# Understanding the Context of Network Traffic Alerts

Bram C.M Cappers, Jarke J. van Wijk

Presented by Stewart Grant

# Motivation

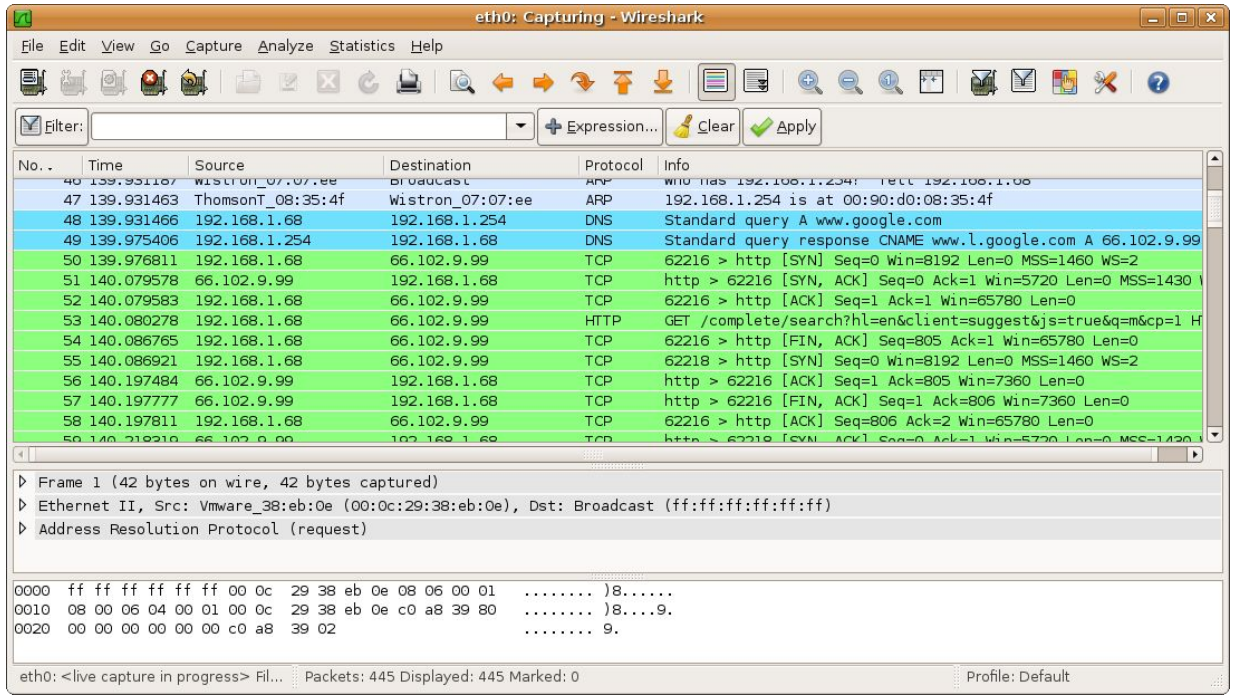
Networks are constantly under attack from malicious users.

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



# 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.

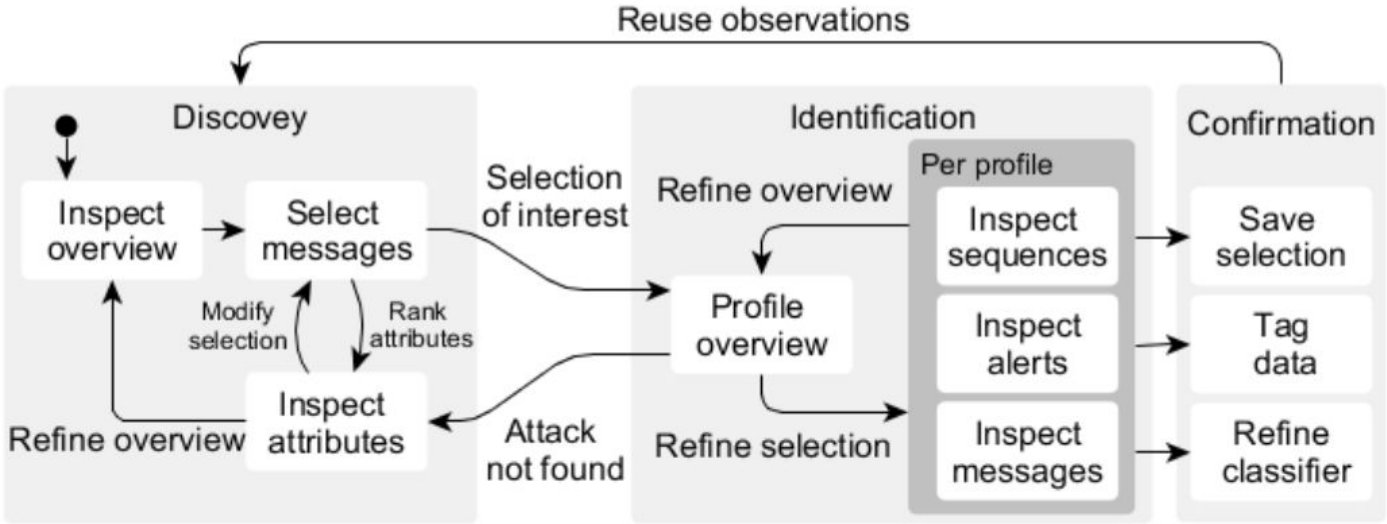
1. Train on sample traffic
2. Monitor streaming traffic
3. Output outliers, and known attack patterns


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

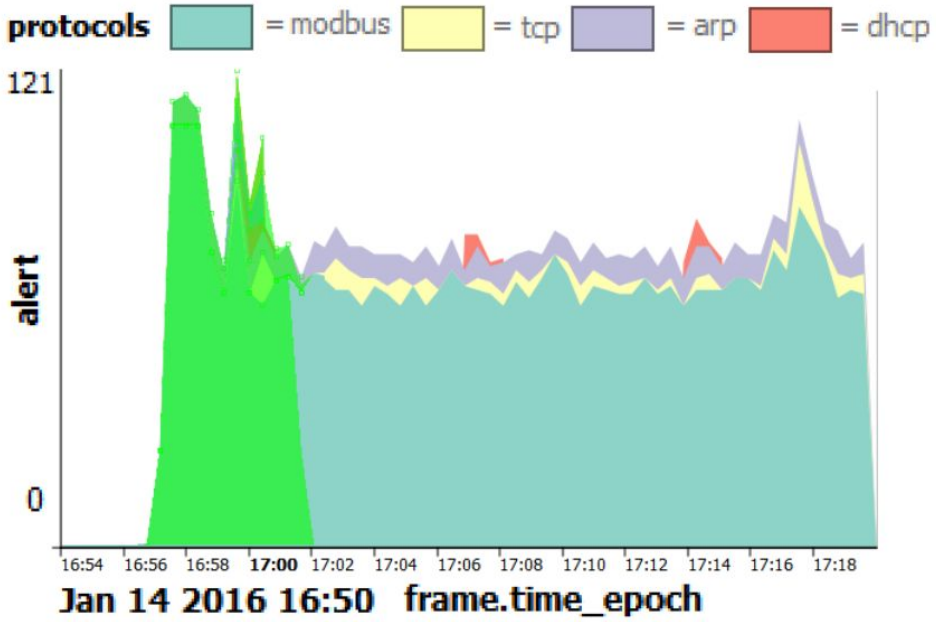
Interactive discovery and refinement of alerts.



# 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

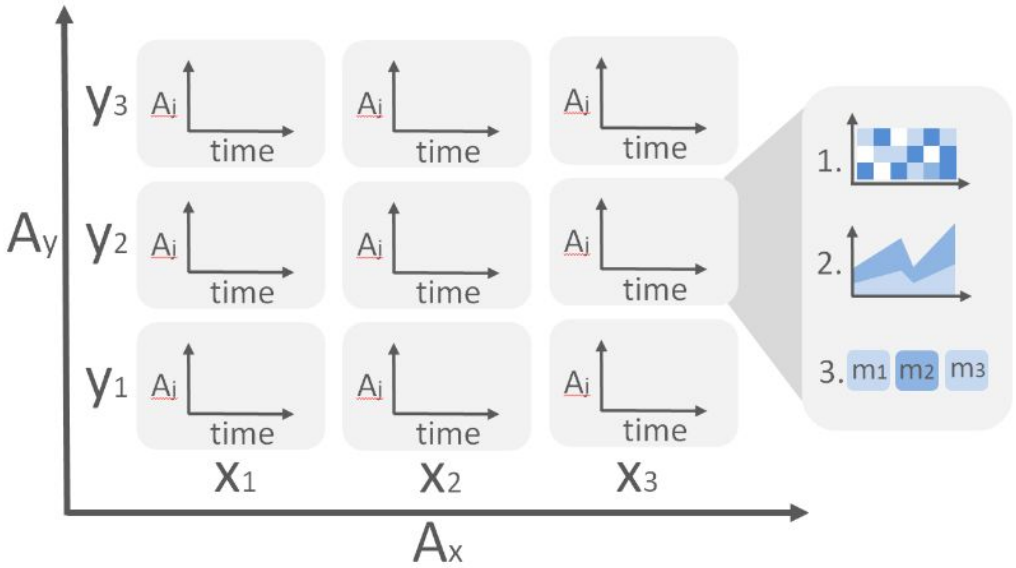


# Time Table grid view View

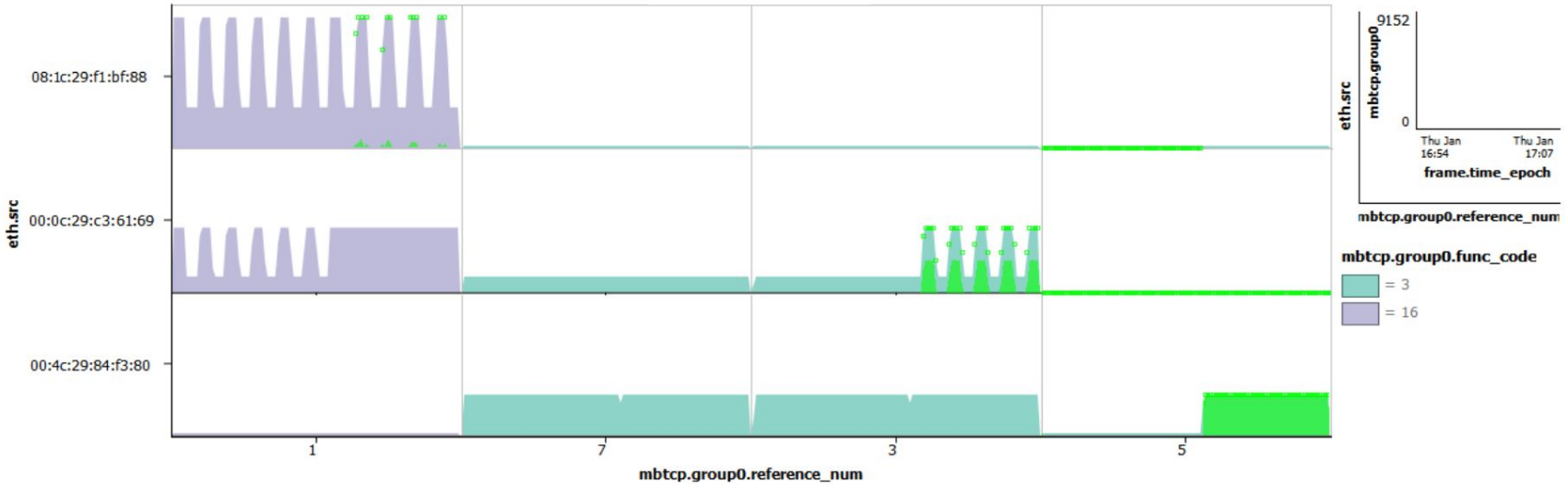
Y axis: Machines

X axis: Attributes

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



# Time Table example 3 separate mac addresses



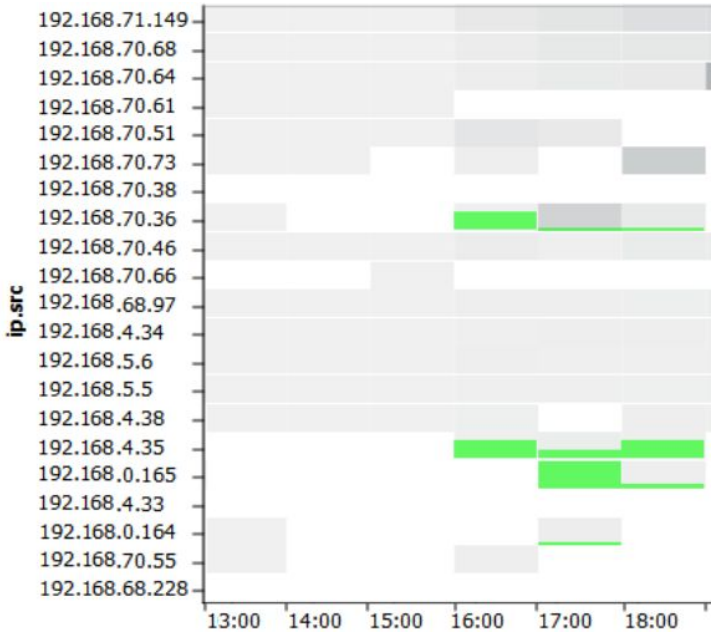
# Discovery: Conversation View

- 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

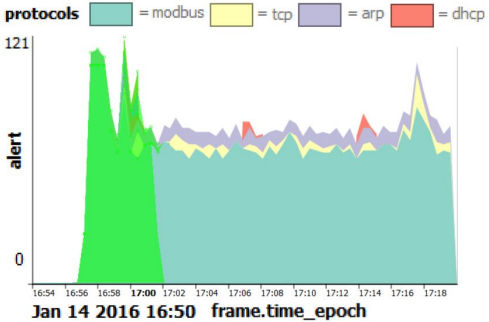


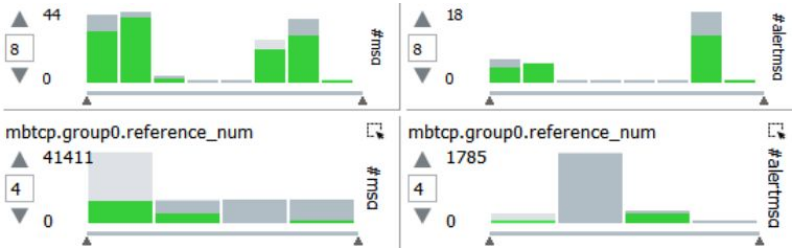
# Identification: Selection

Users select areas of interest

Selected data is highlighted in green

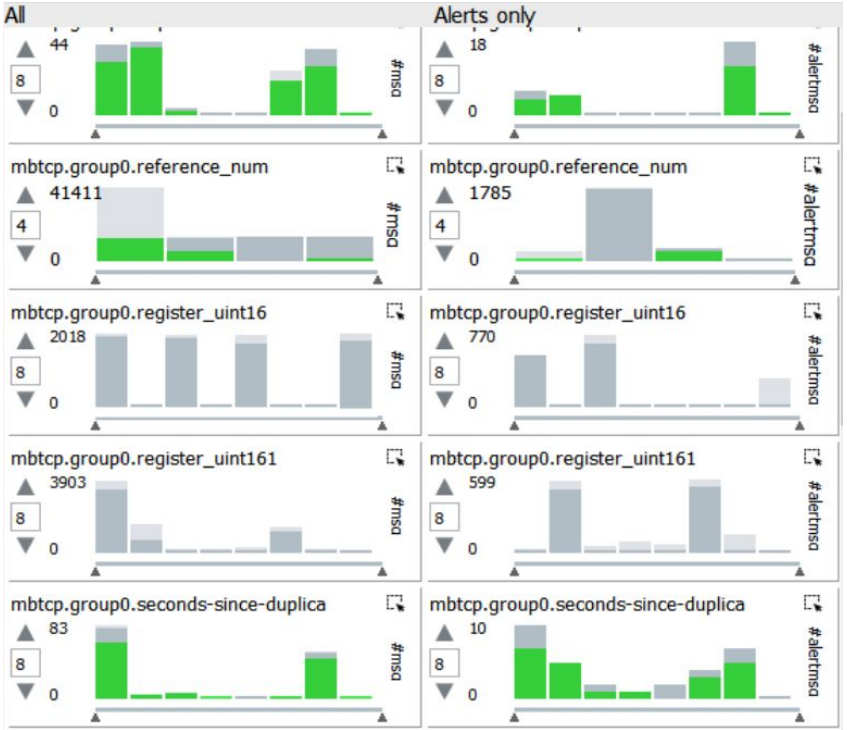
All visuals are updated with selected data




# 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

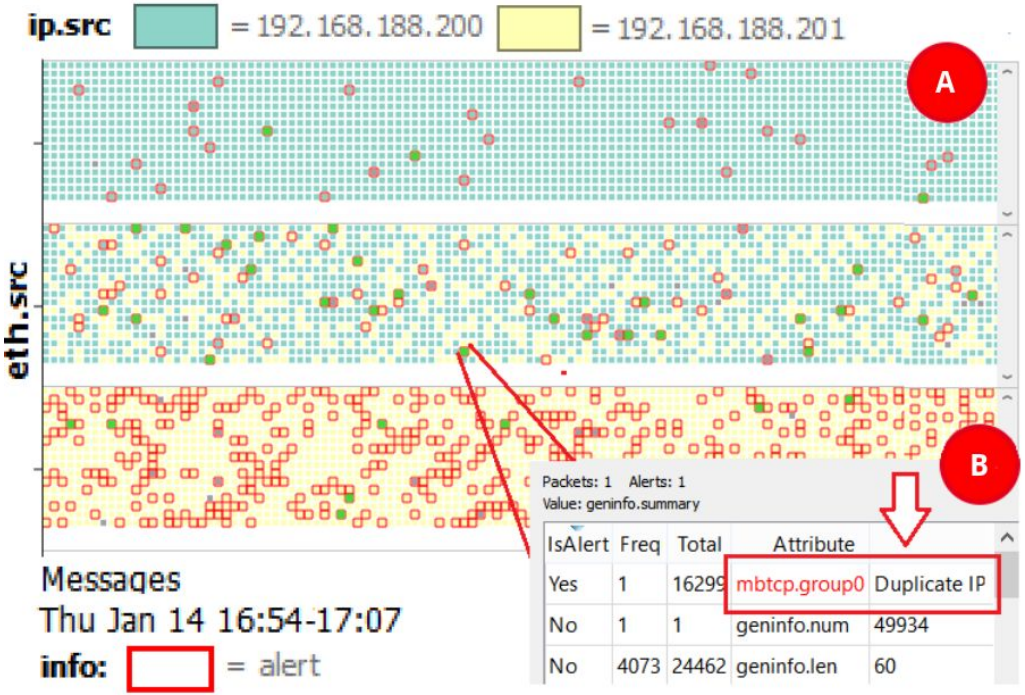
# Interactive Classification

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

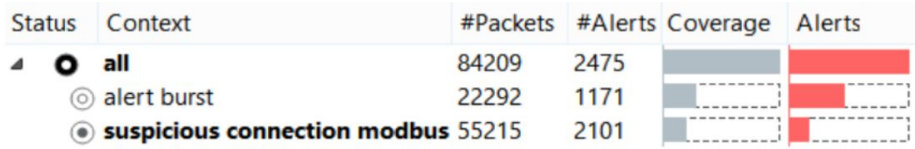
1. Filtering
2. Projection
3. Binning
4. Self training

# 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 |
|---|---|---|---|---|---|
| ◢ | **all** | 84209 | 2475 | | |
| | alert burst | 22292 | 1171 | | |
| | **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 ???