JOURNAL OF CYBERSECURITY

Journal of Cybersecurity, 2016, 1–9
doi: 10.1093/cybsec/tyw007
Research paper

Research paper

Examining signals of trust in criminal markets online

Thomas J. Holt[1]*, Olga Smirnova[2] and Alice Hutchings[3]

[1]School of Criminal Justice, Michigan State University, East Lansing, MI 48824, USA; [2]Department of Political Science, Eastern Carolina University, Greenville, NC 27858, USA; [3]Computer Laboratory, University of Cambridge, Cambridge, UK

*Corresponding author. E-mail: holtt@msu.edu.

Received 13 November 2015; revised 15 July 2016; accepted 27 July 2016

Abstract

This study examines the signals of trust in stolen data advertisements by analysing the structural and situational factors that influence the type of feedback sellers receive. Specifically, this article explores the factors associated with positive and negative buyer feedback from the purchase of stolen credit card data in a series of advertisements from a sample of Russian and English language forums where individuals buy and sell personal information. The results of zero-inflated Poisson regression models suggest that the sellers may influence their likelihood of receiving feedback by specifying the type of payment mechanism, choosing the advertisement language and selecting the type of market they operate within. The implications of this study for our understanding of online illicit markets, criminological theory and policy-making will be explored in depth.

Akerlof's [1] idea that lemon markets are created by information asymmetry has been widely accepted within economics, however little is explored within criminology (exceptions include [41], and [39], both in respect to drug markets). Information asymmetry refers to buyers and sellers having different information about the quality of a product, which creates a lemon market [1]. Buyers in lemon markets are unable to differentiate between sellers offering quality products and those offering poor quality products, or lemons. Buyers are only willing to pay a price somewhere between that of quality and lemon products, which forces quality sellers to either accept a reduced price for their products or to leave the market [1]. Buyers are also less inclined to participate in the market due to the reduced quality of goods which may further reduce the quality of products. Quality sellers are also less active in such markets because they do not get the full price for their products.

The concept of information asymmetry has particular value for criminological inquiry regarding illicit markets for goods and services including drugs [25–28, 30, 49] and prostitution [17, 18]. Within transactional markets, criminals have little opportunity for recompense in the event they experience loss or are intentionally duped into buying poor quality products or are dissatisfied with a service [27, 30, 47]. Individuals cannot contact police due to risk of arrest, and may instead depend on vigilante justice in order to compensate financial losses [27, 44]. As a consequence, some criminals may willingly exploit this lack of social control in order to deceive participants for their own economic advantage.

Researchers have expanded on information asymmetry through the use of signalling theory as established in a wide variety of disciplines, from economics to biology [8, 43]. Gambetta's [9] recent contributions to both criminology and signalling theory expands our understanding of the ways criminals identify themselves to each other and signal trustworthiness in an otherwise untrusted environment. Specifically, when there is information asymmetry, it is in a signaller's best interests to signal their trustworthiness regardless of whether they are actually trustworthy. Untrustworthy actors attempt to mimic the signals used by their trustworthy counterparts, making it in the receiver's best interest to differentiate between those who are trustworthy and those that are not. Legitimate actors use signals that may be too costly for untrustworthy actors to replicate, which provides a potential way for receivers to interpret signals produced.

Within face-to-face transactions, such as drug markets or prostitution, actors determine an individual's legitimacy and reputation based on verbal and non-verbal cues in order to engage in a successful transaction of goods and services [17, 18, 25–30, 47, 49]. For in-

stance, Jacobs [26, 28] found that drug dealers were extremely suspicious of unknown clients, and would look for cues that indicated they might be law enforcement representatives, such as their style of speech and physical posture. Hegghammer [10] examined the signals that terrorist recruiters used, finding that they preferred not to recruit online and instead used visual and verbal signals, such as ethnicity, reactions to prayers and the reciting of poems.

These studies are invaluable to understand how actors in real-world face-to-face illicit transactional markets are structured by internal and external risks. There is, however, minimal research on the ways that risk avoidance techniques are employed and affect the practices of offenders in discrete or low visibility, markets. Few have considered how online markets for illicit products, including drugs, [2] malicious software [15] and stolen financial information [11, 24, 35] are shaped by actor's perceptions of risk and the signalling mechanisms provided. Robust transactional markets have developed online, and researchers like Tzanetakis and associates (2016) argue that online markets provide participants with a substantial amount of information that can be used to minimize risk and harm (see also [3]). At the same time, the faceless nature of the Internet makes it difficult for actors to rely on signals of quality or identity in order to assess trust and reliability. Text-based online mediums make it difficult for individuals to interpret foreground or situational cues, as they can be easily faked (e.g. [7, 11, 21]).

We use information asymmetry and Gambetta's signalling theory to demonstrate the ways that sellers in stolen data markets signal their prospective trustworthy nature within the marketplace, and the ways that buyers respond to those signals via feedback. In stolen data markets, dishonest traders who cheat buyers are referred to as 'rippers' [7]. Legitimate sellers of stolen data compete in the same marketplaces, though they must differentiate themselves from rippers in order to be successful. Dishonesty in data markets has tangible economic costs, creating a prospective 'tax' on all purchases [11]. Essentially, genuine sellers of illegal goods will never be able to receive the true value for their products, as they have to accept reduced payments as a cost of doing business in an environment where malicious actors are present [11]. Data buyers must also discern the legitimacy of a seller in order minimize their risk of loss [18, 22].

Understanding information asymmetry in stolen data markets is essential to improve our knowledge of the signals that demonstrate a seller is trustworthy, and identify the formal and informal factors that encourage vendor success. Recognizing the practices of sellers and their influence on buyer reviews can also increase our understanding of the social relationships that affect individual's position within the market generally (see also [35]). In turn, the findings may enable us to identify the factors that encourage market failure and reduce demand by pushing quality sellers out of the market.

Thus, this study used Gambetta's [8] signalling theory to identify what signals are used by sellers to indicate trustworthiness within their advertisements for stolen data on Russian and English language web forums. A zero-inflated Poisson regression model was conducted using positive feedback as the dependent variable to identify the unique advertiser and market-specific factors that were associated with trust in a seller. Positive feedback was used as it most closely corresponds to the notion of signalling theory in that positive comments are associated with the trust in market actors. The findings demonstrate that signals in online environments are of mixed value and may not be easily interpreted by prospective clients. The implications of this study for our understanding of information asymmetry, signalling theory and public policy are discussed in detail.

Stolen data market operations and signalling theory

Over the last decade, there have been multiple incidents of massive data breaches affecting retailers, payment processors, and government entities by hackers across the globe [12, 37, 42]. Actors gain internal access to sensitive data feeds and systems, and then acquire millions of credit and debit card details, as well as sensitive personally identifiable information. The sheer quantity of information that can be acquired by a small group of hackers and data thieves is beyond the capacity of these groups to use efficiently or effectively without being detected.

As a result, there is now a burgeoning market for individuals to sell the information that they obtain through hacking and other forms of data theft directly to others via web forums and Internet Relay Chat channels, and a corresponding body of scholarship studying this phenomenon [7, 11, 16, 20, 21, 22, 23, 24, 35, 50, 51]. Though these markets are hosted in various countries around the world, many of the most active appear to operate out of Russia and Eastern Europe [15, 36, 45].

The text-based nature of these markets coupled with the range of goods and services offered (e.g. [7, 18]) creates a substantial environment for information asymmetry that can negatively impact buyers and sellers. It is unclear how legitimate sellers and buyers signal identity, intentions, and the quality of goods and services in a way that cannot be easily mimicked by rippers. Rippers may be active in stolen data markets as there is little consequence or cost to them for defrauding buyers in this underground economy [7, 21, 22, 24]. The quality of data sold within stolen data markets is also subject to information asymmetry as rippers can cheat buyers by either not delivering products or selling products of little or no quality. Illegitimate vendors can profit from these exchanges, though their buyers cannot.

The processes of the market also operate in a seller's favour, as they require prospective buyers to send payments first, and wait for data or services to be provided, typically within a 24–48 h period [7, 11, 21, 35, 50]. The buyer cannot determine the quality of the data purchased until they receive it, which in the case of rippers may either be invalid, non-functional or completely absent due to non-delivery [11, 21, 22, 35, 50]. There are no formal dispute resolution mechanisms that can be used by actors within the markets due to the inherently illegal nature of these transactions and products sold, creating opportunities for rippers to post false advertisements for products [21, 22, 50].

To minimize the risk of harm, forums provide informal mechanisms that encourage trust between participants and sanction less reputable actors [16, 21, 22, 50]. Many markets encourage buyers to publicly post feedback on their experience with a vendor, both positive and negative, in order to establish the credibility of a seller and promote trust between market actors generally [21, 22, 35, 50]. Since transactions occur outside of the forum, the ability to provide positive comments about a seller and their services appear to increase their potential share of the market, while those with negative feedback (regarding rippers) may eventually be ostracized and driven out of the community [16, 21, 35].

Characteristics of markets that may affect feedback

Feedback is designed to be a signal by a buyer, which can be received and interpreted by other buyers as an indication of the trustworthiness of a seller. Genuine sellers can expect to be

rewarded from the sale of quality products, with a positive reputation bringing in more frequent sales, and satisfied buyers resulting in repeat customers. Once a buyer has made a successful payment and profited from the purchase, they will trust the same seller in the next transaction [32]. This experiential learning process of product quality creates the trust between the parties. If there is a compromise of quality, negative feedback will go against the seller's reputation and affect the vendor. However, untrustworthy sellers who attempt to mimic their reliable counterparts also operate in stolen data markets. Rippers may generate false accounts in order to generate fake positive feedback, or leave negative feedback for their competitors (e.g. [24]). Moderators are tasked with policing such matters, and in some instances feedback cannot be left without first investing time and interacting in the marketplace [24].

The overall availability of information about both buyers and sellers in online markets has been referred to as 'transparency paradox' [48]. Researchers examining illicit drug cryptomarkets operating on the Tor network argue that the online nature of the exchanges increases the need for measures to anonymize the participants' location and real identity, while at the same time limiting the risks for participants to purchase impure products or experience physical violence [3, 48].

Though stolen data markets share some common elements with cryptomarkets, including the lack of face-to-face interactions and physical violence, there may be a greater risk of being defrauded or the data an individual purchases are outdated and unusable [22]. This may be a function of the nature of the open web, as individuals may be more readily able to access content hosted on forums and websites that can be identified via search engines like Google. Unscrupulous actors may be able to easily gain access to more public markets and prey upon those who are unfamiliar with its processes [11]. In fact, [6] show that outsiders can penetrate forums and post false feedback that disrupts the flow of information and sow distrust that essentially breaks down the relationship between market actors. Thus, more transparent information does not result in better signalling, as rippers may take advantage of this information availability to defraud the buyers.

According to signalling theory, it is in the interest of genuine sellers and buyers to produce signals that are relatively cheap to emit, yet costly to mimic [8]. There is also an incentive for rippers to mimic the signals used by legitimate sellers in order to draw in customers. There are differences in the types of signals that a ripper may be able to employ [9]. According to signalling theory, weak signals are those that are inexpensive for rippers to mimic. For example, simply stating that one is trustworthy comes at no further cost to the seller. Persuading signals are those that are relatively cheap to the genuine seller, but come at a cost to the ripper [9]. For example, some forums have hierarchical structures where vendors achieve certain levels or statuses based on the days of activity and number of public messages [6]. A legitimate vendor will have to invest time in a forum to establish a reputation (e.g. [35]) while a ripper would also need to invest time and then risk losing his status because of ripping.

The competing signals of rippers and genuine sellers broadcasted at the same time creates complexity for the recipients, who must determine their accuracy. The balance between accurate signals and the ability to interpret signals successfully creates three cost conditions [8]. The first is the equilibrium condition, in which there is a clear separation of genuine signallers who successfully signal their legitimacy relative to untrustworthy actors. The second is the uninformative condition, in which the signals produced by both those who can be trusted and those who cannot are unable to be differentiated by recipients. The third condition is semi-sorting, an intermediate condition, in which trustworthy actors signal their legitimacy while others are able to mimic these signals. Recipients in these instances are able to process some of these signals accurately based on the quality of a signal, though not all signals can be successfully interpreted [8]. The less informative the cost condition, the greater the information asymmetry.

Limited research has examined the issue of signals of trust within stolen data markets using samples from single forums [5, 6]. The findings suggest that positive and negative evaluations, the number of ads posted by a seller and the price for data influence the likelihood that buyers and sellers could form relational ties. These studies provide potential direction for research on trust within underground markets for data, though they are based on single forum samples which do not take into account the spectrum of markets in operation across the larger universe of markets generally [22]. In addition, they did not distinguish sellers from buyers, but examined all participants in a single marketplace. This is a key limitation as sellers and buyers may send and receive different signals. Each transaction originates with the advertisement of products, leading sellers to signal to prospective buyers their trustworthiness via different clues in their posts.

As a result, advertisements are the salient signal presented by sellers that are interpreted by other market actors, and serve as the basis for buyers to produce their own signals to the rest of the marketplace as to the trustworthy nature of the vendor. The larger body of research on stolen data markets provides insights regarding the factors within an advertisement, the sellers' practices, and the forum that may serve as signals of trust for the rest of the market. Specifically, the presence of negative feedback posted about a seller may serve as a clear signal to others within the market that an individual is not reliable. Buyers may opt to purchase from vendors with fewer instances of negative feedback because they appear more reliable and minimize the potential for economic harm (e.g. [22]). Negative feedback may be a signal that sellers are unable to directly influence or mimic, making it a costlier signal that is dependent entirely on the participation of genuine sellers.

The status of a seller may also have some association with their overall level of trust. Many forums maintain a user-based naming hierarchy to differentiate between new and established members (e.g. [6, 13]). Users may progress to a higher level based on their time spent in the forum, the number of posts and sales made, and their general reputation within the site [6]. Forums also have administrators or moderators who can participate and adjudicate the disputes on a forum as well as warned or banned users for bad behaviour. Since rank may be an indication of an individual's level of trust, risk-averse buyers may seek out vendors who have been on a forum for some time because they can identify the seller's post history and determine their reputation [6, 35].

The length of time that an individual spends within a site and the total number of posts they make are also pertinent signals of trust [6, 35]. The more time an ad has been available for individuals within the market to view, the more it serves as a marker for others to evaluate the reputation of a user on the basis of publicly posted feedback. Similarly, the number of posts a user makes on the forum indicates their willingness to interact with other users and build trust and a reputation. Such a signal is, however, time-consuming for a ripper to falsify and may serve as a valuable data point affecting trust between market actors [6].

The advertised price of a product is also an immediate point of information about a vendor that could be interpreted by prospective buyers. Such information allows buyers to potentially select a seller

on the basis of their pricing structure alone, due to the economic incentive to gain a large return on investment (e.g. [22]). Decary-Hetu and Laferriere [6] found that users who indicated prices for products in their ads had a lower probability of forming negotiating ties. At the same time, the price for information may be readily falsified by a vendor in order to attract unsuspecting buyers. The desire to pay very little for data may be overwhelming for some (e.g. [11]) while others may be willing to pay a higher price if it guarantees functional products and a greater likelihood of profit [11, 20]. Thus, the price of data may create more noise than signals within stolen data markets.

The use of various payment methods used to exchange money between the actors may also be a signal of trust. Many sellers accept digital currencies of some type, and a small proportion also use real-world payments through Money Gram or Western Union as these are established services for the transfer of hard currency transnationally [21, 35]. Funds transferred through Western Union are guaranteed, making it a more attractive payment mechanism in order to obtain cash. The use of physical payment systems are, however, more risky as individuals must arrive to a wire transfer location and provide identification in order to obtain their currency (e.g. [21, 24]). These conditions may make Western Union a preferred payment mechanism for vendors, but a potential signal for buyers of the inability to trust a seller.

The use of escrow payments may also be a signal used by sellers to guarantee a successful transaction (see [16], Wehinger, 2012). Escrow services are designed to limit buyers' risks because this form of payment operates through the use of an intermediary who holds funds on behalf of a buyer (e.g. [21, 22]). The seller must provide whatever products or services were negotiated in working order to the buyer, who then allows the escrow agent to release the funds to the seller. Typically, a forum selects a single individual to serve as an escrow agent, which is a position of trust in the market. While accepting escrow payments should increase a vendor's perceived trust, they are not required to accept escrow payments by anyone within the forum (e.g. [20]). Sellers could use falsely claim to use escrow, but never actually follow through on that claim, making it an easy signal to fake by untrustworthy actors.

The level of customer service advertised by sellers may be another signal of trust and reliability. Some vendors promote their use of dedicated customer support lines via ICQ and email to answer questions posed by buyers, facilitate purchases and demonstrate their willingness to satisfy customer needs (e.g. [7, 16, 21, 24]). Research suggests that buyers are more likely to praise sellers that maintain rapid and frequent contact with their customer base [16, 21], and thus the use of dedicated customer support services may be a valuable signal of trust that may not be readily falsified by rippers.

In much the same way, sellers may have their products reviewed by a moderator or tester in order to help validate their reputation. Some forums provide product-testing services, whereby an individual designated by the forum obtains a sample of products from a vendor in order to assure the quality of their goods [15, 21, 22, 24]. The tester posts a public review of their product or service to validate any claims the seller made. Some forums verify a seller's reputation through this process, which buyers can observe as a signal of trust and reliability [21, 22]. This practice appears to fit within signalling theory, as the seemingly wasteful action of providing a sample of data with monetary value at no cost for the purpose of verification may actually be useful to the seller [8]. At the same time, forums do not consistently have product testers available, or do not mandate that a seller have their products formally reviewed. Unscrupulous vendors may, therefore, claim they are willing to have their products tested but know they will not be required to do so. This may render product testing to be a signal that can be falsified by vendors within this market.

The type of products sold within data markets may also broadcast their own signals regarding the trustworthy nature of a vendor. The most prevalent product sold across most data markets are dumps, or debit and credit card account numbers, followed by CVVs which include a credit card number and the three-digit Credit Verification Value on the back of the card (see [7, 20, 21, 51]). Rippers may target inexperienced buyers by creating fictitious advertisements for these common products that appear competitively priced compared to legitimate vendors in the market (see [11]).

Similarly, the language used by market actors may serve as a signal of trust. The predominant language used in forums may directly impact the ability of individuals to participate, as those who are not fluent may be unable to effectively communicate. Evidence suggests that trustworthy sellers and reliable products are offered in markets where actors communicate with one another in Russian (e.g. [20, 46]). There is also a relationship between the language used by market actors and the advertised price of data [20]. This may stem from the fact that Russian-speaking nations have difficult extradition relationships with the USA and other European nations, thereby decreasing the risk of detection and prosecution for offenders [4]. Vendors who are unable to communicate in the same language as genuine sellers may be easily identified by others, making the language of a market a difficult signal to mimic. Though individuals may be able to partially mimic foreign language knowledge in forums through the use of machine translation programmes, their inability to appropriately use jargon and slang effectively will lead them to stand out from native speakers [14]. Thus, the language used in advertisements may serve as a costly signal of trust to differentiate legitimate vendors from rippers.

Research design

Taken as a whole, there is evidence that advertisements in stolen data markets produce signals which buyers must interpret in order to successfully complete a transaction. The preponderance of information about participants in online forums makes it possible to know a great deal about an actor [48]. At the same time, there is a negative selection problem as untrustworthy actors are able to cheat customers by mimicking signals produced by genuine sellers and drive legitimate vendors from the market. Buyers should have an inclination to discount all advertisements since they have limited recourse to enforce the terms of any transaction on their end [11, 21, 22, 24].

In order to identify the signals that may have the greatest value in interpreting a seller's reputation, this study uses a sample of Russian and English-language forums to examine the relationship between positive feedback (signals by buyers indicating the worth of advertisements) and signals used within advertisements, as well as the practices of the vendor, and the larger language preferences of the markets., characteristics and activity of vendors, price deviations, the purchase method, the provision of customer support and product testing, types of products sold and the language used.

We test the following model to account for signalling mechanisms within stolen data markets:

$$S_i = f(neg, user\_title, date, posts, price\ deviation, payment, service, product, language)$$

In this study, we use feedback comments as a signal left by buyers to indicate the worth of a product/transaction outcome. We

hypothesise that positive feedback comments (f) are a function of the presence of negative feedback, user titles (new, regular and senior), price deviations on the forum, the payment method (e.g. Western Union), customer service, product type (dumps) and the official language used by forum participants. We focus on advertisements for dumps and CVV data as they are the most frequently sold products in our sample.

Data

To examine the role of signalling in stolen data markets, we use posted advertisements and their corresponding feedback from a sample of six web forums where criminals and hackers buy, sell, and trade stolen financial and personal information. These forums act as online discussion groups where individuals can post advertisements, as well as present issues or discuss problems, and serve as important sources of data for researchers [19]. Each forum is composed of threads, which are a series of posts that centre on a specific topic under a forum's general heading. Threads begin when a registered user creates a post within a forum, asking a question or making a statement, and others respond with posts of their own. The content of threads provides information on the practices of market actors [21, 33, 35].

The initial sample of 13 forums was developed via a snowball sampling procedure (see [13, 14, 21]). Three English language forums were identified through Google.com using common terms in stolen data markets, including 'carding dump purchase sale CVV' [21, 35]. One of these sites was a sub-forum of a larger Russian language forum. After exploring the content of threads from these sites, three Russian language forums were identified via web links provided by forum users. Six additional forums were identified using the same processes to create a total of 10 Russian language sites and 3 English language forums. Of the 13 forums that were identified, 6 were used for this analysis (2 Russian and 4 English language). The criterion for inclusion was that they contained at least five advertisements for dumps and CVV data.

Of the six forums sampled, four were publicly accessible, in that the entire site could be accessed by anyone in the general public. The two remaining sites required that an individual create a registered user account within the site in order to access the content of the sub-forums related to data sales. Registration-restricted forums are thought to differ from that of publicly accessible forums because they add a layer of insularity and protection from outsiders and the general public [14, 34]. Registration systems allow anyone to join by registering a username and password account with the forum. This is not as exclusive or secure as invitation only forums that completely exclude outsiders from access, though registration eliminates the potential for threads to be captured by search engines or identified easily by general public [14, 34]. To capture the forum content across all sites, usernames were created for each forum, but to reduce the potential for contamination there was no interaction with other registered participants [14, 34].

All threads posted from carding or sales related sub-forums were captured to develop a substantive volume of posts. A certified Russian translator with substantive experience with technological jargon and forum communications translated the Russian language content from all forums. Due to the availability of the translator, convenience samples of 25 threads from each Russian forum were selected to capture the most recently posted items for sale in each site. Additional samples of threads were translated from five forums, particularly those that had active posting, to better assess the practices of actors and the network connectivity of participants. Repeat threads were excluded from analysis, but translated to ensure reliability of content. The research team also oversampled threads from the English language forums to capture any variations in the nature of these markets and their organizational composition. This strategy provided a mix of user populations and duration over time, while at the same time creating a relatively matched sample of posts between English and Russian language threads across the forums (see Table 1 for detail).

Table 1. Forum data summary statistics

Number of forums	6
Number of threads per forum	6-590
Total number of threads	1889
Total number of posts	9117
Russian language only forums	2
English language only forums	4
Time frame	5/9/2007-2/25/2012

The unit of analysis of this study is the advertisements for two specific products: (i) dumps, referring to debit or credit card account numbers and personal information, and (ii) CVVs, which include a credit or debit card account number plus the three-digit Credit Verification Value on the signature line of the card used in order to make purchases online or over the phone. These data types were the most prominent in the sample, and had differing amounts of feedback. Dumps were the most common product advertised in this sample of threads ($n = 5732$; 55.15%), in keeping with research from both IRC [7] and forum-based data markets [21]. CVV data was the second most common product ($n = 4481$, 43.12%). Each advertisement was coded using content analysis techniques (see [14, 20]) to create quantitative variables relating to positive and negative feedback (the dependent variables), and the potential signals used within advertisements (the independent variables).

Dependent variables

The dependent variable was the total number of posts featuring positive buyer feedback. Any comments indicating that a buyer had engaged in a transaction and were satisfied with the outcome (including 'good data', 'reasonable price' and 'data worked') were coded as positive feedback. Posts where the user indicated they attempted to complete a transaction but either did not receive their purchase (e.g. 'guy is ripper', 'where is my data'), or the data were non-functional, were coded as negative feedback.

Approximately 78% of all advertisements received no specific feedback. These zeros may stem from two different processes driving forum interactions (Table 2). First, some advertisements prompted interested parties to ask questions of the seller in an attempt to clarify, or strengthen, signals, but had no evidence of observable transaction-based feedback. These exchanges may have led to privately-completed transactions that did not produce observable feedback (e.g. [35]). Secondly, some advertisements did not generate any posts from forum participants, or true zeros due to the absence of feedback. This may be because the advertisement was new, based on a product with no market interest, or was clearly falsified. The zero-inflated regression models allow us to account for these processes by separately estimating the probability of a post receiving zero feedback (logit portion of the model), and the actual regression model that predicts what factors will lead to higher number of positive feedback posts.

Independent variables

Multiple social and market-related variables were coded from the language of each advertisement to consider how they serve as signals

Table 2. Descriptive statistics, total number of observations is 9418 All forums selling dumps and CVV ($n = 6$)

Variable	Mean	Standard Deviation	Min	Max
Positive comments	0.580	2.662	0	32
Negative comments	0.328	1.202	0	15
Senior members	0.034	0.180	0	1
Regular members	0.311	0.463	0	1
New members	0.524	0.499	0	1
Number of days	359.236	277.365	0	688
Number of posts	4.971	11.609	1	68
Price deviation	−0.038	0.671	−0.557	13.981
Western Union	0.253	0.435	0	1
Escrow	0.037	0.188	0	1
Customer support	0.080	0.271	0	1
Product tested	0.025	0.157	0	1
Dumps	0.542	0.498	0	1
Russian	0.023	0.150	0	1

of trust for the seller. Since advertisements may generate both positive and negative feedback, we coded for the presence of 'negative feedback' in the same posts that received positive feedback. Posts where the user indicated they attempted to complete a transaction but either did not receive their purchase (e.g. 'guy is ripper', 'where is my data'), or the data were non-functional, were coded as negative feedback and used as a continuous variable in this model.

To assess the relationship between a forum assigned participation ranking and reputation, three binary variables were created to contain the most common ranks assigned across the forums in this sample. Since the titles used varied by forum, these categories represent users on the basis of their general classification. 'Senior members' were individuals whose user titles containing senior, VIP, experienced or other indications of the highest rank in their title. 'Regular members' were coded on the basis of a lack of superlatives in their title, including: vendors, users and members (across 10 forums). 'New members' were coded on the basis of terms such as noob, new, fresh or any other indicator of a lowest level of a hierarchy. Due to the missing data on some variables (e.g. administrators not participating in posts with any positive feedback), we were able to control only for these three categories. The comparative category in this case becomes users without specific titles, unverified, banned and guest members.

A variable was created to examine the relationship between the duration of an ad and its role as a signal of trust. We calculated the length of time an advertisement may have been observed by participants on the basis of the number of days the post was available on the forum from the first date of posts observed in the sample of threads obtained (number of days). An additional variable, 'number of posts', was calculated to measure the total number of posts a user made throughout the life of the forum during our observation period as a control variable for user activity.

To examine the relationship between price and signals of trust, the variable 'price deviation' was measured by calculating the z-scores for the extent to which an advertised price deviated from the forum's mean price. Though users may not explicitly calculate the extent to which an advertised price differs from the average within a given market, they may associate a product's price to the other prices they have seen in general (see [16, 21]). The mean price z-score is around zero, indicating that indeed, the majority of users have cited prices for dumps/CVV products very close to the forum's mean. None of the users advertised prices that were lower than one standard deviation below a forum's mean across all forums in our data, while some have cited higher prices. This may be an indication of price pegging with the need for further negotiations that took place outside of the forum (e.g. [21]).

Two binary measures (0 = no; 1 = yes) were created to assess the potential influence of two payment mechanisms that may affect seller reputations: (i) 'Western Union' and (ii) 'Escrow' payments. Our comparative category is not indicating a specified payment method in the advertisement. Western Union payments were included due to the potential risk they present for buyers due to the need to physically transfer funds to the seller. Escrow payments were included because they are thought to provide a degree of trust between participants, and may also be an easily falsified signal by unscrupulous vendors.

Two binary measures (0 = no; 1 = yes) for customer service were also included to understand any influence they may have on the positive comments. First, 'customer support' was measured based on whether sellers indicated that they operated specialized customer support lines through ICQ or email to aid customers in case of questions or issues, or provided assistance for buyers after a purchase in order to facilitate the use of purchased data. The second measure, 'product tested', was based on indicators that forum moderators had verified the seller, or tested their products. Additionally, a dummy variable (0 = no; 1 = yes) was created for 'dumps' to control for their large presence in all ads and its relationship to any feedback provided. The 'primary language' used in advertisements was coded as a binary measure (0–Eng; 1–Rus).

Methods

To assess the factors affecting signalling of trust among buyers, we ran a model for positive feedback using a zero-inflated Poisson regression. The zero-inflated models simultaneously estimate the logit model to differentiate two mechanisms that may explain the abundance of zeros in the data, along with a Poisson model, to identify factors affecting the actual count of positive feedback received. This statistical test enabled comparisons between advertisement signals that entice buyers and lead to feedback and those which do not. Additionally, the count model allows for the identification of signals leading to greater instances of positive feedback. The use of a Poisson regression model also eliminates the need for statistical transformation of skewed variables that would otherwise be required through OLS and other regression techniques. Due to the clustered nature of the data within forums, there is some possibility that effects exist due to the non-independence of sellers within, but not necessarily across forums. These issues can result in biases that produce errors in larger statistical models. A common technique to control for such effects is through the use of hierarchical logistical modelling (HLM), though at least 10 units are preferred as the basis to conduct such an analysis. Since there were only six forums that fit the criteria included in the final sample, the models were conducted in Stata software using a Taylor series to account for the sample design and provide accurate testing of the coefficients and standard errors [31, 38].

Findings

The zero-inflated Poisson regression results for positive feedback are shown in Table 3. Overall, the model is significant at 0.001 level, and the Vuong test for zero-inflated model is significant indicating the need for the selected model. The first column shows zero-inflated Poisson regression coefficients, followed by standard errors

Table 3. Survey zero-inflated Poisson regression estimation results for of positive comments

Variables:	Number of positive comments		
	B	St. error	e^B
Model 1: count equation: factor change in expected count for those not always 0			
Negative comments	0.237**	0.037	1.267
Senior members	1.630	0.868	5.106
Regular members	3.069*	0.977	21.519
New members	0.685	0.823	1.985
Number of days	−0.004*	0.002	0.996
Number of posts	0.037**	0.007	1.038
Price deviation	−0.539	0.326	0.583
Western Union	−2.664**	0.532	0.070
Escrow	−0.748	0.412	0.473
Customer support	0.880*	0.356	2.411
Product tested	1.067	0.938	2.906
Dumps	−1.053	0.626	0.349
Russian	2.886**	0.825	17.919
Intercept	−0.441	0.693	0.644
Strata	6		
PSUs	225		
Observations	9117		
F	10.29**		

Note: *significant at $P < 0.05$ level; **significant at $P < 0.001$ level, ^$P = 0.053$. McFadden's Adjusted R^2 is 0.501, but this number can only be calculated for zero-inflated model not adjusted for survey data, hence should be interpreted with caution.

and incidence rate ratios. The Poisson regression models the log of expected count as a function of predictor variables, making the interpretation of coefficients difficult. The Incidence Rate Ratios (IRR) make the interpretation of the Poisson regression coefficients easier.

The results suggest that sellers with negative feedback were more likely to receive positive feedback. The IRR is 1.266, indicating that users with negative feedback in the same posts are 26.6% more likely to receive positive feedback as well. This is a complex finding, as a legitimate vendor may generate both positive and negative feedback if their customer base reaches a certain size. At the same time, this may reflect a trend of rippers posting positive comments in an attempt to obscure signalling from negative feedback.

Vendors with a regular rank assigned by the forum were 21 times more likely to receive positive feedback than other forum participants. The only other variable that provides such high odds is language which also increases the odds of more positive comments by nearly 17 times. Both of these variables are dichotomous, requiring these high incidence rates to be viewed with caution. At the same time, it may be an indication of a qualitative jump in the ability of vendors with somewhat regular participation in a forum and communicating in Russian language markets to be more trustworthy overall (see [6, 11, 20, 50]).

The date of posts was also significant, indicating that more recent posts accumulated less positive feedback than older posts. Similarly, users with a larger number of posts were more likely to receive positive feedback. Such signals are extremely difficult for rippers to falsify and provides a clear marker for identifying trustworthy vendors. This also supports prior research that time spent on forums increases individual network connectivity within the market [6, 35].

Additionally, the more a product's price deviated above the forum's mean price, the less positive feedback that vendor will receive. Though some argue that the lower a product is priced, the less likely that product is functional (e.g. [11]) there has been little research assessing the ceiling for pricing. Advertising data at extremely high prices can be an easily falsified signal by vendors, thus using a more reasonable price point may make that vendor seem more in touch with the market's price tolerances. In fact, this finding provides partial support for the assertion that markets have a general informally accepted price range for a product [16].

Interestingly, vendors who accepted Western Union in their ads were less likely to receive positive feedback (change of factor 0.069 or decrease of about 93%). It is not clear if this is a direct function of the risk buyers face from using this payment type, leading to a need for greater research on this issue. Escrow payments were also non-significant, which may be a reflection of the optional nature of this payment mechanism across the market for data (e.g. [21, 22]). Finally, vendors whose ads included customer service lines and support were more than 2.4 times likely to receive positive feedback, reinforcing the prominence that service plays within data markets (e.g. [20, 21]).

The results of the binary equation estimating the likelihood of not receiving any feedback (feedback = zero group) are not presented here as this model did not perform as well our regression portion. The only significant variable was the use of escrow services, suggesting that the larger range of advertisement-specific and structural factors within the market are unable to differentiate between signals and noise. As a result, there may be too much information available in the text of advertisements for receivers to readily interpret signals provided by vendors in data markets. This notion contradicts some arguments related to the value of transparency in online markets [3, 48] suggesting additional research is required to better understand this issue.

Discussion and conclusion

This study attempted to understand the factors associated with signals of trust observed in advertisements by vendors within stolen data markets by examining the positive feedback posted by individuals who purchased products from vendors. Using Gambetta's signalling theory [8, 9] with a sample of posts from active stolen data forums, the results suggest that there are some clear signals in advertisements that demonstrated seller trust. Examining the count

models for feedback conducted through zero-inflated Poisson regressions indicated that sellers who also received negative feedback received more positive feedback (see [6]). It may be an indication that the quality of the feedback is an important factor on the market, leading participants to seek additional information in order to counteract negative comments received. This may also be a symptom of false feedback posted by rippers in an attempt to shore up their reputation and obfuscate their actual practices. Thus, future research is needed to clarify the proportion of all sellers who receive negative feedback to identify a baseline of customer complaints across stolen data markets.

Signals such as user status indicate the importance of long-term participation in a market in order to identify their potential level of trust. Regular users were more likely to receive positive feedback compared to the other ranks, demonstrating that achieving a certain level within a forum is a more costly signal to mimic compared to others. The linguistic structure of the forum was also a signal of trust, as Russian language forums had fewer negative feedback reported. It is unclear if this is a result of greater insularity and trust in vendors within Russian language markets, or an artefact of the sample of threads acquired for this analysis. Additional research is needed using larger samples of data from forums communicating in multiple languages to understand the relationship between language and trust in cybercrime markets generally (see also [16, 22]).

Several significant variables in this model suggest that stolen data markets comprise a semi-sorted signalling condition, as some signals can be readily interpreted while others are difficult to discern by receivers [8]. For instance, the use of customer service contact points was significant, but any individual could establish an email or ICQ account at no cost and then post it in the language of an advertisement. Similarly, vendors who accepted Western Union payments were less likely to receive positive feedback. Since this language can be inserted into any advertisement, the quantity and transparency of information posted by vendors does not translate into better or more clear signals.

Since stolen data markets operate on the open web, it is possible that they may differ from markets on the dark web and in the real world in that they may provide more opportunities for risk of economic harm. In drug markets (e.g. [26, 28]) and open-air sex markets [18], actors can observe the foreground and situational cues of sellers and prospective clients and use this information to structure their behaviour accordingly. The faceless nature of online spaces increases the value of correctly interpreting advertising-based signals of risk, and discerning trustworthy sellers (e.g. [7, 21, 22]). Though similar issues are present in dark web drug markets, there appears to be less risk of economic harm from vendors seeking to cheat buyers [3, 48]. Thus, the findings of this study suggest that there are only a few factors that serve as strong signals of trust that cannot be manipulated by vendors, such as language and the total number of posts made.

Recognizing the pertinent signals of trust within stolen data markets provides potential applications for market disruption. The findings of this analysis suggest that the conditions within stolen data markets may be manipulated and destabilized through the introduction of information asymmetry. It may be possible to complicate the process of interpreting signals by flooding the market with multiple false posts for products, as well as feedback for sellers. This technique, referred to as a Sybil attack, may be effective in creating too many false signals and increasing the difficulty in determining who is a reliable vendor [6, 7].

In addition to message quantity, the content of each message may also be altered to hinder the ability of individuals to interpret signals within the market. For example, this analysis found sellers who indicated they accepted payments via Western Union was associated with less satisfied customers. By altering the content of Sybil messaging to correspond with law enforcement interventions, they may create further uncertainty among market participants. In turn, Sybil attacks may enable the efficient disruption of markets without the need for arrests or prosecution of offenders. However, future research will need to look at how this signalling mechanism affects future market transactions.

At the same time, the use of Sybil attacks may only be effective against poorly organized forums or for a short period of time in all forums generally [7, 20]. Markets with moderators and managers that directly interact with participants may delete posts at their discretion (e.g. [22, 24]). Better organized forums may be able to discern false posts more efficiently and subvert intervention attempts through Sybil campaigns. In addition, a campaign of this sort is relatively public in nature and may draw attention to the attempt by law enforcement which may lead market actors to harden themselves against external involvement. Thus, there is a need for law enforcement interventions that combine various traditional and novel disruption approaches in order to better affect the market for stolen data.

Though this study explored an under-examined form of crime, there are several limitations that must be addressed. First, the data for this study were drawn primarily from sites operating on the open web with limited restrictions in order to access posts. The findings may not be applicable to more hidden groups, which comprise a deeper portion of the underground economy for stolen data. This includes sites that require participants to be vetted by existing members and pay for access to the forums (e.g. [22]). Additionally, there are a number of markets operating as hidden services, which can only be accessed through anonymity networks, such as Tor [2]. Sampling these forums is complicated by the fact that few search engines, even operating through Tor, index their content. Expanding the sample to include more hidden markets can improve our knowledge of the relationships between advertising signals and feedback. In turn, we may better identify the association between the structure of a forum and the levels of trust between participants.

Second, it is possible that law enforcement or computer security researchers may have generated posts within the forums. There is no way to determine which posts were made by actual vendors and buyers, and those made by individuals posing as criminals in order to blend into, and potentially disrupt, the community. Future research is needed which directly connects researchers with practitioners to better understand and evaluate disruption activities and separate market actor identities from those used by undercover operatives. In turn, this may improve the validity of the models used to identify signals within the market, and the ways that receivers correctly interpret this information.

Finally, this study did not examine signals such as the moniker used by sellers [32], the length of time they had been in a forum or how they interact with each other as prospective signals of trust [5]. Instead, we focused on the language used in advertisements for stolen data to identify how sellers signal their trust and worth. It is possible that monikers used by sellers may be easily mimicked, such as using a similar name, or even taken over through an account hijacking, in order to affect the signals made within the market. Thus, further research is needed in order to consider how other signals may directly affect trust within the market for stolen data. In turn, we may better understand how data markets are structured, as well as their relationship to other illicit markets online and offline.

Funding

This work was supported by the Department of Homeland Security (DHS) Science and Technology Directorate, Cyber Security Division (DHSSandT/

CSD) Broad Agency Announcement 11.02, the Government of Australia and SPAWAR Systems Center Pacific [N66001-13-C-0131 to Hutchings]; and the National Institute of Justice, Office of Justice Programs, US Department of Justice [2010-IJ-CX-1676, 2010, to Holt and Smirnova]. The opinions, findings and conclusions or recommendations expressed are those of the authors and do not reflect those of the aforementioned agencies.

References

1. Akerlof GA. The market for 'lemons': quality uncertainty and the market mechanism. *Quart J Econ* 1970; **84**: 488–500.
2. Barratt MJ. Silk road: eBay for drugs. *Addiction* 2012;**107**: 683–83.
3. Barratt MJ, Ferris JA, Winstock AR. Safer scoring? Cryptomarkets, social supply and drug market violence. *Int J Drug Policy* 2016;**35**: 24–31.
4. Brenner SW. Defining cybercrime: a review of federal and state Law. In: RD Clifford (ed.) *Cybercrime: The Investigation, Prosecution, and Defense of a Computer-Related Crime*, 3rd edn. Raleigh, NC: Carolina Academic Press, 2011, 115–40.
5. Décary-Héto D, Leppänen A. Criminals and signals: an assessment of criminal performance in the carding underworld. *Secur J* 2013;**31**: 1–19.
6. Décary-Hétu D, Laferrière D. Discrediting vendors in online criminal markets. In *Disrupting Criminal Networks: Network Analysis in Crime Prevention*. Boulder, CO: Lynne Rienner, 2015, 129–52.
7. Franklin J *et al*. An inquiry into the nature and causes of the wealth of Internet miscreants. In: *ACM Conference on Computer and Communications Security (CCS)*. Alexandria, VA: ACM, 2007, 275–88.
8. Gambetta D. Signalling. In: P Hedström, P Bearman (eds), *The Oxford Handbook of Analytical Sociology*. New York: Oxford University Press Inc, 2009, 168–94.
9. Gambetta D. *Codes of the Underworld: How Criminals Communicate*. Princeton: Princeton University Press, 2009.
10. Hegghammer T. The recruiter's dilemma: signalling and rebel recruitment tactics. *J Peace Res* 2012;**50**: 3–16.
11. Herley C, Florêncio D. Nobody sells gold for the price of silver: dishonesty, uncertainty and the underground economy. In: T Moore, D Pym, C Ioannidis (eds), *Economics of Information Security and Privacy*. Boston, MA: Springer, 2010, 33–53.
12. Higgins KJ. Target, Neiman Marcus Data breaches tip of the iceberg. *Dark Reading*, 20 July, 2014. http://www.darkreading.com/attacks-breaches/target-neiman-marcus-data-breaches-tip-o/240165363 (20 August 2014, date last accessed).
13. Holt TJ. Subcultural evolution? Examining the influence of on- and off-line experiences on deviant subcultures. *Deviant Behav* 2007;**28**: 171–98.
14. Holt TJ. Exploring strategies for qualitative criminological and criminal justice inquiry using online data. *J Crim Just Educ* 2010;**21**: 466–87.
15. Holt TJ. Examining the forces shaping cybercrime markets online. Soc Sci Computer Rev 2013; **31**: 165–77.
16. Holt TJ. Exploring the social organization and structure of stolen data markets. *Global Crime* 2013; **14**: 155–74.
17. Holt TJ, Blevins KR, Kuhns JB. Examining the displacement practices of johns with online data. J Crim Just 2008;**36**: 522–28.
18. Holt TJ, Blevins KR, Kuhns JB. Examining diffusion and arrest avoidance practices among johns. *Crime Delinquency* 2014; **60**: 261–83.
19. Holt TJ, Bossler AM. An assessment of the current state of cybercrime scholarship. *Deviant Behav* 2014; **35**: 20–40.
20. Holt TJ, Chua Y-T, Holt O. An exploration of the factors affecting the advertised price for stolen data. In: *eCrime Researchers Summit (eCRS), 2013*. IEEE, 2013, 1–10.
21. Holt TJ, Lampke E. Exploring stolen data markets online: products and market forces. *Crim Just Stud* 2010; **23**: 33–50.
22. Holt TJ *et al*. Examining the risk reduction strategies of actors in online criminal markets. *Global Crime* 2015;**16**: 81–103.
23. Holz T, Engelberth M, Freling F. *Learning More About the Underground Economy: A Case-Study of Keyloggers and Dropzones*. Berlin, Heidelberg: Springer, 2009.
24. Hutchings A, Holt TJ. A crime script analysis of the online stolen data market. *Br J Criminol* 2015; **55**: 596–614.
25. Jacobs BA. Undercover deception cues: a case of restrictive deterrence. *Criminology* 1993;**31**: 281–99.
26. Jacobs BA. Crack dealers and restrictive deterrence: identifying narcs. *Criminology* 1996;**34**: 409–31.
27. Jacobs BA. *Robbing Drug Dealers: Violence Beyond the Law*. New York: Aldine de Gruyter, 2000.
28. Jacobs BA. Crack dealers' apprehension avoidance techniques: a case of restrictive deterrence. *Just Quart* 1996; **13**: 359–81.
29. Johnson BD, Natarajan M. Strategies to avoid arrest: crack sellers' response to intensified policing. Am J Police 1995; **14**: 49–69.
30. Knowles GJ. Deception, detection, and evasion: a trade craft analysis of Honolulu, Hawaii's street crack-cocaine traffickers. *J Crim Just* 1999;**27**: 443–55.
31. Levy PS, Lemeshow S. Sampling of Populations: Methods and Applications, 3rd edn. New York: Wiley, 1999.
32. Lusthaus J. Trust in the world of cybercrime. *Global Crime* 2012;**13**: 71–94.
33. Mann D, Sutton M. NETCRIME more change in the Organization of Thieving. *Br J Criminol* 1998;**38**: 201–29.
34. Markham AN. Internet research. in D. Silverman (ed.). *Qualitative Research: Issues of Theory, Method, and Practice*, 3rd edn. Thousand Oaks, CA: SAGE Publications, 2011, 111–27.
35. Motoyama M *et al*. An analysis of underground forums. In: *2011 ACM SIGCOMM Conference on Internet Measurement*. Berlin, Germany: ACM, 2012, 71–80.
36. Peretti KK. Data breaches: what the underground world of carding reveals. *Santa Clara Computer High Tech L J* 2009;**25**: 375–413.
37. Ponemon Institute. *Cost of Data Breach Study: Global Analysis*. Traverse City, MI: IBM, 2014.
38. Rennison CM, Melde C. Gender and robbery: a national test. *Deviant Behav* 2014;**35**: 275–96.
39. Reuter P, Caulkins JP. Illegal 'lemons': Price dispersion in cocaine and heroin markets. Bull Narcotics 2004; **56**: 141–65.
40. Samani R, Paget F, Hart M. *Digital Laundry: An Analysis of Online Currencies, and their Use in Cybercrime*. Santa Clara: McAfee, 2013.
41. Sandberg S. The importance of culture for cannabis markets towards an economic sociology of illegal drug markets. *Br J Criminol* 2012;**52**: 1133–51.
42. Seals T. 2014 so far: The year of the data breach. *Infosecurity 2014*. http://www.infosecurity-magazine.com/news/2014-the-year-of-the-data-breach/ (20 July 2015, date last accessed).
43. Spence M. Job market signaling. *Quart J Econ* 1973;**87**: 355–74.
44. Suroweicki J. Why did criminals trust Liberty Reserve. *The New Yorker*, 31 May 2013. http://www.newyorker.com/online/blogs/newsdesk/2013/05/why-did-criminals-trust-liberty-reserve.html (11 June 2013, date last accessed).
45. Symantec Corporation. *Symantec Internet Security Threat Report, Vol. 17*. Symantec Corporation. http://www.symantec.com/content/en/us/enterprise/other_resources/b-istr_main_report_2011_21239364.en-us.pdf (25 May 2012, date last accessed).
46. Symantec Corporation. *Symantec Internet Security Threat Report, Vol. 19*. Symantec Corporation 2014. http://www.symantec.com/content/en/us/enterprise/other_resources/b-istr_main_report_v19_21291018.en-us.pdf (11 August 2015, date last accessed).
47. Topalli V, Wright R, Fornango R. Drug dealers, robbery and retaliation. Vulnerability, deterrence and the contagion of violence. *Br J Criminol* 2002;**42**: 337–51.
48. Tzanetakis M, Kamphausen G, Werse B *et al*. The transparency paradox. Building trust, resolving disputes and optimising logistics on conventional and online drugs markets. *Int J Drug Policy* 2016;**35**: 58–68.
49. VanNostrand L-M, Tewksbury R. The motives and mechanics of operating an illegal drug enterprise. *Deviant Behav* 1999;**20**: 57–83.
50. Wehinger F. *The Dark Net: Self-regulation Dynamics of Illegal Online Markets for Identities and Related Services*. Intelligence and Security Informatics Conference (EISIC), 2011 European, 2011.
51. Yip M, Webber C, Shadbolt N. Trust among cybercriminals? Carding forums, uncertainty and implications for policing. *Pol Soc* 2013;**23**: 516–39.