# 1
# Introdução

## 1.1
## Motivação

Muitas aplicações para computadores são sistemas distribuídos nos quais (muitos) componentes constituintes estão espalhados através de uma rede, em um regime de controle descentralizado, e que estão sujeitos a mudanças constantes durante o tempo de vida do sistema. Exemplos incluem computação ponto-a-ponto (Ora01), web semântica (Ber01), serviços da web (McI01), e-business (Dei01), m-commerce (Sad02, Vul99), computação independente (*autonomic computing*) (Kep03), computação em grade (*grid computing*) (Fos98) e ambientes de computação pervasivos (Sch02). Em todos esses casos, há a necessidade de se ter componentes autônomos que agem e interagem de maneira flexível de modo a atingir os objetivos para os quais foram criados em ambientes com incertezas e dinâmicos (Sim96). Dado isso, a computação baseada em agentes tem sido defendida como um modelo computacional natural para tais sistemas (Jen01, Ram04).

Mais especificamente, sistemas distribuídos abertos podem ser modelados como sistemas multi-agentes (MAS, *Multi-Agent System*) (Woo02) abertos que são compostos de agentes autônomos que interagem uns com os outros usando mecanismos particulares e protocolos. Portanto, as interações são o núcleo de um sistema multi-agentes. Logo, de maneira não surpreendente, a comunidade de pesquisa em agentes desenvolveu vários modelos de interação entre agentes (Ram04). Entretanto, suas aplicações em sistemas multi-agentes de larga-escala apresentam novos desafios. Primeiramente, os agentes tendem a representar diferentes partes interessadas, cada qual com seus próprios interesses e objetivos. Isso significa que a estratégia de construção mais plausível para um agente é maximizar o seu ganho individual (Neu44, Ram04). Em segundo lugar, dado que o sistema é aberto, agentes podem entrar e sair a qualquer momento, não se conhecendo a priori quem são os agentes que irão interagir nem quem foram os desenvolvedores desses agentes (Les07). Isso significa que um agente pode mudar a sua identidade ao re-entrar no sistema e

se livrar de ser punido por algo errado feito no passado. Em terceiro lugar, um sistema distribuído aberto permite que agentes com características estruturais diferentes (por exemplo, políticas, habilidades e papéis) entrem no sistema e interajam uns com os outros. Em quarto lugar, um sistema distribuído aberto permite que os agentes vendam produtos e serviços e colaborem entre si de diversas maneiras. Logo, os criadores de agentes encaram uma escolha de vários protocolos de interação potenciais que podem ajudar os agentes a atingir os objetivos para os quais foram construídos. Um protocolo define como os agentes devem interagir uns com os outros, e também, através de técnicas de segurança, ajuda a tornar o sistema seguro, impondo restrições nas interações dos agentes. Mais ainda, um protocolo determina um conjunto de regras para a interação dos agentes, e a sua intenção é que a seqüência de movimentos dos agentes e a alocação de recursos, promovidos pelo protocolo, sejam feitas de maneira tal de forma a prevenir que agentes manipulem os outros agentes para satisfazer os seus próprios interesses. Todavia, protocolos muito restritivos podem ser impraticáveis e as técnicas de segurança não garantem a veracidade das mensagens enviadas nem a qualidade das atitudes dos agentes nas suas interações.

É o agente que deve decidir quando, como e com quem interagir sem nenhuma garantia de que a interação de fato vai levar o agente a conseguir os benefícios desejados. Tomar tais decisões, idealmente, requer que o agente esteja completamente informado sobre seus oponentes, ambientes e riscos a correr. Tais informações permitem aos agentes calcular as probabilidades de que certos eventos ocorram e, portanto, possibilitam aos mesmos a atuar de maneira a maximizar o seu ganho esperado (Sav54). Ainda, dadas tais informações, os agentes devem ser capazes de agir estrategicamente calculando a melhor resposta dados os próximos movimentos dos seus oponentes durante o curso da interação (Bin92).

Todavia, tanto o sistema (que impõe o protocolo), quanto os agentes, possuem capacidades computacionais e de armazenamento limitadas que restringem seu controle sobre as interações. Adicionalmente, os limites de banda e de velocidade dos canais de comunicação limitam as capacidades de sensibilidade em aplicações no mundo real. Mais ainda, os agentes não têm como prever como será o comportamento futuro de um agente que acabou de entrar no sistema se ele for aberto (onde agentes podem entrar e sair a qualquer momento). Conseqüentemente, em contextos práticos, é normalmente impossível ao agente atingir uma situação onde possui informações perfeitas (100% de conhecimento) sobre o ambiente e sobre as propriedades, possíveis interesses e estratégias dos parceiros com os quais interage, onde estratégias são atitudes

dos agentes para conseguir o maior ganho possível (Bin92, Rus95a, Axe84). Por isso, os agentes necessariamente encaram níveis significativos de incerteza ao tomar decisões (pode ser muito difícil ou impossível projetar probabilidades para o acontecimento de eventos). Em tais circunstâncias, os agentes devem *confiar* uns nos outros para minimizar a incerteza associada às interações em ambientes abertos distribuídos.

## 1.2 Objetivos

Um sistema multi-agentes aberto é composto por vários agentes, que neste trabalho são definidos formalmente usando o modelo BDI, ou seja, agentes inteligentes que possuem estados mentais de crenças (*Beliefs*), desejos (*Desires*) e intenções (*Intentions*) (Jen01, Woo99, Woo00). O modelo BDI é usado porque possui uma filosofia baseada no raciocínio prático de seres humanos, uma arquitetura de *software* implementável em sistemas reais e uma família de lógicas que suportam uma teoria formal de agentes inteligentes (Woo00).

Como já dito anteriormente, a existência de confiança entre os agentes é fundamental para que haja interações entre os mesmos. Afinal, um agente não só pode não atingir os seus objetivos como pode sofrer danos ao interagir com um agente mal-intencionado. Portanto, ele deve ter um modelo de confiança que indica em quais agentes ele pode confiar e em quais ele não deve confiar. Neste texto, o agente BDI tem uma camada extra, representando a confiança explicitamente, além das crenças, dos desejos e das intenções. Isso é feito porque acredita-se que a confiança em um outro agente não deva ser descrita como uma crença. Embora sejam conceitos parecidos, uma crença é uma representação interna de algo que o agente percebe em seu ambiente. Uma crença vai estar errada se o agente não puder perceber adequadamente o ambiente onde está presente. Embora seja possível dizer o mesmo da confiança, ao contrário da crença, a confiança em um outro agente pode estar errada mesmo que o primeiro possa perceber o comportamento do último corretamente. Basta que o agente que está sendo observado engane o outro, se comportando de uma maneira diferente da esperada de acordo com o comportamento observado anteriormente. Portanto, quando um agente confia ou não em outro, ainda que ele perceba o comportamento do outro corretamente, ele não tem a garantia de que seu comportamento será como o esperado.

Para dar uma semântica a um sistema multi-agentes com confiança aberto, usa-se lógica modal (Che80, Gab84, Gol92, Hug96, Tro00). Usando-se lógica, os diferentes aspectos de um modelo de confiança podem ser expressos

de forma precisa e formal, sem ambigüidades. Com lógica modal, pode-se modelar quais agentes estão presentes em um dado instante de tempo usando-se a semântica de mundos possíveis, onde cada mundo representa quais agentes estão presentes no sistema em cada instante de tempo. Dessa forma, modela-se a parte externa do sistema, mas não os estados mentais de cada agente. Logo, é também preciso dar uma semântica aos mesmos. A abordagem adotada foi deixar as componentes do agente BDI como modalidades e a confiança como um predicado relacionado com as crenças de um agente, que seria definido de acordo com o que pode ser entendido como confiança de um agente em outro.

A partir dessa formalização lógica, é possível se formular propriedades desejáveis e teoremas para que o sistema multi-agentes aberto funcione adequadamente.

## 1.3 Trabalhos Relacionados

Neste trabalho considera-se confiança e crença como sendo dois conceitos distintos. Porém, em (Fal01a, Fal01b, Ram04), a confiança é definida como uma crença que um agente tem em outros de que vão fazer o que dizem (sendo honestos e seguros) ou tendo reciprocidade (agindo para o bem de ambos), dada uma oportunidade de trapacear para conseguir maior ganho. Esses trabalhos abordam modelos de confiança, de agentes em outros agentes, e protocolos e mecanismos que garantem o bom funcionamento do sistema, os quais os agentes são obrigados a cumprir.

Já em (Cas98, Cas00a, Cas00b, Fal01a, Fal01b), também modelando a confiança como crença, ao avaliar a confiança em um oponente, leva-se em conta a percepção subjetiva no mesmo, posto que permite uma análise mais compreensiva das características do oponente. Por subjetiva, entende-se que a percepção é formada de acordo com a avaliação do ambiente e das características do oponente, que podem incluir uma análise de interações passadas. Tais informações são armazenadas em um estado mental do agente e são essenciais para saber a capacidade de um agente fazer o que diz que vai fazer ou a sua vontade de fazer o que diz que vai fazer (ser honesto).

Em particular, esses trabalhos destacam a importância de uma visão cognitiva da confiança, principalmente para agentes BDI. O contexto dos trabalhos é o da delegação de tarefas onde um agente $x$ deseja delegar uma tarefa ao agente $y$. Para fazer isso, o agente $x$ precisa avaliar a confiança que pode depositar em $y$ considerando as diferentes crenças que ele tem sobre as motivações do agente $y$. Afirma-se que as seguintes crenças são essenciais no estado mental de $x$ para determinar a quantidade de confiança que $x$ deve

colocar em $y$: competência, vontade, persistência e motivação.

Já neste texto, quando se analisa a confiança em $y$ em termos de suas motivações (ou capacidades), considera-se que se tem diferentes variáveis de confiança, já que todas são apostas, ou seja, $x$ não tem como ter a certeza do comportamento de $y$ em cada aspecto.

Para calcular o nível de confiança que o agente $x$ pode ter no agente $y$, o agente $x$ precisa considerar cada uma das crenças acima e, possivelmente, outras. Essas crenças de fato causam impacto na confiança, cada uma de uma maneira diferente, e isso precisa ser levado em conta em uma avaliação compreensiva de todas essas crenças. Esses trabalhos, tal como o presente texto, são fortemente motivados nos estados mentais de seres humanos, que nem sempre são racionais (ao contrário do que se espera de um agente).

Em contextos práticos, como na Amazon.com e no eBay, que são *sites* de comércio eletrônico e leilão *online*, a questão da confiança é muito importante pois um agente mal-intencionado pode causar grandes perdas financeiras nos demais e fazer com que os agentes lesados não queiram mais usar os serviços dos *sites* por achá-los inseguros. Neles há um sistema de reputação onde os agentes dão uma nota e fazem comentários com relação aos parceiros com quem interagiram. Qualquer agente pode ver essas notas e comentários, para melhor decidir com quem interagir. Esse é um sistema claramente centralizado, onde os agentes consultam o sistema e não os seus parceiros para saber se devem confiar nos demais ou não. Também não há informações separadas sobre os atributos dos agentes, exibindo em quais capacidades o agente tem maior reputação e, por sua vez, é mais confiável. Porém, há uma maior ênfase no comportamento do passado recente do agente do que no comportamento mais antigo. Ou seja, considera-se que o comportamento futuro do mesmo tenderá a ser semelhante ao mais recente e não ao mais antigo (Gue06, Gue07).

Neste trabalho considera-se que um agente tem que perguntar aos demais o que eles pensam de um terceiro para concluir se deve confiar ou não nesse. Já no artigo (Gue06), essa consulta é feita à organização (Luc04, Luc05) onde o agente se encontra, a qual supostamente possui todas as informações sobre seus agentes e como eles têm agido dentro da mesma. Essa abordagem diminui a busca por informações sobre determinados agentes procurando outros com os quais eles tenham interagido no passado. Contudo, um agente pode pertencer a mais de uma organização, tendo comportamentos diversos nas diferentes organizações. Portanto, o comportamento em uma pode não refletir o comportamento usual do agente. Nesse artigo, ainda tem-se que, se um agente fizer algo errado no passado e se comportar adequadamente depois, sua reputação vai subindo aos poucos novamente.

Nos dois casos acima, tanto ao se pesquisar a reputação de um agente no sistema como numa organização, o agente pode não acreditar nas informações que ele recebeu, por não confiar no sistema e/ou na organização, que podem ser tratados como agentes especiais que representam os mesmos. Portanto, no primeiro caso, embora haja uma centralização da informação da reputação de cada agente, individualmente os agentes podem confiar em outros agentes com um grau diferente do que o sistema lhes informa.

Em (Woo00), explica-se o que é um agente BDI, detalhando cada um de seus componentes e as diferentes implementações para um agente inteligente. Também apresenta-se uma lógica chamada LORA (*Logic of Rational Agents*), contendo uma representação temporal, que permite a representação da dinâmica de como os agentes e os seus ambientes se modificam através do tempo e a representação de ações que os agentes executam e os efeitos dessas ações. LORA pode ser usada para capturar muitos componentes de uma teoria de agentes racionais, incluindo noções de comunicação e cooperação. Esse trabalho lida com sistemas fechados, ou seja, não há a preocupação de se lidar com a entrada e a saída de agentes no sistema. Neste texto LORA foi estendida para lidar com a entrada e a saída de agentes do sistema e a inclusão de um modelo de confiança.

Já em (Ben01, Giu00), é apresentada uma lógica de crenças para protocolos de segurança usando a lógica BAN (Bur89). Nesse trabalho, há tanto um componente temporal, que reflete o comportamento dos agentes que querem se comunicar, como o componente das crenças sobre os demais agentes do sistema. A lógica usada é a CTL e *model checking* é utilizado para verificar a validade de determinadas propriedades no sistema.

Por fim, em (Ama00, Seg89), aborda-se a noção do operador *"bring-it-about"* que, quando aplicado a uma dada proposição, denota o conjunto de ações necessárias para fazer com que essa proposição seja verdadeira. Isso é semelhante a tentativa de um agente de atingir uma intenção. Afinal, para atingi-la, ele precisa executar ações, onde a intenção é também representada por uma fórmula lógica.

## 1.4 Organização do Texto

O capítulo 2 explica o que são agentes inteligentes, explicando primeiramente o que são agentes e posteriormente definindo agentes inteligentes. A seguir, são mostradas algumas arquiteturas para agentes inteligentes, tanto em um nível mais abstrato como em um nível mais concreto. Por fim, são comentadas algumas aplicações para agentes inteligentes.

Os dois primeiros capítulos representam a primeira grande parte desta tese que, basicamente, faz uma introdução para contextualizar o trabalho desenvolvido.

O capítulo 3 explica como se implementar agentes inteligentes usando agentes BDI, que são o foco deste texto. Ele basicamente mostra várias versões de um *loop* de controle de um agente focando diferentes estratégias de comprometimento, ou seja, como o agente se comporta com relação ao seu ambiente e com os seus objetivos a serem atingidos.

O capítulo 4 explica como se colocar confiança no modelo de agentes inteligentes. O capítulo começa explanando o que é confiança e porque ela é importante para um sistema multi-agentes. A seguir, é mostrado por que confiança não deve ser implementada como crença simplesmente, como mostrado nos trabalhos relacionados na sessão 1.3 (Cas00a, Cas00b, Fal01a, Fal01b). Posteriormente, mostra como a camada de confiança se relaciona com as demais de um agente BDI. Ainda, comenta-se as duas abordagens para confiança em um sistema multi-agentes: modelo de confiança e protocolos. Em seguida, é explicado como cada estratégia de comprometimento citada no capítulo três deve ser modificada para incluir a confiança. Por fim, mostra-se como a confiança é implementada e atualizada em um agente.

Os capítulos 3 e 4 representam a segunda grande parte deste texto, explicando basicamente como se implementar agentes sem e, posteriormente, com confiança.

O capítulo 5 explica uma lógica multi-modal para modelar um sistema multi-agentes aberto com o conceito de confiança nos agentes. Primeiramente, ela lida com a modelagem de um sistema multi-agentes aberto, onde os agentes podem entrar e sair. Logo, é necessária uma semântica de mundos possíveis onde cada mundo seja representado por quais agentes estão presentes no sistema em um determinado instante de tempo. Posteriormente, para cada agente do sistema, modela-se a semântica de mundos possíveis dos estados mentais de crenças, desejos e intenções do mesmo, fazendo de cada uma dessas três componentes do agente uma modalidade diferente e, para cada uma delas, uma semântica de mundos possíveis. Para isso, também define-se uma linguagem lógica, apresentando-se a sua respectiva semântica. Já a confiança é modelada como um predicado em lógica de primeira ordem que é relacionado com as crenças de um agente. O capítulo ainda mostra uma série de propriedades possivelmente desejáveis com relação à confiança de um agente em seus parceiros. Ao final, são apresentados dois exemplos simples do uso da lógica para expressar propriedades em um MAS aberto com confiança.

O capítulo 5 representa a terceira e a principal parte deste trabalho, que

é o desenvolvimento de uma lógica para modelar um sistema multi-agentes aberto composto por agentes BDI com um modelo de confiança.

O capítulo 6 apresenta a conclusão deste trabalho, enfatizando o porquê de se explicitar o conceito de confiança no modelo de um agente que está inserido em um ambiente multi-agentes aberto e as contribuições do presente texto. Finalmente, aborda-se os trabalhos futuros que podem vir a ser frutos do presente texto.

O capítulo 6 representa a quarta e a última grande parte do trabalho aqui apresentado.