Mercoledì 9 **Aprile** 2025
ore 9:30 – 18:00

neacademy

Deep Packet Inspection nell'analisi del traffico di rete: Cybersecurity, fingerprinting e classificazione.

Luca Deri <deri@ntop.org>

Who am I

- ntop founder (http://www.ntop.org): company that develops open-source network security and visibility tools.
- Author of various open source software tools and contributor to popular tools (e.g. Suricata and Wireshark).
- Lecturer at the CS Dept, University of Pisa, Italy.

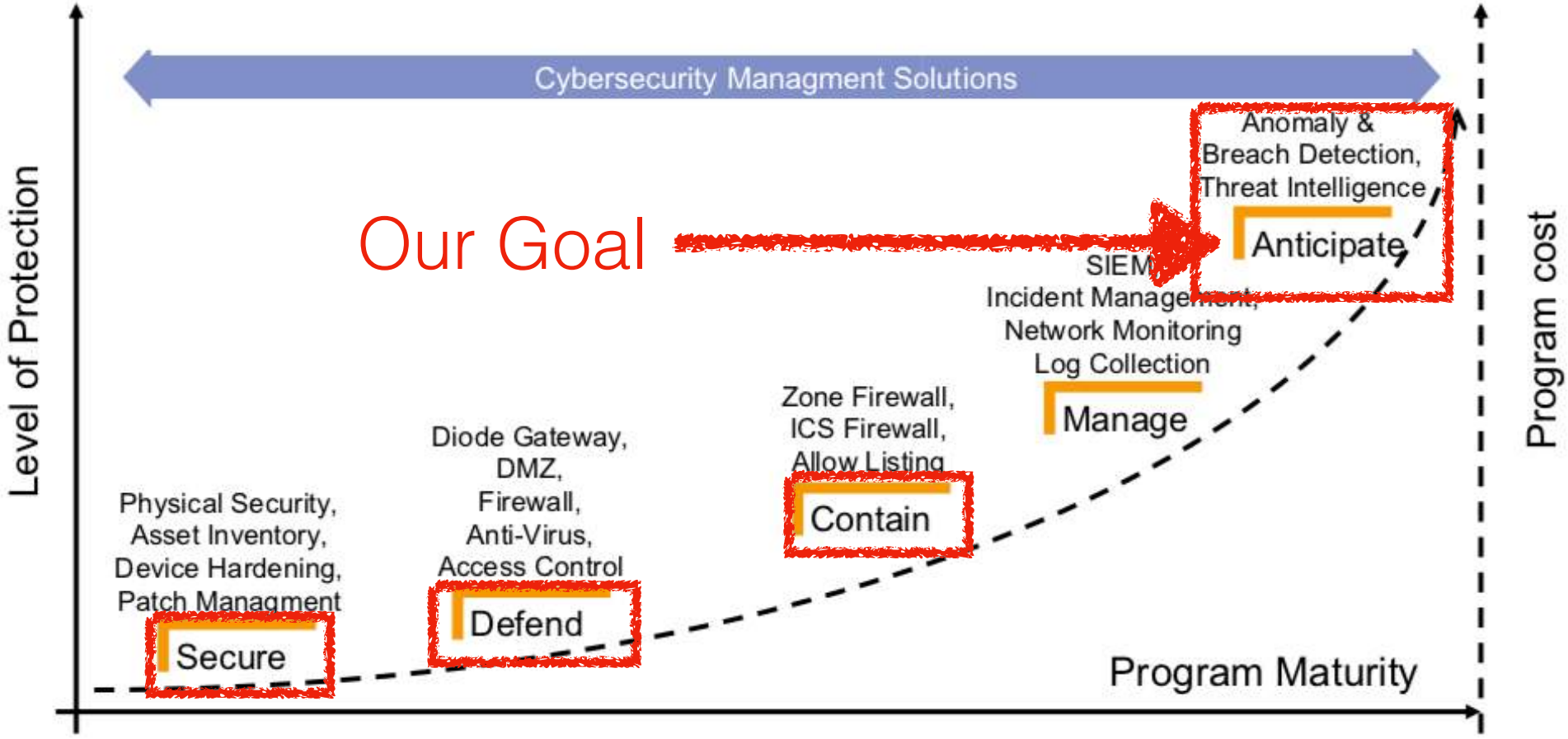
Monitoring Requirements

- Internet Service Providers
  - Prevent the network from collapsing (mostly DDoS).
  - Visibility of the main network activities in order to understand traffic flows (routing/AS-level, not host).
  - Device monitoring (interface drops, state changes).
- Service/Cloud/Hosting Providers
  - Monitor core services (e.g. DNS, email).
  - Detect severe source of troubles (e.g. heavy spammers) in order to avoid decreasing the overall network reputation.

Cybersecurity in Datacenters

- Contrary to companies where everything has to be policed, in ISPs and Providers the goal is NOT to completely cleanup traffic but keep the network infrastructure healthy by:
  - Mitigating volumetric attacks.
  - Identify and quarantine infected hosts that are potentially dangerous for the whole infrastructure.
  - Block/report suspicious activities by providing customers a detailed report in order them to address the issue.

Monitoring Goal: Anticipate



Picture courtesy of switch.ch

What is DPI ?

- DPI (Deep Packet Inspection) enables the inspection of packet payload in order to extract metadata and characterize traffic.
- Commercial DPI libraries are often quite expensive in price, and do not cope with high-speed (> 10 Gbit).
- Network administrators are used (often due to limitations of leading hardware manufacturers) to monitor sampled data with not DPI information.
- In 2025 we need full visibility with DPI and ETA.

Welcome to nDPI [1/2]

- C-based open-source library providing:
  - Deep packet inspection engine for network visibility: protocol classification (440+), metadata extraction, flow risks computation
    - Basic blocks for a cyber-security application
    - Flow risks: an indication that in the flow there is something unusual/ dangerous to pay attention to
      - ~60 different flow risks: self-signed certificate, possible SQL/RCE injection, suspicious DGA domain, invalid character in SNI...
  - Algorithms for data analysis: data forecasting, anomaly detection, clustering and similarity evaluation, (sub-)string searching and IP matching, probabilistic data structures,...
- Available on GitHub, LGPL v3

Welcome to nDPI [2/2]

- Each protocol is identified as <major>.<minor> protocol. Example:
  - DNS.Facebook
  - QUIC.YouTube and QUIC.YouTubeUpload

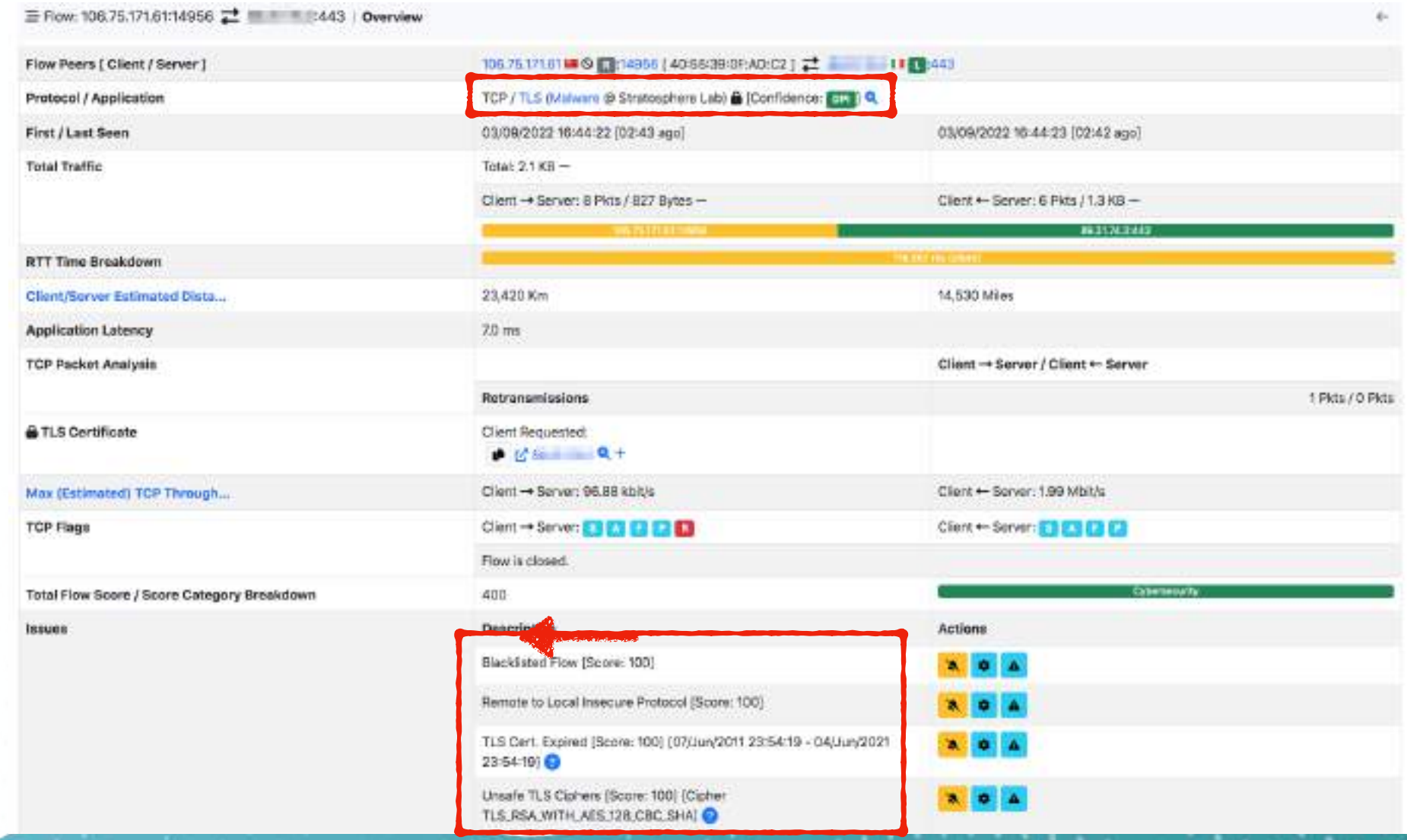
- Caveat: Zoom or WhatsApp are application protocols in the nDPI world but not for IETF.
- The first question people ask when they have to evaluate a DPI toolkit is: how many protocol do you support? This is not the right question.

nDPI Flow Risks

- HTTP suspicious user-agent
- HTTP numeric IP host contacted
- HTTP suspicious URL
- HTTP suspicious protocol header
- TLS connections not carrying HTTPS (e.g. a VPN over TLS)
- Suspicious DGA domain contacted
- Malformed packet
- SSH/SMB obsolete protocol or application version
- TLS suspicious ESNI usage
- Unsafe Protocol used
- Suspicious DNS traffic
- TLS with no SNI
- XSS (Cross Site Scripting)
- SQL Injection
- Arbitrary Code Injection/Execution
- Binary/.exe application transfer (e.g. in HTTP)
- Known protocol on non standard port
- TLS self-signed certificate
- TLS obsolete version
- TLS weak cipher
- TLS certificate expired
- TLS certificate mismatch
- DNS suspicious traffic
- HTTP suspicious content
- Risky ASN
- Risky Domain Name
- Malicious JA3 Fingerprint
- Malicious SHA1 Certificate
- Desktop of File Sharing Session
- TLS Uncommon ALPN
- TLS Certificate Validity Too Long
- Suspicious TLS Extension
- TLS Fatal Alert
- Suspicious Protocol traffic Entropy
- Clear-text Credentials Exchanged
- DNS Large Packet
- DNS Fragmented Traffic
- Invalid Characters Detected
- Possible Exploit Detected
- TLS Certificate Close to Expire
- Punycode/IDN Domain
- Error Code Detected
- Crawler/Bot Detected
- Anonymous Subscriber
- Unidirectional Traffic
- HTTP Obsolete Server
- ALPN/SNI Mismatch
- Client Contacted A Malware Host
- Binary File/Data Transfer (Attempt)
- Probing Attempt
- Obfuscated Traffic

Legenda: Clear Text Only, Encrypted/Plain Text, Encrypted Only

Combining Visibility with ETA



Fingerprinting Methods

- Protocol Fingerprint
  - Analyze a specific protocol (e.g. DHCP fingerprint, or TCP behavior for OS fingerprinting) in order to compute the expected fingerprint. Example: Window hosts do not set the Timestamps option in TCP SYN packets.
- Content Fingerprint
  - Create the fingerprint based on the content of specific protocol. Examples:
    - HTTP User-Agent
    - Android vs iOS vs Windows can be passively detected looking at DNS domain names queries (e.g. thinkdifferent.us and connectivitycheck.android.com)
    - Firefox connects via TLS to firefox.settings.services.mozilla.com

Using Fingerprinting in Real Life

- Browser fingerprinting
  Collects information about a web browser and device where it's running on including browser type, version, operating system, screen resolution, installed plugins. This creates a unique “fingerprint” that can be used to track the user across different sessions and websites.
- Policy Enforcement (OS/Device Fencing)
  Restrict to specific VLANs/block old/specific devices/OSs by looking at the device MAC address or initial DHCP request. This technique plays an important role in securing OT (Operational Technology) networks.
- Traffic Prioritisation
  Disable specific traffic (e.g. Zoom Video) in case of limited available bandwidth.
- Hidden Device Detection
  Spot NAT devices or hotspots

Some Network Fingerprints

- TCP Fingerprint
- Application Fingerprint
  - TLS/QUIC (JA3/JA4) and Web Browser Fingerprint
  - DHCP
  - RDP (Remote Desktop Protocol)
  - SSH (Secure Shell)
  - DHCP (Dynamic Host Configuration Protocol)
  - OpenVPNs (and dialects)
  - Obfuscated TLS (encrypted tunnels based on a TLS dialect)
  - Fully Encrypted Protocols (ShadowSocks, VMess, Trojan,...)

Using Fingerprints

```
static struct os_fingerprint tcp_fps[] = {
    { "2_64_65535_8bf9e292397e",    ndpi_os_freebsd     },

    { "2_64_64800_83b2f9a5576c",    ndpi_os_linux       },
    { "2_64_64240_2e3cee914fc1",    ndpi_os_linux       },
    { "2_64_29200_2e3cee914fc1",    ndpi_os_linux       },
    { "2_64_29200_d853e95bd80f",    ndpi_os_linux       }, /* Sonos */
    { "2_64_14600_8c07a80cc645",    ndpi_os_linux       }, /* QNAP */
    { "2_64_64240_2e3cee914fc1",    ndpi_os_linux       }, /* rPI */
    { "2_64_32120_2e3cee914fc1",    ndpi_os_linux       }, /* rPI */
    { "2_64_29200_90541420d839",    ndpi_os_linux       }, /* Suse Linux */
    { "2_64_64240_41a9d5af7dd3",    ndpi_os_linux       },

    { "2_64_65535_d876f498b09e",    ndpi_os_android     },
    { "2_64_65535_685ad951a756",    ndpi_os_android     },
    { "2_64_65535_41a9d5af7dd3",    ndpi_os_android     },
    { "2_64_65535_148107a0d970",    ndpi_os_android     },
    { "2_64_65535_f518bfb025b0",    ndpi_os_android     },

    { "2_128_64240_6bb88f5575fd",   ndpi_os_windows     },
    { "2_128_8192_4697958db063",    ndpi_os_windows     }, /* Windows 7 */
```

```
local ja4_db = {
   ['02e81d9f7c9f_736b2a1ed4d3'] = 'Chrome',
   ['07be0c029dc8_ad97e2351c08'] = 'Firefox',
   ['07be0c029dc8_d267a5f792d4'] = 'Firefox',
   ['0a330963ad8f_c905abbc9856'] = 'Chrome',
   ['0a330963ad8f_c9eaec7dbab4'] = 'Chrome',
   ['168bb377f8c8_a1e935682795'] = 'Anydesk',
   ['24fc43eb1c96_14788d8d241b'] = 'Chrome',
   ['24fc43eb1c96_14788d8d241b'] = 'Safari',
   ['24fc43eb1c96_845d286b0d67'] = 'Chrome',
   ['24fc43eb1c96_845d286b0d67'] = 'Safari',
   ['24fc43eb1c96_c5b8c5b1cdcb'] = 'Safari',
   ['2a284e3b0c56_12b7a1cb7c36'] = 'Safari',
   ['2a284e3b0c56_f05fdf8c38a9'] = 'Safari',
   ['2b729b4bf6f3_9e7b989ebec8'] = 'IcedID',
   ['39b11509324c_ab57fa081356'] = 'Chrome',
   ['39b11509324c_c905abbc9856'] = 'Chrome',
   ['39b11509324c_c9eaec7dbab4'] = 'Chrome',
   ['41f4ea5be9c2_06a4338d0495'] = 'Chrome',
```

Fingerprints enable accurate device detection

iOS/iPadOS/macOS (Intel)

- Send SYN+ECE+CRW. Others (including macOS Silicon) just SYN.
- Options (iOS but not iPadOS) end with a double EOL.

Windows

- Does not use the timestamp (8) option.
- Has a default TTL of 128, vs 64 used on Linux etc.

Fingerprinting in Cybersecurity

- Fingerprinting plays a crucial role in cyber security as it helps in detecting threats, securing networks, and implementing targeted security measures.
- Defenders:
  - Match malware signatures (e.g. TLS fingerprint or SSL certificate hash) and block malicious traffic.
  - Prevents massive scanners from exploring network services.
- Attackers
  - Use fingerprinting to detect flaws (e.g. CVEs) that can be used to attack the system.
  - During reconnaissance, identify application/OS version in order to target attacks towards weak victims.

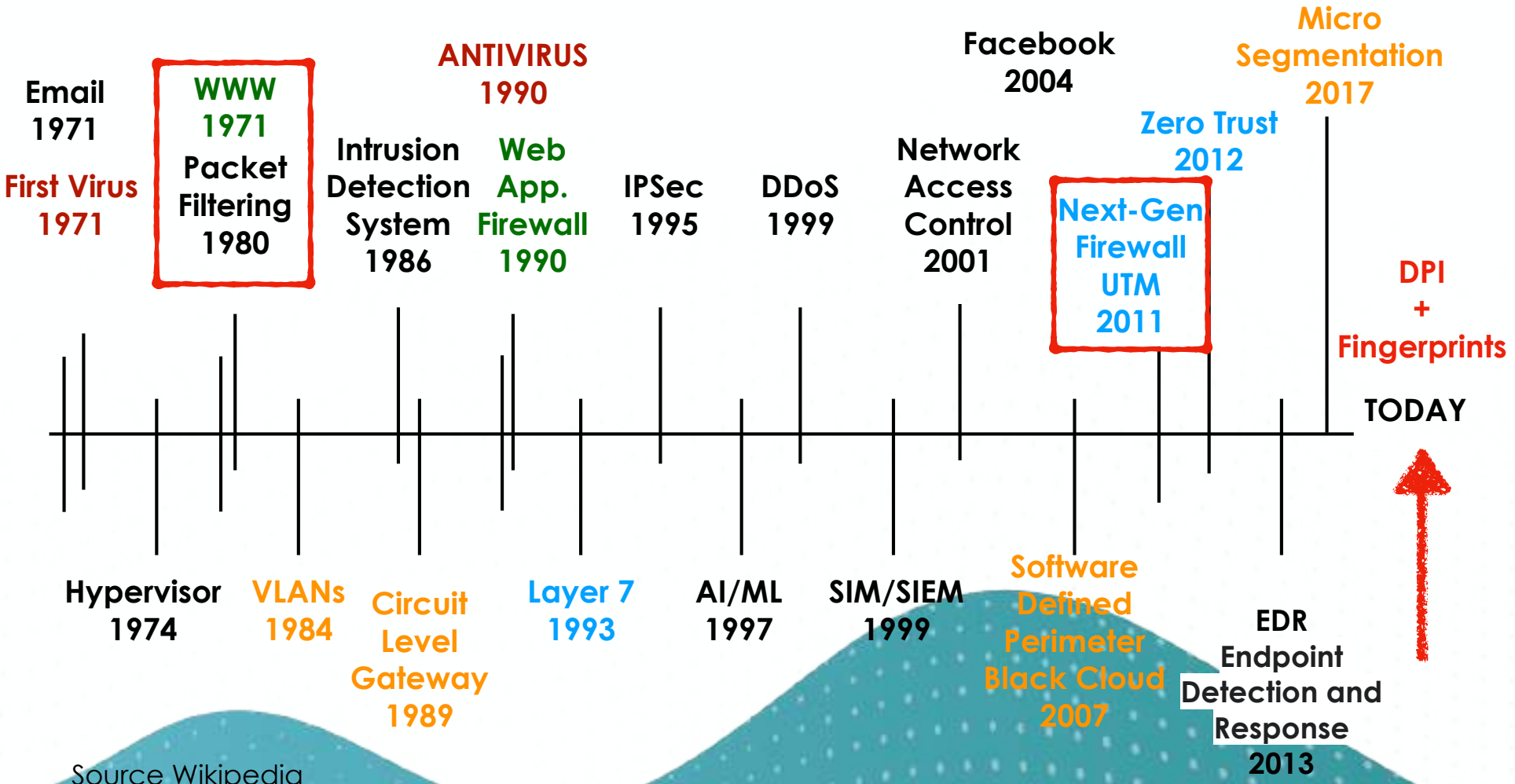
Anticipate Problems [1/4]

- Firewalls evolved:
  - IP-header based rules (ACL) - 1980
  - Next-generation Firewalls (L7 protocol) - 2011
- Traffic fingerprinting refers to the process of identifying and gathering specific information about a system or network to create a (in theory) unique profile or “fingerprint”.
- As fingerprints are created on the initial few traffic bytes, blocking malicious fingerprints means that we can stop threats before they hit the network.

Anticipate Problems [2/4]

- Supported Fingerprints
  - (Anonymous) VPNs (e.g. OpenVPN)
  - Malicious QUIC/TLS applications
  - SSH-based Bots
  - Outdated/unwanted devices (DHCP)
  - Unknown and Encrypted Protocols
  - Cryptominers

Anticipate Problems [3/4]



Source Wikipedia

Anticipate Problems [4/4]

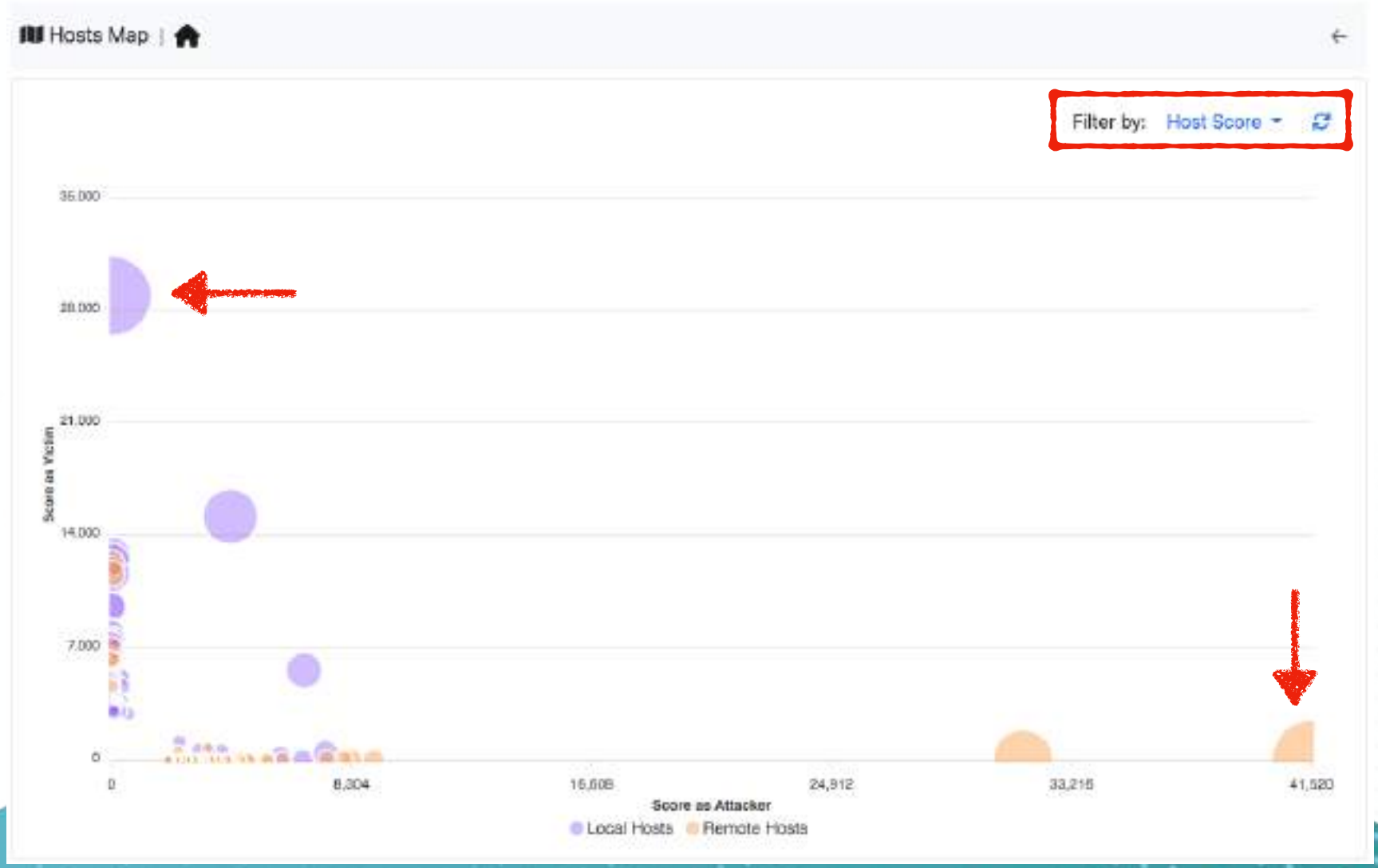
```
Frame 1: 60 bytes on wire (480 bits), 60 bytes captured (480
Ethernet II, Src: 76:ac:b9:35:30:da (76:ac:b9:35:30:da), Dst:
Internet Protocol Version 4, Src: 192.168.10.145 (192.168.10.
Transmission Control Protocol, Src Port: 49175, Dst Port: 88
    Source Port: 49175
    Destination Port: 8888
    [Stream index: 0]
    [Stream Packet Number: 1]
  > [Conversation completeness: Incomplete (35)]
    [TCP Segment Len: 0]
    Sequence Number: 0    (relative sequence number)
    Sequence Number (raw): 253744456
    [Next Sequence Number: 1    (relative sequence number)]
    Acknowledgment Number: 0
    Acknowledgment number (raw): 0
    0101 .... = Header Length: 20 bytes (5)
  > Flags: 0x002 (SYN)
    Window: 65535
    [Calculated window size: 65535]
    Checksum: 0x5297 [unverified]
    [Checksum Status: Unverified]
    Urgent Pointer: 0
  > [Timestamps]
```

```
Frame 1: 60 bytes on wire (480 bits), 60 bytes captured (480 bits)
Ethernet II, Src: 76:ac:b9:35:30:da (76:ac:b9:35:30:da), Dst: PCSSyste
Internet Protocol Version 4, Src: 192.168.10.145 (192.168.10.145), Dst
Transmission Control Protocol, Src Port: 46998, Dst Port: 8888, Seq: 0
    Source Port: 46998
    Destination Port: 8888
    [Stream index: 0]
    [Stream Packet Number: 1]
  > [Conversation completeness: Incomplete (35)]
    [TCP Segment Len: 0]
    Sequence Number: 0    (relative sequence number)
    Sequence Number (raw): 1163206847
    [Next Sequence Number: 1    (relative sequence number)]
    Acknowledgment Number: 0
    Acknowledgment number (raw): 0
    0101 .... = Header Length: 20 bytes (5)
  > Flags: 0x002 (SYN)
    Window: 1024
    [Calculated window size: 1024]
    Checksum: 0xd56b [unverified]
    [Checksum Status: Unverified]
    Urgent Pointer: 0
  > [Timestamps]
```

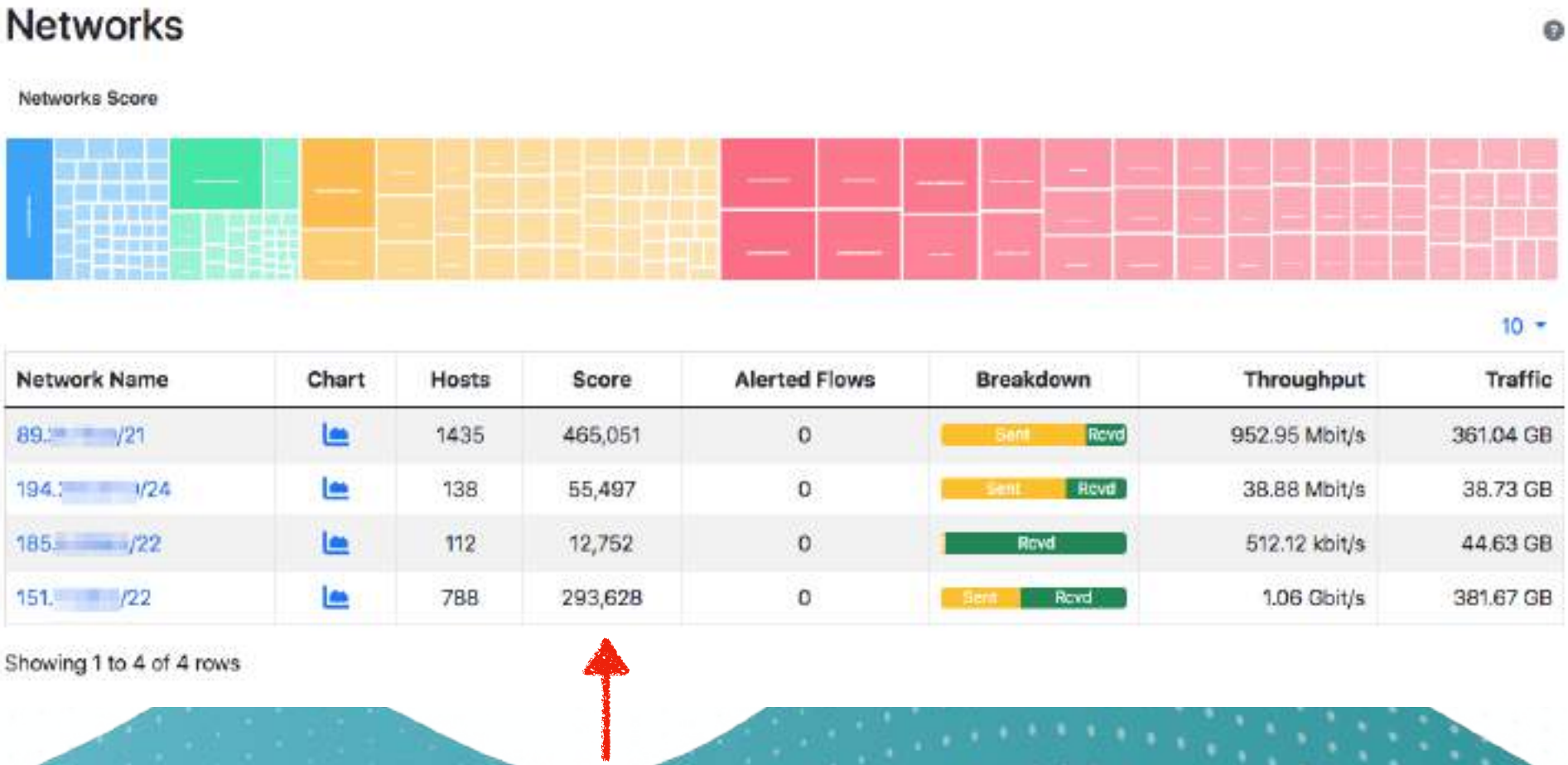
https://zmap.io/

https://github.com/robertdavidgraham/masscan

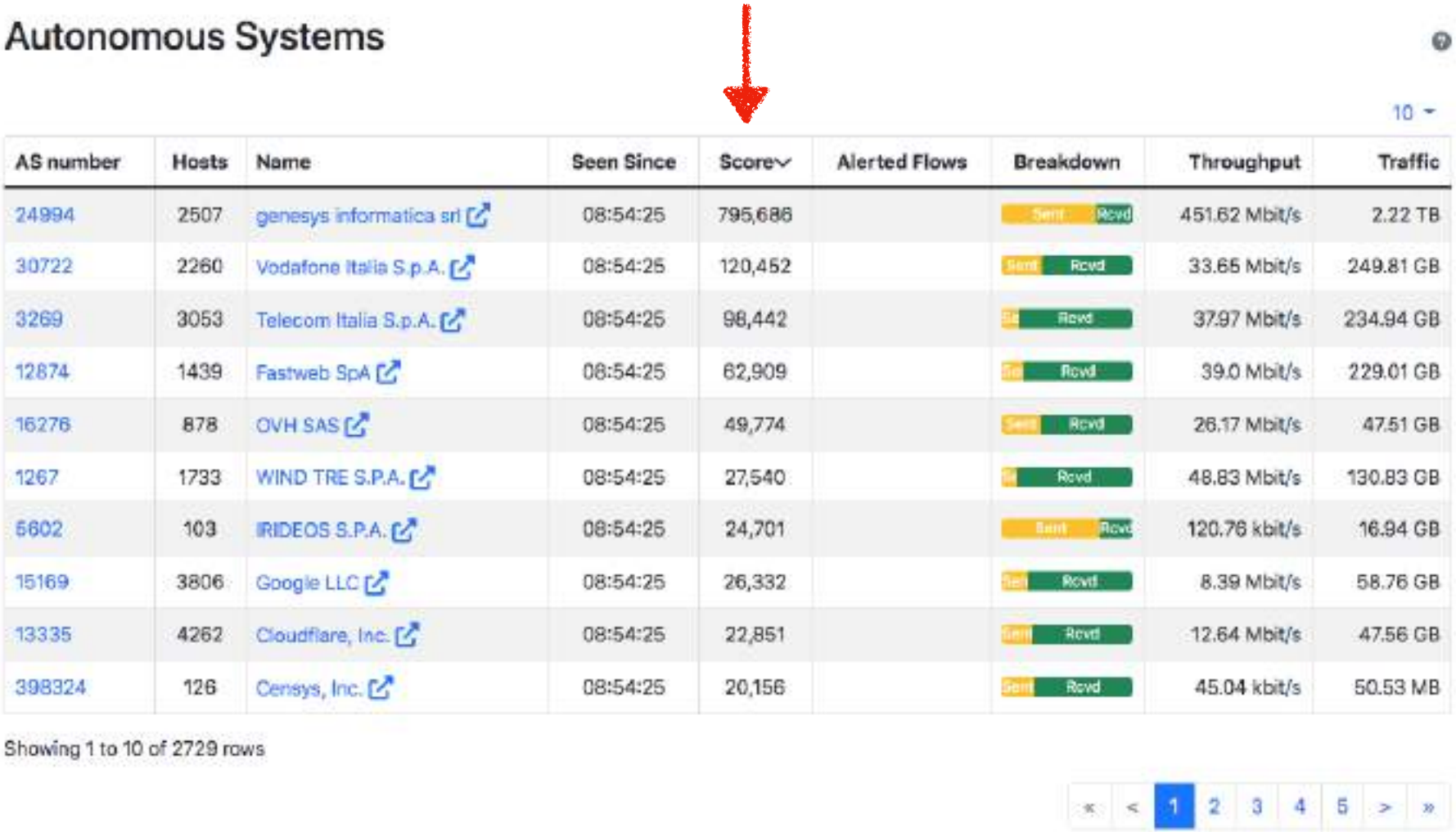
Spotting Issues [1/3]



Spotting Issues [2/3]



Spotting Issues [3/3]



Autonomous Systems

10

AS number	Hosts	Name	Seen Since	Score	Alerted Flows	Breakdown	Throughput	Traffic
24994	2507	genesys informatica srl	08:54:25	795,686		Sent Rcvd	451.62 Mbit/s	2.22 TB
30722	2260	Vodafone Italia S.p.A.	08:54:25	120,452		Sent Rcvd	33.66 Mbit/s	249.81 GB
3269	3053	Telecom Italia S.p.A.	08:54:25	98,442		Sent Rcvd	37.97 Mbit/s	234.94 GB
12874	1439	Fastweb SpA	08:54:25	62,909		Sent Rcvd	39.0 Mbit/s	229.01 GB
16276	878	OVH SAS	08:54:25	49,774		Sent Rcvd	26.17 Mbit/s	47.51 GB
1267	1733	WIND TRE S.P.A.	08:54:25	27,540		Sent Rcvd	48.83 Mbit/s	130.83 GB
5602	103	IRIDEOS S.P.A.	08:54:25	24,701		Sent Rcvd	120.76 kbit/s	16.94 GB
15169	3806	Google LLC	08:54:25	26,332		Sent Rcvd	8.39 Mbit/s	58.76 GB
13335	4262	Cloudflare, Inc.	08:54:25	22,851		Sent Rcvd	12.64 Mbit/s	47.56 GB
398324	126	Censys, Inc.	08:54:25	20,156		Sent Rcvd	45.04 kbit/s	50.53 MB

Showing 1 to 10 of 2729 rows

« < 1 2 3 4 5 > »

What about Traffic Classification ? [1/2]

- Traffic fingerprinting allows network traffic to be clustered according to the sender OS (TCP Fingerprinting) and Application (e.g. JA3/4).

```
194_64_65535_dd5737e4fedb-t13d1516h2_8daaf6152771_9b887d9acb53 [ tiktokv.eu tiktokcdn.com snapchat.com tiktokv.com ]
194_64_65535_dd5737e4fedb-t13d1516ht_8daaf6152771_9b887d9acb53 [ tiktokv.eu ]
2_64_65535_dd5737e4fedb-t13d1516h2_8daaf6152771_e5627efa2ab1 [ googlevideo.com pinimg.com pinterest.com ]
194_64_65535_dd5737e4fedb-t13d1516h2_8daaf6152771_e5627efa2ab1 [ tiktokv.eu tiktokcdn.com snapchat.com tiktokcdn-us.com ]
194_64_65535_dd5737e4fedb-t13d181100_e8a523a41297_d5fe2c511efa [ tiktokcdn.com tiktokv.eu tiktokcdn-eu.com ]
2_64_65535_dd5737e4fedb-t13d1516h2_8daaf6152771_9b887d9acb53 [ tiktokcdn.com ]
2_64_65535_dd5737e4fedb-t12d220700_0d4ca5d4ec72_3304d8368043 [ microsoft.com ryanair.com ]
194_64_65535_dd5737e4fedb-t00d030800_55b375c5d22e_566d5108064c [ facebook.com ]
194_64_65535_dd5737e4fedb-t13d1314h2_f57a46bbacb6_1478d8d241b [ appsflyersdk.com ]
2_64_65535_dd5737e4fedb-t13d2015h2_a09f3c656075_3d00e4afe3b1 [ apple.com ]
2_64_65535_dd5737e4fedb-t00d0310h2_55b375c5d22e_50cc996d9024 [ facebook.com ]
2_64_65535_dd5737e4fedb-t00d030600_55b375c5d22e_8f5d6a331b25 [ facebook.com ]
194_64_65535_dd5737e4fedb-t13d0713gr_04ca88ad2b9b_d8054c94196c [ snapchat.com ]
194_64_65535_dd5737e4fedb-t13d181100_e8a523a41297_ef7df7f74e48 [ tiktokcdn-eu.com ]
194_64_65535_dd5737e4fedb-t13d2015h2_a09f3c656075_3d00e4afe3b1 [ apple.com ]
194_64_65535_dd5737e4fedb-t13d2014ht_a09f3c656075_14788d8d241b [ apple.com icloud.com ]
2_64_65535_dd5737e4fedb-t13d2014ht_a09f3c656075_14788d8d241b [ apple.com spotify.com cdn-apple.com ]
194_64_65535_d3a424420f2a-t13d2015h2_a09f3c656075_3d00e4afe3b1 [ icloud.com apple.com ]
2_64_0_dd5737e4fedb-t13d2014ht_a09f3c656075_14788d8d241b [ apple.com ]
2_64_65535_dd5737e4fedb-t12d220600_0d4ca5d4ec72_3304d8368043 [ ]
2_64_65535_d3a424420f2a-t13d2015h2_a09f3c656075_3d00e4afe3b1 [ apple.com ]
194_64_65535_dd5737e4fedb-t13d0311ap_55b375c5d22e_14aed462abe7 [ apple.com ]
194_64_65535_dd5737e4fedb-t13d181200_e8a523a41297_02c8e53ee398 [ tiktokcdn-eu.com ]
194_64_0_dd5737e4fedb-t13d2014h2_a09f3c656075_14788d8d241b [ icloud.com ]
```

What about Traffic Classification ? [2/2]

- Ok but what is tiktok.com or neacademy.it ?
- Is it possible to classify automatically traffic content ?
- AI can definitely help to do it automatically:
  - Domain name classification:< 5 sec including download time on a host without a GPU.
  - Multi-language support

```
deri@super 218> duckdb ./domains.duck
v1.2.1 8e52ec4395
Enter ".help" for usage hints.
D SELECT domain,category,description FROM domains where domain = 'neacademy.it';
```

domain varchar	category varchar	description varchar
neacademy.it	education	neacademy.it is an educational website that offers a variety of courses focusing on technology, programming, data science, an…

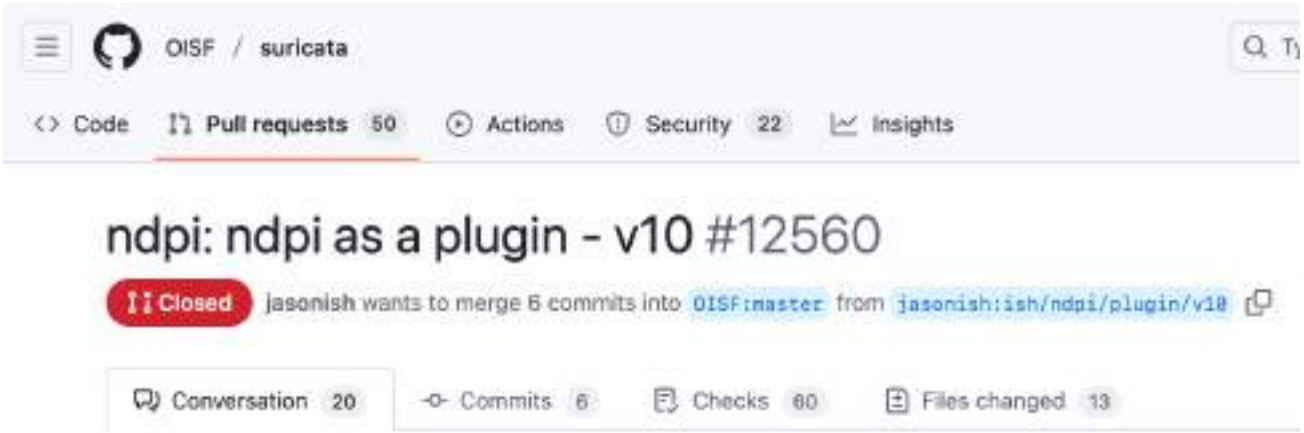
```
D
```

nDPI in Real Life

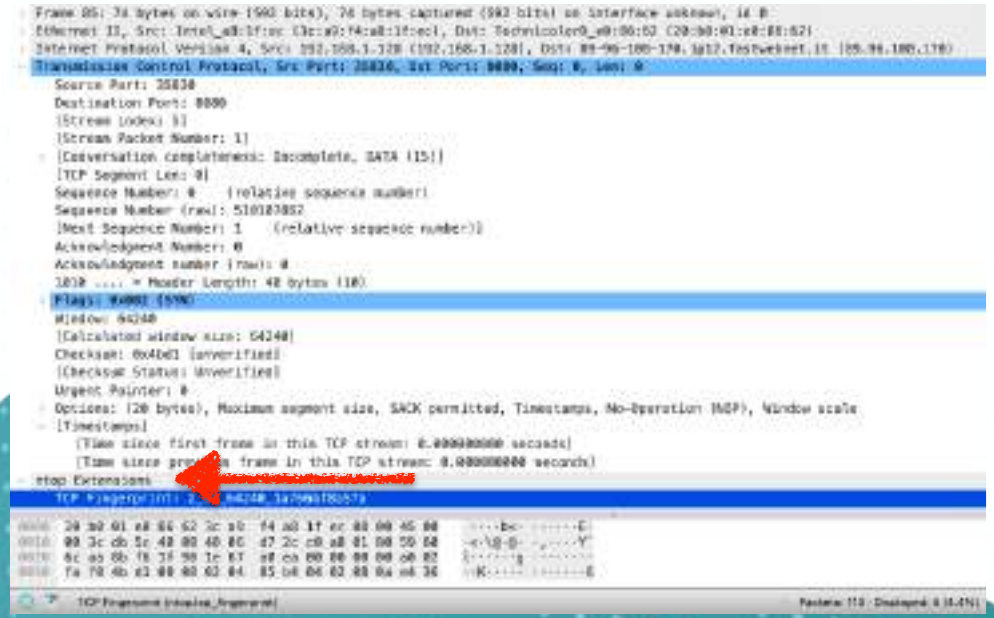
- Block or Prioritize Traffic on Linux
- 100 Gbit Traffic Analysis and Policer
- Policy Traffic based on
  - Application Protocol
  - Operating System
  - Traffic risks
  - Encrypted Traffic Analysis

nDPI in OpenSource Tools

- Suricata IDS (suricata.io)



- Wireshark (wireshark.org)



Thank You, and See you at PacketFest

May 7-9, Zürich, Switzerland

https://www.packetfest.ch