A Survey of Darknet Detection Methodologies: Design, Implementation, and Assessment

MAX GAO, CAIDA/UC San Diego, USA

Network telescope (darknet) traffic has been instrumental in providing visibility into Internet-wide phenomena related to security and availability such as malicious scanning, denial-of-service (DoS) backscatter, and outages. To better operationalize darknets for their monitoring capabilities, researchers have proposed a number of automated detection methodologies which span a diverse set of analytical techniques applied to different features of darknet traffic. Yet despite the abundance of methods, a comprehensive view of their comparative capabilities remains unclear due to gaps across their implementations and inconsistencies across their empirical assessments. In this report, we review the evolution of darknet event detection methodologies over the past decade and examine in-depth the challenges that face a comprehensive, systematic assessment. We conclude with a discussion of future directions to address these challenges.

1 Introduction

Network telescopes, or darknets, have been instrumental to both researchers and practitioners for their ability to observe Internet-wide phenomena in the unsolicited traffic they receive. Over the past two decades, research efforts have shifted from characterizing darknet phenomena and their observable signals towards translating such empirical insights into operationalizable methodologies for monitoring the Internet's security and availability in near real-time. These efforts have culminated in a number of methodologies that, despite sharing a common goal of automatically detecting events from darknet traffic, differ widely in the technique(s) they employ (*e.g.,* dimensionality reduction, forecasting, representation learning), features of the traffic they exploit (*e.g.,* packet inter-arrival times, destination port sequences), and their operational definition of what constitutes an event (*e.g.,* a shift in sender clusters, anomalous traffic volumes towards specific ports). As recent studies propose an increasing number of methods that leverage machine-learning techniques, systematic assessment of their capabilities are needed to enable fair yet rigorous comparisons to determine their effectiveness in-practice. However as our report finds, many of these methods have been developed, implemented, and assessed independently by different research groups without necessary standard procedures to ensure their comparability. As a result, this lack of standardization discourages the development of new and existing methods which we encountered in our prior work [44].

In this report, we survey an extensive body of literature that proposes darknet event detection methodologies to investigate the extent that the challenges we encountered impact the development and assessments of darknet-event detection methodologies at large. Overall, we find a low degree of replicability among proposed methodologies and inconsistencies across datasets and validation approaches in their assessments. Section 1 provides an overview of canonical literature that characterizes darknet traffic and its associated phenomena. We discuss notable findings from past empirical studies and their role in monitoring darknet traffic. Section 2 introduces our taxonomy of detection methodologies with a breakdown of their detected events, employed techniques, and the features and representations of their traffic inputs alongside summarized intuitions guiding their design. Section 3 reviews empirical assessments of prior work separate from methods themselves to highlight specific gaps that result from the lack of standard practices. Finally, we conclude our survey's findings with directions for future work intended to support more robust evidence-informed method selection for darknet event detection tasks by enabling systematic comparative assessments.

Author's Contact Information: Max Gao, CAIDA/UC San Diego, San Diego, CA, USA, magao@ucsd.edu.

2 Background

In this section, we provide a brief overview of network telescopes and discuss their role in Internet measurement research. We first provide a technical explanation of their function as observatories of unsolicited IPv4 Internet traffic. We then summarize findings from past works that have studied the nature and composition of this traffic during landmark Internet-wide events. Finally, we remark on changes in characteristics of such traffic up to present day and speculate on the future development of its traffic and flows.

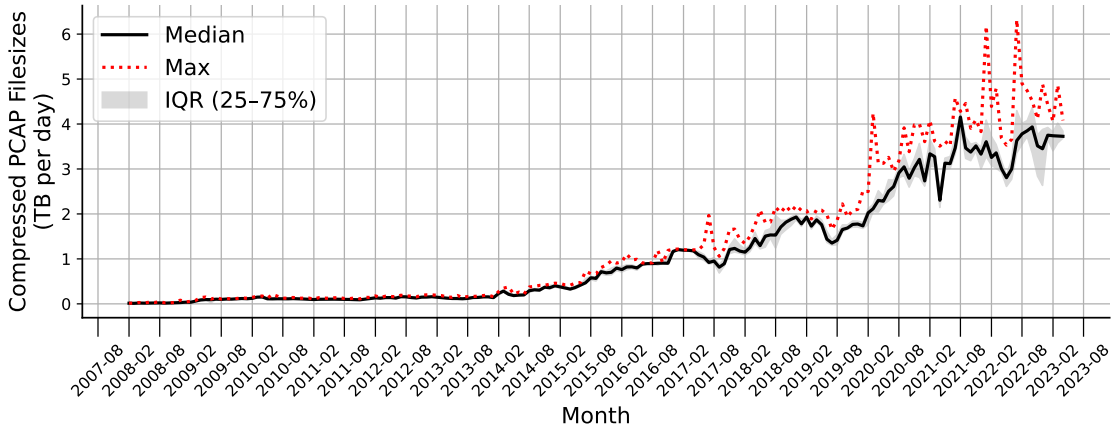

Fig. 1. Growth in volumes of darknet-traffic collected at UCSD-NT over 15 years. Daily compressed PCAP filesizes have exceeded 3.5 terabytes per day.

Network telescopes, or darknets, consist of IPv4 address space that receives, but does not respond to, unsolicited Internet traffic via routes announced through the Border Gateway Protocol (BGP). Researchers have studied this unidirectional traffic to understand its mixtures and origins using darknets as its instrumentation with foundational work attributing its causes to Internet-wide activity that includes malicious scanning [5, 15, 32, 35, 50], residual backscatter from DoS attacks [16, 51, 57, 63, 82, 96], Internet outages [13, 31, 33, 89], and non-trivial network misconfigurations [12].

Large-scale empirical studies have characterized darknet-traffic composition and its changes over the past decades. In 2004, Pang et al. [90] was the first to conduct a systematic analysis of darknet traffic[1] in terms of its activities and source, followed by a study in 2010 by Wustrow et al. [108] that found traffic volumes[2] had grown nearly four-fold along with a reversal in SYN/SYN-ACK trends in the years since. Later in 2014, following the release of high-speed scanning tools (ZMap [36] and Masscan [48]), Durumeric et al. [35] report[3] that horizontal-scans have become common with most malicious scans originating from bullet-proof hosting providers. More recently in 2024, Griffioen et al. [50] conduct a 10-year longitudinal study and find a 30-fold increase, which roughly mirrors the trend pictured in Figure 1 for UCSD-NT, in scan traffic[4] whose sources rapidly change geographic locations and port targets. Parallel to this line of work, IPv4 address space exhaustion pressures have led researchers to study darknet traffic collected at non-traditional

[1]Pang et al. used a /8, two /19, and ten /24 sized darknets.
[2]Wustrow et al. used five different /8 darknets.
[3]Durumeric et al. used a darknet roughly the size of a /9.
[4]Griffioen et al. use their /16 darknet to draw comparisons.

vantage points at IXPs [106], CDNs [94], and cloud infrastructure [92] though the characteristics of this traffic remains to be extensively compared against traditional darknet traffic.

Despite the evolution of darknet traffic and instrumentation for its collection, researchers continue to devise methods and empirically demonstrate their effectiveness at detecting the various kinds of Internet-wide activities known to occur. Our survey reviews the design, implementation, and assessments of these methods with an emphasis on those motivated by recent interest in the application of modern machine-learning based approaches to analyze darknet traffic.

3 Detection methods for darknet scanning activities

In this section, we provide a broad survey of the methods that have been proposed to detect scanning activity in darknet traffic. We construct a body of 35 works published over the past 14 years (summarized in Table 1 which consists of a majority that appear after the official release of ZMap in 2013 [36]. Our single selection criterion includes only works that provided an assessment of their proposed method using real-world darknet traffic data[5] We organize these works by the key technique(s) their detection methods employ, describe major components of their designs, and discuss their strengths and weaknesses based on empirical findings where possible. We briefly define these components before elaborating on individual works.

Detection tasks. Since each work uses different terminology to describe detection tasks of their method, we redefine 3 types of tasks in order to generalize their definitions. *Source characterization* refers to method-specific tasks that range from clustering/classification of darknet traffic sources and detection of potentially coordinated sources. *Target identification* involves identifying specific application-layer ports that anomalous traffic targets which may correspond to reconaissance activity and scans originated from malware-infected hosts. *Event detection* determines specific points in time during which dynamics of darknet traffic exhibit significant change. For example, [61, 64] detect changes in sender behavior by tracking cluster changes over time.

Technique(s) differentiate functionalities undertaken by specific **algorithms(s)** within the scope of each methodology. Across our surveyed works, we identified 7 classes of techniques which include *dimensionality reduction*, *clustering*, *forecasting*, *thresholding*, *representation learning*, *frequent pattern mining*, and *fingerprinting*. Most methodologies employ a combination of techniques (*e.g.*, dimensionality reduction as a prior step to clustering or thresholding similarity metrics computed from intermediary representations) to accomplish analysis subtasks.

Traffic features and representations. Methods consist of a preprocessing step that parses properties of raw darknet traffic into representations usable by their algorithms. Depending on the method's detection task, traffic features are defined per-source, per-target, or per-flow at different levels of aggregation and encoded as one of the four main types of representations: *graphs*, *feature vectors*, *sequences*, *time series*.

3.1 Representation learning methods

Representation learning techniques play a key role in modern machine-learning for their capability to automatically learn lower-dimensional representations of raw data for downstream modeling tasks. While their functionality overlaps with classic dimensionality reduction techniques, for our survey we distinguish these methods by their application of artificial neural network architectures that learn non-linear relationships between features of darknet traffic.

Researchers first devised methodologies that incorporated these techniques into darknet traffic analysis starting in 2020 with the Word2Vec [78] algorithm. Cohen et al. [27] proposed *Dante* to detect scanning trends in TCP ports from

[5]We made an exception for [27] which used greynet traffic in their assessment.

Table 1. Characteristics of event detection methodologies

Work	Detection Tasks			Techniques							Traffic Features	Traffic Representations			
	Source Class.	Target Id.	Event Det.	Clustering	Dim. Reduction	Forecasting	Thresholding	Fingerprinting	Rep. Learning	Pattern Mining		Graph	Feature Vector	Sequences	Time Series
2025 Abduaziz et al. [1]	✓			✓	✓				✓		Per-sIP: <*seq(dport)*>			✓	
2024 Huang et al. [61]	✓		✓	✓	✓				✓		Per-dPort: *cnt(pkt), uniq(saddr)*				✓
2024 Gao et al. [44]	✓	✓	✓				✓			✓	Per-dPort: *cnt(pkt), uniq(saddr)*				✓
2023 Kartsioukas et al. [66]											Per-dPort: *cnt(pkt), uniq(saddr)*				✓
2023 Zakroum et al. [110]	✓		✓	✓		✓									✓
2023 d'Andréa et al. [37]		✓	✓		✓		✓				Per-dPort: *cnt(pkt), uniq(saddr)*				✓
2022 Kallitsis et al. [64]	✓		✓	✓	✓		✓						✓		
2022 Zakroum et al. [109]		✓	✓	✓		✓									✓
2021 Tanaka et al. [103]	✓			✓				✓							
2021 Han et al. [53]	✓	✓	✓		✓						Per-sIP:				✓
2021 Gioacchini et al. [47]	✓			✓	✓						Per-TCP dport: *saddr*			✓	
2020 Griffioen et al. [49]	✓			✓				✓			Per-TCP dport: *saddr*				
2020 Cohen et al. [27]		✓	✓		✓						Per-sIP: *TCP dport*				✓
2020 Han et al. [52]	✓	✓	✓		✓						s16->nt: <*pktcnt*>				✓
2020 Torabi et al. [104]			✓							✓	Per-Flow: *dports, protocol, TTL, TCP flags, IP length, packet count*	✓	✓		
2020 Soro et al. [99]	✓			✓							Per-sAS,dport: *packet count*		✓		
2019 Evrard et al. [39]	✓			✓							Per-saddr: *TCP dports*	✓			
2019 Iglesias et al. [62]	✓		✓	✓				✓			Per-saddr: *22 total features*		✓		
2019 Niranjana et al. [85]	✓			✓							Per-		✓		
2019 Kanehara et al. [65]	✓	✓	✓		✓			✓			Per-sIP,dPort: *ts, pktcnt*				✓
2017 Lagraa et al. [70]	✓			✓								✓			
2016 Ban et al. [9]	✓	✓	✓		✓						*saddr, dport*			✓	✓
2015 Nishikaze et al. [86]	✓			✓							saddr16: *pktcnt, sport, dport, daddr, scantype*		✓		
2012 Fachka et al. [40]	✓									✓	saddr16: *pktcnt, sport, dport, daddr, scantype*				

clustered Word2Vec embeddings trained from *port sequences* per sender. Gioacchini et al. [47] demonstrated that using a modified embedding definition, *i.e., sender sequences* per port, *DarkVec* results in better scalability and more accurate identification of labeled scanner organizations compared to *Dante*. Follow-up works [46, 61] extend *DarkVec* to handle larger volumes of traffic and track changes in clusters across time. Abduaziz et al. [1] replicate *Dante* though with a modified semi-supervised clustering technique to highlight In 2022, Kallitsis et al. [64] demonstrated that their method which employed deep-learning, *i.e.,* by applying autoencoders to embed 12 features selected to represent senders, could more accurately cluster the same labeled senders as *DarkVec* though at a cost of greater implementation complexity and computational requirements.

3.2 Graph-based methods

These methods represent the behaviors of darknet traffic sources and the sequence of their probed targets using different graph formulations. While they leverage the same fundamental features as early Word2Vec methods, graph-mining methods in principle have lower computational requirements depending on the formulation used in analysis.

Methods proposed by Lagraa et al. [70] and Evrard et al. [39] represent sequences of scanned TCP ports as unipartite graphs (ports as nodes and consecutive scans as edges). To identify groups of similarly probed ports, both methods threshold graph metrics (shortest-path similarity in [39], node centrality in [70]). In follow-up work, Lagraa et al. [69] provide an additional graph definition to detect horizontal darknet scanners. Soro et al. [99] propose modeling interactions as weighted bipartite graphs which enables a wider scope of analysis but at the cost scaling limitations which led to mapping sender IPs to a lower spatial grain, Autonomous Systems (ASes). Application of the Greedy Modularity Algorithm resulted in different clusters of ASes that separately contained distributed scanners targeting specific ports, horizontal scanners, and potential misconfiguration-related traffic.

More recent work incorporates representation learning techniques to graph representations of darknet traffic. Using the same graph formulation as [39], d'Andréa et al. [37] apply a Graph Convolutional Network (GCN) to classify senders labeled using AbuseIPDB reports. Zakroum et al. [110] learn graph-embeddings that model packet-level changes in flows.

3.3 Time series-based methods

Time series representations of traffic provide aggregated views of darknet scanning activity and flexibility in the metrics chosen

offer an aggregated view of darknet scanning

supports analysis by a variety of techniques.

Their methods typically operate on a large space of inputs with the goal of localizing time periods that correspond to abnormal traffic activity, defined per-method. Several methods achieve this via dimensionality reduction techniques. *discuss scalability limitations –Max* In 2019, Kanehara et al. [65] applied Nonnegative Tucker Decomposition (NTD) [112] to packet count time series between senders and destination ports they contacted. Their findings from real-world traffic analysis identified groups of behaviorally-similar senders on several ports of applications containing known vulnerabilities. Han et al. proposed *DarkNMF* [53] and *DarkGLASSO* [52], which respectively apply Nonnegative Matrix Factorization (NMF) [72] and Graphical LASSO [43] to time series of packet counts per source host (for *DarkNMF*, also per destination port). Kartsioukas et al. [66] apply Incremental Principal Component Analysis (iPCA) [6] to time series of unique senders per TCP, UDP ports and ICMP.

3.4 Fingerprinting methods

Fingerprinting methods operate on information embedded in packet headers of darknet traffic. The key intuition behind the use of these methods are that distributed yet coordinated hosts likely employ the same stateless scanning tools, *e.g.*, ZMap [36], embed identifying information in the same parts of a packet which can be used to infer their fingerprints. Griffioen et al. [49] first proposed this technique which enabled correlation of similar hosts suspected to use the same scanning tools. Tanaka et al. [102, 103] extend Griffioen et al.'s method to generate candidate fingerprints from a larger set of header fields using a genetic algorithm.

Table 2. Details of method assessments found in our surveyed works. Multiple citations per entry indicate groups of highly similar works; assessment details reflect bolded citations. Table 3 lists additional details of telescopes referenced in this table.

	Replicability			Dataset Attributes					Compar-	Labels
Work	**Code**	**Specs**	**Data**	**Telescope(s)**	**Duration**	**Year**	**Packets**	**Bytes**	**ison**	**Labels**
Abduaziz et al. [1]		✓		NT-1,3	7D,10D	2022,2023	—	1.18, 7.61GB	[27]	✓
Huang et al. [61]				NT-4	50D	2021	—	—		✓
Gao et al. [44]				NT-1	18M	2022	—	—	[52]	
Kartsioukas et al. [66]				NT-2	1M	2016	—	—	[71]	
Zakroum et al. [110]				NT-3,6	4.5Y	2018	—	—	[110]	✓
d'Andréa et al. [37]		✓		Private /20		2021	—	—		✓
Han et al. [54]	✓	✓	✓	NT-3	1M	2018	—	—	[52, 65, 101]	✓
Kallitsis et al. [64]	✓	✓		NT-2	28D, 1D	2016,2022	49B, 3.1B	—	[47]	✓
Zakroum et al. **[109]** [111]		✓		NT-3,6	3.5Y	2017	—	1.5 + TB	[111]	
Tanaka et al. [103]		✓	✓	NT-3	1M	2018	117M	—	[52, 65, 101]	✓
Han et al. **[53]** [54]	✓	✓	✓	NT-3	1M	2018	—	—	[52, 65, 101]	✓
Gioacchini et al. **[47]** [46]	✓	✓	✓	NT-4	30D	2021	63M	—	[27, 95]	✓
Griffioen et al. **[49]**		✓		Private	2M		6.5B	864G	[9]	
Cohen et al. **[27]**		✓		NT-4	1Y		7.9B		[9]	
Han et al. **[52]** [54]	✓	✓	✓	NT-3	1M	2018	—	—	[101]	✓
Torabi et al. [104]	✓	✓	—				—	—		
Soro et al. [99]				NT-4,5	3W,1D	2020	—	—		
Evrard et al. [39]			—	NT-3,6	9M	2015	8M	—		
Iglesias et al. [62]		✓		NT-1	6M	2012	—	2.1 TB		
Niranjana et al. [85]				Private /24	20D	2017		—		
Kanehara et al. **[65]** [54]		✓		NT-3	10M	2017	—	—		
Lagraa et al. **[69]** [70]		✓		NT-6	2Y	2014	2B	500 GB		
Bou-Harb et al. **[20]** [19]		✓	—	NT-1,7	1M,1M	2016,2014	—	670, 240GB	[18]	
Ban et al. [10]			—	NT-3	—	—	—	—	[11]	
Ban et al. [9]			—	NT-3	1y	2015	3×10^7	—		
Nishikaze et al. [86]				NT-3	28D	2014	303M	—		
Bou-Harb et al. [21]			—	NT-7	2D	2013	10^6	30GB		
Fachka et al. [40]				NT-7	2D	2011	—	—		
Aggregate	4/16	7/16	—	8–24	1w–3.5y	2012–2021	—	—	7/16	—

Table 3. Summary of network telescopes referenced in surveyed works.

Telescopes	Country	Size	Data availability
NT-1. UCSD-NT [24]	US	/9+/10	Raw traces and flow data
NT-2. Merit ORION [77]	US	~7 ×/16s	Raw trace, custom-defined event data
NT-3. NICTER [81]	JP	/17, /18, 2×/20	TCP SYN only, anonymized flow data
NT-4. Politecnico di Torino [99]	IT	3 × /24	Unknown
NT-5. Darknet-BR [30]	BR	/19	Unknown
NT-6. LHS Nancy [68]	FR	/20	Raw trace
NT-7. Farsight [97]	US	/13	Private

3.5 Pattern-mining methods

4 Assessments of detection methodologies

To demonstrate the utility of their detection methodologies, prior works conduct assessments using real-world darknet traffic datasets. These assessments and their results provide evidence of a method's performance under realistic settings

and offer empirical insights into its detection capabilities and general scalability. In this section, we characterize prior works along the lines of credibility and reproducibility of their assessments by reviewing: (i) characteristics of the darknet traffic datasets they use; and (ii) their approach to validating detection outputs of their methodology. Furthermore, we consider implementation details of detection methodologies since the reproducibility of assessments and results are closely tied to the availability of method source code and specifications of chosen computational environments.

4.1 Characteristics of darknet traffic datasets.

We consider three characteristics of each dataset used in prior work to assess detection methodologies. These include 1) the timeframe covered by the dataset, 2) the darknet whose traffic comprises the dataset, 3) and the dataset's traffic volume measured in units of packets and bytes.

Across our surveyed works, we identified a total of 7 darknets (listed in Table 3) from which datasets source their traffic. The amount of IPv4 address space occupied by each darknet varies; the largest and smallest sizes respectively span roughly a /9 (approx. 12.5M addresses in NT-1) and three /24 networks (768 addresses in NT-4). While these numbers represent the most up-to-date sizes reported by operators, darknet address space in practice may fluctuate based on ad-hoc BGP announcements (e.g., for leasing purposes [76] or in the case of more permanent changes such as address ownership transfers [7]). Empirical studies such as [30, 100] show that the size and ranges of IP addresses occupied by a darknet have a non-trivial impact on the observability, i.e., in terms of traffic composition and source visibility) of small-scale scanning events. Thus, assessments results may not generalize across datasets sourced from different darknets.

We visualize the combined range of timeframes covered by darknet-traffic datasets from the start of 2012 until late 2023 in Figure 2. The proportion of time covered over this range by at least a single dataset (at a daily granularity) sits at 68%. Most datasets are short snapshots: over half of all datasets span fewer than two months, roughly a third cover under three years of darknet traffic, and less than a fifth cover three or more years. Since a majority of works do not provide clear rationale for their timeframe selection, we infer that selection is largely opportunistic and based on availability of darknet traffic data at the time of assessment.

Some exceptions [64, 66] use known dates of landmark Internet-wide events, *e.g.*, launch of the Mirai botnet, to guide their selection of timeframes. In addition to Mirai, Figure 2 plots the start date of 17 additional events (further summarized in Table 5) identified as either major remote execution vulnerabilities or scans originated by botnets. While the traffic associated with these events, in theory, should have been observed by darknets, besides [64, 66] there have been no explicit studies of their detectability in darknet traffic despite the fact that all but one of the events are covered by the timeframes of datasets used in assessments. We further discuss the implication of this finding in Section 5.

Takeaways: Datasets used throughout method assessments possess different levels of observability in their traffic due to the variety of darknets and timeframes they sample. Generalizability of a specific method's performance on one dataset to another remains unclear.

4.2 Strategies for validating detection results

Since limited definitive "ground truth" exists for attributing darknet traffic to their root causes, prior works rely on improvised strategies to validate their method outputs. The defining practices of these strategies include whether they validate method outputs against labeled data, whether they incorporate external datasets (as used in their label

Table 4. Historic landmark events either purported or confirmed to have been observed by darknets.

ID	Date	Event Name	Type	ID	Date	Event Name	Type
I	2012-04	Carna	Botnet	X	2017-10	Reaper	Botnet
II	2014-03	Heartbleed	Remote Exp.	XI	2018-05	VPNFilter	Botnet
III	2014-09	Shellshock	Remote Exp.	XII	2018-06	Crackonosh	Botnet
IV	2016-08	Mirai	Botnet	XIII	2019-05	BlueKeep	Remote Exp.
V	2017-04	Eternal Blue	Remote Exp.	XIV	2020-06	Ripple20	Remote Exp.
VI	2017-05	WannaCry	Botnet	XV	2020-03	SMBGhost	Remote Exp.
VII	2017-04	BrickerBot	Botnet	XVI	2021-12	Log4JShell	Remote Exp.
VIII	2016-03	Amnesia	Botnet	XVII	2021-06	PrintNightmare	Remote Exp.
IX	2017-06	NotPetya	Botnet				

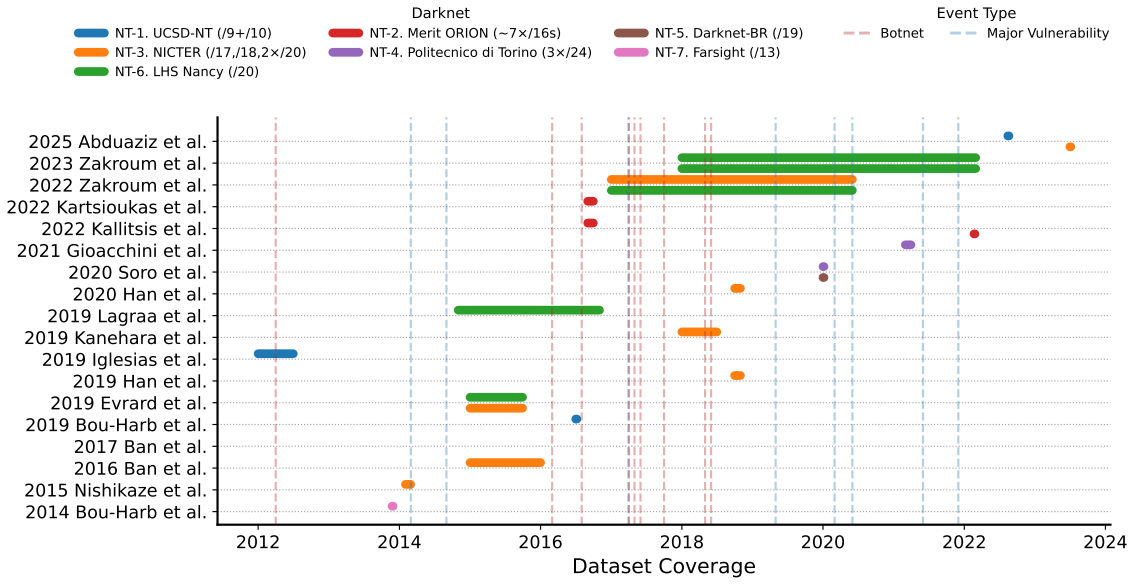

Fig. 2. Timeframes and sources of darknet-traffic datasets used in assessments (Section 3.5) of detection methodologies, overlapped with dates of Internet-wide events (listed in Table 4) observable by darknets.

definitions or more generally to cross-validate results), and whether outputs from baseline methods are used as a reference point for comparison. We review these components, indicated in Table 2 for each work's validation strategy and consider the soundness and robustness of the overall approach.

Among our surveyed works, 7 [9, 39, 52, 62, 65, 66, 86] employ strategies designed to either confirm the results of exploratory analyses or verify that lack practices such as use of labeled data, cross-validation with external datasets, and comparisons against baseline methods. designed to confirm the results of exploratory analyses. A majority of these strategies lack the use of labeled data, cross-validation with external datasets, and comparisons against baseline methods. Furthermore, their scope is limited in-practice by the volume of method outputs and the amount of manual investigation efforts (*e.g.,* application of domain knowledge to interpret network traffic behavior, correlating outputs with third-party reports) required for validation. Most works [9, 39, 52, 62, 65, 66, 86] investigate a limited number of

outputs, linked to targeted ports of botnet scans [14] and remotely exploitable vulnerabilities [4, 8, 107]. More extensive validation is conducted by [52, 62], respectively classifying 1,634 alerts and inspecting 20 traffic clusters.

Validation strategies employed by the remainder of works [1, 10, 20, 21, 53, 64, 69, 99, 109**?** , 110] adopt practices that are more sound compared to those previously discussed. These practices enlarge the scope of validation efforts and strengthen the credibility of results through cross-validation with external datasets, (*e.g.*, national vulnerability disclosures [87], passively-collected domain names [97]), and comparisons against baseline methodologies. Of these, we highlight several works [1, 20, 54, 64, 110] specifically for the rigor of their strategies that combine all three practices. Label definitions for traffic are explicit and well-defined (*e.g.*, malware fingerprints [26] and publicly known addresses of research projects and search engines).

Takeaways: Improvised strategies that prior works use to validate the results of method assessments vary by the scope and soundness of their practices. Nonetheless, we found a small subset of works whose strategies serve as a standard for the degree of rigor that validation strategies should aim towards.

4.3 Replicability of assessments

We consider the replicability of method assessments based on whether prior works provide public access to: 1) the datasets used in experiments; and 2) implementation of methods (as source code or their software artifacts) along with specifications of the computing environment used to run experiments. These two elements are essential for reproducing published results and applying existing methods to new traffic datasets. We consider a methodology replicable only when both conditions are satisified.

Less than a third of surveyed works provide source code of their implementations and roughly half detail specifications of their experimental computing environments. Of readily-available source code, we found Python and R as the programming languages of choice. Both are popular among numeric and scientific communities given their ease-of-use for implementing analytics workflows and abundant open-source algorithmic libraries. Other works such as [21, 69] rely on Java or C, which have become less popular as preferences shift towards higher-languages for analytics use-cases.

Environments used to run experiments range from personal laptop workstations [1, 69, 70] to small-scale research compute clusters [20, 54, 109]. Cluster sizes do not exceed five servers, individually equipped with at most 256GB of memory. Most implementations execute strictly on CPUs while a subset of works that leverage representation-learning techniques employ GPUs for model training and evaluation. Across our surveyed works, release dates of the CPUs and GPUs trail publication dates by as much as 10 years.

Takeaways: Fewer than a third of the proposed detection methodologies qualify as replicable under our two criteria. However, these that do are implemented using widely-adopted programming languages and leverage accessible hardware for experiment execution.

5 Challenges to a comprehensive assessment of methods

Our survey of the current assessments of existing methods revealed inconsistencies across their datasets, implementations, and strategies used to validate their results. In this section, we elaborate on these limitations and discuss how they hinder comprehensive assessment.

Limited replicability of detection methodologies. Despite our findings reported in Section 4.2, we found low reusability for source code released by the small fraction of works. Our own atttempts to execute this code resulted in failures, remediated only with substantial modifications. While the primary goal of these studies is to demonstrate and disseminate research ideas rather than deliver production-grade software, the quality of the code introduces friction for anyone attempting to reproduce results. Cumulatively, this effect produces a replicability gap that to address, requires attention and thought to the quality of released software.

Few direct comparative assessments. A second constraint arises from the scarcity of direct comparisons between detection methodologies. In the ideal case, competing approaches would be evaluated on a shared dataset, enabling clear, controlled comparisons of their relative capabilities. Yet fewer than half of the surveyed works perform such within-dataset comparisons, and only a handful evaluate multiple detection frameworks side-by-side using the same traffic. As a result, the literature offers limited evidence about how methods relate to each other empirically. Instead, most evaluations are siloed within individual studies, making it difficult to reason about comparative performance, robustness, or applicability across deployment contexts. In several instances, no more than two frameworks have ever been demonstrated on the same underlying dataset, leaving open questions about how methodological differences translate into practical differences in detection outcomes.

Non-overlapping datasets used for experimentation. The challenge of comparison is further compounded by the heterogeneity of datasets used across studies. There is very little overlap in the traffic traces employed to evaluate frameworks: timeframes rarely coincide, traffic volumes often differ by multiple orders of magnitude, and data originate from distinct darknet deployments, network telescope sizes, and geographic vantage points. These non-overlapping datasets obscure whether observed performance reflects properties of the methodology or idiosyncrasies of the underlying traffic. Even when studies aim to detect similar classes of events, variation in temporal scope, probe density, and background traffic composition complicates any meaningful cross-paper interpretation. Collectively, this fragmentation of datasets reinforces the difficulty of establishing a unified baseline for evaluating darknet-based detection methods and highlights the need for shared, consistently curated datasets to support future methodological comparison.

6 Future Work

We conclude this report by describing several directions for future work to overcome the barriers that hinder a complete comparative assessment of detection methods.

Curated darknet traffic datasets. One way to lower the barrier to conducting such assessments is to improve the availability of curated reference datasets. Such curation involves intentional selection of darknet traffic and enrichment using well-defined labels to assess method performance. These label definitions may include the malicious senders reported by IP blocklists [2, 41], known scanning organizations [28], and packet fingerprints tied to scanning tools [50] or malware [5].

Consistent method implementations. Consistency across the software used to implement methods and the execution environments that run assessments ensures the validity of performance metrics and improves interpretability of results. While in-practice this is difficult to achieve when multiple parties reproduce assessments, more thorough documentation (*e.g.,* software libraries, CPUs, GPUs, memory, IO, and storage hardware components) can substitute for up to a degree of inconsistency.

Table 5. Algorithms used by each surveyed framework.

Work	Algorithm(s)
Evrard et al. [39]	Dijkstra's [34]; K-NN [29, 42]
Lagraa et al. [69, 70]	Louvain [17, 83]
Kallitsis et al. [64]	Autoencoder Dimensionality Reduction [58]; K-Means [75]
Iglesias et al. [62]	K-Medoids [91]; Fuzzy-Gustafson [67]; MAD-Thresholding [73]
Nishikaze et al. [86]	Hierarchical Clustering [79]
Soro et al. [99]	Louvain Algorithm [17, 83]
Gioacchini et al. [46, 47]	Word2Vec [78]; K-Means [75]; K-NN [29, 42]; Louvain [17, 83]
Abduaziz et al. [1]	Word2Vec [1]; HDBScan [25, 45]
Han et al. [53, 54]	NMF [72]
Han et al. [52, 54]	GLASSO [43]
Kanehara et al. [54, 65]	LRA-NTD [112]; FTSD [23]; Otsu-Thresholding [88]
Kartsioukas et al. [66]	Incremental PCA [6]
Ban et al. [9]	Frequent Pattern Mining [55, 56]; Hierarchical Clustering [79]
Torabi et al. [104, 105]	Association Rule Mining [3]; DBSCAN [38]
Tanaka et al. [102, 103]	*TODO*
Niranjana et al. [85]	PCA [60, 93]
Cabana et al. [22]	Conduction Detection Algorithm [74]; Fastcluster [80]
Shaikh et al. [98]	AdaBoost [**?**]; Gradient Boosting [**?**]; Random Forest [**?**]
Zakroum et al. [109, 111]	Spectral Clustering [84]; LSTM [59]

Standardized validation strategies. Enabled by labeled reference datasets and consistent method implementations, standard metrics to assess detection and computational performance are components of a standardized approach crucial to interpretation of assessment results.

References

[1] Kayumov Abduaziz, Chansu Han, and Ji Sun Shin. Semi-supervised traceability analysis of investigative scanners of darknet traffic. *Computers & Security*, 2025.

[2] AbuseIPDB. Abuseipdb, 2025.

[3] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, 1993.

[4] Akamai Security Team. Memcached udp reflection attacks: A new era for ddos, 2018.

[5] Manos Antonakakis, Tim April, Michael Bailey, Matt Bernhard, Elie Bursztein, Jaime Cochran, Zakir Durumeric, J. Alex Halderman, Luca Invernizzi, Michalis Kallitsis, Deepak Kumar, Chaz Lever, Zane Ma, Joshua Mason, Damian Menscher, Chad Seaman, Nick Sullivan, Kurt Thomas, and Yi Zhou. Understanding the mirai botnet. In *26th USENIX Security Symposium (USENIX Security 17)*, 2017.

[6] Raman Arora, Andrew Cotter, Karen Livescu, and Nathan Srebro. Stochastic optimization for pca and pls. In *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2012.

[7] Attorney Registration & Disciplinary Commission of the Supreme Court of Illinois. Courtesy notice to AG, 2019.

[8] Nadav Avital. Rediswannamine unveiled: New cryptojacking attack powered by redis and nsa exploits, 2018.

[9] Tao Ban, Shaoning Pang, Masashi Eto, Daisuke Inoue, Koji Nakao, and Runhe Huang. Towards early detection of novel attack patterns through the lens of a large-scale darknet. In *2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld)*, 2016.

[10] Tao Ban, Lei Zhu, Jumpei Shimamura, Shaoning Pang, Daisuke Inoue, and Koji Nakao. Detection of botnet activities through the lens of a large-scale darknet. In *Neural Information Processing*, 2017.

[11] Tao Ban, Lei Zhu, Junpei Shimamura, Shaoning Pang, Daisuke Inoue, and Koji Nakao. Behavior analysis of long-term cyber attacks in the darknet. In *Neural Information Processing*, 2012.

[12] K Benson, A Dainotti, k claffy, A Snoeren, and M Kallitsis. Leveraging Internet Background Radiation for Opportunistic Network Analysis. In *ACM Internet Measurement Conference (IMC)*, 2015.

[13] Karyn Benson, Alberto Dainotti, KC Claffy, and Emile Aben. Gaining insight into as-level outages through analysis of internet background radiation. In *2013 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2013.

[14] Olivier Bilodeau and Thomas Dupuy. Dissecting linux/moose, 2015.

[15] Zachary S. Bischof, Kennedy Pitcher, Esteban Carisimo, Amanda Meng, Rafael Bezerra Nunes, Ramakrishna Padmanabhan, Margaret E. Roberts, Alex C. Snoeren, and Alberto Dainotti. Destination unreachable: Characterizing internet outages and shutdowns. In *Proceedings of the ACM SIGCOMM 2023 Conference*, 2023.

[16] Norbert Blenn, Vincent Ghiëtte, and Christian Doerr. Quantifying the spectrum of denial-of-service attacks through internet backscatter. In *Proceedings of the 12th International Conference on Availability, Reliability and Security*, 2017.

[17] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.

[18] Elias Bou-Harb, Chadi Assi, and Mourad Debbabi. Csc-detector: A system to infer large-scale probing campaigns. *IEEE Transactions on Dependable and Secure Computing*, 15(3), 2018.

[19] Elias Bou-Harb, Mourad Debbabi, and Chadi Assi. A time series approach for inferring orchestrated probing campaigns by analyzing darknet traffic. In *2015 10th International Conference on Availability, Reliability and Security*, 2015.

[20] Elias Bou-Harb, Martin Husák, Mourad Debbabi, and Chadi Assi. Big data sanitization and cyber situational awareness: A network telescope perspective. *IEEE Transactions on Big Data*, 5(4), 2019.

[21] Elias Bou-Harb, Nour-Eddine Lakhdari, Hamad Binsalleeh, and Mourad Debbabi. Multidimensional investigation of source port 0 probing. *Digital Investigation*, 11, 2014.

[22] Olivier Cabana, Amr M. Youssef, Mourad Debbabi, Bernard Lebel, Marthe Kassouf, and Basile L. Agba. Detecting, fingerprinting and tracking reconnaissance campaigns targeting industrial control systems. In *Detection of Intrusions and Malware, and Vulnerability Assessment*, 2019.

[23] Cesar F. Caiafa and Andrzej Cichocki. Generalizing the column–row matrix decomposition to multi-way arrays. *Linear Algebra and its Applications*, 433(3):557–573, 2010.

[24] CAIDA. The ucsd network telescope, 2025.

[25] Ricardo J. G. B. Campello, Davoud Moulavi, and Joerg Sander. Density-based clustering based on hierarchical density estimates. In *Advances in Knowledge Discovery and Data Mining*, 2013.

[26] João Marcelo Ceron, Klaus Steding-Jessen, Cristine Hoepers, Lisandro Zambenedetti Granville, and Cíntia Borges Margi. Improving iot botnet investigation using an adaptive network layer. *Sensors*, 19(3), 2019.

[27] Dvir Cohen, Yisroel Mirsky, Manuel Kamp, Tobias Martin, Yuval Elovici, Rami Puzis, and Asaf Shabtai. Dante: A framework for mining and monitoring darknet traffic. In *Computer Security – ESORICS 2020: 25th European Symposium on Research in Computer Security, ESORICS 2020, Guildford, UK, September 14–18, 2020, Proceedings, Part I*, 2020.

[28] M. Patrick Collins, Alefiya Hussain, and Stephen Schwab. Identifying and differentiating acknowledged scanners in network traffic. In *2023 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 2023.

[29] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.

[30] Arthur Vinícius Cunha Camargo, Lisandro Granville, and Leandro M. Bertholdo. Beyond size: Investigating the impact of scaled-down network telescopes on threat detection. *International Journal of Network Management*, 35(3):e70014, 2025.

[31] Alberto Dainotti, Roman Amman, Emile Aben, and Kimberly C. Claffy. Extracting benefit from harm: using malware pollution to analyze the impact of political and geophysical events on the internet. *SIGCOMM Comput. Commun. Rev.*, 42(1), 2012.

[32] Alberto Dainotti, Alistair King, Kimberly Claffy, Ferdinando Papale, and Antonio Pescapé. Analysis of a "/0" stealth scan from a botnet. *IEEE/ACM Transactions on Networking*, 2015.

[33] Alberto Dainotti, Claudio Squarcella, Emile Aben, Kimberly C. Claffy, Marco Chiesa, Michele Russo, and Antonio Pescapé. Analysis of country-wide internet outages caused by censorship. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, 2011.

[34] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numer. Math.*, 1(1):269–271, 1959.

[35] Zakir Durumeric, Michael Bailey, and J. Alex Halderman. An internet-wide view of internet-wide scanning. In *Proceedings of the 23rd USENIX Conference on Security Symposium*, 2014.

[36] Zakir Durumeric, Eric Wustrow, and J. Alex Halderman. Zmap: fast internet-wide scanning and its security applications. In *Proceedings of the 22nd USENIX Conference on Security*, 2013.

[37] Enzo d'Andréa, Jérôme François, Olivier Festor, and Mehdi Zakroum. Multi-label classification of hosts observed through a darknet. In *NOMS 2023-2023 IEEE/IFIP Network Operations and Management Symposium*, 2023.

[38] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 1996.

[39] Laurent Evrard, Jérôme François, and Jean-Noel Colin. Attacker behavior-based metric for security monitoring applied to darknet analysis. In *2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, 2019.

[40] Claude Fachkha, Elias Bou-Harb, Amine Boukhtouta, Son Dinh, Farkhund Iqbal, and Mourad Debbabi. Investigating the dark cyberspace: Profiling, threat-based analysis and correlation. In *2012 7th International Conference on Risks and Security of Internet and Systems (CRiSIS)*, 2012.

[41] FireHOL. All cybercrime ip feeds, 2025.

[42] Evelyn Fix and J. L. Hodges. Discriminatory analysis. nonparametric discrimination: Consistency properties. *International Statistical Review / Revue Internationale de Statistique*, 57(3):238–247, 1989.

[43] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2007.

[44] Max Gao, Ricky Mok, Esteban Carisimo, Eric Li, Shubham Kulkarni, and kc claffy. Darksim: A similarity-based time-series analytic framework for darknet traffic. In *Proceedings of the 2024 ACM on Internet Measurement Conference*, 2024.

[45] Jadson Castro Gertrudes, Arthur Zimek, Jörg Sander, and Ricardo J. G. B. Campello. A unified framework of density-based clustering for semi-supervised classification. In *Proceedings of the 30th International Conference on Scientific and Statistical Database Management*, 2018.

[46] Luca Gioacchini, Luca Vassio, Marco Mellia, Idilio Drago, Zied Ben Houidi, and Dario Rossi. i-darkvec: Incremental embeddings for darknet traffic analysis. *ACM Trans. Internet Technol.*, 23(3), 2023.

[47] Gioacchini, L. et al. DarkVec: automatic analysis of darknet traffic with word embeddings. In *Proc. ACM CoNEXT*, 2021.

[48] Robert Graham. MASSCAN: Mass IP port scanner. https://github.com/robertdavidgraham/masscan, 2021.

[49] Harm Griffioen and Christian Doerr. Discovering collaboration: Unveiling slow, distributed scanners based on common header field patterns. In *NOMS 2020 - 2020 IEEE/IFIP Network Operations and Management Symposium*, 2020.

[50] Harm Griffioen, Georgios Koursiounis, Georgios Smaragdakis, and Christian Doerr. Have you syn me? characterizing ten years of internet scanning. In *Proceedings of the 2024 ACM on Internet Measurement Conference*, 2024.

[51] Harm Griffioen, Kris Oosthoek, Paul van der Knaap, and Christian Doerr. Scan, test, execute: Adversarial tactics in amplification ddos attacks. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, 2021.

[52] Chansu Han, Junpei Shimamura, Takeshi Takahashi, Daisuke Inoue, Jun'ichi Takeuchi, and Koji Nakao. Real-time detection of global cyberthreat based on darknet by estimating anomalous synchronization using graphical lasso. *IEICE TRANSACTIONS on Information*, E103-D(10):2113–2124, 2020.

[53] Chansu Han, Jun'ichi Takeuchi, Takeshi Takahashi, and Daisuke Inoue. Automated detection of malware activities using nonnegative matrix factorization. In *2021 IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2021.

[54] Chansu Han, Jun'ichi Takeuchi, Takeshi Takahashi, and Daisuke Inoue. Dark-tracer: Early detection framework for malware activity based on anomalous spatiotemporal patterns. *IEEE Access*, 10, 2022.

[55] Jiawei Han, Hong Cheng, Dong Xin, and Xifeng Yan. Frequent pattern mining: current status and future directions. *Data Min. Knowl. Discov.*, 15(1), 2007.

[56] Jiawei Han, Jian Pei, and Yiwen Yin. Mining frequent patterns without candidate generation. *SIGMOD Rec.*, 29(2):1–12, 2000.

[57] Raphael Hiesgen, Marcin Nawrocki, Marinho Barcellos, Daniel Kopp, Oliver Hohlfeld, Echo Chan, Roland Dobbins, Christian Doerr, Christian Rossow, Daniel R. Thomas, Mattijs Jonker, Ricky Mok, Xiapu Luo, John Kristoff, Thomas C. Schmidt, Matthias Wählisch, and kc claffy. The age of ddoscovery: An empirical comparison of industry and academic ddos assessments. In *Proceedings of the 2024 ACM on Internet Measurement Conference*, 2024.

[58] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[59] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

[60] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441, 1933.

[61] Kai Huang, Luca Gioacchini, Marco Mellia, and Luca Vassio. Dynamic cluster analysis to detect and track novelty in network telescopes. In *2024 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 2024.

[62] Félix Iglesias and Tanja Zseby. Pattern discovery in internet background radiation. *IEEE Transactions on Big Data*, 5(4):467–480, 2019.

[63] Mattijs Jonker, Alistair King, Johannes Krupp, Christian Rossow, Anna Sperotto, and Alberto Dainotti. Millions of targets under attack: a macroscopic characterization of the dos ecosystem. In *Proceedings of the 2017 Internet Measurement Conference*, 2017.

[64] Michalis Kallitsis, Rupesh Prajapati, Vasant Honavar, Dinghao Wu, and John Yen. Detecting and interpreting changes in scanning behavior in large network telescopes. *IEEE Transactions on Information Forensics and Security*, 17, 2022.

[65] Hideaki Kanehara, Yuma Murakami, Jumpei Shimamura, Takeshi Takahashi, Daisuke Inoue, and Noboru Murata. Real-time botnet detection using nonnegative tucker decomposition. In *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, 2019.

[66] Rafail Kartsioukas, Rajat Tandon, Zheng Gao, Jelena Mirkovic, Michalis Kallitsis, and Stilian Stoev. Detection of sparse anomalies in high-dimensional network telescope signals. 2023.

[67] R. Krishnapuram and Jongwoo Kim. A note on the gustafson-kessel and adaptive fuzzy clustering algorithms. *IEEE Transactions on Fuzzy Systems*, 7(4):453–461, 1999.

[68] Laboratoire de Haute Sécurité. Darknet - Laboratoire de Haute Sécurité, 2024.

[69] Sofiane Lagraa, Yutian Chen, and Jérôme François. Deep mining port scans from darknet. *International Journal of Network Management*, 29(3), 2019.

[70] Sofiane Lagraa and Jérome François. Knowledge discovery of port scans from darknet. In *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, 2017.

[71] Anukool Lakhina, Mark Crovella, and Christiphe Diot. Characterization of network-wide anomalies in traffic flows. In *Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement*, 2004.

[72] Daniel D. Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. In *Proceedings of the 14th International Conference on Neural Information Processing Systems*, 2000.

[73] Hancong Liu, Sirish Shah, and Wei Jiang. On-line outlier detection and data cleaning. *Computers & Chemical Engineering*, 28(9):1635–1647, 2004.

[74] Zongqing Lu, Xiao Sun, Yonggang Wen, Guohong Cao, and Thomas La Porta. Algorithms and applications for community detection in weighted networks. *IEEE Transactions on Parallel and Distributed Systems*, 26(11):2916–2926, 2015.

[75] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66), Vol. I: Statistics*, pages 281–297. Univ. California Press, Berkeley, CA, 1967.

[76] Alexander Männel, Jonas Mücke, K. C. Claffy, Max Gao, Ricky K. P. Mok, Marcin Nawrocki, Thomas C. Schmidt, and Matthias Wählisch. Lessons learned from operating a large network telescope. In *Proceedings of the ACM SIGCOMM 2025 Conference*, 2025.

[77] Merit Network, Inc.,. ORION: Observatory for Cyber-Risk Insights and Outages of Networks, 2025.

[78] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. 2013.

[79] Fionn Murtagh and Pedro Contreras. Algorithms for hierarchical clustering: an overview. *WIREs Data Mining and Knowledge Discovery*, 2(1):86–97, 2012.

[80] Daniel Müllner. fastcluster: Fast hierarchical, agglomerative clustering routines for r and python. *Journal of Statistical Software*, 53(9):1–18, 2013.

[81] National Institute of Information and Communications Technology. Nicterweb - darknet observation, 2025.

[82] Marcin Nawrocki, John Kristoff, Raphael Hiesgen, Chris Kanich, Thomas C. Schmidt, and Matthias Wählisch. Sok: A data-driven view on methods to detect reflective amplification ddos attacks using honeypots. In *2023 IEEE 8th European Symposium on Security and Privacy (EuroS&P)*, 2023.

[83] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.

[84] Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. On spectral clustering: analysis and an algorithm. In *Proceedings of the 15th International Conference on Neural Information Processing Systems: Natural and Synthetic*, 2001.

[85] R. Niranjana, V. Anil Kumar, and Shina Sheen. Darknet traffic analysis and classification using numerical agm and mean shift clustering algorithm. *SN Comput. Sci.*, 1(1), 2019.

[86] Hironori Nishikaze, Seiichi Ozawa, Jun Kitazono, Tao Ban, Junji Nakazato, and Jumpei Shimamura. Large-scale monitoring for cyber attacks by using cluster information on darknet traffic features. *Procedia Computer Science*, 2015.

[87] National Institute of Standards and Technology (NIST). National vulnerability database (nvd), 2025.

[88] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

[89] Ramakrishna Padmanabhan, Arturo Filastò, Maria Xynou, Ram Sundara Raman, Kennedy Middleton, Mingwei Zhang, Doug Madory, Molly Roberts, and Alberto Dainotti. A multi-perspective view of internet censorship in myanmar. In *Proceedings of the ACM SIGCOMM 2021 Workshop on Free and Open Communications on the Internet*, 2021.

[90] Ruoming Pang, Vinod Yegneswaran, Paul Barford, Vern Paxson, and Larry Peterson. Characteristics of internet background radiation. In *Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement*, 2004.

[91] Hae-Sang Park and Chi-Hyuck Jun. A simple and fast algorithm for k-medoids clustering. *Expert Systems with Applications*, 36(2, Part 2):3336–3341, 2009.

[92] Eric Pauley, Paul Barford, and Patrick McDaniel. Dscope: a cloud-native internet telescope. In *Proceedings of the 32nd USENIX Conference on Security Symposium*, SEC '23, USA, 2023. USENIX Association.

[93] Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

[94] P Richter and A Berger. Scanning the Scanners: Sensing the Internet from a MassivelyDistributed Network Telescope. In *IMC 19: Proceedings of the Internet Measurement Conference*, page 14 pages, October 2019.

[95] Markus Ring, Alexander Dallmann, Dieter Landes, and Andreas Hotho. Ip2vec: Learning similarities between ip addresses. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2017.

[96] Christian Rossow. Amplification hell: Revisiting network protocols for ddos abuse. In *Proceedings of the 12th International Conference on Availability, Reliability and Security*, 2014.

[97] Farsight Security. Farsight security, 2025.

[98] Farooq Shaikh, Elias Bou-Harb, Jorge Crichigno, and Nasir Ghani. A machine learning model for classifying unsolicited iot devices by observing network telescopes. In *2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC)*, 2018.

[99] Francesca Soro, Mauro Allegretta, Marco Mellia, Idilio Drago, and Leandro M. Bertholdo. Sensing the noise: Uncovering communities in darknet traffic. In *2020 Mediterranean Communication and Computer Networking Conference (MedComNet)*, pages 1–8, 2020.

[100] Francesca Soro, Idilio Drago, Martino Trevisan, Marco Mellia, João Ceron, and José J. Santanna. Are darknets all the same? on darknet visibility for security monitoring. In *2019 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN)*, 2019.

[101] J. Takeuchi and K. Yamanishi. A unifying framework for detecting outliers and change points from time series. *IEEE Transactions on Knowledge and Data Engineering*, 18(4), 2006.

[102] Akira Tanaka, Chansu Han, and Takeshi Takahashi. Detecting coordinated internet-wide scanning by tcp/ip header fingerprint. *IEEE Access*, 2023.

[103] Akira Tanaka, Chansu Han, Takeshi Takahashi, and Katsuki Fujisawa. Internet-wide scanner fingerprint identifier based on tcp/ip header. In *2021 Sixth International Conference on Fog and Mobile Edge Computing (FMEC)*, 2021.

[104] S Torabi, E Bou-Harb, C Assi, E Karbab, A Boukhtouta, and M Debbabi. Inferring and investigating IoT-generated scanning campaigns targeting a large Network Telescope. *Trans. on Dependable and Secure Computing*, 53, 2020.

[105] Sadegh Torabi, Elias Bou-Harb, Chadi Assi, Mario Galluscio, Amine Boukhtouta, and Mourad Debbabi. Inferring, characterizing, and investigating internet-scale malicious iot device activities: A network telescope perspective. In *2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, 2018.

[106] D Wagner, S Ranadive, H Griffioen, M Kallitsis, A Dainotti, G Smaragdakis, and A Feldmann. How to Operate a Meta-Telescope in your Spare Time. In *Internet Measurement Conference (IMC)*, October 2023.

[107] Hui Wang. Early warning: Adb.miner a mining botnet utilizing android adb is now rapidly spreading, 2018.

[108] Eric Wustrow, Manish Karir, Michael Bailey, Farnam Jahanian, and Geoff Huston. Internet background radiation revisited. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, 2010.

[109] Mehdi Zakroum, Jérôme François, Isabelle Chrisment, and Mounir Ghogho. Monitoring network telescopes and inferring anomalous traffic through the prediction of probing rates. *IEEE Transactions on Network and Service Management*, 19(4):5170–5182, 2022.

[110] Mehdi Zakroum, Jérôme François, Mounir Ghogho, and Isabelle Chrisment. Self-supervised latent representations of network flows and application to darknet traffic classification. *IEEE Access*, 2023.

[111] Mehdi Zakroum, Abdellah Houmz, Mounir Ghogho, Ghita Mezzour, Abdelkader Lahmadi, Jérôme FranÇois, and Mohammed El Koutbi. Exploratory data analysis of a network telescope traffic and prediction of port probing rates. In *2018 IEEE International Conference on Intelligence and Security Informatics (ISI)*, 2018.

[112] Guoxu Zhou, Andrzej Cichocki, Qibin Zhao, and Shengli Xie. Efficient nonnegative tucker decompositions: Algorithms and uniqueness. *IEEE Transactions on Image Processing*, 24(12):4990–5003, 2015.