RANSOMWARE DETECTION USING PROCESS MEMORY

Avinash Singh, Richard Adeyemi Ikuesan, and Hein Venter

17th International Conference on Cyber Warfare and Security, 03/2022

THE CHALLENGE

Ransomware attacks have significantly increased, causing major disruptions to business operations

As of 2025, **over 78% of businesses** were affected by ransomware attacks*.

Traditional detection methods are easily bypassed by sophisticated ransomware

*https://tech.co/news/most-companies-hit-with-ransomware

Methodology

Instead of looking at the software's behavior on the network, we look at how the software uses the computer's memory

Read/Write/Execute/
Copy (RWXC)

analyzing process memory to detect ransomware behavior

Machine Learning algorithms: Decision Tree, Random Forest, Gradient Boosted Trees, etc.

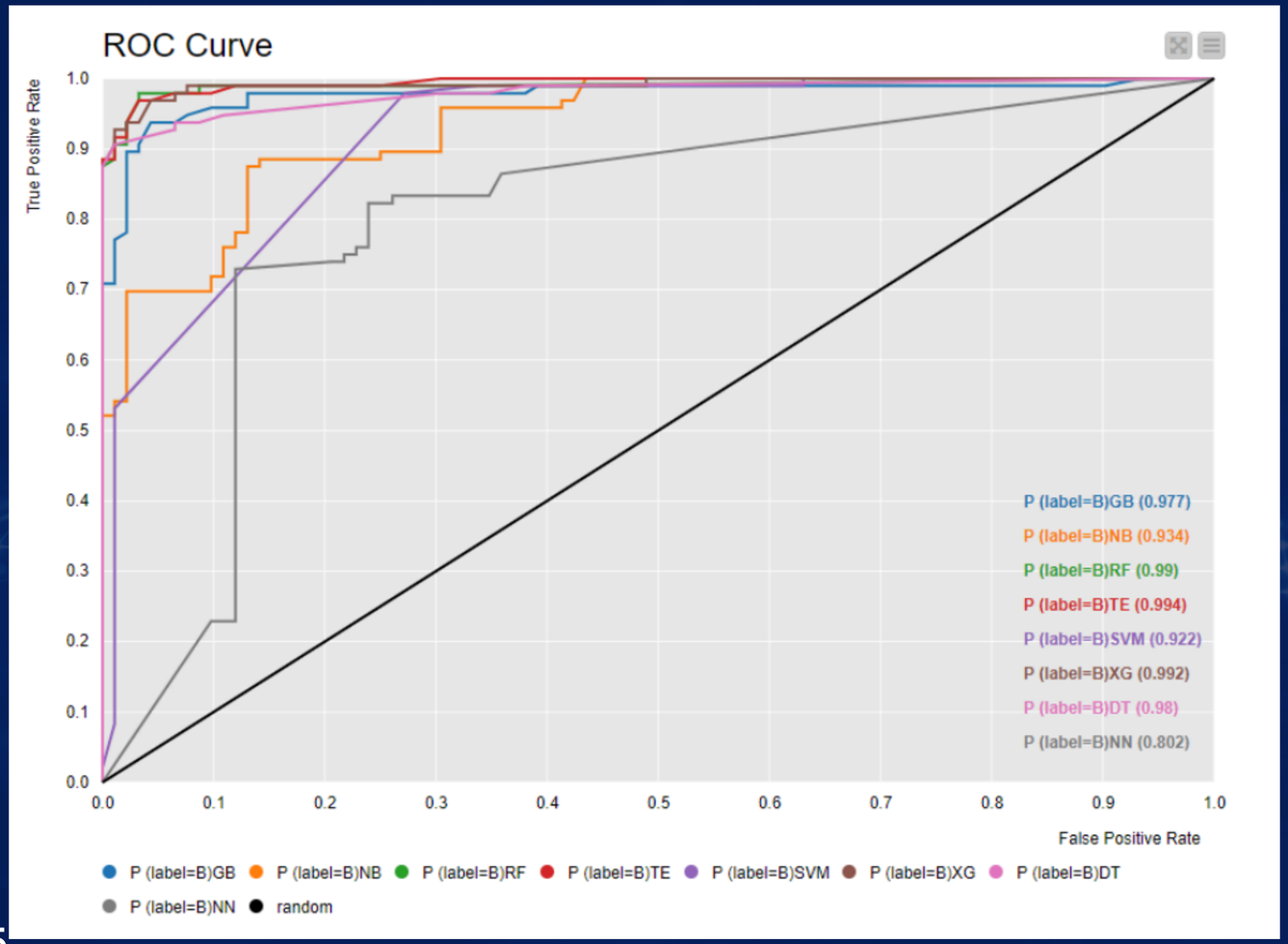
Data Set

https://github.com/icfl-up/rdpm%D7%A5

- Total records: 937
- Benign samples: 476 (354 unique applications).
- Malicious samples: 461 (117 ransomware variants from 70+ families).
- Train-Test Split: 80/20 with stratified sampling to maintain class balance.

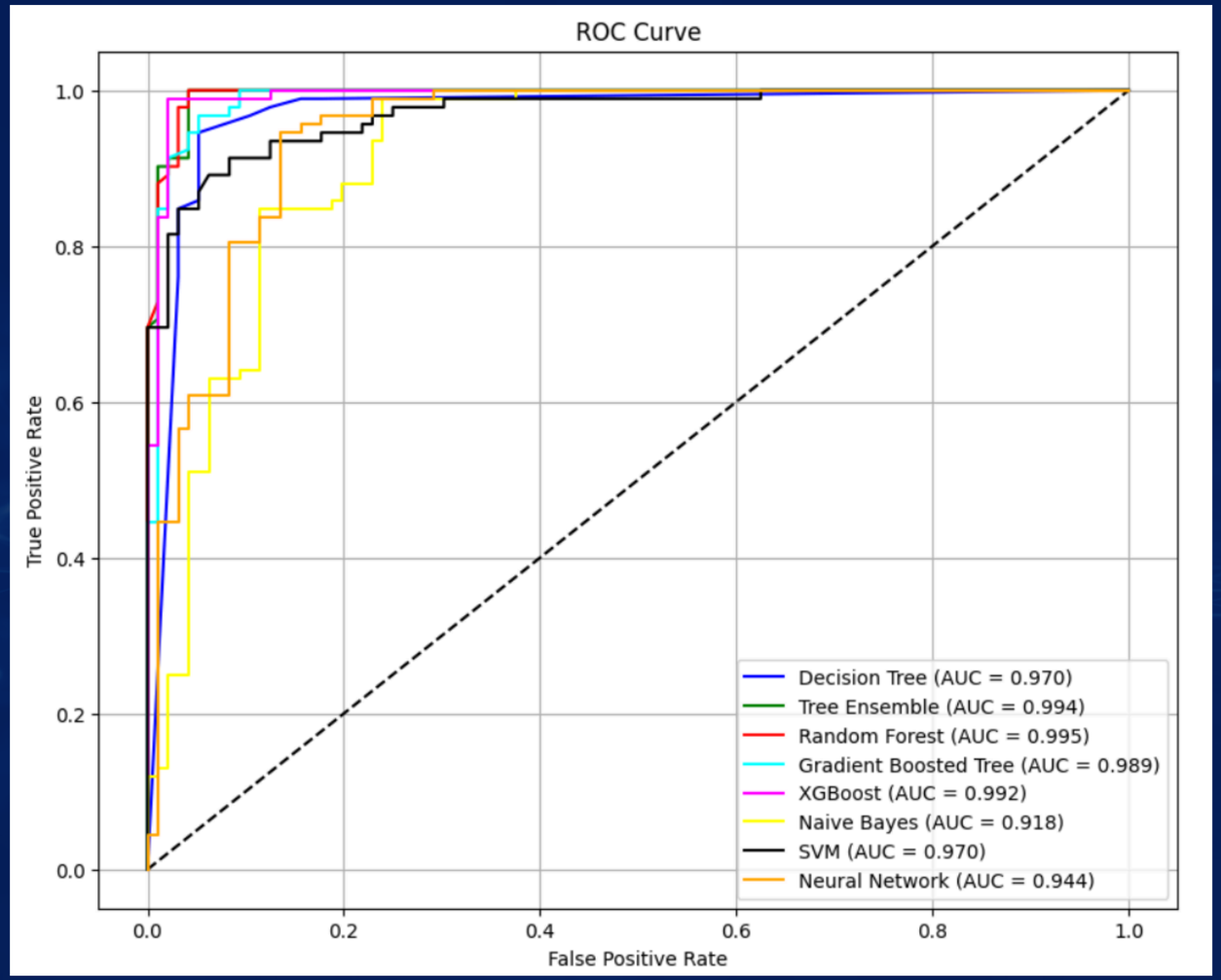
R	RW	RX	RWC	RWX	RWXC	LABEL	CATEGORY
367	307	117	84	70	0	B	Utilities
107	96	43	31	5	0	B	Utilities
62	59	18	21	49	0	B	Utilities
13	13	4	5	4	0	M	Cryptowall
124	94	54	32	6	0	M	DarkSide
239	289	79	104	98	22	M	DeathHiddenTear

EXPECTED OUTCOME:



Model	result
XGBoost (XG)	96.28%
Gradient-Boosted Tree (GB)	94.68%
Decision Tree (DT)	93.62%
Neural Network (NN)	93.62%
Support Vector Machine (SVM)	85.64%
Naive Bayes (NB)	81.38%
Tree Ensemble (TE)	95.74%
Random Forest (RF)	95.21%

OUR RESULTS:



Model	result
XGBoost (XG)	98.6%
Gradient-Boosted Tree (GB)	94.68%
Decision Tree (DT)	90.57%
Neural Network (NN)	86.7%
Support Vector Machine (SVM)	89.36%
Naive Bayes (NB)	84.04%
Tree Ensemble (TE)	94.68%
Random Forest (RF)	96.27%

More detailed look at the results...

Decision Tree

Model Configuration:

- Split Criterion: Gini Index
- Minimum Samples per Leaf: 5 records
- Pruning: No pruning (max_depth unlimited)
- Random State: 42 for reproducibility

Results Evaluation:

paper

Label	TP	FP	TN	FN	Recall	Precision	Sensitivity	Specificity	F-measure	Accuracy
B	90	6	86	6	0.9375	0.9375	0.9375	0.9347	0.9375	
M	86	6	90	6	0.9347	0.9347	0.9347	0.9375	0.9347	
Overall										0.9362

ours

Label	TP	FP	TN	FN	Recall	Precision	F-measure	Accuracy
B	93	14	3	78	0.8692	0.9688	0.9163	
M	78	3	93	14	0.8478	0.9630	0.9017	
Overall								0.9096

Random Forest

Model Configuration:

- Number of Trees: 100 receive the best result
- Criterion: Gini
- Random State: 42

Results Evaluation:

paper

Label	TP	FP	TN	FN	Recall	Precision	Sensitivity	Specificity	F-measure	Accuracy
B	89	2	90	7	0.9271	0.9780	0.9271	0.9783	0.9519	
M	90	7	89	2	0.9783	0.9278	0.9783	0.9271	0.9524	
Overall										0.9521

ours

Label	TP	FP	TN	FN	Recall	Precision	F-measure	Accuracy
B	93	4	3	88	0.9588	0.9688	0.9637	
M	88	3	93	4	0.9565	0.9670	0.9617	
Overall								0.9628

Gradient-Boosted Tree

Model Configuration:

- Boosting Stages: 100 sequential trees - gave the best results
- Learning Rate: 0.1 (shrinkage parameter for regularization)
- Tree Depth: 3 levels per tree - gave the best results
- Loss Function: Logistic regression for binary classification

Results Evaluation:

paper

Label	TP	FP	TN	FN	Recall	Precision	Sensitivity	Specificity	F-measure	Accuracy
B	90	4	88	6	0.9375	0.9574	0.9375	0.9565	0.9474	
M	88	6	90	4	0.9565	0.9362	0.9565	0.9375	0.9462	
Overall										0.9468

ours

Label	TP	FP	TN	FN	Recall	Precision	F-measure	Accuracy
B	91	5	5	87	0.9479	0.9479	0.9479	
M	87	5	91	5	0.9457	0.9457	0.9457	
Overall								0.9468

XG Boost

Model Configuration:

- Boosting Rounds: 1000 trees
- Tree Depth: Max of 6 levels
- objective: binary : logistic
- Random State: 42 for reproducibility

the best result

Results Evaluation:

paper

Label	TP	FP	TN	FN	Recall	Precision	Sensitivity	Specificity	F-measure	Accuracy
B	93	4	88	3	0.9688	0.9588	0.9688	0.9565	0.9637	
M	88	3	93	4	0.9565	0.9670	0.9565	0.9688	0.9617	
Overall										0.9628

ours

Label	TP	FP	TN	FN	Recall	Precision	F-measure	Accuracy
B	94	4	2	88	0.9592	0.9792	0.9691	
M	88	2	94	4	0.9565	0.9778	0.9670	
Overall								0.9681

Naive Bayes

Model Configuration:

- Model Type: Gaussian NB (assumes normal distribution of features)

Results Evaluation:

paper

Label	TP	FP	TN	FN	Recall	Precision	Sensitivity	Specificity	F-measure	Accuracy
B	71	10	82	25	0.7396	0.8765	0.7396	0.8913	0.8023	
M	82	25	71	10	0.8913	0.7664	0.8913	0.7396	0.8241	
Overall										0.8138

ours

```
Results for Naive Bayes:
Label   TP      FP      TN      FN      Recall    Precision   F-measure   Accuracy
B       74      8       22      84      0.9024    0.7708      0.8315
M       84      22      74      8       0.9130    0.7925      0.8485
Overall                                                                   0.8404
```

SVM

Model Configuration:

- Kernel: Radial Basis Function (RBF)
- Gamma: 0.1 (controls influence of a single training)
- Probability : Enabled (=True)
- Random State: 42

Results Evaluation:

paper

Label	TP	FP	TN	FN	Recall	Precision	Sensitivity	Specificity	F-measure	Accuracy
B	94	25	67	2	0.9792	0.7899	0.9792	0.7283	0.8744	
M	67	2	94	25	0.7283	0.9710	0.7283	0.9792	0.8323	
Overall										0.8564

ours

```
Results for SVM:
Label   TP      FP      TN      FN      Recall    Precision   F-measure   Accuracy
B       94      18      2       74      0.8393    0.9792      0.9038
M       74      2       94      18      0.8043    0.9737      0.8810
Overall                                                                   0.8936
```

Tree Ensemble

Model Configuration:

- Number of Trees: 100 models
- Split Criterion: Entropy-Information Gain
- max_features: 'sqrt'

Results Evaluation:

paper

Label	TP	FP	TN	FN	Recall	Precision	Sensitivity	Specificity	F-measure	Accuracy
B	90	2	90	6	0.9375	0.9783	0.9375	0.9783	0.9574	
M	90	6	90	2	0.9783	0.9375	0.9783	0.9375	0.9574	
Overall										0.9574

ours

Label	TP	FP	TN	FN	Recall	Precision	F-measure	Accuracy
B	92	6	4	86	0.9388	0.9583	0.9485	
M	86	4	92	6	0.9348	0.9556	0.9451	
Overall								0.9468

Neural Network

Model Configuration:

- Architecture: Multi-Layer Perceptron (MLP)
- Hidden Layers: 2 layers ,15 neurons each
- Activation Function: ReLU
- Solver: ‘Adam’
- Iterations: 10,000
- Random State: 42

Results Evaluation:

paper

Label	TP	FP	TN	FN	Recall	Precision	Sensitivity	Specificity	F-measure	Accuracy
B	120	8	144	10	0.9231	0.9375	0.9231	0.9474	0.9302	
M	144	10	120	8	0.9474	0.9351	0.9474	0.9231	0.9412	
Overall										0.9362

ours

```
Results for Neural Network:
Label   TP      FP      TN      FN      Recall    Precision   F-measure   Accuracy
B       74      3       22      89      0.9610    0.7708      0.8555
M       89      22      74      3       0.9674    0.8018      0.8768
Overall                                                                   0.8670
```

STEPS & EXECUTION

Task Description	
Data Loading & Preprocessing • aquire dataset • process the data	✓
Implement models: • Decision Tree • Random Forest • Tree Ensemble • Gradient-Boosted Trees • XGBoost • Naive Bayes • Support Vector Machine • Neural Network	✓
Model Comparison & Testing	✓
Documentation & Analysis	✓

REFERENCES &BIBLIOGRAPHY

- Original Paper:
  - Singh, A., Ikuesan, R. A., & Venter, H. (2022). Ransomware Detection using Process Memory. https://www.researchgate.net/publication/359034291_Ransomware_Detection_using_Process_Memory

- Additional References:
  - Malware repositories: Malware Bazaar, TheZoo.
  - Cuckoo Sandbox: https://cuckoosandbox.org.
  - https://github.com/icfl-up/rdpm%D7%A5
  - Machine Learning techniques for malware detection.
  - https://tech.co/news/most-companies-hit-with-ransomware

THANK YOU FOR YOUR ATTENTION!

ANY QUESTIONS?