

Introdução

Cada um dos três séculos anteriores foi dominado por uma única nova tecnologia. O século XVIII foi a época dos grandes sistemas mecânicos que acompanharam a Revolução Industrial. O século XIX foi a era das máquinas a vapor. As principais conquistas tecnológicas do século XX se deram no campo da aquisição, do processamento e da distribuição de informações. Entre outros desenvolvimentos, vimos a instalação das redes de telefonia em escala mundial, a invenção do rádio e da televisão, o nascimento e o crescimento sem precedentes da indústria da informática, o lançamento dos satélites de comunicação e, naturalmente, a Internet.

Como resultado do rápido progresso tecnológico, essas áreas estão convergindo rapidamente no século XXI e as diferenças entre coleta, transporte, armazenamento e processamento de informações estão desaparecendo rapidamente. Organizações com centenas de escritórios dispersos por uma extensa área geográfica normalmente esperam, com um simples pressionar de um botão, poder examinar o status atual até mesmo de suas filiais mais remotas. À medida que cresce nossa capacidade de colher, processar e distribuir informações, torna-se ainda maior a demanda por formas mais sofisticadas de processamento de informação.

Apesar de a indústria de informática ainda ser jovem em comparação a outros setores (por exemplo, o de automóveis e o de transportes aéreos), foi simplesmente espetacular o progresso que os computadores experimentaram em um curto período. Durante as duas primeiras décadas de sua existência, os sistemas computacionais eram altamente centralizados, em geral instalados em uma grande sala, muitas vezes com paredes de vidro, através das quais os visitantes podiam contemplar, embevecidos, aquela grande maravilha eletrônica. Uma empresa de médio porte ou uma universidade contava apenas com um ou dois computadores, enquanto as grandes instituições tinham, no máximo, algumas dezenas. Era pura ficção científica a ideia de que, em quarenta anos, computadores muito mais poderosos, menores que selos postais, seriam produzidos em massa, aos bilhões.

A fusão dos computadores e comunicações teve uma profunda influência na forma como os sistemas computacionais são organizados. O conceito então dominante de ‘centro de computação’ como uma sala com um grande computador ao qual os usuários levam seu trabalho para processamento agora está completamente obsoleto (em-

bora os centros de dados com milhares de servidores de Internet estejam se tornando comuns). O velho modelo de um único computador atendendo a todas as necessidades computacionais da organização foi substituído por outro em que os trabalhos são realizados por um grande número de computadores separados, porém interconectados. Esses sistemas são chamados **redes de computadores**. O projeto e a organização dessas redes são os temas deste livro.

Ao longo do livro, utilizaremos a expressão ‘rede de computadores’ para indicar um conjunto de computadores autônomos interconectados por uma única tecnologia. Dois computadores estão interconectados quando podem trocar informações. A conexão não precisa ser feita por um fio de cobre; também podem ser usadas fibras ópticas, micro-ondas, ondas de infravermelho e satélites de comunicações. Existem redes de muitos tamanhos, modelos e formas, como veremos mais adiante. Elas normalmente estão conectadas para criar redes maiores, com a **Internet** sendo o exemplo mais conhecido de uma rede de redes.

Há uma certa confusão na literatura quanto à distinção entre uma rede de computadores e um **sistema distribuído**. A principal diferença entre eles é que, em um sistema distribuído, um conjunto de computadores independentes parece ser, para os usuários, um único sistema coerente. Em geral, ele tem um único modelo ou paradigma apresentado aos usuários. Com frequência, uma camada de software sobre o sistema operacional, chamada **middleware**, é responsável pela implementação desse modelo. Um exemplo bem conhecido de sistema distribuído é a **World Wide Web**. Ela trabalha em cima da Internet e apresenta um modelo em que tudo tem a aparência de um documento (página Web).

Em uma rede de computadores, essa coerência, esse modelo e esse software estão ausentes. Os usuários ficam expostos às máquinas reais, sem qualquer tentativa, por parte do sistema, de fazer as máquinas parecerem e atuarem de modo coerente. Se as máquinas tiverem hardware e sistemas operacionais distintos, isso será totalmente visível para os usuários. Se um usuário quiser executar um programa em uma máquina remota, ele terá de efetuar o logon nessa máquina e executar o programa lá.

Na prática, um sistema distribuído é um sistema de software instalado na parte superior de uma rede. O software dá ao sistema um alto grau de coesão e transparência. Consequentemente, é o software (e em particular o sistema

operacional) que determina a diferença entre uma rede e um sistema distribuído, não o hardware.

Apesar disso, há uma considerável sobreposição entre os dois assuntos. Por exemplo, os sistemas distribuídos e as redes de computadores precisam movimentar arquivos. A diferença está em quem chama a movimentação: o sistema ou o usuário. Embora este livro seja dedicado principalmente a redes, muitos dos assuntos também são importantes em sistemas distribuídos. Para obter mais informações sobre sistemas distribuídos, consulte Tanenbaum e Van Steen (2007).

1.1 | USOS DE REDES DE COMPUTADORES

Antes de começarmos a examinar detalhadamente as questões técnicas, vale a pena dedicar algum tempo para explicar por que as pessoas estão interessadas em redes de computadores e com que finalidade essas redes podem ser usadas. Afinal, se ninguém estivesse interessado em redes de computadores, poucas teriam sido montadas. Começaremos com os usos tradicionais em empresas, e depois passaremos para as redes domésticas e aos desenvolvimentos mais recentes relacionados a usuários móveis, terminando com as questões sociais.

1.1.1 | APLICAÇÕES COMERCIAIS

Muitas empresas têm um número significativo de computadores. Por exemplo, uma empresa pode ter um computador para cada trabalhador e os usa para projetar produtos, escrever documentos e elaborar a folha de pagamento. Inicialmente, alguns desses computadores podem funcionar isoladamente dos outros, mas, em determinado momento, a gerência pode decidir conectá-los para extrair e correlacionar informações sobre a empresa inteira.

Em termos um pouco mais genéricos, a questão aqui é o **compartilhamento de recursos**. O objetivo é deixar todos os programas, equipamentos e, especialmente, dados ao alcance de todas as pessoas na rede, independentemente da localização física do recurso ou do usuário. Um exemplo óbvio e bastante disseminado é um grupo de funcionários de um escritório que compartilham uma impressora comum. Nenhum dos indivíduos realmente necessita de uma impressora privativa, e uma impressora de grande capacidade conectada em rede muitas vezes é mais econômica, mais rápida e de manutenção mais fácil que um grande conjunto de impressoras individuais.

Porém, talvez mais importante que compartilhar recursos físicos, como impressoras e unidades de fita, seja compartilhar informações. Toda empresa, grande ou pequena, tem uma dependência vital de informações computadorizadas. A maioria das empresas tem registros de clientes, informações de produtos, estoques, extratos financeiros, informações sobre impostos e muitas outras informações on-line. Se todos os computadores de um banco sofressem uma pane, ele provavelmente não

duraria mais que cinco minutos. Uma instalação industrial moderna, com uma linha de montagem controlada por computadores, não duraria nem cinco segundos. Hoje, até mesmo uma pequena agência de viagens ou uma empresa jurídica com três pessoas depende intensamente de redes de computadores para permitir a seus funcionários acessar informações e documentos relevantes de forma quase instantânea.

Para empresas menores, todos os computadores provavelmente se encontram em um único escritório ou talvez em um único prédio; porém, no caso de empresas maiores, os computadores e funcionários podem estar dispersos por dezenas de escritórios e instalações em muitos países. Apesar disso, um vendedor em Nova York às vezes precisa acessar um banco de dados de estoque de produtos localizado em Cingapura. Redes chamadas **VPNs (Virtual Private Networks)** podem ser usadas para unir as redes individuais em diferentes locais em uma rede estendida. Em outras palavras, o mero fato de um usuário estar a 15 mil quilômetros de distância de seus dados não deve impedi-lo de usá-los como se eles fossem dados locais. Resumindo, trata-se de uma tentativa de dar fim à ‘tirania da geografia’.

No mais simples dos termos, é possível imaginar que o sistema de informações de uma empresa consista em um ou mais bancos de dados com informações da empresa e algum número de funcionários que necessitem acessá-los remotamente. Nesse modelo, os dados são armazenados em poderosos computadores chamados **servidores**. Normalmente, eles são instalados e mantidos em um local central por um administrador de sistemas. Ao contrário, os funcionários têm em suas mesas máquinas mais simples, chamadas **clientes**, com as quais acessam dados remotos, por exemplo, para incluir em planilhas eletrônicas que estão elaborando. (Algumas vezes, faremos referência ao usuário humano da máquina cliente como o ‘cliente’, mas deve ficar claro, pelo contexto, se estamos nos referindo ao computador ou a seu usuário.) As máquinas cliente e servidor são conectadas entre si por uma rede, como ilustra a Figura 1.1. Observe que mostramos a rede como uma simples elipse, sem qualquer detalhe. Utilizaremos essa forma quando mencionarmos uma rede no sentido mais abstrato. Quando forem necessários mais detalhes, eles serão fornecidos.

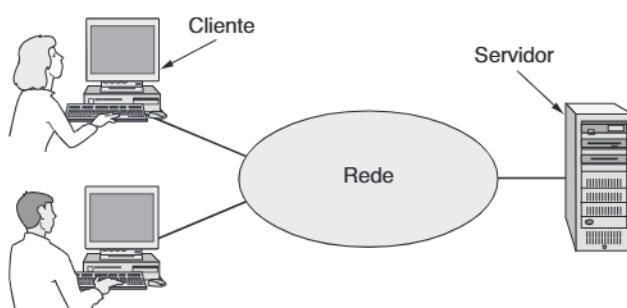


Figura 1.1 | Uma rede com dois clientes e um servidor.

Todo esse arranjo é chamado **modelo cliente-servidor**. Ele é bastante usado e forma a base de grande parte do uso da rede. A realização mais popular é a de uma **aplicação Web**, em que o servidor fornece páginas Web com base em seu banco de dados em resposta às solicitações do cliente. O modelo cliente-servidor é aplicável quando cliente e servidor estão, ambos, no mesmo prédio (e pertencem à mesma empresa), mas também quando estão muito afastados. Por exemplo, quando uma pessoa em casa acessa uma página na World Wide Web, o mesmo modelo é empregado, com o servidor Web remoto fazendo o papel do servidor e o computador pessoal do usuário sendo o cliente. Sob a maioria das condições, um único servidor pode lidar com um grande número (centenas ou milhares) de clientes simultaneamente.

Se examinarmos o modelo cliente-servidor em detalhes, veremos que dois processos (por exemplo, programas em execução) são envolvidos, um na máquina cliente e um na máquina servidora. A comunicação toma a forma do processo cliente enviando uma mensagem pela rede ao processo servidor. Então, o processo cliente espera por uma mensagem de resposta. Quando o processo servidor recebe a solicitação, ele executa o trabalho solicitado ou procura pelos dados solicitados e envia uma resposta de volta. Essas mensagens são mostradas na Figura 1.2.

Um segundo objetivo da configuração de uma rede de computadores está relacionado às pessoas, e não às informações ou mesmo aos computadores. Uma rede de computadores pode oferecer um poderoso **meio de comunicação** entre os funcionários. Praticamente toda empresa com dois ou mais computadores tem o recurso de **e-mail (correo eletrônico)**, que os funcionários utilizam de forma geral para suprir uma grande parte da comunicação diária. De fato, os funcionários trocam mensagens de e-mail sobre os assuntos mais corriqueiros, mas grande parte das mensagens com que as pessoas lidam diariamente não tem nenhum significado, porque os chefes descobriram que podem enviar a mesma mensagem (normalmente, sem conteúdo) a todos os seus subordinados, bastando pressionar um botão.

Ligações telefônicas entre os funcionários podem ser feitas pela rede de computadores, em vez de pela companhia telefônica. Essa tecnologia se chama **telefonia IP ou Voice over IP (VoIP)** quando a tecnologia da Internet é

empregada. O microfone e o alto-falante em cada extremo podem pertencer a um telefone habilitado para VoIP ou a computador do funcionário. As empresas descobriram que essa é uma forma maravilhosa de economizar nas contas telefônicas.

Outras formas de comunicação mais ricas são possíveis com as redes de computadores. O vídeo pode ser acrescentado ao áudio, de modo que os funcionários em locais distantes possam ver e ouvir um ao outro enquanto realizam uma reunião. Essa técnica é uma ferramenta eficiente para eliminar o custo e o tempo anteriormente dedicado às viagens. O **compartilhamento de desktop** permite que trabalhadores remotos vejam e interajam com uma tela de computador. Com isso, duas ou mais pessoas em locais distantes podem participar de uma reunião, vendo e ouvindo uns aos outros e até mesmo escrevendo um relatório em um quadro-negro compartilhado. Quando um funcionário faz uma mudança em um documento on-line, os outros podem ver a mudança imediatamente, em vez de esperar vários dias por uma carta. Essa agilidade facilita a cooperação entre grupos de pessoas dispersas, enquanto anteriormente isso era impossível. Atualmente, estão começando a ser usadas outras formas de coordenação remota mais ambiciosas, como a telemedicina (por exemplo, no monitoramento de pacientes remotos), mas elas podem se tornar muito mais importantes. Algumas vezes, diz-se que a comunicação e o transporte estão disputando uma corrida, e a tecnologia que vencer tornará a outra obsoleta.

Um terceiro objetivo para muitas empresas é realizar negócios eletronicamente, em especial com clientes e fornecedores. Esse novo modelo é chamado de **e-commerce (comércio eletrônico)** e tem crescido rapidamente nos últimos anos. Empresas aéreas, livrarias e outros varejistas descobriram que muitos clientes gostam da conveniência de fazer compras em casa. Consequentemente, muitas empresas oferecem catálogos de seus produtos e serviços e recebem pedidos on-line. Fabricantes de automóveis, aeronaves e computadores, entre outros, compram subsistemas de diversos fornecedores e depois montam as peças. Utilizando redes de computadores, os fabricantes podem emitir pedidos eletronicamente, conforme o necessário. Isso reduz a necessidade de grandes estoques e aumenta a eficiência.

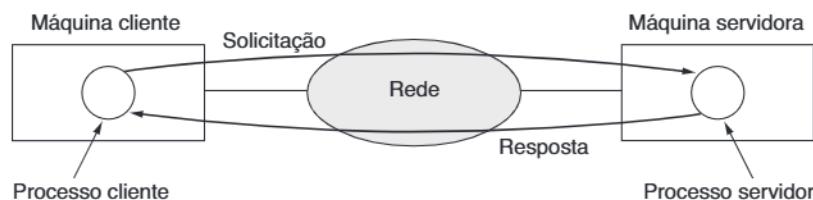


Figura 1.2 | O modelo cliente-servidor envolve solicitações e respostas.

I 1.1.2 APLICAÇÕES DOMÉSTICAS

Em 1977, Ken Olsen era presidente da Digital Equipment Corporation, então o segundo maior fornecedor de computadores de todo o mundo (depois da IBM). Quando lhe perguntaram por que a Digital não estava seguindo a tendência do mercado de computadores pessoais, ele disse: 'Não há nenhuma razão para qualquer indivíduo ter um computador em casa'. A história mostrou o contrário, e a Digital não existe mais. As pessoas inicialmente compravam computadores para processamento de textos e jogos. Nos últimos anos, talvez a maior motivação seja o acesso à Internet. Agora, muitos dispositivos eletrônicos de consumo, como conversores digitais, consoles de jogos e rádio-relógios, já vêm com computadores e redes de computadores embutidas, especialmente redes sem fio, e as redes domésticas são bastante usadas para entretenimento, incluindo escuta, exibição e criação de música, fotos e vídeos.

O acesso à Internet oferece, aos usuários domésticos, **conectividade** a computadores remotos. Assim como as empresas, os usuários domésticos podem acessar informações, comunicar-se com outras pessoas e comprar produtos e serviços com o comércio eletrônico. O principal benefício agora vem da conexão com o exterior da casa. Bob Metcalfe, o inventor da Ethernet, formulou a hipótese de que o valor de uma rede é proporcional ao quadrado do número de usuários, pois esse é aproximadamente o número de conexões diferentes que podem ser feitas (Gilder, 1993). Essa hipótese é conhecida como a 'lei de Metcalfe'. Ela ajuda a explicar como a tremenda popularidade da Internet vem de seu tamanho.

O acesso a informações remotas tem várias formas. Ele pode significar navegar na World Wide Web para obter informações ou apenas por diversão. As informações disponíveis incluem artes, negócios, culinária, governo, saúde, história, passatempos, recreação, ciência, esportes, viagens e muitos outros. A diversão surge sob tantas formas que fica difícil mencioná-las, e também se apresenta em outras formas que é melhor não serem mencionadas.

Muitos jornais são publicados on-line e podem ser personalizados. Por exemplo, às vezes é possível solicitar todas as informações sobre políticos corruptos, grandes incêndios, escândalos envolvendo celebridades e epidemias, mas dispensar qualquer notícia sobre esportes. Algumas vezes, até mesmo é possível transferir os artigos selecionados por download para o disco rígido enquanto você dorme. À medida que essa tendência continuar, ela causará desemprego maciço entre os jovens entregadores de jornais, mas as empresas jornalísticas gostam disso, porque a distribuição sempre foi o elo mais fraco na cadeia de produção inteira. Naturalmente, para que esse modelo funcione, primeiro eles terão de descobrir como ganhar dinheiro nesse novo mundo, algo que não é totalmente óbvio, pois os usuários da Internet esperam que tudo seja de graça.

A próxima etapa, além de jornais (e de revistas e publicações científicas), é a biblioteca digital on-line. Muitas organizações profissionais, como a ACM (www.acm.org) e a IEEE Computer Society (www.computer.org), já têm muitas publicações e anais de conferências on-line. Leitores de e-books (livros eletrônicos) e bibliotecas on-line podem tornar os livros impressos obsoletos. Os cépticos devem observar o efeito que a máquina de impressão teve sobre os manuscritos medievais com iluminuras.

Grande parte dessa informação é acessada por meio do modelo cliente-servidor, mas existe um modelo diferente, popular, para acessar informações, que recebe o nome de comunicação **peer-to-peer** (ou não hierárquica) (Parantheswaran et al., 2001). Nessa forma de comunicação, indivíduos que constituem um grupo livre podem se comunicar com outros participantes do grupo, como mostra a Figura 1.3. Em princípio, toda pessoa pode se comunicar com uma ou mais pessoas; não existe qualquer divisão estrita entre clientes e servidores.

Muitos sistemas peer-to-peer, como BitTorrent (Cohen, 2003), não possuem qualquer banco de dados de conteúdo central. Em vez disso, cada usuário mantém seu próprio banco de dados no local e oferece uma lista de

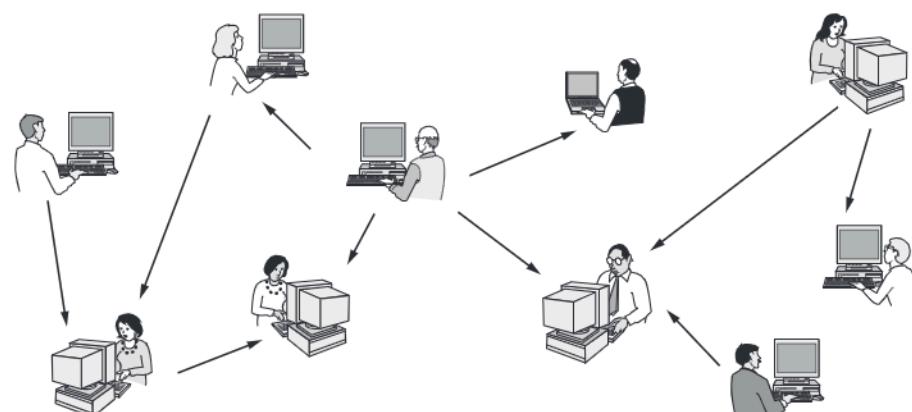


Figura 1.3 | Em um sistema não hierárquico não existem clientes e servidores.

outras pessoas vizinhas que são membros do sistema. Um novo usuário pode, então, ir até qualquer membro existente para ver o que ele tem e obter os nomes de outros membros para inspecionar mais conteúdo e mais nomes. Esse processo de pesquisa pode ser repetido indefinidamente, até criar um grande banco de dados local do que existe no sistema inteiro. Essa é uma atividade que seria tediosa para as pessoas, mas os computadores podem sobressair nisso.

A comunicação peer-to-peer normalmente é usada para compartilhar música e vídeos. Ela realmente alcançou o auge por volta de 2000, com um serviço de compartilhamento de música chamado Napster, que foi encerrado depois daquilo que provavelmente foi a maior violação de direitos autorais em toda a história registrada (Lam e Tan, 2001; Macedonia, 2000). Também existem aplicações legais para a comunicação peer-to-peer. Entre elas estão fãs compartilhando músicas de domínio público, famílias compartilhando fotos e filmes, e usuários baixando pacotes de software públicos. Na verdade, uma das aplicações mais populares de toda a Internet, o correio eletrônico, é basicamente peer-to-peer. Essa forma de comunicação provavelmente crescerá bastante no futuro.

Todas as aplicações anteriores envolvem interações entre uma pessoa e um banco de dados remoto, repleto de informações. A segunda grande categoria de utilização de redes é a comunicação entre pessoas, basicamente a resposta do século XXI ao telefone do século XIX. O correio eletrônico (e-mail) já é usado diariamente por milhões de pessoas em todo o mundo e seu uso está crescendo rapidamente. Em geral, ele já contém áudio e vídeo, além de texto e imagens. O odor pode demorar um pouco mais.

Hoje em dia, qualquer adolescente é fanático pela **troca de mensagens instantâneas**. Esse recurso, derivado do programa *talk* do UNIX, em uso desde aproximadamente 1970, permite que duas pessoas digitem mensagens uma para a outra em tempo real. Também existem serviços de mensagens para várias pessoas, como o serviço **Twitter**, permitindo que as pessoas enviem pequenas mensagens de texto, chamadas ‘tweets’, para seu círculo de amigos ou outras audiências interessadas.

A Internet pode ser usada pelas aplicações para transportar áudio (por exemplo, estações de rádio pela Internet) e vídeo (por exemplo, o YouTube). Além de ser um modo barato de se comunicar com amigos distantes, essas aplicações podem oferecer experiências ricas, como teleaprendizado, o que significa assistir a aulas às 8 da manhã sem a inconveniência de ter de levantar da cama. Com o passar do tempo, o uso das redes para melhorar a comunicação entre os seres humanos poderá ser mais importante do que qualquer outro. Isso pode se tornar extremamente importante para pessoas que estão geograficamente distantes, dando-lhes o mesmo acesso aos serviços que as pessoas morando no meio de um grande centro urbano já têm.

Entre as comunicações interpessoais e o acesso à informação estão as aplicações de **rede social**. Aqui, o fluxo de informações é controlado pelos relacionamentos que as pessoas declaram umas às outras. Uma das redes sociais mais populares é o **Facebook**. Ela permite que as pessoas atualizem seus perfis pessoais e compartilhem as atualizações com outras pessoas de quem elas declararam ser amigas. Outras aplicações de rede social podem fazer apresentações por meio de amigos dos amigos, enviar mensagens de notícias aos amigos, como o Twitter, e muito mais.

Ainda de forma mais livre, os grupos de pessoas podem trabalhar juntos para criar conteúdo. Uma **wiki**, por exemplo, é um website colaborativo que os membros de uma comunidade editam. A wiki mais famosa é a **Wikipedia**, uma encyclopédia que qualquer um pode editar, mas existem milhares de outras wikis.

Nossa terceira categoria é o comércio eletrônico no sentido mais amplo do termo. A atividade de fazer compras em casa já é popular e permite ao usuário examinar catálogos on-line de milhares de empresas. Alguns desses catálogos são interativos, mostrando produtos de diferentes pontos de vista e em configurações que podem ser personalizadas. Depois que um cliente compra um produto eletronicamente, se ele não souber como usá-lo, o suporte técnico on-line poderá ser consultado.

Outra área em que o comércio eletrônico já é uma realidade é o acesso a instituições financeiras. Muitas pessoas já pagam suas contas, administram contas bancárias e manipulam seus investimentos eletronicamente. Sem dúvida, essa tendência crescerá à medida que as redes se tornarem mais seguras.

Uma área que praticamente ninguém previu é a de brechós eletrônicos (e-brechó?). Leilões on-line de objetos de segunda mão tornaram-se uma indústria próspera. Diferente do comércio eletrônico tradicional, que segue o modelo cliente-servidor, os leilões on-line são um tipo de sistema peer-to-peer, no sentido de que os consumidores podem atuar como compradores e vendedores.

Algumas dessas formas de comércio eletrônico utilizam pequenas abreviações baseadas no fato de que ‘to’ e ‘2’ têm a mesma pronúncia em inglês. As mais populares estão relacionadas na Tabela 1.1.

Nossa quarta categoria é o entretenimento. Ela tem feito grande progresso nas residências ao longo dos últimos anos, com a distribuição de música, programas de rádio e televisão, e os filmes pela Internet começando a competir com os meios tradicionais. Os usuários podem localizar, comprar e baixar músicas em MP3 e filmes com qualidade de DVD e depois incluí-los em sua coleção pessoal. Os programas da TV agora alcançam muitos lares via sistemas de **IPTV** (**IP TeleVision**), que são baseados na tecnologia IP, em vez das transmissões de TV a cabo ou rádio. As aplicações de mídia permitem que os usuários sintonizem estações de rádio pela Internet ou assistam a episódios recentes

Abreviação	Nome completo	Exemplo
B2C	Business-to-consumer	Pedidos de livros on-line
B2B	Business-to-business	Fabricante de automóveis solicitando pneus a um fornecedor
G2C	Government-to-consumer	Governo distribuindo eletronicamente formulários de impostos
C2C	Consumer-to-consumer	Leilões on-line de produtos usados
P2P	Peer-to-peer	Compartilhamento de música

Tabela 1.1 | Algumas formas de comércio eletrônico.

de seus programas favoritos na TV. Naturalmente, todo esse conteúdo pode ser passado aos diferentes aparelhos, monitores e alto-falantes de sua casa, normalmente com uma rede sem fio.

Logo, talvez seja possível selecionar qualquer filme ou programa de televisão, qualquer que seja a época ou país em que tenha sido produzido, e exibi-lo em sua tela no mesmo instante. Novos filmes poderão se tornar interativos e, ocasionalmente, o usuário poderá ser solicitado a interferir no roteiro (*Macbeth deve matar Duncan ou esperar pelo momento certo?*), com cenários alternativos para todas as hipóteses. A televisão ao vivo também poderá se tornar interativa, com os telespectadores participando de programas de perguntas e respostas, escolhendo entre correntes e assim por diante.

Outra forma de entretenimento são os jogos eletrônicos. Já temos jogos de simulação em tempo real com vários participantes, como os de esconde-esconde em um labirinto virtual, e simuladores de voo em que os jogadores de uma equipe tentam abater os jogadores da equipe adversária. Os mundos virtuais oferecem um ambiente persistente, em que milhares de usuários podem experimentar uma realidade compartilhada com gráficos tridimensionais.

Nossa última categoria é a **computação ubíqua**, em que a computação está embutida no dia a dia, como na visão de Mark Weiser (1991). Muitos lares já estão preparados com sistemas de segurança que incluem sensores em portas e janelas, e existem muitos outros sensores que podem ser embutidos em um monitor doméstico inteligente, como no consumo de energia. Seus medidores de eletricidade, gás e água também poderiam informar o uso pela rede. Isso economizaria dinheiro, pois não seria preciso enviar funcionários para ler a medição do consumo. E seus detectores de fumaça poderiam ligar para os bombeiros em vez de fazer muito barulho (o que não adianta muito se não houver alguém em casa). À medida que o custo dos sensores e da comunicação diminui, mais e mais aplicações de medição e envio de informações serão disponibilizadas pelas redes.

Cada vez mais, os dispositivos eletrônicos de consumo estão em rede. Por exemplo, algumas câmeras de última

geração já possuem capacidade para rede sem fio, utilizada para enviar fotos a um monitor próximo, para serem exibidas. Fotógrafos profissionais de esportes também podem enviar fotos para seus editores em tempo real, primeiro sem fio, para um ponto de acesso, e em seguida para a Internet. Dispositivos como televisores que se conectam na tomada da parede podem usar a **rede de energia elétrica** para enviar informações pela casa, nos fios que transportam eletricidade. Pode não ser surpresa ter esses objetos na rede, mas objetos que não imaginamos como computadores também podem detectar e comunicar informações. Por exemplo, seu chuveiro poderá registrar o uso de água, dando-lhe um *feedback* visual enquanto você se ensaboá, e informar para uma aplicação de monitoramento ambiental doméstico quando tiver terminado, para ajudá-lo a economizar em sua conta de água.

Uma tecnologia chamada **RFID (Radio Frequency IDentification)** levará essa ideia ainda mais adiante no futuro. Etiquetas RFID são chips passivos (ou seja, não têm bateria), do tamanho de um selo, e já podem ser afixados em livros, passaportes, animais de estimação, cartões de crédito e outros itens em casa e fora dela. Isso permite que os leitores de RFID localizem e se comuniquem com os itens por uma distância de até vários metros, dependendo do tipo de RFID. Originalmente, a RFID foi comercializada para substituir os códigos de barras. Ela ainda não teve muito sucesso porque os códigos de barras são gratuitos e as etiquetas de RFID custam alguns centavos. Naturalmente, as etiquetas de RFID oferecem muito mais e seu preço está caindo rapidamente. Elas podem transformar o mundo real na Internet de coisas (ITU, 2005).

1.1.3 USUÁRIOS MÓVEIS

Computadores móveis, como notebooks e computadores portáteis, constituem um dos segmentos de mais rápido crescimento no setor de informática. Suas vendas já superaram as de computadores desktop. Por que alguém desejaría um? As pessoas que estão em trânsito normalmente desejam usar seus dispositivos móveis para ler e enviar e-mail, ‘tuitar’, assistir a filmes, baixar música, jogar ou simplesmente navegar pelas informações na Web. Elas querem fazer todas as coisas que fazem em casa e no es-

critório. Naturalmente, elas querem fazê-las em qualquer lugar, na terra, no mar ou no ar.

A **conectividade** à Internet habilita muitos desses usos móveis. Como ter uma conexão com fios é impossível nos carros, barcos e aviões, há muito interesse nas redes sem fio. As redes celulares operadas pelas empresas de telefonia são um tipo conhecido de rede sem fio que dá cobertura para telefones móveis. Os **hotspots** sem fio baseados no padrão 802.11 são outro tipo de rede sem fio para computadores móveis. Eles surgem em todo lugar a que as pessoas vão, resultando em uma malha com cobertura em cafés, hotéis, aeroportos, escolas, trens e aviões. Qualquer um com um laptop e um modem sem fio pode simplesmente ligar seu computador e estar conectado à Internet pelo hotspot, como se o computador estivesse conectado a uma rede com fios.

As redes sem fio têm grande valor para frotas de caminhões, táxis, veículos de entrega e funcionários de serviços de assistência técnica, que precisam manter-se em contato com sua base de operações. Por exemplo, em muitas cidades, os motoristas de táxi são homens de negócios independentes, em vez de serem funcionários de uma empresa de táxi. Em algumas dessas cidades, os táxis têm uma tela de vídeo que o motorista pode observar. Ao receber uma chamada de cliente, um despachante central digita os pontos de partida e destino. Essa informação é exibida nas telas de vídeo dos motoristas e um aviso sonoro é emitido. O primeiro motorista a pressionar um botão na tela de vídeo obtém a corrida.

As redes sem fios também são importantes para os militares. Se, de uma hora para outra, for necessário travar uma guerra em qualquer lugar no mundo, talvez não seja possível contar com a possibilidade de usar a infraestrutura de rede local. Será melhor levar seu próprio equipamento de rede.

Embora as redes sem fios e a computação móvel frequentemente estejam relacionadas, elas não são idênticas, como mostra a Tabela 1.2. Aqui, observamos uma distinção entre redes **sem fio fixas** e **sem fio móveis**. Algumas vezes, até mesmo os notebooks podem estar conectados por fios. Por exemplo, se um viajante conecta um notebook à tomada de rede em um quarto de hotel, ele tem mobilidade sem precisar utilizar uma rede sem fio.

Sem fio	Móvel	Aplicações típicas
Não	Não	Computadores desktop em escritórios
Não	Sim	Um notebook usado em um quarto de hotel
Sim	Não	Redes em edifícios que não dispõem de fiação
Sim	Sim	PDA para registrar o estoque de uma loja

Tabela 1.2 | Combinações de redes sem fio e computação móvel.

Por outro lado, alguns computadores sem fio não são portáteis. Em casa e nos escritórios ou hotéis que não possuem cabeamento adequado, pode ser mais conveniente conectar computadores desktop ou aparelhos sem fio do que instalar os fios. A instalação de uma rede sem fio pode exigir pouco mais do que adquirir uma pequena caixa com alguns componentes eletrônicos, retirá-la da embalagem e conectá-la ao equipamento. Essa solução pode ser muito mais barata do que pedir que um trabalhador monte conduites para passar fiação no prédio.

Finalmente, também há as verdadeiras aplicações móveis, sem fios, como pessoas percorrendo lojas com um PDA e registrando o estoque. Em muitos aeroportos mais cheios, os funcionários de devolução de carros alugados trabalham no estacionamento com computadores móveis sem fios. Eles leem os códigos de barras ou chips de RFID dos carros devolvidos, e seu dispositivo móvel, que possui uma impressora embutida, chama o computador principal, recebe a informação da locação e imprime a conta no ato.

Talvez o impulso fundamental das aplicações móveis, sem fio, seja o telefone móvel. O envio de **mensagens de texto**, ou **torpedos**, é tremendamente popular. Ele permite que um usuário de telefone móvel digite uma mensagem curta que é, então, entregue pela rede celular para outro assinante móvel. Poucas pessoas teriam previsto, há dez anos, que adolescentes digitando mensagens de texto curtas em telefones móveis seria uma grande forma de ganhar dinheiro para as companhias telefônicas. Mas o envio de texto (ou **Short Message Service**, como é conhecido fora dos Estados Unidos) é muito lucrativo, pois custa à operadora uma pequena fração de um centavo para repassar uma mensagem de texto, um serviço pelo qual elas cobram muito mais.

A tão aguardada convergência entre telefones e a Internet finalmente chegou, e acelerará o crescimento das aplicações móveis. **Smartphones**, como o popular iPhone, combinam aspectos de telefones e computadores móveis. As redes celulares (3G e 4G) às quais eles se conectam podem oferecer serviços de dados rápidos para usar a Internet, bem como para lidar com ligações telefônicas. Muitos telefones mais avançados também se conectam a hotspots sem fio, e automaticamente alternam entre as redes para escolher a melhor opção para o usuário.

Outros aparelhos eletrônicos também podem usar redes celulares e hotspot para permanecer conectados a computadores remotos. Os leitores de livros eletrônicos podem baixar um livro recém-adquirido, a próxima edição de uma revista ou o jornal de hoje, onde quer que eles estejam. Os porta-retratos eletrônicos podem ser atualizados automaticamente com imagens novas.

Como os telefones móveis têm conhecimento de suas localizações, normalmente porque são equipados com receptores de **GPS** (**Global Positioning System**), alguns serviços são intencionalmente dependentes do local. Ma-

pas móveis e orientações são candidatos óbvios, visto que seu telefone habilitado com GPS e seu carro provavelmente têm uma ideia melhor de onde você está do que você mesmo. O mesmo pode acontecer com as buscas por uma livraria próxima ou um restaurante chinês, ou a previsão do tempo local. Outros serviços podem registrar o local, como a anotação de fotos e vídeos com o local em que foram feitos. Essa anotação é conhecida como ‘geomarcação’.

Uma área em que esses dispositivos podem se destacar é chamada **m-commerce (mobile-commerce)** (Senn, 2000). Mensagens de texto curtas do telefone móvel são usadas para autorizar pagamentos de alimentos em máquinas, ingressos de cinema e outros itens pequenos, ao invés de dinheiro em espécie e cartões de crédito/débito. O débito aparece, então, na conta do telefone celular. Quando equipado com tecnologia **NFC (Near Field Communication)**, o telefone móvel pode atuar como um smartcard com RFID e interagir com um leitor próximo para realizar o pagamento. A força motriz por trás desse fenômeno consiste em uma mistura de fabricantes de PDAs sem fios e operadores de redes que estão tentando descobrir como obter uma fatia do comércio eletrônico. Do ponto de vista da loja, esse esquema pode poupar-lhes a maior parte das tarifas da empresa de cartões de crédito, o que pode significar uma porcentagem elevada. É claro que esse plano pode ter efeito contrário ao desejado, pois os clientes de uma loja poderiam usar os leitores de RFID ou código de barras em seus dispositivos móveis para verificar os preços dos concorrentes antes de comprar e, depois, obter instantaneamente um relatório detalhado de onde mais o item poderia ser adquirido na região e a que preço.

Uma enorme vantagem do m-commerce é que os usuários de telefones celulares se acostumaram a pagar por tudo (ao contrário dos usuários da Internet, que esperam conseguir tudo de graça). Se um website cobrasse uma taxa para permitir a seus clientes efetuar pagamentos com cartão de crédito, haveria uma imensa reclamação dos usuários. Se, no entanto, uma operadora de telefonia celular permitisse às pessoas pagar por itens de uma loja usando o telefone e depois cobrassem uma tarifa por essa conveniência, provavelmente isso seria aceito como algo normal. O tempo dirá.

Não há dúvida de que os usos dos computadores móveis e sem fio aumentarão rapidamente no futuro, à medida que o tamanho dos computadores diminui, provavelmente de maneiras que ninguém é capaz de prever. Vejamos algumas das possibilidades. **Redes de sensores** são compostas de nós que colhem e repassam, sem fios, as informações que eles detectam sobre o estado do mundo físico. Os nós podem fazer parte de itens familiares, como carros ou telefones, ou então podem ser pequenos dispositivos separados. Por exemplo, seu carro poderia colher dados sobre sua localização, velocidade, vibração e economia a partir de seu sistema de diagnóstico de bordo e enviar essa informação para um banco de dados (Hull et al., 2006). Esses

dados podem ajudar a localizar buracos, planejar viagens evitando estradas congestionadas e informar se você é um ‘devorador de combustível’, em comparação com outros motoristas no mesmo trecho de estrada.

Redes de sensores estão revolucionando a ciência oferecendo diversos dados sobre o comportamento que não poderiam ser observados anteriormente. Um exemplo é o rastreamento da migração de zebras individuais, colocando um pequeno sensor em cada animal (Juang et al., 2002). Os pesquisadores inseriram um computador sem fio em um cubo de 1 mm de borda (Warneke et al., 2001). Com computadores móveis desse tamanho, até mesmo pássaros, roedores e insetos podem ser rastreados.

Até mesmo usos rotineiros, como em parquímetros, podem ser significativos, pois utilizam dados que anteriormente não estavam disponíveis. Os parquímetros sem fio podem aceitar pagamentos com cartão de crédito ou débito, com verificação instantânea pelo enlace sem fio. Eles também podem relatar quando estão em uso pela rede sem fio. Isso permite aos motoristas baixar um mapa de estacionamento atual para seu carro, de modo que podem encontrar um ponto disponível mais facilmente. É claro que, quando um parquímetro expira, ele também pode verificar a presença de um carro (emitindo um sinal a partir dele) e informar a expiração ao agente de estacionamento. Estima-se que os municípios dos Estados Unidos poderiam coletar mais de US\$ 10 bilhões apenas dessa maneira (Harte et al., 2000).

Computadores embarcados são outra aplicação promissora. Relógios inteligentes com rádios fazem parte de nosso espaço mental desde seu aparecimento nas tiras de quadrinhos de Dick Tracy em 1946; agora você pode comprá-los. Outros dispositivos desse tipo podem ser implantados, como marca-passos e bombas de insulina. Alguns deles podem ser controlados por uma rede sem fio. Isso permite que os médicos os testem e reconfigurem mais facilmente. Isso também poderia levar a alguns problemas desagradáveis se os dispositivos forem tão seguros como um PC comum, e puderem ser facilmente adulterados (Halperin et al., 2008).

I 1.1.4 QUESTÕES SOCIAIS

As redes de computadores, assim como a imprensa há cerca de 500 anos, permitem que os cidadãos comuns manifestem suas opiniões de maneiras que não eram possíveis anteriormente. Mas, junto com o lado bom vem o lado ruim, pois essa nova liberdade traz consigo uma série de questões sociais, políticas e éticas. Vamos mencionar rapidamente algumas delas; um estudo completo exigiria um livro inteiro, pelo menos.

Redes sociais, quadros de mensagens, sites de compartilhamento de conteúdo e uma série de outras aplicações permitem que as pessoas compartilhem suas ideias com indivíduos de mesmo pensamento. Desde que os assuntos

sejam restritos a assuntos técnicos ou passatempos como jardinagem, não surgirão muitos problemas.

Os problemas começam a vir à tona quando as pessoas abordam temas com os quais as pessoas realmente se preocupam, como política, religião ou sexo. Os pontos de vista divulgados podem ser altamente ofensivos para algumas pessoas. Pior ainda, eles podem não ser politicamente corretos. Além disso, as opiniões não estão obrigatoriamente limitadas ao texto; fotos coloridas de alta resolução e mesmo pequenos videoclipes podem ser facilmente compartilhados pelas redes de computadores. Algumas pessoas adotam a visão de que cada um sabe o que faz, mas outras acham que a publicação de certos tipos de materiais (por exemplo, ataques a determinados países ou religiões, pornografia etc.) é simplesmente inaceitável e tem de ser censurada. Diferentes países têm leis distintas e conflitantes sobre esse assunto. Assim, essa polêmica fica cada vez mais acirrada.

No passado, as pessoas abriram processos contra operadores de redes, partindo do princípio de que, a exemplo do que ocorre com os jornais e revistas, eles têm de assumir a responsabilidade pelo conteúdo do que publicam. A resposta inevitável é que uma rede é como uma companhia telefônica ou uma empresa de correios, e que não se pode esperar que ela censure seus usuários.

Agora, deve ser pouco surpreendente descobrir que alguns operadores de rede bloqueiam o conteúdo por seus próprios motivos. Alguns usuários de aplicações peer-to-peer tiveram seus serviços de rede suspensos porque os operadores de rede não acharam lucrativo transportar as grandes quantidades de tráfego enviadas por essas aplicações. Esses mesmos operadores provavelmente gostariam de tratar diferentes empresas de formas diferentes. Se você é uma empresa grande e paga bem, então receberá um bom serviço, mas se você é um peixe pequeno, receberá um serviço fraco. Os oponentes dessa prática argumentam que o peer-to-peer e outro conteúdo deverá ser tratado da mesma maneira, pois todos são apenas bits para a rede. Esse argumento para as comunicações que não são diferenciadas por seu conteúdo, origem ou por quem está fornecendo o conteúdo é conhecido como **neutralidade da rede** (Wu, 2003). Provavelmente, é seguro dizer que esse debate continuará por algum tempo.

Muitas outras partes estão envolvidas na briga pelo conteúdo. Por exemplo, música e filmes pirateados alimentaram o crescimento maciço das redes peer-to-peer, que não agradaram os proprietários dos direitos autorais, os quais ameaçaram (e às vezes tomaram) ações legais. Atualmente, existem sistemas automatizados que procuram redes peer-to-peer e disparam avisos aos operadores e usuários da rede que são suspeitos de infringir direitos autorais. Nos Estados Unidos, esses avisos são conhecidos como **notas de demolição DMCA** pelo **Digital Millennium Copyright Act**. Essa busca é uma corrida armamentista, pois é difícil

apanhar a infração de direitos autorais de forma confiável. Até mesmo sua impressora poderia ser pega por engano como uma culpada (Piatek et al., 2008).

As redes de computadores facilitam muito a comunicação. Elas também tornam fácil bisbilhotar o tráfego para as pessoas que controlam a rede. Isso cria conflitos por questões como direitos dos funcionários *versus* direitos do empregador. Muitas pessoas leem e escrevem e-mail no trabalho. Muitos empregadores reivindicaram o direito de ler e possivelmente censurar as mensagens dos funcionários, incluindo mensagens enviadas de um computador doméstico fora dos horários de trabalho. Nem todos os funcionários concordam com isso, especialmente com a última parte.

Outro tópico importante é a relação entre o governo e os direitos dos cidadãos. O FBI instalou um sistema em muitos provedores de serviços da Internet para bisbilhotar todas as mensagens de correio eletrônico que entram e saem, em busca de fragmentos de interesse. O sistema foi originalmente chamado Carnivore, mas a publicidade ruim fez com que ele fosse renomeado com a sigla aparentemente mais inocente DCS1000 (Blaze e Bellovin, 2000; Sobel, 2001; e Zack, 2001). No entanto, seu objetivo ainda é espionar milhões de pessoas, na esperança de encontrar informações sobre atividades ilegais. Infelizmente para os espiões, a Quarta Emenda à Constituição dos Estados Unidos proíbe buscas do governo sem um mandado de busca, mas o governo ignora isso com frequência.

Naturalmente, o governo não tem o monopólio das ameaças à privacidade das pessoas. O setor privado também faz sua parte, **traçando perfis** dos usuários. Por exemplo, pequenos arquivos, chamados cookies, que os navegadores da Web armazenam nos computadores dos usuários, permitem que as empresas controlem as atividades desses usuários no ciberespaço e também podem permitir que números de cartões de crédito, números de CPF e outras informações confidenciais vazem pela Internet (Berghel, 2001). As empresas que oferecem serviços baseados na Web podem manter grandes quantidades de informações pessoais sobre seus usuários, permitindo-lhes estudar diretamente as atividades do usuário. Por exemplo, o Google pode ler seu e-mail e mostrar propagandas com base em seus interesses, se você usar seu serviço de e-mail, o **Gmail**.

Uma nova guinada com os dispositivos móveis é a privacidade de local (Beresford e Stajano, 2003). Como parte do processo de oferecer serviços a seu dispositivo móvel, os operadores da rede descobrem onde você está em diferentes momentos do dia. Isso lhes permite rastrear seus movimentos. Eles podem saber qual clube você frequenta e qual centro médico você visita.

As redes de computadores também oferecem o potencial para aumentar a privacidade por meio do envio de mensagens anônimas. Em algumas situações, esse recurso pode ser desejável. Além de impedir que as empresas

descubram seus hábitos, ele proporciona, por exemplo, um meio para alunos, soldados, trabalhadores e cidadãos denunciarem o comportamento ilegal de professores, oficiais, superiores e políticos sem medo de possíveis represálias. Por outro lado, nos Estados Unidos e na maioria dos países democráticos, a lei permite especificamente às pessoas acusadas o direito de se confrontarem com o acusador em juízo, de modo que acusações anônimas não podem ser usadas como evidência.

A Internet torna possível encontrar informações com rapidez, mas uma grande parte dessas informações é incorreta, enganosa ou totalmente errada. Aquele aconselhamento médico que você conseguiu na Internet sobre sua dor no peito pode ter vindo de um ganhador do Prêmio Nobel ou de alguém que abandonou os estudos no ensino médio.

Outras informações constantemente são indesejadas. O lixo de correio eletrônico (spam) se tornou parte de nossa vida, pois os spammers coletam milhões de endereços de e-mail e os pretensos profissionais de marketing podem enviar mensagens geradas pelo computador com um custo muito baixo. A inundação de spam resultante compete com o fluxo de mensagens de pessoas reais. Felizmente, o software de filtragem consegue ler e descartar o spam gerado por outros computadores, com menores ou maiores graus de sucesso.

Outro tipo de conteúdo visa a um comportamento criminoso. As páginas Web e as mensagens de e-mail com conteúdo ativo (basicamente, programas ou macros executados na máquina do receptor) podem conter vírus capazes de devastar seu computador. Eles podem ser usados para roubar suas senhas de conta bancária, ou fazer com que seu computador envie spam como parte de uma **botnet** ou um pool de máquinas infectadas.

O **roubo de identidade** (ou **phishing**) finge estar sendo originado de uma parte confiável — por exemplo, seu banco — para tentar roubar informações confidenciais — por exemplo, números de cartão de crédito. O roubo de identidade está se tornando um problema sério, pois os ladrões coletam informações suficientes sobre a pessoa para obter cartões de crédito e outros documentos em nome da vítima.

Pode ser difícil impedir que os computadores se passem por pessoas na Internet. Esse problema levou ao desenvolvimento de **CAPTCHAs**, em que um computador pede a uma pessoa para resolver uma pequena tarefa de reconhecimento, por exemplo, digitar as letras mostradas em uma imagem distorcida, para mostrar que são humanos (von Ahn, 2001). Esse processo é uma variação do famoso teste de Turing, em que uma pessoa faz perguntas por uma rede para julgar se a entidade que responde é humana.

Muitos desses problemas poderiam ser resolvidos se a indústria de informática levasse a sério a segurança dos computadores. Se todas as mensagens fossem criptogra-

fadas e autenticadas, seria mais difícil haver danos. Essa tecnologia está bem estabelecida e será estudada em detalhes no Capítulo 8. O problema é que os fornecedores de hardware e software sabem que a inclusão de recursos de segurança custa dinheiro, e seus clientes não exigem tais recursos. Além disso, um número substancial dos problemas é causado por bugs de software, que ocorrem porque os fornecedores continuam a acrescentar mais e mais recursos a seus programas, o que inevitavelmente significa mais código e, portanto, mais bugs. Um imposto sobre novos recursos poderia ajudar, mas isso talvez dificultasse as vendas em poucos trimestres. Um programa de reembolso por software defeituoso talvez fosse ótimo, exceto pelo fato de levar à bancarrota toda a indústria de software no primeiro ano.

As redes de computadores levantam novos problemas legais quando interagem com leis antigas. Jogos de apostas eletrônicos são um exemplo. Os computadores têm simulado coisas há décadas, logo, por que não simular máquinas caça-níqueis, roletas, apostas de vinte e um e outros equipamentos de jogo? Bem, porque é ilegal em muitos lugares. O problema é que o jogo é legal em muitos outros lugares (Inglaterra, por exemplo) e os donos de cassinos de lá entenderam o potencial para jogar pela Internet. O que acontece se o jogador, o cassino e o servidor estiverem em países diferentes, com leis em conflito? Boa pergunta.

1.2 | HARDWARE DE REDE

Agora é hora de voltarmos nossa atenção das aplicações e aspectos sociais das redes (a diversão) para as questões técnicas envolvidas no projeto da rede (o trabalho). Não existe nenhuma taxonomia de aceitação geral na qual todas as redes de computadores se encaixam, mas duas dimensões se destacam das demais: a tecnologia de transmissão e a escala. Vamos examinar cada uma delas.

Em termos gerais, há dois tipos de tecnologias de transmissão em uso disseminado nos dias de hoje: enlaces de **broadcast** e enlaces **ponto a ponto**.

Os enlaces ponto a ponto conectam pares de máquinas individuais. Para ir da origem ao destino em uma rede composta de enlaces ponto a ponto, mensagens curtas, chamadas **pacotes** em certos contextos, talvez tenham de visitar primeiro uma ou mais máquinas intermediárias. Como normalmente é possível haver várias rotas de diferentes tamanhos, encontrar boas rotas é algo importante em redes ponto a ponto. A transmissão ponto a ponto com exatamente um transmissor e exatamente um receptor às vezes é chamada de **unicasting**.

Ao contrário, as redes de broadcast têm apenas um canal de comunicação, compartilhado por todas as máquinas da rede; os pacotes enviados por qualquer máquina são recebidos por todas as outras. Um campo de endereço dentro do pacote especifica o destinatário pretendido. Quando re-

cebe um pacote, a máquina processa o campo de endereço. Se o pacote se destinar à máquina receptora, esta o processará; se for destinado a alguma outra máquina, o pacote será simplesmente ignorado.

Uma rede sem fio é um exemplo comum de um enlace de broadcast, com a comunicação compartilhada por uma região de cobertura que depende do canal sem fios e da máquina transmissora. Como uma analogia, imagine uma pessoa em uma sala de reunião, gritando: ‘Watson, venha cá. Preciso de você’. Embora o pacote possa ser recebido (ouvido) por muitas pessoas, apenas Watson responderá; os outros simplesmente o ignoram.

Os sistemas de broadcast normalmente também oferecem a possibilidade de endereçamento de um pacote a *todos* os destinos usando um código especial no campo de endereço. Quando um pacote com esse código é transmitido, ele é recebido e processado por cada máquina na rede; não é à toa que esse modo de operação é chamado **broadcasting**. Alguns sistemas de broadcasting também admitem a transmissão para um subconjunto de máquinas, o que se conhece como **multicasting**.

Um critério alternativo para classificar as redes é por escalabilidade. A distância é importante como métrica de classificação, pois diferentes tecnologias são usadas em diferentes escalas.

Na Figura 1.4, classificamos vários sistemas processadores por seu tamanho físico aproximado. Na parte superior encontram-se as redes pessoais, redes destinadas a uma única pessoa. Depois aparecem as redes de maior tamanho, que podem ser divididas em locais, metropolitanas e a longas distâncias, ambas em escala crescente. Finalmente, a conexão de duas ou mais redes é chamada rede interligada. A Internet mundial certamente é o mais conhecido (mas não o único) exemplo de uma rede interligada. Logo, teremos redes interligadas ainda maiores, com a **Internet interplanetária**, que conecta redes no espaço sideral (Burleigh et al., 2003).

Distância do interprocessador	Processadores localizados no mesmo	Exemplo
1 m	Metro quadrado	Área pessoal
10 m	Cômodo	Rede local
100 m	Prédio	
1 km	Campus	Rede metropolitana
10 km	Cidade	
100 km	País	Rede a longas distâncias
1.000 km	Continente	
10.000 km	Planeta	A Internet

Figura 1.4 | Classificação de processadores interconectados por escala.

Neste livro, tratamos das redes em todas essas escalas. Nas próximas seções, temos uma breve introdução ao hardware de rede por escala.

1.2.1 REDES PESSOAIS

As **redes pessoais**, ou **PANs (Personal Area Networks)**, permitem que dispositivos se comuniquem pelo alcance de uma pessoa. Um exemplo comum é uma rede sem fio que conecta um computador com seus periféricos. Quase todo computador tem monitor, teclado, mouse e impressora conectados. Sem usar tecnologia sem fio, essa conexão deve ser feita com cabos. Tantas pessoas têm dificuldade para encontrar os cabos corretos e encaixá-los nos conectores certos (embora normalmente tenham cores diferentes) que a maioria dos vendedores de computador oferece a opção de enviar um técnico à casa do usuário para fazê-lo. Para ajudar esses usuários, algumas empresas se reuniram para projetar uma rede sem fio de curta distância, chamada **Bluetooth**, para conectar esses componentes sem o uso de fios. A ideia é que, se seu dispositivo tem Bluetooth, então você não precisa de cabos. Você simplesmente os liga e eles funcionam juntos. Para muitas pessoas, essa facilidade de operação é uma grande vantagem.

Na forma mais simples, as redes Bluetooth usam um paradigma mestre-escravo da Figura 1.5. A unidade do sistema (o PC) normalmente é o mestre, falando com o mouse, o teclado etc. como escravos. O mestre diz aos escravos quais endereços usar, quando eles podem transmitir, por quanto tempo, quais frequências eles podem usar e assim por diante.

O Bluetooth também pode ser usado em outros ambientes. Ele normalmente é usado para conectar um fone de ouvido sem cabos, e permite que seu aparelho de música digital se conecte a seu carro simplesmente ao entrar no alcance. Um tipo completamente diferente de rede pessoal é formado quando um dispositivo médico embutido, como um marca-passos, bomba de insulina ou aparelho de audição fala com um controle remoto operado pelo usuário. Discutiremos o Bluetooth com mais detalhes no Capítulo 4.

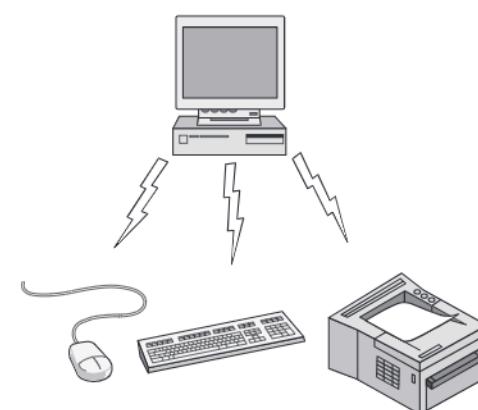


Figura 1.5 | Configuração de rede pessoal Bluetooth.

As redes pessoais também podem ser montadas com outras tecnologias que se comunicam por curtas distâncias, como RFID em smartcards e livros de biblioteca. Estudaremos a RFID no Capítulo 4.

1.2.2 REDES LOCAIS

A próxima etapa é a **rede local**, ou **LAN (Local Area Network)**. Uma LAN é uma rede particular que opera dentro e próximo de um único prédio, como uma residência, um escritório ou uma fábrica. As LANs são muito usadas para conectar computadores pessoais e aparelhos eletrônicos, para permitir que compartilhem recursos (como impressoras) e troquem informações. Quando as LANs são usadas pelas empresas, elas são chamadas **redes empresariais**.

As LANs sem fio são muito populares atualmente, especialmente nas residências, prédios de escritórios mais antigos e outros lugares onde a instalação de cabos é muito trabalhosa. Nesses sistemas, cada computador tem um rádio modem e uma antena, que ele usa para se comunicar com outros computadores. Quase sempre, cada computador fala com um dispositivo no teto, como mostra a Figura 1.6(a). Esse dispositivo, chamado **ponto de acesso (AP – Access Point)**, **roteador sem fio** ou **estação-base**, repassa os pacotes entre os computadores sem fio e também entre eles e a Internet. Ser o AP é como ser o garoto popular na escola, pois todos querem falar com você. Porém, se os outros computadores estiverem próximos o suficiente, eles podem se comunicar diretamente entre si em uma configuração peer-to-peer.

Existe um padrão para as LANs sem fios, chamado **IEEE 802.11**, popularmente conhecido como **WiFi**, que se tornou muito conhecido. Ele trabalha em velocidades de 11 a centenas de Mbps. (Neste livro, vamos aderir à tradição e medir as velocidades de linha em megabits/s, onde 1 Mbps é 1.000.000 bits/s, e gigabits/s, onde 1 Gbps é 1.000.000.000 bits/s.) Discutiremos o padrão 802.11 no Capítulo 4.

As LANs com fios utilizam uma série de tecnologias de transmissão diferentes. A maioria delas usa fios de co-

bre, mas algumas usam fibra óptica. As LANs são restritas em tamanho, o que significa que o tempo de transmissão, no pior caso, é limitado e conhecido com antecedência. Conhecer esses limites ajuda na tarefa de projetar protocolos de rede. Normalmente, as LANs com fios trabalham em velocidades de 100 Mbps a 1 Gbps, têm baixo atraso de transporte de dados (microsegundos ou nanosegundos) e com elas ocorrem muito poucos erros. As LANs mais recentes podem operar em até 10 Gbps. Em comparação com as redes sem fios, as LANs com fios as excedem em todas as dimensões de desempenho. É simplesmente mais fácil enviar sinais por um fio ou por uma fibra do que pelo ar.

A topologia de muitas LANs com fios é embutida a partir de enlaces ponto a ponto. O IEEE 802.3, popularmente chamado **Ethernet**, é de longe o tipo mais comum de LAN com fios. A Figura 1.6(b) mostra uma topologia de exemplo da **Ethernet comutada**. Cada computador troca informações usando o protocolo Ethernet e se conecta a um dispositivo de rede chamado **switch**, com um enlace ponto a ponto. Daí o nome. Um switch tem várias **portas**, cada qual podendo se conectar a um computador. A função do switch é repassar os pacotes entre os computadores que estão conectados a ela, usando o endereço em cada pacote para determinar para qual computador enviá-lo.

Para montar LANs maiores, os switches podem ser conectados uns aos outros usando suas portas. O que acontece se você os conectar em um loop? A rede ainda funcionará? Felizmente, os projetistas pensaram nesse caso. É função do protocolo descobrir que caminhos os pacotes devem atravessar para alcançar o computador pretendido com segurança. Veremos como isso funciona no Capítulo 4.

Também é possível dividir uma LAN física grande em duas LANs lógicas menores. Você pode estar se perguntando por que isso seria útil. Às vezes, o layout do equipamento de rede não corresponde à estrutura da organização. Por exemplo, os departamentos de engenharia e finanças de uma empresa poderiam ter computadores na mesma LAN física, pois estão na mesma ala do prédio, mas poderia ser mais fácil administrar o sistema se engenharia e finanças

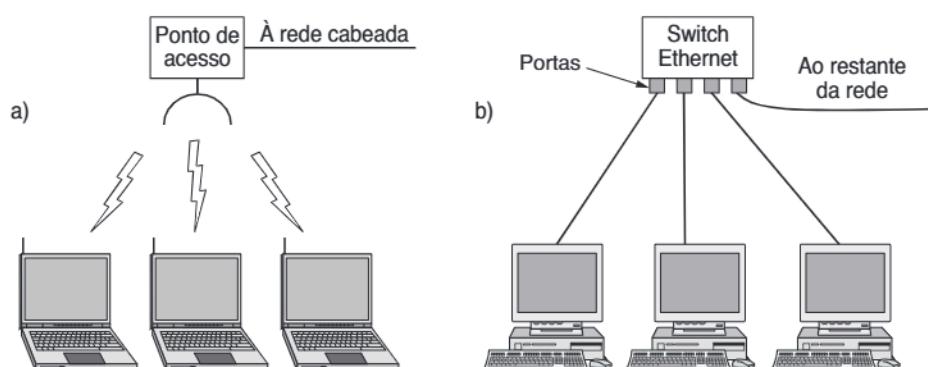


Figura 1.6 | LANs sem fios e cabeadas. (a) 802.11. (b) Ethernet comutada.

tivessem, cada um, sua própria LAN virtual, ou VLAN. Nesse projeto, cada porta é marcada com uma ‘cor’, digamos, verde para engenharia e vermelha para finanças. O switch então encaminha pacotes de modo que os computadores conectados às portas verdes sejam separados dos computadores conectados às portas vermelhas. Os pacotes de broadcast enviados em uma porta de cor vermelha, por exemplo, não serão recebidos em uma porta de cor verde, como se existissem duas LANs diferentes. Estudaremos as VLANs no final do Capítulo 4.

Também existem outras topologias de LAN com fios. Na verdade, a Ethernet comutada é uma versão moderna do projeto Ethernet original, que envia todos os pacotes por um único cabo. No máximo uma máquina poderia transmitir com sucesso de cada vez, e um mecanismo distribuído arbitrava o uso e resolia conflitos da rede compartilhada. Ele usava um algoritmo simples: os computadores poderiam transmitir sempre que o cabo estivesse ocioso. Se dois ou mais pacotes colidissem, cada computador simplesmente esperaria por um tempo aleatório e tentaria mais tarde. Chamaremos essa versão de **Ethernet clássica** para fazer a distinção e, como você já deve imaginar, aprenderá sobre ela no Capítulo 4.

As redes de broadcast, com e sem fios, ainda podem ser divididas em estáticas e dinâmicas, dependendo do modo como o canal é alocado. Em uma alocação estática típica, o tempo seria dividido em intervalos discretos e seria utilizado um algoritmo de rodízio, fazendo com que cada máquina transmitisse apenas no intervalo de que dispõe. A alocação estática desperdiça a capacidade do canal quando uma máquina não tem nada a transmitir durante o intervalo (slot) alocado a ela, e assim a maioria dos sistemas procura alocar o canal dinamicamente (ou seja, por demanda).

Os métodos de alocação dinâmica de um canal comum são centralizados ou descentralizados. No método centralizado de alocação de canal, existe apenas uma entidade, por exemplo, a estação-base nas redes celulares, que determina quem transmitirá em seguida. Para executar essa tarefa, a entidade aceita solicitações e as prioriza de acordo com algum algoritmo interno. No método descentralizado de alocação de canal, não existe nenhuma entidade central; cada máquina deve decidir por si mesma se a transmissão deve ser realizada. Você poderia pensar que isso sempre leva ao caos, mas isso não acontece. Mais tarde, estudaremos muitos algoritmos criados para impedir a instauração do caos potencial.

Vale a pena gastar um pouco mais de tempo discutindo as LANs domésticas. No futuro, é provável que todo dispositivo doméstico seja capaz de se comunicar com cada um dos outros dispositivos, e que todos eles estejam acessíveis pela Internet. Esse é um daqueles conceitos visionários que ninguém solicitou (como os controles remotos de TV ou os telefones celulares) mas, depois que chegaram, ninguém consegue mais imaginar como viver sem eles.

Muitos dispositivos são capazes de se conectar em rede. Entre eles estão computadores, dispositivos de entretenimento como TVs e DVDs, telefones e outros produtos eletrônicos, como câmeras, aparelhos como rádios-relógios e de infraestrutura, como medidores de energia e termostatos. Essa tendência só continuará. Por exemplo, um lar normalmente já tem dezenas de relógios (por exemplo, em aparelhos), todos eles podendo ser ajustados ao horário de verão automaticamente se os relógios estivessem na Internet. O monitoramento remoto da casa é um provável vencedor, pois muitos filhos já crescidos estariam dispostos a gastar algum dinheiro para ajudar seus pais idosos a viverem em segurança em suas próprias casas.

Embora pudéssemos pensar na rede doméstica como apenas outra LAN, ela provavelmente terá propriedades diferentes das outras redes. Primeiro, os dispositivos em rede precisam ser muito fáceis de instalar. Os roteadores sem fio são o item eletrônico de consumo mais devolvido. As pessoas compram um porque querem uma rede sem fio em casa, descobrem que ele não funciona ‘conforme sai da caixa’ e, depois, o devolvem em vez de escutar aquela musiquinha enquanto esperam na linha pela assistência técnica.

Segundo, a rede e os dispositivos têm de ser à prova de falhas quando em operação. Os condicionadores de ar costumavam ter um único botão com quatro ajustes: DESLIGAR, BAIXO, MÉDIO e ALTO. Agora eles têm manuais de 30 páginas. Uma vez que são ligados em rede, espera-se que apenas o capítulo sobre segurança ocupe 30 páginas. Isso é um problema porque somente usuários de computador estão acostumados a montar produtos que não funcionam; os consumidores de carros, televisores e refrigeradores são muito menos tolerantes. Eles esperam que os produtos funcionem 100 por cento sem precisar contratar um gênio da informática.

Em terceiro lugar, o preço baixo é algo essencial para o sucesso. As pessoas não pagarão US\$ 50 a mais por um termostato capaz de se conectar à Internet, porque poucas pessoas consideram importante monitorar a temperatura de sua casa enquanto estão no trabalho. Porém, por US\$ 5 a mais, esse acessório teria boa aceitação.

Quarto, deve ser possível começar com um ou dois dispositivos e expandir o alcance da rede gradualmente. Isso significa nenhuma guerra de padrões oferecidos. Dizer aos clientes para comprar periféricos com interfaces IEEE 1394 (FireWire) e alguns anos depois voltar atrás e dizer que USB 2.0 é a interface do mês e depois dizer que 802.11g — opa, não, é melhor 802.11n — quero dizer, 802.16 (diferentes redes sem fio) — deixará os consumidores muito nervosos. A interface de rede terá de permanecer estável por décadas, como os padrões de radiodifusão da televisão.

Quinto, a segurança e a confiabilidade serão muito importantes. Perder alguns arquivos para um vírus de e-mail é uma coisa; permitir que um assaltante desarme seu

sistema de segurança a partir de seu computador móvel e depois saqueie sua casa é algo muito diferente.

Uma questão interessante é saber se as redes domésticas estarão fisicamente conectadas ou se serão redes sem fios. Conveniência e custo favorecem as redes sem fio, pois não existem fios para encaixar, ou ainda, aperfeiçoar. A segurança favorece as redes fisicamente conectadas, pois as ondas de rádio que elas utilizam passam com facilidade pelas paredes. Nem todo mundo fica satisfeito com a ideia de ter os vizinhos pegando carona em sua conexão à Internet e lendo suas mensagens de correio eletrônico. No Capítulo 8, estudaremos como a criptografia pode ser utilizada para proporcionar segurança, mas, no contexto de uma rede doméstica, a segurança tem de ser infalível, mesmo com usuários inexperientes.

Uma terceira opção que pode ser atraente é reutilizar as redes que já estão na residência. O candidato óbvio são os fios elétricos instalados por toda a casa. As **redes de energia elétrica** permitem que os dispositivos conectados às tomadas transmitam informações por toda a casa. De qualquer forma, você já precisa conectar a TV, e dessa forma ela pode obter conectividade com a Internet ao mesmo tempo. A dificuldade é como transportar energia e sinais de dados ao mesmo tempo. Parte da resposta é que eles usam faixas de frequência diferentes.

Resumindo, as LANs domésticas oferecem muitas oportunidades e desafios. A maior parte dos desafios se relaciona à necessidade de que as redes sejam fáceis de administrar, confiáveis e seguras, especialmente nas mãos de usuários não técnicos, além do baixo custo.

1.2.3 REDES METROPOLITANAS

Uma **rede metropolitana**, ou **MAN (Metropolitan Area Network)**, abrange uma cidade. O exemplo mais conhecido de MANs é a rede de televisão a cabo disponí-

vel em muitas cidades. Esses sistemas cresceram a partir de antigos sistemas de antenas comunitárias usadas em áreas com fraca recepção do sinal de televisão pelo ar. Nesses primeiros sistemas, uma grande antena era colocada no alto de uma colina próxima e o sinal era, então, conduzido até as casas dos assinantes.

Em princípio, esses sistemas eram sistemas *ad hoc* projetados no local. Posteriormente, as empresas começaram a entrar no negócio, obtendo concessões dos governos municipais para conectar cidades inteiras por fios. A etapa seguinte foi a programação de televisão e até mesmo canais inteiros criados apenas para transmissão por cabos. Esses canais costumavam ser bastante especializados, oferecendo apenas notícias, apenas esportes, apenas culinária, apenas jardinagem e assim por diante. Entretanto, desde sua concepção até o final da década de 1990, eles se destinavam somente à recepção de televisão.

A partir do momento em que a Internet atraiu uma audiência de massa, as operadoras de redes de TV a cabo começaram a perceber que, com algumas mudanças no sistema, eles poderiam oferecer serviços da Internet full-duplex (mão dupla) em partes não utilizadas do espectro. Nesse momento, o sistema de TV a cabo começou a se transformar, passando de uma forma de distribuição de televisão para uma rede metropolitana. Em uma primeira aproximação, uma MAN seria semelhante ao sistema mostrado na Figura 1.7. Nessa figura, observamos que os sinais de televisão e de Internet são transmitidos à **central a cabo** centralizada para distribuição subsequente às casas das pessoas. Voltaremos a esse assunto, estudando-o em detalhes no Capítulo 2.

Porém, a televisão a cabo não é a única MAN. Os desenvolvimentos recentes para acesso à Internet de alta velocidade sem fio resultaram em outra MAN, que foi padronizada como IEEE 802.16 e é conhecida popularmente como **WiMAX**. Vamos examiná-la no Capítulo 4.

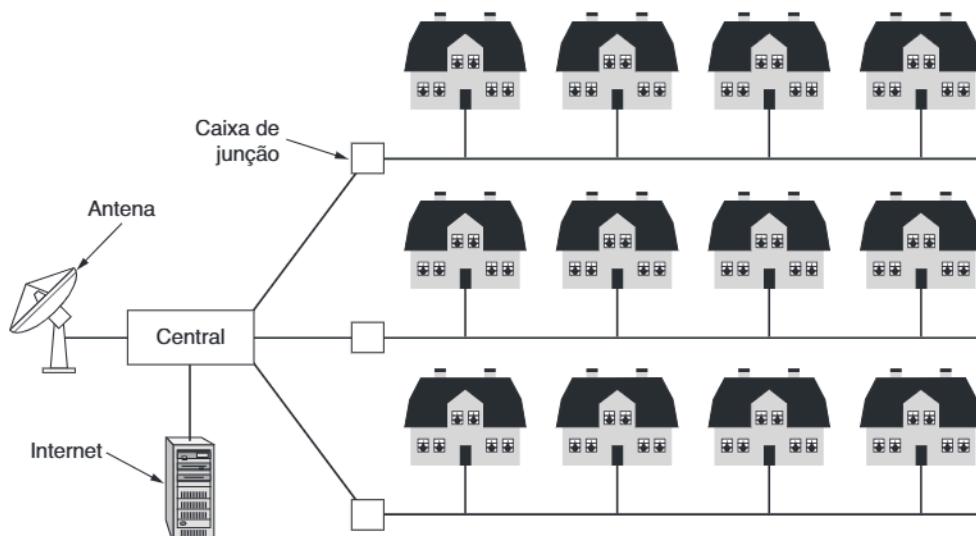


Figura 1.7 | Uma rede metropolitana baseada na TV a cabo.

1.2.4 REDES A LONGAS DISTÂNCIAS

Uma rede a longa distância, ou **WAN (Wide Area Network)**, abrange uma grande área geográfica, com frequência um país ou continente. Vamos começar nossa discussão com as WANs conectadas por fios, usando o exemplo de uma empresa com filiais em diferentes cidades.

A WAN na Figura 1.8 é uma rede que conecta escritórios em Perth, Melbourne e Brisbane. Cada um desses escritórios contém computadores que executam programas (ou seja, aplicações) do usuário. Seguiremos a tradição e chamaremos essas máquinas de **hosts**. O restante da rede que conecta esses hosts é chamada **sub-rede de comunicação** ou, simplificando, apenas **sub-rede**. A tarefa da sub-rede é transportar mensagens de um host para outro, exatamente como o sistema de telefonia transporta as palavras (na realidade, sons) do falante ao ouvinte.

Na maioria das WANs, a sub-rede consiste em dois componentes distintos: linhas de transmissão e elementos de comutação. As **linhas de transmissão** transportam bits entre as máquinas. Elas podem ser formadas por fios de cobre, fibra óptica, ou mesmo enlaces de radiodifusão. A maioria das empresas não tem linhas de transmissão disponíveis, então elas alugam as linhas de uma empresa de telecomunicações. Os **elementos de comutação**, ou apenas comutadores, são computadores especializados que conectam três ou mais linhas de transmissão. Quando os dados chegam a uma interface de entrada, o elemento de comutação deve escolher uma interface de saída para encaminhá-los. Esses computadores de comutação receberam diversos nomes no passado; o nome **roteador** é, agora, o mais comumente usado.

Vale a pena fazermos um breve comentário em relação ao termo ‘sub-rede’. Originalmente, seu **único** significado identificava o conjunto de roteadores e linhas de comunicação que transportava pacotes entre os hosts de origem e de destino. Porém, o termo adquiriu um segundo significado, em conjunto com o endereçamento da rede. Discutiremos esse significado no Capítulo 5 e ficaremos com o significado original (uma coleção de linhas de comunicação de dados e roteadores) até chegarmos lá.

A WAN, conforme a descrevemos, é semelhante a uma grande LAN cabeada, mas existem algumas diferenças importantes que vão além dos extensos cabos de interconexão. Normalmente, em uma WAN, os hosts e a sub-rede são proprietários e administrados por diferentes pessoas. Em nosso exemplo, os funcionários poderiam ser responsáveis por seus próprios computadores, enquanto o departamento de TI da empresa está encarregado do restante da rede. Veremos limites mais claros nos próximos exemplos, em que o provedor da rede ou a companhia telefônica opera a sub-rede. A separação dos aspectos de comunicação estritos da rede (a sub-rede) dos aspectos da aplicação (os hosts) simplifica bastante o projeto geral da rede.

Uma segunda diferença é que os roteadores normalmente conectarão diferentes tipos de tecnologia de rede. As redes dentro dos escritórios podem ser Ethernet comutada, por exemplo, enquanto as linhas de transmissão de longa distância podem ser enlaces SONET (que veremos no Capítulo 2). Algum dispositivo é necessário para juntá-las. O leitor atento notará que isso vai além da nossa definição de uma rede. Isso significa que muitas WANs de fato serão **redes interligadas**, ou redes compostas que são criadas a partir de mais de uma rede. Voltaremos a esse assunto sobre redes interligadas na próxima seção.

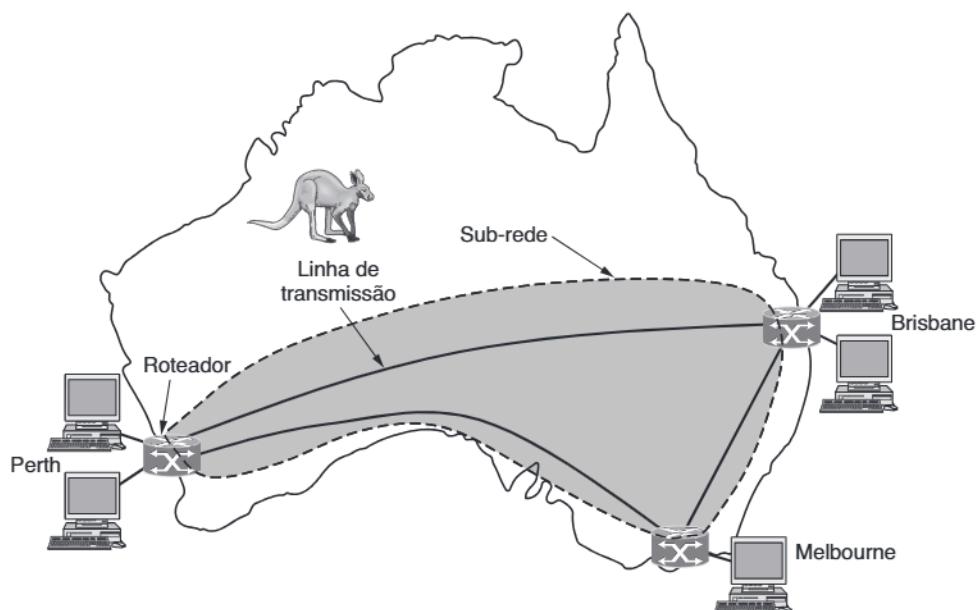


Figura 1.8 | WAN que conecta três escritórios de filiais na Austrália.

Uma última diferença é naquilo que é conectado à sub-rede. Podem ser computadores individuais, como foi o caso para a conexão às LANs, ou podem ser LANs inteiras. É assim que redes maiores são montadas a partir de redes menores. Em relação à sub-rede, ela tem a mesma função.

Agora estamos em uma posição de examinar duas outras variedades de WANs. Primeiro, em vez de alugar linhas de transmissão dedicadas, uma empresa pode conectar seus escritórios à Internet. Isso permite que as conexões sejam feitas entre os escritórios como enlaces virtuais que usam a capacidade de infraestrutura da Internet. Esse arranjo, mostrado na Figura 1.9, é chamado de **rede privada virtual**, ou **VPN** (**Virtual Private Network**). Em comparação com o arranjo dedicado, uma VPN tem a vantagem comum da virtualização, ou seja, ela oferece flexibilidade na reutilização de recurso (conectividade com a Internet). Para entender isso, considere como é fácil incluir um quarto escritório. Uma VPN também tem a desvantagem normal da virtualização, o que é uma falta de controle sobre os recursos subjacentes. Com uma linha dedicada, a capacidade é clara. Com uma VPN, suas milhas estão sujeitas à variação, de acordo com o serviço da Internet.

A segunda variação é que a sub-rede pode ser operada por uma empresa diferente. O operador da sub-rede é conhecido como um **provedor de serviço de rede**, e os escritórios são seus clientes. Essa estrutura aparece na Figura 1.10. O operador da sub-rede também se conectará a outros clientes, desde que eles possam pagar e ela possa oferecer o serviço. Como seria decepcionante um serviço de rede em que os clientes só pudessem enviar pacotes uns aos outros, o operador da sub-rede também se conectará a outras redes que fazem parte da Internet. Esse operador de sub-rede é chamado de **provedor de serviço de Internet**.

net, ou **ISP (Internet Service Provider)**, e a sub-rede é uma **rede ISP**. Seus clientes que se conectam à ISP recebem serviço de Internet.

Podemos usar a rede ISP para prever algumas questões fundamentais que estudaremos em outros capítulos. Na maioria das WANs, a rede contém muitas linhas de transmissão, cada uma conectando um par de roteadores. Se dois roteadores que não compartilham uma linha de transmissão quiserem se comunicar, eles precisam fazer isso indiretamente, por meio de outros roteadores. Pode haver muitos caminhos na rede que conectam esses dois roteadores. O processo em que o roteador toma a decisão sobre qual caminho usar e para onde enviar o pacote em seguida, para a interface adequada, é chamado de **algoritmo de roteamento**. Existem muitos desses algoritmos. Estudaremos alguns tipos em detalhes no Capítulo 5.

Outros tipos de WANs utilizam muito as tecnologias sem fio. Nos sistemas via satélite, cada computador no solo tem uma antena através da qual ele pode enviar e receber dados de e para um satélite em órbita. Todos os computadores podem escutar a saída do satélite, e em alguns casos eles também podem escutar as transmissões que sobem de seus computadores para o satélite da mesma forma. As redes de satélite são inherentemente de radiodifusão, e são mais úteis quando essa propriedade é importante.

A rede de telefonia celular é outro exemplo de uma WAN que usa tecnologia sem fio. Esse sistema já passou por três gerações, e uma quarta está a caminho. A primeira geração era analógica e usada apenas para voz. A segunda geração era digital e apenas para voz. A terceira geração é digital e se destina a voz e dados. Cada estação-base celular cobre uma distância muito maior do que uma LAN sem fio, com um alcance medido em quilômetros, ao invés de dezenas de

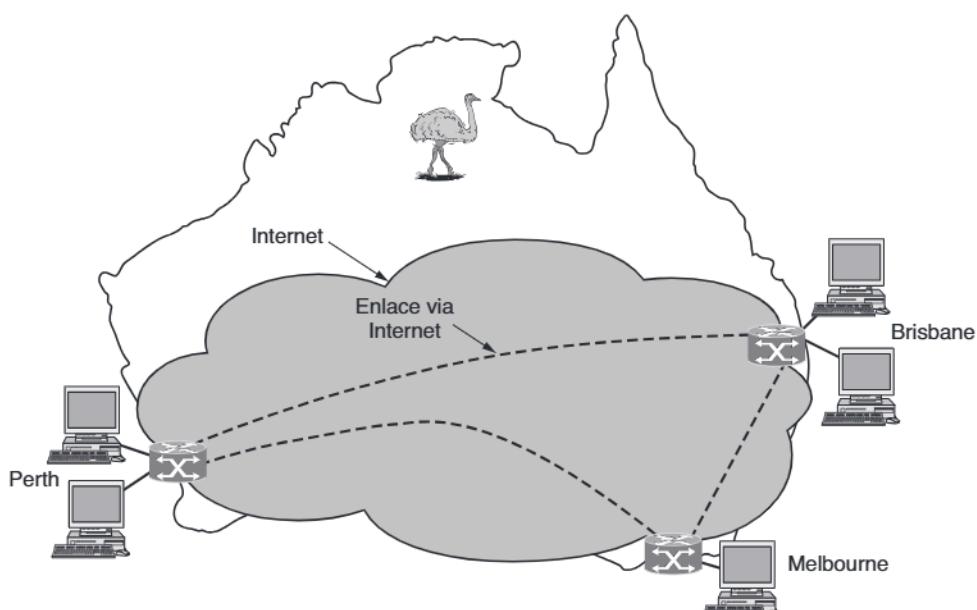


Figura 1.9 | WAN usando uma rede privada virtual.

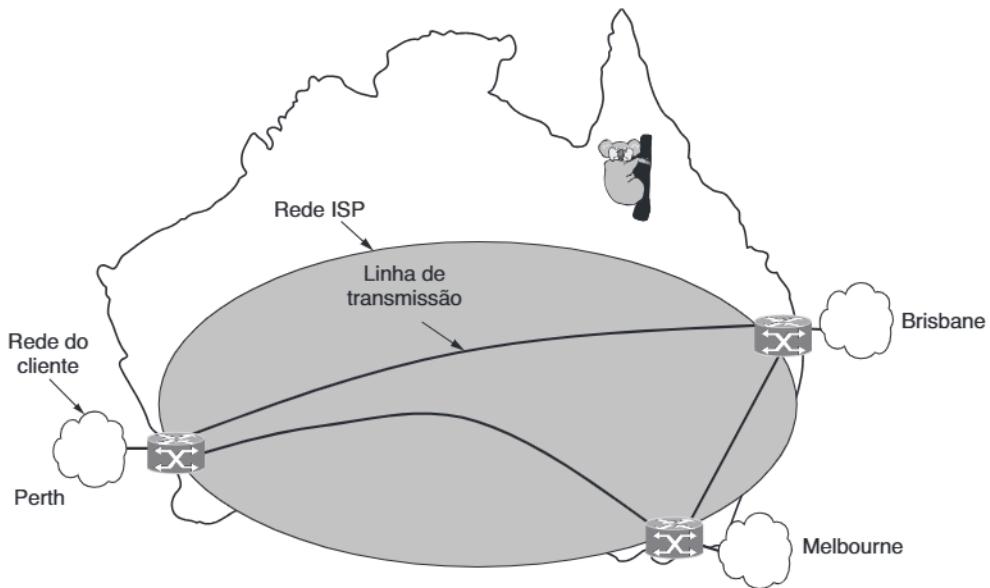


Figura 1.10 | WAN usando uma rede ISP.

metros. As estações-base são conectadas umas às outras por uma rede de backbone que normalmente é conectada por cabos. As taxas de dados das redes celulares normalmente estão na ordem de 1 Mbps, muito menos do que uma LAN sem fio, que pode chegar a uma ordem de 100 Mbps. Falaremos bastante sobre essas redes no Capítulo 2.

I 1.2.5 REDES INTERLIGADAS (INTERNETS)

Existem muitas redes no mundo, frequentemente apresentando diferentes tipos de hardware e software. Normalmente, as pessoas conectadas a redes distintas precisam se comunicar entre si. Para que esse desejo se torne uma realidade, é preciso que se estabeleçam conexões entre redes quase sempre incompatíveis. Um conjunto de redes interconectadas forma uma **rede interligada** ou **internet**. Esses termos serão usados em um sentido genérico, em contraste com a Internet mundial (uma rede interligada em nível mundial), que sempre será representada com inicial maiúscula. A Internet usa redes ISP para conectar redes empresariais, domésticas e muitas outras redes. Veremos a Internet com muito mais detalhes em outro ponto deste livro.

Em geral, sub-redes, redes e redes interligadas se confundem. Uma sub-rede faz mais sentido no contexto de uma rede a longa distância, em que ela se refere ao conjunto de roteadores e linhas de comunicação pertencentes à operadora da rede. Como analogia, o sistema telefônico consiste em estações de comutação telefônica conectadas entre si por linhas de alta velocidade, e com casas e escritórios por linhas de baixa velocidade. Essas linhas e equipamentos, cuja propriedade e gerenciamento são da empresa de telefonia, formam a sub-rede do sistema telefônico. Os telefones propriamente ditos (os hosts, nessa analogia) não fazem parte da sub-rede.

Uma rede interligada é formada pela combinação de uma sub-rede e seus hosts. Entretanto, a palavra ‘rede’ é normalmente usada também em um sentido mais livre. Uma sub-rede poderia ser descrita como uma rede, como no caso da ‘rede ISP’ da Figura 1.10. Uma rede interligada também pode ser descrita como uma rede, como no caso da WAN na Figura 1.8. Seguiremos uma prática semelhante e, se estivermos distinguindo uma rede de outros arranjos, ficaremos com nossa definição original de uma coleção de computadores interconectados por uma única tecnologia.

Falaremos mais sobre o que constitui uma rede interligada. Sabemos que ela é formada quando redes distintas são interconectadas. Em nossa visão, a conexão entre uma LAN e uma WAN ou a conexão de duas LANs é o modo normal de formar uma rede interligada, mas existe pouco acordo no setor sobre a terminologia nessa área. Existem duas regras práticas que são úteis. Primeiro, se diferentes organizações pagam pela construção de partes distintas da rede e cada uma mantém sua parte, temos uma rede interligada, e não uma única rede. Segundo, se a tecnologia subjacente é diferente em partes distintas (por exemplo, broadcast *versus* ponto a ponto e cabeada *versus* sem fio), provavelmente temos uma rede interligada.

Indo mais a fundo, precisamos falar sobre como duas redes diferentes podem ser conectadas. O nome geral para uma máquina que faz uma conexão entre duas ou mais redes e oferece a conversão necessária, tanto em termos de hardware quanto de software, é um **gateway**. Os gateways são distinguidos pela camada em que operam na hierarquia de protocolos. Falaremos mais sobre camadas e hierarquias de protocolos a partir da próxima seção, mas, por enquanto, imagine que as camadas mais altas são mais ligadas às

aplicações, como a Web, e as camadas mais baixas são mais ligadas a enlaces de transmissão, como a Ethernet.

Como o benefício de formar uma rede interligada é conectar computadores pelas redes, não queremos usar um gateway em muito baixo nível, ou então não poderemos fazer conexões entre diferentes tipos de redes. Também não queremos usar um gateway em um nível muito alto, ou então a conexão só funcionará para determinadas aplicações. O nível do meio, que é o mais apropriado, normalmente é chamado de camada de rede, e um roteador é um gateway que comuta pacotes nessa camada. Agora podemos localizar uma rede interligada descobrindo uma rede que tem roteadores.

1.3 | SOFTWARE DE REDE

No projeto das primeiras redes de computadores, o hardware foi a principal preocupação e o software ficou em segundo plano. Essa estratégia foi deixada para trás. Atualmente, o software de rede é altamente estruturado. Nas próximas seções, examinaremos com alguns detalhes a técnica de estruturação do software. O método descrito aqui é de fundamental importância para o livro inteiro e faremos repetidas referências a ele.

1.3.1 | HIERARQUIAS DE PROTOCOLOS

Para reduzir a complexidade de seu projeto, a maioria das redes é organizada como uma pilha de **camadas** (ou **níveis**), colocadas umas sobre as outras. O número, o nome, o conteúdo e a função de cada camada diferem de uma rede para outra. No entanto, em todas as redes o objetivo de cada camada é oferecer determinados serviços às camadas superiores, isolando essas camadas dos detalhes de implementação real desses recursos. Em certo sentido, cada camada é uma espécie de máquina virtual, oferecendo determinados serviços à camada situada acima dela.

Na realidade, esse conceito é familiar e utilizado em toda a ciência da computação, na qual é conhecido por nomes diferentes, como ocultação de informações, tipos de dados abstratos, encapsulamento de dados e programação orientada a objetos. A ideia fundamental é que um determinado item de software (ou hardware) forneça um serviço a seus usuários, mas mantenha ocultos os detalhes de seu estado interno e de seus algoritmos.

Quando a camada *n* de uma máquina se comunica com a camada *n* de outra máquina, coletivamente, as regras e convenções usadas nesse diálogo são conhecidas como o protocolo da camada *n*. Basicamente, um **protocolo** é um acordo entre as partes que se comunicam, estabelecendo como se dará a comunicação. Como analogia, quando uma mulher é apresentada a um homem, ela pode estender a mão para ele que, por sua vez, pode apertá-la ou beijá-la, dependendo, por exemplo, do fato de ela ser uma advogada norte-americana que esteja participando de uma reunião de negócios ou uma princesa europeia presente em um baile de gala. A violação do protocolo dificultará a comunicação, se não torná-la completamente impossível.

A Figura 1.11 ilustra uma rede de cinco camadas. As entidades que ocupam as camadas correspondentes em diferentes máquinas são chamadas pares (ou **peers**). Os pares podem ser processos de software, dispositivos de hardware, ou mesmo seres humanos. Em outras palavras, são os pares que se comunicam utilizando o protocolo.

Na realidade, os dados não são transferidos diretamente da camada *n* de uma máquina para a camada *n* em outra máquina. Em vez disso, cada camada transfere os dados e as informações de controle para a camada imediatamente abaixo dela, até a camada mais baixa ser alcançada. Abaixo da camada 1 encontra-se o **meio físico** por meio do qual se dá a comunicação propriamente dita. Na Figura 1.11, a comunicação virtual é mostrada por linhas pontilhadas e a comunicação física, por linhas contínuas.

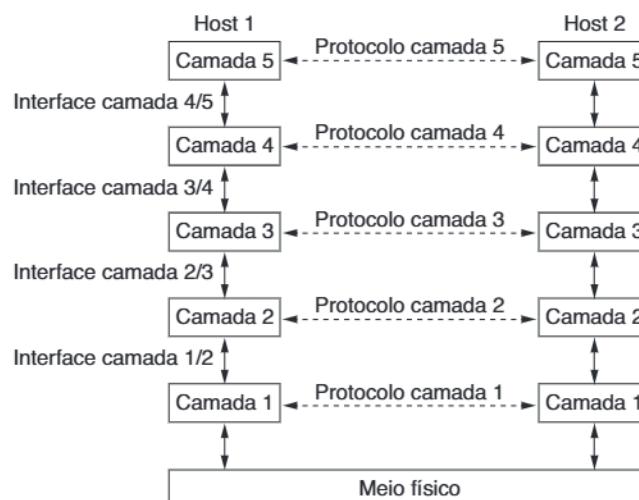


Figura 1.11 | Camadas, protocolos e interfaces.

Entre cada par de camadas adjacentes existe uma **interface**. Esta define as operações e os serviços que a camada inferior tem a oferecer à camada que se encontra acima dela. Quando os projetistas de rede decidem a quantidade de camadas que será incluída em uma rede e o que cada uma delas deve fazer, uma das considerações mais importantes é a definição de interfaces claras entre as camadas. Por sua vez, isso exige que cada camada execute um conjunto específico de funções bem definidas. Além de reduzir o volume de informações que deve ser passado de uma camada para outra, as interfaces bem definidas também simplificam a substituição de uma camada por um protocolo ou implementação completamente diferente (por exemplo, a substituição de todas as linhas telefônicas por canais de satélite), pois o novo protocolo ou a nova implementação só precisa oferecer exatamente o mesmo conjunto de serviços à sua vizinha de cima, assim como era feito na implementação anterior. De fato, é comum que hosts diferentes utilizem implementações distintas do mesmo protocolo (normalmente, escrito por empresas diferentes). De fato, o próprio protocolo pode mudar em alguma camada sem que as camadas acima e abaixo dela sequer percebam.

Um conjunto de camadas e protocolos é chamado **arquitetura de rede**. A especificação de uma arquitetura deve conter informações suficientes para permitir que um implementador desenvolva o programa ou construa o hard-

ware de cada camada de forma que ela obedeça corretamente ao protocolo adequado. Nem os detalhes da implementação nem a especificação das interfaces pertencem à arquitetura, pois tudo fica oculto dentro das máquinas e não é visível do exterior. Nem sequer é necessário que as interfaces de todas as máquinas de uma rede sejam iguais, desde que cada uma delas possa usar todos os protocolos da maneira correta. Uma lista de protocolos usados por um determinado sistema, um protocolo por camada, é chamada **pilha de protocolos**. Arquiteturas de rede, pilhas de protocolos e os próprios protocolos são os principais assuntos deste livro.

Uma analogia pode ajudar a explicar a ideia de uma comunicação em várias camadas. Imagine dois filósofos (processos pares na camada 3), um dos quais fala urdu e português e o outro fala chinês e francês. Como não falam um idioma comum, eles contratam tradutores (processos pares na camada 2), que por sua vez têm cada qual uma secretária (processos pares na camada 1). O filósofo 1 deseja transmitir sua predileção por *oryctolagus cuniculus* a seu par. Para tal, ele envia uma mensagem (em português) através da interface 2/3 a seu tradutor, na qual diz 'Gosto de coelhos', como mostra a Figura 1.12. Como os tradutores resolveram usar um idioma neutro, o holandês, a mensagem foi convertida para 'Ik vind konijnen leuk'. A escolha do idioma é o protocolo da camada 2, que deve ser processada pelos pares dessa camada.

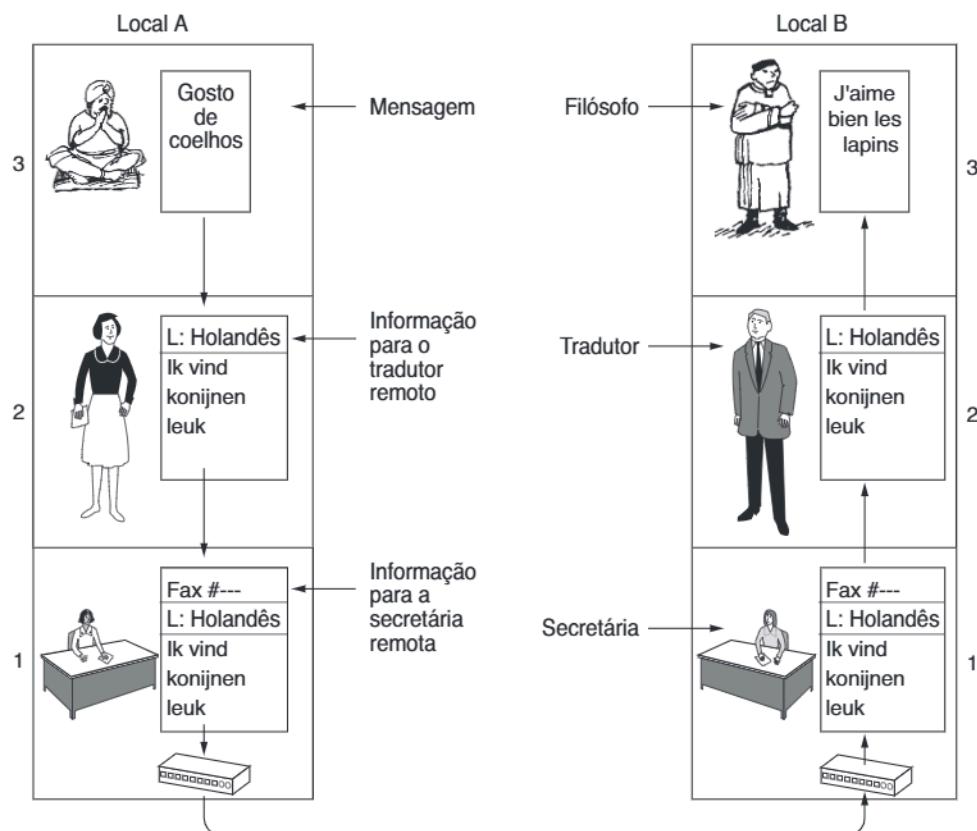


Figura 1.12 | A arquitetura filósofo-tradutor-secretária.

O tradutor entrega a mensagem a uma secretária para ser transmitida, por exemplo, por fax (o protocolo da camada 1). Quando chega, a mensagem é traduzida para o francês e passada através da interface 2/3 para o filósofo 2. Observe que cada protocolo é totalmente independente dos demais, desde que as interfaces não sejam alteradas. Nada impede que os tradutores mudem do holandês para o finlandês, desde que ambos concordem com a modificação e que ela não afete sua interface com a camada 1 ou com a camada 3. De modo semelhante, as secretárias podem passar de fax para correio eletrônico ou telefone sem incomodar (ou mesmo informar) as outras camadas. Cada processo só pode acrescentar informações dirigidas a seu par. Essas informações não são enviadas à camada superior.

Vejamos agora um exemplo mais técnico: como oferecer comunicação à camada superior da rede de cinco camadas na Figura 1.13. Uma mensagem, M , é produzida por um processo de aplicação que funciona na camada 5 e é entregue à camada 4 para transmissão. A camada 4 coloca um **cabeçalho** no início da mensagem para identificá-la e envia o resultado à camada 3. O cabeçalho inclui informações de controle, como endereços, a fim de permitir que a camada 4 da máquina de destino entregue a mensagem. Outros exemplos de informação de controle usados em algumas camadas são números de sequência (caso a camada inferior não preserve a ordem da mensagem), tamanhos e tempos.

Em muitas redes, não há limite para o tamanho das mensagens transmitidas no protocolo da camada 4, mas quase sempre há um limite imposto pelo protocolo da camada 3. Consequentemente, a camada 3 deve dividir as mensagens recebidas em unidades menores, pacotes, anexando um cabeçalho da camada 3 a cada pacote. Nesse

exemplo, M é dividido em duas partes, M_1 e M_2 , que serão transmitidas separadamente.

A camada 3 decide as linhas de saída que serão usadas e transmite os pacotes à camada 2. Esta acrescenta não apenas um cabeçalho a cada fragmento, mas também um final, e fornece a unidade resultante à camada 1 para transmissão física. Na máquina receptora, a mensagem se move de baixo para cima, de camada a camada, com os cabeçalhos sendo retirados durante o processo. Nenhum dos cabeçalhos das camadas abaixo de n é repassado à camada n .

Para entender a Figura 1.13, é importante observar a relação entre a comunicação virtual e a comunicação real, e a diferença entre protocolos e interfaces. Por exemplo, para os processos pares na camada 4, sua comunicação é ‘horizontal’, utilizando o protocolo da camada 4. O procedimento de cada um deles tem um nome semelhante a *EnviarParaOutroLado* e *ReceberDoOutroLado*, muito embora esses procedimentos na realidade se comuniquem com camadas inferiores através da interface 3/4, e não com o outro lado.

A abstração de processos pares (peers) é fundamental para toda a estrutura da rede. Com sua utilização, a tarefa não gerenciável de projetar a rede completa pode ser dividida em diversos problemas de projeto menores e gerenciáveis, ou seja, o projeto das camadas individuais.

Embora o título da Seção 1.3 seja ‘Software de rede’, vale a pena lembrar que as camadas inferiores de uma hierarquia de protocolos costumam ser implementadas no hardware ou como firmware. Apesar disso, algoritmos de protocolo muito complexos estão envolvidos no processo, mesmo se estiverem embutidos (parcial ou totalmente) no hardware.

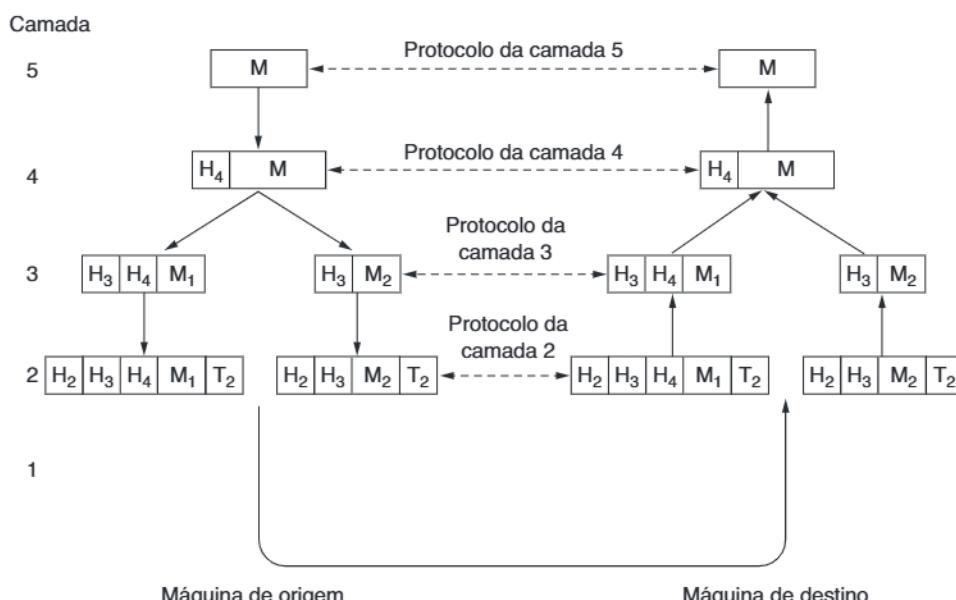


Figura 1.13 | Exemplo de fluxo de informações que admite a comunicação virtual na camada 5.

1.3.2 QUESTÕES DE PROJETO RELACIONADAS ÀS CAMADAS

Algumas questões fundamentais de projeto que ocorrem em redes de computadores estarão presentes em diversas camadas. Mencionaremos a seguir algumas das questões mais importantes.

Confiabilidade é a questão de projeto de criar uma rede que opere corretamente, embora sendo composta de uma coleção de componentes que não são confiáveis. Pense nos bits de um pacote trafegando pela rede. Há uma chance de que alguns desses bits sejam recebidos com problemas (invertidos) em virtude de um ruído elétrico casual, sinais sem fio aleatórios, falhas de hardware, bugs de software e assim por diante. Como é possível encontrar e consertar esses erros?

Um mecanismo para localizar erros na informação recebida usa códigos para **deteção de erros**. As informações recebidas incorretamente podem, então, ser retransmitidas até que sejam recebidas corretamente. Códigos mais poderosos permitem a **correção de erro**, em que a mensagem correta é recuperada a partir de bits possivelmente incorretos, que foram recebidos originalmente. Esses dois mecanismos funcionam acrescentando informações redundantes. Eles são usados nas camadas baixas, para proteger os pacotes enviados por enlaces individuais, e nas camadas altas, para verificar se o conteúdo correto foi recebido.

Outra questão de confiabilidade é descobrir um caminho que funcione através de uma rede. Normalmente, existem vários caminhos entre origem e destino e, em uma rede grande, pode haver alguns enlaces ou roteadores com defeito. Suponha que a rede esteja parada na Alemanha. Os pacotes enviados de Londres a Roma pela Alemanha não passarão, mas poderíamos enviar pacotes de Londres para Roma via Paris. A rede deve tomar essa decisão automaticamente. Esse tópico é chamado **roteamento**.

Uma segunda questão de projeto se refere à evolução da rede. Com o tempo, as redes se tornam maiores e novos projetos aparecem precisando ser conectados à rede existente. Recentemente, vimos o mecanismo-chave de estrutura usado para dar suporte à mudança, dividindo o problema geral e ocultando detalhes da implementação: as **camadas de protocolos**. Mas existem muitas outras estratégias.

Como existem muitos computadores na rede, cada camada precisa de um mecanismo para identificar transmissores e receptores que estão envolvidos em uma determinada mensagem. Esse mecanismo é conhecido como **endereçamento ou nomeação**, nas camadas alta e baixa, respectivamente.

Um aspecto do crescimento é que diferentes tecnologias de rede normalmente têm diferentes limitações. Por exemplo, nem todos os canais de comunicação preservam a ordem das mensagens enviadas neles, ocasionando soluções que numeram mensagens. Outro exemplo são as

diferenças no tamanho máximo de uma mensagem que as redes podem transmitir. Isso ocasiona mecanismos para dividir, transmitir e depois juntar novamente as mensagens. Esse tópico geral é chamado de **interligação de redes**.

Quando as redes ficam muito grandes, novos problemas aparecem. As cidades podem ter engarrafamentos no trânsito, falta de números de telefone, e é fácil se perder pelas ruas. Muitas pessoas não têm esses problemas em sua própria vizinhança, mas a cidade inteira pode ser um grande problema. Os projetos que continuam a crescer bem enquanto a rede cresce são considerados **escaláveis**.

Uma terceira questão de projeto é a alocação de recursos. As redes oferecem um serviço aos hosts a partir de seus recursos subjacentes, como a capacidade de linhas de transmissão. Para fazer isso bem, elas precisam de mecanismos que dividem seus recursos de modo que um host não interfira muito em outro.

Muitos projetos compartilham a largura de banda da rede dinamicamente, de acordo com a necessidade dos hosts a curto prazo, ao invés de dar a cada host uma fração fixa da largura de banda, que ele pode ou não usar. Esse projeto é chamado **multiplexação estatística**, significando compartilhar com base nas estatísticas de demanda. Ele pode ser aplicado às camadas inferiores para um único enlace, nas camadas altas para uma rede ou mesmo em aplicações que usam a rede.

Uma questão de alocação que afeta cada nível é como impedir que um transmissor rápido envie uma quantidade excessiva de dados a um receptor mais lento. Normalmente, usa-se uma espécie de *feedback* do receptor para o transmissor. Esse tópico é chamado **controle de fluxo**. Às vezes, o problema é que a rede é sobrecarregada porque muitos computadores querem enviar muito tráfego e a rede não pode entregar tudo isso. A sobrecarga da rede é chamada **congestionamento**. Uma estratégia é que cada computador reduza sua demanda quando experimentar um congestionamento. Isso também pode ser usado em todas as camadas.

É interessante observar que a rede tem mais recursos a oferecer do que simplesmente largura de banda. Para usos como o transporte de vídeo ao vivo, a prontidão na entrega importa muito. A maioria das redes precisa oferecer serviço às aplicações que desejam essa entrega em **tempo real** ao mesmo tempo que oferece serviço a aplicações que desejam uma alta vazão. A **qualidade de serviço** é o nome dado aos mecanismos que reconciliam essas demandas concorrentes.

A última questão de projeto trata de proteger a rede, defendendo-a contra diferentes tipos de ameaças. Uma das ameaças que mencionamos anteriormente é a de bisbilhagem nas comunicações. Mecanismos que oferecem **confidencialidade** defendem contra essa ameaça e são usados em várias camadas. Os mecanismos para **autenticação** impedem que alguém finja ser outra pessoa. Eles poderiam

ser usados para diferenciar websites falsos de um banco real, ou para permitir verificar se uma chamada da rede celular está realmente vindo de seu telefone, para que você pague a conta correta. Outros mecanismos para **integridade** impedem mudanças clandestinas nas mensagens, como alterar ‘debito US\$ 10 da minha conta’ para ‘debito US\$ 1.000 da minha conta’. Todos esses projetos são baseados em criptografia, que estudaremos no Capítulo 8.

I 1.3.3 SERVIÇOS ORIENTADOS E NÃO ORIENTADOS A CONEXÕES

As camadas podem oferecer dois tipos diferentes de serviços às camadas situadas acima delas: serviços orientados a conexões e serviços não orientados a conexões. Nesta seção, examinaremos esses dois tipos de serviços e as diferenças entre eles.

O serviço **orientado a conexões** se baseia no sistema telefônico. Para falar com alguém, você tira o fone do gancho, digita o número, fala e, em seguida, desliga. Da mesma forma, para utilizar um serviço de rede orientado a conexões, primeiro o usuário do serviço estabelece uma conexão, a utiliza, e depois a libera. O aspecto essencial de uma conexão é que ela funciona como um tubo: o transmissor empurra objetos (bits) em uma extremidade, e esses objetos são recebidos pelo receptor na outra extremidade. Na maioria dos casos, a ordem é preservada, de forma que os bits chegam na sequência em que foram enviados.

Em alguns casos, quando uma conexão é estabelecida, o transmissor, o receptor e a sub-rede conduzem uma **negociação** sobre os parâmetros a serem usados, como o tamanho máximo das mensagens, a qualidade do serviço exigida e outras questões. Em geral, um lado faz uma proposta e a outra parte pode aceitá-la, rejeitá-la ou fazer uma contraproposta. Um **círculo** é outro nome para uma conexão com recursos associados, como uma largura de banda fixa. Isso vem desde a rede telefônica, em que um circuito era um caminho pelo fio de cobre que transportava uma conversa telefônica.

Ao contrário do serviço orientado a conexões, o serviço **não orientado a conexões** se baseia no sistema postal. Cada mensagem (carta) carrega o endereço de destino completo e cada uma delas é roteada pelos nós intermediários através do sistema, independentemente de todas as outras. Existem diferentes nomes para mensagens em diferentes contextos; um **pacote** é uma mensagem na camada de rede. Quando os nós intermediários recebem uma mensagem completa antes de enviá-la para o próximo nó, isso é chamado **comutação store-and-forward**. A alternativa, em que a transmissão de uma mensagem em um nó começa antes de ser completamente recebida por ele, é chamada **comutação cut-through**. Normalmente, quando duas mensagens são enviadas ao mesmo destino, a primeira a ser enviada é a primeira a chegar. No entanto, é possível

que a primeira mensagem a ser enviada esteja atrasada, de modo que a segunda chegue primeiro.

Cada tipo de serviço pode ser caracterizado por sua confiabilidade. Alguns serviços são confiáveis, no sentido de nunca perderem dados. Em geral, um serviço confiável é implementado para que o receptor confirme o recebimento de cada mensagem, de modo que o transmissor se certifique de que ela chegou. O processo de confirmação introduz overhead e atrasos, que normalmente compensam, mas às vezes são indesejáveis.

Uma situação típica em que um serviço orientado a conexões confiável é apropriado é a transferência de arquivos. O proprietário do arquivo deseja se certificar de que todos os bits chegaram corretamente e na mesma ordem em que foram enviados. São poucos os clientes de transferência de arquivos que preferem um serviço que ocasionalmente desorganiza ou perde alguns bits, mesmo que ele seja muito mais rápido.

O serviço orientado a conexões confiável tem duas variações secundárias: sequências de mensagens e fluxos de bytes. Na primeira variação, os limites das mensagens são preservados. Quando duas mensagens de 1.024 bytes são enviadas, elas chegam como duas mensagens distintas de 1.024 bytes, nunca como uma única mensagem de 2.048 bytes. Na segunda, a conexão é simplesmente um fluxo de bytes, sem limites de mensagem. Quando 2.048 bytes chegam ao receptor, não há como saber se eles foram enviados como uma mensagem de 2.048 bytes, duas mensagens de 1.024 bytes ou 2.048 mensagens de 1 byte. Se as páginas de um livro são enviadas por uma rede a uma fotocompositora como mensagens separadas, talvez seja importante preservar os limites da mensagem. Por outro lado, para baixar um filme de DVD, um fluxo de bytes do servidor para o computador do usuário é tudo o que é necessário. Os limites de mensagens dentro do filme não são relevantes.

Para algumas aplicações, os atrasos introduzidos pelas confirmações são inaceitáveis. Uma dessas aplicações é o tráfego de voz digital por **Voice over IP (VoIP)**. Os usuários de telefone preferem ouvir um pouco de ruído na linha ou uma palavra truncada de vez em quando a experimentar um atraso para aguardar confirmações. O mesmo acontece durante a transmissão de uma conferência de vídeo; não haverá problema se aparecerem alguns pixels errados. No entanto, é irritante ver uma imagem parada enquanto o fluxo é interrompido e reiniciado para a correção de erros.

Nem todas as aplicações precisam do serviço orientado a conexões. Por exemplo, spammers enviam lixo eletrônico do serviço orientado a muitos destinatários. Provavelmente, o spammer não deseja enfrentar o problema de configurar e depois desfazer uma conexão apenas para enviar um item. Além disso, não será essencial uma entrega 100 por cento confiável, em especial se o custo for maior. É necessário apenas um modo de enviar uma única mensagem que tenha uma alta probabilidade de chegar, mas sem garantias. O ser-

viço não orientado a conexões não confiável (ou seja, sem confirmação) costuma ser chamado serviço de **datagramas**, em uma analogia com o serviço de telegramas, que também não oferece uma confirmação ao transmissor. Apesar de não ser confiável, essa é a forma dominante na maioria das redes, por motivos que adiante se tornarão mais claros.

Em outras situações, a conveniência de não ter de estabelecer uma conexão para enviar uma única mensagem curta é deseável, mas a confiabilidade é essencial. O serviço de **datagramas confirmados** pode ser oferecido para essas aplicações. Ele é semelhante a enviar uma carta registrada e solicitar um aviso de recebimento. Quando o aviso é devolvido, o transmissor fica absolutamente certo de que a carta foi entregue ao destinatário e não perdida ao longo do caminho. As mensagens de texto em telefones móveis são um exemplo.

Outro serviço é o de **solicitação/resposta**. Nele, o transmissor envia um único datagrama contendo uma solicitação; a resposta contém a réplica. A solicitação/resposta é em geral usada para implementar a comunicação no modelo cliente-servidor: o cliente emite uma solicitação e o servidor responde. Por exemplo, um cliente de telefone móvel poderia enviar uma consulta a um servidor de mapa para receber os dados de mapa para seu local atual. A Figura 1.14 resume os tipos de serviços descritos anteriormente.

O conceito de usar comunicação não confiável pode ser confuso a princípio. Afinal de contas, por que alguém iria preferir uma comunicação não confiável à comunicação confiável? Em primeiro lugar, a comunicação confiável (em nosso sentido, isto é, confirmada) pode não estar disponível em uma determinada camada. Por exemplo, a Ethernet não fornece comunicação confiável. Ocasionalmente, os pacotes podem ser danificados em trânsito. Cabe aos níveis mais altos do protocolo lidar com esse problema. Em particular, muitos serviços confiáveis são montados em cima de um serviço de datagrama não confiável. Em segundo lugar, os atrasos inerentes ao fornecimento de um serviço confiável podem ser inaceitáveis, em especial nas aplicações em tempo real, como as de multimídia. Por essas razões, coexistem tanto a comunicação confiável quanto a não confiável.

1.3.4 PRIMITIVAS DE SERVIÇO

Um serviço é especificado formalmente por um conjunto de **primitivas** (operações) disponíveis para que os processos do usuário acessem o serviço. Essas primitivas informam ao serviço que ele deve executar alguma ação ou relatar uma ação executada por uma entidade par. Se a pilha de protocolos estiver localizada no sistema operacional, como ocorre com frequência, as primitivas serão normalmente chamadas do sistema. Essas chamadas geram uma armadilha para o kernel, que então devolve o controle da máquina ao sistema operacional para enviar os pacotes necessários.

O conjunto de primitivas disponíveis depende da natureza do serviço que está sendo fornecido. As primitivas para um serviço orientado a conexões são diferentes das oferecidas em um serviço não orientado a conexões. Como um exemplo mínimo das primitivas de serviço que poderiam ser fornecidas para implementar um fluxo de bytes confiável, considere as primitivas listadas na Tabela 1.3. Elas serão familiares para os fãs do soquete de Berkeley, pois as primitivas são uma versão simplificada dessa interface.

Essas primitivas podem ser usadas para uma interação de solicitação/resposta em um ambiente cliente/servidor. Para ilustrar como, esboçamos um protocolo simples que implementa o serviço usando datagramas confirmados.

Primeiro, o servidor executa `LISTEN` para indicar que está preparado para aceitar conexões de entrada. Um caminho comum para implementar `LISTEN` é torná-la uma chamada de bloqueio do sistema. Depois de executar a primitiva, o processo servidor fica bloqueado até que surja uma solicitação de conexão.

Em seguida, o processo cliente executa `CONNECT` para estabelecer uma conexão com o servidor. A chamada `CONNECT` precisa especificar a quem se conectar; assim, ela poderia ter um parâmetro fornecendo o endereço do servidor. Então, em geral, o sistema operacional envia um pacote ao par solicitando que ele se conecte, como mostra o item (1) na Figura 1.15. O processo cliente é suspenso até haver uma resposta.

Serviço	Exemplo
Fluxo de mensagens confiável	Sequência de páginas
Fluxo de bytes confiável	Download de filme
Conexão não confiável	VoIP
Datagrama não confiável	Lixo de correio eletrônico
Datagrama confirmado	Mensagem de texto
Solicitação/resposta	Consulta a banco de dados

Figura 1.14 | Seis diferentes tipos de serviço.

Primitiva	Significado
LISTEN	Bloco que espera por uma conexão de entrada
CONNECT	Estabelecer uma conexão com um par que está à espera
ACCEPT	Aceitar uma conexão de entrada de um par
RECEIVE	Bloco que espera por uma mensagem de entrada
SEND	Enviar uma mensagem ao par
DISCONNECT	Encerrar uma conexão

Tabela 1.3 | Seis primitivas de serviço para implementação de uma conexão simples.

Quando o pacote chega ao servidor, o sistema operacional vê que o pacote está solicitando uma conexão. Ele verifica se existe um ouvinte e, se houver, desbloqueia o ouvinte. O processo servidor pode, então, estabelecer uma conexão com a chamada ACCEPT. Isso envia de volta uma confirmação (2) ao processo cliente para aceitar a conexão. A chegada dessa resposta libera o cliente. Nesse momento, o cliente e o servidor estão em execução e têm uma conexão estabelecida entre eles.

A analogia óbvia entre esse protocolo e a vida real ocorre quando um consumidor (cliente) liga para o gerente do serviço de atendimento ao consumidor de uma empresa. No início do dia, o gerente de serviço inicia a sequência ficando próximo ao telefone para atendê-lo caso ele toque. Mais tarde, o cliente efetua a chamada. Quando o gerente levanta o fone do gancho, a conexão é estabelecida.

A próxima etapa é a execução de RECEIVE pelo servidor, a fim de se preparar para aceitar a primeira solicitação. Normalmente, o servidor faz isso imediatamente após ser liberado de LISTEN, antes de a confirmação poder retornar ao cliente. A chamada de RECEIVE bloqueia o servidor.

Depois, o cliente executa SEND para transmitir sua solicitação (3), seguida pela execução de RECEIVE para receber a resposta. A chegada do pacote de solicitação à máquina servidora desbloqueia o processo servidor, para que ele possa processar a solicitação. Depois de terminar o trabalho, ele utiliza SEND para enviar a resposta ao cliente (4). A chegada desse pacote desbloqueia o cliente, que agora pode examinar a resposta. Se tiver solicitações adicionais, o cliente poderá fazê-las nesse momento.

Ao terminar, ele utiliza DISCONNECT para encerrar a conexão (5). Em geral, uma DISCONNECT inicial é uma chamada de bloqueio, suspendendo o cliente e enviando um pacote ao servidor para informar que a conexão não é mais necessária. Quando o servidor recebe o pacote, ele próprio também emite uma DISCONNECT, confirmando o pacote do cliente e liberando a conexão (6). Quando o pacote do servidor retorna à máquina cliente, o processo cliente é liberado e a conexão é interrompida. Em resumo, é assim que funciona a comunicação orientada a conexões.

É claro que a vida não é tão simples assim. Muitos detalhes podem dar errado. O sincronismo pode estar incorreto (por exemplo, CONNECT ser executada antes de LISTEN), os pacotes podem ser perdidos e muito mais. Examinaremos todas essas questões com muitos detalhes mais adiante; porém, por enquanto, a Figura 1.15 resume o funcionamento possível de uma comunicação cliente-servidor com datagramas confirmados, de modo que podemos ignorar os pacotes perdidos.

Considerando-se que são necessários seis pacotes para completar esse protocolo, alguém poderia perguntar por que não é utilizado um protocolo não orientado a conexões. A resposta é que, em um mundo perfeito, esse tipo de protocolo poderia ser usado e, nesse caso, seriam necessários apenas dois pacotes: um para a solicitação e outro para a resposta. Entretanto, em face de mensagens extensas em qualquer sentido (por exemplo, um arquivo com vários megabytes), erros de transmissão e pacotes perdidos, a situação muda. Se a resposta consistisse em centenas de pacotes, alguns dos quais pudessem se

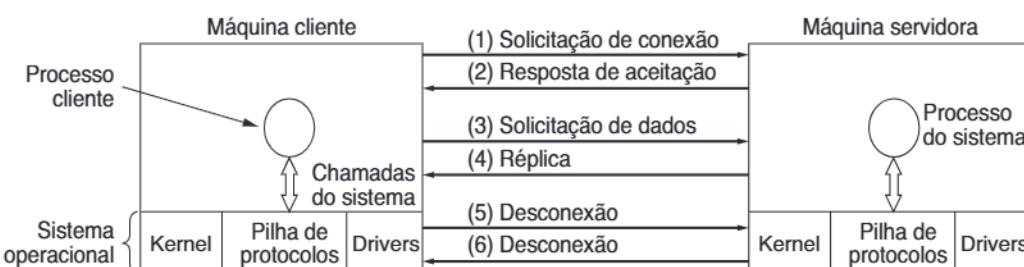


Figura 1.15 | Uma interação cliente-servidor simples, usando datagramas confirmados.

perder durante a transmissão, como o cliente saberia que alguns fragmentos se perderam? Como o cliente saberia que o último pacote recebido foi, de fato, o último pacote enviado? Suponha que o cliente quisesse um segundo arquivo. Como ele poderia distinguir o pacote 1 do segundo arquivo de um pacote 1 perdido do primeiro arquivo que repentinamente tivesse encontrado o caminho até o cliente? Em resumo, no mundo real, um simples protocolo de solicitação/resposta sobre uma rede não confiável normalmente é inadequado. No Capítulo 3, estudaremos em detalhes uma variedade de protocolos que superam esses e outros problemas. Por enquanto, basta dizer que às vezes ter um fluxo de bytes confiável e ordenado entre processos é muito conveniente.

1.3.5 RELACIONAMENTO ENTRE SERVIÇOS E PROTOCOLOS

Serviços e protocolos são conceitos diferentes. Essa distinção é tão importante que vamos enfatizá-la mais uma vez. Um *serviço* é um conjunto de primitivas (operações) que uma camada oferece à camada situada acima dela. O serviço define as operações que a camada está preparada para executar em nome de seus usuários, mas não informa absolutamente nada sobre como essas operações são implementadas. Um serviço se relaciona a uma interface entre duas camadas, sendo a camada inferior o fornecedor do serviço e a camada superior, o usuário do serviço.

Ao contrário, o *protocolo* é um conjunto de regras que controla o formato e o significado dos pacotes ou mensagens que são trocadas pelas entidades pares contidas em uma camada. As entidades utilizam protocolos com a finalidade de implementar suas definições de serviço. Elas têm a liberdade de trocar seus protocolos, desde que não alterem o serviço visível para seus usuários. Portanto, o serviço e o protocolo são independentes um do outro. Esse é um conceito fundamental, que qualquer projetista de rede precisa entender bem.

Em outras palavras, os serviços estão relacionados às interfaces entre camadas, como ilustra a Figura 1.16. Por outro lado, os protocolos se relacionam aos pacotes enviados entre entidades pares em máquinas diferentes. É importante não confundir esses dois conceitos.

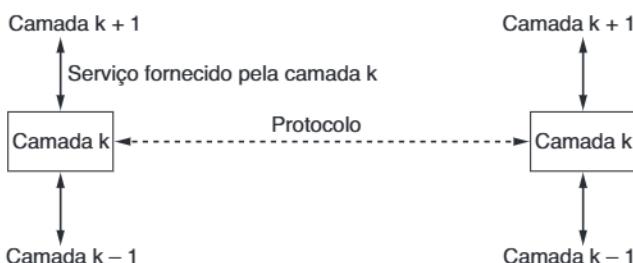


Figura 1.16 | Relacionamento entre um serviço e um protocolo.

Vale a pena fazer uma analogia com as linguagens de programação. Um serviço é como um objeto ou um tipo de dados abstrato em uma linguagem orientada a objetos. Ele define as operações que podem ser executadas sobre um objeto, mas não especifica como essas operações são implementadas. Em contraste, um protocolo se refere à *implementação* do serviço e, consequentemente, não é visível ao usuário do serviço.

Muitos protocolos mais antigos não distinguiam entre o serviço e o protocolo. Na prática, uma camada normal poderia ter uma primitiva de serviço send packet, com o usuário fornecendo um ponteiro para um pacote totalmente montado. Essa organização significava que todas as mudanças no protocolo ficavam imediatamente visíveis para os usuários. Hoje, a maioria dos projetistas de redes considera tal projeto um sério equívoco.

1.4 MODELOS DE REFERÊNCIA

Depois de discutirmos o conceito de redes em camadas em termos abstratos, vamos ver alguns exemplos. Nas duas seções a seguir, examinaremos duas importantes arquiteturas de rede: os modelos de referência OSI e TCP/IP. Embora os *protocolos* associados ao modelo OSI raramente sejam usados nos dias de hoje, o *modelo* em si é de fato bastante geral e ainda válido, e as características descritas em cada camada ainda são muito importantes. O modelo TCP/IP tem características opostas: o modelo propriamente dito não é muito utilizado, mas os protocolos são bastante utilizados. Por essa razão, examinaremos ambos em detalhes. Além disso, às vezes é possível aprender mais com os erros do que com os acertos.

1.4.1 O MODELO DE REFERÊNCIA OSI

O modelo OSI (exceto o meio físico) é representado na Figura 1.17. Esse modelo se baseia em uma proposta desenvolvida pela ISO (International Standards Organization) como um primeiro passo em direção à padronização internacional dos protocolos usados nas várias camadas (Day e Zimmermann, 1983). Ele foi revisado em 1995 (Day, 1995). O modelo se chama **Modelo de Referência ISO OSI (Open Systems Interconnection)**, pois ele trata da interconexão de sistemas abertos — ou seja, sistemas abertos à comunicação com outros sistemas. Para abreviar, vamos chamá-lo simplesmente de **modelo OSI**.

O modelo OSI tem sete camadas. Veja, a seguir, um resumo dos princípios aplicados para chegar às sete camadas.

1. Uma camada deve ser criada onde houver necessidade de outro grau de abstração.
2. Cada camada deve executar uma função bem definida.