

# Distribución de puntos de red inalámbricos utilizando algoritmos genéticos

M. Angélica Pinninghoff J., Ricardo Contreras A. y Rodrigo Aravena A.

Departamento de Ciencias de Computación,  
Universidad de Concepción, Chile  
{`rcontrer,mpinning`}@udec.cl

**Resumen** El suministrar servicios de Internet a un determinado lugar, es una tarea cuya dificultad aumenta a medida que el número de equipos que lo usarán y el área que se abastecerá también lo hacen. Solucionar el problema implica el análisis de factores como la ubicación del conjunto de sensores de red, además de las características y restricciones que estos y su ubicación poseen. Así, se trata un problema de optimización que puede ser abordado mediante una categoría de problemas llamada *Facility Location Problem*. En este trabajo se propone aplicar técnicas de algoritmos genéticos para la resolución del problema, obteniendo una distribución de estos sensores para una ciudad arbitraria de manera que se minimicen los costos de instalación sin disminuir la calidad de la señal.

**Palabras clave:** Algoritmos genéticos, facility location problem, cobertura geográfica.

## 1. Introducción

Hoy en día las tecnologías forman parte de nuestras vidas, sobre todo las tecnologías que van en apoyo a las comunicaciones. De éstas la principal es Internet, a tal punto que es posible considerarla como uno de los servicios básicos para cada hogar. Las apuestas de personajes influyentes en el área de las tecnologías de la información como Negroponte [9], apuntan a lograr que en la mayoría de los hogares exista a lo menos un computador con conexión a Internet. Por lo anterior, resulta de importancia que una comunidad específica tenga las herramientas necesarias para que todos los integrantes de esa comunidad tengan acceso a esta tecnología.

Ofrecer el acceso a Internet en un determinado lugar geográfico, como por ejemplo el centro de una ciudad específica, debe considerar diversos factores. Lo primero es contar con dispositivos de transmisión que lo permitan, los que se denominan sensores de red. Existen dos formas de realizar esta transmisión; la primera es mediante sensores de red conectados, los cuales requieren de un lugar físico en el cual instalar el cableado y habilitar los puntos de red necesarios para conectarse a Internet. La segunda forma es mediante el uso de sensores de red inalámbricos. Estos sensores se comunican mediante la modulación de ondas electromagnéticas, radiaciones, o medios ópticos. Estas se propagan sin

necesidad de dispositivos físicos, salvo en el emisor y el receptor. Esta forma de conexión posee la ventaja de permitir un acceso a Internet de forma libre y sin necesidad de cables, lo que entrega movilidad y conectividad al usuario.

Para acceder a esta forma de comunicación se debe contar con un computador que disponga de un dispositivo inalámbrico integrado o de lo contrario con un computador compatible con los dispositivos inalámbricos externos existentes en el mercado (USB, PCMCIA, PCI). Dichos dispositivos poseen parámetros de configuración que deben mantenerse dentro de los rangos estándares para obtener un servicio aceptable. Uno de ellos es la intensidad de la señal, la cual se ve afectada directamente por la ubicación de cada uno de estos sensores. Es decir, la intensidad de esa señal depende de la topología del lugar, como también de las características de las construcciones cercanas a la ubicación de cada sensor.

Otro parámetro a considerar es el ancho de banda máximo que posee cada sensor. Éste limita la cantidad de usuarios que pueden estar conectados al mismo tiempo a través de un sensor. Si el ancho de banda consumido por todos los usuarios de un sensor es mayor que el máximo, los tiempos de respuesta de la conexión disminuyen notablemente.

El obtener una correcta distribución de los sensores inalámbricos de red para *iluminar* con Internet una determinada área geográfica se convierte en un problema de optimización de recursos, ya que la distribución a obtener debe realizarse minimizando la función de costos asociados al problema, y maximizando el área de cobertura de los sensores involucrados en la solución. Este problema converge a una categoría de *problema tipo* la cual se conoce como *Facility Location Problem* (FLP) [8]. Este tipo de problema se presenta, por ejemplo, en un conjunto de fábricas (bodegas), que almacenan un bien o servicio, y se encuentran distribuidas dentro de un lugar geográfico, y un conjunto de usuarios cuyo objetivo central es consumir los bienes o servicios almacenados en las fábricas. Además de realizar el consumo de estos bienes o servicios, el consumo debe ser cargado a la fábrica cuya distancia entre su ubicación y la ubicación del usuario sea mínima, lo cual se traduce en que los costos invertidos por el usuario con respecto a consumo y transporte sean también mínimos. La analogía para el problema antes descrito respecto de brindar la conexión a Internet, se resume a obtener una distribución óptima de fábricas (o en este caso sensores) distribuidos en un determinado lugar geográfico. La solución para estos problemas considera tanto la distribución óptima de dichas *fábricas*, como la cantidad mínima necesaria de éstas para lograr el objetivo.

Este artículo se estructura de la siguiente manera. La sección 1 consta de la presente introducción, a continuación, la segunda sección hace una revisión del estado del arte asociado al problema particular; la tercera sección presenta el problema en consideración, mientras que la cuarta sección describe el enfoque utilizado. La quinta sección da cuenta de las pruebas realizadas y los resultados obtenidos; finalmente la última sección presenta las conclusiones que se derivan del trabajo.

## 2. Estado del arte

El problema asociado a la localización de servicios (FLP) se ha enfrentado desde diversas perspectivas. Lo habitual es hacerlo por medio de programación lineal, como se muestra en [5], donde se menciona la necesidad de resolver el problema a través de una aproximación del esquema primal-dual. Para enfrentar el desafío, modelan el problema usando grafos bipartitos, permitiendo el manejo separado de instalaciones y ciudades, asociando un costo a la apertura de una instalación, y una función que representa el costo de conectar un punto de demanda con una instalación (abierta). La idea es minimizar la función objetivo resultante. El trabajo considera FLP *no capacitado*, para reducir las restricciones y simplificar el problema, lo que repercute en que sea un enfoque menos adecuado cuando se trata con sistemas distribuidos y a gran escala.

Hasan, Jung y Chwa [3] consideran que la apertura de una nueva instalación tiene los mismos costos, independientemente del lugar en donde se produce, y utilizan un algoritmo de programación lineal (LP-rounding). Esta visión es simplificada, y por lo tanto aproximada, pero permite eliminar las restricciones existentes para la ubicación de nuevas instalaciones. Krivitski et al. [8], atacan el problema minimizando una función de costo total; utilizan como entrada una base de datos de puntos, donde cada punto representa a un cliente, un conjunto que contiene posibles ubicaciones, una función de costo, y una configuración que describe las instalaciones en operación. La idea es que dada una función de costo y un conjunto de instalaciones abiertas, la distancia acumulativa de los puntos con las instalaciones cercanas sea minimizada. El esquema es parecido al enfoque anterior, aunque en este caso se agrega una heurística tipo *hill-climbing* para realizar búsquedas locales, obteniendo efectivamente una mejora en los resultados. También existe un trabajo interesante descrito en [4], donde a través de la manipulación de grafos con distancia heredada se van escogiendo subgrafos analizando para cada uno de ellos la distancia de su vecino más cercano, eligiendo la menor para cada subgrafo que modela el problema. En [7], ya más cercana a nuestra propuesta, se estudia la distribución de sensores de red inalámbricos en una determinada área. Ellos concluyen insistiendo en métodos de programación lineal como se describe en [10] o con heurísticas de búsqueda local, como *hill-climbing*. Sin embargo, esta técnica, a pesar de obtener resultados prometedores, dada su naturaleza de búsqueda local, desecha soluciones que pudieran ser eventualmente mejores, limitando negativamente, en este caso, el espacio de soluciones. Chaudhry et al. [1], reducen el problema a la instalación de  $p$  fábricas que permitan atender a un número determinado de clientes. El objetivo es minimizar la distancia promedio recorrida por los clientes para llegar a la fábrica más cercana. Este trabajo en esencia resuelve un problema tipo FLP utilizando algoritmos genéticos. Este enfoque permite obtener buenos resultados en tiempos reducidos, aunque en algunos casos se generan soluciones no factibles. Finalmente, el trabajo descrito en [6] a pesar de resolver un problema tipo FLP con algoritmos genéticos, introduce consideraciones al modelo que implican el manejo de restricciones de alto nivel, esto permite obtener buenos resultados,

pero al costo de tener tiempos de procesamiento considerables

El trabajo que aquí se presenta combina algoritmos genéticos y la técnica FLP para proponer soluciones al problema de hacer llegar una señal de Internet a un sector geográfico acotado, relajando la restricción de minimizar el número de sensores, pero manteniendo este aspecto en el cálculo del costo. Adicionalmente, utilizando como base una situación real, se realiza una sintonía de los parámetros del algoritmo para disminuir los costos involucrados. Así, se pretende proponer alternativas que, pese a no cubrir la totalidad de una zona geográfica, ofrecen soluciones con relaciones costo/cobertura razonables, y a su vez ajustables.

### 3. El problema

La principal dificultad de brindar una señal de Internet es que se necesita contar con una cantidad variable de sensores la cual es directamente proporcional a la cantidad de edificios que existan en cada calle (los edificios actúan como barreras para la señal), además del flujo de usuarios que se pretende abastecer con el servicio. Esto significa que se debe obtener las ubicaciones óptimas para instalar sensores de red, de modo de alcanzar a un conjunto dado de  $n$  clientes. Para ello se va a evaluar la generación de un conjunto de posibles ubicaciones en las cuales se puede instalar un sensor, y en donde dicha instalación en la ubicación  $i$  incurre en un costo, y cada cliente  $j$  debe estar conectado solamente a un sensor, lo que involucra un costo proporcional a la distancia entre el sensor  $i$  y el cliente  $j$ . El objetivo principal planteado para este problema es obtener la distribución de sensores que minimice la función de costos totales que involucra todos los parámetros antes mencionados.

#### Algoritmos genéticos

Una técnica frecuentemente utilizada para resolver problemas de optimización son los algoritmos genéticos [2]. La principal ventaja que ofrece esta técnica es que no se necesita contar con información específica del problema a resolver, sino que trabaja con modelos de abstracción que asumen este rol. Por otro lado obtienen de forma simultánea otras soluciones, en vez de calcularlas secuencialmente como las otras técnicas. Y por último, esta técnica está basada en principios evolutivos los cuales son implementados mediante el uso de operadores estocásticos, lo que agrega dinamismo y azar al proceso completo de búsqueda dentro del espacio de soluciones.

Los algoritmos genéticos son métodos adaptativos que pueden usarse para la resolución de problemas de optimización, mediante una analogía con los procesos evolutivos biológicos. La emulación de esos procesos biológicos involucra una codificación de una potencial solución al problema en un *cromosoma*, lo que se traduce en una estructura de datos con características específicas, constituyéndose en una representación, que identifica a los individuos a ser considerados por el algoritmo genético. Luego de definir la representación a utilizar, se debe generar la población inicial, cuya fortaleza radica en la diversidad de

sus elementos (cromosomas) para disminuir el riesgo de los óptimos locales. Una forma habitual de obtener la población inicial es generarla de manera aleatoria. A continuación se define una función de adaptación (*fitness*) que permite evaluar la calidad de una solución, lo que normalmente se mide en términos de costo.

**Operadores genéticos:** Para este trabajo se consideraron diferentes operadores genéticos. Estos operadores se describen brevemente a continuación:

- **Selección.** La selección se lleva a cabo utilizando el mecanismo de *ruleta* [2]. Esto significa que los individuos con mejores valores de *fitness* tendrán mayor probabilidad de ser escogidos como padres en el proceso de reproducción.
- **Cruzamiento.** El cruzamiento se usa para intercambiar material genético, permitiendo que parte de la información genética de un individuo se combine con parte de la información genética de otro individuo diferente.
- **Mutación.** Al usar este operador genético, se introduce una pequeña variación en la población de manera que se crea nuevo material genético.

#### 4. El enfoque propuesto

En analogía con la técnica FLP descrita anteriormente, cada sensor inalámbrico se convierte en una fábrica, el ancho de banda de cada sensor se convierte en el tamaño de la fábrica y los clientes que debe satisfacer esa fábrica se transforman en la señal entregada por dicho sensor a su vecindad. Respecto de las restricciones para el problema están el que un usuario se conecta a Internet a través de un solo sensor a la vez, los edificios en un sector influyen de forma negativa respecto de la señal, si hay muchos edificios en un sector se requerirá una mayor cantidad de sensores para poder *iluminar* dicho sector con la señal. También, a medida que hay más edificios en un sector, la cantidad de usuarios que requieren Internet aumenta de forma proporcional a la densidad de población.

Para modelar el problema se tuvo en consideración un conjunto de aspectos que se resumen a continuación:

1. Un sensor  $k$  posee un determinado ancho de banda, denotado por  $W_k$ ; además tiene un radio de alcance  $A_k$  que presenta dos valores, uno referido al alcance al interior de las edificaciones, denotado por  $A_{int(k)}$  y uno para exteriores, denotado por  $A_{ext(k)}$ .
2. El área geográfica considerada se representó por medio de un grafo con  $n$  nodos. Cada nodo representa la ubicación potencial de un sensor; alternativamente, representa a un grupo de usuarios consumidores de la señal de Internet.
3. El número de sensores considerados es una variable crítica, ya que existe un costo asociado a la instalación de cada nuevo sensor. Se denota  $s$  a la cantidad de sensores considerados y  $\delta$  al costo asociado a cada sensor.
4. Todos los sensores están conectados entre sí.

5. Existen lugares predeterminados para la ubicación de los sensores; en el caso de esta presentación, se eligieron los postes de luz como esos lugares, denotados por  $P$ . Una suposición extra es que los postes de luz están separados por una distancia de 20 metros.
6. Cada usuario se conecta a un solo sensor en un instante de tiempo.
7. El alcance de un sensor depende de la cantidad de edificaciones existentes. Si hay edificaciones en torno al sensor, el alcance se calcula utilizando  $A_{int}$ . Si no hay edificaciones cercanas al sensor, su alcance será  $A_{ext}$ .

Para la resolución del problema, se tomó en cuenta factores como cobertura (denotado por  $C$ ), distancia (denotado por  $L$ ) y flujo entre dos nodos, denotado por  $F$ . Sea  $C_{ij} \in \{0, 1\}$  una función de cobertura, cuyo resultado es 1 si el nodo  $i$  es *iluminado* por la señal de un sensor ubicado en el nodo  $j$ , y 0 en caso contrario.  $C_{kk} = 1$  indica que el nodo  $k$  es un sensor. Sea  $L_{ij}$  la distancia entre los nodos  $i$  y  $j$ , y  $F_{ji}$  el flujo desde el nodo  $j$  al nodo  $i$ . Estas consideraciones llevan a las siguientes expresiones:

$Costo = G_0 + \sum_{i=1}^n \sum_{k=1}^n L_{ik} * C_{kk} * \delta$ , donde  $G_0$  son los costos iniciales, y  $\delta$  se asocia a la ubicación de cada sensor.  $\sum_{k=1}^n C_{ik} = 1, \forall i \in \{1, \dots, n\}$  refleja la exclusividad en la conexión de un usuario con un único sensor. Lo anterior sujeto a las restricciones  $\sum_{k=1}^n C_{kk} = s, \forall k \in S, C_{ik} \leq C_{kk}, \forall i, k \in \{1, \dots, n\}$ , y  $C_{ik} \in \{0, 1\}, \forall i, k \in \{1, \dots, n\}$ . Esto refleja además la cantidad de sensores usados,  $s$ ; el hecho de que un sensor es activado a través de la red formada por los sensores activos, evitando la transmisión directa entre nodos que no son sensores. Además se refleja el hecho de que un usuario ubicado en un nodo  $i$  puede ser atendido o no por un sensor ubicado en  $k$  (valores 1 y 0 respectivamente).

Para el problema particular modelado,  $F$  se consideró un valor constante, y se optó por emplear  $A_{int}$  como factor de alcance, por tratarse del modelamiento del centro de una ciudad.

## 5. Pruebas y resultados

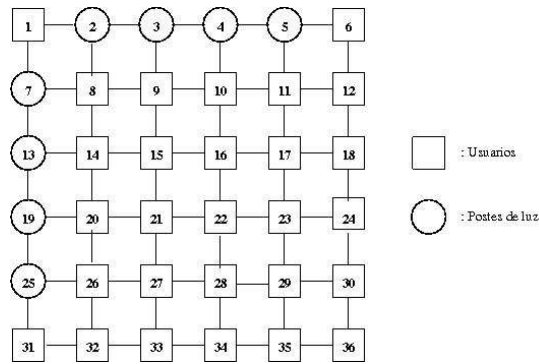
El área utilizada para las pruebas comprende el sector céntrico una ciudad arbitraria de tamaño medio. Se consideró una cantidad de diez calles por un lado y ocho por el otro lado, lo que corresponde aproximadamente a 80 hectáreas (ochocientos mil metros cuadrados). La figura 1 esquematiza el área problema.

Cada cuadra (aproximadamente 100 metros) se divide en 6 tramos que pueden ser solicitantes del servicio de Internet. Aquí también están considerados los postes de luz, que son los lugares en donde se ubican físicamente los sensores. De esta manera, una manzana, delimitada por cuatro cuadras, se traduce en 36 puntos eventuales de demanda o servicio. Luego, este sector comprende un total posible de 2880 puntos a ser considerados como posibles demandantes, de los cuales 640 son postes de luz. Utilizando una representación de grafos, cada uno de estos puntos es un nodo, siendo las aristas del grafo la distancia entre los nodos (ver Figura 2). Se genera una matriz triangular con las distancias entre cada par de puntos del modelo, utilizando el algoritmo de Dijkstra del camino



**Figura 1.** Área problema seleccionada para las pruebas.

mínimo, a la cual se accede durante el proceso posterior de cálculo. El criterio para determinar la bondad de un individuo (solución) está dado por un compromiso entre cobertura y costo de conexión de los  $s$  sensores. Se trata en este caso de minimizar la función de costo que toma en cuenta la inversión material, modulada por la cantidad de usuarios que es posible atender. La figura 2 ilustra una distribución de nodos en una manzana arbitraria, que cubre un área de aproximadamente 10000 metros cuadrados.



**Figura 2.** Distribución de nodos en una manzana.

Cada individuo se representa mediante una estructura de datos que se separa en dos partes: la primera, contiene los índices de la ubicación, en el grafo, de cada uno de los  $s$  sensores considerados; la segunda parte contiene todos los índices restantes, que corresponden a los nodos que no son sensores.

Lo primero que se hace es calcular aquellos valores que permanecerán constantes durante todo el proceso y que son válidos para todos los individuos de la población; la ya mencionada distancia más corta entre cada par de nodos y la ubicación de los postes de luz.

La población de soluciones considera la restricción de que cada sensor está ubicado en un poste de luz (nodos redondos en la figura 2). Inicialmente se distribuyen los sensores de manera aleatoria, obteniendo los índices desde el conjunto  $P$ . El proceso de selección considera el clásico método de la ruleta [2]. Para el cruzamiento, se considera el intercambio de características sólo de la primera parte del cromosoma (que contiene los sensores). Dados dos progenitores escogidos para cruzamiento, si existen sensores repetidos, se reordenan en la cadena, ubicándolos al inicio de la misma, y luego se elige aleatoriamente una subcadena para el intercambio. Esta subcadena no considera la parte inicial de sensores repetidos. Esto permite que los hijos no presenten el problema de asignar un mismo sensor a dos ubicaciones diferentes.

La mutación es implementada cambiando aleatoriamente, con baja probabilidad de ocurrencia, un sensor del cromosoma por otro no considerado en ese individuo (i.e., se desplaza el sensor desde un poste de luz a otro sin sensores). Para conservar los mejores individuos, la estrategia del algoritmo genético considera un 10 % de elitismo. Este 10 % forma parte, junto con los descendientes del cruzamiento, de la siguiente generación de soluciones posibles. Los parámetros utilizados para las pruebas consideraron los valores mostrados en la Tabla 1.

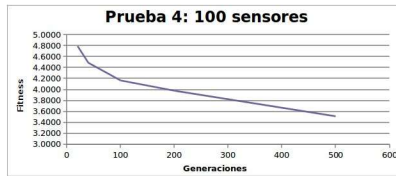
**Tabla 1.** Parámetros de prueba.

Parámetro	Valor(es)
Número de sensores	50, 80, 90, 100, 150, 200
Operador de selección	Método de la ruleta
Probabilidad de cruzamiento	60 %, 80 %
Probabilidad de mutación	2 %
Porcentaje de elitismo	10 %
Tamaño de la población	20, 80 (individuos)
Número de generaciones	500

La tabla a continuación (Tabla 2), muestra los mejores valores obtenidos para las distintas cantidades de sensores considerados. Las unidades de tiempo y las unidades asociadas al costo son arbitrarias y referenciales; su objetivo es simplemente mostrar los cambios en las magnitudes.

Si bien es cierto que a medida que se aumenta el número de sensores en el modelo, se aumenta también la cantidad de usuarios que es posible atender, esto implica un mayor costo en la habilitación de sensores. A partir de cierto número de sensores, el aumento de los costos tiene una velocidad de crecimiento mayor que el aumento en la cantidad de usuarios atendidos. Esto provoca que el





**Figura 4.** Evolución del fitness para el caso de 100 sensores.

resultados obtenidos comienzan a estabilizarse alrededor de la generación 500, los mejoramientos en la calidad de las soluciones, a partir de ese punto, son poco importantes.

## 6. Conclusiones

El enfoque de algoritmos genéticos desarrollado en esta propuesta permite analizar el problema de distribución de sensores de red desde una perspectiva diferente. Al utilizar esta estrategia para abordar un problema del tipo Facility Location Problem se pueden considerar aspectos relevantes como son las funciones de oferta y demanda, la cantidad de sensores utilizados y los alcances de estos sensores.

Ante la ausencia de propuestas similares, el modelo que se propone permite priorizar diferentes aspectos involucrados en el diseño de las soluciones; lo que es uno de los aspectos positivos del mecanismo que aquí se sugiere. Estas diferentes priorizaciones, minimizar el costo o maximizar la cobertura, permiten generar cuadros comparativos cuantificados para apoyar una toma de decisiones más informada.

Finalmente, debido a la forma de modelamiento del problema, es posible realizar cambios tanto en los datos iniciales de configuración del área geográfica como en los cálculos de las funciones involucradas. Esto permite reutilizar este modelo para aplicarlo a áreas de diferente tamaño, o a ciudades distintas a la considerada en las pruebas mostradas.

## Referencias

1. Chaudhry, S., He, S., and Chaudhry, P.: Solving a class of facility location problems using genetic algorithms. *Expert Systems*, 20(2):86-91 (2003)
2. Floreano, D., Mattiussi, C.: *Bio-Inspired Artificial Intelligence. Theories, Methods, and Technologies*. The MIT Press (2008)
3. Hasan, M., Jung, H., Chwa, K.: Approximation algorithms for connected facility location problems. *Journal of Combinatorial Optimization*, 16(2):155-172 (2008)
4. Hsieh, S.: An efficient parallel strategy for the perfect domination problem on distance-hereditary graphs. *The Journal of Supercomputing*, 39(1):39-57, Springer (2007)

5. Jain, K. and Vazirani, V.: Primal-dual approximation algorithms for metric facility location and k-median problems. *Foundations of Computer Science*, 40th Annual Symposium, pp. 2-13, October (1999)
6. Kratica, J., Stanimirovic, Z., Tasic, D., and Filipovic, V.: Two genetic algorithms for solving the uncapacitated single allocation p-hub median problem. *European Journal of Operational Research*, 182(1):15-28 (2007)
7. Krivitski, D., Schuster, A., and Wolf, R.: A local facility location algorithm for sensor networks. *Lecture Notes in Computer Science*, 3560:368-375, Springer (2005)
8. Krivitski, D., Schuster, A., and Wolf, R.: A local facility location algorithm for large-scale distributed systems. *J. Grid Comput.*, 5(4):361-378, Springer (2007)
9. Negroponte, N.: *Being digital*. Random House Inc., New York, NY, USA (1996)
10. Sviridenko, M.: An improved approximation algorithm for the metric uncapacitated facility location problem, *IPCO*, pp. 240-257, Springer (2002)