

Anomaly Extraction in Backbone Networks using Association Rules

Daniela Brauckhoff
Xenofontas (Fontas) Dimitropoulos
Arno Wagner
Kave Salamatian



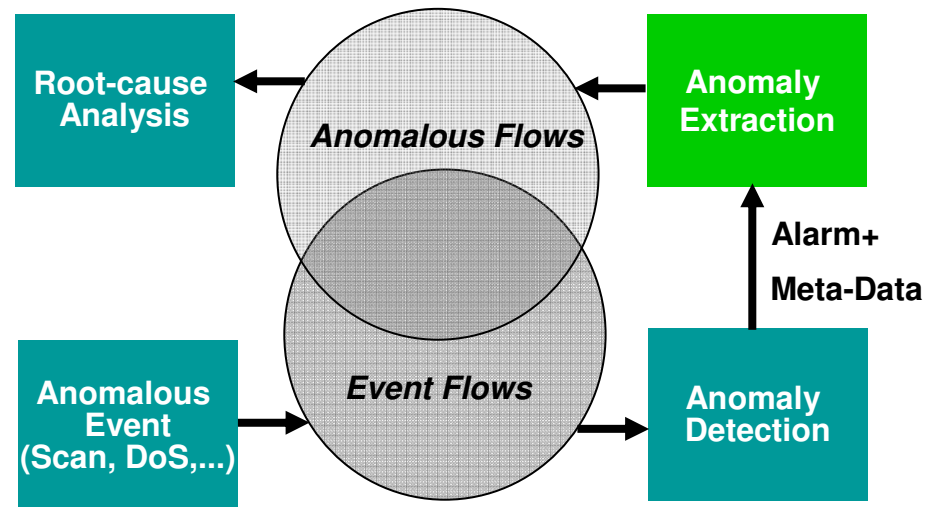
Anomaly extraction

■ Problem statement:

- Given a large set of flows observed during a time interval labeled with an anomaly alert find and summarize the flows involved in the event(s) that triggered the alert.

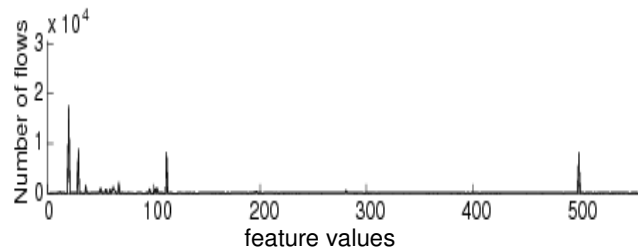
■ Motivation:

- Root cause analysis
- Attack mitigation
- Anomaly modeling



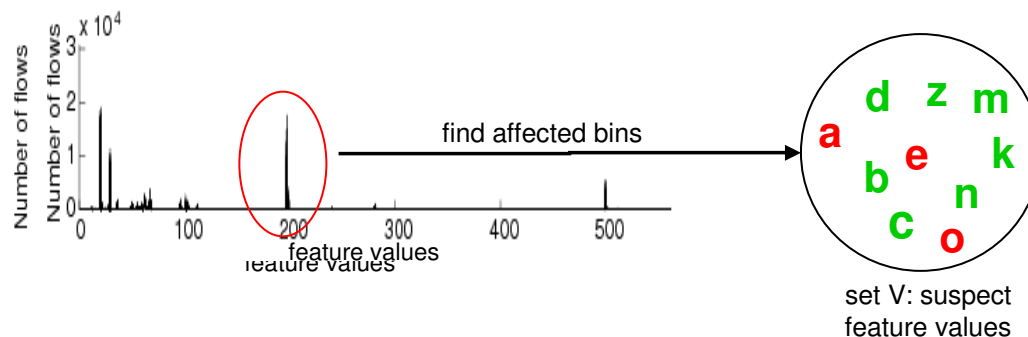
Approach Overview (3 steps)

- **Detection:** Use a number of histogram-based detectors:
 - Identify affected bins and create set V of corresponding feature values
 - Use histogram cloning to reduce collisions and false positives



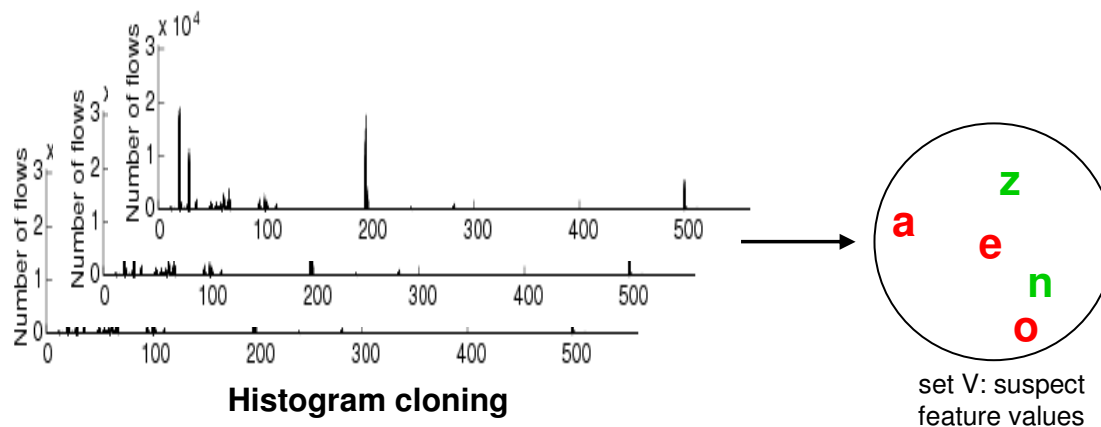
Approach Overview (3 steps)

- **Detection:** Use a number of histogram-based detectors:
 - Identify affected bins and create set V of corresponding feature values
 - Use histogram cloning to reduce collisions and false positives



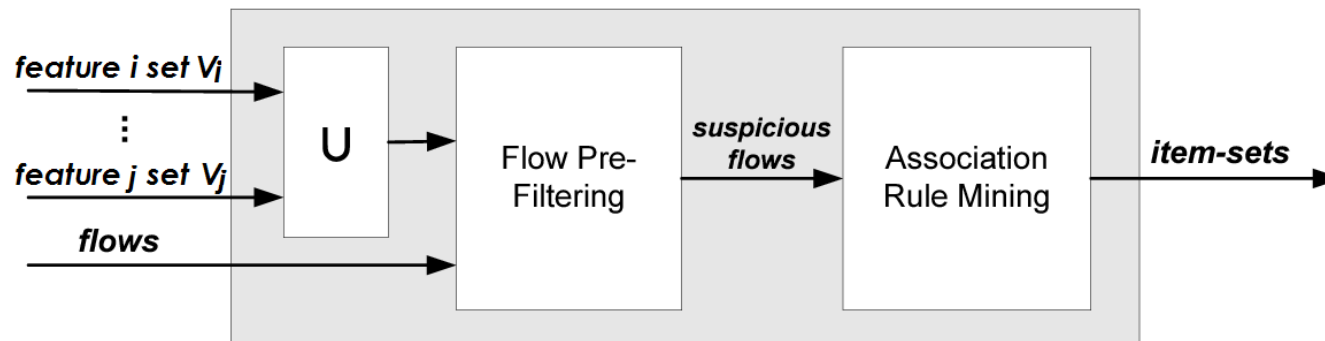
Approach Overview (3 steps)

- **Detection:** Use a number of histogram-based detectors:
 - Identify affected bins and create set V of corresponding feature values
 - Use histogram cloning to reduce collisions and false positives



Approach Overview (3 steps)

- **Detection:** Use a number of **histogram-based detectors**:
 - Identify affected bins and create set V of corresponding feature values
 - Use histogram cloning to reduce collisions and false positives
- **Filtering:** Filter flows that match **union** of meta-data provided by N detectors
 - Filtered flows are called „suspicious“ flows
- **Mining:** Use **association rules** to extract and summarize anomalous flows from the set of suspicious flows



Association Rule Mining

- Given a number of *itemsets*, find frequent subsets which are common to at least a minimum number *s* of the itemsets.
- An itemset is a flow (7-tuple): {srcIP, dstIP, srcPort, dstPort, proto, #packets, #bytes}
- Key intuition:** anomalies trigger a large number of flows with one or more common feature values, e.g., src IP addr, dst port, #packets.
- Use modified Apriori algorithm to find frequent subsets
- Example output:

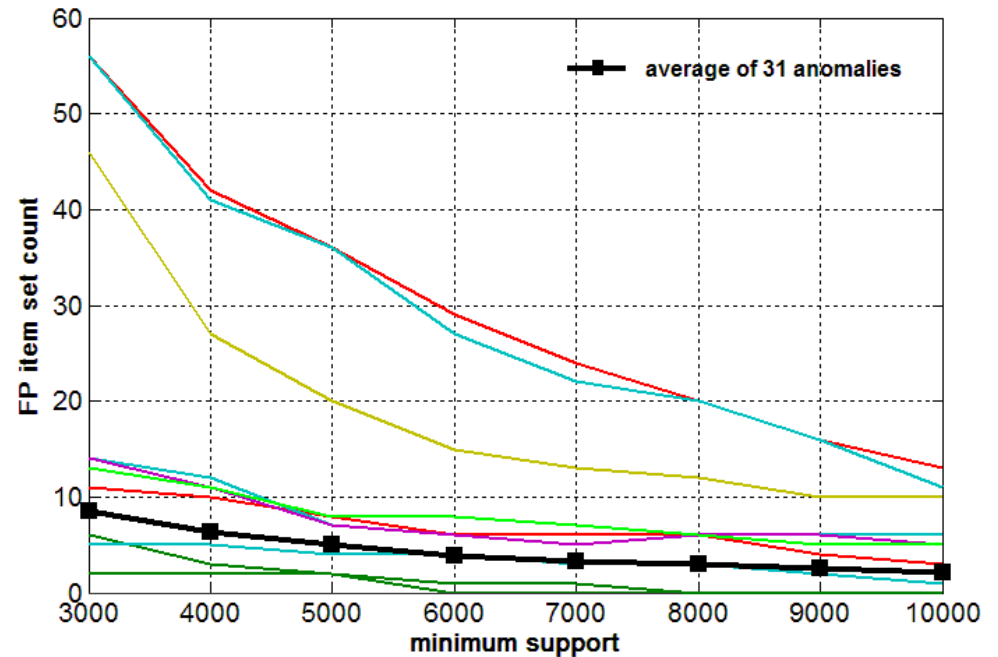
Artificially added port 80 and 25 traffic to illustrate false positives

Anomalous flows

| <i>l</i> | srcIP | dstIP | srcPort | dstPort | #packets | #bytes | support | what |
|----------|--------|--------|---------|---------|----------|--------|---------|----------------|
| 1 | * | * | * | * | 2 | * | 10,407 | |
| 1 | * | * | * | 25 | * | * | 22,659 | |
| 2 | Host A | * | * | 80 | * | * | 11,800 | HTTP Proxy |
| 2 | * | * | * | 80 | 6 | * | 35,475 | |
| 2 | Host B | * | * | 80 | * | * | 14,477 | HTTP Proxy |
| 2 | * | * | * | 80 | 7 | * | 16,653 | |
| 2 | Host C | * | * | 80 | * | * | 15,230 | HTTP Cache |
| 2 | * | * | * | 80 | 5 | * | 58,304 | |
| 3 | * | * | * | 80 | 1 | 46 | 17,212 | |
| 3 | * | * | * | 80 | 1 | 48 | 11,833 | |
| 3 | * | * | * | 80 | 1 | 1024 | 23,696 | |
| 3 | * | * | * | 7000 | 1 | 48 | 12,672 | Dist. Flooding |
| 4 | * | Host D | * | 9022 | 1 | 48 | 22,573 | Backscatter |
| 5 | * | Host E | 54545 | 7000 | 1 | 46 | 23,799 | Dist. Flooding |
| 5 | * | Host E | 45454 | 7000 | 1 | 46 | 15,627 | Dist. Flooding |

Accuracy

- Use a two week NetFlow trace from SWITCH
- Manually classify generated itemsets as true/false positives
- Zero false positive itemsets for 21 anomalies (out of 31)
- False positive itemsets for remaining 10 anomalies →
- On average between 2 and 8.5 false positive itemsets



Conclusions

- Combination of histogram-based detectors and association rule mining works well for extracting anomalous flows

- Further reading:

Daniela Brauckhoff, Xenofontas Dimitropoulos, Arno Wagner, Kave Salamatian “Anomaly Extraction in Backbone Networks using Association Rules” ACM Sigcomm Internet Measurement Conference (IMC), Nov 2009.

Daniela Brauckhoff, Xenofontas Dimitropoulos, Arno Wagner, Kave Salamatian “Anomaly Extraction in Backbone Networks using Association Rules” TIK Technical Report 309, Oct 2009.