

# CÁC GIAO TÁC TRONG THÔNG TIN DI ĐỘNG ỨNG DỤNG THUẬT TOÁN DỰ BÁO VỊ TRÍ TRONG MẠNG DI ĐỘNG

■ Lê Mạnh\*, Giang Minh Đức\*\*, Đỗ Hồng Tuấn\*\*\*

## TÓM TẮT

*Bài báo giới thiệu một vài nghiên cứu của các tác giả trong môi trường di động phục vụ công nghệ mobil internet. Ứng dụng công nghệ Data mining để dự báo vị trí các thuê bao di động, chúng tôi đề xuất một phương pháp cải tiến tính toán sẽ làm giảm thời gian để tính toán mô hình di động. Các nghiên cứu này có thể áp dụng trong giảng dạy cho sinh viên các trường đại học ngành công nghệ.*

*Từ khóa: Giao tác của thuê bao di động, dự báo vị trí, mạng GSM, khai phá dữ liệu.*

## ABSTRACT

### Transactions of mobile network apply location prediction algorithms for mobile networks

*This paper presents some of the author's research in environmental mobile internet mobil. Apply Data mining technology to predict the location of mobile user; we propose an improved method to reduce computational time the cell model. These studies can be applied in teaching students from other university of technology.*

*Keywords: Transactions of MOBILE User; location prediction, GSM network, data mining.*

## I. Giới thiệu tổng quan

Hiện nay với sự phát triển nhanh chóng của các công nghệ truyền thông không dây (wireless communications) và các thiết bị di động đầu cuối hiện đại đã tạo nên một môi trường trao đổi dữ liệu mới, đó chính là môi trường làm việc di động (Mobile Work Environment).

Môi trường làm việc di động khác với môi trường làm việc cố định ở một số đặc điểm sau:

- Sự di chuyển của người sử dụng (Mobility).
- Hay bị ngắt kết nối (Disconnection) do môi trường truyền sóng (bị che chắn hoặc sóng vô tuyến không phủ hết,...).
- Tiến trình chuyển giao kênh truyền dẫn (hand-over) giữa các tế bào trong mạng di động tổ ong (Cellular), tiến trình này sẽ ngắt kết nối (đối với truyền số liệu) một khoảng thời gian ngắn (100ms ÷ 200ms).

Mặt khác, môi trường di động là một môi trường mới, người ta có thể truy xuất dữ liệu và xử lý thông tin khi đang di chuyển. Với môi trường này, người ta có thể làm việc một cách uyển chuyển và linh hoạt. Tuy nhiên, để làm việc tốt trong môi trường này cần phải giải quyết các vấn đề về tranh chấp của các giao tác di động

(Mobile Transactions), thời gian trì hoãn dài, mất kết nối (Disconnect) với cơ sở dữ liệu trung tâm do sự chuyển giao kết nối (hand-over) của các trạm thu phát sóng di động,...

Theo Accenture: “Vào năm 2013 có khoảng 3 tỷ thiết bị di động truy cập vào internet, bỏ xa số lượng các máy tính cá nhân truy cập vào. Hơn 400 triệu thuê bao điện thoại trên toàn thế giới truy cập từ xa thông qua môi trường di động (3G, 4G) để giải quyết công việc” [10].

## II. Các giao tác di động

Trong môi trường làm việc di động, chúng ta luôn phải đối phó với các vấn đề như: vị trí luôn thay đổi của các thiết bị di động, chất lượng kết nối hay thay đổi của mạng di động, hay bị ngắt kết nối, giải quyết các xung đột và các giao tác trong môi trường di động. Người sử dụng có thể truy xuất dữ liệu và xử lý thông tin khi đang di chuyển một cách uyển chuyển và linh hoạt. Tuy nhiên, để làm việc tốt trong môi trường này cần phải giải quyết các vấn đề về tranh chấp của các giao tác di động, thời gian trì hoãn dài, mất kết nối với cơ sở dữ liệu trung tâm do sự chuyển giao kết nối của các trạm thu phát sóng di động.

Việc nâng cao chất lượng dịch vụ trên nền di

\*TS, Trường ĐH Văn Hiến, \*\*ThS, Công ty Viễn thông Bình Dương, \*\*\*TS, Trường ĐH Bách Khoa TP.HCM

động là một yêu cầu không thể bỏ qua.

Bài toán đặt ra là phải giải quyết vấn đề mất kết nối với CSDL trung tâm do sự chuyển giao kết nối của các trạm thu phát sóng di động (BTS). Sau đây là một số loại giao tác di động trên thế giới đã nghiên cứu [6-12]:

**Kiểu giao tác Xúc tiến (Pro-motion transaction model)**

**Đặc điểm:** Giao tác này hỗ trợ 10 mức tách khác nhau. Giao tác cho phép “commit” cục bộ tại các thiết bị di động đầu cuối (mobile hosts); kết quả “commit” của các giao tác có ảnh hưởng đến các giao tác cục bộ khác. Tuy nhiên, kết quả “commit” này chỉ hợp lệ khi mobile host (MH) kết nối trở lại với các máy chủ cơ sở dữ liệu.

**Độ di động (mobility):** Mặc dù bộ quản lý độ di động (mobility manager) hỗ trợ việc truyền thông giữa mobile host và các máy chủ cơ sở dữ liệu, nhưng loại giao tác Pro-motion không đề cập đến độ di chuyển như thế nào.

**Ngắt kết nối (disconnection):** Giao tác Pro-motion hỗ trợ xử lý giao tác khi bị ngắt kết nối thông qua các đối tượng gọi là “compact”. Khi mobile host bị ngắt kết nối từ cơ sở dữ liệu cố định (fixed database), các giao tác con (sub-transactions) được chia nhỏ ra và thực thi ở mobile host. Việc xử lý ngắt kết nối là một tính năng nổi trội của loại giao tác này. Do đó, loại giao tác Pro-motion yêu cầu các nguồn tài nguyên tại mobile host cao.

**Sự thực hiện phân tán (distributed execution):** Hầu hết các giao tác đều thực hiện ở mobile host và kết quả được thống nhất ở các máy chủ cơ sở dữ liệu. Vì vậy, xử lý giao tác phân tán không được hỗ trợ ở loại giao tác này.

**Kiểu giao tác hai lớp (Two-tier transaction model)**

**Đặc điểm:** Loại giao tác này dựa trên lược đồ sao chép dữ liệu. Mỗi đối tượng dữ liệu có một bản sao chủ và những bản sao riêng. Có hai loại giao tác là: *Base* và *Tentative*. Các giao tác *Base* hoạt động trên bản sao chủ; trong khi các giao tác *Tentative* truy cập các bản sao riêng. Một mobile host có thể lưu trữ bản sao chủ hoặc bản sao riêng của các đối tượng dữ liệu. Trong lúc mobile host bị ngắt kết nối, các giao tác *tentative* cập nhật bản sao. Khi mobile host kết nối trở lại với máy chủ cơ sở dữ liệu, các giao tác *tenta-*

*tive* được chuyển đổi trở lại các giao tác gốc, các giao tác đó được thực hiện lại trên bản sao chủ. Nếu một giao tác gốc không hoàn thành qui định chính xác (được chỉ rõ bởi ứng dụng), giao tác *tentative* liên kết bị hủy bỏ.

Giao tác hai lớp không hỗ trợ “mobility”.

Trong lúc mobile host ngắt kết nối từ database servers, các giao tác *tentative* được thực hiện nội bộ dựa vào bản sao của các đối tượng dữ liệu.

Các giao tác được thực hiện thử ở mobile hosts bị ngắt kết nối và thực hiện lại các giao tác gốc ở các máy chủ cơ sở dữ liệu.

**Kiểu giao tác Yếu – Nghiêm ngặt (Weak-Strict transaction model)**

Giao tác Weak-Strict (còn gọi là Clustering) bao gồm 2 loại giao tác: *weak* (hoặc *loose*) và *strict*. Các giao tác này được thực hiện bên trong *clusters*, nó tập hợp các mobile host được kết nối thông qua mạng tốc độ cao và tin cậy. Mỗi *cluster*, dữ liệu liên quan được sao lưu cục bộ. Có 2 loại bản sao: *local consistency (weak)* và *global consistency (strict)*. Bản sao *weak* được dùng khi mobiles hosts bị ngắt kết nối hoặc kết nối thông qua mạng tốc độ chậm hoặc không tin cậy. Các giao tác Weak và Strict truy xuất các bản sao dữ liệu *weak* và *strict* riêng biệt.

Khi mobile hosts kết nối trở lại với máy chủ cơ sở dữ liệu, tiến trình đồng bộ thống nhất những thay đổi giữa dữ liệu cục bộ với dữ liệu toàn cục.

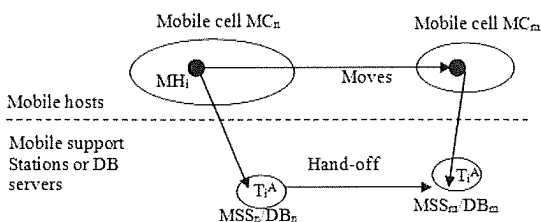
Các giao tác *Weak* được phép “commit” bên trong *cluster* của nó, và kết quả đó có ảnh hưởng đến các giao tác *weak* nội bộ khác. Khi mobile hosts kết nối trở lại, kết quả của các giao tác *weak* được thống nhất với kết quả của các giao tác *strict*. Nếu kết quả của một giao tác *weak* không xung đột với việc cập nhật của các giao tác *strict*, các giao tác *weak* sẽ được “commit” toàn cục, nếu không chúng bị hủy.

**Độ di động (mobility):** Khái niệm “migration” của giao tác được đưa ra để hỗ trợ cho độ di động, và giảm chi phí đường truyền. Khi mobile hosts di chuyển và kết nối với một trạm hỗ trợ di động (MSS – mobile support station) mới, một phần của giao tác thực hiện ở MSS cũ được chuyển đến trạm mới. Tuy nhiên, chi tiết hơn nữa về thiết kế và thực hiện thì không được

cung cấp.

Ngắt kết nối (disconnection): giao tác weak-Strick hỗ trợ xử lý giao tác trong lúc ngắt kết nối và kết nối với sóng yếu thông qua các giao tác weak.

Một số nghiên cứu hiện nay về độ di động của các giao tác khi di chuyển qua các tế bào di động: Vị trí của MH (mobile host) được định danh bởi ID (identity) của tế bào di động (MC-mobile cell). Trong tế bào di động mới, các MH có thể kết nối với trạm hỗ trợ di động (MSS – Mobile Support Station) mới. Trong [6], có định nghĩa một giao tác gọi là “anchor transaction” được cài ở mạng cố định: ở MSS hoặc các máy chủ cơ sở dữ liệu. Các MSS này là các điểm “anchor” của các “anchor transaction” Khi MH di chuyển vào trong các MC mới, một tiến trình “hand-over” (hoặc Hand-off) sẽ xảy ra, khi đó các “anchor transaction” sẽ được chuyển từ điểm “anchor” trước đó tới điểm “anchor” mới. Hình dưới đây thể hiện khi một MH<sub>i</sub> di chuyển từ một MC<sub>n</sub> tới một tế bào di động mới MC<sub>m</sub>, tiến trình hand-over sẽ di chuyển một T<sub>iA</sub> từ một trạm hỗ trợ di động MSS<sub>n</sub> tới MSS<sub>m</sub>. Giao tác “anchor” T<sub>iA</sub> sẽ giữ vết (track) của MSS mà nó di chuyển qua, ví dụ MSS<sub>n</sub> và MSS<sub>m</sub>, và do đó nó hỗ trợ “mobility” của các giao tác di động khi đi qua các tế bào di động khác.



Hình 1: Nguyên lý chuyển giao kết nối di động

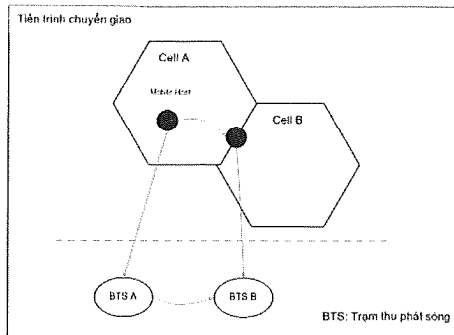
Tuy nhiên, các loại giao tác được đề cập ở trên không giải quyết vấn đề thời gian thực trong tiến trình chuyển giao (hand-over) giữa các tế bào (cell) trong mạng di động tế bào (cellular).

Khi các Mobile Host (MH) di chuyển từ tế bào này sang tế bào khác tiến trình hand-over thực hiện để chuyển MH sang kênh truyền mới.

Vấn đề chuyển giao (Hand-over)

Khi người sử dụng di chuyển từ tế bào này sang tế bào khác à mạng thực hiện việc điều khiển chuyển mạch.

Thời gian thực hiện chuyển giao là 100ms.



Hình 2: Sơ đồ mạng lưới các trạm BTS

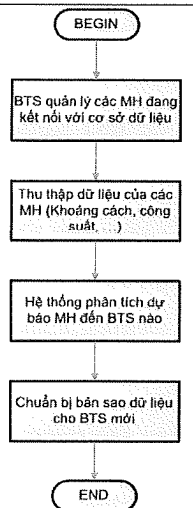
Hướng nghiên cứu để giải quyết vấn đề trong quản lý mạng internet di động:

Khi MH đang di chuyển trong mạng tổ ong, hệ thống sẽ tìm cách dự báo trước MH sẽ di chuyển đến tế bào nào trong số 6 tế bào xung quanh.

Làm cách nào để dự báo trước MH sẽ di chuyển đến tế bào nào trong số 6 tế bào xung quanh.

- Biết được khoảng cách từ MH đến các tế bào lân cận tại từng thời điểm.
- Biết được mức thu tín hiệu của MH tại các tế bào lân cận (BTS lân cận).

Chuẩn bị dữ liệu bản sao đến các BTS đó, không đợi đến tiến trình chuyển giao (Hand-over) hoặc bị mất kết nối (Disconnect) do MH đi vào các vùng tối (vùng không có sóng) giữa các tế bào.



Hình 3: Lưu đồ để dự đoán vị trí trong môi trường di động

### 1. BTS quản lý các MH đang kết nối với CSDL

Tất cả các BTS trong mạng do MSC quản lý sẽ kiểm soát các MH đang kết nối với CSDL xác định nào đó. Ví dụ: Cơ sở dữ liệu hàng không dùng cho khách hàng đặt mua vé qua mạng.

**2. Thu thập dữ liệu của các MH**

Các dữ liệu của các MH bao gồm: Khoảng cách từ MH đến các BTS; Công suất thu được của BTS đối với các MH.

**3. Hệ thống phân tích dự báo MH sẽ đến các BTS nào**

Sau khi thu thập được dữ liệu, hệ thống sẽ dự báo trước rằng MH sẽ được BTS nào tiếp nhận (ví dụ bằng phương pháp phân lớp Bayes).

**4. Chuẩn bị bản sao dữ liệu cho BTS mới**

Sau khi đã dự báo trước BTS nào sẽ tiếp nhận MH thì sẽ sao dữ liệu đến BTS đó. Như thế, sẽ không bị ngắt quãng do tiến trình hand-over tạo ra.

**III. Dự báo trước vị trí Mobil User**

Việc quản lý đường đi của mobile users trong môi trường tính toán di động (Mobile Computing Environments) bao gồm các phương pháp lưu trữ và cập nhật thông tin vị trí phục vụ bởi hệ thống. Một vấn đề quan trọng trong lĩnh vực nghiên cứu quản lý đường đi là dự báo trước đường đi của mobile users [8], [10].

Dự báo trước đường đi có thể được định nghĩa như là việc dự báo sự di chuyển kế tiếp của một mobile user giữa các tế bào (cell) trong mạng GSM. Dự báo trước đường đi của mobile users được dùng để gia tăng hiệu quả của mạng GSM. Bằng cách dự báo trước đường đi, hệ thống có thể cấp phát hiệu quả nguồn tài nguyên tới các cell cần thiết thay vì cấp phát dàn trải đến toàn bộ các cell trên toàn mạng. Ngoài ra, bằng cách dự báo trước đường đi của mobile users các nhà cung cấp dịch vụ có thể tính toán một cách tối ưu khi thiết kế lắp đặt và điều chỉnh cấu trúc mạng di động GSM, điều chỉnh băng thông cho phù hợp với thực tế.

Nhiều ứng dụng như điều hành lưu lượng giao thông, chăm sóc sức khỏe cộng đồng, sinh học, quản lý khách sạn và trong quân sự cũng áp dụng tiến trình xử lý về phụ thuộc vị trí (location – dependent). Với việc dự báo trước vị trí một cách hiệu quả, có thể trả lời các câu hỏi liên quan đến các vị trí trong tương lai của các users.

**3.1. Định nghĩa vấn đề**

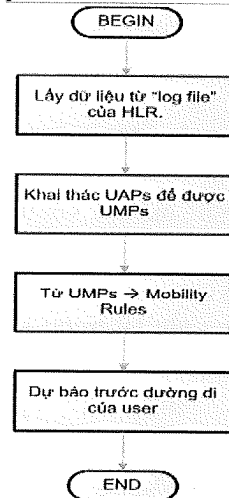
Hiện nay mạng GSM đã rất thông dụng ở Việt Nam và trên thế giới. Vùng phủ sóng của mạng GSM được chia ra những vùng nhỏ hơn gọi là “cell”. Trong mỗi “cell” của mạng GSM có một

trạm thu phát sóng gọi là BTS (Base Transceiver Station) có nhiệm vụ thu và phát sóng vô tuyến đến các Mobile Users [11][12]. Các BTS được kết nối với nhau thông qua mạng truyền dẫn quang hoặc vô tuyến. Các Mobile Users sử dụng các kênh truyền dẫn vô tuyến để truyền thông với các BTS.

Sự di chuyển của một mobile user từ một “cell” hiện tại tới một “cell” khác sẽ được ghi nhận trong một cơ sở dữ liệu gọi là VLR (Visitor Location Register). Sau đó, dữ liệu từ VLR sẽ được chuyển về HLR (Home Location Registry) [13] [14] đặt tại tổng đài MSC (Mobile Switching System). Từ dữ liệu này chúng ta có thể lấy ra lịch sử di chuyển của mobile user để có thể dự báo trước vị trí của họ. Ta gọi những mẫu dữ liệu lịch sử di chuyển này là UAPs (User Actual Paths). UAPs là một nguồn thông tin có giá trị bởi vì sự di chuyển của mobile user bao gồm các mẫu di chuyển ngẫu nhiên và các mẫu di chuyển thường xuyên. Trong bài báo [8] đã đưa ra thuật toán khai phá các mẫu di chuyển của mobile users nhằm dự báo vị trí di chuyển kế tiếp. Trong bài báo của chúng tôi đưa ra cũng dựa theo thuật toán này nhưng được cải tiến hiệu quả hơn về thời gian chạy. Cụ thể sẽ được trình bày ở phần sau.

**3.2. Phương pháp thực hiện dự báo vị trí thuê bao di động**

Theo lưu đồ ở hình 4, ta thấy rằng phương pháp thực hiện để dự báo trước vị trí của mobile users qua 4 bước sau:



**Hình 4:**  
Lưu đồ thực hiện tính toán các giao tác của thuê bao di động

Bước 1: Lấy dữ liệu từ logfile của HLR. Sự di chuyển của mobile user từ cell hiện tại tới cell khác sẽ được ghi nhận trong cơ sở dữ liệu gọi

là HLR. Chúng lưu trữ thường trực thông tin của khách hàng trong mạng di động. Lịch sử di chuyển của user mobil (máy điện thoại có thể được lấy ra từ log file của HLR.

Bước 2: Khai thác UAPs để được UMPs

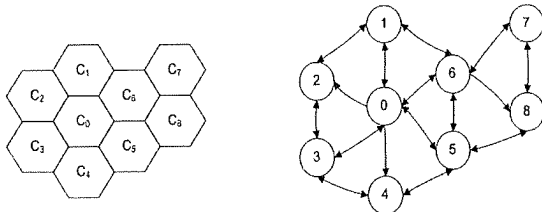
Giả sử, đặt tên đường đi thực tế của mobile user là UAPs (User Actual Paths). UAPs là một nguồn thông tin có giá trị vì nó chứa đựng các mẫu di chuyển thường xuyên và mẫu di chuyển ngẫu nhiên. Do đó, sử dụng UAPs, ta có thể rút ra mẫu di chuyển thường xuyên và sử dụng chúng để dự báo trước.

Giả sử rằng UAPs có dạng:  $C = \{c_1, c_2, \dots, c_n\}$ . Mỗi  $c_k$  biểu thị số ID của "cell" thứ  $k$  trong vùng phủ sóng.

Gọi các mẫu di chuyển của users là: UMPs (User Mobility Patterns). Khai thác UMPs cho ta rút ra được các luật di chuyển (Mobility Rules).

UMP là một chuỗi tuần tự các "cell" lân cận trong vùng phủ sóng. Để khai thác UMPs từ UAPs thực hiện như sau:

Gọi  $G$  là đồ thị có hướng tương ứng với các "cell" trong vùng phủ sóng. Mỗi "cell" là 1 đỉnh của  $G$  như hình 2. Nếu có 2 "cell" gọi là  $A$  và  $B$  lân cận nhau (có chung biên) trong vùng phủ sóng thì có đường đi trực tiếp và không có trọng số từ  $A$  đến  $B$  và ngược lại.



Hình 5: Mô hình mạng tế bào và đồ thị G trong mạng di động

Giả sử có 2 UAPs,  $A = \{a_1, a_2, \dots, a_n\}$  và  $B = \{b_1, b_2, \dots, b_m\}$ .  $B$  là chuỗi con của  $A$ , nếu tồn tại:  $1 \leq i_1 < \dots < i_m \leq n$ ,  $b_k = a_{i_k}$ , với mọi  $k$ , và  $1 \leq k \leq m$ .

Ngoài ra,  $B$  là một chuỗi con của  $A$ , nếu tất cả "cell" trong  $B$  cũng tồn tại trong  $A$  (chúng không cần liên tiếp trong  $A$ ).

Ví dụ: trong hình 5: giả sử  $A = \{c_3, c_4, c_0, c_1, c_6, c_5\}$ , và  $B = \{c_4, c_5\}$  là chuỗi con có chiều dài là 2 của  $A$ . Ngoài ra, UAP  $B$  được chứa bởi UAP  $A$ .

Sau đây là thuật toán UMP Mining trong [8], [10]:

**UMP Mining ()**

Vào: Tất cả các đường đi thực tế (UAPs) của users có trong  $D$ .

Giá trị của độ ủng hộ tối thiểu:  $Supp_{min}$

Đồ thị vùng phủ sóng:  $G$

Ra: Các mẫu di chuyển của users (UMP):  $L$

1.  $C_1 \leftarrow$  các mẫu có chiều dài là 1.
2.  $k = 1$
3.  $L = \emptyset$  // khởi tạo tập các mẫu phổ biến là rỗng
4. while  $C_k \neq \emptyset$  {
5.   foreach UAP  $a \in D$  {
6.      $S = \{s \mid s \in C_k \text{ và } s \text{ là chuỗi con của } a\}$
7.     foreach  $s \in S$  {
8.        $s.count = s.count + s.support$
9.     }
10.   }
11.   // chọn các ứng viên thỏa  $Supp_{min}$ :
12.    $L_k = \{s \mid s \in C_k, s.count \geq supp_{min}\}$
13.    $L = L \cup L_k$
14.   // Tạo các mẫu ứng viên chiều dài  $(k+1)$ .
15.  $C_{k+1} \leftarrow$  CandidateGeneration( $L_k, G$ ),  $\forall c \in$

**Candidate Generation ()**

Input: Các mẫu phổ biến chiều dài  $k$ ,  $L_k$

Đồ thị vùng phủ sóng,  $G$

Output: Các mẫu ứng viên chiều dài  $(k + 1)$ , Candidates

1. Candidates =  $\emptyset$  // Khởi tạo tập candidates bằng rỗng
2. foreach  $L = (l_1, l_2, \dots, l_k)$ ,  $L \in L_k$  { // tìm các cell lân cận của mỗi  $l_k$  trong  $G$
3.    $N^+ = \{v \mid \text{có một cạnh trong } G \text{ như là } l_k \rightarrow v\}$
4.   foreach  $v \in N^+(l_k)$  { // cho mỗi cell lân cận này,
5.     //  $v$  tạo một ứng viên
6.      $C' = (l_1, l_2, \dots, l_k, v)$
7.     // Thêm  $C'$  tới tập các ứng viên (candidates set)
8.     Candidates  $\leftarrow$  Candidates  $\cup C'$
9.   }
10. }
11. return Candidates

UAP ID	UAP
1	{5,6,0,4}
2	{4,5}
3	{3,4,5,0}
4	{1,2,3,4,0,5}
5	{3,2,0}

Bảng 1: Dữ liệu thử

Sau khi chạy thuật toán ta có được tập các mẫu phổ biến như sau (giả sử  $supp_{min} = 1.33$ ):

L			
pattern	support	pattern	support
{0}	4	{3,0}	1.33
{2}	2	{3,4}	2
{3}	3	{4,0}	1.5
{4}	3	{4,5}	2.5
{5}	3	{5,0}	1.5
{0,5}	1.5	{3,4,0}	1.5
{2,0}	1.33	{3,4,5}	1.5

Bảng 2: Các mẫu phổ biến thử nghiệm

*Bước 3: Tạo luật*

Từ UMPs, bây giờ có thể tạo ra tập các luật di chuyển như sau:

Giả sử, ta có một UMP  $C = \{c_1, c_2, \dots, c_k\}$ , với  $k > 1$ . Tất cả luật di chuyển có thể bắt nguồn từ một mẫu như sau:  $\{c_1\} \rightarrow \{c_2, \dots, c_k\}$

$$\{c_1, c_2\} \rightarrow \{c_3, \dots, c_k\}$$

...

$$\{c_1, c_2, \dots, c_{k-1}\} \rightarrow \{c_k\}$$

Cho một luật di chuyển  $R: (c_1, c_2, \dots, c_{i-1}) \rightarrow (c_i, c_{i+1}, \dots, c_k)$ , hệ số confidence được tính như sau:

$$\text{Confidence}(R) = \frac{(c_1, c_2, \dots, c_k) \text{ count}}{(c_1, c_2, \dots, c_{i-1}) \text{ count}} \times 100$$

Bằng cách sử dụng UMPs, tất cả các luật di chuyển được tạo ra và các giá trị confidence cũng phải được tính toán. Các luật có độ confidence  $\geq \text{confmin}$  sẽ được chọn.

*Bước 4: Dự báo trước vị trí của Mobile User*

Đây là các bước dự báo trước mobile users sẽ di chuyển đến “cell” nào trong vùng phủ sóng.

Các luật di chuyển (Mobility rules)	
Rule	Confidence
$\{2\} \rightarrow \{0\}$	66.6
$\{4\} \rightarrow \{0\}$	50
$\{3\} \rightarrow \{4\}$	66.6
$\{5\} \rightarrow \{0\}$	50
$\{4\} \rightarrow \{5\}$	83.33
$\{3,4\} \rightarrow \{0\}$	75
$\{3\} \rightarrow \{4,0\}$	50
$\{3,4\} \rightarrow \{5\}$	75
$\{3\} \rightarrow \{4,5\}$	50

**Bảng 3:** Tất cả các luật di chuyển

Trong bảng 3, tất cả các luật sẽ được dùng để dự báo trước cho “user”. Giả sử “user” đi theo đường  $P = \{2, 3, 0, 4\}$  cho đến bây giờ và “user” hiện đang ở cell 4. Theo luật trên thì có:  $\{4\} \rightarrow \{0\}$ ,  $\{4\} \rightarrow \{5\}$ ,  $\{3,4\} \rightarrow \{0\}$ , và  $\{3,4\} \rightarrow \{5\}$ . Vậy “user” có thể di chuyển đến “cell” 0 hoặc “cell” 5.

**5. Cải tiến thuật toán**

*Định nghĩa về ngữ cảnh khai phá dữ liệu:* Cho tập  $O$  là tập hữu hạn khác rỗng các giao tác (UAP ID) và  $I$  là tập hữu hạn khác rỗng các cell,  $R$  là quan hệ 2 ngôi giữa  $O$  và  $I$  sao cho với  $o \in O$  và  $i \in I$ ,  $(o, i) \in R \Leftrightarrow$  giao tác  $o$  có chứa cell thứ  $i$ . Ngữ cảnh Khai phá dữ liệu (KPD L) là bộ ba  $(O, I, R)$ .

*Định nghĩa về tập phổ biến:* Cho ngữ cảnh KPD L  $(O, I, R)$  và  $S \subset I$ , độ phổ biến của  $S$  được định nghĩa là tỉ số giữa số các giao tác có chứa  $S$  và số lượng cell trong  $O$ . Độ phổ biến của  $S$  ký hiệu là  $SP(S)$  và được tính như sau:  $SP(S) = |p(S)| / |O|$

Với  $| \cdot |$  là lực lượng của tập hợp. Cho  $S \subset I$  và  $\text{minsupp} \in (0, 1]$  là ngưỡng phổ biến tối thiểu,  $S$  là một tập phổ biến theo ngưỡng  $\text{minsupp}$  nếu và chỉ nếu  $SP(S) \geq \text{minsupp}$ .

Ký hiệu  $FS(O, I, R, \text{minsupp})$  là tập hợp các tập phổ biến theo ngưỡng  $\text{minsupp}$  hay  $FS(O, I, R, \text{minsupp}) = \{ S \in P(I) \mid SP(S) \geq \text{minsupp} \}$

Mệnh đề:

Cho  $T \notin FS(O, I, R, \text{minsupp})$ , nếu  $T \subseteq S$  thì  $S \notin FS(O, I, R, \text{minsupp})$ .

Với  $FS(O, I, R, \text{minsupp})$  là tập hợp các tập phổ biến theo ngưỡng  $\text{minsupp}$ .

Tức là: tập ứng viên  $T$  không phổ biến, và  $T \subseteq S$  thì  $S$  cũng không phổ biến.

Áp dụng vào thuật toán tạo các ứng viên  $\text{CandidateGeneration}()$ :

Tại dòng 3:  $N^+ = \{ v \mid \text{có một cạnh trong } G \text{ như là } lk \rightarrow v \}$

Thêm vào:  $v \in L$ , tức là: cạnh  $v$  trong đồ thị  $G$  phải thuộc về tập phổ biến  $L$ .

$N^+ = \{ v \mid \text{có một cạnh trong } G \text{ như là } lk \rightarrow v \text{ và } v \in L \}$

Kết quả tính toán số liệu thực: Chúng tôi trích một phần nhỏ dữ liệu từ logfile của HLR với số trạm BTS là 351 trạm, số UAPs là 1.179.034 record. Sau khi chuẩn hóa dữ liệu số đường đi còn lại là 31.415 record.

Cn	Thuật toán UMPMining		Thuật toán cải tiến 1	
	số lượng Cn	Thời gian chạy	số lượng Cn	Thời gian chạy
C1	351	33	351	32
C2	1492	176	1488	167
C3	5191	511	3340	341
C4	651	63	79	8
Tổng	7685	783	5258	548

Trong đó:

C1: các mẫu ứng viên có chiều dài là 1.

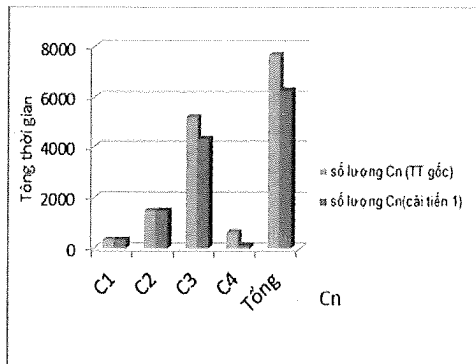
L1: các mẫu phổ biến có chiều dài là 1.

Cn: các mẫu ứng viên có chiều dài n.

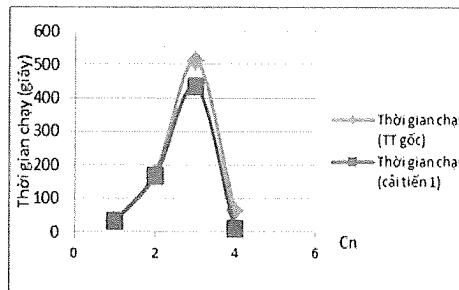
Ln: các mẫu phổ biến có chiều dài n.

**Bảng 5: so sánh kết quả thực nghiệm**

Biểu đồ so sánh Cn và thời gian:



Hình 6: Biểu đồ so sánh Cn



Hình 7: Biểu đồ thời gian chạy của 2 thuật toán

**Nhận xét kết quả thu được:**

Tổng số lượng Cn khi chạy thuật toán cải tiến 1 giảm đi so với thuật toán [8] là:

$$7685 - 5258 = 2427.$$

Cụ thể số lượng Cn giảm như sau:

$$C2 = 1492 - 1488 = 4$$

$$C3 = 5191 - 3340 = 1851$$

$$C4 = 651 - 79 = 572.$$

Đối với C1 không giảm là do thuật toán chạy lần đầu tiên (quét tất cả các cell).

Tổng thời gian chạy thuật toán cải tiến 1 giảm đi so với thuật toán [8] là:

$$783 \text{ giây} - 548 \text{ giây} = 235 \text{ giây}.$$

Cụ thể thời gian giảm như sau:

$$\text{Thời gian chạy C2 giảm: } 176 \text{ giây} - 167 \text{ giây} = 9 \text{ giây}$$

$$\text{Thời gian chạy C3 giảm: } 511 \text{ giây} - 341 \text{ giây} = 170 \text{ giây}$$

$$\text{Thời gian chạy C4 giảm: } 63 \text{ giây} - 8 \text{ giây} = 55 \text{ giây}.$$

**Kết luận**

Ngày nay, xu hướng làm việc trong môi trường di động ngày càng phổ biến, việc nâng cao chất lượng dịch vụ trên nền di động là một yêu cầu không thể bỏ qua. Bài toán đặt ra là khi khách hàng truy cập dữ liệu, khai thác dữ liệu trong môi trường di động cũng giống như trong môi trường cố định. Muốn thế, chúng ta phải giải quyết vấn đề tranh chấp của các giao tác di động, thời gian trì hoãn dài, mất kết nối (Disconnect) với cơ sở dữ liệu trung tâm do sự chuyển giao kết nối (hand-over) của các trạm thu phát sóng di động. Trong phần hai, mục đích của chúng tôi cải tiến thuật toán để giảm thời gian chạy của thuật toán [8]. Kết quả thực nghiệm, chúng ta thấy rằng: với số cell là 351, số UAPs là 31415 thì thời gian chạy của thuật toán là 783 giây và thời gian chạy sau khi cải tiến là 548 giây (giảm đi 30,01%).

**TÀI LIỆU THAM KHẢO**

- [1] Lê Mạnh, Phan Ngọc Khôi (1984), “Một số vấn đề của hệ điều hành mạng máy tính phục vụ nhiều người sử dụng”, *Tạp chí Khoa học kỹ thuật*, Bộ Khoa học và Công nghệ 3/1984, tr.18-34.
- [2] Lê Mạnh, “Hệ GT90 một hệ quản trị dữ liệu trong thời gian thực”, *Tạp chí Khoa học tính toán và điều khiển*, Viện CNTT số 4/1991, tr16-22.
- [3]. Giang Minh Đức (2007), *Nghiên cứu hệ quản trị CSDL được khai thác trong thời gian thực, ứng dụng trong tra cứu danh bạ điện thoại qua tin nhắn ĐTDĐ (SMS)*, Luận án Thạc sỹ Khóa 1 ĐH CNTT – ĐHQG TP.HCM.
- [4] Lê Mạnh và cộng sự (2009), *Nghiên cứu giải pháp tự động trả lời các sinh viên những thông tin nhà trường thông qua tin nhắn (SMS) máy điện thoại di động*, Đề tài cấp ĐHQG TP.HCM.
- [5] Lê Mạnh (1998), “Hệ quản trị cơ sở dữ liệu trong ngành Bưu điện”, *Báo cáo khoa học tổng kết 20 năm thành lập Viện công nghệ thông tin*, tr.455-460.
- [6] G. Gok, O. Ulusoy, *Transmission of continuous query results in mobile computing systems*, Inform. Sci. 125 (1-4) (2000) 37-63.
- [7] S. Mohan, R. Jain, *Two user location strategies for personal communication systems*, IEEE Personal Commun. Mag. 1 (1994) 42-50.
- [8] A. Nanopoulos, D. Katsaros, Y. Manolopoulos, *Effective prediction of web user accesses: a data mining approach*, in: Proceedings of the WebKDD Workshop (WebKDD'01), 2001.
- [9] A. Nanopoulos, D. Katsaros, Y. Manolopoulos, *A data mining algorithm for generalized web prefetching*, IEEE Trans. Knowl. Data Eng. 15 (5) (2003) 1155-1169.



- [10] Accture. The Future of Wireless: Different than You Think, Bolder than You Imagine. Working Paper, June 2001.
- [11] S. K. Madria and B. K. Bhargava: A Transaction Model to Improve Data Availability in Mobile Computing, Distributed and Parallel Databases, 2001, pp127-160.
- [12] R. A. Dirckze and L. Gruenwald: A pre-serialization transaction management technique for mobile multidatabases, Mobile Networks and Applications (MONET), 2000, pp311-321.
- [13] Hien Nam Le: A Transaction Processing System for Supporting Mobile Collaborative Works, Department of Computer and Information Science, Norwegian University of Science and Technology Trondheim, Norway, 2006.
- [14] K.-I. Ku and Y.-S Kim: Moflex Transaction Model for Mobile Heterogeneous Multidatabase Systems, Research Issues in Data Engineering (RIDE), 2000, pp39-46.
- [15] Giang Minh Đức, Báo cáo chuyên đề III NCS “Kết quả thực nghiệm của Thuật toán dự báo trước vị trí của Mobile Users”, 5/2014.
- [16] Manh Le, Duc Giang Minh. Transactions in mobile communication. Sixth International Conference on Information Technology for Education and Research in HCM City. pp120-126 (2010).
- [17] Lê Mạnh, Nguyễn Việt Hùng (2012), “Theo dõi hành vi mã độc trong môi trường mạng ảo cách ly”, Hội thảo Quốc gia lần thứ XV Một số vấn đề chọn lọc của Công nghệ thông tin và Truyền thông, tr443-446.
- [18] Giang Minh Đức (2007), “Sử dụng nhắn tin di động cho hệ thống báo hồng SMS 119”, *Tạp chí Bưu chính Viễn thông và Công nghệ thông tin* kỳ 1, tr70-72.
- [19] Giang Minh Đức (2008), “Nghiên cứu các giao tác trong thông tin di động”, *Tạp chí Công nghệ thông tin và Truyền thông* kỳ 1, tr52-55.
- [20] Lê Mạnh (2014), “Áp dụng nguyên lý máy học để phân tích và phát hiện mã độc trong phần mềm máy tính”, *Tạp chí Khoa học Đại học Văn Hiến* số 4, tr84-92.
- [21] Giang Minh Duc, Le Manh, Do Hong Tuấn, A Novel Location Prediction Algorithm of Mobile User For Cellular Networks. Journal on Information Communications Technology. Reseach and Developmrt on Information & Communications Technology. “Các công trình nghiên cứu và phát triển công nghệ thông tin và truyền thông. ISSN 1859-3534 - Volume E-3, No 8(12) pp.58-66, tháng 8/2015.