

1 Introdução

1.1. Motivação

Grades computacionais são a nova tendência em computação distribuída [Foster 2003]. Originalmente concebido como um conceito para o compartilhamento de recursos computacionais e de comunicação entre nós fixos, o conceito de grade tem, gradativamente, sido estendido para ambientes móveis [KGrid 2002; Yamin et al. 2003; Akogrimo 2004; Kurkovsky et al. 2004; McKnight et al. 2004; Litke et al. 2004; Ihsan et al 2005; Lima et al. 2005]. Existem duas abordagens principais para a integração de redes sem fio em ambientes de grades computacionais. Na primeira abordagem, dispositivos móveis são usados apenas como interfaces para acesso sem fio às grades fixas convencionais [Hwang & Aravamudham 2004]. Na segunda abordagem, dispositivos móveis atuam como elementos computacionais da grade propriamente dita, disponibilizando os seus recursos e serviços, ou seja, esses dispositivos são responsáveis pelo processamento compartilhado e pela execução colaborativa de tarefas. Essa segunda abordagem, denominada nesta tese de **grades móveis**, é propiciada pela tendência atual de utilização de dispositivos móveis com capacidade de processamento e comunicação cada vez maiores, mas cujo uso atual tem sido predominantemente pessoal.

O maior desafio colocado pelas grades móveis é o *compartilhamento de recursos dinâmicos* – como capacidade de processamento e memória – dos dispositivos móveis, considerando-se as limitações dos mesmos em termos desses recursos e de outros recursos envolvidos – como fonte limitada de energia. Esse compartilhamento permite que um usuário, através de seu dispositivo móvel, possa interagir com dispositivos de outros usuários, utilizando os recursos computacionais dinâmicos e de comunicação desses dispositivos para a execução de tarefas complexas, que demandam maior capacidade computacional. Esse

compartilhamento só é possível através da utilização de arquiteturas e protocolos que permitam a identificação e seleção dos recursos e serviços computacionais ofertados na grade móvel. Essa perspectiva de compartilhamento cria formas inovadoras de utilização dos dispositivos sem fio, como, por exemplo, em situações de emergência, com a coordenação da tomada de decisões na gestão de crises provocadas por causas naturais – como terremotos e maremotos – ou pela ação do homem – como em atentados terroristas de grandes dimensões –, onde não existe uma infra-estrutura de comunicação ou a infra-estrutura existente não está disponível. Esses exemplos ilustram a necessidade de grades móveis [McKnight et al. 2004]. Em tais situações, o compartilhamento dos recursos computacionais dos dispositivos móveis é crucial para que se obtenha, em tempo hábil, formas de trabalho colaborativo, como, por exemplo, através da coleta e do processamento automático de informações sobre grandes grupos de pessoas feridas, de modo a promover uma melhor alocação dos recursos médicos disponíveis [Hwang & Aravamudham 2004].

Uma condição extrema de funcionamento de grades móveis ocorre nas redes sem fio *ad hoc* (*Mobile Ad hoc NETWORKS* – MANETs). Essas redes são caracterizadas pela total ausência de infra-estrutura, onde todos os dispositivos participantes são móveis e a topologia da rede é formada temporariamente, de forma dinâmica e independente, com os dispositivos podendo participar ou sair da grade móvel a qualquer momento, formando cenários de compartilhamento temporários. Os desafios relacionados a essas redes são ainda maiores quando se consideram cenários de saltos múltiplos. Em MANETs de saltos múltiplos, tanto a demanda quanto a disponibilidade de serviços e recursos podem apresentar uma alta variabilidade, em comparação com os cenários encontrados nas redes fixas e redes sem fio infra-estruturadas. A Figura 1 ilustra a formação de uma grade móvel *ad hoc* através de uma MANET de saltos múltiplos mostrando como um grupo de dispositivos móveis, localizados em uma mesma região geográfica [Figura 1(a)], interconectam-se. Na Figura 1(b), cada círculo tracejado representa o alcance de transmissão do nó posicionado no seu centro.

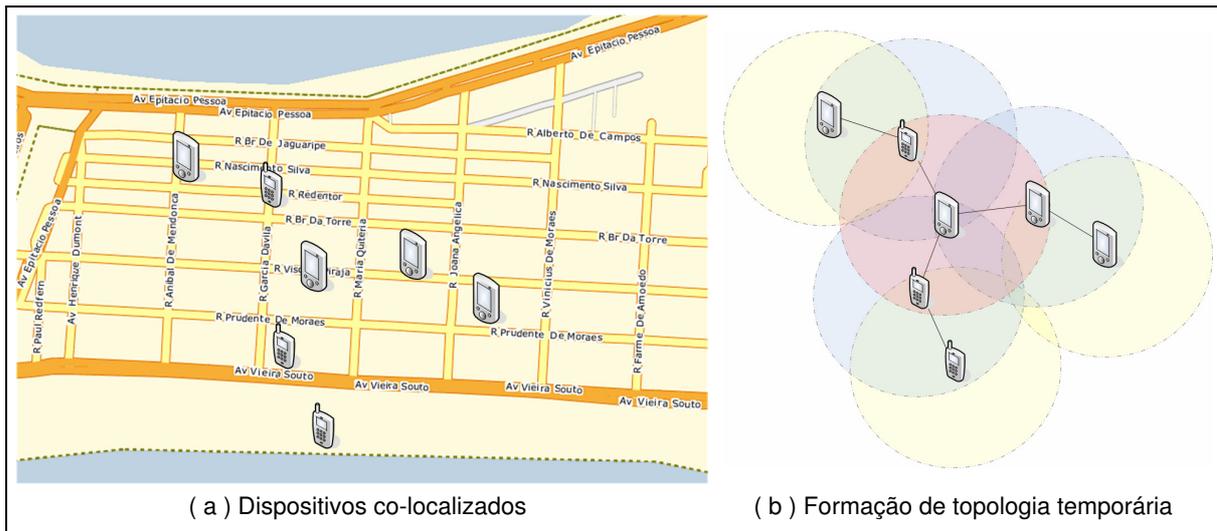


Figura 1 – Formação de uma grade móvel pela proximidade geográfica dos dispositivos.

1.2.

Requisitos para a Descoberta de Serviços em Grades Móveis *ad hoc*

O mecanismo de descoberta de serviços é um elemento fundamental em ambientes computacionais auto-organizáveis e, em particular, nas grades móveis *ad hoc*. De uma forma geral, protocolos de descoberta de serviços viabilizam a detecção automática de dispositivos e dos serviços oferecidos por esses dispositivos em uma rede de computadores [Marin-Perianu et al. 2005; Mian et al. 2006]. Em outras palavras, descoberta de serviços corresponde à ação de encontrar um provedor para um serviço requisitado. Um serviço, em uma rede, pode ser qualquer entidade de *software* ou *hardware* que um usuário deseje utilizar. Quando a localização do serviço solicitado é obtida – tipicamente o endereço do provedor –, o usuário pode, posteriormente, acessá-lo para sua utilização.

Nos ambientes de grade, os serviços são caracterizados como uma composição de recursos computacionais altamente dinâmicos. A disponibilidade desses recursos está sujeita a variações abruptas, como é o caso, por exemplo, da carga de CPU, cuja distribuição é bimodal, apresentando picos de utilização, ao longo do tempo, que variam entre carga total (100% de utilização) e valores próximos à inatividade [Bolosky et al. 2000]. Além disso, a flutuação verificada na disponibilidade dos recursos dinâmicos pode variar consideravelmente em períodos curtos de tempo. Nas grades móveis, particularmente, essa característica é agravada em função da mobilidade dos dispositivos e da sua heterogeneidade, o

que se reflete na capacidade, em termos de recursos, dos mesmos. Em redes sem fio, a taxa de mobilidade pode ser elevada e o tempo de permanência dos dispositivos na grade é variável, caracterizando altas taxas de *churn*.¹ Portanto, além da constituição dos serviços como recursos dinâmicos, típico das grades computacionais, nas grades móveis é importante considerar também aspectos relacionados à mobilidade dos dispositivos, que caracterizam esses ambientes como altamente dinâmicos, onde se verifica uma flutuação constante na topologia da rede. Devido a esses fatores, um protocolo de descoberta de recursos e serviços para grades móveis deve ser primordialmente reativo, baseando-se no envio de requisições sob demanda [Frank & Karl 2004].

Nesse ponto é necessário deixar claro o cenário alvo deste trabalho. A solução aqui apresentada para a descoberta e seleção de recursos dinâmicos, em muitos aspectos, é genérica para ambientes de grades computacionais, considerando-se que algumas adaptações sejam feitas, entretanto, a sua especificação foi concebida tendo como foco um ambiente para execução distribuída de tarefas entre dispositivos sem fio, organizados de forma *ad hoc* em uma grade móvel [McKnight et al. 2004; Kurkovsky et al. 2004; Lima et al. 2005]. Esse trabalho surgiu no contexto da arquitetura MoGrid (*Mobile Grid*) [Lima et al. 2005], apresentada em mais detalhes no Capítulo 2, que foi projetada com o intuito de facilitar o desenvolvimento de aplicações de grade para execução entre dispositivos móveis interconectados através de uma rede sem fio. É importante ressaltar que a solução proposta nesta tese foi avaliada somente no contexto das grades móveis.

Do que foi exposto, pode-se destacar três requisitos principais que devem ser atendidos por um protocolo de descoberta e seleção de recursos e serviços, especificado para uma grade móvel organizada através de uma rede sem fio *ad hoc* de saltos múltiplos, conforme será discutido nos próximos parágrafos: (i) reduzir o impacto do problema da implosão de mensagens de resposta a uma requisição, (ii) minimizar a ocorrência de colisões provocadas pelo aumento do tráfego de dados no enlace sem fio – em virtude da transmissão de mensagens do

¹ Termo comumente empregado em redes P2P para definir a transitoriedade dos dispositivos, que

protocolo de descoberta – e (iii) realizar uma distribuição uniforme das requisições de serviço, promovendo um balanceamento de carga implícito através da seleção dos dispositivos mais aptos.

Redução do impacto do problema da implosão de mensagens de resposta.

É comum que ocorra uma sobrecarga nos dispositivos em protocolos de descoberta de serviços reativos, nos quais as requisições de serviço são feitas sob demanda e dá-se uma interação direta entre clientes e provedores de serviço. Essa sobrecarga é acarretada pela troca de mensagens de descoberta entre eles, tanto no que diz respeito ao seu processamento quanto no aumento do consumo de energia provocado pela sua transmissão. Como nas abordagens reativas a requisição é propagada na rede utilizando um mecanismo de difusão por *broadcast*, é esperado que os dispositivos que oferecem o serviço solicitado respondam prontamente, provocando uma implosão de mensagens de resposta [Duffield et al. 1999]. Esse problema é decorrente do volume potencialmente grande de respostas geradas pelos dispositivos provedores do serviço requisitado, situação essa que se agrava em redes de grande escala. Para aplacar os efeitos colaterais da transmissão por difusão, é necessária a utilização de um mecanismo capaz de reduzir o tráfego de mensagens de resposta na rede, levando em consideração o número de instâncias de serviço solicitadas pelo dispositivo requisitante. Essa redução deve considerar a adequação dos dispositivos respondentes em oferecer o serviço, promovendo uma estratégia de seleção das respostas com maior qualidade ao longo da rede. A solução para essa questão é obtida através de um mecanismo de supressão das respostas excedentes, que atua nos nós intermediários, conforme elas são encaminhadas na rede em direção ao dispositivo que solicitou o serviço.

Redução do número de colisões provocadas pelo aumento do tráfego de mensagens de descoberta. A probabilidade de ocorrência de colisões pode crescer em decorrência do aumento de tráfego no enlace sem fio, provocado pela intensa troca de mensagens de controle – requisição e resposta – do protocolo de descoberta entre os dispositivos móveis. Nesse cenário, é verificado um alto índice de propagação de mensagens de requisição de descoberta na rede e, como conseqüência, um volume potencialmente grande de respostas, caracterizando,

entram e saem do sistema a qualquer instante [Stutzbach & Rejaie 2006].

respectivamente, situações típicas de inundação e implosão de mensagens, o que provoca um aumento na probabilidade da ocorrência de colisões [Ni et al. 1999; Duffield et al. 1999; Tseng et al. 2003]. Para reduzir o impacto do problema descrito, além de adotar um mecanismo de supressão de respostas excedentes, como descrito no item anterior, faz-se necessário adotar um mecanismo de retardo programado que promova, implicitamente, um assincronismo no envio das mensagens de controle do protocolo de descoberta, propiciando, dessa forma, a redução do número de colisões dessas mensagens.

Balanceamento de carga em função da distribuição das requisições de serviços. Para se obter uma distribuição uniforme de serviços entre os dispositivos da grade móvel, com um balanceamento de carga, deve-se considerar a alocação prévia dos recursos, efetuada através do envio de mensagens de resposta e da utilização dos serviços. Portanto, em uma grade móvel, o escalonamento deve adotar uma abordagem distribuída, podendo ser realizado em dois pontos distintos: no envio das requisições por serviços ou no encaminhamento das mensagens de respostas a essas requisições. Na primeira abordagem os dispositivos devem trocar periodicamente informações sobre a sua disponibilidade de recursos e serviços para que a entrega da requisição seja efetuada somente para os dispositivos mais aptos, incorporando um mecanismo de anúncio ao protocolo de descoberta. Essa abordagem é descartada haja vista a defesa realizada no início desta seção por um mecanismo de descoberta exclusivamente reativo para as grades móveis. Na segunda abordagem, cada dispositivo qualifica a sua própria resposta em função da sua disponibilidade de recursos, ou seja, a partir do seu conhecimento sobre os recursos disponíveis localmente, sem depender do conhecimento da disponibilidade de recursos dos demais dispositivos da rede. Em um cenário com tantas restrições e tamanha dinamicidade da topologia de rede, é importante que, ao se efetuar uma requisição de serviço, possa haver uma seleção natural das melhores respostas, isto é, dos dispositivos mais aptos a atender às necessidades definidas na requisição. Sob a óptica desta tese, as melhores respostas são aquelas provenientes dos dispositivos com mais recursos disponíveis. Essas respostas são selecionadas no seu encaminhamento na rede, sendo privilegiadas em detrimento das demais, consideradas excedentes. A implementação do critério de seleção de respostas baseia-se na informação de

contexto de cada dispositivo. Nesta tese, entende-se por *contexto* toda informação que represente o estado dos dispositivos móveis, incluindo a qualidade do enlace sem fio, carga de CPU, carga residual da bateria, memória disponível e espaço de armazenamento em disco.

1.3. Objetivos

O objetivo desta tese é o desenvolvimento e a avaliação de desempenho de um protocolo de descoberta e seleção de recursos e serviços que atenda às necessidades das grades móveis, considerando fatores como a intermitência dos enlaces sem fio e a variação na disponibilidade dos serviços. Para tanto, o protocolo de descoberta deve adotar uma abordagem reativa, em que os serviços são descobertos sob demanda pelos dispositivos requisitantes, através da difusão de mensagens de requisição, e as respostas são selecionadas enquanto são encaminhadas da sua origem até o nó requisitante.

Nesta tese é proposta uma arquitetura descentralizada para atender a demanda das grades móveis organizadas através de redes sem fio *ad hoc*. Essa arquitetura, denominada MoGrid, foi projetada para oferecer um ambiente de descoberta e execução de serviços, definidos como uma composição de recursos dinâmicos, em grades móveis, promovendo a distribuição, entre os dispositivos sem fio, da execução dos serviços requisitados, adotando uma abordagem *peer-to-peer* (P2P). Essa arquitetura compreende uma camada de descoberta e seleção de recursos e serviços e uma camada de transparência que trata as questões relacionadas à conectividade irregular da grade móvel. Durante o processo de especificação e implementação da arquitetura MoGrid, optou-se por concentrar o desenvolvimento desta tese na camada de descoberta e seleção de serviços, que, por si só, já possui uma complexidade considerável.

Nesta tese, é proposto o P2PDP (*Peer-to-Peer Discovery Protocol*), um novo protocolo que integra os mecanismos de descoberta e seleção de recursos e serviços ao mecanismo de roteamento de pacotes de dados, em uma abordagem completamente descentralizada, independente de um endereçamento de rede explícito, operando no nível da camada de aplicação. No caso de serviços não específicos, como ciclos de processamento e espaço de armazenamento, os quais,

tipicamente, podem ser oferecidos por muitos dispositivos, a responsabilidade de prover o serviço é compartilhada entre os dispositivos que constituem a grade móvel. Esse compartilhamento é feito de modo uniforme, através do uso do protocolo P2PDP, tendo como critério de distribuição da tarefa de provisão de serviços a quantidade de recursos disponível em cada dispositivo. Já com os serviços de um tipo específico, como um serviço de impressão, em cuja requisição está implícita a necessidade do cliente por um único provedor do serviço, o objetivo do protocolo P2PDP é selecionar, dentre as possíveis respostas a requisição de serviço, aquela que melhor atenda às necessidades do cliente, considerando, também, em alguns casos, o nível de energia disponível, a qualidade do enlace sem fio e a proximidade física entre provedor e cliente, definida em função da distância em número de saltos.

O protocolo de descoberta P2PDP atende às necessidades das grades móveis *ad hoc*, promovendo a descoberta de serviços sob demanda, em tempo de execução, acionada pela necessidade dos usuários. O processamento introduzido pelo P2PDP não promove gargalos na sua execução e nem gera uma sobrecarga administrativa nos dispositivos da rede, o que se verifica pelo fraco acoplamento entre as entidades envolvidas no processo de descoberta – provedores de serviços e clientes –, as quais entram e saem da rede, a qualquer instante, sem que isso comprometa o funcionamento do mecanismo de descoberta. O protocolo P2PDP foi desenvolvido no contexto da arquitetura MoGrid, com o intuito de facilitar o desenvolvimento de aplicações colaborativas, baseadas no compartilhamento de recursos dinâmicos, para a execução de serviços na grade móvel. A arquitetura MoGrid é apresentada em mais detalhes no Capítulo 2.

A avaliação do protocolo proposto foi realizada em duas frentes. A primeira corresponde a sua implementação, dentro de um *middleware* para grades móveis, e, subseqüentemente, a sua utilização no desenvolvimento de aplicações-protótipo para essas grades; nessa frente, é realizada uma avaliação experimental do protocolo através da análise do seu funcionamento em cenários reais, representados por redes de teste. A segunda frente diz respeito à realização de simulações para avaliar o seu desempenho em cenários de maior escala.

1.4. Sumário das Principais Contribuições

Esta tese apresenta o projeto, implementação e avaliação de desempenho de um protocolo para descoberta e seleção de recursos e serviços, especificamente desenvolvido para atender às necessidades das grades móveis *ad hoc*: o protocolo P2PDP (*Peer-to-peer Discovery Protocol*). As principais contribuições alcançadas na área de descoberta de serviços, com o desenvolvimento deste trabalho, podem ser classificadas em duas categorias: contribuições para o estado da arte e contribuições tecnológicas.

Contribuições para o estado da arte

O desenvolvimento desta tese teve como resultado duas contribuições principais para o estado da arte na área de descoberta de serviços, a especificação do protocolo P2PDP e da arquitetura MoGrid, a saber:

- A concepção, implementação e avaliação de desempenho do algoritmo de supressão de respostas por vizinhança (*Suppression by Vicinity – SbV*), que permite tratar o problema da implosão de respostas em protocolos de descoberta para MANETs, baseados na transmissão por difusão, utilizando *broadcast*;
- A concepção, implementação e avaliação de desempenho do algoritmo de retardo programado no envio de mensagens de resposta (*Delayed Replies – DR*), com o intuito de promover uma seleção natural entre as respostas que melhor se adequam às especificidades de uma dada requisição de serviço;
- Especificação de um conjunto de requisitos para o desenvolvimento de um protocolo para descoberta e seleção de recursos e serviços em grades móveis *ad hoc*. Os requisitos do projeto do protocolo de descoberta levam em consideração as particularidades dos recursos computacionais dinâmicos, que caracterizam os serviços disponíveis em uma grade móvel;
- Definição de uma arquitetura genérica em camadas, denominada MoGrid, que serviu de base para o desenvolvimento de um *middleware*, de mesmo

nome, que trata aspectos relacionados à computação distribuída em grades móveis, organizadas através de redes sem fio infra-estruturadas e *ad hoc*.

Contribuições tecnológicas

Além das contribuições científicas para o estado da arte, foi desenvolvido um conjunto de artefatos de *software* para auxiliar o desenvolvimento de aplicações colaborativas, em grades móveis, baseadas no compartilhamento de recursos e serviços, a saber:

- Implementação do *middleware* MoGrid e do protocolo de descoberta P2PDP;
- Implementação de um *proxy* de colaboração com a funcionalidade de permitir o acesso de dispositivos sem fio aos serviços e recursos disponíveis em uma grade fixa;
- Uso do *middleware* MoGrid para o desenvolvimento de aplicações colaborativas, baseadas no compartilhamento de serviços gerenciado pelo protocolo P2PDP.

1.5. Organização da Tese

O restante desta tese está organizada como se segue:

O Capítulo 2 contextualiza este trabalho apresentando a arquitetura MoGrid, projetada com o intuito de oferecer suporte ao desenvolvimento de aplicações baseadas no compartilhamento de recursos em grades móveis, para a tomada de decisões referentes à descoberta e ao acesso a esses recursos. A arquitetura compreende uma camada de descoberta de serviços *peer-to-peer* – na qual foi desenvolvido o protocolo de descoberta P2PDP, proposto nesta tese – e uma camada de transparência – que lida com problemas relacionados a períodos de desconexão durante os processos de descoberta de recursos, submissão e execução de tarefas. Ao final do capítulo, são apresentados os trabalhos relacionados que dizem respeito a arquiteturas de grades móveis.

O Capítulo 3 é reservado aos trabalhos relacionados à área-tema desta tese – descoberta de serviços para redes sem fio *ad hoc* –, que é feita tendo como base

uma classificação genérica proposta para os protocolos de descoberta de serviços em redes de computadores. Na revisão bibliográfica é dada uma maior ênfase aos trabalhos que tratam, especificamente, da descoberta de serviços em redes sem fio *ad hoc* de saltos múltiplos, abordagem similar à adotada no protocolo P2PDP. Ao final do capítulo, tem-se um resumo comparativo, que utiliza os critérios enumerados na classificação apresentada, incluindo os principais protocolos desenvolvidos para as redes fixas que foram adaptados às redes sem fio infra-estruturadas, os protocolos para as redes sem fio *ad hoc* de salto único e os protocolos para as redes sem fio de saltos múltiplos.

No Capítulo 4, é apresentado o protocolo para descoberta e seleção de recursos em grades móveis *ad hoc*, denominado P2PDP (*Peer-to-Peer Discovery Protocol*), sendo o foco principal deste trabalho. O núcleo do protocolo se baseia em um mecanismo distribuído, que emprega dois algoritmos principais: o algoritmo de supressão de mensagens de resposta por vizinhança (*Suppression by Vicinity*) e o algoritmo de retardo programado no envio de mensagens de resposta (*Delayed Replies*). Esses algoritmos oferecem uma estratégia para reduzir a implosão e a colisão de respostas em grades móveis *ad hoc*, além de proporcionar um mecanismo de seleção das melhores respostas às requisições de serviço.

O Capítulo 5 é dedicado às questões relacionadas à implementação do protocolo P2PDP, trazendo as decisões de projeto quanto às tecnologias utilizadas, além de alguns diagramas de classe e de seqüência, com o intuito de facilitar a compreensão do comportamento das entidades envolvidas no processo de descoberta e seleção de recursos. Esse capítulo traz maiores detalhes sobre a utilização do protocolo P2PDP e da API fornecida pela arquitetura MoGrid no desenvolvimento de três aplicações colaborativas diferenciadas. Um cenário de simulação é utilizado para atestar a correção da implementação da arquitetura MoGrid e do protocolo P2PDP.

No Capítulo 6, é feita a avaliação de desempenho do protocolo P2PDP por meio da interpretação dos resultados obtidos a partir das simulações e da rede de teste. Diferentes cenários foram utilizados com a variação de parâmetros, como o tamanho da rede, a densidade de dispositivos, o raio de transmissão, o impacto de diferentes taxas de requisição, o percentual de dispositivos atuando como provedores de serviço, entre outros.

Finalmente, no Capítulo 7, são tecidas as considerações finais, com destaque para as contribuições decorrentes deste trabalho e a apresentação de problemas não tratados, mas que podem gerar trabalhos futuros interessantes.

O Apêndice traz alguns detalhes sobre as simulações executadas para avaliar a correção e o desempenho do protocolo de descoberta P2PDP e do *middleware* MoGrid.