Tommaso Rescio

Consortium GARR | THE ITALIAN EDUCATION & RESEARCH NETWORK

smartPOT - Analysis of Darknet Traffic Via Smart Honeypots

GIORNATA DI INCONTRO
BORSE DI STUDIO GARR
"ORIO CARLINI"
ROMA

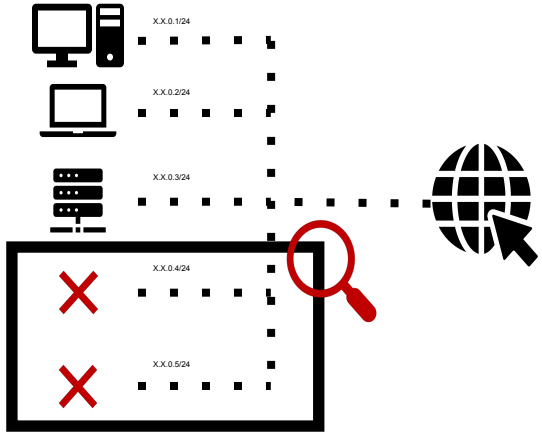
SmartData@PoliTO
Politecnico di Torino

POLITECNICO DI TORINO
SmartData@PoliTO
SmartData

Background

Darknets: IP addresses advertised without hosting any service. **Passive sensors** that highlight several phenomena:

- **Network scans,** both malicious and legitimate
- **Backscattering,** i.e., traffic received from victims of attacks with IP spoofing
- Bugs & **misconfigurations**



Honeypots: intentionally vulnerable hosts used as decoy for attackers in order to record their malicious activities

Motivation and research questions

Increasing the **darknet visibility** with **active responders**

1. How much **extra information** do we get when responding to unsolicited darknet traffic?
2. Do the responses trigger changes on **probed ports and senders**?
3. Do the active services **affect neighbouring darknet ports and addresses**?
4. What if one answers to **services on non-standard ports**?

Our setup

- **L4 responder**

  Negotiate TCP connection, and receive 1st client request
- **L7 responder**

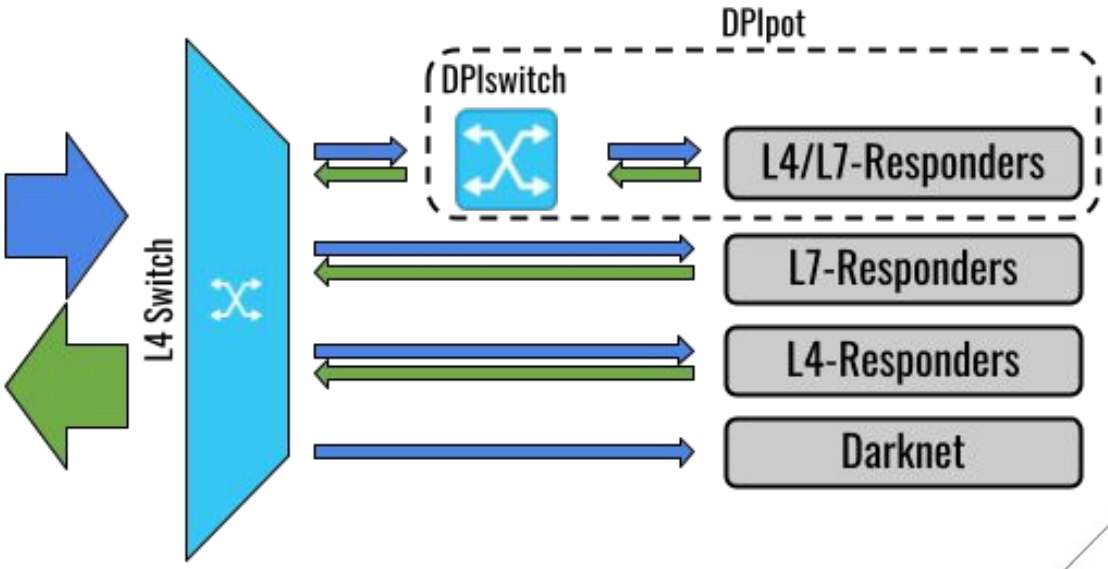
  Vertical honeypots (T-Pot)
- **L4/L7 responder setups**

  Combinations of open ports (standard service ports)
- **DPIPot**

  nDPI to identify L7 protocol + honeypot backends
- **Darknets**

  2 /24 Italy (Polito's IP range)

  2 /24 Italy (GaRR's IP range)



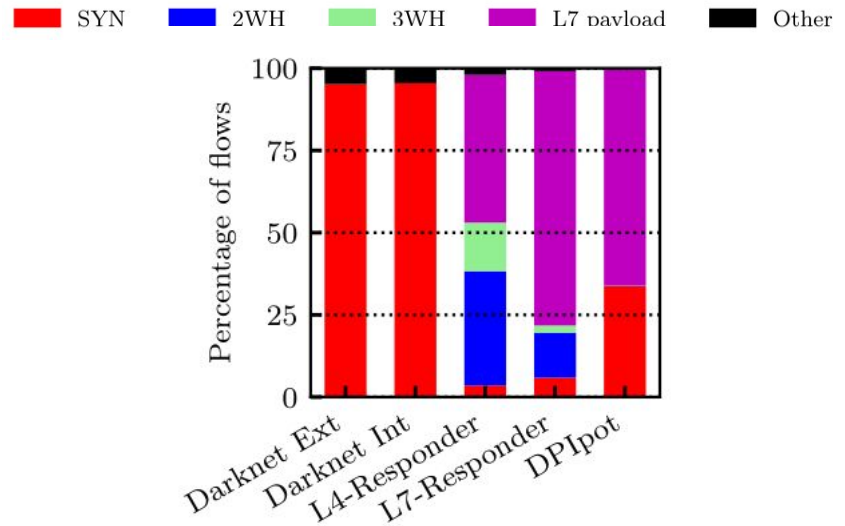
L4 & L7 Responders and DPIPot

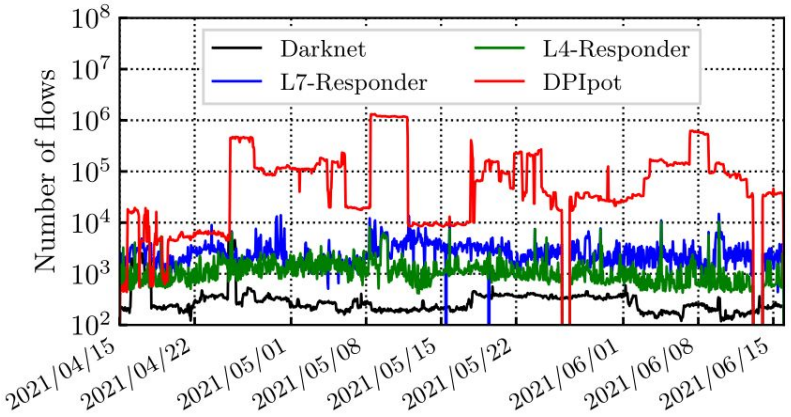
Deployment	Service	Ports	Network size
DPIPot	All	0:65535	/29
L7-Responders	All	All below	/29
	Database	3306, 1433, 27017	/29
	Fileserver	135:139, 445	/29
	Mail	25, 110, 143, 465, 993, 995	/29
	Proxy	8080, 3128	/29
	Remote Desktop	3389, 5900, 5901, 6568	/29
	Terminal	22,23	/29
	Web	80, 443	/29
L4-Responders	All	0:65535	/29
	Database	3306, 33060, 1433, 4022, 1434, 5432, 27017	/29
	Fileserver	135:139, 445	/29
	Mail	25, 110, 143, 465, 993, 995	/29
	Proxy	8080, 8000, 3128	/29
	Remote Desktop	3389, 5900, 5901, 5800, 5801, 5938, 6568	/29
	Terminal	22, 2222, 23, 2323	/29
	Web	80, 443	/29
Darknet	None	0:65535	/24

Macroscopic traffic changes

- How much **extra information** do we get when responding to unsolicited darknet traffic?
- Do the responses trigger changes on **probed ports and senders?**

Macroscopic changes in traffic

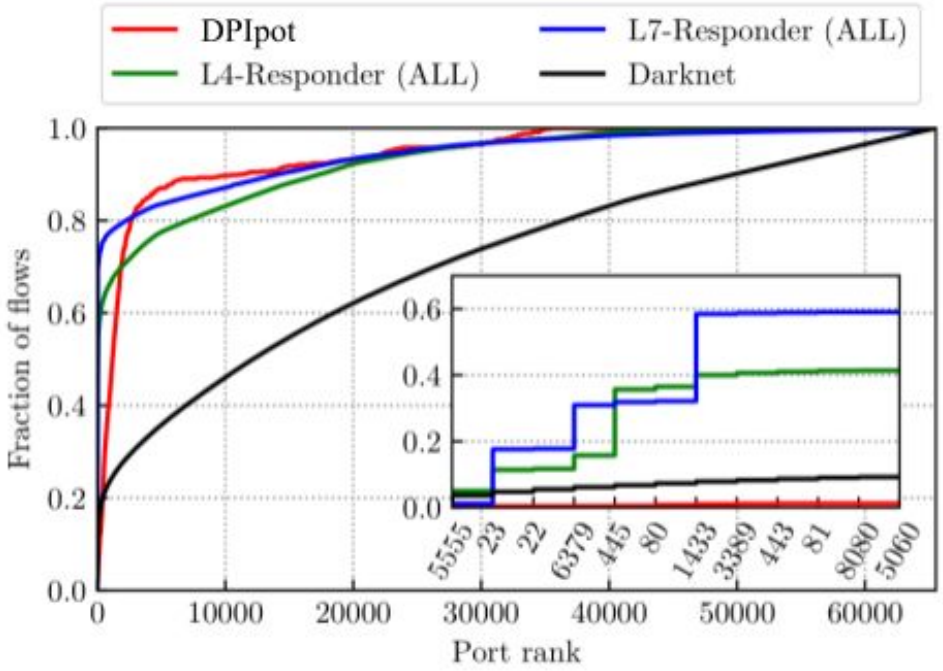




- Darknet: almost only **SYN** messages
- **35%** of the flows hitting the L4-Responder **do not complete the handshake**
- Responding at application level attracts **lots of application layer traffic** (expected)
- DPIPot attracts traffic not seen in **L7-Responders**
  70x increase in volume

Reponders change attackers' behavior?

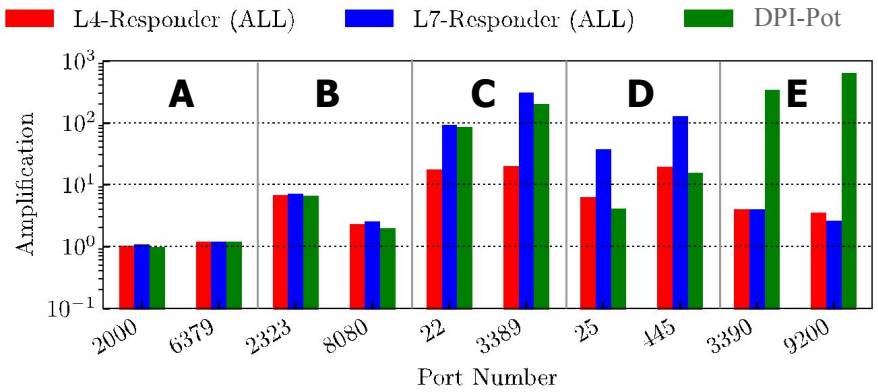
Changes on probed ports



- Well-known ports receive around **20%** of the total traffic hitting the **darknet**
- The top-ports account for the **60%** of the flows on the **L4-Responder**
- The top-ports account for the **70%** of the flows on the **L7-Responder**
- On **DPI-Pot some hundreds of ports** get most of the flows

Service amplification

Amplification factor: ratio between the number of flows seen on the 8 IP addresses of a specific port(s), and the number of flows directed to the same port(s) on the 8 IP addresses belonging to the darknet.



A) **Invariant** (around 50 000 ports): only port scan attempts;

B) **Homogeneous** (around 13 000 ports): senders find possible services on some open ports;

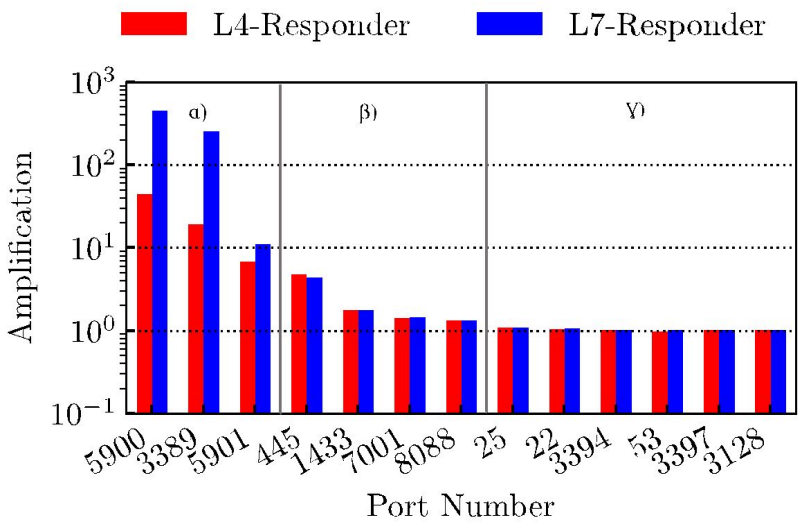
C) **L7 client-initiated** (around 500 ports): these are clear cases of open services on default ports with client-initiated protocols;

D) **L7 server-initiated** (around 10 ports): open services on default ports for which the senders expect the server to initiate the L7 exchange;

E) **Large-scale attacks on non-standard ports** (around 1500 ports): Senders discover particular services on non-standard ports and perform large attacks.

Service amplification

Service-specific deployments (Remote Desktop)



α) Well-known (and open) ports for the category > increase in traffic expected

β) **Side-Scan** ports that suddenly get targeted - despite being blocked > increase in traffic **not** expected

γ) Invariant ports > expected

DPI-Pot

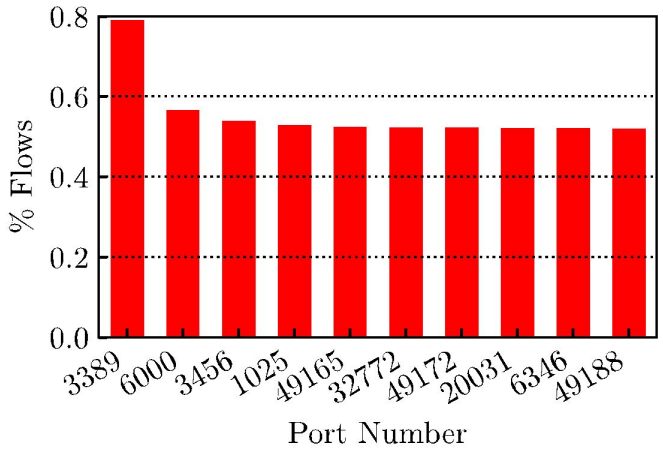
- What if one answers to **services on non-standard ports**?

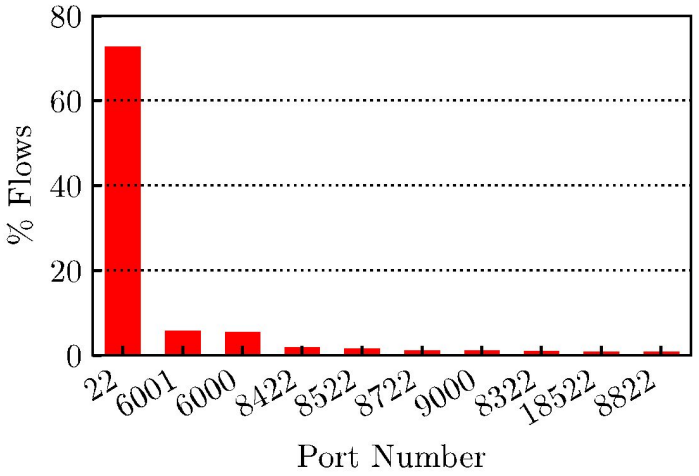
What happens when we do DPI?

Top-5 protocols recognized in DPI-Pot

Protocol	Flows	Sender Addr.	Dest. Ports	% of Flows on Standard Ports
RDP	329 652 678	1 415	28 333	0.8
HTTP	444 715	13 705	9 381	6.2
TLS	221 565	2 806	11 999	4.6
SSH	119 698	1 097	187	72.9
MsSQL-TDS	31 596	3 193	448	92.6

What happens when we do DPI?





RDP protocol:

- Millions of flows on > 28k ports
- 0.8% on port 3389

SSH protocol:

- Thousands of flows on 100s ports
- 72.9% on port 22 *[note the *22*]*

DPIPot attracts new types of scans/attacks that depend on the L7-protocol

Dashboard

How to automate the process?

1. **Raw Data Collection**: we collect raw data through our Honeypots and Darknets

2. **Data Extraction**: we use *Tstat* to process the data (we need Tstat to extract the concept of flow)

3. **Data Adaptation**: we transform the data to create a suitable database for InfluxDB

4. **Data Analysis and Visualization**: we use Grafana

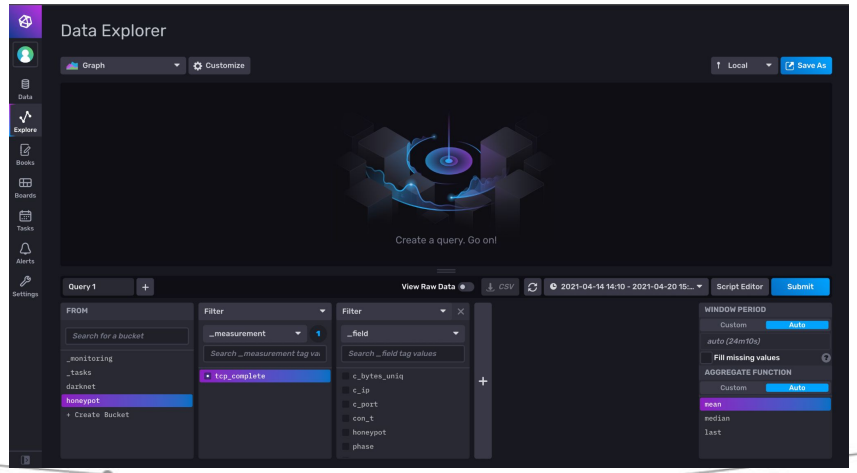
InfluxDB

_time	honeypot	s_port:16_tag	c_port:2_tag	s_ip:15_tag	#00#c_ip:1_tag	phase_tag
2021-04-14 22:47:47.874033920	cannypot	22	47247	130.192.167.144	5.188.86.180	4
2021-04-14 22:47:48.109716992	ip_l4responder_all	8911	38286	130.192.167.11	167.248.133.55	3
2021-04-14 22:47:48.342287872	ip_l4responder_term	22	58243	130.192.167.57	141.98.10.144	2
2021-04-14 22:47:48.207158784	ip_tpot_all	22	30468	130.192.167.78	5.188.86.207	4
2021-04-14 22:47:48.902824960	cannypot	23	36984	130.192.167.146	112.252.221.244	2
...	...	...	...	...	...	...
2021-04-15 22:47:41.909772800	ip_dpipot_all	33917	27223	130.192.167.140	62.197.235.125	4
2021-04-15 22:47:41.923645184	ip_dpipot_all	33925	27225	130.192.167.143	62.197.235.125	4
2021-04-15 22:47:41.920677120	ip_dpipot_all	33930	27224	130.192.167.137	62.197.235.125	4
2021-04-15 22:37:31.022280960	cannypot	22	45766	130.192.167.146	51.89.182.214	2
2021-04-15 22:37:31.022434816	cannypot	22	48280	130.192.167.146	51.89.182.214	2

- Time series database
- Open Source
- Integration with other tools

- Still ongoing

- Horizontal Scalability with cost



Conclusions & Future Works

- Confirm some patterns, e.g., the **increase by 10-100x in traffic when active services are deployed** on the darknet
- Quantify events such as **Side-Scans** attracted by offering different services both on **standard** and **non-standard ports**
- Some services (e.g., RDP, SSH, …) attract **aggressive** (brute-force) attacks
- **InfluxDB** is not the best time series database for our scenario

- Extend the set of responders to mimic behavior of many real system
  - Not just deploying a honeypot
  - Comprehensive simulation of system's behavior, firewalls etc.
- Evolve the responders to avoid detection
  - E.g., our deployment when searched in Shodan
  - **Some IPs have been marked as honeypot, others not.**

Thank you!

Questions?