

2 Trabalhos Relacionados

Neste capítulo são apresentados alguns trabalhos relacionados com modelos de confiança e reputação. Tais modelos fazem parte do próprio agente de software, de forma descentralizada, ou então fazem parte da estrutura do sistema multi-agente, de maneira que todos os agentes do ambiente possam solicitar as reputações de outros agentes. Por isso os trabalhos relacionados estão divididos em mecanismos de reputação centralizados e descentralizados.

Os exemplos utilizados ao longo deste trabalho estão baseados em uma interação dentro de um sistema multi-agente aberto de importação e exportação. Nesta interação agentes importadores enviam mensagens aos agentes exportadores solicitando propostas de compras de mercadorias. Os agentes exportadores interessados enviam suas propostas. Os agentes importadores utilizam mecanismos de confiança e reputação para avaliar a confiabilidade dos agentes exportadores e selecionar a melhor proposta.

2.1. Modelos de Reputação Centralizados

Os mecanismos de reputação centralizados mais famosos são aqueles utilizados no comércio ou leilão eletrônico, tais como Amazon Auctions [2] e eBay. São exemplos de sistemas bem sucedidos, centralizados, que colecionam e agregam avaliações das negociações anteriores. Nestes casos, o sistema centralizado é que calcula a reputação baseando-se na informação que recebe sobre as negociações ocorridas.

Os mecanismos de reputação utilizados no eBay e Amazon Auctions têm a intenção de informar ao comprador a atuação do vendedor em negociações anteriores. No sistema de reputação do eBay, denominado feedback fórum, as avaliações são feitas pelos participantes após uma negociação, e não por agentes de software. A reputação dos participantes é conhecida por todos os usuários, que é uma vantagem, e é representada por uma pontuação global, que corresponde à

soma de todas as avaliações, positivas, neutras ou negativas. Além destas avaliações são incluídos também comentários textuais. Esses textos são fundamentais nestes sistemas porque como a reputação está representada pela soma de avaliações positivas, neutras e negativas, não é possível saber o motivo pelo qual o comprador ficou ou não satisfeito. Um comprador insatisfeito insere comentários e desta forma é possível saber a gravidade do problema que aconteceu na transação. Apesar desta abordagem atender bem as necessidades do eBay, em alguns sistemas multi-agentes é necessário que as avaliações sejam feitas, e analisadas, por agentes de software e não por seres humanos capazes de interpretar o texto que descreve detalhes do que ocorreu nas negociações anteriores.

Segue abaixo algumas vantagens e desvantagens dos mecanismos de reputação centralizados:

Vantagens dos modelos centralizados:

- (i) A reputação de um agente é sempre conhecida.
- (ii) A reputação de um agente está sempre atualizada porque é calculada considerando todas as avaliações ou testemunhos enviados pelos agentes que habitam o ambiente.

Desvantagens dos modelos centralizados:

- (i) Agentes não são capazes de armazenar as reputações com quem interagiram, por isso precisam sempre utilizar as informações armazenadas nos sistemas de reputação centralizados.
- (ii) Tais mecanismos simplesmente reúnem todas as avaliações sobre o mesmo agente. Os agentes, que são pessoas, podem mentir e não existe nenhuma forma de evitar avaliações falsas. Além disso, os agentes podem inclusive não enviar testemunhos sobre o comportamento de outros agentes.
- (iii) Estes mecanismos oferecem uma única reputação global e, portanto, não consideram contextos diferentes. Um agente pode ter boa reputação como vendedor, mas não garante que seja um bom comprador, por exemplo.

2.2. Modelos de Reputação Descentralizados

Os mecanismos de reputação descentralizados permitem que o próprio agente de software calcule a reputação, baseado nas interações diretas ou então consultando outros agentes caso não tenham informação suficiente disponível.

Mecanismos baseados nos modelos FIRE, REGRET, [1] e [31,32] são exemplos de mecanismos de reputação descentralizados. Os modelos têm características semelhantes; todos buscam estimar a confiança e a reputação utilizando diferentes fontes de informação.

Em uma tentativa de padronizar as diferentes denominações utilizadas em mecanismos semelhantes, criamos os termos: Modelo Descentralizado Simples, Modelo Descentralizado Baseado em Testemunhas e Modelo Descentralizado Baseado em Reputação Certificada.

2.2.1. Modelo Descentralizado Simples

O modelo descentralizado simples modela a confiança entre dois agentes e representa o mecanismo básico de reputação descentralizada, pois está presente em todos os modelos descentralizados analisados. O modelo descentralizado simples capacita o agente a avaliar as interações, calcular e armazenar as reputações dos agentes com quem interagiu. A figura 1 representa o funcionamento deste modelo. Um agente A após interagir com o agente B faz uma avaliação e atualiza a reputação deste agente, armazenando em seu próprio repositório. O mesmo ocorre com o agente B. Desta forma, através das interações diretas, os agentes podem saber se os outros agentes com quem interagiram são confiáveis ou não. O problema desta abordagem acontece quando temos, por exemplo, um agente C que nunca interagiu com o agente A e então não tem como saber se ele é confiável ou não.

O sistema REGRET é dividido em três dimensões denominadas: (i) *Individual Dimension*, que modela as interações diretas entre agentes através da reputação subjetiva *Direct Trust*, que é calculada utilizando um conjunto de impressões que um agente tem sobre outro, com base nos resultados de cada item

do contrato, (ii) *Social Dimension*, que utiliza testemunhos enviados por outros agentes através da *Witness Reputation*, onde a reputação é construída pelo ponto de vista de alguns agentes da sociedade e *System Reputation*, que atribui reputações iniciais aos agentes de acordo com o conjunto de impressões que um agente tem sobre outros agentes enquanto desempenhavam os mesmos papéis dentro da mesma instituição e (iii) *Ontological Dimension*, que permite combinar os diferentes tipos de reputação, aqueles relacionados aos itens do contrato, do ponto de vista de um determinado agente.

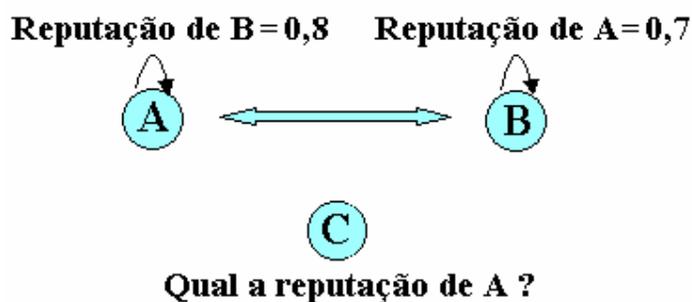


Figura 1. Modelo Descentralizado Simples

O modelo descentralizado simples dentro do sistema REGRET equivale à reputação denominada *Direct Trust* e calcula uma reputação subjetiva obtida através de um conjunto de impressões que um agente tem sobre outro. Uma impressão equivale à avaliação de um item de um contrato. A qualidade do produto, o preço e o prazo de entrega são exemplos de itens de um contrato. A avaliação de um item é feita comparando valores esperados com os valores obtidos após o cumprimento do contrato.

FIRE Model é composto por quatro tipos de reputação denominados: (i) *Interaction Trust*, que modela as interações diretas entre os agentes, (ii) *Role-Based Trust*, que atribui reputações iniciais aos agentes com base nas avaliações diretas anteriores sobre outros agentes ao desempenhar os mesmos papéis, (iii) *Witness Reputation*, que é uma reputação construída baseada nas avaliações enviadas por outros agentes, chamados de testemunhas e (iv) *Certified Reputation*, que é a reputação certificada, calculada com base nas referências fornecidas pelo próprio agente com quem deseja interagir.

O modelo descentralizado simples dentro do modelo FIRE equivale à reputação denominada *Interaction Trust*, que propõe utilizar o mecanismo *Direct Trust* do REGRET, adicionando uma função no cálculo da reputação para garantir que as avaliações mais recentes tenham um peso maior do que avaliações mais antigas.

Em [1] foi identificado um conjunto de características sociais relevantes ao seu modelo de confiança e reputação. Entre elas estão a subjetividade, a dependência de contexto e o dinamismo, uma vez que diferentes agentes têm diferentes percepções sobre a probidade de um determinado agente. A probidade varia de acordo com o contexto e a mesma probidade pode aumentar ou diminuir a medida em que acontecem mais experiências diretas ou recomendações de outros agentes. O modelo é composto por dois tipos de reputação, *Direct Trust*, que corresponde ao modelo descentralizado simples, e *Recommender Trust*. O primeiro tipo modela as experiências diretas e o segundo tipo de reputação modela as recomendações de outros agentes e é utilizada principalmente quando um agente deseja interagir com um agente desconhecido.

O modelo de reputação proposto em [31,32] utiliza a teoria de evidência de *Dempster-Shafer* para saber se um agente é digno de confiança. Também utiliza como fonte de informação as experiências diretas, e na falta delas utiliza testemunhas que já interagiram anteriormente com o agente correspondente. O modelo avalia as informações enviadas pelas testemunhas com objetivo de detectar fraudes.

Para exemplificar o modelo descentralizado simples vamos considerar a interação entre agentes importadores e exportadores. Após cada negociação o agente importador avalia o desempenho do agente exportador. O desempenho é avaliado pelo agente importador considerando as variáveis mais importantes. Esta avaliação é subjetiva, pois cada proprietário programa o seu agente de acordo com a sua estratégia. A avaliação pode estar considerando, por exemplo, se o agente exportador cumpriu o prazo de entrega e se entregou a quantidade solicitada pelo agente importador. Desta forma se o agente importador já interagiu com um agente exportador ele vai considerar esta experiência anterior para considerá-lo confiável ou não.

Vantagens do uso do modelo descentralizado simples:

- (i) A descentralização do mecanismo de reputação ao nível de agente viabiliza seu uso em sistemas multi-agentes de larga escala. Os agentes não precisam acessar uma fonte centralizadora das reputações.
- (ii) O agente constrói sua opinião sobre o comportamento dos agentes com quem interagiu. A descentralização permite que cada proprietário programe seu agente de software com estratégias particulares para avaliação das interações, considerando as variáveis mais importantes segundo seus interesses.

Desvantagens do uso do modelo descentralizado simples:

- (i) Somente é possível saber se um agente é confiável se o agente interessado já tiver interagido com ele anteriormente. Este problema se agrava particularmente em sistemas multi-agentes de larga escala porque pode ser difícil ocorrer interações repetidas entre os mesmos agentes, dificultando a avaliação da reputação do agente de forma consistente.
- (ii) A reputação que um agente conhece de outro agente pode estar desatualizada, pois os agentes podem não ter interagido recentemente. A única forma de saber se o agente continua se comportando de maneira similar a anterior (refletida na reputação armazenada) é interagindo com o agente.

2.2.2.**Modelo Descentralizado Baseado em Testemunhas**

No modelo descentralizado simples, só é possível saber se um agente é confiável interagindo com ele. Para resolver este problema, foi introduzida uma abordagem baseada em testemunhos, na qual agentes de software solicitam a reputação de um determinado agente para uma testemunha que já interagiu anteriormente com ele. A vantagem desta abordagem está na possibilidade de saber a reputação de um agente sem ter interagido com ele no passado. Desta forma, um agente C, que pretende interagir com o agente A, mas não sabe se ele é confiável, pode solicitar a reputação de A para outro agente (agente B), que já

interagiu anteriormente com ele, conforme exemplo da figura 2. O problema desta abordagem é como descobrir quem já interagiu com o agente A, em um universo com muitos agentes.

O sistema REGRET também tem em sua estrutura um modelo descentralizado baseado em testemunhas, chamado de *witness reputation*. Se um agente deseja contratar outro agente, mas nunca interagiu com ele anteriormente, então ele pode construir a reputação coletando informações de um grupo de testemunhas que já interagiram com ele e podem repassar suas impressões. Este mecanismo é chamado de *witness reputation*. As informações coletadas nada mais são do que as reputações subjetivas (*direct trust*) calculadas por outros agentes da sociedade em relação ao agente alvo, de quem se quer obter informações com base no seu desempenho em negociações anteriores.

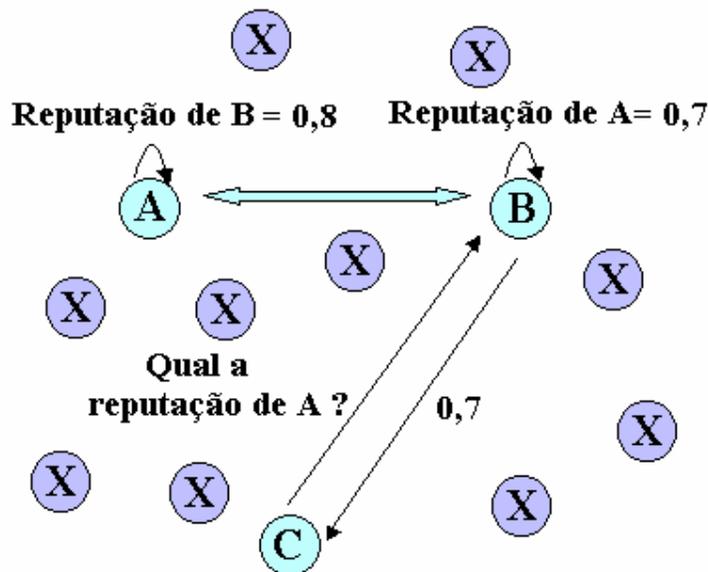


Figura 2. Modelo Descentralizado baseado em testemunhas

O mecanismo *witness reputation* apresenta alguns problemas. As testemunhas são agentes de software que podem mentir, i.e., podem fornecer informações incorretas. Neste caso as testemunhas poderiam fornecer uma reputação que não corresponde com o comportamento real do agente. Por exemplo, um agente importador que pretende interagir com um agente exportador desconhecido poderá utilizar o mecanismo *witness reputation* para obter alguma informação sobre o comportamento anterior do agente com quem pretende interagir. Neste caso as testemunhas são outros agentes importadores. Ao

consultá-la, a testemunha poderá fornecer propositalmente uma reputação incorreta com o objetivo de prejudicar um concorrente ou então para favorecer um exportador que é seu parceiro. Além disso, as testemunhas podem simplesmente não concordar em colaborar, optar por não repassar informações para um agente rival, i.e., um agente concorrente, que oferece o mesmo serviço.

Funcionamento do mecanismo Witness Reputation do Sistema REGRET

No REGRET, inicialmente, o grupo de testemunhas selecionado corresponde ao conjunto de agentes que já interagiram com o agente alvo. Porém este grupo pode ser muito grande então REGRET utiliza uma heurística baseada em [12] para selecionar somente os agentes considerados mais representativos com o objetivo de reduzir o número de solicitações.

Para identificar as testemunhas mais representativas, REGRET utiliza um sociograma (grafo) representando o relacionamento cooperativo entre todos os agentes que já interagiram com o agente alvo. A heurística utilizada por REGRET é a seguinte:

1. Identificar os componentes do grafo. Um componente está definido como um subgrafo conexo maximal;
2. Achar os pontos de articulação de cada componente. Um ponto de articulação (*cut-point*) é um nó cuja eliminação desconectaria o subgrafo, aumentando o número de componentes. Um ponto de articulação pode ser visto do ponto de vista sociológico como um ponto representante do componente. Denominamos como CP o conjunto de pontos de articulação de um componente;
3. Se um componente não tem pontos de articulação então o representante será o vértice com mais arestas. Se houver mais de um vértice com o mesmo número de arestas, então a escolha será aleatória. Este vértice é chamado de ponto central (*central point*). Estamos denominando como LCP o ponto central de um componente;
4. O conjunto de nós selecionados é a união entre o conjunto CP e o conjunto LCP. Quer dizer, $W = CP \cup LCP$.

A figura 3 mostra um exemplo de seleção de testemunhas utilizando o REGRET.

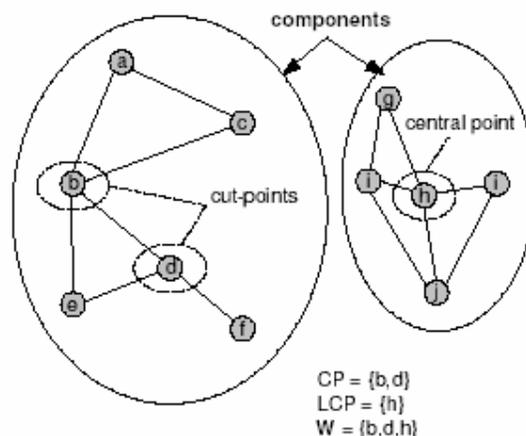


Figura 3. Selecionando testemunhas com REGRET [23]

Funcionamento do mecanismo Witness Reputation do modelo FIRE

O modelo FIRE também utiliza esta abordagem baseada em testemunhos. Um agente pode solicitar avaliações para outros agentes (testemunhas) e calcular a reputação com base nestas avaliações recebidas. O processo de busca por avaliações para estimar a reputação por testemunha está descrito nos 3 passos abaixo:

1. Se um agente A deseja estimar a reputação do agente B com respeito ao termo C então o agente A envia solicitações para vários agentes “conhecidos” (acquaintances) que possivelmente já interagiram com o agente B segundo o termo C.
2. Os “conhecidos” do agente A recebem a solicitação e verificam em sua base de dados local se possuem avaliações do agente B segundo o termo C. Caso afirmativo retornam a reputação ao agente A caso contrário retornam referências de outros agentes para prosseguir com a pesquisa.
3. O processo continua até que o agente A encontre testemunhas suficientes ou até estourar o limite pré-definido para o tamanho desta cadeia de referência.

Funcionamento do mecanismo testemunhos de [1]

O modelo baseado em testemunhos em [1], denominado *Recommender*

Trust, tem um diferencial quando comparado com outros modelos similares porque ao modelar as recomendações de um agente o mecanismo verifica se o resultado da interação foi equivalente à recomendação, de forma que em uma próxima recomendação deste agente esta diferença será considerada. Apesar disso, o mecanismo de recomendações não deve ser utilizado por agentes novos porque não possuem nenhuma informação sobre outros agentes e por isso não sabe sequer em quais recomendações pode confiar.

Vantagem do uso da Reputação Baseada em Testemunhas:

- (i) Possibilidade de conhecer a reputação de um agente mesmo sem ter interagido com ele anteriormente, resolvendo uma desvantagem encontrada nos mecanismos baseados no modelo descentralizado simples.

Desvantagens do uso da Reputação Baseada em Testemunhas:

- (i) O agente não sabe quem já interagiu anteriormente com um determinado agente. Por isso o processo utilizado pelo FIRE e por outras abordagens para encontrar testemunhas pode ser inviável por ser muito lento, considerando sistemas de larga escala.
- (ii) As reputações fornecidas por um agente são baseadas no ponto de vista deste agente. Agentes podem avaliar de maneira diferentes situações iguais, por isso a violação de uma norma pode ser mais grave do ponto de vista de um agente que do ponto de vista de outro agente.
- (iii) Em um sistema multi-agente de larga escala pode ser difícil estabelecer uma quantidade de recomendações suficientes para saber se o agente é uma boa testemunha. Agentes podem mentir, i.e., fornecer informações incorretas, e pode ser difícil saber se a recomendação é confiável.
- (iv) Agentes que estão competindo por clientes, por exemplo, não estarão dispostos a colaborar com seus concorrentes e por isso não irão repassar informações.

- (v) Quando um agente inicia sua participação em um sistema ele não conhece nenhum participante e por isso não pode confirmar as avaliações que recebe dos agentes que questionou.

2.2.3.

Modelo Descentralizado Baseado em Reputação Certificada

Uma terceira abordagem propõe o uso de reputações certificadas. Aqui o agente vai coletando as avaliações que outros agentes fizeram sobre ele mesmo. A figura 4 ilustra esta abordagem, que tem como principal vantagem o fato da reputação estar sempre disponível. Após negociar com o agente B, o agente A solicita ao agente B a avaliação que foi feita. Depois, ao interagir com o agente D, o agente A também solicita que este agente repasse a avaliação que ele fez. Então o agente A tem armazenado as avaliações que outros agentes (B e D) fizeram sobre ele. Estas avaliações são chamadas “avaliações certificadas” e A não tem poder para alterá-las. Desta forma, se um agente C pretende interagir com o agente A, o agente C solicita o conjunto de avaliações certificadas ao agente A. Com base no conjunto de avaliações certificadas enviadas pelo próprio agente A, o agente C calcula a reputação certificada. O problema desta abordagem é que a reputação certificada calculada por C poderá estar superestimada porque as avaliações ruins possivelmente não serão enviadas pelo agente A. No exemplo da figura 4, o agente A não enviou a avaliação feita pelo agente B, enviou somente a avaliação feita pelo agente D. Por exemplo, um agente exportador pode armazenar as avaliações que todos os agentes importadores fizeram dele em negociações anteriores. As avaliações são certificadas, i.e., o agente exportador não tem poder para alterá-las. Mas o seu algoritmo poderá ignorar avaliações menores do que a nota máxima (normalmente os valores máximos são iguais a 1). Um agente importador que pretende interagir com ele pode solicitar as avaliações certificadas e com elas calcular a reputação do agente exportador, que poderão estar superestimadas.

O modelo FIRE introduziu esta abordagem baseada em reputação certificada para complementar os mecanismos baseados nas experiências diretas e nas testemunhas, uma vez que as informações sobre a performance de um agente nem sempre estão disponíveis, por ser um agente desconhecido ou então pela

dificuldade em se obter uma recomendação confiável. O mecanismo de reputação certificada do modelo FIRE funciona da seguinte forma:

1. Após cada transação, o agente solicita ao parceiro que forneça as avaliações certificadas sobre sua performance (sobre cada item do contrato, por exemplo). O agente passa a armazenar um conjunto de avaliações certificadas feitas por todos os parceiros com quem interagiu. As avaliações ruins são descartadas;
2. Quando um agente tem interesse em utilizar os serviços de outro agente, pode então solicitar as avaliações certificadas (de um determinado item, por exemplo) feitas por outros agentes em interações anteriores;
3. O agente recebe um conjunto de avaliações certificadas, como se fossem referências, e então calcula a reputação certificada com base neste conjunto de avaliações feitas por outros agentes.

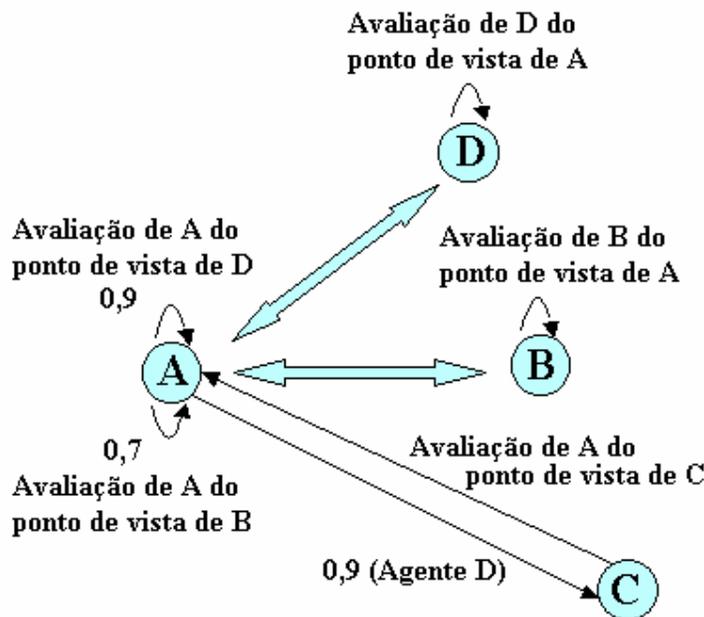


Figura 4. Modelo descentralizado baseado em reputação certificada

Este mecanismo tem similaridades com as ferramentas de gerenciamento de políticas de confiança, como PolicyMaker [9] e Trust-Serv [27]. Nestas ferramentas a confiança em um agente é obtida através de políticas ou regras pré-definidas que determinam quem é confiável ou quem está autorizado a fazer ou a executar uma determinada função através da chave pública. A diferença entre estas ferramentas e o mecanismo de reputação certificada proposto pelo modelo

FIRE é que a primeira determina quais os serviços que os agentes podem executar enquanto que o segundo determina como os agentes se comportaram ao executar o serviço.

Vantagens do uso da Reputação Certificada:

- (i) A reputação de um agente pode ser facilmente conhecida se este agente já interagiu com algum outro agente do sistema.

Desvantagens do uso da Reputação Certificada:

- (i) A reputação de um agente será provavelmente superestimada porque as avaliações ruins possivelmente serão descartadas pelo agente. Este pode fornecer somente as suas avaliações boas quando algum outro agente lhe pergunta sua reputação.
- (ii) As reputações certificadas estão influenciadas pelo ponto de vista dos agentes que forneceram as certificações.

2.3. Considerações

Em [19] diversos trabalhos sobre reputação foram reunidos com o objetivo de construir uma tipologia para representar graficamente as noções de reputação para sistemas multi-agentes. Porém não foram identificados claramente dentro da tipologia quais os conceitos são utilizados por mecanismos de reputação ao nível de sistema e quais são utilizados ao nível do agente de software.

Segundo [20], os mecanismos de reputação ao nível de agente e de sistema se complementam. Mecanismos ao nível de agente, descentralizados e com suas estratégias subjetivas, são prioritários ao decidir com quem interagir. Porém, na falta de informações particulares, um mecanismo de reputação centralizado, ao nível de sistema, pode preencher esta lacuna e diminuir a incerteza, fornecendo informações sobre o comportamento anterior dos agentes desconhecidos.

O nosso modelo de reputação, que será detalhado no próximo capítulo, propõe um mecanismo de reputação ao nível de sistema, distribuído em organizações, para ser utilizado em conjunto com os mecanismos de reputação ao nível do agente de software, que evidentemente irão variar de acordo com a

implementação do agente. A implementação dos mecanismos descentralizados, através de modelos como REGRET, FIRE, [1] e [31,32], não interfere na abordagem centralizada. Tais modelos capacitam o agente a avaliar as interações, que são armazenadas individualmente. Os agentes utilizam diferentes fontes de informação para estimar a confiança e reputação de outros agentes. Modelo sócio-cognitivo como [8], onde confiança é considerada um estado mental, também pode ser utilizado para implementar o mecanismo de reputação descentralizado.

As principais diferenças entre o nosso mecanismo de reputação (semi-) centralizado e os mecanismos centralizados citados na seção 2.1 são as seguintes:

- (i) Nosso mecanismo de reputação não mente, ou seja, não fornece reputações falsas. As reputações fornecidas pelo sistema são imparciais e confiáveis. Nosso mecanismo não recebe avaliações do ponto de vista de um agente sobre a performance do parceiro. Nosso sistema recebe testemunhos sobre violações de normas, que são julgados previamente por um sistema de julgamento [6] minimizando assim a alteração da reputação de um agente por um testemunho mentiroso, embora possam ocorrer erros no julgamento.
- (ii) Nossa abordagem oferece a possibilidade de distribuir o processamento e armazenamento das reputações em uma ou mais organizações distribuídas, sendo adequado para sistemas multi-agentes de larga escala.
- (iii) Nossa abordagem é capaz de fornecer diferentes reputações para cada agente de acordo com diferentes contextos. Três contextos distintos foram definidos estabelecendo três tipos de reputação: reputação global, reputação por papel e por norma. As diferentes reputações ajudam os agentes a prever o comportamento dos parceiros em diferentes situações. É objetivo desta dissertação que o *framework* proposto seja capaz de agrupar reputações para diferentes contextos.

A principal vantagem da nossa abordagem em relação aos mecanismos descentralizados é o uso do mecanismo de reputação ao nível de sistema em conjunto com os mecanismos de reputação descentralizados. Ao propor uma abordagem híbrida estamos reunindo as vantagens das abordagens centralizadas e

descentralizadas. As principais vantagens da solução híbrida são:

- (i) As reputações fornecidas pelo sistema são imparciais (não dependem do ponto de vista de nenhum agente) e confiáveis (os testemunhos que influenciam na reputação são julgados previamente).
- (ii) Agentes podem obter informações sobre outros agentes com quem eles nunca interagiram sem estarem suscetíveis a receber um conjunto de reputações certificadas que podem estar superestimando a reputação do agente;
- (iii) Agentes podem atualizar a reputação dos parceiros com quem eles não interagem há muito tempo consultando a(s) organização(ões) onde estes agentes desempenham papéis.
- (iv) Agentes não precisam se encontrar freqüentemente para obter reputações consistentes dos outros agentes nem precisam procurar por outros agentes que já interagiram anteriormente com o agente alvo.