

**Vinicius da Silva Almendra**

**Um estudo de identificação de fraudadores em mercados eletrônicos através da computação humana**

**Tese de Doutorado**

Tese apresentada como requisito parcial para obtenção do grau de Doutor pelo Programa de Pós-graduação em Informática do Departamento de Informática do Centro Técnico Científico da PUC-Rio.

Orientador: Daniel Schwabe

Rio de Janeiro  
Setembro de 2008

**Vinicius da Silva Almendra**

**Um estudo de identificação de fraudadores em mercados eletrônicos através da computação humana**

Tese apresentada como requisito parcial para obtenção do grau de Doutor pelo Programa de Pós-graduação em Informática do Departamento de Informática do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

**Prof. Daniel Schwabe**

Orientador

Departamento de Informática – PUC-Rio

**Prof. Marco Antonio Casanova**

Departamento de Informática – PUC-Rio

**Prof. Eduardo Sany Laber**

Departamento de Informática – PUC-Rio

**Prof<sup>a</sup>. Ana Cristina Bicharra Garcia**

Departamento de Ciência da Computação – Uff

**Prof. Ricardo Miyashita**

Departamento de Engenharia Industrial – Uerj

**Prof<sup>a</sup>. Bianca Zadrozny**

Departamento de Ciência da Computação – Uff

**Prof. José Eugenio Leal**

Coordenador Setorial do Centro  
Técnico Científico

Rio de Janeiro, 18 de setembro de 2008

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização do autor, do orientador e da universidade.

### **Vinicius da Silva Almendra**

Graduou-se em Engenharia de Computação na Pontifícia Universidade Católica do Rio de Janeiro (1999). Concluiu o Mestrado em Engenharia Elétrica pela Escola Politécnica da Universidade de São Paulo (2003). Atualmente é pós-doutorando do Departamento de Informática da Pontifícia Universidade Católica do Rio de Janeiro.

#### Ficha Catalográfica

Almendra, Vinicius da Silva

Um estudo de identificação de fraudadores em mercados eletrônicos através da computação humana / Vinicius da Silva Almendra; orientador: Daniel Schwabe. – 2008.

122 f. : il. ; 30 cm

Tese (Doutorado em Informática)–Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2008.

Inclui bibliografia

1. Informática – Teses. 2. Fraudes. 3. Comércio eletrônico. 4. Computação humana. 5. Recomendação. I. Schwabe, Daniel. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Aos meus pais, Geraldo e Lucia,  
que dedicaram suas vidas  
de forma abnegada e silenciosa  
a mim e a minha querida irmã.

## Agradecimentos

Ao meu orientador, Professor Daniel Schwabe, que me introduziu nas sendas da pesquisa científica com paciência, tenacidade e generosidade.

Ao CNPq, à CAPES, à PUC-Rio e ao programa UOL Bolsa Pesquisa pelos auxílios concedidos, sem os quais este trabalho não poderia ser realizado.

Aos participantes do teste piloto, que auxiliaram esta pesquisa com uma dedicação generosa do seu tempo.

Aos professores que participaram da Comissão examinadora, por suas valiosas observações.

A todos os companheiros da PUC-Rio – professores, funcionários e colegas – que de um modo ou de outro colaboraram com a realização deste trabalho.

A todos os amigos e irmãos que me apoiaram nas horas difíceis.

Aos meus pais e à minha irmã, por tudo.

A Deus e à Sua Mãe Santíssima, por tantos e inumeráveis benefícios concedidos. Tibi gloria in saecula.

## Resumo

Almendra, Vinicius da Silva; Schwabe, Daniel. **Um estudo de identificação de fraudadores em mercados eletrônicos através da computação humana**. Rio de Janeiro, 2008. 122p. Tese de Doutorado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Comportamento fraudulento é um problema crescente para mercados eletrônicos, particularmente sítios de leilão, causando diversos tipos de perdas. As medidas para redução das perdas por fraude geralmente têm como subproduto indesejável o constrangimento e até mesmo a exclusão de usuários inocentes, criando um difícil equilíbrio entre perdas causadas por fraudadores e perdas causadas por excesso de restrições aos participantes do mercado. O objetivo desta tese é mostrar a viabilidade de uma nova abordagem para redução de perdas por fraude em sítios de leilão: o jogo “pega ladrão”. Esta abordagem leva em conta explicitamente o equilíbrio acima mencionado e é baseada no paradigma da computação humana, no qual pessoas executam tarefas computacionais por diversão ou lucro. A metodologia utilizada foi uma pesquisa exploratória acerca de atividade fraudulenta em um mercado eletrônico real, um teste piloto de detecção de fraudadores por agentes humanos e o desenvolvimento e simulação do elemento central do jogo proposto: o mecanismo de identificação de fraudadores. A pesquisa exploratória apresenta um perfil da fraude de não-entrega no maior sítio de leilão do Brasil, mostrando-a como um problema real, recorrente e mensurável. O teste piloto aponta evidências positivas de que agentes humanos de fato conseguem distinguir fraudadores de vendedores normais por uma margem significativa. A simulação dá suporte para a utilidade do mecanismo proposto na redução de perdas por fraude. Os resultados obtidos confirmam o jogo “pega ladrão” como uma abordagem viável para reduzir perdas por fraude em mercados eletrônicos.

## Palavras-chave

Fraudes; comércio eletrônico; computação humana; recomendação.

## Abstract

Almendra, Vinicius da Silva; Schwabe, Daniel (Advisor). **A study on fraudster identification in electronic markets through human computation**. Rio de Janeiro, 2008. 122p. D.Sc. Thesis – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Fraudulent behavior is an increasing problem for electronic markets, in particular for online auction sites, causing several types of loss. Fraud loss reduction measures generally have as an undesirable by-product the harassment and even exclusion of bona fide users, creating a difficult trade-off between losses with fraudsters and losses due to excessive constraints on market participants. The objective of this thesis is to show the viability of a novel approach to fraud loss reduction in online auction sites, the “catch the thief” game. This approach takes explicitly into account the aforementioned trade-off and is based on the paradigm of human computation, where people do computational tasks for fun or profit. The methodology used was an exploratory research on fraudulent activity in a real electronic market, a pilot test of fraudster detection by human agents, and the development and simulation of the proposed game's core element, the fraudster identification mechanism. The exploratory research presents a profile of non-delivery fraud in the biggest Brazilian online auction site, showing it as real, recurring and measurable problem; the pilot test displays positive evidence that unspecialized human agents can indeed distinguish fraudulent sellers from normal ones by a significant margin; the simulation supports the usefulness of the proposed mechanism for fraud loss reduction. The results obtained confirm “catch the thief” game as a viable approach to reduce fraud loss in electronic markets.

## Keywords

Frauds; e-commerce; human computing; recommendation.

# Sumário

1	Introdução	14
1.1.	Objetivo	17
1.2.	Solução proposta	17
1.3.	Metodologia	20
1.3.1.	Revisão da literatura	20
1.3.2.	Pesquisa exploratória sobre atividade fraudulenta	21
1.3.3.	Teste do uso de pessoas na identificação de fraudadores	22
1.3.4.	Mecanismo de identificação de fraudadores	22
1.3.5.	Limitações da metodologia adotada	23
1.4.	Estrutura do trabalho	23
2	Revisão da literatura	25
2.1.	Comércio eletrônico e sistemas de reputação	25
2.2.	Mercados eletrônicos e atividade fraudulenta	29
2.2.1.	Fraudes em números	30
2.2.2.	Anatomia das fraudes	30
2.2.3.	Impacto das fraudes nos operadores	32
2.2.4.	Identificando fraudes	33
2.3.	Limitações das soluções existentes	34
2.4.	Mecanismos de revelação de informação	36
2.4.1.	Computação humana	38
3	Atividade fraudulenta nos mercados eletrônicos	40
3.1.	Objetivo da coleta de dados	40
3.2.	Descrição do MercadoLivre	40

3.3.	Detalhamento da metodologia	42
3.4.	Resultados	46
4	Teste piloto com agentes humanos	53
4.1.	Medindo o desempenho dos agentes humanos	53
4.2.	A questão dos incentivos	54
4.3.	Projeto do teste piloto	55
4.3.1.	Perfis dos vendedores a serem analisados	56
4.3.2.	Instrumento de coleta de dados	59
4.3.3.	Instruções aos participantes	60
4.3.4.	Amostra utilizada	61
4.4.	Resultados	61
4.5.	Análise dos resultados	66
5	O mecanismo de identificação de fraudadores	67
5.1.	Modelo da atuação fraudulenta	67
5.2.	Modelo da atuação dos jogadores	70
5.3.	Mecanismo de classificação de vendedores	72
5.3.1.	Integrando as respostas de múltiplos classificadores	74
5.3.2.	Melhorando o desempenho do mecanismo	81
5.3.3.	Exemplo	83
5.4.	Mecanismo de incentivo aos classificadores	84
5.4.1.	Comportamentos estratégicos	85
5.4.2.	Avaliando classificadores	90
5.4.3.	Remunerando os classificadores	91
5.4.4.	Equilibrando o orçamento	94
5.4.5.	Exemplo	96
5.5.	Avaliação do mecanismo de identificação de fraudadores	98
5.5.1.	Estimando parâmetros de funcionamento do mecanismo	98

5.5.2.	Medindo o resultado da operação do mecanismo	100
5.5.3.	Simulando a operação do mecanismo	100
6	Conclusões	103
6.1.	Trabalhos futuros	108
6.1.1.	Atuação fraudulenta em mercados eletrônicos	108
6.1.2.	Uso de agentes humanos para detectar fraudadores	109
6.1.3.	O mecanismo de identificação de fraudadores	110
6.1.4.	O jogo “pega ladrão”	110
6.1.5.	Outras possibilidades	110
Anexo 1.	Material utilizado no teste piloto com agentes humanos	114

## Lista de figuras

Figura 1: interações entre operador, jogo e jogadores	18
Figura 2: distribuição das diversas janelas de fraude	48
Figura 3: quantidade de fraudadores ativos a cada dia	49
Figura 4: percentual diário médio de fraudadores entre os vendedores	49
Figura 5: percentual médio de anúncios de fraudadores	50
Figura 6: percentual de vendas fraudulentas	51
Figura 7: seqüência temporal dos eventos	57
Figura 8: planilha para anotar respostas (com apelidos truncados)	60
Figura 9: ciclo de vida de uma identidade fraudulenta	68
Figura 10: ciclo de vida com suspensão prematura do vendedor	69
Figura 11: mecanismo de classificação baseado em computação humana	73
Figura 12: mudança no espaço de eventos	77
Figura 13: distribuições de probabilidade dos parâmetros dos classificadores	83
Figura 14: desempenho do mecanismo vs número de classificadores	84
Figura 15: valor recuperado médio vs número de classificadores	97

## Lista de tabelas

Tabela 1: taxas cobradas pelo MercadoLivre	42
Tabela 2: dados básicos sobre a atividade fraudulenta encontrada	47
Tabela 3: taxa de entrada de fraudadores	48
Tabela 4: dados sobre as perdas com fraudadores	51
Tabela 5: dados demográficos da amostra	61
Tabela 6: resultados da votação	62
Tabela 7: comparação das diversas classes de participantes	63
Tabela 8: desempenho dos participantes	64
Tabela 9: desempenho por categoria	65
Tabela 10: premissas acerca dos classificadores	71
Tabela 11: notação utilizada	71
Tabela 12: probabilidades do exemplo	74
Tabela 13: probabilidades com votação	78
Tabela 14: resultados do mecanismo	80
Tabela 15: parâmetros do exemplo	80
Tabela 16: resultados da aplicação do mecanismo	81
Tabela 17: resultados com o dobro de classificadores	81
Tabela 18: variáveis utilizadas para avaliar um classificador	90
Tabela 19: exemplos de parâmetros do mecanismo	93
Tabela 20: parâmetros adicionais do mec. de identificação de fraudadores	97
Tabela 21: resultados da simulação	101

*You can fool all the people some of the time  
and some of the people all the time, but you  
cannot fool all the people all the time.*

Atribuído a Abraham Lincoln (1809 - 1865)