

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI**

Đinh Thị Hà Ly

**GIẢI THUẬT XÁP XỈ GIẢI BÀI TOÁN BAO PHỦ BARRIER
TRONG MẠNG CẢM BIẾN KHÔNG DÂY**

LUẬN VĂN THẠC SĨ KHOA HỌC
Khoa học máy tính

Hà Nội – Năm 2017

ĐINH THỊ HÀ LY

KHOA HỌC MÁY TÍNH

KHÓA 2016B

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI**

Đinh Thị Hà Ly

**GIẢI THUẬT XÁP XỈ GIẢI BÀI TOÁN BAO PHỦ BARRIER
TRONG MẠNG CẢM BIẾN KHÔNG DÂY**

Chuyên ngành : Khoa học máy tính

LUẬN VĂN THẠC SĨ KHOA HỌC
Khoa học máy tính

NGƯỜI HƯỚNG DẪN KHOA HỌC:
PGS. TS. Huỳnh Thị Thanh Bình

Hà Nội – Năm 2017

MỤC LỤC

LỜI CAM ĐOAN.....	5
DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT.....	6
DANH MỤC CÁC BẢNG	7
DANH MỤC CÁC HÌNH VẼ, ĐỒ THỊ	8
MỞ ĐẦU	9
CHƯƠNG 1: BÀI TOÁN BAO PHỦ BARRIER TRONG MẠNG CẢM BIẾN KHÔNG DÂY.....	11
1.1. Tổng quan bài toán bao phủ trong mạng cảm biến không dây.....	11
1.1.1. Độ bao phủ của cảm biến.....	12
1.1.2. Độ bao phủ của mạng cảm biến	16
1.2. Bài toán bao phủ barrier trong mạng cảm biến không dây	19
1.3. Bài toán tìm đường xâm nhập có khả năng bị phát hiện nhỏ nhất trong mạng cảm biến không dây	28
CHƯƠNG 2: CÁC THUẬT TOÁN XẤP XỈ GIẢI BÀI TOÁN TỐI ƯU	33
2.1. Bài toán tối ưu.....	33
2.1.1. Bài toán tối ưu tổ hợp	34
2.1.2. Bài toán tối ưu liên tục.....	34
2.2. Tổng quan các phương pháp xấp xỉ giải bài toán tối ưu	36
2.3. Giải thuật di truyền.....	43
2.3.1. Mã hóa cá thể	43
2.3.2. Khởi tạo quần thể	44
2.3.3. Hàm thích nghi	45
2.3.4. Các toán tử di truyền	45
2.3.5. Chọn lọc	47
2.3.6. Điều kiện dừng của giải thuật	47
2.3.7. Tính chất của giải thuật di truyền.....	48

CHƯƠNG 3: GIẢI THUẬT ĐỀ XUẤT GIẢI BÀI TOÁN TÌM ĐƯỜNG XÂM NHẬP CÓ KHẢ NĂNG BỊ PHÁT HIỆN NHỎ NHẤT TRONG MẠNG CẢM BIẾN KHÔNG DÂY MÔ HÌNH XÁC SUẤT.....	49
3.1. Mã hóa cá thể.....	49
3.2. Hàm thích nghi.....	52
3.3. Toán tử di truyền.....	52
3.3.1. Lai ghép	52
3.3.2. Đột biến.....	56
3.4. Chọn lọc.....	57
CHƯƠNG 4: KẾT QUẢ ĐẠT ĐƯỢC	59
4.1. Dữ liệu thử nghiệm.....	59
4.2. Thiết lập hệ thống.....	59
4.3. Kết quả thực nghiệm	60
4.4. Kết luận.....	64
DANH MỤC TÀI LIỆU THAM KHẢO	65

LỜI CAM ĐOAN

Tôi xin cam đoan:

1. Những nội dung trong luận văn này là công trình nghiên cứu của tôi dưới sự hướng dẫn trực tiếp của PGS.TS. Huỳnh Thị Thanh Bình
2. Mọi tham khảo dùng trong luận văn đều được trích dẫn rõ ràng tên tác giả, tên công trình, thời gian, địa điểm công bố.
3. Các số liệu, kết quả nêu trong luận văn là trung thực và chưa từng được ai công bố trong bất kỳ công trình nào khác.
4. Mọi sao chép không hợp lệ, vi phạm quy chế đào tạo, hay gian trá, tôi xin chịu hoàn toàn trách nhiệm.

Tác giả luận văn

(Ký và ghi rõ họ tên)

Đinh Thị Hà Ly

DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

Chữ viết tắt	Viết đầy đủ	Ý nghĩa
ACO	Ant Colony Optimization	Tối ưu hóa đàn kiến
PSO	Particle Swarm Optimization	Tối ưu hóa bầy đàn
MEP	Minimal Exposure Path	Đường đi có khả năng bị phát hiện nhỏ nhất
GA	Genetic Algorithm	Giải thuật di truyền
SA	Simulated Annealing	Giải thuật mô phỏng luyện kim
WSN	Wireless Sensor Network	Mạng cảm biến không dây

DANH MỤC CÁC BẢNG

Bảng 1. Tham số mô hình cảm biến xác suất	59
Bảng 2. Các tham số của GAMEP	59
Bảng 3. Bảng so sánh giá trị exposure tốt nhất và giá trị xác suất phát hiện tương ứng (%) tìm được bởi GAMEP sau 5 lần chạy trên các bộ dữ liệu với các giá trị len khác nhau	61
Bảng 4. Bảng so sánh giá trị exposure trung bình và giá trị xác suất phát hiện tương ứng (%) tìm được bởi GAMEP sau 5 lần chạy trên các bộ dữ liệu với các giá trị len khác nhau	62
Bảng 5. Bảng so sánh giá trị độ lệch chuẩn và thời gian chạy trung bình của GAMEP sau 5 lần chạy trên các bộ dữ liệu với các giá trị len khác nhau.....	63

DANH MỤC CÁC HÌNH VẼ, ĐỒ THỊ

Hình 1. Mô hình cảm biến nhị phân.....	13
Hình 2. Mô hình tỉ lệ	14
Hình 3. Bao phủ đối tượng	17
Hình 4. Bao phủ barrier	20
Hình 5. Biểu đồ Voronoi và đường vi phạm tối đa [22]	23
Hình 6. Biểu đồ Delaunay và đường có mức hỗ trợ tối đa [22]	24
Hình 7. Minh họa bài toán tối ưu liên tục: đồ thị hàm số $f(x) = x^3 - 3x^2 + 4$	36
Hình 8. Cơ chế cập nhật cá thể trong PSO	40
Hình 9. Sơ đồ thuật toán tiến hóa.....	42
Hình 10. Lai ghép một điểm cắt.....	46
Hình 11. Biểu diễn cá thể	49
Hình 12. Khởi tạo cá thể.....	50
Hình 13. Phép lai ALX- α	54
Hình 14. Minh họa phép lai một điểm cắt	55
Hình 15. Minh họa quá trình đột biến	57
Hình 16. Độ hội tụ của GAMEP trên bộ dữ liệu S_40_1	63

MỞ ĐẦU

1. Lý do chọn đề tài

Hiện nay, mạng cảm biến không dây đang đóng một vai trò quan trọng trong thực tiễn cuộc sống hàng ngày và là một nhân tố không thể thiếu cho sự hình thành thế giới vạn vật (Internet of Things). Với các ứng dụng liên quan đến giám sát, an ninh của mạng cảm biến không dây, bao phủ rào chắn (barrier) được xem là bài toán then chốt cần giải quyết. Bằng cách đưa ra lời giải cho bài toán này, các nhà xây dựng mạng hay kẻ tấn công sẽ có được các thông tin hữu ích về tình hình chất lượng của mạng, từ đó phía xây dựng mạng có thể đưa các phương án phù hợp cải thiện hay duy trì mạng cảm biến đang có, còn kẻ tấn công sẽ có cơ sở để quyết định tiến hành xâm nhập hoặc tìm con đường khác an toàn hơn. Xa hơn nữa, khi được kết nối vào mạng Internet, việc giải quyết tốt bài toán này sẽ góp phần hình thành nên các ứng dụng như tòa nhà thông minh (smart buildings), thành phố thông minh (smart cities), bảo vệ thông minh (intelligent protection), ...

Có nhiều mô hình khác nhau cho lớp bài toán bao phủ barrier, mỗi mô hình này đại diện cho một ứng dụng và nhu cầu trong thực tế. Luận văn tập trung nghiên cứu mô hình tìm đường xâm nhập có khả năng bị phát hiện nhỏ nhất trong mạng cảm biến không dây, từ đó phát biểu mô hình này trong ngữ cảnh xác suất mà cho phép xét đến sự ảnh hưởng của nhiễu từ môi trường đến quá trình và đưa ra kết quả của các cảm biến. Mặc dù chưa được chứng minh thuộc lớp bài toán NP-khó, nhưng với không gian tìm kiếm lớn và liên tục, việc tìm ra lời giải chính xác cho mô hình này đòi hỏi nhiều thời gian và kỹ thuật. Do đó, tác giả lựa chọn cách tiếp cận giải xấp xỉ, gần đúng.

Như vậy, luận văn sẽ tiến hành nghiên cứu tổng quan về bài toán bao phủ barrier, cụ thể là lớp bài toán tìm đường xâm nhập có khả năng bị phát hiện nhỏ nhất, sau đó phát biểu bài toán này dưới mô hình xác suất và đề xuất một giải thuật xấp xỉ, gần đúng để giải. Trên cơ sở kết quả đạt được, luận văn tổng kết, đánh giá và đưa ra hướng phát triển của đề tài trong tương lai.

2. Mục đích nghiên cứu

Luận văn nhằm nghiên cứu tổng quan bài toán bao phủ rào chắn (barrier coverage) trong mạng cảm biến không dây và đi sâu vào một mô hình cụ thể. Trên cơ sở đó, luận văn đề xuất giải thuật di truyền để giải quyết bài toán. Cài đặt, thử nghiệm thuật toán đề xuất để đưa giá nhận xét, đánh giá.

3. Đối tượng nghiên cứu

Luận văn tập trung nghiên cứu khả năng phát hiện đối tượng xâm nhập của mạng cảm biến không dây, còn gọi là giá trị exposure.

4. Phạm vi nghiên cứu

Mạng cảm biến không dây được nghiên cứu trong luận văn là mạng cảm biến mô hình xác suất, nghĩa là các cảm biến trong mạng có khả năng cảm nhận (phát hiện) sự kiện (đối tượng) tuân theo mô hình xác suất.

5. Phương pháp nghiên cứu

- Phương pháp nghiên cứu tài liệu
 - Đọc và đánh giá các công trình khoa học trong và ngoài nước đã công bố liên quan đến đề tài nghiên cứu của luận văn.
 - Áp dụng kiến thức toán học xác suất
- Phương pháp thực nghiệm
 - Tiến hành thực nghiệm để đánh giá hiệu năng của giải thuật đề xuất

CHƯƠNG 1: BÀI TOÁN BAO PHỦ BARRIER TRONG MẠNG CẢM BIẾN KHÔNG DÂY

Chương này trình bày các vấn đề sau:

- Tổng quan về bài toán bao phủ trong mạng cảm biến không dây
- Lớp bài toán bao phủ barrier trong mạng cảm biến không dây
- Phát biểu bài toán tìm đường xâm nhập có khả năng bị phát hiện thấp nhất trong mạng cảm biến không dây
- Ứng dụng của bài toán

1.1. Tổng quan bài toán bao phủ trong mạng cảm biến không dây

Mạng cảm biến không dây (Wireless Sensor network - WSN) được hình thành bởi một tập các cảm biến năng lượng thấp (low-power sensor), gọi là các nút mạng, được triển khai để thu thập thông tin về thế giới vật lý xung quanh như nhiệt độ, độ ẩm.... tại một khu vực nào đó (Region of Interest). Các thông tin thu thập này sẽ được các cảm biến xử lý và/hoặc gửi về các trạm điều khiển (base station) và/hoặc các nút nguồn (sink node) thông qua kết nối không dây (wireless link) và giao thức định tuyến (routing protocol) đã được thiết lập. Với những tính chất như vậy, mạng cảm biến không dây có thể triển khai được ở nhiều dạng địa hình phức tạp và nhờ đó có nhiều ứng dụng trong các lĩnh vực của đời sống như môi trường, quân sự, y tế... [12].

Tuy nhiên, để mạng cảm biến không dây thực sự có hiệu quả, đáp ứng các yêu cầu từ người sử dụng thì vẫn còn nhiều vấn đề cần được nghiên cứu và giải quyết như bài toán sử dụng năng lượng, độ tin cậy của mạng, lập lịch truyền tin.... Và bao phủ cũng là một trong số những bài toán này.

Theo [3, 5, 20], bao phủ (coverage) là một trong những tiêu chí, thước đo quan trọng nhất được dùng để đánh giá chất lượng của mạng cảm biến không dây. Dựa vào tham số này, con người có thể biết khu vực đặt cảm biến được giám sát tốt như thế nào.

Có 2 mức độ bao phủ trong một mạng cảm biến không dây [5], đó là:

- Độ bao phủ của từng cảm biến (Individual sensor coverage)
- Độ bao phủ của toàn mạng cảm biến (Network-wide sensing coverage)

1.1.1. Độ bao phủ của cảm biến

Đối với mỗi cảm biến, độ bao phủ thể hiện chất lượng giám sát của cảm biến đó và được biểu diễn bởi các mô hình cảm biến (sensing model), còn gọi là mô hình bao phủ. Có nhiều cách khác nhau để phân loại các mô hình này. Theo [5], mô hình bao phủ của cảm biến gồm 3 loại như sau:

- Mô hình nhị phân (Binary model)
- Mô hình tỉ lệ (Index model hay Attenuated disk coverage model)
- Mô hình xác suất (Probabilistic model)

Mô hình nhị phân

Mô hình nhị phân tổng quát biểu diễn miền bao phủ của một cảm biến s bởi một đĩa tròn có bán kính cảm biến R và tâm là vị trí C của cảm biến đó (hình 1a). Đối tượng T tại một điểm l trong không gian hai chiều được coi là bị phát hiện bởi cảm biến đặt tại C nếu khoảng cách Euclide $d(s, l)$ giữa chúng không lớn hơn bán kính cảm biến R . Mô hình này được biểu diễn bởi hàm cảm biến f dựa trên khoảng cách giữa vị trí s và l như sau:

$$f(d(s, l)) = \begin{cases} 1 & \text{khi } d(s, l) \leq R \\ 0 & \text{khi } d(s, l) > R \end{cases} \quad (1)$$

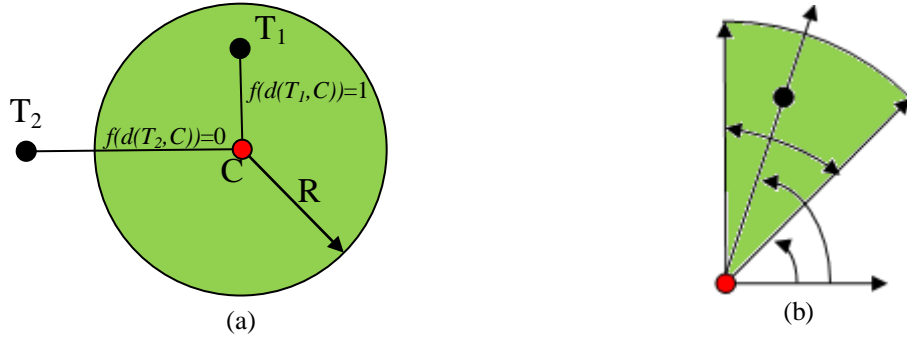
Đây là mô hình cảm biến đơn giản nhất và được sử dụng nhiều nhất trong các bài toán liên quan đến vấn đề bao phủ trong mạng cảm biến không dây.

Một dạng đặc biệt của mô hình này là mô hình nhị phân có hướng (directional boolean model), trong đó miền bao phủ của cảm biến là một miền hình quạt (sector – hình 1b) được xác định bởi bán kính cảm biến R , góc cảm biến ω , hướng của góc cảm biến φ_s với hàm cảm biến (2). Theo đó, nếu góc φ_t nằm ngoài khoảng $[\varphi_s, \varphi_s + \omega]$ hoặc

khoảng cách $d(s, l)$ lớn hơn R thì cảm biến s tại C không cảm nhận được đối tượng T tại l .

$$f(d(s, l), \varphi_t) = \begin{cases} 1 & \text{khi } \varphi_s \leq \varphi_t \leq \varphi_s + \omega \text{ và } d(s, l) \leq R \\ 0 & \text{khi } \varphi_s > \varphi_t \text{ hay } \varphi_t > \varphi_s + \omega \text{ hay } d(s, l) > R \end{cases} \quad (2)$$

Mô hình này thường được sử dụng khi miền bao phủ của cảm biến bị che khuất bởi các yếu tố liên quan đến địa hình.



Hình 1. Mô hình cảm biến nhị phân

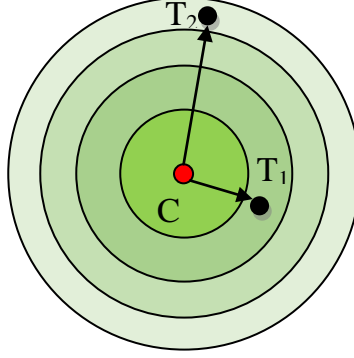
Mô hình tỉ lệ

Cho rằng chất lượng giám sát, bao phủ của cảm biến sẽ giảm đi khi khoảng cách giữa nó với đối tượng T tăng dần, các nhà nghiên cứu đã đề xuất mô hình cảm biến tỉ lệ có công thức như sau:

$$f(d(s, l)) = \frac{c}{d^\lambda(s, l)} \quad (3)$$

với $d(s, l)$ là khoảng cách Euclide giữa cảm biến s đặt tại C và đối tượng T tại l ; c là giá trị hằng được xác định bởi các thuộc tính của mạng, λ hệ số suy giảm tín hiệu theo khoảng cách. Hàm f này trả về giá trị thực không âm. Những điểm càng gần vị trí của cảm biến s cho giá trị hàm f càng cao, nghĩa là càng dễ được phát hiện bởi cảm biến.

Hình 2 là một ví dụ minh họa cho mô hình tỉ lệ với khả năng phát hiện của cảm biến được thể hiện qua sắc độ của màu trong từng vòng tròn. Các vòng tròn càng gần vị trí C có màu càng đậm.



Hình 2. Mô hình tỉ lệ

Mô hình xác suất

Có nhiều định nghĩa khác nhau cho mô hình cảm biến xác suất trong mạng cảm biến không dây, phù hợp với từng ứng dụng cụ thể. Trong luận văn này, tác giả chỉ trình bày hai mô hình thường được sử dụng như sau.

Mô hình xác suất đơn giản, được xem là dạng mở rộng của mô hình nhị phân [25, 5]. Trong đó định nghĩa một số thực R_u nhỏ hơn bán kính bao phủ R của cảm biến và đoạn mở $(R - R_u, R + R_u)$ với xác suất phát hiện đối tượng trên đoạn này là p . Dựa trên các tham số vừa được định nghĩa, hàm cảm biến được tính toán như sau:

$$f(d(s, l)) = \begin{cases} 0 & \text{khi } d(s, l) \leq R - R_u \\ e^{-\omega a^\beta} & \text{khi } R - R_u < d(s, l) \leq R \\ 1 & \text{khi } R < d(s, l) \end{cases} \quad (4)$$

với l là vị trí của đối tượng, $d(s, l)$ là khoảng cách Euclide giữa cảm biến s và l ; $a = d(s, l) - (R - R_u)$ và ω, β là các tham số định lượng cho xác suất phát hiện đối tượng trong một khoảng cách nhất định so với vị trí đặt cảm biến. Xác suất phát hiện đối tượng càng cao thì cảm biến đó có chất lượng và khả năng giám sát càng tốt.

Mô hình trên chỉ đơn giản dựa vào khoảng cách và bán kính cảm biến để đưa ra xác suất phát hiện. Trong trường hợp muốn xét đến sự ảnh hưởng của của nhiễu lên quá trình và đưa ra kết quả của cảm biến, thì mô hình thứ hai sau đây được sử dụng. Theo [5, 1], mô hình xác suất này được phát biểu như sau:

Xét cảm biến s và một đối tượng ở vị trí l trong không gian hai chiều. Bên cạnh việc tín hiệu bị suy giảm theo khoảng cách như trong mô hình tỉ lệ, mô hình xác suất còn tính đến sự ảnh hưởng của nhiễu từ môi trường bằng cách đưa thêm tham số η vào hàm cảm biến như trong công thức (4) dưới đây.

$$f(d(s,l)) = \frac{C}{d^\lambda(s,l)} + \eta \quad (5)$$

Trong đó, η là một biến ngẫu nhiên tuân theo phân phối Gauss với trung vị μ và phương sai σ^2 . Khi đó, xác suất phát hiện P_s^l của cảm biến s cho một sự kiện (đối tượng) tại vị trí l được định nghĩa như sau.

$$P_s^l = \Pr[f(d(s,l)) \geq A] = \Pr\left[\frac{C}{d^\lambda(s,l)} + \eta \geq A\right] \quad (6)$$

với A là một ngưỡng cho trước. Công thức (5) chỉ ra rằng một cảm biến ra quyết định một đối tượng (sự kiện) có ở vị trí l hay không khi độ đo của nó bao gồm cả nhiễu không nhỏ hơn ngưỡng A và ngược lại.

Do $\eta \sim \mathcal{N}(\mu, \sigma^2)$ nên để đơn giản mà không làm mất tính tổng quát, đặt $\mu=0$, khi đó, công thức (5) được tính như sau:

$$\begin{aligned} P_s^l &= \Pr\left[\frac{C}{d^\lambda(s,l)} + \eta \geq A\right] = \Pr\left[\eta \geq A - \frac{C}{d^\lambda(s,l)}\right] \\ &= Q\left(\frac{A - \frac{C}{d^\lambda(s,l)}}{\sigma}\right) \end{aligned} \quad (7)$$

Trong đó, $Q(\cdot)$ là hàm Q-function [6] và được định nghĩa như sau:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{t^2}{2}} dt \quad (8)$$

Đây cũng là mô hình mà luận văn sử dụng để giải quyết bài toán được trình bày trong mục 1.3 dưới đây.

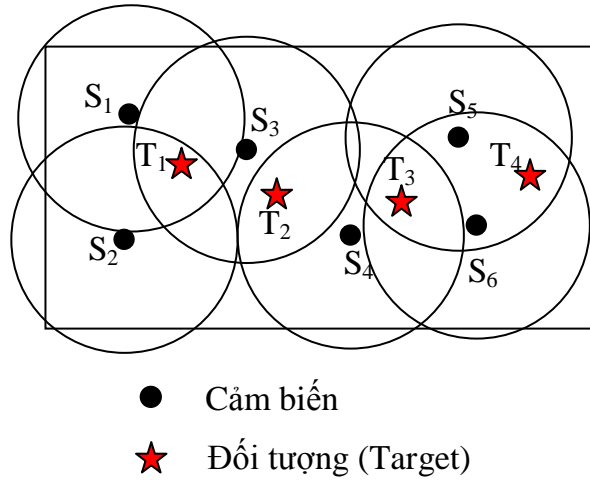
1.1.2. ***Độ bao phủ của mạng cảm biến***

Xét đến độ bao phủ ở mức toàn mạng cảm biến, trong bài khảo sát năm 2011, Bang Wang [5] đã chia các mô hình bao phủ trong mạng cảm biến không dây thành 3 nhóm theo tiêu chí được giám sát, gồm:

- Bao phủ đối tượng (Target coverage hay Point coverage)
- Bao phủ barrier
- Bao phủ diện tích (Area coverage)

Bao phủ đối tượng

Trong mô hình bao phủ đối tượng của mạng cảm biến không dây, một tập các đối tượng (target) đã biết vị trí sẽ được giám sát bởi các cảm biến. Và để giám sát toàn bộ các đối tượng này, vị trí của các cảm biến có thể được tính toán trước hoặc được triển khai một cách ngẫu nhiên. Đây cũng là hai hướng tiếp cận của bài toán này.



Hình 3. Bao phủ đối tượng

Cách tiếp cận tìm vị trí của cảm biến trước khi triển khai được gọi là bài toán tối ưu vị trí cảm biến và có thể ứng dụng với các mạng có kích thước không quá lớn. Mục tiêu là tối thiểu số lượng cảm biến (chi phí của mạng) trong khi đảm bảo mọi đối tượng (target) đều được bao phủ bởi ít nhất một cảm biến. Với việc định nghĩa khái niệm *site* là tập các vị trí của một cảm biến tại đó đối tượng phủ bởi cảm biến này không thay đổi, bài toán tìm vị trí trở thành bài toán tìm *site* để đặt cảm biến. Như vậy, thay vì phải tìm giá trị tọa độ - là những giá trị liên tục cho từng cảm biến, ta chỉ cần tìm *site* - là những tập rời rạc, hữu hạn và có thể giải quyết bằng cách vét cạn mọi khả năng có thể khi số lượng đối tượng và *site* không quá lớn. Tuy nhiên, với những trường hợp lớn hơn phương pháp này không hiệu quả. Một số phương pháp khác có thể sử dụng là tham lam, di truyền, mô phỏng luyện kim.

Trong cách tiếp cận triển khai ngẫu nhiên, các cảm biến sẽ được rải (scatter) xung quanh các đối tượng. Khi đó, một cảm biến có thể bao phủ nhiều đối tượng và ngược lại, một đối tượng có thể được phủ bởi nhiều cảm biến. Vấn đề đặt ra là cần phân hoạch các cảm biến thành từng tập thỏa mãn yêu cầu bao phủ (tức là mỗi đối tượng được bao phủ bởi ít nhất một cảm biến) để lần lượt kích hoạt các tập nhằm kéo dài thời

gian sống của mạng. Thời gian sống được tính từ thời điểm mạng bắt đầu hoạt động cho đến khi mạng không hoàn thành nhiệm vụ giám sát của nó (thường là do cảm biến hết năng lượng). Ở đây, thời gian sống sẽ bằng tổng thời gian hoạt động của các tập sau khi được phân hoạch. Ví dụ ở hình 3, mạng cảm biến có thể phân hoạch thành 2 tập $C_1 = \{s_1, s_3, s_6\}$, $C_2 = \{s_2, s_4, s_5\}$, sau đó tập C_1 sẽ được cho hoạt động trước đến khi hết năng lượng thì tập C_2 sẽ được kích hoạt. Nhờ vậy mà thời gian sống của mạng sẽ được kéo dài hơn. Bài toán này có thể quy về bài toán phủ tập lớn nhất và có thể giải quyết bằng thuật toán tham lam.

Bao phủ barrier

Lớp bài toán bao phủ barrier quan tâm đến nhiệm vụ phát hiện xâm nhập của mạng cảm biến không dây, một ứng dụng tiêu biểu trong lĩnh vực quân sự. Mục tiêu của nó là xác định liệu các đặc tính bao phủ mong muốn của một mạng cảm biến cho trước có tồn tại hay không? Theo [5], lớp bài toán này được chia thành 2 bài toán con, đó là:

- Bài toán xây dựng các rào chắn xâm nhập (intrusion barrier) để phát hiện sự xâm nhập của một đối tượng khi nó đi qua miền giám sát. Nói cách khác, bài toán này nhằm đảm bảo các điểm được bao phủ có khả năng hình thành nên một barrier kéo dài từ đầu này sang phía đối diện của miền giám sát và giao với mọi đường xâm nhập của đối tượng.
- Bài toán tìm đường xâm nhập (penetration path). Đường xâm nhập là một đường cong liên tục với hình dạng bất kỳ, đi từ đầu này sang phía đối diện của miền giám sát. Mục đích của bài toán này là để xác định một đường xâm nhập như vậy với các điểm trên đó có thỏa mãn các độ đo bao phủ được yêu cầu hay không?

Chi tiết về hai lớp bài toán con này sẽ được trình bày trong mục 1.2 dưới đây.

Bao phủ diện tích

Cho đến năm 2011, khi nhắc đến bài toán bao phủ diện tích trong mạng cảm biến không dây, người ta nghĩ ngay đến việc toàn bộ khu vực đặt cảm biến đều được giám sát. Có thể xem đây là trường hợp mở rộng của bài toán bao phủ đối tượng, trong đó tất cả các điểm thuộc khu vực cần giám sát đều được coi là các target.

Lúc này vấn đề đặt ra trong bài toán bao phủ diện tích là số lượng cảm biến cần dùng là bao nhiêu để yêu cầu bao phủ được thỏa mãn, tức là mọi vị trí trong khu vực cần giám sát đều được phát hiện bởi ít nhất một cảm biến. Nếu bài toán chỉ yêu cầu mỗi vị trí trong miền giám sát cần được phát hiện bởi 1 cảm biến thì ta gọi đó là bài toán 1-coverage, còn nếu yêu cầu phải có ít nhất k cảm biến cùng giám sát 1 vị trí thì đó là bài toán k -coverage.

Đến năm 2013, bài toán bao phủ diện tích được tiếp cận theo một cách khác. Thay vì xác định số lượng cảm biến tối thiểu để bao phủ toàn bộ một vùng diện tích cho trước, số cảm biến sẽ được xác định từ đầu và người ta quan tâm đến cách đặt (triển khai) chúng để độ bao phủ thu được trên một miền diện tích cụ thể là lớn nhất [26].

1.2. Bài toán bao phủ barrier trong mạng cảm biến không dây

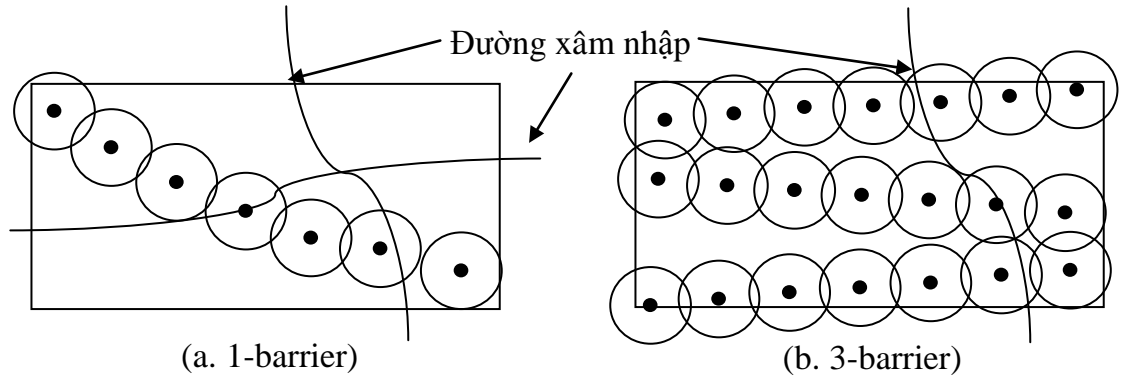
Như đã đề cập ở trên, nội dung chi tiết của hai lớp bài toán con của bao phủ barrier trong mạng cảm biến không dây sẽ được trình bày chi tiết dưới đây.

1.2.1. Xây dựng rào chắn xâm nhập

Phát hiện xâm nhập là một ứng dụng tiêu biểu của mạng cảm biến. Theo đó, mạng cảm biến sau khi được triển khai phải có khả năng phát hiện đối tượng di chuyển vào hoặc băng qua khu vực được giám sát. Một phương án dễ thấy cho bài toán này là đặt các cảm biến sao cho chúng bao quát toàn bộ khu vực cần bảo vệ. Tuy nhiên, việc phát hiện một đối tượng tại mọi bước di chuyển của nó là không cần thiết. Thay vào đó, chỉ cần có các hàng rào (barrier) sao cho đối tượng dù đi theo con đường nào băng qua khu

vực cần bảo vệ cũng đều bị phát hiện. Như vậy vừa giúp giảm số lượng cảm biến cần dùng, vừa đảm bảo nhiệm vụ của mạng.

Dựa vào số barrier cần xây dựng ít nhất k , nghĩa là có ít nhất k cảm biến phân biệt phát hiện đối tượng khi băng qua khu vực được giám sát theo bất kỳ con đường nào, các bài toán bao phủ barrier được gọi với tên riêng là k -barrier. Hình vẽ dưới đây là hai ví dụ minh họa cho trường hợp 1-barrier (hình 4(a)) và 3-barrier (hình 4(b)) với khu vực hình chữ nhật cần bảo vệ.



Hình 4. Bao phủ barrier

Trong hình 4(a), mọi đường băng từ biên trên qua biên dưới cũng như từ biên trái sang biên phải của khu vực cần giám sát đều được phát hiện bởi barrier. Cách triển khai barrier như vậy được gọi là barrier mạnh (strong barrier). Còn trong hình 4(b), 3-barrier chỉ đảm bảo khi đường băng qua đi từ biên trên xuống biên dưới của hình chữ nhật, và được gọi là barrier yếu (weak barrier).

Có hai vấn đề thường được đặt ra trong lớp bài toán này là: Thứ nhất, làm thế nào để xây dựng được bao phủ k -barrier? Thứ hai là sau khi triển khai các cảm biến một cách ngẫu nhiên, người ta muốn biết vị trí của các cảm biến có đảm bảo k -barrier hay không?

Câu hỏi đầu tiên đã được nghiên cứu bởi [4, 14, 15, 21]. Năm 2007, [14] đã nghiên cứu vấn đề này cho bao phủ barrier yếu trong đó chỉ xem xét trường hợp các đường

bằng qua vuông góc với biên của miền giám sát. Tuy nhiên, kết quả của họ không thể cung cấp cái nhìn thực tế bởi các đường xâm nhập thương là những đường cong, không vuông góc với biên của miền giám sát. Năm 2008, vấn đề này được nghiên cứu cho bao phủ barrier mạnh bởi [15], theo đó, một mạng cảm biến được cho là bao phủ k -barrier mạnh nếu xác suất bị bao phủ bởi k -barrier của bất kỳ đường băng qua nào bằng 1 với khả năng cao. Các tác giả này cũng đưa ra một lý thuyết rằng việc tồn tại bao phủ barrier mạnh trong miền giám sát hình chữ nhật phụ thuộc vào tỉ lệ giữa chiều dài và chiều rộng của miền đó. Cụ thể, nếu chiều rộng của miền giám sát nhỏ hơn không tiệm cận giá trị logarithm của chiều dài, thì mạng triển khai trên miền đó không tồn tại tính chất bao phủ barrier mạnh, ngược lại nếu chiều rộng lớn hơn không tiệm cận giá trị logarithm của chiều dài thì tồn tại khả năng bao phủ barrier mạnh. Các tiếp cận phân tích này được [21] mở rộng vào năm 2009, hướng tới một ngữ cảnh thực tế hơn khi mà tất cả các nút cảm biến được triển khai ngẫu nhiên trong miền giám sát hẹp.

Để trả lời cho câu hỏi thứ hai, [14] đề xuất sử dụng đồ thị bao phủ toàn cục (global coverage graph) để kiểm tra tính chất bao phủ k -barrier. Đồ thị bao phủ $CG = (V, E)$ của mạng cảm biến được xây dựng như sau. Tập đỉnh V tương ứng với tập các nút cảm biến của mạng. Ngoài ra, nó còn có hai đỉnh ảo, s và t , tương ứng với biên trái, phải của miền giám sát. Tồn tại cạnh giữa hai đỉnh của đồ thị nếu miền cảm biến của chúng chồng nhau, và tồn tại cạnh giữa đỉnh của đồ thị và đỉnh ảo nếu miền cảm biến của nó tương ứng chồng lên biên của miền giám sát. Sau đó, các tác giả đã xây dựng một lý thuyết: *Một mạng cảm biến N được triển khai trên một miền giám sát. Miền này được bao phủ k -barrier bởi N khi và chỉ khi tồn tại k đường đi phân biệt đỉnh giữa hai đỉnh ảo s và t trên đồ thị bao phủ CG .*

1.2.2. Tìm đường xâm nhập

Lớp bài toán tìm đường xâm nhập có nguồn gốc từ các ứng dụng truy vết và phát hiện xâm nhập. Đường xâm nhập là một đường cong liên tục với hình dáng bất kỳ đi xuyên qua miền giám sát từ biên này sang phía đối diện. Mục tiêu của bài toán là xác

định một đường xuyên qua với mọi điểm trên đó có độ đo bao phủ thỏa mãn một yêu cầu bao phủ đã được định nghĩa trước đây. Đây là sự khác biệt so với bài toán xây dựng rào chắn barrier mà chủ yếu đảm bảo rằng một vài điểm của đường xuyên qua cần đáp ứng yêu cầu bao phủ cụ thể.

Có hai vấn đề được đặt ra trong lớp bài toán con này đó là: Có những đường xâm nhập nào và cách để tìm ra chúng?

Đến nay, có 3 loại đường xâm nhập được nghiên cứu là:

- Đường có độ vi phạm tối đa (Maximal Breach Path)
- Đường có mức hỗ trợ tối đa (Maximal Support Path)
- Đường exposure (Exposure path)

Đường có độ vi phạm tối đa

Megerian [22] được xem là người đầu tiên nghiên cứu về bài toán tìm đường xâm nhập. Họ đã sử dụng khoảng cách Euclid giữa một điểm với nút cảm biến gần nó nhất như một độ đo về sự bao phủ tại điểm đó và đưa ra bài toán tìm đường có độ vi phạm tối đa như sau.

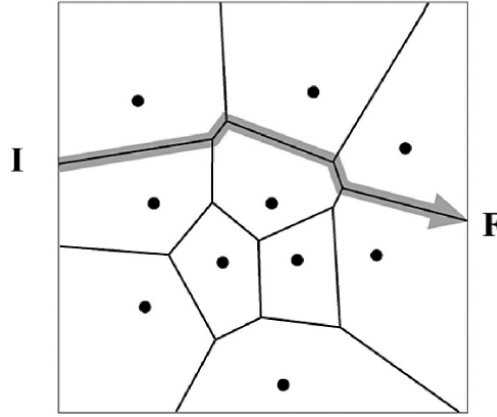
Ký hiệu S là tập các nút cảm biến, \wp là đường đi nối điểm đầu I với điểm cuối F cho trước. Khi đó mức độ vi phạm của đường \wp được định nghĩa là khoảng cách Euclid nhỏ nhất từ \wp tới bất kỳ cảm biến nào thuộc S . Nghĩa là, với mọi điểm trên \wp , ta sẽ tính khoảng cách Euclid giữa nó và cảm biến gần nhất của nó rồi lấy giá trị nhỏ nhất trong các khoảng cách này làm độ vi phạm của đường \wp .

Khi đó, bài toán tìm đường có độ vi phạm tối đa yêu cầu rằng: Trong số tất cả các đường đi nối I với F , hãy tìm đường có độ vi phạm lớn nhất và gọi đường đó là đường vi phạm tối đa, ký hiệu là P_B .

Từ quan điểm của kẻ xâm nhập, đường P_B là đường an toàn nhất bởi khoảng cách tới cảm biến gần nhất dọc theo đường này là lớn nhất. Nhưng với những người bảo vệ,

đây lại là đường tối nhất bởi cơ hội phát hiện kẻ xâm nhập bị tối thiểu, tương đương với trường hợp bao phủ tối nhất.

Có vô số các đường đi nối I với F , do đó việc tìm kiếm chúng bằng phương pháp vét cạn là không khả thi. [22] đã áp dụng biểu đồ Voronoi để có cái nhìn trực quan về miền giám sát và lập luận rằng có ít nhất một đường có độ vi phạm tối đa đi dọc theo các cạnh của biểu đồ Voronoi. Sau đó, họ thực hiện tìm kiếm nhị phân và tìm kiếm theo chiều rộng để tìm P_B trên đồ thị xây dựng từ biểu đồ Voronoi (hình 5).



Hình 5. Biểu đồ Voronoi và đường vi phạm tối đa [22]

Sau này, có thêm nhiều nghiên cứu mở rộng bài toán này, theo đó, độ đo vi phạm được tính là tổng khoảng cách từ một điểm tới k cảm biến gần nó nhất. Mô hình này tương ứng với ngữ cảnh một kẻ xâm nhập bị phát hiện chỉ khi nó ở gần k cảm biến tại thời điểm đó.

Đường có mức hỗ trợ tối đa

Cũng trong [22], các tác giả đã đề xuất vấn đề tìm đường có mức hỗ trợ tối đa sử dụng khoảng cách Euclid giữa một điểm và cảm biến gần nó nhất làm độ đo giá trị bao phủ tại điểm đó.

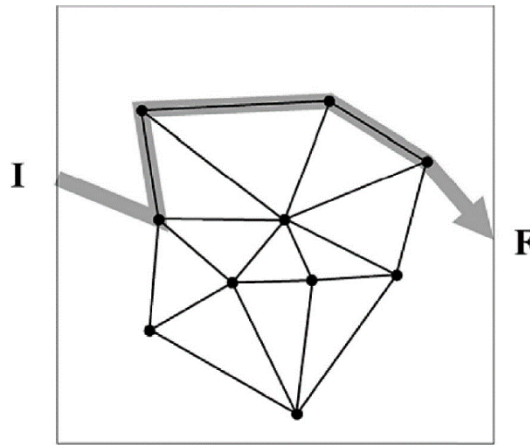
Xét tập cảm biến S , điểm đầu I và điểm cuối F . Mức hỗ trợ của một đường φ nối I và F là khoảng cách Euclid lớn nhất từ φ đến cảm biến gần nhất thuộc tập S . Nghĩa là,

với mỗi điểm thuộc đường φ , ta tìm khoảng cách Euclid giữa nó và cảm biến gần nhất, rồi lấy giá trị lớn nhất trong những khoảng cách này và gọi là mức hỗ trợ.

Khi đó, bài toán tìm đường có mức hỗ trợ lớn nhất phát biểu rằng: Trong tất cả các đường nối I với F , tìm đường có giá trị hỗ trợ nhỏ nhất và gọi là đường có mức hỗ trợ tối đa, ký hiệu P_S .

Từ góc độ của kẻ xâm nhập, đây là đường tồi nhất để đi qua vì khả năng bị phát hiện là lớn nhất. Tuy nhiên, với người bảo vệ, đây là đường cho chất lượng giám sát cao nhất, ứng với trường hợp bao phủ tốt nhất.

Để giải quyết bài toán này, [22] đã áp dụng phương pháp tam giác phân Delaunay và lập luận có ít nhất một đường có mức hỗ trợ tối đa đi dọc theo các cạnh của biểu đồ Delaunay (hình 6).



Hình 6. Biểu đồ Delaunay và đường có mức hỗ trợ tối đa [22]

Đường Exposure

Đường exposure là một dạng khác của đường xâm nhập, được định nghĩa để đánh giá miền giám sát được bao phủ tốt như thế nào theo khía cạnh khả năng phát hiện một đối tượng di chuyển. Giá trị exposure càng cao, độ bao phủ của mạng càng tốt.

Tương tự với đường có độ vi phạm tối đa và đường có mức hỗ trợ tối đa, khoảng cách cũng được sử dụng làm độ đo để định nghĩa giá trị exposure tại một điểm. Tuy

nhien, trong khi hai đường này sử dụng trực tiếp giá trị cực trị của khoảng cách của tất cả các điểm thuộc đường đi để làm thước đo, thì đường exposure thông qua giá trị trung bình của chúng để định nghĩa độ đo.

Năm 2002, [16] sử dụng hàm cảm biến mô hình tỉ lệ $f(3)$ đã trình bày ở trên và đưa ra hai cách để đo cường độ cảm biến I mà miền cảm biến S tác dụng lên một điểm p , đó là:

- Cường độ miền cảm biến gần nhất (Closest sensor field intensity)
- Cường độ của toàn miền cảm biến (All sensor field intensity)

Trong đó, cường độ miền cảm biến gần nhất, ký hiệu I_C , được định nghĩa bởi (8).

$$I_C(S, p) = f(d(s_{\min}, p)) = \frac{c}{d^\lambda(s_{\min}, p)} \quad (9)$$

với s_{\min} là cảm biến gần điểm p nhất. Còn cường độ của toàn miền cảm biến, ký hiệu I_A , được tính bởi công thức (9).

$$I_A(S, p) = \sum_{i=1}^n f(d(s_i, p)) \quad (10)$$

với n là số cảm biến thuộc tập S . Theo cách tính này, mọi cảm biến s_i của tập S đều đóng góp một lượng cảm biến nhất định tại điểm p phụ thuộc vào khoảng cách đến điểm đó.

Từ đây, các tác giả giả sử một kẻ xâm nhập di chuyển trong miền giám sát được bao phủ bởi tập cảm biến S (gọi tắt là miền cảm biến S) từ điểm $p(t_1)$ tới điểm $p(t_2)$ dọc theo đường $p(t)$ và đưa ra định nghĩa giá trị exposure E của đường này như sau.

$$E(p(t), t_1, t_2) = \int_{t_1}^{t_2} I(S, p(t)) \left| \frac{dp(t)}{dt} \right| dt \quad (11)$$

Trong đó, cường độ cảm biến $I(S, p(t))$ có thể là $I_C(S, p(t))$ hoặc $I_A(S, p(t))$ và $|dp(t)/dt|$ là độ dài cung di chuyển. Ví dụ, nếu $p(t)=(x(t), y(t))$ thì

$$\left| \frac{dp(t)}{dt} \right| = \sqrt{\left(\frac{dx(t)}{dt} \right)^2 + \left(\frac{dy(t)}{dt} \right)^2}.$$

Lúc này, bài toán tìm đường exposure nhỏ nhất (Minimal Exposure Problem - MEP) nhằm tìm một đường có giá trị exposure nhỏ nhất, ký hiệu là P_E , trong số tất cả các đường nối điểm đầu I với điểm cuối F . Với kẻ xâm nhập, đường exposure nhỏ nhất là đường xâm nhập có khả năng bị phát hiện thấp nhất. Nhưng với người bảo vệ, đây lại là trường hợp bao phủ tồi nhất, nghĩa là họ cần thực hiện một số biện pháp để làm tăng giá trị exposure như thêm nút cảm biến.

Theo [5], việc tìm P_E trong mạng cảm biến không dây được triển khai ngẫu nhiên là một bài toán tối ưu vô cùng khó. Đến nay, có hai phương pháp xấp xỉ thường được sử dụng để giải bài toán này, đó là: phương pháp dựa trên lưới (grid-based method) và phương pháp hình học tính toán dựa trên biểu đồ Voronoi [17].

[18] là các tác giả đầu tiên đề xuất sử dụng lưới để chuyển bài toán từ miền giá trị liên tục sang miền giá trị rời rạc. Trọng số của một đoạn thẳng chính là giá trị exposure trên đoạn đó với giả sử tốc độ di chuyển của kẻ xâm nhập là hằng số. Khi đó, việc tìm P_E được giới hạn chỉ trên các đoạn thẳng của lưới và chính là tìm đường đi ngắn nhất nối I với F trên đồ thị lưới này. Nhược điểm của phương pháp dựa trên lưới này là kích thước lưới bị cố định và đối tượng chỉ có thể di chuyển trên các điểm lưới và đặc biệt kích thước của lưới có ảnh hưởng lớn đến độ chính xác của lời giải và thời gian tính toán. Nếu số lượng các ô lưới quá nhỏ thì lời giải thu được cho bài toán MEP có độ chính xác thấp, và để có chất lượng lời giải cao hơn thì số lượng ô lưới phải lớn, đồng nghĩa với thời gian tính toán sẽ lâu hơn, tăng theo hàm mũ [17]. Do đó, phương pháp này không phù hợp để giải MEP với yêu cầu độ chính xác cao và mạng cảm biến có kích thước lớn (large-scale).

Trong phương pháp sử dụng biểu đồ Voronoi [21], miền cảm biến sẽ được phân hoạch thành m đa giác lồi được gọi là các ô Voronoi với m là số lượng cảm biến sao cho mỗi điểm trong ô đó gần với cảm biến của nó hơn tất cả các cảm biến thuộc ô khác. Đỉnh của các đa giác lồi này được gọi là đỉnh Voronoi, còn các cạnh của chúng được gọi là các cạnh Voronoi. Các đỉnh và cạnh Voronoi tạo thành một đồ thị, gọi là đồ thị Voronoi. Trọng số trên mỗi cạnh Voronoi chính là giá trị exposure khi đối tượng di chuyển trên cạnh đó. Dựa trên lý thuyết của biểu đồ Voronoi, [21] chứng minh rằng khi cường độ cảm biến chỉ phụ thuộc vào một cảm biến (mô hình cường độ cảm biến gần nhất đã trình bày ở trên) thì đường MEP phải nằm trên cạnh Voronoi. Sau đó, các tác giả sử dụng thuật toán tìm đường đi ngắn nhất như Dijkstra để thu được lời giải cho bài toán MEP. Tuy nhiên, khi cường độ miền cảm biến phụ thuộc vào tất cả các cảm biến thì phương pháp này sẽ không còn chính xác. Ngoài ra, khi điểm đầu và điểm kết thúc của đường xâm nhập không nằm trên cạnh Voronoi thì phương pháp này không thể được áp dụng. Cuối cùng, khi khả năng bao phủ của các cảm biến là khác nhau thì MEP sẽ không đi qua các cạnh Voronoi và do đó, phương pháp dựa trên biểu đồ Voronoi không thể tìm được lời giải cho bài toán MEP.

Trong thời gian gần đây, khi các nhà nghiên cứu quan tâm đến việc xác định khả năng phát hiện đối tượng xâm nhập của mạng cảm biến không dây, thì giá trị exposure còn được định nghĩa như xác suất phát hiện đối tượng khi nó di chuyển trên một đường xuyên qua miền giám sát [10, 23]. Tuy nhiên, [10] chỉ dừng lại ở việc đưa ra cách tính giá trị exposure của một đường xâm nhập cho trước trong mạng cảm biến không dây có nhiều trạng thái. Năm 2015, [17] giải quyết bài toán MEP cho cả 3 mô hình cảm biến, nhưng mô hình xác suất họ sử dụng là mô hình xác suất không tính đến sự ảnh hưởng của nhiễu được định nghĩa trong công thức (4) trên đây. Trong khi đó, mặc dù sử dụng mô hình xác suất (4), nhưng [23] lại quan tâm đến bài toán tìm cách triển khai các cảm biến để thỏa mãn hiệu năng về việc giám sát, không phải bài toán tìm đường xâm nhập.

Với mong muốn mở rộng nghiên cứu về bài toán MEP trong mạng cảm biến không dây sử dụng mô hình xác suất, luận văn sẽ đưa ra cách tính giá trị exposure cho mô hình xác suất có xét đến sự ảnh hưởng của nhiễu và phát biểu bài toán MEP với mô hình xác suất này trong mục 1.3 dưới đây, gọi là bài toán tìm đường xâm nhập có khả năng bị phát hiện nhỏ nhất trong mạng cảm biến không dây. Sau đó, một thuật toán xấp xỉ sẽ được đề xuất để giải quyết mô hình bài toán này.

1.3. Bài toán tìm đường xâm nhập có khả năng bị phát hiện nhỏ nhất trong mạng cảm biến không dây

1.3.1. Xây dựng độ đo exposure

Công thức (6) đưa ra khả năng phát hiện đối tượng tại vị trí l của một cảm biến s . Khi đó, xác suất không phát hiện được đối tượng tại l được tính như sau:

$$\overline{P_s^l} = 1 - P_s^l = 1 - Q\left(\frac{A - \frac{C}{d^\lambda(s,l)}}{\sigma}\right) \quad (12)$$

Để tính xác suất không phát hiện đối tượng tại l của N cảm biến, [1] đưa ra công thức dưới đây.

$$\overline{P^l} = \prod_{i=1}^N \overline{P_{s_i}^l} = \prod_{i=1}^N (1 - P_{s_i}^l) = \prod_{i=1}^N \left(1 - Q\left(\frac{A - \frac{C}{d^\lambda(s_i,l)}}{\sigma_i}\right)\right) \quad (13)$$

Kí hiệu \wp là một đường xâm nhập từ biên trái sang biên phải của miền giám sát \mathfrak{R} , L_\wp là tập hữu hạn điểm của \wp , thì xác suất không phát hiện đối tượng đi theo đường \wp được cho bởi công thức (14) với $|L_\wp|$ là lực lượng của L_\wp .

$$\overline{P^\wp} = \prod_{j=1}^{|L_\wp|} \overline{P^{l_j}} \quad (14)$$

Kết hợp công thức (13), (14), ta thu được công thức sau đây.

$$\overline{P^{\wp}} = \prod_{j=1}^{|L_{\wp}|} \overline{P^{l_j}} = \prod_{j=1}^{|L_{\wp}|} \prod_{i=1}^N \left(1 - Q \left(\frac{A - \frac{C}{d^{\lambda}(s_i, l_j)}}{\sigma_i} \right) \right) \quad (15)$$

Bởi vì giá trị của biểu thức $(1 - Q(.))$ nhỏ hơn 1, nên việc tính trực tiếp (15) không đơn giản. Thay vào đó, (15) có thể được tính thông qua cách lấy logarithm cả hai vế như sau.

$$\ln(\overline{P^{\wp}}) = \ln \left(\prod_{j=1}^{|L_{\wp}|} \prod_{i=1}^N \left(1 - Q \left(\frac{A - \frac{C}{d^{\lambda}(s_i, l_j)}}{\sigma_i} \right) \right) \right) \quad (16)$$

hay

$$\ln(\overline{P^{\wp}}) = \sum_{j=1}^{|L_{\wp}|} \sum_{i=1}^N \ln \left(1 - Q \left(\frac{A - \frac{C}{d^{\lambda}(s_i, l_j)}}{\sigma_i} \right) \right) \quad (17)$$

Từ đó, xác suất phát hiện của N cảm biến trên cả đường đi \wp được tính như sau:

$$P^{\wp} = 1 - \overline{P^{\wp}} = 1 - e^{\ln(\overline{P^{\wp}})} \quad (18)$$

Rõ ràng, giá trị của $\ln(\overline{P^{\wp}})$ càng lớn thì giá trị của P^{\wp} càng nhỏ. Bên cạnh đó, hàm logarithm tự nhiên là một hàm đồng biến tăng và $\ln(\overline{P^{\wp}})$ là một số âm với $\overline{P^{\wp}}$ nằm trong đoạn $[0, 1]$.

Dựa vào những diễn giải trên, giá trị exposure của đường xâm nhập \wp tính được bởi N cảm biến dưới mô hình bao phủ xác suất (6), ký hiệu là $E(\wp)$ được định nghĩa như sau:

$$E(\wp) = -\ln(\overline{P^\wp}) = -\sum_{j=1}^{|L_\wp|} \sum_{i=1}^N \ln \left(1 - Q \left(\frac{A - \frac{C}{d^\lambda(s_i, l_j)}}{\sigma_i} \right) \right) \quad (19)$$

Như vậy, giá trị exposure của đường xâm nhập \wp được trực tiếp định nghĩa bởi xác suất phát hiện đối tượng khi đi qua miền giám sát. Khi đó, bài toán tìm MEP trong mạng cảm biến không dây sử dụng mô hình xác suất tính đến sự ảnh hưởng của nhiễu nhằm tìm một đường xâm nhập có giá trị $E(\wp)$ cho bởi công thức (19) là nhỏ nhất. Bài toán này sẽ được phát biểu chi tiết dưới đây.

1.3.2. *Phát biểu bài toán*

Bài toán tìm đường xâm nhập có khả năng bị phát hiện nhỏ nhất được xây dựng cho mạng cảm biến không dây sử dụng mô hình cảm biến xác suất xét đến sự ảnh hưởng của nhiễu, gọi là bài toán PM-based-MEP có nội dung như sau: Cho một tập các cảm biến được triển khai ngẫu nhiên trong vùng giám sát \mathfrak{R} và hai điểm I, F nằm trên hai cạnh đối diện của miền \mathfrak{R} , lần lượt gọi là điểm xuất phát và điểm kết thúc. Bài toán nhằm tìm ra một đường xâm nhập \wp xuất phát từ I đi đến F sao cho nếu đối tượng xâm nhập đi theo con đường này xuyên qua \mathfrak{R} sẽ có khả năng bị phát hiện là nhỏ nhất.

Mô hình bài toán được phát biểu cụ thể dưới đây:

Đầu vào:

- W, H : chiều rộng và chiều dài của miền cần giám sát
- N : số lượng các cảm biến

- $S = \{s_i = (x_{s_i}, y_{s_i}, r_i) \mid i = 1..N\}$: tập các cảm biến với tọa độ (x_{s_i}, y_{s_i}) và bán kính cảm biến r_i tương ứng.
- $(0, y_I)$: tọa độ điểm xuất phát của đối tượng xâm nhập
- (W, y_F) : tọa độ điểm kết thúc của đối tượng
- σ_i : độ lệch chuẩn của nhiễu tại cảm biến $s_i (i = 1..N)$
- A : ngưỡng phát hiện
- C : hằng số
- λ : hệ số suy giảm tín hiệu

Đầu ra:

- Tập điểm có thứ tự hình thành nên đường đi \wp kết nối I với F .

Mục tiêu:

- Tối thiểu hóa giá trị exposure của đường đi \wp , tức là:

$$E(\wp) = -\ln(\overline{P^\wp}) = -\sum_{j=1}^{|L_\wp|} \sum_{i=1}^N \ln \left(1 - Q \left(\frac{A - \frac{C}{d^\lambda(s_i, l_j)}}{\sigma_i} \right) \right) \rightarrow \min$$

Ràng buộc:

- Đối tượng luôn luôn di chuyển trong vùng giám sát từ điểm xuất phát I đến điểm kết thúc F với vận tốc không đổi (*)
- Đối tượng không được phép quay đầu lại (**)

Ràng buộc đầu tiên để đảm bảo rằng khoảng cách giữa hai điểm liên tiếp trong quá trình di chuyển của đối tượng luôn bằng nhau. Điều này sẽ cho phép đánh giá độ bao phủ trung bình của mạng cảm biến không dây. Để đơn giản và không làm mất tính tổng

Đề tài: Giải thuật xấp xỉ giải bài toán bao phủ barrier trong mạng cảm biến không dây

quát, khoảng cách này được thiết lập bằng 1 đơn vị độ dài, nghĩa là sau mỗi một đơn vị thời gian, đối tượng đi được một quãng đường có độ dài là 1 đơn vị.

Ràng buộc thứ hai phù hợp với thực tế rằng kẻ xâm nhập luôn có tham vọng đi qua vùng giám sát càng nhanh càng tốt.

CHƯƠNG 2: CÁC THUẬT TOÁN XẤP XỈ GIẢI BÀI TOÁN TỐI ƯU

Chương này sẽ trình bày các vấn đề sau:

- Tổng quan về lớp bài toán tối ưu
- Tổng quan các phương pháp xấp xỉ giải bài tối ưu
- Tổng quan về mô hình giải thuật di truyền

2.1. Bài toán tối ưu

Trong thực tế có nhiều tình huống với nhiều phương án giải quyết khác nhau khiến chúng ta phải lựa chọn để tìm ra phương án tốt nhất. Quan điểm “tốt nhất” này phụ thuộc vào từng tình huống và từng mục đích, hoặc tối đa (maximize) hoặc tối thiểu (minimize) một tiêu chí nào đó. Ví dụ như khi có nhiều cách để đi từ nhà đến nơi làm việc, có người chọn cách đi ngắn nhất, có người lại chọn cách đi ít tắc đường nhất hay gặp ít đèn tín hiệu giao thông nhất.

Việc tìm phương án tốt nhất trong số những phương án có thể như trên chính là việc giải quyết bài toán tối ưu (optimization problem) trong lĩnh vực toán và công nghệ thông tin. Nhìn chung, các bài toán tối ưu luôn có hai đặc điểm sau:

- Số lời giải nhiều hơn một. Tập các lời giải này được gọi là không gian lời giải hay không gian tìm kiếm.
- Mọi lời giải của một bài toán đều được định lượng độ “tốt” bởi cùng một hàm tính theo tiêu chí của bài toán đó.

Và để giải quyết những bài toán này (cũng như những bài toán khác), trước hết cần phải phát biểu chúng theo mô hình toán học (problem formulation). Trong đó liệt kê dữ liệu đầu vào, yêu cầu đầu ra, mục tiêu cụ thể bằng công thức toán và các ràng buộc (nếu có) một cách rõ ràng và logic. Từ đó, lời giải sẽ được biểu diễn thành một hay nhiều biến quyết định (decision variable) sao cho ứng với mỗi giá trị khác nhau sẽ thu được lời giải không giống nhau. Dựa vào miền giá trị (rời rạc hay liên tục) của các biến này, bài toán tối ưu được chia thành hai loại:

- Bài toán tối ưu rời rạc, còn gọi là bài toán tối ưu tổ hợp (combinatorial optimization problem).
- Bài toán tối ưu liên tục (continuous optimization problem)

2.1.1. **Bài toán tối ưu tổ hợp**

Bài toán tối ưu được gọi là tối ưu tổ hợp khi các biến quyết định nhận giá trị trong một tập rời rạc, được giới hạn bởi một số ràng buộc. Theo [2], bất kỳ một bài toán tối ưu tổ hợp nào cũng có thể được định nghĩa hình thức bởi một bộ 4:

$$I = (U, P, f, extr)$$

Trong đó,

- U : không gian lời giải
- P : các ràng buộc
- f : hàm mục tiêu, là một ánh xạ $U \rightarrow$ tập số thực \mathbf{R}
- $extr$: cực trị, thường là cực đại hoặc cực tiểu

Ở đây, không gian lời giải U là hữu hạn, rời rạc. Các lời giải trong không gian U nếu thỏa mãn các ràng buộc P được gọi là lời giải khả thi hay lời giải chấp nhận được.

Một số bài toán tiêu biểu thuộc lớp bài toán tối ưu tổ hợp này là:

- Bài toán người du lịch (Traveling Salesman Problem)
- Bài toán cái túi (Knapsack Problem)
- Bài toán phân công (Assignment Problem)
- Cây khung nhỏ nhất (Minimum Spanning Tree)
- ...

2.1.2. **Bài toán tối ưu liên tục**

Khi các biến biểu diễn lời giải nhận giá trị liên tục trong không gian số thực thì bài toán tối ưu trở thành tối ưu liên tục. Mô hình chuẩn của bài toán tối ưu này như sau:

$$\underset{x}{\text{minimize / maximize}} \quad f(x)$$

sao cho :

$$g_i(x) \leq 0, \quad i=1, \dots, m$$

$$h_i(x) = 0, \quad i=1, \dots, p$$

Trong đó:

- $f(x): R^n \rightarrow R$ là hàm mục tiêu cần được tối thiểu (minimize) hoặc tối đa (maximize) với giá trị của biến x
- $g_i(x) \leq 0$: là các ràng buộc bất đẳng thức
- $h_i(x) = 0$: là các ràng buộc đẳng thức
- $m, p \in N$

Trường hợp $m = p = 0$ thì ta gọi đó là bài toán tối ưu (liên tục) không ràng buộc.

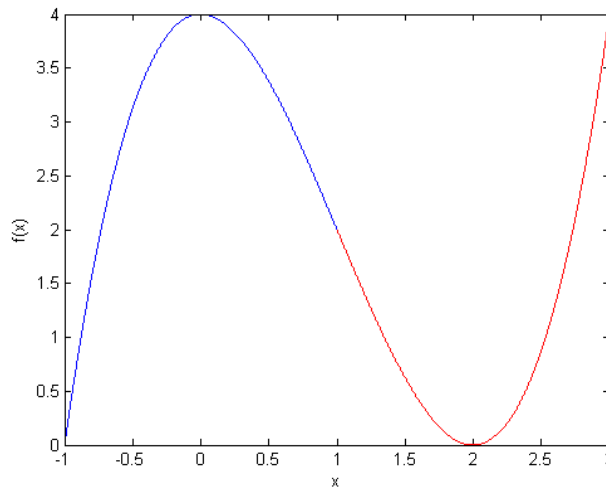
Tìm cực trị (cực đại hoặc cực tiểu) của hàm số là một ví dụ điển hình cho lớp bài toán này. Chẳng hạn:

$$\underset{x}{\text{minimize}} \quad f(x) = x^3 - 3x^2 + 4$$

$$\text{sao cho : } 1 \leq x \leq 3, \quad x \in R$$

Đầu vào của bài toán chính là đoạn số thực $[1, 3]$ còn đầu ra là một giá trị thuộc đoạn đó sao cho hàm mục tiêu $f(x)$ có giá trị nhỏ nhất.

Rõ ràng, có rất nhiều giá trị x nằm trong đoạn $[1, 3]$ nhưng chỉ có duy nhất một giá trị $x = 2$ (hình 7) làm hàm $f(x)$ tối thiểu. Khi đó ta nói $x = 2$ là lời giải tối ưu còn $f^* = f(x^*) = 0$ là giá trị tối ưu của bài toán.



Hình 7. Minh họa bài toán tối ưu liên tục: đồ thị hàm số $f(x) = x^3 - 3x^2 + 4$

Sau khi bài toán đã được mô hình hóa, vấn đề tiếp theo được đặt ra là giải quyết chúng như thế nào?

2.2. Tổng quan các phương pháp xấp xỉ giải bài toán tối ưu

Phương pháp giải xấp xỉ được đưa ra để khắc phục hạn chế của giải chính xác khi có sự bùng nổ tổ hợp, nghĩa là không gian lời giải trở nên quá lớn khi kích thước dữ liệu đầu vào tăng lên. Mục đích của phương pháp này không phải tìm cho ra lời giải tối ưu mà là tìm lời giải gần tối ưu nhất trong một khoảng thời gian chấp nhận được.

Đến nay đã có nhiều thuật toán theo hướng giải xấp xỉ được đề xuất và có thể áp dụng cho các bài toán khác nhau. Tiêu biểu là thuật toán tham lam, mô phỏng luyện kim, tìm kiếm địa phương và các thuật toán tiến hóa.

Thuật toán tham lam (Greedy Algorithm)

Thuật toán tham lam được xem là phương pháp giải xấp xỉ đơn giản nhất. Ý tưởng của thuật toán này là luôn lựa chọn giá trị tốt nhất (tối ưu địa phương) tại mỗi bước trong quá trình xây dựng lời giải với hy vọng rằng tìm được lời giải tối ưu toàn cục.

Ví dụ với bài toán người du lịch, tại mỗi bước thuật toán tham lam sẽ chọn đến thăm thành phố (chưa được thăm trước đó) gần nhất (chi phí ít nhất) từ thành phố hiện tại.

Nhìn chung, giải thuật tham lam gồm có 5 thành phần sau:

- Tập hợp các ứng cử viên để tạo ra lời giải
- Hàm đánh giá để lựa chọn ứng cử viên tốt nhất, bổ sung vào lời giải đang có.
- Hàm khả thi (feasibility function) quyết định một ứng cử viên có thể được dùng để xây dựng lời giải hay không?
- Hàm giá trị (mục tiêu) để tính giá trị của lời giải và/hoặc lời giải bộ phận (lời giải chưa hoàn chỉnh)
- Điều kiện dừng chỉ ra thời điểm một lời giải là hoàn chỉnh.

Vấn đề của thuật toán này nằm ở chỗ tìm cách để lựa chọn phương án tốt nhất (chính là xây dựng hàm đánh giá) trong mỗi bước xây dựng lời giải. Thông thường, tương tự như thuật toán nhánh cận, tham lam sẽ sử dụng tiêu chí tối ưu của bài toán để tiến hành lựa chọn. Ta có thể lựa chọn phương án được cho là tốt nhất ở thời điểm hiện tại và sau đó giải bài toán con nảy sinh từ lựa chọn vừa rồi. Nghĩa là, lựa chọn trong thuật toán tham lam chỉ phụ thuộc vào các lựa chọn trước đó. Như vậy, thuật toán sẽ tiến triển bằng cách thực hiện các lựa chọn theo một vòng lặp, cùng lúc đó, thu nhỏ bài toán ban đầu về bài toán con nhỏ hơn. Đây là sự khác biệt giữa thuật toán này là quy hoạch động. Trong khi giải thuật quy hoạch động duyệt hết và luôn đảm bảo tìm thấy lời giải chính xác (nếu có) của bài toán gốc, thì thuật toán tham lam không có sự đảm bảo này do thuật toán này không bao giờ xét lại các quyết định cũ, dẫn đến không quét hết tất cả các lời giải khả thi.

Mô phỏng luyện kim (Simulated Annealing - SA)

Mô phỏng luyện kim là một thuật toán tìm kiếm dựa trên xác suất được đề xuất bởi Kirkpatrick, Gelatt, Vecchi và Cerny năm 1982 [8] nhằm tìm ra giá trị tối ưu toàn cục của hàm mục tiêu. Phương pháp này mô phỏng quá trình luyện kim trong lĩnh vực cơ học. Theo đó, khi tiến hành luyện, kim loại được đốt nóng tới nhiệt độ cao T rồi làm lạnh từ từ để nó kết tinh ở cấu hình năng lượng thấp. Mục đích là để tăng kích thước của tinh thể và làm giảm những khuyết điểm của kim loại đó (tối ưu toàn cục). Nếu việc làm lạnh không xảy ra từ từ thì kim loại sẽ bị đông lạnh ở một trạng thái không ổn định (tối ưu hóa địa phương) [26].

Thuật toán luyện kim sử dụng T với vai trò là biến điều khiển toàn cục. Ban đầu, T được gán một giá trị rất lớn và sau đó được giảm dần xuống. Trong quá trình tìm kiếm, SA chọn ngẫu nhiên lời giải láng giềng của lời giải hiện tại để thay thế với một xác suất phụ thuộc vào sự chênh lệch giữa giá trị hàm mục tiêu và T . Công việc này được gọi là quá trình tối ưu. Quá trình này sẽ được tiếp diễn cho đến khi cực tiểu toàn cục được tìm thấy hoặc đến khi tổng số bước thay thế vượt quá một giá trị tối đa đã định nghĩa từ trước. Giá trị của T được giảm sau khi có sự cân bằng nhiệt xảy ra.

Thuật toán luyện kim có thể vượt qua các giá trị tối ưu địa phương nhờ sử dụng tham số T . Tuy nhiên, giá trị T ban đầu nên là bao nhiêu, giảm xuống như thế nào lại tùy thuộc vào từng bài toán cụ thể và thường phải tiến hành thực nghiệm để lựa chọn ra các thông số phù hợp.

Tìm kiếm địa phương (Local search)

Tương tự như phương pháp mô phỏng luyện kim, tìm kiếm địa phương cũng khám phá không gian lời giải bằng cách tại mỗi bước cố gắng chuyển đến các lời giải láng giềng của lời giải hiện tại. Tuy nhiên, khác với SA, phương pháp tìm kiếm địa phương chỉ chuyển đến lời giải láng giềng khi lời giải đó có giá trị hàm mục tiêu tốt hơn.

Cụ thể, các bước của phương pháp này như sau:

- Xuất phát từ một phương án nào đó
- Áp dụng một phép biến đổi lên phương án hiện hành để được một phương án mới *tốt hơn* phương án đã có
- Lặp lại việc áp dụng phép biến đổi lên phương án hiện hành cho đến khi không còn có thể cải thiện được phương án nữa.

Thông thường một phép biến đổi chỉ thay đổi một bộ phận nào đó của phương án hiện hành để được một phương án mới nên phép biến đổi được gọi là phép biến đổi địa phương và do đó phương pháp này có tên là tìm kiếm địa phương [31]. Vấn đề của phương pháp này chính là cách tìm phép biến đổi lời giải hiện tại sang lời giải mới có giá trị hàm mục tiêu tốt hơn.

Tìm kiếm Tabu (Tabu Search)

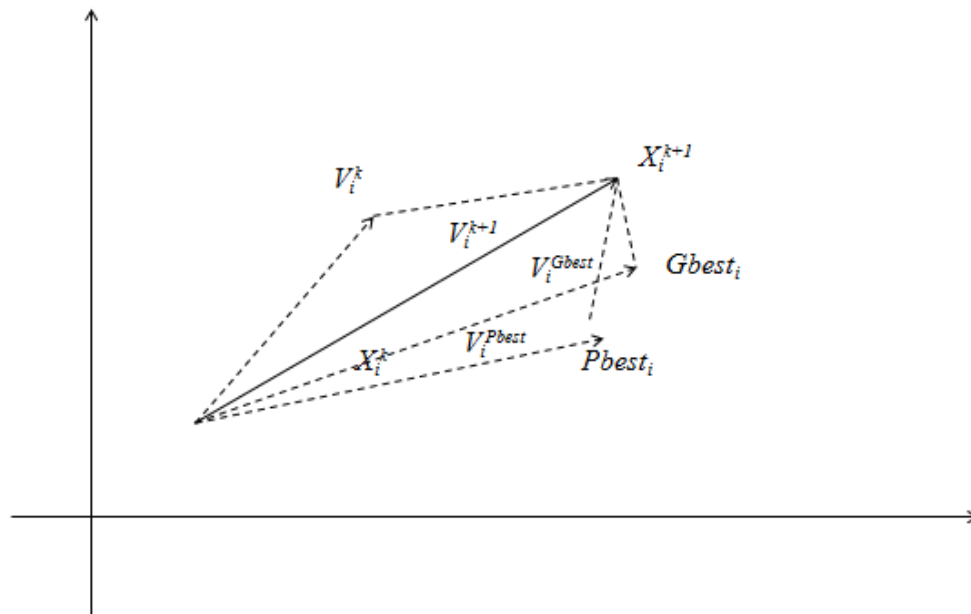
Năm 1986, Glover đã tạo ra thuật toán Tabu search và được công bố vào năm 1989 [11]. Tabu search thừa kế ý tưởng từ thuật giải tìm kiếm địa phương và cải tiến hiệu suất của thuật toán. Ý tưởng chính của sự cải tiến này là tại mỗi bước di chuyển xấu, lời giải có thể được chấp nhận nếu không có lời giải nào tốt hơn và một phương án đã được khảo sát thì phương án đó sẽ được đánh dấu là “tabu” (cấm) để thuật toán không xem xét lời giải đó nhiều lần. Theo [24], kỹ thuật sử dụng bộ nhớ ngắn hạn này giúp bài toán tránh phải các tối ưu địa phương cũng như vòng lặp nhiều lần cùng vị trí. Bộ nhớ ngắn hạn được biểu diễn như là danh sách tabu, danh sách này theo dõi các phương án gần đây nhất và cấm quay lại các phương án này. Nếu sau một số vòng lặp liên tiếp mà lời giải không được cải thiện thì danh sách tabu sẽ được giải phóng.

Thuật toán tối ưu hóa bầy đàn (Particle Swarm Optimization - PSO)

Giải thuật PSO [13] là một dạng của giải thuật tiến hóa quần thể cũng như GA, tuy nhiên PSO thiên về sử dụng các tương tác giữa các cá thể trong quần thể để khảo sát không gian tìm kiếm dựa trên quá trình mô hình hóa đàn chim bay tìm kiếm thức ăn. Tại thời điểm đầu, mỗi cá thể sẽ ngẫu nhiên tìm thức ăn ở một số nơi ngẫu nhiên. Sau một thời gian, sẽ có các cá thể tìm được thức ăn và báo thông tin lại cho các cá thể

khác. Tùy thuộc vào số lượng thức ăn tìm được mà các cá thể sẽ bay về hướng nơi có thức ăn nhiều nhất. Tương tự như vậy, PSO bắt đầu với việc khởi tạo quần thể. Một cá thể thường lưu giữ ba thuộc tính là vị trí hiện tại của cá thể đó, vận tốc di chuyển và vị trí tốt nhất của cá thể đó tính tới thời điểm hiện tại. Trong mỗi thế hệ, mỗi cá thể được cập nhật theo ba hướng vận tốc. Hướng thứ nhất dựa trên giá trị tốt nhất tính tới hiện tại tại V_i^{Pbest} , hướng thứ hai dựa trên giá trị tốt nhất của toàn quần thể hiện tại tại V_i^{Gbest} (hình 8) và hướng thứ ba dựa trên vận tốc hiện tại của cá thể V_i^k .

Quá trình tìm kiếm dừng lại khi tìm được phương án tối ưu hoặc đạt tiêu chuẩn hội tụ.



Hình 8. Cơ chế cập nhật cá thể trong PSO

Thuật toán tối ưu hóa đàn kiến (Ant Colony Optimization - ACO)

Theo [24], ACO được giới thiệu bởi Dorigo năm 1991 và đang được áp dụng rộng rãi trong nghiên cứu và giải các bài toán tối ưu. Thuật toán dựa trên ý tưởng mô phỏng cách tìm đường của đàn kiến trong tự nhiên. ACO [9] kết hợp các thông tin kinh nghiệm và cập nhật thông qua vết mùi của các con kiến. Để thực hiện được điều này, các cá thể kiến thực hiện các bước đi ngẫu nhiên trên một đồ thị cấu trúc $G = (C, L)$,

trong đó, C là các thành phần lời giải, L là tập các kết nối. Trong thuật toán, sau khi khởi tạo các tham số, ma trận mùi và các cá thể kiến của đàn kiến, tại mỗi bước, mỗi cá thể xây dựng một lời giải riêng. Những lời giải sau đó được sử dụng để cập nhật lại các giá trị mùi và lựa chọn lời giải tốt nhất.

Để xây dựng một thuật giải ACO thích hợp, ta cần lưu ý tới ba yếu tố. Thứ nhất, việc xây dựng một đồ thị cấu trúc thích hợp. Thứ hai, xây dựng lời giải cho mỗi cá thể kiến một cách hiệu quả. Cuối cùng là chọn quy tắc cập nhật mùi hợp lý, có nhiều chiến lược cập nhật mùi và tùy thuộc vào bài toán chúng ta cần lựa chọn cách tối ưu nhất.

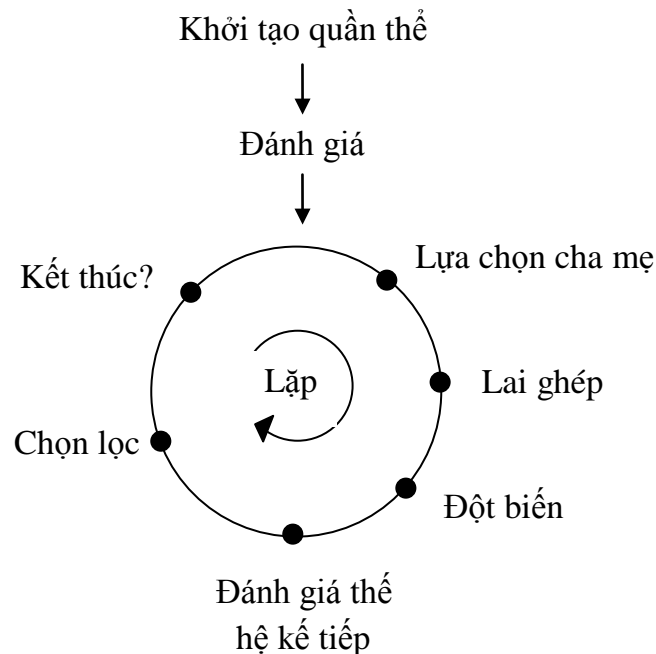
Các thuật toán tiến hóa (Evolutionary Algorithms)

Thuật toán tiến hóa là một lớp các thuật toán tìm kiếm dựa trên “quần thể” [19]. Khái niệm “quần thể” được sử dụng để chỉ một tập các lời giải khả thi của bài toán. Các thuật toán này được xây dựng dựa trên thuyết tiến hóa của Darwin với quan niệm: “Quá trình tiến hóa tự nhiên là quá trình hoàn hảo nhất, hợp lý nhất và tự nó đã mang tính tối ưu” [27], nghĩa là những cá thể nào thích nghi nhất sẽ có nhiều cơ hội tồn tại và phát triển hơn. Quan niệm này được xem như một tiên đề đúng, không chứng minh được nhưng phù hợp với thực tế khách quan [19]. Quá trình tiến hóa thể hiện tính tối ưu ở chỗ, thế hệ sau bao giờ cũng tốt hơn (phát triển hơn, hoàn thiện hơn) thế hệ trước và được duy trì bởi hai quá trình cơ bản: sinh sản và chọn lọc tự nhiên. Trong suốt quá trình tiến hóa, nhờ lai ghép các cá thể (lời giải) ở thế hệ cha – mẹ mà thế hệ mới luôn được sinh ra để bổ sung thay thế thế hệ cũ. Một đặc điểm khác của thuyết tiến hóa là một cá thể mới có thể mang những đặc tính của cha mẹ, đó là biểu hiện của di truyền và cũng có thể mang thuộc tính hoàn toàn khác, chính là biểu hiện của đột biến trong tự nhiên. Cá thể nào phát triển hơn, thích nghi hơn với môi trường thì cá thể đó sẽ tồn tại, ngược lại sẽ bị đào thải. Đây chính là quá trình chọn lọc tự nhiên.

Thuật toán tiến hóa bao gồm giải thuật di truyền (genetic algorithm), lập trình di truyền (genetic programming), chiến lược tiến hóa (evolutionary strategy) ... Các thuật

toán tiến hóa này, tuy có nhiều điểm khác biệt nhưng đều mô phỏng ba quá trình cơ bản: lai ghép, đột biến và chọn lọc tự nhiên (hình 9), trong đó:

- Phép lai ghép là phép sinh ra cá thể mới bằng cách ghép các nguyên liệu di truyền từ hai hay nhiều cá thể cha mẹ.
- Phép đột biến là phép sinh trong đó cá thể con mang thuộc tính không có trong cả cha và mẹ. Xác suất của phép đột biến này nhỏ hơn nhiều so với phép lai ghép.
- Chọn lọc chính là loại bỏ cá thể xấu ra khỏi quần thể nhờ quá trình đánh giá. Thời điểm quá trình chọn lọc hoàn thành thì ta nói kết thúc một thế hệ tiến hóa.



Hình 9. Sơ đồ thuật toán tiến hóa

Trong số những thuật toán tiến hóa, giải thuật di truyền được biết đến nhiều hơn cả. Đây cũng là giải thuật mà tác giả chọn lựa để giải quyết bài toán được ra trong chương 1. Vì lẽ đó, phần tiếp theo của chương sẽ được dành để trình bày rõ hơn về giải thuật di truyền, các đặc tính và các thành phần xây dựng nên giải thuật này.

2.3. Giải thuật di truyền

Vào đầu những năm 1960s, nhà khoa học người Mỹ gốc Án, John Henry Holland, lần đầu tiên đưa ra ý tưởng về giải thuật di truyền [7]. Mục đích ban đầu của ông là để hiểu rõ hiện tượng thích nghi (adaptation) xảy ra trong tự nhiên từ đó phát triển các kỹ thuật đưa tính thích nghi tự nhiên vào hệ thống tính toán. Sau đó, ông cùng các sinh viên của mình phát triển ý tưởng này và đến năm 1975, giải thuật di truyền chính thức được trình bày trong cuốn sách *Adaptation in Natural and Artificial Systems* (Thích nghi trong tự nhiên và trong các hệ thống nhân tạo).

Sau này, giải thuật di truyền được xếp vào lớp thuật toán tiến hóa bởi giải thuật này mô phỏng tiến hóa sinh học trong tự nhiên và cũng tuân theo thuyết tiến hóa của Darwin với mục đích tìm ra lời giải tối ưu hoặc gần tối ưu.

Các bước của giải thuật di truyền gần tương tự như đã chỉ ra ở sơ đồ hình 9 trên đây. Theo đó, trước khi khởi tạo quần thể để tiến hành quá trình tiến hóa gồm lai ghép, đột biến, đánh giá, chọn lọc cần thực hiện một bước quan trọng là: mã hóa cá thể. Các khái niệm quần thể, lai ghép, đột biến, chọn lọc, mã hóa đều được mượn từ thuyết tiến hóa của Darwin. Cụ thể như sau:

2.3.1. Mã hóa cá thể

Mã hóa cá thể chính là cách biểu diễn lời giải khả thi của bài toán. Nói cách khác, một lời giải được gọi là cá thể khi và chỉ khi đó là lời giải chấp nhận được.

Có rất nhiều cách mã hóa khác nhau và tùy thuộc vào từng bài toán mà lựa chọn cách mã hóa cho phù hợp. Đây cũng chính là một vấn đề cần được xem xét kỹ trước khi thực hiện các bước tiếp theo của giải thuật di truyền. Bởi nó quyết định phép lai ghép, đột biến nào nên được sử dụng để đảm bảo lời giải thu được vẫn là một cá thể. Có thể nói, mã hóa cá thể là bước quan trọng nhất, quyết định sự thành công của giải thuật di truyền đến 50%.

Một số cách mã hóa truyền thống hay được sử dụng là:

- Mã hóa nhị phân
- Mã hóa hoán vị
- Mã hóa số thực
- ...

Mã hóa nhị phân sử dụng một chuỗi các bit (0, 1) để biểu diễn lời giải. Ví dụ như bài toán tìm x để tối thiểu hàm f trong mục 1.2, tùy thuộc vào độ chính xác mong muốn mà ta có thể sử dụng một số lượng bit (0, 1) để biểu diễn giá trị của x trong đoạn $[1, 3]$.

Mã hóa hoán vị thường được sử dụng trong những bài toán tìm hoán vị tối ưu như bài toán người du lịch, bài toán phân công đã đề cập trong mục 1.1. Khi đó, giả sử lời giải cần tìm là một hoán vị của 5 số tự nhiên từ 1 đến 5, thì ta có thể mã hóa cá thể bởi một dãy số như: 2, 3, 1, 4, 5.

Với bài toán tối thiểu hàm f ở mục 1.2, ta cũng có thể sử dụng mã hóa số thực, nghĩa là coi mỗi số thực bất kỳ trong đoạn $[1, 3]$ là một cá thể. Tuy nhiên, cách mã hóa này sẽ làm cho phép lai ghép gặp khó khăn để sinh ra một cá thể mới như được trình bày trong mục 3.3 dưới đây.

Trong nhiều bài toán, khi lời giải có nhiều thành phần, thì những thành phần đó thường được gọi là nhiễm sắc thể hay gen để thể hiện rõ cấu trúc. Tuy nhiên, về bản chất, các khái niệm này chỉ là cách gọi để phân biệt các mức cấu tạo nên một lời giải. Ta hoàn toàn có thể gọi cá thể là gen hay nhiễm sắc thể nếu điều đó không gây hiểu lầm cho người đọc.

2.3.2. Khởi tạo quần thể

Sau khi tìm được cách mã hóa cá thể, bước tiếp theo trong giải thuật di truyền là khởi tạo quần thể. Đây chính là bước thể hiện rõ nhất đặc điểm của giải thuật này, đó là tìm kiếm dựa trên quần thể. Và như đã giải thích ở trên, quần thể chính là cách gọi một tập các lời giải chấp nhận được của bài toán. Hay chính là một tập các cá thể. Như vậy,

Có thể thấy, việc khởi tạo quần thể chính là việc sinh ra các cá thể với một số lượng nhất định, được gọi là kích thước quần thể.

Và cũng tương tự như cách mã hóa cá thể, có rất nhiều cách để khởi tạo quần thể. Thông thường, người ta thường chọn cách khởi tạo ngẫu nhiên, nghĩa là tạo ra các lời giải một cách ngẫu nhiên, miễn sao đảm bảo đó là một cá thể.

Ngoài ra, khởi tạo hướng tốt (heuristic) cũng được sử dụng để đưa vào quần thể ban đầu những cá thể có chất lượng không quá tồi với hy vọng quần thể sẽ tìm ra những cá thể mới tốt hơn từ những cá thể này và như vậy sẽ có nhiều cơ hội thu được kết quả tối ưu hơn. Tuy nhiên, phương pháp khởi tạo hướng tốt không nên quá phức tạp vì thông thường lời giải tốt hơn sẽ được tìm thấy trong quá trình tối ưu sau đó. Thuật toán tham lam có thể được sử dụng để thực hiện điều này bởi tính đơn giản và hiệu quả lời giải của nó.

Người ta đôi khi cũng kết hợp cả hai phương pháp này để khởi tạo quần thể lời giải ban đầu với mục đích tạo ra sự đa dạng cho quần thể.

2.3.3. Hàm thích nghi

Để đánh giá cá thể là tốt hay không tốt, giải thuật di truyền sử dụng hàm thích nghi. Hàm thích nghi chính là hàm thể hiện mục tiêu của bài toán. Cá thể nào có giá trị hàm thích nghi nhỏ hơn thì tốt hơn (đối với bài toán cực tiểu) hoặc ngược lại (đối với bài toán cực đại).

Thông thường, hàm mục tiêu của bài toán cũng chính là hàm thích nghi của giải thuật. Tuy nhiên cũng có trường hợp, người ta xây dựng một hàm thích nghi khác nhưng vẫn đảm bảo tiêu chí của bài toán bởi lý do tính toán theo hàm mục tiêu gặp khó khăn hay có độ phức tạp lớn.

2.3.4. Các toán tử di truyền

Phép lai ghép và đột biến được gọi chung là toán tử di truyền. Đây cũng là một giai đoạn quan trọng của giải thuật. Nhờ có lai ghép, đột biến mà quần thể được mở rộng,

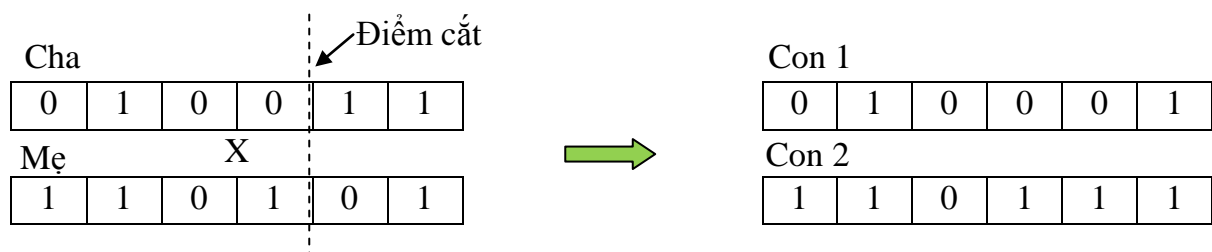
nghĩa là khám phá thêm được không gian lời giải và do đó, có nhiều cơ hội tìm được lời giải tối ưu toàn cục hơn.

Hai toán tử này mặc dù đều có vai trò sinh ra những cá thể mới cho quần thể nhưng lại có ý nghĩa khác nhau đối với giải thuật di truyền. Phép lai ghép có nhiệm vụ sinh ra những cá thể mới bằng cách tận dụng những nguyên liệu di truyền tốt có ở cả cá thể cha và cá thể mẹ. Trong khi đó, cá thể được đột biến nhằm sinh ra cá thể mới có sự khác biệt so với cá thể cũ hay cá thể cha mẹ của nó. Mục đích là để tránh tối ưu địa phương.

Phép lai ghép

Các phép lai ghép thường được thực hiện với sự tham gia của hai cá thể cha – mẹ nhưng đôi khi có thể nhiều hơn. Một số phép lai ghép cơ bản gồm:

- Lai ghép một điểm cắt
- Lai ghép hai điểm cắt



Hình 10. Lai ghép một điểm cắt

Trong phép lai ghép một điểm cắt, Sau khi chọn lựa hai cá thể cha, mẹ, một điểm cắt sẽ được quyết định. Khi đó, phần nằm phía sau điểm cắt của cá thể cha, mẹ sẽ được trao đổi cho nhau sinh ra hai con mới (hình 10). Tương tự với phép lai ghép hai điểm cắt với số điểm cắt là 2 và phần trao đổi nằm giữa hai điểm cắt. Không có cách xác định điểm cắt cụ thể nào được đưa ra, phần lớn đều lựa chọn ngẫu nhiên.

Với cách mã hóa hoán vị như đã trình bày trên đây, phép lai ghép theo điểm cắt này không hiệu quả vì không đảm bảo con sinh ra là một cá thể. Vì vậy cần có cách lai ghép khéo léo hơn. Mã hóa số thực mặc dù có nhiều cách để sinh ra con vẫn là số thực

nằm trong đoạn cho trước nhưng rất khó kiểm soát sự biến động của nó, và trong trường hợp này, lai ghép giống với đột biến nhiều hơn.

Đột biến

Phép đột biến thay đổi giá trị của một hay một số giá trị của gen theo cách nào đó để sinh ra một cá thể mới. Với toán tử này, có thể có sự tham gia của hai cá thể cha mẹ hoặc không.

Ví dụ với cách mã hóa nhị phân, phép đột biến đơn giản chỉ là đảo bit ở một vị trí bất kỳ của cá thể đó. Hay với mã hóa hoán vị, đột biến sẽ tiến hành đổi giá trị ở hai gen trong cá thể cho nhau. Còn trong mã hóa số thực, đột biến có thể thực hiện bằng cách thêm vào hoặc bớt đi một giá trị của cá thể đó một lượng nhỏ bất kỳ.

2.3.5. Chọn lọc

Sau khi thực hiện toán tử di truyền, quần thể ban đầu sẽ có thêm nhiều thành viên mới khiến kích thước của quần thể tăng lên. Tuy nhiên, một đặc điểm của giải thuật này là kích thước quần thể luôn được cố định sau mỗi thế hệ. Điều này được thực hiện nhờ quá trình chọn lọc. Quần thể sau khi tiến hành lai ghép, đột biến sẽ được sắp xếp theo độ thích nghi của cá thể rồi sau đó giữ lại những cá thể tốt nhất.

Cũng có một phương pháp chọn lọc khác hay được sử dụng đó là chọn lọc theo vòng tròn Roulette. Ý tưởng của phương pháp này sắp xếp các cá thể vào một vòng tròn với độ rộng của cung tròn phản ánh độ thích nghi của chúng. Sau đó, ta tiến hành quay vòng tròn bằng cách sinh ngẫu nhiên 1 số trong khoảng từ 0 đến 1 và chọn ra cá thể tương ứng.

2.3.6. Điều kiện dừng của giải thuật

Điều kiện dừng được đưa ra nhằm đảm bảo tính dừng của giải thuật di truyền. Có nhiều chiến lược để thực hiện điều này. Thông thường, GA sẽ dừng sau một số thế hệ được định nghĩa từ trước hoặc dừng lại sau một số thế hệ mà chất lượng lời giải không tốt lên. Số thế hệ này cũng là một tham số của giải thuật di truyền. Tùy từng bài toán

và quá trình thực nghiệm mà giá trị này sẽ được thiết lập cho phù hợp sao cho đảm bảo tìm được lời giải tốt mà chi phí thời gian chấp nhận được.

2.3.7. Tính chất của giải thuật di truyền

Có thể nói, giải thuật di truyền rất đa dạng vì chỉ với một sơ đồ giải thuật (hình 9) như trên, nếu khởi tạo quần thể khác nhau, chiến lược chọn cha mẹ (số lượng, cách chọn) khác nhau, sử dụng phép lai ghép, đột biến khác nhau hay thậm chí cách đánh giá cá thể khác nhau cũng tạo nên một giải thuật di truyền mới. Vì lẽ đó, khi tiến hành so sánh hai giải thuật di truyền cùng giải quyết một bài toán, người ta thường xem xét dựa trên các tiêu chí này.

CHƯƠNG 3: GIẢI THUẬT ĐỀ XUẤT GIẢI BÀI TOÁN TÌM ĐƯỜNG XÂM NHẬP CÓ KHẢ NĂNG BỊ PHÁT HIỆN NHỎ NHẤT TRONG MẠNG CẢM BIẾN KHÔNG DÂY MÔ HÌNH XÁC SUẤT

Chương này trình bày giải thuật di truyền đề xuất GAMEP theo các nội dung:

1. Mã hóa cá thể
2. Hàm thích nghi
3. Các toán tử di truyền
4. Chọn lọc

3.1. Mã hóa cá thể

Trong giải thuật di truyền, mỗi cá thể là một lời giải của bài toán đang xét đến. Với bài toán PM-based-MEP, một lời giải là một tập các điểm có thứ tự L_\varnothing mà hình thành nên đường đi \varnothing từ I đến F . Do đó, một cách trực quan về cách biểu diễn cá thể là sử dụng một mảng lưu các điểm trong không gian hai chiều, cụ thể là tọa độ thực tương ứng của những điểm này trong đó phần tử đầu tiên và cuối cùng của mảng luôn là $(0, y_I)$ và (W, y_F) . Một cách hình ảnh của biểu diễn này được thể hiện ở hình 11.

$(0, y_I)$	(x_1, y_1)	(x_1, y_2)	...	(W, y_F)
------------	--------------	--------------	-----	------------

Hình 11. Biểu diễn cá thể

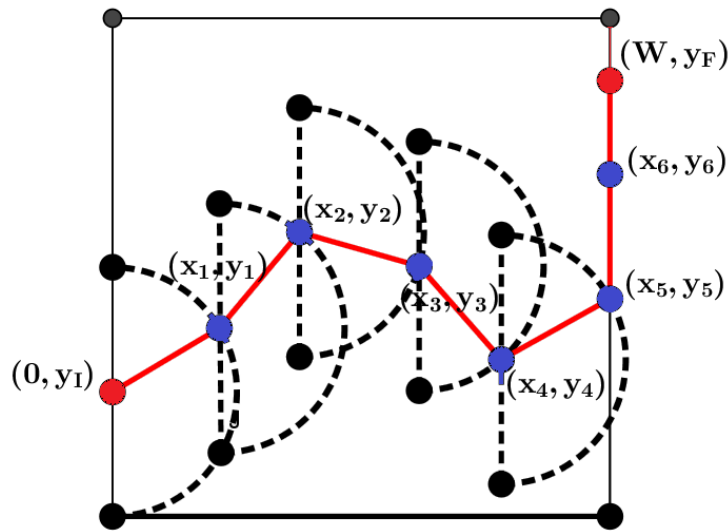
Để thỏa mãn các ràng buộc (*, **), các cá thể của GAMEP có các đặc điểm sau:

- Các gen luôn biểu diễn các điểm thuộc miền \mathcal{R} , nghĩa là miền giá trị của tọa độ x và tọa độ y lần lượt là $[0, W]$ và $[0, H]$
- Tọa độ x của gen thứ i không lớn hơn tọa độ x của gen thứ j với mọi $j > i$
- Khoảng cách giữa hai điểm biểu diễn bởi 2 gen liên tiếp, ngoại trừ 2 gen cuối cùng của một cá thể, luôn bằng len . Tham số len là một số nguyên dương không nhỏ hơn 1. Như vậy, chỉ khi len bằng 1 thì ràng buộc (*) mới được thỏa mãn, trong các trường hợp khác, cá thể tốt nhất tìm được sẽ cần

tính lại giá trị exposure bằng cách chia các đoạn $[(x_i, y_i), (x_{i+1}, y_{i+1})]$ thành len đoạn có độ dài 1 đơn vị.

Để khởi tạo cá thể với các đặc điểm trên đây, tác giả sử dụng phương pháp ngẫu nhiên. Ý tưởng của phương pháp này là một điểm biểu diễn bởi một gen (sau đây sẽ gọi tắt là gen) thì luôn nằm trên nửa đường tròn bên phải có tâm tại gen phía trước với bán kính len và nằm trong miền giám sát \mathcal{R} .

Hình 12 dưới đây biểu diễn quá trình tạo ra một cá thể *indi*, trong đó, gen thứ 2 (x_1, y_1) là một điểm ngẫu nhiên nằm trên nửa đường tròn bên phải có tâm tại gen đầu tiên $(0, y_I)$ được biểu diễn bởi nét đứt. Quá trình này lặp lại cho đến khi có nửa đường tròn bên phải nào đó giao với biên phải của miền \mathcal{R} . Như trong hình 12, nửa đường tròn bên phải của gen (x_4, y_4) giao với biên phải của \mathcal{R} tại hai điểm và một trong hai điểm này được chọn làm gen của cá thể đang khởi tạo. Từ đây, nếu khoảng cách của gen hiện tại của cá thể với điểm cuối (W, y_F) còn lớn hơn len thì cá thể sẽ có thêm gen nằm trên biên phải của \mathcal{R} , theo hướng về F ; còn nếu không thì điểm (W, y_F) sẽ được thêm vào cá thể và quá trình khởi tạo kết thúc. Đường thẳng màu đỏ trong hình 12 biểu diễn một ví dụ về kiểu hình của cá thể.



Hình 12. Khởi tạo cá thể

Đoạn mã giả của quá trình khởi tạo này được mô tả dưới đây.

Thuật toán 1: Khởi tạo cá thể

Đầu vào:

Chiều rộng và chiều dài của miền giám sát \mathcal{R} : W, H

Tham số độ dài: len

Tọa độ điểm đầu: $(0, y_I)$

Tọa độ điểm cuối: (W, y_F)

Đầu ra: Một cá thể $indi = [(0, y_I), (x_1, y_1), \dots, (W, y_F)]$

Begin

1. $indi \leftarrow (0, y_I)$
 2. $x \leftarrow 0$
 3. $y \leftarrow y_I$
 4. **while** $x < W$ **do**
 5. $\delta X \leftarrow \text{random}(0, len)$
 6. $\delta Y \leftarrow \text{random} \left\{ -\sqrt{len^2 - \delta X^2}, \sqrt{len^2 - \delta X^2} \right\}$
 7. **if** $(y + \delta Y > H \parallel y + \delta Y < 0)$ **do**
 8. $\delta Y = -\delta Y$
 9. **end if**
 10. $x \leftarrow x + \delta X$
 11. $y \leftarrow y + \delta Y$
 12. $indi \leftarrow indi \cup (x, y)$
 13. **end while**
 14. $\delta X \leftarrow x - W$
 15. $\delta Y \leftarrow \text{random} \left\{ -\sqrt{len^2 - \delta X^2}, \sqrt{len^2 - \delta X^2} \right\}$
 16. **if** $(y + \delta Y > H \parallel y + \delta Y < 0)$ **do**
 17. $\delta Y = -\delta Y$
 18. **end if**
-

```
19.  $x \leftarrow W$ 
20.  $y \leftarrow y + \text{delta}Y$ 
21.  $\text{delta}Y \leftarrow y - y_F$ 
22. while  $|\text{delta}Y| > \text{len}$  do
23.   if  $\text{delta}Y < 0$  do
24.      $y \leftarrow y + \text{len}$ 
25.      $\text{delta}Y \leftarrow \text{delta}Y + \text{len}$ 
26.   else
27.      $y \leftarrow y - \text{len}$ 
28.      $\text{delta}Y \leftarrow \text{delta}Y - \text{len}$ 
29.   end if
30. end while
31.  $\text{indi} \leftarrow \text{indi} \cup (W, y)$ 
32. return  $\text{indi}$ 
```

End.

Trong đó, hàm $\text{random}(a, b)$ để lấy một số ngẫu nhiên thuộc đoạn $[a, b]$, còn hàm $\text{random}\{a, b\}$ trả về giá trị a hoặc b .

Bằng cách tạo ra một lượng cố định cho trước các cá thể theo phương thức trên đây, ta thu được quần thể ban đầu của thuật toán di truyền.

3.2. Hàm thích nghi

Mỗi cá thể trong quần thể có một “độ tốt” của nó được tính thông qua hàm thích nghi đã được định nghĩa trong công thức (19). Cá thể có giá trị exposure càng nhỏ thì cá thể đó càng tốt.

3.3. Toán tử di truyền

3.3.1. Lai ghép

GAMEP sử dụng hai phép lai ghép, đó là: BLX- α [7] và lai ghép một điểm cắt. Nội dung hai phép lai này được trình bày dưới đây.

Lai ghép BLX- α

Phép lai ghép BLX- α [7] được áp dụng để tạo ra thế hệ con. Tuy nhiên, do cách mã hóa đặc trưng của GAMEP, nên tác giả đã điều chỉnh phép lai này. Theo đó, các gen của con lai sẽ được tạo ra thông qua giá trị góc tạo bởi 2 gen cha, mẹ thay vì được lấy trực tiếp từ các đoạn (interval). Phép lai được hiệu chỉnh này được gọi là ALX- α .

Xét một cặp cha mẹ như sau:

$$indi1 = [(0, y_I), (x_1, y_1), (x_2, y_2), \dots, (W, y_F)]$$

$$indi2 = [(0, y_I), (x'_1, y'_1), (x'_2, y'_2), \dots, (W, y_F)]$$

và con sinh ra:

$$child = [(0, y_I), (x''_1, y''_1), (x''_2, y''_2), \dots, (W, y_F)]$$

Để thuận tiện, ký hiệu P_{i1} biểu diễn gen (x_i, y_i) , P'_{i2} biểu diễn (x'_i, y'_i) và Z_i biểu diễn (x''_i, y''_i) . Bởi vì mọi cá thể đều có gen đầu tiên là $(0, y_I)$ nên ALX- α sẽ được áp dụng từ gen thứ 2 trở đi. Kí hiệu φ_1, φ_2 lần lượt là góc tạo bởi các vecto $\overrightarrow{Z_i P_{i1}}$ và $\overrightarrow{Z_i P'_{i2}}$ với trục x trong đó P_{i1} và P'_{i2} là các gen có chỉ số nhỏ nhất có tọa độ x lớn hơn tọa độ x của Z_i . Sau đó, chọn ngẫu nhiên một góc φ trong đoạn $[lb, ub]$ với

$$lb = \max \left\{ -\frac{\pi}{2}, \min \{ \varphi_1, \varphi_2 \} \right\} - I\alpha,$$

$$ub = \min \left\{ \frac{\pi}{2}, \max \{ \varphi_1, \varphi_2 \} \right\} - I\alpha$$

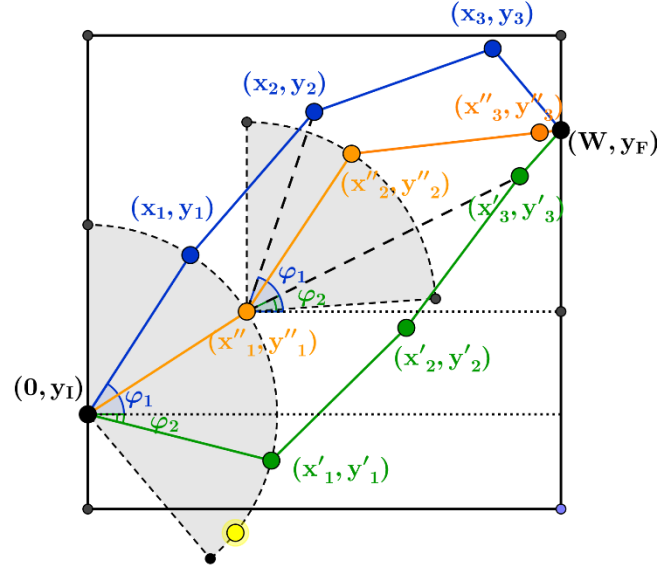
Trong đó, $I = |\varphi_1 - \varphi_2|$. Dựa trên giá trị góc φ chọn được, gen Z_{i+1} được tạo ra với

$$x''_{i+1} = len * \cos \varphi,$$

$$y''_{i+1} = len * \sin \varphi$$

Trong trường hợp y''_i vượt quá biên của miền \Re thì nếu giá trị này âm, nó sẽ được gán lại bằng 0; ngược lại, gán bằng H . Khi giá trị của x''_i lớn hơn W , nó sẽ được gán lại bằng W và quá trình lai ghép sẽ dừng lại. Trong cả hai trường hợp này, giá trị của y''_i

và x''_i đều được điều chỉnh để đảm bảo khoảng cách giữa hai gen liên tiếp bằng len . Sau khi phép lai ALX- α kết thúc, nếu khoảng cách giữa gen cuối cùng của con lai và F còn lớn hơn len thì quá trình thêm gen sẽ được thực hiện tương tự như trong quá trình khởi tạo biểu diễn từ dòng 19 đến dòng 31 trong thuật toán 1.



Hình 13. Phép lai ALX- α

Lai ghép một điểm cắt

Với mỗi cặp cha mẹ, $indi1$ và $indi2$, một điểm cắt idc , tức là vị trí gen thứ idc trên cá thể cha, mẹ, được sinh ngẫu nhiên thỏa mãn $0 < idc < \min\{|indi1|, |indi2|\}$. Điểm cắt này chia số lượng gen của mỗi cha, mẹ làm 2 phần, phần thứ nhất từ gen đầu tiên đến gen ở vị trí thứ idc , phần thứ hai là số gen con lại. Sau đó, hai con lai, $child1$ và $child2$, sao chép toàn bộ phần gen đầu của cha $indi1$, và mẹ $indi2$ một cách tương ứng. Phần gen còn lại của mỗi con lai được tạo ra như sau để đảm bảo các đặc tính của một cá thể như đã được mô tả ở mục 3.1.

Xét con lai $child1$, từ vị trí gen idc của cha $indi1$, ta tìm cách nối với gen ở vị trí idc trên mẹ $indi2$. Có 3 trường hợp xảy ra:

- Trường hợp 1: Khoảng cách giữa 2 gen này vừa đúng bằng giá trị len

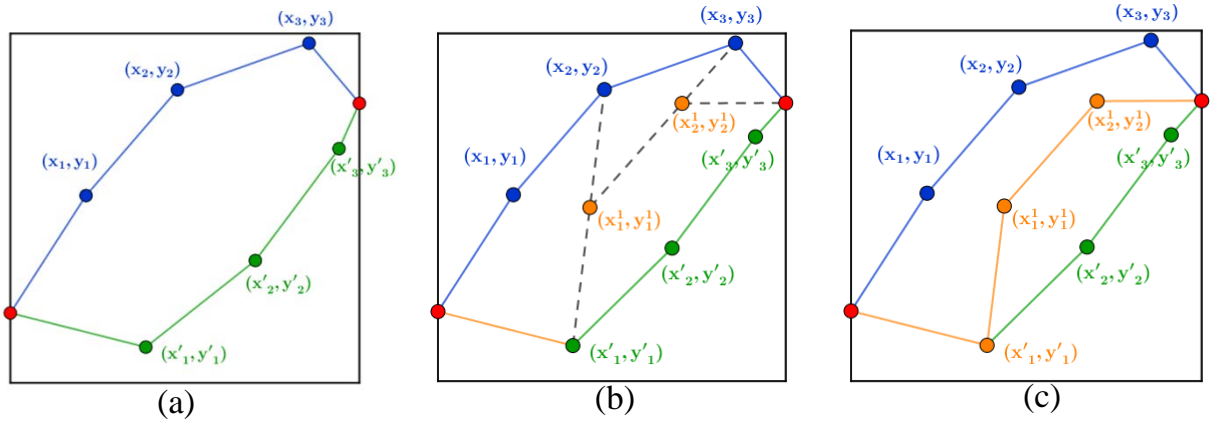
- Trường hợp 2: Khoảng cách giữa 2 gen này nhỏ hơn len
- Trường hợp 3: Khoảng cách giữa 2 gen lớn hơn len

Trong trường hợp thứ nhất, ta chỉ đơn giản sao chép toàn bộ phần thứ 2 của mẹ *indi2* vào sau con *child1* hiện tại để sinh ra con *child1* hoàn chỉnh.

Nếu xảy ra trường hợp 2, ta tiến hành tăng dần chỉ số *idc* trên mẹ *indi2*, cho đến khi khoảng cách giữa gen hiện tại của con với gen đang xét trên mẹ bằng hoặc lớn hơn len . Nếu bằng, ta làm tương tự như trường hợp thứ nhất: sao chép toàn bộ phần gen từ vị trí đang xét trên mẹ *indi2* sang con *child1* và kết thúc quá trình tạo *child1*. Nếu lớn hơn, thực hiện tương tự như trường hợp 3.

Trong trường hợp 3, ta tìm điểm thuộc đoạn thẳng nối gen hiện tại của con với gen đang xét trên mẹ sao cho khoảng cách từ nó và gen hiện tại của con bằng len .

Quá trình trên được thực hiện cho đến khi gen cuối cùng trên cá thể mẹ được xét đến và khoảng cách giữa gen hiện tại của con và gen cuối cùng của mẹ nhỏ hơn len .



Hình 14. Minh họa phép lai một điểm cắt

Hình 14 minh họa quá trình tạo ra con lai *child1* (màu cam) từ cha *indi1* (màu xanh lá) và mẹ *indi2* (màu xanh dương) qua phép lai một điểm cắt tại điểm lai $idc = 1$ (hình 14(a)). Theo đó, phần đầu của cha $\{(0, y_1), (x'_1, y'_1)\}$ được sao chép sang con *child1*. Do khoảng cách giữa (x'_1, y'_1) và (x_1, y_1) nhỏ hơn len , ta xét vị trí tiếp theo (x_2, y_2) trên

mẹ. Lúc này, khoảng cách giữa gen hiện tại của con (x'_1, y'_1) và (x_2, y_2) lớn hơn len , nên ta chọn điểm (x_1^1, y_1^1) trên đoạn thẳng $[(x'_1, y'_1), (x_2, y_2)]$ cách (x'_1, y'_1) một khoảng bằng len làm gen tiếp theo của *child1*. Tương tự như vậy cho việc tạo gen (x_2^1, y_2^1) (hình 14(b)). Đến đây, do khoảng cách giữa gen hiện tại của *child1* và gen cuối cùng của mẹ *indi2* nhỏ hơn len nên quá trình sinh con *child1* kết thúc. Hình 14(c) là hình ảnh của con *child1* (màu cam) thu được sau quá trình lai ghép một điểm cắt.

Con lai *child2* được tạo ra hoàn toàn tương tự con *child1* với việc đảo ngược vai trò *indi1* và *indi2*.

3.3.2. Đột biến

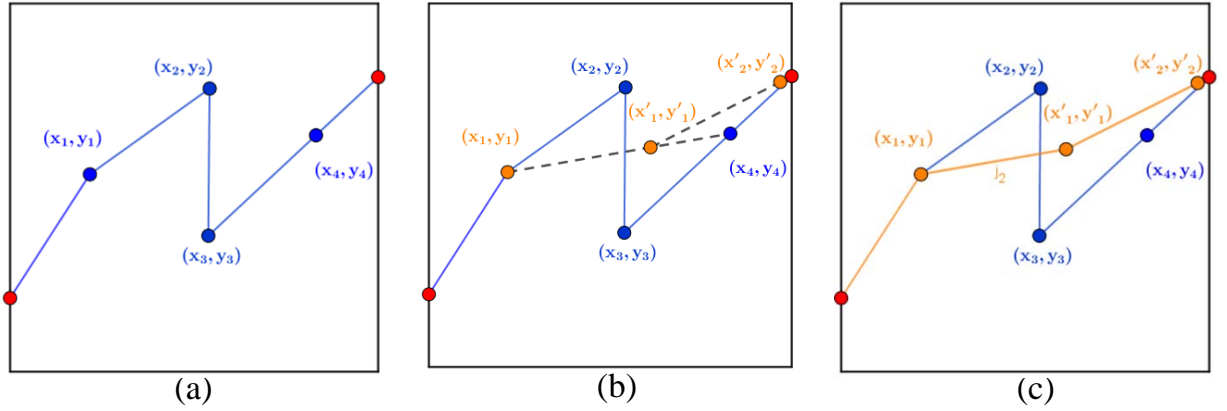
Với mỗi cá thể được chọn *indi*, một gen ngẫu nhiên (x_i, y_i) , trừ gen cuối, được chọn ngẫu nhiên để tiến hành quá trình đột biến như sau.

Từ gen đầu tiên đến gen (x_i, y_i) của cá thể *indi* được sao chép sang con đột biến *child*. Theo đó, gen tiếp sau gen được chọn của *indi*, tức gen (x_{i+1}, y_{i+1}) , sẽ bị lược bỏ. Kể từ gen (x_{i+2}, y_{i+2}) , ta xét khoảng cách giữa nó với gen hiện tại của con với 3 trường hợp sau:

- Trường hợp 1 – Khoảng cách giữa hai gen bằng len : sao chép toàn bộ gen từ (x_{i+2}, y_{i+2}) đến gen cuối cùng của *indi* sang *child* và kết thúc quá trình đột biến.
- Trường hợp 2 – Khoảng cách giữa hai gen nhỏ hơn len : tăng dần chỉ số của gen trên *indi* cho đến khi khoảng cách giữa gen đang xét của *indi* với gen hiện tại của con lớn hơn hoặc bằng len . Nếu bằng, thực hiện tương tự trường hợp 1 với đoạn sao chép bắt đầu từ vị trí đang xét của *indi*. Nếu lớn hơn, thực hiện quá trình đột biến như trường hợp 3 dưới đây.

- Trường hợp 3 – Khoảng cách giữa hai gen lớn hơn len : tìm điểm trên đoạn thẳng nối gen hiện tại của con với gen đang xét trên *indi* sao cho khoảng cách giữa nó với gen hiện tại của con bằng len .

Quá trình đột biến kết thúc khi gen cuối cùng của *indi* được xét đến và khoảng cách giữa gen hiện tại của con *child* với gen cuối cùng này nhỏ hơn len .



Hình 15. Minh họa quá trình đột biến

Hình 15 minh họa quá trình đột biến cá thể *indi* (hình 15(a)) tại vị trí (x_1, y_1) . Bỏ qua gen (x_2, y_2) , khoảng cách giữa (x_1, y_1) và (x_3, y_3) được xét. Tuy nhiên, khoảng cách này nhỏ hơn len nên ta xét đến đoạn $[(x_1, y_1), (x_4, y_4)]$. Đoạn thẳng này có độ dài lớn hơn len , do đó, điểm (x'_1, y'_1) thuộc đoạn này cách (x_1, y_1) một khoảng bằng len được chọn làm gen tiếp theo của con *child*. Quá trình được diễn ra cho đến gen cuối cùng và khoảng cách giữa gen hiện tại tại (x'_2, y'_2) với gen cuối này nhỏ hơn len và ta thu được con đột biến với kiểu hình thu được ở hình 15(c).

3.4. Chọn lọc

Một điểm đặc biệt của thuật toán di truyền là sự ổn định kích thước quần thể, nghĩa là số lượng cá thể trong quần thể không thay đổi sau từng thế hệ. Tuy nhiên, các toán tử di truyền, lai ghép và đột biến, làm tăng số lượng này. Do đó, phép toán chọn lọc được sử dụng để chọn ra đủ số lượng cá thể đưa vào quá trình tiến hóa tiếp theo. Trong GAMEP, sau quá trình sản xuất, các cá thể mới sẽ được tính độ thích nghi. Sau đó,

Đề tài: Giải thuật xấp xỉ giải bài toán bao phủ barrier trong mạng cảm biến không dây

những cá thể tốt nhất sẽ được giữ lại và tiếp tục tiến hóa cho đến khi đạt đến số thế hệ đã định nghĩa từ trước hoặc tìm được cá thể có giá trị exposure bằng 0.

CHƯƠNG 4: KẾT QUẢ ĐẠT ĐƯỢC

4.1. Dữ liệu thử nghiệm.

Tác giả tạo ra 15 bộ topo mạng đặt tên là S_N_k , trong đó N là số cảm biến có giá trị thuộc tập $\{20, 40, 60, 80, 100\}$ và k là số thứ tự của bộ dữ liệu để phân biệt các topo có cùng số lượng cảm biến và nhận giá trị trong tập $\{0, 1, 2\}$.

Tất cả các cảm biến được triển khai ngẫu nhiên trong miền giám sát \mathcal{R} có kích thước 500×500 và sử dụng mô hình cảm biến xác suất với các tham số cho trong bảng dưới đây.

Bảng 1. Tham số mô hình cảm biến xác suất

Tham số	Giá trị
Ngưỡng A	6
Hằng số C	100
Hệ số suy giảm tín hiệu λ	1
Độ lệch chuẩn của nhiễu σ^2	1
Bán kính cảm biến r	100

Bởi vì hàm Q-function được sử dụng để tính giá trị xác suất như đã trình bày, các tham số của mô hình cảm biến xác suất được lựa chọn ở trên để thỏa mãn đầu vào của hàm Q có thể phủ được đoạn $[-5, 5]$.

Điểm xuất phát và điểm kết thúc được cố định là $(0, 150)$ và $(500, 450)$.

4.2. Thiết lập hệ thống

Để đánh giá sự ảnh hưởng của tham số *len* đến kết quả tìm được và thời gian chạy giải thuật, tác giả tiến hành thực nghiệm với một số giá trị *len* bằng: 1, 5 và 10. Tất cả các thực nghiệm này đều chạy với cùng một bộ các tham số của giải thuật di truyền được cho trong bảng dưới đây.

Bảng 2. Các tham số của GAMEP

Tham số	Giá trị
Số lần chạy trên một bộ dữ liệu	5
Số thế hệ	300
Kích thước quần thể	2000

Tỷ lệ lai ghép ALX- α	72%
Tỷ lệ lai ghép một điểm cắt	18%
Tỷ lệ đột biến	10%
Giá trị α trong ALX- α	0.5

Giá trị exposure của các đường đi sau khi tìm được sẽ được sử dụng để tính giá trị xác suất mà kẻ xâm nhập bị phát hiện P nếu đi trên đường đó bằng công thức sau.

$$P^{\rho} = 1 - e^{\overline{\ln(P^{\rho})}} = (1 - e^{-E(\rho)}) * 100\% \quad (20)$$

Chương trình thực nghiệm được cài đặt trên ngôn ngữ lập trình Java, sử dụng máy Intel® Core™ i5-3230M 2.60 GHz với 4GB RAM chạy hệ điều hành Windows 10 64 bit.

4.3. Kết quả thực nghiệm

Từ bảng 3 và bảng 4, ta có thể thấy rằng:

- Độ dài *len* bằng 1 và 10 có hiệu quả tương đương trong thuật toán GAMEP để tìm ra đường có giá trị exposure nhỏ nhất, trong đó, độ dài *len* 10 có xu hướng cho lời giải tốt với các bộ dữ liệu nhỏ ứng với các trường hợp mạng thưa còn *len* 1 lại tỏ ra tốt hơn với trường hợp mạng dày (số lượng cảm biến lên tới 100).
- Độ dài *len* 5 chỉ tốt với các bộ dữ liệu vừa có 60, 80 cảm biến tương ứng mạng có mật độ trung bình.
- 12/15 bộ dữ liệu có sự chênh lệch giữa xác suất phát hiện được của mạng cảm biến khi đối tượng di chuyển trên đường có giá trị exposure nhỏ nhất tương ứng của các giá trị *len* khác nhau là không lớn, chỉ từ 0.00% (bộ d_20_2) đến 6.63 % (bộ d_80_2). Chỉ 3 bộ d_40_1, d_60_2 và d_100_1 có sự chênh lệch này lớn hơn 13%.
- Trong trường hợp tính trung bình sau 5 lần chạy, thì có đến 13/15 dữ liệu có sự chênh lệch giữa xác suất phát hiện được của mạng tương ứng với các

giá trị *len* khác nhau không nhiều, từ 0.00% (bộ d_20_2) đến 4.82% (bộ d_80_2). Chỉ còn 2 bộ d_40_1 và d_100_1 vẫn có sự chênh lệch này lớn trên 16%.

- Khi giá trị exposure đạt đến 1 ngưỡng thì xác suất phát hiện đạt đến 100%. Lúc này, dù exposure tăng lên bao nhiêu đi nữa thì xác suất phát hiện vẫn luôn là 100%. Ngưỡng này có thể tính toán được, dựa vào công thức (19) và bảng khoảng 17. Điều này giải thích cho trường hợp bộ dữ liệu d_100_0 trong bảng 3 và bảng 4, mặc dù giá trị exposure là khác nhau nhưng đều lớn hơn 17 nên xác suất phát hiện tương ứng luôn là 100%.

Bảng 3. Bảng so sánh giá trị exposure tốt nhất và giá trị xác suất phát hiện tương ứng (%) tìm được bởi GAMEP sau 5 lần chạy trên các bộ dữ liệu với các giá trị len khác nhau

Dữ liệu	1		5		10	
	Best	Prob	Best	Prob	Best	Prob
d_20_0	0.00	0.22	0.00	0.22	0.00	0.19
d_20_1	0.25	21.70	0.26	22.50	0.26	22.67
d_20_2	0.00	0.12	0.00	0.12	0.00	0.12
d_40_0	0.19	17.29	0.17	15.98	0.17	15.94
d_40_1	0.51	39.58	0.58	44.00	0.77	53.57
d_40_2	0.03	3.37	0.04	3.57	0.04	3.73
d_60_0	4.76	99.13	4.06	98.25	4.03	98.20
d_60_1	0.13	12.15	0.18	16.53	0.18	16.46
d_60_2	1.73	82.18	1.33	73.51	2.16	88.37
d_80_0	0.55	42.10	0.53	41.00	0.44	35.68
d_80_1	12.21	100.00	12.21	100.00	12.43	100.00
d_80_2	1.52	78.04	1.26	71.41	1.30	72.60
d_100_0	27.41	100.00	27.51	100.00	27.31	100.00
d_100_1	0.39	32.28	0.67	48.84	0.46	36.51
d_100_2	4.38	98.73	4.41	98.77	4.52	98.90

Bảng 4. Bảng so sánh giá trị exposure trung bình và giá trị xác suất phát hiện tương ứng (%) tìm được bởi GAMEP sau 5 lần chạy trên các bộ dữ liệu với các giá trị len khác nhau

Dữ liệu	1		5		10	
	Avg	Prob	Avg	Prob	Avg	Prob
d_20_0	0.00	0.22	0.00	0.22	0.00	0.19
d_20_1	0.25	21.70	0.26	22.50	0.26	22.67
d_20_2	0.00	0.12	0.00	0.12	0.00	0.12
d_40_0	0.20	18.38	0.18	16.44	0.18	16.03
d_40_1	1.39	75.10	2.72	93.38	2.97	94.81
d_40_2	0.03	3.40	0.04	4.00	0.05	5.15
d_60_0	4.79	99.16	4.11	98.34	4.24	98.54
d_60_1	0.17	15.80	0.18	16.57	0.18	16.58
d_60_2	14.87	100.00	8.19	99.97	8.31	99.97
d_80_0	0.56	42.74	0.56	42.76	0.49	38.49
d_80_1	12.34	100.00	12.32	100.00	12.45	100.00
d_80_2	1.56	78.78	1.35	73.97	1.44	76.20
d_100_0	28.16	100.00	27.87	100.00	27.71	100.00
d_100_1	0.39	32.48	0.68	49.16	0.42	34.11
d_100_2	4.59	98.97	6.30	99.81	4.92	99.26

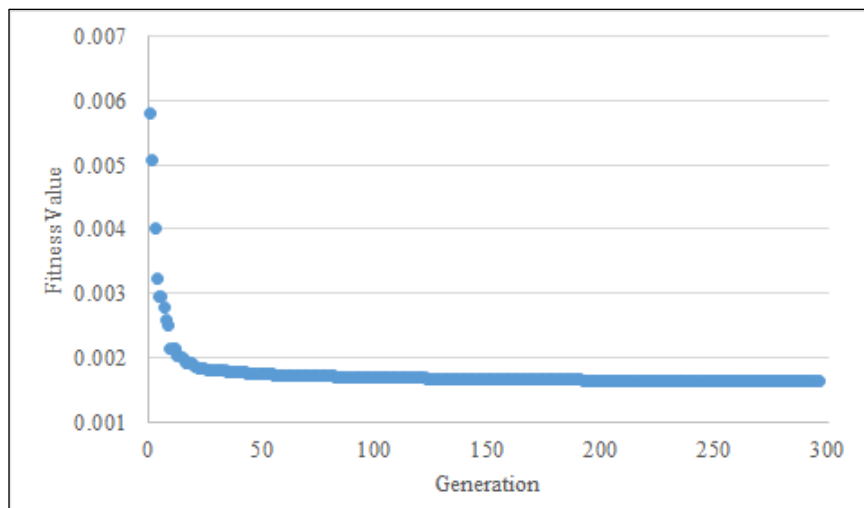
Từ bảng 5, ta có những nhận xét sau đây:

- Độ dài len 1 cho kết quả chạy ổn định nhất so với hai giá trị còn lại trong 9/15 bộ dữ liệu, sau đó là độ dài len 10 với 4/15 bộ, cuối cùng là len 5 với chỉ 2/15 bộ.
- Hầu hết các giá trị len đều cho kết quả ổn định, với độ lệch chuẩn rất nhỏ, đặc biệt có những bộ độ lệch chuẩn chỉ 0.00 như d_20_0, d_100_1. Tuy nhiên, có duy nhất bộ d_60_2 có độ lệch chuẩn tương đối lớn.
- Thời gian chạy của GAMEP sử dụng giá trị len 10 là ít nhất và nhanh gấp khoảng 10 lần so với độ dài len 1 và khoảng 2 lần so với độ dài len 5. Trong

khi đó, kết quả tìm được như đã trình bày ở những nhận xét trên không chênh lệch nhau quá lớn.

Bảng 5. Bảng so sánh giá trị độ lệch chuẩn và thời gian chạy trung bình của GAMEP sau 5 lần chạy trên các bộ dữ liệu với các giá trị len khác nhau

Instance	1		5		10	
	StdDev	Time(s)	StdDev	Time(s)	StdDev	Time(s)
d_100_0	0.78	1309.00	0.34	272.48	0.36	130.58
d_100_1	0.00	1238.14	0.00	250.68	0.08	120.75
d_100_2	0.19	1620.41	3.03	324.83	0.47	160.31
d_20_0	0.00	353.96	0.00	70.44	0.00	37.45
d_20_1	0.00	257.32	0.00	52.75	0.00	28.36
d_20_2	0.00	338.29	0.00	66.79	0.00	35.69
d_40_0	0.01	727.18	0.01	140.99	0.00	72.91
d_40_1	0.81	797.76	3.77	156.73	3.89	78.69
d_40_2	0.00	712.09	0.01	135.63	0.01	70.12
d_60_0	0.04	869.91	0.06	165.23	0.29	86.93
d_60_1	0.02	627.06	0.00	122.83	0.00	64.89
d_60_2	11.79	927.55	6.44	178.07	6.28	92.18
d_80_0	0.01	1210.04	0.02	231.52	0.02	118.47
d_80_1	0.10	1260.54	0.14	236.02	0.02	120.37
d_80_2	0.05	1400.39	0.08	269.70	0.20	137.10



Hình 16. Độ hội tụ của GAMEP trên bộ dữ liệu S_40_1

4.4. Kết luận

Về mặt lý thuyết, luận văn đã trình bày được các nội dung sau:

- Tổng quan về bài toán bao phủ rào chắn trong mạng cảm biến không dây.
- Đề xuất cách tính giá trị exposure trong mạng cảm biến không dây mô hình xác suất, từ đó phát biểu bài toán tìm đường xâm nhập có giá trị exposure nhỏ nhất trong mạng cảm biến không dây sử dụng mô hình xác suất.
- Tổng quan về các giải thuật xấp xỉ giải quyết bài toán tối ưu.
- Đề xuất giải thuật di truyền giải bài toán tìm đường xâm nhập có giá trị exposure nhỏ nhất trong mạng cảm biến không dây sử dụng mô hình xác suất.

Về mặt thực nghiệm, luận văn đã thu được một số kết quả:

- Cài đặt thành công giải thuật di truyền đề xuất
- Thực nghiệm giải thuật đề xuất với các tham số và đưa ra đánh giá, nhận xét

Tuy nhiên, do còn nhiều hạn chế về kiến thức, kinh nghiệm của bản thân và thời gian thực hiện, luận văn còn một số thiếu sót sau:

- Chưa áp dụng các kỹ thuật heuristic để tăng hiệu quả cho giải thuật đề xuất
- Chưa cài đặt, thực nghiệm các giải thuật xấp xỉ khác để so sánh, đánh giá tổng quan hơn.

DANH MỤC TÀI LIỆU THAM KHẢO

- [1] Ahmed, Nadeem, Salil S. Kanhere, Sanjay Jha, *Probabilistic coverage in wireless sensor networks*, Local Computer Networks, The IEEE Conference on, IEEE, 2005.
- [2] Alexander Souza, *Combinatorial Algorithms*, Lecture notes in Winter Term 10/11, Humboldt University Berlin, 2011.
- [3] Anju Sangwan, Rishi Pal Singh, *Survey on coverage problems in wireless sensor networks*, Wireless Personal Communications 80.4, pp. 1475-1500, 2015.
- [4] Balister, Paul, et al, *Reliable density estimates for coverage and connectivity in thin strips of finite length*, Proceedings of the 13th annual ACM international conference on Mobile computing and networking, ACM, pp. 75-86, 2007.
- [5] Bang Wang, *Coverage problems in sensor networks: A survey*, ACM Computing Surveys (CSUR) 43.4, pp. 32-85, 2011.
- [6] Chiani, Marco, Davide Dardari, *Improved exponential bounds and approximation for the Q-function with application to average error probability computation*, Global Telecommunications Conference, GLOBECOM'02. IEEE, pp. 1399-1402, 2002.
- [7] Colin R. Reeves, Jonathan E. Rowe, *Genetic algorithms-principles and perspectives*, Kluwer Academic Publishers, 2003.
- [8] Dimitris Bertsimas, John Tsitsiklis, *Simulated annealing*, Statistical science 8.1, pp. 10-15, 1993.
- [9] Dorigo, Marco, Mauro Birattari, Thomas Stutzle, *Ant colony optimization*, IEEE computational intelligence magazine 1.4, pp. 28-39, 2006.
- [10] Elmorsy, Mohammed, Ehab S. Elmallah, Hosam MF AboElFotoh, *On path exposure in probabilistic wireless sensor networks*, Local Computer Networks (LCN), 2013 IEEE 38th Conference, IEEE, 2013.
- [11] Glover, Fred, *Tabu search: A tutorial*, Interfaces 20.4, pp. 74-94, 1990.
- [12] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, *Wireless sensor networks: a survey*, Computer networks 38.4, pp. 393-422, 2002.
- [13] Kennedy, James, *Particle swarm optimization*, Encyclopedia of machine learning, Springer US, pp. 760-766, 2011.
- [14] Kumar, Santosh, Ten H. Lai, Anish Arora, *Barrier coverage with wireless sensors*, Proceedings of the 11th annual international conference on Mobile computing and networking, ACM, pp. 284-298, 2005.

- [15] Liu, Benyuan, et al, *Strong barrier coverage of wireless sensor networks*, Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing, ACM, pp. 411-420, 2008.
- [16] Megerian, Seapahn, et al., *Exposure in wireless sensor networks: Theory and practical solutions*, Wireless Networks 8.5, pp. 443-454, 2002.
- [17] Miao, Ye, Yuping Wang, Cai Dai, Wei Jing–Xuan, *A hybrid genetic algorithm for the minimum exposure path problem of wireless sensor networks based on a numerical functional extreme model*, IEEE Transactions on Vehicular Technology 65(10), pp. 8644-8657, 2016.
- [18] Meguerdichian, Seapahn, et al., *Exposure in wireless ad-hoc sensor networks*, in Proceedings of the 7th annual international conference on Mobile computing and networking, ACM, 2001.
- [19] Rafal Kicinger, Tomasz Arciszewski, Kenneth De Jong, *Evolutionary Computation and Structural Design: a Survey of the State of the Art*, Computers & Structures, vol. 83, pp. 23-24, 1943 – 1978.
- [20] S. Meguerdichian, F. Koushanfar, M. Potkonjak, and M. Srivastava, *Coverage problems in wireless ad-hoc sensor networks*, in Proceedings of INFOCOM, pp. 1380–1387, April 2001.
- [21] Saipulla, Anwar, et al, *Barrier coverage of line-based deployed wireless sensor networks*, INFOCOM 2009, IEEE, pp. 127-135, 2009.
- [22] Seapahn Megerian, Farinaz Koushanfar, Miodrag Potkonjak, and Mani B. Srivastava. *Worst and Best-Case Coverage in Sensor Networks*, IEEE Transactions on Mobile Computing, Vol. 4, No. 1, pp. 84-92, Jan/ Feb 2005.
- [23] T. Clouqueur, V. Phipatanasuphorn, P. Ramanathan, and K. K. Saluja. *Sensor deployment strategy for detection of targets traversing a region*, Mobile Network. vol. 8, no. 4, pp. 453–461, Apr 2003.
- [24] Thomas Weise, *Global Optimization Algorithms – Theory and Application*, 2nd edition, 2009.
- [25] Wang, Duoqiang, Qi Qin, Jiying Yu, *The Algorithm of Coverage Base on Probability Sensing Model in Wireless Sensor Network*, Wireless Communications, Networking and Mobile Computing (WiCOM), 2012 8th International Conference on, IEEE, pp.1-4, 2012.
- [26] Đinh Thị Hà Ly, *Đồ án tốt nghiệp*, Viện Công nghệ thông tin và truyền thông, Đại học Bách Khoa Hà Nội, 2015.
- [27] Nguyễn Đình Thúc, *Lập trình tiến hóa*, NXB Giáo dục, 2009