) *Mutación*: El operador de mutación alterará de acuerdo a un valor aleatorio elegido en un determinado rango.

$$x = x + k_1, y = y + k_2, d_i = d_i + k_{3_i}, k_i \in (n, m)$$

4) *Búsqueda Local (sí? no? por qué?)*:

F. Mecanismo de reparación

Cabe señalar que los individuos son válidos desde el momento de su inicialización y los operadores aplicados no los corrompen, por lo que no hace falta un mecanismo de reparación o de descarte de individuos inválidos.

VI. ANÁLISIS EXPERIMENTAL

A. Entorno de prueba

B. Instancias de prueba

C. Ajustes de parámetros

D. Resultados obtenidos

Para evaluar el algoritmo se harán 30 ejecuciones independientes para el caso de 3, 4, 5, ..., 10 nodos. Adicionalmente se harán 3 ejecuciones en un entorno real, usando 3, 4 y 5 routers.

Para todos los casos se reportará:

- Tiempo de ejecución
- Generación del mejor individuo
- Fitness del mejor individuo
- Valor promedio del fitness
- Desviación estándar del fitness

Para los casos reales se reportará además:

- El error entre la posición del mejor individuo y la posición real que debió encontrar
- La diferencia entre dicha posición y la posición devuelta por servicios existentes

VII. CONCLUSIÓN

The conclusion goes here.

APPENDIX A

PROOF OF THE FIRST ZONKLAR EQUATION

Appendix one text goes here.