

2009

OPTIMIERUNG VON REPUTATIONSSYSTEMEN - WECHSELWIRKUNGEN ZWISCHEN MANIPULATIONEN UND LÖSUNGSKONZEPTEN

Ivo Reitzenstein

Martin-Luther-Universität Halle-Wittenberg

Ralf Peters

Martin-Luther-Universität Halle-Wittenberg

Follow this and additional works at: <http://aisel.aisnet.org/wi2009>

Recommended Citation

Reitzenstein, Ivo and Peters, Ralf, "OPTIMIERUNG VON REPUTATIONSSYSTEMEN - WECHSELWIRKUNGEN ZWISCHEN MANIPULATIONEN UND LÖSUNGSKONZEPTEN" (2009). *Wirtschaftsinformatik Proceedings 2009*. 106.
<http://aisel.aisnet.org/wi2009/106>

This material is brought to you by the Wirtschaftsinformatik at AIS Electronic Library (AISeL). It has been accepted for inclusion in Wirtschaftsinformatik Proceedings 2009 by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

OPTIMIERUNG VON REPUTATIONSSYSTEMEN - WECHSELWIRKUNGEN ZWISCHEN MANIPULATIONEN UND LÖSUNGSKONZEPTEN

Ivo Reitzenstein, Ralf Peters¹

Kurzfassung

Um die Robustheit von Reputationssystemen gegenüber strategischem Verhalten und Manipulationen herzustellen, werden oft verschiedene Lösungskonzepte miteinander kombiniert. Dieser Beitrag untersucht die Wirksamkeit eines solchen Vorgehens. Die Untersuchung erfolgt auf der Basis einer Multi-Agenten-Simulation, die Konzepte der Evolutionären Spieltheorie anwendet. Die vorliegenden Simulationsergebnisse belegen Wechselwirkungen zwischen Lösungskonzepten und decken zudem neue Aspekte bei der Konstruktion robuster Reputationssysteme auf.

1. Einführung

Im E-Business bieten Plattformen wie eBay, Amazon, Wikipedia und Youtube ihren Nutzern zahlreiche neue Interaktionsmöglichkeiten. Dem Vertrauen zwischen den Teilnehmern kommt dabei eine große Bedeutung zu. Da zumeist einander unbekannte, anonyme Teilnehmer aufeinandertreffen, bestehen häufig erhebliche Unsicherheiten hinsichtlich der Fähigkeiten oder Absichten der Interaktionspartner [4, S. 48-49]. Unsicherheit und fehlendes Vertrauen können jedoch das reguläre Plattformgeschehen behindern oder sogar zum Erliegen bringen, wie beispielsweise auch das Modell des Lemons-Market von Akerlof zeigt [2].

Reputationssysteme bieten eine Lösung dieses Vertrauensproblems, indem sie Bewertungen der Teilnehmer sammeln, aggregieren und wiederum anderen Teilnehmern zur Verfügung stellen [14, S. 46]. Die *Signalisierungsfunktion* macht dabei das bisherige Verhalten des Interaktionspartners sichtbar. Die *Sanktionierungsfunktion* führt dazu, dass Teilnehmer mit einer negativen Reputation gemieden werden. So ist es möglich, sich über potenzielle Interaktionspartner zu informieren und solche mit einer negativen Reputation zu meiden. Dieser als „shadow of the future“ bekannte Effekt schafft einen Anreiz zu vertrauenswürdigem Verhalten [3, S. 126-133].

In der Praxis beeinträchtigen verschiedene Formen strategischen Verhaltens die korrekte Funktion von Reputationssystemen [12, S. 1359]. In der Forschung wurden bereits verschiedene Konzepte zur Lösung dieses Problems vorgeschlagen [5-7, 9, 12]. Die Arbeiten konzentrieren sich zumeist auf die Robustheit gegenüber einzelnen Manipulationsstrategien. Für den praktischen Einsatz stellt sich demgegenüber die Frage nach einem Reputationssystem, das gegen alle relevanten Manipula-

¹ Lehrstuhl für E-Business, Martin-Luther-Universität Halle-Wittenberg, 06099 Halle (Saale)

tionen robust ist. Ein vielversprechender Ansatz besteht darin, verschiedene Lösungskonzepte in ein Reputationssystem zu integrieren, um damit eine breitere Robustheit zu erlangen. Hierbei stellt sich jedoch auch die Frage nach den Wechselwirkungen der miteinander kombinierten Konzepte. So könnte ein Teilkonzept zwar eine bestimmte Manipulationsstrategie unterbinden, im Gegenzug jedoch andere Manipulationen begünstigen. Diese Frage wurde bislang noch nicht untersucht.

In diesem Beitrag sollen die Wechselwirkungen zwischen verschiedenen Lösungskonzepten im Hinblick auf die Robustheit eines Reputationssystems untersucht werden. Als interessanter Vertreter eines aus verschiedenen Einzelkonzepten zusammengesetzten Reputationssystems wird der Sporas-Algorithmus von Zacharia et al. [17] untersucht. Die Untersuchung erfolgt auf der Basis eines evolutionären Bewertungsansatzes. Im Rahmen eines evolutionären Modells wird dazu betrachtet, welche Strategien sich erfolgreich in einer Population durchsetzen. Es wird analysiert, welche Manipulationsstrategien ein bestimmtes Reputationssystem erfolgreich verhindern kann. Der Beitrag ergänzt damit auch den Ansatz von Zacharia et al. [17] um eine bisher, auch von anderen Autoren bemängelte, fehlende Untersuchung der Robustheit des Sporas-Algorithmus [5, S. 1].

Der folgende Abschnitt erläutert Manipulationen von Reputationssystemen und zeigt, wie verschiedene Lösungsansätze im Rahmen des Sporas-Algorithmus kombiniert werden. Anschließend wird im dritten Abschnitt der evolutionäre Bewertungsansatz vorgestellt. Im vierten Abschnitt wird dann das Verhalten der darin tätigen Akteure modelliert. Danach werden im fünften Abschnitt die Ergebnisse der darauf basierenden Multi-Agenten-Simulation diskutiert. Der sechste Abschnitt fasst den Beitrag zusammen und gibt einen Ausblick auf weitere Forschungsfragen.

2. Manipulationen und Lösungskonzepte für robuste Reputationssysteme

Das grundsätzliche Ziel einer Manipulation besteht darin, den bereits durch einige Studien [6, S. 1411-1413] belegten Wert einer guten Reputation auszunutzen. In der Praxis lassen sich verschiedene Manipulationen unterscheiden. Bei der Rest-on-the-Lorels (ROTL)-Strategie baut ein Teilnehmer zunächst eine gute Reputation auf, um deren Wert schließlich in einer betrügerischen Transaktion abzuschöpfen [6, S. 1419]. In Scheintransaktionen mit anderen Teilnehmern können positive Bewertungen gesammelt werden (Ballot-Stuffing), um einen ROTL-Angriff vorzubereiten oder die Folgen einer negativen Bewertung auszugleichen [5]. Konkurrenten können außerdem mit negativen Bewertungen denunziert werden (Bad-Mouthing), um diese im Wettbewerb zu benachteiligen. Bei der Sybil-Attack erlangt ein Teilnehmer verschiedene Scheinidentitäten, um einen Ballot-Stuffing (BS) oder Bad-Mouthing (BM) Angriff durchzuführen [8]. Im Rahmen eines Whitewashing kann ein Teilnehmer eine negative Reputation durch einen Identitätswechsel abstreifen um sich so den negativen Konsequenzen seines unerwünschten Verhaltens zu entziehen [9]. Trittbrettfahrerprobleme [12, S. 1359] und Rachebewertungen [11, S. 24] können schließlich dazu führen, dass Teilnehmer auf die Abgabe einer Bewertung verzichten.

In der Literatur finden sich zahlreiche Lösungskonzepte zur Verbesserung der Robustheit von Reputationssystemen. Zur Verhinderung der ROTL-Strategie wird eine stärkere Gewichtung neuer Bewertungen vorgeschlagen, sodass aktuelles Verhalten einen stärkeren Einfluss auf die Reputation eines Teilnehmers hat [6, S. 1419f.]. Ein Konzept gegen Whitewashing besteht darin, Neulinge generell mit dem geringst möglichen Reputationswert beginnen zu lassen, um so den Anreiz zum Identitätswechsel zu beseitigen [7]. Ein weiterer Ansatz besteht darin, die Möglichkeit eines Identitätsmissbrauches durch eine Anmeldegebühr zu begrenzen oder mittels starker Authentifizierung generell zu beseitigen [9]. Hiermit werden auch die Möglichkeiten einer Sybil-Attack begrenzt.

Andere Lösungskonzepte zielen darauf ab, die Folgen manipulierter Bewertungen im Rahmen der BS- oder BM-Strategie zu mindern. Eine Methode besteht darin, auffällige Bewertungen mittels Cluster-Analyse aufzudecken [5]. Andere Autoren schlagen die Einführung eines Metareputations-systems vor, mit dem die „Richtigkeit“ der Bewertungen wiederum von anderen Teilnehmern bewertet werden kann. Ein verbreiteter Ansatz ist, den Einfluss Einzelner auf die Reputation anderer Teilnehmer zu begrenzen, indem nur deren jeweils neueste Bewertung gezählt wird [17]. Ein anderes Konzept besteht darin, die durch einen Teilnehmer abgegebene Bewertung mit dessen Reputation zu gewichten [13]. Die Zuverlässigkeit der abgegebenen Bewertungen soll dadurch erhöht werden, da man davon ausgeht, dass erfahrene Teilnehmer mit einer guten Reputation tendenziell korrekte Bewertungen abgeben. Außerdem wird so die Effizienz einer Sybil-Attack eingeschränkt, da neue Identitäten nur wenig Einfluss auf die Reputation eines Teilnehmers ausüben können [17].

Bei Sporas werden einige dieser Lösungsansätze kombiniert, um die Robustheit gegenüber verschiedenen Manipulationen sicherzustellen. Die Reputation R eines Teilnehmers wird auf Grundlage der durch einen Interaktionspartner abgegebenen Bewertung W in Abhängigkeit von dessen Reputation R^{other} mit der Vorschrift

$$R_{t+1} = \frac{1}{\theta} \sum_1^t \Phi(R_i) \cdot R_i^{other} \cdot (W_{i+1} - E(R_{i+1}))$$

$$\text{mit } \Phi(R) = 1 - \frac{1}{1 + e^{-\frac{-(R-D)}{\sigma}}} \quad \text{und} \quad E(R_{t+1}) = \frac{R_t}{D} \quad (1)$$

angepasst [17]. Der Parameter D gibt hierbei die maximal erreichbare Reputation an und führt zu einem Wertebereich $R \in [0; D]$ der Reputationsfunktion

Der Parameter θ bewirkt eine Gewichtung der abgegebenen Bewertungen im Zeitverlauf und soll so ein ROTL verhindern. Um den Anreiz zum Whitewashing zu beseitigen, erhalten neue Teilnehmer die geringst mögliche Reputation $\min(R)=0$. Zur Vermeidung der Sybil Attack sehen die Autoren vor, bei der mehrfachen Bewertung seitens eines Teilnehmers nur die jeweils neueste Bewertung zu berücksichtigen. Zusätzlich wird die von einem Teilnehmer abgegebene Bewertung mit dessen Reputation gewichtet, sodass neue Identitäten im Rahmen einer Sybil-Attack wenig Einfluss auf die Reputation ausüben können. Fraglich ist, ob dies ein BS oder BM auch dann verhindert, falls der Angriff durch Kooperation mit „echten“ Teilnehmern zustande kommt. Die Sporas-Funktion enthält darüber hinaus eine Dämpfungsfunktion Φ , deren Zweck nicht genau beschrieben ist. Diese Funktion bewirkt, dass eine Bewertung von Teilnehmern mit einer guten Reputation nur sehr geringe Auswirkungen hat und eignet sich so prinzipiell zur Verhinderung eines BM. Es stellt sich jedoch die Frage, ob dies eventuell einen ROTL-Angriff begünstigt und damit den Gewichtungparameter θ außer Kraft setzt.

3. Evolutionärer Bewertungsansatz

3.1. Basismodell

Die Untersuchung der Robustheit von Sporas erfolgt anhand eines hypothetischen Marktplatzes, auf dem ein homogenes Gut zwischen einer Menge von Käufern K und einer Menge von Verkäufern V gehandelt wird, wobei die Akteure genau einer Marktseite zugeordnet sind, also $K \cap V = \emptyset$ gilt. Das Marktgeschehen findet in Runden $t=1, 2, \dots$ statt.

Entsprechend der Grundannahme evolutionärer Modelle [15, S. 84] werden in jeder Runde t im Rahmen eines random matching zufällige Paare $(k_i^t, v_j^t) \in K \times V$ gebildet, die sich in der in Abbildung 1 dargestellten Entscheidungssituation gegenüber stehen. Der Verkäufer hat für das Produkt eine Wertschätzung $r > 0$ und bietet es zu einem von ihm individuell festgelegten Preis p_j^t an. Der Käufer mit der Zahlungsbereitschaft $z > 0$ entscheidet, ob er das Angebot des Verkäufers annimmt und bezahlt oder ablehnt. Für den Fall der Angebotsannahme entscheidet der Verkäufer anschließend, ob er das Produkt wie vereinbart ausliefert oder den Vertrag bricht, den vom Käufer gezahlten Preis einbehält und das Produkt selbst konsumiert. Es entsteht eine Moral Hazard-Situation, da der Verkäufer in jeder Runde die Möglichkeit hat, seinen eigenen Gewinn durch unehrliches Verhalten auf Kosten des Käufers um $p - (p - r) = r$ zu erhöhen.

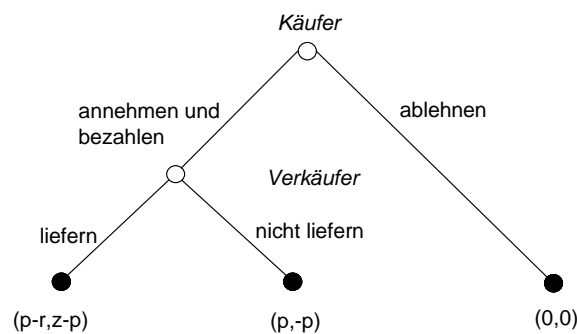


Abbildung 1: Einfaches Vertrauensspiel

Als Lösung dieses Problems wird ein Reputationssystem R betrachtet, bei dem die Käufer eine Bewertung $r = \{1, -1\}$ für das Verhalten der Verkäufer abgeben, wobei $r = 1$ vertrauenswürdigen Verhalten und $r = -1$ betrügerisches Verhalten des Verkäufers kennzeichnet. Für jeden Verkäufer j werden die Bewertungen in einem Reputationsprofil $R_j = \{r_1, r_2, \dots, r_t\}$ gesammelt und als aggregierter Reputationswert $f(R_j)$ auf dem Marktplatz veröffentlicht. Ein wesentliches Merkmal des zu untersuchenden Reputationssystems bildet hierbei die Ausgestaltung der Reputationsfunktion $f(\cdot)$. In diesem Beitrag wird die Sporas-Funktion (1) untersucht.

Das Verhalten der Akteure wird durch die von ihnen verfolgte Strategie festgelegt. Für Käufer und Verkäufer existieren jeweils unterschiedliche Strategiemengen S_k und S_v , von denen jeder Akteur jeweils genau eine Strategie s verfolgt. Der Interaktionsverlauf eines Paares (k_i^t, v_j^t) ergibt sich aus den Strategien beider Akteure und führt zu den in Abbildung 1 dargestellten Auszahlungen. Der Erfolg einer Strategie ergibt sich aus der durchschnittlich erzielten Auszahlung und wird als Fitness bezeichnet [10, S. 934-935].

3.2. Imitation und Robustheit

Im Rahmen des evolutionären Simulationsmodells wird angenommen, dass die Akteure von Zeit zu Zeit ihre Strategie überdenken und diese wechseln, falls eine andere Strategie erfolgreicher ist [15, S. 87]. Für diesen Imitationsprozess vergleichen die Akteure in regelmäßigen Abständen ihre Fitness mit der eines zufällig gewählten anderen Teilnehmers und übernehmen dessen Strategie, falls dieser eine höhere Auszahlung erzielt [1, S. 48]. So ergibt sich ein Selektionsprozess, in dem sich Strategien mit einer überdurchschnittlichen Fitness in der Population ausbreiten und Strategien mit unterdurchschnittlicher Fitness abnehmen. Dies kann schließlich zur Bildung von Gleichgewichten führen, bei denen sich der Populationsanteil der Strategien nicht mehr ändert [10, S. 948ff.].

Ausgehend von diesem Konzept lässt sich die Robustheit eines Reputationssystems untersuchen. Ein Reputationssystem ist gegenüber einer Manipulation robust, falls der beschriebene Selektionsprozess zur Verdrängung einer Manipulationsstrategie führt und sich ein Gleichgewicht kooperativen Verhaltens einstellt. Ein perfektes Reputationssystem hat die Eigenschaft, gegenüber allen relevanten Manipulationen robust zu sein.

4. Strategien

4.1. Verhalten der Käufer

In Anlehnung an die Praxis elektronischer Marktplätze wie eBay wird für das Zustandekommen einer Transaktion nur die Reputation auf der Verkäuferseite berücksichtigt. Unter der Annahme, dass die Akteure im Modell nur auf jeweils einer Marktseite agieren, wird keine Reputation auf der Käuferseite betrachtet. Alle Käufer verwenden somit dieselbe Strategie, und es findet damit auch kein Selektionsprozesses auf der Käuferseite statt.

Ein Käufer steht in jeder Runde vor der Entscheidung, ob er das Angebot des Verkäufers annimmt oder ablehnt. Da das Verhalten des Verkäufers ex ante unbekannt ist, stellt dies eine Entscheidung unter Unsicherheit dar. So führt die in Abbildung 1 dargestellte Entscheidungssituation für den Käufer im Fall der Ablehnung zu einer sicheren Auszahlung von Null. Die Annahme des Angebotes führt entweder zu einer Auszahlung von $(z - p_j^t)$ bei der vertragsgemäßen Auslieferung des Gutes oder einem Verlust in Höhe des Kaufpreises im Falle eines Betruges durch den Verkäufer.

Als Grundlage seiner Entscheidung verwendet der Käufer die erwartete Auszahlung für den Fall der Angebotsannahme:

$$E(a) = (z - p_j^t) \cdot p_{\text{erfolg}} - p_j^t \cdot (1 - p_{\text{erfolg}}) \quad (2)$$

Der Erwartungswert basiert auf der Wahrscheinlichkeit p_{erfolg} für den Erfolg der Transaktion. Hierzu liefert das Reputationssystem eine auf dem Reputationsprofil des Verkäufers basierende Erfolgswahrscheinlichkeit $p_{\text{rep}} = f(R_j)$. Zusätzlich sammelt der Käufer auch eigene Erfahrungen auf dem Marktplatz und kann auf Basis seiner Erfahrungen eine subjektive Erfolgswahrscheinlichkeit $p_{\text{sub}} = |\text{Anzahl erfolgreiche Transaktionen}| / |\text{Anzahl aller Transaktionen}|$ berechnen.

Als Erfolgswahrscheinlichkeit verwendet die Käuferstrategie einen gewichteten Mittelwert

$$p_{\text{erfolg}} = g \cdot p_{\text{rep}} + (1 - g) \cdot p_{\text{sub}}, \quad 0 \leq g \leq 1, \quad g \in \mathfrak{R} \quad (3)$$

beider Größen, wobei g das „Vertrauen“ des Käufers in die Korrektheit des Reputationssystems abbildet. Besitzt der Käufer uneingeschränktes Vertrauen in das Reputationssystem, so gilt $g = 1$. Verhält sich ein Verkäufer trotz guter Reputation betrügerisch, so ist dies möglicherweise ein Anzeichen dafür, dass das Reputationssystem manipuliert wurde. In diesem Fall erfolgt eine Verringerung von g und der Einfluss des Reputationssystems auf die Entscheidung des Käufers wird reduziert. Analog erfolgt eine Erhöhung von g , falls sich der Verkäufer entsprechend seiner guten Reputation vertragskonform verhält.

Insgesamt ergibt sich als Entscheidungsregel (4) für den Käufer, dass das Angebot des Verkäufers genau dann angenommen wird, falls die zu erwartende Auszahlung größer als die sichere Auszahlung bei der Ablehnung des Angebotes ist.

$$s_k = \begin{cases} \text{annehmen} & \text{falls } E(a) > 0 \\ \text{ablehnen} & \text{sonst} \end{cases} \quad (4)$$

Diese Entscheidungsregel hat die Eigenschaft, dass der Käufer bei einer guten Reputation des Verkäufers einen höheren Preis akzeptiert. Analog impliziert eine schlechte Reputation ein höheres Ausfallrisiko und reduziert damit den maximal akzeptierten Preis.

4.2. Strategien der Verkäufer

Ein Verkäufer entscheidet in jeder Runde, ob er vertragskonform kooperiert oder aber den Vertrag bricht und die Option „nicht liefern“ wählt. Zusätzlich zu dieser Kooperationsentscheidung legt der Verkäufer zu Beginn jeder Runde den Preis fest, zu dem das Produkt angeboten wird.

Unter dem Ziel der Gewinnmaximierung versucht jeder Verkäufer den Angebotspreis an die aktuelle Zahlungsbereitschaft der Käufer anzupassen. Die Anpassung erfolgt im Simulationsmodell anhand der „Derivative-Follower (DF)-Strategie“ [16, S. 89-90], indem der Angebotspreis der nächsten Runde nach einer erfolgreich durchgeführten Transaktion um den Betrag b erhöht bzw. bei einer Ablehnung um b gesenkt wird. Der Betrag b wird in jeder Runde zufällig aus dem Intervall $[0, s]$ gezogen, wobei s den maximal möglichen Anpassungsschritt angibt.

Bezüglich der Kooperationsentscheidung stehen dem Verkäufer neben einem vertragskonformen Verhalten grundsätzlich die in Abschnitt 2 aufgezählten Manipulationen zur Verfügung. Für die Simulation werden die ROTL-, Whitewashing-, BS- und BM-Strategie umgesetzt.

Die ROTL-Strategie zielt darauf ab, den ökonomischen Wert einer guten Reputation mit einer betrügerischen Transaktion abzuschöpfen. Die hier verwendete ROTL-Strategie kooperiert zunächst für eine Anzahl n Runden, wählt in der folgenden Runde die Handlungsoption nicht liefern und kooperiert daraufhin für weitere n Runden. Diese Strategie führt dazu, dass ein Verkäufer im Simulationsverlauf negative Bewertungen in R_j ansammelt. Die Whitewashing-Strategie erweitert die ROTL-Strategie dahingehend, dass der Verkäufer nach dem Betrug einen Identitätswechsel vornimmt und so ein leeres Reputationsprofil erhält. Damit ist es dem Verkäufer möglich, sich den negativen Folgen einer schlechten Bewertung zu entziehen. Im Rahmen der BS-Strategie versucht ein Verkäufer, seine Reputation durch zusätzliche positive Bewertungen zu verbessern. Diese Strategie kann beispielsweise zum Ausgleich der durch die ROTL-Strategie verursachten, negativen Bewertungen dienen. Die in der Simulation verwendete BS-Strategie erweitert die ROTL-Strategie dahingehend, dass der Verkäufer sein eigenes Reputationsprofil in jeder Runde $t < n$ durch eine zusätzliche positive Bewertung verbessert. Im Gegensatz zu den bisher beschriebenen Strategien zielt die BM-Strategie darauf ab, die Reputationsprofile anderer Teilnehmer zu manipulieren. Die hier verwendete BM-Strategie wählt in jeder Runde einen Verkäufer zufällig aus, der jeweils eine negative Bewertung erhält. Die BM-Strategie führt dazu, dass die Reputation konkurrierender Verkäufer geschädigt wird.

5. Simulationsergebnisse

Zur Untersuchung der Robustheit von Sporas wurde der vorgestellte evolutionäre Bewertungsansatz auf ein Multi-Agenten-System (MAS) übertragen und anhand des *Java Agent Development Frameworks (JADE)* implementiert. Das MAS besteht aus einer Menge von Käufer- und Verkäufer-Agenten, die sich entsprechend der in Abschnitt 4 dargestellten Strategien verhalten. Außerdem

wird ein „Marktplatz“-Agent verwendet, der die Aufgabe des Marktplatzes und der Simulationssteuerung übernimmt. Die Zahlungsbereitschaft der Käufer wird auf $z = 1$ und der Reservationspreis der Verkäufer wird auf $r = 0,1$ festgelegt. Entsprechend der Publikation von Zacharia et al. wird in der Simulation eine Variante von Sporas verwendet, mit der sich die Reputation eines Agenten im Simulationsverlauf mit Hilfe der Funktion

$$R_{t+1} = \frac{1}{\theta} \Phi(R_t) R_{i+1}^{other} \left(W_{i+1} - \frac{R_t}{D} \right) + R_t \quad (5)$$

iterativ berechnen lässt [17]. Der Wertebereich dieser Reputationsfunktion wurde in der Publikation von Zacharia et al. auf $R \in [0; 3000]$ festgelegt. Um das Reputationssystem im Rahmen der Entscheidungsregel (4) verwenden zu können, wird der von R gelieferte Wert in der Simulation auf das Intervall $[0; 1] \in \mathcal{R}$ abgebildet. Dazu wird für die einwandfreie Reputation $\max(R) = 3000$ eine Transaktionswahrscheinlichkeit von Eins, für $\min(R) = 0$ eine Transaktionswahrscheinlichkeit von Null angenommen. Für die Experimente sind außerdem der Gewichtungparameter θ sowie die Dämpfungskonstante σ der Sporas-Funktion festzulegen. Entsprechend der Evaluation von Zacharia et al. wird zunächst $\theta = 10$ gewählt. Da die Auswirkung der Dämpfungsfunktion nicht genau beschrieben ist, wird zunächst eine geringe Dämpfungskonstante von $\sigma = 1,0$ verwendet.

In den nachfolgenden Experimenten werden verschiedene einfache Reputationsszenarien mit einer Populationsgröße von $n=100$ Agenten betrachtet. Zu Beginn des Simulationslaufes verfolgen zunächst alle Agenten die Strategie „kooperieren“. In Runde 50 wird dieses Gleichgewicht durch eine kleine Mutation gestört, in der einzelne Agenten zu einer der beschriebenen Manipulationsstrategien wechseln.

Als Ausgangspunkt wird die Populationsentwicklung bei einer simultanen Mutation jeweils eines Agenten zur BM-, BS-, ROTL- und Whitewashing-Strategie betrachtet. Abbildung 2 zeigt, dass die ROTL- und die BS-Strategie alle anderen Strategien im weiteren Simulationsverlauf verdrängen. Insgesamt zeigt sich, dass das Sporas-Reputationssystem nur gegenüber einem Teil der betrachteten Manipulationen robust ist. Um den Selektionsprozess näher zu untersuchen, wurden anschließend Mutationen zu einzelnen Manipulationsstrategien isoliert betrachtet.

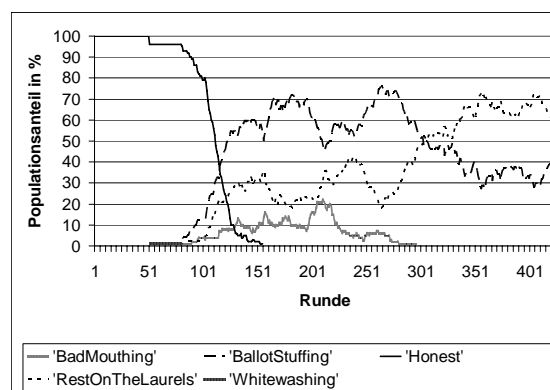


Abbildung 2: Populationsentwicklung im Ausgangsszenario

Bei einer einzelnen Mutation zur Whitewashing-Strategie ist wie im Ausgangsszenario eine relativ schnelle Verdrängung zu beobachten. Der Whitewashing-Agent kann sich nach dem Betrug durch eine Neuansmeldung zwar erfolgreich der negativen Bewertung entziehen. Hierdurch sinkt jedoch seine Reputation auf den Wert Null und der Agent kann keine weiteren Transaktionen durchführen.

Bei einer einzelnen Mutation zur BM-Strategie ist das Reputationssystem insoweit robust, als dass keine Verbreitung der Strategie stattfindet. Die Ursache liegt hier darin, dass die Reputation der kooperierenden Verkäufer aufgrund des niedrigen Dämpfungsfaktors nicht ausreichend beschädigt wird, um der BM-Strategie einen Wettbewerbsvorteil zu verschaffen.

In Abbildung 3 ist der Selektionsprozess für die ROTL-Strategie dargestellt. Die geringe Dämpfungskonstante bewirkt eine schnelle Reaktion des Reputationssystems auf Änderungen und führt zum Erfolg der ROTL-Strategie. Zunächst nutzt der ROTL-Agent seine gute Reputation in Runde 69 für einen Betrug. Die darauf folgende negative Bewertung und die Verschlechterung seiner Reputation führen zu einer Ablehnung seines Angebotes in der folgenden Runde. Aufgrund des DF-Verhaltens senkt der ROTL-Agent daraufhin seinen Angebotspreis und kann anschließend erfolgreich eine weitere Transaktion durchführen. Dies führt dazu, dass sich der Reputationswert des ROTL-Agenten verbessert und sich auch der Angebotspreis in der folgenden Runde wieder erhöht. Aufgrund der geringen Dämpfung der Sporas-Funktion verbessert sich die Reputation des ROTL-Agenten in den folgenden Interaktionen so stark, dass bereits nach 10 Runden ein weiterer Betrug möglich ist und der erneute Reputationsverlust wiederum durch eine kurzfristige Senkung des Angebotspreises ausgeglichen werden kann. Dieser Prozess führt analog auch zum Erfolg der BS-Strategie.

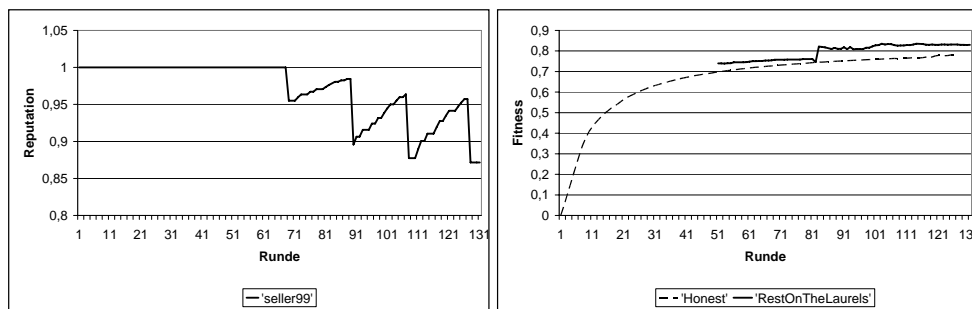


Abbildung 3: Entwicklung der Reputation eines ROTL-Agenten (a) und der durchschnittlichen Fitness der ROTL-Strategie (b) □

Die degressive Gewichtung von Bewertungen im Sporas-Reputationssystem kann die ROTL- und die BS-Strategie entgegen bisheriger Annahmen nicht erfolgreich verhindern, da eine negative Reputation aufgrund der geringen Dämpfung relativ schnell wieder ausgeglichen wird. In weiteren Experimenten wurde daher untersucht, ob eine Anpassung der Dämpfungsfunktion die Robustheit von Sporas gegenüber der ROTL- bzw. BS-Strategie herstellt. Dabei wurde festgestellt, dass die ROTL- und die BS-Strategie bei einem Dämpfungsfaktor von $\sigma = 350$ im Simulationsverlauf verdrängt werden. Die Ursache besteht darin, dass nun auch geringere Reputationswerte von der Dämpfungsfunktion betroffen sind und die Reputationsfunktion mit einer erhöhten Trägheit auf Änderungen reagiert. Die Fitness der ROTL- bzw. BS-Strategie sinkt, da eine höhere Rundenzahl zum Ausgleich des Reputationsverlustes bei einem Betrug benötigt wird.

Abbildung 4 zeigt jedoch, dass unter diesen veränderten Bedingungen die BM-Strategie erfolgreich ist. Der Reputationsverlust der ungerechtfertigt negativ bewerteten Verkäufer kann mit dem DF-Verhalten nur sehr langsam ausgeglichen werden. Dies führt zunächst zu einem Fitnesszuwachs der BM-Strategie gegenüber der kooperierenden Strategie. Aufgrund des Imitationsverhaltens wechseln vermehrt Agenten mit beschädigter Reputation zur BM-Strategie, womit eine Verringerung der durchschnittlichen Reputation der BM-Strategie einhergeht. Dies führt schließlich auch zu einem Fitnessverlust der BM-Strategie, und ein Teil der Agenten wechselt wieder zur kooperierenden Strategie. Der Populationsanteil der BM-Strategie stabilisiert sich bei ungefähr 50 Prozent und zeigt, dass das Reputationssystem grundsätzlich anfällig gegenüber der BM-Strategie ist. Die An-

passung des Dämpfungsfaktors führt somit nur bedingt zu einer Erhöhung der Robustheit von Sporas. Es stellt sich heraus, dass die Lösung eines Problems im Gegenzug eine andere Manipulation erleichtern kann.

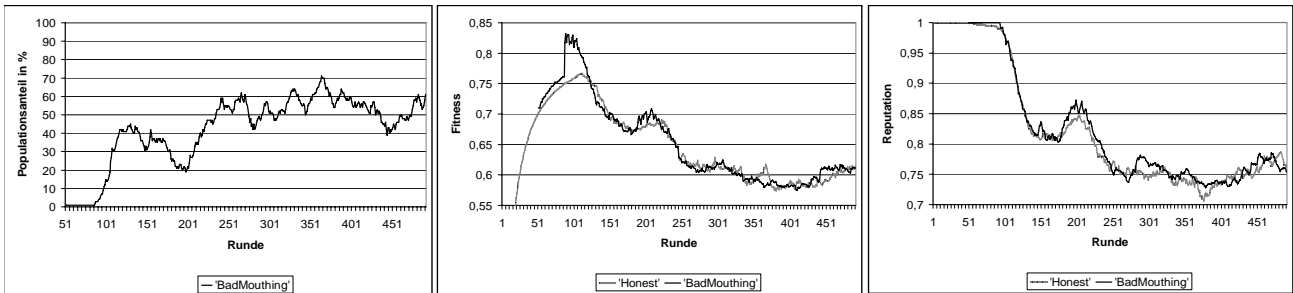


Abbildung 4: Entwicklung des Populationsanteils (a), der Fitness (b) und Reputation (c) der BM-Strategie bei einem Dämpfungsfaktor $\sigma = 350$ □

Da sich die Anpassung des Dämpfungsfaktors nur bedingt zur Erhöhung der Robustheit von Sporas eignet, wurde in einem weiteren Experiment die Auswirkung des Gewichtungsfaktors θ untersucht. Es zeigt sich, dass die Robustheit gegenüber der ROTL- und BM-Strategie mit einem Wert von $\theta = 5$ hergestellt werden kann. Die stärkere Gewichtung führt hier nach einem Betrug der ROTL-Strategie zu einem höheren Reputationsverlust. Die Preisanpassungen im Rahmen des DF-Verhaltens kompensieren den Betrugsgewinn, führen zu einer geringeren Fitness der ROTL-Strategie und das System kehrt zum kooperierenden Gleichgewicht zurück. Trotz des geringen Dämpfungsfaktors bleibt das Reputationssystem unter dieser Bedingung gegenüber der BM-Strategie robust. Der geringe Dämpfungsfaktor führt dazu, dass die kooperierende Strategie den durch die negative Bewertung der BM-Strategie verursachten höheren Reputationsverlust hinreichend schnell ausgleichen kann.

Insgesamt deuten die Ergebnisse auf ein mögliches Optimum des Parameters θ hin. So konnte in der Simulation beobachtet werden, dass bei $\theta < 5$ eine so starke Gewichtung der Bewertungen erfolgt, dass die kooperierende Strategie eine ungerechtfertigte negative Bewertung nicht mehr erfolgreich ausgleichen kann und schließlich durch die BM-Strategie verdrängt wird. Dagegen ist bei $\theta > 5$ die Auswirkung einer negativen Bewertung auf die Reputation geringer, sodass die BM-Strategie verdrängt wird. Jedoch kann die ROTL-Strategie unter dieser Bedingung die negativen Folgen eines Betruges erfolgreich kompensieren und verdrängt die kooperierende Strategie.

6. Zusammenfassung und Ausblick

In diesem Beitrag wurde die Robustheit des Sporas-Reputationssystems anhand eines evolutionären Bewertungsansatzes untersucht. Im Rahmen einer Multi-Agenten-Simulation konnte gezeigt werden, dass Sporas in Abhängigkeit der gewählten Parameter robust gegenüber Manipulationen ist. Dies bestätigt auch, dass sich die aus der Literatur bekannten Lösungskonzepte erfolgreich zu einem robusten Reputationssystem kombinieren lassen. Darüber hinaus konnten Wechselwirkungen zwischen einzelnen Lösungskonzepten aufgezeigt werden. Insbesondere zeigte sich, dass einzelne Lösungskonzepte wiederum andere Manipulationen begünstigen können. Die Ergebnisse der Simulation deuten darauf hin, dass bei der Gewichtung der Bewertungen eine ausgewogene Sanktionierung und Rehabilitierung bedeutende Eigenschaften für die Robustheit des untersuchten Reputationssystems sind.

Weitere Untersuchungen sollen zeigen, ob die in diesem Beitrag identifizierten Aspekte der optimalen Gewichtung auch bei anderen Reputationssystemen gültig sind. Insgesamt zeigt sich, dass

der vorgestellte Bewertungsansatz die bereits vorhandenen Forschungsansätze sinnvoll ergänzen kann. Mit der Ausweitung der Untersuchung auf andere Reputationssysteme und weitere Manipulationsstrategien soll das vorgestellte Forschungsprojekt mit fortschreitendem Verlauf einen weiteren Beitrag zur Konstruktion und Bewertung robuster Reputationssysteme liefern.

7. Literaturverzeichnis

- [1] AMANN, E., Evolutionäre Spieltheorie: Grundlagen und neue Ansätze, Heidelberg 1999.
- [2] AKERLOF, G. A., The Market for „Lemons“: Quality and Uncertainty and the Market-Mechanism, in: Quarterly Journal of Economics. Bd. 84 (1970), S. 488-500.
- [3] AXELRODT, R., The Evolution of Cooperation, London et al. 1990.
- [4] BIERHOFF, H. W.; VORNEFELD, B.: The Social Psychology of Trust with Applications in the Internet, in: Analyse & Kritik. Zeitschrift für Sozialtheorie, Bd. 1 (2004), S. 48-62.
- [5] DELLAROCAS, C., Immunizing online reputation reporting systems against unfaier ratings and discriminating behavior, in: Proceedings of the 2nd ACM conference on electronic commerce, Minnesota 2000.
- [6] DELLAROCAS, C., The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms, in: Management Science, Bd. 49 (2003), S. 1407-1424.
- [7] DELLAROCAS, C., Efficiency and Robustness of eBay-like Online Feedback-Mechanisms in Environments with pure Moral Hazards, Working-Paper, MIT Sloan School of Management 2003.
- [8] Douceur, J. R., The Sybil Attack, in: Proceedings of the IPTPS02 Workshop, Cambridge (MA) 2002.
- [9] FRIEDMANN, E.; RESNICK, P., The Social Cost of Cheap Pseudonyms, in: Journal of Economics and Management Strategy, Bd. 10 (2001), S. 173-199.
- [10] HAMMERSTEIN, P.; SELTEN, R., Game Theory and Evolutionary Biology, in: R.Aumann; S. Hart Handbook of Game Theory (Bd 2), 2. Auflage, Amsterdam et al. 2003, S. 929-993.
- [11] KENNES, J.; SCHIFF, A., The Value of a Reputation System, Working Paper, University of Auckland 2003.
- [12] MILLER, N.; RESNICK, P.; ZECKHAUSER, R., Eliciting Honest Feedback: The Peer Prediction Method, in: Management Science, Bd. 51 (2005), S. 1359-1373.
- [13] MUI, L.; MOHTASHEMI, M; HALBERSTADT, A., A Computational Model of Trust and Reputation, in: Proceedings of the 35th Annual Hawaii International Conference of System Science 2002.
- [14] RESNICK, P.; ZECKHAUSER, R.; FRIEDMANN, E.; KUWABARA, K., Reputation Systems, in: Communications of the ACM, Bd. 43 (2000), S. 45-48.
- [15] SAMUELSON, L., Evolutionary Games and Equilibrium Selection, Cambridge et al. 1997.
- [16] ZACHARIA, G. et al., Economics of Dynamic Pricing in a Reputation Brokered Agent Mediated Marketplace, in: Proceedings of the 23rd Hawaii International Conference on System Science 1999.
- [17] ZACHARIA, G., MOUKAS, A., MAES, P., Collaborative reputation mechanisms in electronic Marketplaces, in: Proceedings of the 32nd Hawaii International Conference on System Science 1999.