Unsupervised Network Anomaly Detection

Juliette Dromard

LAAS-CNRS, Toulouse-France

July 6, 2015

LAAS CNRS
INSTITUT CARNOT LAAS CNRS

ONTIC

SEVENTH FRAMEWORK PROGRAMME

Outline

1. Network security
2. Unsupervised Network Anomaly Detection
    - Some Terms
    - Incremental Unsupervised Network Anomaly Detector
    - Some results
3. Conclusion

Outline

1. Network security

2. Unsupervised Network Anomaly Detection
   - Some Terms
   - Incremental Unsupervised Network Anomaly Detector
   - Some results

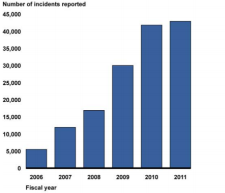
3. Conclusion

Network Security

Network Security

- Network's attacks are increasing
- These attacks are costly

Figure 2: Incidents Reported to US-CERT: Fiscal Years 2006-2011



Source: GAO analysis of US-CERT data for fiscal years 2006-2011.

Existing solutions: knowledge-based detection

- Signature-based detection
  - can't detect attacks they don't know: many false negatives
- Behavior-based detection
  - detect as an attack a new normal behavior: many false positives

Outline

1. Network security
2. Unsupervised Network Anomaly Detection
    - Some Terms
    - Incremental Unsupervised Network Anomaly Detector
    - Some results
3. Conclusion

Some Terms

Network Anomaly

- Rare flow which pattern is different from other flows (normal network traffic)
- Of interest for network's administrators as it may be induced by an attack or a network failure

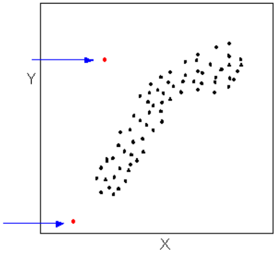
Unsupervised Network Anomaly Detection

- Detect anomalies in an unsupervised way, i.e. without previous knowledge on the anomalies
- Solve the problem of knowledge-based detectors as signature -based detectors and behavioral-based detectors

Incremental Unsupervised Network Anomaly Detector

How to detect anomalies ?

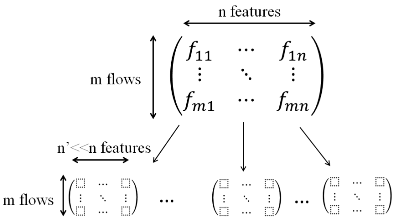
- Use of clustering techniques
- Group similar flows into clusters
- Flows that are rare and different from the others are isolated, they represent anomalous flows.



Incremental Unsupervised Network Anomaly Detector

How can I cluster flows ? Subspace clustering

- Flows are represented by a set of features (nbSyn, nbPackets, nbICMP, ...) around 15 features
- Its is not possible to cluster high dimensional space because of the curse of dimensionality
- Need to divide the space in many subspaces and cluster each subspace independently



Incremental Unsupervised Network Anomaly Detector

How can I cluster flows ? Subspace clustering

- Flows are represented by a set of features (nbSyn, nbPackets, nbICMP, ...) around 15 features
- Its is not possible to cluster high dimensional space because of the curse of dimensionality
- Need to divide the space in many subspaces and cluster each subspace independently

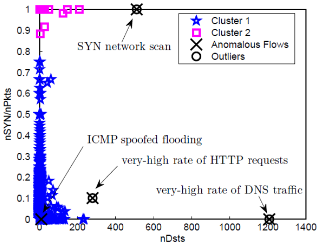
How can I then identify anomalous flows ? Evidence Accumulation

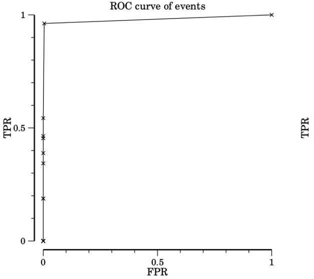
- Apply evidence accumulation techniques in order to group the results obtained in the different subspaces
- Anomalous flows are flows which are outliers in many subspaces and very far from normal flows

Results

Description of the evaluation

- Results obtained on labelled network traces which were collected between Japan and the states (MAWI traces)
- These labelled traces are used as ground truth for the evaluation





(a) ROC curves of events

Outline

1. Network security
2. Unsupervised Network Anomaly Detection
   - Some Terms
   - Incremental Unsupervised Network Anomaly Detector
   - Some results
3. Conclusion

Conclusion

IUNAD

- Based on subspace clustering
- Allow to detect anomalies in an unsupervsied manner

Future Works

- Make more evaluation
- Root cause analysis
  - Identify whether an anomaly is an attack, a network failure or a benign flow
  - No current literature on this subject
  - Analysing anomalies in time