

**ĐẠI HỌC QUỐC GIA HÀ NỘI  
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**

**DƯƠNG VIỆT HUY**

**TỔNG HỢP DỮ LIỆU NHIỀU SENSOR  
TRONG MẠNG CẢM BIẾN KHÔNG DÂY**

**Chuyên ngành: TRUYỀN DỮ LIỆU VÀ MẠNG MÁY TÍNH  
Mã số: 62.48.15.01**

**TÓM TẮT LUẬN ÁN  
TIẾN SĨ CÔNG NGHỆ THÔNG TIN**

**NGƯỜI HƯỚNG DẪN KHOA HỌC:  
PGS. TS. Nguyễn Đình Việt**

**HÀ NỘI, 2017**

Công trình được hoàn thành tại:

Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội

Người hướng dẫn khoa học: **PGS. TS. Nguyễn Đình Việt**

Phản biện: .....

.....

Phản biện: .....

.....

Phản biện: .....

.....

Luận án sẽ được bảo vệ trước Hội đồng cấp Đại học Quốc gia chấm luận án  
tiến sĩ họp tại .....  
vào hồi        giờ        ngày        tháng        năm

Có thể tìm hiểu luận án tại:

- Thư viện Quốc gia Việt Nam
- Trung tâm Thông tin - Thư viện, Đại học Quốc gia Hà Nội

## MỞ ĐẦU

Sự xuất hiện của các mạng cảm biến không dây – WSNs (*Wireless Sensor Networks*) và các ứng dụng của chúng là một trong những xu hướng công nghệ chiếm ưu thế, phù hợp với xu thế phát triển của cuộc cách mạng công nghiệp lần thứ 4 hiện nay và trong những thập kỷ tới. Các mạng này được thiết kế bởi số lượng nút cảm biến (*sensor*), kích thước và chức năng mỗi nút tùy thuộc vào từng ứng dụng cụ thể, chúng có thể hoạt động độc lập hoặc theo nhóm. Trong nhiều trường hợp, các nút cảm biến chỉ sử dụng nguồn năng lượng dự trữ là pin mà chưa sử dụng nguồn năng lượng tái tạo, khi năng lượng lưu trữ của nút thấp hơn ngưỡng nào đó, nút sẽ không hoạt động và không thể tham gia với tư cách là một nút trong mạng. Các nút cảm biến hoạt động có nhiệm vụ theo dõi mục tiêu, mỗi nút có thể theo dõi được nhiều tham số và gửi (trực tiếp hoặc gián tiếp thông qua các nút trung gian) kết quả này đến trạm đích - BS (*Base Station*). Mỗi nút có thể tự nhận biết vị trí địa lý của nó so với các nút lân cận cũng như trên toàn mạng. Nút cũng có thể tự kiểm soát được mức năng lượng của chính nó và điều chỉnh công suất phát sóng tùy theo khoảng cách và dung lượng truyền dữ liệu đến nơi nhận.

Khi nhiều nút cùng theo dõi một mục tiêu và cùng gửi kết quả này đến BS sẽ có hiện tượng dư thừa dữ liệu (do có cùng thông tin theo dõi về mục tiêu) gây lãng phí năng lượng dự trữ của nút đồng thời tăng nguy cơ nghẽn mạng. Tổng hợp dữ liệu (*data fusion hoặc data aggregation*) trên đường truyền từ nút cảm biến trực tiếp theo dõi mục tiêu đến BS là một trong những giải pháp khắc phục nhược điểm nêu trên.

Cho đến nay, việc tổng hợp dữ liệu thông qua mạng có phân cụm được nhiều nhóm lựa chọn để nghiên cứu bởi ưu điểm: phù hợp với mạng cảm biến tĩnh với việc các nút cảm biến được rải ngẫu nhiên và vị trí không thay đổi trong quá trình hoạt động cho đến lúc hết năng lượng, nút cảm biến sẽ ngưng hoạt động (trạng thái “die”); thuận tiện trong kiểm soát định tuyến nhằm tiết kiệm năng lượng truyền dữ liệu đến đích; phân bố tiêu hao năng lượng trên toàn mạng bằng việc thiết lập lại cụm (*cluster*) và nút cụm trưởng – CH (*cluster head*). Nút CH được lựa chọn theo một giải thuật nhất định, có thể được chỉ định bởi BS hoặc được bầu bởi các nút trong cụm. CH sẽ chịu trách nhiệm tổng hợp dữ liệu từ các nút trong cụm và gửi trực tiếp kết quả này đến BS hoặc gửi gián tiếp đến BS thông qua CH (hoặc nút) khác trên tuyến truyền.

Hiện nay, các hướng nghiên cứu ở trong và ngoài nước về chủ đề tổng hợp dữ liệu chủ yếu hướng đến các mục tiêu chính sau:

- Sử dụng hiệu quả năng lượng trên từng nút mạng nhằm kéo dài tuổi thọ của nút đồng nghĩa với việc kéo dài “thời gian sống” của WSNs.
- Giải pháp đồng bộ hóa giữa các nút cảm biến nhằm chia sẻ vai trò của mỗi nút trong việc theo dõi về mục tiêu và truyền dữ liệu đó đến đích. Việc đồng bộ hóa có thể là điều khiển nút thức – ngủ một cách thông minh hay đo lường mục tiêu khi có sự biến động mà không theo chu kỳ cố định...

- An toàn dữ liệu trên các nút và đường truyền từ nút gửi đến nút nhận hoặc nút gửi đến BS. Các yếu tố dữ liệu được xem xét như: Tính bảo mật, tính toàn vẹn, tính xác thực, tính sẵn sàng và tính tươi mới của dữ liệu.

Như vậy, hầu hết các nghiên cứu đó chưa đề xuất được một giải pháp tổng thể nhằm tiệm cận giá trị tối ưu của WSNs (trường hợp mạng lý tưởng, dữ liệu được truyền như với mạng có dây trong điều kiện lý tưởng) hoặc chỉ áp dụng thuật toán để giải quyết một công đoạn nào đó trong ứng dụng cụ thể. Do vậy, cần rất nhiều nghiên cứu chuyên sâu khác để theo kịp và phù hợp với nhu cầu phát triển mạnh của ứng dụng của WSNs. Hướng nghiên cứu tổng hợp dữ liệu nhiều cảm biến hầu như chưa được nghiên cứu ở Việt Nam.

Với những đặc điểm cơ bản về tình hình nghiên cứu ở trong và ngoài nước đối với WSNs nói chung và vấn đề tổng hợp dữ liệu nhiều cảm biến trong mạng cảm biến không dây nói riêng đã đặt ra một số hướng cần được nghiên cứu có tính thời sự trong thực tiễn hiện nay. Được sự hướng dẫn tận tình của PGS. TS. Nguyễn Đình Việt và các thầy cô giáo Bộ môn Mạng và Truyền thông máy tính, tôi đã lựa chọn đề tài nghiên cứu "**Tổng hợp dữ liệu nhiều sensor trong mạng cảm biến không dây**" để tập trung giải quyết và làm sáng tỏ một số nội dung có liên quan, đồng thời gợi mở một số hướng nghiên cứu khác có liên quan.

**Đối tượng nghiên cứu:** Luận án nghiên cứu các vấn đề liên quan đến mô hình tổng hợp dữ liệu nhiều nút cảm biến trong mạng cảm biến không dây, bao gồm:

- Kỹ thuật phân chia WSNs thành cụm theo chu kỳ;
- Vấn đề theo dõi mục tiêu của nút cảm biến;
- Vấn đề lấy mẫu dữ liệu bằng cách lựa chọn dữ liệu của một số nút;
- Vấn đề tiền xử lý dữ liệu làm đầu vào tổng hợp dữ liệu;
- Kỹ thuật tổng hợp dữ liệu theo cụm tại nút cụm trưởng - CH.

**Phạm vi nghiên cứu được đề cập ở Luận án như sau:**

- Mạng cảm biến bao gồm số lượng nút hữu hạn có vị trí cố định sau khi được rải ngẫu nhiên trong phạm vi cần giám sát.

- Cụm nút cảm biến được phân chia (từ mạng cảm biến) bằng xác suất và đã biết trước. Số lượng cụm, số lượng nút của mỗi cụm có thể khác nhau tại mỗi thời điểm mạng đã được phân chia.

- Nút cảm biến đồng nhất (về cấu tạo, lượng pin dự trữ) tại thời điểm mạng bắt đầu hoạt động. Năng lượng dự trữ của nút giảm dần khi hoạt động.

- Nút cụm trưởng (CH) được lựa chọn từ các nút trong cụm. Vì vậy sau khi phân chia mạng, số lượng nút của cụm sẽ bao gồm cả nút CH.

- Mục tiêu theo dõi là ngẫu nhiên trong phạm vi giám sát và được tập trung hóa thành một điểm. Trong thực tế, mục tiêu có thể là đối tượng để trinh sát, phát hiện đột nhập...

**Mục tiêu nghiên cứu chính của đề tài:** (1) Tiết kiệm năng lượng cho nút bằng cách giảm việc có nhiều (từ hai nút trở lên) phải truyền trên mạng các gói tin dư thừa do có cùng thông tin mà các nút đó cùng đo lường được về mục tiêu; (2) đảm bảo tính đúng đắn về giá trị đo của dữ liệu từ nút cảm nhận sự kiện của mục tiêu cho đến BS; (3) hướng đến tối ưu độ hội tụ của không gian và thời gian.

**Bố cục của Luận án** gồm các phần Mở đầu, Kết luận và 4 chương, trong đó:

Chương 1 có nội dung tổng quan vấn đề cần nghiên cứu trong đó đề cập đến lịch sử phát triển, kiến trúc của mạng cảm biến, nút cảm biến; vấn đề tiêu thụ năng lượng nút cảm biến; tóm lược tình hình sử dụng công cụ mô phỏng mạng cảm biến hiện nay, giới thiệu bộ mô phỏng NS-2 và mã nguồn của MIT mà Luận án sẽ thừa kế.

Chương 2 với nội dung chính về định tuyến và tổng hợp dữ liệu nhiều cảm biến. Định tuyến và tổng hợp dữ liệu đối với WSNs có quan hệ rất mật thiết với nhau, sự kết hợp giữa chúng tạo nên hiệu quả tốt hơn trong việc tiết kiệm năng lượng; phân tích giao thức định tuyến phân cụm, tổng hợp dữ liệu đối với mạng có phân cụm đồng thời đặt ra một số vấn đề cần giải quyết; Đề xuất mô hình tổng hợp dữ liệu nhiều nút cảm biến trong WSNs với các nhóm bài toán thành phần, làm cơ sở để trình bày chi tiết ở Chương 3 và Chương 4.

Chương 3 là nhóm giải pháp đề xuất, cải tiến việc theo dõi mục tiêu và lựa chọn dữ liệu của mô hình tổng hợp dữ liệu đã đặt ra ở Chương 2. Các bài toán con gồm theo dõi mục tiêu dựa vào khoảng cách giữa nút cảm biến, CH và mục tiêu; theo dõi thích nghi với biến động của mục tiêu và trạng thái đo tốt nhất của nút cảm biến; lựa chọn một số nút cảm biến theo cơ chế cửa sổ trượt để lấy dữ liệu làm đầu vào tổng hợp dữ liệu.

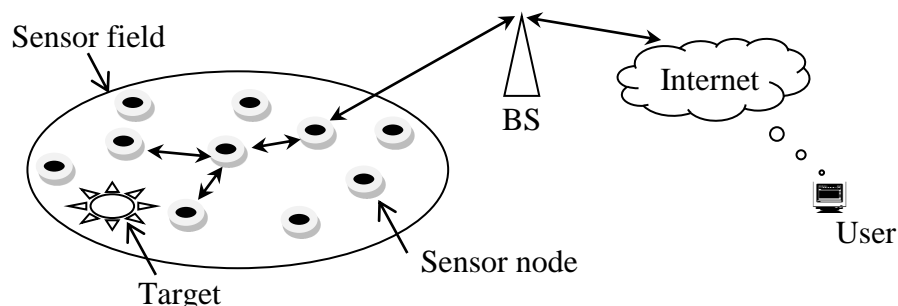
Chương 4 là nhóm giải pháp tổng hợp dữ liệu tại nút CH theo mô hình đã đặt ra ở Chương 2, các giải pháp đề xuất, cải tiến gồm: Áp dụng sự phù hợp và các phép tính toán của lý thuyết tập thô để đưa ra quyết định về tổng hợp dữ liệu; tiền xử lý tại CH để tạo bộ dữ liệu tốt hơn phục vụ tổng hợp; đề xuất phương pháp tính toán đơn giản để tổng hợp dữ liệu phù hợp với tài nguyên và khả năng tính toán thấp của nút cảm biến.

**Phương pháp nghiên cứu:** Sử dụng kết hợp phương pháp nghiên cứu lý thuyết, mô phỏng và thực nghiệm đối với nút cảm biến và WSNs trong phạm vi nghiên cứu.

## Chương 1. TỔNG QUAN VẤN ĐỀ NGHIÊN CỨU

### 1.1. Mạng cảm biến không dây

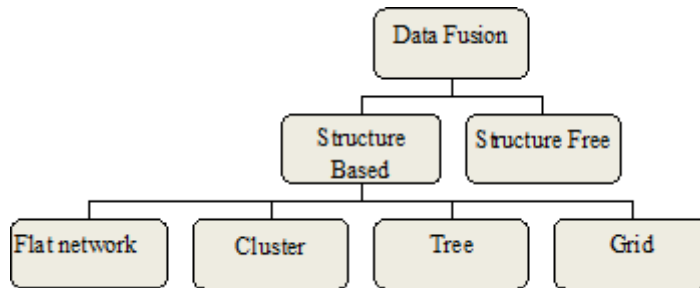
Mô hình chung của một WSNs bao gồm 6 thành phần (xem Hình 1.1): *Target* là mục tiêu hay nguồn sinh sự kiện cần cảm biến; *Sensor node* ghi nhận thay đổi của mục tiêu; *Sensor field*: Vùng cảm biến được giới hạn bởi đường biên chứa toàn bộ *Target* và *Sensor node*; *BS* là trạm gốc, chịu trách nhiệm điều khiển, giao tiếp với *Sensor field* để truyền dữ liệu đó đến *User*; *User*: Là người sử dụng kết quả cảm biến; *Internet*: Là môi trường truyền dẫn giữa *User* và *BS*.



Hình 1.1. Mô hình mạng cảm biến không dây

### Một số cách phân loại mạng

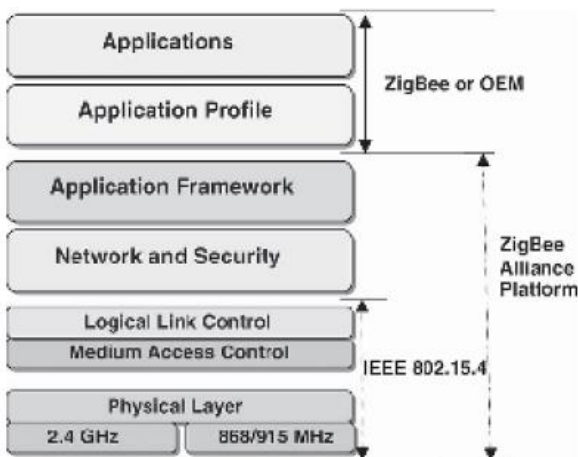
Hiện nay, có nhiều tiêu chí để phân loại mạng, ví dụ: Theo cấu trúc mạng, theo chức năng của mỗi lớp mô hình OSI, theo dữ liệu đầu vào/đầu ra, theo chức năng framework... Tuy nhiên, phân loại WSNs theo cấu trúc mạng được nhiều nhóm tiếp cận. Theo cách chia này, vấn đề tổng hợp dữ liệu sẽ dựa trên mạng có cấu trúc (*structure based*) và mạng không có cấu trúc (*structure free*) xem Hình 1.2.



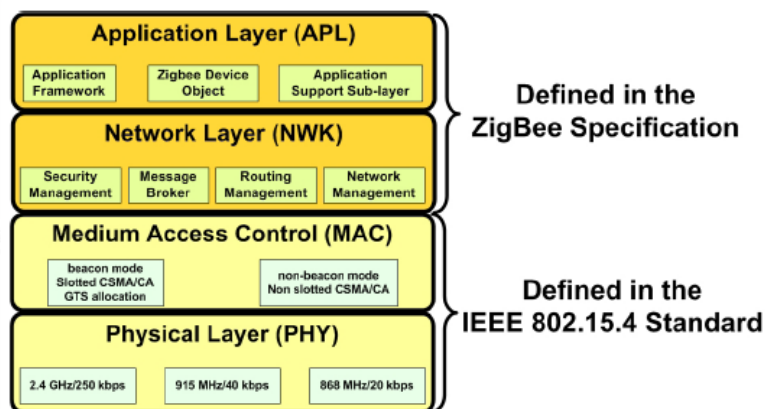
Hình 1.2. Hướng tiếp cận theo kiến trúc mạng

### Giao thức truyền dữ liệu

Tiêu chuẩn IEEE 802.15.4 được bổ sung vào họ tiêu chuẩn mạng không dây IEEE 802. Giao thức Zigbee là chuẩn bổ sung cho IEEE 802.15.4 (vì vậy thường được viết tắt Zigbee/IEEE 802.15.4) bao gồm tập hợp các giao thức không những ở tầng vật lý (PHY) và MAC mà còn từ tầng mạng (*Network*) đến ứng dụng (*Application*) để giao tiếp mạng không dây khoảng cách ngắn có tốc độ truyền dữ liệu thấp. Các thiết bị không dây dựa trên chuẩn ZigBee hoạt động trên 3 dải tần số là 868MHz, 915MHz và 2.4GHz. Tốc độ dữ liệu là 250Kbps ở dải tần 2.4 GHz sử dụng toàn cầu, 40 Kbps ở dải tần 915 MHz (đối với Mỹ, Nhật) và 20kbps ở dải tần 868 MHz ở Châu Âu.



Hình 1.3. Mô hình tham chiếu Zigbee và 802.15.4



Hình 1.4. Kiến trúc Zigbee/IEEE 802.15.4

Đặc điểm của công nghệ ZigBee là tốc độ truyền tin thấp, tiêu hao ít năng lượng, chi phí thấp, là giao thức mạng không dây hướng tới các ứng dụng điều khiển từ xa và tự động hóa. Mục tiêu của công nghệ ZigBee là hướng tới việc truyền tin với mức tiêu hao năng

lượng nhỏ và công suất thấp cho những thiết bị chỉ có thời gian sống hữu hạn và không yêu cầu cao về tốc độ truyền tin.

Hình 1.3, Hình 1.4 là mối quan hệ chuẩn IEEE 802.15.4 và Zigbee trong các lớp kiến trúc mạng không dây Zigbee. Thừa kế chuẩn IEEE 802.15.4, Zigbee tập trung vào thiết kế liên quan đến lớp mạng, bảo mật và lớp ứng dụng; cung cấp các thông số cho khả năng tương thích giữa nút cảm biến với các lớp mạng.

## 1.2. Các vấn đề cần giải quyết

### 1.2.1. Tiêu thụ năng lượng

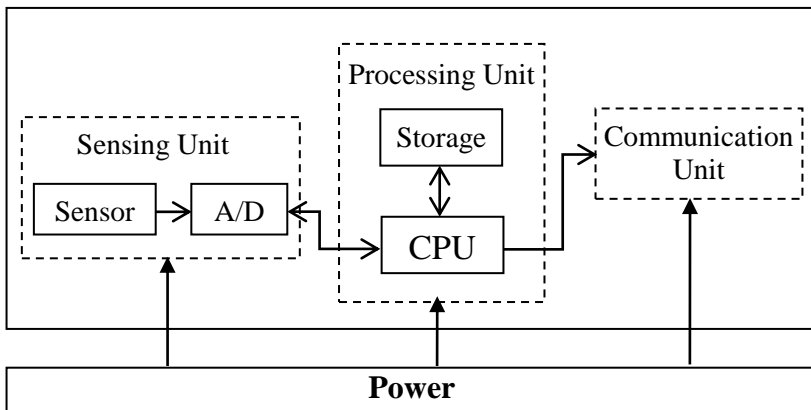
#### a) Năng lượng tiêu thụ của nút cảm biến

Tùy thuộc vào yêu cầu ứng dụng để nhà thiết kế nút cảm biến lựa chọn tích hợp công nghệ phù hợp. Nhìn chung, nút có cấu tạo các khối và tiêu thụ năng lượng như Hình 1.5:

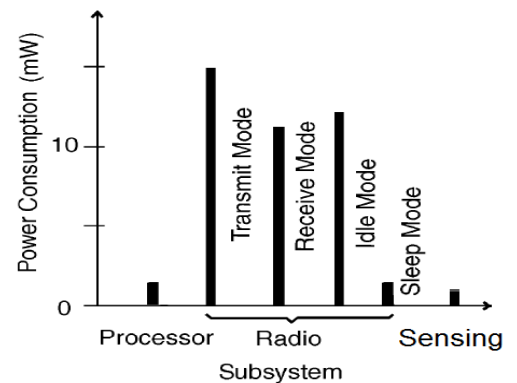
Đơn vị xử lý - PU (*Processing unit*): PU gồm phần lưu trữ dữ liệu (*storage*) và CPU để điều khiển và xử lý toàn bộ hoạt động của nút.

Đơn vị cảm biến - SU (*Sensing unit*): Cung cấp cho khối cảm nhận và truyền thông tin về sự kiện đến PU. SU gồm phần tử cảm nhận (*sensor*) và bộ chuyển đổi tín hiệu A/D (A→D từ SU → PU và D→A từ PU → SU).

Đơn vị truyền thông (*Communication unit*): Tiêu thụ năng lượng để thực hiện việc truyền thông tín hiệu từ nút mạng này đến nút mạng khác hoặc truyền thông với BS..



Hình 1.5. Sơ đồ cung cấp năng lượng cho nút cảm biến



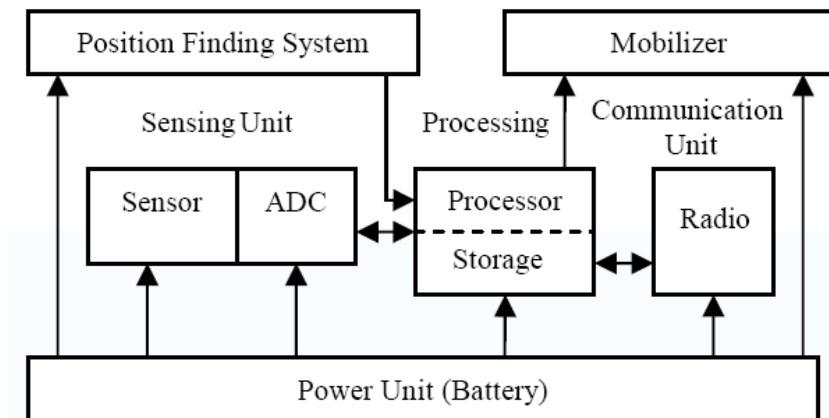
Hình 1.6. Mức tiêu thụ năng lượng của các chức năng của nút cảm biến

#### b) Mối quan hệ giữa tiêu thụ năng lượng của nút và mạng

Năng lượng tiêu thụ của WSNs chủ yếu là của các nút cảm biến của mạng. Năng lượng tổn hao do quá trình thu phát tín hiệu bằng sóng vô tuyến lớn hơn gấp nhiều lần so với tổn hao năng lượng để xử lý các công đoạn khác, trong đó có việc tính toán trên nút. Sơ đồ so sánh tỉ lệ mức tiêu thụ năng lượng trong quá trình nút cảm biến hoạt động ở Hình 1.6.

Ngoài ra, năng lượng tiêu thụ của toàn bộ mạng và nút cảm biến được liên hệ ở Hình 1.7. Theo đó, nút cảm biến còn phải tiêu hao năng lượng để phối hợp với BS trong việc dò tìm vị trí (*position finding*) cũng như phục vụ tính di chuyển (*mobilizer*) nếu có của nút cảm biến. Trong khi khả năng dự trữ năng lượng của nút cảm biến là có giới hạn thì việc sử dụng

năng lượng không bị hạn chế của BS là điều kiện để tối ưu hóa năng lượng tiêu thụ của toàn mạng nhằm đạt hiệu quả tốt nhất về hiệu năng của mạng.



Hình 1.7. Môi liên hệ trong tiêu thụ năng lượng của nút và BS

### 1.2.2. Thiết kế nút mạng cảm biến

Việc thiết kế phần cứng nút cảm biến có thể phụ thuộc vào nhà sản xuất nhưng chúng phải đảm bảo theo tiêu chuẩn để dùng chung. Tùy vào mục tiêu chính cần được ưu tiên nhưng thông thường, nhà thiết kế phải cố gắng cân bằng giữa các mục tiêu như hiệu suất, tuổi thọ của pin, kích thước của nút, độ tin cậy, lưu lượng dữ liệu, phạm vi phủ sóng để hoạt động tốt ... và đặc biệt là giá thành sản phẩm.

### 1.2.3. Tổ chức mạng và định tuyến

Với mục tiêu nâng cao hiệu năng hoạt động của mạng bằng việc tối ưu hiệu quả sử dụng năng lượng của nút cảm biến, việc tổ chức mạng và định tuyến để truyền dữ liệu có ích đến đích luôn được đặt ra. Hiện nay, việc các nút mạng cũng có thể di chuyển làm thay đổi sơ đồ mạng đặt ra cho các nhóm nghiên cứu hướng đến giải pháp tối ưu hóa khả năng tự điều chỉnh/tự cấu hình (*auto-reconfigurable*) của WSNs để tổ chức mạng theo sơ đồ (*topology*) linh động như ad-hoc, mesh, star...

### 1.2.4. Truyền và xử lý dữ liệu

Truyền dữ liệu và xử lý dữ liệu bao gồm các giai đoạn từ thời điểm nút cảm biến theo dõi mục tiêu, chuyển đổi thông tin theo dõi thành dạng số, đóng gói, xử lý trên nút cảm biến đó, gửi trực tiếp hoặc qua các nút trung gian (tùy mô hình mạng, có thể là nút cảm biến tiếp theo, CH...) để đến đích BS. Dữ liệu qua các nút cảm biến trên tuyến truyền có thể được xử lý để đạt mục tiêu nào đó như: giảm dữ liệu dư thừa, nén để giảm dung lượng...

### 1.2.5. Tổng hợp dữ liệu

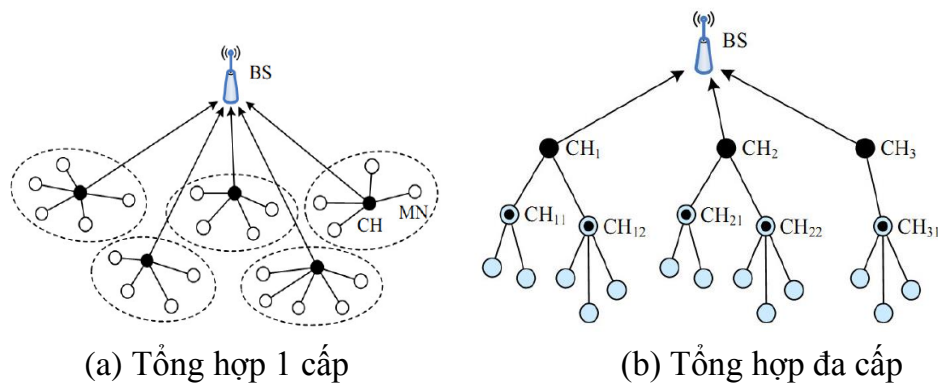
Để tổng hợp dữ liệu, WSNs được chia thành nhiều cụm, mỗi cụm có một nút cụm trưởng (CH) được bầu theo một giải thuật nhất định, việc tổng hợp dữ liệu từ các nút trong cụm này được diễn ra ở nút cụm trưởng (CH) của cụm. Kết quả tổng hợp dữ liệu được gửi trực tiếp đến đích BS (xem Hình 1.8a) nếu tổng hợp dữ liệu theo mô hình phân nhóm một cấp *hoặc* đến BS thông qua một nút CH kế tiếp trong trường hợp mạng được tổ chức thành nhiều cụm, nhiều cấp; Hình 1.8b là ví dụ với mô hình tổng hợp 2 cấp.



Công trình của nhóm tác giả Y. Takama and D. Ursino đã hệ thống hóa được tiêu chí phân loại theo đầu vào đầu ra. Trong đó dữ kiện đầu vào để tổng hợp dữ liệu có thể là dữ liệu (*data*), tính năng (*feature*) hoặc quyết định (*decision*).

Cho đến nay, nhiều cá nhân và nhóm nghiên cứu đã thực hiện một số lượng khá lớn các công trình về tổng hợp dữ liệu nhiều cảm biến trong mạng cảm biến không dây và đã đề xuất được 5 nhóm cơ sở để giải quyết bài toán:

- Dựa vào độ quan trọng của hệ số đo (*weight coefficient*);
- Dựa vào ước lượng tham số (*parameter estimation*);
- Dựa trên bộ lọc, xác suất, máy học: Kalman (*Kalman filtering*); Bayesian, Dempster-Shafer, Fuzzy logic, Neural Network,...
- Dựa trên lý thuyết tập thô (*rough set theory*);
- Dựa trên độ đo mức độ ngẫu nhiên của thông tin (*information entropy*).

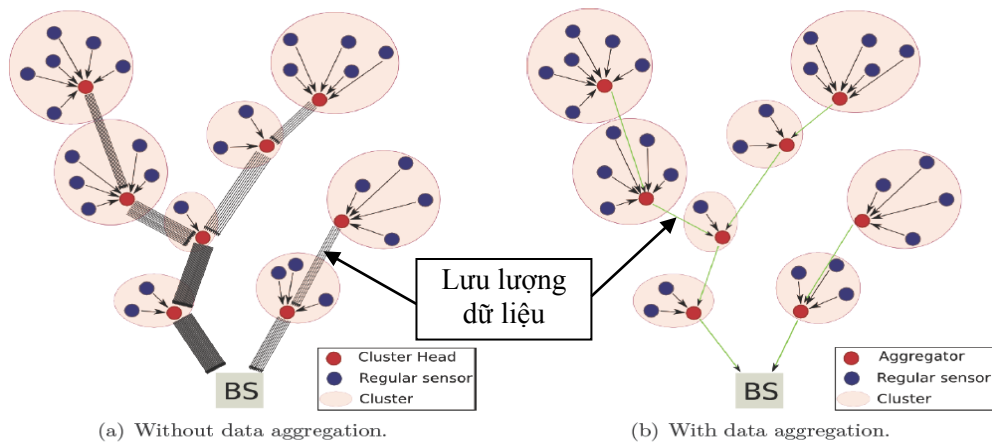


Hình 1.8. Tổng hợp dữ liệu nhiều cảm biến không dây

### 1.2.6. Xử lý vấn đề dữ liệu dư thừa

Trong quá trình cảm nhận, truyền và xử lý thông tin từ nguồn gây ra sự kiện (là mục tiêu cần theo dõi) đến đích cuối cùng (BS), thường có nhiều dữ liệu dư thừa nhưng vẫn được truyền trong mạng, gây nên sự lãng phí về năng lượng cũng như tăng nguy cơ nghẽn mạng. Dữ liệu dư thừa là dữ liệu có cùng giá trị và cùng thông tin. Ví dụ, có hai sensor  $S_1, S_2$ , mỗi sensor được chế tạo để đo 3 tham số  $\{x_1, x_2, x_3\} = \{\text{nhiệt độ, độ ẩm, tốc độ gió}\}$ . Kết quả đo lường của mỗi sensor là bộ số  $\{x_1, x_2, x_3\}$ . Giả sử tại một thời điểm  $t_1$  nào đó, kết quả đo mục tiêu của  $S_1$  và  $S_2$  tương ứng là  $\{x_1S_1=30, x_2S_1=60, x_3S_1=5\}$  và  $\{x_1S_2=30, x_2S_2=60, x_3S_2=6\}$  với  $x_iS_j$  là giá trị đo tham số  $x_i$  của sensor  $S_j$  thì có thể nói rằng dữ liệu  $S_1$  và  $S_2$  bị dư thừa một phần vì  $x_1S_1=x_1S_2=30$  và  $x_2S_1=x_2S_2=60$ . Trong trường hợp có thêm  $x_3S_1=x_3S_2$  thì có thể xem dữ liệu  $S_1$  (hoặc  $S_2$ ) bị dư thừa hoàn toàn. Để giải quyết vấn đề này, giải pháp tổng hợp dữ liệu trên đường truyền từ nguồn đến đích là điều cần thiết.

Hình 1.9 mô tả lưu lượng dữ liệu của mạng cảm biến khi không thực hiện tổng hợp dữ liệu (Hình 1.9a) và khi được tổng hợp dữ liệu (Hình 1.9b). Rõ ràng khi áp dụng giải pháp tổng hợp dữ liệu thì lưu lượng dữ liệu của mạng giảm đáng kể thể hiện ở mức độ “đầy” của luồng dữ liệu.

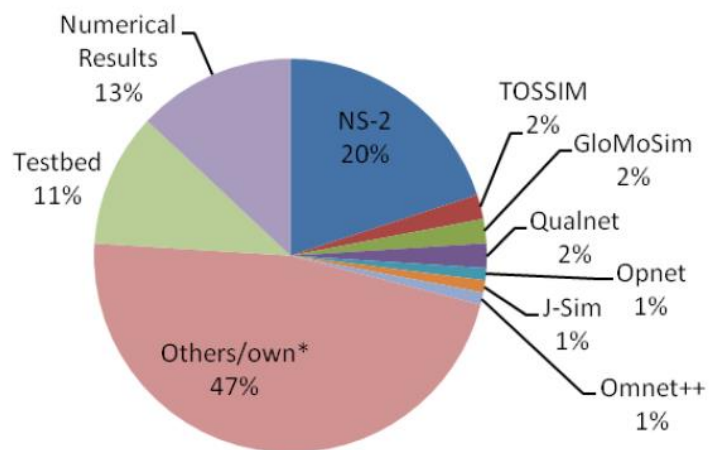


Hình 1.9. So sánh mô hình truyền dữ liệu của WSNs

### 1.3. Mô phỏng mạng cảm biến

Hiện nay, có nhiều công cụ để mô phỏng WSNs như: OPNET, OMNet++, NS-2, J-Sim (JavaSim), Mannasim (base NS-2), SensorSim (base NS-2), GloMoSim... Mỗi công cụ đều được các nhóm phát triển mở rộng với mục tiêu cụ thể và có lợi thế riêng. Tuy nhiên phần mềm mô phỏng NS-2 được các nhóm nghiên cứu chú ý hơn cả vì có lịch sử lâu đời cũng như hỗ trợ của cộng đồng.

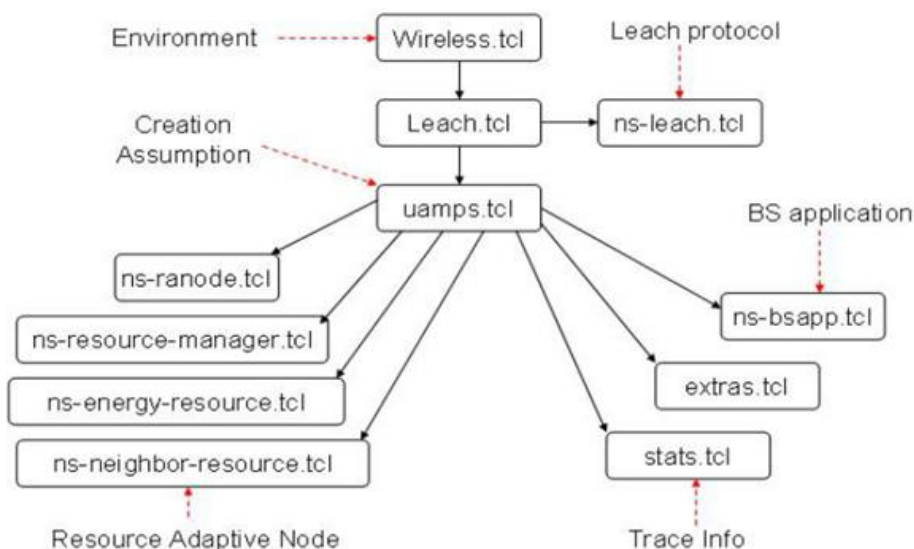
(\*) Bao gồm MATLAB, phương pháp Monte Carlo và một số chương trình mô phỏng hiếm như Prowler, Emstar, JiST/SWANS, ROSS, GTNetS



Hình 1.10. Tỷ lệ sử dụng các phần mềm mô phỏng

Theo thống kê từ năm 2000 đến nay, có khoảng 20% nhà nghiên cứu về mạng không dây lựa chọn chương trình mô phỏng NS-2 để phân tích, định lượng và đánh giá hiệu năng mạng không dây (xem Hình 1.10).

**NS-2 và phần mở rộng mô phỏng WSNs của MIT:** Nhóm nghiên cứu của MIT (*Massachusetts Institute of Technology*) đã phát triển mã nguồn để mô phỏng WSNs. Ngoài việc sử dụng nguồn lực, đối tượng và các lớp có sẵn của NS-2, MIT đã thêm một số nhóm đối tượng. Kiến trúc của MIT ở Hình 1.11:



Hình 1.11. Kiến trúc mô phỏng WSNs của MIT

## Chương 2. ĐỊNH TUYẾN PHÂN CỤM VÀ TỔNG HỢP DỮ LIỆU

### 2.1. Giao thức định tuyến phân cụm thích ứng với năng lượng thấp

Định tuyến phân cụm của mạng phân cụm (*cluster-based network*) thích ứng với năng lượng thấp LEACH được xem là thuật giải điển hình. Đặc điểm chính của LEACH: Chia nhỏ mạng thành các cụm (*cluster*) không giao nhau để phân phối mức tiêu thụ năng lượng, phân cấp trong vấn đề tổng hợp và định tuyến. Mỗi cụm có 1 nút cụm trưởng CH (*cluster head*). Các nút trong cụm gửi dữ liệu đến CH kiểu đa chặng (*multihop*), CH chịu trách nhiệm tổng hợp để giảm dữ liệu dư thừa và gửi dữ liệu tổng hợp đến BS (hoặc CH cấp cao hơn) kiểu đơn chặng (*singlehop*). Do đó, CH sẽ tiêu tốn nhiều năng lượng hơn các nút khác trong cụm của nó. Các nút trong mỗi cụm sẽ sử dụng kênh truyền kiểu TDMA và gán cho mỗi nút 1 khe thời gian để tránh xung đột khi truyền dữ liệu đến CH. Các nút cụm trưởng sử dụng CDMA (đa phân chia theo mã) để quảng bá/giao tiếp với các nút trong cụm nhằm tránh xung đột với các nút khác ngoài cụm.

LEACH có 2 pha: Pha thiết lập cụm (gồm lựa chọn nút CH, tạo thông tin của cụm) và pha ổn định trạng thái (gồm tiếp nhận dữ liệu, tổng hợp dữ liệu và truyền dữ liệu đến BS).

Vấn đề năng lượng của LEACH: Giải thuật LEACH quy ước các nút trong mạng có năng lượng đồng nhất tức lúc mạng bắt đầu hoạt động, toàn bộ các nút trong mạng có nguồn dự trữ năng lượng như nhau. Kênh truyền tin là đối xứng, năng lượng để truyền (và nhận) 1 bit dữ liệu giữa 2 nút lúc gửi (và nhận) là như nhau. Lựa chọn CH bằng xác suất.

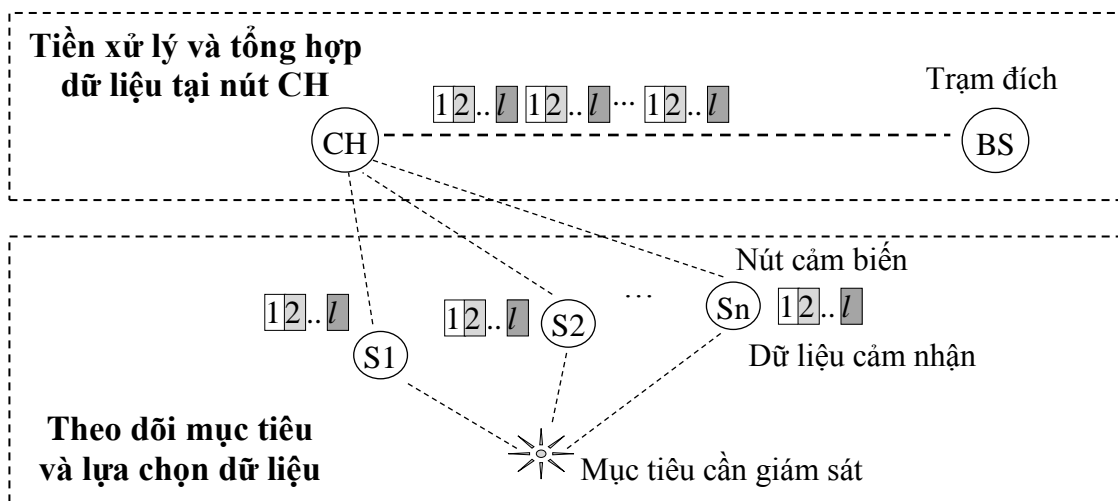
Do đó, LEACH có hai ưu điểm: (1) phân tán sự tiêu hao năng lượng ra các nút trong mạng đồng nghĩa với có thể kéo dài thời gian sống của mạng; (2) Có thể tổng hợp ở CH để giảm việc truyền dư thừa thông tin đến BS.

Tuy nhiên, LEACH tồn tại một số nhược điểm như: (1) CH được chọn ngẫu nhiên mà không xem xét yếu tố năng lượng dự trữ của nút do đó khả năng nút ít năng lượng hơn vẫn

có thể được chọn làm CH: (2) Trong pha thiết lập cụm, ngoài chọn ngẫu nhiên CH còn chọn các nút thành viên của cụm. CH sẽ gửi bản tin mời các nút lân cận tham gia nhóm, có thể tồn tại một số nút ở xa CH (ví dụ biên của mạng) không nhận được “lời mời” này nên trong vòng này, các nút đó sẽ không thuộc CH nào và đương nhiên không tham gia mạng; (3) Với pha ổn định trạng thái: Việc lấy thông tin cảm nhận theo chu kỳ sẽ không hợp lý nếu tần số xảy ra sự kiện bé hơn tần số lấy mẫu. Điều này có nghĩa là mạng đã tiêu tốn năng lượng để thu được các mẫu dữ liệu cảm biến như nhau; (4) Trong mỗi vòng, BS sẽ thiết lập N khe thời gian để CH gửi thông tin đã được tổng hợp đến BS (theo TDMA), nhưng chưa chắc vòng đó có N cụm (hay CH), nghĩa là tồn tại những khe thời gian, thông tin không được truyền, điều này gây hiện tượng trễ.

Vì lý do đó, rất nhiều đề xuất cải tiến nhằm tận dụng hết năng lượng phân tán ở các nút mạng. Nhiều nghiên cứu thêm, bớt, cải tiến nhằm hướng một số mục tiêu sau: Chọn nút có năng lượng lớn nhất CH như E-Leach (*energy leach*), gắn yếu tố khoảng cách/vị trí giữa nút (sẽ chọn làm CH) với BS (*Leach-C*), dùng CH khác làm trung gian chuyển tiếp giữa CH với BS TL-Leach (*two-lever Leach*), truyền dữ liệu từ nút đến CH kiểu Multihop (*multihop Leach*), thêm 1 CH dự phòng (*V-Leach*)... Các nhóm tác giả cũng chỉ rõ những yếu điểm cần cải tiến tiếp.

## 2.2. Mô hình tổng hợp dữ liệu và bài toán thành phần



Hình 2.1. Mô hình tổng hợp dữ liệu và các bài toán thành phần

Kết quả tổng hợp dữ liệu là tổ hợp nhiều công đoạn thành phần. Bắt đầu từ việc các nút cảm biến theo dõi mục tiêu, nút cảm biến truyền dữ liệu đến CH (có thể theo điều kiện do CH hoặc BS đặt ra), CH tiếp nhận dữ liệu, tiền xử lý và thực hiện tổng hợp dữ liệu và gửi dữ liệu này đến BS. Mô hình tổng hợp dữ liệu ở Hình 2.1 với 02 nhóm bài toán con: *Thứ nhất*, nhóm các bài toán theo dõi mục tiêu và lựa chọn dữ liệu; *thứ hai*, tiền xử lý dữ liệu và tổng hợp dữ liệu tại nút CH. Đề xuất giải pháp giải quyết bài toán nhóm thứ nhất ở Chương 3, bài toán nhóm thứ hai ở Chương 4.

## Chương 3. THEO DÕI MỤC TIÊU VÀ LỰA CHỌN DỮ LIỆU

### 3.1. Theo dõi mục tiêu dựa vào khoảng cách

Giải pháp theo dõi mục tiêu dựa vào khoảng cách là một bài toán con thuộc mô hình tổng hợp dữ liệu đã trình bày ở Chương 2 (Mục 2.2). Giải pháp này sử dụng mô hình theo dõi mục tiêu kiểu cạnh tranh (*competitive*) và theo dõi định hướng mục tiêu (*target oriented tracking*) vì số lượng mục tiêu biết trước và kết quả đo lường được tính theo chu kỳ. Việc lựa chọn nút cảm biến dựa vào vị trí tương đối giữa nút cảm biến đó với CH và mục tiêu theo dõi để giải quyết được một phần nhược điểm vừa nêu trên, giải pháp có tên là *ETR-DF* (*Efficiency in TRacking to target in multi-sensor Data Fusion*).

Giả sử có một cụm nút cảm biến ( $S$ ) gồm  $n$  nút được rải ngẫu nhiên trên mặt phẳng, biết trước vị trí của 01 mục tiêu (Tag) và 01 nút cụm trưởng (CH). Ban đầu, năng lượng dự trữ của các nút cảm biến bằng nhau, trong quá trình sử dụng năng lượng của các nút sụt giảm và mức độ dự trữ có thể không bằng nhau. Giải pháp ETR-DF lựa chọn các nút nằm trên đường đi ngắn nhất giữa CH và Tag.

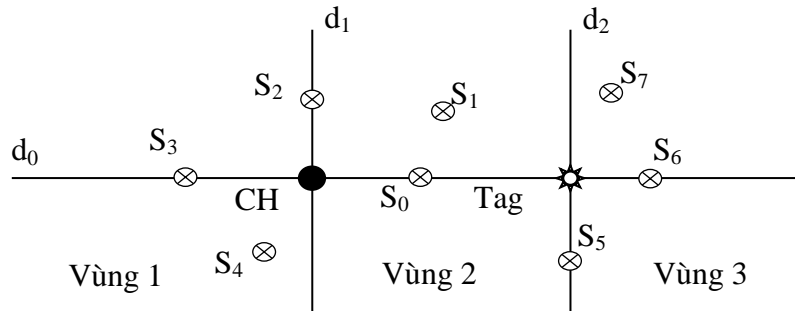
Không mất tính tổng quát, trong đề xuất này sử dụng khoảng cách trong hình học phẳng. Khi đó nút cảm biến, Tag, CH là các điểm trong mặt phẳng, tọa độ của các điểm là  $\text{Node}(x_{\text{node}}, y_{\text{node}})$ ,  $\text{Tag}(x_{\text{tag}}, y_{\text{tag}})$ ,  $\text{CH}(x_{\text{CH}}, y_{\text{CH}})$ . Gọi  $d_{\text{node-CH}}$ ,  $d_{\text{node-tag}}$ ,  $d_{\text{CH-tag}}$  lần lượt là khoảng cách giữa nút với CH, giữa nút với Tag, giữa CH với tag và được tính như sau:

$$d_{\text{node-CH}} = \sqrt{(x_{\text{node}} - x_{\text{CH}})^2 + (y_{\text{node}} - y_{\text{CH}})^2} \quad (\text{CT 3.1})$$

$$d_{\text{node-tag}} = \sqrt{(x_{\text{node}} - x_{\text{tag}})^2 + (y_{\text{node}} - y_{\text{tag}})^2} \quad (\text{CT 3.2})$$

$$d_{\text{CH-tag}} = \sqrt{(x_{\text{CH}} - x_{\text{tag}})^2 + (y_{\text{CH}} - y_{\text{tag}})^2} \quad (\text{CT 3.3})$$

Giữa CH và Tag luôn tồn tại 1 đường thẳng  $d_0$ , các đường thẳng  $d_1$ ,  $d_2$  vuông góc lần lượt đi qua CH và Tag chia không gian thành các phần như ở Hình 3.1. Ví dụ vị trí của nút cảm biến  $S_0..S_7$  so với CH và Tag tương ứng với 8 trường hợp có thể xảy ra.



Hình 3.1. Vị trí của nút cảm biến so với CH và Tag.

Nếu tại thời điểm xét, năng lượng dự trữ của các nút như nhau, cùng đo lường Tag và gửi đến CH cùng một đơn vị dữ liệu thì các nút nằm trên đường thẳng nối CH với Tag (ví dụ nút  $S_0$  trong Hình 3.1) có thể tiêu thụ ít năng lượng hơn vì khoảng cách  $d = d_{\text{node-CH}} + d_{\text{node-tag}}$

$= d_{CH-tag} = d_{min}$ . Gọi  $Ed_{node-tag}$  và  $Ed_{node-CH}$  lần lượt là năng lượng tiêu thụ của nút cảm biến lúc đo lường mục tiêu và gửi dữ liệu đến CH, khi đó:  $Ed_{node-CH} > Ed_{node-tag}$  theo và đúng với trường hợp  $d_{node-CH} = d_{node-tag}$ . Vì vậy trong trường hợp này, nút  $S_0$  ở gần CH có thể mang lại hiệu quả tốt hơn về tiết kiệm năng lượng.

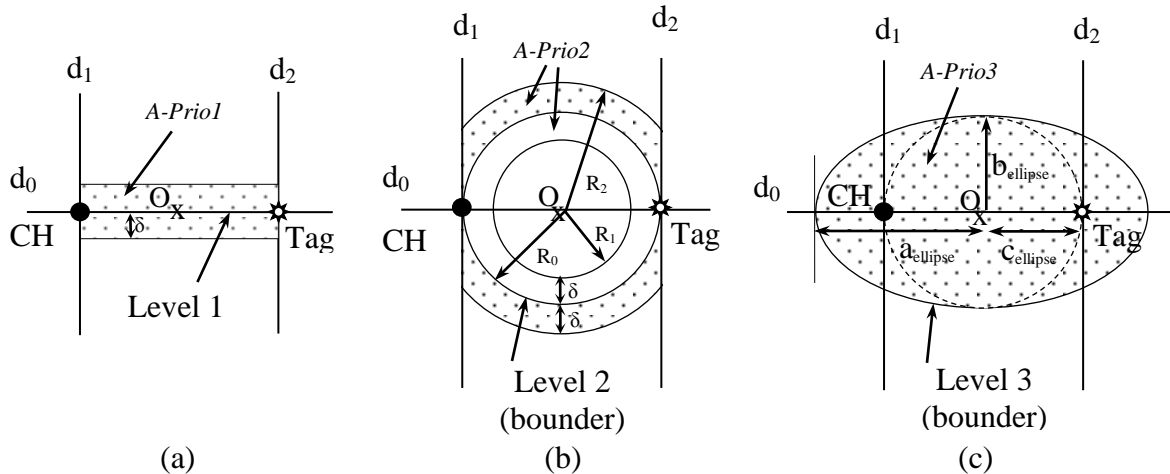
### 3.1.1. Sai số

Đề xuất sử dụng sai số  $\delta \geq 0$  để xác định giới hạn khoảng cách sai lệch vị trí của nút so với đường biên xác định vùng ưu tiên, mức ưu tiên. Sai số  $\delta$  chỉ được sử dụng trong vùng bị chặn bởi  $d_1, d_2$ . Có nghĩa là nếu trục hoành ( $Ox$ ) chứa  $d_0$ , gốc tọa độ  $O$  là trung điểm của CH và Tag thì chỉ xét các nút có tọa độ trên trục  $Ox$  trong khoảng bị chặn (hay khoảng đóng)  $[-(d_{CH-Tag})/2, (d_{CH-Tag})/2]$  Trường hợp lý tưởng  $\delta = 0$  khi nút nằm trên đường biên. Vì  $\delta \geq 0$  và có nhiều mức ưu tiên nên một nút có thể thuộc nhiều mức ưu tiên khác nhau, vị trí của nút đó nằm trong vùng giao nhau của các mức ưu tiên.

### 3.1.2. Vùng ưu tiên

Dựa vào phân tích về các khoảng cách giữa nút cảm biến, CH và Tag, đề xuất ETR-DF tập trung phân tích vùng 2 - vùng được giới hạn bởi  $d_1, d_2$  và bao gồm cả  $d_1, d_2$  (xem hình Hình 3.1). Vùng 2 được chia thành các vùng ưu tiên và mức ưu tiên ở Hình 3.2. Mức độ ưu tiên từ cao về thấp được CH sử dụng trong trường hợp lựa chọn thứ tự lấy kết quả đo lường về mục tiêu phục vụ tổng hợp dữ liệu. Điều này có nghĩa là, trong cùng một chu kỳ hoạt động của cụm, nút CH có thể lựa chọn nút bất kỳ trong cụm thuộc vùng ưu tiên có mức độ ưu tiên cao hơn, sử dụng kết quả đo đó để tổng hợp dữ liệu. Trong các vùng ưu tiên, tiêu chí để CH lựa chọn nút ngoài mức độ ưu tiên của vùng còn tiêu chí khác của nút cảm biến như năng lượng dự trữ, số gói tin cần phải tiếp tục gửi đến CH để hoàn thành dữ liệu đo mục tiêu, tỉ lệ  $d_{node-CH}/d_{node-tag} \dots$

Các nút cảm biến có mức ưu tiên từ cao đến thấp nếu vị trí thuộc phạm vi sau: *Mức 1* là đường thẳng CH-Tag; *Mức 2* giới hạn bởi đường tròn đường kính CH-Tag; *Mức 3* là vùng giới hạn bởi Elip có 2 tiêu điểm CH, Tag và tiêu cự  $d_{CH-tag}$ .



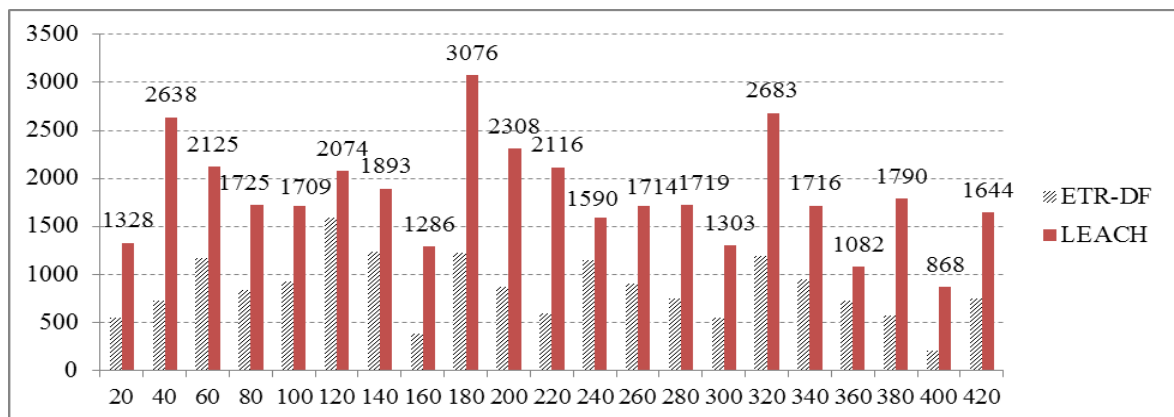
Hình 3.2. Các vùng ưu tiên và các mức ưu tiên

Vùng ưu tiên mức 1 (*A-Prio1*) ở Hình 3.2 a) là hình chữ nhật với diện tích  $(d_{CH-Tag} * 2\delta)$ , tọa độ 4 đỉnh  $(-x_{CH} + x_{tag})/2, -\delta)$ ,  $((x_{CH} + x_{tag})/2, -\delta)$ ,  $((x_{CH} + x_{tag})/2, \delta)$ ,  $(-x_{CH} + x_{tag})/2, \delta)$ . Mức ưu tiên 2 (*A-Prio2*) là hình vành khăn giới hạn bởi 2 đường tròn tâm  $O$  xem Hình

3.2 b), bán kính ( $d_{CH-Tag}/2$ ) -  $\delta$  (giới hạn bên trong đường tròn) và tâm  $O$ , bán kính ( $d_{CH-Tag}/2$ ) +  $\delta$  (giới hạn bên ngoài đường tròn). Vùng ưu tiên mức 3 ( $A-Prio3$ ) là vùng giới hạn bởi Elip xem Hình 3.2 c). Elip có 2 tiêu điểm CH, Tag. Như vậy, vùng để chọn nút là hợp của ( $A-Prio1$ ,  $A-Prio2$ ,  $A-Prio3$ ). Tất nhiên sẽ có trường hợp 1 nút cảm biến sẽ thuộc 2 (hoặc 3) vùng ưu tiên, khi đó việc lựa chọn nút sẽ dựa vào độ cân bằng giữa mức độ ưu tiên và các thuộc tính khác của nút như năng lượng còn lại của nút, số gói tin cần truyền đến CH,...

Bảng 3.1. Hiệu quả việc giảm gói tin của ETR-DF và LEACH

<b>Time (s<sup>th</sup>)</b>	<b>20</b>	<b>40</b>	<b>60</b>	<b>80</b>	<b>100</b>	<b>120</b>	<b>140</b>	<b>160</b>	<b>180</b>	<b>200</b>	<b>220</b>
<b>ETR-DF</b>	553	730	1175	833	928	1587	1236	377	1228	870	589
<b>LEACH</b>	1328	2638	2125	1725	1709	2074	1893	1286	3076	2308	2116
<b>Efficient (%)</b>	41.64	27.67	55.29	48.29	54.30	76.52	65.29	29.32	39.92	37.69	27.84
<b>Time (s<sup>th</sup>)</b>	<b>240</b>	<b>260</b>	<b>280</b>	<b>300</b>	<b>320</b>	<b>340</b>	<b>360</b>	<b>380</b>	<b>400</b>	<b>420</b>	
<b>ETR-DF</b>	1152	902	752	547	1193	950	722	569	204	747	
<b>LEACH</b>	1590	1714	1719	1303	2683	1716	1082	1790	868	1644	
<b>Efficient (%)</b>	72.45	52.63	43.75	41.98	44.47	55.36	66.73	31.79	23.50	45.44	



Hình 3.3. So sánh việc sử dụng năng lượng giữa ETR-DF và LEACH

Bằng cách phân tích số liệu đối với tất cả các cụm trong mỗi chu kỳ  $T = 20s$  và so sánh với thuật toán LEACH trong thời gian mô phỏng đến giây thứ 420 kết quả như sau: tại từng cụm, tỉ lệ nút được lựa chọn so với tổng số nút trong cụm dao động với khoảng rất lớn, từ 0% đến 100%. Tuy nhiên nếu tính trong từng chu kỳ  $T$  thì hiệu quả khoảng từ 23.5% đến 76.52%. Tổng hợp hiệu quả trung bình của các cụm theo chu kỳ  $T$  trong thời gian mô phỏng giữa ETR-DF và LEACH ở Bảng 3.1. Hiệu quả tiết kiệm năng lượng nhờ hạn chế việc gửi dữ liệu bằng sóng vô tuyến được biểu diễn ở Hình 3.3.

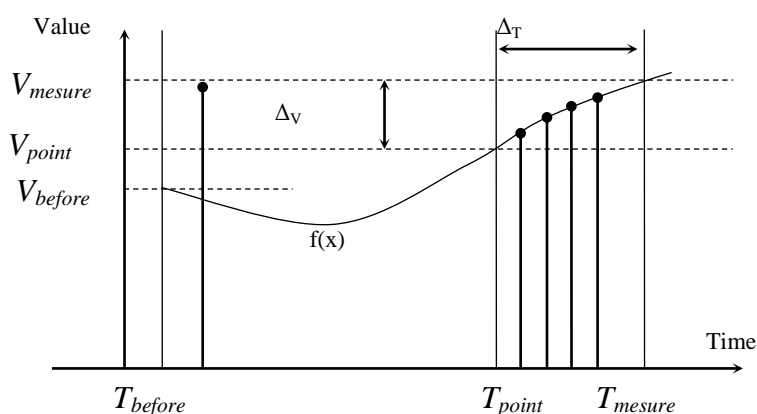
Thuật toán ETR-DF phát huy hiệu quả với điều kiện năng lượng dự trữ của các nút tương đối đồng đều. Khi đó, yếu tố về khoảng cách được xem là tiêu chí quyết định để lựa chọn. Ngoài ra, yêu cầu về độ tin cậy của nút cũng được đặt ra vì trong một số trường hợp, số liệu đo từ một nút có thể được xem là tốt hơn so với kết quả tổng hợp từ nhiều nút. Điều này cũng rất tự nhiên đối với vấn đề tổng hợp dữ liệu từ nhiều nút cảm biến không dây.

### 3.2. Theo dõi mục tiêu thích nghi theo thời gian ATTS-DF

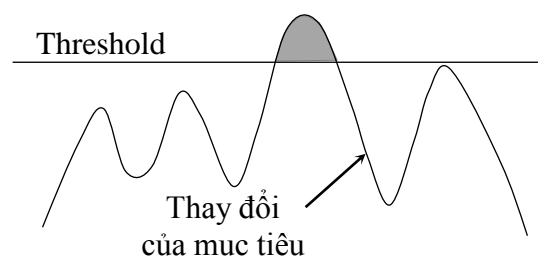
Giải pháp ATTS-DF (*Adaptive Target Tracking Solution for multi-sensor Data Fusion in WSNs*) đề xuất được một phương pháp theo dõi mục tiêu thích nghi với biến động của mục tiêu, khi mục tiêu biến động vượt ngưỡng, nút đang ở chế độ chờ và ngủ sẽ được bật; đề xuất một số khái niệm: Điểm đo biến động, trạng thái ổn định đo lường, thời gian đo thích ứng và phương pháp dự đoán giá trị đo mục tiêu theo xác suất giả thiết đã biết trước.

#### 3.2.1. Điểm đo biến động

Giải pháp sử dụng 3 mốc thời gian  $T_{before}$ ,  $T_{point}$ ,  $T_{measure}$ . Trong đó,  $T_{before}$  là mốc thời gian nút bắt đầu chuyển sang trạng thái *idle* (hoặc *sleep*);  $T_{point}$  là thời điểm nút xem xét để chuyển trạng thái từ *idle* (hoặc *sleep*) sang sẵn sàng đo lường và được **gọi là điểm đo biến động** (sau mốc thời gian này, việc theo dõi biến động của mục tiêu được sử dụng để điều khiển quyết định chuyển trạng thái của nút);  $T_{measure}$  là thời điểm quyết định đo lường. Các mốc thời gian và trạng thái làm việc của nút ở Hình 3.4.



Hình 3.4. Các mốc thời gian và trạng thái làm việc của nút



Hình 3.5. Thay đổi của thuộc tính khi vượt ngưỡng

#### 3.2.2. Thời gian đo thích ứng

Giả sử giá trị đo mục tiêu của nút cảm biến có thể biểu diễn bởi hàm số  $f(t)$  theo thời gian (*time*). Tại thời điểm  $T_{point}$ , giá trị đo là  $f(T_{point}) = V_{point}$  có xu hướng thay đổi giá trị để đạt  $f(T_{measure}) = V_{measure}$  tại thời điểm  $T_{measure}$ . Độ chênh lệch  $\Delta_V = |V_{measure} - V_{point}| \geq 0$ . Gọi thời gian đo thích ứng là  $\Delta_T$ . Trong khoảng  $\Delta_T$ ,  $f(t)$  giảm khi  $V_{measure} < V_{point}$  và  $f(t)$  tăng khi  $V_{measure} > V_{point}$ . Không mất tính tổng quát, giả sử  $f(t)$  tăng như ở Hình 3.4, khi đó hàm giá trị thích ứng là  $f(t) \in [V_{measure}, V_{point}]$  với  $\forall t \in [T_{measure}, T_{point}]$ .

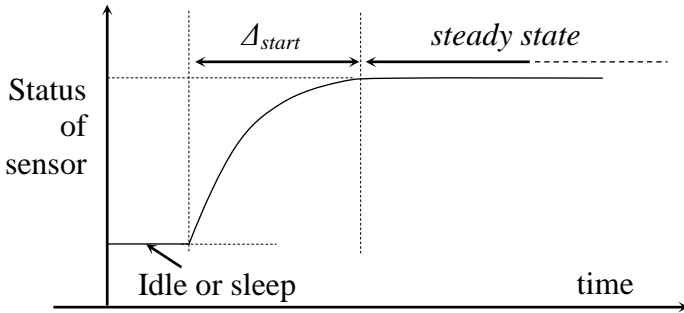
#### 3.2.3. Ngưỡng đo

Ngưỡng đo thể hiện trên 02 đối tượng đó là mục tiêu cần đo lường và nút (là thiết bị đo). Một mục tiêu có thể có nhiều thuộc tính đo lường ví dụ độ âm, nhiệt độ, ánh sáng... nút cũng phải có khả năng đo lường các thuộc tính đó. Khi mục tiêu xảy ra sự kiện và một thuộc tính cần đo lường của mục tiêu có biến động với biên độ (độ lệch) vượt một giá trị nào đó đặt trước và xem giá trị đó như là một ngưỡng thì nút sẽ đo lường (xem Hình 3.5). Khi biến động của mục tiêu làm cho giá trị đo của nút dưới ngưỡng thì nút vẫn ở trạng thái không đo lường (là trạng thái “*idle*” hoặc “*sleep*”). Giải pháp đề xuất giá trị ngưỡng đo lường (*threshold*) của nút, ký hiệu là  $\delta$ .

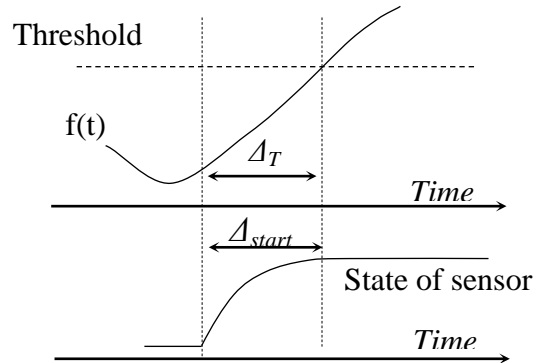


### 3.2.4. Trạng thái ổn định đo lường

Đo lường mục tiêu của nút cảm biến dựa trên việc điện tử hóa các đại lượng không có tính chất điện thành các đại lượng có thể đo và xử lý được bằng tín hiệu điện tử. Các kích thích của mục tiêu sẽ tác động đến bộ phận cảm nhận của nút, bộ phận này là linh kiện điện tử vì vậy có độ trễ (*response time*) nhất định  $\Delta_{start}$  để có thể đạt trạng thái hoạt động ổn định (*steady state*) như Hình 3.6.



Hình 3.6. Mô hình chuyển trạng thái của nút cảm biến

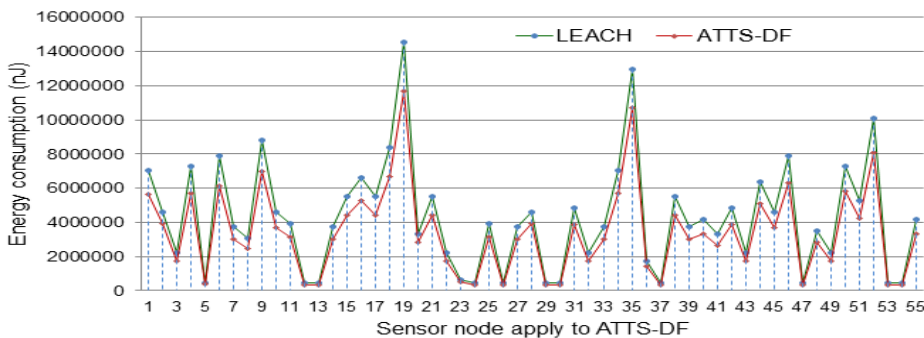


Hình 3.7. Mô hình trạng thái thích ứng của giải pháp ATTS-DF

### 3.2.5. Dự đoán

Giả sử độ trễ đó là  $\Delta_{start}$ . Giải pháp ATTS-DF hướng đến mối quan hệ giữa  $\Delta_T$  (thời gian đo thích ứng) và  $\Delta_{start}$ . Tính thích ứng (*adaptive*) của giải pháp này đạt lý tưởng khi  $\Delta_T = \Delta_{start}$ . Các trường hợp khác, đặt  $\Delta_{Adap} = |\Delta_T - \Delta_{start}|$ . Với  $\delta$  là giá trị ngưỡng đo lường, khi đó  $f(T_{measure}) = V_{measure} = \delta$ . Giả thiết biết trước hàm giá trị  $f(t)$  sao cho trong khoảng  $\Delta_T$  thì  $f(t)$  tăng theo giả thiết và  $f(T_{measure}) \geq \delta$  đồng thời  $\Delta_T = \Delta_{start}$ . Trạng thái thích ứng có nghĩa là nút khởi động và đạt trạng thái đo bình thường đúng lúc mục tiêu biến động vượt ngưỡng đo buộc nút phải đo lường (xem Hình 3.7).

Đánh giá hiệu quả bằng mô phỏng việc tiết kiệm năng lượng đối với các nút áp dụng ATTS-DF và so sánh với việc sử dụng năng lượng của các nút tương ứng khi áp dụng LEACH. Kết quả cho thấy, ATTS-DF tiết kiệm được từ 13,3% đến 20% năng lượng của nút tương ứng khi áp dụng LEACH. Biểu đồ tiêu thụ năng lượng của các nút áp dụng thuật toán ATTS-DF và LEACH ở Hình 3.8.



Hình 3.8. Mức thụ năng lượng của các nút giữa ATTS-DF và LEACH

## Kết luận về giải pháp ATTS-DF

Giải pháp ATTS-DF đề xuất được một phương pháp theo dõi mục tiêu theo thời gian, thích nghi với biến động của mục tiêu; đề xuất được khái niệm: Đo lường biến động, trạng thái ổn định đo lường, thời gian đo thích ứng và phương pháp dự đoán giá trị đo mục tiêu theo xác suất (đã biết trước). Hiệu quả của ATTS-DF so với LEACH gồm: *thứ nhất*, đo lường mục tiêu không theo chu kỳ cố định mà có điều chỉnh theo mục tiêu đã hạn chế được dung lượng dữ liệu đo lường giống nhau và tiết kiệm được năng lượng do không gửi dữ liệu dư thừa (vì có cùng thông tin) này đến CH, BS; *thứ hai*, đề xuất việc chuyển trạng thái đo lường của nút cảm biến từ “idle” (hoặc “sleep”) sang “active” đúng vào thời điểm nút cảm biến có thể đo lường ở trạng thái bình thường, điều này đã hạn chế tối đa thời gian nút cảm biến được bật và hoạt động nhưng chưa đạt trạng thái đo lường tốt nhất gây tổn hao năng lượng vô ích. Ngoài ra, giải pháp đã hướng đến việc đảm bảo độ hội tụ về thời gian.

### 3.3. Lựa chọn dữ liệu bằng cửa sổ trượt

Giải pháp này cài đặt trên CH để CH lựa chọn nút cảm biến trong cụm dựa trên ngữ nghĩa của nút cảm biến như khoảng cách, năng lượng còn lại, số gói tin cần truyền... Để tiết kiệm năng lượng của nút, CH có thể yêu cầu các nút trong cụm gửi *header* của nó với dung lượng nhỏ làm thuộc tính điều kiện. Từ các kết luận về ngữ nghĩa, CH sẽ lựa chọn nút thỏa mãn điều kiện và sử dụng dữ liệu đo của các nút đó phục vụ tổng hợp dữ liệu. Ngoài ra, thông tin về ngữ nghĩa giúp CH điều chỉnh kích thước cửa sổ trượt theo số lượng nút trong cụm để thay vì CH sẽ lấy và tổng hợp tất cả dữ liệu từ các nút trong cụm thì chỉ lấy và tổng hợp dữ liệu từ một số nút tại thời điểm xem xét đã thỏa mãn điều kiện.

#### 3.3.1. Cửa sổ trượt, kích thước cửa sổ trượt

Gọi  $\ell$  là hệ số trượt của cửa sổ, kích thước cửa sổ là bộ  $(H_{sw}, W_{sw})$  trong đó  $H_{sw}$  là chiều cao,  $W_{sw}$  là chiều rộng của cửa sổ.  $H_{sw}$  thay đổi theo hệ số trượt  $\ell$ . Các hệ số  $\ell$ ,  $H_{sw}$ ,  $W_{sw}$  được tính như sau:

Hệ số trượt  $\ell$  là chiều cao của cửa sổ ( $H_{sw}$ ) và được tính toán và sử dụng một lần cho mỗi cụm (*cluster*) và cho mỗi vòng. Có nghĩa là trong một vòng, tương ứng với mỗi *cluster* sẽ có một  $\ell$ . Sau khi bảng dữ liệu thuộc tính đã được sắp xếp hoàn chỉnh, hệ số trượt  $\ell$  được tính theo 2 trường hợp:

- Nếu có từ một nửa số nút của nhóm trở lên có năng lượng lớn hơn năng lượng trung bình của nhóm thì:

$$\ell = N_S/2 \text{ nếu } N_S \text{ chẵn hoặc } \ell = (N_S+1)/2 \text{ nếu } N_S \text{ lẻ}$$

- Nếu có dưới nửa số nút của nhóm có năng lượng lớn hơn năng lượng trung bình của nhóm thì:

- ✓ Lựa chọn các nút của nhóm có năng lượng lớn hơn năng lượng trung bình của nhóm.
- ✓ Tính trung bình khoảng cách giữa các nút (đã chọn) so với CH.
- ✓  $\ell$  = số nút cảm biến có khoảng cách đến CH nhỏ hơn hoặc bằng khoảng cách trung bình.

#### 3.3.2. Dữ liệu để tổng hợp

Dựa vào hệ số trượt  $\ell$  để lựa chọn nút cảm biến. Bảng dữ liệu để tổng hợp có  $\ell$  hàng và  $N_{P.mes}$  cột, với  $N_{P.mes}$  là số tham số đo lường của nút. Như vậy, khi áp dụng DF-SWin, chỉ

một số nút thuộc cửa sổ có hệ số  $\xi$  được chọn và dữ liệu đo lường của các nút đó sẽ được gửi đến CH thay vì toàn bộ dữ liệu của tất cả các nút trong cụm sẽ gửi đến CH (của LEACH).

### Xử lý dữ liệu

Sau mỗi vòng hoạt động, mạng cảm biến có khả năng tự tổ chức lại, phân chia lại nhóm, bầu nút CH cho mỗi nhóm. Giải thuật DF-SWin gồm 2 pha. *Pha 1*, từ bảng dữ liệu thuộc tính chứa các thông tin header về nút cảm biến như: khoảng cách giữa nút và CH, năng lượng còn lại (*residual energy*) của nút, gói tin còn lại (*residual package*) cần phải truyền đến đích của nút. Dựa vào các giá trị đó để xác định cửa sổ trượt bao gồm  $\xi$ ,  $H_{sw}$ ,  $W_{sw}$ . *Pha 2*, từ cửa sổ để xác định các nút và giá trị các tham số đo phục vụ tổng hợp dữ liệu. Giải pháp này chưa đề cập đến việc tính toán để có kết quả tổng hợp dữ liệu cuối cùng trước khi gửi BS. Như vậy, kết quả của giải thuật là bộ dữ liệu đo lường từ một số nút trong cụm phục vụ tổng hợp dữ liệu tại nút CH.

Áp dụng giải thuật khi năng lượng của nút còn nhiều (so với mức  $2J$  lúc khởi tạo) ví dụ ở giây thứ 80 (mạng có 4 cụm) và khi năng lượng của nút đã tiêu tốn phần lớn, ví dụ ở giây thứ 320 (mạng có 4 cụm), kết quả tính toán hệ số  $\xi$ ,  $H_{sw}$ ,  $W_{sw}$  ở Bảng 3.2.

Bảng 3.2. Kết quả mô phỏng tại thời điểm 80 giây và 320 giây

Thời điểm (giây)	Chỉ số cụm	Số nút của cụm	Avg(res.E <sub>Si</sub> , n) (J)	n <sub>1</sub> {res.E <sub>Si</sub> ≥ Avg (res.E <sub>Si</sub> , n)}	ξ (Hws)	Dung lượng dữ liệu tại CH để DF (sig_size)		Năng lượng tiết kiệm (J)
						DF-SWin	LEACH	
80	1	48	1.73834295	40	24	179	474	0.059
	2	25	1.69793760	20	12	75	177	0.0204
	3	11	1.76963819	10	5	65	126	0.0122
	4	16	1.40296564	8	8	72	139	0.0134
320	1	22	0.25857710	9	6	89	323	0.0468
	2	20	0.49958504	7	3	54	374	0.064
	3	21	0.45585774	5	3	39	286	0.0494
	4	31	0.41489440	9	4	67	420	0.0706

Bảng 3.3. Kết quả áp dụng trong thời gian mô phỏng

Time (s <sup>th</sup> )	20	40	60	80	100	120	140	160	180	200	220
DF-SWin	120	223	226	179	90	143	318	101	270	50	138
LEACH	246	477	528	474	334	256	705	206	543	212	522
Efficient (%)	51.22	53.25	57.20	62.24	73.05	44.14	54.89	50.97	50.28	76.42	73.56
Time (s <sup>th</sup> )	240	260	280	300	320	340	360	380	400	420	
DF-SWin	78	90	102	65	67	72	15	54	18	15	
LEACH	301	391	474	258	420	499	115	315	135	216	
Efficient (%)	74.09	76.98	78.48	74.81	84.05	85.57	86.96	82.86	86.67	93.06	

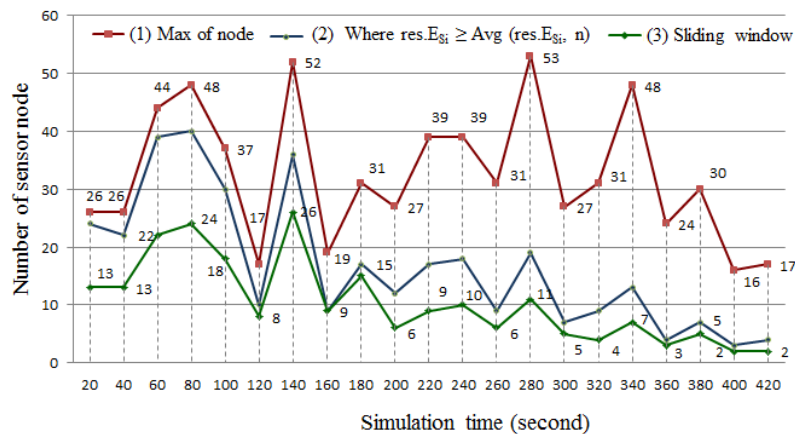
Ở giây thứ 80, ví dụ phân tích cụm số 1 có 48 nút,  $N_s = 47$  vì không bao gồm CH. Sau khi CH nhận và tách dữ liệu header ( $hdr\_size = 25$  Byte) từ các nút để làm đầy bảng dữ liệu. Sau đó áp dụng DF-SWin, số nút thỏa mãn  $res.E_{Si} \geq Avg(res.E_{Si}, n)$  là 40 bởi vì mạng mới

hoạt động, phần lớn nút cảm biến còn năng lượng dư trữ khá lớn so với mức năng lượng khởi tạo  $2J$ . Hệ số cửa sổ trượt  $\ell = 24$  (bởi vì  $N_s$  lẻ). Từ việc phân tích dữ liệu *header* và  $\ell$ , CH chỉ lựa chọn dữ liệu đo của các nút trong cửa sổ với dung lượng là  $179 \text{ data}$  thay vì  $474 \text{ data}$  của LEACH. Như vậy, trong trường hợp này DF-SWin tiết kiệm được so với LEACH là  $295 \text{ data}$  (62.23%). Với dung lượng một data là  $\text{sig\_size} = 500 \text{ Byte}$ , năng lượng tiết kiệm được cho cả cụm trong việc chuyển dữ liệu đến CH là:  $(295 \text{ sig\_size} * 500 \text{ Byte/sig\_size} * 8\text{bit} / \text{Byte} * 50\text{nJ} / \text{bit}) = 59.000.000 \text{ nJ} = 0.059 \text{ J}$ .

Tương tự, ở giây thứ 320, ví dụ phân tích cụm 1 với 22 nút,  $N_s = 21$ , vì  $n_l = 9$  nghĩa là chỉ có 9 sensor thỏa mãn  $\text{res.E}_{Si} \geq \text{Avg}(\text{res.E}_{Si}, n)$  và  $n_l < (N_s + 1)/2 = 11$  sensor, đưa 11 sensor này vào tập *Set1*. Tính trung bình khoảng cách của 11 sensor này so với CH theo  $\text{Avg}(\text{dist.S}_i, n_l) = 25$ . Có 6/9 sensor thỏa mãn điều kiện  $\text{dist.S}_i \leq \text{Avg}(\text{dist.S}_i, n_l)$  và được đưa vào tập *Set2*. Hệ số cửa sổ trượt  $\ell = 6$ . Số dữ liệu DF-SWin tiết kiệm được so với LEACH là  $234 \text{ sig\_size}$  (72.44%). Năng lượng tiết kiệm được của cụm 21 sensor bởi việc chỉ chọn 6/21 nút để gửi dữ liệu đến CH là:  $(234 \text{ sig\_size} * 500 \text{ Byte/sig\_size} * 8\text{bit} / \text{Byte} * 50\text{nJ} / \text{bit}) = 46.800.000 \text{ nJ} = 0.0468 \text{ J}$ .

Bảng 3.3 là khảo sát việc áp dụng DF-SWin đối với các cụm có nhiều nút nhất trong các vòng  $T = 20\text{s}$ . Kết quả mô phỏng cho thấy, giải thuật DF-SWin có xu hướng phát huy hiệu quả tiết kiệm năng lượng khi năng lượng của nút giảm và mạng còn ít nút hoạt động. Cụ thể, tại giây thứ 420, còn 50 nút hoạt động chia thành 5 cụm, cụm có nhiều nút nhất là 17 (bao gồm cả CH), khi áp dụng DF-SWin đã tiết kiệm được 93.06% năng lượng của cụm.

So sánh số lượng nút của cụm có số lượng nút lớn nhất, số lượng nút có mức năng lượng không thấp hơn năng lượng trung bình của cụm, số lượng nút cửa sổ trượt. Kết quả khảo sát trong 420 s ở Hình 3.9.



Hình 3.9. Kết quả mô phỏng đối với các phương án

### Kết luận về giải pháp DF-SWin

Ý tưởng của phương pháp DF-SWin dựa trên sự kết hợp một số khái niệm của lý thuyết tập thô (*Rough Set*), khai phá dữ liệu (*data mining*) và cửa sổ trượt (*sliding window*) trong giao thức TCP/IP. Giải pháp này đề xuất việc lựa chọn dữ liệu để tổng hợp thông qua thuộc tính điều kiện của nút trước khi lựa chọn giá trị tham số đo lường của nút. Để tiết kiệm năng lượng, CH có thể yêu cầu các nút trong cụm gửi *header* (có dung lượng nhỏ) để

làm thuộc tính điều kiện từ đó có thể điều chỉnh kích thước cửa sổ trượt và lựa chọn nút đồng nghĩa với việc lựa chọn số lượng *sensor* theo điều kiện, đây là giải pháp hiệu quả để tiết kiệm năng lượng của nút trong cụm.

Phương pháp DF-SWin phù hợp với mạng cảm biến gồm các nút hoạt động trong trạng thái mức năng lượng còn lại thấp và không đồng đều.

## **Chương 4. TỔNG HỢP DỮ LIỆU TẠI NÚT CỤM TRƯỞNG**

### **4.1. Lý thuyết tập thô và tổng hợp dữ liệu nhiều cảm biến**

#### **4.1.1. Cơ sở áp dụng**

Sự phù hợp để chọn RST làm giải pháp tổng hợp dữ liệu nhiều cảm biến thể hiện ở những quan điểm sau:

- Tính chất rời rạc của tín hiệu và xử lý liên tục của nút cảm biến: Giao thức IEEE 802.15.4 khi áp dụng cho mạng WSN sẽ điều khiển việc lấy dữ liệu theo chu kỳ thức-ngủ (*active-sleep*) nên dữ liệu CH thu được từ nút cảm biến sẽ rời rạc. Khi nút cảm biến ở trạng thái thức (*active*), nó sẽ xử lý và truyền dữ liệu đến nút cảm biến tiếp theo trên tuyến thì dữ liệu đó là liên tục.

- Hỗ trợ để xử lý mô tả không chắc chắn: Khi sensor cảm nhận về đối tượng, tín hiệu có thể bị nhiễu dẫn đến tính chân lý của dữ liệu truyền đi không được bảo toàn. Dựa trên dữ liệu thuộc tính, CH có thể xác định lại sự đúng đắn của dữ liệu cảm nhận bằng cách loại bỏ thông tin nhiễu, giữ lại thông tin hữu ích, ít bị nhiễu phục vụ tổng hợp.

- Hỗ trợ xử lý vấn đề mất dữ liệu: Dữ liệu thu thập được từ các nút cảm biến khi truyền đến CH có thể không đầy đủ, nghĩa là CH không nhận đủ dữ liệu từ một hoặc nhiều nút trong nhóm để làm dữ kiện cho quá trình tổng hợp. Tình huống để mất dữ liệu có thể là: Lúc cần cảm nhận thì nút cảm biến đang trạng thái ngủ, lúc đang truyền dữ liệu đến CH thì nút cảm biến hết năng lượng, đang truyền thì đến chu kỳ ngủ của nút cảm biến...

- Hỗ trợ để xử lý vấn đề dư thừa dữ liệu: Đây là một vấn đề rất quan trọng trong bài toán tổng hợp. Khi các nút cảm biến cùng cảm nhận về một đối tượng và cùng truyền một loại thông tin đó trực tiếp đến BS hoặc qua nút cảm biến trung gian (là CH nếu mạng có phân cụm) để truyền đến BS thì việc loại bỏ các dữ liệu dư thừa này là điều rất cần thiết.

- RST hỗ trợ tổng hợp dữ liệu được chính xác hơn thông qua ngữ nghĩa, “tri thức” của thông tin chứ không thông qua xử lý trực tiếp toàn bộ dữ liệu “kiến thức” của thông tin. Trên thực tế, rất khó để có thể tổng hợp được dữ liệu đúng tuyệt đối (100%) với tình huống diễn ra ở thực địa do đặc tính phân cứng của sensor khó định lượng chính xác giá trị của đại lượng đặc trưng cho sự kiện cần giám sát. Do đó, đôi khi phải tính định lượng thông qua ngữ nghĩa của dữ liệu thay vì định lượng giá trị đo của từng loại tham số cụ thể.

#### **4.1.2. Quy trình ứng dụng RST để tổng hợp dữ liệu**

Với WSNs, nhiều lớp bài toán đã được đặt ra như: điều khiển chu kỳ cảm nhận dữ liệu, lấy mẫu cảm biến, phân nhóm, chọn nút cụm trưởng, định tuyến, tổng hợp... Không mất tính tổng quát có thể giả sử những bài toán đó đã được giải quyết. Riêng bài toán tổng hợp dữ liệu, mục tiêu chính chỉ tập trung vào vấn đề (điều khiển) tính toán trên nút CH. Như

vậy, ứng dụng lý thuyết tập thô để tổng hợp dữ liệu nhiều cảm biến sẽ được thực hiện ở Lớp 3 (xem Hình 4.1). Quy trình ứng dụng RST để tổng hợp dữ liệu được đề xuất như sau:

Bước 1: Tiền xử lý, tách thuộc tính từng nút cảm biến (của dữ liệu cảm nhận và của nút cảm biến), giả sử được  $m$  thuộc tính;

Bước 2: Lập bảng, nếu  $n$  nút cảm biến thì bảng sẽ có  $n$  hàng,  $(m + 1)$  cột;

Bước 3: Tìm lớp con tương đương theo giá trị thuộc tính quyết định;

Bước 4: Tìm tập các lớp con tương đương của các tập con thuộc tính điều kiện;

Bước 5: Tìm các tập xấp xỉ trên, xấp xỉ dưới, tập biên, hệ số xấp xỉ;

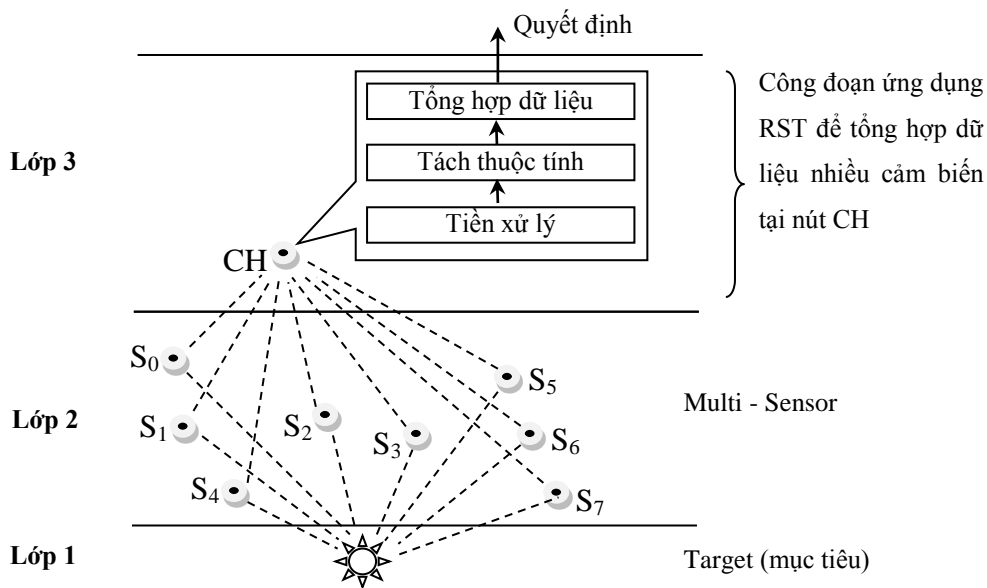
Bước 6: Tìm tập các tập thuộc tính rút gọn;

Bước 7: Tìm tập thuộc tính lõi;

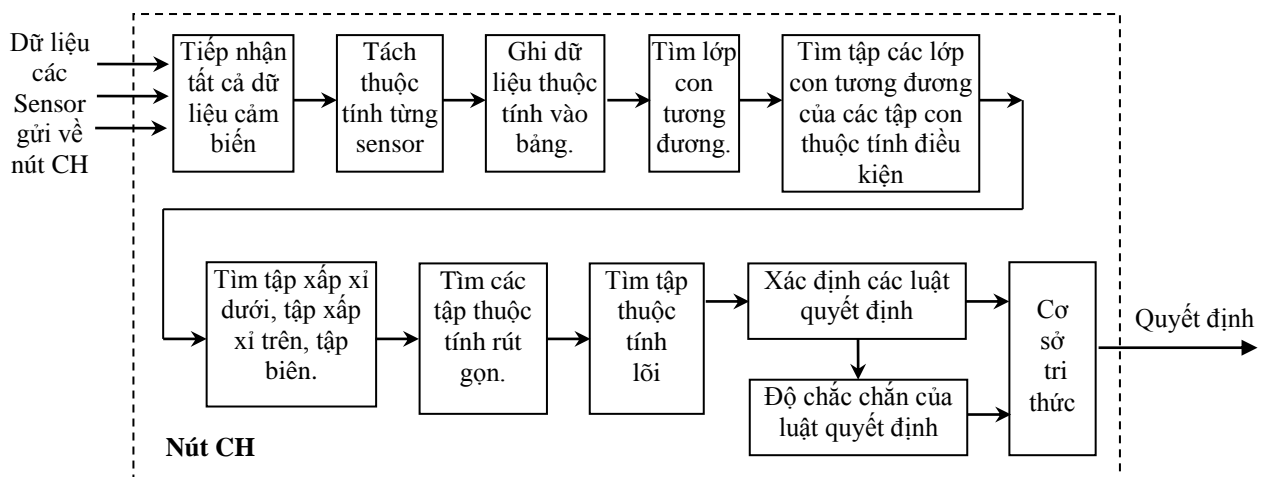
Bước 8: Xác định các luật quyết định, độ chắc chắn của mỗi luật để làm cơ sở tri thức;

Bước 9: Quyết định.

Đề xuất mô hình xử lý dữ liệu tại CH như sau (xem Hình 4.2):



Hình 4.1. Mô tả bài toán DF có sử dụng RST.



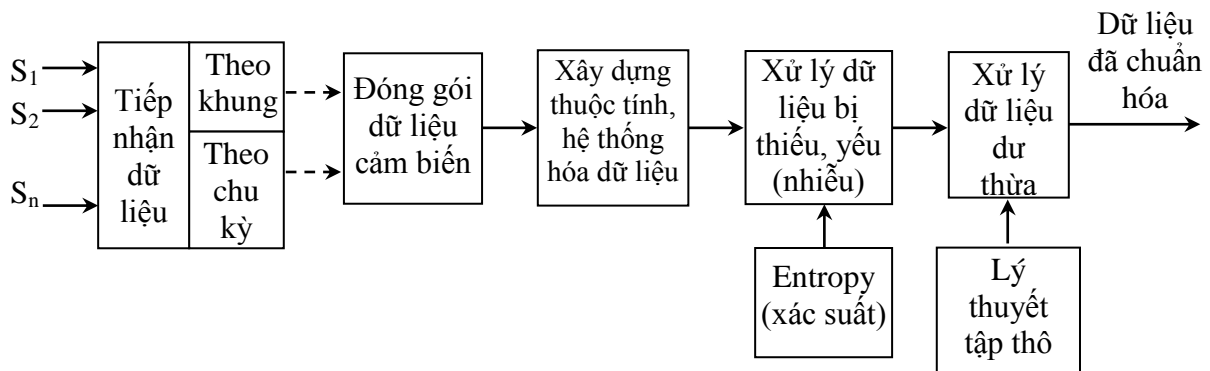
Hình 4.2. Mô hình xử lý, tổng hợp dữ liệu tại nút CH

Việc tìm tập thuộc tính rút gọn có ý nghĩa quyết định đối với vấn đề ứng dụng RST để tổng hợp. Tối ưu các cột trong bảng thông tin là một bài toán khó, có độ phức tạp hàm mũ của thuộc tính điều kiện. Xác định *tập thuộc tính lõi, các luật quyết định và độ chắc chắn của luật quyết định là cơ sở để* lọc “kiến thức” thành “tri thức”, hỗ trợ CH quyết định.

## 4.2. Tiền xử lý dữ liệu đầu vào

### 4.2.1. Quy trình xử lý dữ liệu

Sau khi CH đóng gói dữ liệu cảm biến theo khung tin hoặc theo chu kỳ, CH sẽ sử dụng dữ kiện này làm đầu vào để áp dụng giải pháp tiền xử lý. Tương tự kỹ thuật tiền xử lý trong khai phá dữ liệu (*data mining*) giai đoạn tiền xử lý tại nút CH gồm các công đoạn và thứ tự xử lý như ở Hình 4.3:



Hình 4.3. Quá trình tiền xử lý dữ liệu cảm biến tại nút CH của DP-DF

- *Xây dựng thuộc tính (attribute/feature construction)*: Là các thuộc tính ngữ nghĩa của nút cảm biến và ngữ nghĩa của dữ liệu cảm nhận. Thuộc tính là các cột của bảng dữ liệu cảm biến.

- *Hệ thống hóa dữ liệu*: là quá trình nhận diện đặc điểm chung của dữ liệu cảm biến và sự hiện diện của dữ liệu nhiễu, dữ liệu thiếu hoặc các phần tử kì dị (*outliers*) khi nút cảm biến đo lường; định lượng hóa thành giá trị để đưa vào bảng dữ liệu gồm  $n$  hàng,  $m$  cột tương ứng với  $n$  nút cảm biến của mạng và  $m$  thuộc tính của mỗi nút cảm biến.

- *Xử lý dữ liệu bị thiếu (missing data)*: Khi CH không nhận đủ dữ liệu từ một hoặc nhiều nút trong nhóm để làm dữ kiện cho quá trình tổng hợp dữ liệu. Dữ liệu bị thiếu có thể là dữ liệu đo của tất cả các tham số đo về mục tiêu hoặc của một vài tham số đo thành phần của mục tiêu. Do đó, xử lý dữ liệu bị thiếu là bước quan trọng trong giai đoạn tiền xử lý. Luận án áp dụng xác suất để xử lý dữ liệu bị thiếu.

- *Xử lý dữ liệu bị nhiễu (noisy data)*: Khi nút cảm biến cảm nhận về mục tiêu, tín hiệu có thể bị nhiễu dẫn đến tính chân lý của dữ liệu truyền đi không được bảo toàn. Tiền xử lý tại CH có thể xác định lại sự đúng đắn của dữ liệu cảm nhận bằng cách loại bỏ thông tin nhiễu thay thế giá trị nhiễu theo xác suất, giữ lại thông tin hữu ích, ít bị nhiễu..

- *Xử lý dữ liệu dư thừa (redundancy)*: Đây là một vấn đề rất quan trọng trong bài toán tổng hợp dữ liệu. Khi các nút cảm biến cùng cảm nhận về một đối tượng và cùng truyền một

loại thông tin đó trực tiếp đến BS hoặc qua nút cảm biến trung gian (là CH nếu mạng có phân cụm) để truyền đến BS thì việc loại bỏ các dữ liệu dư thừa này là điều rất cần thiết.

#### 4.2.2. Ví dụ minh họa

Giả sử một mạng cảm biến có 5 nút cảm biến  $S_1 \rightarrow S_5$  với 4 thuộc tính  $A_1 \rightarrow A_4$  với kết quả đo được ở Bảng 4.1. Sau tiên xử lý được kết quả ở Bảng 4.2.

Bảng 4.1. Dữ liệu đo của mạng cảm biến

Nút cảm biến	Thuộc tính			
	$A_1$	$A_2$	$A_3$	$A_4$
$S_1$	$\emptyset$	5	$\sim 2$	4
$S_2$	3	4	$\emptyset$	4
$S_3$	3	3	3	4
$S_4$	4	$\emptyset$	4	4
$S_5$	$\emptyset$	$\sim 2$	4	$\emptyset$

Bảng 4.2. Các phương án dữ kiện đã được rút gọn để tổng hợp dữ liệu

(a). Dữ liệu RG1

Nút	$A_1$	$A_2$
$S_1$	4	5
$S_2$	3	4
$S_3$	3	3
$S_4$	4	4
$S_5$	6	1

(b). Dữ liệu RG2

Nút	$A_1$	$A_3$
$S_1$	4	1
$S_2$	3	4
$S_3$	3	3
$S_4$	4	4
$S_5$	6	4

(c). Dữ liệu RG3

Nút	$A_1$	$A_3$	$A_4$
$S_1$	4	1	4
$S_2$	3	4	4
$S_3$	3	3	4
$S_4$	4	4	4
$S_5$	6	4	5

#### Kết luận về giải pháp DF-DP

Tiên xử lý dữ liệu nhằm hướng đến bộ dữ liệu đầu vào tốt nhất để CH tổng hợp là một trong những bài toán con của mô hình tổng hợp dữ liệu. Ví dụ, Bảng 4.2 gồm 03 phương án dữ liệu đầu vào CH, vừa đảm bảo vấn đề tối giản trong xử lý dữ liệu, vừa tạo ra bộ dữ liệu gồm nhiều sự lựa chọn cho CH. Một bộ dữ liệu đầu vào CH chất lượng, bảo toàn dữ liệu là điều kiện cần để việc tổng hợp dữ liệu tại CH có hiệu quả. Tùy thuộc vào mục đích và thuật giải cụ thể để CH lựa chọn dữ liệu đầu vào đảm bảo cân bằng giữa tính toán và bảo toàn dữ liệu cảm nhận của nút cảm biến trong cụm.

Đối với giải pháp tiên xử lý dữ liệu nhiều cảm biến của mạng cảm biến không dây bằng lý thuyết tập thô nói chung và các phương pháp khác sử dụng bảng dữ liệu nói riêng thì việc tìm bảng rút gọn (bảng cách tìm thuộc tính rút gọn, thuộc tính lõi) là một trong những bước thực hiện quan trọng làm cơ sở để giảm không gian tính toán đối với dữ liệu.

#### 4.3. Tổng hợp dữ liệu tại nút CH

Giải pháp tổng hợp dữ liệu kết hợp cực đại, trung bình, trung vị được đặt tên là DF-AMS (*Data Fusion – Average Median Sampling*) gồm 02 công đoạn chính: lấy mẫu nút cảm biến (trong cụm) thông qua thuộc tính của nút cảm biến; CH tổng hợp dữ liệu nhiều tham số của những nút đã được lựa chọn bằng phương pháp trung vị (*median*) và cực đại (*max*) các độ đo của nút.

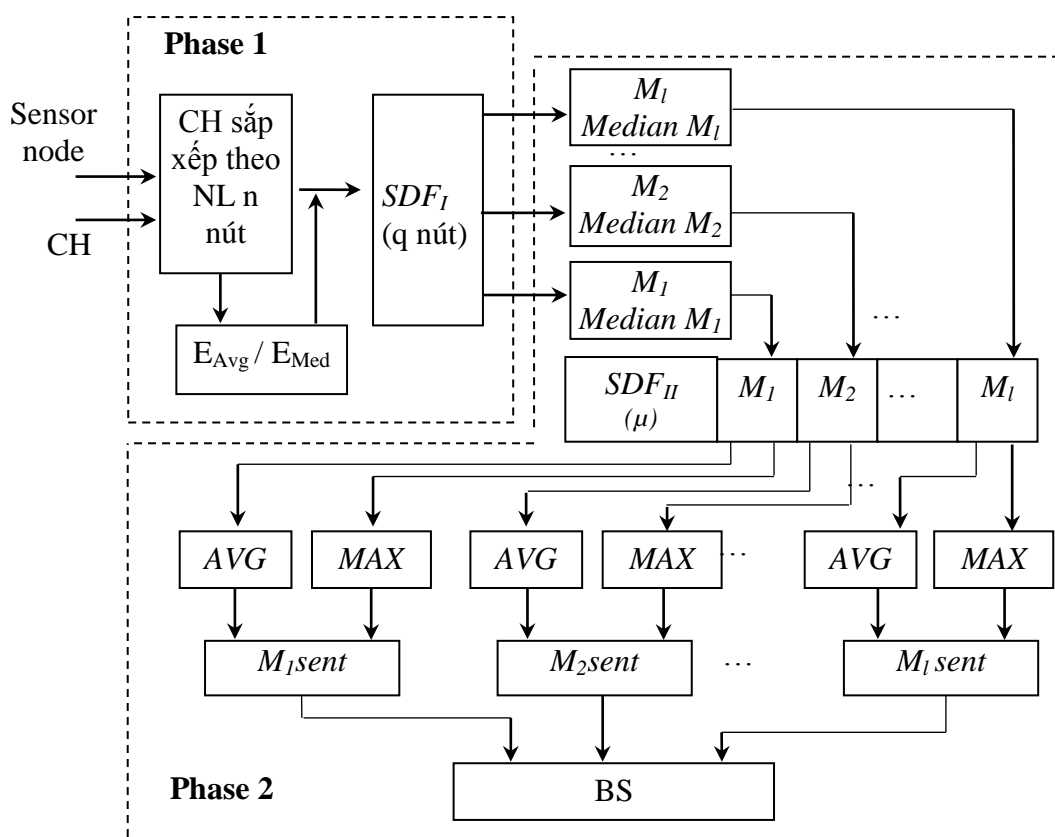


### Mô hình xử lý dữ liệu

Thuật toán gồm 2 pha xử lý chính ở Hình 4.4:

*Pha 1*, ngay sau khi cụm được thiết lập, dựa vào năng lượng của nút để CH chọn các nút có năng lượng lớn hơn (lấy mốc là mức trung bình hoặc trung vị) đưa vào tập  $SDF_I$ , chỉ các nút được chọn mới tiếp tục gửi dữ liệu cảm nhận đến CH.

*Pha 2*, ngay trước thời điểm kết thúc vòng, CH lựa chọn các kết quả đo  $M_i$  của  $q$  nút, sắp theo chiều tăng giá trị đo lường và chọn các *Median*  $M_i$  tương ứng. Lựa chọn nút theo giá trị đo *Median*  $M_i$  đưa vào tập  $SDF_{II}$  (có  $\mu$  nút). CH sẽ tổng hợp giá trị theo trung bình và Max theo từng tham số đo  $M_i$  và gửi kết quả này đến BS.



Hình 4.4. Mô hình xử lý dữ liệu của DF-AMS

Giải pháp DF-AMS có giá trị tổng hợp tiệm cận giá trị lớn nhất có xu hướng đảm bảo tính đúng đắn của kết quả đo lường; CH lọc dữ liệu thô và loại bỏ các giá trị đo yếu (chứa ít thông tin) và chỉ gửi đi 1 kết quả duy nhất đến BS đã mang lại hiệu quả lớn về tiết kiệm năng lượng của nút cảm biến trong cụm trong vòng đó. Do vậy DF-AMS phù hợp với mạng cảm biến có nhiều nút, năng lượng của nút không đồng đều, cảm biến đo nhiều tham số của mục tiêu nhưng giá trị đo có sự khác biệt cao.

## KẾT LUẬN

Luận án đề xuất được mô hình tổng hợp dữ liệu nhiều nút cảm biến trong mạng cảm biến không dây gồm 02 nhóm bài toán với một số đề xuất, cải tiến cụ thể:

### 1.1. Nhóm bài toán theo dõi mục tiêu và lựa chọn dữ liệu

1.1.1. Đề xuất giải pháp theo dõi để lấy dữ liệu thích nghi với biến động của mục tiêu. Nút chỉ đo lường khi biến động mục tiêu có xu hướng vượt ngưỡng. Thuật toán đề xuất việc chuyển trạng thái đo lường khi giá trị đo mục tiêu vượt ngưỡng của nút từ *idle* (hoặc *sleep*) sang *active* đúng vào thời điểm nút có thể đo lường ở trạng thái bình thường bởi vì để đạt trạng thái này, sensor phải mất một khoảng thời gian nhất định để khởi động.

1.1.2. Đề xuất giải pháp theo dõi mục tiêu và ưu tiên tổng hợp dữ liệu từ các nút nằm trên đường đi ngắn nhất giữa mục tiêu theo dõi và CH. Đề xuất này hạn chế được đáng kể các sensor ở xa mục tiêu và ở xa CH tham gia quá trình theo dõi, truyền dữ liệu đến CH.

1.1.3. Đề xuất giải pháp lựa chọn một số nút trong cụm bằng cơ chế cửa sổ trượt. Xuất phát từ ý tưởng sử dụng cửa sổ trượt để truyền dữ liệu trong giao thức TCP/IP, giải pháp DF-SWin sử dụng cửa sổ trượt 1 chiều theo số lượng nút trong cụm. Tại mỗi thời điểm sử dụng, kích thước cửa sổ có thể thay đổi tùy thuộc số lượng nút trong cụm, năng lượng của mỗi nút, khoảng cách từ các nút trong cụm đến CH, số gói tin nút cần phải truyền đến CH ....

### 1.2. Nhóm bài toán tiền xử lý dữ liệu và tổng hợp dữ liệu tại nút CH.

1.2.1. Đề xuất tiếp cận phương pháp tổng hợp dữ liệu nhiều cảm biến trong mạng cảm biến không dây bằng lý thuyết tập thô. Đề tài đã chứng minh sự phù hợp để chọn lý thuyết tập thô - RST (*Rough Set Theory*) làm giải pháp tổng hợp dữ liệu nhiều cảm biến.

1.2.2. Đề xuất giải pháp tiền xử lý dữ liệu tại nút CH, xử lý dữ liệu bị mất (giá trị null), dữ liệu ít giá trị (ít chứa tri thức) và dữ liệu dư thừa (là các giá trị giống nhau khi đo cùng 1 tham số mục tiêu). Kết quả đầu ra sau khi tiền xử lý là bảng dữ liệu đầy đủ (không chứa giá trị null), rút gọn (cả hàng và cột) làm bộ dữ liệu đầu vào CH phục vụ tổng hợp.

1.2.3. Đề xuất việc kết hợp giải pháp lựa chọn nút và tính toán dữ liệu tại nút CH. Việc lựa chọn nút thông qua ngữ nghĩa là đặc tính vật lý của nút cảm biến (như năng lượng còn lại, khoảng cách đến CH, khoảng cách so với láng giềng, chỉ số khe truyền tin TDMA do CH cấp...) và đặc tính dữ liệu cảm nhận (như số gói tin cần truyền để hoàn tất thông tin cảm nhận, cường độ tín hiệu, nhiễu...) để CH loại bỏ nút cảm biến trong cụm, chỉ lấy dữ liệu của một số nút để tổng hợp bằng việc sử dụng hợp lý và kết hợp các tính toán đơn giản như *Min*, *Max*, *Average*, *Median* tùy thuộc vào giá trị năng lượng của nút và của mạng. Trong mỗi chu kỳ, sau khi tổng hợp dữ liệu mỗi tham số đo của tất cả các nút trong cụm được tổng hợp thành 1 kết quả tương ứng, kết quả này được gửi đến BS. Như vậy, trong 1 chu kỳ đo, nếu cụm có  $N$  nút, mỗi nút có thể đo lường  $M$  tham số về mục tiêu thì sau khi tổng hợp, CH chỉ gửi  $M$  kết quả đến BS.

## DANH MỤC CÔNG TRÌNH CỦA TÁC GIẢ

1. Nguyen Duy Tan, Ho Duc Ai, **Duong Viet Huy**, Nguyen Dinh Viet, “*An Improved LEACH Routing Protocol for Energy-Efficiency of Wireless Sensor networks*”, Proceedings of the 6<sup>th</sup> National Conference on Fundamental and Applied IT Research (**FAIR’6, 2013**), pp. 33-39.
2. **Dương Viết Huy**, Nguyễn Duy Tân, Hồ Đức Ái, Nguyễn Đình Việt, “*Tiếp cận phương pháp tổng hợp dữ liệu nhiều cảm biến trong mạng cảm biến không dây bằng lý thuyết tập thô*”, Kỷ yếu Hội nghị quốc gia lần thứ VII về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (**FAIR 2014**), 19-20 June 2014, Thái Nguyên, Việt Nam. ISBN: 978-604-913-300-8, DOI 10.15625/FAIR VII.2014-0397, pp. 668-677.
3. **Dương Viết Huy**, Nguyễn Đình Việt, “*Đề xuất giải pháp tổng hợp dữ liệu nhiều cảm biến trong mạng cảm biến không dây*”, Kỷ yếu Hội thảo Quốc gia - @ XVII (**2014**) chủ đề Một số vấn đề chọn lọc của Công nghệ thông tin và truyền thông, trang 50-55.
4. **Duong Viet Huy**, Nguyen Dinh Viet, “*DF-AMS: Proposed solutions for multi-sensor data fusion in wireless sensor networks*”, 2015 Seventh International Conference on Knowledge and Systems Engineering (**KSE 2015**), ISBN 978-1-4673-8013-3/15, DOI 10.1109/KSE.2015.28, pp 1-6. (*Scopus, IEEE Xplore*).
5. **Dương Viết Huy**, Nguyễn Đình Việt, “*Đề xuất giải pháp tiên xử lý để tổng hợp dữ liệu nhiều cảm biến trong mạng cảm biến không dây*”, Proceeding: the 8th National Conference on Fundamental and Applied IT Research, **2015**. ISBN: 978-604-913-397-8, DOI: 10.15626/vap.2015.000149, pp 165-170.
6. **Duong Viet Huy**, Nguyen Dinh Viet, “*Target tracking solution for multi-sensor data fusion in wireless sensor networks*”, VNU Journal of Science: Comp. Science & Com. Eng., Vol. 32, No. 2 (**2016**) 63-71.
7. **Duong Viet Huy**, Nguyen Dinh Viet, “*ATTS-DF: Adaptive tracking solution to the target for data fusion in wireless sensor networks*”, International Conference on Systems Science and Engineering 2017 (**ICSSE 2017**), Ho Chi Minh city, Vietnam, July 21-23, 2017. ISBN: 978-1-5386-3421-9. DOI: 10.1109/ICSSE.2017.8030925, pp 526-531. (*Scopus, IEEE Xplore*).
8. **Huy. Duong-Viet**, Viet. Nguyen-Dinh, “*DF-SWin: Sliding Windows for Multi-Sensor Data Fusion in Wireless Sensor Networks*”, The 9th International Conference on Knowledge and Systems Engineering (**KSE 2017**), Hue, Vietnam October 19-21, 2017. DOI: 10.1109/KSE.2017.8119434. Pp 62-67 (*IEEE Xplore*).