

**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

---



**LÊ NGỌC HÙNG**

**TỐI ƯU HÓA QUẢN LÝ DI ĐỘNG TRONG  
MẠNG VÔ TUYẾN HỖN HỢP ĐA DỊCH VỤ**

Chuyên ngành: **Hệ thống thông tin**

Mã số: **9.48.01.04**

**LUẬN ÁN TIẾN SĨ KỸ THUẬT**

**HÀ NỘI - 2020**

BỘ THÔNG TIN VÀ TRUYỀN THÔNG  
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

---

## LUẬN ÁN TIẾN SĨ

*ĐỀ TÀI*

**TỐI ƯU HÓA QUẢN LÝ DI ĐỘNG TRONG MẠNG VÔ TUYẾN HỖN  
HỢP ĐA DỊCH VỤ**

Chuyên ngành: **Hệ thống thông tin**

Mã số: **9.48.01.04**

Nghiên cứu sinh : Lê Ngọc Hưng

Người hướng dẫn khoa học: GS. TSKH. Nguyễn Xuân Quỳnh

Hà Nội - năm 2020

## **LỜI CAM ĐOAN**

Tôi xin cam đoan đây là công trình nghiên cứu của tôi, các số liệu và kết quả trình bày trong luận án là trung thực và chưa được công bố ở bất kỳ công trình nào khác.

Tác giả

**Lê Ngọc Hưng**

# LỜI CẢM ƠN

Luận án Tiến sỹ này được thực hiện tại Học viện Công nghệ Bưu chính Viễn thông dưới sự hướng dẫn của GS.TSKH. Nguyễn Xuân Quỳnh. Nghiên cứu sinh bày tỏ lòng biết ơn sâu sắc tới GS.TSKH. Nguyễn Xuân Quỳnh, thầy trực tiếp hướng dẫn, giúp đỡ tạo mọi điều kiện thuận lợi để nghiên cứu sinh hoàn thiện công trình nghiên cứu của mình. Thầy đã có rất nhiều ý kiến gợi mở về hướng nghiên cứu để tôi thực hiện thành công đề tài.

Tôi cũng xin cảm ơn Ban lãnh đạo Học viện Công nghệ Bưu chính Viễn thông, Khoa Đào tạo sau đại học, Công ty TNHH VKX (nơi tôi đang công tác), cũng như các đồng nghiệp đã tạo điều kiện và giúp đỡ tôi hoàn thành được đề tài nghiên cứu của mình.

Cuối cùng là sự biết ơn tới gia đình, bạn bè đã thông cảm, động viên giúp đỡ cho tôi có đủ nghị lực để hoàn thành luận án.

*Hà Nội, tháng 12 năm 2020*

Lê Ngọc Hưng

## MỤC LỤC

<b>MỤC LỤC</b> .....	iii
<b>DANH MỤC BẢNG MẪU</b> .....	xi
<b>DANH MỤC HÌNH VẼ</b> .....	xii
<b>BẢNG KÝ HIỆU TOÁN HỌC</b> .....	xiv
<b>MỞ ĐẦU</b> .....	1
1. Lý do nghiên cứu .....	1
2. Mục đích nghiên cứu .....	2
3. Đối tượng và phạm vi nghiên cứu .....	2
4. Phương pháp và công cụ nghiên cứu .....	3
5. Ý nghĩa khoa học và thực tiễn của đề tài .....	3
6. Cấu trúc của Luận án .....	3
<b>Chương 1: Tổng quan các vấn đề cần nghiên cứu</b> .....	5
1.1 Mạng thông tin vô tuyến băng rộng đa dịch vụ BcN .....	5
1.2 Yêu cầu chuyển giao trong mạng BcN .....	7
1.3 Phân tích và đánh giá các nghiên cứu liên quan đến Luận án .....	9
1.4 Các vấn đề còn tồn tại .....	12
1.4.1 Phân loại chuyển giao .....	12
1.4.2 Các yêu cầu về hiệu suất .....	14
1.4.3 Phân tích trễ chuyển giao .....	15
1.4.4 Tối ưu hoá trễ HO .....	16
1.5 Kết luận chương 1 .....	17
<b>Chương 2: Xây dựng tập tham số phân tích và đánh giá hiệu suất của các kỹ thuật chuyển giao dựa trên MIP, TCP-M, và SIP</b> .....	19
2.1 Mở đầu .....	19
2.2 Phân tích hiệu suất chuyển giao dựa trên các giao thức quản lý di động hiện tại .....	19
2.2.1 Các giao thức quản lý di động ở lớp liên kết (Lớp 2) .....	21
2.2.2 Các giao thức quản lý di động ở lớp mạng (Lớp 3) .....	21
2.2.3 Các giao thức quản lý di động ở lớp giao vận (Lớp 4) .....	22

2.2.4	Các giao thức quản lý di động ở lớp ứng dụng (lớp 5).....	23
2.3	Xác định các tham số cơ bản của mô hình phân tích .....	24
2.3.1	Xác suất gói tin bị thất lạc từ điểm đến điểm .....	25
2.3.2	Độ trễ truyền bản tin điểm tới điểm.....	27
2.3.3	Trung bình độ trễ truyền gói tin báo hiệu khi sử dụng giao thức UDP .....	27
2.4	Xây dựng phương thức đánh giá hiệu suất chuyển giao của các ứng dụng dạng B và dạng C (MIP và TCP-M).....	30
2.4.1	Đánh giá hiệu suất chuyển giao của kết nối TCP khi sử dụng MIP .....	32
2.4.2	Đánh giá hiệu suất chuyển giao của kết nối TCP khi sử dụng TCP-M.....	36
2.5	Xây dựng phương thức đánh giá hiệu suất chuyển giao của các ứng dụng dạng D và dạng E (MIP và SIP) .....	38
2.5.1	Đánh giá hiệu suất chuyển giao của một kết nối UDP khi sử dụng MIP .....	39
2.5.2	Đánh giá hiệu suất chuyển giao của một kết nối UDP khi sử dụng SIP .....	41
2.6	Phân tích đánh giá tương quan giữa tiêu hao nguồn điện và hiệu suất chuyển giao .....	43
2.6.1	Mô hình phân tích .....	43
2.6.2	Mô tả giao thức .....	43
2.6.3	Thuật toán tìm tuyến.....	44
2.6.4	Cấu trúc bản tin RREQ .....	46
2.6.5	Mô phỏng và đánh giá .....	47
2.7	Kết luận chương 2 .....	50
Chương 3:	Xây dựng phương thức định trước băng thông chuyển giao trong mạng BcN.....	53
3.1	Mở đầu .....	53
3.2	Xây dựng mô hình hệ thống.....	53
3.2.1	Kiến trúc trao đổi tin.....	54
3.2.2	Đánh số tế bào.....	55
3.2.3	Mô hình hoá di chuyển .....	55
3.2.4	Phân lớp các yếu tố ảnh hưởng đến chuyển giao.....	56
3.2.5	Tính toán xác suất chuyển giao.....	62
3.3	Thuật toán định trước băng thông và điều khiển đăng nhập.....	64

3.3.1 Định trước băng thông không thông tin di chuyển .....	64
3.3.2 Định trước băng thông với thông tin di chuyển .....	66
3.3.3 Điều khiển cửa sổ ước tính thời gian di chuyển.....	67
3.3.4 Điều khiển đăng nhập.....	69
3.3.5 Kết quả mô phỏng .....	71
3.4 Thuật toán dự báo chuyển giao đi và đến.....	72
3.4.1 Ba trạng thái xác suất.....	72
3.4.2 Điều khiển đăng nhập .....	74
3.5 Thuật toán dự báo băng thông theo kết nối.....	74
3.5.1 Điều khiển AG .....	74
3.5.2 Định trước băng thông theo kết nối sau khi đăng nhập.....	75
3.5.3 Định trước băng thông theo kết nối trước khi đăng nhập.....	76
3.6 Kết luận chương 3 .....	77
<b>Chương 4: Xây dựng phương thức quản lý chuyển giao linh hoạt trong mạng BcN .....</b>	<b>78</b>
4.1 Mở đầu .....	78
4.2 Phân tích hiệu suất của các giao thức định tuyến cho MANETs.....	78
4.2.1 Xu hướng nghiên cứu trong thời gian gần đây.....	79
4.2.2 Các giao thức định tuyến truyền thống.....	80
4.2.3 Mô phỏng và phân tích kết quả .....	82
4.3. Giải pháp Quản lý chuyển giao linh hoạt (AMMS) .....	87
4.3.1 Cấu trúc của AMMS.....	87
4.4 Phân tích mô hình đánh giá hiệu suất của AMMS .....	92
4.4.1 Mật số liệu của MIP và SIP .....	96
4.4.2 Thời gian giảm thông lượng của TCP-M .....	96
4.5 Kết luận chương 4 .....	96
<b>KẾT LUẬN .....</b>	<b>98</b>
<b>HƯỚNG NGHIÊN CỨU TIẾP THEO .....</b>	<b>100</b>
<b>DANH MỤC CÁC CÔNG TRÌNH CÔNG BỐ.....</b>	<b>101</b>
<b>TÀI LIỆU THAM KHẢO .....</b>	<b>102</b>

## CÁC TỪ VIẾT TẮT

<b>Viết tắt</b>	<b>Nghĩa tiếng Anh</b>	<b>Nghĩa tiếng Việt</b>
3/4/5G	Mobile 3/4/5 <sup>th</sup> Generation	Mạng di động thế hệ 3/4/5
A_WCETT	Advance Weighted Cumulative Expected Transmission Time	Thời gian truyền với trọng số mở rộng
AAA	Authentication, Authorization, and Accounting	Xác thực, Phân quyền, và Giám sát
AIMD	Additive Increase Multiplicative-Decrease (AIMD)	Thuật toán điều khiển nghẽn trong TCP
AMC	Architecture for ubiquitous Mobile Communications	Kiến trúc thông tin di động phổ biến
AMMS	Adaptive Multi-layer Mobility management Solution	Giải pháp quản lý di động đa lớp thích ứng
AODV	Ad hoc On-Demand Distance Vector	Véc tơ khoảng cách theo yêu cầu
APME	Access Point Management Entity	Thực thể quản lý điểm truy nhập
AR	Access Router	Router lớp truy nhập
AU	Authentication Unit	Đơn vị xác thực
AC	Authentication Center	Trung tâm xác thực
BLG	Boundary Location Register	Đăng ký vị trí biên
BS	Base Station	Trạm cơ sở
CARD	Candidate Access Router Discovery	Khám phá bộ định tuyến lớp truy nhập
CDR	Call Detail Records	Bản tin cước
CDMA	Code Division Multiple Access	Tiêu chuẩn đa truy nhập theo mã
CH	Correspondent Host	Thuê bao di động trong mạng



		IP
CMP	Cross-layer (Layer 2 + 3) handoff Management Protocol	Giao thức quản lý chuyển giao lớp 2-3
CN	Correspondent Node	Thuê bao cố định trong mạng IP
CA	Care-of-Addresses	Địa chỉ IP tạm thời cấp phát cho thuê bao di động
DECT	Digital Enhanced Cordless Telephone	Điện thoại cố định không
DHCP	Dynamic Host Configuration Protocol	Giao thức cấu hình máy chủ
DSR	Dynamic Source Routing	Định tuyến nguồn động
EAP	Extensible Authentication Protocol	Giao thức xác thực mở rộng
EAPOL	EAP over LAN	
EEMA	Efficient Energy and High-Performance Routing Protocol	Giao thức định tuyến cân bằng năng lượng và hiệu suất
ETX	Expected Transmission Count	Số lần truyền dự kiến
ETT	Expected Transmission Time	Thời gian truyền dự kiến
FA	Foreign Agent	Hệ thống xử lý yêu cầu truy nhập trên mạng khách
FN	Foreign Network	Mạng xử lý yêu cầu truy nhập trên mạng khách
FEC	Forward Error Correction	Sửa lỗi
FER	Frame Error Rate	Tỷ lệ lỗi khung
GEO	Geostationary Earth Orbit	Quỹ đạo địa tĩnh
GFA	Gateway Foreign Agent	Công trạm khách
GGSN	Gateway GPRS Support Node	Điểm hỗ trợ công GPRS
GPRS	General Packet Radio Services	Dịch vụ chuyển mạch gói vô tuyến
GPS	Global Positioning System	Hệ thống định vị toàn cầu

GRX	GPRS Roaming eXchange	Chuyển giao GPRS
GSM	Global System for Mobile communications	Công nghệ thông tin di động thế hệ 2 của châu Âu
HA	Home Agent	Hệ thống xử lý yêu cầu truy nhập trên mạng đăng ký thuê bao
HM	Hysteresis Margin	Biên độ trễ
HMIP	Hierarchical Mobile IP	Giao thức quản lý di động cục bộ theo mạng phân cấp
HN	Home Network	Mạng xử lý yêu cầu truy nhập trên mạng đăng ký thuê bao
HHO	Horizontal Handover	Chuyển giao trong mạng cùng công nghệ
HO	Handover	Chuyển giao
HOF	Hand Over Failure	Lỗi chuyển giao
IAPP	Inter-Access Point Protocol	Giao thức điểm liên truy nhập
IG	Interworking Gateway	Cổng liên mạng
IP	Internet Protocol	Giao thức lớp mạng
ISHO	Inter-System Handover	Chuyển giao liên mạng
LEO	Low Earth Orbit	Quỹ đạo trái đất thấp
LTE	Long-Term Evolution	Thông tin di động tốc độ cao
MAC	Message Authentication Code	Mã xác thực bản tin
MAHO	Mobile Assisted network controlled Handoff	Mạng hỗ trợ di động điều khiển chuyển giao
MANET	Mobile Ad Hoc Network	Mạng di động tùy biến
MAP	Mobile Application Part	Ứng dụng di động
MBCR	Minimum Battery Cost Routing	Chi phí định tuyến nguồn tối ưu
MH	Mobile Host	Trạm chủ di động

MIP	Mobile IP	IP di động
MN	Mobile Node	Node di động
MSE	Mobility State Estimation	Ước lượng trạng thái di động
MT	Mobile Terminal	Đầu cuối di động
NAHO	Network Assisted mobile controlled Handoff	Mạng hỗ trợ điều khiển chuyển giao
NAI	Network Access Identifier	Bộ định dạng mạng truy nhập
NBS	New Base Station	Trạm cơ sở mới
NGWS	Next-Generation Wireless Systems	Hệ thống vô tuyến thế hệ sau
NIA	Network Inter-operating Agent	Trạm liên mạng
NN	New Network	Mạng mới
OBS	Old Base Station	Trạm cơ sở cũ
PCF	Packet Control Function	Chức năng điều khiển gói
PDSN	Packet Data Serving Node	Điểm hỗ trợ dữ liệu gói
PLMN	Public Land Mobile Network	Mạng di động mặt đất
PSD	Power Spectral Density	Mật độ phổ năng lượng
QoS	Quality of Service	Chất lượng dịch vụ
RAN	Radio Access Network	Mạng truy nhập vô tuyến
RFC	Request For Comment	Đề nghị nhận xét
RLP	Radio Link Protocol	Giao thức liên kết vô tuyến
RREP	Route Reply	Trả lời tuyến
RREQ	Route Request	Yêu cầu tuyến
RRER	Route Error	Lỗi tuyến
RSS	Received Signal Strength	Cường độ tín hiệu nhận
RT	Retransmission Timeout	Quá thời gian truyền lại
RTP	Real-time Transport Protocol	Giao thức giao vận thời gian thực

RTT	Round-Trip Time	Thời gian nhận dữ liệu phản hồi
SCTP	Stream Control Transmission Protocol	Giao thức điều khiển luồng
SIM	Subscriber Identity Module	Khởi nhận dạng thuê bao
SIP	Session Initiation Protocol	Giao thức thoại trên Internet
SLA	Service Level Agreement	Thỏa thuận lớp dịch vụ
TDMA	Time Division Multiple Access	Tiêu chuẩn đa truy nhập kênh vô tuyến theo thời gian
TTT	Time To Trigger	Thời điểm kích hoạt HO
UMP	User Mobility Profile	Hồ sơ người dung
UMTS	Universal Mobile Telecommunication System	Hệ thống thông tin di động thế hệ 3 của châu Âu
VEPSD	Velocity estimation using the PSD	Ước lượng vận tốc bằng PSD
VGSN	Virtual GPRS Support Node	Khởi hỗ trợ GPRS ảo
VHO	Vertical Handover	Chuyển giao giữa các mạng công nghệ khác nhau
WLAN	Wireless Local Area Network	Mạng không dây cục bộ

## DANH MỤC BẢNG MẪU

BẢNG	NỘI DUNG	TRANG
Bảng 2.1	Hiệu suất quản lý di động của các giao thức	24
Bảng 2.2	Phương pháp tính cho phí tuyến	45
Bảng 2.3	Các tham số mô phỏng	47
Bảng 3.1	Tổng hợp các cơ chế điều kiện đăng nhập được trình bày ở trên	70
Bảng 4.1	Các giao thức định tuyến trên IEEE giai đoạn 2011-2020	78
Bảng 4.2	Các tham số mô phỏng	83

## DANH MỤC HÌNH VẼ

HÌNH	NỘI DUNG	TRANG
Hình 1.1	Cấu trúc ví dụ của một hệ thống vô tuyến tích hợp	6
Hình 1.2	Sự di chuyển trong các hệ thống vô tuyến thế hệ tiếp theo	7
Hình 1.3	Các mô hình quản lý di động trong MobileIP, MIP-RR và SIP	9
Hình 1.4	Các mô hình quản lý di động trong 3GPP, 3GPP2 và Micro mobility	10
Hình 2.1	Các kết nối Vô tuyến và Hữu tuyến giữa CH/HA, BS và MH	25
Hình 2.2	Cấu trúc gói số liệu	28
Hình 2.3	Cơ chế chuyển giao của kết nối TCP sử dụng giao thức MIP	32
Hình 2.4	Sơ đồ hoạt động của TCP-M	36
Hình 2.5	Chuyển giao kết nối TCP sử dụng MIP và SIP	39
Hình 2.6	So sánh trễ chuyển giao giữa MIP và SIP khi không có RLP (a) và có RLP (b)	42
Hình 2.7	Mô hình mạng MANET	43
Hình 2.8	Thủ tục kiểm tra tiêu hao năng lượng	43
Hình 2.9	Tập tuyến khả dụng giữa các cặp nút sau pha khám phá	44
Hình 2.10	Cấu trúc của bản tin RREQ	46
Hình 2.11	Thời gian duy trì mạng	48
Hình 2.12	Tỷ lệ phân phối bản tin	48
Hình 2.13	Trung bình trễ kết cuối	49
Hình 2.14	Trung bình thông lượng	49
Hình 3.1	Giao tiếp giữa MSC và BS	53
Hình 3.2	Đánh số tế bào	54
Hình 3.3	Mô hình Handoff sử dụng Lý thuyết	56
Hình 3.4	Vùng dịch vụ của 2 cell mỗi cell 3 véc tơ	56
Hình 3.5	Hàm thuộc của BTS thứ $k$ xác định theo khoảng cách vật lý	58
Hình 3.6	Số lần chuyển vùng trung bình ứng với các điều kiện khác nhau	60
Hình 3.7	Số lần chuyển vùng trung bình khi sử dụng SVAV với $T=10$	60

Hình 3.8	Thuật toán RSSI và SVAV đối với (a) Bất định đơn giản, và (b) Bất định có trọng số	60
Hình 3.9	Các cửa sổ thời gian thu được các hàm ước tính chuyển giao với $N_{win\_days}=2$	61
Hình 3.10	Ví dụ về đồ thị của hàm ước tính chuyển giao với $prev=1$	62
Hình 3.11	ví dụ về tính toán $p_h(C_{0,j} \rightarrow next)$ khi $prev(C_{0,j})=1$ và $next=4$ sử dụng đồ thị của $F_{HOE}(t_0, prev(C_{0,j}), next', T_{soj})$	64
Hình 3.12	Mã giả của thuật toán điều chỉnh $T_{est}$ trong mỗi BS	67
Hình 3.13a	Xác suất rớt cuộc gọi HO theo lưu lượng trên mỗi kênh.	70
Hình 3.13b	Xác suất khóa cuộc gọi mới theo lưu lượng trên mỗi kênh.	70
Hình 3.14a	Xác suất khóa cuộc gọi mới theo lưu lượng trên mỗi kênh với C khác nhau	70
Hình 3.14b	Xác suất rớt cuộc gọi HO theo lưu lượng trên mỗi kênh với C khác nhau	70
Hình 4.1	Giai đoạn khám phá tuyến	79
Hình 4.2	Quá trình xác định nút MPR của giao thức OLSSR	80
Hình 4.3	Tỷ lệ phân phối gói – Di động	83
Hình 4.4	Bảng thông trung bình - Di động	83
Hình 4.5	Trễ trung bình - Di động	83
Hình 4.6	Tải định tuyến chuẩn hóa - Di động	83
Hình 4.7	Tỷ lệ phân phối gói – Lưu lượng	84
Hình 4.8	Bảng thông trung bình - Lưu lượng	84
Hình 4.9	Trễ trung bình - Lưu lượng	84
Hình 4.10	Tải định tuyến chuẩn hóa - Lưu lượng	84
Hình 4.11	Cấu trúc của AMMS	86
Hình 4.12	Vùng phủ sóng của OBS và NBS	92
Hình 4.13	Sơ đồ hoạt động của AMMS	93

## BẢNG KÝ HIỆU TOÁN HỌC

Ký hiệu	Ý nghĩa
$\{x, y, z\}$	Tập chứa các phần tử $x, y, z$
$\{a_1, a_2, \dots, a_n\}$	Tập chứa các số nguyên từ $a_1$ tới $a_n$
$[a, b]$	Tập các số thực trong khoảng $a \leq b$
$(a, b)$	Tập các số thực trong khoảng $a < b$
$[a, b)$	Tập các số thực trong khoảng $a < b$ gồm cả $a$
$(a, b]$	Tập các số thực trong khoảng $a < b$ gồm cả $b$
$x^{(i)}$	Đầu vào thứ $i$
$a$	Số thực $a$
$A$	Ma trận $A$
$[a_i]_n$ Hoặc $(a_1, \dots, a_n)$	Véc tơ hàng $a$ cấp $n$
$(a_1, \dots, a_n)^T$	Véc tơ cột $a$ cấp $n$
$[A_{ij}]_{mn}$	Ma trận $A$ cấp $m \times n$
$f : A \rightarrow B$	Hàm $f$ với tập xác định $A$ và tập giá trị $B$
$f(x)$	Hàm 1 biến $f$ theo biến $x$
$f(x, y)$	Hàm 2 biến $f$ theo biến $x$ và $y$
$f(x)'$ hoặc $\frac{df}{dx}$	Đạo hàm của $f$ theo $x$
$\frac{\delta f}{\delta x}$	Đạo hàm riêng của $f$ theo $x$



$\int_a^b f(x)dx$	Tích phân theo x trong khoảng [a, b]
$\log x$ hoặc $\ln x$	Lo ga rít tự nhiên

# MỞ ĐẦU

## 1. Lý do nghiên cứu

Cùng với sự phát triển của công nghệ Thông tin và Viễn thông, đặc tính di động đã và đang trở thành tính năng cơ bản của các mạng thông tin hiện tại và tương lai. Ngày càng có nhiều hình thức truyền thông mới ra đời như Mạng xã hội, Điện toán đám mây di động, IoT, ... làm nền tảng để xây dựng xã hội 5.0 [55], làm đa dạng các loại hình dịch vụ, tăng độ phức tạp và phạm vi hoạt động của mạng,... Do vậy, các kỹ thuật quản lý di động hiện tại đang gặp nhiều vấn đề trong việc xử lý yêu cầu chuyển giao, đáp ứng nhu cầu sử dụng dịch vụ và di chuyển của người dùng với QoS đảm bảo. Vấn đề đặt ra là phải xây dựng phương án lựa chọn hệ thống chuyển giao tối ưu, vừa nâng cao hiệu suất sử dụng tài nguyên mạng vừa đảm bảo chất lượng dịch vụ cam kết.

Ngày nay các thiết bị và ứng dụng di động phát triển mạnh mẽ, kéo theo nhu cầu sử dụng dịch vụ di động tăng cao. Theo dự báo của IDC (tháng 6 năm 2019), số lượng thiết bị IOT sẽ đạt 41.6 tỷ trên toàn thế giới vào năm 2025, trong khi năm 2017 mới chỉ là 27 tỷ [41] [42]. Do vậy, nhu cầu truyền tải dữ liệu hàng tháng sẽ tăng từ 20EB năm 2018 lên 119EB năm 2023 [71]. Từ thực tế này, các ứng dụng di động như mạng xã hội, điện toán đám mây, IoT,... mở ra mô hình kiến trúc hạ tầng mạng mới, đối tượng mới bao gồm đầu cuối, dịch vụ, mạng, nội dung, tính toán, mã hóa... tất cả đều di động [16].

Xu hướng hội tụ công nghệ và nhu cầu sử dụng dịch vụ đa dạng mọi lúc, mọi nơi của người sử dụng đòi hỏi các nhà cung cấp dịch vụ phải thiết lập các hệ thống hạ tầng mạng có khả năng tích hợp nhiều loại hình dịch vụ, các loại lưu lượng khác nhau lên trên một hạ tầng truyền thông duy nhất. Hệ thống hạ tầng truyền thông đó về cơ bản phải hỗ trợ các cơ chế truyền dữ liệu tốc độ cao bằng dịch vụ best-effort, đồng thời vẫn đảm bảo QoS cho các loại hình dịch vụ khác nhau, nhất là các dịch vụ thời gian thực. Hạ tầng truyền thông như vậy được gọi là một mạng vô tuyến hỗn hợp băng rộng đa dịch vụ [29] (BcN).

Mạng BcN cho phép người dùng sử dụng các dịch vụ đồng thời mọi lúc, mọi nơi, thông qua các loại đầu cuối khác nhau mà không cần quan tâm tới công nghệ của mạng mình đang kết nối tới, chỉ cần ký thoả thuận với nhà cung cấp dịch vụ (SLA). Vì thế một trong những yêu cầu đối với mạng BcN là phải tối ưu hoá chức năng xử lý chuyển giao liên mạng (ISHO) để đáp ứng các yêu cầu này của khách hàng. ISHO phải có khả năng chuyển hướng dòng dữ liệu của ứng dụng giữa các mạng khác nhau mà vẫn duy trì tính liên tục cho các ứng dụng đó theo QoS thoả thuận. Do vậy, hiện tại có nhiều Viện nghiên cứu, trường đại học, các tổ chức trong và ngoài nước,... đã và đang nghiên cứu tìm các giải pháp tối ưu khác nhau xử lý yêu cầu này. Tuy nhiên, cho đến nay các nghiên cứu về ISHO phần lớn chỉ tập trung vào việc giải quyết vấn

đề kích hoạt ISHO khi dịch vụ người dùng bị ngắt quãng do di chuyển ra khỏi vùng phủ sóng [53][73]. ISHO cho mạng BcN không chỉ đáp ứng được yêu cầu trên, mà phải xử lý được các vấn đề như dự báo gián đoạn kết nối, cải thiện QoS, giảm trễ do chuyển giao, và các yêu cầu của người dùng như loại hình dịch vụ, tiết kiệm nguồn tiêu thụ, cước phí,...

Điều khiển chuyển giao là kỹ thuật cung cấp khả năng cho các đối tượng di động trao đổi thông tin và truy cập dịch vụ mọi lúc, mọi nơi. Thời gian đầu chỉ được áp dụng cho hệ thống thông tin di động tế bào (Cellular), sau đó nó không ngừng được phát triển để ứng dụng cho các mạng khác như Internet, Mobile Internet, Ubiquitous, ICN (Information Centric Network),... và mạng tương lai [66].

Với nhu cầu sử dụng dịch vụ di động ngày càng cao và sự phát triển không ngừng của các hệ thống truy nhập có công nghệ khác nhau (LTE, NB, WLAN, Bluetooth, Wifi,...), yêu cầu cung cấp dịch vụ tích hợp liên tục cho thuê bao di động ngày càng tăng, do đó việc nghiên cứu đưa ra các giải pháp ISHO mới có khả năng hoạt động xuyên suốt giữa các hệ thống khác nhau vẫn đang là cơ hội và thách thức cho nhiều nhà nghiên cứu. Hơn nữa, các mạng di động trong tương lai vẫn sẽ sử dụng hạ tầng IP như là mạng lõi [32]. Do vậy, bước đột phá kiến trúc mạng trong tương lai sẽ là sự hội tụ giữa Internet (như là mạng lõi) và mạng truy nhập vô tuyến với các công nghệ khác nhau. Để đạt được mục tiêu này, cần có các kỹ thuật ISHO mới, tối ưu đáp ứng nhu cầu dịch vụ và di chuyển của ứng dụng.

Tại nước ta, Chính phủ đã ban hành Quy hoạch phát triển viễn thông và công nghệ thông tin Việt Nam đến năm 2020 và định hướng tới năm 2030 [42], theo đó hệ thống mạng viễn thông đang từng bước được chuyển sang mạng hội tụ băng rộng đa dịch vụ (BcN) theo từng giai đoạn phù hợp với yêu cầu thực tế, nhằm cung cấp đa dịch vụ trên một hạ tầng viễn thông thống nhất. Vì vậy, các vấn đề liên quan đến việc xây dựng các cơ chế quản lý, điều khiển trên BcN cũng đang được giới khoa học chuyên ngành trong nước quan tâm nghiên cứu.

## **2. Mục đích nghiên cứu**

Mục đích nghiên cứu của luận án là đề xuất một phương pháp quản lý chuyển giao linh hoạt nhằm tối ưu hóa quản lý chuyển giao trong mạng Vô tuyến hỗn hợp đa dịch vụ.

## **3. Đối tượng và phạm vi nghiên cứu**

Đối tượng nghiên cứu của luận án là cơ chế quản lý chuyển giao trong mạng BcN.

Phạm vi nghiên cứu của luận án này được giới hạn trong việc nghiên cứu các giao thức (protocol) và các giải thuật (procedure, algorithm) quản lý chuyển giao. Luận án chỉ tập trung vào việc phân tích hiệu suất chuyển giao dựa trên các giao thức đã có như MIP, TCP-M, SIP từ đó xây dựng các tham số có ảnh hưởng tới chất lượng dịch vụ khi thực hiện chuyển giao,

xây dựng phương thức dự báo và đặt trước băng thông để đi đến việc đề xuất phương pháp quản lý chuyển giao linh hoạt nhằm tối ưu hóa quản lý chuyển giao trong mạng BcN. Mọi vấn đề khác liên quan đến quản lý vị trí của ứng dụng, cập nhật và thống kê các thông tin liên quan đến hướng di chuyển của người dùng,... coi như đã được giải quyết bằng các công cụ khác nằm ngoài phạm vi nghiên cứu của luận án này.

#### **4. Phương pháp và công cụ nghiên cứu**

Luận án sử dụng phương pháp nghiên cứu tổng hợp và phân tích: tổng hợp các kết quả nghiên cứu đã có của các tác giả khác có liên quan đến đề tài và phân tích các yếu tố liên quan đến vấn đề quản lý di động trong mạng BcN, qua đó đề xuất phương pháp quản lý di động linh hoạt nhằm tối ưu hóa quản lý di động trong mạng BcN.

Luận án sử dụng các công cụ toán học và các công cụ của lý thuyết hệ thống, lý thuyết điều khiển để giải quyết yêu cầu nghiên cứu. Do chưa có các chuẩn chung thống nhất về kết cấu, giao thức và công nghệ của BcN, nên trước tiên luận án đưa ra một mô hình chung được chấp nhận rộng rãi của mạng BcN, và lấy đó làm cơ sở để xây dựng và đề xuất cơ chế điều khiển chuyển giao.

#### **5. Ý nghĩa khoa học và thực tiễn của đề tài**

Ý nghĩa khoa học của đề tài được thể hiện qua các đóng góp mới của luận án, bao gồm: Thông qua phân tích hiệu suất chuyển giao dựa trên các giao thức đã có để xây dựng các tham số cơ bản ảnh hưởng nhiều nhất tới QoS khi chuyển giao; Sử dụng lý thuyết Bayes để xác định xác suất chuyển giao, từ đó xây dựng phương pháp đặt trước băng thông nhằm đảm bảo QoS và nâng cao hiệu suất sử dụng tài nguyên mạng; Cuối cùng dựa trên việc xác định được các yếu tố ảnh hưởng tới QoS và phương pháp đặt trước băng thông, luận án đề xuất phương pháp quản lý chuyển giao linh hoạt, tùy thuộc vào loại ứng dụng để đạt được mục tiêu tối ưu hóa quản lý chuyển giao.

#### **6. Cấu trúc của Luận án**

Luận án gồm phần Mở đầu, phần Kết luận và 04 chương nội dung. Sau phần Mở đầu, Chương 1, giới thiệu tổng quan về cấu trúc, đặc trưng và yêu cầu trong việc quản lý chuyển giao của mạng BcN; Phân tích và đánh giá các nghiên cứu liên quan đến luận án, nêu lên các vấn đề còn tồn tại cần phải giải quyết, từ đó đưa ra hướng nghiên cứu của luận án. Chương 2, phân tích và đánh giá hiệu suất chuyển giao dựa trên các giao thức đã có, từ đó xây dựng tập các tham số cơ bản ảnh hưởng đến QoS khi chuyển giao. Chương 3, tính toán xác suất chuyển giao  $P_b$  và xây dựng phương pháp đặt trước băng thông cho các ứng dụng. Chương 4, xây

dụng phương pháp quản lý chuyển giao linh hoạt. Cuối cùng, phần Kết luận trình bày các kết quả mới của luận án và các định hướng nghiên cứu tiếp theo, cụ thể luận án đã:

- Đề xuất giao thức định tuyến theo yêu cầu - EEMA cho MANET. EEMA chọn tuyến tối ưu cho chuyển giao dựa trên: số bước nhảy và hàm chi phí, và cân đối giữa trễ và năng lượng tiêu thụ. Đề xuất này được công bố trên tạp chí JCM, tên công trình.

- Đề xuất tính toán *xác suất chuyển giao*  $P_b$ . Xây dựng cơ chế *đặt trước băng thông* cho các ứng dụng có  $P_b$  lớn hơn ngưỡng chuyển giao, nhằm duy trì QoS cho các ứng dụng dạng A-E. Công bố tại hội nghị NICS 12/2019.

- Đề xuất sử dụng các giao thức định tuyến chủ động (OLSR, DSDV ) phù hợp với các mạng có cấu trúc ổn định, và các giao thức định tuyến theo yêu cầu (AODV) phù hợp với các mạng có tính di động cao. Công bố trên tạp chí JCM 5/2020.

- Đề xuất giải pháp quản lý chuyển giao linh hoạt (AMMS) nhằm khai thác hiệu quả tài nguyên mạng, đáp ứng yêu cầu QoS và mức tiêu thụ năng lượng cho các loại ứng dụng A,B,C,D,E.

# Chương 1: Tổng quan các vấn đề cần nghiên cứu

## 1.1 Mạng thông tin vô tuyến băng rộng đa dịch vụ BcN.

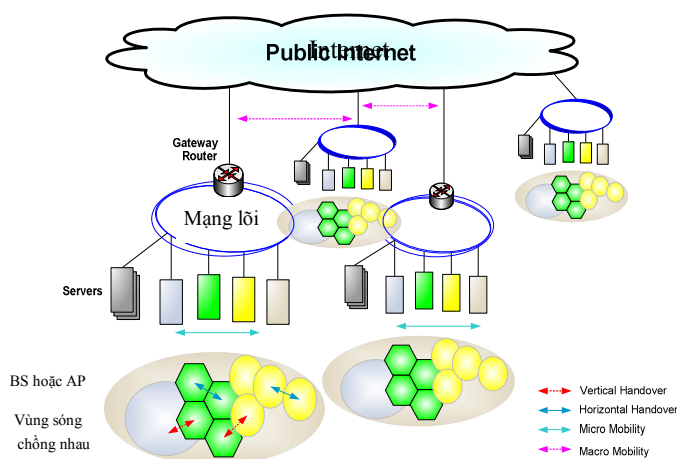
Trong những năm gần đây, với sự phát triển vượt bậc của các công nghệ truy nhập vô tuyến và thiết bị đầu cuối di động (như máy tính xách tay, Smart phone, IOT,...) làm cho ứng dụng di động đa dạng hơn, nhu cầu về kết nối, tốc độ, băng thông tăng nhanh hơn về thoại và các dịch vụ cơ bản [10]. Trong 5 năm từ 2014 – 2018, lưu lượng dữ liệu thông qua các thiết bị di động tăng gấp gần 5 lần [91], dự kiến tăng 6 lần trong giai đoạn 2017 – 2022 [10]. Công nghệ truy cập vô tuyến bao gồm nhiều cơ sở hạ tầng khác nhau, như công nghệ Bluetooth cho người sử dụng cá nhân, IEEE 802.11 cho mục đích mạng nội bộ, LTE/5G cho thông tin di động diện rộng, NB-IOT cho kết nối IOT, các mạng vệ tinh cho mục đích kết nối toàn cầu,... Những mạng này được thiết kế độc lập với nhau để cung cấp các dịch vụ cụ thể theo yêu cầu. Sự khác biệt giữa các mạng này thông qua các thông số dịch vụ [52][73], được tóm tắt như sau:

- **Tốc độ truyền tải dữ liệu:** các mạng vệ tinh và mạng tế bào có thể truyền với tốc độ dữ liệu vào khoảng 2 Mbps, LTE tối đa là 100Mbps, 5G tối đa là 1Gbps,...
- **Độ trễ truy cập:** độ trễ truy cập một chiều trong kết nối mạng nội bộ vô tuyến (WLAN) là không lớn, cỡ một vài trăm mili-giây cho một round-trip-time (RTT), và vào khoảng một giây trong các kết nối 3G do các xử lý thêm ở lớp vật lý như chuyển tiếp lỗi đã chỉnh (FEC), độ trễ khi truyền [31]. Độ trễ truy cập sẽ còn lớn hơn rất nhiều ở các kết nối vệ tinh, có thể lớn tới 270 ms [3].
- **Độ che phủ:** các mạng vệ tinh hay các mạng di động tế bào có thể cung cấp độ che phủ lớn (toàn thành phố, toàn bộ lãnh thổ 1 quốc gia) đến rất lớn (toàn cầu). Tuy nhiên các mạng 802.11 và các mạng nội bộ khác chỉ che phủ được ở một vùng giới hạn.

Do đó, hiện tại chưa có một mạng vô tuyến nào có thể cùng lúc thỏa mãn các yêu cầu như là tốc độ truyền dữ liệu cao, độ trễ thấp, độ che phủ mọi nơi để phục vụ cho các yêu cầu về dịch vụ từ phía người sử dụng di động [73]. Các nghiên cứu về HO hiện tại chưa giải quyết được sự tương tác giữa chi phí HO và chức năng tăng công suất của mạng [61].

Tuy nhiên, vì các mạng vô tuyến này bù đắp và bổ sung cho nhau [1], nên sự kết hợp giữa chúng có thể cung cấp các kết nối tốt nhất [34] cho người sử dụng. Hình 1.1 đưa ra một cấu trúc ví dụ về sự kết hợp của một hệ thống vô tuyến, bao gồm mạng 3G/LTE/5G/NR, mạng vệ tinh, và mạng WLAN được kết hợp chung với nhau. Ngoài ra, các mạng khác như là mạng Bluetooth, Home RF, MANETs, IOT,... cũng có thể được tích hợp vào trong mạng ở hình 1.1 này. Trong cấu trúc mạng này, người dùng sử dụng các thiết bị đa giao diện vô tuyến, nên có thể kết nối được với các loại mạng có công nghệ khác nhau. Bằng việc sử dụng các thiết bị này

cho phép người sử dụng di động luôn luôn có được kết nối tốt nhất đến một hoặc nhiều mạng. Ví dụ, khi người dùng đang trong vùng phủ sóng của mạng WLAN (mạng WLAN ở văn phòng, sân bay, ...), họ sẽ truy cập thông qua các mạng WLAN này. Trong trường hợp người dùng ở xa mạng WLAN, ví dụ trên đường cao tốc, họ sẽ được kết nối đến các mạng 3G/LTE. Trong trường hợp khi không có kết nối đến mạng WLAN hay mạng 3G/LTE, người dùng có thể sử dụng mạng vệ tinh. Và trong tất cả các trường hợp, khi người dùng di chuyển từ mạng này sang mạng khác, thiết bị của họ sẽ tự động chuyển sang các mạng mới này trong khi các ứng dụng đang chạy không bị ngắt kết nối. Do đó, người dùng có thể sử dụng các mạng vô tuyến khác nhau này như là một hệ thống vô tuyến tích hợp đa dịch vụ hay còn gọi là BcN.



Hình 1.1: Cấu trúc ví dụ của một hệ thống vô tuyến tích hợp

Những vấn đề khó khăn khi tổ chức mạng BcN, đó là:

- **Công nghệ truy cập:** các mạng khác nhau sử dụng các công nghệ truy cập vô tuyến khác nhau, ví dụ GPRS sử dụng công nghệ đa truy cập phân chia thời gian (TDMA), công nghệ UMTS/LTE sử dụng công nghệ đa truy cập phân chia mã (CDMA), và công nghệ WLAN sử dụng các mô hình truy cập ngẫu nhiên như là CSMA/CA.
- **Các giao thức mạng:** các mạng khác nhau sử dụng các giao thức khác nhau cho việc vận chuyển, định tuyến, quản lý di động, xác thực, tính cước, ...
- **Dịch vụ cung cấp:** những mạng này thuộc về các nhà cung cấp dịch vụ khác nhau nên sẽ khó khăn trong việc xây dựng thoả thuận dịch vụ (SLA) dùng chung.
- **Yêu cầu lưu lượng cao:** 5G đáp ứng được yêu cầu này bằng cách thu hẹp phạm vi phủ sóng của BS, số lượng người dùng trên mỗi BS sẽ giảm và cải thiện được hiệu suất tái sử dụng tần số. Tuy nhiên do nhu cầu di chuyển và trao đổi dữ liệu của MT không ngừng tăng lên, làm cho yêu cầu xử lý HO càng tăng cao hơn, dẫn đến lưu lượng dành cho báo hiệu tăng, làm giảm lưu lượng dành cho MT [61]

Do đó, việc liên tục nghiên cứu cải tiến các giải pháp HO, tích hợp các mạng truy cập vô tuyến này với nhau để cung cấp kết nối thông suốt cho người dùng là cần thiết.

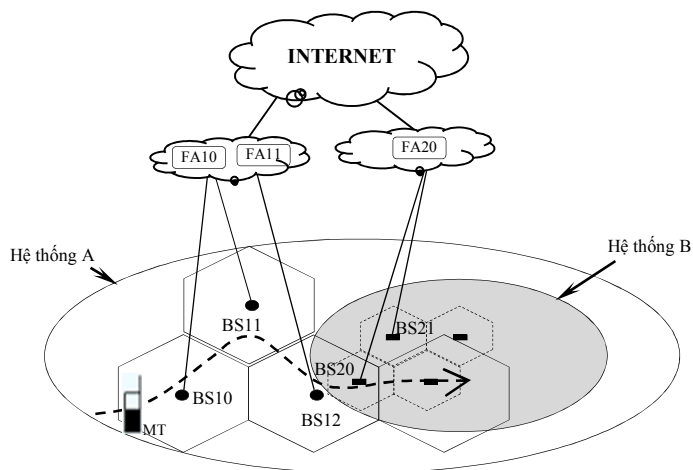
## 1.2 Yêu cầu chuyển giao trong mạng BcN

Trong nhiều năm qua, các kỹ thuật xử lý HO phức tạp đã được nghiên cứu phát triển và triển khai để quản lý việc đăng ký, xác thực, di chuyển,... của MT. Tuy nhiên các kỹ thuật này phù hợp với các mạng đồng nhất, áp dụng cho BcN thì hiệu suất sẽ không cao vì kích thước tế bào giảm và tốc độ di chuyển của MT tăng [52]. Do vậy, việc đáp ứng được các vấn đề nêu trên là vấn đề còn nan giải, cần giải quyết [77] [66]:

- Nhiều ứng dụng và mô hình truyền tin xuất hiện làm cho hạ tầng mạng thêm phức tạp, đa dạng, và có quan hệ với nhau. Nhiều giải pháp quản lý di động hiện có đều ứng dụng cho các hệ thống (mạng) cụ thể, chưa đáp ứng được yêu cầu dự phòng và khả năng xử lý cho mạng hỗn hợp phức tạp;

- Đang tồn tại nhiều mạng truy cập vô tuyến có công nghệ khác nhau;
- Nhiều loại dịch vụ khác nhau và không linh hoạt;
- Nhiều kỹ thuật HO đã được triển khai;
- Thiếu cơ chế thích hợp để giải quyết khác biệt giữa các hệ thống thông tin nêu trên.

Bên cạnh các vấn đề nêu trên, chi phí báo hiệu liên quan đến tính di động, ví dụ: đo lường / báo cáo quản lý tài nguyên vô tuyến, tăng độ trễ chuyển giao, lỗi mạng và tốc độ là các vấn đề cần quan tâm trong mạng di động ngày nay. Theo đó, mạng BcN yêu cầu một phương pháp quản lý di động mới, chính xác để đối phó với những thách thức này [77].



Hình 1.2: Sự di chuyển trong các hệ thống vô tuyến thế hệ tiếp theo



Ở mạng BcN, có hai kịch bản chuyển giao, đó là: chuyển giao trong mạng (Horizontal Handover) và chuyển giao ngoài mạng (Vertical Handover) như trình bày ở hình 1.2. Sự dịch chuyển của người sử dụng di động giữa hai BS của một hệ thống (từ BS10 sang BS11 trong hình 1.2) gọi là chuyển giao trong mạng, và sự dịch chuyển của người sử dụng di động giữa hai hệ thống khác nhau (từ BS12 sang BS20 trong hình 1.2) gọi là chuyển giao ngoài mạng. Một yếu tố cần thiết là các ứng dụng đang chạy trên thiết bị di động sẽ không bị tác động của sự chuyển giao này, cả chuyển giao trong và ngoài mạng, để tránh gián đoạn dịch vụ và đảm bảo QoS. Để đáp ứng được yêu cầu này phải xử lý được việc giảm lỗi, trễ và mất số liệu trong quá trình chuyển giao.

Ở mạng đồng nhất (cùng công nghệ vô tuyến), các MT sử dụng tập các tham số TTT, HM, A3,... để xử lý HO trong toàn mạng. Tuy nhiên, nếu sử dụng bộ tham số này để xử lý HO trong BcN thì hiệu suất quản lý di động sẽ bị giảm [17]. Bởi vì khi MT cố gắng kết nối với một tế bào đang quá tải (trong khi tải ở các tế bào lân cận khác có thể còn dư), thì QoS sẽ không đảm bảo và tỷ lệ HOF tăng do thiếu hụt tài nguyên. Do đó, quản lý HO linh hoạt trong mạng di động là cần thiết để nâng cao hiệu suất mạng. Các yếu tố sau cần quan tâm khi xây dựng phương án xử lý HO [52]:

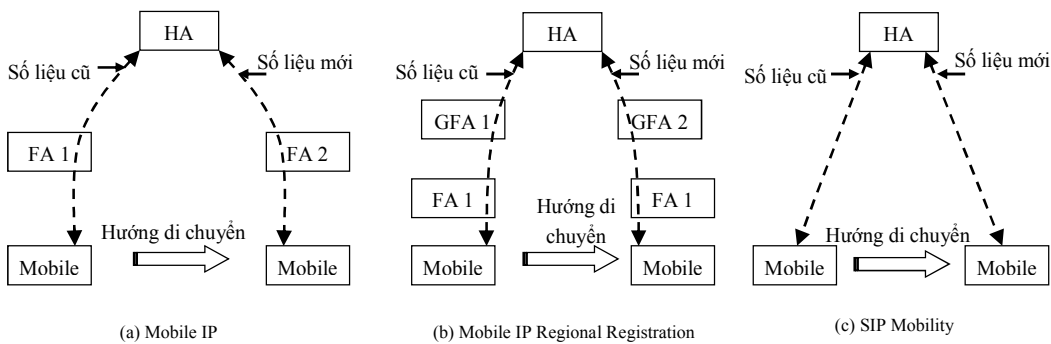
- MSE có vai trò quan trọng trong việc cải thiện hiệu suất HO. Thông qua MSE ta có thể lựa chọn một cách linh hoạt tập tham số HM, A3, TTT. Sử dụng thông tin về “số lần chọn lại tế bào”, “Suy hao tần số và RSS” và “mức độ di chuyển của MT là chậm, vừa hay nhanh” làm tăng độ chính xác về dự báo HO. Khi MT di chuyển với tốc độ thấp trong khu vực có mật độ tế bào Micro cao, thì nên tắt nguồn tế bào Macro trong khu vực này để tiết kiệm năng lượng.
- Khi MT đang có kết nối vô tuyến tốt, thì chỉ khi thủ tục HO yêu cầu mới kích hoạt EHOP và EHO để giảm các trường hợp RLF trước khi thực hiện HO.
- Ước lượng RSS tại biên tế bào và tiến hành catching dữ liệu để đề phòng việc hết tài nguyên vô tuyến và đảm bảo QoS.
- Lựa chọn tế bào ít sử dụng nhất trong số 3 tế bào lân cận. Việc này sẽ giúp cho cân bằng tải một cách tự động.
- Kích hoạt chế độ DC và kết nối tới tế bào Macro khi MT di chuyển ở tốc độ cao. Ước lượng thời gian MT lưu trú tại tế bào Micro. Khi tế bào Micro có kết nối vô tuyến tốt, và thời gian MT lưu trú tại đây lớn hơn ngưỡng tối thiểu thì chuyển kết nối của MT sang tế bào này.
- Đặt trước băng thông tại các tế bào lân cận đang thừa tải để giảm thời gian thực hiện HO khi có yêu cầu.

### 1.3 Phân tích và đánh giá các nghiên cứu liên quan đến Luận án.

Phần này luận án trình bày về sự giống và khác nhau trong phương pháp quản lý di động giữa các mạng IP và mạng số liệu gói, do các tổ chức nghiên cứu 3GPP và 3GPP2 đề xuất. Đặc biệt, chúng tôi quan tâm đến giao thức MIP, Đăng ký miền cho MIP, tính di động của đầu cuối SIP, quản lý di động trong mạng 3GPP, 3GPP2, Cellular IP và HAWAII.

Trước hết Luận án so sánh các cấu trúc quản lý di động cơ bản, được sử dụng để vận chuyển và quản lý sự thay đổi của các gói số liệu chuyển tới MT. Ở đây, các phần tử mạng tham gia vào quá trình xử lý bản tin chính là các phần tử giao thức quản lý di động. Ở giao thức MIP thì Home Agent (HA), Foreign Agent (FA) và MT là những phần tử của giao thức di động. Các nút mạng khác (như các Router chuyển tiếp) nằm trên toàn tuyến kết nối với các thực thể giao thức di động này đều không phải là thực thể giao thức di động.

Hình 1.3 dưới đây minh họa các cấu trúc quản lý di động cơ bản dùng trong MIP, đăng ký miền cho MIP, SIP mobility.



Hình 1.3: Các mô hình quản lý di động trong MIP, MIP-RR và SIP

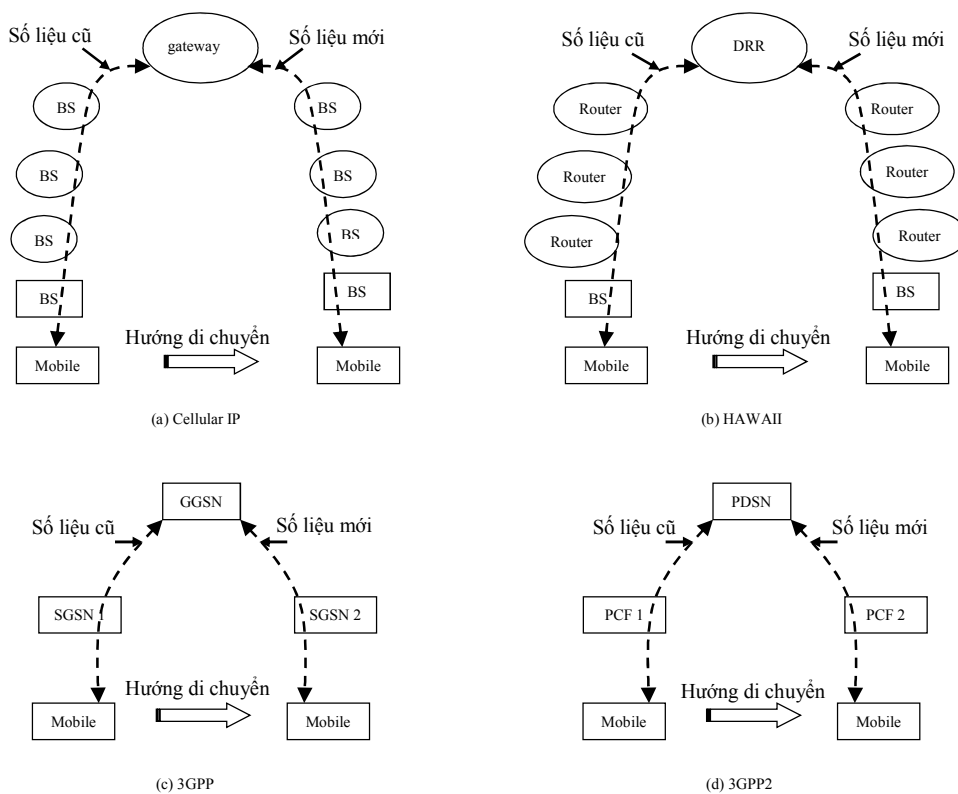
Hình 1.3 chỉ thể hiện một máy chủ SIP (SIP server), mà không thể hiện các máy chủ proxy SIP có thể được dùng đến trong các mạng khách. Cấu trúc quản lý di động cơ bản dùng trong mạng gói 3GPP, 3GPP2, Cellular IP và HAWAII được thể hiện trong hình 1.4.

Sự giống nhau cơ bản của các phương pháp quản lý di động được thể hiện trong hình 1.3 và hình 1.4 đó là, tất cả các phương pháp đều sử dụng chính sách chuyển hướng giao vận (relayed delivery) [36] [92] như là một cơ chế cơ bản để phân phát các bản tin báo hiệu, các gói số liệu của người dùng, hoặc đồng thời cả các bản tin báo hiệu và ứng dụng của người dùng đến MT. Đặc biệt, điểm neo di động (mobility anchor) được dùng để giám sát vị trí của các MT và chuyển hướng truyền gói số liệu đến MT. Cụ thể:

- Với giao thức MIP và MIP - RR, gói số liệu trước hết được định tuyến tới các HA quản lý trực tiếp MT, sau đó mới chuyển tới MT, nếu MT lúc này không có mặt ở trạm HA, khi đó gói số liệu được chuyển tới FA đang tạm thời quản lý MT, sau đó mới chuyển tới đích cuối

cùng là MT. MIP có chức năng định tuyến tối ưu cho phép Host tương ứng nhận biết vị trí hiện tại của MT rồi sử dụng phương pháp phân phát trực tiếp [36] để truyền các gói số liệu đến MT.

- Với giao thức SIP Mobility, trước khi Server tương ứng biết được địa chỉ đích, nó liên tục gửi bản tin SIP khởi động đến Server đích. Tùy thuộc vào loại SIP server được dùng trong mạng đích mà nó có thể hoặc chuyển bản tin báo hiệu đến MT hoặc thông báo cho Server tương ứng vị trí hiện tại của MT để Server này liên lạc trực tiếp với MT. Trường hợp Server liên lạc trực tiếp được với MT, nó gửi tiếp các bản tin báo hiệu và số liệu của người dùng trực tiếp tới MT mà không cần thông qua Server đích.



Hình 1.4: Các mô hình quản lý di động trong 3GPP, 3GPP2 và Micro mobility

- Với Cellular IP, các gói số liệu đến và đi từ MT trước hết được định tuyến đến router đầu tiên, router này có nhiệm vụ chuyển các gói số liệu đến đích cuối cùng.
- Với HAWAII, các gói số liệu đi và đến MT trước hết được định tuyến đến một Router miền gốc (Domain root router), router này cũng có nhiệm vụ chuyển các gói số liệu đến đích cuối cùng như trong Cellular IP.
- Trong mạng 3GPP, tất cả các gói số liệu của người dùng được định tuyến đến GGSN, GGSN có nhiệm vụ chuyển các gói số liệu đến đích cuối cùng.

- Trong mạng 3GPP2, tất cả các gói số liệu đầu tiên cũng được định tuyến đến PDSN, PDSN có nhiệm vụ chuyển các gói số liệu đến đích cuối cùng...

Sự khác nhau cơ bản đầu tiên giữa các phương pháp quản lý di động được thể hiện trên hình 1.3 và hình 1.4 là cách thức các gói số liệu được chuyển đi từ một thực thể giao thức di động tới thực thể khác. Chẳng hạn:

- Với MIP và MIP-RR, các HA sử dụng các tuyến IP-IP để truyền các gói số liệu đến COA hiện tại của MT.

- Với Mobile SIP, SIP server dùng một tuyến IP nào đó và chuyển các gói số liệu đến một SIP server khác hoặc là tới đích ứng dụng của người dùng.

- Mạng 3GPP và 3GPP2 thì dùng một tuyến nhất định để trao đổi gói số liệu của người dùng giữa một điểm neo di động (GGSN trong 3GPP và PDSN trong 3GPP2) và MT. Tuy nhiên, các mạng này thiết lập kênh truyền theo các cách khác nhau [92], đó là:

- 3GPP dùng giao thức GPRS để thiết lập tuyến liên lạc bên trong mạng lõi và giữa mạng lõi với RAN.

- 3GPP2 dùng giao thức IP Tunnel, GRE, được định nghĩa trong IETF để thiết lập tuyến IP chuyển tải các gói số liệu của người dùng qua mạng lõi, sau đó các tuyến này được nối với các kênh vô tuyến trong RAN theo kiểu kết nối PPP để trao đổi các gói số liệu giữa MT và điểm neo di động của MT đó, ở đây chính là PDSN.

- Cellular IP và HAWAII sử dụng lớp IP để định tuyến và truyền số liệu từ một điểm neo di động (là Gateway trong Cellular IP hoặc là DRR trong HAWAII) đến MT.

Một điểm khác nhau cơ bản giữa các phương pháp quản lý di động nữa đó là mối liên hệ giữa quản vị trí và quản lý tuyến.

- Đối với MIP, MIP-RR và Mobile SIP, việc quản lý vị trí được tách biệt với việc định tuyến lớp ở lớp IP. Các server trung tâm (như MIP, HA, SIP server) được dùng để chỉnh sửa thông tin về vị trí. Các gói số liệu từ một thực thể giao thức di động đến một thực thể khác hoặc là được định tuyến qua route IP (trong trường hợp SIP mobility) hoặc là được truyền qua các tuyến IP ( trong trường hợp MIP và MIP-RR).

- Cellular IP và HAWAII lại tích hợp quản lý vị trí với định tuyến lớp IP. Không có server chuyên quản lý vị trí. Thay vào đó, mạng duy trì vị trí của MT bằng cách duy trì một tuyến cụ thể tới mỗi MT đang hoạt động.

- Mạng 3GPP và 3GPP2 cũng dùng các tuyến host định trước. Nhưng các tuyến host cụ thể trong mạng lõi IP được thiết lập giống như các tuyến trên lớp truyền tải IP thay vì trên lớp IP

nhu trong Cellular IP hay HAWAII. Do vậy, quản lý vị trí ở đây được tích hợp với quản lý tuyến host định trước.

Điểm khác biệt cơ bản thứ 3 trong quản lý di động được thể hiện trong hình 1.3 và hình 1.4 đó là chức năng paging có được hỗ trợ không và nếu có thì cách thức hoạt động như thế nào.

- Với MIP và MIP-RR không hỗ trợ paging, mặc dù MIPv4 gần đây đã được mở rộng để hỗ trợ chức năng này.
- Với mạng gói 3GPP, các router biên (như SGSN) trong mạng lõi có nhiệm vụ khởi động chức năng paging. Quản lý di động được tích hợp với định tuyến host trong mạng lõi để loại bỏ bớt các GGSN liên quan đến quá trình paging.
- Với 3GPP2, tích hợp quản lý vị trí với định tuyến host để mạng lõi IP không phải liên quan đến paging. Thay vào đó, khi các gói số liệu đi vào mạng RAN CDMA2000, mạng RAN sẽ thực hiện paging khi cần bằng cách dùng các thủ tục và giao thức paging bên trong RAN.
- Với Cellular IP và HAWAII, cả hai đều xác định thủ tục paging cho riêng mình, chúng dùng cơ chế định tuyến host riêng biệt.

#### **1.4 Các vấn đề còn tồn tại**

Chuyển giao là quá trình đầu cuối di động (MT) di chuyển từ trạm phát (AP) này tới AP khác trong cùng mạng. Về cơ bản, có thể phân thành 2 loại chuyển giao: là chuyển giao đồng nhất (homogeneous - giữa các AP cùng công nghệ) và chuyển giao hỗn hợp (heterogeneous - giữa các AP có công nghệ khác nhau). Có nhiều vấn đề cần giải quyết để hỗ trợ cho MT thực hiện chuyển giao hỗn hợp giữa các mạng truy nhập (AN) có công nghệ khác nhau như 802.11, UMTS, CDMA, GPRS,... Mỗi AN đều có các đặc tính về QoS, Security, Băng thông,... khác nhau. Tương tự như vậy, việc MT di chuyển giữa các mạng của các nhà cung cấp dịch vụ (SP – Service Provider) khác nhau cũng có nhiều vấn đề cần quan tâm, như sự Xác thực (Authentication), sự Cho phép (Authorization) cũng như với các SP khác nhau thì các yêu cầu về thoả thuận dịch vụ (Service Level Agreement – SLA) cũng khác nhau. Để có thể cung cấp giải pháp H-HO tối ưu thì cần phải xác định những yếu tố (thông số) chính làm trễ quá trình HO.

Trong luận án có giới thiệu khái niệm về H-HO, minh hoạ các đặc tính cơ bản cần thiết trong truyền thông thời gian thực, trích dẫn các thủ tục H-HO giữa các AN và SP khác nhau, phân tích các yếu tố như trễ, suy hao, jitter,...nảy sinh trong quá trình H-HO. Các vấn đề này có thể được sử dụng để xây dựng giải pháp H-HO tối ưu.

##### **1.4.1 Phân loại chuyển giao**

Việc MT di chuyển giữa các AN và thay đổi kết nối có thể dẫn đến sự gián đoạn dịch vụ đang dùng. Trong suốt quá trình chuyển giao, do sự thay đổi kết nối mạng nên MT có thể kết thúc việc truyền tin sử dụng giao diện thứ hai ở mạng mới, thay đổi subnet hoặc domain hiện tại. Dựa trên hình thức di chuyển và kiểu mạng AN, ta có thể phân loại chuyển giao thành “Liên mạng - Intersubnet” và “Nội mạng – Intr subnet”.

Chuyển giao Inter Subnet có thể là chuyển giao:

- Công nghệ AN khác nhau (Ví dụ: CDMA, WLAN), Domain khác nhau;
- Khác công nghệ, cùng Domain;
- Cùng công nghệ, khác Domain;
- Cùng công nghệ, cùng Domain.

Còn chuyển giao Intra Subnet có thể là chuyển giao:

- Cùng công nghệ, cùng Domain;
- Khác công nghệ, cùng Domain.

HHO là loại chuyển giao cho phép cung cấp hoặc sửa đổi tài nguyên của MT, tương tác với khối Xác thực của Domain.

Intra-Subnet: Khi MT di chuyển giữa 2 mạng vô tuyến AN khác nhau, nhưng trong cùng Subnet và không thay đổi bộ định dạng thuộc lớp 3 của nó (ví dụ: Địa chỉ IP), thì gọi là chuyển giao trong cùng Subnet với công nghệ mạng khác nhau (VD: WLAN, UMTS,...). Tuy nhiên đối với HO giữa các AN có công nghệ khác nhau, trong cùng Subnet thì bộ định dạng L3 của MT có thể thay đổi nếu nó sử dụng giao diện với mạng AN khác để truyền tin.

Inter-Subnet: Chuyển giao giữa các Subnet khác nhau xảy ra khi MT di chuyển giữa các mạng khác nhau, các mạng này thuộc vào Subnet khác nhau. Khi đó bộ định dạng L3 (VD: Địa chỉ IP) sẽ thay đổi, điều này yêu cầu giao thức quản lý di động như MIP [58][59], MIP6 [35], SIP Mobility [72],... cần phải duy trì các ứng dụng mà MT đang sử dụng. Chuyển giao liên Subnet là nguyên nhân tiềm tàng làm mất số liệu và Jitter do trở liên quan tới sự chuyển dịch tại L2 và L3.

Inter-Technology: MT có thể được thiết kế tích hợp nhiều giao diện vô tuyến khác nhau, mỗi giao diện hỗ trợ một loại công nghệ (CDMA, WLAN, UMTS,...). Khi chuyển giao, MT có thể di chuyển từ AN công nghệ này sang AN có công nghệ khác (VD: từ WLAN sang CDMA). Loại chuyển vùng này còn được gọi là chuyển giao theo chiều đứng (Vertical HO). Do băng thông của các mạng AN khác nhau thì thường cũng khác nhau, nên loại chuyển giao này có thể ảnh hưởng tới QoS của các dịch vụ Multimedia [23].

Intra-Technology: Là chuyển giao giữa các mạng AN cùng công nghệ. MT có thể là loại đa hoặc đơn giao diện vô tuyến. Nếu chuyển giao thực hiện giữa các Subnet khác nhau, khi đó bộ định dạng L3 sẽ thay đổi.

Inter-Domain: Một Domain ở đây được hiểu là một vùng bao gồm một hoặc nhiều mạng (có thể cùng hoặc khác công nghệ) do một SP quản lý. Chuyển vùng xảy ra khi MT di chuyển giữa các vùng của các SP khác nhau gọi là chuyển vùng Inter-Domain. Loại chuyển giao này có thể làm nảy sinh trễ do chuyển giao Subnet lặp lại.

Intra-Domain: Là chuyển vùng khi MT di chuyển giữa các mạng AN trong cùng một SP.

#### 1.4.2 Các yêu cầu về hiệu suất

Để cung cấp QoS đảm bảo cho các dịch vụ Multimedia thì cần phải giới hạn trễ kết cuối, jitter, suy hao số liệu ở một ngưỡng nhất định. ITU đã định nghĩa các giới hạn này. Ví dụ đối với trễ một chiều, ITU-T G114 khuyến nghị giới hạn trên cho hầu hết các ứng dụng là 150ms, không chấp nhận trễ lên tới 400ms,... Tương tự như vậy, gói số liệu xem như là bị mất nếu nó đến đích sau quãng thời gian vượt quá giới hạn quy định. Các khuyến nghị của IETF như [4], [5], [6] mô tả một vài kỹ thuật đo lường xác định trễ và jitter.

Trễ kết cuối bao gồm trễ mạng, trễ OS, trễ CODEC, và trễ ứng dụng. Trễ mạng bao gồm các loại trễ đường truyền, môi trường, hàng đợi ở các router trung gian. Trễ OS gồm trễ lập lịch cho bộ gửi và nhận. Trễ CODEC là trễ hình thành do quá trình đóng gói và mở gói tại bộ phát và nhận. Đối với các lưu lượng VoIP tương tác, thì trễ kết cuối ảnh hưởng đến jitter là **vấn đề quan trọng** cần quan tâm. Trong suốt thời gian chuyển giao thường xuyên của MT, lưu lượng tạm thời không thể tới được MT, điều này góp phần tạo nên jitter. Nếu hệ thống cuối có buffer phát thì jitter này được gộp vào trễ buffer phát, tuy nhiên nếu không có buffer phát thì jitter được gộp vào trễ lưu lượng tương tác. Tồn thất số liệu chủ yếu do nghẽn gây ra, định tuyến không ổn định, kết nối lỗi. Chuyển giao là nguyên nhân gây nên mất số liệu do MT thay đổi kết nối mạng. Do vậy việc mất số liệu ở cả lưu lượng streaming và tương tác VoIP ảnh hưởng tới chất lượng dịch vụ của các ứng dụng thời gian thực. Số lượng gói số liệu bị mất tương đương với trễ chuyển giao và tốc độ lưu lượng mà MT đang nhận. Do đặc tính truyền lại, nên việc mất gói số liệu sẽ làm tăng nghẽn lưu lượng TCP, nhưng nó không làm gia tăng nghẽn trong trường hợp lưu lượng truyền dựa trên RTP/UDP. **Do vậy, cần thiết phải có các cơ chế quản lý di động đáp ứng được yêu cầu giảm việc mất số liệu và ảnh hưởng của trễ chuyển giao.** Dưới đây là các ví dụ liên quan đến yêu cầu về hiệu suất. Theo ETSI TR 101[ETSI], một cuộc thoại bình thường có thể chấp nhận mất 2% gói số liệu. Trong khi đàm thoại mà MT di chuyển giữa các AN dẫn đến hàng loạt chuyển giao sẽ diễn ra, mỗi chuyển giao đều gây nên việc mất số liệu. Do vậy, số liệu bị mất tối đa trong cuộc đàm thoại

cần phải giảm tới mức chấp nhận được. Chưa có ngưỡng rõ ràng cho các ứng dụng streaming, nhưng phải giảm tối đa tới mức có thể được để cung cấp QoS tốt nhất cho các ứng dụng định trước.

### 1.4.3 Phân tích trễ chuyển giao

Ảnh hưởng của HHO: MT tham gia vào quá trình HHO là chủ thể làm trễ HO bởi vì sự liên kết lại tài nguyên tại một vài lớp của giao thức. Có một vài tài nguyên chung tham gia vào quá trình liên kết các lớp này. Tài nguyên này phần lớn được cho là các đặc tính truy nhập ( Băng thông, đặc tính kênh), cơ chế truy nhập (CDMA, TDMA, CSMA/CA), cấu hình tham số lớp 3, xác thực lại, liên kết lại các mức security ở tất cả các lớp.

Bất kỳ ứng dụng multimedia nào đang chạy trong quá trình chuyển giao đều chịu ảnh hưởng của trễ bên trong mỗi lớp của stack giao thức. Do vậy cần thiết phải quan sát tại tất cả các lớp và phân tích sự liên kết lại một vài tài nguyên chung tại mỗi lớp của một MT tham gia vào HHO.

**Trễ lớp 2:** Tùy theo kiểu truy nhập (802.11, CDMA) mà MT thực hiện một vài bước chuyển dịch trước khi liên kết lớp 2 mới được thiết lập lại. Ví dụ: một kết nối 802.11 thực hiện quá trình dò tìm, xác thực và liên kết trong suốt thời gian liên kết tới kết nối mạng AN 802.11 mới. Tương tự như vậy, các mạng truy nhập khác như CDMA, GPRS thực hiện một loạt chuyển dịch trạng thái trong suốt thời gian liên kết tới điểm truy nhập mới. Mỗi chuyển dịch trạng thái này đều tạo ra trễ. Sau khi một liên kết mới được thiết lập, tùy theo loại HO mà các lớp khác của giao thức cần phục hồi. Ví dụ trường hợp HO cùng công nghệ và cùng subnet, chỉ có lớp 2 bị ảnh hưởng và trễ HO nảy sinh là do sự chuyển dịch trạng thái chỉ ở lớp 2. Một vài thủ tục xác thực cần được thực hiện khi MT chuyển kết nối sang mạng AN mới. Nếu các đặc tính lớp 2 như QoS giữa các mạng AN mới và cũ là khác nhau thì thủ tục xác thực này có thể cần nhiều bước hơn là trường hợp giống nhau. Vì thế sự khác nhau của các đặc tính ở lớp 2 có thể là điều kiện cho HO cùng công nghệ cũng như HHO.

**Trễ lớp 3:** Quá trình chuyển dịch ở lớp 3 thực hiện thông qua vài bước như tìm kiếm địa chỉ IP mới, dò địa chỉ trùng lặp, cập nhật ARP, và xác thực cục bộ. Vùng xác thực cục bộ thành công sẽ cho phép MT gửi số liệu theo HOP của router đầu tiên. Ở các mạng AN khác nhau và giao thức quản lý di động khác nhau (Ipv4, Ipv6) thì cách thức tìm kiếm địa chỉ IP mới cũng khác nhau. Có một vài giao thức như DHCP, DHCPv6, PPP hay cơ chế tự định cấu hình có thể được sử dụng để thực hiện tìm kiếm địa chỉ IP. Thời gian cần thiết để hoàn thành việc chuyển dịch tại lớp 3 đối với các phương pháp khác nhau là khác nhau. Địa chỉ lớp 3 cần thiết lập lại sau khi hoàn thành việc chuyển dịch tại lớp 3, có vài chức năng nào đó ở lớp trên cần được hoàn thành. Các chức năng này bao gồm việc gửi cập nhật liên kết từ MT, và định



hướng lại media tại host gửi. Trễ do các chức năng này tạo ra có thể xem như là do trễ đường truyền, trễ xử lý tại hệ thống cuối,... Đối với lớp di động với Subnet khác nhau nhưng cùng Domain, thì chỉ cần thủ tục xác thực tối thiểu đáp ứng được việc cấp và sử dụng địa chỉ IP trong Subnet mới. Tuy nhiên đối với lớp khác Subnet, khác Domain có thể phải cần thêm các xử lý xác thực nữa.

**Trễ lớp ứng dụng:** Đối với việc quản lý di động trong Inter-Domain điển hình, tính độc lập của HO cùng hoặc khác công nghệ, thủ tục xác thực cần được thực hiện để thiết lập tài nguyên lớp 2 và lớp 3 trong domain mới. Điều này có thể nảy sinh trễ bất thường không mong muốn trong quá trình HO. Nếu HO Inter-Subnet tham gia vào tính di động của Inter-Domain, thì trễ sản sinh là do thủ tục xác thực, nó ảnh hưởng tới các phiên dịch vụ multimedia đang thực hiện. Thủ tục xác này có thể gồm xác thực EAP nơi mà server AAA có thể tham gia vào bản tin EAP trong thời gian HO. Tùy theo kiến trúc mà các tín hiệu AAA đi qua tất cả các hướng tới server AAA trong vùng gốc (home domain) của MT cũng như trước khi dịch vụ mạng cấp cho MT trong mạng mới. Như vậy sự tương tác giữa MT và server AAA cần phải được ngăn ngừa hoặc giảm trong thời gian HO.

Trễ ở lớp ứng dụng là trễ cần để thiết lập lại hoặc sửa đổi tài nguyên tại lớp này, như các phiên định trước SIP. Quá trình cập nhật số liệu hoặc báo hiệu tại các phiên trung gian có thể được xem như là trễ, do vậy địa chỉ IP và thông số CODEC cho các phiên trung gian này có thể bị thay đổi.

#### 1.4.4 Tối ưu hoá trễ HO

Theo phân tích ở trên, trễ trong khi MT tham gia vào quá trình HO phụ thuộc nhiều vào loại HO. Các loại HO cùng Subnet, cùng công nghệ sẽ trễ ít hơn các loại khác Subnet, khác công nghệ bởi vì sự thay đổi trạng thái của MT trong quá trình HO cùng Subnet, cùng công nghệ ít hơn khác Subnet, khác công nghệ. Như vậy, sẽ có một lượng tài nguyên chung nhất định như đặc tính mạng AN, an toàn mạng tại mỗi lớp trong quá trình kết nối lại, cách thức liên kết tại lớp 3 (phân bổ địa chỉ IP, cập nhật liên kết, định tuyến lại dịch vụ), xác thực tại mạng mới đóng vai trò quan trọng trong việc xác định tổng thể trễ HO. **Như vậy, cần thiết phải có nghiên cứu lựa chọn mạng chuyển giao nhằm tối ưu hoá trễ nảy sinh trong quá trình HHO, đảm bảo QoS cung cấp cho các ứng dụng của MT mà người dùng đã thoả thuận với nhà cung cấp dịch vụ.** Hai hướng nghiên cứu chính để giảm trễ HO đã được nhiều chuyên gia quan tâm đó là:

- xử lý song song nhiều quá trình chuyển dịch trạng thái, và
- thực hiện trước một số chuyển dịch trạng thái trước khi quá trình HHO bắt đầu.

Một số bước (Xác thực, đặt trước băng thông,...) trong quá trình HO có thể được thực hiện trước nhằm giảm trễ. Cụ thể, nếu MT thực hiện HO giữa các mạng trong cùng Domain, thì ngay khi MT di chuyển vào Domain sẽ được cấp quyền truy cập tới tất cả các mạng thuộc Domain này. Nhờ vậy, thủ tục HO giữa các mạng sẽ giảm được công đoạn xác thực quyền truy cập, nghĩa là giảm được trễ HO.

## 1.5 Kết luận chương 1

Theo kết quả phân tích và đánh giá các nghiên cứu về quản lý di động ở trên (phần 1.3 và 1.4) cho thấy, còn nhiều vấn đề cần giải quyết để vừa đáp ứng nhu cầu sử dụng dịch vụ theo thỏa thuận giữa SP và người dùng, vừa đảm bảo khai thác tối ưu tài nguyên mạng. Việc tích hợp mạng di động và cố định thông qua hạ tầng IP sẽ thuận lợi và hiệu quả hơn việc cải thiện kết nối giữa chúng. Một đặc điểm của các mạng di động hiện tại là có công nghệ khác nhau (LTE, 4G/5G, NB-IOT, WLAN,...), cùng cung cấp các dịch vụ tương tự nhau (thoại, video, data,...), vấn đề đặt ra là làm thế nào để khai thác hiệu quả hạ tầng mạng đã đầu tư mà vẫn đáp ứng nhu cầu sử dụng dịch vụ chất lượng, mọi lúc, mọi nơi của người dùng. Những kỹ thuật quan trọng nhất đã giới thiệu để giải quyết một phần yêu cầu này là thiết kế các AP mà MT đa giao tiếp có khả năng tự cấu hình để tương thích lẫn nhau. Điều này làm tăng khả năng lựa chọn hệ thống chuyển giao cho MT. Hơn nữa, do đang tồn tại nhiều loại mạng công nghệ khác nhau, cùng với nhu cầu sử dụng dịch vụ và di chuyển ngày càng tăng, dẫn đến yêu cầu chuyển giao sẽ tăng. Đây là nguyên nhân cơ bản làm gia tăng tổng trễ trong suốt quá trình cung cấp dịch vụ, do vậy việc giảm tối đa trễ trong quá trình chuyển giao cần được đặc biệt quan tâm.

Xây dựng mô hình di chuyển để khảo sát hiệu suất của các giao thức quản lý di động hiện tại cho các loại ứng dụng khác nhau, giúp cho việc quản lý di chuyển trong hệ thống BcN được hiệu quả. Các giao thức quản lý di động thích ứng với ứng dụng sẽ được phát triển dựa trên các kết quả của những phân tích này.

Để duy trì và đảm bảo chất lượng dịch vụ cho các kết nối di động trong mạng BcN, tác giả tập trung nghiên cứu các cơ chế dự báo và đăng ký trước băng thông, hỗ trợ cho việc điều khiển chuyển giao cho kết nối hiện hữu và điều khiển đăng nhập cuộc gọi cho kết nối mới.

Do đó, luận án tập trung nghiên cứu ba vấn đề sau:

**1.5.1 Phân tích hiệu suất của các kỹ thuật chuyển giao dựa trên MIP, TCP-M, và SIP:** Các giao thức quản lý chuyển giao thực hiện trên các lớp mạng khác nhau (Liên kết, mạng, giao vận và ứng dụng) đều hỗ trợ cho các thể hệ thông tin di động. Hiệu suất xử lý chuyển giao của các giao thức này khi áp dụng cho các ứng dụng khác nhau là khác nhau. Để nghiên cứu kỹ hơn ảnh hưởng của việc chuyển giao, các ứng dụng di động được phân loại thành năm dạng, từ

A đến E, dựa trên các yêu cầu và đặc tính quản lý di chuyển riêng của chúng. Sau đó phát triển mô hình phân tích để khảo sát hiệu suất của các giao thức quản lý di chuyển hiện tại dùng cho những dạng ứng dụng này. Qua đó ta thấy đối với ứng dụng của một lớp cụ thể, khi sử dụng các giao thức khác nhau thì hiệu suất chuyển giao cũng khác nhau.

**1.5.2 Dự báo và định trước băng thông:** Mục đích của phần này là duy trì kết nối và đảm bảo QoS cho các ứng dụng di động. Trong nghiên cứu này, chúng tôi xây dựng cơ chế dự báo xác suất chuyển giao dựa trên thông tin di chuyển của người dùng được cập nhật định kỳ, từ đó đưa ra quyết định đăng ký trước băng thông cho kết nối.

**1.5.3 Quản lý di động thích ứng cho các ứng dụng trong mạng BcN:** các ứng dụng khác nhau thì có các yêu cầu khác nhau về tính di động, tuy nhiên hiện tại chưa có giao thức quản lý di động nào có khả năng hỗ trợ hiệu quả việc quản lý chuyển giao cho từng loại của ứng dụng. Cơ chế quản lý di động thích ứng (AMMS) được đề xuất để sử dụng giao thức quản lý di động phù hợp với yêu cầu chuyển giao của mỗi loại ứng dụng cụ thể. Để nâng cao hiệu suất chuyển giao, AMMS sử dụng các thông tin từ các lớp khác nhau để nhận biết được sự tương tác giữa chúng trong quá trình chuyển giao. Do đó, loại trừ được ảnh hưởng bất lợi của các lớp khác khi thực hiện xử lý chuyển giao trên một lớp cụ thể. Đầu tiên, nguyên lý hoạt động cơ bản của AMMS được phát triển dựa trên thiết kế kiến trúc của nó. Tiếp đến, xây dựng mô hình phân tích để khảo sát hiệu suất chuyển giao của AMMS. Cuối cùng, các giả lập sẽ được đưa ra thông qua việc sử dụng mô hình phân tích để đánh giá hiệu suất chuyển giao của AMMS cho các lớp khác nhau của ứng dụng.

## **Chương 2: Xây dựng tập tham số phân tích và đánh giá hiệu suất của các kĩ thuật chuyển giao dựa trên MIP, TCP-M, và SIP.**

### **2.1 Mở đầu**

Trong chương này luận án tập trung phân tích hiệu suất chuyển giao dựa trên các giao thức quản lý di động ở các lớp liên kết, lớp mạng, lớp giao vận và lớp ứng dụng. Những đóng góp cơ bản của chương này đó là:

- Xác định các tham số cơ bản có ảnh hưởng đến chất lượng dịch vụ, gồm *xác suất gói tin bị thất lạc; trễ báo hiệu và truyền tin; thời gian suy giảm thông lượng; bảo mật; tiêu hao nguồn điện*. Luận án cũng đưa ra và chứng minh định lý “*Giao thức kết nối vô tuyến quyết định chất lượng của hệ thống*”, và “*Độ trễ trung bình của truyền gói tin báo hiệu một chiều giữa MH và HA bằng tổng có trọng số của độ trễ truyền gói tin với trọng số là xác suất truyền thành công*”.
- Đặc điểm của các lớp ứng dụng A, B, C, D, E được nêu tại [50], trong đó ứng dụng dạng A là ứng dụng TCP hoặc UDP, khởi nguồn tại MN, có thời gian tồn tại ngắn, nên Luận án bỏ qua hỗ trợ chuyển giao cho dạng ứng dụng này.
- Sử dụng các tham số cơ bản nêu trên để xây dựng nên phương thức đánh giá hiệu suất chuyển giao của các ứng dụng dạng B (các ứng dụng TCP mà có thời gian tồn tại dài được khởi nguồn từ MN như truy cập web hay các phiên telnet), C (các ứng dụng TCP mà tồn tại lâu và được kết thúc bởi MN như là các phiên telnet), D (các ứng dụng UDP tồn tại lâu và được khởi nguồn bởi MN như là dịch vụ thoại và MN là người gọi) và E (các ứng dụng UDP tồn tại lâu và được khởi nguồn bởi MN như là dịch vụ thoại và MN là người được gọi).
- Đề xuất giao thức lựa chọn tuyến chuyển giao, cân bằng 2 tiêu chí tiêu thụ năng lượng (Pin MT) và hiệu suất mạng.

Các kết quả của chương này sẽ là tiền đề để xây dựng cơ chế *dự báo chuyển giao* và *đặt trước băng thông* cho các ứng dụng chuẩn bị chuyển giao ở chương tiếp theo.

### **2.2 Phân tích hiệu suất chuyển giao dựa trên các giao thức quản lý di động hiện tại.**

Ảnh hưởng chuyển giao trên các dạng ứng dụng này có thể cụ thể hóa bằng các tham số:

- *Độ trễ chuyển giao*: đây là tiêu chí mấu chốt trong KPI đánh giá chất lượng dịch vụ [43], là khoảng thời gian từ lúc bắt đầu cho đến khi hoàn thành việc chuyển giao. Các ứng dụng thời gian thực sử dụng giao thức giao vận thời gian thực (RTP) trên UDP, như dịch vụ thoại qua

Internet, và các ứng dụng đa phương tiện thuộc về dạng D và dạng E yêu cầu độ trễ chuyển giao tối thiểu.

- *Mất gói dữ liệu khi chuyển giao*: những ứng dụng dạng D và dạng E chạy trên UDP. Do UDP không phải là giao thức đáng tin cậy cho việc truyền dữ liệu trên mạng, các gói tin bị thất lạc trong quá trình chuyển giao sẽ không thể lấy lại được. Do đó, các ứng dụng dạng D và E sẽ mất gói dữ liệu trong quá trình chuyển giao. Các ứng dụng dạng B và C có thể chạy trên TCP. Do giao thức TCP là giao thức tin cậy, các gói tin bị thất lạc trong quá trình chuyển giao có thể được truyền lại qua phương thức truyền lại gói tin của TCP. Cho nên những ứng dụng của hai dạng này sẽ không bị mất gói tin trong quá trình chuyển giao.

- *Thời gian suy giảm thông lượng*: đối với các ứng dụng dạng B và C sử dụng TCP để truyền tin, các gói tin bị thất lạc trong quá trình chuyển giao là nguyên nhân kích hoạt cơ chế slow-start của TCP làm giảm thông lượng. Thời gian suy giảm thông lượng phải được giữ ở mức tối thiểu.

- *Độ trễ điểm đến điểm*: một số các ứng dụng yêu cầu các host đang kết nối phải trao đổi các gói tin trực tiếp mà không thông qua bất cứ sự can thiệp nào của các phần tử khác trong mạng. Những ứng dụng này yêu cầu độ trễ điểm đến điểm ở mức tối thiểu. Độ trễ dạng này sẽ tăng lên trong trường hợp giao thức quản lý di chuyển định tuyến lại các gói tin như trong MIP, khi đó độ trễ điểm tới điểm sẽ tăng lên. Các ứng dụng dạng D và dạng E là các ứng dụng thời gian thực do đó chúng yêu cầu độ trễ điểm tới điểm thấp.

- *Sự trong suốt ở lớp vận chuyển*: các ứng dụng (dạng B và C) chạy trên TCP yêu cầu phải có phương pháp khôi phục lại kết nối bị ngắt trong quá trình chuyển giao mà không ảnh hưởng tới QoS. Do đó, các giao thức quản lý di chuyển sẽ ẩn sự thay đổi địa chỉ IP của mobile host dựa trên việc chuyển giao, giống như MIP và TCP-M.

- *Bảo mật*: một ứng dụng có thể có nhiều mức độ khác nhau về bảo mật trong các môi trường mạng khác nhau. Ví dụ, trong khi truyền thông bên trong một home network, ứng dụng sẽ không yêu cầu cao về phương pháp bảo mật. Nhưng trong trường hợp khác, khi đang ở một foreign domain hay đang truyền thông với CN ở foreign domain, ứng dụng này sẽ yêu cầu cao về bảo mật cho dù nó là một ứng dụng tương tự như ở ví dụ trên. Do đó, bảo mật là rất quan trọng cho tất cả các dạng của ứng dụng.

- *Tiêu hao năng lượng*: Các thông tin về mức độ tiêu thụ nguồn điện được mạng định kỳ cập nhật cho MT [20].

Các phân tích phía trên cho thấy rằng các dạng khác nhau của ứng dụng có thể có các mong muốn và yêu cầu khác nhau từ một giao thức quản lý di chuyển. Ở phần tiếp theo, chúng tôi sẽ thảo luận về những giao thức quản lý di chuyển và đưa các đánh giá hiệu suất định tính liên quan đến những tiêu chí hiệu suất chuyển giao ở trên.

Các giao thức quản lý di chuyển được thực thi trên các lớp khác nhau như lớp liên kết [1][49], lớp mạng [58], lớp giao vận [74], và lớp ứng dụng [86] đã được đề xuất trong nhiều nghiên cứu cách đây nhiều năm. Luận án sẽ xem xét, đánh giá các giao thức này dựa trên các tham số hiệu suất chuyển giao trình bày ở trên.

### **2.2.1 Các giao thức quản lý di động ở lớp liên kết (Lớp 2)**

Các giao thức quản lý di chuyển lớp kết nối tập trung vào các vấn đề liên quan tới chuyển giao giữa các mạng truy cập không đồng nhất và các kỹ thuật quản lý mạng khác nhau [2]. Khi một MN dịch chuyển từ một mạng truy cập vô tuyến này đến một mạng khác nơi có hỗ trợ giao diện vô tuyến và phần ứng dụng di động (MAP) giống với mạng cũ, các dịch vụ sẽ được cung cấp một cách liền mạch. Tuy nhiên, khi MAP ở hai hệ thống khác nhau, các phần tử bổ sung và lưu lượng báo hiệu sẽ được yêu cầu trong quá trình MN chuyển giao giữa hai hệ thống [2]. Thông tin di chuyển của người sử dụng (UMP) sẽ được sử dụng trong [1] để hỗ trợ cho việc nâng cao khả năng quản lý di chuyển. Khái niệm về các tế bào ranh giới trong cùng mạng được sử dụng trong [49] để chuẩn bị trước cho người sử dụng một tiến trình chuyển giao trong mạng khả thi nhất, làm giảm đáng kể nguy cơ lỗi chuyển giao trong mạng. Hiệu suất làm việc của các giao thức di chuyển lớp kết nối được tóm tắt như sau:

- Độ trễ chuyển giao trong cùng mạng cao vì một số chức năng như là cấu hình lại và thông dịch địa chỉ, lấy thông tin di chuyển (quá khứ) của người dùng, truyền bản tin điều khiển và tạo kết nối, các thông tin di chuyển liên quan đến việc lưu lại vết di chuyển trong cùng mạng, dàn xếp QoS, và xác thực giữa các hệ thống được lưu ý trong quá trình chuyển giao trong cùng mạng [2].
- Độ trễ trong quá trình chuyển giao càng cao sẽ càng tăng số lượng gói tin bị thất lạc.
- Sau khi quá trình chuyển giao trong mạng hoàn tất, một MN kết nối đến một hệ thống mới mà không cần các yêu cầu nào đối với việc chuyển hướng. Do đó, yêu cầu về độ trễ điểm đến điểm của ứng dụng là được bảo đảm.
- Do một MN kết nối đến một địa chỉ mới trong hệ thống mới, một kết nối lớp vận chuyển phải được thiết lập lại sau khi chuyển giao nội mạng hoàn tất. Các giao thức quản lý di chuyển mạng lớp kết nối sẽ không tương thích với các ứng dụng TCP và UDP.
- Do việc xác thực được thực hiện trong quá trình chuyển giao trong mạng, việc chuyển giao này sẽ được bảo mật.

### **2.2.2 Các giao thức quản lý di động ở lớp mạng (Lớp 3)**

MIP [58] được đề xuất để hỗ trợ di chuyển trong các mạng IP. MIP chuyển tiếp các gói tin đến người sử dụng di động ở ngoài mạng chủ của họ bằng việc sử dụng kỹ thuật kênh IP-IP [58].

Kỹ thuật này chèn lượng đáng kể thông tin vào đầu vào các gói dữ liệu. Hơn nữa, MIP bị ảnh hưởng bởi các vấn đề về định tuyến vòng (triangular), tổng tải báo hiệu và độ trễ chuyển giao cao [2]. Hiệu suất của giao thức MIP được tóm tắt như sau:

- Khi đăng kí, MIP sử dụng một lượng trễ lớn trong quá trình chuyển giao. MIP phân cấp [25] và những giao thức micro-mobility như Cellular IP [80], IDMP [51], và HAWAII [63] giảm độ trễ chuyển giao bằng việc đưa những lớp phân cấp khác vào cấu trúc MIP cơ bản để dồn các bản tin báo hiệu vào một miền xác định.
- Độ trễ lớn của MIP gây ra một lượng đáng kể số gói tin bị thất lạc trong quá trình chuyển giao.
- Định tuyến vòng của MIP gây nên đường không đối xứng giữa một CN và một MN. Độ trễ tăng thêm do việc định hướng lại đường từ CN đến MN phía Trạm chủ (HA). Theo tính toán trong [90] cho thấy rằng độ trễ điểm đến điểm khi dùng MIP tăng thêm 45% trong một khu vực (từ CN đến MN), độ trễ này sẽ tăng lên lớn hơn nhiều lần trong mạng WAN. Điều này là khó có thể chấp nhận được đối với những ứng dụng nhạy cảm về độ trễ [86].
- Thông qua việc định hướng lại gói tin trong quá trình chuyển giao, MIP không làm thay đổi địa chỉ IP từ các ứng dụng. Do đó, chuyển giao MIP là trong suốt đến các ứng dụng và các kết nối lớp vận chuyển được giữ nguyên vẹn trong quá trình chuyển giao.
- Việc xác thực của việc đăng kí trong MIP được thực hiện như là một phần của việc đăng kí MIP [11]. Do đó, chuyển giao MIP được bảo mật.

### **2.2.3 Các giao thức quản lý di động ở lớp giao vận (Lớp 4)**

Với việc sử dụng phương thức quản lý di động tại lớp này, một đầu TCP có thể tạm dừng một kết nối mở và kích hoạt lại kết nối đó từ một địa chỉ IP khác. Việc phục hồi lại kết nối TCP được thực hiện giống như các ứng dụng có thể tiếp tục sử dụng kết nối TCP trước đó thông qua chuyển giao [74]. Các giao thức quản lý di động ở lớp vận chuyển được đưa ra để hỗ trợ cho việc di chuyển giữa các mạng mà không cần phải thay đổi các mạng riêng lẻ. Do đó, việc không cần thay đổi công nghệ mạng mà vẫn đảm bảo khả năng chuyển giao giữa các mạng, giúp cho việc duy trì tính tự nhiên phi trạng thái của Internet và các mạng IP khác. Tuy nhiên, cần phải sửa lại lớp vận chuyển, và các ứng dụng hiện tại để đảm bảo tính tương thích với nhau. Việc này chắc chắn sẽ rất tốn kém, vì liên quan tới rất nhiều các ứng dụng đã phát triển từ trước và đang hoạt động trên mạng Internet. Một kiến trúc MSOCKS đã được đề xuất ở [47][48] để xử lý việc quản lý di động tại lớp giao vận. Phương thức kết cuối quản lý di động thông qua các lớp đã được đề xuất ở [28]. [74] đề xuất TCP-M để hỗ trợ cho việc quản lý di động lớp giao vận. Hơn nữa, vấn đề này hiện đang được thảo luận tại IETF để sửa SCTP,

cho phép nó có thể thay đổi các địa chỉ đích một cách linh động ở giữa một kết nối [13][14]. Hiệu suất của các giao thức quản lý di động tại lớp giao vận được tóm tắt như sau:

- Do chỉ có các điểm cuối truyền thông với nhau tham gia vào quá trình chuyển giao, nên độ trễ thường thấp hơn so với MIP [74]. Cần lưu ý là việc sử dụng các thành phần thứ ba như HA của MIP sẽ làm tăng độ trễ trong quá trình xử lý chuyển giao.
- Trong quản lý di động ở lớp vận chuyển, một kết nối TCP duy trì chung khối quản lý và trạng thái, bao gồm dải số tuần tự [74]. Do đó, việc truyền lại số liệu cần phải tuân theo chuẩn. Các gói tin đã bị thất lạc trong quá trình chuyển giao có thể được phục hồi. Các giao thức quản lý di động ở lớp vận chuyển có thể được thiết kế để xác thực là không có một gói tin nào bị thất lạc trong quá trình chuyển giao.
- Do không định hướng lại gói tin, nên tuyến giữa các trạm chủ đang kết nối (ví dụ CN và MN) là đối xứng. Do đó, độ trễ điểm đến điểm không bị tăng sau khi chuyển giao. Việc này trái ngược với việc xử lý chuyển giao của MIP ở lớp mạng, do định tuyến vòng nên làm tăng độ trễ điểm đến điểm từ CN đến MN khi MN cách xa mạng chủ của nó.
- Khi một kết nối lớp vận chuyển được khởi động lại, các ứng dụng không bị ảnh hưởng bởi việc di chuyển.
- Việc xác thực tuyệt đối được tích hợp vào quá trình chuyển giao lớp vận chuyển, điều này làm tăng cao tính bảo mật. Các phương pháp quản lý di động kết cuối làm đơn giản hoá yêu cầu bảo mật so với các phương pháp ở lớp mạng như MIP [74].

#### **2.2.4 Các giao thức quản lý di động ở lớp ứng dụng (lớp 5)**

Quản lý di động lớp ứng dụng sử dụng giao thức SIP được đề xuất ở [86]. Giao thức SIP dựa trên tính di động không yêu cầu bất cứ sự thay đổi nào đến các lớp IP của đầu cuối di động. Thêm nữa, tính di động và vùng dịch vụ do SIP hỗ trợ. Hiệu suất của giao thức di động SIP được tóm tắt như sau:

- Do các trạm định hướng lại như SIP proxy và SIP server được sử dụng trong quá trình chuyển giao, độ trễ chuyển giao của SIP có thể tương đương với MIP nhưng lại cao hơn so với các giao thức quản lý di động ở lớp vận chuyển.
- Khi thực hiện thủ tục báo hiệu chuyển giao, số lượng bản tin bị thất lạc có thể ngang với phương thức chuyển giao sử dụng MIP.
- Khi quá trình báo hiệu chuyển giao kết thúc, các trạm chủ đang trao đổi với nhau, như là CN và MN, sẽ trao đổi trực tiếp mà không cần trạm định hướng lại. Do đó, độ trễ điểm đến điểm sẽ không bị tăng khi một MN cách xa mạng chủ của nó.



- SIP không thể hỗ trợ cho các kết nối TCP [86] . Do đó, di động SIP sẽ không trong suốt với giao thức TCP.

- Các bản tin báo hiệu được sử dụng quản lý di động SIP được bảo mật bằng việc sử dụng các phương pháp bảo mật khác nhau. Do đó, SIP dựa trên quản lí di động là được bảo mật.

Các giao thức quản lý di động ở các lớp khác nhau trong TCP/IP được tóm tắt trong bảng 2.1, trong đó không có một giao thức quản lí di động nào có thể hỗ trợ cho việc quản lý di chuyển trong suốt đến các loại ứng dụng khác nhau. Do không có khả năng để hỗ trợ cho tất cả các loại ứng dụng sử dụng một giao thức cụ thể trong việc quản lý di động, việc sử dụng một cấu trúc quản lý di động được đề cập trong đề tài này để lựa chọn linh hoạt một giao thức quản lý di động dựa trên các yêu cầu của ứng dụng cụ thể. Để xác định rõ sự phù hợp cho một lớp ứng dụng cụ thể, cần nghiên cứu và nắm rõ hiệu suất chuyển giao của các giao thức quản lý di động sử dụng cho các lớp ứng dụng khác nhau.

Bảng 2.1: Hiệu suất quản lý di động của các giao thức

T.Số Hiệu suất	Lớp 2	Lớp 3	Lớp 4	Lớp 5
Trễ HO	Rất kém	Rất kém	Kém	Rất kém
Mất số liệu HO	Rất kém	Rất kém	Kém	Rất kém
Trễ kết cuối	Tốt	Kém	Tốt	Tốt
Tính trong suốt lớp vận chuyển	Kém	Tốt	Tốt	Kém
Bảo mật	Tốt	Tốt	Tốt	Tốt

Trong phần này, các mô hình phân tích được đề cập để khảo sát hiệu suất chuyển giao đối với các giao thức quản lý chuyển giao đã có theo các ứng dụng loại B, C, D, và E. Như đã đề cập ở trên, ứng dụng loại A không yêu cầu hỗ trợ về mặt di động. Dựa trên các kết quả phân tích toán học, có khả năng lựa chọn giao thức quản lý di động phù hợp cho một loại ứng dụng cụ thể. Hơn nữa, các phân tích này còn cung cấp một cái nhìn chi tiết về các tham số ảnh hưởng đến hiệu suất chuyển giao của giao thức.

### 2.3 Xác định các tham số cơ bản của mô hình phân tích

Để xây dựng mô hình phân tích hiệu suất của các giao thức quản lý di động đã có, Trạm chủ di động (MH) được xem xét đang ở ngoài mạng chủ (HM) của nó và di chuyển từ mạng cũ (ON) sang một mạng mới (NN) trong khi đang kết nối với một trạm trung gian (CH). Các phần tử mạng hỗ trợ cho HM trong việc quản lý di chuyển như máy chủ SIP [65], máy chủ tên miền (DNS), và trạm gốc (HA) được đặt ở HN.

### 2.3.1 Xác suất gói tin bị thất lạc từ điểm đến điểm

Do UDP không phải là giao thức đáng tin cậy cho việc truyền dữ liệu trên mạng, các gói tin bị thất lạc trong quá trình chuyển giao sẽ không thể lấy lại được. Do đó, các ứng dụng dạng D và E sẽ mất gói dữ liệu trong quá trình chuyển giao. Các ứng dụng dạng B và C có thể chạy trên TCP. Do giao thức TCP là giao thức tin cậy, các gói tin bị thất lạc trong quá trình chuyển giao có thể được truyền lại qua phương thức truyền lại gói tin của TCP. Cho nên những ứng dụng của hai dạng này sẽ không bị mất gói tin trong quá trình chuyển giao.

Trong mạng BcN, một kết nối hoàn chỉnh được thực hiện thông qua nhiều chặng, với các môi trường vật lý khác nhau, có thể bao gồm cả hữu tuyến và vô tuyến. Do vậy chất lượng của dịch vụ sẽ phụ thuộc vào chất lượng kết nối toàn tuyến, và các giao thức được sử dụng trên các kết nối đó.

*Định lý 2.1:* Giao thức kết nối vô tuyến quyết định chất lượng của hệ thống, với xác suất thất lạc gói tin điểm tới điểm giữa MH và HA được xác định như sau:

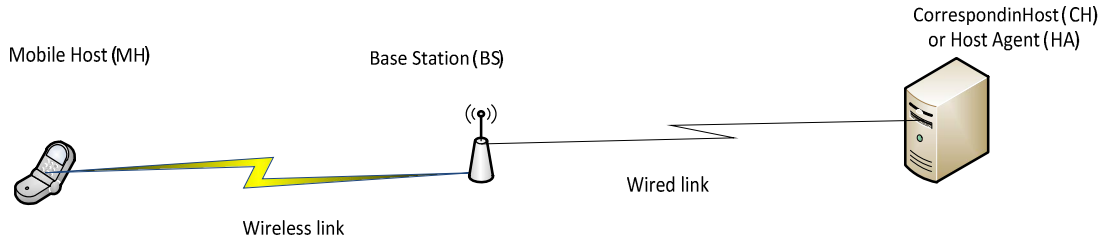
$$p_r = 1 - \left[ 1 - p_f \left( (2 - p_f) p_f \right)^{\frac{(n^2+n)}{2}} \right]^K (1 - p_c)$$

*Chứng minh:* Để tìm được xác suất gói tin bị thất lạc từ điểm đến điểm giữa MH và HA (hay CH), đường dẫn từ HM đến HA (hay CH) sẽ được chia ra hai phần: kết nối vô tuyến từ MH đến BS, và kết nối hữu tuyến từ BS đến HA (hay CH). Gọi  $p_w$  và  $p_c$  lần lượt là xác suất lỗi gói của kết nối vô tuyến giữa MH và BS và của kết nối hữu tuyến giữa BS và HA. Theo định luật tổng xác suất, xác suất lỗi gói toàn bộ của hệ thống là:

$$p = 1 - (1 - p_w)(1 - p_c) \quad (2.1)$$

Trong thực tế, kết nối vô tuyến thường chịu ảnh hưởng của các hiệu ứng fading, shadowing và thường được mô hình hóa bởi phân bố Rayleigh trong khi kết nối hữu tuyến được mô hình hóa bởi kênh nhiễu trắng (AWGN), do đó ta có  $p_w \ll p_c$ , dẫn đến:

$$\begin{aligned} p &= 1 - (1 - p_w)(1 - p_c) \\ &= 1 - (1 - p_w - p_c + p_w p_c) \\ &= p_w + p_c + p_w p_c \\ &\approx p_w \end{aligned}$$



Hình 2.1 : Các kết nối Vô tuyến và Hữu tuyến giữa CH/HA, BS và MH

Tiếp theo, để xác định biểu thức  $p$  trong trường hợp không có giao thức kết nối vô tuyến (RLP) và có kết nối vô tuyến RLP. Đặt  $L_p$  và  $L_f$  là độ dài của gói tin (thông thường là gói tin IP) và độ dài của frame ở lớp kết nối. Số lượng frame trên gói tin là  $K = \left\lceil \frac{L_p}{L_f} \right\rceil$ . Trong trường hợp không sử dụng giao thức RLP, xác suất gói tin bị thất lạc trên kết nối vô tuyến sẽ là:

$$p_{wr} = 1 - (1 - p_f)^K$$

ở đây  $p_f$  là tỉ lệ các frame lớp kết nối bị lỗi. Do đó, xác suất gói tin bị thất lạc điếm tới điếm  $p_{nr}$  giữa MH và HA (hay CH) trong trường hợp không sử dụng RLP có thể tính được bằng  $p = p_{nr}$  và  $p_w = p_{wr}$  trong công thức (2.1). Do đó,  $p_{nr}$  sẽ là:

$$p_{nr} = 1 - (1 - p_f)^K (1 - p_c) \quad (2.2)$$

Khi RLP được sử dụng, xác suất gói tin bị thất lạc trên kết nối vô tuyến  $p_{wr}$  được tính như sau:

$$p_{wr} = 1 - \left[ 1 - p_f \left( (2 - p_f) p_f \right)^{\frac{(n^2+n)}{2}} \right]^K \quad (2.3)$$

Ở đây,  $n$  là số lần tối đa truyền lại frame bị lỗi qua lớp kết nối. Thông thường,  $n=3$  cho RLP [8].

Thay  $p_w = p_{wr}$  và  $p = p_r$  vào (2.1), ta có xác suất gói tin bị thất lạc điếm đến điếm giữa MH và HA (hay CH)  $p_r$  sẽ là:

$$p_r = 1 - \left[ 1 - p_f \left( (2 - p_f) p_f \right)^{\frac{(n^2+n)}{2}} \right]^K (1 - p_c) \quad (2.4)$$

Ở đây,  $p_f$  là FER ở lớp kết nối và  $K$  là số lượng của frame lớp vật lý trên số gói tin.

Ta dễ dàng thấy được  $p_r < p_{nr}$ , do vậy xác suất mất gói tin khi sử dụng giao thức kết nối vô tuyến RLP bé hơn khi không sử dụng RLP, nghĩa là chất lượng hệ thống (QoS) khi sử dụng RLP sẽ tốt hơn.

### 2.3.2 Độ trễ truyền bản tin điểm tới điểm

Độ trễ (tính bằng giây (s)) khi truyền bản tin điểm tới điểm từ MH đến HA (hay CH) được tính bằng tổng của toàn bộ độ trễ khi truyền của gói tin trên kết nối vô tuyến từ MH đến BS và độ trễ truyền gói tin trên kết nối hữu tuyến giữa BS và HA (hay CH). Khi không sử dụng RLP, sẽ không phải truyền lại số liệu trên lớp kết nối. Do đó, độ trễ khi truyền gói tin từ điểm đến điểm,  $T_{nr}$ , giữa MH và HA (hay CH) được xác định như sau:

$$T_{nr} = D + t_w \quad (2.5)$$

Ở đây,  $D$  là độ trễ truy cập lớp kết nối và  $t_w$  là độ trễ trên kết nối hữu tuyến giữa BS và HA (hay CH).

Độ trễ khi truyền frame một chiều  $T_f$  giữa MH và BS với RLP được tính bằng [8]:

$$T_f = D(1 - p_f) + \sum_{i=1}^n \sum_{j=1}^i P(C_{i,j})(2iD + 2(j-1)t) \quad (2.6)$$

Ở đây,  $p_f$  là FER lớp kết nối và  $t$  là khoảng thời gian trao đổi giữa các frame ở lớp kết nối, thông thường là 20ms [8].  $P(C_{i,j})$  là xác suất mà frame đầu tiên nhận được chính xác ở BS từ MH, là frame thứ  $i$  trong lần truyền lại thứ  $j$ . Công thức cho  $P(C_{i,j})$  được tính như sau [8]:

$$P(C_{i,j}) = p_f(1 - p_f)^2 \left( (2 - p_f)p_f \right)^{\frac{i^2-i}{2} + j - 1} \quad \text{trong đó } i = 1, 2, \dots, n \text{ và } j = 1, 2, \dots, i \quad (2.7)$$

Do đó, khi RLP được sử dụng, độ trễ khi truyền gói tin từ điểm đến điểm,  $T_r$ , giữa MH và HA (hay CH) sẽ là:

$$T_r = T_f + (K-1)t + t_w \quad (2.8)$$

Ở đây,  $K$  là số lượng các frame lớp kết nối trên gói tin đã định nghĩa ở phần 2.3.1

### 2.3.3 Trung bình độ trễ truyền gói tin báo hiệu khi sử dụng giao thức UDP

*Định lý 2.2:* Độ trễ trung bình của việc truyền gói tin báo hiệu một chiều giữa MH và HA bằng tổng có trọng số của các độ trễ truyền gói tin với trọng số là xác suất truyền thành công.

$$D_p = (1 - q) \left\{ B + A \sum_{i=2}^m q^{i-1} (\gamma^{i-1} - 1) + \sum_{i=m+1}^{\infty} q^{i-1} \left[ A(\gamma^{m-1} - 1) + (i - m)\gamma^{m-2}\Delta \right] \right\} \quad \text{voi } A = \frac{\Delta}{\gamma - 1}$$

*Chứng minh:*

Gọi  $p_i$  là xác suất truyền thành công sau lần thử thứ  $i$ , ta có:

$$\sum_i p_i = 1.$$

Gọi  $T_i$  là độ trễ tương ứng và áp dụng định lý tổng xác suất với giả sử rằng MH sẽ là:

$$D_p = \sum_i^{\infty} p_i T_i \quad (2.9)$$

Khi  $p_i$  là xác suất truyền thành công sau lần truyền lại thứ  $i$ , có nghĩa là  $i-1$  lần truyền trước bị thất bại. Trong điều kiện các lần truyền là độc lập nhau,  $p_i$  được biểu diễn như sau:

$$\begin{aligned} p_i &= \underbrace{qq \dots q}_{i-1} (1-q) \\ &= q^{i-1} (1-q) \end{aligned} \quad (2.10)$$

ở đây  $p$  là xác suất mất gói (truyền thất bại) giữa MH và CH. Do kết nối giữa MH và CH bao gồm 2 kết nối vô tuyến và hữu tuyến nối tiếp nhau, ta có

$$p = 1 - (1 - p_w)(1 - p_c)$$

Hai trường hợp sau được xem xét:

- Có sử dụng giao thức kết nối vô tuyến (radio link protocol)
- Không sử dụng giao thức kết nối vô tuyến

Lý do của việc sử dụng giao thức kết nối vô tuyến là giao thức TCP/IP được thiết kế để sử dụng cho kênh hữu tuyến mà thông thường tỷ lệ lỗi frame (frame error rate) là rất nhỏ. Khi áp dụng trực tiếp giao thức TCP/IP vào kênh vô tuyến, do các hiệu ứng fading, tỷ lệ lỗi khung của kênh vô tuyến trở nên rất lớn. Do đó, để đảm bảo chất lượng dịch vụ đầu cuối (FER nhỏ), giao thức kết nối vô tuyến thường được sử dụng để giảm tỷ lệ lỗi bit trước khi chuyển các khung dữ liệu cho lớp TCP/IP. Ở đây, cho đơn giản, ta có thể hiểu việc sử dụng giao thức kết nối vô tuyến như việc sử dụng  $N$ -go back ARQ.

Frame #1	Frame #2	...	Frame #k
----------	----------	-----	----------

Hình 2.2: Cấu trúc gói số liệu

#### a. Không sử dụng giao thức kết nối vô tuyến

Giả sử rằng trong một gói dữ liệu chứa  $K$  khung dữ liệu, và một gói dữ liệu được xem là nhận sai khi chỉ cần chứa một khung sai. Hay nói cách khác, một gói dữ liệu được xem là nhận đúng nếu nó chứa  $K$  khung dữ liệu đúng. Tỷ lệ lỗi gói cho kênh vô tuyến có thể được biểu diễn dưới dạng tỷ lệ lỗi khung như sau:

$$p_w = 1 - (1 - p_f)^K$$

với  $p_f$  là tỷ lệ lỗi khung (Frame error rate).

Khi không có truyền lại (do không sử dụng giao thức kết nối vô tuyến), thì độ trễ khi truyền một gói tin từ MH đến CH chỉ đơn thuần là tổng độ trễ tiêu tốn ở lớp kết nối ở kênh vô tuyến ( $D$ ) và độ trễ trên kết nối hữu tuyến ( $T_w$ )

$$B = D + T_w$$

b. Có sử dụng giao thức kết nối vô tuyến

Khi có sử dụng kết nối vô tuyến, thì tỷ lệ lỗi gói trên kênh vô tuyến được tính theo tỷ lệ lỗi khung và số lần truyền lại  $n$  của kênh vô tuyến như sau [8]

$$p_w = 1 - \left[ 1 - p_f \left[ (2 - p_f) p_f \right]^{\frac{(n^2+n)}{2}} \right]^K$$

Độ trễ khi truyền gói tin,  $T_i$ , giữa MH và CH sẽ là:

$$B = T_f + (K - 1)\tau + t_w$$

Trong công thức trên  $\tau$  là khoảng thời gian giữa các frame ở lớp kết nối và  $T_f$  là độ trễ khi truyền khung một chiều giữa MH và BS với RLP được tính bằng [8]:

$$T_f = D(1 - p_f) + \sum_{i=1}^n \sum_{j=1}^i P(C_{i,j}) (2iD + 2(j-1)t)$$

Gọi  $m$  là số lần truyền lại tối đa mà hệ thống được phép,  $T_i$  trong cả 2 trường hợp sử dụng và không sử dụng giao thức kết nối vô tuyến là:

$$T_i = \begin{cases} \Delta + \gamma\Delta + \gamma^2\Delta + \dots + \gamma^{i-2}\Delta + B, & i \leq m \\ \Delta + \gamma\Delta + \gamma^2\Delta + \dots + \gamma^{m-2}\Delta + (i-m)\gamma^{m-2}\Delta + B, & i > m \end{cases} \quad (2.11)$$

Trong đó  $\Delta$  là giá trị khởi tạo của bộ đếm thời gian truyền lại, giá trị này vào khoảng gấp đôi thời gian truyền đi và lại giữa MH và HA (hay CH), và cộng thêm ít nhất 100ms để cho phép các xử lý ở MH và HA (hay CH).  $\gamma$  là hệ số để điều chỉnh khoảng thời gian giữa 2 lần truyền lại, thông thường  $\gamma$  được chọn bằng 2.

Thay  $p_i$  và  $T_i$  từ (2.10) và (2.11) vào (2.9) ta thu được độ trễ trung bình (tính bằng giây (s)):

$$\begin{aligned}
D_p &= \sum_i^{\infty} p_i T_i = p_1 T_1 + \sum_{i=2}^m p_i T_i + \sum_{i=m+1}^{\infty} p_i T_i \\
&= (1-q)B + \sum_{i=2}^m q^{i-1} (1-q) [\Delta + \gamma \Delta + \gamma^2 \Delta + \dots + \gamma^{i-2} \Delta + B] + \\
&\quad \sum_{i=m+1}^{\infty} q^{i-1} (1-q) [\Delta + \gamma \Delta + \gamma^2 \Delta + \dots + \gamma^{m-2} \Delta + (i-m) \gamma^{m-2} \Delta + B] \\
&= (1-q) \left\{ B + A \sum_{i=2}^m q^{i-1} (\gamma^{i-1} - 1) + \sum_{i=m+1}^{\infty} q^{i-1} [A(\gamma^{m-1} - 1) + (i-m) \gamma^{m-2} \Delta] \right\} \quad khi \quad A = \frac{\Delta}{\gamma - 1}
\end{aligned} \tag{2.12}$$

## 2.4 Xây dựng phương thức đánh giá hiệu suất chuyển giao của các ứng dụng dạng B và dạng C (MIP và TCP-M)

Vì các ứng dụng dạng B và dạng C sử dụng giao thức TCP, ở đây Luận án sẽ xem xét một kết nối TCP giữa một CH và MH để nghiên cứu về hiệu suất chuyển giao. Hiệu suất chuyển giao của các ứng dụng dạng B và dạng C đồng nghĩa với hiệu suất của một kết nối TCP. Với kịch bản khi MH đang ở mạng cũ (ON) và bắt đầu tải một tệp sử dụng FTP từ CH và di chuyển đến một mạng mới (NN) trong quá trình tải này. Giả sử kích cỡ của file đủ lớn để kết nối TCP có thể tiếp tục từ ON đến NN. Ngoài ra, giả sử thêm rằng ứng dụng FTP của CH tạo ra các gói tin liên tục và đủ nhanh để cửa sổ nghẽn cho phép. Hơn nữa, kích cỡ của window này do bên nhận (trong trường hợp này là MH) thông báo sẽ luôn luôn lớn hơn kích cỡ của cửa sổ nghẽn. Do đó, việc gửi kích cỡ của cửa sổ sẽ luôn luôn bị giới hạn bởi cửa sổ nghẽn. Giả sử rằng khi MH ở trong mạng ON, kết nối TCP giữa CH và MH sẽ hoạt động ở trạng thái ổn định. Trong trạng thái này, các thông số trạng thái TCP, như là kích cỡ của cửa sổ nghẽn và RTT được quyết định bởi đường giữa CH và MH. Để duy trì hiệu suất thông lượng cao nhất ở các mạng vô tuyến khác nhau đặc tả bởi  $p_f$  và  $D$  khác nhau và độ lưu thông của các nguồn TCP hữu tuyến cùng chia sẻ điểm nghẽn (bottleneck), lưu ý phương thức điều khiển nghẽn thích ứng được đề xuất ở [36] điều chỉnh các tham số AIMD là  $\alpha$  và  $\beta$  tùy theo điều kiện kết nối vô tuyến hiện tại. Theo [36]  $\alpha$  được xác định như sau:

$$\alpha = \frac{bp(1-\beta)}{2(1+\beta)} \left[ \hat{T} (2R + 3T_0 p (1 + 32 p^2) (1 + \beta)) \right] \tag{2.13}$$

Trong đó  $p$  là xác suất mất gói số liệu kết cuối,  $T$  là thông lượng đạt được của nguồn TCP hữu tuyến với  $p_c$  là xác suất mất gói số liệu do nghẽn mạng hữu tuyến, và  $R_c$  là RTT kết cuối trong mạng hữu tuyến.  $R$  là RTT kết cuối giữa CH và MH,  $T_0$  là ngưỡng thời gian truyền lại (RTO) đối với kết nối TCP, và  $b$  là số gói dữ liệu được ghi nhận bởi tín hiệu ACK đơn. Theo [36]  $\beta$  có thể là 0.75, 0.80 và 0.85 tương ứng đối với WLAN, 3G và mạng Vệ tinh.  $p$  là xác suất mất gói số liệu kết cuối giữa MH và CH.

Thông lượng của kết nối TCP với các tham số AIMD  $\alpha$  và  $\beta$  sẽ là [87]

$$T_{\alpha,\beta}(p, R, T_0, b) = \frac{1}{R\sqrt{\frac{2b(1-\beta)p}{\alpha(1+\beta)}} + T_0 \min\left(1, 3\sqrt{\frac{(1-\beta^2)bp}{2\alpha}}\right) p(1+32p^2)} \quad (2.14)$$

$\alpha$  và  $\beta$  tương ứng sẽ là 1 và  $\frac{1}{2}$ , từ (2.13) và (2.14) ta có:

$$\hat{T} = \frac{1}{R_c\sqrt{\frac{2bp_c}{3}} + T_{oc} \min\left(1, 3\sqrt{\frac{3bp_c}{8}}\right) p_c(1+32p_c^2)} \quad (2.15)$$

ở đây  $T_{oc}$  là RTO khi khởi động.

Kích thước cửa sổ nghẽn của TCP ở Trạng thái ổn định tùy thuộc vào xác suất mất gói số liệu kết của, và được tính toán như sau [87]:

$$E[W] = \frac{\alpha + b(1-\beta)}{2b(1-\beta^2)} + \sqrt{\left(\frac{\alpha + b(1-\beta)}{2b(1-\beta^2)}\right)^2 + \frac{2\alpha(1-p)}{bp(1-\beta^2)}} \quad (2.16)$$

Sử dụng (2.13) để xác định tham số tăng bổ sung của kết nối TCP và (2.16) để xác định kích cỡ cửa sổ nghẽn ở trạng thái ổn định khi MH đang ở mạng ON và NN. Khi MH đang ở ON, giả sử kết nối TCP đang hoạt động ở trạng thái ổn định tương ứng với tỷ lệ lỗi khung lớp liên kết ( $p_f$ ) và trễ kết cuối lớp vận chuyển (D) của mạng ON. Sau khi MH chuyển giao sang NN, thì kết nối TCP sẽ đạt được trạng thái ổn định tương ứng với NN ngay khi có thể được. Lý tưởng là, TCP sẽ hoạt động ở trạng thái ổn định ngay sau khi hoàn thành chuyển giao.

*Định nghĩa 2.1: Thời gian giảm thông lượng là thời gian cần thiết để kết nối TCP khôi phục lại hoạt động ở trạng thái ổn định sau khi chuyển giao.*

Hơn nữa, giao thức quản lý chuyển giao lý tưởng cần đảm bảo để các ứng dụng chạy trên TCP ở MH không cảm nhận được trễ chuyển giao khi MH chuyển từ mạng ON sang NN.

*Định nghĩa 2.2: Trễ chuyển giao là thời gian kể từ khi MH nhận gói số liệu cuối cùng ở ON đến khi nó nhận gói số liệu đầu tiên ở NN.*

TCP là giao thức hoạt động ổn định, số liệu không bị mất khi chuyển giao nhờ cơ chế phát lại. Do vậy, hiệu suất chuyển giao của kết nối TCP có thể được xác định thông qua 2 tham số “**Thời gian giảm thông lượng**” và “**Trễ chuyển giao**”. Tiếp theo, luận án nghiên cứu hiệu suất chuyển giao của kết nối TCP khi sử dụng MIP [58] và TCP-M.

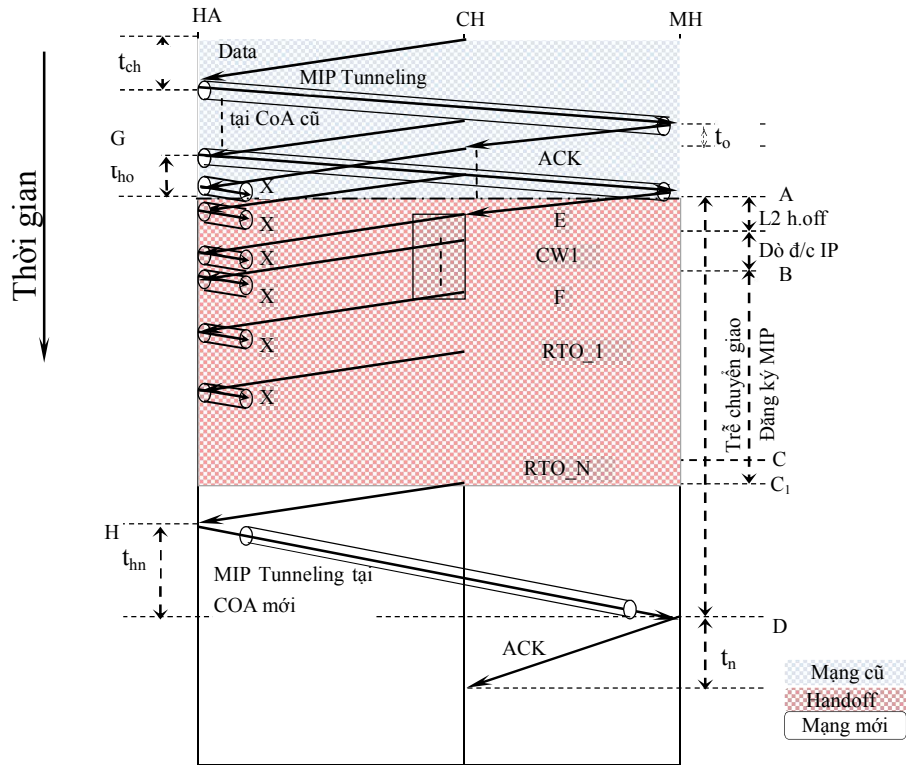


#### 2.4.1 Đánh giá hiệu suất chuyển giao của kết nối TCP khi sử dụng MIP

Trong MIP, HA dùng để chặn các gói số liệu cho MH khi nó không ở HN của mình. HA tạo ra kênh để truyền các gói số liệu tới MH thông qua FA khi dùng COA trạm khách hoặc chuyển trực tiếp tới NH khi dùng chung COA ở trọng mạng FN [58]. Hình 2.3 minh họa quá trình chuyển giao của kết nối TCP khi MH di chuyển từ ON sang NN.  $t_{ch}$  là trễ một chiều giữa CH và HA,  $t_{ho}$  là trễ một chiều giữa MH và HA của nó khi MH đang ở ON,  $t_{hn}$  là trễ một chiều giữa MH và HA của nó khi MH đang ở NN,  $t_0$  là trễ một chiều giữa MH và CH khi MH đang ở ON,  $t_n$  là trễ một chiều giữa MH và CH khi MH đang ở NN.

Ở hình 2.3, MH di chuyển vào NN tại thời điểm A. Do vậy lúc này MH bắt đầu chuyển giao lớp 2 tới NN. Như đã trình bày ở trên, trước thời điểm A, kết nối TCP đang hoạt động ở trạng thái ổn định tương ứng với ON. Ký hiệu kích thước cửa sổ nghẽn ở trạng thái ổn định là  $CW_1$ . Giả sử tất cả các gói số liệu do MH nhận được trước thời điểm A đều được xác nhận bởi CH. Gọi  $n$  là số tuần tự của gói số liệu nhận được tại thời điểm A. Như vậy, gói số liệu tiếp theo MH đang đợi là  $n+1$ . Theo hình 2.3, MH bắt đầu chuyển giao lớp 2 và đăng ký địa chỉ IP từ NN tại thời điểm A. Các thủ tục này kết thúc tại thời điểm B. Sau đó, tại thời điểm B, MH bắt đầu đăng ký MIP với HA của nó. Việc đăng ký COA mới của MH tại HA sẽ thành công tại thời điểm C. Như vậy các gói số liệu HA nhận được sau thời điểm C được chuyển tiếp chính xác tới MH trong NN. Các gói số liệu do HA của MH nhận từ CH trong khoảng thời gian từ thời điểm G đến C là bị mất do chúng được chuyển tiếp tới COA cũ của MH. Tại thời điểm E, CH sẽ nhận được tín hiệu ACK cuối cùng do MH gửi tại ON. Do vậy CH chuyển tất cả các gói số liệu trong cửa sổ nghẽn của nó sau thời điểm E và đợi tín hiệu xác nhận ACK, khi đó có thể xảy ra mấy khả năng sau:

- Trường hợp A: CoA mới của MH được đăng ký tại HA sau khi HA nhận được gói số liệu do CH truyền tại thời điểm F. Trong trường hợp này tất cả các gói số liệu ở cửa sổ nghẽn (từ E tới F) sẽ bị mất do HA tạo kênh truyền các gói số liệu này tới CoA cũ của MH. Do vậy, CH không nhận được các bản tin ACK xác nhận đối với các gói số liệu này và đợi cho đến khi RTO của gói số liệu được truyền tại thời điểm E. Sau đó nó giảm cửa sổ nghẽn tới 1 và truyền lại các gói số liệu mà RTO phát hiện tại thời điểm  $RTO_1$ . Nếu CoA mới của MH không được đăng ký ở HA tại thời điểm mà gói số liệu đã truyền lại này tới HA, thì HA sẽ gửi số liệu tới CoA cũ của MH và số liệu lại bị mất. Sau đó TCP của CH tăng 2 lần giá trị RTO, đợi đến khi RTO thứ 2, và truyền lại gói số liệu khi RTO thứ 2 hết hạn. Nếu các gói số liệu đã truyền lại sau khi RTO thứ N tới được HA sau khi CoA mới của MH được đăng ký tại HA thì HA sẽ tạo kênh số liệu tới CoA mới của MH. Trong trường hợp này gói số liệu đã truyền lại được MH nhận thành công ở NN.



Hình 2.3: Cơ chế chuyển giao của kết nối TCP sử dụng giao thức MIP

- Trường hợp B: CoA mới của MH được đăng ký tại HA trước khi HA nhận được gói số liệu từ CH tại thời điểm F. Trong trường hợp này, các gói số liệu thuộc cửa sổ nghẽn (từ E tới F) và xuất hiện sau việc đăng ký địa chỉ IP mới của MH tại HA thì được định tuyến tới CoA mới của MH. TCP cần 1 RTT để truyền tất cả các đoạn trong một cửa sổ nghẽn. Trễ chuyển giao lớn hơn RTT, do vậy trường hợp B này rất hiếm khi xảy ra.

Thời gian giảm thông lượng và Trễ chuyển giao ở trường hợp A được mô tả như sau:

#### 2.4.1.1 Trễ chuyển giao

Để tính toán trễ chuyển giao theo định nghĩa 2.2 (khoảng thời gian từ thời điểm MH nhận được gói số liệu thứ n tại mạng ON và thứ n+1 tại mạng NN), trước hết ta phải xác định được khoảng thời thất lạc các gói số liệu do MH truyền đi, đó là:

$$T = C - A = \tau_{L2} + \tau_a + \tau_m \quad (2.17)$$

Ở đây,  $\tau_{L2}$  là thời gian cần để MH chuyển giao lớp 2 tới NN,  $\tau_a$  Thời gian cần thiết để MH đăng ký được địa chỉ IP mới tại NN,  $\tau_m$  là thời gian đăng ký MIP [58]. Nếu hết thời gian T (thường xảy ra), thì TCP của MH sẽ rơi vào trạng thái Timeout, giảm cửa sổ nghẽn đi 1 và sau đó truyền lại số liệu mà RTO hết hiệu lực. Các gói số liệu sẽ lại bị mất nếu chúng đến HA

trước khi HA có CoA mới của MH. TCP MH sẽ tăng gấp đôi giá trị RTO và đợi cho đến khi Timeout để truyền lại. Theo cách này thì nếu N lần Timeout xảy ra trước khi HA nhận được CoA mới của MH, thì trễ chuyển giao  $T_{h1}$  sẽ là:

$$T_{h1} = D - A \quad (2.18)$$

$$T_{h1} = D - C_1 + C_1 - A \quad (2.19)$$

$C_1 - A$  phụ thuộc vào số lần Timeout xảy ra trước khi HA nhận được CoA mới. Theo hình 2.3,  $C_1 - A$  sẽ là:

$$\begin{aligned} C_1 - A &= C_1 - E + E - A \\ &= \begin{cases} TO_1 + \gamma TO_1 + \dots + \gamma^N TO_1 + t_o & \text{Khi } N \leq m \\ TO_1 + \gamma TO_1 + \dots + \gamma^m TO_1 + (N - m)\gamma^m TO_1 + t_o & \text{Khi } N > m \end{cases} \\ &= \begin{cases} TO_1 \frac{\gamma^{N+1} - 1}{\gamma - 1} + t_o & \text{Khi } N \leq m \\ TO_1 \frac{\gamma^{m+1} - 1}{\gamma - 1} + (N - m)\gamma^m TO_1 + t_o & \text{Khi } N > m \end{cases} \end{aligned} \quad (2.20)$$

Thay (2.20) vào (2.19) ta có:

$$T_{h1} = \begin{cases} TO_1 \frac{\gamma^{N+1} - 1}{\gamma - 1} + t_o + t_{ch} + t_{hn} & \text{Khi } N \leq m \\ TO_1 \frac{\gamma^{m+1} - 1}{\gamma - 1} + (N - m)\gamma^m TO_1 + t_o + t_{ch} + t_{hn} & \text{Khi } N > m \end{cases} \quad (2.21)$$

$TO_1$  là RTO đối với kết nối TCP khi MH ở ON, và  $TO_1 = \xi RTT_0$ ,  $\xi$  là hằng số,  $RTT_0$  là RTT khi MH ở ON. Từ hình 2.3 ta thấy,  $RTT_0 = t_{ch} + t_{h0} + t_0$ , trong đó  $t_{ch}$  là trễ một chiều giữa CH và HA,  $t_{h0}$  là trễ một chiều giữa MH và HA khi MH ở ON, và  $t_0$  là trễ một chiều giữa MH và CH khi MH ở ON.

T ở (2.17) yêu cầu xác định số lần Timeout truyền lại (N) mà TCP của MH đã trải qua trước khi HA nhận được CoA mới của MH. Khi đã xác định được N, ta có thể tính được trễ chuyển giao thông qua (2.21). Ở các mạng vô tuyến dùng riêng như WLAN, 3G, Vệ tinh,...  $\tau_{L2}$  và  $\tau_a$  trong (2.17) thường là hằng số. Mặt khác  $\tau_m$  phụ thuộc vào khoảng cách giữa MH và HA của nó, và điều kiện của kết nối vô tuyến.

$\tau_m$  Chính là thời gian cần để bản tin yêu cầu đăng ký MIP của MH [58] tới HA và bản tin trả lời đăng ký MIP của HA tới MH. Ví dụ,  $\tau_m = 2D_{mh}$ , ở đây  $D_{mh}$  là trễ một chiều trung bình để chuyển các gói tin báo hiệu giữa MH và HA (MIP sử dụng UDP để truyền các bản tin báo hiệu [58]). Tương tự, từ (2.12) ta có:

$$D_{mh} = (1 - q_1) \left\{ B_1 + A_1 \sum_{i=2}^m q_1^{i-1} (\gamma^{i-1} - 1) + \sum_{i=m+1}^{\infty} q_1^{i-1} [A_1 (\gamma^{m-1} - 1) + (i - m) \gamma^{m-2} \Delta_1] \right\} \quad (2.22)$$

Trong đó  $B_1$  là trễ truyền số liệu kết cuối giữa MH và HA.  $B_1 = B_{1nr}$  khi không sử dụng RLP, và  $B_1 = B_{1r}$  khi sử dụng RLP.  $B_{1nr}$  được xác định ở (72) khi sử dụng  $T_{nr} = B_{1nr}$  và  $t_w = t_{whn}$  và  $B_{1r}$  được xác định ở [74] khi sử dụng  $T_r = B_{1r}$ ,  $K = K_m$ , và  $t_w = t_{whn}$ .  $K_m = \left\lceil \frac{L_m}{L_f} \right\rceil$  là số khung lớp kết nối trên 1 bản tin yêu cầu hoặc trả lời đăng ký MIP,  $L_m$  là độ dài bản tin yêu cầu hoặc trả lời đăng ký MIP,  $L_f$  là độ dài khung lớp kết nối.  $t_{whn}$  là trễ một chiều trong mạng cố định, giữa NBS và HA.

$q_1$  là xác suất mất số liệu kết cuối giữa MH và HA,  $q_1 = q_{1nr}$  khi không sử dụng RLP và  $q_1 = q_{1r}$  khi sử dụng RLP.  $Q_{1nr}$  và  $q_{1r}$  xác định thông qua [73] và khi sử dụng  $p_{nr} = q_{1nr}$  và  $K = K_m$  và  $p_r = q_{1r}$  và  $K = K_m$

$\Delta_1$  Là giá trị đầu của Timer truyền lại đối với các bản tin báo hiệu MIP, nó phải đủ lớn để đếm được kích cỡ các bản tin báo hiệu MIP, bằng 2 lần RTT giữa NH và HA, và bé nhất là = 100ms đủ để xử lý các bản tin ở MH và HA.  $\gamma$  là thừa số biểu diễn khoảng thời gian RTO, và tăng lên 1 sau mỗi lần truyền lại bị lỗi, thường = 2.  $A_1 = \frac{\Delta_1}{\gamma - 1} \cdot m$  là số nguyên.

#### 2.4.1.2 Thời gian giảm thông lượng

Như trình bày ở trên, HA nhận được gói số liệu truyền lại thứ N sau khi việc đăng ký CoA mới của MH thành công. Do vậy, HA tạo ra kênh để truyền lại gói số liệu thứ N và sau đó được TCP của CH truyền tới CoA mới của MH. TCP của CH khôi phục lại hoạt động khởi động chậm (low start) tại thời điểm  $C_1$  như hình 2.3. Sau đó nó tăng giá trị cửa sổ nghẽn lên trạng thái ổn định của NN, biểu thị bởi  $CW_2$ . Thời gian  $\tau_s$  cần để kích cỡ cửa sổ nghẽn tăng từ 1 lên  $CW_2$  được xác định như sau:

$$\tau_s = [1 + \log CW_2] RTT_h \quad (2.23)$$

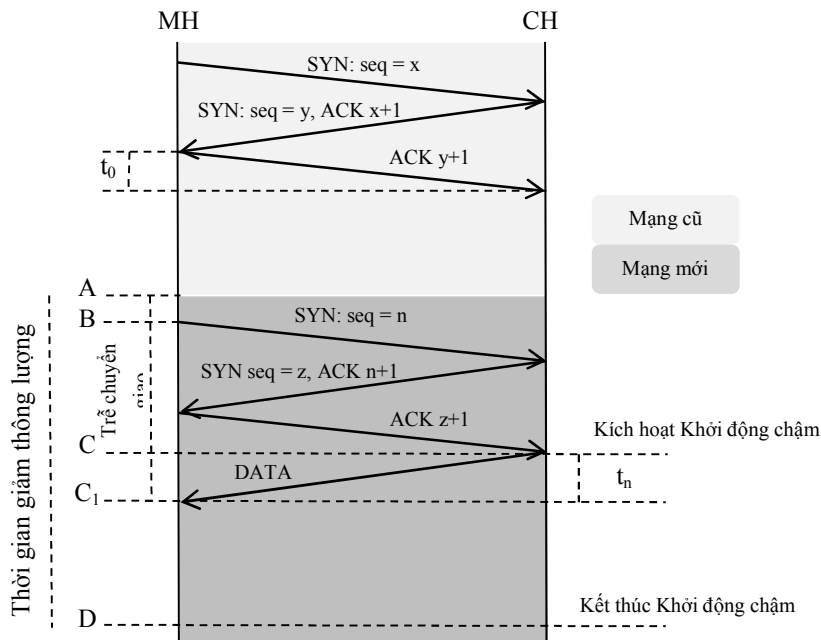
$RTT_n$  là RTT khi MH thuộc NN, và  $RTT_n = t_{ch} + t_{hn} + t_n$ , trong đó  $t_{ch}$  là trễ một chiều giữa CH và HA,  $t_{hn}$  là trễ một chiều giữa MH và HA khi MH ở NN, và  $t_n$  là trễ một chiều giữa MH và CH khi MH ở NN.

Thời gian để việc giảm thông lượng kết nối TCP ( $T_{tl}$ ) bằng  $T_t = (C_1 - A) + \tau_s$ . Từ (2.20) và (2.23) ta có:

$$T_{tl} = \begin{cases} TO_1 \frac{\gamma^{N+1} - 1}{\gamma - 1} + t_o + [1 + \log_2 CW_2] RTT_n & \text{Khi } N \leq m \\ TO_1 \frac{\gamma^{m+1} - 1}{\gamma - 1} + (N - m)\gamma^m TO_1 + t_o + [1 + \log_2 CW_2] RTT_n & \text{Khi } N > m \end{cases} \quad (2.24)$$

### 2.4.2 Đánh giá hiệu suất chuyển giao của kết nối TCP khi sử dụng TCP-M

Chọn TCP-M [64] làm giao thức quản lý di động lớp vận chuyển vì nó giảm tối đa yêu cầu thay đổi cấu trúc mạng. Ngược lại, các giải pháp khác như MSOCKS [47][48] yêu cầu phải bổ sung thêm phần tử mạng mới như Proxy để phân chia kết nối TCP. Dưới đây là phần giải thích nguyên tắc hoạt động của TCP-M khi chuyển giao.



Hình 2.4: Sơ đồ hoạt động của TCP-M

MH và CH đăng ký một Token thông qua chức năng mở rộng (migrate) mô tả ở [74] trong suốt quá trình thiết lập kết nối TCP. Do vậy, kết nối TCP có thể được định nghĩa duy nhất tại MH và CH bởi  $\langle \text{Địa chỉ MH}, \text{Cổng MH}, \text{địa chỉ CH}, \text{Cổng CH} \rangle$  hoặc  $\langle \text{địa chỉ CH}, \text{Cổng CH}, \text{Token} \rangle$  [74]. Khi MH di chuyển tới NN và nhận được địa chỉ IP mới, nó gửi địa chỉ IP mới

của mình có chứa SYN và một lựa chọn Lưu trữ (Migrate) tới CH. Đoạn SYN này bao gồm Token được xác định trong quá trình thiết lập kết nối ban đầu tại trường Token. CH nhận dạng kết nối tương ứng với Token này và thay đổi địa chỉ và cổng cho khớp với địa chỉ IP mới của MH. Sau đó CH đặt trạng thái nghẽn liên quan của kết nối về giá trị ban đầu và khôi phục lại kết nối từ trạng thái khởi động chậm (low start) của TCP.

Hình 2.4 minh họa quá trình chuyển giao TCP-M của kết nối TCP khi MH di chuyển từ ON sang NN. Tại thời điểm A, MH bắt đầu tiến trình chuyển giao sang NN. Giả sử rằng, tất cả số liệu do MH nhận được trước thời điểm A đã được xác nhận (ACK). Gọi n là số thứ tự của gói số liệu nhận được tại thời điểm A, như vậy gói số liệu tiếp theo MH nhận được là n+1. Như hình 2.4, MH bắt đầu chuyển giao lớp 2 tới NN và đăng nhập địa chỉ IP ở NN tại thời điểm A. Các thủ tục này sẽ hoàn thành tại thời điểm B, Sau đó MH bắt đầu quá trình chuyển giao TCP-M và kết thúc tại thời điểm C<sub>1</sub>, lúc này CH khôi phục lại trạng thái kết nối từ giai đoạn khởi động chậm (low start) và kết thúc tại D. Như vậy, kết nối TCP giành được trạng thái ổn định tương ứng với NN.

#### 2.4.2.1 Trễ chuyển giao

Từ hình 2.4 ta thấy trễ chuyển giao TCP-M được xác định như sau:

$$T_{h2} = C - A = \tau_{L2} + \tau_a + E[L] + t_n$$

Ở đây  $\tau_{L2}$  và  $\tau_a$  là thời gian cần thiết để MH chuyển giao lớp 2 tới NN và đăng nhập được địa chỉ IP ở đây.  $t_n$  là trễ một chiều giữa CH và MH khi MH ở NN.  $E[L]$  Trễ trung bình để chuyển bản tin báo hiệu TCP-M.

Để biểu diễn  $E[L]$ , đầu tiên MH gửi gói tin SYN cùng với lựa chọn TCP-M chứa địa chỉ IP mới của MH tới CH  $i \geq 0$  lần không thành công cho đến gói SYN thứ  $i+1$  (mới thành công). Tiếp đến CH sẽ phát lặp lại SYN/ACK của nó cho tới khi nó nhận được ACK từ MH. Giả sử CH gửi SYN/ACK tới MH  $j \geq 0$  lần không thành công, và lần  $j+1$  là thành công. Sau đó MH truyền lại ACK tới CH mà thành công ở lần truyền thứ  $k+1$  ( $k \geq 0$ ). Do vậy, xác suất  $ph(i,j,k)$  chuyển giao TCP-M được hoàn thành sau  $i$  lần trao đổi SYN không thành công rồi mới có 1 lần trao đổi SYN thành công,  $j$  lần SYN/ACK lỗi tiếp đến 1 thành công, và  $k$  lần ACK lỗi tiếp theo 1 thành công, được biểu diễn bởi:

$$P_h(i, j, k) = p_1^i (1 - p_1) p_2^j (1 - p_2) p_2^k (1 - p_2) \quad \text{với } i, j, k = 0, 1, 2, \dots, N_m - 1 \quad (2.25)$$

$N_n$  là số lần truyền lại,  $p_1$  là xác suất mất số liệu kết cuối giữa CH và MH đối với gói SYN,  $p_2$  là xác suất mất số liệu kết cuối giữa CH và MH đối với gói SYN/ACK hoặc ACK.  $p_1 = p_{1nr}$  khi không sử dụng RLP,  $p_1 = p_{1r}$  khi sử dụng RLP. Từ [73] ta tính được  $p_{1nr}$  với  $p_{nr} = p_{1nr}$  và  $K=K_1$ .

Từ [75] ta tính được  $p_{1r}$  với  $p_r = p_{1r}$  và  $K=K_1$ .  $K_l = \left\lceil \frac{L_1}{L_f} \right\rceil$  là số khung lớp kết nối trên một gói SYC.  $L_1$  là chiều dài gói SYC và  $L_f$  là chiều dài khung lớp kết nối. Tương tự ta cũng xác định được  $p_2$ . Theo nguyên tắc trên, trễ chuyển giao sẽ là:

$$\begin{aligned} L_h(i, j, k) &= 1.5RTT_n + \sum_{m=0}^{i-1} 2^m RTO + \sum_{m=0}^{j-1} 2^m RTO + \sum_{m=0}^{k-1} 2^m RTO \\ &= 1.5RTT_n + (2^i + 2^j + 2^k - 3)RTO \end{aligned}$$

Với  $i, j, k = 0, 1, 2, \dots, N_m-1$  (2.26)

Trong đó  $RTO = \mathcal{RTT}_j$  và khi ở ON thì  $RTT_0$  chính là  $RTT$ . Do vậy trễ chuyển giao TCP-M trung bình sẽ là:

$$E[L] = \sum_{i=0}^{N_m-1} \sum_{j=0}^{N_m-1} \sum_{k=0}^{N_m-1} P_h(i, j, k) L_h(i, j, k). \quad (2.27)$$

#### 2.4.2.2 Thời gian giảm thông lượng

Như mô tả ở trên, TCP của CH khôi phục lại trạng thái khởi động chậm (low start) tại thời điểm  $C_1$  (hình 2.4). Sau đó nó tăng cửa sổ nghẽn lên giá trị trạng thái ổn định của NN biểu diễn bởi  $CW_2$ . Giả sử rằng, tại thời điểm D thì kích thước cửa sổ nghẽn đạt được giá trị trạng thái ổn định, số liệu không bị mất. Do vậy, từ thời điểm A đến D là khoảng thời gian giảm thông lượng (kí hiệu  $T_{l2}$ ) của kết nối TCP, được xác định như sau:

$$T_{l2} = D - A = \tau_{L2} + \tau_a + E[L] + [1 + \log_2 CW_2] RTT_n \quad (2.28)$$

$\tau_{L2}$  và  $\tau_a$  thời gian để MH chuyển giao lớp 2 tới NN và đăng nhập địa chỉ IP tại NN. tn trễ một chiều giữa CH và MH khi MH ở NN.  $RTT_n$  là  $RTT$  khi MH ở NN.  $E[L]$  xác định tại [78].

### 2.5 Xây dựng phương thức đánh giá hiệu suất chuyển giao của các ứng dụng dạng D và dạng E (MIP và SIP)

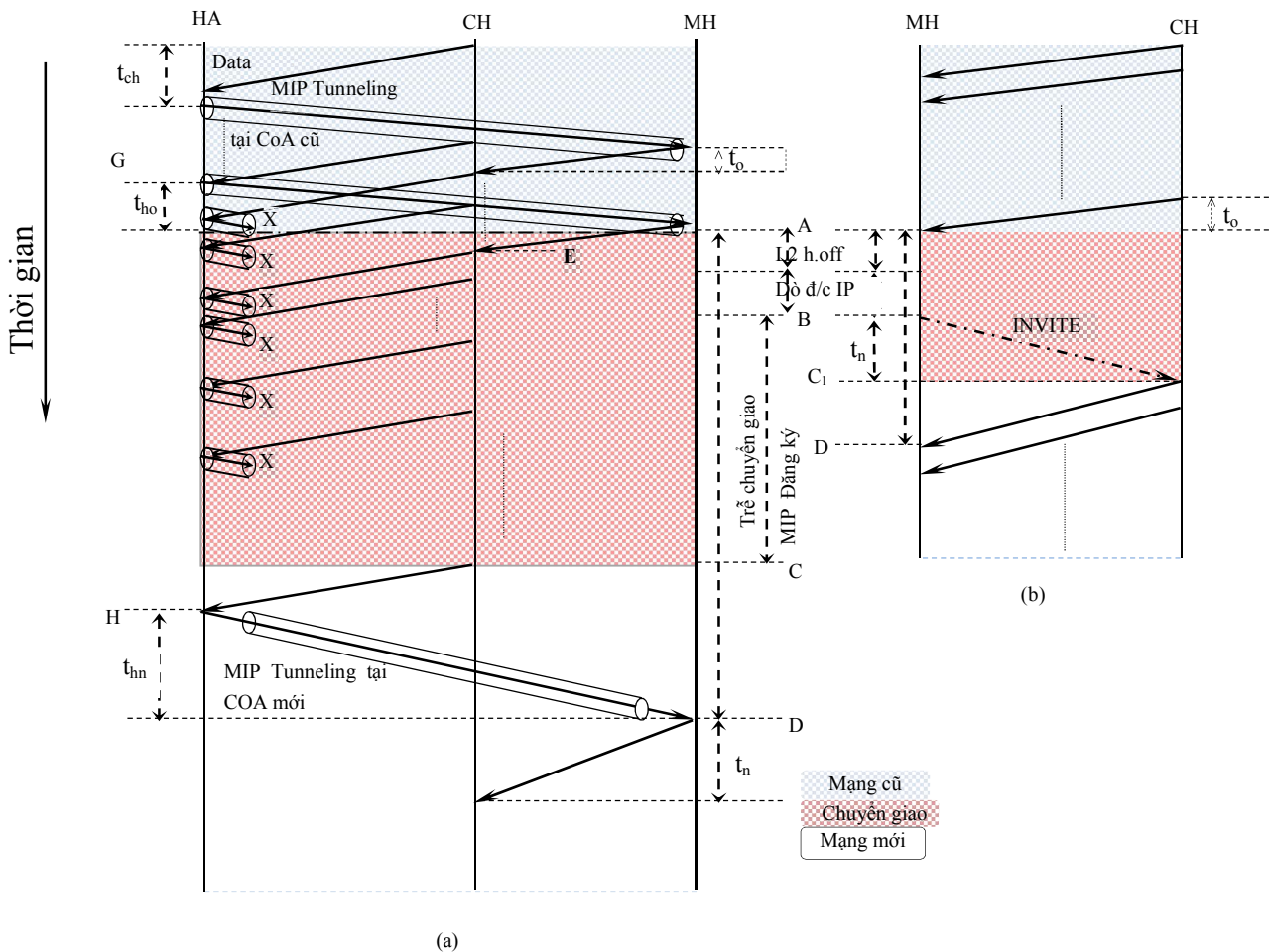
Do các ứng dụng dạng D và E sử dụng giao thức kết nối là UDP giữa MH và CH, hiệu suất chuyển giao của ứng dụng dạng D và dạng E phụ thuộc vào hiệu suất chuyển giao của kết nối UDP. Do vậy luận án lưu ý đến việc phân tích ứng dụng VoIP sử dụng RTP trên UDP.

Việc phân tích này dựa trên các ứng dụng UDP thời gian thực và không thời gian thực. UDP không phải là giao thức vận chuyển tin cậy, do đó các gói tin bị thất lạc trong quá trình chuyển giao sẽ không được lấy lại. Hơn nữa trong trường hợp khi ta khảo sát một ứng dụng thời gian thực VoIP, lợi ích của việc lưu trữ vào bộ đệm và truyền các bản tin khi quá trình chuyển giao



kết thúc sẽ không được áp dụng. MIP và SIP đều hỗ trợ bảo mật cho việc chuyển giao, cho nên vấn đề bảo mật cũng sẽ không được đề cập trong phần này. Dựa vào các kết quả đã thu thập, ba thành phần chính được khảo sát đối với hiệu suất của MIP, SIP cho ứng dụng VoIP là: *độ trễ chuyển giao, gói tin bị thất lạc khi chuyển giao và độ trễ điểm đến điểm*. Độ trễ điểm đến điểm tương ứng với độ trễ truyền của các gói tin dữ liệu VoIP.

### 2.5.1 Đánh giá hiệu suất chuyển giao của một kết nối UDP khi sử dụng MIP



Hình 2.5: Chuyển giao kết nối TCP sử dụng MIP và SIP

Ở hình 2.5(a) cho thấy rằng quá trình chuyển giao MIP của một kết nối UDP khi MH dịch chuyển từ mạng ON đến mạng NN. Ở đây,  $t_{ch}$  là độ trễ một chiều giữa CH và HA,  $t_{hn}$  là độ trễ một chiều giữa MH và HA khi MH đang trong mạng NN, và  $t_o$  là độ trễ một chiều giữa MH và CH trong khi MH đang trong mạng ON. Như thể hiện trong hình 2.5(a), MH khởi tạo chuyển giao ở lớp 2 đến mạng NN và lấy được một địa chỉ IP từ mạng NN tại thời gian A. Các thủ tục này được thực hiện tại thời gian B. Do đó, tại thời gian B MH khởi tạo việc đăng ký MIP với mạng HA của nó. Một CoA của MH sẽ biểu thị cho việc đăng ký thành công tại HA vào một thời gian cụ thể là C. Các bản tin nhận được ở HA sau thời gian C sẽ được chuyển tiếp tới MH



chính xác ở mạng mới NN. Các gói tin nhận được ở mạng HA của MH từ CH giữa thời gian G và C sẽ bị thất lạc vì chúng được chuyển tiếp đến CoA cũ của MH. Ở đây, độ trễ chuyển giao sẽ được tính bằng khoảng thời gian khi MH nhận được bản tin cuối cùng khi nó trong mạng ON cho đến khi MH nhận được bản tin đầu tiên khi nó ở trong mạng NN. Đối với độ trễ chuyển giao, gói tin bị thất lạc khi chuyển giao và độ trễ điềm tới điềm của một kết nối UDP khi sử dụng MIP sẽ được thể hiện thông qua các phân tích toán học ở phần dưới đây.

### 2.5.1.1 Độ trễ chuyển giao

Từ hình 2.5(a), độ trễ chuyển giao của kết nối UDP được tính như sau:

$$T_{h3} = D - A = \tau_{L2} + \tau_a + \tau_m + t_{ch} + t_{hn} \quad (2.29)$$

Với  $\tau_{L2}$ ,  $\tau_a$  và  $\tau_m$  được định nghĩa ở phần 2.4.1

### 2.5.1.2 Gói tin bị thất lạc

Từ hình 2.5, các gói tin bị chặn bởi HA giữa thời gian G và C sẽ bị thất lạc. Do đó, nếu tỉ lệ truyền gói tin của CH là R, số lượng các gói tin bị thất lạc trong quá trình chuyển giao sẽ là:

$$P_h = R(C - G) = R(\tau_{L2} + \tau_a + \tau_m + t_{ho}) \quad (2.30)$$

### 2.5.1.3 Độ trễ khi truyền gói tin từ điềm đến điềm

Độ trễ khi truyền gói tin từ điềm đến điềm của các gói tin dữ liệu VoIP theo đường từ MH đến CH là  $D_{fm}$  khi không có RLP và là  $D_{fmnr}$  khi có RLP.

$$D_{fmnr} = D + t_{wcn} \quad (2.31)$$

$$D_{fmr} = D + (K_p - 1)\tau + t_{wcn} \quad (2.32)$$

D là độ trễ truy cập lớp kết nối và  $t_{wcn}$  là độ trễ một chiều trong mạng vô tuyến giữa BS mới (NBS) và CH.  $K_p = \frac{L_p}{L_f}$  là số lượng của các frame lớp kết nối trên một bản tin dữ liệu VoIP,

mà  $L_p$  là độ dài của một gói tin dữ liệu VoIP và  $L_f$  là độ dài của một frame lớp kết nối. Tương tự, độ trễ khi truyền gói tin từ điềm đến điềm từ MH đến CH là ( $D_{rm}$ ) khi không có RLP và là  $D_{rmnr}$  khi có RLP.

$$D_{rmnr} = D + t_{ch} + t_{whn} \quad (2.33)$$

$$D_{rnr} = D + (K_p - 1)\tau + t_{ch} + t_{whn} \quad (2.34)$$

$t_{ch}$  là độ trễ thời gian một chiều giữa CH và HA.  $t_{whn}$  là độ trễ thời gian một chiều trong mạng vô tuyến giữa NBS và HA.

## 2.5.2 Đánh giá hiệu suất chuyển giao của một kết nối UDP khi sử dụng SIP

Việc quản lý di chuyển dựa trên SIP được đề xuất để loại bỏ các điểm yếu của MIP [79]. Khi MH dịch chuyển từ mạng ON sang mạng NN, nó sẽ gửi một bản tin INVITE [59] đến CH sử dụng bộ nhận biết cuộc gọi giống với khi thiết lập cuộc gọi ban đầu như thể hiện ở hình 2.5(b). MH đặt một địa chỉ IP mới trong trường contact của bản tin SIP INVITE [79]. Địa chỉ IP mới này sẽ thông báo cho CH về việc chuyển mạng của MH. Do đó, sau khi nhận được địa chỉ IP mới của MH, CH gửi các gói tin dữ liệu VoIP cho MH theo địa chỉ IP mới. Ở hình 2.5(b) thể hiện quá trình chuyển giao với SIP từ mạng ON đến mạng NN. Ở thời điểm A, MH bắt đầu quá trình chuyển giao đến mạng NN. MH bắt đầu với chuyển giao lớp 2 đến mạng NN và lấy được địa chỉ IP từ mạng NN tại thời điểm A. Những thủ tục này được hoàn thành vào thời gian B. Sau đó, tại thời gian B MH gửi bản tin INVITE đến CH và nhận được ở CH bản tin này ở thời gian  $C_1$ . Do đó, các gói tin được gửi bởi CH trong khoảng thời gian thời gian  $A - t_0$  và  $C_1$  sẽ bị thất lạc vì chúng được gửi đến địa chỉ IP cũ của MH.

### 2.5.2.1 Độ trễ khi chuyển giao

Từ hình 2.5(b) độ trễ khi chuyển giao khi sử dụng giao thức SIP được tính như sau:

$$T_{h4} = D_1 - A = \tau_{L2} + \tau_a + 2D_{mc} \quad (2.35)$$

Với  $\tau_{L2}$  và  $\tau_a$  là thời gian cần thiết để chuyển giao lớp 2 của MH đến mạng NN và được cấp phát địa chỉ IP mới từ mạng NN.  $D_{mc}$  là độ trễ một chiều trung bình để truyền các bản tin báo hiệu SIP giữa MH và CH. Các bản tin báo hiệu SIP có thể được truyền bằng cách sử dụng giao thức UDP hoặc TCP [59]. Để phân tích, ở đây luận án quan tâm đến bản tin báo hiệu được truyền bằng giao thức UDP. Với các bước tương tự như ở biểu thức (2.12),  $D_{mc}$  được tính như sau:

$$D_{mc} = (1 - q_2) \left\{ B_2 + A_2 \sum_{i=2}^m q_2^{i-1} (\gamma^{i-1} - 1) + \sum_{i=m+1}^{\infty} q_2^{i-1} [A_2 (\gamma^{m-1} - 1) + (i - m) \gamma^{m-2} \Delta_2] \right\} \quad (2.36)$$

Trong đó  $B_2$  là trễ truyền số liệu kết cuối giữa MH và CH.  $B_2 = B_{2nr}$  khi không sử dụng RLP, và  $B_2 = B_{2r}$  khi sử dụng RLP.  $B_{2nr}$  được xác định ở (4.16) khi sử dụng  $T_{nr} = B_{2nr}$  và  $t_w = t_{wcn}$ ,  $t_{wcn}$  là trễ một chiều trong mạng cố định giữa BS mới và CH. và  $B_{2r}$  được xác định ở (4.17) khi

sử dụng  $T_r = B_{2r}$ ,  $K = K_s$  và  $t_w = t_{wcn}$ .  $K_s = \left\lceil \frac{L_s}{L_f} \right\rceil$  là số khung lớp kết nối trên 1 bản tin SIP

INVITE,  $L_s$  là độ dài bản tin SIP INVITE,  $L_f$  là độ dài khung lớp kết nối.

$q_2$  là xác suất mất số liệu kết cuối giữa MH và CH.  $q_2 = q_{2nr}$  khi không sử dụng RLP và  $q_2 = q_{2r}$  khi sử dụng RLP.  $Q_{2nr}$  tính được theo [73] khi sử dụng  $p_{nr} = q_{2nr}$  và  $K = K_s$ .  $q_{2r}$  tính được theo (4.19) khi  $p_r = q_{2r}$  và  $K = K_s$ .

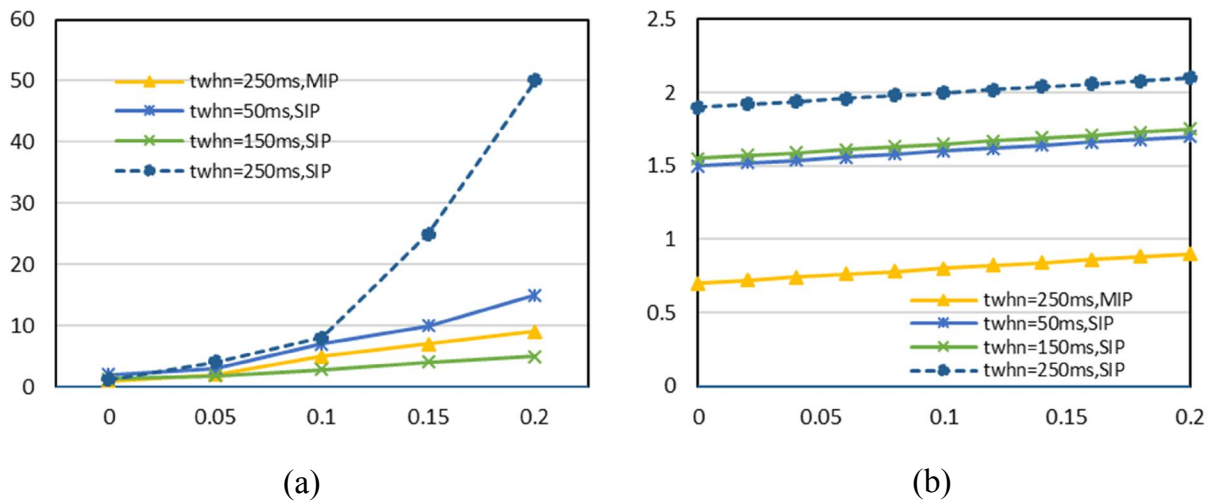
$\Delta_2$  là giá trị đầu của đồng hồ truyền lại cho các bản tin báo hiệu SIP, nó đủ lớn để đếm được kích cỡ của các bản tin báo hiệu SIP, 2 RTT giữa MH và CH, và ít nhất là 100ms để xử lý các bản tin tại MH và CH.  $\gamma$  là thừa số, biểu diễn số lần truyền lại do Timeout. Thông thường  $\gamma = 2$ ,  $A_2 = \frac{\Delta_2}{\gamma - 1}$ ,  $m$  là số nguyên biểu diễn thời điểm đồng hồ truyền lại hết tác dụng.

### 2.5.2.2 Mất số liệu

Theo hình 2.5(b), trong khoảng thời gian giữa  $A-t_0$  và  $C_1$  tất cả các gói số liệu do CH truyền lại đều bị mất. Do vậy nếu  $R$  là tốc độ truyền lại số liệu của CH thì số gói dữ liệu bị mất trong quá trình chuyển giao sẽ là:

$$P_{hl} = R(C_1 - A + t_0) = R(\tau_{L2} + \tau_a + D_{mc} + t_0)$$

### 2.5.2.3 Trễ vận chuyển dữ liệu kết cuối



Hình 2.6: So sánh trễ chuyển giao giữa MIP và SIP khi không có RLP (a) và có RLP (b)

Trễ vận chuyển dữ liệu kết cuối của các gói số liệu VoIP giữa MH và CH là như nhau trong cả hai trường hợp có và không có RLP, và xác định bởi:

$$D_{fsnr} = D_{rsnr} = D + t_{wcn} \quad (2.37)$$

$$\text{Và } D_{fsr} = D_{rsr} = D + (K_p - 1)\tau + t_{wcn} \quad (2.38)$$

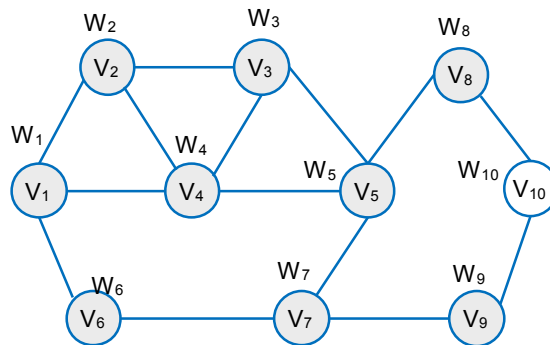
Trong đó  $D_{sfnr}$  và  $D_{sfr}$  là trễ vận chuyển số liệu kết cuối của dữ liệu VoIP từ MH tới CH cho trường hợp có và không có RLP. Và  $D_{rsnr}$  và  $D_{rsr}$  là trễ theo chiều ngược lại từ CH tới MH. Do vậy, khi sử dụng SIP thì trễ loại này là đối xứng cho cả hai chiều (vì trong SIP không định hướng lại số liệu).

## 2.6 Phân tích đánh giá tương quan giữa tiêu hao nguồn điện và hiệu suất chuyển giao

Gần đây, một số nghiên cứu đã đề xuất việc định tuyến lựa chọn mạng chuyển giao dựa trên sự kết hợp giữa mức năng lượng (nguồn điện) còn lại và số bước nhảy. Theo đó, trong [24], các tác giả đã đề xuất một giao thức định tuyến, gọi là HCESDSR (Định tuyến nguồn động tiết kiệm năng lượng dựa trên Hop Count) để cải thiện tuổi thọ cho mạng. Ý tưởng chính ở đây là sử dụng tham số mới đó là tuổi thọ tối thiểu dự kiến của tuyến đường. Dựa trên tham số này, hàm chi phí sẽ chọn tuyến đường có tuổi thọ trung bình dự kiến cao nhất.

### 2.6.1 Mô hình phân tích

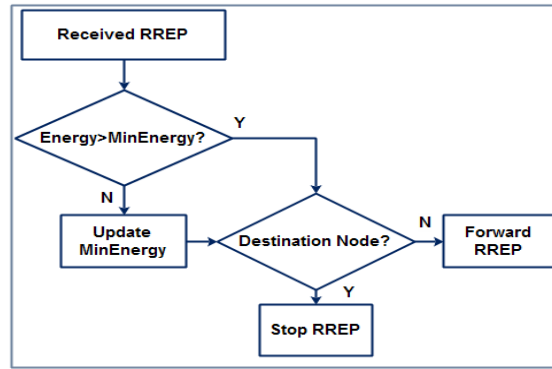
Như trong hình 2.7. mỗi nút di động có trọng số biểu diễn dung lượng pin còn lại. Từ mô hình hệ thống, ta có thể xây dựng được mô hình mạng kết nối. Định nghĩa  $G = (V, E)$  là đồ thị truyền tin của MANET, trong đó  $V = \{V_1, \dots, V_n\}$  là tập hợp các đầu cuối di động,  $E$  là tập hợp các kết nối,  $L_{i,j} = (V_i, V_j)$  nếu chúng có kết nối trực tiếp. Để giải quyết các vấn đề định tuyến hiệu suất cao cho MANET, kiến trúc mạng kết nối bằng đồ thị có trọng số nút được mô hình hóa như sau. Mỗi nút di động  $V_i$  được biểu diễn thành một cặp  $(V_i, W_i)$ , trong đó  $i$  là lượng Pin còn lại của  $V_i$ . Một nút di động có thể kết nối trực tiếp hoặc gián tiếp (thông qua các nút trung gian) với các nút khác.



Hình 2.7: Mô hình mạng MANET

### 2.6.2 Mô tả giao thức

Giống như giao thức AODV [62] của IETF, chúng tôi xây dựng giao thức hoạt động theo nguyên tắc mỗi khi có yêu cầu truyền số liệu thì nút nguồn sẽ tìm đường tới nút cuối.



Hình 2.8: Thủ tục kiểm tra tiêu hao năng lượng

Đầu tiên, nút nguồn gửi các gói RREQ với Header thay đổi là  $\{MinEnergy, AODV RREQ Header\}$ . Sau đó các gói này sẽ được chuyển tiếp tới các nút trung gian trước khi đi tới nút đích. Điểm khác biệt với các giao thức khác là thực hiện việc kiểm tra năng lượng (Energy-check) tại mỗi nút trung gian như mô tả trong hình 2.8. Thuật toán này thực hiện việc kiểm tra và cập nhật dung lượng Pin còn lại tại mỗi đường.

Cuối cùng, nút đích sẽ gửi RREP (route reply) với Header được sửa đổi  $\{MinEnergy, AODV RREP Header\}$  tới nút nguồn.

Ngoài ra, tương tự như AODV, thủ tục duy trì tuyến của giao thức sử dụng bản tin RRER. Nút nguồn sẽ nhận tất cả các tuyến khả dụng khi hoàn thành thủ tục tìm tuyến này (Hình 2.9).

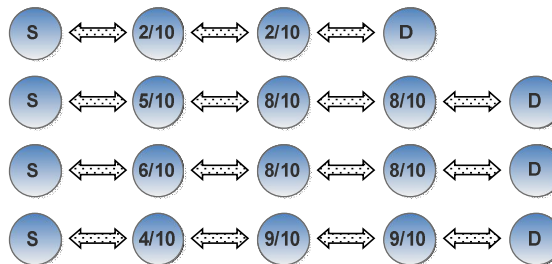
### 2.6.3 Thuật toán tìm tuyến

Sau khi nhận được các tuyến khả dụng, chúng tôi đưa ra hai ràng buộc:

1) Số bước nhảy của tuyến phải thuộc  $[Hopmin, Hopmax]$ . Tuyến có số bước nhảy nằm ngoài dải này sẽ bị loại bỏ.

$$Hopcount = [Hopmin, Hopmax] \quad (2.39)$$

ở đây Hopmin là số bước nhảy của tuyến khả dụng ngắn nhất giữa cặp nút (S,D). Định nghĩa  $Hopmax = Hopmin + k$ , với k là số tự nhiên để hạn chế số tuyến khả dụng cần quan tâm. Khi mô phỏng, đặt  $k = 2$ .



Hình 2.9: Tập tuyến khả dụng giữa các cặp nút sau pha khám phá.

2) Để đạt được các tuyến khả dụng nhiều năng lượng, hàm chi phí EEMA được định nghĩa như sau:

Gọi  $MEP_{(i)}$  là năng lượng tối thiểu còn lại của tuyến  $i$  nhận được thông qua trường  $MinEnergy$  của bản tin RREP.

Gọi  $N$  và  $MinEnergySet$  (MES) là tổng số tuyến và tập chi phí tuyến của các tuyến khả dụng thỏa mãn các điều kiện trên, ta có:

$$EnergySet = \begin{bmatrix} MES_{(1)} \\ MES_{(2)} \\ \vdots \\ MES_{(n-1)} \\ MES_{(n)} \end{bmatrix} \quad (2.40)$$

$$Optionalroute = \text{Max} (EnergySet) \quad (2.41)$$

Thông qua (2.41) ta xác định được tuyến tốt nhất. Thuật toán lựa chọn tuyến EEMA có thể tóm tắt như sau:

1. routeset=shortest-route(S,D)
2. minhop=min(shortest-route(S,D))
3. maxhop=minhop+2; cons1valid=∅
4. // Equation (1)
5. **for** i=1 to maxsizeof(routeset) **do**
6.     **if** minhop ≤ numhop(routeset(i)) ≤ maxhop **then**
7.         cons1valid <- route(i)
8.     **Endif**
9. **Endfor**
10. // Equation (2)
11. Cost= ∞, weight=0
12. **for** i=1 to sizeof(cons1valid) **do**
13.     weight= MinEnergy(cons1valid(i))
14.     **if** Cost > weight **then**
15.         Cost=weight
16.         selectedroute=cons1valid(i)
17.     **Endif**
18. **Endfor**
19. **Return** (selectedroute, cost)

Giả sử tồn tại 4 tuyến giữa cặp nút nguồn (S) và nút đích (D), với dung lượng Pin của mỗi nút như hình 2.9. Dung lượng Pin còn lại của các nút S và D có giá trị là 5/10, Hopmin = 3 và

Hopmax = 5. Dựa vào thông tin có được, sử dụng hàm chi phí và công thức (2.41), ta thấy EEMA sẽ chọn tuyến số 3 với chi phí là 0.6 như bảng 2.2 dưới đây.

Bảng 2.2: Phương pháp tính cho phí tuyến

<i>Route</i>	<i>P</i>	<i>Min Energy</i>	<i>EEMA</i>
1	3	0.20	0.2
2	4	0.50	0.5
3	4	0.60	0.6
4	4	0.40	0.4

Bằng phương pháp phân phối tải động được đề xuất, khả năng cao tìm thấy tuyến giữa một cặp nút đảm bảo sự cân đối giữa hai tiêu chí: tuổi thọ và hiệu suất của mạng.

#### 2.6.4 Cấu trúc bản tin RREQ

Trong giao thức EEMA, quá trình tìm kiếm và xác định tuyến giữa nút nguồn và nút đích dựa trên các bản tin RREQ và RREP được mô tả trong Phần 2.6.2. Để có được thông tin và xác định tuyến tối ưu theo phương trình (2.41), các trường dành riêng trong header của bản tin RREQ để lưu giá trị năng lượng (trường MinEnergy) được sử dụng. Phương pháp này đã được đề xuất trong nhiều nghiên cứu gần đây [7][81][84]. Sử dụng các trường dành riêng trong header của bản tin RREQ giúp xác định số liệu của tuyến mà không tăng tải định tuyến và mức tiêu thụ năng lượng. Do đó, bản tin RREQ được mở rộng như trong Hình 2.10.

Loại	Dự phòng	Hop cuối	Bộ đếm Hop
RREQ ID			
Địa chỉ IP đích			
Số tuần tự đích			
Địa chỉ IP nguồn			
Số tuần tự nguồn			

Hình 2.10: Cấu trúc của bản tin RREQ

## 2.6.5 Mô phỏng và đánh giá

Trong phần này, Luận án thiết lập môi trường mô phỏng để đánh giá và so sánh hiệu suất của EEMA với hai giao thức định tuyến điển hình là AERP [82] và AODV [62], trên phiên bản phần mềm NS2 2.34.

### 2.6.5.1 Chỉ số hiệu suất

Sử dụng các số liệu sau đây để đánh giá hiệu suất của các giao thức định tuyến được thử nghiệm.

- 1) *Tỷ lệ chuyển gói (PDR) (tính bằng%)*: tỷ lệ số bản tin được gửi đến các nút đích  $P_r$  vượt quá số lượng bản tin do các nút nguồn  $P_s$  gửi:

$$PDR = \frac{P_r}{P_s} \times 100\% \quad (2.42)$$

- 2) *Trung bình độ trễ kết cuối*: Thời gian cần thiết để truyền bản tin từ nút nguồn tới nút đích:

$$\text{Delay}_{\text{avg}} = \frac{\sum_{i=1}^n (t_r - t_s)}{P_r} \quad (2.43)$$

- 3) *Thông lượng*: Thông lượng trên một kết nối là tích của số lượng bản tin được truyền và kích thước nó trong một giây

$$\text{Throughput} = \frac{P_r \times KT}{T} \quad (2.44)$$

- 4) *Thời gian duy trì mạng*: là khoảng thời gian mà mạng bắt đầu hoạt động cho đến khi nút đầu tiên hết năng lượng, hay chính xác hơn là năng lượng của nút nhỏ hơn ngưỡng.

Trong đó:

$P_r$  là số bản tin nút đích nhận được;

$P_s$  là số bản tin do nút nguồn gửi đi;

$t_r$  là thời gian bản tin nhận được tại nút đích;

$t_s$  là thời gian bản tin gửi đi tại nút nguồn;

$T$  là thời gian của quá trình đo;

$KT$  là kích thước của bản tin.

### 2.6.5.2 Các tham số mô phỏng

Luận án sử dụng loại lưu lượng CBR với 100 giá trị ngẫu nhiên gán cho nút di động (sử dụng mô hình di động Waypoint ngẫu nhiên) trong phạm vi  $1000 \times 1000$  (m). Dải truyền của nút di động là 250 m. Vận tốc của các nút di động đặt ở mức 2 (m / s). Thời gian mô phỏng là 500



(s). Số lượng kết cuối (end to end) được đo là 10, 20,..., 90, 100. Các tham số mô phỏng được tóm tắt trong bảng 2.3 dưới đây.

Bảng 2.3: Các tham số mô phỏng

<b>Tham số</b>	<b>Giá trị</b>
Diện tích mô phỏng	1000×1000 (m)
Thời gian mô phỏng	500 (s)
Số Nút	100
MAC Layer	802.11b
Loại lưu lượng	CBR
Lớp giao vận	UDP
Kích thước gói	512 (byte)
Khoảng cách truyền	250 (m)
Tốc độ di chuyển	2 (m/s)
Năng lượng ban đầu của nút	7 (J)
Công suất truyền	1.0 (W)
Công suất nhận	0.2 (W)
Overhearing Power	0.01 (W)
Hopmax	$Hop_{min} + 2$
Loại di động	Random Waypoint

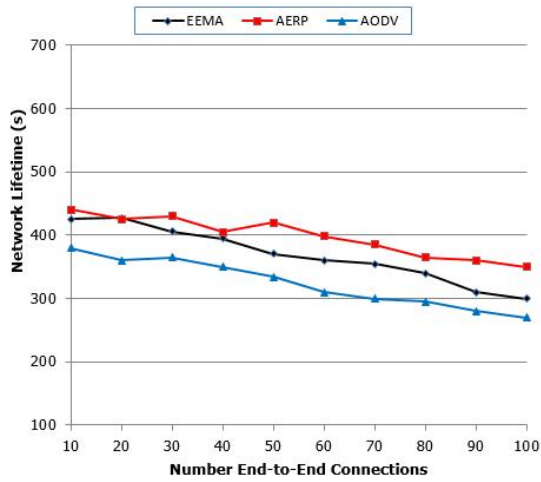
### 2.6.5.3 Kết quả mô phỏng

Hình 2.11 cho thấy hiệu suất mạng theo tiêu chí thời gian duy trì (sống) của mạng. Quan sát kết quả, tác giả thấy rằng thời gian duy trì mạng của AERP luôn được cải thiện tốt hơn giao thức EEMA và AODV trong tất cả các mô phỏng. Bởi vì, các giao thức truyền thống (như AODV) sử dụng phương thức định tuyến dựa trên số bước nhảy, do đó lưu lượng mạng tập trung trên tuyến có số bước nhảy thấp nhất. Đây là lý do chính làm tiêu hao năng lượng nhiều, nhanh hết năng lượng. Ngược lại, AERP sử dụng phương pháp định tuyến dựa trên năng lượng còn lại của nút, do đó tất cả các nút trong mạng có thể cân bằng mức năng lượng tiêu thụ, giúp cho thời gian duy trì mạng của AERP luôn cao hơn giao thức khác.

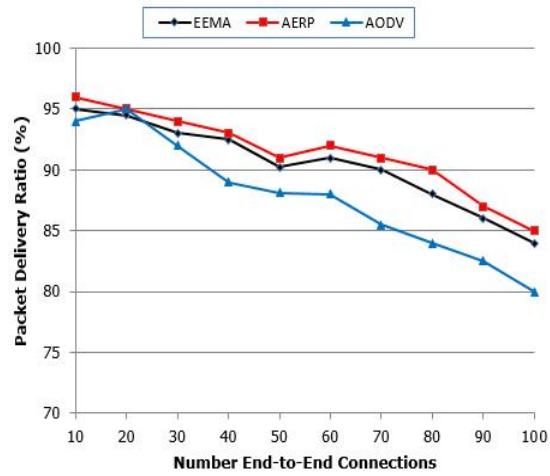
Với mục tiêu tăng thời gian sống của mạng mà vẫn đảm bảo hiệu suất, đặc biệt là theo tiêu chí trễ kết cuối, EEMA thực hiện việc phân chia lưu lượng trên một số tuyến có giới hạn độ trễ (các tuyến có số bước nhảy trong phạm vi  $[hop_{min}, hop_{max}]$ ). Nói cách khác, EEMA không đánh đổi tiêu chí để làm tăng tuổi thọ mạng bằng mọi cách. Do vậy, tuổi thọ mạng khi sử dụng EEMA cao hơn so với AODV nhưng thấp hơn AERP.

Hình 2.12 biểu diễn hiệu suất mạng theo tiêu chí tỷ lệ phân phối bản tin. Kết quả mô phỏng cho thấy, khi lưu lượng mạng thấp (số lượng kết nối đầu cuối thấp), thì tỷ lệ phân phối bản tin

của 3 giao thức là khá cao và không khác nhau nhiều. Khi lưu lượng mạng tăng lên (số lượng kết nối đầu cuối cao), thì tỷ lệ phân phối gói của cả 3 giao thức đều giảm.

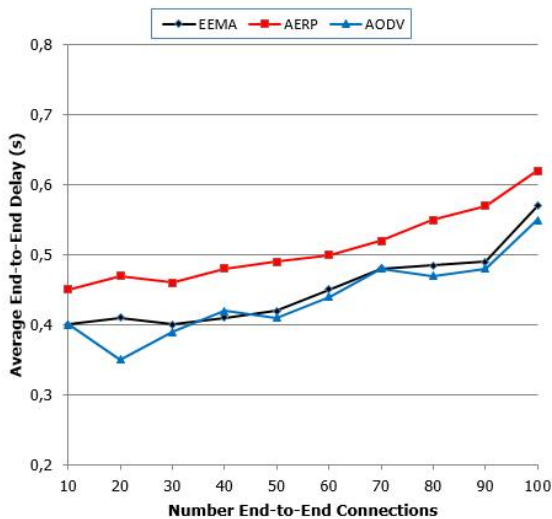


Hình 2.11: Thời gian duy trì mạng

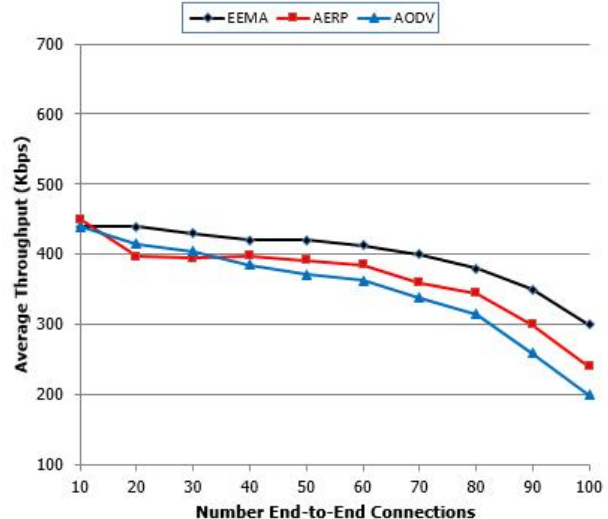


Hình 2.12: Tỷ lệ phân phối bản tin

Với phương pháp định tuyến dựa trên số bước nhảy của AODV, các tuyến ngắn nhất được chọn để truyền dữ liệu. Do đó, một số nút mạng có thể bị quá tải. Hơn nữa, tình trạng xung đột giữa các bản tin tăng lên, cao hơn thời gian thiết lập lại tuyến, đồng thời số bản tin truyền lại càng nhiều hơn. Do vậy, tỷ lệ phân phối bản tin của AODV là thấp nhất so với các giao thức khác. Ngược lại, với phương pháp định tuyến cân bằng động, các tuyến của EEMA và AERP sẽ được chọn dựa trên năng lượng Pin tối thiểu còn lại của tuyến, làm giảm trạng thái xung đột giữa các bản tin, nghĩa là tỷ lệ phân phối bản tin của các giao thức EEMA và AERP được cải thiện. Tuy nhiên, do cơ chế sử dụng giới hạn số bước nhảy để cải thiện hiệu suất mạng, nên tỷ lệ phân phối bản tin của EEMA thấp hơn AERP chút ít.



Hình 2.13: Trung bình trễ kết cuối



Hình 2.14: Trung bình thông lượng

Trong hình 2.13 chỉ ra kết quả mô phỏng dựa trên độ trễ trung bình cho 3 giao thức. Kết quả mô phỏng cho thấy độ trễ trung bình của 3 giao thức tăng lên khi số lượng kết nối đầu cuối tăng, tuy nhiên trễ trung bình của AERP là cao nhất và của EEMA và AODV là xấp xỉ bằng nhau. Lý do là: (1) phương trình tính toán độ trễ trung bình không bao gồm các bản tin truyền lại và (2) AERP luôn chọn tuyến có năng lượng còn lại tối đa để truyền dữ liệu, trong khi AODV chọn tuyến ngắn nhất.

EEMA được đề xuất nhằm cân bằng hai phương pháp kia. Như vậy, trung bình trễ kết cuối của EEMA thấp hơn so với AERP và xấp xỉ bằng AODV.

Hình 2.14 biểu diễn kết quả mô phỏng dựa trên thông lượng trung bình đối với ba giao thức. Kết quả mô phỏng cho thấy, thông lượng trung bình của ba giao thức ngày càng giảm khi số lượng kết nối đầu cuối tăng lên. Tuy nhiên thông lượng trung bình của EEMA được cải thiện tốt hơn so với hai giao thức còn lại. Điều này là phù hợp cho các tính toán lý thuyết của luận án. Với phương pháp định tuyến được đề xuất, EEMA cải thiện tỷ lệ phân phối bản tin, tuổi thọ mạng, và làm giảm tình trạng tắc nghẽn và truyền lại. Do đó, thông lượng trung bình của EEMA được cải thiện hơn các giao thức khác.

## 2.7 Kết luận chương 2

Chương này luận án đã tập trung phân tích các yếu tố ảnh hưởng tới chất lượng dịch vụ khi chuyển giao, mức độ ảnh hưởng của các loại giao thức tới các loại ứng dụng khác nhau, tương quan giữa năng lượng tiêu thụ và hiệu suất mạng, từ đó rút ra được các yếu tố ảnh hưởng tới hiệu suất chuyển giao của một giao thức quản lý di động đó là:

a) *Xác suất thất lạc gói tin số liệu*: các phân tích ở trên đã cho thấy rằng độ trễ chuyển giao, độ trễ truyền gói tin từ điểm đến điểm, và tỉ lệ gói tin bị thất lạc phụ thuộc vào xác suất lỗi khung lớp kết nối ( $p_f$ ) kể cả trong trường hợp có và không sử dụng RLP.

b) *Độ trễ báo hiệu và truyền bản tin kết cuối*: độ trễ chuyển giao và thất lạc gói tin trong quá trình chuyển giao phụ thuộc vào độ trễ báo hiệu (*bằng tổng có trọng số của các độ trễ truyền gói tin với trọng số là xác suất truyền thành công*) giữa các phân tử mạng tham gia vào quá trình chuyển giao. Độ trễ khi truyền bản tin điểm tới điểm từ MH đến HA (hay CH) được tính bằng tổng của toàn bộ độ trễ khi truyền của gói tin trên kết nối vô tuyến từ MH đến BS và độ trễ truyền gói tin trên kết nối hữu tuyến giữa BS và HA (hay CH). Khi không sử dụng RLP, sẽ không phải truyền lại số liệu trên lớp kết nối.

c) *Các công nghệ truy cập lớp kết nối*: như đã xem xét ở phần phân tích phía trên, giá trị các tham số ảnh hưởng đến chuyển giao phụ thuộc vào loại công nghệ truy cập lớp kết nối. Hơn nữa công nghệ truy cập mạng khác nhau nghĩa là độ trễ truy cập lớp kết nối sẽ khác nhau, hiệu suất chuyển giao cũng bị ảnh hưởng.

d) *Loại ứng dụng*: các ứng dụng khác nhau sử dụng các giao thức lớp vận chuyển khác nhau. Các nguyên lý hoạt động cơ bản của các giao thức lớp vận chuyển là khác nhau, nên ảnh hưởng tới chuyển giao cũng khác nhau. Do đó, hiệu suất của một giao thức quản lý chuyển giao cũng sẽ khác đối với từng loại ứng dụng. Ví dụ, như trình bày ở trên, độ trễ chuyển giao sử dụng MIP với giao thức TCP sẽ lớn hơn các ứng dụng dạng này khi sử dụng giao thức UDP, vì khi các gói tin bị thất lạc trong quá trình chuyển giao, TCP sẽ gửi lại các gói tin này dựa vào thông số timeout.

e) *Năng lượng tiêu thụ và hiệu suất mạng*: các MT có nhiều giao diện vô tuyến để kết nối tới các mạng phù hợp nhất trong quá trình định tuyến chuyển giao. Việc dò tìm mạng phù hợp để đáp ứng QoS và hiệu suất mạng sẽ làm tăng mức tiêu thụ năng lượng. Do vậy, cân bằng giữa năng lượng và hiệu suất mạng là tiêu chí ảnh hưởng đến lựa chọn tuyến chuyển giao.

Dựa vào các điều tra; nghiên cứu về hiệu suất chuyển giao; và đề xuất giao thức (EEMA) định tuyến chuyển giao thỏa mãn hai tiêu chí quan trọng là năng lượng tiêu thụ và hiệu suất mạng ở chương này, luận án xây dựng *cơ chế đặt trước băng thông cho các ứng dụng có khả năng chuyển giao* trong chương 3, và *giải pháp lựa chọn các giao thức phù hợp cho các ứng dụng tương ứng và chia sẻ thông tin ở các lớp khác nhau để nâng cao hiệu suất quản lý di động* mà được trình bày trong chương 4. Đó là sử dụng TCP-M cho các ứng dụng sử dụng kết nối TCP, như các ứng dụng lớp B và C. SIP sẽ phù hợp đối với các ứng dụng thời gian thực sử dụng UDP. Tuy nhiên, SIP đã được chuẩn hóa cho các ứng dụng thời gian thực, cho nên MIP

có thể được sử dụng cho các ứng dụng sử dụng giao thức UDP mà không phải là các ứng dụng thời gian thực.

Trong chương này, luận án đã:

- Phân tích và đánh giá hiệu suất quản lý di động cho nhiều loại ứng dụng khác nhau (A,B,C,D,E) của các giao thức đã có như MIP, TCP-M, SIP. Từ đó chứng minh được các định lý liên quan tới việc *thất lạc gói tin* và *độ trễ trung bình*, làm tiền đề xác định các yếu tố cơ bản ảnh hưởng tới QoS khi chuyển giao, đó là: *xác suất gói tin bị thất lạc*; *độ trễ truyền tin*; và *trung bình độ trễ truyền bản tin báo hiệu*.
- Đề xuất giao thức định tuyến theo yêu cầu - EEMA cho MANET. EEMA chọn tuyến tối ưu cho chuyển giao dựa trên: số bước nhảy và hàm chi phí, và cân đối giữa trễ và năng lượng tiêu thụ.

Thể hiện qua 3 công trình sau:

- 1) Lê Ngọc Hưng, Nguyễn Xuân Quỳnh, “Nhận dạng và phân lớp các yếu tố ảnh hưởng tới điều khiển chuyển giao”, Hội nghị FAIR lần 10, tháng 8/2017, Đà Nẵng.
- 2) Vu Khanh Quy, Le Ngoc Hung, Nguyen Dinh Han , “CEPRM: A Cloud-assisted Energy-Saving and Performance-Improving Routing Mechanism for MANETs” JCM 15 Nov. 2019
- 3) Vu Khanh Quy, Le Ngoc Hung, “A Trade-off between Energy Efficient and High-Performance in Routing for Mobile Ad hoc Networks”, JCM Mar. 2020

## Chương 3: Xây dựng phương thức định trước băng thông chuyển giao trong mạng BcN.

### 3.1 Mở đầu

Một trong các vấn đề quan trọng đối với QoS ở mức kết nối đó là làm thế nào để điều khiển hoặc hạn chế được việc rơi các cuộc gọi chuyển giao trong điều kiện băng thông có sẵn bị hạn chế. Luận án phân tích hai tham số liên quan tới chất lượng dịch vụ mức kết nối, đó là: *xác suất khóa các yêu cầu kết nối mới  $P_{CB}$  và xác suất rơi các kết nối chuyển giao  $P_{HD}$* . Bài toán đặt ra là cả nhà cung cấp dịch vụ và người dùng đều mong muốn kết nối sẽ được duy trì liên tục mọi lúc mọi nơi, không phụ thuộc vào quá trình chuyển giao. Như vậy, để thực hiện được yêu cầu này thì phải đặt trước băng thông trong tất cả các tế bào mà thuê bao có thể đi qua, điều này là không thể thực hiện được trong tất cả các trường hợp vì việc di chuyển là không được biết trước. Hơn nữa, nếu đặt trước băng thông trên tất cả các hướng mà thuê bao có thể đi qua sẽ gây nên việc sử dụng băng thông không hiệu quả, dẫn đến băng thông suy giảm nhanh, gây ra xác suất  $P_{CB}$  cao.

Thay vào đó, ta có thể đặt trước một lượng băng thông ở các tế bào liền kề với tế bào đang có kết nối, và phần băng thông này chỉ được sử dụng cho các kết nối chuyển giao, không dùng cho các yêu cầu kết nối mới. Vấn đề đặt ra là đặt trước bao nhiêu băng thông cho đủ. Trong chương này luận án đề xuất cơ chế **dự báo chuyển giao, đặt trước băng thông linh hoạt và điều khiển đăng nhập để đảm bảo xác suất rơi kết nối chuyển giao thấp hơn giá trị mong muốn  $P_{HD,Target}$**  (theo thiết kế). Vì không thể hạn chế hoàn toàn việc rơi các kết nối chuyển giao, nên để đảm bảo chất lượng dịch vụ, luận án đề xuất giữ  $P_{HD}$  bé hơn một giá trị định trước.

Để so sánh và đánh giá kết quả, luận án xét tới 5 cơ chế định trước băng thông hiện có là STATIC, NAG, AG, BHARG và NCBF. Các cơ chế được so sánh với nhau thông qua các tham số (1) *xác suất rơi chuyển giao, xác suất khóa kết nối mới và hiệu suất sử dụng băng thông*; (2) *phụ thuộc vào các tham số thiết kế*; (3) *phụ thuộc vào độ chính xác dự báo khả năng di chuyển*; (4) *tính phức tạp*.

Nội dung chương này bao gồm: Xây dựng tham số hệ thống và các trạng thái giả định được sử dụng; Xây dựng các thuật toán dự báo và đặt trước băng thông, thuật toán dự báo chuyển giao.

### 3.2 Xây dựng mô hình hệ thống

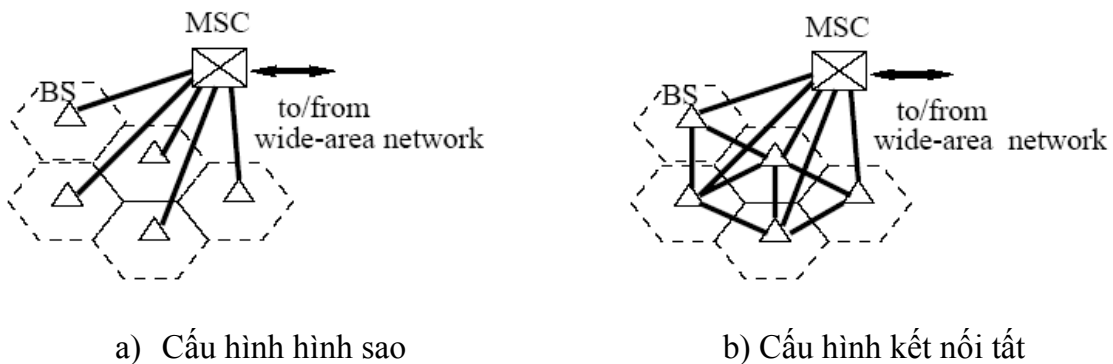
Giả sử các kết nối yêu cầu một lượng băng thông nào đó để sử dụng. Nếu không đủ băng thông thì kết nối chuyển giao sẽ gián đoạn, do vậy quá trình chuyển giao thất bại. Ở đây luận án hạn chế (1) *các ứng dụng không nhạy cảm với trễ* (Ví dụ: Thoại, truyền số liệu,...), nó có

thể chịu được trễ chuyển giao lớn trong trường hợp không đủ băng thông trong tế bào mới tại thời điểm chuyển giao; (2) *chuyển giao mềm trong hệ thống CDMA*, nó cho phép đầu cuối di động giao tiếp đồng thời với 2 trạm gốc BS trong một khoảng thời gian trước khi chuyển giao thực sự diễn ra. Luận án đề xuất dành một lượng băng thông trong mỗi tế bào cho các khả năng chuyển giao từ các tế bào bên cạnh. Băng thông này chỉ được sử dụng cho các chuyển giao từ tế bào bên cạnh và không được sử dụng cho các kết nối mới. Do đó việc điều khiển đăng nhập cho một kết nối mới chỉ được thực hiện khi:

$$\sum_i b_i + b_{new} \leq C - B_r \quad (3.1)$$

Trong đó  $C$  là dung lượng kết nối vô tuyến,  $B_r$  là băng thông định sẵn cho chuyển giao,  $b_i$  là băng thông đang được kết nối thứ  $i$  sử dụng và  $b_{new}$  là băng thông do kết nối mới yêu cầu. Khi có một yêu cầu kết nối mới,  $B_r$  được cập nhật một cách có dự báo và thích nghi trước khi thực hiện việc kiểm tra đăng nhập và phụ thuộc vào trạng thái lưu lượng trong các tế bào bên cạnh. Chú ý rằng  $B_r$  là mục tiêu, chứ không phải là băng thông định trước thực tế, vì một tế bào có thể không có khả năng định trước băng thông mục tiêu, ví dụ  $\sum_i b_i + b_r > C$   $\sum_i b_i + B_r > C$ . Điều này có thể xảy ra vì một BS có thể điều khiển đăng nhập chỉ cho các kết nối mới, không cho các kết nối chuyển giao từ các tế bào bên cạnh.

### 3.2.1 Kiến trúc trao đổi tin

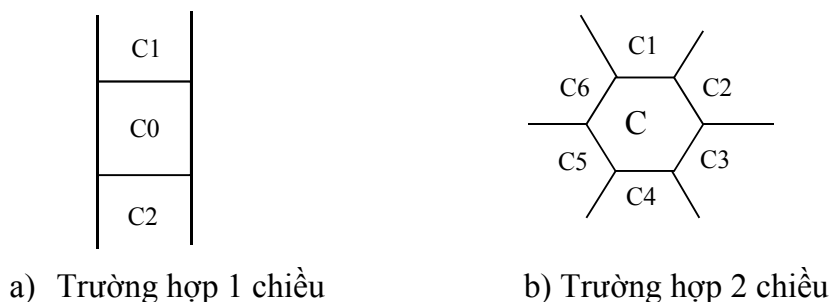


Hình 3.1: Giao tiếp giữa MSC và BS

Cơ chế định trước băng thông đề xuất ở đây dựa trên thông tin từ các tế bào bên cạnh bao gồm số các kết nối đang diễn ra và yêu cầu băng thông của chúng. Do đó, việc duy trì trao đổi thông tin giữa các BS là rất quan trọng. Kiến trúc mạng của BS có thể có hai cấu hình như trong hình 3.1. Một nút mạng MSC điều khiển một số BS, đóng vai trò cổng giao tiếp với mạng ngoài. Hình 3.1 (a) biểu diễn kiến trúc kết nối hình sao giữa MSC và BS, trong đó không có kết nối trực tiếp giữa các BS. Đây là kiến trúc điển hình trong các mạng tế bào hiện tại. Trong môi trường này, mỗi BS truyền thông tin về các kết nối hiện tại trong tế bào của nó tới MSC. Sau đó MSC sẽ xác định băng thông định trước mục tiêu trong mỗi tế bào, và tương ứng

sẽ tiến hành kiểm tra đăng nhập cho mỗi kết nối mới được yêu cầu trong vùng phủ sóng của tế bào. Mặt khác, hình 3.1 (b) biểu diễn trường hợp tất cả các BS được kết nối đầy đủ. Trong kiến trúc này, BS có thể giao tiếp trực tiếp với nhau không cần qua MSC, và mỗi BS có thể tự xác định băng thông định trước đích, và do đó nó thực hiện kiểm tra đăng nhập cho mỗi kết nối mới được yêu cầu trong tế bào của nó.

### 3.2.2 Đánh số tế bào



Hình 3.2: Đánh số tế bào

Tất cả các tế bào bên cạnh tế bào A được đánh số: A là số 0 và các tế bào khác được đánh số bắt đầu từ số 1 như hình 3.2. Gọi  $C_{i,j}$  là kết nối j trong tế bào i và  $b(C_{i,j})$  là băng thông yêu cầu của nó. Để đơn giản, giả sử một đầu cuối di động không thể có nhiều kết nối đồng thời, do đó một đầu cuối di động hoạt động có nghĩa là nó có một kết nối. Hệ thống thông tin tế bào sử dụng cơ chế cấp phát kênh định sẵn, và tế bào i có dung lượng liên kết vô tuyến là  $C(i)$ . Đơn vị băng thông là BU, đó là băng thông được yêu cầu cho một kết nối thoại. Một kết nối sẽ phải trải qua nhiều liên kết hữu tuyến và vô tuyến, do đó ta phải xem xét băng thông định trước trên cả hai liên kết vô tuyến và hữu tuyến cho chuyển giao. Tuy nhiên, luận án chỉ hạn chế việc định trước băng thông cho liên kết vô tuyến trong mỗi tế bào bởi vì trọng tâm của phần này là định tuyến và định tuyến lại một kết nối chuyển giao.

### 3.2.3 Mô hình hoá di chuyển

Luận án mô hình hóa các đáp ứng của chuyển giao một cách gần đúng và ước tính việc di động của chúng dựa trên tập hợp các chuyển giao đã diễn ra trong mỗi tế bào. Để hiểu được cơ sở của việc ước tính di chuyển, chúng tôi nghiên cứu lưu lượng giao thông thông thường qua ví dụ:

- O1. Có các hạn chế tốc độ trên hầu hết các con đường, và tốc độ di chuyển thường không nhỏ hơn hoặc lớn hơn tốc độ hạn chế
- O2. Các tuyến trong phố, tín hiệu giao thông ảnh hưởng quan trọng tới việc di chuyển
- O3. Trong giờ cao điểm, tốc độ của các di chuyển trong một khu vực nhất định là tương quan gần gũi với nhau.



O4. Trong nhiều trường hợp, hướng di chuyển có thể được dự báo thông qua đường di chuyển trước đó.

Từ những quan sát trên, chúng tôi mong muốn việc đáp ứng chuyển giao cho kết nối sẽ tương tự như các kết nối xuất hiện trong cùng một tế bào trước đó, và bây giờ lưu trú trong tế bào hiện tại. Do đó, ta có thể dự báo tế bào tiếp theo của kết nối và ước tính thời điểm chuyển giao của nó bởi các giám sát trong mỗi tế bào. Thậm chí những quan sát trên được thực hiện từ lưu lượng giao thông, phương pháp tương tự có thể được thực hiện cho người đi bộ bởi vì tốc độ đi bộ sẽ không khác nhau nhiều. Trong mạng tế bào ngoài trời, sẽ có thuê bao đi bộ hoặc đi xe trong khi đối với trường hợp trong nhà thì phần lớn là đi bộ hoặc không di chuyển.

Nói chung, vị trí và hướng di chuyển của đầu cuối di động là không xác định đối với mạng hữu tuyến (hoặc BS). Đây là trường hợp thông thường khi thuê bao di động giao tiếp thông qua máy cầm tay hoặc máy tính xách tay. Mặt khác, cũng có thể là một trường hợp đặc biệt khi mạng hữu tuyến biết hướng và đường di chuyển, ví dụ, tế bào mà di động sẽ đi qua. Khi một thuê bao di động có một hệ thống định vị, nó định vị đầu cuối di động tới một điểm cụ thể sử dụng GPS và ITS, và có một giao diện giao tiếp giữa hệ thống định vị và mạng hữu tuyến, thông tin đường di chuyển có thể có sẵn cho mạng hữu tuyến. Khi thông tin đường đi có sẵn cho mạng hữu tuyến, cơ chế này sẽ hoạt động tốt hơn trường hợp không có thông tin đường đi.

Một khả năng khác là sử dụng các quá trình cũ cụ thể của di động được đưa ra trong [56]. Đó là mỗi chuyển động của di động cụ thể được giám sát liên tục, sau đó hướng di chuyển trong một tế bào cụ thể có thể được dự báo bằng cách sử dụng các giám sát này. Tuy nhiên, giám sát di chuyển liên tục rất tốn kém, và trong nhiều trường hợp, các quá trình di chuyển cũ cụ thể thường không đủ chính xác để thực hiện việc dự báo tốt. Do đó luận án hạn chế các thông tin như vậy.

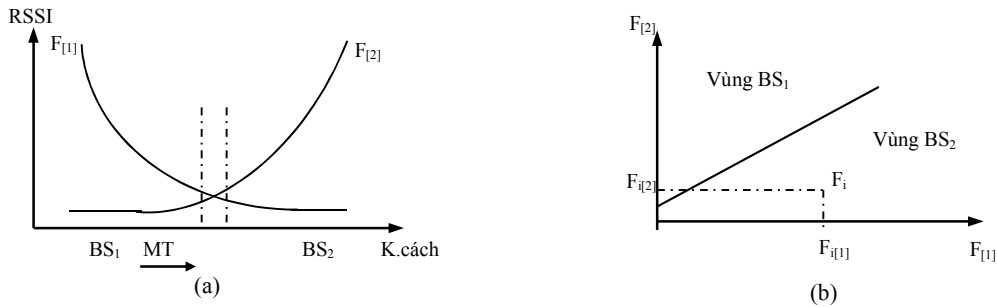
### 3.2.4 Phân lớp các yếu tố ảnh hưởng đến chuyển giao

Nếu quyết định chuyển vùng chỉ dựa trên cường độ tín hiệu thu được (Received Signal Strength Intensity - RSSI), thì trường hợp tín hiệu của trạm phát hiện tại hoặc lân cận yếu sẽ gây ra tình trạng chuyển vùng không cần thiết. Trong nhiều trường hợp, rất khó xác định vị trí chính xác của thuê bao, vì vậy thông qua kỹ thuật nhận dạng phân tích giá trị cường độ tín hiệu thu được góp phần làm giảm những quyết định chuyển vùng không chính xác (tăng tính tối ưu).

Mỗi MT đại diện bởi một véc tơ đặc trưng  $F_i = [F_{i[1]}, F_{i[2]}, \dots, F_{i[k]}, \dots, F_{i[c-1]}, F_{i[c]}]$  thuộc vùng trong không gian  $c$  chiều. Mỗi véc tơ thành phần  $F_{i[k]}$  biểu diễn cường độ tín hiệu thu được từ trạm BTS thứ  $k$  của MT thứ  $i$ , do vậy  $c$  biểu diễn số trạm BTS có thể cung cấp dịch vụ cho  $MT_i$ .

Quá trình chuyển vùng tương đương với việc xác định vị trí của véc tơ đặc trưng trong miền định trước, điều này có thể thực hiện thông qua véc tơ thành phần gán cho MT thứ  $i$ . MT được gán cho trạm phát thứ  $k$  nếu quan hệ thành phần của nó với cell  $k$  lớn hơn so với các cell khác, nghĩa là:

$$U_{i[k]} = \max U_{i[j]}, \text{ trong đó } j = 1, 2, 3, \dots, c \quad (3.2)$$



Hình 3.3 Mô hình Handoff sử dụng Lý thuyết mờ

Ví dụ, xét trường hợp có hai trạm phát  $BS_1$  và  $BS_2$  như Hình 3.3a.  $MT_i$  di chuyển từ trạm  $BS_1$  sang trạm  $BS_2$ , nhận được cường độ tín hiệu :

$$F_{[i]} = [F_{i[1]}, F_{i[2]}] \quad (3.3)$$

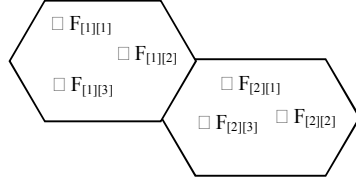
trong đó  $F_{i[1]}$  của  $BS_1$  và  $F_{i[2]}$  của  $BS_2$ . Theo hướng di chuyển của  $MT_i$ , thì  $F_{i[1]}$  giảm dần còn  $F_{i[2]}$  tăng dần. Véc tơ hai chiều  $F_{[i]}$  biểu diễn cường độ tín hiệu thu được (hình 3.3b) từ hai vùng dịch vụ do  $BS_1$  và  $BS_2$  cung cấp.

#### 3.2.4.1 Thuật toán phân lớp Fuzzy

Các thuật toán phân lớp hỗ trợ ra quyết định ứng dụng lý thuyết mờ thường dựa trên cơ chế thu nạp thông tin trước (learning), nghĩa là các điều kiện đầu vào, ngưỡng ra quyết định đều không cố định, mà thay đổi và điều chỉnh theo số liệu thống kê và sai số cho phép của các thông số liên quan tới quyết định chuyển vùng như RSSI, BER, Khoảng cách, SLA,.... Để tăng hiệu suất mạng và giảm chi phí chuyển vùng không cần thiết, nhất là đối với mạng di động băng rộng hỗn hợp, thì áp dụng cơ chế learning ra quyết định chuyển vùng dựa trên tri thức động đang được quan tâm rộng rãi. Trước hết, cung cấp tập véc tơ training  $F_{k(m)}$  phân bố trên miền dịch vụ của hệ thống. Các véc tơ  $c$  chiều là đại diện cho các MT trong mỗi cell. Như vậy, mỗi véc tơ  $F_{k(m)}$  sẽ được biểu diễn bởi:

$$F_{k(m)} = [F_{k(m)[1]}, F_{k(m)[2]}, F_{k(m)[3]}, \dots, F_{k(m)[c]}] \quad (3.4)$$

Phần tử thứ  $n$   $F_{k(m)[n]}$  của véc tơ training thứ  $m$  tương ứng với RSSI thu được từ trạm BTS  $n$ , đo được tại vị trí  $m$  của cell  $k$ . Giá trị khởi đầu của số lượng véc tơ training tại cell  $k$  là  $t_k$ . Hình 3.4 minh họa 3 véc tơ training được phân bố đều trên 2 cell, nghĩa là  $t_1 = t_2 = 3$ .



Hình 3.4 Vùng dịch vụ của 2 cell mỗi cell 3 véc tơ

Ta xem như  $F_{k(m)[n]}$  là các phần tử ngẫu nhiên, dung sai là  $\sigma_{k[n]}^2$  và trị trung bình là  $\overline{F_{k[n]}}$ , trong đó:

$$\overline{F_{k[n]}} = \frac{1}{t_k} \sum_{m=1}^{t_k} F_{k(m)[n]} \quad (3.5)$$

$$\text{Và } \sigma_{k[n]}^2 = \frac{1}{t_k} \sum_{m=1}^{t_k} (F_{k(m)[n]})^2 - \overline{F_{k[n]}}^2 \quad (3.6)$$

Thuật toán trình bày ở đây có thể biểu diễn ở 2 dạng “Thích ứng – ADT” và “Không thích ứng – NADT” tùy thuộc vào véc tơ training có được cập nhật hay không cập nhật.

Ở dạng Thích ứng, nếu có một mẫu mới  $F_i$  thuộc lớp  $k$  nêu trên thì ta cập nhật thông tin tương ứng với lớp này. Công thức sau dùng để cập nhật các thành phần của véc tơ trung bình [9]:

$$\overline{F_{k[n]}} = \frac{t_k}{t_k + 1} \overline{F_{k[n]}} + \frac{1}{t_k + 1} F_{i[n]} \quad (3.7)$$

Bổ sung biến phụ  $C_{k[n]}$  để cập nhật thành phần  $\sigma_{k[n]}^2$  của dung sai  $\sigma_k^2$ .  $C_{k[n]}$  được xác định theo công thức sau:

$$C_{k[n]} = \sum_{m=1}^{t_k} (F_{k(m)[n]})^2 \quad (3.8)$$

Mỗi khi một MT được chuyển sang sử dụng dịch vụ ở BS mới,  $C_{k[n]}$  sẽ được cập nhật bởi:

$$C_{k[n]} = C_{k[n]} + (F_{i[n]})^2 \quad (3.9)$$

Thay 3.8 vào 3.6 ta có giá trị mới của dung sai sẽ là:

$$\sigma_{k[n]}^2 = \frac{1}{t_k + 1} C_{k[n]} - \overline{F_{k[n]}}^2 \quad (3.10)$$

$F_i$  có thể xem như là véc tơ mẫu thuộc cell  $k$ , nghĩa là mỗi khi một MT đi vào vùng dịch vụ của cell  $k$  thì:

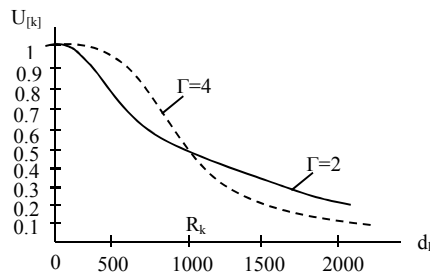
$$t_k = t_{k+1} \quad (3.11)$$

### 3.2.4.2 Phân lớp dựa trên khoảng cách

Ở phương pháp này, hàm thuộc  $U_{i[k]}$  của MT do  $F_i$  đại diện, có thể biểu diễn như sau:

$$U_{i[k]} = [1 + (d_{ik} / \xi)^\Gamma]^{-1} \quad (3.12)$$

Ở đây  $d_{ik}$  biểu diễn khoảng cách từ véc tơ  $F_i$  tới véc tơ đại diện của Cell  $k$  là  $\hat{F}_k$ .



Hình 3.5 Hàm thuộc của BTS thứ  $k$  xác định theo khoảng cách vật lý

Đồ thị hàm thuộc tính theo (3.12) tùy thuộc vào các hằng số (lớn hơn 1) là  $\xi$  và  $\Gamma$ . Hình 3.5 chỉ ra đồ thị hàm thuộc với  $d_{ik}$  là khoảng cách giữa MT và BTS [18]. Ở đây,  $\xi$  chính là bán kính vùng phủ sóng.

Các tham số cường độ tín hiệu RSSI và khoảng cách vật lý  $d$  thường được sử dụng trong quá trình ra quyết định chuyển vùng, tuy nhiên RSSI dễ đo lường hơn việc xác định  $d$ . Khoảng cách  $d_{ik}$  được xác định theo công thức sau:

$$d_{ik} = \left[ \sum_{n=1}^c \frac{1}{S_{k[n]}} (F_{i[n]} - \hat{F}_{k[n]})^2 \right]^{0.5} \quad (3.13)$$

$\hat{F}_k$  (Véc tơ đại diện cho cell  $k$ ) có thể bằng  $\overline{F}_k$  hoặc  $V_k$  (Véc tơ trung tâm của cell  $k$ ).  $V_k = [V_{k[1]}, V_{k[2]}, \dots, V_{k[j]}, \dots]$ , trong đó  $V_{k[j]}$  là cường độ tín hiệu thu được từ cell  $k$ .

### 3.2.4.3 Phép đo khoảng cách

Ở kỹ thuật nhận dạng dựa trên khoảng cách, các phép đo oclit được dùng để ước lượng tính đồng dạng giữa mẫu và không gian đặc trưng. Mỗi phần tử véc tơ số liệu biểu diễn đặc tính của mẫu, và có thể có các đơn vị vật lý để phân biệt với các phần tử khác [68].

Thông thường, các mẫu cùng lớp có tính đồng dạng cao hơn so với mẫu khác lớp. Tính đồng dạng giữa các véc tơ đặc trưng có thể xem là phép đo khoảng cách:  $d: \mathbb{R}^n \times \mathbb{R}^n \rightarrow [0, \infty)$ . Do vậy, các mẫu đồng được xem là các mẫu đồng dạng. Khoảng cách  $d(F_i, F_j)$  cần đáp ứng các đặc tính sau [9]:

$$d(F_i, F_j) = 0$$

$$d(F_i, F_j) = d(F_j, F_i)$$

$$d(F_i, F_j) < d(F_i, F_k) + d(F_k, F_j)$$

Dạng chung của khoảng cách  $d()$  là phép đo  $l^p$  như sau:

$$l^p(F_i, F_j) = \left( \sum_{n=1}^c |F_{i[n]} - F_{j[n]}|^p \right)^{1/p}, p=1 \quad (3.14)$$

Khi  $p=2$  thì phép đo sẽ là khoảng cách ơclit.

Một phép đo khoảng cách khác là Mahanalobis [9].

$$d(F_i, F_j) = (F_i - F_j) A_k^{-1} (F_i - F_j)^T \quad (3.15)$$

$A_k$  là ma trận nguyên dương, xác định như sau:

$$A_k = E \left[ (F_i - \bar{F})^T (F_i - \bar{F}) \right] \quad (3.16)$$

$\bar{F}$  là véc tơ trung bình của lớp  $k$ . Khoảng cách tính theo (3.15) liên quan với Log tự nhiên của hàm mật độ (pdf) Gaussian  $n$  chiều.

$$p(F) = \frac{1}{(2\pi)^{n/2} (\det A_k)^{1/2}} \exp \left( -\frac{1}{2} (F_i - \bar{F})^T A_k^{-1} (F_i - \bar{F}) \right) \quad (3.17)$$

Phép đo khoảng cách này rất tiện lợi trong việc phân lớp các mẫu có phân phối Gaussian [18].

#### 3.2.4.4 Các kết quả số

Chọn mô hình truyền tín hiệu radio Log-Normal để mô phỏng, cường độ tín hiệu từ trạm  $k$  được xác định [69]:

$$F_k = K_1 - K_2 \log(d) + \delta \quad (3.18)$$

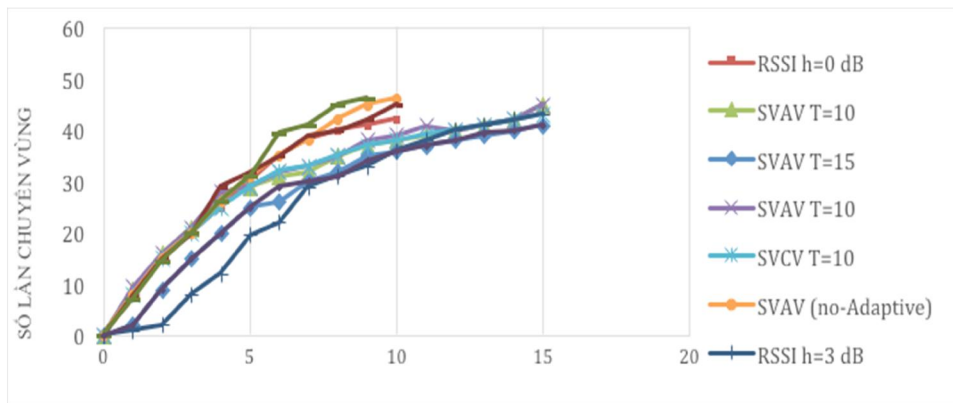
$d$  - Khoảng cách từ MT tới trạm phát thứ  $k$ ;

$K_1$  và  $K_2$  - Hằng số môi trường truyền tín hiệu;

$\delta$  - Biến ngẫu nhiên biểu diễn Fading

Giá trị  $K_1$  và  $K_2$  đối với Micro Cell trong môi trường thành thị là 120 và 35 [69]. Mô hình này dễ dàng nội suy ra mô hình với kích thước cell khác nhau bằng cách thay đổi giá trị  $K_1$  và  $K_2$  và sai số ( $d$ ) của  $\partial$ . Hiệu suất của thuật toán dựa trên sai số ( $d$ ), trong khoảng 0 – 10dB. Giá định, số trạm phát là 3 với bán kính phủ sóng là 0,5 KM, 1 KM và 1,5 KM, cường độ tín hiệu ở biên của Cell là 15dB.

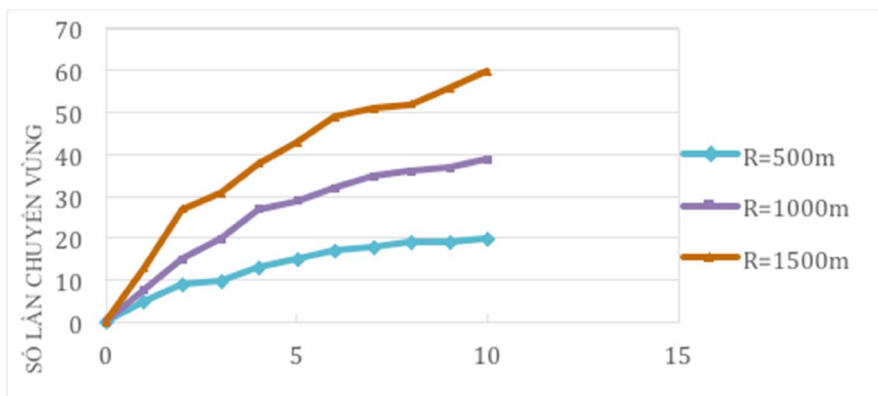
Việc đánh giá thuật toán được thực hiện ngẫu nhiên cho 1000 kết nối. Bảng thông của kết nối là tham số ngẫu nhiên để mô phỏng các cuộc gọi với thời gian gọi khác nhau. Số lần chuyển vùng dựa trên thuật toán phân lớp mờ sẽ được so sánh với phương pháp dựa trên RSSI sử dụng mức trễ  $h$ .



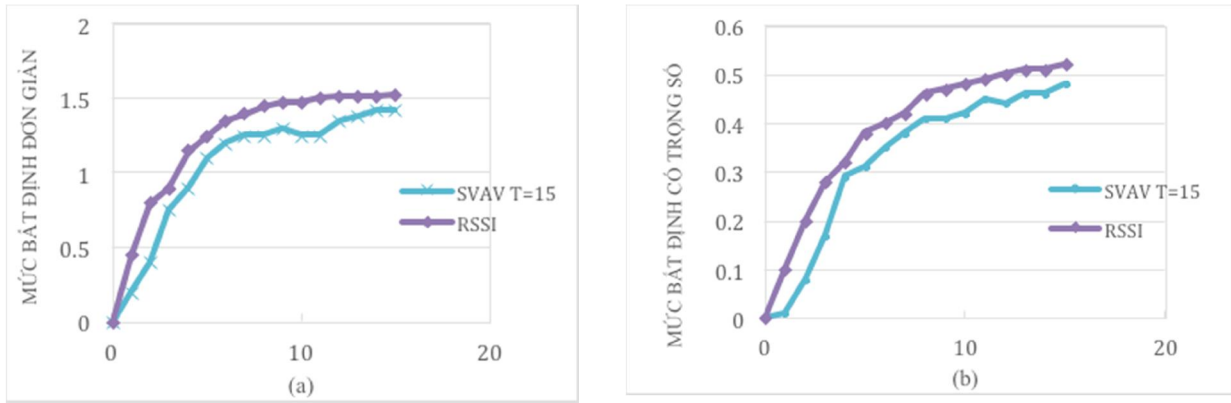
Hình 3.6 Số lần chuyển vùng trung bình ứng với các điều kiện khác nhau

Hình 3.6 chỉ ra số lần chuyển vùng trung bình của MT theo phương pháp ADT với ngưỡng là 10 và 15dB. Khi dùng thuật toán SVAV và SVCV thì kết quả tốt nhất đạt được khi đặt  $T = 15$ .

Hình 3.7 chỉ ra số lần chuyển vùng trung bình của MT đối với hệ thống có bán kính phủ sóng là 0,5 KM, 1 KM và 1,5 KM sử dụng thuật toán ADT SVAV với  $T = 15$ . Chúng ta thấy bán kính Cell càng bé thì số lần chuyển vùng càng ít.



Hình 3.7 Số lần chuyển vùng trung bình khi sử dụng SVAV với  $T=10$

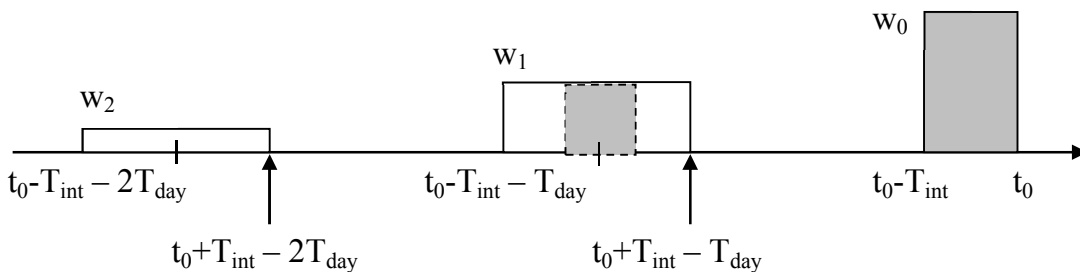


Hình 3.8 Thuật toán RSSI và SVAV đối với (a) Bất định đơn giản, và (b) Bất định có trọng số

Hình 3.8 chỉ ra tính bất định đơn giản và có trọng số khi sử dụng thuật toán phân lớp đồng dạng theo Véc tơ trung bình SVAV ( $T = 15$ ) và thuật toán dựa trên cường độ tín hiệu RSSI với  $h = 3\text{dB}$ . Kết quả mô phỏng cho thấy, thuật toán SVAV cho kết quả tốt hơn RSSI trong cả hai hướng tiếp cận của phép đo bất định.

### 3.2.5 Tính toán xác suất chuyển giao

Phần này, tác giả phát triển cơ chế ước tính và dự báo di chuyển. Cơ chế này được thực hiện ở BS của mỗi tế bào theo phương thức phân tán. Mỗi một đầu cuối khi di chuyển từ tế bào 0 hiện tại sang một tế bào bên cạnh, BS của tế bào 0 sẽ ghi lại bộ tham số của đầu cuối ( $T_{\text{event, prev, next, } T_{\text{soj}}}$ ), được gọi là bộ tham số chuyển giao, trong đó  $T_{\text{event}}$  là thời điểm đầu cuối xuất phát từ tế bào hiện tại, prev là chỉ số của tế bào trước đó mà đầu cuối đã cư trú trước khi vào tế bào hiện tại, next là chỉ số của tế bào mà đầu cuối chuyển sang sau khi xuất phát từ tế bào hiện tại,  $T_{\text{soj}}$  là thời gian tạm trú của đầu cuối trong tế bào hiện tại, ví dụ quãng thời gian giữa lúc vào và lúc đi khỏi tế bào hiện tại. Chú ý, nếu prev=0 có nghĩa là đầu cuối bắt đầu kết nối trong tế bào hiện tại.



Hình 3.9: Các cửa sổ thời gian thu được các hàm ước tính chuyển giao với  $N_{\text{win\_days}}=2$

Từ các tham số ghi lại, BS xây dựng hàm tạm tính chuyển giao, nó mô tả phân bố tạm tính của tế bào tiếp theo và thời gian tạm trú của đầu cuối, phụ thuộc vào tế bào mà đầu cuối cư trú

trước đó. Thời gian cư trú của MT phụ thuộc vào thời điểm trong ngày (cao điểm, bình thường, thấp điểm, ban ngày, buổi tối,...). Hàm ước tính chuyển giao tại thời điểm  $t_0$  với bộ tham số  $(T_{event}, prev, next, T_{soj})$  được biểu diễn như sau:

$$t_0 - T_{int} - nT_{day} \leq T_{event} < t_0 + T_{int} - nT_{day} \quad (3.19)$$

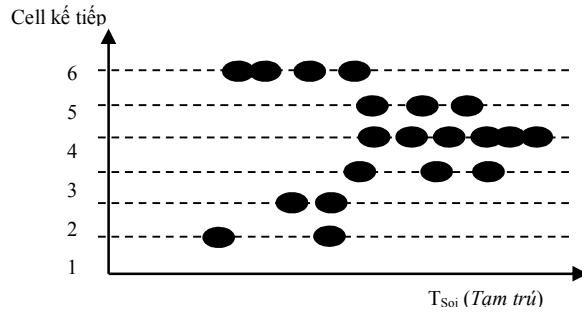
trong đó  $T_{int}$  khoảng thời gian ước tính, đây là tham số thiết kế,  $T_{day}$  là khoảng thời gian trong ngày, ví dụ 24 giờ,  $n \geq 0$  là số nguyên

$$F_{HOE}(T_0, prev, next, T_{soj}) := w_n \quad (3.20)$$

Trong đó  $1 \geq w_n \geq w_{n+1}$  và  $w_0 = 0$  đối với tất cả  $n > N_{win\_days}$ . Trọng số  $w_n$  thu được từ thực tế điều kiện lưu lượng trong tế bào trong một khoảng thời gian cụ thể nào đó của ngày và có thể thay đổi theo thời gian.  $N_{win\_days}$  là tham số thiết kế do đó bộ tham số chuyển giao (4 tham số) được giám sát trước ( $N_{win\_days}, T_{day} + T_{int}$ ) được xác định là quá hạn, và sẽ không được sử dụng trong hàm ước tính chuyển giao. Có thể dễ dàng nhận ra rằng các hàm ước tính chuyển giao bị ảnh hưởng bởi các tham số chuyển giao trong cửa sổ thời gian  $2T_{int}$  như trong hình 3.9. Chú ý rằng khoảng thời gian  $[t_0, t_0 + T_{int}]$  không có trong hình vì nó biểu diễn khoảng thời gian trong tương lai, nó không có nhiều ý nghĩa trong định nghĩa bộ tham số chuyển giao.

Thực tế mong muốn hạn chế số các bộ tham số (1) *sử dụng cho hàm ước tính chuyển giao* và (2) *hiện tại không sử dụng cho hàm ước tính chuyển giao nhưng được ghi lại để sử dụng trong tương lai*, ví dụ các tham số trong khoảng  $t_0 + T_{int} - T_{day} < T_{event} < t_0 - T_{int}$   $t_0 + T_{int} - T_{day} < T_{event} < t_0 - T_{int}$  trong hình 3.9 để giảm bộ nhớ và sự phức tạp tính toán. Luận án định nghĩa kích thước hàm tạm tính chuyển giao lớn nhất  $N_{quad}$  là số lớn nhất các tham số chuyển giao được sử dụng cho hàm ước tính chuyển giao cho mỗi prev. Điều này có nghĩa là không cần các bộ tham số từ các ngày trước nếu ta đã quan sát đủ trong khoảng thời gian  $T_{int}$  cuối cùng. Tới tận  $N_{quad}$  bộ tham số được ghi lại được sử dụng cho hàm với luật ưu tiên sau. Đầu tiên, bộ tham số thỏa mãn công thức (3.19) với  $n$  nhỏ hơn sẽ được ưu tiên cao hơn. Thứ hai, trong số các bộ tham số thỏa mãn công thức (3.19) với  $n$  giống nhau, bộ tham số nào có  $|T_{event} - nT_{day}|$  nhỏ hơn sẽ được ưu tiên cao hơn. Hình 3.9 biểu diễn một ví dụ chỉ có các bộ tham số với các thời điểm  $T_{event}$  trong vùng được bôi đen là được sử dụng cho hàm tương ứng với luật ưu tiên, có nghĩa là tổng số các bộ tham số trong vùng là  $N_{quad}$ . Để giảm kích thước bộ nhớ ghi lại (cached), các bộ tham số được giám sát tại thời điểm  $t'$ , ví dụ  $T_{event} = t'$ , khi hàm ước tính tại thời điểm  $t'$  không sử dụng bất cứ bộ tham số nào được quan sát trong các ngày trước đó không được ghi vào bộ nhớ để sử dụng trong tương lai, bởi vì có vẻ nó sẽ không được dùng cho hàm ước tính ngày hôm sau. Chú ý rằng các bộ tham số (1) *trong khoảng*  $T_{event} < t_0 - T_{int} - N_{win\_days} T_{day}$  và (2) *không được sử dụng cho hàm ước tính trong khoảng cuối*  $(T_{day} + T_{int})$  có thể được xóa khỏi bộ nhớ ghi lại.





Hình 3.10: Ví dụ về đồ thị của hàm ước tính chuyển giao với  $prev=1$

Có các loại khác của mẫu tuần hoàn và không tuần hoàn để nghiên cứu ước tính di động. Chúng dùng để giám sát trong các ngày cuối tuần và ngày nghỉ, và các mẫu di động sẽ khác rất nhiều so với các ngày trong tuần. Do đó một bộ các tham số sẽ được ghi lại cho những ngày đặc biệt này, và hàm ước tính chuyển giao cho cuối tuần, ví dụ sẽ được xây dựng bằng cách sử dụng công thức (3.19) và (3.20) bằng cách thay  $T_{day}$  và  $N_{win\ days}$  với  $T_{week}=7$  và  $N_{win\_weeks}$ . Hình 3.10 biểu diễn các điểm của hàm ước tính chuyển giao với  $prev=1$  không biểu diễn giá trị của  $w_n$ . Số các điểm trong đồ thị, mỗi điểm tương ứng với một bộ tham số chuyển giao được ghi lại, được giới hạn bởi  $N_{quad}$ . Đối với hàm ước tính chuyển giao trong không gian 3 chiều, hàm biểu diễn  $n \times n$  có các độ cao khác nhau, phụ thuộc vào giá trị của  $w_n$ . Hình 3.2 (b) biểu diễn một ví dụ được vẽ từ cùng một chỉ số. Từ hình vẽ ta thấy tế bào 4 là tế bào xa nhất từ tế bào 1 (tức là tế bào trước đây) qua tế bào 0 (nghĩa là tế bào hiện tại) trong số các tế bào bên cạnh tế bào 0 vì các thời gian cư trú trước khi vào tế bào 4 được chỉ ra nằm trong số lớn nhất. Chú ý rằng hàm ước tính chuyển giao đối với một giá trị cụ thể của  $prev$  có thể tạo ra một hàm xác suất tổng quát cho vecto ngẫu nhiên hai chiều ( $next, T_{soj}$ ), trong đó  $next$  là tế bào dự báo tiếp theo và  $T_{soj}$  là thời gian tạm trú ước tính trong tế bào hiện tại. Phần sau sẽ biểu diễn hàm ước tính chuyển giao được sử dụng để ước tính di chuyển của người sử dụng.

### 3.3 Thuật toán định trước băng thông và điều khiển đăng nhập

Dưới đây luận án mô tả phương pháp định trước băng thông và điều khiển đăng nhập để giữ xác suất rơi chuyển giao  $P_{HD}$  thấp hơn  $P_{HD,target}$  bằng cách sử dụng hàm ước tính chuyển giao được trình bày ở trên.

#### 3.3.1 Định trước băng thông không thông tin di chuyển

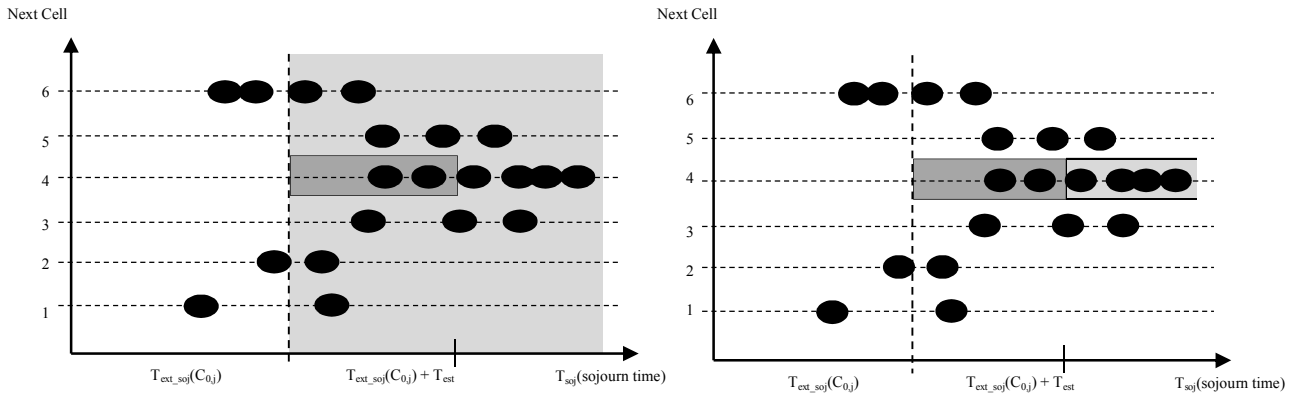
Cơ chế được nghiên cứu ở đây là trường hợp tổng quát khi mạng hữu tuyến (hoặc BS) không có thông tin về đường đi của đầu cuối di động. Trường hợp đặc biệt khi BS biết thông tin về đường đi sẽ được nghiên cứu ở phần tiếp theo. Cách tiếp cận của luận án dựa trên ước tính di động trong cửa sổ thời gian  $[t_0, t_0+T_{est}]$ , trong đó  $t_0$  là thời điểm hiện tại. Di chuyển của đầu

cuối di động với kết nối  $C_{0,j}$  được ước tính với  $p_h(C_{0,j} \rightarrow i)$ , xác suất có  $C_{0,j}$  chuyển giao sang tế bào  $i$  trong  $T_{est}$ .

Xác suất chuyển giao có thể được tính toán bằng cách sử dụng hàm ước tính chuyển giao như sau. BS của tế bào giám sát đầu cuối di động hoạt động trong tế bào của nó thông qua thời gian cư trú hiện tại. Thời gian cư trú hiện tại  $T_{ext\_soj}(C_{0,j})$  của kết nối  $C_{0,j}$  là thời gian trôi qua khi đầu cuối di động hoạt động với kết nối  $C_{0,j}$  di chuyển vào tế bào hiện tại. Sử dụng lý thuyết Bayes [74], xác suất  $p_h(C_{0,j} \rightarrow next)$  tại thời điểm  $t_0$  được tính:

$$p_h(c_{0,j} \rightarrow next) := \frac{\sum_{T_{ext\_soj}(c_{0,j}) < t_{soj} \leq T_{ext\_soj}(c_{0,j}) + T_{ext}} F_{HOE}(t_0, prev(c_{0,j}), next, t_{soj})}{\sum_{next' \in A_0} \sum_{t_{soj} > T_{ext\_soj}(c_{0,j})} F_{HOE}(t_0, prev(c_{0,j}), next', t_{soj})} \quad (3.21)$$

trong đó  $prev(c_{0,j})$   $prev(C_{0,j})$  là tế bào có  $c_{0,j}$   $C_{0,j}$  cư trú trong đó trước khi di chuyển vào tế bào hiện tại và  $A_i$  là tập các chỉ số của các tế bào bên cạnh của tế bào  $i$ . Công thức biểu diễn xác suất mong muốn kết nối  $C_{0,j}$  chuyển giao vào tế bào  $next$  với thời gian cư trú  $t_{soj}$  nhỏ hơn hoặc bằng  $T_{ext\_soj}(C_{0,j}) + T_{ext}$  đưa ra điều kiện  $t_{soj} > T_{ext\_soj}(C_{0,j})$ , đó là xác suất chuyển giao  $p_h(C_{0,j} \rightarrow next)$ .



(a) Không có thông tin đường

(b) Có thông tin đường

Hình 3.11: ví dụ về tính toán  $p_h(C_{0,j} \rightarrow next)$   $p_h(C_{0,j} \rightarrow next)$  khi  $prev(C_{0,j})=1$  và  $next=4$  sử dụng đồ thị của  $F_{HOE}(t_0, prev(C_{0,j}), next', T_{soj})$

Hình 3.11 (a) biểu diễn một ví dụ tính toán  $p_h(C_{0,j} \rightarrow 4)$   $p_h(C_{0,j} \rightarrow 4)$ , khi kết nối  $C_{0,j}$  chuyển vào tế bào 0 từ tế bào 1, sử dụng đồ thị của hàm ước tính di động đối với  $prev(C_{0,j})=1$  được biểu diễn trong hình 3.10. Trong hình vẽ, các giá trị của  $F_{HOE}(t_0, 1, next', T_{soj})$  từ các điểm ở bên phải của đường thẳng đứng tại  $T_{soj} = T_{ext\_soj}(C_{0,j})$  (nghĩa là trong cả vùng bôi đậm và bôi nhạt) được cộng lại để thu được mẫu số trong công thức (3.21). Bởi vì giá trị này khác 0, các giá trị

của  $F_{HOE}(t_0, 1, 4, T_{soj})$  từ hai điểm trong vùng bồi đậm được cộng lại để thu được tử số trong công thức (3.21). Do đó ta có thể tính toán được  $p_h(C_{0,j} \rightarrow 4)$ . Chú ý rằng đầu cuối di động với kết nối  $C_{0,j}$  được ước tính là đứng yên trong tế bào 0 nếu không có sự kiện chuyển giao trong hàm ước tính chuyển giao với thời gian cư trú lớn hơn thời gian cư trú hiện tại của kết nối  $C_{0,j}$ . nghĩa là mẫu số trong công thức (3.21) bằng 0.

Bây giờ, sử dụng các xác suất chuyển giao kết nối vào tế bào 0 từ tế bào bên cạnh  $i$  của nó trong khoảng thời gian  $T_{est}$  (tức là các xác suất chuyển giao  $p_h(C_{i,j} \rightarrow 0)$ ), bảng thông yêu cầu  $B_{r,0}^i$  được định trước trong tế bào 0 cho các chuyển giao mong muốn từ tế bào  $i$  thu được như sau:

$$B_{r,0}^i = \sum_{j \in S_i} b(C_{i,j}) p_h(C_{i,j} \rightarrow 0) \quad (3.22)$$

Trong đó  $S_i$  là tập các chỉ số của các kết nối trong tế bào  $i$  và  $b(C_{i,j})$  là bảng thông của kết nối  $C_{i,j}$ . Cuối cùng, bảng thông tập hợp được định trước trong tế bào 0 cho các cuộc gọi chuyển giao mong muốn từ các tế bào bên cạnh trong khoảng thời gian  $T_{est}$  được tính như sau:

$$B_{r,0} = \sum_{i \in A_n} B_{r,0}^i \quad B_{r,0} = \sum_{i \in A_0} B_{r,0}^i \quad (3.23)$$

Trong đó  $A_i$  là tập các chỉ số của các tế bào bên cạnh tế bào  $i$ .

Chú ý rằng bảng thông định trước mong muốn là một hàm tăng dần của thời gian ước tính  $T_{est}$  vì  $p_h(C_{i,j} \rightarrow 0) p_h(C_{i,j} \rightarrow 0)$  là một hàm tăng dần của  $T_{est}$ . Có thể có một giá trị tối ưu của  $T_{est}$  đối với trạng thái lưu lượng hoặc di chuyển cụ thể để đảm bảo xác suất khóa cuộc gọi mới nhỏ nhất trong khi giữ xác suất rơi cuộc gọi chuyển giao thấp hơn mức mong muốn. Trong cơ chế ở đây, thời gian ước tính được điều chỉnh một cách thích hợp trong mỗi tế bào và độc lập với các tế bào khác, phụ thuộc vào các sự kiện rơi cuộc gọi chuyển giao trong tế bào như được trình bày trong phần 3.3.3. Khi đó, thời gian tạm tính  $T_{est}$  của tế bào next (hoặc  $T_{est,next}$ ) sẽ được sử dụng trong công thức (3.21). Do đó, khi BS của tế bào 0 cần phải cập nhật giá trị của  $B_{r,0}$ , BS sẽ thông báo giá trị hiện tại của  $T_{est,0}$  cho các tế bào bên cạnh, khi đó BS của mỗi tế bào bên cạnh sẽ tính toán bảng thông yêu cầu cho các cuộc gọi chuyển giao mong muốn từ tế bào đó, tức là  $B_{r,0}^i$   $B_{r,0}^i$  cho tế bào  $i$ , sử dụng công thức (3.22), và thông báo giá trị này cho BS của tế bào 0. Cuối cùng BS của tế bào 0 sẽ tính toán  $B_{r,0}$  sử dụng công thức (3.23).

### 3.3.2 Định trước bảng thông với thông tin di chuyển

Cơ chế được phát triển ở trên được điều chỉnh để sử dụng thông tin đường đi. Trường hợp đặc biệt này xảy ra, ví dụ như khi thuê bao sử dụng ô tô có yêu cầu hệ thống dẫn đường, nó sẽ hướng dẫn khách hàng tới một đích cụ thể từ vị trí hiện tại sử dụng hệ thống dẫn đường GPS

và ITS, và hệ thống dẫn đường sẽ trao đổi với BS của tế bào hiện tại để thông báo tế bào tiếp theo mà khách hàng đang dự định di chuyển tới. Khách hàng có thể đôi lúc không theo chỉ dẫn của hệ thống dẫn đường, ví dụ bất ngờ họ bị mất thông tin chỉ dẫn. Tuy nhiên luận án không quan tâm tới những trường hợp như thế, giả sử thuê bao đi theo đường được thông báo cho BS.

Trong phần trước, một BS sử dụng hàm ước tính chuyển giao theo 2 cách: cách thứ nhất là phán đoán tế bào tiếp theo mà đầu cuối di động sẽ chuyển tới, cách thứ hai là ước tính thời gian cư trú của di động trong tế bào hiện tại. Với thông tin đường đi, BS biết tế bào tiếp theo, và do đó hàm ước tính chuyển giao được sử dụng duy nhất cho mục đích thứ hai. Bây giờ, giả sử đầu cuối di động với kết nối  $C_{0,j}$  trong tế bào hiện tại sẽ chuyển tới tế bào tiếp theo, do đó

$$p_h(C_{0,j} \rightarrow next') := 0, \text{ nếu } next' \neq next \quad p_h(C_{0,j} \rightarrow next') := 0, \text{ if } next' \neq next, \quad (3.24)$$

Và công thức (3.21) được biểu diễn dưới dạng khác như sau:

$$p_h(C_{0,j} \rightarrow next) := \frac{\sum_{t_{soj} < T_{ext\_soj}(C_{0,j}) \leq T_{ext\_soj}(C_{0,j}) + T_{ext}} F_{HOE}(t_0, prev(C_{0,j}), next, t_{soj})}{\sum_{t_{soj} > T_{ext\_soj}(C_{0,j})} F_{HOE}(t_0, prev(C_{0,j}), next, t_{soj})}$$

$$\begin{cases} \frac{\sum_{T_{ext\_soj}(C_{0,j}) < t_{soj} \leq T_{ext\_soj}(C_{0,j}) + T_{est}} F_{HOE}(t_0, prev(C_{0,j}), next, t_{soj})}{\sum_{t_{soj} > T_{ext\_soj}(C_{0,j})} F_{HOE}(t_0, prev(C_{0,j}), next, t_{soj})}, \\ \text{if } \sum_{t_{soj} > T_{ext\_soj}(C_{0,j})} F_{HOE}(t_0, prev(C_{0,j}), next, t_{soj}) \neq 0, \\ 0, \text{ Khác} \end{cases} \quad (3.25)$$

Hình 3.11 (b) biểu diễn ví dụ về tính toán  $p_h(C_{0,j} \rightarrow 4)$  với thông tin đường đi, khi kết nối  $C_{0,j}$  đến từ tế bào 1, sử dụng đồ thị của hàm ước tính chuyển giao đối với  $prev(C_{0,j})=1$ , được biểu diễn trong hình 3.11. Trong hình, giá trị của  $F_{HOE}(t_0, 1, 4, T_{soj})$  từ các điểm trong vùng bôi đậm và bôi nhạt được cộng lại để thu được mẫu số trong công thức (3.25). Bởi vì giá trị này khác 0, giá trị của  $F_{HOE}(t_0, 1, 4, T_{soj})$  từ hai điểm trong vùng bôi đậm được cộng để thu được tử số trong công thức (3.25). Khi đó thu được giá trị của  $p_h(C_{0,j} \rightarrow 4)$ . Bây giờ, các xác suất chuyển giao sử dụng công thức (3.24) và (3.25) được sử dụng trong công thức (3.22) và (3.23) để tính toán bằng thông định trước mục tiêu. Cơ chế điều chỉnh này với mong muốn thực hiện tốt hơn cơ chế trước về độ chính xác ước tính và do đó định trước bằng thông có hiệu quả.

### 3.3.3 Điều khiển cửa sổ ước tính thời gian di chuyển

Sử dụng cơ chế ở đây, bằng thông dành cho chuyển giao sẽ được định trước cao hơn (hoặc thấp hơn) nếu  $T_{est}$  là rất lớn (hoặc nhỏ). Do đó có thể tồn tại một giá trị tối ưu của  $T_{est}$  cho lưu lượng tải nhất định và sự di chuyển của người sử dụng, nhưng trong thực tế các tham số này

biến đổi theo thời gian. Hơn nữa, hàm ước tính di chuyển được sử dụng có thể không mô tả tốt đáp ứng của di động, do đó gây ra ước tính di chuyển không chính xác thậm chí là đối với giá trị tối ưu của  $T_{est}$ . Luận án đề xuất một thuật toán tương thích để điều khiển cửa sổ thời gian ước tính di chuyển dựa trên các sự kiện rơi chuyển giao trong mỗi tế bào để tính gần đúng giá trị tối ưu  $T_{est}$  theo thời gian. Hình 3.12 biểu diễn thuật toán được thực hiện bởi BS trong mỗi tế bào để điều chỉnh giá trị của  $T_{est}$ .

Trước khi chạy thuật toán, kích thước cửa sổ tham chiếu  $w (= \lceil 1/P_{HD,target} \rceil)$  được xác định và gán cho kích thước cửa sổ quan sát  $w_{obs}$ . Thêm vào đó,  $T_{est}$  được khởi tạo với giá trị  $T_{start}$ , một tham số thiết kế, và bộ đếm cho kết nối chuyển giao  $n_H$  và rơi kết nối chuyển giao  $n_{HD}$  được điều chỉnh về 0. Có thể nhận ra trong mã giả,  $w_{obs}$  tăng hoặc giảm bởi  $w$ , và với điều kiện  $P_{HD} < P_{HD, target}$  có thể được hiểu là để giữ bộ đếm  $n_{HD}$  các kết nối chuyển giao bị rơi trong suốt cửa sổ  $w_{obs}$  giám sát các kết nối chuyển giao thấp hơn  $w_{obs}/w$ . Trong quá trình chạy, bất cứ lúc nào có một kết nối chuyển giao rơi sau khi có  $w_{obs}/w$  kết nối bị rơi,  $T_{est} := T_{est} + 1$  và  $w_{obs} := w_{obs} + w$ . Mặt khác, khi mà có ít hơn hoặc bằng  $w_{obs}/w$  kết nối chuyển giao bị rơi trong cửa sổ  $w_{obs}$ ,  $T_{est} := T_{est} - 1$  và  $w_{obs} := w$ .  $T_{est}$  không được lớn hơn  $T_{soj,max}$  trong hình 3.12, đó là giá trị  $T_{soj}$  lớn nhất thu được từ các hàm ước tính chuyển giao trong các tế bào bên cạnh, bởi vì bất kể giá trị nào lớn hơn nó đều không có nghĩa. Thiết lập giá trị nhỏ nhất của  $T_{est} = 1$ , vì nếu giá trị quá nhỏ thì sẽ gần như không định trước bằng thông kê cả là đang có các kết nối tồn tại trong tế bào bên cạnh.

```

01. if( $w = \lceil 1/P_{HD,target} \rceil$ ) then  $w_{obs} := w$ ;
02.  $T_{est} := T_{start}$ ;  $n_H := 0$ ;  $n_{HD} := 0$ ;
03. while (Thời gian tăng) {
04. if (Chuyển giao ở cell hiện tại) then {
05.  $n_H := n_H + 1$ ;
06. if (rơi kết nối) then {
07.  $n_{HD} := n_{HD} + 1$ ;
08. if ( $n_{HD} > w_{obs}/w$ ) then {
09.  $w_{obs} := w_{obs} + w$ ;
10. if ( $T_{est} < T_{soj,max}$ ) then  $T_{est} := T_{est} + 1$ ;
11. }
12. }
13. else if ( $n_H \geq w_{obs}$ ) then {
14. if ( $n_{HD} \leq w_{obs}/w$  and  $T_{est} > 1$ ) then  $T_{est} := T_{est} - 1$ ;

```

15.  $w_{\text{obs}} := w; n_{\text{H}} := 0; n_{\text{HD}} := 0;$
16. }
17. }
18. }

Hình 3.12: Mã giả của thuật toán điều chỉnh  $T_{\text{est}}$  trong mỗi BS

Những vấn đề sau cần xem xét khi thiết kế cửa sổ thời gian ước tính điều khiển thuật toán:

- C1. Khi số kết nối chuyển giao bị rơi nhiều hơn giới hạn, thuật toán phải bắt đầu tăng  $T_{\text{est}}$  thật nhanh bởi băng thông định trước thấp; nếu không sẽ tiếp tục rơi kết nối chuyển giao
- C2. Việc tăng  $T_{\text{est}}$  không thể quá lớn. Nó có thể gây nên phản ứng quá và do đó định trước băng thông quá nhiều
- C3. Do phản ứng quá hoặc giảm lưu lượng tải suốt thời gian, có thể có ít kết nối chuyển giao bị rơi so với giới hạn cho phép, do đó giá trị của  $T_{\text{est}}$  có thể giảm nhanh. Nếu không thì băng thông sẽ bị định trước quá nhiều, dẫn tới giảm hiệu suất sử dụng hệ thống
- C4.  $T_{\text{est}}$  không nên giảm quá nhiều. Nếu không thì nó có thể gây ra phản ứng thái quá, do đó định trước không đủ băng thông

Mục đích chính của lựa chọn này là phản ứng nhanh đối với rơi chuyển giao, tức là C1 và C3. Tuy nhiên, những lựa chọn này cũng gây ra phản ứng thái quá và gây nên băng thông định trước dao động mạnh giữa việc đặt trước quá nhiều và quá ít băng thông.

### 3.3.4 Điều khiển đăng nhập

Sau khi tính toán băng thông định trước mong muốn theo công thức (3.1) thì thực hiện kiểm tra đăng nhập, cụ thể:

T1. Kiểm tra nếu  $\sum_{j \in S_0} b(C_{0,j}) + b_{\text{new}} \leq C(0) - B_{r,0}$   $\sum_{j \in S_0} b(C_{0,j}) + b_{\text{new}} \leq C(0) - B_{r,0}$

T2. Nếu việc kiểm tra trên là dương, kết nối được đăng nhập

Trong đó  $C(0)$  và  $b_{\text{new}}$  lần lượt là dung lượng liên kết của tế bào 0 và băng thông của kết nối mới yêu cầu. Cơ chế điều khiển đăng nhập đơn giản này từ nay về sau sẽ được tham chiếu tới AC1. Tuy nhiên, khi không còn đủ băng thông thông rồi bởi các kết nối đang thực hiện thì không thể định trước được băng thông cho chuyển giao, việc tính toán băng thông định trước mong muốn là vô nghĩa. Nếu tình trạng này duy trì trong một khoảng thời gian dài vì những chuyển giao tiếp tục xuất hiện, vấn đề trở nên nghiêm trọng vì một số chuyển giao sẽ bị rơi vì không có băng thông, tiến hành tăng  $T_{\text{est}}$  nhanh hơn nữa. Tiếp theo nó lại yêu cầu định trước

băng thông không tồn tại. Tình huống này xảy ra khi các tế bên cạnh chấp nhận kết nối mới chỉ dựa trên mỗi kiểm tra đăng nhập theo công thức (3.1) và các kết nối này tiếp tục được chuyển giao vào tế bào hiện tại thậm chí nó không đủ băng thông.

Để xử lý vấn đề này, kiểm tra băng thông nên kiểm tra băng thông rỗi của các tế bào bên cạnh cũng như tế bào hiện tại. Chính xác hơn, phải kiểm tra nếu có đủ băng thông trong tế bào hiện tại và tế bào tiếp theo cho mỗi kết nối mới được yêu cầu, nhưng không có thông tin đường đi thì không thể biết được tế bào nào tiếp theo. Do đó, dựa vào việc liệu có thông tin về đường đi hay không để áp dụng cơ chế điều khiển đăng nhập khác nhau. Đầu tiên, khi không có thông tin đường đi:

T1. Đối với tất cả  $i \in A_0$ , kiểm tra nếu  $\sum_{j \in S_i} b(C_{i,j}) \leq C(i) - B_{r,i}$   $\sum_{j \in S_i} b(C_{i,j}) \leq C(i) - B_{r,i}$

T2. Kiểm tra nếu  $\sum_{j \in S_0} b(C_{0,j}) + b_{new} \leq C(0) - B_{r,0}$   $\sum_{j \in S_0} b(C_{0,j}) + b_{new} \leq C(0) - B_{r,0}$

T3. Nếu tất cả các kiểm tra trên là đúng, thì kết nối được đăng nhập

Có thể gọi cơ chế này là AC2. Thứ hai, với thông tin về đường đi, tức là biết kết nối mới sau đó sẽ chuyển tới tế bào *next*:

T1. Nếu  $next \neq 0$  , thì kiểm tra nếu  $\sum_{j \in S_{next}} b(C_{next,j}) \leq C(next) - B_{r,next}$

$$\sum_{j \in S_{next}} b(C_{next,j}) \leq C(next) - B_{r,next}$$

T2. Kiểm tra nếu  $\sum_{j \in S_i} b(C_{i,j}) + b_{new} \leq C(0) - B_{r,0}$   $\sum_{j \in S_i} b(C_{i,j}) + b_{new} \leq C(0) - B_{r,0}$

T3. Nếu các việc kiểm tra trên là đúng, thì kết nối được chấp nhận

Cơ chế này được tham chiếu như là AC2w/PI. Chú ý rằng *next* có thể là 0, và trong trường hợp này bước đầu tiên có thể bỏ qua.

Chú ý rằng sử dụng kiểm tra đăng nhập trong trường hợp không có thông tin đường đi, tế bào hiện tại và tất cả các tế bào bên cạnh của nó phải tính  $B_{r,i}$  cho tất cả các yêu cầu đăng nhập mới, và việc này là tốn kém. Trong thực tế, tình huống không mong muốn được mô tả trong phần đầu của phần này được mong muốn chỉ xảy ra trong mạng tải cao. Do đó, Luận án trình bày một cơ chế lai trong đó chỉ yêu cầu các tế bào bên cạnh, những tế bào không thể định trước được băng thông mong muốn tính toán lại băng thông mong muốn và tham gia vào việc kiểm tra đăng nhập. Chú ý rằng  $B_{r,i}$  là hàm biến thiên theo thời gian và được cập nhật trên cơ sở kiểm tra đăng nhập. Trên cơ sở xuất hiện yêu cầu kết nối mới tại tế bào 0, nếu băng thông định trước mong muốn hiện tại của tế bào bên cạnh  $i$ ,  $B_{r,i}^{curr}$ , băng thông được tính toán cho lần kiểm tra đăng nhập trước, không được định trước hoàn toàn, tế bào này sẽ tính toán lại  $B_{r,i}$ , và tham gia vào kiểm tra đăng nhập. Đầu tiên, khi không có thông tin đường đi (AC3):

T1. Đối với tất cả  $i \in A_0, i \in A_0, \sum_{j \in S_i} b(C_{i,j}) + b_{r,i}^{curr} > C(i) \sum_{j \in S_i} b(C_{i,j}) + B_{r,i}^{curr} > C(i)$ , kiểm tra

$$\text{nếu } \sum_{j \in S_i} b(C_{i,j}) + b_{new} \leq C(i) - B_{r,i} \sum_{j \in S_i} b(C_{i,j}) \leq C(i) - B_{r,i}$$

T2. Kiểm tra nếu  $\sum_{j \in S_0} b(C_{0,j}) + b_{new} \leq C(0) - B_{r,0} \sum_{j \in S_0} b(C_{0,j}) + b_{new} \leq C(0) - B_{r,0}$

T3. Nếu tất cả việc kiểm tra trên là đúng, thì kết nối được chấp nhận

Với thông tin đường đi, hoặc khi kết nối mới yêu cầu sẽ chuyển tới tế bào *next* (AC3w/PI):

T1. Nếu  $next \neq 0$  và  $\sum_{j \in S_{next}} b(C_{next,j}) + b_{r,next}^{curr} > C(next)$  thì kiểm tra nếu

$$\sum_{j \in S_{next}} b(C_{next,j}) \leq C(next) - B_{r,next} \sum_{j \in S_{next}} b(C_{next,j}) \leq C(next) - B_{r,next}$$

T2. Kiểm tra nếu  $\sum_{j \in S_i} b(C_{i,j}) + b_{new} \leq C(0) - B_{r,0} \sum_{j \in S_i} b(C_{i,j}) + b_{new} \leq C(0) - B_{r,0}$

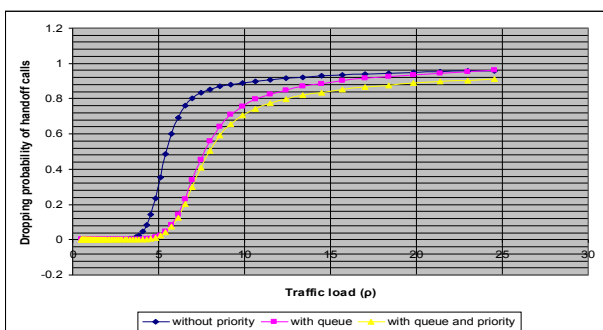
T3. Nếu tất cả kiểm tra trên là đúng, kết nối được chấp nhận

Bảng 3.1 tổng hợp các cơ chế điều kiện đăng nhập được trình bày ở trên.

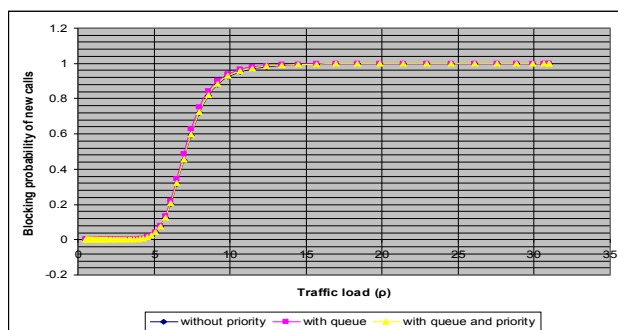
Tên	Thông tin đường đi	Mô tả
AC1	không	Tính $B_r$ chỉ trong tế bào hiện tại
AC2	không	Tính $B_r$ cho tế bào hiện tại và tất cả tế bào bên cạnh
AC2w/PI	có	
AC3	không	Tính $B_r$ trong tế bào hiện tại và chỉ trong vài tế bào bên cạnh
AC3w/PI	có	

### 3.3.5 Kết quả mô phỏng

Gọi  $C$ ,  $C_h$ , và  $K$  lần lượt là tổng số kênh, số kênh HO và kích thước bộ đệm. Giả sử tỷ lệ xuất hiện cuộc gọi mới là  $\lambda_n = 1/15$ , tỷ lệ cuộc gọi HO là  $\lambda_h = 1/20$



(a)



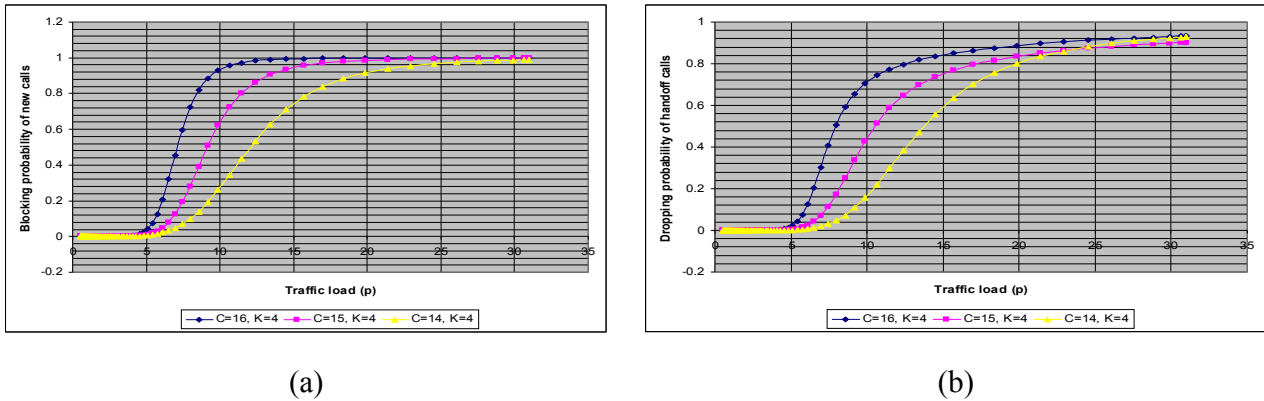
(b)

Hình 3.13: a) Xác suất rớt cuộc gọi HO theo lưu lượng trên mỗi kênh.

b) Xác suất khóa cuộc gọi mới theo lưu lượng trên mỗi kênh.



Ở hình 3.13a, khi tải tăng thì xác suất rớt cuộc gọi tăng, trong khi đó ở hình 3.13b, xác suất khóa cuộc gọi mới là không đổi.



Hình 3.14: a) Xác suất khóa cuộc gọi mới theo lưu lượng trên mỗi kênh với C khác nhau.

b) Xác suất rớt cuộc gọi HO theo lưu lượng trên mỗi kênh với C khác nhau.

Hình 3.14 cho ta thấy, với số kênh bảo vệ và hàng đợi như nhau thì xác suất rớt cuộc gọi HO và khóa cuộc gọi mới đều tăng, bởi vì tỷ lệ kênh bảo vệ cho cuộc gọi HO và tổng số kênh của hệ thống là giảm.

### 3.4 Thuật toán dự báo chuyển giao đi và đến

Luận án mô tả cơ chế điều khiển đăng nhập phân tán (NAG), xuất phát từ [70], trong đó sử dụng ước tính tế bào dựa trên quá trình di chuyển của người sử dụng cho mục đích so sánh. Việc mô tả ở đây mang tính tổng quát hơn cơ chế gốc với việc hỗ trợ các kết nối hỗn tạp (trong phạm vi các băng thông kết nối). [56] trình bày một phiên bản khác được tổng quát hóa từ cơ chế gốc với một số băng thông kết nối. Tất cả việc tính toán hoạt động được ghi lại dựa trên phân bố mũ theo thời gian chiếm giữ của đầu cuối trong mỗi tế bào và tỷ lệ chuyển giao hoặc kết thúc được biết trước.

#### 3.4.1 Ba trạng thái xác suất

Điểm khác nhau chính giữa cơ chế đề xuất trong luận án và NAG là cơ chế của luận án chỉ nghiên cứu chuyển giao vào hệ thống, trong khi đó NAG nghiên cứu cả chuyển giao tới và đi khỏi tế bào. NAG cũng dựa trên khoảng thời gian di chuyển ước tính  $[t_0, t_0 + T_{est}]$ , trong đó  $t_0$  là thời gian hiện tại. Sau khoảng thời gian  $T_{est}$ , kết nối  $C_{0,j}$  có thể ở một trong ba trạng thái với xác suất tương ứng như sau: (1) chuyển giao sang tế bào bên cạnh  $i$  ( $p_h(C_{0,j} \rightarrow i)$ ); (2) kết thúc sau khi đã hoàn thành kết nối ( $p_e(C_{0,j})$ ); (3) ở lại trong tế bào hiện tại ( $p_s(C_{0,j})$ ). Có thể tính toán xác suất của mỗi sự kiện bằng cách sử dụng ước tính di chuyển của khách hàng.

Đầu tiên, các xác suất chuyển giao  $p_h(C_{0,j} \rightarrow i)$  được định nghĩa trong công thức (3.21). Tiếp theo, tính toán xác suất kết nối  $C_{0,j}$  sẽ kết thúc trong khoảng thời gian  $T_{est}$ ,  $p_e(C_{0,j})$ . BS sử dụng thời gian kết nối trung bình  $T_{ave\_life}$  của mỗi đầu cuối, nó được tính theo thời gian:

$$T_{ave\_life} := (1 - \alpha)T_{ave\_life} + \alpha T_{last\_life} \quad (3.26)$$

Trong đó  $\alpha (<1)$  là tham số thiết kế,  $T_{last\_life}$  là thời gian kết nối thu được từ kết nối cuối cùng của khách hàng đó. Giả sử thời gian kết nối của  $C_{0,j}$  theo phân bố mũ với trung bình  $T_{ave\_life}(C_{0,j})$ . Trong thực tế, thời gian kết nối không hoàn toàn tuân theo phân bố mũ, nhưng điều này phụ thuộc vào mỗi đầu cuối, nó không phụ thuộc vào tế bào mà đầu cuối đang ở trong. Do đó, việc giả sử này không ảnh hưởng lớn tới kết quả. Do đó, xác suất được biểu diễn là:

$$p_e(C_{0,j}) = 1 - e^{-T_{est}/T_{ave\_life}(C_{0,j})} \quad (3.27)$$

Cuối cùng, xác suất để kết nối  $C_{0,j}$  nằm trong tế bào với thời gian là  $T_{est}$  được biểu diễn như sau:

$$p_s(C_{0,j}) = (1 - p_e(C_{0,j})) \left(1 - \sum_{i \in A_0} p_h(C_{0,j} \rightarrow i)\right) \quad (3.28)$$

Trong đó  $A_i$  là tập các chỉ số của tế bào bên cạnh tế bào  $i$ .

Giả sử (1) *đáp ứng của mỗi kết nối là độc lập nhau*, (2) *xác suất số lần chuyển giao của đầu cuối trong khoảng thời gian  $T_{est}$  lớn hơn 1 lần là không đáng kể*. Do đó, bảng thông yêu cầu  $B_{T_{est},0}$  cho chuyển giao và các kết nối tồn tại trong tế bào 0 trong khoảng thời gian  $T_{est}$  sẽ là tổng của các bảng thông của (1) *các kết nối trong tế bào 0 trong khoảng thời gian  $T_{est}$*  và (2) *các kết nối chuyển giao vào tế bào 0 từ các tế bào lân cận trong khoảng thời gian  $T_{est}$* . Sử dụng lý thuyết “hạn chế trung tâm” [79], có thể xấp xỉ để có phân bố Gaussian như sau:

$$Pr_{B_{T_{est},0}}(k) \approx B(m_{B,0}, \sigma_{B,0}) \quad (3.29)$$

Trong đó

$$m_{B,0} = \sum_{i \in A_0} \sum_{j \in S_i} b(C_{i,j}) p_h(C_{i,j} \rightarrow 0) + \sum_{j \in S_0} b(C_{0,j}) p_s(C_{0,j}) \quad (3.30)$$

Và biến

$$\begin{aligned} \sigma_{B,0}^2 = & \sum_{i \in A_0} \sum_{j \in S_i} b^2(C_{i,j}) p_h(C_{i,j} \rightarrow 0) (1 - p_h(C_{i,j} \rightarrow 0) \\ & + \sum_{j \in S_0} b^2(C_{0,j}) p_s(C_{0,j}) (1 - p_s(C_{0,j})) \end{aligned} \quad (3.31)$$

Lưu ý  $b(C_{ij})$  là bảng thông của kết nối  $C_{ij}$ .  $S_i$  là tập các chỉ số của kết nối trong tế bào  $I$ , và  $A_i$  là tập các chỉ số của các tế bào lân cận tế bào  $i$ .

### 3.4.2 Điều khiển đăng nhập

Để quyết định đăng nhập, chúng tôi định nghĩa xác suất quá tải khi hết khoảng thời gian  $T_{est}$  trong tế bào  $i$  như sau:

$$\begin{aligned} P_{O,i} &= \Pr(B_{T_{est},i} > C) \\ &\approx Q\left(\frac{C - m_{B,i}}{\sigma_{B,i}}\right) \end{aligned} \quad (3.32)$$

Trong đó  $C$  là dung lượng kết nối,  $m_{B,i}$  và  $\sigma_{B,i}$  lần lượt thu được từ công thức (3.30) và (3.31), sau khi thay  $i = k$ , sau đó thay  $i = 0$  vào các công thức này. Bây giờ, đối với yêu cầu kết nối mới, kiểm tra đăng nhập được thực hiện như sau:

T1. Đối với tất cả  $i \in A_0 \cup \{0\}$ , kiểm tra nếu  $P_{O,i} \leq P_{HD,target}$

T2. Nếu tất cả việc kiểm tra là dương, kết nối được chấp nhận

Chú ý rằng đối với cơ chế này, bảng thông định trước là không xác định. Do đó, liên hệ giữa giá trị  $T_{est}$  và bảng thông định trước cho chuyển giao là không rõ ràng. Về cơ bản,  $T_{est}$  lớn hơn thì  $P_h$  và  $P_e$  lớn hơn, do đó  $P_s$  nhỏ hơn. Nhưng sẽ không rõ ràng liệu  $m_{B,i}$  và  $\sigma_B^2$  sẽ tăng hay giảm khi  $T_{est}$  tăng. Có thể tồn tại một giá trị tối ưu  $T_{est}$  cho phép thu được  $P_{CB}$  nhỏ nhất trong khi giữ  $P_{HD}$  dưới giá trị hướng tới, nhưng không thể thu được một cơ chế tương tự sử dụng cơ chế này để điều khiển cửa sổ ước tính di chuyển. Luận án sử dụng mô phỏng để đánh giá ảnh hưởng của giá trị  $T_{est}$ .

## 3.5 Thuật toán dự báo bằng thông theo kết nối

### 3.5.1 Điều khiển AG

Cơ chế này được mô tả như là cơ chế đảm bảo không có việc rơi cuộc gọi chuyển giao. Việc này có thể được thực hiện bởi kiểm tra bảng thông tại tất cả các tế bào mà đầu cuối yêu cầu kết nối sẽ đi qua, sau đó định trước bảng thông trong mỗi tế bào này. Do đó, cơ chế đăng nhập này liên quan tới định trước bảng thông cho từng kết nối trong mỗi tế bào. Việc định trước bảng thông theo kết nối và điều khiển đăng nhập tương ứng được thực hiện theo điều khiển đăng nhập dựa trên đo kiểm [88]

Để thực hiện cơ chế này, mỗi đầu cuối thông báo cho mạng lõi hoặc trạm gốc tương ứng về thông số di chuyển bao gồm các tế bào mà đầu cuối sẽ đi qua trong thời gian kết nối. Điều này là hoàn toàn không thể để xác định trước đường đi của đầu cuối. Hệ thống định vị của hệ

thống ITS có thể được sử dụng để phán đoán đường đi chính xác của đầu cuối, và từ đó dự báo thông số di chuyển. Vấn đề ở chỗ là nếu sử dụng hệ thống định vị thì có thể biết được tế bào mà đầu cuối tương ứng sẽ di chuyển vào, nhưng ta không biết liệu kết nối sẽ tiếp tục hay không khi đầu cuối di chuyển vào tế bào tiếp theo. Do đó, không thể biết chính xác thông số di chuyển tại thời điểm điều khiển đăng nhập. Nên luận án mô tả cơ chế điều khiển đăng nhập với điều kiện đã có thông tin di chuyển như trong [88].

Đối với thông số di chuyển  $M_{sp}$  của kết nối mới, bao gồm tập các tế bào, và băng thông yêu cầu  $b_{new}$ , điều khiển đăng nhập và băng thông định trước theo kết nối như sau:

- T1. Đối với mỗi tế bào  $i$  trong thông số di chuyển  $M_{sp}$ , kiểm tra nếu
- $$\sum_{j \in S_i} b(C_{i,j}) + b_{new} \leq C - B_{r,i}$$
- T2. Nếu tất cả việc kiểm tra trên là đúng, đối với mỗi tế bào  $i$  trong thông số di chuyển  $M_{sp}$ ,  $B_{r,i} := B_{r,i} + b_{new}$ , kết nối được chấp nhận

Trong đó  $B_{r,i}$  là tổng các băng thông định trước theo cuộc gọi trong tế bào  $i$ . Khi đầu cuối di chuyển vào tế bào, băng thông định trước cho chuyển giao của tế bào sẽ bị giảm:

Trên cơ sở chuyển giao của kết nối  $C_{i,j}$  vào tế bào  $i$ ,  $B_{r,i} := B_{r,i} - b(C_{i,j})$

Chú ý chỉ số  $i$  của tế bào sử dụng trong cơ chế này là mang ý nghĩa khác với chỉ số  $i$  sử dụng ở các cơ chế khác, chỉ số  $i$  là số thứ tự của tế bào trong hệ thống mạng tế bào. Thông qua việc định trước băng thông theo kết nối trong các tế bào trong thông số di chuyển, nó có thể thực hiện để xác suất rơi kết nối chuyển giao bằng 0, nhưng việc sử dụng băng thông là không hiệu quả và xác suất khóa kết nối cao.

### 3.5.2 Định trước băng thông theo kết nối sau khi đăng nhập

Cơ chế này không chủ đích hạn chế xác suất rơi chuyển giao hoặc hạn chế rơi chuyển giao, nhưng nó đặt trước băng thông cho mỗi kết nối trong tế bào dự tính tiếp sau của đầu cuối có kết nối đang tiến hành. Điểm chính của cơ chế này là làm thế nào dự đoán được tế bào tiếp theo của đầu cuối, nó được đề xuất cho môi trường di động trong nhà [64][66]. Điều khiển đăng nhập và định trước băng thông theo kết nối thực hiện như sau:

- T1. Kiểm tra nếu  $\sum_{j \in S_0} b(C_{0,j}) + b_{new} \leq C - B_{r,0}$
- T2. Nếu kiểm tra trên là đúng, đối với tế bào dự kiến tiếp theo của kết nối,  $B_{r,next} := B_{r,next} + b_{new}$ , kết nối được chấp nhận

Trong đó  $B_{r,i}$  là tổng của tất cả băng thông đặt trước theo kết nối trong tế bào  $i$ . Khi đầu cuối di chuyển vào một tế bào, băng thông đặt trước cho chuyển giao của tế bào sẽ giảm:

Trên cơ sở chuyển giao của kết nối  $C_{i,j}$  vào tế bào  $i$ ,  $B_{r,i} := B_{r,i} - b(C_{i,j})$

Chú ý rằng đo kiểm đăng nhập chỉ kiểm tra băng thông có sẵn trong tế bào hiện tại của đầu cuối. Do đó, BS trong tế bào dự kiến tiếp theo của đầu cuối sẽ cố đặt trước băng thông kết nối của đầu cuối. Tuy nhiên, việc này không phải lúc nào cũng thực hiện được vì băng thông có sẵn trong tế bào tiếp theo này không phải là điều kiện đối với đăng nhập hệ thống. Do vậy,  $B_{r,i}$  trong tế bào  $i$  không phải là băng thông định trước thật, nhưng nó là băng thông định trước mong muốn. Thậm chí cơ chế này không nhằm tới không rơi kết nối chuyển giao, nó chỉ thực hiện không rơi kết nối chuyển giao ảo, nhưng nếu so với cơ chế AG thì tốn kém hơn nhiều.

### 3.5.3 Định trước băng thông theo kết nối trước khi đăng nhập

Cơ chế này dự đoán tế bào tiếp theo của đầu cuối yêu cầu kết nối mới trong quá trình điều khiển đăng nhập, và đầu cuối đó được chấp nhận chỉ khi cả tế bào hiện tại và tế bào tiếp theo có đủ băng thông hỗ trợ kết nối. Điểm khác nhau giữa cơ chế này và cơ chế AG là cơ chế AG chỉ định trước băng thông trong tế bào tiếp theo, và điểm khác giữa cơ chế này với cơ chế đặt trước băng thông sau khi đăng nhập là cơ chế định trước băng thông sau khi đăng nhập không cho phép kết nối mới nếu tế bào tiếp theo không có đủ băng thông hỗ trợ kết nối đang yêu cầu. Cơ chế dự báo tế bào tiếp theo được đề xuất cho cơ chế định trước sau khi đăng nhập cũng được sử dụng cho cơ chế này vì cơ chế này chỉ khác một chút so với cơ chế trên. Giả sử có một công cụ dự báo tế bào tiếp theo hoàn hảo (công cụ có khả năng thông báo cho BS là đầu cuối kết thúc kết nối của nó trong tế bào hiện tại hoặc chuyển sang tế bào bên cạnh với kết nối) là có sẵn để đánh giá hoạt động của cơ chế định trước băng thông này. Điều khiển đăng nhập và định trước băng thông thực hiện như sau:

T1. Đối với  $i \in \{0, next\}$  trong đó  $next$  là chỉ số của tế bào dự kiến tiếp theo, kiểm tra nếu

$$\sum_{j \in S_i} b(C_{i,j}) + b_{new} \leq C - B_{r,i}$$

T2. Nếu cả hai việc kiểm tra trên là đúng, kết nối được chấp nhận và  $B_{r,next} := B_{r,next} + b_{new}$

Trong đó  $B_{r,i}$  là tổng tất cả các băng thông định trước trong tế bào  $i$ . Khi đầu cuối chuyển vào một tế bào, băng thông định trước cho chuyển giao của tế bào sẽ giảm:

Trên cơ sở chuyển giao kết nối  $C_{i,j}$  vào tế bào  $i$ ,  $B_{r,i} := B_{r,i} - b(C_{i,j})$

Chú ý rằng  $B_{r,i}$  trong tế bào  $i$  đối với cơ chế này là băng thông định trước thật (do đó nó khác với cơ chế định trước băng thông sau khi đăng nhập). Mặc dù vậy, cơ chế này không nhằm tới tránh rơi chuyển giao, nó cũng thực hiện không rơi chuyển giao ảo, nhưng chi phí đắt hơn so với AG. Cơ chế này nhằm tới cơ chế định trước băng thông theo kết nối tốt nhất.

### 3.6 Kết luận chương 3

Nội dung chính của chương này là dựa trên thông tin di chuyển của người dùng được cập nhật định kỳ để xây dựng cơ chế dự báo xác suất chuyển giao, từ đó đưa ra quyết định đăng ký trước băng thông cho kết nối nhằm đảm bảo QoS cam kết.

Để thực hiện được nội dung nêu trên, luận án tập trung xây dựng và đánh giá *phương pháp dự báo và đặt trước băng thông linh hoạt* để điều khiển chuyển giao và thiết lập cuộc gọi mới, đảm bảo xác suất rơi cuộc gọi thấp hơn giá trị định trước. Phương pháp này sử dụng hai phần tử để đặt trước băng thông cho chuyển giao là: (1) Hàm ước lượng chuyển giao để dự báo tế bào tiếp theo của thuê bao và ước lượng thời gian lưu trú dựa trên tế bào trước đó của nó và thông tin chuyển giao được lưu lại tại mỗi tế bào; (2) cơ chế điều khiển cửa sổ thời gian ước lượng di động mà phụ thuộc vào việc rớt chuyển giao quan sát được, kích cỡ cửa sổ thời gian ước lượng được điều khiển linh hoạt để sử dụng hiệu quả băng thông, và thông tin phản hồi về (i) sự thay đổi lưu lượng (ii) độ chính xác của việc ước lượng tính di động. Cơ chế này được gọi là *dự báo* vì nó ước tính hướng và số lần chuyển giao của các kết nối ở các tế bào bên cạnh, và nó được gọi là *linh hoạt* vì nó điều chỉnh băng thông định trước một cách mềm dẻo tương ứng với kết quả ước tính và các sự kiện rớt chuyển giao được giám sát.

Luận án cũng tập trung phân tích ba cơ chế điều khiển tiếp nhận cuộc gọi phụ thuộc vào việc có bao nhiêu BS liên kết tham gia vào quyết định đăng nhập một yêu cầu kết nối mới. Thông qua việc so sánh về hiệu suất và độ phức tạp của các cơ chế này, luận án đưa ra kết luận là cơ chế AC3 mang lại hiệu suất cao nhất với độ phức tạp chấp nhận được.

Phương pháp dự báo chuyển giao và đặt trước băng thông này sẽ là cơ sở để xây dựng nên *phương thức quản lý chuyển giao linh hoạt* trong chương sau.

Công trình công bố:

- 1) Lê Ngọc Hưng, Nguyễn Xuân Quỳnh, “Nhận dạng và phân lớp các yếu tố ảnh hưởng tới điều khiển chuyển giao”, Hội nghị FAIR lần 10, tháng 8/2017, Đà Nẵng
- 2) Ngọc Hưng Le, “An approach to handover bandwidth reservation in the wireless Future Convergence Network”, NICS Dec.2019.

## **Chương 4: Xây dựng phương thức quản lý chuyển giao linh hoạt trong mạng BcN**

### **4.1 Mở đầu**

Khi chuyển giao, các ứng dụng khác nhau sẽ bị ảnh hưởng khác nhau. Để hiểu được ảnh hưởng của chuyển giao lên ứng dụng di động, dựa trên yêu cầu quản lý di động luận án phân chúng thành 5 loại khác nhau là A, B, C, D, và E [50]. Ảnh hưởng của chuyển giao lên các lớp ứng dụng khác nhau được phân tích tại chương 2, thể hiện thông qua các tham số *Trễ chuyển giao và kết cuối; Tính trong suốt của lớp giao vận; Bảo mật; Tiêu thụ năng lượng và hiệu suất mạng*.

Trong số các tham số ảnh hưởng tới hiệu suất chuyển giao trình bày ở trên, thì Trễ chuyển giao và Bảo mật là các tham số quan trọng đối với tất cả các lớp ứng dụng di động. Trễ kết cuối quan trọng đối với lớp D và E trong khi Tính trong suốt lớp vận chuyển lại quan trọng đối với lớp B và C [50].

Phân tích của luận án trong chương 2 chỉ ra rằng hiệu suất của các giao thức ở trên là khác nhau xét trên khía cạnh các tham số chuyển giao là khác nhau. Cụ thể, trễ kết cuối là nhỏ nhất khi sử dụng SIP [65] và TCP-M [74], tính trong suốt ở lớp vận chuyển khi sử dụng TCP-M và MIP [58]. Việc giảm hiệu suất mà các ứng dụng ghi nhận được trong suốt thời gian chuyển giao chỉ phụ thuộc vào trễ chuyển giao và mức tiêu thụ năng lượng. Như vậy giảm được trễ thì hiệu suất chuyển giao của các của các giao thức khác nhau sẽ được cải thiện, và việc giảm trễ này cần đạt được bất chấp sự biến động của các lớp khác.

Như vậy, không thể có được hiệu suất quản lý di động tối ưu cho các lớp ứng dụng khi chỉ sử dụng một giao thức nào đó.

Để trả lời câu hỏi “Giao thức quản lý di động nào là thích hợp cho lớp ứng dụng riêng biệt” luận án đã nghiên cứu và phân tích hiệu suất chuyển giao của các giao thức hiện tại khi sử dụng cho các lớp ứng dụng khác nhau [50]. Dựa trên kết quả phân tích toán học, luận án đề xuất việc sử dụng TCP-M [74] cho các ứng dụng lớp B và C, SIP [65] cho các ứng dụng thời gian thực lớp D và E, và MIP [58] cho các ứng dụng không thời gian thực lớp D và E. Trong trường hợp này các giao thức quản lý di động chỉ hoạt động từ một lớp và là bất khả thi về tính động của các lớp khác. Do vậy hiệu suất chuyển giao của chúng thay đổi dựa trên tính động của các lớp khác.

### **4.2 Phân tích hiệu suất của các giao thức định tuyến cho MANETs.**

Hiệu suất của MANETs phụ thuộc chủ yếu vào các giao thức định tuyến [83], do vậy phần này luận án đánh giá các giao thức truyền thống từ đó quyết định việc sử dụng các giao thức này trong các ứng dụng cụ thể.

#### 4.2.1 Xu hướng nghiên cứu trong thời gian gần đây.

Để làm rõ xu hướng nghiên cứu cải thiện hiệu suất MANET trong những năm gần đây, chúng tôi đã thực hiện một cuộc khảo sát về các giao thức định tuyến được đề xuất để cải thiện hiệu suất cho MANET trong giai đoạn 2011 – 2020, được công bố trên cơ sở dữ liệu của Thư viện

BẢNG 4.1: CÁC GIAO THỨC ĐỊNH TUYẾN TRÊN IEEE GIAI ĐOẠN 2011-2020

TT	Giao thức đề xuất	Năm	Hệ mô phỏng	So sánh với	Trễ	PDR	Overh.
1	LBRP [13]	2011	GloMoSim	AODV, DSR, ZRP	Có	Có	Có
2	CA-AOMDV[14]	2011	Matlab	AOMDV	Có	Có	Có
3	EDRP [15]	2011	NS2	AODV, PGP	Có	Có	Có
4	D-ODMRP [90]	2012	NS2	ODMRP	Không	Có	Có
5	3DLIS [43]	2013	NS-2	MDART	Có	Có	Có
6	OANTGPS [44]	2013	NS2	AODV, AOMDV,	Có	Có	Không
7	PSR [19]	2014	NS-2	OLSR, DSDV, DSR	Có	Có	Có
8	IAR [8]	2016	NS-3	CLWPR, PIAR,	Có	Có	Có
9	DCFP [45]	2016	Designed	NCPR, AODV	Có	Có	Có
10	TLRC [39]	2016	NS2	GyTAR, STAR	Có	Có	Không
11	iCAR-II [23]	2016	MATLAB	GPSR, GSR, GyTAR	Có	Có	Có
12	3MRP [24]	2017	NS2	GPSR, VIRTUS	Có	Có	Không
13	MoZo [38]	2017	NS2	CBDRP, Brave	Có	Có	Có
14	CBLTR [26]	2017	MATLAB	CBVANET, AODV-	Có	Không	Có
15	RARP [27]	2017	Designed	Traditionnal Protocols	Không	Có	Không
16	CLOLSR [12]	2018	OPNET	Traditionnal Protocols	Có	Có	Có
17	RDBTMA [29]	2019	NS2	EDCA, DBTMA	Có	Có	Có
18	CBMLB [16]	2019	NS2	AODV, AOMDV	Có	Có	Có
19	QTAR [31]	2020	QualNet	RTAR, iCar-II, GyTAR, GPSR, LAR	Có	Có	Có

số Xplore. Kết quả khảo sát được trình bày trong Bảng 4.1.

Kết quả khảo sát cho thấy nhiều giao thức định tuyến đã được đề xuất trong mười năm gần đây dựa trên những cải tiến từ các giao thức truyền thống. Khoảng 80% các nghiên cứu đã so sánh hiệu suất của giao thức được đề xuất với các giao thức truyền thống như AODV hoặc



DSR. Khoảng 90% các nghiên cứu đã thực hiện mô phỏng và đánh giá hiệu suất trên phần mềm mô phỏng như NS2 và Matlab. Các tiêu chí hiệu suất thường được sử dụng để so sánh là thời gian trễ (khoảng 88%), tỷ lệ phân phối gói (khoảng 95%) và tải định tuyến (khoảng hơn 75%). Số lượng nghiên cứu trong những năm gần đây đã tăng lên. Đặc biệt, hơn 65% tổng số công trình được đề xuất trong năm năm gần đây (giai đoạn 2016 - 2020). Kết quả khảo sát cũng cho thấy hầu hết các giao thức định tuyến mới được đề xuất đều dựa trên những cải tiến từ các giao thức định tuyến truyền thống đã biết.

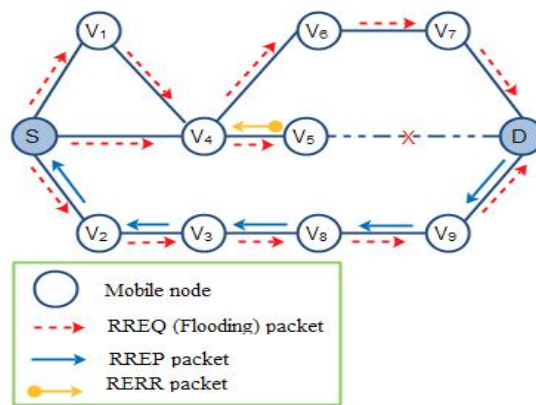
Các kết quả khảo sát trên cho thấy có sự quan tâm và xu hướng nghiên cứu lớn trong lĩnh vực cải thiện hiệu suất của MANETs. Để giải quyết vấn đề này, một chỉ dẫn quan trọng, chúng ta có thể cải thiện giao thức định tuyến truyền thống để có được giao thức định tuyến mới. Do đó, chúng ta cần xem xét hiệu suất của các giao thức truyền thống cho MANET.

### 4.2.2 Các giao thức định tuyến truyền thống

#### 4.2.2.1 Giao thức định tuyến theo nhu cầu

Trong MANETs, hai giao thức định tuyến phản ứng điển hình đã được IETF chuẩn hóa là AODV [26] và DSR [30]. Các giao thức định tuyến theo yêu cầu hoạt động theo nguyên tắc là bất cứ khi nào cần dữ liệu, nút nguồn sẽ khám phá và tìm đường đến nút đích.

Quá trình khám phá tuyến bắt đầu với việc gửi các gói tin Tuyến đường REQuest (RREQ) để phát đường dẫn. Sau đó, các gói này sẽ được chuyển tiếp qua các nút trung gian để đến nút đích (Hình 4.1, đường  $\text{---}\rightarrow$ ).



Hình 4.1: Giai đoạn khám phá tuyến

Nút đích hoặc nút trung gian (nút biết tuyến đến đích) sẽ trả lời bằng cách gửi gói trả lời Tuyến (RREP) đến nút nguồn. Khi nút nguồn nhận được RREP, đường dẫn được đặt và dữ liệu có thể được truyền (Hình 4.1, đường  $\text{---}\rightarrow$ ). Bên cạnh quy trình khám phá tuyến đường, AODV và DSR cũng có quy trình bảo trì tuyến đường bằng cách sử dụng các gói Lỗi tuyến

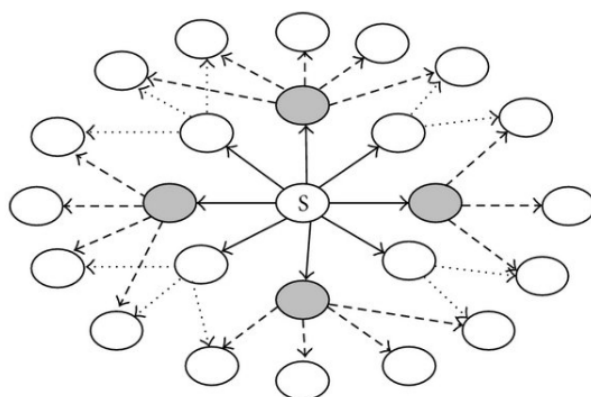
đường (RERR) (Hình 4.1, đường ●→). Mặc dù nó được thiết kế để phù hợp với các đặc tính của MANET, nhưng có một sự khác biệt giữa AODV và DSR. AODV không xây dựng trước một tuyến đường để truyền dữ liệu từ nguồn đến đích. Lộ trình truyền sẽ được xác định bởi mỗi nút mạng khi có sẵn dữ liệu đến, dựa trên thông tin trạng thái hệ thống mà nút mạng thu được. Đồng thời, AODV cũng sử dụng số thứ tự đích / nguồn để xác định tuyến mới cũng như tránh định tuyến vòng lặp. Trong khi đó, DSR xây dựng tuyến đường tại nút nguồn. Nút nguồn sẽ xác định đầy đủ tuyến đường từ nút nguồn đến nút đích để truyền. Do đó, cấu trúc của các gói RREQ và RREP trong DSR phải được mở rộng hơn nữa để chứa thông tin địa chỉ của các nút trung gian. Ngoài ra, DSR duy trì bộ nhớ tạm thời để lưu trữ các tuyến đường và sử dụng chúng cho đến khi nó không còn hiệu lực.

Cả AODV và DSR đều sử dụng tài nguyên hiệu quả, tiết kiệm năng lượng và hỗ trợ các đặc điểm của kiến trúc / tổ chức mạng ad hoc di động như tự tổ chức, tự cấu hình và di động. Tuy nhiên, với khả năng hạn chế, các giao thức định tuyến linh hoạt và hiệu quả hơn, phù hợp với môi trường MANET cần được nghiên cứu và đánh giá thêm.

#### 4.2.2.2 Giao thức định tuyến chủ động

Các giao thức định tuyến chủ động sử dụng bảng định tuyến để xác định tuyến đến tất cả các nút trong mạng. Các nút được cập nhật thường xuyên với thông tin định tuyến về kiến trúc mạng và trạng thái liên kết để làm mới bảng định tuyến. Điều này cho phép bảng định tuyến kiểm soát trạng thái chung của các liên kết trong mạng. Tuy nhiên, trong các mạng có cấu trúc động, việc cập nhật bảng định tuyến liên tục với tần suất cao có thể ảnh hưởng đến hiệu suất của mạng.

##### 1) Giao thức định tuyến trạng thái liên kết được tối ưu hóa (OLSR)



Hình 4.2: Quá trình xác định nút MPR của giao thức OLSSR

OLSR [34] là một cải tiến từ giao thức định tuyến trạng thái liên kết dựa trên multi-hop, sử dụng ba cơ chế để định tuyến: (1) Gửi các gói Hello đến các nút lân cận theo định kỳ, (2)

các gói điều khiển được phát trên mạng bằng MPR (Multi - Điểm chuyển tiếp) và (3) xác định tuyến bằng thuật toán đường đầu tiên ngắn nhất. OLSR được đề xuất để giảm quá tải các gói truyền đi bằng cách chọn một số lượng nhỏ các nút hoạt động như các nút chuyển tiếp đa điểm (MPR). Chỉ các nút này có thể chuyển tiếp các gói, điều này làm giảm số lượng gói truyền đi cũng như kích thước của các gói điều khiển. Để chọn một nút MPR, các gói Hello được gửi trong vòng hai chặng để xác định các nút lân cận. Sau đó, các nút này thực hiện để chọn nút MPR trong vùng (Hình 4.2). Giao thức OLSR có hiệu suất tốt hơn trong môi trường mạng dày đặc và lưu lượng dữ liệu lớn, tuy nhiên, hạn chế của OLSR là chiếm nhiều tài nguyên mạng.

## 2) DSDV (Destination Sequenced Distance Vector)

DSDV [25] là một giao thức định tuyến chủ động sử dụng số bước nhảy để chọn tuyến đường. Giao thức được đề xuất để giải quyết vấn đề lặp lại bằng cách thêm trường số thứ tự vào bảng định tuyến. Không giống như các giao thức định tuyến trạng thái liên kết, DSDV không có bảng định tuyến để xác định tuyến đến tất cả các nút trong mạng. Mỗi nút duy trì một bảng định tuyến đến các nút đích mà nó biết và thông tin này được trao đổi và cập nhật định kỳ. Khi chọn tuyến, DSDV ưu tiên sử dụng tuyến có số tuần tự cao nhất, trong trường hợp có nhiều tuyến có cùng số sê-ri, giao thức chọn tuyến có chi phí thấp hơn. Do nguyên tắc trao đổi định kỳ và cập nhật thông tin định tuyến, trong môi trường MANETs, DSDV thường lãng phí tài nguyên hệ thống nếu kiến trúc mạng ít thay đổi cũng như quá tải khi các tuyến mạng tồn tại lâu trong các bảng định tuyến không được sử dụng.

### 4.2.3 Mô phỏng và phân tích kết quả

#### 4.2.3.1 Tiêu chí đánh giá hiệu suất

Để thực hiện các mô phỏng xác định hiệu suất của các giao thức định tuyến trong các tình huống lưu lượng và lưu lượng khác nhau, luận án xác định các tiêu chí đánh giá hiệu suất như sau:

*Tỷ lệ phân phối gói:* Được xác định là tỷ lệ số lượng gói được nhận bởi nút đích chia cho tổng số gói được gửi từ nút nguồn. Tỷ lệ phân phối gói (PDR) phản ánh cụ thể số gói bị mất trong quá trình truyền và chứng minh tính hiệu quả của các giao thức định tuyến.

$$PDR_{avg} = \frac{P_r}{P_s} \times 100\% \quad (4.1)$$

Trong đó:

$P_r$  là số gói được nhận bởi nút đích

$P_s$  là số gói được gửi bởi nút nguồn

*Thời gian trễ*: Được xác định là thời gian để chuyển gói từ nút nguồn sang nút của lớp ứng dụng, đơn vị: giây (s). Độ trễ đầu cuối trung bình (Độ trễ) được xác định như sau:

$$Delay_{avg} = \frac{\sum_{i=1}^n (t_r - t_s)}{P_r} \quad (4.2)$$

Trong đó:

$t_r$  là thời gian gói tin được nhận tại nút đích

$t_s$  là thời gian gói được gửi tại nút nguồn

*Thông lượng*: Được xác định là phép nhân của số gói dữ liệu và kích thước gói trong một đơn vị thời gian, đơn vị: bit / giây (bps). Công thức sau đây xác định thông lượng:

$$Throughput_{avg} = \frac{P_r \times KT}{T} \quad (4.3)$$

Trong đó:

$T$  là thời gian của quá trình đo

$KT$  là kích thước của gói.

*Tải định tuyến chuẩn hóa (NRL)*: Được xác định là tỷ lệ của tổng số gói điều khiển và tổng số gói dữ liệu nhận được trong một mô phỏng. Tham số NRL cho thấy có bao nhiêu gói điều khiển là cần thiết, bao gồm các gói khám phá và bảo trì tuyến để truyền thành công một gói dữ liệu từ nút nguồn đến nút đích. Công thức sau đây xác định NRL:

$$NRL_{avg} = \frac{P_{Control}}{P_r} \times 100\% \quad (4.4)$$

#### 4.2.3.2 Tham số và kịch bản mô phỏng

Để đánh giá hiệu suất của bốn giao thức định tuyến được phân tích trong Phần 3, AODV, DSR, OLSR và DSDV, tác giả đã thiết lập một hệ thống mô phỏng trên phần mềm NS2, phiên bản 2.34 và so sánh hiệu suất của các giao thức với sự thay đổi tốc độ di động của các nút (kịch bản di động) và lưu lượng mạng (kịch bản giao thông). Trong các mô phỏng này, tác giả đã sử dụng lưu lượng CBR với 200 nút di động được phân phối ngẫu nhiên trong diện tích  $1000m \times 1000m$ . Vùng truyền của các nút mạng được đặt thành 150m.

Tốc độ di chuyển của mỗi nút được đặt ngẫu nhiên trong phạm vi  $[0, V_{max}]$ .

Trong kịch bản di động, đặt số lượng kết nối đầu cuối là 50 trong tất cả các mô phỏng. Tốc độ di chuyển của nút mạng được đặt ngẫu nhiên trong phạm vi  $[0, V_{max}]$ , trong đó,  $V_{max}$  lần lượt là: 5, 10, 15 và 20 (m / s), tương ứng với tốc độ di chuyển trong phạm vi  $[15 - 70]$  (km / h) là tốc độ di chuyển thực sự của các phương tiện trong khu vực đô thị. Việc đánh giá hiệu suất của các giao thức với chuyển động được đề cập ở trên cũng nhằm mục đích tìm ra một giao thức truyền thông thích hợp cho các phương tiện trong thành phố thông minh.

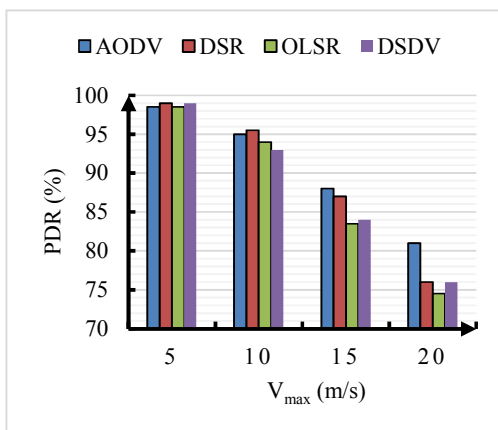
Trong kịch bản lưu lượng mạng, chúng tôi đặt số lượng kết nối đầu cuối là: 20, 40, 60 và 80, tương ứng. Tốc độ di chuyển tối đa của nút mạng trong tất cả các mô phỏng được đặt ở cùng giá trị  $V_{max} = 10$  (m / s). Các thông số mô phỏng chi tiết được trình bày trong Bảng 4.2.

Bảng 4.2: Tham số mô phỏng

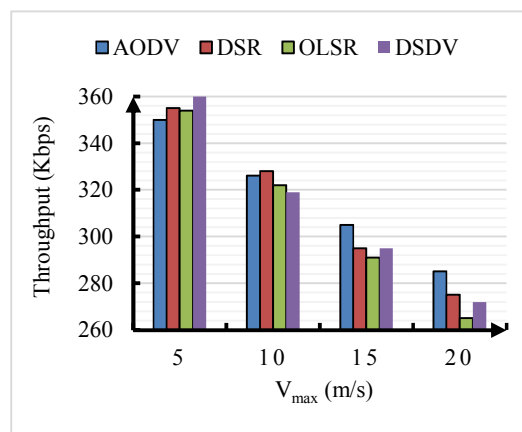
Tham số	Giá trị
Vùng mô phỏng	1000m×1000m
Số nút	200
Thời gian	600 (s)
Loại lưu lượng	CBR
Băng thông	2 (Mbit/s)
Kích thước gói	512 (byte)
MAC Layer	802.11
Transport Layer	UDP
Mô hình di động	Random Way Point
Tốc độ di chuyển	[5-20] (m/s)
Phạm vi truyền	150 (m)

#### 4.2.3.3 Kịch bản di động

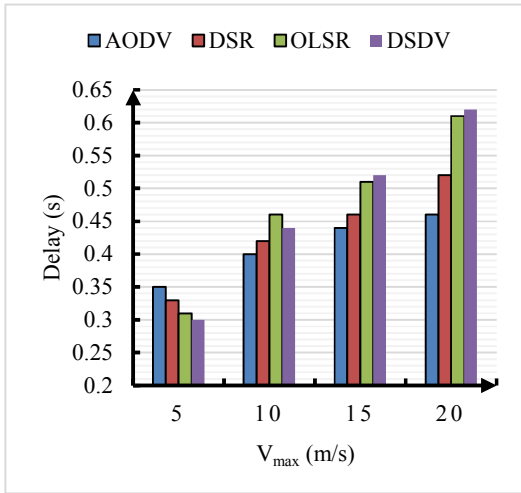
Kết quả mô phỏng cho thấy, trong môi trường MANET, các đặc tính di động của các nút mạng có ảnh hưởng đáng kể đến hiệu năng hệ thống. Khi các nút di chuyển với vận tốc cao, các liên kết có xác suất ngắt cao hơn, bởi vì các tuyến có thể được định tuyến lại và các gói dữ liệu được truyền lại ngày càng nhiều. Do đó, tỷ lệ phân phối gói (Hình 4.3) và thông lượng (Hình 4.4) sẽ giảm nhanh trong khi độ trễ (Hình 4.5) và tải định tuyến (Hình 4.6) sẽ nhanh chóng tăng.



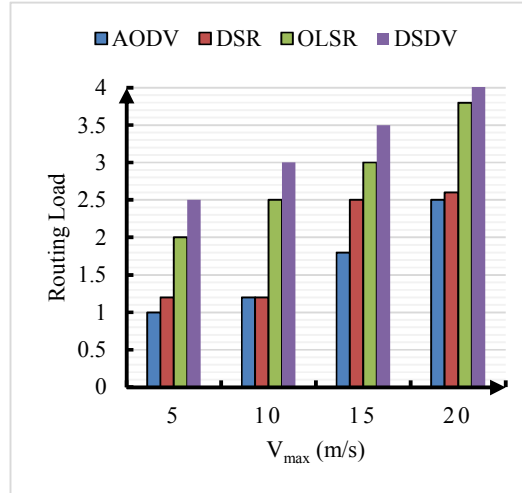
Hình 4.3: Tỷ lệ phân phối gói – Di động



Hình 4.4: Băng thông trung bình - Di động



Hình 4.5: Trễ trung bình - Di động



Hình 4.6: Tải định tuyến chuẩn hóa - Di động

Nói chung, trong điều kiện, các nút mạng có tốc độ di chuyển thấp, các giao thức định tuyến chủ động như OLSR hoạt động khá tốt. Tuy nhiên, khi tốc độ di chuyển của các nút mạng tăng lên khiến cấu trúc mạng liên tục thay đổi, các giao thức này bị hạn chế với độ trễ cao hơn, tỷ lệ phân phối gói thấp hơn và thông lượng giảm đáng kể so với các giao thức định tuyến theo yêu cầu. Hơn nữa, số liệu tải định tuyến của các giao thức theo yêu cầu cũng thấp hơn nhiều so với các giao thức chủ động. Điều này phản ánh hiệu quả năng lượng của các giao thức định tuyến theo yêu cầu.

#### 4.2.3.4 Kích bản lưu lượng

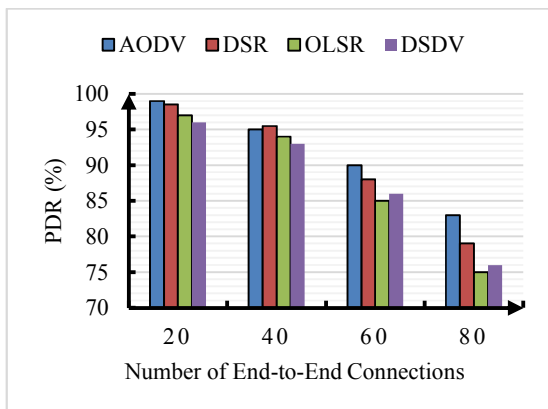
Kết quả mô phỏng cho thấy, tương tự như kịch bản di động, khi lưu lượng mạng tăng (số lượng kết nối đầu cuối tăng), tỷ lệ phân phối gói (Hình 4.7) và thông lượng (Hình 4.8) cho thấy xu hướng giảm. Ngược lại, độ trễ (Hình 4.9) và tải định tuyến (Hình 4.10) hiển thị xu hướng tăng với tất cả các giao thức.

Khi lưu lượng mạng thấp (kết nối số đầu cuối bằng 20), sự khác biệt giữa các giao thức là không nhiều, tỷ lệ phân phối gói của tất cả các giao thức là rất cao, trên 97%, trong khi độ trễ của OLSR là thấp nhất. Tuy nhiên, khi lưu lượng mạng tăng, tỷ lệ phân phối gói và thông lượng của giao thức OLSR và DSDV giảm và độ trễ tăng nhanh so với giao thức AODV và DSR.

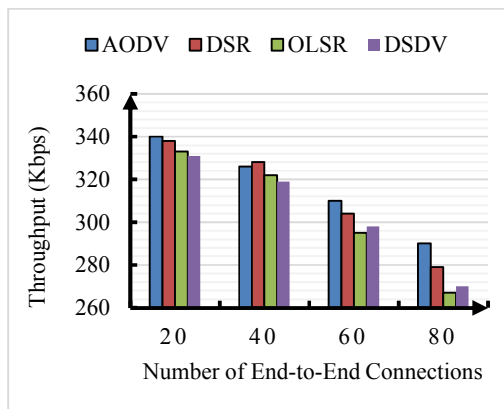
Ở số kết nối đầu cuối bằng 80, hiệu suất của AODV được cải thiện đáng kể nhất. Những vấn đề này cho thấy, khi mạng có lưu lượng truy cập cao, khả năng tắc nghẽn và va chạm xảy ra thường xuyên hơn. Đây là nguyên nhân chính của việc truyền lại, do đó làm tăng thời gian trễ, tải định tuyến cũng như giảm thông lượng và tỷ lệ phân phối gói của toàn bộ hệ thống.

Nói chung, khi mạng có lưu lượng thấp, các giao thức định tuyến chủ động như OLSR hoạt động khá tốt. Tuy nhiên, khi hệ thống có lưu lượng truy cập cao, các giao thức định tuyến theo

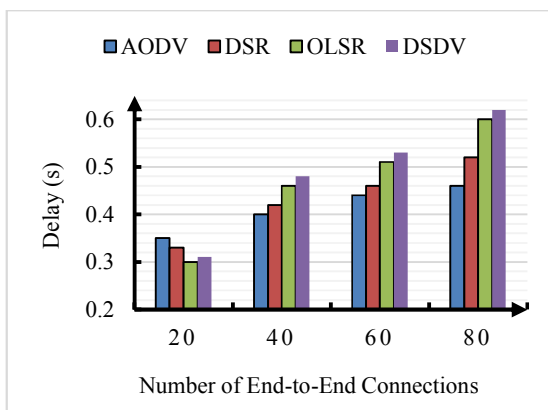
yêu cầu như AODV có các chỉ số hiệu suất vượt trội so với các giao thức còn lại. Hơn nữa, số liệu tải định tuyến của các giao thức theo yêu cầu cũng thấp hơn nhiều so với các giao thức chủ động. Các vấn đề phản ánh khả năng tiết kiệm năng lượng của các giao thức định tuyến phản ứng so với các giao thức định tuyến chủ động.



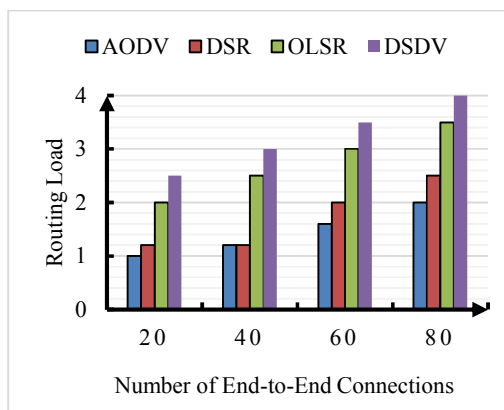
Hình 4.7: Tỷ lệ phân phối gói – Lưu lượng



Hình 4.8: Băng thông trung bình – Lưu lượng



Hình 4.9: Trễ trung bình – Lưu lượng



Hình 4.10: Tải định tuyến chuẩn hóa – Lưu lượng

Như vậy, luận án đã thực hiện một cuộc khảo sát về xu hướng nghiên cứu hiệu suất cao cho các mạng ad hoc di động trong thời gian gần đây để làm rõ các phương pháp và phương pháp tiếp cận. Sau đó, luận án đã chọn và phân tích 04 giao thức điển hình được đề xuất cho MANET, đó là AODV, DSR: sử dụng phương pháp định tuyến theo yêu cầu và OLSR, DSDV: sử dụng phương pháp định tuyến chủ động. Để làm rõ hiệu suất mạng của từng giao thức trong các kịch bản khác nhau về lưu lượng và tính di động, chúng tôi đã thực hiện đánh giá hiệu suất của các giao thức với các tình huống mạng di động và lưu lượng khác nhau.

Kết quả mô phỏng cho thấy khi hệ thống mạng có tính lưu động và lưu lượng truy cập thấp, các giao thức định tuyến chủ động như OLSR hoặc DSDV hoạt động tốt. Mặt khác, khi hệ thống mạng có tính lưu động và lưu lượng cao (tốc độ di chuyển của các nút trên 10 m / s), các giao thức định tuyến theo yêu cầu đạt được tiêu chí hiệu suất tốt hơn các giao thức định

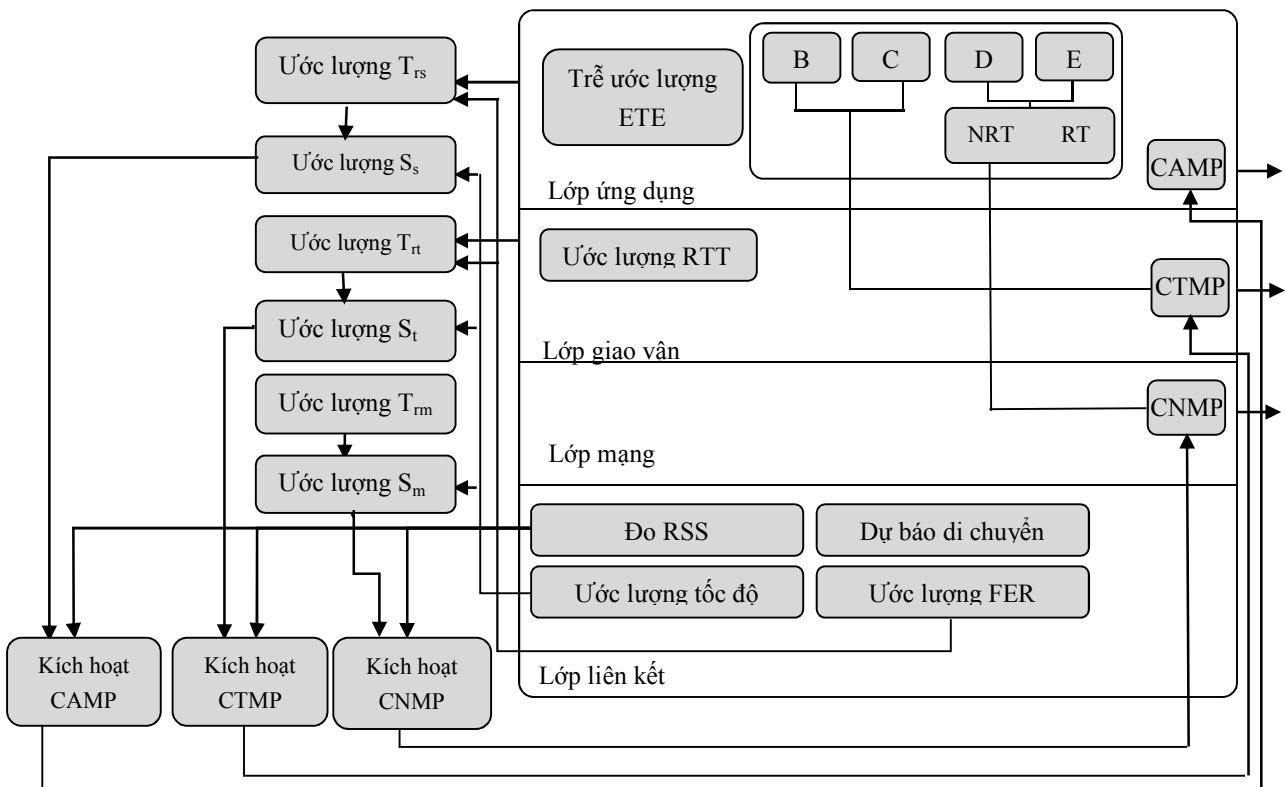
tuyến chủ động, đặc biệt là giao thức AODV. Điều đó cho thấy các giao thức định tuyến chủ động phù hợp với các mạng có cấu trúc ổn định và các giao thức định tuyến theo yêu cầu phù hợp với các mạng có tính di động cao.

### 4.3. Giải pháp Quản lý chuyển giao linh hoạt (AMMS)

AMMS hoạt động theo 2 cách tiếp cận. Đầu tiên, chọn giao thức quản lý di động thích hợp nhất cho ứng dụng. Tiếp đến, xác định trước trễ báo hiệu chuyển giao và khởi động các thủ tục chuyển giao tại thời điểm thích hợp để giảm tối đa trễ. Cấu trúc của AMMS được minh họa như hình 4.11.

#### 4.3.1 Cấu trúc của AMMS

Hình 4.11 minh họa cấu trúc của AMMS. NRT là các ứng dụng không thời gian thực lớp D và E, còn RT là các ứng dụng thời gian thực lớp D và E. Như hình 4.11, việc sử dụng thông tin từ các lớp khác nhau cho phép hoạt động xuyên suốt giữa các lớp trong các giao thức quản lý di động.



Hình 4.11: Cấu trúc của AMMS

- Giao thức quản lý di động lớp ứng dụng (CAMP): Sử dụng đối với các ứng dụng thời gian thực lớp D và E.  $\tau_s$  (Trễ đăng ký địa chỉ của SIP) sử dụng FER, công nghệ truy nhập lớp kết



nối, và thông tin trễ kết cuối (ETE) như trình bày ở mục 4.3.1.6. CAMP khởi động các thủ tục chuyển giao khi RSS do MH thu được từ BS bé hơn  $S_s$ .

- Giao thức quản lý di động lớp vận chuyển (CTMP): sử dụng cho các ứng dụng lớp B và C.

$\tau_r$  (Trễ đăng ký địa chỉ khi dùng TCP-M) sử dụng FER, công nghệ truy nhập lớp kết nối, thông tin RTT của TCP như ở mục 4.3.1.7. CTMP khởi động các thủ tục chuyển giao khi RSS do MH thu được từ BS bé hơn  $S_t$ .

- Giao thức quản lý di động lớp mạng (CNMP): sử dụng cho các ứng dụng không thời gian thực lớp D và E.  $\tau_{rm}$  (Trễ đăng ký địa chỉ khi dùng MIP) như ở mục 4.3.1.8. CNMP khởi động các thủ tục chuyển giao khi RSS do MH thu được từ BS bé hơn  $S_m$ .

Chức năng của các khối thuộc AMMS được mô tả như sau.

#### 4.3.1.1 Khối dự báo di chuyển.

Tác giả sử dụng [1] để dự báo NN.

#### 4.3.1.2 Khối xác định tốc độ di chuyển

Khối này có nhiệm vụ xác định tốc độ di chuyển của MH. Ở đây chúng tôi sử dụng thuật toán VEPSD [47] để xác định tốc độ di chuyển của MH. Sử dụng thông tin về tần số Doppler ( $f_m$ ) có được từ tín hiệu thu được.  $f_m$  liên quan tới vận tốc ( $v$ ) của ánh sáng trong vũ trụ ( $c$ ) và tần số ( $f_c$ ) mang tín hiệu thu được thông qua:

$$v = \left( \frac{c}{f_c} \right) f_m \quad (4.7)$$

VEPSD xác định  $f_m$  thông qua việc sử dụng độ dốc mật độ phổ công suất (PSD) của tín hiệu thu được. Độ dốc PSD của tín hiệu thu được có giá trị cực đại tại các tần số  $f_c \pm f_m$  trong môi trường di động [47]. VEPSD xác định giá trị cực đại của PSD tương ứng với véc tơ tần số cao nhất  $f_c + f_m$  để xác định  $f_m$ .

#### 4.3.1.3 Khối xác định FER

Khối này xác định tỷ lệ lỗi khung (FER) lớp kết nối. Thực tế, các giao thức MAC có thông tin về FER [33]. Khối này lựa chọn thông tin FER từ lớp MAC.

#### 4.3.1.4 Khối xác định RTT

Khối này xác định RTT của kết nối TCP thông qua việc sử dụng các biến trạng thái của TCP.

#### 4.3.1.5 Khối xác định trễ kết cuối ETE

Khối này xác định trễ kết cuối lớp ứng dụng của ứng dụng thời gian thực khi sử dụng SIP. Trễ ETE này phụ thuộc vào việc có hay không có RLP ở lớp kết nối, và được xác định thông qua [50]:

$$ED_{nr} = D + t_{wco} \quad (4.8)$$

$$\text{Và } ED_{nr} = T_f + (K - 1)\tau + t_{wco} \quad (4.9)$$

Trong đó, D là trễ truy nhập lớp kết nối,  $t_{wco}$  là trễ ở kết nối hữu tuyến giữa OBS và CH, và  $K = \left\lceil \frac{L_p}{L_f} \right\rceil$  là số khung lớp kết nối trên một gói số liệu RTP.  $L_p$  là độ dài của một gói RTP, và  $L_f$

là độ dài của một khung lớp kết nối.  $T_f$  được xác định như sau [8] :

$$T_f = D(1 - p_f) + \sum_{i=1}^n \sum_{j=1}^i P(C_{i,j})(2iD + 2(j-1)\tau) \quad (4.10)$$

Trong đó,  $p_f$  là FER và  $\tau$  là khoảng cách giữa các khung trong lớp kết nối, thường là 20ms [8].  $P(C_{ij})$  là xác suất để BS nhận được khung thứ i do MH truyền lại ở lần truyền thứ j. Theo [8] ta có:

$$P(c_{i,j}) = p_f(1 - p_f)^2 ((2 - p_f)p_f)^{\binom{i-1}{2} + j - 1}, \text{ với } i=1,2,\dots,n \text{ và } j=1,2,\dots,i \quad (4.11)$$

Khi MH nhận gói RTP mới, nó lấy được giá trị mẫu mới cho trễ kết cuối bằng cách so sánh trường thời gian của RTP với đồng hồ chủ của nó. Sau đó nó cập nhật trễ kết cuối như sau:

$$ED_{nr} = (1-x)(ED_{nr} + xT_{nr}) \quad (4.12)$$

$$ED_r = (1-x)ED_r + xT_r \quad (4.13)$$

Ở đây,  $T_{nr}$  và  $T_r$  là giá trị lấy mẫu tức thì của trễ ETE có và không có RLP. Giá trị điển hình của x là 0.125.

#### 4.3.1.6 Khối xác định $\tau_{rs}$

Khối này xác định trễ đăng ký địa chỉ của SIP ( $\tau_{rs}$ ). Theo [50] ta có:

$$\tau_{rs} = 2D_{mc} \quad (4.14)$$

Với  $D_{mc}$  Là trễ một chiều trung bình truyền tới các bản tin báo hiệu SIP (có thể thông qua TCP hoặc UDP [65]) giữa MH và CH. Trong phân tích này, luận án quan tâm tới việc sử dụng UDP để truyền các bản tin báo hiệu SIP,  $D_{mc}$  được xác định theo [50]

$$D_{mc} = (1-q) \left\{ B + A \sum_{i=2}^m q^{i-1} (\gamma^{i-1} - 1) + \sum_{i=m+1}^{\infty} q^{i-1} [A(\gamma^{m-1} - 1) + (i-m)\gamma^{m-2} \Delta] \right\} \quad (4.15)$$

Trong đó, B là trễ truyền số liệu kết cuối giữa MH và CH.  $B = B_{nr}$  khi không có RLP và  $B = B_r$  khi có RLP.

$$B_{nr} = D + t_{wco} \quad (4.16)$$

$$\text{Và } B_r = T_f + (K_s - 1)\tau + t_{wco} \quad (4.17)$$

ở đây  $t_{wco}$  là trễ trong kết nối hữu tuyến giữa OBS và CH. D là trễ truy nhập lớp kết nối.

$K_s = \left\lceil \frac{L_s}{L_f} \right\rceil$  số khung lớp kết nối vô tuyến trên một bản tin SIP INVITE,  $L_s$  là độ dài bản tin và

$L_f$  được tính theo (4.10),  $q$  là xác suất mất số liệu kết cuối giữa Mh và CH.  $Q = q_{nr}$  khi không có RLP và  $q = q_r$  khi có RLP.

$$p_{nr} = 1 - (1 - p_f)^{K_s} (1 - p_c) \quad (4.18)$$

$$p_r = 1 - \left[ 1 - p_f \left( (2 - p_f) p_f \right)^{\frac{(n^2+n)}{2}} \right]^{K_s} (1 - p_c) \quad (4.19)$$

Với  $p_f$  là FER,  $p_c$  là xác suất mất số liệu trong mạng cố định giữa MH và CH, và  $n$  là số lần thử tối đa mà RLP thực hiện trước khi hủy bỏ việc truyền khung qua lớp kết nối., thông thường  $n = 3$  đối với RLP [8].

$\Delta$  là giá trị đầu của đồng hồ truyền lại đối với bản tin báo hiệu SIP, nó phải đủ lớn để đếm được kích cỡ bản tin báo hiệu SIP, bằng 2 RTT giữa MH và CH, và bé nhất là 100ms để cho phép xử lý các bản tin giữa MH và CH.  $\gamma$  là hệ số khoảng thời gian RTO được tăng lên sau mỗi lần truyền hỏng, thường = 2,  $A = \frac{\Delta}{\gamma - 1}$ ,  $m$  là số nguyên đếm số lần truyền lại.

Do vậy  $\tau_{rs}$  có thể xác định được thông qua các thông tin sau:

- Có hay không có RLP ở lớp kết nối
- FER lớp kết nối
- B

FER và B tính được thông qua khối xác định FER và khối xác định trễ kết cuối như trình bày ở trên. Từ (4.8), (4.9), (4.16) và (4.17) ta có:

$$B_{nr} = ED_{nr} \quad (4.20)$$

$$\text{Và } B_r = ED_r - (K - 1)\tau + (K_s - 1)\tau \quad (4.21)$$

$\tau_{rs}$  tính được thông qua (4.14).

#### 4.3.1.7 Khối xác định trễ đăng ký địa chỉ TCP-M ( $\tau_{rs}$ )

Theo [50] ta có:

$$\tau_{rt} = \sum_{i=0}^{N_m-1} \sum_{j=0}^{N_m-1} \sum_{k=0}^{N_m-1} P_h(i, j, k) L_h(i, j, k) \quad (4.22)$$

$N_m$  là số lần truyền lại trước khi TCP hủy kết nối.  $P_h(i, j, k)$  được xác định như sau:

$$P_h(i, j, k) = p_1^i (1 - p_1) p_2^j (1 - p_2) p_2^k (1 - p_2) \quad \text{với } (i, j, k = 0, 1, 2, \dots, N_m - 1) \quad (4.23)$$

$p_1$  là xác suất mất số liệu kết cuối giữa MH và CH đối với bản tin SYN,  $p_2$  là xác suất mất số liệu kết cuối giữa MH và CH đối với bản tin SYN/ACK hoặc ACK.  $p_1 = p_{1nr}$  khi không có RLP,  $p_1 = p_{1r}$  khi có RLP. Từ (4.18) và (4.19) ta tính được  $p_{1nr}$  và  $p_{1r}$ . Tương tự ta có  $p_2 = p_{2nr}$  khi không có RLP và  $p_2 = p_{2r}$  khi có RLP.

Từ [50] ta có:

$$L_h(i, j, k) = 1.5RTT_o + \sum_{m=0}^{i-1} 2^m RTO + \sum_{m=0}^{j-1} 2^m RTO = 1.5RTT_o + (2^i + 2^j + 2^k - 3)RTO$$

Với  $(i, j, k = 0, 1, 2, \dots, N_m - 1)$  (4.24)

RTO là giá trị đầu của Timeout truyền lại đối với kết nối TCP và  $RTO = \xi RTT_o$ .  $\xi$  là hằng số.

$RTT_o$  là RTT của TCP ở ON. Do vậy,  $\tau_{rt}$  có thể xác định được thông qua các thông tin sau:

- Có hay không có RLP ở lớp kết nối
- FER lớp kết nối
- $RTT_o$

FER và  $RTT_o$  xác định thông qua khối FER và khối RTT.  $\tau_{rt}$  xác định htoong qua (4.22)

#### 4.3.1.8 Khối xác định trễ đăng ký địa chỉ của MIP ( $\tau_{rm}$ )

Khối này tính toán trễ đăng ký địa chỉ khi dùng MIP ( $\tau_{rm}$ ).  $\tau_{rm}$  là thời gian cần để xử lý việc đăng ký MIP, nó phụ thuộc vào trễ giữa MH và HA của nó. Luận án đề xuất một cơ chế đơn giản sử dụng MIP để xác định  $\tau_{rm}$  như sau.

MH gửi bản tin đăng ký MIP tới HA với một mở rộng xác thực HA không có hiệu lực. Mục đích của việc sử dụng mở rộng xác thực không có hiệu lực là chỉ để ghi nhận trễ báo hiệu đăng ký địa chỉ mà không thay đổi mobility binding tại HA. Khi HA nhận được bản tin đăng ký MIP và ghi nhận sự hiện diện của mở rộng xác thực không có hiệu lực nó trả lại bản tin trả lời đăng ký địa chỉ với mã tương ứng [58] biểu thị việc xác thực bị lỗi. Sau đó thông qua việc so sánh sự khác nhau về thời gian giữa thời gian truyền yêu cầu đăng ký và thời gian nhận trả lời đăng ký để xác định  $\tau_{rm}$ . Cơ chế này tạo ra một lượng bản tin làm tăng tải hệ thống. Tuy nhiên chúng tôi vẫn sử dụng cơ chế này vì nó đơn giản. Hơn nữa, cơ chế có thể cài đặt được trên giao thức MIP hiện tại mà không cần thay đổi bổ sung nào khác.

#### 4.3.1.9 Khối kích hoạt CAMP

Khối này lựa chọn thông tin  $\tau_{rs}$  và  $v$  từ khối ước lượng  $\tau_{rs}$  và khối ước lượng tốc độ di chuyển ở trên. Nó xác định giá trị  $L_s$  bằng cách sử dụng  $L = L_s$  và  $\tau = \tau_{rs}$  ở (4.4). Sau đó nó tính toán ngưỡng RRS động cho việc đăng ký địa chỉ SIP, sử dụng  $S = S_s$  và  $L = L_s$  ở (4.5) để xác định  $S_s$  và  $L_s$ . Khi RSS của MH nhận được từ BS bé hơn  $S_s$  thì khối CAMP sẽ kích hoạt CAMP để thực hiện chuyển giao.

#### 4.3.1.10 Khối kích hoạt CTMP

Khối này lựa chọn thông tin  $\tau_{rt}$  và  $v$  từ khối xác định  $\tau_{rt}$  và xác định tốc độ di chuyển ở trên. Nó xác định giá trị  $L_t$  bằng cách sử dụng  $L = L_t$  và  $\tau = \tau_{rt}$  ở (4.4). Sau đó nó tính toán ngưỡng RRS động cho việc đăng ký địa chỉ TCP-M, sử dụng  $S = S_t$  và  $L = L_t$  ở (4.5) để xác định  $S_t$  và  $L_t$ . Khi RSS của MH nhận được từ BS bé hơn  $S_t$  thì khối CTMP sẽ kích hoạt CTMP để thực hiện chuyển giao.

#### 4.3.1.11 Khối kích hoạt CNMP

Khối này lựa chọn thông tin  $\tau_{rm}$  và  $v$  từ khối xác định  $\tau_{rm}$  và xác định tốc độ di chuyển ở trên. Nó xác định giá trị  $L_m$  bằng cách sử dụng  $L = L_m$  và  $\tau = \tau_{rm}$  ở (4.4). Sau đó nó tính toán ngưỡng RRS động cho việc đăng ký địa chỉ SIP, sử dụng  $S = S_m$  và  $L = L_m$  ở (4.5) để xác định  $S_m$  và  $L_m$ . Khi RSS của MH nhận được từ BS bé hơn  $S_m$  thì khối CNMP sẽ kích hoạt CNMP để thực hiện chuyển giao.

Sự hoạt động của AMMS được minh họa ở hình 4.12. Đầu tiên MH thấy trước việc chuyển giao và dự báo được NN. Tiếp theo, dựa trên loại ứng dụng nó chọn một giao thức quản lý di động thích hợp. Tiếp đến nó xác định trễ báo hiệu chuyển giao đối với giao thức quản lý di động đã chọn, xác định thời gian khởi động chuyển giao, cuối cùng MH sẽ khởi động việc đăng nhập địa chỉ từ NN tại thời điểm khởi động chuyển giao.

### 4.4 Phân tích mô hình đánh giá hiệu suất của AMMS

Hình 4.12 minh họa quá trình xử lý chuyển giao của AMMS khi MH di chuyển giữa hai tế bào. Khi sử dụng các giao thức hiện tại, và MH di chuyển qua điểm Q (hình 4.12a) thì nó sẽ khởi động chuyển giao từ OBS sang NBS, ta sử dụng (4.2) để xác định trễ chuyển giao. Cơ chế AMMS bắt đầu xử lý chuyển giao trước khi MH đi vào vùng giao nhau giữa OBS và NBS, ở đây giả sử rằng MH sẽ di chuyển vào BS dự báo trước, trường hợp này có thể xảy ra các tình huống sau:

Trường hợp 1: Dự báo NN không chính xác, nghĩa là MH di chuyển vào BS khác với BS dự báo. MH ghi nhận việc dự báo BS không thành công bằng việc so sánh ID của NBS (MH ghi nhận ID của NBS thông qua các bản tin quảng bá BS mà nó nhận được sau khi đi vào vùng giao nhau giữa OBS và NBS) với ID của BS dự báo. Trong trường hợp này việc đăng nhập địa chỉ IP đối với BS dự báo là không hữu ích, và MH khởi động chuyển giao tới NBS sau khi đi qua điểm Q (hình 4.12a). Ký hiệu  $p_{c1}$  là xác suất xảy ra trường hợp này, tức là xác suất dự báo NN không thành công.

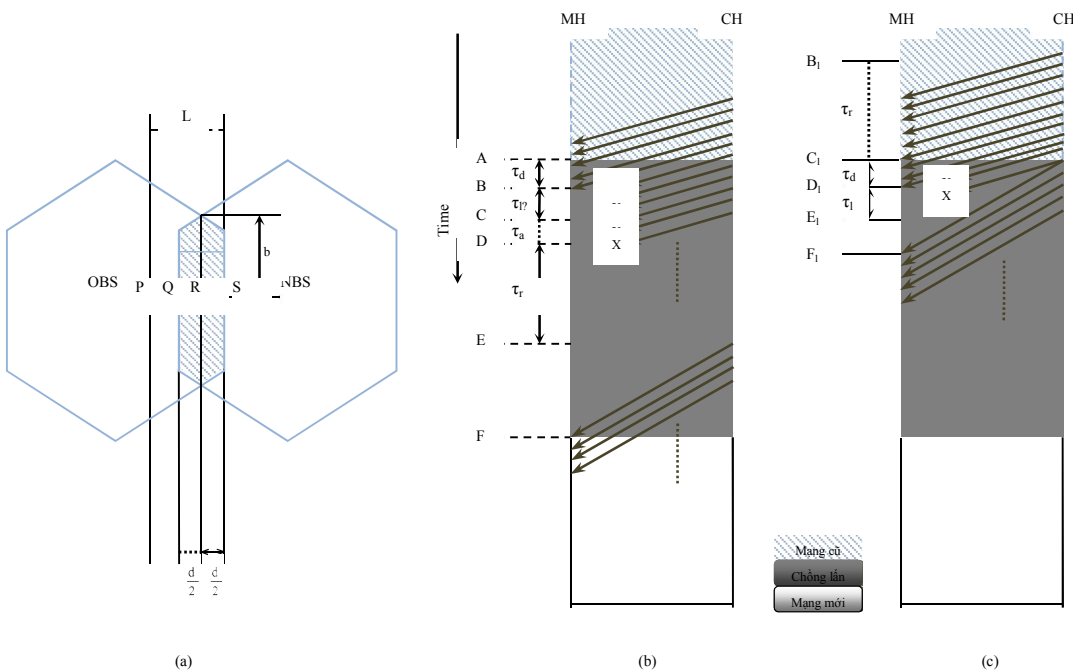
Trường hợp 2: MH di chuyển vào BS do khối dự báo di chuyển đưa ra. Ký hiệu xác suất xảy ra trường hợp này là  $p_{c2}$ , tức là xác suất dự báo NN thành công. Khi đó:

$$p_{c2} = 1 - p_{c1}$$

Trong trường hợp 2, có thể xảy ra các tình huống sau:

i) Quá trình MH đăng nhập và đăng ký địa chỉ IP của CAMP được hoàn thành trước khi MH đi vào vùng giao nhau. Gọi  $P_e$  là xác suất xảy ra khả năng này, ta có:

$$p_e = p(t < \tau_{rs}) \tag{4.25}$$



Hình 4.12: Vùng phủ sóng của OBS và NBS

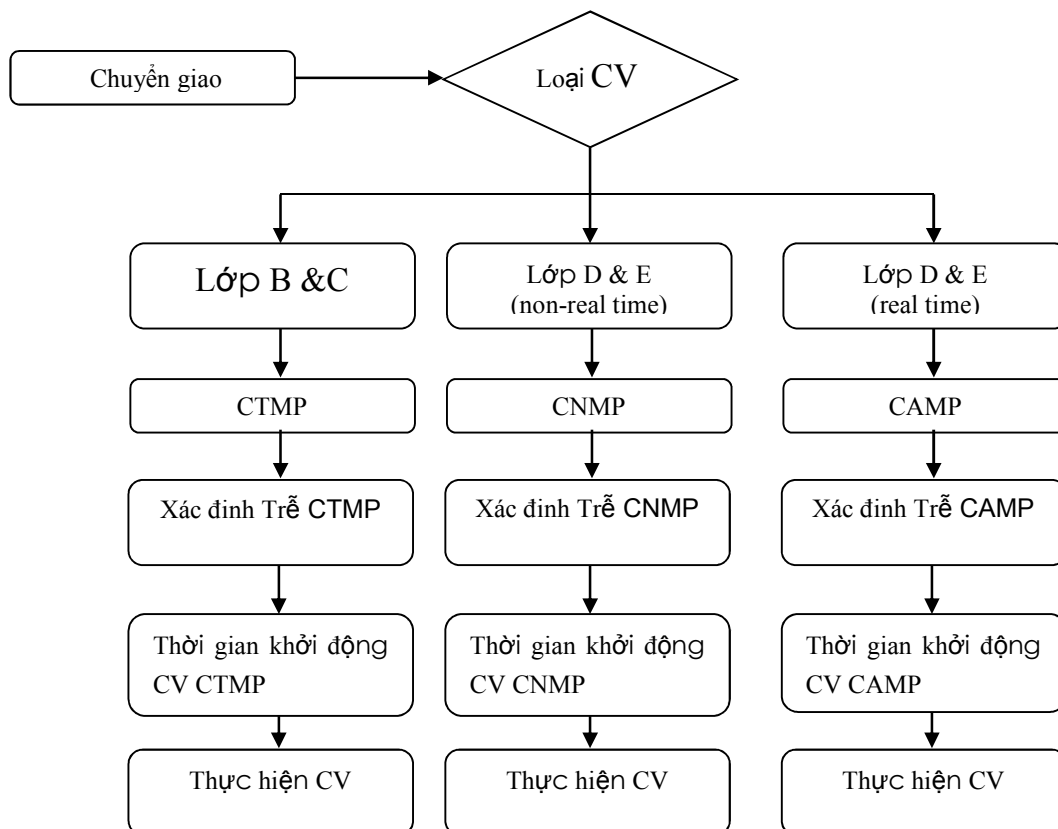
Với  $\tau_{rs}$  là thời gian cần thiết để đăng ký địa chỉ,  $t$  là thời gian để MH di chuyển từ P tới Q. Từ [15] ta có:

$$f_{i(t)} = \begin{cases} \frac{L-d}{\theta\sqrt{v^2t^2 - (L-d)^2}}, & \frac{L-d}{v} < t < \frac{\sqrt{b^2 + (L-d)^2}}{v} \\ 0 & \text{otherwise} \end{cases} \quad (4.26)$$

Với  $\theta = \arctan\left(\frac{b}{l-d}\right)$ ,  $b = \frac{a}{2} + \frac{d}{\sqrt{3}}$  theo hình 4.9a, L là khoảng cách từ P tới S, d là độ dài của

vùng giao nhau, v là tốc độ di chuyển của M. Do đó:

$$\begin{aligned} pe &= \int_0^{\tau_{rs}} f_t(t) \\ &= \int_{\frac{L-d}{v}}^{\tau_{rs}} \frac{L-d}{\pi t \sqrt{v^2t^2 - (L-d)^2}} dt \\ &\approx \frac{1}{\theta} \arccos\left(\frac{L-d}{v\tau_{rs}}\right) \end{aligned} \quad (4.27)$$



Hình 4.13: Sơ đồ hoạt động của AMMS

Trong trường hợp này thì các gói số liệu định trước cho MH bắt đầu đến NN trước khi MH di chuyển vào NN. Do vậy nếu không có tác động thêm nữa, thì các gói số liệu này sẽ bị rớt tại NN cho đến khi MH di chuyển vào vùng giao nhau (hình 4.12a). Trễ chuyển giao của trường hợp này là:

$$T_{he} = \frac{L}{v} - \tau_{rs} + \tau_d + \tau_l \quad (4.28)$$

Với  $\tau_d$  là thời gian cần để dò tìm được NN và  $\tau_l$  là thời gian cần để chuyển giao L2 tới NN.

ii) Hoàn thành việc đăng ký địa chỉ sau khi MH đi vào vùng giao nhau. Ký hiệu  $p_u$  là xác suất xảy ra trường hợp này, và trễ tương ứng là  $T_{hu}$ . Trường hợp này có thể xảy ra 3 tình huống sau:

+ Việc đăng ký địa chỉ hoàn thành trong khi dò tìm hướng di chuyển. Ký hiệu  $p_{u1}$  là xác suất xảy ra trường hợp này, khi đó

$$p_{u1} = p \left( \frac{L-d}{v} < t < \frac{L-d}{v} + \tau_d \right) \quad (4.29)$$

Trễ chuyển giao là

$$T_{u1} = \tau_l + \frac{L-d}{v} + \tau_d - \tau_{rs} \quad (4.30)$$

+ Việc đăng ký địa chỉ hoàn thành sau khi dò tìm hướng di chuyển nhưng trước khi hoàn thành chuyển giao L2. Ký hiệu  $p_{u2}$  là xác suất xảy ra trường hợp này, khi đó

$$p_{u2} = p \left( \frac{L-d}{v} + \tau_d < t < \frac{L-d}{v} + \tau_d + \tau_l \right) \quad (4.31)$$

Trễ chuyển giao là

$$T_{u2} = \tau_l \quad (4.32)$$

+ Việc đăng ký địa chỉ hoàn thành sau khi hoàn thành chuyển giao L2. Ký hiệu  $p_{u3}$  là xác suất xảy ra trường hợp này, khi đó

$$p_{u3} = p \left( \frac{L-d}{v} + \tau_d + \tau_l < t < \frac{L}{v} \right) \quad (4.33)$$

Trễ chuyển giao là

$$T_{u3} = \tau_{rs} - \frac{L-d}{v} - \tau_d \quad (4.34)$$

Do vậy  $p_u$  và  $T_{hu}$  sẽ là

$$p_u = p_{u1} + p_{u2} + p_{u3} \quad (4.35)$$

và

$$T_{hu} = p_{u1}T_{u1} + p_{u2}T_{u2} + p_{u3}T_{u3} \quad (4.36)$$

$p_{u1}$ ,  $p_{u2}$ ,  $p_{u3}$  có thể xác định được thông qua (4.27).

Từ (4.27) và (4.35) ta suy ra

$$P_{c2} = p_e + p_u \quad (4.37)$$



Trễ chuyển giao khi NN dự báo thành công  $T_{hc2}$  sẽ là

$$T_{hc2} = p_e T_{he} + p_u T_{hu} \quad (4.38)$$

Trễ chuyển giao trung bình của CAMP là

$$\hat{T} = p_{cl} T_{hc1} + (1 - p_{cl}) T_{hc2} \quad (4.39)$$

với  $T_{hc1}$  xác định bởi (4.2), và  $T_{hc2}$  xác định bởi (4.38). Tương tự ta xác định được trễ trung bình của CTMP  $\hat{T}_{ht}$  và CNMP  $\hat{T}_{hm}$ . Việc mất số liệu của CNMP và CAMP, và thời gian giảm thông lượng của CTMP phụ thuộc vào trễ chuyển giao như sau.

#### 4.4.1 Mất số liệu của MIP và SIP

Số liệu bị mất trong thời gian CNMP ( $P_{hm}$ ) và CAMP ( $P_{hs}$ ) dựa trên chuyển giao ở sử dụng AMMS sẽ là:

$$P_{hm} = R \hat{T}_{hm} \quad (4.40)$$

và

$$P_{hs} = R \hat{T}_{hs} \quad (4.41)$$

Với  $R$  là tốc độ dữ liệu của kết nối giữa CH và MH

#### 4.4.2 Thời gian giảm thông lượng của TCP-M

Thời gian giảm thông lượng của CTMP là

$$T_t = \hat{T}_{th} + [1 + \log_2 CW_n] RTT_n \quad (4.42)$$

Với  $CW_n$  là kích thước cửa sổ nghẽn ở trạng thái ổn định của TCP trong NN và  $RTT_n$  là RTT của TCP trong NN.

### 4.5 Kết luận chương 4

Chương này, luận án đề xuất *giải pháp AMMS - sử dụng các giao thức khác nhau cho các lớp ứng dụng khác nhau, và phù hợp với cấu trúc mạng truy cập trong quản lý di động*. Qua đó hiệu suất quản lý chuyển giao sẽ được cải thiện dựa trên bộ tham số chuyển giao sử dụng cho các lớp ứng dụng đó. Như vậy, việc sử dụng giao thức chuyển giao sẽ phụ thuộc vào đặc tính của ứng dụng và cấu trúc mạng truy cập.

Mặc dù AMMS hỗ trợ chuyển giao rất tốt cho các lớp ứng dụng, nhưng:

- Với MANET, các giao thức định tuyến chủ động (OLSR, DSDV) phù hợp với các mạng có cấu trúc ổn định, và các giao thức định tuyến theo yêu cầu (AODV) phù hợp với các mạng có tính di động cao.

- Không đủ để hỗ trợ chuyển giao liền mạch (Seamless – là loại chuyển giao yêu cầu trễ rất bé, gần bằng 0. Nghĩa là số liệu không bị mất và thông lượng không bị ảnh hưởng trong quá trình chuyển giao). Để giải quyết vấn đề này luận án đề xuất *phương án chia sẻ thông tin*,

*dự báo lỗi khung (FER) và trễ báo hiệu* để quyết định thời điểm thích hợp kích hoạt chuyển giao, và sử dụng thông tin này để tăng hiệu suất mạng. Phương án này loại trừ được các yếu tố ảnh hưởng đến chất lượng dịch vụ như lỗi khung (FER) và trễ báo hiệu chuyển giao.

Như vậy giải pháp AMMS này có hai ưu điểm nổi bật đó là: (1) Phát triển được ứng dụng hỗ trợ tính di động thích ứng, và (2) Cải thiện hiệu suất chuyển giao thông qua sự tương tác giữa các lớp, và cấu trúc mạng truy cập.

Công trình công bố:

- 1) Le Ngoc Hung, Vu Khanh Quy, “A Review: Performance Improvement Routing Protocols for MANETs”, JCM May 2020

## KẾT LUẬN

Mục tiêu của bài toán Tối ưu hóa quản lý chuyển giao trong mạng vô tuyến hỗn hợp băng rộng đa dịch vụ là xây dựng được giải pháp quản lý chuyển giao mới, linh hoạt cho các loại ứng dụng sử dụng đồng thời, đáp ứng các yêu cầu về QoS và năng lượng tiêu thụ mà không phụ thuộc vào sự di chuyển của thuê bao. Luận án thực hiện khảo sát các công trình nghiên cứu trước đây về lĩnh vực quản lý di động trong mạng di động nói chung và BcN nói riêng ở chương 1 và cho thấy, nhiều công nghệ vô tuyến khác nhau (LTE, 5G, NB-IOT, WLAN, MANETs...) cùng tồn tại trong hạ tầng mạng di động, cùng cung cấp các dịch vụ tương tự nhau (thoại, video, data,...). Vấn đề đặt ra là làm thế nào để khai thác hiệu quả hạ tầng mạng đã đầu tư mà vẫn đáp ứng nhu cầu sử dụng dịch vụ đa dạng, chất lượng, mọi lúc, mọi nơi của người dùng.

Luận án đã phân tích và đánh giá các yếu tố ảnh hưởng tới chất lượng dịch vụ khi chuyển giao, mức độ ảnh hưởng của các loại giao thức tới các loại ứng dụng khác nhau, tương quan giữa năng lượng tiêu thụ và hiệu suất mạng, từ đó rút ra được các yếu tố ảnh hưởng tới hiệu suất chuyển giao của một giao thức quản lý di động đó là (i) *Xác suất thất lạc gói tin số liệu*; (ii) *Độ trễ báo hiệu và truyền bản tin kết cuối*; (iii) *Các công nghệ truy cập lớp kết nối*; (iv) *Loại ứng dụng*; (v) *Năng lượng tiêu thụ và hiệu suất mạng*. Phân tích và đánh giá hiệu suất quản lý di động cho nhiều loại ứng dụng khác nhau (A,B,C,D,E) của các giao thức đã đề xuất như MIP, TCP-M, SIP. Từ đó chứng minh được các định lý liên quan tới việc *thất lạc gói tin* và *độ trễ trung bình*, làm tiền đề xác định các yếu tố cơ bản ảnh hưởng tới QoS khi chuyển giao, đó là: *xác suất gói tin bị thất lạc*; *độ trễ truyền tin*; và *trung bình độ trễ truyền bản tin báo hiệu*.

### Các kết quả đạt được của luận án:

1. Đề xuất giao thức định tuyến theo yêu cầu - EEMA cho MANETs. EEMA chọn tuyến tối ưu cho chuyển giao dựa trên: số bước nhảy và hàm chi phí, và cân đối giữa trễ và năng lượng tiêu thụ.
2. Dựa trên các yếu tố cơ bản ảnh hưởng tới QoS khi chuyển giao và sử dụng lý thuyết Bayes để tính toán *xác suất chuyển giao*  $P_b$ . Xây dựng cơ chế *đặt trước băng thông* cho các ứng dụng có  $P_b$  lớn hơn ngưỡng chuyển giao, nhằm duy trì QoS cho các ứng dụng này.
3. Lựa chọn các giao thức phù hợp với cấu trúc mạng, đó là OLSR và DSDV cho mạng có cấu trúc ổn định, tính di động thấp, và AODV cho cấu trúc mạng có tính di động cao.
4. Đề xuất giải pháp *quản lý chuyển giao linh hoạt* (AMMS) nhằm khai thác hiệu quả tài nguyên mạng, đáp ứng yêu cầu QoS và mức tiêu thụ năng lượng cho các loại ứng dụng

A,B,C,D,E. Phương án này loại trừ được các yếu tố ảnh hưởng đến chất lượng dịch vụ như lỗi khung (FER), xác suất mất số liệu, trễ báo hiệu chuyển giao. Như vậy phương pháp AMMS này có hai ưu điểm nổi bật đó là: (1) *Phát triển được ứng dụng hỗ trợ tính di động thích ứng*, và (2) *Cải thiện hiệu suất chuyển giao thông qua sự tương tác giữa các lớp*.

Trong thời gian thực hiện luận án, với sự hỗ trợ của người hướng dẫn, tác giả cùng các cộng sự đã công bố một số công trình nghiên cứu trên các hội nghị, tạp chí trong và ngoài nước. Các công trình này được liệt kê tại phần DANH MỤC CÁC CÔNG TRÌNH CÔNG BỐ dưới đây, và được sử dụng trong việc hoàn thành luận án tiến sĩ này.

## HƯỚNG NGHIÊN CỨU TIẾP THEO

Đối với cơ chế điều khiển đăng nhập dựa trên các mức ưu tiên được trình bày trong luận án mới là bước đầu trong việc xây dựng một thuật toán điều khiển chuyển giao liên mạng, trong tương lai phân tích toán học và mô phỏng sẽ được phát triển cho hệ thống với môi trường di động không đồng nhất, các lưu lượng hỗn hợp được tạo ra từ các nguồn lưu lượng di động với tốc độ di chuyển thay đổi (mô phỏng với các môi trường như đường cao tốc hay trong một khu công nghiệp- tốc độ người đi bộ).

Cơ chế điều khiển đăng nhập kết hợp với SLA đã đề xuất biện pháp điều khiển dựa trên điều khiển các bộ đệm phụ thêm cho các mức SLA, trong tương lai sẽ phát triển phương pháp thống nhất tham số SLA toàn hệ thống (SLA giữa mạng-mạng và người sử dụng-mạng) để đảm bảo hỗ trợ tối đa giữa các mức.

Đảm bảo chất lượng dịch vụ trong mạng lõi là vấn đề lớn còn rất nhiều tồn tại. Tuy đã đề xuất biểu diễn toán học đối với môi trường hỗn hợp, tuy nhiên khả năng áp dụng để điều khiển mạng là khó khăn do yêu cầu năng lực tính toán cao và thời gian tính toán dài. Trong tương lai phương pháp rút gọn toán học với một số các điều kiện biên để giảm tính hỗn hợp của nguồn lưu lượng sẽ được nghiên cứu.

Tác giả cũng sẽ tập trung nghiên cứu vấn đề bảo mật khi chia sẻ thông tin giữa các lớp mạng và lựa chọn các giao thức chuyển giao cho các lớp ứng dụng, đánh giá hiệu suất của AMMS dựa trên mức độ di động cao hơn.

## DANH MỤC CÁC CÔNG TRÌNH CÔNG BỐ

1. Lê Ngọc Hưng, Nguyễn Xuân Quỳnh, “Nhận dạng và phân lớp các yếu tố ảnh hưởng tới điều khiển chuyển giao”, Hội nghị FAIR lần 10, tháng 8/2017, Đà Nẵng.
2. Dzung Van Dinh, Byeong-Nam Yoon, Hung Ngoc Le, Uy Quoc Nguyen, Khoa Dang Phan, Lam Dinh Pham, “ICT Enabling Technologies for Smart Cities”, 2018 20th International Conference on Advanced Communication Technology (ICACT), IEEE, Feb. 2018, Korea.
3. Vu Khanh Quy<sup>1</sup>, Le Ngoc Hung<sup>2</sup>, Nguyen Dinh Han<sup>3</sup>, “CEPRM: A Cloud-assisted Energy-Saving and Performance-Improving Routing Mechanism for MANETs” JCM 15 Nov. 2019
4. Ngoc Hung Le, “An approach to handover bandwidth reservation in the wireless Future Convergence Network”, NICS Dec.2019.
5. Vu Khanh Quy, Le Ngoc Hung, “A Trade-off between Energy Efficient and High-Performance in Routing for Mobile Ad hoc Networks”, JCM Mar. 2020.
6. Le Ngoc Hung, Vu Khanh Quy, “A Review: Performance Improvement Routing Protocols for MANETs”, JCM May 2020

## TÀI LIỆU THAM KHẢO

- [1] Akyildiz, I. F. and Wang, W., "A Predictive User Mobility Profile for Wireless Multimedia Networks," *IEEE/ACM Transactions on Networking*, vol. 12, no. 6, pp. 1021-1035, 2004.
- [2] Akyildiz, I. F., Xie, J., and Mohanty, S., "A survey on mobility management in next generation all-IP based wireless systems," *IEEE Wireless Communications*, vol. 11, no. 4, pp. 16-28, 2004.
- [3] Akyildiz, I. F., Morabito, G., and Palazzo, S., "TCP-Peach: a new congestion control scheme for satellite IP networks," *IEEE/ACM Trans. Networking*, vol. 9, no. 3, pp. 307-321, 2001.
- [4] Almes, G., Kalidindi, S., and M. Zekauskas, "A One-way Delay Metric for IPPM", RFC 2679, September 1999. A Predictive User Mobility Profile
- [5] Almes, G., Kalidindi, S., and M. Zekauskas, "A One-way Packet Loss Metric for IPPM", RFC 2680, September 1999.
- [6] Almes, G., Kalidindi, S., and M. Zekauskas, "A Round-trip Delay Metric for IPPM", RFC 2681, September 1999.
- [7] Aqeel Taha, Raed Alsaqour, Mueen Uddin et al., "Energy Efficient Multipath Routing Protocol for Mobile Ad-Hoc Network Using the Fitness Function", *IEEE Access*, Vol. 5, pp. 10369-10381, 2017.
- [8] Bao, G., "Performance evaluation of TCP/RLP protocol stack on Correlated Fading DS-CDMA wireless links," *IEEE Veh. Tech. Conf.*, Ottawa, May 1998.
- [9] B.Kosko, "Digital Signal Processing and Neural Networks", Englewood Cliffs, NJ: Prentice-Hall, 1991.
- [10] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update 2017–2022," Cisco, San Jose, CA, USA, White Paper c11738429, Feb. 2019.
- [11] C. Perkins, P. R. C. and Bharatia, J., "Mobile IPv4 challenge/response extensions (revised)," RFC 4721, Jan 2007.
- [12] Cung Trong Cuong, Vo Thanh Tu, Nguyen Thuc Hai, "MAR-AODV: Innovative Routing Algorithm in MANET Based on Mobile Agent". *IEEE WAINA*, Spain, 2013, pp. 62-66.
- [13] Conrad, P. and et al, "SCTP in battle field networks," in *Proc. of IEEE MILCOM*, pp. 289-95.
- [14] C.C Wang, K.H Chi, "An IP-Decoupling Approach to Host Mobility," *Journal of information science and engineering* 23, 91-112, 2007.
- [15] Chung Hung Liao, "A Cross-Layer (Layer 2 + 3) Handoff Management Protocol for Next Generation Wireless Systems," *IEEE Trans. on Mobile Computing*, 2007.
- [16] D.S.J. De Couto, D. Aguayo, J. Bicket, R. Morris, "A High Throughput Path Metric for Multi-Hop Wireless Routing". In *Proceedings of ACM MobiCom*, USA, 2003, pp. 134-146.

- [17] D. Lopez-Perez, I. Guvenc, and X. Chu, "Mobility management challenges in 3GPP heterogeneous networks," *IEEE Commun. Mag.*, vol. 50, no. 12, pp. 70–78, Dec. 2012.
- [18] D. Munoz and K.W. Cattermole, "Hand off procedure for fuzzy delined radio cells," in *Proc. IEEE 37th VTC*, 1987, pp. 38–42.
- [19] D.Oliva and J.I.Alonso, "A two-hop MIMO relay architecture using LTE and millimeter wave bands in high speed trains," in *Proc. IEEE WCNC Conf.*, Apr. 2018, pp. 1–6.
- [20] Evolved Universal Terrestrial Radio Access (E-UTRA): Requirements for Support of Radio Resource Management, 3GPP TS 36.133, v. 13.3.0, Rel. 13, May. 2016
- [21] E. Onggosanusi, S. Rahman, L. Guo, Y. Kwak, H. Noh, Y. Kim, S. Faxer, M. Harrison, M. Frenne, S. Grant, R. Chen, R. Tamrakar, and Q. Gao, "Modular and high-resolution channel state information and beam management for 5G new radio," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 48–55, Mar. 2018.
- [22] Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) Procedures in Idle Mode, document 3GPP 36.304, Mar. 2019
- [23] F. Baboescu, S. Kanugovi, J. Zhu, "Multi-Access Management Services (MAMS)", RFC 8743, Mar. 2020
- [24] Geetha N., Sankar A., "Hop Count Based Energy Saving Dynamic Source Routing Protocol for Ad Hoc Network", *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, Vol. 108, Springer, Berlin, Heidelberg, 2012.
- [25] Gustafsson, E., Jonsson, A., and Perkins, C. E., "Mobile IPv4 regional registration," RFC 4857, Jun 2007.
- [26] Hashimoto et al., "Evaluation of Mobile Agent-Based Service Dissemination Schemes in MANETs". In *Proceedings of IEEE ICNC, China*, 2011, pp. 257-260.
- [27] Helal S., Lee C., Zhang Y., G.G. Richard I Q "An Architecture for Wireless LANIWAN Integration," *IEEE Wireless Communication and Networking Conference (WCNC) 2000*. Chicago, Illinois, September 2000.
- [28] Hsieh, H.-Y., Kim, K.-H., and Sivakumar, R., "An End-to-End Approach for Transparent Mobility across Heterogeneous Wireless Networks," *ACM/Kluwer Mobile Networks and Applications Journal (MONET)*, Special Issue on Integration of Heterogeneous Wireless Technologies, vol. 9, no. 4, pp. 363-378, 2004.
- [29] I.V Hwang, R.G Harry, X.Z Wu, "The Global Information Technology Report", 2012
- [30] Ishizuka et al., "A Mobile Agent Creation Mechanism for Service Collection and Dissemination in Heterogeneous MANETs". In *Proceedings of IEEE ICNC, USA*, 2012, pp. 321-322.



- [31] Inamura, H., Montenegro, G., Ludwig, R., Gurtov, A., and Khafizov, F., "TCP over second (2.5G) and third (3G) generation wireless networks," RFC 3481 Feb. 2003.
- [32] ITU, Q.1704 **and** M.1645, 10/2008
- [33] J. Rodriguez, A. Radwan, C. Barbosa, F. H. P. Fitzek, R. A. Abd-Alhameed, J. M. Noras, S. M. R. Jones, I. Politis, P. Galiotos, G. Schulte, A. Rayit, M. Sousa, R. Alheiro, X. Gelabert, and G. P. Koudouridis, "SECRET—Secure network coding for reduced energy next generation mobile small cells: A European training network in wireless communications and networking for 5G," in Proc. IEEE Internet Technol. Appl. (ITA) Conf., Sep. 2017, pp. 329–333.
- [34] J.Kellokoski, "Managing Mobility in an Always-Best-Connected IP Network," Copyright © 2013, by University of Jyväskylä.
- [35] Johnson, D., Perkins, C., and J. Arkko, "Mobility Support in IPv6", RFC 3775, June 2004.
- [36] Jyh-Cheng Chen and Tao Zhang, "IP-Based Next-Generation Wireless Networks: Systems, Architectures, and Protocols", John Wiley & Sons Inc., 2004.
- [37] J. Liu, K. Au, A. Maaref, J. Luo, H. Baligh, H. Tong, A. Chassaigne, and J. Lorca, "Initial access, mobility, and user-centric multi-beam operation in 5G new radio," IEEE Commun. Mag., vol. 56, no. 3, pp. 35–41, Mar. 2018.
- [38] K. Kanwal and G. A. Safdar, "Energy efficiency and superlative TTT for equitable RLF and ping pong in LTE networks," Mobile Netw. Appl., vol. 23, no. 6, pp. 1682–1692, 2018.
- [39] LTE; Evolved Universal Terrestrial Radio Access Network (E-UTRAN); X2 General Aspects and Principles, document ETSI TS 136 420 V10.2.0 (2011-10), Jun. 2018.
- [40] LTE; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall Description; Stage 2, document ETSI TS 136 300 V15.3.0 (2018-10), Mar. 2019, pp. 105–158.
- [41] Ministry of Internal Affairs and Communications, Japan, "Outline of the 2018 White Paper on Information and Communications in Japan", Jul. 2018.
- [42] Ministry of Information and Communications, "White book of Vietnam Information and Communication technology", 2017.
- [43] M.auridsen, L. C. Giménez, I. Rodriguez, T. B. Sorensen, and P.Mogensen, "From LTE to 5G for connected mobility," IEEE Commun. Mag., vol. 55, no. 3, pp. 156–162, Mar. 2017.
- [44] M. Polese, M. Giordani, M. Mezzavilla, S. Rangan, and M. Zorzi, "Improved handover throughl connectivity in 5G mm Wave mobile networks," IEEE J. Sel. Areas Commun., vol. 35, no. 9 pp. 2069–2084, Sep. 2017.

- [45] M. Joud, M. García-Lozano, and S. Ruiz, "User specific cell clustering to improve mobility robustness in 5G ultra-dense cellular networks," in Proc. 14th Annu. Conf. Wireless On-Demand Netw. Syst. Services (WONS), Feb. 2018, pp. 45–50.
- [46] M. Tayyab, G. P. Koudouridis, and X. Gelabert, "A simulation study on LTE handover and the impact of cell size," in Proc. Broadband Commun., Netw., Syst. (BROADNETS), 2018, pp. 398–408.
- [47] Mohanty, S., "VEPSD: Velocity estimation using the PSD of the received signal envelope in next generation wireless systems," IEEE Transactions on Wireless Communications, 2005.
- [48] Maltz, D. and Bhagwat, P., "MSOCKS: an architecture for transport layer mobility," in Proc. of IEEE INFOCOM'98, pp. 1037-45.
- [49] McNair, J., Akyildiz, I. F., and Bender, M., "An Inter-System Handoff Technique for the IMT-2000 System," in Proc. of IEEE INFOCOM'00.
- [50] Mohanty, S. and Akyildiz, I. F., "Performance Analysis of Handoff Techniques Based on Mobile IP, TCP-Migrate, and SIP," IEEE Trans. on Mobile Computing, Jul 2007.
- [51] Misra, A., Das, S., Dutta, A., McAuley, A., and Das, S., "IDMP-based fast handoffs and paging in IP-based 4G mobile networks," IEEE Communications Magazine, 2002.
- [52] M. Tayyab, X. Gelabert, and J. Riku, "A Survey on Handover Management: From LTE to NR" IEEE, 2019.
- [53] M. Alhabo, L. Zhang, and O. Oguejiofor, "Inbound handover interference-based margin for load balancing in heterogeneous networks," in Proc. Int. Symp. Wireless Commun. Syst. (ISWCS), Bologna, Italy, Aug. 2017, pp. 146–151.
- [54] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, "A tutorial on beam management for 3GPP NR at mmWave frequencies," IEEE Commun. Surveys Tuts., vol. 21, no. 1, pp. 173–196, 1st Quart., 2018.
- [55] Outline of the 2019 White Paper on Information and Communications in Japan, Jul. 2019
- [56] Ohba, Y., "A Framework of Media-Independent Pre- Authentication (MPA)", draft-ohba-mobopt-mpa-framework-01 (work in progress), July 2005.
- [57] Papoulis, A. and Pillai, S. U., Probability, random variables, and stochastic processes. Mc Graw Hill, 4<sup>th</sup> edition, 2002.
- [58] Perkins, C., "IP mobility support for IPv4," RFC 3220, IETF, 2002.
- [59] Perkins, C., "IP Mobility Support for IPv4", RFC 3344, August 2002.
- [60] P. T. Dat, A. Kanno, K. Inagaki, F. Rottenberg, N. Yamamoto, and T. Kawanishi, "High-speed and uninterrupted communication for high-speed trains by ultrafast WDM fiber–wireless backhaul system," J. Lightw. Technol., vol. 37, no. 1, pp. 205–217, Jan. 1, 2019.

- [61] R. Arshad, H. Elsayy, S. Sorour, T. Y. Al-Naffouri, and M.-S. Alouini, "Handover management in 5G and beyond: A topology aware skipping approach," *IEEE Access*, vol. 4, pp. 9073–9081, 2016.
- [62] RFC3561, "<https://www.ietf.org>", accessed 10/5/2019
- [63] Ramjee, R., Varadhan, K., Salgarelli, L., Thuel, S. R., Wand, S.-Y., and Porta, T. L., "HAWAII: A domain-based approach for supporting mobility in wide-area wireless networks," *IEEE/ACM Transactions on Networking*, vol. 10, no. 3, pp. 396-410, 2002.
- [64] Rigney, C. and et al, "Remote Authentication Dial In User Service (RADIUS)," RFC 2865, 2000.
- [65] Rosenberg, J. and et al., "SIP: Session Initiation Protocol," RFC 3261, IETF, 2002.
- [66] Shanzhi Chen, Yan Shi, Bo Hu, and Ming Ai, "Mobility-Driven Networks (MDN): From Evolution to Visions of Mobility Management", *IEEE*, Aug. 2014.
- [67] S. Chaudhuri, I. Baig, and D. Das, "Self organizing method for handover performance optimization in LTE-advanced network," *Comput. Commun.*, vol. 10, pp. 151–163, Sep. 2017.
- [68] S. Banks, "Signal Processing, Image Processing and Pattern Recognition", Englewood Cliffs, NJ: Prentice-Hall, 1990, pp. 262–254.
- [69] S Chakraborty, S Sen, "Quantitative analysis on Log-Normal Indoor Propagation Model for WLAN", *International Journal of Computer science and Technology*, Vol.5, Issue 1, Mar. 2014.
- [70] Salkintzis, A. K., Fors, C., and Pazhyannur, R., "WLAN-GPRS Integration for Next-Generation Mobile Data Networks," *IEEE Wireless Communications*, 2002.
- [71] Shan Jaffry, Rasheed Hussain, Xiang Gui, "A Comprehensive Survey on Moving Networks", *Computer Science*, 2020.
- [72] Schulzrinne, H. and E. Wedlund, "Application Layer Mobility Using SIP", *ACM MC2R*.
- [73] Stemm, M. and Katz, R. H., "Vertical handoffs in wireless overlay networks," *ACM/Springer Journal of Mobile Networks and Applications (MONET)*, vol. 3, no. 4, pp. 335-350, 1998.
- [74] Snoeren, A. C. and Balakrishnan, H., "An end-to-end approach to host mobility," *ACM 2000 1-58113-197-6/00/08*.
- [75] S. Kutty and D. Sen, "Beamforming for millimeter wave communications: An inclusive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 949–973, 2nd Quart., 2016.
- [76] Stuber, G. L., "Principles of Mobile Communication", Springer International Publishing AG 2017
- [77] T. Bilen, B. Canberk, and K. R. Chowdhury, "Handover management in software-defined ultra-dense 5G networks," *IEEE Netw.*, vol. 31, no. 4, pp. 49–55, Jul./Aug. 2017.
- [78] Tepedelenlioglu, C. and Giannakis, G. B., "On Velocity Estimation and Correlation Properties of Narrow-Band Mobile Communication Channels," *IEEE Trans. on Vehicular Technology*, vol. 50, no. 4, pp. 1039-1052, 2001.

- [79] Tsao, S. and Lin, C., "Design and Evaluation of UMTS-WLAN Interworking Strategies," in Proc. IEEE VTC 2002.
- [80] Valko, A., "Cellular IP: A new approach to Internet host mobility," ACM SIGMOBILE Computer Communication Review, vol. 29, no. 1, pp. 50-65, 1999.
- [81] Vu Khanh Quy, Nguyen Tien Ban, Nguyen Dinh Han, "A High-Performance Routing Protocol for Multimedia Applications in MANETs," Journal of Communications, Vol. 14, No. 4, pp. 267-274, 2019.
- [82] Vu Khanh Quy, Nguyen Tien Ban, Nguyen Dinh Han, "An Advanced Energy Efficient and High-Performance Routing Protocol for MANET in 5G," Journal of Communications, Vol. 13, No.12, pp. 743-749, 2018.
- [83] V. K. Quy, N. D. Han, and N. T. Ban, "PRP: A highperformance routing protocol for mobile ad-hoc networks," in Proc. International Conference on Advanced Technologies for Communications (ATC), Ho Chi Minh City, 2018, pp. 226-231.
- [84] Wei Sun, Zheng Yang, Xinglin Zhang et al., "Energy-Efficient Neighbor Discovery in Mobile Ad Hoc and Wireless Sensor Networks: A Survey", IEEE Communications Surveys & Tutorials, Vol. 16, Iss. 3, pp. 1448-1459, 2014.
- [85] W. Guo, W. Zhang, P. Mu, F. Gao, and H. Lin, "High-mobility wideband massive MIMO communications: Doppler compensation, analysis and scaling laws," IEEE Trans. Wireless Commun., vol. 18, no. 6, pp. 3177–3191, Jun. 2019.
- [86] Wedlund, E. and Schulzrinne, H., "Mobility support using SIP," in Proc. of Second ACM/IEEE International Conference on Wireless and Mobile Multimedia (WoWMoM'99).
- [87] Wisely, D. and Mitjana, E., "Paving the Road to Systems Beyond 3G-The IST BRAIN and MIND Projects," Journal of Communications and Networks, vol. 4, no. 4, pp. 292-301, 2002.
- [88] Xie, J. and Akyildiz, I. F., "A hybrid control resource allocation scheme for policy-enabled handoff in wireless heterogeneous overlay networks," submitted for publication, 2005.
- [89] Y.S.Hussein, B. M. F. A. Ali, A.Sali, and A. M. Mansoor, "A novel cell selection optimization handover for long-term evolution (LTE) macro cell using fuzzy TOPSIS," J. Comput. Commun., vol. 73, pp. 22–33, Jan. 2016.
- [90] Zhao, X., Castelluccia, C., and Baker, M., "Flexible network support for mobility," in Proc. of ACM Mobicom
- [91] 5GMF White Paper, Sep. 2017
- [92] 5G; NR; Radio Resource Control (RRC); Protocol Specification, document 3GPP TS 38.331 Version 15.3.0 Release 15, Jan. 2018.