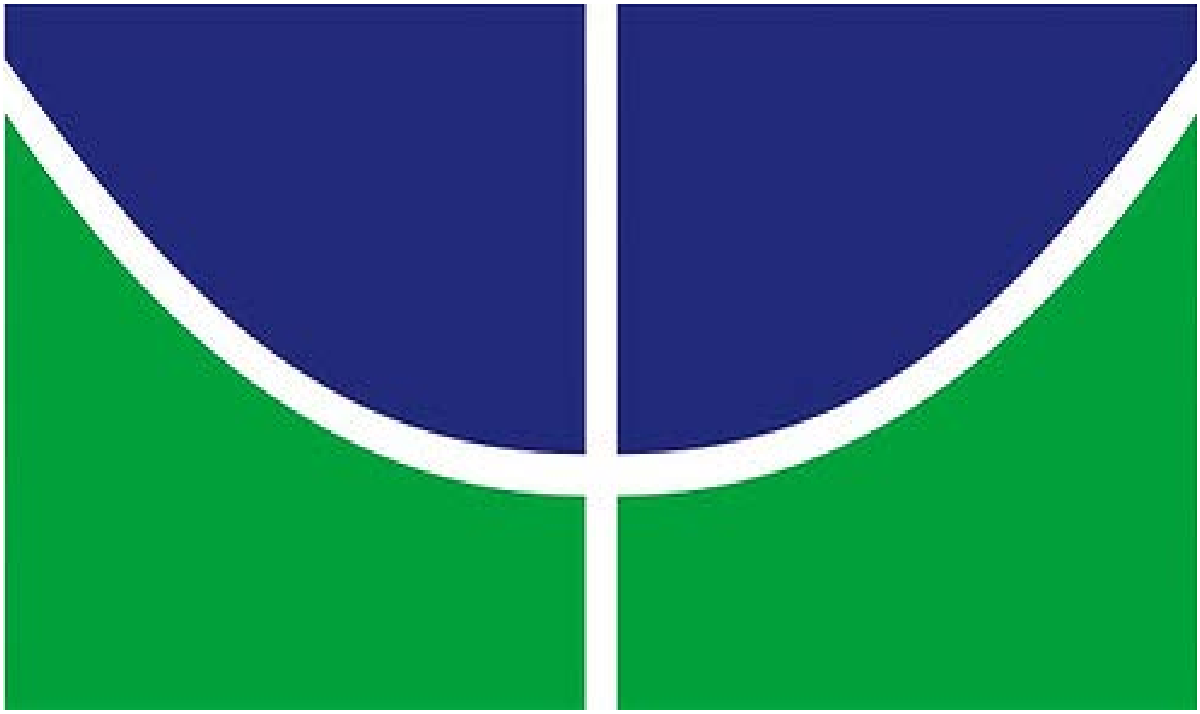

Arquitetura de Rede com Suporte à Mobilidade de Fluxo IP e ao Gerenciamento de Mobilidade Distribuído (DMM)

YARISLEY PEÑA LLERENA

DISSERTAÇÃO DE MESTRADO EM ENGENHARIA ELÉTRICA

DEPARTAMENTO DE ENGENHARIA ELÉTRICA

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASÍLIA

UNIVERSIDADE DE BRASÍLIA

FACULDADE DE TECNOLOGIA

DEPARTAMENTO DE ENGENHARIA ELÉTRICA

Arquitetura de Rede com Suporte à Mobilidade de Fluxo IP e ao Gerenciamento de Mobilidade Distribuído (DMM)

YARISLEY PEÑA LLERENA

ORIENTADOR: PAULO ROBERTO DE LIRA GONDIM

DISSERTAÇÃO DE MESTRADO EM ENGENHARIA ELÉTRICA

PUBLICAÇÃO: PPGEE.DM – 619/2016

BRASÍLIA/DF: FEVEREIRO - 2016

UNIVERSIDADE DE BRASÍLIA

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

ARQUITETURA DE REDE COM SUPORTE À MOBILIDADE DE FLUXO IP E AO GERENCIAMENTO DE MOBILIDADE DISTRIBUÍDO (DMM)

YARISLEY PEÑA LLERENA

DISSERTAÇÃO DE MESTRADO SUBMETIDA AO DEPARTAMENTO DE ENGENHARIA ELÉTRICA DA FACULDADE DE TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE.

APROVADA POR:

PAULO ROBERTO DE LIRA GONDIM, Dr., ENE/UNB
(ORIENTADOR)

PAULO HENRIQUE PORTELA DE CARVALHO, Dr., ENE/UNB
(EXAMINADOR INTERNO)

ANDRÉ COSTA DRUMMOND, Dr., CIC/UNB
(EXAMINADOR EXTERNO)

Brasília, 25 de fevereiro de 2016.

FICHA CATALOGRÁFICA

LLERENA, YARISLEY PEÑA

Arquitetura de Rede com Suporte à Mobilidade de Fluxo IP e ao Gerenciamento de Mobilidade Distribuído (DMM) [Distrito Federal] 2016.

xvi, 132p.,210 x 297 mm (ENE/FT/UnB, Mestre, Engenharia Elétrica, 2016).

Dissertação de Mestrado – Universidade de Brasília.

Faculdade de Tecnologia.

Departamento de Engenheira Elétrica.

1.Escoamento de dados | 2. *Handover*

3.DMM | 4. Mobilidade de fluxo IP

REFERÊNCIA BIBLIOGRÁFICA

LLERENA, Y. P. (2016). Arquitetura de Rede com Suporte à Mobilidade de Fluxo IP e ao Gerenciamento de Mobilidade Distribuído (DMM). Dissertação de Mestrado em Engenharia Elétrica, Publicação PPGEE.DM – 619/2016, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 132p.

CESSÃO DE DIREITOS

AUTOR: Yarisley Peña Llerena.

TÍTULO: "Arquitetura de Rede com Suporte à Mobilidade de Fluxo IP e ao Gerenciamento de Mobilidade Distribuído (DMM)".

GRAU/ANO: Mestre/2016

Yarisley Peña Llerena
UNIVERSIDADE DE BRASÍLIA
Faculdade de Tecnologia.
Departamento de Engenharia Elétrica.
70910-900 – Brasília – DF – Brasil.

Dedicado a mi esposo Noel y a mi mamá Maricela, por el amor y el apoyo que me han brindado siempre, por la confianza que tuvieron em mi, por saber soportar mis momentos de estrés y alentarme a seguir hasta el final. Con todo mi amor para los dos.

AGRADECIMENTOS

Agradeço a Deus por ter me dado a oportunidade de cumprir o sonho de ser mestre.

Ao meu esposo por estar ao meu lado, pela sua compreensão, pela sua ajuda, pela sua confiança em min e sobre tudo pelo seu amor.

A minha mãe, meus sogros, minha tia Teresa, meus primos e resto da família em Cuba por ter me apoiado neste tempo que tenho estado longe deles, mas mesmo estando longe me fizeram sentir que estavam por perto, me cuidando e me acompanhando.

Ao meu orientador, pela sua paciência, pelo apoio, por ter me recebido como sua orientanda e ter me guiado no desenvolvimento do meu projeto de mestrado, Prof. Paulo muito obrigada.

Às pessoas que conheci em Brasília e às que conviveram comigo no dia a dia fazendo me sentir como em família com elas. Obrigada Ruben, Oisys, Elier, Matilde pelo seu apoio, carinho e bons conselhos.

Ao pessoal do Laboratório de Redes da UnB e ao Programa de Pós-graduação em Engenharia Elétrica da UnB.

Ao Programa PEC-PG do Conselho Nacional de Desenvolvimento a Pesquisa (CNPq), pelo apoio financeiro.

A todos os que de algum jeito se interessaram pelo sucesso deste trabalho.

Muito obrigada, Yaris.

RESUMO

Arquitetura de Rede com Suporte à Mobilidade de Fluxo IP e ao Gerenciamento de Mobilidade Distribuído (DMM)

Nos últimos anos o desenvolvimento de aplicações e novos dispositivos móveis tem trazido um aumento no tráfego de dados nas redes celulares. Para resolver esta problemática o escoamento é apresentado como uma solução viável, com o fim de controlar a sobrecarga das redes e oferecer melhor qualidade dos serviços. Algumas arquiteturas tem sido propostas para fazer escoamento de tráfego, suportando mobilidade de fluxo.

Por outro lado, diversos protocolos para gerenciamento de mobilidade têm sido propostos, sendo os principais os protocolos MIP (*Mobile IP*) [1-2] (que é baseado no host) e PMIPv6 (*Proxy MIP*) [3] (que é baseado na rede). Ditos protocolos são centralizados, onde todo o tráfego de dados atravessa a mesma entidade de rede, a qual também é encarregada do controle de mobilidade. Esta característica leva a problemas como: ter um ponto único de falha, escalabilidade e confiabilidade. Com o propósito de resolver esses problemas, surgiu o gerenciamento distribuído de mobilidade (DMM, do inglês *Distributed Mobility Management*) [4], tendendo a se mostrar uma melhor alternativa, considerando os problemas em questão.

Nesta dissertação propomos uma arquitetura para suporte à mobilidade de fluxos IP e ao gerenciamento distribuído de mobilidade, SIFDMM (*Seamless IP Flow and Distributed Mobility Management*), que se caracteriza por: i) é baseada no protocolo PMIPv6, e portanto, baseada na rede; ii) o gerenciamento de mobilidade é totalmente distribuído; iii) é assistida pelo terminal móvel, que atua provendo informações de localização relativas a roteadores de acesso recentemente visitados; iv) integra funcionalidades de gerenciamento de mobilidade e de escoamento de tráfego de dados em uma entidade da rede; v) o escoamento é feito na camada de rede movimentando fluxos IP.

Uma comparação entre a arquitetura proposta e duas outras arquiteturas (centralizada e parcialmente distribuída) é feita, por meio de modelagem analítica e simulação, sendo mostrada redução da latência de escoamento de fluxo (seletivo ou total), em relação a soluções centralizadas ou parcialmente distribuídas baseadas em PMIPv6. Os resultados mostraram também, que a arquitetura proposta apresenta um bom desempenho na transmissão de vídeo streaming.

ABSTRACT

In the recent years, the development of new applications and mobile devices, has brought an increase in data traffic on mobile networks. To solve this problem, the offloading is presented as a viable solution in order to control the overload of the network and offer better quality of services. Some architectures have been proposed to make flow traffic, supporting flow mobility.

On the other hand, several protocols for mobility management as MIP (Mobile IP) [1-2] (which is host-based) and PMIPv6 (Proxy MIP) [3] (which is network based) have been proposed. Said protocols are centralized, where all the data traffic goes through the same network entity, which is also in charge of mobility control. This feature leads to problems such as having a single point of failure, scalability and reliability. In order to solve these problems, came the distributed mobility management (DMM) [4], tending to show a better alternative, considering the problems relative to centralized proposals.

In this dissertation we propose an architecture, SIFDMM (Seamless IP Flow and Distributed Mobility Management), to support distributed mobility management and IP flows mobility. The architecture is characterized by: i) is based on PMIPv6 protocol, and therefore, based on the network; ii) the mobility management is fully distributed; iii) it is assisted by the mobile terminal that supplies location information related to recently visited access routers; iv) integrates features of mobility management and data offloading on a network entity; v) the offloading is done at the network layer moving the IP flows.

A comparison of the proposed architecture and two other architectures (centralized and partially distributed) is performed through analytical modeling and simulation. It is displayed a reduce in the flow handover latency (selective or total) in the proposed architecture in relation to the partially and centralized ones. The result also showed that the proposed architecture has good performance in video streaming transmission.

SUMÁRIO

LISTA DE TABELAS

LISTA DE FIGURAS

LISTA DE ABREVIAÇÕES

3GPP: 3rd Generation Partnership Project. ,

AMA: Access Mobility Anchors.

AP: AccessPoint.

ATT: Access Technology Type.

BA: Binding Acknowledgments.

BC: Binding Cache.

BID: Binding Identification number.

BSA: Basic Service Area.

BSS: Basic Service Set.

BU: Binding Update.

CMM: Centralized Mobility Management.

CN: Correspondent Node.

CoA: Care of Address.

DMM: Distributed Mobility Management.

DSMIPv6: Dual-Stack Mobile IPv6.

eNB: evolved NodeB.

EPC: Evolved Packet Core Network.

EPS: Evolved Packet System.

E-UTRAN: Evolved Universal Terrestrial Access Network.

FA: Foreing Agent.

FB: Flow Binding.

FBT: Flow Binding Table.

F-DMM: Fully - Distributed Mobility Management.

FMA: Flow and Mobility Anchor.

FMIPv6: Fast Mobile IPv6.

FPMIPv6: Fast Proxy Mobile IPv6.

HA: Home Agent.

HMIPv6: Hierarchical Mobile IPv6.

HoA: Home Address.

HSPA: High Speed Packet Access.

IEEE: Institute of Electrical and Electronics Engineers.

IFOM: IP Flow Mobility.

IFT: InterfaceTable.

IP: Internet Protocol.

LabTVDi: Laboratório de Televisão Digital Interativa.

LIPA: Local IP Access.

LL-ID: Link Layer Identifier.

LMA: Local Mobility Anchor.

LTE: Long Term Evolution.

MAG: Mobile Access Gateway.

MCF: Mobile Control Funtion.

MFR: Mobile Function Routers.

MIH: Media Independent Handover.

MIP: Mobile IP.

MIPv4: Mobile IPv4.

MIPv6: Mobile IPv6.

MN: Mobile Node.

P- DMM: Partially - Distributed Mobility Management.

PBA: Proxy BindingAcknowledgment.

PBU: Proxy Binding Update.

PCoA: Proxy Care of Address.

PGW: Packet Data Network-Gateway.

PMIPv6: ProxyMobile IPv6.

QoS: Quality of Service.

RA: Router Advertisement.

RS: Router Solicitation.

SIFDMM: Seamless IP Flow and Distributed MobilityManagement.

SIPTO: Selected IP Traffic Offload.

STA: Stations.

TS: Traffic Selector.

UE: User Equipment.

WiFi: Wireless Fidelity.

WLAN: Wireless Local Area Networks.

1 - INTRODUÇÃO

Atualmente existe um elevado número de usuários nas redes de acesso sem fio, havendo uma tendência de crescimento exponencial desse número. Os usuários utilizam diversos serviços, que incluem comunicação por voz, troca de mensagens (correio eletrônico, SMS, MMS) e "streaming" de vídeo, dentre outros. A participação em sessões de comunicação por meio da Internet se dá graças ao suporte da pilha de protocolos baseada em IP, brindado pelos sistemas celulares mais recentes. Da parte dos usuários, um crescimento da demanda tem ocorrido por novos serviços de dados, em tempo real e, em boa parte dos casos, com requisito de mobilidade dos terminais. Mais ainda, as redes de próxima geração (NGN, do inglês *Next Generation Networks*), baseadas totalmente no protocolo IP, proverão leques de serviços mais amplos e melhorados, onde sejam inclusos fatores como a convergência global, a interoperabilidade e a mobilidade.

Esse cenário traduz-se em um grande aumento do tráfego de dados nas redes móveis, o que pode ocasionar deterioração da qualidade dos serviços (QoS, do inglês *Quality of Service*) oferecidos aos usuários e da qualidade de experiência (QoE, do inglês *Quality of Experience*) vivenciada pelos mesmos, produto da sobrecarga na rede. Tal sobrecarga se caracteriza por situações de congestionamento na rede, detectadas por eventos de perdas de pacotes ou pela expiração de tempo-limite de espera de pacotes de reconhecimento (*acknowledgements*). Observe-se ainda que, para alguns tipos de tráfego (como por exemplo voz e vídeo), requisitos de tempo real tornam-se mais difíceis de serem atendidos durante situações de congestionamento, podendo levar a uma baixa qualidade dos serviços prestados, assim como a uma redução do nível de qualidade de experiência por parte dos usuários.

No tocante a redes sem fio, o meio de transmissão e a mobilidade comumente associada impõem dificuldades adicionais, expressas por exemplo pela elevada variabilidade das condições dos meios de transmissão não-confinados, pela possibilidade de desconexões frequentes e pelo aumento da taxa de perdas de pacotes. Adicionalmente, a possibilidade de perda de continuidade de sessões de comunicação, especialmente em decorrência de eventos de *handoff (*tanto intra-rede quanto inter-redes), leva à necessidade de arquiteturas, técnicas e protocolos efetivamente capazes de se contrapor a essas condições desfavoráveis de operação, ao mesmo tempo em que se busca atender a uma demanda que se mostra cada vez mais elevada, com possibilidade de congestionamento na rede. Considere-se ainda que tais

redes sem fio, para se fazerem aptas ao transporte de serviços multimídia (como voz e vídeo, por exemplo) e com mobilidade, devem ser capazes de atender a tráfegos dotados de requisitos de tempo real, com QoS e QoE em níveis adequados.

Para minimizar os problemas decorrentes de congestionamento, as operadoras de rede tentam buscar novas soluções tecnológicas que, por exemplo, aumentem a capacidade da rede. Algumas das soluções consideradas envolvem o aumento do número de estações rádio base, a melhora seletiva de algumas delas, o aumento da cobertura através de femtocélulas e-ou a melhoria das tecnologias de acesso de rádio para melhorar o emprego da banda disponível. Tais soluções envolvem, comumente, custos elevados.

Por outro lado, verifica-se que os fabricantes de terminais móveis já tem incorporado ao mercado dispositivos multimodais, com diferentes interfaces de rádio integradas, gerando um ambiente heterogêneo no que diz respeito às tecnologias de acesso. Essas redes de acesso sem fio podem conviver na mesma área geográfica, sobrepondo ou complementando suas áreas de cobertura. Então uma solução viável é fazer o escoamento dos dados (*data offloading*), para uma rede com maior largura de banda e menor congestão, tentando obter um balanceamento de carga nas redes e melhorando assim a qualidade da experiência percebida pelo usuário.

Outro aspecto importante, é permitir aos usuários se movimentar entre as diferentes tecnologias de acesso, sem experimentar interrupções nas sessões de comunicação em andamento, ao tempo em que, nas áreas de sobreposição, tomam vantagem da possibilidade de conexão simultânea de suas interfaces, aumentando por exemplo a vazão. O principal desafio para dar solução a esta problemática é que, ao se conectar em uma rede diferente, pode ocorrer perda de continuidade das sessões. Ocorre que o endereçamento IP impõe uma associação entre a localização e a identificação do host. Então é preciso tratar essa associação para terminais móveis (cuja localização é alterada sem alterar a identificação), o que tem sido feito por protocolos de gerenciamento de mobilidade padronizados pela IETF (*Internet Engineering Task Force*) [5].

Diversos protocolos para gerenciamento de mobilidade têm sido propostos, sendo os principais os protocolos MIP (*Mobile IP*) [1-2] (que é baseado no host) e PMIPv6 (*Proxy MIP*) [3] (que é baseado na rede). Ditos protocolos são centralizados, onde todo o tráfego de dados atravessa a mesma entidade de rede, a qual também é encarregada do controle de

mobilidade. Esta característica leva a problemas como: ter um ponto único de falha, escalabilidade e confiabilidade. Com o propósito de resolver esses problemas, surgiu o gerenciamento distribuído de mobilidade (DMM, do inglês *Distributed Mobility Management*) [4], podendo ser parcialmente distribuído (plano de dados distribuído e plano de controle centralizado), ou totalmente distribuído (ambos planos distribuídos).

Como apoio à solução do problema de congestão, e para facilitar o escoamento de dados, é preciso que a rede suporte eficientemente o tratamento de mobilidade IP. Do ponto de vista de escoamento de dados, os protocolos mencionados provêm opções para movimentar todos ou nenhum dos fluxos através da rede (*handover*), mas para ter melhor balanceamento de carga na rede, são necessárias opções para movimentar determinados fluxos IP de um dado usuário, o que é conhecido como mobilidade de fluxo IP (do inglês *IP Flow Mobility*) [6-7].

Assim, novos protocolos devem ser desenvolvidos, para atender à mobilidade, ao escoamento de dados e aos requisitos de tempo real impostos por aplicações de grande interesse para os usuários (como vídeo streaming, por exemplo). Em especial, “vídeo streaming” representa uma aplicação de enorme interesse para grande parte dos usuários, ao mesmo tempo em que envolve necessidades de compressão de forma a reduzir o elevado consumo de banda, além de requerer adaptabilidade dos recursos da rede e das técnicas de codificação/compressão.

Para uma eficiente prestação de serviços aos usuários móveis, as redes sem fio requerem mecanismos de gestão de mobilidade os quais são capazes de movimentar os fluxos de maneira total ou parcial. Tais mecanismos devem permitir o escoamento de fluxo, onde a localização de cada usuário e a alocação de recursos deve se dar de forma proativa. Por outro lado, para a elaboração de um protocolo de comunicação adaptável, vários planos de gestão de mobilidade podem ser considerados, sendo necessários mecanismos eficazes de *handover* e escoamento de fluxo, para garantir a conectividade ininterrupta e a prestação de serviços ininterrupta [8].

1.1 - MOTIVAÇÃO

Situações de congestionamento acontecem na Internet, havendo expectativa de que ocorram com frequência cada vez maior devido ao aumento da demanda. Isso é ocasionado em boa parte pelo aprofundado interesse dos usuários em "vídeo streaming", vista como uma aplicação que demanda muito dos recursos de rede (banda e "buffers" em roteadores, por

exemplo). Tal demanda elevada soma-se aos problemas associados ao meio sem fio e à mobilidade.

Por outro lado, observa-se, nos últimos anos, que o emprego da Internet com base em redes sem fio heterogêneas começa a ocorrer pelos usuários, face à disponibilidade de terminais multimodo e à possibilidade de integração entre essas redes. Quando combinadas com protocolos eficientes para a gestão de mobilidade e escoamento de dados, as redes heterogêneas podem ajudar a melhorar a vazão e o balanceamento de carga na rede e, com isto, a qualidade percebida pelo usuário. Considerando o protocolo IP para a gestão de mobilidade, ou ainda algumas de suas variantes (sejam as centradas no cliente, como o MIP, sejam as centradas na rede, como o PMIP), é possível movimentar seletivamente fluxos IP e assim oferecer uma elevada QoS aos usuários.

Ainda no tocante ao gerenciamento de mobilidade, outro aspecto importante é evitar os problemas existentes nas propostas de mobilidade IP centralizada, como ponto único de falha, escalabilidade e confiabilidade. Portanto, novas soluções baseadas em DMM estão sendo propostas na literatura [5],[8-19], algumas das quais tem maior granularidade [5],[8],[13-16], permitindo movimentar seletivamente fluxos IP, entre redes de acesso com diferentes tecnologias. Portanto, existe motivação para desenvolver uma arquitetura distribuída que permita o direcionamento do tráfego de dados, movimentando seletivamente fluxos IP entre redes de acesso com diferentes tecnologias, provendo suporte para a possível implementação de algoritmos dinâmicos de balanceamento de carga e de controle de congestão. Com isto é possível oferecer maior vazão e evitar congestão nas redes de acesso, permitindo ao usuário se movimentar sem experimentar interrupção em suas sessões de comunicação.

Adicionalmente, é desejável reduzir a latência e a perda de pacotes de dados, durante e após o processo de movimentação dos dados, a fim de garantir as exigências mínimas de qualidade para o usuário final [5],[8],[14]. Neste sentido é possível propor um esquema que permita movimentar fluxos seletivamente entre redes de acesso com diferente tecnologias, utilizando protocolos baseados em DMM. Com isto é possível lograr melhorias de desempenho no atraso de transferência e na taxa de perda de pacotes, aumentando a vazão e obtendo um adequado balanceamento de carga nas redes de acesso.

1.2 - OBJETIVOS

1.2.1 - Objetivo Geral

Propor uma arquitetura para escoamento de tráfego e gerenciamento de mobilidade com base em redes sem fio heterogêneas, que possua bom desempenho em comparação com outras arquiteturas já publicadas.

1.2.2 - Objetivos Específicos

- Conhecer conceitos básicos e avançados de integração de redes sem fio, integração de redes WiFi e redes LTE.
- Conhecer e aplicar conceitos básicos e avançados sobre o funcionamento e projeto de protocolos, bem como buscar propor novos protocolos e arquiteturas para a gestão de mobilidade.
- Conhecer e avaliar diferentes propostas de protocolos de gerência de mobilidade, que permitam escoamento de tráfego, para posterior seleção e comparação.
- Avaliar a qualidade do serviço oferecido por redes sem fios heterogêneas a partir da preparação para a movimentação de fluxos IP, bem como durante e depois desse processo, com a utilização de diferentes parâmetros de qualidade de serviço (QoS).
- Propor e avaliar uma arquitetura distribuída que permita direcionar o tráfego de dados com um bom desempenho.
- Avaliar a arquitetura proposta com vistas à transmissão de vídeo (*vídeo streaming*), em comparação com as demais arquiteturas, e considerando tráfego concorrente e configurações comumente utilizadas para as camadas física, de enlace, rede, transporte e aplicação.

1.3 - CONTRIBUIÇÕES

As seguintes contribuições podem ser destacadas:

- Levantamento bibliográfico e discussão sobre as soluções e padrões existentes para o gerenciamento de mobilidade;
- Levantamento bibliográfico e discussão sobre as soluções que realizam escoamento de dados através da movimentação de fluxos IP;
- Proposta de uma arquitetura para o gerenciamento de mobilidade distribuído, que permite escoamento de dados entre redes de acesso com diferente tecnologia;

- Proposta de um modelo de avaliação analítica para comparação entre as arquiteturas consideradas.
- Avaliação do desempenho da arquitetura proposta, em termos de QoS e QoE, quando considerada para vídeo streaming em conjunto com tráfego concorrente, e considerando configurações comumente utilizadas para as camadas física, de enlace, rede, transporte e aplicação.

1.4 - ORGANIZAÇÃO

O restante desta dissertação está organizado da seguinte forma: o Capítulo 2 apresenta conceitos básicos em redes de acesso sem fio, bem como considerações importantes dos processos de *handover* e escoamento de dados.

No Capítulo 3 é apresentado o estado da arte das soluções e padrões existentes para o gerenciamento de mobilidade, sendo feita uma compreensiva categorização e comparação das mesmas.

No Capítulo 4 é descrita a arquitetura proposta, sendo apresentado seu funcionamento geral e as considerações para escoamento de dados, assim como as estruturas de dados utilizadas, juntamente com o formato e fluxo das mensagens de controle.

O Capítulo 5 apresenta o modelo analítico proposto, para a avaliação do desempenho da arquitetura, bem como são apresentados e discutidos os resultados numéricos. Também são descritos os cenários de simulação no NS-3 e são apresentados e discutidos os resultados obtidos, relativos a métricas de desempenho de rede (QoS) e a qualidade de vídeo (QoV, do inglês *Quality of Video*), utilizando a ferramenta Evalvid.

Finalmente, no Capítulo 6 são apresentadas as conclusões do trabalho desenvolvido e as propostas de trabalhos futuros para a continuidade desta dissertação.

2 - FUNDAMENTOS TEÓRICOS

Neste capítulo apresentamos alguns conceitos básicos importantes para o entendimento deste trabalho, tais como redes de acesso sem fio LTE e WiFi, suas principais características e arquitetura. Além disso, apresentamos conceitos sobre gerenciamento de mobilidade, abordando o processo de *handover* e, ao final, abordamos conceitos referentes ao escoamento de dados (*data offloading* do inglês), apresentando os principais padrões e técnicas existentes.

2.1 - REDES DE ACESSO SEM FIO HETEROGÊNEAS

Atualmente existe um elevado número de usuários nas redes de acesso móveis, o que evidencia a grande utilidade dos dispositivos sem fio, tais como celulares ou *tablets*, na vida das pessoas. Os usuários não só utilizam os serviços de voz e de troca de mensagens mas também são capazes de participar em sessões de Internet graças ao suporte ao protocolo IP brindado pelos sistemas celulares atuais. Mais ainda, as redes de próxima geração NGN, totalmente baseadas no IP, proverão serviços mais variados e melhorados, onde sejam inclusos fatores como a convergência global, a interoperabilidade e a mobilidade.

Tudo isto traz um grande aumento do tráfego de dados nas redes móveis, o que pode ocasionar deterioração da qualidade dos serviços oferecidos aos usuários, produto da sobrecarga na rede. Para tratar este inconveniente, as operadoras de rede tentam buscar novas soluções tecnológicas para resolver a necessidade do aumento da capacidade da rede. Algumas das soluções consideradas são o aumento do número de estações base ou melhorar algumas delas seletivamente. Outra proposta é a melhora da cobertura através de femtocélulas ou melhorar as tecnologias de acesso de rádio para incrementar a largura de banda [1]. Mas todas estas soluções incorrem em grandes inversões, convertendo-se em soluções custosas.

Logo é preciso aproveitar o fato de que, os fabricantes de terminais móveis já tem incorporado ao mercado dispositivos multimodais com diferentes interfaces de rádio integradas, gerando um ambiente heterogéneo no que diz respeito às tecnologias de acesso. Essas redes de acesso sem fio, convivem na mesma área geográfica, sobrepondo suas áreas de cobertura. Então uma solução viável é fazer o escoamento dos dados móveis para uma rede com maior largura de banda e menor congestão, tentando obter um balanceamento de

carga nas redes e buscando aumentar assim a qualidade de experiência percebida pelo usuário.

As redes WiFi (*Wireless Fidelity* – 802.11) parecem ser a melhor solução, devido ao fato que existem muitos pontos de acesso (AP – *Access Point*) disponíveis em casas, universidades, lojas. Ademais os operadores podem instalar seus próprios AP, onde a demanda seja alta e eles operem em uma banda diferente das estações base, não gerando interferências.

2.1.1 - IEEE 802.11: WiFi (*Wireless Fidelity*)

Nos últimos anos, a maioria das comunicações sem fio de área local é baseada no padrão IEEE 802.11. O padrão é dominante no mercado internacional e um grande número de terminais (*smartphones, laptops, tablets* ...) estão equipados com a interface WiFi *(Wireless Fidelity*, nome pelo qual é conhecido o padrão), que define as operações da rede local, em uma área especifica. O padrão 802.11 [2] foi publicado em 1997 pelo *Institute of Electrical and Electronics Engineers* (IEEE) e proporciona conectividade sem fio, para estações fixas, portáteis e móveis dentro de uma área local. O mesmo opera na faixa de frequência ISM (do inglês *Industrial, Scientific and Medical*) nas bandas de 2.4, 3.6, 5 e 60 GHz.

O padrão define o uso das camadas física e de enlace do modelo OSI/ISO e inclui várias revisões, que diferem principalmente na especificação da camada física. Entre as principais revisões temos:

IEEE 802.11a: foi aprovada em 1999, trabalha na banda de 5 GHz e utiliza 52 subportadoras OFDM (*Orthogonal Frequency-Division Multiplexing*) com uma velocidade teórica máxima de 54 Mbits/s. Tem 12 canais sem sobreposição, 8 para rede sem fio e 4 para conexões ponto a ponto.

IEEE 802.11b: foi aprovada em 1999, opera na banda de 2.4 GHz. Foi a primeira revisão que teve uma ampla implementação no mercado, tomando o nome de WiFi. 802.11b tem uma velocidade máxima de transmissão de 11 Mbit/s, mas devido ao protocolo CSMA/CA (*Carrier Sense Multiple Access With Collision Avoidance*), na pratica, a velocidade máxima de transmissão é de aproximadamente 5.9 Mbit/s sobre TCP e 7.1 Mbit/s sobre UDP (*User Datagram Protocol*).

IEEE 802.11e: visa dar possibilidades em termos de qualidade de serviço na camada de enlace. Assim, esta norma tem como objetivo definir as necessidades dos diferentes pacotes em termos de banda concorrida e prazo de transmissão, de maneira a permitir uma melhor transmissão da voz e de vídeo. Em suma, 802.11e permite a transmissão de diferentes classes de tráfego, além de trazer o recurso de *Transmission Oportunity* (TXOP), que permite a transmissão em rajadas, otimizando a utilização da rede.

IEEE 802.11g: Foi aprovada em 2003, sendo a evolução de 802.11b. A revisão utiliza a banda 2.4 GHz e opera a uma velocidade máxima teórica de 54Mbit/s, que na pratica é de aproximadamente 22 Mbit/s. É compatível com o padrão b e utiliza as mesmas frequências.

IEEE 802.11n: foi aprovada em 2009 com uma velocidade máxima teórica de 600 Mbit/s na capa física, utilizando antenas MIMO (*Multiple-Input Multiple-Output*). Na atualidade existem vários produtos que cumprem com a revisão com um máximo de 300 Mbit/s.

2.1.1.1 - IEEE 802.11 Arquitetura

A arquitetura de IEEE 802.11 consiste de um grupo de componentes que interatuam para prover comunicação sem fio, formando uma rede sem fio de área local (*Wireless Local Area Networks* -WLAN), que suporta mobilidade das estações (STA - *Stations*), de um jeito transparente para as capas superiores [3]. A componente essencial da arquitetura de 802.11 é o BSS (*Basic Service Set*). O BSS é definido como um conjunto de estações que podem estabelecer comunicação entre elas, diretamente ou utilizando um AP (*Access Point*), em uma área geográfica especifica. Essa área de cobertura é conhecida como BSA (*Basic Service Area*) [2-4].

O padrão suporta dois modos de operação:

Ad-Hoc: onde as estações se comunicam umas com as outras diretamente no âmbito de um BSS, baseadas na comunicação ponto-a-ponto (*peer-to-peer*), como mostra a Figura 2.1. Este conjunto de serviços é chamado de IBSS (*Independent Basic Service Set*).

Figura 2.1: Rede *Ad-Hoc* IEEE 802.11.

Infraestrutura: a rede utiliza APs, como unidade centralizada, para estabelecer a comunicação com a rede fixa e entre os terminais sem fio, como mostra a Figura 2.2. O AP é o ponto obrigatório de comunicação, por onde passa todo o tráfego da rede. Em uma rede infra estruturada, as estações devem efetuar a função de associação ao AP de modo a obter os serviços de rede. Um terminal inicia-se na rede sempre com esta função mas o AP decide se permite ou não o seu registo. A função de associação é exclusiva do terminal que só pode estar associado a um AP.

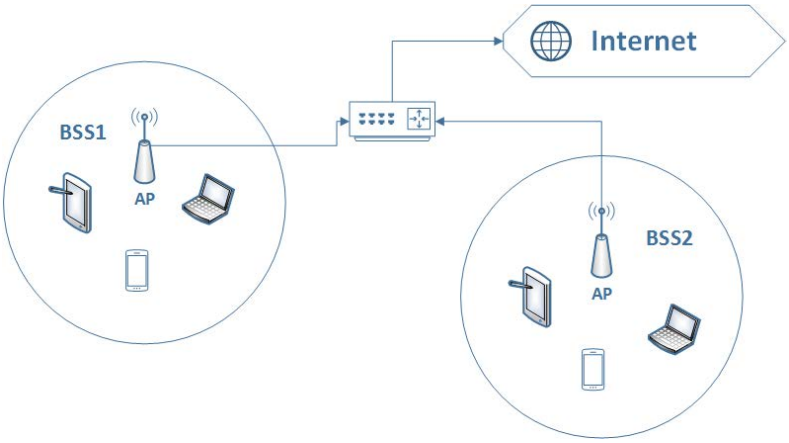

Figura 2.2: IEEE 802.11 Infraestrutura.

2.1.2 - LTE (*Long Term Evolution*)

LTE é uma tecnologia de redes para as comunicações móveis em banda larga proposta pelo 3GPP (*3rd Generation Partnership Project*), e introduzida no Release 8. Esta tecnologia permite atingir valores de eficiência espectral e vazão maiores do que os conseguidos pelo protocolo *High Speed Packet Access* (HSPA), com a finalidade de suportar a demanda de novos serviços móveis com altos requisitos de taxa de bit, retardos e capacidade do sistema [5, 6].

LTE é projetado para aumentar as taxas de dados na borda da célula, melhorar a eficiência do espectro (*unicast*, bem como *broadcast*) e permitir flexibilidade do espectro (1,25, 2,5, 5, 10, 15 e 20 MHz) para o planejamento de rádio flexível [6], apresentando na interface de rádio, as seguintes características chaves:

- O uso de OFDM/OFDMA como técnica de acesso, permite a utilização de larguras de banda flexíveis entre 1,4 e 20 MHz (até 100 MHz em LTE-A), além de combater eficientemente os efeitos produzidos pela propagação multipercurso, entre eles a interferência inter-simbólica.
- O uso de modulações e codificações adaptativas, de acordo ao estado de canal experimentado pelo usuário (em termos da razão sinal-interferência), permite melhorar a vazão do sistema ao utilizar modulações/codificações com diferentes eficiências espectrais.
- A arquitetura de rede totalmente baseada no Protocolo de Internet (IP- *Internet Protocol*) e o uso de técnicas de agendamento de pacotes (*packet scheduling*) melhoram o aproveitamento dos recursos de rádio ao permitir alocá-los conforme as necessidades dos serviços e estados do canal de cada usuário.
- O uso de técnicas multi-antena (MIMO), consiste no uso de estruturas de transmissão/recepção que utilizam diversas antenas acompanhadas de técnicas de processamento de sinais. Estas técnicas permitem melhorar o desempenho do sistema aumentando a capacidade das células, melhorando a cobertura e a velocidade de transmissão por usuário.

2.1.2.1 - Arquitetura do LTE

Nas especificações do 3GPP é chamada a arquitetura de um sistema LTE como EPS (*Evolved Packet System*), tendo o sistema três elementos fundamentais: o terminal dos usuários UE (*User Equipment*), a rede de acesso E-UTRAN (*Evolved Universal Terrestrial Access Network*) e o coração da rede chamado EPC (*Evolved Packet Core Network*), como o da Figura 2.3.

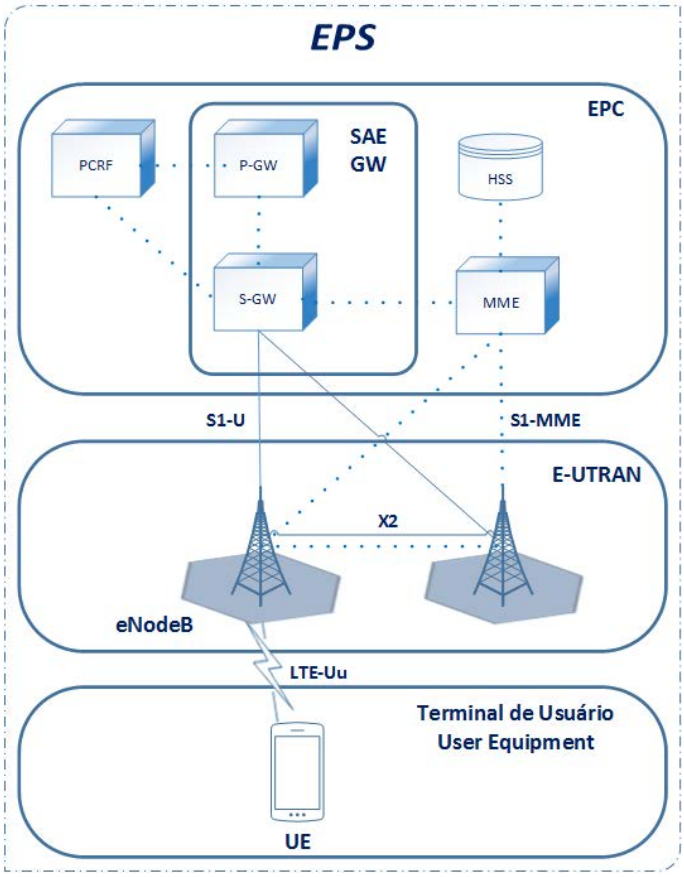

Figura 2.3: Arquitetura de um Sistema LTE.

A rede de acesso LTE ou E-UTRAN só tem um componente, que é a estação base, chamada *evolved NodeB* (eNB), que gera uma arquitetura plana como a da Figura 2.4. Esta estação base integra todas as funcionalidades da rede de acesso.

A rede de acesso E-UTRAN, ao estar formada só pelas eNB, as mesmas serão as encarregadas de proporcionar conectividade entre os usuários e o coração da rede. As eNBs têm três interfaces: para comunicar-se com os usuários, com o EPC e com outra eNB, como mostra a Figura 2.4.

E-UTRAN Uu é a interface de rádio que comunica o EU com a eNB, utilizando o canal de rádio. Todas as funções e protocolos necessários para o envio de dados e as operações de controle são implementadas na eNB.

Ao EPC se comunica utilizando a interface S1, que se divide em outras dois: S1-MME que utiliza-se para o plano de controle e a S1-U para o plano de usuário.

A interface X2 é utilizada para conectar as eNB. Através desta interface as eNB intercambiam tanto mensagens de sinalização, quanto tráfego de usuários do sistema durante um processo de *handover*.

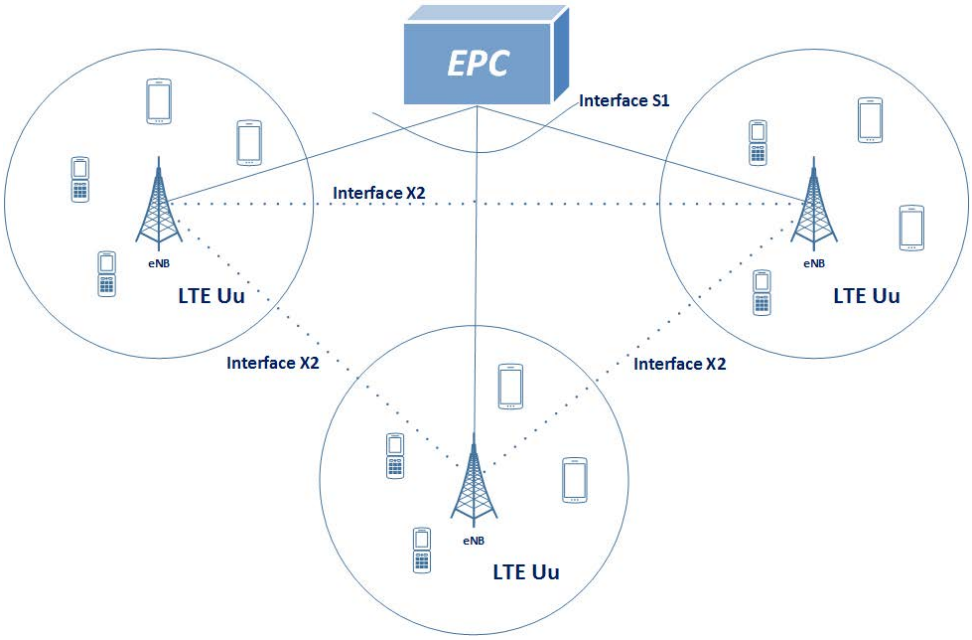

Figura 2.4: Rede de acesso LTE.

2.2 - PROCESSO DE *HANDOVER*

As redes sem fio oferecem a possibilidade aos usuários de se movimentar, mantendo a continuidade das sessões abertas, o que traz consigo a necessidade de *handover* rápido e transparente. O *handover* é a ação de um terminal móvel (MT – *Mobile Terminal*) trocar seu ponto de acesso à rede, também chamado *handoff*. Os MT, na rede, podem cambiar de um ponto de acesso a outro do mesmo tipo, como UMTS - UMTS ou WiFi – WiFi. Esse processo é conhecido como *Horizontal Handover* (HHO). Mas atualmente é comum existir sobreposição da cobertura de distintas tecnologias de rede de acesso, o que leva ao *Vertical Handover* (VHO). Este é um processo de maior complexidade, onde um MT troca de pontos de acesso com diferentes tecnologias, como LTE – WiFi [7, 8]. A Figura 2.5 ilustra os processos de *handover*.

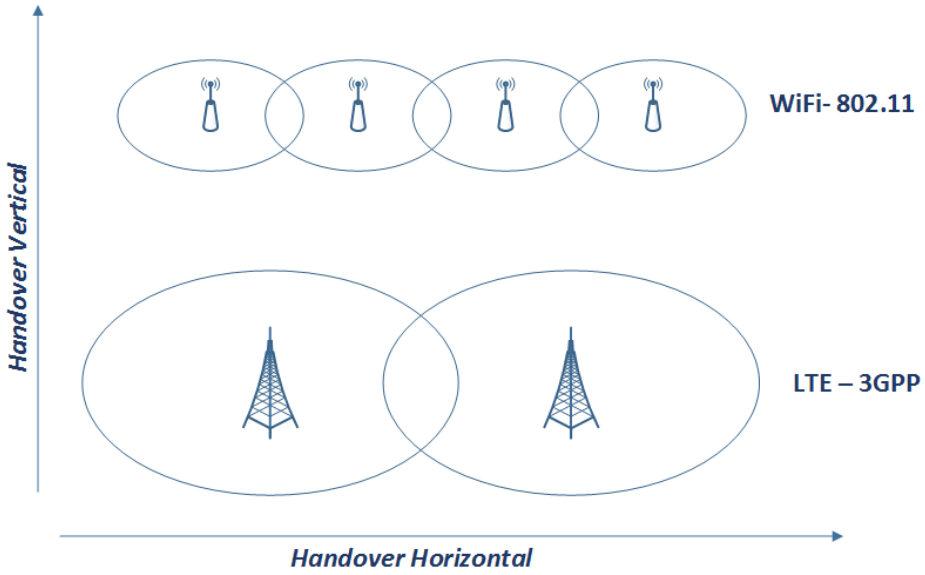

Figura 2.5: *Handover* Horizontal e Vertical.

2.2.1 - *Handover* Vertical (VHO)

Motivados pelos desafios que traz a complexidade das redes heterogêneas, e as novas tendências de oferecer a melhor QoS (*Quality of Service*), vamos centrar nossa pesquisa no VHO. Hoje existem diferentes tecnologias de acesso que convivem na mesma área geográfica, sobrepondo sua cobertura. Isto possibilita aos operadores aproveitar as características de cada uma das redes, com o fim, de se adaptar as preferências do usuário e aplicações. Um dos principais problemas é fazer um *handover* transparente e com mínimos retardos e perdas de pacotes para poder suportar, por exemplo, aplicações sensíveis ao tempo com a QoS requerida. Para dar solução a essa problemática é preciso otimizar o processo do *handoff*.

A gestão do processo de *handover* consiste de três etapas:

1. Descoberta do sistema: durante a descoberta do sistema é coletada a informação requerida para identificar a necessidade de um *handover* e saber quais são as redes de acesso disponíveis. Tanto o MT como a rede podem participar.
2. Seleção de rede: a informação coletada na primeira etapa é avaliada com o objetivo de determinar as disponibilidades das redes de acesso disponíveis, fazendo possível a escolha da que melhor satisfaça as preferências do usuário ou aplicações.
3. Execução do *handover*: finalmente uma nova conexão é estabelecida e os recursos da antiga são liberados.

Nas pesquisas muitos frameworks têm sido desenvolvidos, para assistir ao terminal móvel na descoberta e seleção da rede de acesso, com o objetivo de executar VHO. Entre os padronizados temos:

- O padrão IEEE 802.21 ou MIH (*Media Independent Handovers*), desenvolvido pela IEEE em 2009 [9]. O padrão propõe três serviços principais que podem ser usados para assistir ao nó móvel no processo do *handover* entre redes sem fio heterogêneas. Os serviços são: *Media Independent Information Service* (MIIS), *Media Independent Event Services* (MIES) e *Media Independent Command Services* (MICS). Uma versão de código aberto do padrão 802.21 chamada *Open Dot Twenty One* (ODTONE) [10], foi desenvolvida pelo *Heterogeneous Networking Group* (HNG). ODTONE prove um framework para a implementação do MIH em múltiplas plataformas, considerando tanto hardware como software.
- *Access Network Discovery and Selection Function* (ANDSF) [11], desenvolvido pelo 3GPP, pretende assistir ao UE para descobrir as redes de acesso, não 3GPP, disponíveis que podem ser utilizadas para comunicação de dados. Além de prover regras ao UE que permitam a conexão com essas redes.

2.3 - ESCOAMENTO DE DADOS (*DATA OFFLOADING*)

Nos últimos anos o desenvolvimento de aplicações e os novos dispositivos móveis têm trazido um aumento no tráfego de dados nas redes celulares. Para resolver esta problemática o escoamento é apresentado como uma solução viável. Nos últimos Release do 3GPP, tem despertado grande interesse no escoamento do tráfego de redes celulares, com o fim de controlar a sobrecarga das redes e oferecer melhor QoE. Uma das principais soluções propostas, é movimentar fluxos IP para femtocélulas ou redes WiFi. Em ambos casos, as macro células são utilizadas para oferecer uma cobertura continua e as femtocélulas ou redes WiFi para balanceamento de carga em áreas congestionadas. Isto é consequência de as últimas poderem manejar grandes volumes de dados, mas sua área de cobertura ser pequena [20].

Logo têm sido propostos três mecanismos: *Local IP Access* (LIPA), *Selected IP Traffic Offload* (SIPTO) e *IP Flow Mobility* (IFOM), para realizar o escoamento eficiente e evitar assim a deterioração na QoS oferecido ao usuário final. Os mecanismos LIPA e SIPTO [21] são baseados em descarregar o trafego IP antes que alcance o núcleo da rede, e requerem

uma HeNB ou femtocélula , e o mecanismo IFOM [6-7] assume o uso de duas interfaces de rádio no UE, uma para a rede celular e outra para a redes WLAN (i.e. WiFi), capazes de trabalhar simultaneamente.

2.3.1 - LIPA - *Local IP Access*

LIPA [21] é um método que oferece a possibilidade ao UE, ligado a uma HeNB (*Home node B or Home enodeB*), de transferir dados à rede local conectada ao mesmo HeNB, sem que os dados atravessem a rede macro celular. LIPA também permite ao UE acessar qualquer rede externa que estiver ligada à rede local [22]. LIPA só é aplicável a redes privadas, podendo ser utilizado em redes sem fio residenciais e coorporativas, para fazer o roteamento local ou acessar à Internet através de um gateway privado.

O 3GPP define o *breakout* como a localização na arquitetura 3GPP onde o escoamento dos dados deve ocorrer. No LIPA o ponto de *breakout* deve estar dentro da rede privada. Nesta arquitetura o núcleo da rede é evitado e o trafego do usuário é roteado através do *Local Gateway* (L-GW), localizado na rede privada, como mostra a Figura 2.6.

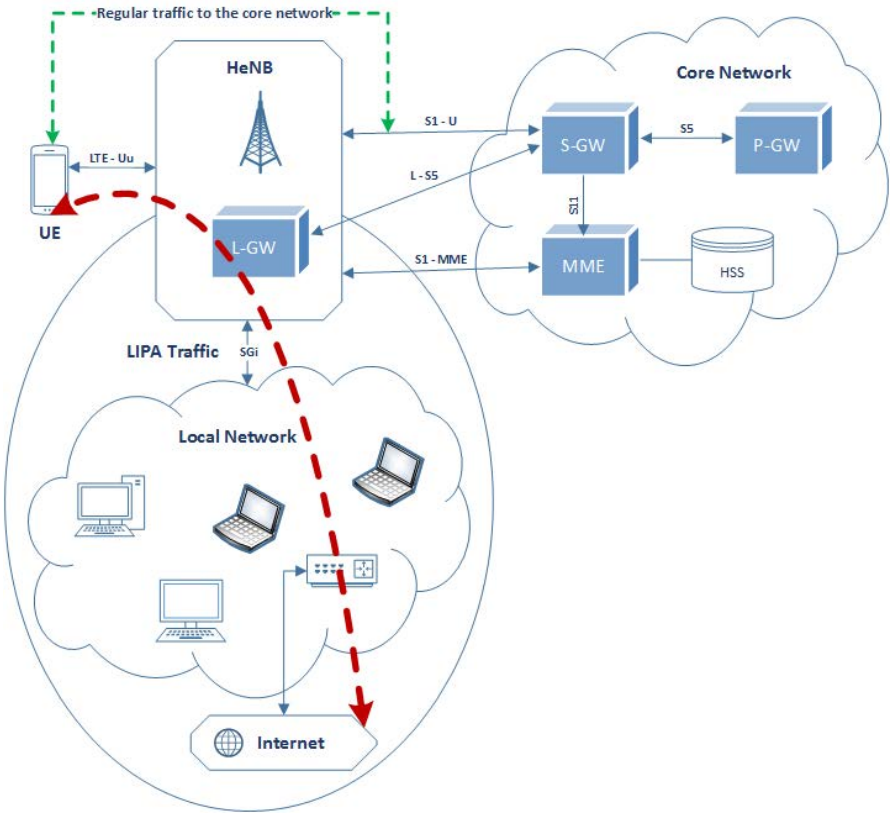

Figura 2.6: Arquitetura LIPA.

Na especificação 3GPP [21], referem-se ao HeNB e o L-GW como o subsistema HeNB. A seguir é apresentada uma breve descrição dos principais elementos da topologia LIPA.

O L-GW é essencialmente um PGW (*Packet Data Network-Gateway*) que implementa funções como alocação de endereço IP para o UE, filtragem de pacotes, túnel direto com o HeNB, entre outras. Está localizado logicamente no HeNB e se conecta aos dispositivos da rede privada diretamente através da interface SGi. Além de trocar informação com o S-GW da rede macro através da interface L-S5. O núcleo da rede celular, para estabelecer uma conexão LIPA PDN, encaminha mensagens de sinalização que permitem fazer o registro e o gerenciamento das seções, logo o L-GW é o encarregado de rotear os pacotes de dados [22]. Por outro lado, o UE pode também manter suas seções de dados com o núcleo da rede celular macro, enquanto a conexão LIPA é estabelecida.

O HSS (*Home Subscriber Server*) contém informação da autorização ou não de acesso local, para cada APN (*Access Point Name*) e cada subscritor. Se o acesso local é permitido, o MME faz a seleção do endereço habilitando um plano de usuário que representa um caminho direto entre o L-GW e o HeNB.

O primeiro conjunto de funções LIPA são definidos pelo 3GPP no *Release 10*, mas o suporte para a mobilidade é oferecido no *Release 11*.

2.3.2 - SIPTO - *Selected IP Traffic Offload*

No método SIPTO [21] porções do tráfego IP, em uma HeNB ou eNB, são escoados a uma rede local com o fim de reduzir a carga do sistema. A Figura 2.8 mostra soluções para SIPTO utilizando a rede celular ou uma HeNB, onde o ponto de *breakout*, é localizado acima da RAN (*Radio Access Network*). Um conjunto de gateways (S-GW e P-GW) e MME, localizados perto do ponto de conexão do UE à rede, são selecionados para fazer o escoamento dos dados. Com esta solução são diminuídos a quantidade de saltos até o destino, selecionando caminhos de menor custo.

Entre os elementos da topologia estão o HSS, MME e os gateways. O HSS tem funções similares no LIPA, mantendo informação de permissão por cada subscritor e APN. O MME em SIPTO deve realizar as seguintes funções: selecionar um conjunto de S-GW e P-GW que estão topologicamente perto do UE, autorizar ou não o SIPTO baseado nos dados de HSS, e

tomar decisões de nova localização de um *gateway* devido à mobilidade do UE. O S-GW e P-GW tem as mesmas funções [23].

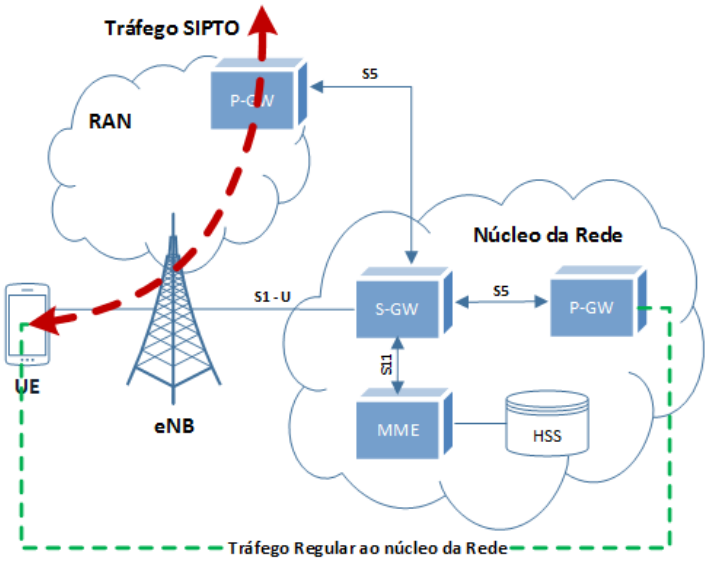

Figura 2.7: Arquitetura SIPTO em uma macro célula.

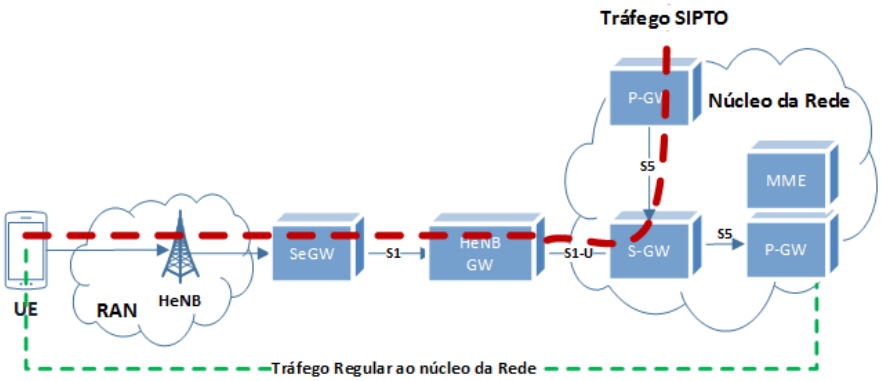

Figura 2.8: Arquitetura SIPTO em uma femtocélula.

2.3.3 - IFOM – *IP Flow Mobility*

IFOM [7] é uma extensão especial de mobilidade IP, padronizada pela 3GPP. O padrão, permite movimentar fluxos seletivos de comunicações em andamento, de uma rede de acesso para outra, sem interrupção dos fluxos modificados, enquanto mantém outros fluxos na rede de acesso atual [7],[24]. Baseados nas especificações [6-7], os UEs poderão estabelecer conexão simultaneamente com um acesso 3GPP e um WiFi, e podem trocar diferentes fluxos IP da mesma conexão PDN entre as diferentes redes de acesso, como mostra a Figura 2.9. As soluções são baseadas nos protocolos DSMIPv6, PMIPv6 e GTP, permitindo a

preservação do endereço IP e a continuidade das seções, durante a movimentação dos fluxos IP entre os sistemas de acesso.

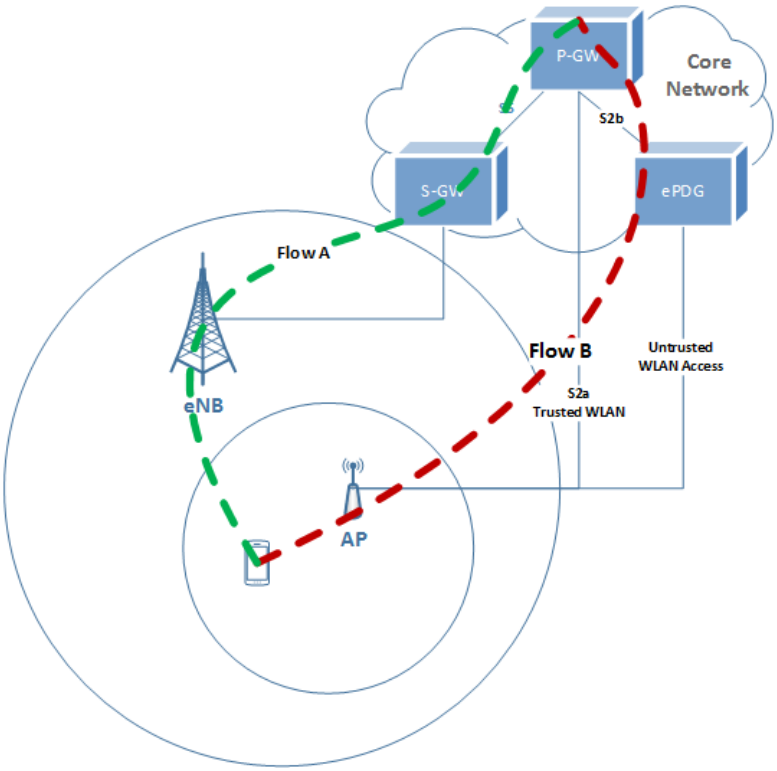

Figura 2.9: IFOM em uma arquitetura I-WLAN.

As soluções baseadas em femtocélulas, apresenta alguns problemas como: (i) SIPTO permite tráfego sobre a RAN, logo soluções utilizando “SIPTO para macro células” não reduzem a congestão no lado da célula ou interface de rádio, (ii) usam bandas licenciadas, que poderiam trabalhar na mesma frequência utilizada na macro infraestrutura existente, o que pode gerar interferência. Enquanto as redes celulares e WiFi operam em bandas de frequência diferentes e não geram interferência. Também as redes WiFi têm uma grande largura de banda, permitindo altas taxas de transferências. A diferença das soluções baseadas em femtocélulas, o operador não tem que implantar todos os pontos de acesso, pode tirar partido, das redes públicas WiFi já estabelecidas [20]. Logo soluções utilizando redes de acesso WiFi parecem ter maior aceitação que as baseadas em femtocélulas.

2.4 - CONSIDERAÇÕES FINAIS

Este capítulo apresentou um breve resumo dos principais conceitos que serão tratados no trabalho. Foram apresentadas as principais características e arquitetura das redes de acesso LTE e WiFi, com o fim de compreender as possibilidades de integração das mesmas,

permitindo um melhor aproveitamento do ambiente sem fio heterogêneo existente na atualidade. Além disso foi apresentado o processo de *handover*, dando ênfase no *handoff* vertical, pois neste trabalho serão abordadas arquiteturas com redes de acesso diferente, gerando um ambiente heterogêneo no que respeita a tecnologias de acesso. Também foram discutidas as etapas do processo de *handover*, e os padrões existentes, para assistir ao terminal móvel na descoberta e seleção da rede de acesso. Não aprofunda-se nas etapas de seleção da rede de acesso, por estar fora do escopo do trabalho, onde é utilizado um mecanismo simples e estático.

Os últimos Release do 3GPP têm despertado grande interesse no escoamento do tráfego de redes celulares, com o fim de controlar a sobrecarga das redes e oferecer melhor QoE. Uma das principais soluções propostas, é movimentar fluxos IP para femtocélulas ou para redes WiFi. Neste capítulo foram estudados os três mecanismos propostos pelo 3GPP, com o fim de ter um melhor entendimento de cada um deles. Também foram discutidos os problemas das soluções baseadas em femtocélulas, concluindo que soluções baseadas em redes de acesso WiFi, podem despertar maior interesse. Então vários pesquisadores têm realizado propostas que permitem mobilidade seletiva de fluxo IP, de um jeito similar a IFOM, permitindo fazer escoamento de dados, mecanismo adotado neste trabalho. No próximo capítulo é apresentado o estado da arte de ditas soluções, classificando-as com base ao gerenciamento de mobilidade que elas utilizam. Isso é devido à influência que tem o gerenciamento de mobilidade IP, por exemplo na continuidade da sessão, latência, perda de pacotes e escalabilidade.

3 - TRABALHOS RELACIONADOS

Na literatura podemos encontrar várias estratégias para o gerenciamento de mobilidade em redes celulares. Neste capítulo, são analisadas as principais estratégias e é provida uma compreensiva categorização das soluções existentes, envolvendo os seguintes aspectos:

- Entidades envolvidas;
- Mensagens envolvidas no processo de gerenciamento de mobilidade;
- Tecnologias de acesso;
- Gerenciamento de mobilidade baseado no host ou na rede;
- Gerenciamento de mobilidade centralizado ou distribuído;
- Possiblidade de movimentação seletiva de fluxos IP (*data offloading*);
- Entidades responsáveis pela decisão de escoamento de dados.

3.1 - CATEGORIAS RELATIVAS À MOBILIDADE

É importante precisar, que um aspecto chave no suporte de mobilidade são as entidades envolvidas no controle do processo de mobilidade. O nó móvel (MN) pode ter uma participação ativa no gerenciamento, mas uma alternativa para não sobrecarregar o dispositivo do usuário, é utilizar uma entidade da rede para realizar as operações relativas à mobilidade no seu nome. Com isto temos uma primeira categorização.

Mobilidade baseadas no host: neste caso, o MN tem uma participação ativa no gerenciamento de mobilidade. Ele é o encarregado de atualizar sua localização na rede caseira e com seu par comunicante se for o caso. Isto tem um impacto negativo na sua complexidade, desempenho e consumo de energia.

Mobilidade baseada na rede: neste caso, o MN não tem participação na sinalização relacionada à mobilidade. A rede é responsável pelo gerenciamento de mobilidade IP em seu nome. As entidades na rede são as encarregadas de fazer, o seguimento dos movimentos do MN e iniciar a sinalização requerida, com o fim de atualizar sua localização e manter a continuidade das sessões em andamento [3].

Outro aspecto a considerar é o incremento do volume de dados nas redes celulares. Isto tem levado à pesquisa de soluções, que suportem escoamento de dados, com o fim de fazer balanceamento de carga e aumentar a vazão na rede de acesso. Uma das principais soluções

propostas, é movimentar fluxos IP para redes WiFi. Então em nossa classificação será considerado o suporte ou não de mobilidade de fluxos IP, para redes de acesso WiFi:

Com suporte para mobilidade de fluxo IP: neste caso a infraestrutura de rede, tem suporte para movimentar fluxos IP de uma rede de acesso para outra, sem interrupção nas sessões em andamento. Quando o MN, este em uma área de sobreposição de redes de acesso com tecnologias diferentes, poderá receber fluxos IP, pelas duas interfaces ao mesmo tempo. Com isto tem uma granularidade fina, permitindo movimentar seletivos ou todos os fluxos IP.

Sem suporte para mobilidade de fluxo IP: neste caso o MN só poderá estabelecer conexão com uma rede de acesso. O MN poderá fazer *handover* entre redes de acesso com a mesma ou diferente tecnologia, mas não receberá fluxos IP pelas duas simultaneamente. É dizer, a infraestrutura de rede só permitirá movimentar todos os fluxos.

O mapeamento da localização do MN, pode ser realizado por só uma entidade da rede, ou ser distribuído entre vários *anchors*. Com isto temos nossa última categorização, onde consideramos se o gerenciamento de mobilidade é centralizado ou distribuído.

CMM (Centralized Mobility Management): neste caso, um dado *mobility anchor* mantem os *bindings*, de todos os MNs na rede. O tráfego de dados é então encapsulado entre o *mobility anchor* e o MN ou seu *router* de acesso. Esta abordagem é usualmente implementada em arquiteturas centralizadas onde, tanto a localização do MN, quanto o encapsulamento do tráfego precisam ser processados pela entidade central da rede (i.e. *mobility anchor*).

DMM (Distributed Mobility Management): neste caso as funções na rede, encarregadas do controle de mobilidade, são separadas do plano de dados dos nós, que é responsável principalmente pelo reenvio dos pacotes de dados. Esta abordagem, em soluções baseadas na rede, de acordo com os níveis de distribuição do plano de controle, pode também ser categorizada como [4],[12]:

P- DMM (Partially - DMM): são esquemas onde unicamente o plano de dados é distribuído entre *routers* de acesso, em quanto o plano de controle é centralizado em um nó da rede, usado para armazenar informação, mas não realiza funções de reenvio de pacotes de dados.

F-DMM (Fully- DMM): são esquemas onde tanto o plano de dados, quanto o plano de controle são distribuídos entre os *routers* de acesso.

Finalmente na Tabela 3.1, se estabelece uma relação entre as estratégias utilizadas no gerenciamento de mobilidade e as soluções existentes na literatura, que as utilizam.

Tabela 3.1: Estratégias utilizadas no gerenciamento de mobilidade.

Categoria	**Suporte para mobilidade de fluxo IP**	
	Sem suporte	**Com suporte**
Mobilidade baseada no host	CMM: [1-2],[25-28] DMM: [10],[29-35]	CMM: [6-7] DMM: [9]
Mobilidade baseada na rede	CMM: [3],[27],[36-38] P –DMM: [10],[12],[17-19],[39-47] F – DMM: [10-12],[35],[42],[45],[48-50]	CMM: [6],[51-56] P – DMM: [5],[8],[13-16] F – DMM: capítulo 4

3.2 - SEM SUPORTE PARA MOBILIDADE DE FLUXO IP

A maioria das soluções existentes para gerenciamento de mobilidade, não foram concebidas para suportar mobilidade seletiva de fluxo na camada de rede. As arquiteturas não suportam conexões simultâneas dos MN através de mais de uma interface. Elas só podem movimentar todos os fluxos IP, quando um processo de *handover* ocorre [24].

Nesta Seção apresentamos o estado da arte das soluções existentes, que brindam suporte para mobilidade, mas não permitem fazer escoamento de tráfego. Ditas propostas podem ser baseadas no usuário ou na rede, como será analisado nas próximas seções.

3.2.1 - Mobilidade baseada no host

Nas primeiras tentativas de solucionar o problema em questão, foi proposto pela IETF o protocolo MIP [1-2]. Neste protocolo o MN tem uma participação ativa no gerenciamento de mobilidade. Isto traz grandes latência de *handover* e perda de pacotes, produto aos problemas do enlace sem fio. Logo foram feitas propostas para tentar solucionar esses problemas que serão descritas na próxima sub-seção.

Com o incremento do usuários nas redes sem fio, os operadores começaram a ter problemas de congestionamento nas suas redes. Logo, surgiram novas soluções para tentar aumentar a quantidade de usuários gerenciados pela rede, mantendo uma adequada QoS (*Quality of*

Service). Uma das principais estratégias foi o gerenciamento de mobilidade distribuído [4],[57-58] logo, nós apresentamos um estado da arte de soluções deste tipo baseadas no host, na sub-seção 3.2.1.2.

3.2.1.1 - CMM

No gerenciamento de mobilidade centralizado, existe uma entidade da rede que é encarregada das funções de localização, *handover* e reenvio dos dados. A IETF em [1-2] padronizou o gerenciamento de mobilidade baseado no host para IPv4 e IPv6, sendo os protocolos chamados, MIPv4 e MIPv6 respetivamente. Neles um endereço HoA (*Home Address*) é atribuído ao MN na sua rede caseira, pela entidade centralizada HA (*Home Agent*). Esse endereço é mantido durante toda a sessão de comunicação, sendo o MN alcançável através dela. Cada vez que o MN troca seu ponto de conexão, assinasse-lhe um endereço IP dinâmico CoA (*Care of Address*), que pode ser dado localmente pelo FA (*Foreing Agent*). MIP estabelece no primeiro lugar a descoberta do agente pelo MN. Logo existe um processo de registro e finalmente se estabelecem os túneis e *bindings* entre o HA e o FA para redirecionar os pacotes encaminhados ao MN. O MN é o encarregado de atualizar sua localização com o HA, ou seu par corresponde, utilizando para isto as mensagens BU (*Binbing Update*) e BA (*Binding Acknowledgments*). Este protocolo presenta elevados retardos de *handover*, mensagens de sinalização e perda de pacotes, pelo que foram desenvolvidas outras soluções baseadas nele, que tentam dar solução a ditas desvantagens.

Por outro lado MIPv6, não suporta trabalhar com endereços IPv4. Isto levou a uma extensão do padrão, apresentada em [26], chamada *Mobile IPv6 Support for Dual Stack Hosts and Routers* (DSMIPv6). Dita solução permite o registro de endereços e prefixos IPv4, assim como o transporte de pacotes IPv4 e IPv6 através do túnel até o HA. O padrão também permite ao MN se movimentar entre redes de acesso com qualquer um dos endereços, incluindo o caso onde o NAT (*Network Address Translation*) está presente no caminho entre o MN e seu HA.

HMIPv6 (*Hierarchical Mobile IPv6*), padrão descrito na RFC5380 [28], é uma melhora do protocolo MIPv6, com o fim de reduzir a quantidade de mensagens de sinalização, e melhorar assim a latência no processo de *handover*. No protocolo MIPv6 as atualizações da localização, com o HA e o CN (*Correspondent Node*), ocorrem de igual forma no movimento do MN em um domínio local ou global. Para resolver dita ineficiência, HMIPv6 introduz

uma nova entidade, chamada MAP (*Mobile Anchor Point*). Agora se o MN está em um domínio local, só precisa atualizar à entidade MAP local, diminuindo com isto a latência de *handoff*. Dita entidade atua como um HA local, recebendo todos os pacotes em representação do MN, encapsulando-os e reenviando-os para o endereço atual do MN. Com isto o MN, só terá que atualizar sua localização, com o HA e o CN quando se movimenta fora do domínio MAP.

FMIPv6 (*Fast handover for MIPv6*), apresentado em [25],[27], é outra melhora do protocolo MIPv6, que também tem o objetivo de reduzir a latência do *handover*. Dito padrão utiliza como estratégia, a entrega dos pacotes no novo ponto de conexão o mais rápido possível. A ideia básica do FMIPv6 é que o MN possa antecipar o processo de *handover* e informar ao o novo *router* de acesso (*newAR*) sobre o *handover*, utilizando o L2 *trigger* para iniciar o processo de *handover* L3, enquanto o MN ainda está no velho *router* de acesso (*oldAR*), e o L2 *handover* não tenha sido completado. Isto reduziria, o tempo necessário para que o MN detecte o movimento (i.e., *proactive handover based*) ou o início do procedimento de L3 *handover* depois finalizado o L2 *handover* (i.e. *reactive handover based*).

3.2.1.2 - DMM

Com o gerenciamento de mobilidade distribuído se pretendem eliminar os problemas de ponto único de falha, congestão da rede pelo processamento centralizado, demoras devido á ubiquação geográfica (a entidade centralizada fica longe da ubiquação do MN), problemas de escalabilidade, etc. As melhoras de desempenho são logradas pela repartição das funcionalidades das entidades centralizadas do CMM em diferentes entidades distribuídas no DMM. Como consequência negativa se tem o aumento da quantidade de sinalização de controle e que a arquitetura seja mais complexa em termos de definição das entidades.

Em [35] Bernardos, Carlos J. Zunniga and J. C. Reznik, A, propõem uma solução DMM baseada no host que evidencia o reuso de elementos de arquitetura e interfaces. Na proposta novas entidade de rede chamadas D-GW (*Distributed Gateways*) são distribuídas na borda da rede, perto dos UE na rede de acesso. Os D-GW implementam as funcionalidades do DSMIPv6 *Home Agent*, interatuam diretamente com os UE e portanto desenvolvem as funcionalidades de acesso e roteamento. Tem a vantagem que para as conexões PDN na HPLDN (*Home Public Land Mobile Network*) não se emprega nenhuma funcionalidade extra, ficando transparente para o UE e o resto das entidades da rede. O D-GW também

substitui ao ePDG fornecendo as funcionalidades de *IpSec tunneling* para o UE, em acessos *untrusted non-3GPP*. Nos casos de acesso 3GPP trabalha como um simples *relay* entre o eNB e o SGW. A semelhança do D-GW com o PGW faz possível reutilizar a implementação do software *stack*. A proposta baseada no host tem como desvantagem que os UEs devem desenvolver funcionalidades extras, como gerenciamento do endereço IP, seleção de endereço destino ou associação dos tráfegos com o tipo de gerenciamento (alguns tráfegos distribuídos e outros centralizados).

Em [31] os autores propõem uma técnica conhecida como HIMALIS (*Heterogeneity Inclusion and Mobility Adaptation Through Locator ID Separation*). O gerenciamento de mobilidade é baseado no conceito da separação do identificador (ID) e localizador. O tráfego final é encaminhado por uma rota ótima que varia com a movimentação do usuário na seleção dos diferentes localizadores envolvidos, mantendo o identificador constante. Tem como desvantagem que o *stack* de protocolos no host tem que ser desenvolvido quase a totalidade.

Outra proposta focada na otimização de rotas pode ser olhada em [29]. Chung-Sheng Li, Fred Lin and Han-Chieh Chao, propõem um esquema tipo DMM para o protocolo básico de mobilidade de rede NEMO. A ideia é criar um conjunto de agentes caseiros distribuídos (DHA) que proveem roteamento ótimo entre o MN e o CN. Além dos DHA também são criadas as entidades HALRA (*Home Agents Location Registration Agents*) que ajudam no processo de seleção da rota. Propõem uma alternativa para otimização da rota que evita o processo *NEMO Routing Optimization*, diminuindo os problemas derivados do mesmo, os retardos e a perda de pacotes, economizando largura de banda.

Uma extensão DMM do MIPv6, para dar solução ao problema de roteamento triangular e à diminuição da latência de *handover*, tem sido proposta em [32]. Conhecida como ADA (*Asymmetric Double-Agents*), introduz dois agentes móveis para tratar a comunicação fim a fim. Um agente vai ser localizado perto do CN conhecido como CMP (*Correspondente Mobile Proxy*) e o outro perto do MN conhecido como LMP (*Local Mobile Proxy*). O principal objetivo do CMP é diminuir a distância entre o CN e o HA do MN, e minimizar o *overhead* de roteamento. O LMP atua como a entidade MAP do HMIPv6, fazendo o gerenciamento local no domínio. Esta proposta além de diminuir a latência de *handover* e a latência fim a fim também prove mecanismo específicos de avaliação da QoS.

Liu, Y., et al.[30] propõem uma arquitetura plana para o gerenciamento de mobilidade distribuído. Empregando o conceito de separação da identificação (MN ID) e localização (MN *location*) do MN, são criadas *HashTables* distribuídas que junto a mecanismos de *bi-casting* locais, otimizam o processo de *handoff*. É uma proposta centrada no cliente pois as entidades das bordas da rede tem que trocar uma serie de mensagens com o MN. Na solução as entidades empregam um modulo de *handoff* próprio, o que pode ser uma desvantagem na implementação.

Outro esquema baseado em DHT (*Distributed Hash Table*) é proposto em [34]. Tanto os MN como os CN tem que disponibilizar os serviços de MIPv6. Os autores tratam a rede como um sistema P2P no qual as funcionalidades do HA são distribuídas nas diferentes entidades da rede. Também é proposto um algoritmo para lograr um roteamento eficiente, otimizando as rotas. A solução reduz a inconsistência de topologias física e logica no gerenciamento de mobilidade, otimizando o performance de atualização e *query*, em redes de grande escala.

Em [33] outra proposta de arquitetura plana é desenvolvida. A solução é baseada na atualização entre os *routers* do domínio. Para isto empregam um DNS, atualizado constantemente, para determinar o prefixo de rede associado aos nós móveis. O emprego do BGP (*Border Gateway Protocol*) permite a atualização entre os *routers* intra domínio. Esta proposta entra na categoria das baseadas no host devido a que o mesmo tem a tarefa de se atualizar no DNS mas, também pode ser considerada como uma proposta hibrida.

Finalmente uma solução DMM hibrida é apresentada por, F. Giust, C. Bernardos, e A. de la Oliva, em [10]. A proposta é composta de duas partes independentes, uma solução baseada no cliente e outra baseada na rede, que será comentada em seções posteriores. A solução baseada no cliente, é uma evolução da arquitetura MIPv6, para implantação em redes flat. As funcionalidades do HA são distribuídas e movidas para a borda da rede, então uma instancia dela é implantada em cada *default gateway* (DAR - *Distributed Anchor Router*), ao que o MN vai se ligar. Quando o MN estabelece uma ligação com um novo DAR, as novas sessões serão roteadas diretamente, atuando o novo DAR como HA, só as comunicações em andamento serão roteadas através dos túneis entre o prévio DAR e o DAR que está servindo ao MN.

3.2.2 - Mobilidade baseada na rede

Os protocolos de mobilidade baseados na rede pretendem atingir os mesmos objetivos que os baseados no host. Eles diferem em que as decisões do gerenciamento de mobilidade são feitas sem a participação ativa do MN. A ideia principal desta classe de protocolos é que seja mantida a mesma pilha de protocolos IP nos UE pelo que mudanças e otimizações em termos de infraestrutura física e de software não levam a atualizações ou ao desenvolvimento de novos terminais.

Ao igual que os protocolos baseados no host, as diferentes propostas podem se classificar em centralizadas o distribuídas. A continuação é feito um resumo dos trabalhos relativos a esta classe de arquiteturas.

3.2.2.1 - CMM

No gerenciamento de mobilidade centralizado baseada na rede se tem duas entidades no núcleo da rede: LMA (*Local Mobility Anchor*) e MAG (*Mobile Access Gateway*). O LMA trabalha como entidade centralizada onde em um domínio PMIP todos os MAG tem saída através do LMA. Quando o usuário faz um deslocamento uma troca de mensagens entre os MAG envolvidos e o LMA é feita para assim estabelecer as novas políticas de roteamento.

A proposta mais relevante desta categoria é conhecida como PMIPv6 [3] (*Proxy Mobile IP Version 6*) e é a contrapartida baseada na rede do MIPv6. Neste protocolo a rede é a encarregada de fazer qualquer atualização referente à localização do MN. O LMA tem como função principal a manutenção do *tracking* do MN mediante o emprego de registro de *binding*. As entidades MAG situadas nos *routers* de acesso notificam mediante uma troca de mensagens PBU (*Proxy Binding Update*) e PBA (*Proxy Binding Acknowlegment*) que o MN vai começar a ser atendido pela mesma. O LMA atualiza sua *Binding Cache* e os novos pacotes que chegassem para o MN vão ser reencaminhados para a MAG que atende atualmente ao mesmo. Se o MN se desloca de novo, na entrada na nova MAG se faz o mesmo procedimento e o tráfego da sessão em andamento vai ser encaminhado via um túnel IP para a nova MAG. Uma atualização do PMIPv6 é provida em [37] onde se estabelecem políticas para a assinação de endereços de camada de enlace reservando um conjunto de endereços associados a cada um dos diferentes tipos de entidades da rede.

Para PMIPv6 também existe uma otimização conhecida como FPMIPv6 [38] *(Fast PMIPv6*) mediante o emprego do protocolo FMIPv6 mas em uma arquitetura centrada na rede. Os

autores propõem um conjunto de técnicas para tratar o cenário onde o MN não tem nenhuma funcionalidade de mobilidade IP. Em [27] é proposta uma atualização para o FMIPv6 brindando suporte especifico para clientes *Multicast* (*IPTV Clients*), melhorando ainda mais o desempenho do FMIPv6.

Em [36] é proposto um mecanismo para balanceamento de carga nas MAG, evitando sobrecarga nelas. No esquema proposto as MAGs periodicamente trocam sua informação de carga e fazem uma lista de MAGs candidatas. No processo de balanceamento de carga, o MAG que está servindo, comanda ao MN fazer *handover* para outra entidade MAG, reduzindo sua carga. A MAG que está servindo não seleciona uma MAG que tenha valores de carga pertos de seu umbral, evitando com isto o efeito *ping-pong*. Esta solução mesmo as MAGs troquem mensagens de controle entre elas, ainda são centralizadas, pois o LMA é a entidade que gerencia os *bindings* e todo o tráfego de dados. É importante ressaltar que o mecanismo de balanceamento de carga proposto, não suporta escoamento seletivo dos fluxos, os MN não vão receber fluxos simultaneamente por duas interfaces, elas serão orientados a fazer *handover* (movimentar todas as seções), em áreas de sobreposição.

3.2.2.2 - DMM

Nas propostas distribuídas baseadas na rede, se persegui fazer todo o gerenciamento de mobilidade no lado da rede, e distribuir as funções de localização e encapsulamento dos dados, em entidades na borda da rede. Com isto evitam ter pontos únicos de falhas e ataque, melhorando a escalabilidade e confiabilidade na rede, sim mudanças no *stack* de protocolos do MN. As soluções DMM podem ser classificadas em dois grupos, parcialmente distribuídas e totalmente distribuídas.

- P- DMM

Nas soluções P-DMM, são separadas as funções de localização e *handover* (plano de controle), do roteamento de tráfego de dados (plano de dados). O plano de dados é distribuído e o plano de controle centralizado. Com isto, o objetivo principal que se persegue é a otimização das rotas.

No artigo [40], os autores apresentam uma extensão para PMIPv6, com o fim de otimizar o roteamento dos dados. A ideia é dar aos LMA a habilidade de controlar o caminho de cada fluxo, selecionando um conjunto intermédio de data *anchors* a ser atravessados. Logo a proposta desassocia a sinalização e o caminho dos dados, usando o LMA como *anchor* de

sinalização e os IAs (*Intermediate Anchors*) como *anchor* de dados. A solução faz escoamento de dados, mas no núcleo da rede, não na rede de acesso, é por isso que não é considerada como uma solução que suporte IFOM. O principal problema da proposta é que não pode fornecer o caminho ótimo, mas apenas uma aproximação.

Outra extensão de PMIPv6 para otimização da rota é proposta em [41]. Nesta solução, os MAGs que servem ao MN e ao CN, aproveitam a informação armazenada no LMA, para estabelecer um túnel direito entre eles, então o melhor caminho poder ser usado na comunicação. Dita solução só é aplicável se o CN está no mesmo domínio PMIPv6 que o MN.

Em [12],[42] duas soluções DMM são descritas, uma totalmente distribuída que será discutida no próximo apartado, e outra parcialmente distribuída. A solução distribui os *mobility anchors* na borda da rede, com o que ficam perto do usuário. Também introduze duas entidades de rede *Anchor - Mobility Anchor and Access Router* (A-MAAR) e *Serving - Mobility Anchor and Access Router* (S-MAAR). O S-MAAR é o MAAR onde, o MN envolvido na comunicação, está ligado, enquanto o A-MAAR é MAAR que assigna o prefixo de rede. Em relação à semântica de PMIPv6, um S-MAAR atua como um MAG, e um A-MAAR como um LMA. No esquema parcialmente distribuído, o *Central Mobility Database* (CMD) é adicionada para armazenar os *mobility bindings* dos MNs. Essa entidade mantem as funções de controle do LMA, mas não participa no reenvio dos pacotes no plano de dados. Logo na solução os MAARs gerenciam o plano de dados o que é distribuído, enquanto o plano de controle é centralizado utilizando o CMD.

Outros três esquemas para controle de mobilidade distribuídos, são propostos em [45]: *Signal-driven PMIP* (S-PMIP), *Data-driven Distributed PMIP* (DD-PMIP), and *Signal-driven Distributed*. S-PIMIP é parcialmente distribuído, enquanto DD-PMIP e SD-PMIP são totalmente distribuídos. No S-PMIP o plano de controle é separado do plano de dados. As operações de atualização dos *bindings* são realizadas como em PMIPv6. Para a entrega dos dados, o MAG do CN realiza *binding query operation*, com o LMA para encontrar a localização do MN. Logo envia os pacotes direto ao MAG que serve ao MN, sendo o plano de dados distribuído.

Em [17] é descrita uma proposta que sugere dividir as funcionalidades do LMA de PMIPv6 em dois nós distintos: o CLMA (*Control plane LMA*) e o DLMA (*Data plane LMA*). O

primeiro mantem o controle dos *bindings* dos MNs, e o segundo o tráfego de dados. Essa proposta em geral não representa uma arquitetura plana, pois a hierarquia do DLMA/MAG e CLMA/MAG é preservada. Porém, prevê um modo de operação, no qual se o MN e o CN estão embaixo a administração do mesmo CLMA, uma otimização de rotas pode ser configurado entre as MAGs correspondentes.

Em [18],[46] foi proposta a solução ePMIP (*enhanced Proxy Mobile IPv6*) para suportar DMM. Neste esquema, duas funções logicas foram introduzidas: *Location Management Function* (LMF) e *Distributed Anchoring Function* (DAF). O LMF mantem o mapeamento entre o endereço IP e a informação de localização dos MNs, e pode ser implementado no LMA constituindo o eLMA (*evolved LMA*). O DAF inclui, *Distributed Routing sub-Function* (DRF) que habilita otimização de rotas entre o MN e o CN, e o *Distributed Mobility sub-Function* (DMF) que garante a mobilidade do MN, quando a otimização de rota é habilitada. O DAF pode ser implementado no MAG de PMIPv6 constituindo um eMAG (*evolved MAG*). O draft principalmente considera otimização de rotas e comunicações inter-eLMA. Porém, ainda existe um ponto único de falha, devido a que o esquema precisa a função de controle LMF, estando todas as mensagens de controle concentradas no eLMA. Também no draft [19] outro método para otimização de rotas em uma arquitetura distribuída, baseado em PMIPv6, é proposto.

Em [39], os autores realizam uma análises das diferentes funções de mobilidade fornecidas pelo PMIPv6.Propõem uma divisão destas funções através de vários nós na rede. No entanto, a solução proposta utiliza para o roteamento real dos fluxos uma abordagem centralizada, portanto, não há uma verdadeira mobilidade distribuída.

Hahn, W. propõe em [43-44] mecanismos para a detecção de rotas ótimas. O objetivo não é fazer uma realocação para todos os fluxos, mas, classificar quais aplicações detectam o impacto de rotas não ótimas. O foco destes trabalhos não é um protocolo de gerenciamento de mobilidade mesmo, porém, podem se situar nesta categoria pois empregam protocolos de gerenciamento parcialmente distribuídos onde o plano de dados segue uma arquitetura distribuída e o plano de controle ainda é centralizado.

Como discutido em seções anteriores os autores, F. Giust, C. Bernardos, e A. de la Oliva, em [10] descrevem uma solução DMM hibrida, que tem uma abordagem independente baseada na rede. Dita abordagem é uma extensão do padrão PMIPv6 para operar em uma

arquitetura distribuída. Os autores propõem uma solução parcialmente distribuída, onde no plano de controle utilizam uma entidade centralizada, chamada CDM (*Central Mobility Database*). O CMD implementa todas as funções relativas à atualização da *Binding Cache*, como é feito pelo LMA em PMIPv6.

Finalmente uma abordagem similar à descrita em [31], é seguida no draft [47] onde a divisão da localização e o identificador, são obtidos a partir do uso do protocolo LISP (*Locator Identifier Separation Protocol*).

Todas as soluções antes discutidas, ainda tem entidades centralizadas no plano de controle, onde seguem presentando problemas de escalabilidade, confiabilidade e ponto único de falha e ataque.

- F- DMM

Em [10] os autores também descrevem uma solução totalmente distribuída baseada em PMIPv6. Para a solução, propõem dois mecanismos alternativos: (i) fazendo *multicasting* do PBU enviado pelo S-DAR (*Serving-DAR*), no grupo formado por todos os DAR no domínio, (ii) utilizando os serviços da especificação *Media Independent Handover* (IEEE 802.21).

Além da proposta parcialmente distribuída tratada em [12],[42] os autores apresentam um totalmente distribuída onde os MAARs vão tratar tanto o plano de dados como o plano de controle. Para resolver o problema da descoberta dos MAARs no processo de atualização do plano de controle distribuído, os autores propõem uma arquitetura P2P ou fazer mediante consultas *Unicast, Multicast* ou *BroadCast*.

Em [50] é proposto um esquema chamado D-NEMO, que é a combinação de PR-NEMO (*Proxy Router NEMO*) [59] e DMA [16], mas a diferencia da proposta [16], a solução não dá suporte para seletivamente movimentar fluxos IP. Uma nova entidade é definida, chamada *Proxy Router*, que é utilizada, para realizar registro da localização em nome do MN com o *router* de acesso e gerenciar o ID e prefixo de cada MN.

Jaehwoon Lee e Younghan Kim, [11] apresentam outro mecanismo de gerenciamento totalmente distribuído. As funcionalidades do LMA são replicadas em todos os AR no domínio PMIPv6. Tem a particularidade de que todas as MAGs compartem o mesmo prefixo de rede. Isto facilita que as MAG tenham conhecimento se um MN vai entrar no domínio

pela primeira vez dependendo se faz uma solicitação DHCP ou então é que está fazendo um *handover*.

Os autores do trabalho [35], discutido em seções anteriores, também fazem uma proposta F-DMM baseada na rede. Neste cenário os D-GW se comportam como AR e como agentes de sinalização de mobilidade. Ao mesmo tempo, incluem as funcionalidades do LMA para os casos de redes baseadas em PMIPv6.

Em [48] outro trabalho referente a F-DMM destaca pela sua caraterística dinâmica nos AR. Os *routers* de acesso vão trabalhar de dois jeitos diferentes dependendo o tipo de fluxo gerado pelo MN. A ideia de trabalho do esquema é a seguinte: os nós de acesso (AN), trabalham como AAN (*Anchor Access Node*) quando eles tem o MN associado. Se o MN se registra em outro AN produto do deslocamento, então este novo AN vai funcionar como VAN (*Visited Access Node*). A ideia é reencaminhar os fluxos IP existentes no AAN para o VAN mediante um túnel IP mas sem precisar de trocar mensagens de controle entre os AN. Para isto o VAN faz uma leitura dos pacotes que chegam a ele, olhando os cabeçalhos dos pacotes do *downlink* e estabelece uma relação com os pacotes que chegam do *uplink* do MN. O problema principal desta proposta é que o MN tem que estar fazendo upload de pacotes *void*, no caso que o fluxo seja só no sentido CN para MN, pois o VAN precisa desses pacotes para poder estabelecer a relação entre os fluxos.

DIMA (*Distributed IP Mobility Approach*) é apresentada em [49] como outra vertente F-DMM. O protocolo é baseado no MIP onde o HA é distribuído nos nós da rede. Também tem um banco de dados distribuído (comum para todos os nós) que se atualiza com a chegada dos pacotes na rede. Neste último sentido o MN não participa na atualização da localização sendo os HA distribuídos os encarregados da troca de mensagens PBU e PBA. Esta proposta também faz uso do conceito de DHA (*Distributed Hash Tables*).

Finalmente em [45] são exploradas propostas para manutenção do plano de dados e o plano de controle distribuídos. A ideia é empregar redes P2P ou *multicast* para encaminhar os pacotes a través da rota ótima. Uma vez conseguida a otimização o resto dos pacotes pertencentes a esse MN serão re encaminhados seguindo a mesma rota.

3.2.3 - Considerações parciais

Nesta Seção foram analisadas as soluções e padrões existente para o gerenciamento de mobilidade na camada de rede. As propostas distribuídas baseadas na rede, são apresentadas como as melhores vertentes, já que evitam problemas como ponto único de falha, escalabilidade e confiabilidade, assim como mudanças nos dispositivos de usuário e diminuição da complexidade dos mesmo. As últimas vantagens se devem a que uma entidade da rede é encarregada do gerenciamento de mobilidade em lugar do MN.

Não obstante, na atualidade existe um crescimento exponencial de usuários móveis o que leva ao desenvolvimento de novas estratégias, com o fim de aumentar a vazão e lograr balanceamento de carga na rede de acesso, tentando oferecer melhor QoS. Os fabricantes de terminais móveis já tem desenvolvido dispositivos multimodais, com diferentes interfaces de rádio integradas, o que gera um ambiente heterogêneo nas redes de acesso. Logo, os MNs podem fazer uso de duas redes de acesso com diferentes tecnologia simultaneamente. Isto levou ao desenvolvimento de extensões ou novas soluções de gerenciamento de mobilidade que permitem seletivamente movimentar fluxos IP entre redes de acesso com diferentes tecnologias. Na próxima Seção será feito um estudo de ditas propostas, as que além de permitir ao usuário movimentar-se mantendo as sessões em andamento, aumentam a qualidade da experiência, possibilitando ao usuário utilizar em qualquer lugar e momento a melhor rede de acesso disponível.

3.3 - COM SUPORTE PARA MOBILIDADE DE FLUXO IP

Para facilitar escoamento de dados, na camada de rede, é preciso que a rede suporte eficientemente mobilidade IP. Os protocolos descritos na Seção anterior, provem opções para movimentar todos ou nenhum dos fluxos IP através da rede (*handover*), mas para ter melhor balanceamento de carga na rede, são necessárias opções para movimentar determinados fluxos IP de um dado usuário, o que é conhecido como mobilidade de fluxo IP (*flow mobility*). Logo, o escoamento de dados pode-se realizar através da movimentação de fluxos na camada de rede. Na literatura existem trabalhos que permitem a movimentação seletiva de fluxo baseados em soluções, para o gerenciamento de mobilidade, centralizadas ou parcialmente distribuídas. Nesta Seção apresentamos um compreensivo *survey* das técnicas para o gerenciamento de mobilidade de fluxo, que suportam escoamento de dados, através da movimentação seletiva de fluxos IP.

3.3.1 - Mobilidade baseadas no host

Nas primeiras tentativas de solucionar o problema de mobilidade dos usuários na camada IP, foi proposto pela IETF o protocolo MIP [1-2]. Neste protocolo o MN tem uma participação ativa no gerenciamento de mobilidade. Isto traz grandes latência de *handover* e perda de pacotes, produto aos problemas do enlace sem fio. Logo, na atualidade, não tem sido desenvolvidas muitas soluções deste tipo.

A continuação discutimos três propostas baseadas no host, com suporte para escoamento de dados, que achamos na literatura. Ditas soluções são classificadas em centralizadas ou distribuídas, dependendo de sua arquitetura.

3.3.1.1 - CMM

O 3GPP em [6-7] especifica uma descrição para mobilidade de fluxo IP entre redes 3GPP e WLAN. A solução técnica é baseada nos princípios de trabalho do protocolo DSMIPv6 [26], e é aplicável a arquiteturas EPC (*Evolved Packet System*) e I-WLAN (*Interworking Wireless Local Area Network*) [60]. No processo de mobilidade de fluxo IP, o UE é assumido que estará conectado simultaneamente via um acesso 3GPP e um acesso WLAN. O UE usara ambos acesso para a mesma conexão PDN (*Packet Data Networks*). Subsequentemente, o UE pode adicionar, modificar, eliminar, ou mover os fluxos IP, entre as redes de acesso utilizando mensagens DSMIPv6. Também o escoamento de fluxo pode ser iniciado pela entidade centralizada na rede, PGW, ou pelo UE, sempre que seja aprovado por ambas partes.

Esta solução permite fazer balanceamento de carga na rede, e oferece maior vazão ao UE, mas pode ter problemas de ponto único de falha e escalabilidade, pela sua característica centralizada. Também os UE enviam mensagens de sinalização no acesso sem fio, o que pode aumentar a perda de ditos mensagens, precisando retransmissões, o que tem um impacto negativo na latência do processo.

3.3.1.2 - DMM

Com o fim de dar solução aos problemas de CMM, existem propostas baseadas em gerenciamento de mobilidade distribuído. Jong-Hyouk, L., et al. em [9] propõem um protocolo para suporte de mobilidade IP baseada no usuário, que utiliza *mobility anchors* na rede de acesso. O protocolo não adota a entidade centralizada (HA – *Home Agent*), como

alternativa, utiliza entidades distribuídas denominadas *Access Mobility Anchors* (AMAs). Um AMA funciona como um *router* de acesso e aloca o prefixo de rede ao MN. No artigo utilizam o esquema de integração das tecnologias de acesso 3GPP e WLAN proposto pelo 3GPP [60], onde as funcionalidade do HA são removidas do PGW e distribuídas nos SGW, gateway de acesso (A-GW, *Access Gateway*), e ePDG (*Evolved Packet Data Gateway*). O protocolo construí um sistema de escoamento onde o MN toma a decisão de movimentar fluxos IP entre as suas interfaces. Nesta proposta a complexidade do MN pode aumentar, produto de algoritmos de decisão de escoamento dinâmicos, que em geral precisam maior processamento. Então propostas baseadas na rede, que minimizem as mudanças no dispositivo de usuário podem ser preferíveis.

3.3.2 - Mobilidade baseada na rede

Nos protocolos de mobilidade baseados na rede, uma entidade da rede faz o controle da mobilidade em nome do MN. Com isto a ideia principal é que seja mantida a mesma pilha de protocolos IP nos dispositivos de usuário. Logo as mudanças e otimizações em termos de infraestrutura física e de software, não levam a atualizações ou ao desenvolvimento de novos terminais.

A continuação apresentamos um estado da arte desse tipo de esquema, que oferecem possibilidades de escoamento de dados. Ditas soluções utilizam a vantagem, dos terminais multimodo, de ter duas interfaces, para melhorar a QoS oferecida aos usuários. Em nossa pesquisa as classificamos segundo o jeito de fazer gerenciamento de mobilidade em: centralizadas, e distribuídas, como sugerido na Tabela 3.1.

3.3.2.1 - CMM

Um esquema de mobilidade de fluxo baseado em PMIPv6 é proposto por Choi, H.-Y., et al. em [51],[53]. Desenham o suporte para mobilidade de fluxo baseado em uma interface logica no MN. Introduzem duas componentes para o gerenciamento de mobilidade de fluxo: o gerenciador de interface de fluxo (*Flow Interface Manager*) e o gerenciador da ligação de fluxo (*Flow Binding Manager*). O gerenciador de interface de fluxo é colocado na camada de interface logica no MN (*Mobile Node*) e o gerenciador da ligação de fluxo, na camada de rede do LMA (*Local Mobility Anchor*), sendo par um do outro. Ditas componentes estabelecem políticas de fluxo, que são utilizadas para selecionar a tecnologia de acesso pela que devem ser enviados os pacotes. Os autores dividem o procedimento de mobilidade de

fluxo em três casos, o MN estabelece uma nova conexão, decisão do LMA e decisão do MN aprovada pela LMA. Com isso a decisão de movimentar um fluxo pode ser tomada tanto pelo usuário quanto pela rede, mas sempre com o consentimento da última.

Em [52] Melia, T., et al. centram-se na concepção e implementação de extensões de mobilidade de fluxo para PMIPv6. Descrevem os componentes funcionais necessários na rede para suportar o direcionamento inteligente do tráfego, minimizando o impacto sobre os dispositivos móveis e aumentando a QoE do usuário. Na proposta, a rede (em particular o LMA) é a entidade de controle de decisão. Ela executa mobilidade de fluxo com base em políticas do operador da rede, que podem reagir de forma dinâmica sobre a carga da rede.

Makaya, C., et al. [55] trabalham em redes de comunicações veiculares e propõem um esquema centralizado chamado *Multilink Striping Manager* (MSM). Este esquema permite escoamento de dados entre diferentes tecnologias de redes de acesso. Quando um evento de ligação é detectado, o MN utiliza primitivas MIH (*Media Independent Handover*) para iniciar o pedido de mobilidade de fluxo IP. A qualidade da ligação sem fio e o estado da rede são monitorados, para ajudar na toma de decisão. Também Meneguette, R. I., et al. em [56] propõem uma arquitetura para gerenciamento de mobilidade transparente de fluxo, baseada em classes de aplicações de redes veiculares, com gerenciamento de mobilidade baseada na rede.

A 3GPP em [6] apresenta cenários de estudo, requisitos e soluções para os UE com multiplex interfaces, que poderão estar simultaneamente conectados a um acesso 3GPP e um acesso WLAN não 3GPP (i.e. WiFi). O escoamento de fluxo pode ser iniciado pela entidade centralizada na rede, PGW, ou pelo UE, sempre que seja aprovado por ambas partes. A especificação técnica é baseada nos protocolos PMIPv6 e GTP, e como tratado em seções anteriores, no protocolo baseado no host DSMIPv6. Quando são utilizados os protocolos baseados na rede, o UE não participa na sinalização relativa ao gerenciamento de mobilidade, mas pode solicitar escoamento de fluxo.

Um mecanismo optimizado para *handover* transparente de fluxo IP, com base no protocolo FPMIPv6 [38] e inicialização do móvel, é proposto por Jinho, K., et al. em [54] . O mecanismo aumenta o desempenho de *handover*, permitindo a utilização simultânea de várias interfaces durante a mobilidade de fluxo.

As propostas mencionadas, brindam a possibilidade de escoamento seletivo de tráfego, mas tem os problemas próprios do gerenciamento de mobilidade centralizado.

3.3.2.2 - DMM

Produto do aumento de tráfego nas redes móveis, na atualidade existe uma preferência por desenvolver, soluções distribuídas para o gerenciamento de mobilidade IP. Esso é devido aos problemas de escalabilidade derivados de soluções centralizadas. Na literatura existem propostas parcialmente distribuídas, que também suportam mobilidade seletiva de fluxos IP, com o que pode aumentar a qualidade dos serviços percebida pelos usuários.

Soluções de gerenciamento de mobilidade distribuída, podem incluir separar o plano de dados e o plano de controle como D-PMIPv6 [17], que divide o LMA em duas entidades, CLMA (*Control plane Local Mobility Anchor*) e DLMA (*Data plane Local Mobility Anchar*). Keqiang, X., et al. em [13] acrescentam, a esta solução, suporte para mobilidade de fluxo IP. Os autores utilizam roteamento baseado em fluxos e uma interface lógica no MN. Estabelecem políticas de roteamento, suportando diferenciação de serviços através do marcado de pacotes. No artigo, são propostas duas abordagens para iniciar o escoamento, uma baseada em mudanças da rede e outra baseadas nos desejos do terminal móvel. Esta proposta pode diminuir os problemas de escalabilidade, demoras de processamento e fiabilidade das propostas centralizadas, mas não os eliminam. As entidades CLMA e DMLA são centralizadas em cada plano e precisam ter a capacidade para gerenciar todos os nós da rede, além de ser um ponto único de falha.

Outro esquema parcialmente distribuído baseado na rede é proposto por Sun, K. e Y. Kim em [14]. Neste esquema, as funções de gerenciamento de mobilidade são realocadas para a borda da rede. Para fornecer roteamento e gerenciamento de mobilidade, as entidades de rede trocam mensagens de sinalização e fazem túneis sem ter nenhuma entidade centralizada. Mas para implementar o esquema de fluxo, utilizam uma entidade central chamada MCF (*Mobile Control Funtion*) que funciona como uma base de dados, coletando informação de todos os MN na rede, incluindo identificador, endereço e tipos de tráfego. Como o MCF tem informação de todos os MN no domínio é encarregada de tomar as decisões relativas ao escoamento de fluxo. Embora o MCF não participa no reenvio de dados, ainda pode ter problemas de escalabilidade e é um ponto único de falha, que armazena informação de todo

os MN na rede. Com o propósito de dar solução a este problema, os autores propõem ter vários MCF na rede, que compartilhem informação entre eles.

Por outro lado P. Seite, P. Bertin, and J. Lee, em [16] descrevem uma solução parcialmente distribuída baseada em PMIPv6. Nesta proposta os túneis entre os MARs *(Mobility capable Access Router*), só são utilizados para seções em andamento que foram iniciadas antes de *handover*. Os pacotes de dados das novas seções abertas no MAR que serve ao MN, serão roteados direto. Para otimizar as rotas, cada MAR prévio que ainda tem seções do MN, em andamento, estabelecerão um túnel com o MAR que serve ao MN, com o fim de manter a continuidade das seções. O plano do controle é centralizado utilizando uma base de dados centralizada, mas a interação entre os MAR e a base de dados não é especificada no draft. Os autores também descrevem o suporte de mobilidade de fluxo IP, em terminais com multiplex interfaces, habilitando o escoamento de dados na solução.

Perras, M. and J. Cartmell em [15] consideram mobilidade em uma rede de gateways de células pequenas, com a capacidade de trabalhar em bandas celulares e não licenciadas, combinando um ponto de acesso WiFi e estações base celular dentro de um dispositivo único, denominado *Converged Gateway* (CGW). Descrevem métodos para suportar gerenciamento local de fluxos IP entre tecnologias de acesso WiFi e celular, assim como soluções IFOM (*IP Flow Mobility*) baseados no CGW. A solução pode ser estendida para suportar gerenciamento de mobilidade distribuído utilizando vários CGWs. Esta proposta implica utilizar os dispositivos CGW, o que pode aumentar o custo de implantação da rede.

Também uma arquitetura escalável para o gerenciamento de mobilidade baseada na rede e um esquema de gestão de mobilidade de fluxo, no contexto *multi-access* e *multi-homing* é apresentada em [8]. Esta arquitetura não é baseada nos protocolos para gerenciamento de mobilidade padronizados pela IETF. A mesma, consiste de quatro entidades funcionais: *Mobility Information Control Server* (MICS), *Handover Control Agent* (HCA), *Point of Attachment* (PoA) e MN. O MICS está localizado no núcleo da rede e o HCA no gateway na rede de acesso. Esses dois nós, são os encarregados do gerenciamento de mobilidade. No esquema de gestão de mobilidade IP utilizam tabelas de políticas no MN e no MICS, que determina a interface correta para o envio de um fluxo, baseado na prioridade das tecnologia de acesso. As tabelas de políticas podem ser modificas pelo usuário ou pela rede, assim como a lista de prioridades dos ATT (*Access Tecnology Type*).

Uma outra solução parcialmente distribuída é proposta por Purohith, D. R., et al. em [5]. Eles apresentam uma arquitetura que utiliza os conceitos de PMIPv6 e redes definidas por software (SDN – *Software Defined Networking*), chamada *Seamless Internetwork Flow Mobility* (SIFM). Dita arquitetura define um *Flow Controller* (FC) similar ao controlador *OpenFlow*. Quando um switch recebe um pacote que ele não tinha processado antes, o reencaminha ao controlador. O controlador toma as decisões de roteamento, e instrui ao switch de como ele deve reenviar pacotes similares, adicionando uma entrada na tabela de fluxo do switch. O FC só realiza funcionalidades relacionadas com mobilidade. O PGW em redes LTE e os WAG (*Wireless Access Gateway*) em redes WiFi atuam como switches *OpenFlowhybrid* que executam a sinalização relativa à mobilidade em nome do UE. Eles seguem as instruções do FC quando o MN se movimenta de uma rede LTE para uma WiFi, com o fim de prover transição transparente. O FC é o encarregado de tomar as decisões de movimentação de fluxo, estabelecer as regras e informar as outras entidades da rede.

Todas as soluções anteriormente mencionadas, oferecem possibilidade de movimentar fluxos IP, mas de um jeito parcialmente distribuído. Precisando de uma entidade centralizada com capacidade de processar informação de todos os nó da rede. Logo existe uma carência de soluções totalmente distribuídas, com suporte para mobilidade seletiva de fluxo IP, que aproveite das novas tecnologias para oferecer a melhor qualidade de experiências aos usuários.

No Capitulo 4, é apresentada uma solução para o gerenciamento de mobilidade distribuído denominada SIFDMM, que permite fazer escoamento de dados na camada de rede, movimentando fluxos IP. A proposta é baseada na rede, mas assistida pelo usuário e minimiza os problemas de confiabilidade, escalabilidade e ponto único de falhas das soluções com entidades centralizadas, pois tanto o plano de dados quanto o plano de controle são distribuídos.

A Tabela 3.2 resume as seguintes características de os trabalhos discutidos nesta Seção: (i) se o gerenciamento de mobilidade é centralizado, parcialmente distribuído ou totalmente distribuído, (ii) as entidades de rede que conformam a arquitetura, (iii) qual é a entidade encarregada da toma da decisão de escoamento de fluxo, (iv) novas mensagens definidas na arquitetura e (v) as tecnologias de rede consideradas na solução.

Tabela 3.2: Comparação dos esquemas para o gerenciamento de mobilidade de fluxos IP.

Proposta	**Gerenciamento de Mobilidade**					**Entidades de Rede**	**Decisão de escoamento**	**Novas Mensagens**	**Tecnologia de Acesso**
	Baseado no *host*		Baseado na rede						
	CMM	DMM	CMM	DMM					
				P-DMM	F-DMM				
3GPP [6-7]	X		X			SGW, PGW, ePDG	UE, PGW	-	- 3GPP, - WLAN
L. Jong-Hyouk, et al. [9]		X				AMA	MN	*Access Binding Update* (ABU) *Access Binding Acknowledgment*(ABA)	- LTE, - WLAN (Untrusted)
H.-Y. Choi, et al. [51],[53]			X			MAG, LMA	MN, LMA	*HNP Update Request* (HUR) *HNP Update Acknowledge* (HUA)	- 3G, - WLAN, - WiMax
K. Jinho, et al. [54]			X			MAG, LMA	MN	*Handover Initiate for Flow mobility* (HIF) *Handover Acknowledge for Flow mobility* (HAF)	- 3G, - WLAN
C. Makaya, et al. [55]			X			OBU, RSU, MSM	MSM	Mensagens MIH	- 3G/LTE (RSU - *Roadside Units*), - WLAN (OBU - *Onboard Unit*)
R. I. Meneguette, et al. [56]			X			MAG, LMA	MN, MAG, LMA	Mensagens MIH	- LTE, - WLAN

T. Melia, et al. [52]			X			MAG, LMA	LMA	-	- 3G, - WLAN
X. Keqiang, et al. [13]				X		DLMA, CLMA, MAG	MN	*Flow Move Update* (FMU) Mensagem RS é estendida contendo informação sobre os fluxos.	- 3G, - WLAN
K. Sun and Y. Kim [14]				X		MCF *Mobile Function Routers* (MFR)	MCF	*Flow request message* *Flow response message*	- 3G, - WLAN
P. Seite, et al. [16]				X		MAR	Não especificado	-	Genérico
M. Perras and J. Cartmell [15]				X		CGW	CGW	-	- 3G, - LTE, - WLAN
H.-B. Lee, et al. [8]				X		MICS HCA	MN, MICS	-	- 3G, - WiMax, - WLAN
D. R. Purohith, et al. [5]				X		FC MA	FC	*Flow Modification Message* *Port Status Update*	- LTE, - WLAN
SIFDMM					X	FMA (*Flow and Mobility Anchor*)	FMA	*Flow PBU, Flow PBA* *Handover PBU, Handover PBA* Mensagem RS é estendida contendo informação sobre os fluxos.	- LTE, - WLAN

3.4 - CONSIDERAÇÕES FINAIS

No capítulo foi feito um estudo do estado da arte dos protocolos existentes para o gerenciamento de mobilidade na camada de rede. Primeiro foram estudadas as soluções e padrões que permitem movimentar todos os fluxos IP entre redes de acesso, com a mesma ou diferente tecnologia, durante um processo de *handover*. Mas o aumento de tráfego de dados vem trazendo novos desafios, o que tem levado à necessidade de desenvolver soluções com maior granularidade na camada de rede, que permitam movimentar seletivamente fluxos IP, com o fim de balancear carga na rede e aumentar por exemplo a vazão. Então também foram estudadas as soluções deste tipo que existem na literatura.

Os protocolos de gerenciamento de mobilidade baseados no host, apresentam problemas tais como elevada latência de *handover*, perdas de pacotes e *overhead* de sinalização. Também requerem modificações no *stack* do protocolos dos MN, em ordem de suportá-los. Ditos requisitos de modificação nos MNs podem incrementar sua complexidade e introduzir maior consumo de bateria e gasto de recursos aéreos. Além disso os túneis que são estabelecidos entre o HA e o MN aumentam as restrições de largura de banda no enlace sem fio e a carga de processamento no nó móvel. Ao acrescentar a soluções deste tipo escoamento de dados, a complexidade dos dispositivos de usuário aumentaria em grande medida. Agora os MN devem implementar algoritmos para decidir quando e que fluxos são trocados entre as interfaces.

Por outro lado, em abordagens que utilizam gerenciamento de mobilidade baseada na rede, uma entidade da mesma faz o relativo à mobilidade no nome do MN. Assim, o MN não requer participar em qualquer sinalização relativa à mobilidade, com o que a implantação de tais soluções é mais simples e os custos são menores.

Para os dispositivos de usuário suportar escoamento de dados precisam algumas modificações, com o fim, por exemplo, de tratar fluxos simultâneos de uma mesma conexão, por duas interfaces. Esses requisitos, sem dúvida, podem aumentar sua complexidade. Também, trocas de mensagens de sinalização no acesso sem fio podem piorar o congestionamento na rede de acesso, problema que se pretende evitar. Por isso, pode ser preferível deixar do lado da rede as operações relativas à mobilidade e as decisões de escoamento.

Outros problemas muito discutidos são os apresentados pelo gerenciamento de mobilidade centralizado e hierárquica como:

- *Encaminhamento sub-ótimo:* o tráfego sempre atravessa uma entidade central da rede, o que leva a caminhos que são, em geral, mais longos do que aqueles diretos entre o nó móvel e seus pares. Isto adiciona demoras desnecessárias e desperdício de recursos do operador.
- *Escalabilidade:* a entidade central da rede precisa ter capacidade de processamento e encaminhamento suficientes, para ser capaz de lidar com o tráfego de todos os nós, simultaneamente.
- *Confiabilidade:* soluções centralizadas são mais propensas a problemas de confiabilidade, pelo fato de que, uma entidade central é um ponto único de falha.

Como alternativa para dar solução a ditos problemas surgiu o gerenciamento de mobilidade distribuído. As soluções separam o plano de dados do de controle, deixando na borda da rede as funções de reenvio dos dados, com o fim de obter caminhos ótimos para o tráfego de dados. Logo, a maioria das proposta na literatura, até onde temos conhecimento, tem acrescentado suporte para escoamento de dados a soluções parcialmente distribuídas. Com isto o problema de encaminhamento sub-ótimo fica resolvido, mas ainda no plano de controle, seguem os problemas de escalabilidade e ponto único de falha e ataque, diminuindo o desempenho da rede.

Então as novas pesquisas para a Internet do futuro precisam prover escoamento de dados em soluções de gerenciamento de mobilidade totalmente distribuído ou hibrido. Com isto são unificadas as novas tecnologia, para oferecer a melhor QoS, à maior quantidade de usuários. Atendendo dita necessidade, no próximo Capítulo será descrita uma solução para o gerenciamento de mobilidade totalmente distribuída, baseada na rede e assistida pelo usuário, que permite movimentar seletivamente fluxos IP.

Também é importante considerar nas soluções existentes e nas próximas, desafios como: mecanismos de segurança, implementação e avaliação de mecanismos de escoamento dinâmico.

4 - ARQUITETURA SIFDMM (SEAMLESS IP FLOW AND DISTRIBUTED MOBILITY MANAGEMENT)

Neste capítulo são apresentados o desenho, as componentes e o funcionamento da arquitetura proposta "SIFDMM - *Seamless IP Flow and Distributed Mobility Management*", para suporte à mobilidade de fluxos IP e ao gerenciamento distribuído de mobilidade.

O presente capítulo inicia-se com a definição de alguns termos utilizados, seguida da apresentação geral da arquitetura, das estruturas de dados consideradas, das mensagens, dos cenários de emprego da arquitetura e dos estados correspondentes. Uma discussão sobre o atendimento a requisitos de DMM é realizada. Por fim, nas considerações finais, uma síntese das características da arquitetura SIFDMM é apresentada.

4.1 - CONCEITOS BÁSICOS

A seguir é feita a definição dos principais termos utilizados.

Número de identificação do *binding* (BID – *Binding Identification number*) [61]: número atribuído univocamente para distinguir entre vários *bindings* registrados para o MN, com o fim de permitir-lhe registrar múltiplas entradas na *Binding Cache.*

Seletor de Tráfego (TS – *Traffic Selector*): Descrito por um o mais parâmetros que podem ser comparados com os campos nos cabeçalhos dos pacotes. Tem o propósito de classificar um pacote [62] ou fluxo. Neste trabalho será utilizada uma 5-tupla com os seguintes parâmetros: endereço IP fonte, endereço IP destino, porta fonte, porta destino e protocolo de camada de transporte.

Binding de fluxo (*Flow Binding* - FB): consiste de um seletor de tráfego, e um ou mais identificadores de *Binding* (BID). Os pacotes IP de um o mais fluxos que correspondem com o seletor de tráfego associado com o FB, são reenviados aos BIDs associados com o mesmo FB [63].

Identificador de fluxo (FID - *Flow Identifier*): o FID identifica de forma unívoca, a um FB associada com o MN.

4.2 - ARQUITETURA

A arquitetura SIFDMM proposta para redes LTE e WiFi é representada na Figura 4.1. As principais componentes são: as entidades de rede *Flow and Mobility Anchor* (FMA) e os

dispositivos de usuário, *Mobile Node* (MN) e *Correspondent Node* (CN). SIFDMM é baseada no protocolo para a gestão de mobilidade PMIPv6 [3], sendo uma arquitetura baseada na rede, mas assistida pelo usuário. Esta última caraterística permite o tratamento do plano de controle distribuído. Logo o gerenciamento de mobilidade é totalmente distribuído.

Na Figura 4.1 as FMA podem ter tecnologias de acesso LTE ou WiFi. Estas são as encarregadas de conectar os MN com os CN. O MN ao fazer o registro em uma FMA, informará o último PCoA (*Proxy Care of Address* – IP da FMA) associado a cada uma de suas interfaces. Com esta informação as entidades de rede terão conhecimento das FMA que tiveram ou têm ligação com o MN. Neste sentido poderão estabelecer túneis para fazer escoamento de tráfego ou *handover*, podendo ser implementados algoritmos que determinem dinamicamente a movimentação dos fluxos IP, total ou seletivamente. Logo o tráfego pode passar de uma tecnologia de acesso para outra, baseando-se por exemplo: nas condições de rede, perfis de usuário, carga na rede, políticas de prioridades, entre outras.

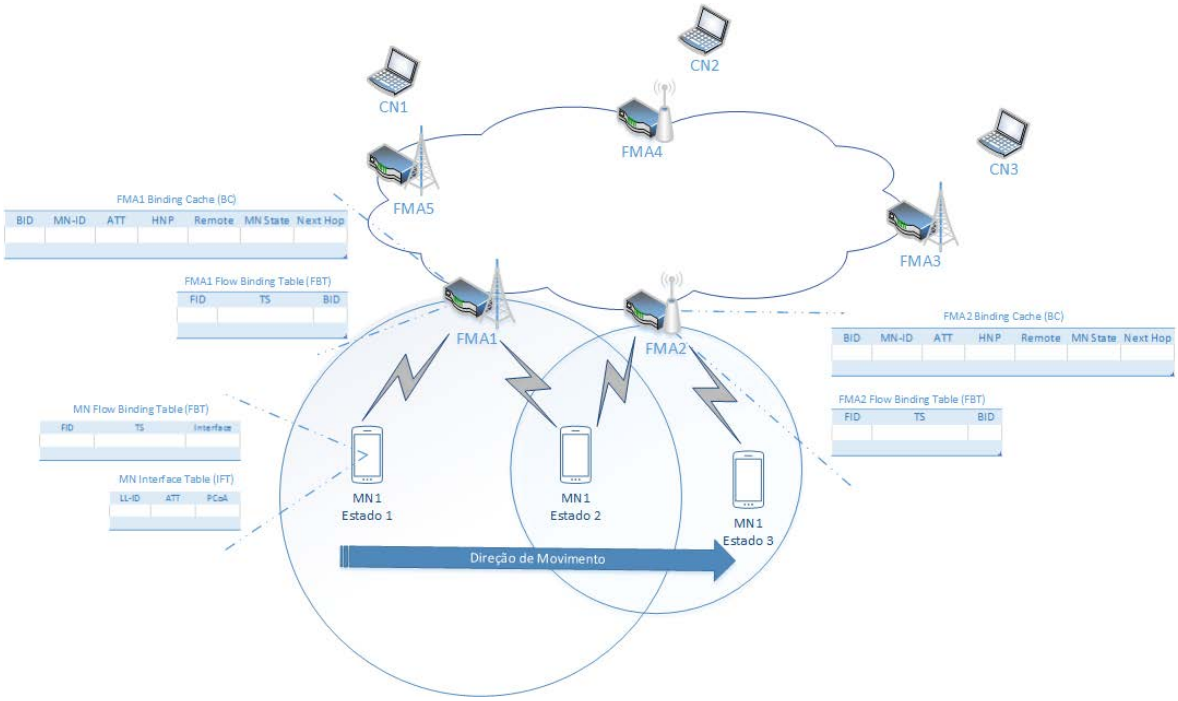

Figura 4.1: Arquitetura SIFDMM para redes LTE e WiFi.

1. *Flow and Mobility Anchor (FMA)*: serão *anchor point* e a sua vez *router* de acesso, tendo funções similares as entidades LMA e MAG do PMIPv6 [3]. São as responsáveis de atribuir o prefixo de rede (HNP – *Home Network Prefix*) ao MN, para que possa configurar o endereço IP e prover conexão à Internet. Também detectam o movimento do MN entre as

diferentes redes de acesso e mantem a continuidade das sessões abertas. Baseadas em estruturas de dados locais e na informação das condições da rede, devem implementar algoritmos para a decisão de *handover* ou escoamento de tráfego (i.e. para balanceamento de carga na rede ou garantir parâmetros de QoS). Uma vez detectada a necessidade de movimentar algum fluxo, são as encarregadas de fazer a troca de mensagens de sinalização, com o propósito de criar túneis que permitam ao MN manter suas sessões ativas.

Quando um MN sai da rede de acesso de uma FMA, ela deve armazenar os pacotes recebidos para esse móvel por um tempo. Nesse período deve chegar uma mensagem de outra FMA indicando que o MN agora é atendido por ela, caso contrário os pacotes serão descartados. As mensagens de controle utilizados, são similares ao PBU e PBA de PMIPv6, incluindo opções de mobilidade de fluxo, e identificação do estado do MN, como será descrito posteriormente.

2. *Mobile Node (MN)*: são dispositivos de usuário multimodo, que podem trabalhar nos modos "*weak host*" [64-65] ou "*logical interface*" [66]. Eles têm a capacidade de processar os pacotes que estão chegando por uma interface, com um endereço IP diferente ao associado a ela. Também armazenam informação dos fluxo, com o fim de envía-os pela interface correta, mas não tomam decisões de movimentação de fluxo. Eles ao receber um fluxo por uma interface diferente, atualizam sua estrutura de dados e seguem a decisão da rede. Também devem guardar o endereço IP (PCoA) da última FMA ao que estiveram ligados, por cada uma das interfaces, com o fim de assistir as entidades da rede no processo de *handover* ou escoamento.

3. *Correspondent Node (CN):* são os dispositivos extremos da comunicação estabelecida com os MN. Eles podem ter as mesmas características e funciones que um MN, ou podem ser qualquer dispositivo de usuário final.

4.3 - ESTRUTURA DE DADOS

Três estruturas de dados são definidas para suportar SIFDMM, como descrito a continuação: *Binding Cache* (BC), *Flow Binding Table* (FBT) e *Interface Table* (IFT).

1. *Binding Cache* (BC)*:* está contida nos FMA e tem informação sobre os MN que estão na sua rede de acesso, ou que estiveram, mas ainda existem sessões abertas, através dela. Cada entrada na BC contém os seguintes campos:

- BID: identificador único do *binding*.

- MN – ID: identificador único do MN.
- ATT (*Access Technology Type*): tecnologia de acesso.
- HNP: prefixo de rede atribuído ao MN.
- Remote:

  True: indica que o *binding* é feito com outra FMA.

  False: indica que o *binding* é feito com a própria FMA.

- MN State:

  1: indica que o MN está na rede de acesso do FMA (possibilidade de mobilidade de fluxo).

  2: indica que o MN não está conectado com a rede de acesso do FMA, mas na BC existe uma/várias entradas para o MN (ainda tem sessões abertas com o HNP atribuído, indicando a necessidade de um *handover*).

  3: indica que a BC do FMA não tem nenhuma entrada para o MN.

- *Next Hop*: Indica o próximo salto.

2.*Flow Binding Table* (FBT): está contida nos FMA e no MN e armazena informação do identificador, seletor de tráfego e interface (no MN) ou BID (na FMA) associadas a cada fluxo. Cada entrada na FBT contém os seguintes campos:

- FID: Identificador único do fluxo.
- TS: seletor de tráfego, 5-tupla <*ip fonte, ip destino, porta fonte, porta destino, protocolo*>.
- BID: identificador único do *binding* no caso da FMA.
- Interface: identificador único da interface no caso do MN.

3. *Interface Table* (IFT): está contida no MN e guarda informação do identificador, tecnologia de acesso e endereço IP (PCoA) associada a cada uma de suas interfaces. Cada entrada na IFT contém os seguintes campos:

- LL-ID (*Link Layer Identifier*): identificador único da interface.
- ATT: tecnologia de acesso.
- PCoA: endereço IP da entidade de rede associada à interface.

4.4 - MENSAGENS

A mensagem RS (*Router Solicitation*) têm opções de mobilidade, que vão conter informação sobre a tecnologia de acesso e o PCoA de suas interfaces.

A mensagem RA (*Router Advertisement*) terá um campo para informar o PCoA da FMA com que está estabelecendo conexão.

Nas mensagens PBU (*Proxy Binding Update*) e PBA (*Proxy Binding Acknowledgement*) para registro, é preciso acrescentar o campo MN State (no mínimo de 2bits) para indicar o estado do MN na FMA. As mensagens também incluem o campo PCoA, que no caso de ser diferente de zero, indica o endereço da FMA que tem uma ligação com o MN (utilizado para otimizar rota).

As mensagens *Flow PBU* e *Flow PBA* terão também habilitadas opções de fluxo, para indicar o seletor de tráfego e o status na resposta, como indicado na RFC 6089 [63].

As mensagens *Handover PBU* e *Handover PBA* fazem referência às mesmas mensagens PBU e PBA do registro, só que é atribuído o valor correspondente ao tipo de *handover*, no campo *Handoff Indicator*.

4.5 - CENÁRIO 1

No primeiro cenário, mostrado na Figura 4.1, temos duas FMA com diferentes tecnologias de acesso, FMA1 LTE e FMA2 WiFi. O MN neste cenário vai passar por três estados:

1. Primeiro Estado: O MN estabelece uma conexão com o FMA1 (LTE), sendo seu primeiro registro no domínio.
2. Segundo Estado: o MN se movimenta para uma área de sobreposição das cobertura das FMAs (FMA1 e FMA2) e estabelece conexão com a FMA2.
3. Terceiro Estado: o MN sai da zona de sobreposição, e só tem cobertura da FMA2 (WiFi).

4.5.1 - Primeiro Estado (E1)

Quando o MN entra no domínio e faz seu primeiro registro são precisas as seguintes ações, como mostra a Figura 4.2:

1. O MN encaminha ao FMA1 a mensagem RS que não terá opções de mobilidade, pois as interfaces não têm ainda valores de PCoA associados.
2. Ao FMA1 receber a mensagem RS, faz uma nova entrada na *Binding Cache*, atribuindo um HNP1 ao MN, e colocando os valores dos campos: Remote em false e MN State em 1.
3. O FMA1 forma a mensagem RA, e informa ao MN seu PCoA1 no campo PCoA.

4. O MN ao receber a mensagem RA associa o PCoA1 com essa interface, e configura seu endereço IP.
5. Estabelece comunicação com o CN1 (Flow A) e com o CN2 (Flow B).

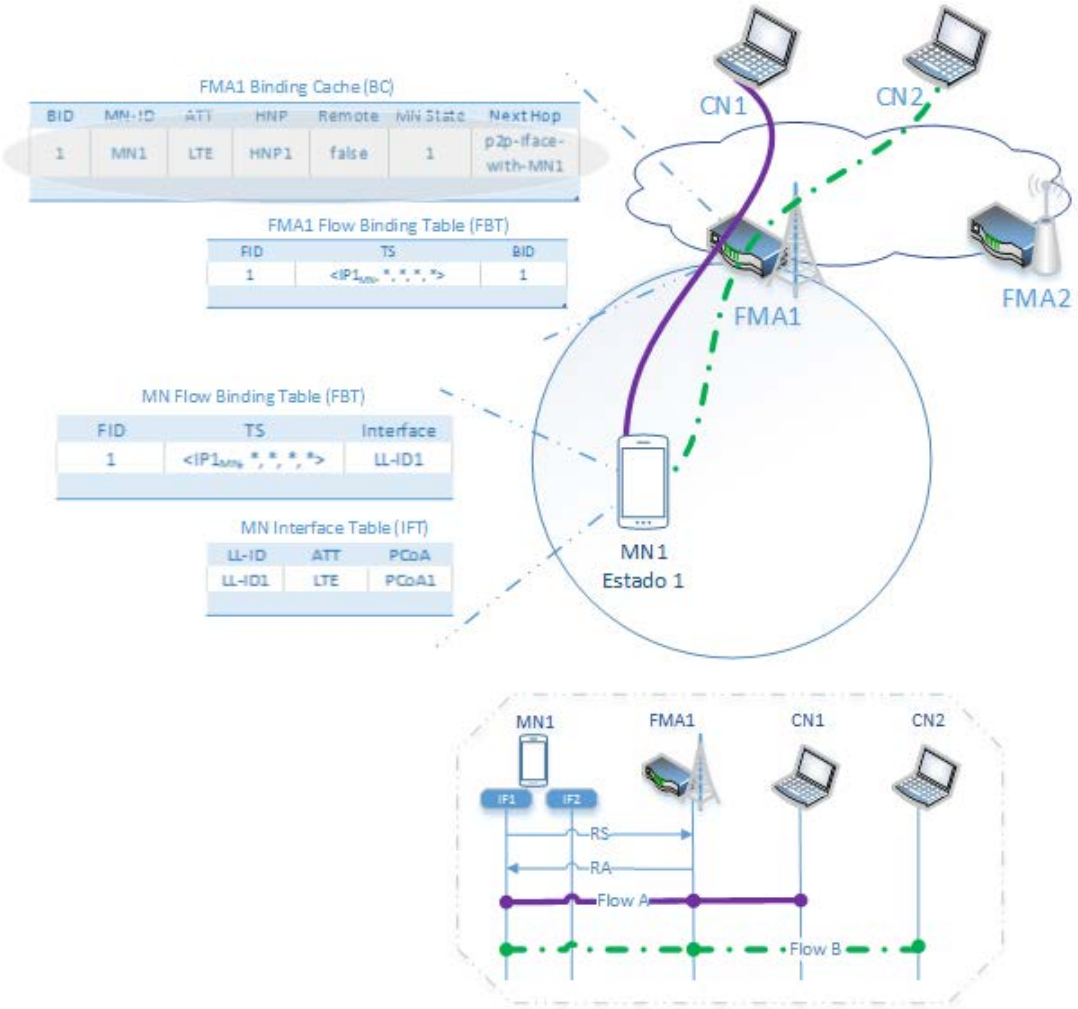

Figura 4.2: Primeiro Estado - MN estabelece uma conexão com o FMA1 (LTE).

4.5.2 - Segundo Estado (E2)

No segundo estado o MN1 chega a uma aérea de sobreposição da cobertura das duas FMAs. Então as FMAs terão a possibilidade de movimentar todos ou seletivos fluxos IP. Em outras palavras, poderão fazer escoamento de tráfego (i.e. para balanceamento de carga) o que pode melhorar a QoE percebida pelos usuários. Neste estado para estabelecer conexão com a FMA2 é preciso a seguinte troca de mensagens, como mostra a Figura 4.3:

1. O MN encaminha ao FMA2 a mensagem RS. As opções de mobilidade da mesma, tem a ATT e o PCoA associados a sua outra interface (interface LTE onde o MN tem uma conexão com a FMA1)
2. Ao FMA2 receber a mensagem RS, faz uma nova entrada na *Binding Cache*, atribuindo um HNP2 ao MN1, e colocando os valores dos campos: Remote em false e MN State em 1.
3. O FMA2 conforma a mensagem RA, e informa ao MN seu PCoA2 no campo PCoA.
4. O MN ao receber a mensagem RA associa o PCoA2 com essa interface, e configura seu endereço IP.
5. O FMA2 envia uma mensagem PBU ao FMA1 (obtendo o PCoA1 da mensagem RS encaminhada pelo MN), com os campos MN State em 1 e HI em 1 (*Attachment over a new interface*), indicando que ele também tem uma conexão estabelecida com o MN. Esta informação pode ser logo utilizada caso que precise movimentar fluxo.
6. O FMA1 ao receber o PBU, adiciona uma nova entrada na BC, e envia uma mensagem PBA, com os campos MN State em 1 e HI em 1 (*Attachment over a new interface*), indicando que ele tem um *binding* com o MN.
7. O FMA2 ao receber o PBA, adiciona uma nova entrada na BC, tendo agora conhecimento da possibilidade de um escoamento de tráfego ou *handover* com a FMA1, e vice-versa.

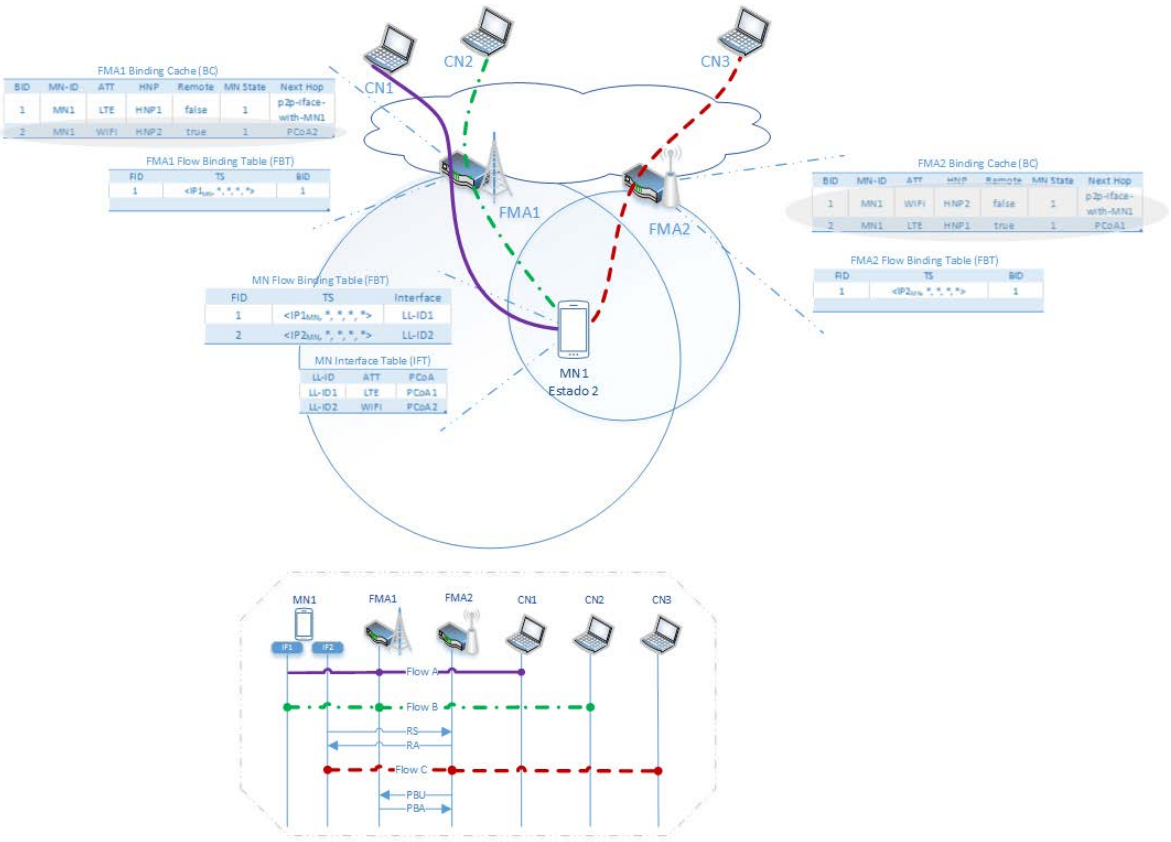

Figura 4.3: Segundo Estado - MN está em uma área de sobreposição das cobertura das FAMs.

Quando o MN está neste estado podem ocorrer dois eventos:

1. O MN estabelece comunicação com um CN3, e começa um fluxo de dados (Flow C) através da FM2. Neste evento não é preciso nenhuma sinalização de controle adicional, já o *binding* está estabelecido, e o MN pode enviar e receber pacotes (Figura 4.3).

2. Uma das duas FMA, precisa fazer escoamento de tráfego, movimentando determinados fluxos à outra interface do MN. Neste evento é preciso uma troca de mensagens de controle entre as FMA. Para explicar: assumimos que a FMA1 precisa movimentar o fluxo B para a FMA2, como mostra a Figura 4.4, então terão lugar as seguintes ações:

1. O FMA1 encaminha uma mensagens *Flow* PBU, para solicitar a reserva de recursos e o estabelecimento de um túnel, com o fim de movimentar o fluxo B.
2. O FMA2, se tem recursos, inclui na FC as opções de fluxo e associa com o BID existente na sua BC para a FMA1 (PCoA1), e envia uma mensagem *Flow* PBA aceitando a solicitação.
3. Queda estabelecido o túnel bidirecional e começa o escoamento do fluxo B.

4. Quando o MN1 recebe pacotes encaminhados ao IP1 (endereço IP configurado com o HNP1 atribuído pela interface LTE) pela interface WiFi, processa o pacote e atualiza sua *Flow Binding Table*. Com a atualização, o MN1 pode enviar os pacotes, que coincidam com o TS pela interface correta.

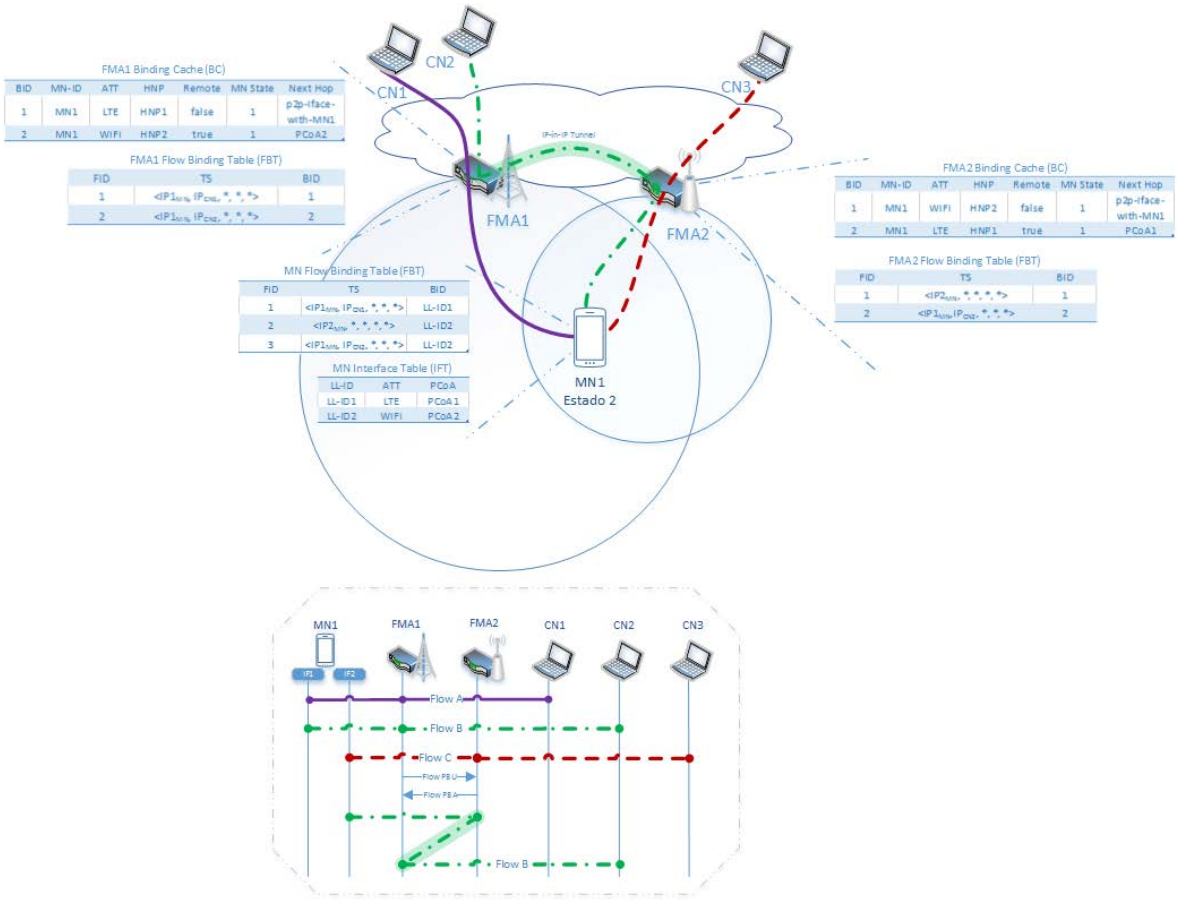

Figura 4.4: Escoamento de fluxo B da rede de acesso LTE para a WiFi.

4.5.3 - Terceiro Estado (E3)

Quando o MN sai da área de cobertura da FMA1 é necessário realizar um *handover* (movimentar todos os fluxo do MN1 da FMA1 para a FMA2), como mostra a Figura 4.5. Quando a FMA1 detecta que o MN1 não está na sua rede de acesso:

1. Atualiza a entrada na BC onde Remote é false (entrada do *binding* local), colocando o campo *MN State* em 2. Isto indica que já o MN1 não está na sua rede de acesso, mas ainda tem sessões abertas, e precisa um *handover*.
2. Encaminha uma mensagem *Handover* PBU à FMA2, com o valor de campo *MN State* em 2. Com isto faz uma solicitação de reserva de recursos, para fazer *handover*.

3. Na FMA2, se é possível atender a solicitação, modifica a entrada relacionada com o FMA1 na BC, e envia um *Handover* PBA aceitando o *handoff*.
4. Agora todos os fluxos serão enviados pelo túnel, com o fim de manter continuidade nas sessões em andamento.

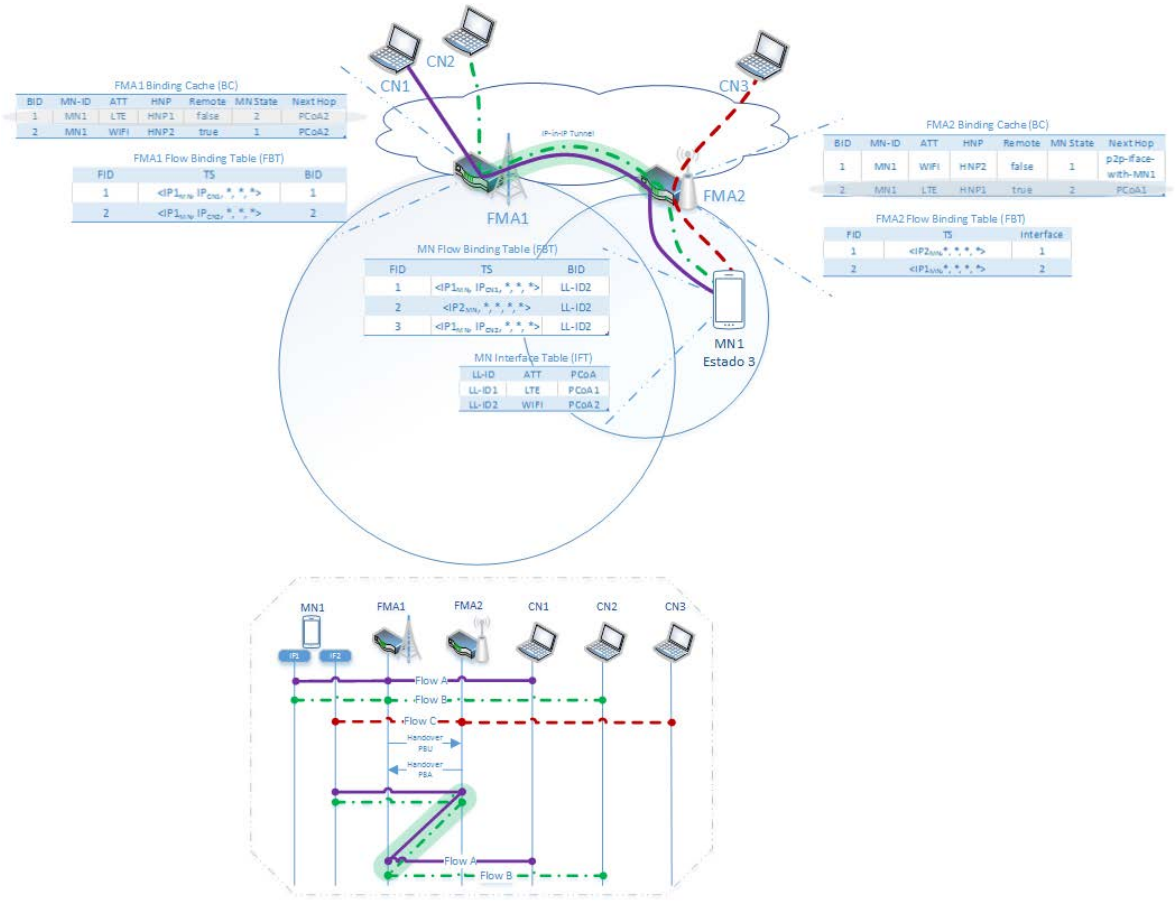

Figura 4.5: *Handover* da rede LTE para a WiFi.

4.6 - CENÁRIO 2

Neste cenário temos três FMA, duas com redes de acesso LTE e uma com WiFi como mostra a Figura 4.6. Agora o MN ira percorrer cinco estados, onde os três primeiros são iguais aos do cenário um, e os outros serão explicados a continuação:

4. Quarto Estado (E4): o MN1 se movimenta para uma área de sobreposição das coberturas da FMA2 e FMA3, e estabelece uma comunicação com o CN4 através da FMA3.
5. Quinto Estado (E5): o MN1 fica só com a cobertura da FMA3, mas ainda tem sessões abertas através das FMA1 e FMA2.

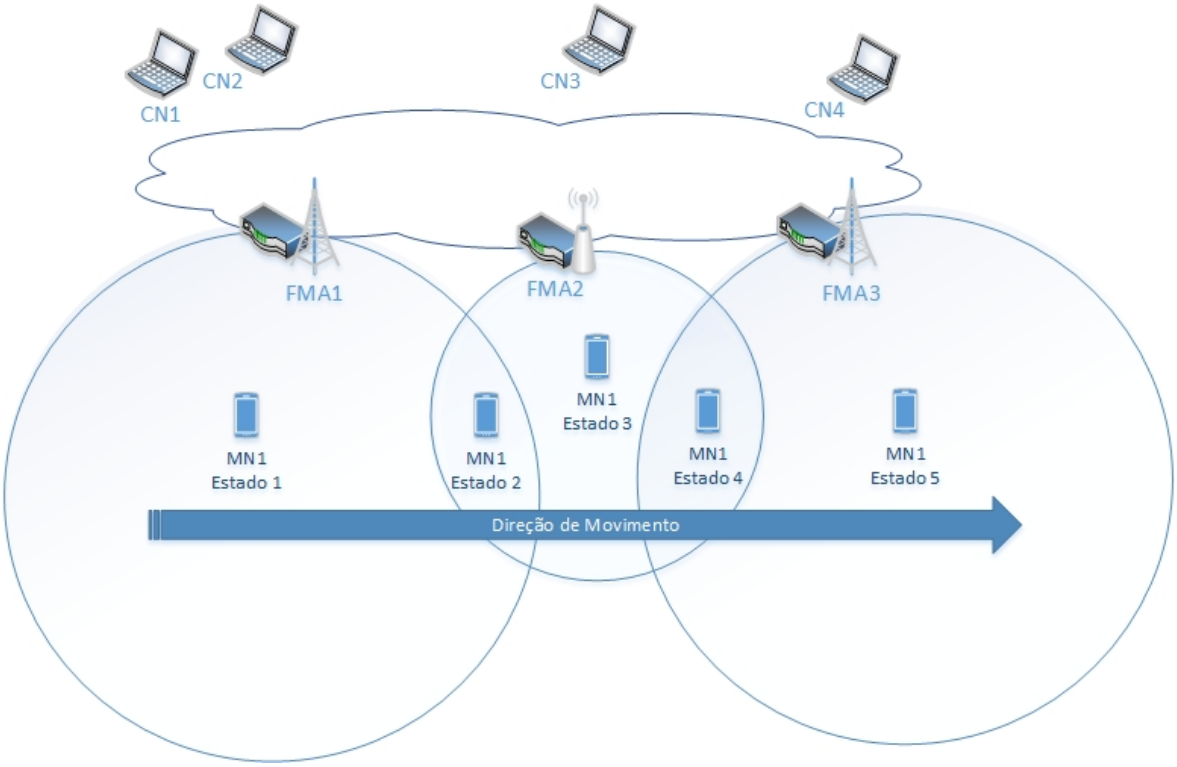

Figura 4.6: Exemplo com três FMAs (2 LTE - 1 WiFi).

4.6.1 - Quarto Estado (E4)

Quando o MN chega à área de sobreposição das FMA2 e FMA3, para estabelecer o *binding* acontecem as seguintes ações (Figura 4.7):

1. O MN1 envia uma mensagem RS, que contém os endereços associados a cada uma de sus interfaces (PCoA1 e PCoA2).
2. Quando a FMA3 recebe a mensagem cria uma entrada na sua BC, atribui um HNP3 ao o MN e encaminha o RA com seu endereço PCoA3.
3. A FMA3 encaminha uma mensagem PBU para atualizar as FMA1 e FMA2, informando que agora ele está atendendo ao MN1, onde o valor do campo *MN State* será 1. Com isto as outras entidades tem conhecimento que o MN1 está na sua rede de acesso, existindo a possibilidade de fazer escoamento ou *handover*.
4. A FMA2 (WiFi e ainda atendendo ao MN) cria uma nova entrada na sua BC e encaminha em resposta o PBA.
5. A FMA1 (LTE que não tem ao MN1 na sua rede de acesso, mas ainda tem sessões em andamento) cria uma nova entrada na BC e encaminha em resposta o PBA. Dependendo das políticas de *handover* e escoamento poderá: trocar mensagens de

controle com o FMA3 com o propósito de movimentar alguns ou todos os fluxos para a LTE, ou deixar todo pelo túnel com o FMA2.

6. O MN ao receber o RA atualiza sua *Interface Table.*
7. O MN estabelece comunicação com o CN4 (*Flow* D).

Neste estado as FMA2 e FMA3 podem fazer escoamento de tráfego, com o mesmo mecanismo explicado no estado dois, do cenário 1. Também FMA1 pode estabelecer dois túneis (um com FMA2 e outro com FMA3) com o fim de fazer escoamento de fluxos ou *handoff*, segundo as políticas da rede.

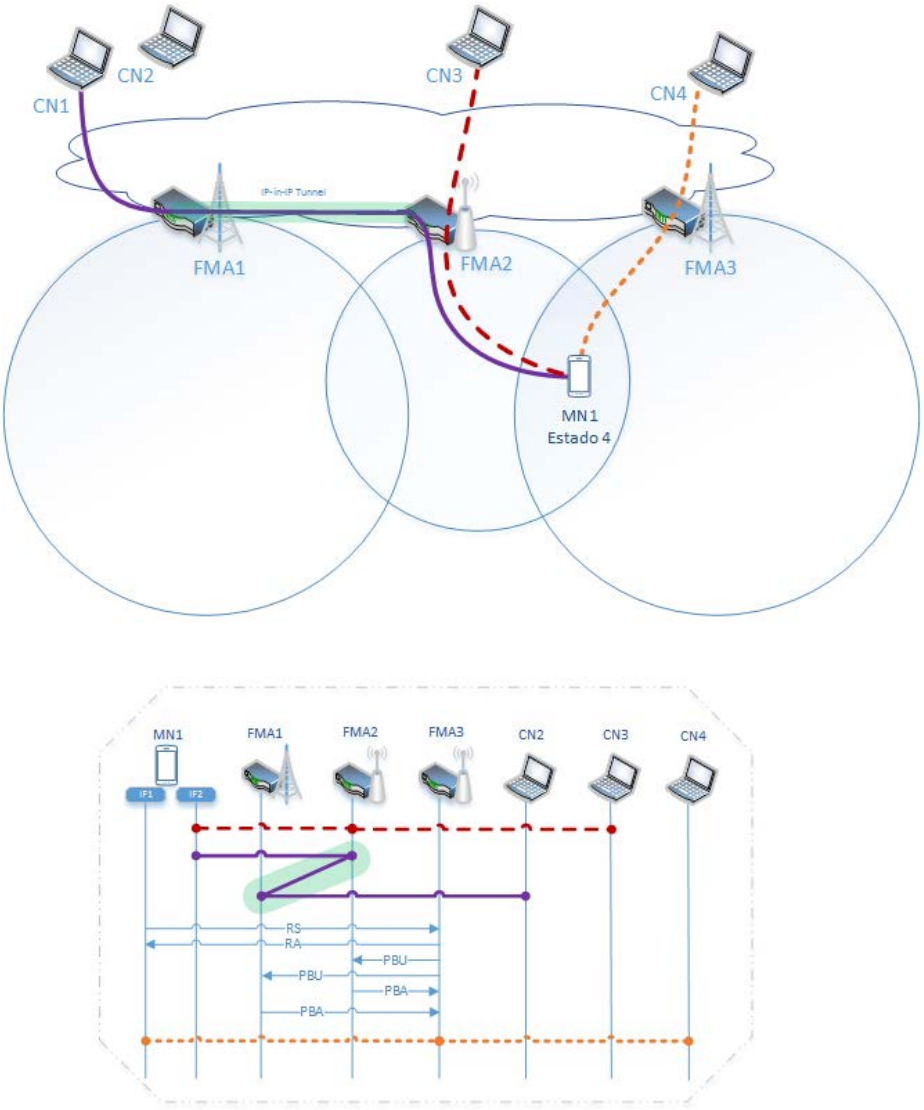

Figura 4.7: Quarto Estado - Fluxos do MN na cobertura das FMA2 e FMA3.

4.6.2 - Quinto Estado (E5)

Neste estado o MN só está na rede de acesso da FMA3 (LTE), então será preciso fazer *handover* das seções abertas através das FMA1 e FMA2. A continuação explicaremos as ações que ocorrem em cada uma das entidades de rede, que tem fluxos, com o endereço do MN1 como destino, através delas (Figura 4.8).

- FMA2 (com rede de acesso WiFi):

Quando a FMA2 detecta que o MN2 não está na sua rede de acesso, como mostra a Figura 4.8:

- Atualiza a entrada do *binding* local (*Remote == false*), colocando o valor 2, no campo *MN State.*
- Espera a chegada da mensagem PBU da nova FMA que agora atende ao MN (FMA3). Neste período, todos os pacotes que chegam ao FMA2 com o endereço do MN1 como destino, são armazenados no buffer.
- Uma vez, recebida a mensagem PBU do FMA3, a FMA2 verifica as entradas remotas para o MN1 onde o valor do campo *MN State* é 2, e envia um PBU às FMA associadas. Essa mensagem de controle coloca o PCoA3 no campo PCoA, para indicar que deve cerrar o túnel com a FMA2 e estabelecê-lo com a FMA3. Note-se que isto é preciso para otimizar a rota até o FMA1, e não criar vários túneis IP-in-IP.

Neste cenário a informação poderá parecer redundante, mas em cenários com maior quantidade de redes LTE, por exemplo, será necessário para otimizar rotas e não ter perdida de pacotes. É valido lembrar que o MN, só informará o PCoA da última FMA ligada a essa interface, mas ainda poderá ter fluxos através de outras FMA com rede de acesso LTE. Então o PCoA dessas entidades de rede o terá a última FMA, que será a encarregada de informar com qual FMA deve agora criar um túnel, evitando assim a existência de vários túneis IP-in-IP.

- Ao FMA2 receber o PBA da FMA1, eliminará a entrada remota na sua BC, associada com ela.

• FMA1 (LTE):

Neste caso quando a FMA1 recebe o PBU da FMA2, já ela tem uma entrada remota, indicando que o FMA3 está atendendo ao MN (criada no processo de registro do MN com a FMA3). Então trocará mensagens de controle com FMA3 para estabelecer um túnel e manter

as sessões ativas. No caso de não ter ainda nenhuma entrada associada com o FMA3, cria uma nova entrada e tenta estabelecer o túnel.

- FMA3 (LTE):

Ao receber os PBU das FMA1 e FMA2, se poder atender as solicitações, cria as entradas na suas estruturas de dados e os túneis ficam estabelecidos.

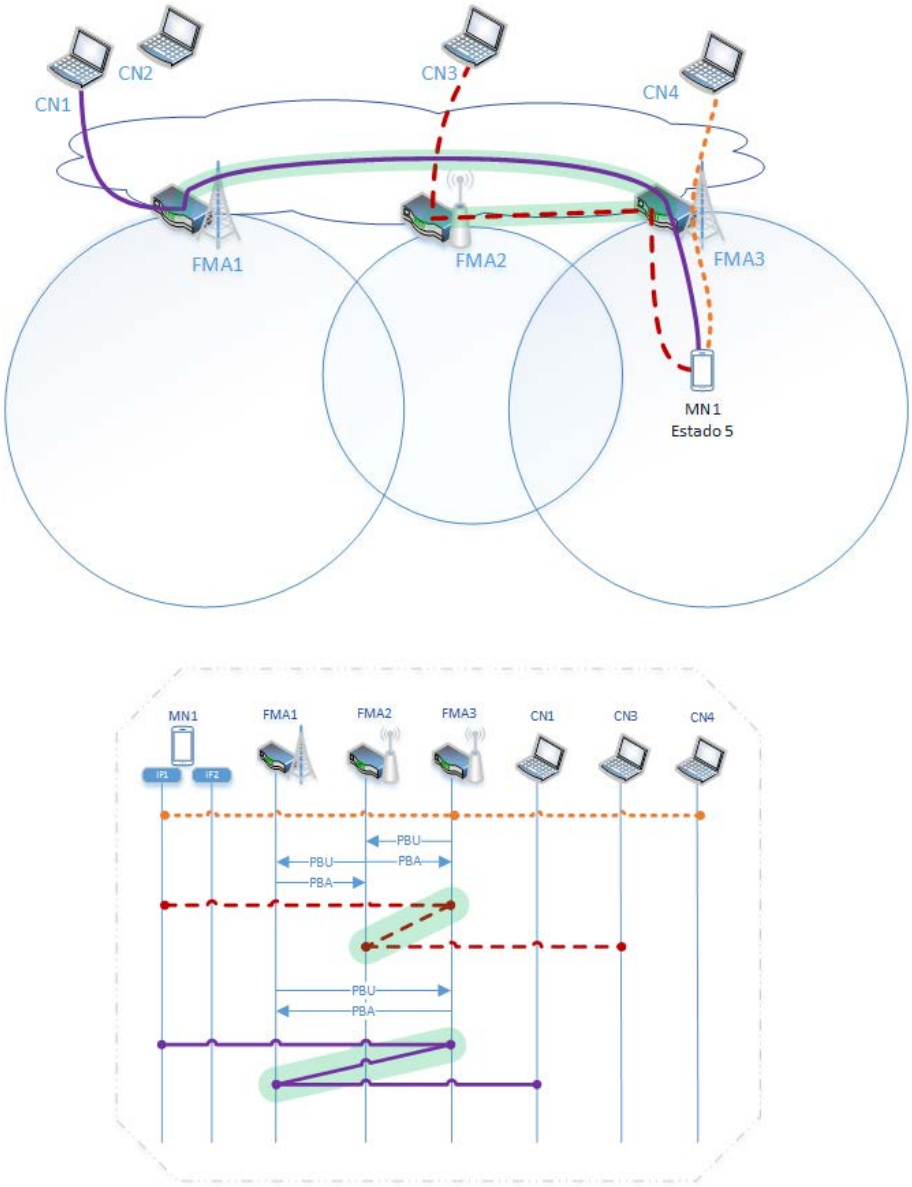

Figura 4.8: *Handover* para FMA3.

Com todo o descrito até agora a arquitetura é capaz de prover gerenciamento de mobilidade IP, sem que os usuários experimentem interrupções nas comunicações. Também é possível

movimentar seletivamente fluxos IP, o que aumenta a vazão, podendo utilizar duas redes de acesso simultaneamente, e permite balanceamento de carga nas redes de acesso. Na próxima Seção será discutido o atendimento da solução com os requisitos de DMM.

4.7 - ATENDIMENTO A REQUISITOS DE DMM

Na arquitetura proposta, as funções de mobilidade (i.e. MAG e LMA) são co-localizadas nos routers de acesso similar às soluções [10] e [16]. As entidades chamadas *Distributed Anchor Router* (DAR) em [10], *Mobility capable Access Router* (MAR) em [16], e FMA (*Flow and Mobility Anchor*) em nossa solução, são similares. Elas representam também o primeiro router IP ao qual o MN vai se ligar, suportando funcionalidades de gerenciamento da localização do MN e *traffic anchoring*.

Em soluções baseadas na rede é considerado que o MN não participa na sinalização relativa à mobilidade, mas em arquiteturas de mobilidade distribuída pode ser complicado porque: 1) o MN pode ser servido por mais de um LMA e 2) o *mobility anchor* depende do ponto de *attachment* onde a comunicação IP tem sido iniciada [16]. Então temos os seguintes caminhos para direcionar este problema:

1. Apoiar-se em uma entidade centralizada no plano de controle: quando o MN estabelece uma conexão com o novo *router* de acesso, a função MAG recupera as sessões em andamento do MN a partir da base de dados centralizada. A base de dados deverá ser atualizada cada vez que um novo prefixo é atribuído ao MN, e também quando o prefixo é liberado. Dita abordagem é utilizada nas soluções descritas em [10],[16], mas em [16] não é tratada a interação das entidades MAR com a base de dados. A abordagem pode ter problemas de escalabilidade e confiabilidade, representando a base de dados um ponto único de falhas e ataque.
2. *Handover* de camada de enlace (Layer-2 *handover*) suportado através da especificação IEEE 802.21 (MIH- *Media Independent Handover*). Esta abordagem, permite de acordo com a filosofia *make-before-break*, fazer o gerenciamento do *handover* com uma infraestrutura do plano de controle dedicada, onde o movimento é preparado, executado e completado de um jeito controlado e assistido, mas precisa da entidade centralizada MIIS (*Media Independent Information Services*), que pode trazer os mesmo problemas que a abordagem 1.

3. *Multicasting* das mensagens PBU enviadas pelo *mobility anchor* que serve ao MN, a um grupo formado por todos os *mobility anchor* do domínio. Com isto pode ter informação dos fluxos ativos do MN em prévios *mobility anchor*. Esta abordagem é uma das propostas em [10], mas desafortunadamente poderia não prover um bom desempenho em termos de latência de *handover* e adiciona sinalização desnecessária na rede.
4. Envolver o MN no processo de gerenciamento de mobilidade:
   - Durante o processo de *attachment* com o novo *router* de acesso, o MN deve comunicar suas sessões em andamento ao MAG (i.e. uma lista dos HNP associada aos *mobility anchors*). Essa informação pode ser provida através de uma opção da mensagem RS. Dita solução é sugerida em [16], mas não desenvolvida, pois pode quebrar o conceito básico de abordagens do gerenciamento de mobilidade baseada na rede, onde o MN deve permanecer agnóstico do suporte de mobilidade.
   - Durante o processo de *attachment* com o novo *router* de acesso, o MN deve comunicar, o endereço IP (PCoA - *Proxy Care of Address*) do último *router* de acesso associado com cada uma de suas interfaces. A informação é provida por uma opção na mensagens RS. Se for a sua primeira conexão no domínio, a opção pode ser zero é ignorada no MAG, logo na mensagem RA o MN receberá o endereço IP e associará com a interface, para quando se movimentar oferecer a informação. Se o MN ou o *router* de acesso não tem suporte para gerenciamento de mobilidade, ignorarão a opção na mensagem RS. Dita opção pode ser considerada híbrida ou baseada na rede mas com assistência do MN. As mudanças no MN são mínimas, só é preciso armazenar um endereço IP associado a cada uma de suas interfaces, como um identificador, e logo prover essa informação à rede no caso se movimentar.

   Se o prévio *mobility anchor* estava servindo como estrangeiro (MAG) para alguns fluxos do MN, ele é o responsável por informar as entidades que fazem do LMA (onde os fluxos iniciaram), para que estabeleçam túneis com o *mobility anchor* que serve ao MN. Assim são evitados vários túneis IP-in-IP. Dita abordagem evita os problemas de confiabilidade, escalabilidade e sinalização desnecessária das abordagens descritas antes. Ela apresenta o problema de precisar da participação do MN, mas isso poderia ter menor influência no desempenho geral do sistema, com modificações mínimas no dispositivo do usuário. Esta é a abordagem utilizada na solução SIFDMM.

Na RFC 7333 [4] são definidos os requisitos para gerenciamento de mobilidade distribuído na camada de rede. Na continuação, é discutido como a arquitetura SIFDMM atende ditos requisitos:

REQ1: Na solução SIFDMM o roteamento IPv6 regular é aplicado quando uma comunicação IP é iniciada. Por exemplo, se o MN estando conectado ao FMA1, inicia uma comunicação, o tráfego será roteado através do FAM1 sem requerer qualquer operação de mobilidade especifica. Quando o MN se move e estabelece conexão com o FMA2, o tráfego *anchored* no FMA1, atravessa o túnel entre FMA1 e FMA2, no caso de um *handover* ou escoamento de dados. Nesse caso o FMA1 joga o rol do *mobility anchor* (LMA), para seções iniciadas com o HNP1 atribuído por ele, e o FMA2 joga o rol de MAG dessas seções.

Se uma comunicação inicia com o HNP2 atribuído pelo FMA2, o roteamento IPv6 regular será aplicado a ditos fluxos, sempre que eles não estejam no processo de escoamento. Os túneis sempre serão estabelecidos entre o FMA donde iniciou o fluxo e o FMA que está servindo ao MN, garantido otimização das rotas.

REQ2: As comunicações são iniciadas sem requerer *mobility anchor* e túneis. Opções de mobilidade só serão utilizadas em caso de precisar para escoamento de dados ou *handover*. Para escoamento de dados os identificadores de fluxos (sessões) são utilizados de acordo com a RFC 6089 [63].

REQ3: considera o endereço IPv6 como primário, baseado no PMIPv6.

REQ4: baseada no protocolo PMIPv6, e considera outras soluções existentes como [10],[16].

REQ5: Se as entidades de rede não suportam a solução, as opções de mobilidade das mensagens de controle serão ignoradas, e pode ser utilizado qualquer outro protocolo para o gerenciamento de mobilidade. Se o *host* não suporta a solução, uma base de dados no plano de controle ou *multicasting* deverá ser utilizado para oferecer as opções de mobilidade e garantir continuidade das sessões.

4.8 - CONSIDERAÇÕES FINAIS

Neste capitulo, propomos uma nova arquitetura chamada “SIFDMM - *Seamless IP Flow and Distributed Mobility Management*”, baseada no protocolo PMIPv6 (e portanto na rede). Foram analisados dois cenários, um simples com duas FMA e um mais complexo com três

FMAs, que pode ser estendido para trabalhar com vários domínios, onde foi sugerida uma estratégia para optimização da rota, evitando a criação de vários túneis IP-in-IP. Nossa solução suporta mobilidade seletiva de fluxos IP e com isto escoamento de dados (*data offloading*). SIFDMM, pelo fato de ser totalmente distribuída, evita os problemas de soluções centralizadas como ponto único de falha, escalabilidade e confiabilidade. Logo com a arquitetura proposta é possível obter maior vazão e fazer balanceamento de carga, dada a possibilidade dos MN de estar conectados a dois AR no mesmo tempo. Com tudo isto, a arquitetura permite implementar algoritmos dinâmicos de balanceamento de carga, baseados por exemplo nas condições da rede, preferências do usuário ou politicas de QoS, para oferecer aos usuários níveis aceitáveis de QoE.

Outras características da arquitetura SIFDMM são aqui destacadas: i) é assistida pelo terminal móvel, que atua provendo informações de localização relativas a roteadores de acesso recentemente visitados; ii) integra funcionalidades de gerenciamento de mobilidade e de escoamento de tráfego de dados em uma entidade da rede; iii) o escoamento é feito na camada de rede movimentando fluxos IP.

5 - AVALIAÇÃO DE DESEMPENHO

Para avaliar a arquitetura SIFDMM, foi proposta uma modelagem analítica que nos possibilita o estudo da latência no processo de escoamento de dados, apresentando o conceito de latência de *handover* de fluxo. Neste capítulo são descritos os modelos de rede e de atrasos considerando o tempo de enfileiramento e de processamento nos *anchors* da rede. Utilizando a modelagem são avaliadas três arquiteturas baseadas no protocolo PMIPv6.

Também são descritas as simulações realizadas usando o *Network Simulator* versão 3 (NS-3), com o fim de estudar o funcionamento das arquiteturas, quando transmissões de vídeo são feitas através da rede, sujeitas a tráfegos concorrentes dos tipos voz e FTP, tanto para usuários estáticos quanto usuários móveis. Por último, são apresentados os resultados dos parâmetros de QoS e QoE obtidos com a utilização da ferramenta Evalvid.

5.1 - MODELAGEM ANALÍTICA

Nesta Seção se realiza, similar aos trabalhos [67-68], uma avaliação analítica e numérica do desempenho de três propostas de protocolos de mobilidade, que suportam escoamento seletivo de fluxos IP. No trabalho assumimos que será feito um escoamento de fluxo imediatamente depois de estabelecida a conexão com uma interface diferente. Com isto, para a análise definimos a latência de *handover* de fluxo (FHL), como o tempo entre a conexão pela nova interface e a chegada do primeiro pacote pela mesma. A FHL é expressa como segue:

$$FHL(.) = T_{L2} + T_{MD} + T_{LU} + T_p \qquad (5.1)$$

Onde T_{L2} é a latência do *attachment* de camada de enlace, T_{MD} é a latência de detecção de movimento, T_{LU} é a latência de atualização da localização, T_p é a latência de envio do primeiro pacote.

É valido ressaltar que este processo pode se referir tanto à movimentação de fluxos selecionados, quando à movimentação de todos os fluxos, dependendo da decisão da rede.

Na análise serão consideradas: uma solução centralizada, uma parcialmente distribuída e a SIFDMM totalmente distribuída, baseadas em PMIPv6. Selecionamos a solução apresentada em [53] como uma proposta centralizada (PMIPv6), e a [14] como parcialmente distribuída

(PMIPv6_PD). Elas representam o funcionamento geral, permitindo-nos avaliar o desempenho de soluções de seu tipo.

5.1.1 - Modelo de rede

Os três esquemas de mobilidade são baseados na rede. Logo, consideramos uma rede com M *routers* de acesso (AR – MAG, MFR ou FMA, dependendo da arquitetura) e M-1 áreas de superposição de redes de acesso com diferentes tecnologia, como mostra a Figura 5.1. São distribuídos uniformemente N nós móveis, que têm duas interfaces de rede. Assumimos que em média um percentual K de nós móveis entrarão em uma área de superposição simultaneamente. Então $(K * N)/100$ MNs vão fazer handover de fluxo ao mesmo tempo.

Denotamos por $h_{x,y}$ distância média em saltos entre duas entidades de rede x e y, assumindo que é simétrica ($h_{x,y} = h_{y,x}$). Definimos uma escala de rede ξ, como a taxa entre o número de saltos entre dois AR e o número de saltos entre um AR e a entidade centralizada (LMA ou MCF) [67].

$$\xi = \mathrm{h_{ar,ar}}/\mathrm{h_{ar,lma}} \qquad (5.2)$$

O modelo geral da rede e a troca de mensagens necessária em cada proposta são mostrados na Figura 5.1 e Figura 5.2 respectivamente.

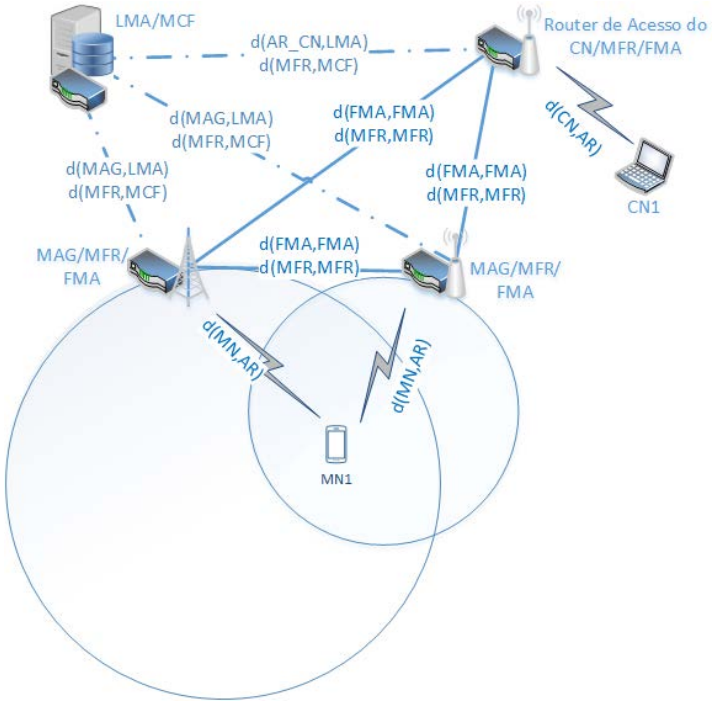

Figura 5.1: Modelo de Rede para avaliação.

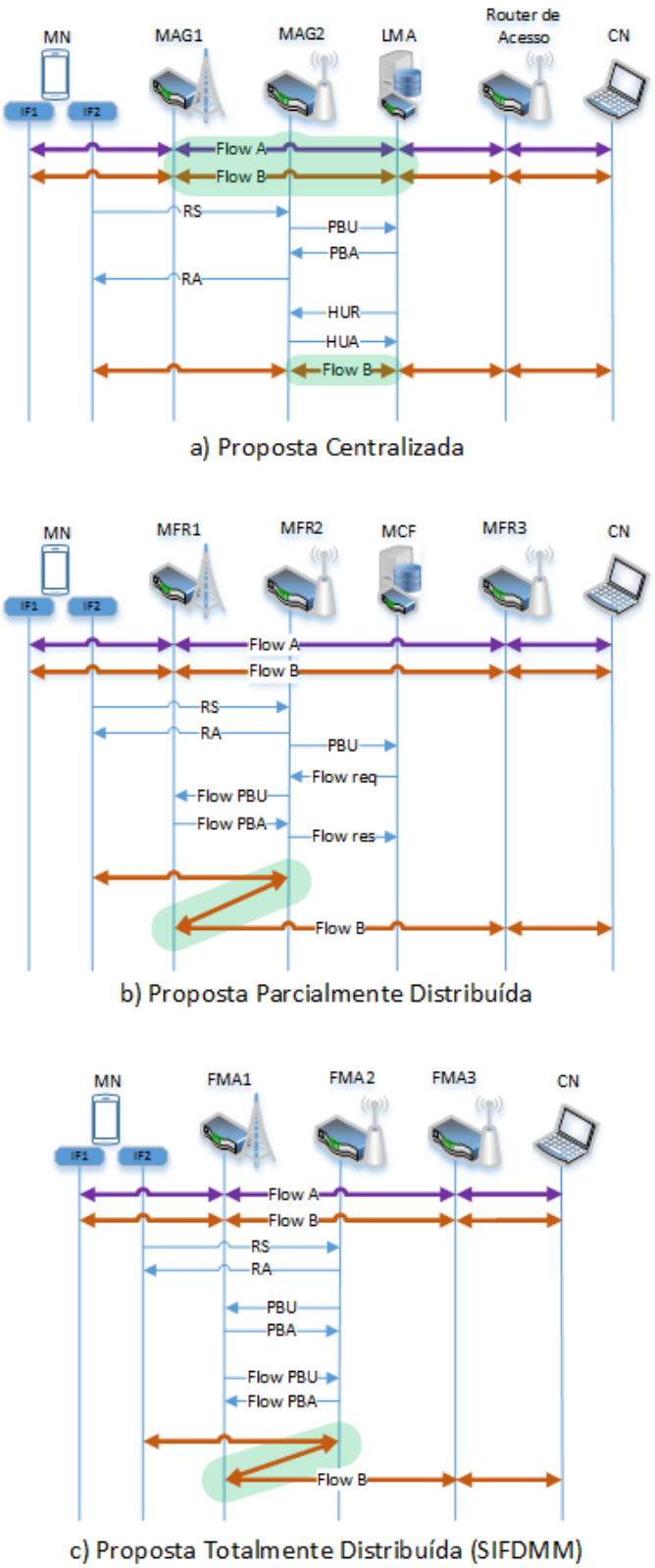

Figura 5.2: Troca de mensagens para escoamento de fluxo.

5.1.2 - Modelo de atraso nos *anchors* da rede

Consideramos um cenário onde os MN podem ter várias sessões ativas simultaneamente, por diferentes interfaces. As filas nos *anchors* da rede são modeladas como um sistema M/M/1. Assumimos que a chegada de pacotes de dados para o MN segue um processo de Poisson com taxa média λ. Logo a taxa média de chegada de pacotes de dados, em um *anchor* da rede (λ_d) é obtida pela equação (5.3).

$$\lambda_{\mathrm{d}} = \sum_{\mathrm{i}=1}^{\mathrm{N}} \lambda_{\mathrm{i}} = \mathrm{N}\lambda \tag{5.3}$$

Onde N é o número de MN que é gerenciado pela entidade. Os tempos de processamento dos pacotes de dados e de controle, em cada nó *anchor*, assumimos que segue uma distribuição de Poisson com taxa média μ_d e μ_c respetivamente. Com isto obtemos, que o tempo médio gasto no sistema (T_s, tempo de enfileiramento e de processamento nos anchors da rede), pode ser expresso como:

$$\mathrm{T_{sd}} = \frac{1}{\mu_{\mathrm{d}} - \lambda_{\mathrm{d}}} \tag{5.4}$$

$$\mathrm{T_{sc}} = \frac{1}{\mu_{\mathrm{c}} - \lambda_{\mathrm{c}}} \tag{5.5}$$

Onde $\rho = \frac{\lambda}{\mu}$ é a utilização do servidor (nó anchor), λ_c é a taxa média de chegada de pacotes de controle, e T_{sd} *e* T_{sc} são os tempos médios gastos em cada nó anchor da rede, de um pacote de dados e controle respetivamente.

5.1.3 - Modelo de atraso nos enlaces sem fio e cabeado

Consideramos que o tempo de envio de um pacote sobre enlaces sem fio e cabeados incluí tempos de transmissão e propagação [67] .

Assumimos que os enlaces cabeados são confiáveis e não é preciso retransmissão. Então o atraso de um pacote de tamanho S_p enviado de um nó x para um no y é expressa pela equação:

$$d_{x,y}(S_p) = h_{x,y} * \left(\frac{S_p}{B_{wd}} + L_{wd}\right) \quad (5.6)$$

Onde B_{wd} e L_{wd} são a largura de banda e o atraso de propagação em enlaces com fio, respetivamente.

Os enlaces sem fio são considerados não confiáveis, o que pode introduzir falhas na transmissão de um pacote e precisar várias retransmissões. Logo temos que P_f é a probabilidade de falha do enlace sem fio. Com isto o atraso de um pacote de tamanho S_p em um enlace sem fio é [67]:

$$d_{wl}(S_p) = \left(\frac{S_p}{B_{wl}} + L_{wl}\right) * \left(\frac{1}{1 - P_f}\right) \quad (5.7)$$

Onde B_{wl} e L_{wl} são a largura de banda e o atraso de propagação no enlace sem fio, respetivamente.

5.1.4 - Análises da latência de *handover* de fluxo

Quando o MN chega a uma área de sobreposição de duas redes de acesso com tecnologias diferentes, é preciso realizar o *attachment* de camada de enlace, pela nova interface. A latência deste processo (T_{L2}) não depende do esquema de mobilidade utilizado, pelo que é considerado igual nas três propostas em analises.

A latência de detecção de movimento (T_{MD}) é composta pela troca das mensagens *Router Solicitation* (RS) e *Router Advertisement* (RA), entre o MN e o novo AR, sobre o enlace sem fio, sendo expressa como:

$$T_{MD}^{PMIPv6} = T_{MD}^{PMIPv6-PD} = 2d_{wl}(S_{RS}) \quad (5.8)$$

$$T_{MD}^{SIFDMM} = 2d_{wl}(S_{RS'}) \quad (5.9)$$

A latência de atualização da localização (T_{LU}) é dividida na latência da transmissão (L_{tc}) e a latência nos anchors da rede (T_{sc}), das mensagens de controle.

$$T_{LU} = L_{tc} + T_{sc} \quad (5.10)$$

O plano de controle das propostas PMIPv6 e PMIPv6_PD é centralizado. Logo a entidade centralizada da rede (LMA/MCF), vão ter que processar as solicitações dos $\frac{K*N}{100}$, MNs que chegam as áreas de sobreposição simultaneamente. Sendo $\lambda_c = \frac{K*N}{100}$. Com isto, T_{sc} é expresso como:

$$\mathrm{T_{sc}^{PMIPv6}} = \mathrm{T_{sc}^{PMIPv6-PD}} = \frac{1}{\mu_c - \frac{K*N}{100}} \tag{5.11}$$

O plano de controle da proposta SIFDMM é distribuído, logo os FMAs terão que processar as solicitações de $\frac{K*N}{100*(M-1)}$ MNs que chegam à área de superposição simultaneamente, sendo $\lambda_c = \frac{K*N}{100*(M-1)}$. Então T_{sc} é expressa como:

$$\mathrm{T_{sc}^{SIFDMM}} = \frac{1}{\mu_c - \frac{K*N}{100*(M-1)}} \tag{5.12}$$

Finalmente a T_{LU} das soluções em analises é expressa como:

$$\mathrm{T_{LU}^{PMIPv6}} = 2\mathrm{d_{mag,lma}}(\mathrm{S_{PBU}}) + 2\mathrm{d_{mag,lma}}(\mathrm{S_{flow}}) + \mathrm{T_{sc}^{PMIPv6}} \tag{5.13}$$

$$\mathrm{T_{LU}^{PMIPv6-PD}} = \mathrm{d_{mfr,mcf}}(\mathrm{S_{PBU}}) + 2\mathrm{d_{mfr.mcf}}(\mathrm{S_{flow}}) + 2\mathrm{d_{mfr,mfr}}(\mathrm{S_{flow}}) + \mathrm{T_{sc}^{PMIPv6-PD}} \tag{5.14}$$

$$\mathrm{T_{LU}^{SIFDMM}} = 2\mathrm{d_{fma,fma}}(\mathrm{S_{PBU}}) + 2\mathrm{d_{fma,fma}}(\mathrm{S_{flow}}) + \mathrm{T_{sc}^{SIFDMM}} \tag{5.15}$$

A latência de envio do primeiro pacote pela nova interface (T_p), é dividido na latência de transmissão (L_{td}) e a latência nos *anchors* da rede (T_{sd}) do pacote de dado.

$$\mathrm{T_p} = \mathrm{L_{td}} + \mathrm{T_{sd}} \tag{5.16}$$

O plano de dados da proposta PMIPv6 é centralizado. Logo o LMA, deve processar os pacotes de dados dos N nó móveis da rede. Com isto, a taxa média de chegada de pacotes no LMA é $\lambda_d = N\lambda$. Então T_{sd} é expresso como:

$$\mathrm{T_{sd}^{PMIPv6}} = \frac{1}{\mu_d - \mathrm{N}\lambda} \tag{5.17}$$

Nas soluções PMIPv6-PD e SIFDMM, o plano de dados é distribuído. Logo, assumindo que os MNs são distribuídos uniformemente, cada nó *anchor* da rede, deve processar os pacotes de $\frac{N}{M}$ nós móveis . Com isto, a taxa média de chegada de pacotes no FMA/MFR é $\lambda_d = \frac{N}{M}\lambda$. Então T_{sd} é expresso como:

$$\mathrm{T_{sd}^{PMIPv6-PD}} = \mathrm{T_{sd}^{SIFDMM}} = \frac{1}{\mu_d - \frac{N}{M}\lambda} \tag{5.18}$$

Finalmente a T_p das soluções em estudo é expressa como:

$$\mathrm{T_p^{PMIPv6}} = 2\mathrm{d_{wl}}(\mathrm{S_{data}}) + \mathrm{d_{ar,lma}}(\mathrm{S_{data}}) + \mathrm{d_{lma,mag}}(\mathrm{S_{data}} + \tau) + \mathrm{T_{sd}^{PMIPv6}} \tag{5.19}$$

$$\mathrm{T_p^{PMIPv6-PD}} = 2\mathrm{d_{wl}}(\mathrm{S_{data}}) + \mathrm{d_{mfr.mfr}}(\mathrm{S_{data}}) + \mathrm{d_{mfr,mfr}}(\mathrm{S_{data}} + \tau) + \mathrm{T_{sd}^{PMIPv6-PD}} \tag{5.20}$$

$$\mathrm{T_p^{SIFDMM}} = 2\mathrm{d_{wl}}(\mathrm{S_{data}}) + \mathrm{d_{fma,fma}}(\mathrm{S_{data}}) + \mathrm{d_{fma.fma}}(\mathrm{S_{data}} + \tau) + \mathrm{T_{sd}^{SIFDMM}} \tag{5.21}$$

5.1.5 - Resultados numéricos

A continuação apresentamos e discutimos os resultados numéricos. Estudamos o impacto de alguns parâmetros na latência de *handover* de fluxo e a quantidade de mensagens a serem transmitidas, em soluções baseadas em PMIPv6. Os valores por default dos parâmetros do sistema são mostrados na Tabela 5.1 [67-68]. Geralmente o número de saltos médio entre dois AR vizinhos é menor que entre um AR e uma entidade centralizada da rede. Isto significa que a escala de rede é $\xi \leq 1$. Na literatura o valor por default é geralmente considerado entre 0.2 e 0.5 [67-68].

Tabela 5.1: Valores por Default dos Parâmetros do Sistema.

Parâmetro	Valor	Unidade	Descrição
M	3		*Router* de Acesso
N	40		Número de MN no domínio
K	50	%	Percentual de MN que entram em uma área de sobreposição simultaneamente
$\boldsymbol{h_{fma,fma} = h_{mfr,mfr}}$	$\sqrt{M} \cong 2$	hops	Saltos entre *Router* de Acesso

$h_{mag,lma} = h_{ar,lma} = h_{mfr,mfc}$	10	hops	Saltos entre AR e a entidade centralizada
ξ	0.2		Escala de rede
λ	0.01	pct/ms	Taxa média de chegada de pacotes do MN
μ_d	5	pct/ms	Taxa média de processamento dos pacotes de dados
μ_c	0.1	pct/ms	Taxa média de processamento dos pacotes de controle
B_{wd}	100	Mbps	Largura de banda do enlace cabeado
L_{wd}	0.5	ms	Retardo de propagação do enlace cabeado
B_{wl}	54	Mbps	Largura de banda do enlace sem fio
L_{wl}	2	ms	Retardo de propagação de enlace sem fio
P_f	0.5		Probabilidade de falha do enlace sem fio
T_{L2}	2	ms	Tempo de *attachment* da camada de enlace
τ	40	bytes	Tamanho do cabeçalho IP nos túneis
S_{rs}	50	bytes	Tamanho das mensagens RS e RA das propostas PMIPv6 e PMIPv6-PD
$S_{rs'}$	70	bytes	Tamanho das mensagens RS e RA da proposta SIFDMM
S_{PBU}	76	bytes	Tamanho das mensagens PBU e PBA
S_{flow}	100	bytes	Tamanho das mensagens que contém opções de fluxo
S_{data}	400	bytes	Tamanho dos pacotes de dados

Considerando os valores da Tabela 5.1 e a troca de mensagens mostradas na Figura 5.2, é possível estimar o número de bytes (B_{Tx}) a serem transmitidos em cada arquitetura, para estabelecer comunicação com o novo *router* de acesso e solicitar o processo de escoamento, como:

$$B_{Tx}^{\text{PMIPv6}} = 2S_{rs} + 2S_{PBU} + 2S_{flow} = 452\ bytes \quad (5.22)$$

$$B_{Tx}^{\text{PMIPv6-PD}} = 2S_{rs} + S_{PBU} + 4S_{flow} = 576\ bytes \quad (5.23)$$

$$B_{Tx}^{\text{SIFDMM}} = 2S_{rs'} + 2S_{PBU} + 2S_{flow} = 492\ bytes \quad (5.24)$$

Os resultados mostram que nas arquiteturas distribuídas ocorre o aumento do número de bytes a serem transmitidos, mas como será analisado à continuação, apresentam menor latência no processo de escoamento de dados. A arquitetura SIFDMM utiliza a mesma quantidade de mensagens que a solução centralizada, aumentando só o tamanho das mensagens RS e RA, mostrando o melhor desempenho.

A latência de *handover* de fluxo pode estar afetada pela quantidade de MNs que entram em uma área de superposição simultaneamente e fazem escoamento de fluxo. Investigamos o impacto variando o parâmetro K de 0 a 100%. A Figura 5.4 mostra que na medida que aumentam a quantidade de nós que estão fazendo solicitações simultaneamente, a FHL das propostas onde o plano de controle é centralizado, começa a aumentar. Isto é devido a demoras nas filas nos *anchors* da rede, quando solicitações de controle são feita por vários nós simultaneamente, como se pode apreciar na Figura 5.3.

A Figura 5.3 mostra que a proposta PMIPv6-PD é a que maior latência de atualização da localização apresenta. Isto é produto que a troca de mensagens de sinalização é maior, tendo que atualizar mais entidades na rede. Mas o maior valor de FHL é apresentado pela proposta PMIPv6 totalmente centralizada, e o melhor comportamento é da solução totalmente distribuída SIFDMM.

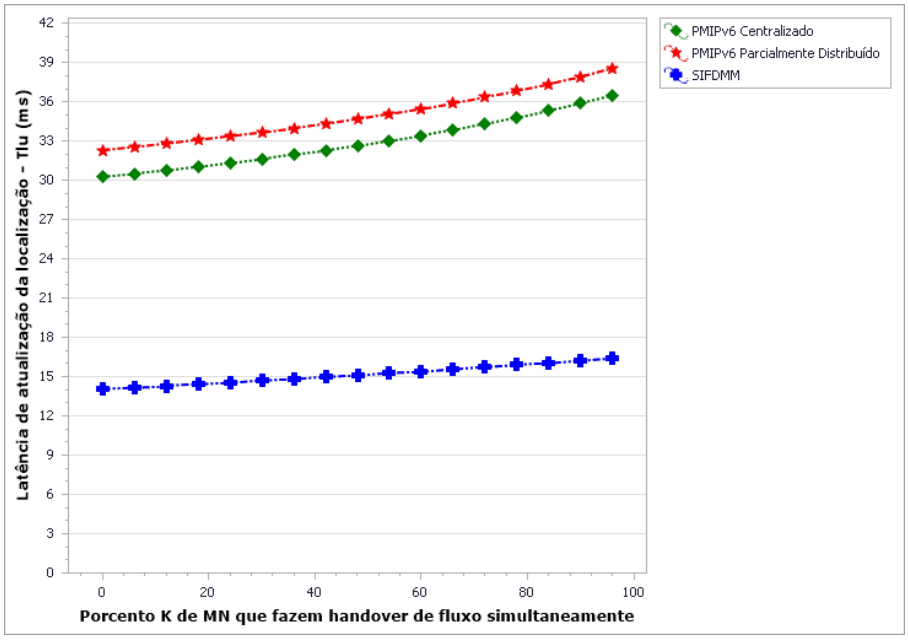

Figura 5.3: Impacto de K em TLU.

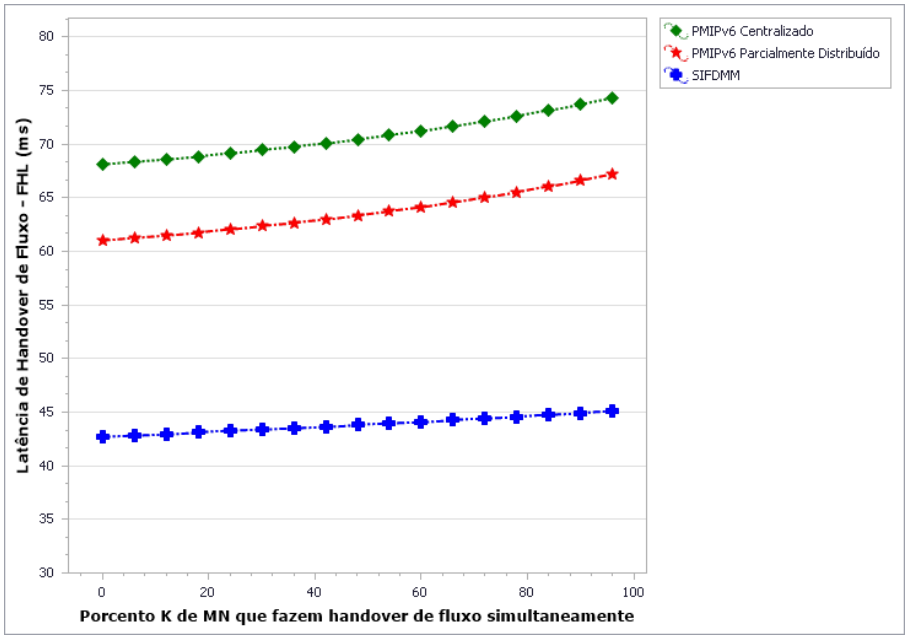

Figura 5.4: Impacto de K em FHL.

O aumento da quantidade de MNs no domínio também afeta a FHL. Para a análise variamos a quantidade de MN de 1 até 100. Podemos observar nas Figura 5.5, Figura 5.6 e Figura 5.7, que o LMA é a primeira entidade em ficar congestionada. A proposta PMIPv6-PD mesmo precisando maior sinalização de controle, tem menor tempo FHL que a solução centralizada, pois o plano de dados é distribuído e pode gerenciar maior quantidade de MN. Mas a proposta SIFDMM sempre apresenta os menores tempos, pois não aumenta a sinalização de controle, senão o tamanho das mensagens RS e RA, mas o impacto na FHL não é significativo.

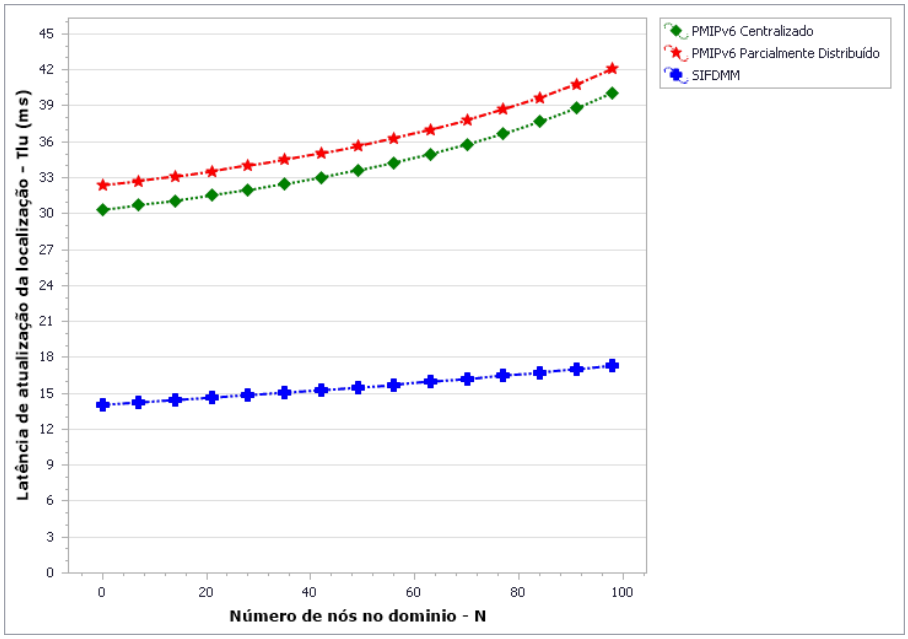

Figura 5.5: Impacto de N em TLU.

Nas Figura 5.6 e Figura 5.7 pode-se observar que quando a quantidade de nós no domínio é aproximadamente 50, o sistema, na proposta centralizada, torna-se altamente congestionado. Então, a entidade está recebendo mais petições que as que podem processar, o que pode ocasionar tempos de espera nas filas muito longos e perdas de pacotes, caso não ser considerado buffer infinito. As propostas distribuídas mantêm um comportamento estável, sem aumentar muito a FHL. É importante destacar que com maior quantidade de nós o plano de controle poderia ficar congestionado também. Nesta situação a solução parcialmente distribuída, onde o plano de controle é centralizado, apresentaria piores resultados tornando-se o sistema altamente congestionado para esse N. Resultados como esses são mais demorados na solução SIFDMM, pois tanto o plano de dados quanto o plano de controle são distribuídos e a escalabilidade é maior.

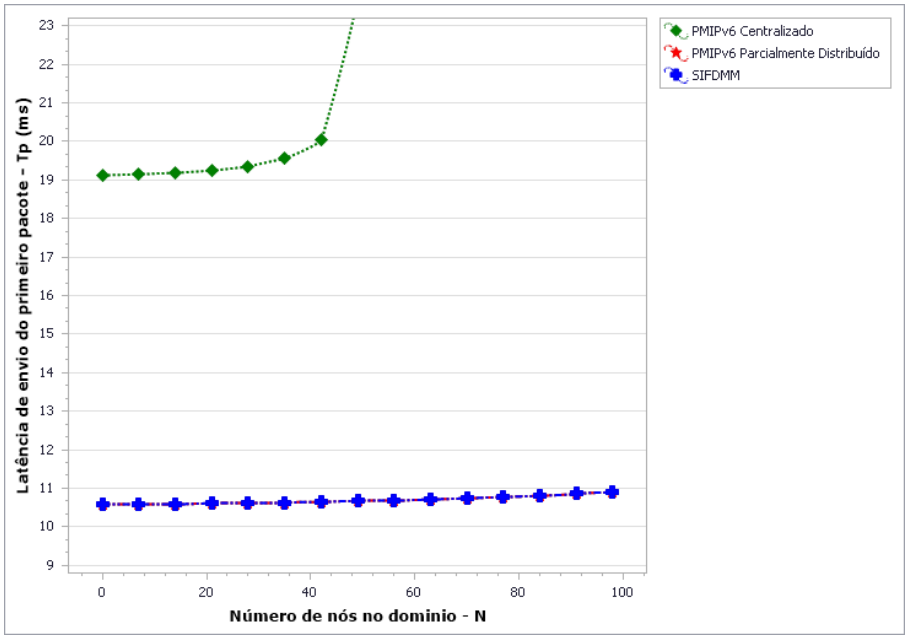

Figura 5.6: Impacto de N em Tp.

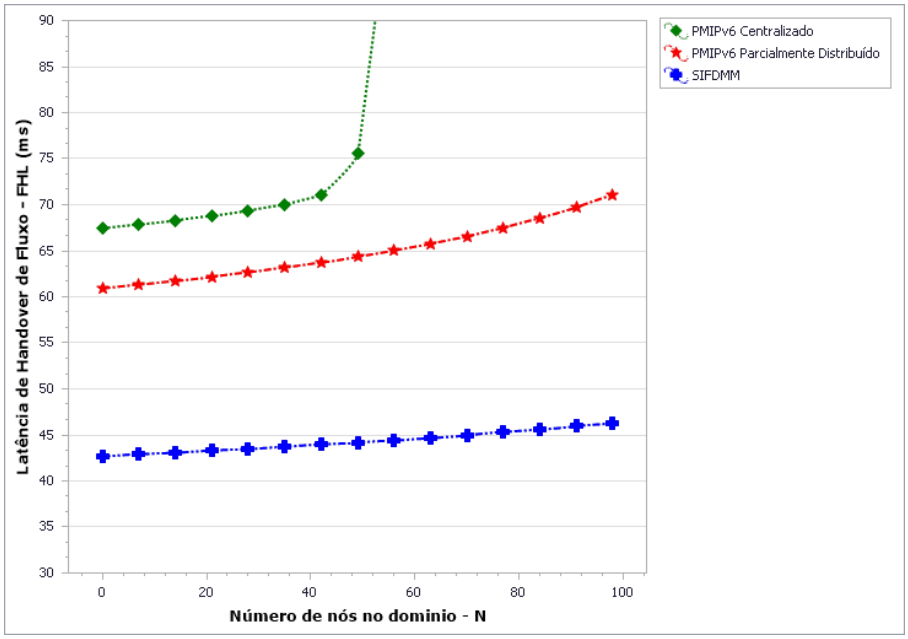

Figura 5.7: Impacto de N em FHL

Outro parâmetro que pode afetar a FHL é a transmissão dos pacotes sobre os enlaces. Logo, nos investigamos o impacto, da probabilidade de falha no enlace sem fio P_f e a escala de rede ξ, na FHL. Primeiro variamos P_f de 0 até 0.8. A Figura 5.8 mostra a variação de FHL como função de P_f. Os resultados mostram que o enlace sem fio, tem um impacto reduzido nas três propostas. Isto se deve que as três são baseadas na rede, e o MN não tem participação

na sinalização relativa ao gerenciamento de mobilidade. Embora com o aumento da P_f aumenta a FHL produto das retransmissões no enlace sem fio.

A Figura 5.9 mostra o impacto de ξ variando desde 1/10 até 1. A escala de rede não tem impacto em PMIPv6 devido a que é uma arquitetura centralizada. Em contrapartida, as soluções distribuídas presentam melhor desempenho quando, os ARs estão perto um dos outros, que é o caso típico previsto.

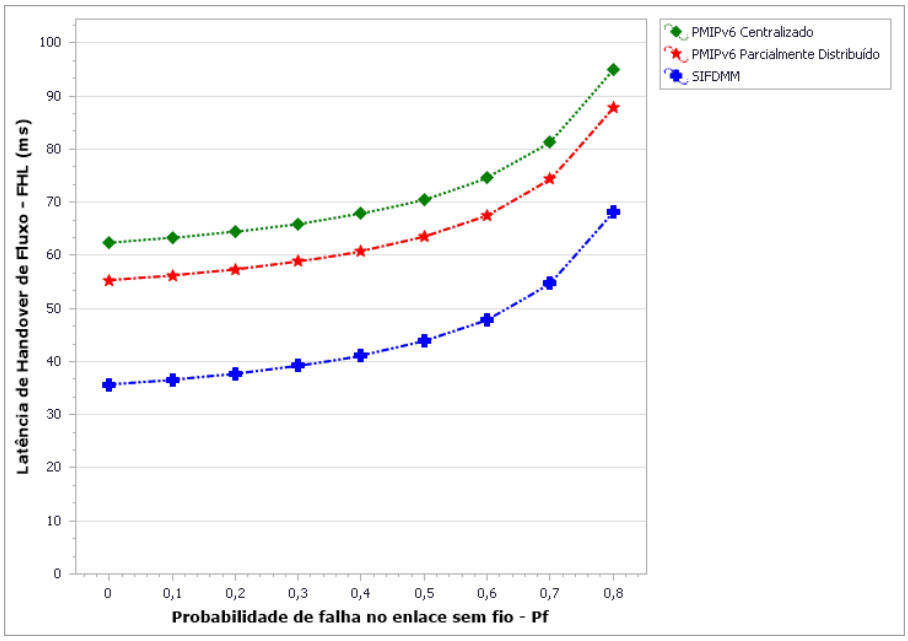

Figura 5.8: Impacto de Pf em FHL.

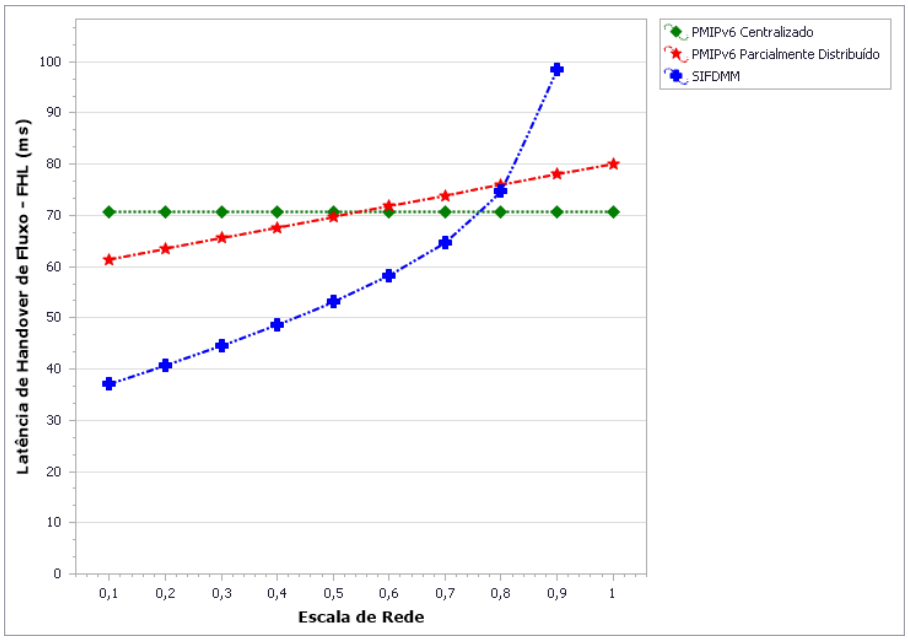

Figura 5.9: Impacto da Escala de Rede em FHL.

Os resultados mostram que as arquiteturas distribuídas superam à centralizada na maioria dos casos. Embora, quando a distância entre os ARs é grande, a FHL aumenta nas soluções distribuídas, podendo ser bem maior que na centralizada.

5.2 - SIMULAÇÃO

O tráfego na Internet, especificamente multimídia, é especialmente atraente para os usuários finais dos sistemas sem fio. Na atualidade, diferentes métodos tem sido estudados e desenvolvidos com o fim de avaliar a qualidade dos sistemas multimídia, sendo alguns deles baseados na visão humana e/ou em analises matemático, tendo em conta o fato que muitos fatores podem afetar a qualidade do conteúdo. Entre esses fatores encontram-se as condições da transmissão, sendo preciso manter na rede, uma adequada QoS. Logo, é de interesse avaliar o desempenho de streaming de vídeo nas arquiteturas. Com o fim de atender dita necessidade, realizamos simulações no *Network Simulator* versão 3 (NS-3) [69], utilizando o Evalvid [70] como ferramenta para a avaliação da qualidade do vídeo transmitido e a qualidade dos serviços nas três arquiteturas em estudo.

NS3 é um simulador para sistemas de Internet baseado em eventos discretos. Desenvolvido especialmente para pesquisa e uso educacional. O projeto ns-3 iniciou em 2006 e tem seu código aberto, desenvolvido nas linguagens C++ e Python [69],[71-72]. No estudo em questão foi utilizada a versão ns-3.23, que foi a mais recente e estável no decorrer desta dissertação. O NS-3 foi instalado no sistema operacional Ubuntu 12.04 LTS, que não apresentou problemas de compatibilidade com as dependências requeridas pelo simulador.

Para a análise do desempenho, foram utilizados três cenários que correspondem com as arquiteturas de rede avaliadas na modelagem analítica, como mostra a Figura 5.10. Na simulação foi necessária a modificação de alguns módulos do NS-3 como descrito na última Seção do APÊNDICE I.

O modelo de rede é formado por três redes de acesso, uma com tecnologia LTE e duas com tecnologia WiFi – 802.11g e o núcleo de rede, com acesso à Internet. Conforme será observado, para a simulação serão considerados aspectos e configurações de 5 (cinco) camadas (física, MAC, rede, transporte e aplicação), diferentemente do que foi feito na modelagem analítica, que se restringiu à camada de rede. As configurações consideradas no NS-3 nas camadas física, enlace, rede, transporte e aplicação utilizadas nos diferentes cenários, são descritas no APÊNDICE I.

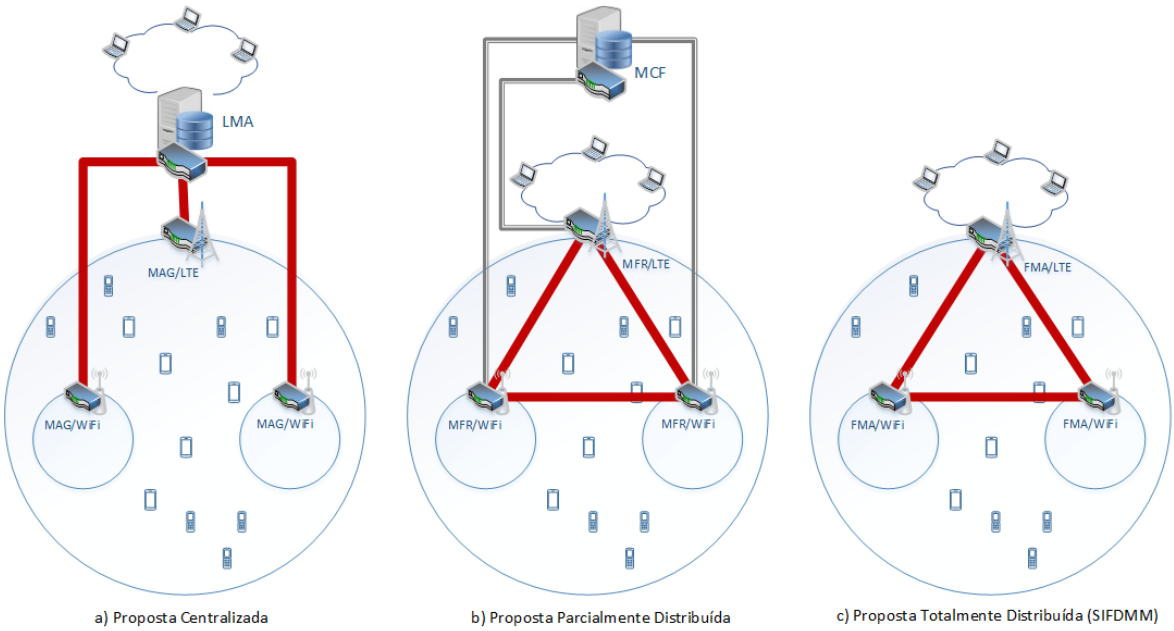

Figura 5.10: Modelo de rede na simulação.

5.2.1 - Cenários simulados

O estudo em questão é focado no escoamento de dados através da movimentação de fluxos IP, entre redes de acesso com diferentes tecnologias, e no gerenciamento de mobilidade. Logo, na implementação é considerada uma rede de acesso LTE com uma elevada demanda de transmissão de dados, o que gera um ambiente congestionado. Isto traz a necessidade de movimentar, assim que for possível, parte do tráfego de dados para uma rede de acesso complementar, sendo neste caso as redes WiFi. Assim é considerada uma simples política de decisão de escoamento: sempre que estabeleça conexão com uma interface WiFi, movimentar os fluxos, conhecida como *on-the-spot offloading* [73-74]. Neste tipo de escoamento, o tráfego é transmitido através da rede celular unicamente quando não há disponibilidade WiFi. Quando os usuários se movem para fora da cobertura WiFi, o escoamento é interrompido e todas as transferências incompletas são completadas em redes celulares. A maioria dos *smartphones* com Wi-Fi, já estão realizando *on-the-spot offloading* como padrão [75], no caso deste trabalho a decisão de escoamento é tomada pela rede.

São enviados *streams* de vídeos, utilizando a integração do modulo de Evalvid no ns-3, de forma concorrente com tráfego *backgroung* CBR e FTP, como descrito no APÊNDICE I.

Foram desenvolvidos três cenários que correspondem a cada uma das arquiteturas. Para cada cenário foram feitas simulações variando a quantidade de nós móveis (terminais

multimodais) entre 10 e 100. Os parâmetros das camadas física, de enlace, rede, transporte e aplicação utilizados na simulação são descritos na Tabela 5.2, e o modelo de mobilidade utilizado e o tráfego gerado por cada MN, assim como os serviços de QoS são explicados na Tabela 5.3.

Tabela 5.2: Parâmetros de configuração da simulacão (Pilha TCP/IP).

Parâmetro		**Valor**
WiFi	Padrão	802.11g e 802.11e/QoS
	Modulação	ERP-OFDM
	Taxa de transmissão	54Mbps
	Potência de transmissão	15 dbm
	Limiar de detecção para recepção de sinais	-96 dbm
	Limiar de detecção de uso de canal	-99 dbm
	Ganho da antena	2.2
	Figura de ruído	7 dbm
	Quantidade de nós no inicio	5%
	Modelo de perdas do caminho	RangePropagationLossModel
LTE	Modulação	DL – OFDM UL - FDMA
	Taxa de transmissão	DL- 6 blocos de recursos UL - 6 blocos de recursos
	Potência de transmissão	eNodeB - 30 dbm UE – 10 dbm
	Ganho da antena	0
	Figura de ruído	eNodeB - 5 dbm UE – 9 dbm
	Quantidade de nós no inicio	90%
	Modelo de perdas do caminho	FriisPropagationLossModel
PMIPv6 – camada de rede		centralizedPmipv6/IPv6 pdPmipv6/IPv6 fdPmipv6/IPv6

Camada de Transporte	TCP/UDP
Camada de Aplicação	FTP/CBR/Vídeo

Tabela 5.3: Modelo de tráfego e mobilidade dos terminais na simulação.

Percentagem dos MN		**Modelo de tráfego**	**AC**	**QCI**	**Modelo de mobilidade**
Estático	10%	FTP	AC_BE	NGBR_VIDEO_TCP_OPERATOR	*ns3::ConstantPositionMobilityModel*
	20%	CBR	AC_VO	GBR_CONV_VOICE	
	20%	Vídeo	AC_VI	GBR_CONV_VIDEO	
Móveis	10%	FTP	AC_BE	NGBR_VIDEO_TCP_OPERATOR	*ns3:: WaypointMobilityModel*
	20%	CBR	AC_VO	GBR_CONV_VOICE	
	20%	Vídeo	AC_VI	GBR_CONV_VIDEO	*ns3::WaypointMobilityModel* (assegura que no segundo 5 da simulação os nós fazem *handover* de fluxo)

Finalmente podemos resumir que o 50% dos MN vão ficar estático e outro 50% vão se movimentar, onde se assegura que um 20% dos MN, gerando tráfego de vídeo, entra na cobertura WiFi no segundo 5 da simulação.

5.2.2 - Resultados da simulação

A continuação são apresentadas e discutidas as métricas de QoS, obtidas com a ferramenta Evalvid, e os valores de latência de *handover* de fluxo (FHL), para cada uma das arquiteturas em estudo. Também são obtidos parâmetros de QoV, que mostram a influência da transmissão do vídeo, através da rede, na qualidade do vídeo recebido. Isto facilita a seleção da arquitetura com melhor desempenho, na criação por exemplo de sistemas multimídia.

5.2.2.1 - Latência de *handover* de fluxo (FHL)

Na modelagem analítica foi apresentada a métrica FHL, que permite avaliar as arquiteturas considerando a latência entre a detecção da movimentação e a chegada do primeiro pacote

pela nova interface. Logo, na simulação, foram utilizadas variáveis de tempo, que nos permitem obter a latência entre o envio da mensagem RS, para estabelecer conexão pelo WiFi, até a chegada do primeiro pacote pela interface. A obtenção do FHL na simulação permite-nos ter uma melhor compreensão do desempenho das arquiteturas em estudo, pois contempla parâmetros de toda a pilha de protocolos IP, logrando com o NS-3 um cenário para avaliação mais perto da realidade. É importante esclarecer que na modelagem analítica são contemplados principalmente parâmetros da camada de rede.

Como discutido na modelagem analítica, a FHL está afetada pela quantidade de nós no domínio (N). A Figura 5.11 mostra a variação da FHL, quando o parâmetro N varia entre 1 e 100. Pode-se observar um resultado similar ao obtido através da modelagem analítica, onde a solução centralizada mostra o pior desempenho produto da congestão no LMA. As soluções distribuídas mostram um desempenho estável produto da maior escalabilidade das mesmas. A solução parcialmente distribuída tem maior latência que a solução SIFDMM, devido às demoras no enlace com a entidade de controle centralizada.

Os resultados obtidos por simulação não são exatamente os mesmos obtidos na modelagem analítica; tal divergência é esperada devido à inclusão, na simulação, de parâmetros de outras camadas, como por exemplo: modelos de propagação, modelos de perda, custos de processamento diferente nas entidades, buffers finitos e tratamento de QoS na camada MAC, bem como aspectos decorrentes da adoção de protocolos com configurações específicas, entre outros.

Adicionalmente, destaca-se que o tratamento analítico não considerou detalhes específicos de cada protocolo nem de outras camadas diferentes da camada de rede. Uma das razões é a dificuldade de tratar analiticamente todo o conjunto de variáveis e parâmetros envolvidos em um conjunto de camadas. A simulação permitiu, assim, explorar de forma mais rica as particularidades de cada camada, sem tornar inválidos os resultados obtidos por meio da modelagem analítica.

Assim, apesar da divergência verificada, é lícito considerar que a comparação entre as arquiteturas leva a conclusões semelhantes às obtidas por meio do modelo analítico.

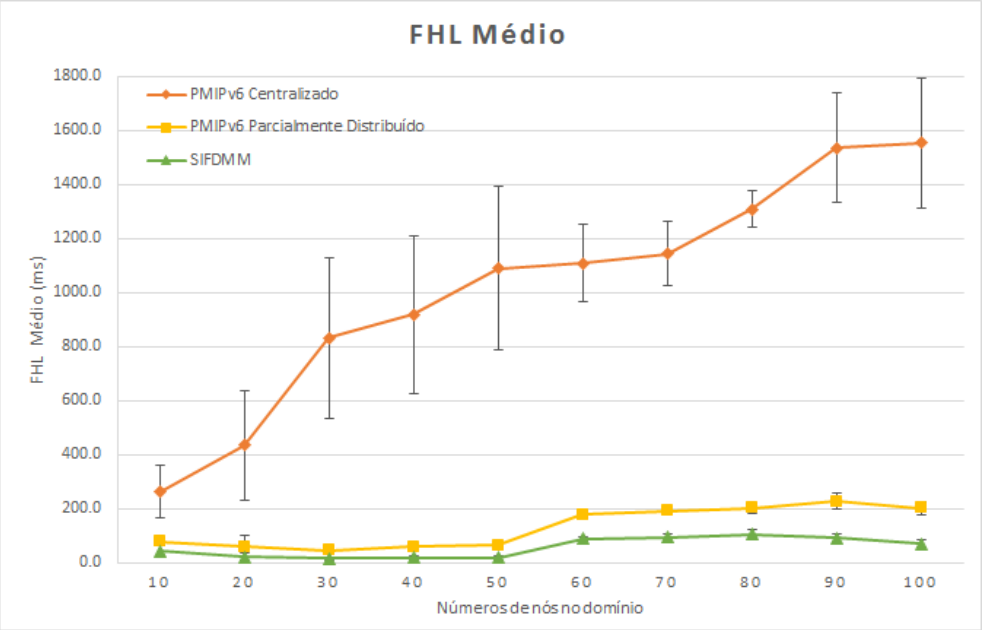

Figura 5.11: Impacto de N em FHL na simulação.

5.2.2.2 - Métricas de QoS

- Atraso: É o tempo entre o envio de um pacote desde o origem e a chegada ao destino. Os pacotes enviados normalmente passam por diferentes equipamentos de rede (exemplo: roteadores), sofrendo diversos tipos de atrasos até chegar ao host destino.

A Figura 5.12 mostra os resultados do atraso médio obtidos na simulação para as três arquiteturas. Pode-se apreciar que na arquitetura centralizada o atraso aumenta consideravelmente com o aumento da quantidade de nós no domínio. Esse resultado é esperado produto da baixa escalabilidade de sistemas centralizados. O LMA (entidade centralizada) tem que processar os pacotes de tráfego e de controle de todos os nós do domínio, comportando-se como um ponto de gargalho na rede. As propostas distribuídas mostram melhor comportamento, pois separam o plano de dados e o plano de controle, tendo maior escalabilidade. As soluções distribuídas tem um comportamento similar, sendo melhor a solução SIFDMM, pelo fato de que o controle é totalmente distribuído, e não introduz os atrasos do link com a entidade centralizada. É importante destacar que na medida que aumentem as entidades da rede o plano de controle da arquitetura parcialmente distribuída vai ficar com maior congestão, o que introduz maior atraso na rede. A partir de 50 nós começa a crescer o atraso produto da alta congestão na rede LTE.

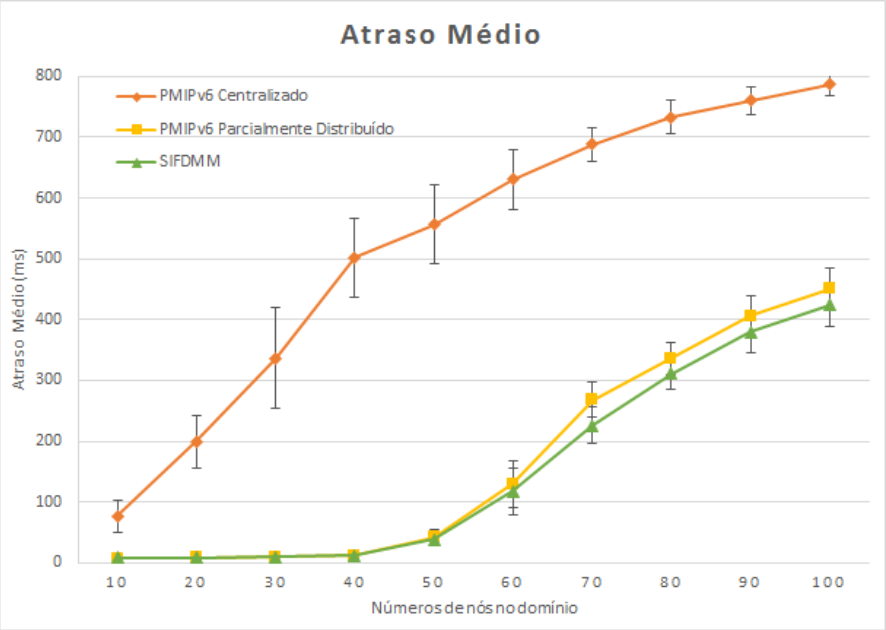

Figura 5.12: Atraso médio.

- Perda de Pacotes: Nas aplicações de vídeo, a retransmissão de pacotes não é tolerável, exemplos de protocolos que podem ser utilizados são o UDP e RTP então este parâmetro representa um indicador de qualidade na transmissão de pacotes entre dois pontos da rede, e que tem influência na qualidade de serviço.

A Figura 5.13 mostra a perda média de pacote fim a fim, nas três arquiteturas em estudo com o aumento de MN no domínio. Ao LMA ter que processar todos os pacotes no domínio, as filas ficam cheias muito mais cedo que nas arquiteturas distribuídas e maior quantidade de pacotes são descartados. A arquitetura SIFDMM consegue ter a menor perda de pacotes devido a que o *handover* de fluxo é feito com maior eficácia diminuindo a congestão na rede LTE antes que a arquitetura parcialmente distribuída.

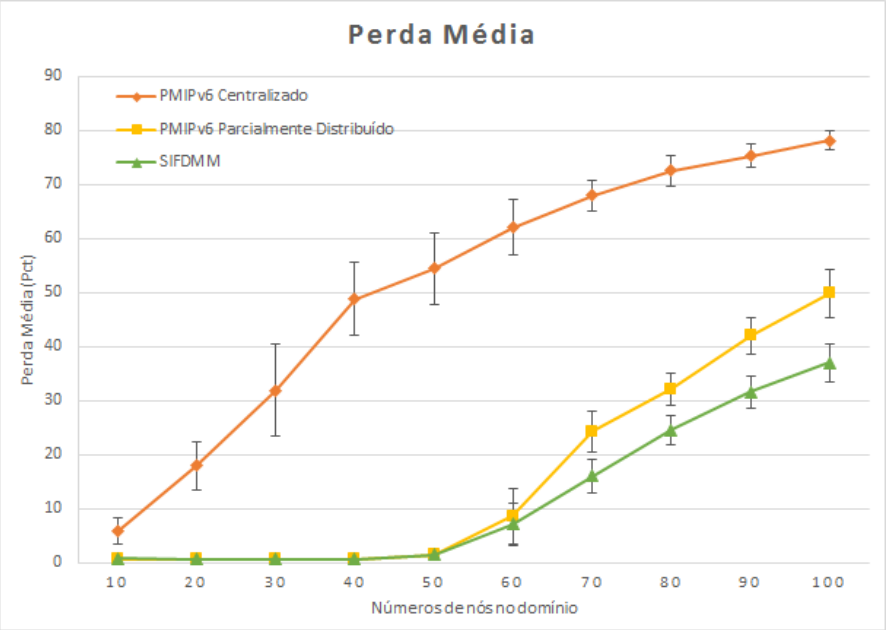

Figura 5.13: Perda média de pacotes.

5.2.2.3 - Métricas de QoE

- *Peak Signal to Noise Ratio* (PSNR): métrica objetiva mais simples e comum para estimar a qualidade de vídeo, somente compara a taxa de erro do vídeo recebido em relação ao vídeo original, o que pode resultar em uma elevada correlação no que respeita à percepção da qualidade subjetiva. A medida de grandeza que utiliza é decibéis (dB).

- *Structural Similarity Index Metric* (SSIM) [76]: avalia o vídeo recebido considerando outros fatores como o SVH (*Sistema Visual Humano*). O SSIM analisa similaridade de cores, luminosidade e estrutura. Os valores extraídos do frame recebido pelo usuário e do frame original são armazenados em vetores separadamente, um vetor para luminosidade, outro vetor para estrutura e um vetor para cor. Posteriormente, obtém-se a média de cada vetor e a combinação dessas três médias gera o valor do SSIM, indicando a qualidade do vídeo com valores entre 0 e 1.

- *Mean Opinion Score* (MOS) [70]: métrica clássica para avaliar a qualidade da experiência do jeito subjetivo. Pode ser calculado como a média de todas as opiniões dos usuários envolvidos nas provas, sendo esta métrica a avaliação mais exata do funcionamento do sistema. Esse processo pode resultar complicado pela necessidade de ter um grupo de sujeitos avaliando a qualidade de todas as sequencias dos vídeos obtidos na simulação. Logo

neste trabalho foi utilizado o resultado oferecido pelo Evalvid que obtém a métrica a partir do mapeamento do PSNR como mostra a Tabela 5.4.

Tabela 5.4: Mapeamento PSNR → MOS [70].

PSNR [dB]	MOS
> 37	5 (Excelente)
31 – 37	4 (Bom)
25 – 31	3 (Médio)
20 – 25	2 (Pobre)
< 20	1 (Ruim)

O resultados das métricas PSNR, SSIM e MOS obtidas na simulação são mostradas nas Figura 5.14, Figura 5.15 e Figura 5.16, respectivamente. É apreciável que a qualidade da experiência é melhor nas arquiteturas distribuídas, pelo fato que a transmissão do vídeo apresenta menores atrasos e perdas como tratado na Seção anterior. Os gráficos mostram que nas soluções distribuídas começam a ser perceptível os erros na transmissão a partir de 90 nós aproximadamente. Isto mostra que na medida que melhora a qualidade dos serviços oferecida, melhora a qualidade da experiência percebida pelo usuário final na transmissão de vídeo. Sendo a arquitetura SIFDMM a que melhor desempenho apresenta, logo pode ser preferível para o emprego em um sistema sem fio de transmissão de vídeo.

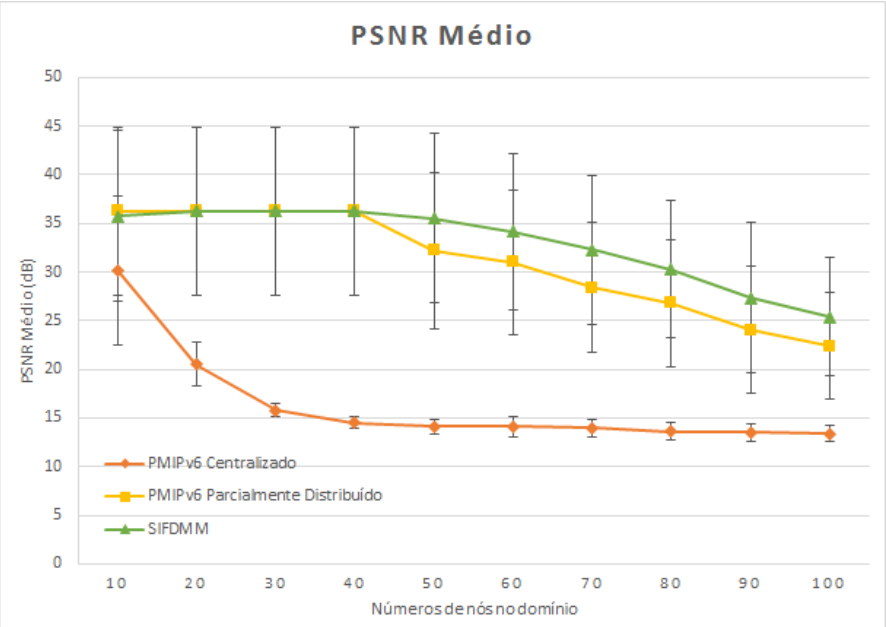

Figura 5.14: PSNR médio.

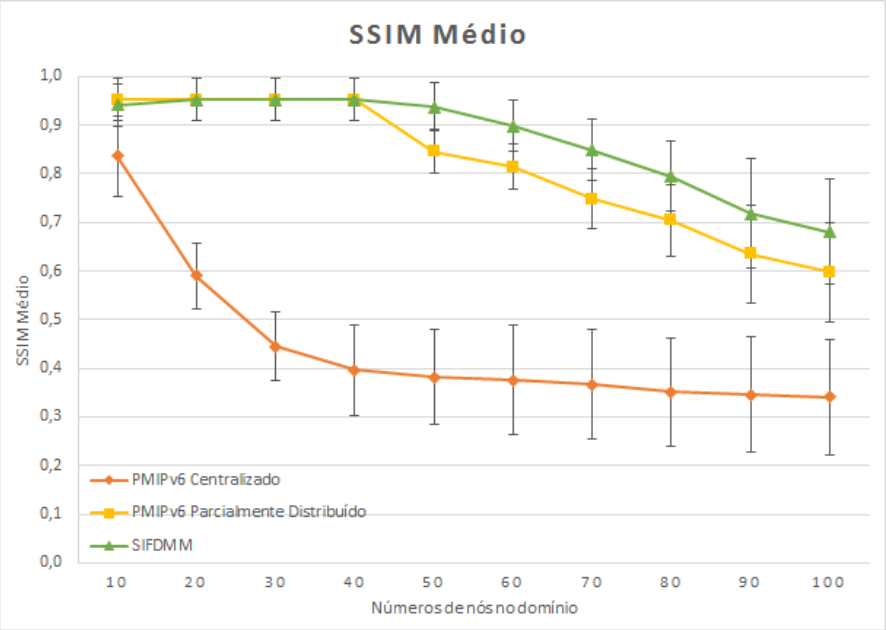

Figura 5.15: SSIM médio.

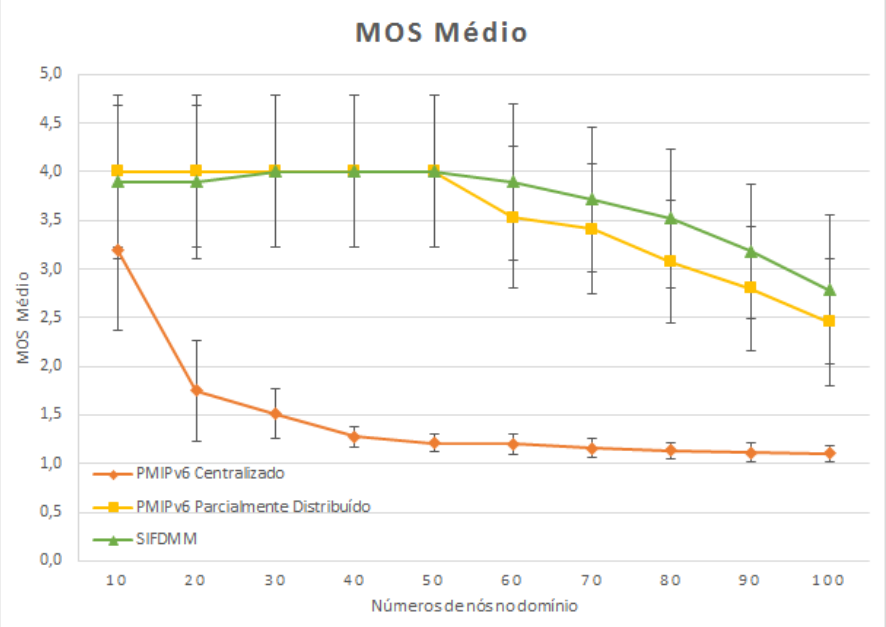

Figura 5.16: MOS médio.

5.3 - CONSIDERAÇÕES FINAIS

Neste capítulo, utilizando a modelagem analítica proposta e simulações a eventos discretos, foram avaliadas três arquiteturas de rede que fazem o gerenciamento de mobilidade do terminal móvel, permitindo movimentar fluxos IP, e assim fazer escoamento de dados entre redes de acesso com diferentes tecnologias.

Com base nos resultados obtidos verifica-se que a arquitetura centralizada sempre mostra o pior desempenho, apresentando maior latência no processo de escoamento de dados e no atraso fim a fim, assim como uma elevada perda de pacotes. Isto é refletido nas transmissões de vídeo, degradando a qualidade da experiência percebida pelo usuário final. Esse comportamento é devido a que as arquiteturas centralizadas possuem uma única entidade, no caso LMA, que é a encarregada de fazer o controle de gerenciamento de mobilidade, assim como o encaminhamento do tráfego de dados, para todos os usuários no domínio. Logo, dita entidade representa um ponto de gargalo na rede. Na medida em que aumenta a quantidade de usuários no domínio, o LMA deverá ter a capacidade para realizar o gerenciamento de cada um deles, o que pode levar a elevados tempos de enfileiramento, e descarte de pacotes, produto dos *buffers* finitos.

Arquiteturas centralizadas são mais simples que arquiteturas distribuídas, toda a informação da rede é armazenada em um só local, o que pode facilitar as tarefas administrativas. Também pode facilitar o monitoramento de segurança, visto que somente um ponto de concentração de dados é mais simples de ser vigiado contra possíveis falhas de segurança. Porém, nas arquiteturas centralizadas é necessário por exemplo replicar a informação de controle, com o fim de ter um certo nível de tolerância a falhas, e se o LMA fica inoperante, o sistema pode ficar inoperante. Esses problemas incrementam-se à medida em que aumenta o número de usuários na rede, podendo até colapsar a rede, produto por exemplo da congestão no LMA. Logo arquiteturas distribuídas, podem ser preferíveis em sistemas com uma elevada quantidade de usuários, embora aumente a complexidade da rede.

As arquiteturas distribuídas podem separar o plano de dados do plano de controle, onde o plano de dados será distribuído, mas o plano de controle ainda será gerenciado por uma entidade centralizada que executará as tarefas de controle, para todos os nós no domínio. Ditas arquiteturas são conhecidas como parcialmente distribuídas. Sendo o plano de dados distribuído é possível otimizar rotas e diminuir o congestionamento nas entidades da rede, o que aumenta a escalabilidade; se alguma entidade na rede falha, os reflexos serão provavelmente menores e recairão predominantemente sobre a parte do sistema que tem a ver com dita entidade que ficará inoperante; boa parte do restante do sistema não será afetada.

Apesar de possibilitarem melhoria do desempenho do sistema, pois o gerenciamento de mobilidade fica em entidades próximas da rede de acesso, observa-se que as arquiteturas

parcialmente distribuídas, por contarem com um ponto único de falha no plano de controle, estão sujeitas à ocorrência de problemas de segurança que levem ao má funcionamento de toda a rede. Uma falha na entidade de controle centralizada pode levar à queda do sistema, sendo, com o aumento de usuários no domínio, mais difícil ter duplicados das informações referente ao controle.

Então, como mostram os resultados obtidos, ditas arquiteturas parcialmente distribuídas superam as centralizadas, mas os parâmetros de QoS começam a piorar na medida que aumentam os nós no domínio, produto da congestão no plano de controle centralizado, repercutindo nos parâmetros de QoE na transmissão de vídeo.

Na arquitetura SIFDMM, totalmente distribuída, tanto o plano de dados quanto o plano de controle são distribuídos. Assim, arquiteturas totalmente distribuídas, serão preferíveis, sempre que a quantidade de usuários no domínio justifique o aumento da complexidade nas entidades da rede de acesso.

Por outro lado, a opção por uma arquitetura totalmente distribuída faz que as entidades FMA sejam mais complexas do que entidades MAG e MFR, pois tem funções para o controle e encaminhamento de dados unificadas em uma só entidade. Os FMA comportam-se como LMA para fluxos iniciados através delas e como MAG no caso de um escoamento de fluxos. Porém, não tem problemas de ponto único de falha, e propicia aumento da escalabilidade, apresentando, como mostram os resultados obtidos, menor latência no processo de escoamento de dados e atrasos fim a fim, assim como menor perda de pacotes. Isto é produto que cada FMA vai gerenciar menor quantidade de nós, diminuindo os tempo de enfileiramento, e com isso o descarte de pacotes. Logo aumenta a QoS na rede, o que vai se refletir em uma melhor QoE na transmissão de vídeo.

6 - CONCLUSÕES E TRABALHOS FUTUROS

Esta dissertação teve por objetivo propor uma nova arquitetura chamada SIFDMM - *Seamless IP Flow and Distributed Mobility Management*, baseada na rede e assistida pelo usuário, é destinada a suportar mobilidade de fluxo IP e escoamento de dados. Com o fim de avaliar o desempenho, foi realizado tratamento analítico da latência de *handover* de fluxo, em uma solução centralizada, uma parcialmente distribuída e a SIFDMM. Em tal modelagem, apenas aspectos da camada de rede foram considerados. Os resultados mostram que SIFDMM, pelo fato de ser totalmente distribuída, é capaz de gerenciar maior quantidade de MN sem congestão nas entidades do núcleo da rede. Também apresenta uma menor latência de *handover* de fluxo na maioria dos cenários. Logo com a solução proposta é possível obter maior vazão e realizar balanceamento de carga, dada a possibilidade dos MN de estarem conectados a dois AR ao mesmo tempo. Além disso, não apresenta os problemas de confiabilidade, escalabilidade e ponto único de falha, encontrados nas outras arquiteturas consideradas.

A arquitetura SIFDMM proposta não foi somente avaliada em termos analíticos. O simulador de redes NS-3 e módulos como *LENA*, *PMIPv6 e Evalvid* permitiram a obtenção de parâmetros de QoS como atraso fim a fim e perda de pacotes, assim como métricas de QoE relativas à transmissão de vídeo como PSNR, SSIM e MOS. Nos cenários simulados a rede de acesso LTE foi progressivamente congestionada, enquanto as WLANs estiveram submetidas a uma baixa demanda para transferência de dados, produto que menor quantidade de nós estão na sua área de cobertura.

Com base nos resultados analíticos, a arquitetura SIFDMM, apresenta o melhor desempenho. Ela é capaz de movimentar os fluxos com uma latência menor, o que se reflete em menores atrasos e menor descarte de pacotes, produto do congestionamento das entidades da rede.

No tocante à simulação, foram considerados aspectos e configurações não somente da camada de rede, mas também das camadas física, de enlace (MAC), de transporte e de aplicação, e considerando tráfegos de vídeo (com codificação H.264), voz e transferência de arquivos (FTP). Algumas conclusões podem ser ressaltadas:

- A simulação apresenta resultados com algumas diferenças em relação aos resultados da modelagem analítica, o que é aceitável por se tratar de duas técnicas de avaliação bem

diferenciadas. Apesar disso, a forma das curvas permite observar um comportamento similar das arquiteturas.

- Por meio da simulação, foram comtemplados parâmetros de toda a pilha de protocolos IP, o que faz com que os resultados fiquem mais próximos da realidade.

- Através da simulação foi possível também aplicar métricas de QoS, assim como métricas de QoE, avaliando o desempenho das arquiteturas na transmissão de vídeo.

Os parâmetros de QoE obtidos mostram um desempenho superior da arquitetura SIFDMM, em comparação com as propostas parcialmente distribuídas e centralizadas. Logo, em sistemas de transmissão de vídeo em redes sem fio, verifica-se que a arquitetura apresenta um desempenho aceitável, permitindo fazer gerenciamento de mobilidade e a movimentação seletiva de fluxos IP.

Em trabalhos futuros, mecanismos para determinar quando se deve movimentar um fluxo, podem ser estudados, assim como a atualização dinâmica das tabelas de regras de fluxo, baseadas no estado da rede e nas preferências do usuário. Também, pretendemos realizar simulações de maior complexidade no que respeita a topologia da rede incluindo novos pontos de acesso. Sugere-se também avaliar a transmissão de diferentes sequências de vídeo que permitam um melhor estudo do desempenho da arquitetura, em aplicações com requisitos de temporização estritos. Finalmente, pode-se considerar em trabalhos futuros comparações da arquitetura com outras soluções totalmente distribuídas.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] C. Perkins, "RFC 5944: IP Mobility Support for IPv4," November 2010.

[2] C. Perkins, D. Johnson, and J. Arkko, "RFC 6275: Mobility support in IPv6," July 2011.

[3] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil, "RFC 5213: Proxy Mobile IPv6 " August 2008.

[4] H. Chan, D. Liu, P. Seite, H. Yokota, and J. Korhonen, "RFC 7333: Requirements for Distributed Mobility Management," August 2014.

[5] D. R. Purohith, A. Hegde, and K. M. Sivalingam, "Network architecture supporting seamless flow mobility between LTE and WiFi networks," in *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2015 IEEE 16th International Symposium on a*, pp. 1-9, 2015.

[6] 3GPP, "TR 23.861 V1.14.1, Network based IP flow mobility (Release 13)," 2015.

[7] 3GPP, "TS 23.261 V12.0.0, IP flow mobility and seamless Wireless Local Area Network (WLAN) offload; Stage 2 (Release 12)," 2014.

[8] H.-B. Lee, S.-G. Min, Y.-H. Han, K.-H. Lee, H.-W. Lee, *et al.*, "IP flow mobility scheme in scalable network-based mobility management architecture," *Telecommunication Systems*, pp. 1-11, 2015.

[9] L. Jong-Hyouk, K. D. Singh, J. M. Bonnin, and P. Sangheon, "Mobile Data Offloading: A Host-Based Distributed Mobility Management Approach," *IEEE Internet Computing, Institute of Electrical and Electronics Engineers*, vol. 18, pp. 20-29, 2014.

[10] F. Giust, C. Bernardos, and A. de la Oliva, "HDMM: deploying client and network-based distributed mobility management," *Telecommunication Systems*, vol. 59, pp. 247-270, 2015.

[11] J. Lee and Y. Kim, "PMIPv6-based Distributed Mobility Management," *IETF draft, draft-jaehwoon-dmm-pmipv6-04*, June 18 2015.

[12] C. Bernardos, A. d. l. Oliva, and F. Giust, "A PMIPv6-based solution for Distributed Mobility Management," *IETF draft, draft-bernardos-dmm-pmip-04*, March 5 2015.

[13] X. Keqiang, L. Jun, and W. Lei, "Design and Implementation of Flow Mobility Based on D-PMIPv6," in *Computational Science and Engineering (CSE), 2014 IEEE 17th International Conference on*, pp. 1344-1349, 2014.

[14] K. Sun and Y. Kim, "Flow Mobility Management in PMIPv6-based DMM (Distributed Mobility Management) Networks," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA)*, vol. 5, pp. 120-127, 2014.

[15] M. Perras and J. Cartmell, "Mobility for heterogeneous SmallNets," *Telecommunication Systems*, pp. 1-18, 2015.

[16] P. Seite, P. Bertin, and J. Lee, "Distributed Mobility Anchoring," *IETF draft, draft-seite-dmm-dma-07*, February 6 2014.

[17] L. Yi, H. Zhou, D. Huang, and H. Zhang, "D-PMIPv6: A distributed mobility management scheme supported by data and control plane separation," *Mathematical and Computer Modelling*, vol. 58, pp. 1415-1426, 2013.

[18] W. Luo and J. Liu, "PMIP Based DMM Approaches," *IETF draft, draft-luo-dmm-pmip-based-dmm-approach-02*, July 29 2013.

[19] K. Xue, L. Li, P. Hong, and P. McCann, "Routing optimization in DMM," *IETF draft, draft-xue-dmm-routing-optimization-02*, June 20 2013.

[20] Y. Khadraoui, X. Lagrange, and A. Gravey, "A Survey of Available Features for Mobile Traffic Offload," in *European Wireless 2014; 20th European Wireless Conference; Proceedings of*, pp. 1-4, 2014.

[21] 3GPP, "TR 23.829 V10.0.1, Local IP Access and Selected IP Traffic Offload (LIPA-SIPTO) (Release 10)," 2011.

[22] C. B. Sankaran, "Data offloading techniques in 3GPP Rel-10 networks: A tutorial," *IEEE Communications Magazine*, vol. 50, pp. 46-53, 2012.

[23] R. Gupta and N. Rastogi, "LTE Advanced - LIPA and SIPTO," *Aricent*, 2012.

[24] L. Bokor, J. Kovács, and C. A. Szabó, "A Home Agent Initiated Handover Solution for Fine-Grained Offloading in Future Mobile Internet Architectures: Survey and Experimental Evaluation," *International Journal of Agent Technologies and Systems (IJATS)*, vol. 6, pp. 1-27, 2014.

[25] R. Koodli, "RFC 5568: Mobile IPv6 Fast Handovers," *Network Working Group*, July 2009.

[26] H. Soliman, "RFC 5555: Mobile IPv6 Support for Dual Stack Hosts and Routers," *Network Working Group* June 2009.

[27] T. Schmidt, M. Waehlisch, R. Koodli, G. Fairhurst, and D. Liu, "RFC 7411: Multicast Listener Extensions for Mobile IPv6 (MIPv6) and Proxy Mobile IPv6 (PMIPv6) Fast Handovers," *Internet Engineering Task Force (IETF)*, November 2014.

[28] H. Soliman, C. Castelluccia, K. ElMalki, and L. Bellier, "RFC 5380: Hierarchical Mobile IPv6 (HMIPv6) Mobility Management," *Network Working Group*, October 2008.

[29] C.-S. Li, F. Lin, and H.-C. Chao, "Routing optimization over network mobility with distributed home agents as the cross layer consideration," *Telecommunication systems*, vol. 42, pp. 63-76, 2009.

[30] Y. Liu, Z. Zhao, T. Lin, and H. Tang, "Distributed mobility management based on flat network architecture," in *Wireless Internet Conference (WICON), 2010 The 5th Annual ICST*, pp. 1-6, 2010.

[31] V. P. Kafle, Y. Kobari, and M. Inoue, "A distributed mobility management scheme for future networks," in *Kaleidoscope 2011: The Fully Networked Human? - Innovations for Future Networks and Services (K-2011), Proceedings of ITU*, pp. 1-7, 2011.

[32] M. Liu, X. Guo, A. Zhou, S. Wang, Z. Li, *et al.*, "Low latency IP mobility management: protocol and analysis," *EURASIP Journal on Wireless Communications and Networking*, pp. 1-16, 2011.

[33] P. J. McCann, "Design of a flat wireless Internet Service Provider network," in *Wireless Personal Multimedia Communications (WPMC), 2011 14th International Symposium on*, pp. 1-5, 2011.

[34] Z. Yujia, W. Yue, Y. Ilsun, Y. Jian, R. Yong, *et al.*, "A DHT and MDP-based mobility management scheme for large-scale mobile internet," in *Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on*, pp. 379-384, 2011.

[35] C. J. Bernardos, J. C. Zunniga, and A. Reznik, "Towards flat and distributed mobility management: A 3GPP evolved network design," in *Communications (ICC), 2012 IEEE International Conference on*, pp. 6855-6861, 2012.

[36] K. Hyunjin, J. Yoonchang, and C. Hyunseung, "An Efficient Load Balancing of Mobile Access Gateways in Proxy Mobile IPv6 Domains," in *Computational Science and Its Applications (ICCSA), 2010 International Conference on*, pp. 289-292, 2010.

[37] S. Gundavelli, "RFC 6543: Reserved IPv6 Interface Identifier for Proxy Mobile IPv6," *Internet Engineering Task Force (IETF)* May 2012.

[38] H. Yokota, K. Chowdhury, R. Koodli, B. Patil, and F. Xia, "RFC 5949: Fast Handovers for Proxy Mobile IPv6," September 2010.

[39] H. A. Chan, "Proxy mobile IP with distributed mobility anchors," in *GLOBECOM Workshops (GC Wkshps), 2010 IEEE*, pp. 16-20, 2010.

[40] M. Boc, A. Petrescu, and C. Janneteau, "Anchor-based routing optimization extension for Proxy Mobile IPv6 in flat architectures," in *Wireless Personal Multimedia Communications (WPMC), 2011 14th International Symposium on*, pp. 1-5, 2011.

[41] P. P. Ernest and H. A. Chan, "Enhanced handover support and routing path optimization with distributed mobility management in flattened wireless networks," in *Wireless Personal Multimedia Communications (WPMC), 2011 14th International Symposium on*, pp. 1-5, 2011.

[42] F. Giust, A. de la Oliva, C. J. Bernardos, and R. P. F. Da Costa, "A network-based localized mobility solution for Distributed Mobility Management," in *Wireless Personal Multimedia Communications (WPMC), 2011 14th International Symposium on*, pp. 1-5, 2011.

[43] W. Hahn, "3GPP Evolved Packet Core support for distributed mobility anchors: Control enhancements for GW relocation," in *ITS Telecommunications (ITST), 2011 11th International Conference on*, pp. 264-267, 2011.

[44] W. Hahn, "Flat 3GPP Evolved Packet Core," in *Wireless Personal Multimedia Communications (WPMC), 2011 14th International Symposium on*, pp. 1-5, 2011.

[45] H. Jung, M. Gohar, J.-I. Kim, and S.-J. Koh, "Distributed Mobility Control in Proxy Mobile IPv6 Networks," *IEICE Transactions on Communications*, pp. 2216-2224, 2011.

[46] D. Liu, J. SONG, and W. Luo, "PMIP Based DMM Approaches," *IETF draft, draft-liu-dmm-pmip-based-approach-02*, March 13 2012.

[47] H. Zhang, F. Qiu, H. Zhou, X. Li, and F. Song, "A Distributed Mobility Management Solution in LISP networks," *IETF draft, draft-zhang-dmm-lisp-00*, September 4 2012.

[48] P. Bertin, S. Bonjour, and J. Bonnin, "A Distributed Dynamic Mobility Management Scheme Designed for Flat IP Architectures," in *New Technologies, Mobility and Security, 2008. NTMS '08.*, pp. 1-5, 2008.

[49] M. Fischer, F. U. Andersen, A. Kopsel, G. Schafer, and M. Schlager, "A Distributed IP Mobility Approach for 3G SAE," in *Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008. IEEE 19th International Symposium on*, pp. 1-6, 2008.

[50] D. Truong-Xuan and K. Younghan, "Distributed network mobility management," in *Advanced Technologies for Communications (ATC), 2012 International Conference on*, pp. 319-322, 2012.

[51] H.-Y. Choi, S.-G. Min, and Y.-H. Han, "PMIPv6-based Flow Mobility Simulation in NS-3," in *Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2011 Fifth International Conference on*, pp. 475-480, 2011.

[52] T. Melia, C. Bernardos, A. de la Oliva, F. Giust, and M. Calderon, "IP Flow Mobility in PMIPv6 Based Networks: Solution Design and Experimental Evaluation," *Wireless Personal Communications*, vol. 61, pp. 603-627, 2011.

[53] H.-Y. Choi, S.-G. Min, Y.-H. Han, and R. Koodli, "Design and Simulation of a Flow Mobility Scheme Based on Proxy Mobile IPv6," *Journal of Information Processing Systems (JIPS)*, vol. 8, pp. 603-620, 2012.

[54] K. Jinho, Y. Morioka, and J. Hagiwara, "An optimized seamless IP flow mobility management architecture for traffic offloading," in *Network Operations and Management Symposium (NOMS), 2012 IEEE*, pp. 229-236, 2012.

[55] C. Makaya, S. Das, and F. J. Lin, "Seamless data offload and flow mobility in vehicular communications networks," in *Wireless Communications and Networking Conference Workshops (WCNCW), 2012 IEEE*, pp. 338-343, 2012.

[56] R. I. Meneguette, L. F. Bittencourt, and E. R. M. Madeira, "A seamless flow mobility management architecture for vehicular communication networks," *Journal of Communications and Networks*, vol. 15, pp. 207-216, 2013.

[57] T. Condeixa and S. Sargento, "Studying the integration of distributed and dynamic schemes in the mobility management," *Computer Networks*, vol. 60, pp. 46-59, 2014.

[58] P. Bertin, S. Bonjour, and J. Bonnin, "Distributed or Centralized Mobility?," in *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, pp. 1-6, 2009.

[59] S. Jeon and Y. Kim, "Cost-efficient network mobility scheme over proxy mobile IPv6 network," *IET Communications*, vol. 5, pp. 2656-2661, 2011.

[60] 3GPP, "TS 23.234 V12.0.0, 3GPP system to Wireless Local Area Network (WLAN) interworking; System description (Release 12)," 2014.

[61] R. Wakikawa, V. Devarapalli, G. Tsirtsis, T. Ernst, and K. Nagami, "Multiple care-of addresses registration," 2009.

[62] G. Tsirtsis, G. Giarreta, H. Soliman, and N. Montavont, "Traffic selectors for flow bindings," *RFC6088*, January 2011.

[63] G. Tsirtsis, H. Soliman, N. Montavont, G. Giaretta, and K. Kuladinithi, "Flow Bindings in Mobile IPv6 and Network Mobility (NEMO) Basic Support (RFC 6089)," *Available from: datatracker. ietf. org/doc/rfc6089*, 2011.

[64] R. Braden, "Requirements for Internet Hosts -- Communication Layers," *RFC 1122*, October 1989.

[65] M. Wasserman and P. Seite, "Current Practices for Multiple-Interface Hosts," *RFC 6419*, November 2011.

[66] T. Melia and S. Gundavelli, " Logical-interface Support for Multi-access enabled IP Hosts," *draft-ietf-netext-logical-interface-support-11*, 2015.

[67] H. Ali-Ahmad, M. Ouzzif, P. Bertin, and X. Lagrange, "Performance Analysis on Network-Based Distributed Mobility Management," *Wireless Personal Communications*, vol. 74, pp. 1245-1263, 2014.

[68] H. Ali-Ahmad, M. Ouzzif, P. Bertin, and X. Lagrange, "Distributed Mobility Management: Approaches and analysis," in *Communications Workshops (ICC), 2013 IEEE International Conference on*, pp. 1297-1302, 2013.

[69] ns-3 project, "ns-3 Model Library. Release ns-3.23," August 13, 2015, https://www.nsnam.org/ns-3-23/, Acessado em: 12/10/2015.

[70] J. Klaue, B. Rathke, and A. Wolisz, "EvalVid – A Framework for Video Transmission and Quality Evaluation," in *Computer Performance Evaluation. Modelling Techniques and Tools: 13th International Conference, TOOLS 2003, Urbana, IL, USA, September 2-5, 2003. Proceedings*, P. Kemper and W. H. Sanders, Eds., ed Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 255-272, 2003.

[71] ns-3 project, "ns-3 Manual. Release ns-3.23," August 13, 2015, https://www.nsnam.org/ns-3-23/, Acessado em: 12/10/2015.

[72] ns-3 project, "ns-3 Rastreamento. Versão ns-3.23," August 13, 2015, https://www.nsnam.org/ns-3-23/, Acessado em: 12/10/2015.

[73] F. Mehmeti and T. Spyropoulos, "Performance analysis of "on the-spot" mobile data offloading," in *Global Communications Conference (GLOBECOM), 2013 IEEE*, pp. 1577-1583, 2013.

[74] F. Rebecchi, M. Dias de Amorim, V. Conan, A. Passarella, R. Bruno, *et al.*, "Data Offloading Techniques in Cellular Networks: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 17, pp. 580-603, 2014.

[75] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile data offloading: how much can WiFi deliver?," presented at the Proceedings of the 6th International Conference, Philadelphia, Pennsylvania, 2010.

[76] V. Jonnalagadda and V. Musti, "Evaluation of Video Quality of Experience using EvalVid," 2012.

[77] WG802.11 - Wireless LAN Working Group, "802.11-2012 - IEEE Standard for Information technology--Telecommunications and information exchange between systems Local and metropolitan area networks--Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," 2012.

[78] Centre Tecnológico de Telecomunicacions de Catalunya (CTTC), "The LENA ns-3 LTE Module. Documentation. Release v8," January 21, 2014, http://lena.cttc.es/manual/, Acessado em: 12/07/2015.

[79] 3GPP, "TS 23.203 V11.6.0, Policy and charging control architecture (Release 11)," 2012.

[80] L. Valtulina, "Seamless Distributed Mobility Management (DMM) solution in cloud based LTE Systems," Master Thesis, Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS), Design and Analysis of Communication Systems (DACS), University of Twente, The Netherlands, 2013.

[81] C. Hyon-Young, M. Sung-Gi, H. Youn-Hee, P. Jungsoo, and K. Hyoungjun, "Implementation and Evaluation of Proxy Mobile IPv6 in NS-3 Network Simulator," in *Ubiquitous Information Technologies and Applications (CUTE), 2010 Proceedings of the 5th International Conference on*, pp. 1-6, 2010.

[82] V. J. R. Galvis, "Qualidade de serviço e de experiência na distribuição de vídeo escalável em redes Par-a-Par (P2P)," Mestrado, Universidade de Brasília, 2013.

APÊNDICE I - Configurações da simulação no NS-3

A seguir são descritos os parâmetros de configuração considerados no NS-3 para avaliação das arquiteturas em estudo mediante a simulação. Adicionalmente, são descritos aspectos de configuração do Evalvid, bem como aspectos atinentes às alterações e complementos feitos ao NS-3.

I.1- REDE DE ACESSO WIFI – MODELO FÍSICO E MAC

A rede de acesso WiFi foi desenhada em modo infraestrutura utilizando o módulo WiFi do ns-3.23, que é baseado na revisão IEEE Std 802.11-2012 [77], para criar um *WiFiNetDevice* (dispositivo de rede WiFi). Na configuração é preciso seguir 4 passos principais:

1. Configuração do canal WiFi – ***WiFiChannel***
2. Configuração da camada física – ***WifiPhy***
3. Configuração da camada MAC – ***WifiMac***
4. Criação do dispositivo WiFi – ***WifiDevice***

Para uma melhor compreensão esses passos serão descritos nas próximas seções.

I.1.1- Configuração do canal WiFi – *WiFiChannel*

O canal WiFi é o encarregado da comunicação entre dois dispositivos via a interface aérea. As principais configurações de um canal WiFi são as relativas ao modelo de perda de propagação e ao modelo do atraso de propagação. Na simulação foi utilizada, para a configuração, a classe *ns3::YansWifiChannelHelper*. Foi configurado um modelo de canal com atrasos de propagação constantes à velocidade da luz (*ns3::ConstantSpeedPropagationDelayModel*) e as perdas de propagação são baseadas em um modelo dependente somente da distância entre o transmissor e o receptor (*ns3::RangePropagationLossModel*).

O modelo *RangePropagationLossModel* só tem um atributo *MaxRange* em metros que determina as perdas no caminho. Os nós receptores dentro da circunferência com centro no AP e radio *MaxRange*, vão receber a transmissão com a mesma intensidade do sinal transmitido, enquanto os receptores fora dessa circunferência vão receber com uma potência de -1000 dBm (efetivamente cero).

Finalmente os parâmetros de configuração foram fixados como:

```
YansWifiChannelHelper wifiChannel = YansWifiChannelHelper::Default ();
wifiChannel.SetPropagationDelay ("ns3::ConstantSpeedPropagationDelayModel");
wifiChannel.AddPropagationLoss ("ns3::RangePropagationLossModel",
                                "MaxRange", DoubleValue (27));
```

I.1.2- Configuração da camada física – *WifiPhy*

O *WifiPhy* cuida do envio e recebimento do sinal sem fio do canal WiFi. O modelo de camada física decide se o *frame* será decodificado com êxito ou não dependendo da intensidade do sinal recebido e o ruído. Assim a configuração principal é o modelo de taxa de erro, que é o que realmente calcula a probabilidade de decodificar um *frame* com sucesso. Os dispositivos físicos (baseados na classe *ns3::WifiPhy*) conectam-se com o modelo *ns3::WifiChannel*. Logo é preciso criar apropriadamente os objetos *WifiPhy* para o canal *YansWifiChannel*, para o que é utilizada a classe *YansWifiPhyHelper*. Dita classe permite configurar o *ErrorRateModel*, que neste trabalho, é utilizado o modelo por *default ns3::NistErrorRateModel*. No modelo também são configurados os padrões de camada físicas como 802.11g, com as modulações e taxa de transmissão. As últimas aparecem como propriedades privadas, tendo acesso somente quando o dispositivo de rede é criado com a classe *ns3::WifiHelper*, como será descrito na sub-seção I.1.4.

Outros parâmetros de interesse configurados na classe são: o limiar de detecção para a recepção de sinais, o limiar de detecção de uso de canal, o ganho das antenas, a figura de ruído, e a potência de transmissão. Finalmente o código de configuração utilizado foi:

```
YansWifiPhyHelper wifiPhy = YansWifiPhyHelper::Default ();
wifiPhy.Set ("EnergyDetectionThreshold", DoubleValue(-96));
wifiPhy.Set ("CcaMode1Threshold", DoubleValue(-99));
wifiPhy.Set ("TxGain", DoubleValue(2.2));
wifiPhy.Set ("RxGain", DoubleValue(2.2));
wifiPhy.Set ("TxPowerStart", DoubleValue(15));
wifiPhy.Set ("TxPowerEnd", DoubleValue(15));
wifiPhy.Set ("RxNoiseFigure", DoubleValue(7));
wifiPhy.SetChannel (wifiChannel.Create());
```

I.1.3- Configuração da camada MAC – *WifiMac*

Neste passo é configurado o modo de trabalho “infraestrutura” na arquitetura WiFi e é habilitada a utilização do padrão 802.11e provendo serviços de QoS, com o uso da classe *ns3::WifiMacHelper*.

No ns3 existem três modelos que permite a configuração dos três elementos da topologia WiFi:

(i) *Access Point (AP*) - *ns3::ApWifiMac*;

(ii) non-AP Station (STA) - *ns3::StaWifiMac*;

(iii) STA em um *Independent Basic Service Set* (IBSS – conhecido como redes ad hoc) - *ns3::AdhocWifiMac*.

Esses três modelos herdam da classe *ns3::RegularWifiMac*, a qual expõe, entre diferentes configurações MAC, um atributo para suporte de QoS, que permite a configuração do padrão 802.11e/WMM (*Wi-Fi Multimedia*). Com os modelos MAC *QoS-enabled*, é possível trabalhar com 4 categorias de acesso (AC – *Access Categories*) diferentes, mostradas na Tabela I.1. Com o fim de determinar o AC apropriado, os pacotes encaminhados para a camada MAC, deverão ser marcados usando a etiqueta *ns3::QosTag*, fixando um identificador de tráfego (TID - *traffic id*) para o pacote. Logo os geradores de tráfego na camada de aplicação do *ns3*, tratados na Seção I.6, deverão ser modificados.

Tabela I.1: Categorias de Acesso para suporte de QoS [77].

AcIndex	**Categoria de Acesso (AC)**	**Descrição**
0	AC_BE	Melhor esforço – *Best effort*
1	AC_BK	*Background*
2	AC_VI	Vídeo
3	AC_VO	Voz

Finalmente a camada MAC foi configurada como segue:

```
Ssid ssid = Ssid ("wifi");
QosWifiMacHelper wifiMac = QosWifiMacHelper::Default ();
wifiMac.SetType ("ns3::ApWifiMac",
                        "Ssid", SsidValue (ssid),
                        "BeaconGeneration", BooleanValue (true),
                        "BeaconInterval", TimeValue (Seconds (2.5)));

//------------------Instalar no AP–WiFi -------------------------

wifiMac.SetType ("ns3::StaWifiMac",
                        "Ssid", SsidValue (ssid),
                        "ActiveProbing", BooleanValue (false));

//------------------Instalar no terminal móvel -------------------------
```

I.1.4- Criação do dispositivo WiFi – *WifiDevice*

Neste passo são configurados o padrão WiFi 802.11g com modulação ERP-OFDM (*Extended Rate PHY - Orthogonal frequency-division multiplexing*) na cada física e o algoritmo de controle de taxa de transmissão. Para a configuração utiliza-se a classe *ns3::WifiHelper* como descrito na seguinte sequência de código:

```
WifiHelper wifi = WifiHelper::Default ();
wifi.SetStandard (WIFI_PHY_STANDARD_80211g);
wifi.SetRemoteStationManager ("ns3::ConstantRateWifiManager",
                              "DataMode", StringValue("ErpOfdmRate54Mbps"),
                              "ControlMode", StringValue("ErpOfdmRate54Mbps "));

//------------------Configurar MAC do AP–WiFi -------------------------

NetDeviceContainer apDevice;
staDevice = wifi.Install (wifiPhy, wifiMac, wifiApNode);

//------------------ Configurar MAC do terminal móvel -------------------------

NetDeviceContainer staDevice;
staDevice = wifi.Install (wifiPhy, wifiMac, wifiStaNode);
```

I.2- REDE DE ACESSO LTE – MODELO FÍSICO E MAC

A rede de acesso LTE, foi simulada empregando o módulo LTE, também conhecido como LENA, que foi desenvolvido pelo *Centre Tecnológico de Telecomunicacions de Catalunya* (CTTC) [78]. O Modelo de simulação LTE-EPC (*Evolved Packet Core*), mostrado na Figura I.1, tem duas componentes principais:

- Modelo LTE: este modelo inclui o *stack* de protocolos de rádio LTE (como MAC, PHY). Esta entidade reside completamente dentro dos nós UE (*User Equipment*) e eNB (*Evolved Base Stations – eNodeB*).

- Modelo EPC: este modelo inclui as interfaces do núcleo da rede, protocolos e entidades da rede. Essas entidades e protocolos residem dentro dos nós SGW, PGW e MME, e parcialmente nos nós eNB [78].

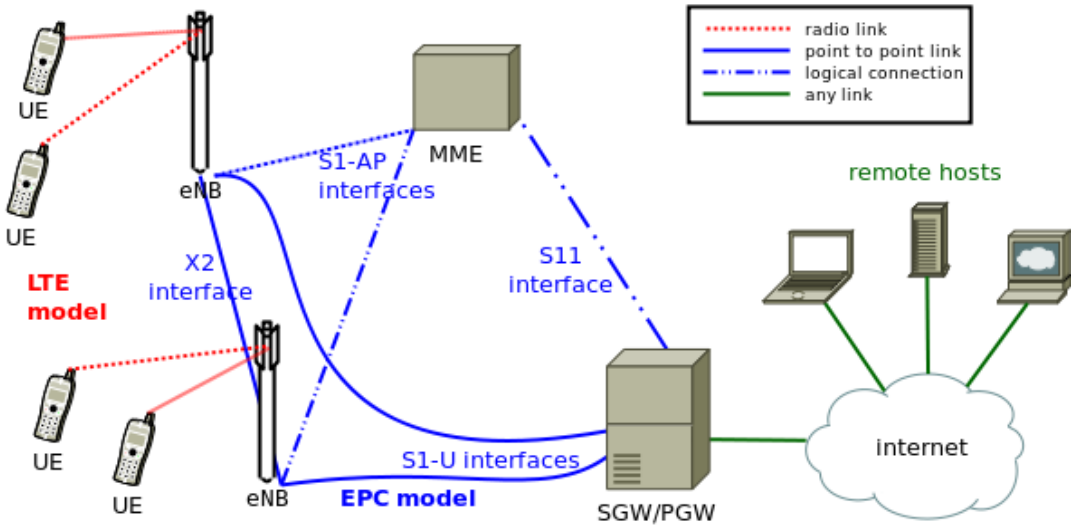

Figura I.1: Modelo EPC-LTE do LENA [78].

Esses modelos permitem a criação da infraestrutura 3GPP E-UTRAN e redes LTE. No modelo LTE são implementados os equipamentos de usuários (UEs) e as estações base (eNBs). Também são implementadas as entidades RRC (*Radio Resource Control*) para UE e eNB, esquemas de modulação para *dowlink*, filas MAC, instâncias RLC (*Radio Link Control*), gestão do Indicador de Qualidade de Canal (CQI – *Channel Quality Indicator*), agendamento de pacotes para *downlink* e *uplink*, entre outros recursos. Além disso, o módulo também inclui um modelo de camada PHY e de canal com perda de propagação EUTRAN exterior.

Na simulação foram utilizadas as classes *ns3::LteHelper* e *ns3::PointToPointEpcHelper*. A classe *LteHelper* é responsável pela criação de diferentes objetos LTE e a união destes para conformar o sistema LTE. A configuração global seguiria a seguinte composição:

- Canal de *downlink – Downlink spectrum channel:*
    - Modelo de perda do caminho – *Path loss model*: *FriisPropagationLossModel*
    - Modelo de desvanecimento – *Fading model:* nenhum modelo é utilizado

- Canal de *uplink – Uplink spectrum channel:*
    - Modelo de perda do caminho – *Path loss model: FriisPropagationLossModel*
    - Modelo de desvanecimento – *Fading model:* nenhum modelo é utilizado
- eNodeB node(s)
    - eNodeB device(s)
        - Antenna model: *IsotropicAntennaModel*, modelado na classe *ns3::AntennaModel*, onde o ganho das antenas é 0 dB em todas as direções.
        - eNodeB PHY (includes spectrum PHY, interference model, HARQ model):
            - DL *Bandwidth* = 6 em número de blocos de recursos;
            - UL *Bandwidth* = 6 em número de blocos de recursos;
            - Potência de transmissão = 30 dBm;
            - Figura de ruído = 5
        - eNodeB MAC
        - eNodeB RRC (includes RRC protocol): *true*, para sinalização RRC é empregado *LteRRCProtocolIdeal*
        - Scheduler: *PfFfMacScheduler*
        - Handover algorithm: *NoOpHandoverAlgorithm*
        - FFR (*frequency reuse*) algorithm: *LteFrNoOpAlgorithm*
        - ANR (*automatic neighbour relation*): *false*, não é preciso na simulação dado que só se tem uma eNB.
    - EPC related models (EPC application, Internet stack, X2 interface)
- UE node(s)
    - UE device(s)
        - Antenna model: *IsotropicAntennaModel*
        - UE PHY (includes spectrum PHY, interference model, HARQ model):
            - Potência de transmissão = 10 dBm;
            - Figura de ruído = 9
- EPC helper, utilizando na simulação *PointToPointEpcHelper:*
    - Interface S1-U

- Taxa de dados = 1Gbps
- MTU = 2000
- Delay = 0

- A interface X2 não é utilizada pois só existe um eNodeB na simulação.

Os canais são criados automaticamente: um para DL (*downlink*) e outro para UL (*uplink*), os dispositivos eNodeB são criados com o chamado a função *InstallEnbDevice*, enquanto os UE são criados chamando a função *InstallUeDevice.*

As entidades SGW/PGW, são consideradas como uma única entidade de rede, contida dentro das MAGs na arquitetura, e nos terminais multimodais foi instalado uma interface LTE. Para prover QoS na rede LTE são utilizados os indicadores de classe de QoS (QCI – *QoS Class Indicator*), mostrados na Tabela I.2, como descrito na especificação técnica 3GPP 23.203 Seção 6.1.7.2 [79], referenciado no ns-3.23 . Esses indicadores podem ser configurados através da classe *ns3::EpsBearer*. Na implementação original da classe ns3::*LteHelper*, dito parâmetro é configurado com o valor estático *EpsBearer::NGBR_VIDEO_TCP_DEFAULT* e não é permitida a mudança desde os *script* de simulação. Então foi adicionada na classe *LteHelper* uma função *Attach* que permite a configuração dos indicadores de qualidade, sendo posteriormente configurados na simulação dependendo dos modelos de tráfego utilizados.

Tabela I.2: Características QCI padrões [79].

QCI	**Tipo de Recurso**	**Prioridade**	**Atraso de Pacote**	**Taxa de erro e perda**	**Exemplo de serviços**	**enum QCI no NS-3**
1	GBR	2	100 ms	10^{-2}	Voz	GBR_CONV_VOICE
2		4	150 ms	10^{-3}	Vídeo Conferencia (*Live Streaming*)	GBR_CONV_VIDEO
3		3	50 ms	10^{-3}	Jogos Tempo Real	GBR_GAMING
4		5	300 ms	10^{-6}	Vídeo não vídeo conferencia (*Buffered Streaming*)	GBR_NON_CONV_VIDEO

5	Non-GBR	1	100 ms	10^{-6}	Sinalização IMS	NGBR_IMS
6		6	300 ms	10^{-6}	Vídeo (*Buffered Streaming*) TCP-based (e.g., www, e-mail, chat, ftp,p2p file sharing, progressive video, etc.)	NGBR_VIDEO_TCP_OPERADOR
7		7	100 ms	10^{-3}	Voz, Vídeo (*Live Streaming*) Jogos Interativos	NGBR_VOICE_VIDEO_GAMING
8		8	300 ms	10^{-6}	Vídeo (*Buffered Streaming*), TCP-based (e.g., www, e-mail, chat, ftp, p2p *file sharing, progressive video,* etc.)	NGBR_VIDEO_TCP_PREMIUM
9		9				NGBR_VIDEO_TCP_DEFAULT

A versão do módulo LTE contida no ns-3.23, não tem implementado na camada de rede o protocolo IPv6. As arquiteturas em questão são baseada em PMIPv6, protocolo que precisa suporte para IPv6. Logo foi feita uma ampliação no módulo LTE, para trabalhar com IPv6.

I.3- NÚCLEO DA REDE

Os link no núcleo da rede e na Internet foram criados a partir da configuração de canais csma, conforme alternativa apresentada para simulação no NS-3, na dissertação [80]. Foi utilizado o modelo *ns3::CsmaHelper* que construí um conjunto de dispositivos de rede *ns3::CsmaNetDevice*. Os parâmetros de cada link utilizado na simulação foram configurados como segue:

(i) CSMA da Internet (enlace entre LMA-CN na arquitetura centralizada, enlace entre MFR-CN na arquitetura parcialmente distribuída e enlace entre FMA-CN na arquitetura SIFDMM):

```
CsmaHelper csma;
csma.SetChannelAttribute ("DataRate", DataRateValue (DataRate ("100Mb/s")));
csma.SetDeviceAttribute ("Mtu", UintegerValue (1500));
csma.SetChannelAttribute ("Delay", TimeValue (MilliSeconds(0.5)));
```

(ii) CSMA entre a entidade centralizada e os *gateway* de acesso (LMA-MAG na arquitetura centralizada, MCF-MFR na arquitetura parcialmente distribuída):

```
CsmaHelper csma;
csma.SetChannelAttribute ("DataRate", DataRateValue (DataRate ("20Mb/s")));
csma.SetDeviceAttribute ("Mtu", UintegerValue (1500));
csma.SetChannelAttribute ("Delay", TimeValue (MilliSeconds(3.5)));
```

(iii) CSMA entre os *gateway* de acesso (MFR-MFR na arquitetura parcialmente distribuída e FMA-FMA na arquitetura SIFDMM):

```
CsmaHelper csma;
csma.SetChannelAttribute ("DataRate", DataRateValue (DataRate ("100Mb/s")));
csma.SetDeviceAttribute ("Mtu", UintegerValue (1500));
csma.SetChannelAttribute ("Delay", TimeValue (MilliSeconds(0.5)));
```

Com o fim de representar um número de saltos maior entre a entidade centraliza e os *gateway* de acesso foi penalizado o link entre elas na configuração (ii) acima descrita.

I.4- MOBILIDADE

Na simulação são utilizados dois cenários de mobilidade. Para os nós estáticos na rede (como AP, eNodeB, CN, e alguns UE) é usada a classe *ns3::ConstantPositionMobilityModel*. Para os dispositivos de usuário que vão-se movimentar é usada a classe *ns3::WaypointMobilityModel*, a que recebe como parâmetros as coordenadas do próximo ponto até onde o dispositivo deve movimentar-se (*waypoint*), e o tempo de chegada ao mesmo. Com esses parâmetros o nó vai se movimentar a uma velocidade constante entre dois *waypoint* consecutivos.

Os MN foram distribuídos aleatoriamente nas redes de acesso LTE e WiFi utilizando o modelo de localização *ns3::RandomDiscPositionAllocator*, que gera posições aleatória dentro de um disco centrado na posição do eNB ou AP, e um rádio que varia de 10-200 m. As posições aleatórias foram geradas utilizando *ns3::UniforRandomVariable*, com a semente e o número de corrida (*run*) por *default* do ns-3.23, definidos na classe *ns3::RngSeedManager*.

I.5- CAMADA DE REDE

As arquiteturas em estudo são baseadas no protocolo PMIPv6, logo a implementação da camada de rede foi feita tendo como base o modelo *ns3::Internet* e o modelo PMIPv6 implementado por Hyon-Young Choi [51],[81] e integrado neste trabalho ao ns-3.23.

O modelo PMIPv6 permite a configuração das entidades LMA e MAG definidas no protocolo, utilizando as classes *Pmipv6LmaHelper* e *Pmipv6MagHelper*. O módulo original só implementava entidades MAG como redes WiFi, logo foi ampliado para integrar suporte para MAG com redes de acesso LTE.

Cada arquitetura em estudo utiliza uma versão diferente no gerenciamento de mobilidade, o que leva a ter três implementações diferentes do PMIPv6. Os cenários simulados corresponde com as arquiteturas avaliadas no modelo analítico, e três versões do modelo PMIPv6 são utilizadas. Foram denominadas como *centralizedPmipv6*, *pdPmipv6* e *fdPmipv6*, correspondendo com a arquitetura centralizada [53], a arquitetura parcialmente distribuída [14] e a arquitetura SIFDMM, descrita no capítulo 4, respetivamente.

I.6- CAMADAS DE TRANSPORTE E APLICAÇÃO

O protocolo de camada de transporte utilizado foi configurado dependendo das aplicações empregadas para gerar os diferentes modelos de tráfego utilizados na simulação. O estudo em questão pretende avaliar o impacto dos parâmetros de QoS nas arquiteturas, sobre transmissões de vídeo. Logo, é gerado tráfego de vídeo e tráfego concorrente CBR e FTP, que são tratados na redes WiFi e LTE com diferentes serviços de QoS, como descrito na Tabela I.3. As características e configuração no NS-3 são descritas nas próximas seções.

Tabela I.3: Classificação de tráfego para habilitar QoS.

Tipo de tráfego	AC – Rede WiFi	QCI – Rede LTE
Vídeo	AC_VI	GBR_CONV_VIDEO
CBR	AC_VO	GBR_CONV_VOICE
FTP	AC_BE	NGBR_VIDEO_TCP_OPERATOR

I.6.1- Tráfego CBR

O modelo de tráfego CBR (*Constant Bit Rate*) simula tráfego VoIP (*Voice over IP*), caracterizado no NS-3 por dois estados: ON e OFF. O estado ON é dedicado ao tempo no qual os usuários falam e um fluxo constante de pacotes é transmitido em intervalos regulares. O estado OFF corresponde ao tempo onde os usuários permanecem em silencio e que provoca que nenhum pacote seja transmitido.

Os parâmetros utilizados nesta simulação são apresentados na Tabela I.4. Os intervalos de tempo referentes a ON e OFF são 0,352 e 0,650 segundos respectivamente. Foi empregado o codificador G.711 o qual não desenvolve nenhum sistema de compressão, garantindo a melhor qualidade de voz. A ausência de técnicas de compressão leva a altos requisitos de largura de banda. A classe *ns3::OnOffHelper* é utilizada para gerar esse tipo de tráfego, onde os parâmetros são configurados como a seguir:

```
PacketSinkHelper clientcbr ("ns3::UdpSocketFactory", Inet6SocketAddress(
ueAddress,dlPort));
client.Add(clientcbr.Install(args.ueNode));

OnOffHelper servercbr ("ns3::UdpSocketFactory", Inet6SocketAddress(ueAddress,dlPort));
servercbr.SetAttribute("PacketSize", UintegerValue(200));
servercbr.SetAttribute("DataRate", StringValue("64kb/s"));
servercbr.SetAttribute("OnTime", StringValue("ns3::ConstantRandomVariable[Con
stant=0.352]"));
servercbr.SetAttribute("OffTime", StringValue ("ns3::ConstantRandomVariable[Con
stant=0.65]"));
server.Add(servercbr.Install(args.cnNode));
```

Tabela I.4: Configuração do tráfego CBR.

Parâmetros	**Valor**
ON	0.352 segundos
OFF	0.650 segundos
Codec	G.711
Output	64 kbps
***Bandwidth* para ligações de voz**	200 bytes
***Bandwidth* na camada IP**	80 kbps por ligação
***Bandwidth* médio**	28.1 kbps

I.6.2- Tráfego FTP

O FTP (*File Transfer Protocol*), é um protocolo para a transferência de arquivos, que trabalha sobre o protocolo TCP, sendo categorizado como um tráfego de melhor esforço (*Best Effort Traffic*). Para sua implementação no NS-3 foi utilizada a classe *ns3::BulkSendApplication*, sendo configurada como a seguir:

```
uint8_t data_mbytes = 4;
PacketSinkHelper clientftp ("ns3::TcpSocketFactory", Inet6SocketAddress(ueAd
dress,dlPort));
client.Add(clientftp.Install(args.ueNode));

BulkSendHelper serverftp("ns3::TcpSocketFactory", Inet6SocketAddress(ueAddress,
dlPort));
serverftp.SetAttribute ("SendSize", UintegerValue (600));
serverftp.SetAttribute ("MaxBytes", UintegerValue(int(data_mbytes*1000000)));
server.Add(serverftp.Install(args.cnNode));
```

I.6.3- Tráfego de vídeo

Para a geração de tráfego de vídeo e posterior avalição de parâmetros de QoS e QoE, foi utilizada a ferramenta Evalvid [70], ver ANEXO I, para uma melhor compreensão. O Evalvid foi instalado no Ubuntu 12.04 e integrado no NS-3. O módulo de Evalvid no NS-3 gera os "logs" do vídeo transmitido e o vídeo recebido, necessários para a obtenção de métricas de referência completa com o Evalvid.

Na simulação foram utilizadas 6 sequências de vídeo com diferentes características temporais e espaciais, em formato YUV, 4:2:0 e resolução QCIF (176x144), disponíveis em http://trace.eas.asu.edu/yuv/. A Figura I.2 mostra uma classificação das sequências similar à dissertação [82], em sequências de movimento leve (*Slight Moviment* – SM), de movimento moderado (*Gentle Walk* – GW), e de movimento rápido (*Rapid Moviment* – RM).

O diagrama de fluxo da Figura I.3 mostra as etapas necessárias desde a obtenção do vídeo original em formato YUV até a obtenção do vídeo reconstruído (possivelmente danificado) no mesmo formato, para comparação utilizando métricas com referência.

A primeira etapa (Etapa 1) tem como objetivo a codificação do vídeo no formato YUV (QCIF (176x144)), para a codificação no padrão H.264, com a utilização da biblioteca *libx264*. Neste processo os seguintes parâmetros foram configurados: taxa de codificação (KBPS = 128), *frames* por segundos (FPS = 30) que é rápida o suficiente para causar a percepção de movimento continua ao observador (valor mínimo 24), e GOP (*Group Of Pictures* = 30) pelo que vamos ter um *frame* I cada 1segundo.

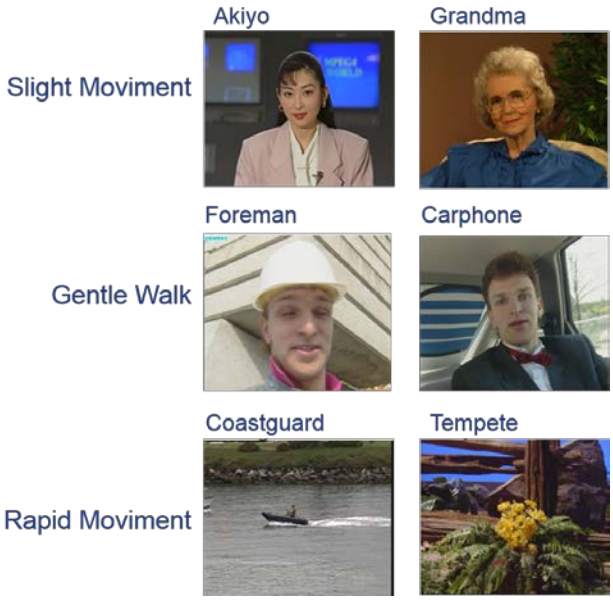

Figura I.2: Sequências de vídeo.

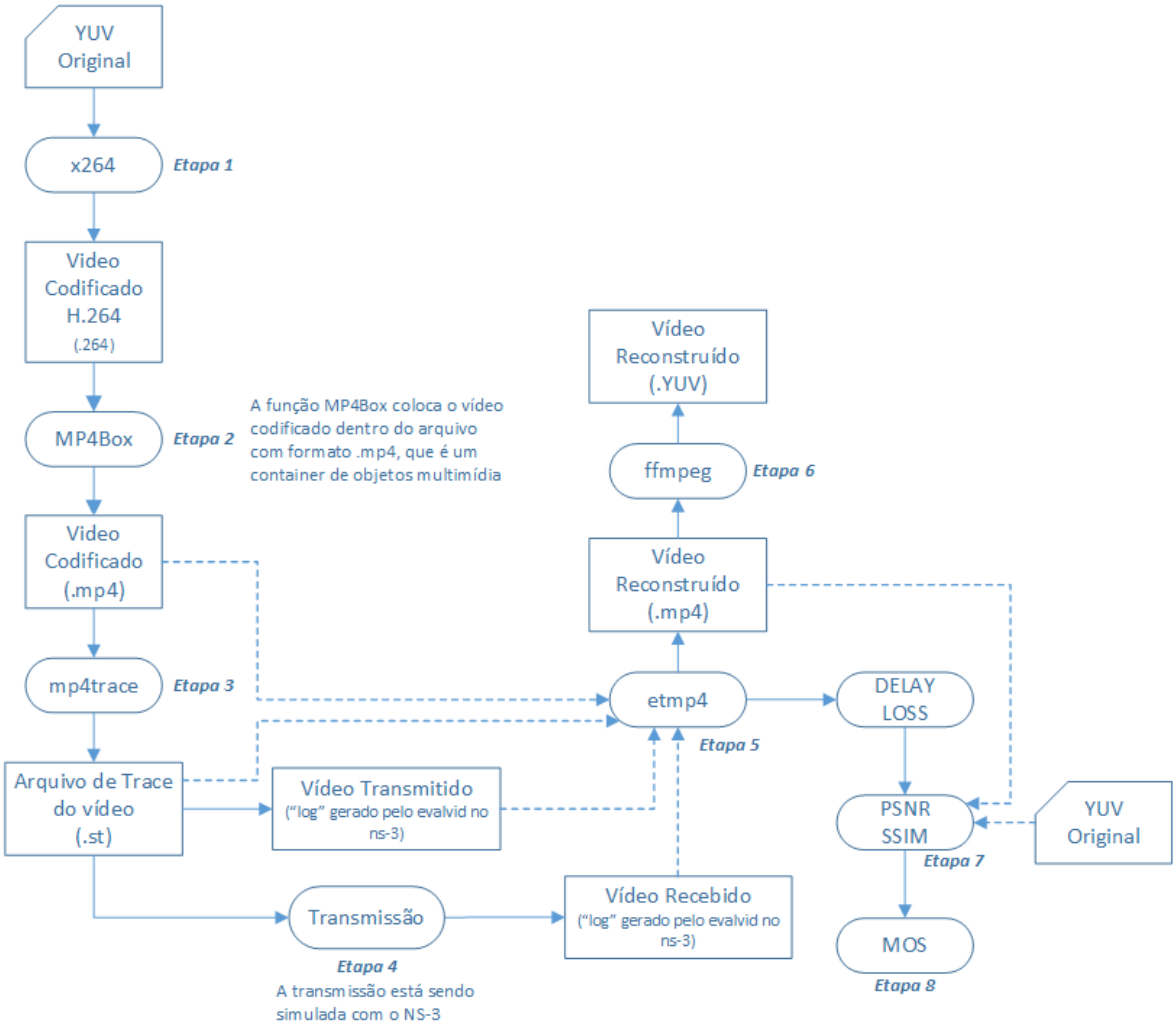

Figura I.3: Diagrama de fluxo do experimento.

A segunda etapa (Etapa 2) utilizando o comando MP4BOX, é gerado o vídeo em formato .mp4, permitindo o *hint track.* O *hint track* descreve como os quadros serão distribuídos em pacotes para a transmissão do vídeo junto ao protocolo RTP *(do inglês, Real-time Transport Protocol).* No processo são passados como parâmetros de configuração o FPS = 30 e o MTU (*Maximum Transmission Unit*) = 1024, valores comumente utilizados na literatura.

A terceira etapa (Etapa 3) utiliza o comando mp4trace do Evalvid para gerar o arquivo de *trace* .st do vídeo. O comando recebe como entrada o vídeo no formato mp4 e um endereço IP para gerar os instantes de tempo presentes no trace. Aos efeitos da simulação este endereço não tem utilidade, pelo que nós passamos o valor do host (127.0.0.1). O arquivo de *trace* está conformado pela quantidade de bytes necessários para transmitir cada *frame,* e em quantos pacotes eles devem ser divididos.

Na quarta etapa (Etapa 4) o vídeo e transmitido mediante o emprego do módulo Evalvid do NS-3, obtendo os arquivos "*logs"* referentes às informações sobre o envio dos pacotes e ao recebimento dos mesmos. O tamanho de cada pacote (*payload*) foi configura com o valor de 1014 coincidindo com a divisão feita no *trace* da etapa anterior. Foram utilizadas as classes *EvalvidClientHelper* e *EvalvidServerHelper* para a configuração dos aplicativos cliente e servidor instalados nos nós da simulação, como segue:

É importante lembrar, ainda, que, como o EvalVid abstrai o uso do protocolo RTP (do inglês, *Real-time Transport Protocol*), 14 bytes são adicionados em todos os pacotes como cabeçalho da camada de aplicação. O módulo cria uma aplicação, na qual cliente e servidor abrem uma conexão por meio de um *socket*. O cliente conhece o endereço IP do servidor a princípio, mas o contrário não ocorre. No início, o servidor lê o arquivo de *trace* do vídeo e armazena as informações de identificação, tipo de cada quadro, tamanho, número de pacotes UDP para cada quadro e *timestamp* em formas de *struct*. Então, fica em modo de escuta esperando por um pacote de requisição de vídeo, que deve ser enviado pelo cliente.

Quando o pacote de requisição de vídeo chega, o servidor identifica o endereço IP presente no campo do pacote e inicia o fluxo de transmissão para aquele endereço. Os quadros são divididos de acordo com o (*payload*) configurado e a leitura do trace já realizada pelo servidor. Os pacotes são, então, criados e enviados um de cada vez, utilizando as informações previamente armazenadas nas *structs*. Conforme cada pacote é montado e transmitido, os

dados do tempo de início desta transmissão, id e tamanho (do pacote, e não do quadro) são inseridos no arquivo de log do transmissor.

À medida em que o cliente recebe os pacotes, este monta o arquivo de log do receptor, com as informações de id e tamanho de cada pacote (mesmo do trace do transmissor) e o momento da recepção. O cliente não tem informação do final do fluxo, que eventualmente acaba quando o servidor termina de enviar os quadros, logo, ele ficará monitorando a possível chegada de novos pacotes até o final da simulação.

Na quinta etapa (Etapa 5) se reconstruí o vídeo a partir do uso do etmp4 e dos arquivos obtidos nas etapas 3 e 4 (*trace* e *logs*), assim como o vídeo mp4 gerado na etapa 2. Como resultado obtém-se um vídeo onde os pacotes perdidos foram substituídos por cero, e são gerados pelo comando Evalvid os arquivos que contem informação referente à perda e atrasos.

Na etapa 6 é decodificada a sequência de vídeo com a ferramenta ffmepg convertendo ao formato inicial (YUV). Nas etapas 7 e 8 são obtidos arquivos que contem informação referentes às métricas para avaliação da qualidade do vídeo PSNR, SSIM e MOS, utilizando os comandos *psnr, ssim* e *mos* do Evalvid.

Finalmente foram implementados diferentes *scripts* para automatizar as etapas descritas anteriormente, onde atendem-se cada uma das arquiteturas em estudo, fazendo simulações onde o vídeo é transmitido através de cada uma delas. Com isto pretende-se avaliar as influencias dos para metros de QoS, nos parâmetros de QoV.

I.7- Modificações realizas nos módulos do simulador NS-3

Foi necessária, na implementação dos cenários estudados, a modificação de alguns módulos do NS-3, tais como:

- As arquiteturas em estudo são baseadas no protocolo PMIPv6, precisando na camada de rede IPv6. Logo foi necessário, no módulo LTE acrescentar funções para suportar IPv6 na camada de rede, além de incluir um *call-back* com o fim de informar ao protocolo PMIPv6 a conexão dum novo usuário.
- No módulo WiFi foi acrescentado, na implementação do AP, um *call-back* com o fim de informar ao protocolo PMIPv6 a associação dum novo usuário.
- O modelo PMIPv6 implementado por Hyon-Young Choi [51],[81] foi integrado ao ns-3.23 e modificado para implementar cada uma das soluções em estudo. Foram

acrescentadas classes para a implementação de políticas de fluxos, que podem ser definidas desde os *scripts* na simulação. As políticas permitem o correto preenchimento das tabelas de roteamento das entidades de rede.

- O módulo Evalvid [70] foi integrado ao NS-3 e foram acrescentadas opções para a marcação dos pacotes utilizados na implementação de QoS na rede de acesso WiFi.
- No módulo de aplicação foram acrescentadas opções para o marcado dos pacotes utilizados na implementação de QoS na rede de acesso WiFi.
- Com o fim de medir a FHL foram incluídas algumas variáveis na interface WiFi, nos pacotes e nas aplicações.

As modificações anteriores serão disponibilizadas via repositório git, no Laboratório de Televisão Digital Interativa (LabTVDi), da Faculdade de Tecnologia da Universidade de Brasília.

APÊNDICE II - Carta de Aceitação de artigo no evento 7th International Conference on Information and Multimedia Technology (ICIMT 2015)

Acceptance Notification of ICIMT 2015

Paper ID: NB008

Paper Title: IP flow mobility management in mobile networks

Authors: Paulo Roberto L. Gondim and Yarisley Peña Llerena

To whom it may concern,

Congratulations! The review processes for 2015 7th International Conference on Information and Multimedia Technology (ICIMT 2015) has been completed. The conference received submissions from 34 different countries and regions, which were reviewed by international experts, and about 50 papers have been selected for presentation and publication. Based on the recommendations of the reviewers and the Technical Program Committees, we are pleased to inform you that your paper identified above has been accepted for publication and oral presentation. You are cordially invited to present the paper orally at ICIMT to be held in Barcelona, Spain, during December 21-22, 2015.

If the registration procedure **(available on the next page)** is completed before/on the set deadline, the paper will be published in **Journal of Advances in Computer Networks (ISSN: 1793-8244, DOI: 10.18178/JACN)**, Abstracting/Indexing: EI (INSPEC, IET), Engineering & Technology Digital Library, DOAJ, Electronic Journals Library, Ulrich's Periodicals Directory, International Computer Science Digital Library (ICSDL), ProQuest, and Google Scholar.

Please strictly follow the instructions of the format specified in the conference template while preparing your final paper. If you have any problem in preparing the final paper, please feel free to contact us via icimt@iacsit.org. For the most updated information on the conference, please check the conference website at http://www.icimt.org/. The Conference Program will be available at the website in early December, 2015.

We are looking forward to meeting you in Barcelona, Spain.

Yours sincerely,
ICIMT 2015 Organizing Committees
icimt@iacsit.org
Barcelona, Spain

APÊNDICE III - Artigo "IP flow mobility management in mobile networks", submetido no evento (ICIMT 2015)

IP flow mobility management in mobile networks

Yarisley Peña Llerena
University of Brasília/ Electrical Engineering, Brasília, Brazil
Email: yarisleyllerena@gmail.com

Paulo Roberto L. Gondim
University of Brasília/ Electrical Engineering, Brasília, Brazil
Email: pgondim@unb.br

Abstract— **One of the biggest challenges in today's cellular networks concerns the management of the exponential growth of data traffic. Data offloading is a promissory and low-cost solution for the reduction of overload in access networks. However, a new paradigm of hybrid networks that take advantage of alternative communication channels is required. This implies changes in the way data are processed and also affects the behavior of the existing network protocols. This manuscript reports on a study of the existing solutions for mobility management that support data offloading. These solutions add IP flow mobility, enable the maintenance of the continuity of user´s sessions and load balancing in the network and provide higher throughput to multi-mode terminals.**

Index Terms— **data offloading, mobility management, IP flow mobility**

I. INTRODUCTION

Presently, a large number of users accesses mobile networks. They not only use the voice and messaging services, but also participate in Internet sessions because of the IP support, provided by the latest cellular systems. Furthermore, the next generation networks (NGN), totally based on the IP protocol, will provide a wide range of broader and improved services, which include factors, such as global convergence, interoperability and mobility.

The data traffic largely increases in mobile networks, which may deteriorate the quality of services (product of the overload in the network) offered to users. Therefore, network operators aim at new technological solutions to increase the network capacity. Increases in the number of radio base stations and selective improvement in some of them, increase in the coverage by femtocells, or improvement in radio access technology to increase bandwidth are some of the solutions considered [1]. However, all of them require large investments and become costly.

Manufacturers of mobile terminals have already brought multimodal devices into the market, with different integrated radio interfaces, which has generated a heterogeneous environment regarding access technologies. These wireless networks coexist in the same geographical location, overlaying their coverage areas. A viable solution is to offload mobile data to a network with increased bandwidth and less congestion, for a balance in terms of network load and increase in the user´s quality of experience.

WiFi networks (Wireless Fidelity - 802.11) appear to be the best solution, because of the many access points (AP) available in homes, universities and stores, as well as the free use of spectrum. Moreover, operators can install their own AP in areas of high demand. APs operate in a different cellular band than the base stations and cause no interference, which represents an advantage over the use of small cellular cells.

Users should also move among different access technologies with no interruptions in the ongoing sessions and take advantage of a simultaneous connection of their interfaces in overlapping areas to increase the throughput. The main challenge is, when connecting to a different network, the user must receive a new IP address, in spite of maintaining the same identification, which can result in loss of the continuity of sessions.

An efficient IP mobility network support is required for an easier data offloading. The major protocols are MIP (Mobile IP) [2, 3], based on the host, and PMIPv6 (Proxy Mobile IPv6) [4], based on the network. Such protocols and those based on them(for example, FMIP --- Fast MIP, and HMIP – Hierarchical MIP) are centralized and all data traffic runs through the same network entity, also responsible for the mobility control. This feature leads to problems, such as single point of failure, scalability and reliability. Distributed mobility management (DMM) [5], which may be partially distributed (data plan distributed and centralized control plane) or fully distributed (both plans distributed) has arisen as a possible solution to those problems.

From the point of view of data offloading, the mentioned protocols provide options, so that all or none of the flows through the network can be moved. However, for a better load balancing in the network, options for the moving of certain flows of a given user, known as flow mobility, are required. The literature reports studies on the selective flow movement, based on centralized (i.e. [6-13]) or distributed solutions (i.e. [14-20]). This paper addresses a comprehensive survey of techniques for the mobility management that support data offloading through the selective movement of IP flows.

The remainder of the paper is organized as follows: Section II provides an overview of mobile data offloading techniques proposed by the 3GPP (3rd Generation

Partnership Project); Section III discusses user-based and network-based mobility management and data offloading proposals; Section IV addresses some problems and challenges; finally, Section V provides the conclusions.

II. Mechanisms of Data Offloading (3GPP)

The last releases of 3GPP have reported the great interest the offloading of mobile network traffic has drawn regarding the control of the overhead of network and delivery of a better QoE (Quality of Experience). One of the main solutions proposed is the moving of IP flows to femtocells or WiFi networks.

Three mechanisms, namely Local IP Access (LIPA), Selected IP Traffic Offload (SIPTO) and IP Flow Mobility (IFOM) have been proposed to provide efficient offloading and avoid the deterioration in the QoS offered to the end user. LIPA and SIPTO [21] are based on the unloading of the IP traffic before it reaches the core network and require an HeNB (Home eNode B) or femtocell. IFOM [6, 7] assumes the use of two radio interfaces in the UE (User Equipment), one for the cellular network and another for the WLAN (Wireless Local Area Network) (i.e. WiFi), capable of working simultaneously.

A. *LIPA - Local IP Access*

LIPA [21] is a method that enables UE, connected to an HeNB (in a femtocell), to transfer data to the LAN connected to the same HeNB, with no data passing through the macro cellular network. It also enables the UE to access any external network connected to the local one [22]. LIPA is applicable only to private networks and can be used in residential and corporative wireless networks for local routing and access to the Internet through a private gateway.

3GPP defines breakout as the location in the architecture where the data offloading should occur. In LIPA, the breakout point must be within the private network. In such an architecture, the network core is avoided and the user traffic is routed through the Local Gateway (LGW), located on the private network, as shown in "Fig. 1".

The 3GPP specification [21] refers to the set composed by HeNB and LGW as subsystem HeNB. Below is a brief description of the main elements of LIPA topology.

LGW is essentially a PGW (Packet Data Network-Gateway) that implements functions, such as IP address allocation for the UE, packet filtering and direct tunnel with the HeNB. It is logically located in the HeNB and connects to the private network devices directly through the SGi interface. It also exchanges information with the S-GW (Serving Gateway) of the macro network via the L-S5 interface. The core of the cellular network for the establishment of an LIPA PDN connection forwards signaling messages that enable the registration and management of the sessions, therefore, LGW routes data packets [22]. On the other hand, the UE can also keep its data sessions with the core

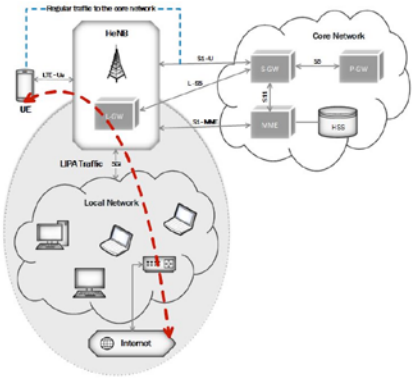

Figure 1. LIPA Architecture.

of the macro cell network, while the LIPA connection is being established.

The HSS (Home Subscriber Server) contains information from the local access authorization for each APN (Access Point Name) and subscriber. If local access is allowed, the MME (Mobility Management Entity) selects the address and enables a user plan to represent a direct path between the L-GW and the HeNB.

B. *SIPTO - Selected IP Traffic Offload*

In the SIPTO method [21], IP traffic portions, in an HeNB (in a femtocell) or eNB are offloaded to a local network for a reduction in the system's load. "Fig. 2" shows solutions for SIPTO over HeNB, where the point of breakout is located above the RAN (Radio Access Network). A set of gateways (S-GW and P-GW) and MME, located near the UE network connection point, are selected to offload the data. This solution enables a reduction in the amount of hops to the destination and lower-cost paths are selected.

The elements of the topology are HSS (Home Subscriber Server), MME and the gateways. HSS has similar functions in LIPA and keeps permission information for each subscriber and APN. The MME in SIPTO should perform the following functions: selection of a set of S-GW and P-GW, topologically close to the UE, authorization or refusal of the SIPTO traffic, based on HSS data, and making of decisions over the new location of a gateway due to the UE's mobility. S-GW and P-GW have the same functions [23].

C. *IFOM – IP Flow Mobility*

IFOM [6] is a special extension of IP mobility standardized by 3GPP. The pattern enables the movement of selective ongoing communication flows, from one access network to another, without interruption of the transferred flow while it keeps other flows in the current access network [6, 24]. Based on the specifications [6, 7], the UEs may simultaneously establish a connection with a 3GPP and a WiFi access and switch between different IP flows, in the same PDN connection, among the different access networks, as shown in "Fig. 3". The solutions are based on DSMIPv6 (Dual Stack MIPv6), PMIPv6 and GTP (GPRS Tunnelling Protocol) protocols, in which the IP address and continuity of the sessions are

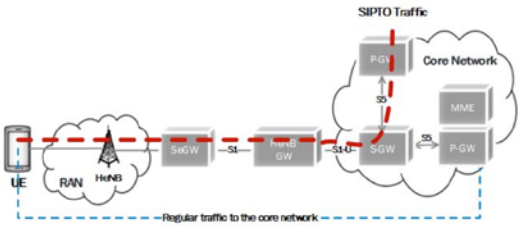

Figure 2. SIPTO Architecture in a femtocell.

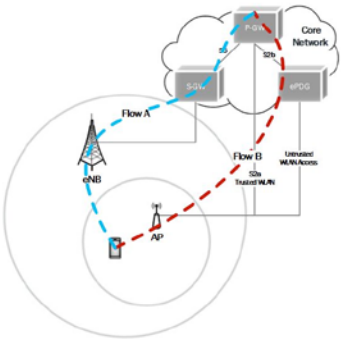

Figure 3. IFOM in the I-WLAN architecture.

preserved during the movement of IP flows between access systems.

III. MOBILITY MANAGEMENT

A key aspect of the mobility support is related to the participation of entities involved in the control of localization of mobile stations. Host-based mobility and network-based mobility are two possibilities commonly considered.

Host based mobility: the MN actively participates in the mobility management. It updates its location on the home network and connects with its partner if necessary, which impacts on their complexity, performance and power consumption negatively.

Network based mobility: the MN does not participate in the signaling related to mobility. The network is responsible for the IP mobility management. The entities in the network are the ones in charge of the follow-up of the MN moves and start the required signaling for updating its location, initializing the data offloading, if necessary, and maintaining the continuity of the sessions in progress [4].

The mapping of the MN's location can be either performed by only one network entity, or distributed among several anchors. Consequently, we have another important categorization, described as follows:

CMM (Centralized Mobility Management): a given mobility anchor maintains bindings of all MNs in the network. The data traffic is then encapsulated between the mobility anchor and the MN or its access router. This approach is usually implemented in centralized architectures where the location of the MN and the encapsulation of traffic must be processed by the central entity of the network (i.e. mobility anchor).

DMM (Distributed Mobility Management): the network´s functions, responsible for the mobility control, are separated from the nodes data plane. In network-based solutions, this approach can be categorized as partially or fully distributed [5, 25], according to the control plane distribution levels.

A. Host based mobility management

As first attempts towards solving the problem of user mobility at the IP layer, the MIP protocol [2, 3] was proposed by the IETF. In this protocol the MN actively participates in the mobility management, which causes long handover latency and packet loss, products of problems related to the wireless link. To date, not many solutions of this type have been developed.

Below we discuss three proposals based on the host and with support for data offloading reported in the literature. The solutions are classified into centralized and distributed, depending on the architecture type.

1) CMM

The 3GPP [6, 7] specifies a description for IP flow mobility between 3GPP and WLAN networks. The technical solution is based on the principles of work of the DSMIPv6 protocol and applicable to EPC (Evolved Packet Core) and I-WLAN (Interworking Wireless Local Area Network) architectures [27]. The UE uses both accesses for the same PDN connection and, subsequently, can add, modify, delete, or move the IP flows between access networks using DSMIPv6 messages. The offloading can also be initiated by either a centralized entity in the network, PGW, or the UE, whenever it is approved by both parties.

Although this solution enables network´s load balancing and provides higher throughput to the UE, it may cause problems of single point of failure and scalability, due to its central characteristic. Moreover, the UE sends signaling messages in wireless access, which can increase the loss and requires retransmissions and impacts on the process latency negatively.

2) DMM

Proposals based on the distributed mobility management have been developed for the solution of CMM problems. Hyouk L. Jong et al. [14] designed a host-based protocol for IP mobility support that uses anchors in the access network. The protocol does not adopt the centralized entity (HA - Home Agent), but uses distributed entities, called Access Mobility Anchors (AMAs). An AMA serves as an access router and allocates the network prefix to MN. The authors used the integration scheme of 3GPP access technologies and WLAN proposed by 3GPP [27], where the HA functionality is removed from the PGW and distributed in the SGW, Access Gateway (AGW, Access Gateway) and ePDG (Evolved Packet Data Gateway). The protocol built an offloading system in which the MN decides to move IP flows between their interfaces. The complexity of the MN can increase, as a product of the dynamic decision algorithms that usually require further processing. Therefore, proposals based on the network that minimize changes in the user device may be preferable.

B. *Network based mobility Management*

This subsection addresses the state of the art of solutions of network-based mobility management with data offloading possibilities. They were classified into centralized and distributed, according to the mobility management.

1) CMM

The 3GPP [7] presents study scenarios, applications and solutions for the EU with multiple interfaces, which can be simultaneously connected to a 3GPP access and a non 3GPP WLAN access (i.e. WiFi). The technical specification is based on PMIPv6 and GTP protocols and the host-based DSMIPv6 protocol. When network-based protocols are used, the UE does not participate in the signaling of the mobility management, however, it can request data offloading.

H.-Y. Choi et al. [8, 9] proposed a PMIPv6-based flow mobility scheme. The support for flow mobility is drawn according to a logic interface in MN and two components, namely Flow Manager Interface and Flow Binding Manager are introduced for the management of the mobility flow. The flow interface manager is placed in the logical layer interface on the MN and the flow binding manager is located in the network layer of the LMA (Local Mobility Anchor). These components, which work as a pair, establish flow policies used for the selection of the access technology by which the packets should be sent. The authors divided the flow mobility procedure into three cases: the MN establishes a new connection, decision of the LMA and the MN decision approved by the LMA.??? Therefore, both the user and the network can decide on moving a flow, but always with the consent of the latter.

K. Jinho et al. [10] proposed a mechanism optimized for seamless IP flow handover based on the FPMIPv6 (Fast PMIPv6) [28] protocol and initialization of the MN. It increases the performance of handover and enables a simultaneous use of multiple interfaces during the flow mobility. The above proposals highlight a possible selective IP flow mobility, however, they pose their own problems of centralized mobility management.

On the other hand, C. Makaya et al. [11] worked on vehicular communication networks and proposed a centralized scheme, called Multilink Striping Manager (MSM) that enables data offloading between different network access technologies. When a binding event is detected, the MN uses primitive MIH (Media Independent Handover) to start the application for IP flow mobility. The quality of the wireless connection and status of the network are monitored to help in the decision making process. R. I. Meneguette et al. [12] proposed an architecture for seamless flow mobility management based on the class of vehicle networking applications, with a network-based mobility management approach.

T. Melia et al. [13] focused on the design and implementation of flow mobility extensions for PMIPv6. They described the functional components required in the network to support intelligent driving traffic and minimize the impact on mobile devices and increase Quality of Experience (QoE). The network (particularly LMA) is the decision control entity that performs flow mobility based on the network operator's policies, which can react dynamically over the network load.

2) DMM

Product of increased traffic on mobile networks, today there is a preference for developing distributed IP mobility management solutions. ??? (qual é o produto?)This is due to scalability issues derived from centralized solutions. Below we discuss distributed proposals, which also support selective IP flow mobility.

Distributed mobility management solutions may include separate data plane and control plane as D-PMIPv6 [29], which divides the LMA into two entities, namely CLMA (Control plane Local Mobility Anchor) and DLMA (Data plane Local Mobility Anchor). CLMA manages signaling messages for the update binding, allocation of the HNP and maintenance of entries from BC (Binding Cache) for the MN, whereas DLMA establishes the tunnels with the MAGs for the forwarding of the data packets. X. Keqiang et al. [15] added IP flow mobility support to this solution using routing based on flows and a logical interface in the MN. They established a routing policy supporting differentiated services through the marked packages. Two approaches were proposed for the data-offloading start, one based on network changes and another based on wishes of the mobile terminal. The latter can reduce the scalability issues, processing delays and reliability of centralized proposals, however, it does not eliminate them. The CLMA and DMLA entities are centralized in each plane and must manage all network nodes and be a single point of failure.

Another network-based partially distributed scheme is proposed by K. Sun and Y. Kim [16]. The mobility management functions are relocated at the edge of the network. To provide routing and mobility management, network entities exchange signaling messages and make tunnels with no centralized entity. Nevertheless, to implement the flow scheme, the proposal uses a central entity called MCF (Mobile Control Funtion). It acts as a database, collecting information from all the MN in the network, including identifier, address and traffic types, and making offloading decisions. Although the MCF does not participate in the data forward, it may face scalability problems and is a single point of failure, which stores information from all the MN in the network. The authors proposed the existence of several MCFs in the network that share information.

On the other hand, P. Seite, P. Bertin, and J. Lee [17] described a partially distributed solution based on PMIPv6. The tunnels between MARs (Mobility capable Access Router) are used only for ongoing sections initiated before the handover. The data packets of the new sections opened at serving MAR will be directly routed. To optimize the routes, each previous MAR that still has the ongoing sections of the MN establishes a tunnel with the serving MAR to maintain the continuity of the sections. The control plane is centered by a centralized database, however, the interaction between MAR and the

database is not specified in the draft. The authors also describe the IP flow mobility support in multi-interface terminals that enables the data offloading in the solution.

M. Perras and J. Cartmell [18] consider mobility in a network of gateways of small cells that work in cellular and non licensed bands, combining an access point WiFi and base stations within a single device, called Converged Gateway (CGW). They describe methods to support local IP flow management between WiFi and cellular access technologies, as well as IFOM solutions based on CGW. The solutions can be extended to support distributed mobility management using several CGWs. This proposal involves the use of CGW devices, which can increase the cost of network deployment.

A scalable architecture for the network-based mobility management and an IP flow mobility management scheme, in the multi-access and multi-homing context, are presented in [19]. The architecture is not based on the mobility management protocols standardized by the IETF and consists of four functional entities, namely Mobility Information Control Server (MICS), Handover Control Agent (HCA), Point of Attachment (PoA) and MN. MICS is located in the core network and HCA is placed in the gateway in the access network and both manage mobility. The scheme for the IP mobility management uses the policy tables in MN and MICS, which determine the correct interface for the sending of a flow, based on the priority of access technology. The policy tables, as well as the list of priorities of the ATT (Access Technology Type) can be modified by either the user, or the network.

D. R. Purohith et al. [20] proposed another partially distributed solution. They designed an architecture, called Seamless Internetwork Flow Mobility (SIFM), that uses the concepts of PMIPv6 and SDN (Software Defined Networking). It defines a Flow Controller (FC) similar to the OpenFlow controller. When a switch receives a non-processed packet, it forwards it to the controller. The controller makes the routing decisions and instructs the switch on how to forward similar packages by adding an entry in the switch's flow table. FC performs functions related only to the mobility. PGW in LTE networks and WAG (Wireless Access Gateway) in WiFi networks act as OpenFlow hybrid switches that execute the relative mobility signaling on behalf of the EU. They follow the FC instructions when the MN moves from an LTE network to a WiFi network to provide seamless transition. FC makes the flow mobility decisions, establishes rules and informs other network entities.

All the above solutions offer the possibility of flow mobility, but in a partially distributed way. They require a centralized entity that processes information from all network nodes. Therefore, the literature lacks fully distributed solutions with support for selective IP flow mobility, which take advantage of new technologies to deliver the highest quality experience to users.

Table I summarizes the following characteristics of the above-mentioned studies: (i) mobility management: host-based or network-based and whether it is centralized or distributed, (ii) network entities that make up the architecture, (iii) which is the entity in charge of taking the data offloading decision, (iv) new messages defined in the architecture and (v) access technologies considered in the solution. Ainda está confuso!!

IV. PROBLEMS AND CHALLENGES

Host-based mobility management protocols face problems, such as high latency handover, packet loss and signaling overhead and also require changes in the MN IP stack to support them. Such requirements may increase its complexity and cause higher battery consumption and waste of air resources. Furthermore, the tunnels established between the HA and the MN increase the bandwidth constraints on the wireless link, the load processing on the mobile node, and the complexity of user devices. The MN should implement algorithms to decide when and what flows are exchanged between interfaces.

On the other hand, in approaches that use network-based mobility management, the network entity acts on behalf of the MN. The MN does not participate in any mobility-related signaling, therefore, the solutions are simpler and costs are lower.

Some modifications are required for user devices to support data offloading and treat concurrent flows from the same connection by two interfaces. These requirements will certainly increase their complexity. Exchanges of signaling messages in wireless access can worsen the congestion in the access network, which must be avoided. Therefore, operations related to mobility and offloading decisions may be performed on the network side.

Other problems faced by the hierarchical or centralized mobility management are:

Sub-optimal routing path: the traffic always crosses a central entity network, which leads to paths generally longer than those between the mobile node and its peers and causes unnecessary delays and waste of operator resources.

Scalability: the central entity of the network must have sufficient processing capacity and routing to handle the traffic of all nodes simultaneously.

TABLE I. COMPARISONS OF IP FLOW MOBILITY MANAGEMENT SCHEMES

Proposal	Mobility Management				Network Entities	Offloading Decision	New Messages	Access technology
	Host-based		Network-based					
	CMM	DMM	CMM	DMM				
3GPP [6, 7]	X		X		SGW, PGW, ePDG	UE, PGW	-	- 3GPP, - WLAN
L. Jong-Hyouk, et al. [14]		X			AMA	MN	Access Binding Update (ABU) Access Binding Acknowledgment(ABA)	- LTE, - WLAN (Untrusted)
H.-Y. Choi, et al. [8, 9]			X		MAG, LMA	MN, LMA	HNP Update Request (HUR) HNP Update Acknowledge (HUA)	- 3G, - WLAN, - WiMax
K. Jinho, et al. [10]			X		MAG, LMA	MN	Handover Initiate for Flow mobility (HIF) Handover Acknowledge for Flow mobility (HAF)	- 3G, - WLAN
C. Makaya, et al. [11]			X		OBU, RSU, MSM	MSM	MIH messages	- 3G/LTE (RSU - Roadside Units), - WLAN (OBU - Onboard Unit)
R. I. Meneguette, et al. [12]			X		MAG, LMA	MN,MAG,LMA	MIH messages	- LTE, - WLAN
T. Melia, et al. [13]			X		MAG, LMA	LMA	-	- 3G, - WLAN
X. Keqiang, et al. [15]				X	DLMA, CLMA, MAG	MN	Flow Move Update (FMU) RS message is extended to contain information about flow.	- 3G, - WLAN
K. Sun and Y. Kim [16]				X	MCF Mobile Function Routers (MFR)	MCF	Flow request message Flow response message	- 3G, - WLAN
P. Seite, et al. [17]				X	MAR	Not specified	-	Generic
M. Perras and J. Cartmell [18]			X	X	CGW	CGW	-	- 3G, - LTE, - WLAN
H.-B. Lee, et al. [19]				X	MICS HCA	MN, MICS	-	- 3G, - WiMax, - WLAN
D. R. Purohith, et al. [20]				X	FC Mobility Agent (MA)	FC	Flow Modification Message Port Status Update	- LTE, - WLAN

Reliability: centralized solutions are more prone to reliability issues, as a central entity is a single point of failure.

Distributed mobility management has emerged as an alternative for the solution of such problems. The solutions split the control and data planes and leave the forwarding functions at the edge of the network, so that optimum paths can be obtained for the data traffic. To the best of our knowledge, most proposals have added support for data offloading to partially distributed solutions. This sub-optimal routing path problem can be solved, however, the scalability problem and single point of failure and attack are still in the control plane, which worsens the network performance.

Research for the Internet of the future must provide data offloading on fully distributed or hybrid mobility management solutions, so that new technologies can be unified and provide better QoS to the largest number of users. Challenges, such as security mechanisms, implementation and evaluation of dynamic offloading algorithms, collection of suitable data from the networks and the MNs to support decision making, and performance when applications with QoS requirements are addressed must also be considered in the current and future solutions.

V. CONCLUSIONS

Mobile data offloading reduces network congestion at a minimal cost and enables users to experience high quality in access networks. This manuscript has addressed the state of the art of solutions for mobility management that support data offloading moving IP flows. The analyses revealed the current schemes implement offloading mainly in centralized or partially distributed architectures, therefore, problems of scalability and reliability still remain unsolved. In most proposals, the offloading decision is based on static policies, which enables no overall assessment of the system´s performance. New research results must be achieved for the solution of the mobility management problem and offer of best quality of services to multi-mode users. Such results include evaluation of the system by, for example, decision and selection dynamic algorithms, applications with QoS requirements, and aspects related to security.

ACKNOWLEDGMENT

The authors acknowledge the financial support from the National Council for Scientific and Technological Development (CNPq) to Yarisley Peña Llerena through the PEC/PG Program. Comments and suggestions from the anonymous reviewers are also gratefully acknowledged.

REFERENCES

[1] E. Bulut and B. K. Szymanski, "WiFi access point deployment for efficient mobile data offloading," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 17, 2013, pp. 71-78.

[2] C. Perkins, "RFC 5944: IP Mobility Support for IPv4," *Internet Engineering Task Force (IETF)* November 2010,

[3] C. Perkins, D. Johnson, and J. Arkko, "RFC 6275: Mobility support in IPv6," *Internet Engineering Task Force (IETF),* July 2011,

[4] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil, "RFC 5213: Proxy Mobile IPv6 " *Network Working Group,* August 2008,

[5] H. Chan, D. Liu, P. Seite, H. Yokota, and J. Korhonen, "RFC 7333: Requirements for Distributed Mobility Management," *Internet Engineering Task Force (IETF),* August 2014,

[6] 3GPP, "TS 23.261 V12.0.0, IP flow mobility and seamless Wireless Local Area Network (WLAN) offload; Stage 2 (Release 12)," 2014,

[7] 3GPP, "TR 23.861 V1.14.1, Network based IP flow mobility (Release 13)," 2015,

[8] H.-Y. Choi, S.-G. Min, and Y.-H. Han, "PMIPv6-based Flow Mobility Simulation in NS-3," in *Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2011 Fifth International Conference on*, 2011, pp. 475-480.

[9] H.-Y. Choi, S.-G. Min, Y.-H. Han, and R. Koodli, "Design and Simulation of a Flow Mobility Scheme Based on Proxy Mobile IPv6," *JIPS,* vol. 8, 2012, pp. 603-620.

[10] K. Jinho, Y. Morioka, and J. Hagiwara, "An optimized seamless IP flow mobility management architecture for traffic offloading," in *Network Operations and Management Symposium (NOMS), 2012 IEEE*, 2012, pp. 229-236.

[11] C. Makaya, S. Das, and F. J. Lin, "Seamless data offload and flow mobility in vehicular communications networks," in *Wireless Communications and Networking Conference Workshops (WCNCW), 2012 IEEE*, 2012, pp. 338-343.

[12] R. I. Meneguette, L. F. Bittencourt, and E. R. M. Madeira, "A seamless flow mobility management architecture for vehicular communication networks," *Communications and Networks, Journal of,* vol. 15, 2013, pp. 207-216.

[13] T. Melia, C. Bernardos, A. de la Oliva, F. Giust, and M. Calderon, "IP Flow Mobility in PMIPv6 Based Networks: Solution Design and Experimental Evaluation," *Wireless Personal Communications,* vol. 61, 2011/12/01 2011, pp. 603-627.

[14] L. Jong-Hyouk, K. D. Singh, J. M. Bonnin, and P. Sangheon, "Mobile Data Offloading: A Host-Based Distributed Mobility Management Approach," *Internet Computing, IEEE,* vol. 18, 2014, pp. 20-29.

[15] X. Keqiang, L. Jun, and W. Lei, "Design and Implementation of Flow Mobility Based on D-PMIPv6," in *Computational Science and Engineering (CSE), 2014 IEEE 17th International Conference on*, 2014, pp. 1344-1349.

[16] K. Sun and Y. Kim, "Flow Mobility Management in PMIPv6-based DMM (Distributed Mobility Management) Networks," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA),* vol. 5, 2014, pp. 120-127.

[17] P. Seite, P. Bertin, and J. Lee, "Distributed Mobility Anchoring," *IETF draft, draft-seite-dmm-dma-07,* February 6 2014,

[18] M. Perras and J. Cartmell, "Mobility for heterogeneous SmallNets," *Telecommunication Systems,* 2015/03/26 2015, pp. 1-18.

[19] H.-B. Lee, S.-G. Min, Y.-H. Han, K.-H. Lee, H.-W. Lee, *et al.*, "IP flow mobility scheme in scalable network-based mobility management architecture," *Telecommunication Systems,* 2015/04/02 2015, pp. 1-11.

[20] D. R. Purohith, A. Hegde, and K. M. Sivalingam, "Network architecture supporting seamless flow mobility between LTE and WiFi networks," in *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2015 IEEE 16th International Symposium on a*, 2015, pp. 1-9.

[21] 3GPP, "TR 23.829 V10.0.1, Local IP Access and Selected IP Traffic Offload (LIPA-SIPTO) (Release 10)," 2011,

[22] C. B. Sankaran, "Data offloading techniques in 3GPP Rel-10 networks: A tutorial," *IEEE Communications Magazine,* vol. 50, 2012, pp. 46-53.

[23] R. Gupta and N. Rastogi, "LTE Advanced - LIPA and SIPTO," *Aricent,* 2012,

[24] L. Bokor, J. Kovács, and C. A. Szabó, "A Home Agent Initiated Handover Solution for Fine-Grained Offloading in Future Mobile Internet Architectures: Survey and Experimental Evaluation," *Int. J. Agent Technol. Syst.,* vol. 6, 2014, pp. 1-27.

[25] C. Bernardos, A. d. l. Oliva, and F. Giust, "A PMIPv6-based solution for Distributed Mobility Management," *IETF draft, draft-bernardos-dmm-pmip-04,* March 5 2015,

[26] H. Soliman, "RFC 5555: Mobile IPv6 Support for Dual Stack Hosts and Routers," *Network Working Group* June 2009,

[27] 3GPP, "TS 23.234 V12.0.0, 3GPP system to Wireless Local Area Network (WLAN) interworking; System description (Release 12)," 2014,

[28] H. Yokota, K. Chowdhury, R. Koodli, B. Patil, and F. Xia, "RFC 5949: Fast Handovers for Proxy Mobile IPv6," *Internet Engineering Task Force (IETF)* September 2010,

[29] L. Yi, H. Zhou, D. Huang, and H. Zhang, "D-PMIPv6: A distributed mobility management scheme supported by data and control plane separation," *Mathematical and Computer Modelling,* vol. 58, 9// 2013, pp. 1415-1426.

APÊNDICE IV - Carta de Aceitação do Artigo no evento 2016 IEEE Wireless Communications and Networking Conference (WCNC)

Dear Mrs. Yarisley Llerena:

We are pleased to inform you that your paper #1570220933, 'Network Architecture to Support IP Flow Mobility and DMM (Distributed Mobility Management)', has been accepted for presentation at the 2016 IEEE Wireless Communications and Networking Conference (WCNC) -- http://wcnc2016.ieee-wcnc.org/

A total of 964 papers were submitted to IEEE WCNC 2016. Your paper is among the 467 papers accepted after careful consideration -- congratulations! Among the 467 accepted papers, a majority will be presented in a lecture style and some will be presented as poster papers. This distinction has no relationship with the quality of the accepted papers whatsoever. Accepted and presented papers will be published in the IEEE WCNC 2016 conference proceedings and submitted to IEEE Xplore® without any indication of the presentation mode.

The reviews are given below; they can also be found at http://edas.info/showPaper.php?m=1570220933. We recommend that you revise your paper to address the reviewers' comments and suggestions.

The Final Paper Submission Instructions are posted at http://wcnc2016.ieee-wcnc.org/authors.

To submit your Final paper, please go to EDAS, click on the "My Papers" button, and then upload your paper on the link provided. DO NOT go to Papers/Submit paper, because you will not find IEEE WCNC 2016 there. Please use IEEE PDF eXpress to check if your paper is IEEE Xplore® compatible as described at the website.

To be published in the IEEE WCNC 2016 conference proceedings and submitted to IEEE Xplore®, an author of an accepted paper is required to register for the conference at the full (member or non-member) rate and the paper must be presented by an author of that paper at the conference unless the TPC co-chairs grant permission for a substitute presenter before the conference opens. Non-refundable registration fees must be paid prior to uploading the final IEEE formatted, publication-ready version of the paper. For authors presenting multiple papers, one full registration is valid for up to three papers. Accepted and presented papers will be published in the IEEE WCNC 2016 conference proceedings and submitted to IEEE Xplore®.

Your final paper must be submitted by 12 January 2016. The maximum number of pages is seven (additional 100$ for the seventh page).

Congratulations once again for having your paper accepted to IEEE WCNC 2016, a flagship conference of the IEEE Communications Society. We look forward to seeing you in Qatar.

Sincerely,

Marwan Krunz, Chadi Assi, Sunghyun Choi, and Michele Zorzi
IEEE WCNC 2016 Conference - Network Track Chairs

APÊNDICE V - Artigo "Network Architecture to Support IP Flow Mobility and DMM (Distributed Mobility Management)", submetido no evento WCNC 2016

Network Architecture to Support IP Flow Mobility and DMM (Distributed Mobility Management)

Yarisley Peña Llerena
Electrical Engineering Dept.
University of Brasilia, UnB
Brasilia, Brazil
yarisleyllerena@aluno.unb.br

Paulo Roberto L. Gondim
Electrical Engineering Dept.
University of Brasilia, UnB
Brasilia, Brazil
pgondim@unb.br

Abstract- **The development of new applications and mobile devices has increased the data traffic in cellular networks. Offloading is a viable solution to alleviate network congestion and offer better quality of services. Some architectures have been proposed to offload data traffic and support IP flow mobility. Centralized or partially distributed architectures possess single point of failure and other problems, related to scalability and reliability. This manuscript proposes a fully distributed architecture, called Seamless IP Flow and Distributed Mobility Management (SIFDMM), with network-based mobility management, but assisted by the mobile terminal. Evaluation using analytical modelling revealed the proposed architecture leads to a better performance than centralized or partially distributed solutions based on PMIPv6.**

Keywords- handover, DMM, offloading, IP flow mobility

I. INTRODUCTION

Presently, a large number of users accesses mobile networks, which increases the data traffic in mobile networks, leading to possible overload in the network and deterioration of the quality of services offered. Therefore, network operators aim at new technological solutions to increase the network capacity. Increases in the number of radio base stations and selective improvement in some of them, increase in the coverage by femtocells, or improvement in radio access technology for larger bandwidth are some of the solutions considered [1]. However, all of them require large investments and become costly.

The design of multimodal devices with different integrated radio interfaces by manufacturers of mobile terminals has generated a heterogeneous environment regarding access technologies. These wireless networks coexist in the same geographical location, which causes their coverage areas to overlay. A viable solution is to offload mobile data to a network with increased bandwidth and less congestion, for a balance in terms of network load and increase in the user´s quality of experience.

Wireless networks based on IEEE 802.11 standard appear to be a good solution, because of the many access points (AP) available in homes, stores, etc., as well as the free use of spectrum. Operators can install their own AP in areas of high demand. APs operate in a different cellular band than the base stations and cause no interference, which represents an advantage over the use of small cellular cells.

Users should also move among different access technologies with no interruptions in the ongoing sessions and take advantage of a simultaneous connection of their interfaces in overlapping areas to increase the throughput. The main challenge is, when connecting to a different network, the user must receive a new IP address and maintain the same identification, which may result in loss of the continuity of sessions.

An efficient IP mobility network support is required for an easier data offloading. The major protocols are MIP (Mobile IP) [2, 3], based on the host, and PMIPv6 (Proxy Mobile IPv6) [4], based on the network. Such protocols are centralized and all data traffic runs through the same network entity, also responsible for the mobility control. This feature leads to problems, such as single point of failure, scalability and reliability. Distributed Mobility Management (DMM) [5], which may be partially distributed (data plan distributed and centralized control plane) or fully distributed (both plans distributed) has arisen as a possible solution to such problems.

From the point of view of data offloading, the above-mentioned protocols provide options, so that all or none of the flows through the network can be moved. However, for a better load balancing in the network, mechanisms for the moving of certain flows of a given user, known as flow mobility, are required. This paper proposes a new fully distributed architecture based on PMIPv6 and assisted by the mobile node (MN), for the IP flow mobility management in heterogeneous networks (WiFi and LTE), which enables data offloading.

The remainder of the paper is organized as follows: Section II discusses terminal-based and network-based mobility management, as well as data offloading proposals; Section III describes the SIFDMM architecture; Section IV analyzes the behavior of existing solutions in terms of handover flow latency; finally, Section V provides the conclusions.

II. RELATED WORK

This section describes the main characteristics of relevant related works, initially focused on centralized proposals for IP flow mobility.

3GPP [6] specifies a description for IP flow mobility between 3GPP and WLAN networks. The technical specification is based on PMIPv6, GTP and DSMIPv6 protocols and applicable to EPC (Evolved Packet Core) and I-WLAN (Interworking Wireless Local Area Network)

architectures. The UE uses both accesses for the same PDN connection and, subsequently, can add, modify, delete, or move the IP flows.

H.-Y. Choi et al. [7, 8] proposed a PMIPv6-based flow mobility scheme, considering two components, namely Flow Interface Manager and Flow Binding Manager. The Flow Interface Manager is placed in the logical layer interface on the MN and the Flow Binding Manager is located in the network layer of the LMA (Local Mobility Anchor). These components, which work as a pair, establish flow policies used for the selection of the access technology. Both user and network can decide on moving a flow, but always with the consent of the latter.

K. Jinho et al. [9] proposed a mechanism optimized for seamless IP flow handover based on the FPMIPv6 (Fast PMIPv6) protocol and initialization of the MN. It increases the performance of handover and enables a simultaneous use of multiple interfaces during the flow mobility. T. Melia et al. [10] focused on the design and implementation of flow mobility extensions for PMIPv6. They described the functional components required by the network to support intelligent driving traffic, minimize the impact on mobile devices, and increase Quality of Experience (QoE).

C. Makaya et al. [11] proposed a scheme called Multilink Striping Manager (MSM), which enables data offloading. When a binding event is detected, the MN uses primitive MIH (Media Independent Handover) to start the application for the IP flow mobility. The quality of the wireless connection and status of the network are monitored to help in the decision-making process.

R. I. Meneguette et al. [12] proposed an architecture for seamless flow mobility management based on the class of vehicle networking applications. It is based on PMIPv6 and MIH (Media Independent Handover – IEEE 802.21) protocols.

The above proposals enable data offloading, however they raise issues of centralized mobility management, namely sub-optimal routing, scalability and reliability.

Proposals based on the partially distributed mobility management have been developed to solve such problems. We begin presenting network-based proposals, using PMIPv6.

D- PMIPv6 [13] includes separate data plane and control plane dividing the LMA into two entities, namely CLMA (Control plane Local Mobility Anchor) and DLMA (Data plane Local Mobility Anchor). CLMA manages signaling messages, whereas DLMA establishes the tunnels with the MAGs for the forwarding of the data packets. X. Keqiang et al. [14] added IP flow mobility support to this solution. They established a routing policy to support differentiated services through the marked packets. Both CLMA and DMLA entities are centralized in each plane and must manage all network nodes, thus representing relevant points of possible failure.

K. Sun and Y. Kim [15] propose a scheme where the device location and information flow are managed by a centralized control function, whereas data packets are forwarded by a distributed MFR (Mobile Function Routers). The central entity is called MCF (Mobile Control Function) and acts as a database, collecting information from all the MN in the network and making offloading decisions.

P. Seite, P. Bertin, and J. Lee [16] described a solution where the tunnels between MARs (Mobility capable Access Routers) are used only for ongoing sections initiated prior to the handover. The data packets of the new sections opened at serving MAR will be directly routed. To optimize the routes, each previous MAR that still has the ongoing sections of the MN establishes a tunnel with the serving MAR to maintain the continuity of the sections. The control plane is centered, however, the interaction between MAR and the database is not specified in the draft.

A scalable architecture for the network-based mobility management and an IP flow mobility management scheme are presented in [17]. The architecture is not based on the mobility management protocols standardized by the IETF.

D. R. Purohith et al. [18] proposed another partially distributed solution, considering an architecture, called Seamless Internetwork Flow Mobility (SIFM), which uses the concepts of PMIPv6 and SDN (Software Defined Networking and defines a Flow Controller (FC) similar to the OpenFlow controller. The controller makes the routing decisions and instructs the switch on how to forward similar packets. FC performs functions related only to the mobility.

The above solutions enable flow mobility, but in a partially distributed way, requiring a centralized entity that processes information from all network nodes.

L. Jong-Hyouk et al. [19] designed a host-based protocol for IP mobility support which uses distributed entities, called Access Mobility Anchors (AMAs). An AMA serves as an access router and allocates the network prefix to MN. The protocol built an offloading system in which the MN decides to move IP flows between their interfaces. The complexity of the MN can increase, as a product of the dynamic decision algorithms that usually require further processing.

The literature lacks fully distributed solutions with support for selective IP flow mobility that take advantage of the heterogeneous access network and multimodal nodes, to deliver the highest quality experience to users. This paper proposes a fully distributed architecture based on PMIPv6, where the MN assists the network entities in the handover process and in the IP flow mobility.

III. Seamless IP Flow and Distributed Mobility Management (SIFDMM) Architecture

This section presents the design, components and operation of the proposed architecture "Seamless Mobility of Flow Distributed Management" for LTE and WiFi networks, as shown in Figure 1. The main components are the flow and mobility anchor (FMA), mobile node (MN), and corresponding node (CN). SIFDMM is based on the PMIPv6 protocol [4]; it is a network-based architecture, but assisted by the user, which enables the treatment of the distributed control plane. Therefore, the mobility management is fully distributed.

Flow and Mobility Anchor (FMA)

FMA is an anchor point and access router of similar functions to the MAG and LMA entities of the PMIPv6 protocol [4]. It assigns the network prefix (HNP - Home Network Prefix) to the MN, provides internet connection, detects the MN movement between different access networks and maintains the continuity of the open sections. Based on local data structures and information on network conditions, it must implement algorithms for the handover or data offloading decision. Once the necessity of moving

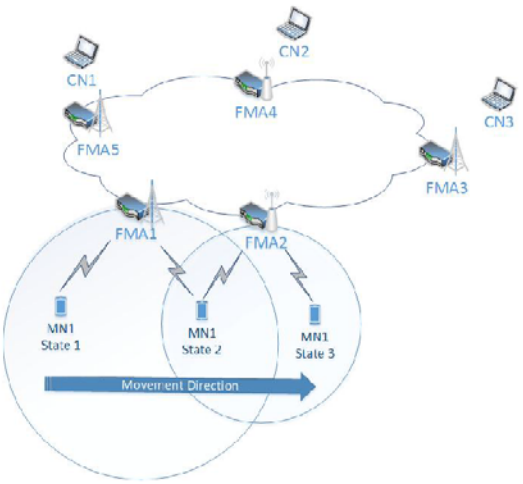

Figure 1. SIFDMM architecture for LTE and WiFi networks.

some flow has been detected, FMA exchanges signaling messages, so that tunnels are created, which enables the MN to keep its active sections.

When an MN leaves the network access of an FMA, it must store the received packets for that mobile for some time. During this period, a message must be obtained from another FMA, but rather packets sent to the MN will be discarded. The control messages used are similar to PBU and PBA of PMIPv6, including flow mobility options and identification of the MN state.

The FMA can act as a home and/or visited flow and mobility anchor (H-FMA and V-FMA , respectively) for a given MN with different HNPs. H-FMA denotes the flow and mobility anchor that has been previously visited by the MN and has anchored an IP address used by one or more active MN flows. A given MN may have several H-AMF. Term V-FMA refers to the FMA that is serving the MN. It handles the signaling related to mobility for the MN and establishes tunnels with H-AMF to maintain ongoing sessions.

Mobile Node (MN)

MN is a multi-user device that can work in either "a weak host" [20], or "a logical interface" [21]. It processes packets coming through an interface with a different IP address associated with it and stores information of the flow to send to the correct interface. However, it does not make flow movement decisions. When it receives a flow for a different interface, it updates its data structure following the decision of the network. It must also keep the PCoA of the last linked FMA for each interface to help the network entities in the handover or data offloading process.

Correspondent Node (CN)

CN is the end device from the communication established with the MN. It may either display the same characteristics and functions of an MN, or be any end-user device.

During the process of attachment to the FMA, the MN should communicate the PCoA on the last FMA associated with each of its interfaces. The information is provided by an option in RS messages. If it is the first connection of MN in the domain, the option can be zero and is ignored in FMA. The MN will receive the PCoA in the RA message and associate with the interface.

The regular IPv6 routing is applied to the SIFDMM solution when an IP communication has been initiated. When the MN connects to a new FMA, the traffic anchored in a previous FMA through the tunnel between them while new open sessions will be routed directly. In case of an offloading process, only the selected flows pass through the tunnel. However, in the handover process, all flows will be forwarded to the serving FMA. In such cases, the serving FMA plays the role of both H-FMA for sessions beginning with the HNP assigned by it and V-FMA for sessions initiated in the previous FMAs.

Therefore, FMAs may offload data, if necessary, as shown in Figure 2. The flow and mobility anchors can act as V-FMA for the data flow involved in the offloading process. The tunnels are always established between the FMA where the flow started (H-FMA) and the serves the MN (V-HMA), which guarantees the optimization of routes, as shown in Figure 3.

IV. PERFORMANCE EVALUATION

This section addresses an analytical and numerical evaluation of the performance of three proposals for mobility protocols that support selective IP flows mobility. It is assumed there will be a data offloading immediately after the connection has been established with a different interface. We define the flow handover latency (FHL), as the time elapsed between the start of the connection establishment with an access router and the arrival of the first packet through the new link. The FHL is expressed as follows:

$$FHL(.) = T_{L2} + T_{MD} + T_{LU} + T_p \quad (1)$$

T_{L2}: latency of the link layer attachment
T_{MD}: movement detection latency
T_{LU}: update location latency
T_p: latency of the delivery of the first packet

The analysis will consider a centralized solution, [7], a partially distributed [15] and a fully distributed (SIFDMM) based on PMIPv6, since they hey represent the general operation and enable an evaluation of the performance of solutions of each type.

A. Network Model

We model a network with M access routers (MAG, MFR or FMA, depending on the architecture) and M-1 areas of overlap of access networks of different technologies, as shown in Figure 4. N mobile nodes, with two network interfaces, are uniformly distributed and it is assumed on average K percent of the MNs enters an overlapping area simultaneously. Therefore, $(K * N)/100$ MNs will make flow handover simultaneously.

Denote by $h_{x,y}$ the average distance in hops between two network entities x and y, assuming it is symmetrical ($h_{x,y} = h_{y,x}$). We define a network scale ξ the ratio between the number of hops between two AR and the number of hops between an AR and the centralized entity (LMA or MCF), like in [22] $\xi = h_{ar,ar}/h_{ar,lma}$.

The general model of the network and the exchange of messages required for each proposal are shown in Figure 4 and 5, respectively.

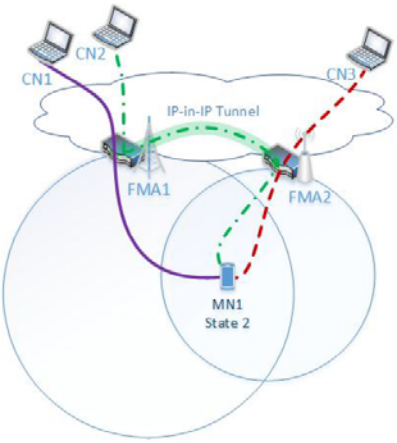

Figure 2. Data offloading of the LTE access network for the WiFi access network.

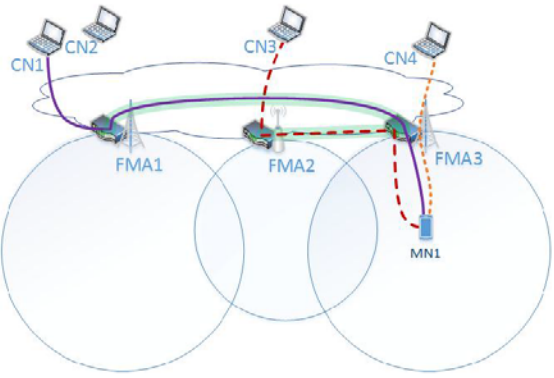

Figure 3. Handover to FMA3.

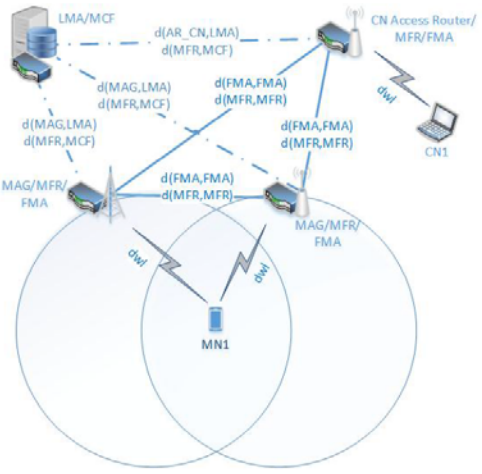

Figure 4. Network model for evaluation.

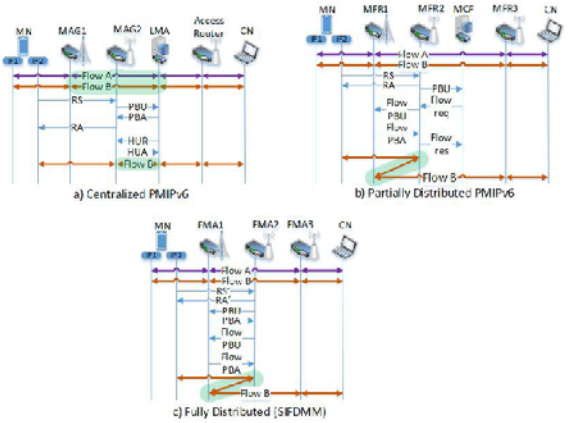

Figure 5. Exchange of messages for data offloading.

B. *Delay model in the anchors entities*

We consider a scenario where the MN may have multiple active sessions simultaneously by different interfaces. The queues at the anchors are modeled as an M/M/1 system. We assume the arrival of data packets to the MN follows a Poisson process with mean rate λ. The average arrival rate of data packets in the anchor entity (λ_d) is obtained by (3).

$$\lambda_d = \sum_{i=1}^{N} \lambda_i = N\lambda \qquad (2)$$

where N is the number of MN managed by the FMA entity. We assume the processing times of data and control packets in each anchor node follows a Poisson distribution with average rates μ_d and μ_c, respectively. Therefore, the average time spent in the system (T_s, queuing and processing time in the anchors) can be expressed as:

$$T_{sd} = \frac{1}{\mu_d - \lambda_d} \qquad (3)$$

$$T_{sc} = \frac{1}{\mu_c - \lambda_c} \qquad (4)$$

where λ_c is the average rate of arrival of control packets and T_{sd} **and** T_{sc} are the average time spent of a data and control packet respectively, in each anchor node of the network.

C. *Delay model in the wired and wireless links*

The time of delivery of a packet of wireless and wired links includes transmission and propagation times [22] . We assume the wired links are reliable and do not need retransmission. Therefore, the delay in the sending of a packet of S_p size, from an x node to a y node is expressed by $d_{x,y}(S_p) = h_{x,y} * (S_p/B_{wd} + L_{wd})$, where B_{wd} and L_{wd} are the bandwidth and propagation delay of wired links, respectively.

The wireless links can introduce errors in the transmission of a packet and need several retransmissions. Let P_f be the probability of failure of the wireless link. With this the delay of a packet with S_p size in a wireless link is $d_{wl}(S_p) = (S_p/B_{wl} + L_{wl}) * (1/(1 - P_f))$ [22], where B_{wl} and L_{wl} are the bandwidth and the propagation delay in the wireless link, respectively.

D. Analyses of the flow handover latency

When the MN arrives at an overlapping area of two access networks with different technologies, it must perform the attachment link layer, by the new interface. The latency of this process (T_{L2}), does not depend on the mobility scheme used, and it is considered to be equal in the three proposals analysis.

The movement detection latency (T_{MD}), is made up of exchange of Router Solicitation (RS) and Router Advertisement (RA) messages, between the MN and the new AR on the wireless link and is expressed as:

$$T_{MD}^{PMIPv6} = T_{MD}^{PMIPv6-PD} = 2d_{wl}(S_{RS}) \quad (5)$$

$$T_{MD}^{SIFDMM} = 2d_{wl}(S_{RS'}) \quad (6)$$

The update location latency (T_{LU}) is divided in the latency of transmission (T_{tc}) and latency in the anchor entity (T_{sc}), of the control messages.

$$T_{LU} = T_{tc} + T_{sc} \quad (7)$$

The control plane of PMIPv6 and PMIPv6_PD proposals is centralized. Soon the centralized network entity (LMA/MCF) will have to process the requests of the $(K * N)/100$ MNs arriving the overlapping areas simultaneously. It is $\lambda_c = (K * N)/100$. With this, T_{sc} is expressed as:

$$T_{sc}^{PMIPv6} = T_{sc}^{PMIPv6-PD} = \frac{1}{\mu_c - \frac{K * N}{100}} \quad (8)$$

The proposal SIFDMM the control plane is distributed as soon as FMA will have to process the requests of the $(K * N)/(100 * (M-1))$ MNs arriving in overlapping area simultaneously being $\lambda_c = (K * N)/(100 * (M-1))$. So it is expressed as:

$$T_{sc}^{SIFDMM} = \frac{1}{\mu_c - \frac{K * N}{100 * (M-1)}} \quad (9)$$

Finally the T_{LU} of solutions in analyzes is expressed as:

$$T_{LU}^{PMIPv6} = 2d_{mag,lma}(S_{PBU}) + 2d_{mag,lma}(S_{flow}) + T_{sc}^{PMIPv6} \quad (10)$$

$$T_{LU}^{PMIPv6-PD} = d_{mfr,mcf}(S_{PBU}) + 2d_{mfr,mcf}(S_{flow}) + 2d_{mfr,mfr}(S_{flow}) + T_{sc}^{PMIPv6-PD} \quad (11)$$

$$T_{LU}^{SIFDMM} = 2d_{fma,fma}(S_{PBU}) + 2d_{fma,fma}(S_{flow}) + T_{sc}^{SIFDMM} \quad (12)$$

The latency of sending the first packet for the new interface (T_p), is divided in the transmission latency (T_{td}) and latency in the anchor entity (T_{sd}), of the data packet.

$$T_p = T_{td} + T_{sd} \quad (13)$$

The PMIPv6 proposal data plane is centralized. Soon the LMA should process the data packets of the N mobile node on the network. With this, the average rate of packet arrival in the LMA is $\lambda_d = N\lambda$. So T_{sd} is expressed as:

$$T_{sd}^{PMIPv6} = \frac{1}{\mu_d - N\lambda} \quad (14)$$

In PMIPv6-PD and SIFDMM solutions, the data plane is distributed. Therefore, assuming that the MNs are uniformly distributed, each anchor node of the network, it should process the packets of $\frac{N}{M}$ MNs . With this, the average rate of arrival of packets on the FMA/MFR is $\lambda_d = \frac{N}{M}\lambda$. So T_{sd} is expressed as:

$$T_{sd}^{PMIPv6-PD} = T_{sd}^{SIFDMM} = \frac{1}{\mu_d - \frac{N}{M}\lambda} \quad (15)$$

Finally the T_p of solutions in analyzes is expressed as:

$$T_p^{PMIPv6} = 2d_{wl}(S_{data}) + d_{ar,lma}(S_{data}) + d_{lma,mag}(S_{data} + \tau) + T_{sd}^{PMIPv6} \quad (16)$$

$$T_p^{PMIPv6-PD} = 2d_{wl}(S_{data}) + d_{mfr,mfr}(S_{data}) + d_{mfr,mfr}(S_{data} + \tau) + T_{sd}^{PMIPv6-PD} \quad (17)$$

$$T_p^{PMIPv6-PD} = 2d_{wl}(S_{data}) + d_{fma,fma}(S_{data}) + d_{fma,fma}(S_{data} + \tau) + T_{sd}^{SIFDMM} \quad (18)$$

E. Numerical results

We studied the impact of some parameters in the flow handover latency at the solutions based on PMIPv6. The default values of the system parameters are assumed to be as follows.

Network model: M = 4, N = 450, K = 40%, $\mathbf{h_{fma,fma} = h_{mfr,mfr} = \sqrt{M} = 2\ hops}$ [22, 23], $\mathbf{h_{ar,lma} = h_{mfr,mfc} = 10\ hops}$. Usually the average number of hops between two neighboring AR is smaller than AR and a centralized entity. This means that the network scale is $\xi \leq 1$. In the literature the default value is generally considered to be between 0.2 and 0.5 [22, 23]. In our case, the default value of the network scale is $\xi = 0.2$. However, we investigate later the impact of this parameter on different costs.

Delay model in the anchor entity: $\lambda = 0.01$ pct/ms, $\mu_d = 5$ pct/ms, $\mu_c = 0.360$ pct/ms.

Delay model in the wired and wireless links [22, 23]: $B_{wd} = 100$ Mbps, $L_{wd} = 0.5$ ms, $B_{wl} = 10$Mbps, $L_{wl} = 2$ms, $P_f = 0.5$, $T_{L2} = 2$ ms, $\tau = 40$ bytes, $S_{rs} = 50$ bytes, $S_{rs'} = 70$ bytes, $S_{PBU} = 76$ bytes, $S_{flow} = 100$ bytes, $S_{data} = 400$ bytes.

The flow handover latency may be affected by the amount of MNs, which enter within overlapping area simultaneously and makes data offloading. We investigated the impact of this, varying the parameter K from zero to 100%. The Figure 6 shows that when more than half of MNs are doing requests simultaneously, the FHL of proposals in

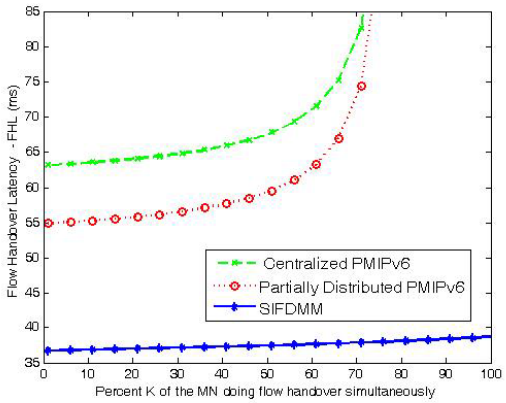

Figure 6. Impact of K parameter in the FHL.

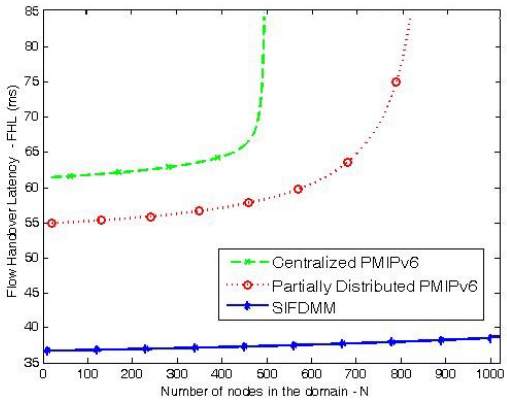

Figure 7. Impact of amount of MNs in the FHL.

which the control plane is centered, starts to increase. This is due to delays in the queues in the mobility anchors.

When the K parameter passes the 70%, the system gets out of balance, and centralized entities will be congested, which can cause timeouts in very long queues and packet loss, should not be considered infinite buffer. However the distributed proposal maintains a stable performance without greatly increase the FHL. This is because in the control plan there is no centralized entity for flow mobility management. Each FMA meets the requests of MN that are in its coverage area.

The increased quantity of MNs in the domain also affects FHL. For the analysis varies the amount of MNs from 10 to 1000. We can observe in Figure 7 that the LMA is the first entity to be congested. The proposal PMIPv6-PD even needing greater control signaling, has less FHL that the centralized solution, because the data plane is distributed and can manage larger amount of MNs. But the proposal SIFDMM always presents the smaller times, because it does not increase the control signaling, and the increase in the size of RS and RA messages do not have a significant impact on FHL.

Another parameter that may affect the FHL is the transmission of packets over the links. Therefore, we investigated the impact of the probability of failure in the wireless link P_f and network scale ξ, in the FHL. First we varied P_f from 0 to 0.8. The Figure 8 shows the variation as a function of FHL. The results show that the wireless link, has little impact on the three proposals. This is because the three are based on the network, and the MN has no involvement in signaling relative to the mobility management. Although with the increase of P_f increases the FHL product of retransmissions in wireless link.

The Figure 9 shows the impact of ξ varying from 1/10 to 1.The network scale has no impact on PMIPv6 because it is a centralized architecture. In contrast, the distributed solutions have better performance when ARs are close to each other, which is the typical case.

The results show that the distributed architectures outperform the centralized in most cases. Although, when the distance between the ARs is large, the FHL increases in distributed solutions, which may be much larger than in centralized.

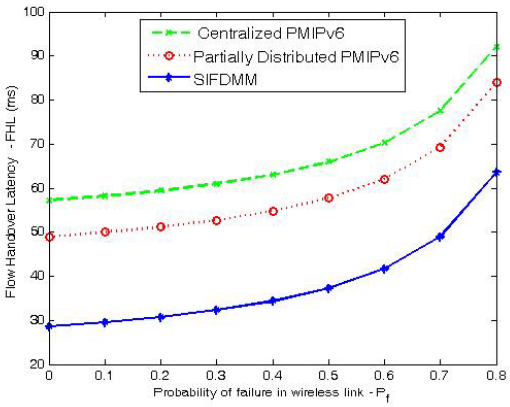

Figure 8. Impact of Pf in the FHL.

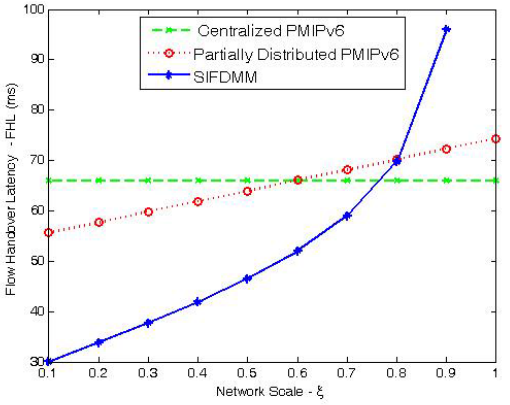

Figure 9. Impact of the network scale in the FHL.

V. Conclusions

In this paper, we propose a new architecture for mobility management and data offloading. It is network-based, assisted by the terminal and fully distributed. The solution supports IP flow mobility and thus data offloading. In order to evaluate the performance, we used an analytical model of the flow handover latency for centralized, partially distributed and SIFDMM solutions. The results show that SIFDMM is capable of handling larger amount of MNs without congestion in the mobility anchor entities. It also has a lower flow handover latency in most scenarios. With the proposed solution, it is possible to obtain higher throughput and load balance, given the possibility of MN to be connected to two AR at the same time. In addition, it does not have the problems of reliability, scalability and single point of failure.

In future work, mechanisms to determine when to move a flow can be studied, as well as the dynamic update of the flow tables rules based on the network status and user preferences. Also, we intend to run simulations and evaluate the performance of SIFDMM in applications with timing requirements, such as video streaming.

Acknowledgment

The authors acknowledge the financial support from the National Council for Scientific and Technological Development (CNPq) to Yarisley Peña Llerena through the PEC/PG Program.

References

[1] E. Bulut and B. K. Szymanski, "WiFi access point deployment for efficient mobile data offloading," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 17, 2013, pp. 71-78.

[2] C. Perkins, "RFC 5944: IP Mobility Support for IPv4," *Internet Engineering Task Force (IETF)* November 2010,

[3] C. Perkins, D. Johnson, and J. Arkko, "RFC 6275: Mobility support in IPv6," *Internet Engineering Task Force (IETF),* July 2011,

[4] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, and B. Patil, "RFC 5213: Proxy Mobile IPv6 " *Network Working Group,* August 2008,

[5] H. Chan, D. Liu, P. Seite, H. Yokota, and J. Korhonen, "RFC 7333: Requirements for Distributed Mobility Management," *Internet Engineering Task Force (IETF),* August 2014,

[6] 3GPP, "TR 23.861 V1.14.1, Network based IP flow mobility (Release 13)," 2015,

[7] H.-Y. Choi, S.-G. Min, Y.-H. Han, and R. Koodli, "Design and Simulation of a Flow Mobility Scheme Based on Proxy Mobile IPv6," *JIPS,* vol. 8, 2012, pp. 603-620.

[8] H.-Y. Choi, S.-G. Min, and Y.-H. Han, "PMIPv6-based Flow Mobility Simulation in NS-3," in *Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2011 Fifth International Conference on*, 2011, pp. 475-480.

[9] K. Jinho, Y. Morioka, and J. Hagiwara, "An optimized seamless IP flow mobility management architecture for traffic offloading," in *Network Operations and Management Symposium (NOMS), 2012 IEEE*, 2012, pp. 229-236.

[10] T. Melia, C. Bernardos, A. de la Oliva, F. Giust, and M. Calderon, "IP Flow Mobility in PMIPv6 Based Networks: Solution Design and Experimental Evaluation," *Wireless Personal Communications,* vol. 61, 2011/12/01 2011, pp. 603-627.

[11] C. Makaya, S. Das, and F. J. Lin, "Seamless data offload and flow mobility in vehicular communications networks," in *Wireless Communications and Networking Conference Workshops (WCNCW), 2012 IEEE*, 2012, pp. 338-343.

[12] R. I. Meneguette, L. F. Bittencourt, and E. R. M. Madeira, "A seamless flow mobility management architecture for vehicular communication networks," *Communications and Networks, Journal of,* vol. 15, 2013, pp. 207-216.

[13] Y. Li, Z. Huachun, and Z. Hongke, "An Efficient Distributed Mobility Management Scheme Based on PMIPv6," in *Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2012 Sixth International Conference on*, 2012, pp. 274-279.

[14] X. Keqiang, L. Jun, and W. Lei, "Design and Implementation of Flow Mobility Based on D-PMIPv6," in *Computational Science and Engineering (CSE), 2014 IEEE 17th International Conference on*, 2014, pp. 1344-1349.

[15] K. Sun and Y. Kim, "Flow Mobility Management in PMIPv6-based DMM (Distributed Mobility Management) Networks," *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA),* vol. 5, 2014, pp. 120-127.

[16] P. Seite, P. Bertin, and J. Lee, "Distributed Mobility Anchoring," *IETF draft, draft-seite-dmm-dma-07,* February 6 2014,

[17] H.-B. Lee, S.-G. Min, Y.-H. Han, K.-H. Lee, H.-W. Lee, *et al.*, "IP flow mobility scheme in scalable network-based mobility management architecture," *Telecommunication Systems,* 2015/04/02 2015, pp. 1-11.

[18] D. R. Purohith, A. Hegde, and K. M. Sivalingam, "Network architecture supporting seamless flow mobility between LTE and WiFi networks," in *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2015 IEEE 16th International Symposium on a*, 2015, pp. 1-9.

[19] L. Jong-Hyouk, K. D. Singh, J. M. Bonnin, and P. Sangheon, "Mobile Data Offloading: A Host-Based Distributed Mobility Management Approach," *Internet Computing, IEEE,* vol. 18, 2014, pp. 20-29.

[20] M. Wasserman and P. Seite, "Current Practices for Multiple-Interface Hosts," *RFC 6419,* November 2011,

[21] T. Melia and S. Gundavelli, " Logical-interface Support for Multi-access enabled IP Hosts," *draft-ietf-netext-logical-interface-support-11,* 2015,

[22] H. Ali-Ahmad, M. Ouzzif, P. Bertin, and X. Lagrange, "Performance Analysis on Network-Based Distributed Mobility Management," *Wireless Personal Communications,* vol. 74, 2014/02/01 2014, pp. 1245-1263.

[23] H. Ali-Ahmad, M. Ouzzif, P. Bertin, and X. Lagrange, "Distributed Mobility Management: Approaches and analysis," in *Communications Workshops (ICC), 2013 IEEE International Conference on*, 2013, pp. 1297-1302.

ANEXO I – FERRAMENTA EVALVID [70]

O EvalVid é uma ferramenta utilizada para avaliar a qualidade de um vídeo transmitido em uma rede real ou simulada. Além de fornecer dados sobre a rede, como atrasos e taxas de perdas, oferece também suporte a uma avaliação objetiva baseada no cálculo quadro a quadro de uma métrica chamada de PSNR. Sua utilização é feita em conjunto com ferramentas gratuitas como o FFMPEG, para realizar tarefas de conversão e codificação de vídeo. Uma grande vantagem do EvalVid é a possibilidade de realizar reconstruções e avaliações de vídeos em que houve perdas de pacotes. O funcionamento do EvalVid é ilustrado na Figura 1. Este diagrama de blocos mostra uma transmissão completa desde a gravação e codificação do vídeo na fonte até sua decodificação no receptor.

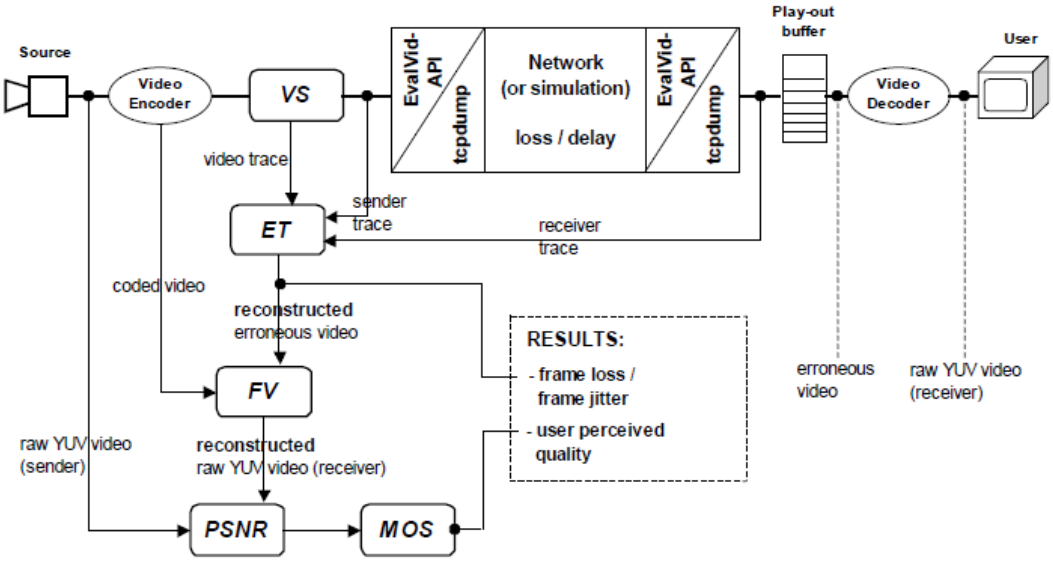

Figura 1: Funcionamento e design do EvalVid

Para realizar os cálculos de atraso, taxas de perdas e qualidade do vídeo, alguns dados se fazem necessários. A nível de transmissor tem-se:

- Vídeo original sem codificação;
- Vídeo codificado;
- Timestamp e tipo de todos os pacotes enviados.

A nível de receptor, são necessários:

- Timestamp e tipo de todos os pacotes recebidos;
- Vídeo codificado reconstruído (possivelmente com erros);

- Vídeo decodificado a ser mostrado.

Os blocos VS (do inglês, Video Sender), ET (do inglês, Evaluate Traces) e FV (do inglês, Fix Video), mostrados na Figura 1, são os recursos do EvalVid, e serão explicados nas Subseções seguintes. Os blocos PSNR, SSIM e MOS são as métricas utilizadas para a avaliação de qualidade dos vídeos recebidos.

❖ **O Bloco VS**

O propósito do VS é gerar um arquivo de trace a partir do vídeo codicado. Opcionalmente, o vídeo pode ser transmitido via UDP. Os resultados produzidos pelo VS são dois arquivos contendo informações a respeito de todos os quadros no vídeo e todos os pacotes (transmitidos e recebidos).

O arquivo de trace do vídeo contém as seguintes informações a respeito de cada quadro no vídeo:

- ID;
- Tipo (I, P, B, S);
- Tamanho;
- Número de pacotes UDP (um pacote UDP possui 1014 bytes);
- Timestamp.

Os arquivos de trace do transmissor e receptor contém as seguintes informações a respeito de cada pacote:

- Timestamp;
- ID (do pacote);
- Tamanho (do pacote).

Para realizar a simulação no NS-3 é necessário apenas o arquivo de trace do vídeo. Os outros dois são gerados durante a simulação, indicando os pacotes que estão sendo enviados e recebidos.

❖ **O Bloco ET**

É onde os cálculos de perdas e atrasos acontece. Para realiza-los são necessários os arquivos de trace descritos na sub-seção anterior. O cálculo de perdas é simples, já que os IDs presentes nos arquivos de trace são únicos. Se dá por meio da Equação (1):

$$PL_t = 100\frac{nT_{recv}}{nT_{sent}}, \qquad \text{Equação (1)}$$

onde Lt é o tipo do pacote (I, P, B, S), nTsent é o número de pacotes enviados e nTrecv é o número de pacotes recebidos.

Se o primeiro dos segmentos (pacotes) de um quadro é perdido, este é considerado perdido. Isso porque o decodificador não pode decodificar um quadro cujo primeiro segmento está faltando. A equação para calcular perdas de quadro se dá da mesma maneira, porém contabilizando quadros ao invés de pacotes.

O cálculo dos atrasos é feito por meio do uso dos timestamps presentes nos arquivos de trace do transmissor e receptor, através das Equações 2 a 4.

$$it_{p0} = 0 \qquad \text{Equação (2)}$$

$$it_{pn} = t_{pn} - t_{pn-1} \qquad \text{Equação (3)}$$

t_{pn} : timestamp do pacote n

$$j_P = \frac{1}{N}\sum_{i=1}^{N}(it_i - \bar{it}_N)^2 \qquad \text{Equação (4)}$$

N : número de pacotes

$\bar{it}_N$: média dos tempos entre pacotes

Estas fórmulas, porém, não podem ser usadas exatamente dessa maneira no caso de pacotes perdidos, já que não existirá no arquivo de trace o timestamp correspondente. O ET, neste caso, faz um palpite na hora do cálculo do atraso de um pacote perdido, baseado no valor mais provável de tempo em que o pacote chegaria no destino. Isso gera poucos problemas se a taxa de erros for baixa.

Uma outra tarefa realizada pelo ET é a reconstrução de um vídeo defeituoso devido à perda de pacotes. Este arquivo corrompido, junto ao arquivo de vídeo original, serão usados para efetuar a avaliação de qualidade. Além disso, pode ser especificado um tamanho limite de

buffer, e assim, pacotes que chegam com atraso superior serão considerados perdidos, para aumentar a semelhança com uma situação real. No caso por exemplo de um streaming de vídeo em tempo real, não haveria sentido em mostrar ao usuário quadros que chegaram muito atrasados.

Após a reconstrução do vídeo recebido, podem ser calculadas as métricas PSNR e SSIM para a avaliação objetiva da qualidade percebida pelo usuário. Além disso, pode ser calculado o MOS, que possui por padrão no EvalVid os intervalos descritos na seguinte Tabela.

PSNR [dB]	MOS
> 37	5 (Excellent)
31 - 37	4 (Good)
25 - 31	3 (Fair)
20 - 25	2 (Poor)
< 20	1 (Bad)

No MOS, a comparação entre os vídeos original e recebido é feita por meio da fração de quadros em cada nível. O impacto na rede é facilmente visível e a performance do sistema pode ser medida em termos da qualidade percebida pelo usuário. A Figura 2 mostra um exemplo de como é simples perceber se um vídeo transmitido possui qualidade próxima à máxima alcançável por meio das porcentagens de quadros na escala MOS definida. A barra da direita mostra os quadros do vídeo original, a do centro os quadros de um vídeo com 5% de perdas e a da esquerda os quadros de um vídeo com 25% de perdas.

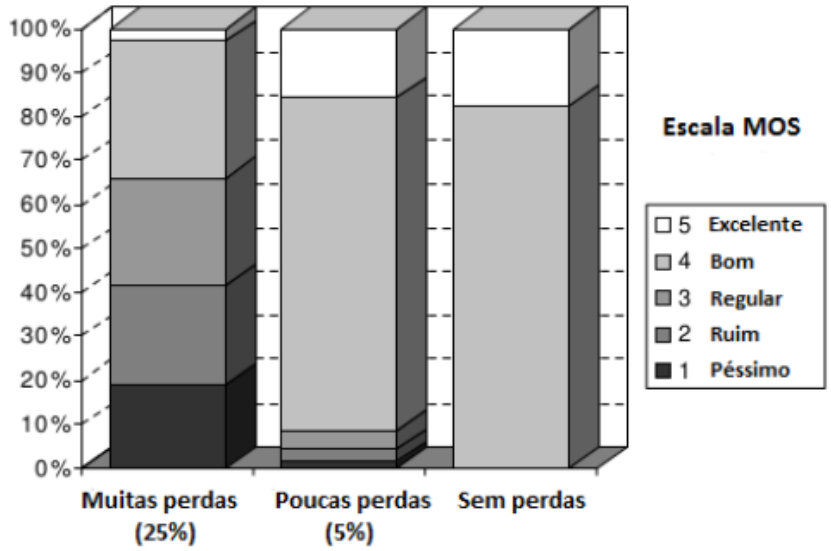

Figura 2: Comparação da qualidade de vídeos utilizando o MOS

❖ **O Bloco FV**

A avaliação de qualidade do vídeo é feita quadro a quadro. Isto significa que o número de pacotes recebidos deve ser igual ao número de pacotes transmitidos, gerando um problema no caso de uma situação em que houve perdas. Assim, o FV só é necessário quando o decodificador não consegue gerar quadros vazios para os pacotes perdidos.

O FV, para realizar esta tarefa de adequar o vídeo ao decodificador, poderia usar duas estratégias: inserir um quadros vazio, isto é, sem informação, no lugar de um quadro perdido ou ainda inserir uma cópia do quadro anterior. Na primeira alternativa, este quadro vazio resulta em um ponto preto (ou branco) na imagem. A segunda alternativa é a utilizada pelo FV, devido ao fato de que dois quadros consecutivos normalmente são muito parecidos, e por isso, este quadro inserido torna-se uma boa aproximação daquele que foi perdido.