Evaluation of Evidence in Internet Auction Fraud Investigations

Michael Kwan, Richard Overill, Kam-Pui Chow, Jantje Silomon, Hayson Tse, Frank Law, Pierre Lai

Chapter 9

EVALUATION OF EVIDENCE IN INTERNET AUCTION FRAUD INVESTIGATIONS

Michael Kwan, Richard Overill, Kam-Pui Chow, Jantje Silomon, Hayson Tse, Frank Law and Pierre Lai

Abstract Internet auction fraud has become prevalent. Methodologies for detecting fraudulent transactions use historical information about Internet auction participants to decide whether or not a user is a potential fraudster. The information includes reputation scores, values of items, time frames of various activities and transaction records. This paper presents a distinctive set of fraudster characteristics based on an analysis of 278 allegations about the sale of counterfeit goods at Internet auction sites. Also, it applies a Bayesian approach to analyze the relevance of evidence in Internet auction fraud cases.

Keywords: Internet auction fraud, Bayesian network, relevance of evidence

1. Introduction

According to the Data Center of the China Internet [5], Chinese users spent 2.56 trillion Renminbi ($698 billion) on the Internet during the first half of 2008, a 58.2% increase over the same period in 2007. Of the total amount, 35% was spent on purchases made via the Internet; the remaining 65% was spent on on-line games and network communities. China already has more Internet users than any other country in the world, and the number of users is expected to nearly double from 253 million in 2008 to 480 million in 2010 [2]. By 2010, the volume of online transactions in China will exceed those in Japan and South Korea [2].

Internet auctions offer buyers unparalleled selections of products and the opportunity to make great deals. They also provide sellers with a means to reach millions of potential buyers. Meanwhile, criminals are attracted by the low entry costs and tremendous profits of Internet auc-

tions. Unscrupulous sellers take advantage of buyers by misrepresenting the quality or condition of their goods. Some have no intention of delivering the goods that are offered for sale. As a result, Internet auction fraud is the most common type of fraud reported in the electronic commerce domain [16].

This paper examines the characteristics of Internet auction fraud in Hong Kong related to the sale of counterfeit goods (i.e., goods bearing false trade descriptions or forged trademarks). In addition, it uses Bayesian network models representing the prosecution and defense viewpoints in conjunction with the likelihood ratio as a criterion to determine the relevance of digital evidence in Internet auction fraud cases.

2. Background and Related Work

This section reviews the nature of Internet auction fraud. Also, it surveys approaches for detecting fraud in online auctions.

2.1 Internet Auctions

Internet auctions are successful for many reasons. Potential buyers have sufficient time to search for items of interest and they can bid for items 24 hours a day, seven days a week. The Internet does not impose geographical constraints on buyers and sellers and they are not required to be physically present at an auction. The large numbers of buyers and sellers tend to reduce selling costs as well as sales prices. Many users describe their online auction experience as comparable to gambling. Offering the highest bid provides the same thrill as winning a game.

Ochaeta [16] lists six basic features of Internet auctions:

- **Initial Buyer and Seller Registration:** This step helps authenticate the trading parties. It involves the exchange of cryptographic keys and the creation of a profile for each trader. The profile reflects the trader's interest in products and possibly his/her authorized spending limits.

- **Auction Set Up:** This step sets up the auction protocol and rules such as item descriptions, auction type (e.g., open cry, sealed bid or Dutch), negotiated auction parameters (e.g., price, delivery dates, terms of payment), auction starting time and duration, and auction closing conditions.

- **Scheduling and Advertising:** In order to attract potential buyers, items in a given category (e.g., art or jewelry) are generally

auctioned together on a regular schedule. Popular items are sometimes mixed with less popular items. Items to be sold in upcoming auctions and the dates of upcoming auctions are advertised.

- **Bidding:** The bidding step handles the collection of bids from potential buyers. It implements the bid control rules (e.g., minimum bid, bid increment, bid deposit). It also notifies auction participants when higher bids are received.

- **Bid Evaluation and Auction Closing:** This step implements the auction closing rules and notifies the winners and losers.

- **Trade Settlement:** This final step handles payments to sellers and the transfer of goods to buyers. If the seller is not the auctioneer, this final step also includes the payment of fees to the auctioneer and other agents.

2.2 Internet Auction Fraud

Criminals have discovered the Internet to be a highly profitable venue for conducting illicit business activities [7]. Organized crime groups are involved in numerous technology-enabled crimes, including Internet auction fraud [3].

Sakurai and Yokoo [18] have observed that anonymity is an important factor in perpetrating Internet fraud and that the existence of indivisible bids causes difficulty in matching supply and demand. This is because a buyer or seller can submit a false name bid by pretending to be a potential buyer or seller, thereby manipulating the balance of supply and demand. Chae, *et al.* [1] have confirmed these observations, concluding that online auction fraud is successful due to information asymmetry and anonymity.

Chua and Wareham [4] have listed some of the reasons for the proliferation of Internet auction fraud. The high degree of anonymity is at the top of the list; it is easy for dishonest users to evade prosecution. Second on the list is the low cost of entry and exit. Interestingly, these are precisely the reasons for the success of Internet auctions.

According to the 2008 Internet Crime Report [9], the median loss per Internet fraud complaint in the United States was $931 in 2008; the total loss was $264.6 million. In all, there were 275,284 Internet crime complaints – auction fraud, non-delivery of purchased goods, credit/debit card fraud, computer intrusions, spam and child pornography. However, Internet auction fraud was the most commonly reported offense, comprising 25.5% of all complaints and 16.3% of the total reported loss. The average median loss per auction fraud complaint was $610.

Table 1. Internet fraud taxonomy.

Seller as Fraudster	
Bid Shilling	Seller bids on his own items to drive up the price
Misrepresentation	Seller intentionally misrepresents an item
Fee Stacking	Seller adds hidden costs such as handling charges after the auction
Failure to Ship	Seller does not send the items to the buyers
Reproductions and Counterfeits	Seller advertises counterfeit items as the real thing
Triangulation Fencing	Stolen items are sold
Shell Auction	Seller sets up an auction solely to obtain bank account and credit card information
Buyer as Fraudster	
Bid Shielding	Two buyers collude – one makes a low bid, while the other makes an inflated bid; the higher bidder withdraws before the auction ends
Failure to Pay	Buyer does not pay for the items
Buy and Switch	Buyer refuses the items, but keeps the original items and returns inferior items
Loss or Damage Claims	Buyer claims the items are damaged, disposes of them and requests a refund

Gregg and Scott [8] discovered that Internet auction fraud takes various forms, such as delivering goods that are different, of low quality, without ancillary components, defective, damaged or black market items.

Morzy [15] describe other practices, including bid shielding and bid shilling. Bid shielding is the offering of an artificially high bid for an item to discourage other bidders from competing for the item. At the last moment, the "shielder" withdraws the high bid, enabling the second highest bidder, who is usually an accomplice, to win the auction. Bid shilling involves the use of a false bidder identity to drive up the price of an item on behalf of the seller.

Gregg and Scott [8] note that accumulation fraud is on the increase. In this type of fraud, a seller builds his reputation by selling large quantities of low-value merchandise over a long period of time. Having earned a good reputation, the seller offers expensive goods, but does not send the goods to buyers after receiving payment for them.

Chua and Wareham [4] created the auction fraud taxonomy presented in Table 1. According to Chua and Wareham, all the types of fraud

listed are very damaging to Internet auction houses. They undermine user trust, which is disastrous for business.

Ku, *et al.* [12] note that while both buyers and sellers can be victims of fraud, a buyer is more easily targeted than a seller. They observed that 89% of seller-buyer pairs conducted just one transaction during the time period of their study; at most, there were four transactions between a seller-buyer pair. This means that the repeated transaction rate for the same seller-buyer pair is lower than 2%. If the transaction rate is much higher than 2%, then the transactions between the seller-buyer pair are suspect and could involve bid shilling or bid shielding.

Kobayashi and Ito [11] observed that many fraudsters tend to make honest deals during the early stages of their auction lives. However, they commit fraud soon after earning good reputations.

Ochaeta [16] also observed that fraudsters tend to establish good reputations prior to committing fraudulent acts. Therefore, the reputation building process of fraudsters is different from that of legitimate users. Specifically, fraudsters attempt to gain as much one-time profit as possible and as quickly as practicable. Consequently, fraudsters can be identified based on their reputation-building activities.

Fraudsters attempt to build their reputations by buying or selling numerous cheap items from sellers with good reputations. Additionally, they may buy or sell moderately priced or expensive items to accomplices. These buying and selling activities generally take place over a short period of time.

In order to build good reputations over a short period of time, most Internet auction fraudsters tend to sell large amounts of low-priced products. These sales take place at the beginning of their fraudulent auction lives. Also, fraudsters may attempt to bid for inexpensive items from sellers with good reputations. This is done to establish a favorable reputation by conducting many legitimate transactions.

3. Internet Auction Fraud in Hong Kong

We conducted a statistical analysis of 278 cases in Hong Kong to reveal the characteristics of Internet auction fraud related to the sale of counterfeit goods. The cases were the result of complaints lodged with the Hong Kong Customs and Excise Department. The following characteristics were observed in the analyzed cases:

- Fake goods are sold at unreasonably low prices, about 10% of the prices of legitimate goods.
- In about two-thirds of the cases (180 out of 278), the goods are sold within seven days of account creation.

- Fraudsters have multiple auction accounts that do not carry high trust values or reputation scores (8 out of 10 or higher).
- Fraudulent accounts are short lived (less than ten days) and fraudsters tend to switch to other auction accounts before the auction period expires.
- Many categories of goods (more than five) are sold (e.g., watches, mobile phones, footwear and sportswear).

4. Investigative Model

This section describes an investigative model for online auction fraud involving the sale of counterfeit goods. The model employs a Bayesian network to support the reasoning about evidentiary hypotheses.

4.1 Hypotheses and Evidentiary Traces

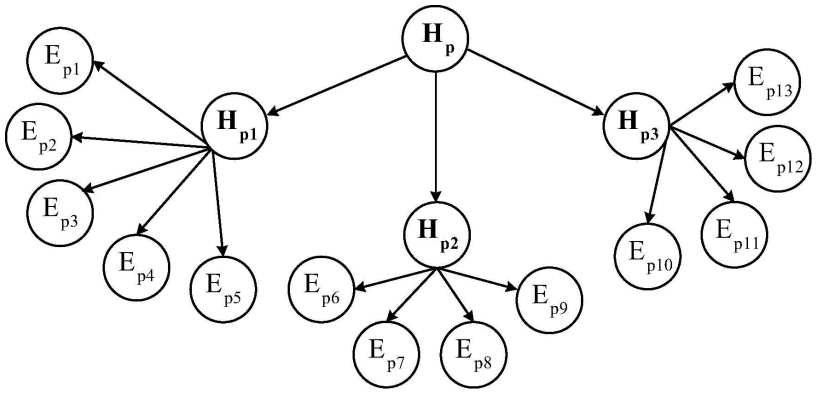
Digital evidence related to twenty prosecuted cases from the 278 complaints of selling counterfeit goods in Internet auctions was used to frame three sub-hypotheses about the actions taken by fraudsters. Because detailed judgments were not available for the prosecuted cases, digital forensic examiners who worked on the cases were interviewed to elicit the sub-hypotheses. The three sub-hypotheses are:

- Auction-related materials (e.g., images and item descriptions) were downloaded.
- The auction item (e.g., price of the item) was manipulated.
- The buyer and seller communicated (e.g., via email or instant messaging) about the counterfeit item.

These three sub-hypotheses substantiate the overall prosecution hypothesis that an online auction fraud crime was committed in the twenty prosecuted cases. The sub-hypotheses are supported by thirteen distinct evidentiary traces, which were obtained from the responsible digital forensic examiners. The various hypotheses and evidentiary traces are expressed using a Bayesian network model shown in Figure 1.

This investigative model does not of itself substantiate the entire prosecution case. The auctioned item has to be procured by the investigator and then be examined by the trademark owner to ascertain whether or not the item is counterfeit.

In order to evaluate the relevance of the digital evidential traces, a second simple Bayesian network model is created to express the defense



Hypotheses
H_p: Seized computer was used as a transaction tool for the auction of the counterfeit item
H_{p1}: Uploading of auction material related to the counterfeit item was performed
H_{p2}: Manipulation of the corresponding auction item took place
H_{p3}: Communication between the seller and buyer about the counterfeit item occurred

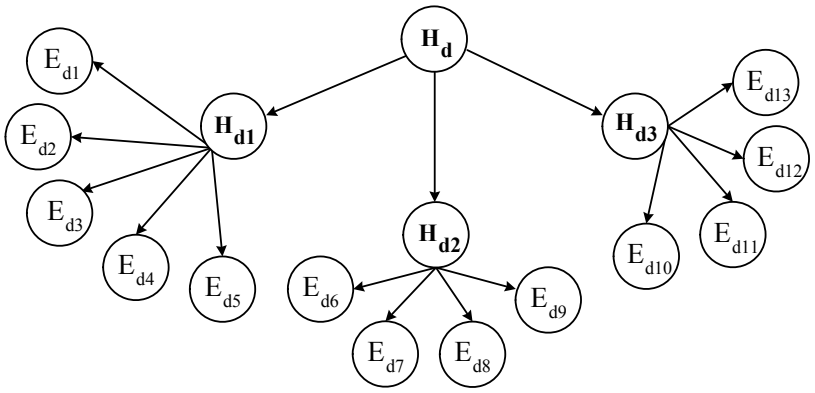
Evidence
E_{p1}: Information about the counterfeit item (e.g., image, description) was found on the seized computer
E_{p2}: Seller's account login record was retrieved from the auction site
E_{p3}: File metadata found on the seized computer matched the metadata found on the auction site
E_{p4}: IP address assigned to the seized computer matched the IP address used for data transfer
E_{p5}: Internet history/cache contents on the seized computer indicated the transfer of the counterfeit item
E_{p6}: Seller's account login record was retrieved from the auction site
E_{p7}: IP address assigned to the seized computer matched the IP address used for data transfer
E_{p8}: Editing of the auction item (e.g., price adjustment) occurred on the auction site
E_{p9}: Information about the auction item (e.g., image, description) was found on the seized computer
E_{p10}: Messages from the auction site related to the auction item were found on the seized computer
E_{p11}: Messages to/from the buyer related to the auction item were found on the seized computer
E_{p12}: Address book containing the covert investigator's email address was found on the seized computer
E_{p13}: IP address assigned to the seized computer matched the IP address used for email communication

Figure 1. Bayesian network model for prosecution hypotheses and evidentiary traces.

viewpoint. Figure 2 presents the defense hypotheses and their associated evidentiary traces. Although the root hypotheses of the defense and prosecution models appear to be the same, they are, in fact, different because of the different supporting sub-hypotheses that express the defense and prosecution viewpoints. However, the same set of evidentiary traces is used in both models.

4.2 Evidence Evaluation

We use the likelihood ratio (LR) to evaluate the evidence in Internet auction fraud cases. LR is a general technique that can be applied to any scenario with decision uncertainty. In particular, it is very effective for quantifying the value or relevance of evidence [14]. The closer the LR



Hypotheses
H_d: Seized computer was used as a transaction tool for the auction of the counterfeit item
H_{d1}: Downloading of auction material related to the counterfeit item was performed
H_{d2}: Manipulation of the non-counterfeit auction item took place
H_{d3}: Communication between the seller and buyer about the non-counterfeit item occurred

Evidence
E_{d1}: Information about the counterfeit item (e.g., image, description) was found on the seized computer
E_{d2}: Seller's account login record was retrieved from the auction site
E_{d3}: File metadata found on the seized computer matched the metadata found on the auction site
E_{d4}: IP address assigned to the seized computer matched the IP address used for data transfer
E_{d5}: Internet history/cache contents on the seized computer indicated the transfer of the counterfeit item
E_{d6}: Seller's account login record was retrieved from the auction site
E_{d7}: IP address assigned to the seized computer matched the IP address used for data transfer
E_{d8}: Editing of the auction item (e.g., price adjustment) occurred on the auction site
E_{d9}: Information about the auction item (e.g., image, description) was found on the seized computer
E_{d10}: Messages from the auction site related to the auction item were found on the seized computer
E_{d11}: Messages to/from the buyer related to the auction item were found on the seized computer
E_{d12}: Address book containing the covert investigator's email address was found on the seized computer
E_{d13}: IP address assigned to the seized computer matched the IP address used for email communication

Figure 2. Bayesian network model for defense hypotheses and evidentiary traces.

value is to one, the less relevant is the evidence. Evett [6] generalized the LR approach to represent a situation where it is uncertain if the evidence is the result of the activities of a suspect. The general form proposed by Evett is:

$$LR = \frac{Pr(E|H_p)}{Pr(E|H_d)}$$

where E is the total digital evidence related to the crime, and H_p and H_d are the overall prosecution hypothesis and the overall defense hypothesis, respectively.

In our simple Bayesian network model, the existence of each individual trace of digital evidence does not imply the existence of any other traces. Since the evidentiary traces are mutually independent, their individual

Table 2. Conclusions drawn on LR values.

Likelihood Ratio	Evidentiary Support
1 to 10	Limited
10 to 100	Moderate
100 to 1,000	Moderately Strong
1,000 to 10,000	Strong
More than 10,000	Very Strong

probabilities can be multiplied together to determine the probability of E given a root hypothesis. The prior probability values of the individual evidentiary traces for the Internet auction fraud models (prosecution and defense) were obtained by surveying digital forensic examiners with the Hong Kong Customs and Excise Department, and are generally accepted values within this community of experts.

Evaluation of the Individual Sub-Hypotheses To evaluate the evidentiary relevance or LR values of the individual sub-hypotheses, it is necessary to set the individual sub-hypotheses to "Yes" separately and then multiply the prior probability values of their associated evidence. Thus, the LR value of evidence for hypothesis H_p against the evidence for hypothesis H_d ($Pr(E|H_p)/Pr(E|H_d)$) is given by:

$$\begin{aligned} & \frac{Pr(E_{p1}|H_{p1}) \times Pr(E_{p2}|H_{p1}) \times Pr(E_{p3}|H_{p1}) \times Pr(E_{p4}|H_{p1}) \times Pr(E_{p5}|H_{p1})}{Pr(E_{d1}|H_{d1}) \times Pr(E_{d2}|H_{d1}) \times Pr(E_{d3}|H_{d1}) \times Pr(E_{d4}|H_{d1}) \times Pr(E_{d5}|H_{d1})} \\ \approx & \frac{0.9 \times 0.75 \times 0.6 \times 0.75 \times 0.85}{0.9 \times 0.05 \times 0.6 \times 0.01 \times 0.01} \approx \frac{0.258}{0.0000027} \approx 95,600 \end{aligned}$$

Applying the interpretation adopted by the U.K. Forensic Science Service [10], an LR value of 95,600 indicates very strong support of the evidence for the prosecution's claim over the defense's claim. Table 2 illustrates the interpretation used by the Forensic Science Service.

Similarly, the LR values for H_{p2} against H_{d2} and for H_{p3} against H_{d3} are given by:

$$\begin{aligned} \frac{Pr(E|H_{p2})}{Pr(E|H_{d2})} & \approx \frac{0.247}{0.000319} \approx 774 \\ \frac{Pr(E|H_{p3})}{Pr(E|H_{d3})} & \approx \frac{0.190}{0.000938} \approx 203 \end{aligned}$$

The computed LR value indicates very strong evidentiary support for the prosecution's sub-hypothesis H_{p1}. On the other hand, the LR values indicate that the evidence supports the prosecution's sub-hypotheses H_{p2} and H_{p3} moderately strongly.

A limitation exists in the application of the LR approach to evaluate the evidentiary relevance of individual sub-hypotheses. In order to compute LR values, the corresponding sub-hypotheses should exist in the Bayesian network models expressing the prosecution and defense viewpoints. This requirement renders the LR approach inapplicable when the sub-hypotheses in two models do not correspond to each other (e.g., the number of sub-hypotheses in the defense Bayesian network model is larger than the number in the prosecution model).

However, under normal circumstances, the sub-hypotheses in both models will correspond because most of the sub-hypotheses in the defense model stem from the sub-hypotheses in the prosecution model. Evaluating the evidentiary relevance of individual sub-hypotheses can identify the strongest and weakest sub-hypotheses in the models. This enables digital forensic practitioners to identify the most significant and/or the most insignificant groups of evidence that are encompassed by the individual sub-hypotheses.

Evaluation of the Overall Hypotheses To compute $Pr(E|H_p)$, it is necessary to set the root hypothesis H_p of the prosecution Bayesian network to "Yes" and then multiply the resulting probability values of E_{p1} to E_{p13}. Similarly, to compute $Pr(E|H_d)$, it is necessary to set the root hypothesis H_d of the defense Bayesian network to "No" and multiply the resulting probability values of E_{d1} to E_{d13}. Specifically, we have:

$$Pr(E|H_p) \approx 0.000293; \quad Pr(E|H_d) \approx 0.00000000179$$

Hence,

$$LR = \frac{Pr(E|H_p)}{Pr(E|H_d)} \approx \frac{0.000293}{0.00000000179} \approx 164,000$$

The LR value of 164,000 indicates very strong evidentiary support for the prosecution's claim over the defense's claim.

5. Conclusions

The analysis of allegations of counterfeit goods at Internet auction sites provides interesting insights into fraudster behavior. Bayesian networks and likelihood ratio values offer a powerful mechanism for analyzing the relevance of evidence in such cases. If all the evidentiary

traces are initially assumed to be present, the LR values computed from the prosecution and defense models can be used as criteria to determine whether or not it is worthwhile to proceed with the search for evidentiary traces. Specifically, if the LR value is relatively large (greater than 1,000) the search for the implied digital evidence should proceed. This would be followed by applying a cost-effective digital forensic investigation model [17] to identify the evidentiary traces and then applying the Bayesian network model [13] with the retrieved traces. On the other hand, if the LR value is found to be relatively small, the evidence does not strongly support the chosen hypotheses. Therefore, the prosecution should review its hypotheses and/or the implied evidentiary traces.

Acknowledgements

The authors wish to thank Dr. Jeroen Keppens of the Department of Computer Science, King's College London for his technical assistance.

References

[1] M. Chae, S. Shim, H. Cho and B. Lee, Empirical analysis of online auction fraud: Credit card phantom transactions, *Expert Systems with Applications*, vol. 37(4), pp. 2991–2999, 2010.

[2] China Internet Network Information Center, Statistical Survey Report on Internet Development in China (Abridged Edition), Beijing, China (www.cnnic.net.cn/uploadfiles/pdf/2008/8/15/145744.pdf), 2008.

[3] R. Choo, Organized crime groups in cyberspace: A typology, *Trends in Organized Crime*, vol. 11(3), pp. 270–295, 2008.

[4] C. Chua and J. Wareham, Self-regulation for online auctions: An analysis, *Proceedings of the Twenty-Third International Conference on Information Systems*, pp. 115–125, 2002.

[5] Data Center of the China Internet, The First Half of 2008 China Internet User Measurement Data IUI Index Report, Beijing, China, 2008.

[6] I. Evett, Establishing the evidential value of a small quantity of material found at a crime scene, *Journal of the Forensic Science Society*, vol. 33(2), pp. 83–86, 1993.

[7] S. Gajek and A. Sadeghi, A forensic framework for tracing phishers, *Proceedings of the Third International Conference on the Future of Identity in the Information Society*, pp. 19–33, 2008.

[8] D. Gregg and J. Scott, A typology of complaints about eBay sellers, *Communications of the ACM*, vol. 51(4), pp. 69–74, 2008.

[9] Internet Crime Complaint Center, 2008 Internet Crime Report, National White Collar Crime Center, Richmond, Virginia, 2008.

[10] J. Keppens, Towards qualitative approaches to Bayesian evidential reasoning, *Proceedings of the Eleventh International Conference on Artificial Intelligence and Law*, pp. 17–25, 2007.

[11] M. Kobayashi and T. Ito, A transactional relationship visualization system in Internet auctions, *Studies in Computational Intelligence*, vol. 110, pp. 87–99, 2008.

[12] Y. Ku, Y. Chen and C. Chiu, A proposed data mining approach for Internet auction fraud detection, *Proceedings of the Pacific Asia Workshop on Intelligence and Security Informatics*, pp. 238–243, 2007.

[13] M. Kwan, K. Chow, F. Law and P. Lai, Reasoning about evidence using Bayesian networks, in *Advances in Digital Forensics IV*, I. Ray and S. Shenoi (Eds.), Springer, Boston, Massachusetts, pp. 275–289, 2008.

[14] D. Lucy, *Introduction to Statistics for Forensic Scientists*, Wiley, Chichester, United Kingdom, 2005.

[15] M. Morzy, New algorithms for mining the reputation of participants of online auctions, *Algorithmica*, vol. 52(1), pp. 95–112, 2008.

[16] K. Ochaeta, Fraud Detection for Internet Auctions: A Data Mining Approach, Ph.D. Thesis, College of Technology Management, National Tsing-Hua University, Hsinchu, Taiwan, 2008.

[17] R. Overill, M. Kwan, K. Chow, P. Lai and F. Law, A cost-effective model for digital forensic investigations, in *Advances in Digital Forensics V*, G. Peterson and S. Shenoi (Eds.), Springer, Heidelberg, Germany, pp. 231–240, 2009.

[18] Y. Sakurai and M. Yokoo, A false-name-proof double auction protocol for arbitrary evaluation values, *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 329–336, 2003.