TECHNISCHE UNIVERSITÄT DARMSTADT

ICML International Conference On Machine Learning

Long Range Dependency Understanding in State Space Models

Srividya Ravikumar, Shweta Verma, Abhinav Anand, Mira Mezini

SOFTWARE TECHNOLOGY GROUP

Software Technology Group, TU Darmstadt

INTRODUCTION

SSMs are emerging as efficient alternatives for source code vulnerability detection, but their internal kernel behavior remains poorly understood.

- Kernels capture local syntax or long-range dependencies?
- A systematic kernel interpretability study on S4D, trained on real world task.
- Analyze CNN–S4D hybrid kernels in time & frequency domain to guide better SSM design.

Dataset & setup:

22,734 functions	2,240 Vulnerable	20,494 Non- vulnerable

ReVeal dataset: Linux debian kernel + chromium source code. Binary classification.

Training: Adam (lr= 10^{-3}), dropout of 0.5, sequence length of 256. BCEWithLogitsLoss with class weights for imbalance

MODEL ARCHITECTURES

Convolution 1D	Parallel kernel sizes for local features
Depth Wise separable convolutions	Depthwise + pointwise convolutions: fewer parameters
S4D (standalone)	Diagonal SSM layer, FFT based kernel
Depthwise + S4D	Local features + global context
Conv 1D + S4D	Local features + broader dependencies
SMR + S4D	Conv1D fused with embeddings

KERNEL ANALYSIS:

1. Time domain: Peak amplitude, transition sharpness. Reveals local & smooth long range structures.
2. Frequency domain analysis via FFT to obtain the spectral representation of the kernel.

Fourier transform of the kernel:

$$\hat{K}(f) = \mathcal{F}\{K(t)\}$$

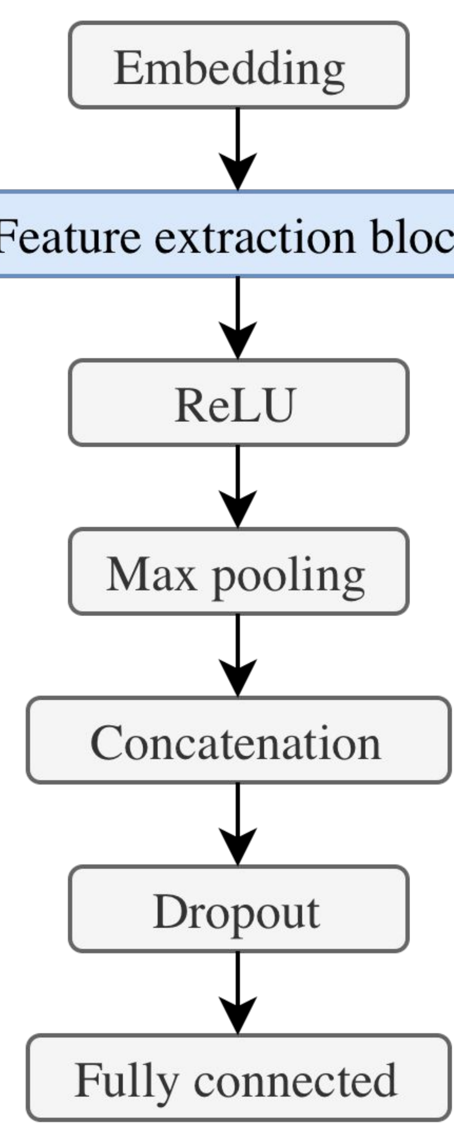
Normalized power spectral density:

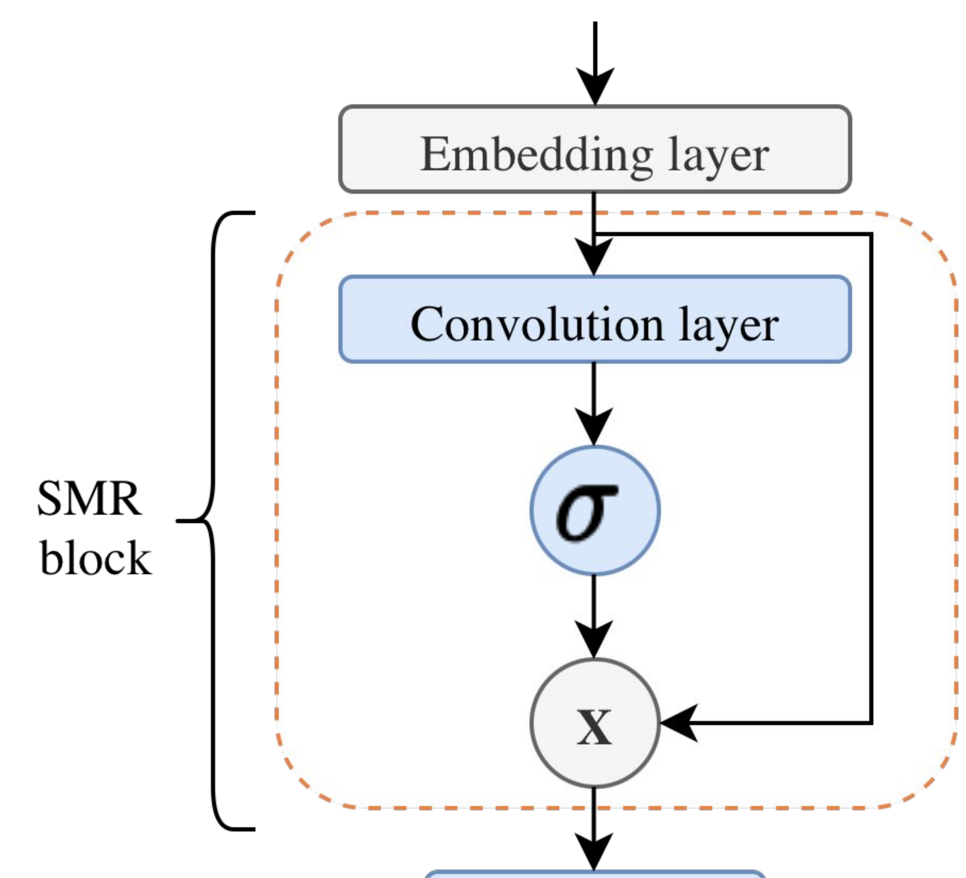
$$P(f) = \frac{|\hat{K}(f)|^2}{\sum_f |\hat{K}(f)|^2}$$

Spectral entropy:

$$H = -\sum_f P(f) \log P(f)$$

ARCHITECTURAL PIPELINE



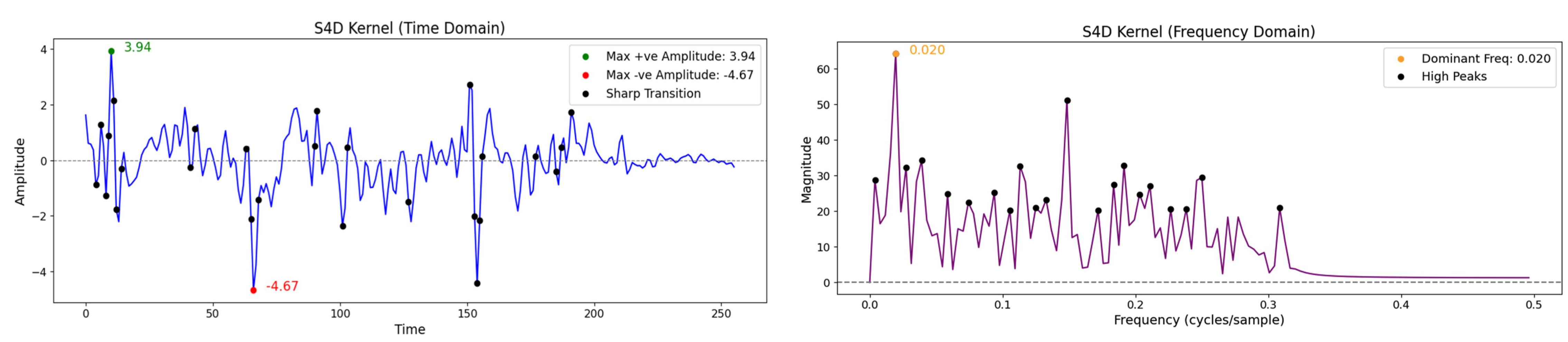


SMR + S4D Feature block:

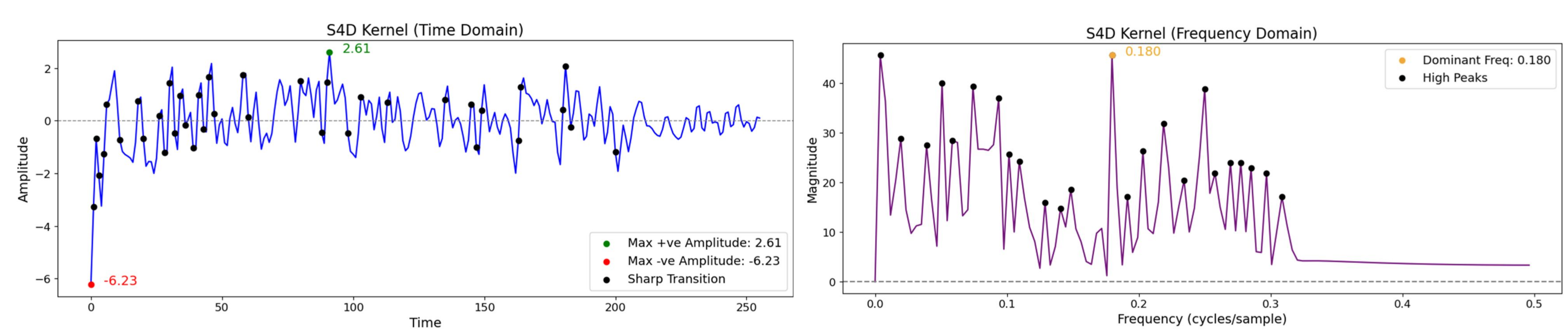
Feature extraction block is the only component that varies across six model architectures.

INTERPRETABILITY ANALYSIS

Time domain and frequency domain impulse responses across architectures:



1) S4D standalone model



2) SMR + S4D model

RESULTS

Model architecture	Acc	F1
Convolution 1D	87.64	87.66
Depth Wise separable conv	87.12	87.57
S4D	87.07	87.02
Depthwise + S4D	88.17	87.17
Conv 1D + S4D	85.05	86.22
SMR + S4D	88.26	88.03

Model	Time domain		Freq domain		
	Max +amp	Max -amp	Dom freq (cycle/s)	Token range	Spectral entropy
S4D	+2.6	-6.2	0.18	5-6	3.94
Depthwise+S4D	+2.7	-3.8	0.086	12	3.97
Conv1D+S4D	+6.7	-5.2	0.14	7	4.21
SMR + S4D	+3.9	-4.7	0.02	50	3.79

REFERENCES

[1] Chakraborty, S., Krishna, R., Ding, Y., and Ray, B. Deep learning based vulnerability detection: Are we there yet?, 2020.

[2] Gu et al. (2022). On the parameterization and initialization of diagonal state space models. NeurIPS.

[3] Qi et al. (2024). SMR: State memory replay for long sequence modeling. ACL Findings.

[4] Verma, S., Anand, A., and Mezini, M. CodeSSM: Towards state space models for code understanding. In The 2025 Conference on Empirical Methods in Natural Language Processing, 2025