

User-Centered Visual Analysis of PCAP Data

Dr. Alex Ulmer



Dr. Alex Ulmer

Research: User-Centered Data Science and Progressive Visual Analytics

Application Area: Cybersecurity

Visual Analysis of PCAP Data

1. Introduction & Agenda	1 min	
2. Fraunhofer – Applied Research	2 min	< 10
3. Data Science and Visual Analytics	6 min	
4. Visual Analysis of PCAP Data - Live	15 min	
5. Application Classification of Packets	7 min	
6. Explainable AI	9 min	
7. Visualization Plugin for Wireshark	5 min	
		<hr/> 45 ✓

2. Fraunhofer – Applied Research



Explore research ideas and get them market ready

Collaborate with businesses
Create an advantage in the market

76 Institutes
@ 192 locations in Germany

Topics:

- Digitalization
- Energy
- Health and Medicine
- Mobility and Transport
- Work and Management
- Mechanics and Production
- Society and Security





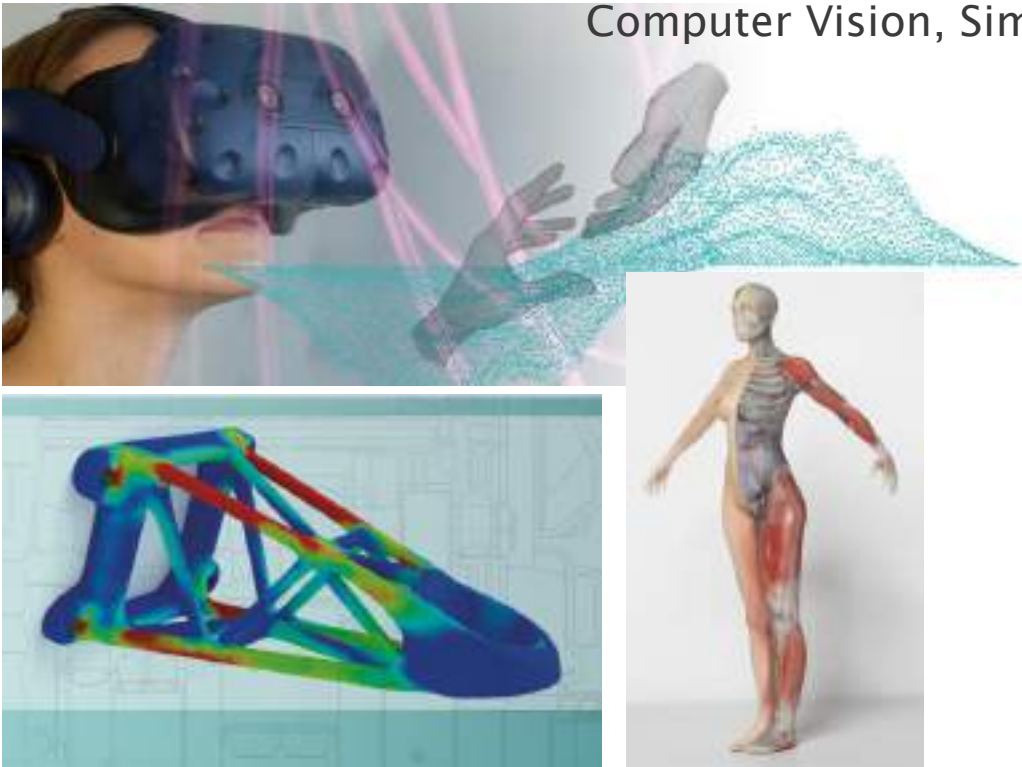
Fraunhofer

IGD

40 Years of Computer Graphics Research

Computer Vision, Simulation, 3D Printing,

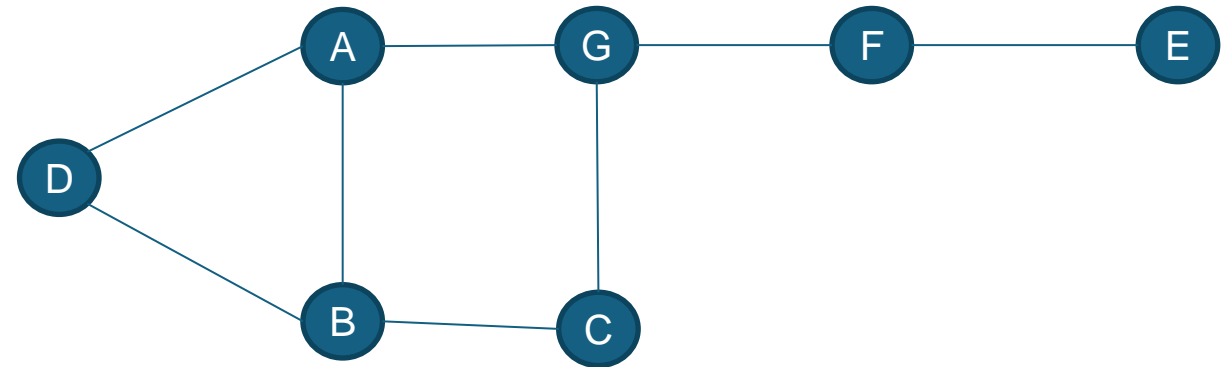
User Centered Data Science and Visual Analytics



3. Data Science and Visual Analytics

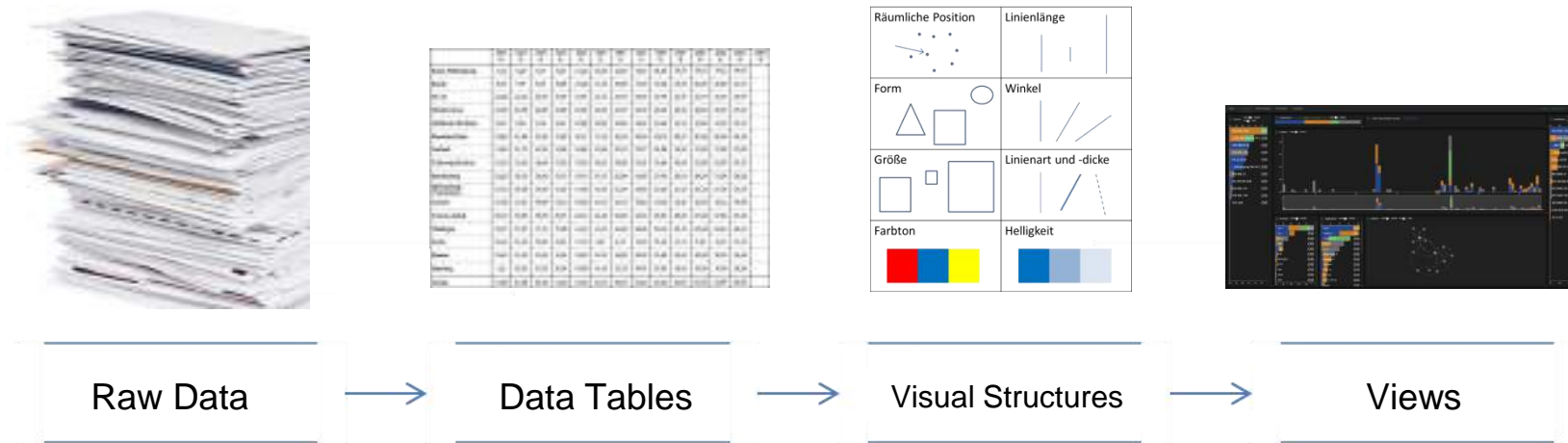
Why Visualization?

A is connected to B
B is connected to C
D is connected to A and B
E is connected to F
G is connected to A and C
F is connected to G

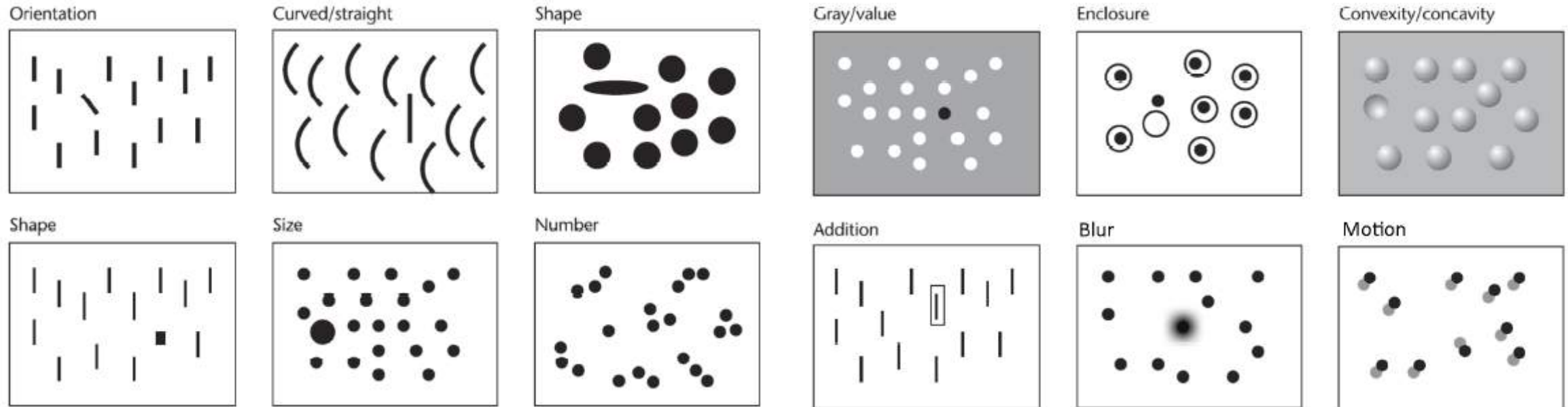




From Raw Data to Visualizations

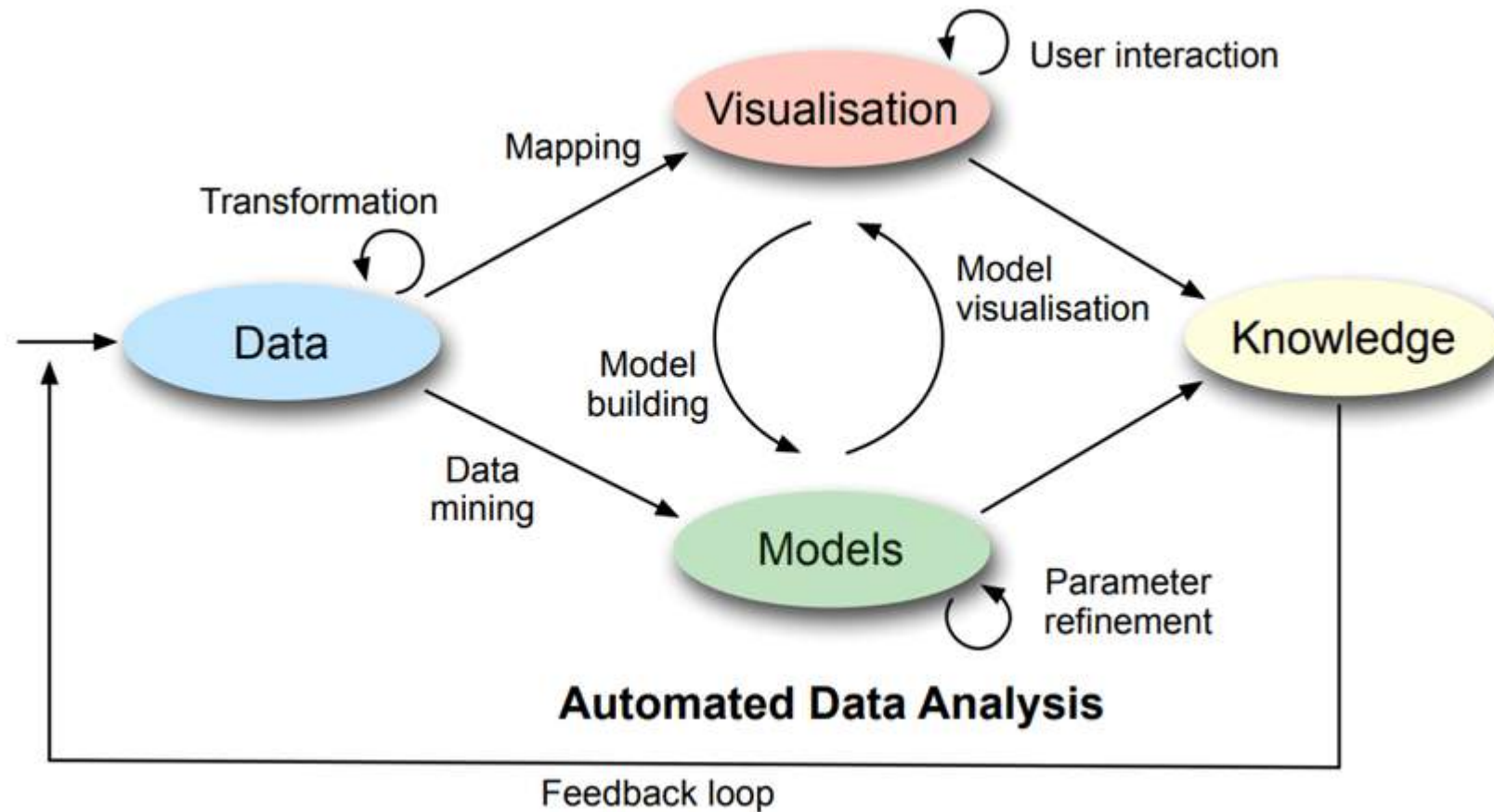


adapted from Card et al. 1999

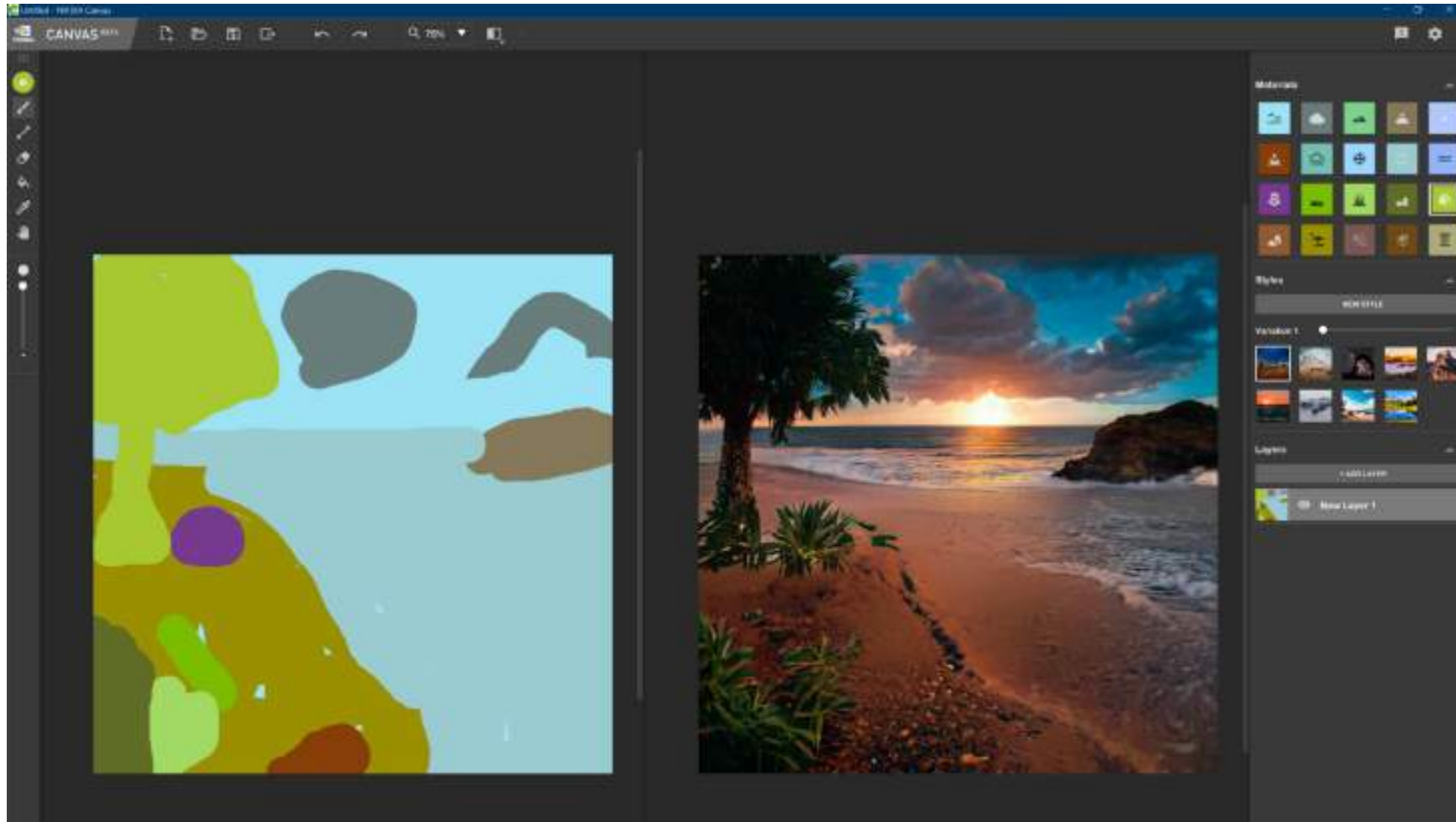


Preattentive Attributes for visual perception, adapted from Colin Ware, 2013

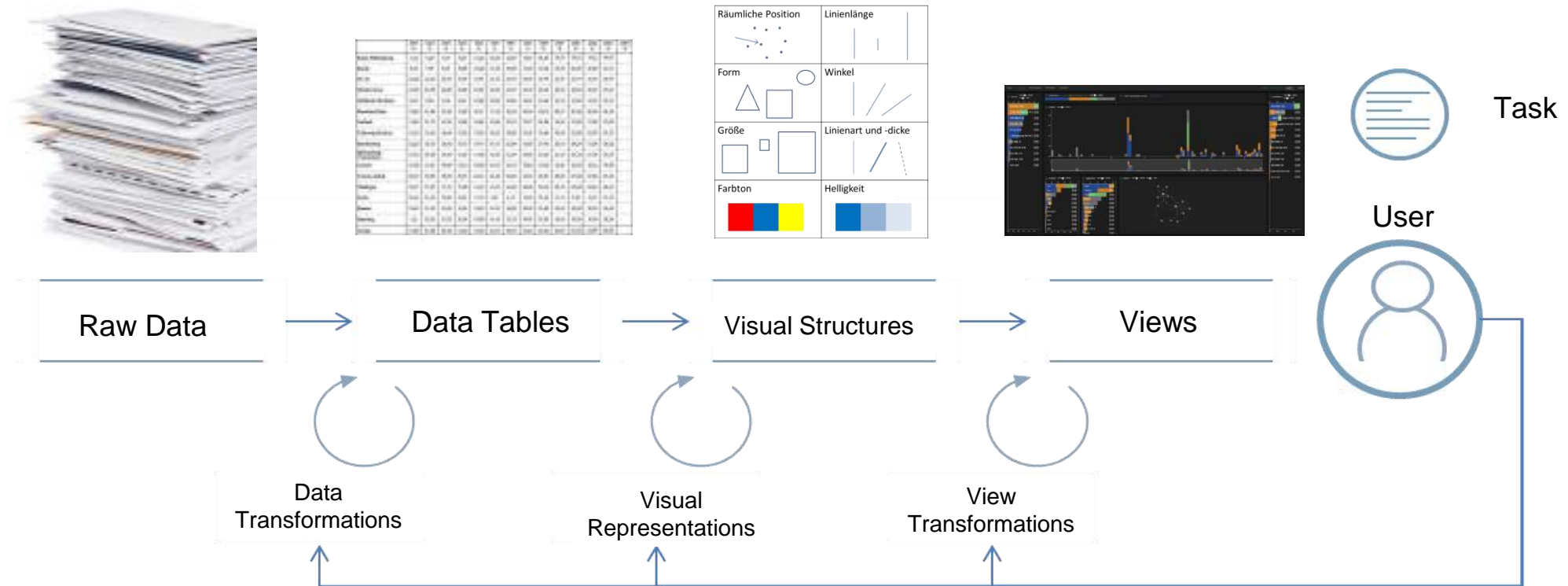
Bridging Data, Humans, and Artificial Intelligence



Visual Data Exploration, adapted from Keim et al. 2008



Change of direction - Which task needs to be completed?



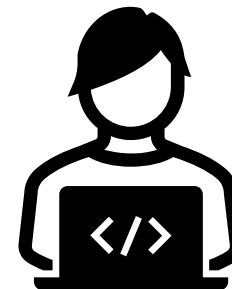
adapted from Card et al. 1999

4. Visual Analysis of PCAP Data - Live

Why did we start NetCapVis?



Wireshark is fantastic for experts



But what about beginners?
Can we make it easier to use?



<https://netcapvis.igd.fraunhofer.de>



Wireshark filter export

The screenshot shows the NetCapVis application interface. A modal dialog titled "Wireshark Filter Query:" is open, displaying the following filter query: `(frame.time >= "Aug 13, 2009 07:57:35") && (frame.time <= "Aug 13, 2009 07:58:06") && (ip.dst == 192.168.1.159) && (udp || tcp)`. Below the query is a large black button labeled "COPY TO CLIPBOARD". The background interface includes a "Wireshark Filter" header with settings and login icons, a "Count" toggle, and a list of active filters on the right. At the bottom, a timeline visualization shows data points for the specified time range.

Wireshark Filter

Wireshark Filter Query:

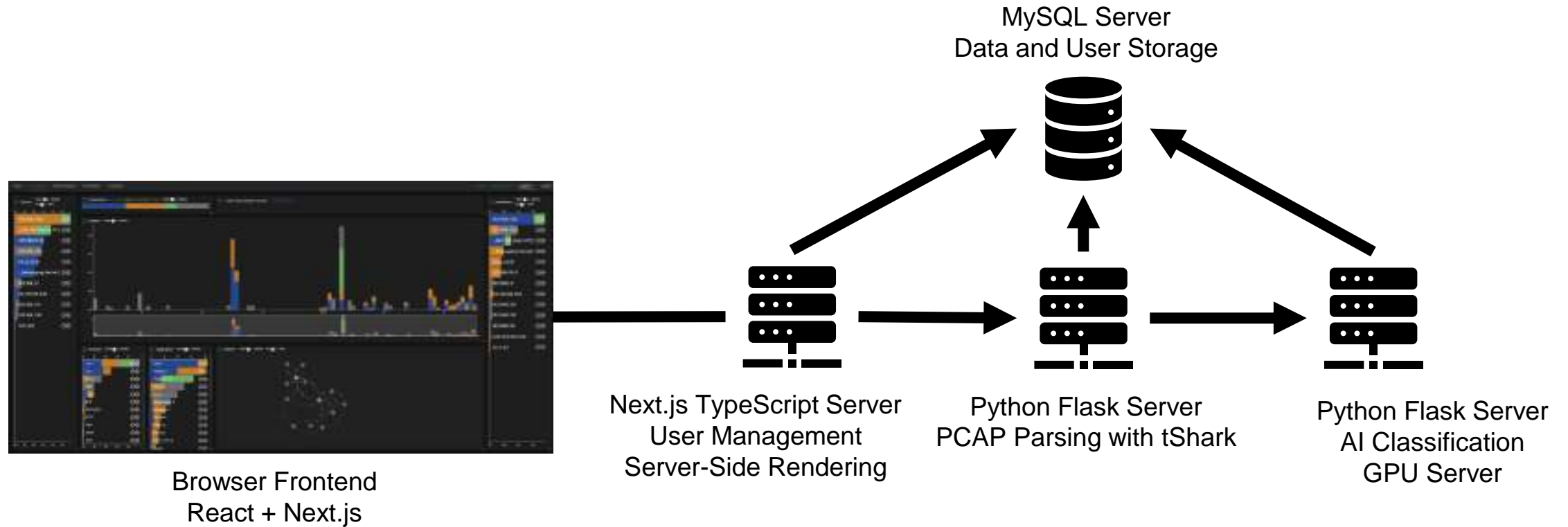
```
(frame.time >= "Aug 13, 2009 07:57:35") && (frame.time <= "Aug 13, 2009 07:58:06") && (ip.dst == 192.168.1.159) && (udp || tcp)
```

COPY TO CLIPBOARD

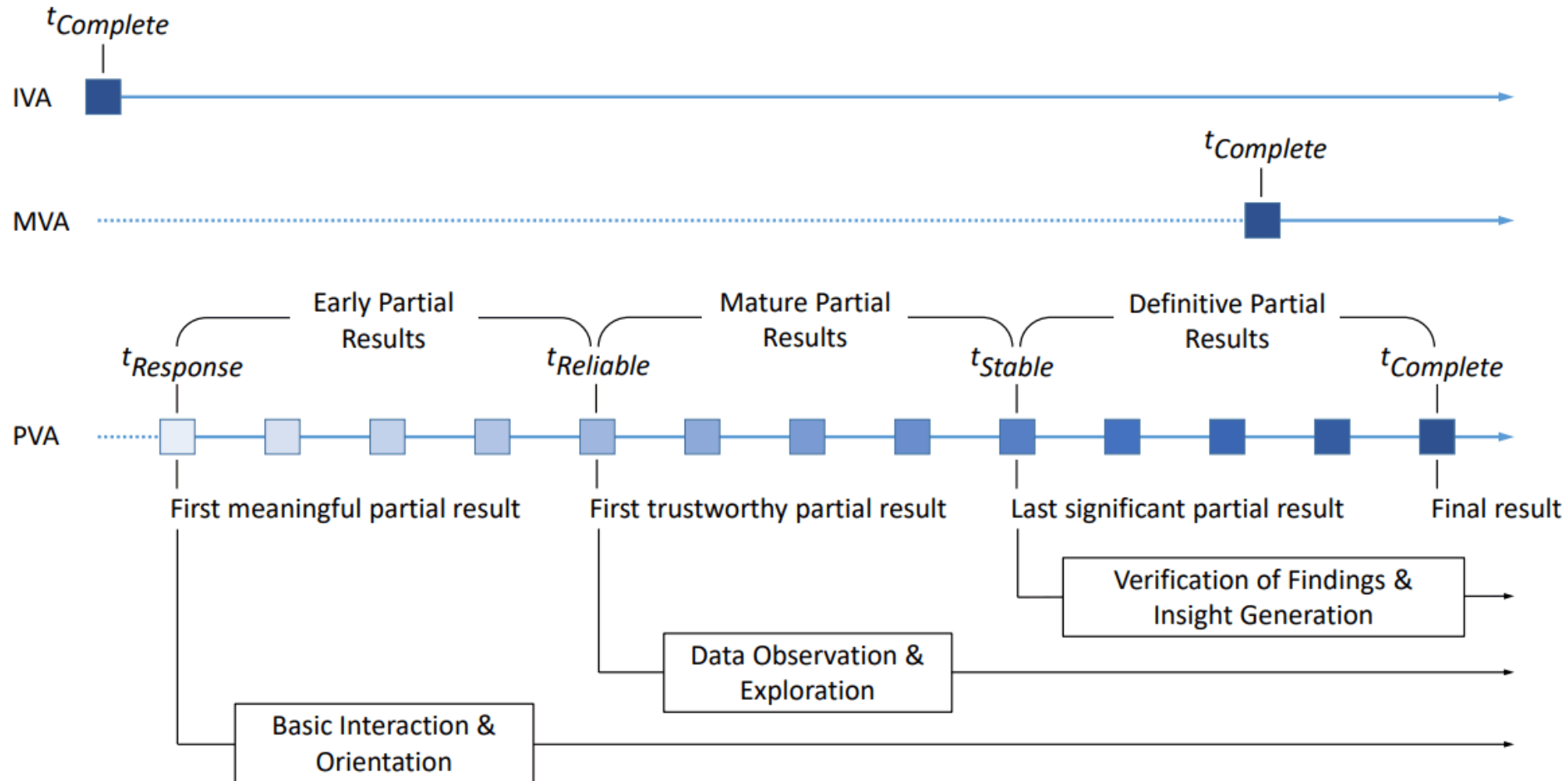
Parsing

- From Java Backend to Python with tShark
- DNS resolving included
- Working on GeoIP
- tShark optimizations needed

MySQL database to access uploaded data



Progressive Data Analysis Approach –
Load large captures
Load streams of data
with interactive visualizations at all times



Angelini, Marco, et al. "A review and characterization of progressive visual analytics" 2018.

Progressive Visual Analytics Example



Chunking a PCAP file is difficult
cutting and merging leads to corrupted data

Priority sampling for the filter that the user is setting

5. Application Classification of Packets

Machine Learning (ML)

- uncover hidden relationships in multidimensional data

Goals

- extract meaningful patterns from data
- make decisions based on those discoveries
- make predictions with AI systems

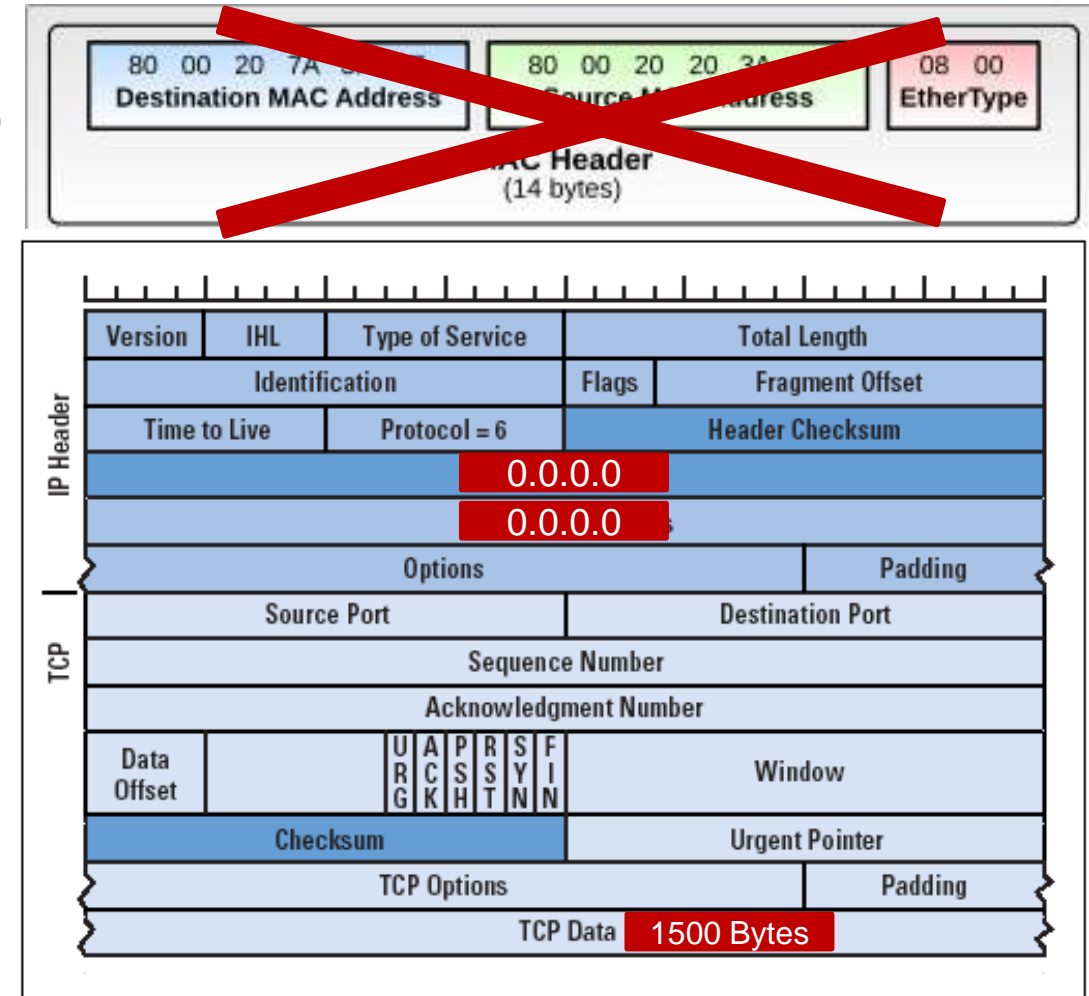
Unsupervised Learning

- Find structure in unlabeled data
- For PCAPs:
Distinguish between normal traffic vs anomaly

Supervised Learning

- Learn from labeled data (prior knowledge)
- For PCAPs:
Define specific classes, gather training data and label it

1. Remove small packets (SYN/ACK, ...)
2. Remove ethernet header
3. Set src and dst address to 0.0.0.0
4. Cut or pad payload to 1500 Bytes
5. Similar approach for other protocols



Goal:

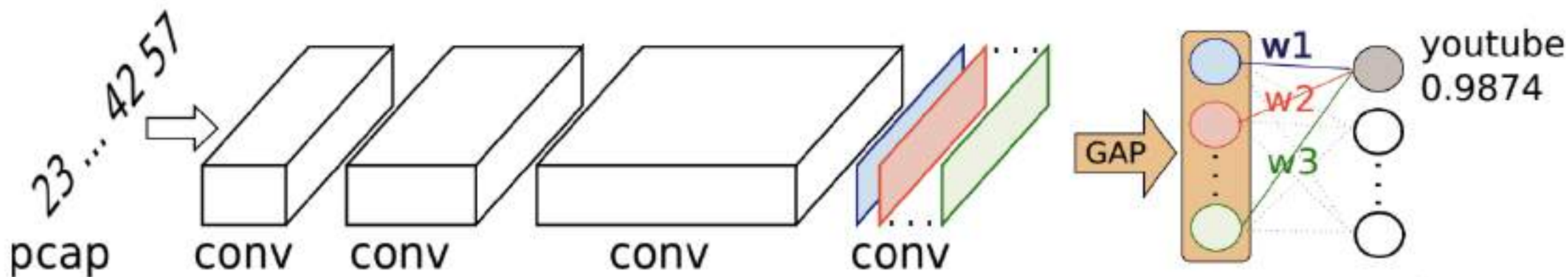
Predict which application caused a packet

Training data for the AI system



Application	Size (IPv4)	Size (IPv6)
Big Blue Button	122672	150474
Email	66564	52041
Facebook Video	111950	62030
FTPS+SFTP	103701	50589
Google Meet	90214	55059
Amazon Prime Video	123329	174930
Reddit	119221	157737
Telegram Files	42252	41187
TikTok	128739	138929
Twitch	105531	138029
Vimeo	129652	177734
Youtube	139248	179525
Zoom	53029	114624
Instagram	69872	153085
Facebook Feeds	117894	93891

Convolutional Neural Network (CNN)



NetCapVis AI Predictions

Home Dashboards Packet Analysis **AI Predictions** AI Analysis Pricing

Active Filter (260/260 Packets) [Clear All Filter](#)

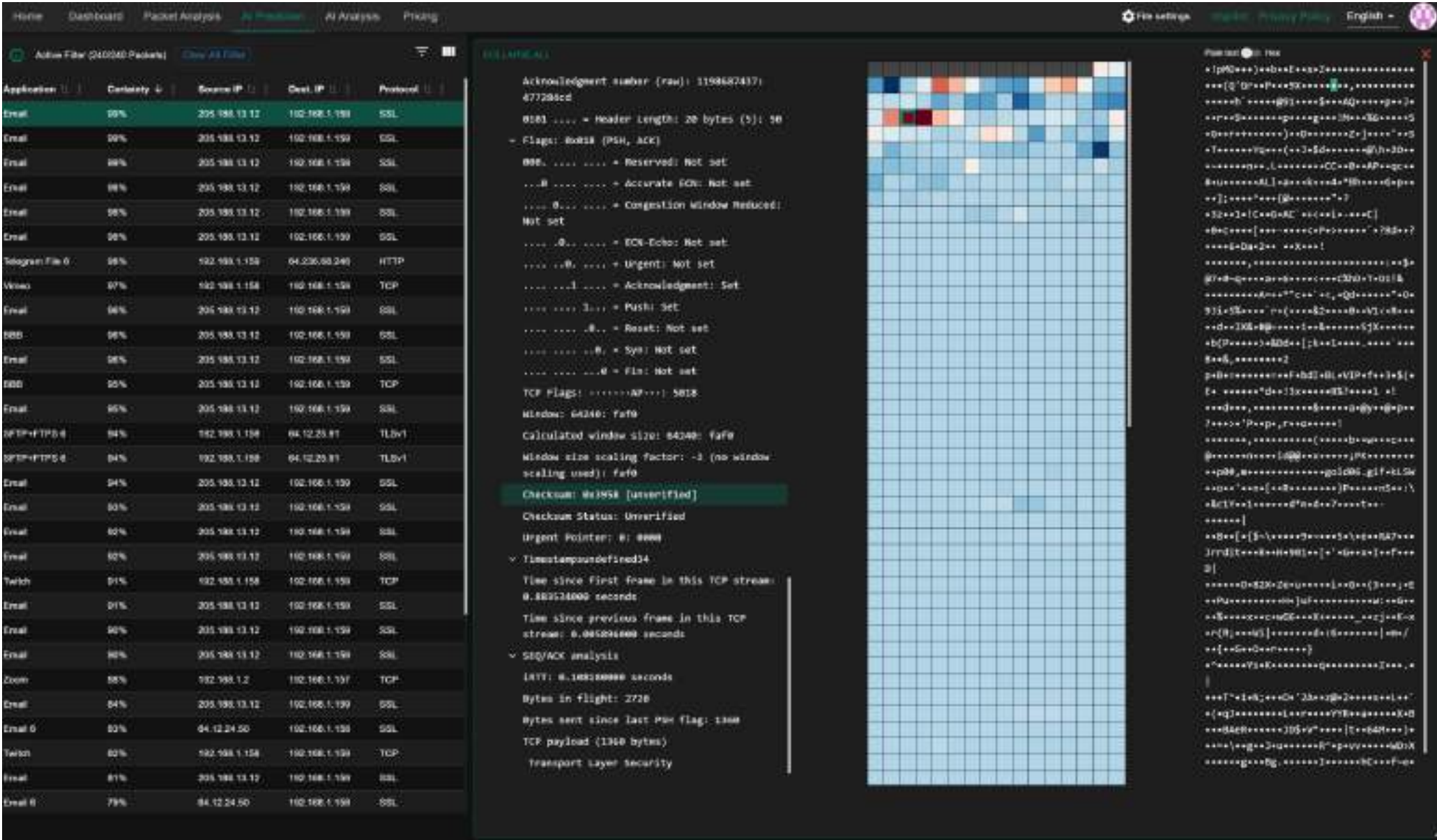
Application	Confidity	Source IP	Dest. IP	Protocol
Email	99%	205.188.11.12	102.160.1.158	SSL
Email	99%	205.188.11.12	102.160.1.158	SSL
Email	99%	205.188.11.12	102.160.1.158	SSL
Email	99%	205.188.11.12	102.160.1.158	SSL
Email	98%	205.188.11.12	102.160.1.158	SSL
Email	98%	205.188.11.12	102.160.1.158	SSL
Email	98%	205.188.11.12	102.160.1.158	SSL
Telegram File 0	98%	192.188.1.158	84.236.89.246	HTTP
Video	97%	192.188.1.158	102.160.1.158	TCP
Email	96%	205.188.11.12	102.160.1.158	SSL
BBB	96%	205.188.11.12	102.160.1.158	SSL
Email	96%	205.188.11.12	102.160.1.158	SSL
BBB	95%	205.188.11.12	102.160.1.158	TCP
Email	95%	205.188.11.12	102.160.1.158	SSL
SMTP+FTPS 8	94%	192.188.1.158	84.12.25.81	TLSv1
SMTP+FTPS 8	94%	192.188.1.158	84.12.25.81	TLSv1
Email	94%	205.188.11.12	102.160.1.158	SSL
Email	93%	205.188.11.12	102.160.1.158	SSL
Email	93%	205.188.11.12	102.160.1.158	SSL
Email	92%	205.188.11.12	102.160.1.158	SSL
Twitch	91%	192.188.1.158	102.160.1.158	TCP
Email	91%	205.188.11.12	102.160.1.158	SSL
Email	90%	205.188.11.12	102.160.1.158	SSL
Email	90%	205.188.11.12	102.160.1.158	SSL
Zoom	88%	192.188.1.2	102.160.1.157	TCP
Email	84%	205.188.11.12	102.160.1.158	SSL
Email 0	83%	84.12.24.50	102.160.1.158	SSL
Twitch	82%	192.188.1.158	102.160.1.158	TCP
Email	81%	205.188.11.12	102.160.1.158	SSL
Email 8	79%	84.12.24.50	102.160.1.158	SSL

last

now

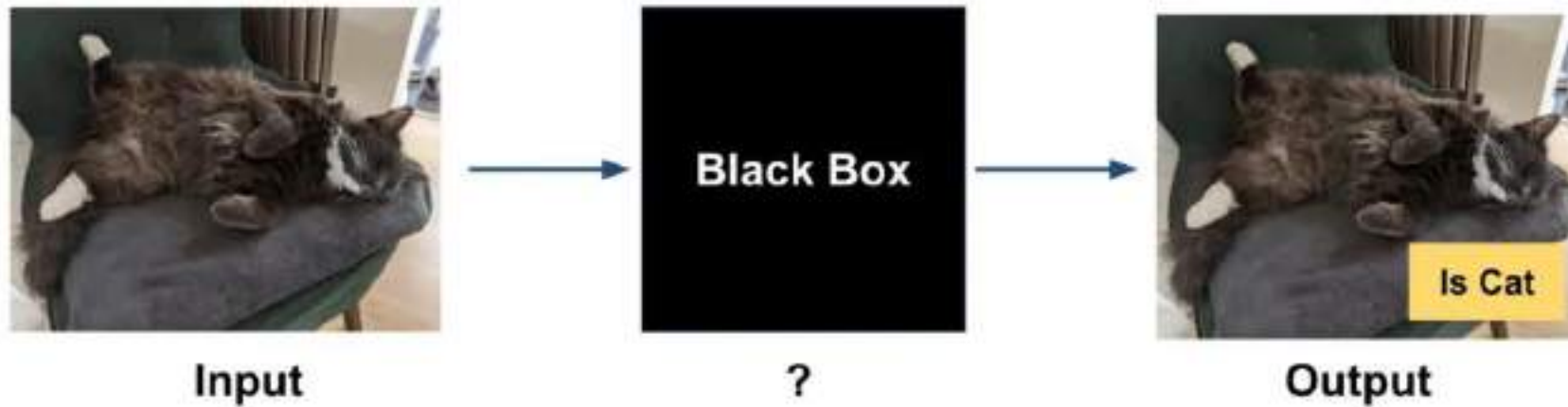
up next

NetCapVis AI Predictions

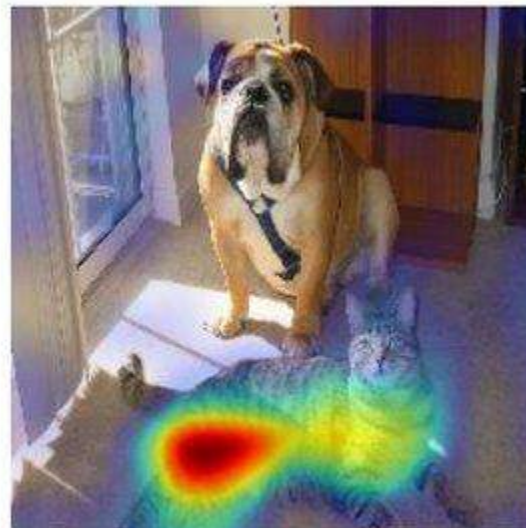


6. Explainable AI

What is XAI?



(a) Original Image



(b) Grad-CAM 'Cat'



(c) Grad-CAM 'Dog'

Trustworthy versus Explainable AI in Autonomous Vessels, Glomsrud et al. (2019)

Local Class Activation Maps (CAMs)

The image displays the Wireshark network protocol analyzer interface. The top menu bar includes Home, Dashboard, Packet Analysis, **File**, Analysis, and Pricing. The main window is divided into four panes:

- Packet List:** A table of captured packets. The first packet is an Email packet (100000) with a 100% capture rate, source IP 205.188.13.12, and destination IP 100.168.1.159. Other packets include Telegram File 0, Video, Email, and HTTP.
- Packet Details:** A hierarchical view of the selected packet's structure. It shows the Email packet structure, including the SMTP envelope, MIME parts, and the actual email content (HTML).
- Packet Bytes:** A hex dump of the packet data, showing the raw bytes of the email content.
- Packet Hex:** A hex dump of the packet data, showing the raw bytes of the email content.

Local Class Activation Maps (CAMs)

The screenshot displays the Wireshark network protocol analyzer interface. The top navigation bar includes 'Home', 'Dashboard', 'Packet Analysis', 'AI Prediction', 'AI Analysis', and 'Pricing'. The 'Packet Analysis' pane on the left shows a list of captured packets with columns for Application, Certainty, Source IP, Dest. IP, and Protocol. The main pane on the right is divided into three sections, each showing the details of a selected packet (Frame 77, 78, and 80) and its corresponding Local Class Activation Map (CAM).

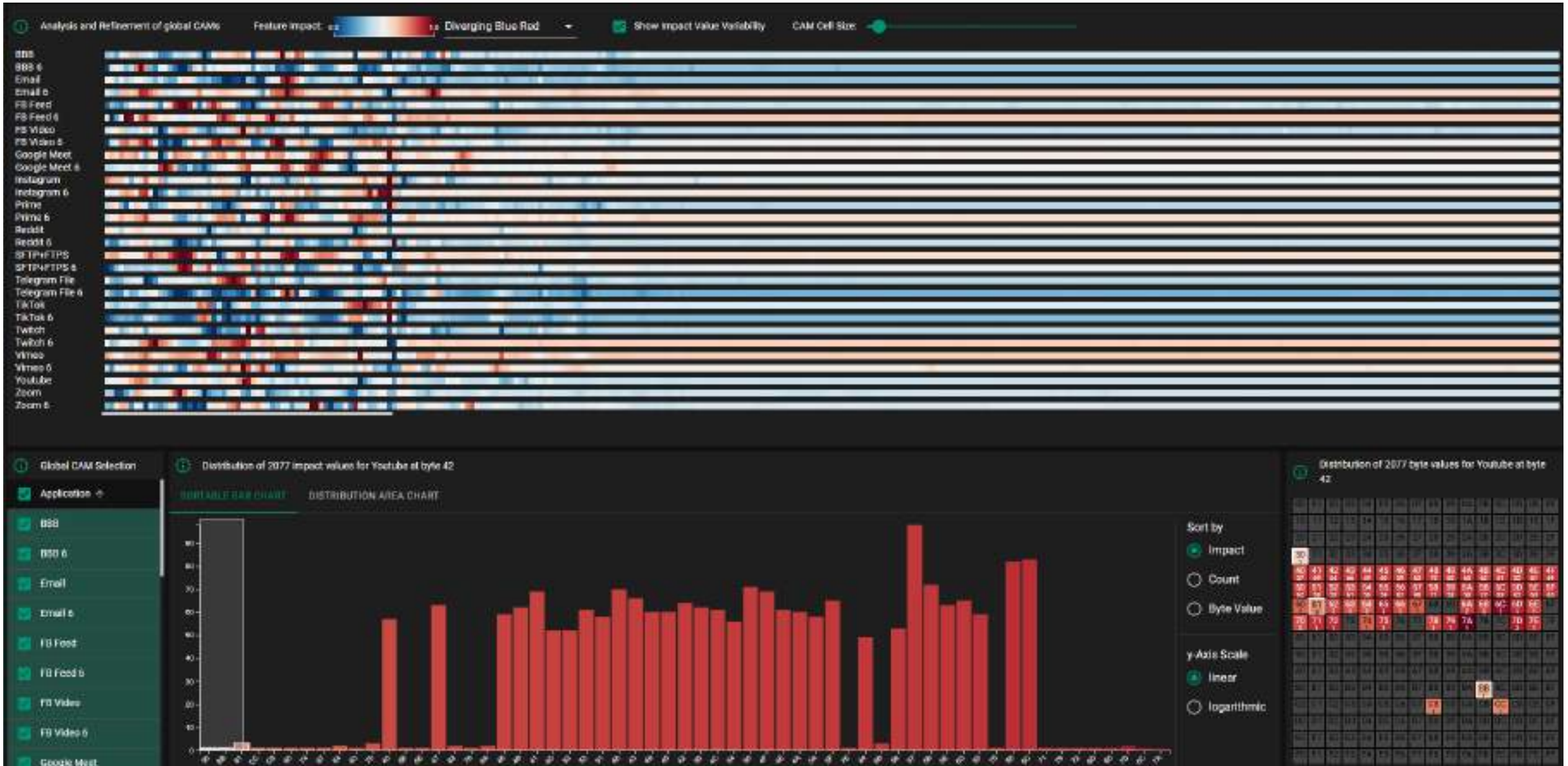
Packet 77: Frame 77: 3434 bytes on wire (11312 bits), 1434 bytes captured (11312 bits) on interface 0. Ethernet II, Src: VMware_b0:0d:8d:62, Dst: Dell_4d:4f:ae (08:00:27:00:0d:62). Internet Protocol Version 4, Src: 205.188.13.12, Dst: 192.168.1.159. Transmission Control Protocol, Src Port: 443, Dst Port: 1271, Seq: 19880, Ack: 126, Len: 1500.

Packet 78: Frame 78: 3434 bytes on wire (11312 bits), 1434 bytes captured (11312 bits) on interface 0. Ethernet II, Src: VMware_b0:0d:8d:62, Dst: Dell_4d:4f:ae (08:00:27:00:0d:62). Internet Protocol Version 4, Src: 205.188.13.12, Dst: 192.168.1.159. Transmission Control Protocol, Src Port: 443, Dst Port: 1271, Seq: 21240, Ack: 126, Len: 1500.

Packet 80: Frame 80: 3434 bytes on wire (11312 bits), 1434 bytes captured (11312 bits) on interface 0. Ethernet II, Src: VMware_b0:0d:8d:62, Dst: Dell_4d:4f:ae (08:00:27:00:0d:62). Internet Protocol Version 4, Src: 205.188.13.12, Dst: 192.168.1.159. Transmission Control Protocol, Src Port: 443, Dst Port: 1271, Seq: 22680, Ack: 126, Len: 1500.

Each packet detail is accompanied by a heatmap visualization (CAM) showing the activation of different classes across the packet's structure. The CAMs are color-coded, with red indicating high activation and blue indicating low activation. The rightmost pane shows the raw packet data in hexadecimal and ASCII format.

Global Class Activation Maps (CAMs)



Data Drift and Active Learning



last

now

up next

Insights

- Encrypted payload has almost no contribution to classification
- Training data preparation requires more expertise
- Sequence and ack number need to be cut
- Variable length of TCP options

Future Work

- Feedback system: network expert corrects the AI system
- Aggregate bytes to actual fields of the protocol
- Active Learning: split or combine application classes

Bringing visualization, machine learning and network experts together

Come talk to us!



Alex



Igor



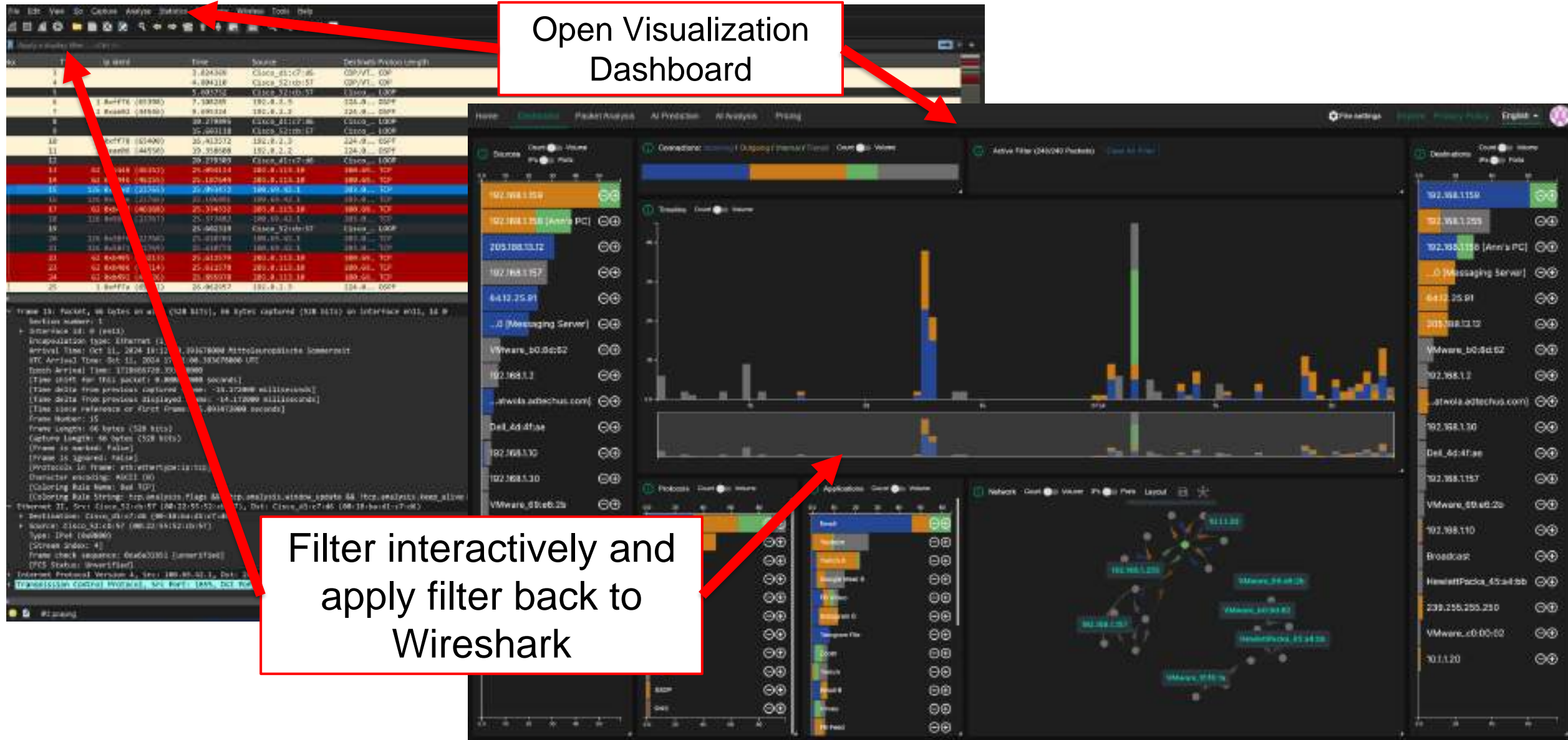
David



Felix

7. Visualization Plugin for Wireshark

Idea: Vizualization Dashboard

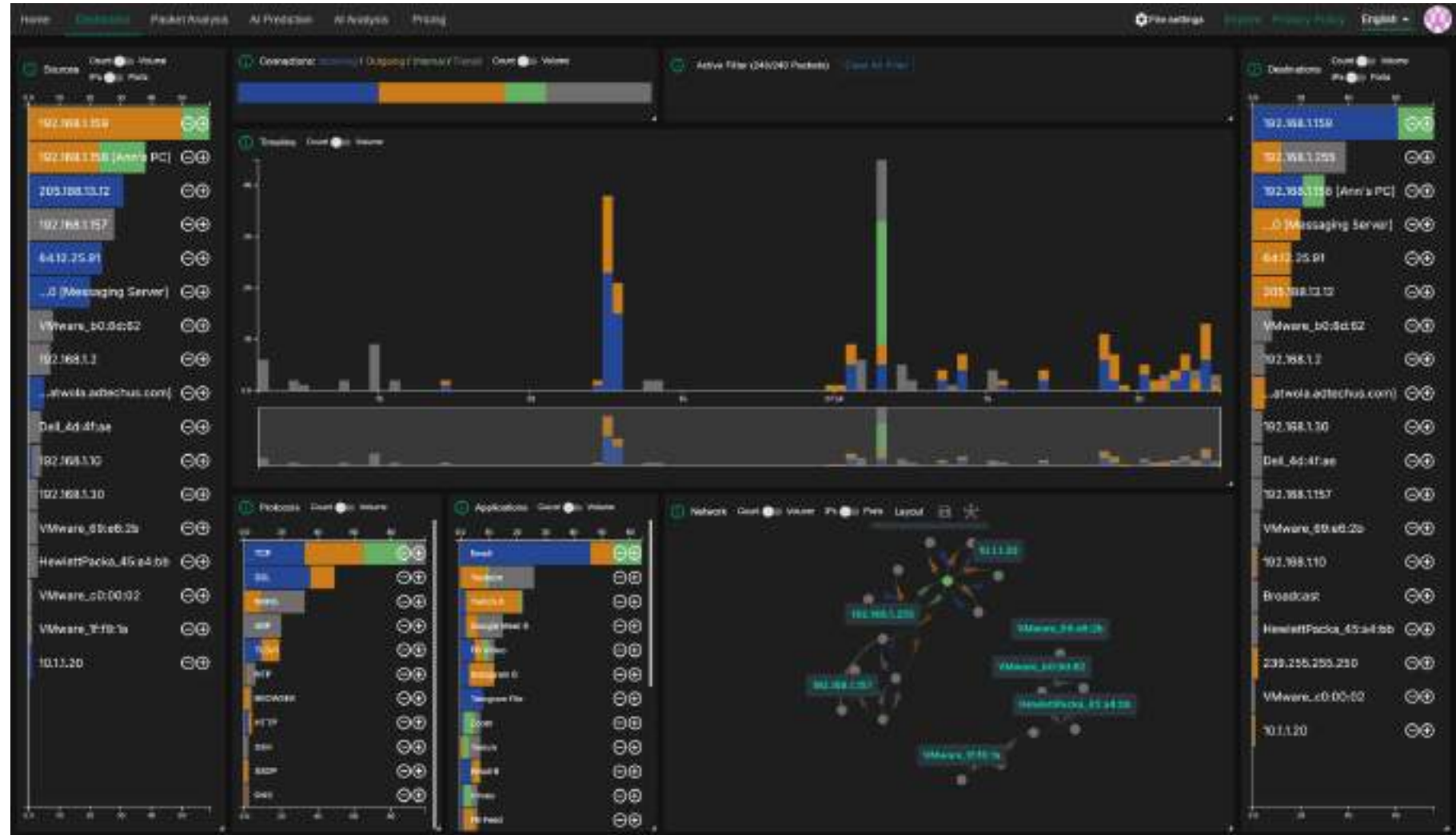


Idea: Vizualization Dashboard

Interactive
Dashboard only

No Packet Table
No AI Functions

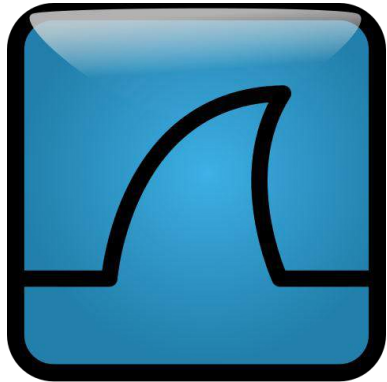
For free



Option A: Cloud-hosted Dashboard

Reduced Packet Header Data

```
[{"packetId":1,"srcMac":"00:50:56:c0:00:02","dstMac":"00:0c:29:69:e6:2b","size":66,"timestamp_unix":1250143821151,"protocol":"TCP","srcIp":"192.168.1.2","dstIp":"192.168.1.30","srcPort":55488,"dstPort":22,"udpStream":null,"tcpStream":0}, {"packetId":2,"srcMac":"00:0c:29:69:e6:2b","dstMac":"00:50:56:c0:00:02","size":114,"timestamp_unix":1250143821151,"protocol":"SMB","srcIp":"192.168.1.30","dstIp":"192.168.1.2","srcPort":22,"dstPort":55488,"udpStream":null,"tcpStream":0}, {"packetId":3,"srcMac":"00:50:56:c0:00:02","dstMac":"00:0c:29:69:e6:2b","size":66,"timestamp_unix":1250143821154,"protocol":"TCP","srcIp":"192.168.1.2","dstIp":"192.168.1.30","srcPort":55488,"dstPort":22,"udpStream":null,"tcpStream":0}, {"packetId":4,"srcMac":"00:0c:29:69:e6:2b","dstMac":"00:50:56:c0:00:02","size":170,"timestamp_unix":1250143821154,"protocol":"TCP","srcIp":"192.168.1.30","dstIp":"192.168.1.2","srcPort":22,"dstPort":55488,"udpStream":null,"tcpStream":0}, {"packetId":5,"srcMac":"00:0c:29:69:e6:2b","dstMac":"00:21:70:4d:4f:ae","size":60,"timestamp_unix":1250143824169,"protocol":"ARP","srcIp":"VMware_00:0d:60:62","dstIp":"Dell_A4:4f:ae","srcPort":null,"dstPort":null,"udpStream":null,"tcpStream":null}, {"packetId":6,"srcMac":"00:21:70:4d:4f:ae","dstMac":"00:0c:29:69:e6:2b","size":60,"timestamp_unix":1250143824169,"protocol":"ARP","srcIp":"Dell_A4:4f:ae","dstIp":"VMware_00:0d:60:62","srcPort":null,"dstPort":null,"udpStream":null,"tcpStream":null}, {"packetId":7,"srcMac":"00:0c:29:69:e6:2b","dstMac":"00:0c:29:69:e6:2b","size":90,"timestamp_unix":1250143826117,"protocol":"NTP","srcIp":"192.168.1.30","dstIp":"192.168.1.10","srcPort":123,"dstPort":123,"udpStream":0,"tcpStream":null}, {"packetId":8,"srcMac":"00:0c:29:69:e6:2b","dstMac":"00:0c:29:69:e6:2b","size":90,"timestamp_unix":1250143826117,"protocol":"NTP","srcIp":"192.168.1.10","dstIp":"192.168.1.30","srcPort":123,"dstPort":123,"udpStream":0,"tcpStream":null}]
```



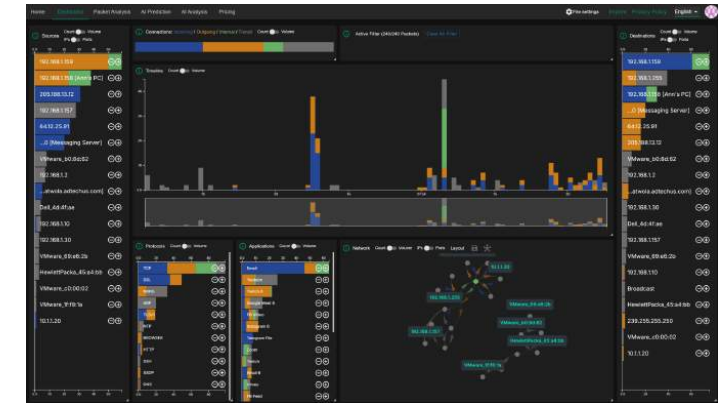
json



<https://...../dashboard>

filter query

```
(frame.time >= "Aug 13, 2009 07:57:35") && (frame.time <= "Aug 13, 2009 07:58:06") && (ip.dst == 192.168.1.159) && (tcp || udp)
```



Option B: Local Dashboard

Reduced Packet Header Data

```
[{"packetId":1,"srcMac":"00:50:56:c0:00:02","dstMac":"00:0c:29:69:e6:2b","size":66,"timestamp_unix":1250143821151,"protocol":"TCP","srcIp":"192.168.1.2","dstIp":"192.168.1.30","srcPort":55488,"dstPort":22,"udpStream":null,"tcpStream":0}, {"packetId":2,"srcMac":"00:0c:29:69:e6:2b","dstMac":"00:50:56:c0:00:02","size":114,"timestamp_unix":1250143821151,"protocol":"SYN","srcIp":"192.168.1.30","dstIp":"192.168.1.2","srcPort":22,"dstPort":55488,"udpStream":null,"tcpStream":0}, {"packetId":3,"srcMac":"00:50:56:c0:00:02","dstMac":"00:0c:29:69:e6:2b","size":66,"timestamp_unix":1250143821154,"protocol":"TCP","srcIp":"192.168.1.2","dstIp":"192.168.1.30","srcPort":55488,"dstPort":22,"udpStream":null,"tcpStream":0}, {"packetId":4,"srcMac":"00:0c:29:69:e6:2b","dstMac":"00:50:56:c0:00:02","size":170,"timestamp_unix":1250143821154,"protocol":"TCP","srcIp":"192.168.1.30","dstIp":"192.168.1.2","srcPort":22,"dstPort":55488,"udpStream":null,"tcpStream":0}, {"packetId":5,"srcMac":"00:0c:29:69:e6:2b","dstMac":"00:21:70:4d:4f:ae","size":60,"timestamp_unix":1250143824169,"protocol":"ARP","srcIp":"VMware_00:0d:60:62","dstIp":"Dell_A4:4f:ae","srcPort":null,"dstPort":null,"udpStream":null,"tcpStream":null}, {"packetId":6,"srcMac":"00:21:70:4d:4f:ae","dstMac":"00:0c:29:69:e6:2b","size":60,"timestamp_unix":1250143824169,"protocol":"ARP","srcIp":"Dell_A4:4f:ae","dstIp":"VMware_00:0d:60:62","srcPort":null,"dstPort":null,"udpStream":null,"tcpStream":null}, {"packetId":7,"srcMac":"00:0c:29:69:e6:2b","dstMac":"00:0c:29:69:e6:2b","size":90,"timestamp_unix":1250143826117,"protocol":"NTP","srcIp":"192.168.1.30","dstIp":"192.168.1.10","srcPort":123,"dstPort":123,"udpStream":0,"tcpStream":null}, {"packetId":8,"srcMac":"00:0c:29:69:e6:2b","dstMac":"00:0c:29:69:e6:2b","size":90,"timestamp_unix":1250143826117,"protocol":"NTP","srcIp":"192.168.1.10","dstIp":"192.168.1.30","srcPort":123,"dstPort":123,"udpStream":0,"tcpStream":null}]
```

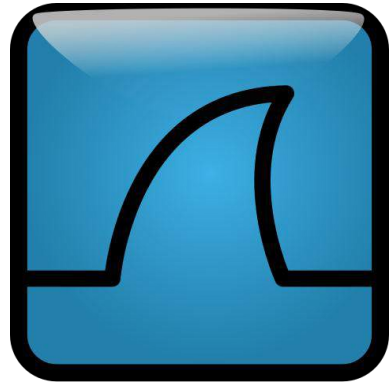
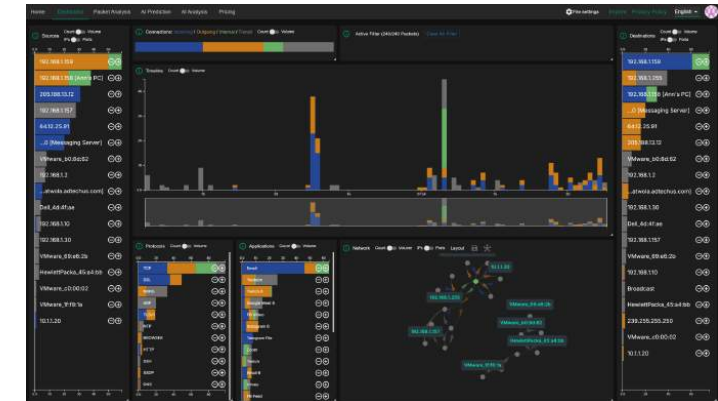
json



localhost:3000/dashboard

filter query

```
(frame.time >= "Aug 13, 2009 07:57:35") && (frame.time <= "Aug 13, 2009 07:58:06") && (ip.dst == 192.168.1.159) && (tcp || udp)
```



Option A: Cloud-hosted Dashboard

- + easier integration
- + no extra install
- internet connection required
- data upload

Option B: Local Dashboard

- + data stays on the device
- + no service dependency
- more install requirements
- start own localhost service

Would you use this plugin?

Which option would you prefer?

Are there more options?

Contact and Feedback

Dr. Alex Ulmer

User Centered Data Science, Visual Analytics
Fraunhofer IGD
Darmstadt, Germany

alex.ulmer@igd.fraunhofer.de

LinkedIn:



Thank
you!

SharkFest Feedback Form



#sf25eu