Then and Now: On The Maturity of the Cybercrime Markets The lesson that black-hat marketeers learned

Luca Allodi Student Member, IEEE,, Marco Corradin, Fabio Massacci Member, IEEE

Abstract-Cybercrime activities are supported by infrastructures and services originating from an underground economy. The current understanding of this phenomenon is that the cybercrime economy ought to be fraught with information asymmetry and adverse selection problems. They should make the effects that we observe every day impossible to sustain. In this paper we show that the market structure and design used by cyber criminals have evolved towards a market design that is similar to legitimate, thriving, on-line forum markets such as eBay. We illustrate this evolution by comparing the 'market regulatory mechanisms' of two underground forum markets: a failed market for credit cards and other illegal goods and another, extremely active marketplace for vulnerabilities, exploits, and cyber attacks in general. The comparison shows that cybercrime markets evolved from unruly, 'scam for scammers' market mechanisms to mature, regulated mechanisms that greatly favors trade efficiency.

Index Terms—Cybercrime, black markets, security economics

I. INTRODUCTION

Cybercrime is gaining more and more momentum as a source of threats for final users. Credit card, banking and financial frauds are continuously reported in the news and often studied in the literature [1], [2], and recent studies have uncovered a whole infrastructure of services that are available to cyber criminals to deploy their attacks [3], [4], [5]. Exploitation tools, automated redirection of user connections to arbitrary domains [6], and trading of new malware or vulnerabilities are only example of a multitude of *measured* effects of what is notoriously called "cybercrime". These infrastructures and services, on the other hand, must be sustained and provided by an underlying economy.

Market design is a problem of great interest in economics, as a successful market necessarily involves an equilibrium of forces that on one side encourages trading, and on the other discourages "cheaters". Obviously, a market where everybody cheats is not a sustainable market and is doomed to fail because nobody would eventually initiate a trade. Cybercrime markets represent, intuitively, a fascinating case study for these issues: they are run by criminals (who are not trustworthy by definition), are typically run on-line, and are to a degree anonymous. How can anonymous criminals trust other anonymous criminals in delivering the promised service or good after the payment has been issued? And even if the buyer gets 'something', how can she be sure that what she thinks she is buying is effectively what she will end up with? If a trade goes sour, a buyer cannot call the police to apprehend the scammer.

Florêncio et al. [7] showed that IRC cybercrime markets (Markets run through Internet Relay Chats) may be no different from the notorious market for lemons captured by Akerlof [8], where effectively the asymmetry of information between the seller and the buyer is such that "bad sellers" are incentivized in participating in the market to the point that it makes no sense for the "good sellers" to remain active. In Akerlof's case, a "bad seller" is a seller that trades 'lemons'. A 'lemon' is a defective car that is advertised as a good one. If the customer can not assess the quality of the car before buying it (e.g. because she knows little about cars), then she will buy the cheapest she can find on the market. Since 'lemons' are cheaper than good cars, 'good sellers' are ultimately forced out of the market. In Florencio et al.'s case, a 'lemon' was a credit card number with (allegedly) a certain amount of money ready to be used by the buyer. As shown in Akerlof's work, discerning 'good sellers' from 'bad sellers' is therefore a critical point of a market design. Florencio et al. clearly demonstrated that it is virtually impossible to do so in the IRC cybercrime markets.

Yet, empirical evidence from numerous studies shows that the attack tools traded in these markets do work [4], [3], [9], and the losses caused by cybercrime are real [10]. How can these observations be reconciled with the understanding that cybercrime markets *cannot* work? The explanation is that current markets are run under a different structure than IRC markets: rather than anonymous, free-to-join, unregulated communities of criminals, modern cybercrime markets are run as virtual forums [5], [4], [11], [12]. Forums provide an easy way for the community administrators to control the flow of users into the community and to enforce, through moderation, a number of rules that can be aimed - in a coherent market design structure - at mitigating the issues of information asymmetry [11]. Proper regulation is, therefore, the key to a successful market.

In this paper, we show how cyber criminals may have learned market design by analyzing two different cybercrime markets: the first, **Carders.de**, is a (failed) market in German for credit cards numbers and other illegal goods, whose database leaked in 2010. We are able to reproduce and analyze the market in its entirety and we show how the systematic failure of its regulatory mechanisms led to a market where rippers and 'legitimate users' are indistinguishable from each other. Secondly, we discuss the case of a functioning, segregated, on-line, underground community for cyber attacks in Russian. For the purpose of this discussion, we label this market **HackMarket.ru**. We do not disclose the real name of the market not to hinder future studies. We infiltrated this community and analyzed its rules and their enforcement. The comparison between the two markets is centered on the

Luca Allodi and Fabio Massacci are at the DISI Department at the University of Trento, Italy. emails: name.surname@unitn.it

Marco Corradin is a consultant for the Dutch Government. Part of this work has been carried during his MSc Thesis at the University of Twente, NL. email: mc.corradin@belastingdienst.nl

regulatory issues that arise in the virtual, anonymous and criminal IRC market first underlined in [7]. In an environment operated by criminals, 'law enforcement' by police officers and judges is clearly not possible. Differently from other online communities such as eBay, in the criminal market there is no clear authority that enforces regulation and incentivizes the 'good behavior' of users. In this respect, the analysis results for HackMarket.ru are in sharp contrast with those of Carders.de and clearly show *prima-facie* evidence that underground cybercrime communities can be mature (and functioning) market.

Section II discusses current relevant literature and sets the stage for the discussion of the paper. In Section III we formulate our working hypotheses to test for the proper enforcement of the market regulation and its reputation mechanisms. Section IV presents the Carders.de data and describes the market and the designed regulatory mechanisms. In Section V we provide an analysis of Carders.de with respect of our hypotheses. We then present in Section VI the second market, HackMarket.ru, and discuss its differences with the first market. Finally, Section VII provides a discussion of our findings, and Section VIII concludes the paper.

II. BACKGROUND & LITERATURE REVIEW

Current literature on underground markets can be clustered in two categories: studies that (indirectly) provide factual evidence of the workability of the underground markets, and studies that analyze the structure and economics of the markets.

Fact Finding. Efficiency is key for underground markets to increase workability. One way to achieve this has been shown by Grier et al. [3] in which they described the Exploit-as-a-Service (EaaS) model. In the EaaS model the cyber criminal can rent a service in which the contractor provides a full service that supports all the necessities to infect computers for the buyer. More efficient markets have been studied by Sood and Enbody [13]. They have shown evidence that also the Crimeware-as-a-Service (CaaS) model is present in underground markets. Where EaaS only provides a full service for exploitation and infection of machines, CaaS provides a full service that provides the cyber criminal with all the resources he/she may need. This means that the service offered contains all necessary tools and services to commit the cybercrime such as frameworks, settings, machine infections and identity masking. Another study on the quality of offered products in underground markets has been conducted by Allodi et al. [4]. By analyzing exploit kits they measured the resiliency and efficacy of cybercrime tools in delivering attacks. Numerous other studies analyzed the technical details behind these infection processes [6], [5] and the creation of botnets [14], [15]. A similar line of research also gave insights on the mechanics of spam [16] and diffusion of attacks [17]. Fallman et al. [18] and Yang et al. [19] discuss automated ways to probe online (possibly underground) markets and automatically extract information from the ongoing discussion.

Economics Studies. The annual internet security threat report by Symantec [20] published in 2013 estimated the value

of the goods offered throughout 2012 in underground markets at \$276 million. Vömel et al. [21] monitored an Internet Relay Chat (IRC) channel for credit card advertisements. Studies with a social approach towards the underground markets analyzed cyber criminals who operated in successful markets[22], [23]. They showed that criminals prefer trading in a more secured and hierarchical system to further increase trading efficiency and stability of the market. Given this increased need of a more structured hierarchy most markets moved towards forums, resulting into studies that try to infiltrate and analyze these forums [12], [4].

Still, running an efficient underground economy in which criminals trade goods and services with other criminals is not a trivial exercise. Florêncio et al. [7] showed that underground markets feature scammers who try to scam other members of the market. The market they studied was largely a 'market for lemons', in contrast with the efficient markets described by Yip et al.[22] and Zhao et al. [23]. The work of Florêncio et al. was a first step in identifying the mechanisms responsible for market failure:

- 1) Users could join the market freely and with an arbitrary identity. Feedback mechanisms (e.g. reputation) on the 'reliability' of the users are not effective.
- There is no history of transactions available, so it is impossible to look back at a users' trades or communityprovided feedbacks.
- 3) The community is largely unregulated and no assurance for the buyer or the seller exists that they are engaging with is a "legitimate" trader and not a scammer.

IRC markets are, however, an 'outdated' model of markets, for cybercrime or otherwise. Recent markets moved towards a forum-like environment [5], [11], which provides many advantages over the IRC model: first of all, users must register and are therefore assigned a unique ID. The forum structure provides a well-defined technological means for users to leave permanent and easily-searchable feedback on another user, and many forum platforms allow for the assignment of 'reputation points' to different users which may directly reflect a members' role in the community. Finally, a forum can be easily moderated and administered, meaning that *some* regulation of the market activities is possible. This makes 'forum markets' potentially different from the IRC markets that have been shown to be irremediably flawed.

III. HYPOTHESES ON FORUM MARKETS

Both Carders.de and HackMarket.ru are forum-based markets. They have administrators, moderators, users' registration procedures, reputation mechanisms and so on. The major difference with Alibaba, eBay, or Craiglist is that they mostly advertise 'illegal' goods. Carders.de specialized mostly in credit cards, while HackMarket.ru specialized mostly in cyber-crime tools, albeit some transactions were also about monetary goods (e.g. credentials for Skype accounts).

At first, notice that even legitimated forum markets are rife with scams. After 20 years since eBay's foundation, many frauds reported by FBI's 2013 Internet Crime Reports [24] rely on legitimate forum markets to perform scams: good old lemons are advertised and sold via eBay [24, pag. 8]; bogus real estates are sold via Craiglist; failed delivery or payment of goods are common places; etc.

To create 'safe trading places' where only experienced and trustworthy users participate, forum-based markets have created a number of mechanisms aimed at distinguishing 'good' and 'bad' users. A system to effectively manage reputation is a key issue in the trust of an on-line market place. For example, eBay filed its own reputation based mechanisms for patenting in 2000 [25] and at the beginning of 2015 has almost 200 patents listed on Google's patent with the keyword 'user reputation".

The forum mechanisms in legal on-line markets have provided a 'satisfycing', in the sense of Simon [26], protection to legitimate users to make those markets thrive. For example, Melnik and Alm showed that reputation does matter in sales [27]; Resnick and Zeckhauser showed that buyers and sellers actively and deliberatively provide positive or negative ratings, with positive ratings being the majority [28].

From a legal perspective, reputation mechanisms only provide partial coverage. Law scholars have discussed the issue at length (see e.g. [29], [30] for some of the earliest papers). However, if the reputation mechanism fails, and a 'lemon' is sold via eBay, a customer can always resort to the FBI Internet Crime Center which will pass the complain to the local prosecutor [24, pag. 18]. Similar protections are available to customers in other countries. Such last resort is not available to victims of trades gone sour in Carders.de or HackMarket.ru.

Therefore, illegal markets must either make the reputation mechanism more robust or compensate for the failure of the mechanism with prosecution procedures. Absence or failure of these additional enforcement mechanisms would intuitively re-create the same conditions that Florêncio et al. [7] identified for the IRC markets: information asymmetry would favor 'ripping' behaviour and eventually bring the market to fail.

We formulate a number of hypotheses from the description of the forum regulatory mechanisms (reputation being just one of them). If evidence for the validity of the hypotheses is not found in the data, we conclude that the regulation was not effectively enforced. Vice versa, if most hypotheses are supported by the data, we conclude that the forum administrators applied the stated rules.

A. Effectiveness of Reputation mechanism

If the reputation mechanism works, known scammers should have the lowest reputation among all user.

Hypothesis 1 Banned users have on average lower reputation than normal users.

If Hypothesis 1 is true, it is evidence that the regulatory mechanism for reputation is effectively enforced, and provides to forum users an instrument to evaluate traders' historical trustworthiness. If the data does not support this, "reputation" in the forum is not a good *ex-ante* indicator of a users' trustworthiness.

For amay present a hierarchy of roles or status groups that each user can 'escalate' to. In a functioning system the status should be reflected in the reputation rating.

Hypothesis 2 Users with a higher status should on average have a higher reputation than lower status users.

If Hypotheses 1 and 2 do not hold, it may as well be because moderators left a part of the market to its own and concentrated all regulatory efforts on the higher market tiers. For example, in the **Carders.de** market, there are three Tiers of traders and the first Tier may just represent noise in the data.

To check this possibility, we can restrict Hyp1 to holds only for users that are higher in the hierarchy.

Hypothesis 3 *Banned users who happened to have a higher status have a lower reputation than other users with the same status.*

If even Hyp. 3 does not hold, we conclude that the reputation mechanisms even after controlling for market alleged 'status' provide no meaningful way for the forum users to distinguish between "bad traders" and "good traders".

B. Enforcement of Rules

Reputation may fail to provide effective information, but the hard-wired categories of the forum users (the ones under the direct control of the administrators) may provide a better indicator of quality. Normally, access to the higher market tiers should be subject to some rules. The market is reliable if such rules are consistently enforced.

To see whether this regulation is enforced we can test the following hypothesis:

Hypothesis 4 The ex-ante rules for assigning a user to a category are enforced.

Once transactions fail, Carders.de and HackMarket.ru users cannot turn to legitimate law enforcement agencies for a redress. Therefore, the forum must have some alternative rules to manage trades gone sour.

Hypothesis 5 *There are ex-post rules for enforcing trades contemplating compensation or banning violators.*

C. Market Existence

An obvious, but important question to ask is whether the market actually exists. In other words, whether actual transactions take place (*took* place for Carders.de). Indeed, the role of the forum boards is to provide a platform for sellers and buyers to advertise their merchandise. The actual finalization of the trade usually happens through the exchange of *private messages* between the trading parties [1], [7].

Hypothesis 6 Users finalize their contracts in the private messages market.

If Hyp 6 holds, than the exchange of private messages would be a good proxy for us to measure the successfulness of 'normal' users and 'rippers' in closing trades. To check whether 'normal users' are significantly more successful than 'rippers' we test the following hypothesis:

Hypothesis 7 Normal users receive more trade offers than known rippers do.

For Carders.de, where we have access to the whole forum, a suitable proxy is counting the number of times a forum user initiates a trade with another forum user i.e. the number of *unsolicited incoming private messages* a user receives. The proportion of private messages that are trade-initiation can be calculated to answer the previous hypothesis. For HackMarket.ru such analysis must be qualitative as downloading the whole forum would reveal our presence.

We would expect the results for Hyp. 7 to be coherent with the results obtained so far for the forum. In other words, if the reputation mechanism works, the tier system is properly enforced, and the exchange of private messages is used to conclude the trading process, then we would expect normal users to conclude more trades than rippers do. This is because the consistent enforcement of the forum rules would give the users an instrument to discern rippers from normal users. Otherwise, if the evidence gathered so far suggests a systematic failure in the market regulation, then we would expect rippers to be indistinguishable from normal users because the user cannot do better than randomly picking a seller from the whole population.

IV. THE CARDERS.DE MARKET

In 2010 an online underground market for credit cards and other illegal goods, Carders.de, have been exposed by a hacking team named "inj3ct0r". The leaked package contains a Structured Query Language (SQL) dump of the database, a copy of the Owned and Exp0sed Issue no. 1 (documenting the leak) and an added text file containing all private messages on the forum. By examining the added notes Owned and Exp0sed Issue no. 1 we were able to create what we believe can be considered a close to perfect replica of the original Carders.de forum. It is important to recreate the original settings of Carders.de in order to gain precise insights on the operations of the market, including the reputation mechanisms that were implemented at that time, users' posting history and dates. Appendix A provides an overview of this work. The actual dataset can be obtained from the authors of the paper (http://securitylab.disi.unitn.it/doku.php?id=datasets).

The data consists of forum posts and private message records spanning 12 months from 1 May, 2009 to May 1, 2010 containing a total of 215.328 records.

A. Market Description

The forum has a strict separation of trade related boards and non-trade related boards. Advertisement of (illegal) goods is permitted in the dedicated trading section. Members in this section are also allowed to request specific goods. The non-trade related boards serve the purpose of providing a discussion forum for the members where they can share thoughts, ask questions, publish tutorials and offer free goods on a specific subject. A third area of the forum, of little interest here, is dedicated to discussion of technical forumrelated matters (e.g. maintenance). **Carders.de** allows both English and German speaking members on their forum. Figure 1 shows a schema of the two forum sections for English and German Speakers.

TABLE I CARDERS.DE USER ROLES

Role	Forum	Admins	Other
Newbie	Х		
Normal user	×		
2nd Tier user	×		
3rd Tier user	×		
Verified Vendor	×		
Redaktion		×	
Moderator		×	
Global Moderator		×	
Administrator		×	
Scammers and banned			×

Since we are interested in the market characteristics of the forum, we exclude from the analysis users who have never participated in the trading sections. Further, the Germanspeaking part of the community is clearly the most developed one: the English section has 8% of all market posts while the remaining 92% are found in the German market. For this reason, we will focus on the German market.

Users that join the community for selling or buying products are active in one of the market tiers within the forum. A user can advertise a product by creating a topic in the designated board in which this specific product falls.

In this newly created thread, other users discuss the product, ask questions and when a user shows interest as a potential buyer they contact the advertiser. According to the forum regulation, product trading should be finalized via private messages between the two parties.

1) Member roles: An important part of our study is to distinguish between different types of users. A user's status in the forum is also reflected by its membership in one of 12 user roles identified by the forum administrators. Table I shows these roles with the category to which they belong. The entry rank Newbie labels a newly registered user in the forum. After passing this role a newbie gets the role of normal user. Further up in the hierarchy, the user becomes a 2nd and 3rd tier user and have access to more specialized marketplaces. A verified vendor sells goods that are verified by the administrative team and therefore ought to be more trusted by market participators. In contrast to other forum roles, a verified vendor does not require to climb up the rank ladder to achieve this entitlement.

Users with an administrative role manage, maintain and administer the forum. Members of the 'Redaktion' are editors of the forum. They publish news, events, regulation and other administrative information. The moderators maintain the forum and enforce regulation.

Administrative users are also responsible for banning users who have been reported for "ripping" other users in a transaction, or who have violated some internal rules.

Another important distinction to make is among banned users, which may have been excluded from the forum for a variety of reasons. Banned users are usually assigned an (arbitrary) string tag that describes the reason of the ban. By manual inspection we identified five categories of banned users: *Rippers, Double accounts, Spammers, Terms of Service violators* and an additional "Uncategorized" group for users



Fig. 1. Categories of the Carders.de forum. The German market comprises more discussion sections and more market levels than the English market. Similarly, we found most of the activity to happen in the German section of Carders.de.

 TABLE II

 CARDERS.DE NUMBER OF USERS PER IDENTIFIED GROUP

User group	no. users
Normal users	2468
Rippers	205
Double accounts	148
Spammers	42
ToS	5
Uncategorized	40
Total	2908

banned without a reported reason. Table II shows the number of users for each group.

Each user in **Carders.de** can assign positive or negative *reputation points* to other forum users. Higher reputation points should correspond to a higher "crowd-sourced trustwor-thiness" for the user. In the data there is no historical record of reputation points per users; we only have the reputation level at the moment of the dump. This prevents us from studying the evolution of a user's reputation level with time. For our stated hypothesis this is not necessary.

B. Carders.de's Regulation

The administrators of **Carders.de** published the guiding rules of the community in the regulation section. What follows is an overview of the regulatory structure of the community that will be central to our analysis as it identifies rules to access the trading areas of the forum and provides a principled distinction between "good" and "bad" users.

The forum regulation distinguishes three different *trading areas* (namely *Tiers*) in the forum, the access to which is constrained by increasingly selective sets of rules.

Tier 1 The lowest accessible tier is considered the public market on **Carders.de**. Newly registered users on the forum (Newbies, above) are not permitted to join the public market in Tier 1. the forum regulation statement reports that users that have obtained the role of "normal user" can access this area. *Access rule: To become a normal user a newbie has to have posted at least 5 messages on the board.*

Tier 2 This market section is intended to be reserved to the 'elite' of the forum. More restrictive rules limit access to higher tiers. Access rules: 1) Only users with at least 150 posts are allowed in Tier 2. 2) Users must have been registered to the forum for at least 4 months.

Tier 3 This tier is an invitation-only section of the market. Access rules: 1) The user has been selected by a team member of the forum to be granted access to Tier 3. 2) Access to Tier 2 is required. This division clearly aims at creating 'elitist' sub-communities within the forum where the most reliable and active users participate. One would also assume that users of Tier 2 and 3 would be generally considered, in a working market, more trustworthy than users with Tier 1 only access. We however exclude Tier 3 from our analysis because it features only 5 users, including one administrator, and 17 posts. It is a negligible part of the overall market.

V. CARDERS.DE ANALYSIS

A. A failure of reputation mechanisms

To test our hypotheses we analyze reputation values for users in the **Carders.de** market. Figure 2 summarizes the distribution between banned and normal users, possibly accounting for the respective tiers. The data is on a logarithmic scale. The distribution of outliers suggests that reputation points make little sense with respect to user categories.

A Mann-Whitney unpaired test (chosen for its robustness to outliers and non-normality assumption) with null hypothesis "The difference in reputation between banned and normal users is zero" and alternative hypothesis "banned users have higher reputation than normal users" rejects the null (p = 5.2e - 15). We conclude that banned users have on average higher reputation than normal users. Hypothesis 1 is therefore rejected.

The Mann-Whitney test rejects the null "*Tier 1 and Tier 2* users have the same reputation distribution" and accepts the alternative "*Tier 1 users have a higher reputation than Tier 2* users" (p = 4.8e - 06). Hyp. 2 is rejected as well: reputation levels do not reflect membership in a "higher market level" and are effectively misleading.



Fig. 2. From left to right: 1) Reputation levels for normal users and banned users (whole market). 2) Users active in the tier 1 markets and tier 2 market. 3) Reputation of banned and normal users in tier 2. Banned users showed consistently higher reputation than normal users, even when considering only those active in the tier 2 market. The reputation mechanism is ineffective in both market sections.

Finally, we check whether reputation is at least a satisfactory indicator of user trustworthiness in Tier 2. It is not: Tier 2's normal users have on average a *lower* reputation than banned users. Hyp. 3 is rejected (p = 4.9e - 16).

All evidence suggests that the reputation mechanism in the forum did not work. We therefore exclude that reputation could have been a significant and useful instrument in the hands of the user to identify trustworthy trading partners. This also means that cheaters, or rippers, had no "fear" of having reputation points decreased by a disgruntled costumer, as reputation itself had no meaning whatsoever in the market. The only evidence is that it was used by bad users to inflate their own ratings.

B. A failure of regulations

Carders.de had no ex-post system of regulations (Hyp. 5) and therefore we concentrate on the presence of ex-ante enforcement rules (Hyp. 4). To test the validity of Hypothesis 4 we need to check each individual rule.

If rules are enforced in the first tier this would mean that no user with less than 5 posts is able to participate in Tier 1. We find that more than 50% of the users in Tier 1 accessed it before their fifth post in the community. Despite this being a very simple and straightforward rule to automate, there is no evidence of its implementation in the forum.

The first rule for access to Tier 2 states that users should have at least 150 posts before posting their first message in Tier 2. Figure 3 reports a breakdown of the posting history for each user category. The totality of users with *double accounts* posts in Tier 2 before reaching the 150 post limit threshold. This may suggest that users already familiar with the forum (e.g. previously banned users) were accessing Tier 2 more quickly than others, possibly purposely exploiting the lack of controls. In general, the great majority of users in Tier 2 accessed it before the set limit of 150 posts.

Figure 4 shows a density plot of posts in Tier 2 along the months for which a user is registered to the forum. This also supports the previous conclusion that users had access to Tier 2 immediately when registered. Therefore we also reject Hyp. 4.



Fig. 3. Users in tier 2 with more and less than 150 posts at the moment of their first post in tier 2. Most users had access to tier 2 before reaching the declared 150 posts threshold. D=Double accounts; N=Normal Users; R=Rippers; S=Spammers; U=Unidentified banned users.



Fig. 4. Time Distribution of Posts for Users in Tier 2. Most of the posting activity of users in Tier 2 happened well before they reached the required 4 months waiting period.



Fig. 5. Initiated trades for Ripper users and Normal users. There is no difference in the number of trades the users of the two categories are involved in. Consistently with the analysis so far, this indicates that market participants are not able to distinguish good traders from bad traders.

C. Market existence ... for rippers

Finally, we now measure the effects of these regulatory inefficiencies within the market. We first verify Hypothesis 6. Given the unstructured nature of the data at hand, we proceed with a manual inspection of a sample of 50 randomly picked threads in the Private Message (PM) market and classify them as "trade related" or "not trade related". The goal is to understand whether the ratio of Private Message threads aimed at finalizing a trade supports Hyp. 6 or not.

Table III reports that almost 90% of the manually examined sample threads are trade related. 54% of the trade-related PM

 TABLE III

 CLASSIFICATION OF 50 PRIVATE MESSAGE THREADS IN CARDERS.DE

Туре	# Threads		
Trade Initiated	43	86%	
Trade Initiated & Concluded	27	54%	

Almost all threads in the PM section of Carders.de are about finalizing trades and more than half of them come to a close.

threads also contained contact information between the parties (e.g. ICQ, Post Address and PayPal) and led to a concluding contract between the two. The evidence therefore supports Hyp. 6: there has actually been a market.

We are now interested in seeing whether users that have been banned for explicitly *ripping* other users are more or less successful than normal users. Given the results we obtained so far, we expect the two types to be indistinguishable: if there is no available tool to distinguish between 'good' and 'bad' users (as the evidence indicates up to here), then choosing with whom to trade can be no better than randomly picking from the population of traders. Figure 5 is a boxplot representation of initiated trades for Rippers and Normal users in the forum. The two distributions overlap significantly. A Mann-Whitney test accepts the null hypothesis "There is no difference in the average number of received private messages for rippers and normal users" (p = 0.98). As expected in light of the evidence so far, the systematic failure of the forum mechanisms made rippers and normal users effectively indistinguishable to the trade initiator.

VI. THE COMPARISON WITH HACKMARKET.RU

In this section we provide an introductory overview of the HackMarket.ru market which is still an active and arguably well-functioning cybercrime market.

A. Market Description

HackMarket.ru is a market for exploits, botnets and malware. It is also one of the main markets that introduced exploitas-a-service [3] in the cyberthreat scenario, as we find there the main players and products that the industry reports be driving the majority of reported web-attacks [20]. Indirect evidence of this markets' efficacy is the recent burst in cyberattacks driven by means of tools, services and infrastructures traded or rented in these markets [3], [14], [4]. HackMarket.ru appeared in 2009 in the Russian underground. Differently from Carders.de, HackMarket.ru has a flat trading structure, whereby traders all participate in the same marketplace. In contrast to other hacker for sstudied in the literature [31], it is not public. HackMarket.ru is run in Russian, and very little interaction happens in English. The trading sections in this market are, like in Carders.de, organised by 'topic of interest'. The virus-related area of the market is by far the most popular one, with tens of thousands of posts at the time of writing. Other goods of interest for the marketeers of HackMarket.ru are 'Internet traffic' (i.e. redirectable user connections for spam or infection purposes), stolen access credentials, access to infected servers, spam, bank accounts, credit cards and other compromised financial services. To access the market the forum administrators perform a background check on the participant, that has to provide additional profiles that provably belong to him/herself on other underground communities. We joined this community in 2011 and remained undercover up top the time of going to press (2014).

In this case we do not have an SQL dump of the market, but we will provide instead first-hand evidence that the problems we highlighted for Carders.de are not present here.

For the purpose of this paper, we only focus on some characteristics of this market, which serve as a comparison to our analysis on Carders.de: the reputation mechanism and the punishment mechanism. These characteristics are documented and referenced in the format [ID n], with ID being an internal code we use to classify the evidence and n being the document number.

Interested researchers can contact the authors to access the data (http://securitylab.disi.unitn.it/doku.php?id=datasets).

B. A successful reputation mechanism

The forum regulation outlines seven user groups [DMN 5]. The following list presents these groups in descending order of trustworthiness, i.e. those on top of the list are the most reliable users in the community.

- 1) Admin.
- 2) Moderator.



Fig. 6. Boxplot representation of reputation distribution among categories. Reputation levels are statistically higher for higher categories when compared to reputation at lower categories. Only the categories Trustee and Specialist do not show statistical difference; these two are *elective* categories to which belong users deemed noteworthy by the administrator.

- 3) Trustee: members of the community that "own important services, or are moderators or administrators of other forums" [DMN 5].
- 4) Specialist: Users elected in this group are considered "advanced" users" with a "high level of literacy".
- 5) User: Normal users.
- 6) Rippers: users that have been reported and have been found guilty of "scamming". It is explicitly recommended "to have no deals (business, work) with users of this group" [DMN 5].
- 7) Banned: Users that have been precluded access to the forum.

Reputation points are attributed to users by other users after a positive or negative interaction between the two [DMN 6]. Of course, such system is subject to abuse; for example, a user may want to lower his competitors' reputation level to improve the competitiveness of their own business, or create fake accounts on the market to provide "collective" negative feedback. This adversarial behavior is limited by the mechanism's implementation rules: "Only users with more than 30 posts can change reputation. Only 5 +/- reputation points per day can be assigned by any user to any other users." [DMN 6]. This effectively places an upper bound in the number of reputation points one may assign in a given day and decreases one's influence over the overall distribution of reputation points in the market.

Figure 6 reports a boxplot representation of the distribution of reputation scores among user categories. Categories are listed in ascending order. It is here clear that higher rankings are reflected in higher reputation levels of the users. We run a Mann-Whitney unpaired test to check if the difference in reputation levels between categories is significant, and we find that reputation levels significantly increase with higher categories. The only exception is for the Trustee and Specialist categories, for which no difference is found (which is explained by the elective nature of these categories). While this does not mean that higher reputation results in a higher ranking (as a number of endogenous factors other than reputation may be related to the inclusion in a user group - i.e. there is a selfselection problem), it does show that the reputation mechanism is effectively enforced and results in coherent distributions among users. For HackMarket.ru we accept Hypothesis 1-3.

C. Enforced ex-post regulations

Since there is no market hierarchy, Hypothesis 4 does not apply to HackMarket.ru.

With regard to the ex-post type of regulations (Hyp 5), users can effectively report other users to the board of administrators when they think they have been scammed. The administrators remark that "*We expose [cheaters] with pleasure.*" [ADM 6]. The exposure of a user in the list of cheaters is a fairly refined process, that requires a report to be filed, an investigation to be carried, and that allows the 'alleged scammer' the right to defend himself before the decision by the moderators. The whole phase takes place in a dedicated sub-community of the market, a sort of 'court of justice' where the offended reports the (alleged, at this point) offender.

The reporting is to be filed according to a specific procedure established in the market regulation, that includes the "name, contacts, a proof of the fact (log, screenshot of correspondence, money transfers,..) and a link to the user's profile." Following the filing, an actual 'trial' takes place. The defendant has the obligation of replying to the accusation, as not doing so within seven days from the filing results in the accuser automatically winning the case. The investigation can be carried both by moderators and administrators, while the final decision usually belongs to the administrator. The community is also often active in the discussion, reporting further evidence or personal experience with the accused, or helping in the investigations. An example of regulation during a trial is reported in the following, where the administrator is stating clearly the points of dispute:

Key issues, without which it would be impossible to objectively consider [to put the accused in the] Black [list of scammers]:

- 1) Whether the transfer happened at all
- 2) Whether the transfer was cashed
- 3) Exactly who received/took off with the money. [DMN 1]

A key point is to understand how the punishment mechanism is applied in practice. In particular, we are interested in understanding whether trials unfold with significant discussions, and whether the final decision is ultimately enforced.

To this aim, in Table IV we illustrate three example trials held in the market, two of which ended with a user being 'black listed', and one where the accused is acquitted and no punishment is imposed. We define 'accuser' the user that reports the complaint, and 'defender' the reported user. All three cases were filed by disgruntled clients who paid the sellers but did not receive the goods. All trials above took place within an observation year. In every case, the HackMarket.ru community joins in into the investigation, either providing

 TABLE IV

 ENFORCEMENT OF REGULATION MECHANISMS IN HACKMARKET.RU.

Case	Challenged amount	#Users involved	Evidence	#Messages	Duration	Outcome	Reason
Defender no show	390\$	7	Chat transcripts	11	7 days	Defender banned	Defender never showed up.
Defender loses	2800\$	7	Screenshots, transaction logs, chat transcripts.	29	29 days	Defender banned.	Defender did not provide exhaustive evi- dence that the payment was ultimately com- mitted in favor of the accuser.
Defender wins	1400\$	3	Chat transcripts, screenshots.	9	11 days	Defender found not guilty, no action taken.	The defender demonstrated that good was not delivered because the payment happened during a technical malfunction of his Inter- net connection, and he therefore could not acknowledge it.

Trial regulation is strictly enforced. Evidence brought in support to the case of either the defender or the accuser is always critically analyzed; more controversial trials require longer time to be concluded, and the final decision can be in favor of either participant, depending on how convincing the evidence supporting one's case was.

additional details on the current status of the users involved in the case, or as witnesses with past experience in dealing with the accuser or the defender. As expected, controversial cases take more time than easier ones. In Table IV, the first case is quickly closed as simply the defender does not show up in time. This complies with the forum regulation noted above. The second case is the most controversial of the three, with the defender aggressively participating in the discussion and providing more and more (unsatisfactory) evidence of his innocence. The amount of evidence provided, and the intricacy of the discussion require time for the administrator to come to a verdict, which happens after a month. In the third case, the defender was able to show that he never "cashed" the sent payment. The accuser stops replying soon after that and the administrator closes the case.

Evidence is carefully analyzed by the forum administrator as the following excerpt shows:

Judging from the screen from post #num, there is a transfer, and it was received. Double-check that, you can verify online with Western. But I haven't seen proof of receipt. To get the answer for the third question, we need to ask to whom the money was sent through Western. If I am not mistaken, upon request of the sender they can provide full information.

Therefore, we will do as follows. Sender, i.e. #buyer nickname get all details and full information from Western, report here the result before Friday #date.[DMN 1]

In some cases, the administrator tries to arbitrate the question as s/he clearly values both buyer and seller: *It would be* great if you two [buyer and seller] contact each other and sort this matter out. We only need to know the details for the recipient, and it will immediately be clear who is at fault, even without [proceeding with] the Black [list]. [DMN 1]

On a qualitative note we observed what follows:

- 1) the defender always reports detailed information on the accused user and on the case of complaint.;
- many witnesses appear in 'court' giving opinions on the evolution of the case, or providing supporting evidence for either the accuser and the defender;
- 3) the moderators and the administrators are always present in each report, and actively moderate the discussion;
- 4) when the defender does not show up within the time limit specified by the administrator [DMN 6], the case

always goes to the defender;

- when the defender shows up, he/she always publishes evidence of his/her case, being those screenshots of chats with the accuser or Webmoney transaction logs;
- some cases last several months, with all parties actively participating in the discussion and new evidence being examined or asked for iteratively;
- when the evidence provided by either of the defender or the accuser is not conclusive, the case goes to the opponent or a 'null' is thrown (when neither of the two is convincing, nobody wins);
- users that end up being found guilty are *always* exposed in the list of cheaters and/or are banned from the forum. The latter is a harsh punishment: in contrast to IRC markets,re-entry into the forum is neither easy in effort nor short in time.

We therefore accept Hypothesis 5 for HackMarket.ru.

D. Market existence

We have not direct access to the private conversation of participants in HackMarket.ru, but we collected exhaustive evidence on their private transactions through the conversation logs reported in the trials. In every case reported, the finalization of the contract and the transaction always happen through some type of private communication, usually thought the ICQ chat messaging system or Jabber.ru. We therefore accept Hypothesis 6.

Participants initiating a trade also often declare to have performed a background check on the seller by either contacting the administrators or by checking the official blacklist of the forum. One example of this is given in [NTL 12]: "[*The*] admin [of the forum] confirmed me that you [the seller] are not a rookie trader". Evidence for background checks such as this is frequent. We therefore accept Hypothesis 7.

VII. DISCUSSION

"Regulation" is the main advantage that a forum-based community has over an IRC-based community: it provides the forum users with a set of rules and mechanisms to assess the information they can collect on a particular trade. The analyzed markets attempted to enforce this by providing a regulatory

 TABLE V

 COMPARISON OF RESULTS FOR CARDERS.DE AND HACKMARKET.RU.

Hypothesis	Description	Hyp #	Carders.de	HackMarket.ru
	Banned users have lower reputation than normal users.	Hyp 1	Rejected	Accepted
Reputation mechanisms work	Higher status users have a higher reputation than lower status users	Hyp 2	Rejected	N.A.
	Banned users with a higher status have a lower reputation than	Hyp 3	Rejected	N.A.
	other users with the same status			
Pagulations are enforced	Preventive (ex-ante) rules are enforced	Hyp 4	Rejected	N.A.
Regulations are enforced	Punishment (ex-post) rules are enforced	Hyp 5	N.A.	Accepted
The mentat menta	Users privately finalize their contracts	Нур б	Accepted	Accepted
	Normal users receive more trade offers than known rippers	Hyp 7	Rejected	Accepted

Hypotheses aimed at assessing the reliability of the reputation mechanism, the enforcement of regulation, and market fairness are all rejected for Carders.de. In contrast, HackMarket.ru appears to be a well-functioning market.

mechanism for user reputation and access to "elite" market tiers. This may be not sufficient for the user to have complete information on the transaction; yet, it could provide her with some baseline information on her trading partner, ruling out part of the *information asymmetry* problem identified for other markets [7], and precisely by mitigating the *adverse selection* problem [32]. For legitimate markets, reputation proved to be an effective mechanism albeit not a definitive solution.

Table V reports the summary of Hypothesis testing for the two markets. The organizational and structural differences of HackMarket.ru with respect to Carders.de is evident.

In Carders.de, each of the regulation mechanisms has been faultily implemented and the potential means for a user to assess ex-ante a trade are pointless or even misleading. The systematic failure of the regulatory mechanisms clearly led to a market were users had no incentives in conducting fair transactions and had no means to distinguish "good traders" from "bad traders". We showed that there is in fact no difference in the number of trades initiated with a ripper and trades initiated with a normal user. This effect alone may have brought to the failure of the market, which we show being effectively of the same nature of Florêncio et al.'s IRC market.

In HackMarket.ru the reputation and punishment mechanisms generate meaningful information for the user:

- Evidence supports the hypothesis that reputation points are meaningfully assigned to users and this arguably results in a useful tool for the user to asses potential trading partners.
- The punishment mechanism is a well-regulated one and direct evidence suggests that 'trials' are conducted in a fair manner. This boost market activity and incentivizes 'honest' behavior.
- 3) Users that have been found guilty are, if not banned, publicly exposed and assigned to the 'scammers' group. This allows other users to clearly assess a scammer's trading history and make an informed decision with whom to trade.

It appears that the punishlment mechanism is enforced coherently with the stated rules (e.g. the time frame for the defendant to show up is firmly enforced). We find evidence that trials in the market involve an in-depth discussion on the issue raised by the accuser, and witnesses are called to support one's claims. Importantly, evidence supporting the case of both the defender and the accuser (e.g. transaction logs and previous exchanges between the two parties) is always requested and analyzed. This shows that the forum administrators tend to take well-informed decisions. This is in accordance with the overall reputation levels among categories (Figure 6).

The very fact that defendants do show up is a proof that they see a value in preserving their reputation as users and do not just register with a new account. The difficulties of the registration process makes dropping and re-registering a costly and lengthy process.

The analysis of the HackMarket.ru market also sheds light on the mechanisms and organizational robustness of recent cybercrime communities. These insights can be a basis for future work by capturing the economic mechanisms driving such markets or exploring effective policies to mitigate or discourage these online aggregation of criminals and their operations. For example, in [33] Allodi et al. used information from the black markets to evaluate risk-based patching policies.

VIII. CONCLUSIONS AND FUTURE WORK

The contribution of this paper is twofold. On one side, it replicates and confirms the findings of Florêncio et al. [7] by showing that a badly regulated cybercrime *forum* community is virtually no different from an unregulated *IRC* community. As a result, users participating in those markets have no means to safely assess the characteristics of the user they are trading with. As predicted by Florêncio et al., this leads to a chaotic market where rippers and legitimate sellers are indistinguishable, and therefore there is no incentive for the rippers to not scam other users.

The second contribution of this article provides an example of regulation in a successful underground community, the (indirect) effects of which are daily reported in security news and industry reports. While the evidence presented in this paper is limited by the scope of the article, it does show that rigorously and well maintained underground markets are possible and do exist. These markets are key to explain the economics behind the empirical observations by Gier et al [3] and Allodi et al. [4] among many others: the underground economy should be seen, rather than a confused and unorganized group of criminals scamming other criminals, as a well-organized and administered source of risk that makes for an interesting venue for future research. We leave a formal and proper characterization in economic terms of the working markets for future work.

ACKNOWLEDGEMENTS

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 285223 - SECONOMICS. This work is also supported by the Italian PRIN Project TENACE.

REFERENCES

- J. Franklin, A. Perrig, V. Paxson, and S. Savage, "An inquiry into the nature and causes of the wealth of internet miscreants," in *Proc. of CCS'07*, 2007, pp. 375–388.
- [2] R. Anderson, C. Barton, R. Böhme, R. Clayton, M. van Eeten, M. Levi, T. Moore, and S. Savage, "Measuring the cost of cybercrime." in *Proc.* of WEIS'12, 2012.
- [3] C. Grier, L. Ballard, J. Caballero, N. Chachra, C. J. Dietrich, K. Levchenko, P. Mavrommatis, D. McCoy, A. Nappa, A. Pitsillidis, N. Provos, M. Z. Rafique, M. A. Rajab, C. Rossow, K. Thomas, V. Paxson, S. Savage, and G. M. Voelker, "Manufacturing compromise: The emergence of exploit-as-a-service," in *Proc. of CCS'12*, 2012, pp. 821–832.
- [4] L. Allodi, V. Kotov, and F. Massacci, "Malwarelab: Experimentation with cybercrime attack tools," in *Proc. of USENIX CSET*'13, 2013.
- [5] V. Kotov and F. Massacci, "Anatomy of exploit kits. preliminary analysis of exploit kits as software artefacts," in *Proc. of ESSOS'13*. Springer, 2013.
- [6] N. Provos, P. Mavrommatis, M. A. Rajab, and F. Monrose, "All your iframes point to us," in *Proc. of USENIX Security*'08, 2008, pp. 1–15.
- [7] C. Herley and D. Florêncio, "Nobody sells gold for the price of silver: Dishonesty, uncertainty and the underground economy," in *Proc. of WEIS'10.* Springer, 2010, pp. 33–53.
- [8] G. A. Akerlof, "The market for "lemons": Quality uncertainty and the market mechanism," *Quart. J. of Econ.*, vol. 84, no. 3, pp. 488–500, 1970.
- [9] Symantec, Analysis of Malicious Web Activity by Attack Toolkits, Symantec, 2011, accessed on June 2012. [Online]. Available: http://www.symantec.com/threatreport/topic.jsp?\\id=threat_activity_trends\&aid=analysis_of_malicious_web_activity
- [10] M. Greisiger, "Cyber liability & data breach insurance claims a study of actual claim payouts," NetDiligence, Tech. Rep., 2013.
- [11] M. Yip, N. Shadbolt, and C. Webber, "Why forums? an empirical analysis into the facilitating factors of carding forums," in *Proc. of ACM Web Sci'13*, 2013.
- [12] M. Motoyama, D. McCoy, K. Levchenko, S. Savage, and G. M. Voelker, "An analysis of underground forums," in *Proc. of ACM IMC'11*, 2011, pp. 71–80.
- [13] A. K. Sood and R. J. Enbody, "Crimeware-as-a-service—a survey of commoditized crimeware in the underground market," *Int. J. of Crit. Infrastr. Prot.*, vol. 6, no. 1, pp. 28 – 38, 2013.
- [14] B. Stone-Gross, M. Cova, B. Gilbert, R. Kemmerer, C. Kruegel, and G. Vigna, "Analysis of a botnet takeover," *IEEE Sec. & Priv. Mag.*, vol. 9, no. 1, pp. 64–72, 2011.
- [15] J. B. Grizzard, V. Sharma, C. Nunnery, B. B. Kang, and D. Dagon, "Peer-to-peer botnets: Overview and case study," in *Proc. of USENIX HotBots* '07, 2007, pp. 1–1.
- [16] C. Kanich, C. Kreibich, K. Levchenko, B. Enright, G. M. Voelker, V. Paxson, and S. Savage, "Spamalytics: An empirical analysis of spam marketing conversion," in *Proc. of ACM CCS'08*, 2008, pp. 3–14.
- [17] D. Dagon, C. Zou, and W. Lee, "Modeling botnet propagation using time zones," in *In Proc. of NDSS'06*, 2006.
- [18] H. Fallmann, G. Wondracek, and C. Platzer, "Covertly probing underground economy marketplaces," in *Proc. of DIMVA'10*, 2010, pp. 101– 110.
- [19] J.-M. Yang, R. Cai, Y. Wang, J. Zhu, L. Zhang, and W.-Y. Ma, "Incorporating site-level knowledge to extract structured data from web forums," in *Proc. of WWW'09*, 2009, pp. 181–190.
- [20] Symantec, "Symantec corporation internet security threat report 2013," 2012 Trends, Tech. Rep. 18, April 2013.
- [21] S. Vömel, T. Holz, and F. Freiling, "I'd like to pay with your visa card: an illustration of illicit online trading activity in the underground economy," Universität Mannheim/Institut für Informatik, Tech. Rep. TR-2010-004, 2010.
- [22] M. Yip, N. Shadbolt, and C. Webber, "Structural analysis of online criminal social networks," in *Proc. of IEEE ISI'12*, 2012, pp. 60–65.

- [23] Z. Zhao, G.-J. Ahn, H. Hu, and D. Mahi, "Socialimpact: Systematic analysis of underground social dynamics," in *Proc. of ESORICS'12*, 2012, pp. 877–894.
- [24] FBI, "Internet crime report 2013," Internet Crime Complaint Center, Tech. Rep., 2013. [Online]. Available: http://www.ic3.gov/media/ annualreport/2013_IC3Report.pdf
- [25] R. Ratterman, R. Maltzman, and J. Knepfle, "Determining a community rating for a user using feedback ratings of related users in an electronic environment," 2000, US Patent 8,290,809.
- [26] H. A. Simon, "Theories of decision-making in economics and behavioral science," Am. Econ. Rev., vol. 49, no. 3, pp. 253–283, 1959.
- [27] M. I. Melnik and J. Alm, "Does a seller's ecommerce reputation matter? evidence from ebay auctions," *J. of Industr. Econ.*, vol. 50, no. 3, pp. 337–349, 2002.
- [28] P. Resnick and R. Zeckhauser, "Trust among strangers in internet transactions: Empirical analysis of ebay's reputation system," Adv. in Appl. Microecon., vol. 11, pp. 127–157, 2002.
- [29] M. M. Calkins, "My reputation always had more fun than me: The failure of ebay's feedback model to effectively prevent online auction fraud," *Richmond J. Law & Tech.*, vol. 7, pp. 33–34, 2001.
- [30] M. R. Albert, "E-buyer beware: Why online auction fraud should be regulated," Am. Bus. Law J., vol. 39, no. 4, pp. 575–644, 2002.
- [31] T. J. Holt, D. Strumsky, O. Smirnova, and M. Kilger, "Examining the social networks of malware writers and hackers," *Int. J. of Cyber Criminology*, vol. 6, no. 1, pp. 891–903, 2012.
- [32] K. M. Eisenhardt, "Agency theory: An assessment and review," Acad. of Manag. Rev., vol. 14, no. 1, pp. 57–74, 1989.
- [33] L. Allodi and F. Massacci, "Comparing vulnerability severity and exploits using case-control studies." ACM TISSEC, vol. 17, no. 1, August 2014.



Luca Allodi got his Master Degree from the University of Milan. In 2006 he co-founded a software consultancy Area Software, of which he was Executive Director for five years. He is now a PhD candidate at the University of Trento. Luca's scientific production has been published and presented at many academic and industry venues, among which BlackHat USA and ACM TISSEC. Luca's work has been instrumental for some features of the Common Vulnerability Scoring System v3 (CVSS), the international standard for vulnerability criticality estimation.



Marco Corradin received his bachelor degree in computer networking in 2010. In 2013 he received his master's degree in Computer Science from the University of Twente. His main areas of research interest are criminal behaviour on the internet and forensics in IT. He is currently working as a forensic IT specialist for the Dutch government where he does research in the field of tax fraud, criminal activities on the internet and big data visualisation.



Fabio Massacci is a full professor at the University of Trento (IT). He has a Ph.D. in Computing from the University of Rome La Sapienza in 1998. He has been in Cambridge (UK), Toulouse (FR) and Siena (IT). He has published more than 250 articles in peer reviewed journals and conferences and his h-index is 35. His current research interest is in empirical methods for cyber security. He is the European Coordinator of the multi-disciplinary research project SECONOMICS (www.seconomics.org) on socio-economic aspects of security.