

Utilisation de normes et de réputations pour détecter et sanctionner les contradictions

—
Contribution au Contrôle Social des interactions dans les
Systèmes Multi-Agents Ouverts et Décentralisés

Guillaume Muller

Directeur	Olivier	BOISSIER
Co-directeur	Laurent	VERCOUTER
Rapporteurs	Yves	DEMAZEAU
	Carles	SIERRA

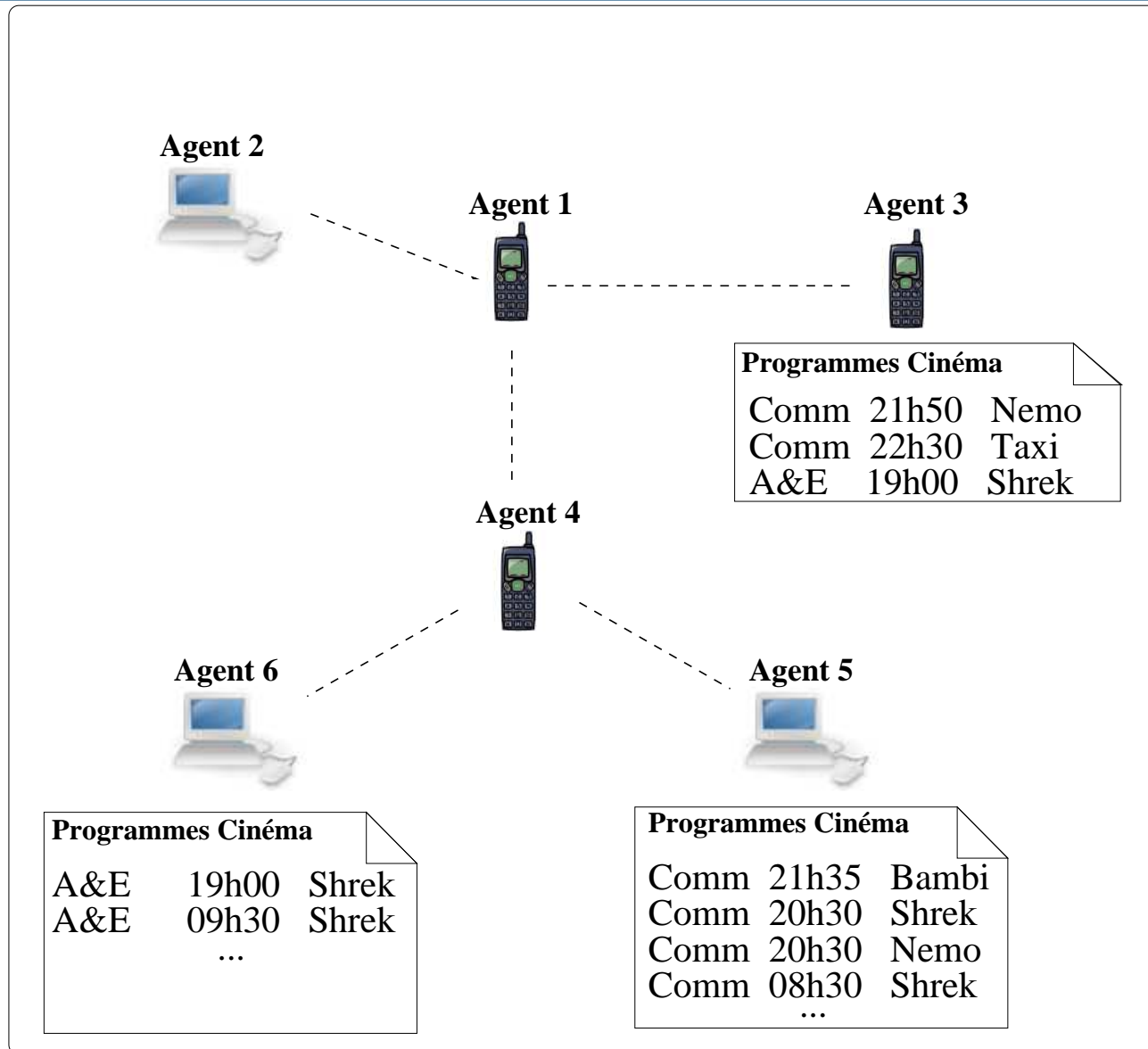
Examineurs	Salima	HASSAS
	Andreas	HERZIG
	Juliette	ROUCHIER

- ✓ **Cadre** : Systèmes Multi-Agents Ouverts et Décentralisés (SMAOD)
 - ✗ Modification, ajout ou retrait dynamique d'entités
 - ✗ Absence de vision globale sur le système

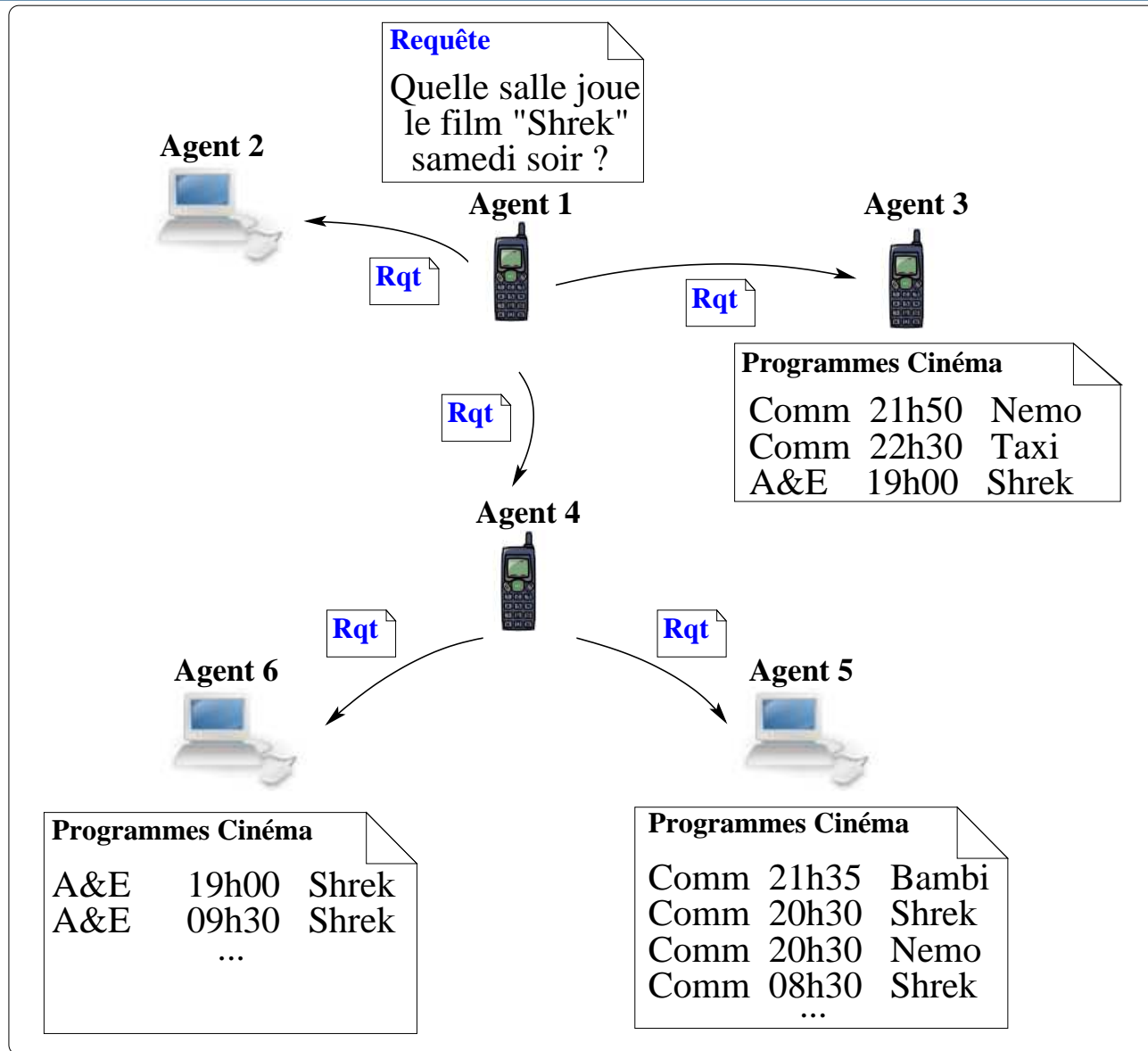
- ✓ **Problématique** : Identifier et isoler les agents nuisibles

- ✓ **Objectif** : Contrôle des interactions entre agents
 - ✗ Contrainte 1 : décentralisé
 - ✗ Contrainte 2 : non intrusif

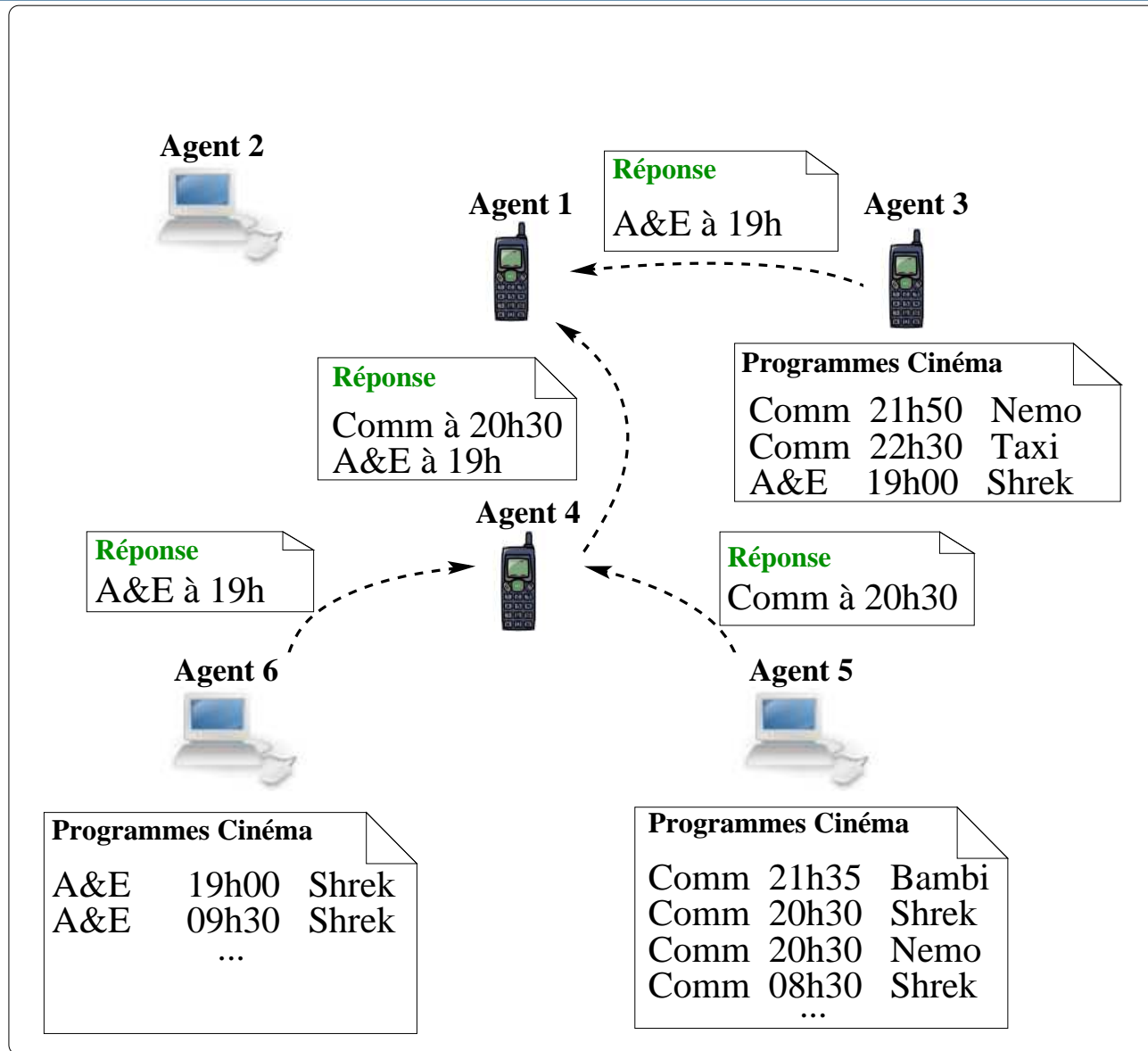
Problématique : identifier et isoler les agents nuisibles dans un SMA Ouvert et Décentralisé



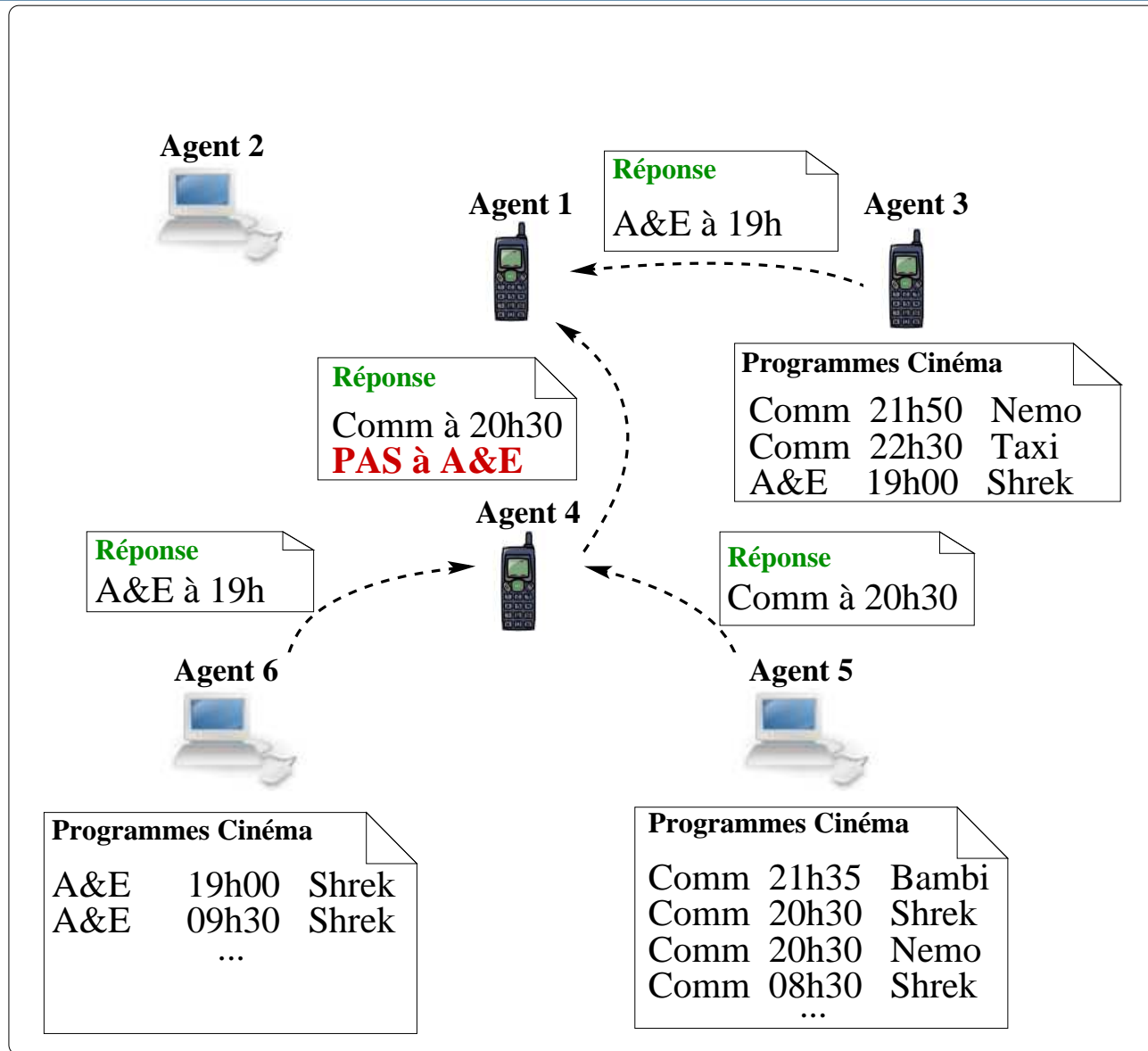
Problématique : identifier et isoler les agents nuisibles dans un SMA Ouvert et Décentralisé



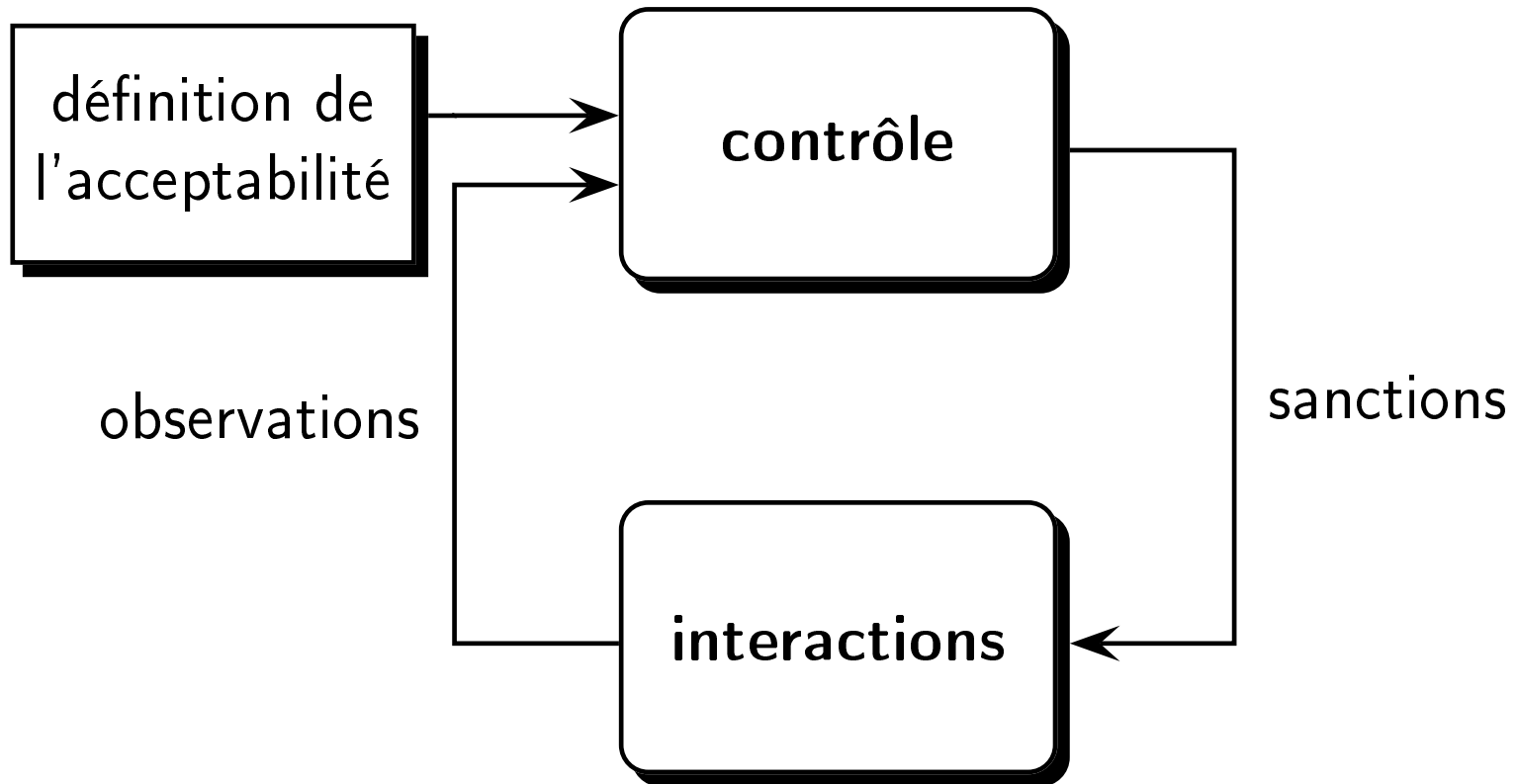
Problématique : identifier et isoler les agents nuisibles dans un SMA Ouvert et Décentralisé



Problématique : identifier et isoler les agents nuisibles dans un SMA Ouvert et Décentralisé



Objectif : contrôle des interactions entre agents



Contrôle des interactions entre agents : approches

✓ Approche sécuritaire

PolicyMaker, Keynote... [Blaze96,Blaze98,Seigneur02]

- ✗ *But* : Empêcher les agents nuisibles d'entrer
- ✗ *Principes* : Cryptographie + Protocoles d'interaction
- ✗ ⊖ *Limites* : *intrusif, ne traite pas les mensonges*

✓ Approche institutionnelle

[Plaza98,Camino05,Gateau05]

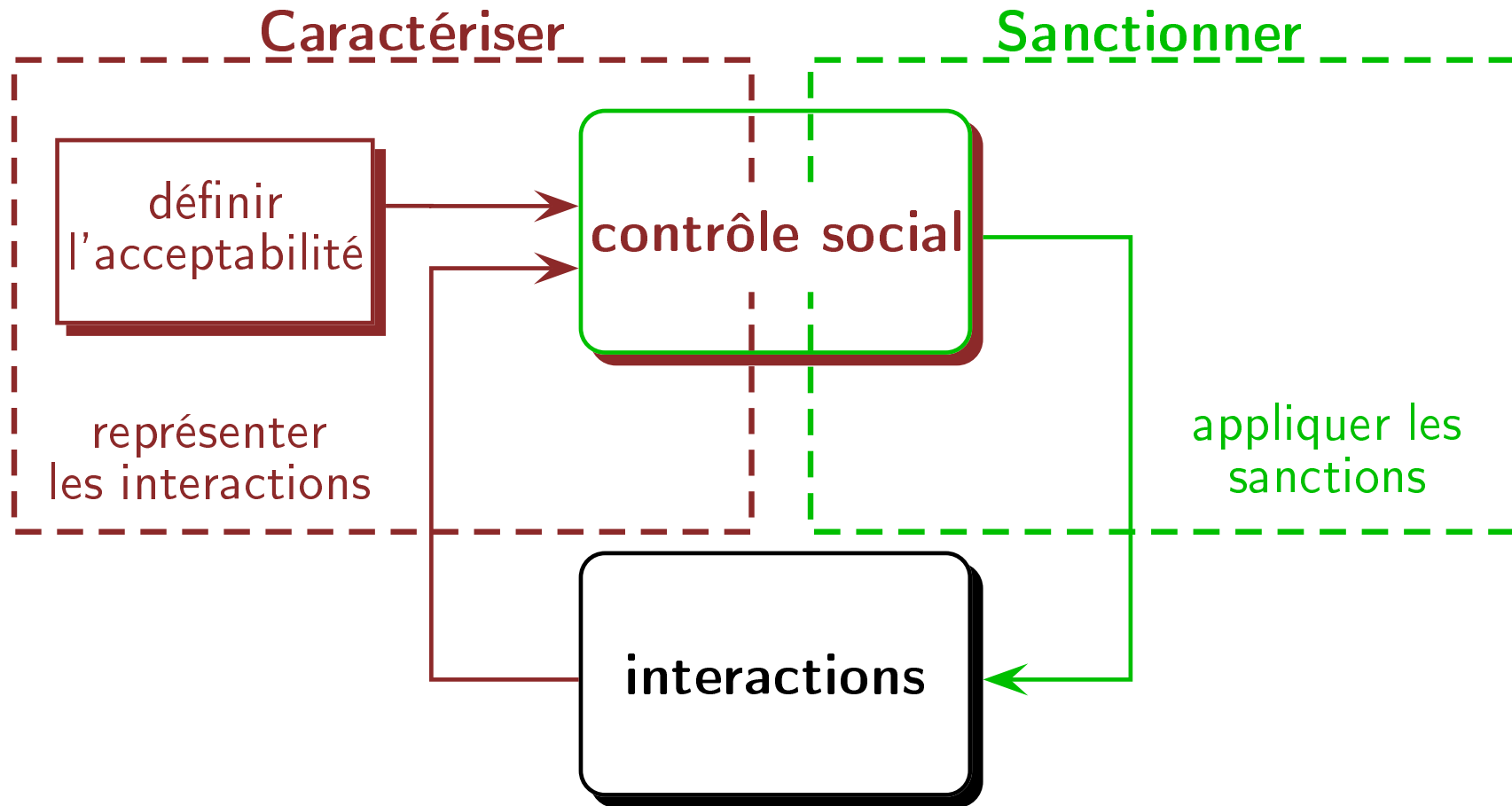
- ✗ *But* : Définir les organisations où les agents interagissent
- ✗ *Principes* : Définition des rôles, groupes, missions...
- ✗ ⊖ *Limite* : *centralisé*

✓ Approche sociale

[Castelfranchi98,Rouchier00,McKnight01,Sabater02,Conte02,Melaye05]

- ✗ *But* : « Définir un contrôle adaptatif et auto-organisé, mis en place par les agents eux-mêmes » [Castelfranchi00]
- ✗ *Principe* : **Décentraliser le contrôle (confiance et réputation)**

Sous-objectifs du contrôle social des interactions



Chaque agent doit pouvoir **représenter** les interactions qu'il perçoit, **définir l'acceptabilité** des interactions, **évaluer** les interactions perçues et **sanctionner** les agents pour qu'ils modifient leurs interactions.

Plan

Introduction

Cadre

Problématique

Objectif

Approches

Sous-objectifs

Plan

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

LIAR

✓ Introduction

✓ État de l'art

✗ Modèles pour le Contrôle Social des interactions

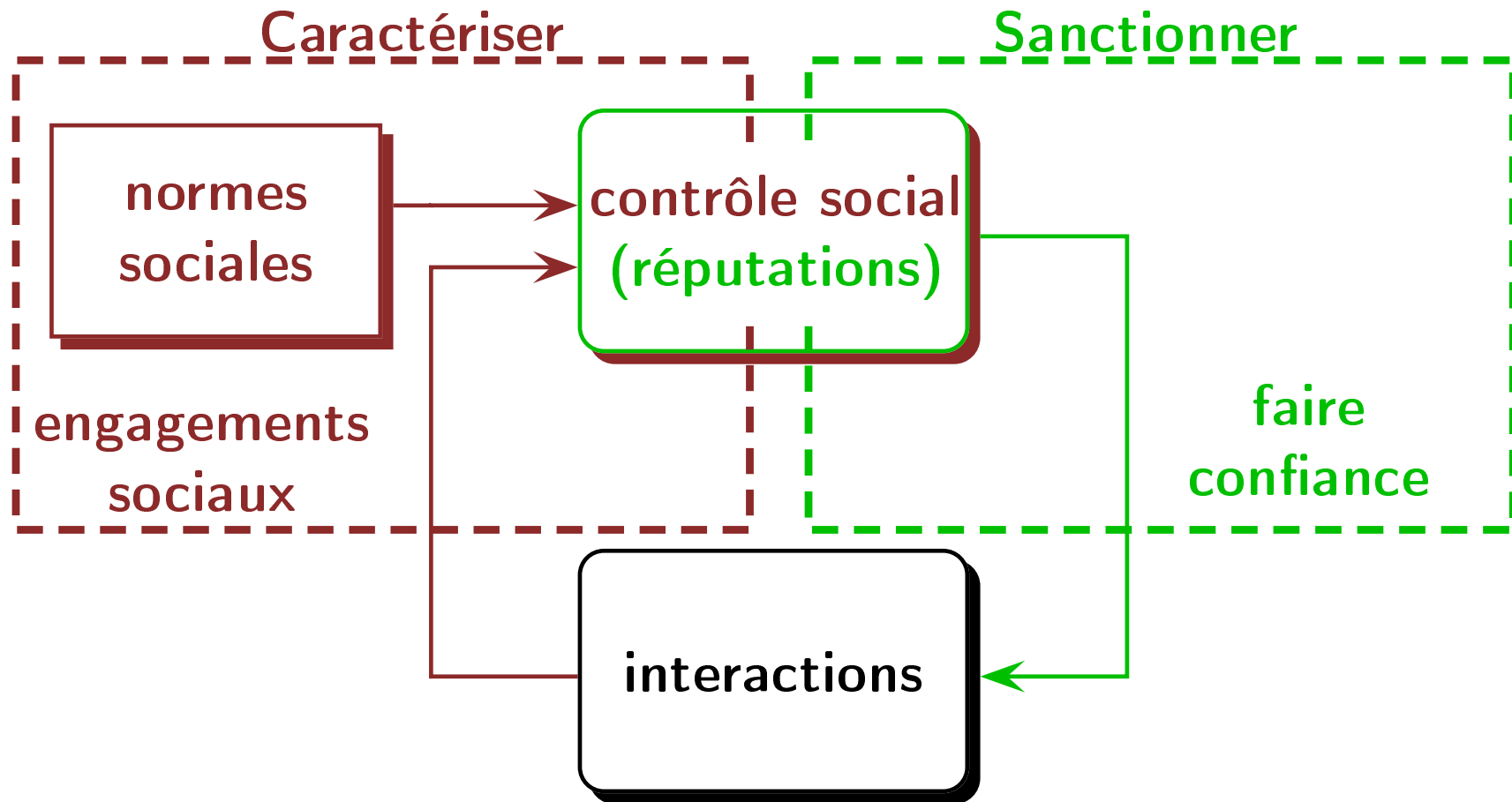
✗ Concepts et Modèles de Confiance et de Réputation

✓ Modèle L.I.A.R.

✓ Application & Expérimentations

✓ Conclusions & Perspectives

Modèles pour le Contrôle Social des interactions



- ✓ Engagement social [Bentahar03,Pasquier04,Fornara05]
- ✓ Normes sociales [Dignum96,Esteva01,Lopez02,Vàzquez03]

Définitions

Introduction

État de l'Art

Techniques

Concepts

Processus

Informatique/SMA

Social ReGreT

Synthèse

Plan

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

LIAR

Pas de définition communément admise [Barber83,Shapiro87]

Définition : « faire confiance »

Le processus de décision de la confiance, qui aboutit à un comportement en confiance.

Définition : « réputations »

Les croyances sur lesquelles un agent s'appuie pour prendre sa décision d'agir en confiance.

Construction et utilisation des réputations

Introduction

État de l'Art

Techniques

Concepts

Processus

Informatique/SMA

Social ReGreT

Synthèse

Plan

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

LIAR

Processus impliquant la réputation (extension de [Casare05]) :

- ✓ **Initialisation**
Affectation d'un niveau par défaut
- ✓ **Évaluation**
Juger un comportement
- ✓ **Punition**
Établir les niveaux de réputation
- ✓ **Raisonnement**
Déduire les intentions de faire confiance
- ✓ **Décision**
Décider de faire confiance
- ✓ **Propagation**
Décider quand, à qui, pourquoi et comment diffuser la réputation ?

Modèles computationnels de réputation

	Initialisa°	Evalua°	Puni°	Raisonnem ^t	Décis°	Propaga°
OpenPGP	O	X	H	H	H	H
eBay, onSale	O	H	O	H	H	H
Zacharias	O	H	O	N	H	N
AbdulRahman	O	H	O	N	O	I
Marsh	N	N	O	N	O	N
AFRAS	N	H	O	N	O	O
Schillo	N	O	O	N	N	I
Sen	O	O	O	N	O	I
Wang	N	H	O	O	O	I
Melaye	N	N	N	O	O	N
Sabater	O	O	O	O	O	O

Légende :			
O	oui	N	non
H	humain	I	inféré
X	non pertinent		

Modèles computationnels de réputation

	Initialisa°	Evalua°	Puni°	Raisonnem ^t	Décis°	Propaga°	- ↑ autonomie ↓ +
OpenPGP	O	X	H	H	H	H	
eBay, onSale	O	H	O	H	H	H	
Zacharias	O	H	O	N	H	N	
AbdulRahman	O	H	O	N	O	I	
Marsh	N	N	O	N	O	N	
AFRAS	N	H	O	N	O	O	
Schillo	N	O	O	N	N	I	
Sen	O	O	O	N	O	I	
Wang	N	H	O	O	O	I	
Melaye	N	N	N	O	O	N	
Sabater	O	O	O	O	O	O	

Légende :			
O	oui	N	non
H	humain	I	inféré
X	non pertinent		

Social ReGreT, le modèle de [Sabater02]

Introduction

État de l'Art

Techniques

Concepts

Processus

Informatique/SMA

Social ReGreT

Synthèse

Plan

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

LIAR

- ✓ Cadre : commerce électronique
- ✓ Processus d'évaluation
 - ✗ Négociation d'un contrat entre deux agents
 $(a_1 = v_1) \wedge (\dots) \wedge (a_n = v_n)$
 - ✗ Exécution du contrat
 $(a_1 = v'_1) \wedge (\dots) \wedge (a_n = v'_n)$
 - ✗ \ominus *Évaluation triviale grâce au cadre contractuel*
- ✓ Modèle de réputation
 - ✗ Richesse sémantique des types de réputation
Réputation Collective et Stéréotypée
 - ✗ \ominus *Perte de la richesse lors des raisonnements/décisions*

Synthèse de l'état de l'art

Introduction

État de l'Art

Techniques

Concepts

Processus

Informatique/SMA

Social ReGreT

Synthèse

Plan

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

LIAR

- ✓ Il existe différents formalismes pour le Contrôle Social
- ✓ Ils ne sont **pas adaptés** aux SMA Ouverts et Décentralisés
- ✓ Ils ne sont **pas intégrés** les uns avec les autres
- ✓ L'évaluation n'est pas déléguée aux agents ou est **triviale**

Plan

Introduction

État de l'Art

Techniques

Concepts

Processus

Informatique/SMA

Social ReGreT

Synthèse

Plan

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

LIAR

✓ Introduction

✓ État de l'art

✓ Modèle L.I.A.R.

✗ Modèles et processus pour caractériser

✗ Modèle de réputation pour sanctionner

✓ Application & Expérimentations

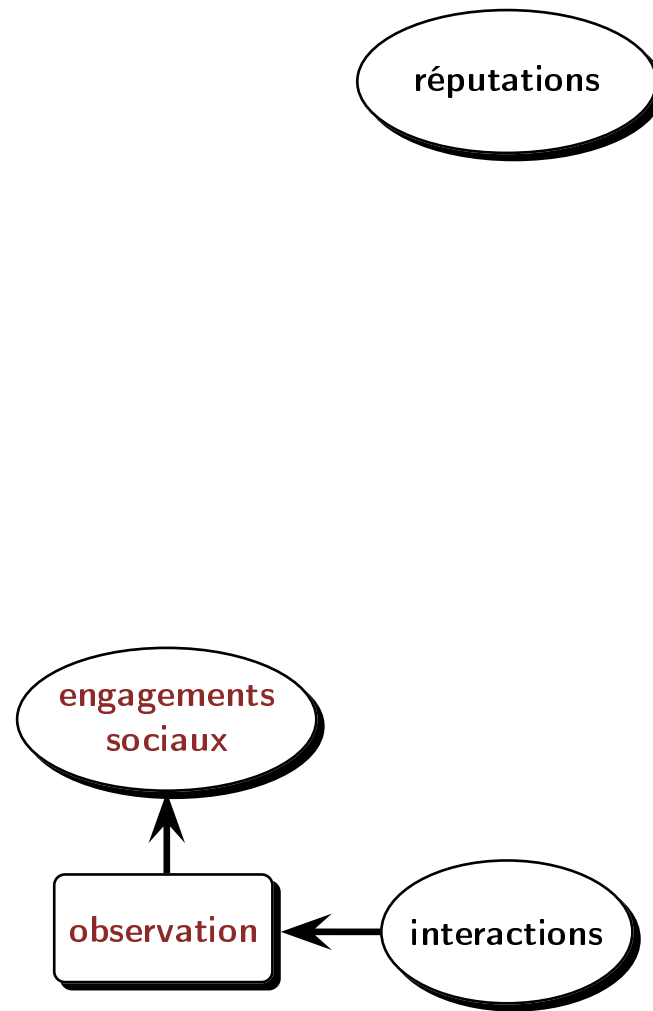
✓ Conclusions & Perspectives

Architecture du modèle L.I.A.R.

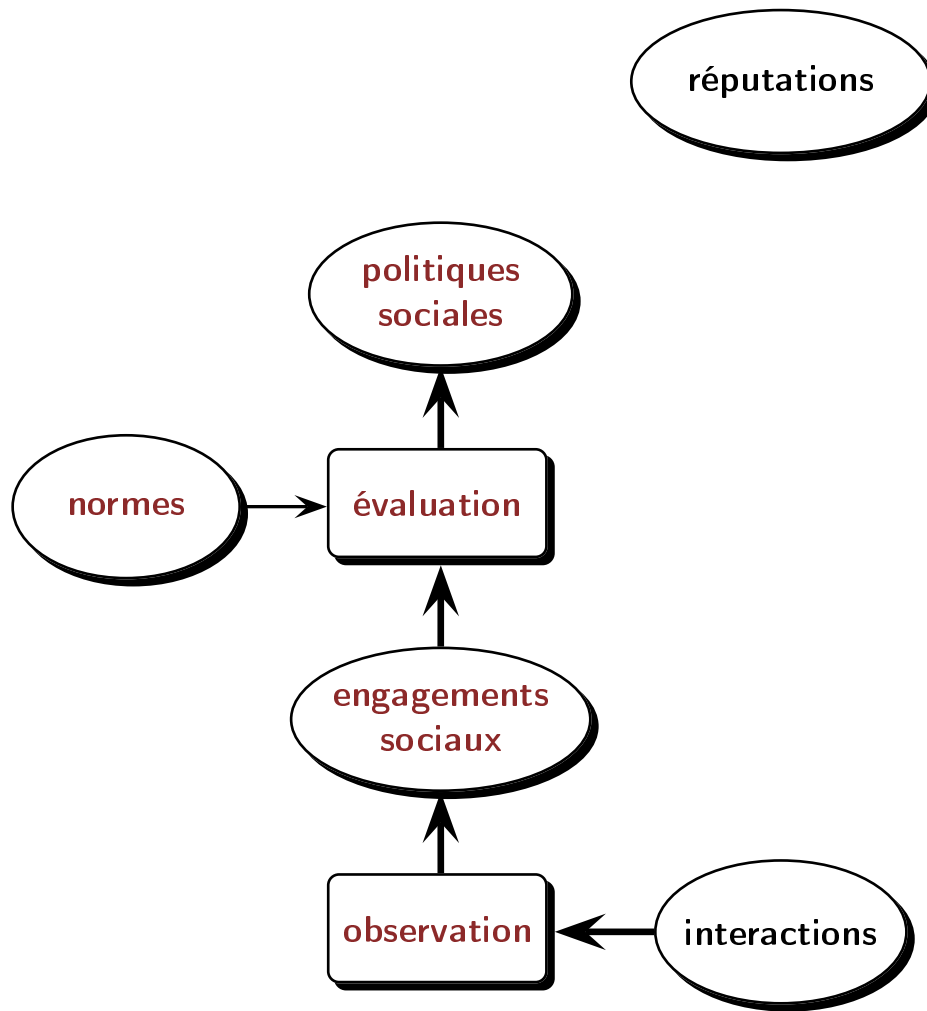
réputations

interactions

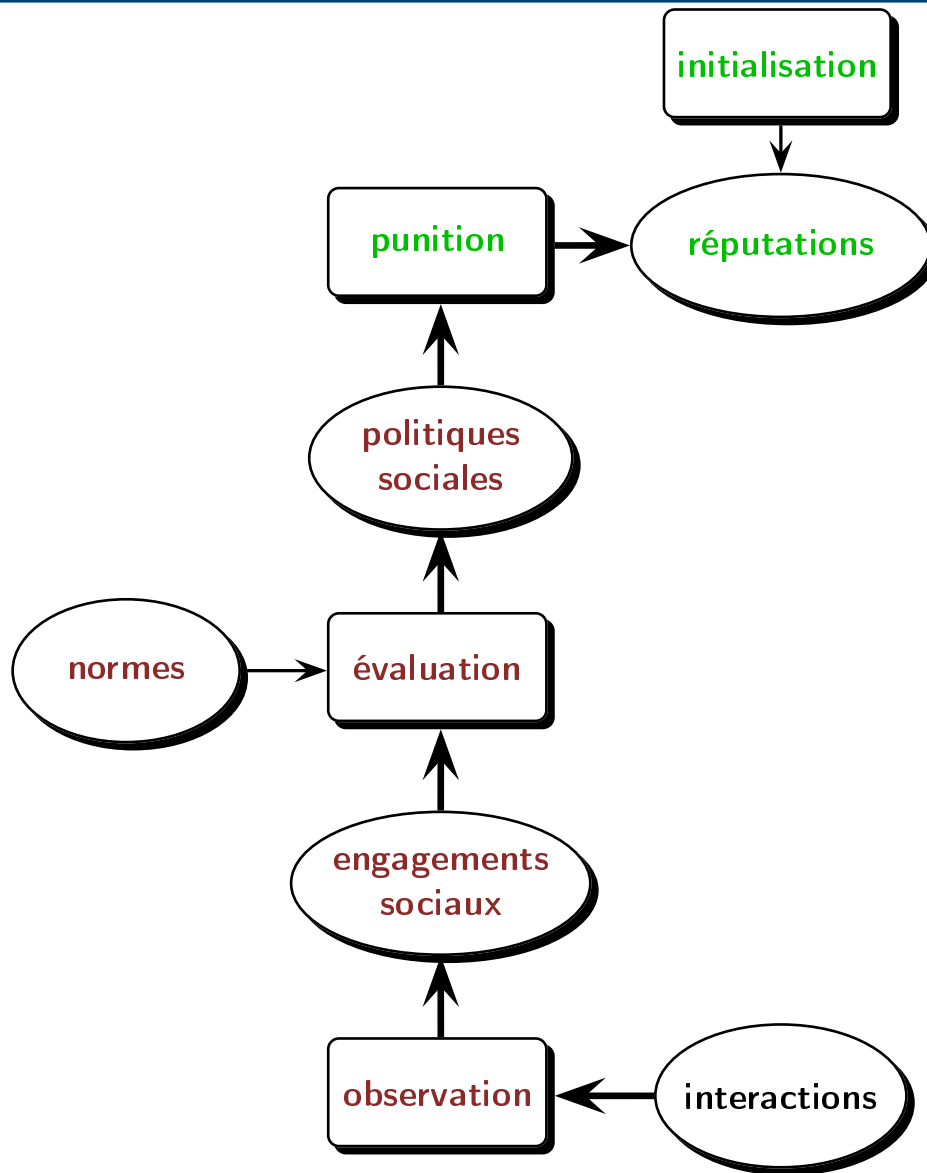
Architecture du modèle L.I.A.R.



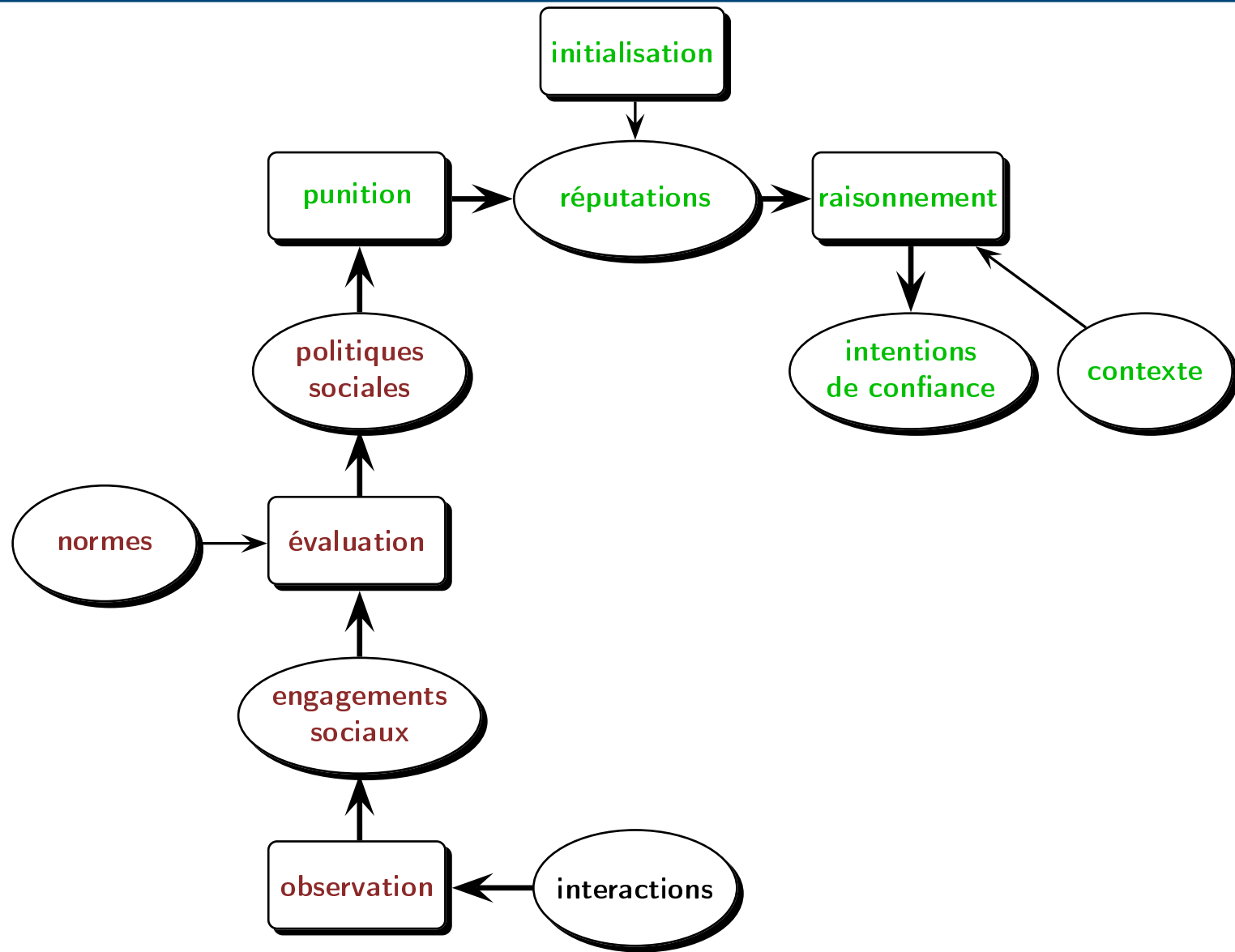
Architecture du modèle L.I.A.R.



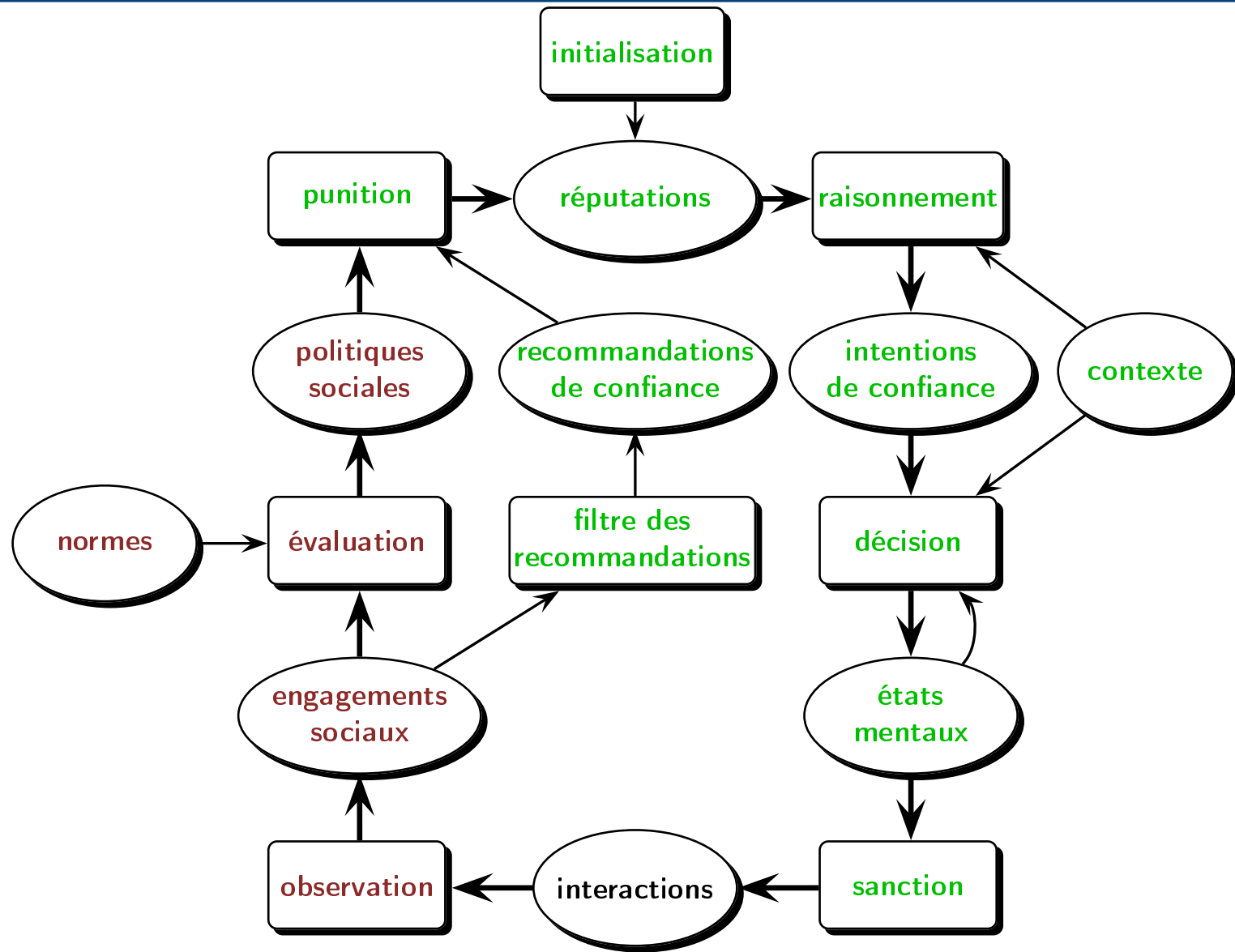
Architecture du modèle L.I.A.R.



Architecture du modèle L.I.A.R.



Architecture du modèle L.I.A.R.



Modèle d'engagement social

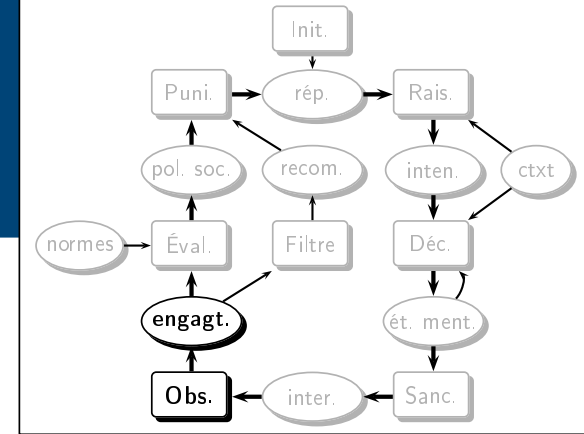
Définition :

$obSCom(db, cr, t_e, st, [cond,] cont)$

- ✓ ob : observateur, agent qui modélise l'engagement
- ✓ db : débiteur, agent est engagé
- ✓ cr : créateur, agent envers qui l'engagement est pris
- ✓ t_e : moment d'émission
- ✓ st : état de l'engagement social
- ✓ cond : conditions (optionnel)
- ✓ cont : contenu, ce sur quoi le débiteur est engagé

Exemple :

$obSCom(a, b, 13 : 00, active, shows(theater1, Shrek, 19 : 00, room1))$



Modèle de norme

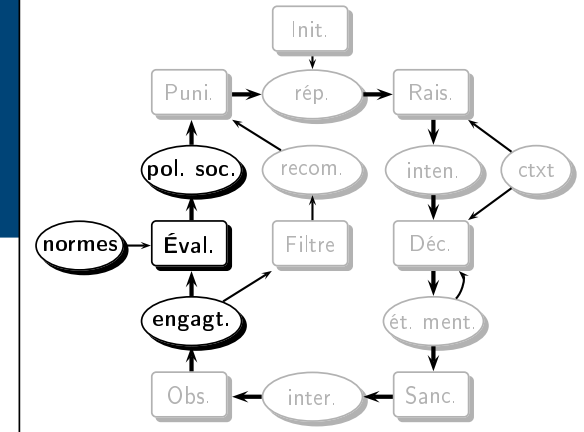
Définition :

$\text{snorm}(\text{op}, \text{Tg}, \text{Ev}, \text{Pu}, \text{cond}, \text{cont}, \text{st})$

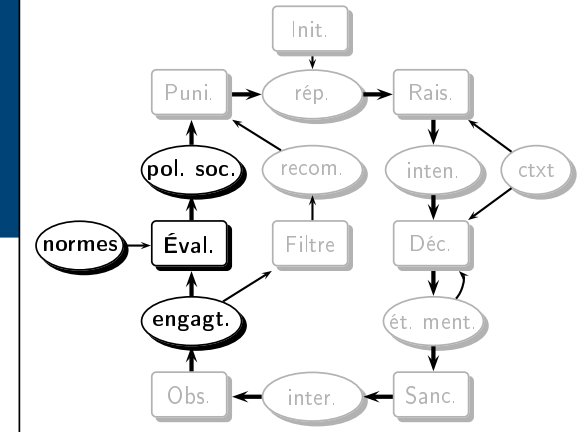
- ✓ $\text{op} \in \{ \text{I}, \text{O}, \text{P} \}$ opérateur déontique
- ✓ Tg : cible
- ✓ Ev : évaluateurs
- ✓ Pu : punisseurs
- ✓ cond : conditions de validité
- ✓ cont : contenu
- ✓ st : état (parmi : { inactive, active })

Exemple :

$\text{snorm}(\text{I}, G(t), \Omega(t), \Omega(t), \text{true},$
 $\forall x \in G(t), \text{talk_about_showtimes}(x), \text{active})$



Instanciation des normes



- ✓ Les normes sont définies du point de vue du système
- ✓ Instanciation en **politiques sociales** pour que chaque évaluateur puisse :
 - ✗ Spécifier le contenu dans son référentiel
 - ✗ Gérer la violation d'une même norme par différentes cibles
 - ✗ Gérer la violation répétée d'une même norme par une cible donnée
 - ✗ Faire apparaître des pénalités

Politique sociale

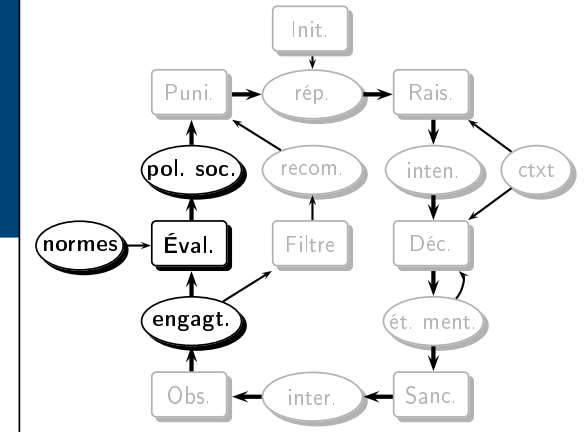
Définition :

${}_{ev}SPol(db, cr, t_e, st, [cond,] cont)$

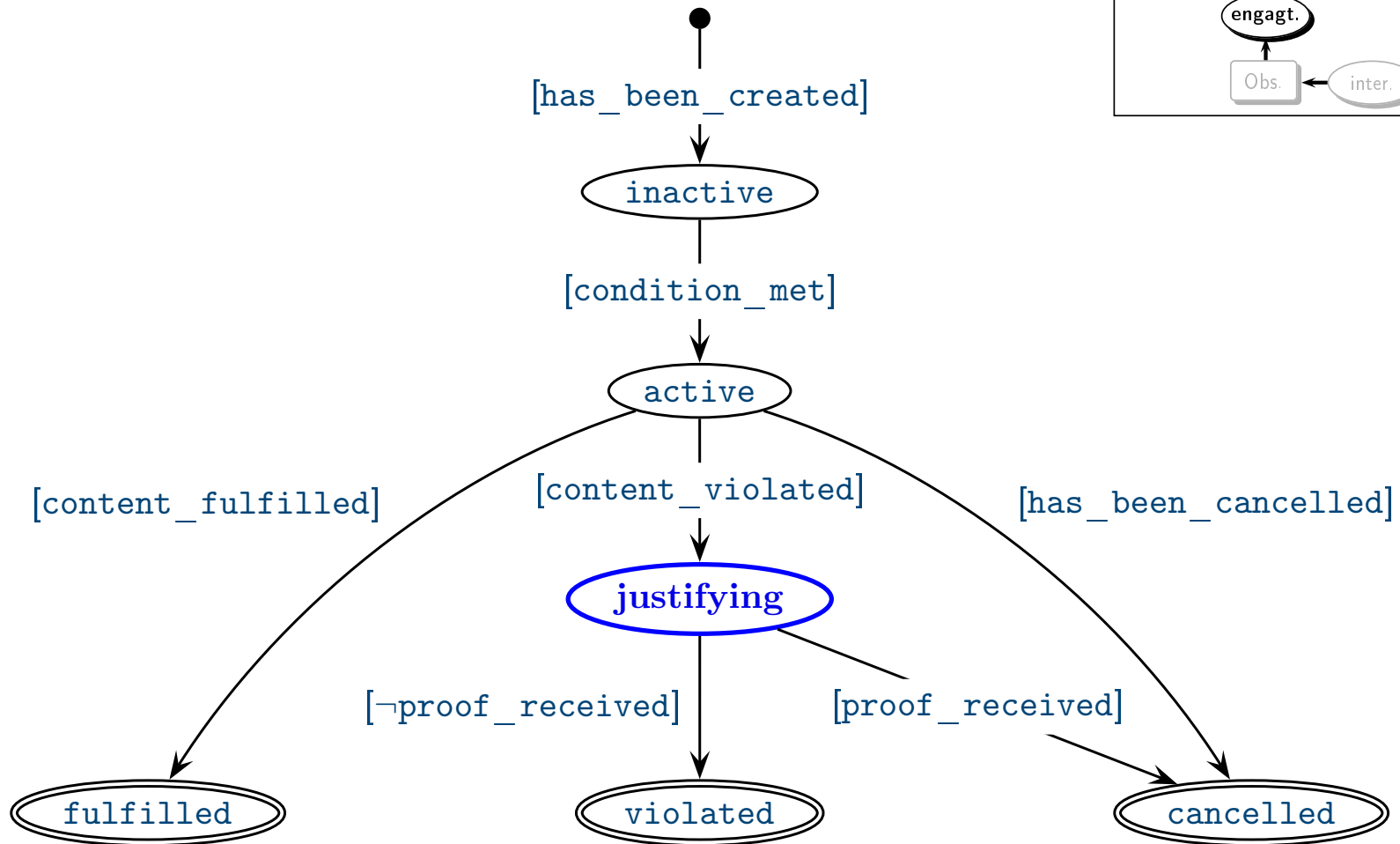
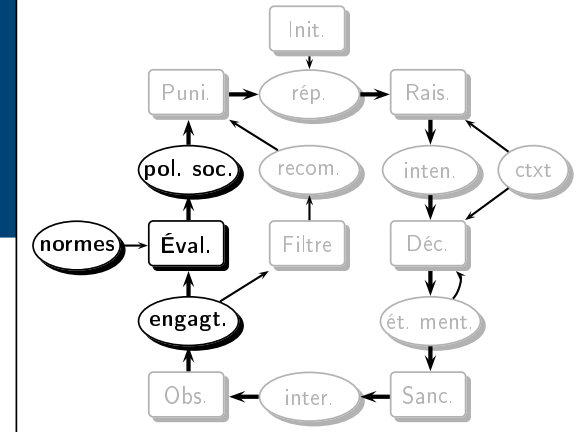
- ✓ ev : évaluateur qui instancie la norme
- ✓ db : débiteur
- ✓ cr : créateur
- ✓ t_e : moment de création
- ✓ st : état
- ✓ $cond$: conditions d'activation (optionnel)
- ✓ $cont$: contenu
- ✓ $punishes : \mathcal{S} \times \mathcal{T} \mapsto [0, +1]$

Exemple :

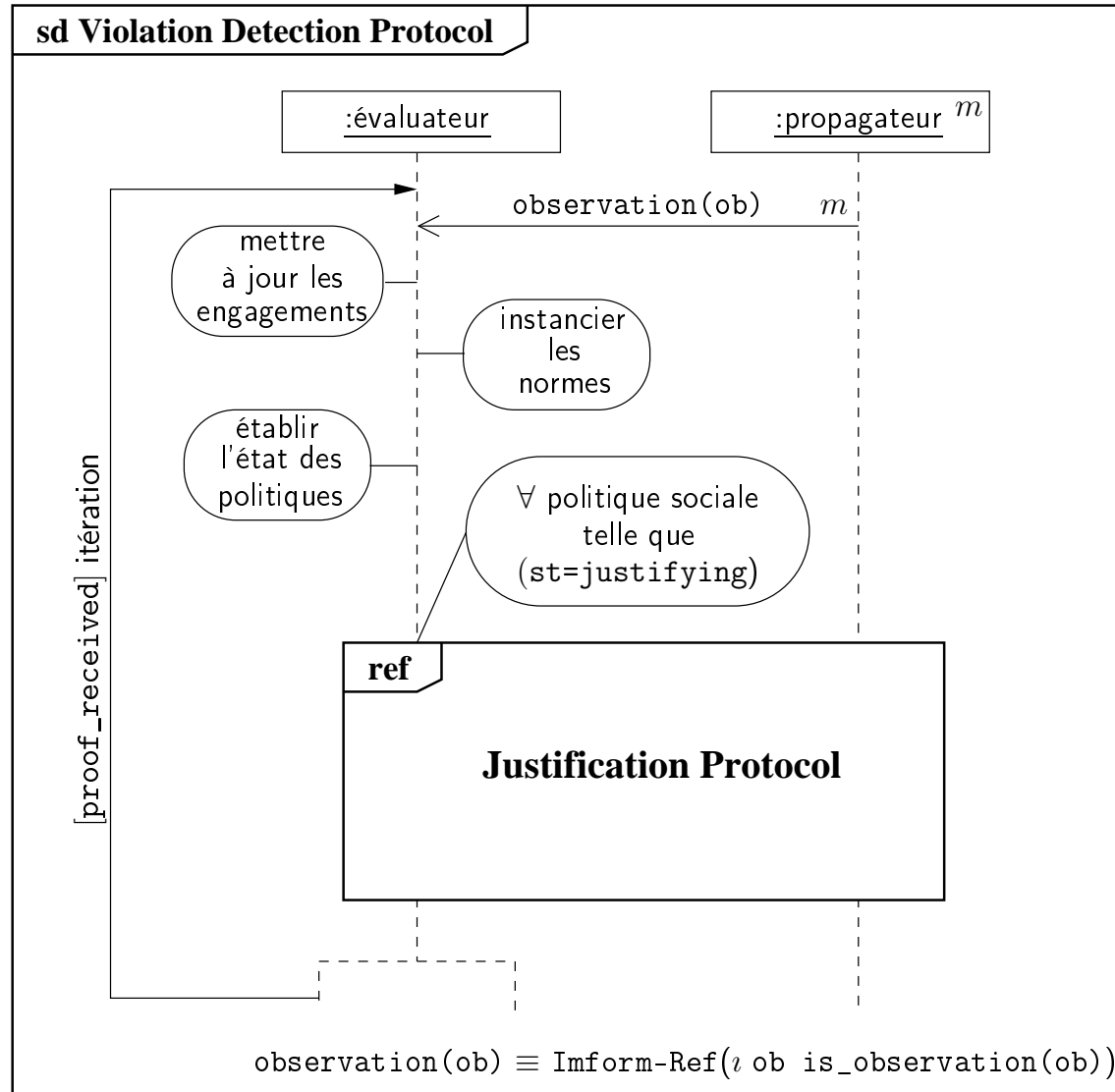
e : engagement social pris par x et perçu par ev
 ${}_{ev}SPol(x, ev, 13 :00, active, x \in G(t) \wedge ev \in Ev,$
 $\neg(ev.facets(e.cont) \supseteq \{ "theater showtimes" \}))$



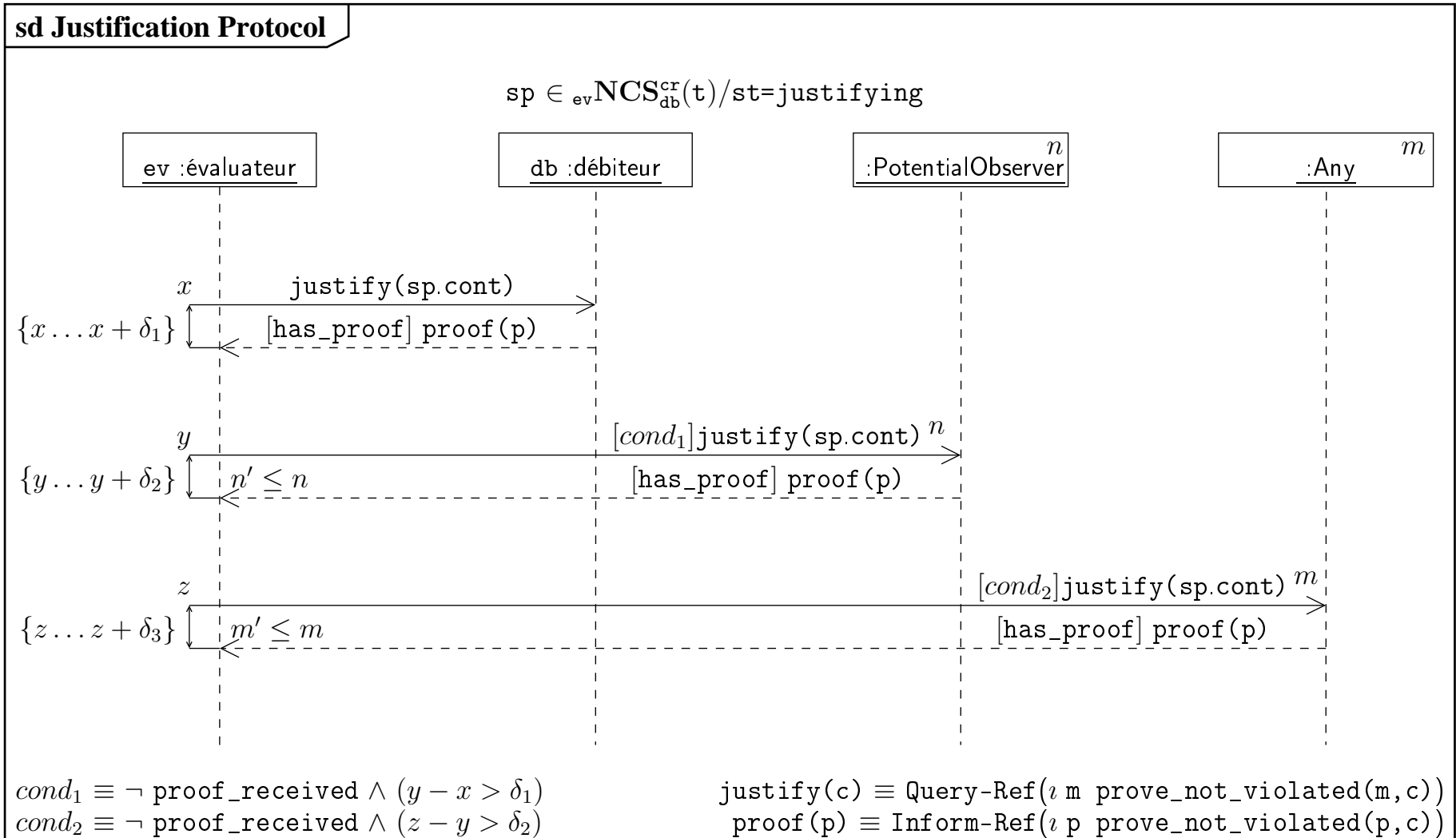
Cycle de vie d'une politique sociale



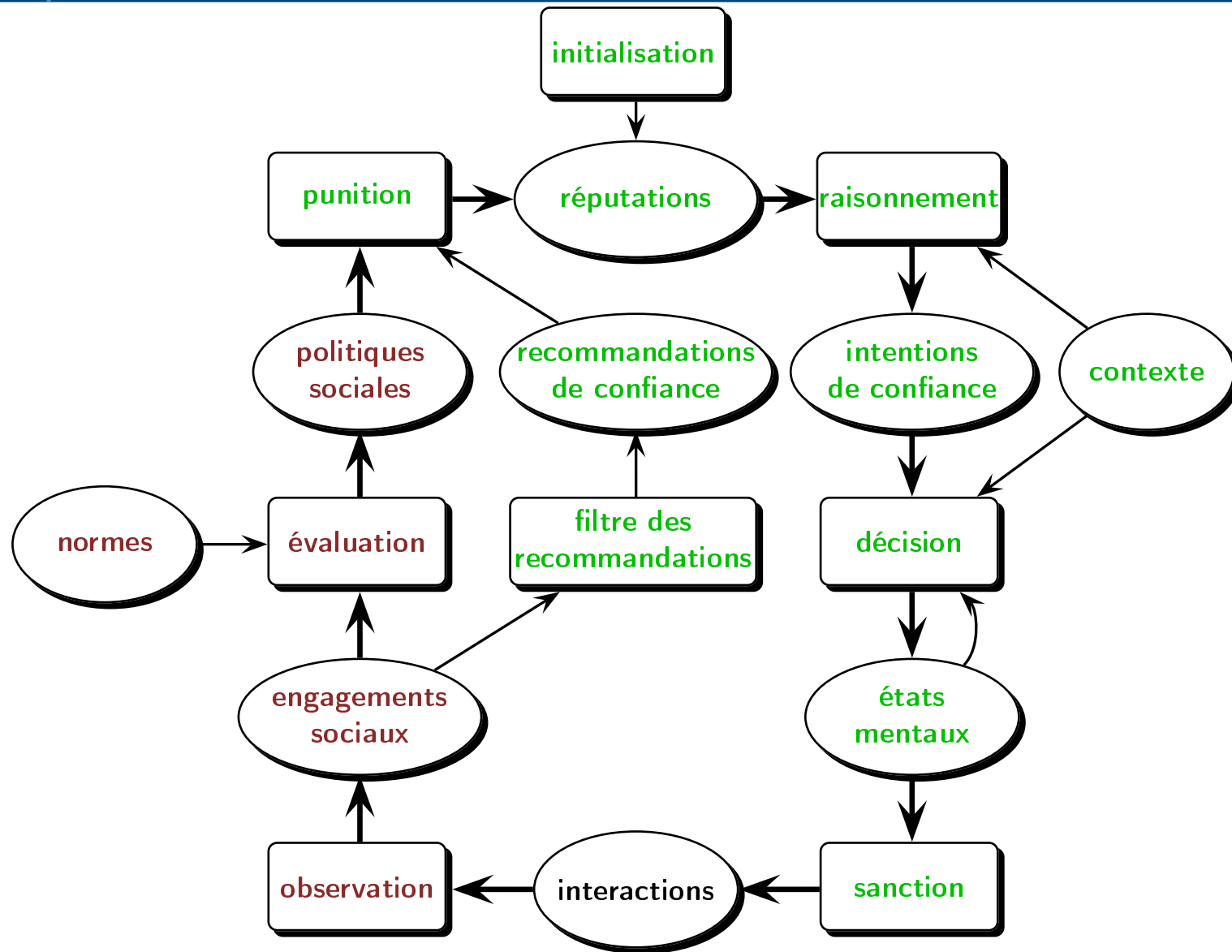
Détection de violations – Processus Général



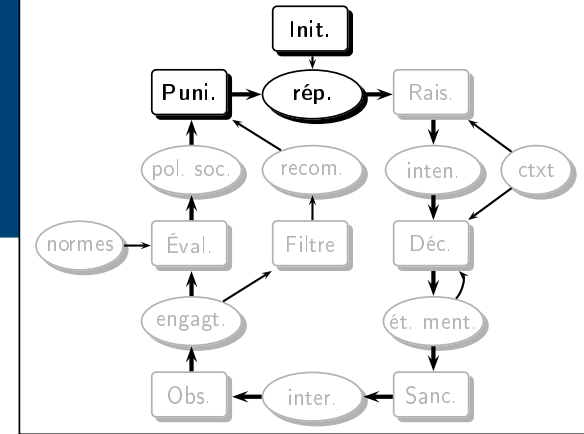
Détection de violations – Processus de Justification



Architecture du modèle L.I.A.R.



Types de Réputation



$R_{\text{bénéficiaire}}^{\text{cible}}(\text{facette, dimension, instant}) \in [-1, +1] \cup \{\text{unknown}\}$

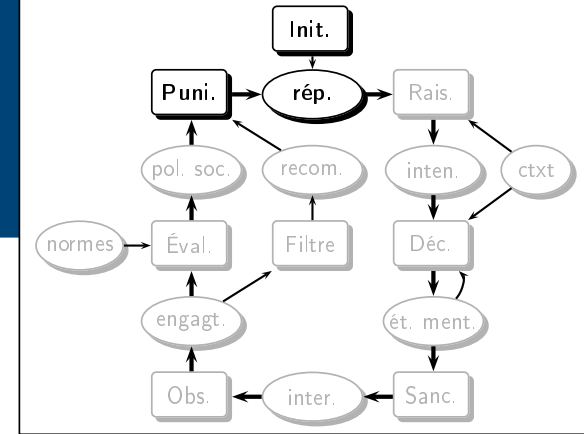
7 rôles (extension de [Conte02]) :

- ✓ Cible
- ✓ Participant
- ✓ Observateur
- ✓ Évaluateur
- ✓ Punisseur
- ✓ Bénéficiaire
- ✓ Propagateur

5 types de réputation, fondée sur :

- ✓ Interaction directe (DIbRp)
- ✓ Interaction indirecte (IIbRp)
- ✓ Recommandation d'observation (ObsRcbRp)
- ✓ Recommandation d'évaluation (EvRcbRp)
- ✓ Recommandation de réputation (RpRcbRp)

Types de Réputation



$R_{\text{bénéficiaire}}^{\text{cible}}(\text{facette, dimension, instant}) \in [-1, +1] \cup \{\text{unknown}\}$

7 rôles (extension de [Conte02]) :

- ✓ Cible
- ✓ Participant

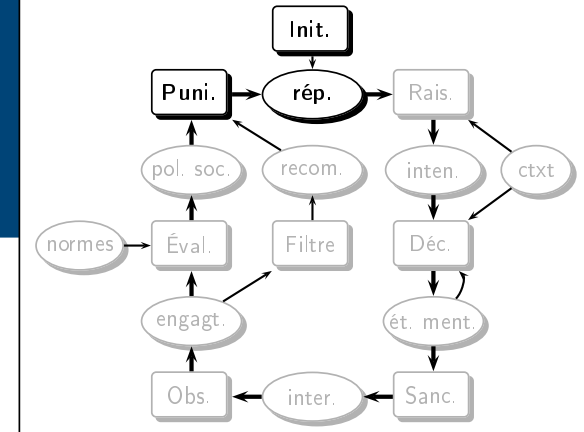
- ✓ Observateur
- ✓ Évaluateur
- ✓ Punisseur
- ✓ Bénéficiaire
- ✓ Propagateur

5 types de réputation, fondée sur :

- ✓ **Interaction directe** (DIbRp)
- ✓ Interaction indirecte (IIbRp)

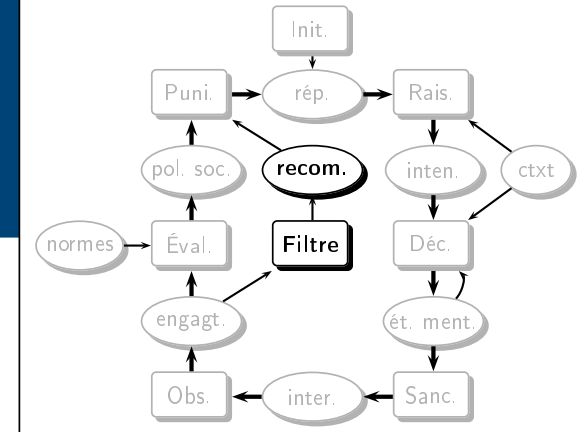
- ✓ Recommandation d'observation (ObsRcbRp)
- ✓ Recommandation d'évaluation (EvRcbRp)
- ✓ **Recommandation de réputation** (RpRcbRp)

Processus de Puniton



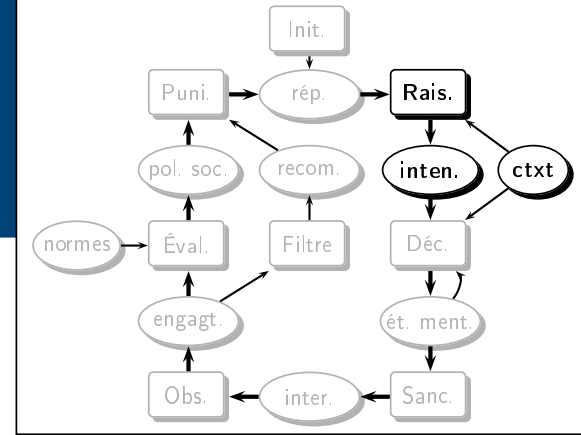
- ✓ par la Réputation fondée sur les Interactions Directes
- ✗ (Établir l'état des politiques sociales)
- ✗ Séparer les ensembles de politiques sociales par état
- ✗ Associer une valeur à chaque ensemble grâce aux pénalités
- ✗ Réputation = Moyenne pondérée des valeurs

Processus de Puniton

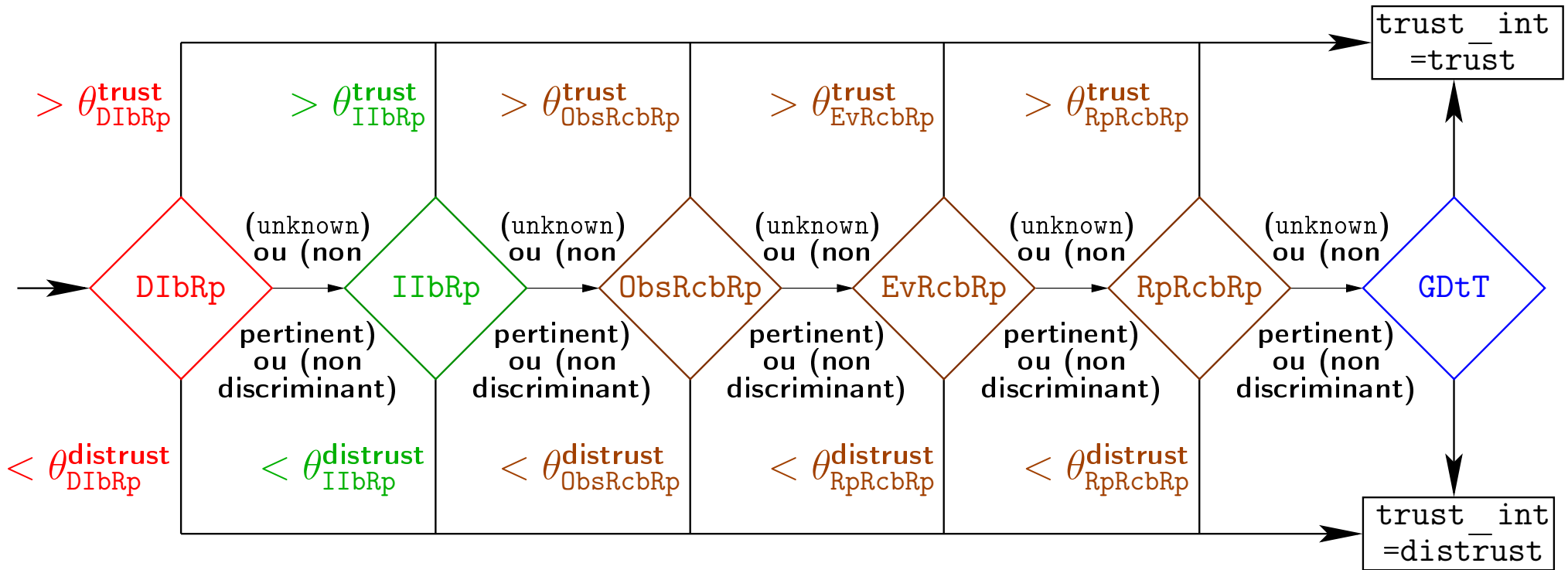


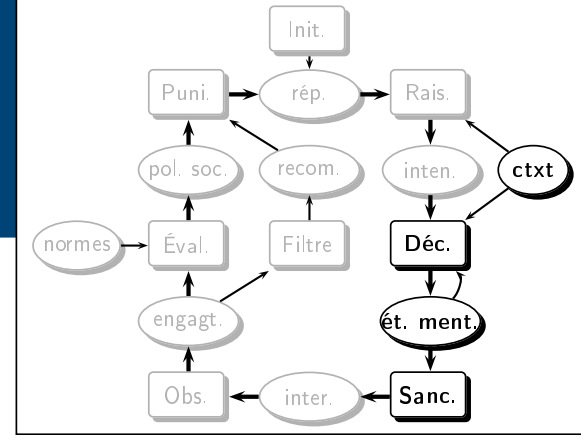
- ✓ par la Réputation fondée sur les Interactions Directes
 - ✗ (Établir l'état des politiques sociales)
 - ✗ Séparer les ensembles de politiques sociales par état
 - ✗ Associer une valeur à chaque ensemble grâce aux pénalités
 - ✗ Réputation = Moyenne pondérée des valeurs

- ✓ par la Réputation fondée sur les Recommandations de Réputation
 - ✗ Filtrage par la réputation du propagateur/recommandation/intégrité
 - ✗ Réputation = Moyenne pondérée des niveaux reçus
 - ✗ Pondération par la réputation du punisseur/recommandation/compétence

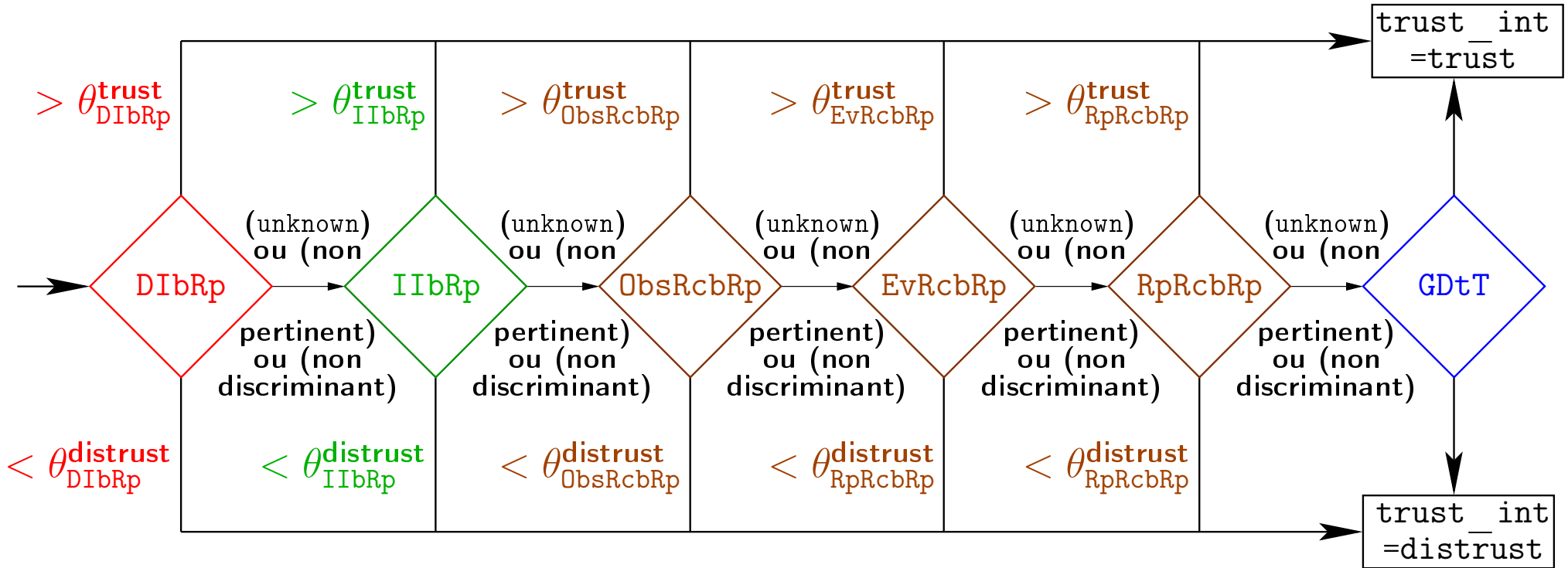


✓ Raisonnement :

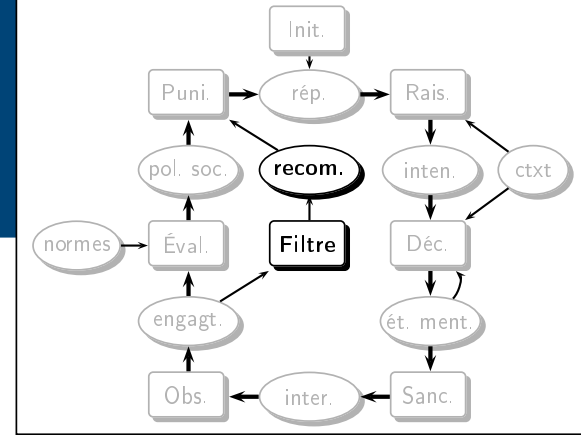




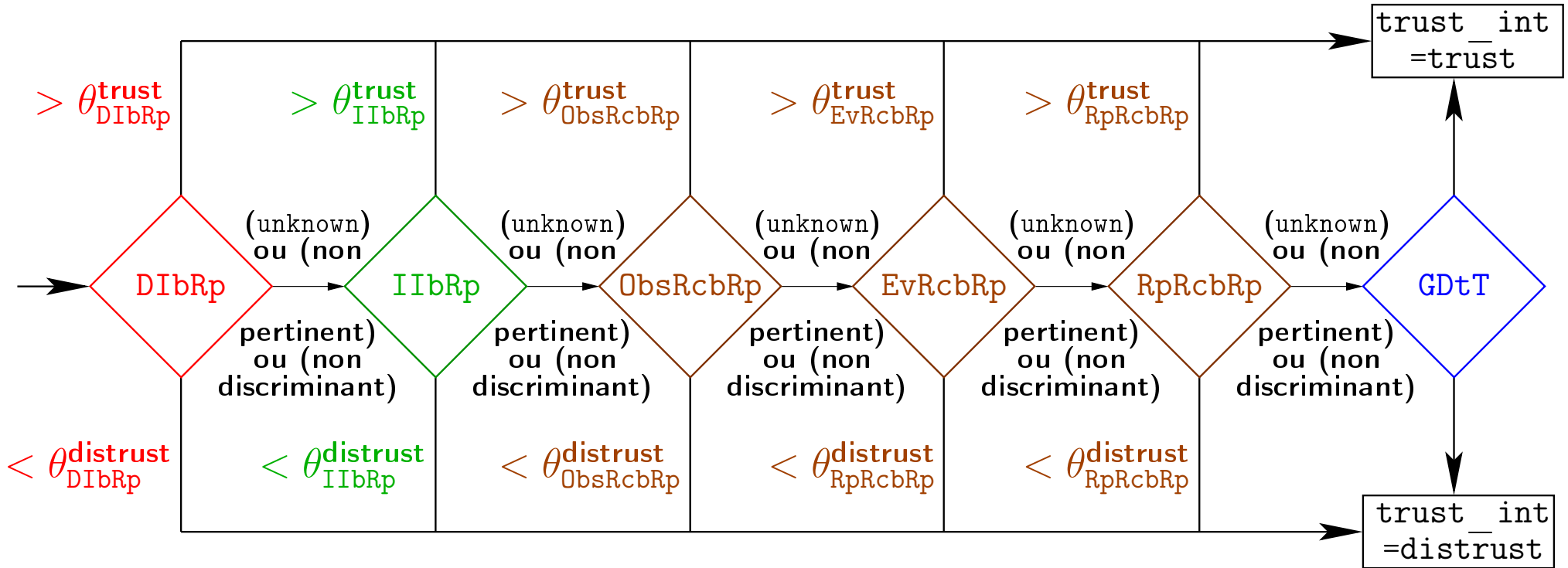
✓ Raisonnement :



✓ Décision : sélection ou tri



✓ Raisonnement :



✓ Décision : sélection ou tri

✓ Propagation : « push » / « pull »

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Scénario

Grille

Résultats

Synthèse

Conclusion

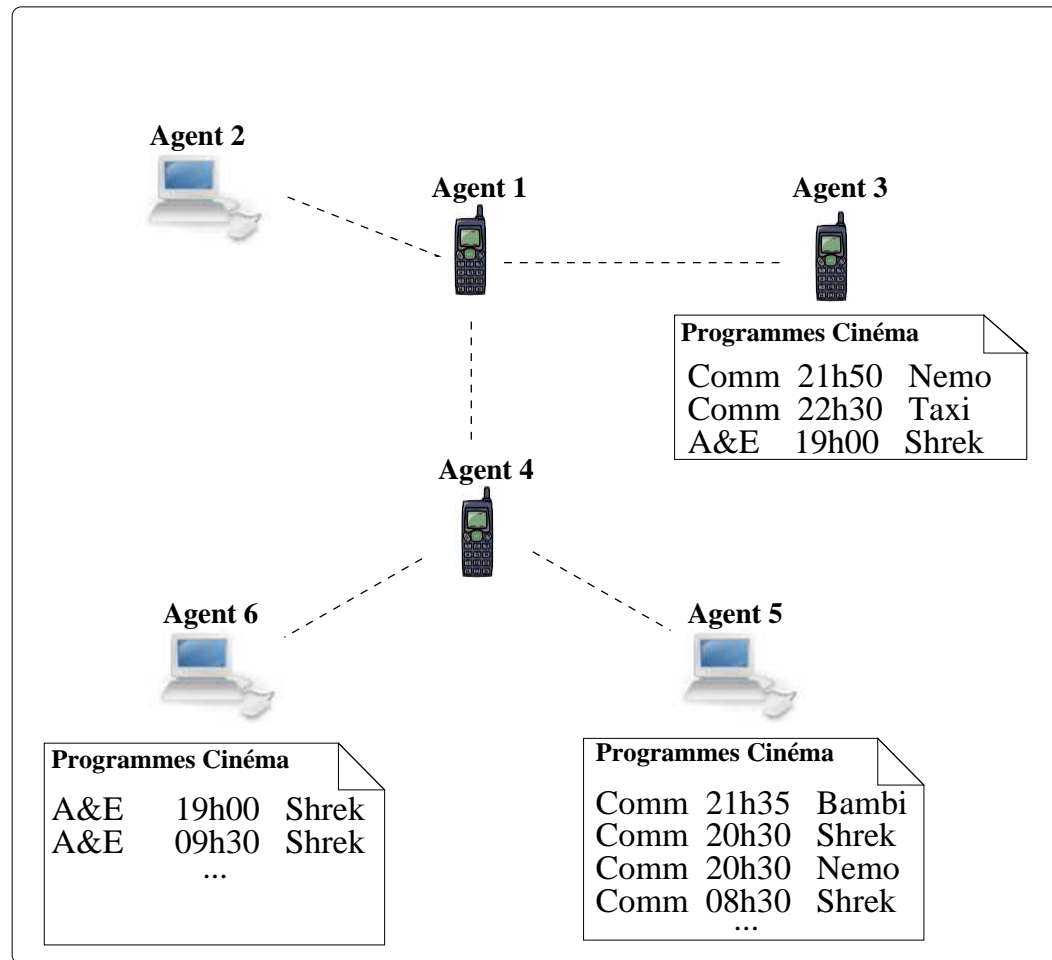
État de l'Art

LIAR

Expérimentations

Régulation des communications dans un réseau P2P

- ✓ Réseau pair-à-pair (P2P) « agentifié »
- ✓ Les agents implémentent le modèle de **réputation** L.I.A.R.
- ✓ **Engagements sociaux** pris / observés par les agents

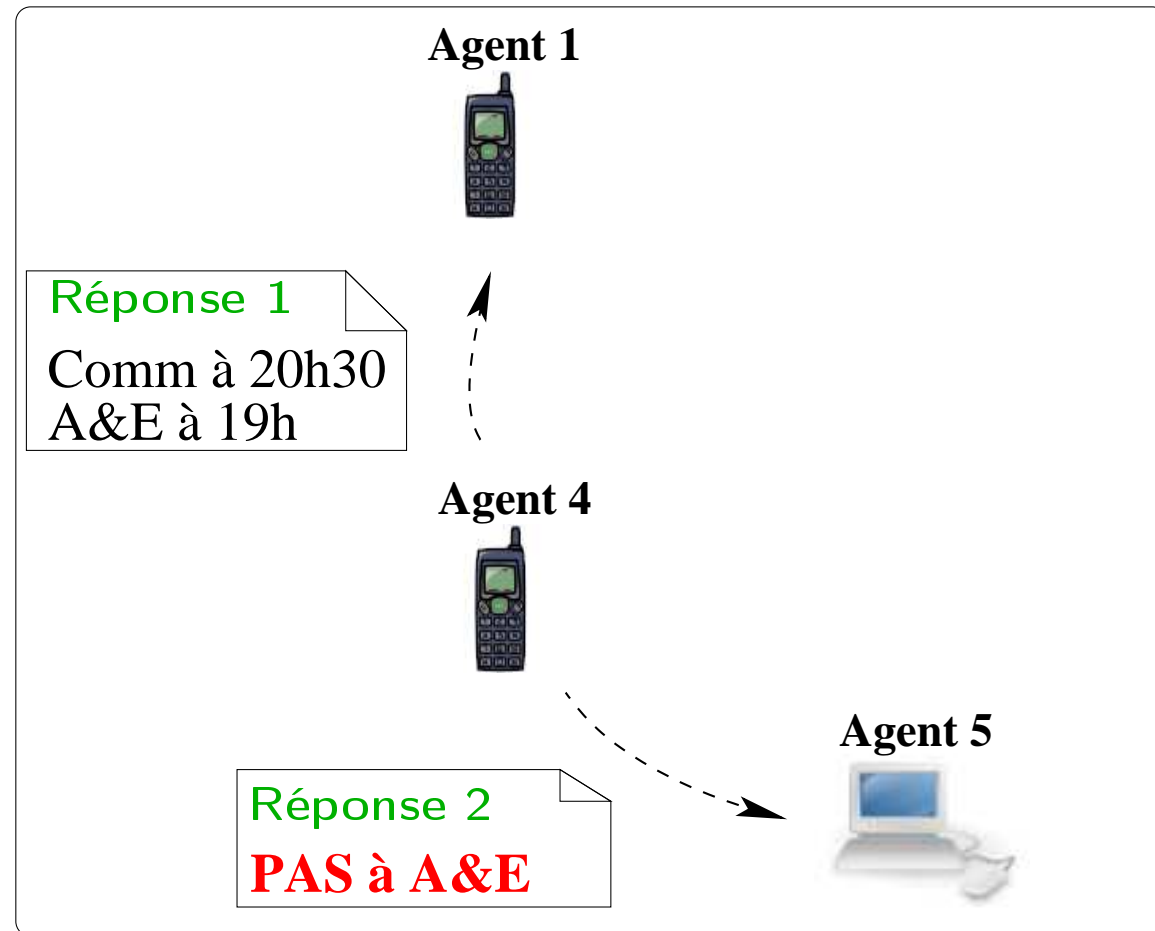


- Introduction
- État de l'Art
- Modèle L.I.A.R.
- Expérimentations**
- Scénario**
- Grille
- Résultats
- Synthèse
- Conclusion
- État de l'Art
- LIAR

Régulation des communications dans un réseau P2P

- ✓ Normes interdisent les contradictions

Contradiction en émission



Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Scénario

Grille

Résultats

Synthèse

Conclusion

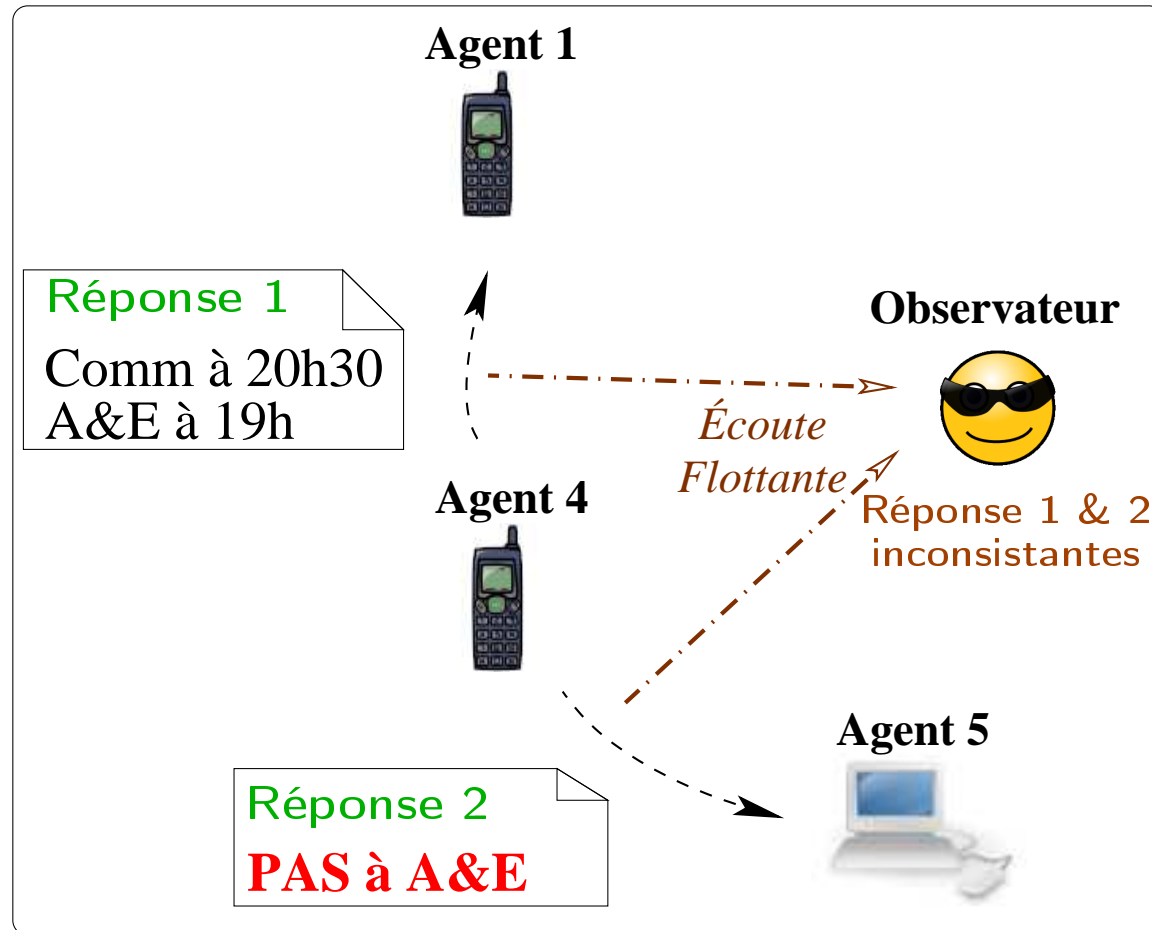
État de l'Art

LIAR

Régulation des communications dans un réseau P2P

- ✓ Normes interdisent les contradictions

Contradiction en émission

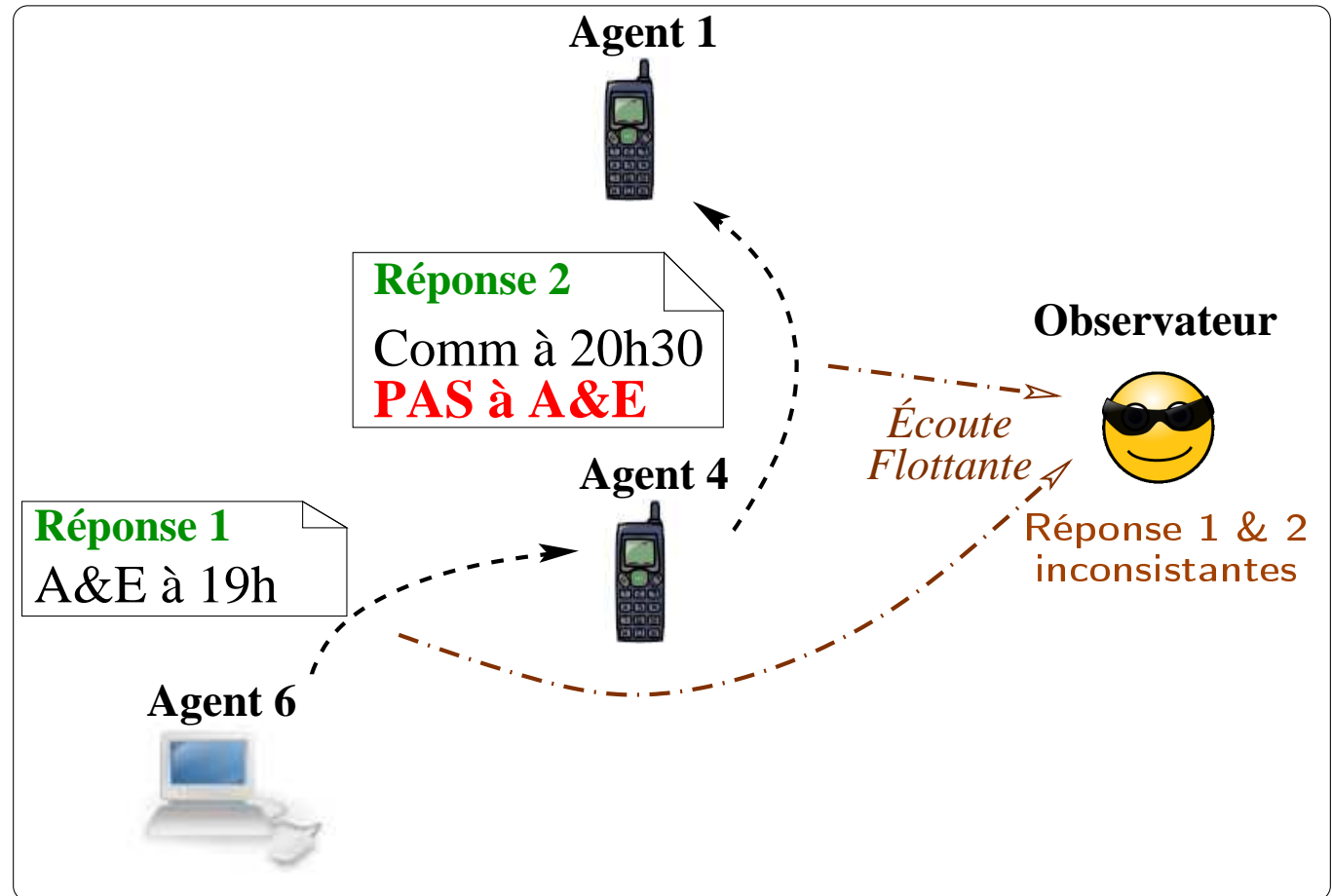


- Introduction
- État de l'Art
- Modèle L.I.A.R.
- Expérimentations**
- Scénario**
- Grille
- Résultats
- Synthèse
- Conclusion
- État de l'Art
- LIAR

Régulation des communications dans un réseau P2P

✓ Normes interdisent les contradictions

Contradiction en transmission



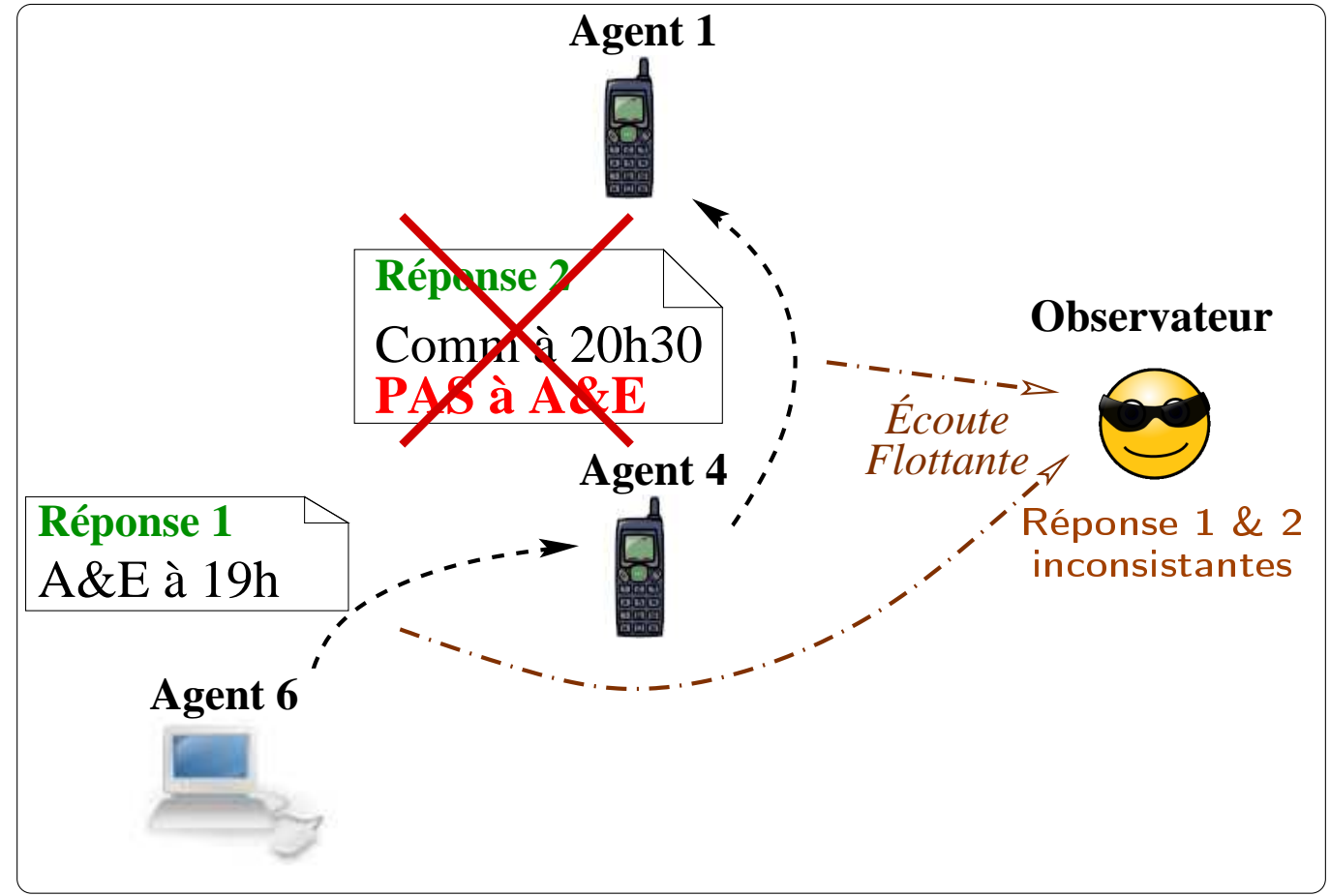
- Introduction
- État de l'Art
- Modèle L.I.A.R.
- Expérimentations**
- Scénario**
- Grille
- Résultats
- Synthèse
- Conclusion
- État de l'Art
- LIAR

Régulation des communications dans un réseau P2P

- Introduction
- État de l'Art
- Modèle L.I.A.R.
- Expérimentations**
- Scénario**
- Grille
- Résultats
- Synthèse
- Conclusion
- État de l'Art
- LIAR

✓ Normes interdisent les contradictions

Stratégie d'instanciation



✓ Stratégie d'instanciation : indulgente vs. rancunière

Grille d'expérimentations

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Scénario

Grille

Résultats

Synthèse

Conclusion

État de l'Art

LIAR

✓ Réputations

✗ *Multi-★*

✗ Rapidement convergent [Ding04]

✗ Précis [Fullam03, Klos04, Whitby04]

✗ Adaptatif [Fullam04]

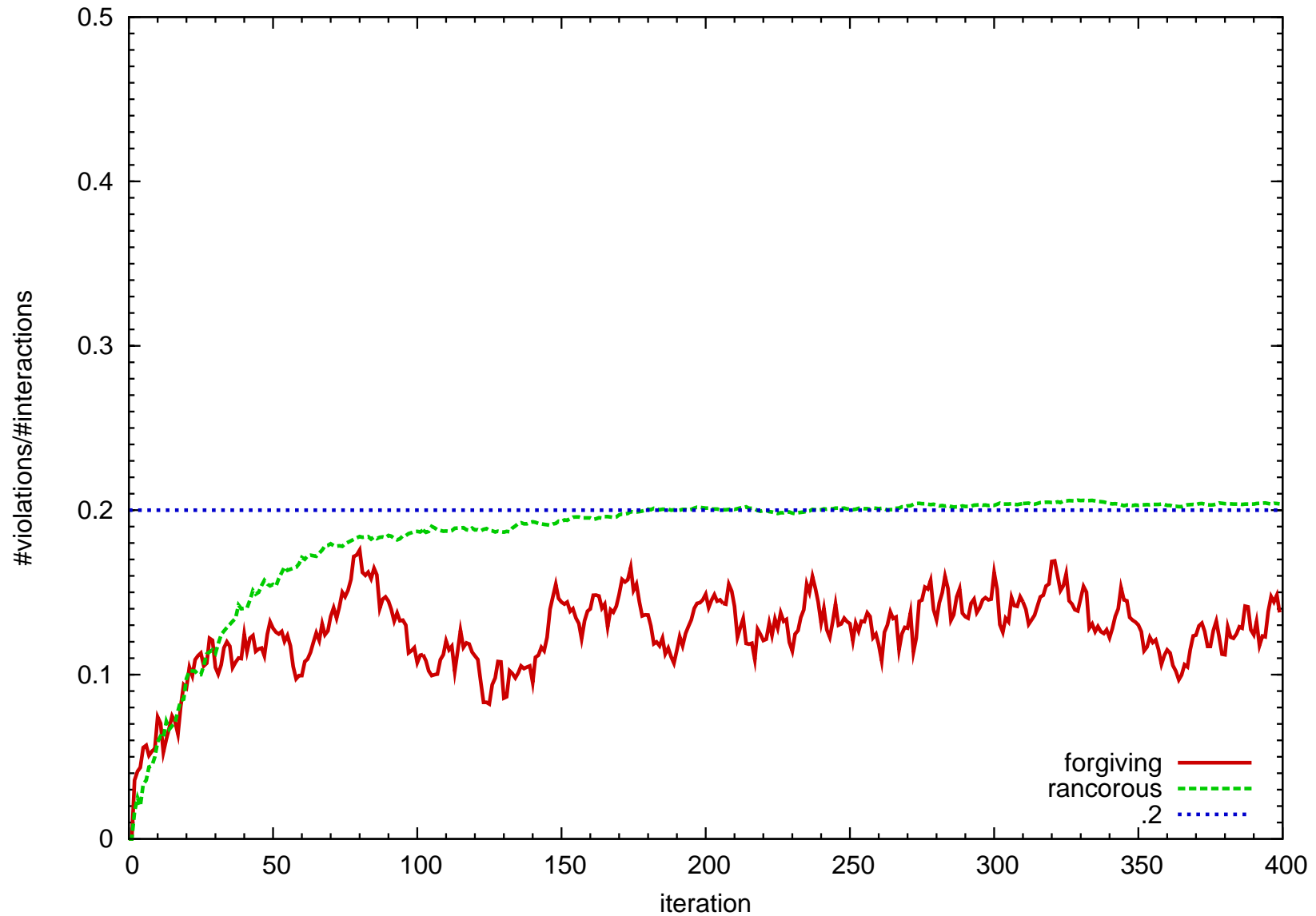
✗ Efficace [Ghanea04, Yamamoto04]

✓ Décision

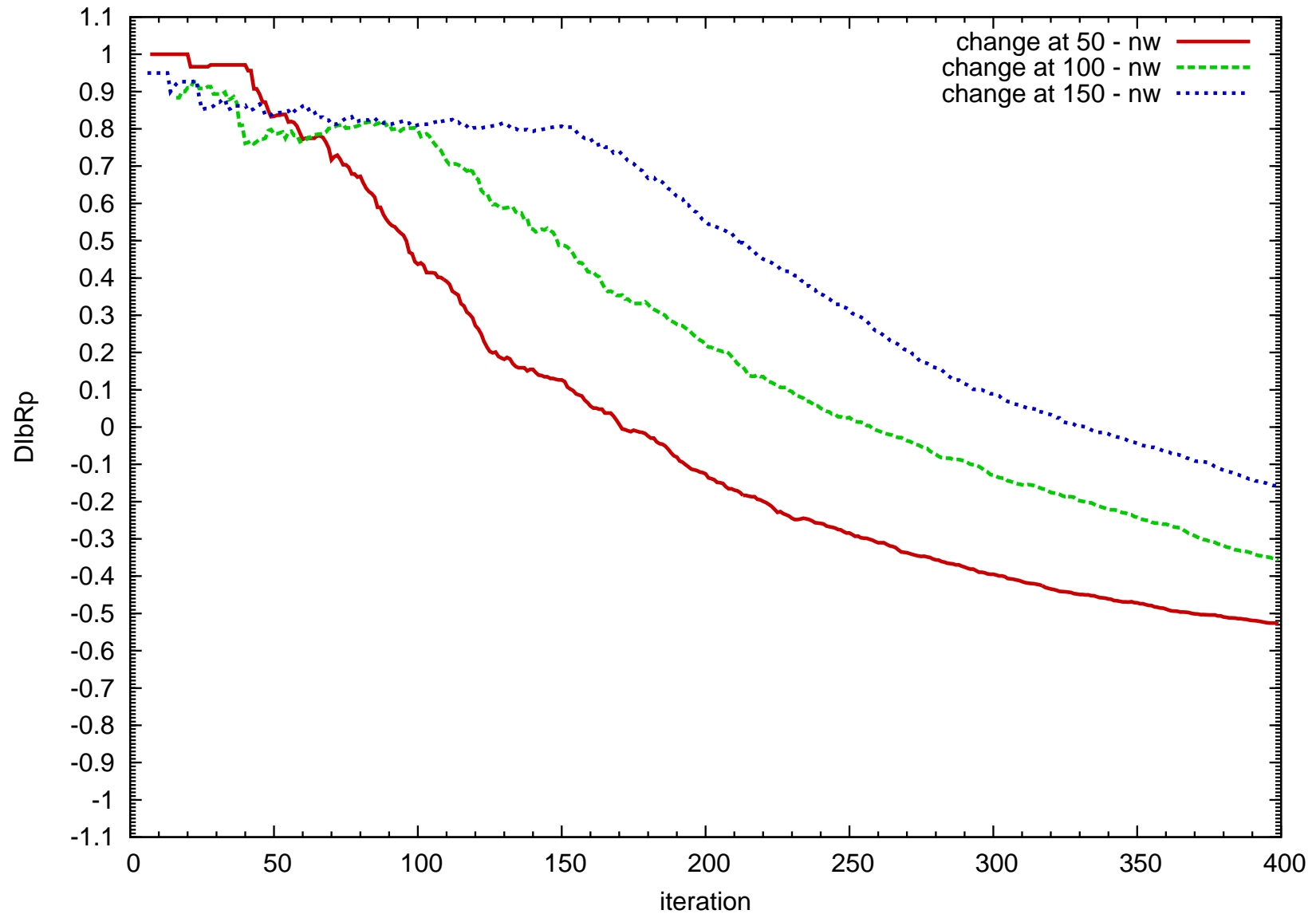
✗ Identifier et isoler les agents malveillants
[Barber02, Biswas00]

✗ Décider si et avec qui interagir [Falcone04, Schillo00]

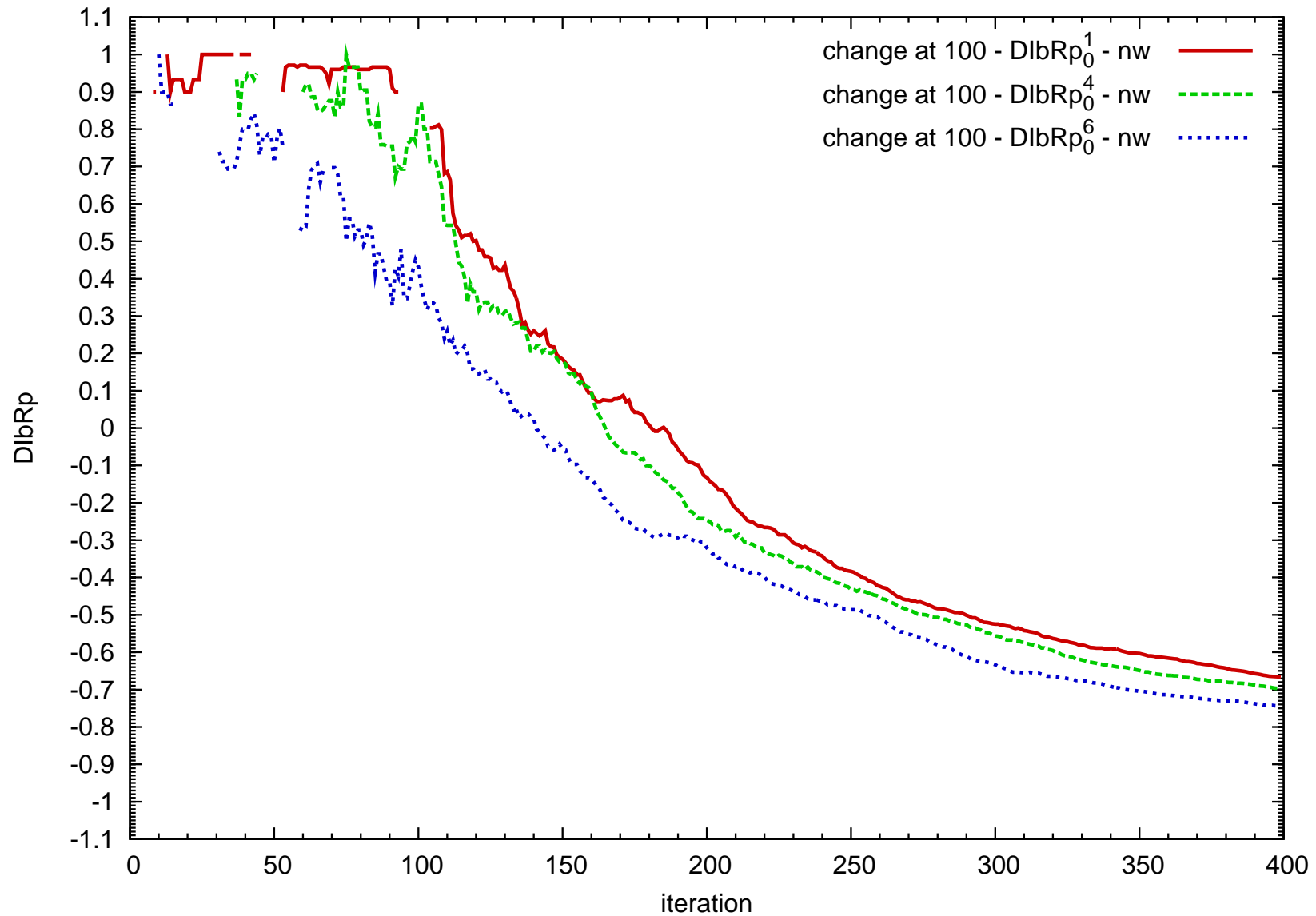
Évaluation du modèle : précision et convergence



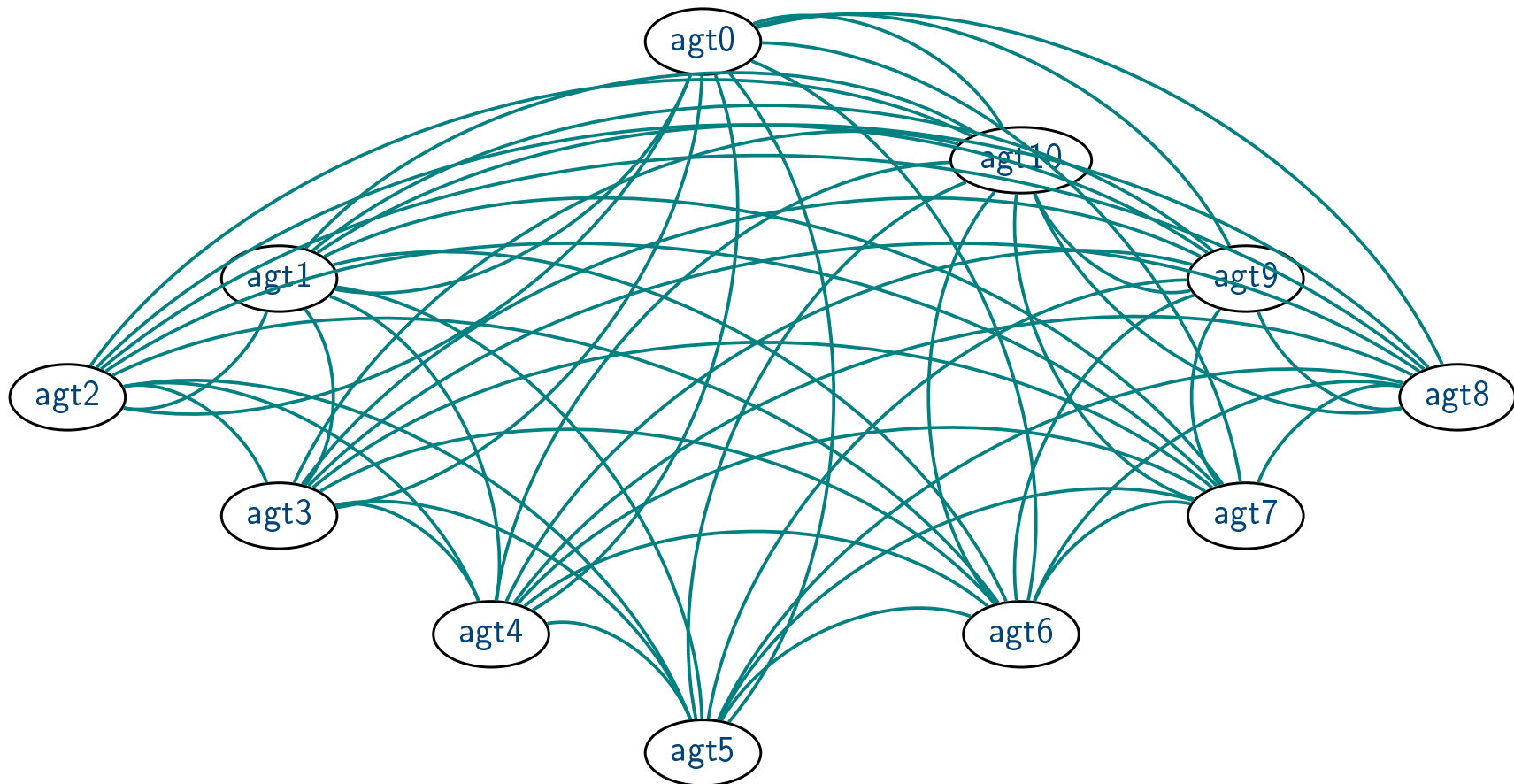
Évaluation du modèle : inertie



Évaluation du modèle : fragilité

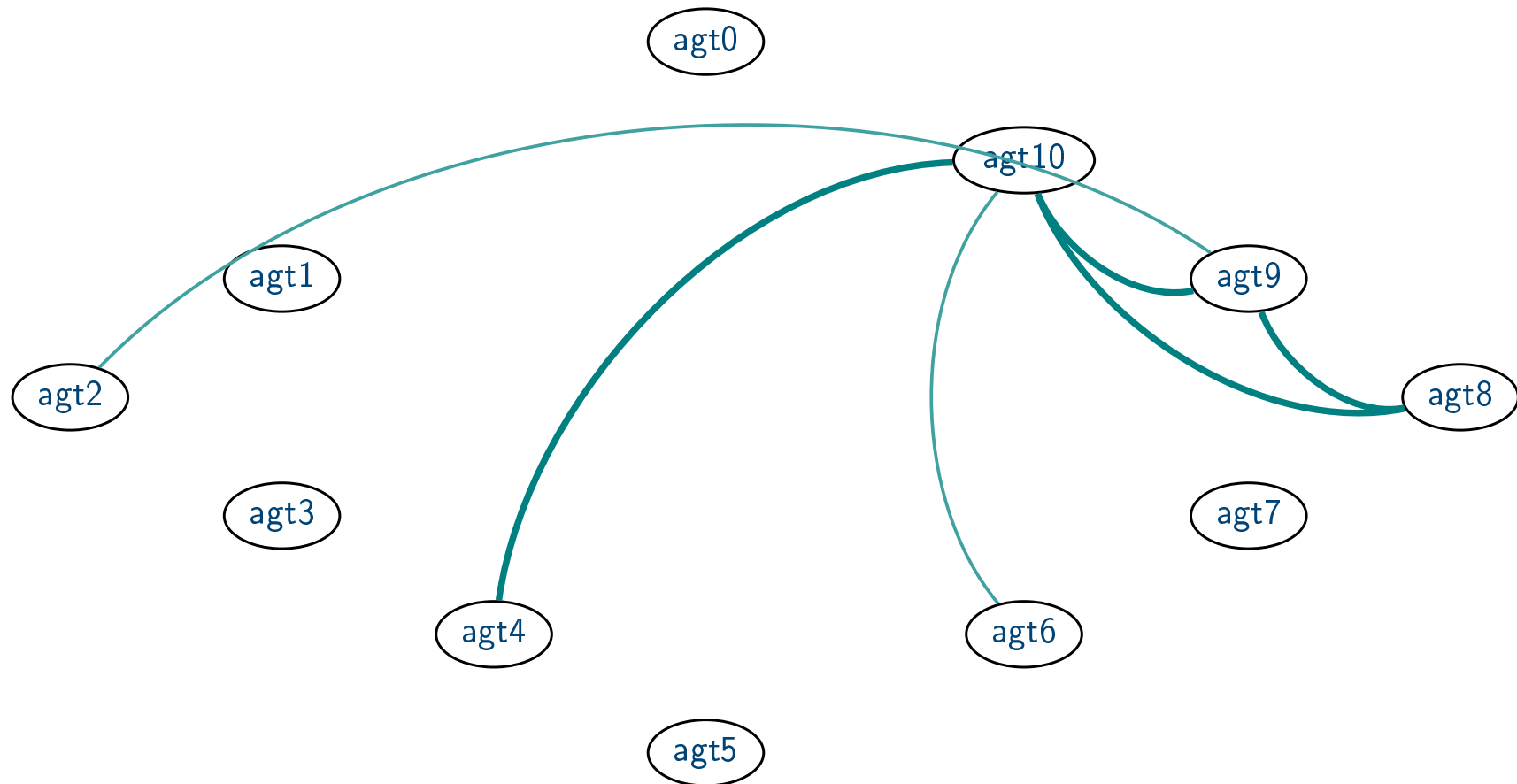


Évaluation du modèle : décision



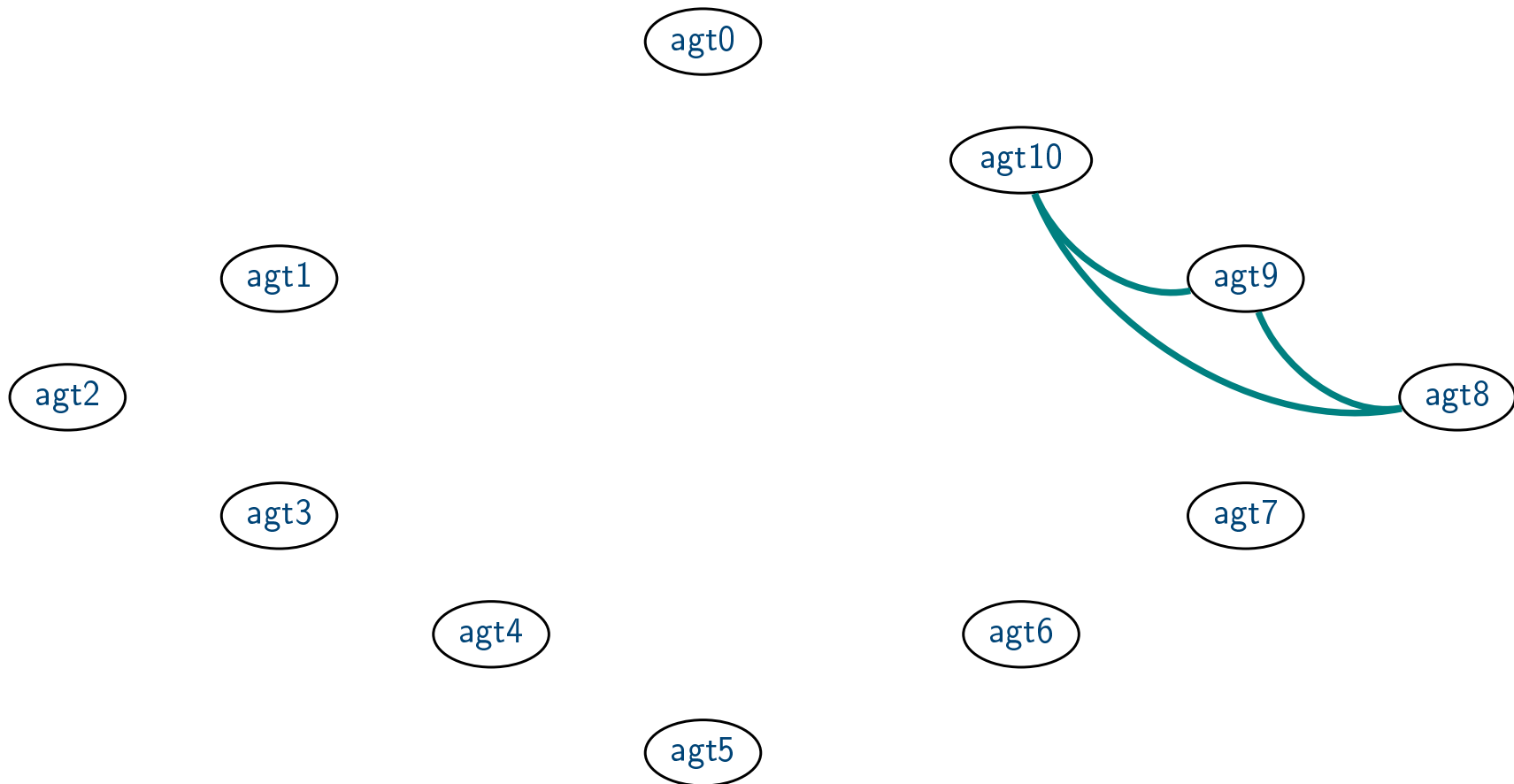
Étape 1

Évaluation du modèle : décision



Étape 50

Évaluation du modèle : décision



Étape 100

Expérimentations

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Scénario

Grille

Résultats

Synthèse

Conclusion

État de l'Art

LIAR

✓ Réputations

- ✗ Rapidement convergent
dépend de la stratégie d'instanciation et du type de réputation
- ✗ Précis
avec la stratégie rancunière
- ✗ Adaptatif
inertie et de fragilité paramétrables
- ✗ Efficace
temps de calcul linéaire (empirique)

✓ Décision

permet aux agents de :

- ✗ **identifier et isoler les agents malveillants**
- ✗ **décider si et avec qui interagir**

Conclusion

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

Contributions

Perspectives

État de l'Art

LIAR

- ✓ Modèle d'engagement social **adapté aux SMAOD**
- ✓ Modèle de normes et de politiques sociales **intanciation, placement des sanctions, adapté aux SMAOD**
- ✓ Processus d'évaluation **explicite, adapté aux SMAOD**
- ✓ Modèle de réputation **analyse approfondie de l'état de l'art, finesse sémantique, intégration de tous les processus**

Conclusion

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

Contributions

Perspectives

État de l'Art

LIAR

- ✓ Modèle d'engagement social
adapté aux SMAOD
- ✓ Modèle de normes et de politiques sociales
intanciation, placement des sanctions, adapté aux SMAOD
- ✓ Processus d'évaluation
explicite, adapté aux SMAOD
- ✓ Modèle de réputation
analyse approfondie de l'état de l'art, finesse sémantique, intégration de tous les processus



**Contrôle social des interactions entre agents dans un
SMAOD**

Limites et perspectives

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

Contributions

Perspectives

État de l'Art

LIAR

✓ Théoriques

- ✗ Enrichir le processus de justification avec l'argumentation
- ✗ Prendre en compte la présence d'institutions [Grizard06]
- ✗ Prendre en compte l'hétérogénéité des modèles de réputation

✓ Applicatives

- ✗ Trouver des applications réelles en adéquation avec la richesse du modèle
- ✗ Établir des benchmarks [ARTTestbed04]

Productions Bibliographiques – de la thèse

- G. Muller, L. Vercouter & O. Boissier. Towards a general definition of trust and its application to openness in MAS. In Workshop on "Deception, Fraud and Trust in Agent Societies" at AAMAS'03, pages 49–56, 2003
- G. Muller & L. Vercouter. Liar Detection Within Agent Communication. In Workshop LCMAS'04, pages 4–16, 2004
- G. Muller & L. Vercouter. Détection Décentralisée d'Agents menteurs. In Actes des JFSMA'04, pages 243–248, 2004
- G. Muller & L. Vercouter. Decentralized Monitoring of Agent Communications with a Reputation Model. In Workshop on "Trust in Agent Societies" at AAMAS'04, volume LNCS-3577 of Trusting Agents for trusting Electronic Societies, pages 144–161. Springer-Verlag, 2005
- G. Muller & L. Vercouter. Using Social Commitments to Control the Agents' Freedom of Speech. In Workshop on "Agent Communications, Languages and Conversation Policies" at AAMAS'05, volume LNAI-3859, pages 109–123, Utrecht, The Netherlands, July 2005
- G. Muller, L. Vercouter & O. Boissier. A Trust Model for the Reliability of Agent Communications. In Workshop on "Trust in Agent Societies" at AAMAS'05, 2005
- A. Grizard, L. Vercouter, T. Stratulat & G. Muller. A peer-to-peer normative system to achieve social order. In Workshop on "Coordination, Organization, Institutions and Norms" (COIN) at AAMAS'06, volume LNCS (in press), 2006

Productions Bibliographiques – associées

K. K. Fullam, T. B. Klos, G. Muller, J. Sabater, Z. Topol, K. S. Barber, J. S. Rosenschein & L. Vercouter. Le banc d'essais ART (Agent Reputation and Trust) pour les modèles de confiance. In Actes des JFSMA'05, pages 175–179, 2005

K. Fullam, T. Klos, G. Muller, J. Sabater, Z. Topol, K. S. Barber, J. S. Rosenschein & L. Vercouter. The Agent Reputation and Trust (ART) Testbed Architecture. In Workshop on "Trust in Agent Societies" at AAMAS'05, pages 50–62, 2005

K. Fullam, T. Klos, G. Muller, J. Sabater, A. Schlosser, Z. Topol, K. S. Barber, J. S. Rosenschein, L. Vercouter & M. Voss. A Specification of the Agent Reputation and Trust (ART) Testbed : Experimentation and Competition for Trust in Agent Societies. In AAMAS'05, 2005

K. Fullam, T. Klos, G. Muller, J. Sabater, A. Schlosser, Z. Topol, K. S. Barber, J. S. Rosenschein, L. Vercouter & M. Voss. A Demonstration of The Agent Reputation and Trust (ART) Testbed : Experimentation and Competition for Trust in Agent Societies. In AAMAS'05, 2005

K. K. Fullam, T. B. Klos, G. Muller, J. Sabater i Mir, Z. Topol, K. S. Barber, J. S. Rosenschein & L. Vercouter. The Agent Reputation and Trust (ART) Testbed Architecture. In CCIA'05, volume 131 of Frontiers in Artificial Intelligence and Applications, pages 389–396, 2005

K. K. Fullam, T. B. Klos, G. Muller, J. Sabater-Mir, K. S. Barber & L. Vercouter. The Agent Reputation and Trust (Art) Testbed. In iTrust'06, volume LNCS-3986, pages 439–442, 2006

Karen K. Fullam, Tomas B. Klos, Guillaume Muller, Jordi Sabater-Mir, K. Suzanne Barber & Laurent Vercouter. The Agent Reputation and Trust (ART) Testbed. In BNAIC'06, pages 449–450, 2006

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

Annexes

État de l'Art

LIAR

Annexes

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

SCom

Confiance

LIAR

État de l'Art

Sémantique des communications

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

SCom

Confiance

LIAR

Il existe différentes approches pour la spécification des langages de communication entre agent (ACL) [Guerin02] :

- ✓ Langue naturelle
 - ⊖ *trop ambiguë*
- ✓ Protocoles [Smith80, Durfee90, Campbell90]
 - ⊖ *trop rigide, seul l'enchaînement compte*
- ✓ Mentalistique [Cohen95, Finin93, FIPA02]
 - ⊖ *hypothèses sur l'implémentation interne*
- ✓ Sémantique sociale [Singh91, Castelfranchi95]
formel, flexible, externe
- ✓ Argumentative [Dung94, Amgoud02]
 - ⊖ *modélise le dialogue (s'appuie sur la sémantique sociale)*

Types de Normes

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

SCom

Confiance

LIAR

- ✓ Les **normes** permettent de définir l'acceptabilité
- ✓ Différents types de **normes sociales**
[Tuomela95, Wright93, Dignum00, Lopez02] :
 - ✗ *r-normes* :
 - ✓ Règles juridiques, plutôt écrites : lois, codes...
 - ✓ Agent ↔ Autorité
 - ✓ Autorité définit, détecte violations, sanctionne (imposer)
 - ✓ Sanctions explicites, connues *a priori* et toujours appliquées
 - ✗ *s-normes* :
 - ✓ Standards de fait, plutôt orales : politesse...
 - ✓ Agent ↔ Groupe social
 - ✓ Groupe définit, détecte, sanctionne (émerger)
 - ✓ Sanctions implicites, non connues *a priori*, pas toujours appliquées

Formalismes de Norme

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

SCom

Confiance

LIAR

- ✓ Logiques déontiques
[Dignum99AAwithNorms]
problèmes : décidabilité, richesse d'expression
- ✓ Formalismes descriptifs
[Sallee02,Stratulat02,Lopez04,Vazquez05]
avantages : logique 1^{er} ordre, richesse d'expression
- ✓ Politiques Sociales
[Singh99,Fornara05]
engagements sociaux, richesse d'expression
- ✓ Implémentation : Enrégimentation vs. Autonomie/Normes

Classes de Confiance

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

SCom

Confiance

LIAR

- ✓ Dispositionnelle [McKnight01]
- ✓ Institutionnelle [McKnight01]
- ✓ Interpersonnelle [McKnight01]
- ✓ De groupe [Dasgupta00, Quere03]
- ✓ Dans les objets, lieux, événements, activités [Casare05]

Classes de Confiance

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

SCom

Confiance

LIAR

- ✓ Dispositionnelle [McKnight01]
 - ✓ Institutionnelle [McKnight01]
 - ✓ Interpersonnelle [McKnight01]
 - ✓ De groupe [Dasgupta00, Quere03]
 - ✓ Dans les objets, lieux, événements, activités [Casare05]
-] [McKnight01]

Classes de Confiance

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

SCom

Confiance

LIAR

- ✓ Dispositionnelle [McKnight01]
 - ✓ Institutionnelle [McKnight01]
 - ✓ Interpersonnelle [McKnight01]
- [McKnight01]
- ✓ De groupe [Dasgupta00, Quere03]
 - ✓ Dans les objets, lieux, événements, activités [Casare05]

Classes de Confiance

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations


Conclusion

État de l'Art

SCom

Confiance

LIAR

- ✓ Dispositionnelle [McKnight01]
- ✓ Institutionnelle [McKnight01] 
- ✓ Interpersonnelle [McKnight01] ← [Rouchier00]
[McKnight01]
- ✓ De groupe [Dasgupta00, Quere03]
- ✓ Dans les objets, lieux, événements, activités [Casare05]

Classes de Confiance

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations


Conclusion

État de l'Art

SCom

Confiance

LIAR

- ✓ Dispositionnelle [McKnight01]
 - ✓ Institutionnelle [McKnight01]
 - ✓ Interpersonnelle [McKnight01]
 - ✓ De groupe [Dasgupta00, Quere03]
 - ✓ Dans les objets, lieux, événements, activités [Casare05]
- [Castelfranchi98]
- 

Classes de Confiance

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

SCom

Confiance

LIAR

- ✓ Dispositionnelle [McKnight01]
 - ✓ Institutionnelle [McKnight01]
 - ✓ Interpersonnelle [McKnight01]
 - ✓ De groupe [Dasgupta00, Quere03]
 - ✓ Dans les objets, lieux, événements, activités [Casare05]
- [Brousseau00]
[Tuomela95]

Classes de Confiance

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations


Conclusion

État de l'Art

SCom

Confiance

LIAR

- ✓ Dispositionnelle [McKnight01]
 - ✓ Institutionnelle [McKnight01]
 - ✓ Interpersonnelle [McKnight01]
 - ✓ De groupe [Dasgupta00, Quere03]
 - ✓ Dans les objets, lieux, événements, activités [Casare05]
- [Castelfranchi98]
- 

Classes de Confiance

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

SCom

Confiance

LIAR

- ✓ Dispositionnelle [McKnight01]
- ✓ Institutionnelle [McKnight01]
- ✓ Interpersonnelle [McKnight01] — [Dasgupta00]
- ✓ De groupe [Dasgupta00, Quere03] — [Dasgupta00]
- ✓ Dans les objets, lieux, événements, activités [Casare05]

Classes de Confiance

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

SCom

Confiance

LIAR

- ✓ Dispositionnelle [McKnight01]
- ✓ Institutionnelle [McKnight01]
- ✓ Interpersonnelle [McKnight01]
- ✓ De groupe [Dasgupta00, Quere03]
- ✓ ~~Dans les objets, lieux, événements, activités~~ [Casare05]
⊖ [Orstom98, Conte02, LikMui02, Castelfranchi03] :
pas de réciprocité ⇒ qualité [Quere03]

Classes de Confiance

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

SCom

Confiance

LIAR

- ✓ Dispositionnelle [McKnight01]
- ✓ Institutionnelle [McKnight01]
- ✓ **Interpersonnelle** [McKnight01]
- ✓ De groupe [Dasgupta00, Quere03]
- ✓ Dans les objets, lieux, événements, activités [Casare05]

Réputation Directe

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

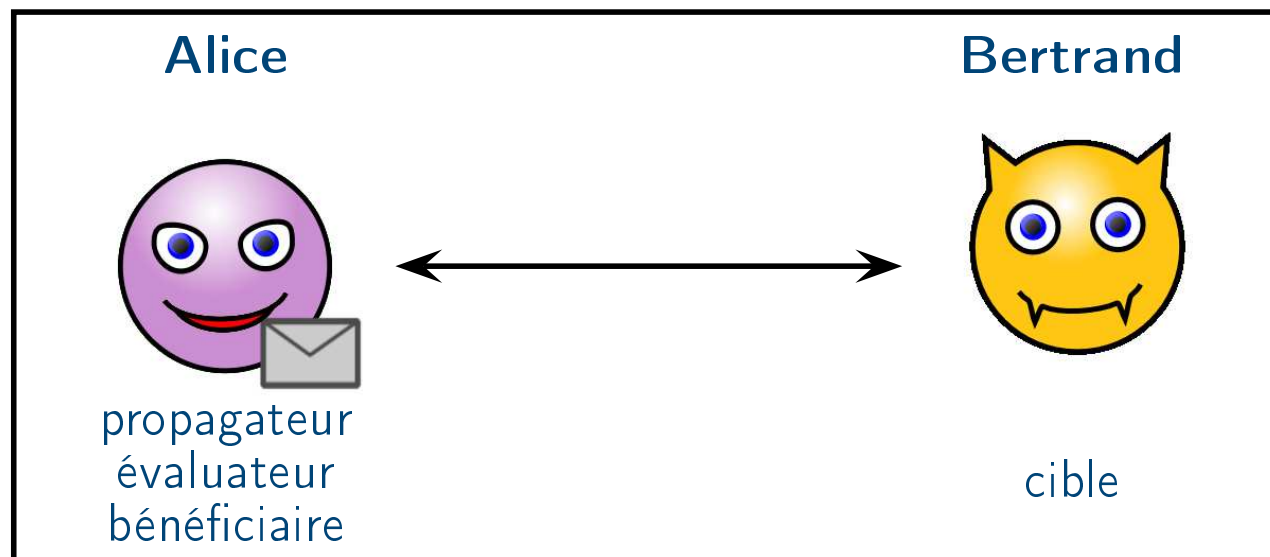
Conclusion

État de l'Art

SCom

Confiance

LIAR



Réputation Observée

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

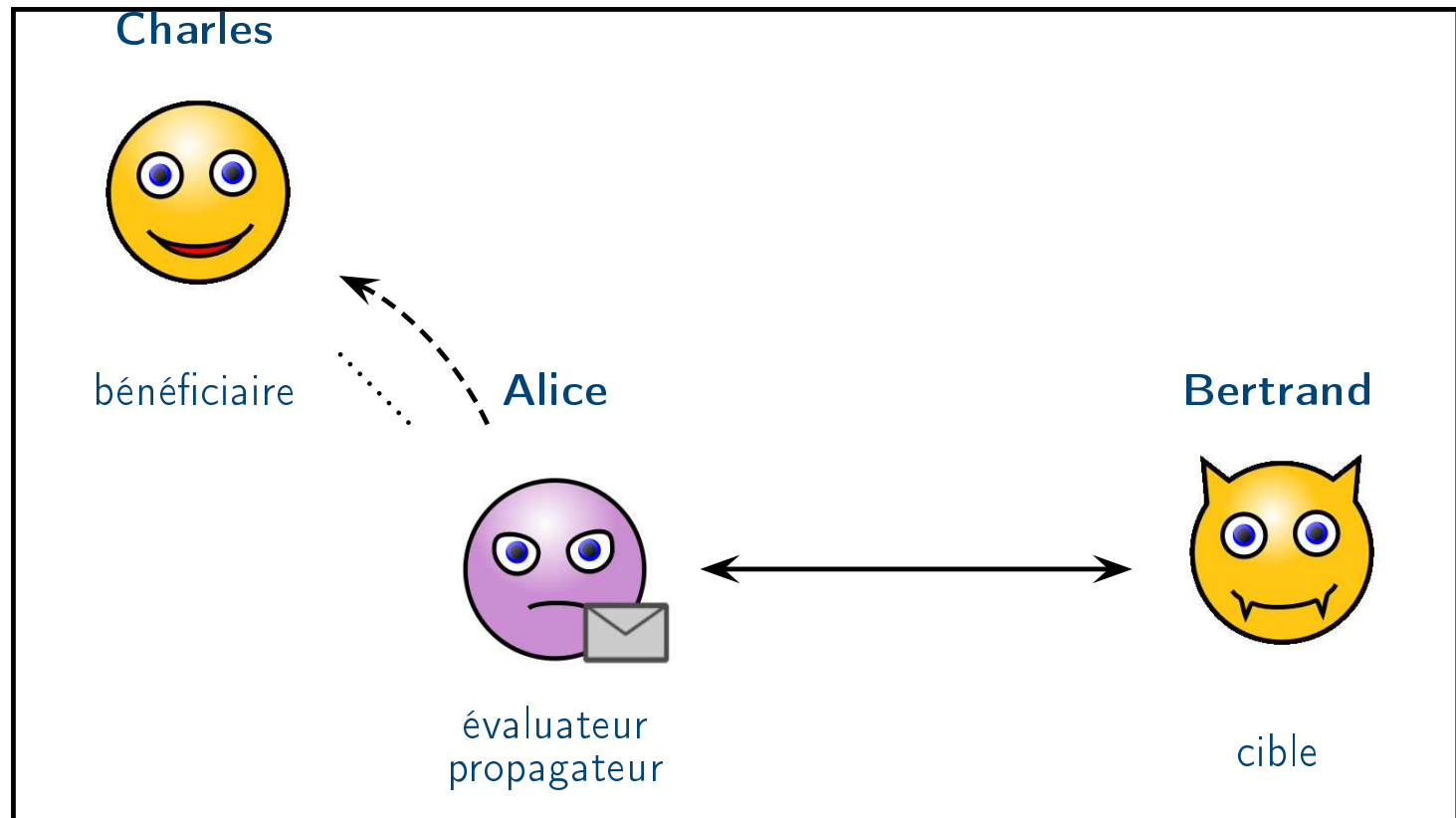
Conclusion

État de l'Art

SCom

Confiance

LIAR



Réputation Propagée

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

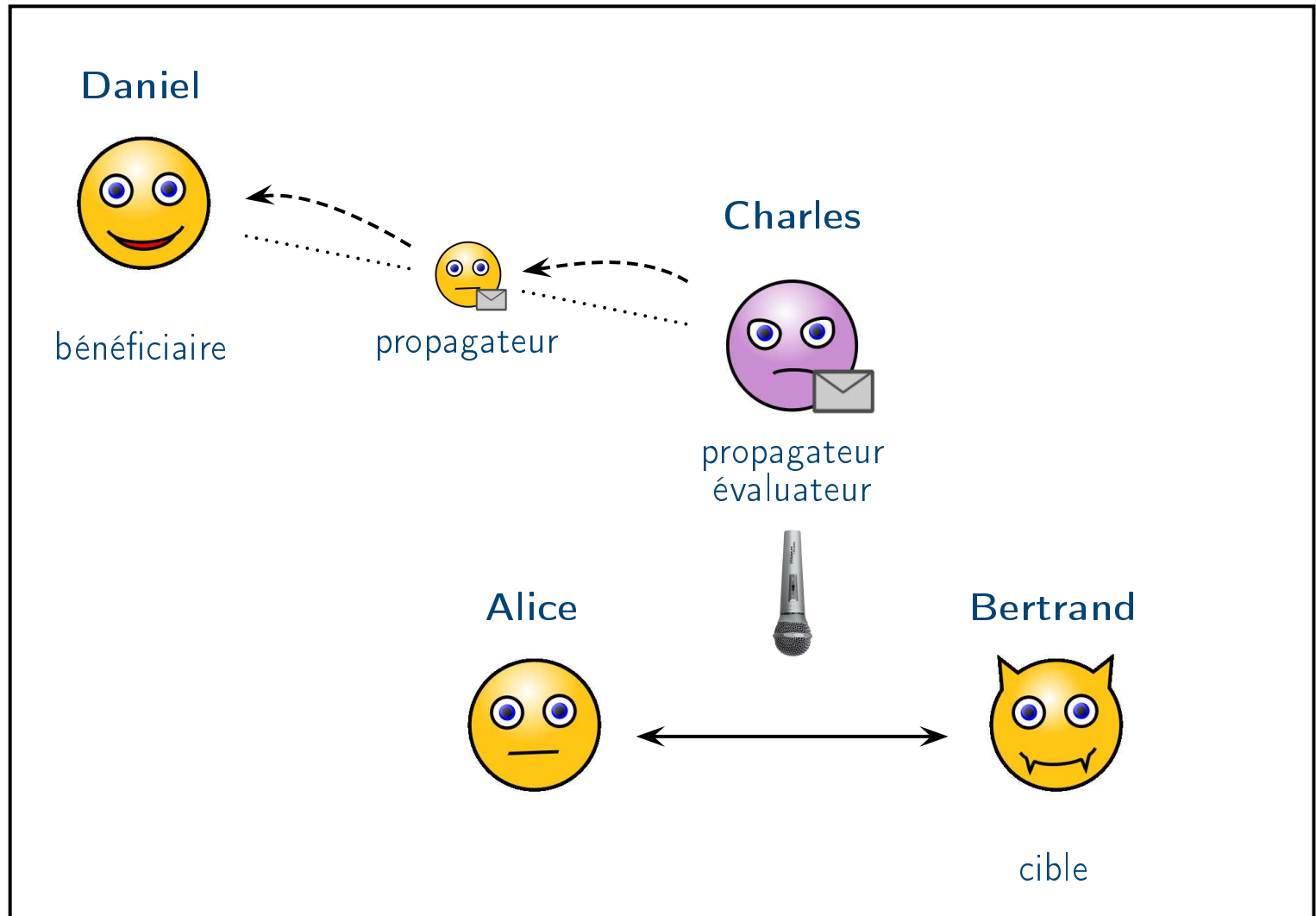
Conclusion

État de l'Art

SCom

Confiance

LIAR



Propriétés de la Réputation fondée sur les Interactions

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

SCom

Confiance

LIAR

- ✓ **Subjectivité** [Gambetta00, Castelfranchi98, Sabater02] *et al.*
Asymétrie [Grandison00, Rouchier00, Castelfranchi94, Quere03]

Propriétés de la Réputation fondée sur les Interactions

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

SCom

Confiance

LIAR

✓ **Subjectivité**
Asymétrie

✓ **Facettes**

unique : [Zacharia99,Carbo03]

multiples : [Sabater02,Wang03,Melaye05]

Propriétés de la Réputation fondée sur les Interactions

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

SCom

Confiance

LIAR

✓ **Subjectivité**
Asymétrie

✓ **Facettes**

✓ **Dimensions**

compétence, sincérité [Castelfranchi98]

compétence, prévisibilité, bonne volonté, intégrité [McKnight01]

Propriétés de la Réputation fondée sur les Interactions

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

SCom

Confiance

LIAR

- ✓ **Subjectivité**
Asymétrie
- ✓ **Facettes**
- ✓ **Dimensions**
- ✓ **Dynamisme** [Rouchier01, Grandison00, Marsh94, Ramchurn03]

Propriétés de la Réputation fondée sur les Interactions

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

SCom

Confiance

LIAR

- ✓ **Subjectivité**
Asymétrie
- ✓ **Facettes**
- ✓ **Dimensions**
- ✓ **Dynamisme**
- ✓ **Graduation** [AbdulRahman00,Rouchier00,Melek98,Quere03]
neutre \neq inconnu [Marsh94]

Propriétés de la Réputation fondée sur les Interactions

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

SCom

Confiance

LIAR

✓ **Subjectivité**
Asymétrie

✓ **Facettes**

✓ **Dimensions**

✓ **Dynamisme**

✓ **Graduation**

✓ **Évolution**

monotonie [AbdulRahman00,Jonker02]

fragilité [Gambetta00,Bhargava03,Melaye05]

Propriétés de la Réputation fondée sur les Interactions

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

SCom

Confiance

LIAR

✓ **Subjectivité**
Asymétrie

✓ **Facettes**

✓ **Dimensions**

✓ **Dynamisme**

✓ **Graduation**

✓ **Évolution**

✓ **Transitivité** de la confiance, des réputations

oui : [OpenPGP05,AbdulRahman97,Zacharia99,Josang05]

non : [Christian-

son96,Grandison00,Firozabadi00,AbdulRahman97,AbdulRahman97,Josa

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

LIAR

Rôles

Information

Réputations

Formules

Inconsistance

Stratégies

Convergence

Décision

Modèle L.I.A.R.

Rôles des agents vis-à-vis de la réputation

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

LIAR

Rôles

Information

Réputations

Formules

Inconsistance

Stratégies

Convergence

Décision

- ✓ **cible**
est évaluée et interagit directement avec le participant
- ✓ **participant**
interagit directement avec la cible
- ✓ **observateur**
observe des interactions : messages → engagements sociaux
- ✓ **évaluateur**
évalue des observations :
engagements + normes → état d'une politique sociale
- ✓ **punisseur**
punit un autre agent : politiques → niveau de réputation
- ✓ **bénéficiaire**
raisonne et prend des décisions :
niveaux des réputations → intentions de confiance → états mentaux
- ✓ **propagateur**
transmet des recommandations

Types d'information fondement de la réputation

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

LIAR

Rôles

Information

Réputations

Formules

Inconsistance

Stratégies

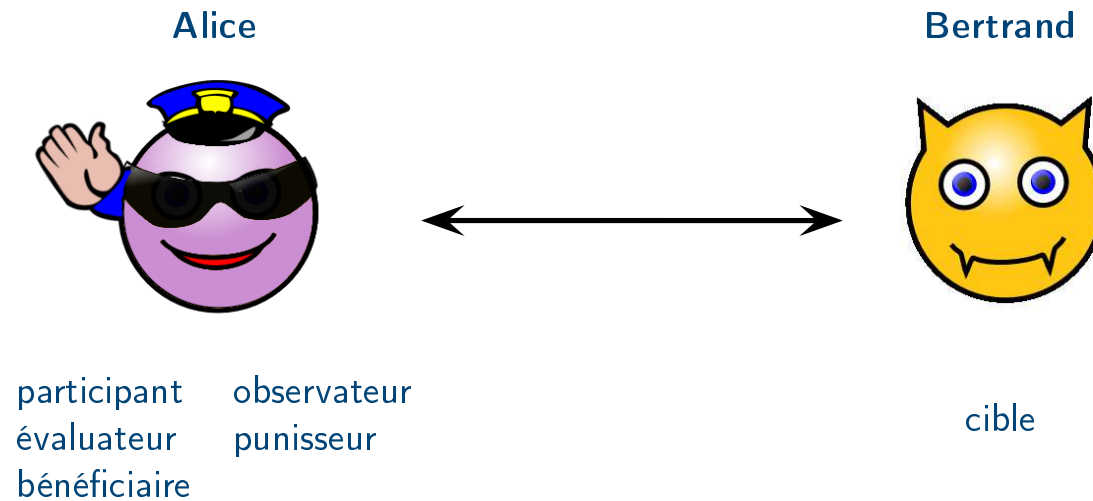
Convergence

Décision

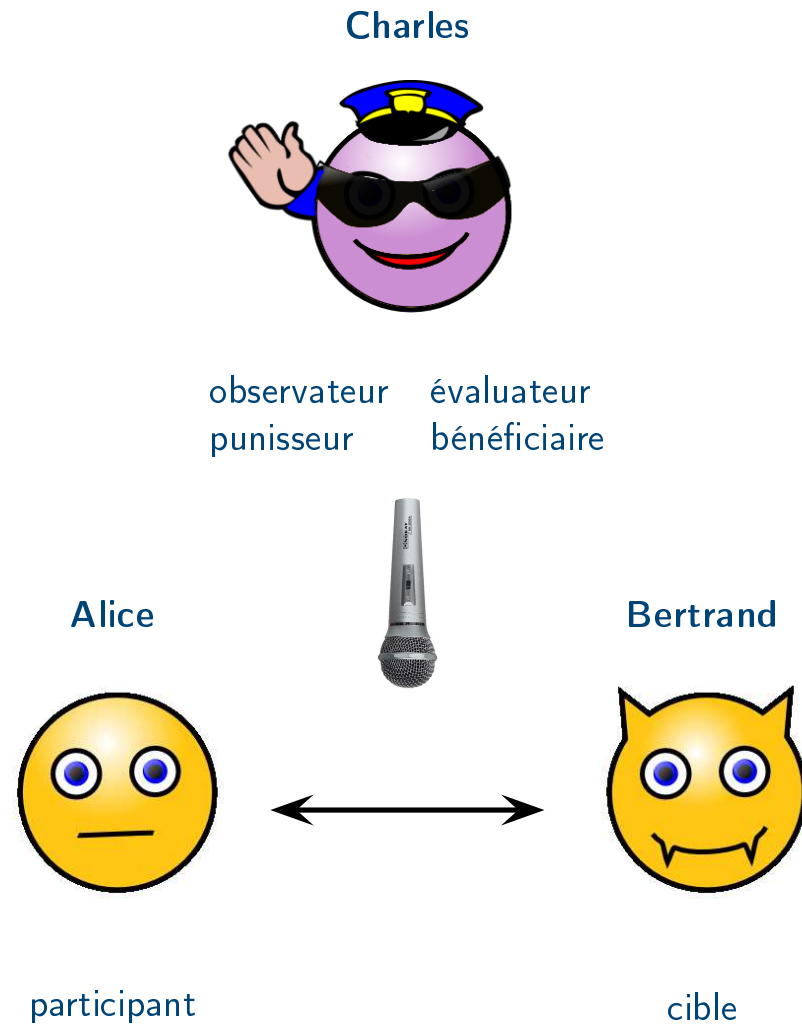
- ✓ **Interaction :**
 - ✗ Interaction **directe**
 - ✗ Interaction **indirecte**

- ✓ **Recommandation, portant sur :**
 - ✗ **Observation**
 - ✗ **Évaluation**
 - ✗ **Niveau de réputation**

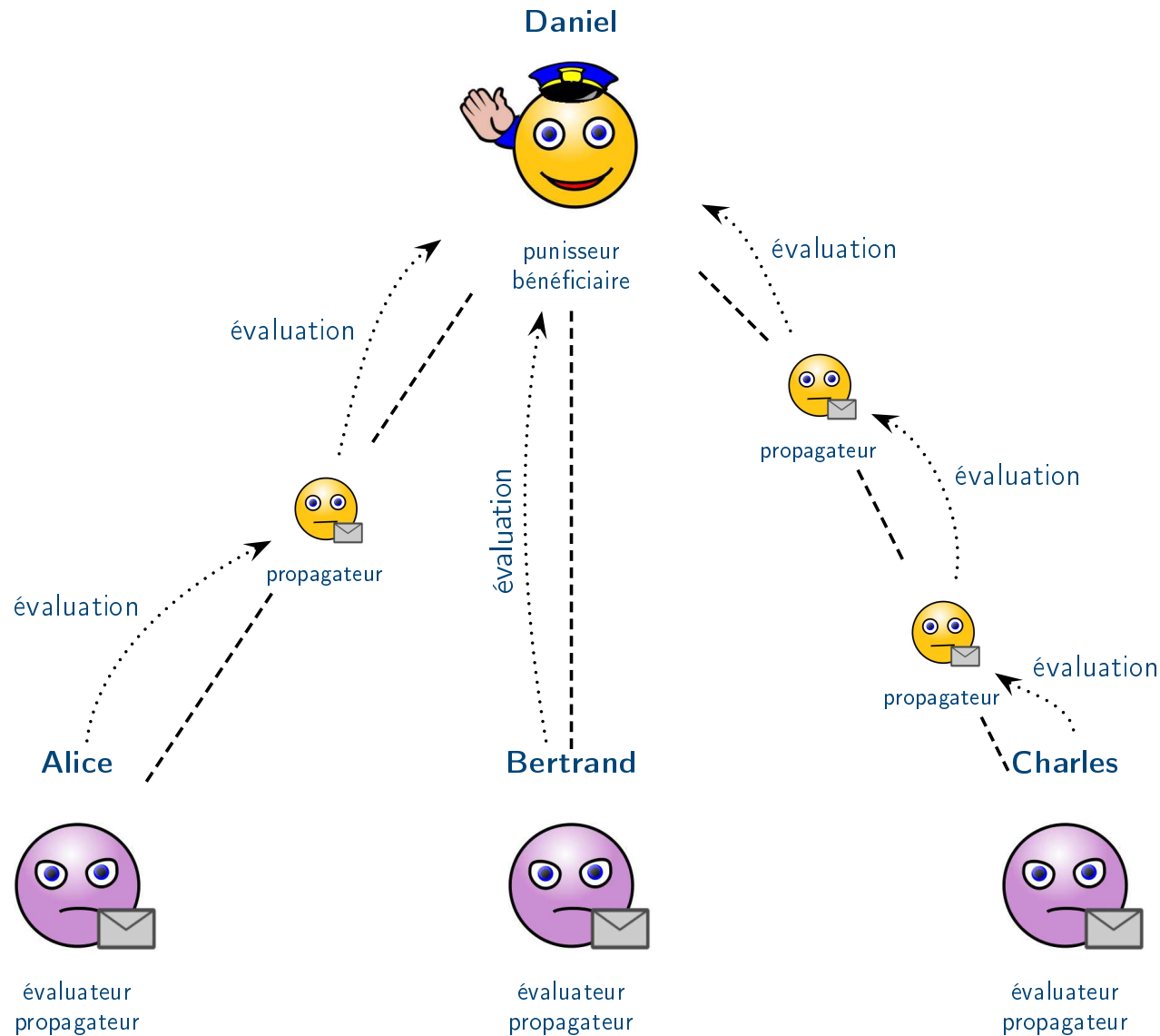
Réputation fondée sur les Interactions Directes



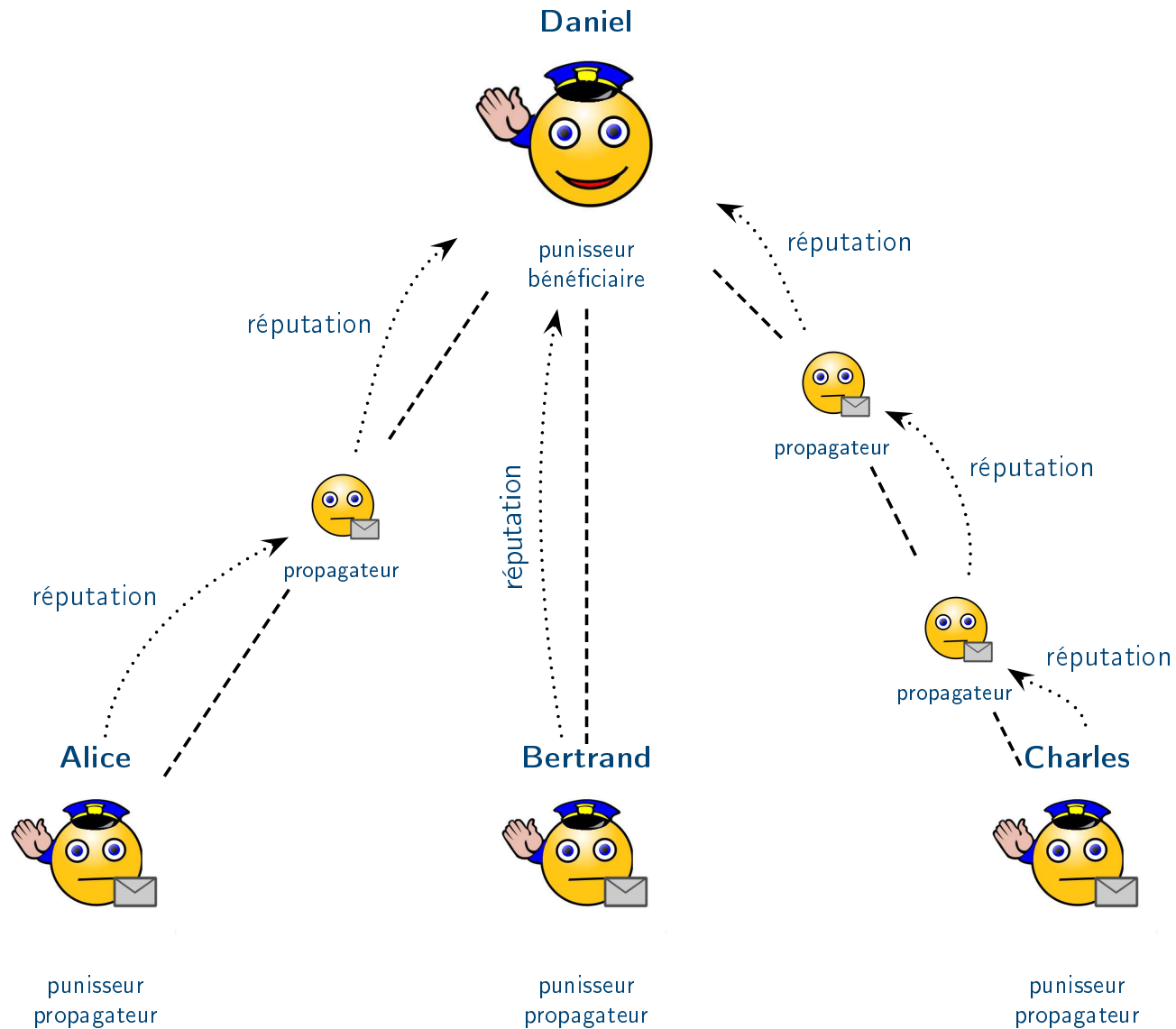
Réputation fondée sur les Interactions Indirectes



Réputation fondée sur les Recommandations d'Évaluations



Réputation fondée sur les Recommandations de Réputation



Calculs – Processus de Puniton

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

LIAR

Rôles

Information

Réputations

Formules

Inconsistance

Stratégies

Convergence

Décision

$$DIbRp \stackrel{\text{def}}{=} \frac{\tau_F \times \text{Imp}(\mathbf{FNCS}) + \tau_V \times \text{Imp}(\mathbf{VNCS}) + \tau_C \times \text{Imp}(\mathbf{CNCS})}{|\tau_F| \times \text{Imp}(\mathbf{FNCS}) + |\tau_V| \times \text{Imp}(\mathbf{VNCS}) + |\tau_C| \times \text{Imp}(\mathbf{CNCS})}$$

- ✓ \mathcal{XNCS} = ensembles de politiques sociales
- ✓ $\text{Imp}(A)$ = somme des pénalités
- ✓ τ_F, τ_V, τ_C : paramètres propres à chaque agent

$$RcbRp \stackrel{\text{def}}{=} \frac{\sum \text{lev} \times \text{pu.reasons}(\text{pu}, \alpha', \delta', \text{RcLev}, t). \text{trust_val}}{\sum \text{pu.reasons}(\text{pu}, \alpha', \delta', \text{RcLev}, t). \text{trust_val}}$$

- ✓ lev = niveau de réputation communiqué
- ✓ pu = punisseur d'origine
- ✓ α' = facette de recommandation
- ✓ δ' = dimension de compétence

Définition de l'inconsistance

Inconsistance d'engagements sociaux :

$ob.inconsistent(t, c, \mathcal{A})$

def

$c \in ob.CCS(t) \wedge \forall c_i \in \mathcal{A}, c_i \in ob.CCS(t) \wedge$

$c.st \in \{active, fulfilled\} \wedge$

$\exists \{c_1, \dots, c_n\} \subseteq \mathcal{A} /$

$\forall c_i \in \{c_1, \dots, c_n\}, c_i.st \in \{active, fulfilled\} \wedge$

$ob.inconsistent_content(t, c.cont, c_1.cont, \dots, c_n.cont)$

→ Tous les engagements sont observés par le même agent ob

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

LIAR

Rôles

Information

Réputations

Formules

Inconsistance

Stratégies

Convergence

Décision

Définition de l'inconsistance

Inconsistance d'engagements sociaux :

$$\text{ob.inconsistent}(t, c, \mathcal{A})$$

def

$$c \in {}_{\text{ob}}\text{CCS}(t) \wedge \forall c_i \in \mathcal{A}, c_i \in {}_{\text{ob}}\text{CCS}(t) \wedge$$
$$c.\text{st} \in \{\text{active}, \text{fulfilled}\} \wedge$$
$$\exists \{c_1, \dots, c_n\} \subseteq \mathcal{A} /$$
$$\forall c_i \in \{c_1, \dots, c_n\}, c_i.\text{st} \in \{\text{active}, \text{fulfilled}\} \wedge$$
$$\text{ob.inconsistent_content}(t, c.\text{cont}, c_1.\text{cont}, \dots, c_n.\text{cont})$$

Tous les engagements sont observés par le même agent ob

Ils sont en état « positif »

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

LIAR

Rôles

Information

Réputations

Formules

Inconsistance

Stratégies

Convergence

Décision

Définition de l'inconsistance

Inconsistance d'engagements sociaux :

$$\text{ob.inconsistent}(t, c, \mathcal{A})$$
$$\stackrel{\text{def}}{=}$$
$$c \in {}_{\text{ob}}\text{CCS}(t) \wedge \forall c_i \in \mathcal{A}, c_i \in {}_{\text{ob}}\text{CCS}(t) \wedge$$
$$c.\text{st} \in \{\text{active}, \text{fulfilled}\} \wedge$$
$$\exists \{c_1, \dots, c_n\} \subseteq \mathcal{A} /$$
$$\forall c_i \in \{c_1, \dots, c_n\}, c_i.\text{st} \in \{\text{active}, \text{fulfilled}\} \wedge$$

→ $\text{ob.inconsistent_content}(t, c.\text{cont}, c_1.\text{cont}, \dots, c_n.\text{cont})$

Tous les engagements sont observés par le même agent ob
Ils sont en état « positif »

→ Leurs contenus sont inconsistants

Introduction

État de l'Art

Modèle L.I.A.R.

Expérimentations

Conclusion

État de l'Art

LIAR

Rôles

Information

Réputations

Formules

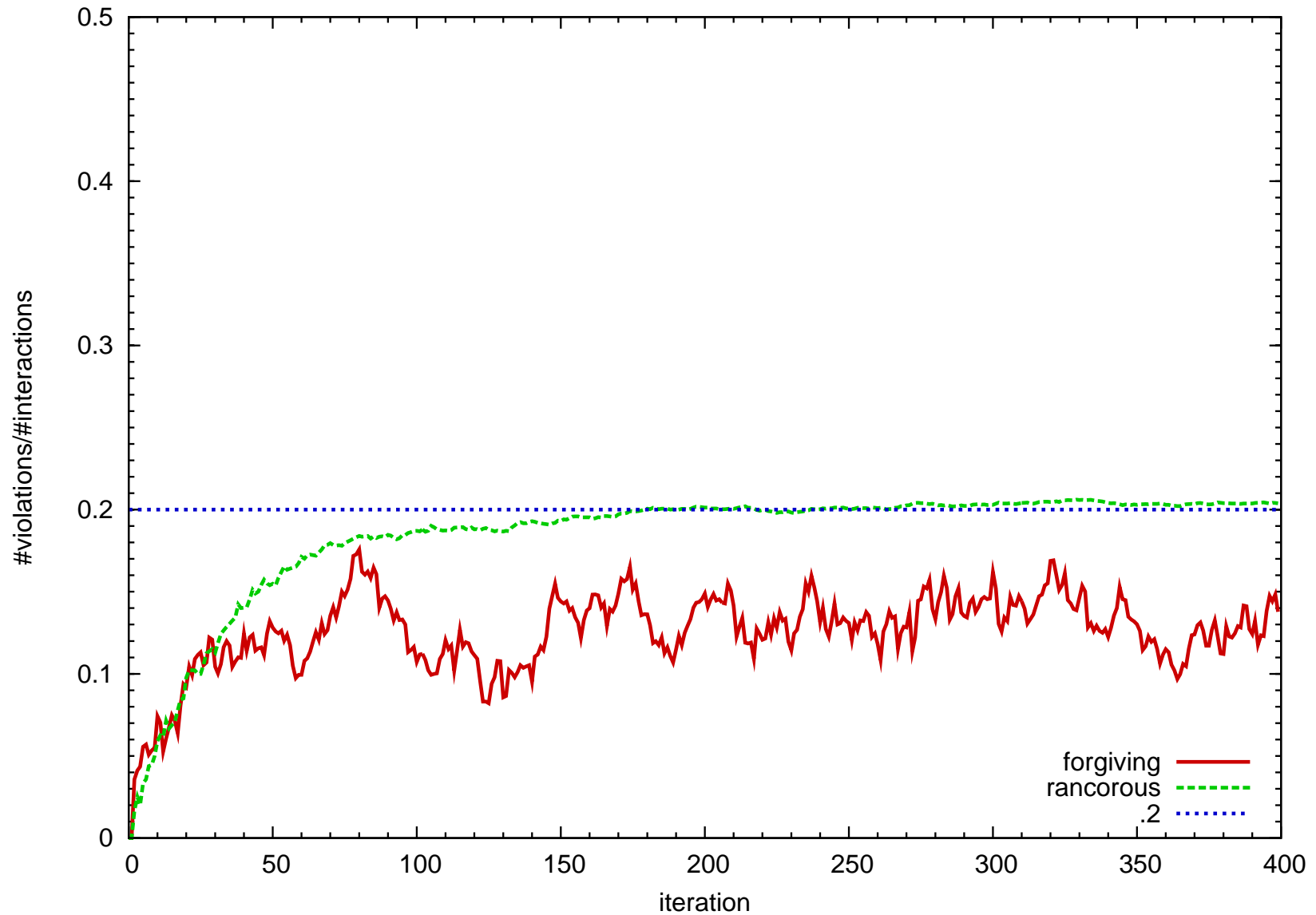
Inconsistance

Stratégies

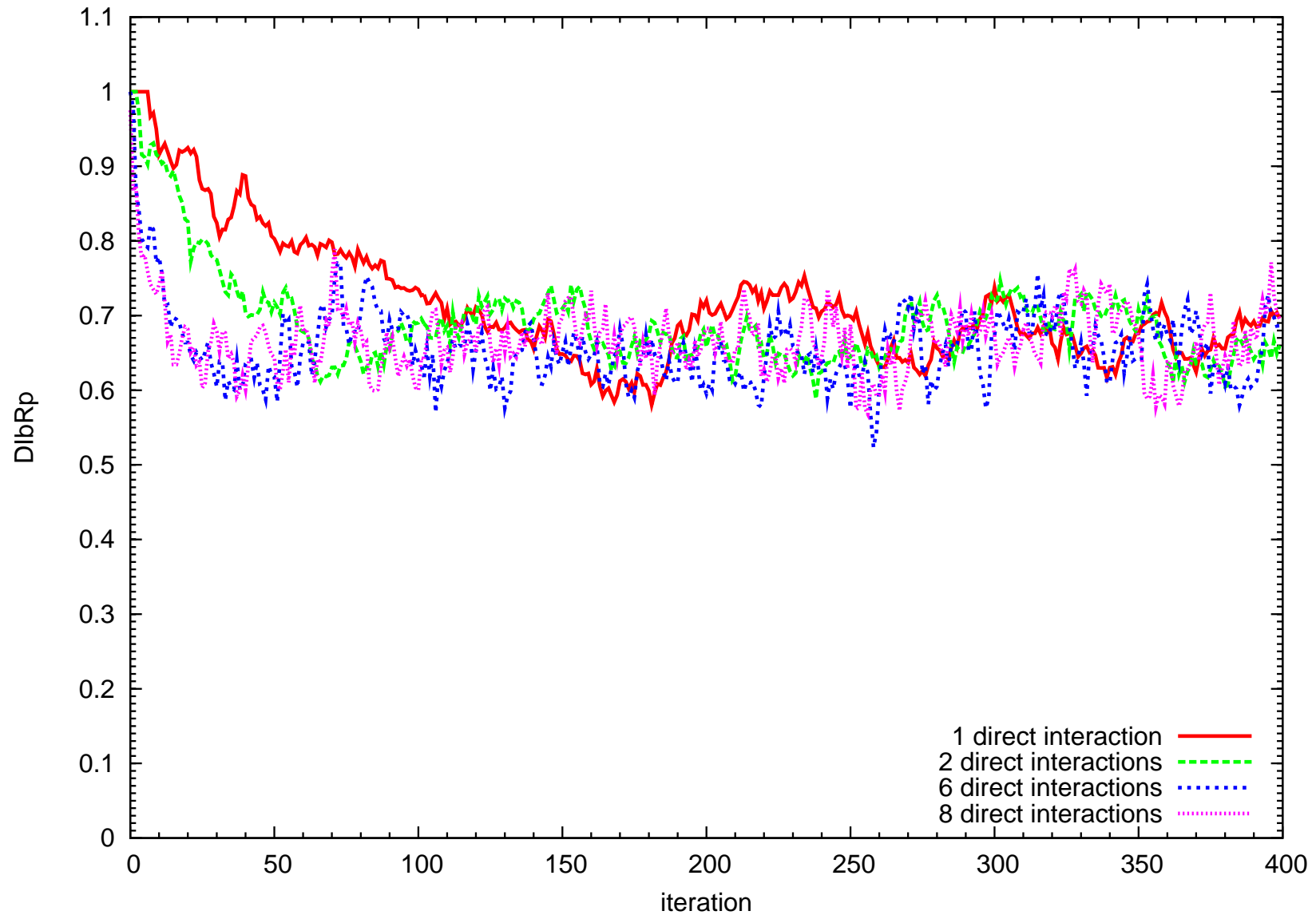
Convergence

Décision

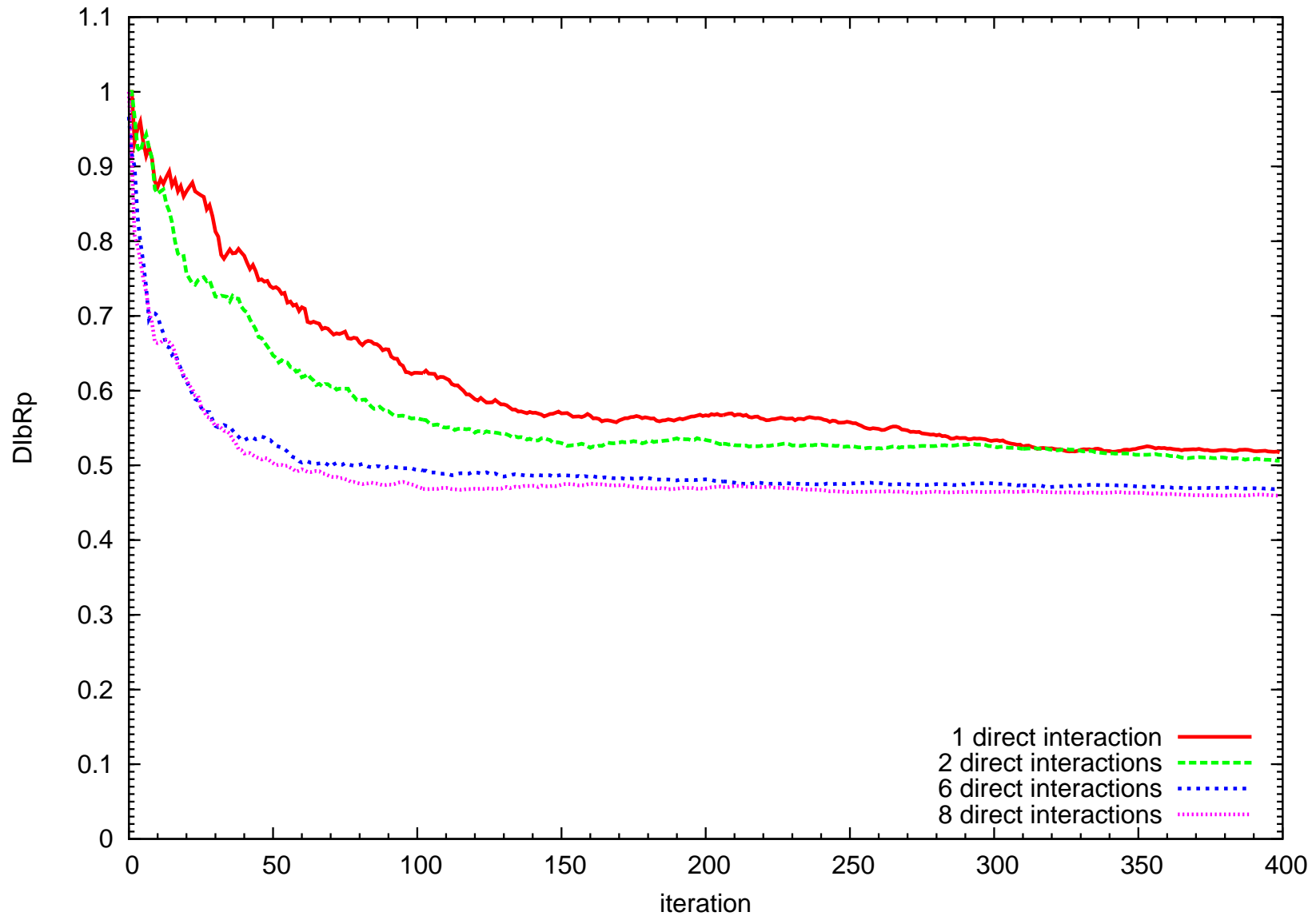
Comparaison des stratégies



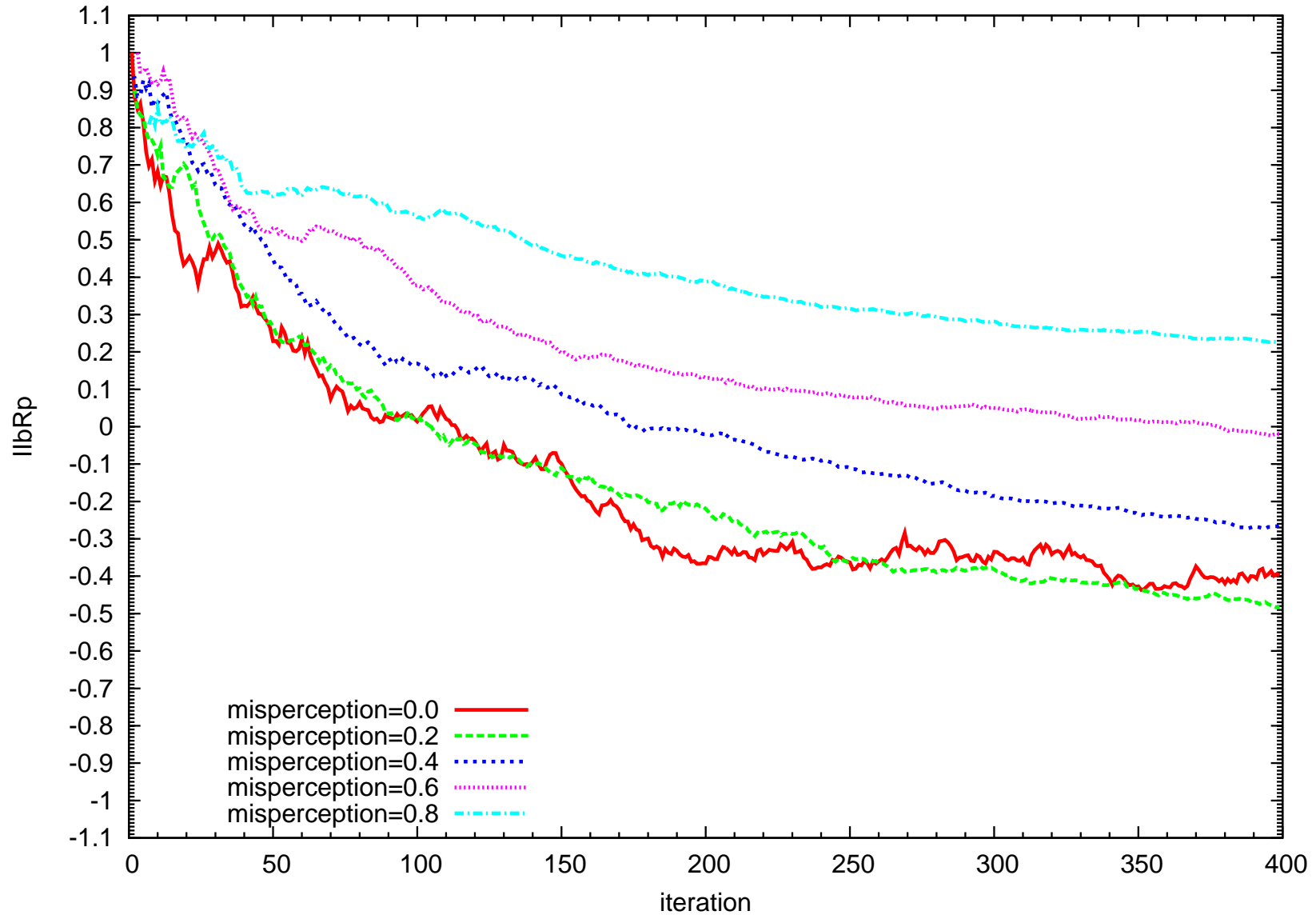
Convergence et Précision – DIbRp indulgents



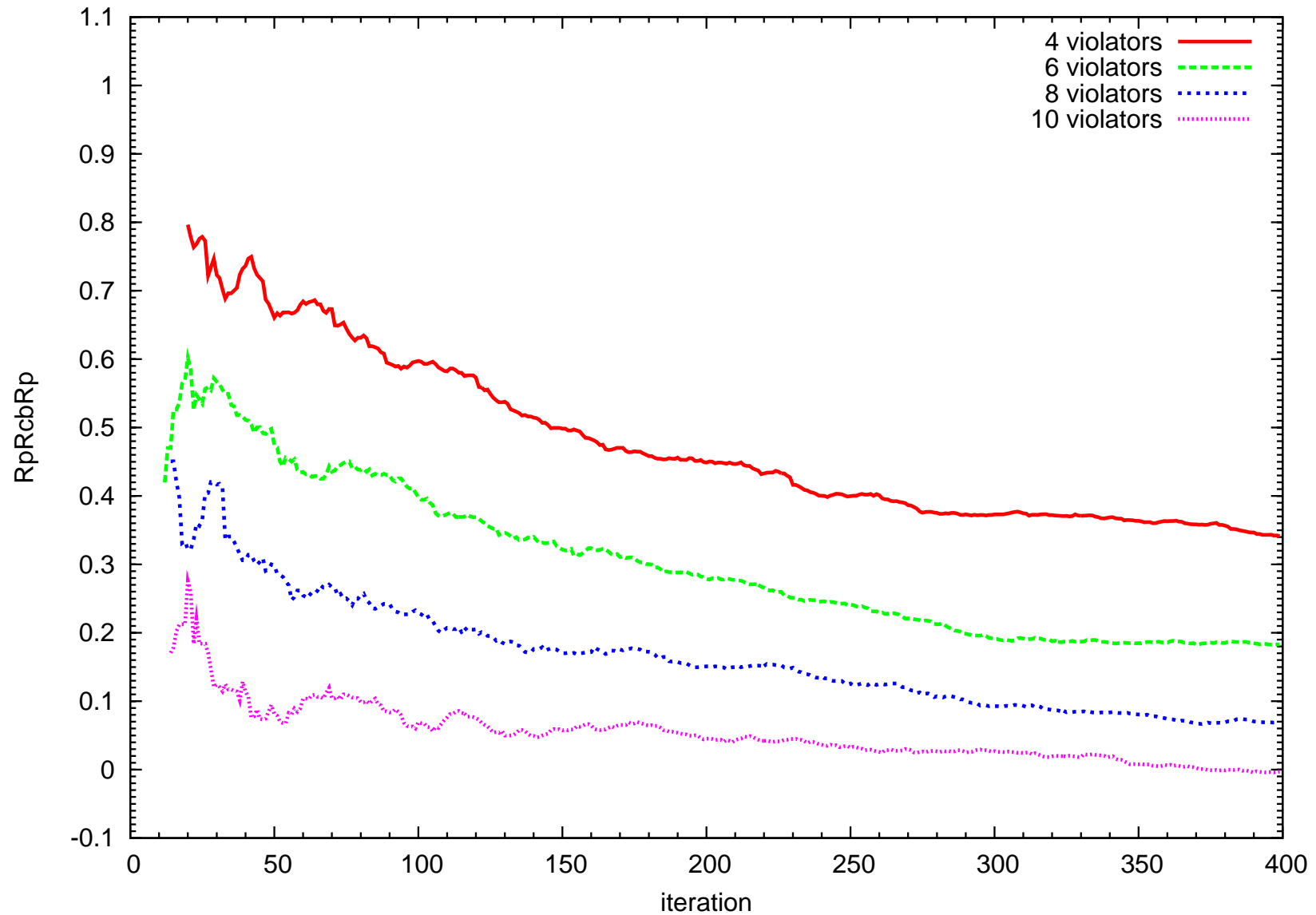
Convergence et Précision – DIbRp rancuniers



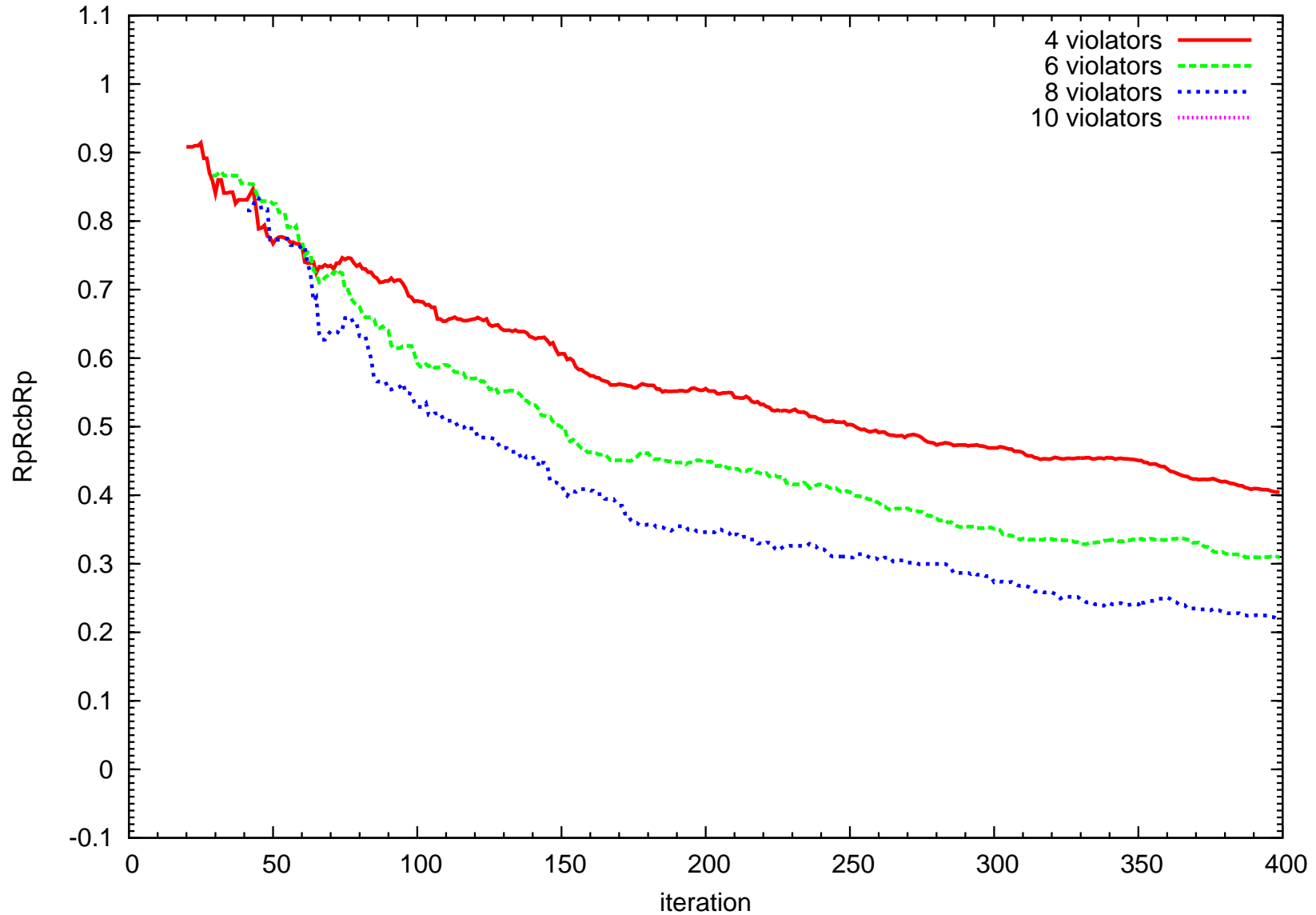
Convergence et Précision – IIbRp indulgents



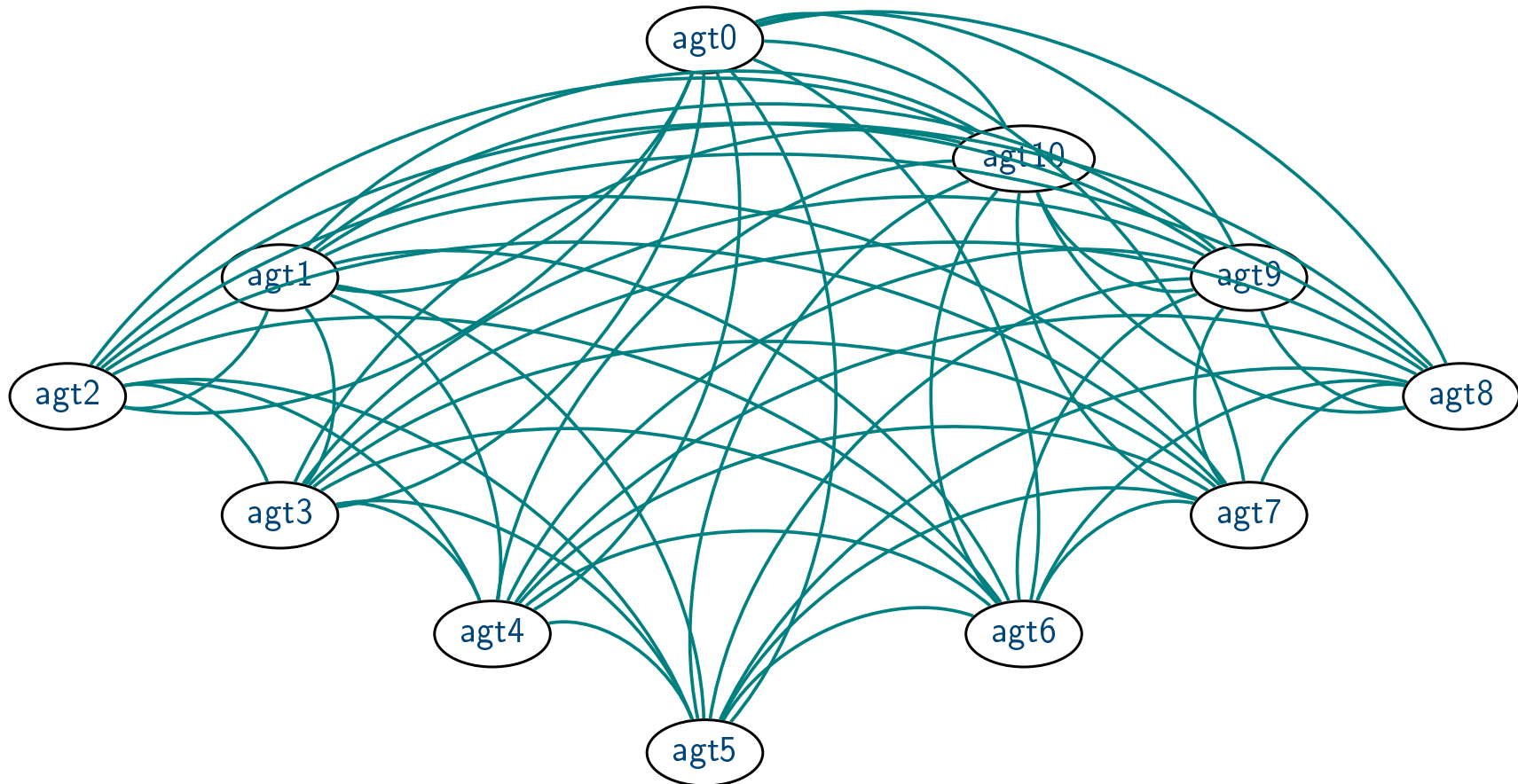
Convergence et Précision – $R_p R_{cb} R_p$ indulgents sans filtre



Convergence et Précision – $R_p R_{cb} R_p$ indulgents, avec filtre

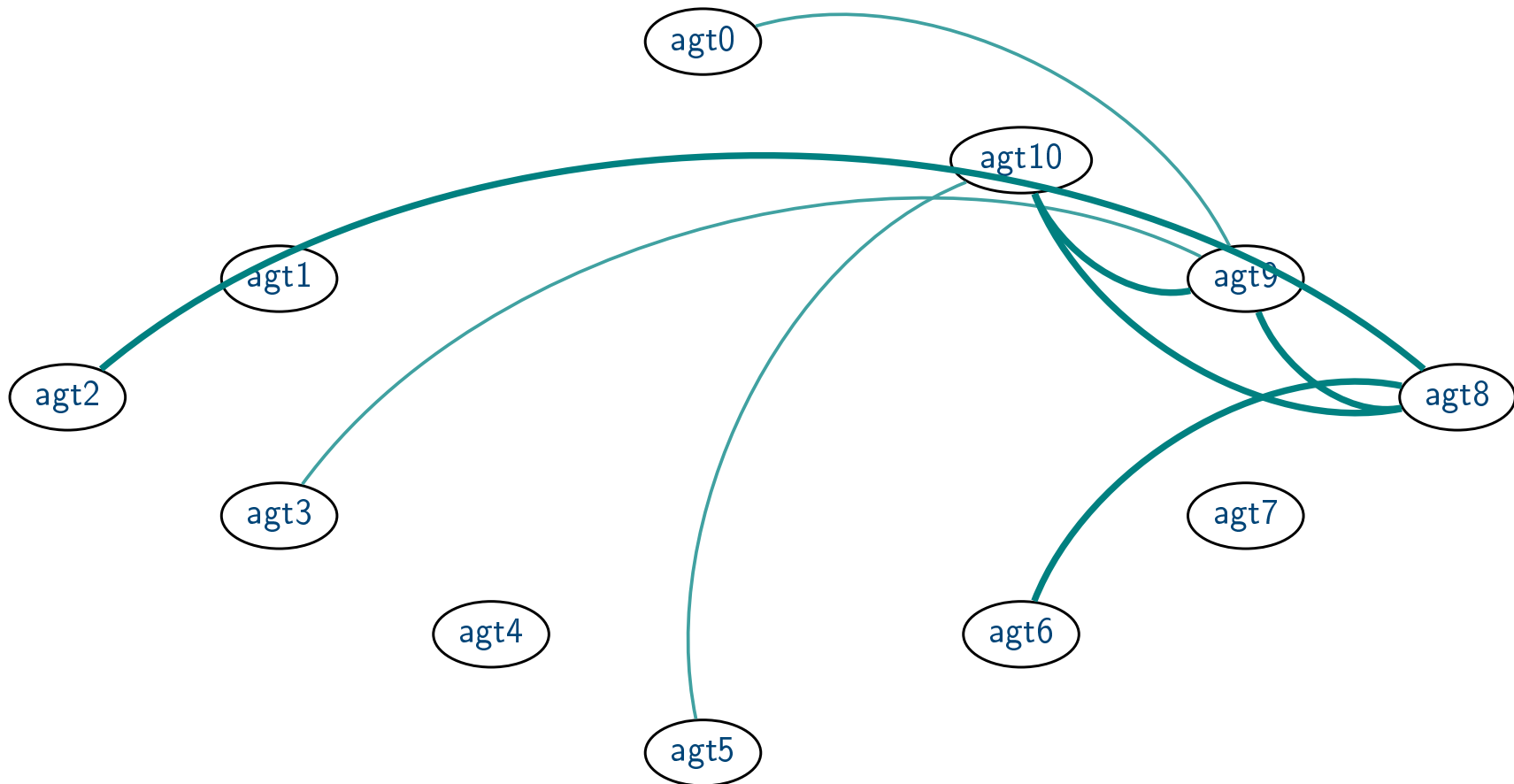


Décision – indulgents



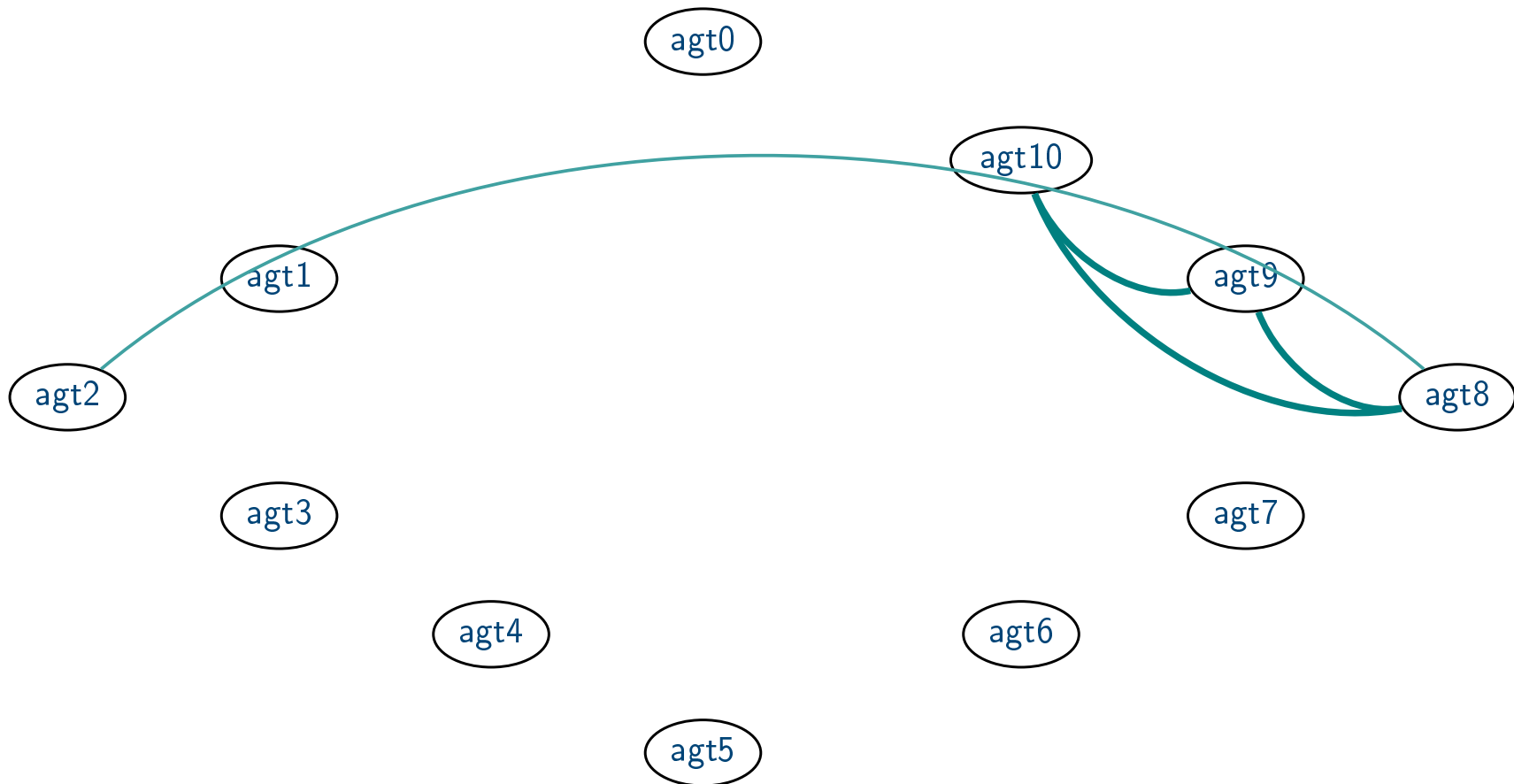
Étape 1

Décision – indulgents



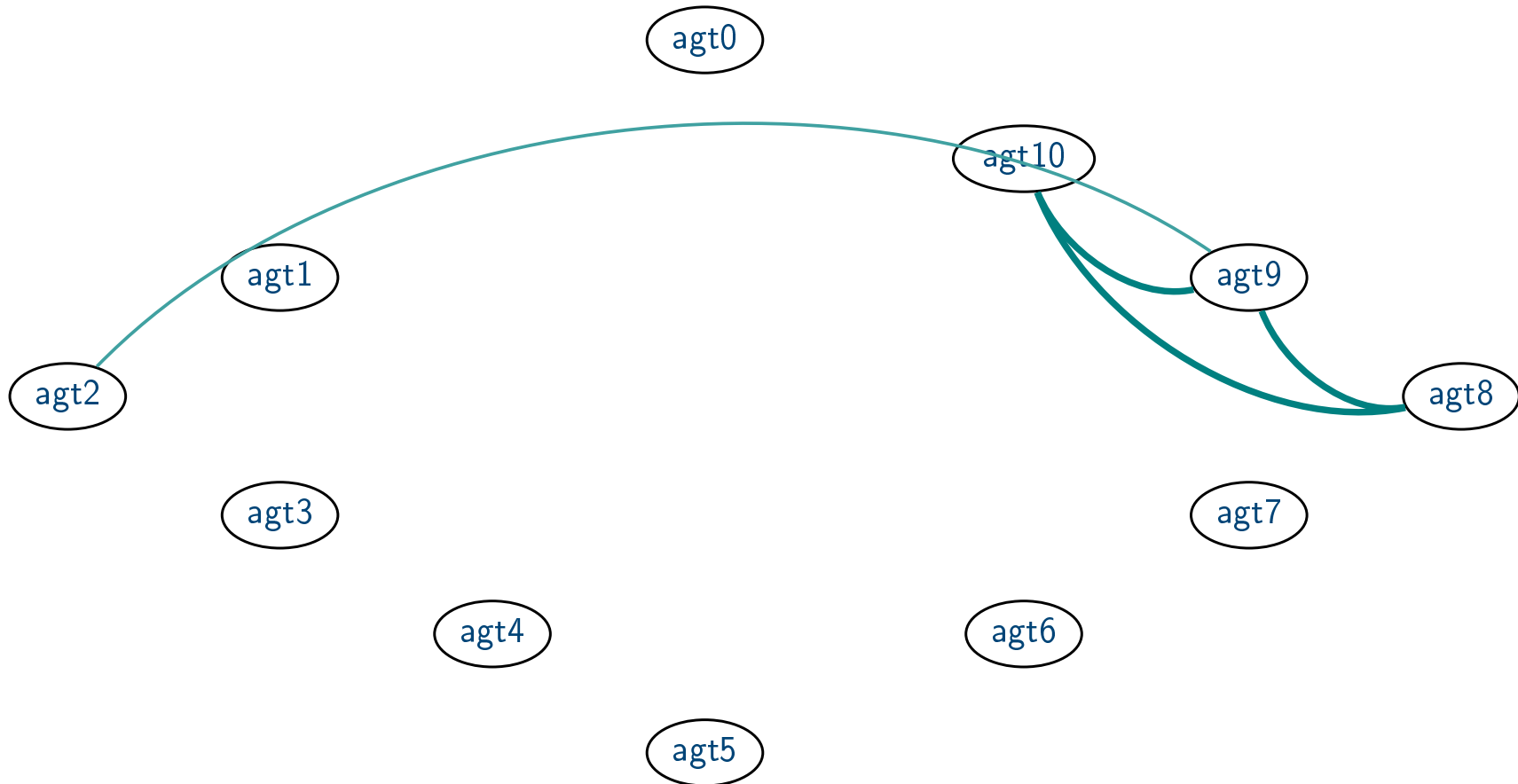
Étape 50

Décision – indulgents



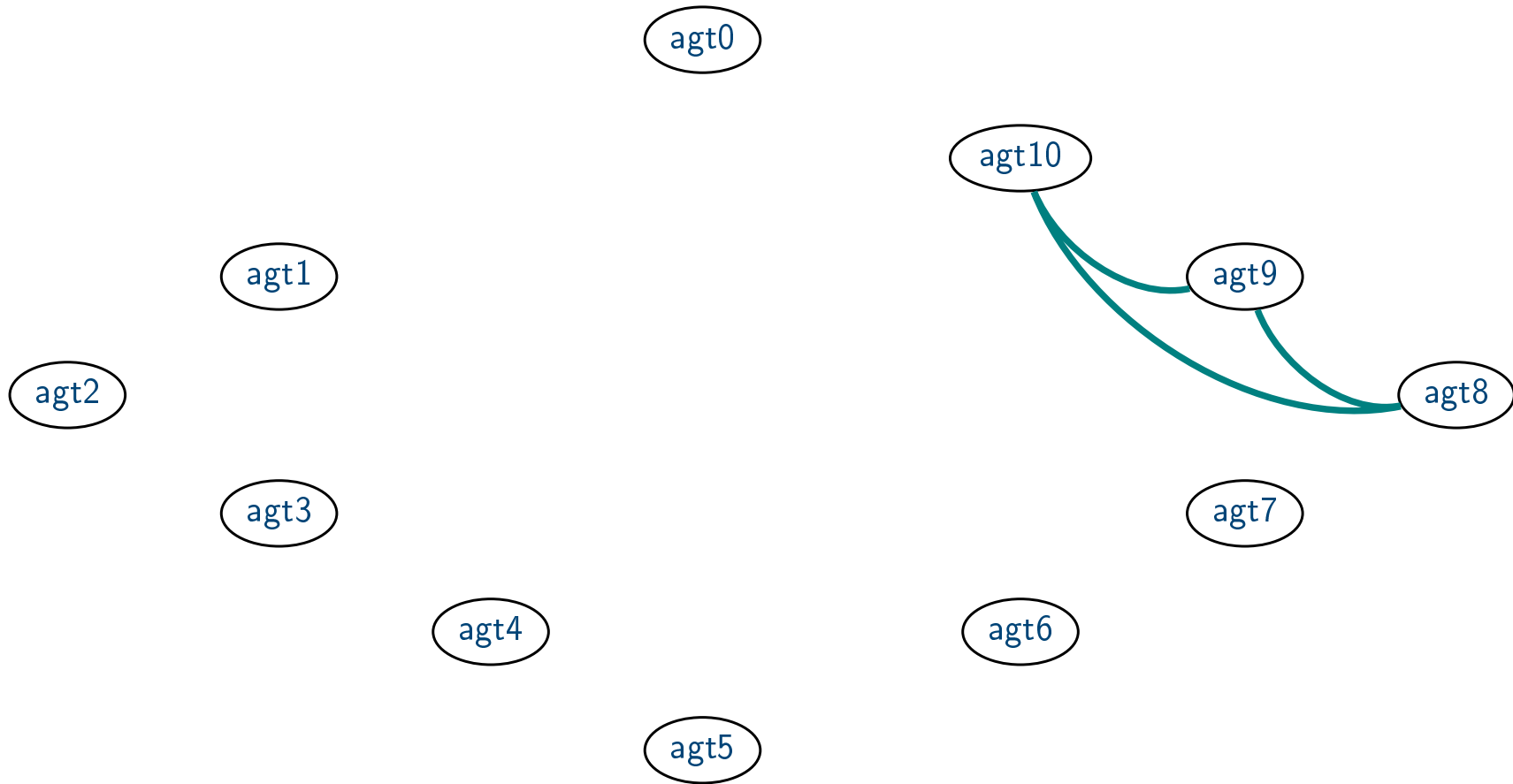
Étape 100

Décision – indulgents



Étape 150

Décision – indulgents



Étape 200