Fighting the "blackheart airports": internal policing in the Chinese censorship circumvention ecosystem

Yi Ting Chua
University of Cambridge
yiting.chua [@] cl.cam.ac.uk

Ben Collier
University of Cambridge
ben.collier [@] cl.cam.ac.uk

***Abstract*—The "Great Firewall of China" (GFW) has become a familiar trope in information security circles. China is comparatively shut off from the rest of the Internet, with extensive censorship and blocking of websites, especially those hosted outside China. The Chinese public use a variety of methods to circumvent the extensive online censorship practised by the Chinese state. In scaling the "Great Firewall", these ordinary users of the Internet now make use of so-called "airport services" which allow tunnelling through to censored websites. This has resulted in the emergence of both an illicit ecosystem of censorship circumvention providers and a population of scammers who set up fake or unreliable services to con these would-be everyday Internet freedom enthusiasts. We have discovered, through scraping the public chat channels used by "airport" providers, that this anti-censorship ecosystem is beginning to develop highly organised methods of self-policing, making co-ordinated use of technicques associated with malicious cybercrime to force these scammers offline and promote a healthy market. While so-called "booter" services - websites where users can purchase Denial of Service attacks for a small fee - have generally been used for malicious purposes, amusement, protest, or extortion, our research suggests that in the Chinese censorship circumvention market they are being used for internal policing.**

***Index Terms*—China, cybercrime, censorship, DDoS**

I. Introduction

In this paper, we document a novel phenomenon – the emergence of what appears to be organised internal policing mechanisms in the market for censorship circumvention services in China. Illicit online markets for drugs [1], hacking tools [2], [3], personal data [4], credit cards [5], and other services have been well-documented in the cybercrime literature, however little attention has been paid to markets for censorship circumvention – services which allow users freer access to the Internet outside their home nation. In China, these so-called "airport" services are a popular way for users to access services which are blocked by their Internet Service Providers (ISPs), such as Netflix, Facebook, or a range of online games. Although these services are prohibited in Chinese law, they complicate the traditional dynamics of harm associated with online illegal markets, and hence constitute a particularly interesting subject of study.

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) [grant number EP/M020320/1].

978-1-7281-6383-3/19/$31.00 ©2019 IEEE

The same issues of trust affect this market as have been identified in other kinds of illegal online markets, and recent years have seen a proliferation of scammers offering circumvention services which do not work, or simply trying to steal money from prospective users [6], [7]. Discussions of these scams are common on Chinese Twitter – for example, users of Lantern (a popular circumvention service) were complaining about the influx of scammers as they reacted to a Twitter post by the service's owner in August 2018, which offered affected customers freebies as compensation for unstable services [8].

These scams make it hard for everyday users of the Internet in China to trust vendors in the circumvention market, and present a further barrier to their attempts to use the Internet free from censorship. Scammers are a perennial problem of illicit online markets, and the different ways in which these markets attempt to combat fraud, self-regulate and promote trust have been of particular interest to cybercrime researchers [9]–[12]. Studies of cryptomarkets have revealed the development of complex internal mechanisms for regulating trust which mimic the reputation systems and escrow services popularised by successful legal online markets such as eBay and Amazon [13]. We have observed the Chinese market for censorship circumvention services also beginning to fight back against scammers and develop mechanisms for promoting trust, however these appear to differ from the "marketplace design" adaptations of cryptomarkets. Instead, a number of highly-organised groups have appeared who carry out vigilante justice against scammers. These groups fulfil a self-policing function for the market, using tools and techniques usually associated with malicious cybercrime to enforce order and ensure stability and trust.

We present here an initial, exploratory study of these groups, documenting the mechanisms by which they attempt to police these markets and how these work in practice.

II. Internet censorship in China – from climbing the wall to scientific web browsing

The Internet in China is subject to heavy censorship and surveillance, with connections to other nations subject to extensive filtering by the so-called Great Firewall of China. Much like the Great Wall of China, the construction of the Great Firewall of China (GFW) spanned over decades. The project allegedly began in 1998 and was launched officially in 2003. The most well-known techniques employed by the

GFW include Internet protocol blocking, domain name system (DNS) poisoning and transmission control protocol (TCP) resets. The GFW is capable of preventing access to specific pages or images, tracking requests from China to foreign websites, and affects the speed of access to certain websites [14]. The scale and magnitude of the GFW is achieved through the cooperation of stakeholders from multiple agencies and the incorporation of surveillance technologies and mechanisms within many different levels of the infrastructures on which the Internet depends. Within China, access to the Internet is controlled by the Ministry of Industry and Information Technology and all eight ISPs are state-licensed [15]–[18]. However, the power to determine objectionable content and content that can be disseminated lie in other agencies such as the Chinese Communist Party Propaganda Department [14].

Curious individuals attempting to research methods of circumventing online censorship in China meet resistance from the state, with related search terms being banned or filtered. The Chinese characters associated with circumventing the "Great Firewall of China" have evolved as a result. One of the earliest general terms describing circumvention is "翻墙" [6], which translates to 'climbing the wall'. As these terms have been censored, there has been a shift in vocabulary used to describe tools that bypass the GFW. Rather than climbing the wall, "翻墙", users now attempt "科学上网", which translates to "scientific web-browsing" and refers to a wide range of censorship circumvention tools and techniques. This type of evolution in vocabulary, staying one step ahead of the censors, is very common in the heavily-monitored Chinese cyberspace. Despite this monitoring, there is a vibrant ecosystem of mechanisms and services for Internet censorship circumvention, or "突破网络审查/封锁" in Chinese.

The wide reach of their control over the infrastructure itself means that it is relatively easy for the Chinese government to adapt to advances in circumvention technology. In September 2009, the Chinese government mandated the installation of a system named Blue Dam, or 蓝坝, at the ISP level. The system included "a graphic-filtering system, administration-management system, and Internet-behaviour manager" with the intention to police online behaviour and access to content [17]. The introduction of Blue Dam came after the failed mandate to install Green Dam on all personal computers during the same year. The Green Dam was nominally introduced as an automated detection and filtering program aimed to prevent children's and adolescents' access to pornographic and obscene materials, but in practice was far more effective in censoring political and religious content [17].

Another well-known state project for Internet control in China is the Golden Shield Project, or 金盾工程. This project has been on-going since 1998 when it was first approved. The goal of the project is to create a nationwide surveillance network as well as a database with information on all citizens. The relationship between the Golden Shield Project and the GFW is unclear. Some claim that it is part of the GFW [14], while others see it as a separate method to achieve Internet censorship [17], [18]. Nonetheless, the overarching goal of these existing mechanisms of surveillance and censorship is to establish Chinese government control over Internet access.

Despite the scope of the controls on the Chinese Internet, laws targeting those who attempt to circumvent these have generally not been strictly enforced in practice. A large majority of netizens who post content in violation of the Chinese laws which forbid the writing and sharing of information that harm national interests are often not legally punished [14]. It is not until the past two years that the Chinese government began targeting the censorship circumvention ecosystem and personal users. For instance, in 2018, an ordinary user was arrested for using a well-known wall-climbing VPN service LanternPro [6]. LanternPro is a "software application for desktop and mobile that delivers fast, reliable and secure access to blocked websites and application." [19].

Across the range of available tools and techniques, the most popular are Shadowsocks (SS) and ShadowsocksR (SSR) [15], [20]. These SS and SSR are essentially socks5 proxy that can be built by oneself or be rented out by business. A market has sprung up of providers who have set up an infrastructure of servers which use these methods to circumvent the GFW, with these providers known as "机场", or airports, as the logo for both applications is a paper airplane [15]. A typical airport service provider would have the options of monthly or annual subscription plans of different tiers. The tiers dictate the amount of allotted traffic and server slots available to customers. For example, the platinum annual subscription plan on one airport allows customers to use one terabyte of traffic per month with more than 80 server slots while the gold annual subscription plan only allows for 500 megabytes of traffic with more than 50 server slots [20].

This underground ecosystem of providers which the Chinese public use to tunnel through the Great Firewall has inevitably created opportunities for less civically-minded groups to engage in scams, taking money for services which do not work or are unreliable. As a result, a range of 黑心机场, or "black-hearted airport" service providers have sprung up to scam customers, decreasing trust in the market for users and frustrating their attempts to circumvent censorship. The emergence of a scammer market is unsurprising. With other online illicit markets, such as online stolen data markets, rippers and scammers are common [11]. One explanation is the relative ease in scamming due to the uncertainty with online transactions where buyers have no way to ensure the quality of products and identities of sellers [21], [22]. In the context of airport service providers, it is very easy for sellers to evade customers either by banning customers or simply shutting down the service, and there are no conventional methods of complaint as the market is already illegal.

These "black-hearted" (scammer) airport services have become more active as the Chinese government has begun to crackdown on virtual private networks (VPNs) and airport providers. This has given rise to organised self-policing mechanisms within the airport community, largely administered through publicly-accessible Telegram channels. In this paper, we conduct an initial study of this emerging phenomenon.

III. Methods

This research grew out of attempts to study so-called booter services (websites providing Denial of Service attacks for hire) and the different ways in which they engage with their communities of users. When scraping chat channels associated with these booters, we included a number of Chinese booter services which appeared to link to other channels. In investigating these channels further, we discovered that these services had links to organised groups, also hosted on chat channels, which were set up for the purpose of policing the illicit market for censorship circumvention tools within China. In researching this emerging phenomenon, we made use of web scraping scripts to collect data from these Telegram channels, numbering 11 in total. Although they are fairly new, all appearing since April 2019, these channels are extremely active, with thousands of posts a week and hundreds of users. We chose four channels from this collection to study in depth, with a total of 175,519 messages. The channel with the largest number of messages (n = 174,608) was a channel dedicated to social interactions and exchanges. These datasets were then subject to extensive qualitative analysis. One of the author, whose first language was Mandarin, read through all messages within all four channels and took detailed field notes. The author also translated selected quotes and figures for the purposes of presentation in this paper. Analysis was conducted on the original Mandarin text. This is not a large systematic study, and we are not making claims about the Chinese censorship circumvention ecosystem as a whole. Instead, we document this emerging phenomenon in its early stages.

In analysing this data, we drew on digital ethnographic approaches from Pink [23] and Kozinets [24]. As the volume of data were too great for qualitative systematic coding, we instead read through the chat channel logs at length and made copious field notes as to interesting findings, themes, and content. This approach has been well-established within the digital society literature. To ensure the robustness of our findings, we triangulated [25] particular findings between the different channels, attempting to document multiple instances of particular phenomena where possible in order to establish their broader salience to the censorship circumvention market. Our focus was on documenting practices, community dynamics, and matters of fact about this novel phenomenon, rather than an in-depth study of discourse or culture, and so a lighter-touch, fieldnotes-based approach was indicated rather than the generation of a codebook. This is appropriate for an exploratory study, however for a more in-depth follow-up investigation we plan to engage in interviews and more systematic coding of a larger number of channels.

In this paper, we set out our initial findings in studying these groups, outlining the key activities in which they are involved, how they attempt to promote trust and assert authority in this market, the dynamics within this community and the motivations underlying what they do. We gave substantial thought to ethical considerations throughout this research, in particular the risk to participants. We obtained ethical approval from our institution to carry out this research, and followed the required standards and provisions mandated by our department. While informed consent has not been established for the use of the data collected, we believe that our research follows the best practice guidance provided by the British Society of Criminology, which states that such research may be justified if individual privacy is protected, and outputs reflect only collective behaviour, and is in accordance with established practice within this field [26]–[28]. Especially keeping in mind the potential threat of police action against individuals or groups we discuss, we have taken care to anonymise the quotes and outputs we present here, and do not name any providers, channels, or individuals. We present quotes translated from the original Mandarin to mitigate attempts to search for the text included, aside from images, which we believe will be difficult to attribute to individual conversation participants in practice.

IV. Results

A. Friendly testing

A key way in which these groups facilitate order and good conduct in the market for "airport services" is through testing. These channels orchestrate so-called 'friendly-stress testing' of airport services, using Denial of Service attacks to direct large amounts of traffic to their servers. While in many online markets this "stress-testing" is often given as a bad-faith excuse for knocking competitors offline, in this case it appears to be genuinely carried out in an attempt to improve the market. These services follow their stress-testing attacks with lengthy reports (see Figure 1), providing tips for improvement where they find services which are poorly-run. Any services which scam, or are unreliable, are added to the list of "black-hearted" airports – an example of how this community is attempting to overcome the problems of information diffusion observed in many illicit markets, which we describe above.

The emergence of these groups appears to be a fairly recent phenomenon. The Telegram channel focused on performing such friendly-testing was created in May 2019, and the report from the first friendly-testing was shared to the channel on the day after the channel was created. The channel advertises its purpose openly, which was to provide updates and target any "black-hearted" airports. The channel also clarified that airports who had been friendly-tested would not be added to the list of black-hearted airports. It is unsure, based on the conversation, whether if these airport service providers gave consent to or were aware of their participation in such tests. Currently, this channel has 1,294 members and has conducted friendly-testing on 37 airport services.

Another Telegram channel conducting similar friendly-testing was created in June 2019. This channel is smaller in size, with 402 members. In its channel description, there was no statement on targeting "black-hearted" airports or any link to other Telegram channels. However, its channel name does translate to website defence testing. It occasionally conduct testing on non-airport websites, but majority of the targets

of the friendly stress-testing were airport service providers. Since its creation, the channel has tested 26 airport services, averaging about four airports a month since its creation. Between these two channels, a total of 58 unique airport services were subject of friendly stress-testing. There was a small degree of overlap between the services tested where five airport services were tested by both channels.

B. Enforcement

Where services appear to not work at all, or there is evidence that they have been set up to scam potential customers, individuals can file complaints in these chat channels against these 'black-hearted' airport providers, showing proof of scams through screenshots of conversations. When the complaints are found to be legitimate, then the channel owner hands out punishment, usually in the form of a DDoS attack which knocks the scam service offline. Examples of these complaints can be found in Figures 1 and 2.

These attacks are not carried out on a whim – the evidence which complainants need to provide is substantial. Figure 2 showed the start of a complaint where the channel owner asked the complainant to begin his/her description. The complainant proceeded to share a screenshot of the conversation between themself and the airport service. In Figure 3, the channel owner said that they will punish the airport once the complaint has been validated, and asked the complainant to provide favourable evidence ("请提供一些有利证据"). The complainant stated that they promoted the airport service at the risk of being prosecuted. The channel owner prompted the complainant to continue, to which the complainant provided additional screenshots of conversation history with the scammer airport service provider. On the same day, screenshots of the scammer airport service being subject to a DDoS attack from the channel owner's booter service was shared on the channel.

Since we began observing this channel, there have been six instances of successful complaints being filed, and punishment carried out. The first complaint occurred seven days after the channel was created, and the punishment was carried out the same day the complaint was received. Three complaints were filed by third-party complainants and three complaints appeared to be originate from the admins of the channel itself. The public dimension of this kind of punishment is a significant factor beyond the harm to the particular service in question – it acts as a powerful statement of norms and an assertion of authority over the ecosystem as a whole.

Occasionally, there are instances where punishments were unwarranted. In one instance, an airport service was attacked because an user asked in an instigating tone for their airport to be attacked for the purposes of testing its resistance to DDoS attacks. The disrespect was noted and a DDoS attack was subsequently carried out as punishment. Later, the channel owner showed that the instigator was actually not the airport owner but rather a customer of the airport service. The actual airport owner apologised and claimed to have no relationship with the instigator. The channel owner warned the airport

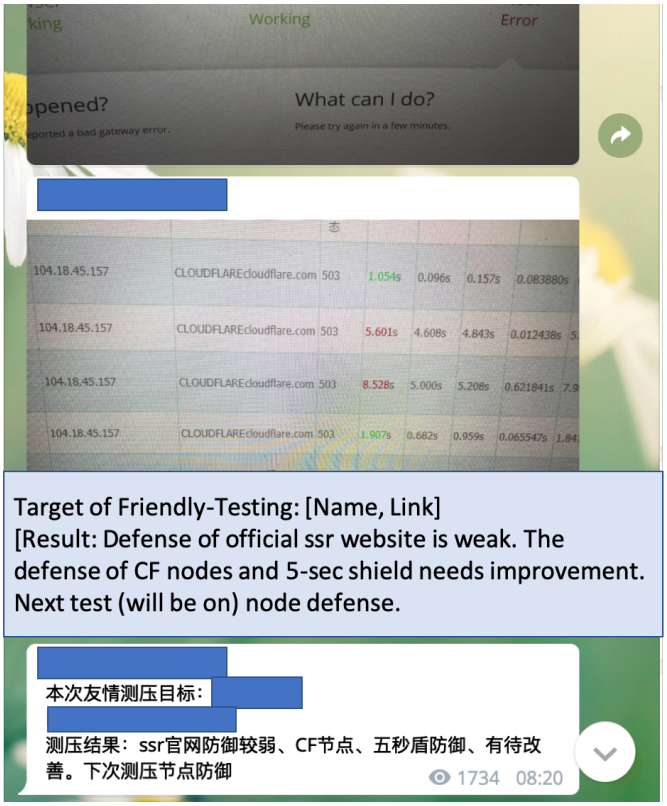

Fig. 1. Part of a "friendly testing" report, (CF refers to Cloudflare)

owner that if there was indeed a relationship between the instigator and the airport owner, then they could expect further punishment.

This semi-formalised policing goes beyond simply the punishment of scammer services. Sometimes this group utilises their hacking skills and conducts a "human flesh search", or 人肉搜索, of individuals who tried to scam legitimate airport service providers. This is the Chinese term for what in the United States information security circles is often known as doxxing – the collection and malicious release of personal information. In one instance, the channel owner leaked scammer's personal information such as date of birth, past mobile phone numbers and email addresses, and records of his ownership of a shell company. The channel had also attempted to telephone the scammer. Infuriated, the scammer threatened to report these hackers to the police. The seriousness of the threat escalated when the scammer shared a video of a police officer threatening to make an arrest. Surprisingly, the police officer became the next target of the "human flesh search"; the channel owner released the officer's personal information, along with his spouse's, the next day. This indicates that, despite their protestations about the risks they take in carrying out this unofficial enforcement role, the operators of these channels are fairly confident in their ability to evade law enforcement attention and arrest.

Fig. 2. Example of a complaint in one of the groups' Telegram channel

Fig. 3. Another example complaint

C. Community

These channels are organised by well-established central members of the community, who wield considerable power. This is reflected in a case where a scammer targeted a legitimate airport service provider, and was subject to "human flesh search" (doxxing). The scammer, who was introduced to the channel as a possible complainant, was infuriated about his information being leaked by hackers while he continued his refusal to provide proof of a transaction. The third-party individual who instigated the complaint eventually responded saying that having brought the scammer to the attention of the group it is now outside their power to stop the attack, and that the scammer could expect the human flesh search to continue as a punishment for upsetting the "talented and skilled" admins of the group by degrading trust in the market. As might be expected, individuals within these groups are particularly cautious to avoid upsetting the administrators as a result.

There are at least three other channels that appear to be connected with the particular enforcement group we studied in depth. One of the channels is dedicated to their DDoS service. It appears to serve as a customer service platform with a side-line in selling DDoS services to users, with constant updates on server performance, service maintenance and lottery drawings for users. This constitutes a mechanism of income generation for the group, with its airport testing service serving the useful ancillary function of advertising the power of their "booter" service to potential users. Besides the channels associated with the group, one can find several other Telegram channels used by various players involved in the airport underground ecosystem. For example, there is a channel that claims to be the official channel for legitimate airport service providers to communicate with each other. Other channels either function as announcement bulletin boards or resource-sharing platforms, and other channels are dedicated purely to fun and social interactions.

These channels operate as important social sites for this community, and they engage in other displays of resistance to Chinese state censorship. Due to the heavy filtering and tagging of online communication in China, a range of strategies exist which permit those attempting to discuss taboo or prohibited topics to evade detection. The heavy use of memes and gifs which incorporate puns on these channels is one example, and in some cases a conversation might be carried out entirely in these memes. A typical image might show a duck standing on another animal, bearing the caption '感到鸭力' (see Figure 3). This is an excellent example of word play in Chinese language. The phrase in the picture translates to "feeling the pressure". However, instead of using the proper word for pressure, 压(yā)力(lì), the first character is switched out with 鸭(yā), the character for duck with the same pronunciation. This kind of wordplay is a common strategy in

Fig. 4. An example of pun-based censor evasion

China for creatively circumventing filtering and monitoring by social media platforms, allowing individuals to communicate discreetly [29].

D. Understandings and motivations

In the ecosystem we have been describing, the scammers disrupting the market for censorship circumvention services are in fact operating in the interests of the Chinese state by frustrating attempts to tunnel through the GFW and degrading trust in the market. Although there is undoubtedly a profit motive at play for the providers of real airport services, there is also a more civic-minded component to their action. Rather than operating on a solely self-interested basis, the operators of these anti-scammer groups appear to have an genuine interest in creating a stable economy for these services and permitting Chinese citizens to use the Internet freely.

Such motivation is highlighted in the regular messages which these groups share about the status of the market as a whole within the past month. For instance, the Telegram channel with nearly 1,300 members included the following message at the end of a friendly-test report: "the current airport market is unstable, there are all kinds of scamming mechanisms out there, all customers/audiences please take care of your purse/wallet". This warning illustrates the motivation to ensure the survival of the market in a period of turmoil, specially with recent crackdown on the use of VPN services in China [6]. The channel has also launched an bulletin board system for users to suggest airport services for friendly-tests as well as reporting "black-hearted" airports.

This focus on the overall health of the censorship circumvention market, rather than attempting to promote a single service, is also evident in one of the early announcements at the start of one of the channels: "Channel Disclaimer: We are not currently testing nodes, including aforementioned two friendly-tests. We only tested the official website. This channel will ignore any defamation against the channel. Second, we have no relationship with airports and are not afraid of anyone causing trouble. If someone really wish to do something, they can try and we will await [your coming] at any time."

This demonstrates at least an initial desire to establish authority and legitimacy over the market as a whole rather than support any particular service. The bravado and relative lack of concern about negative consequences from the Chinese state reflects the shared values underpinning these groups, and they appear to be wielding their authority over parts of the circumvention market with increasing ease.

There are already signs that this may change, however. As of the time of writing, one of these channels has established a "sponsorship" arrangement with one of these policing groups, indicating that their altruistic aims may be short-lived in practice.

V. Discussion

A. Trust issues in online markets

The policing of the censorship circumvention market which document is ultimately a question of trust. It benefits both users of these services, and legitimate providers, if trust in the market is maintained, as without trust, the marketplace is ultimately doomed. Two main sources of uncertainty exist which are unique to online markets. These are quality of product and the identity of sellers and buyers [21], [22]. Quality of product is a source of uncertainty stemming from an inherent asymmetry in information between buyers and sellers [30], [31]. In a real world market, in most cases, buyers would have the opportunity to examine the actual products and determine the quality. With online markets, however, only the sellers have such information. In such cases, buyers need to know that a seller is trustworthy in terms of meeting the buyers' expectations before initiating a transaction.

The second source of uncertainty is the identity of sellers because sellers in online markets are often linked to an e-mail address [30], [31]; there is a lack of valid mechanisms for finding a seller's true identity. In addition, the lack of enforceable legal systems means that buyers will have to rely on informal mechanisms to ensure that sellers fulfil their end of the transactions. Overall, the two sources of uncertainties place buyers in online markets with higher risks of losses due to information asymmetry and the lack of enforceable mechanisms against sellers who cheat. As a result, buyers are more likely to initiate transactions with sellers that are widely considered trustworthy.

B. Trust mechanisms

In legitimate online markets, two categories of mechanism exist to establish trust. The first category contains mechanisms that facilitate institution-based trust and is defined as "trust that is based on guarantees and recommendations from third

parties" [31]. In other words, institution-based trust refers to trust in structures and mechanisms beyond trust at the dyadic level (i.e. trust between sellers and buyers). Some examples of third-parties include third-party services, feedback mechanisms, and guarantee services. Third-party services, such as escrow, ensure safe methods for transferring products and payments [30].

The second category, trust in intermediaries, addresses buyers' trust toward the marketplace itself in general [31]. An intermediary is an organisation that assists the economic exchanges between sellers and buyers [32]. The intermediary can be wholesalers or retailers, or organisations that provide functions such as online storefront management. Some well-known examples of intermediaries for online auction markets are Amazon and eBay [33]. Trust in intermediaries is built upon buyers' perception on the effectiveness of the mechanisms for the facilitation of institution-based trusts [31].

C. Illicit online markets

Much like online legitimate markets, online illicit markets rely on trust to facilitate commerce. Our research findings demonstrate a range of efforts by anti-censorship groups to solve the of trust issues which affect both legitimate markets and markets for illegal services online.

The more well-known types of online illicit markets such as cryptomarkets have developed complex mechanisms for maintaining trust. These tend to fall into two main categories - reputation-based mechanisms, (either through word of mouth associated with the usernames individuals use on these platforms, or through more formal feedback recording services built into marketplaces); and escrow services, which enable the marketplaces to assure purchases by holding funds until the goods requested have been received [9], [13]. Other types of online illicit markets such as online stolen data markets employed similar mechanisms where users provide public reviews and ratings on sellers and untrustworthy sellers are publicly declared as "rippers" and may face repercussions such as being banned from a forum [5], [11], [12].

In the case of cryptomarkets, multiple sellers are drawn together on a single platform, which is often characterised by a shared culture, enacted on the forums and other social spaces linked to the market. The platform, with its built-in trust mechanisms, meant there is often little incentive to develop internal regulation of the marketplace through self-policing. By contrast, in the market for circumvention services, the implementation of reputation, escrow services or feedback system proves to be challenging with the lack of centralised platform or hierarchical structure. Within the airport market, the risk of being detected as scammers is low as information does not diffuse in the same way as it does through a centralised marketplace [34]. If other participants of the marketplace do not learn of scammers quickly, then it increases the incentives for individuals to engage in scamming, eventually leading to market collapse.

The absence of ways to build in trust mechanisms also highly increases the possibility of intermediate fraud. Intermediate fraud refers to businesses and organisations that began as legitimate but later choose to become fraudulent [35]. Equally, where illicit marketplaces accumulate funds or act as escrow services for users, they can engage in "exit scams", where they shut down and walk away with users money. All in all, the dynamics and structures of the airport market suggest that sellers are more likely to violate trust with buyers than on open platforms or cryptomarkets, and as a result, there is a greater demand for alternative trust-facilitating mechanisms in these marketplaces.

The groups we have documented, however, employ different strategies for cultivating trust in the market than are generally observed in online criminal markets. Their effort to police internally not only facilitates trust within services, but it also attempts to cultivate intermediary trust by sending out the message that scammers are not tolerated within the marketplace. Ultimately, these groups attempt to promote the overall health of the market for censorship circumvention. This dynamic, rooted in a belief in Internet freedom, is distinct from an entirely profit-focused enterprise. They bear more resemblance to the forms of non-state order maintenance associated with organised crime groups, which often operate as alternative sources of authority in communities with less access to justice from conventional sources, governing conduct and maintaining social order through violent punishment, but still attempting to cultivate legitimacy within the communities they 'police' [36]–[39].

D. Cybercrime tools

Unlike the regular police, these online anti-censorship vigilante groups do not have access to conventional methods of enforcement such as arrest, imprisonment, or fines. Conversely (as we describe in the Results section), they do have a range of enforcement methods to which the police do not have access – tools more commonly associated with cybercrime activity. In particular, we observed their use of Denial of Service and doxxing.

Denial of Service attacks are a common tool used in the commission of online illegal activities. These involve the generation of large amounts of Internet traffic which is directed at a target, overwhelming them and knocking them offline. As these attacks have become more sophisticated, using botnets and "reflector" servers to generate attack traffic, a substantial market has grown up around them, with providers who amass this attack capacity, selling their service through professionalised "booter" websites. Customers can use these websites to launch attacks with little or no technical knowledge of their own for a small fee. There have been a number of studies of the market for booter services, however this largely documents their use for amusement, protest or malice – knocking competitors on online games offline, retaliation, holding web services offline as a form of direct action activism, threatening or extorting online services, or targeting rival businesses [40]–[42].

Doxxing is another well-established tool of retaliation used to cause harm in communities involved in illegal online

behaviour. This involves collecting personal information about a target through searching public records and attempting to compromise email and social media accounts for compromising photos, videos, and information. This is then used either to publicly humiliate the target, or to attempt to harm them in other ways. By posting this information online, targets can be opened up to a wide range of other attacks, as their address and other details can be used to attempt to steal credit card details, to harrass them or members of their family, or in extreme cases to "SWAT" them by calling armed police to their house on false pretences (such as claiming that they are engaged in terrorist activities) in order to initiate a possibly-fatal confrontation [43]–[45]

While DDoS and doxxing are common features of the cryptomarket economy, we believe this to be the first documented example of their systematic use for internally policing an illicit market. These tend to be used exclusively by services in order to take out their competitors or retaliation, rather than to promote the health of the online market for, for example, prohibited drugs as a whole. By contrast, the kind of semi-formalised internal policing we have observed in these Chinese anti-censorship groups differs from the mechanisms through which other illicit online markets maintain trust. Although there have been some examples of groups like Artists Against 419 [46] previously using DDoS against scammers, we are not aware of any previous research which shows this used as a mechanism for order maintenance within illicit markets.

VI. Conclusions: Cybercrime tools as order maintenance mechanisms

Understanding these kinds of groups poses some challenges for criminologists and cybercrime researchers with an interest in trust mechanisms in illicit online markets. While they are engaged in an attempt to solve many of the key trust problems associated with other online illicit markets, the usual dynamics of harm appear to be reversed from many of the classic examples on which the research literature focuses. Of particular interest is the use, which we document in this paper, of tools associated with cybercrime and malicious harm for order maintenance and market governance.

This reflects the odd space which this illicit market occupies. Although illegal, this market is facilitating a freer Internet for Chinese citizens, with the scammers in fact acting in the interests of Chinese state control. While there is clearly a profit motive in providing these services and maintaining market trust, there is a further interest in this community in growing a stable, sustainable ecosystem of censorship circumvention for the greater good, promoting Internet freedom, (and access to Netflix and gaming) for the Chinese public.

It remains too early to establish whether this self-policing activity is in fact having an effect on the prevalence of scammers, or on the overall health of the market for censorship circumvention services. This is similar to any other illegal online market, due to the barriers to collecting representative data on online services. Further research on this community could attempt to find evidence of this. Tracking how this phenomenon evolves from these early beginnings is also a key future area of research – whether these groups proliferate or die out, whether their struggles for authority over the markets falter, and the reaction of the Chinese state, will all be important factors shaping the evolution of Internet censorship and resistance in China.

Acknowledgements

We wish to thank the Cambridge Cybercrime Centre for access to the booter chat channel datasets. Finally, we thank our colleagues, particularly [removed for peer review], for their invaluable feedback.

References

[1] Judith Aldridge and David Décary-Hétu. Hidden wholesale: The drug diffusing capacity of online drug cryptomarkets. *International Journal of Drug Policy*, 35:7–15, 2016.

[2] Gunter Ollmann. The evolution of commercial malware development kits and colour-by-numbers custom malware. *Computer Fraud & Security*, 2008(9):4–7, 2008.

[3] Juan Caballero, Chris Grier, Christian Kreibich, and Vern Paxson. Measuring pay-per-install: the commoditization of malware distribution. In *Usenix security symposium*, pages 13–13, 2011.

[4] Alice Hutchings and Thomas J Holt. A crime script analysis of the online stolen data market. *British Journal of Criminology*, 55(3):596–614, 2014.

[5] Thomas J Holt and Eric Lampke. Exploring stolen data markets online: products and market forces. *Criminal Justice Studies*, 23(1):33–50, 2010.

[6] Siying. 中国vpn用户被罚"翻墙"怎么会违法. bbc中文[Chinese VPN users punsihed. How is "wall-climbing" illegal]. https://www.bbc.com/zhongwen/simp/chinese-news-46823319, 2019. Accessed on 08.14.19.

[7] VPNDada.com. 中国下令全面禁止vpn ? (2019更新)[China orders to ban VPN?(Updated 2019)]. https://www.vpndada.com/china-ban-vpn/, 2019. Accessed on 08.30.19.

[8] Lantern 蓝 灯. Tweet. https://twitter.com/getlantern/status/1025459212468531202, 2018. Accessed on 08.19.19.

[9] Monica J Barratt, Jason A Ferris, and Adam R Winstock. Safer scoring? cryptomarkets, social supply and drug market violence. *International Journal of Drug Policy*, 35:24–31, 2016.

[10] Cormac Herley and Dinei Florêncio. Nobody sells gold for the price of silver: Dishonesty, uncertainty and the underground economy. In *Economics of information security and privacy*, pages 33–53. Springer, 2010.

[11] M. Yip, C. Webber, and N. Shadbolt. Trust among cybercriminals? Carding forums, uncertainty and implications for policing. *Policing and Society*, 23(4):516–539, 2013.

[12] Marti Motoyama, Damon McCoy, Kirill Levchenko, Stefan Savage, and Geoffrey M Voelker. An analysis of underground forums. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pages 71–80. ACM, 2011.

[13] Meropi Tzanetakis, Gerrit Kamphausen, Bernd Werse, and Roger von Laufenberg. The transparency paradox. building trust, resolving disputes and optimising logistics on conventional and online drugs markets. *International Journal of Drug Policy*, 35:58–68, 2016.

[14] Margaret E Roberts. *Censored: distraction and diversion inside China's Great Firewall*. Princeton University Press, 2018.

[15] Phoebe Cross. Bypass gfw china 2019. https://medium.com/@phoebecross/bypass-gfw-china-2019-9d293b322e20, 2019. Accessed on 08.15.19.

[16] Great firewall father speaks out. http://english.sina.com/china/p/2011/0217/360409.html, 2011.

[17] OpenNet Initiative. Country profile: China. http://access.opennet.net/wp-content/uploads/2011/12/accesscontested-china.pdf, 2012. Accessed on 08.19.19.

[18] Bin Liang and Hong Lu. Internet development, censorship, and cyber crimes in china. *Journal of Contemporary Criminal Justice*, 26(1):103–120, 2010.

[19] Lantern. Frequently asked question, 2019. Accessed on 08.14.19.

[20] DUYAOSS. 浅谈部分机场（ss/ssr提供商）的使用感受–毒药笔记持续更新中[Brief discussion on user experience on some airports (SS/SSRservice providers – continuous update from PoisonNote]. http://387099.blogspot.com/2018/03/ssssr.html, 2018. Accessed on 08.30.19.

[21] David Gefen, Elena Karahanna, and Detmar W Straub. Trust and tam in online shopping: an integrated model. *MIS quarterly*, 27(1):51–90, 2003.

[22] P. A. Pavlou. Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. *International journal of electronic commerce*, 7(3):101–134, 2003.

[23] Sarah Pink. Digital ethnography. *Innovative methods in media and communication research*, pages 161–165, 2016.

[24] Robert V Kozinets. *Netnography: Doing ethnographic research online*. Sage publications, 2010.

[25] Uwe Flick. Triangulation in qualitative research. *A companion to qualitative research*, 3:178–183, 2004.

[26] James Martin and Nicolas Christin. Ethics in cryptomarket research. *International Journal of Drug Policy*, 35:84–91, 2016.

[27] Daniel R Thomas, Sergio Pastrana, Alice Hutchings, Richard Clayton, and Alastair R Beresford. Ethical issues in research using datasets of illicit origin. In *Proceedings of the 2017 Internet Measurement Conference*, pages 445–462. ACM, 2017.

[28] British Society of Criminology. Statement of ethics, 2015. Accessed on 08.30.19.

[29] Gary King, Jennifer Pan, and Margaret E Roberts. How censorship in china allows government criticism but silences collective expression. *American Political Science Review*, 107(2):326–343, 2013.

[30] Sulin Ba, Andrew B Whinston, and Han Zhang. Building trust in online auction markets through an economic incentive mechanism. *Decision Support Systems*, 35(3):273–286, 2003.

[31] Paul A Pavlou and David Gefen. Building effective online marketplaces with institution-based trust. *Information systems research*, 15(1):37–59, 2004.

[32] Mitrabarun Sarkar, Brian Butler, and Charles Steinfield. Cybermediaries in electronic marketspace: toward theory building. *Journal of Business Research*, 41(3):215–221, 1998.

[33] Yannis Bakos. The emerging landscape for retail e-commerce. *Journal of economic perspectives*, 15(1):69–80, 2001.

[34] Werner Raub and Jeroen Weesie. Reputation and efficiency in social interactions: An example of network effects. *American journal of sociology*, 96(3):626–654, 1990.

[35] Wayne E Baker and Robert R Faulkner. Diffusion of fraud: Intermediate economic crime and investor dynamics. *Criminology*, 41(4):1173–1206, 2003.

[36] Aldo Civico. "we are illegal, but not illegitimate." modes of policing in medellin, colombia. *PoLAR: political and legal anthropology review*, 35(1):77–93, 2012.

[37] Diego Gambetta. *The Sicilian Mafia: the business of private protection*. Harvard University Press, 1996.

[38] Jan Van Dijk. Mafia markers: assessing organized crime and its impact upon societies. *Trends in organized crime*, 10(4):39–56, 2007.

[39] Andrew Silke. Rebel's dilemma: The changing relationship between the ira, sinn féin and paramilitary vigilantism in northern ireland. *Terrorism and Political Violence*, 11(1):55–93, 1999.

[40] Mohammad Karami, Youngsam Park, and Damon McCoy. Stress testing the booters: Understanding and undermining the business of ddos services. In *Proceedings of the 25th International Conference on World Wide Web*, WWW, pages 1033–1043, Montréal, Québec, Canada, 2016. International World Wide Web Conferences Steering Committee.

[41] Alice Hutchings and Richard Clayton. Exploring the provision of online booter services. *Deviant Behavior*, 37(10):1163–1178, 2016.

[42] José Jair Santanna, Ricardo de O. Schmidt, Daphne Tuncer, Joey de Vries, Lisandro Z. Granville, and Aiko Pras. Booter blacklist: Unveiling ddos-for-hire websites. In *12th International Conference on Network and Service Management (CNSM)*, pages 144–152, Montréal, Québec, Canada, 2016. IEEE.

[43] Duncan Philpot. Symbolic interactionism and technocrime: Swating as episodic and agentic. In *Technocrime and Criminological Theory*, pages 85–101. Routledge, 2017.

[44] Nellie Veronika Binder. From the message board to the front door: Addressing the offline consequences of race-and gender-based doxxing and swatting. *Suffolk UL Rev.*, 51:55, 2018.

[45] Jean Burgess and Ariadna Matamoros-Fernández. Mapping sociocultural controversies across digital media platforms: One week of# gamergate on twitter, youtube, and tumblr. *Communication Research and Practice*, 2(1):79–96, 2016.

[46] Artists Against 419. New bandwidth policy. https://wiki.aa419.org/index.php/New_Bandwidth_Policy, 2009. Accessed on 08.14.19.