INVESTIGATING THE CHARACTERISTICS OF DARK WEB MARKETS AND THEIR SECURITY IMPLICATIONS

A THESIS SUBMITTED TO

THE UNIVERSITY OF KENT

IN THE SUBJECT OF COMPUTER SCIENCE

FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY.

By

Yichao Wang

January 2025

Abstract

The dark web and the markets on it are often shrouded in mystery. However, the fact that they are little known does not mean that they have no impact on society. On the contrary, these markets have evolved into platforms that may even facilitate cybercriminal activities. Furthermore, dark web markets take advantage of technology to hide themselves on the hard-to-trace dark web, making the fight against criminal activity increasingly challenging. Despite the rapid development of modern dark web markets in recent years, knowledge about them remains limited and needs to be expanded urgently.

Therefore, the research presented in this thesis aims to expand the understanding of dark web markets and explore their characteristics. To achieve this, a custom crawler was developed to help collect data from 21 dark web markets (over different time periods), including quantitative and qualitative data and information.

An in-depth analysis of the current status of the English and Chinese dark web markets was conducted, indicating that markets in different languages could be influenced by cultural factors. Following, security mechanisms on dark web markets were further examined and documented, which not only laid the groundwork for encountering the challenges of data collection in this area but also discussed the role and impact of the market's security mechanisms as a key part of the market's operation. Additionally, an analysis of data and events before and after the closure of markets was conducted, offering deeper insights into the perspectives of

market operators, vendors, and users, as well as the impact of this phenomenon. A case study was also carried out focusing on the availability of child sexual abuse material on dark web markets, obtaining evidence of the sale/distribution of this specific serious criminal material on the dark web market. While it is fortunately not found in the mainstream English dark web markets, there are indications that it is widely available on the Chinese dark web markets.

Overall, the dark web market continues to be an active, often dynamic, and unpredictable environment. Despite the fact that the dark web market has been around for many years (even decades), the lack of up-to-date information and a clear vision of it has meant that we have always been one step behind the criminals. This thesis provides a thorough investigation of the current state of dark web markets in both English and Chinese contexts. It highlights how cultural differences affect market operations, summarises the common security mechanisms employed, and offers insights for effective data collection. Additionally, it categorises three types of market closures and examines the role and impact of dark web markets in facilitating specific criminal activities.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my PhD supervisors, Dr. Budi Arief and Prof. Julio Hernandez-Castro, for their unwavering support throughout my PhD journey. I would like to thank them especially for their patience, empathy, and help. They have also provided me with many excellent opportunities to grow as a mature researcher. Their guidance and feedback, both academically and professionally, have been invaluable. Their mentorship has profoundly influenced my life and daily work. I have learnt so much from them and am deeply grateful for their encouragement and wisdom.

I would also like to thank my collaborators, peers, colleagues, and university staff for creating a supportive working environment. Their support, inspiration and resources are truly heartwarming. I am especially grateful for the unique experiences provided by the Institute of Cyber Security for Society (iCSS) and the University of Kent through their events and opportunities.

Lastly, I would like to express my heartfelt gratitude to my family and friends for their unwavering support and encouragement throughout this journey. I am especially thankful to my parents for their continuous support and financial assistance, which provided me with the foundation and strength to pursue my goals. To my friends, I am sincerely grateful for their constant presence, their encouragement during challenging times, and the many joyful moments we have shared. I truly value the memories we have created together.

Broadstairs, UK & Jinan, China

Contents

List of Tables

List of Figures

Abbreviations

The table below summarises and expands the abbreviations used throughout this thesis in alphabetical order.

Abbreviation	Definition
2FA	Two-Factor Authentication
AI	Artificial Intelligence
API	Application Programming Interface
AUD	Australian Dollar
BTC	Bitcoin
CAD	Canadian Dollar
CAPTCHA	Completely Automated Public Turing test to tell Computers and Humans Apart
CSAM	Child Sexual Abuse Material
CSEA	Child Sexual Exploitation and Abuse
CVV	Card Verification Value
DDoS	Distributed Denial of Service
EUR	Euro
GB	Gigabyte
GBP	Pound Sterling
GPT	Generative Pre-training Transformer
HTML	Hypertext Markup Language

HTTP	Hypertext Transfer Protocol
I2P	Invisible Internet Project
ID	Identification
IOCTA	Internet Organised Crime Threat Assessment
IP	Internet Protocol
IRC	Internet Relay Chat
ISP	Internet Service Provider
JSON	JavaScript Object Notation
LEA	Law Enforcement Agency
LLM	Large Language Model
LTC	Litecoin
LTS	Long-Term Support
MFA	Multi-Factor Authentication
N/A	Not Applicable
NAT	Network Address Translation
P2P	Peer-to-Peer
PGP	Pretty Good Privacy
PIN	Personal Identification Number
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RaaS	Ransomware-as-a-Service
RAM	Random-Access Memory
RQ	Research Question
RUB	Russian Ruble
SOCKS	Socket Secure
Tor	The Onion Router
UI	User Interface
URL	Uniform Resource Locator

USD	United States Dollar
USDT	Tether
VPN	Virtual Private Network
XMR	Monero
XPath	XML Path Language

Publications

The research leads to the following publications, which are presented in chronological order. Chapters in this thesis use content from these publications. All papers underwent peer reviews before publication. All conference papers were double-blind reviewed.

- "Toad in the Hole or Mapo Tofu? Comparative Analysis of English and Chinese Darknet Markets" [142]
  - Authors: **Yichao Wang**, Budi Arief, Julio Hernandez-Castro
  - 2021 APWG Symposium on Electronic Crime Research (pp. 1-13)
  - Year: 2021
- "Dark Ending: What Happens when a Dark Web Market Closes down" [143]
  - Authors: **Yichao Wang**, Budi Arief, Julio Hernandez-Castro
  - 9th International Conference on Information Systems Security and Privacy (pp. 106-117)
  - Year: 2023
  - *Award: Best Student Paper*
- "Investigating the Availability of Child Sexual Abuse Materials in Dark Web Markets: Evidence Gathered and Lessons Learned" [146]

  - Authors: **Yichao Wang**, Budi Arief, Virginia N. L. Franqueira, Anna Grace Coates, Caoilte Ó Ciardha
  - 2023 European Interdisciplinary Cybersecurity Conference (pp. 59-64)
  - Year: 2023

- "The Social and Technological Incentives for Cybercriminals to Engage in Ransomware Activities" [147]
  - Authors: **Yichao Wang**, Sophia Roscoe, Budi Arief, Lena Connolly, Hervé Borrion, Sanaa Kaddoura
  - 9th International Symposium on Security and Privacy in Social Networks and Big Data (pp. 149-163)
  - Year: 2023

- "Analysis of Security Mechanisms of Dark Web Markets" [144]
  - Authors: **Yichao Wang**, Budi Arief, Julio Hernandez-Castro
  - 2024 European Interdisciplinary Cybersecurity Conference (pp. 120-127)
  - Year: 2024

- "Secure in the Dark? An In-Depth Analysis of Dark Web Markets Security" [145] (an extended version of [144])
  - Authors: **Yichao Wang**, Budi Arief, Julio Hernandez-Castro
  - International Journal of Information Security
  - Year: 2025

Chapter 1

Introduction

1.1 Background and Motivation

The dark web always seems to give people a sense of mystery in the past. However, with the development of modern technology and the influence of various films and media, more people have started to become interested and even try to get involved with the dark web nowadays. As mentioned in the latest Europol's Internet Organised Crime Threat Assessment (IOCTA) 2023 and 2024 reports, dark web markets have been identified as a venue often used for advertising and selling illicit services and products, and continue to be facilitators of cybercrime [57, 58].

Speaking of dark web markets, one of the first and most notorious dark web markets is *Silk Road*, but it was shut down in October 2013 [133]. Nevertheless, shutting down a market is not the end of the "game" for many of the people involved. Despite multiple successful crackdowns by law enforcement agencies (e.g., [53] and [56]), these markets continue to demonstrate resilience, indicating that dark web markets are no longer isolated digital underground spaces. They have shifted towards more decentralised and elusive structures [76, 103]. Even now, many newer markets are succeeding and continuously innovating to evade law enforcement efforts [23].

Silk Road ran on the Tor network. Tor (which is short for "the onion router") is an open-source software that uses volunteer nodes to hide users' IP addresses through multiple hops [42]. Tor was initially used for privacy protection [116], but cybercriminals also try to benefit from the Tor network to hide their real identities and illegal activities. Hence, *dark web* markets – also commonly known as "anonymous markets" – naturally put in place measures for protecting the privacy of their users (both sellers and buyers), as this is a key priority that can attract users worldwide. In contrast, we call the regular Internet that we use on a daily basis, that is, the websites indexed by mainstream search engines (like Google and Bing), the *clear web* (also known as "surface web").

In addition to Tor, the Invisible Internet Project (I2P) [81] is another anonymous network that has emerged recently. Some mainstream dark web markets also support dual-network operations. I2P's technical implementation is slightly different to Tor, but in practice, its reliability and performance (i.e. access speed) have certain advantages over Tor [68, 144]. Although there are a few other dark web network platforms (such as Freenet and Riffle), Tor and I2P remain the most widely known and popular [59, 97]. In this thesis, we focus mainly on dark web markets on the two most popular anonymous networks (i.e. Tor and I2P).

From a broader perspective, dark web markets serve primarily as the platform for monetisation or trading within the larger cybercrime ecosystem. Previous studies have extensively explored the structure and development of dark web markets and their role in the trade of illegal goods. For example, Christin [26] conducted a detailed analysis of the development of the *Silk Road*, revealing different transaction mechanisms and user behaviours in dark web markets in 2013. Van Wegberg et al. [136] tracked the evolution of commoditisation on eight English dark web markets over six years. The paper concluded that retail has a large share in the dark web markets, and the overall revenue for cybercrime commodities was at least $15 million between 2011 and 2017. More recently, Kermitsis et al. [86] identified

six types of dark web markets: weapons, drugs, jewellery and gold, fraud and counterfeiting (e.g. fake documents, stolen credit cards), guides and tutorials, and digital goods and services (e.g. software, botnets, currency exchange).

While these studies provide us with a preliminary understanding, they mainly focus on early market platforms or snapshots of multiple platforms. In recent years, with the continuous development of new markets and the emergence of more complex trading models, in-depth and a combination of observational, retrospective, and longitudinal approaches to the analysis of modern dark web markets still need to be expanded. In particular, the coverage of existing literature needs to be more comprehensive in terms of studying how modern markets operate, along with their long-term trends.

Furthermore, unlike earlier market environments that were dominated by a few individual platforms, the modern dark web market has diversified [114], which has likely changed the way the market operates. Because of these changes, we urgently need newer findings and results to gain insights into how these platforms have evolved and grown, how they are linked to other criminal behaviour and so on. In general, we find that staying up-to-date with the latest work in this field is essential. Although we are often one step behind cybercriminals, understanding the problem itself is a prerequisite to finding solutions.

Therefore, we believe that the work presented in this thesis, along with the research directions explored, is both timely and valuable in contributing to the fight against cybercrime. This thesis aims to provide a more comprehensive understanding of dark web markets by exploring and documenting the mechanisms of their operation, trends, and behaviours adopted by their stakeholders (i.e. vendors, buyers, and market owners). Most importantly, a deeper understanding of the different characteristics of dark web markets is fundamental to disrupting these illegal trading platforms.

1.2 Research Questions

Unlike most research areas, where a clear direction is often established at the outset, this area is characterised by numerous unknowns and constant change. The dynamic nature of the domain makes it inherently unpredictable, and new developments occur continuously. Nevertheless, the main aim of this thesis is to investigate and understand the characteristics of dark web markets and explore their security implications. Benefiting from our preliminary data collection, the following research questions (RQs) have been formulated to start this investigation:

RQ1: *Do cultural differences influence the operation and structure of dark web markets in different language communities?*

We initially examined extensive discussions concerning dark web markets found in dark web forums. Often, popular dark web markets advertise themselves on the forums. We also gathered information from several news websites (in both English and Chinese) related to the dark web, many of which often include links to dark web markets. We then conducted a comparative analysis of English and Chinese dark web markets in Chapter 4.

After a period of conducting data collection and observations, we found that gathering comprehensive data proved challenging due to the inherent security mechanisms of the market itself. Additionally, as several of the observed markets continued to shut down one after another, we were compelled to invest considerable effort in continuously adjusting and refining our web crawlers. As a result, we formulated the following two research questions:

RQ2: *What mechanisms do dark web markets employ to ensure the protection of their operations and the safety of their users?*

RQ3: *What trends can be identified in the development of dark web markets*

over time, and what underlying factors contribute to these trends?

To answer these questions, we studied prior works on the security mechanisms and the development trends of cybercrime underground forums. We found these to be different from what we observed on dark web markets. Thanks to our periodic data collection, we could further analyse and document the latest security mechanisms of dark web markets (in Chapter 5) and traced back to what happened before the market was closed (in Chapter 6).

Finally, although our previous questions have consistently focused on dark web markets, we do not yet know the impact of dark web markets on a particular type of cybercrime. Therefore, we posed the following research question:

RQ4: *What role does the existence and development of dark web markets play in specific cybercrime?*

To achieve this, we collaborated with psychology scholars to conduct a case study on the availability of child sexual abuse material (CSAM) in mainstream dark web markets (in Chapter 7).

1.3 Thesis Contributions

Based on the research questions we raised in Section 1.2, this thesis makes the following contributions:

- **Towards RQ1:** An in-depth comparative analysis of English and Chinese dark web markets was conducted, involving three English and two Chinese dark web markets, to understand the characteristics of the emerging and popular-choices (at the time) related to dark web markets operating in these two languages. This work described, analysed, and compared the results of our investigation in six main aspects: (i) *operation model and structures*, (ii) *product categories*, (iii) *market policies*, (iv) *payment methods*, (v) *security*

mechanisms, and (vi) *vendors' characteristics*. This work can be further extended to encompass dark web market communities in other languages (e.g., Russian, French, Spanish, etc.), offering valuable insights that would enhance the academic community's ability to explore and analyse this landscape more comprehensively.

- **Towards RQ2:** A comprehensive analysis and documentation of the security mechanisms utilised in twelve mainstream dark web markets were completed. We classified and described the security mechanisms used in these markets into three main categories:

    1. *web security*, which includes accessibility, waiting queues, anti-phishing, CAPTCHAs, secret phrases, warrant canaries, bug bounties, rate limiting, and distributed denial-of-service (DDoS) protection;
    2. *account security*, which includes areas such as username, password and personal identification number (PIN) requirements, mnemonics, multi-factor authentication (MFA), account kill-switch; and
    3. *financial security*, which covers strategies such as the choice of (crypto) currency being used, specific transaction concepts such as the use of multi-signature, escrow and finalise early, as well as the handling of complaints and general user support.

    This work also shares some insights into underlying trends and data collection and raises some ethical considerations that may be relevant in this research area.

- **Towards RQ3:** After capturing the trend of dark web market closure, we conducted an analysis based on the data collected before the market closure. Through our investigation, we classified the ending of dark web markets into three categories: *exit scams*, *voluntary closures*, and *taken down by law*

enforcement agencies (LEAs). We tracked some indicators and came up with insights into dark web market development and life-cycle.

- **Towards RQ4:** A case study was conducted, presenting the findings and insights from an investigation into the characteristics and features of CSAM in both English and Chinese dark web markets. This analysis aimed to enhance the understanding of the role these dark web markets play in the broader cybercrime economy. Building on the results of previous research, we reviewed and compiled four sets of keywords used to detect or identify potential CSAM within dark web markets. Using these keywords, we created a new text-based dataset for data collection. The analysis of this dataset offers novel perspectives on the policies and trends of two Chinese dark web markets involved in CSAM transactions, as well as the protections implemented by eight identified English-language markets.

- All available datasets will be shared with the academic community, security researchers, and LEAs *upon request*, as their potentially criminal content may not be appropriate for public release. Those datasets are intended to encourage further research in this area and mitigate the challenges encountered during data collection.

1.4 Thesis Outline

The rest of the thesis is organised as follows:

- **Chapter 2** provides an overview of the relevant background and a literature review of the field of work. It includes key research in cybercrime, not just on the dark web, such as underground forums on the clear web. It also covers research specific to dark web markets and communities.

- **Chapter 3** details the research methodology, covering the overall approaches,

the crawler's design and technical environment, basic information of our dataset, and ethical considerations in this study.

- **Chapter 4** compares the characteristics of the English and Chinese dark web markets to provide an overview of the current market landscape and explain how they operate. It also explicitly contributes results from non-English markets to fill the research gaps, since the majority of existing research has been focusing on English-language markets only.

- **Chapter 5** delves into the implementation of security mechanisms on dark web markets, aiming to gain a comprehensive understanding of the core characteristics of the dark web markets. It also documents the challenges encountered during data collection and provides valuable insights. In addition, it explores how these security mechanisms may impact the operation of the dark web markets.

- **Chapter 6** records and analyses pertinent data from six dark web markets before their closure. We also categorise the reasons for the closure of dark web markets to help the academic community better understand the trends in market operations.

- **Chapter 7** presents a case study examining the availability of CSAM in English and Chinese dark web markets. We focus on this severe type of cybercrime (i.e. the sale/distribution of CSAM), using the case study to understand the role of dark web markets in such cybercrime.

- **Chapter 8** concludes the thesis, highlighting the contributions of this work, pointing out directions for future research, and outlining insights and reflections on the challenges in this field.

Chapter 2

Literature Review

2.1 Introduction

This chapter provides a literature review to introduce some background and explore the current state of research in the area of dark web markets. Figure 1 presents an overview of the structure of this literature review. We start with the background of the dark web, introducing the basic characteristics of Tor and I2P networks. Then, we explain some of the terms used in this thesis to avoid any confusion. In Section 2.3, we focus on basic cybercrime and underground economy research. Finally, we provide an overview of research related to dark web markets, which is divided into four key areas that correspond to the main chapters of this thesis. These areas are market measurement and documentation, security mechanisms, dynamics of market closures, and the studies of items sold on markets.

2.2 Dark Web Background

The scope of this thesis is to focus mainly on dark web markets on these two anonymous networks: Tor and I2P. Nonetheless, all markets we observed operated

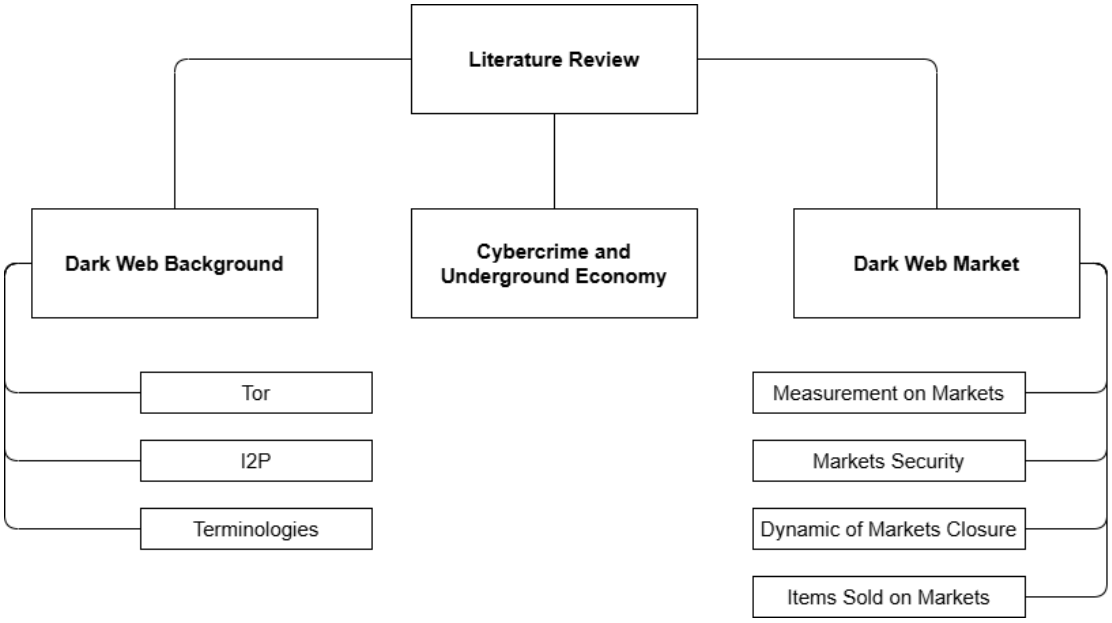


Figure 1: An overview of the structure of the literature review in this thesis

on Tor, and some of them were hosted on the I2P network. It is also worth mentioning that with the development of dark web markets, more markets expand their accessibility through I2P. Traditionally, most dark web markets have relied heavily on the Tor network due to its robust anonymity and widespread use. However, from late 2022 to early 2023, the Tor network has been suffering from various performance issues [128]. Several markets have built I2P-based websites to ensure stable and reliable user access to their markets. This dual-access strategy provides users with an alternative when Tor suffers from downtime. In some cases, the Tor network suffers from network performance issues, and our data collection also moved to the I2P network as an alternative solution. We discuss this phenomenon in more detail in Chapter 5. To begin with, we present the technical implementations of these two networks, i.e. how they provide anonymisation to their users.

The Tor network, known as the onion router, protects data through multiple layers of encryption like the layers of an onion. The user's traffic passes through multiple relay nodes distributed across the world. Each node only knows the

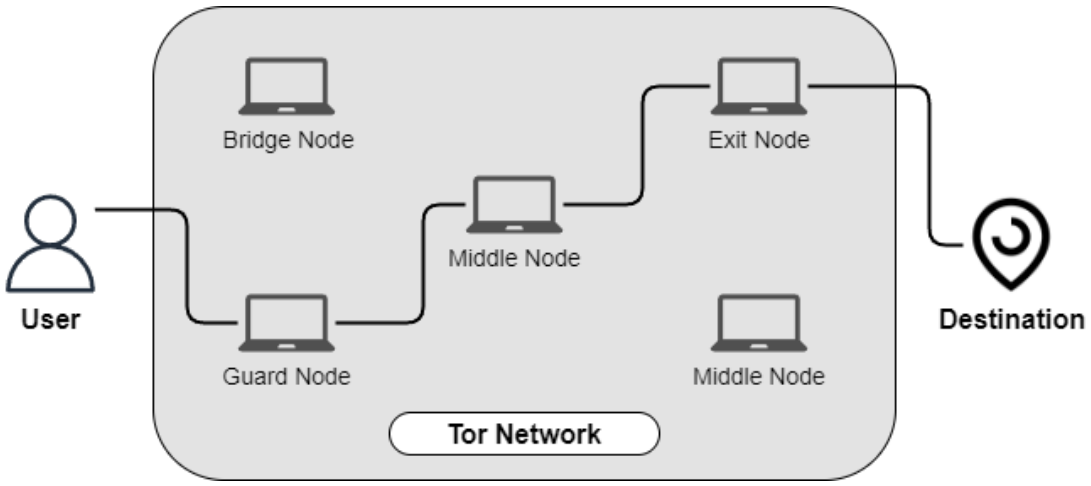


Figure 2: Topology of the Tor circuit

previous and next hops of the data, and does not know the complete path. The data is encrypted layer by layer until it reaches its final destination. Figure 2 shows the topology of the Tor circuit. There are four types of nodes: guard node, bridge node, middle node and exit node [130]. The guard node is the first hop in a Tor circuit. The bridge node is a relay (hop) in the network that is not listed in the public Tor directory and uses special techniques to obfuscate traffic, making it more difficult for ISPs and governments to block when users establish a Tor circuit. The guard node and bridge node can be collectively referred to as the entry node. The middle node is the second hop in a Tor circuit. Finally, the exit node is the final hop. The destination end will see the IP address of the exit node instead of the Tor user's real IP address. The Tor network consists of thousands of volunteer-owned servers, referred to as Tor relays (nodes) [129].

In addition to understanding the topology of Tor, the following features of Tor are worth highlighting:

1. Multi-layer encryption: When a user wants to connect to a target website, the Tor client first encrypts the packet several times. The outermost layer of encryption is for the first guard node, and each layer after encryption corresponds to a relay node. In this way, each node can only decrypt its own

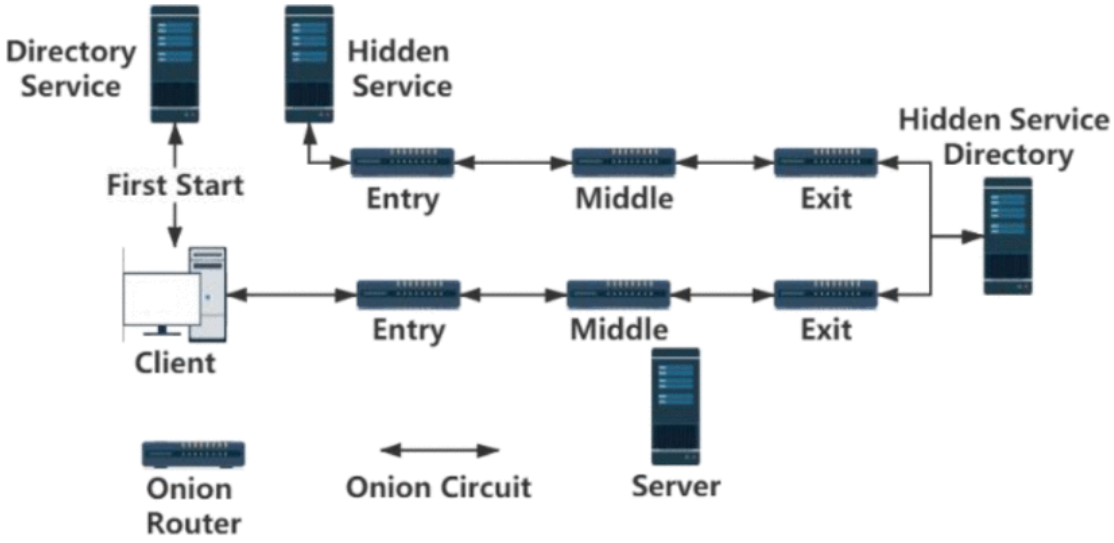


Figure 3: Topology of Tor hidden service, figure obtained from [160]

layer of encryption and get the address of the next node without knowing the original sender or the final destination.

2. Path selection: Tor clients randomly select multiple relay nodes to construct a virtual path, just like shown in Figure 2.

3. Hidden server: It is the server in the Tor network that provides content such as operating a dark web market. When it communicates with a Tor client, it needs to go through a Tor circuit as well. Thus, when a client accesses the hidden server via Tor, the data does not leave the Tor network.

4. Hidden service directory server: It has and maintains a distributed hash table that is used to record the corresponding public keys and introduction points.

Zhang and Zou [160] drew Figure 3 to further explain the topology of users accessing the hidden server on Tor, in which the directory server is used to make the path selection.

I2P is another anonymous network we utilised in this thesis. Compared to Tor, it is not as out-of-the-box as Tor (to use Tor, a user only needs to download and

install the Tor browser). Instead, I2P requires some simple configuration to make it run. The I2P network is a peer-to-peer (P2P) network, which consists of many nodes with unidirectional inbound and outbound virtual tunnels. Each tunnel uses a separate path, ensuring that the routes for outgoing and incoming data are distinct, which enhances anonymity. Transit nodes along the path forward the encrypted data, only knowing the next hop but not the full route or the data's origin and destination. I2P uses garlic routing [159], an enhanced version of onion routing. It packages multiple encrypted messages (or "cloves") together. Each clove contains routing information for a specific segment of the path. This approach improves efficiency and anonymity. For address lookup, I2P has no centralised directory server. Instead, users can maintain their own address books. If a user does not know an address, they can query peers through the distributed network to resolve it.

In the literature, the usage of some terminologies may cause some confusion. Therefore, we make some clarification here.

First, the *dark web* is publicly accessible. Even though some websites require registration, the registration process does not require any personal information. Compared to the clear web, account registration in some dark web markets and forums does not even require an email address. This mainly depends on the service provider, who can advertise that no personal information is required to use their services. However, the content on the dark web needs to be searched manually and requires access using specialised software/techniques we mentioned above. These servers are not indexed by mainstream search engines. The *deep web* is accessible through the public internet (without any software needed) but is usually not indexed by mainstream search engines for a variety of reasons, such as invitation-only forums, private chat messages, etc.

Second, there are multiple types of places that can sell items on the dark web. In this thesis, we call a place that serves a trading platform a *market*. The place

should have proper infrastructure for transactions between vendors and buyers. For example, platforms such as *eBay* and *Amazon* attract independent vendors to join and sell products to buyers. In this thesis, we use the term *market* to refer to such selling platforms and websites. There are other types of websites that sell items, but those are beyond the scope of our discussion. When an independent vendor operates their own website for sales, we refer to it as a *vendor shop*. Additionally, some forums have a sub-board dedicated to transactions between users, which is often collectively called a market in the literature.

2.3 Cybercrime and Underground Economy

Cybercrime is an extensive concept, and its research usually involves multiple disciplines, such as computing, sociology, criminology, and economics [22]. The cybercrime community is a platform for cybercriminals to carry out cybercrime activities, both in the dark web and surface web [79]. In the early days, Franklin et al. [64] studies an active underground economy in Internet Relay Chat (IRC). As time goes by, more recent communities also adapt to a more modern "bullet-proof" infrastructure [106].

A model of cybercrime from a stakeholder perspective was proposed by Arief, Adzmi and Gross [3] [4]. The authors refer to cybercrime as a combination of crime and cyberspace, where cyberspace can be a target or a medium. The stakeholders of cybercrime are divided into attacker, victim, and defender. The attacker is defined as someone who initiates a threat. They usually have tools and techniques for attempting the attack with certain motives, such as financial gain or psychological factors. The victim is a potential target of cybercrime. The defender aims to prevent attacks from occurring through technology and their capabilities. It is important to note that attackers can be ethical, which means they warn the potential victim after discovering vulnerabilities to help the defender fix and

implement new and effective strategies and techniques. Although the proposed model is primarily an attack-defence type of model, the key role of the human factor in cybercrime is highlighted. In addition, the National Academies of Science [104] noted that one area of cyber security research focuses on technical or data-related challenges, such as identifying malware and mitigating the impact of malicious attacks. Another area, cyber security for society, aims to identify, understand, and predict the societal impacts and changes brought about by the digital online environment. This thesis focuses more on the latter.

Although there is still no universal definition of cybercrime, Wall [140] summarised the four categories of cybercrime as follows: (1) cyber-trespass; (2) cyber-deception/theft; (3) cyber-porn and obscenity; and (4) cyberviolence. The challenges in these areas have evolved over the past two decades and up to the present. In 2013, Holt and Bossler [75] reviewed the current status of the four areas mentioned above. The authors highlight several questions that still need to be considered. Researchers need to assess emerging forms of cybercrime and victimisation that have not been thoroughly examined. At that time, there were some active stolen data markets operated primarily on websites hosted abroad, where users communicated in Russian rather than English [27]. Therefore, both qualitative and quantitative studies are required to gain a deeper understanding of these markets. In fact, not only that, in recent years, cybercrime has also developed rapidly along with the development of the Internet.

Unlike more technical cybercrimes such as cyberattacks and malware, contraband and illegal resource trading have also received academic attention in recent years. In 2013, Yip, Webber and Shadbolt [157] studied the trust issues among cybercriminals in the underground markets. The authors examine the structure of cybercrime by analysing data from online underground markets. They noted that cooperation between criminals is possible due to the potential benefit even

in the absence of trust. There are two main sources of uncertainty in the underground economy: the quality of goods and services, and the trading partner's identity (i.e. whether this partner is a real/honest trader, a dishonest trader, or even a law enforcement agent). To support their findings, they also provided a case study – *ShadowCrew* – a market to trade stolen credit cards. The involved theories were drawn from social psychology, organised crime, and transaction cost economics. They argued that the law enforcement community must prevent cybercriminals from building trust. When trust develops to a particular stage, it becomes difficult for law enforcement agencies to infiltrate criminals and carry out operations. For example, some communities have moved to an invitation-only format. Also, cracking down on similar intermediary platforms (i.e. underground markets, underground forums, trading groups) should be a priority.

More recently, Vu et al. [139] studied the evolution of a well-known cybercrime underground market – *Hack Forums* – on the surface web from June 2018 to June 2020. This market offers a range of features to foster trust between sellers and buyers. A contract system is encouraged to be used in the *Hack Forums*, to gain a certain level of protection (e.g. opening disputes) for both parties. Thanks to the contract system being publicly accessible at the time, the authors analysed nearly 190,000 real contracts over the two years. They divided the entire observation period into three eras: the set-up era, the stable era, and the Covid-19 era. In the set-up era, the contract system is established but is not mandatory. Users started making small transactions with each other. However, over time, disputes between users began to increase. In the stable era, the contract system is mandatory for trading. Larger vendors were starting the trend in retail. Reputation and trust began to attract more attention. In the Covid-19 era, the number of transactions had increased across all product categories, and the Covid-19 pandemic had a positive influence on the market sales. The paper comprehensively describes the period from inception to the prosperity of an underground forum. The authors

suggest that the best time to strike an illegal market is in the early era when user trust has not yet been built with others.

From the perspective of underground markets/forums, reputation and trust are very important to vendors. Moreover, researchers tried to identify key actors by using their characteristics and assets in underground forums.

Benjamin and Chen [8] examined the relationship between various hacker posting behaviours and their reputation in two major underground hacker forums in the United States and China in 2012. Both forums can add code and tools to messages, and both have peer-assessed reputation systems. The results show that those who made outstanding contributions (i.e., provided more tools and code) to the forum generally have a better reputation, while active time and discussion quality are not significantly correlated with reputation. Similarly, Samtani, Chinn and Chen [121] explored the hack assets (i.e., tools, code, and tutorials) in underground forums. The results show that most of the code is not related to a specific attack, but is mainly related to the post's topic, with tutorials being the main resource. Li and Chen [95] proposed a framework for identifying the top malware/credit card sellers in a Russian forum. The method they used was based on a deep learning-based sentiment analysis framework, so the automated solution can save time. Peersman, Pencheva and Rashid [110] applied conversation analysis to interactions between underground forum users. They found four main types of users: passive users, the most common type of user (95%), who do not contribute much to the forum; entrepreneurs who commit to increasing income through communication; influencers who provide advice or more technical details; and gatekeepers who may include administrators or reputable users who support the forum's social network.

Although we have very briefly outlined some of the research ideas in the area of underground economy above, there are still some limitations and challenges. Hughes et al. [79] systematically sorted out the current status and challenges

in this area. They summarised the inherent challenges of studying cybercrime communities into six categories:

1. Data collection is limited to the specific community/site/market being monitored. In the time dimension, the availability of a site is influenced by many factors (e.g., network congestion, server load).

2. Unstructured data thwarts large-scale analysis.

3. The leaked dataset is a single snapshot, so the results may not be generalisable, and the patterns of cybercriminals can change.

4. When measuring entire communities at scale, the completeness of the dataset may be biased.

5. Datasets often lack ground truth, mainly due to the complex nature of the data, and labelling usually relies on human effort.

6. Machine learning has certain limitations when applied in this area, especially in terms of training time and the difficulty of reusing the model on other data sources. Another paper argued that heuristic approaches – such as learning from conversations in underground forums – are better alternatives [109].

Pastrana et al. [108] realised that most of the previous datasets were outdated and did not provide a clear understanding of the underground economy at the time. Therefore, they built a customised crawler and presented the CrimeBB dataset. Subsequently, they also expanded the scope to other well-known underground forums and shared a large number of valuable datasets with the research community [20]. To encourage more researchers with interdisciplinary backgrounds, Pete et al. [111] also developed a toolset that allows researchers with non-computer backgrounds to easily access these datasets.

In summary, many excellent studies covering various aspects of cybercrime exist under the broad concept. Although some inherent challenges exist, researchers point out ways to improve future work.

2.4 Dark Web Market

In the previous section, we briefly introduce and provide an overview of cybercrime and the underground economy. In this section, we focus more specifically on the dark web.

As mentioned before, the dark web is able to provide technically enhanced anonymity. As a result, unlike the pretence on the surface web (i.e., people often claim that they are just cheating at games, exchanging software techniques, etc. [78]), on the dark web, cybercriminal activities are more obvious.

Bermudez Villalva et al. [10] compared the usage of leaked account credentials on the dark web and the surface web by monitoring using a honeypot infrastructure. The authors first manually created 100 honey accounts on Gmail, and then posted those account credentials on both the dark web and surface web and monitored their activity. The results showed that the compromised accounts that were leaked on the dark web had been accessed more times. This may be because the administrators of the surface web actively deleted the leaked information. This paper serves as an example, highlighting the potential for a higher proportion of malicious activities on the dark web.

In recent years, with the rapid development of information technology, including the popularity of cryptocurrencies, and the development of secure browser technologies such as Tor, people have gradually begun to pay attention to privacy issues. The dark web has become a new platform for cybercrime, even organised crime [148]. Nowadays, the development of the underground economy has shifted from IRC to more modern platforms, including the underground forums

mentioned above, more private end-to-end encrypted chat software [29], and the dark web markets we are concerned about. There is a high possibility that dark web market transactions are related to organised crime. However, Weber and Kruisbergen [148] also acknowledged that the dark web is still an environment with unknown risks.

In the following four subsections, we provide a detailed review of research on dark web markets from four perspectives: measurement and documentation of market characteristics, security mechanisms within the markets, dynamics of market closure, and studies of items sold.

2.4.1 Measurement on Markets

In earlier times, Christin [26] performed a comprehensive measurement analysis of the *Silk Road*, which is a well-known dark web market. Data covered eight months between the end of 2011 and 2012, containing over 24,000 items and parsed feedback messages. The author analysed the data and covered three aspects: market characteristics, vendor characteristics, and economic implications. The results indicated that drugs and controlled substances were the most frequently listed items on the market, with most remaining on the list for less than three weeks. The top three regions for vendors are the United States, the United Kingdom, and the Netherlands. This study represents the first comprehensive measurement analysis of a comprehensive dark web market. Shortly after, Dolliver [45] conducted a similar study on *Silk Road 2*. Although *Silk Road 2* attempted to fill the void left by the closure of *Silk Road*, it is much smaller than the *Silk Road*, with only 1,834 items for sale.

Similarly, Soska and Christin [123] conducted a long-term measurement analysis of 16 different dark web markets within the online anonymous market ecosystem from 2013 to 2015. The paper also shows how to crawl data to ensure completeness, soundness, and instantaneousness. In addition to the attributes of

listing items, the authors analysed the vendor's characteristics in the dark web market, including the number of vendors over time, volumes per vendor, and vendor aliases. The detection of vendor aliases is novel, and the authors utilised it to analyse the survival time of vendors on dark web markets. As an extension, analysis of aliases can be used to examine and track top vendors across different dark web markets. This paper presents some useful data points that we can incorporate into future market measurements, such as an analysis of sales volumes, product categories and vendor characteristics. The authors also present the basis and requirements when collecting data.

In 2018, Van Wegberg et al. [136] measured the development of the commodification of cybercrime on the dark web market between 2011 and 2017. The dataset leveraged parsed and analysed dataset from Soska and Christin [123], and then expanded it further using 16 snapshots from *AlphaBay* between 2016 and 2017. The commodification of cybercrime could drive cybercrime further in some ways. The authors summarised eight existing types of cybercrime business models: spamvertised products (e.g., sending spam e-mail for advertising), extortion (e.g. ransomware), click fraud (e.g. hijacked traffic), social engineering scams, fraud, mining, carding, and accounts.

More recently, Kermitsis et al. [86] analysed various characteristics of the newest and most recently popular dark web markets and independent vendor shops. For instance, the authors mentioned that vendor trust is an important characteristic of overall market trust. Escrow remains a popular trading model, which solves the trust issue between buyers and vendors, but in turn, makes the market more "convenient" for exit scams. Nevertheless, when some market exit scams, its users usually move to other alternative dark web markets within a short period. As the platform provider, the market provides cyberspace while also ensuring anonymity and security. Therefore, market operators often charge

vendors various fees to make profits, including registration fees for vendors to enter the market, deposits to establish the vendor's creditworthiness, commissions for transactions through the market, and other shop services fees (e.g. vendor promotions). These characteristics help to understand the current state of the dark web markets. The authors learned from three of the most popular information sites (*DeepDotWeb*, *Dark Web News*, and *Darknet Markets*[1]), which specialise in dark web information, that there were 14 markets active in May 2019.

Georgoulias et al. [69] comprehensively documented the features and functionality of existing dark web markets in 2021. Their paper describes and summarises the operations of 41 markets and 35 independent vendor shops and details the mechanisms for those markets' framework, which also include some security mechanisms, such as CAPTCHAs, during the registration and payment process.

While the papers mentioned above comprehensively analysed the development and characteristics of dark web markets in different time periods, there is a lack of analysis of dark web markets in non-English communities. Furthermore, Zhuge et al. [163] pointed out that the underground economy (in terms of malicious websites) in China is different from that in communities in the United States or Europe. These observations inspired our first research question stated in Section 1.2, that is, more specifically, whether there are key differences between markets operating in English and those in another language, such as Chinese.

Zhou and Zhuge [161] analysed and compared three Chinese dark web markets and one English dark web market in 2019. They mainly looked at three aspects: market operation, security, and items sold. They also considered and discussed the potential impact of policy and law enforcement on different language communities. The authors pointed out that personal information data is a popular commodity in Chinese markets, while drugs and controlled substances are popular in English markets. The results also showed that the operation of the English market was

[1] All three sites are now unavailable.

more functionally-ready, while the Chinese market was still in the process of being developed. The authors also mentioned about Chinese vendors who trade in English-speaking markets, which provides a theoretical basis for future tracking of cross-market actors.

In comparison to the work by Zhou and Zhuge [161], we have built on this work to update and expand data points to more broad aspects such as market policies, payment methods, crawling restrictions, and vendors' characteristics. Our work is presented in Chapter 4.

2.4.2 Markets Security

Data collection on the dark web is widely considered a challenge in the current studies, which is not limited to the dark web markets but also includes various underground forums [92, 131, 156]. Bergman and Popov [9] systematically reviewed 34 published research papers that contained web crawlers on the dark web. They also developed a new dark web crawler based on the knowledge in the literature, able to achieve automatic web content classification through a novel toolset. Although the results show that the crawler they developed is usable, some human inputs are still required to ensure the reliability of the data collection process.

In another study, Labrador and Pastrana [92] evaluated multiple characteristics of selling products, vendors, and markets by implementing a customised crawler. Although we noticed that the authors described a flexible crawler capable of dealing with some anti-crawling techniques for their study, the time it took for the crawler to traverse the entire market still depends on market constraints (in their case, a single market can take up to 61 hours). This is often the reason why it is challenging to conduct large-scale longitudinal studies on dark web markets. Cuevas et al. [34] also pointed out a problem that existing literature did not focus on methods of data collection. Obviously, when the aim of the research was to measure the market, the method of collecting data would affect the quantity

and the quality of the results. This is a challenge that is often faced by academic researchers. To make the challenge more complicated, DDoS attacks on dark web markets make the time period for data collection inconsistent [162], and mostly, those situations are unpredictable.

Furthermore, Campobasso and Allodi [21] proposed a trainable, scalable crawler tool that would make it possible for researchers without computer backgrounds to collect data from underground forums. In the experiments conducted by the authors, they overcame some restrictions (e.g., rate limiting and page loading time) by using different strategies when crawling different websites. It is both interesting and extra challenging that they also noticed a new and complex anti-crawling measure in the study, consisting of the dynamic obfuscation within HTML attributes (like IDs or the name of tags/classes). However, how these ideas and strategies can be deployed for mainstream dark web markets remains unclear.

On top of the papers mentioned above, while proposing a newly designed crawler, Turk, Pastrana and Collier [131] studied and summarised the anti-crawling techniques used in 26 underground forums, covering both websites in the Tor network and the clear web. The paper also classified these anti-crawling techniques and discussed some methods for researchers to mitigate these challenges. This study concluded that data collection in "adversarial" environments can be particularly challenging. Even though there are some ways to mitigate their effect, there are no easy solutions to avoid it completely. It is recommended that the academic communities should actively share datasets.

In fact, some similar anti-crawling techniques are still unseen in dark web markets in our experience. Since dark web markets tend to focus more on sales rather than forum-like discussions, we speculate that security mechanisms may be quite different in those cases. In relation to this, we feel there is a real need to investigate these security mechanisms further.

The work of Georgoulias et al. [69] also touched on the security mechanisms

of dark web markets. In our work presented in Chapter 5, we focus more on security mechanisms using more recent and broad data points. The insights and experiences we share are mostly regarding data collection.

Finally, Yoon et al. [158], as well as Güldenring and Roth [71] investigated the important problem of phishing domains/websites on the dark web. Although the research topics and objectives are slightly different, the results of these independent works consolidate our results, especially in the anti-phishing aspect of the dark web market.

2.4.3 Dynamic of Markets Closure

The European Monitoring Centre for Drugs and Drug Addiction and Europol published a poster in 2018 indicating the lifetimes and reasons for the closure of more than 100 dark web markets that offer drugs around the world [49]. The results showed that 13 markets were operating for one to two years. Nine markets were in operation for two to three years, while 14 were still active at the end of their study. The majority of markets (n=78) did not last more than a year. Similarly, Branwen [15] also maintained a table and a figure to count the number of closed-down markets, last updated in 2020. The results also show that only a small portion of markets have a lifespan greater than one year.

In 2015, *DeepDotWeb* interviewed the administrators of some of the then-active dark web markets, in order to gain their views on the state of the dark web market at that time [39, 40, 41]. The administrator of *AlphaBay* mentioned that when other markets exit-scammed, trading continued anyway, with many vendors and buyers moving to alternative markets. This was reflected in the growth in the number of users, posts and transactions after the closure of a particular market. In comparison, *TheRealDeal* was forced to close due to the arrest of some of the operators of their operation team, but relaunched after a period of time. Moreover, *Aurora Market* administrators said in a *DarkNetDaily* interview that

greedy administrators would run away with three to five million USD in around five months [36]. Ironically, this market did an exit scam after about three months.

In the early days, studies by Van Buskirk et al. [135] and Lacson and Jones [93] showed that the closure of *Silk Road* not only did not cause dark web market users to lose confidence in dark web markets, but instead led to a surge in the number of active vendors on alternative dark web markets. ElBahrawy et al. [48] investigated how the dark web market ecosystem was affected by unexpected market closures between 2013 and 2019. Their research was based on a dataset of Bitcoin transactions from 31 major dark web markets, 24 of which were closed down by scams or police raids. They also noted a rapid migration following market closures, which mainly affected smaller vendors.

More recently, Labrador and Pastrana [92] referred to a brief case study of one market closure in their paper. They analysed the trends in prices and volumes of products in the period leading up to the closure. They also mentioned the DDoS attack that preceded the closure of this market and possibly affected the economics of the market – causing prices to fall while losing trust from buyers – leading to the closure of the market.

Operations by law enforcement agencies have also had an impact on dark web markets. Décary-Hétu and Giommoni [38] analysed the effects of an operation that happened in November 2014. They found that police crackdowns were not an effective measure to reduce the volume of sales on the dark web market. However, the authors suggested that future research should expand the study of dark web markets and continue research to provide new insights into how markets are organised. Bradley and Stringhini [14] evaluated two innovative law enforcement operations targeting dark web market users in 2019. The first, Operation Hyperion, involved contacting buyers and vendors with warnings rather than arrests. The second, Operation Bayonet, utilised deceptive tactics, including shutting down markets under the guise of an exit scam and operating the market

as a honeypot. Qualitative data from *Reddit* forums revealed more extensive discussions and higher user concern during the second operation, particularly over financial losses and exposure of personal identities. Users who were affected or highly likely affected were believed to have fewer precautions, such as not applying pretty-good-privacy (PGP) encryption on the market. In summary, this paper offers some perspectives to describe the impact of law enforcement actions on the dark web community. Chan et al. [23] discussed the impact of law enforcement operations on drug dealers on dark web markets. The results showed that on *Silk Road 2*, the arrest of a major drug dealer reduced the overall transaction level and the number of dealers who remained active in the market. These three papers have shown that law enforcement operations can have an impact on existing cybercriminals in the short term.

Related to LEA operations, Ursani et al. [132] introduced the concept of anomaly detection in the context of dark web markets. It uses machine learning methods based on unsupervised learning to help analyse and understand events occurring on dark web markets. They argued that the closure of one market could cause anomalies in other markets, which law enforcement could use to thwart the dark web market community. Although this is a very promising effort, the limitation is that it can not do much about emerging markets.

While the reasons for market closures vary, we realise that the trend toward market closures has remained the same since the early years. However, there is a lack of recent research on new modern markets on this topic. Therefore, using data from the dark web market itself before the market closure could provide more accurate results. In addition, we have also increased the data collection of the market-associated forums to expand our insights with more details. Chapter 6 shows our findings to contribute to this research gap.

2.4.4 Items Sold on Markets

There is a wide variety of illegal online activities or cybercrime taking place on dark web markets. Some researchers have done an in-depth study focusing on a single type of item sold in the dark web markets. For example, drugs are the type of item widely sold on dark web markets [46, 99, 119]; another type of popular item is firearm [16, 32, 120]; stolen data is also a major sales category on dark web markets [33, 77, 107, 137]. During the worst period of the Covid-19 pandemic in 2020, Bracci et al. [12] analysed 788 listings directly related to Covid-19 products from 30 dark web markets over a period of approximately eleven months. The authors highlighted the importance of continuous monitoring of dark web markets, which are also rapidly adapting to global events.

In addition, the following studies focused on different types of items sold on dark web markets. Gañán, Akyazi and Tsvetkova [67] studied business fraud related items on dark web markets through longitudinal data on eight major markets (*Agora*, *Alphabay*, *Black Market Reloaded*, *Evolution*, *Hydra*, *Pandora*, *Silk Road 1* and *Silk Road 2*) between 2011 and 2017. The dataset from the IMPACT project [82] includes 44,671 product listings and over 56,000 transactions, of which 2,318 listings correspond to business fraud. The authors estimated a total loss of revenue of up to $7.5 million over the entire analysis period. Compared to other cybercrime activities, the victims of business fraud are usually enterprises rather than individuals. This paper provides insight into the economic impact of a single type of cybercrime on victims. Georgoulias, Yaben and Vasilomanolakis [68] comprehensively studied the cybercrime related products in the dark web markets. Those products include stolen user account information, fraudulent documents, malicious software, and other cybercrime-oriented services (i.e. cybercrime-as-a-services [98]). Georgoulias et al. [70] also focused specifically on botnet-related products on 26 dark web markets. These works demonstrate that items sold on dark web markets have the potential to facilitate a variety of criminal activities.

Those approaches can also be applied to a broader range of cybercrime on the dark web in order to measure the impact of the dark web markets.

In addition, current child sexual exploitation and abuse (CSEA) literature on the dark web mainly relates to the analysis of online forums, communities and networks of offenders and their group dynamics. There is still a gap in understanding the markets of child sexual abuse material (CSAM) not protected by private channels, e.g., in terms of trends, types of products on offer and their economic aspects.

Kloess and Bruggen [88] conducted a systematic literature review on the dynamics of CSEA's forums on the dark web from the perspective of trust. They found that a delicate balance takes place between distrust (i.e., the risk of betrayal among forum participants) and trust building. The latter is evidenced by cooperation (e.g., exchange of advice), reputation (e.g., elaboration on past criminal activities), and engagement (e.g., production and sharing of novel CSAM). Woodhams et al. [153] analysed the behaviour of 53 suspected CSEA offenders, who were active in four invite-only Tor forums on the dark web, which enforced specific requirements for membership. Their dataset was composed of forum postings, chat and private messages released by law enforcement spanning across two years (2014-2016).

The pandemic has increased the prevalence of "live distance child abuse" including the capture of self-generated CSAM resulting from manipulation and blackmailing of children, contributing to offerings on the dark web [54]. Findings by Liggett et al. [96] suggested that a big portion of CSAM is distributed in the dark web free of charge. A smaller portion, however, is commercially exploited in markets, including services related to physical offences (such as in the context of tourism or travel and trafficking), purchases per download (where monetary value is proportional to novelty) and revenue per clicked advertisements. Dalins, Wilson and Carman [35] identified that 31% of 232k pages from 7,651 Tor virtual

domains crawled in the dark web were related to CSAM "market/for sale". They focused on motivation and broad categorisation of content, but did not analyse any actual markets.

The use of dark web markets for distributing CSAM is obviously an appalling crime that is very challenging to address. As such, we endeavour to contribute in the effort to combat this crime through a case study investigation presented in Chapter 7, which reports our latest findings and areas for further research.

2.5 Chapter Conclusion

This chapter has presented various aspects of the literature related to dark web markets. The ways in which criminals continuously adapt to challenging environments and innovate to achieve their goals are undoubtedly dynamic. In summary, we found the following key research gaps:

- Most research in this area focuses on English-speaking communities. While this is understandable, it may limit the generalisability of the findings and introduce potential bias. Thanks to the increasing openness and accessibility of the internet, examining online communities in different languages could potentially provide valuable insights. This exploration helps the academic community better understand the dynamics of cybercriminal groups within diverse cultural and social contexts.

- Data collection within the domain of underground economies and communities presents a significant challenge, regardless of both the dark web and surface web. In this context, the academic community lacks a clear understanding of the security mechanisms that operate within dark web markets. Gaining this understanding is essential for developing effective targeted countermeasures and further understanding the operation of the dark web market.

- The closure of one market only indicates the end of that one market's lifecycle. However, from a broader perspective, the economy of dark web markets has continued to grow, seemingly unaffected by the closure of individual markets. More investigation would be valuable in describing what happens in this dynamic.

- There is a wide variety of items traded on dark web markets, but there is no conclusion on whether serious illegal items, such as CSAM, are being sold on these platforms. This situation highlights the need for further exploration into the prevalence of those items and any compelling evidence that supports or refutes their existence in these markets would be valuable.

Chapter 3

Methodology

3.1 Introduction

The previous chapter described the background of the dark web, including its technology and terminology. It also introduced the underground ecosystem and its community, and finally focused on the dark web market community. In this chapter, the main focus is on the overall methodology of the study, which includes the overall design, the dataset and the infrastructure for data collection. We begin with a short discussion of the strengths of the chosen approaches. We describe the scope of the data collection. We show the design and logical flow of the crawler we implemented. We present an overview of our dataset, although more detailed information is provided in the following main chapters. The challenges of this research approach are also briefly discussed. Lastly, and equally important, we present our ethical considerations regarding this research.

3.2 Research Design

In the literature mentioned in the previous chapter, many studies rely on datasets shared by other researchers. While this approach may sometimes reduce barriers

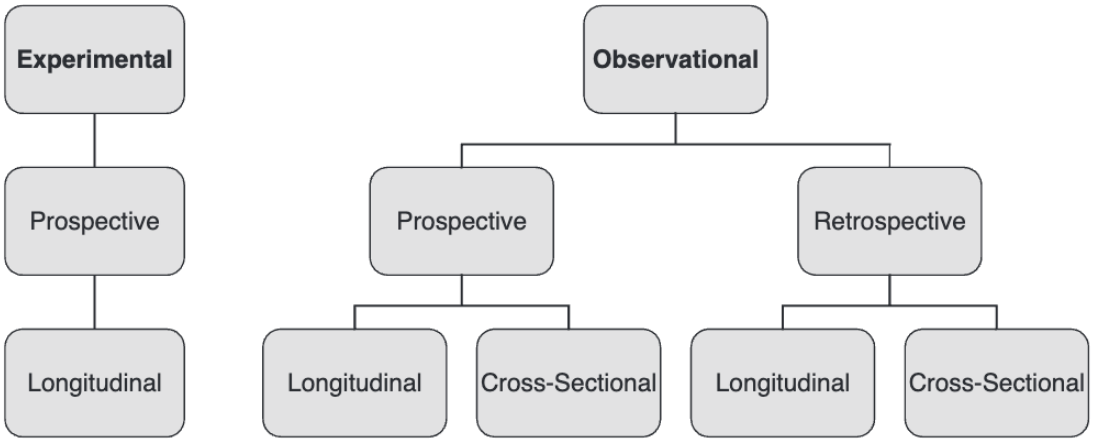


Figure 4: An overview of general research design methods, figure obtained from [1]

to conducting research, it also narrows the scope, timeframe, and flexibility of the study. Naturally, this depends on the research objectives. We adopted a longitudinal approach and collected our own dataset, hoping to make new discoveries that could be used to address the research questions posed in Chapter 1.

Albery and Munafò [1] outlined various major research designs used in health psychology that can also be commonly used in wider areas. Figure 4 shows these main ways in which a study can be designed. We defined and carried out our study based on a combination of observational, retrospective, and longitudinal approaches. **Observational** refers to the data collected from participants during the research process without influencing or intervening in these data. **Retrospective** means we are more interested in what has already happened. **Longitudinal** means our study consists of multiple "snapshots" on dark web markets over time, rather than a single "snapshot" at a single moment. Similarly, the studies by Christin [26], Soska and Christin [123], and Van Wegberg et al. [136] confirm that those approaches are effective and promising in the area of studying dark web markets.

In terms of data and data analysis, we used both quantitative and qualitative approaches. We mainly used quantitative data to present changes in market data

over time, such as product listings, sales, etc. Moreover, in some cases, we employed qualitative analysis to gain insights into market operations and policies. Quantitative data provided meaningful insights into trends through descriptive statistics. Qualitative methods are used for analysing market security mechanisms, policies, and functions. The combination of these approaches complemented each other, leading to more comprehensive findings.

To build our dataset, we conducted weekly data collection between July 2021 and August 2024, which included both quantitative and qualitative data. This was done approximately every Monday, but under certain unexpected factors (e.g., market server unavailable, Tor network performance issues, etc.), we tried the next day or skip a week and resumed as soon as possible. Although the weekly data volume fluctuates, we consistently documented valuable notes for later review and analysis as part of the qualitative data. We also designed and implemented a customised crawler for data collection. In Section 3.3, we introduce this crawler and its technical environments.

Our market selection was primarily based on the popularity (reputation and discussion) observed on the dark web forum *Dread* [87], a well-known *Reddit*-like dark web forum. We also tried to obtain the name and basic information of the dark web market through public onion service directory websites (e.g., *dark.fail* and *TorTaxi*). While this approach was relatively subjective, which could introduce some bias, the highly unpredictable nature of the dark web markets made it challenging to forecast their development. Therefore, prior to data collection, we relied on commonly mentioned, well-known mainstream markets as the basis for our selection. This approach has also been commonly used in the literature, such as [68] and [92].

Dread operates independently from specific markets. Similar to *Reddit*, *Dread* features various sub-forums, some of which are related to those markets we observed. Typically, market operators also serve as moderators for these sub-forums,

but the admins of the *Dread* are actively monitoring those sub-forums. Therefore, in addition to the dark web markets data, we also collected data on market-associated forums on a one-time basis in some of our works. The aim of getting data from forums is mainly to obtain additional information and offer more valuable insights. The reason for taking only a single snapshot is that the content in the forum tends to remain quite stable. Furthermore, if the market ever shuts down for any reason, the forum archives all threads, ensuring that the content remains unchanged and is preserved with timestamps. Detailed information is provided in Chapter 6. The process of data collection will also be described in Section 3.3.

3.3 Customised Crawler

In order to apply the research approach we described above, we implemented a customised crawler. The crawler was written in Python with the Scrapy web-crawling framework [90] and the Selenium suite of tools [122]. The Scrapy framework was used to implement the crawling, downloading, and partly parsing functions, while Selenium was integrated into Scrapy to provide an interface for login, CAPTCHA solving, and other processes that required human interaction. This solution ensured that Selenium performance limitations would not significantly slow down the actual crawling speed. Furthermore, in some cases, by leveraging Selenium's ability to handle sessions automatically, the crawler could deal with the use of dynamic cookies in certain markets (i.e. each request would apply a new cookie based on the previous request). Figure 5 depicts the workflow of our crawler. There are nine steps that our crawler had to perform before obtaining the data:

1. The crawler was hosted in a virtual machine container (i.e. the host). In this step, we needed to establish a Tor connection from the host. Privoxy was used to relay between Tor connections and Scrapy, because they support

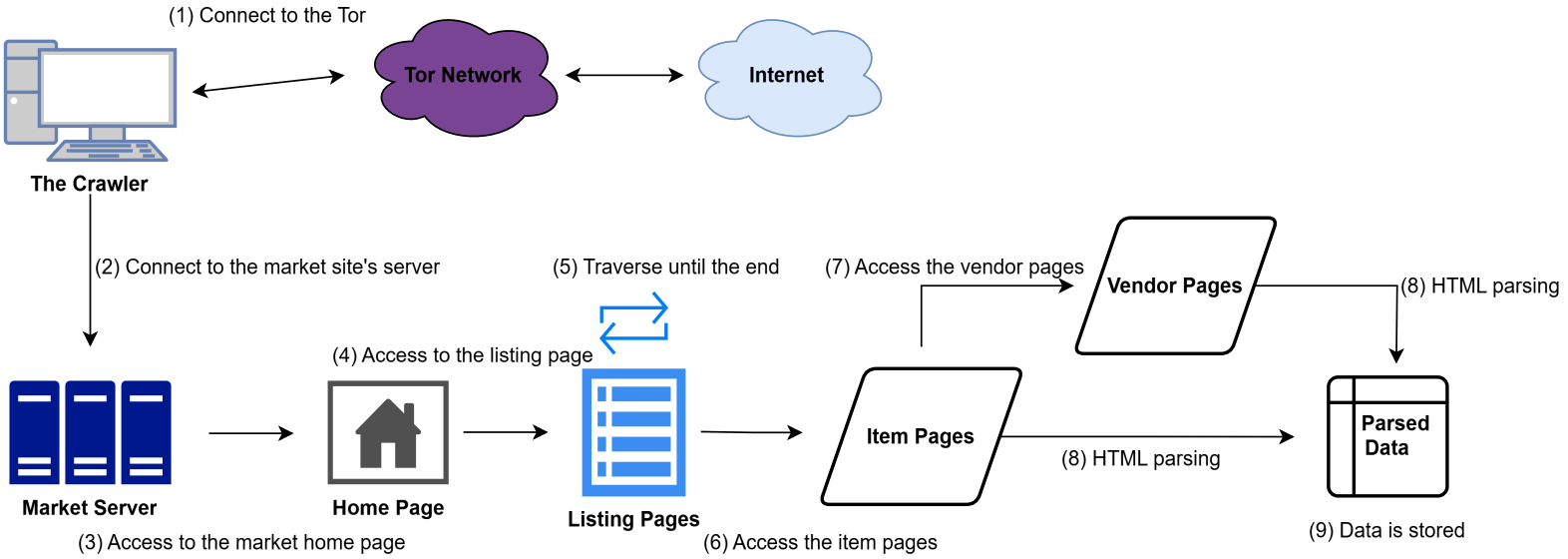


Figure 5: The workflow of the customised crawler

different proxy network protocols.

2. Selenium was integrated into Scrapy. Then, Selenium was used to initialise and establish a connection to the market server.

3. Human interaction was required here to solve the CAPTCHAs, to log in to an account, etc. After that, Selenium passed the web session back to Scrapy to automate the subsequent steps.

4. Depending on the structure of different websites, the market usually has its own peculiar style and logic in arranging and displaying the products on sale. The crawler would look for the entry to the listing page on the home page. The URL of the listing page could also be defined in the crawler, in advance.

5. This step sorted the products appropriately on the listing page, and then traversed through all of the listing pages of all items on the website.

6. Depending on the market, most of the information that we cared about (such as price, sales volume, feedback, and product description) would be included on the item page.

7. The crawler would then find the vendor information page on the item page and access it via Scrapy.

8. This step parsed our target data (depending on the purpose of the study) in the HTML code using methods such as XPath.

9. Finally, the data was saved in the host computer for later analysis.

Thanks to the flexibility provided by Scrapy, the crawler can be easily modified as requested. For example, in our study, we can access all the links on the homepage (based on HTML) in Step 4, and then further confirm whether those

pages contain any valuable information. Scrapy can also handle URLs that are repeatedly visited and supports customised filters. When downloading data, we can use the built-in middleware function of Scrapy to determine whether the response code of the website is valid (i.e. between Steps 7 and 8). When the website returns a non-valid response, the crawler can call the Selenium component for manual interaction and inspection. Moreover, to the best of our knowledge, the class names used in steps 5-8 are static in terms of path or ID in the HTML code, thus indicating that the sites do not apply crawler obfuscation traps (unlike those used in underground forums [131]). Similarly, we did not find any malicious/phishing links targeting the crawler, which may benefit from how the crawler obtains links (i.e.∼using predefined HTML tags instead of getting all URLs in the code). The described process also prevents crawlers from getting stuck in program deadlocks (infinite loops) [37].

In practice, the customisation component required efforts in the following three parts:

1. Complete various challenges designed by the market and log in before landing on the home page

2. Design an effective crawling path for the crawler based on the internal structure of the market

3. Store data (web pages) in a virtual machine, or parse the target data when needed

The completeness of the data collection would also depend on the market's security mechanisms, which are part of the market characteristics. We will discuss this matter in Chapters 4 and 5. For minimal working examples implemented in Scrapy, see Appendix A.

For data collection on *Dread*, the procedure was slightly different but more efficient and streamlined compared to the procedure for market data collection.

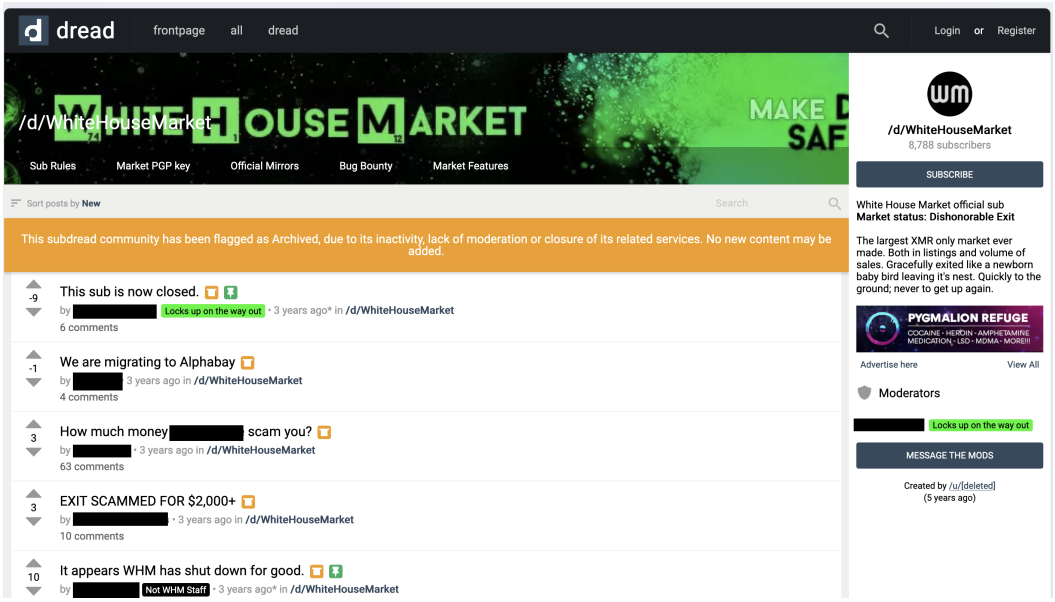


Figure 6: An example of the sub-forum structure from */d/WhiteHouseMarket*

We started with locating the names and links to the sub-forums of those markets. Once the crawler entered the sub-forum, it would get the structure shown in Figure 6, where there were multiple pages consisting of multiple threads in the sub-forum. Then, all the crawler needed to do was visit each thread and get the data within the thread.

In addition, we also performed some experiments with our crawler while learning about the security mechanisms of dark web markets. For datasets with specific requirements (such as a case study about CSAM in Chapter 7), we made some extra efforts to ensure that our ethical considerations were met.

3.3.1 Manual Intervention

While crawlers are designed to automate data collection, there are cases where manual intervention is still necessary. After connecting to the market server, we manually entered the website's credentials. Although this process can be automated, the web structure may vary for each market (e.g., having multiple pages

before entering the homepage). Therefore, it is often more efficient and straightforward to manually enter the information before logging into the market's homepage. During log-in, we also needed to solve CAPTCHAs, which typically involve more than one. Additionally, due to unforeseen circumstances, such as market availability issues, the crawler may require a new session during data collection, necessitating a repetition of the manual steps mentioned above. All other tasks can be automated, and we can monitor the access rate (pages per minute) in real-time. More details and discussion can be found in Chapter 5, where we discuss the market's security mechanisms.

3.3.2 Technical Environment

Figure 7 shows the data flow of our data collection scheme. We connected to the Tor and/or I2P network using a virtual machine for secure and anonymous access to the dark web markets. The virtual machine ran Ubuntu LTS with four cores, eight-thread, 8 GB RAM, 200 GB storage and a NATed network from the host computer. A VPN tunnel was established on the host computer and forwarded to the virtual machine through the NATed network. The data pipeline was therefore wrapped in different layers. In practice, the VPN traffic wrapped the Tor traffic, i.e. onion-over-VPN, which reduced the chance that the host's real IP address would be seen at the Tor entry. At the same time, security was further enhanced by NAT the internal network (i.e., the virtual machine was not actively discoverable). Since Tor uses SOCKS5 proxy and Scrapy uses HTTP proxy, Privoxy was used to relay Tor and Scrapy. For I2P, we simply replaced the Tor component with I2P. Once collected, the data was password-protected, saved into local devices, and backed up regularly.

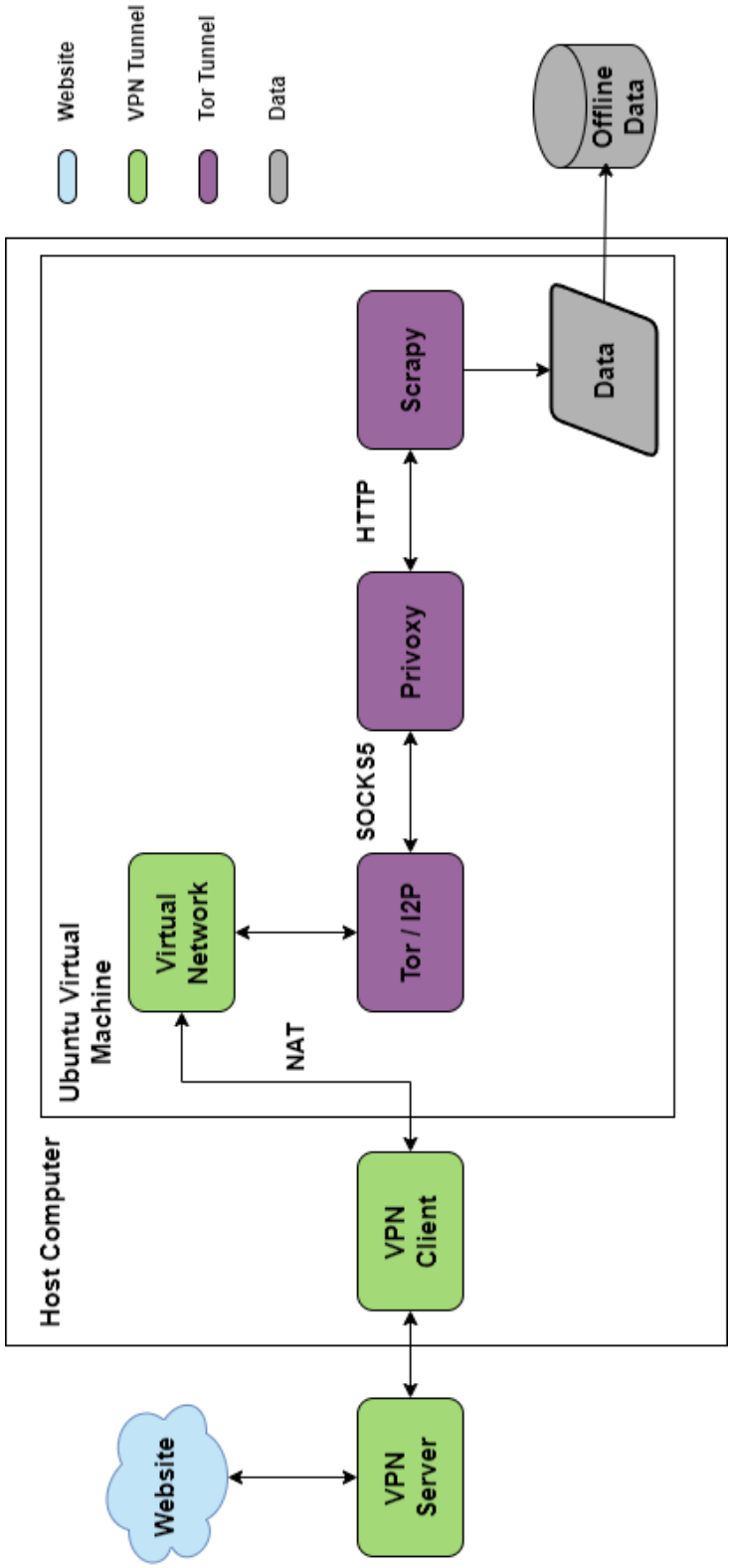


Figure 7: The data flow of the data collection scheme

Table 1: Summary of the 14 dark web markets obtained through our crawler for quantitative data

Markets	Observation dates (from/to)	Number of snapshots
AlphaBay Market	2022-06-06 / 2023-02-13	36
Archetyp	2023-04-17 / 2023-08-29	19
ASAP Market	2022-01-18 / 2022-04-20	15
Bohemia	2023-05-02 / 2024-01-29	40
Cartel Marketplace	2021-07-05 / 2021-12-20	24
Chang'an Nocturnal City	2022-03-01 / 2024-08-26	130
Chinese Exchange Market	2021-07-19 / 2024-08-26	162
Dark0de Reborn	2021-07-05 / 2022-02-21	33
Loulan City	2022-01-11 / 2022-02-14	6
Monopoly Market	2021-10-18 / 2021-12-27	11
Nemesis Market	2023-08-31 / 2024-03-18	30
Tea Horse Road	2021-07-09 / 2021-11-16	18
Versus Project	2021-10-18 / 2022-05-16	31
White House Market	2021-07-05 / 2021-10-11	14

3.4 Dataset

Table 1 shows the quantitative data obtained using our crawler, which includes 14 dark web markets. Those data can also be used for qualitative analysis. We collected five sub-forums from *Dread* related to the markets we observed in August 2022. Later, we took another snapshot of the entire *Dread* in December 2023.

We also collected qualitative data from seven other dark web markets to enrich our dataset for answering our research questions. The seven additional dark web markets are: *Abacus Market*, *Colombia Connection*, *Incognito*, *Kingdom Market*, *Royal Market*, *Tor2door Market*, and *Vice City Market*. This brings the total number of dark web markets we observed to 21.

Most of the data was not parsed when it was collected, instead, we quickly saved each market's HTML file. Technically, this would ensure that we have

relatively rich data points when solving our research problems, and we could also process the data as needed. The dataset used in this thesis will be shared ethically with the academic community and security researchers (see Appendix B).

Overall, we collected long-term data spanning three years from 2021 to 2024. However, depending on the research question, the datasets we used in the following chapters cover smaller subsets (and potentially different parts) of the whole dataset. These subset datasets were chosen to be scientifically meaningful for each specific research question. The specific parts of the dataset used are detailed accordingly in the following chapters.

3.5 Challenges

We faced several challenges when implementing the research method presented above, in order to address the research questions we raised in Chapter 1. First, the dark web market uncertainty could lead to less efficient data collection. Sometimes, when markets frequently updated their anti-crawler or other security mechanisms, we had to adapt the crawler in time to cope with these changes. In turn, due to time constraints, it was difficult for us to verify and quantify their changes in a short period of time. The dark web markets might shut down or suspended their operations at any time for various reasons, leading to potential discontinuities in our data collection. In addition, since the establishment of a market is typically quite secretive, it was difficult for us to obtain data on the entire life cycle of a market from the very beginning.

3.6 Ethical Considerations

Ethical considerations in research are extremely important in this area. Since we collect data on the dark web (Tor network and I2P network), which may be

related to potential cybercriminal activities, we must handle the ethical issues of our research very carefully.

Our datasets contain items such as product information and discussions from the dark web markets and their associated forums. Since this information is inherently public, we consider it part of the public digital space. Even though most dark web markets require registration, the process is open to the public. When registering, our usernames and other information are not linked to any individual or organisation. We disclosed the names of these dark web markets because they are well-known within the dark web community. We believe this will help academia and law enforcement agencies better understand the trends in mainstream dark web markets.

Moreover, due to the anonymous nature of the dark web, we were unable (nor interested) to collect the personal information of users, or track their real identities. Nonetheless, we still had to anonymise the usernames of the users (if mentioned), because it might be possible to use these usernames to connect back to their real identities.

During data collection, we applied dynamic delays to our crawler, in order to prevent additional server stress to the observed sites (i.e. we did not want to disrupt or interfere with their operation as researchers).

The ethical aspects of this study have been reviewed and approved by the Research Ethics Coordinator of the University of Kent's Central Research Ethics Advisory Group (reference number: 057-04-2021), as detailed in Appendix C. The final ethics approval was made in July 2021.

More specific considerations that are directly relevant to the study (such as CSAM topics) are pointed out in the methodology of the corresponding chapters.

3.7 Chapter Conclusion

This chapter presents the overall research methodology, including the reasons for choosing a mixed-methods approach. The combination of quantitative and qualitative methods also helps ensure the robustness of the analyses and the results. The implementation of our customised crawler was informed by the proposed research methodology. We also briefly describe some of the challenges faced. Ethical considerations are also taken into account to minimise potential data collection risks. In the following four chapters, we present the results and insights gained from applying this approach to answer the four research questions outlined in Chapter 1.

Chapter 4

Comparative Analysis of English and Chinese Dark Web Markets

This chapter is based on the content of previous publication:
"Toad in the Hole or Mapo Tofu? Comparative Analysis of English and Chinese darknet Markets" [142]

4.1 Introduction

The popularity of online shopping and cryptocurrency has contributed to driving the economy of dark web markets in recent years. These markets are often perceived to be conducive to (or may even facilitate) cybercrime activities. Therefore, it is worthwhile to have a deeper understanding of how various dark web markets operate so that researchers and law enforcement agencies can test and deploy appropriate countermeasures to fight online crime.

Previous studies have measured a wide range of dark web markets, but they are heavily skewed towards those conducting their business mainly in English. However, English dark web markets are not the only platforms of interest here. Investigations into other language dark web markets – such as those in Chinese

or Russian – are still scarce, and this is a gap that needs to be addressed. Furthermore, and due to the rise of cryptocurrency despite their hard stance towards bitcoin, Chinese LEAs have also increased their effort to fight criminal activities on dark web markets [28]. The development and acceleration of economic globalisation have also made it necessary to study the diversity and impact of dark web markets in different regions of the world.

The study presented in this chapter aims *to investigate, analyse and compare several current popular English and Chinese dark web markets*. We collected datasets from five active and popular dark web markets, three operating in English and two in Chinese. Through statistical analysis and in-depth investigation, we track some indicators and come up with a summary of key characteristics, and how these characteristics compare between the two different languages markets. We described and compared the results of our investigation in six main aspects: (i) operation model and structures, (ii) product categories, (iii) market policies, (iv) payment methods, (v) security mechanisms, and (vi) vendors' characteristics. We share our insights into market development and vendors' behaviour, which are helpful for future investigations.

4.2 Methodology

This section outlines our approach and provides an overview of the data collection process, including a description of the two crawling strategies and a detailed outline of our datasets. We also briefly discuss ethical considerations at the end of this section.

4.2.1 Approach

We constructed our datasets during a seven-week period between 17 July and 30 August 2021, containing data from 384 vendors on the English dark web markets

and 4,429 on the Chinese dark web markets.

Data on the *operation model and structures* include the markets' basic information, as well as changes in the market size and the possible reasons behind such changes. The *product category* displays the proportion of different types of items on the markets based on the average number of each snapshot. The more comprehensive data available on the Chinese dark web markets means we can also estimate the revenue generated by the main product categories. The *market policy* mainly focuses on what goods or services are explicitly banned. The *payment methods* describe what currencies are accepted. The *security mechanisms* describe the crawling restrictions in each dark web market, as well as general account security. In terms of *vendors' characteristics*, we analysed vendor location, trust level and active/inactive status and time. We also selected some top and cross-market vendors on the English dark web markets, and some top vendors on the Chinese dark web markets. We define "top" as the ones that make more profits, have more sales and have more positive feedback or better reputations during the observation period. Top vendors are sometimes defined and shown on the home page of the dark web markets. Top vendors are mostly calculated and selected based on the materials mentioned above. Cross-market vendor refers to a vendor that sells at the same time on multiple markets. We were particularly interested in the vendors' behaviours and operating model of cross-market vendors when comparing. Showcase ratings from other markets are a market function used by vendors to build their brand. Some markets also allow vendors' activities in other markets to be included in the calculation of the reputation system [69]. This feature works as a way to attract vendors. This feature also helped us to identify the cross-market vendor. Finally, we highlight some stark differences between English and Chinese dark web markets based on the six aspects above.

Table 2: Summary of the observed dark web markets (for the comparative analysis of English and Chinese markets)

Market	First seen	# Listings	# Vendors	Lang
Dark0de Reborn	2020-05	45876	1648	EN
White House Market	2019-08	44740	3453	EN
Cartel Marketplace	2020-06	2596	195	EN
Chinese Exchange Market	2018-03	10949	2636	CN
Tea Horse Road	2020-04	8302	1793	CN

4.2.2 Data Collection

We discuss the design of our crawler in detail in Chapter 3. Depending on the restriction policies of different markets and the information contained in each market, we used two sets of strategies for this chapter:

1. If the market had stringent anti-crawl measures, the website would take a long time to crawl, or/and the session might expire during crawling. In this case, we would only collect data based on what is shown on the website home pages. For instance, most websites display highly rated vendors, promotional products and featured listings. Therefore, we could get selected vendors' data.

2. If the market had less stringent protections and restrictions, we would try to get as much information as possible through the listing pages.

The listing page URLs can usually be traversed easily in both strategies because their URLs are generally sequential. The product pages and vendor pages are partly obtained and parsed depending on the dark web market website structure. Crawling restriction details are described and compared further in Section 4.3.

The data was collected once a week to avoid stressing the markets' website and being as inconspicuous as possible. In some circumstances, e.g. a DDoS attack on the website, the data collection was slightly delayed.

Table 2 provides a summary of the observed dark web market names, when they were first seen, the number of active listings of products, the number of active vendors, and the language used. The active numbers are as of 30 August 2021.

Markets Introduction

Dark0de Reborn is one of the English-based dark web markets, and it started in the early days of the Covid-19 outbreak, 24 May 2020. It has the most number of listings, and vendors are able to import feedback scores from other popular markets. Therefore, it has attracted a large number of vendors with a good reputation. The market has strict crawling restrictions. A single session will expire in about an hour, and the number of requests is also limited. We focus on the top vendors (an average of 20 top vendors per week) and collect data from their product pages. Data includes the vendors' profile pages and some of the feedback received. We also collected some statistical data to study market trends.

White House Market (WHM) is one of the most popular dark web markets in the English language. The market has been in operation since 24 August 2019. A decline in *Empire Market*'s reputation led to a rise of *WHM* [115], which has a very high reputation. It also has the strictest crawling restrictions. The market only allowed a limited number of requests in ten minutes. We used six accounts to crawl to ensure that enough data was collected before the session expired. We were only able to collect top vendors (an average of 25 top vendors per week) mentioned on the homepage and some statistical data. Data includes vendors' profile pages and their product listings.

Cartel Marketplace is a medium size dark web market in English. Although it is not as large as the previously mentioned markets, it still has a good reputation in dark web forums. So we think it should be included in the research. It has a relatively less stringent crawling policy, which allows about 300 pages per session within about 40 minutes to 50 minutes. We used three accounts to crawl. The

product URLs are more likely random or coded, which means we had to traverse the whole listing with the page numbers. We saved all product URLs in a list, then sent requests accordingly. We were able to collect all vendor's information (an average of 186 vendors per week) and listing pages on this market.

Chinese Exchange Market is the most active and oldest dark web market in Chinese. It was developed from a forum. It has a less stringent crawling policy, and the cookie structure is also very simple. The market has no restrictions on the number of requests per session and is given a long lifetime of a single session. Therefore, we collected all active listing pages and parsed them. The product URLs are sequential. It can be traversed easily using a brute force approach.

Tea Horse Road represents the new generation of Chinese dark web markets. It has a user-friendly interface with an innovative "request-to-buy" model. It is a market with several historical versions, the earliest can be traced back to October 2019. The current version of the market was launched around April 2020. It also has a less stringent crawling policy (similar to the Chinese Exchange Market), although the structure of the cookies is a bit complicated. Dynamic cookies are used, which means we need to change the cookie every time we send a request (this cookie value is obtained from the previous response). We managed to collect all active listing pages because the product URLs are, again, sequential so they can be traversed exhaustively.

Table 3 compares the indicators and features of the collected markets. The symbol ✖ means either the indicator does not exist on the market or it was not collected due to crawling limitations. Overall, a total of 1,968 pages and 168,398 listings were collected. They contain 143 pages of vendor information and seven status pages (one per week) on *Dark0de Reborn*; 516 pages of vendor information on *WHM*; 1,302 pages of vendor information and 20,776 listings on *Cartel Marketplace*; 75,224 listings on *Chinese Exchange Market*; 72,398 listings on *Tea Horse Road*.

Table 3: Comparison of the indicators and features of the collected markets

		Dark0de Reborn	White House Market	Cartel Marketplace	Chinese Exchange Market	Tea Horse Road
Vendor Characteristics	Username / ID	✔	✔	✔	✔	✔
	Profile	✔	✔	✔	✖	✖
	Member Since	✔	✔	✔	✔	✖
	Last Seen	✔	✔	✔	✔	✖
	Total #Sales	✖	✔	✔	✔	✔
	Rating	✔	✔	✔	✖	✖
	Disputes	✖	✔	✔	✖	✖
	Feedback	✔	✔	✖	✖	✖
	PGP	✔	✔	✔	✖	✖
	#Listing	✔	✔	✔	✔	✔
Product	Title	✔	✔	✔	✔	✔
	Price	✔	✔	✔	✔	✔
	Category*	✔	✔	✔	✔	✔
	Sales	✔	✖	✔	✔	✔
	Shipping From	✔	✔	✔	✖	✖
	Shipping To	✔	✔	✖	✖	✖
Market	Total #Vendors	✔	✔†	✔	✔	✔
	Total #Products	✔	✔	✔	✔	✔
	BTC	✔	✖	✔	✔	✔
	XMR	✔	✔	✔	✖	✖
	Fiat Currency	✖	✖	✖	✖	✔
	CAPTCHA	✔	✔	✔	✔	✔

*** This is not for each product, but as an overall number. † Shown only in real-time**

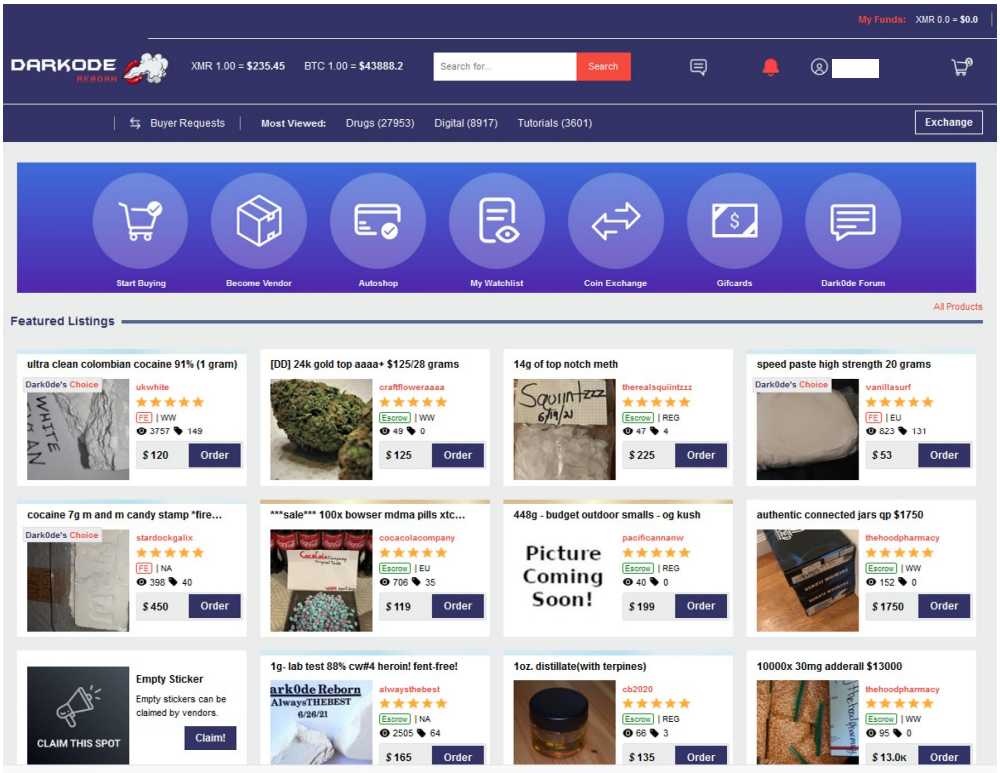


Figure 8: *Dark0de Reborn* homepage

4.2.3 Ethical Considerations

Since our study collected data from activities that could potentially be related to cybercrime – such as drug dealing, sexual abuse and exploitation of vulnerable groups and other criminal activities – we had to ensure that we obtained ethical clearance before we commenced our study. Please refer to Section 3.6 (in Chapter 3) for further details on ethical considerations during data collection.

4.3 Results

In this section, we present the results of our comparison of the dark web markets in English and in Chinese. The findings are divided into market, vendors, market policies, payment methods and security mechanisms.

4.3.1 Dark Web Markets in English

Market

Operation structures. All three English dark web markets have a website interface similar to *Silk Road*. Figure 8 shows the *Dark0de Reborn* homepage. They all have multiple sections on the homepage, including promotional products, trending products and recommended top vendors. There is usually a category list on the homepage, and users can browse all products under this category. Most web pages also support a mobile-friendly interface.

Number of listings and vendors. The number of listings and vendors indicates the state of a market. Figure 9 shows the number of listings on *WHM* between 18 July and 30 August 2021[2]. Figure 10 shows the number of listings and vendors on *Dark0de Reborn*. The number of vendors on *Dark0de* contains all vendors, even if the vendor does not have any active listings. The overall listing number keeps rising, accompanied by some fluctuations on both markets. *WHM* and *Dark0de Reborn* introduced a new "Product Quality and Harm Reduction Program" on 17 June 2021. The program aims to provide high-quality products and lower levels of risk to buyers. Vendors who send products and receive positive test results can get badges and display them on the product page to increase credibility, facilitate trust and increase sales. Also, the vendors can apply for a reduced fee on *WHM* if they test regularly. This program causes some non-compliant products to be removed. Figure 11 shows the number of listings and vendors on *Cartel Marketplace* over the observation period. Those two lines are consistent in most weeks. As the number of vendors increases, the number of listings also rises. On *Cartel Marketplace*, the data collection strategies may cause fluctuations in the number of vendors. If the vendors did not list anything, are on

[2]The vendor information on *WHM* is shown only in real-time. Hence we do not have historical information on the number of vendors on *WHM*. Furthermore, in early May 2021, *WHM* stopped accepting any new vendor applications, making the number of vendors plateau at about 3,450. In comparison, *Dark0de Reborn* maintained a continuous vendor growth.

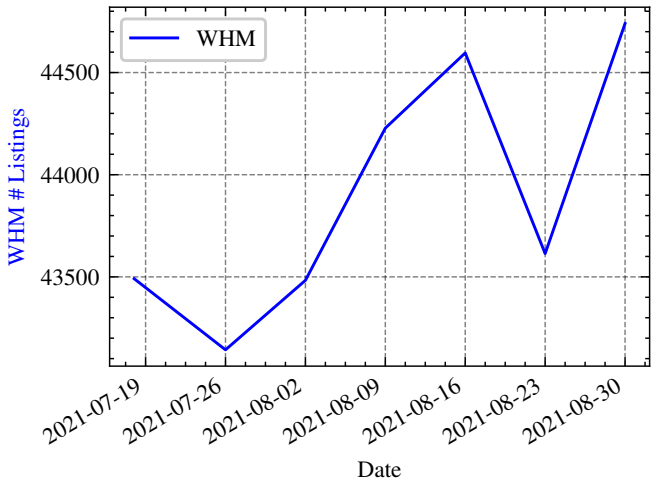


Figure 9: Number of listings on *White House Market*

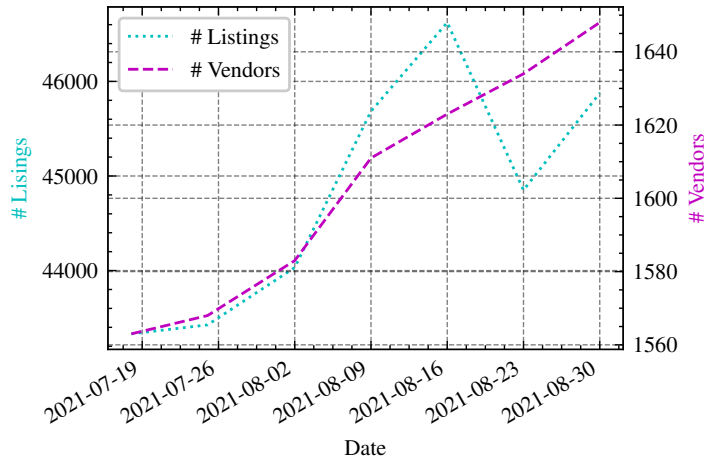


Figure 10: Number of listings and vendors on *Dark0de Reborn*

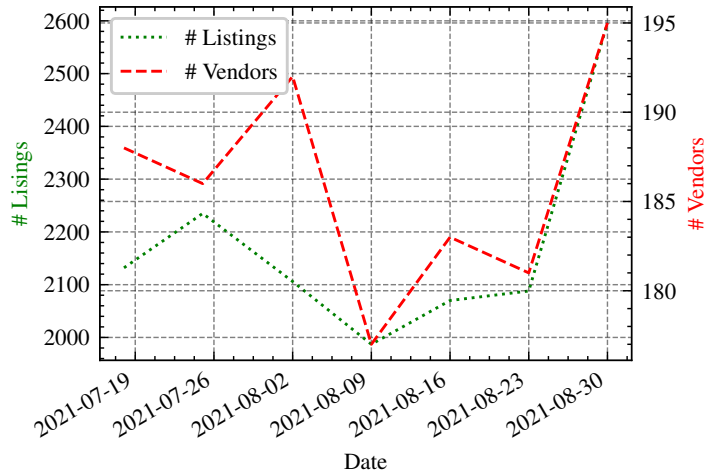


Figure 11: Number of listings and vendors on *Cartel Marketplace*

vacation status, or sell items that are out of stock, their posts would not appear on the listing pages. Hence, the number of vendors should be considered as the number of active vendors when the data is collected.

Product category. Figure 18 shows that drugs take the largest proportion of the three dark web markets. We count those by using the website information listed in the navigation interface. Drugs account for over 66% of products, followed by digital products and fraud-related materials. Digital products contain pirated software, exploit kits, digital services (e.g., DDoS services), botnets and malware. Another category of trending products in all three markets is physical items such as smart devices, jewellery, and watches. According to the vendors' descriptions, most of them are fakes or reshipping drops.

Vendor

Inactive days. Figure 12 shows the joining date and inactive days of 264 vendors from *Cartel Marketplace*. We define the number of inactive days as the number of days from the last seen date to the date when the data was collected. A few vendors were not very active over 50 days, but most vendors were still active, even if they had been registered for more than one year. Over 89% of the vendors have appeared within the last ten days. Even though the total number of vendors on the market is small, they seem to be very active.

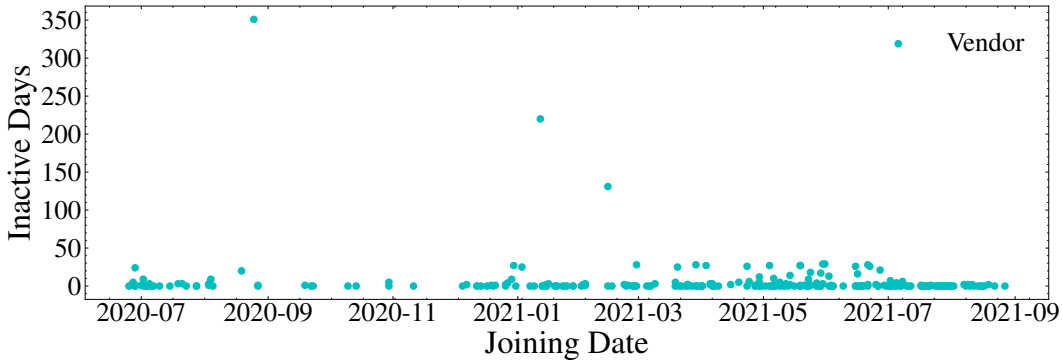


Figure 12: Vendors' inactive days and joining date on *Cartel Marketplace*

Table 4: The proportion of vendors' locations on English dark web markets

Location	Number	Percentage
United States	124	32.29%
Worldwide	65	16.93%
United Kingdom	38	9.90%
Europe	37	9.64%
Netherlands	32	8.33%
Germany	19	4.95%
Australia	19	4.95%
N/A	11	2.86%
France	7	1.82%
North America	7	1.82%

Location. On all three English dark web markets, we collated the location of 384 vendors based on their profiles. Table 4 shows the top ten countries and regions where those vendors are located, covering 359 vendors, or 93.49% of the 384 vendors. The United States and Europe are the main supply locations. Approximately 62% of European vendors are from the United Kingdom, the Netherlands and Germany. This result is consistent with previous reports [50]. In terms of "Worldwide" or "N/A", these mean that some vendors do not want to reveal their location, particularly when they only sell virtual items. According to observations on dark web forums, buyers tend to buy drugs in local countries and regions to reduce the risk of being discovered by customs and LEAs. Therefore, we have reason to believe that the location information provided by the vendor is mostly accurate, in order to attract target customers better.

Trust Level. This is usually the way to show the credibility of vendors on the English dark web markets. On different dark web markets, the trust level may be calculated by their own algorithms designed by the operators. Figure 13 shows the relationship between trust level and sales, feedback, number of disputes, joining date and vendor population distribution on *Cartel Marketplace*. The numbers

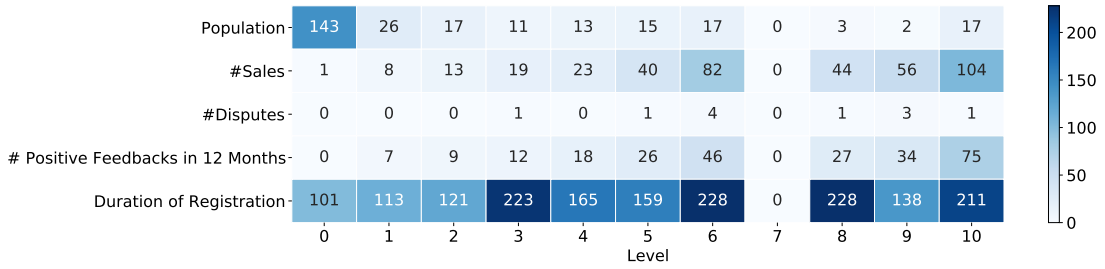


Figure 13: Trust level correlation matrix on *Cartel Marketplace*

displayed are the median values at each level, except for the population distribution. For example, in level 10, the median sales volume is 104. Level changes are affected by a combination of these aspects. High-level vendors usually have higher sales volumes and better reviews. The duration of registration has a limited impact on the level reached. For example, *Vendor_EN_1* is a level 10 vendor who joined the market in July 2021 with 185 sales, only one dispute, and 141 positive feedbacks. Please note that not all orders compulsorily require feedback. In comparison, *Vendor_EN_2*, trust level 5, joined the market in March 2021 with 526 sales but 15 disputes and only 187 positive feedbacks. As a result, in addition to sales, the number of complaints and positive feedback also influence the trust level.

Behaviour. We selected some vendors to describe and analyse their behaviours from *WHM*. *Vendor_EN_3* joined the market in July 2020, with currently 3,680 sales and 95% positive feedback. In the seven weeks included in our datasets, the number of orders was at least 830. The vendor had over 7,000 transactions in the previous *Empire Market*, with 99% positive feedback. This vendor mainly sells cocaine products in the United States. The largest displayed unit can reach 500 grams, and the price is close to $20,000. According to the smallest sales unit, we calculated that their total estimated profit is at least $2.7 million with around $207,500 made during the observation period. As a successful big vendor, they

can usually use their name as a symbol of their brand. *Vendor_EN_4* also confirms this point. The vendor has good reviews in its own region and sells drugs on multiple dark web markets simultaneously. Most successful vendors sell on different English dark web markets at the same time. They use the same format and language style in their profile. Since the feedback rating of English dark web markets can be imported into another, buyers can easily identify cross-market vendors. Some vendors also regularly update the product or their own situation in their profile. This can be used as a signal if they do not appear for a long time. They may be arrested or just quit, warning past and potential customers that they may be at risk.

Market Policy

This subsection describes what products are banned on the three English dark web markets, and the regulations regarding communication between users.

In *Dark0de Reborn*, the policy is called "selling policies and seller code of conduct". It is strictly forbidden to sell any images of sexual abuse of children, fentanyl-related products and any product or service related to terrorism. Fentanyl is an analgesic generally used in surgery; excessive use can quickly lead to addiction, hypotension and death due to respiratory depression. The website stipulates the use of on-site messenger as the means of communication between buyers and sellers and for customer service. Also, the exchange of large amounts of communications should be avoided unless paid for via *Dark0de*'s services. To protect the competitiveness and security of the market, external links and external dissemination of user information are not allowed.

In comparison to *Dark0de Reborn*, *WHM* is more comprehensive and detailed. The policy strictly forbids any child abuse, human or animal abuse, murder for hire, weapons, Fentanyl and terrorism-related products. During the Covid-19 pandemic, products that purportedly can cure the virus are banned, but discounts

on related tests and promotional codes are available. In terms of communication, they also ban external links and any external contact information.

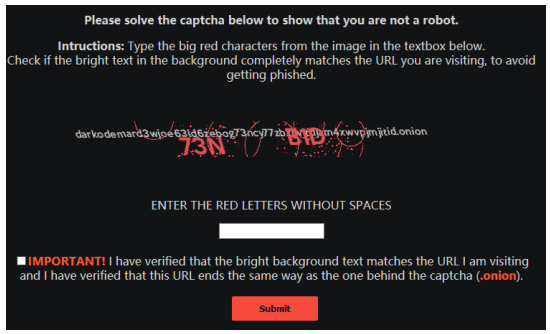
Cartel Marketplace has tough market rules. It forbids child abuse, biological, radiological, or chemical weapons, murder for hire, scamming, and deceptive tutorials. The market also bans searching for and publishing private or identifying information. In terms of communication, it bans direct deals but does not limit external links or contact.

Payment Method

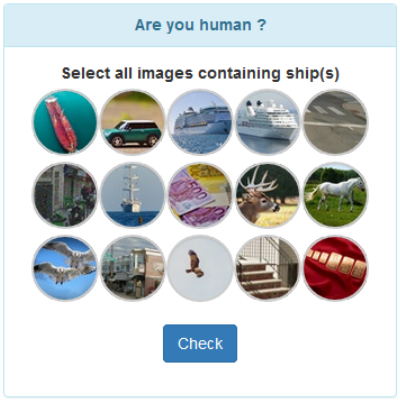
All three markets use the on-site wallet mode. Users are given a deposit address, and they then add funds by using Bitcoin or Monero. *Dark0de Reborn* and *Cartel Marketplace* allow both Bitcoin and Monero. On *WHM*, since Bitcoin is much easier to track, only Monero is allowed, for security reasons. On *Dark0de Reborn*, users can also use the website balance to purchase gift cards and send them to other accounts. That means *Dark0de Reborn* supports the transfer of funds within the market. On the website, the default currency is USD. However, users can choose between CAD, EUR, GBP, RUB, AUD, etc. The website also displays real-time market exchange rates.

Security Mechanisms

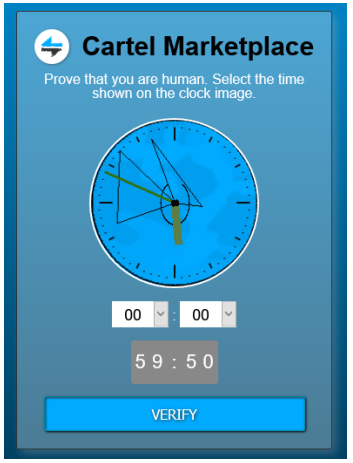
All three markets studied have strict crawling restrictions, such as the use of CAPTCHA. Figure 14 shows the CAPTCHAs on the three English dark web markets. *Dark0de Reborn* only implements a simple letter and number combination verification code. On *WHM*, the bot-check consists of choosing the images that match a specific description out of 15 pictures. Those are randomly rotated, for added security. On *Cartel Marketplace*, the user needs to indicate the time shown by the analogue clock. Then, there is another simple verification code to be entered when logging in. *WHM* and *Cartel Marketplace* allow users to choose



(a) *Dark0de Reborn* CAPTCHA



(b) *White House Market* CAPTCHA



(c) *Cartel Marketplace* CAPTCHA

Figure 14: CAPTCHA samples from English dark web markets

how long the session will be kept alive – available options range from ten minutes to 48 hours. After those verification processes are completed, a user can access the market's home page.

The number of pages that can be accessed per minute or per hour is strictly limited as well. *WHM* only allows a limited number of requests in ten minutes. Even if a (human) user tries to open multiple tabs at once, it is easy to trigger the detection system which will force logout and require re-login. Moreover, the trigger conditions may vary, but approximately 40 requests are allowed within 30 minutes. We applied a dynamic delay of 16 seconds to 90 seconds with six accounts. The dynamic delay is based on the corresponding time of the server to ensure that the page is returned. Sixteen seconds is the minimum interval between each request in an ideal network situation. One of the six accounts was randomly used for each request. The probability of one account being used continuously is relatively low. This method simulates the real situation where humans browse the web, and at the same time, collect data most efficiently without being detected. On *Cartel Marketplace*, the limit of requests is 300 pages per session within about 40-50 minutes. Once a user reaches such a threshold, the session expires automatically. We applied a dynamic delay of three seconds to ten seconds with three accounts. An account was used to traverse the listing page of the market. The remaining two were randomly used to request vendor pages. For *Dark0de Reborn*, the threshold is unclear. We noticed a large number of requests that cannot be parallelised in the crawler, and we applied a dynamic delay between 5 to 60 seconds. We were able to use one account to collect the required data. In terms of the structure of cookies, they seem to be static on all three markets, which means the cookies of a session do not change.

In terms of the user account, all three markets use PGP public key to ensure account security. Once a user sets up the PGP public key in their profile, they are able to use Two-Factor Authentication (2FA). PGP has also been used widely

in on-site communication. At the registration stage, the user needs to set up a username and password, and then the markets will send a set of English words to create a wallet.

Key Takeaway

The English dark web markets have a complete ecosystem, which means the features are fully supported and implemented. Vendors on the English dark web markets are active, which also includes cross-market vendors. The English dark web markets tend to have more restrictive policies in terms of what products can be sold. Drugs are the most popular category. Bitcoin and Monero are commonly used as the main payment methods. The English dark web markets usually have strict crawling restrictions.

Figure 15: *Chinese Exchange Market* homepage

4.3.2 Dark Web Markets in Chinese

Market

Operation model and structures. On *Chinese Exchange Market*, the website structure is simple and similar to a community forum. Figure 15 shows the *Chinese Exchange Market* homepage. It displays the latest posts under each category on the homepage. On each category page, each post contains the title, price, post time and vendor. On *Tea Horse Road*, the interactive interface is more modern, and users can also browse by category. The Tea Horse Road also supports the request-to-buy mode, where users post the products they want to buy and describe their requisites, and then vendors can provide quotes. Even though the structure of the two websites is relatively simple, the user experience is good: there are no redundant functions, and the products are easy to browse based on their category.

Number of listings and vendors. Figures 16 and 17 show the number of listings and vendors over the observation period. In terms of the number of listings, both markets maintain an overall upward trend. The number of listings only includes the normal sale mode on *Tea Horse Road*. For the request-to-buy mode, the number of listings remains stable at 2,400 during the observation period. Due to the operating mode of both Chinese dark web markets, once the advertisement is posted on the website, the post will be kept for a long time even if the vendor is not active. On the contrary, if the vendors do not show up for a long time on the English dark web markets, their status could be changed to inactive. In this case, the listing page will not cover their posts. Moreover, according to the data, the number of vendors remains stable. Even if the number of listings keeps growing, the number of vendors remains at a certain level. A new vendor may bring more than one listing.

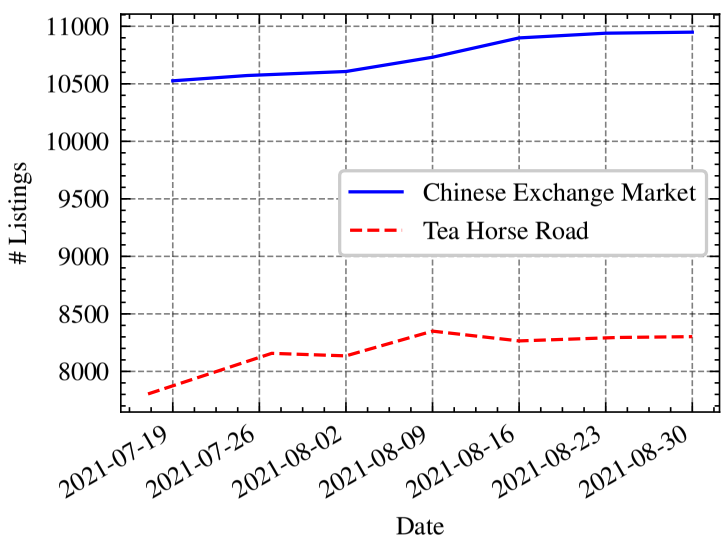


Figure 16: Number of listings on observed Chinese dark web markets

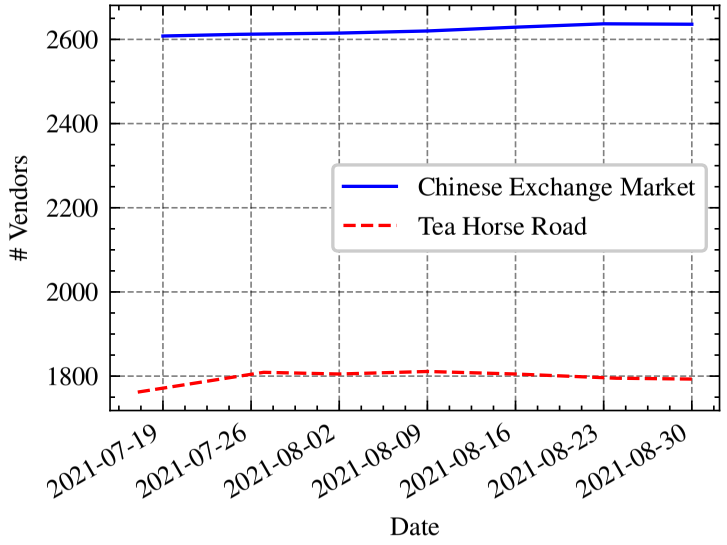


Figure 17: Number of vendors on observed Chinese dark web markets

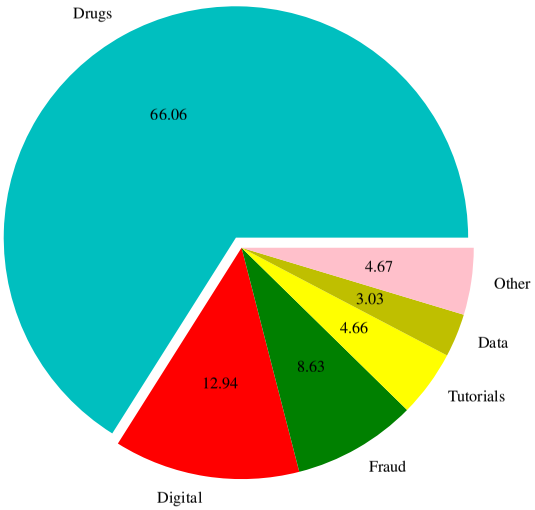


Figure 18: Item categories breakdown on English dark web markets

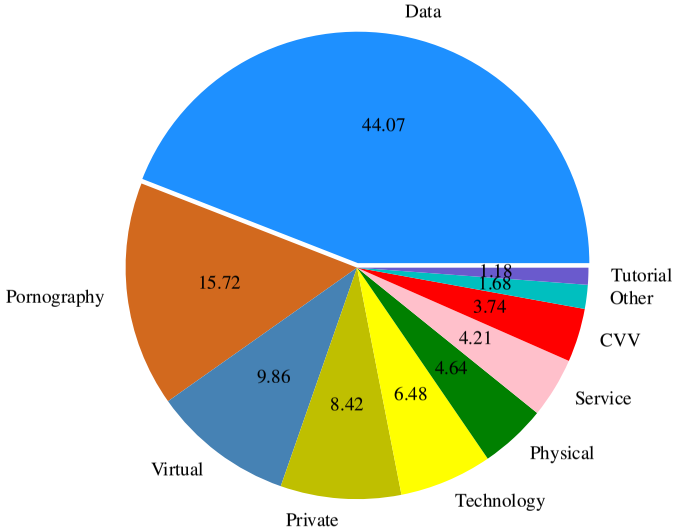


Figure 19: Item categories breakdown on *Chinese Exchange Market*

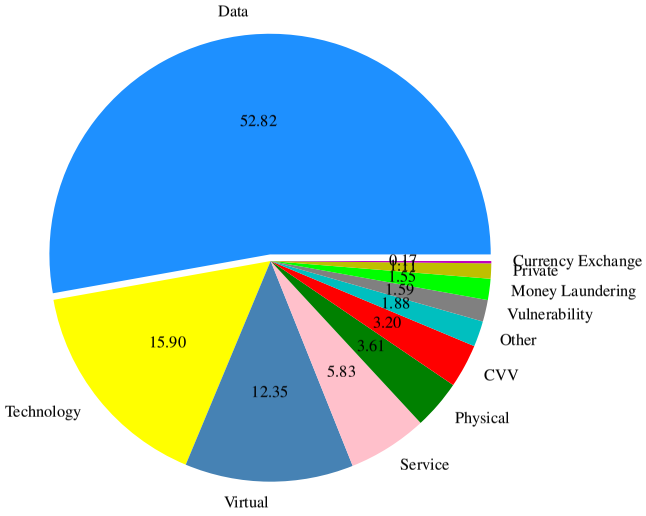


Figure 20: Item categories breakdown on *Tea Horse Road* with sale mode

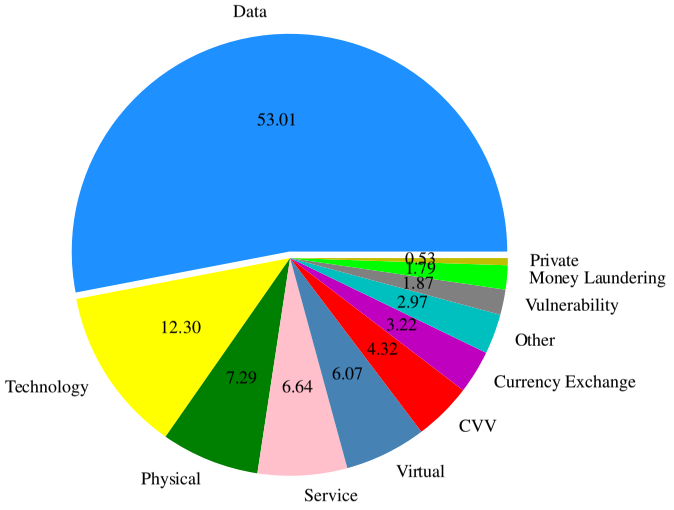


Figure 21: Item categories breakdown on *Tea Horse Road* with request-to-buy mode

Product category. We described the breakdown of categories separately for those two Chinese dark web markets as they used different criteria to classify categories. Figure 19 shows the percentage of items that belong to each category on the *Chinese Exchange Market*. Leaked data and personal information accounted for the largest proportion, which is 44.07%. Pornography is the second most sold category with 15.72%, which also potentially includes CSAM. Private is a category that allows buyers and vendors to use the market as a secure way to carry out their transactions. They usually have previously communicated and agreed to a deal.

Two figures show the percentage of each category in all items on the *Tea Horse Road*. Figure 20 shows the normal selling mode. Figure 21 shows the request-to-buy mode. Leaked data and personal information accounted for the largest proportion in both sections, approximately 53%. Pornography is classified as virtual items, and drugs are classified as physical items. Buyers request more physical items for sale. On the Chinese dark web markets, the most profitable category is related to leaked personal data. We used the data from the first week and the seventh week to estimate the revenue. If a product does not appear in the first week, we mark the initial sales as zero. Otherwise, We calculate the difference in seven weeks. After obtaining the profit of a single product, it was taken into account according to the category. On the *Chinese Exchange Market*, all vendors are estimated to have made profits of at least $86,984.98 during the observation period. The profit related to personal data is $35,749.21, or 41.1% of all profits. The second is service, which usually are personal information query services or hires to harm others. Following is pornographic except private section, because the private section is usually priced as $1 or in the smallest unit (e.g., per gram, record, picture, video), which causes the sales number to be inaccurate. The pornographic category generated 1489 sales with $10,211.75 profits during the observation period. On *Tea Horse Road*, the sales volume is displayed only

within a certain period instead of cumulative (that is, the numbers counted to this period will be deleted after a certain period of time), so the estimated profit is not as accurate.

Vendor

Inactive days Figure 22 shows the vendors' joining date and inactive days from the first post to the last seen on the *Chinese Exchange Market* with 2,537 vendors. At the end of 2019, there is a clear dividing line. We speculate that the market was maybe temporarily closed, banned many vendors, scam exited or updated for a long period of time. It seems like a large number of vendors exited the market at this point. Only a small part of the vendors before the end of 2019 have survived to the present times, and some vendors that joined after 2020 are still active nowadays. After 2020, some people remained active until now.

Behaviour On both Chinese dark web markets, we selected some typical vendors to describe their behaviours. *Vendor_CN_1* has 113 sold items over a month with $6,773 estimated profits. The vendor first posted in June 2019, and generated a total of $29,021 in estimated profits. The vendor lists 88 posts about personal data, including pornography, virtual items and CVV-related items. This personal data contains express delivery information in China, Japanese phone numbers, a set of Chinese ID card numbers with names, phone numbers and emails, email accounts with plaintext passwords, personal information in loan

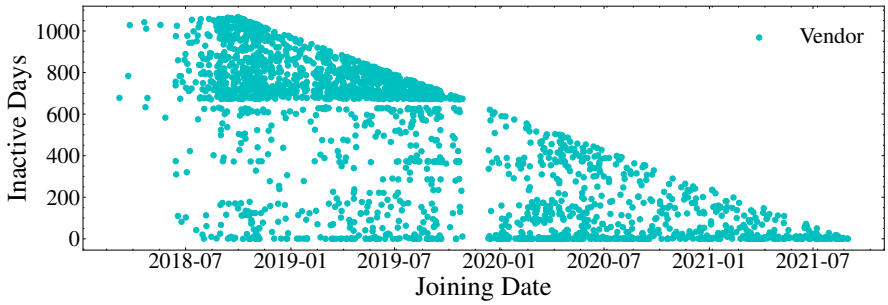


Figure 22: Vendors' inactive days and joining date on *Chinese Exchange Market*

databases, motor vehicle registration information and much more, covering data all over the world. Pornography includes child abuse material and hidden camera shots. Virtual items include (stolen) VPN accounts and software. The CVV-related items include card information and carding methods. In almost all posts, this vendor has screenshots to showcase items.

Vendor_CN_2 also sells personal data. The vendor sold 59 items with $2,124 estimated profits over the observation period. The most valuable product is the records of student data who attend university. The personal data contains name, age, phone number, e-mail address, school name and subject. *Vendor_CN_3* and *Vendor_CN_4* also provide services for e-whoring and hiring to hurt, respectively. However, there is not too much information in their posts, but they recommended using on-site communication for details. Since they do not have a clear price, the profit is difficult to estimate.

Market Policy

On the *Chinese Exchange Market*, there are no official policy pages, just user instructions for new users and vendors. The instructions mainly ask the vendor to describe the product as accurately as possible when posting and give some tips for buyers to prevent fraud. As a guarantee intermediary, the market operation does not prohibit any products or services. Posting tutorial products needs to explain that why vendors cannot do this by themselves and describe the potential risks. Otherwise, they will not be allowed to advertise. In a prominent position on the website, it is reminded that external connections and external communications are not allowed. Tea Horse Road forbids child abuse, unethical resources, materials to subvert state power, and political and political leaders related resources. In terms of communication, the website allows users to use off-site communication but avoid off-site transactions. Also, the website does not explain whether it is possible to use external websites, but we did find clear web links and other external

Figure 23: CAPTCHA samples from Chinese dark web markets

websites in some product descriptions.

Payment Method

On the *Chinese Exchange Market*, the on-site wallet mode is used. The user needs to transfer Bitcoin to the Bitcoin wallet of the market administrator. The corresponding amount will be shown in their account on the market. On the advertisement post, the price is displayed in USD and Bitcoin. *Tea Horse Road* also uses the on-site wallet mode. The market allowed adding credit to their wallet by Bitcoin and Tether. Moreover, a difference with the *Chinese Exchange Market* is that users are allowed to use common fiat currency with daily payment methods, including Alipay, WeChat, and even debit cards, by contacting market-authorised exchanges via Telegram. On the advertisement post, the price is displayed in USD and Bitcoin.

Security Mechanisms

On both Chinese language markets, the crawling restrictions are easy to bypass. Both use simple four character letters and numbers combination CAPTCHAs when login-in. Figure 23 shows the CAPTCHAs on both Chinese dark web markets. Both markets lack request limitations. In order to avoid any bandwidth stress, we also applied a dynamic delay of 0.5 seconds to 2 seconds with a low number of concurrent requests. Hence, this causes the total crawling time to vary from four hours up to 12 hours. In terms of the structure of the cookies, *Tea Horse Road* uses a dynamic approach in which it will change every time a request is sent. It does not seem to affect the crawling process, but it was noticed when setting up the initial cookie, and we let the program follow up. We also noticed that the post page URLs are consecutive. This may lower the website's security level because of the potential for exhaustive cracking and crawling. Overall, both markets do not have strict crawling restrictions.

In terms of the user account, the two Chinese dark web markets have different mechanisms. On the *Chinese Exchange Market*, the user only needs to set a password. Then the website will assign an ID to the user. Interestingly, this user ID is a counter, which means that every time a new account is registered, the ID is increased by one. We tracked back our dataset and found that the ID is very likely to increment by one. Our latest data shows that this ID is currently 677653. Although this is certainly not the number of active users on the market, it is the cumulative number of registered users on the market since its inception. The user needs to subscribe to the "post" and "reply" types of activity, respectively, to post and buy. The "post" cost 0.00024 (∼$20) and 0.0006 Bitcoin (∼$60) for three months and one year, respectively. The "reply" cost 0.00012 (∼$10) and 0.0003 (∼$30) Bitcoin for three months and one year, respectively. On the *Tea Horse Road*, the user needs to set a username, password and payment password. The payment password is used when a user purchases a product. Accounts need to be

activated before publishing or purchasing products. The user needs to pay $10 in equivalent Bitcoin for the activation. There are no account password recovery options on either market.

Key Takeaway

The Chinese dark web markets remain incomplete and less developed, but at the same time, they try to be innovative, for example by adding the request-to-buy mode. Vendors are less active than the English dark web markets. The cross-market vendors are hard to track as the vendor ID is made of immutable numbers (i.e. they cannot be personalised, which would allow more obvious cross-market tracking). Leaked/personal data is the most popular category. There is no official policy to prohibit any products or services on the Chinese dark web markets we studied. Bitcoin is the main payment method, however, fiat currency has also been accepted on *Tea Horse Market*. The Chinese dark web markets usually have less strict crawling restrictions.

4.3.3 Comparison of Dark Web Markets in English and Chinese

The differences between the Chinese and English dark web markets are reflected in many aspects. This section focuses on the market and vendors' characteristics.

Market

Operation model and structures. English dark web markets tend to be real online markets, while Chinese dark web markets tend to be more similar to forums. Some built-in functions on the English dark web markets, such as credit systems and feedback systems, build a mature ecosystem. The Chinese dark web markets only serve as an intermediate platform for posting. It is, therefore, quite difficult

for users to judge the credibility of the vendor. In terms of novelty, one of the Chinese dark web markets provides a request-to-buy mode. Users are no longer limited to browsing the displayed products but can make a request, which can be customised. The website structure of the Chinese dark web markets is simple but functional and helps locate products faster. Nevertheless, considering the profitability of the market itself, the English dark web markets are usually mixed with promotional products in all lists, and the promotional items are difficult to distinguish. On the other hand, on homepages, the English dark web markets use both pictures and text, while the Chinese dark web markets only have text. Physical items often need more pictures to display, while virtual items could use descriptive text only. This phenomenon reflects that different strategies apply to different types of popular products in different language dark web markets.

Product category. Drugs dominate the English dark web markets, while on the Chinese dark web markets, the most popular category is personal data. Due to the convenience of express delivery in North America and Europe, and the different laws and regulations of each state or country, drug shipments are difficult to spot and stop. However, Chinese law enforcement is characterised by a heavy crackdown, deep inspection, severe sentencing, and heavy propaganda against drug abuse. As of 2020, the number of drug abuse users in China has kept falling for three years in a row [25]. Regarding personal data, Chinese LEAs have been cracking down on telecom fraud in recent years. Most of them are related to the leakage of personal information. Due to the convenience of disseminating virtual items such as personal information, it is difficult to stop it. Pornography is the second-largest category on the Chinese dark web markets, which also contains child abuse material. On the English dark web markets, they are clearly stated that child abuse material is not allowed.

Market policy. English dark web markets restrict most high-risk products, such as arms, chemical weapons, child abuse, animal abuse, etc. There are no

special restrictions on items on the Chinese dark web markets, but fraudulent behaviours with fake products on the market will be banned.

Payment method. Bitcoin is still the main currency on both English and Chinese dark web markets. Most English dark web markets also accept Monero for better privacy. Users can use fiat currency for small deposits on Chinese dark web markets.

Security mechanisms. The English dark web markets usually have stricter security measures than Chinese ones. More complex CAPTCHAs are used on the English dark web markets. The session time on the English dark web market usually has different time-window options, but it must be within a certain number of requests. Otherwise, users will be kicked out. However, since there is usually no limit on the number of requests on Chinese dark web markets, as long as the session is active, it will not be automatically logged out. In terms of account security, the Chinese dark web markets use the pay-to-activation method to control malicious registration, while the English dark web markets use the PGP public key.

Vendors

Inactive days. Vendors on the English dark web markets are more active than the Chinese dark web markets. The English dark web markets usually have a shorter life cycle. The three English dark web markets were established later than the *Chinese Exchange Market*. In the past, most English dark web markets were being shut down or exited scams at the end, and then those market operators would usually change their identities and operate new markets. The new vendors on English dark web markets will remain active and establish their brand in dark web forums. Even if the market is closed, they can quickly sell on the new market because buyers usually follow good vendors. However, on Chinese dark web markets, even if the operator's scam, they would not exit the market. They keep the operation as new vendors will not know because of the lack of

communication. Vendors on the Chinese dark web markets do not have to consider such feedback, so they also spend less time on the market for such customer service. By comparing Figures 12 and 22, we can clearly find that some vendors are not active within a few months after posting their posts on the Chinese dark web markets. On the contrary, most vendors were still active on the English dark web market, even over a year after they first registered.

Location. On the English dark web markets, the proportion of international vendors is greater because of the widespread use of English. For instance, some non-English speaking countries in Europe have more lax drug regulations, contributing to some vendors from such countries. In comparison, we found that most of the vendors on the Chinese dark web markets are native speakers, as indicated by the jargon being used. However, since most virtual products and pornography are sold, the vendor's real geographic location is difficult to measure.

Behaviour. Vendors on the English dark web markets pay more attention to building their own brands. They usually use their vendor profile or description section to advertise themselves. Successful vendors claim and show their sales numbers and ratings on other well-known English dark web markets. They also explain the return policy, e.g. what will happen if the items are lost in transit. Vendors sometimes update new products or their personal status. On the Chinese dark web markets, it is difficult for the vendor to do the same on the English dark web market because of the lack of functionality. However, we noticed that vendors on the Chinese dark web markets sometimes have their own language style, but it is still difficult to define the cross-market actors.

4.4 Discussion

4.4.1 Insights

In comparison to their English counterpart, even though the ecosystem on the Chinese dark web markets remains incomplete and less developed, it tries to be innovative, for example by adding new selling modes. The lack of any feedback and rating system likely contributes to making the vendors slightly less active. Chinese markets seem to serve mainly as a safer first point of contact rather than a full trading platform. Goods exchanges and price negotiations are likely to be carried out of the market, using other means of communication. We observed that both of the analysed Chinese dark web markets have a "private deal" section, allowing vendors to trade with specific target buyers. With the request-to-buy mode, buyers have more options than on English dark web markets. Chinese dark web markets may be improving and upgrading the functionality and security of their services by learning from the practices of English dark web markets. Moreover, as automated language translation becomes more and more accurate, markets in different languages can be accessed without the need for complex applications. That could become a new challenge for us. On the other hand, we can delve into how to use those technologies against crimes, for example, by creating a system that can detect illegal cross-market activities.

Chinese dark web markets have less stringent policies than English dark web markets. On the Chinese dark web markets, there are resources for child abuse, weapons, and hire-to-harm. The administrators usually do not care about the products sold, which is not the case on the English dark web markets. English dark web market administrators have also begun to focus on the quality of their products, for example, by implementing the "Product Quality and Harm Reduction" programmes.

All markets suffer from reputation issues. On English dark web forums, we can

see discussions or comments for each English dark web market. New markets will always appear, and most of the old markets will always gradually lose vendors and buyers for some reason. Some closed down, either being seized or the operators performing exit scams. The Chinese dark web navigation website also displays comments from anonymous users on the Chinese dark web market. They usually complain about customer service and potential scamming activities. We may be able to explore such methods using specific indicators to predict scams before they widely occur.

Cross-market actors are active. We have seen the same vendors on all major English dark web markets. We speculate that international vendors are likely, especially for personal data, on both English and Chinese dark web markets. We found that the personal information data sold on markets contains leaked data from all over the world. Cross-market behaviour exists, and even on dark web forums, they use the vendor's identity to participate in discussions. It should be noted that this may also be one of the means to promote their own items. We also observed that they use other accounts to assume other identities. If we can track and link these accounts, we will be able to understand e-crime operations better.

The main reason items advertised in Chinese and English markets differ might be due to legislation and law enforcement. China has very strict drug control legislation, which has a certain deterrent effect on potential criminals [72]. Vendors may not want to take the greater risk (e.g., sending packages requires using the real-name system in China). From our observations on the English markets, we found that buyers also prefer local vendors because there is less risk for both parties, and the packages do not need to go through customs. Compared to drug trafficking, trading personal data carries less risk, and the potential profits can be realised more quickly. Furthermore, China's larger population may provide cybercriminals with easier access to large amounts of personal data, making the

exploitation of this information potentially more profitable. Since China does not have a content rating system similar to that of the West, adult content is also a big part of the advertising on the Chinese dark web market.

The results indicate that cyber-enabled crimes are the main type of cybercrime found on the dark web market in both languages. This is largely due to the concealed nature of the dark web, which facilitates the expansion of sales channels and enhances efficiencies for activities such as drug dealing and personal information trading. While there are tutorials and resources available concerning cyber-dependent crimes (e.g., malware, hacking tools), an estimate suggests that these constitute less than 20% of all items (as per the possible categories). Therefore, dark web markets in different languages do not seem to differ much in terms of cyber-enabled and cyber-dependent crimes. Future studies could benefit from a more detailed classification of listing items in different language markets from a criminological perspective.

4.4.2 Challenges

The main technical challenge we faced was the many restrictions these markets implemented to stop or at least slow down the automatic crawling and scraping of their websites. This is, obviously, a serious challenge for data collection. This is compounded by the instability of the Onion service, resulting in frequent but irregular interruptions to the data collection process. In addition, we must manually log in to each of our accounts before starting the crawler in order to get the cookies needed for the session. English markets are severely more restrictive in their anti-crawling measures, so we needed to maintain and operate multiple accounts simultaneously. In addition, market operators frequently change the structure and the design of their websites, sometimes causing crawlers to fail. At times, market operators update their security mechanisms without any notice,

typically in ways that are damaging for bots but transparent for humans. For instance, *Cartel Marketplace* updated their CAPTCHAs, and the *Tea Horse Road* also redacted the number of requests per session. In each of those cases, we must reconfigure our crawler. So data gathering becomes a continuous cat-and-mouse game, costly to maintain for large periods of time.

4.4.3 Limitations

Due to technical and time issues, there is an imbalance in the number of vendors between the English and the Chinese dark web markets. The more stringent crawling restrictions on the English dark web markets caused our crawler not to be able to scrape all of the market content, resulting in a smaller amount of data being collected. In comparison, the Chinese dark web markets have less stringent restrictions, leading to more data points. Our dataset also contains a relatively short period of time on both English and Chinese dark web markets, so further work to expand our dataset would be worthwhile.

We have considered and dealt with bad data in our analysis. However, despite our best efforts, we still cannot guarantee the integrity of all the figures, such as prices and sales, on any dark web market. As such, the figures presented in this work are based on our best estimation, which can and shall be improved in follow-up research.

4.5 Chapter Conclusion

In conclusion, this work has investigated and analysed the differences between dark web markets using English and Chinese as their main languages. The differences found are, at times, quite interesting and have a basis that is not only linguistic but also cultural. For this research, we collected data from five trending dark web markets, comprising three English and two Chinese dark web markets. Data

collection was carried out for seven consecutive weeks.

English dark web markets generally seem to offer a more mature and complete ecosystem, with more active vendors than their Chinese counterparts. We found that the multiple differences between English and Chinese dark web markets are reflected across many aspects, including selling modes, product categories, market policies, payment methods, security mechanisms and vendors. On Chinese dark web markets, vendors are on average less active than on English dark web markets, but the demand and number of sales of personal data and pornography are relatively large. On English dark web markets, the main products sold are drugs. Moreover, the policies of Chinese dark web markets show that there are very few products banned, and the problem of child abuse material is extremely serious. On one of the Chinese dark web markets, fiat currency can be used, and the anti-crawling restrictions on both of the observed Chinese markets are easy to bypass.

Some interesting insights from our research are the existence of request-to-buy modes and some uncommon policy issues. We believe that these provide a way to gain comparative insights into dark web markets, and attract the attention of law enforcement agencies. In particular, this request-to-buy mode could be a good way to launch sting operations – where allowed by law. We hope that our study will provide a better understanding of dark web markets, particularly Chinese dark web markets. We also discussed that the main reason for the difference in items advertised in Chinese and English markets could be legislation and enforcement. Future work can focus on vendor behaviour and cross-market operations on different language dark web markets. For instance, by using natural language processing techniques, we may be able to discover connections between vendors and identify potential cross-market international large-scale operators. Finally, the tracking of payment methods and profits is an additional interesting research path for the future.

Chapter 5

An Analysis of Dark Web Markets Security

This chapter is based on the content of previous publications:
"Analysis of Security Mechanisms of Dark Web Markets" [144], and an extended version, "Secure in the Dark? An In-Depth Analysis of Dark Web Markets Security" [145]

5.1 Introduction

As the name implies, dark web markets – also commonly known as anonymous markets – have put in place measures for protecting the privacy of its users, both sellers and buyers, as this is a key priority that can attract users worldwide. With the rapid growth of dark web markets, competition among them has become more intense. In this environment, malicious attacks targeting competitors – for instance, aimed at reducing the availability of rivals' services – have also become more common. These attacks not only affect other services' availability and accessibility, but they may also lead to personal and private information being

leaked. As such, it is understandable that dark web markets may want to implement strong security mechanisms to protect themselves and their users from both law enforcement and other operators. This is particularly true as good security can be a matter of survival but also help to gain a competitive edge over rivals. We also investigate and present what typically would happen before and after the closing down of dark web markets in Chapter 6.

Previous studies have mainly focused on social aspects, including analysis of products sold, emerging criminal patterns, criminal ecosystems, and key actors [48, 68, 70, 86, 92, 142]. However, the security mechanisms used by the dark web markets have not been addressed in enough depth.

In this chapter, we investigate the security elements of different dark web markets. Furthermore, we understand the challenges of data collection on the dark web, and therefore, we expect to gain some valuable insights from the data collection process in this work in order to help improve the efficacy of current crawlers.

5.2 Methodology

We selected twelve existing mainstream dark web markets when we started our research in May 2023. The market selection was based on observations of the *Dread* forum (a dark web forum), searches on the clear web using keywords such as "dark web market", and manual verification of markets that were active at the time. Over a span of four months (until August 2023), we collected data pertinent to these markets, paying close attention to any security mechanisms they have in place. Please note that there were some variations to the timing and quantity of the data collected from each market, due to some factors beyond our control (such as markets being down for short periods of time, etc.).

Initially, we tried to use an existing crawler (mentioned in Chapter 3) to obtain

Table 5: A summary of the selected dark web markets (for the analysis of market security, as of 31 August 2024)

Market Names	Type	First Seen	Last Seen	Status
Abacus Market	Comprehensive	2021-09	(still active)	Live
Archetyp	Drugs-only	2020-05	(still active)	Live
ASAP Market	Comprehensive	2020-03	2023-07	Retired
Bohemia	Comprehensive	2021-05	2024-01	Closed/Seised
Incognito	Drugs-only	2020-10	2024-03	Closed/Seised
Kingdom Market	Comprehensive	2021-05	2023-12	Seised
Nemesis Market	Comprehensive	2021-05	2024-03	Seised
Royal Market	Comprehensive	2021-03	2023-08	Closed
Tor2door Market	Comprehensive	2020-07	2023-09	Closed
Vice City Market	Comprehensive	2020-08	2023-07	Closed
Chinese Exchange Market	Comprehensive	2018-03	(still active)	Live
cabyc*	Comprehensive	2022-02	(still active)	Live

**cabyc* is the initials of Chang'an Nocturnal City in Chinese

(sales-related) data for all markets. However, we encountered some difficulties, which also shows the dynamic nature of most market security mechanisms. While dealing with these issues, we were able to jot down the security mechanisms we encountered and run some small experiments against them using the crawler.

Table 5 provides a key summary of these twelve markets, which either sell drugs-only items or sell many different categories of items (labelled as "Comprehensive"). In this table, we also state the date they were first and last seen online, and the market status at the time of writing this chapter (August 2024). There are some markets marked as closed, with the reason for closure – which could be "retired" (where the market operators voluntarily closed down their market), or "seised" (where it is understood that some law enforcement agencies took down the market). It should be noted that there are some uncertainties associated with this area of research. Also, we aim to reflect market conditions and characteristics at the time the research was conducted, which means the closure should not affect our results. Interestingly, we will discuss in the following sections whether security mechanisms have likely affected the operational longevity of the market. Those markets have been selected based on good representation and reputation

in the dark web community (i.e. included and recommended by dark web forums and information websites).

The various security mechanisms are identified based primarily on previous work and our previous experience. Turk, Pastrana and Collier [131] and Georgoulias et al. [69] mentioned the anti-crawler technology in underground forums and the characteristics of dark web markets respectively. According to their different functionalities, we define and group the security mechanisms in those different markets into three main aspects: *web security*, *account security* and *financial security*. Following, thematic analysis was used to code for different security mechanisms in different markets. In *web security*, we focus on the technical implementations and strategies that the market applies to their websites to protect users and themselves, such as CAPTCHAs, secret phrases, rate limiting, etc. In *account security*, we focus on the security mechanisms and policies that keep the account secure, such as username and password requirements, etc. In *financial security*, we describe the mechanisms associated with transactions, such as the accepted currencies, the type of transaction, etc.

Information is gathered either while running a customised crawler or manually accessing the markets. The next subsection provides an overview of the customised crawler we implemented for conducting this research. While the crawler is helpful in certain scenarios, such as aiding in the understanding of rate limiting, sometimes its utility can be limited during investigations, such as when testing CAPTCHAs and account security. In these cases, manual interaction is generally more effective and precise, allowing for a deeper understanding of the implementation quirks of security mechanisms.

5.3 Results

In this section, we describe and present our results. First, we describe the security mechanisms used by dark web markets in web security. We walk through the process of accessing a market to explore security mechanisms, and cover an open source software commonly used by the dark web markets. Following this, we describe account security, which covers the username, password & PIN requirements, mnemonics, MFA and account kill-switch. Finally, we describe the details of their implementation of financial security, including the supported currencies, transaction types, as well as the handling of complaints and general user support.

5.3.1 Web Security

Table 6 presents an overview of whether selected dark web markets implement specific web security mechanisms. We describe each mechanism in the following sections.

Accessibility

Accessing the dark web requires the users to know the address of the server (or its mirror). To help promote their markets, market operators usually advertise the address of their servers on the "website directories" pages. Users also typically share them on general forums. In the case of a registered or reputable user, you may also receive a private address, which is used to increase the availability of the market in the event that the main address suffers a successful DDoS attack. In general, all markets support the Tor network, but we also found that some additionally support the I2P network. One of the main reasons for this is that from late 2022 to early 2023, the Tor network suffered from many performance issues [66, 128], making it significantly harder for users to access these markets. Therefore, some decided to operate on both networks for redundancy. In addition,

Table 6: An overview of the selected dark web markets' web security mechanisms (●= yes, ○= no, ◑= partial)

Markets	Accessibility TOR	Accessibility I2P	Waiting Queue	Anti-Phishing	CAPTCHAs	Secret Phrase	Canary	Bug Bounty
Abacus Market	●	○	●	●	text and image	◑	○	●
Archetyp	●	○	○	●	image	◑	○	●
ASAP Market	●	○	●	◑	interactive text	○	○	○
Bohemia	●	●	●	●	interactive text	○	●	●
Incognito	●	●	●	●	image	●	●	●
Kingdom Market	●	●	○	●	interactive text	●	○	◑
Nemesis Market	●	○	○	◑	image	○	●	○
Royal Market	●	○	○	◑	image	○	●	●
Tor2door Market	●	●	○	◑	text	○	●	◑
Vice City Market	●	○	●	○	color	○	○	●
Chinese Exchange Market	●	○	●	○	text	○	○	○
cabyc	●	○	○	○	Chinese, math, text	○	○	○

we noticed that most of the time, I2P seems to be faster than Tor when it comes to response times. This observation is in agreement with what Georgoulias, Yaben and Vasilomanolakis [68] reported in their paper. They measured the response times in *AlphaBay*, and highlighted that the average response time for I2P was 5.6 seconds, which is faster than the 9.1 seconds for Tor. In [47], other researchers have shown that although I2P generally features better latency results than Tor, the latter seems to offer, in general, better throughput and download times than I2P. For example, Tor got to an average speed of 51 kB/s while I2P only got to approximately 13 kB/s. This research was published more than a decade ago (2011) and can easily no longer hold true, which would be in agreement with what we observed during our work. In addition, and regarding congestion, Tor uses circuit switching, whereas I2P uses packet switching. Hence, Tor often has to cope with high congestion rates [31] leading to high latency. Whereas in I2P, the packet switching leads to some implicit load balancing and helps to avoid congestion and service interruptions. This is specifically important for large file transfers, and therefore, I2P is more suitable for such purposes. We also noticed that three (*ASAP Market*, *Nemesis Market*, *Chinese Exchange Market*) markets allow access to product pages even if the user is not logged in or registered.

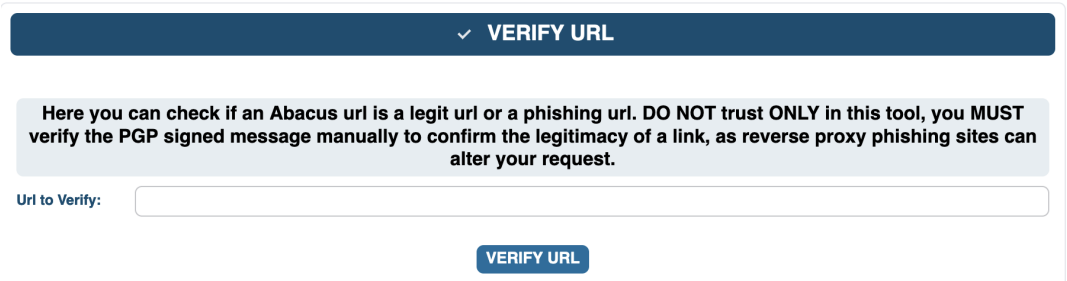
Waiting Queue

The first screen users usually see after entering a market will be queuing, which is mainly used to protect the website from DDoS attacks. The market first puts the user into a queue and then automatically redirects to the next screen after waiting for a period of time. We also found that this mechanism should also include some sort of load balancing feature on the server side. We expect that most markets would implement this mechanism at the first point of entry to the site, but this is surprisingly not the case. Only half of the markets we selected apply this security mechanism. Actually, CAPTCHAs could also have a somewhat protective effect

on DDoS (which we describe in Section 5.3.1).

(a) Anti-phishing page on *Bohemia*

(b) Anti-phishing page on *Archetyp*



(c) Anti-phishing page on *Abacus Market*

Figure 24: Three anti-phishing pages on different markets

Anti-phishing

Depending on the market, users may see this screen before or after logging in. This is mainly due to the fact that the Tor network is flooded with fake mirror links used for phishing.

Figure 24a shows an example of an anti-phishing page. Users need to compare the URL in the browser's address bar and fill in the missing letters or numbers in the spaces.

Other markets have similar strategies. For example, Figure 24c shows that users can verify their address on the website by entering the complete URL address.

There are four markets marked as half-filled circles under the "anti-phishing" column of Table 6. This means that these four markets alert users to check and compare whether the URL being accessed is the same one showing on the page, but without any form of verification. For example, Figure 24b shows another way to remind the user in the background of the CAPTCHA to check that the starting and ending characters of the URL address should match. Admittedly, those measures do not completely prevent phishing from occurring. Once an attacker completely clones a website and replaces the engine behind this mechanism (i.e. the method of verification), completely unsuspecting users can still be easily deceived. This mechanism is more like a reminder to force users to check the URL. We also note that, interestingly, in certain markets, this security mechanism is missing if users access the market via the I2P network.

CAPTCHAs

As the most widely used security mechanism in this sector, CAPTCHAs are ubiquitous on the dark web. The lack of access (due to their illegal operations) to a somewhat standardised solution such as *reCAPTCHA* forces each dark web

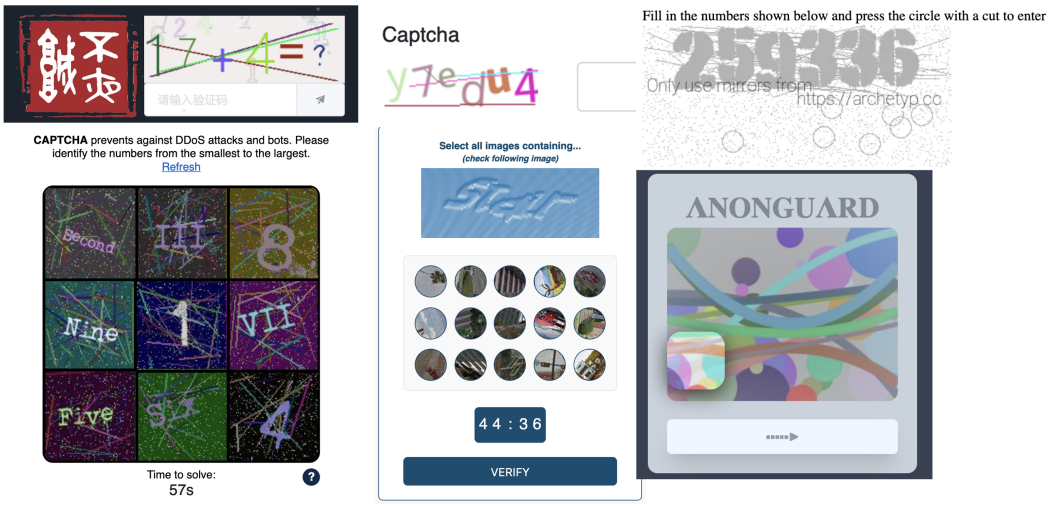


Figure 25: Examples of CAPTCHAs from six dark web markets

market to develop its own implementations. This approach is strongly considered dangerous in computer security circles. This naturally implies that all of the CAPTCHAs deployed in dark web markets are, without exceptions, flawed and trivial to break automatically, especially after the recent developments in AI capabilities in text and image recognition and classification.

Figure 25 shows examples of CAPTCHAs from six dark web markets. As shown in Table 6, in addition to the common static text input and image recognition, interactive text (i.e., the user needs to click/drag the correct answer instead of typing) and even colour and math-based questions appear on dark web markets as CAPTCHAs. These are all old and now trivial approaches that the CAPTCHA community has been abandoned years ago. The most popular modern approach is based on adversarial examples that fool AI systems. Users not only need to solve a CAPTCHA when entering the website, but also sometimes they need to solve another one when logging in.

Generally speaking, some CAPTCHAs seem to be difficult to solve, and some even have strict time limits (i.e. users need to solve them correctly within a limited time). However, if we bring in the latest knowledge on CAPTCHA generation and breaking from the security community [91, 155], these CAPTCHAs are all trivial to break. It seems that CAPTCHA designers for dark web markets, fortunately (and unsurprisingly), do not follow recent academic literature on the topic, and it absolutely shows in their designs.

Secret Phrase

This is actually another type of anti-phishing mechanism. Users can specify a secret phrase when registering. When the user logs in to the market, the secret will be displayed on the homepage. When the user realises that the secret phrase is displayed incorrectly (i.e. the user is on a fake website), the user has the opportunity to change the password and other information on the real market to

prevent further losses.

In Table 6, there are two markets marked with half-filled circles. Their implementation of this mechanism may be unintentional, but they have the same effect. Users are asked to fill in their nickname when registering. However, users only need their username and password when logging in. Therefore, users have the same opportunity to check whether their nickname is showing correctly after logging in.

Warrant Canary

This is a more traditional security mechanism that states that the market is still controlled by specific operators. This statement (canary) is usually displayed on a page on the market and is signed with the operator's PGP signature. The statement has the date of the next update and proof of the date the current statement was signed (e.g., this could be the latest Bitcoin block hash). Users will be aware that operators may lose control of the market if the canary is not updated by the mentioned date. We have noticed that some markets have canaries that are out of date, but operators usually update them after a few days.

Bug Bounty

While some may consider that bug bounty programs are not a core security mechanism, bug bounty programs do usually have an impact on market security, by engaging users in discovering and reporting potential vulnerabilities for improvement. In Table 6, markets marked with full-filled circles are those that have a proper bug bounty program and clearly mention that certain rewards will be obtained after discovering bugs. There are two markets marked with half-filled circles, meaning that the market mentions that a support ticket should be submitted to the market when a bug is discovered, but without further information on rewards. That is not the way to run a bug bounty program. For markets marked

with open circles, we believe users are still able to report directly to the market operator via in-site messages or support tickets. For example, one of the markets we observed has an optional subject in their ticket system – bug bounty, which further categorises the priorities into low, medium, and high. However, markets with open circles do not mention any guidance about what users should do if a bug is discovered.

In one of the markets, we observed a very detailed list of potential rewards available to the bug reporter. The market operators have graded the importance of different bugs, including info, low, medium, high, and critical. These tiered rewards range from $5 to more than $5,000. In terms of scope, this program covers various types of vulnerabilities, such as UI issues, server-side disclosure, sensitive data disclosure, authentication bypass, command execution, etc.

Rate Limiting

This limitation may be affected by many factors in real-world data collection (e.g., the internet service provider, physical network interface, etc.). In practice, it is difficult for users to tell whether they are suffering from rate limits by the market servers or if there are other bottlenecks in the user's network. Here, we only discuss the potential use of security mechanisms on the dark web market side. A typical approach is to preset a threshold on the server side. Once the frequency or number of requests reaches this threshold within a given period of time, the server will refuse to return the result page and perform additional security checks. Those additional security checks may include additional CAPTCHAs (i.e. making the session expire) and killing the current Tor circuit (i.e. changing the Tor identity needed). In practice, some markets have very restrictive thresholds, making it very challenging to obtain data for the entire market. For example, on *Bohemia*, our crawler operates at an average of ten requests per minute without the need for manual intervention for a long time. If we doubled up the speed (or even faster),

we might be able to collect data more quickly in bursts, but if we continued for several minutes, we would trigger the anti-crawling mechanisms and this would interrupt our data collection. As such, we have to make a compromise in this situation. The slower method would lead to a lengthy data collection time – e.g., taking more than four hours just to retrieve the listing pages data, without the item details yet – but it would allow our crawler to operate autonomously without requiring our intervention.

EndGame DDoS Filter Toolset

EndGame[3] is an open-source and widely used front-end system for DDoS protection on the dark web. This toolset is used to easily deploy some of the security mechanisms we mentioned above (e.g., two rate limiting methods based on Tor service circuit ID and cookies, customised randomly generated CAPTCHAs, time-based queue system, packet filtering, load-balancing etc.) Due to the nature of open source, market operators can easily customise functions without requiring an advanced technical background, and the setup process is highly scripted. Compared with DDoS protection services on the clear web, EndGame can be deployed locally rather than hosted on a third party (like Cloudflare), which fits well with the privacy requirements of market operators.

5.3.2 Account Security

In this subsection, we explore the security mechanisms applied to the account. As market operators, it is necessary to apply appropriate mechanisms to protect user accounts, which can help users avoid potential account loss, theft, scams, etc. Table 7 provides an overview of selected dark web markets' account security mechanisms implementation.

[3] `https://github.com/onionltd/EndGame`

Table 7: An overview of the selected dark web markets' account security mechanisms (●= yes, ○= no)

Markets	Username	Password	PIN	MFA	Mnemonic	Kill-switch
Abacus Market	alphanumeric*	min. 1 chars	6 digits	●	●	○
Archetyp	alphanumeric	min. 4 chars	4 to 16 chars	●	○	○
ASAP Market	any	min. 6 chars	min. 4 chars	●	○	○
Bohemia	any	min. 7 chars‡	4 to 10 digits	●	●	○
Incognito	alphanumeric	min. 8 chars	system assigned	●	●	●
Kingdom Market	alphanumeric	min. 5 chars	6 digits	●	●	○
Nemesis Market	alphanumeric†	min. 5 chars	min. 4 chars	●	●	○
Royal Market	alphanumeric	min. 8 chars	4 to 6 digits	●	●	○
Tor2door Market	alphanumeric†	min. 8 chars	6 to 10 digits	●	●	○
Vice City Market	alphanumeric	min. 6 chars	4 to 12 chars	●	○	○
Chinese Exchange Market	system assigned	min. 8 chars‡	8 to 25 chars	○	●	○
cabyc	alphanumeric	min. 8 chars‡	8 to 24 chars	○	●	○

* lowercase only; † underscore and dash allowed; ‡ combination of uppercase, lowercase, number and/or special characters

Username

The username is used as part of the login credentials. It is not only used to display the identity on the market, but also is part of the account security mechanism. Most markets support alphanumeric only, but there are exceptions, one market's username is automatically assigned by the system, and two markets have almost no restriction (i.e. any character including special characters). The minimum length of usernames is one character, but four characters is the most common minimum requirement. The maximum length of usernames is sometimes set at 16 or 20 characters, but half of the markets in our study had no length limit (i.e. we achieved successful registration with more than 64 characters). Special characters are mostly not supported, but underscore and dash are accepted in two markets.

Password & Personal Identification Number (PIN)

Table 7 shows the minimum password requirements of the twelve studied dark web markets. In addition to minimum password length requirements, only three markets force users to set more complex passwords (i.e. a combination of uppercase, lowercase and numbers or/and special characters). Surprisingly, in two of the markets, there is an obvious maximum password length limit. This is not a good strategy, with a market having a maximum length limit of only 16 characters. In terms of PINs, all markets in our study have PINs for payment-related activities. But they employ quite different policies. Some requested numbers only, while some can be set to be as complex as the password. One exception is that a market gives a secret word after user registration, but functions similarly to a PIN. In other markets, the PIN is set when the user registers.

Mnemonic & Multi-factor authentication (MFA)

Due to the highly anonymous nature of the dark web market, the user registration process does not use any identifiable personal information, including the email

address and phone number we commonly use on the clear web. Mnemonic and PGP keys are used for the same purposes on dark web markets. Mnemonics are given by the market when registered and are usually a set of English words or a long, meaningless string. Users need to save the mnemonic phrase in a safe place to use it to recover their account in certain situations. Nine out of twelve markets have mnemonics for account recovery. MFA is mostly implemented through the PGP key. Users need to set up a PGP public key in the market first, and then the market will send verification information and encrypt it using that public key. Users can use the private key to obtain this verification information. MFA is often optional when browsing listings but mandatory for purchasing items. There is a market that uses a third method, which requires the users to enter the mnemonic phrase every time they log in. *Incognito* also requires users to enter their mnemonic phrase each time they log in, in addition to using PGP as MFA.

Account Kill-switch

This allows users to set a time limit in advance. When the account is inactive over the time set, the account will be automatically deleted by the market. Currently, this feature is a one-off, meaning the countdown will stop when the account is logged in again, and the user will need to set a new time limit. We only noticed this feature in one market (*Incognito*), but some markets support manual account deletion. However, we are not aware of any mechanism in the market to delete user accounts that have been inactive for a long time, although we believe this may exist but not be reported or well-known.

5.3.3 Financial Security

In this subsection, we explore information about financial security related to doing transactions while using the market. Financial security is essential for both users and vendors, as it could attract and maintain their loyalty by offering more

Table 8: An overview of the selected dark web markets' financial security (●= yes, ○= no)

Markets	Currency Allowed			Transaction		
	BTC	XMR	Others	Multisig	Escrow	FE
Abacus Market	●	●	○	●	●	●
Archetyp	○	●	○	○	●	●
ASAP Market	●	●	○	○	●	●
Bohemia	●	●	○	●	●	●
Incognito	●	●	○	○	●	●
Kingdom Market	●	●	●	○	●	●
Nemesis Market	●	●	○	○	●	●
Royal Market	●	●	○	○	●	●
Tor2door Market	●	●	○	●	●	●
Vice City Market	●	●	○	●	●	●
Chinese Exchange Market	●	○	○	○	●	○
cabyc	●	○	●	○	●	○

FE: finalise early

selections. Table 8 presents an overview of allowed currencies and transaction types in our observed markets.

Currency

Not surprisingly, Bitcoin (BTC) remains the dominant currency in the markets we observed, despite its poor anonymity properties and easy traceability. On the other hand, Monero (XMR) has become a popular choice due to it being significantly more anonymity-focused than Bitcoin [7, 11]. In practice, Monero is considered the best choice by the dark web community. With Bitcoin, buyers and vendors are advised to use *mixers* to avoid tracking and enhance their anonymity [69, 154]. There are many markets that offer a variety of options (i.e. two or more currencies) to provide flexibility for users and vendors. The market usually has a real-time exchange rate for currency conversion, even though most item prices are shown in USD. That means the buyer will pay the corresponding cryptocurrency according to the exchange rate. Additionally, Litecoin (LTC) and

Tether (USDT) are also available in some markets, but in a very small percentage. Litecoin has the same anonymity features as Bitcoin, so next to zero, and using USDT in this context is clearly a mistake because USDT is controlled and tracked by a private institution. Dai (cryptocurrency), with its decentralised features would be a much better option than USDT, but both pale in comparison with Monero.

Transaction

In this study, we examine and describe three different transaction functions: *Multisig*, *Escrow*, and *Finalise early*. The market offers various transaction functions for vendors to select, allowing them to pick one (or more) that best fits their operations. Buyers can then select from the transaction functions supported by the vendors during trading.

Multisig, also known as multi-signature, refers to a transaction only made after being agreed upon by two or more parties. This method is quite attractive within the dark web community because of the extra security it provides, especially when it comes to avoiding market exit scams. Thanks to this, even if the market is closed, transactions can still be completed (as long as the buyer and vendors have another communication channel, which is usually a public dark web forum) if the items have been shipped. At the same time, when using this mechanism, the transaction funds do not need to go through the market's wallet address.

Escrow is the most basic transaction model, that is, the market acts as a somewhat trusted "third party" between buyers and vendors. Typically, buyers will need to deposit funds into the market's wallet first, which then will be reflected on the market's interface. After purchasing an item, the funds will still be held in the market wallet. When the buyer confirms that the item has been delivered, the market operator will release the funds to the vendor. At this time, the funds are still (possibly) in the market address, and the vendors need to withdraw them

to their private address. Escrow seems to solve the trust issue between buyers and vendors relatively well, but introduces extra security risks due to the needed trust on the market operators, which frequently exit scam.

Finalise early refers to the common possibility that, once the buyer places an order, the funds could reach the vendor's wallet directly, even before the items have arrived. This process shortens the escrow time. On the other hand, this increases the risk of fraud to buyers by vendors. Therefore, the market usually only allows vendors with a certain reputation (or that pay a certain deposit to the market in advance) to enable this function. This mechanism is, in general, more beneficial to vendors than to buyers, because it ensures a faster turnover of funds and mitigates the impact of market exit scams.

5.3.4 Support and Complaints

At the core of dark web markets, a user support system capable of dealing effectively and efficiently with complaints is quite necessary. Every market has this mechanism where it acts as a middleman to resolve any disputes between buyers and vendors. The whole process is integrated with the market and is done on a case-by-case basis. In theory, data on dark web markets, including private messages related to support and complaints (as well as transaction details, customer information, etc.), is encrypted using one or more encryption algorithms. This is to protect their users and themselves from potential criminal evidence. But few markets have made any statement about this. Therefore, while this is implemented as a functional security mechanism, its technical security remains highly unclear.

5.4 Discussion

We see that, unsurprisingly, the actual implementation of security mechanisms is actually very relevant to almost every aspect of market operations. This is reasonable, as some users go to dark web markets with the intent of hiding their true identities, in order to conduct potential illegal activities without legal consequences. Market operators also know the importance of security mechanisms, and often build their infrastructure with security at its core.

This section discusses the implications of these security mechanisms, especially with regard to market closure and the challenges these mechanisms pose to data collection. We also describe the deployment of market-associated forums. Finally, we explain some ethical considerations that need to be taken into account in this line of research.

5.4.1 Implication of Security Mechanisms on Market Closure

Dark web markets have always been very dynamic and full of potentially unknown features or quirks. We noted that three markets (i.e. *Royal Market*, *Tor2door Market* and *Vice City Market*) were closed at the time of writing the first draft (December 2023) of this work. Following on, when we were revising and extending this chapter (August 2024), we noted two more markets (*Bohemia* and *Incognito*) were closed down. We also mentioned that the closure of a dark web market has some correlation with the security issues of the market in the next chapter (Chapter 6). One reason why this correlation might exist is that markets are definitely more prone to shut down when under attack. Also, it is reasonable to expect that markets offering very poor levels of security can become low hanging fruit for competitors and law enforcement.

For the first three markets with unknown reasons for closure (i.e. *Royal Market*,

Tor2door Market and *Vice City Market*), a review of the security mechanisms listed in Table 6 suggests that relatively weak DDoS protection (either the lack of a waiting queue, or the use of weak CAPTCHAs, or both) could have been a contributing factor. This weakness would make them more prone to being disrupted by a third party. This would often result in the loss of data and trust, which in turn might encourage users to switch to competitor markets.

On the other hand, the implementation of security mechanisms reflects the operator's business philosophy, even though this may change at any time. On the *Bohemia* market, there was betrayal and division within the operations team. However, not long after (August 2024), the dark web community believed that law enforcement agencies arrested the administrator(s) based on a news report [19, 102]. On the *Incognito* market, even more shocking to the dark web community, the market operators turned to extorting and blackmailing both users and vendors, claiming they had the on-site transaction details, shipment information, private messages, etc. The operators threatened to disclose them to authorities. Nevertheless, according to information from law enforcement agencies [134], the founder of the market was arrested a few months later.

From the perspective of dark web users and vendors, there are almost no security mechanisms to prevent or mitigate such events because the operators are the ones who implement those mechanisms. The rich security mechanisms implemented actually attract users and vendors to join. However, driven by the huge potential profits, the market almost always has its way and closes at a specific and maximally profitable time.

While our results may not be exhaustive, we believe the security mechanisms implemented by the market interact somehow with potential market closures. That is to say, markets that want long-term stable operations will pay close attention to user experience and security. These factors will, in turn, attract users

to trade on the market. We exclude impounded retirement and voluntary retirement because these two reasons for closure are often affected by more complex factors [80, 143].

5.4.2 Implication of Security Mechanisms on Data Collection

Data collection on dark web markets has always been a challenge in this field of research. Most mechanisms are not present after login to affect crawler access, with the exception of CAPTCHAs and rate limits. There are also other case-by-case solutions that work for certain market websites.

Regarding using automation (including machine learning) to solve CAPTCHAs on the dark web, Audran et al. [5] have verified that this is feasible with decent accuracy and performance. However, the authors focus on the clock CAPTCHAs and its variations. There are many other types and variants of CAPTCHAs used on the dark web market, but we agree with their conclusions based on our experience and the data we collected. There is a trade-off here, after the time we spend in an effort to crack a given CAPTCHA, the market may no longer operate or change to new CAPTCHAs. This is why we need a general CAPTCHA solver based on AI, but we strongly believe the state of the art is very close to achieve this.

Moreover, we argue that cracking (or knowing) rate limiting thresholds is more valuable, even if this requires some upfront experimentation. We can obtain complete data faster by using multiple threads to run the crawler simultaneously. Nevertheless, very aggressive rate limiting settings can also affect access by real customers. We wonder if we can design a crawler that can use an adaptive method to adjust the request rate dynamically instead of a set of predetermined values or a range of values.

In addition, and somewhat surprisingly, in our experience the use of I2P makes

data collection easier. Tor's network performance can cause the market server to take significantly longer to respond than over the I2P network. When considering data collection, the crawler's downloader will require additional time to wait for the data to be downloaded locally [68]. Moreover, certain markets implement security mechanisms slightly differently on the Tor network than on I2P. For example, the *Bohemia* market on the I2P network does not use anti-phishing mechanisms although it does on the Tor network.

There may also actually be specific solutions for ad-hoc crawlers (rather than universal crawlers) of certain dark web markets. To our knowledge, most crawlers benefit from cookies obtained after manually solving CAPTCHAs, thereby simulating human visits to pages to obtain target data.

During our research, we noticed that there is a market where the product listing data can be obtained by submitting a single request to the server API. Since the way this market obtains data on the front end of the web page is through a simple API, we are able to pass larger parameters to this said API to obtain all the data in JSON at once. Considering the file format characteristics of JSON, data transmission and formatting are very efficient and do not put more stress onto the server.

We also found that since JavaScript is generally not used on the dark web for security reasons (e.g. JavaScript can be used to execute some malicious code), the structure of the web page is simpler than those on the clear web (i.e. there is less dynamically loaded content).

5.4.3 Market-Associated Forums

Forums associated with dark web markets, though often overlooked, provide crucial platforms for users to report scams and receive important updates, boosting transparency and trust. They also boost market security through open communication, user feedback, and effective problem-solving.

Most markets have associated forums on *Dread* or on their own servers. Those *Dread*-based forums are usually moderated by market operators and often jointly moderated by *Dread* administrators. The purpose is to have a relatively open platform for exchanging information, which includes release announcements, promotions, feedback, complaints, etc. Since such *Dread*-based forums are not part of a specific market, feedback and complaints, for example, are more independent, or their purpose is to be a communication infrastructure with other users in the same market. A common example of this is when a user is treated unfairly by a vendor, the user could take to the forums to tell what happened to them and use it as a warning to other users. Additionally, when the market website goes down for various reasons, the market operators are able to issue announcements and solutions timely. Another example is when a market operator disappears (e.g., due to a scam exit or arrest), *Dread* administrators often come together with other users to discuss the matter, verify possible truths, and mitigate losses.

The other type of forum is internal to the market and is therefore owned and moderated by the market operators. Some markets also have Telegram group chats, which are moderated by the market administrators, even though it is no longer considered a traditional dark web scope.

5.4.4 Ethical Considerations

Ethical considerations are very important for research in this area. When conducting security experiments (e.g., access restrictions and rate limiting), we carefully adjust parameters to ensure that our experiments do not affect the market's servers. We only visit the market from an observer's perspective and do not attempt any unnecessary actions out of this study. Please refer to Section 3.6 (in Chapter 3) for further details on ethical considerations during data collection.

In fact, there are many more aggressive security tests we could have done,

such as the rate limiting and the OWASP Top Ten[4]. But for ethical reasons, we figured we did not want to be a potential attacker. On the other hand, some markets offer bug bounty programs that actually allow for a certain level of agreement to conduct some security testing on the market. Bug bounty programs are discussed in Section 5.3.1. A well-implemented bug bounty programme has the potential to enhance the security of dark web markets significantly. Meanwhile, such programmes may also influence their longevity and success. However, these vulnerabilities are often exploited by competitors or law enforcement to disrupt operations. Our intention is not to advocate for improving the security of such markets, particularly as this could inadvertently impede law enforcement efforts.

Further dialogue and discussion should be necessary within the law enforcement agencies and academic communities to reduce barriers. This will also help both parties better understand the responsibilities and needs of the other party, and further promote understanding of the dark web area.

5.4.5 Limitations and Future Work

It will be beneficial to expand the data sources, which will help the academic community gain a broader understanding of the security measures of the dark web markets and design crawlers for data collection in a more targeted manner. Moreover, a better and more detailed understanding of how rate limiting works may greatly improve the efficiency of crawlers and reduce manual labour. There are promising signs that we will soon be able to use a collection of scripts that can universally solve the kind of CAPTCHAs encountered on dark web markets, which will be a very useful breakthrough. It is envisaged that machine learning techniques can be used to solve them quite easily in the near future. We are only able to cover the end-user side of security mechanisms, which may not be comprehensive. But we also raise ethical considerations for academics on how to

[4]`https://owasp.org/www-project-top-ten`

properly and ethically improve research in this area.

5.5 Chapter Conclusion

In conclusion, this chapter presents the outcome of our investigation into how security mechanisms are implemented on mainstream dark web markets. In particular, we highlight that the twelve dark web markets we observed have different levels of security to protect themselves and their users. At this stage, we believe that using manual labour in data collection on the dark web market is unavoidable though we would not be surprised if soon enough this can be automated.

Our results reveal that the security mechanisms implemented by mainstream dark web markets include web security, account security and financial security. Web security includes accessibility, waiting queue, anti-phishing, CAPTCHAs, secret phrase, warrant canary, bug bounty, rate limiting. Account security includes username requirement, password & PIN requirement, mnemonic/seed phrases, MFA, account kill-switch. Financial security includes the choice of (crypto)currency being used, specific transaction functions such as multi-signature, escrow and finalise early, as well as support and complaints handling. We also discuss how security mechanisms being implemented (or not) by market operators may reflect the operators' business philosophy (for instance, whether they plan to stay long in the business). We share some insights regarding data collection and the key challenges associated with it. We also describe the deployment of market-related forums and the roles they play. Finally, we discuss some ethical issues that need to be considered in this line of research.

Chapter 6

An Analysis of Closure of Dark Web Markets

This chapter is based on the content of previous publication:
"Dark Ending: What Happens when a Dark Web Market Closes down" [143]

6.1 Introduction

Dark web markets are one of the main economic hubs of illegal online activity. Similar to the legitimate online markets, as time goes by, some dark web markets flourish, some wither, new ones are opened and some close down. However, unlike legitimate online markets, the ending of dark web markets is usually unannounced, difficult to predict, and frequently shrouded in mystery. At times, even disinformation might take place. This opens up an interesting challenge for cybercrime researchers, and we try to address this through the work presented in this chapter.

There have been several instances of high-profile dark web markets being closed down. For example, *Hydra*, a Russian-language dark web market, was shut down by law enforcement agencies (LEAs) on 5 April 2022 [127]. The LEAs involved in this operation have indicated that, even after shutting down the servers and

confiscating around €23 million in Bitcoin, they fear this will not end the *Hydra* cybercrime gang, as it has proved quite difficult to identify who was behind it [127].

Apart from LEA operations, most closures are referred to as "exit scams", in which the market operators chose to close the market without prior notice, thus stealing any funds in temporary escrow from both vendors and buyers. In 2020, for example, the operators of the largest dark web market at the time, *Empire Market*, performed an exit scam and got away with around $30 million in Bitcoin [115].

In rare occurrences, the operators would "gracefully" close down the market, i.e. they would inform all customers in advance, allowing extra time for ongoing orders to be completed and any remaining funds to be transferred to the appropriate parties. In 2021, *White House Market* did just that, via an announcement on their website stating that the project had already reached their goal and that they were retiring as planned [152]. The market operator immediately stopped the registration of new users, and they ceased to accept new orders on the site. They finally closed the site down after existing vendors fulfilled their open orders.

Nevertheless, new markets steadily appear to compete with existing ones – and to replace closed-down ones. We may never know whether the same people behind the existing or closed-down markets are running those new ones. For instance, an operator that previously performed an exit scam could launch another market with the same objective of exit scamming. In contrast, those operators that closed down their old market gracefully might transfer the reputation and skills they have built up to the new market.

Previous studies have investigated various aspects of dark web markets, but to our knowledge, none has specifically focused on the data collection and analysis of how, why or when dark web markets closed down. Thus, it is important to dig further into the ending phase of dark web markets, not only to improve our understanding, but also to help reduce the risk of people getting exit scammed, and to assist LEAs in securing evidence before these markets disappear.

While previous work has examined user records of Bitcoin transactions to analyse the unexpected closure of dark web markets [92], our work collected data directly from six markets and their associated forums due to the trend of not using Bitcoin (Monero instead) on existing dark web markets. By including multiple markets, we aim to increase the breadth of our understanding. This will also allow us to conduct meaningful comparisons among various instances of closed-down dark web markets, which can lead to more useful insights.

As such, the study presented in this chapter aims to *understand what typically would happen before and after the closing down of dark web markets*, and *whether they have any common characteristics*. If we were able to identify some common features, we would also like to know *whether we can use these to predict whether a market is about to close down*.

6.2 Methodology

This section explains our approach, mainly covering the data collection process and the ethical considerations. We also describe the crawling strategy of our customised crawler software, and provide an overview of our datasets[5].

6.2.1 Approach

In order to understand what happened before and after the dark web markets being closed down, we collected data weekly[6], and analysed data from six dark web markets over a period of time before they closed. These six dark web markets are *Cartel Marketplace*, *Dark0de Reborn*, *The Versus Project*, *White House Market*, *Monopoly Market*, and *Tea Horse Road*. We also collected data from five of their

[5]Due to the potentially criminal content of the datasets, we had to choose an appropriate and ethical way to share them. We are happy to share our datasets with the academic community, security researchers and LEAs.

[6]Some weeks' data may not be collected for unexpected circumstances, such as server down-time or crawler errors.

Table 9: Reasons for the closure of 21 major dark web markets since September 2019

Market Names	Reasons	Closure
Apollon Market	Exit scam	2020-01
Aurora Market	Exit scam	2021-04
BitBazaar	Exit scam	2020-07
Cartel Marketplace	Exit scam	2021-12
Dark0de Reborn	Exit scam	2022-02
Empire Market	Exit scam	2020-08
Grey Market	Exit scam	2019-12
Silk Road 3.1	Exit scam	2020-07
World Market	Exit scam	2022-03
Yellow Brick Market	Exit scam	2021-01
CannaHome Market	Voluntary closure	2022-04
Cannazon Market	Voluntary closure	2021-11
Dream Market	Voluntary closure	2019-04
The Versus Project	Voluntary closure	2022-05
ToRReZ Market	Voluntary closure	2021-12
White House Market	Voluntary closure	2021-10
Big Blue Market	Taken down by LEAs	2021-04
CanadaHQ	Taken down by LEAs	2022-01
Dark Market	Taken down by LEAs	2021-01
Hydra Market	Taken down by LEAs	2022-04
Monopoly Market	Taken down by LEAs	2021-12

associated forums on *Dread*. The data from forums are only collected once, as those forums have been marked as archived, which means no new threads would be made after the archived date (usually a few days after the associated market has been closed down).

Cartel Marketplace, *Dark0de Reborn*, *The Versus Project*, *White House Market* and *Tea Horse Road* are comprehensive markets where drugs, fraud-related material, stolen data and ransomware are all listed. All of these markets use some sort of escrow mechanism to maintain the operation of the market. *Monopoly Market* was advertised as a wallet-less and drug-focused market. Nevertheless, it seems that direct payment to vendors is only available to a select group of vendors

with a good reputation at the time [44].

We categorised these six markets into three categories based on the different endings in which they were closed down. The criteria used to determine the category of each market are based on publicly available information. For reference, we roughly counted the reasons and time for the closure of major markets since September 2019, including 21 markets in English. Ten of them were considered exit scams, six voluntarily closed down, while five were raided by LEAs, as shown in Table 9. Due to the timing of this study and other limitations, we have only got data from six markets.

Dark0de Reborn and *Cartel Marketplace* shut down their sites and deleted the administrator's accounts on the forum without any prior notice. With some users complaining on the forums and no statement from the LEAs, we believe this is a classic exit scam. Voluntary closures include two markets, which are the *Versus Project* and the *White House Market*. The former chose to retire due to potential security concerns, and the operator sent private links to the vendors to access the market to complete transactions. The latter announced its retirement in a post on the website and immediately stopped accepting new orders. The admins claimed that the public link to the site would no longer work after all orders were completed. *Tea Horse Road* is a Chinese dark web market. A few months after its abrupt closure, screenshots of its home page appeared in reports about the fight against cybercrime. It is therefore identified as having been shut down by the LEAs. Similarly, the reason for the closure of *Monopoly* was due to the servers being seized, as claimed by the operator of *dark.fail*[7] and [56].

Smaller markets may not be very active in forums, but operators may introduce "cross-posts" to keep the market buzzing. The forums in *Dread* allow "cross-post", which means there are threads that can appear in one or more forums. For example, someone may post a review about a vendor in */d/review*, which can

[7] `https://twitter.com/darkdotfail`

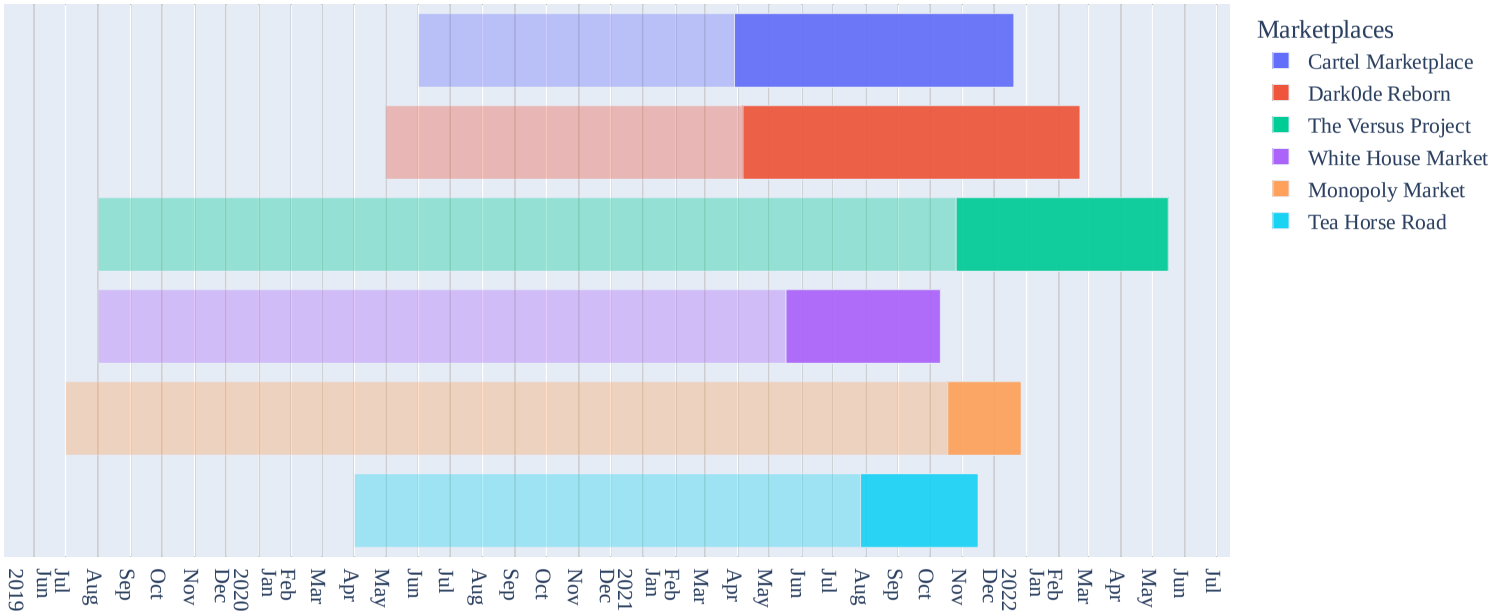


Figure 26: The time period of the data collection for each of the six dark web markets observed and their entire lifecycle

Table 10: A summary of the datasets obtained (for the analysis of market closure) *This market does not have a forum in *Dread*

Market Names	Dates Covered (from/to)	# snap.	#Dread Threads	Size
Cartel Marketplace	2021-03-29/2021-12-20	38	701	352.3 MB
Dark0de Reborn	2021-04-06/2022-02-21	46	4976	542.1 MB
The Versus Project	2021-10-26/2022-05-16	31	3713	5.3 GB
White House Market	2021-05-17/2021-10-11	21	8793	315.0 MB
Monopoly Market	2021-10-18/2021-12-27	11	599	658.4 MB
Tea Horse Road	2021-07-27/2021-11-16	16	N/A*	117.8 MB

later be re-shared in */d/versus* as well. Therefore, when calculating the number, we count all the data in the forum, i.e. including the "cross-posts". We then comb through the results to find more meaningful insights, such as how users shift between markets and discuss them.

6.2.2 Data Collection

We discuss the design of our crawler in detail in Chapter 3. For this study, our crawler employs two strategies to collect data on dark web markets:

- In situations where the site's security mechanisms would allow crawlers to operate at higher speeds with no restrictions – i.e. the site's sessions would not (or rarely) expire after a certain time, as long as the crawler keeps interacting with the site – our crawler would access and collect the details of each product through the listing page.

- In situations where the site would apply a strict security mechanism – whereby the session would expire after a specific number of requests, and then a CAPTCHA would be enforced – we tried to use multiple accounts to crawl in parallel and did our best to get statistical data other than text.

Figure 26 shows the timeline for our data collection for different dark web markets. The light-coloured area represents the overall life cycle of the market,

Table 11: A summary of the observed dark web markets (for the analysis of market closure)

Market Names	First Seen	Last Seen	Lifetime	#Vendors
Cartel Marketplace	2020-06	2021-12	18 Months	237
Dark0de Reborn	2020-05	2022-02	21 Months	2640
The Versus Project	2019-08	2022-05	33 Months	937
White House Market	2019-08	2021-10	26 Months	3450
Monopoly Market	2019-07	2021-12	29 Months	162
Tea Horse Road	2020-04	2021-11	19 Months	3275

i.e., the light-coloured left edge, which is when this market was first seen. The darker areas represent the time periods covered by our quantitative data. The start time of collection varies for each market, but the end time is the last time it is accessible. During these periods we obtained data once a week, so we could analyse the differences over time. For the *Dread* data, they were collected once on 14 August 2022, as those forums have been marked as archived. Table 10 provides information on the dataset obtained for each market. Table 11 shows some of the basic characteristics of the markets observed. In the dataset, we note that some markets use the EUR to display prices, and some use the USD. Given the volatility of exchange rates, we have not converted them as the analysis of trends is limited to individual markets. Still, we do make high-level comparisons based on trends on individual markets.

6.2.3 Ethical Considerations

Since we had to collect data on the dark web (the Tor network), and the data could potentially be related to cybercrime activities, we had to be very careful in dealing with the ethical issues of our research. Please refer to Section 3.6 (in Chapter 3) for further details on ethical considerations during data collection.

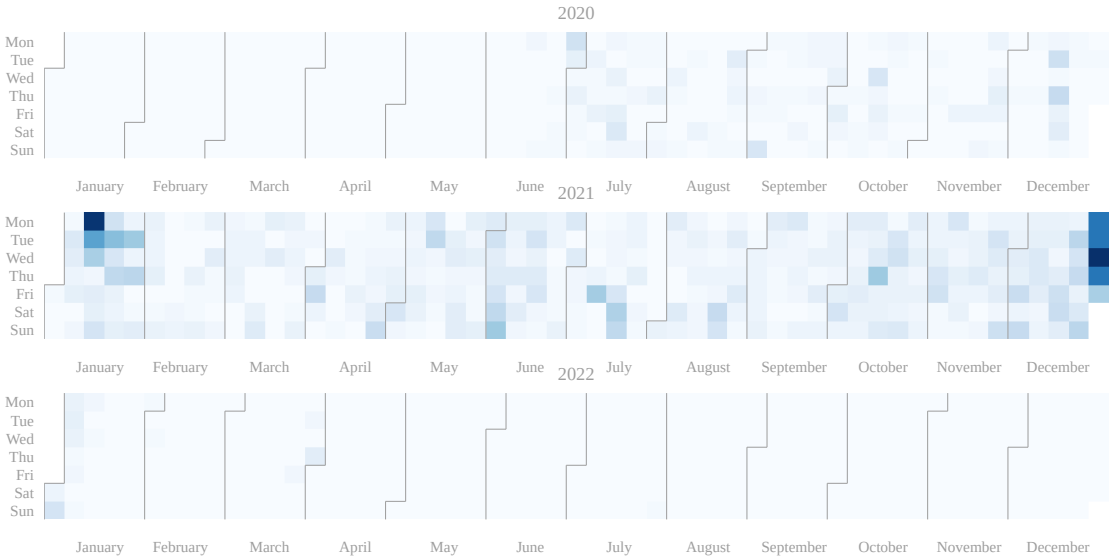


Figure 27: A heat map of the number of comments on *Cartel Marketplace Dread* forum (darker colours mean higher numbers).

6.3 Results

In this section, we categorised six markets into three categories based on the different endings in which they were closed down. We describe some of the key things that happened before the closure, and also try to analyse different indicators depending on the availability of the data.

6.3.1 Exit Scams

Exit scams appear to be the most common type of closure, where the operator closes the site without any notice and takes all of the user's funds in their wallet. This happens when markets operate with escrow mechanisms. The escrow mechanism means that the market is a third party for vendors and buyers. The buyer deposits a certain amount of money into a cryptocurrency wallet provided by the market, and the fund is only released to the vendor's wallet when the transaction is completed.

Cartel Marketplace was launched in June 2020 and closed down in December

2021. The lifetime is about 18 months. Figure 27 shows the number of posts in the *Cartel Marketplace* sub-forum. In January 2021, that actually had an official announcement from *Dread* dominating the discussion. At the same time, there were plenty of advertisements from *Cartel* operators to attract new vendors and users. December 2021 is the month when the market closes and disappears. The problem was first identified on the 21st of December, when the market was suspected to be under DDoS attacks and down for a few hours. Users also started asking in the forums for a time for the market to return. On the 24th, probably the last appearance of the market operator. On the 29th, the forum administrator announced the market had been exit scammed, as the market operator did not reappear again, and the market website had not been online since the 21st.

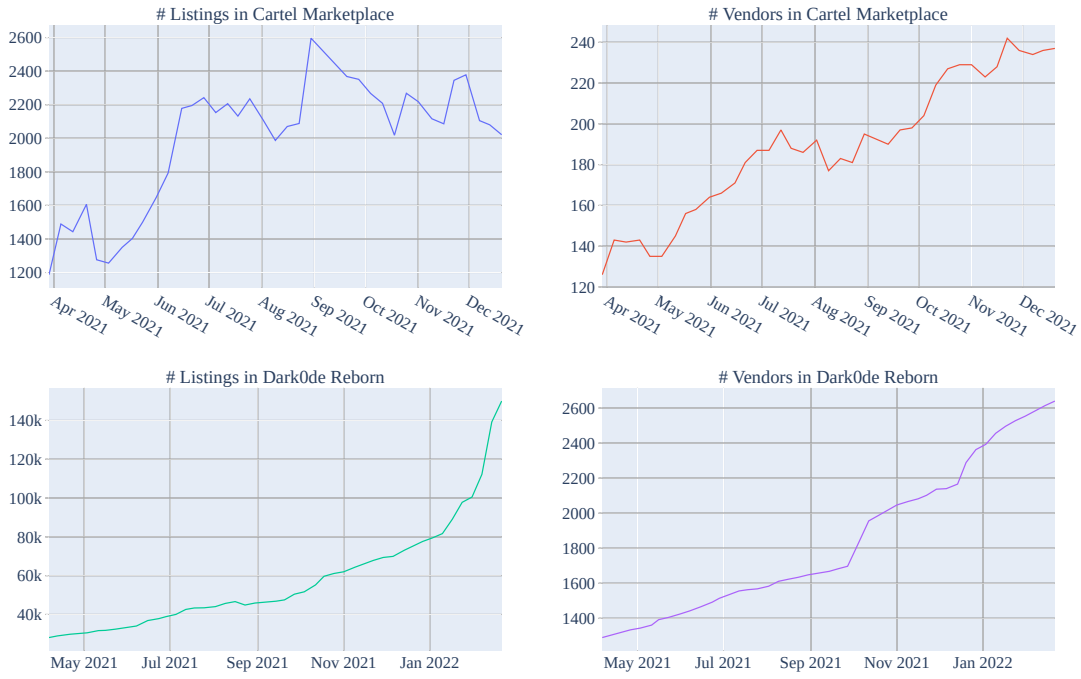


Figure 28: The number of listings on *Cartel Marketplace* (top left), the number of vendors on *Cartel Marketplace* (top right), the number of listings on *Dark0de Reborn* (bottom left), and the number of vendors on *Dark0de Reborn* (bottom right).

Dark0de Reborn was launched in May 2020 and closed down in February 2022. The lifetime is about 21 months. Almost the same thing happens in this market. With the DDoS attack at the beginning of the month, it seemed they had the ability to bring the site back to normal. When the end of the month came, the operators disappeared. Unlike *Cartel Marketplace*, we did not observe many complaints, but people moved quickly to other alternative markets.

Figure 28 shows the number of listings and vendors on both markets, which both markets keep increasing overall. An exception is in the number of listings on the *Cartel Marketplace*. The number peaked in September 2021, and then it started to decline. However, we did not find any interesting factors that could affect the number, and it was very quiet in the forum instead. In October, *Cartel Marketplace* operators began advertising for the recruitment of new vendors, while the closure of *White House Market* led some to transfer to this market, which is reflected in the growth of the charts. Therefore, we suspect that the drop in figures may be due to a small number of "dishonest" vendors (or "rippers" called in dark web communities) being banned from the market.

Interestingly, vendor numbers rose rapidly about two months before the *Dark0de* market closed. However, this was seemingly due to the closure of other markets leading to vendors changing places. In addition to *Cartel Marketplace*, another larger market closed at that time. Similarly, the closure of the *White House Market* is reflected in the increase in vendor numbers at the end of September and the beginning of October. It was also from this time (about four months before the market closed) that the discussion volume on the forum increased rapidly.

On the economic side, we have tried to analyse the median price and number of sales of the *Cartel Marketplace* in Figure 29a. The number of sales has maintained consistent growth. The median price maintained a downward trend. In particular, median prices fell rapidly in August, and listing numbers did improve at that time, which should have influenced the overall results. It is worth noting that its

(a) Median price and number of sales on *Cartel Marketplace*

(b) Median price and number of sales on *Versus Project*

Figure 29: Median price and number of sales comparison on two markets with different exit types

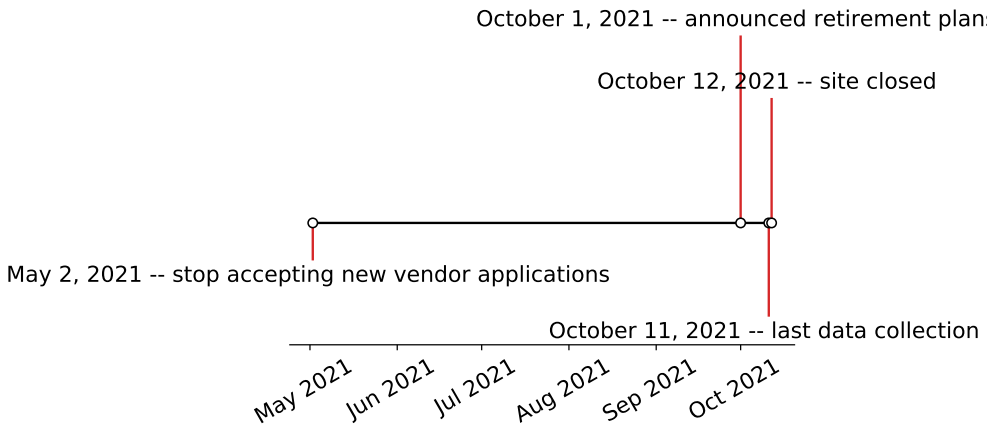


Figure 30: Key events occurring on *White House Market* before it was closed down

median price fell again in the final weeks of the *Cartel Marketplace*. Interestingly, a vendor claimed the market operators have secretly revised stock quantities and reduced the prices of productions in the back end of the server. This behaviour is considered profitable for the market operators, as attracting more orders means more funds go into the market wallet (due to the application of escrow mechanisms on the market).

6.3.2 Voluntary Closures

Voluntary closures are usually a "win-win" for users and operators, which inform all customers in advance and allow extra time for ongoing orders to be completed. However, although this ending is usually less common in the past, it happens more frequently nowadays.

White House Market was launched in August 2019 and closed down in October 2021. The lifetime is roughly 26 months. Figure 30 shows the timeline of some key events before the market's closure. The market first announced in early May 2021 that it would no longer accept new vendors. On 1st October 2021, the market owners claimed their retirement, i.e. a voluntary closure of the market. We were

allowed to access the site for the last time on the 11th, and then the site was shut down on the 12th. It took about 12 days from the announcement to the market's closing. Everything looks graceful from an observer's point of view, yet the truth could be different. On the same day that the market was closed, many vendors and users complained they did not get their coins back. These people have lost money either due to open orders or open disputes. Therefore, the forum's administrators marked the market as a dishonourable exit.

A different story took place in another market. The *Versus Project* was launched in August 2019 and closed down in May 2022. The lifetime is roughly 33 months. On 5th May 2022, the market operators claimed to have transformed the market into an invite-only community to maintain the quality of support, including invite-only vendors and invite-only buyers. Over the next few days, other forums appeared to discuss a major security breach on the market. Finally, in a statement dated 22nd May 2022, the operator described the fact that the market had a security breach and decided to close the market down. Unlike *White House Market*, the administrator did not disappear from the forum after the website was closed directly. Instead, after about four weeks, market administrators announced a link to complete all transactions.

The number of listings on both markets is growing steadily. The number of vendors on the *Versus Project* market has also continued to grow without many surprises. Figure 29b shows the median price and number of sales on the *Versus Project*. It should be noted that the currency unit of the price is EUR. Sales volumes are steadily increasing, but there are fluctuations in the median price. The median price is in a downward trend from November 2021 to January 2022, and then begins to rise until mid-March 2022. Sales also increased faster at that time, which may be the possibility that *Dark0de* closed at that time and caused many users to move in. After that time, the median price fluctuated between €75 to €85.

Figure 31: Median sales volume and number of active disputes on *Monopoly Market*

6.3.3 Taken Down by LEAs

This is usually the hardest type to define, as it is difficult to establish authenticity across different sources of information other than the LEAs making a statement. The LEA may operate a market as a honeypot for a period of time after taking control of it before shutting it down, which sometimes looks like a voluntary closure.

Monopoly Market was launched in July 2019 and closed down in December 2021. The lifetime is about 29 months. It did not seem to have any attacks or exceptions until it was shut down. After closing, it was identified as sized by LEAs, claimed by the operator of *dark.fail*. The numbers of vendors and listings were quite stable, with an upward trend. As some vendors withdrew, the number of listings on this market began to decline in early December. But we can see in Figure 31 that there were still some disputes resolved at the end of November, while the median sales volume was still growing. This is considered a fairly normal pattern, and the market was growing rapidly.

Tea Horse Road is a Chinese dark web market which was launched in April 2020 and closed down in November 2021. The lifetime is about 19 months. The screenshots of its home page appeared in reports about the fight against cybercrime a few months after its abrupt closure. The numbers of vendors and listings were shown, where both numbers were rising continuously. The median price rose from $5 to $20 in the two months before the market closed, then remained flat.

Monopoly Market has been developing for over two years, and it has developed rapidly in the last two months, benefiting from the *White House Market* exit bringing some users. The market has a good reputation, and even with a slight loss of vendors, sales are still stable. *Tea Horse Road* is also in a very smooth development stage, and all indicators are developing in a good place. Therefore, the LEAs have reason to crack down on fast-growing markets to deter criminals in their infancy.

6.4 Discussion

Based on the results we observed, there are no significant indicators to show whether a market is heading for closure. However, we have gained some insights that may be useful for warning users that a market is going through some "difficulties", and these dynamics may lead to further moves by its market operators (including closure).

6.4.1 Insights

The life cycles of the six markets we observed were all greater than 18 months, with the largest being 33 months. This may be a bias caused by the fact that we picked the more popular markets when crawling, but the markets we picked contain different sizes. Therefore, we have reason to believe that, in the early days of some little-known reputable markets establishment, there is a high probability

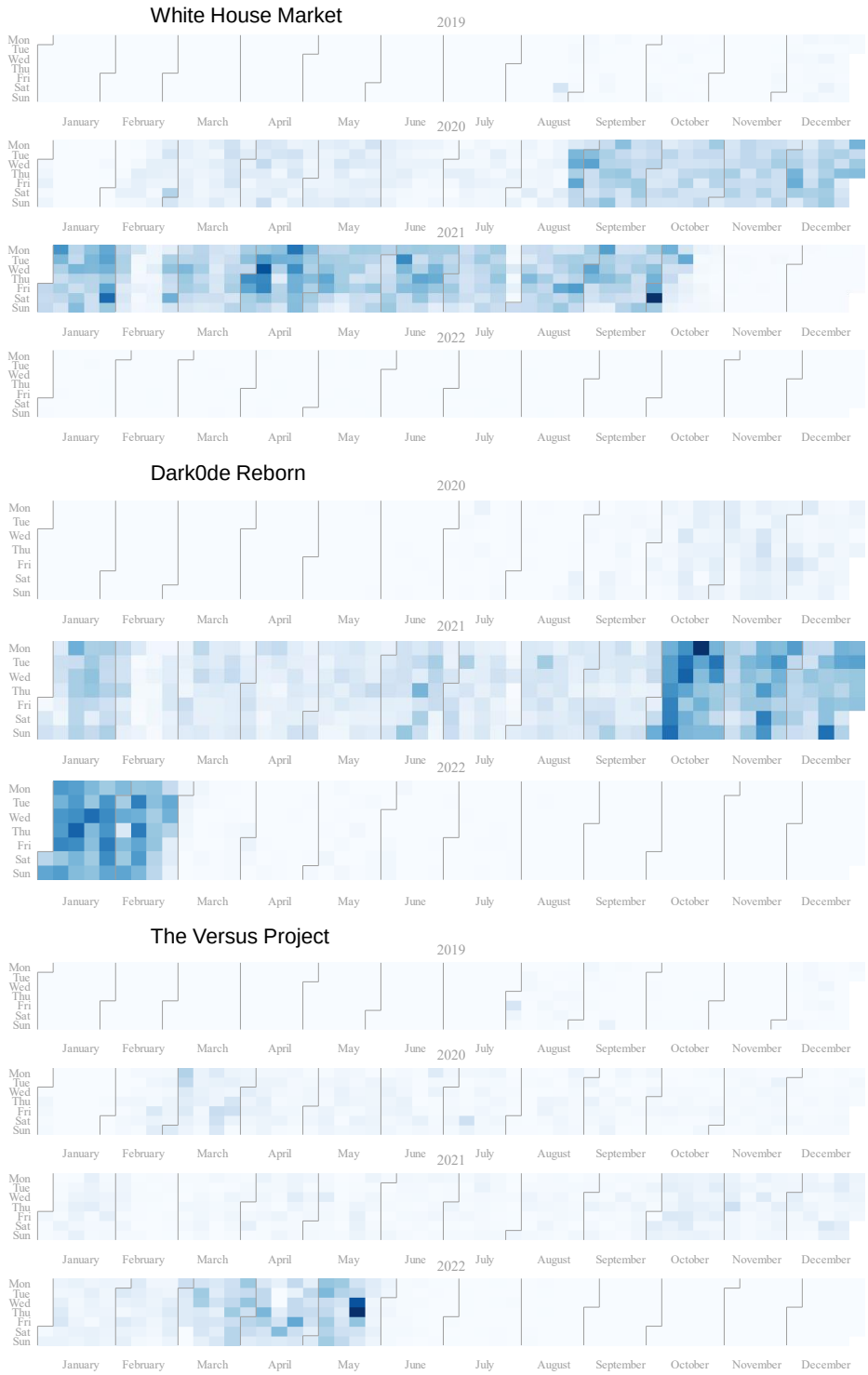


Figure 32: A heatmap of the volume of comments on *Dread* sub-forums of *White House Market*, *Dark0de* and *Versus Project*, showing clear transitions of users (reflected by the comments' volume – the darker the square, the higher the volume) from one market to the next; Note that the timelines are synchronised, although *Dark0de* only covered the last three years

that they will not suddenly disappear. However, after a certain number of users, sales and profits have been achieved, the risk of closure becomes greater.

LEAs may be more interested in fast-growing markets, since the larger dark web market has a greater negative impact on society. Also, the market operators may have certain psychological expectations. For example, when a certain amount of profit is reached, they will try to prevent the market from becoming too exposed and uncontrolled – for instance, the market operators may choose to close it down in order to keep themselves safe from LEAs' take-down.

On the other hand, once growth is slowing down, the risk of a market closure begins to increase. Based on our results, when the median price falls, this may indicate a decline in the market economics to attract more customers. Dark web market operators may choose to exit at this time, meaning they try to get the last profit. However, it requires further studies to solidify the explanation of this observation.

For similar reasons, we believe that markets that are not accepting new vendors are trying to become more "closed" communities because they may have significant circular revenues and do not want to take more risks. Nevertheless, several security issues (e.g. DDoS attacks), and possibly other reasons, have led market administrators to opt for the more conservative side – either exit scam or voluntary closure.

We also notice that people usually move to other popular alternative markets when a market closes. This is reflected in the data collected from our study, in which the increase in the number of comments associated with one market appears to coincide with the closure of another. Figure 32 shows the heat map of the volume of comments on *White House Market*, *Dark0de* and *Versus Project*, where darker colours mean more comments. We notice clear boundaries where people moved to the *Dark0de* market after *White House Market* exited, and to *Versus Project* after *Dark0de* was closed down. As users become active, conflicts

and problems may arise, but within two to four months, the market would either calm down and settle, or disappear into the darkness.

From another previous study, Bradley [13] identified that the dark web market ecosystem had shown resilience to market closures. This study used statistics to confirm that the relocation of vendors after the closure of the market was not randomly selected. While the age of the market is not a significant variable, the size of the market is. This actually confirms our similar findings. That is, in our cases, after a mainstream market is closed, users' discussion volume in other mainstream markets increases, while the number of posts on the affiliated forums of the closed market drops sharply. The vendors who are less affected will "forget" the closure after a period of time. However, there are also variables that are harder to measure, such as the reliability of the market (i.e., uptime and functionality) and the diversity of available products. Also, the study found that this was related to where buyers seemed to be turning and community consensus. Future work is expected to attempt more qualitative analysis, such as textual analysis, in a collection of forums related to dark web markets.

6.4.2 Challenges

Data collection is considered to be a challenge in this study. Firstly, we do not yet have the ability to predict which markets will close in the near future, so we can only do our best to collect data on some markets and then analyse them after they have closed down. Secondly, data collection is influenced by the accessibility of the market website. Dark web markets are often attacked by various parties, which may be LEAs or competitors. This makes the downtime for some markets very long, causing the crawling process to be interrupted and making the data incomplete.

Moreover, the security mechanisms of some markets result in a limited number of requests being sent at a time. For instance, *Cartel* market only allows 300

requests to be made in a session over a period of time (approximately 40-60 minutes). Therefore, we tried to use multiple accounts for parallel crawling, but were still limited by the site's measures not being able to access the full content of the market. In addition, we used two different software packages and two different strategies for data collection (see Section 6.2).

6.4.3 Limitations and Future Work

The data points obtained are not very comprehensive due to the security mechanisms implemented on some of the markets' sites. The main problem was due to the CAPTCHA employed on these sites causing our crawler to be disrupted. Additionally, our dataset contains relatively short snapshots (approximately 2-9 months) of the observed markets' data, even though the markets we observed all had lifetime greater than 18 months. Finally, the markets in our dataset represent only a small number of existing markets; as such, some bias might have been introduced as a result.

It would be interesting to look further into the behaviour of cross-market vendors when a market is closed down. We observed that many vendors are selling in different markets simultaneously, which means they would suffer some losses when a market they are operating in was closed down. However, they do not seem to be too concerned about these losses and try to maintain their reputation by, for example, actively seeking out purchasers in relevant forums.

We also expect more long-term observational research on the dark web markets in general, for instance, to better understand the development and evolution of a dark web ecosystem due to its dynamic and unpredictable nature.

6.5 Chapter Conclusion

In our study, we collected data from six dark web markets and five associated forums to investigate what happens when dark web markets are closed down. We describe and analyse several indicators for such events. The results showed that even though the markets may be closed down for various reasons, they still have some interesting commonalities.

Both exit scams and voluntary closures are more likely to happen when the market's economy starts to change (i.e. not in line with its own "normal" economic pattern). Measuring the stage of development of a market may depend on indicators such as the number of vendors, the median price, sales volume and the number of disputes.

It is also important to note whether the market continues to accept new vendors or not. If the market administrators are not looking to accept new vendors, they might want to be more stealthy or the periodic profit has likely met their expectations (which could mean they might try to become an invite-only community or simply shut down at some point). As for markets being shut down by LEAs, those markets seem to be in a period of rapid growth and showing no signs of slowing down – then suddenly disappear.

After a market closure, users and vendors will quickly move to other markets with a good reputation. However, after two to four months, these alternative markets will most likely go into the next darkness. We believe that these insights provide a way to gain a more comprehensive understanding of the development of dark web markets. We also hope our research will draw the attention of the academic community to this often-forgotten dynamic on the dark web market.

Chapter 7

Case Study: Investigating the Availability of Child Sexual Abuse Materials in Dark Web Markets

This chapter is based on the content of previous publication:
"Investigating the Availability of Child Sexual Abuse Materials in Dark Web Markets: Evidence Gathered and Lessons Learned" [146]

7.1 Introduction

Child sexual exploitation and abuse (CSEA) is recognised as a serious crime, especially in Western countries. CSEA affects millions of children worldwide [126] and is continuously growing [105], although still under-reported [54, 84]. It is considered a European Union priority in the "fight against organised and serious international crime" for the period of 2022-2025 [55], and is the subject of a National Strategy in the United Kingdom [74]. The first known mega forum of CSEA

operating on the dark web was taken down in 2017 by the Taskforce Argos, led by the Australian Police [89]. Since then, the dark web has been increasingly used for CSEA-related activities by providing a higher level of anonymity compared to the surface web. The dark web is currently the main platform for discussion, consumption and distribution of CSAM [60, 96]. In particular, the dark web has become the hub for like-minded individuals to share CSAM, exchange tips and manuals to avoid detection, encourage abuse, and reward contributions under a structure and *modus-operandi* akin to other organised crimes [51].

It is therefore important to understand this harmful ecosystem better in terms of markets, in order to devise further effective countermeasures to fight CSAM. Previous studies have focused on the motivation and the broad categorisation of CSAM content on the dark web [17, 18, 83]. However, a question remains regarding whether there are some relevant (and revealing) CSAM-related details on the dark web markets' characteristics, such as CSAM products on offer and the words used in advertising or describing these CSAM products, market policies and trends, as well as other operational aspects. This is the key research question that the study presented in this chapter aims to address.

In this chapter, we report findings and lessons learned from our investigation into the features and characteristics of CSAM on both English and Chinese dark web markets. A couple of prior studies have found that CSAM is most likely to be found on Chinese dark web markets [142, 161]. We aim to confirm or refute those preliminary findings by conducting a more in-depth investigation of this distressing (yet interesting) social and cultural observation.

The rest of the chapter is organised as follows. We explain the methodology used in this work, including how we screened and obtained our keywords, our data collection process, and the ethical considerations in Section 7.2. We show our key results and findings in Section 7.3, while we discuss the implications of those results, as well as some limitations of our current research, along with ideas

for future work in Section 7.4. Finally, we conclude this chapter in Section 7.5.

7.2 Methodology

In this section, we explain the methodology we followed in our study. We also describe our ethical considerations and the careful handling of the collected data.

7.2.1 Approach

In order to understand what terms are used in practice to refer to CSAM and related activities, we used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) approach [101] to build a list of related terms. We incorporated these terms into a customised crawler (that we developed and configured specifically for this study) to collect the pertinent data from the dark web, and analysed the data obtained.

PRISMA Eligibility Criteria

Literature related to the search terms and/or keywords used to find or sell CSAM on the dark web was selected for inclusion. The creation, discovery, or refinement of search terms and/or keywords for CSAM on the dark web had to be included in the methodology, results, or discussion of relevant literature. Given the anonymity of the dark web, there were no inclusion criteria specific to age or demographics. Book chapters, dissertations/theses, commentaries, conference papers and publications, annual reports, government and official documents, and legal materials were included.

Literature would be excluded if the method for finding CSAM on the dark web was solely done through hash values (e.g., PhotoDNA), file or folder names, age or body part detection, or IP addresses. Literature would also be excluded if the focus – while related to CSAM – was unrelated to the finding or selling of

it on the dark web (e.g., longevity of CSAM, structure of dark web). Further, literature would be excluded if its ultimate focus was on the ethics or philosophy of censorship on the dark web. Trade publications, magazines, news articles, newspapers, wire feeds, and encyclopedias were excluded.

Search Strategy

The following databases were searched from May to July of 2022: SCOPUS (`https://www.scopus.com/`), IEEE Xplore (`https://ieeexplore.ieee.org/`), ACM Digital Library (`https://dl.acm.org/`), and EBSCO (`https://www.ebsco.com/`), which included PsycNet (`https://psycnet.apa.org/`) and PsycInfo (`https://www.apa.org/pubs/databases/psycinfo`). Reference screening and contact with authors furthered the identification of relevant literature and retrieval of keywords and/or search terms used to find or sell CSAM on the dark web. Following data extraction, searches were completed on Google Scholar to ensure that pertinent literature was not missed.

Search terms were chosen according to their presence in the relevant literature. They are intentionally broad to ensure the inclusion of pertinent literature and to account for the varied terminology for CSAM in particular. The boolean operator "AND" was added between the two separate search categories. Table 12 shows the search terms used in our search queries. Translated search terms in Chinese were also used in the China Online Journals database [141] (`https://www.wanfangdata.com.cn/`).

CSAM Related Keywords

Sixteen articles (all in English) were selected in the end [6, 61, 62, 63, 65, 73, 85, 94, 112, 113, 124, 125, 138, 149, 150, 151]. The authors from eight sources were contacted for the full list of CSAM keywords or search terms mentioned in their articles. Two articles were subsequently removed due to a lack of response

Table 12: Search terms used in the literature search queries

	("child sexual abuse material" OR "child sexual exploitation material" OR "indecent images of children" OR "illicit images of children" OR "child pornography" OR "child abuse material" OR "child abuse images" OR "child sexual abuse markets" OR "child exploitation market" OR "online child sexual exploitation" OR "child sexual abuse images" OR "child sexual abuse images online" OR "online child sexual exploitation and abuse" OR "child sexual exploitation and abuse")
AND	("dark web" OR "dark net" OR "darknet" OR "deep web" OR "network" OR "livestream" OR "peer to peer network" OR "surface web" OR "peer to peer sharing" OR "The Onion Router" OR "webcrawl" OR "webscrape")

or inability to produce the requested materials. Combining the search terms and keywords from all sources resulted in 669 unique words, terms, or phrases associated with CSAM in varying degrees of proximity.

We then screened, selected and divided the obtained English keywords into three groups: (1) keywords directly related to CSAM, (2) keywords related to young-age, and (3) keywords related to sex.

To the best of our knowledge, there is no literature reporting on Chinese keywords related to CSAM. Therefore, an effort was made to translate English keywords into Chinese to fill this gap. These keywords were snowballed and updated during the screening period. It is worth mentioning that several keywords in English may be covered by one keyword in Chinese. Let's take the English term "paedophile" which can be translated in Chinese to "恋童" or "恋童癖". The first translation is composed of two terms: "love" ("恋") and "child" ("童"). Therefore, the latter Chinese term is enough to represent "child", "paedophile", "kid", "baby" and so on.

Table 13 provides a summary of the 198 CSAM-related keywords (and their

Table 13: The four groups of keywords that we obtained and screened (n=198, case insensitive) – the asterisk (*) represents a wild card denoting any letter(s)

CSAM related keywords (n=53)	babyj, baby cduk, babyjdog, babyshivid, childfugga, childlover, childporn, childsex, ddoggprn, eurololita, fallenangelfuns, halyavapictures, hussyfan, kdquality, kdv, kidzilla, kinder, kinderficker, kingpass, lola*, loli*, loll*, lsm, lso, lsobar, mafiasex, nimphets, nympet, nymphets, paedo*, pedfilia, pedo*, pedphilia, phtc, pthc, ptsc, qqazz, QWERTY, R@ygold, raygold, reelkiddymov, yamad, youngvideomodels, pjk, rbv, hmv, komorka, jagget, gomon, cjb, propthc, t4c, tihjj
Young-age related keywords (n=31)	{"X y", "X-yo", "Xyo", "X year", "X years", "X years old", "X y", "X ano", "X y/o"} (where X is between 0 and 17, or "one", "two", "three" etc.), adolescent, babe*, baby, bebe, boy*, chaby, child*, diaper, enfant, florian boy, gamine, girl*, infant*, kid*, kindergarten, post-pubescent, postpubescent, pre-pubescent, prepubescent, preteen*, pretten, toddler, tween, underage, young, hairless, little, smooth, school, angels
Sex related keywords (n=100)	anal, anale, animalsex, anus, ass, assfuck, asslick, BDSM, bitch, blowjob, bondage, defloration, dfloration, dildo, doggy, doggysex, ejac, ejaculation, erotic, eurosex, exhib, facia, fellation, fetish, fisting, fuck*, gangbang, gay, groupsex, handjob, hardcore, hentai, incest, inzest, jailbait, JOP, lesbian, lickin*, liluplanet, lingerie, masterbate, masterbating, mastur*, naked, nakie, naturist, necrofilia, nude, nudity, nudist*, oral, orgasm, orgy, penetrat*, penis, pnis, pntration, porn*, prostitu*, purenudism, pussy, rape*, sex*, shemale, sodom*, soumise, spank*, sperm*, suce, suck*, swallow, transexual, twink, vagina, virgin, voyeur, webcam, whore, xxx, yasuda, zofilia, zoofilia, zoophilia, cum*, upskirt, sado*, nua, nse, jizz, jeezy, abuse, spread*, torture, shower, erection, pee, piss, cunt, jacking off, cock*
CSAM related keywords in Chinese (n=14)	萝莉(lori), 幼(young), 初中(junior school), 高中(high/senior school), 中学(middle/secondary school), 小(young/little/small), 岁(age/year), 童(child), 年轻(young/teen), 少(young), 妹(sister), 弟(brother), 孩(child), 未成年(underage)

classification into four groups), in both English and Chinese languages. The English keywords were collated from our literature survey, while the Chinese keywords were compiled based on the translation of the English keywords.

Data Collection & Analysis

We used Python with the Scrapy web-crawling framework [90] to implement a customised crawler, which ran weekly as part of previous studies. In the initial phase, we validated the data collected to ensure that valuable textual information was obtained rather than media files (i.e., images and videos). We identified eight dark web markets in English and two in Chinese after the keywords had been sorted. The eight English markets were: *AlphaBay Market, ASAP Market, Tor2door Market, Vice City Market, Colombia Connection, Nemesis Market, Royal Market, Kingdom Market*. The two Chinese markets were: *Chinese Exchange Market* and *cabyc* (initials of Chang'an Nocturnal City in Chinese). From this initial phase, we found that those eight English markets did not contain relevant textual information related to CSAM. Therefore, as per our preliminary results, we resumed data collection and processing for both Chinese markets, creating a new text-based dataset for further analysis.

7.2.2 Ethical Considerations

Due to the nature of the dark web and CSAM, no images were collected during the data collection process, and the process was automated. This means that we only got the textual content of the pages through a pre-defined process. We also removed the tags with *¡img¿* in the HTML code and then replaced all *src* attributes with invalid values. Please refer to Section 3.6 (in Chapter 3) for general details on ethical considerations during data collection.

7.3 Results

In this section, we present our findings and results, including market policies, market trends and the characteristics of CSAM products.

7.3.1 Markets' Policy

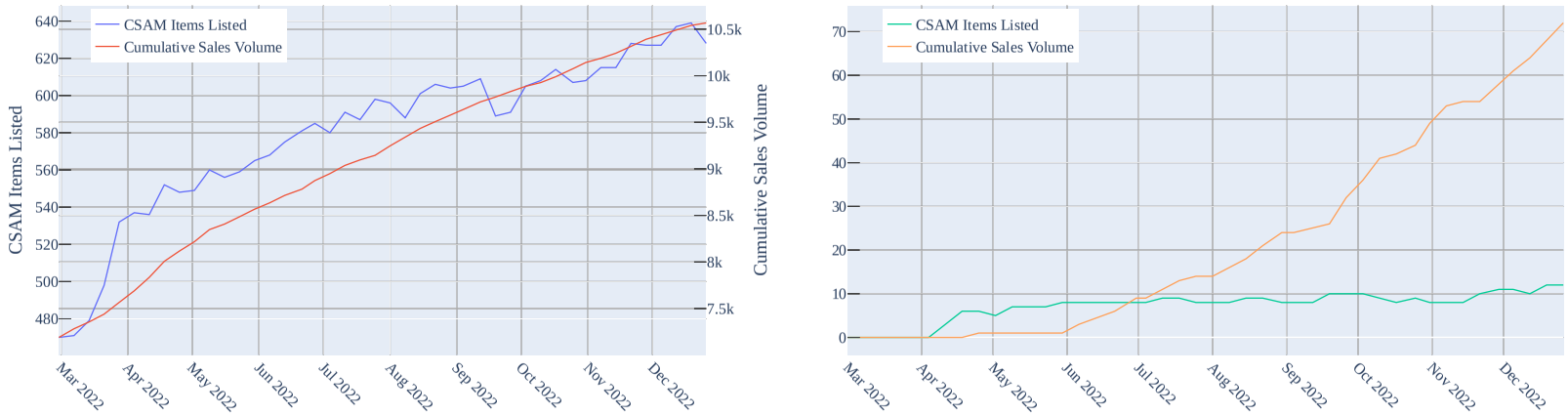
We reviewed and checked ten popular dark web markets, as detailed in the previous section. We also attempted preliminary data collection at this stage. We quickly found that there are stricter policies on the English markets that prohibit the sale of CSAM products. However, CSAM products were present on the Chinese markets.

The mainstream English dark web markets usually ban pornographic materials of any kind. For example, the *AlphaBay Market* states the following market policy: "No erotica/porn/softcore of any sorts (logins for major sites are okay). Child porn, animal abuse videos cover this rule - we have never previously allowed anything like this and we will not start doing now as it is something we strongly oppose." [2].

In this case, "logins" refer to login credentials for accessing third-party clear web (pornographic) sites; such credentials are often sold on the dark web markets. The *ASAP Market* is an exception in terms of this kind of policy; its market rules only vaguely specify that no threats or actual violence are allowed.

We attempted further checks but no CSAM products were returned. As a result, we did not find any items related to CSAM on those major English-language dark web markets. Instead, there were many vendors who sell "premium accounts" (usually hacked accounts or for cashing stolen credit cards) from clear web porn sites – but not the CSAM items themselves.

In comparison, we found a number of CSAM items on offer on the Chinese dark web markets. Regarding the market policy, *Chinese Exchange Market* does not specify nor restrict which specific products are available or unavailable for sale. *cabyc* do claim that "This site prohibit child pornography and fake cryptocurrency transactions, [...]" [24]. However, we still found CSAM items being sold on this market. This means that operators in this market may not be strict in enforcing their policy when reviewing listed items.



(a) The number of CSAM items listed and their cumulative sales volume on *Chinese Exchange Market* during the time period observed; Note that there are two different scales

(b) The number of CSAM items listed and their cumulative sales volume on *cabyc* during the time period observed

Figure 33: The number of CSAM items listed and their cumulative sales volume comparison on both Chinese dark web markets during the time period observed

Table 14: The monthly numbers of CSAM items listed and sold on *Chinese Exchange Market* and *cabyc*

	Chinese Exchange Market			*cabyc*		
Month	**# Listed**	**# Sold**	**% Sold**	**# Listed**	**# Sold**	**% Sold**
Mar-22	537	376	70%	0	0	0%
Apr-22	569	551	97%	6	1	17%
May-22	579	450	78%	8	0	0%
Jun-22	594	307	52%	8	8	100%
Jul-22	603	271	45%	9	5	56%
Aug-22	617	429	70%	9	10	111%
Sep-22	622	256	41%	10	8	80%
Oct-22	621	312	50%	11	17	155%
Nov-22	635	249	39%	11	9	82%
Dec-22	660	176	27%	13	14	108%

7.3.2 Markets' Trend

We found 724 unique CSAM-related items listed during the 44-week period between March and December 2022[8], of which 704 unique items are from the *Chinese Exchange Market* and the remaining 20 from the *cabyc* market. Given that both markets are not CSAM-focused (i.e. there are other types of items listed), the results come from their "pornography" category.

Figure 33 shows the number of CSAM items listed (when the time of each data collection was done) and their cumulative sales volume (i.e., the sum of the cumulative sales since the item was first listed, for all items) over the time period observed. The *Chinese Exchange Market* is a more established dark web market with a total sales volume of 3,377, where the earliest CSAM-related item was listed in July 2018 (see Figure 33a). The *cabyc* market, on the other hand, is new and only appeared in early 2022, therefore, its sales volume is 72 during the observation period (see Figure 33b).

Table 14 shows the monthly numbers of CSAM items listed and sold on *Chinese Exchange Market* and *cabyc*. On *Chinese Exchange Market*, the sales were greater in the first few months, and then showed a slight downward trend. The average

[8]Please note that the *cabyc* market has 43 weeks' worth of data due to website maintenance.

monthly sales volume is greater than 300. We speculate that people tend to buy newly listed items. There are two potential factors that may have affected sales: (i) there was only a small net increase in items from May 2022, compared to the previous two months, and (ii) DDoS issues on the Tor network in October 2022 caused low accessibility[9] [128].

On *cabyc*, as this was an emerging market at the time, there was only one sale in the first three months of observation. However, as the market operated longer, more consistent transactions began to be made, even with the limited number of items listed compared to *Chinese Exchange Market*. Tor issues do not appear to have had an impact on this market.

It is worth highlighting that these two markets have slightly different policies: *Chinese Exchange Market* has no strict restriction on selling CSAM items; whereas *cabyc* does but it is not enforced well.

7.3.3 Characteristics of Items on Sale

We collected 724 records from 156 different vendors. Figure 34 shows the number of newly listed CSAM items (monthly) on the *Chinese Exchange Market*, based on our observed data. The earliest items appeared on the market in July 2018. The market became more active after July 2021 with at least ten items being added almost every month.

Product descriptions often use indecent language to describe scenes. A large number of vendors also include images in their product descriptions, as we found the filename extension for images in their descriptions. Regarding delivery, vendors use web-based cloud storage services for distribution. For example, the most commonly used is MEGA.

Multiple CSAM items are usually included in one purchase. Moreover, some

[9]Even though such issues still exist at the time of writing, there was a significant impact on accessibility on this particular market in October 2022.

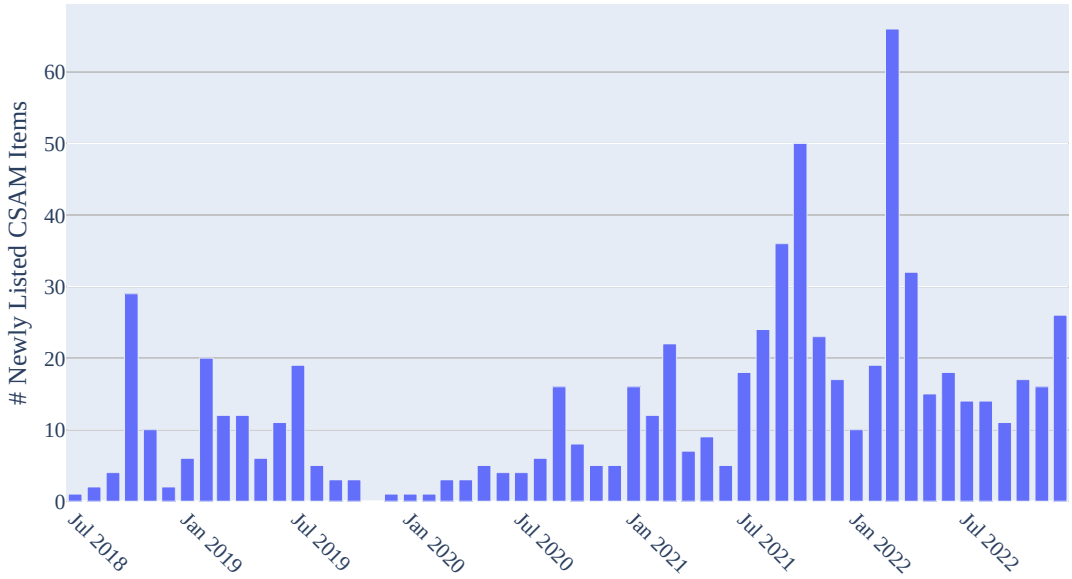


Figure 34: The number of newly listed CSAM items on *Chinese Exchange Market* monthly

vendors offer a list of different types of CSEA videos to promote. We also noticed that even though these materials are sold on the Chinese dark web market, there are a large number of materials from other countries and regions on the studied Chinese markets. Although the majority of items indicated the age and gender of the child involved in the title of the post, sometimes more ambiguous terms were used, such as *lolita*, *underage*, and *secondary school student*. In some cases, the vendor claimed that the CSEA videos came from Japanese or European forums. This would suggest the presence of materials for reselling.

The median price of items on offer was very low. For the *Chinese Exchange Market*, the median price was only $5, while for the *cabyc* market it was $3. On the other hand, the median sales volumes for all CSAM items were 6.5 and 1 on those two markets, while on average, the sales volumes were 15 and 3.6, respectively.

7.4 Discussion

The results of this research provide some useful insights and evidence to better understand the current stage of CSAM trends on the dark web market.

English language CSEA dark web markets seem to adopt measures aiming to minimise the risk of being detected by LEAs, by making channels more private and restricting access to members that fulfil certain requirements [88, 153]. In China, however, since LEAs may be more focused on dealing with such harmful activities on the surface web, the Chinese dark web markets seem to operate more openly. This explains why we could collect data more easily for the latter but not the former.

Based on the nature of the observed Chinese dark web markets, the aim of the vendors appears to be more about profit-making, rather than the traditional sharing and exchange of CSAM. To some extent, since accessing the Tor network has become pretty easy (i.e. there is less need for complex network knowledge to configure and use Tor), CSAM has been commercialised on the Chinese dark web market. Furthermore, although the results show fluctuations in sales, the general trend is upward. Almost every month at least ten items have been added to *Chinese Exchange Market* since July 2021. Remember that this is a comprehensive market, so other items (e.g. stolen data, drugs, other digital items) are also listed on that market. This means that CSAM items are being exposed to a wider range of people.

It is worth noting that CSAM items from diverse countries and regions have already been appearing on the Chinese dark web markets. This means that multi-lingual speakers (or even, someone who might have been using existing automated language translation technologies) are distributing materials that were originally in their own language into foreign language markets. However, there is a lack of multicultural studies in this field. We noticed that existing studies have mainly revolved around English-speaking cyberspace, yet there are limited studies on the

dark web markets using other languages such as Chinese or Russian.

Finally, some of the CSAM items are shipped using web-based cloud storage services such as MEGA, Baidu Web Drive (Baidu Wangpan), or even Microsoft OneDrive. Those storage providers face many challenges in detecting CSAM in their platforms. First, fuzzy hashing technology (e.g., PhotoDNA) is only able to detect variations of known CSEA images, which means they do not have the ability to recognise novel images. Second, end-to-end encryption prevents even the detection of known CSEA images.

7.4.1 Limitations and Future Work

Our Chinese keywords may not be very comprehensive as they were not supported by the literature. The application of linguistics and machine learning has the potential to provide more opportunities in the future. This is because these techniques may lead to the discovery of more slang terms commonly used by criminals.

For future work, the effectiveness of different types of keywords can be assessed by the results returned. Moreover, more complex query conditions can be added when crawling data. For example, when both "pupil" and "data" are present in returned results, it is more likely to be leaked data rather than CSEA-related.

Similarly, at the time of writing, ChatGPT is becoming popular. This tool can be utilised to generate a number of keywords for retrieving CSAM and has the ability to be improved by users' input. This could pose a huge challenge for the future. For example, users could potentially abuse similar tools to avoid keyword detection; more emerging, obscure terms may appear in the CSEA community.

7.5 Chapter Conclusion

In this chapter, we aimed to study the CSEA landscape on the dark web markets. For that, we systematically compiled a list of CSEA-related keywords from the literature to identify potential English and Chinese dark web markets trading CSAM. These keywords were then utilised by our scraper to collect text-based datasets over a 44-week period between March and December 2022, resulting in the identification of 724 CSAM items being listed on two Chinese dark web markets but none on the eight English dark web markets observed.

In particular, we found that CSAM items were being sold on both Chinese dark web markets without the need for memberships or any other gatekeeping practice in place. The two Chinese markets selling CSAM items adopted a relaxed policy with regard to this type of product. Sales figures on both of these Chinese markets showed an upward trend with regard to the volume of the items being sold. It is worth noting that these CSAM items were sold cheaply, with a median price of $5 on the *Chinese Exchange Market*, for example. Moreover, there was an indication that CSAM items originating from other countries and regions have been appearing on the Chinese dark web markets for reselling. Finally, we noticed that mainstream web-based cloud storage services were being used for the distribution and sharing of CSAM.

Chapter 8

Conclusion

8.1 Introduction

This final chapter provides a summary of the contributions of the thesis, states the limitations of this work, and points out future research opportunities in relation to dark web markets. Finally, some final words and thoughts are presented.

8.2 Summary of Contributions

In this section, we revisit the research questions posed in Section 1.2 and discuss how our findings presented in the main chapters address those research questions.

RQ1: *Do cultural differences influence the operation and structure of dark web markets in different language communities?*

Regarding RQ1, we identified five active (at the time) mainstream dark web markets, three in English and two in Chinese, to compare and analyse their differences and similarities. We documented and described six different perspectives in terms of their organisational structure and functionality. Several differences were found in terms of website structure, products sold, payment methods, and

policies. Nevertheless, innovative sales models on the Chinese dark web market, request-to-buy, cannot be ignored. In terms of the vendors, vendors on the English dark web markets are more active than those on the Chinese dark web markets. Due to the language barrier, only people with Chinese language skills sell on the Chinese market, while English market vendors sell in a wider range of locations. The main reason for variations in advertised items may be differences in legislation and law enforcement across regions. Additionally, the language barrier also causes the dark web market in different languages to be isolated from others to a certain extent. Those markets also have different operating procedures because of cultural differences. For instance, the policies of Chinese dark web markets show that there are very few products banned, and the problem of child abuse material is extremely serious.

RQ2: *What mechanisms do dark web markets employ to ensure the protection of their operations and the safety of their users?*

Regarding RQ2, as mentioned in both Chapter 4 and Chapter 5, dark web markets apply a bulk of security measures in the three aspects we categorised, namely web security, account security and financial security. Web security includes accessibility, waiting queue, anti-phishing, CAPTCHAs, secret phrase, warrant canary, bug bounty, rate limiting. Account security includes username requirement, password & PIN requirement, mnemonic/seed phrases, MFA, and account kill-switch. Financial security includes the choice of (crypto)currency being used, specific transaction functions such as multi-signature, escrow and finalise early, as well as support and complaints handling. We believe that the security mechanisms implemented by the market reflect the thinking of the market operators. The implementation of comprehensive security mechanisms attracts both users and vendors to join certain markets. However, driven by the prospect of significant profits, the market is likely to exit scams during the most profitable periods. Our

findings not only provide insights into the characteristics of the studied markets, but also point out possible directions for future crawler designs.

RQ3: *What trends can be identified in the development of dark web markets over time, and what underlying factors contribute to these trends?*

Regarding RQ3, we mainly captured the trend of dark web market closures and reviewed the data before the closure of six dark web markets to analyse the patterns, in Chapter 6. As we see in the previous research question, profits drive the closure (most likely exit scam) of markets. Besides that, we also identified two other ways of closing the market: voluntary closures and those taken down by LEAs. We found that when the market economy begins to shift away from its inherent "normal" economic pattern, it becomes more susceptible to exit scams and voluntary closures. But at the same time, we found it unlikely that markets closed by law enforcement can be predicted in advance. This is a good sign as the LEAs would be able to seize and obtain more evidence as a result. We believe that the main underlying factor is still profit. At the core of the dark web market, everyone is trying to make more profits. In addition, the increase in law enforcement operations (e.g., Operation DisrupTor [52] and Operation SpecTor [56]) against dark web markets in recent years has made the dark web markets themselves more dynamic, so market operators are becoming more cautious.

RQ4: *What role does the existence and development of dark web markets play in specific cybercrime?*

Regarding RQ4, we conducted an interdisciplinary case study to investigate the availability of CSAM on the dark web markets, in Chapter 7. Firstly, we did not find CSAM for sale on the mainstream English dark web markets we observed; however, we found many on the Chinese dark web markets. This shows how the dark web market has become a trading platform for profiting from CSAM on some

of the Chinese dark web markets. The anonymity of the dark web is crucial for both vendors and buyers, reducing the risk of detection by law enforcement. Due to the lack of regulation by its market operators, CSAM still exists despite the indication in the market policy that it cannot be listed. We also found that the prices at which CSAM was sold were very low, which may have further facilitated its distribution. In short, dark web markets provide a safe and anonymous environment for the dissemination of serious offences such as these. To that end, addressing the issue may require a range of innovative enforcement strategies and international cooperation to undermine and disrupt the existence of these markets.

8.3 Implications of Research

Overall, it is essential to conduct a more thorough and detailed investigation into these emerging markets. This study investigated the characteristics of modern dark web markets, examining the human aspect of cybercrime and dark web market operations. This study gathered data from various online sources, mainly the dark web, to extract insights into how these people interact and operate. These insights inform the development of effective countermeasures and protective tools.

- The characteristics of the Chinese and English dark web markets described in Chapter 4 offer a unique perspective, providing a detailed account of the differences in modern markets across various cultural contexts. This work fills the gap in the literature where there is a lack of understanding of non-English dark web market communities.

- Chapter 5 builds upon the foundation established in the previous chapter by exploring the security mechanisms used on dark web markets. By enhancing our understanding of how the market works, we will have a better chance to address the challenges encountered in the data collection process, paving the way for more effective research methods and future work directions.

- By documenting and analysing the events and data from these markets, Chapter 6 contributes to the academic understanding of dark web market dynamics, particularly in the context of market closures. The insights gained from this study not only enhance our understanding of the factors leading to market closures, but also open avenues for future research. Specifically, this work paves the way for exploring the potential of automated data analysis techniques to predict or provide early warnings of market closure risks, offering practical implications for both researchers and policymakers. This could enable policymakers to anticipate and mitigate the effects of market disruptions, such as the migration of illicit activities to alternative platforms, and help develop proactive measures to disrupt criminal networks operating on the dark web.

- The findings on the availability of CSAM on dark web markets are presented in Chapter 7. This study explores the role and impact of dark web markets in facilitating specific criminal activities, providing concrete evidence of their involvement. The compilation of English and Chinese keywords derived from our findings offers a valuable resource for future research and investigations in this domain. Furthermore, the study underscores a critical observation: the extremely low unit prices of CSAM on Chinese dark web markets. This economic factor is likely to accelerate the proliferation of such materials, raising urgent concerns for policy and enforcement strategies.

- The lessons learned from experiments and data collection can significantly enhance future research efforts. Additionally, these findings contribute meaningfully to the academic literature by documenting the characteristics of recent dark web markets. The detailed methodologies outlined in each chapter allow future researchers to replicate the research approach, even if certain dark web markets are no longer operational. All available datasets used

during this study can be shared ethically with the academic community and security researchers (see Appendix B).

8.4 Limitations

While this thesis documents and demonstrates the characteristics and trends of dark web markets in many aspects, we still need to consider the limitations. In this section, we summarise the limitations of this work.

The main limitation of this study is that it is a retrospective study based on historical data. Even though the observed market characteristics are supported by a large amount of data and evidence, it is not enough for us to infer their validity for the future. Therefore, data collected over different time periods may have different conclusions. Also, the time span of the results of this study is shorter compared to studies in other areas. Fortunately, we found some independent parallel works (which different research aims, such as [68] and [71]) reporting similar results to consolidate further our results, which are included in Chapter 2.

In addition, the crawler also potentially imposes some limitations in this study. The crawler can be detected by market operators and enter traps without our knowledge. Although we have not noticed this happening, we are not able to verify it. We did data collection on a weekly basis. As a result, we may miss items that were sold during the one-week crawl time window but were taken off the market before the next data collection. This situation could affect the completeness of the data collection.

Due to the unpredictable nature of dark web markets, it is difficult to have a hypothesis about the problem before conducting research on it, making it difficult to evaluate our results with a baseline or standard. This research is exploratory. Similarly, as academics, we may have a different perspective than law enforcement agencies, and therefore, may miss valuable findings or insights, even though we

have reported all findings in as much detail as possible.

Data from 20 dark web markets was used in this study. Data from *Loulan City* (a Chinese market) was excluded due to having too few snapshots (only 6-weeks' worth of data) to provide a comprehensive range of the data required. Furthermore, *Loulan City* disappeared shortly after, around 14 February 2022. Even with the data from these 20 dark web markets, the sample size may be a bit small to cover the whole dark web market landscape. Having said that, we made sure that representative samples were collected from the more visible dark web markets. Our choice of markets was also influenced by factors such as several markets not being accessible at all for a period of time but then becoming available again. We could miss some snapshots in those cases. Also, the selected markets are all located on the Tor and I2P networks, which may introduce biases as dark web markets on other networks are not covered.

There is more discussion of topic-related limitations in the four technical chapters; these can be found in Sections 4.4.3, 5.4.5, 6.4.3 and 7.4.1.

8.5 Future Research Opportunities

Overall, this study solved several research gaps in the context of understating the dark web market characteristics. In this section, we summarise promising future work and present insights based on our experiences.

From the findings of Chapter 4, it is noticed that cross-market actors are active and exist. Those cross-market actors usually have a relatively large influence on the dark web markets. That means they typically conduct large-scale illegal sales for profit, mostly drugs. Some markets also provide cross-market rating system, further promoting the development of cross-market actors. These actors also highlight their own ratings, especially when a market is closed, so that they can direct customers to new markets. By following up on dark web forums, we could

gain valuable insights into their operations and can even track the development of organised cybercriminal groups. Therefore, even when certain markets are closed, we are still able to track and measure the profits of these actors over the long term based on the forum data. It is also possible to track their payment and profits to attract the attention of LEAs [117, 118], enhancing the capacity to intervene in illicit activities at key points. For future work, it would be interesting to track those actors' behaviours across different markets due to their large profits.

Moreover, we are also interested in specifically measuring the economic characteristics and impacts of the market. One key observation is that vendors on the market come from various regions globally, raising questions about how prices compare for similar items across different locations. Furthermore, exploring whether issues such as currency inflation manifest on dark web markets would be a particularly intriguing area of study. During the study period, several major global events occurred around the world, which were also closely related to dark web markets. For example, [139] described the evolution of the *Hack Forums* market through the set-up, stable state and Covid-19 pandemic eras. They found that the market centralised heavily around influential users and threads. To further extend similar ideas to dark web markets, for example, we could learn about the dynamics of the dark web market (possibly prices, new products, or even new types of scams) during the Russo-Ukrainian conflict. This would rely on a more comprehensive longitudinal dataset.

In a more technical flavour, based on our findings in Chapter 5, crawlers targeting dark web markets can be designed more specifically. A script that can automatically solve CAPTCHAs is promising to be integrated into the crawler [5, 155], even if we need more efforts to adapt to more dark web markets. Campobasso and Allodi [21] proposed a promising extensible crawler designed for data collection in underground forums. Although intended to be used on forums, the crawler can be adapted with minimal effort for dark web markets, demonstrating potential

usability even for researchers without a computer science background. Future crawlers could be more flexible crawlers that are able to handle larger-scale data collections, allowing researchers to quickly deploy and adapt them as the scope of their collections changes.

Chapter 6 discusses some of the reasons for the closure of the dark web market. Whilst we have briefly mentioned the possible reasons behind market operators' decisions to close markets, further research is needed to provide stronger evidence. Direct sources, such as posts, announcements and discussions in forums, are needed to understand their decisions. These materials can provide valuable insights into operators' perspectives, especially when closures are due to internal conflict and economic reasons. Moreover, we have found that using information from affiliated forums is promising for providing early warnings regarding market closures. Ursani et al. [132] proposed an unsupervised learning method to monitor adverse events in dark web markets. Therefore, combining statistical analysis with textual analyses, such as sentiment and emotion, to monitor posts for use in early warning of potential market closures is a valuable direction for future research.

In another technical aspect, it is noticeable that the structure of some dark web markets is somehow very similar. Therefore, the source code may be open-sourced or likely to be coming from the same group of operators. In particular, as per our findings in Chapter 6, when there is a dark web market that closes down, there are always some new markets that show up. It is possible to find that the same operator manages different dark web markets or dark web forums.

On top of this research, other research directions can be carried out by combining emerging technologies such as generative AI and machine learning. With the rapid development of large language models (LLM) in the past two years, there are some potential opportunities to leverage the latest technology to aid data analysis or gain valuable insights, although this is not guaranteed yet. These models, such as GPT-3.5, GPT-4, and others, have demonstrated impressive capabilities

in understanding and generating human-like text, making them great tools for processing vast amounts of unstructured data. This capability can be used to summarise and categorise market vendor profiles, which could significantly reduce manual labour. Nevertheless, further research and testing of the capabilities of these models are needed to ensure the robustness of the results. Potential research directions can also include leveraging generative AI to interact with market operators and vendors (note that this requires further ethical considerations). For example, generative AI can be used to obtain fake security analysis reports for interaction with market operators to compare differences between different markets. Machine learning can be utilised for long-term monitoring of key vendors, such as automated categorisation of sold items to monitor sales traffic and changes in their business strategy. Machine learning techniques can also be used to quickly classify, determine and analyse the sentiment of forum posts, which are valuable for closely observing the dynamics of cybercriminals. Moreover, a data analysis framework can be developed to enable a more automated approach.

As a platform for transactions, dark web markets are closely related to many specific cybercrimes. For example, we have conducted a case study in Chapter 7 to investigate the availability of CSAM on dark web markets. Ransomware is another emerging type of cybercrime in recent years, and it is also active on the dark web markets. Previous work has identified the Ransomware-as-a-Service (RaaS) economy within the dark web [100]. Their findings show that the RaaS only has a few offered for sale, often with questionable authenticity in their measurement between 2018 and 2019. One of our works [147] also revealed the social and technical incentives that may encourage certain actors to engage in ransomware crimes. Although we have seen more and more ransomware in the news in recent years, our understanding of it is still limited. It is worth noting that some malware and toolkits, which contain source code, have appeared on the Chinese dark web

market [142, 161]. Overall, it may be worth exploring and measuring the development of ransomware on the dark web more broadly (on both dark web markets and forums), and in different languages, including (but not limited to) English, Chinese, and Russian.

Since our research mainly focuses on Tor and I2P networks, future work can also be extended to other networks such as Freenet and Riffle. Most of the current research in this area has focused on Tor, leaving other networks relatively under-explored. Figueras-Martín, Magán-Carrión and Boubeta-Puig [59] show that Freenet has better connectivity than I2P and is not affected much by anti-crawling systems such as CAPTCHA. This suggests that Freenet may host more resilient or less-documented online communities. A deeper investigation into these networks could reveal additional niche communities, their structures, and how they interact within the broader ecosystem of anonymous communities.

8.6 Final Words

It is difficult to forecast when and how this "cat-and-mouse" game will end. Perhaps, just like the trend of those dark web markets, similar cybercriminal platforms will not disappear but will only continue to exist in another form. But at the same time, more importantly, we believe it is crucial to continue supporting academics, industry professionals, policymakers and law enforcement officials in their efforts to understand and combat cybercrime and protect public privacy while effectively monitoring and regulating potential cybercrime risks.

Moreover, there is a wealth of opportunities for interdisciplinary research that are worth exploring that apply to the broader topic of cybercrime. Collaboration between sociology, psychology, criminology, law, economics and computer science can provide new insights into the nature of cybercrime. For example, benefiting from researchers with a background in computer science, experts in other fields can

gain access to higher-quality and larger-volume data [111]. Sociological research can help reveal the social dynamics that underpin organised cybercrime [147], while psychological and criminology research can reveal the individual incentives that drive people to engage in such activities [30]. Economic and public policy analysis can further help us understand the market drivers and regulatory environments that influence cybercriminal activities [43], while legal research can help improve prevention frameworks [29]. Such interdisciplinary research deepens the theoretical understanding and promises to develop more robust, practical strategies to prevent and mitigate harm in the rapidly evolving digital environment.

Finally, cybersecurity is never just a technical matter. When humans get involved, things start to get complicated because humans, by their very nature, are complex and they represent an essential part of society.

Appendix A

Crawler Code Examples

This study uses Scrapy as a framework for crawlers, benefiting from its efficiency and flexible customisability. The following code relies on Scrapy to implement only the part that needs to be customised, rather than the entire crawler engine.

The following code shows a minimal working example of a crawler that combines Scrapy and Selenium. The example has shortened the onion address.

```
import scrapy
from selenium import webdriver
from selenium.webdriver.firefox.options import Options
import time
from datetime import datetime

class DarkwebSpider(scrapy.Spider):
    name = 'nameOfTheSpider'
    allowed_domains = ['onion']
    custom_settings = {
        'AUTOTHROTTLE_START_DELAY': 11,
        'AUTOTHROTTLE_MAX_DELAY': 12,
        'CONCURRENT_REQUESTS': 6,
        # 'DOWNLOAD_DELAY': 0.5, # This is not applicable in this example because the Scrapy
        ↪  download engine is bypassed.
        'RETRY_TIMES': 9
    }

    def start_requests(self):
```

```
19          profile = webdriver.FirefoxProfile()
20          options = Options()
21          options.set_preference('network.proxy.type', 1)
22          options.set_preference('network.proxy.http', '127.0.0.1')
23          options.set_preference('network.proxy.http_port', 8118)
24
25          driver = webdriver.Firefox(options=options)
26          driver.get("http://.onion")
27          time.sleep(180)
28
29          print(driver.current_url)
30          url = str(driver.current_url)
31          global base_url
32          base_url = "http://"+url.split('/')[2]
33
34          url = 'http://.onion/listings?page=1&type=all'
35          yield scrapy.Request(url=url, cookies=driver.get_cookies(), callback=self.parse_listing)
36
37          driver.quit()
38
39      def parse_listing(self, response):
40          url = response.url
41
42          pageNum = url.split('&')[-2].split('=')[-1]
43          filename = "Listing_list-" + pageNum + "-" + datetime.now().strftime("%Y-%m-%d-%H%M%S")
            ↪  + ".html"
44          open(filename, 'wb').write(response.body)
45          if response.xpath("//a[@class='page-link']/text()[contains(.,'Next')]"):
46              l = 'http://.onion/listings?page=%s&type=all' % str(
47                  int(pageNum) + 1)
48              yield scrapy.Request(url=l,callback=self.parse_listing)
```

- Lines 10 – 16: Override default settings to adjust crawl rate.
- Line 18: Integrate Selenium to enable processes that require human intervention, such as logging in, and then pass the cookie back to Scrapy.
- Lines 19 – 26: Connect to the Tor network through a local proxy.
- Line 27: Pause the programme to allow enough time for manual operation.

- Lines 29 – 37: Pass the URL and cookie back to Scrapy.
- Lines 39: Parse the listing page and, in some cases, use XPath to extract the required data. In this example, it only shows how to save the page directly to the virtual machine.
- Lines 40 – 44: Save the listing page locally.
- Lines 45 – 48: Utilise XPath to check if there is a next page; if so, jump to the next page and repeat the above steps. If not, the crawler will stop until there are no more unrequested pages in Scrapy's queue.

The following code shows a minimal working example of a crawler that uses Scrapy only (i.e. we need to manually copy and paste the session information into the code before running the crawler). The example has shortened the onion address.

```
1  import scrapy
2  from mySpider.items import MyspiderItem
3  from datetime import datetime
4
5  class DarkwebSpider(scrapy.Spider):
6      name = 'nameOfTheSpider'
7      allowed_domains = ['onion']
8
9      custom_settings = {
10         'AUTOTHROTTLE_START_DELAY': 0.2,
11         'AUTOTHROTTLE_MAX_DELAY': 0.7,
12         'CONCURRENT_REQUESTS': 8,
13         'DOWNLOAD_DELAY' : 0,
14         'CONCURRENT_REQUESTS_PER_DOMAIN':8
15     }
16
17     header = {
18         'Accept': 'text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,*/*;q=0.8',
19         'Accept-Encoding': 'gzip, deflate',
20         'Accept-Language': 'en-US,en;q=0.5',
21         'Connection': 'keep-alive',
```

```
        'Host': '.onion',
        'Referer': 'http://.onion/index.php',
        'Upgrade-Insecure-Requests': '1',
        'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; rv: 78.0) Gecko/20100101 Firefox / 78.0'
    }

    cookie = {  # This is the initial Cookie.
        'PHPSESSID': '',
        'uuid': '012345'}

    def start_requests(self, header=header, cookie=cookie):
        for i in range(1,300):
            yield scrapy.Request(
                url='http://.onion/ea.php?area=10001&pagea=%d#pagea' % i,
                cookies=cookie, headers=header, callback=self.get_listing)

    def get_listing(self, response, header=header, cookie=cookie):
        links = []
        for each in response.xpath('//div[@class="div_length_500"]/a'):
            link = each.xpath('@href').extract()[0]
            links.append('http://.onion/' + link)

        for l in links:
            yield scrapy.Request(url=l,
                                 headers=header, cookies=cookie, callback=self.parse_page)

    def parse_page(self, response):
        #  "tbody" is tag automatically added by browsers like Firefox and Chrome.
        product = response.xpath("//table[2]/tr[2]/td/a[3]/text()").extract()
        category = response.xpath("//table[2]/tr[2]/td/a[2]/text()").extract()
        description = response.xpath('//div[@class="div_view_goods_reply"]/*/text()').extract()
        str = ';'
        des = str.join(description).strip()
        infolist = response.xpath('//table[3]//*/text()').extract()
        vendor = infolist[10]
        price = infolist[4]
        sold = infolist[-1]
        product_url = infolist[2]
        lastseen = response.xpath('//table[3]//*//td/text()').extract()[-1]
        posttime = infolist[7]

        item = MyspiderItem()
```

```
64          item['vendor'] = vendor
65          item['sold'] = sold
66          item['product'] = product[0]
67          item['price'] = price
68          item['product_url'] = product_url  # can be use as Product ID No.
69          item['category'] = category[0]
70          item['description'] = des
71          item['lastseen'] = lastseen
72          item['posttime'] = posttime
73
74          yield item
75          pass
```

- Lines 17 – 30: Before running the crawler for the first time, manually import the header and cookie into the code. Only the cookies need to be changed every time the crawler is run.
- Lines 32 – 36: This example uses a simple number loop to traverse the entire listing pages. We can also use the XPath method to determine if there is a next page.
- Lines 38 – 46: Use XPath to get the link of each product page and request it.
- Lines 48 – 75: Extract and save the data of the target data point via XPath.

Appendix B

Dataset Availability

Most of the datasets mentioned and used in this thesis can be shared ethically with researchers and law enforcement agencies. The sharing of these datasets facilitates cybercrime research and provides scientific reproducibility. Please contact the author at `yw300@kent.ac.uk`.

Appendix C

Ethical Considerations

The following document provides confirmation of the ethical review approval conducted at the University of Kent.

Central Research Ethics Advisory Group
Room 120, Rutherford Annexe
University of Kent
Canterbury
Kent CT2 7NZ

Email: centralresearchethics@kent.ac.uk

1 July 2021

To: Yichao Wang
Email address: yw300@kent.ac.uk
By email alone

Dear Yichao

Research ethics application: (Ref: 057-04-2021) `Understanding the Evolution of Cybercrime in Darknet Markets'

Thank you for your email of 30 June 2021.

I can confirm the Central Research Ethics Advisory Group (CREAG) has received the amended documents listed below and that these now comply with the conditions detailed in our letter dated 15 June 2021.

As a **positive consideration** of your research proposal was concluded by the Group, in our letter dated 15 June 2021 you may now commence with your project.

Documents received and approved

The documents received and approved were as follows:

Document	*Version*	*Date*
Checklist	V1	27 April 2021
Full Application Form	V2	21 June 2021
Research Proposal	V2	Undated

After ethical review

You are obliged to seek consideration of any changes which might affect the ethics of the project.

We wish you best wishes for the success of this project

Yours sincerely

Wendy

Wendy Atkins
Research Ethics Coordinator
Research and Innovation Services

Enc. Documents listed above

Bibliography

[1] Albery, I. and Munafò, M. (2008). *Key concepts in health psychology*. Sage.

[2] AlphaBay Market (2022). Terms of service alphabay market.

[3] Arief, B., Adzmi, M. A. B. and Gross, T. (2015). Understanding cybercrime from its stakeholders' perspectives: Part 1–attackers. *IEEE Security & Privacy*, 13(01), pp. 71–76.

[4] Arief, B. and Bin Adzmi, M. A. (2015). Understanding cybercrime from its stakeholders' perspectives: Part 2–defenders and victims. *IEEE Security & Privacy*, 13(02), pp. 84–88.

[5] Audran, D., Andersen, M., Hansen, M., Andersen, M., Frederiksen, T., Hansen, K., Georgoulias, D. and Vasilomanolakis, E. (2022). Tick tock break the clock: Breaking captchas on the darkweb. In *Proceedings of the 19th International Conference on Security and Cryptography - SECRYPT*, INSTICC, Lisbon, Portugal: SciTePress, pp. 357–365.

[6] Bagley, C. (2003). Diminishing incidence of internet child pornographic images. *Psychological reports*, 93(1), pp. 305–306.

[7] Ben Sasson, E., Chiesa, A., Garman, C., Green, M., Miers, I., Tromer, E. and Virza, M. (2014). Zerocash: Decentralized anonymous payments from bitcoin. In *2014 IEEE Symposium on Security and Privacy*, pp. 459–474.

[8] Benjamin, V. and Chen, H. (2012). Securing cyberspace: Identifying key actors in hacker communities. In *2012 IEEE International Conference on Intelligence and Security Informatics*, pp. 24–29.

[9] Bergman, J. and Popov, O. B. (2023). Exploring dark web crawlers: a systematic literature review of dark web crawlers and their implementation. *IEEE Access*, 11, pp. 35914–35933.

[10] Bermudez Villalva, D. A., Onaolapo, J., Stringhini, G. and Musolesi, M. (2018). Under and over the surface: a comparison of the use of leaked account credentials in the dark and surface web. *Crime Science*, 7(1), pp. 1–11.

[11] Biryukov, A., Khovratovich, D. and Pustogarov, I. (2014). Deanonymisation of clients in bitcoin p2p network. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, New York, NY, USA: Association for Computing Machinery, CCS '14, pp. 15—-29.

[12] Bracci, A., Nadini, M., Aliapoulios, M., McCoy, D., Gray, I., Teytelboym, A., Gallo, A. and Baronchelli, A. (2021). Dark web marketplaces and covid-19: before the vaccine. *EPJ data science*, 10(1), p. 6.

[13] Bradley, C. (2019). *On the resilience of the Dark Net Market ecosystem to law enforcement intervention*. Ph.D. thesis, UCL (University College London).

[14] Bradley, C. and Stringhini, G. (2019). A qualitative evaluation of two different law enforcement approaches on dark net markets. In *2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, IEEE, pp. 453–463.

[15] Branwen, G. (2019). Darknet market mortality risks. `https://www.gwern.net/DNM-survival`.

[16] Broadhurst, R., Foye, J., Jiang, C. and Ball, M. (2021). Illicit firearms and other weapons on darknet markets. *Trends and Issues in Crime and Criminal Justice*, pp. 1–20.

[17] Bruggen, M. V. D. and Blokland, A. (2021). A crime script analysis of child sexual exploitation material fora on the darkweb. *Sexual Abuse*, 33(8), pp. 950–974.

[18] Bruggen, M. V. D. and Blokland, A. (2022). Profiling darkweb child sexual exploitation material forum members using longitudinal posting history data. *Social Science Computer Review*, 40(4), pp. 865–891.

[19] Burnhill, E. (2024). Man charged with money laundering of cryptocurrencies. `https://www.rte.ie/news/ireland/2024/0808/1463995-cryptocurrencies/`.

[20] Cambridge Cybercrime Centre (2024). Description of available datasets. `https://www.cambridgecybercrime.uk/datasets.html`.

[21] Campobasso, M. and Allodi, L. (2022). Threat/crawl: a trainable, highly-reusable, and extensible automated method and tool to crawl criminal underground forums. In *2022 APWG Symposium on Electronic Crime Research (eCrime)*, Boston, MA, USA: IEEE, pp. 1–13.

[22] Carley, K. M. (2020). Social cybersecurity: an emerging science. *Computational and mathematical organization theory*, 26(4), pp. 365–381.

[23] Chan, J., He, S., Qiao, D. and Whinston, A. (2023). Shedding light on the dark: The impact of legal enforcement on darknet transactions. *Info Sys Research*, 35(1), pp. 145—-164.

[24] Chang'an Nocturnal City (2022). Trade rule - cabyc.

[25] China Anti-drug (2021). China's drug situation 2020 report. http://www.nncc626.com/2021-07/16/c_1211244064.htm.

[26] Christin, N. (2013). Traveling the silk road: A measurement analysis of a large anonymous online marketplace. In *Proceedings of the 22nd International Conference on World Wide Web*, New York, NY, USA: Association for Computing Machinery, WWW '13, pp. 213–224.

[27] Chu, B., Holt, T. J. and Ahn, G. J. (2010). Examining the creation, distribution, and function of malware on-line. *Department of Justice Abstract*, pp. 1–183.

[28] CNWest (2021). The "clean net" operation launched, six departments worked together to rectify harmful information on the internet. http://news.cnwest.com/tianxia/a/2021/06/08/19728920.html.

[29] Collier, B., Thomas, D. R., Clayton, R., Hutchings, A. and Chua, Y. T. (2022). Influence, infrastructure, and recentering cybercrime policing: evaluating emerging approaches to online law enforcement through a market for cybercrime services. *Policing and Society*, 32(1), pp. 103–124.

[30] Connolly, L., Borrion, H., Arief, B. and Kaddoura, S. (2023). Applying neutralisation theory to better understand ransomware offenders. In *2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, IEEE, pp. 177–182.

[31] Conrad, B. and Shirazi, F. (2014). A survey on tor and i2p. In *Ninth International Conference on Internet Monitoring and Protection (ICIMP2014)*, pp. 22–28.

[32] Copeland, C., Wallin, M. and Holt, T. J. (2020). Assessing the practices and products of darkweb firearm vendors. *Deviant Behavior*, 41(8), pp. 949–968.

[33] Covrig, B., Mikelarena, E. B., Rosca, C., Goanta, C., Spanakis, G. and Zarras, A. (2022). Upside down: Exploring the ecosystem of dark web data markets. In *IFIP International Conference on ICT Systems Security and Privacy Protection*, Springer, pp. 489–506.

[34] Cuevas, A., Miedema, F., Soska, K., Christin, N. and van Wegberg, R. (2022). Measurement by proxy: On the accuracy of online marketplace measurements. In *31st USENIX Security Symposium (USENIX Security 22)*, pp. 2153–2170.

[35] Dalins, J., Wilson, C. and Carman, M. (2018). Criminal motivation on the dark web: a categorisation model for law enforcement. *Digital Investigation*, 24, pp. 62–71.

[36] DarkNetDaily.com (2021). Interview with dark web marketplace aurora market. `https://darknetdaily.com/2021/01/08/interview-with-dark-web-marketplace-aurora-market/`.

[37] David, B., Delong, M. and Filiol, E. (2021). Detection of crawler traps: formalization and implementation—defeating protection on internet and on the tor network. *Journal of Computer Virology and Hacking Techniques*, 17(3), pp. 185–198.

[38] Décary-Hétu, D. and Giommoni, L. (2017). Do police crackdowns disrupt drug cryptomarkets? a longitudinal analysis of the effects of operation onymous. *Crime, Law and Social Change*, 67, pp. 55–75.

[39] DeepDotWeb (2015). Interview with alphabay market admin. `https://gir-pub.github.io/deepdotweb/2015/04/20/interview-with-alphabay-admin/`.

[40] DeepDotWeb (2015). Interview with german-plaza admin. `https://gir-p`

ub.github.io/deepdotweb/2015/11/04/interview-with-german-plaza-admin.

[41] DeepDotWeb (2015). Therealdeal: This long-dead market was just relaunched! https://gir-pub.github.io/deepdotweb/2015/12/01/therealdeal-this-dead-market-was-just-relaunched/.

[42] Dingledine, R., Mathewson, N. and Syverson, P. (2004). Tor: The second-generation onion router. In *Proceedings of the 13th Conference on USENIX Security Symposium*, *SSYM'04*, vol. 4, USA: USENIX Association, pp. 303–320.

[43] Dodson, M., Beresford, A. R. and Thomas, D. R. (2020). When will my plc support mirai? the security economics of large-scale attacks against internet-connected ics devices. In *2020 APWG Symposium on Electronic Crime Research (eCrime)*, IEEE, pp. 1–14.

[44] Dolejška, D., Koutenskỳ, M., Veselỳ, V. and Pluskal, J. (2023). Busting up monopoly: Methods for modern darknet marketplace forensics. *Forensic Science International: Digital Investigation*, 46, p. 301604.

[45] Dolliver, D. S. (2015). Evaluating drug trafficking on the tor network: Silk road 2, the sequel. *International Journal of Drug Policy*, 26(11), pp. 1113–1123.

[46] Dolliver, D. S. and Kenney, J. L. (2016). Characteristics of drug vendors on the tor network: A cryptomarket comparison. *Victims & Offenders*, 11(4), pp. 600–620.

[47] Ehlert, M. (2011). I2p usability vs. tor usability a bandwidth and latency comparison. In *Seminar Report, Humboldt University of Berlin*, pp. 129–134.

[48] ElBahrawy, A., Alessandretti, L., Rusnac, L., Goldsmith, D., Teytelboym, A. and Baronchelli, A. (2020). Collective dynamics of dark web marketplaces. *Scientific reports*, 10(1), pp. 1–8.

[49] EMCDDA and Europol (2018). Darknet markets ecosystem – lifetimes and reasons for closure of over 100 global darknet markets offering drugs, sorted by date. https://www.euda.europa.eu/publications/posters/2018/darknet-markets-ecosystem_en.

[50] European Monitoring Centre for Drugs and Drug Addiction and Europol (2017). Drugs and the darknet: perspectives for enforcement, research and policy.

[51] Europol (2020). Exploiting isolation: Offenders and victims of online child sexual abuse during the covid-19 pandemic. https://www.europol.europa.eu/cms/sites/default/files/documents/europol_covid_report-cse_jun2020v.3_0.pdf.

[52] Europol (2020). International sting against dark web vendors leads to 179 arrests. https://www.europol.europa.eu/media-press/newsroom/news/international-sting-against-dark-web-vendors-leads-to-179-arrests.

[53] Europol (2021). Darkmarket: world's largest illegal dark web marketplace taken down. https://www.europol.europa.eu/media-press/newsroom/news/darkmarket-worlds-largest-illegal-dark-web-marketplace-taken-down.

[54] Europol (2021). European union serious and organised crime threat assessment (socta). https://www.europol.europa.eu/cms/sites/default/files/documents/socta2021_1.pdf.

[55] Europol (2022). Eu policy cycle - empact. https://www.europol.europa.eu/crime-areas-and-trends/eu-policy-cycle-empact.

[56] Europol (2023). 288 dark web vendors arrested in major marketplace seizure. https://www.europol.europa.eu/media-press/newsroom/news/288-dark-web-vendors-arrested-in-major-marketplace-seizure.

[57] Europol (2023). Internet organised crime assessment (iocta) 2023. https://www.europol.europa.eu/cms/sites/default/files/documents/IOCTA%202023%20-%20EN_0.pdf.

[58] Europol (2024). Internet organised crime assessment (iocta) 2024. https://www.europol.europa.eu/cms/sites/default/files/documents/Internet%20Organised%20Crime%20Threat%20Assessment%20IOCTA%202024.pdf.

[59] Figueras-Martín, E., Magán-Carrión, R. and Boubeta-Puig, J. (2022). Drawing the web structure and content analysis beyond the tor darknet: Freenet as a case of study. *Journal of Information Security and Applications*, 68, p. 103229.

[60] Fonhof, A. M. P., Van Der Bruggen, M. and Takes, F. W. (2019). Characterizing key players in child exploitation networks on the dark net. In L. Aiello, C. Cherifi, H. Cherifi, R. Lambiotte, P. Lió and L. Rocha, eds., *Complex Networks and Their Applications VII*, Cham: Springer, pp. 412–423.

[61] Fournier, R. and Latapy, M. (2015). Temporal patterns of pedophile activity in a p2p network: First insights about user profiles from big data. *International Journal of Internet Science*, 10(1), pp. 8–19.

[62] Fournier, R., Cholez, T., Latapy, M., Chrisment, I., Magnien, C., Festor, O. and Daniloff, I. (2014). Comparing pedophile activity in different p2p systems. *Social Sciences*, 3(3), pp. 314–325.

[63] Frank, R., Westlake, B. and Bouchard, M. (2010). The structure and content of online child exploitation networks. In *ACM SIGKDD Workshop on Intelligence and Security Informatics*, New York, NY, USA: ACM, pp. 1–9.

[64] Franklin, J., Perrig, A., Paxson, V. and Savage, S. (2007). An inquiry into the nature and causes of the wealth of internet miscreants. *Ccs*, 7, pp. 375–388.

[65] G. Westlake, B. and Bouchard, M. (2016). Criminal careers in cyberspace: Examining website failure within child exploitation networks. *Justice Quarterly*, 33(7), pp. 1154–1181.

[66] Gatlan, S. (2023). Tor and i2p networks hit by wave of ongoing ddos attacks. `https://www.bleepingcomputer.com/news/security/tor-and-i2p-networks-hit-by-wave-of-ongoing-ddos-attacks/`.

[67] Gañán, C. H., Akyazi, U. and Tsvetkova, E. (2020). Beneath the radar: Exploring the economics of business fraud via underground markets. In *2020 APWG Symposium on Electronic Crime Research (eCrime)*, pp. 1–14.

[68] Georgoulias, D., Yaben, R. and Vasilomanolakis, E. (2023). Cheaper than you thought? a dive into the darkweb market of cyber-crime products. In *Proceedings of the 18th International Conference on Availability, Reliability and Security*, New York, NY, USA: Association for Computing Machinery, ARES '23, pp. 1–10.

[69] Georgoulias, D., Pedersen, J. M., Falch, M. and Vasilomanolakis, E. (2021). A qualitative mapping of darkweb marketplaces. In *2021 APWG Symposium on Electronic Crime Research (eCrime)*, Boston, MA, USA: IEEE, pp. 1–15.

[70] Georgoulias, D., Pedersen, J. M., Falch, M. and Vasilomanolakis, E. (2023). Botnet business models, takedown attempts, and the darkweb market: A survey. *ACM Comput Surv*, 55(11).

[71] Güldenring, B. and Roth, V. (2024). Protecting onion service users against phishing. *arXiv preprint arXiv:240807787*.

[72] Guo, J.-y. and Kong, Y. (2023). "i sometimes pretended to get groceries": Restrictive deterrence in drug dealing. *International journal of offender therapy and comparative criminology*, 67(16), pp. 1681–1698.

[73] Hammond, S., Quayle, E., Kirakowski, J., O'Halloran, E. and Wynne, F. (2009). An examination of problematic paraphilic use of peer to peer facilities.

[74] HM Government (2021). Tackling child sexual abuse strategy 2021. `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/973236/Tackling_Child_Sexual_Abuse_Strategy_2021.pdf`.

[75] Holt, T. J. and Bossler, A. M. (2014). An assessment of the current state of cybercrime scholarship. *Deviant Behavior*, 35(1), pp. 20–40.

[76] Hoptrail (2024). The re-rise of alphabay: a decentralised darknet market. `https://www.hoptrail.io/post/the-re-rise-of-alphabay-a-decentralised-darknet-market`.

[77] Howell, C. J., Fisher, T., Muniz, C. N., Maimon, D. and Rotzinger, Y. (2023). A depiction and classification of the stolen data market ecosystem and comprising darknet markets: a multidisciplinary approach. *Journal of Contemporary Criminal Justice*, 39(2), pp. 298–317.

[78] Hughes, J., Collier, B. and Hutchings, A. (2019). From playing games to committing crimes: A multi-technique approach to predicting key actors on an online gaming forum. In *2019 APWG Symposium on Electronic Crime Research (eCrime)*, IEEE, pp. 1–12.

[79] Hughes, J., Pastrana, S., Hutchings, A., Afroz, S., Samtani, S., Li, W. and Santana Marin, E. (2024). The art of cybercrime community research. *ACM Computing Surveys*, 56(6), pp. 1–26.

[80] Hutchings, A., Clayton, R. and Anderson, R. (2016). Taking down websites to prevent crime. In *2016 APWG Symposium on Electronic Crime Research (eCrime)*, Toronto, ON, Canada: IEEE, pp. 1–10.

[81] I2P (2024). The invisible internet project. `https://geti2p.net/en/`.

[82] Impact - Cyber Trust (2020). Information marketplace for policy and analysis of cyber-risk trust. `https://www.impactcybertrust.org/`.

[83] Insoll, T., Ovaska, A. K., Nurmi, J., Aaltonen, M. and Vaaranen-Valkonen, N. (2022). Risk factors for child sexual abuse material users contacting children online: Results of an anonymous multilingual survey on the dark web. *Journal of Online Trust and Safety*, 1(2).

[84] Interpol (2022). Interpol global crime trend summary report. `https://www.interpol.int/en/content/download/18350/file/Global%20Crime%20Trend%20Summary%20Report%20EN.pdf`.

[85] Joffres, K., Bouchard, M., Frank, R. and Westlake, B. (2011). Strategies to disrupt online child pornography networks. In *2011 European Intelligence and Security Informatics Conference*, IEEE, pp. 163–170.

[86] Kermitsis, E., Kavallieros, D., Myttas, D., Lissaris, E. and Giataganas, G. (2021). Dark web markets. *Dark web investigation*, pp. 85–118.

[87] KimCrawley (2021). Dread forums: The dark web's reddit. `https://www.hackthebox.com/blog/dread-forums-dark-web-reddit`.

[88] Kloess, J. A. and Bruggen, M. V. D. (2021). Trust and relationship development among users in dark web child sexual exploitation and abuse

networks: A literature review from a psychological and criminological perspective. *Trauma, Violence, & Abuse*, p. 15248380211057274.

[89] Knaus, C. (2017). Australian police sting brings down paedophile forum on dark web. `https://www.theguardian.com/society/2017/oct/07/australian-police-sting-brings-down-paedophile-forum-on-dark-web`.

[90] Kouzis-Loukas, D. (2016). *Learning Scrapy*. Packt Publishing Ltd.

[91] Kumar, M., Jindal, M. and Kumar, M. (2022). A systematic survey on captcha recognition: types, creation and breaking techniques. *Archives of Computational Methods in Engineering*, 29(2), pp. 1107–1136.

[92] Labrador, V. and Pastrana, S. (2022). Examining the trends and operations of modern dark-web marketplaces. In *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, Genoa, Italy: IEEE, pp. 163–172.

[93] Lacson, W. and Jones, B. (2016). The 21st century darknet market: lessons from the fall of silk road. *International Journal of Cyber Criminology*, 10(1), p. 40.

[94] LeGrand, B., Guillaume, J., Latapy, M. and Magnien, C. (2009). Technical report on dynamics of paedophile keywords in edonkey queries. measurement and analysis of p2p activity against paedophile content project.

[95] Li, W. and Chen, H. (2014). Identifying top sellers in underground economy using deep learning-based sentiment analysis. In *2014 IEEE joint intelligence and security informatics conference*, IEEE, pp. 64–67.

[96] Liggett, R., Lee, J. R., Roddy, A. L. and Wallin, M. A. (2020). The dark web as a platform for crime: an exploration of illicit drug, firearm, csam,

and cybercrime markets. *The Palgrave handbook of international cybercrime and cyberdeviance*, pp. 91–116.

[97] Magán-Carrión, R., Abellán-Galera, A., Maciá-Fernández, G. and García-Teodoro, P. (2021). Unveiling the i2p web structure: a connectivity analysis. *Computer Networks*, 194, p. 108158.

[98] Manky, D. (2013). Cybercrime as a service: a very modern business. *Computer Fraud & Security*, 2013(6), pp. 9–13.

[99] Martin, J., Munksgaard, R., Coomber, R., Demant, J. and Barratt, M. J. (2020). Selling drugs on darkweb cryptomarkets: differentiated pathways, risks and rewards. *The British Journal of Criminology*, 60(3), pp. 559–578.

[100] Meland, P. H., Bayoumy, Y. F. F. and Sindre, G. (2020). The ransomware-as-a-service economy within the darknet. *Computers & Security*, 92, p. 101762.

[101] Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. and Group, T. P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The prisma statement. *PLOS Medicine*, 6(7), pp. 1–6.

[102] Moore, A. (2024). Man charged over €6.5m cryptocurrency seizure in dublin. `https://www.bbc.co.uk/news/articles/cql3vx15e2wo`.

[103] Nadini, M., Bracci, A., ElBahrawy, A., Gradwell, P., Teytelboym, A. and Baronchelli, A. (2022). Emergence and structure of decentralised trade networks around dark web marketplaces. *Scientific reports*, 12(1), p. 5425.

[104] National Academies of Sciences, Division of Behavioral and Social Sciences, Board on Behavioral and Sensory Sciences and Committee on a Decadal Survey of Social and Behavioral Sciences for Applications to National Security

(2019). A decadal survey of the social and behavioral sciences: A research agenda for advancing intelligence analysis. *National Academies Press*.

[105] National Crime Agency (2022). National crime agency annual report and accounts 2021-2022. https://www.nationalcrimeagency.gov.uk/who-we-are/publications/606-national-crime-agency-annual-report-2021-2022/file.

[106] Noroozian, A., Koenders, J., Van Veldhuizen, E., Ganan, C. H., Alrwais, S., McCoy, D. and Van Eeten, M. (2019). Platforms in everything: Analyzing ground-truth data on the anatomy and economics of bullet-proof hosting. In *28th USENIX Security Symposium (USENIX Security 19)*, pp. 1341–1356.

[107] Ouellet, M., Maimon, D., Howell, J. C. and Wu, Y. (2022). The network of online stolen data markets: How vendor flows connect digital marketplaces. *The British Journal of Criminology*, 62(6), pp. 1518–1536.

[108] Pastrana, S., Thomas, D. R., Hutchings, A. and Clayton, R. (2018). Crimebb: Enabling cybercrime research on underground forums at scale. In *Proceedings of the 2018 World Wide Web Conference*, Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, WWW '18, pp. 1845—-1854.

[109] Pastrana, S., Hutchings, A., Thomas, D. and Tapiador, J. (2019). Measuring ewhoring. In *Proceedings of the Internet Measurement Conference*, pp. 463–477.

[110] Peersman, C., Pencheva, D. and Rashid, A. (2021). Tokyo, denver, helsinki, lisbon or the professor? a framework for understanding cybercriminal roles in darknet markets. In *2021 APWG Symposium on Electronic Crime Research (eCrime)*, IEEE, pp. 1–12.

[111] Pete, I., Hughes, J., Caines, A., Vu, A. V., Gupta, H., Hutchings, A., Anderson, R. and Buttery, P. (2022). Postcog: A tool for interdisciplinary research into underground forums at scale. In *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, IEEE, pp. 93–104.

[112] Prichard, J., Watters, P. A. and Spiranovic, C. (2011). Internet subcultures and pathways to the use of child pornography. *Computer Law & Security Review*, 27(6), pp. 585–600.

[113] Prichard, J., Spiranovic, C., Watters, P. and Lueg, C. (2013). Young people, child pornography, and subcultural norms on the internet. *J of the American Society for Information Science and Technology*, 64(5), pp. 992–1000.

[114] Raman, R., Nair, V. K., Nedungadi, P., Ray, I. and Achuthan, K. (2023). Darkweb research: Past, present, and future trends and mapping to sustainable development goals. *Heliyon*, 9(11).

[115] Redman, J. (2020). After empire's exit scam, darknet market patrons scramble to find alternatives. *bitcoincom*.

[116] Reed, M., Syverson, P. and Goldschlag, D. (1998). Anonymous connections and onion routing. *IEEE Journal on Selected Areas in Communications*, 16(4), pp. 482–494.

[117] Remy, C., Rym, B. and Matthieu, L. (2017). Tracking bitcoin users activity using community detection on a network of weak signals. In *Int'l Conference on Complex Networks and Their Applications*, Springer, pp. 166–177.

[118] Reynolds, P. and Irwin, A. S. (2017). Tracking digital footprints: anonymity within the bitcoin system. *Journal of Money Laundering Control*.

[119] Rhumorbarbe, D., Staehli, L., Broséus, J., Rossy, Q. and Esseiva, P. (2016). Buying drugs on a darknet market: A better deal? studying the online

illicit drug market through the analysis of digital, physical and chemical data. *Forensic science international*, 267, pp. 173–182.

[120] Ryan, N., Persi Paoli, G., Aldridge, J. and Warnes, R. (2017). Behind the curtain: The illicit trade of firearms, explosives and ammunition on the dark web. *Rand Corporation*.

[121] Samtani, S., Chinn, R. and Chen, H. (2015). Exploring hacker assets in underground forums. In *2015 IEEE international conference on intelligence and security informatics (ISI)*, IEEE, pp. 31–36.

[122] Software Freedom Conservancy (2022). Selenium project. `https://www.selenium.dev/`.

[123] Soska, K. and Christin, N. (2015). Measuring the longitudinal evolution of the online anonymous marketplace ecosystem. In *24th USENIX security symposium (USENIX security 15)*, pp. 33–48.

[124] Steel, C. M. (2009). Child pornography in peer-to-peer networks. *Child Abuse & Neglect*, 33(8), pp. 560–568.

[125] Steel, C. M. (2009). Web-based child pornography: Quantification and qualification of demand. *International Journal of Digital Crime and Forensics (IJDCF)*, 1(4), pp. 58–69.

[126] Stoltenborgh, M., Bakermans-Kranenburg, M. J., Alink, L. R. and van IJzendoorn, M. H. (2015). The prevalence of child maltreatment across the globe: Review of a series of meta-analyses. *Child Abuse Review*, 24(1), pp. 37–50.

[127] Tidy, J. (2022). Hydra: How german police dismantled russian darknet site. `https://www.bbc.co.uk/news/technology-61002904`.

[128] Tor Project (2022). Network ddos — tor project status. https://status.torproject.org/issues/2022-06-09-network-ddos/.

[129] Tor Project (2024). Servers. https://metrics.torproject.org/networksize.html.

[130] Tor Project (2024). Types of relays on the tor network. https://community.torproject.org/relay/types-of-relays/.

[131] Turk, K., Pastrana, S. and Collier, B. (2020). A tight scrape: methodological approaches to cybercrime research data collection in adversarial environments. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, Genoa, Italy: IEEE, pp. 428–437.

[132] Ursani, Z., Peersman, C., Edwards, M., Chen, C. and Rashid, A. (2021). The impact of adverse events in darknet markets: an anomaly detection approach. In *2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pp. 227–238.

[133] U.S. Attorney's Office (2015). Ross ulbricht, the creator and owner of the silk road website, found guilty in manhattan federal court on all counts. https://www.fbi.gov/contact-us/field-offices/newyork/news/press-releases/ross-ulbricht-the-creator-and-owner-of-the-silk-road-website-found-guilty-in-manhattan-federal-court-on-all-counts.

[134] U.S. Attorney's Office (2024). "incognito market"' owner arrested for operating one of the largest illegal narcotics marketplaces on the internet. https://www.justice.gov/usao-sdny/pr/incognito-market-owner-arrested-operating-one-largest-illegal-narcotics-marketplaces.

[135] Van Buskirk, J., Roxburgh, A., Farrell, M. and Burns, L. (2014). The closure of the silk road: what has this meant for online drug trading?

[136] Van Wegberg, R., Tajalizadehkhoob, S., Soska, K., Akyazi, U., Gañán, C., Klievink, B., Christin, N. and Van Eeten, M. (2018). Plug and prey? measuring the commoditization of cybercrime via online anonymous markets. In *Proceedings of the 27th USENIX Conference on Security Symposium*, USA: USENIX Association, SEC'18, pp. 1009–1026.

[137] Vargas, V.-M. (2019). The new economic good: Your own personal data. an integrative analysis of the dark web. In *Proceedings of the International Conference on Business Excellence*, vol. 13, pp. 1216–1226.

[138] Vehovar, V., Ziberna, A., Kovacic, M., Mrvar, A. and Dousak, M. (2009). An empirical investigation of paedophile keywords in edonkey p2p network. Tech. rep., Citeseer.

[139] Vu, A. V., Hughes, J., Pete, I., Collier, B., Chua, Y. T., Shumailov, I. and Hutchings, A. (2020). Turning up the dial: The evolution of a cybercrime market through set-up, stable, and covid-19 eras. In *Proceedings of the ACM Internet Measurement Conference*, New York, NY, USA: Association for Computing Machinery, IMC '20, pp. 551—-566.

[140] Wall, D. (2001). *Crime and the Internet*. Routledge London.

[141] WanFang Data (2023). China online journals (coj). `https://www.wanfangdata.com.cn/`.

[142] Wang, Y., Arief, B. and Hernandez-Castro, J. (2021). Toad in the hole or mapo tofu? comparative analysis of english and chinese darknet markets. In *2021 APWG Symposium on Electronic Crime Research (eCrime)*, IEEE, Boston, MA, USA: IEEE, pp. 1–13.

[143] Wang, Y., Arief, B. and Hernandez-Castro, J. (2023). Dark ending: What happens when a dark web market closes down. In *Proceedings of the 9th*

International Conference on Information Systems Security and Privacy - ICISSP, INSTICC, Lisbon, Portugal: SciTePress, pp. 106–117.

[144] Wang, Y., Arief, B. and Hernandez-Castro, J. (2024). Analysis of security mechanisms of dark web markets. In *Proceedings of the 2024 European Interdisciplinary Cybersecurity Conference*, New York, NY, USA: Association for Computing Machinery, EICC '24, pp. 120—-127.

[145] Wang, Y., Arief, B. and Hernandez-Castro, J. (2025). Secure in the dark? an in-depth analysis of dark web markets security. *International Journal of Information Security*, 24(3), pp. 1–15.

[146] Wang, Y., Arief, B., Franqueira, V. N. L., Coates, A. G. and Ó Ciardha, C. (2023). Investigating the availability of child sexual abuse materials in dark web markets: Evidence gathered and lessons learned. In *Proceedings of the 2023 European Interdisciplinary Cybersecurity Conference*, New York, NY, USA: Association for Computing Machinery, EICC '23, pp. 59–64.

[147] Wang, Y., Roscoe, S., Arief, B., Connolly, L., Borrion, H. and Kaddoura, S. (2023). The social and technological incentives for cybercriminals to engage in ransomware activities. In *Security and Privacy in Social Networks and Big Data*, Singapore: Springer Nature Singapore, pp. 149–163.

[148] Weber, J. and Kruisbergen, E. W. (2019). Criminal markets: the dark web, money laundering and counterstrategies-an overview of the 10th research conference on organized crime. *Trends in Organized Crime*, 22(3), pp. 346–356.

[149] Westlake, B., Bouchard, M. and Frank, R. (2017). Assessing the validity of automated webcrawlers as data collection tools to investigate online child sexual exploitation. *Sexual Abuse*, 29(7), pp. 685–708.

[150] Westlake, B. G. and Bouchard, M. (2016). Liking and hyperlinking: Community detection in online child sexual exploitation networks. *Social science research*, 59, pp. 23–36.

[151] Westlake, B. G., Bouchard, M. and Girodat, A. (2017). How obvious is it? the content of child sexual exploitation websites. *Deviant behavior*, 38(3), pp. 282–293.

[152] WIRED (2021). The demise of white house market will shake up the dark web. *WIRED*.

[153] Woodhams, J., Kloess, J. A., Jose, B. and Hamilton-Giachritsis, C. E. (2021). Characteristics and behaviors of anonymous users of dark web platforms suspected of child sexual offenses. *Frontiers in Psychology*, 12, pp. 1–11.

[154] Wu, L., Hu, Y., Zhou, Y., Wang, H., Luo, X., Wang, Z., Zhang, F. and Ren, K. (2021). Towards understanding and demystifying bitcoin mixing services. In *Proceedings of the Web Conference 2021*, New York, NY, USA: Association for Computing Machinery, WWW '21, p. 33–44.

[155] Yannikos, Y. and Heeger, J. (2024). Captchas on darknet marketplaces: Overview and automated. *Electronic Imaging*, 36, pp. 1–6.

[156] Yannikos, Y., Heeger, J. and Steinebach, M. (2022). Data acquisition on a large darknet marketplace. In *Proceedings of the 17th International Conference on Availability, Reliability and Security*, New York, NY, USA: ACM, ARES '22, pp. 1–6.

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

[158] Yoon, C., Kim, K., Kim, Y., Shin, S. and Son, S. (2019). Doppelgängers on the dark web: A large-scale assessment on phishing hidden web services. In *The World Wide Web Conference*, New York, NY, USA: Association for Computing Machinery, WWW '19, p. 2225–2235.

[159] Zantout, B., Haraty, R. et al. (2011). I2P data communication system. In *Proceedings of ICN*, Citeseer, pp. 401–409.

[160] Zhang, H. and Zou, F. (2020). A survey of the dark web and dark market research. In *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, pp. 1694–1705.

[161] Zhou, G. and Zhuge, J. (2020). Adapting to local conditions: Similarities and differences in anonymous online market between chinese and english speaking communities. In *International Conference on Digital Forensics and Cyber Crime*, Springer, pp. 164–181.

[162] Zhou, G., Zhuge, J., Fan, Y., Du, K. and Lu, S. (2020). A market in dream: the rapid development of anonymous cybercrime. *Mobile Networks and Applications*, 25, pp. 259–270.

[163] Zhuge, J., Holz, T., Song, C., Guo, J., Han, X. and Zou, W. (2009). Studying malicious websites and the underground economy on the chinese web. *Managing Information Risk and the Economics of Security*, pp. 225–244.