Ensembles for intrusion detection

Alexandre Balon-Perin[1,2] **Björn Gambäck**[1,3] **Lillian Røstad**[1]

[1]Department of Computer and Information Science
Norwegian University of Science and Technology (NTNU)
Trondheim, Norway

[2]Ecole Polytechnique
Université Libre de Bruxelles (ULB)
Brussels, Belgium

[3]SICS — Swedish Institute of Computer Science AB
Kista, Sweden

November 21, 2012

Goals

1. Develop the state-of-the-art for ensemble-based methods applied to intrusion detection
2. Show that, when trying to detect attacks on a network, each class of attacks should be treated separately
   $\Rightarrow$ Apply one algorithm with one set of features to one class of attacks
3. Compare ensemble-based methods with more standard approaches

Overview

1. Security
   - Intrusion detection systems
   - Classes of attacks
2. Machine learning
   - Machine learning and its drawbacks
   - The KDD99 dataset
   - Ensemble approaches
   - Feature selection
3. Experiments
   - Experiment 1: Feature selection
   - Experiment 2: Model assessment
4. Conclusion
   - Final Model

Intrusion detection systems

Generalities

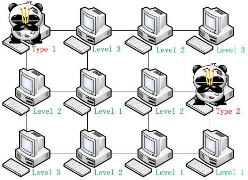
1. Devices monitoring a network to detect anomalous behaviours
2. Network-based IDS

1. Misuse-based detection
2. Anomaly-based detection

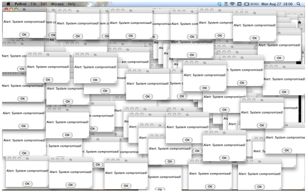
Problem

Need 100% accuracy → 0 False Positives AND 0 False Negatives

False Negative



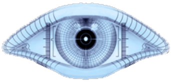
False Positive

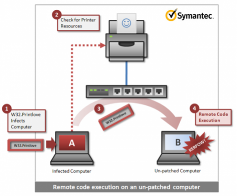


Classes of attacks

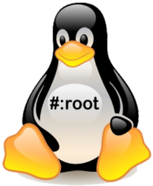
Attacks on a network can be divided into four classes:

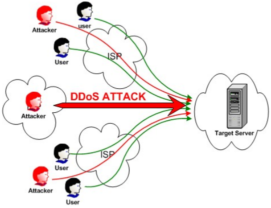
1. Probe
2. Remote to local (R2L)
3. User to root (U2R)
4. Denial of Service (DoS)





(from Symantec)





Machine learning

Goal

Classify unseen examples as normal or anomalous traffic

Why machine learning?

Inability for misuse-based IDSs to detect

- new attacks
- variants of known attacks

Drawbacks

1. Performance degradation when examples are very different from the ones in the training set
2. Only application of ML where users try to fool or attack the system

Machine learning

Goal

Classify unseen examples as normal or anomalous traffic

Why machine learning?

Inability for misuse-based IDSs to detect

- new attacks
- variants of known attacks

Drawbacks

1. Performance degradation when examples are very different from the ones in the training set
2. Only application of ML where users try to fool or attack the system

Machine learning

Goal

Classify unseen examples as normal or anomalous traffic

Why machine learning?

Inability for misuse-based IDSs to detect

- new attacks
- variants of known attacks

Drawbacks

1. Performance degradation when examples are very different from the ones in the training set
2. Only application of ML where users try to fool or attack the system

The KDD99 dataset

1. Knowledge Discovery and Data Mining 1999
2. Modified version of the dataset developed by the Defense Advanced Research Projects Agency (DARPA) in 1998
3. 4,898,431 entries for the training set
4. 311,029 entries for the test set
5. 41 variables including time-related and content-related features
6. Labels representing the type of attack of the example or "normal" if the traffic is considered harmless

The KDD99 dataset

Pros & Cons

- Cons:
  1. Developed in 1998 $\rightarrow$ many attacks are obsolete
  2. Unbalanced distribution of examples
  3. Developed in a simulated environment different from the real world
  4. The test set contains many unseen types of attacks

- Pros:
  1. The only labelled dataset publicly available
  2. Used in many studies $\rightarrow$ good comparison tool
  3. IDSs should at least perform well on these attacks to be useful

The KDD99 dataset

Pros & Cons

- Cons:
  1. Developed in 1998 → many attacks are obsolete
  2. Unbalanced distribution of examples
  3. Developed in a simulated environment different from the real world
  4. The test set contains many unseen types of attacks

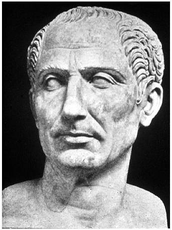
- Pros:
  1. The only labelled dataset publicly available
  2. Used in many studies → good comparison tool
  3. IDSs should at least perform well on these attacks to be useful

Why an ensemble?

Properties

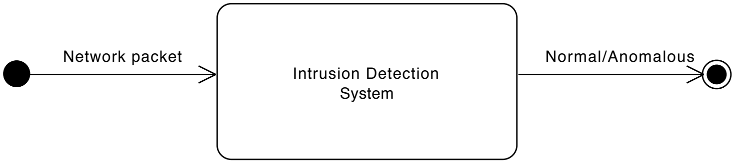
1. The ensemble approach is a machine learning paradigm which combines several algorithms
2. Two properties make ensembles suitable for the problem of intrusion detection:

"Divide and Conquer"

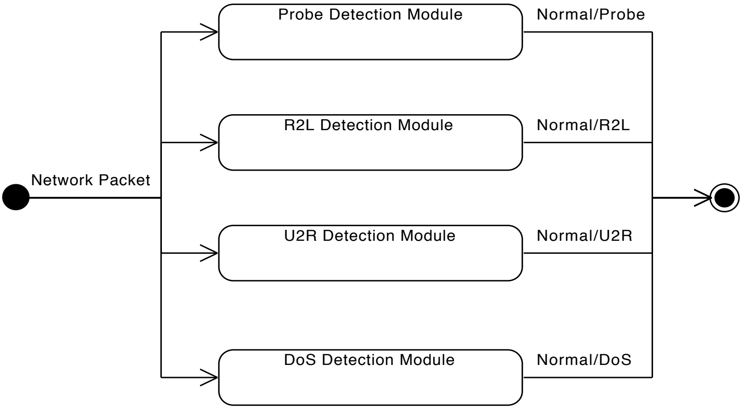


"Unity is Strength"

Why an ensemble?
"Divide and conquer"

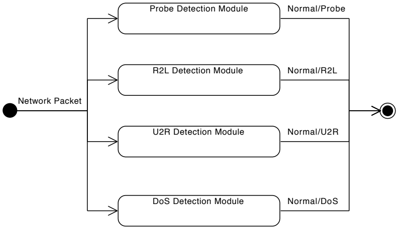


Why an ensemble?
"Divide and conquer"



Why an ensemble?

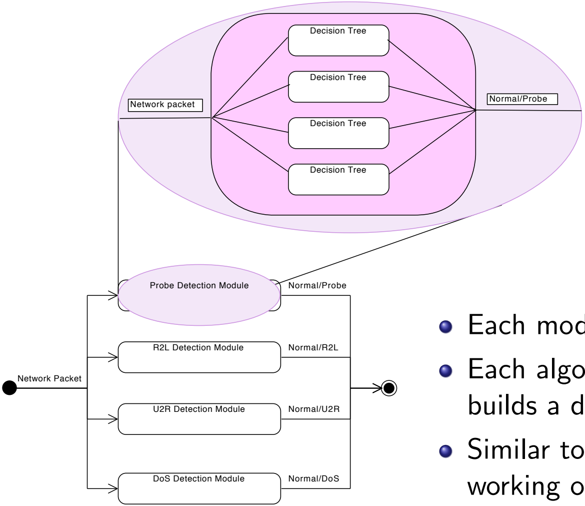
"Unity is Strength"



- Each module is an ensemble
- Each algorithm of the ensemble builds a different model
- Similar to several specialists working on the same problem

Why an ensemble?

"Unity is Strength"



- Each module is an ensemble
- Each algorithm of the ensemble builds a different model
- Similar to several specialists working on the same problem

Feature selection

Principle

Select a subset of variables from the dataset or transform the variables into a lower dimensional space

Main goals

- Remove irrelevant information
- Speed up the computation
- Speed up the preprocessing phase

The idea

1. Select a different set of features for each class of attacks
2. Three feature selection algorithms: SVM, LGP and MARS

(Support Vector Machines, Linear Genetic Programming, Multivariate Adaptive Regression Splines)

Feature selection

Principle

Select a subset of variables from the dataset or transform the variables into a lower dimensional space

Main goals

- Remove irrelevant information
- Speed up the computation
- Speed up the preprocessing phase

The idea

1. Select a different set of features for each class of attacks
2. Three feature selection algorithms: SVM, LGP and MARS

(Support Vector Machines, Linear Genetic Programming, Multivariate Adaptive Regression Splines)

Feature selection

Principle

Select a subset of variables from the dataset or transform the variables into a lower dimensional space

Main goals

- Remove irrelevant information
- Speed up the computation
- Speed up the preprocessing phase

The idea

1. Select a different set of features for each class of attacks
2. Three feature selection algorithms: SVM, LGP and MARS

   (Support Vector Machines, Linear Genetic Programming, Multivariate Adaptive Regression Splines)

Several decision trees were trained with different sets of features.

The evaluation was performed on the training set using a 10-fold cross-validation

Goal

Conclude on how well the algorithms perform with a smaller set of features

Remarks

1. "combined" set of features
2. ensemble$_{\mathrm{max}}$

Several decision trees were trained with different sets of features.
The evaluation was performed on the training set using a 10-fold cross-validation

Goal

Conclude on how well the algorithms perform with a smaller set of features

Remarks

1. "combined" set of features
2. $\text{ensemble}_{\max}$

Table: Accuracy of the feature selection assessment

Classifier	Probe	U2R	R2L	DoS
DT: 41 features	99.86	93.00	99.02	99.95
DT: 5 SVM features	99.82	96.00	98.58	93.35
DT: 5 LGP features	99.32	90.00	97.38	98.69
DT: 5 MARS features	99.75	97.00	98.04	99.86
DT: combined features	99.90	96.00	98.93	99.95
Peddabachigari et al.	99.86	68.00	84.19	96.83
Wu and Banzhaf	97.29	76.30	80.22	99.70

Experiments

Feature selection assessment - Results

Table: False positives and false negatives

	Probe		U2R		R2L		DoS	
Classifier	FP	FN	FP	FN	FP	FN	FP	FN
DT: 41 features	12.0	17.0	4.0	3.0	17.0	10.0	6.0	8.0
$\texttt{ensemble}_{\max}$	0.7	3.0	0.3	0.3	6.6	0.5	0.0	1.6

Several decision trees were trained with different sets of features.

The evaluation was performed on the test set

Goal

Assess if the ensemble could generalize to new types of attacks

Table: Accuracy of the model assessment

Classifier	Probe	U2R	R2L	DoS
DT: 41 features	93.09	90.00	50.00	79.34
DT: 5 SVM features	77.63	40.00	50.00	87.70
DT: 5 LGP features	87.48	83.57	61.03	76.10
DT: 5 MARS features	84.04	85.00	50.00	82.20
DT: combined features	79.97	94.29	50.00	85.36

Table: False positives and false negatives

	Probe		U2R		R2L		DoS	
Classifier	FP	FN	FP	FN	FP	FN	FP	FN
DT: 41 features	86.0	490.0	3.0	11.0	0.0	16,347.0	69.0	7,268.0
$\text{ensemble}_{\max}$	11.4	524.0	1.6	1.0	1.0	7,779.0	16.6	688.0

Concluding Remarks

Conclusions

- The ensemble improved the accuracy
- Ensemble approaches help reducing FP and FN
- The features selected by Mukkamala et al. are mostly appropriate
- Most misclassifications were caused by very specific features

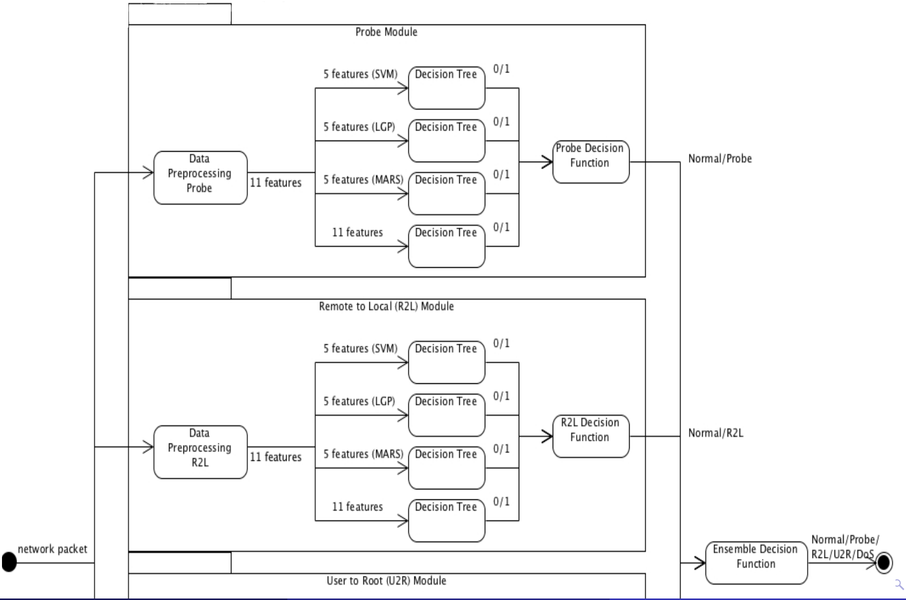
Warning

- Accuracy not yet good enough for real-world applications
- Results obtained in Experiment 2 were less interesting because of inappropriate distribution of examples

Future work

1. Framework for ensemble approaches applied to intrusion detection
   1. Testing centre
   2. Multi core architecture
2. Make the system reactive
3. Active learning to quickly create datasets

Final Model



Probe Module
5 features (SVM)
Decision Tree
0/1
5 features (LGP)
Decision Tree
0/1
Data Preprocessing Probe
11 features
5 features (MARS)
Decision Tree
0/1
Probe Decision Function
Normal/Probe
11 features
Decision Tree
0/1
Remote to Local (R2L) Module
5 features (SVM)
Decision Tree
0/1
5 features (LGP)
Decision Tree
0/1
R2L Decision Function
Normal/R2L
Data Preprocessing R2L
11 features
5 features (MARS)
Decision Tree
0/1
11 features
Decision Tree
0/1
network packet
Ensemble Decision Function
Normal/Probe/ R2L/U2R/DoS
User to Root (U2R) Module