SAMSUNG

NEURAL INFORMATION PROCESSING SYSTEMS

CLEAR: Command Level Annotated Dataset for Ransomware Detection

Barak Bringoltz, Elisha Halperin, Ran Feraru, Evgeny Blaichman, Amit Berman

Samsung Semiconductor Israel Research and Development Center

SAMSUNG

Motivation – Ransomware Detection

NEURAL INFORMATION PROCESSING SYSTEMS

- Major cybersecurity threat to organizations worldwide
- Active research challenge
- No publicly available datasets currently provide per-command labeling

SAMSUNG

Contribution

NEURAL INFORMATION PROCESSING SYSTEMS

- Presents the largest dataset to date - **1,045 TiB of data and 137 ransomware variants**
- Enhances per-command labeling to distinguish ransomware from benign
- Auxiliary features derived from the data improve **(1-AUC) by up to 30%**

SAMSUNG

Contribution

NEURAL INFORMATION PROCESSING SYSTEMS

Raw Data

Timestamp	OpCode	Offset	Size
89	Read	600000	256
91	Write	600250	512
⋮			
605	Read	620170	256

Old Labeling

Chunk Label
1

New Attributes

Auxiliary	Higher-level Semantics
0	Process ID=4
OV_{WaR}=6	Process ID=4
⋮	
0	Process ID=9

New Labeling

Command Label
0
1
⋮
0

SAMSUNG

NEURAL INFORMATION PROCESSING SYSTEMS

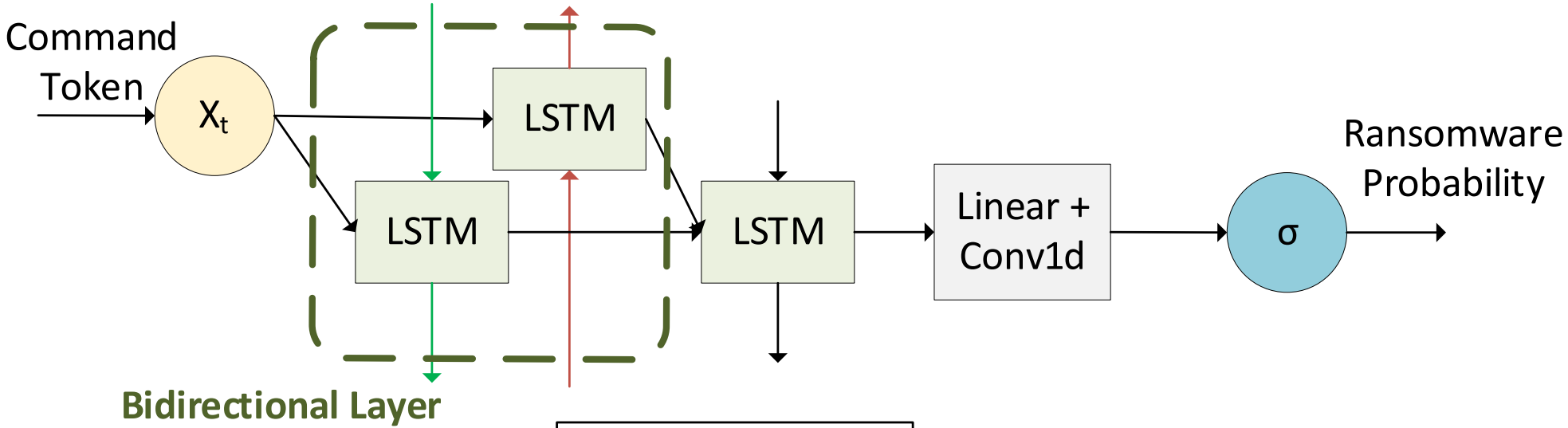
Key Benefits of the Dataset

- Enables per-command sequential models that outperform SoTA
- Presents a benchmark for out-of-distribution model performance

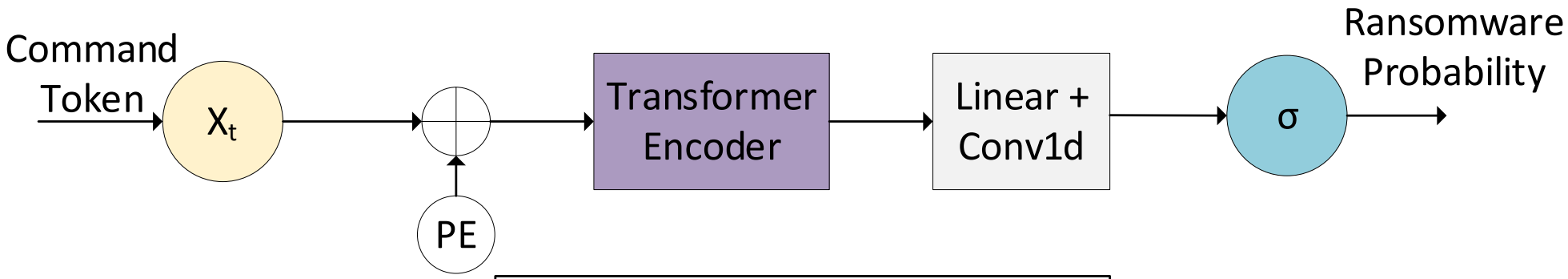
SAMSUNG

NEURAL INFORMATION PROCESSING SYSTEMS

Command Based Models



biLSTM / uLSTM



Command-level Transformer (CLT)

SAMSUNG

NEURAL INFORMATION PROCESSING SYSTEMS

Results

Model	Missed Ransomware Volume [%]	Volume Corrupted Until Detection [MB]	Per-Command Labeling
Random Forest [1]	9.07 ± 1.8	286 ± 33	✗
DeftPunk [2]	6.74 ± 1.2	282 ± 34	
UNET	1.34 ± 0.4	195 ± 41	
Patch Level Transformer	1.16 ± 0.2	157 ± 18	
Command Level Transformer	1.42 ± 0.4	77 ± 10	✓
biLSTM	1.24 ± 0.3	76 ± 07	
uLSTM	**0.36 ± 0.1**	**50 ± 03**	

~3.24x gain **~3.14x gain**

[1] SungHa Baek et al. (2018). "Ssd-insider: Internal defense of solid-state drive against ransomware with perfect data recovery". In: 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS), p 875–884

[2] Zhongyu Wang et al. (2024). "Ransom access memories: Achieving practical ransomware protection in cloud with DeftPunk". In: 18th USENIX Symposium on Operating Systems Design and Implementation (OSDI 24), pages 687–702

SAMSUNG

NEURAL INFORMATION PROCESSING SYSTEMS

Unseen Ransomware Benchmark