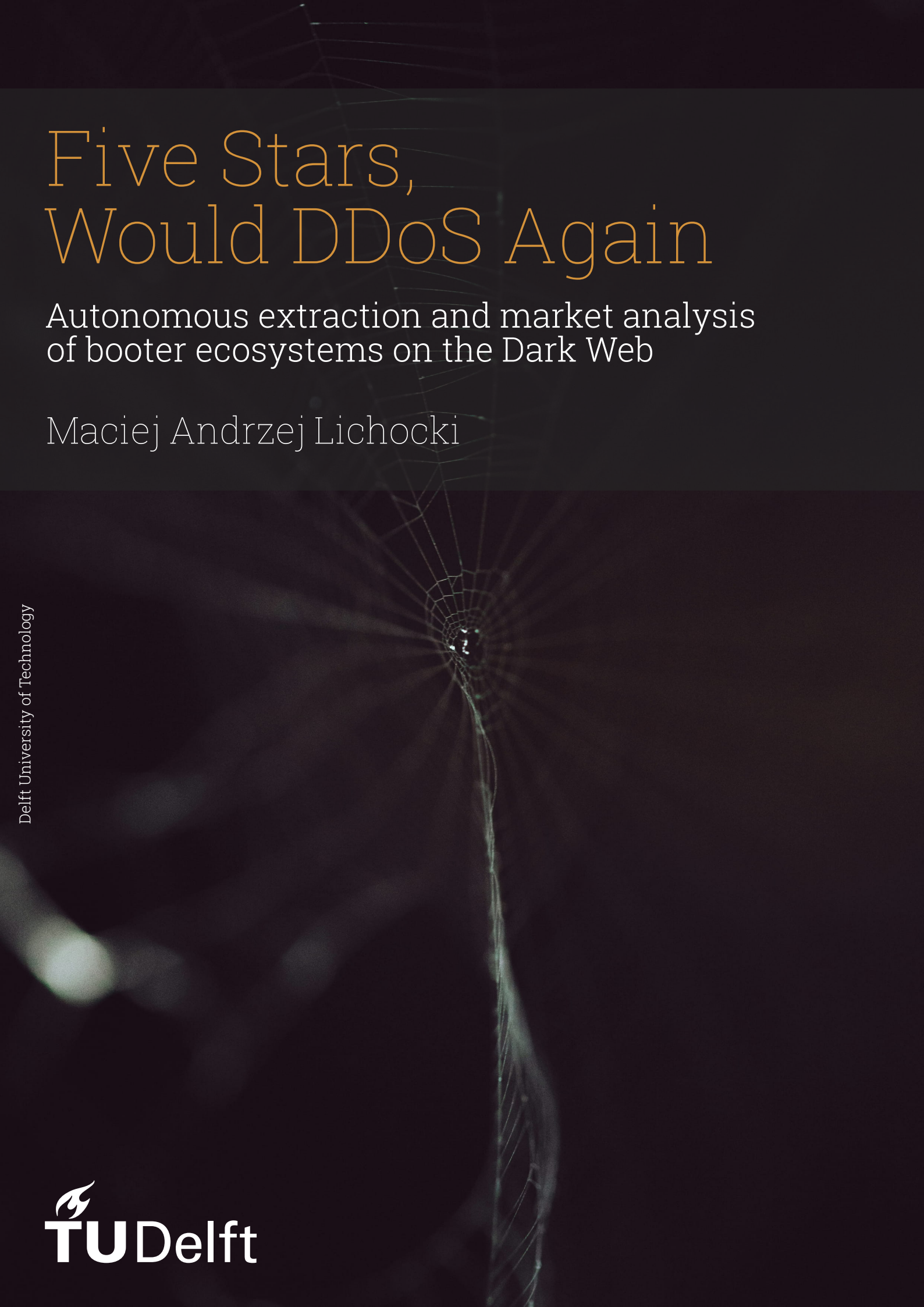
Five Stars, Would DDoS Again

Autonomous extraction and market analysis of booter ecosystems on the Dark Web

Maciej Andrzej Lichocki



Delft University of Technology

TU Delft

Five Stars, Would DDoS Again

Autonomous extraction and market analysis of booter ecosystems on the Dark Web

by

Maciej Andrzej Lichocki

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on 16 June 2026 at 11:30.

Student number:	5488400	
Project duration:	November 2025 – June 2026	
Thesis committee:	Harm Griffioen	TU Delft, Advisor
	Maarten Weyns	TU Delft, Daily Supervisor
	Johan Pouwelse	TU Delft

Style: TU Delft Report Style, with modifications by Daan Zwaneveld

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

TU Delft

Abstract

This thesis investigates the observable structure of dark web DDoS-for-hire ecosystems to support operational disruption planning and presents an automated CTI pipeline for repeatable measurement. The pipeline integrates seeded discovery, Tor-based crawling, harmful-content filtering, deterministic indicator extraction, and LLM-assisted structuring. In the fixed reporting snapshot, 389,006 raw pages were reduced to 304,949 sentry-passed pages, 3,919 extracted pages, and 301 DDoS-classified URLs across 59 domains. Linkage analysis in this dataset found fragmented but actionable cross-domain signals, while market profiling showed sparse pricing and attack-vector fields. These results support triage and monitoring, rather than full attribution or complete market quantification. The main contribution is a transparent and reproducible measurement workflow that supports independent replication and keeps evidence interpretation within explicit confidence boundaries.

Contents

Nomenclature

API Application Programming Interface. 3

CSAM Child Sexual Abuse Material. 6
CTI Cyber Threat Intelligence. i, ii, 1–4, 6, 26, 28

DDoS Distributed Denial of Service. i, ii, 1–4, 26, 28

HTTP Hypertext Transfer Protocol. 3–5

IOC Indicator of Compromise. 2, 4, 26

JSON JavaScript Object Notation. 6

LLM Large Language Model. i, 1, 2, 5, 6, 26

NLP Natural Language Processing. 5, 6

Tor The Onion Router. i, ii, 1, 2, 4

UDP User Datagram Protocol. 3
URL Uniform Resource Locator. i, 1, 28

1

Introduction

As society becomes increasingly reliant on connected digital systems, the threat surface for cyberattacks has expanded significantly. Among these threats, DDoS attacks, which overwhelm target infrastructure with malicious traffic until it becomes unavailable, remain a persistent large-scale threat in recent operational reporting [1, 2]. Because digital availability underpins public services, commerce, and critical operations, understanding how attack capacity is supplied has become essential for prevention, disruption, and policy response.

Today, launching a large-scale, disruptive attack no longer requires advanced technical skills. Cybercriminals have commoditised attack capabilities as an affordable service, commonly marketed as "booters" or "stressers." While many of these services operate openly on the clear web [3, 4], recent measurements also report DDoS-for-hire activity on dark web hidden-service infrastructure [5]. Dark web DDoS-for-hire operators work through hidden `.onion` domains within the Tor network, forming a fragmented and evasive landscape. Operators advertise their capabilities and sell attack services across scattered underground platforms, making it difficult to observe the market as a whole.

For cybersecurity researchers and law enforcement, mapping this shadow economy is critical for prioritising interventions: who runs these platforms, their pricing structures, and what kinds of attacks they sell. However, manual review at this scale is resource-intensive and can increase analyst exposure to illegal or harmful material [6]. Generic web crawlers are usually not configured for robust Tor collection, and adapting Tor crawling to indicator-focused CTI workflows across unstable hidden-service ecosystems remains non-trivial [7]. Prior work on DDoS-for-hire services has largely addressed clear web platforms [8, 4, 3], where page content and domain records are comparatively accessible. The dark web segment remains poorly characterised.

This thesis aims to fill that gap. The central goal is to produce evidence-based answers about the dark web DDoS-for-hire ecosystem: how many active platforms exist, their pricing structures, what attack capabilities they offer, and whether independent storefronts can be linked to common operators. The primary scientific contribution is the market analysis itself. To gather the data needed at scale, this thesis develops an automated CTI pipeline as a data-collection instrument. The pipeline applies risk-reduction controls, including harmful-content filtering before persistence, and uses Large Language Model (LLM)-assisted extraction to convert unstructured text into structured pricing and capability indicators.

Compared with prior work, the novelty is integration rather than a single new detection model. This study contributes three methodological elements in one auditable workflow: stage-wise contraction reporting from raw intake to DDoS-classified records under fixed query logic, explicit denominator separation at Uniform Resource Locator (URL), domain, and extracted-row levels, and mirror-aware linkage interpretation that separates screening value from attribution strength. Together, these elements support transparent comparison and bounded interpretation for the observable subset.

The thesis-level research question is:

> *How can the observable dark web DDoS-for-hire ecosystem be measured in a repeatable and auditable way, and what can be inferred from that measurement about market surface, cross-domain linkage, and advertised service characteristics?*

This question is addressed in the embedded article through three research questions: **RQ1** on detection and classification (stage-wise contraction from raw intake to DDoS-classified records), **RQ2** on ecosystem linkage (repeated technical indicators and descriptive identifiers across distinct domains), and **RQ3** on market characteristics (pricing models, advertised Layer 4/7 capabilities, and service-availability signals). Within this workflow, sentry scoring is used as a weighted admission gate that combines threat-relevant keyword classes with hard-indicator bonuses to prioritise candidate pages for extraction.

1.1. Contributions

This thesis makes the following contributions:

- Provides empirical answers to the thesis research questions by characterising the observable structure of the dark web DDoS-for-hire market, including active platform prevalence, pricing norms, and advertised Layer 4 and Layer 7 capability distribution.
- Analyses potential relationships between storefronts through fuzzy hashing and shared technical indicators, delivering a mirror-aware assessment with explicit reporting when evidence is insufficient for strong operator-linkage claims.
- Constructs a structured dataset of active dark web booter services, capturing pricing tiers, advertised attack capabilities, and operator contact handles to support reproducible market analysis.
- Designs and implements an automated dark web CTI pipeline as a purpose-built measurement instrument, integrating a seeder, crawler, extractor, LLM parser, screenshot worker, and a human-in-the-loop review dashboard. The pipeline routes all traffic through Tor and handles anti-bot protections via a headless browser fallback.
- Devises a hazmat filtering mechanism that detects and discards content matching potentially illegal or harmful indicators before storage or analyst presentation, reducing exposure risk, enabling safer and more ethical dark web investigation.

1.2. Outline of the Thesis

The remainder of this thesis is structured as follows. Chapter 2 provides a focused background overview covering the concepts needed to understand the pipeline design and research analysis: DDoS attacks and the booter service model, the Tor network, CTI and Indicator of Compromises (IOCs), dark web crawling, anti-bot protections, fuzzy hashing, LLMs, information extraction from unstructured text, and the ethical considerations and hazmat filtering that govern data collection.

Chapter 3 presents the article that forms the core contribution of this thesis. It details the pipeline architecture and methodology, reports the results of the dark web DDoS-for-hire market analysis, and discusses the findings, limitations, and directions for future work.

Chapter 4 reflects on the contribution to CTI research and practice and synthesises broader implications of the work. Chapter 5 summarises the empirical findings, revisits the thesis-level research question, and outlines directions for future work.

2

Background

This chapter introduces the core concepts needed to understand the automated CTI pipeline. It breaks down how denial-of-service attacks work, explores the dark web infrastructure where they are sold, and outlines the techniques used to extract structured intelligence from these hidden environments.

2.1. Distributed Denial of Service attacks

A DDoS attack is intended to make a server, service, or network unavailable by saturating it with hostile traffic. Attackers coordinate multiple compromised systems to send concurrent traffic volumes that exceed the target's bandwidth or processing limits, which blocks legitimate access and can cause significant operational and financial harm to affected organisations [9].

Recent vendor and operational reporting describes evolving DDoS attack vectors, including amplification and botnet-based traffic [1]. Crucially, orchestrating such high-volume disruptions no longer requires advanced technical skills or custom infrastructure. Not only are the underlying botnets leased out on underground markets, but fully managed DDoS-for-hire services are also readily available, lowering the barrier to entry and transforming network disruption into a highly accessible commercial enterprise [3, 10].

These operations are generally categorised by the network layer they exploit. Layer 4 attacks target the transport layer by flooding the victim with raw data, such as User Datagram Protocol (UDP) amplification traffic, to saturate network bandwidth. In contrast, Layer 7 attacks target the application layer by simulating resource-intensive client behaviours, such as initiating continuous Hypertext Transfer Protocol (HTTP) GET requests. While Layer 4 attacks rely on massive bandwidth to overwhelm network infrastructure, Layer 7 attacks are often stealthier and require fewer resources from the attacker to exhaust the memory and processing power of the victim's server [9, 1, 11].

2.2. Booters and Stressers (DDoS-as-a-Service)

As DDoS attacks evolved into a profitable business, criminals developed web platforms commonly known as booters or stressers. These websites act as DDoS-as-a-Service storefronts, hiding the technical complexity of botnet management behind easy-to-use web dashboards and Application Programming Interface (API) access [3]. These interfaces lower the expertise barrier, allowing users with limited technical skill to initiate disruptive attacks through simplified web controls [3, 8, 12]. To avoid immediate legal consequences, operators frequently market these services as legitimate "server stressers" meant for network administrators to test their own defences [8]. However, their design and features clearly target malicious use against third parties, such as gaming servers or competing businesses [12]. Behind the scenes, the administrators must continuously scan the internet for vulnerable devices to maintain their attack capacity and update their web infrastructure to keep customers satisfied [13].

Customers pay for these attacks through subscription models that closely resemble legitimate software services. Users purchase plans that limit the maximum duration of an attack, the total network volume,

and how many attacks can run at the same time [14]. Reported pricing can be low enough to make these services accessible to low-skill users, in some cases for only a few dollars [14]. Payment handling is operationally central, and case-study evidence shows that interventions targeting payment channels can disrupt specific providers [15]. Because of the severe damage these platforms cause, global law enforcement agencies have conducted coordinated takedown operations to seize their domains [4]. Despite these efforts, recent measurements still report related DDoS-for-hire activity on hidden services [5].

2.3. The Tor Network and Hidden Services

A practical way to frame the web is as clear web content indexed by common search engines, deep web content that is typically unindexed but often legitimate behind forms or logins, and dark web content that is deliberately concealed within anonymity networks and accessed with specialised software [7, 16, 17]. While estimates differ, review literature generally describes most deep web content as outside conventional indexing and the dark web as a comparatively smaller subset of that broader space [7, 17].

In this thesis, the relevant anonymity network is Tor. According to Tor project documentation, hidden services use `.onion` addresses resolved within the overlay via onion-service descriptors and rendezvous circuits, rather than the public naming infrastructure used by ordinary HTTP websites [18]. These properties reduce the direct exposure of server location and complicate straightforward internet mapping. For booter operators, this provides resilience after disruption: infrastructure can be moved, mirrored, and re-announced with relatively low overhead. Measurement studies show that, as hidden services appear and disappear, the observable population is inherently unstable, so longitudinal analysis requires repeated discovery and recrawl rather than a single snapshot [5, 7].

2.4. Cyber Threat Intelligence (CTI)

CTI is the process of converting raw threat observations into evidence that supports defensive decisions. Standards guidance treats CTI as processed threat data with enough context to guide operational and strategic security decisions [19, 20, 21]. In practice, this means turning noisy forum and marketplace material into structured knowledge about DDoS-for-hire activity.

At ecosystem scale, this transformation depends on consistent collection and curation rather than one-off inspection of individual pages. It includes tracking who advertises what, where services are promoted, and how offers evolve across time and channels [22, 13, 17]. A key analytical unit is the IOC: a technical or behavioural signal that supports tracking of actors, infrastructure, or campaigns. In booter analysis, examples include cryptocurrency wallets, Telegram handles, service names, and recurring payment instructions. Individual indicators are often ambiguous, but combinations across observations can reveal operational reuse, mirror relationships, and can suggest possible rebranding hypotheses [12, 15, 21].

2.5. Automated Dark Web Crawling

Automated dark web crawling refers to the use of software agents to discover and download content from services that are only reachable through anonymity networks such as Tor. Compared to clear web crawling, these targets are slower to respond, often temporarily offline, and more unstable over time: hidden services appear, disappear, and reappear under new addresses, and many advertised links are short-lived or intentionally misleading [23, 22, 7]. Discovery is further complicated by the lack of a comprehensive search index for `.onion` domains, so crawlers must rely on seed lists, hyperlink traversal, and community directories rather than exhaustive enumeration [7, 24].

For measurement and threat-intelligence purposes, this environment changes the emphasis from visiting as many pages as possible once to maintaining a reasonably stable view over time. Crawling is better understood as an ongoing sampling process in a volatile network. Repeated visits are needed to track content evolution, failures and route changes are expected rather than exceptional, and near-duplicate content must be controlled so that retries, mirrors, and minor edits do not dominate the collected dataset [23, 22, 24]. These characteristics form the backdrop for the pipeline design described later in this thesis.

2.6. Anti-Bot Protections and Access Queues

Across web services, anti-automation controls are commonly used to preserve availability and limit abuse. Industry descriptions of bot management emphasise layered defences that include challenge pages, script-based checks, browser fingerprinting, rate limiting, and queue or waiting-room mechanisms that throttle access when load or perceived attack risk is high [25, 26]. When comparable controls are encountered on dark web services, they can affect crawler visibility and continuity in practice [24].

For automated measurement, these protections introduce practical observability constraints beyond basic network reachability. Simple HTTP clients that do not execute JavaScript, handle cookies robustly, or respect queue tokens may be repeatedly challenged or diverted, which can under-represent protected paths in collected data. Conversely, more capable automated clients incur higher resource costs and can still be selectively throttled or blocked. As a result, any longitudinal study of dark web services must treat anti-bot controls as part of the measurement environment, recognising that what is observable is shaped by both crawler behaviour and access controls on target services.

2.7. Fuzzy Hashing and Data Deduplication

Booter-related pages are frequently mirrored, cloned, or incrementally updated, which creates substantial near-duplicate text and imagery. Studies of web and dark web content show that a non-trivial fraction of pages or marketplace listings are near-identical copies, differing only in minor layout, boilerplate, or media changes [27]. Exact deduplication with cryptographic digests is still useful for byte-identical content, but it cannot recognise similar pages that differ by minor edits, reordered blocks, or small markup changes.

Fuzzy hashing addresses this gap by scoring similarity rather than strict equality. SimHash, for example, maps textual features into compact fingerprints where near-duplicates have small Hamming distance, a technique that has been widely applied in large-scale web crawling and deduplication [28, 29]. For a forum thread or marketplace listing scraped repeatedly over time, this allows incremental processing policies: if a page version is very similar to the previous one, small edits such as a single new comment can be recorded without triggering full downstream reprocessing, while larger changes that push similarity below a threshold are treated as meaningful updates [28, 29, 27].

2.8. Large Language Models

Modern LLMs are general-purpose language models built on the transformer architecture [30, 31]. They are pre-trained to predict the next token from prior context using large text datasets and can be steered through prompting to support generation, classification, transformation, and extraction tasks in a single model family [31, 32]. In threat-intelligence workflows, this makes them useful for reading advertisements or forum posts and extracting fields such as service names, prices, and contact identifiers.

In extraction pipelines, LLMs can act as flexible converters from semi-structured text to structured records [33, 34]. Their limits are also important: each request has a finite context window, and longer inputs increase computational overhead and can increase latency [32, 35]. Practical systems therefore select relevant passages, chunk long pages, and validate outputs before storage.

2.9. Information Extraction from Unstructured Text

Classical extraction based on regular expressions or fixed parsers works best when syntax is stable and document templates change slowly. Dark web advertisements and forum posts, however, vary in wording, order, and formatting, so rigid rules degrade quickly as templates drift. Within Natural Language Processing (NLP), this is usually handled with models that learn meaning from context rather than exact word matches, for example in named-entity recognition and relation extraction [33]. These approaches remain a good conceptual fit for the task, but require annotated data and dedicated model training that are beyond the scope of this thesis.

Recent empirical work shows that general-purpose LLMs can perform useful information extraction in a zero- or few-shot setting, given carefully designed prompts and schemas [33, 34]. To make this usable in analytics pipelines, outputs must be constrained to machine-checkable structure. A common pattern

is schema-guided prompting with strict validation of required keys and value types in JavaScript Object Notation (JSON), followed by rejection and retry when constraints are violated. Studies of structured extraction workflows describe consistency-focused patterns under schema-constrained prompting and validation in their evaluated settings [36, 33]. This thesis adopts that family of approaches, focusing on schema-constrained extraction rather than on training new task-specific NLP models.

This is a deliberate trade-off. Compared with task-specific supervised models, schema-constrained LLM extraction in this thesis improves adaptability to drifting templates but can reduce field-level precision and recall under sparse or ambiguous text [33, 34].

2.10. Ethical Considerations and Content Safety Filtering

Dark web collection introduces legal and ethical risk because automated systems can retrieve harmful or illegal material without warning. In academic settings, this risk extends beyond individual researchers to institutional compliance, infrastructure handling procedures, and data-governance obligations [6, 37, 17]. A defensible CTI workflow therefore needs clear collection, storage, and disposal boundaries. Cybercrime research guidelines emphasise proportionality, data minimisation, and the need to avoid unnecessary retention of material that may be illegal to possess [6, 37].

Content safety filtering operationalises these boundaries by screening for prohibited categories, including potential Child Sexual Abuse Material (CSAM) and extreme-violence indicators, before persistence and downstream analysis. In child-safety practice, some resources are controlled-access, such as Internet Watch Foundation hash and keyword services. Others are openly published to support detection workflows, including Thorn's CSAM Keyword Hub [38, 39, 40]. For controlled-access child-safety resources, providers describe vetting as a safeguard against misuse. As a complement, broader open-source profanity and toxicity lists, including the LDNOOBW "List of Dirty, Naughty, Obscene and Otherwise Bad Words", were used only for coarse lexical triage of general abuse content when controlled resources were unavailable or when extra lexical coverage was needed. These lists were not used for CSAM detection [41, 42].

The goal of these controls is risk reduction and scope control, not perfect semantic classification. In line with existing guidance, content that triggers content safety filters is typically dropped or, where governance allows, escalated for tightly controlled review rather than retained by default [6, 37]. This reduces the likelihood of storing illegal material while still enabling lawful, aggregate analysis of cybercrime ecosystems.

3

Scientific Article

Autonomous extraction and market analysis of booter ecosystems on the dark web

Maciej Andrzej Lichocki
Delft University of Technology, Faculty of Electrical Engineering, Mathematics and Computer Science

Measurement of denial-of-service-for-hire activity on hidden services is still fragmented, which makes evidence-based disruption planning difficult. This article analyses dark web booter ecosystems through three observable dimensions: market-surface contraction, cross-domain linkage signals, and advertised service characteristics. A repeatable measurement pipeline is used, with discovery, Tor-based crawling, sentry admission filtering, deterministic indicator extraction, and language-model-assisted structuring. Sentry scoring combines weighted Critical Threat, High Interest, and Commerce keywords with wallet and public-IPv4 bonuses, and pages scoring at least 50 are marked as sentry-passed. In the fixed reporting snapshot, 389,006 raw pages are reduced to 304,949 sentry-passed pages, 3,919 URLs with persisted extraction output, and 301 DDoS-classified URLs. At domain level, the same pipeline resolves to 59 domains with DDoS-classified pages. In the restricted RQ2 group (301 URLs across 59 domains), linkage analysis produced a mixed domain-indicator graph with 30 nodes, 28 edges, and 7 clusters. The largest component contains 23.33% of nodes, indicating a fragmented linkage pattern with limited but actionable screening value. Market output is informative but incomplete: from 681 structured rows, specific attack-vector coverage is 3.96% and price field coverage is 7.64% (52/681). These findings support repeatable screening and coarse market profiling, with interpretation grounded in passive public-surface observation and transparent boundaries around sparse pricing and vector fields.

I. Introduction

In denial-of-service-for-hire markets, operators advertise attack capability as a purchasable service, which lowers the practical barrier to launching attacks [1–4]. For operational Cyber Threat Intelligence (CTI), empirical market observation can contribute to situational awareness and analytic prioritisation [5]. Clear web booter and stresser ecosystems have been studied for more than a decade [1–4]. Several studies in the reviewed source set focus on broader dark web collection problems rather than booter-specific hidden-service measurement [6–8].

The main gap is twofold: limited dark web booter measurements and limited integrated evidence pipelines that make each filtering and extraction boundary clear. Existing strands often evaluate one part of the problem at a time, for example market measurement on clear web surfaces, broad dark web collection without booter-focused interpretation, or extraction performance without ecosystem-level analysis [2, 6, 9]. As a result, it is difficult to compare how much observable signal survives each stage from intake to final Distributed Denial of Service (DDoS)-specific evidence.

This article examines the observable structure of dark web booter ecosystems in one reproducible reporting snapshot. Prior dark web CTI studies motivate broad collection under partial observability [5, 7, 8]. This article explicitly defines the filtering and extraction boundaries used within that broader collection context. To operationalise this objective, the automated pipeline is treated as a measurement instrument that converts passive hidden-service collection into structured records.

This objective is addressed through three research questions:

1) **RQ1:** quantify observable market surface through selectivity from raw intake to DDoS-classified records,
2) **RQ2:** assess cross-domain linkage signals through repeated technical indicators and descriptive identifiers,
3) **RQ3:** characterise observable market signals on pricing, advertised attack capabilities (Layer 4/7), and service availability.

This article combines these questions within a single traceable evidence pipeline. For RQ1, clear stage contraction is reported at Uniform Resource Locator (URL) and domain level under fixed query logic. For RQ2, linkage concentration is quantified and screening value is separated from attribution strength. For RQ3, market-facing signals are reported together with field-coverage limits that constrain interpretation. The pipeline produces repeatable evidence, and the article provides an empirical description of the resulting dark web booter snapshot with clear interpretation limits.

II. Related Work

A. Booter and denial-of-service-for-hire measurements

The literature describes denial-of-service-for-hire as a lasting service market rather than a set of isolated incidents [1–4, 10, 11]. Early work documented core service features and abuse workflows [1], while later studies tracked infrastructure, customer-facing business models, and intervention effects over time [2–4, 10]. Complementary criminological analysis examined provider signalling and service presentation in booter offerings, further supporting a market-oriented interpretation [11]. Taken together, this literature provides an observational baseline for booter ecosystems, largely from clear web surfaces [1–4, 10, 11]. These studies also show that market visibility depends strongly on collection surface and record standardisation. Because much evidence in this strand is drawn from openly indexed services, court-exposed infrastructure, or focused intervention windows, transferability to hidden-service market structure should be treated cautiously [2, 4, 10].

B. Cyber threat intelligence collection on the dark web

Dark web collection research emphasises scalable collection, noise control, and analyst usefulness [5, 7]. Gupta et al. and Shakarian outlined practical workflows for dark web cyber threat intelligence under mixed source conditions [5, 7]. Pastrana et al. showed that forum-scale datasets can reveal ecosystem structure when collection and standardisation are controlled [6]. This operational line builds on earlier hidden-service marketplace measurements, which demonstrated that longitudinal collection remains feasible under platform churn and partial observability [12, 13]. De Pascale et al. proposed CRATOR as a reusable architecture for sustained dark web crawling and monitoring under practical access constraints [8]. Across the reviewed dark web collection literature, scope is often broad, and the cited studies primarily emphasise collection coverage, platform churn, and reusable crawling infrastructure [6–8].

C. Automated extraction and linkage from unstructured sources

Recent work on model-assisted parsing emphasises extraction quality and validation requirements [9]. Schema and interoperability discussions emphasise clear entity-relation representations in standards such as STIX 2.1 and in taxonomy comparisons [14, 15]. NIST guidance complements this with recommendations for CTI sharing and reporting practice [16]. Near-duplicate techniques such as SimHash remain useful for consolidating crawl outputs [17]. Recent large-language-model extraction studies report exploratory flexibility gains on noisy text in evaluated cases, while emphasising clear tracking and validation requirements [9]. Within the studies discussed above, evaluation emphasis is on extraction accuracy and reliability. This article additionally reports how filtering-stage decisions and sparse field completion affect market-level interpretation in hidden-service ecosystems [6–9].

D. Gap addressed by this study

These strands are well developed but are usually evaluated in isolation: booter measurement [1, 2,

4, 11], dark web collection [6–8, 12, 13], and automated extraction [9, 14]. This article addresses the integration challenge by reporting a single end-to-end snapshot where intake contraction, linkage concentration, and market-signal coverage can be interpreted together, with explicit uncertainty limits.

III. Methodology

A. Study design and scope

This study uses passive collection of publicly reachable hidden services. It does not include account creation, purchases, credential use, or direct interaction with operators. The objective is ecosystem description rather than intervention.

Three analytical units are used throughout the manuscript. The first is the URL-level page record, defined as one fetched and normalised page state linked to a canonical URL key and crawl metadata. The second is the domain-level record, defined by onion host and used for ecosystem breadth and linkage aggregation. The third is the extracted-row record produced by `llm_parser`, where one source page can produce multiple structured rows. This separation avoids mixing page-volume claims with listing-volume claims.

The implementation is a research prototype with explicit operational boundaries. Stages and controls are specified in operational terms below so that reported counts and figures can be regenerated from one fixed configuration. One practitioner interview was used only as contextual discussion input and not as a primary data source for RQ metrics.

B. Pipeline implementation

The implementation follows a queue-based modular worker design in Go, with Redis queues linking seeding, crawling, deterministic extraction, and language-model parsing.

Analyst-defined subscribed keywords are stored in Redis and consumed by the seeder service as the canonical input for discovery. The fixed subscribed seed set used for the final reporting run is listed in Appendix C.

Operationally, flow proceeds from `seeder` to `crawler` to `extractor` to `llm_parser`, then to `screenshot capture` and `dashboard Application Programming Interface (API)` services. Each stage consumes from an upstream queue, writes stage outputs into `ClickHouse` as a shared analytical store, and emits tasks for the next stage. URL normalisation and recrawl controls are applied before queue insertion, so repeated links are handled consistently across stages. The crawler also feeds discovered hidden-service URLs back into the crawl queue after normalisation and recrawl checks. This creates a controlled discovery loop rather than a strictly one-pass pipeline. Figure 1 summarises this stage flow.

- **Passive collection scope:** the seeding stage runs periodically, reads subscribed keywords from Redis, and queries predefined dark web search engines keyword by keyword to extract candidate `.onion` links. Recrawling is handled through a minimum revisit interval and a per-URL last-crawl timestamp. A known URL is re-enqueued only after the interval has elapsed, which allows the system to capture newly published posts while avoiding unnecessary repeated fetches within short time windows.
- **The Onion Router (Tor) crawling with dynamic-page fallback:** all hypertext transfer protocol traffic is routed through a SOCKS5 Tor client [18]. The crawler downloads reachable pages, converts hypertext markup language to markdown for normalisation, stores page-level records in `raw`, and self-enqueues newly discovered hidden-service links after normalisation and recrawl checks. For revisited pages, downstream duplicate controls compare content fingerprints, so unchanged content can be suppressed while changed content is retained for extraction. When a page is blocked by waiting-room delay pages, browser-side script challenges, or bot-protection checks, the crawler can enqueue a headless Chromium path. This path waits for rendered content, then returns a normalised page record to the same downstream stages.
- **Deterministic extraction and relevance scoring:** regular-expression extractors identify hard indicators such as cryptocurrency wallet

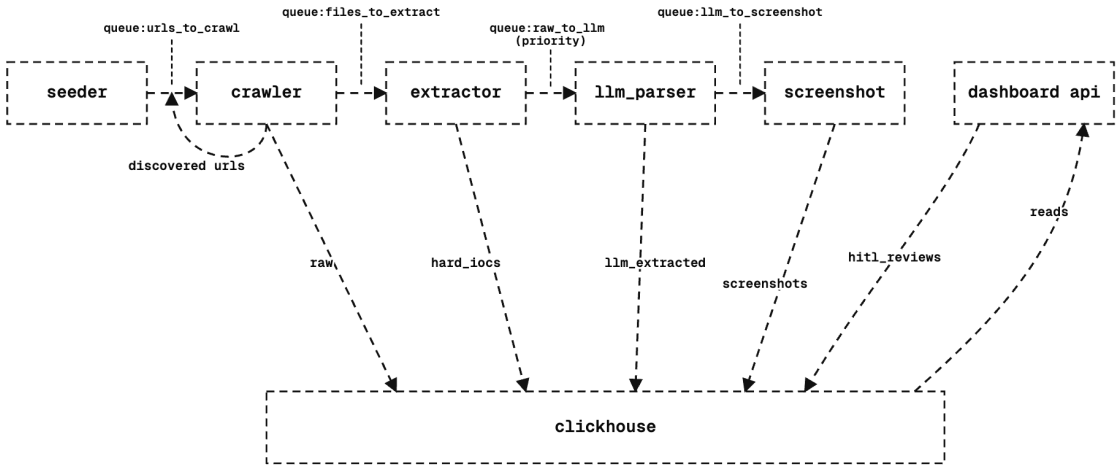


Figure 1. Queue-based architecture used in this study, showing staged collection, filtering, extraction, parsing, and storage flow for reproducible analysis.

addresses (Bitcoin, Monero, Ethereum), email addresses, onion links, Telegram handles, and public IPv4 addresses, and write outputs to `hard_iocs`. In parallel, the sentry stage evaluates content with a curated keyword list organised into three classes: Critical Threat, High Interest, and Commerce. The exact critical keyword set used in the final reporting configuration is listed in Appendix D. Critical Threat terms represent direct service-risk cues. High Interest terms capture adjacent offensive and infrastructure language. Commerce terms capture offer-context language such as pricing and sales framing. The motivation for this structure is to preserve broad candidate coverage under noisy hidden-service text while preventing weak commercial language from dominating relevance decisions. Appendix C and Appendix D provide term-level selection rationale, including direct service labels, technical descriptors, multilingual variants, and choices made to control ambiguity. Scoring applies class-specific weights and caps: Critical Threat (weight 50, cap 200), High Interest (weight 20, cap 100), and Commerce (weight 1, cap 10). Indicator additions contribute +30 per detected wallet and +30 per detected public IPv4 address.

$$
\begin{aligned}
S = \min(50\,c_{\text{critical}}, 200) \\
+ \min\big(20\,c_{\text{high}}, 100\big) \\
+ \min(c_{\text{commerce}}, 10) \\
+ 30\,n_{\text{wallet}} + 30\,n_{\text{ipv4}} \\
\text{sentry-passed} \iff S \geq 50
\end{aligned}
$$

where c_{critical}, c_{high}, and c_{commerce} are per-page keyword-match counts for the Critical Threat, High Interest, and Commerce classes, and n_{wallet} and n_{ipv4} are per-page detected indicator counts for wallet addresses and public IPv4 addresses, respectively. Records with a sentry score ≥ 50 are given priority for `llm_parser`. Operationally, this stage reduces weakly related crawl noise and keeps admission broad enough that parser-stage exclusion does not depend on a single brittle rule. Equal bonus values keep admission broad. Wallet and IPv4 signals are supporting cues, not standalone evidence. Final DDoS specificity is left to parser labels because public IPv4 mentions can be incidental.

- **Language-model parsing gate and full parse:** `llm_parser` is the throughput bottleneck, so parsing uses a three-stage policy. Gate 1 applies keyword prefiltering on normalised text. Gate 2 runs a quick relevance classifier on leading content. This leading-content sampling is a

screening trade-off: routing cues usually appear early in page text, while full-page pre-screening increases latency. In the fixed reporting snapshot used for all reported RQ results, Gate 2 used the first 3,000 characters. A separate later engineering test evaluated a 1,000-character preflight for throughput, but those outputs are not part of the reported results. Only accepted pages proceed to full extraction. If Gate 2 classification fails, the page is allowed to proceed to full extraction (fail-open), which favours recall and queue continuity at some precision cost. Full extraction uses sliding-window chunking with a 2,000-character window and a 5,000-character merged-chunk cap, and a per-chunk parse timeout of 300 s. The 2,000-character window was selected to keep a complete service-offer segment together in most pages, including surrounding context for price, vector, and contact fields. The 5,000-character merge cap was selected to reduce request overhead on short pages while limiting prompt-size variance and response latency on longer pages. The 300 s timeout bounds worst-case stalls and reduces queue starvation. For accepted pages, the parser extracts structured threat information and labels, including item type, vendor, contacts, price, currency, service categories, threat level, confidence, and chunk metadata. Parsed outputs are stored in `llm_extracted` and used in the analysis stage. This stage makes labels more specific. Results can vary with model, quantisation, and prompt configuration, so deterministic pre-filters and post-run verification are retained around it. The extraction role of language models follows current evidence on information-extraction utility [9].

- **Screenshot worker:** a dedicated screenshot worker consumes URLs processed by `llm_parser`, launches headless Chromium sessions, and captures screenshots to preserve analyst-facing context that may not be evident from extracted text alone.
- **Dashboard integration and human verification:** the dashboard combines screenshot artefacts with language-model outputs and supports manual human verification and correction before downstream interpretation.
- **Harmful-content filtering:** harmful-content filtering is implemented as a reusable hazmat block-list loaded from `FILTER_HAZMAT_FILE` (default `data/hazmat_terms.txt`). Terms are lower-cased. Empty and comment lines are discarded. Text checks use whole-word matching to reduce accidental hits. URL checks use a separate path. Host, path, query, and fragment are decoded, normalised, and then evaluated with both token-aware and compact matching to catch obfuscated concatenations. In this deployment, the term list was limited to English expressions as a scope-bounding implementation choice. This scope was also chosen to avoid a broader multilingual list that would increase matching overhead and reduce throughput. The same filter is applied at four control points: crawler URL pre-checks before fetch, crawler content checks before storage, `llm_parser` URL pre-checks before database reads, and pre-analysis cleanup/requeue tooling. This repeated placement was introduced during iterative development as a hardening step after stage-specific bypass cases were observed. It provides defence in depth, although a simpler single-pass prototype would not require this level of redundancy. Hazmat term filtering is based on the LDNOOBW[1] bad-words list and adapted to project scope.
- **Deduplication:** URL-level deduplication alone was insufficient because mirrors, template-spam sites, and NSFW-tag variants repeatedly changed endpoints while reusing the same layout and near-identical text. URL-only suppression misses this behaviour, while SimHash content fingerprints suppress near-duplicates across endpoint churn and minor text edits. The primary duplicate control therefore uses SimHash for near-duplicate detection, following established web-crawling principles [17]. In this deployment, a 64-bit fingerprint with Hamming distance ≤ 3 and a Redis time-to-live of 7 days was implemented as a project-specific calibration choice. The ≤ 3 threshold was selected as a conservative near-duplicate boundary to retain small template

[1] `https://github.com/LDNOOBW/List-of-Dirty-Naughty-Obscene-and-Otherwise-Bad-Words`

edits while reducing duplicate processing load. A known limitation is that very short texts can produce unstable or weakly discriminative SimHash signatures, so short-page duplicate decisions may be less reliable than long-page decisions. The 7-day retention window reduces repeated processing under endpoint churn while keeping memory and recrawl behaviour bounded in this deployment context. In the final run, this control reduced repeated template pages, while remaining non-exhaustive under project time limits and competing priorities.

C. Operationalisation of research questions

The analyses map directly to three research questions and are generated from rerunnable analysis queries.

1) **RQ1, detection and classification:** quantify contraction from raw intake to sentry-passed, persisted-output, and DDoS-classified sets at URL and domain level.
2) **RQ2, ecosystem linkage:** identify repeated technical indicators and descriptive identifiers across distinct domains, then summarise component concentration and top reused attributes.
3) **RQ3, market characteristics:** analyse pricing models, advertised Layer 4/7 attack capabilities, and observable service-availability signals, together with field coverage for each output field.

RQ1 outputs are evaluated as stage-wise retention and absolute loss between adjacent stages. For RQ1, the persisted-output stage is defined as URLs with at least one parser row stored in `llm_extracted` within the sentry-passed cohort. This is an extraction-yield measure, not a direct parser-throughput measure. RQ2 outputs are evaluated as graph size, component concentration, and cross-domain reuse counts for top attributes. RQ3 outputs are evaluated as composition and distribution summaries with explicit denominator tracking, so low-coverage fields are interpreted as bounded subsets rather than market totals.

D. Operational definition of DDoS-classified records

In the final analysis run, a URL is counted as DDoS-classified only if it appears in the sentry-passed URL set and has at least one associated parser record with `lower(item_type) LIKE '%ddos%'`. Operationally, this is implemented as a sentry-constrained semi-join from canonical URLs to parser rows, followed by distinct URL and distinct domain aggregation. This query logic is applied to the sentry-passed subset before RQ1 totals and downstream RQ2 and RQ3 subsets are derived. A sentry-passed URL can still produce no persisted parser row, so extracted-stage counts should not be interpreted as total parser attempts. Consequently, DDoS counts reflect parser-labelled evidence within filtered candidates, not all crawled pages, and should not be interpreted as how common DDoS pages are in the full hidden-service population.

E. Quality controls

Quality assurance was implemented as a layered control framework, with each layer tied to a specific error mode and an auditable output.

At intake, keyword-based sentry scoring controlled thematic scope by giving priority to pages that combined explicit threat terms with supporting context signals. This reduced drift from the research objective while preserving broad candidate coverage for ambiguous pages that required downstream model judgement.

At collection and parsing boundaries, harmful-content filtering removed clearly out-of-scope unsafe material before storage and before model processing. This reduced contamination risk in later analytical tables and lowered unnecessary processing load.

At the content-normalisation stage, duplicate suppression based on SimHash reduced repeated processing of mirrored or template-derived pages across changing URLs [17]. This control limited inflation of counts driven by endpoint churn rather than substantive new content.

At interpretation stage, DDoS-classified reporting was restricted to sentry-passed URLs with at least one matching parser label. This constraint preserved consistency between admission logic and final reported subsets.

Collectively, these controls define an auditable path from raw collection to reported tables and figures. This design reflects project traceability objectives and documents stage boundaries to support reproducible

interpretation within the study scope. They improve robustness of relative patterns, but they do not provide exhaustive coverage of all hidden-service activity.

F. Validation boundaries

This study does not report formal precision and recall estimates for the full corpus because a complete hand-labelled benchmark was not available.

Validation is therefore based on internal consistency and boundary checks. First, stage-wise contraction is monotonic and stable at both URL and domain level, which supports consistency of filtering behaviour. Second, sentry-score overlap between DDoS-classified and other sentry-passed pages is interpreted as expected behaviour for an inclusive admission gate, with parser labels providing final specificity. Third, manual review of high-frequency shared indicators indicates substantial mirror effects, so linkage outputs are treated as screening signals rather than attribution evidence.

Accordingly, the reported findings are strongest for relative comparisons within the observed dataset, for example contraction patterns and concentration effects. They are less definitive for population-level frequency, operator-level attribution, and complete economic description. These limitations reflect passive public-surface observation, sparse structured fields in some categories, and the absence of active engagement with gated markets.

G. Ethical considerations

The workflow is passive by design and follows established cybercrime web-scraping ethics guidance on minimising harm, limiting unnecessary personal data processing, and reporting aggregate behaviour rather than operationally sensitive detail [19]. The study does not purchase illicit services, does not execute attacks, and does not provide actionable instructions for misuse. Reporting is restricted to aggregate, non-operational findings that describe ecosystem behaviour rather than service-deployment detail. Active engagement with operators, account-gated workflows, and transactional validation is excluded from this study's methodological scope.

IV. Experiment and Analysis Design

This section explains how iterative development work was consolidated into one reporting analysis pass for final findings and figures.

A. Repeated development context and final reporting protocol

The implementation and cleanup process followed repeated development cycles. Development cycles were used to stabilise the controls that most affect measurement quality: duplicate suppression, harmful-content filtering, parser admission behaviour, and handling of dynamic pages. These activities were treated as engineering calibration, not as final evidence for interpretation.

After stage behaviour stabilised, one reporting dataset was generated and analysed without further parameter changes. This separation between calibration and reporting reduces the risk of tuning decisions against final outcomes.

The reporting procedure followed three steps:

1) Execute the full pipeline with the final control stack (sentry admission, harmful-content filtering, duplicate suppression, and parser classification) to produce the analysis-ready dataset.
2) Run data-readiness checks to confirm required tables are populated, stage totals are internally coherent, and analysis joins can be reproduced.
3) Execute analysis blocks in fixed order: detection and classification (RQ1), linkage structure (RQ2), and market summaries (RQ3).

For repeatability, this reporting pass used the fixed subscribed seeder keyword seed set and fixed critical keyword set documented in Appendix C and Appendix D.

This protocol keeps development tuning and reported evidence distinct, while preserving one consistent analytical basis.

B. Reporting terminology and scope

The reported counts follow the working definitions in the methodology section. In particular, DDoS-classified records are derived from sentry-passed URLs with at least one parser label matching the

DDoS criterion. URL-level, domain-level, and extracted-row summaries are reported separately to keep denominators consistent across research questions.

For transparency, some internal exports and plot titles use the label "Confirmed DDoS". In this manuscript, the same subset is consistently referred to as DDoS-classified, following the working definition in the methodology section. The Results section follows this reporting protocol and excludes exploratory calibration outputs.

V. Results

This section presents findings from the final fixed reporting snapshot defined in the experiment design section. Deduplication was stabilised in earlier development runs, and harmful-content filtering was applied in stages during execution and repeated during cleanup. All values reported below come from one validated run. This validated run used the fixed subscribed seeder keyword seed set and fixed critical keyword set documented in Appendix C and Appendix D.

A. Detection and classification evidence (RQ1)

RQ1 examines selectivity from raw intake to DDoS-classified records. The final funnel starts with 389,006 raw pages. Of these, 304,949 are sentry-passed (score ≥ 50). Within this subset, 3,919 URLs have at least one persisted parser-output row in `llm_extracted`, and 301 are DDoS-classified. At domain level, 3,224 raw domains are reduced to 2,426 sentry-passed domains, then 263 domains with persisted parser output, and finally 59 domains with DDoS-classified pages. In absolute terms, sentry filtering removes 84,057 pages from raw intake. The persisted-output stage contains 3,919 URLs before the DDoS-specific subset is identified at 301 URLs. The final DDoS-classified set therefore represents 0.08% of raw pages and 7.68% of URLs with persisted parser output. At domain level, the final 59-domain subset represents 1.83% of the 3,224 raw domains, indicating clear breadth contraction in addition to page-volume contraction. Across these 59 domains, the 301 DDoS-classified URLs yield 681 individual DDoS service listings in row-level parser output used for market analysis.

Table 1. Stage-wise counts and retention relative to the previous stage.

Stage	URLs	URL ret.	Domains	Dom. ret.
Raw intake	389,006	100.00%	3,224	100.00%
Sentry passed	304,949	78.39%	2,426	75.25%
Persisted parser output	3,919	1.29%	263	10.84%
DDoS-classified	301	7.68%	59	22.43%

Table 1 reports stage-to-stage retention. Percentages quoted elsewhere against raw intake are explicitly raw-baseline percentages. The extracted stage in this table denotes persisted parser-output yield within sentry-passed candidates. It is not a direct measure of total parser attempts or completion throughput.

Table 1 and Figure 2 show strong stage-wise reduction from broad intake to a compact DDoS-classified set. The matching contraction across pages and domains indicates narrowing ecosystem breadth, not only removal of repeated pages from a small set of domains. These results indicate strong contraction from broad intake to a compact DDoS-classified subset at both URL and domain level, narrowing the observable cross-domain market surface under the applied reporting rules.

Figure 3 shows strong overlap between DDoS-classified pages and the remaining sentry-passed pages, with the largest outliers in the non-DDoS sentry-passed group. This pattern indicates that sentry scoring acts as a broad admission gate for potentially relevant content, while parser output provides final DDoS specificity. Because no fully labelled negative set was audited in this snapshot, recall is not estimated here. The overlap also implies that sentry score level alone cannot be interpreted as a DDoS-classification proxy without parser confirmation.

B. Linkage network findings (RQ2)

RQ2 examines repeated indicators and identifiers across domains after filtering to should-run URLs with at least one parser item type label matching DDoS, booter, or stresser. In the current run, RQ2 is restricted to 301 URLs across 59 domains. Within this group, shared-indicator links connect 16 domains.

Figure 4 reports a mixed domain-indicator graph

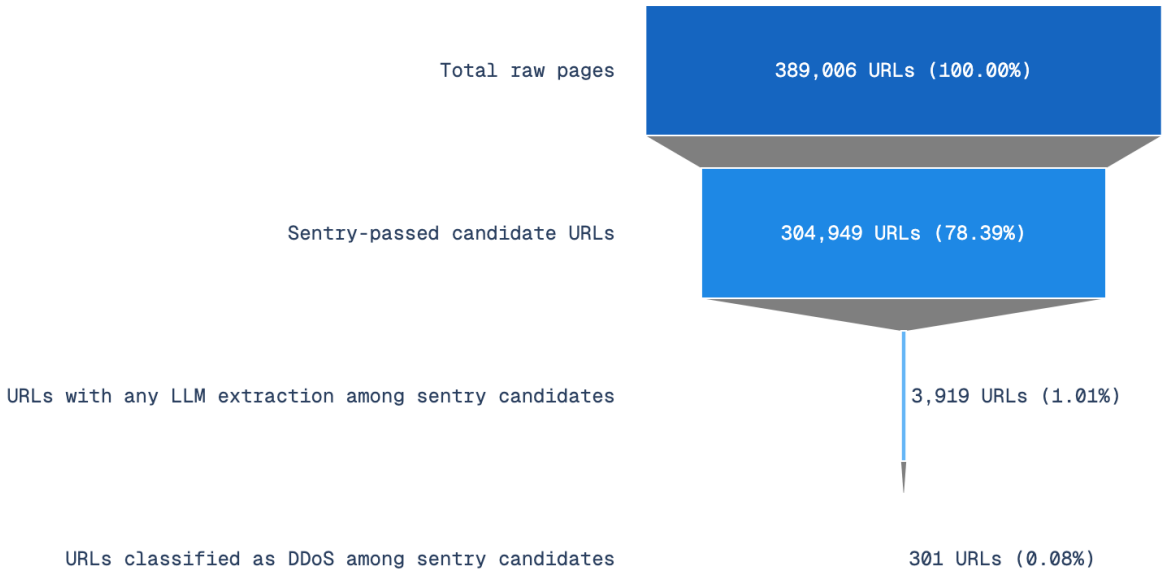


Figure 2. Retention funnel from raw collection to DDoS-classified records.

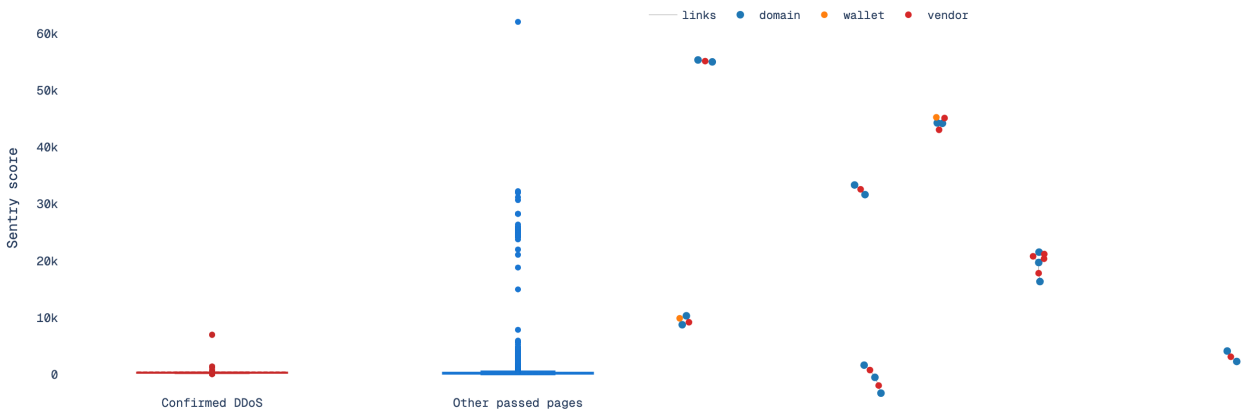


Figure 3. Distribution of sentry scores for DDoS-classified pages versus other sentry-passed pages, showing broad overlap and high-score outliers in the non-DDoS group.

with 30 nodes and 28 edges in 7 clusters. The 30 nodes represent 16 linked domains and 14 shared technical indicators (12 vendor handles and 2 cryptocurrency wallets). The largest connected component contains 23.33% of all nodes, which indicates a fragmented structure rather than one dominant linkage core.

RQ2 therefore delivers a fragmented but actionable linkage structure for screening and follow-up analysis.

Figure 4. Network linkage structure derived from shared indicators across domains.

Figure 5 shows DDoS-classified listing counts for the top 20 domains, with bar colour encoding average `llm_parser` confidence. Here, listing count is a row-level measure from `llm_parser` outputs rather than a distinct-URL count, so one URL can contribute multiple listings. The distribution is skewed, with a small subset of domains accounting for most classified listings. This skew supports analyst priority setting, but it does not by itself imply that high-volume domains are distinct operators.

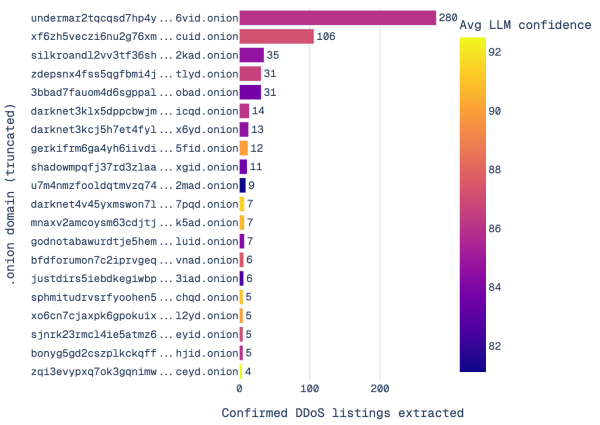


Figure 5. Top 20 domains by DDoS-classified listing count, with bar colour encoding average `llm_parser` confidence.

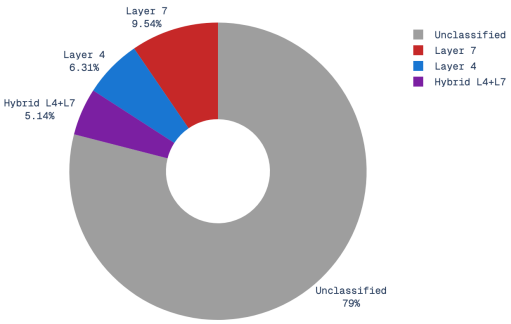


Figure 6. Attack-layer composition across structured market rows.

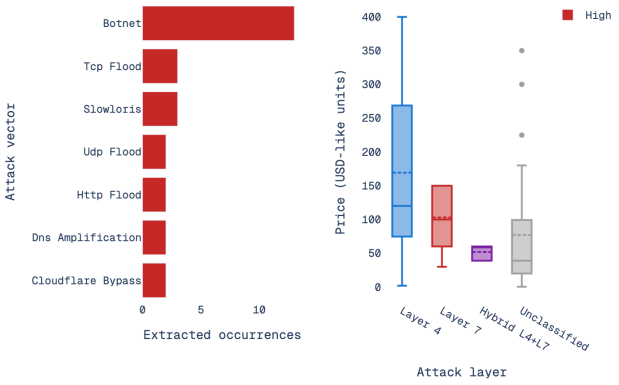


Figure 7. Two-panel market summary: stacked attack-vector counts by threat-level tag (left) and price-by-attack-layer boxplots for parseable rows (right).

C. Market characteristics from observed coverage (RQ3)

RQ3 analyses pricing, advertised attack capabilities, and service-availability signals across 681 structured rows. These rows are row-level listings: one DDoS-classified URL can contain multiple advertisements. The 301 DDoS-classified URLs across 59 domains expand to 681 individual service listings for market analysis. Figure 6 shows that layer composition is dominated by unclassified rows (79.0%), followed by Layer 7 (9.54%), Layer 4 (6.31%), and hybrid Layer 4+Layer 7 (5.14%). This supports broad capability profiling of the observable subset, while also showing that most records do not expose a specific layer label.

In quantitative terms, 21.0% of rows provide an explicit layer assignment. Therefore, layer-distribution claims are precise for a minority subset of visible listings.

Vector evidence follows a similar pattern. In Figure 7, the left panel reports stacked vector counts by threat-level tags, with dominant-vector concentration at 48.15% of vector-labelled rows, while the right panel shows price-by-attack-layer boxplots for the parseable subset. Specific-vector coverage is 3.96%, which is sufficient to surface recurring vectors but not to support robust interpretation of vector-level market diversity.

This coverage level is sufficient to identify repeated advertised vectors and guide analytical prioritisation. Prevalence estimates across the full observed market remain preliminary.

Pricing coverage is likewise focused: 7.64% of rows are parseable for price, so Figure 8 should be read as a subset summary rather than a complete market estimate. In the current snapshot, the dense-currency boxplot panel (USD) reports min 0.3, Q1 30.0, median 60.0, Q3 105.0, an upper fence at 180.0, and a maximum observed value of 400.0.

Consequently, this dataset supports targeted price-pattern screening, while robust market-wide estimates of median price, revenue concentration, or stable package taxonomy require broader coverage.

Within the parseable subset, some listings use duration-based prices, such as one-hour, one-day, or one-week attack windows, rather than the tiered or package-style service plans commonly reported in

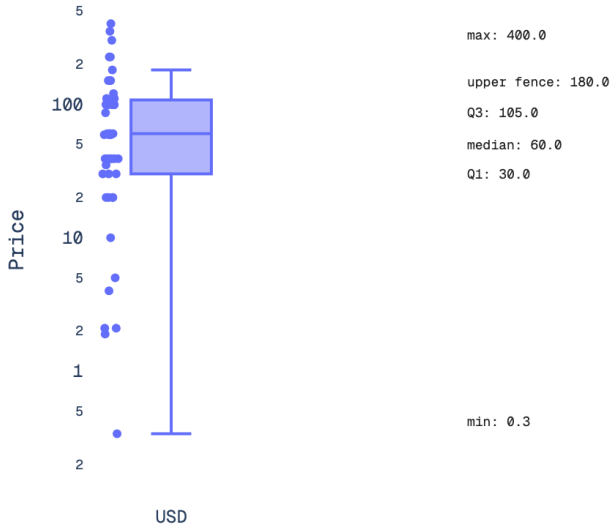


Figure 8. Price distribution for the parseable subset (52/681, 7.64%), shown on a logarithmic y-axis.

prior clear web booter studies [1, 20]. This pattern is consistent with mixed listing formats, and the current sample is not yet dense enough for stronger generalisation.

Overall, the outputs provide strong support for screening and broad market profiling, with clearly defined limits for detailed attribution and pricing inference.

VI. Discussion

A. Practical meaning for screening and analysis

The implementation shows that a layered pipeline can turn large hidden-service intake into structured records suitable for investigation and repeatable reporting. The reduction from 389,006 raw pages to 301 DDoS-classified pages, and from 3,224 raw domains to 59 domains with DDoS-classified pages, is consistent with substantial filtering selectivity in this run. At the same time, Figure 3 shows broad sentry-score overlap between DDoS-classified pages and other sentry-passed pages, with high-score outliers in the non-DDoS group. This pattern is consistent with sentry filtering as an inclusive intake filter, after which parser output determines DDoS-specific classification. For practitioners, this means the pipeline is best used as a staged screening system: sentry output for broad candidate intake, parser-labelled output for DDoS-focused queues, and analyst review for final judgement.

Linkage outputs show that repeated indicators and descriptive identifiers produce a usable relationship map. Structural concentration remains modest. In the current run, the restricted RQ2 group covers 301 URLs across 59 domains, with 16 linked domains and a mixed domain-indicator graph of 30 nodes and 28 edges in 7 clusters (16 domains, 12 vendor handles, and 2 cryptocurrency wallets). The largest connected component contains 23.33% of nodes. This supports screening and priority setting, particularly when analysts must review large numbers of candidate service brands. Manual review of top shared indicators suggests that several prominent links may reflect mirror relationships rather than independent Indicator of Compromise (IOC)-based cross-domain links. This attribution boundary does not remove the screening value of the map, but it narrows attribution strength for the most connected pairs in the current snapshot. Linkage signals are best used to form and rank hypotheses, not direct actor attribution or enforcement action without external confirmation.

The main societal implication is prioritisation: analysts can use limited review capacity on higher-signal candidates without claiming how common these services are across the full population. Ethically, these outputs are aggregate, non-operational evidence for defensive screening and should not be treated as standalone proof for attribution or intervention.

Market outputs provide a well-defined snapshot. Layer composition and dominant vectors are observable, while detailed interpretation remains focused on the observable subset because specific-vector coverage is 3.96% and price field coverage is 7.64%. The market view is therefore strong for aggregate profiling and directional analyst decisions. The strongest practical output is therefore directional intelligence, for example which domains and service descriptions are repeatedly visible, rather than complete commercial intelligence on package structure or revenue. Compared with one prior booter study that includes price-related observations [20], the observed dark web price signals in this dataset are too sparse for robust cross-surface inference.

B. Practitioner insight from Rapid7 interview

A practitioner interview conducted for this study with Christian Beek (Principal Threat Intelligence Researcher at Rapid7) provided contextual observations [21]. The summary below reports points from that session as interpretive context rather than primary empirical evidence. The participant described operations that combine continuously running automation with manual investigators, rather than fully autonomous processing.

For denial-of-service-for-hire monitoring, interview observations described a repeated workflow: detect a signal, enrich it with actor context, history, and reputation, launch additional collection or honeypots when needed, notify the client, and repeat. The interview also described text extraction as a combination of natural language processing outputs with prior case context and actor reputation. Interview observations further indicated that no single fixed threat-score framework should be inferred from one post in isolation, because scoring depends on broader confirming evidence. According to the interviewee, even high model accuracy may remain insufficient for high-impact cyber threat intelligence workflows when model-generated errors propagate into consequential analytical decisions. These interview observations align with this study's analyst-in-the-loop dashboard, calibrated parser interpretation, and clearly scoped claims on attribution and market detail.

C. Confidence boundaries

The current evidence supports three claims with different confidence strengths. First, confidence is strongest for RQ1 stage contraction and DDoS-specific narrowing, because counts are directly produced by fixed query logic in one fixed run. Second, confidence is moderate for RQ2 linkage screening value, because repeated attributes and network concentration are observable, while mirror effects temper attribution certainty. Third, confidence for RQ3 market detail is lower than for RQ1 and RQ2 because most rows lack parsable layer, vector, or price fields. These boundaries position the dataset as strong for screening and monitoring baselines, while conclusive estimates of frequency, operator attribution and control relationships, or full economic estimates require additional coverage.

D. Limitations and threats to validity

Several limitations define the scope of interpretation. Coverage uncertainty begins with multi-stage hazmat harmful-content filtering: this filtering reduces non-target noise, but it can also remove genuine denial-of-service-for-hire pages when their wording overlaps with blocked categories. Relatedly, overlap with hazmat terms can be incidental and may also reflect evasive wording patterns in collected content, and the current dataset cannot separate these mechanisms. At the same time, the filtering stack can still pass DDoS-related pages that are not direct threat offerings, including informational blog posts, technical guides, or outage announcements in which a site reports DDoS disruption. These effects can bias both recall and precision in opposite directions: relevant listings can be excluded, while non-offer pages can remain in candidate sets.

Pipeline mechanics also introduce validity constraints. Language-model extraction is currently the throughput bottleneck: the Large Language Model (LLM) stage is comparatively slow, which caps how much material can be processed under fixed runtime and infrastructure constraints. Accordingly, the RQ1 extracted stage is interpreted as persisted output yield rather than a direct measure of parser-attempt throughput. Runtime is therefore configuration-dependent and infrastructure-dependent, and this study reports interpretation-relevant settings rather than exhaustive performance benchmarking. In addition, screenshot capture and HTML extraction are not always synchronous, so visual evidence and extracted text may refer to different page states. This temporal mismatch can occasionally affect manual verification when short-lived pages or rotating landing content are involved.

External verification and market-access boundaries further restrict interpretation. Documented studies in other illicit online market contexts report law enforcement interventions that can disrupt market visibility [22]. Other work examines strategic law enforcement approaches, including deception-oriented operations [23]. In this context, apparent service visibility can reflect either genuine operator activity, mirror replication, or deliberate decoy exposure.

Distinguishing these mechanisms for specific services requires transactional or identity-grounded validation, including controlled test-purchase approaches used in darknet-market research [24]. Active engagement remains out of scope for legal and ethical reasons [19]. Authentication-gated forums and markets were also not explored in depth, so the observed activity primarily reflects publicly reachable surfaces and may under-represent closed communities. The findings therefore provide high-value visibility on the open hidden-service perimeter, rather than full ecosystem coverage.

For RQ2 specifically, linkage analysis used both LLM-structured records and hard IOC records extracted using regular expressions, constrained to analysis-scope URLs with at least one parser `item_type` label matching DDoS, booter, or stresser. This design can still permit weaker links in cases where DDoS is peripheral rather than central: in mixed markets, a shared vendor label can appear on pages that mainly advertise non-DDoS services, with DDoS only appearing in a hyperlink or secondary reference. Mirror infrastructure also remains dominant in the linkage graph even after duplicate suppression, which makes it difficult in the current snapshot to separate true cross-service or cross-server IOC reuse from cross-domain cloning. As a result, linkage density is best used as an investigation-priority signal, not as a direct measure of unique operator coordination.

Finally, market-detail completeness is limited because many DDoS listings do not expose price or technical delivery details in-page, such as target layer, duration, or execution parameters. Many listings in this dataset omit these details on-page. Off-page negotiation through private contact channels is a plausible interpretation, but it remains unverified because direct vendor engagement was intentionally not performed. This constrains claims about average pricing or feature standardisation to the observable subset.

E. Future work

Near-term improvements are bounded to changes that fit the current passive design. First, implement keyword-priority frontier scheduling so pages likely to contain target terminology are visited earlier. Second, increase extraction coverage for pricing and vectors to improve downstream market resolution. Third, strengthen anti-bot and dynamic-page handling through systematic browser-automation hardening and challenge-handling evaluation under strict reliability and safety constraints. Fourth, compare overlap between clear web and dark web denial-of-service-for-hire listings to describe migration and mirroring patterns across surfaces.

VII. Conclusion

This study addresses RQ1 to RQ3 with one fixed and reproducible measurement snapshot, providing a clear evidence baseline within a defined observation window. For RQ1, the workflow shows strong and traceable contraction from broad intake to DDoS-classified evidence at both URL and domain level. The persisted-output stage should be read as extraction yield within admitted candidates, not as total parser throughput.

For RQ2, cross-domain linkage signals are observable and practically useful for prioritisation. At the same time, mirror prevalence and limited multi-signal confirmation require conservative interpretation, so these links are treated as screening evidence rather than attribution proof.

For RQ3, market signals are observable across 681 structured rows, with uneven field completeness. Layer composition and dominant vectors can be summarised, while specific-vector coverage is 3.96% and price field coverage is 7.64% (52/681), so fine-grained claims remain focused on the observable subset. Accordingly, pricing and capability findings are strongest as evidence for the observable subset rather than full-market totals.

Taken together, RQ1 to RQ3 provide a concrete answer to the main research question on the observable structure of dark web booter ecosystems. The observable surface can be narrowed reproducibly from noisy intake to a small DDoS-classified subset, yielding a manageable measurement target. Cross-domain linkage signals are present and actionable for prioritisation, although they are fragmented and often mirror-heavy, so attribution still requires external corroboration. Market-facing signals are also observable, with sparsity primarily affecting fine-grained economic and practical interpretation.

In practice, the pipeline is well suited to repeatable screening and monitoring, and it establishes a strong foundation for deeper attribution and market measurement as coverage improves.

Overall, the main contribution is a transparent measurement chain that supports repeatable evidence production for dark web booter analysis and makes interpretation boundaries explicit. The conclusions are strongest for stage-wise contraction and screening utility, while sparse market-detail fields and surface-level visibility remain the main constraints. The observations primarily represent publicly reachable hidden-service surfaces and likely under-represent authentication-gated communities. The most direct next step is to improve evidence quality and coverage through keyword-priority crawl scheduling, richer extraction of pricing and attack-vector fields, stronger mirror-aware duplicate suppression, and longer observation windows. A complementary step is to analyse overlap between clear web and dark web listings and broaden sampling of authentication-gated sources under legal and ethical constraints.

A. Prompts Used in the Reported Run

The parser applies two LLM prompts in sequence: a short classification gate and a full extraction prompt. The templates below reflect the prompt content used in the implementation, with runtime placeholders shown explicitly.

Gate 2 quick classification prompt.

Instruction. Classify the page in one word from the labels below.

Labels.

- `DIRECTORY`: page whose primary purpose is listing or indexing links, such as Hidden Wiki pages, link lists, and search indexes.
- `LOGIN`: login form, registration page, 404 page, or empty placeholder.
- `THREAT`: illegal or criminal cybersecurity content, such as malware, ransomware, exploits, hacking-for-hire, DDoS services, botnets, stolen credentials, leaked databases, and zero-day content.
- `MARKET`: dark web marketplace or forum advertising illegal goods or services, such as fake documents, stolen bank cards, drugs, forged IDs, SIM-swapping, money laundering, illegal software, or criminal-for-hire services.
- `OTHER`: clearly unrelated and fully legal content.

Decision rule. A page selling hacking, DDoS attacks, fake documents, bank-card fraud, or similar illegal services must be classified as `MARKET` or `THREAT`, not `OTHER`.

Input placeholder. `<first 3000 characters from page content>`

Required output. One word only.

Full extraction prompt template.

System instruction. The model acts as a threat-intelligence analyst and extracts actionable intelligence from dark web content.

Critical rule. Link directories, hidden wikis, and index pages are not threats and must be filtered out.

Page-context fields.

- `Page URL: <url>`
- `Page Title: <page_title_if_available>`
- `Page Introduction: <page_intro_if_available>`

Input fields.

- `Keywords to Monitor: <merged keyword list>`
- `Note: chunk <n> of <N> when applicable`
- `Content Chunk: <chunk text>`

Reasoning sequence.

1) Type analysis: determine whether the page is primarily a directory or a login page. If yes, return `{"relevant": false}`.
2) Relevance check: determine whether content contains monitored keywords or strong semantic equivalents in meaningful context. If no, return `{"relevant": false}`.
3) Extraction: if relevant, extract threat details.

Output contract.

- Output must be valid JSON only, with no markdown.
- Irrelevant content returns `{"relevant": false}`.
- One threat returns one object with fields: `relevant`, `category`, `actor`, `intent`, `price`, `target`, `botnet_name`, `method`, `port`, `confidence`, `summary`.
- Multiple distinct threats in one chunk return an array of threat objects.

B. Seeder Search Engines and Sources

The seeder search-engine list was informed by ROBIN and extended during iterative testing.[2]

- Ahmia: `http://juhanurmihxlp77nkq76byazcldy2hlmovfu2epvl5ankdibsot4csyd.onion/search/?q=\{query\}`
- OnionLand: `http://3bbad7fauom4d6sgppalyqddsqbf5u5p56b5k5uk2zxsy3d6ey2jobad.onion/search?q=\{query\}`

[2] `https://github.com/apurvsinghgautam/robin`

- Torgle: http://iy3544gmoeclh5de6gez2256v6pjh4omhpqdh2wpeeppjtvqmjhkfwad.onion/torgle/?query=\{query\}
- Amnesia: http://amnesia7u5odx5xbwtpnqk3edybgud5bmiagu75bnqx2crntw5kry7ad.onion/search?query=\{query\}
- Kaizer: http://kaizerwfvp5gxu6cppibp7jhcqptavq3iqef66wbxenh6a2fklibdvid.onion/search?q=\{query\}
- Anima: http://anima4ffe27xmakwnseih3ic2y7y3l6e7fucwk4oerdn4odf7k74tbid.onion/search?q=\{query\}
- Tornado: http://tornadoxn3viscgz647shlysdy7ea5zqzwda7hierekeuokh5eh5b3qd.onion/search?q=\{query\}
- TorNet: http://tornetupfu7gcgidt33ftnungxzyfq2pygui5qdoyss34xbgx2qruzid.onion/search?q=\{query\}
- Torland: http://torlbmqwtudkorme6prgfpmsnile7ug2zm4u3ejpcncxuhpu4k2j4kyd.onion/index.php?a=search&q=\{query\}
- Find Tor: http://findtorroveq5wdnipkaojfpqulxnkhblymc7aramjzajcvpptd4rjqd.onion/search?q=\{query\}
- Excavator: http://2fd6cemt4gmccflhm6imvdfvli3nf7zn6rfrwpsy7uhxrgbypvwf5fad.onion/search?query=\{query\}
- Onionway: http://oniwayzz74cv2puhsgx4dpjwieww4wdphsydqvf5q7eyz4myjvyw26ad.onion/search.php?s=\{query\}
- Tor66: http://tor66sewebgixwhcqfnp5inzp5x5uohhdy3kvtnyfxc2e5mxiuh34iid.onion/search?q=\{query\}
- OSS: http://3fzh7yuupdfyjhwt3ugzqqof6ulbcl27ecev33knxe3u7goi3vfn2qqd.onion/oss/index.php?search=\{query\}
- Torgol: http://torgolnpeouim56dykfob6jh5r2ps2j73enc42s2um4ufob3ny4fcdyd.onion/?q=\{query\}
- The Deep Searches: http://searchgf7gdtauh7bhnbyed4ivxqmuoat3nm6zfrg3ymkq6mtnpye3ad.onion/search?q=\{query\}
- Haystak: http://haystak5njsmn2hqkewecpaxetahtwhsbsa64jom2k22z5afxhnpxfid.onion/?q=\{query\}
- Torch: http://torchdeedp3i2jigzjdmfpn5ttjhthh5wbmda2rr3jvqjg5p77c54dqd.onion/search.php?query=\{query\}&action=search

C. Initial Subscribed Keyword Seed Set for Reporting

The final reporting run started from a fixed subscribed keyword seed set stored in Redis and consumed by the seeder during discovery. This set was held constant in the reporting configuration so that source coverage and downstream comparisons remained reproducible across runs.

```
ddos
booter
stresser
botnet
mirai
ддос
стрессер
ботнет
положить сайт
僵尸网络
ddos攻击
layer7
layer4
amplification
udp flood
boatnet
```

This seed set was a project-specific design choice that combines direct booter terms, attack-descriptor terms, multilingual terms, and an anti-automatic-flagging obfuscation variant of `botnet` (`boatnet`). The mix was intended to broaden retrieval across service-advertising language, operational jargon, and language-specific postings in this dataset.

Term-level inclusion followed four rules. First, direct service labels (`ddos`, `booter`, `stresser`) were included as explicit offer cues. Second, technical descriptors (`layer4`, `layer7`, `amplification`, `udp flood`, `botnet`, `mirai`) were included because operators can advertise these as package characteristics. Third, Russian and Chinese variants (ддос, стрессер, ботнет, положить сайт, 僵尸网络, ddos攻击) were included to reduce English-only retrieval bias. Fourth, `boatnet` was retained as an anti-automatic-flagging obfuscation variant of `botnet`, not as an accidental misspelling. Highly generic single-word terms were avoided in this subscribed seed set when they broadened crawl noise without improving denial-of-service-for-hire specificity.

D. Critical Keyword Set and Provenance

The critical keyword set used in this prototype is project-specific and was manually curated by the author. It was assembled from analyst seed terms informed by common cyber threat intelligence and underground-market terminology, rather than imported from a single external repository. The list shown below is the exact `CRITICAL_KEYWORDS_FILE` content used in the final reporting configuration.

```
# [Cyber threats]
ddos
ransomware
botnet
exploit
malware
phishing
c2
c&c
0day
zero-day
backdoor
trojan
rat
rce
payload

# [Data / credentials]
credential
dump
leak
breach
stolen
hacked
cve-

# [Crypto / commerce]
bitcoin
btc
monero
xmr
crypto
escrow

# [Carding / fraud]
carding
cvv
fullz
dumps
bins
```

The critical list is intended as a high-signal gate rather than an exhaustive cybercrime lexicon. Direct offensive cues (`ddos`, `botnet`, `exploit`, `rce`, `payload`) capture explicit attack-service language. Variant pairs (`c2/c&c`, `0day/zero-day`, `bitcoin/btc`, `monero/xmr`) reduce misses from stylistic variation. Data-compromise terms (`credential`, `dump`, `leak`, `breach`, `cve-`) and fraud-market terms (`carding`, `cvv`, `fullz`, `dumps`, `bins`) were retained as context signals because denial-of-service-for-hire content can appear on mixed underground market pages. Payment tokens (`bitcoin`, `btc`, `monero`, `xmr`, `crypto`, `escrow`) capture transaction framing and are interpreted with lower class weight in sentry scoring so they do not dominate admission by themselves.

For harmful-content screening terms, the project used and adapted the LDNOOBW[3] bad-words list to study scope.

Acknowledgements

The author thanks Assistant Professor Harm Griffioen for overall project supervision and PhD researcher Maarten Weyns for frequent checkups and support.

References

[1] Karami, M., and McCoy, D., "Understanding the Emerging Threat of DDoS-as-a-Service," *Proceedings of the 6th USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET)*, USENIX Association, 2013. URL `https://www.usenix.org/conference/leet13/workshop-program/presentation/karami`.

[2] Santanna, J. J., Rijswijk-Deij, R. v., Hofstede, R., Sperotto, A., Wierbosch, M., Granville, L. Z., and Pras, A., "Booters — An analysis of DDoS-as-a-service attacks," *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, 2015. doi: `10.1109/INM.2015.7140298`.

[3] Santanna, J. J., Schmidt, R. O. D., Tuncer, D., Vries, J. d., Granville, L. Z., and Pras, A., "Booter blacklist: Unveiling DDoS-for-hire websites," *Conference on Network and Service Management*, 2016. doi: `10.1109/CNSM.2016.7818410`.

[4] Kopp, D., Wichtlhuber, M., Poese, I., Santanna, J., Hohlfeld, O., and Dietzel, C., "DDoS Hide & Seek:

[3] `https://github.com/LDNOOBW/List-of-Dirty-Naughty-Obscene-and-Otherwise-Bad-Words`

On the Effectiveness of a Booter Services Takedown," *ACM/SIGCOMM Internet Measurement Conference*, 2019. doi: 10.1145/3355369.3355590.

[5] Shakarian, P., "Dark-Web Cyber Threat Intelligence: From Data to Intelligence to Prediction," *Information*, Vol. 9, No. 12, 2018, p. 305. doi: 10.3390/info9120305.

[6] Pastrana, S., Thomas, D. R., Hutchings, A., and Clayton, R., "CrimeBB: Enabling Cybercrime Research on Underground Forums at Scale," *The Web Conference*, 2018. doi: 10.1145/3178876.3186178.

[7] Gupta, A., Maynard, S. B., and Ahmad, A., "The Dark Web Phenomenon: A Review and Research Agenda," *Proceedings of the 30th Australasian Conference on Information Systems (ACIS)*, ACIS, 2019. URL https://aisel.aisnet.org/acis2019/1.

[8] De Pascale, D., Cascavilla, G., Tamburri, D. A., and Van Den Heuvel, W.-J., "CRATOR: A Dark Web Crawler," *arXiv preprint arXiv:2405.06356*, 2024. URL https://arxiv.org/abs/2405.06356.

[9] Han, R., Yang, C., Peng, T., Tiwari, P., Wan, X., Liu, L., and Wang, B., "An Empirical Study on Information Extraction using Large Language Models," 2024. URL https://arxiv.org/abs/2409.00369.

[10] Santanna, J. J., Vries, J. d., Schmidt, R. d. O., Tuncer, D., Granville, L. Z., and Pras, A., "Booter list generation: The basis for investigating DDoS-for-hire websites," *Int. J. Netw. Manag.*, 2018. doi: 10.1002/NEM.2008.

[11] Hutchings, A., and Clayton, R., "Exploring the Provision of Online Booter Services," *Deviant Behavior*, Vol. 37, No. 10, 2016, pp. 1163–1178. doi: 10.1080/01639625.2016.1169829, URL https://doi.org/10.1080/01639625.2016.1169829.

[12] Christin, N., "Traveling the Silk Road: A Measurement Analysis of a Large Anonymous Online Marketplace," *Proceedings of the 22nd International Conference on World Wide Web (WWW 2013)*, ACM, 2013, pp. 213–224. doi: 10.1145/2488388.2488408.

[13] Soska, K., and Christin, N., "Measuring the Longitudinal Evolution of the Online Anonymous Marketplace Ecosystem," *Proceedings of the 24th USENIX Security Symposium (USENIX Security 15)*, USENIX Association, 2015, pp. 33–48. URL https://www.usenix.org/conference/usenixsecurity15/technical-sessions/presentation/soska.

[14] Mavroeidis, V., and Bromander, S., "Cyber Threat Intelligence Model: An Evaluation of Taxonomies, Sharing Standards, and Ontologies within Cyber Threat Intelligence," *2017 European Intelligence and Security Informatics Conference (EISIC)*, IEEE, 2017, pp. 91–98. doi: 10.1109/EISIC.2017.20.

[15] OASIS Cyber Threat Intelligence (CTI) Technical Committee, "STIX Version 2.1," , 2021. URL https://docs.oasis-open.org/cti/stix/v2.1/stix-v2.1.html.

[16] Johnson, C., Badger, M., Waltermire, D., Snyder, J., and Skorupka, C., "Guide to Cyber Threat Information Sharing," , 2016. URL https://csrc.nist.gov/publications/detail/sp/800-150/final.

[17] Manku, G. S., Jain, A., and Das Sarma, A., "Detecting Near-Duplicates for Web Crawling," *Proceedings of the 16th International Conference on World Wide Web*, ACM, 2007, pp. 141–150. doi: 10.1145/1242572.1242592.

[18] Dingledine, R., Mathewson, N., and Syverson, P., "Tor: The Second-Generation Onion Router," *Proceedings of the 13th USENIX Security Symposium*, USENIX Association, 2004. URL https://www.usenix.org/conference/13th-usenix-security-symposium/presentation/tor-second-generation-onion-router.

[19] Brewer, R., Westlake, B., Hart, T., and Arauza, O., "The Ethics of Web Crawling and Web Scraping in Cybercrime Research: Navigating Issues of Consent, Privacy, and Other Potential Harms Associated with Automated Data Collection," *Researching Cybercrimes: Methodologies, Ethics, and Critical Approaches*, edited by A. Lavorgna and T. J. Holt, Palgrave Macmillan, 2021, pp. 435–456. doi: 10.1007/978-3-030-74837-1_22.

[20] Hyslip, T. S., and Holt, T. J., "Assessing the Capacity of DRDoS-For-Hire Services in Cybercrime Markets," *Deviant Behavior*, 2019. doi: 10.1080/01639625.2019.1616489.

[21] Beek, C., "Personal communication on analyst-in-the-loop operations in threat intelligence," , 2026. Practitioner interview. Christian Beek is Principal Threat Intelligence Researcher at Rapid7.

[22] Décary-Hétu, D., and Giommoni, L., "Do police crackdowns disrupt drug cryptomarkets? A longitudinal analysis of the effects of Operation Onymous," *Crime, Law and Social Change*, Vol. 67, No. 1, 2017, pp. 55–75. doi: 10.1007/s10611-016-9644-4.

[23] Bradley, C., and Stringhini, G., "A Qualitative Evaluation of Two Different Law Enforcement Approaches on Dark Net Markets," *2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, IEEE, 2019, pp. 453–463. doi: 10.1109/EuroSPW.2019.00057.

[24] Rhumorbarbe, D., Staehli, L., Broséus, J., Rossy, Q., and Esseiva, P., "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*, Vol. 267, 2016, pp. 173–182. doi: 10.1016/j.forsciint.2016.08.032.

4

Reflection on impact of the work in the field

This chapter reflects on the contribution of the thesis to cyber threat intelligence research and practice, and it synthesises implications from the embedded article in Chapter 3. The work unifies three strands that are often handled separately: measurement of DDoS-for-hire ecosystems [10, 3, 4], dark web collection for CTI [16, 22], and LLM-assisted information extraction [33]. The core outcome is an auditable pipeline with explicit stage boundaries from intake through classification, linkage, and market summarisation.

For operational CTI, the architecture pairs deterministic filtering with model-based extraction, so it can support higher-throughput screening while keeping analyst judgement central. This enables a repeatable screening process in which candidate services are narrowed in stages and reviewed with traceable provenance. Linkage outputs based on repeated descriptors and shared IOC patterns are therefore most useful for prioritisation and hypothesis setting, rather than as direct evidence of actor coordination.

Consultation with Christian Beek at Rapid7, conducted as part of this work, provided practitioner context consistent with an analyst-in-the-loop approach: automation should provide scalable screening, while analysts retain control over consequential decisions [43]. The same practitioner perspective cautions against inferring stable threat assessments from isolated posts, where sparse evidence may lead analysts to overstate confidence.

Societally, this thesis improves visibility into a harmful service economy without normalising offensive use. Better measurement can help defenders prioritise disruption, but interpretation must avoid overclaiming attribution from mirror-heavy data and sparse fields. Legally and ethically, the workflow is limited to passive collection of publicly reachable content, excludes purchases and direct engagement, and applies harmful-content controls before storage, in line with cybercrime research ethics guidance [6]. These limits reduce exposure risk but also constrain inference about closed or authentication-gated communities.

The implications differ across stakeholder groups. For defenders and incident-response teams, the practical value is faster triage and queue prioritisation. For law-enforcement and policy stakeholders, the outputs are more suitable for hypothesis generation than direct intervention decisions without corroboration. For researchers, the main value is methodological transparency and denominator-consistent reporting. For platform operators and civil-society stakeholders, the primary concern is dual-use leakage, which motivates constrained disclosure and governance controls.

Dual-use risk remains explicit: methods that support defensive monitoring could also aid abusive reconnaissance if released without safeguards. Mitigation relies on constrained disclosure, omission of operationally sensitive exploitation detail, institutional governance, and analyst oversight for consequential outputs. Policy implications are therefore cautious: evidence pipelines can inform disruption

and regulation, but decisions should account for uncertainty and treat exploratory linkage signals as hypotheses, not proof of coordinated criminal control.

For the research community, the main impact is methodological. The study presents dark web monitoring as a fixed, stage-wise measurement protocol with explicit uncertainty boundaries rather than ad hoc crawl snapshots. This improves comparability and highlights priorities for future work: stronger validation sets, better mirror discrimination, broader coverage of sparse market fields, and clearer reporting for reproducibility.

5

Conclusion

This thesis examined how the dark web DDoS-for-hire ecosystem can be measured with an automated and auditable data-collection pipeline. The central objective was to produce evidence-based answers about observable market size, cross-domain linkage signals, and advertised service characteristics, while keeping legal and ethical constraints explicit throughout the workflow. The thesis-level research question was operationalised through article RQ1 (detection and classification), RQ2 (ecosystem linkage), and RQ3 (market characteristics).

The empirical results show a strong stage-wise contraction from broad intake to DDoS-specific evidence, directly addressing RQ1. The pipeline's filtering stages reduced 389,006 raw pages to a final set of 301 DDoS-classified URLs across 59 domains, with 304,949 sentry-passed and 3,919 extracted pages at intermediate stages. At domain level, the same process reduced 3,224 raw domains to 59 with DDoS-classified pages. This contraction demonstrates that the proposed pipeline can narrow a noisy hidden-service landscape into a tractable analysis subset.

For linkage analysis, repeated indicators provided actionable screening evidence, but the graph remained fragmented and mirror-heavy, which answers RQ2 with bounded attribution confidence. Signals therefore support analyst triage and hypothesis ranking, rather than direct operator attribution. For market profiling, useful aggregate signals were observable, but field sparsity remained a major constraint, which limits RQ3 to subset-level interpretation: in the analysed sample of 681 service records, parseable price fields were found in 7.64% (52/681) of records and specific attack-vector coverage was 3.96% (27/681). These limitations constrain the conclusions that can be drawn about detailed market structure and pricing behaviour.

Taken against the contribution set defined in Chapter 1, the thesis contributes: an empirical characterisation of observable dark web DDoS-for-hire activity, a mirror-aware linkage analysis based on shared indicators, a structured dataset for repeatable analysis, an implemented CTI pipeline as a measurement instrument, and a safety-oriented filtering workflow for risk-reduced collection. In practice, the resulting system is best suited for repeatable monitoring, prioritisation, and analyst-in-the-loop investigation support. It is not a basis for complete ecosystem attribution or full economic quantification without stronger coverage and external validation.

Future work should prioritise higher extraction coverage for pricing and vector fields, improved linkage methods that reliably distinguish content mirrors from genuine operator connections, and longer observation windows to assess temporal stability. A second priority is stronger validation, including labelled quality checks for classification and extraction outputs. Future work should also extend multilingual and governance-aware safety filtering, and formalise dashboard review protocols, so that the hazmat-filtering and analyst-in-the-loop contributions mature alongside extraction and linkage components. Together, these steps would increase confidence in detailed interpretation and improve the utility of this pipeline for operational and research settings.

References

[1] Cisco Systems. *The Continued Evolution of DDoS Attack Vectors*. White paper. Developed in conjunction with Radware. Cisco Systems, Inc., 2021. URL: https://www.cisco.com/c/en/us/products/collateral/security/evolution-ddos-attack-vectors-wp.pdf.

[2] Richard Hummel and ASERT Team. *2024 DDoS-for-Hire Landscape Part 1*. Blog post on NETSCOUT Blog. Accessed: 2025-11-16. Dec. 2024. URL: https://www.netscout.com/blog/asert/2024-ddos-hire-landscape-part-1.

[3] José Jair Santanna et al. "Booters: An Analysis of DDoS-as-a-Service Attacks". In: *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)* (2015). DOI: 10.1109/INM.2015.7140298.

[4] Daniel Kopp et al. "DDoS Hide & Seek: On the Effectiveness of a Booter Services Takedown". In: *ACM/SIGCOMM Internet Measurement Conference* (2019). DOI: 10.1145/3355369.3355590.

[5] Anh V. Vu et al. "Assessing the Aftermath: the Effects of a Global Takedown against DDoS-for-hire Services". In: *arXiv (Cornell University)* (2025). DOI: 10.48550/ARXIV.2502.04753.

[6] Russell Brewer et al. "The Ethics of Web Crawling and Web Scraping in Cybercrime Research: Navigating Issues of Consent, Privacy, and Other Potential Harms Associated with Automated Data Collection". In: *Researching Cybercrimes: Methodologies, Ethics, and Critical Approaches*. Ed. by Anita Lavorgna and Thomas J. Holt. Palgrave Macmillan, 2021, pp. 435–456. DOI: 10.1007/978-3-030-74837-1_22.

[7] Mohammed Khalafallah Alshammery and Abbas Fadhil Aljuboori. "Crawling and Mining the Dark Web: A Survey on Existing and New Approaches". In: *Iraqi Journal of Science* 63.3 (2022), pp. 1339–1348. DOI: 10.24996/ijs.2022.63.3.36. URL: https://ijs.uobaghdad.edu.iq/index.php/eijs/article/view/4003.

[8] Mohammad Karami, Young-Sam Park, and Damon McCoy. "Stress Testing the Booters: Understanding and Undermining the Business of DDoS Services". In: (2016). DOI: 10.1145/2872427.2883004.

[9] Gulshan Kumar. "Denial of Service Attacks – An Updated Perspective". In: *Systems Science & Control Engineering* 4.1 (2016), pp. 285–294. DOI: 10.1080/21642583.2016.1241193.

[10] Mohammad Karami and Damon McCoy. "Understanding the Emerging Threat of DDoS-as-a-Service". In: *Proceedings of the 6th USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET)*. USENIX Association, 2013. URL: https://www.usenix.org/conference/leet13/workshop-program/presentation/karami.

[11] NCC Group. *Application Layer Attacks – The New DDoS Battleground*. Tech. rep. White paper. NCC Group Plc, 2014. URL: https://www.nccgroup.com/media/d4cojg4z/_the-new-ddos-battleground-white-paper-final.pdf.

[12] José Jair Santanna et al. "Booter list generation: The basis for investigating DDoS-for-hire websites". In: *Int. J. Netw. Manag.* (2018). DOI: 10.1002/NEM.2008.

[13] Benjamin Collier et al. "Cybercrime is (often) boring: maintaining the infrastructure of cybercrime economies". In: (2020). DOI: 10.17863/CAM.53769.

[14] Thomas S. Hyslip and Thomas J. Holt. "Assessing the Capacity of DRDoS-For-Hire Services in Cybercrime Markets". In: *Deviant Behavior* (2019). DOI: 10.1080/01639625.2019.1616489.

[15] Ryan Brunt. "Booted: An Analysis of a Payment Intervention on a DDoS-for-Hire Service". In: (2017).

[16] Abhineet Gupta, Sean B. Maynard, and Atif Ahmad. "The Dark Web Phenomenon: A Review and Research Agenda". In: *Proceedings of the 30th Australasian Conference on Information Systems (ACIS)*. ACIS, 2019. URL: https://aisel.aisnet.org/acis2019/1.

[17] Randa Basheer and Bassel Alkhatib. "Threats from the Dark: A Review over Dark Web Investigation Research for Cyber Threat Intelligence". In: *International Journal of Intelligent Systems and Applications in Engineering* 9.3 (2021), pp. 305–317. DOI: 10.1155/2021/1302999.

[18] The Tor Project. *Tor: Onion Service Protocol*. https://web.archive.org/web/20181223204417/https://www.torproject.org/docs/onion-services.html. Accessed 13 April 2026.

[19] Christopher Johnson et al. *Guide to Cyber Threat Information Sharing*. NIST Special Publication 800-150. National Institute of Standards and Technology, 2016. URL: https://csrc.nist.gov/publications/detail/sp/800-150/final.

[20] European Union Agency for Cybersecurity (ENISA). *ENISA Threat Landscape 2020: Cyber Threat Intelligence Overview*. Tech. rep. ENISA, 2020. URL: https://www.enisa.europa.eu/publications/cyberthreat-intelligence-overview.

[21] Paulo Shakarian. "Dark-Web Cyber Threat Intelligence: From Data to Intelligence to Prediction". In: *Information* 9.12 (2018), p. 305. DOI: 10.3390/info9120305.

[22] Sergio Pastrana et al. "CrimeBB: Enabling Cybercrime Research on Underground Forums at Scale". In: *The Web Conference* (2018). DOI: 10.1145/3178876.3186178.

[23] José Jair Santanna et al. "Booter blacklist: Unveiling DDoS-for-hire websites". In: *Conference on Network and Service Management* (2016). DOI: 10.1109/CNSM.2016.7818410.

[24] Daniel De Pascale et al. "CRATOR: A Dark Web Crawler". In: *arXiv* 2405.06356 (2024). URL: https://arxiv.org/abs/2405.06356.

[25] Cloudflare, Inc. *What Is Bot Management? How Bot Managers Work*. Accessed 13 April 2026. 2024. URL: https://www.cloudflare.com/learning/bots/what-is-bot-management/.

[26] Distil Networks. *Bad Bot Report 2018: The Year Bad Bots Went Mainstream*. https://www.distilnetworks.com/library/bad-bot-report/. White paper. 2018.

[27] Philipp Winter et al. "DarkDiff: Explainable Web Page Similarity of TOR Onion Sites". In: *Proceedings of the 2023 ACM Web Conference Companion*. Preprint arXiv:2308.12134. 2023.

[28] Gurmeet Singh Manku, Arvind Jain, and Anish Das Sarma. "Detecting Near-Duplicates for Web Crawling". In: *Proceedings of the 16th International Conference on World Wide Web*. ACM, 2007, pp. 141–150. DOI: 10.1145/1242572.1242592.

[29] Muhammad Saeed Hussain, Muhammad Hussain, and Muhammad Afzal. "Detection of Duplicate and Near-Duplicate Content for Web Crawlers". In: *Journal of Independent Studies and Research – Computing* 13.1 (2015), pp. 27–37. URL: https://jisrc.szabist.edu.pk/ojs/index.php/jisrc/article/view/126.

[30] Ashish Vaswani et al. "Attention Is All You Need". In: *Advances in Neural Information Processing Systems* 30 (2017). URL: https://arxiv.org/abs/1706.03762.

[31] Tom B. Brown, Benjamin Mann, Nick Ryder, et al. "Language Models Are Few-Shot Learners". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 1877–1901. URL: https://arxiv.org/abs/2005.14165.

[32] Yi Tay et al. "Efficient Transformers: A Survey". In: *ACM Computing Surveys* 55.6 (2022), pp. 1–28. DOI: 10.1145/3530811.

[33] Ridong Han et al. "An Empirical Study on Information Extraction using Large Language Models". In: (2024). arXiv: 2409.00369 [cs.CL]. URL: https://arxiv.org/abs/2409.00369.

[34] Jordan Friedman et al. "Small Language Models Enable Rapid and Accurate Extraction of Structured Data from Unstructured Text: An Example with Plants and Their Specialized Metabolites". In: *Patterns* (2025). In press. URL: https://www.biorxiv.org/content/10.1101/2025.05.19.654923v1.

[35] Iz Beltagy, Matthew E. Peters, and Arman Cohan. "Longformer: The Long-Document Transformer". In: *arXiv* 2004.05150 (2020). URL: https://arxiv.org/abs/2004.05150.

[36] Max Moundas, Jules White, and Douglas C. Schmidt. "Prompt Patterns for Structured Data Extraction from Unstructured Text". In: *Proceedings of the 31st Conference on Pattern Languages of Programs, People, and Practices (PLoP 2024)*. Skamania Lodge, Columbia River Gorge, WA, USA. ACM, 2025. DOI: 10.64346/PLoP2024p24. URL: https://www.plopcon.org/proceedings/plop/2024/24.html.

[37] Russell Brewer et al. "Establishing a Framework for the Ethical and Legal Use of Web Scrapers by Cybercrime and Cybersecurity Researchers: Learnings from a Systematic Review of Australian Research". In: *International Journal of Law and Information Technology* 31.3 (2023), pp. 186–212. DOI: 10.1093/ijlit/eaad023.

[38] Internet Watch Foundation. *Internet Watch Foundation Tools and Services*. https://www.kirkleessafeguardingchildren.co.uk/wp-content/uploads/2019/09/Presentation-5-Internet-Watch-Foundation-FINAL-002.pdf. Includes description of hash list, URL list, and keywords list. 2019.

[39] Thorn / Safer. *CSAM Keyword Hub*. Accessed 13 April 2026. 2024. URL: https://safer.io/resources/csam-keyword-hub/.

[40] Safer by Thorn. *Introducing the CSAM Keyword Hub: A Free Collection of Words Related to Child Safety Risks*. Blog post, accessed 13 April 2026. 2024. URL: https://safer.io/resources/introducing-the-csam-keyword-hub-a-free-collection-of-words-related-to-child-safety-risks/.

[41] LDNOOBW. *List of Dirty, Naughty, Obscene, and Otherwise Bad Words*. https://github.com/LDNOOBW/List-of-Dirty-Naughty-Obscene-and-Otherwise-Bad-Words. Open-source profanity and obscenity word list, accessed 13 April 2026. 2012.

[42] Michael Wiegand and Josef Ruppenhofer. “Inducing a Lexicon of Abusive Words – A Feature-Based Approach”. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Vol. 1. New Orleans, LA, USA: Association for Computational Linguistics, 2018, pp. 1046–1056. URL: https://www.aclweb.org/anthology/N18-1095.pdf.

[43] Christian Beek. *Personal communication on analyst-in-the-loop operations in threat intelligence*. Practitioner interview. Christian Beek is Principal Threat Intelligence Researcher at Rapid7. 2026.

A

Use of Artificial Intelligence

Artificial intelligence tools (GitHub Copilot) were used as assistant support during pipeline development and manuscript preparation. Use was limited to productivity and drafting support, not autonomous scientific decision-making.

- **Pipeline engineering support:** code cleanup and refactoring suggestions, bug review assistance, unit-test scaffolding for edge cases, and small maintenance scripts.
- **Writing support:** wording alternatives, sentence-level clarity checks, and spelling and grammar correction.
- **Literature support:** rapid first-pass screening of sources for likely topical relevance before full manual reading.

The author retained full responsibility for all technical and written content. Code suggestions were manually reviewed before integration and verified through local runs. Citations and substantive claims were validated against sources read by the author.