

MỘT CẢI TIẾN PHÂN CỤM RFID ĐỘNG NHẪM LỌC DỮ LIỆU HIỆU QUẢ NĂNG LƯỢNG

Võ Viết Minh Nhật¹, Lê Văn Hòa¹, Huỳnh Quốc Phương², Nguyễn Văn Tùng³

¹Đại học Huế

²Khoa CNTT, Đại học An Giang

³Khoa CNTT, Đại học Công nghệ Thực phẩm – TP. Hồ Chí Minh

Tóm tắt: Lọc dữ liệu trong mô hình tích hợp nhận dạng đối tượng bằng sóng vô tuyến (RFID) với mạng cảm biến (SN) là một vấn đề thời sự đang thu hút nhiều sự quan tâm của các nhà nghiên cứu trên thế giới. Một trong những hướng tiếp cận về lọc dữ liệu hiệu quả năng lượng là dựa trên phân cụm trong đó các điểm lọc dữ liệu chỉ được thực hiện bởi các nút chủ cụm. Tuy nhiên, đa số các đề xuất đều xem xét trong môi trường mà ở đó các đầu đọc được giả sử là không di chuyển và vai trò chủ cụm là cố định tại một nút. Điều này làm cho các chủ cụm tiêu tốn quá nhiều năng lượng, mà kết quả là thời gian sống của chúng giảm nhanh. Bài viết này sẽ đề xuất một cải tiến về phân cụm động đối với các đầu đọc RFID trong đó việc phân cụm được thực hiện lại một cách định kỳ và vai trò chủ cụm được thay đổi một cách linh hoạt giữa các nút sao cho năng lượng được tiêu thụ được chia sẻ hợp lý giữa chúng và do đó làm tăng thời gian sống của toàn hệ thống.

Từ khóa: Tích hợp RFID với SN, lọc dữ liệu, hiệu quả năng lượng, phân cụm động.

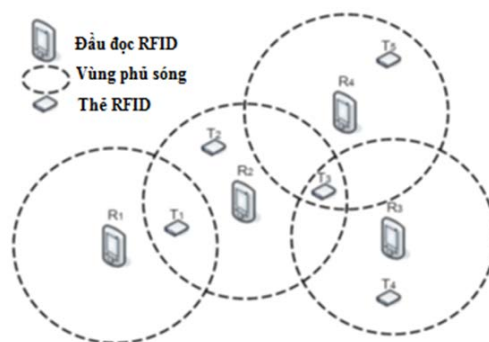
I. GIỚI THIỆU

Tích hợp công nghệ nhận dạng đối tượng theo tần số vô tuyến (*Radio Frequency Identification - RFID*) [1] với mạng cảm biến (*Sensor Networks - SN*) [2] đang là một xu thế hiện nay bởi nó có một phạm vi ứng dụng rộng rãi và đa dạng mà ở đó những ưu điểm của cả hai công nghệ được khai thác và sử dụng. Mô hình tích hợp này đã tạo ra một cơ sở hạ tầng tuyệt vời để xử lý và phân phối dữ liệu trong môi trường động, được phân cấp. Tuy nhiên, mô hình tích hợp cũng đối mặt với nhiều thách thức trong đó việc làm giảm dữ liệu dư thừa là hết sức phức tạp vì nó còn đi kèm với các yếu tố như độ trễ truyền thông, năng lượng tiêu thụ và lãng phí các loại tài nguyên khác.

Về cơ bản, mạng cảm biến là một mô hình mạng gồm nhiều nút sink hay còn được gọi là trạm cơ sở (base station) và nhiều nút cảm biến có kích thước bé, trọng lượng nhỏ. Các nút cảm biến có thể cảm nhận điều kiện môi trường như: nhiệt độ, độ ẩm, áp suất,

ánh sáng, âm thanh hay các rung động ... mà phù hợp cho việc thu thập thông tin [3]. Các nút cảm biến còn có khả năng tính toán và cho phép xử lý các thông tin thu thập được. Thông tin này sau đó được chuyển đến các trạm cơ sở. Mạng cảm biến cung cấp cơ chế giám sát chi phí hiệu quả cho các ứng dụng quan trọng, bao gồm các ứng dụng giám sát biên giới, hải đảo, điều khiển hoạt động trong các nhà máy công nghiệp, giám sát môi trường, quân sự và cả các ứng dụng về y tế, du lịch.

Với công nghệ RFID, nó cho phép phát hiện và nhận diện các đối tượng trong một môi trường. Một hệ thống RFID bao gồm các thiết bị (*reader*) đọc dữ liệu từ các thẻ (*tag*) như Hình 1. Một thẻ bao gồm một chip và một ăng ten được gắn trên một đối tượng mục tiêu cần đọc. Thông tin thu thập được bằng cách các thiết bị đọc quét qua các thẻ và sau đó truyền thông tin đọc được đến server ở trạm cơ sở. Các ứng dụng của RFID đã được phát triển khá nhiều trong thời gian gần đây như trong quản lý chuỗi cung ứng, thu phí đường cao tốc, quản lý giao thông, phát triển nhà thông minh... [4].



Hình 1. Mô hình phủ sóng chồng lấp của các đầu đọc RFID đối với các thẻ

Công nghệ RFID đã được chấp nhận trong nhiều ứng dụng công nghiệp, trong khi mạng cảm biến có thể phát hiện thông tin trong các điều kiện môi trường khác nghiệt. Tuy nhiên, cũng có nhiều ứng dụng mà thông tin thu thập từ môi trường là không đủ để xử lý;

do đó việc xác định thêm thông tin như vị trí của đối tượng là hết sức cần thiết [3]. Việc sử dụng mạng cảm biến cho các ứng dụng về môi trường, quản lý các điểm danh lam thắng cảnh... đang là một xu thế của du lịch thông minh. Trong trường hợp này mô hình tích hợp RFID với mạng cảm biến là giải pháp tối ưu, trong đó chúng vừa bổ sung, hỗ trợ cho nhau [4].

Tuy nhiên, mô hình tích hợp cũng đối mặt với nhiều thách thức khác nhau như: xác định hiệu suất thời gian thực, hiệu quả năng lượng, lọc dữ liệu dư thừa, chống va chạm (*anti-collision*) và hiệu quả xác thực [5]. Trong số các thách thức trên thì vấn đề lọc dữ liệu để vừa làm sạch dữ liệu và vừa hiệu quả năng lượng là một vấn đề quan trọng nhằm sử dụng hiệu quả các nguồn tài nguyên mạng, giảm tiêu thụ năng lượng [5]. Trong một hệ thống RFID, thiết bị đọc thường xuyên kiểm tra các thẻ nhiều lần để tăng tốc độ đọc. Điều này đã tạo ra nhiều bản sao về một đối tượng duy nhất, mà dẫn đến thừa dữ liệu tại các đầu đọc cùng đọc một thẻ. Như được chỉ ra trong Hình 1, vấn đề trùng lặp dữ liệu sẽ xuất hiện ở các đầu đọc R_2 , R_3 và R_4 vì cùng đọc thẻ T_3 . Thực tế việc dư thừa dữ liệu do 2 nguyên nhân chính: (1) một thẻ được bao phủ bởi nhiều đầu đọc (như thẻ T_3) và (2) đầu đọc đọc thẻ nhiều lần nên tạo ra nhiều các bản sao không cần thiết. Việc loại bỏ các dư thừa này là cần thiết vì nó không đem lại bất kỳ một thông tin hữu ích nào.

Việc loại bỏ dư thừa góp phần sử dụng hiệu quả các nguồn tài nguyên hơn. Quá trình lọc và loại bỏ các thông tin dư thừa, được gọi là quá trình làm sạch dữ liệu (*data cleaning*). Cụ thể, loại bỏ dữ liệu dư thừa là một quá trình thay thế, sửa đổi hoặc xóa những phần không liên quan, không chính xác hoặc không chính xác một phần. Hầu hết các vấn đề loại bỏ dữ liệu dư thừa đều tập trung vào phương pháp phân cụm. Việc phân cụm sẽ hạn chế các điểm lọc, do chỉ có nút chủ cụm mới chịu trách nhiệm lọc; do đó tiết kiệm được năng lượng trong quá trình lọc. Tuy nhiên do nút chủ cụm lọc dữ liệu nên nó sẽ phải tiêu tốn năng lượng lớn hơn; kết quả là có thời gian sống ít hơn. Hơn nữa, việc phân cụm chưa được xem xét trong môi trường mà các nút di chuyển tự do. Bài báo này sẽ đề xuất một giải pháp phân cụm động các đầu đọc RFID nhằm nâng cao hiệu quả về năng lượng đối với việc lọc dữ liệu và do đó tăng thời gian sống của toàn hệ thống.

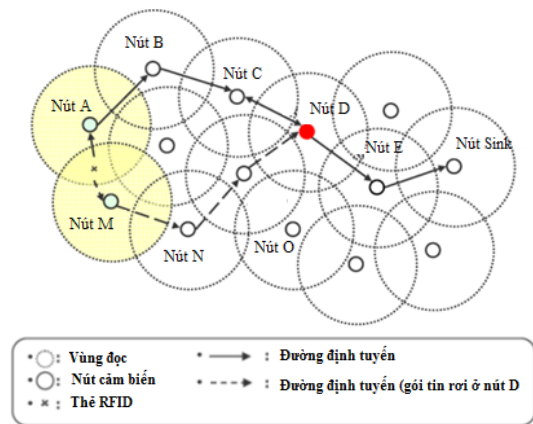
Các phần tiếp theo của bài báo được tổ chức như sau: Phần 2 tóm lược và phân tích các công trình nghiên cứu liên quan. Trên cơ sở các đánh giá, Phần 3 trình bày mô hình lọc dữ liệu hiệu quả năng lượng được đề xuất. Cài đặt mô phỏng và phân tích kết quả sẽ được mô tả ở Phần 4. Cuối cùng kết luận ở Phần 5.

II. CÁC ĐỀ XUẤT VỀ LÀM SẠCH DỮ LIỆU

Lọc dữ liệu là một vấn đề quan trọng trong mạng cảm biến không dây tích hợp với RFID. Các ứng dụng dựa trên mô hình tích hợp này thường chỉ quan tâm đến một bản dữ liệu duy nhất, nhưng việc trùng lặp dữ liệu trong khi đọc đã tạo ra nhiều các bản sao không mong muốn.

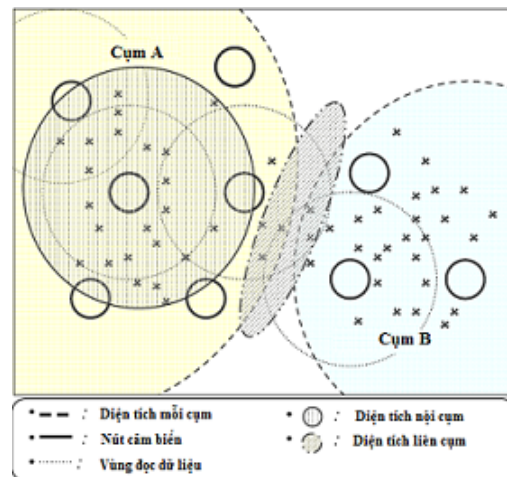
Wonil và cộng sự trong [6] đã đề xuất kỹ thuật INPFM (*In-Network Phased Filtering Mechanism*) trong đó dữ liệu chỉ được lọc ở nút thứ k vì họ cho rằng lọc dữ liệu tại tất cả các nút sẽ gây chậm trễ trên

toàn hệ thống. Cách tiếp cận này được thể hiện dưới dạng cấu trúc cây theo nguyên tắc định tuyến đa chặng (*multi-hops*), trong đó các nút cha sẽ đóng vai trò nút lọc trong khi các nút con phát hiện sự trùng lặp dữ liệu. Như được chỉ ra trong Hình 2, nút A và M cùng đọc dữ liệu ở vùng chồng lấp "x" và sau đó truyền dữ liệu đến trạm cơ sở qua nhiều chặng. Trong [6], việc lọc dữ liệu được đề xuất ở khoảng cách k chặng (trong Hình 2 thì $k = 3$) và x được truyền theo 2 hành trình được định tuyến khác nhau để đến nút D. Nút D lúc này đóng vai trò nút lọc dữ liệu và sẽ loại bỏ bớt một bản sao trước khi chúng được gửi đến trạm cơ sở.



Hình 2. Lọc dữ liệu theo phương pháp INPFM [6]

Trong [7], Kim và cộng sự đã đề xuất phương pháp CLIF (*Cluster-based In-network phase Filtering scheme*) dựa trên phân cụm và việc lọc dữ liệu được xây ra tại nút chủ cụm (*Cluster Head*). Cụ thể, các nút gần nhau được gom thành một cụm và một nút được chọn để đóng vai trò chủ cụm. Nút chủ cụm sẽ chịu trách nhiệm lọc dữ liệu cho cụm. Như được chỉ ra trong Hình 3, có 2 cụm A và B. Dữ liệu thuộc cụm A sẽ được lọc bởi nút chủ cụm A, nhưng dữ liệu nằm trong vùng chồng lấn của 2 cụm A và B sẽ được lọc bởi một nút chủ cụm trung gian. Nút chủ cụm này sẽ phát hiện sự trùng lặp dữ liệu (tức là nhận được từ 2 bản sao trở lên).



Hình 3. Lọc dữ liệu theo phương pháp CLIF [7]

Bashir và cộng sự trong [8] đã đề xuất sơ đồ EIFS (*Energy efficient In-network RFID data Filtering Scheme*), trong đó trùng lặp dữ liệu cũng được chia

thành hai loại là nội cụm và liên cụm. Việc lọc đối với hai loại này cũng được thực hiện tách biệt tương tự như trong CLIF. Tuy nhiên, sau khi nhận được dữ liệu, chủ cụm sẽ xác định loại trùng lặp dựa trên giá trị trường f được lưu trong cấu trúc của gói dữ liệu. Nếu giá trị của trường f là 1, nút gửi được xác định là nút nội cụm và chủ cụm sẽ thực hiện việc lọc dữ liệu. Sau khi việc lọc dữ liệu đã thực hiện xong, giá trị trường f được thiết lập bằng 0. Như vậy, chủ cụm sẽ không lọc các gói tin có $f = 0$ và do đó làm giảm đáng kể chi phí tính toán. Sau bước lọc dữ liệu nội cụm, các chủ cụm gửi dữ liệu của chúng về phía trạm cơ sở. Trong trường hợp lọc dữ liệu liên cụm, EIFS đầu tiên sẽ tìm và phát hiện dữ liệu trùng lặp liên cụm. Nếu có dữ liệu trùng lặp được phát hiện, các chủ cụm trung gian sẽ gửi một thông tin phản hồi để thông báo cho các chủ cụm nơi sinh ra các gói dữ liệu trùng lặp tránh việc truyền gói không cần thiết.

Bashir và cộng sự trong [9] tiếp tục mở rộng EIFS thành giải thuật có tên gọi là IRDF (*In-network RFID Duplicate data Filtering*), trong đó việc lọc nội cụm được tiến hành với phương pháp EIFS, nhưng việc lọc ngoại cụm được tiến hành tại những cụm lân cận, thay vì tại nút chủ cụm trung gian như trong EIFS. Một khác biệt khác của IRDF là loại bỏ cơ chế phản hồi thông tin vì nó cho rằng việc làm này làm gia tăng độ trễ của quá trình truyền dữ liệu.

Tóm lại, các phương pháp nêu trên đã loại bỏ được đa số dữ liệu dư thừa trước khi được truyền đến trạm cơ sở. Tuy nhiên vẫn tồn tại 4 vấn đề sau: (1) nút chủ cụm phải chịu hao tổn năng lượng đáng kể vì lọc dữ liệu; (2) nếu nút chủ cụm không nằm trên tuyến đường được định tuyến đến trạm cơ sở, các nút có dữ liệu trùng lặp phải chuyển hướng đến nút này; điều này làm tăng độ dài hành trình và dẫn đến chậm trễ trong việc lọc dữ liệu; (3) khi dữ liệu đến nút chủ cụm lớn, nó có khả năng rơi vào tình trạng quá tải; (4) các đầu đọc và thẻ được giả thiết là cố định, trường hợp chúng di chuyển chưa được xem xét đến. Đề xuất sau đây của chúng tôi sẽ giải quyết 4 vấn đề này.

III. MÔ HÌNH LỌC DỮ LIỆU HIỆU QUẢ NĂNG LƯỢNG

A. Giới thiệu mô hình

Trong mô hình được đề xuất của chúng tôi, việc trùng lặp dữ liệu cũng được chia thành 2 loại: nội cụm và liên cụm. Việc lọc đối với hai loại này cũng được thực hiện tách biệt như trong IRDF. Tuy nhiên, chúng tôi xem xét trường hợp các đầu đọc và các thẻ di chuyển. Do đó, việc phân cụm các đầu đọc được thực hiện một cách động, mà chi tiết về thuật toán phân cụm này sẽ trình bày trong Mục III.B. Khi tiến hành phân cụm, nút chủ cụm được xác định một cách động, nghĩa nó sẽ được thay đổi luân phiên theo 2 tiêu chí: (1) năng lượng hiện tại và (2) xem xét trên hành trình đến trạm cơ sở. Việc thay đổi nút chủ cụm như vậy sẽ giúp cho các bộ đọc chia sẻ năng lượng bị tiêu hao. Hơn nữa, việc ưu tiên chọn nút chủ cụm nằm trên hành trình đến trạm cơ sở sẽ giúp rút ngắn hành trình truyền tải dữ liệu. Chi tiết của giải thuật xác định nút chủ cụm linh động này sẽ được trình bày trong Mục III.C. Mô hình lọc dữ liệu cải tiến của chúng tôi có tên

gọi là DCDF (*Dynamic Clustering-based in-network Data Filtering*) và được trình bày trong Mục III.D.

B. Phương pháp phân cụm các đầu đọc di chuyển

Giải thuật phân cụm được chúng tôi đề xuất có tên gọi là CMR (*Clustering Moving Readers*) dựa trên ý tưởng như sau. Bước 1, các đầu đọc được phân cụm bằng phương pháp K-mean [10]; một danh sách các đầu đọc di chuyển được lưu lại sau từng khoảng thời gian (tương tự như vấn đề đọc dữ liệu của đầu đọc). Bước 2, khoảng cách từ mỗi đầu đọc di chuyển đến các tâm cụm được tính toán; một đầu đọc sẽ được phân vào một cụm mới nếu khoảng cách từ nó đến tâm cụm mới là bé nhất. Cụ thể, 2 bước của phương pháp CMR là như sau:

Bước 1: Xác định danh sách các đầu đọc di chuyển

Với mỗi đầu đọc, tọa độ của nó được duy trì bởi một vector $R_i(x_i, y_i)$, $i = 1..N$ trong đó N là số lượng các đầu đọc. Với một mạng cảm biến được triển khai, vị trí của các đầu đọc là được xác định một cách dễ dàng. Dựa trên các tọa độ này, các đầu đọc được phân cụm dựa trên giải thuật K-mean. Mỗi khi có thay đổi vị trí (x_i, y_i) sau từng khoảng thời gian cố định, đầu đọc R_i được đưa vào một danh sách cần phân cụm lại (như được mô tả từ dòng 4 đến 10 trong giải thuật CMR).

Bước 2: Phân bố đầu đọc di chuyển vào cụm mới

Với mỗi đầu đọc R_i nằm trong danh sách cần phân cụm lại, khoảng cách Eucliden từ nó đến các tâm cụm được tính toán lại. Đặt $D(i, j)$ là khoảng cách từ R_i đến tâm cụm j . R_i được phân vào cụm j nếu $D(i, j)$ là bé nhất (như được mô tả từ dòng 13 đến 25 trong giải thuật CMR). Lưu ý rằng tâm cụm j thay đổi một cách động và giải thuật xác định tâm cụm động được mô tả trong Mục III.C.

Sau đây là mô tả chi tiết của giải thuật CMR:

Giải thuật CMR (*Clustering Moving Readers*)

Input:

- danh sách các đầu đọc đã phân cụm $C = \{C_j | j=1..K\}$, $C_j = \{R_i | i=1..m\}$ và $0 < m < N$, với N là số lượng đầu đọc;
- tâm cụm c_j , $j = 1..K$. (được xác định ở giải thuật CCR)

Output:

- danh sách các cụm sau khi phân cụm lại C_j , $j = 1..K$;

Process:

```

1  i ← 1;
2  r ← 0;
3  list_change ← ∅; // danh sách các đầu đọc di chuyển
4  while (i ≤ N) do
5      // khi Ri có sự thay đổi vị trí xi hoặc yi
6      if (change(xi) or change(yi)) then
7          // bổ sung Ri vào danh sách đầu đọc di chuyển
8          list_change ← Ri;
9          r++; // số lượng đầu đọc trong danh sách
10     end if
11     i++;
12 end while
13 t ← 1;
14 j ← 1;
15 while (t ≤ r) do
16     min ← 0;
```

```

15  temp ← ∞ ;
16  while (j ≤ K) do
17      if (eucliden(Ri, ccj) < temp) then
18          temp ← eucliden(Ri, ccj);
19          // lưu vị trí tâm cụm có khoảng cách bé nhất
20          min ← j;
21      end if
22      j++;
23  end while
24  // phân bổ Ri vào cụm có khoảng cách bé nhất
25  Cmin ← Ri;
26  t++;
27 end while

```

Độ phức tạp của giải thuật là $O(r \times K)$ với r là số lượng các đầu đọc di chuyển và K số cụm trong hệ thống.

C. Phương pháp xác định lại tâm cụm

Việc xác định tâm cụm là quan trọng vì nó phải chịu trách nhiệm lọc dữ liệu. Phương pháp xác định lại tâm cụm mà chúng tôi đề xuất, có tên gọi là CCR (*Cluster Center Recomputing*), dựa trên 2 tiêu chí: (1) năng lượng hiện tại của nút trong cụm; (2) nằm trên hành trình định tuyến đến trạm cơ sở. Quá trình xác định lại tâm cụm được chia thành 2 bước:

Bước 1: Xác định danh sách các đầu đọc tiềm năng

Năng lượng hiện tại của các đầu đọc trong một cụm được so sánh để xác định một danh sách các đầu đọc tiềm năng làm chủ cụm (từ dòng 12 đến dòng 18 trong giải thuật CCR). Danh sách này gồm các đầu đọc có mức năng lượng cao nhất và có độ lệch không vượt quá một giá trị ΔE được xác định trước. ΔE được gọi là khoảng chênh lệch năng lượng tiềm năng giữa các nút; nó cần đủ nhỏ để không gây chênh lệch quá lớn giữa các mức năng lượng trong danh sách.

Bước 2: Xác định lại tâm cụm

Các đầu đọc tiềm năng trong danh sách được xem xét về khả năng định tuyến đến trạm cơ sở, trong đó đầu đọc được chọn là đầu đọc đi qua ít nút trung gian nhất để đến trạm cơ sở (như được mô tả từ dòng 22 đến dòng 29 trong giải thuật CCR).

Giải thuật xác định lại tâm cụm CCR được mô tả chi tiết như sau:

Giải thuật CCR (*Cluster Center Recomputing*)

Input:

- danh sách các đầu đọc trong một cụm $R_i(x_i, y_i)$, $i = 1..n$;
- năng lượng các đầu đọc E_i , $i = 1..n$;
- độ lệch năng lượng ΔE .

Output:

- tâm cụm cc .

Process:

```

1  i ← 1;
2  list_energy ← ∅;
3  max_energy ← ∅;
4  while (i ≤ n) do
5      if (Ei > max_energy) then
6          // xác định mức năng lượng cao nhất trong cụm
          max_energy ← Ei;

```

```

7      end if
8      i++;
9  end while
10 i ← 1;
11 t ← 0;
12 while (i ≤ n) do
13     // kiểm tra mức năng lượng của các đầu đọc
14     if (Ei > max_energy - ΔE) then
15         // DS các đầu đọc có khả năng làm chủ cụm
16         list_energy ← Ri;
17         // số lượng đầu đọc có khả năng làm chủ cụm
18         t++;
19     end if
20     i++;
21 end while
22 min ← ∞;
23 i ← 1;
24 temp ← ∅;
25 while (i ≤ t) do
26     // so sánh số nút trung gian của các nút có khả năng
27     // làm chủ cụm, trong đó count(Ri) số nút trung gian
28     // để đến được trạm cơ sở của đầu đọc Ri
29     if (count(Ri) < min) then
30         temp ← Ri;
31         min ← count(Ri);
32     end if
33     i++;
34 end while
35 cc ← temp; // xác định tâm cụm

```

Độ phức tạp của giải thuật CCR là $O(n)$ với n là số các đầu đọc trong một cụm.

D. Mô hình lọc dữ liệu hiệu quả năng lượng

Đầu tiên chúng tôi sử dụng phương pháp phân cụm K-mean (từ dòng 2 đến 6 trong giải thuật DCDF) để phân bổ các đầu đọc tương ứng vào từng các cụm. Sau từng khoảng thời gian cố định chúng tôi sử dụng giải thuật CCR dòng 8 để xác định lại các tâm cụm và sử dụng giải thuật CMR dòng 9 để phân các đầu đọc di chuyển vào các cụm mới.

Giải thuật DCDF được mô tả chi tiết như sau:

Giải thuật DCDF (*Dynamic Clustering-based in-network Data Filtering*)

Input:

- danh sách các đầu đọc $R_i(x_i, y_i)$, $i = 1..N$;
- số cụm K ; khoảng thời gian xác định tâm cụm t ; thời gian kết thúc mô phỏng t_{end} ;
- năng lượng các đầu đọc E_i , $i = 1..N$;
- độ lệch năng lượng ΔE .

Output:

- năng lượng trung bình của các nút chủ cụm HE

Process:

```

1  HE ← 0;
2  Khởi tạo K cụm {ccj; j = 1..K};
3  // nếu có sự thay đổi giá trị tâm
4  while (change(ccj))
5     // Cj tập các đầu đọc trong cụm j, j*≠j và j*=1..K
6     Cj ← {Ri | eucliden(Ri, ccj) ≤ eucliden(Ri, ccj*)};
7     // xác định lại tâm cụm theo K-mean và
8     // average(Ri) giá trị trung bình của các đầu đọc
9     // trong cụm j, tâm cụm là nút gần với giá trị trung

```

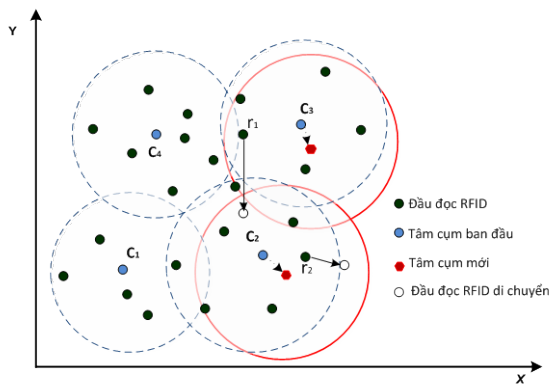
```

    bình nhất.
     $cc_j \leftarrow \text{average}(R_i | R_i \in C_j)$ ;
6   end while
7   while ( $t < t_{end}$ )
8     call(CCR); // gọi giải thuật CCR
9     call(CMR); // gọi giải thuật CMR
10  end while
    // xác định năng lượng trung bình của các tâm cụm
11   $HE \leftarrow \text{average}(cc_j)$ ;

```

E. Ví dụ minh họa

Với việc xác định chủ cụm một cách động, phương pháp DCDF đảm bảo thích ứng được với những di chuyển ngẫu nhiên của các đầu đọc, giúp việc chọn chủ cụm hợp lý và cân bằng năng lượng tiêu thụ của các đầu đọc trong mạng. Hơn nữa, phương pháp DCDF cũng giúp việc truyền dữ liệu nhanh hơn vì tuyến đường được chọn đi qua ít nút trung gian nhất.



Hình 4. Một ví dụ về phân cụm lại khi các đầu đọc di chuyển (với số cụm $K=4$)

Để làm rõ hơn vấn đề này hãy xem xét một ví dụ như trong Hình 4, trong đó các đầu đọc được phân cụm theo thuật toán K-mean. Có 4 cụm được hình thành với các tâm lần lượt là C_1 , C_2 , C_3 và C_4 (những đường tròn đứt nét). Sau từng khoảng thời gian cố định, một số tâm cụm được xác định lại (theo giải thuật CCR), như trong Hình 4 là C_2 và C_3 (được thể hiện bằng các đường tròn liền nét). Khi các đầu đọc di chuyển chúng được đưa vào một danh sách cần phân cụm lại (theo giải thuật CMR) sau từng khoảng thời gian xác định. Các đầu đọc vẫn có thể thuộc cụm ban đầu (chẳng hạn r_2) nhưng cũng có thể chuyển sang cụm mới (chẳng hạn r_1).

IV. MÔ PHÒNG VÀ PHÂN TÍCH KẾT QUẢ

Chúng tôi tiến hành cài đặt mô phỏng trên máy tính 2.4 GHz Intel Core 2 CPU, 2G RAM. Các tham số mô phỏng được mô tả trong Bảng I, với 101 đầu đọc trong có 1 đầu đọc đóng vai trò là trạm cơ sở. Các đầu đọc di chuyển ngẫu nhiên trong vùng bán kính là $100 \times 100 \text{m}^2$; mức năng lượng ban đầu của các đầu đọc là 2KJ (Kilo Jun); số lượng các cụm theo K-mean thay đổi từ 5, 7, 9 cụm. Định kỳ 1s chúng tôi tiến hành xác định lại tâm cụm theo giải thuật CCR và phân cụm lại các đầu đọc theo giải thuật CMR.

Bảng I. Các tham số mô phỏng

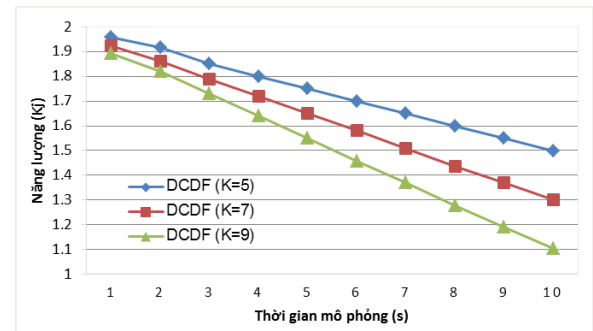
Tham số	Giá trị
Vùng mô phỏng	$100 \times 100 \text{m}^2$
Số lượng đầu đọc	101
Số lượng cụm	5, 7, 9
Năng lượng ban đầu của đầu đọc	2KJ
Độ lệch năng lượng ΔE	0.1KJ
Thời gian xác định phân cụm lại	1s
Thời gian mô phỏng	10s

Các mục tiêu mô phỏng bao gồm:

- Xem xét ảnh hưởng của việc thay đổi số cụm K đến năng lượng trung bình của các đầu đọc;
- So sánh hiệu quả năng lượng của giải thuật DCDF so với IRDF.

A. Ảnh hưởng của số cụm đến năng lượng trung bình của các đầu đọc

Chúng tôi tiến hành thay đổi số cụm K từ 5 đến 7 và 9 để xem xét ảnh hưởng của số cụm đến mức tiêu hao năng lượng trung bình của các đầu đọc trong cụm. Kết quả thu được như Hình 5, ở đó chúng ta nhận thấy rằng năng lượng trung bình của các đầu đọc giảm khi số cụm tăng. Điều này có được là do khi tăng số cụm thì số điểm lọc dữ liệu và số điểm truyền dữ liệu tăng lên dẫn đến việc mất mát nhiều dữ liệu hơn.



Hình 5. Mức tiêu hao năng lượng trung bình của các đầu đọc khi thay đổi số cụm

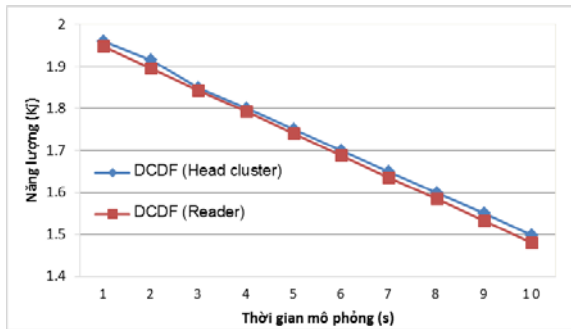
Để xem xét ảnh hưởng của số cụm đến mức năng lượng của các nút chủ cụm, chúng tôi tiến hành so sánh năng lượng trung bình của các nút chủ cụm khi số cụm thay đổi; Kết quả thu được như Bảng II.

Bảng II. Mức tiêu hao năng lượng trung bình của các nút chủ cụm khi số cụm thay đổi (đơn vị KJ)

Thời gian (s)	2	4	6	8	10
DCDF ($K=5$)	1.916	1.800	1.700	1.600	1.498
DCDF ($K=7$)	1.860	1.720	1.580	1.434	1.303
DCDF ($K=9$)	1.820	1.640	1.456	1.276	1.105

Rõ ràng, việc tăng số cụm không chỉ ảnh hưởng đến mức tiêu hao năng lượng của các đầu đọc trong cụm mà còn gây ảnh hưởng đến các nút chủ cụm. Do trong chính sách của chúng tôi luôn sử dụng nút chủ

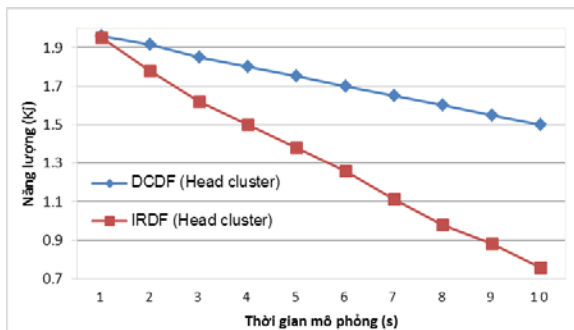
cụm thay đổi nhằm bảo toàn năng lượng nên nút chủ cụm luôn là nút có mức năng lượng cao hơn so với mức năng lượng trung bình của các đầu đọc trong cụm như chỉ ra ở Hình 6.



Hình 6. So sánh năng lượng trung bình nút chủ cụm và các nút trong cụm khi $K=5$

B. So sánh hiệu quả năng lượng giữa DCDF và IRDF

Chúng tôi tiếp tục so sánh năng lượng trung bình của các nút chủ cụm khi $K=5$ giữa giải thuật DCDF và IRDF; Kết quả thu được như Hình 7.



Hình 7. So sánh năng lượng trung bình của các nút chủ cụm khi $K=5$ giữa DCDF và IRDF

Từ Hình 7 chúng ta nhận thấy rằng giải thuật IRDF tiêu hao năng lượng diễn ra nhanh hơn, điều này là do giải thuật IRDF không luân phiên sử dụng các nút chủ cụm, làm cho nút chủ cụm tiêu tốn năng lượng nhiều hơn. Chẳng hạn, tại thời điểm 2s IRDF giảm năng lượng trung bình khoảng 7.8% so với DCDF, trong khi tại 10s việc giảm năng lượng này là 50%. Với trường hợp tăng số cụm lên 7 và 9, các kết quả Hình 5 và Hình 7 có thể giúp suy ra rằng IRDF sẽ làm giảm nhanh chóng năng lượng trung bình và thời gian sống của toàn hệ thống sẽ bị ảnh hưởng nghiêm trọng.

C. Nhận xét

Do giải thuật DCDF sử dụng phương pháp lọc dữ liệu như IRDF nên lượng dữ liệu truyền đến trạm cơ sở không có nhiều thay đổi. Tuy nhiên, dựa trên kết quả mô phỏng, DCDF là hiệu quả trong việc bảo toàn năng lượng, khi tiến hành thay đổi động các nút chủ cụm. Hơn nữa, với chính sách phân cụm được đề xuất đối với các đầu đọc di chuyển, DCDF bảo đảm được hiệu quả năng lượng như được chỉ ra trong Bảng III. Việc triển khai giải thuật DCDF là khả thi vì các điều kiện xem xét phù hợp với điều kiện thực tế là các đầu đọc và thẻ có thể di chuyển.

Bảng III. Phân bố các đầu đọc trong mỗi cụm theo thời gian, khi các đầu đọc và thẻ di chuyển (trường hợp $K=5$)

Thời gian (s)	2	4	6	8	10
Cụm 1	16	11	19	17	23
Cụm 2	19	34	15	47	34
Cụm 3	25	21	32	6	12
Cụm 4	14	10	9	19	19
Cụm 5	26	24	25	11	12

V. KẾT LUẬN

Trong bài báo này chúng tôi đã đề xuất một phương pháp phân cụm RFID động nhằm lọc dữ liệu hiệu quả năng lượng có tên là DCDF. Giải thuật DCDF bao gồm 2 giải thuật chính: (1) CMR nhằm phân bổ các đầu đọc di chuyển vào các cụm mới và (2) CCR nhằm tính toán lại nút đóng vai trò là chủ cụm sau từng khoảng thời gian xác định. Kết quả mô phỏng trong Mục 4 đã chỉ ra rằng phương pháp mà chúng tôi đề xuất cho hiệu quả năng lượng tốt hơn IRDF. Đồng thời DCDF cũng phù hợp với yêu cầu thực tế của các mô hình tích hợp RFID với mạng cảm biến mà ở đó các đầu đọc và các thẻ có thể tự do di chuyển.

TÀI LIỆU THAM KHẢO

- [1] A. S. Korotkov. "Radio frequency identification systems: Survey," *Radio electronics and Communications Systems*, vol. 59, no. 3, pp. 97-108, 2016.
- [2] Kamal D. Singh, Hakima Chaouchi, Jean M. Bonnin. "Wireless sensor networks: a survey on recent developments and potential synergies," *The Journal of Super computing*, vol. 68, no. 1, pp. 1-48, 2014.
- [3] Fadi M. Al-Turjman, Ashraf E. Al-Fagih1 and Hossam S. Hassanein, "A Novel Cost-Effective Architecture and Deployment Strategy for Integrated RFID and WSN Systems", *In: International Conference on Computing, Networking and Communications, Network Architecture P2P Protocol Symposium*, 2012.
- [4] Hai Liu, Miodrag Bolic, Amiya Nayak and Ivan Stojmenovic, "Taxonomy and Challenges of the Integration of RFID and Wireless Sensor Networks," *IEEE Network*, vol. 22, no.6, pp.26-35, November-December 2008.
- [5] Li Wang, Li Da Xu. "Data Cleaning for RFID and WSN Integration," *IEEE Transactions On Industrial Informatics*, vol. 10, no. 1, February 2014.
- [6] Choi. W., Park. M.S. "In-network Phased Filtering Mechanism for a Large-Scale RFID Inventory Application," *In: Proceedings of the 4th International Conference on IT & Applications (ICITA)*, Harbin China (January 2007) pp. 401-405.
- [7] Kim. D.S., Kashif. A., Ming. X., Kim. J.H., Park. M.S. "Energy Efficient In-Network Phase RFID Data Filtering Scheme," *In: Proceedings of the 5th International Conference on Ubiquitous Intelligence and Computing, UIC 2008*, Oslo, Norway (23-25 June 2008) pp. 311-322.
- [8] Ali Kashif Bashir, Se-Jung Lim, Chauhdary Sajjad Hussain and Myong-Soon Park. "Energy Efficient In-network RFID Data Filtering Scheme in Wireless Sensor Networks", *Sensors*, vol. 11, pp.7004-7021,

- 2011.
- [9] Ali Kashif Bashir, Se-Jung Lim, Chauhdary Sajjad Hussain and Myong-Soon Park. "In-network RFID Data Filtering Scheme in RFID-WSN for RFID Applications", *In: International Conference on Intelligent Robotics and Applications (ICIRA): Intelligent Robotics and Applications (2013)* pp 454-465.
- [10] G. Sandhiya, and Ramya Jothikumar. "Enhanced K-means with dijkstra algorithm for energy consumption in wireless sensor network," *In: Intelligent Systems and Control (ISCO), 10th International Conference on*, 7-8 Jan. 2016.



Nguyễn Văn Tùng, nhận học vị thạc sỹ năm 2015; Hiện đang công tác tại Khoa CNTT, trường Đại học Công Nghiệp Thực Phẩm, Thành phố Hồ Chí Minh. Lĩnh vực nghiên cứu: Mạng truyền dẫn quang (OBS), mạng cảm biến tích hợp với RFID.

Email: tungnv@cntp.edu.vn

AN IMPROVEMENT OF DYNAMIC RFID CLUSTERING FOR ENERGY-EFFICIENT DATA FILTERING

Abstract: Data filtering in the model of integrated RFID and sensor networks is a current issue that is drawing much interest from researchers around the world. One of the approaches of energy-efficient data filtering is based on clustering in which data filter points are only performed by cluster nodes. However, most of the proposals are studied in an environment where readers are assumed to be non-moving and the clustering role is fixed at clusters. This causes clusters to consume too much energy, which results in a dramatic reduction in their living time. This article will propose an improvement in dynamic clustering for RFID readers in which clustering is replayed periodically and clustering roles are flexibly changed between readers so that the consumed energy is appropriately shared among them and thus increases the lifetime of the system.



Võ Viết Minh Nhật, nhận học hàm Phó Giáo sư năm 2016, học vị Tiến sỹ năm 2007 tại đại học Quebec Canada; Hiện công tác tại Đại học Huế. Lĩnh vực nghiên cứu bao gồm: mạng truyền dẫn quang (OBS), mạng cảm biến tích hợp với RFID, tính toán mềm (mạng nơ ron nhân tạo, giải thuật tiến hóa, di truyền).

Email: vvmnhat@hueuni.edu.vn



Lê Văn Hòa, nhận học vị thạc sỹ năm 2013; Hiện làm nghiên cứu sinh tại Khoa CNTT, Đại học Khoa học, Đại học Huế. Lĩnh vực nghiên cứu: Mạng truyền dẫn quang (OBS), mạng cảm biến tích hợp với RFID.

Email: lvhoa@hueuni.edu.vn



Huỳnh Quốc Phương, nhận học vị thạc sỹ năm 2016; Hiện đang công tác tại Khoa CNTT, trường Đại học An Giang; Lĩnh vực nghiên cứu: mạng cảm biến tích hợp với RFID, mạng nơ ron nhân tạo.

Email: hqphuong@agu.edu.vn