Decoding Identifying Printer Information

Seth Schoen
Staff Technologist
Electronic Frontier Foundation

Chaos Communication Camp
Luftfahrtmuseum Finowfurt

August 7, 2007

Introduction

- EFF is a 28-person member-supported non-profit based in San Francisco, age 16
- We advocate individual freedom in tech., often via impact litigation (“test cases”) on free speech, privacy, copyright, etc.
- We're also interested in how technology itself works – “architecture is politics”
- Color laser printers and photocopiers are designed to track their users; what can we find out?

Tracking and forensics

- As privacy advocates, we wish that artifacts were less traceable and that users had better options for anonymity
- We oppose the decision of firms to make communications media more traceable; we'd like technologies like those I will describe here to be eliminated
- It's important to note that this would not guarantee absolutely impenetrable anonymity in every scenario, since there are many other forensic methods

Forensics and disclosure

- As advocates of full disclosure and open publication, we've tried to investigate and disclose how tracking technologies work
- We want the public to be well-informed about what's possible, and also to rebut the common claim/intuition that tracking could only be done by law enforcement
- We hope to expand the open literature about this and similar technology

Yellow dots

- Almost all color laser printers and color photocopiers ever made embed patterns of small yellow dots for tracking on every output page printed in color mode
- This is certainly not the only way to get forensic information about printers...
- but it's unusual because the dots are *intentionally added* for tracking, and easy to see without special equipment

Where do they come from? (1)

- The United States Secret Service may have privately negotiated this practice with manufacturers as early as the 1980s
- Statements about the nature of this relationship from both government and manufacturers remain circumspect
- Our Freedom of Information Act request to USSS (2005) is mired in bureaucracy
- Motive is most often described as counterfeit deterrence

Where do they come from? (2)

- Note in Xerox documentation:

  “Das digitale Farbdrucksystem DocuColor 5252 ist entsprechend der Forderung **zahlreicher Regierungen** mit einem fälschungssicheren Kennzeichnungs- und Banknotenerkennungssystem ausgerüstet.”

- “Zahlreicher Regierungen”?? (=CBCDG?)
- Printer/copier import restriction threats?

Which devices? When?

- Almost all color laser printers and color laser photocopiers; known exceptions at <http://www.eff.org/Privacy/printers/list.php>
- **Not** color inkjet printers
- Usually **not** on pages printed in B/W
- It is possible that printers that don't print yellow dots nonetheless print some other kind of tracking information not known to the public (or can be identified by other means)

What information is coded?

- Printer/copier serial number
  - often corresponding to user-visible serial number on device chassis
- Date and time of printing
  - only for devices that have this information (most often high-end devices)
- Device manufacturer/model is possibly coded or can be inferred
- Some data bits remain unidentified

Means of viewing dots (1)

- Dots are yellow and repeated across entire page, in grid or staggered grid
- Repeated unit is small enough to allow for multiple repetitions on any currency-sized rectangle (and likely machine-readable)
- Yellow has very low visual contrast against white for the human eye
  - because our Sun is a yellow star?
- Of C, M, Y, K inks, Y (yellow) is visually closest to white background

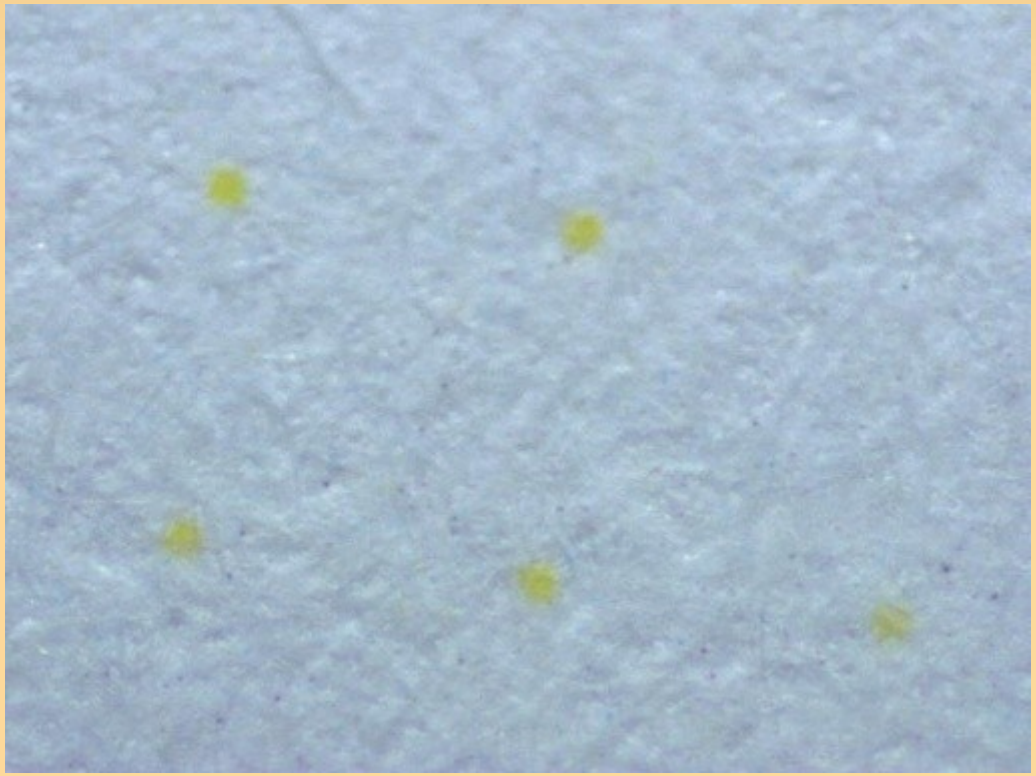
Special equipment not needed

- For example, either of these is enough:
  - DigitalBlue toy computer microscope
  - Pocket LED flashlight

Images from EFF and manufacturer websites

Means of viewing dots (2)

- Simple optical magnification (10x – 100x) by microscope, hand lens, or camera lens



Magnification by toy computer microscope

Means of viewing dots (3)

- Increasing contrast by illumination with blue light, e.g. by a blue LED flashlight, typically makes patterns visible to the naked eye

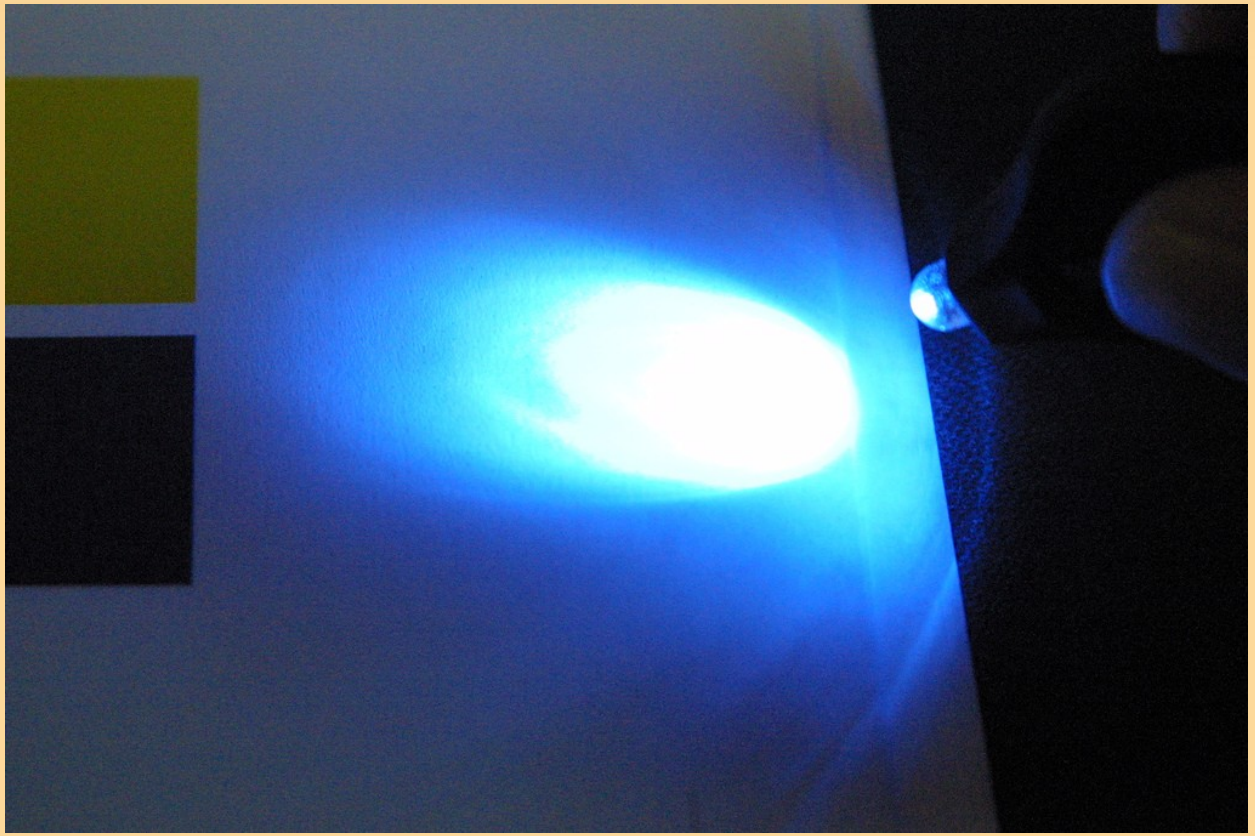
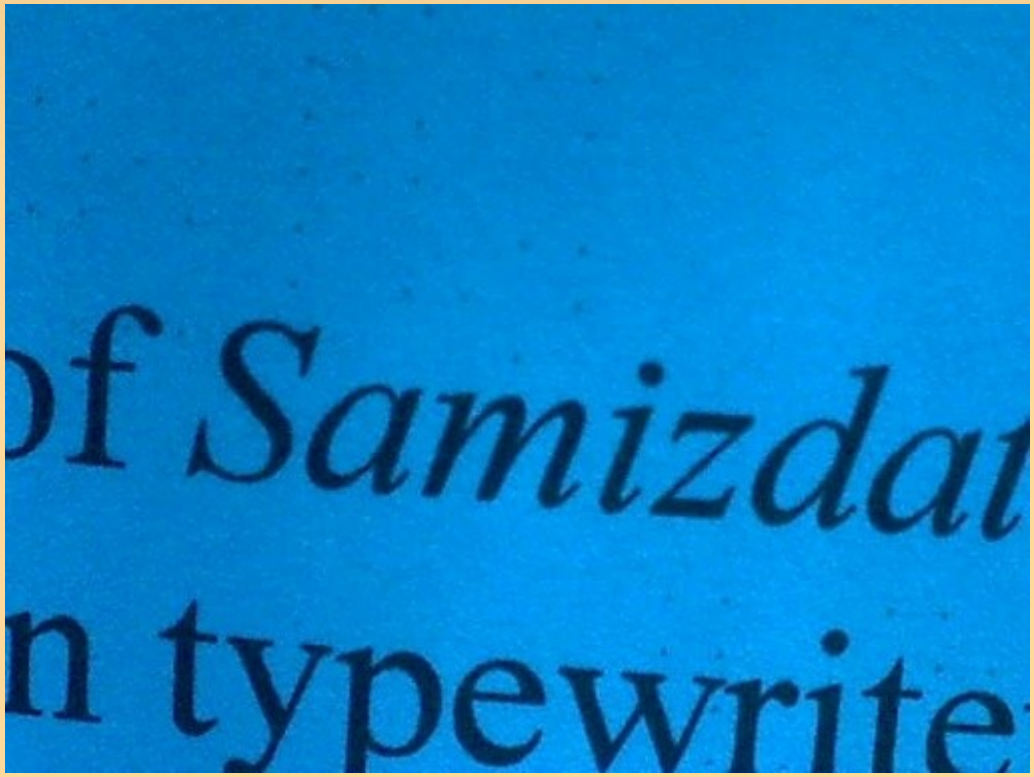


photo:

Quinn Norton

Means of viewing dots (4)

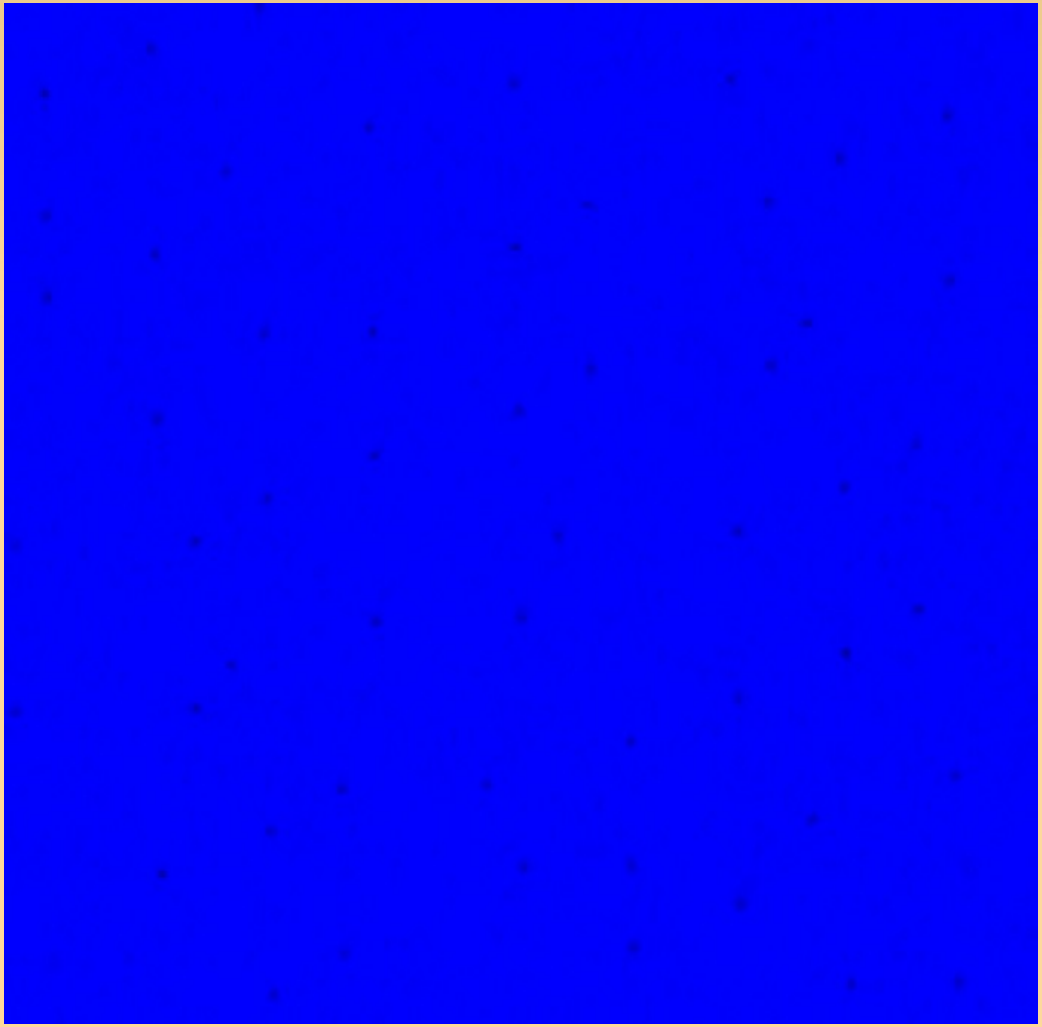
- Blue illumination *and* magnification



Left: text + dots; right: dots only

Means of viewing dots (5)

- Conventional color flatbed scanning (24 bit depth, 600 dpi) with image processing
  - select only blue channel in image
- Modifying scanner is not necessary, because it already has a blue light



Selecting blue channel

- I always forget how to do this, so...
- In GIMP:
  - In Layers/Channels/Paths window's channels tab, deselect Red and Green channels
- In ImageMagick:

```
convert -channel RG -fx 0 scan.tiff blue.png
```

- You can do some amazing things with ImageMagick and a flatbed scanner!
  - Though PIL is a lot faster for this

Mapping to grayscale

- Alternative: select blue channel and convert to grayscale (map blue channel's value to intensity)

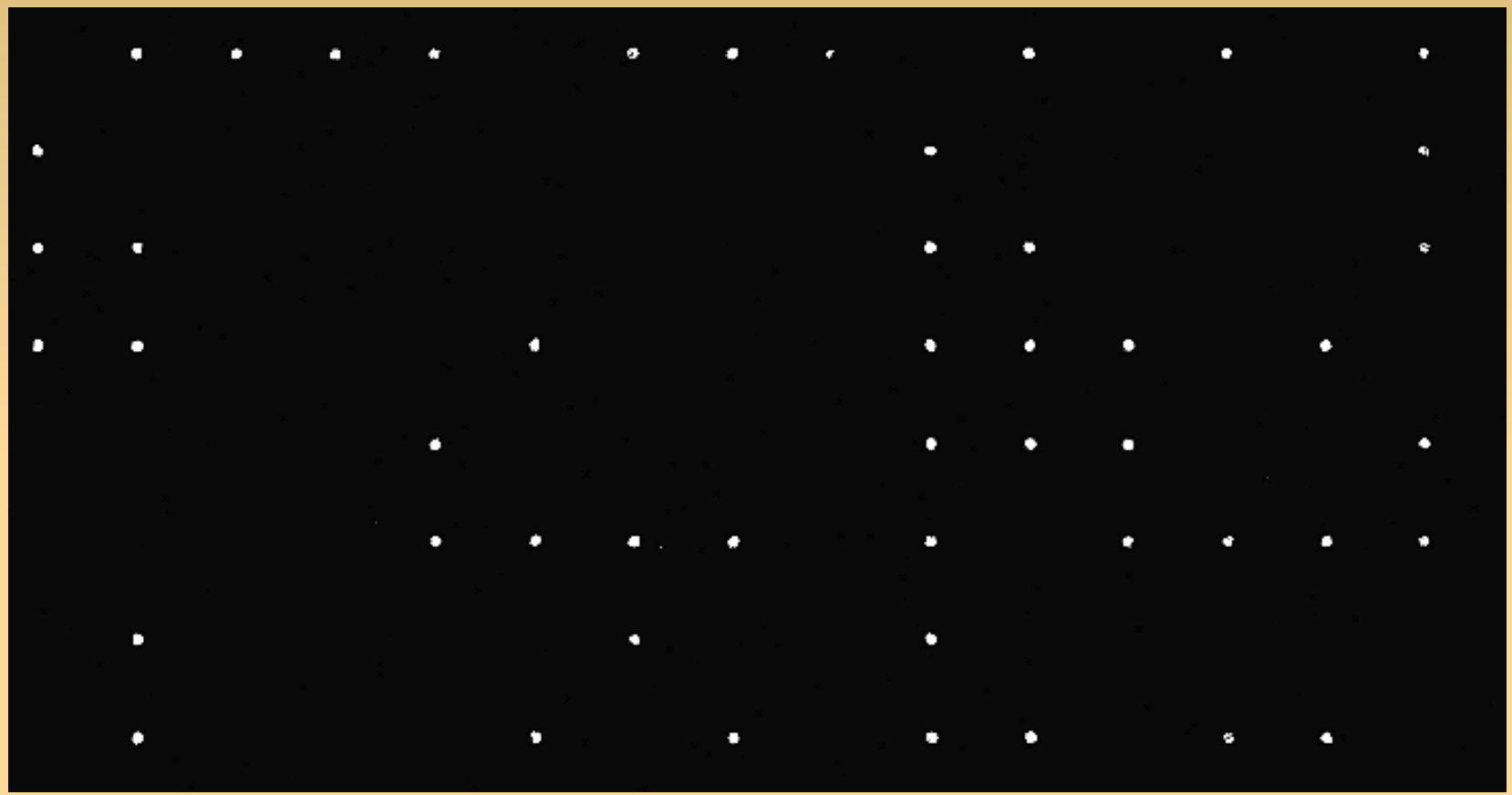
- ImageMagick:

```
convert -fx b
```

- PIL:

```
Image.open("img.tif").split()[2].show()
```

Better contrast



- PIL:

```
blue = Image.open("img.tif").split()[2]
blue.point(lambda x: (256-x)**2).show()
```

- http://www.pythonware.com/library/pil/

How many codes are there? (1)

- Different mfrs. use different codes
- We can often visually distinguish the output of different printers even without breaking the code
  - Xerox and Dell printers have staggered rectangular grids (Xerox in horizontal orientation, Dell in vertical)
  - Canon is most chaotic, seems to form diagonal bands
  - Konica/Minolta has staggered grid; dot alignments matter

How many codes are there? (2)

- The codes fall into certain families
- One hypothesis is that the codes are actually designed or implemented, not by individual printer manufacturers, but by manufacturers of imaging subassemblies
- For example, the Dell code is a 90° rotation of the Xerox DocuColor code

How to read the codes

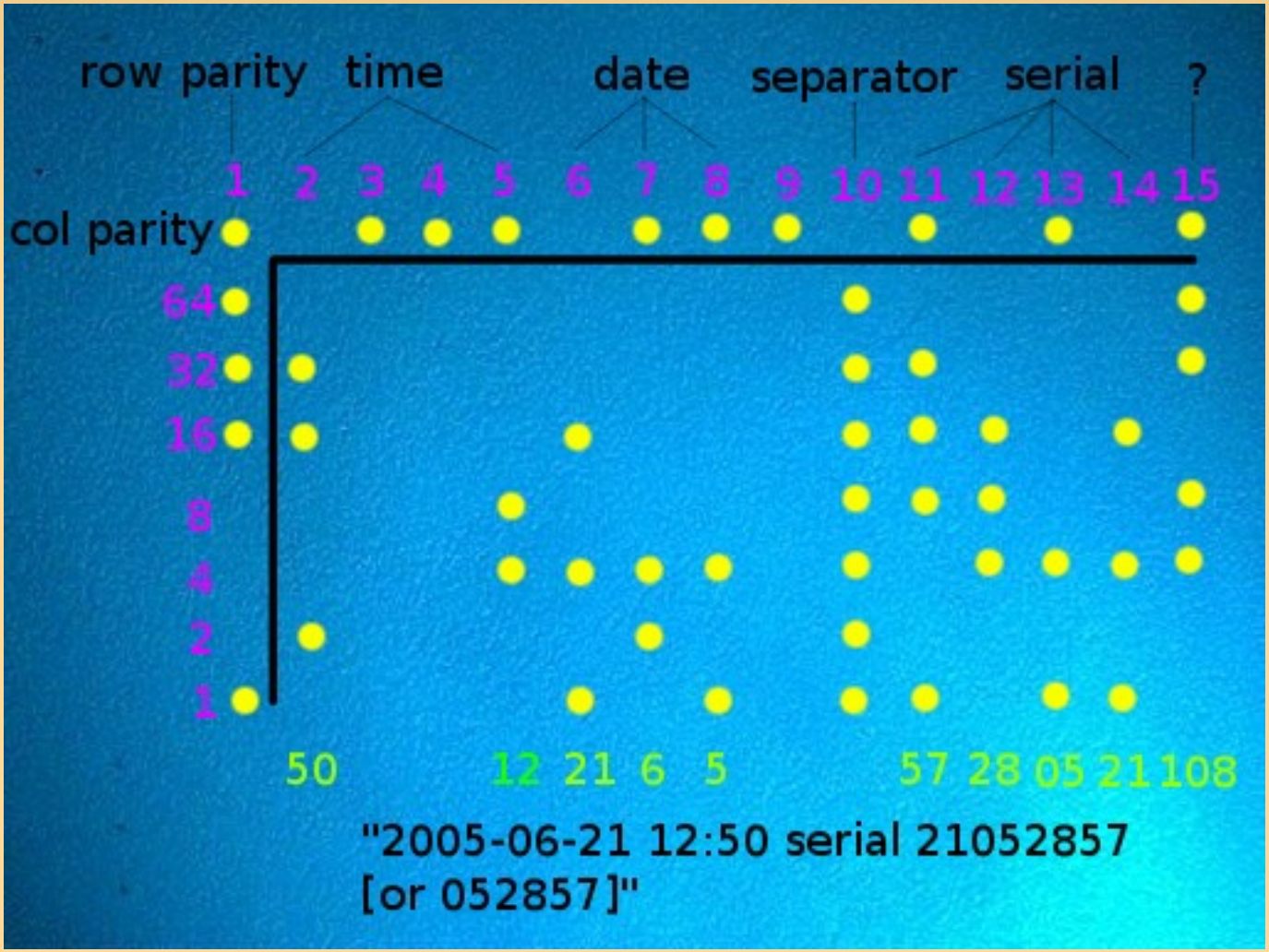
- What we know so far:
- Xerox DocuColor/Dell Color Laser
  - Broken (thanks to Joel Alwen, Patrick Murphy)
- Epson / Konica/Minolta
  - Partly broken (thanks to “P”)
- HP Color LaserJET
  - Structure analyzed
- Several other codes remain unanalyzed, but can be visually distinguished

Xerox DocuColor, Dell Color Laser

- Explained at http://www.eff.org/Privacy/printers/docucolor
- 15x8 rectangular grid
- = 14 7-bit data bytes plus odd row and column parity checks for error correction
- 3-4 bytes unused, 1 byte unknown
- 4 bytes user-visible device serial number
- 1 byte each year/month/day/hour/minute

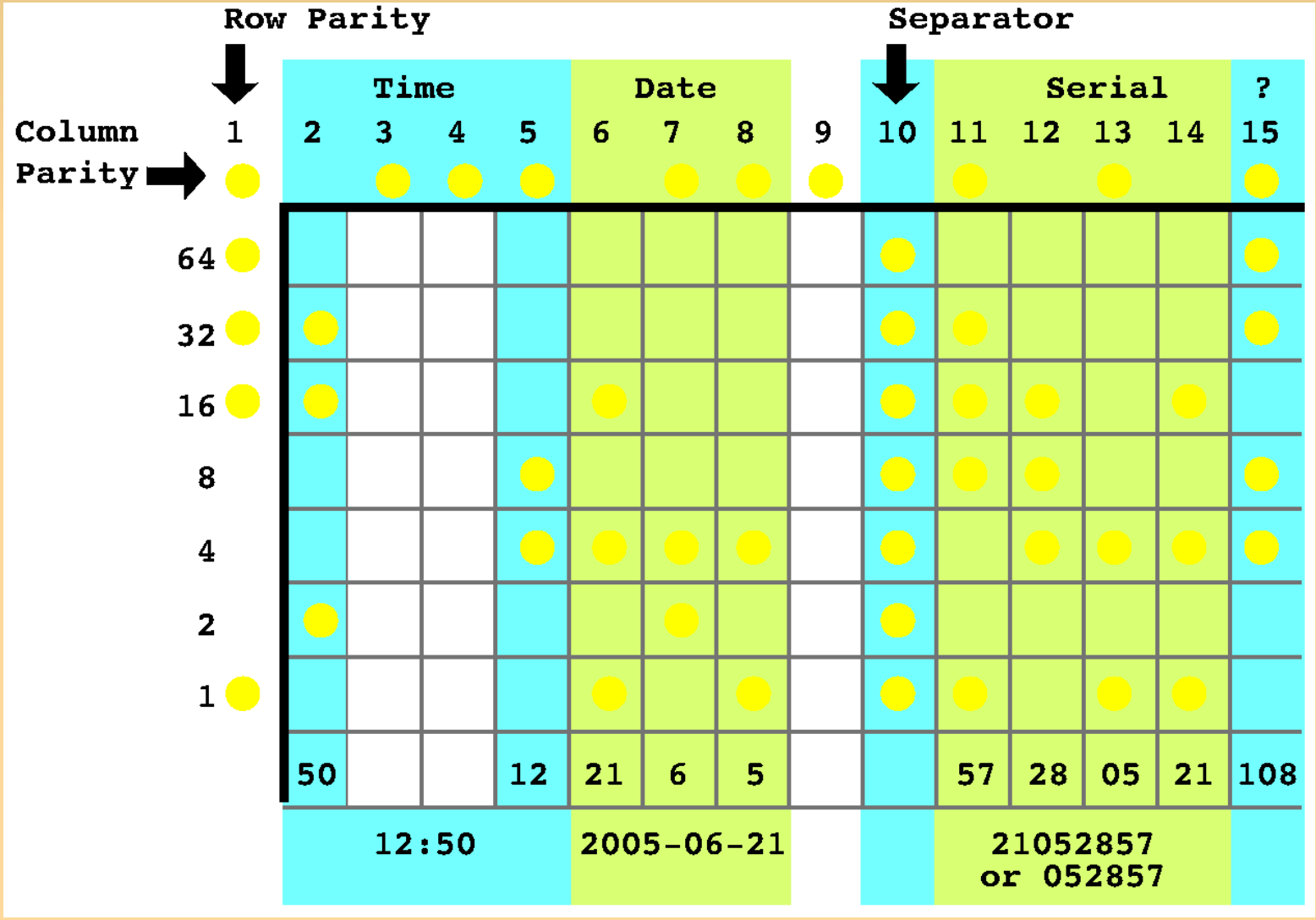
Example

Annotated microscope photograph
(false color, dots not to scale)



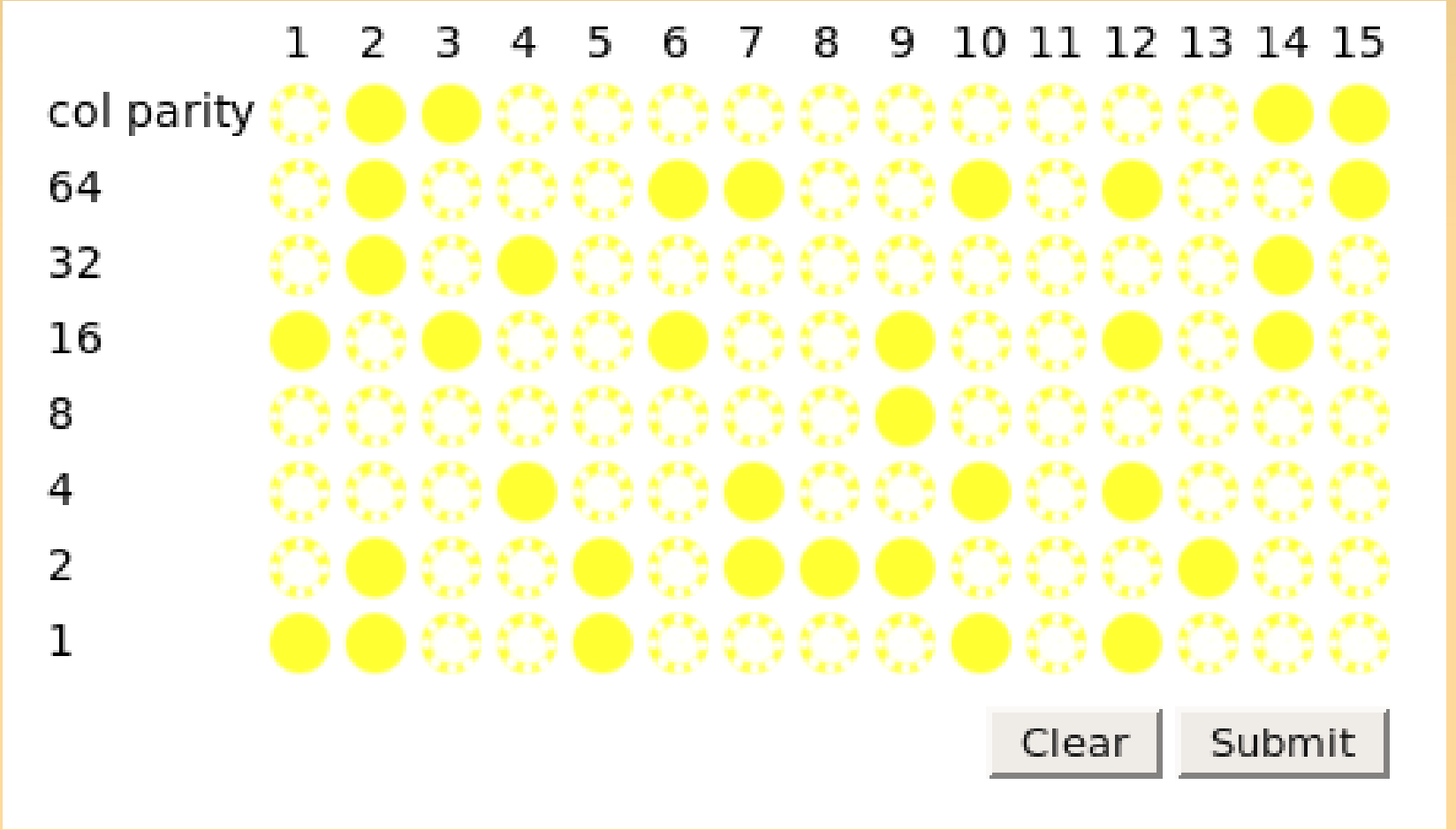
Clearer diagram (on paper)

Redrawn from scratch by Hugh D'Andrade



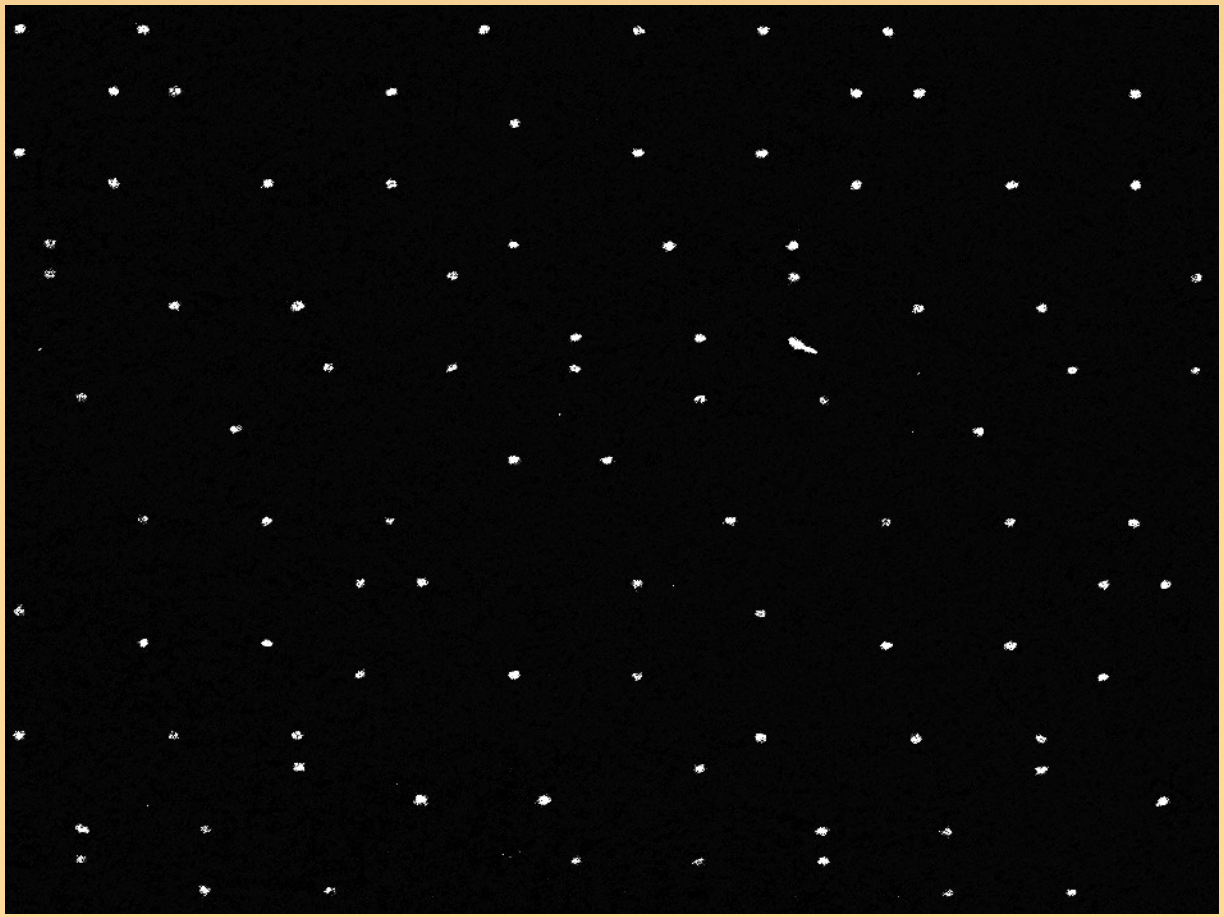
Web-based decoder application

Implements this interpretation on-line



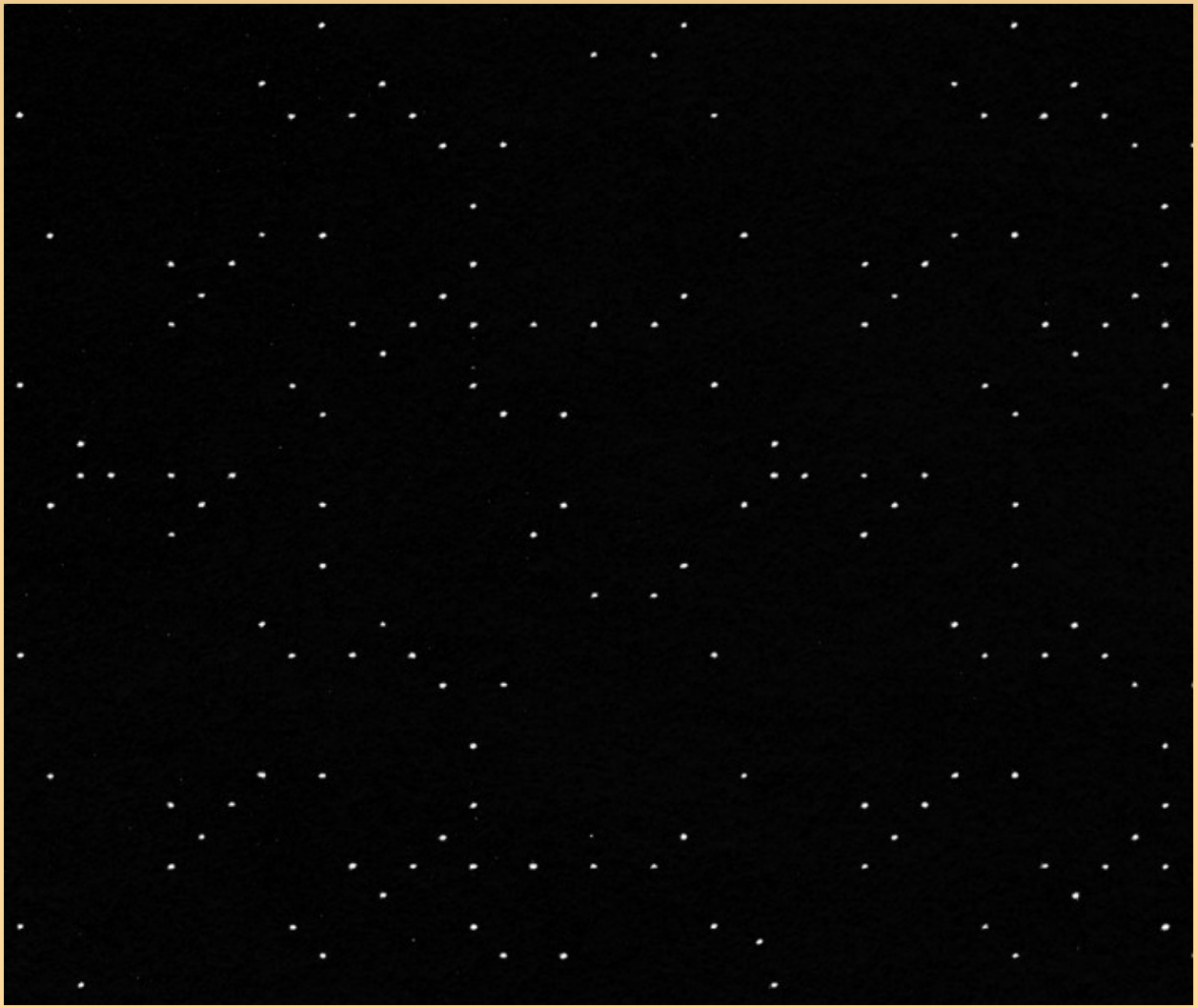
Epson, Konica/Minolta

- Values coded by position of dots within 2x3 rectangles, aligned to larger grid
- Date and time partly decoded
- Serial number not yet decoded



HP Color LaserJET

- Periodic on a near-square grid, no gaps



Canon

- Appears highly chaotic (but periodic)

Other manufacturers

- Some information at <http://www.eff.org/Privacy/printers/list.php>
  - Mostly empirical, from our sample library (which I'll discuss momentarily)
- Not classified by code family yet
- Trying to encourage consumers to buy printers without tracking dots, but can't promise that such printers (e.g. Xerox Phaser) do not contain tracking codes other than yellow dot patterns

Photocopiers

- Xerox DocuColor is also a photocopier (some models); seem to use same code
- Copiers are harder to study than printers
  - Poorer control of (analog) input data, though you can color photocopy blank pages
  - Today, many fewer people own color copiers than printers; tiny number of samples sent in
- They may all or mostly turn out to have identical codes to some printer models
- What happens to **serial color copies** (on the same copier or different copiers)?

Even without breaking the code…

- We might be able to tell whether a given printer produced a given document if we have access to the printer to print new test data! *(Matching vs. decoding)*
- Also, some printers don't code date and time at all, and many such printers will print a completely identical dot pattern on every output page, forever
- This merely requires recognizing a forensic mark by its shape without understanding its structure

How can we learn to decode other printer codes? (1)

- Empirical study (want to help out?)
  - EFF has a large (though imperfect) library of sample data

photo:

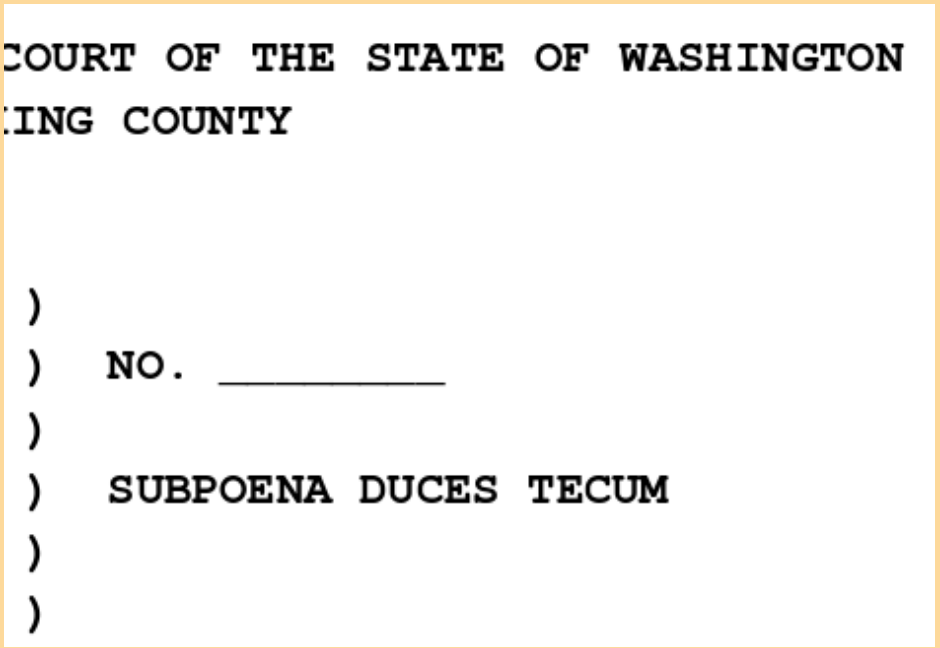
Quinn Norton

How can we learn to decode other printer codes? (2)

- We can likely share this library at no charge with researchers for the purpose of discovering *and publishing* details of how printer codes work
  - must agree to safeguard any personal information!

How can we learn to decode other printer codes? (3)

- In litigation in which details of codes are material to the interpretation of forensic evidence, it might be possible to subpoena a printer manufacturer as a third-party witness
- We'd be interested to hear about the results of trying this approach...



COURT OF THE STATE OF WASHINGTON
ING COUNTY

)
) NO. _____
)
) SUBPOENA DUCES TECUM
)
)

How can we learn to decode other printer codes? (4)

- We've filed a Freedom of Information Act request to the United States Secret Service
  - topics sought include technical details, history, cooperation of printer companies
- USSS has already missed statutory deadline to respond to our request by ca. two years, but continues to process it
- Possible claims of “investigative sources and methods”

How can we learn to decode other printer codes? (5)

- Reverse engineering of printers
- If imaging is handled in printer firmware (code written for a well-known embedded microprocessor), well and good (and we might even be able to figure out how to disable the tracking features by modifying the firmware)
- If imaging is handled in specialized imaging hardware, reverse engineering will be a lot harder!

Other printer tracking mechanisms

- NSF-funded Purdue research on inkjet watermarking and forensics (PSAPF) http://cobweb.ecn.purdue.edu/~prints/
- Broad project includes forensics of existing inkjet printers (based on mechanical differences, among other things) and hypothetical methods for making printer output more traceable
- Intentional vs. unintentional forensics
  - compare RID code, EXIF

Countermeasures

- Firmware modifications?
- Overprinting a decoy pattern?
  - Suggested by many people; I was skeptical
    - Problems: are tracking dots distinguishable from user-generated pixels? is their offset to the edge of the page predictable? etc.
    - Adding random noise can't work (at least if its frequency is different from tracking codes)
  - HP Color LaserJET experiment: offsets are predictable, 2px Y square at 600dpi wasn't *visually distinguishable* from tracking codes at small magnifications, so maybe effective

Decoy pattern strategies

- **Not clear if these work, how to tell!**
- Overprint all possible dot locations for a given printer model
- Calculate/observe dot locations that are not printed and print those (possibly hard if your printer codes date and time)
- Print *extremely* high-intensity noise
- Add several false candidate patterns (??)
- Print on the same page using several or many different printers (??)

Seeing Yellow project

- An MIT Media Lab project that asks people to complain to printer vendors about this feature and ask how to disable it; thousands of people have done so already
- Responds to experience of one person who was visited by Secret Service after asking how to disable tracking

<http://www.seeingyellow.com/>

Live demo

- QX5 microscope on Linux with and without blue illumination
- Scanner images: selecting blue channel, enhancing contrast
- Identifying and decoding Xerox DocuColor code
- DocuColor print samples:
  - FedEx Kinko's 100 California Street: 620350
  - FedEx Kinko's 369 Pine Street: 685956

Contact information

http://www.eff.org/

http://www.eff.org/Privacy/printers/

<schoen@eff.org>

9B36 BCFA 4DE0 8ADE 8A17 D091 56B0 315F 0167 CA38

Thanks to Rob Lee, Joel Alwen, and other volunteers!