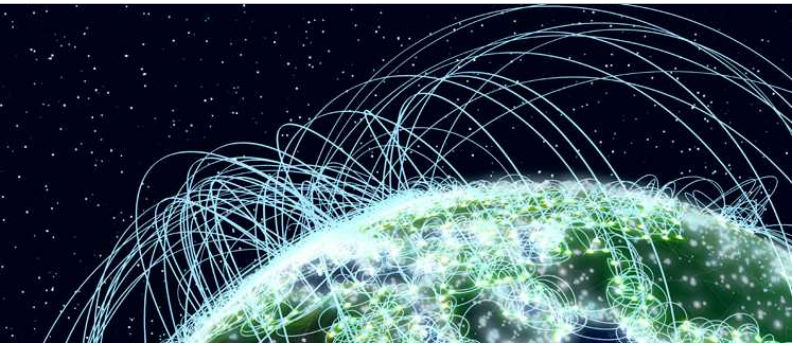
SAND2015-10095C

Exceptional service in the national interest

Sandia National Laboratories



A signal processing approach for cyber data classification with deep neural networks

Jonathan A. Cox, Conrad D. James, James B. Aimone

U.S. DEPARTMENT OF ENERGY

National Nuclear Security Administration

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND NO. 2011-XXXXP

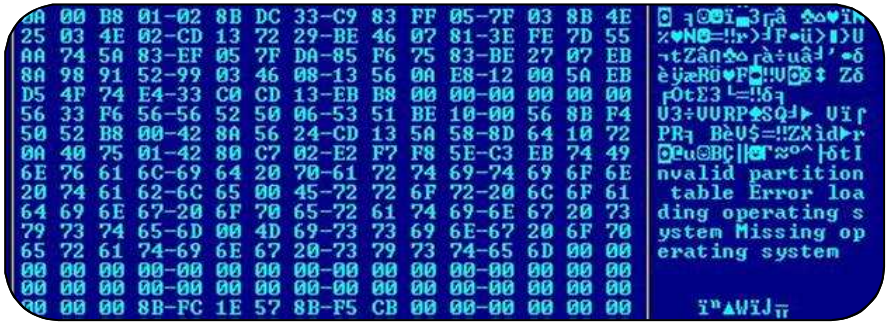
Evolving Cyber Threats

Sandia National Laboratories

- Cyber threats continue to expand
  - Outpacing ability to train and deploy skilled human analysts
  - Network traffic far exceeds capacity for human processing
- Attacks becoming more sophisticated
- Signature-based defenses not adequate
  - Polymorphic code (self-modifying programs)
  - Steganography (hiding data in plain sight)
  - “Zero-day” attacks

The Human Analyst

- Large knowledge of prior threat “signatures”
  - Specific code or knowledge of “bad” websites
- Recognizes previously observed attack methods
  - Man-in-the-middle, “phishing” or Distributed Denial of Service (DDoS)
- Discovers anomalous patterns
  - Temporal
  - “Spatial” (within individual temporal events)
- Can ***infer*** new threats based on evidence
  - e.g. discovered new class of computer worm



Problem Statement and Hypothesis

Sandia National Laboratories

- Analyst often faced with untrusted BLOBs of data
  - What is this data, really?
- Can we learn a model of normal, trusted data?
  - Without dictionaries, signatures, specific features or other prior knowledge?
- Can this model reliably identify unknown data: **distinguish file types?**
- *We propose:* **(1)** A set of transformations for representing data. **(2)** Automated feature discovery and learning with a neural network.

Background

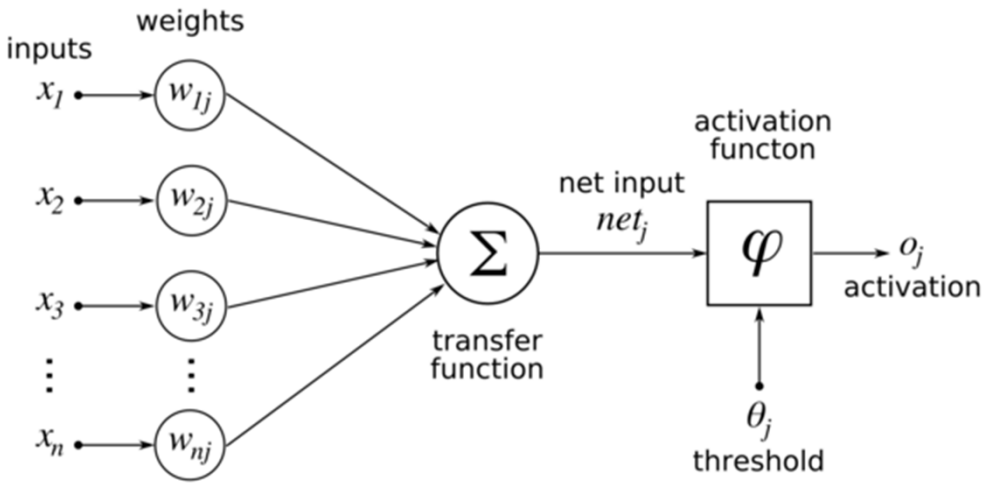
Sandia National Laboratories

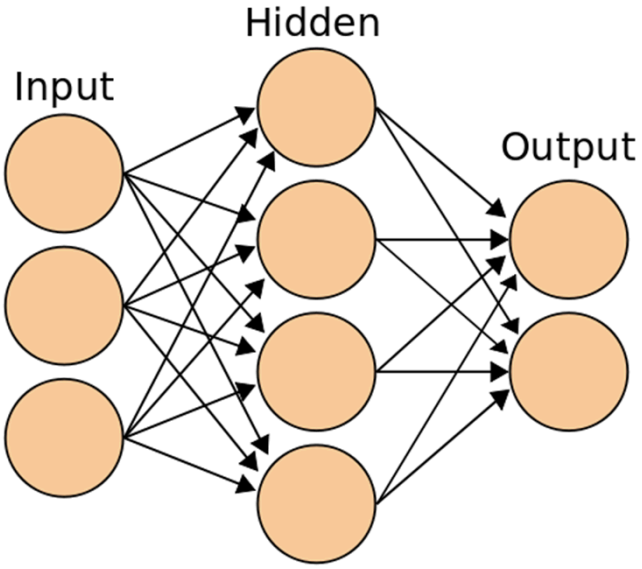
- Predominantly binary classification (good/bad executables)
- Feature sets
  - Entropy within sliding window[1]
    - Short-time Fourier Transform of entropy (proposed)[2]
  - Byte frequency[3]
  - Byte n-grams (e.g. CPU instructions)[4]
- Classifiers
  - Support Vector Machines[1,3,4,5]
  - Decision Trees[5]
  - Naïve Bayes[5]
  - Neural Networks[6]

1. Hall, Gregory A. "Sliding window measurement for file type identification." (2006).
2. Thomas C. Schmidt et al. "Context-adaptive Entropy Analysis as a Lightweight Detector of Zero-day Shellcode Intrusion for Mobiles." ACM WiSec (2011).
3. Zhang, Like, and Gregory B. White. *Parallel and Distributed Processing Symposium, 2007. IPDPS*.
4. Reddy, D. et al. "N-gram analysis for computer virus detection." Journal in Computer Virology 2.3 (2006).
5. J.Z. Kolter et al. "Learning to detect malicious executables in the wild." Proc. of the 10th ACM SIGKDD (2004).
6. Wright, Jason L., and Milos Manic. "Neural Network Approach to Locating Cryptography in Object Code." *ETFA*. 2009.

Neural Networks: Brief Overview

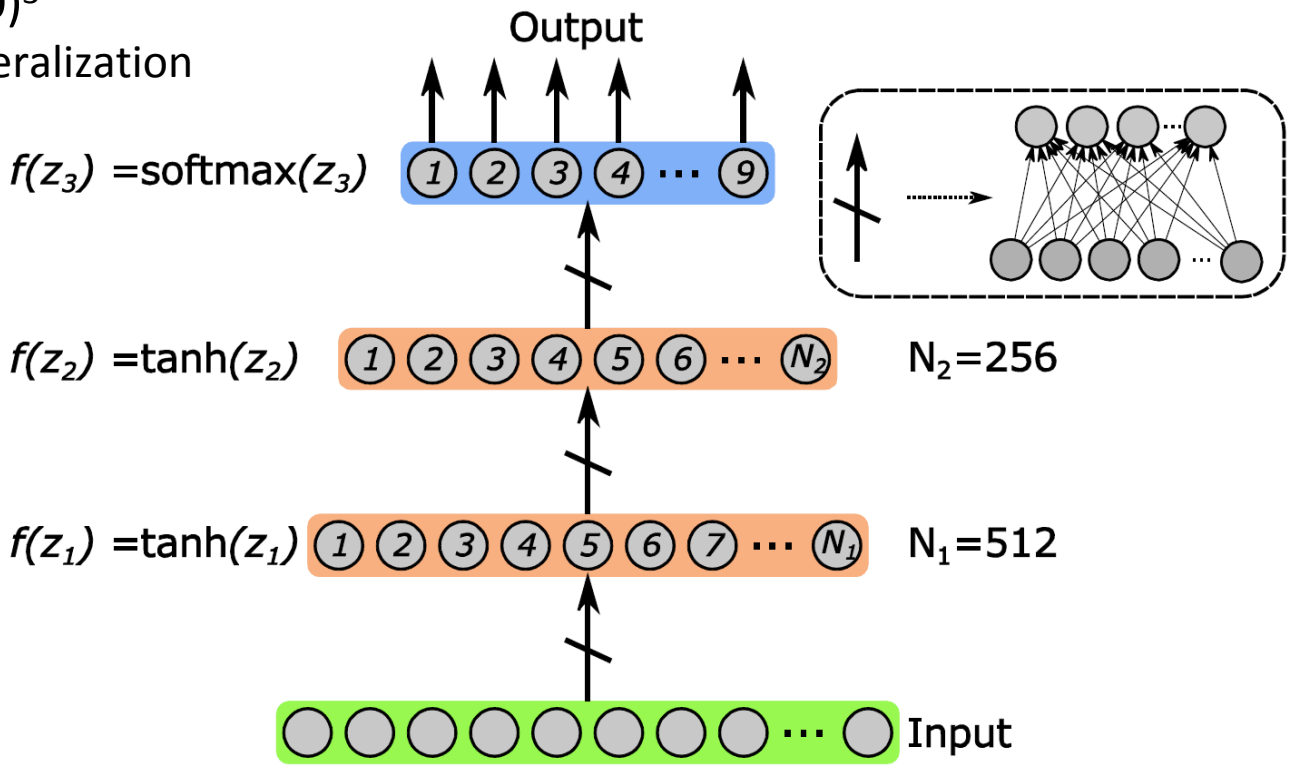
- Well suited for wide range of data (learns features):
  - Images, video, speech and other signals
  - Word or sentiment vectors in a hyper-space
    - e.g. “Germany” is close to “Austria” in the vector space
  - There are many types of data (with more everyday). We do not want to develop features for all of them.





Neural Network Classifier

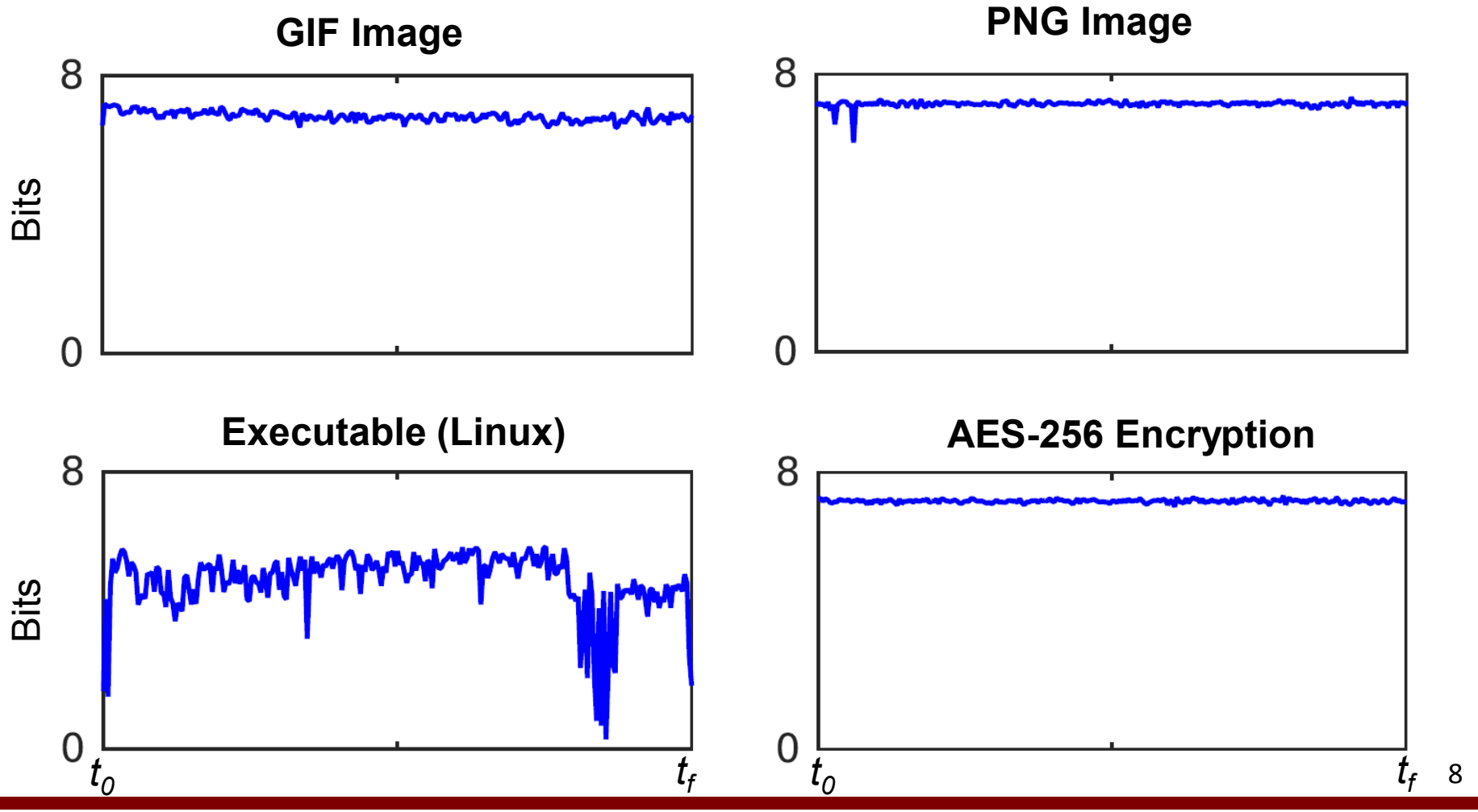
- Pre-training (2006)[1]
  - Initializes network closer to good minimum
- Dropout regularization (2014)[2]
  - Prevents overfitting on deep networks
- De-noising (2010)[3]
  - Improves generalization



1. G.E. Hinton et al. "Reducing the dimensionality of data with neural networks." Science 313.5786 (2006): 504-507.
2. N. Srivastava, et al. "Dropout: A simple way to prevent neural networks from overfitting." J. Machine Learning Research 15.1 (2014).
3. P. Vincent et al. "Stacked denoising autoencoders" J. Machine Learning Research 11 (2010).

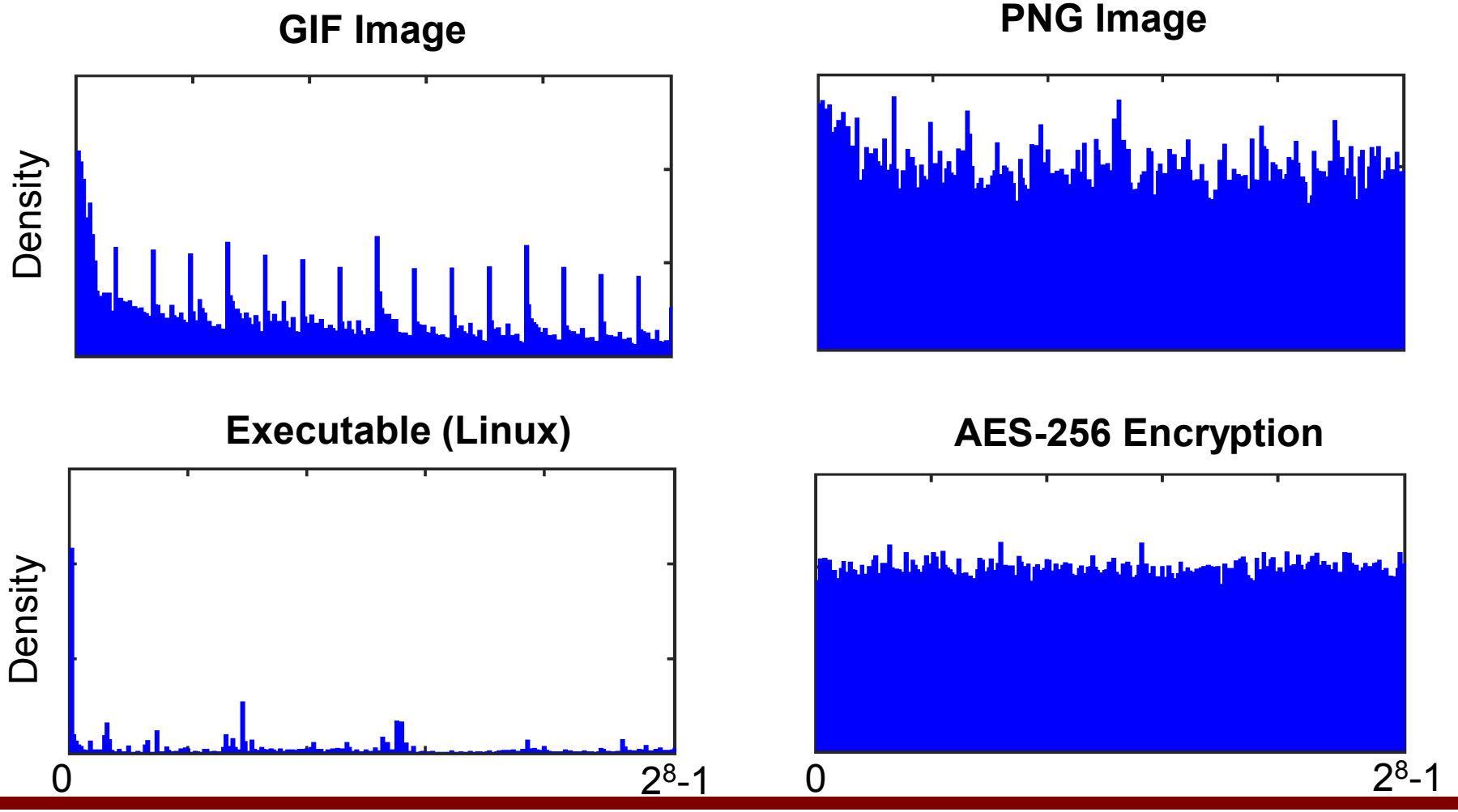
Entropy within Sliding Window

$$H(X,t) = -\sum_{i=0}^{255} P(x_i,t)\log_2 P(x_i,t)$$



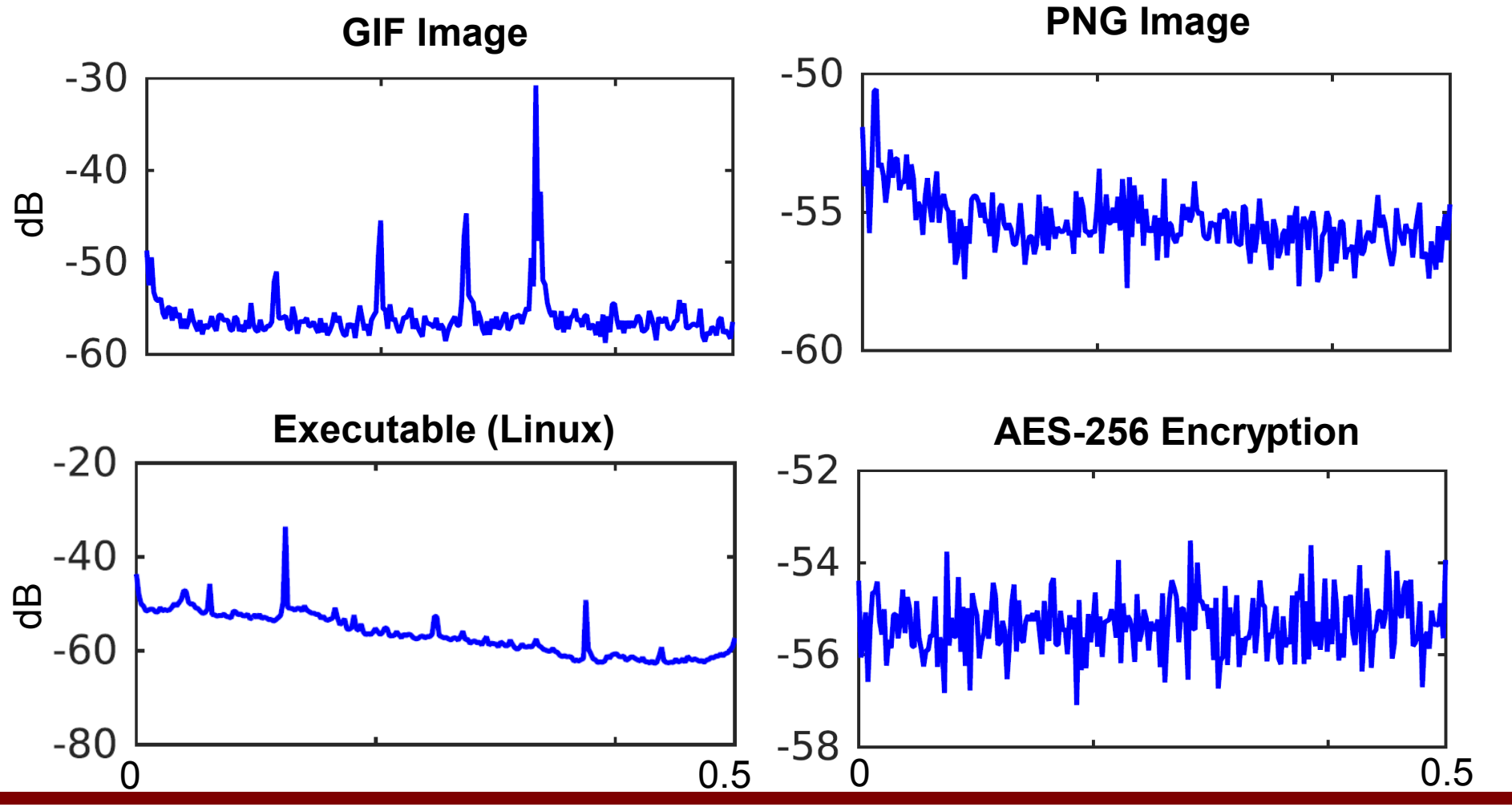
Byte Frequency Distribution

$$P(x_i) = \frac{\text{hist}(x_i)}{N}$$



Power Spectral Density

$$S[\omega] \approx \sum_t \left\{ \left| \sum_n x[n] w[n-t] e^{-j\omega n} \right|^2 \right\}$$



Performance

- Widely varying performance between representations
- Entropy performs worst
  - Many compressed formats look similar
- Byte frequency distribution performs best
  - Reveals common characters, instructions, etc.
- Combining all three yields best performance (97.4%)
- Good performance despite small training set[2]
  - Dropout and denoising help

	P	H	S	P, H	P, S	H, S	P,H,S
Accuracy[1,2,3] (%)	95.30	77.11	81.26	96.74	96.63	94.44	97.44

File types (9): HTML, PNG, JPEG, GIF, PDF, DOC, ELF, GZIP, AES

P: Byte freq. distribution
H: entropy within window
S: power spectrum

1. Average of three trainings.
2. Training set size is 500 per class.
3. Cross-validation set size is 100 per class.

Confusion Matrix

- Six of nine types achieve ≥99%
- AES readily distinguished (95%)
  - Only confused with PNG
- PNG has worst performance (91%)
  - Can be very high entropy or have large amount of metadata

Actual Class \ Predicted Class	HTML	PNG	JPEG	GIF	PDF	DOC	ELF	GZIP	AES
HTML	100								
PNG		91	1		1	1	1	1	4
JPEG		1	99						
GIF				100					
PDF		1	3	1	95				
DOC		1				99			
ELF							100		
GZIP								100	
AES		5							95

Concluding Remarks

- Recent advances improve neural network performance
  - Pre-training, dropout and denoising
- Minimal feature engineering or tuning
- Benefits from variety of input representations
- Effective for identifying unknown data
  - e.g. distinguishes between high-entropy formats

- Future work: automating temporal tasks
  - Recent advances in recurrent neural networks are promising
  - More powerful optimization techniques[1] and pre-training[2]

1. Martens, James, and Ilya Sutskever. "Learning recurrent neural networks with hessian-free optimization." ICML. (2011).
2. Hermans, Michiel, and Benjamin Schrauwen. "Training and analysing deep recurrent neural networks." NIPS. (2013).

Confusion Matrix – Entropy Alone

Sandia National Laboratories

Predicted Class

Actual Class	HTML	PNG	JPEG	GIF	PDF	DOC	ELF	GZIP	AES
HTML	100								
PNG		46	4	2	2	3			**43**
JPEG		5	86	2	1	1		5	
GIF		14	2	37				2	**45**
PDF	3	7	4	2	83		1		
DOC					1	98	1		
ELF						5	95		
GZIP				3				68	**29**
AES									100