DarkVec: Automatic Analysis of Darknet Traffic with Word Embeddings

Luca Gioacchini, Luca Vassio, Idilio Drago*, Marco Mellia

SmartData@PoliTO, Politecnico di Torino, *Università di Torino

Zied Huidi, Dario Rossi

Huawei Datacom FRC lab, Paris (FR)

WORKSHOP GARR 2021

NET MAKERS

Politecnico di Torino

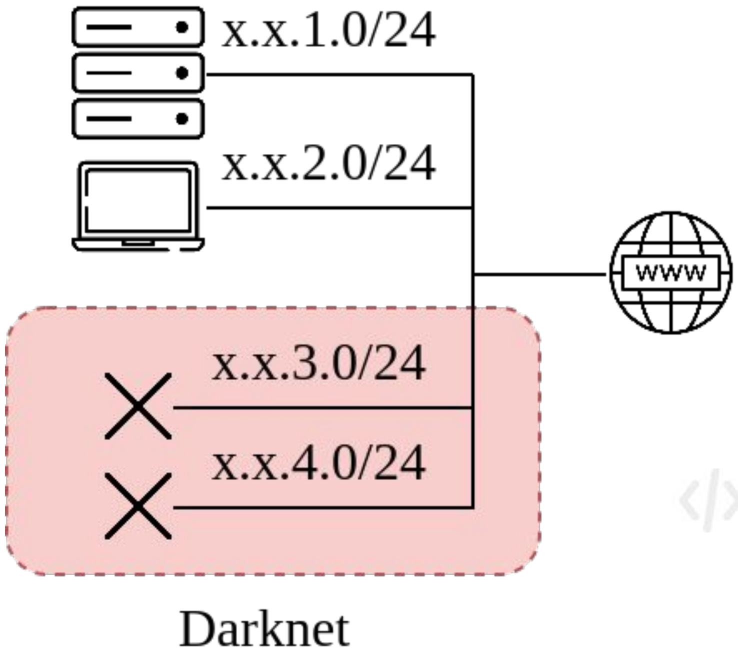
SmartData@PoliTO

Darknets

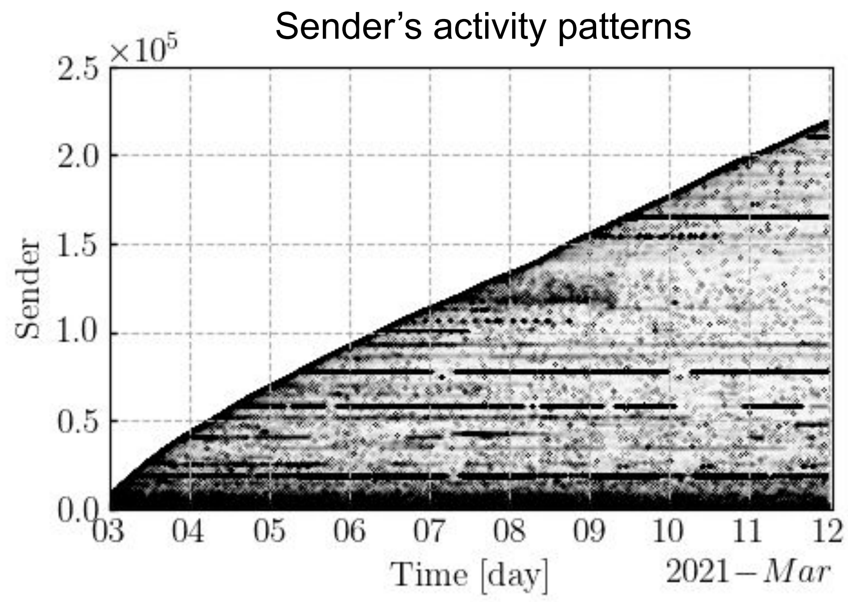
- Darknets are sets of **passive IP addresses** not hosting any services

- Darknets **receive only unsolicited traffic** by definition:
  - **Privileged point of view** for cybersecurity applications
  - But observe a very **noisy picture**

- **Coordinated** senders targeting darknets may be a threat (e.g. **botnets** running distributed attacks), or not (**project scanning the internet** IP address space, backscattering, …)



Problem Definition



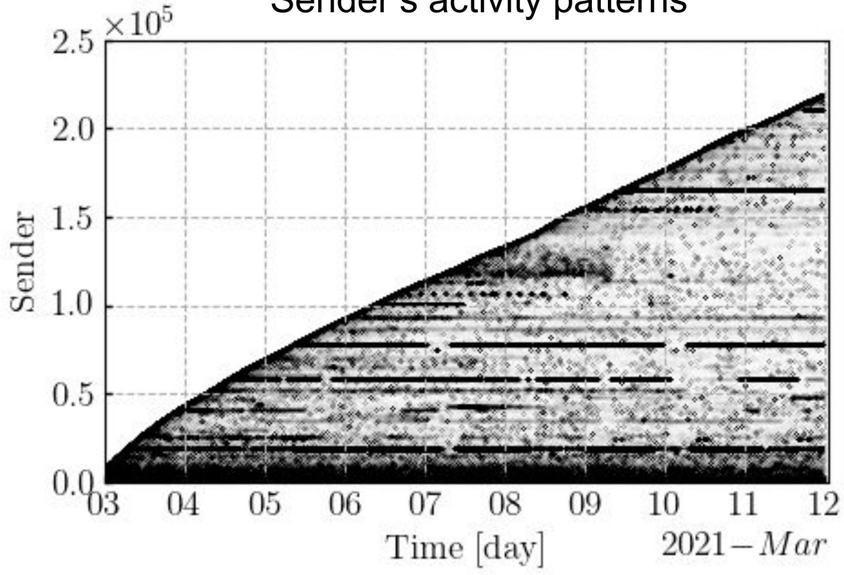
- **10^5** senders target a /24 darknet in one month making **manual** analysis **infeasible**

- Need of **automate the analysis process**

- DarkVec: **Methodology** to automatically identify **clusters** of senders engaged in **similar activities** on darknets relying on **word embeddings**

Problem Definition

Sender's activity patterns

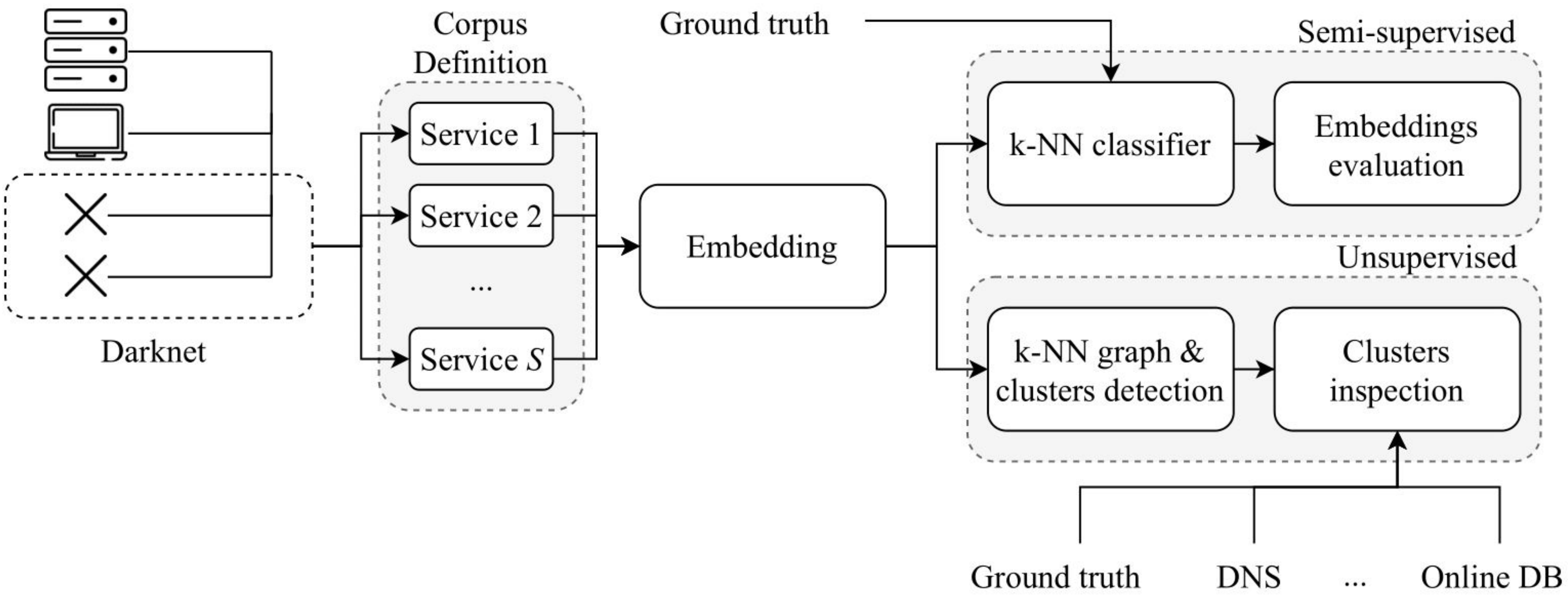


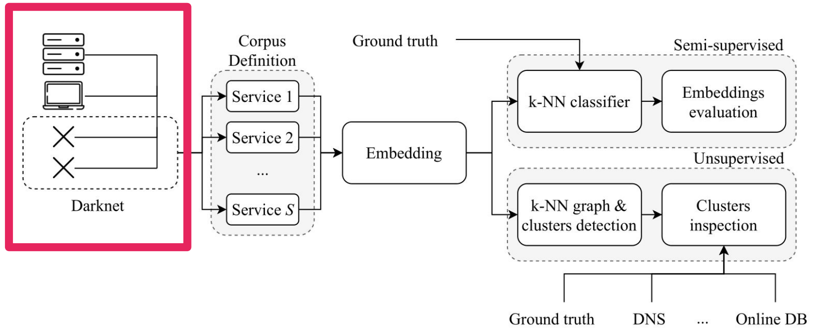
- How can we highlight similar behaviors among senders?

- Main assumption: senders engaged in similar behaviors are expected to exhibit similar **temporal and spatial** patterns in reaching darknets

RQ: Is it possible exploiting temporal co-occurrences between senders reaching darknet to highlight similar behaviors?

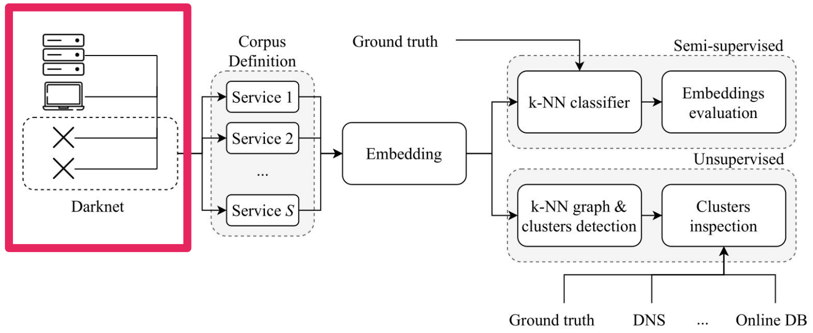
DarkVec Overview




Darknet
Corpus Definition
Service 1
Service 2
...
Service S
Embedding
Ground truth
Semi-supervised
k-NN classifier
Embeddings evaluation
Unsupervised
k-NN graph & clusters detection
Clusters inspection
Ground truth
DNS
...
Online DB

Dataset

Data Collection



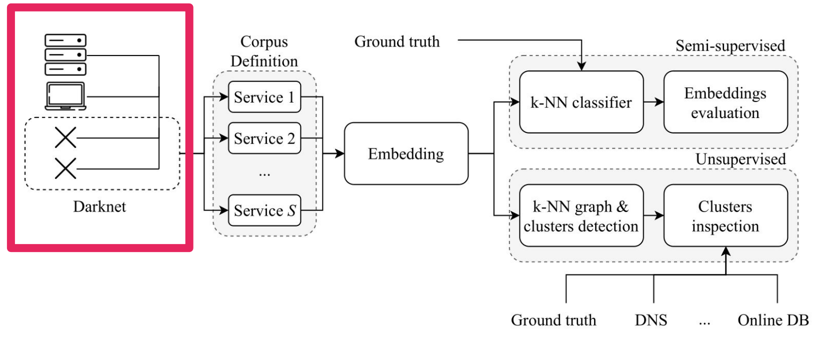
- ***/24*** darknet
- **30 days** of traffic for training
  - From 03-2021
  - > 500k senders
  - > 63M packets
- **Testing** dataset:
  - **Last day** of the collection 2021-03-01
  - > 43k senders
  - > 3M packets

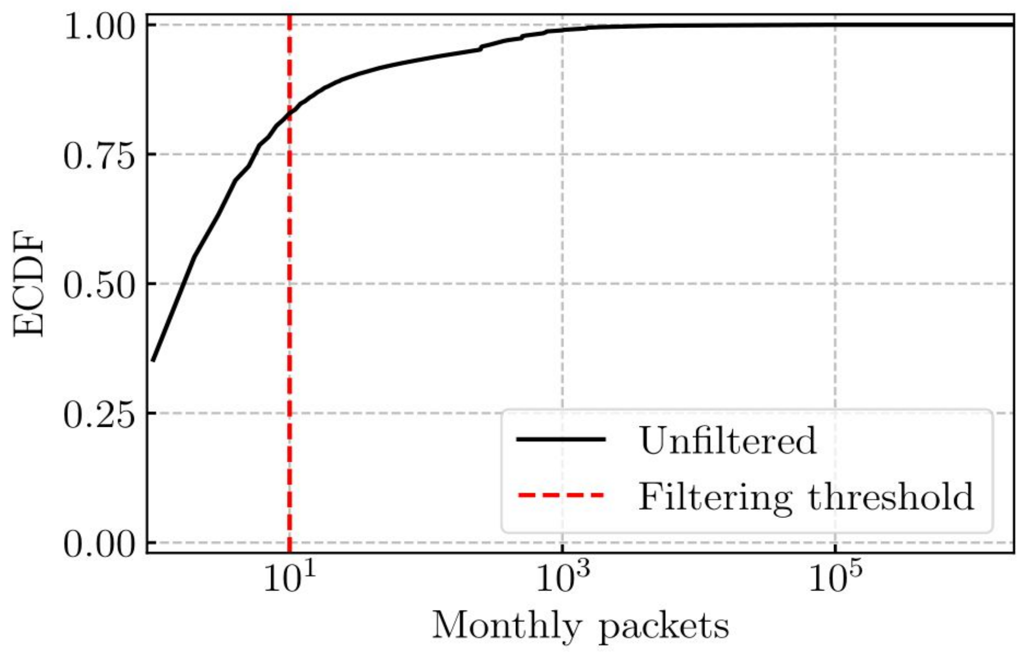
Data Filtering

- **/24** darknet

- **30 days** of traffic for training
  - From 03-2021
  - **> 500k senders**
  - **> 63M packets**

- **Testing** dataset:
  - **Last day** of the collection 2021-03-01
  - > 43k senders
  - > 3M packets

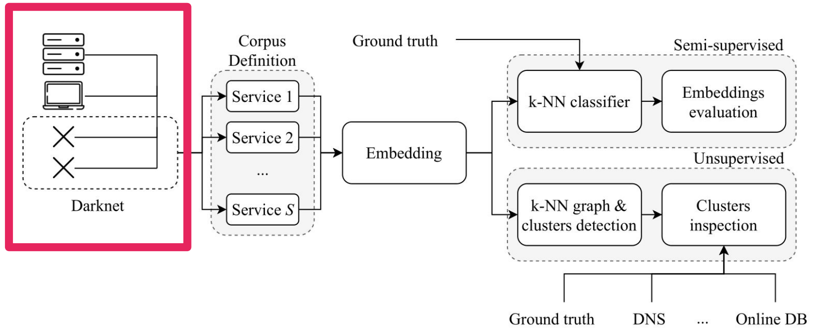


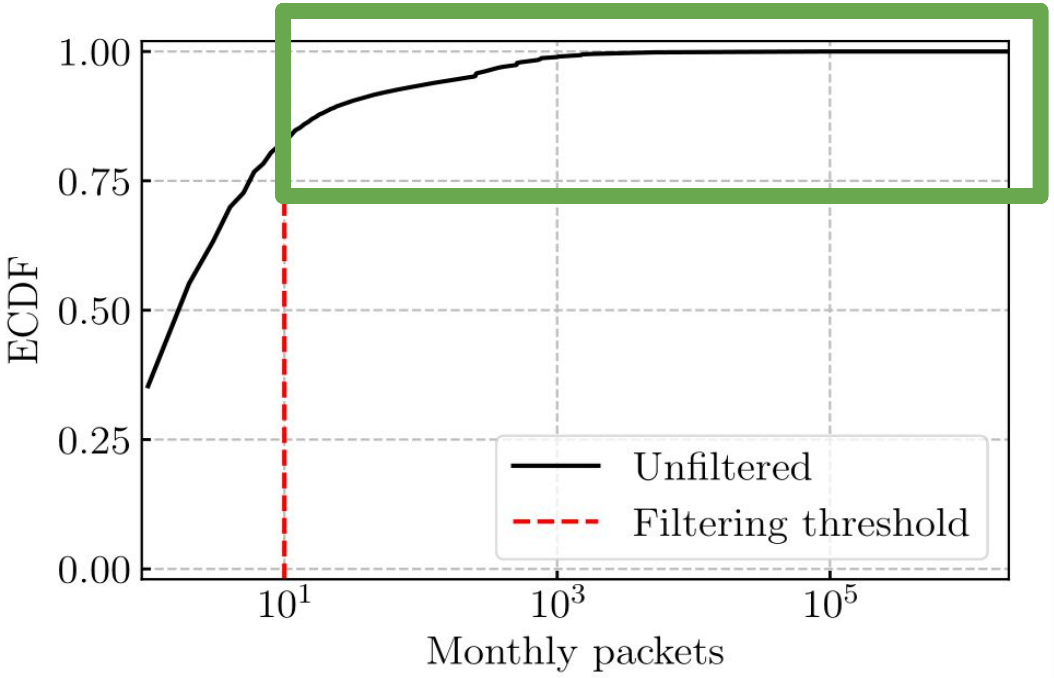


Few senders generate the most of the traffic

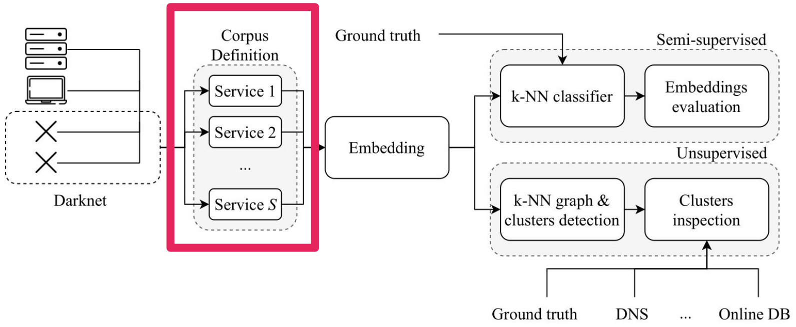
Data Filtering

- **/24** darknet
- **30 days** of traffic for training
    - From 03-2021
    - **> 500k senders**
    - **> 63M packets**
- **Testing** dataset:
    - **Last day** of the collection 2021-03-01
    - > 43k senders
    - > 3M packets



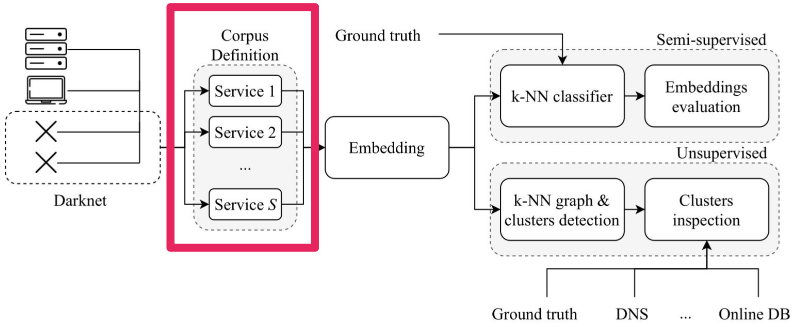


Few senders generate the most of the traffic



Methodology:
time-series by services

Word2Vec Embeddings: Services

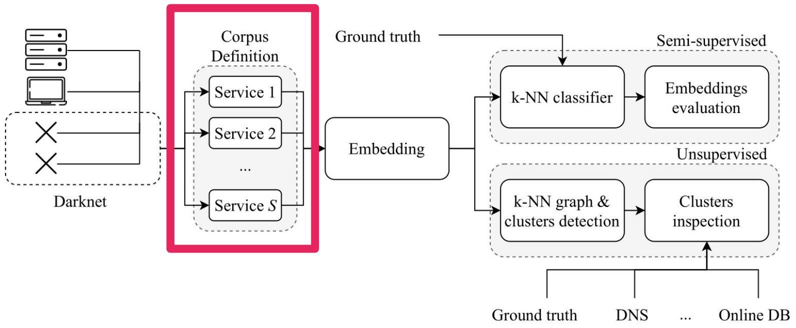


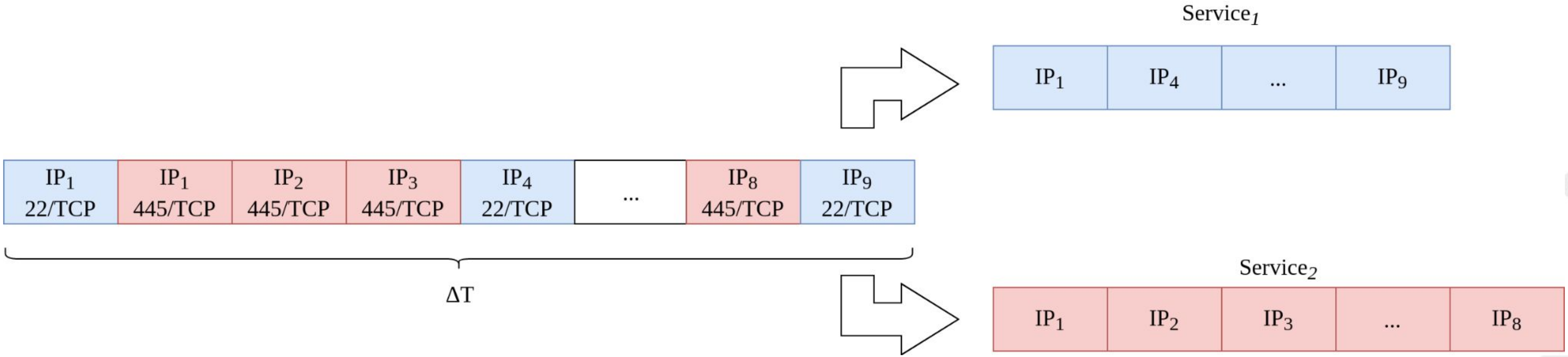
- Service: set of **(destination ports, used protocol)**
  - Port 22/TCP -> SSH service
  - Port 445/TCP -> NetBIOS service
  - Ports 80/TCP, 8080/TCP -> HTTP service

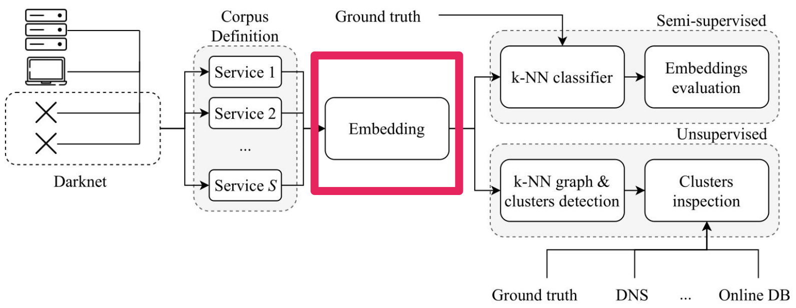
- Extract time-series
  - **Sequence of senders** reaching the darknet
  - **Targeting a given service**

- Three scenarios:
  - **Single** service (original darknet traces)
  - **Auto-defined** services (Top-10 destination ports + 1 as ‘others’)
  - **Domain knowledge based** (15 known services + 1 as ‘others’)

Word2Vec Embeddings: Services

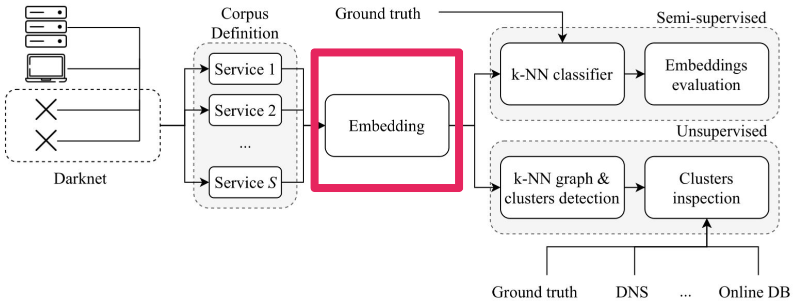




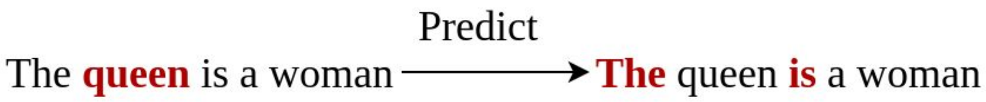


Methodology: *Embeddings*

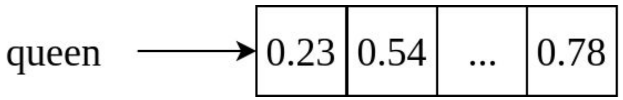
Word2Vec Embeddings



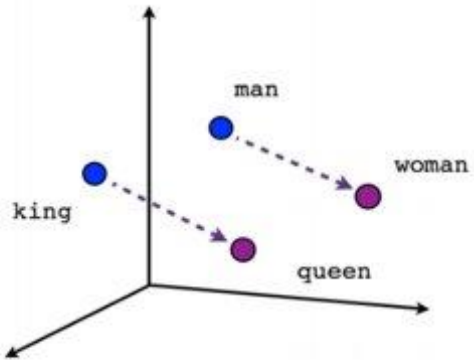
- **Natural Language Processing** technique applied to text documents



- Artificial Neural Network trained to **predict a word in a sentence**

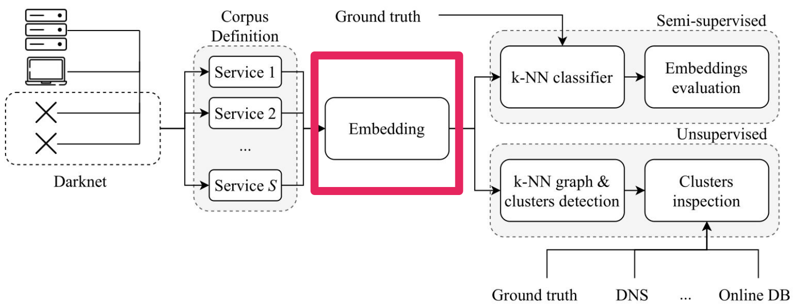


- Word embedding: Latent space with **numerical representation** of a word



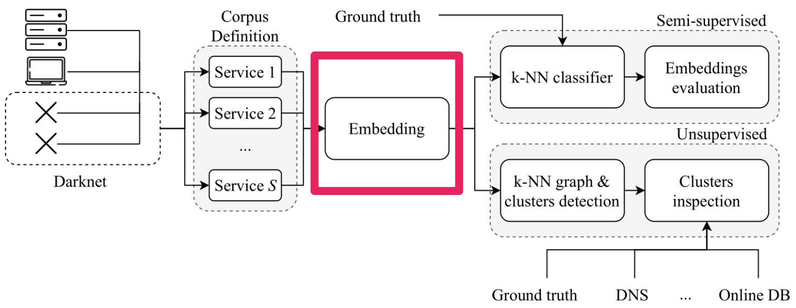
- Words belonging to **similar context** appear **close** in the embedding

Word2Vec Embeddings: NLP Analogies



NLP Applications	DarkVec
Word	Sender

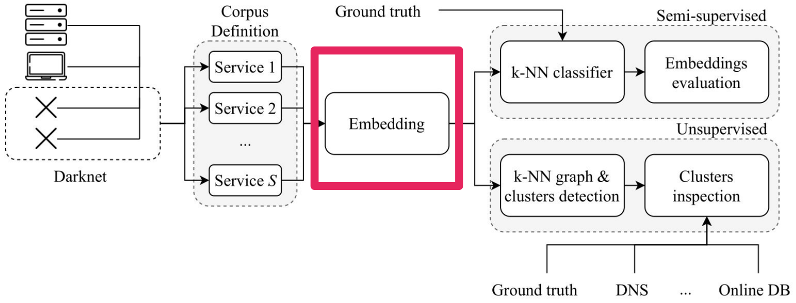
Word2Vec Embeddings: NLP Analogies



NLP Applications	DarkVec
Word	Sender
Semantic context provided by text	Co-occurrence in time and services

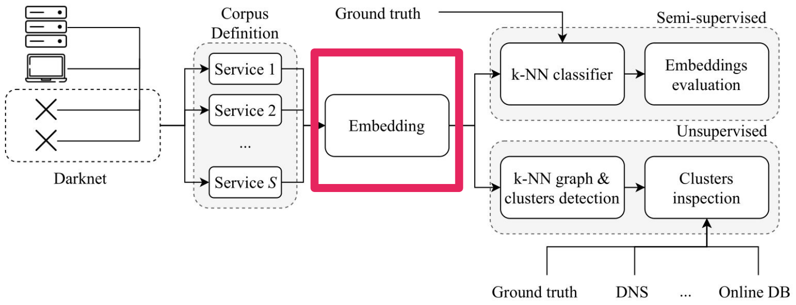
WORKSHOP GARR 2021 NET MAKERS

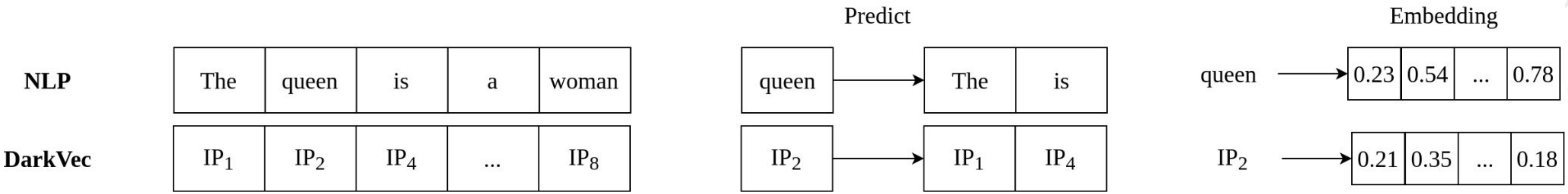
Word2Vec Embeddings: NLP Analogies

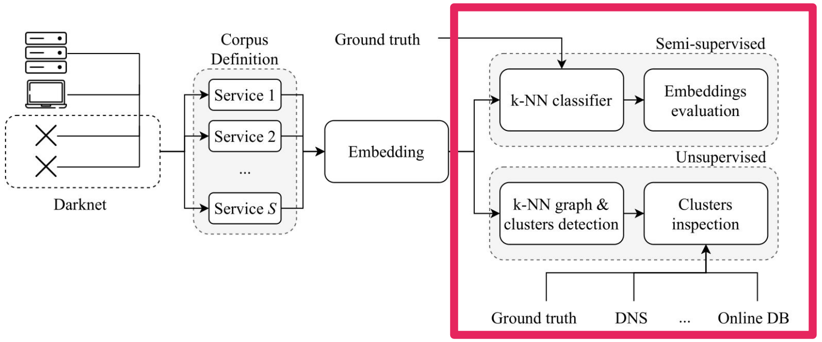


NLP Applications	DarkVec
Word	Sender
Semantic context provided by text	Co-occurrence in time and services
Sentence: sequence of words	Sentence: sequence of senders within a time interval

Word2Vec Embeddings: NLP Analogies







Experiments

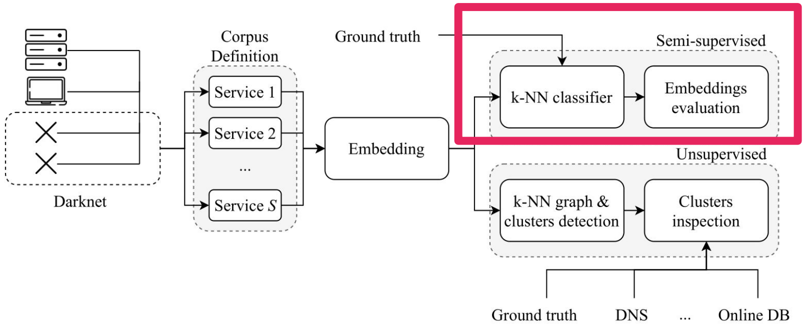
Ground Truth

- Groups of senders whose **coordination is known *a-priori***
  - Fingerprint of the well known Mirai-like malwares
  - Reverse DNS Lookup
  - Publicly available IP addresses of security search engines and research projects (e.g. Shodan.io)

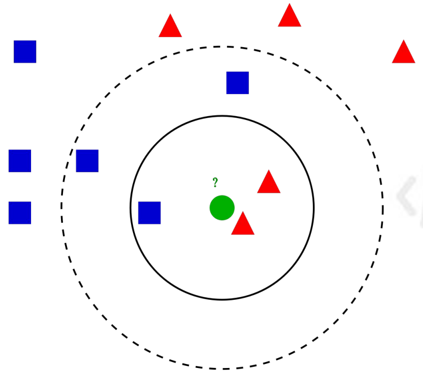
Label	Source	Senders	Packets	Ports	Top-5 Ports Traffic [%]
GT1	Mirai-like [25]	7 351	88 192	75	97.34
GT2	Censys [4]	336	233 004	11 118	7.5
GT3	Stretchoid [15]	104	57 144	91	14.2
GT4	Internet Census [8]	103	9 396	231	37.3
GT5	BinaryEdge [3]	101	7 646	21	38.7
GT6	Sharashka [12]	50	5 436	485	2.28
GT7	Ipip [2]	49	17 342	41	58.9
GT8	Shodan [13]	23	13 566	349	4.1
GT9	Engin-Umich [9]	10	506	1	100
Unknown	–	14 272	2 971 687	10 520	23.7
Total		22 399	3 403 959	19 882	22.8

Identified GT classes active in the last day of our collection

Semi-Supervised Approach

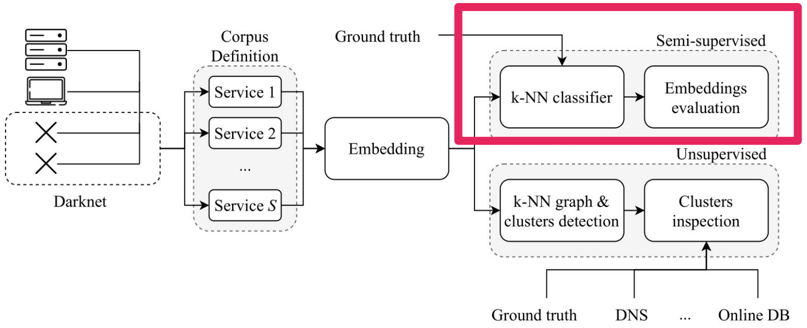


- *Question*: given a partial GT, can we build a classifier to **extend** our knowledge and label more senders?
- Classic **classification** problem
- *Idea*: use a simple ***k*-nn classifier** to label unlabelled senders in the latent space
  - Leave-One-Out validation



WORKSHOP GARR 2021 NET MAKERS

Semi-Supervised Approach



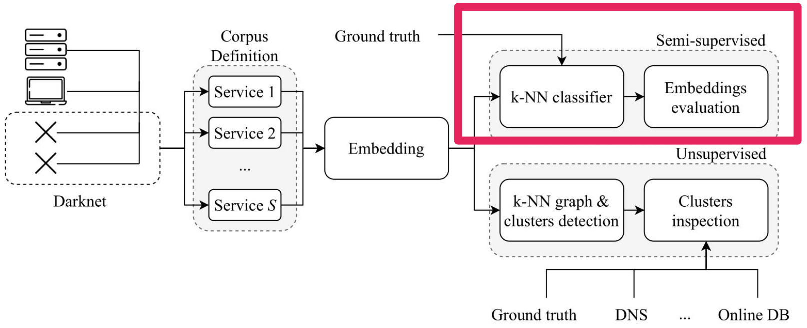
k-NN classifier report

	Single service (c=75, V=50)	
	Precision	Recall
Mirai-like	0.98	0.86
Censys	0.63	0.91
Stretchoid	**0.03**	**0.01**
Internet-census	**0.41**	0.50
Binaryedge	**0.44**	0.74
Sharashka	**0.12**	**0.02**
Ipip	**0.42**	0.92
Shodan	**0.00**	**0.00**
Engin-umich	0.67	1.00
Accuracy	0.84	

Single service is biased on larger class

Best scenario: domain knowledge based services 96% of accuracy

Semi-Supervised Approach



k-NN classifier report

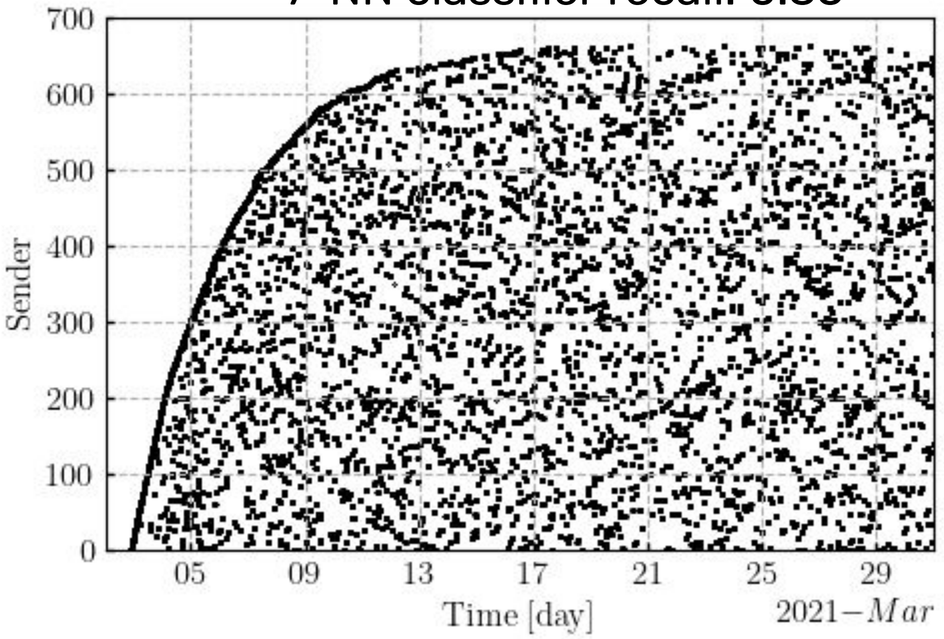
	Single service (c=75, V=50)		Auto-defined services (c=50, V=50)		Domain knowledge based (c=25, V=50)	
	Precision	Recall	Precision	Recall	Precision	Recall
Mirai-like	0.98	0.86	1.00	0.98	1.00	0.97
Censys	0.63	0.91	0.96	1.00	0.91	0.94
Stretchoid	**0.03**	**0.01**	0.94	**0.30**	1.00	**0.35**
Internet-census	**0.41**	0.50	0.79	0.86	0.94	1.00
Binaryedge	**0.44**	0.74	0.98	0.87	0.94	1.00
Sharashka	**0.12**	**0.02**	0.92	0.72	0.96	1.00
Ipip	**0.42**	0.92	0.51	0.86	**0.34**	0.84
Shodan	**0.00**	**0.00**	0.94	0.70	0.93	0.61
Engin-umich	0.67	1.00	0.62	1.00	1.00	1.00
Accuracy	0.84		0.96		0.96	

Single service is biased on larger class

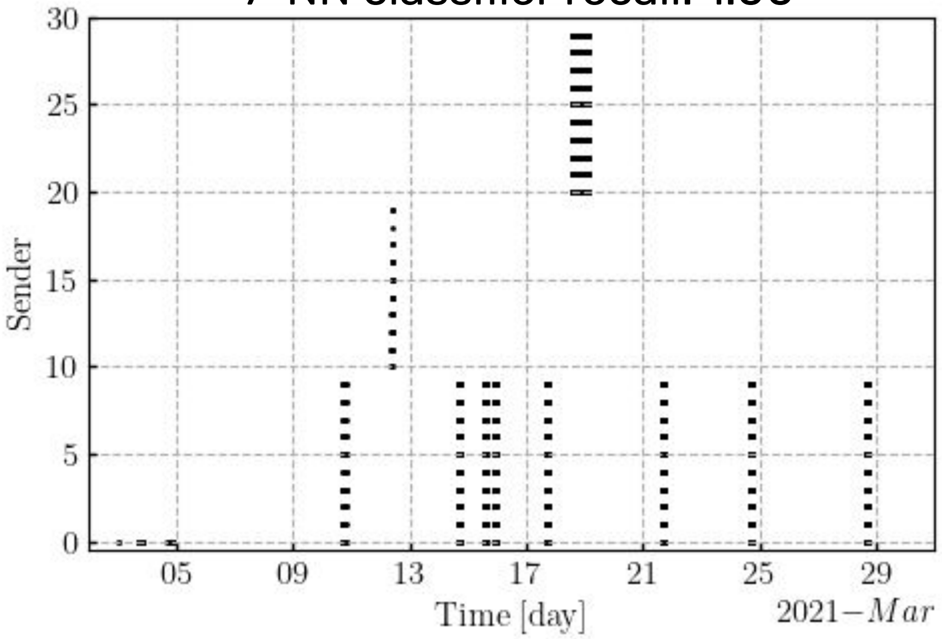
Best scenario: domain knowledge based services 96% of accuracy

Semi-Supervised Approach

Stretchoid activity pattern
7-NN classifier recall: 0.35



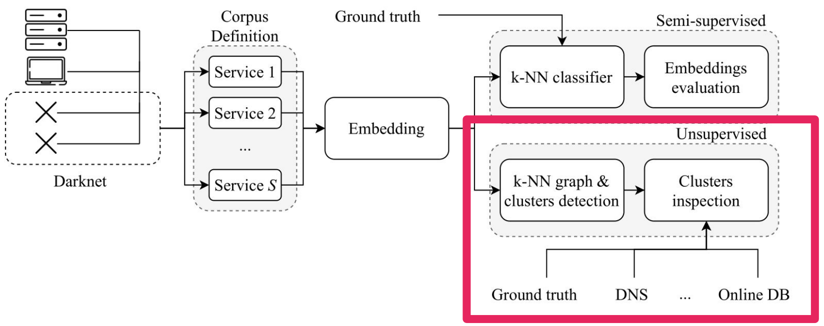
Engin-umich activity pattern
7-NN classifier recall: 1.00



Random patterns are not highlighted by the embeddings

Regular pattern classes are projected into the same portion of the embedding space.

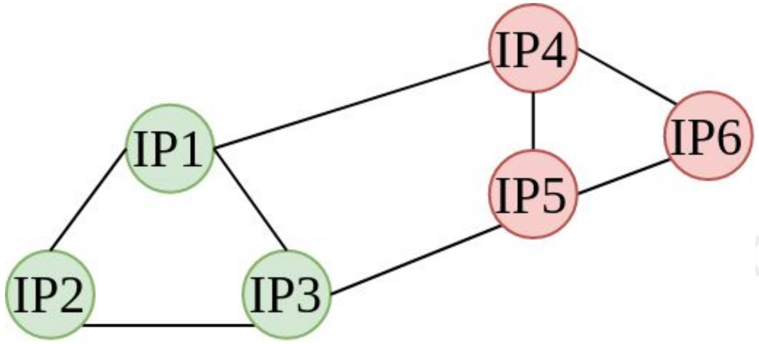
Unsupervised Approach



- *Question*: without knowing anything about the labels, can we **automatically** group senders running **similar activities** on darknets?

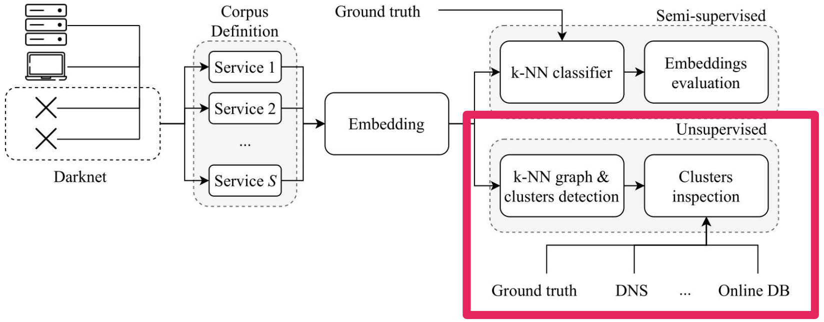
- Clustering problem

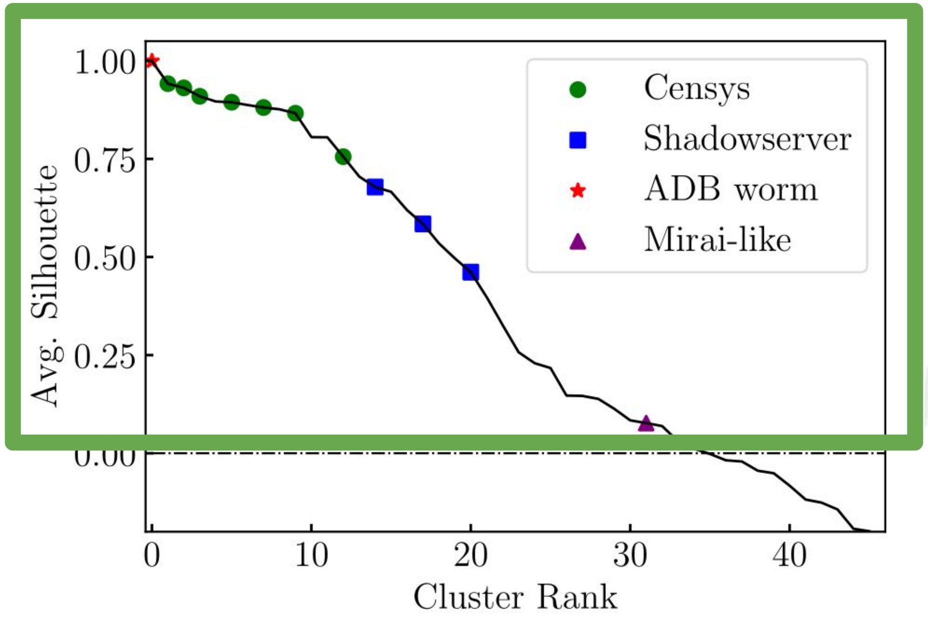
- Cluster senders in the latent space (embeddings)



Unsupervised Approach

- 22k senders grouped in **46 clusters**
- Clusters quality metric: **Silhouette (*Sh*)**
  - For each point it evaluates the within-cluster similarity (cohesion) compared to other clusters (separation)
  - *Sh* = 1 -> sender is well clustered
  - *Sh* = -1 -> sender is bad clustered
  - *Sh* = 0 -> sender is on the border of two clusters





74% of found clusters have *Sh*>0

WORK SHOP GARR 2021 NET MAKERS

Clusters Inspection

- Manual inspection of found clusters
- Main findings:
  - **Sub-clusters** in known scanners
  - **New scanners** from security services
  - New scanners **unknown** to security databases

Name/Type	Cluster	IP	Ports	*Sh*	Description
Censys known scanner sub-clusters	C5 C28 C33 C34 C39 C42 C44	14 16 17 16 16 16 16	19 21 31 25 13 27 26	0.91 0.94 0.76 0.87 0.93 0.88 0.89	Senders of the Censys ground truth class fall into different groups according to the set of ports they target.
ShadowServer known scanner sub-clusters	C25 C29 C37	61 36 16	47 42 51	0.68 0.46 0.58	Senders belonging to the ShadowServer /16 subnet and targeting the same set of ports.
unknown1 NetBios scanner	C40	85	18	0.62	Same /24 subnet in Congent Communications AS.
unknown2 SMTP scanner	C30	10	12	0.89	Same /24 subnet in the Google cloud. >1 600 packets, 76% to SMTP port 25/TCP.
unknown3 SMB scanner	C13	61	5	0.33	>10 900 packets (99.5% of group traffic) is directed to port 445/TCP.
unknown4 ADB massive scanner	C41	525	141	1.00	75% of traffic to 5555/TCP. The IPs activity pattern is coherent with the spreading of a known ADB worm. (Fig.15)
unknown5 Mirai-like massive scanner	C18	1412	212	0.08	71% of senders has Mirai fingerprint. The most of traffic is towards typical botnet ports 23/TCP and 2323/TCP (85%)
unknown6 SSH brute-force	C26	623	116	0.40	>400 000 packets. 88% of group traffic is directed to SSH port 22/TCP.
unknown7 Massive scanner	C31	158	148	0.03	Mostly 'Unknown'. Daily regular activity pattern. Equal share on 148 ports
unknown8 Massive scanner	C45	22	69	0.80	Mostly 'Unknown'. Regular pattern. Almost equal share on 69 ports

Clusters Inspection

- Manual inspection of found clusters

- Main findings:
  - **Sub-clusters** in known scanners
  - **New scanners** from security services
  - New scanners **unknown** to security databases

Name/Type	Cluster	IP	Ports	*Sh*	Description
Censys known scanner sub-clusters	C5 C28 C33 C34 C39 C42 C44	14 16 17 16 16 16 16	19 21 31 25 13 27 26	0.91 0.94 0.76 0.87 0.93 0.88 0.89	Senders of the Censys ground truth class fall into different groups according to the set of ports they target.
ShadowServer known scanner sub-clusters	C25 C29 C37	61 36 16	47 42 51	0.68 0.46 0.58	Senders belonging to the ShadowServer /16 subnet and targeting the same set of ports.
unknown1 NetBios scanner	C40	85	18	0.62	Same /24 subnet in Congent Communications AS.
unknown2 SMTP scanner	C30	10	12	0.89	Same /24 subnet in the Google cloud. >1 600 packets, 76% to SMTP port 25/TCP.
unknown3 SMB scanner	C13	61	5	0.33	>10 900 packets (99.5% of group traffic) is directed to port 445/TCP.
unknown4 ADB massive scanner	C41	525	141	1.00	75% of traffic to 5555/TCP. The IPs activity pattern is coherent with the spreading of a known ADB worm. (Fig.15)
unknown5 Mirai-like massive scanner	C18	1412	212	0.08	71% of senders has Mirai fingerprint. The most of traffic is towards typical botnet ports 23/TCP and 2323/TCP (85%)
unknown6 SSH brute-force	C26	623	116	0.40	>400 000 packets. 88% of group traffic is directed to SSH port 22/TCP.
unknown7 Massive scanner	C31	158	148	0.03	Mostly 'Unknown'. Daily regular activity pattern. Equal share on 148 ports
unknown8 Massive scanner	C45	22	69	0.80	Mostly 'Unknown'. Regular pattern. Almost equal share on 69 ports

Clusters Inspection

- Manual inspection of found clusters
- Main findings:
  - **Sub-clusters** in known scanners
  - **New scanners** from security services
  - New scanners **unknown** to security databases

Name/Type	Cluster	IP	Ports	*Sh*	Description
Censys known scanner sub-clusters	C5 C28 C33 C34 C39 C42 C44	14 16 17 16 16 16 16	19 21 31 25 13 27 26	0.91 0.94 0.76 0.87 0.93 0.88 0.89	Senders of the Censys ground truth class fall into different groups according to the set of ports they target.
ShadowServer known scanner sub-clusters	C25 C29 C37	61 36 16	47 42 51	0.68 0.46 0.58	Senders belonging to the ShadowServer /16 subnet and targeting the same set of ports.
unknown1 NetBios scanner	C40	85	18	0.62	Same /24 subnet in Congent Communications AS.
unknown2 SMTP scanner	C30	10	12	0.89	Same /24 subnet in the Google cloud. >1 600 packets, 76% to SMTP port 25/TCP.
unknown3 SMB scanner	C13	61	5	0.33	>10 900 packets (99.5% of group traffic) is directed to port 445/TCP.
unknown4 ADB massive scanner	C41	525	141	1.00	75% of traffic to 5555/TCP. The IPs activity pattern is coherent with the spreading of a known ADB worm. (Fig.15)
unknown5 Mirai-like massive scanner	C18	1412	212	0.08	71% of senders has Mirai fingerprint. The most of traffic is towards typical botnet ports 23/TCP and 2323/TCP (85%)
unknown6 SSH brute-force	C26	623	116	0.40	>400 000 packets. 88% of group traffic is directed to SSH port 22/TCP.
unknown7 Massive scanner	C31	158	148	0.03	Mostly 'Unknown'. Daily regular activity pattern. Equal share on 148 ports
unknown8 Massive scanner	C45	22	69	0.80	Mostly 'Unknown'. Regular pattern. Almost equal share on 69 ports

MAKERS

Clusters Inspection

- Manual inspection of found clusters

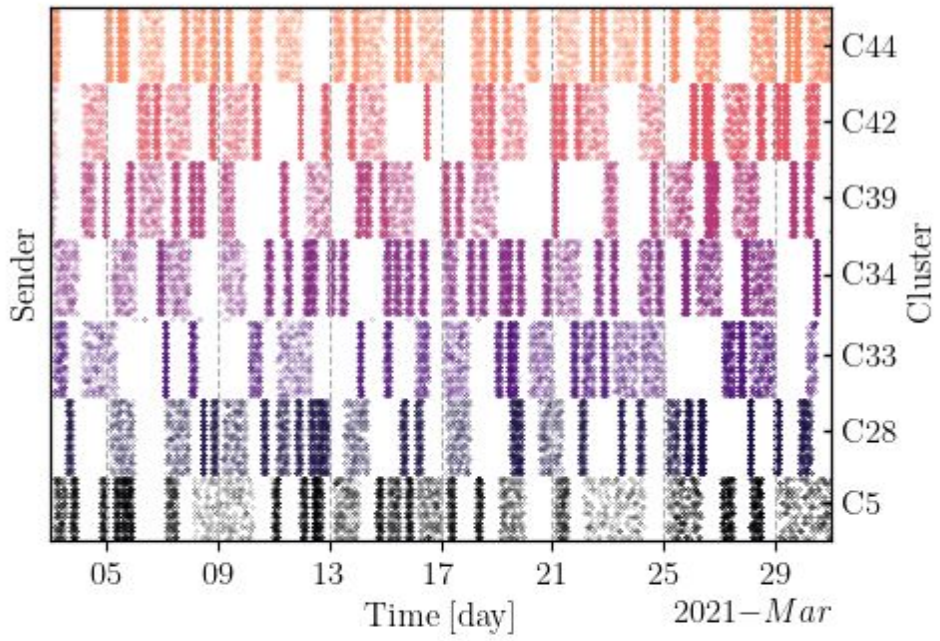
- Main findings:
  - **Sub-clusters** in known scanners
  - **New scanners** from security services
  - New scanners **unknown** to security databases

Name/Type	Cluster	IP	Ports	*Sh*	Description
Censys known scanner sub-clusters	C5 C28 C33 C34 C39 C42 C44	14 16 17 16 16 16 16	19 21 31 25 13 27 26	0.91 0.94 0.76 0.87 0.93 0.88 0.89	Senders of the Censys ground truth class fall into different groups according to the set of ports they target.
ShadowServer known scanner sub-clusters	C25 C29 C37	61 36 16	47 42 51	0.68 0.46 0.58	Senders belonging to the ShadowServer /16 subnet and targeting the same set of ports.
unknown1 NetBios scanner	C40	85	18	0.62	Same /24 subnet in Congent Communications AS.
unknown2 SMTP scanner	C30	10	12	0.89	Same /24 subnet in the Google cloud. >1 600 packets, 76% to SMTP port 25/TCP.
unknown3 SMB scanner	C13	61	5	0.33	>10 900 packets (99.5% of group traffic) is directed to port 445/TCP.
unknown4 ADB massive scanner	C41	525	141	1.00	75% of traffic to 5555/TCP. The IPs activity pattern is coherent with the spreading of a known ADB worm. (Fig.15)
unknown5 Mirai-like massive scanner	C18	1412	212	0.08	71% of senders has Mirai fingerprint. The most of traffic is towards typical botnet ports 23/TCP and 2323/TCP (85%)
unknown6 SSH brute-force	C26	623	116	0.40	>400 000 packets. 88% of group traffic is directed to SSH port 22/TCP.
unknown7 Massive scanner	C31	158	148	0.03	Mostly 'Unknown'. Daily regular activity pattern. Equal share on 148 ports
unknown8 Massive scanner	C45	22	69	0.80	Mostly 'Unknown'. Regular pattern. Almost equal share on 69 ports

Clusters Inspection

Name/Type	Cluster	IP	Ports	*Sh*	Description
Censys known scanner sub-clusters	C5 C28 C33 C34 C39 C42 C44	14 16 17 16 16 16 16	19 21 31 25 13 27 26	0.91 0.94 0.76 0.87 0.93 0.88 0.89	Senders of the Censys ground truth class fall into different groups according to the set of ports they target.
ShadowServer known scanner sub-clusters	C25 C29 C37	61 36 16	47 42 51	0.68 0.46 0.58	Senders belonging to the ShadowServer /16 subnet and targeting the same set of ports.
unknown1 NetBios scanner	C40	85	18	0.62	Same /24 subnet in Congent Communications AS.
unknown2 SMTP scanner	C30	10	12	0.89	Same /24 subnet in the Google cloud. >1 600 packets, 76% to SMTP port 25/TCP.
unknown3 SMB scanner	C13	61	5	0.33	>10 900 packets (99.5% of group traffic) is directed to port 445/TCP.
unknown4 ADB massive scanner	C41	525	141	1.00	75% of traffic to 5555/TCP. The IPs activity pattern is coherent with the spreading of a known ADB worm. (Fig.15)
unknown5 Mirai-like massive scanner	C18	1412	212	0.08	71% of senders has Mirai fingerprint. The most of traffic is towards typical botnet ports 23/TCP and 2323/TCP (85%)
unknown6 SSH brute-force	C26	623	116	0.40	>400 000 packets. 88% of group traffic is directed to SSH port 22/TCP.
unknown7 Massive scanner	C31	158	148	0.03	Mostly 'Unknown'. Daily regular activity pattern. Equal share on 148 ports
unknown8 Massive scanner	C45	22	69	0.80	Mostly 'Unknown'. Regular pattern. Almost equal share on 69 ports

Sender's activity patterns



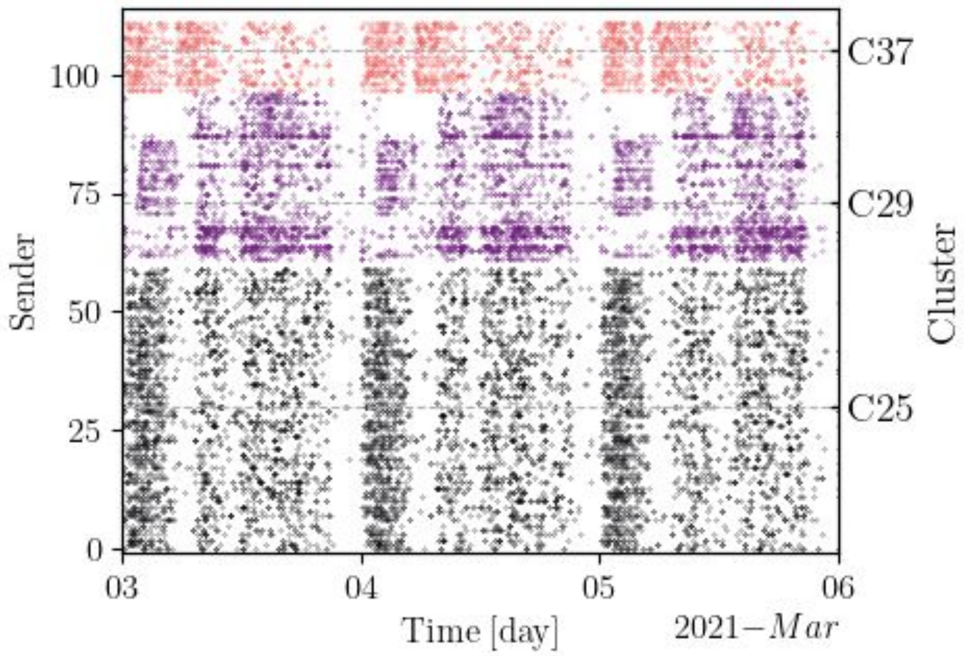
Sub-clusters in known scanners
(*Censys* GT label)

WORKSHOP GARR 2021 NET MAKERS

Clusters Inspection

Name/Type	Cluster	IP	Ports	*Sh*	Description
Censys known scanner sub-clusters	C5	14	19	0.91	Senders of the Censys ground truth class fall into different groups according to the set of ports they target.
	C28	16	21	0.94	
	C33	17	31	0.76	
	C34	16	25	0.87	
	C39	16	13	0.93	
	C42	16	27	0.88	
	C44	16	26	0.89	
ShadowServer known scanner sub-clusters	C25	61	47	0.68	Senders belonging to the ShadowServer /16 subnet and targeting the same set of ports.
	C29	36	42	0.46	
	C37	16	51	0.58	
unknown1 NetBios scanner	C40	85	18	0.62	Same /24 subnet in Congent Communications AS.
unknown2 SMTP scanner	C30	10	12	0.89	Same /24 subnet in the Google cloud. >1 600 packets, 76% to SMTP port 25/TCP.
unknown3 SMB scanner	C13	61	5	0.33	>10 900 packets (99.5% of group traffic) is directed to port 445/TCP.
unknown4 ADB massive scanner	C41	525	141	1.00	75% of traffic to 5555/TCP. The IPs activity pattern is coherent with the spreading of a known ADB worm. (Fig.15)
unknown5 Mirai-like massive scanner	C18	1412	212	0.08	71% of senders has Mirai fingerprint. The most of traffic is towards typical botnet ports 23/TCP and 2323/TCP (85%)
unknown6 SSH brute-force	C26	623	116	0.40	>400 000 packets. 88% of group traffic is directed to SSH port 22/TCP.
unknown7 Massive scanner	C31	158	148	0.03	Mostly 'Unknown'. Daily regular activity pattern. Equal share on 148 ports
unknown8 Massive scanner	C45	22	69	0.80	Mostly 'Unknown'. Regular pattern. Almost equal share on 69 ports

Sender's activity patterns

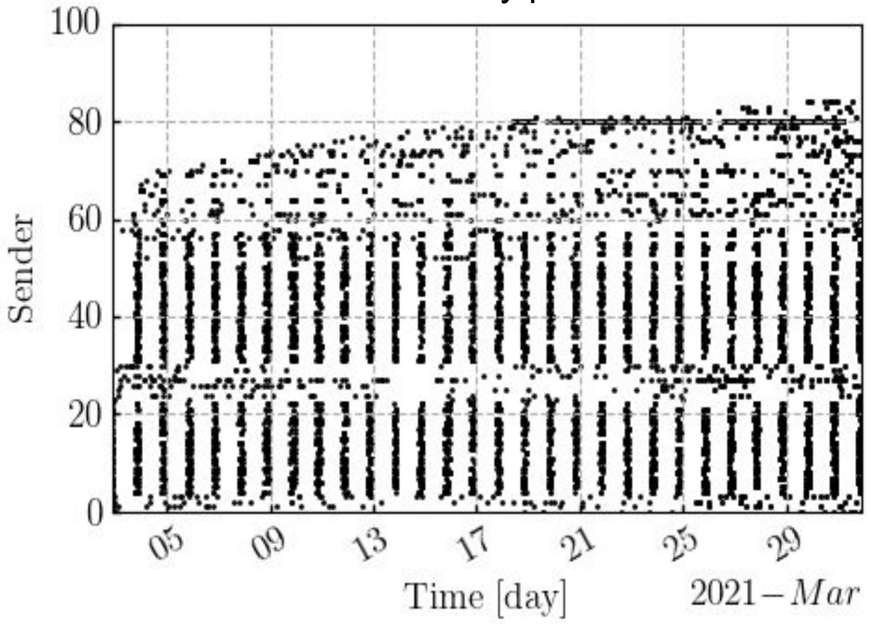


Sub-clusters in **new unknown** security scanners (*Shadowserver.org*)

Clusters Inspection

Name/Type	Cluster	IP	Ports	*Sh*	Description
Censys known scanner sub-clusters	C5 C28 C33 C34 C39 C42 C44	14 16 17 16 16 16 16	19 21 31 25 13 27 26	0.91 0.94 0.76 0.87 0.93 0.88 0.89	Senders of the Censys ground truth class fall into different groups according to the set of ports they target.
ShadowServer known scanner sub-clusters	C25 C29 C37	61 36 16	47 42 51	0.68 0.46 0.58	Senders belonging to the ShadowServer /16 subnet and targeting the same set of ports.
unknown1 NetBios scanner	C40	85	18	0.62	Same /24 subnet in Congent Communications AS.
unknown2 SMTP scanner	C30	10	12	0.89	Same /24 subnet in the Google cloud. >1 600 packets, 76% to SMTP port 25/TCP.
unknown3 SMB scanner	C13	61	5	0.33	>10 900 packets (99.5% of group traffic) is directed to port 445/TCP.
unknown4 ADB massive scanner	C41	525	141	1.00	75% of traffic to 5555/TCP. The IPs activity pattern is coherent with the spreading of a known ADB worm. (Fig.15)
unknown5 Mirai-like massive scanner	C18	1412	212	0.08	71% of senders has Mirai fingerprint. The most of traffic is towards typical botnet ports 23/TCP and 2323/TCP (85%)
unknown6 SSH brute-force	C26	623	116	0.40	>400 000 packets. 88% of group traffic is directed to SSH port 22/TCP.
unknown7 Massive scanner	C31	158	148	0.03	Mostly 'Unknown'. Daily regular activity pattern. Equal share on 148 ports
unknown8 Massive scanner	C45	22	69	0.80	Mostly 'Unknown'. Regular pattern. Almost equal share on 69 ports

Sender's activity patterns

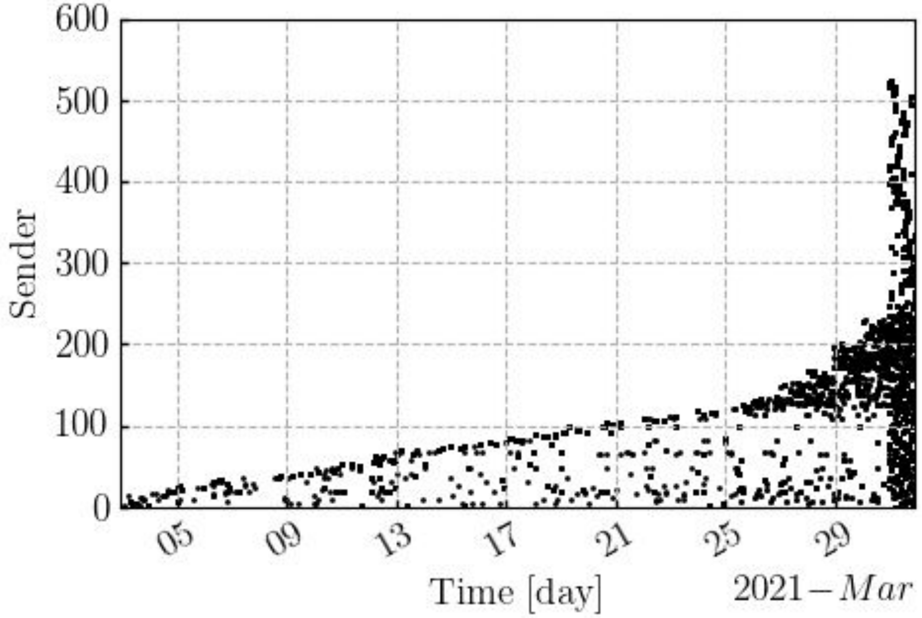


Unknown massive scanners
NetBIOS scan

Clusters Inspection

Name/Type	Cluster	IP	Ports	*Sh*	Description
Censys known scanner sub-clusters	C5 C28 C33 C34 C39 C42 C44	14 16 17 16 16 16 16	19 21 31 25 13 27 26	0.91 0.94 0.76 0.87 0.93 0.88 0.89	Senders of the Censys ground truth class fall into different groups according to the set of ports they target.
ShadowServer known scanner sub-clusters	C25 C29 C37	61 36 16	47 42 51	0.68 0.46 0.58	Senders belonging to the ShadowServer /16 subnet and targeting the same set of ports.
unknown1 NetBios scanner	C40	85	18	0.62	Same /24 subnet in Congent Communications AS.
unknown2 SMTP scanner	C30	10	12	0.89	Same /24 subnet in the Google cloud. >1 600 packets, 76% to SMTP port 25/TCP.
unknown3 SMB scanner	C13	61	5	0.33	>10 900 packets (99.5% of group traffic) is directed to port 445/TCP.
unknown4 ADB massive scanner	C41	525	141	1.00	75% of traffic to 5555/TCP. The IPs activity pattern is coherent with the spreading of a known ADB worm. (Fig.15)
unknown5 Mirai-like massive scanner	C18	1412	212	0.08	71% of senders has Mirai fingerprint. The most of traffic is towards typical botnet ports 23/TCP and 2323/TCP (85%)
unknown6 SSH brute-force	C26	623	116	0.40	>400 000 packets. 88% of group traffic is directed to SSH port 22/TCP.
unknown7 Massive scanner	C31	158	148	0.03	Mostly 'Unknown'. Daily regular activity pattern. Equal share on 148 ports
unknown8 Massive scanner	C45	22	69	0.80	Mostly 'Unknown'. Regular pattern. Almost equal share on 69 ports

Sender's activity patterns



Unknown massive scanners
ADB worm-like

Conclusions

- DarkVec exploits word embeddings to highlight similar behaviors among senders targeting darknets
- It is able to automatically clusters senders performing known activity (semi-supervised learning)
- It lets previously unknown coordinated activity to emerge (unsupervised learning)
  - Sub-clusters in known scanners
  - New scanners from security services
  - New scanners unknown to security databases
- Open source code available at https://github.com/SmartData-Polito/darkvec

Future Work

- Improve DarkVec scalability
- Apply to other temporal sequences (e.g., honeypots)
- Study temporal evolution of senders’ embeddings and clusters structure to detect drifts and new patterns
- Understand if a transfer learning is possible
  - Use the same embedding in i) different vantage points, at ii) different time

THANK YOU FOR YOUR ATTENTION

QUESTIONS?

Luca Gioacchini

luca.gioacchini@polito.it