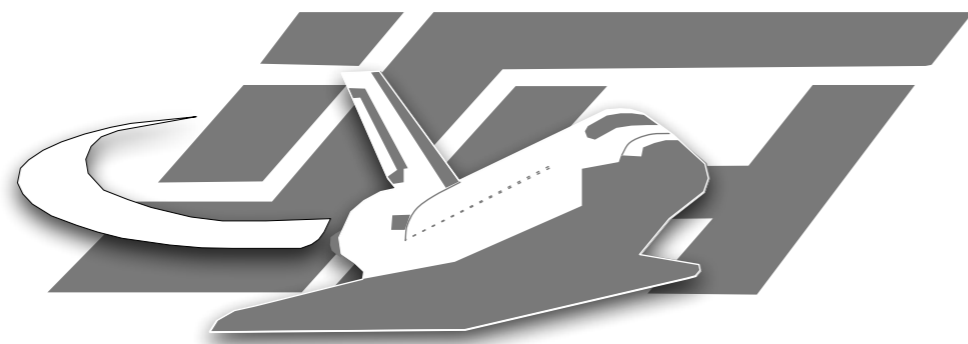


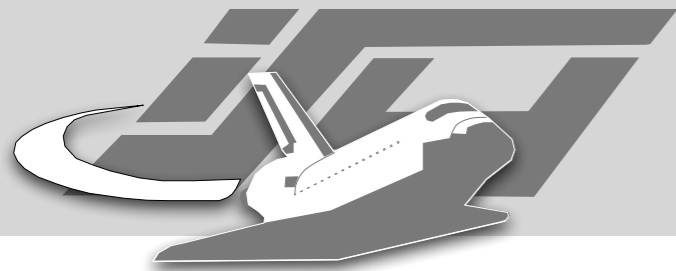
Far More Than You Ever Wanted To Tell

Hidden Data In Document Formats

Maximillian Dornseif
at Defcon 2004, Las Vegas

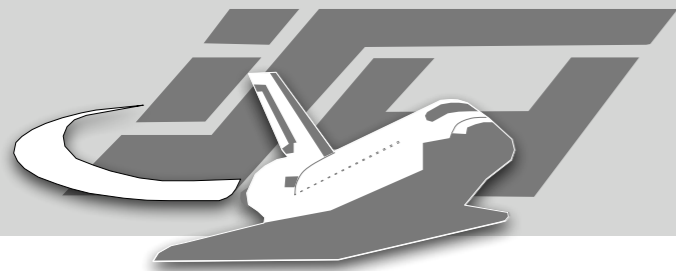


RWTHAACHEN
RHEINISCH-WESTFÄLISCHE TECHNISCHE HOCHSCHULE AACHEN



Warning

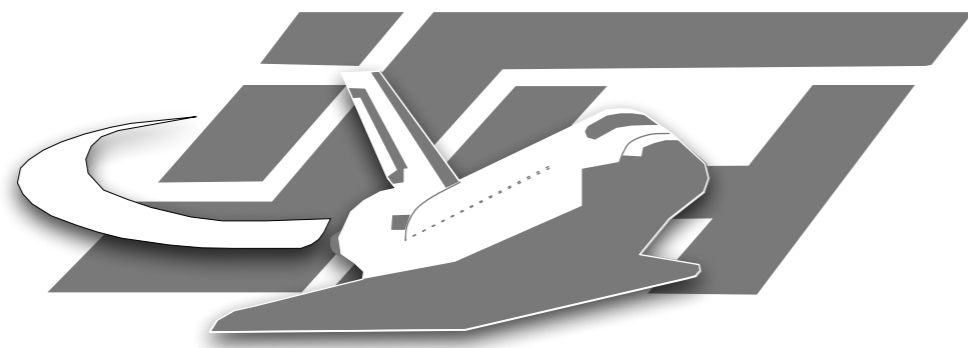
- When you read this this slides are probably outdated.
- Get the slides used in the presentation at <http://md.hudora.de/presentations/#hiddendata-dc>
- The PDF version has no demos. Go for the Quicktime version,



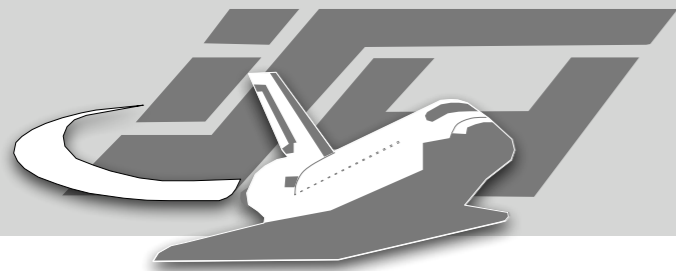
The Problem

- Complex Dataformats
- We are not supposed to understand
- or we are not willing to understand
- Covert channels everywhere!

Examples

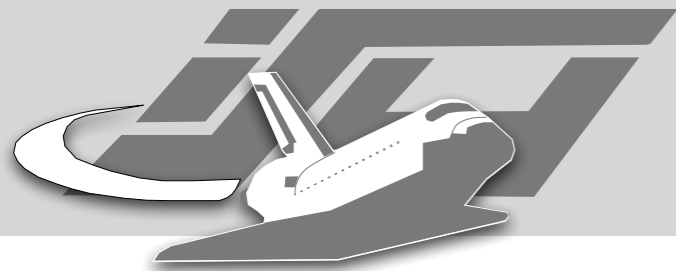


RWTHAACHEN
RHEINISCH-WESTFÄLISCHE TECHNISCHE HOCHSCHULE AACHEN



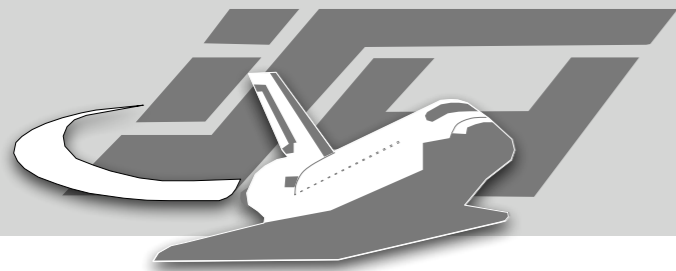
Mail- & News-Headers

- RfC 822 and friends are well known in the techie community but a mystery to everybody else.
- Data in there possibly include: OS, IP, server, software and their versions, organisation, time, customer number at isp / telephone number (!), etc.



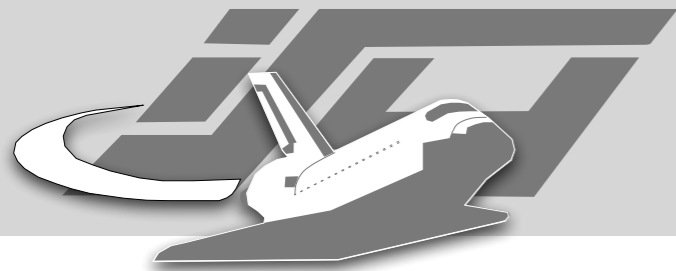
Incidents

- T-Online



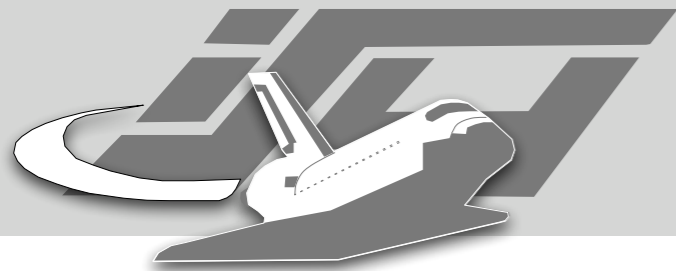
Config Files

- Config files which are not well understood are a security issue...
- ... but also can result in disclosure of information which is not to be disclosed



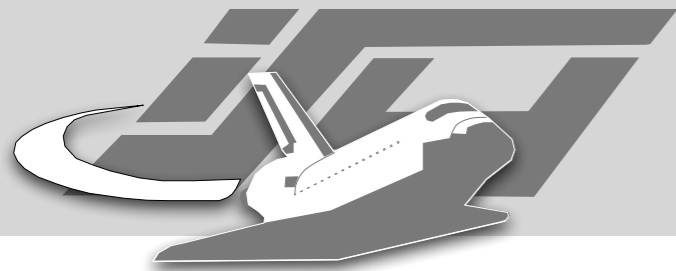
Incidents

- Apache
- BitchX



HTML

- Complex programs generate complex HTML
- Most obvious:
 - META generator
 - Paths to local files

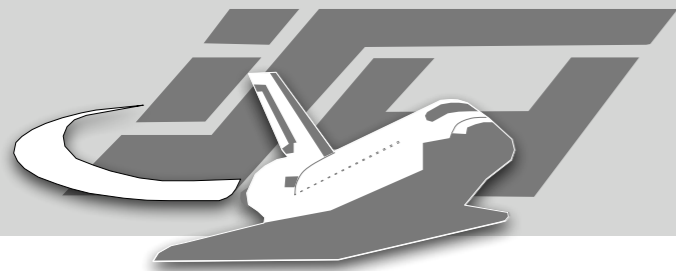


Incidents

- Defaced web pages (attrition.org)

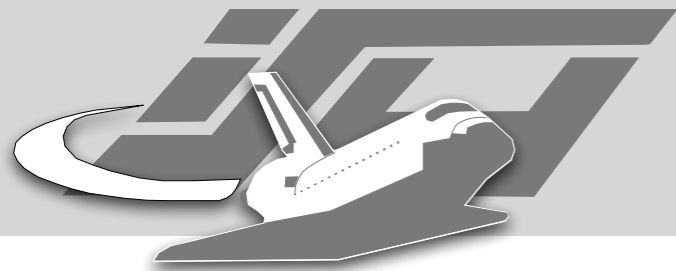
```

```



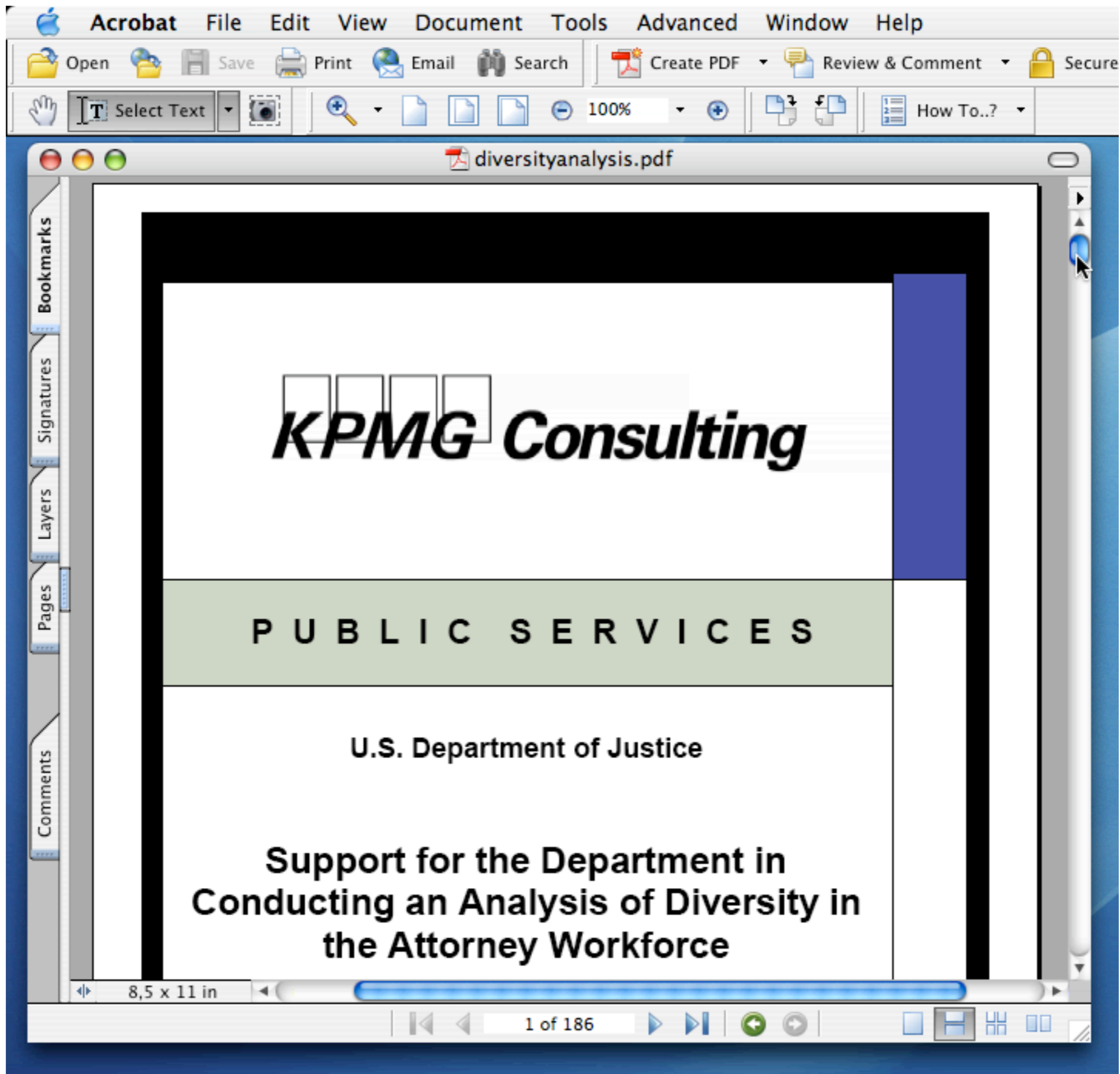
PDF

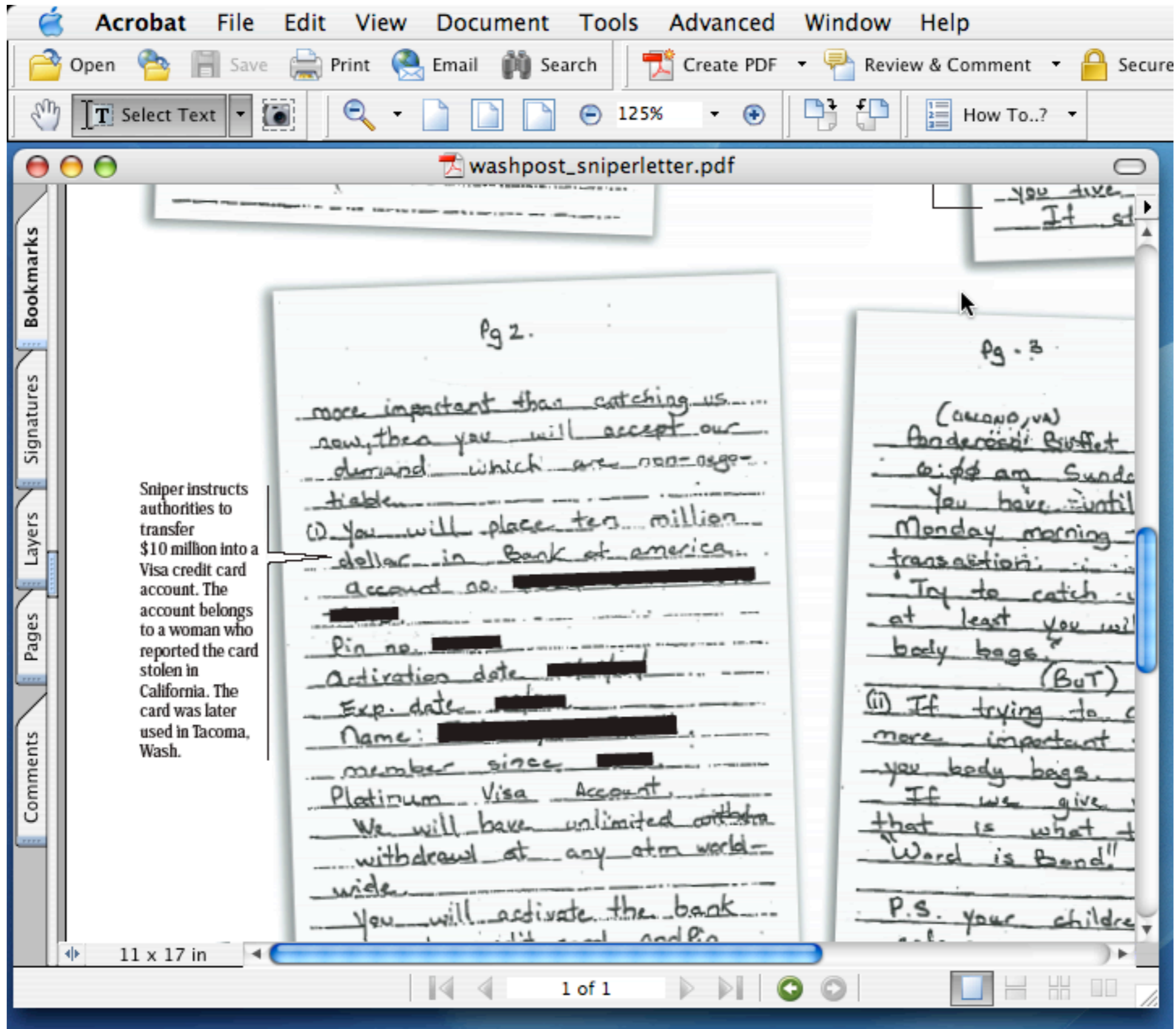
- Looks like an “open standard” ...
- ... but very hard to decode in depth
- The Problem of [REDACTED] / redaction.

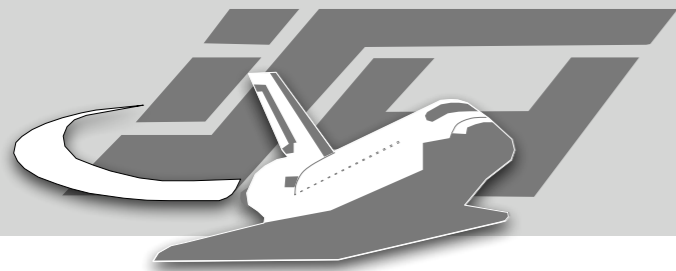


Exploiting hidden data

- Copy black text on black ground
- Copy underlying graphics

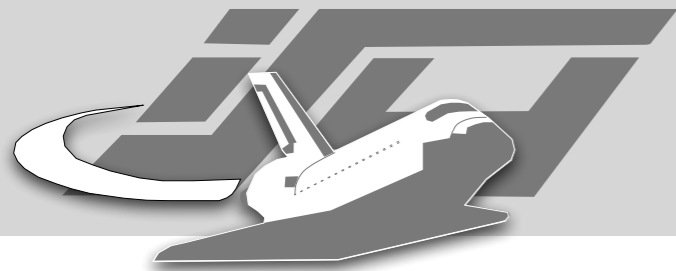






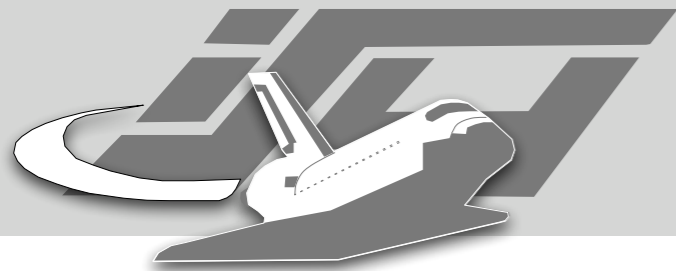
Incidents

- Sniper Letter
- The Justice Dept's Attorney Workforce Diversity Study
- "Secrets of History: The C.I.A. in Iran"

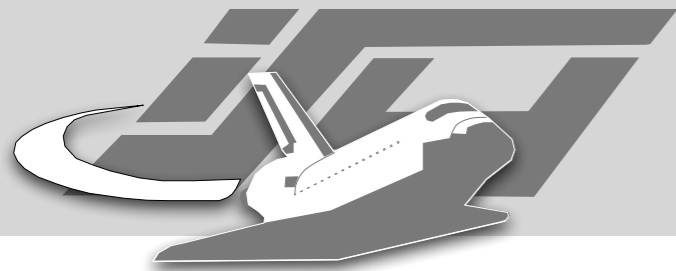


MS Office

- The MS Office document format is incredibly complex, undocumented and ever changing

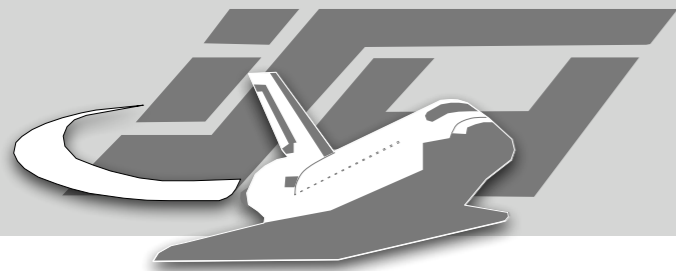


- Documented incidents include:
 - Text from a completely unrelated document edited before appears in the file.
 - Data deleted from the document or overwritten is appears in the file.



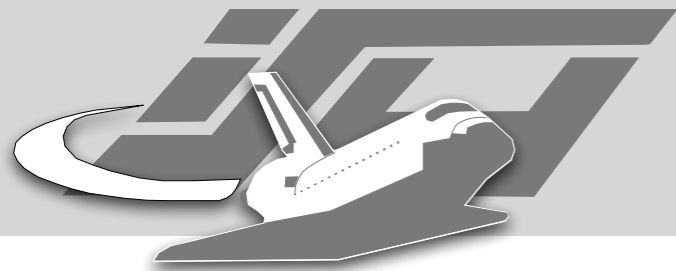
Incidents

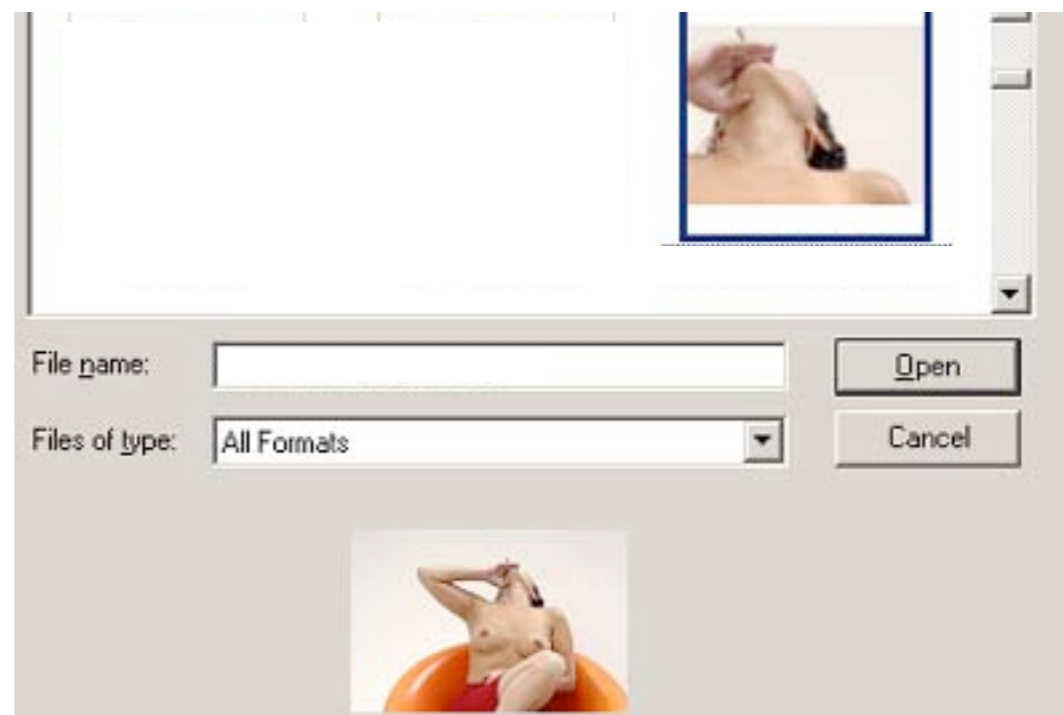
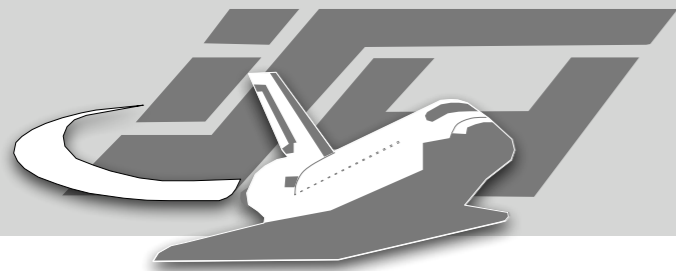
- UK Irak Dossier
- Transrapid / Rheinbraun / Managment / Machbarkeitsstudie
- Melissa

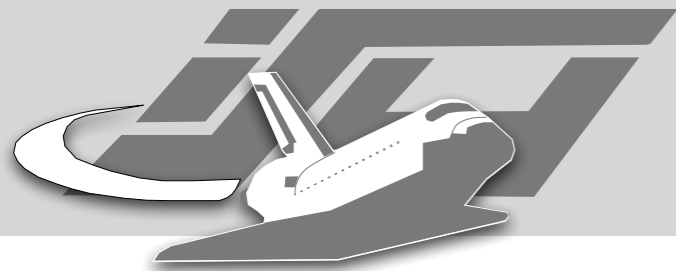


JPEG/EXIF

- Many image formats contain comment fields which might disclose unwanted data.
- JPEG has the extensible EXIF format for meta data.
- There was a remarkable incident with EXIF thumbnails

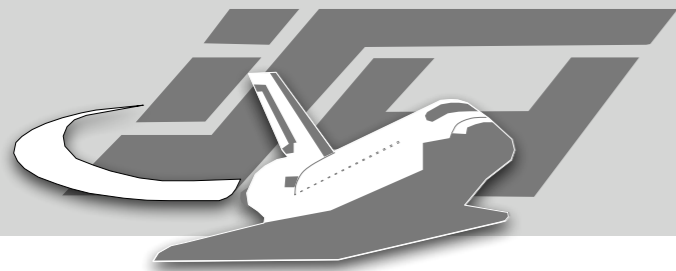






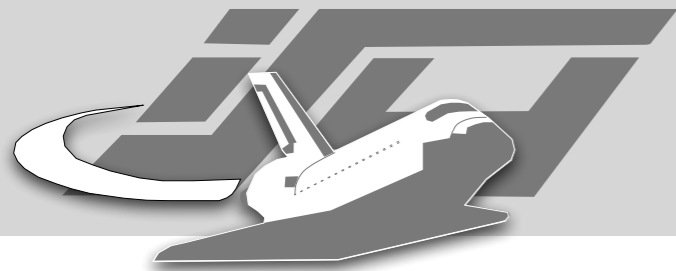
Misc

- Starr Report
- Embedded Serials / GUIDS
- unregistered marks
- ...



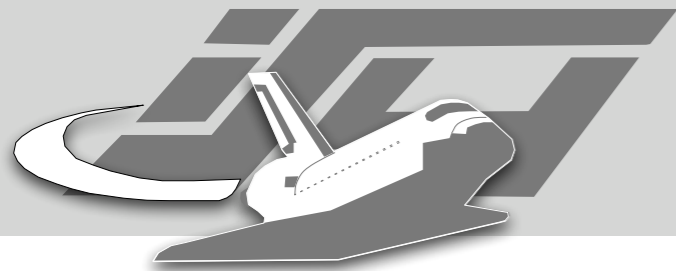
An Experiment

- Idea
 - Crawl the Web
 - Download Documents
 - Find the ones with hidden data.
- Problem:
 - How to detect hidden data?



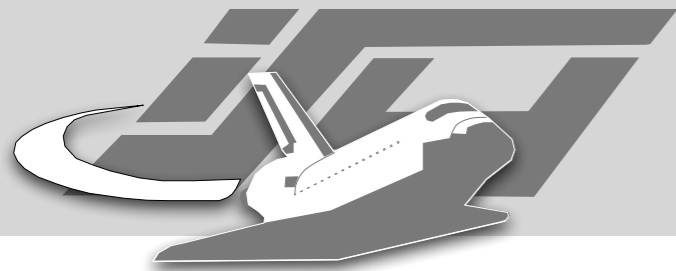
The Byers Experiment

- Scalable Exploitation of, and Responses to Information Leakage Through Hidden Data in Published Documents, Simon Byers, IEEE Security & Privacy pp. 23-27, March / April 2004



Our Experiment

- Technical details and demo



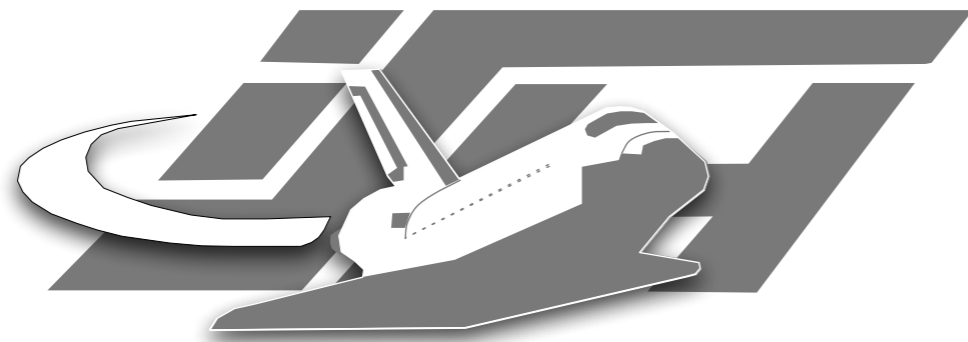
Conclusions

- You never know what proprietary formats carry
- Open formats are only part of a solution
- Spider the web and enjoy

Thank You!

Maximillian Dornseif <dornseif@informatik.rwth-aachen.de>

Slides at <http://md.hudora.de/presentations/#hiddendata-dc>



RWTHAACHEN
RHEINISCH-WESTFÄLISCHE TECHNISCHE HOCHSCHULE AACHEN