De l'utilisation des politiques de bandits manchots dans les systèmes de réputation

Thibaut Vallée Grégory Bonnet François Bourdon
prénom.nom@unicaen.fr
Normandie Université, GREYC, CNRS UMR 6072, F-14032 Caen, France

Résumé

La robustesse des systèmes de réputation face aux manipulations a été largement étudiée. Cependant, les études portant sur l'utilisation des valeurs de réputation sont, elles, beaucoup plus rares. Dans cet article, nous dressons une analogie entre les systèmes de réputation et les problèmes de bandits manchots. Nous proposons alors d'utiliser les politiques de sélection de ces derniers dans le cadre des systèmes de réputation afin d'augmenter leur robustesse face à des agents malveillants. Pour cela, nous proposons un modèle général de système d'échange de services utilisant un système de réputation. Par une étude empirique, nous montrons que certaines des politiques de bandits manchots sont robustes aux manipulations mais peu coûteuses pour les agents malveillants tandis que d'autres sont manipulables mais avec un coût élevé.

Mots Clefs

Systèmes de réputation, Manipulations, Bandits manchots

Abstract

The robustness of reputation systems against manipulations has been widely studied. However, the studies of how to use the reputation values computed by those systems are rare. In this paper, we draw an analogy between reputation systems and multi-armed bandit problems. We investigate how to use the multi-armed bandit selection policies in order to increase the robustness of reputation systems against malicious agents. To this end, we propose a model of an abstract service sharing system using such a bandit-based reputation system. In an empirical study, we show that some policies are more robust against manipulations but cost-free for the malicious agents whereas some other policies are manipulable but costly.

Keywords

Reputation System, Manipulation, Multi-Armed Bandits

1 Introduction

Dans un système multi-agent, lorsqu'un agent n'est pas capable de réaliser une tâche seul, il peut la déléguer à un autre agent. Pour cela, les agents doivent se partager leurs compétences et leurs connaissances. Dans un tel système, les agents sont à la fois fournisseurs et utilisateurs de services. Toutefois, comme un tel système permet à des agents hétérogènes d'intéragir entre eux, certains agents peuvent fournir des services de mauvaise qualité, que cela soit du à des erreurs de calcul, des défaillances ou encore à un comportement malveillant. Afin d'aider les agents à décider avec qui ils doivent interagir [1], il est possible de mettre en place un système de réputation. Un tel système permet à un agent de demander des services à un autre agent qui lui a été recommandé par des agents tiers. Pour cela, les agents évaluent leurs interactions passées et calculent une valeur représentant un niveau de confiance qu'ils ont envers les autres. Ces valeurs de confiance sont ensuite partagées entre les agents par le biais de témoignages et agrégées entre elles. Cette agrégation des témoignages produit une valeur de réputation pour chaque agent, qui est supposée refléter leur capacité respective à fournir des services de qualité. De nombeux systèmes de réputation ont été proposés mais un agent malveillant peut y manipuler sa valeur de réputation en fournissant de faux témoignages, en formant des coalitions avec d'autres agents malveillants, en introduisant de fausses identités (appelées Sybil) ou en changeant subitement de comportement. Certains travaux proposent des systèmes de réputation robustes à des manipulations spécifiques. Cependant, ces études portent sur le calcul des valeurs de confiance et de réputation et non sur la manière dont les agents les utilisent. Or, la politique d'utilisation de ces valeurs a une influence sur le fonctionnement du système. En effet, si l'ensemble des agents décide de ne consommer des services qu'auprès des agents ayant la plus haute valeur de réputation, il sera difficile pour un agent malveillant seul de fournir des mauvais services sur le long terme. Cependant, une telle politique risque de surcharger de requêtes ces agents réputés et d'empêcher l'ouverture du système. Par ailleurs, les systèmes de réputation sont sensibles aux coalitions d'agents malveillants changeant soudainement de comportement. Par exemple, sur eBay [9], un agent peut vendre de nombreux biens de faible valeur afin d'augmenter sa valeur de réputation et ensuite vendre un objet de forte valeur et ne jamais le fournir. De manière intéressante, le problème qui consiste à décider avec qui interagir en utilisant des observations passées a été largement

1. Deux agents interagissent lorsque l'un fournit un service à l'autre.

étudié dans le contexte des bandits manchots (*multi-armed bandits* ou MAB). Dans cet article, nous proposons d'étudier comment l'utilisation des politiques de MAB dans un système de réputation peut influencer sa robustesse. Nous présentons en section 2 l'état de l'art dans le domaine des systèmes de réputation et introduisons le problème des bandits manchots. En section 3, nous proposons un modèle général de système d'échange de services et faisons l'analogie entre ce modèle et celui des bandits manchots. Nous présentons en section 4 différentes politiques de sélection. Pour finir, nous présentons en section 5 une étude empirique des performances de ces politiques lorsqu'une coalition d'agents malveillants manipule le système.

2 État de l'art

La notion de confiance a été introduite dans le contexte des systèmes multi-agents par Marsh [16]. Elle formalise une estimation du comportement futur d'un agent lorsqu'il existe un risque que celui-ci ait un comportement inattendu. Resnick [17] propose trois axiomes fondamentaux pour définir un système de réputation : (1) les agents doivent interagir ensemble dans le futur ; (2) les agents doivent partager leurs valeurs de confiance par le biais de témoignages ; (3) ces témoignages doivent être utilisés par les agents pour déterminer à qui demander un service souhaité. Ainsi, la valeur de confiance d'un agent envers un autre est une évaluation des services que ce dernier a fourni au premier. La réputation d'un agent est, elle, une agrégation des valeurs de confiance des autres agents vis-à-vis de celui-ci. De nombreux systèmes de réputation ont été proposés [17, 12, 13, 19, 11, 2] et peuvent être classés en trois familles : symétrique (eBay [17]), assymétrique global (EigenTrust [13]) et assymétrique personnalisé (BetaReputation [12]). La robustesse des systèmes de réputation aux manipulations a été aussi fortement étudiée [11, 2, 10, 8, 1]. Cheng et Friedman [8] ont prouvé qu'il n'existe pas de système de réputation symétrique robuste aux attaques Sybil et que les système assymétrique ne peuvent l'être que s'ils satisfont des conditions restrictives. Altman *et al.* [2, 1] ont prouvé que la plupart des systèmes de réputation ne peuvent pas simultanément satisfaire un axiome de robustesse et d'autres axiomes fondamentaux. Cependant, ces travaux considèrent la robustesse aux manipulations à un instant donné : un système est robuste s'il n'existe pas de manipulation capable de modifier les valeurs (ou le rang) de réputation à l'instant où la manipulation se déroule. Or, certaines manipulations telles que l'attaque oscillante [19] sont définies sur le long terme. Enfin, ces études ne s'intéressent pas à la manière dont ces valeurs de réputation sont utilisées. Il convient donc de se poser une autre question : comment utiliser les valeurs de réputation pour déterminer à qui demander un service souhaité ? Usuellement, dans les systèmes de réputation, un agent demande le service désiré à l'agent ayant la meilleure réputation, même si certains auteurs proposent des politiques de sélection probabilistes [13]. Ce problème de décision a été étudiée dans un autre contexte, celui des bandits manchots (MAB) [18]. La définition canonique d'un problème MAB est la suivante : considérons une machine à sous avec plusieurs bras, chacun ayant une fonction de gain suivant une loi de distribution a priori inconnue, quelle séquence de bras un agent doit-il tirer afin de maximiser son gain ? De nombreux modèles de MAB ont été étudiés, avec plusieurs joueurs [15], une fonction de gain stationnaire ou non [14], la possibilité de tirer plusieurs bras simultanément [3], ou même avec un adversaire [5]. Dans tous les cas, l'agent dispose d'une politique de sélection lui permettant de minimiser son regret, c'est-à-dire la différence entre le gain obtenu et le gain qu'aurait eu l'agent si à chaque pas de temps il avait choisit le meilleur bras. Toutes ces politiques, telles que UCB, Poker, ε-glouton [20, 6], sont des compromis entre l'exploitation (tirer le bras qui possède le plus haut gain espéré) et l'exploration (tirer un autre bras afin d'améliorer les estimations des espérances de gain). Awerbuch et Kleinberg [7] ont proposé un lien entre systèmes de réputation et bandits manchots du point de vue de la mise à jour des valeurs de confiance, mais non pas à leur utilisation dans une politique de sélection. Dans cet article, nous faisons l'analogie entre les deux problèmes de décisions : la sélection d'agents évalués par un système de réputation et la sélection de bras évalués par une estimation de leur fonction de gain. Nous étudions alors comment les politiques peuvent influencer la robustesse des systèmes de réputation.

3 Modèle

Nous proposons une application où les agents doivent interagir avec les autres et utilisent pour cela un système de réputation. Dans ce système, les agents utilisent une politique afin de sélectionner avec qui ils doivent interagir, par analogie avec un problème de bandits manchots.

3.1 Système d'échange de services

Considérons un système multi-agent où chaque agent peut fournir des services et demander à d'autres de lui en fournir. Afin de ne pas perdre en généralité, nous considérons ces services comme abstrait. Un tel système est appelé un *système d'échange de services* : lorsqu'un agent a besoin d'un service qu'il ne peut pas réaliser lui-même, il demande à un autre agent de le lui fournir.

Définition 1 *Un* système d'échange de services *est un tuple $\langle N, S\rangle$ où N est l'ensemble des agents et S l'ensemble des services qui peuvent être fournis. Notons par $N_x \subseteq N$ l'ensemble des agents qui peuvent fournir le service $s_x \in S$.*

Définition 2 *Un agent $a_i = \langle \overrightarrow{\varepsilon_i}, v_i, T_i, f_i, \pi_i\rangle$ est une entité autonome qui peut fournir et recevoir des services où $\overrightarrow{\varepsilon_i}$ désigne son vecteur d'expertise ; v_i sa fonction d'évaluation ; T_i sa matrice de confiance ; f_i sa fonction de réputation ; π_i sa politique.*

L'*expertise* de $a_i \in N$ pour le service $s_x \in S$, notée $\varepsilon_{i,x}$, est la capacité de a_i à fournir le service s_x avec

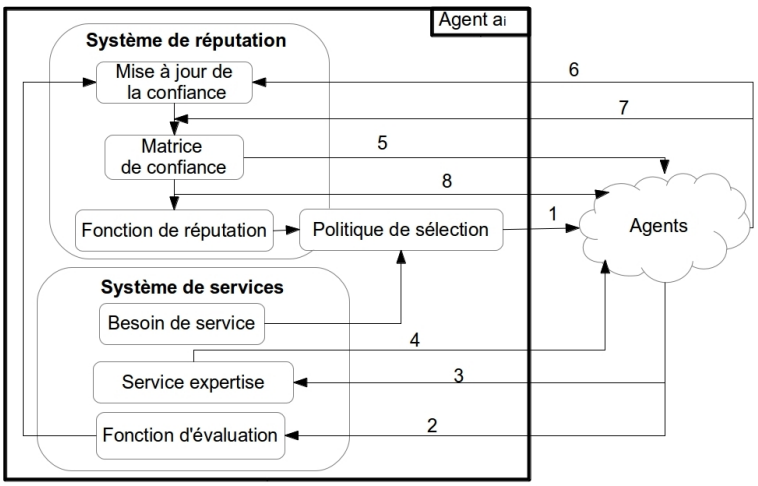

FIGURE 1 – Système d'échange de services

une bonne qualité lorsqu'un autre agent le lui demande. Même si la qualité d'un service dépend de l'expertise de son fournisseur, elle est sujette à l'évalutation du demandeur. Afin de rester général, nous supposons que tout agent $a_i \in N$, évalue les services qu'il reçoit avec sa propre *fonction d'évaluation* $v_i : S \rightarrow V$ où V une échelle d'évaluation commune à l'ensemble des agents. Nous supposons que les agents partagent avec les autres leurs expériences passées, modélisée par une matrice de confiance $T_i : N \times N \times S \rightarrow V$ propre à chaque agent. Ainsi, $T_i[j,k,s]$ désigne la confiance de l'agent $a_j \in N$ envers l'agent $a_k \in N$ pour fournir le service $s_x \in S$. La réputation d'un agent $a_j \in N$ pour un service $s_x \in S$ est une agrégation de toutes les valeurs de confiances des agents vis-à-vis de a_j pout le service s_x. Dans cet article, nous ne nous intéressons pas à la manière dont sont calculées les valeurs de réputation et nous supposons que chaque agent utilise sa *fonction de réputation* $f_i : N \times S \rightarrow \mathbb{R}$. Nous ne faisons pas d'hypothèse sur la fonction de réputation utilisée par les agents et permettons même que deux agents utilisent des fonctions de réputation différentes. Enfin, la *politique de sélection* d'un agent $a_i \in N$ définit comment cette valeur de réputation est utilisée pour déterminer à quel agent demander un service : $\pi_i : S \rightarrow N_x$. La figure 1 résume les différentes interactions entres les agents du système. La flèche 1 est une demande de service de a_i à un autre agent. La flèche 2 est la réalisation de ce service. Les flèches 3 et 4 sont respectivement la demande de service d'un agent a_j et la réalisation de ce service par a_i. Les témoignages des agents sont désignés par les flèches $5, 6, 7$ et 8 (respectivement une demande de témoignage d'un agent a_i à a_j, la réponse de a_j, une demande de témoignage d'un agent a_j à a_i et la réponse de a_i).

3.2 Analogie avec un bandit manchot

L'objectif de la politique de sélection dans un système d'échange de services est de décider à quel agent demander un service. Ce problème est similaire à celui des bandits manchots. Le tableau 1 résume cette analogie entre système d'échange de services et MAB. En effet, considérons un joueur et une machine à sous ayant plusieurs bras. Chacun de ces bras est associé à une fonction de gain dont la loi est a priori inconue. Le problème est de décider quelle séquence de bras utiliser afin de maximiser le gain cumulé. Dans les deux modèles, bandit manchot et échange de services, les expériences passées sont utilisées pour estimer la qualité d'un service futur (le gain) d'un agent (d'un bras) s'il est sélectionné. Ainsi, nous pouvont modéliser un système d'échange de services par un ensemble de bandits manchots, où chaque bandit correspond à un service et chaque bras à un agent capable de fournir ce service.

Définition 3 *Soit $\langle N, S \rangle$ un système d'échange de services. Le MAB correspondant est un ensemble M de bandits manchots tel que $|M| = |S|$ et que $\forall s_x \in S$, la machine à sous $m_x \in M$ dispose de $|N_x|$ bras. Le gain espéré du bras $m_{x,i}$ est l'expertise de l'agent a_i pour le service s_x : $\varepsilon_{i,x}$.*

Ainsi, un agent a_i désirant recevoir le service s_x doit décider quels bras du bandit $m_x \in M$ tirer afin de maximiser son gain cumulé. Dans ce MAB, les agents communiquent pour partager leurs observations et ainsi mieux estimer le gain espéré de chaque bras. Cependant, certains témoignages peuvent être faux. Un système de réputation aide alors les agents à agréger les observations en donnant une valeur de réputation à chaque bras. Cette valeur ne correspond pas au gain espéré (par exemple la réputation d'EigenTrust correspond au ratio du gain fourni par un agent sur l'ensemble des gains obtenus) mais nous faisons l'hypothèse que pour deux bras $m_{x,j}$ et $m_{x,k}$, l'inégalité $f_i(a_j, s_x) > f_i(a_k, s_x)$ implique que le gain espéré du premier bras est meilleur que celui du second. Comme le gain espéré et la réputation sont corrélés (le bras ayant le

	Système d'échange de services	MAB
Objectif	Maximiser la qualité des services reçus	Maximiser le gain
Acteurs	Consommateurs et fournisseurs	Joueurs et bandits
Interactions	Demander un service	Tirer un bras
Capacité	Expertise	Fonction de distribution des gains
Gain	Qualité d'un service	Gain
Observations	Matrice de confiance	Observations passées
Communication	Témoignage sur un autre agent	Témoignage sur un bras
Réputation	Comportement futur espéré	Gain espéré
Politique	Déterminer le futur fournisseur de service	Déterminer le futur bras à utiliser
Manipulations	Agents malveillants	Adversaire

TABLE 1 – Analogie entre système d'échange de services et MAB

meilleur gain espéré correspond à celui de l'agent ayant la meilleure réputation pour le service s_x), nous considérons que la réputation d'un agent pour un service est une approximation du gain espéré du bras correspondant. Ce MAB est un bandit manchot non-stationnaire [14] en présence d'un adversaire [5] car certains agents, appelés agents malveillants, choisissent à chaque pas de temps l'espérance de gain des bras qu'ils contrôlent et peuvent ainsi volontairement fournir des services de mauvaise qualité.

4 Stratégies des agents

La présence d'agents malveillants dans le système nécéssite de définir les stratégies utilisées par les agents, qu'ils soient honnêtes en cherchant à maximiser leurs gains ou malveillants en cherchant à fournir de mauvais services .

4.1 Stratégies malveillantes

Dans un système d'échange de services, un groupe d'agents malveillants formant une coalition peut choisir la qualité des services qu'ils fournissent et donner de faux témoignages. Nous faisons deux hypothèses : (1) tous les agents malveillants sont dans une même coalition (notée $M \subset N$) et (2) leur objectif est de fournir le maximum de mauvais services. Il existe de nombreuses stratégies de manipulation telles que la diffamation ou la promotion dont le but est de modifier les valeurs de réputation des agents à un instant donné. Ces manipulations consistent à fournir de faux témoignages lors du partage des expériences passées entre les agents. Un agent malveillant peut aussi quitter puis entrer à nouveau dans le système avec une nouvelle identité. Ceci lui permet de bénéficier d'une valeur de réputation initiale et de faire oublier sa mauvaise réputation. Une telle manipulation est appelé blanchiment [11]. D'autres manipulations telles que l'attaque oscillante [19] s'appliquent sur le long terme. Nous considérons alors que les agents malveillants suivent la stratégie globale suivante.

Définition 4 *Soit une coalition d'agents malveillants M partionnée en deux sous-groupes M_1 et M_2. À chaque pas de temps, (1) M_1 diffame $N \setminus M$ tout en fournissant de mauvais services ; (2) les agents de M_2 promeuvent ceux de M_1 tout en fournissant des services de bonne qualité selon leurs facteurs d'expertises. Si un agent de M_1 atteint une valeur de réputation trop basse, il rejoint M_2 sous une fausse identité tandis qu'un agent de M_2 rejoint M_1.*

4.2 Politiques de sélection

Nous proposons d'utiliser les politiques canoniques des bandits manchots comme politiques de sélection d'un système d'échange de services. Ces politiques sont des compromis entre l'exploitation des connaissances des agents et l'exploration du système permettant d'affiner ces connaissances. Nous adaptons ici deux d'entre elles, UCB (*Upper Confidence Bound*) et la politique ε-gloutonne, et en proposons une troisième : l'ε-élitisme. Rappelons que nous supposons que la réputation d'un agent est une approximation de la qualité d'un futur service qu'il peut fournir. Notre adaptation d'UCB est définie comme suit :

Définition 5 *Un agent $a_i \in N$ qui désire le service $s_x \in S$ suit la politique UCB s'il sélectionne l'agent $a_j \in N_x$ qui maximise $f_i(a_j, s_x) + \sqrt{\frac{2\ln(1+n_x)}{1+n_{j,x}}}$ où $n_{j,x}$ est le nombre de fois que l'agent a_j a déjà fourni le service s_x à a_i et n_x le nombre de fois que a_i a reçu le service s_x.*

Notons que cette adaptation d'UCB permet de conserver la propriété d'ouverture du système en incitant les agents à interagir avec ceux qu'ils ne connaissent pas. Cependant, ce bonus d'exploration s'applique également aux agents malveillants effectuant un blanchiment. La politique ε-gloutonne [4] est la suivante :

Définition 6 *Un agent $a_i \in N$ suit une* politique ε-gloutonne *s'il demande le service $s_x \in S$ à l'agent $a_j \in N_x$ ayant la meilleure valeur de réputation avec une probabilité $1 - \varepsilon$ ou, avec une probabilité ε, le sélectionne uniformément parmi N_x.*

Nous proposons une troisième politique appelée ε-élitisme. Un agent suivant cette politique sélectionne le futur fournisseur de services uniformément parmi les $\lceil \varepsilon \times |N_x| \rceil$ agents de N_x ayant les meilleures valeurs de réputation :

Définition 7 *Soit $N_{x,\varepsilon} \subseteq N_x$ tel que $|N_{x,\varepsilon}| = \lceil \varepsilon \times |N_x| \rceil$ et que $\forall a_j \in N_{x,\varepsilon}, \nexists a_k \in N_x \setminus N_{x,\varepsilon} : f_i(a_j, s_x) < f_i(a_k, s_x)$. Un agent $a_i \in N$ suit une* politique ε-élitiste *s'il sélectionne aléatoirement uniformément a_j dans $N_{x,\varepsilon}$.*

4.3 Évalutation du système

Afin d'évaluer les politiques face aux manipulations, nous proposons trois métriques : l'efficacité du système, le coût de la manipulation et l'équilibre de charge. Une mesure classique de la performance des politiques de bandits manchots est le *regret* [4, 21]. Comme l'objectif du système d'échange de services est de maximiser le nombre de bons services fournis, l'efficacité du système est définie comme le complémentaire du regret. De plus, afin de maintenir une bonne valeur de réputation, les agents malveillants doivent parfois fournir de bons services. Nous définissons donc un coût de la manipulation. Nous désirons aussi maintenir la propriété d'ouverture de notre système afin d'éviter que seul un petit sous-ensemble des agents soit sollicité, ce qui peut produire des surcharges.Ainsi, nous mesurerons l'équilibre des charges au sein du système comme [13].

Définition 8 *Soit $N^t \subseteq N$ les agents fournisseurs de service à l'instant t. Soit R_i les services reçus par a_i et $R_i^+ \subseteq R_i$ ceux de bonne qualité. Soit F_i les services fournis par a_i et $F_i^+ \subseteq F_i$ ceux de bonne qualité. L'*efficacité du système *est donné par* $\sum_{a_i \in N} |R_i^+| \ / \ \sum_{a_i \in N} |R_i|$. *Le* coût de la manipulation *est donné par* $\sum_{a_i \in M} |F_i^+| \ / \ \sum_{a_i \in N} |F_i|$. *L'*équilibre des charges *est donné par* $|N^t| \ / \ |N|$.

5 Étude empirique

Nous considérons les deux systèmes de réputation EigenTrust [13] et BetaReputation [12] en comparant les politiques présentées précédements et deux politiques utilisées classiques : la politique 0-gloutonne qui sélectionne toujours l'agent ayant la meilleure réputation classiquement utilisé et la politique probabiliste proposée pour EigenTrust [13]. Cette dernière sélectionne l'agent avec qui interagir à partir d'une roue biaisée proportionnellement à sa valeur de réputation avec une probabilité $0,9$ ou un agent n'ayant pas encore fourni de service avec une probabilité $0,1$.

5.1 Protocole

Afin de simplifier notre étude, nous ne considérons qu'une seule catégorie de service. À chaque pas de temps, chaque agent demande à un autre de lui fournir ce service qui est rendu en un seul pas de temps (un agent peut fournir plusieurs services simultanément). Le facteur d'expertise des agents est tiré aléatoirement uniformément. Nous considérons 100 agents sur 100 pas de temps avec une probabilité $0,01$ qu'un nouvel agent honnête rejoigne ou quitte le système à chaque pas de temps. Puis, nous introduisons 10 agents malveillants sur 1000 pas de temps. Nous réitérons nos simulation 50 fois et calculons les moyennes de nos différentes mesures présentées en section 4.3. Les agents malveillants suivent la stratégie donnée en section 4.1 et les politiques de sélection considérées sont la politique uniforme, UCB, les politiques $0,2$-gloutonnes, $0,2$-élitistes [2].

2. Des expérimentations avec $\varepsilon \in [0,1]$ ont été réalisées. Nous ne présentons que les résultats représentatifs par soucis de lisibilité.

5.2 Résultats et analyse

Le principal résultat de cette étude est que la politique de sélection a une large influence sur la robustesse du système d'échange de services. La politique UCB (qui minimise le regret dans le cadre des bandits manchots) est clairement sensible aux manipulations. Cependant, elle est également très coûteuse pour les agents malveillants. À l'opposé, la robustesse d'un sytème utilisant une politique ε-gloutonne dépend principalement de la fonction de réputation utilisée. De plus, manipuler un tel système est peu coûteux pour les agents malveillants. Enfin, une politique ε-élitiste trace un compromis entre le fort coût de manipulation de UCB et le faible taux de mauvais services fournis avec une politique ε-gloutonne. Les figures 2.1 et 3.1 montrent la performance du système en fonction des politiques. UCB est clairement sensible aux coalitions d'agents malveillants, et ceci même avec BetaReputation qui est plus robuste qu'EigenTrust. À l'opposé, la politique $0,2$-gloutonne sur BetaReputation est robuste, ce qui ne pas le cas avec EigenTrust. Comme EigenTrust est une fonction de réputation sensible aux manipulations, les agents malveillants peuvent y avoir une haute valeur de répution et ils sont sélectionnés par la politique gloutonne. Ainsi, il semblerait que la robustesse d'un système utilisant une politique gloutonne est fortement liée à sa fonction de réputation utilisée, ce qui n'est pas le cas avec UCB. Notons que la politique $0,2$-élitiste est moins efficace que la politique $0,2$-gloutonne avec BetaReputation mais est également moins sensible aux manipulations avec EigenTrust. La politique $0,2$-gloutonne est légérement plus performante qu'une politique 0-gloutonne. Ceci est du à son facteur d'exploration qui permet de découvrir des agents ayant une forte expertise mais n'ayant pas encore suffisament interagit pour avoir une haute valeur de réputation. En revanche, ce facteur d'exploration rend la politique $0,2$-gloutonne plus sensible aux manipulations. De plus, si ce facteur d'exploration aussi présent dans UCB facilite la manipulation de BetaReputation, il permet sur EigenTrust aux agents honnêtes diffamés de continuer à interagir. Par ailleurs, même si UCB est sensible aux manipulations, les figures 2.2 et 3.2 nous montrent qu'elle est également coûteuse pour les agents malveillants. En effet, afin de maintenir une bonne valeur de réputation, les agents malveillants doivent fournir d'avantage de bons services que de mauvais. La manipulabilité d'UCB provient du fait qu'elle sélectionne les agents ayant le moins souvent interagi avec les autres afin de mieux les évaluer. Ainsi, afin de manipuler le système, les agents malveillants doivent fréquement effectuer des blanchiments, ce qui est coûteux. La politique $0,2$-gloutonne, elle, entraîne des manipulations à faible coût. Avec EigenTrust, les agents malveillants peuvent fournir un grand nombre de mauvais services sans avoir besoin d'en fournir de bons afin de maintenir leur réputation. La politique $0,2$-élitiste est un compromis entre robustesse et coût : les agents malveillants peuvent fournir de mauvais services mais ils doivent également en fournir de bons afin de conserver leur réputation

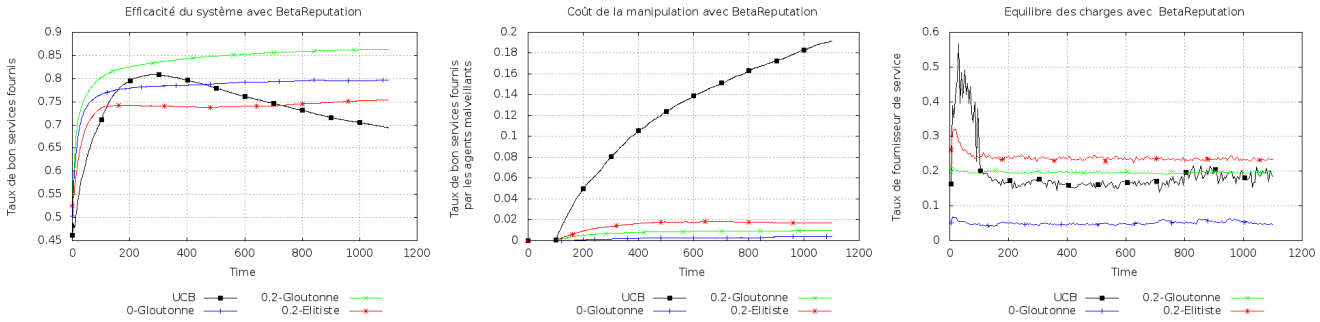

FIGURE 2 – (1) Effacité du système ; (2) coût de la manipulation ; (3) équilibre de charge sous BetaReputation

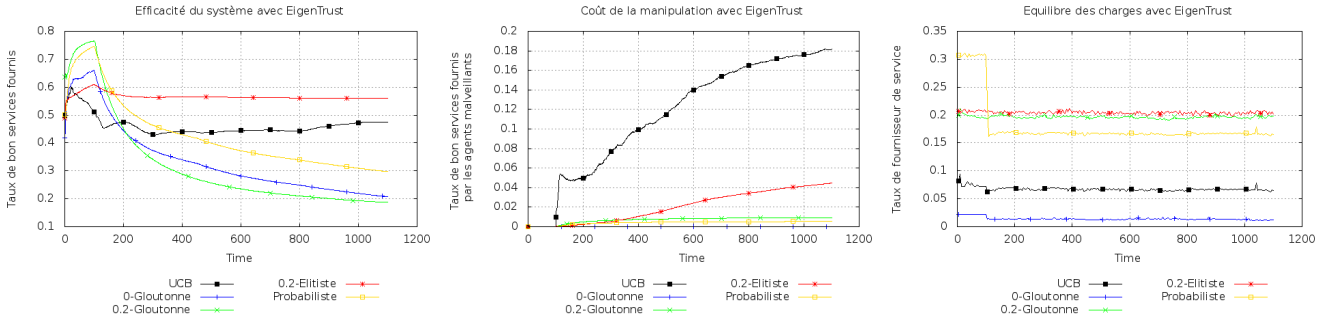

FIGURE 3 – (1) Effacité du système ; (2) coût de la manipulation ; (3) équilibre de charge sous EigenTrust

à un haut niveau. Tout comme la politique $0,2$-gloutonne, les politiques 0-gloutonne et probabiliste ont un coût de manipulation très faible (voir même nul pour la politique 0-gloutonne sur EigenTrust). L'équilibre des charges présenté en figures 2.3 et 3.3 nous montre le degré d'ouverture du système. Notons que la politique gloutonne sélectionne toujours le même sous-ensemble d'agents. Ainsi, la probabilité qu'un nouvel agent puisse être sélectionné est faible. Utiliser une politique gloutonne produit donc une robustesse au blanchiment mais réduit l'ouverture du système. Par ailleurs, si un agent malveillant réussit à obtenir une meilleure valeur de réputation que les agents honnêtes, cet agent malveillant sera fréquemment sélectionné et pourra fournir des mauvais services. C'est en particulier le cas de la politique 0-gloutonne sur EigenTrust. La politique probabiliste d'EigenTrust tend à équilibrer les charge en fonction de la distribution des valeurs de réputation. Plus l'écart-type entre les valeurs de répuation est important, plus le système va se cantonner à un petit sous-ensemble de fournisseurs. Remarquons qu'UCB sélectionne aussi un petit sous-ensemble de fournisseurs de services. Cependant, de part son facteur d'exploration supérieur, la propriété d'ouverture du système est conservée. Pour conclure cette étude, nous avons montré que suivre la politique UCB rend le système certes manipulable mais coûteux par les agents malveillants. À l'opposé, la robustesse d'un système d'échange de services utilisant une politique gloutonne dépend de la robustesse de la fonction de réputation utilisée. De plus, cette politique est quasiment sans coût pour les agents malveillants. La politique élitiste est alors un compromis entre robustesse et coût des manipulations. Enfin, nous avons également montré que la robustesse au blanchiment a un coût sur la propriété d'ouverture du système.

6 Conclusion

Le problème de sélection dans les systèmes de réputation étant similaire à celui des bandits manchots, nous avons étudié l'influence des politique UCB, ε-gloutonne et ε-élitiste sur la robustesse de ces systèmes. Nous montrons que ces politiques sont soit sensibles aux manipulations mais coûteuses pour les agents malveillants, soit robustes (en fonction du mécanisme de réputation utilisé) mais peu coûteuses à manipuler. Trouver une politique à la fois robuste et coûteuse qui ne réduit pas l'ouverture du système reste toutefois un problème ouvert. Une solution consisterait à définir une échelle commune entre les différents paramètres (efficacité du système, coût de la manipulation, degrés d'ouverture du système) afin de se ramener à un problème d'optimisation monocritère. Il serait aussi intéressant d'étudier l'agrégation de différentes politiques de sélection. En effet, cette approche a déjà été considérée sur les bandits manchots [5] où les joueurs disposent de plusieurs politiques qui sont utilisées selon leurs connaissances. Enfin, nous souhaitons pousser plus loin l'analogie entre bandits manchots et systèmes de réputation. Pour cela, il serait intéressant de modéliser un système où la confiance dans la capacité à réaliser une tâche et la confiance envers les témoignages des agents sont distincts.

Références

[1] Alon Altman and Moshe Tennenholtz. Ranking systems : the PageRank axioms. In *6th ACM EC*, pages 1–8, 2005.

[2] Alon Altman and Moshe Tennenholtz. An axiomatic approach to personalized ranking systems. *JACM*, 57(4) :26, 2010.

[3] Venkatachalam Anantharam, Pravin Varaiya, and Jean Walrand. Asymptotically efficient allocation rules for the multiarmed bandit problem with multiple plays. *IEEE AC*, 32(11) :968–976, 1987.

[4] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2-3) :235–256, 2002.

[5] Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. Gambling in a rigged casino : the adversarial multi-armed bandit problem. In *36th FOCS*, 1995.

[6] Peter Auer and Ronald Ortner. UCB revisited : Improved regret bounds for the stochastic multi-armed bandit problem. *Periodica Mathematica Hungarica*, 61(1-2) :55–65, 2010.

[7] Baruch Awerbuch and Robert D Kleinberg. Competitive collaborative learning. In *Learning Theory*, pages 233–248. Springer, 2005.

[8] Alice Cheng and Eric Friedman. Sybilproof reputation mechanisms. In *3rd P2PECON*, pages 128–132, 2005.

[9] Federico Dini and Giancarlo Spagnolo. Buying reputation on eBay : Do recent changes help ? *IJEB*, 7(6) :581–598, 2009.

[10] Michal Feldman, Kevin Lai, Ion Stoica, and John Chuang. Robust incentive techniques for peer-to-peer networks. In *5th ACM EC*, pages 102–111, 2004.

[11] Kevin Hoffman, David Zage, and Cristina Nita-Rotaru. A survey of attack and defense techniques for reputation systems. *CSUR*, 42(1) :1, 2009.

[12] Audun Josang and Roslan Ismail. The Beta reputation system. In *15th Bled EC*, pages 41–55, 2002.

[13] Sepandar D Kamvar, Mario T Schlosser, and Hector Garcia-Molina. The EigenTrust algorithm for reputation management inP2P networks. In *12th WWW*, pages 640–651, 2003.

[14] DE Koulouriotis and A Xanthopoulos. Reinforcement learning and evolutionary algorithms for non-stationary multi-armed bandit problems. *Applied Mathematics and Computation*, 196(2) :913–922, 2008.

[15] Keqin Liu and Qing Zhao. Distributed learning in multi-armed bandit with multiple players. *IEEE SP*, 58(11) :5667–5681, 2010.

[16] Stephen Paul Marsh. Formalising trust as a computational concept. 1994.

[17] Paul Resnick, Ko Kuwabara, Richard Zeckhauser, and Eric Friedman. Reputation systems. *ACM Communications*, 43(12) :45–48, 2000.

[18] Herbert Robbins. Some aspects of the sequential design of experiments. *AMS*, 58(5) :527–535, 1952.

[19] Mudhakar Srivatsa, Li Xiong, and Ling Liu. TrustGuard : countering vulnerabilities in reputation management for decentralized overlay networks. In *14th WWW*, pages 422–431, 2005.

[20] Joannes Vermorel and Mehryar Mohri. Multi-armed bandit algorithms and empirical evaluation. In *16th ECM*, pages 437–448. 2005.

[21] Yizao Wang, Jean-Yves Audibert, and Rémi Munos. Algorithms for infinitely many-armed bandits. *NIPS*, pages 1729–1736, 2008.