Labubusec Project Overview

By: Judson, Sam, Jonathan, Toren, Alex

01

Attack

Methodology & Tooling

- Loaders are the way
  - load malware into a bmp, then generate dropper
- Iterative transformation
- Batch outputs (don't put all eggs in one basket)
- post compilation
  - UPX
  - Astral PE
  - BOAZ

Figure: bitmap for 7.exe

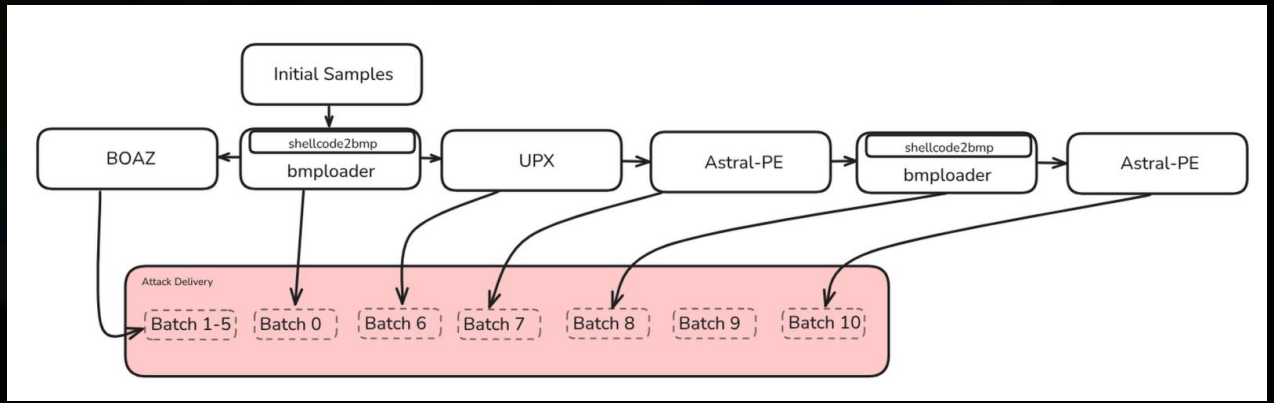


Figure: Attack Flow Diagram*

*Batch 9 is Astral PE, Batch 10 is Batch 8 with low entropy data appended

Attack Verification & Evaluation

Methodology

- VirusTotal
- Hybrid Analysis
- Local Testing

Results

- Kinda Bad
  - Models are inconsistent
  - Other batches more promising
- Not too bad against VT...
  - Hard to evaluate when different samples perform differently. E.x. 4 was in the teens, 11 is in the 20s
- Sandbox proves malware works

Batch	File 11
Control	58/72
0	26/71
8	27/71
10	28/71

Figure: VT for file 11

Batch	Group 1	Group 2	Group 3	Group 5	Group 7	Group 8
0	0 Evaded	0 Evaded	50 Evaded	0 Evaded	0 Evaded	2 Evaded
9	0 Evaded	0 Evaded	50 Evaded	0 Evaded	0 Evaded	0 Evaded
10	0 Evaded	0 Evaded	50 Evaded	0 Evaded	0 Evaded	7 Evaded

Figure: Local testing results

02

Defense

Black box Model

- LightGBM Model trained on ember 2018 EMBER subset
- Stateful nearest neighbor
- Poorly configured thresholds

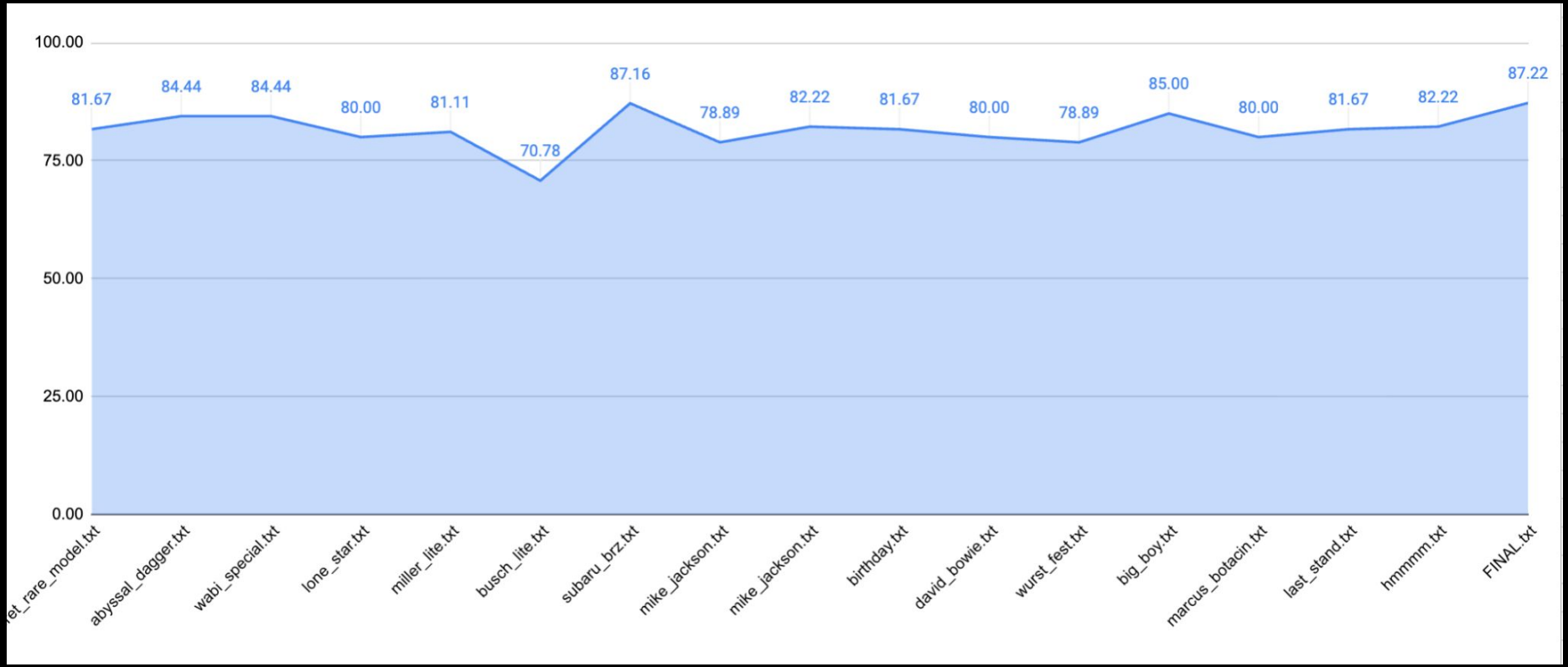
GW1	GW2	GW3	GW4	GW5	GW6	GW7	MW1	MW2	MW3	MW4	MW5	MW6	MW7	MW8	MW9	MW10	MW11
0.00%	0.00%	70.00 %	40.00 %	30.00 %	0.00%	0.00%	80.00 %	60.00 %	100.00 %	100.00 %	100.00 %	100.00 %	80.00 %	100.00 %	90.00 %	90.00 %	50.00 %

White box Model

- LightGBM Model trained on ~25% of 2024 EMBER Dataset
- Stateful Nearest Neighbor
- ChatGPT enhanced Thresholds
- 85% adversary detection rate, 80% overall accuracy
- Timeout problem. But still good results?

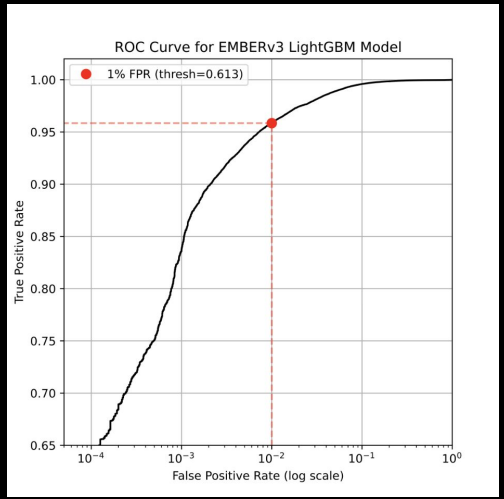
GW1	GW2	GW3	GW4	GW5	GW6	GW7	MW1	MW2	MW3	MW4	MW5	MW6	MW7	MW8	MW9	MW10	MW11
100%	100%	80%	60%	100%	20%	0%	70%	80%	80%	100%	100%	10%	100%	100%	100%	100%	90%

Moving Forward



Final Model

- **1st challenge:** Optimizing Training to use full dataset on a 32gb system.
- **2nd challenge**: Feature Engineering (helpful...?)
- **3rd challenge:** Automated Threshold tuning
- **4th challenge:** Optuna hyperparameter tuning
- **5th challenge:** Time out problem → file truncation + failsafe
- Final results → 87.22% overall acc.
- Future Ideas/Challenges: Multi-model system, actual feature engineering, malware 6...



FINAL.txt	100	100	70	90	90	80	100	70	70	90	100	100	10	100	100	100	100	100	87.22

THANK YOU