

# Analysis of Fraudulent Activity in a Brazilian Auction Site

Vinicius da Silva Almendra

Departamento de Informática, PUC-Rio

Av. Marquês de São Vicente, 225

Rio de Janeiro – RJ

Brazil

55 21 3527-1500

vinicius.almendra@gmail.com

Daniel Schwabe

Departamento de Informática, PUC-Rio

Av. Marquês de São Vicente, 225

Rio de Janeiro – RJ

Brazil

55 21 3527-1500

dschwabe@inf.puc-rio.br

## ABSTRACT

Online auction fraud is the most common complaint according to Internet Crime and Complaint Center. Despite that, there are not many published empirical studies about fraud occurrence in online auction sites, and the existing ones mostly target eBay. This lack of research is even worse in Latin American countries. In this paper we present the results of an exploratory research on fraud occurrence in MercadoLivre, the biggest Brazilian auction site, which suggest that non-delivery fraud is a real, recurrent and measurable problem.

## Categories and Subject Descriptors

K.4.4 [Computing Milieus]: Computing and Society – *Electronic Commerce*.

J.4 [Computer Applications]: Social and Behavioral Sciences – *Economics*.

## General Terms

Measurement, Economics

## Keywords

E-commerce, Fraud

## 1. INTRODUCTION

In the last years we have witnessed a tremendous growth in electronic markets, of which online auction sites have an important share. Fraud levels have been increasing at a similar pace [3,7,9]. A key element to b2c and c2c negotiations in auction sites is the use of reputation systems [11], which are trust building mechanisms based on the dissemination of agents' prior performance, in order to incentive future cooperation. These mechanisms did a good job in enabling the initial growth of e-commerce; however, fraudsters reached these markets, quickly revealing the weaknesses of reputation systems.

In spite of that, there is still little published empirical research on the issue of fraudulent activities in online auction sites. The operators of these sites certainly have a great amount of empirical knowledge on this issue, but, for understandable reasons, this is not commonly publicly shared. This kind of research is further complicated by the absence of "ground truth": with few exceptions, there is no authority that confirms a given agent has indeed committed fraud.

Copyright is held by the International World Wide Web Conference Committee (IW3C2). Distribution of these papers is limited to classroom use, and personal use by others.

WWW'09, April 20–24, 2009, Madrid, Spain.

ACM 978-1-60558-487-4/09/04.

This lack of research is even worse in Latin America. We did not find any study about fraudulent activity in this region, notwithstanding the growing dissemination of electronic commerce. MercadoLibre, the biggest Latin American auction site, had a gross merchandise volume (GMV) of US\$ 1.5 billion in 2007 (excluding real estate, vehicles, ships, airplanes and services). It operates in 12 Latin American countries, including Brazil, where it is called MercadoLivre.

## 1.1 Objective

As part of an investigation on fraud prevention mechanisms [1], we needed to obtain several parameters about such markets. In order to fill this gap, we did an exploratory study of fraudulent activity in MercadoLivre, seeking answers to the following questions related to non-delivery fraud, the most common swindle in electronic markets:

1. Does MercadoLivre exhibit a perceivable level of non-delivery fraud?
2. Is there a set of signs that can be used to characterize systematically this fraud scheme in MercadoLivre?
3. How frequent are frauds?
4. How much profit a fraudster can make of one successful swindle?

As we said before, we are sure auction sites have answers to these questions (or at least have means to answer them). Nonetheless, as they do not share information with sufficient level of detail to foster scientific investigation, we have to obtain these answers using data that is publicly available.

## 1.2 Scope of Research

There are several kinds of fraudulent behavior, as we can see from Gregg and Scott's typology of buyer complaints [6]: non-delivery of products or services, misrepresentation, shill bidding, fee stacking, trouble to get refund, just to name the most frequent. There is also seller deception: for example, a buyer may send a forged email to a seller telling payment through an escrow service was made; seller sends the product and later discovers the trick.

A complaint from a buyer (or a seller) does not automatically mean that the other part defrauded or behaved with malice. There may be communications problems, misunderstandings, false expectations etc. In order to avoid dubious situations, we decided to restrict our research to frauds committed by sellers who list many products, get payments and do not deliver promised merchandise ("quick buck" method [3]). Given that (i) non-delivery fraud is the most usual one [4-6], (ii) it is easy to characterize (merchandise arrived or not) and (iii) we assume the feedback from buyers in this situation will accurately reflect what

happened, as it is improbable that a seller fails to deliver promised merchandise to many clients in a short period of time.

In section 2 we present related work on fraud in auction sites; in section 3 we describe MercadoLivre, the subject of our research. In section 4 we explain the adopted methodology. In section 5 we describe data collection procedures used. In section 6 we analyze the obtained data and draw conclusions about fraud in MercadoLivre. In section 7 we close our work, pointing limitations and future research.

## 2. RELATED WORK

There is a lot of research about reputation systems and how they induce cooperative behavior in strategic settings. Dellarocas [2] has done a good review on this topic. While providing incentive to good behavior, reputation systems discussed in literature may also help eliciting deceptive behavior, as the counterpart will try to publicize it through the feedback mechanism. In fact, some fraud-related studies rely on reputational information as evidence of fraud [5,9,10].

Gregg and Scott's works [5,6] focus on the analysis of reputational data available in auction sites: overall rating – positive, negative or neutral – and textual comments. They collected data from 40,976 eBay users in April, 2003 and used this data to test three hypotheses: that “the negative feedback reported in on-line auction reputation systems related to on-line auction fraud will exceed that reported through official channels”, “recent negative feedback will be a better predictor of fraud accusations than the overall feedback score” and that “buyers with less on-line auction experience are more likely to be victims of on-line auction fraud”.

To confirm these hypotheses, Gregg and Scott analyzed textual comments associated with negative feedback, in order to find signs of fraudulent behavior. Negative comments containing this kind of evidence were counted as fraud occurrences. The data obtained confirmed their hypotheses, and they estimated a fraud rate of 0.21%, an order of magnitude greater than the 0.01% reported by eBay [8].

An important difference among their work and ours is the assessment of a fraud incident. They considered each feedback comment with a complaint that signaled fraudulent behavior as a fraud occurrence. We took a different approach, as our unit of observation was the *seller*: if a user profile displayed enough signs of fraudulent activity, then all of his/her listings were considered fraudulent. Another difference is the scope: while they considered several types of fraud, we restricted our research to sellers committing non-delivery fraud and deceiving multiple buyers. We also only considered those sellers that were suspended by MercadoLivre. Due to these differences, we believe our approach offers a tighter lower bound to fraudulent activity.

Gavish and Tucci [3,4] also studied the fraud phenomenon in major auction site. They sent 1,298 questionnaires to users asking whether they had been defrauded, collecting 98 answers. They estimated that at least 0.62% of negotiations are fraudulent. As the methodology used was not clearly stated, we do not know if their results can be directly compared to ours.

Nikitkov and Stone [9] provide a model of auction fraud based on the literature of deception. While offering interesting information in order to guide the search from fraud patterns, they do not present empirical data of fraudulent activity.

## 3. DESCRIPTION OF MERCADOLIVRE

MercadoLivre<sup>1</sup> is the biggest Brazilian auction site, online since 1999. It has 32 million registered users in Latin America. It is affiliated with eBay and has similar functionality, although offering fewer options. It has two formats for listings: auction format (named “Arremate”) and fixed-price format (named “Compre já”). Each listing can be used to sell up to 999 items.

Sellers can improve listing appeal and relevance in search results through payment of extra fees. MercadoLivre has its own escrow service, called MercadoPago, which is tightly integrated into the negotiation workflow.

When a transaction takes place, MercadoLivre charges the seller with fees varying from 3% to 9% of the transaction value, depending on product type and listing characteristics. When the escrow service is used, buyer is charged with fees ranging from 3% to 20% of transaction value.

MercadoLivre has a reputation system: it requests sellers and buyers to give feedback of each other, which can be positive, negative or neutral. A textual comment may also be supplied. This information is displayed in the user's profile page. There is also an aggregated feedback score, which is the number of positive feedback from unique users minus the number of negative feedback from unique users.

Like eBay, MercadoLivre also suspends user accounts under certain conditions: non-payment of fees, infringement of MercadoLivre's policies, fraudulent behavior, and attempt to register more than once with similar personal data. A suspended account has its listing withdrawn, but its profile remains available online. Depending on the reason for suspension, MercadoLivre offers the possibility of account reinstatement if some conditions are met (usually, the user has to provide further documentation in order to confirm his identity).

An important difference between MercadoLivre and eBay is the preferred format of a listing: while in eBay the auction format is still prevalent, in MercadoLivre around 89% of transactions are done with fixed-price format. New items are also prevalent in MercadoLivre: almost 80% of transactions.

Another important difference is the requirements to open an account, specially a seller one: while in eBay every seller needs to verify its identity (normally giving a valid credit card number and providing matching personal data), and also needs to provide a secure payment method, in MercadoLivre a seller is requested only to inform a document number, which can be easily forged and it is difficult to check automatically; s/he must supply more information only if they wish to become a “powerseller”.

## 4. RESEARCH METHODOLOGY

### 4.1 Characterizing Fraudulent Behavior

In order to do our exploratory research, we needed examples of fraudulent sales. We visited sites dedicated to fraud discussion and found a great number of complaints related to non-delivery fraud in MercadoLivre. So, we considered our first research question answered.

In these sites, besides descriptions of swindles, we found some complaints with enough information to locate profiles of alleged fraudsters in MercadoLivre. From these profiles we gathered hints

---

<sup>1</sup> [www.mercadolivre.com.br](http://www.mercadolivre.com.br)

to find more suspicious sellers. We then selected sellers with profiles and listings according to those hints and visited their profiles daily, observing which of them were suspended and received feedback that pointed to fraudulent behavior. We found a clear recurring pattern of what seemed to be a fraudulent behavior:

1. Fraudster obtains an account with some reputation. S/he can hijack another seller’s account or may build some reputation buying goods. However, the scheme that seemed more popular was the one of opening some fresh new user accounts and using them to “buy” products from the fraudster “main” user account. As no documents have to be forged, this is an inexpensive method to build reputation. Nevertheless, some fraudster profiles had no reputation at all, showing that many buyers have a very limited knowledge of reputation systems.
2. Fraudster lists popular products (mostly cellular phones and digital cameras) with attractive prices, big quantities (hundreds of items sometimes), and with payment through wire transfer.
3. Within a short period (few days or sometimes even hours) many users close the deal with the fraudster.
4. Some users might complain very fast, while the fraudster account is still active. When this happens, the fraudster usually replies accusing the buyer of something or employs a variety of stalling tactics to delay negative feedback, in order to get additional time to sell products before eventually being suspended.
5. After some time, MercadoLivre suspends the fraudster account, probably because s/he was trapped in some security check or because someone complained of fraud. Many negative feedback starts to appear in the fraudster profile, some of which accusing of non-delivery fraud. But it is too late: those who paid lost their money.
6. After some time, no new negative feedback arrives anymore and account continues suspended.

This pattern is similar to the “quick buck” method reported by Gavish and Tucci [3]. We did not have any “official” confirmation of fraud in these cases. Nonetheless, we assumed that the above sequence of facts together with the following set of signs was enough to characterize occurrence of fraud, as their joint presence is unlikely when the seller is not a fraudster:

- Seller account is not reinstated: this rules out most attacks against honest sellers, as they will try to recover their account.
- There are several successive transactions with negative feedback as a seller *from different buyers* and this feedback is not withdrawn after some time: this also rules out badmouthing by other sellers, as MercadoLivre has a mechanism to remove undeserved negative feedback and we have evidence from Internet forums that it really works.
- Textual feedback comments mention that payment was made but nothing was received yet and that seller stopped answering emails or phone. This rules out most situations where seller had some problems of delivery.

There are other signs that something wrong has occurred: prior reputation was acquired buying products or through transactions with low reputation buyers, product location is highly unusual (this “justifies” a longer delivery time and avoids attempts to retrieve the product in person).

We consider answered our second research question: we have signs that systematically appear in seller profiles exhibiting evidences of non-delivery fraud.

This first exploratory research also called attention to an interesting fact: sellers that met the above conditions, i.e. probably were fraudsters, sold their products very fast. Looking at product descriptions, we concluded that the reason for this was a combination of low price and attractive products: fraudsters would probably find out what were the best-selling products and list them with prices lower than other sellers’.

Notice that our criteria did not include several other kinds of fraudulent behavior, as we are focusing on non-delivery fraud against many buyers. This means our results will typically be a lower bound to actual fraud levels.

From now on when we use the word fraudster, we mean a seller whose behavior and profile fits in the above descriptions.

## 4.2 Designing Data Collection

In order to meet our research objectives and given its exploratory nature, we opted for a one-month longitudinal data collection of MercadoLivre’s public data, tracking evolution over time of seller profiles and listings. We restricted the set of products we would monitor in order to keep data collection manageable, while trying to maintain sample representativeness. First, we only monitored listings in the fixed-price format, as it is the predominant one (almost 89% of transactions) and the auction format offers no advantage to a fraudster trying to swindle many buyers at the same time.

Then we focused on best-selling product categories, since from our first exploratory research we noticed fraudsters tended to list popular items, attracting attention of many buyers and maximizing sales in the short time window they had. Chosen categories are shown in Table 1; they comprise around 78% of products listed in fixed-price format.

**Table 1: product categories monitored**

Sports and fitness	Musical instruments
Computer equipment	Agro industry
Cameras and photo	Toys and hobbies
Electronic appliances	Jewelry and watches
Games	Décor and furniture
Clothing and apparel	Car accessories
Animals	Health and beauty
Baby products	

An important absence in this list is the category of cellular phones. Due to a mistake it was omitted in the data collection. As cellular phones are top-selling products and we have plenty of evidence in Internet forums that fraudsters indeed have an important presence in this segment, our results certainly show fraud levels below real ones.

The third restriction was to monitor products with price equal or above R\$ 100 (around US\$ 60 at the time data collection was done). This reduced significantly the number of listings to be watched. In the categories of computer equipment and cameras,

almost 80% of listings had prices above this limit. Again we are constrained to find a fair lower bound on fraud occurrence.

We did data collection on a daily basis. Each day all listings in each of the chosen categories were examined in order to see if they fit the other criteria (price and format). If so, we saved listing data (title, price, quantity and category) and saved seller data (nickname and registration date). We also took a daily snapshot of each saved listing, containing the total number of units already sold. Finally, we took daily snapshots of seller profiles, containing the following data: whether or not the account was in suspended state, the total score of positive and negative feedback, amount of recent positive, negative and neutral feedback (in the last week, in the last month and in the last six months), number of recently sold items (in the last week, in the last month and in the last six months). All saved information was annotated with the date it was collected.

In order to find fraudulent behavior, we adopted a semiautomatic approach. In the end of data collection, we selected all sellers that had the following characteristics:

- Had received in a single week more than five negative feedback points.
- Had been suspended for three days or more.

Then we manually inspected the profile pages of these sellers, in order to see whether the negative points given by buyers and the textual comments displayed the characteristics we expected (were from different buyers, asserted that some people paid and did not receive what was promised). Those who displayed these signs were classified as fraudsters.

## 5. DATA COLLECTION

We developed a crawler in Java in order to extract information from MercadoLivre’s website. We did not use an account for this: we crawled only public pages. We checked the terms of use and there was nothing that prevented us to do this data collection.

In order to reduce bandwidth, we only downloaded HTML pages, ignoring other resources attached to them. Information was extracted using regular expressions and stored in an Oracle Express Database.

The data collection lasted from July 18<sup>th</sup> to August 20<sup>th</sup>, 2008. We monitored 199,305 unique product listings and 28,690 unique seller profiles. In Figure 1, we show how many new product listings and seller profiles were found each day of the data collection. We omitted the first three days, as they are outliers: the huge amount of data prevented us to crawl all sellers and active product listings in a single day.

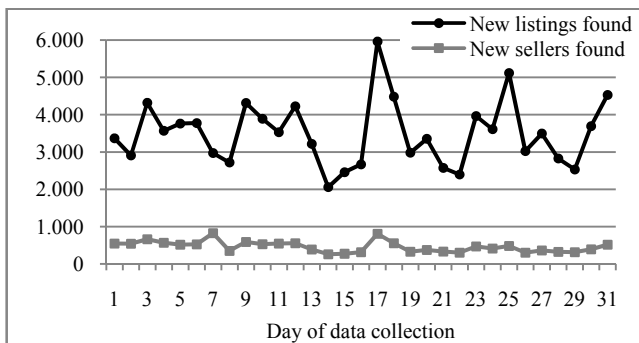


Figure 1: amount of data crawled per day

In Table 2 we show the number of saved listings per product category. Notice that products of categories other than the ones we crawled were also present. We conjecture that MercadoLivre put some product listings in more than one category; another possibility is that categories are not disjoint and share some sub-categories.

One limitation of our data collection scheme was that we only updated the number of sold units of *active listings*, that is, those which had not expired yet. This means that units sold in the last day of listing permanence might be missed, as the listing might end before the next daily update.

Table 2: number of saved listings per product category

Category	Number of saved listings	% of total number of saved listings
Computer equipment	40,679	20,4%
Clothing and apparel	23,141	11,6%
Electronic appliances	18,219	9,1%
Toys and hobbies	17,780	8,9%
Sports and Fitness	14,369	7,2%
Jewelry and watches	13,800	6,9%
Musical Instruments	13,776	6,9%
Car accessories	11,029	5,5%
Camera and photo	9,510	4,8%
Health and beauty	8,960	4,5%
Games	7,865	3,9%
Décor and furniture	7,646	3,8%
Cellular phones and telephony	5,372	2,7%
Agro industry	2,435	1,2%
Animals	1,679	0,8%
Baby products	1,668	0,8%
Other	1,377	0,7%

In Table 4 we see a summary of the fraudulent activity spotted using our methodology. Money values were converted to US dollars using the exchange rate at the time of data collection. We estimated the amount of money defrauded inspecting feedback comments, in order to count how many explicitly said payment was made. Then we looked the price of the item in question, which was the profit the fraudster obtained with that single buyer. Of the 33 fraudsters found, 6 were responsible for 63% of the money loss, indicating that few seller accounts were enough to cause a high impact on buyers.

Table 3: summary of fraudulent activity

Number of fraudsters found	33
Total listings	199,305
Listings belonging to fraudsters	139 (0.07%)
Total of individual items sold	549,605
Items sold belonging to fraudulent listings	1,244 (0.23%)
Buyers that allegedly paid for the products	321
Money allegedly paid for these products	US\$ 54,081
Average loss per buyer that allegedly paid	US\$ 169
Average fraudster profit per day (while active)	US\$ 536

An interesting metric is the *fraud window*: the time between the first listing of a fraudster and the moment his account is suspended. To maximize profit, the fraudster has to “convince” many people to buy his products in this interlude.



spread a negative opinion about the auction site in question, which may eventually end up known as a dangerous place.

Numbers obtained are comparable to other ones found in the literature [4,5,7], although a direct comparison cannot be reliably done, due to methodological differences and to the obvious fact that they refer to different countries, with different cultures, different legal systems etc.

The use of systematic crawling and of a longitudinal approach gave those numbers a more solid empirical grounding; properly speaking, we did not sample the population: we tracked daily the entire subset of sellers where we expected fraudsters to concentrate in.

Even though, there are some weaknesses in data collection: the last sold units of products were missed. We estimated a reduction of 12% on the total number of transactions. This reduction affected both normal sellers and fraudsters, so we ignored it, as we were mainly interested in rates.

Another limitation is the short time period. We do not know if the results can be generalized, although Figure 2 shows a regular phenomenon that does not resemble to be a “fraud spike”. We speculate the situation observed reflects reality, given information about frauds in MercadoLivre we obtained in public Internet forums. Fraud window is also affected by this short period, as some fraudsters probably were already in operation before data collection started. In any case, this would just enlarge some fraud windows, aggravating the problem.

The price criteria used to select which products to track left too many items outside data collection. As we targeted listings with a higher probability of being fraudulent, the overall fraud level could be smaller than the one we found. In fact, we tracked near 20% of all available listings. If the other 80% had very small fraud levels, the numbers above would be diluted.

Given the absence of formal confirmation, we cannot say that all sellers we considered to be fraudsters have really committed fraud. Nonetheless, we regard the results as very convincing, given the methodology used and the textual comments examined.

Textual analysis gave some insights about agents involved with fraud (auction site, buyer and fraudster). However, these results must be treated as a starting point for a more systematic study, as they were a by-product of fraud detection and did not obey a well-defined research objective.

## 7. CONCLUSIONS

We investigated the fraud problem on MercadoLivre, the biggest Brazilian online auction site. We restricted ourselves to non-delivery fraud with multiple buyers and done an extensive data collection during a month. The predicted pattern of fraudulent behavior was really found and we found several sellers exhibiting it. Based on this pattern, we calculated fraud rates, buyer losses and fraudster profit, answering our research questions for the population of sellers analyzed. We also elicited the concept of fraud window – the time a fraudster remains active in the market – as an important factor to check security measures effectiveness and fraudster successfulness. We also provided descriptions of typical statements appearing in fraud complaints. As far as we know this is the first work that explored this kind of fraud in Brazil and that used longitudinal data collection to assess fraudulent behavior.

The restrictions adopted to keep data collection manageable reduced the possibility of generalization. The fraud numbers presented can only be reliably used as lower bounds of fraudulent activity in the examined subset of sellers. Further studies are needed, in order to generalize those numbers.

Future work includes a replication of this study for a longer time period and for a more comprehensive set of sellers. In order to make this possible, textual patterns found could be used to automate fraud detection. This research could also be replicated in other Latin American countries where MercadoLivre operates.

## 8. ACKNOWLEDGMENTS

This research was sponsored by UOL ([www.uol.com.br](http://www.uol.com.br)), through its UOL Bolsa Pesquisa program, process number 200606012154 00a, and by CNPq, process number 140768/2004-1

## 9. REFERENCES

- [1] V. Almendra and D. Schwabe. 2009. Improving Precision of Fraudster Detection through Human Computation (submitted to EC'09).
- [2] C. Dellarocas. 2006. Reputation Mechanisms. In *Handbook on Economics and Information Systems*. Elsevier Publishing, pp. 629-660.
- [3] B. Gavish and C. Tucci. 2006. Fraudulent auctions on the Internet. *Electronic Commerce Research*. Vol. 6. April 17. pp. 127-140. DOI=10.1007/s10660-006-6954-0.
- [4] B. Gavish and C. Tucci. 2008. Reducing Internet Auction Fraud. *Communications of the ACM*. Vol. 51 (5).
- [5] D.G. Gregg and J.E. Scott. 2006. The Role of Reputation Systems in Reducing On-Line Auction Fraud. *International Journal of Electronic Commerce*. Vol. 10. pp. 95-120.
- [6] D.G. Gregg and J.E. Scott. 2008. A typology of complaints about ebay sellers. *Communications of the ACM*. Vol. 51 (4).
- [1] Internet Crime Complaint Center. 2007 Internet Crime Report. [http://www.ic3.gov/media/annualreport/2007\\_IC3Report.pdf](http://www.ic3.gov/media/annualreport/2007_IC3Report.pdf). Accessed June 10, 2008.
- [8] F. Manjoo. 2001. EBay Fraud Law: Any Takers? *Wired.com*. June 27. <http://www.wired.com/politics/law/news/2001/06/44831>. Accessed February 7, 2009.
- [9] A. Nikitkov and D. Stone. 2006. Towards a Model of Online Auction Deception. In *Proceedings of the ECAIS 2006* (Dublin, Ireland).
- [10] S. Pandit, D.H. Chau, S. Wang, and C. Faloutsos. 2007. NetProbe: A Fast and Scalable System for Fraud Detection in Online Auction Networks. In *Proceedings of the 16th international conference on World Wide Web* (Banff, Alberta, Canada). ACM.
- [11] P. Resnick, R. Zeckhauser, E. Friedman, and K. Kuwabara. 2000. Reputation Systems. *Communications of the ACM*. Vol. 43. pp. 45-48.