Using the Forensic Case Generator in a Group-based CSI

MICHAEL JONES

The Assignment Context

- Computing Framework with 4 pathways
  - One of which is Forensic Computing & Security (FCS)
- Common year 2 unit – theme: 'integration'
  - Common element: project management
  - Pathway-specific element: chosen by pathway leader
- FCS specific element: Group-based CSI
- Organisation:
  - 2 weeks full-time at end of second term
  - Group size – 5 or 6

Pathway Specific Element

- Tools available: FTK, EnCase, …
- Activities
  - Seizure *
  - Capture
  - Investigation
  - Report writing (witness, expert witness)
  - Simulated court event

  - * in week prior

Aims of the Assignment

- To make the students think
  - What are we looking for?
  - Where could it be (hidden)?
  - What does it mean?
- To provide challenges for all students
- To provide limited guidance, assistance
- To provide lots of routine activities

The Cases

- Cases for each group are unique, but share a common theme
- (possibly) the same crime
  - In this case: terrorist plot linked to the Olympics
    - Each case linked to different events in London
- Similarity means:
  - Students can collaborate
  - Collusion, copying are more difficult
  - Commissioning?

The Data

- Contacts
  - Names, mobile phone numbers
- Locations
  - Meeting and target
  - Actual and triangulated
- Times and dates
  - Rehearsal, warning, target (zero)
- ‘labelling’

Data Sources

- Examples:
  - String
  - ‘regular’ expression
  - File
  - Directory
- Examples:
  - 07[0-9]{9}
  - $namesDirectory?column=firstName

GPS Data Manipulation

- Options
  - None
  - Triangulation
  - Translation
- Triangulation
  - Four points generated from the target location
    - Endpoints of two intersecting lines
- Translation
  - Location ‘moved’ a short distance

Data ‘Labels’

- ‘label’ (attribute) associated with each data item
- Example:
  - Data: 07123456789
  - Label: Contact 1 of 3 Mobile
- Labelling options:
  - ‘of N’ included/omitted
  - Label included/omitted
  - Clues to omitted labels included/omitted

Encoding

- Plaintext can be easily found using tools
- Example encodings used: base64, hexadecimal
  - Can be combined
  - Can be applied to: label, value, ‘all’
  - Example rule:
    - base64|hex=30-50% & base64+hex= & loops=1-2

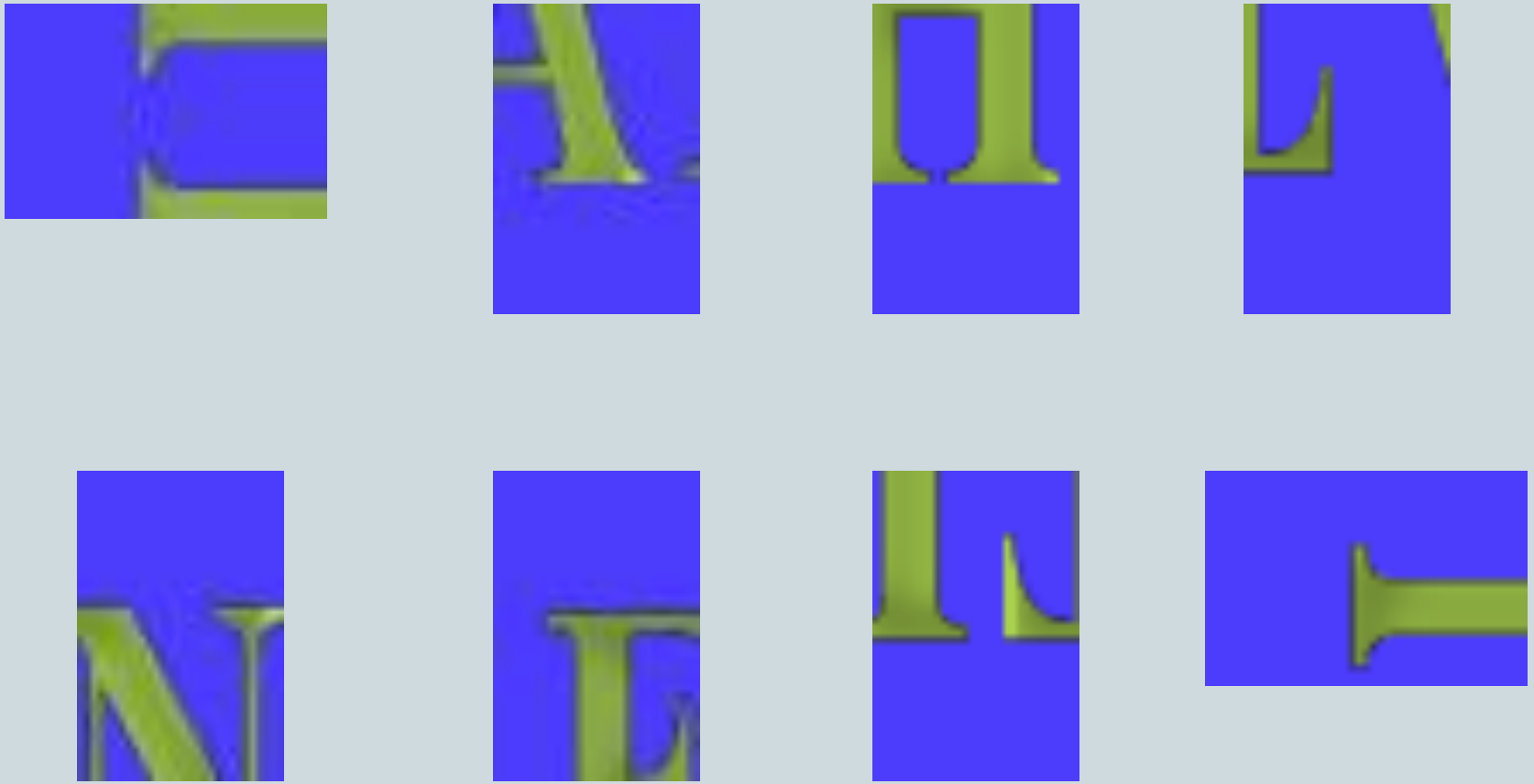
Steganography

- 20+locations associated with files
- Examples:
  - Metadata
  - Body
    - E.g., LSB, HTML/XML comments
    - E.g., as sounds
    - E.g., across multiple files
  - Appended
  - Generated file(s)

Groups

- Multiple file systems produced
  - 4 devices involved per group
  - Mobile phone, SD cards, memory sticks
- Many file types involved
- ‘Graded’ steganography
  - JPG metadata
  - Simple ‘non-tools’ example: image tiling
  - ‘Programming’ example:
    - Embedding single characters in HTML tags within a thousand+ files

Image Tiles



Transcript

Attribute: Driver 1 of 1
    value: Disco Stu
    encoding regime
    encoded
        all: Driver 1|Disco Stu
        attribute: Driver 1
        value: Disco Stu
    data embedded
        all: <!-- Driver 1|Disco Stu -->
        attribute: Driver 1
        file: 03/Hermione/35_Mathilda.jpg
        file type: docx
        location: xml file body
        original file: Generated Latin (0017) docx

The Experience

- Each case:
  - 80+ data items to locate
  - Around 20 techniques involved
  - ‘of N’ omitted from labels
- Some techniques employed had not been covered before
  - E.g., image ‘tiles’
  - E.g., finding terror organisation

The Terror Organisation

- Organisation name encoded
- Embedded in JPG metadata
  - In one image in a sequence of photographs
- File appended to another JPG in the sequence
- Bytes modified at the ‘join point’
- Modified file placed in directory along with others from the image sequence

The Results

- Data Retrieval
  - Each group found (almost) all the data
- Report Writing
  - (reasonably) thorough and well-written
- Crime Inferring
  - Moderate
- Reflection on Experience
  - Fair

Errors in Inferring Crimes

- Example 1:
  - Found: location, date, time
    - Lord's cricket ground August 2012
  - Conclusion: no association with Olympics
- Example 2:
  - Incorrect GPS triangulation -> near Florence, Italy
  - Conclusion: 'Mafia' involvement
- Example 3:
  - Conclusion: no crime and no suggestion of other lines of enquiry or other data sources that might be consulted

Conclusions

- Good student engagement
  - Throughout the two weeks
  - Throughout the groups
- Effective collaboration
  - Techniques communicated to other groups
- Limitations in data analysis highlighted

Commentary

- ‘Shallow’ problem solving
  - Limited dependencies
  - Multiple techniques involved
  - Abstraction may be more challenging
- Development: linking to other databases
  - Example: stealing cars to order
    - Additional database: ‘DVLA’ database