Implementation and Evaluation of Network Anomaly Detection Algorithms

Modules

Control Unit

- The Control Unit centrally manages communication between the rest of the modules.

Traffic Capture Module

- The Traffic Capture Module captures traffic required for learning phase, and computes traffic statistics as required for the Learning Module.

Learning Module

- The LM computes the model probabilities which is later used by the Detection Module (DM) for classification of normal traffic/anomalous traffic.

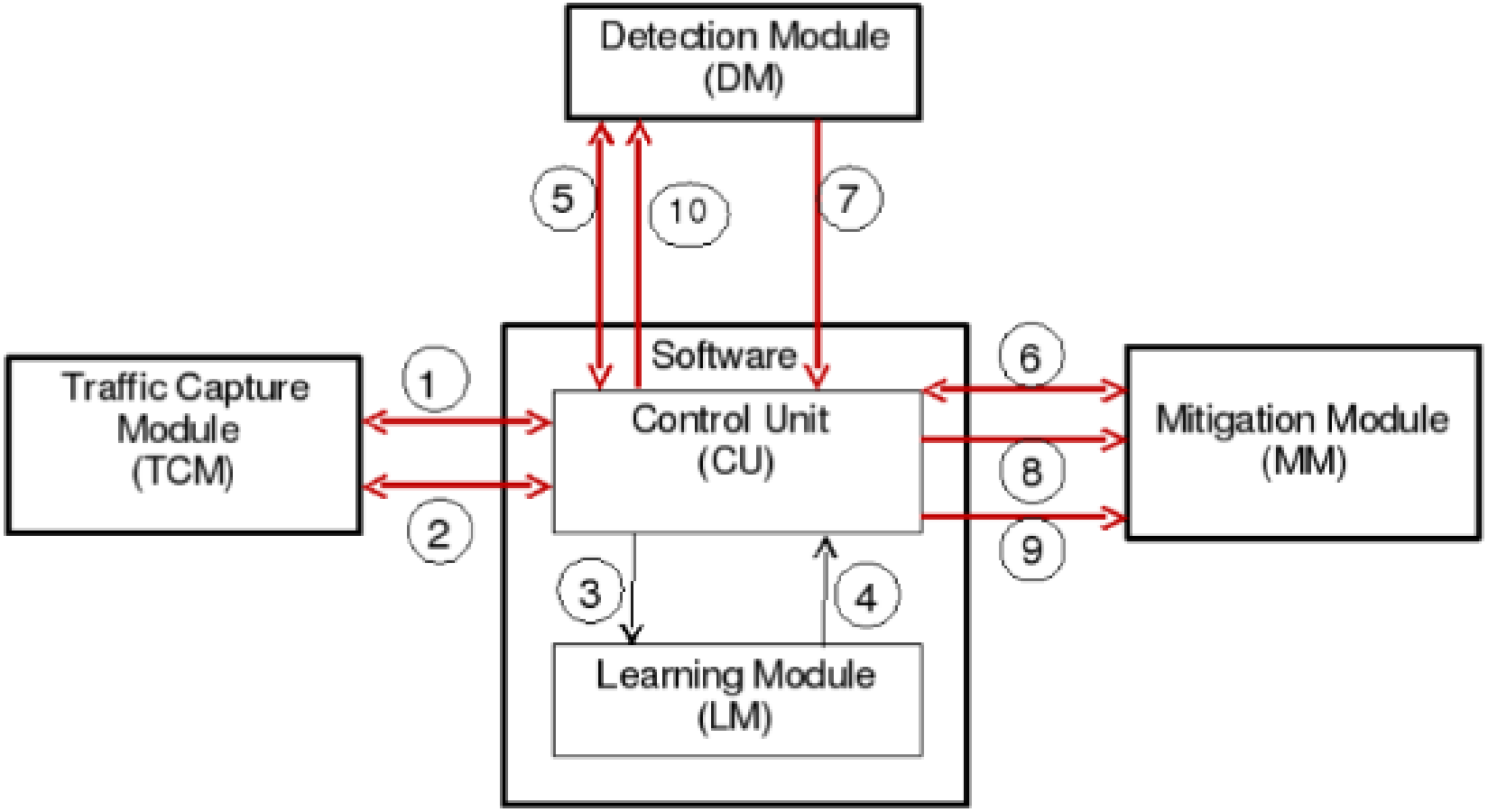
Detection Module

- The Detection Module determines the state of the network (normal or abnormal) in real-time.

Mitigation Module

- The Mitigation Module
  - When attack flag is not set in any stream, it allows input traffic unaltered.
  - When attack flag is set in one or more streams, it filters traffic for the stream(s) on the basis of a white list of source IP's corresponding to the stream.

Modules


Detection Module
(DM)
5
10
7
Software
Traffic Capture
Module
(TCM)
1
2
Control Unit
(CU)
6
8
9
Mitigation Module
(MM)
3
4
Learning Module
(LM)

Algorithms

Deploy

- This module tests the input traffic and computes if a window is an attack window or not.
- Input: Stream S, filename, Abnormal Window Count (AWC)
- Output: State of the system and the window counts

Train Phase Functionality

- This procedure initiates training. It calls the training module that trains the NBC against the training dataset. It also calls the module that determines the threshold probability.
- Input: Window size N, number of bands K, group count, error proportion t and the filename containing the dataset.
- Output: Updated Stream S with learnt probabilities and threshold probability for stream.

Train

- This procedure trains the Naive Bayes Classifier using the TCP flags retrieved from the dataset.
- Input: Stream S, Flag Array TF, Window size for stream N , number of bands K.
- Output: Updated Stream S with learnt probabilities.

Determine Optimal Bands

- This procedure groups the events into bands. The band formation is based on the clustering algorithm.
- Input: Event array A, window size for stream N, number of bands K.
- Output: 2-dimensional array band, containing indices of per group elements in A.

Update Probabilities

- This procedure computes and updates the probability based on the events.
- Input: TCP Stream S, 2-D array band containing band indices of groups of A, number of bands K, Index of observable type L
- Output: Probabilities of all events updated into corresponding probability array W index in stream S

Determine Threshold Probability

- This module computes the threshold probability which will be used to classify if a particular window is either an attack window or not
- Input: Stream S, Window size for stream N , group count , flag array
- Output: Threshold probability Pt for S

Datasets used in the paper

- DARPA
  - Training data - 70,603 packets to port 23.
  - Attack traffic ->

Packet Numbers	Nature of Attack
1-1046	SYN flooding
1047-2277	FIN/ACK flooding
2278-2543	SYN/ACK flooding
2544-2911	PSH/ACK flooding
2912-3629	SYN flooding
3630-4806	RST flooding
4807-6077	FIN/ACK flooding
6078-16115	SYN flooding
16116-22934	ACK flooding
22935-31864	PSH/ACK flooding
31865-42349	RST flooding
42350-97649	FIN/ACK flooding
97650-192549	SYN flooding
192550-545549	ACK flooding
545550-1158549	SYN flooding

Our datasets

- DARPA Train – 219600 packets = 1999 Week 1, Monday + Thursday (Dest-172.16.112.50:23)
- DARPA 1
  - No Attack – 61400 packets - 1999 Week 1, Wednesday (Dest-172.16.112.50:23)
  - Attack – 74200 packets - 1999 Week 1, Tuesday + attacks (Dest-172.16.112.50:23)

Our Datasets (contd.)

- DARPA 2
  - No Attack – 126400 packets - 1999 Week 3, Tuesday + Wednesday (Dest-172.16.112.50:23)
  - Attack – 1164 packets - 1999 Week 3, Friday + attacks (Dest-172.16.112.50:23)
- DARPA 3
  - Attack – 244107 packets (39303 NA / 204804A) - 1998 Week 3, Thursday (Dest-172.16.112.50 all ports)

Accuracy results (t = 0.08)

		Actual		Detected		Accuracy %
		Normal	Attack	Normal	Attack	
DARPA 1	**No Attack file**	614	0	614	0	100.00
	Attack File	371	371	377	365	98.38
DARPA 2	**No Attack file**	1264	0	1241	23	98.18
	Attack File	582	582	556	608	95.53
DARPA 3	**Attack File**	393	2048	656	1785	87.16
Thesis	**Attack File**	0	11594	6	11588	99.95

Accuracy results (t = 0.091)

		Actual		Detected		Accuracy %
		Normal	**Attack**	**Normal**	**Attack**	
DARPA 1	**No Attack file**	614	0	607	7	98.86
	Attack File	371	371	371	371	100.00
DARPA 2	**No Attack file**	1264	0	1210	54	95.73
	Attack File	582	582	550	614	94.50
DARPA 3	**Attack File**	393	2048	347	2094	88.30
Thesis	**Attack File**	0	11594	2	11592	99.98

Accuracy Comparison