Tommaso Rescio

Consortium GARR | THE ITALIAN EDUCATION & RESEARCH NETWORK

smartPOT - Analysis of Darknet Traffic Via Smart Honeypots

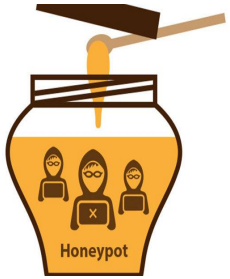
GIORNATA DI INCONTRO
BORSE DI STUDIO GARR
"ORIO CARLINI"
ROMA
21/04/2021

SmartData@PoliTO
Politecnico di Torino

Background

- **Network monitoring for cyber-security purposes;**
- **Darknets** are defined as sets of IP addresses that are advertised without answering any traffic;
  - **Passive traffic only;** ☹

- **Honeypots** are intentionally vulnerable hosts used as decoy for attackers in order to record their malicious activities;



  - ➢ **Active engagement of possible attacker;** ☺
  - ➢ **Protocol-specific;**
  - ➢ **No flexibility.** ☹

Objective

- Engineering of a **novel solution of honeypot**: DPIpot
  - Smart and efficient classification of the application protocol by means of **Deep Packet Inspection** (DPI)
- Engineering of a **flexible framework** of **honeypots** whose configuration can be changed **dynamically:** smartPOT

1. **Analysis of Deep Packet Inspection (DPI)** tools
2. **Design** of the complete infrastructure
3. **Preliminary analysis** of the traffic reaching our infrastructure

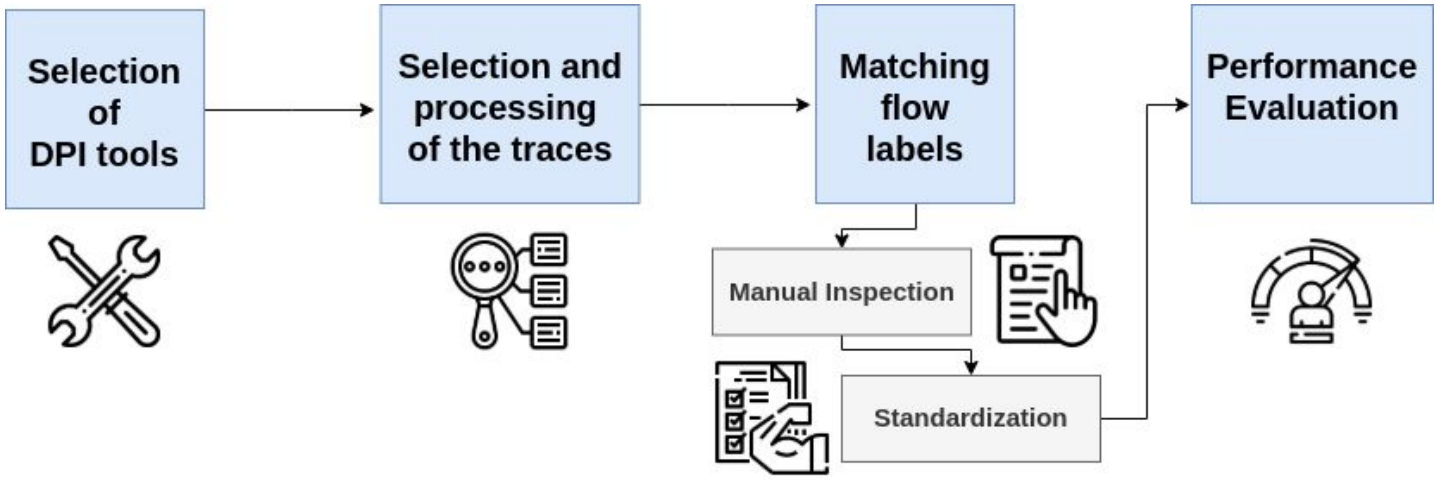
Analysis of Deep Packet Inspection Tools

Research Questions

Find a methodology to **quickly** reply **on-the-fly** to a possible attack with the suitable protocol:

a. Which is the best DPI tool that is able to understand the protocol with the **minimum number of packets**?
b. How do we validate the protocol recognition? Which is the best library among the ones available?

Methodology



Requirements

- open-source
- flexibility
- documentation

Tools

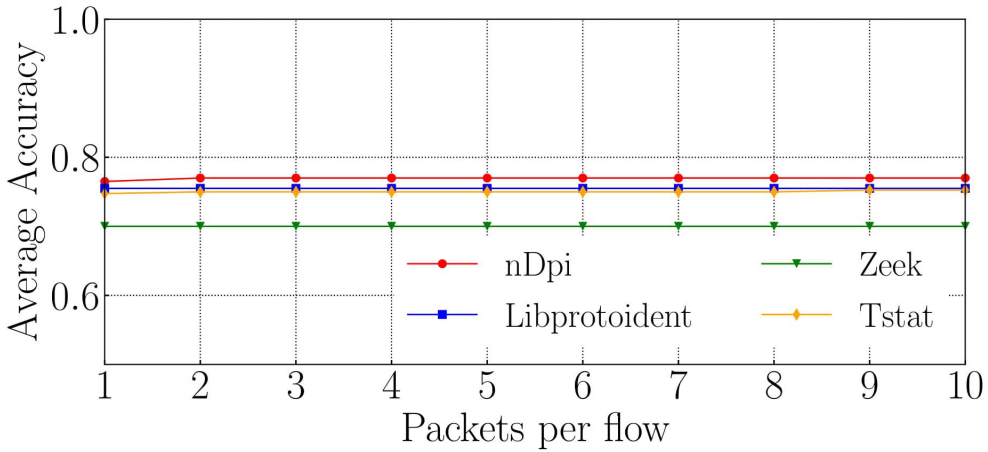
- nDPI
- Libprotoident
- Tstat
- Zeek

Real traces

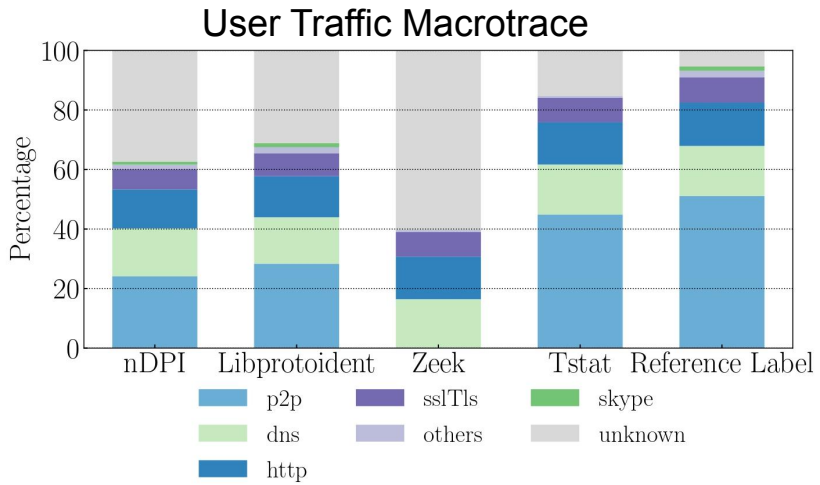
- User
- Media & Games
- Malware
- IoT

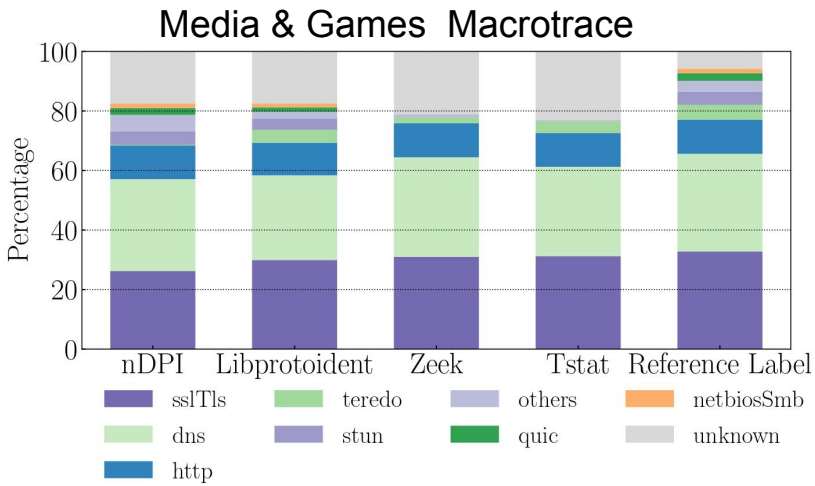
Which is the minimum number of packets per flow?

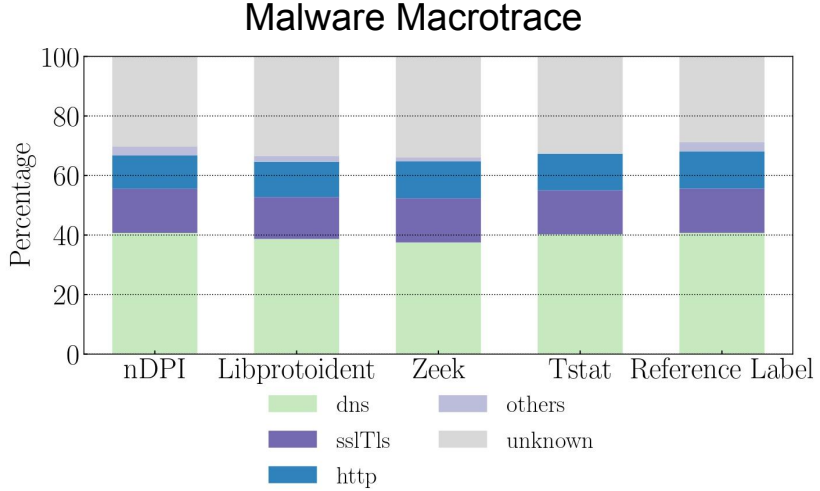
- **Average accuracy** when increasing the number of packets per flow
- Tools reach a final classification already in the **first packet** with payload

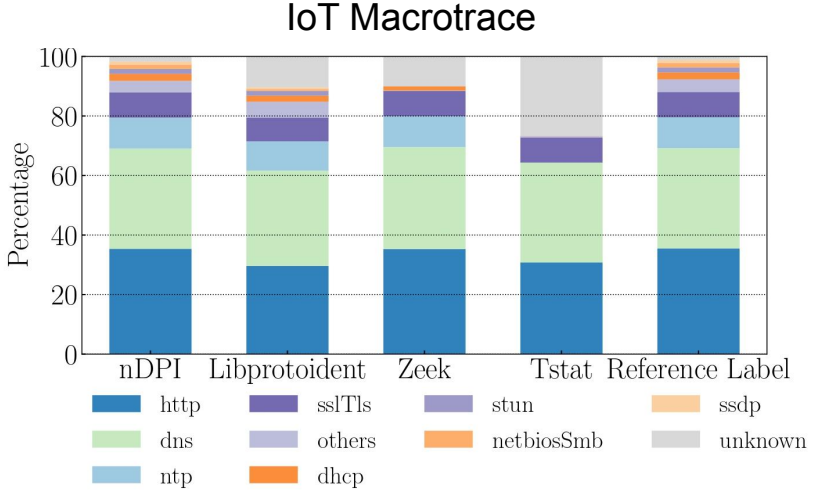


Which is the most accurate library?



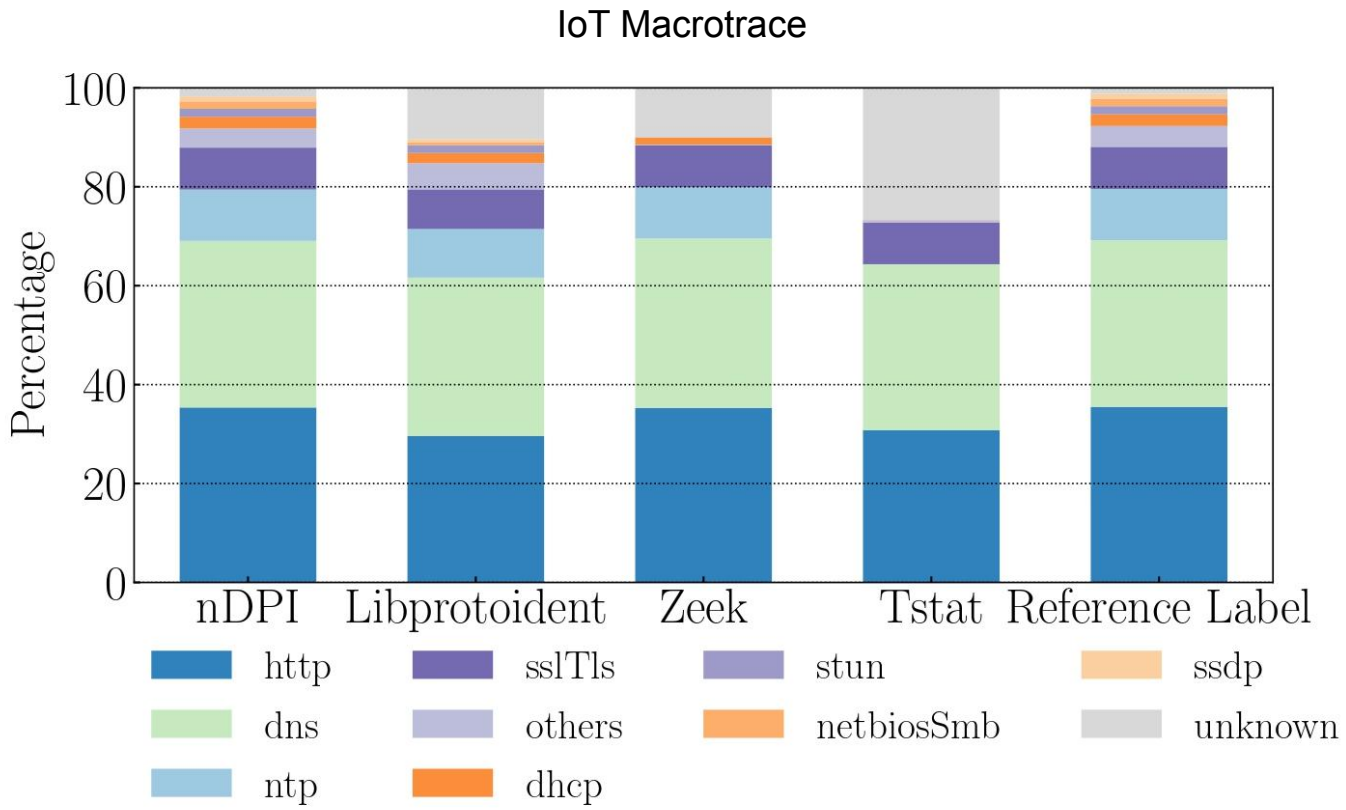






Percentage of labelled flows for each tool

Which is the most accurate library?



Percentage of labelled flows for each tool

Which is the most accurate library?

- **Summary** of classification results
- **nDPI** is the winning tool

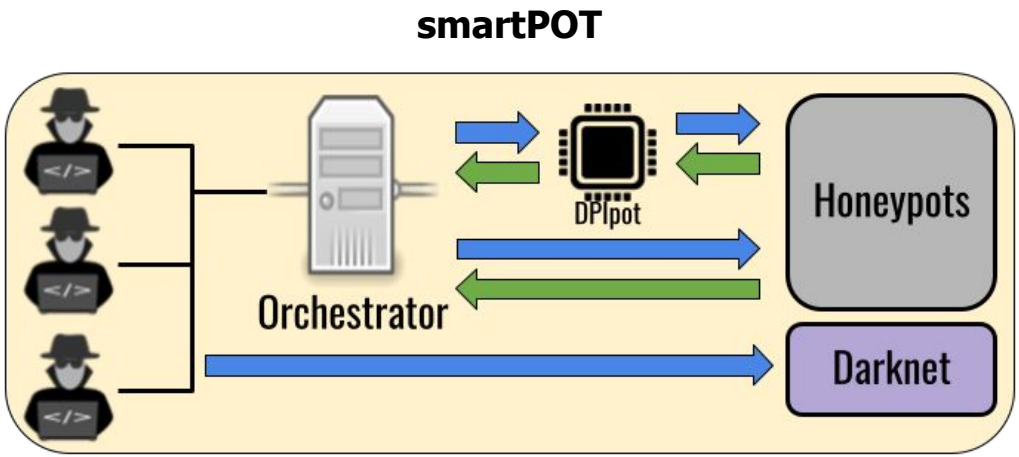
	User Traffic	*Games & Media*	*Malware*	*IoT*
Tstat	**0.85**	0.77	0.67	0.73
Libprotoident	0.69	**0.82**	0.66	0.85
nDPI	0.62	0.79	**0.70**	**0.98**
Zeek	0.40	0.78	0.66	0.89

Design of the infrastructure

Infrastructure

smartPOT: flexible framework of **honeypots**, whose configuration can be changed **dynamically**

DPIpot: redirect the attacks to the **most suitable honeypot**



Infrastructure

- **Darknets** are our **baseline**
- **L4-Responder:** it completes only the three way handshake
- **L7-Responder:** SoA honeypots

Preliminary analysis of the collected traffic

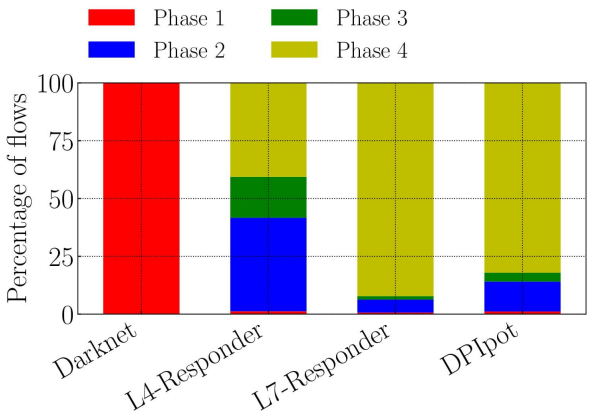
Research Questions

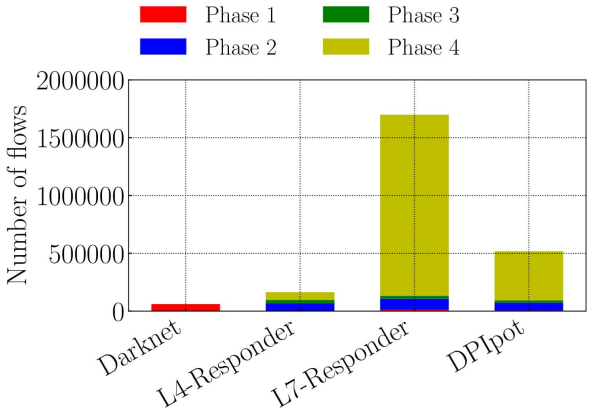
Preliminary analysis of the collected traffic:

a. What is the **share of the traffic** reaching the honeypots that arrives to **different attack phases**?
b. Does the **attack pattern** change if **we start replying** to all the connection requests?
c. Does the **attack pattern** change depending on the kind of services we expose?
d. Does **identifying protocols on-the-fly** before replying, even when traffic reaches non-standard ports, influence the attack patterns?

What is the share of the traffic that arrives to different attack phases?

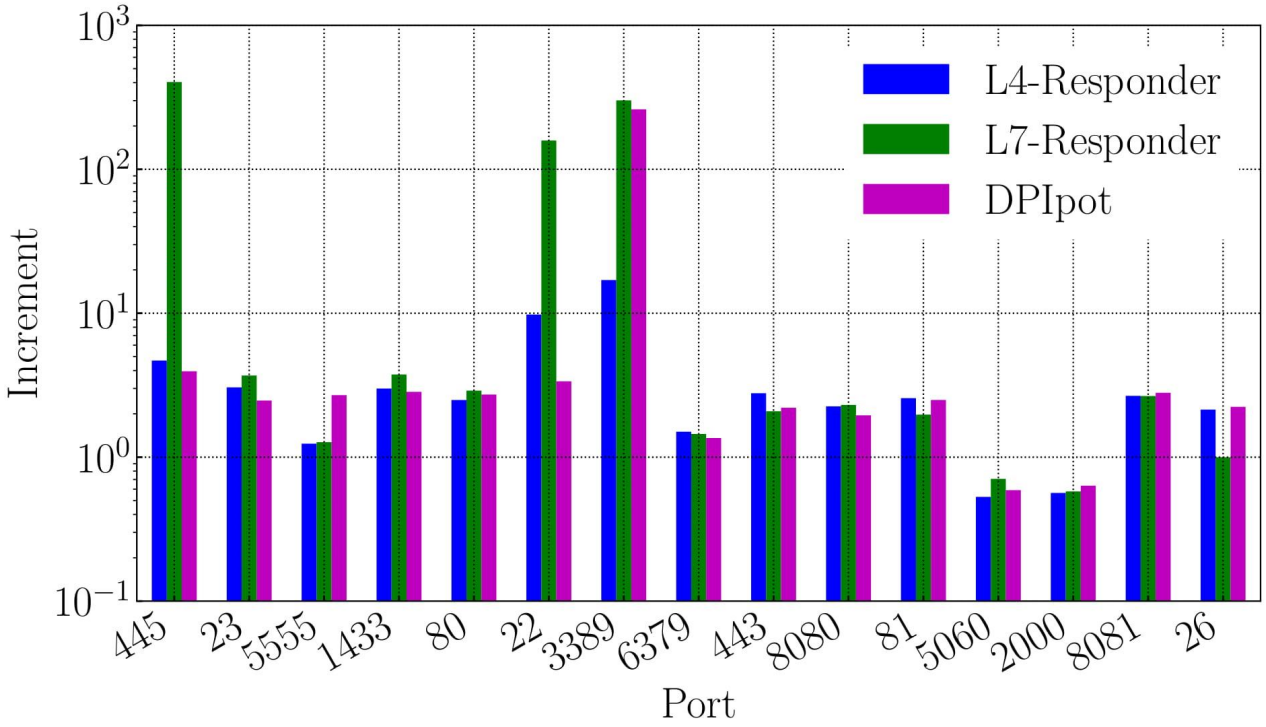
- **Phase 1**: only SYN
- **Phase 2**: three way handshake **incomplete** [SYN + SYN/ACK only]
- **Phase 3:** three way handshake complete **without** payload
- **Phase 4**: three way handshake complete **with** payload





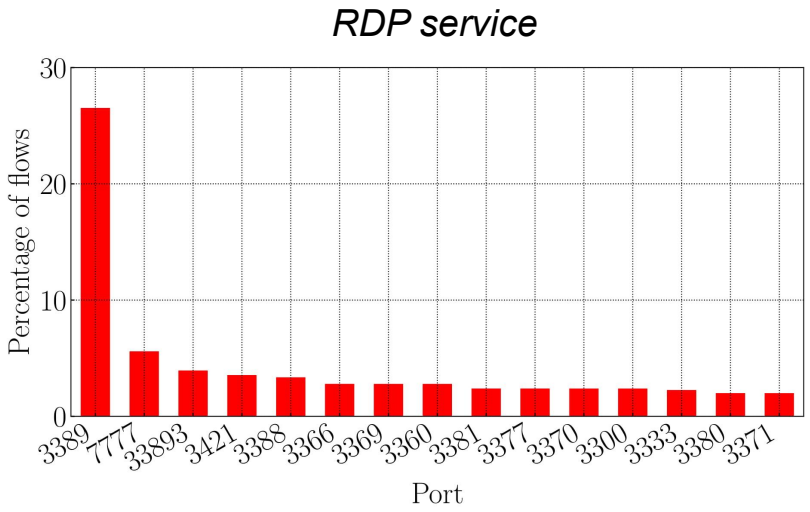
Increment in traffic when we start replying

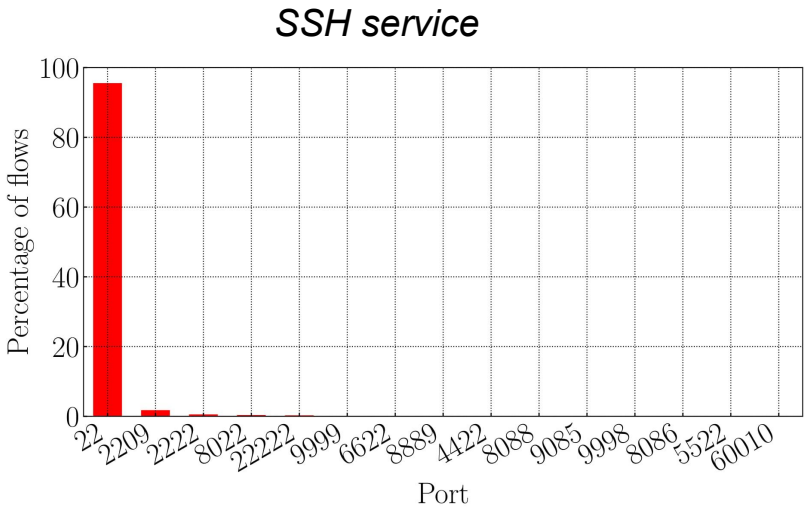
Does the attack pattern change if we start replying to all requests?



Significant increase w.r.t. darknet

Does identifying protocols on-the-fly influence the attack patterns?





We observe that attacks on non-standard ports are very common for some services

Future works

- Enlarge our observation window to a **larger period of time**
- Definition of **new scenarios**
- **Extend the set of protocols** supported by DPIpot to all the 100 protocols implemented by nDPI
- **Improve the performances** of DPIpot in order to support more connections simultaneously

Thank you!

Questions?