

4 Confiança no Modelo

4.1 Introdução

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 (Mul05). 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 verificou 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 para outro e levar mais tempo do que o razoável para a sua execução. Ou então, a qualidade da ação executada pelo 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 (Ram04). 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.

4.2

O que é Confiança

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 (Dig05, For03, Rol00). 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 (Dig05).

A confiança é definida de várias maneiras diferentes (For03) 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* (Gam00, Rie03)¹. 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 (For03, Gam00), 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* (McK96). 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* (May95). 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

¹Em inglês, *the willingness to be vulnerable based on positive expectations about the action of others*.

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. (Smi05).

Como já dito na introdução deste capítulo (seção 4.1), 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 do agente, o qual é compartilhado com os seus parceiros, 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.

4.3 **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 (Cas98, Cas00a, Cas00b).

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 (Bur89, Bur96). Afinal, o agente 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.

4.4 **Confiança em um Sistema Multi-Agentes**

Basicamente, há duas abordagens principais para se inserir confiança em sistemas multi-agentes (Ram04). Primeiramente, para permitir que agentes confiem uns nos outros é necessário fornecer-lhes a habilidade de raciocinar sobre a natureza da reciprocidade (agir para o bem comum), confiança e segurança (ter atitudes que não causam danos) 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 possam 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 adulterar 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.

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 (Huy05). Dessas características, pode-se assumir que em um sistema multi-agentes aberto:

- os agentes tendem a ser 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 (Huy05, Ram04).

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?

Como foi mostrado na seção 4.2, a confiança está em todos os níveis das interações entre agentes. Com respeito à 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.

- $frcr : \mathcal{P}(Bel) \times \mathcal{P}(Tru) \times P \rightarrow \mathcal{P}(Bel)$
- $opcoes : \mathcal{P}(Bel) \times \mathcal{P}(Tru) \times \mathcal{P}(Int) \rightarrow \mathcal{P}(Des)$
- $filtro : \mathcal{P}(Bel) \times \mathcal{P}(Tru) \times \mathcal{P}(Des) \times \mathcal{P}(Int) \rightarrow \mathcal{P}(Int)$

Já a função $execucao()$, que não depende da confiança, não precisa ser redefinida. Para a implementação de um algoritmo que faz o controle de um agente BDI, tem-se a função $plano()$, que também deve ser redefinida para levar em conta a confiança:

$$plano : \mathcal{P}(Bel) \times \mathcal{P}(Tru) \times \mathcal{P}(Int) \rightarrow Plan$$

Já os algoritmos para a implementação do controle de um agente inteligente são modificados como a seguir, onde $T \subseteq Tru$. Para o comprometimento cego, tem-se a figura 4.2.

1. $B := B_0$
2. $T := T_0$
3. $I := I_0$
4. *enquanto verdadeiro faça*
5. *pegue a proxima percepcao p*
6. $B := frcr(B, T, P)$
7. $T := frco(B, T)$
8. $D := opcoes(B, T, I)$
9. $I := filtro(B, T, D, I)$
10. $\pi := plano(B, T, I)$
11. $execucao(\pi)$
12. *fim_enquanto*

Figura 4.2: *Loop* de controle de um agente BDI com confiança com comprometimento cego

Os dois próximos algoritmos são as respectivas modificações para o comprometimento cabeça-dura. onde no primeiro o agente procurará um novo plano quando ele acreditar que o atual falhou (figura 4.3) e, no segundo, além disso descartará intenções quando elas forem impossíveis ou já tiverem sido alcançadas (figura 4.4). Nos dois casos, a função $consis()$, que verifica se o plano é consistente com as crenças e intenções atuais, agora também deve levar em conta o modelo de confiança, pois o plano deve ser consistente com a confiança em outros agentes que são relevantes para a execução do plano. O mesmo pode-se dizer com relação às funções $sucesso()$ (por exemplo, se um agente tiver como intenção testar se um outro agente é confiável e tiver como condição de sucesso um grau de confiabilidade no mesmo maior do que determinado valor) e $impossivel()$ (por exemplo, se um agente não confiar nenhum pouco em um outro, tentar atingir uma intenção onde envolvam finanças pode ser considerado impossível).

1. $B := B_0$
2. $T := T_0$
3. $I := I_0$
4. *enquanto verdadeiro faça*
5. *pegue a proxima percepcao p*
6. $B := frcr(B, T, P)$
7. $T := frco(B, T)$
8. $D := opcoes(B, T, I)$
9. $I := filtro(B, T, D, I)$
10. $\pi := plano(B, T, I)$
11. *enquanto $\neg vazio(\pi)$ faça*
12. $\alpha = cab(\pi)$
13. *execucao(α)*
14. $\pi := resto(\pi)$
15. *pegue a proxima percepcao p*
16. $B := frcr(B, T, P)$
17. $T := frco(B, T)$
18. *se $\neg consis(\pi, I, B, T)$ entao $\pi := plano(B, T, I)$ fim_se*
19. *fim_enquanto*
20. *fim_enquanto*

Figura 4.3: Agente cabeça-dura com confiança que procurará um novo plano quando ele acreditar que um plano falhou por qualquer razão

O próximo algoritmo (figura 4.5) é a implementação para o agente cabeça-aberta. Como a motivação de um agente também depende da confiança dele em outros agentes, a função *motivado()* passa a depender do modelo de confiança do agente. Isso é necessário se, por exemplo, o grau de confiança em um outro agente for muito pequeno, o que pode desmotivar um agente a realizar uma transação financeira com o primeiro.

Tem-se também, na figura 4.6, o agente cauteloso, que sempre reconsidera as intenções toda a vez que o agente executar o *loop* interno de controle. Por fim, na figura 4.7, para evitar que o agente reconsidere as suas intenções desnecessariamente, tem-se um componente de meta-nível. É também necessário que esse componente leve em conta a confiança.

4.6 Implementação da Confiança

Há basicamente quatro maneiras de se implementar o modelo de confiança que um determinado agente tem em outro (Ram04):

- 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

1. $B := B_0$
2. $T := T_0$
3. $I := I_0$
4. *enquanto verdadeiro faça*
5. *pegue a proxima percepcao p*
6. $B := frcr(B, T, P)$
7. $T := frco(B, T)$
8. $D := opcoes(B, T, I)$
9. $I := filtro(B, T, D, I)$
10. $\pi := plano(B, T, I)$
11. *enquanto $\neg vazio(\pi) \wedge \neg sucesso(I, B, T)$*
12. $\wedge \neg impossivel(I, B, T)$ *faça*
13. $\alpha = cab(\pi)$
14. *execucao*(α)
15. $\pi := resto(\pi)$
16. *pegue a proxima percepcao p*
17. $B := frcr(B, T, P)$
18. $T := frco(B, T)$
19. *se $\neg consis(\pi, I, B, T)$ entao $\pi := plano(B, T, I)$ fim_se*
20. *fim_enquanto*
21. *fim_enquanto*

Figura 4.4: Descartando intenções quando elas forem impossíveis ou já tiverem sido atingidas no modelo BDI com confiança

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 (Ram04)), que é o caso binário; com três níveis: 0, 0.5 e 1 (ou -1 , 0 e 1 (Ram04)), e assim sucessivamente. Com n níveis: 0, $\frac{1}{n-1}$, $\frac{2}{n-1}$, ..., $\frac{n-2}{n-1}$, 1. Se por um lado essa abordagem é bem mais flexível do

1. $B := B_0$
2. $T := T_0$
3. $I := I_0$
4. *enquanto verdadeiro faça*
5. *pegue a proxima percepcao p*
6. $B := frcr(B, T, P)$
7. $T := frco(B, T)$
8. $D := opcoes(B, T, I)$
9. $I := filtro(B, T, D, I)$
10. $\pi := plano(B, T, I)$
11. *enquanto $\neg vazio(\pi) \wedge \neg sucesso(I, B, T) \wedge \neg impossivel(I, B, T)$*
12. $\wedge motivado(I, B, T)$ *faça*
13. $\alpha = cab(\pi)$
14. *execucao*(α)
15. $\pi := resto(\pi)$
16. *pegue a proxima percepcao p*
17. $B := frcr(B, T, P)$
18. $T := frco(B, T)$
19. *se $\neg consis(\pi, I, B, T)$ entao $\pi := plano(B, T, I)$ fim_se*
20. *fim_enquanto*
21. *fim_enquanto*

Figura 4.5: Um agente cabeça-aberta com confiança

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 (Ram04)). 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 (Ram04).

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

1. $B := B_0$
2. $T := T_0$
3. $I := I_0$
4. *enquanto verdadeiro faca*
5. *pegue a proxima percepcao p*
6. $B := frcr(B, T, P)$
7. $T := frco(B, T)$
8. $D := opcoes(B, T, I)$
9. $I := filtro(B, T, D, I)$
10. $\pi := plano(B, T, I)$
11. *enquanto $\neg vazio(\pi) \wedge \neg sucesso(I, B, T) \wedge \neg impossivel(I, B, T)$*
12. $\wedge motivado(I, B, T)$ *faca*
13. $\alpha = cab(\pi)$
14. *execucao*(α)
15. $\pi := resto(\pi)$
16. *pegue a proxima percepcao p*
17. $B := frcr(B, T, P)$
18. $T := frco(B, T)$
19. $D := opcoes(B, T, I)$
20. $I := filtro(B, T, D, I)$
21. *se $\neg consis(\pi, I, B, T)$ entao $\pi := plano(B, T, I)$ fim_se*
22. *fim_enquanto*
23. *fim_enquanto*

Figura 4.6: Um agente com confiança que sempre reconsidera as intenções

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

1. $B := B_0$
2. $T := T_0$
3. $I := I_0$
4. *enquanto verdadeiro faça*
5. *pegue a proxima percepcao p*
6. $B := frcr(B, T, P)$
7. $T := frco(B, T)$
8. $D := opcoes(B, T, I)$
9. $I := filtro(B, T, D, I)$
10. $\pi := plano(B, T, I)$
11. *enquanto $\neg vazio(\pi) \wedge \neg sucesso(I, B, T) \wedge \neg impossivel(I, B, T)$*
12. $\wedge motivado(I, B, T)$ *faça*
13. $\alpha = cab(\pi)$
14. *execucao*(α)
15. $\pi := resto(\pi)$
16. *pegue a proxima percepcao p*
17. $B := frcr(B, T, P)$
18. $T := frco(B, T)$
19. *se reconsiderar*(I, B, T) *entao*
20. $D := opcoes(B, T, I)$
21. $I := filtro(B, T, D, I)$
22. *fim_se*
23. *se $\neg consis(\pi, I, B, T)$ entao $\pi := plano(B, T, I)$ fim_se*
24. *fim_enquanto*
25. *fim_enquanto*

Figura 4.7: Um agente com confiança que tenta atingir um balanço entre teimosia e cautela

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 tiver sido pedido a um outro agente e não tiver sido respondido ou a tarefa solicitada não tiver sido 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.