



SPOLM 2007

ISSN 2175-6295

Rio de Janeiro- Brasil, 08 e 09 novembro de 2007.

CONFIANÇA EM AGENTES INTELIGENTES

Juliana Carpes Imperial

Departamento de Informática – Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)
Rua Marquês de São Vicente, 225 – Gávea -- 22.453-900 -- Rio de Janeiro – RJ

juliana@inf.puc-rio.br

Edward Hermann Haeusler

Departamento de Informática – Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)
Rua Marquês de São Vicente, 225 – Gávea -- 22.453-900 -- Rio de Janeiro – RJ

hermann@inf.puc-rio.br

Resumo

Confiança é um aspecto fundamental em sistemas distribuídos abertos de larga-escala. Ela está no núcleo de todas as interações entre as entidades que precisam operar em ambientes com muita incerteza e que se modificam constantemente. Dada essa complexidade, esses componentes, e o sistema resultante, são cada vez mais contextualizados, projetados e construídos usando técnicas baseadas em agentes. Portanto, confiança é fundamental em um sistema multi-agentes (MAS) aberto. Logo, este trabalho investiga como se ter um modelo de confiança explicitamente em um agente inteligente, que possui crenças (*Beliefs*), desejos (*Desires*) e intenções (*Intentions*), chamado de agente BDI. Ou seja, o agente passa a ter um quarto componente chamado confiança (*Trust*).

Palavras-Chaves: Agentes Inteligentes; Sistemas Multi-Agentes Abertos; BDI; Confiança; Inteligência Artificial.

Abstract

Trust is a fundamental concern in large-scale open distributed systems. It lies at the core of all interactions between the entities that have to operate in such uncertain and constantly changing environments. Given this complexity, these components, and the ensuing system, are increasingly being conceptualised, designed, and built using agent-based techniques. Therefore, the presence of trust is imperative in a multi-agent system (MAS). Consequently, this work studies how to have a explicit trust model in an intelligent agent, which has beliefs, desires and intentions (BDI agent). That is, the agent now has a fourth component called Trust.

Keywords: Intelligent Agents; Open Multi-Agent Systems; BDI; Trust; Artificial Intelligence.

1. INTRODUÇÃO

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 [1], web semântica [2], serviços da web [3], e-business [4], m-commerce [5, 6], computação independente (*autonomic computing*) [7], computação em grade (*grid computing*) [8] e ambientes de computação pervasivos [9]. 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 [10]. Dado isso, a computação baseada em agentes tem sido defendida como um modelo computacional natural para tais sistemas [11, 12].

Mais especificamente, sistemas distribuídos abertos podem ser modelados como sistemas multi-agentes (MAS, *Multi-Agent System*) [13] 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 [12]. 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 [12, 14]. 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 [15]. 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 diferentes (por exemplo, policiamentos, 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 ajudá-los 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, promovido pelo protocolo, seja feita de maneira tal de forma a prevenir que agentes manipulem os outros 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 [16]. 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 [17].

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. Conseqüentemente, em contextos práticos, é normalmente

impossível ao agente atingir um estado onde possui informações perfeitas sobre o ambiente e sobre as propriedades, possíveis estratégias e interesses dos parceiros com os quais interage [17, 18, 19]. 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.

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*) [11, 20, 21]. O modelo BDI é usado porque possui uma filosofia baseada no raciocínio prático de seres humanos e uma arquitetura de *software* implementável em sistemas reais [21].

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.

O resto do trabalho é organizado como a seguir. A próxima sessão define o conceito de confiança. A sessão a seguir explica porque confiança é considerado como sendo diferente de crença. A sessão 4 mostra como incluir confiança em sistema multi-agentes. Já a sessão 5 faz o mesmo para um agente BDI. Na sessão a seguir é mostrado como se implementar confiança em um agente BDI. Já na sessão 6 exhibe-se os trabalhos relacionados. Por fim, este artigo termina com as conclusões e os trabalhos futuros.

2. O QUE É CONFIANÇA

A comunicação entre agentes é uma questão muito importante em sistemas multi-agentes. Os agentes precisam interagir entre si para resolver um problema, trocar informações, etc. As atividades coletivas dentro do sistema dependem fortemente da boa funcionalidade das comunicações e podem falhar se as comunicações são, voluntariamente ou não, erradas. Algumas garantias tais como autenticação, integridade, confidencialidade, etc, podem ser obtidas pelo uso de técnicas de segurança. Entretanto, também há ameaças relacionadas à veracidade do conteúdo das mensagens. Se o sistema for aberto, que é o de interesse deste trabalho, agentes maliciosos podem ser introduzidos para mentir aos outros agentes e perturbar o bom funcionamento do sistema [22]. Ou então, podem passar informações errôneas, por outros motivos além da falta de honestidade. Alguns exemplos desse último caso seriam os seguintes:

- Um agente repassa uma mensagem que ele não verificara e que foi passada por um agente malicioso;
- Um agente não consegue perceber adequadamente o seu ambiente e passa uma informação que não corresponde à realidade do ambiente a um parceiro;
- Um erro é introduzido na mensagem durante o processo de transmissão, por causa de

problemas no meio de transmissão, ou

- O próprio agente comete um erro ao construir a mensagem.

Algumas vezes, o problema não está em erros na comunicação e sim na qualidade da mesma. Por exemplo, um agente pode ter que executar uma tarefa e levar mais tempo do que o razoável para a sua execução. Ou então, a qualidade da ação do agente ser abaixo do esperado. Por exemplo, em comércio eletrônico, um agente pode vender um produto a outro com uma qualidade inferior à esperada pelo comprador [12]. Também pode ser o caso de o agente encarregado de executar uma tarefa não informar ao que a solicitou o andamento da ação quando solicitado. Por exemplo, tem-se, em comércio eletrônico, um comprador querendo saber o status do seu pedido e o vendedor simplesmente não o responde.

Agentes com frequência precisam confiar uns nos outros quando engajados em ações conjuntas. Confiança é fundamental para se tomar decisões e compartilhar informações durante a comunicação entre agentes [23, 24, 25]. Pode ser considerada o elemento mais importante para que a comunicação entre agentes seja um sucesso. Sem confiança, as informações não serão compartilhadas. Mais ainda, confiança não é uma entidade estática e também não é simplesmente presente ou ausente [23].

A confiança é definida de várias maneiras diferentes [24] e possui diferentes significados dependendo da perspectiva tomada. Uma definição que se pode dar é *o desejo de se ficar vulnerável baseado em expectativas positivas sobre a ação dos outros* (em inglês, *the willingness to be vulnerable based on positive expectations about the action of others*). Isso quer dizer que um agente, ao interagir com outro, por não ter a certeza de como aquele irá agir, está se colocando em risco, embora esse seja um risco calculado quando se analisa o comportamento dos outros agentes. Afinal, essa interação pode lhe causar danos se o outro agente do qual se espera um comportamento adequado for malicioso ou não for capaz de agir adequadamente, por exemplo.

Na sociologia [24, 26], três tipos de confiança foram identificadas: disposicional, interpessoal e institucional. A primeira, a confiança disposicional ou baseada na personalidade, descreve a disposição geral de confiança daquele que confia. Ela foi definida como *um senso de confiança básico, o qual é uma atitude penetrante com respeito a si próprio e ao ambiente* [27]. Essa propensão à confiança é independente de qualquer grupo ou contexto.

O segundo tipo de confiança é a interpessoal, que é a confiança que um agente inteligente tem em outro agente inteligente. Ela é definida como: *o desejo de um agente de aumentar a sua vulnerabilidade com relação às ações de outro agente* [28]. Esse tipo de confiança é específica ao contexto e ao agente.

Por fim, há a confiança institucional, que se refere à confiança que é baseada nas propriedades percebidas ou na confiança no sistema ou instituição dentro da qual a confiança existe, ao invés das características do objeto de confiança. Ela pode ser definida como o sentimento de confiança de que a instituição vai criar e especificar leis e regulamentações para proteger os direitos individuais e não prejudicá-los. Dentro da confiança institucional, pode-se distinguir a confiança impessoal, ou confiança baseada em papéis, que é a confiança baseada na posição daquele que confia dentro da instituição, e não no agente que ocupa a posição. De maneira similar à confiança pessoal, a competência e a integridade são fatores importantes na confiança institucional.

Em muitos casos, nenhuma arquitetura de um time de agentes tem autoridade para assegurar que os agentes irão cooperar entre si. Confiança é necessário quando os agentes possuem valores potencialmente diferentes ou objetivos conflitantes. Os agentes devem possuir meios para decidir quando confiar uns nos outros ao se engajarem em ações conjuntas [29].

Como já dito, pode-se dizer que a confiança possui dois componentes distintos:

- Integridade: O agente é honesto em suas intenções e compromentimentos?
- Competência: O agente é capaz de cumprir os compromissos estabelecidos?

A competência de um agente em realizar ações é a sua habilidade em modificar de maneira inconstante o estado do ambiente compartilhado do agente, ou seja, atingir de maneira correta as suas intenções. A integridade é baseada em uma reputação de honestidade e em um compromisso contínuo de ser confiável, isto é, honrando um compromisso de realizar uma ação. Integridade e competência são distintas e devem ser tratados como componentes separados em decisões baseadas em modelos de confiança e reputação.

Em um sistema multi-agentes, um agente precisa:

- Observar o comportamento dos agentes em um ambiente e inferir conhecimento sobre suas competências e integridades, e
- Aplicar esse conhecimento para tomar decisões efetivas quando interagir com os agentes observados.

O ambiente onde os agentes se encontram pode ser classificado das seguintes formas:

- O ambiente pode incluir ou não um mecanismo de verificação de identificação;
- Pode haver ou não mecanismos de policiamento e coerção;
- Os agentes podem ter racionalidade limitada ou ilimitada;
- Pode haver ou não conluio entre os agentes, e
- Pode haver ou não um oráculo que avisa sobre a reputação de um agente.

A questão da confiança é ainda mais crítica em um sistema multi-agentes aberto. Muitos sistemas em computação em uma rede são desse tipo. Nesses sistemas os agentes podem entrar e sair livremente a qualquer momento e os mesmos pertencem a várias partes interessadas com diferentes metas e objetivos [30]. Dessas características, pode-se assumir que em um sistema multi-agentes aberto:

- os agentes tendem a serem egoístas e podem não ser confiáveis;
- nenhum agente pode saber tudo sobre o seu ambiente, e
- nenhuma autoridade central pode controlar os agentes.

Dadas tais incertezas, a confiança é de vital importância para efetivar interações entre os agentes [12, 30].

Como confiar em um novo agente do qual nada se sabe a respeito? Deve-se simplesmente não confiar nele? Ou cada agente que interage com ele deve lhe atribuir um determinado grau de confiança inicial e depois modificá-lo de acordo com as ações tomadas pelo novo agente?

3. CONFIANÇA X CRENÇA

Uma crença, como já foi definido, é uma representação interna de algo do ambiente. O agente captura uma percepção do ambiente e mapeia-a como uma crença. Essa crença não necessariamente corresponderá ao que de fato ocorre no ambiente, pois ele pode não ser capaz de percebê-lo com a precisão necessária.

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. 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. Ou seja, a confiança é uma espécie de aposta, que pode resultar em ganho ou em perda [31].

Obviamente, a confiança em um agente poderia simplesmente ser tratada como uma crença de que um determinado agente é confiável. Mas optou-se por colocá-la num nível mais abstrato, separado das demais componentes de um agente BDI. Mais ainda, se a confiança for colocada no nível de crença, dizer que se confia em um agente é dizer que se acredita que um agente seja confiável, tornando a definição circular. Outra diferença é que em muitas situações a confiança não é estabelecida por causa de um comportamento observado do outro agente, e sim através de suposições sobre o comportamento alheio, que podem estar corretas

ou não. Isso é verdade principalmente quando um agente tem que lidar com outro sobre o qual nada sabe (quando um novo agente entra no sistema, por exemplo).

Confiança, neste texto, será sempre entre agentes, embora pudesse se ter confiança em ações ou mensagens de outros agentes. Essa abordagem escolhida é devido a forma como o assunto é tratado na literatura de sistemas multi-agentes. Quando a confiança for de um agente em algo que não seja um outro agente ou com relação ao comportamento do mesmo, isso será simplesmente tratado como uma crença nesse algo se o agente não puder ser enganado. Se for algo com em relação ao comportamento de um outro agente, como o envio de uma mensagem, a crença passará a ter uma representação de que esse agente disse algo. Essa representação é semelhante ao que ocorre na lógica para protocolos de autenticação BAN [32]. Afinal, ele não pode ter a certeza de que a mensagem enviada é verdadeira. A representação de que um agente enviou uma mensagem é colocada como crença pois não deixa de ser uma percepção do ambiente onde está o agente.

Para finalizar, somente agentes cognitivos, tais como os agentes BDI, podem confiar em um outro agente, que não necessariamente precisa ser cognitivo. Isso porque apenas agentes com objetivos (intenções) e crenças podem confiar em outros.

4. CONFIANÇA EM UM SISTEMA MULTI-AGENTES

Basicamente, há duas abordagens principais para se inserir confiança em sistemas multi-agentes [12]. Primeiramente, para permitir que agentes confiem uns nos outros é necessário fornecer-lhes a habilidade de raciocinar sobre a natureza da reciprocidade, confiança e segurança de outros agentes. Essa habilidade é capturada por modelos de confiança. Esses modelos têm por objetivo calcular o nível de confiança que os agentes podem atribuir aos seus parceiros que com eles interagem. Um grau de confiança alto em um agente significa que é provável que ele seja escolhido para uma interação e (possivelmente), uma estratégia com reciprocidade será usada com ele através de múltiplas interações. Por outro lado, um grau de confiança baixo em um agente significa que ele não será selecionado (se outros parceiros mais confiáveis estiverem disponíveis para a interação) ou uma estratégia sem reciprocidade será adotada com relação ao mesmo através de múltiplas interações (se não houver alternativa melhor). Dessa maneira, os modelos de confiança têm por objetivo guiar a tomada de decisão do agente para decidir como, quando e com quem interagir. Pode-se conseguir isso de várias maneiras, incluindo inferências concluídas através do resultado de diversas interações com esses parceiros ou através de informações indiretas fornecidas por outros. O caso da interação direta torna necessária a consideração de métodos pelos quais os agentes podem aprender e desenvolver estratégias para lidar com agentes honestos e desonestos de forma a maximizar o lucro a longo prazo. O caso da interação indireta requer que os agentes sejam capazes de desenvolver métodos para conseguir e raciocinar sobre as informações conseguidas de outros agentes com segurança. Esta é a abordagem deste trabalho, desenvolver um modelo de confiança para agentes BDI.

Enquanto os modelos de confiança têm a ver com a habilidade dos agentes de raciocínio e coleta de informações, a segunda abordagem principal à confiança tem a ver com a criação de protocolos e mecanismos de interação (ou seja, regras de encontro). Esses mecanismos de interação devem ser planejados para garantir que os envolvidos possam ter a certeza de que eles sairão ganhando se eles o merecerem (por exemplo, um agente malicioso não pode adular a alocação de ganho correta do mecanismo). Portanto, espera-se que os agentes interajam usando um determinado mecanismo apenas se esse puder ser confiável. Isso torna necessário a existência de protocolos que garantam que os participantes não encontrarão outra alternativa a não ser serem honestos uns com os outros.

Como foi mostrado, a confiança está em todos os níveis das interações entre agentes. Com respeito a criação de agentes e sistemas multi-agentes abertos, contextualiza-se a confiança das seguintes maneiras:

- Confiança no nível individual:

Através da qual o agente possui informações sobre a honestidade ou da natureza recíproca de seus parceiros de interação, a de interesse do presente texto.

- **Confiança no nível de sistema:**
Através da qual os atores do sistema são forçados a serem confiáveis através de regras (protocolos ou mecanismos) que regulam o sistema.

5. CONFIANÇA EM UM AGENTE BDI

Os modelos de sistemas multi-agentes em geral são distribuídos [23]. Logo, a modelagem da confiança também é distribuída.

Um agente, ao detectar um comportamento de comunicação bom ou ruim, atualiza seu modelo de confiança e, quando ele precisa escolher entre interagir ou não com outro agente, ele utiliza esse modelo de confiança para tomar a sua decisão. A figura 1 mostra como se incluir confiança em um agente BDI genérico [20].

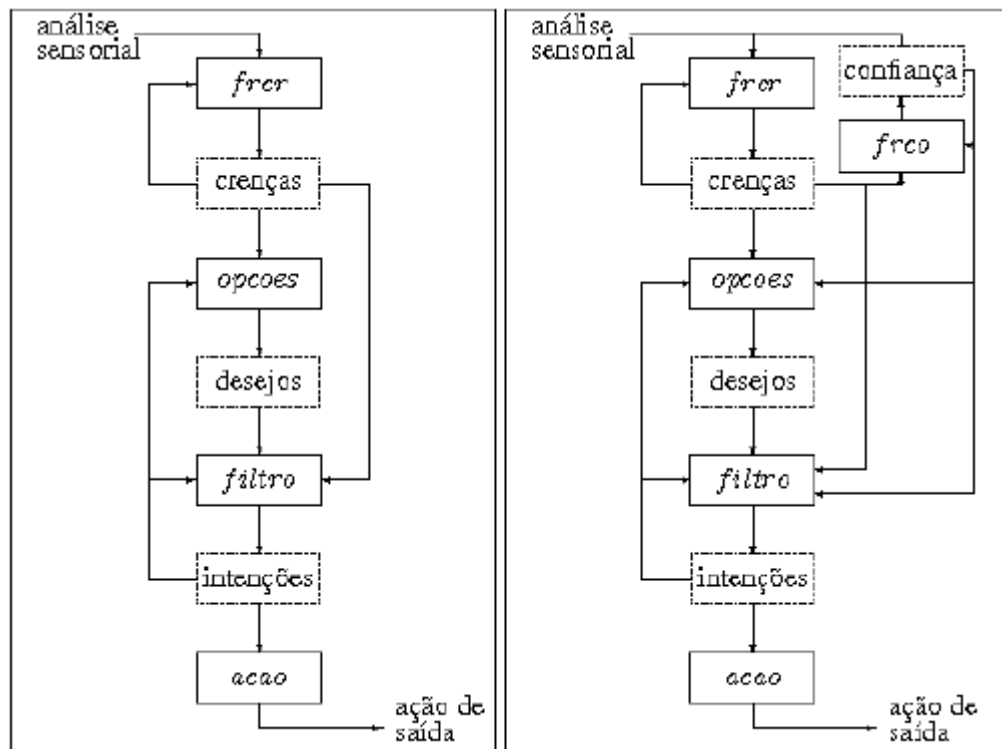


Figura 1 - Diagrama esquemático de uma arquitetura de crenças, desejos e intenções genérica sem e com confiança, respectivamente

O processo de raciocínio prático em um agente BDI sem confiança é resumido à esquerda. Como esta figura ilustra, há sete componentes principais em um agente BDI:

- Um conjunto de *crenças* atuais, representando as informações que o agente tem de seu ambiente atual;
- Uma *função de revisão de crenças* ($frer()$), que pega uma análise sensorial e as crenças atuais do agente, e, com base nelas, determina o novo conjunto de crenças;
- Uma *função de geração de opções* ($opcoes()$), que determina as opções disponíveis para o agente (seus desejos), com base nas suas crenças sobre o ambiente e suas intenções atuais;
- Um conjunto de *opções atuais* (desejos), representando as possíveis ações disponíveis ao agente;
- Uma função de *filtro* ($filtro()$), que representa o processo de deliberação do agente e que determina as intenções do agente baseando-se nas crenças, nos desejos e nas intenções atuais;

- Um conjunto de *intenções* atuais, representando o foco atual do agente, metas com as quais o agente se comprometeu em atingir, e
- Uma *função de seleção de ação* ($execucao()$), que determina uma ação a ser realizada com base nas intenções atuais.

Com a inserção da confiança e de uma função $frco()$ (função de revisão de confiança) para atualizar o modelo de confiança de um agente, as funções existentes devem ser modificadas, pois agora a atualização das crenças, dos desejos e das intenções dependem da confiança. Por sua vez, a atualização da confiança depende das crenças atuais do agente. As crenças dependem da confiança porque o agente vai atualizar suas crenças com o conteúdo de uma mensagem somente se o nível de confiança em seu parceiro for maior do que determinado nível. Já a atualização dos desejos e das intenções depende da confiança porque se eles dependerem de outros agentes para serem realizados, desejos e intenções podem vir a ser inatingíveis se o agente não tiver parceiros confiáveis. Quanto a atualização da confiança, ela depende das crenças porque a percepção do comportamento dos parceiros é que vai determinar como a confiança neles será atualizada.

A definição formal desses componentes é direta. Primeiramente, seja Bel o conjunto de todas as crenças possíveis; Des , o conjunto de todos os desejos possíveis; Int , o conjunto de todas as intenções possíveis, e Tru , o conjunto de todas as confianças possíveis do agente em questão em relação aos seus parceiros. Quaisquer que sejam os conteúdos desses conjuntos, é importante notar que eles devem ter alguma noção de consistência definida sobre eles, de forma que alguém possa responder, por exemplo, a questão: ter uma intenção de atingir x é consistente com a crença em y ? O estado de um agente BDI em um dado momento é, obviamente, uma tripla (B, D, I) , onde $B \subseteq Bel$, $D \subseteq Des$ e $I \subseteq Int$ no caso sem confiança. Adicionando-se a confiança, o estado do agente é uma tupla (B, T, D, I) , onde $T \subseteq Tru$.

Sem a confiança, as funções são definidas como mostrado a seguir, onde P é uma percepção e A , uma ação:

- $frcr: \wp(Bel) \times P \rightarrow \wp(Bel)$
- $opcoes: \wp(Bel) \times \wp(Int) \rightarrow \wp(Des)$
- $filtro: \wp(Bel) \times \wp(Des) \times \wp(Int) \rightarrow \wp(Int)$
- $execucao: \wp(Int) \rightarrow A$

$filtro()$ deve satisfazer a seguinte restrição:

$$\forall B \in \wp(Bel) \wedge \forall D \in \wp(Des) \wedge \forall I \in \wp(Int) \quad filtro(B, D, I) \subseteq D \cup I$$

Com a confiança, $frco()$ é definida e as demais funções com a exceção de $execucao()$ são redefinidas como mostrado a seguir, pois agora a atualização das crenças, dos desejos e das intenções dependem da confiança. Por sua vez, a atualização da confiança depende das crenças atuais do agente.

- $frcr: \wp(Bel) \times \wp(Tru) \times P \rightarrow \wp(Bel)$
- $opcoes: \wp(Bel) \times \wp(Tru) \times \wp(Int) \rightarrow \wp(Des)$
- $filtro: \wp(Bel) \times \wp(Tru) \times \wp(Des) \times \wp(Int) \rightarrow \wp(Int)$
- $fcro: \wp(Bel) \times \wp(Tru) \times P \rightarrow \wp(Tru)$

Agora, $filtro$ deve satisfazer a restrição:

$$\forall B \in \wp(Bel) \wedge \forall T \in \wp(Tru) \wedge \forall D \in \wp(Des) \wedge \forall I \in \wp(Int) \quad filtro(B, D, I) \subseteq D \cup I$$

A função de decisão de um agente BDI $acao()$ é a seguinte para agentes com ou sem confiança:

$$acao: P \rightarrow A$$

Ela é definida pelo pseudo-código na figura 2.

<ol style="list-style-type: none"> 1. função $acao(p:P) : A$ 2. início 3. $B := frc(B, P)$ 4. $D := opcoes(B, I)$ 5. $I := filtro(B, D, I)$ 6. retorne $execução(I)$ 7. fim_acao 	<ol style="list-style-type: none"> 1. função $acao(p:P) : A$ 2. início 3. $B := frc(B, T, P)$ 4. $T := frco(B, T)$ 5. $D := opcoes(B, T, I)$ 6. $I := filtro(B, T, D, I)$ 7. retorne $execução(I)$ 8. fim_acao
--	--

Figura 2 - Função $acao()$ de um agente BDI sem e com confiança, respectivamente

Refina-se agora as funções $acao()$ acima com a inserção da noção de plano. Os planos e as intenções estão intimamente ligados. Frequentemente se usa os dois termos no dia a dia como se tivessem o mesmo significado. Porém, neste trabalho, planos se referem a *receitas* para se atingir intenções. Agora, precisa-se definir a função $plano()$ no agente sem confiança:

$plano: \wp(Bel) \times \wp(Int) \rightarrow Plan$

onde $Plan$ é o conjunto de planos. Quando se insere a confiança no agente BDI, tem-se que:

$plano: \wp(Bel) \times \wp(Tru) \times \wp(Int) \rightarrow Plan$

Afinal, se o agente precisar de outro para atingir suas intenções, na elaboração do seu plano ele deve levar em conta a confiança nos seus parceiros para escolher com quem interagir.

Pode-se, agora, reescrever os algoritmos para a função $acao()$ descritos na figura 2 usando a definição de plano. Note-se que a função $execucao()$ agora executa um plano e não mais uma intenção.

<ol style="list-style-type: none"> 1. função $acao(p:P) : A$ 2. início 3. $B := frc(B, P)$ 4. $D := opcoes(B, I)$ 5. $I := filtro(B, D, I)$ 6. $\pi := plano(B, I)$ 7. retorne $execução(\pi)$ 8. fim_acao 	<ol style="list-style-type: none"> 1. função $acao(p:P) : A$ 2. início 3. $B := frc(B, T, P)$ 4. $T := frco(B, T)$ 5. $D := opcoes(B, T, I)$ 6. $I := filtro(B, T, D, I)$ 7. $\pi := plano(B, T, I)$ 8. retorne $execução(\pi)$ 9. fim_acao
--	---

Figura 3 - Função $acao()$ de um agente BDI com planos sem e com confiança, respectivamente

Ao invés de se usar a função $acao()$ explicitamente, os próximos algoritmos a transformam em um *loop* de controle do agente. Então, os algoritmos acima são equivalentes aos da figura 3, onde B_0 , T_0 e I_0 são as crenças, a confiança e as intenções iniciais, respectivamente. Estes *loops* de controle mostram um comprometimento cego dos agentes.

<ol style="list-style-type: none"> 1. $B := B_0$ 2. $I := I_0$ 3. enquanto verdadeiro faça 4. pegue a proxima percepcao p 5. $B := frc(B, P)$ 6. $D := opcoes(B, I)$ 7. $I := filtro(B, D, I)$ 8. $\pi := plano(B, I)$ 9. retorne $execução(\pi)$ 10. fim_enquanto

```

1.  B := B0
2.  T := T0
3.  I := I0
4.  enquanto verdadeiro faça
5.    pegue a proxima percepcao p
6.    B := frc(B, T, P)
7.    T := frco(B, T)
8.    D := opcoes(B, T, I)
9.    I := filtro(B, T, D, I)
10.  π := plano(B, T, I)
11.  retorne execucao(π)
12.  fim_enquanto

```

Figura 4 - Loop de controle de um agente BDI com planos sem e com confiança, respectivamente

Outras versões de *loops* presentes em [21] também podem ser modificadas para a inclusão da confiança.

6. IMPLEMENTAÇÃO DA CONFIANÇA

Há basicamente quatro maneiras de se implementar a confiança que um determinado agente tem em outro:

- Na primeira opção, diz-se que a confiança de um agente é binária, ou seja, ele simplesmente confia ou não confia em um outro agente. Isso pode ser implementado de maneira bem simples, adicionando-se ao banco de dados do agente somente as referências aos agentes confiáveis. Mas esse modelo não é realista; como já dito anteriormente, ao se afirmar que a confiança não é simplesmente presente ou ausente. Dessa maneira, como qualificar um agente que acabou de entrar no sistema? Não seria razoável dizer que ele é simplesmente confiável. Nem simplesmente não-confiável, já que ele não realizou nenhuma ação que pudesse classificá-lo como tal. Mais ainda, na vida real, não simplesmente se confia ou se desconfia de outras pessoas. Há graus de confiança; há aqueles nos quais se confia mais e aqueles nos quais se confia menos. Porém, pode-se alegar que a confiança é sim binária e, que quando se interage com um determinado agente, naquele momento se está confiando nele. O mesmo pode-se dizer quando não se interage com um outro agente por achar que vai se ter perda ao invés de ganho com a interação, ou seja, simplesmente não se confia.
- Na segunda opção, a confiança de um agente em outro não é mais binária, mas sim há níveis fixos de confiança dependendo do modelo a ser implementado. Por exemplo, um modelo com dois níveis teria os níveis 0 e 1 (ou -1 e 1 [12]), que é o caso binário; com três níveis: 0, 0.5 e 1 (ou -1, 0 e 1 [12]), e assim sucessivamente. Com n níveis: 0, $1/(n-1)$, $2/(n-1)$, ..., $(n-2)/(n-1)$, 1. Se por um lado essa abordagem é bem mais flexível do que a anterior, por outro lado ela ainda é muito rígida quanto aos níveis de confiança, por só permitir um número fixo de níveis pré-determinado.
- Tem-se também a opção onde o nível de confiança é simplesmente um valor entre 0 e 1 (ou -1 e 1 [12]). Essa é uma situação mais flexível. Porém, ela lida com um valor global de confiança.
- Finalmente, pode-se ter vários graus de confiança para vários aspectos do que se espera de um agente, como integridade e competência, onde esse último pode ser dividido em diferentes capacidades que o agente pode ter. Afinal, um agente que executa bem determinada tarefa pode não ser capaz de executar uma outra adequadamente. Esses aspectos podem ser em maior ou menor número dependendo do sistema. Já cada um desses graus poderia ser implementado com uma das três maneiras citadas acima [12].

É possível se ter mais de um tipo de modelo de confiança implementado em um sistema. Nesses casos, cada tipo seria usado para determinadas maneiras de interação entre os

agentes. Por exemplo, se um sistema tiver dois tipos de interação entre os agentes e se para determinado tipo de interação apenas uma capacidade (a honestidade, por exemplo) estiver envolvida, não há razão para se usar um modelo parametrizado. Se para a outra mais de uma capacidade estiver em questão, essa pode ter como modelo uma confiança parametrizada. Porém, apenas um modelo deve ser usado para um tipo de interação. Por exemplo, não se deve usar a confiança binária se em algum momento for necessário falar da confiança parametrizada para interagir com um agente. Nesse caso, sempre deve ser usado um modelo específico de confiança parametrizada.

O grau de confiança é atualizado toda a vez que um agente recebe uma mensagem vinda de outro agente. O fato de que um agente mandou uma mensagem é considerado como uma percepção do ambiente, a qual é utilizada para atualizar as crenças do agente que recebeu a mensagem. A atualização das crenças obviamente depende da confiança no agente que enviou a mensagem, pois o conteúdo de uma mensagem de uma fonte não-confiável não deve ser levado em conta para a atualização das crenças. A atualização das crenças com o conteúdo da mensagem só deve ocorrer se o grau de confiança na fonte for maior do que determinado valor, dependendo da aplicação e de que informação estiver sendo passada. Com essa atualização das crenças, dependendo da coerência da informação passada com as crenças do agente e de acordo com o seu modelo de confiança, ele irá atualizar o seu modelo de confiança, aumentando ou diminuindo o grau de confiança com relação ao agente que enviou a mensagem e/ou com relação a outros agentes, se a mensagem for em referência ao comportamento de um outro agente. Por exemplo, um agente pode querer avisar outro de que um terceiro agente está se comportando de forma maliciosa, informação que pode ser verdadeira ou não.

O agente sempre deve levar em consideração a confiança que ele tem nos demais para decidir com quem vai interagir, preferindo um agente honesto e com a capacidade necessária para executar a tarefa desejada, se o modelo de confiança tiver diferentes graus para diferentes capacidades. Ao delegar uma tarefa para outro agente, ele espera que o outro tenha a intenção de executá-la, isso é, que seja benevolente.

Quando um agente decide interagir com outros para atingir uma intenção, algumas ações de seu plano devem incluir mensagens para outros, algumas com tarefas a serem delegadas. Para que a intenção seja atingida, os outros agentes devem fazer o que lhes foi pedido com a qualidade desejada. O agente, no caso geral, precisa verificar se o seu parceiro fez o que lhe fora pedido e da maneira desejada. Essa verificação pode ser parte do plano para atingir uma determinada intenção. Ou seja, determinadas ações do plano devem ser verificações para ver se a atividade delegada fora concluída adequadamente. Ou então, o registro da delegação pode ficar registrado na própria intenção e o agente, ao verificar se ela já foi atingida, se ainda é possível ser atingida ou se ainda está motivado para atingir a intenção, verifica se a atividade delegada foi concluída da maneira desejada. Dependendo do que o outro agente fizer, o plano pode ser descartado e até mesmo a intenção ser descartada, se não tiver um terceiro agente que possa ajudá-lo a atingir a intenção desejada.

Adicionalmente, deve-se registrar a mensagem enviada e o pedido de delegação nas crenças, pois não deixa de ser uma modificação no ambiente. Mais ainda, essa informação pode ser útil para a tomada de outras decisões e a atualização da confiança. Por exemplo, se algo foi pedido a um outro agente e não foi respondido ou a tarefa solicitada não foi feita de maneira adequada, o grau de confiança no outro agente deve ser atualizado de modo negativo.

Finalizando, na maioria dos casos, a confiança vai ser atualizada apenas quando um agente interagir com outros agentes. Ela pode ser atualizada com relação ao agente com o qual se está interagindo no momento ou sobre outro agente, se a intenção for saber informações sobre a reputação de terceiros. Todavia, em casos especiais, é possível que um agente também consiga saber informações sobre outros agentes analisando o sistema ou as interações dos outros.

7. TRABALHOS RELACIONADOS

Neste trabalho considera-se confiança e crença como sendo dois conceitos distintos. Porém, em [12, 31], 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 (para o bem de ambos), dada uma oportunidade de trapacear para conseguir maior lucro. 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 [31], 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 [33].

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 em [33], essa consulta é feita à organização 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 ele tenha 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 ou 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 lhe informa.

Por fim, em [21], explica-se o que é um agente BDI, detalhando cada um de seus componentes e as diferentes implementações para um agente inteligente.

8. CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho apresentou-se como se ter confiança em um agente cognitivo com crenças, desejos e intenções explicitamente, tornando tal componente mais um estado mental para um agente inteligente.

Primeiramente, enfatizou-se o porquê de se ter um modelo de confiança em um agente que interage com outros em um sistema multi-agentes aberto e por que a confiança difere da crença.

Para se colocar confiança no modelo, inicialmente se acrescentou a confiança à definição de um agente BDI. Posteriormente, colocou-se como a confiança interage com os outros componentes do agente durante o funcionamento de seu *loop* de controle em suas várias versões.

Uma modificação na abordagem é que, da mesma forma que se criou um estado mental para confiança, talvez fosse interessante criar um estado mental sobre o que os agentes dizem para os outros, já que tal tipo de informação é encontrada em muita quantidade nas crenças do agente. Afinal, os agentes, ao interagir, trocam muitas mensagens. Esse estado mental teria todas as trocas de informação que o agente consegue capturar do ambiente e que lhe são relevantes. Afinal, tanto no caso das crenças quanto das mensagens trocadas, ele não precisa armazenar informações que não lhe interessam.

Outro estado mental que poderia ser adicionado ao agente é o conhecimento [21]. Conhecimento é diferente de crença porque uma vez que um agente sabe algo, ele não deixa de saber. Seria interessante estabelecer relações entre as crenças e o conhecimento, e os demais estados mentais do agente, incluindo a confiança.

Seria interessante também a modelagem de fatos, que são verdades, os quais poderiam ser sobre propriedades do ambiente, ações e comportamento de agentes, por exemplo. Isso permitiria modelar precisamente quando um agente está mentindo, falando a verdade, cometendo omissões em suas mensagens, etc.

Mais ainda, assim como a confiança é algo probabilístico no caso não-binário, poderia se fazer o mesmo com as crenças. Ou seja, um agente, ao invés de simplesmente colocar uma representação do que ele percebe do ambiente em suas crenças, ele poderia atribuir-lhe um valor probabilístico, que significaria a chance de ele estar tendo uma percepção correta do que foi analisado. Afinal, nem sempre um agente, assim como os seres humanos, conseguem perceber o ambiente corretamente e, em alguns casos, poderiam ter consciência de suas limitações. Resumindo, as crenças poderiam ter graus variados de “achismos”, que é o que o agente “acha” que está percebendo do seu ambiente. Note-se que isso não é o mesmo do que confiar em algo do ambiente, já que supõe-se que apenas a percepção do agente em algo do seu ambiente esteja em questão. Ou seja, o agente não está apostando que o ambiente vai se comportar de uma determinada forma e sim está fazendo suposições com maior ou menor grau de certeza sobre o estado atual do ambiente.

9. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] ORAM, A. **Peer-to-Peer: Harnessing the Power of Disruptive Technologies**. O'Reilly & Associates, Inc, Sebastopol, CA, 2001.
- [2] BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. **The semantic web**. Scientific American, 284(5):34--43, May 2001.
- [3] MCILRAITH, S. A.; SON, T. C.; ZENG, H. **Semantic web services**. IEEE Intelligent Systems, 16(2):46--53, 2001.
- [4] DEITEL, M. H; DEITEL, P. J.; NIETO, T. R. **e-business and e-commerce: how to program**. Prentice Hall, NJ, USA, 2001.
- [5] SADEH, N. **M-Commerce: technologies, services and business models**. Wiley computer publishing, 2002.
- [6] VULKAN, N. **Economic implications of agent technology and e-commerce**. The Economic Journal, 109(453):67--90, 1999.
- [7] KEPHART, J. O.; CHESS, D. M. **The vision of autonomic computing**. IEEE Computer, 36(1):41--50, 2003.
- [8] FOSTER, I.; KESSELMAN C., Editors. **The Grid, blueprint for a new computing infrastructure**. Morgan Kaufmann Inc, 1998.
- [9] SCHMECK, H.; UNGERER, T.; WOLF, L. C., Editors. In: TRENDS IN NETWORK AND PERVASIVE COMPUTING - ARCS 2002, INTERNATIONAL CONFERENCE ON ARCHITECTURE OF COMPUTING SYSTEMS, LECTURE NOTES IN COMPUTER SCIENCE, VOLUME 2229. Springer, April 2002.
- [10] SIMON, H. A. **The sciences of the artificial**. MIT Press, third edition, 1996.
- [11] JENNINGS, N. R. **An agent-based approach for building complex software systems**. Communications of the ACM, 44(4):35--41, 2001.
- [12] RAMCHURA, S. D.; HUYNH, T. D.; JENNINGS, N. R. **Trust in multi-agent systems**. The Knowledge Engineering Review, 19(1), March 2004.
- [13] WOOLDRIDGE, M. **An introduction to multiAgent systems**. Wiley, Liverpool, UK, 2002.
- [14] VON NEUMAN, J.; MORGENSTERN, O. **The theory of games and economic behaviour**. Princeton University Press, NJ, USA, 1944.
- [15] **Sistemas multiagentes**. URL: http://web.teccomm.les.inf.puc-rio.br/index.php/Sistemas_Multiagentes, 2007.
- [16] SAVAGE, L. **The foundations of statics**. John Wiley and Sons, 1954.
- [17] BINMORE, K. **Fun and games: a text on game theory**. D. C. Health and Company, 1992.
- [18] RUSSEL, S.; NORVIG, P. N. **Artificial intelligence: a modern approach**. Prentice-Hall, 1995.
- [19] AXELROD, R. **The evolution of cooperation**. Basic Books, New York, USA, 1984.
- [20] WOOLDRIDGE, M. **Intelligent agents**. In: Weiss, G., editor, MULTIAGENT SYSTEMS - A MODERN APPROACH TO DISTRIBUTED ARTIFICIAL INTELLIGENCE. The MIT Press, April 1999.
- [21] WOOLDRIDGE, M. **Reasoning about rational agents**. The MIT Press, Cambridge, Massachusetts, 2000.

- [22] MÜLLER, J. P.; PISCHEL, M.; THIEL, M. **Modelling reactive behaviour in vertically layered agent architectures.** In: Wooldridge, M.; Jennings, N. R., Editors, INTELLIGENT AGENTS: THEORIES, ARCHITECTURES AND LANGUAGES (LNAI VOLUME 890), p. 261--276. Springer-Verlag, Berlin, Germany, January 1995.
- [23] DIGNUM, V.; VAN EIJK, R. M. **Towards a model to understand the influence of trust in knowledge sharing decisions.** In: PROCEEDINGS OF TRUST IN AGENT SOCIETIES (TRUST 2005), 2005.
- [24] FORD, D. **Trust and knowledge management: the seeds of success.** In: Holsapple, C., Editor, HANDBOOK ON KNOWLEDGE MANAGEMENT 1, INTERNATIONAL HANDBOOKS ON INFORMATION, p. 553--576. Springer, 2003.
- [25] ROLLAND, N.; CHAUVEL, D. **Knowledge transfer in strategic alliances.** In: Despres, C.; Chauvel, D., Editors, KNOWLEDGE HORIZONS: THE PRESENT AND PROMISE OF KNOWLEDGE MANAGEMENT, INTERNATIONAL HANDBOOKS ON INFORMATION SYSTEMS, p. 225--236. Butterworth Heineman, 2000.
- [26] GAMBETTA, D. **Can we trust trust?** In: Gambetta, D., Editor, TRUST: MAKING AND BREAKING COOPERATIVE RELATIONS, chapter 13, p. 213--237. Department of Sociology, University of Oxford, 2000.
- [27] MCKNIGHT, D.; CHERVANY, N. **The meanings of trust.** Technical report, Carlson School of Management, University of Minnesota, Minnesota, USA, 1996.
- [28] MAYER, R.; DAVIS, J.; SCHOORMAN, F. **An integrative model of organizational trust.** Academy of Management Review, 20:709--734, 1995.
- [29] SMITH, M. J.; DESJARDINS, M. **A model for competence and integrity in variable payoff games.** In: PROCEEDINGS OF TRUST IN AGENT SOCIETIES (TRUST 2005), 2005.
- [30] HUYNH, T. D.; JENNINGS, N. R.; SHADBOLT, N. R. **On handling inaccurate witness reports.** In: PROCEEDINGS OF TRUST IN AGENT SOCIETIES (TRUST 2005), 2005.
- [31] CASTELFRANCHI, C.; FALCONE, R. **Principles of trust for MAS: cognitive anatomy, social importance, and quantification.** In: PROCEEDINGS OF THE INTERNATIONAL CONFERENCE OF MULTI-AGENT SYSTEMS (ICMAS'98), p. 72 -- 79, 1998.
- [32] BURROWS, M.; ABADI, M.; NEEDHAM, R. **A logic of authentication.** PROCEEDINGS OF THE ROYAL SOCIETY, 1989.
- [33] GUEDES, J. S. P.; DA SILVA, V. T.; DE LUCENA, C. L. P. **A reputation model based on testimonies.** In: PROCEEDINGS OF THE AOIS2006@CAISE WORKSHOP, Luxembourg, Grand-Duchy of Luxembourg, June 2006.