CAPÍTULO 7

Redes sem fio e redes móveis

No mundo da telefonia, pode-se dizer que os últimos 25 anos foram os anos dourados da telefonia celular. O número de assinantes de telefones móveis no mundo inteiro aumentou de 34 milhões em 1993 para 8,3 bilhões em 2019. Hoje, há mais assinaturas de telefones móveis do que pessoas no nosso planeta. As muitas vantagens dos telefones celulares são evidentes para todos – em qualquer lugar, a qualquer hora, acesso desimpedido à rede global de telefonia por meio de um equipamento leve e totalmente portátil. Mais recentemente, smartphones, tablets e notebooks se conectaram sem fio à Internet por meio de redes celulares ou WiFi. E, cada vez mais, dispositivos como consoles de videogame, termostatos, sistemas de segurança, eletrodomésticos, relógios, óculos, automóveis, sistemas de controle de tráfego e muito mais estão estabelecendo conexões sem fio à Internet.

Do ponto de vista de rede, os desafios propostos pela ligação em rede desses dispositivos móveis e sem fio, em particular nas camadas de enlace e de rede, são tão diferentes dos desafios das redes de computadores cabeadas que é necessário um capítulo inteiro (*este* capítulo) devotado ao estudo de redes sem fio e redes móveis.

Iniciaremos este capítulo com uma discussão sobre usuários móveis, enlaces e redes sem fio e sua relação com as redes maiores (normalmente cabeadas) às quais se conectam. Traçaremos uma distinção entre os desafios propostos pela natureza *sem fio* dos enlaces de comunicação nessas redes e pela *mobilidade* que os enlaces sem fio habilitam. Fazer essa importante distinção – entre sem fio e mobilidade – nos permitirá isolar, identificar e dominar melhor os conceitos fundamentais em cada área.

Começaremos com um resumo da infraestrutura de acesso sem fio e da terminologia associada. Então, consideraremos as características desse enlace sem fio na Seção 7.2. Nessa seção, incluímos uma breve introdução ao acesso múltiplo por divisão de código (CDMA, do inglês *code division multiple access*), um protocolo de acesso ao meio compartilhado que é utilizado com frequência em redes sem fio. Na Seção 7.3, estudaremos com certa profundidade os aspectos da camada de enlace do padrão da rede local (LAN, do inglês *local area network*) sem fio IEEE 802.11 (WiFi); também falaremos um pouco sobre redes pessoais sem fio Bluetooth. Na Seção 7.4, daremos uma visão geral do acesso à Internet por telefone celular, incluindo 4G e as tecnologias celulares emergentes 5G, que fornecem acesso à Internet por voz e em alta velocidade. Na Seção 7.5, voltaremos nossa atenção à mobilidade, focalizando os problemas da localização de um usuário móvel, do roteamento até o usuário

móvel e da transferência (*hand-off*) do usuário móvel que passa dinamicamente de um ponto de conexão com a rede para outro. Estudaremos como esses serviços de mobilidade são implementados nas redes celulares 4G/5G e no padrão IP móvel na Seção 7.6. Por fim, na Seção 7.7, consideraremos o impacto dos enlaces e da mobilidade sem fio sobre protocolos de camada de transporte e aplicações em rede.

7.1 INTRODUÇÃO

A Figura 7.1 mostra o cenário no qual consideraremos os tópicos de comunicação de dados e mobilidade sem fio. Começaremos mantendo nossa discussão dentro de um contexto geral o suficiente para abranger uma ampla faixa de redes, entre elas LANs sem fio, WiFi e redes celulares 4G/5G; em outras seções, passaremos então para uma discussão mais detalhada de arquiteturas sem fio específicas. Podemos identificar os seguintes elementos em uma rede sem fio.

- *Hospedeiros sem fio.* Como no caso de redes cabeadas (ou com fio), hospedeiros são os equipamentos de sistemas finais que executam aplicações. Um **hospedeiro sem fio** pode ser um notebook, um tablet, um smartphone ou um dispositivo da Internet das Coisas (IoT, do inglês *Internet of Things*), como um sensor, eletrodoméstico, automóvel ou qualquer um de inúmeros aparelhos conectados à Internet. Os hospedeiros em si podem ser móveis ou não.
- *Enlaces sem fio.* Um hospedeiro se conecta a uma estação-base (definida mais adiante) ou a outro hospedeiro sem fio por meio de um **enlace de comunicação sem fio**. Tecnologias diferentes de enlace sem fio têm taxas de transmissão diversas e podem transmitir a distâncias variadas. A Figura 7.2 mostra duas características fundamentais, taxas de transmissão de enlaces e faixas de cobertura, dos padrões de enlace sem fio mais

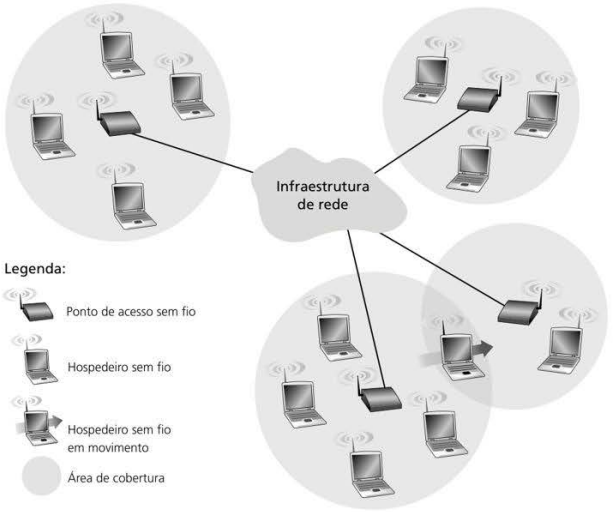


Figura 7.1 Elementos de uma rede sem fio.

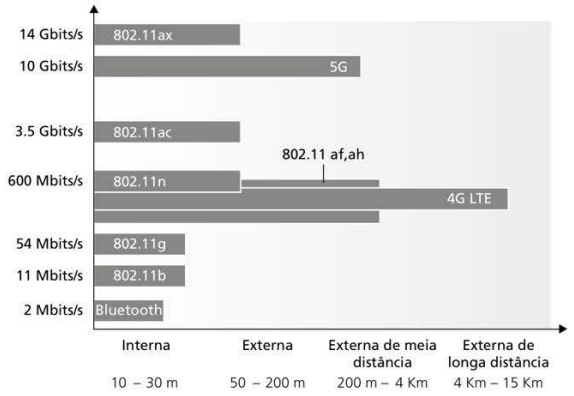


Figura 7.2 Alcance e taxas de transmissão sem fio para padrões WiFi, 4G/5G celular e Bluetooth (obs.: eixos não estão em escala).

populares. (A figura serve apenas para dar uma ideia aproximada dessas características. P. ex., alguns desses tipos de redes só estão sendo empregados agora, e algumas taxas de enlace podem aumentar ou diminuir além dos valores mostrados, dependendo da distância, das condições do canal e do número de usuários na rede sem fio.) Abordaremos esses padrões mais adiante, na primeira metade deste capítulo; consideraremos também outras características de enlaces sem fio (como suas taxas de erros de bit e as causas desses erros) na Seção 7.2.

Na Figura 7.1, enlaces sem fio conectam hospedeiros localizados na borda da rede com a infraestrutura da rede de maior porte. Não podemos nos esquecer de acrescentar que enlaces sem fio às vezes também são utilizados *dentro* de uma rede para conectar roteadores, comutadores e outros equipamentos de rede. Contudo, neste capítulo, focalizaremos a utilização da comunicação sem fio nas bordas da rede, pois é aqui que estão ocorrendo muitos dos desafios técnicos mais interessantes e a maior parte do crescimento.

- *Estação-base.* A **estação-base** é uma parte fundamental da infraestrutura de rede sem fio. Diferentemente dos hospedeiros e enlaces sem fio, uma estação-base não tem nenhuma contraparte óbvia em uma rede cabeada. Uma estação-base é responsável pelo envio e recebimento de dados (p. ex., pacotes) de e para um hospedeiro sem fio que está associado a ela. Uma estação-base frequentemente será responsável pela coordenação da transmissão de vários hospedeiros sem fio com os quais está associada. Quando dizemos que um hospedeiro sem fio está "associado" a uma estação-base, isso quer dizer que (1) o hospedeiro está dentro do alcance de comunicação sem fio da estação-base, e (2) o hospedeiro usa a estação-base para retransmitir dados entre ele (o hospedeiro) e a rede maior. **Torres celulares** em redes celulares e **pontos de acesso** em LANs sem fio 802.11 são exemplos de estações-base.

  Na Figura 7.1, a estação-base está conectada à rede maior (i.e., à Internet, à rede corporativa ou residencial); portanto, ela funciona como uma retransmissora da camada de enlace entre o hospedeiro sem fio e o resto do mundo com o qual o hospedeiro se comunica.

  Quando hospedeiros estão associados com uma estação-base, em geral diz-se que estão operando em **modo de infraestrutura**, já que todos os serviços tradicionais de rede (p. ex., atribuição de endereço e roteamento) são fornecidos pela rede com a qual estiverem conectados por meio da estação-base. Em **redes ad hoc**, hospedeiros sem fio não dispõem de qualquer infraestrutura desse tipo com a qual possam se conectar. Na ausência de tal infraestrutura, os próprios hospedeiros devem prover serviços como roteamento, atribuição de endereço, tradução de endereços semelhante ao sistema de nomes de domínio (DNS, do inglês *domain name system*) e outros.

  Quando um hospedeiro móvel se desloca para fora da faixa de alcance de uma estação-base e entra na faixa de outra, ele muda seu ponto de conexão com a rede maior (i.e., muda a estação-base com a qual está associado) – um processo denominado **transferência (*handoff* ou *handover*)**. Essa mobilidade dá origem a muitas questões desafiadoras. Se um hospedeiro pode se mover, como descobrir sua localização atual na rede de modo que seja possível lhe encaminhar dados? Como é realizado o endereçamento, visto que um hospedeiro pode estar em um entre muitos locais possíveis? Se o hospedeiro se movimentar *durante* uma conexão (do inglês *Transmission Control Protocol* – Protocolo de Controle de Transmissão) ou ligação telefônica, como os dados serão roteados para que a conexão continue sem interrupção? Essas e muitas (mas muitas!) outras questões fazem das redes sem fio e móveis uma área de pesquisa muito interessante sobre redes.
- *Infraestrutura de rede.* É a rede maior com a qual um hospedeiro sem fio pode querer se comunicar.

Após discutir sobre as "partes" da rede sem fio, observamos que essas partes podem ser combinadas de diversas maneiras diferentes para formar diferentes tipos de redes sem fio. Você pode achar uma taxonomia desses tipos de redes sem fio útil ao ler este capítulo, ou ler/aprender mais sobre redes sem fio além deste livro. No nível mais alto, podemos classificar as redes sem fio de acordo com dois critérios: *(i)* se um pacote na rede sem fio atravessa exatamente *um salto único sem fio ou múltiplos saltos sem fio*, e *(ii)* se há *infraestrutura* na rede, como uma estação-base:

- *Salto único, com infraestrutura.* Essas redes têm uma estação-base conectada a uma rede cabeada maior (p. ex., a Internet). Além disso, toda a comunicação é feita entre a estação-base e um hospedeiro sem fio através de um único salto sem fio. As redes 802.11 que você utiliza na sala de aula, na lanchonete ou na biblioteca; e as redes de dados 4G LTE, que aprenderemos em breve, encaixam-se nesta categoria. A grande maioria das nossas interações diárias é com redes sem fio de salto único, com infraestrutura.
- *Salto único, sem infraestrutura.* Nessas redes, não existe estação-base conectada à rede sem fio. Entretanto, como veremos, um dos nós nessa rede de salto único pode coordenar as transmissões dos outros nós. As redes Bluetooth (que conectam pequenos dispositivos sem fio, como teclados, alto-falantes e fones de ouvido, e que serão estudadas na Seção 7.3.6) são redes de salto único, sem infraestrutura.
- *Múltiplos saltos, com infraestrutura.* Nessas redes, está presente uma estação-base cabeada para as redes maiores. Entretanto, alguns nós sem fio podem ter que restabelecer sua comunicação através de outros nós sem fio para se comunicarem por meio de uma estação-base. Algumas redes de sensores sem fio e as chamadas **redes em malha sem fio**, usadas em residências, se encaixam nesta categoria.
- *Múltiplos saltos, sem infraestrutura.* Não existe estação-base nessas redes, e os nós podem ter de restabelecer mensagens entre diversos outros nós para chegar a um destino. Os nós também podem ser móveis, ocorrendo mudança de conectividade entre eles – uma categoria de redes conhecida como **redes móveis *ad hoc*** (**MANETs**, do inglês ***mobile ad hoc networks***). Se os nós móveis forem veículos, essa rede é denominada **rede veicular *ad hoc*** (**VANET**, do inglês ***vehicular ad hoc network***). Como você pode imaginar, o desenvolvimento de protocolos para essas redes é desafiador e constitui o assunto de muita pesquisa em andamento.

Neste capítulo, vamos nos limitar às redes de salto único e, depois, principalmente às redes baseadas em infraestrutura.

Agora vamos nos aprofundar um pouco mais nos desafios técnicos que surgem em redes sem fio e móveis. Começaremos considerando, em primeiro lugar, o enlace sem fio individual, deixando nossa discussão sobre mobilidade para outra parte deste capítulo.

7.2 CARACTERÍSTICAS DE ENLACES E REDES SEM FIO

Os enlaces sem fio diferem das suas contrapartes com fio em diversos aspectos importantes:

- *Redução da força do sinal.* Radiações eletromagnéticas são atenuadas quando atravessam algum tipo de matéria (p. ex., um sinal de rádio ao atravessar uma parede). O sinal se dispersará mesmo ao ar livre, resultando na redução de sua força (às vezes denominada **atenuação de percurso**) à medida que aumenta a distância entre emissor e receptor.
- *Interferência de outras fontes.* Várias fontes de rádio transmitindo na mesma banda de frequência sofrerão interferência umas das outras. Por exemplo, telefones sem fio de 2,4 GHz e LANs sem fio 802.11b transmitem na mesma banda de frequência. Assim, o usuário de uma LAN sem fio 802.11b que estiver se comunicando por um telefone sem fio de 2,4 GHz pode esperar que nem a rede nem o telefone funcionem particularmente bem. Além da interferência de fontes transmissoras, o ruído eletromagnético presente no ambiente (p. ex., um motor ou um equipamento de micro-ondas próximo) pode causar interferência. Por esse motivo, uma série de padrões 802.11 mais recentes operam na banda de frequência de 5 GHz.
- *Propagação multivias.* A **propagação multivias** (ou multicaminhos) ocorre quando partes da onda eletromagnética se refletem em objetos e no solo e tomam caminhos de comprimentos diferentes entre um emissor e um receptor. Isso resulta no embaralhamento do sinal recebido no destinatário. Objetos que se movimentam entre o emissor e o receptor podem fazer a propagação multivias mudar ao longo do tempo.

Para obter uma discussão detalhada sobre as características, modelos e medidas do canal sem fio, consulte Anderson (1995) e Almers (2007).

A discussão anterior sugere que erros de bit serão mais comuns em enlaces sem fio do que em enlaces com fio. Por essa razão, talvez não seja nenhuma surpresa que protocolos de enlace sem fio (como o protocolo 802.11 que examinaremos na seção seguinte) empreguem não só poderosos códigos de detecção de erros por verificação de redundância cíclica (CRC, do inglês *cyclic redundancy check*), mas também protocolos de transferência de dados confiáveis em nível de enlace, que retransmitem quadros corrompidos.

Tendo considerado as falhas que podem ocorrer em um canal sem fio, vamos voltar nossa atenção para o hospedeiro que recebe o sinal sem fio. Esse hospedeiro recebe um sinal eletromagnético que é uma combinação de uma forma degradada do sinal original transmitido pelo remetente (degradada pelos efeitos da atenuação e da propagação multivias, discutidas acima, entre outros) e um ruído de fundo no ambiente. A **relação sinal-ruído** (**SNR**, do inglês ***signal-to-noise ratio***) é uma medida relativa da potência do sinal recebido (i.e., a informação sendo transmitida) e o ruído. A SNR costuma ser calculada em unidades de decibéis (dB), uma unidade de medida que, segundo alguns, é utilizada por engenheiros elétricos principalmente para confundir cientistas da computação. A SNR, medida em dB, é vinte vezes o logaritmo de base 10 entre a amplitude do sinal recebido e a amplitude do ruído. Para nossos fins, precisamos saber apenas que uma SNR maior facilita ainda mais para o destinatário extrair o sinal transmitido de um ruído de fundo.

A Figura 7.3 (adaptada de Holland [2001]) mostra a taxa de erro de bits (BER, do inglês *bits error rate*) – em termos simples, a probabilidade de um bit transmitido ser recebido com erro no destinatário – *versus* a SNR para três técnicas de modulação diferentes para codificar informações para a transmissão em um canal sem fio idealizado. A teoria da modulação

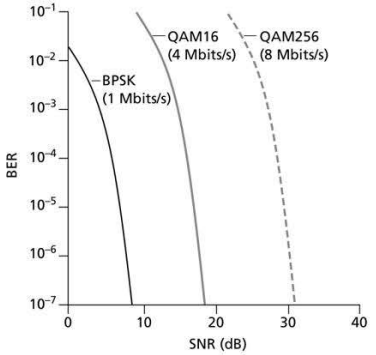


Figura 7.3 Taxa de erro de bits, taxa de transmissão e SNR.

e da codificação, bem como a extração do sinal e a BER, estão além do escopo deste livro (consulte Schwartz [1980] e Goldsmith [2005] para obter uma discussão sobre esses assuntos). Não obstante, a Figura 7.3 ilustra diversas características da camada física que são importantes para entender os protocolos de comunicação sem fio da camada superior.

- *Para um determinado esquema de modulação, quanto mais alta for a SNR, mais baixa será a BER.* Visto que um remetente consegue aumentar a SNR elevando sua potência de transmissão, ele pode reduzir a probabilidade de um quadro ser recebido com erro aumentando tal potência. Observe, entretanto, que há um pequeno ganho prático no aumento da potência além de certo patamar, digamos que para diminuir a BER de 10^{-12} para 10^{-13}. Existem também *desvantagens* associadas com o aumento da potência de transmissão: mais energia deve ser gasta pelo remetente (uma consideração importante para usuários móveis, que utilizam bateria), e as transmissões do remetente têm mais probabilidade de interferir nas transmissões de outro remetente (consulte Figura 7.4(b)).
- *Para determinada SNR, uma técnica de modulação com uma taxa de transmissão de bit maior (com erro ou não) terá uma BER maior.* Por exemplo, na Figura 7.3, com uma SNR de 10 dB, a modulação BPSK com uma taxa de transmissão de 1 Mbit/s possui uma BER menor do que 10^{-7}, enquanto para a modulação QAM16 com uma taxa de transmissão de 4 Mbits/s, a BER é 10^{-1}, longe de ser útil na prática. Entretanto, com uma SNR de 20 dB, a modulação QAM16 possui uma taxa de transmissão de 4 Mbits/s e uma BER de 10^{-7}, enquanto a modulação BPSK possui uma taxa de transmissão de apenas 1 Mbit/s e uma BER tão baixa como estar (literalmente) "fora do gráfico". Se é possível suportar uma BER de 10^{-7}, a taxa de transmissão mais alta apresentada pela modulação QAM16 faria desta a técnica de modulação preferida nesta situação. Tais considerações dão origem à característica final, descrita a seguir.
- *A seleção dinâmica da técnica de modulação da camada física pode ser usada para adaptar a técnica de modulação para condições de canal.* A SNR (e, portanto, a BER) pode mudar, como resultado da mobilidade ou em razão das mudanças no ambiente. A modulação adaptativa e a codificação são usadas nas redes de dados WiFi 802.11 e nas redes de dados celulares 4G e 5G, que estudaremos nas Seções 7.3 e 7.4. Isso permite, por exemplo, a seleção de uma técnica de modulação que ofereça a mais alta taxa de transmissão possível sujeita a uma limitação na BER, para as características de determinado canal.

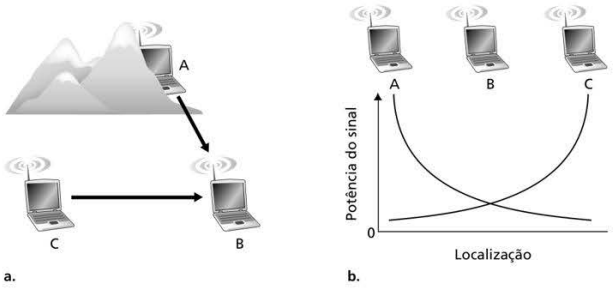


Figura 7.4 Problema do terminal oculto (a) e do desvanecimento (b).

Taxas de erros de bits mais altas e que variam com o tempo não são as únicas diferenças entre um enlace com fio e um enlace sem fio. Lembre-se de que, no caso de enlaces de difusão cabeados, cada nó recebe as transmissões de todos os outros nós. No caso de enlaces sem fio, a situação não é tão simples, conforme mostra a Figura 7.4. Suponha que a estação A esteja transmitindo para a estação B. Suponha também que a estação C esteja transmitindo para a estação B. O denominado **problema do terminal oculto**, obstruções físicas presentes no ambiente (p. ex., uma montanha ou um prédio), pode impedir que A e C escutem as transmissões um do outro, mesmo que as transmissões de A e C interfiram no destino, B. Isso é mostrado na Figura 7.4(a). Um segundo cenário que resulta em colisões que não são detectadas no receptor é causado pelo **desvanecimento** da força de um sinal à medida que se propaga pelo meio sem fio. A Figura 7.4(b) ilustra o caso em que a localização de A e C é tal que as potências de seus sinais não são suficientes para que eles detectem as transmissões um do outro, mas, mesmo assim, *são* fortes o bastante para interferir uma com a outra na estação B. Como veremos na Seção 7.3, o problema do terminal oculto e o desvanecimento tornam o acesso múltiplo em uma rede sem fio consideravelmente mais complexo do que em uma rede cabeada.

7.2.1 CDMA

Lembre-se de que dissemos, no Capítulo 6, que, quando hospedeiros se comunicam por um meio compartilhado, é preciso um protocolo para que os sinais enviados por vários emissores não interfiram nos receptores. No mesmo capítulo, descrevemos três classes de protocolos de acesso ao meio: de partição de canal, de acesso aleatório e de revezamento. O CDMA pertence à família de protocolos de partição de canal. Ele predomina em tecnologias de LAN sem fio e celulares. Por ser tão importante no mundo sem fio, examinaremos o CDMA rapidamente agora, antes de passar para tecnologias específicas de acesso sem fio nas próximas seções.

Com um protocolo CDMA, cada bit que está sendo enviado é codificado pela multiplicação do bit por um sinal (o código) que muda a uma velocidade muito maior (conhecida como **taxa de chipping**) do que a sequência original de bits de dados. A Figura 7.5 mostra um cenário simples e idealizado de codificação/decodificação CDMA. Suponha que a velocidade com que bits de dados originais cheguem ao codificador CDMA defina a unidade de tempo; isto é, cada bit original de dados a ser transmitido requer um intervalo de tempo de um bit. Seja d_i o valor do bit de dados para o i-ésimo intervalo de bit. Por conveniência do cálculo matemático, representamos o bit de dados com valor 0 por –1. Cada intervalo de bit é ainda subdividido em M mini-intervalos. Na Figura 7.5, $M = 8$, embora, na prática, M seja muito maior. O código CDMA usado pelo remetente consiste em uma sequência de M valores,

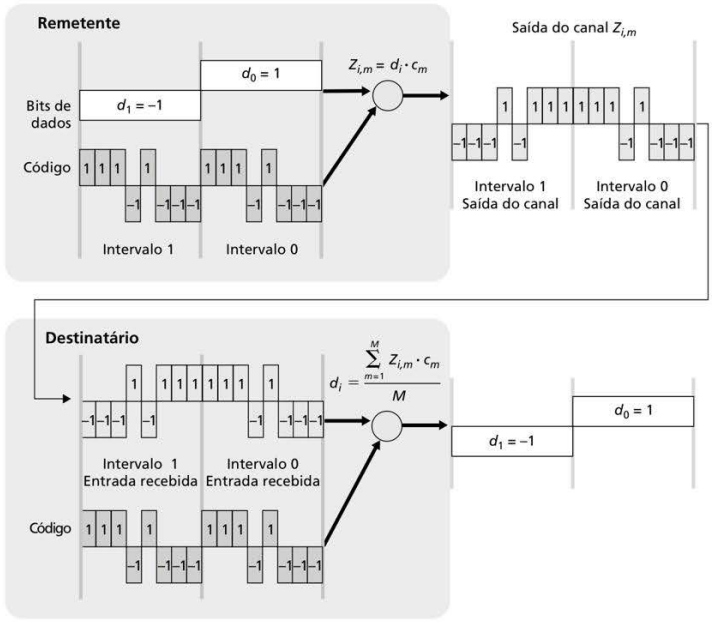


Figura 7.5 Um exemplo simples de CDMA: codificação no remetente, decodificação no receptor.

c_m, $m = 1, \ldots, M$, cada um assumindo um valor de +1 ou –1. No exemplo da Figura 7.5, o código CDMA de M bits que está sendo usado pelo remetente é (1, 1, 1, –1, 1, –1, –1, –1).

Para ilustrar como o CDMA funciona, vamos focalizar o i-ésimo bit de dados, d_i. Para o m-ésimo mini-intervalo do tempo de transmissão de bits de d_i, a saída do codificador CDMA, $Z_{i,m}$, é o valor de d_i multiplicado pelo m-ésimo bit do código CDMA escolhido, c_m:

$$Z_{i,m} = d_i \cdot c_m \tag{7.1}$$

Se o mundo fosse simples e não houvesse remetentes interferindo, o receptor receberia os bits codificados, $Z_{i,m}$, e recuperaria os bits de dados originais, d_i, calculando:

$$d_i = \frac{1}{M} \sum_{m=1}^{M} Z_{i,m} \cdot c_m \tag{7.2}$$

Talvez o leitor queira repassar os detalhes do exemplo da Figura 7.5 para verificar se os bits originais de dados são, de fato, corretamente recuperados no receptor usando a Equação 7.2.

No entanto, o mundo está longe de ser ideal e, como mencionamos antes, o CDMA deve funcionar na presença de remetentes que interferem e que estão codificando e transmitindo seus dados usando um código designado diferente. Mas como um receptor CDMA pode recuperar bits de dados originais de um remetente quando estes estão sendo embaralhados com bits que estão sendo transmitidos por outros remetentes? O CDMA trabalha na hipótese de que os sinais de bits interferentes sendo transmitidos são aditivos. Isso significa, por

exemplo, que, se três remetentes enviam um valor 1 e um quarto remetente envia um valor –1 durante o mesmo mini-intervalo, então o sinal recebido em todos os receptores durante o mini-intervalo é 2 (já que 1 + 1 + 1 − 1 = 2). Na presença de vários remetentes, s calcula suas transmissões codificadas, $Z^s_{i,m}$, exatamente como na Equação 7.1. O valor recebido no receptor durante o m-ésimo mini-intervalo do i-ésimo intervalo de bit, contudo, é agora a *soma* dos bits transmitidos de todos os N remetentes durante o mini-intervalo:

$$Z^*_{i,m} = \sum_{s=1}^{N} Z^s_{i,m}$$

Surpreendentemente, se os códigos dos remetentes forem escolhidos com cuidado, cada receptor pode recuperar os dados enviados por um dado remetente a partir do sinal agregado apenas usando o código do remetente, como na Equação 7.3:

$$d_i = \frac{1}{M}\sum_{m=1}^{M} Z^*_{i,m} \cdot c_m \tag{7.3}$$

A Figura 7.6 ilustra um exemplo de CDMA com dois remetentes. O código CDMA de M bits usado pelo remetente que está acima é (1, 1, 1, –1, 1, –1, –1, –1), ao passo que o código CDMA usado pelo que está embaixo é (1, –1, 1, 1, 1,–1, 1, 1). A Figura 7.6 ilustra um receptor recuperando os bits de dados originais do remetente que está acima. Note que o receptor pode extrair os dados do remetente 1, a despeito da transmissão interferente do remetente 2.

Voltando à analogia do coquetel apresentada no Capítulo 6, um protocolo CDMA é semelhante à situação em que os convidados falam vários idiomas; nessa circunstância, os seres humanos até que são bons para manter conversações no idioma que entendem e, ao mesmo tempo, continuar filtrando (rejeitando) outras conversações. Vemos aqui que o CDMA é um protocolo de partição, pois reparte o espaço de código (e não o tempo ou a frequência) e atribui a cada nó uma parcela dedicada do espaço de código.

Nossa discussão do código CDMA aqui é necessariamente breve; na prática, devem ser abordadas inúmeras questões diferentes. Primeiro, para que receptores CDMA consigam extrair o sinal de um emissor qualquer, os códigos CDMA devem ser escolhidos cuidadosamente. Segundo, nossa discussão considerou que as intensidades dos sinais recebidos de vários emissores são as mesmas; na realidade, isso pode ser difícil de conseguir. Existe muita literatura abordando essas e outras questões relativas ao CDMA; veja Pickholtz (1982) e Viterbi (1995) se quiser mais detalhes.

7.3 WIFI: LANS SEM FIO 802.11

Presentes no local de trabalho, em casa, em instituições educacionais, em cafés, aeroportos e esquinas, as LANs sem fio agora são uma das mais importantes tecnologias de rede de acesso na Internet de hoje. Embora muitas tecnologias e padrões para LANs sem fio tenham sido desenvolvidos na década de 1990, uma classe particular de padrões surgiu claramente como a vencedora: a **LAN sem fio IEEE 802.11**, também conhecida como **WiFi**. Nesta seção, estudaremos em mais detalhes as LANs sem fio 802.11, examinando a estrutura do quadro 802.11, o protocolo 802.11 de acesso ao meio e a interconexão de LANs 802.11 com LANs Ethernet cabeadas.

Como resumido na Tabela 7.1, existem diversos padrões 802.11 (IEEE 802.11, 2020). Os padrões 802.11 b, g, n, ac e ax são gerações sucessivas da tecnologia 802.11, voltada para redes locais sem fio (WLANS, do inglês *wireless local area networks*), em geral com alcance de menos de 70 m, para escritórios domésticos, escritórios comerciais ou empresas. Os padrões 802.11 n, ac e ax foram batizados recentemente de WiFi 4, 5 e 6, respectivamente, sem dúvida nenhuma

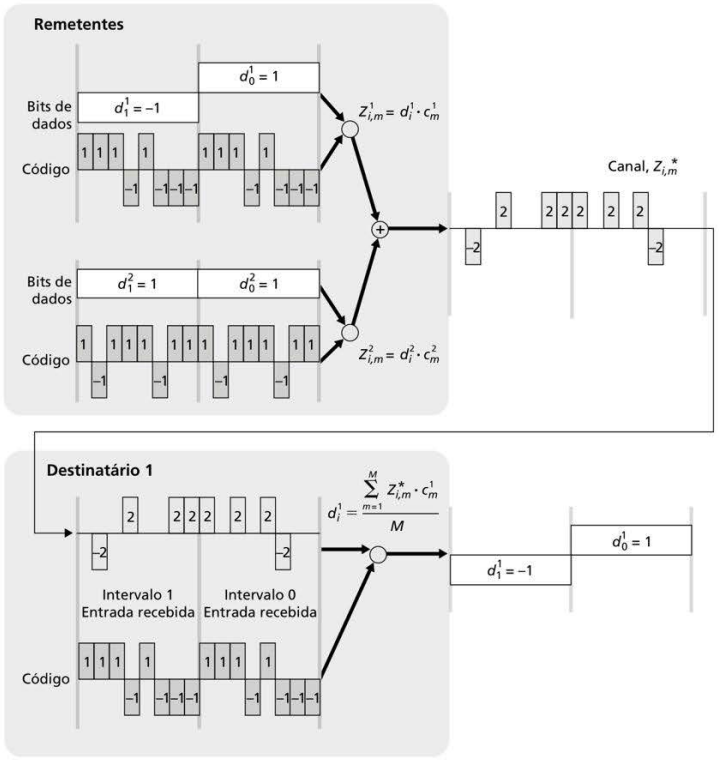


Figura 7.6 Um exemplo de CDMA com dois remetentes.

TABELA 7.1 Resumo dos padrões IEEE 802.11

Padrão IEEE 802.11	Ano	Taxa de dados máx.	Alcance	Frequência
802.11 b	1999	11 Mbits/s	30 m	2,4 Ghz
802.11 g	2003	54 Mbits/s	30 m	2,4 Ghz
802.11 n (WiFi 4)	2009	600	70 m	2,4, 5 Ghz
802.11 ac (WiFi 5)	2013	3,47 Gbits/s	70 m	5 Ghz
802.11 ax (WiFi 6)	2020 (esperado)	14 Gbits/s	70 m	2,4, 5 Ghz
802.11 af	2014	35–560 Mbits/s	1 km	bandas de TV não utilizadas (54–790 MHz)
802.11 ah	2017	347 Mbits/s	1 km	900 Mhz

para competir com as marcas de redes celulares 4G e 5G. Os padrões 802.11 af e ah operam em distâncias maiores e são direcionados para a IoT, redes de sensores e aplicações de medição.

Os diferentes padrões 802.11 b, g, n, ac e ax compartilham algumas características comuns, incluindo o formato de quadro 802.11 que estudaremos a seguir, e têm compatibilidade reversa, o que significa, por exemplo, que um dispositivo móvel que tem capacidade apenas para o 802.11 g ainda pode interagir com uma estação-base mais nova que use os padrões 802.11 ac ou 802.11 ax. Todos usam também o mesmo protocolo de acesso ao meio, o CSMA/CA, que também discutiremos a seguir, e o 802.11 ax também suporta o escalonamento centralizado pela estação-base de transmissões de dispositivos sem fio associados.

Contudo, como vemos na Tabela 7.1, os padrões têm algumas diferenças importantes na camada física. Os dispositivos 802.11 operam em duas faixas de frequência diferentes: 2,4 a 2,485 GHz (chamada de faixa de 2,4 GHz) e 5,1 a 5,8 GHz (chamada de faixa de 5 GHz). A faixa de 2,4 GHz é uma banda de frequência não licenciada, na qual os dispositivos 802.11 podem competir pelo espectro de frequência com telefones 2,4 GHz e eletrodomésticos, tais como fornos de micro-ondas. Na faixa de 5 GHz, as LANs 802.11 têm distâncias de transmissão mais curtas para um determinado nível de potência e sofrem mais com a propagação multivias. Os padrões 802.11 n, 802.11 ac e 802.11 ax utilizam antenas de entrada múltipla e saída múltipla (MIMO, do inglês *multiple input multiple-output*); ou seja, duas ou mais antenas no lado remetente e duas ou mais antenas no lado destinatário que estão transmitindo/recebendo sinais diferentes (Diggavi, 2004). As estações-base 802.11 ac e 802.11 ax podem transmitir para múltiplas estações simultaneamente, e usam antenas "inteligentes" para formar feixes (*beamforming*) adaptativamente de modo a direcionar transmissões para um receptor. Isso reduz a interferência e aumenta a distância alcançada em uma determinada taxa de dados. As taxas de dados listadas na Tabela 7.1 se referem a um ambiente idealizado, como o de um receptor próximo à estação-base, sem interferência, algo que dificilmente veremos na prática! Assim como a quilometragem dos automóveis, a taxa de dados sem fio também pode variar.

7.3.1 A arquitetura da LAN sem fio 802.11

A Figura 7.7 ilustra os principais componentes da arquitetura de LAN sem fio 802.11. O bloco de construção fundamental da arquitetura 802.11 é o **conjunto básico de serviço** (**BSS**, do inglês ***basic service set***). Um BSS contém uma ou mais estações sem fio e uma **estação-base** central, conhecida como um **ponto de acesso** (**AP**, do inglês ***access point***) na

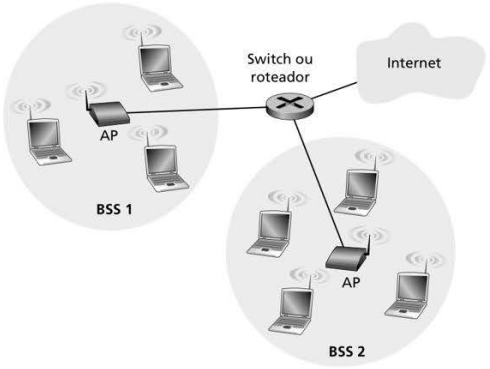


Figura 7.7 A arquitetura de LAN IEEE 802.11.

terminologia 802.11. A Figura 7.7 mostra o AP em cada um dos dois BSSs conectando-se a um dispositivo de interconexão (tal como um switch ou um roteador), que, por sua vez, leva à Internet. Em uma rede residencial típica, há apenas um AP e um roteador (normalmente integrados como uma unidade), que conecta o BSS à Internet.

Como acontece com dispositivos Ethernet, cada estação sem fio 802.11 tem um endereço MAC de 6 bytes que é armazenado no firmware do adaptador da estação (i.e., na placa de interface de rede 802.11). Cada AP também tem um endereço MAC (do inglês *media access control* – controle de acesso ao meio) para sua interface sem fio. Como na Ethernet, esses endereços MAC são administrados pelo IEEE e são (em teoria) globalmente exclusivos.

Como observamos na Seção 7.1, LANs sem fio que disponibilizam APs em geral são denominadas **LANs sem fio de infraestrutura**, e, nesse contexto, "infraestrutura" significa os APs junto com a infraestrutura de Ethernet cabeada que interconecta os APs e um roteador. A Figura 7.8 mostra que estações IEEE 802.11 também podem se agrupar e formar uma rede ad hoc – rede sem nenhum controle central e sem nenhuma conexão com o "mundo exterior". Nesse caso, a rede é formada conforme a necessidade, por dispositivos móveis que, por acaso, estão próximos uns dos outros, têm necessidade de se comunicar e não dispõem de infraestrutura de rede no lugar em que se encontram. Uma rede ad hoc pode ser formada quando pessoas que portam notebooks se reúnem (p. ex., em uma sala de conferências, um trem ou um carro) e querem trocar dados na ausência de um AP centralizado. As redes ad hoc estão despertando um interesse extraordinário com a contínua proliferação de equipamentos portáteis que podem se comunicar. Porém, nesta seção, concentraremos nossa atenção em LANs sem fio com infraestrutura.

Canais e associação

Em 802.11, cada estação sem fio precisa se associar com um AP antes de poder enviar ou receber dados da camada de rede. Embora todos os padrões 802.11 usem associação, discutiremos esse tópico especificamente no contexto da IEEE 802.11 b, g, n, ac e ax.

Ao instalar um AP, um administrador de rede designa ao ponto de acesso um **Identificador de Conjunto de Serviços** (**SSID**, do inglês ***Service Set Identifier***) composto por uma ou duas palavras. (Quando você escolhe WiFi na opção Configurações no seu iPhone, p. ex., é apresentada uma lista que mostra o SSID de todos os APs ordenados por faixa.) O administrador também deve designar um número de canal ao AP. Para entender números de canal, lembre-se de que as redes 802.11 operam na faixa de frequência de 2,4 a 2,485 GHz. Dentro dessa faixa de 85 MHz, o padrão 802.11 define 11 canais que se sobrepõem em parte. Não há sobreposição entre quaisquer dois canais se, e somente se, eles estiverem separados por quatro ou mais canais. Em particular, o conjunto dos canais 1, 6 e 11 é o único de três canais não sobrepostos. Isso significa que um administrador poderia criar uma LAN sem fio com uma taxa máxima de transmissão agregada de três vezes a taxa de transmissão máxima

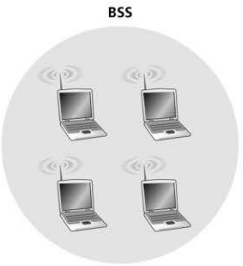


Figura 7.8 Uma rede *ad hoc* IEEE 802.11 BSS.

mostrada na Tabela 7.1 instalando três APs 802.11 na mesma localização física, designando os canais 1, 6 e 11 aos APs e interconectando cada um desses APs com um switch.

Agora que já entendemos o básico sobre canais 802.11, vamos descrever uma situação interessante (e que não é completamente fora do comum) – uma selva de WiFis. Uma **selva de WiFis (WiFi *jungle*)** é qualquer localização física na qual uma estação sem fio recebe um sinal suficientemente forte de dois ou mais APs. Por exemplo, em muitos cafés da cidade de Nova York, uma estação sem fio pode captar um sinal de diversos APs próximos. Um deles pode ser o AP gerenciado pelo café, enquanto os outros podem estar localizados em apartamentos vizinhos. Cada ponto de acesso provavelmente estaria localizado em uma sub-rede IP diferente e teria sido designado independentemente a um canal.

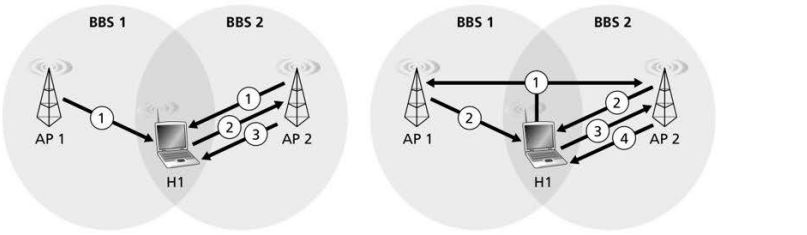
Agora suponha que você entre nessa selva de WiFis com seu smartphone, tablet ou notebook, em busca de acesso à Internet sem fio e de um cafezinho. Suponha que há cinco APs na selva de WiFis. Para conseguir acesso à Internet, seu dispositivo sem fio terá de se juntar a exatamente uma das sub-redes e, portanto, precisará se **associar** com exatamente um dos APs. Associar significa que o dispositivo sem fio cria um fio virtual entre ele mesmo e o AP. De modo específico, só o AP associado enviará quadros de dados (i.e., quadros contendo dados, tal como um datagrama) ao seu dispositivo sem fio, e este enviará quadros de dados à Internet apenas por meio do AP associado. Mas como seu dispositivo sem fio se associa com um determinado AP? E, o que é mais fundamental, como seu dispositivo sem fio sabe quais APs estão dentro da selva, se é que há algum?

O padrão 802.11 requer que um AP envie periodicamente **quadros de sinalização**, cada qual incluindo o SSID e o endereço MAC do AP. Seu dispositivo sem fio, sabendo que os APs estão enviando quadros de sinalização, faz uma varredura dos 11 canais em busca de quadros de sinalização de quaisquer APs que possam estar por lá (alguns dos quais talvez estejam transmitindo no mesmo canal – afinal, estamos na selva!). Ao tomar conhecimento dos APs disponíveis por meio dos quadros de sinalização, você (ou seu dispositivo sem fio) seleciona um desses pontos de acesso para se associar.

O padrão 802.11 não especifica um algoritmo para selecionar com quais dos APs disponíveis se associar; esse algoritmo é de responsabilidade dos projetistas do firmware e do software 802.11 em seu dispositivo sem fio. Em geral, o dispositivo escolhe o AP cujo quadro de sinalização é recebido com a intensidade de sinal mais alta. Embora uma intensidade alta do sinal seja algo bom (p. ex., veja a Figura 7.3), esta não é a única característica do AP que determinará o desempenho que um dispositivo recebe. Em particular, é possível que o AP selecionado tenha um sinal forte, mas pode ser sobrecarregado com outros dispositivos associados (que precisarão compartilhar a largura de banda sem fio naquele AP), enquanto um AP não tão carregado não é selecionado em razão de um sinal levemente mais fraco. Diversas formas alternativas de escolher os APs foram propostas recentemente (Vasudevan, 2005; Nicholson, 2006; Sudaresan, 2006). Para obter uma discussão interessante e prática de como a intensidade do sinal é medida, consulte Bardwell (2004).

O processo de varrer canais e ouvir quadros de sinalização é conhecido como **varredura passiva** (veja a Figura 7.9(a)). Um dispositivo sem fio pode também realizar uma **varredura ativa**, transmitindo um quadro de investigação que será recebido por todos os APs dentro de uma faixa do dispositivo sem fio, como mostrado na Figura 7.9(b). Os APs respondem ao quadro de requisição de investigação com um quadro de resposta de investigação. O dispositivo sem fio pode, então, escolher o AP com o qual irá se associar entre os APs que estão respondendo.

Após selecionar o AP ao qual se associará, o dispositivo sem fio envia um quadro de solicitação de associação ao AP, e este responde com um quadro de resposta de associação. Observe que essa segunda apresentação de solicitação/resposta é necessária com a varredura ativa, visto que um AP de resposta ao quadro de solicitação de investigação inicial não sabe quais dos (possivelmente muitos) APs de resposta o dispositivo escolherá para se associar, do mesmo modo que um cliente DHCP (do inglês *Dynamic Host Configuration Protocol* – Protocolo de Configuração Dinâmica de Hospedeiros) pode escolher entre múltiplos servidores DHCP (veja a Figura 4.21). Uma vez associado ao AP, o dispositivo desejará entrar



a. Varredura passiva
1. Quadros de sinalização enviados pelos APs
2. Quadro de Solicitação de Associação enviado: H1 para AP selecionado
3. Quadro de Resposta de Associação enviado: AP selecionado para H1

a. Varredura ativa
1. Difusão do quadro de Investigação de H1
2. Quadro de Resposta de Investigação enviado pelos APs
3. Quadro de Pedido de Associação enviado ao AP selecionado: AP selecionado para H1
4. Quadro de Resposta de Associação enviado: AP selecionado para H1

Figura 7.9 Varredura passiva e ativa para pontos de acesso.

na sub-rede (no sentido do endereçamento IP da Seção 4.3.3) à qual pertence o AP. Assim, o dispositivo normalmente enviará uma mensagem de descoberta DHCP (veja a Figura 4.21) à sub-rede por meio de um AP a fim de obter um endereço IP na sub-rede. Logo que o endereço é obtido, o resto do mundo, então, vê esse dispositivo apenas como outro hospedeiro com um endereço IP naquela sub-rede.

Para criar uma associação com um determinado AP, o dispositivo sem fio talvez tenha de se autenticar perante o AP. LANs sem fio 802.11 dispõem de várias alternativas para autenticação e acesso. Uma abordagem, usada por muitas empresas, é permitir o acesso a uma rede sem fio com base no endereço MAC de um dispositivo. Uma segunda abordagem, usada por muitas "LAN houses", emprega nomes de usuários e senhas. Em ambos os casos, o AP em geral se comunica com um servidor de autenticação e transmite informações entre o dispositivo sem fio e o servidor de autenticação usando um protocolo como o RADIUS (RFC 2865) ou o DIAMETER (RFC 6733). Separar o servidor de autenticação do AP permite que um servidor de autenticação atenda a muitos APs, centralizando as decisões de autenticação e acesso (quase sempre delicadas) em um único servidor e mantendo baixos os custos e a complexidade do AP. Veremos, no Capítulo 8, que o novo protocolo IEEE 802.11 i, que define aspectos de segurança da família de protocolos 802.11, adota exatamente essa técnica.

7.3.2 O protocolo MAC 802.11

Uma vez associada com um AP, uma estação sem fio pode começar a enviar e receber quadros de dados de e para o ponto de acesso. Porém, como múltiplos dispositivos sem fio podem querer transmitir quadros de dados ao mesmo tempo sobre o mesmo canal, é preciso um protocolo de acesso múltiplo para coordenar as transmissões. Aqui, chamaremos esses dispositivos ou o AP como "estações" sem fio que compartilham o canal de acesso múltiplo. Como discutimos no Capítulo 6 e na Seção 7.2.1, em termos gerais, há três classes de protocolos de acesso múltiplo: partição de canal (incluindo CDMA), acesso aleatório e revezamento. Inspirados pelo enorme sucesso da Ethernet e seu protocolo de acesso aleatório, os projetistas do 802.11 escolheram um protocolo de acesso aleatório para as LANs sem fio 802.11. Esse protocolo de acesso aleatório é denominado **CSMA com prevenção de colisão** ou, mais sucintamente, **CSMA/CA** (do inglês ***carrier sense multiple access with collision avoidance* – acesso múltiplo com detecção de portadora com prevenção de colisão**). Do mesmo modo que o CSMA/CD da Ethernet, o "CSMA" de CSMA/CA quer dizer "acesso múltiplo por detecção de portadora", o que significa que cada estação sonda o canal antes de transmitir e abstém-se

de transmitir quando percebe que o canal está ocupado. Embora tanto a Ethernet quanto o 802.11 usem acesso aleatório por detecção de portadora, os dois protocolos MAC apresentam diferenças importantes. Primeiro, em vez de usar detecção de colisão, o 802.11 usa técnicas de prevenção de colisão. Segundo, em virtude das taxas relativamente altas de erros de bits em canais sem fio, o 802.11 (ao contrário da Ethernet) usa um esquema de reconhecimento/retransmissão (ARQ, do inglês *Automatic Repeat reQuest* – solicitação automática de repetição) de camada de enlace. Mais adiante, descreveremos os esquemas usados pelo 802.11 para prevenção de colisão e reconhecimento na camada de enlace.

Lembre-se de que, nas seções 6.3.2 e 6.4.2, dissemos que, com o algoritmo de detecção de colisão, uma estação Ethernet ouve o canal à medida que transmite. Se, enquanto estiver transmitindo, a estação detectar que alguma outra estação também está, ela abortará sua transmissão e tentará novamente após uma pequena unidade de tempo aleatória. Ao contrário do protocolo Ethernet 802.3, o protocolo MAC 802.11 *não* implementa detecção de colisão. Isso se deve a duas razões importantes:

- A capacidade de detectar colisões exige as capacidades de enviar (o próprio sinal da estação) e de receber (para determinar se alguma outra estação está transmitindo) ao mesmo tempo. Como a potência do sinal recebido em geral é muito pequena em comparação com a potência do sinal transmitido no adaptador 802.11, é caro construir um hardware que possa detectar colisões.
- Mais importante, mesmo que o adaptador pudesse transmitir e ouvir ao mesmo tempo (e, presumivelmente, abortar transmissões quando percebesse um canal ocupado), ainda assim ele não seria capaz de detectar todas as colisões, devido ao problema do terminal escondido e do desvanecimento, como discutimos na Seção 7.2.

Como LANs 802.11 sem fio não usam detecção de colisão, uma vez que uma estação comece a transmitir um quadro, *ela o transmite integralmente*; isto é, tão logo uma estação inicie, não há volta. Como é de se esperar, transmitir quadros inteiros (em particular, os longos) quando existe grande possibilidade de colisão pode degradar significativamente o desempenho de um protocolo de acesso múltiplo. Para reduzir a probabilidade de colisões, o 802.11 emprega diversas técnicas de prevenção de colisão, que discutiremos em breve.

Antes de considerar prevenção de colisão, contudo, primeiro devemos examinar o esquema de **reconhecimento na camada de enlace** do 802.11. Lembre-se de que dissemos, na Seção 7.2, que quando uma estação em uma LAN sem fio envia um quadro, este talvez não chegue intacto à estação de destino, por diversos motivos. Para lidar com essa probabilidade não desprezível de falha, o protocolo MAC 802.11 usa reconhecimentos de camada de enlace. Como ilustrado na Figura 7.10, quando a estação de destino recebe um quadro que passou na CRC, ela espera um curto período, conhecido como **Espaçamento Curto Interquadros** (**SIFS**, do inglês ***Short Inter-frame Spacing***), e então devolve um quadro de reconhecimento. Se a estação transmissora não receber um reconhecimento em dado período, ela admitirá que ocorreu um erro e retransmitirá o quadro usando de novo o protocolo CSMA/CA para acessar o canal. Se a estação transmissora não receber um reconhecimento após certo número fixo de retransmissões, desistirá e descartará o quadro.

Agora que já discutimos como o 802.11 usa reconhecimentos da camada de enlace, estamos prontos para descrever o protocolo CSMA/CA 802.11. Suponha que uma estação (pode ser uma estação sem fio ou um AP) tenha um quadro para transmitir.

1. Se inicialmente a estação perceber que o canal está ocioso, ela transmitirá seu quadro após um curto período conhecido como **Espaçamento Interquadros Distribuído** (**DIFS**, do inglês ***Distributed Inter-frame Space***); ver Figura 7.10.
2. Caso contrário, a estação escolherá um valor aleatório de recuo usando o recuo exponencial binário (conforme encontramos na Seção 6.3.2) e fará a contagem regressiva a partir desse valor quando perceber que o canal está ocioso. Se a estação perceber que o canal está ocupado, o valor do contador permanecerá congelado.

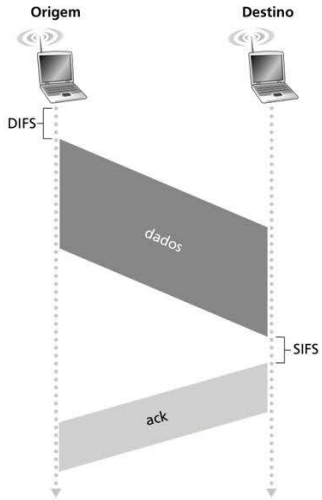


Figura 7.10 802.11 usa reconhecimentos da camada de enlace.

3. Quando o contador chegar a zero (note que isso pode ocorrer somente quando a estação percebe que o canal está ocioso), a estação transmitirá o quadro inteiro e então ficará esperando um reconhecimento.
4. Se receber um reconhecimento, a estação transmissora saberá que o quadro foi corretamente recebido na estação de destino. Se a estação tiver outro quadro para transmitir, iniciará o protocolo CSMA/CA na etapa 2. Se não receber um reconhecimento, a estação entrará de novo na fase de recuo na etapa 2 e escolherá um valor aleatório em um intervalo maior.

Lembre-se de que, no protocolo de acesso múltiplo com detecção de colisão (CSMA/CD, do inglês *carrier sense multiple access with collision detection*) (Seção 6.3.2), uma estação começa a transmitir tão logo percebe que o canal está ocioso. Com o CSMA/CA, entretanto, a estação priva-se de transmitir enquanto realiza a contagem regressiva, mesmo quando percebe que o canal está ocioso. Por que o CSMA/CD e o CDMA/CA adotam essas abordagens diferentes aqui?

Para responder a essa pergunta, vamos considerar um cenário com duas estações em que cada uma tem um quadro a transmitir, mas nenhuma transmite imediatamente porque percebe que uma terceira estação já está transmitindo. Com o CSMA/CD da Ethernet, cada uma das duas estações transmitiria tão logo detectasse que a terceira estação terminou de transmitir. Isso causaria uma colisão, o que não é um problema sério em CSMA/CD, já que ambas as estações abortariam suas transmissões e assim evitariam a transmissão inútil do restante dos seus quadros. Entretanto, com 802.11, a situação é bem diferente. Como o 802.11 não detecta uma colisão nem aborta transmissão, um quadro que sofra uma colisão será transmitido integralmente. Assim, a meta do 802.11 é evitar colisões sempre que possível. Com esse protocolo, se duas estações perceberem que o canal está ocupado, ambas entrarão imediatamente em backoff aleatório e, esperamos, escolherão valores diferentes de backoff. Se esses valores forem, de fato, diferentes, assim que o canal ficar ocioso, uma das duas começará a transmitir antes da outra, e (se as duas não estiverem ocultas uma da outra)

a "estação perdedora" ouvirá o sinal da "estação vencedora", interromperá seu contador e não transmitirá até que a estação vencedora tenha concluído sua transmissão. Desse modo, é evitada uma colisão dispendiosa. É claro que ainda podem ocorrer colisões com 802.11 nesse cenário: as duas estações podem estar ocultas uma da outra ou podem escolher valores de backoff aleatório próximos o bastante para que a transmissão da estação que se inicia primeiro tenha ainda de atingir a segunda. Lembre-se de que já vimos esse problema antes em nossa discussão sobre algoritmos de acesso aleatório no contexto da Figura 6.12.

Tratando de terminais ocultos: RTS e CTS

O protocolo 802.11 MAC também inclui um esquema de reserva inteligente (mas opcional) que ajuda a evitar colisões mesmo na presença de terminais ocultos. Vamos estudar esse esquema no contexto da Figura 7.11, que mostra duas estações sem fio e um ponto de acesso. Ambas as estações estão dentro da faixa do AP (cuja área de cobertura é representada por um círculo sombreado) e ambas se associaram com o AP. Contudo, pelo desvanecimento, as faixas de sinal de estações sem fio estão limitadas ao interior dos círculos sombreados mostrados na Figura 7.11. Assim, cada uma das estações está oculta da outra, embora nenhuma esteja oculta do AP.

Agora vamos considerar por que terminais ocultos podem ser problemáticos. Suponha que a estação H1 esteja transmitindo um quadro e, a meio caminho da transmissão, a estação H2 queira enviar um quadro para o AP. O H2, que não está ouvindo a transmissão de H1, primeiro esperará um intervalo DIFS para, então, transmitir o quadro, resultando em uma colisão. Por conseguinte, o canal será desperdiçado durante todo o período da transmissão de H1, bem como durante a transmissão de H2.

Para evitar esse problema, o protocolo IEEE 802.11 permite que uma estação utilize um quadro de controle **RTS** (do inglês ***Request to Send* – solicitação de envio**) curto e um quadro de controle **CTS** (do inglês ***Clear to Send* – pronto para envio**) curto para *reservar* acesso ao canal. Quando um remetente quer enviar um quadro DATA, ele pode enviar primeiro um quadro RTS ao AP, indicando o tempo total requerido para transmitir o quadro DATA e o quadro de reconhecimento (ACK, do inglês *acknowledgement*). Quando o AP recebe o quadro RTS, responde fazendo a transmissão por difusão de um quadro CTS. Esse quadro CTS tem duas finalidades: dá ao remetente uma permissão explícita para enviar e também instrui as outras estações a não enviar durante o tempo reservado.

Assim, na Figura 7.12, antes de transmitir um quadro DATA, H1 primeiro faz uma transmissão por difusão de um quadro RTS, que é ouvida por todas as estações que estiverem dentro do seu círculo de alcance, incluindo o AP. O AP então responde com um quadro CTS, que é ouvido por todas as estações dentro de sua faixa de alcance, incluindo H1 e H2. Como ouviu o CTS, a estação H2 deixa de transmitir durante o tempo especificado no quadro CTS. Os quadros RTS, CTS, DATA e ACK são mostrados na Figura 7.12.

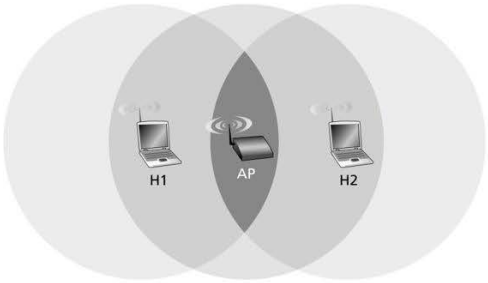


Figura 7.11 Exemplo de terminal oculto: H1 está oculto de H2, e vice-versa.

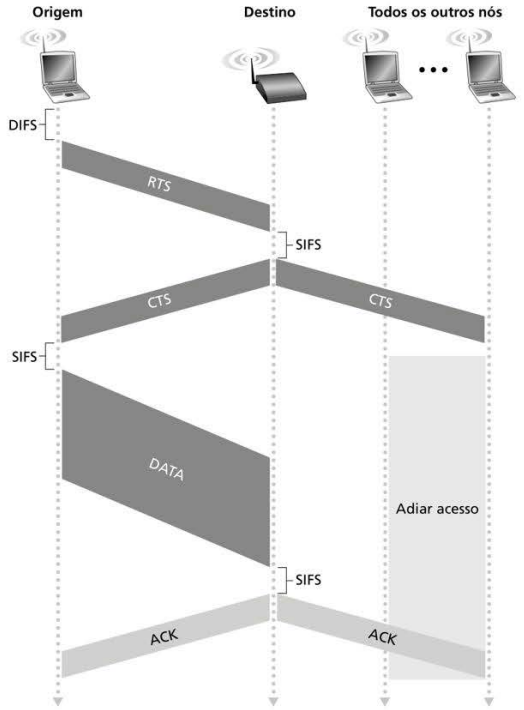


Figura 7.12 Prevenção de colisão usando os quadros RTS e CTS.

A utilização dos quadros RTS e CTS pode melhorar o desempenho de dois modos importantes:

- O problema da estação oculta é atenuado, visto que um quadro DATA longo é transmitido apenas após o canal ter sido reservado.
- Como os quadros RTS e CTS são curtos, uma colisão que envolva um quadro RTS ou CTS terá apenas a duração dos quadros RTS ou CTS curtos. Desde que os quadros RTS e CTS sejam corretamente transmitidos, os quadros DATA e ACK subsequentes deverão ser transmitidos sem colisões.

Aconselhamos o leitor a verificar a animação sobre 802.11 no site deste livro. Essa animação interativa ilustra o protocolo CSMA/CA, incluindo a sequência de troca RTS/CTS.

Embora a troca RTS/CTS ajude a reduzir colisões, também introduz atraso e consome recursos do canal. Por essa razão, a troca RTS/CTS é utilizada (se for utilizada) apenas para reservar o canal para a transmissão de um quadro DATA longo. Na prática, cada estação sem fio pode estabelecer um patamar RTS tal que a sequência RTS/CTS seja utilizada somente quando o quadro for mais longo do que o patamar. Para muitas estações sem fio, o valor default do patamar RTS é maior do que o comprimento máximo do quadro, de modo que a sequência RTS/CTS é omitida para todos os quadros DATA enviados.

Usando o 802.11 como enlace ponto a ponto

Até aqui, nossa discussão focalizou a utilização do 802.11 em um cenário de múltiplo acesso. Devemos mencionar que, se dois nós tiverem, cada um, uma antena direcional, eles poderão dirigir suas antenas um para o outro e executar o protocolo 802.11 sobre o que é, essencialmente, um enlace ponto-a-ponto. Dado o baixo custo comercial do hardware 802.11, a utilização de antenas direcionais e uma maior potência de transmissão permitem que o 802.11 seja utilizado como um meio barato de prover conexões sem fio ponto a ponto por dezenas de quilômetros. Raman (2007) descreve uma das primeiras redes sem fio multissaltos, que operou nas planícies rurais do rio Ganges, na Índia, e que usava enlaces 802.11 ponto a ponto.

7.3.3 O quadro IEEE 802.11

Embora o quadro 802.11 tenha muitas semelhanças com um quadro Ethernet, ele também contém vários campos que são específicos para sua utilização em enlaces sem fio. O quadro 802.11 é mostrado na Figura 7.13. Os números acima de cada campo no quadro representam os comprimentos dos campos em *bytes*; os números acima de cada subcampo no campo de controle do quadro representam os comprimentos dos subcampos em *bits*. Agora vamos examinar os campos no quadro, bem como alguns dos subcampos mais importantes no campo de controle do quadro.

Campos de carga útil e de CRC

No coração do quadro está a carga útil, que consiste, tipicamente, em um datagrama IP ou em um pacote ARP (do inglês *Address Resolution Protocol* – Protocolo de Resolução de Endereços). Embora o comprimento permitido do campo seja 2.312 bytes, em geral ele é menor do que 1.500 bytes, contendo um datagrama IP ou um pacote ARP. Como um quadro Ethernet, um quadro 802.11 inclui uma CRC, de modo que o receptor possa detectar erros de bits no quadro recebido. Como já vimos, erros de bits são muito mais comuns em LANs sem fio do que em LANs cabeadas, portanto, aqui, CRC é ainda mais útil.

Campos de endereço

Talvez a diferença mais marcante no quadro 802.11 é que ele tem *quatro* campos de endereço, e cada um pode conter um endereço MAC de 6 bytes. Mas por que quatro campos de endereço? Um campo de origem MAC e um campo de destino MAC não são suficientes como são na Ethernet? Acontece que aqueles três campos de endereço são necessários para finalidades de interconexão em rede – especificamente, para mover o datagrama de camada de enlace de uma estação sem fio, passando por um AP, até uma interface de roteador. O quarto campo de endereço é usado quando APs encaminham quadros uns aos outros em

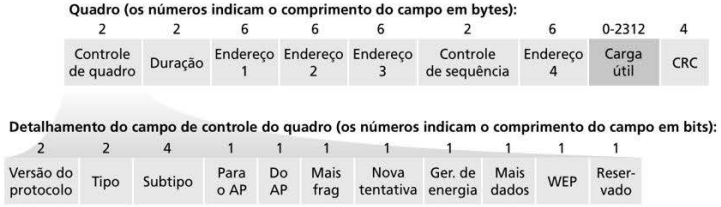


Figura 7.13 O quadro 802.11.

modo ad hoc. Visto que estamos considerando apenas redes de infraestrutura, vamos concentrar nossa atenção nos três primeiros campos de endereço. O padrão 802.11 define esses campos da seguinte forma:

- Endereço 2 é o endereço MAC da estação que transmite o quadro. Assim, se uma estação sem fio transmitir o quadro, o endereço MAC daquela estação será inserido no campo de endereço 2. De modo semelhante, se um AP transmitir o quadro, o endereço MAC do AP será inserido no campo de endereço 2.
- Endereço 1 é o endereço MAC da estação sem fio que deve receber o quadro. Assim, se uma estação móvel sem fio transmitir o quadro, o endereço 1 conterá o endereço MAC do AP de destino. De modo semelhante, se um AP transmitir o quadro, o endereço 1 conterá o endereço MAC da estação sem fio de destino.
- Para entender o endereço 3, lembre-se de que o BSS (que consiste no AP e em estações sem fio) faz parte de uma sub-rede, e que esta se conecta com outras sub-redes, por meio de alguma interface de roteador. O endereço 3 contém o endereço MAC dessa interface de roteador.

Para compreender melhor a finalidade do endereço 3, vamos examinar um exemplo de interconexão em rede no contexto da Figura 7.14. Nessa figura, há dois APs, cada um responsável por certo número de estações sem fio. Cada AP tem uma conexão direta com um roteador que, por sua vez, se liga com a Internet global. Devemos ter sempre em mente que um AP é um dispositivo da camada de enlace e, portanto, não "fala" IP nem entende endereços IP. Agora, considere mover um datagrama da interface de roteador R1 até a estação sem fio H1. O roteador não está ciente de que há um AP entre ele e H1; do ponto de vista do roteador, H1 é apenas um hospedeiro em uma das sub-redes às quais ele (o roteador) está conectado.

- O roteador, que conhece o endereço IP de H1 (pelo endereço de destino do datagrama), utiliza ARP para determinar o endereço MAC de H1, exatamente como aconteceria em uma LAN Ethernet comum. Após obter o endereço MAC de H1, a interface do roteador R1 encapsula o datagrama em um quadro Ethernet. O campo de endereço de origem desse quadro contém o endereço MAC de R1, e o campo de endereço de destino contém o endereço MAC de H1.
- Quando o quadro Ethernet chega ao AP, este converte o quadro Ethernet 802.3 para um quadro 802.11 antes de transmiti-lo para o canal sem fio. O AP preenche o endereço 1

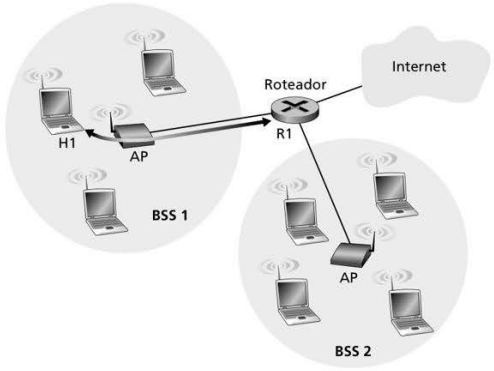


Figura 7.14 A utilização de campos de endereço em quadros 802.11: movendo um quadro entre H1 e R1.

e o endereço 2 com o endereço MAC de H1 e seu próprio endereço MAC, respectivamente, como descrito. Para o endereço 3, o AP insere o endereço MAC de R1. Dessa maneira, H1 pode determinar (a partir do endereço 3) o endereço MAC da interface de roteador que enviou o datagrama para a sub-rede.

Agora considere o que acontece quando a estação sem fio H1 responde movendo um datagrama de H1 para R1.

- H1 cria um quadro 802.11, preenchendo os campos de endereço 1 e 2 com o endereço MAC do AP e com o endereço MAC de H1, respectivamente, como descrito. Para o endereço 3, H1 insere o endereço MAC de R1.
- Quando recebe o quadro 802.11, o AP o converte para um quadro Ethernet. O campo de endereço de origem para esse quadro é o endereço MAC de H1, e o campo de endereço de destino é o endereço MAC de R1. Assim, o endereço 3 permite que o AP determine o endereço de destino MAC apropriado ao construir o quadro Ethernet.

Em resumo, o endereço 3 desempenha um papel crucial na interconexão do BSS com uma LAN cabeada.

Campos de número de sequência, duração e controle de quadro

Lembre-se de que, em 802.11, sempre que uma estação recebe corretamente um quadro de outra, devolve um reconhecimento. Como reconhecimentos podem ser perdidos, a estação emissora pode enviar várias cópias de determinado quadro. Como vimos em nossa discussão sobre o protocolo rdt2.1 (Seção 3.4.1), a utilização de números de sequência permite que o receptor distinga entre um quadro recém-transmitido e a retransmissão de um quadro anterior. Assim, o campo de número de sequência no quadro 802.11 cumpre aqui, na camada de enlace, exatamente a mesma finalidade que cumpria na camada de transporte do Capítulo 3.

Não se esqueça de que o protocolo 802.11 permite que uma estação transmissora reserve o canal durante um período que inclui o tempo para transmitir seu quadro de dados e o tempo para transmitir um reconhecimento. Esse valor de duração é incluído no campo de duração do quadro (tanto para quadros de dados quanto para os quadros RTS e CTS).

Como mostrado na Figura 7.13, o campo de controle de quadro inclui muitos subcampos. Diremos apenas umas poucas palavras sobre alguns dos mais importantes; se o leitor quiser uma discussão mais completa, aconselhamos que consulte a especificação 802.11 (Held, 2001; Crow, 1997; IEEE 802.11, 1999). Os campos *tipo* e *subtipo* são usados para distinguir os quadros de associação, RTS, CTS, ACK e de dados. Os campos *de* e *para* são usados para definir os significados dos diferentes campos de endereço. (Esses significados mudam dependendo da utilização dos modos ad hoc ou de infraestrutura, e, no caso do modo de infraestrutura, mudam dependendo de o emissor do quadro ser uma estação sem fio ou um AP.) Finalmente, o campo WEP (do inglês *Wired Equivalent Privacy* – Privacidade Equivalente à de Rede com fios) indica se está sendo ou não utilizada criptografia. (A WEP é discutida no Capítulo 8.)

7.3.4 Mobilidade na mesma sub-rede IP

Para ampliar a faixa física de uma LAN sem fio, empresas e universidades frequentemente distribuirão vários BSSs dentro da mesma sub-rede IP. Isso, claro, levanta a questão da mobilidade entre os BSSs – como estações sem fio passam imperceptivelmente de um BSS para outro enquanto mantêm sessões TCP em curso? Como veremos nesta subseção, a mobilidade pode ser manipulada de uma maneira relativamente direta quando os BSSs são parte de uma sub-rede. Quando estações se movimentam entre sub-redes, são necessários protocolos de gerenciamento de mobilidade mais sofisticados, tais como os que estudaremos nas Seções 7.5 e 7.6.

Agora vamos examinar um exemplo específico de mobilidade entre BSSs na mesma sub-rede. A Figura 7.15 mostra dois BSSs interconectados e um hospedeiro, H1, que se move

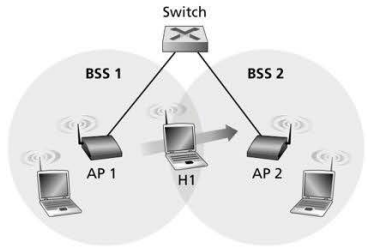


Figura 7.15 Mobilidade na mesma sub-rede.

entre BSS1 e BSS2. Como nesse exemplo o dispositivo de interconexão entre os dois BSSs *não* é um roteador, todas as estações nos dois BSSs, incluindo os APs, pertencem à mesma sub-rede IP. Assim, quando H1 se move de BSS1 para BSS2, ele pode manter seu endereço IP e todas as suas conexões TCP em curso. Se o dispositivo de interconexão fosse um roteador, então H1 teria de obter um novo endereço IP na sub-rede na qual estava entrando. Essa mudança de endereço interromperia (e, em consequência, finalizaria) qualquer conexão TCP em curso em H1. Na Seção 7.6, veremos como um protocolo de mobilidade da camada de rede, como o IP móvel, pode ser usado para evitar esse problema.

Mas o que acontece especificamente quando H1 passa de BSS1 para BSS2? À medida que se afasta de AP1, o H1 detecta um enfraquecimento do sinal de AP1 e começa a fazer uma varredura em busca de um sinal mais forte. H1 recebe quadros de sinalização de AP2 (que em muitos ambientes empresariais e universitários terão o mesmo SSID do AP1). Então, H1 se desassocia de AP1 e se associa com AP2, mantendo, ao mesmo tempo, seu endereço IP e suas sessões TCP em curso.

Isso resolve o problema de transferência do ponto de vista do hospedeiro e do AP. Mas e quanto ao switch na Figura 7.15? Como é possível saber que o hospedeiro se locomoveu de um AP a outro? O leitor talvez se lembre de que, no Capítulo 6, dissemos que switches são "autodidatas" e constroem automaticamente suas tabelas de repasse. Essa característica de autoaprendizagem funciona bem para movimentações ocasionais (p. ex., quando um profissional é transferido de um departamento para outro); contudo, switches não são projetados para suportar usuários com alto grau de mobilidade, que querem manter conexões TCP enquanto se movimentam entre BSSs. Para avaliar este problema aqui, lembre-se de que, antes da movimentação, o switch tem um registro em sua tabela de repasse que vincula o endereço MAC de H1 com a interface de saída do switch por meio da qual o H1 pode ser alcançado. Se H1 estiver inicialmente em BSS1, então um datagrama destinado a H1 será direcionado a ele via AP1. Contudo, tão logo H1 se associe com BSS2, seus quadros deverão ser direcionados para AP2. Uma solução (na verdade um tanto forçada) é o AP2 enviar ao switch um quadro Ethernet de difusão com o endereço de origem de H1 logo após a nova associação. Quando o switch recebe o quadro, atualiza sua tabela de repasse, permitindo que H1 seja alcançado via AP2. O grupo de padrões 802.11 f está desenvolvendo um protocolo entre APs para cuidar dessas e outras questões associadas.

Nossa discussão acima se concentrou na mobilidade com a mesma sub-rede de LAN. Lembre-se que as redes locais virtuais (VLANs, do inglês *virtual local area networks*), que estudamos na Seção 6.4.4, podem ser usadas para unificar ilhas de LANs para formar uma grande LAN virtual, abrangendo uma região geográfica maior. A mobilidade entre estações-base dentro dessa VLAN pode ser gerenciada exatamente da maneira descrita acima (Yu, 2011).

HISTÓRICO DO CASO

DESCOBERTA DE LOCALIZAÇÃO: GPS E POSICIONAMENTO WIFI

Muitos dos aplicativos para smartphone mais úteis e importantes da atualidade são aplicações móveis baseadas em localização, incluindo Foursquare, Yelp, Uber, Pokémon Go e Waze. Todos esses aplicativos de software utilizam uma interface de programação de aplicação (API, do inglês *application programming interface*) que lhes permite extrair a sua posição geográfica atual diretamente do smartphone. Você já se perguntou como o seu smartphone obtém a sua posição geográfica? Hoje, o processo combina dois sistemas, o **Sistema de Posicionamento Global (GPS, do inglês *Global Positioning System*)** e o **Sistema de Posicionamento WiFi (GPS, do inglês WPS – *WiFi Positioning System*)**.

O GPS, com uma constelação de mais de 30 satélites, transmite por difusão informações de temporização e localização do satélite, que, por sua vez, são utilizadas por cada receptor de GPS para estimar a sua geolocalização. O governo dos Estados Unidos criou, mantém e disponibiliza gratuitamente o sistema para todos com um receptor GPS. Os satélites possuem relógios atômicos muito estáveis, sincronizados entre si e com relógios no solo. Os satélites também sabem suas localizações com alto nível de precisão. Todo satélite de GPS transmite continuamente um sinal de rádio que contém seu horário e posição atuais. Se obtém essas informações de, no mínimo, quatro satélites, um receptor de GPS pode resolver equações de triangulação para estimar a sua posição.

O GPS não consegue, entretanto, sempre fornecer geolocalizações exatas se não possui linha de visada com pelo menos quatro satélites de GPS ou quando há interferência de outros sistemas de comunicação de alta frequência. Isso vale particularmente em ambientes urbanos, onde edifícios altos muitas vezes bloqueiam os sinais de GPS. É aqui que os sistemas de posicionamento WiFi vêm em apoio. Estes usam bancos de dados de pontos de acesso WiFi, mantidos de forma independente por diversas empresas da Internet, incluindo Google, Apple e Microsoft. Cada banco de dados contém informações sobre milhões de pontos de acesso WiFi, incluindo a SSID de cada ponto de acesso e uma estimativa da sua localização geográfica. Para entender como um sistema de posicionamento WiFi utiliza um banco de dados desse tipo, considere um smartphone Android combinado com o serviço de localização da Google. De cada ponto de acesso próximo, o smartphone recebe e mede a intensidade de sinais de sinalização (ver Seção 7.3.1), que contêm a SSID do ponto de acesso. Assim, o smartphone pode enviar mensagens continuamente ao serviço de localização da Google (na nuvem) que incluem as SSIDs de pontos de acesso próximos e as intensidades de sinal correspondentes. Ele também envia o seu posicionamento pelo GPS (obtido por meio de sinais de satélite, como descrito acima), quando disponível. Usando as informações de intensidade de sinal, a Google estima a distância entre o smartphone e cada um dos pontos de acesso WiFi. O sistema então usa essas distâncias estimadas para resolver equações de triangulação para estimar a geolocalização do smartphone. Por fim, essa estimativa baseada em WiFi é combinada com a estimativa baseada no satélite de GPS para formar uma estimativa agregada, que é então enviada de volta ao smartphone e utilizada pelos aplicativos móveis baseados em localização.

Mas você pode estar se perguntando como a Google (e a Apple, a Microsoft, etc.) consegue obter e manter o banco de dados de pontos de acesso, especialmente a localização geográfica de cada um. Lembre-se que, para um determinado ponto de acesso, cada smartphone Android próximo envia para o serviço de localização da Google informações sobre a intensidade do sinal recebido do ponto de acesso, assim como a localização geográfica estimada do smartphone. Dado que milhares de smartphones podem passar pelo ponto de acesso a cada dia, o serviço de localização da Google possui *muitos* dados à disposição para aplicar em estimativas sobre a posição do ponto de acesso e, mais uma vez, resolver equações de triangulação. Assim, os pontos de acesso ajudam os smartphones a determinar as suas localizações, e os smartphones, por sua vez, ajudam os pontos de acesso a determinar as suas próprias localizações!

7.3.5 Recursos avançados em 802.11

Finalizaremos nossa abordagem sobre 802.11 com uma breve discussão sobre capacidades avançadas encontradas nas redes 802.11. Como veremos, essas capacidades *não* são completamente especificadas no padrão 802.11, mas, em vez disso, são habilitadas por mecanismos especificados no padrão. Isso permite que fornecedores diferentes implementem essas capacidades usando suas próprias abordagens, aparentemente trazendo vantagens sobre a concorrência.

Adaptação da taxa 802.11

Vimos antes, na Figura 7.3, que as diferentes técnicas de modulação (com as diferentes taxas de transmissão que elas fornecem) são adequadas para diferentes cenários de SNR. Considere, por exemplo, um usuário 802.11 móvel que está, inicialmente, a 20 metros de distância da estação-base, com uma alta relação sinal-ruído. Dada a alta SNR, o usuário pode se comunicar com essa estação usando uma técnica de modulação da camada física que oferece altas taxas de transmissão enquanto mantém uma BER baixa. Esse usuário é um felizardo! Suponha agora que o usuário se torne móvel, se distanciando da estação-base, e que a SNR diminui à medida que a distância da estação-base aumenta. Neste caso, se a técnica de modulação usada no protocolo 802.11 que está operando entre a estação-base e o usuário não mudar, a BER será inaceitavelmente alta à medida que a SNR diminui e, por conseguinte, nenhum quadro transmitido será recebido corretamente.

Por essa razão, algumas execuções de 802.11 possuem uma capacidade de adaptação de taxa que seleciona, de maneira adaptável, a técnica de modulação da camada física sobreposta a ser usada com base em características atuais ou recentes do canal. Se um nó enviar dois quadros seguidos sem receber confirmação (uma indicação implícita de erros de bit no canal), a taxa de transmissão cai para a próxima taxa mais baixa. Se dez quadros seguidos forem confirmados, ou se um temporizador (que registra o tempo desde o último fallback) expirar, a taxa de transmissão aumenta para a próxima taxa mais alta. Esse mecanismo de adaptação da taxa compartilha a mesma filosofia de "investigação" (refletida por recebimentos de ACK); quando algo "ruim" acontece, a taxa de transmissão é reduzida. A adaptação da taxa 802.11 e o controle de congestionamento TCP são, desse modo, semelhantes a uma criança: está sempre exigindo mais e mais de seus pais até eles por fim dizerem "Chega!" e a criança desistir (para tentar novamente após a situação melhorar!). Diversos métodos também foram propostos para aperfeiçoar esse esquema básico de ajuste automático de taxa (Kamerman, 1997; Holland, 2001; Lacage, 2004).

Gerenciamento de energia

A energia é uma fonte preciosa em aparelhos móveis, e, assim, o padrão 802.11 provê capacidades de gerenciamento de energia, permitindo que os nós 802.11 minimizem o tempo de suas funções de percepção, transmissão e recebimento, e outros circuitos necessários para "funcionar". O gerenciamento de energia 802.11 opera da seguinte maneira. Um nó é capaz de alternar entre os estados "dormir" e "acordar" (como um aluno com sono em sala de aula!). Um nó indica ao ponto de acesso que entrará no modo de dormir ajustando o bit de gerenciamento de energia no cabeçalho de um quadro 802.11 para 1. Um temporizador localizado no nó é, então, ajustado para acordar o nó antes de um AP ser programado para enviar seu quadro de sinalização (lembre-se de que, em geral, um AP envia um quadro de sinalização a cada 100 ms). Uma vez que o AP descobre, pelo bit de transmissão de energia, que o nó vai dormir, ele (o AP) sabe que não deve enviar nenhum quadro àquele nó, e armazenará qualquer quadro destinado ao hospedeiro que está dormindo para transmissão posterior.

Um nó acordará logo após o AP enviar um quadro de sinalização, e logo entrará no modo ativo (ao contrário do aluno sonolento, esse despertar leva apenas 250 μs [Kamerman, 1997]!). Os quadros de sinalização enviados pelo AP contêm uma relação de nós cujos quadros foram mantidos em buffer no AP. Se não houver quadros mantidos em buffer para o nó, ele pode voltar a dormir. Caso contrário, o nó pode solicitar explicitamente o envio dos quadros armazenados, enviando uma mensagem de polling ao AP. Com um tempo de 100 ms entre as sinalizações, um despertar de 250 μs e um tempo semelhantemente pequeno para receber um quadro de sinalização e verificar que não haja quadros em buffer, um nó que não possui quadros para enviar ou receber pode dormir 99% do tempo, resultando em uma economia de energia significativa.

7.3.6 Redes pessoais: Bluetooth

As redes **Bluetooth** parecem ter se tornado parte do cotidiano rapidamente. Talvez você já tenha usado uma rede Bluetooth como tecnologia de "substituição de cabos" para interconectar seu computador a um teclado, mouse ou outro periférico sem fio. Ou talvez tenha usado uma rede Bluetooth para conectar seus fones de ouvido, alto-falantes, relógios ou pulseira de monitoramento de saúde ao seu smartphone ou o próprio telefone ao sistema de som do automóvel. Em todos esses casos, o Bluetooth opera a curtas distâncias (dezenas de metros ou menos), baixa potência e baixo custo. Por esse motivo, as redes Bluetooth também são chamadas de **redes pessoais sem fio** (**WPANs**, do inglês ***wireless personal area networks***) ou **picorredes.**

Embora sejam pequenas e relativamente simples pela natureza do projeto, as redes Bluetooth estão repletas de muitas das técnicas de rede do nível da camada de enlace que estudamos anteriormente, incluindo multiplexação por divisão de tempo (TDM, do inglês *time-division multiplexing*) e divisão de frequência (Seção 6.3.1), recuo aleatório (Seção 6.3.2), polling (seleção; Seção 6.3.3), detecção *e* correção de erros (Seção 6.2) e transferência confiável de dados por ACKs e NAKS (Seção 3.4.1). E estamos pensando só na camada de enlace do Bluetooth!

As redes Bluetooth operam na banda de frequência Industrial, Científica e Médica (ISM, do inglês *Industrial, Scientific and Medical*) de 2,4 GHz não licenciada, junto com outros eletrodomésticos, tais como micro-ondas, abridores de portão de garagem e telefones sem fio. Por consequência, as redes Bluetooth são projetadas explicitamente com ruídos e interferência em mente. O canal sem fio Bluetooth é operado no modo TDM, com intervalos de tempo de 625 microssegundos. Durante cada intervalo de tempo, um emissor transmite por um entre 79 canais, sendo que, de intervalo para intervalo, o canal muda de uma maneira conhecida, porém pseudoaleatória. Essa forma de saltar de canal em canal, conhecida como **espectro espalhado com salto de frequência** (**FHSS**, do inglês ***frequency-hopping spread spectrum***), é usada para que a interferência de outros dispositivos ou eletrodomésticos que operam na banda ISM interfira apenas com comunicações Bluetooth em, no máximo, um subconjunto dos intervalos. As taxas de dados Bluetooth podem alcançar 3 Mbits/s.

As redes Bluetooth são redes ad hoc: não é preciso que haja uma infraestrutura de rede (p. ex., um ponto de acesso). Os dispositivos Bluetooth precisam organizar *a si mesmos* em uma picorrede (*piconet*: pequena rede) de até oito dispositivos ativos, como ilustra a Figura 7.16. Um desses dispositivos é designado como o mestre, e os outros agem como clientes. O nó mestre comanda a picorrede – seu relógio determina o tempo na picorrede (p. ex., determina os limites de intervalo TDM), determina a sequência de salto de frequência entre intervalos, controla a entrada de dispositivos clientes na picorrede, controla a potência (100 mW, 2,5 mW ou 1mW) à qual cada dispositivo cliente transmite, e usa polling para

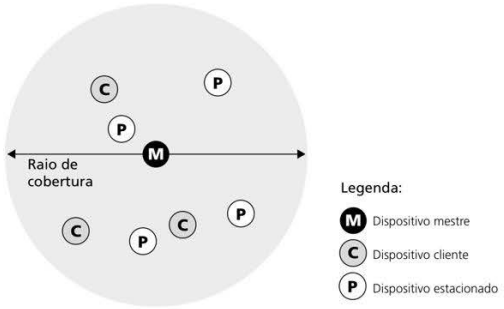


Figura 7.16 Uma picorrede Bluetooth.

conceder permissão aos clientes para transmitir após admiti-los à rede. Além dos dispositivos ativos, a rede também pode conter até 255 dispositivos "estacionados". Os dispositivos estacionados muitas vezes se encontram em alguma forma de modo de suspensão para conservar energia (como vimos com o gerenciamento de energia nas redes 802.11) e acordam periodicamente, de acordo com o cronograma do dispositivo mestre, para receber mensagens de sinalização deste. Os dispositivos estacionados não podem se comunicar até que o nó mestre tenha mudado seu estado de estacionado para ativo.

Como as redes ad hoc Bluetooth devem se **auto-organizar**, vale a pena analisar como constroem sua estrutura de rede em tempo real. Quando um nó mestre deseja formar uma rede Bluetooth, ele deve antes determinar quais outros dispositivos Bluetooth estão ao seu alcance; é o problema da **descoberta de vizinhos**. Para isso, o mestre transmite por difusão uma série de 32 mensagens de *inquiry* (pergunta), cada uma em um canal de frequência diferente, e repete a sequência de transmissão até 128 vezes. O dispositivo cliente escuta na sua frequência escolhida, na expectativa de ouvir as mensagens de *inquiry* do mestre nela. Quando recebe uma mensagem, o dispositivo cliente recua por um período aleatório de 0 a 0,3 segundos (para evitar colisões com outros nós que estejam respondendo, semelhante ao recuo binário da Ethernet), e então responde ao mestre com uma mensagem que contém a ID do dispositivo.

Após ter descoberto todos os clientes em potencial ao seu alcance, o Bluetooth mestre convida-os para se juntar à picorrede. A segunda fase é chamada de **paginação Bluetooth**, e se assemelha ao modo como clientes 802.11 se associam a uma estação-base. Por meio do processo de paginação, o mestre informa o cliente sobre o padrão de salto de frequência a ser utilizado e o relógio do remetente. O mestre recomeça o processo de paginação enviando 32 mensagens de convite de paginação idênticas, agora endereçando cada uma a um cliente específico, mas novamente usando frequências diferentes, pois o cliente ainda não aprendeu o padrão de salto de frequência. Após o cliente responder com uma mensagem ACK à mensagem de convite de paginação, o mestre envia as informações de salto de frequência e de sincronização dos relógios e um endereço de membro ativo para o cliente; por fim, o mestre seleciona (polling) o cliente, já usando o padrão de salto de frequência, para garantir que o cliente está conectado à rede.

Na nossa discussão acima, mencionamos apenas brevemente as redes sem fio Bluetooth. Protocolos de níveis superiores permitem a transferência confiável de pacotes de dados, streaming de áudio e vídeo semelhante a circuito, alterações de níveis de potência de transmissão, alteração do estado ativo/estacionado (e outros estados) e muito mais. Versões mais recentes do Bluetooth resolveram questões de baixa energia e segurança. Para mais informações sobre o Bluetooth, o leitor interessado deve consultar Bisdikian (2001), Colbach (2017) e Bluetooth (2020).

7.4 REDES CELULARES: 4G E 5G

Na seção anterior, vimos como um hospedeiro pode acessar a Internet quando estiver nas vizinhanças de um ponto de acesso (AP) WiFi 802.11. Mas, como vimos, os APs têm áreas de cobertura pequenas, e o hospedeiro certamente não será capaz de se associar com cada AP que encontrar. Por consequência, o acesso WiFi sequer chega perto de ser onipresente para um usuário em movimento.

Por outro lado, o acesso a redes celulares 4G se disseminou rapidamente. Um estudo de medição mais recente com mais de um milhão de assinantes de redes celulares móveis nos Estados Unidos descobriu que estes encontram sinais 4G mais de 90% do tempo, com velocidades de download de 20 Mbits/s ou maiores. Os usuários das três maiores operadoras da Coreia do Sul encontram um sinal 4G entre 95 e 99,5% do tempo (Open Signal, 2019). Por consequência, hoje é normal assistir streaming de vídeo em HD ou participar de

videoconferências no carro, ônibus ou trem de alta velocidade. A onipresença do acesso à Internet 4G também permitiu o surgimento de inúmeras novas aplicações da IoT, como patinetes e bicicletas compartilhadas ligadas à Internet, e aplicativos para smartphones, como pagamentos móveis (comuns na China desde 2018) e mensagens de texto via Internet (WeChat, WhatsApp e mais).

O termo *celular* refere-se ao fato de que uma área geográfica é dividida em várias áreas de cobertura geográfica, conhecidas como **células**. Cada célula contém uma **estação-base** que transmite e recebe sinais de **dispositivos móveis** dentro de sua célula. A área de cobertura de uma célula depende de muitos fatores, incluindo potência de transmissão da estação-base e de aparelhos do usuário, obstáculos na célula e altura e tipo das antenas da estação-base.

Nesta seção, apresentamos uma visão geral das redes celulares 4G (corrente) e 5G (emergente). Consideraremos o primeiro salto sem fio entre o dispositivo móvel e a estação-base, assim como o núcleo da rede toda em IP da operadora de celular que conecta o primeiro salto sem fio à rede da operadora, outras redes de operadoras e a Internet como um todo. Talvez surpreenda saber (dadas as origens das redes celulares móveis no mundo da telefonia, que tinha arquitetura de rede *muito* diferente da Internet) que encontraremos muitos dos mesmos princípios arquitetônicos nas redes 4G que vimos nos estudos focados na Internet nos Capítulos 1 a 6, incluindo camadas de protocolos, distinção entre borda e núcleo, interconexão de múltiplas redes de provedores para formar uma "rede de redes" global e a delimitação clara entre planos de dados e de controle com controle logicamente centralizado. Agora veremos esses princípios pela ótica das redes celulares móveis (e não pela ótica da Internet), e logo encontraremos realizações diferentes desses princípios. E, claro, como a rede da operadora possui um núcleo todo em IP, também encontraremos muitos dos protocolos da Internet que conhecemos tão bem. Analisaremos tópicos 4G adicionais (gerenciamento da mobilidade na Seção 7.6 e segurança 4G na Seção 8.8) posteriormente, após desenvolvermos os princípios básicos necessários para esses tópicos.

Nossa discussão sobre as redes 4G e 5G será relativamente breve. As redes celulares móveis são uma área de grande amplitude e profundidade, e muitas universidades oferecem diversos cursos sobre o assunto. O leitor que desejar um conhecimento mais profundo da questão deve consultar Goodman (1997); Kaaranen (2001); Lin (2001); Korhonen (2003); Schiller (2003); Scourias (2012); Turner (2012) e Akyildiz (2010), bem como as referências particularmente excelentes e completas em Mouly (1992) e Sauter (2014).

Assim como os RFCs da Internet definem os protocolos e arquitetura padrão da Internet, as redes 4G e 5G também são definidas por documentos padrão, chamados de Especificações Técnicas (*Technical Specifications*). Esses documentos estão disponíveis gratuitamente online em (3GPP, 2020). Assim como os RFCs, as especificações técnicas são um material denso e detalhado, difícil de ler. Mas quando você tem uma pergunta, elas são a fonte definitiva para as respostas!

7.4.1 Redes celulares 4G LTE: arquitetura e elementos

Em 2020, quando este livro estava sendo escrito, as redes 4G eram amplamente disseminadas e implantavam o padrão 4G Long-Term Evolution, mais conhecido como **4G LTE**. Nesta seção, descreveremos as redes 4G LTE. A Figura 7.17 mostra os principais elementos da arquitetura de rede 4G LTE. Em linhas gerais, a rede se divide em uma rede de rádio na borda da rede celular e o núcleo da rede. Todos os elementos da rede se comunicam entre si usando o protocolo IP que estudamos no Capítulo 4. Assim como as redes 2G e 3G anteriores, o 4G LTE é repleto de siglas e nomes de elementos obscuros. Para tentar desfazer esse nó, primeiro enfocaremos as funções dos elementos e como os diversos elementos da rede 4G LTE interagem entre si no plano de dados e no de controle.

- **Dispositivo móvel.** Este é um smartphone, tablet, notebook ou aparelho IoT que se conecta à rede de uma operadora de celular. É aqui que são executadas aplicações como

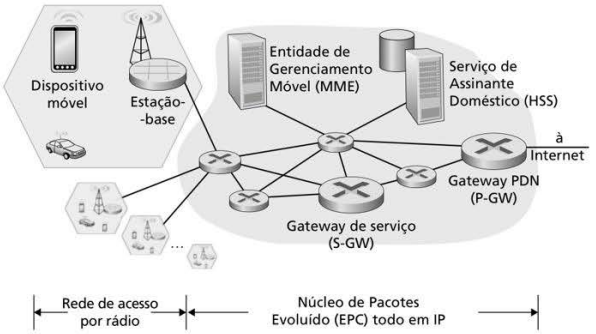


Figura 7.17 Elementos da arquitetura 4G LTE.

navegadores, mapas, voz, videoconferência e muito, muito mais. Em geral, o dispositivo móvel implementa todas as cinco camadas da pilha de protocolos da Internet, incluindo as camadas de transporte e aplicações, como vimos com os hospedeiros na borda da rede da Internet. O dispositivo móvel é um ponto final da rede, com um endereço IP (obtido por NAT, como veremos). O dispositivo móvel também possui um identificador de 64 bits globalmente único chamado de **identidade internacional de assinante móvel** (**IMSI**, do inglês ***International Mobile Subscriber Identity***), armazenado no seu cartão SIM (do inglês *Subscriber Identity Module* – módulo de identificação do assinante). A IMSI identifica o assinante no sistema mundial de redes de operadoras de celular, incluindo o país e a rede nativa da operadora à qual o assinante pertence. Em certos sentidos, a IMSI é análoga a um endereço MAC. O cartão SIM também armazena informações sobre os serviços que o assinante pode acessar e informações da chave criptográfica referentes ao assinante. No jargão oficial do 4G LTE, o dispositivo móvel é chamado de **Equipamento do Usuário** (**UE**, do inglês ***User Equipment***). Contudo, neste livro, usaremos o termo "dispositivo móvel", que é mais simpático. Também observamos aqui que um dispositivo móvel não é sempre móvel; por exemplo, ele pode ser um sensor de temperatura fixo ou uma câmera de vigilância.

- **Estação-base.** A estação-base fica na "borda" da rede da operadora e é responsável por gerenciar os recursos de rádio sem fio e os dispositivos móveis na sua área de cobertura (mostrada como uma célula hexagonal na Figura 7.17). Como veremos, o dispositivo móvel interage com uma estação-base para se ligar à rede da operadora. A estação-base coordena a autenticação do dispositivo e a alocação de recursos (acesso a canais) na rede de acesso por rádio. Nesse sentido, a estação-base celular funciona de forma comparável (mas absolutamente não idêntica) aos APs em LANs sem fio. Mas as estações-base celulares têm vários outros papéis importantes que não ocorrem nas LANs sem fio. Em especial, as estações-base criam túneis de IP específicos para cada dispositivo que vão destes aos gateways e interagem entre si para gerenciar a mobilidade de dispositivos entre células. Na terminologia oficial do 4G LTE, a estação-base é chamada de "**eNode-B**", um termo um tanto opaco e pouco descritivo. Neste livro, daremos preferência ao termo "estação-base", por ser mais fácil para o leitor.

  Uma breve digressão: se você acha a terminologia do LTE um tanto obscura, não está sozinho! A etimologia de "eNode-B" vem da terminologia anterior do 3G, na qual os pontos de função da rede eram chamados de "nós" (*nodes*), com "B" se referindo à terminologia anterior do 1G, de "estação-base (BS, do inglês *Base Station*)" ou "Base Transceiver Station (BTS)" na terminologia do 2G. O 4G LTE é uma "e"volução em

relação ao 3G, de onde vem o "e" antes de "Node-B". E esse obscurantismo na nomenclatura não dá nenhum sinal de parar: Nos sistemas 5G, as funções do eNode-B agora são chamadas de "ng-eNB". Só tente adivinhar o que significa essa sigla!

- **Serviço de Assinante Doméstico (HSS**, do inglês ***Home Subscriber Server***). Como mostrado na Figura 7.18, o HSS é um elemento do plano de controle. O HSS é um banco de dados que armazena informações sobre os dispositivos móveis para os quais o HSS é a sua rede nativa. Ele é usado junto com a MME (discutida abaixo) para a autenticação do dispositivo.
- **Gateway de serviço (S-GW**, do inglês ***Serving Gateway***), **Gateway da rede de pacote de dados (P-GW**, do inglês ***Packet Data Network Gateway***) **e outros roteadores de rede.** Como mostrado na Figura 7.18, o gateway de serviço e o gateway da rede de pacote de dados (PDN) são dois roteadores (muitas vezes, colocalizados na prática) que ficam no caminho de dados entre o dispositivo móvel e a Internet. O gateway de PDN também oferece endereços IP NAT para dispositivos móveis e cumpre funções de tradução de endereços de rede (NAT, do inglês *network address translation*) (ver Seção 4.3.4). O gateway de PDN é o último elemento do LTE que um datagrama originário de um dispositivo móvel encontra antes de entrar na Internet como um todo. Para o mundo externo, o P-GW se parece com todos os outros roteadores de borda; a mobilidade dos nós móveis dentro da rede LTE da operadora de celular é ocultada do mundo externo pelo P-GW. Além desses roteadores de borda, o núcleo todo em IP da operadora terá roteadores adicionais cuja função é semelhante à dos roteadores IP tradicionais: repassar datagramas IP entre si ao longo de caminhos que normalmente terminam em elementos do núcleo da rede LTE.
- **Entidade de gerenciamento móvel (MME**, do inglês ***Mobility Management Entity***). A MME também é um elemento do plano de controle, como mostra a Figura 7.18. Junto com o HSS, tem um papel importante em autenticar um dispositivo que deseja se conectar à sua rede. Ela também estabelece os túneis no caminho de dados entre o dispositivo e o roteador de borda da Internet PDN e mantém informações sobre a localização celular de um dispositivo móvel ativo dentro da rede da operadora. Mas, como mostra a Figura 7.18, ela não fica no caminho do repasse para os datagramas do dispositivo móvel enviados de e para a Internet.
  - *Autenticação.* É importante que a rede e o dispositivo móvel que se liga a ela se autentiquem *mutuamente*, pois a rede precisa saber que o dispositivo que se liga a ela é mesmo aquele associado a uma determinada IMSI, e o dispositivo móvel precisa saber que a rede à qual está se ligando é também a rede legítima de uma operadora de celular. Trabalharemos a autenticação no Capítulo 8 e a autenticação 4G na Seção 8.8. Aqui, observamos simplesmente que a MME atua como intermediário entre o dispositivo móvel e o serviço de assinante doméstico (HSS) na rede nativa do dispositivo móvel. Mais especificamente, após receber uma requisição de ligação do dispositivo móvel, a MME local contata o HSS na rede nativa do dispositivo. A seguir, o HSS da rede nativa retorna informações criptografadas suficientes para a MME local para provar para o dispositivo móvel que o HSS nativo está realizando a autenticação através dessa MME e para que o dispositivo móvel prove para a MME que é mesmo o dispositivo associado àquela IMSI. Quando um dispositivo móvel se liga à sua rede nativa, o HSS a ser contatado durante a autenticação é localizado na mesma rede nativa. Contudo, quando um dispositivo móvel está em roaming em uma rede visitada por uma operadora de celular diferente, a MME na rede transitada precisará contatar o HSS na rede nativa do dispositivo móvel.
  - *Estabelecimento do caminho.* Como mostrado na metade inferior da Figura 7.18, o caminho de dados do dispositivo móvel até o roteador de borda da operadora é composto por um primeiro salto sem fio entre o dispositivo móvel e a estação-base, e túneis IP concatenados entre a estação-base e o gateway de serviço e entre o gateway de serviço e o gateway de PDN. Os túneis são configurados sob o controle da MME e usados para repasse de dados (e não repasse direto entre roteadores de rede) para facilitar a mobilidade do dispositivo – quando um dispositivo se movimenta, apenas a extremidade do túnel que termina na estação-base precisa ser alterada, enquanto as outras extremidades do túnel e a qualidade de serviço associada ao túnel permanecem inalteradas.
  - *Rastreamento de local de celular.* À medida que o dispositivo se move entre células, as estações-base atualizam a MME sobre a localização do dispositivo. Se o dispositivo móvel está em modo de suspensão, mas ainda assim se move entre as células, as estações-base não podem mais rastrear a localização do dispositivo. Nesse caso, é responsabilidade da MME localizar o dispositivo para despertá-lo, usando um processo conhecido por **paginação.**

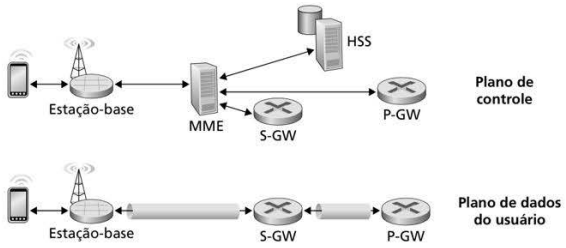


Figura 7.18 Elementos do plano de dados e do plano de controle LTE.

A Tabela 7.2 resume os principais elementos arquitetônicos da LTE discutidos acima e compara essas funções com aquelas que encontramos no nosso estudo das LANs sem fio WiFi (WLANs).

TABELA 7.2 Elementos LTE e funções semelhantes da WLAN (WiFi)

Elemento LTE	Descrição	Função(ões) semelhante(s) da WLAN
Dispositivo móvel (UE: equipamento do usuário)	Dispositivo móvel/sem fio habilitado para IP do usuário final (p. ex., smartphone, tablet, notebook)	Hospedeiro, sistema final
Estação-base (eNode-B)	Lado da rede do enlace de acesso sem fio à rede LTE	Ponto de acesso (AP), apesar de a estação-base LTE realizar muitas funções não encontradas em WLANs
A Entidade de Gerenciamento Móvel (MME)	Coordenador para serviços de dispositivos móveis: autenticação, gerenciamento da mobilidade	Ponto de acesso (AP), apesar de a MME realizar muitas funções não encontradas em WLANs
Serviço de assinante doméstico (HSS)	Localizado na rede *nativa* do dispositivo móvel, oferece privilégios de acesso e autenticação nas redes nativa e visitada	Sem equivalente na WLAN
Gateway de serviço (S-GW), Gateway de PDN (P-GW)	Roteadores na rede da operadora de celular, coordenam o repasse para fora da rede da operadora	Roteadores iBGP e eBGP na rede do provedor local
Rede de acesso por rádio	Enlace sem fio entre dispositivo móvel e estação-base	Enlace sem fio 802.11 entre o dispositivo móvel e o AP

HISTÓRICO DO CASO

A EVOLUÇÃO DA ARQUITETURA DE 2G PARA 3G PARA 4G

Em um período relativamente curto de 20 anos, as redes das operadoras celulares completaram uma transição incrível, passando quase exclusivamente de redes por comutação de circuitos para redes de dados por comutação de pacotes todas em IP que incluem a voz como apenas uma entre as muitas aplicações. Como foi essa transição do ponto de vista da arquitetura? Houve um "dia da conversão" (*flag day*), no qual as redes anteriores, orientadas para telefonia, foram "desligadas" e as redes celulares todas em IP foram "ligadas"? Ou elementos das redes anteriores começaram a adotar funcionalidades duplas de circuitos (legado) e pacotes (novo), como vimos na transição do IPv4 para o IPv6 na Seção 4.3.5?

A Figura 7.19 veio da 7ª edição deste livro, que analisava as redes celulares 2G e 3G (esse material histórico foi aposentado e substituído por uma cobertura mais profunda do 4G LTE nesta 8ª edição, mas ainda está disponível no site do livro, em inglês). Apesar de a rede 2G ser uma rede de telefonia móvel por comutação de circuitos, a comparação entre as Figuras 7.17 e 7.19 ilustra uma estrutura conceitual semelhante, ainda que para serviços de voz, não de dados: uma borda sem fio controlada por uma estação-base, um gateway da rede da operadora para o mundo externo e pontos de agregação entre as estações-base e o gateway.

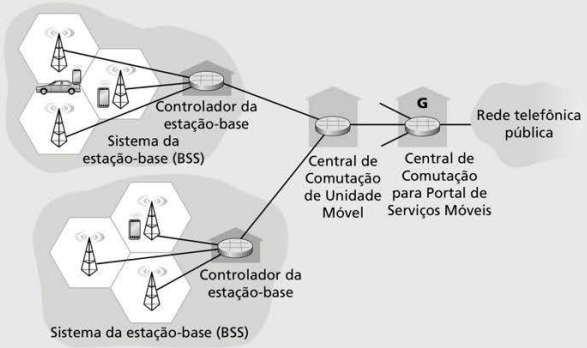


Figura 7.19 Elementos da arquitetura celular 2G, que suportam serviço de voz por comutação de circuitos com o núcleo da rede da operadora.

A Figura 7.20 (também advinda da 7ª edição deste livro) mostra os principais componentes arquitetônicos da arquitetura celular 3G, que suporta o serviço por comutação de circuitos e serviços de dados por comutação de pacotes. Aqui, fica clara a transição de uma rede apenas de voz para uma rede que combinava voz e dados: os elementos nucleares existentes da rede de voz celular 2G permaneceram intocados. *Contudo, funcionalidades de dados celulares adicionais foram agregadas em paralelo e funcionavam de forma independente de, ao núcleo existente da rede de voz na época.* Como mostra a Figura 7.20, o ponto de divisão para esses dois núcleos de rede independentes, voz e dados, ocorreu na borda da rede, na estação-base na rede de acesso por rádio. A alternativa, de integrar novos serviços de dados diretamente aos elementos nucleares da rede celular de voz existente, teria provocado os mesmos desafios encontrados na integração de tecnologias novas (IPv6) e de legado (IPv4) na Internet. As operadoras também queriam aproveitar e explorar seu investimento significativo na infraestrutura existente (e seus serviços lucrativos) na rede celular de voz existente.

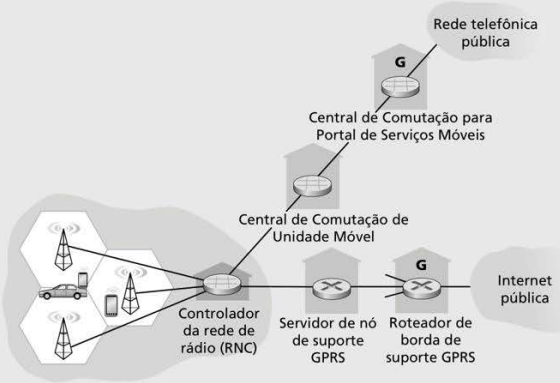


Figura 7.20 Arquitetura do sistema 3G: suporte a serviço de voz por comutação de circuitos e serviço de dados por comutação de pacotes com o núcleo da rede da operadora.

7.4.2 Pilhas de protocolos LTE

Como a arquitetura 4G LTE é toda em IP, já conhecemos os protocolos das camadas superiores da pilha LTE, especialmente os protocolos IP, TCP, UDP e os diversos protocolos da camada de aplicação, dados os nossos estudos nos Capítulos 2 a 5. Por consequência, os novos protocolos LTE que analisaremos aqui se encontram principalmente nas camadas de enlace e física e no gerenciamento da mobilidade.

A Figura 7.21 mostra as pilhas de protocolos do plano do usuário no nó móvel LTE, a estação-base e o gateway de serviço. Entraremos em diversos dos protocolos do plano de controle da LTE posteriormente, quando estudarmos o gerenciamento da mobilidade (Seção 7.6) e segurança (Seção 8.8) LTE. Como vemos na Figura 7.21, a maioria das

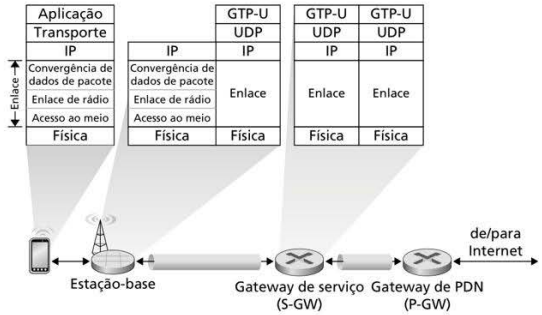


Figura 7.21 Pilhas de protocolos do plano de dados LTE.

atividades de protocolos do plano do usuário novas e interessantes ocorre no enlace de rádio sem fio entre o dispositivo móvel e a estação-base.

O LTE divide a camada de enlace do dispositivo móvel em três subcamadas:

- *Convergência de dados de pacote.* É a subcamada superior da camada de enlace, logo abaixo do IP. O protocolo de convergência de dados de pacote (PDCP, do inglês *Packet Data Convergence Protocol*) (3GPP PDCP, 2019) realiza a compressão do cabeçalho IP para reduzir o número de bits enviados pelo enlace sem fio e a criptografia/decriptação do datagrama IP usando chaves estabelecidas por mensagens de sinalização entre o dispositivo móvel LTE e a MME quando o dispositivo móvel se liga à rede inicialmente; analisaremos aspectos da segurança do LTE na Seção 8.8.2.
- *Controle de enlace de rádio.* O protocolo de controle de enlace de rádio (RLC, do inglês *Radio Link Control*) (3GPP RLCP, 2018) desempenha duas funções importantes: *(i)* fragmentar (no lado do remetente) e remontar (no lado do destinatário) os datagramas IP grandes demais para caberem nos quadros da camada de enlace subjacentes; e *(ii)* transferência confiável de dados da camada de enlace pelo uso de um protocolo ARQ baseado em ACK/NAK. Lembre-se que estudamos os elementos básicos dos protocolos ARQ na Seção 3.4.1.
- *Controle de acesso ao meio (MAC).* A camada MAC é responsável pelo escalonamento da transmissão, ou seja, a solicitação e o uso dos intervalos de transmissão por rádio descritos na Seção 7.4.4. A subcamada MAC também tem funções adicionais de detecção e correção de erros, incluindo o uso de transmissão redundante de bits como técnica de correção de erros antecipada. O nível de redundância pode ser adaptado às condições do canal.

A Figura 7.21 também mostra o uso de túneis no caminho de dados do usuário. Como discutido acima, esses túneis são estabelecidos, sob controle da MME, quando o dispositivo móvel se liga à rede inicialmente. Cada túnel entre dois pontos finais possui um identificador de ponto final do túnel (TEID, do inglês *tunnel endpoint identifier*) único. Quando recebe datagramas do dispositivo móvel, a estação-base encapsula-os usando o protocolo de túnel GPRS (3GPP GTPv1-U, 2019), incluindo o TEID, e envia-os nos segmentos UDP para o Gateway de Serviço na outra extremidade do túnel. No lado do receptor, a estação-base desencapsula os datagramas UDP enviados por túnel, extrai o datagrama IP encapsulado destinado para o dispositivo móvel e repassa o datagrama IP pelo enlace sem fio até o dispositivo móvel.

7.4.3 Rede de acesso por rádio LTE

O padrão LTE usa uma combinação de multiplexação por divisão de frequência e multiplexação por divisão de tempo no canal descendente, conhecida como multiplexação por divisão de frequência ortogonal (OFDM, do inglês *orthogonal frequency division multiplexing*) (Hwang, 2009). (O termo "ortogonal" vem do fato de que os sinais enviados em diferentes canais de frequência são criados de modo que interfiram muito pouco uns nos outros, mesmo quando as frequências de canal são pouco espaçadas.) No LTE, cada nó móvel ativo recebe um ou mais intervalos de tempo de 0,5 ms em uma ou mais das frequências do canal. A Figura 7.22 mostra a alocação de oito intervalos de tempo sobre quatro frequências. Recebendo cada vez mais intervalos de tempo (seja na mesma frequência ou em frequências diferentes), um dispositivo nó móvel pode alcançar velocidades de transmissão cada vez mais altas. A (re)alocação de intervalo entre os dispositivos móveis pode ser realizada até mesmo a cada milissegundo. Diferentes esquemas de modulação também podem ser usados para alterar a taxa de transmissão; veja nossa discussão anterior da Figura 7.3 e a seleção dinâmica de esquemas de modulação em redes WiFi.

A alocação em particular de intervalos de tempo a dispositivos móveis não é exigida pelo padrão LTE. Em vez disso, a decisão de quais dispositivos móveis terão permissão para

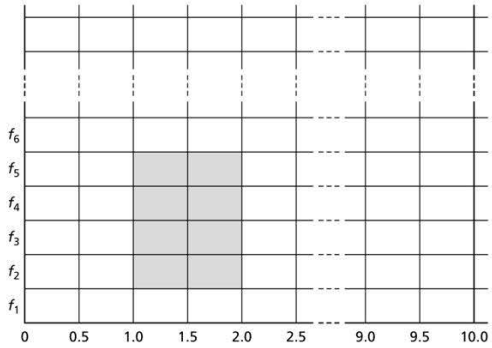


Figura 7.22 Vinte intervalos de 0,5 milissegundos organizados em quadros de 10 milissegundos em cada frequência. A região sombreada indica uma alocação de oito intervalos.

transmitir em dado intervalo de tempo em dada frequência é determinada pelos algoritmos de escalonamento fornecidos pelo fornecedor de equipamento LTE e/ou operador da rede. O escalonamento oportunista (Bender, 2000; Kolding, 2003; Kulkarni, 2005), combinando o protocolo da camada física e as condições do canal entre remetente e destinatário e escolhendo os destinatários aos quais os pacotes serão enviados com base nas condições do canal, permite que a estação-base faça o melhor uso do meio sem fio. Além disso, as prioridades do usuário e os níveis de serviço contratados (p. ex., prata, ouro ou platina) podem ser usados no escalonamento das transmissões descendentes de pacotes. Além das capacidades do LTE descritas aqui, LTE-Advanced permite larguras de banda descendentes de centenas de Mbits/s, com a alocação de canais agregados a um dispositivo móvel (Akyildiz, 2010).

7.4.4 Funções adicionais do LTE: ligação à rede e gerenciamento de energia

Vamos concluir nosso estudo sobre o 4G LTE com a consideração de duas funções adicionais importantes desse padrão: *(i)* o processo pelo qual um dispositivo móvel se liga inicialmente à rede e *(ii)* as técnicas usadas pelo dispositivo móvel, em conjunto com elementos do núcleo da rede, para gerenciar o seu consumo de energia.

Ligação à rede

O processo pelo qual um dispositivo móvel se liga à rede da operadora de celular se divide, em linhas gerais, em três fases:

- *Ligação a uma estação-base.* A primeira fase da ligação do dispositivo é semelhante em propósito ao protocolo de associação 802.11 que estudamos na Seção 7.31, mas muito diferente dele na prática. Um dispositivo móvel que deseja se ligar à rede de uma operadora de celular começa um processo de inicialização (*bootstrapping*) para aprender sobre uma estação-base próxima e então se associar a ela. Inicialmente, o dispositivo móvel procura todos os canais, em todas as bandas de frequência, em busca de um sinal de sincronização primário transmitido periodicamente a cada 5 ms pela estação-base. Após encontrar o sinal, o dispositivo móvel permanece nessa frequência e localiza o sinal de sincronização secundário. Usando as informações obtidas com esse segundo

sinal, o dispositivo pode localizar (após vários passos subsequentes) informações adicionais, como a largura de banda do canal, configurações do canal e as informações da operadora de celular da estação-base. Armado com essas informações, o dispositivo móvel pode selecionar uma estação-base com a qual se associar (se ligando preferencialmente à rede nativa, se disponível) e estabelecer uma conexão de sinalização do plano de controle com a estação-base através do enlace sem fio. Esse canal entre o dispositivo móvel e a estação-base será usado no restante do processo de ligação à rede.

- *Autenticação mútua.* Na nossa descrição anterior sobre a entidade de gerenciamento móvel (MME) na Seção 7.4.1, observamos que a estação-base contata a MME local para realizar autenticação mútua, um processo que estudaremos em mais detalhes na Seção 8.8.2. É a segunda fase da ligação à rede, que permite que esta saiba que o dispositivo que se liga a ela é mesmo aquele associado a uma determinada IMSI, e que o dispositivo móvel saiba que a rede à qual está se ligando também é a rede de uma operadora de celular legítima. Completada essa segunda fase da ligação à rede, a MME e o dispositivo móvel se autenticaram mutuamente, e a MME sabe também a identidade da estação-base à qual o dispositivo móvel está ligado. Armada com essas informações, a MME está então pronta para configurar o caminho de dados de dispositivo móvel para o gateway de PDN.
- *Configuração de caminho de dados de dispositivo móvel para o gateway de PDN.* A MME contata o gateway de PDN (que também fornece um endereço NAT para o dispositivo móvel), o gateway de serviço e a estação-base para estabelecer os dois túneis mostrados na Figura 7.21. Após esta fase estar completa, o dispositivo móvel consegue enviar/receber datagramas IP por meio da estação-base através desses túneis de e para a Internet!

Gerenciamento de energia: Modos de suspensão

Na nossa discussão anterior sobre os recursos avançados das redes 802.11 (Seção 7.3.5) e Bluetooth (Seção 7.3.6), vimos que um rádio em um dispositivo sem fio pode entrar em estado de suspensão para poupar energia enquanto não transmite ou recebe, minimizando o tempo que os circuitos do dispositivo móvel precisam estar "ligados" para enviar/receber dados e para detecção de canais. No 4G LTE, um dispositivo móvel em modo de suspensão pode estar em um de dois estados de suspensão diferentes. No estado de recepção descontínua, normalmente ativado após várias centenas de milissegundos de inatividade (Sauter, 2014), o dispositivo móvel e a estação-base agendam de antemão momentos periódicos (em geral, com várias centenas de milissegundos de distância entre si) nos quais o dispositivo móvel acordará para monitorar ativamente o canal em busca de transmissões *downstream* (da estação-base para o dispositivo móvel); fora desses momentos programados, no entanto, o rádio do dispositivo móvel fica suspenso.

Se o estado de recepção descontínua pode ser considerado um "sono leve", o segundo estado de suspensão, o estado inativo, que ocorre após períodos mais longos, de 5 a 10 segundos de inatividade, pode ser considerado um "sono profundo". Nesse estado, o rádio do dispositivo móvel acorda e monitora o canal com ainda menos frequência. Na verdade, esse sono é tão profundo que, se entra em uma nova célula da rede da operadora nessa situação, o dispositivo móvel não precisa informar a estação-base à qual estava associado anteriormente. Assim, quando acorda periodicamente desse sono profundo, o dispositivo móvel precisa restabelecer uma associação com uma estação-base (possivelmente nova) de modo a verificar as mensagens de paginação transmitidas pela MME para as estações-base próximas à última estação-base à qual o dispositivo móvel se associou. Essas mensagens de paginação do plano de controle, transmitidas por difusão por essas estações-base para todos os dispositivos móveis nas suas células, indicam quais dispositivos móveis devem acordar completamente e restabelecer uma nova conexão no plano de dados com uma estação-base (ver Figura 7.18) de modo a receber os pacotes enviados.

7.4.5 A rede celular global: uma rede de redes

Tendo estudado a arquitetura da rede celular 4G, vamos nos afastar um pouco e analisar como se organiza a rede celular global, ela própria uma "rede de redes", assim como a Internet.

A Figura 7.23 mostra o smartphone de um usuário conectado através de uma estação-base 4G à sua **rede nativa**. A rede móvel nativa do usuário é operada por uma operadora de celular como Verizon, AT&T, T-Mobile ou Sprint, nos Estados Unidos; Orange, na França; SK Telecom, na Coreia do Sul; ou Vivo, Claro ou TIM, no Brasil. A rede nativa do usuário, por sua vez, está conectada às redes de outras operadoras de celular e à Internet global por meio de um ou mais roteadores de borda na rede nativa, como mostrado na Figura 7.23. As redes móveis em si se interconectam umas às outras através da Internet pública ou de uma rede IPX (Internet Protocol Packet eXchange) (GSMA, 2018a). Uma IPX é uma rede gerenciada especificamente para interconectar operadoras de celular, semelhante aos pontos de troca da Internet (IXPs, do inglês *Internet exchange points*; ver Figura 1.15) para parcerias (*peering*) entre ISPs. Na Figura 7.23, vemos que a rede celular global é mesmo uma "rede de redes", assim como a Internet (lembre-se da Figura 1.15 e da Seção 5.4). As redes 4G também podem formar parcerias com redes de voz/dados celulares 3G e com as redes anteriores apenas para voz.

Voltaremos em breve a tópicos adicionais do 4G LTE (gerenciamento da mobilidade na Seção 7.6 e segurança no 4G na Seção 8.8.2) após desenvolvermos os princípios básicos necessários para esses tópicos. Agora vamos examinar rapidamente as redes 5G emergentes.

7.4.6 Redes celulares 5G

A versão definitiva do serviço de dados de longa distância teria velocidades de conexão de Gbit/s onipresentes, latência extremamente baixa e zero limitações no número de usuários e dispositivos suportados em cada região. Um serviço como esse abriria as portas para

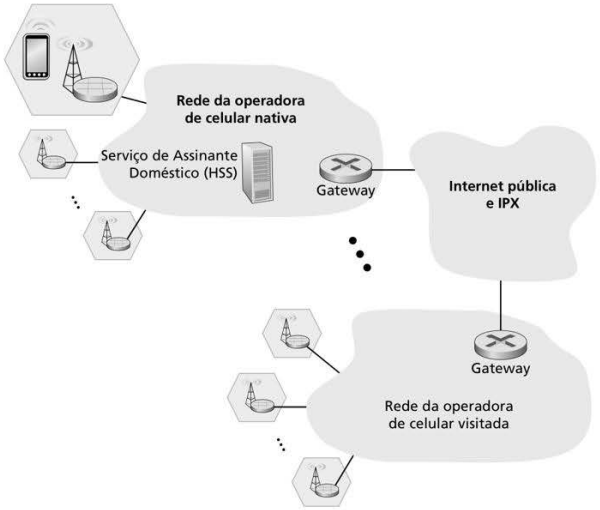


Figura 7.23 A rede de dados celulares global: uma rede de redes.

inúmeras novas aplicações, incluindo realidade aumentada e realidade virtual disseminadas, controle de veículos autônomos usando conexões sem fio, controle de robôs em fábricas por conexões sem fio e substituição das tecnologias de acesso residencial, como DSL e cabo, por serviços de Internet sem fio fixa (i.e., conexões sem fio residenciais a partir de estações-base para modems domésticos).

Espera-se que o 5G, para o qual versões progressivamente melhores provavelmente serão implantadas na década de 2020, dê um grande passo na direção de atingir os objetivos desse serviço definitivo de dados de longa distância. Prevê-se que o 5G levará um aumento de cerca de 10x na taxa de bits máxima, uma redução de 10x na latência e um aumento de 100x na capacidade de tráfego em relação ao 4G (Qualcomm, 2019).

A sigla 5G se refere principalmente a "5G NR (*New Radio*)," que é o padrão adotado pelo 3GPP. Existem outras tecnologias 5G além do NR, no entanto. Por exemplo, a rede proprietária 5G TF da Verizon opera nas frequências 28 e 39 GHz e é usada apenas para serviço de Internet sem fio fixa, não em smartphones.

Os padrões 5G dividem as frequências em dois grupos: FR1 (450 MHz–6 GHz) e FR2 (24 GHz–52 GHz). As primeiras implementações ocorreram no espaço FR1, apesar de, já em 2020, terem ocorrido implementações iniciais no espaço FR2 para acesso fixo à Internet residencial, como mencionado acima. É importante ressaltar que os aspectos da camada física (i.e., sem fio) do 5G *não* têm compatibilidade reversa com sistemas de comunicação móvel 4G, como o LTE: em especial, eles não podem ser entregues aos smartphones existentes com atualizações das estações-base ou de software. Assim, na transição para o 5G, as operadoras sem fio precisarão realizar investimentos significativos na infraestrutura física.

As frequências FR2 também são chamadas de **frequências de ondas milimétricas**. Apesar de permitirem velocidades de dados muito maiores, as frequências de ondas milimétricas têm duas desvantagens importantes:

- As frequências de ondas milimétricas têm alcance muito menor entre a estação-base e os receptores. Assim, a tecnologia de ondas milimétricas não é apropriada para zonas rurais e exige a instalação de uma maior densidade de estações-base em áreas urbanas.
- Comunicações por ondas milimétricas são altamente suscetíveis à interferência atmosférica. A folhagem nas proximidades e a chuva podem criar problemas para o uso em áreas externas.

O 5G não é um padrão coeso, sendo na verdade composto por três padrões coexistentes (Dahlman, 2018):

- *eMBB* (do inglês *Enhanced Mobile Broadband – banda larga móvel melhorada*). As implantações iniciais do 5G NR se concentraram no eMBB, que permite maior largura de banda para velocidades de download e upload maiores, além de uma redução moderada na latência em comparação com o 4G LTE. O eMBB permite aplicações de mídia interativa, tais como realidade virtual e realidade aumentada móvel, assim como resolução 4K móvel e streaming de vídeo 360°.
- *URLLC* (do inglês *Ultra Reliable Low-Latency Communications – comunicações ultraconfiáveis de baixa latência*). O URLLC é direcionado para aplicações com alta sensibilidade à latência, como automação industrial e veículos autônomos. O URLLC busca latências de 1 ms. Na época da redação deste livro, as tecnologias que permitem o URLLC ainda estavam sendo padronizadas.
- *mMTC* (do inglês *Massive Machine Type Communications – comunicações massivas de tipo de máquina*). O mMTC é um tipo de acesso de banda estreita para aplicações de detecção, medição e monitoramento. Uma prioridade no projeto das redes 5G é reduzir as barreiras à conectividade de rede para dispositivos IoT. Além de reduzir a latência, as tecnologias emergentes para as redes 5G se concentram na redução dos requisitos de energia, tornando o uso de dispositivos IoT mais disseminado do que ocorreu com o 4G LTE.

5G e frequências de ondas milimétricas

Muitas inovações do 5G serão o resultado direto do trabalho nas frequências de ondas milimétricas na banda de 24 a 52 GHz. Por exemplo, essas frequências oferecem o potencial de multiplicar por 100 a capacidade em relação ao 4G. Para entender melhor esse fenômeno, a capacidade pode ser definida como o produto de três termos (Björnson, 2017):

$$\text{capacidade} = \text{densidade celular} \times \text{espectro disponível} \times \text{eficiência espectral}$$

em que a unidade da densidade celular é em células/km^2, a do espectro é Hertz, e a eficiência espectral é uma medida da eficiência com a qual a estação-base consegue se comunicar com os usuários, cuja unidade é bits/s/Hz/célula. Multiplicando essas unidades, é fácil ver que a unidade de capacidade é de bits/s/km^2. Para cada um desses três termos, os valores serão maiores para o 5G do que para o 4G:

- Como as frequências de ondas milimétricas têm alcance muito menor do que as frequências 4G LTE, mais estações-base são necessárias, o que, por sua vez, aumenta a densidade celular.
- Como o 5G FR2 opera em uma banda de frequência muito maior (52 − 24 = 28 GHz) do que o 4G LTE (até cerca de 2 GHz), o espectro disponível é maior.
- Com relação à eficiência espectral, a teoria da informação afirma que para dobrar a eficiência espectral, é preciso multiplicar a potência por 17 (Björnson, 2017). Em vez de aumentar a potência, o 5G usa tecnologia MIMO (a mesma que encontramos no nosso estudo sobre as redes 802.11 na Seção 7.3), que utiliza múltiplas antenas em cada estação-base. Em vez de transmitir o sinal por difusão em todas as direções, cada antena MIMO emprega **formação de feixes** (*beam forming*) e direciona o sinal para o usuário. A tecnologia MIMO permite que uma estação-base transmita para 10 a 20 usuários ao mesmo tempo na mesma banda de frequência.

Ao aumentar todos os três termos na equação de capacidade, espera-se que o 5G multiplique por 100 a capacidade nas áreas urbanas. Da mesma forma, devido à largura muito maior da banda de frequência, espera-se que o 5G ofereça taxas de download máximas de 1 Gbit/s ou maiores.

Contudo, os sinais de ondas milimétricas são facilmente bloqueados por árvores e edifícios. **Estações celulares pequenas** são necessárias para "tapar os buracos" entre as estações-base e os usuários. Em uma região muito populosa, a distância entre duas células pequenas pode variar de 10 a 100 metros (Dahlman, 2018).

O núcleo da rede 5G

O **núcleo da rede 5G** é a rede de dados que gerencia todas as conexões móveis 5G de voz, dados e Internet. O núcleo da rede 5G está sendo reprojetado para se integrar melhor com serviços de Internet e baseados em nuvem, e inclui caches e servidores distribuídos espalhados pela rede, o que reduz a latência. A virtualização da função de rede (discutida nos Capítulos 4 e 5) e o fatiamento da rede (*network slicing*) para diferentes aplicações e serviços serão gerenciados no núcleo.

A nova especificação do núcleo 5G introduz mudanças importantes no modo como as redes móveis suportam uma ampla variedade de serviços, com diversos níveis de desempenho. Assim como no caso do núcleo da rede 4G (lembre-se das Figuras 7.17 e 7.18), o núcleo 5G transmite tráfego de dados dos dispositivos finais, autentica dispositivos e gerencia a mobilidade dos dispositivos. O núcleo 5G também contém todos os elementos de rede que encontramos na Seção 7.4.2: os dispositivos móveis, as células, a estação-base e a entidade de gerenciamento móvel (agora dividida em dois subelementos, como discutido abaixo), o HSS e os gateways de serviço e de PDN.

Apesar de os núcleos 4G e 5G terem funções semelhantes, há diferenças importantes na nova arquitetura do núcleo 5G, que foi projetado para separação completa entre o plano de

controle e o plano do usuário (ver Capítulo 5). O núcleo 5G é totalmente composto por funções de rede baseadas em software virtualizadas. Essa nova arquitetura dará aos operadores a flexibilidade necessária para atender aos diversos requisitos de diferentes aplicações 5G. Algumas das funções do núcleo da rede 5G incluem (Rommer, 2019):

- *Função do plano do usuário (UPF, do inglês user-plane function).* A separação entre o plano de controle e o plano do usuário (ver Capítulo 5) permite que o processamento de pacotes seja distribuído e transferido para a borda da rede.
- *Função de gerenciamento de acesso e mobilidade (AMF, do inglês access and mobility management function).* O núcleo 5G essencialmente decompõe a entidade de gerenciamento móvel (MME) 4G em dois elementos funcionais: AMF e SMF. A AMF recebe todas as informações de sessão e conexão do equipamento do usuário final, mas só lida com tarefas de conexão e de gerenciamento da mobilidade.
- *Função de gerenciamento de sessão (SMF, do inglês session management function).* O gerenciamento de sessão é administrado pela função de gerenciamento de sessão (SMF). A SMF é responsável por interagir com o plano de dados desacoplado. A SMF também cuida do gerenciamento de endereços IP e desempenha o papel do DHCP.

Na época da redação deste livro (2020), o 5G estava nos seus estágios iniciais de implantação, e muitos dos padrões 5G ainda não haviam sido finalizados. Só o tempo dirá se o 5G se tornará um serviço de banda larga sem fio disseminado, se terá sucesso na concorrência com o WiFi para serviço sem fio em ambientes internos, se tornar-se-á um componente crítico da infraestrutura para veículos autônomos e da automação industrial e se nos ajudará a dar um grande passo adiante em direção a uma versão definitiva do serviço de dados de longa distância.

7.5 GERENCIAMENTO DA MOBILIDADE: PRINCÍPIOS

Após estudarmos a natureza sem fio dos enlaces de comunicação em uma rede sem fio, é hora de voltarmos nossa atenção à mobilidade que esses enlaces sem fio possibilitam. No sentido mais amplo, um dispositivo móvel é aquele que muda seu ponto de conexão com a rede ao longo do tempo. Como o termo mobilidade adquiriu muitos significados nos campos da computação e da telefonia, será proveitoso, antes de tudo, considerarmos formas de mobilidade.

7.5.1 Mobilidade de dispositivos do ponto de vista da camada de rede

Do ponto de vista da camada de rede, um dispositivo fisicamente móvel pode representar um conjunto muito diferente de desafios para ela, dependendo de quanto o dispositivo está ativo enquanto se move entre os pontos de ligação à rede. Em uma extremidade do espectro, o cenário *(a)* na Figura 7.24, é o próprio usuário móvel que fisicamente se move entre as redes, mas desliga o dispositivo móvel enquanto está em movimento. Por exemplo, um estudante poderia se desconectar de uma sala de aula sem fio e desligar seu dispositivo, ir para o refeitório e se conectar à rede de acesso sem fio enquanto come, então desconectar e desligar o aparelho dessa rede, caminhar até a biblioteca e conectá-lo à rede sem fio da biblioteca enquanto estuda. De uma perspectiva de rede, este dispositivo *não* é móvel, pois se liga a uma rede de acesso e permanece nela enquanto está ligado. Nesse caso, o dispositivo se associa serialmente, e se dissocia posteriormente de cada rede de acesso sem fio que encontra. Esse caso de (i)mobilidade de dispositivo pode ser gerenciada totalmente pelos mecanismos de rede que já estudamos nas Seções 7.3 e 7.4.

No cenário *(b)* da Figura 7.24, o dispositivo está fisicamente móvel, mas permanece ligado à mesma rede de acesso. O dispositivo também *não* é móvel do ponto de vista da camada de rede. Além disso, se o dispositivo permanece associado à mesma estação-base 802.11 AP ou LTE, ele sequer é móvel da perspectiva da camada de enlace.

De uma perspectiva de rede, nosso interesse por mobilidade de dispositivos começa de fato com o caso *(c)*, no qual um dispositivo muda de rede de acesso (p. ex., WLAN 802.11 ou célula LTE) enquanto continua a enviar e receber datagramas IP e enquanto mantém conexões de mais alto nível (p. ex., TCP). Aqui, a rede precisa realizar uma **transferência** (*handover*), a passagem da responsabilidade pelo repasse de datagramas de/para um AP ou estação-base para o dispositivo móvel, à medida que o dispositivo se move entre WLANs ou entre células LTE. Analisaremos a transferência em detalhes na Seção 7.6. Se essa transferência ocorrer dentro de redes de acesso que pertencem a uma única operadora, esta pode orquestrar a transferência por conta própria. Quando o dispositivo móvel transita entre múltiplas redes de operadoras (*roaming*), como no cenário *(d)*, estas devem orquestrar a transferência juntas, o que aumenta significativamente a complexidade do processo.

7.5.2 Redes nativas e roaming em redes visitadas

Como aprendemos nas nossas discussões sobre redes 4G LTE na Seção 7.4.1, cada assinante possui uma "casa" com uma operadora de celular. Vimos que o serviço de assinante doméstico (HSS) armazena informações sobre cada um dos seus assinantes, incluindo uma identidade de dispositivo globalmente única (integrada ao cartão SIM do assinante), informações sobre os serviços que o assinante poderia acessar, chaves criptográficas para uso na comunicação e informações de cobrança. Quando um dispositivo é conectado a uma rede celular que não é a sua **rede nativa**, diz-se que o dispositivo está em **roaming** (ou transitando) em uma **rede visitada**. Quando um dispositivo móvel se liga a uma rede visitada e transita por ela, é preciso alguma coordenação entre a rede nativa e a rede visitada.

A Internet não possui uma noção igualmente forte de rede nativa ou de rede visitada. Na prática, a rede nativa do aluno pode ser a rede operada pela sua escola; para profissionais, a rede nativa pode ser a rede da empresa. A rede visitada pode ser operada pela rede de uma escola ou empresa que estão visitando. Mas não há uma ideia de rede nativa/visitada profundamente integrada à arquitetura da Internet. O protocolo de IP Móvel (Perkins, 1998; RFC 5944), que analisaremos brevemente na Seção 7.6, foi uma proposta que incorporava fortemente a ideia de redes nativas/visitadas. Mas o IP Móvel foi pouco implantado ou usado na prática. Também estão em andamento atividades construídas sobre a infraestrutura de IP existente, para oferecer acesso autenticado entre redes IP visitadas. A Eduroam é um exemplo dessas atividades (Eduroam, 2020).

A noção de um dispositivo móvel ter uma rede nativa oferece duas vantagens importantes: a rede nativa funciona como local único do qual obter informações sobre o dispositivo e (como veremos) pode servir como ponto de coordenação para comunicação com/de um dispositivo móvel em roaming.

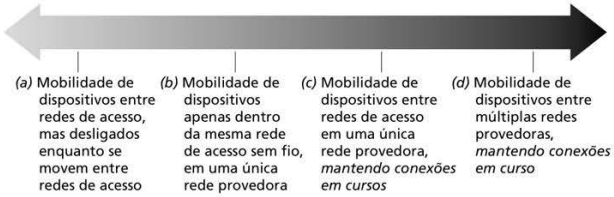


Figura 7.24 Vários graus de mobilidade do ponto de vista da camada de rede.

Para entender o valor em potencial do ponto central para informação e coordenação, considere uma analogia com seres humanos. Bob, um adulto de 20 e poucos anos que sai da casa dos pais torna-se móvel, pois passa a morar em uma série de quartos e/ou apartamentos e está sempre mudando de endereço. Se Alice, uma velha amiga, quiser entrar em contato com ele, como conseguirá o endereço atual de Bob? Uma maneira comum de fazer isso é entrar em contato com a família, já que um jovem móvel costuma informar seus novos endereços (nem que seja só para que os pais possam lhe enviar dinheiro para ajudar a pagar o aluguel!). A residência da família torna-se o único lugar a que outros podem se dirigir como uma primeira etapa para estabelecer comunicação com Bob. Além disso, as comunicações postais posteriores de Alice podem ser *indiretas* (p. ex., pelo envio de uma carta primeiro à casa dos pais, que a encaminharão a Bob) ou *diretas* (p. ex., Alice utilizaria o endereço informado pelos pais e enviaria uma carta diretamente a Bob).

7.5.3 Roteamento direto e indireto de/para um dispositivo móvel

Consideremos agora o dilema enfrentado pelo hospedeiro conectado à Internet (que chamaremos de *correspondente*) na Figura 7.25, que deseja se comunicar com um dispositivo móvel que pode estar localizado na rede celular nativa do dispositivo móvel, ou pode estar em trânsito em uma rede visitada. No nosso cenário abaixo, adotaremos uma perspectiva de rede celular 4G/5G, pois tais redes possuem um longo histórico de suporte à mobilidade de dispositivos. Como veremos, no entanto, os desafios fundamentais e as abordagens de solução básicas para suportar a mobilidade de dispositivos se aplicam igualmente às redes celulares e na Internet.

Como mostrado na Figura 7.25, supomos que o dispositivo móvel possui um identificador globalmente único associado a si. Nas redes celulares 4G LTE (ver Seção 7.4), este seria

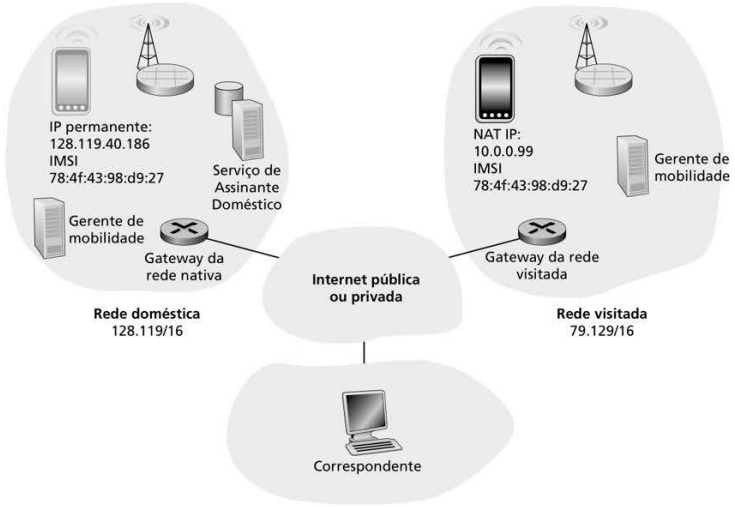


Figura 7.25 Elementos de uma arquitetura de rede móvel.

a identidade internacional de assinante móvel (IMSI) e um número de telefone associado, que fica armazenado no cartão SIM do dispositivo móvel. Para usuários da Internet móvel, este seria o endereço IP permanente na faixa de endereço IP da sua rede nativa, como no caso da arquitetura de IP Móvel.

Qual abordagem seria usada em uma arquitetura de rede móvel que permitiria que um datagrama enviado pelo correspondente alcançasse esse dispositivo móvel? Três abordagens básicas podem ser identificadas e são discutidas a seguir. Como veremos, as duas últimas são adotadas na prática.

Aproveitando a infraestrutura existente de endereço IP

A abordagem mais simples ao roteamento para um dispositivo móvel em uma rede visitada é simplesmente usar a infraestrutura de endereçamento IP existente, sem adicionar nada de novo a ela. O que poderia ser mais fácil?!

Voltando à discussão sobre a Figura 4.21, lembre-se que um Provedor de Serviços de Internet (ISPs, do inglês *Internet Service Provider*) usa Protocolo de Roteador de Borda (BGP, do inglês *Border Gateway Protocol*) para anunciar rotas para redes de destino por meio da enumeração das faixas de endereço "ciderizadas" das redes alcançáveis. Assim, uma rede visitada poderia anunciar para todas as outras redes que um determinado dispositivo móvel está residente na sua rede ao simplesmente anunciar um endereço altamente específico (o endereço IP permanente de 32 bits completo do dispositivo móvel), o que essencialmente informaria as outras redes de que possui o caminho a ser usado para repassar datagramas para tal dispositivo móvel. Essas redes vizinhas então propagariam essas informações de roteamento por toda a rede como parte do procedimento normal de BGP de atualizar as informações de roteamento e tabelas de repasse. Como os datagramas serão sempre repassados ao roteador que anuncia o destino mais específico para o endereço (ver Seção 4.3), todos os datagramas endereçados ao dispositivo móvel serão repassados para a rede visitada. Se o dispositivo móvel deixa uma rede visitada e entra em outra, a nova rede visitada pode anunciar uma nova rota até o dispositivo móvel, altamente específica, e a rede visitada antiga pode retirar suas informações de roteamento relativas ao dispositivo móvel.

Esse procedimento resolve dois problemas de uma vez só e o faz sem promover mudanças na infraestrutura da camada de rede! Outras redes conhecem a localização do dispositivo móvel, e é fácil rotear datagramas para o dispositivo móvel, visto que as tabelas de repasse dirigirão datagramas à rede externa. A grande desvantagem, contudo, é a da capacidade de expansão: os roteadores de rede precisariam manter linhas da tabela de repasse para, possivelmente, bilhões de dispositivos móveis, e atualizar a linha do dispositivo cada vez que este transitasse por uma rede diferente. Claramente, essa abordagem não funcionaria na prática. Algumas desvantagens adicionais serão exploradas nos problemas ao final deste capítulo.

Um método alternativo (que tem sido adotado na prática) é passar a funcionalidade de mobilidade do núcleo da rede para a borda – um tema recorrente em nosso estudo da arquitetura da Internet. Um modo natural de fazer isso é por meio da rede nativa do dispositivo móvel. De maneira muito semelhante ao modo como os pais daquele jovem de 20 e poucos anos monitoram a localização do filho, uma MME na rede nativa do dispositivo móvel pode monitorar a rede visitada na qual o dispositivo móvel reside. Essas informações podem estar em um banco de dados, mostrado como o banco de dados do HSS na Figura 7.25. Um protocolo que opere entre a rede visitada e a rede nativa será necessário para atualizar a rede na qual o dispositivo móvel reside. Lembre-se que encontramos os elementos de MME e HSS em nosso estudo sobre o 4G LTE. Reutilizaremos os nomes desses elementos aqui, pois são tão descritivos e são implantados em larga escala nas redes 4G.

A seguir, vamos considerar em mais detalhes os elementos da rede visitada mostrados na Figura 7.25. O dispositivo móvel claramente precisará de um endereço IP na rede visitada. Aqui, as possibilidades incluem usar um endereço permanente associado à rede nativa do dispositivo móvel, alocar um novo endereço na faixa de endereços da rede visitada e fornecer um endereço IP através da NAT (ver Seção 4.3.4). Nos dois últimos casos, o

dispositivo móvel possui um identificador transiente (um endereço IP recém-alocado) além dos seus identificadores permanentes armazenados no HSS da sua rede nativa. Esses casos são análogos a alguém que escreve uma carta para o endereço da casa na qual o nosso adulto móvel de 20 e poucos anos mora atualmente. No caso de um endereço NAT, os datagramas destinados ao dispositivo móvel chegariam por fim ao roteador do gateway NAT na rede visitada, que então realizaria a tradução do endereço NAT e repassaria o datagrama para o dispositivo móvel.

Agora já vimos diversos elementos de uma solução ao dilema do correspondente na Figura 7.24: redes nativas e visitadas, MME, HSS e endereçamento do dispositivo móvel. Mas como datagramas devem ser endereçados e repassados para o dispositivo móvel? Já que apenas o HSS (e não roteadores no âmbito da rede) conhece a localização do dispositivo móvel, o correspondente não pode simplesmente endereçar um datagrama ao endereço permanente do dispositivo móvel e enviá-lo à rede. Algo mais precisa ser feito. Duas abordagens podem ser identificadas: roteamento indireto e direto.

Roteamento indireto para um dispositivo móvel

Vamos considerar primeiro um correspondente que quer enviar um datagrama a um dispositivo móvel. Na abordagem de **roteamento indireto**, o correspondente apenas endereça o datagrama ao endereço permanente do dispositivo móvel, envia o datagrama para a rede, e nem precisa saber se o dispositivo móvel reside em sua rede nativa ou está em uma rede visitada; assim, a mobilidade é completamente transparente para o correspondente. Esses datagramas são primeiro roteados, como sempre, para a rede local do dispositivo móvel. Isso é ilustrado na etapa 1 da Figura 7.26.

Agora vamos voltar nossa atenção ao HSS, responsável por interagir com as redes visitadas para registrar a localização do dispositivo móvel e o roteador de borda da rede nativa. Uma função desse roteador é estar atento à chegada de um datagrama endereçado a um

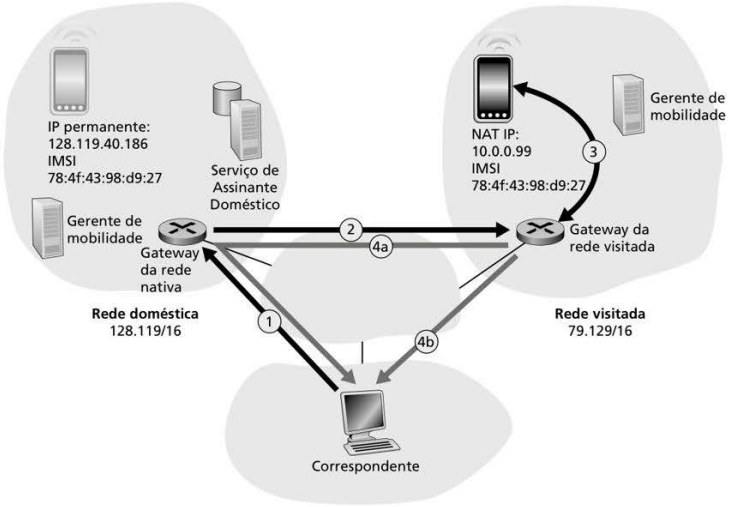


Figura 7.26 Roteamento indireto para um dispositivo móvel.

dispositivo para o qual aquela é a rede nativa, mas que atualmente reside em uma rede visitada. O gateway da rede nativa intercepta o datagrama, consulta o HSS para determinar a rede visitada na qual o dispositivo móvel está residindo e repassa o datagrama para o roteador de borda da rede visitada (a etapa 2 na Figura 7.26). O roteador de borda da rede visitada então repassa o datagrama na direção do dispositivo móvel (a etapa 3 na Figura 7.26). Se a tradução NAT é utilizada, como na Figura 7.26, esta é responsabilidade do roteador de borda da rede visitada.

É instrutivo considerar esse redirecionamento com mais detalhes. Claramente, o gateway da rede nativa precisará repassar o datagrama para o roteador de borda na rede visitada. Por outro lado, é desejável deixar intacto o datagrama do correspondente, pois a aplicação que recebe o datagrama deve desconhecer que este foi repassado por meio da rede nativa. Ambas as metas podem ser cumpridas fazendo o agente nativo encapsular o datagrama original completo do correspondente dentro de um novo (e maior) datagrama. Este é endereçado e entregue ao roteador de borda da rede visitada, que desencapsula o datagrama – isto é, remove o datagrama original do correspondente de dentro daquele datagrama maior de encapsulamento – e repassa o datagrama original para o dispositivo móvel (etapa 3 na Figura 7.26). O leitor atento notará que o encapsulamento/desencapsulamento descrito aqui é idêntico à noção de implementação de túnel discutida na Seção 4.3, no contexto do IPv6; na verdade, também discutimos o uso da implementação de túnel no contexto da Figura 7.18, quando introduzimos o plano de dados 4G LTE.

Por fim, vamos considerar como o dispositivo móvel envia datagramas para o correspondente. No contexto da Figura 7.26, o dispositivo móvel claramente precisará repassar o datagrama por meio do roteador de borda da rede visitada de modo a realizar a NAT. Mas como esse roteador deve repassar o datagrama para o correspondente? Como mostrado na Figura 7.26, temos aqui duas opções: *(4a)* o datagrama poderia ser enviado por túnel de volta ao roteador de borda da rede nativa e enviado de lá para o correspondente, ou *(4b)* o datagrama poderia ser transmitido da rede visitada diretamente para o correspondente, uma abordagem conhecida como **local breakout** no LTE (GSMA, 2019a).

Vamos resumir o que discutimos sobre roteamento indireto revisando as novas funcionalidades da camada de rede exigidas para dar suporte à mobilidade.

- *Um protocolo de associação do dispositivo móvel para a rede visitada.* O dispositivo móvel precisará se associar com a rede visitada e se dissociar quando sair da rede visitada.
- *Um protocolo de registro no HSS da rede visitada para a rede nativa.* A rede visitada precisará registrar o local do dispositivo móvel junto ao HSS na rede nativa e possivelmente usar as informações obtidas do HSS para autenticar o dispositivo.
- *Um protocolo de túnel de datagramas entre o gateway da rede nativa e o roteador de borda da rede visitada.* O lado remetente realiza o encapsulamento e repassa o datagrama original do correspondente; no lado do destinatário, o roteador de borda realiza o desencapsulamento, NAT e repasse do datagrama original para o dispositivo móvel.

A discussão anterior oferece todos os elementos necessários para que um dispositivo móvel mantenha uma conexão contínua com um correspondente enquanto se move entre redes. Quando um dispositivo transita de uma rede visitada para outra, as informações da nova rede visitada precisam ser atualizadas no HSS da rede nativa, e o ponto final do túnel entre os roteadores de borda da rede nativa e da visitada precisa ser movido. Mas o dispositivo móvel verá um fluxo interrompido de datagramas enquanto se move entre as redes. Desde que o tempo entre a desconexão do dispositivo móvel de uma rede visitada e a sua ligação à próxima seja pequeno, poucos datagramas se perderão. No Capítulo 3, vimos que as conexões fim a fim podem sofrer perda de datagramas devido ao congestionamento na rede. Assim, a perda ocasional de datagramas dentro de uma conexão quando um dispositivo se move entre redes não representa um problema catastrófico. Se é preciso ter comunicação sem perdas, mecanismos das camadas superiores

se recuperam da perda de datagramas, seja ela resultante do congestionamento da rede ou da mobilidade do dispositivo.

Nossa discussão foi conscientemente um pouco genérica. Uma abordagem indireta de roteamento é utilizada no padrão IP móvel (RFC 5944), assim como nas redes 4G LTE (Sauter, 2014). Seus detalhes, especialmente os procedimentos de implementação de túnel, diferem pouco da explanação genérica acima.

Roteamento direto para um dispositivo móvel

A abordagem do roteamento indireto ilustrada na Figura 7.26 sofre de uma ineficiência conhecida como **problema do roteamento triangular** – datagramas endereçados ao dispositivo nó móvel devem ser roteados primeiro para a rede nativa e em seguida para a rede visitada, mesmo quando existir uma rota muito mais eficiente entre o correspondente e o dispositivo móvel em roaming. No pior caso, imagine um usuário móvel que está transitando na mesma rede que a rede nativa de um colega estrangeiro que nosso usuário móvel está visitando. Os dois estão sentados lado a lado e trocando dados. Os datagramas entre o usuário móvel e seu colega estrangeiro serão repassados para a rede nativa do primeiro, e então retornarão à rede visitada!

O **roteamento direto** supera a ineficiência do roteamento triangular, mas o faz à custa de complexidade adicional. Na abordagem do roteamento direto, mostrada na Figura 7.27, o correspondente primeiro descobre a rede visitada na qual o dispositivo móvel reside. Para tanto, ele consulta o HSS na rede nativa do dispositivo móvel, pressupondo (assim como no caso do roteamento indireto) que a rede visitada do dispositivo móvel esteja registrada no HSS. É o que mostram as etapas 1 e 2 da Figura 7.27. O correspondente então envia por túnel os datagramas da sua rede *diretamente* para o roteador de borda na rede visitada do dispositivo móvel.

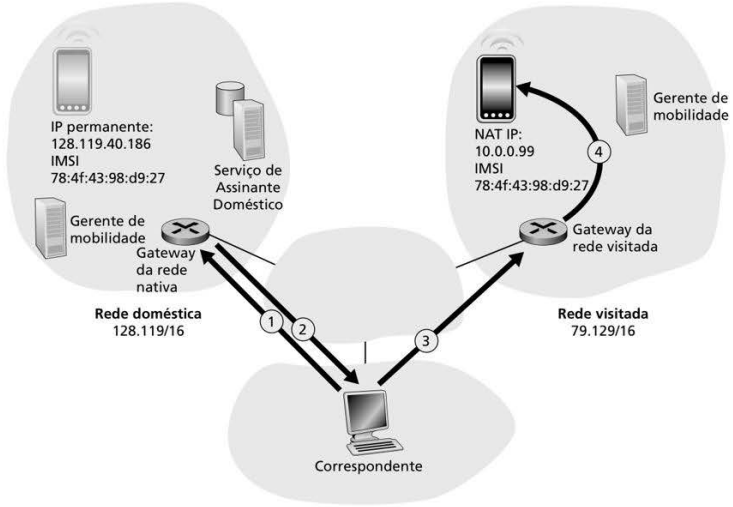


Figura 7.27 Roteamento direto para um dispositivo móvel.

Embora supere o problema do roteamento triangular, o roteamento direto introduz dois importantes desafios adicionais:

- É necessário um protocolo de localização do usuário móvel para que o correspondente consulte o HSS para obter a rede visitada do dispositivo móvel (etapas 1 e 2 na Figura 7.27), além do protocolo necessário para que o dispositivo móvel registre seu local junto ao seu HSS.
- Quando o dispositivo móvel passa de uma rede visitada para outra, como o correspondente sabe que deve repassar os datagramas para a nova rede visitada? No caso do roteamento indireto, o problema é resolvido facilmente pela atualização do HSS na rede nativa e pela alteração da extremidade do túnel para que termine no roteador de borda da nova rede visitada. Com o roteamento direto, contudo, essa mudança nas redes visitadas não é tão fácil, pois o HSS é consultado pelo correspondente apenas no início da sessão. Assim, seriam necessários mecanismos de protocolo adicionais para atualizar proativamente o correspondente todas as vezes que o dispositivo móvel se mover. Dois problemas no final deste capítulo exploram soluções para esse problema.

7.6 GERENCIAMENTO DA MOBILIDADE NA PRÁTICA

Na seção anterior, identificamos os desafios fundamentais e soluções em potencial para o desenvolvimento de uma arquitetura de rede que suporte a mobilidade de dispositivos: as ideias de redes nativa e visitada; o papel da rede nativa como ponto central de informações e controle para dispositivos móveis da qual são assinantes; as funções do plano de controle necessárias para que a entidade de gerenciamento móvel da rede nativa controle o roaming de um dispositivo móvel entre redes visitadas; e abordagens do plano de dados para roteamento direto e indireto de modo a capacitar o intercâmbio de datagramas entre um correspondente e um dispositivo móvel. Agora, vamos ver como esses princípios são colocados em prática. Na Seção 7.6.1, estudaremos o gerenciamento da mobilidade nas redes 4G/5G; na Seção 7.6.2, estudaremos o IP Móvel, padrão proposto para a Internet.

7.6.1 Gerenciamento da mobilidade em redes 4G/5G

Nosso estudo anterior sobre a arquitetura 4G e a 5G emergente, na Seção 7.4, nos apresentou todos os elementos de rede que têm um papel crítico no gerenciamento da mobilidade 4G/5G. Agora, vamos ilustrar como esses elementos operam uns com os outros para prestar serviços de mobilidade nas redes 4G/5G atuais (Sauter, 2014; GSMA 2019b), que têm suas origens nas redes celulares de voz e dados 3G anteriores (Sauter, 2014) e nas redes 2G ainda mais anteriores, que ofereciam apenas serviços de voz (Mouly, 1992). Isso nos ajudará a sintetizar aquilo que aprendemos até aqui e nos permitirá introduzir mais alguns tópicos avançados, além de oferecer uma lente sobre o que o gerenciamento da mobilidade que o 5G pode ter a oferecer.

Consideremos um cenário simples, no qual um usuário móvel (p. ex., o passageiro em um carro) liga seu smartphone a uma rede 4G/5G visitada, começa a assistir streaming de vídeo em HD de um servidor remoto e então passa da cobertura celular de uma estação-base 4G/5G para a de outra. A Figura 7.28 apresenta as quatro etapas principais nesse cenário.

1. *Associação entre dispositivo móvel e estação-base.* O dispositivo móvel se associa com uma estação-base na rede visitada.
2. *Configuração do plano de controle de elementos de rede para o dispositivo móvel.* As redes visitada e nativa estabelecem um estado do plano de controle indicando que o dispositivo móvel reside na rede visitada.

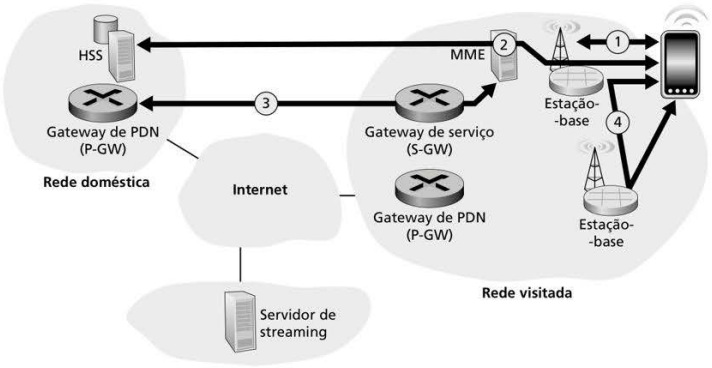


Figura 7.28 Exemplo de cenário de mobilidade 4G/5G.

3. *Configuração do plano de dados de túneis de repasse para o dispositivo móvel.* A rede visitada e a rede nativa estabelecem túneis através dos quais o dispositivo móvel e o servidor de streaming podem enviar/receber datagramas IP, usando roteamento indireto através do gateway da rede de pacote de dados (P-GW) da rede nativa.
4. *Transferência do dispositivo móvel de uma estação-base para outra.* O dispositivo móvel muda o seu ponto de ligação à rede visitada por meio da transferência de uma estação-base para outra.

Consideremos agora cada um desses quatro passos em mais detalhes.

1. Associação à estação-base. Lembre-se que, na Seção 7.4.2, estudamos os procedimentos pelos quais um dispositivo móvel se associa a uma estação-base. Aprendemos que o dispositivo móvel busca em todas as frequências os sinais primários transmitidos pelas estações-base na sua área. O dispositivo móvel adquire progressivamente mais informações sobre essas estações-base e seleciona aquela à qual se associará, então inicia um canal de sinalização de controle com a estação-base relevante. Durante essa associação, o dispositivo móvel fornece à estação-base a sua identidade internacional de assinante móvel (IMSI), um identificador único do dispositivo móvel e da sua rede nativa, além de informações do assinante adicionais.

2. Configuração do plano de controle de elementos da rede LTE para o dispositivo móvel. Após o estabelecimento do canal de sinalização entre o dispositivo móvel e a estação-base, esta pode contatar a MME da rede visitada. A MME consulta e configura diversos elementos 4G/5G nas redes nativa e visitada para estabelecer o estado em nome do nó móvel:

- A MME usará a IMSI e outras informações fornecidas pelo dispositivo móvel para recuperar informações de autenticação, criptografia e serviços de rede disponíveis para o assinante. Essas informações podem estar no cache local da MME, ser recuperadas de outra MME que o dispositivo móvel contatou recentemente ou recuperadas do HSS na rede nativa do dispositivo móvel. O processo de autenticação mútua (que analisaremos em mais detalhes na Seção 8.8) garante que a rede visitada tem certeza sobre a identidade do dispositivo móvel e que o dispositivo pode autenticar a rede à qual está se ligando.
- A MME informa o HSS na rede nativa do dispositivo móvel que este agora reside na rede visitada; o HSS atualiza o seu banco de dados.

- A estação-base e o dispositivo móvel selecionam parâmetros para o canal do plano de dados a ser estabelecido entre o dispositivo móvel e a estação-base (lembre-se que um canal de sinalização do plano de controle já está em operação).

3. Configuração do plano de dados de túneis de repasse para o dispositivo móvel. A seguir, a MME configura o plano de dados para o dispositivo móvel, como mostra a Figura 7.29. Dois túneis são estabelecidos. Um túnel fica entre a estação-base e um gateway de serviço na rede visitada. O segundo túnel fica entre esse gateway de serviço e o roteador do gateway de PDN *na rede nativa do dispositivo móvel*. O 4G LTE implementa essa forma de roteamento indireto simétrico – todo o tráfego de e para o dispositivo é enviado por túnel através da rede nativa do dispositivo. Os túneis 4G/5G usam o Protocolo de Implantação de Túnel GPRS (GTP, do inglês *GPRS Tunneling Protocol*), especificado em (3GPP GTPv1-U, 2019). O TEID no cabeçalho GTP indica a qual túnel um datagrama pertence, o que permite que múltiplos fluxos sejam multiplexados e demultiplexados pelo GTP entre os pontos finais dos túneis.

Seria instrutivo comparar a configuração de túneis na Figura 7.29 (o caso do roaming de um celular em uma rede visitada) com a da Figura 7.18 (o caso da mobilidade apenas dentro da rede nativa do dispositivo móvel). Vemos que, em ambos os casos, o gateway de serviço reside na mesma rede que o dispositivo móvel, mas o gateway de PDN (que é sempre o gateway de PDN na rede nativa do dispositivo móvel) pode estar em uma rede diferente daquela na qual o dispositivo móvel se encontra. Este é exatamente o roteamento indireto. Foi especificada uma alternativa ao roteamento indireto, conhecida como **local breakout** (GSMA, 2019a), na qual o gateway de serviço estabelece um túnel para o gateway de PDN na rede visitada local. Na prática, entretanto, o local breakout não é uma prática muito difundida (Sauter, 2014).

Após os túneis terem sido configurados e ativados, o dispositivo móvel pode então repassar pacotes de/para a Internet através do gateway de PDN na sua rede nativa!

4. Gerenciamento de transferência. Uma **transferência** (***handover***) ocorre quando um dispositivo móvel muda a sua associação de uma estação-base para outra. O processo de transferência descrito abaixo continua o mesmo independentemente do dispositivo móvel residir na sua rede nativa ou estiver em trânsito em uma rede visitada.

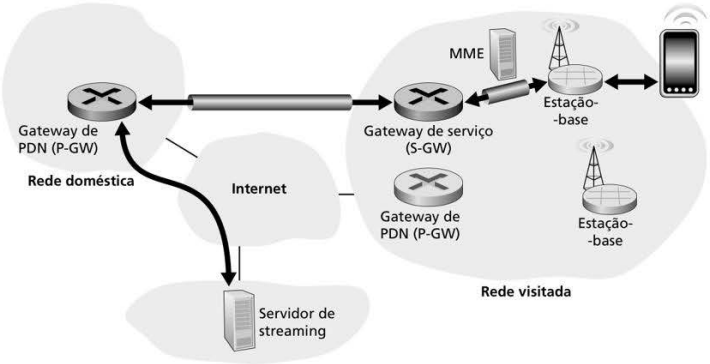


Figura 7.29 Implantação de túnel nas redes 4G/5G entre o gateway de serviço na rede visitada e o gateway de PDN na rede nativa.

Como mostra a Figura 7.30, datagramas de/para o dispositivo inicialmente (antes da transferência) são repassados a ele através da estação-base (que chamaremos de estação-base *de origem*), e, após a transferência, são roteados para o dispositivo móvel através de outra estação-base (que chamaremos de estação-base *de destino*). Como veremos, uma transferência entre estações-base resulta, além da transmissão/recebimento de/para uma nova estação-base por parte do dispositivo móvel, em uma mudança no lado da estação-base em relação ao túnel entre o gateway de serviço e a estação-base na Figura 7.29. No caso mais simples de transferência, quando as duas estações-base estão próximas uma à outra e na mesma rede, todas as mudanças que ocorrem devido à transferência são, assim, relativamente locais. Em especial, o gateway de PDN usado pelo gateway de serviço continua sem saber nada sobre a mobilidade do dispositivo. Obviamente, cenários de transferência mais complicados exigem o uso de mecanismos mais complexos (Sauter, 2014; GSMA, 2019a).

Pode haver diversos motivos para que ocorra a transferência. Por exemplo, o sinal entre a estação-base corrente e o dispositivo móvel pode se deteriorar tanto que a comunicação fica gravemente prejudicada. Ou a célula pode estar sobrecarregada, tamanha a quantidade de tráfego; transferir dispositivos móveis para células vizinhas menos congestionadas pode aliviar o problema. O dispositivo móvel mede periodicamente as características de um sinal de sinalização emitido por sua estação-base corrente, bem como de sinais de sinalização emitidos por estações-base próximas que ele pode "ouvir". Essas medições são passadas uma ou duas vezes por segundo para a estação-base corrente do dispositivo móvel. Com base nessas medições, nas cargas correntes de usuários móveis em células próximas e outros fatores, a estação-base de origem pode optar por iniciar a transferência. Os padrões 4G/5G não especificam o algoritmo específico a ser utilizado por uma estação-base para decidir se realiza ou não uma transferência ou qual estação-base de destino escolher; esta é uma área de pesquisa em atividade (Zheng, 2008; Alexandris, 2016).

A Figura 7.30 ilustra as etapas envolvidas quando uma estação-base de origem decide transferir um dispositivo móvel para a estação-base de destino.

1. A estação-base corrente (origem) seleciona a estação-base de destino e envia uma mensagem de Solicitação de Transferência para a estação-base de destino.
2. A estação-base de destino verifica se possui os recursos para suportar o dispositivo móvel e seus requisitos de qualidade de serviço. Em caso positivo, pré-aloca para o dispositivo recursos de canal (p. ex., intervalos de tempo) na sua rede de acesso por rádio e outros recursos. Essa pré-alocação de recursos libera o dispositivo móvel de precisar realizar o demorado protocolo de associação à estação-base discutido anteriormente e

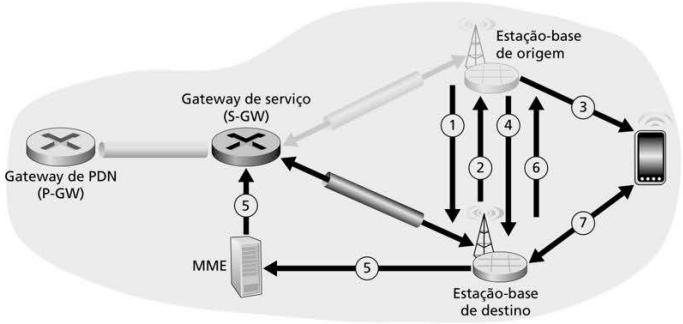


Figura 7.30 Etapas na transferência de um dispositivo móvel da estação-base de origem para a estação-base de destino.

acelera ao máximo a execução da transferência. A estação-base de destino responde à estação-base de origem com uma mensagem de Reconhecimento da Solicitação de Transferência contendo toda as informações na estação-base de destino das quais o dispositivo móvel precisará para se associar à nova estação-base.
3. A estação-base de origem recebe a mensagem de Reconhecimento da Solicitação de Transferência e informa o dispositivo móvel sobre a identidade da estação-base de destino e fornece informações de acesso ao canal. Neste ponto, o dispositivo móvel pode começar a enviar/receber datagramas de/para a nova estação-base de destino. Do ponto de vista do dispositivo móvel, a transferência está completa! Contudo, ainda há bastante trabalho pela frente dentro da rede.
4. A estação-base de origem também parará de repassar datagramas para o dispositivo móvel e passará então a repassar quaisquer datagramas enviados por túnel que receber para a estação-base de destino, que posteriormente repassará esses datagramas para o dispositivo móvel.
5. A estação-base de destino informa a MME que ela será a nova estação-base atendendo o dispositivo móvel. A MME, por sua vez, sinaliza para o gateway de serviço e a estação-base de destino para reconfigurar o túnel do gateway de serviço para a estação-base, de modo que este termine na segunda, não na estação-base de origem.
6. A estação-base de destino confirma de volta para a estação-base de origem que o túnel foi reconfigurado, permitindo que a estação-base de origem libere recursos associados com o dispositivo móvel.
7. Nesse ponto, a estação-base de destino também pode começar a entregar datagramas para o dispositivo móvel, incluindo datagramas repassados para a estação-base de destino pela estação-base de origem durante a transferência, além de datagramas recém-chegados ao túnel reconfigurado a partir do gateway de serviço. Ela também pode repassar datagramas de saída recebidos do dispositivo móvel para o túnel até o gateway de serviço.

As configurações de roaming nas redes 4G LTE da atualidade, como analisado acima, também serão utilizadas nas redes 5G emergentes do futuro (GSMA, 2019c). Lembre-se, no entanto, pela nossa discussão na Seção 7.4.6, que as redes 5G serão mais densas, com células de tamanho significativamente menor. Isso significa que a transferência será uma função de rede ainda mais crítica. Além disso, a baixa latência nas transferências será fundamental para muitas aplicações do 5G em tempo real. A migração do plano de controle da rede celular para a estrutura de rede definida por software (SDN, do inglês *software-defined networking*), estudada anteriormente no Capítulo 5 (GSMA, 2018b; Condoluci, 2018), promete capacitar implementações de um plano de controle da rede celular 5G de maior capacidade e menor latência. A aplicação da SDN no contexto 5G é o tópico de diversos esforços de pesquisa (Giust, 2015; Ordonez-Lucena, 2017; Nguyen, 2016).

7.6.2 IP móvel

A Internet da atualidade não possui uma infraestrutura ampla e disseminada que oferece o tipo de serviço para usuários móveis que encontramos nas redes celulares 4G/5G. Mas certamente não é por falta de soluções técnicas para oferecer esses serviços no contexto da Internet! Na verdade, a arquitetura e os protocolos do IP Móvel (RFC 5944), que discutimos brevemente a seguir, foram padronizados por RFCs da Internet há mais de 20 anos, e os pesquisadores continuam a produzir novas soluções de mobilidade mais seguras e generalizadas (Venkataramani, 2014).

Em vez disso, talvez tenha sido a falta de casos de negócios e de uso relevantes (Arkko, 2012) e o desenvolvimento e implantação tempestivos de soluções alternativas de mobilidade nas redes celulares que impediram a implementação do IP Móvel. Lembre-se que, 20 anos atrás, as redes celulares 2G já ofereciam uma solução para serviços móveis de voz (o "aplicativo obrigatório" para usuários móveis); além disso, as redes 3G da próxima geração, que apoiavam voz *e* dados, estavam no horizonte. Talvez a solução tecnológica dupla

– serviços móveis por redes celulares quando estamos realmente móveis e em movimento (i.e., na extremidade direita do espectro de mobilidade na Figura 7.24) e serviços de Internet através de redes 802.11 ou cabeadas quando estamos estacionários ou nos movemos apenas localmente (i.e., na extremidade esquerda do espectro de mobilidade na Figura 7.24) – que tínhamos 20 anos atrás e ainda temos hoje continuará a existir no futuro.

Ainda assim, pode ser instrutivo considerar brevemente o padrão IP Móvel, pois oferece muitos dos mesmos serviços que as redes celulares e implementa muitos dos mesmos princípios básicos de mobilidade. As edições anteriores deste livro ofereceram um estudo mais aprofundado sobre o IP Móvel do que temos aqui (o leitor interessado encontra o material no *site* do livro). Os protocolos e a arquitetura da Internet para suportar mobilidade, conhecidos coletivamente pelo nome IP Móvel, foram definidos principalmente no RFC 5944 para o IPv4. O IP móvel, assim como o 4G/5G, é um padrão complexo, e seria preciso todo um livro para detalhá-lo; Perkins (1998b), aliás, foi um que escreveu um livro com esses detalhes. Nosso objetivo aqui é mais modesto: oferecer uma visão geral sobre os aspectos mais importantes do IP Móvel.

Os elementos e a arquitetura geral do IP Móvel são incrivelmente semelhantes aos das redes de operadoras de celular. Há uma noção forte de rede nativa, na qual um dispositivo móvel possui um endereço IP permanente, e de redes visitadas (chamadas de redes "externas" no IP Móvel), nas quais é alocado ao dispositivo móvel um endereço administrado (COA, do inglês *care-of-address*). O agente nativo no IP Móvel possui uma função semelhante ao HSS no LTE: rastrear o local de um dispositivo móvel, assim como o HSS recebe atualizações das MMEs nas redes visitadas nas quais reside um dispositivo móvel 4G. Ambos os padrões, 4G/5G e IP Móvel, utilizam roteamento indireto para um nó móvel, usando túneis para conectar os roteadores de borda nas redes nativas e visitadas/externas. A Tabela 7.3 resume os elementos da arquitetura do IP Móvel e inclui uma comparação com elementos semelhantes nas redes 4G/5G.

O padrão IP móvel consiste em três partes principais:

- *Descoberta de agente.* O IP móvel define os protocolos utilizados por um agente externo para anunciar seus serviços a dispositivos móveis que desejam se ligar à sua rede. Esses serviços incluem o fornecimento de um endereço administrado ao dispositivo móvel para uso na rede externa, registro do dispositivo móvel junto ao agente nativo na rede nativa do dispositivo móvel e repasse de datagramas de e para o dispositivo móvel, entre outros.
- *Registro no agente nativo.* O IP móvel define os protocolos usados pelo dispositivo móvel e/ou agente externo para registrar e anular os registros de COAs no agente local de um dispositivo móvel.
- *Roteamento indireto de datagramas.* O IP móvel também define a maneira pela qual datagramas são repassados para dispositivos móveis por um agente nativo, incluindo regras para repassar datagramas, regras para manipular condições de erro e diversas formas de implementação de túnel (RFC 2003; RFC 2004).

TABELA 7.3 Semelhanças entre arquiteturas 4G/5G e IP móvel

Elemento 4G/5G	Elemento do IP móvel	Discussão
Rede doméstica	Rede doméstica	
Rede visitada	Rede externa	
Identificador IMSI	Endereço IP permanente	Informação de endereço roteável globalmente único
Serviço de Assinante Doméstico (HSS)	Agente nativo	
Entidade de Gerenciamento Móvel (MME)	Agente externo	
Plano de dados: repasse indireto através da rede nativa, com túnel entre a rede nativa e a visitada e túnel dentro da rede na qual o dispositivo móvel reside	Plano de dados: repasse indireto através da rede nativa, com túnel entre a rede nativa e a visitada	
Estação-base (eNode-B)	Ponto de acesso (AP)	Nenhuma tecnologia de AP específica é definida no IP móvel
Rede de acesso por rádio	WLAN	Nenhuma tecnologia de WLAN específica é definida no IP móvel

Mais uma vez, nossa cobertura sobre o IP móvel foi propositalmente breve. O leitor interessado deve consultar as referências nesta seção ou as discussões mais detalhadas sobre o IP móvel em edições anteriores deste livro.

7.7 SEM FIO E MOBILIDADE: IMPACTO SOBRE PROTOCOLOS DE CAMADAS SUPERIORES

Neste capítulo, vimos que redes sem fio são significativamente diferentes de suas contrapartes cabeadas tanto na camada de enlace (como resultado de características de canais sem fio como desvanecimento, propagação multivias e terminais ocultos) quanto na camada de rede (como resultado de usuários móveis que mudam seus pontos de conexão com a rede). Mas há diferenças importantes nas camadas de transporte e de aplicação? É tentador pensar que essas diferenças seriam pequenas, visto que a camada de rede provê o mesmo modelo de serviço de entrega de melhor esforço às camadas superiores tanto em redes cabeadas quanto em redes sem fio. De modo semelhante, se protocolos como TCP ou UDP são usados para oferecer serviços da camada de transporte a aplicações tanto em redes cabeadas como em redes sem fio, então a camada de aplicação também deve permanecer inalterada. Nossa intuição está certa em um sentido – TCP e UDP podem operar, e de fato operam, em redes com enlaces sem fio. Por outro lado, protocolos de transporte em geral e o TCP em particular às vezes podem ter desempenhos muito diferentes em redes cabeadas e em redes sem fio, e é neste particular, em termos de desempenho, que as diferenças se manifestam. Vejamos por quê.

Lembre-se de que o TCP retransmite um segmento que é perdido ou corrompido no caminho entre remetente e destinatário. No caso de usuários móveis, a perda pode resultar de congestionamento de rede (esgotamento de buffer de roteador) ou de transferência (p. ex., de atrasos no redirecionamento de segmentos para um novo ponto de conexão do usuário à rede). Em todos os casos, o ACK do destinatário ao remetente do TCP indica apenas que um segmento não foi recebido intacto; o remetente não sabe se o segmento foi perdido por congestionamento, durante a transferência, ou por erros de bits detectados. Em todos os casos, a resposta do remetente é a mesma – retransmitir o segmento. A resposta do controle de congestionamento do TCP *também* é a mesma em todos os casos – o TCP reduz sua janela de congestionamento, como discutimos na Seção 3.7. Reduzindo de modo incondicional sua janela de congestionamento, o TCP admite implicitamente que a perda de segmento resulta de congestionamento e não de corrupção ou transferência. Vimos na Seção 7.2 que erros de bits são muito mais comuns em redes sem fio do que nas cabeadas. Quando ocorrem esses erros de bits ou quando há perda na transferência, na realidade não há razão alguma para que o remetente TCP reduza sua janela de congestionamento (reduzindo, assim, sua taxa de envio). Na verdade, é bem possível que os buffers de roteador estejam vazios e que pacotes estejam fluindo ao longo de caminhos fim a fim desimpedidos, sem congestionamento.

Entre o início e meados da década de 1990, pesquisadores perceberam que, dadas as altas taxas de erros de bits em enlaces sem fio e a possibilidade de perdas pela transferência de usuários, a resposta do controle de congestionamento do TCP poderia ser problemática

em um ambiente sem fio. Há três classes gerais de abordagens possíveis para tratar esse problema:

- *Recuperação local.* Os protocolos de recuperação local recuperam erros de bits quando e onde (p. ex., no enlace sem fio) eles ocorrem (p. ex., o protocolo ARQ 802.11, que estudamos na Seção 7.3, ou técnicas mais sofisticadas que utilizam ARQ e FEC (Ayanoglu, 1995) que vimos em uso nas redes 4G/5G na Seção 7.4.2).
- *Remetente TCP ciente de enlaces sem fio.* Em técnicas de recuperação locais, o remetente TCP fica completamente desavisado de que seus segmentos estão atravessando um enlace sem fio. Uma técnica alternativa é o remetente e o destinatário ficarem cientes da existência de um enlace sem fio, para distinguir entre perdas por congestionamento na rede cabeada e corrupção/perdas no enlace sem fio, e invocar o controle de congestionamento somente em resposta a perdas por congestionamento na rede cabeada. Liu (2003) investiga técnicas para diferenciar entre perdas nos segmentos cabeados e sem fio de um caminho fim a fim. Huang (2013) apresenta ideias sobre o desenvolvimento de aplicações e mecanismos de protocolos de transporte que se adaptem melhor a LTE.
- *Técnicas de conexão dividida.* Nesta técnica de conexão dividida (Bakre, 1995), a conexão fim a fim entre o usuário móvel e o outro ponto terminal é dividida em duas conexões da camada de transporte: uma do hospedeiro móvel ao ponto de acesso sem fio, e uma do ponto de acesso sem fio ao outro ponto terminal de comunicação (admitiremos, aqui, um usuário cabeado). A conexão fim a fim é, então, formada por uma concatenação de uma parte sem fio e uma parte cabeada. A camada de transporte sobre um segmento sem fio pode ser uma conexão-padrão TCP (Bakre, 1995), ou principalmente um protocolo de recuperação de erro personalizado em cima do UDP. Yavatkar (1994) analisa o uso de um protocolo de repetição seletiva da camada de transporte por uma conexão sem fio. As medidas relatadas em Wei (2006) indicam que conexões TCP divididas são bastante usadas em redes de dados celulares, e que aperfeiçoamentos significativos podem ser feitos com o uso dessas conexões.

Aqui, nosso tratamento do TCP em enlaces sem fio foi necessariamente breve. Estudos mais aprofundados dos desafios e soluções do TCP nessas redes podem ser encontrados em Hanabali (2005) e Leung (2006). Aconselhamos o leitor a consultar as referências se quiser mais detalhes sobre essa área de pesquisa em curso.

Agora que já consideramos protocolos de camada de transporte, vamos analisar em seguida o efeito do sem fio e da mobilidade sobre protocolos da camada de aplicação. Devido à natureza compartilhada do espectro sem fio, aplicações que operam por enlaces sem fio, em particular por enlaces celulares sem fio, devem tratar a largura de banda como uma mercadoria escassa. Por exemplo, um servidor Web que serve conteúdo a um navegador Web que está rodando em um telefone 4G talvez não consiga prover o mesmo conteúdo rico em imagens que oferece a um navegador que está rodando sobre uma conexão cabeada. Embora enlaces sem fio proponham desafios na camada de aplicação, a mobilidade que eles criam também torna possível um rico conjunto de aplicações dependentes de localização e de contexto (Baldauf, 2007). Em termos mais gerais, redes sem fio e redes móveis desempenharão um papel fundamental na concretização dos ambientes de computação onipresentes do futuro (Weiser, 1991). É justo dizer que vimos somente a ponta do iceberg quando se trata do impacto de redes sem fio e móveis sobre aplicações em rede e seus protocolos!

7.8 RESUMO

As redes sem fio e móveis revolucionaram a telefonia e agora estão causando um impacto cada vez mais profundo no mundo das redes de computadores. Com o acesso à infraestrutura da rede global que oferecem – desimpedido, a qualquer hora, em qualquer lugar –, estão

não só aumentando a onipresença do acesso a redes, mas também habilitando um novo conjunto muito interessante de serviços dependentes de localização. Dada a crescente importância das redes sem fio e móveis, este capítulo focalizou os princípios, as tecnologias de enlace e as arquiteturas de rede para suportar comunicações sem fio e móveis.

Iniciamos o capítulo com uma introdução às redes sem fio e móveis, traçando uma importante distinção entre os desafios propostos pela natureza *sem fio* dos enlaces de comunicação desse tipo de rede e pela *mobilidade* que permitem. Isso nos possibilitou isolar, identificar e dominar melhor os conceitos fundamentais em cada área. Focalizamos primeiro a comunicação sem fio, considerando as características de um enlace sem fio na Seção 7.2. Nas Seções 7.3 e 7.4, examinamos os aspectos de camada de enlace do padrão IEEE 802.11 (WiFi) para LANs sem fio, redes Bluetooth e redes celulares 4G/5G. Depois, voltamos nossa atenção para a questão da mobilidade. Na Seção 7.5, identificamos diversas formas de mobilidade, e verificamos que há pontos nesse espectro que propõem desafios diferentes e admitem soluções diferentes. Consideramos os problemas de localização e roteamento para um usuário móvel, bem como técnicas para transferir o usuário móvel que passa dinamicamente de um ponto de conexão com a rede para outro. Examinamos como essas questões foram abordadas nas redes 4G/5G e no padrão IP móvel. Por fim, na Seção 7.7, consideramos o impacto causado por enlaces sem fio e pela mobilidade sobre protocolos de camada de transporte e aplicações em rede.

Embora tenhamos dedicado um capítulo inteiro ao estudo de redes sem fio e redes móveis, seria preciso todo um livro (ou mais) para explorar completamente esse campo tão animador e que está se expandindo tão depressa. Aconselhamos o leitor a se aprofundar mais nesse campo consultando as muitas referências fornecidas neste capítulo.

Exercícios de fixação e perguntas

Questões de revisão do Capítulo 7

SEÇÃO 7.1

R1. O que significa para uma rede sem fio estar operando no "modo de infraestrutura"? Se a rede não estiver nesse modo, em qual modo ela está e qual é a diferença entre esse modo de operação e o de infraestrutura?

R2. Quais são os quatro tipos de redes sem fio identificadas em nossa taxonomia na Seção 7.1? Quais desses tipos de rede sem fio você usou?

SEÇÃO 7.2

R3. Quais são as diferenças entre os seguintes tipos de falhas no canal sem fio: atenuação de percurso, propagação multivias, interferência de outras fontes?

R4. Um nó móvel se distancia cada vez mais de uma estação-base. Quais são as duas atitudes que uma estação-base poderia tomar para garantir que a probabilidade de perda de um quadro transmitido não aumente?

SEÇÃO 7.3

R5. Descreva o papel dos quadros de sinalização em 802.11.

R6. Verdadeiro ou falso: antes de uma estação 802.11 transmitir um quadro de dados, ela deve primeiro enviar um quadro RTS e receber um quadro CTS correspondente.

R7. Por que são usados reconhecimentos em 802.11, mas não em Ethernet cabeada?

R8. Verdadeiro ou falso: Ethernet e 802.11 usam a mesma estrutura de quadro.

R9. Descreva como funciona o patamar RTS.

R10. Suponha que os quadros RTS e CTS IEEE 802.11 fossem tão longos quanto os padronizados DATA e ACK. Haveria alguma vantagem em usar os quadros CTS e RTS? Por quê?

R11. A Seção 7.3.4 discute mobilidade 802.11, na qual uma estação sem fio passa de um BSS para outro dentro da mesma sub-rede. Quando os APs estão interconectados com um switch, um AP pode precisar enviar um quadro com um endereço MAC fingido para fazer o switch transmitir quadros adequadamente. Por quê?

R12. Quais são as diferenças entre o dispositivo mestre em uma rede Bluetooth e uma estação-base em uma rede 802.11?

R13. Qual é o papel da estação-base na arquitetura celular 4G/5G? Com quais outros elementos de rede 4G/5G (dispositivo móvel, MME, HSS, roteador do Gateway de Serviço, roteador do Gateway de PDN) ela se comunica *diretamente* no plano de controle? E no plano de dados?

R14. O que é uma identidade internacional de assinante móvel (IMSI)?

R15. Qual é o papel do serviço de assinante doméstico na arquitetura celular 4G/5G? Com quais outros elementos de rede 4G/5G (dispositivo móvel, estação-base, MME, roteador do Gateway de Serviço, roteador do Gateway de PDN) ele se comunica *diretamente* no plano de controle? E no plano de dados?

R16. Qual é o papel da entidade de gerenciamento móvel (MME) na arquitetura celular 4G/5G? Com quais outros elementos de rede 4G/5G (dispositivo móvel, estação-base, HSS, roteador do Gateway de Serviço, roteador do Gateway de PDN) ela se comunica *diretamente* no plano de controle? E no plano de dados?

R17. Descreva o propósito de dois túneis no plano de dados da arquitetura celular 4G/5G. Quando um dispositivo móvel está ligado à sua própria rede nativa, em qual elemento de rede 4G/5G (dispositivo móvel, estação-base, HSS, MME, roteador do Gateway de Serviço, roteador do Gateway de PDN) cada extremidade de cada um dos dois túneis termina?

R18. Quais são as três subcamadas na camada de enlace da pilha de protocolos LTE? Descreva brevemente as suas funções.

R19. A rede de acesso sem fio LTE usa FDMA, TDMA ou ambos? Explique sua resposta.

R20. Descreva os dois modos de suspensão possíveis de um dispositivo móvel 4G/5G. Em cada um deles, o dispositivo móvel permanecerá associado à mesma estação-base entre o momento em que é suspenso e aquele no qual acorda e envia/recebe um novo datagrama pela primeira vez?

R21. O que são a "rede visitada" e a "rede nativa" na arquitetura celular 4G/5G?

R22. Liste três diferenças importantes entre redes celulares 4G e 5G.

SEÇÃO 7.5

R23. O que significa dizer que um dispositivo móvel está em roaming?

R24. O que significa falar da "transferência" (*handover*) de um dispositivo de rede?

R25. Qual é a diferença entre roteamento direto e indireto de datagramas de/para um hospedeiro móvel em modo de roaming?

R26. O que significa "roteamento triangular"?

SEÇÃO 7.6

R27. Descreva a semelhança e as diferenças na configuração do túnel quando um dispositivo móvel reside na sua rede nativa em comparação quando está transitando (em roaming) em uma rede visitada.

R28. Quando um dispositivo móvel é transferido de uma estação-base para outra em uma rede 4G/5G, qual elemento da rede toma a decisão de iniciar a transferência? Qual elemento de rede escolhe a estação-base de destino para a qual o dispositivo móvel será transferido?

R29. Descreva como e quando o caminho de repasse dos datagramas que entram na rede visitada e destinados ao dispositivo móvel muda antes, durante e após a transferência.

R30. Considere os seguintes elementos da arquitetura do IP Móvel: rede nativa, endereço permanente na rede externa, agente nativo, agente externo, repasse do plano de dados, ponto de acesso (AP) e WLANs na borda da rede. Quais os elementos equivalentes mais próximos na arquitetura das redes celulares 4G/5G?

SEÇÃO 7.7

R31. Quais são os três métodos que podem ser realizados para evitar que um único enlace sem fio reduza o desempenho de uma conexão TCP fim a fim da camada de transporte?

Problemas

P1. Considere o exemplo do remetente CDMA único na Figura 7.5. Qual seria a saída do remetente (para os 2 bits de dados mostrados) se o código do remetente CDMA fosse (1, –1, 1, –1, 1, 1, 1, –1)?

P2. Considere o remetente 2 na Figura 7.6. Qual é a saída do remetente para o canal (antes de ser adicionada ao sinal vindo do remetente 1), $Z^2_{i,m}$?

P3. Suponha que o receptor na Figura 7.6 queira receber os dados que estão sendo enviados pelo remetente 2. Mostre (por cálculo) que o receptor pode, na verdade, recuperar dados do remetente 2 do sinal agregado do canal usando o código do remetente 2.

P4. Para o exemplo sobre dois remetentes, dois destinatários, dê um exemplo de dois códigos CDMA contendo 1 e 21 valores, que não permitem que dois destinatários extraiam os bits originais transmitidos por dois remetentes CDMA.

P5. Suponha que dois ISPs fornecem acesso WiFi em um determinado local, e que cada um deles opera seu próprio AP e tem seu próprio bloco de endereços IP.

   a. Suponha ainda, que, por acidente, cada ISP configurou seu AP para operar no canal 11. O protocolo 802.11 falhará totalmente nessa situação? Discuta o que acontece quando duas estações, cada uma associada com um ISP diferente, tentam transmitir ao mesmo tempo.

   b. Agora suponha que um AP opera no canal 1 e outro, no canal 11. Como você mudaria suas respostas?

P6. Na etapa 4 do protocolo CSMA/CA, uma estação que transmite um quadro com sucesso inicia o protocolo CSMA/CA para um segundo quadro na etapa 2, e não na 1. Quais seriam as razões que os projetistas do CSMA/CA provavelmente tinham em mente para fazer essa estação não transmitir o segundo quadro de imediato (se o canal fosse percebido como ocioso)?

P7. Suponha que uma estação 802.11b seja configurada para sempre reservar o canal com a sequência RTS/CTS. Imagine que essa estação de repente queira transmitir 1.500 bytes de dados e que todas as outras estações estão ociosas nesse momento. Calcule o tempo requerido para transmitir o quadro e receber o reconhecimento como uma função de SIFS e DIFS, ignorando atraso de propagação e admitindo que não haja erros de bits.

P8. Considere o cenário mostrado na Figura 7.31, no qual existem quatro nós sem fios, A, B, C e D. A cobertura de rádio dos quatro nós é mostrada pelas formas ovais mais escuras; todos os nós compartilham a mesma frequência. Quando A transmite, ele pode ser ouvido/recebido por B; quando B transmite, ele só pode ser ouvido/recebido por A e C; quando C transmite, B e D podem ouvir/receber de C; quando D transmite, somente C pode ouvir/receber de D.

Agora suponha que cada nó possua um estoque infinito de mensagens que ele queira enviar para os outros nós. Se o destinatário da mensagem não for um vizinho imediato, então a mensagem deve ser retransmitida. Por exemplo, se A quer enviar para D, uma mensagem de A deve ser primeiro enviada a B, que, então, envia a mensagem a C, e este, a D. O tempo é dividido em intervalos, com um tempo de transmissão de mensagem de exatamente um intervalo de tempo, como em um slotted Aloha, por exemplo. Durante um intervalo, um nó pode fazer uma das seguintes opções: (*i*) enviar uma mensagem; (*ii*) receber uma mensagem (se, exatamente, uma mensagem estiver sendo enviada a ele); (*iii*) permanecer silencioso. Como sempre, se um nó ouvir duas ou mais transmissões simultâneas, ocorrerá uma colisão, e nenhuma das mensagens transmitidas é recebida com sucesso. Você pode admitir aqui que não existem erros de bits, e, dessa forma, se uma mensagem for enviada, ela será recebida corretamente pelos que estão dentro do raio de transmissão do emissor.

a. Suponha que um controlador onisciente (i.e., que sabe o estado de cada nó na rede) possa comandar cada nó a fazer o que ele (o controlador onisciente) quiser, isto é, enviar uma mensagem, receber uma mensagem, ou permanecer silencioso. Dado esse controlador onisciente, qual é a taxa máxima à qual uma mensagem de dados pode ser transferida de C para A, sabendo que não existem outras mensagens entre nenhuma outra dupla remetente/destinatária?

b. Suponha que A envie uma mensagem a B, e D envie uma mensagem a C. Qual é a taxa máxima combinada à qual as mensagens de dados podem fluir de A a B e de D a C?

c. Considere agora que A envie uma mensagem a B, e C envie uma mensagem a D. Qual é a taxa máxima combinada à qual as mensagens de dados podem fluir de A a B e de C a D?

d. Suponha agora que os enlaces sem fio sejam substituídos por enlaces cabeados. Repita as questões de "a" a "c" neste cenário cabeado.

e. Agora imagine que estamos de novo em um cenário sem fio, e que para cada mensagem de dados enviada do remetente ao destinatário, este envie de volta uma mensagem ACK para o remetente (p. ex., como no TCP). Suponha também que cada mensagem ACK ocupe exatamente um slot. Repita as questões de "a" a "c" para este cenário.

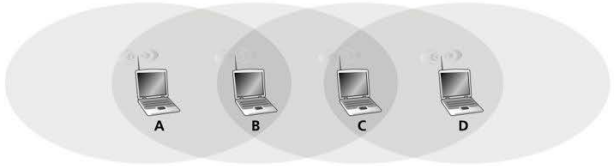


Figura 7.31 Cenário para o problema P8.

P9. Descreva o formato do quadro Bluetooth. Você precisará de uma leitura complementar para encontrar essa informação. Existe algo no formato do quadro que basicamente limite o número de nós ativos para oito em uma rede? Explique.

P10. Considere o seguinte cenário LTE ideal. O quadro "*downstream*", isto é, na direção do usuário móvel, é dividido em compartimentos de tempo e utiliza F frequências. Existem quatro nós, A, B, C e D, alcançáveis da estação-base a taxas de 10 Mbits/s, 5 Mbits/s, 2,5 Mbits/s e 1 Mbit/s, respectivamente, no canal *downstream*. Essas taxas pressupõem que a estação-base utiliza todos os intervalos de tempo disponíveis em todas as frequências F para transmitir para uma única estação. A estação-base possui infinitos dados para enviar a cada nó e pode enviar para qualquer um dos quatro nós usando qualquer uma das F frequências durante qualquer intervalo de tempo no subquadro *downstream*.

a. Qual é a taxa máxima à qual a estação-base pode enviar aos nós, admitindo que ela pode enviar a qualquer nó de sua escolha durante cada compartimento de tempo? Sua solução é justa? Explique e defina o que você quis dizer com "justo".

b. Se há requisito de equidade que todos os nós devem receber uma quantidade igual de dados durante cada quadro de *downstream*, qual é a taxa média de transmissão pela estação-base (para todos os nós) durante o subquadro de *downstream*? Explique como você chegou a essa resposta.

c. Suponha que, como critério de equidade, qualquer nó possa receber, no máximo, duas vezes tantos dados quanto qualquer outro nó durante o subquadro. Qual é a taxa média de transmissão pela estação-base (para todos os nós) durante o subquadro de *downstream*? Explique como você chegou a esta resposta.

P11. Na Seção 7.5, uma solução proposta que permitia que usuários móveis mantivessem seu endereço IP à medida que transitavam entre redes externas era fazer uma rede externa anunciar ao usuário móvel uma rota altamente específica e usar a infraestrutura de roteamento existente para propagar essa informação por toda a rede. Uma das preocupações que identificamos foi a escalabilidade. Suponha que, quando um usuário móvel passa de uma rede para outra, a nova rede externa anuncie uma rota específica para o usuário móvel e a antiga rede externa retire sua rota. Considere como informações de roteamento se propagam em um algoritmo vetor de distâncias (em particular para o caso de roteamento interdomínios entre redes que abrangem o globo terrestre).

a. Outros roteadores conseguirão rotear datagramas imediatamente para a nova rede externa tão logo essa rede comece a anunciar sua rota?

b. É possível que roteadores diferentes acreditem que redes externas diferentes contenham o usuário móvel?

c. Discuta a escala temporal segundo a qual outros roteadores na rede finalmente aprenderão o caminho até os usuários móveis.

P12. Em redes 4G/5G, que efeito terá a transferência sobre atrasos fim a fim de datagramas entre a fonte e o destino?

P13. Considere um dispositivo móvel que se liga a uma rede visitada *A*, e suponha que utiliza-se roteamento indireto para o dispositivo móvel a partir da sua rede nativa *H*. Posteriormente, em modo de roaming, o dispositivo sai do alcance da rede visitada *A* e entra no alcance da rede visitada LTE *B*. Projete um processo de transferência de uma estação-base *BS.A* na rede visitada *A* para uma estação-base *BS.B* na rede visitada *B*. Diagrame a série de passos que precisariam ser executados, tomando cuidado para identificar os elementos da rede envolvidos (e as redes às quais pertencem) para realizar essa transferência. Suponha que, após a transferência, o túnel da rede nativa à rede visitada termina na rede visitada *B*.

P14. Considere mais uma vez o cenário no Problema P13, mas agora suponha que o túnel da rede nativa *H* à rede visitada *A* continuará a ser usado. Em outras palavras, a rede visitada *A* servirá como ponto de âncora após a transferência. (A propósito: esse realmente é o processo usado para chamadas de voz por comutação de circuitos para um telefone móvel em modo de roaming em redes GSM 2G.) Nesse caso, túneis adicionais precisarão ser construídos para atingir o dispositivo móvel na sua rede visitada residente *B*. Mais uma vez, diagrame a série de passos que precisaria ser executada, tomando cuidado para identificar os elementos da rede envolvidos (e as redes às quais pertencem) para completar essa transferência.

Cite uma vantagem e uma desvantagem dessa abordagem em relação àquela adotada na sua solução para o Problema P13.

Wireshark Lab: WiFi

No site do livro, e também reproduzido no site para instrutores, você encontrará um Wireshark Lab, em inglês, para este capítulo, que captura e estuda os quadros 802.11 trocados entre um notebook sem fio e um ponto de acesso.