

Understanding the Context of Network Traffic Alerts

Bram C.M Cappers, Jarke J. van Wijk

Presented by Stewart Grant

Motivation

Networks are constantly under attack from malicious users.

Attacks are hidden inside a massive amount of harmless traffic.

Experts require tools to detect and correlate malicious messages in a sea of traffic.

Data: Wireshark log

- Fine grained
- Dense
- Attacks are hidden

The screenshot shows the Wireshark interface with a live capture on the eth0 interface. The packet list pane displays the following traffic:

No.	Time	Source	Destination	Protocol	Info
46	139.931187	wistron_07:07:ee	broadcast	ARP	who has 192.168.1.254? tell 192.168.1.68
47	139.931463	ThomsonT_08:35:4f	wistron_07:07:ee	ARP	192.168.1.254 is at 00:90:d0:08:35:4f
48	139.931466	192.168.1.68	192.168.1.254	DNS	Standard query A www.google.com
49	139.975406	192.168.1.254	192.168.1.68	DNS	Standard query response CNAME www.l.google.com A 66.102.9.99
50	139.976811	192.168.1.68	66.102.9.99	TCP	62216 > http [SYN] Seq=0 Win=8192 Len=0 MSS=1460 WS=2
51	140.079578	66.102.9.99	192.168.1.68	TCP	http > 62216 [SYN, ACK] Seq=0 Ack=1 Win=5720 Len=0 MSS=1430
52	140.079583	192.168.1.68	66.102.9.99	TCP	62216 > http [ACK] Seq=1 Ack=1 Win=65780 Len=0
53	140.080278	192.168.1.68	66.102.9.99	HTTP	GET /complete/search?hl=en&client=suggest&js=true&q=m&cp=1 H
54	140.086765	192.168.1.68	66.102.9.99	TCP	62216 > http [FIN, ACK] Seq=805 Ack=1 Win=65780 Len=0
55	140.086921	192.168.1.68	66.102.9.99	TCP	62218 > http [SYN] Seq=0 Win=8192 Len=0 MSS=1460 WS=2
56	140.197484	66.102.9.99	192.168.1.68	TCP	http > 62216 [ACK] Seq=1 Ack=805 Win=7360 Len=0
57	140.197777	66.102.9.99	192.168.1.68	TCP	http > 62216 [FIN, ACK] Seq=1 Ack=806 Win=7360 Len=0
58	140.197811	192.168.1.68	66.102.9.99	TCP	62216 > http [ACK] Seq=806 Ack=2 Win=65780 Len=0
59	140.212310	66.102.9.99	192.168.1.68	TCP	http > 62218 [SYN, ACK] Seq=0 Ack=1 Win=5720 Len=0 MSS=1430

The packet details pane for the selected packet (No. 46) shows:

- Frame 1 (42 bytes on wire, 42 bytes captured)
- Ethernet II, Src: Vmware_38:eb:0e (00:0c:29:38:eb:0e), Dst: Broadcast (ff:ff:ff:ff:ff:ff)
- Address Resolution Protocol (request)

The packet bytes pane shows the raw hex and ASCII data:

```
0000 ff ff ff ff ff ff 00 0c 29 38 eb 0e 08 06 00 01 ..... )8.....
0010 08 00 06 04 00 01 00 0c 29 38 eb 0e c0 a8 39 80 ..... )8....9.
0020 00 00 00 00 00 00 c0 a8 39 02 ..... 9.
```

At the bottom of the window, it indicates: eth0: <live capture in progress> Fil... Packets: 445 Displayed: 445 Marked: 0 Profile: Default

Data attributes

1. Timestamp
2. Ip address
3. Mac address
4. Protocol
5. Protocol flags
6. Message size
7. Many more metadata attributes

Anatomy of an attack

- Localized to a time interval
- Composed of many messages
- Can span multiple machines
- Exhibit uncommon behaviour

Example: Man in the middle

Malicious user intercepts, potentially modifies, and relays messages.

ML tools

Tools such as snort and Bro use ML to detect attacks.

1. Train on sample traffic
2. Monitor streaming traffic
3. Output outliers, and known attack patterns

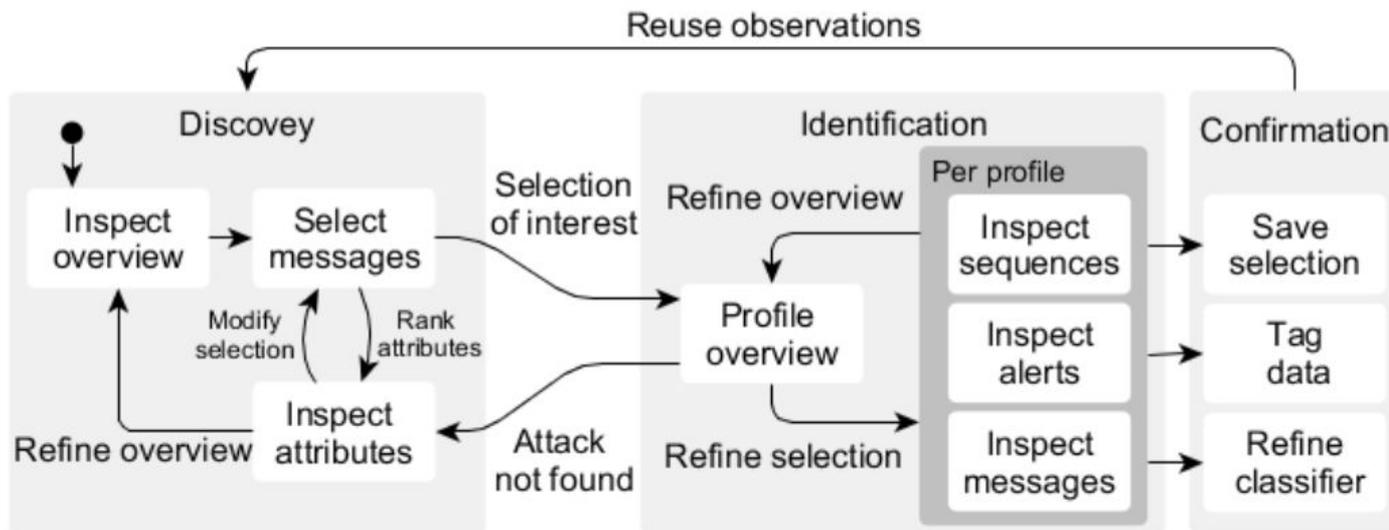
Downsides

1. Lots of alerts
2. Large number of false positives
3. Difficult to query
4. No intuitive feedback



CoNTA - Contextual Analysis of Network Traffic Alerts

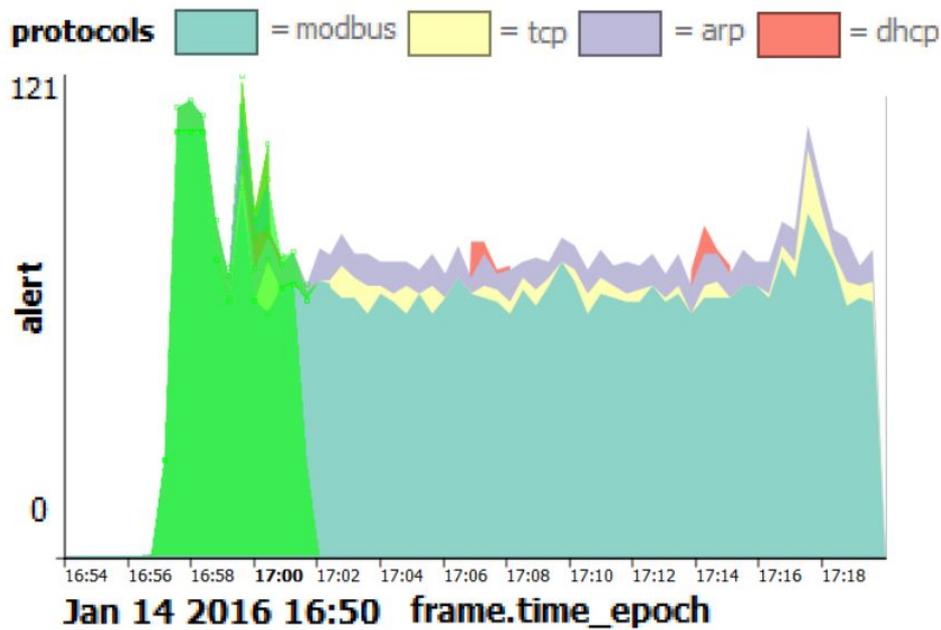
Interactive discovery and refinement of alerts.



Discovery (Time table)

Experts identify anomalies by examining high level trends such a bursts.

- IDS alerts are plotted as a stacked line chart
- X axis Time
- Y axis Message attribute

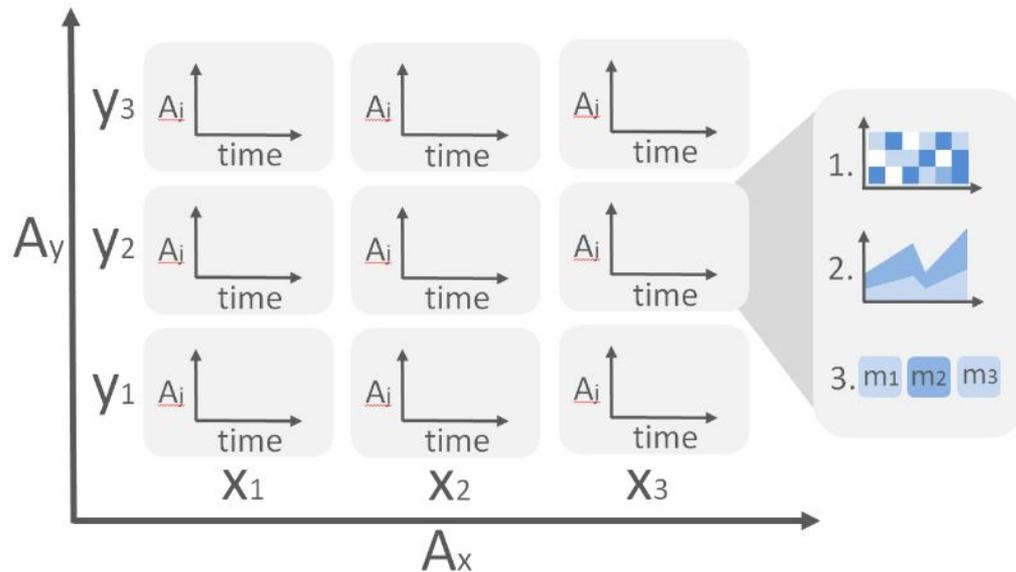


Time Table grid view View

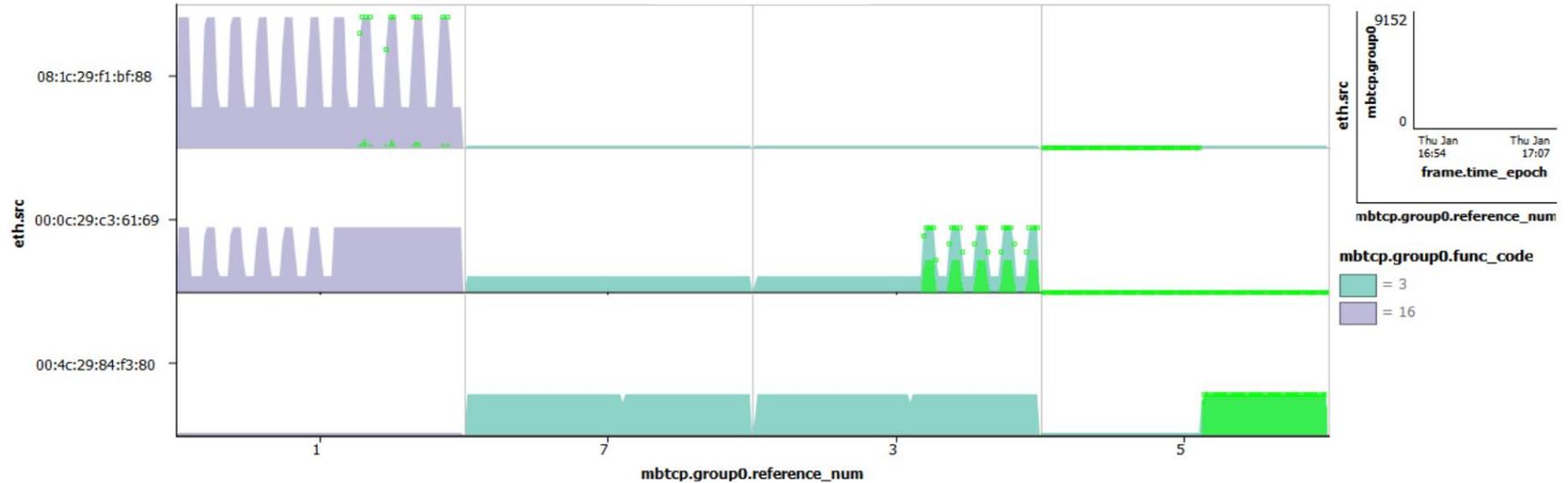
Y axis: Machines

X axis: Attributes

Display allows for the visual correlation of alerts which compose an attack.



Time Table example 3 separate mac addresses



Discovery: Conversation View

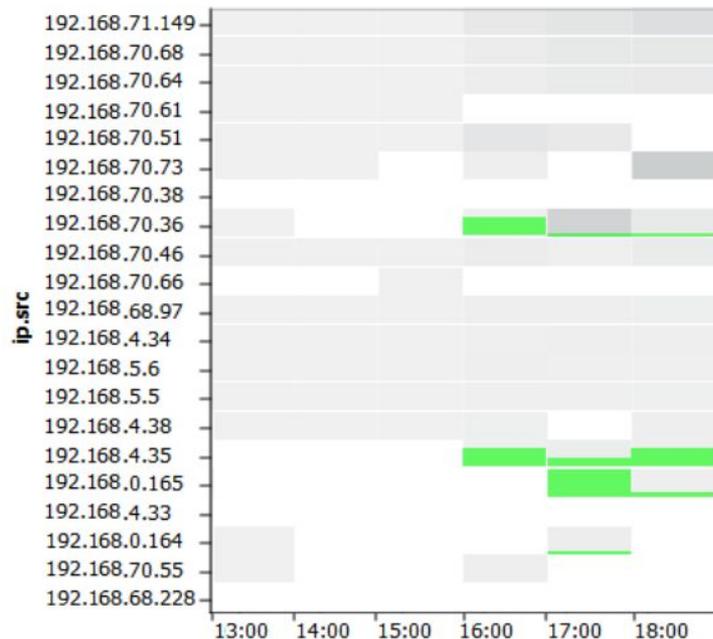
- Communication topology is useful for identifying malicious and colluding machines.
- Thickness of edges corresponds to bandwidth



Discovery: Heat map

- General aggregate information displayed as heatmap
- Useful for detecting high level trends

*Alerts per IP over time

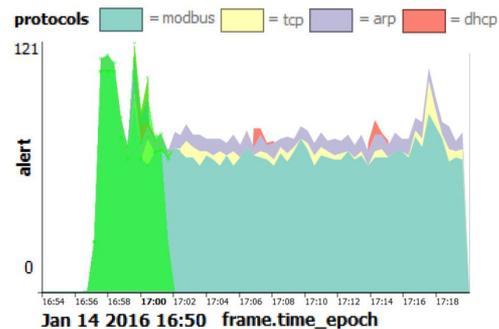
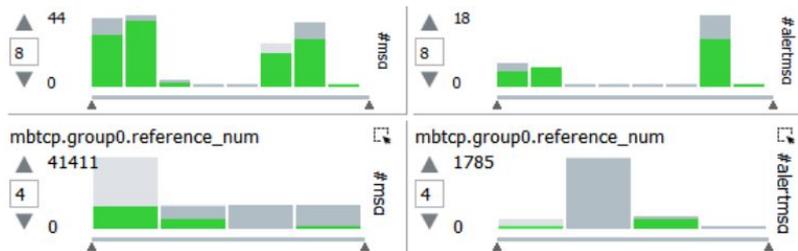


Identification: Selection

Users select areas of interest

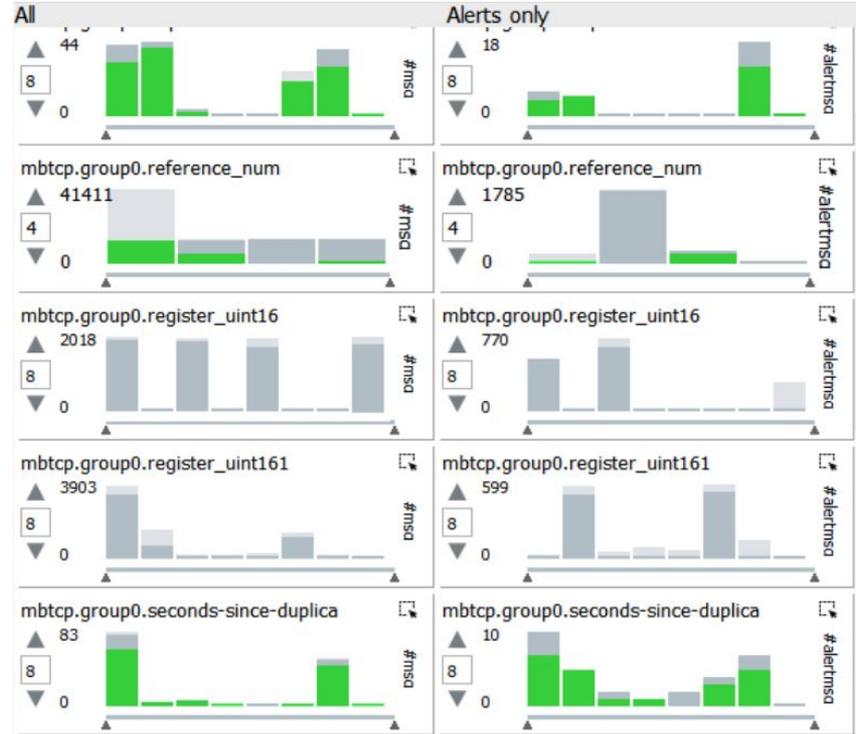
Selected data is highlighted in green

All visuals are updated with selected data



Identification (Attribute Histograms)

- Traffic is bucketed into histograms based on attributes.
- Left column shows the distribution of all traffic
- Right column shows the distribution of alert traffic only



Interactive Identification

A data packets can have hundreds of attribute values, experts need to filter and search for information relevant to an attack.

- Manually set #of attribute buckets
- Sort histogram
 - Alphabetical
 - Most alerts
 - Relevance

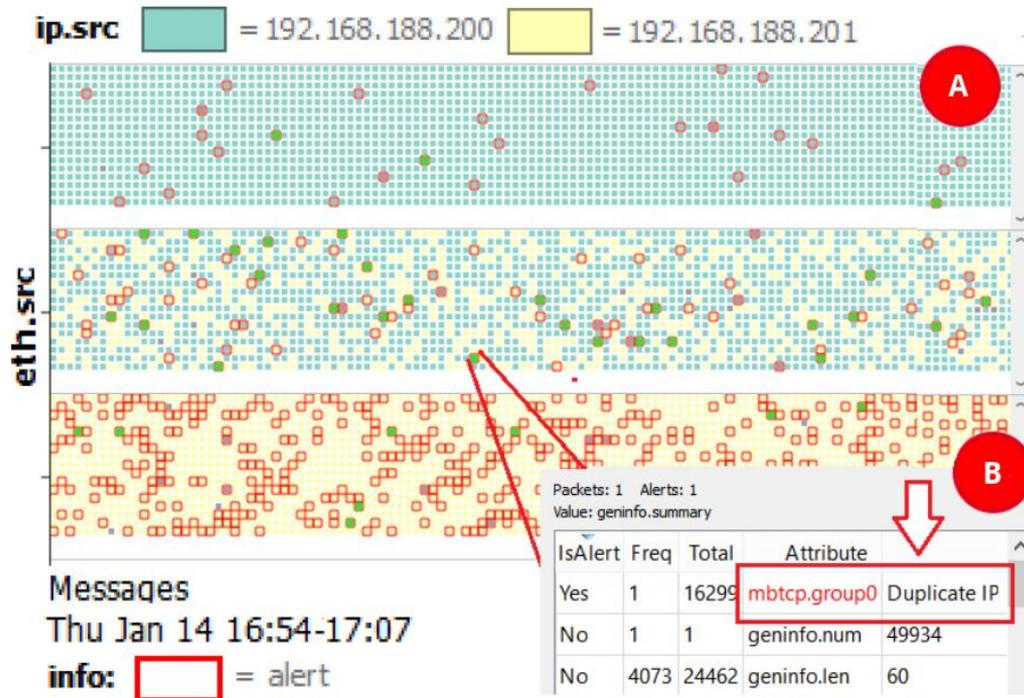
Interactive Classification

IDS false positives are high, CoNTA uses 4 interactive training strategies:

1. Filtering
2. Projection
3. Binning
4. Self training

Confirmation (inspecting messages)

- Pixel map encodes individual messages
- Attributes are color coded
- Red outlines encode alert messages
- Upon selection raw message data is displayed



Context Preservation

Status	Context	#Packets	#Alerts	Coverage	Alerts
<input checked="" type="radio"/>	all	84209	2475		
<input type="radio"/>	alert burst	22292	1171		
<input checked="" type="radio"/>	suspicious connection modbus	55215	2101		

- Selecting and zooming can cause a loss of context
- Iterative selections form a hierarchy
- The hierarchy is displayed as a tree
- Histogram display the coverage of the selection

Evaluation (use cases)

Water Plant (synthetic)

- Detected man in the middle attack
- Used [Line chart, histograms, pixel map]

University (real logs)

- Detected a user remotely installing software
- Used [heatmap, manual filtering, conversation topology]

Time for a demo video ??

[link](#)

What Why How

What

- messages and alerts from IDS
- Communication graph

Why

- Attack detection
- Correlation between malicious messages

How

- High level heat maps, and line charts to detect patterns
- Mouse selection, and queries to facet data
- Pixel map for inspecting individual packets

Limitations

- Histograms don't scale well if the number of attributes are high
- Number of visible attributes is constrained by the number of histograms
- Requires an IDS that supports interactive learning

Critique

- Information about malicious packets are spread across many histograms perhaps clustering (tSNE) alerts on their attributes would help detect trends.
- Selections must be contiguous, making correlations between multiple features difficult.
- All visualizations are built for a 5 - 20 machine network, so they do not apply to data centers which desperately need them.
- Pixel packet view is not temporally aligned which could cause confusion.
- Color contrast (green on gray) has great pop out and is effective at maintaining context between views.

Conclusion

- CoNTA provides an interactive attack detection framework
- Helps experts translate high level phenomenon to packet attributes
- Has a very nice selection interface

Questions ???