CERIAS

the center for education and research in information assurance and security

Printer Security and Forensics

Aravind K. Mikkilineni, Nitin Khanna, Edward J. Delp

Goals

- Use *intrinsic signature* of the printing device to identify as much information as possible from printed document about the device that produced it
- Embed information in printed document at hardware level via *extrinsic signature embedding*
- Intrinsic and extrinsic signatures are based on extraction and modulation of physical characteristics of the printer mechanism
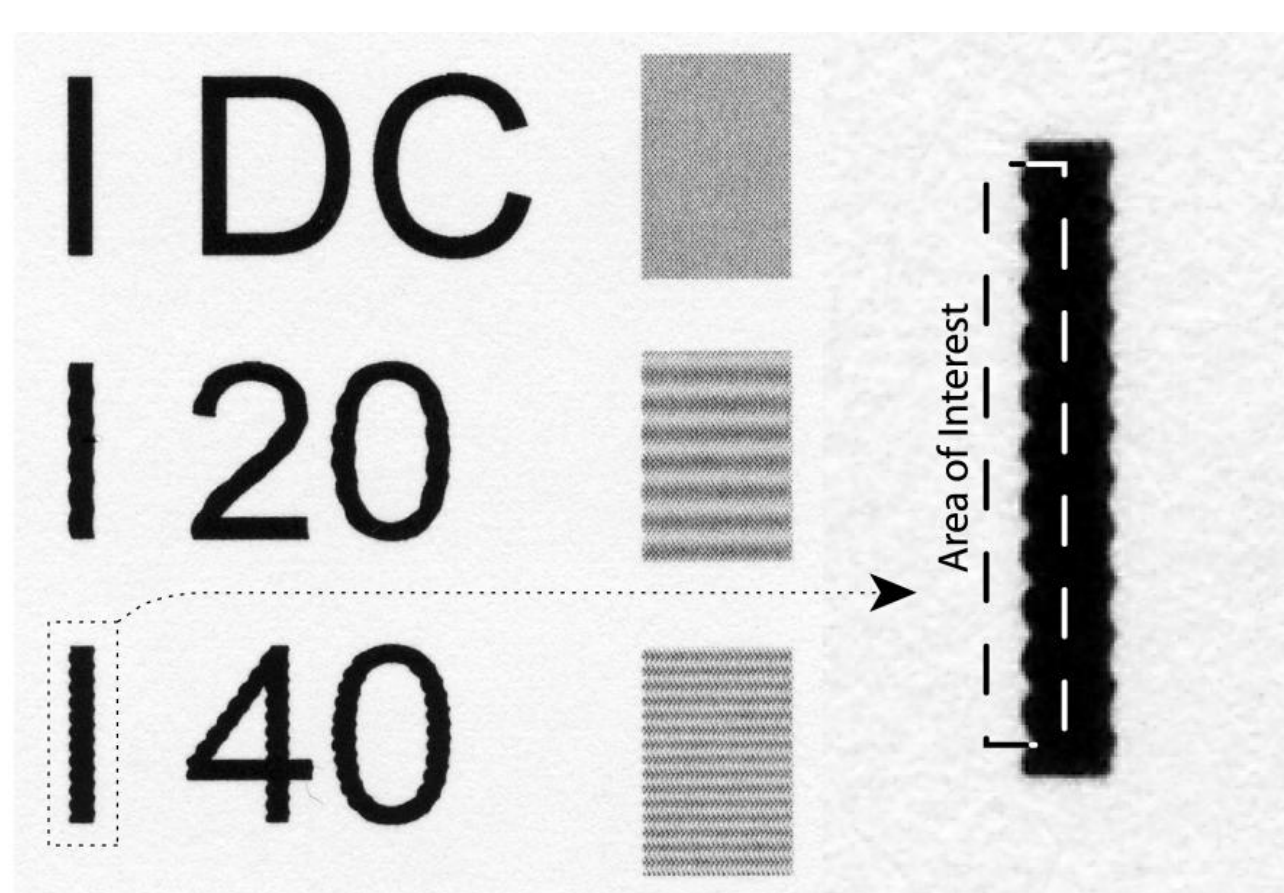
Protect and Prevent

- Copying - scan and print
- Fingerprint and trace the "scan-print" attack
- Authentication
- Forgery/Alterations – additions and deletions

Extrinsic Signatures

- Vary laser power to embed signal in edges of text
- Use DFT detector to decode embedded signals
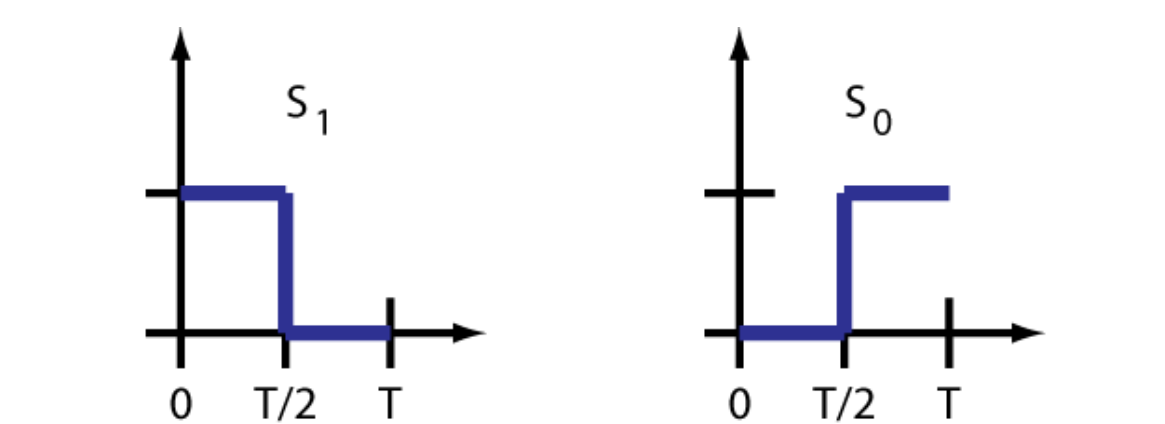- Approximately 8 bits per text line

$$B(y) = \sum_{i=0}^{n} b_i A_i sin(\frac{2\pi f_i y}{R_p})$$

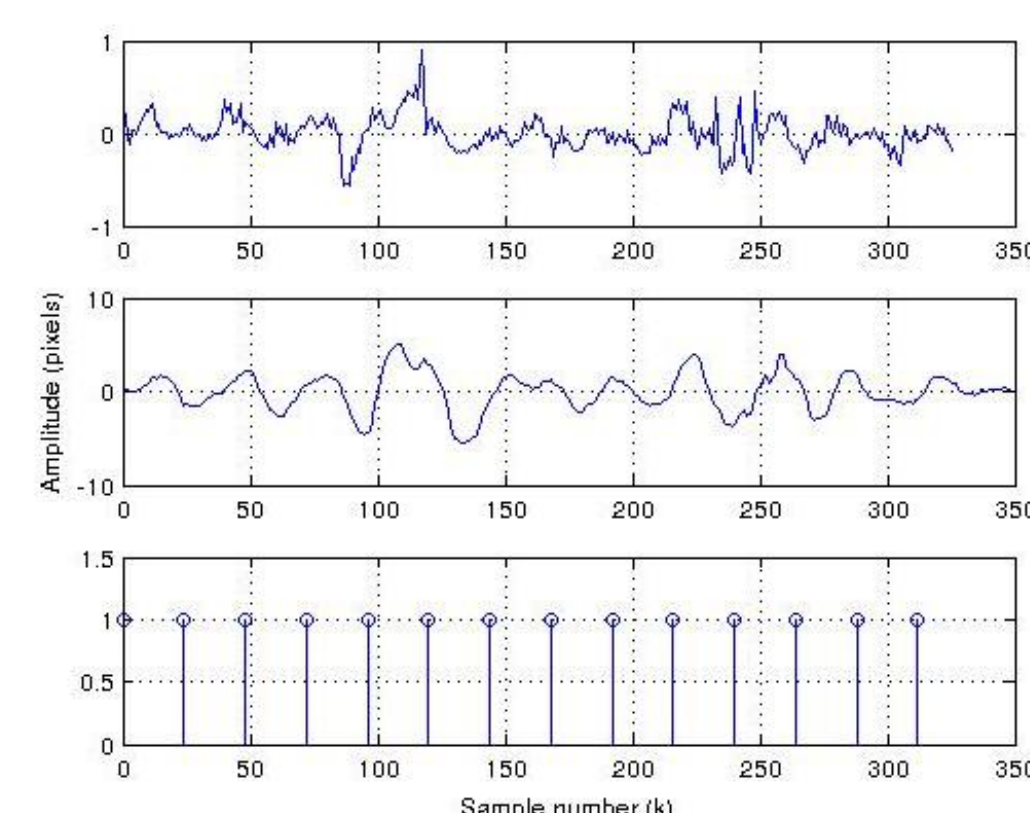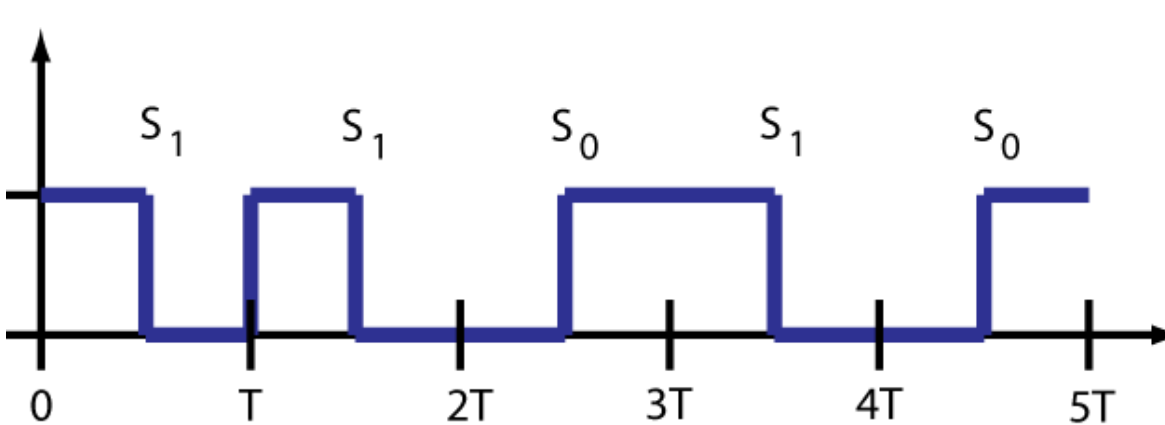$$\mathbf{f} = f_0, f_1, \ldots, f_n,$$

$$A_i = \frac{n-i}{n} A_{max} + \frac{i}{n} A_{min}.$$

- Modification of printer hardware is not always an option
- Perform similar embedding without printer modification
  - Embed in document before sending to printer

$$b_i \epsilon \{0, 1\}$$

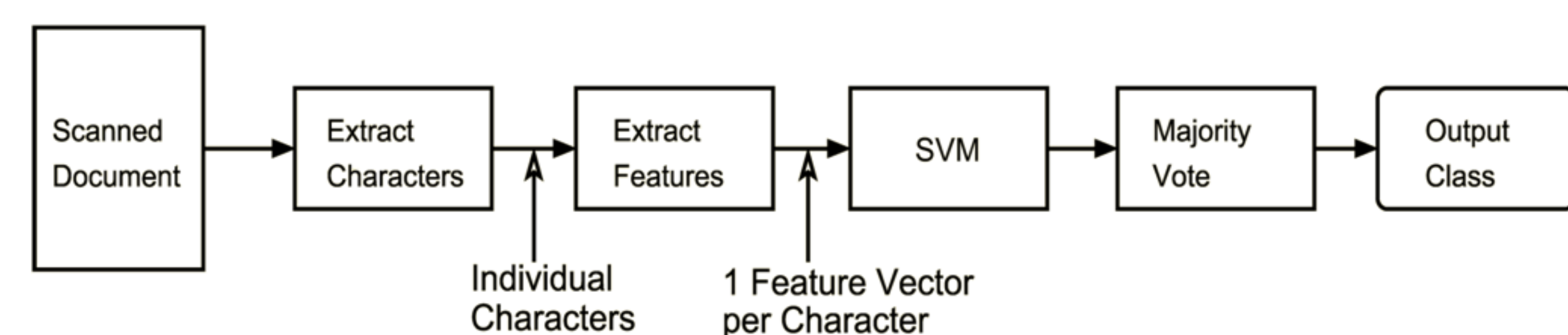$$s[k] = \sum_{j=1}^{14} s_{b_i}[k - iT_p]$$

Marked TIFF - Marked Scan - Non-marked Scan

Extracted signal from edge of text character.

- Use matched filter detector to estimate embedded signal
- Embedded characters: [bBDEFhHkKLmMnNpPrRuU]
- 14 bits/character or approximately 12.7kbits/page for 12 point text

Intrinsic Signatures

Scanned Document → Extract Characters → Extract Features → SVM → Majority Vote → Output Class

Individual Characters

1 Feature Vector per Character

- Graylevel co-occurrence texture features estimated from printed regions within individual text characters
- System works across various font types and sizes, paper types, and consumables age when trained with same font and paper type
- 90% classification accuracy when training on new data and testing on old data
- 100% classification accuracy among ten printers using 22 features and majority voting by over 300 'e's printed with 12 pt. Times Roman font
- For each of the testing documents more than 250 (out of 300) 'e's were classified correctly
- Including banding features in the feature set increases robustness of the intrinsic signature
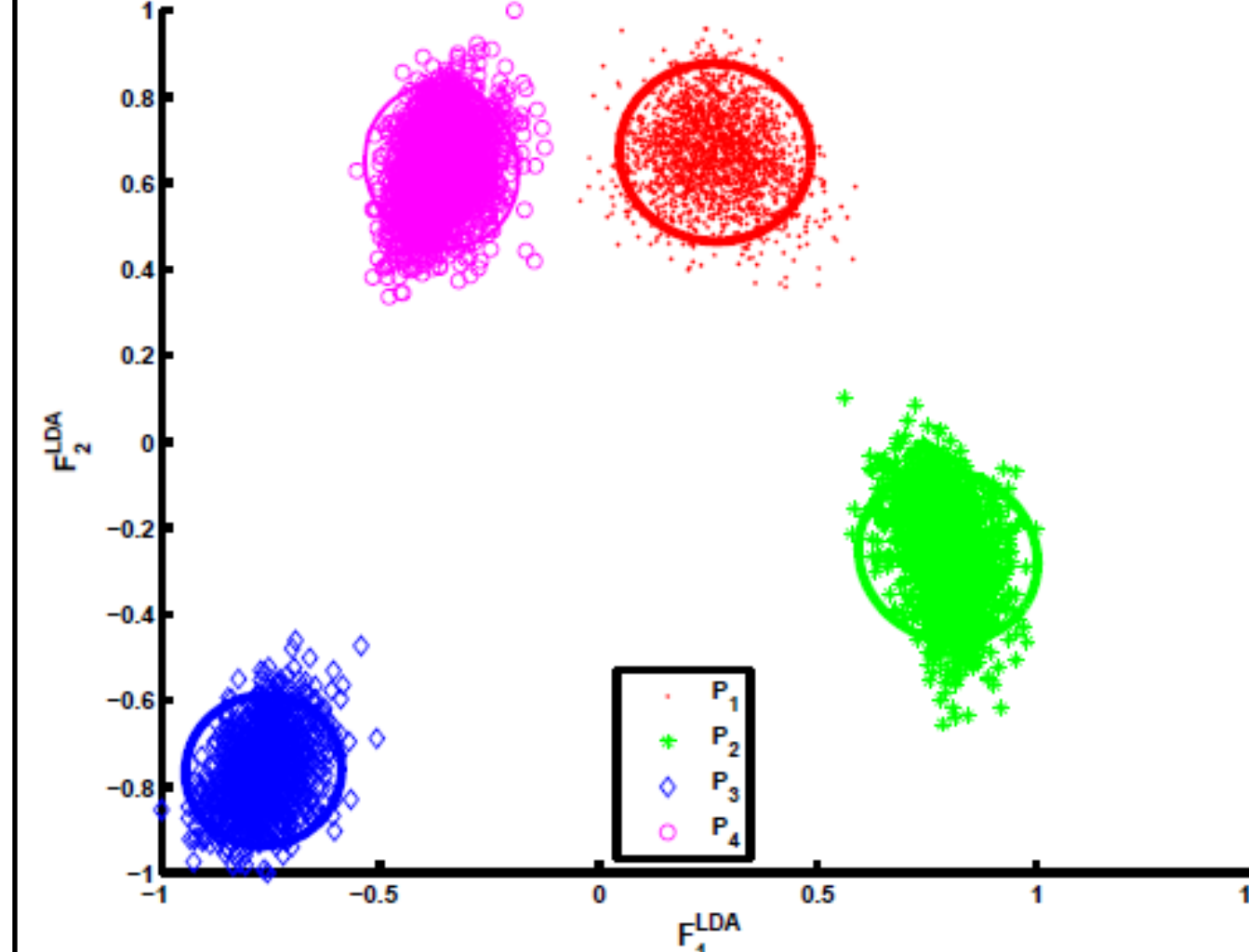
(a) LDA on GLCM Features
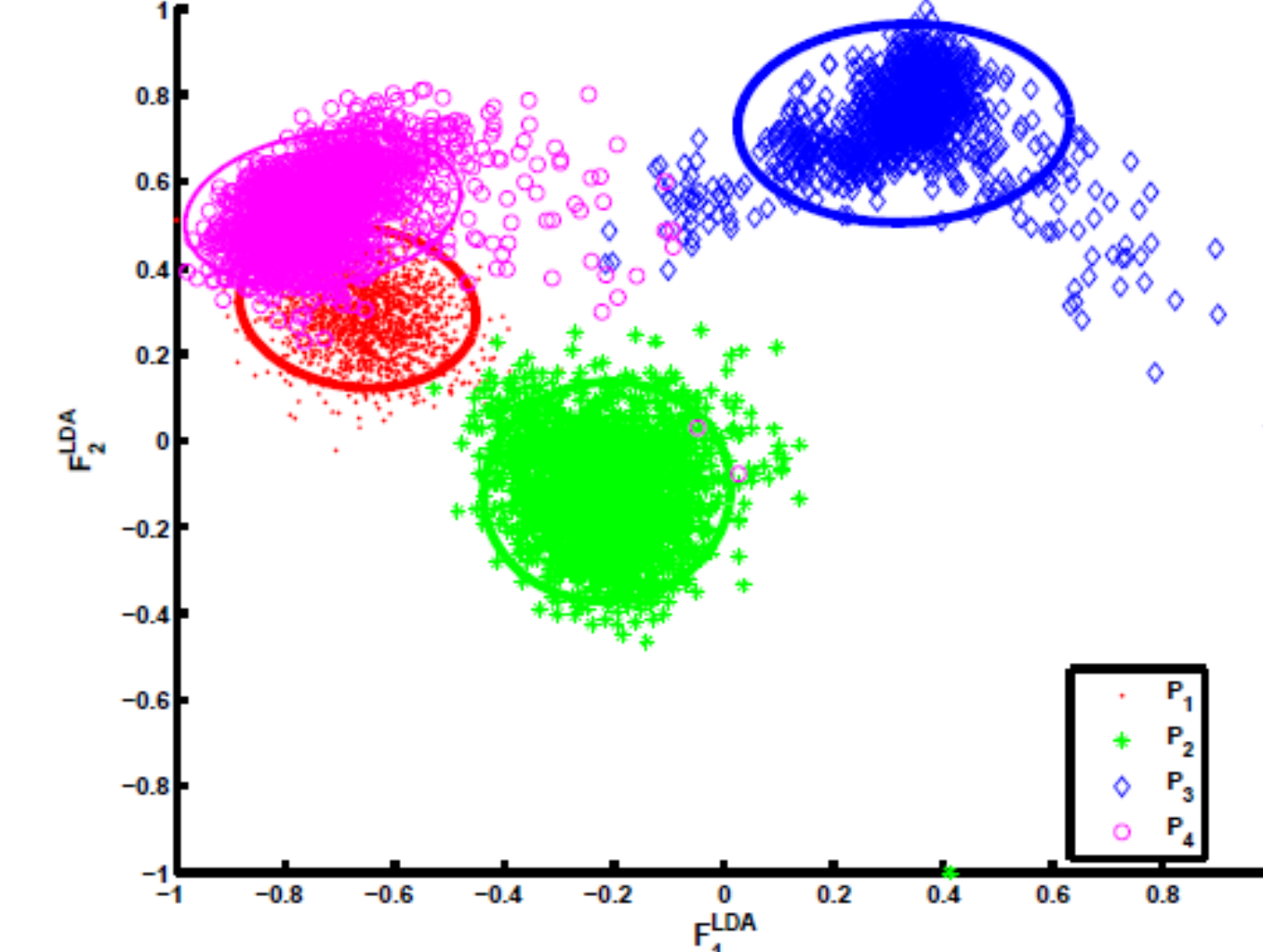
(b) LDA on DFT Features

- Intrinsic signature for printer identification from text documents performs well under several attack models
- Effectiveness of this system starts to break down only when the perceptual quality of the text is greatly affected
- These features are scalable to a larger number of printers using a distance based classifier, as was shown in the results using the reduced feature sets

PURDUE UNIVERSITY

CERIAS

Discovery Park
e-Enterprise Center