The Use of Contextual Information to Detection of Fraud on On-line Auctions

LADISLAV BERANEK, PhD
MBA, Associate Professor, University of South Bohemia, Faculty of Economics, Ceske Budejovice, Czech republic
Postal Address: **Studentska 13, 37005 Ceske Budejovice, Czech Republic**
Email: **beranek@ef.jcu.cz**
Prof. Ladislav Beranek is a member of Department of Applied Mathematics and Informatics at the Faculty of Economics, University of South Bohemia, Czech Republic. His research interests are e-commerce, web technologies, data mining and social network analysis.

JIRI KNIZEK, PhD
Research Associate, Charles University in Prague, Faculty of Medicine in Hradec Kralove, Department of Medical Biophysics, Hradec Kralove, Czech Republic
Postal Address: **Simkova 870/13, 50038, Hradec Kralove, Czech Republic**
Email: **knizekj@lfhk.cuni.cz**
Dr. Jiri Knizek is a member of Department of Medical Biophysics, Faculty of Medicine in Hradec Kralove, Charles University in Prague. His main research interests are statistic methods and their application especially in medicine and biology.

Abstract

Currently, Internet auction portals are an integral part of business activities on the Internet. Anyone can easily participate in online auctions, either as a seller or a buyer (bidder), and the total turnover on Internet auction portals represents billions of dollars. However, the amount of fraud in these Internet auctions is related to their popularity. To prevent discovery, fraudsters exhibit normal trading behaviors and disguise themselves as honest members. It is therefore not easy to detect fraud in online auctions. There are some papers and approaches dealing with this problem with varying results. Most of

them concentrate on the selection of the attributes available within online auction portals and on computational methods for their processing. This study proposes extended the fraud detection approach by using certain contextual information whose origin is outside online auctions portals. The suggested model integrates information from auctions and relevant contextual information with the aim to evaluate the behavior of certain sellers in an online auction and determine whether it is legal or not. Experimental results show that this approach based on the use of contextual information from other Internet sources provides good results and enhances significantly the accuracy of detection of certain types of fraud in online auctions.

Keywords: **online auctions; Internet fraud; information integration; reasoning under uncertainty**

© Ladislav Beranek and Jiri Knizek, 2013

INTRODUCTION
Currently, many users participate in online auctions operated by various Internet online auction portals. Besides eBay (eBay, 2013), the Aukro auction portal (Aukro, 2013), a Czech auction company with a turnover of USD 250 million and 2.5 million users as of January 2012, is an example of such an Internet auction system. Online auctions allow their users to buy or sell products and services. A large number of users use Internet auctions as their main source of income. For example the Aukro states (Aukro, 2013) that from 2.5 million users, 9,300 users are professional traders. On the other hand, the amount of a fraud in online auctions is also increasing (Gavish and Tucci, 2006).

The most common types of fraud are incorrect (purposely) description of goods, undelivered goods, irredeemable payments, sale of stolen goods, and others (Gregg and Scott 2008). Fraudsters are attracted by low admission costs and high profit potential. A typical procedure for the detection of fraudulent behavior on an online auction consists of the two basic steps: (1) a set of attributes is proposed and their values are extracted from a transaction history to distinguish between normal traders and scammers, and (2) a detection model based on these attributes is built using various machine learning techniques.

This study aims to suggest an improvement of detection methods on the basis of utilization of additional sources of information (contextual information). For this purpose, various sources of information on the Internet, which could serve as sources of contextual information and hence could be used to improve the detection of fraud, have been explored. More than 195 different Internet forums and discussions and similar Internet sources were surveyed. Specific pieces of information relating to some Internet auctions were extracted from these information sources and examined in detail. Together with these information sources, we investigated 424 online auctions on the Aukro auction portal (Aukro, 2013). Our proposed fraud detection model was tested with the help of this information. The results of our experiments show that the proposed approach improves the accuracy of fraud detection in online auctions.

The remainder of this paper is organized as follows: Section 2 summarizes some concepts and techniques related to the detection of fraud in online auctions including works using the belief function approach. Section 3 presents the basic principles of the belief theory, including a description of a situation when there might be some doubts about the reliability of information sources. Section 4 presents our approach. It outlines the key definitions of belief functions which represent the degree of potential fraudulent behavior. This degree is calculated on the basis of information about the respective online auction and on the basis of contextual information found using other Internet information sources. Contextual discounting is used here to express the influence of contextual information on the degree of potential fraudulent behavior. Section 5 presents the results of the experiments, followed by the conclusion and suggestions for future work in Section 6.

BACKGROUND AND RELATED WORK

Successful Internet auction portals must perform many different activities. In addition to the creation, implementation and operation of their trading system (including users login, displaying data about items being sold, including initial price and duration of auction, displaying of bids of buyers and other data) they must also solve problems with trust and trustfulness. This follows from the fact that transactions in the online auction predominantly take place in a situation when the users do not know each other. However, if users want to do business here they must decide whether they will trust each other. Online auction portals, such as eBay (eBay, 2013), Aukro (Aukro, 2013) and others are successful primarily because they are able to create a trusted environment for users of online auctions (Hoffman et al., 1999; Bryant and Colledge, 2002; Hsu and Wang, 2008). Most of the mechanisms (reputation systems), which create a trusted environment, use a variety of attributes associated with users and their roles.

Fraudulent Behavior

Although reputation systems and certain methods of user identification are functional, a lot of fraud takes place in online auctions. The sellers and bidders are not in physical contact, and bidders are not able even physically see the auctioned item. This situation provides many opportunities for cheating (Chae et al, 2010; Chua and Warenham, 2008; Sukurai and Yokoo, 2003).

From the perspective of a seller, online auctions bring about certain risks (Trevathan and Read, 2008, 2009; Gregg and J. Scott, 2008; Dong et al, 2009), particularly:
− Bidder does not pay for the goods supplied.
− Bidder wrongly claims that the goods have not been delivered.

From the perspective of the bidder (Trevathan, 2005):
− The seller refuses to send the goods (Gregg and Scott, 2008).
− The description of the auctioned object is false.
− Seller sends different goods or goods of lower quality (Da Silva Almendra and Schwabe, 2009).
− Goods intended for auction are fake or stolen.
− Seller can influence the course of the auction (seller can collude with other bidders or can bid on his own items to drive up the price of the item being auctioned) (Trevathan and Read, 2007). We denote this as shill behavior.

Fraudulent behavior is relatively widespread in online auctions. The main reason is that engaging in fraudulent behavior is relatively easy because:

− Online auction participants are largely anonymous - they frequently act under pseudonyms. Internet auction systems use different methods for verifying the identity of users. These methods, however, may not be sufficiently reliable (Beranek, 2010a).

− Online auction houses do not always exhibit full commitment to actively engage in combating fraud.

− Law enforcement is often difficult with regard differences of legal systems in different countries.

Related Work
Fraudulent behavior which occurs in online auctions is not easy to detect mainly due to the use of various techniques by fraudsters to camouflage their behavior and due to the pseudonymity of users participating in the auction. The most widely used detection approaches are based on various statistical methods, data mining techniques, methods of analysis of online users behavior or social networks analysis, see for example (Hu and Panda, 2005; Gregg and Scott, 2006, 2008; Chang and Chang, 2012; Ku et al., 2007; Pandit et al., 2007; Zhang et al. 2008; Ku et al.,2007; Zhang et al., 2008) and others. Works that deal with the detection of fraudulent behavior in online auctions focus mostly on certain types of fraudulent behavior, for example shill behavior (Dong et al., 2010; (Trevathan et al., 2009). A general approach would be very complicated due to the complexity of the issue.

The selection of an appropriate set of attributes is crucial for constructing a detection model. The simplest way to devise an attribute set for fraud detection is to enumerate all the features of tricks that have already occurred. Attributes of fraudulent behaviors are taken directly from statistics related to past transactions. These attributes include the count of positive ratings and negative ratings, the median, the standard deviation, the average of all labeled prices during a specific time period (Chau and Faloutsos, 2005), the starting labeled price of a bid and some Boolean variables (Wang and Chiu, 2005). Trevathan and Read (2007) deal with fraudulent behavior in online auctions. They propose an algorithm to detect shill behavior based on comparisons of patterns of behavior in online auctions. Trevathan and Read present in another paper (Trevathan and Read 2009) a method for detecting colluding shill users. Chau et al. (2006) use methods based on a data mining approach to detect shill behavior. They apply this approach on the user level and on the level of interaction among users. They link these two levels to detect suspicious behavior patterns using Markov random field methods. Xu and Cheng (2007) introduce a dynamic auction model for shill detection in real online auctions and use formally specified shilling behavior by the help of linear temporal logic to verify the shill behavior. The other authors (Goel et al, 2010, Ford et al, 2010,) propose using Bayesian networks or decision trees to detect fraudulent behavior in online Internet auctions.

The use of belief functions to detect shill behavior is presented in the work of Dong et al. (2009, 2010, 2012). The authors indicate a conceptual design framework for calculating the belief functions. They demonstrate the correctness of their approach in an eBay auctions case study. The paper (Beranek et al., 2010b, 2012) describes some features of shill behavior which are then expressed by belief functions and combined with the aim to classify users into categories of shill, suspect and trustworthy. Pandit et al. (2007)

designed and implemented an online auction fraud detection system named NetProbe. The NetProbe system models auction data as a network graph in which sellers and bidders are represented by nodes, and transactions between sellers and bidders are represented by edges. The Markov random field and the belief propagation algorithms are utilized to unearth suspicious trading patterns created by fraudsters and thus to detect possible fraudsters. An online auction fraud detection systems was also presented in the works (Chau and Faloutsos, 2005, Chau et al., 2006a,b, Pandit et al. 2007, Zhang et al., 2008, Chang and Chang, 2011). Chang et al. (2012) propose the data mining methods for early detection of fraudulent behavior. Forty-four attributes are defined and analyzed in this paper with the aim to build a model for early detection of fraud. Kwan et al. (2010) focus on the detection of selling fake products. They define attributes of this fraudulent behavior and use the Bayes approach in their evaluation.

Generally, the detection accuracy is closely related to the suitability of the attributes and the choice of modeling method. It is obvious that previous work in this area has provided good level of progress but some problems still remain. The use of a greater number of measured attributes may not bring substantial improvement. The accuracy of detection can even deteriorate when irrelevant attributes are incorporated into the model. A major improvement in fraud detection based on data from auction portals cannot be expected at the present. We therefore suggest an improvement of fraud detection on the basis of the utilization of additional sources of information available on the Internet.

THE PROPOSAL OF OUR MODEL

This paper aims to detect specific Internet auction fraud related to the selling of stolen goods (i.e., goods being stolen and subsequently sold in an online auction). The detection model is based on chosen attributes of this fraud and it also uses contextual information, i.e., information found on various public Internet forums and discussions to improve the prediction accuracy of the detection model. The model is based on the belief function theory. The advantage of the use of this theory is the possibility to represent partial knowledge and the possibility to combine pieces of evidence concerning possible fraudulent behavior with pieces of contextual information from the Internet sources.

Basic Principles of the Belief Function Theory

Our model is a particular application of the belief function theory. The belief function theory (Shafer, 1976) is designed to deal with the uncertainty and incompleteness of available information. It is a powerful tool for combining evidence and changing prior knowledge in the presence of new evidence. The belief function theory can be considered as a generalization of the Bayesian theory of subjective probability. In the following paragraphs, we give a brief introduction of the basic notions of the belief function theory (frequently called Dempster-Shafer theory or theory of evidence).
Considering a finite set referred to as *the frame of discernment* Ω a *basic belief assignment* (*BBA*) is a function *m*: 2^Ω [0,1] so that

$$\sum_{A \subseteq \Omega} m(A) = 1, \tag{1}$$

where $m(\varnothing) = 0$, see (Shafer, 1976). The subsets of 2^Ω which are associated with non-zero values of *m* are known as focal elements and the union of the focal elements is called the core. The value of *m(A)* expresses the proportion of all relevant and available evidence that supports the claim that a particular element of belongs to the set A but not

to a particular subset of *A*. This value pertains only to the set A and makes no additional claims about any subsets of *A*. We denote this value also as a *degree of belief* (or *basic belief mass - BBM*).

Shafer further defined the concepts of *belief* and *plausibility* (Shafer, 1976) as two measures over the subsets of as follows:

$$Bel(A) = \sum_{B \subseteq A} m(B), \tag{2}$$

$$Pl(A) = \sum_{B \cap A = \varnothing} m(B). \tag{3}$$

A *BBA* can also be viewed as determining a set of probability distributions *P* over so that $Bel(A) \le P(A) \le Pl(A)$. It can be easily seen that these two measures are related to each other as $Pl(A) = 1 - Bel(A)$. Moreover both of them are equivalent to *m*. Thus one needs to know only one of the three functions *m*, *Bel*, or *Pl* to derive the other two. Hence we can speak about belief function using corresponding *BBA*s in fact.

Dempster's rule of combination can be used for pooling evidence represented by two belief functions Bel_1 and Bel_2 over the same frame of discernment coming from independent sources of information. The Dempster's rule of combination for combining two belief functions Bel_1 and Bel_2 defined by (equivalent to) *BBA*s m_1 and m_2 is defined as follows (the symbol $\oplus$ is used to denote this operation):

$$(m_1 \oplus m_2)(A) = \frac{1}{1-k} \sum_{B \cap C = A} m_1(B) \cdot m_2(C), \tag{4}$$

$$k = \sum_{B \cap C = \varnothing} m_1(B) \cdot m_2(C). \tag{5}$$

Here *k* is frequently considered to be *a conflict measure* between two belief functions m_1 and m_2 or a measure of conflict between m_1 and m_2 (Shafer, 1976). Demspter's rule is not defined when *k* = 1, i.e. when cores of m_1 and m_2 are disjointed. This rule is commutative and associative; as the rule serves for the cumulation of beliefs, it is not idempotent.

Belief Function Correction
When receiving a piece of information represented by a belief function, some metaknowledge regarding the quality or reliability of the source that provides some information can be available. In the following paragraphs, we describe briefly some possibilities how to adjust the information according to this metaknowledge.

Discounting. To handle the lower reliability of information sources, a discounting scheme has been introduced by Shafer (1976). It is expressed by equations:

$$^{\alpha}m(A) = \begin{cases} (1-\alpha)m(A) & \text{if } A \subset \Omega \\ \alpha + (1-\alpha)m(\Omega) & \text{if } A = \Omega \end{cases}, \tag{6}$$

where $\alpha \in [0,1]$ is a *discounting factor* and $^{\alpha}m(A)$ denotes the discounted mass of *m(A)*.

The larger α is, the more mass *m(A)* is withdrawn from $A \subset \Omega$ and assigned to the frame of discernment Ω.

Thus, the principle of discounting is *transferring parts of basic belief masses BBMs of all focal elements which are proper subsets of the frame of discernment to the entire frame*.

This process is the result of additional information which indicates that the source is not entirely reliable. The transfer of *BBM*s from a source to the framework reflects an increase of the degree of uncertainty of the data that the source produces.

De-discounting (reinforcement). In some cases we need to perform an opposite operation, e.g., transfer parts of basic belief mass (*BBM*) from the entire frame to all focal elements. This can be the result of a situation when we, for example, obtain information that the source of the information is more reliable than we had anticipated at the beginning. We can then re-compute *m* by reversing the discounting operation (Smets, 1993; Mercier et al, 2012). We denote this operation as *reinforcement* (or de-*discounting*):

$$m(A) = \frac{{}^{\alpha}m(A)}{1-\alpha} \qquad \forall A \subset \Omega (7), \tag{7}$$

$$m(\Omega) = \frac{{}^{\alpha}m(\Omega) - \alpha}{1-\alpha}, \tag{8}$$

where $\alpha \in [0, {}^{\alpha}m(\Omega)]$. We denote here α as a reinforcement coefficient. The result of maximal de-discounting is the totally reinforced belief function. It is noted ${}^{tr}m$ and defined as follows:

$$ {}^{tr}m(A) = \begin{cases} \dfrac{m(A)}{1 - {}^{\alpha}m(\Omega)} & \forall A \subset \Omega \\ 0 & \text{otherwise} \end{cases}, \tag{9}$$

The scenarios associated with this idea can often be found in the multi-evidence pooling systems, where decisions are made based on a set of existing pieces of evidence and the corresponding confidence in (or evaluation of) these pieces of evidence. Evidence and corresponding confidence may be elicited in different manners, e.g., drawn by different experts, or based on different viewpoints.

PROPOSED FRAUD DETECTION METHOD BASED ON CONTEXTUAL INFORMATION

We have chosen the Dempster-Shafer theory for the mathematical representation of the fraudulent behavior - the selling of stolen goods. The theory makes the expression of uncertainty in our model possible. Our uncertainty in the sense that we are not able to say assuredly that certain behavior on certain auction is or is not fraudulent is significant here and we need to express this ignorance, this partial knowledge. Therefore, the Dempster-Shafer theory is particularly suitable for the modeling of evaluation of fraudulent behavior.

Basics of Our Model

Our model consists of four steps:

Step 1. *Definition of belief functions representing analyzed fraudulent behavior*. In the first step, determinative attributes of fraudulent behavior, the selling of stolen goods, are specified: an inadequately low price m_L, goods sold mostly at fixed price m_F and variety of goods being sold m_V.

Step 2. *Definition of the reinforcement coefficient on the basis of contextual information*. In this step, certain characteristics of contextual information are used to define the reinforcement coefficient α described in Section 3.2.

Step 3. *Assessment of the influence of contextual information on the evidence about fraudulent behavior*. In this third step, contextual information is used to eventually increase the value of the belief function expressing our conviction that stolen goods are sold in analyzed online auction. The aim of this step is to assess the effect of "additional" contextual information about stolen goods.

Step 4. *Categorization of sellers*. In the last step, sellers are categorized according to the resulting belief functions representing the selling of stolen goods into three categories: *proper seller*, *suspect seller* and *fraudulent seller*.

Definition of Belief Functions Representing Illegal Behavior
We explored in detail data about auctions related to five prosecuted cases out of 54 complaints of stolen goods being sold in Internet auctions. Our aim was to determine the characteristics of online auction fraud related to the sale of stolen goods. The cases were the result of complaints lodged with the Czech Trade Inspection (Czech e-shops inspections, 2012). Because detailed judgments were not available for the prosecuted cases, examiners who worked on the cases were interviewed to elicit some additional information.

The following attributes of the online auction offering stolen goods were specified:
1. Stolen goods were sold at inadequately low prices (at least about 20% below the price of legitimate goods).
2. Fraudsters prefer to sell stolen goods for a fixed price.
3. A variety of goods were sold via fraudulent accounts (such as car accessories, footwear, sporting goods etc.).
4. Life span of such fraudulent accounts was very short (often less than twelve days).
5. In most cases, the goods were sold within several days of creating the account. In most cases (31 out of 54), the goods were sold within six days of account creation.
6. Fraudsters had accounts on multiple auction systems, and the value of their reputational score is low.

Our aim was to define belief functions corresponding to the chosen attributes and then to combine these belief functions to assess whether the respective bidder sells stolen goods or not. Based on our analysis, we have chosen the following attributes as indicators that stolen goods are being sold (the other ones were too difficult to verify or express mathematically):
1. Goods sold at inadequately low prices;
2. Goods sold mostly at fixed prices;
3. A variety of goods being sold.

We denoted the frame of discernment concerning the analyzed fraudulent behavior Ω = {*stolen*, ¬*stolen*}. Here *stolen* represents the hypothesis that stolen goods are sold in the analyzed online auction, ¬*stolen* represents the hypothesis that the analyzed online auction is conducted properly. The power set of the set Ω (the set of all subsets) 2^{Ω} has three elements (we do not consider the empty set here): 2^{Ω} = {{*stolen*}, {¬*stolen*}, {*stolen*, ¬*stolen*}}, where {*stolen*, ¬*stolen*} = Ω denotes our ignorance. That means that we are not able to assess whether stolen goods are sold in the online auction or not. The belief functions expressing our belief concerning single evidences of this fraudulent behavior are described in the next paragraphs.

Inadequate low price. This attribute shows that the seller i sells stolen goods for lower prices than the average price of the legitimate goods. The belief functions have the following form:

$$m_L(\{stolen_i\}) = \begin{cases} v_{LA} \dfrac{\overline{P} - P_i}{\overline{P}} & \text{for } P_i \le \overline{P} \\ 0 & \text{for } P_i > \overline{P} \end{cases}$$

$$m_L(\{\neg stolen_i\}) = 0 \qquad\qquad (10)$$

$$m_L(\Theta_i) = \begin{cases} 1 - v_{LA} \dfrac{\overline{P} - P_i}{\overline{P}} & \text{for } P_i \le \overline{P} \\ 1 & \text{for } P_i > \overline{P} \end{cases}$$

where v_L is the weight of this evidence. We can intuitively read this weight as a reliability of this evidence; P_i - is the price at which the seller i sells certain goods. $\overline{P}$ is the average price of the same goods offered through online auction system.

With this equation, we have expressed our belief that the lower the price of goods offered by seller i compared to the average price of respective goods, the higher the suspicion that the seller offers stolen goods. We also assume that the equation reflecting the offering of legitimate goods does not show that the seller does not offer "stolen" goods, i.e $m_L(\{\neg stolen_i\}) = 0$.

Goods sold mostly at fixed price. The sellers (fraudsters) want to sell their stolen goods as quickly as possible. They want to dispose of them quickly and easily. Therefore they prefer to sell the goods at a fixed price (Internet auction systems have the option "*buy now*"). It is the fastest way to sell goods on an online auction. When a bidder purchases goods at a fixed price, the auction ends, and the seller does not have to wait for the end of the auction. The belief functions have the following forms:

$$m_F(\{stolen_i\}) = v_F \frac{N_{Fi}}{N_i},$$

$$m_F(\{\neg stolen_i\}) = 0, \qquad\qquad (11)$$

$$m_F(\Theta_i) = 1 - v_F \frac{N_{Fi}}{N_i},$$

where N_{Fi} is the number of goods sold by seller i for the fixed price; N_i is the total number of goods sold by this seller. It is valid that the higher the number of goods sold at fixed price, compared to the total number of goods sold by this seller i, the higher the suspicion that this seller sells "stolen" goods. Therefore, we also assume that the presented equation does not indicate that the seller does not sells stolen goods, i.e. $m_F(\{\neg stolen\}) = 0$. The parameter v_F is in these equations the weight of evidence. We can intuitively interpret this parameter as the reliability of respective evidence.

A variety of goods being sold. Let's suppose that the seller sells stolen goods in online auctions. He sells any kinds of goods that he "gets". The variety of goods being sold is then higher than that of the average proper seller. The belief functions of this attribute have the following form:

$$m_V(\{stolen_i\}) = \begin{cases} v_V \dfrac{V_i - \overline{V}}{V_i} & \text{for } V_i \le \overline{V} \\ 0 & \text{for } V_i > \overline{V} \end{cases},$$

$$m_V(\{\neg stolen_i\}) = 0, \tag{12}$$

$$m_V(\Theta_i) = \begin{cases} 1 - v_V \dfrac{V_i - \overline{V}}{V_i} & \text{for } V_i \le \overline{V} \\ 1 & \text{for } V_i > \overline{V} \end{cases},$$

where V_i is the amount of different types of goods sold by seller i, and $\overline{V}$ is the average amount of different types of goods sold by proper sellers in a respective category. The v_V parameter is the weight of evidence. We can intuitively interpret this weight as the reliability. It is valid that the higher the variety of goods the seller i sells, compared to the average types of goods sold by proper seller, the higher the suspicion that this seller i sells stolen goods. Therefore, we assume that the given equation does not indicate that the bidder does sell stolen goods, i.e. $m_V(\{\neg stolen_i\}) = 0$.

Combination of characteristic signs (evidences) of fraudulent behavior. A single characteristic is not enough to identify fraudulent behavior. Thus, once we have obtained more belief functions expressing our belief regarding fraudulent behavior, we combine them in a consistent manner to get a more complete assessment of what the whole group of evidences indicates. The combination of belief functions is done with the help of the Dempster combination rule (4). We express the assumption that a given seller i sells stolen goods with the help of belief function $m(\{stolen_i\})$. We calculate the value $m(\{stolen_i\})$ using the combination of single belief functions expressing appropriate evidence:

$$m(\{stolen_i\}) = (m_L \oplus m_F \oplus m_V)(\{stolen_i\}), \tag{13}$$

The operator $\oplus$ is the Dempster's rule of belief function combination (see equation 4). We perform the combination of multiple evidences according to the Dempster rule first we combine two belief functions, then we combine the result with the third belief function, fourth belief function and so forth. For example, the following rules combine the first and second belief functions:

$(m_L \oplus m_F)(\{stolen_i\}) = 1/K \ [m_L(\{stolen_i\})m_F(\{stolen_i\}) + m_L(\{stolen_i\})m_F(\Theta) + m_L(\Theta)m_F(\{stolen_i\})]$,
$(m_L \oplus m_F)(\{\neg stolen_i\}) = 1/K \ [m_L(\Theta)m_E(\Theta)]$,
where $K = 1 - (m_L(\{\neg stolen_i\})m_F(\{stolen_i\}) + m_L(\{stolen_i\})m_F(\{\neg stolen_i\}))$.

The Definition of the Reinforcement Coefficient α
We explored various sources of information on the Internet (on Internet forums, discussions etc.) that could potentially serve as a source of contextual information and could be used to improve the detection of fraud. We extracted pieces of information from these information sources and examined them in detail within the context of online auctions conducted on the Aukro auction portal (Aukro, 2013). For examples, we found complaints discussed in various Internet forums and discussions. For example, some users on these forums complained that somebody had stolen from them a certain object and this object appeared later on an Internet auction.

An example is the internet forum (Stolen Columbus, 2011). Here, a user mentions the theft of car navigation device Columbus. At the same forum (Stolen Columbus, 2011), other users inform him that they recently saw this navigation device offered for sale on the Internet auction portal Aukro (Aukro, 2013). Further conversation on this forum relates to whether it could be the stolen device mentioned by the first user. Some users say that it is for certain the stolen device because a) this particular navigation is factory-fitted, b) the price in the online auction is suspiciously low, and c) the online auction in which this device has been sold started few hours after the theft. This case belonged to the five mentioned prosecuted cases from the complaints of selling stolen goods in Internet auctions.

We performed further analysis of about 245 different Internet forums and found out that similar conversations occurred in 25 cases which we tried to connect to Internet auctions conducted on the Aukro online auction portal (Aukro, 2013). It is clear that every theft is not mentioned in an Internet forum. However, when we find out information about the theft in Internet forums or discussions, we consider this information as contextual information that can help identify fraudulent behavior - selling stolen goods in the internet auction.

We have defined the reinforcement coefficient α as follows: we find out that the stolen object appears in an online auction a short time (some hours) after the theft occurred. The thieves want to dispose of stolen goods as soon as possible. We have expressed the reinforcement coefficient α as a function of time:

$$\alpha = \begin{cases} Ke^{-k \cdot t} & \text{if time information about the theft is specified} \\ \alpha_0 & \text{if time information about the theft is not specified} \end{cases} \tag{14}$$

where K is equal to $m(\Omega)$ (see Section 4.2), k is a coefficient that is to be determined on the basis of statistical analysis and t is the difference between the start time of an online auction with goods which have been discussed as stolen on some Internet forum and the time of publication of the complaint on the respective Internet forum. We found out that the person who reports the theft usually provides the approximate time information as well. In the event that the time information about the theft is not specified, the α value is set to α_0. The value of α_0 is also to be qualified on the basis of statistical evaluations of analyzed auctions.

The Assessment of the Influence of Contextual Information on the Evidence about Fraudulent Behavior

Let's suppose that the chosen attributes of an online auction have been examined. The calculated belief assignment $m(\{stolen_i\})$ using the equation 13 of this auction indicates that the seller may be selling a stolen good. At the same time we find out that one of the users of an Internet forum complained that the same good was stolen from him. We consider this information as additional context information that reinforces our belief that stolen goods are sold within this auction. This reflects the fact that part of *BBM* is transferred from the total frame of discernment $m(\Omega)$, denoting our uncertainty, into the element $m(\{stolen_i\})$. This operation reflects our belief related to the sale of stolen goods (we denote this operation as de-discounting or reinforcement, see equations 7 and 8.

We can calculate the resulting belief about fraudulent behavior $m_R(stolen_i)$ by using the equation:

$$m_R(stolen_i) = \frac{m(stolen_i)}{1-\alpha} \qquad (15)$$

We have calculated the belief concerning the certain fraudulent online auction (stolen goods are sold within this auction) $m_R(\{stolen_i\})$ with the use of additional contextual information available from Internet sources. This additional information reinforces our confidence that stolen goods are sold within this auction. On the other side, our uncertainty concerning analyzed online auction that stolen goods are sold here will decrease.

Categorization of Sellers According to the Resulting Belief Function Representing the Behavior - Selling of Stolen Goods
We will divide users into categories according to the degree of belief that a certain user i sells stolen goods, i.e. $m_R(\{stolen_i\})$. These categories are: "*Proper seller*", "*Suspect seller*" and "*Seller sells stolen goods*".
We define two thresholds -η and ξ. The first threshold η is the threshold for determining whether a seller *i* is a proper seller. If the value of $m_R(\{stolen_i\})$ is below η, the seller *i* will be considered a proper seller. The second threshold ξ is the threshold for determining that a seller *i* sells stolen goods. If the value of $m_R(\{stolen_i\})$ exceeds ξ, the respective seller *i* will be considered a seller selling stolen goods. If the value of $m_R(\{stolen_i\})$ is between η and ξ, we will consider the seller "only" suspect of selling stolen goods. The thresholds η and ξ will be qualified on the basis of statistical evaluations of analyzed auctions.

The schema of the proposed model. Major task of our proposed model is to identify sellers who sell stolen goods and identify proper sellers. Figure 1 depicts the schema of our proposed model.
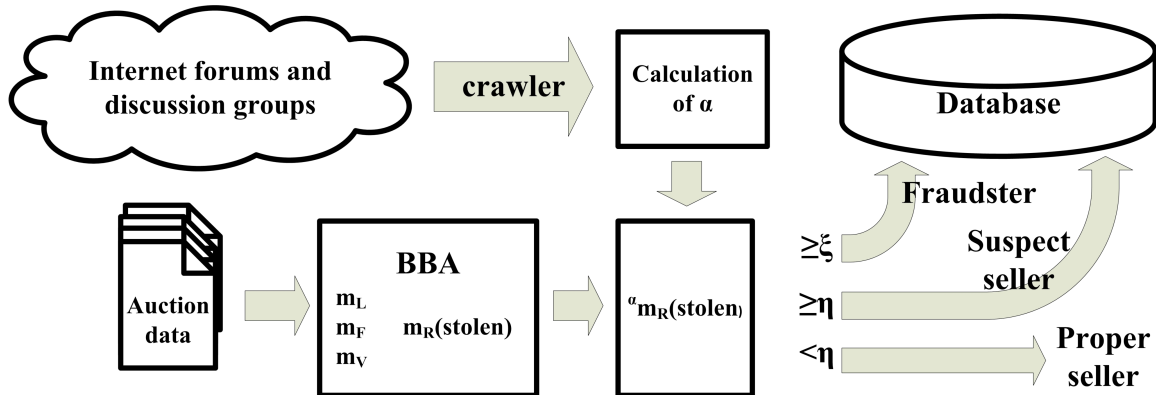

Figure 1: Schema of proposed model

Sellers are evaluated mathematically using a data fusion method that combines information from different information sources on the Internet and auction-level features. The threshold ξ of certifying sellers as Fraudsters, who sells stolen goods, should be fairly high to reduce the number of false positives.

For the sellers that are certified as Suspect, the values of their $m_R(\{stolen_i\})$ must be lower than ξ but greater than the values of threshold η. This means that the evidence is not sufficient enough to support a conclusion that a seller sells stolen goods, even though the seller behaved more like a fraudster than an honest proper seller. As a result, the seller is considered Suspect. When new additional independent evidence is presented, the Suspect certification will be revalidated. If a seller's certification changes, the new certification is committed to the database. If a seller's certification is labeled as Fraudster, the seller is subject to further investigation and possible punishment, but this fraudster-handling step is outside the scope of this paper.

CASE STUDY AND ANALYSIS OF RESULTS

The motivation for our study was an actual case with which we were familiar: a valuable specific brand-name radio was stolen from a car. The radio owner mentioned the theft including the information about the time of the event on an Internet forum. Others members of this forum reported, as reaction to this information, that an auction selling the same type of radio was taking place at the same time. The radio owner took legal steps then and the perpetrator of the theft was apprehended as the police co-operated with the operator of the online auction portal. Based on the experience from this case, we started a detailed examination of Internet forums and online auctions. We created a simple crawler and searched Czech Internet forums and discussion groups. The aim was to find notices about theft on these forums. Analyses of these notices were performed manually. We performed an analysis of 245 different pages of forums and discussions with the occurrence of words (or their English equivalents) relevant to theft presented in Table 1:

Table I. Searched word occurrence on scanned 256 Internet forums and discussion.

Searched word	Word occurrence (No.)
Theft	16
Thievery	4
Burglary	5
Abstraction	2
Pilferage	3
Robbery	6

Along with the analysis of Internet forums and discussions, we carried out analyses of online auctions in which objects mentioned on these Internet sources were being sold. We explored bidding history of auctions, prices, goods sold within these auctions and history of seller's transactions. We had to investigate all information manually because the Czech online auction portal Aukro (Aukro, 2013) does not have (in contrast to eBay) an API interface enabling automatic gathering of information.

Table II. Chosen auction data collected from Czech online auction Aukro (Aukro, 2013).

Seller i	Average price at which seller i sells certain goods (in CZK)	Average price of these goods throughout the online auction system (in CZK)	Number of goods which the seller i sells at fixed price	Total number of goods sold by seller i	Amount of different type of goods offered by seller i	Average amount of different type sold goods sold by sellers in respective category

D***r	1500	2525	2	2	2	2
O***2	700	1850	1	1	1	2
m***k	1250	1600	1	2	2	2
d***l	650	750	2	6	6	2
2***j	1420	1540	1	7	5	2
b***s	1200	1450	0	8	5	2

The values of belief that stolen goods are being sold are calculated using equations 10, 11, 12 and 13. The calculations are presented in Table III. The weights of evidence v_L, v_F and v_V were set in agreement with our experiments at 0.9, 0.7 and 0.8. We consider the character "Inadequate low price" as the most predicative. The character "Goods sold at fixed prices" is the less reliable in determining a seller selling of stolen goods.

Table III. The basic masses assigned to single "selling stolen goods" characteristics (equations 10, 11, 12 and 13).

Seller i	$m_L(\{stolen_i\})$	$m_F(\{stolen_i\})$	$m_V(\{stolen_i\})$	$m(\{stolen_i\})$	$m(\Omega)$
D***r	0.405941	0.7	0	0.821782	0.178218
O***2	0.621622	0.7	0	0.886486	0.113514
m***k	0.21875	0.35	0	0.492188	0.507813
d***l	0.133333	0.233333	0.533333	0.689926	0.310074
2***j	0.077922	0.1	0.48	0.568468	0.431532
b***s	0.172414	0	0.48	0.569655	0.430345

Table IV. The influence of contextual information on belief function (equation 15) and categorization of sellers according to the resulting belief function.

Seller i	$m(\{stolen_i\})$	$m(\Omega)$	t (hours)	α	$m_R(\{stolen_i\})$	$m_R(\Omega)$	Result
D***r	0.821782	0.178218	28	0,010837	0,830786	0,169214	*Suspect seller*
O***2	0.886486	0.113514	11	0,037785	0,921298	0,078702	*Seller sells stolen goods*
m***k	0.492188	0.507813	-	0	0,492188	0,507813	*Proper seller*
d***l	0.689926	0.310074	12	0,093393	0,760997	0,239003	*Suspect seller*
2***j	0.568468	0.431532	38	0,009654	0,574009	0,425991	*Proper seller*
b***s	0.569655	0.430345	42	0,006453	0,573355	0,426645	*Proper seller*

The value t is the difference between the time when the online auction selling goods which have been mentioned as stolen on some Internet forum began and the time of publication of this complaint on the respective Internet forum. The value of k (equation 14) was set on the basis of our experiments at 0.10. The value of threshold ξ was set on the basis of our experiments at 0.90 and the value of threshold η at 0.75. The values α and $m_R(\{stolen_i\})$ were calculated using equations 14 and 15. The value $m_R(\{stolen_i\})$ expresses our belief that the seller sells stolen goods. On the other side, values $m_R(\Omega)$ represent our uncertainty or rather ignorance concerning the classification if the seller i sells legal goods or it the seller i sells stolen goods. The values of $m_R(\{stolen_i\})$ are high in the case of seller D***r. They are greater than the threshold ξ, hence we consider him/her as seller who sells stolen goods. This seller must then be subject to further investigation. The value of $m_R(\{stolen_i\})$ is greater than the threshold η in the case of the sellers d***l and D***r.

These sellers are "only" suspected of selling stolen goods. It is recommended to monitor the behavior of these sellers. The values of $m_R(\{stolen_i\})$ of the sellers m***k, 2***j and b***s is less than the threshold η. They are considered to be proper sellers. Their behavior on the auction corresponds to the average behavior pattern.

CONCLUSION AND FUTURE WORK

Based on the conceptual framework of Dempster-Shafer theory, a practical approach for detection of a specific type of fraudulent behavior (selling stolen goods) has been proposed. This method in essence takes into consideration evidence found from different information sources, in this case from online auction system and from Internet information sources. Pieces of knowledge about auctions including bidding behavior were processed and quantified. Using the Dempster rule of combination, we combined evidence that enforces each other and resolved the conflicts between different pieces of evidence. We also took into account contextual information from various Internet sources. We consider this information as a factor that can influence (reinforce) our belief concerning the analyzed fraudulent behavior. The case study shows that our proposed approach is quite accurate and practical for real world deployment.

We verified our model on the Czech online auction Aukro (Aukro, 2013). We performed a number of experiments on this auction. We made certain that we can increase the detection of fraudulent behavior by using additional information from Internet sources. Nevertheless, we are also aware that the mathematical formalization of parameters used in our model (especially the parameters v_L, v_F, v_V; η and ξ) is necessary to increase the practical usefulness of our model.

In our future work, we want to define these parameters with the help of mathematical formulas. We want to perform further statistical analyses of online auctions and Internet sources to verify these formulas and the values of the parameters used in our model.

We are convinced that the presented approach represents a promising line of research. Similar methods and systems can be used in particular by auction portals that can monitor suspicious auctions and then provide warnings to users to simply pay attention to what they buy. In addition, law enforcement may use the system to investigate reports of fraud. Substantial limitation is that not all information concerning suspicious behavior on the auctions is available. However, the integration of information from various sources to detection of online fraud is a promising direction for future work.

REFERENCES

Aukro (2013). Aukro. URL: http://aukro.cz/

eBay (2013). eBay. URL: http://www.ebay.cz/.

Stolen goods on Aukro (2010). Cesi stale kupuji kradene veci. EuroZpravy.cz. URL: http://ekonomika.eurozpravy.cz/ceska-republika/6145-cesi-stale-kupuji-kradene-veci/ (in Czech).

Czech e-shops inspection (2012). Kontrola e-shop v roce 2012: Vedou nekalé obchodn praktiky. In: Oversito.cz URL: http://www.oversito.cz/aktualne/kontrola-e-shopu-v-roce-2012-vedou-nekale-obchodni-praktiky-92/ (in Czech).

Stolen Columbus (2011). Odcizeni Columbuse. Forum.Skodahome.cz. URL: http://forum.skodahome.cz/topic/114051-odcizeni-columbuse/ (in Czech).

Beranek, L. (2010a). Auditing Electronic Auction Systems: Knowing the Risks. ISACA Journal: Journal Online [online suplement], 4, URL: http://www.isaca.org/ Journal.

Beranek, L., Tlusty, P., Remes, R. (2010b). The Usage of Belief Functions for an Online Auction Reputation Model. In: Proceedings of the 28th Conference on Mathematical Method in Economy 2010. Economical faculty. University of South Bohemia, Ceske Budejovice, Czech republic, 49-54.

Beranek, L., Knizek, J. (2012). The Usage of Contextual Discounting and Opposition in Determining the Trustfulness of Users in Online Auctions. Journal of Theoretical and Applied Electronic Commerce Research, 7 (1), 34-50.

Bryant, A., Colledge, C. (2002). Trust in Electronic Commerce Business Relationships. Journal of Electronic Commerce Research, 3, 32-39.

Chae, M., Shim, S., Cho, H., Lee, B. (2010). Empirical analysis of online auction fraud: Credit card phantom transactions, Expert Systems with Applications,37, 2991-2999.

Chang, W.H., Chang, J.S. (2010a). A multiple-phased modeling method to identify potential fraudsters in online auctions. In Proceedings of the 2nd International Conference on Computer Research and Development, Kuala Lumpur, Malaysia, 186-190.

Chang, W.H., Chang, J.S. (2010b). Using clustering techniques to analyze fraudulent behavior changes in online auctions. In Proceedings of 2010 International Conference on Networking and Information Technology, Manila, Philippines, 34-38.

Chang, W.H., Chang, J.S. (2010c). An online auction fraud screening mechanism for choosing trading partners. In Proceedings of the 2nd International Conference on Education Technology and Computer, Shanghai, China, pp. V5-56-V5-60.

Chang, W.H., Chang, J.S. (2011). A novel two-stage phased modeling framework for early fraud detection in online auctions. Expert Systems with Applications, 38, 11244-11260.

Chang, W.H., Chang, J.S. (2012). An effective early fraud detection method for online auctions. Electronic Commerce Research and Applications, 11, 346-360.

Chau, D.H., Faloutsos, C. (2005). Fraud Detection in Electronic Auction. In: Proceedings of EWMF'05: European Web Mining Forum, at ECML/PKDD'05. October 3-7, Porto, Portugal, 87-97.

Chau, D.H., Pandit, S., Faloutsos, C. (2006). Detecting Fraudulent Personalities in Networks of Online Auctioneers. Principles and Practice of Knowledge Discovery in Database 2006, Springer, Lecture Notes in Computer Science Volume 4213, 103-114.

Chua, C., Wareham, J. (2002). Self Regulation for on-line Auctions: An Analysis. In 23rd International Conference on Information Systems (ICIS), Barcelona, Spain, December 2002, 115-125.

Da Silva Almendra, V., Schwabe, D. (2009). Analysis of Fraudulent Activity in a Brazilian Auction Site. In International World Wide Web Conference, WWW'09, ACM, Madrid, Spain.

Dong, F., Shatz, S.M., Xu, H. (2009). Combating online in-auction fraud: Clues, techniques and challenges. Computer Science Review, 3, 245-258.

Dong, F., Shatz, S.M., Xu, H. (2010). Reasoning under Uncertainty for Shill Detection in Online Auctions Using Dempster-Shafer Theory. International Journal of Software Engineering and Knowledge Engineering, 20, 943-973.

Dong, F., Shatz, S.M., Xu, H., Majumdar, D. (2012). Price comparison: A reliable approach to identifying shill bidding in online auctions? Electronic Commerce Research and Applications, 11, 171-179.

Ford, B.J., Xu, H., Valova, I. (2010). Identifying Suspicious Bidders Utilizing Hierarchical Clustering and Decision Trees. In Proceedings of the 2010 International Conference on Arti_cial Intelligence, ICAI 2010, July 12-15, 2010, Las Vegas Nevada, USA, vol. 2, 195-201.

Gavish, B., Tucci, C. (2006). Fraudulent auctions on the internet. Electronic Commerce Research, 6, 127-140.

Goel,A., Xu,H., Shatz S.M. (2010). A Multi-State Bayesian Network for Shill Verification in Online Auctions. In Proceedings of the 22nd International Conference on Software Engineering & Knowledge Engineering (SEKE'2010), Redwood City, San Francisco Bay, CA, USA, 279-285

Gregg, D. G., Scott, J. E. (2006). A typology of complaints about eBay sellers. Communications of the ACM, 51, 69-74.

Gregg, D.G., Scott, J.E. (2006). The role of reputation systems in reducing on-line auction fraud. International Journal of Electronic Commerce, 10, 95-120.

Gregg, D. Scott, J. (2008). A typology of complaints about eBay sellers. Communications of the ACM, 51, 69-74.

Hoffman, D. L., Novak, T. P., and Peralta, M. (1999). Building consumer trust online. Communications of the ACM, 42, 80-85.

Hsu, L.C., Wang, CH.H. (2008). A Study of E-trust in Online Auctions. Journal of Electronic Commerce Research, 9, 310-321.

Hu, Y., Panda, B. (2005). Design and analysis of techniques for detection of malicious activities in database systems. Journal of Network and Systems Management, 13, 269-291.

Ku, Y., Chen, Y., Chiu, C. (2007). A proposed data mining approach for Internet auction fraud detection. Intelligence and Security Informatics, Lecture Notes in Computer Science, vol. 4430, Springer, New York, NY, 238-243.

Kwan, M.Y.K., Overill, R.E., Chow, K.P., Silomon, M.A., Tse, H., Law, F., Lai P. (2010). Evaluation of Evidence in Internet Auction Fraud Investigations. In IFIP Int. Conf. Digital Forensics 2010, 121-132.

Mercier, D., Lefevre, E., Delmotte, F. (2012). Belief functions contextual discounting and canonical decompositions. International Journal of Approximate Reasoning, 53, 146-158.

Pandit, S., Chau, D. H., Wang, S., Faloutsos, C. (2007). NetProbe: A fast and scalable system for fraud detection in online auction networks. In Proceedings of the 16th international Conference on the World Wide Web, Banf, Canada, 201-210.

Sakurai, Y., Yokoo, M. (2003). A false-name-proof double auction protocol for arbitrary evaluation values. In Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems, ACM New York, NY, 329-336.

Shafer, G. (1976) A mathematical theory of evidence. Princeton: Princeton University Press.

Smets, P. (1993). Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem. International Journal of Approximate Reasoning, 9, 1-35.

Trevathan, J., Read, W. (2009). Detecting Shill Bidding in Online English Auctions. In Handbook of Research on Social and Organizational Liabilities in Information Security, M. Gupta and R. Sharman (eds.), Information Science Reference, 446-470.

Trevathan, J., Read, W. (2008). Secure online English auctions. In Proceedings of the Conference E-Business and Telecommunication Networks, Porto 2008, Springer-Verlag Berlin, 119-133.

Trevathan, J. (2005). Security, anonymity and trust in electronic auctions. Crossroads, 11, 3-9.

Trevathan, J., Read, W. (2007). Detecting Collusive Shill Bidding. In Fourth International Conference on Information Technology: New Generations (ITNG 2007), Las Vegas, Nevada, USA, IEEE Computer Society, 799-808.

Wang, J., Chiu, C.Q. (2005). Detecting online auction inflated-reputation behaviors using social network analysis. In Proceedings of North American Association for Computational Social and Organizational Science, Notre Dame, Indiana.

Xu, H., Cheng, Y. (2007). Model Checking Bidding Behaviors in Internet Concurrent Auctions. International Journal of Computer Systems Science & Engineering, 22, 179-191.

Zhang, B., Zhou, Y., Faloutsos, C. (2008). Toward a Comprehensive Model in Internet Auction Fraud Detection. In Proc. of the 47th Hawaii International Conference on System Sciences, 79.