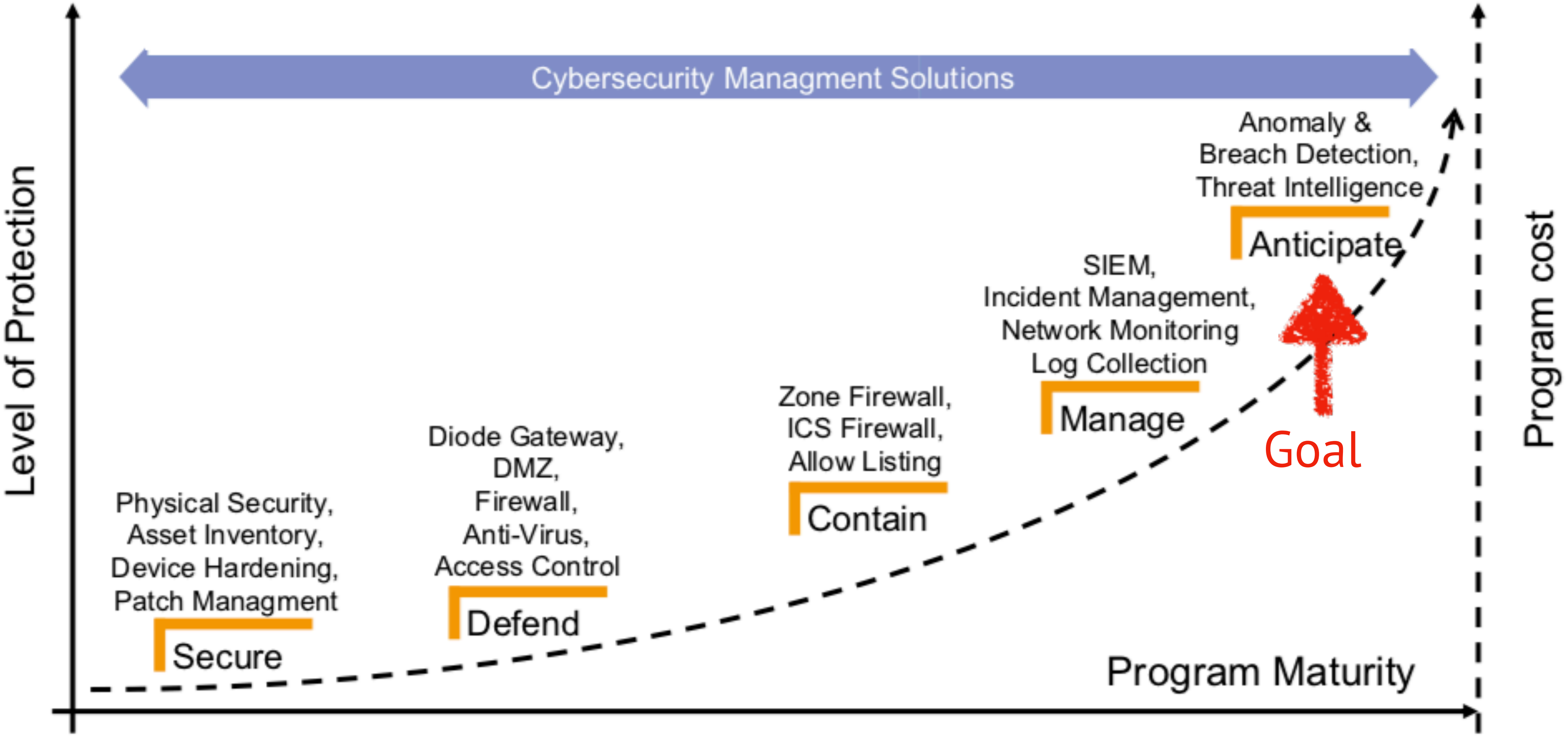
Evaluating IP Blacklists Effectiveness

Luca Deri <deri@ntop.org>, @lucaderi
Francesco Fusco <ffu@zurich.ibm.com>

Introduction [1/2]



Introduction [2/2]

- Reputation systems have been extensively used in network security and network management to maintain networks and networked services secure and reliable.
- A blacklist is an access control mechanism which denies access to selected network resources to peers belonging to a curated list.
- Blacklists often represent the first line of defence for many networks as they can reduce internal hosts’ risk of establishing communications with peers with a bad reputation.

Blacklist Families

- IP Address Blacklists (e.g. malware)
- Domain names
- Malicious TLS Certificates (e.g. JA3)
- IoC (Indicators of Compromise)
- URLs
- Hashes of Malware Files
- …

Blacklist Limitations

- Blacklists are only effective when maintained in a timely manner.
- Blacklists might not equally effective across the planet: a blacklist built and maintained for a specific region (e.g., North America) is not guaranteed to be effective when deployed in another region (e.g., Europe).
- Blacklists do not necessarily cover the traffic seen in the network where they are deployed.

Paper Outline

- In-depth study of publicly available IP malware blacklists: file and cloud-based blacklists.
- Assess the effectiveness of the blacklists by evaluating them against malicious activities on the *easy cases*, i.e., for hosts that can be detected as malicious with certainty, even using simple mechanisms (e.g. a port/network scan).
- All tools used in this paper are open source and their code can be found on github.com/ntop

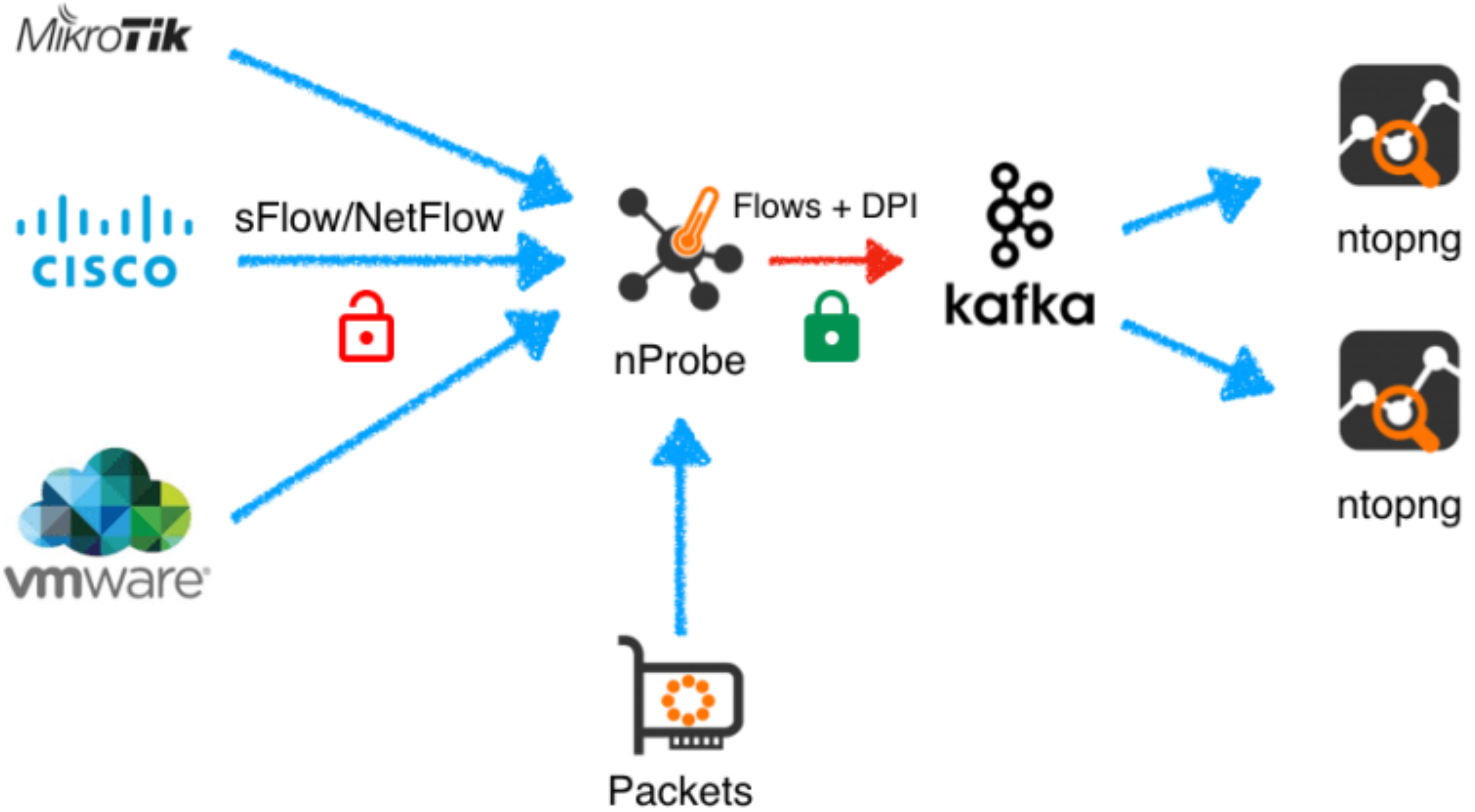
Paper Contributions

- We perform a large-scale study evaluating IP blacklists on real-world production networks of more than *hundred thousand* IP hosts belonging to multiple production networks.
- We describe an effective instrumentation approach to detect IP scanning and suspicious activities toward network servers.
- We showcase that blacklists are optimized for precision, leaving much of the malicious traffic undetected and a false sense of security.

Data Collection Architecture [1/2]

- IP blacklists have been evaluated in three distinct European locations using both IPv4/IPv6:
  - An Italian service provider with about 5’000 hosted servers.
  - A university located in northern Europe with over 100’000 assigned public IP addresses.
  - A leading hosting provider located in the Netherlands hosting corporate servers. Each server is monitored using log files instead of live traffic (packet monitoring with cloud providers isn’t feasible) that include authentication, web administration, email and TCP/UDP ports monitoring.

Data Collection Details [1/2]

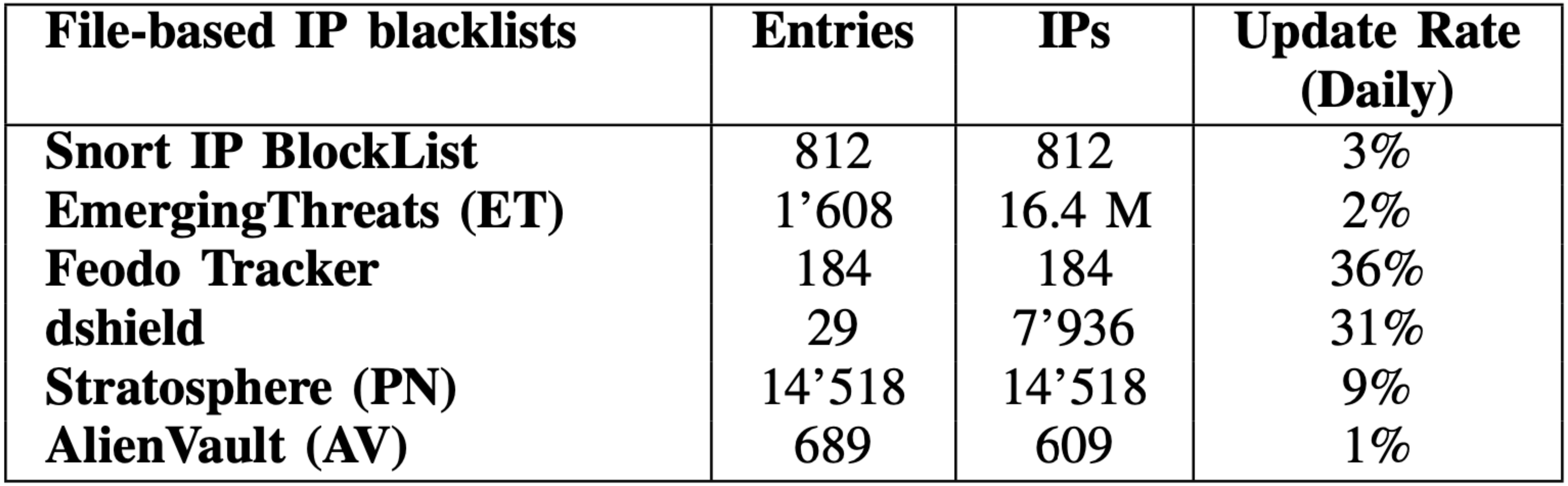


https://github.com/ntop/ntopng/

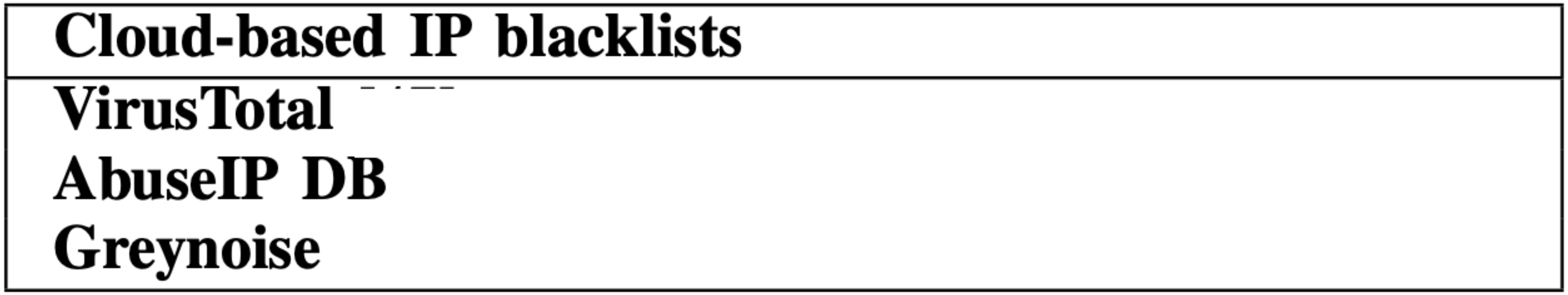
Data Collection Details [2/2]

- The main goal was to collect the IPs that corresponded to attackers with high confidence in order to evaluate the existing blacklists and understand their strengths and limitations.
- Collection Architecture setup in December 2022.
- Paper Experiments Duration: three weeks of traffic between February 27th and March 19th, 2023 (out of 5 months of consistent data collected).
- Scans vs Holidays: scans seems to be affected by holidays. During the period Dec 27th - Jan 3rd a typical host using in our experiments received an average of 30k scans/day with a low of 22k on Jan 1st.

File vs Cloud-Based Blacklists

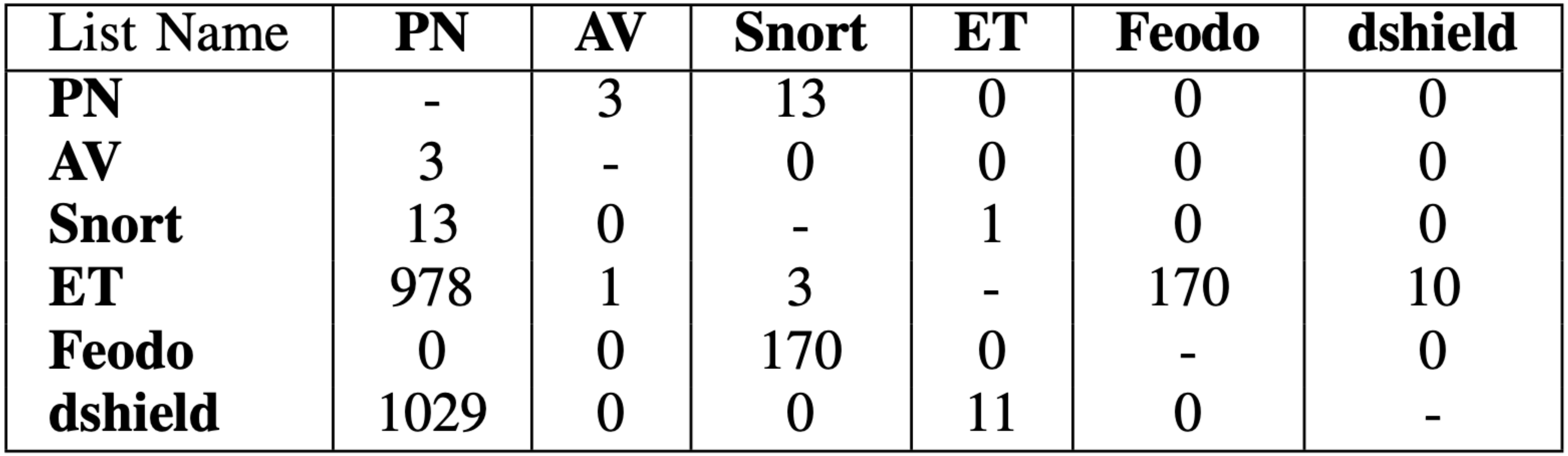


File-based IP blacklists	Entries	IPs	Update Rate (Daily)
Snort IP BlockList	812	812	3%
EmergingThreats (ET)	1’608	16.4 M	2%
Feodo Tracker	184	184	36%
dshield	29	7’936	31%
Stratosphere (PN)	14’518	14’518	9%
AlienVault (AV)	689	609	1%



Cloud-based IP blacklists
VirusTotal
AbuseIP DB
Greynoise

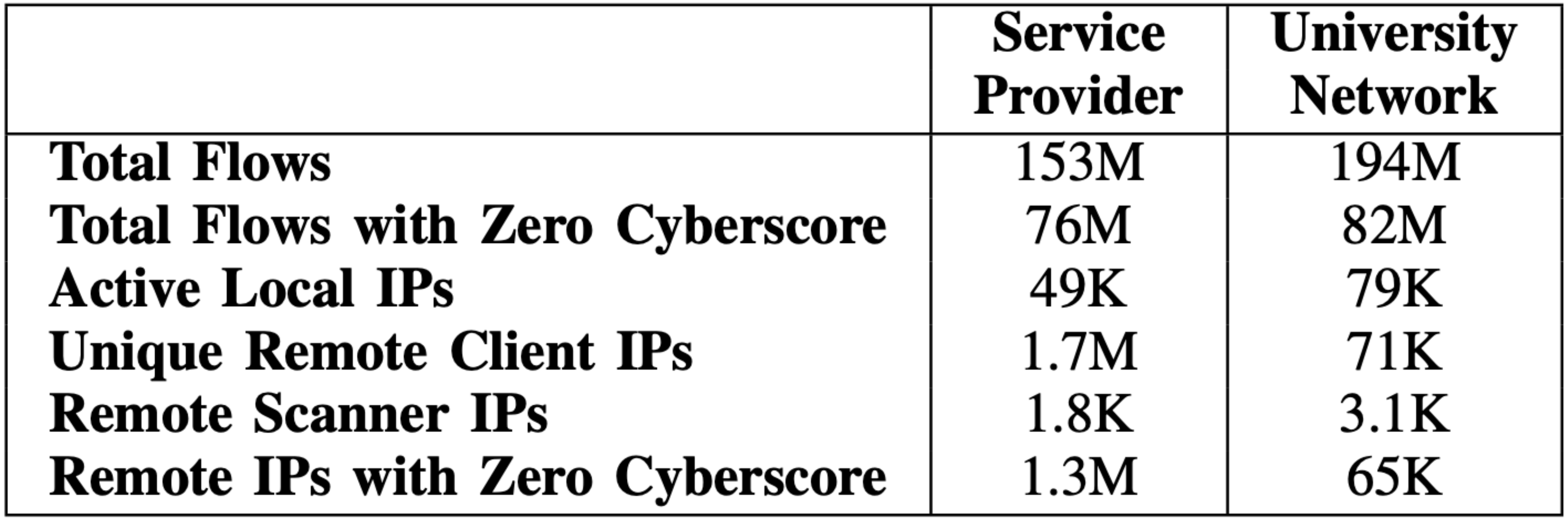
Intersecting Blacklists



List Name	PN	AV	Snort	ET	Feodo	dshield
PN	-	3	13	0	0	0
AV	3	-	0	0	0	0
Snort	13	0	-	1	0	0
ET	978	1	3	-	170	10
Feodo	0	0	170	0	-	0
dshield	1029	0	0	11	0	-

- The intersection of the lists (i.e. an IP address that is included in the intersection of two lists if it is contained in both of them) is limited.
- Note: The results are biased for lists that contain subnets as it is unlikely that all subnet IPs are reported in other lists.

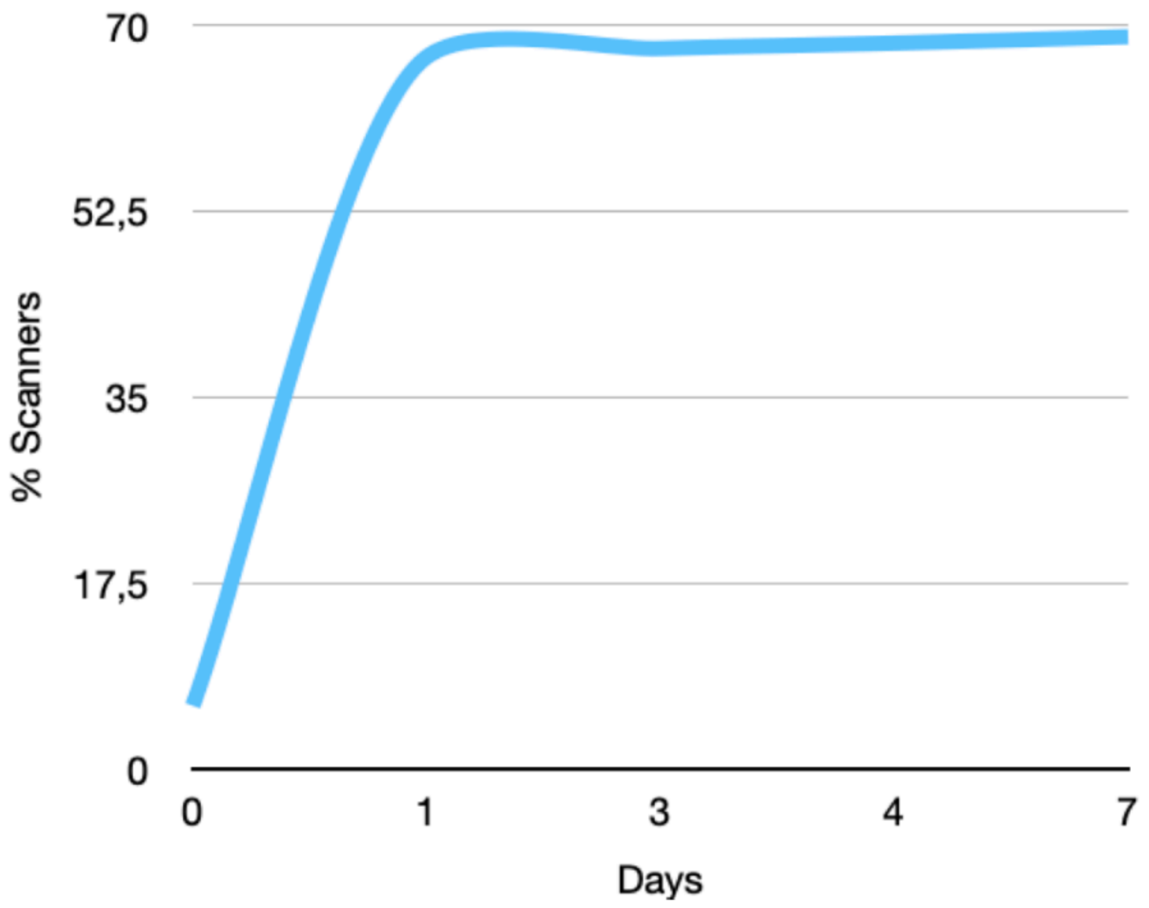
Dataset Size: Service Provider vs University



	Service Provider	University Network
Total Flows	153M	194M
Total Flows with Zero Cyberscore	76M	82M
Active Local IPs	49K	79K
Unique Remote Client IPs	1.7M	71K
Remote Scanner IPs	1.8K	3.1K
Remote IPs with Zero Cyberscore	1.3M	65K

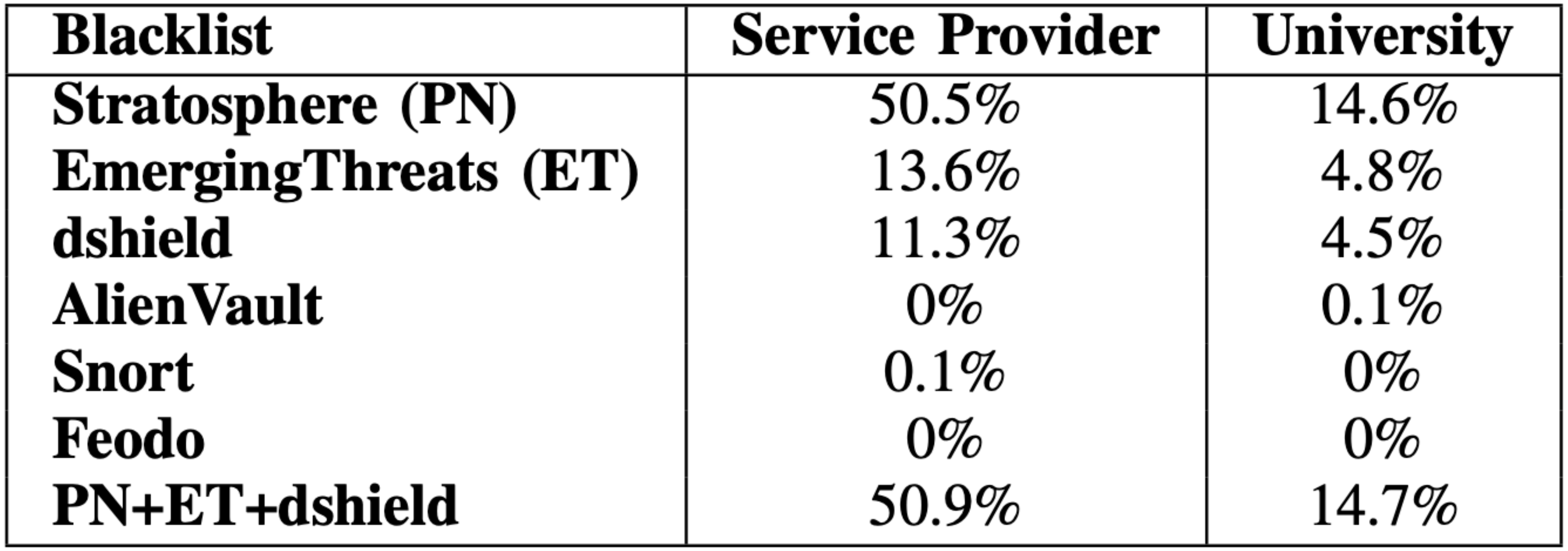
Daily Dataset Size (7 Days Average)

Scanners Propagation



- This slide focuses only on IP/port scanners.
- Only 6% scanner IPs visited both networks on the same day.
- After one day 67% of the university scanners also visited the service provider network. After 13 days this percentage slowly increased to 70%.

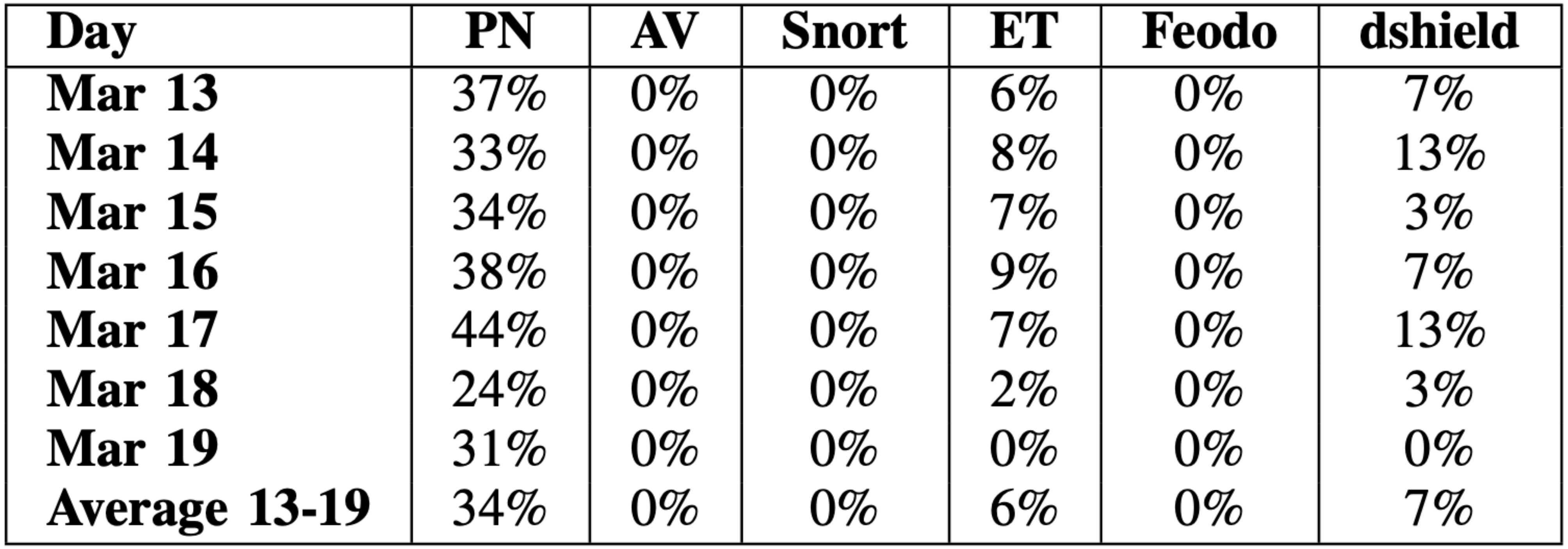
Scanners vs IP Blacklist Match



Blacklist	Service Provider	University
Stratosphere (PN)	50.5%	14.6%
EmergingThreats (ET)	13.6%	4.8%
dshield	11.3%	4.5%
AlienVault	0%	0.1%
Snort	0.1%	0%
Feodo	0%	0%
PN+ET+dshield	50.9%	14.7%

- Providers such as AlienVault, Feodo and Snort are ineffective in marking scanner IPs.
- Stratosphere is the best blacklist to identify scanner IPs, and augmenting it with all the other blacklists improves the detection only marginally (by 1%).

Malicious Hosts vs IP Blacklist Match



Day	PN	AV	Snort	ET	Feodo	dshield
Mar 13	37%	0%	0%	6%	0%	7%
Mar 14	33%	0%	0%	8%	0%	13%
Mar 15	34%	0%	0%	7%	0%	3%
Mar 16	38%	0%	0%	9%	0%	7%
Mar 17	44%	0%	0%	7%	0%	13%
Mar 18	24%	0%	0%	2%	0%	3%
Mar 19	31%	0%	0%	0%	0%	0%
Average 13-19	34%	0%	0%	6%	0%	7%

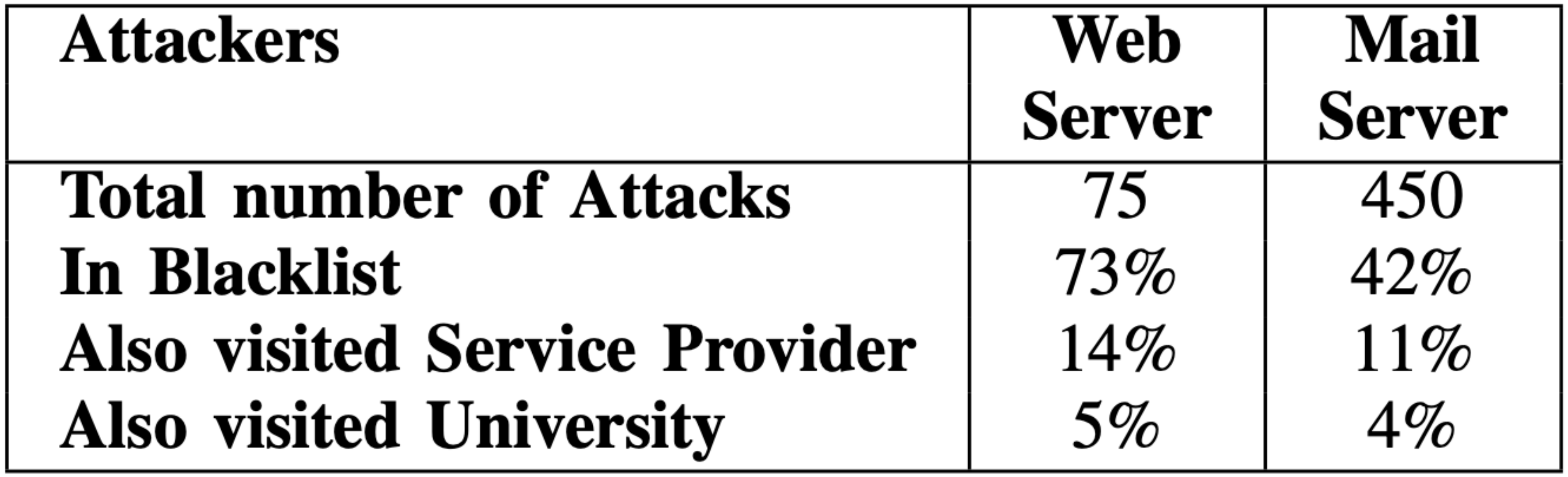
- We have evaluated IP blacklists when detecting cyberthreats and not just simple scanning.
- We have considered only those hosts that ntopng considers malicious with high confidence.
- Results confirm that most IP blacklists are not really effective.

File vs Cloud-based Blacklists

Blacklist	Service Provider	University
Virus Total	80%	63%
Abuse IP DB	25%	18%
Grenoise (50 IP)	10%	10%
Stratosphere	12%	12%
EmergingThreats	4%	4%

- Analysis is limited to 500 IP addresses due to cloud-API limitations.
- Results are divided into two columns: the first one reports a match if the IP is listed, and the second only if there is also a consensus (5 or more matches for VirusTotal, 100% accuracy for Abuse IP, and ’malicious’ for GreyNoise).
- Cloud-based blacklists outperform file-based ones.

Server Monitoring Analysis [1/2]



Attackers	Web Server	Mail Server
Total number of Attacks	75	450
In Blacklist	73%	42%
Also visited Service Provider	14%	11%
Also visited University	5%	4%

- We evaluate intrusion attempts on two hosts, one used as an email and the other as a web server.
- All attempts are recorded by looking at multiple (3+) failures in authentication and application (e.g. email and web server) log files and connection attempts on closed TCP ports.
- The above table reports results for attackers found in IP blacklists.

Server Monitoring Analysis [2/2]

- Blacklists are more effective for the web server host rather than the mail server, even if the total number of attacks on the mail server is higher.
- Comparing attackers with the list of IPs that visited the same-day service provider and the university network, we observe that only a small minority visited such networks.
- While both hosts are dual stack, 99% of the attackers use IPv4 addresses.

Conclusions and Future Work

- Blacklists can only capture a small fraction of scanning activities, and the recall does not significantly improve when blacklists maintained by distinct parties are combined.
- Cloud-based blacklists are more effective than most file-based blacklists, but the limitation in the API queries prevents these types of blacklists from being used as the first line of defence.
- Blacklists are optimised for precision (low recall), leaving much of the malicious traffic undetected and a false sense of security.
- Future work: use ntop open-source tools to dynamically create blacklists and promptly share attackers IPs.

https://github.com/ntop/