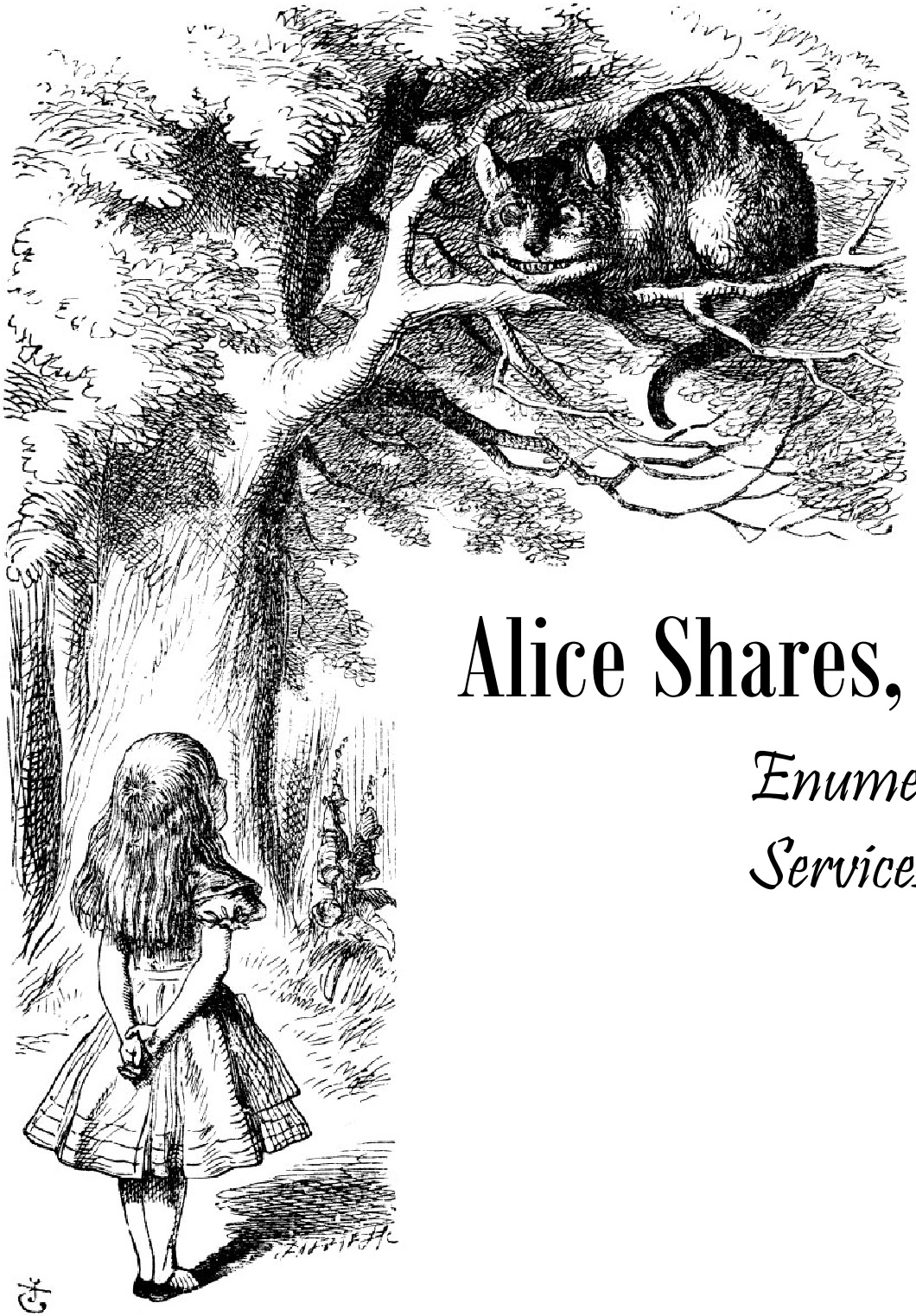


Alice Shares, Eve Reads

Enumerating File Hosting Services

Nick Nikiforakis
Katholieke Universiteit Leuven, Belgium

Outline

- The Cloud
- File Hosting Services
  - Workings
  - Token Generation
  - Privacy
- Enumeration
- Results
- Security issues
- Conclusion

The Cloud...


GOGRID


xls


CLOUDMARK®


amazon
web services™


3PAR
Serving Information


3LEAFSYSTEMS


Appistry


Akamai


ElasticHosts
Flexible servers in the cloud

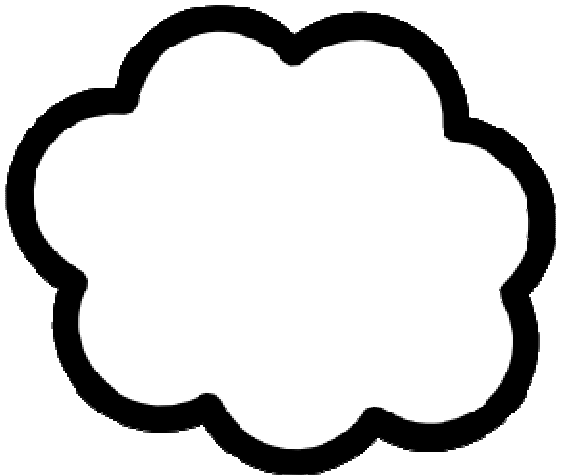

FileFactory


sendspace


RAPIDSHARE
EASY FILEHOSTING

What is the cloud?

- Shared resources on demand
- Pay-as-you-go style
- No need for small/medium company to buy and create their own infrastructures
  - Great for startups
  - Not so great for private data



When the cloud turns gray

- Data losses
  - 2007: Amazon EC2
    - Customer Application Data
  - 2009: Microsoft: T-Mobile Sidekick data
    - SMS, Calendars, pictures etc.
- Privacy
  - Chrome OS
  - Google Printing Service

File Hosting Services

- Cloud storage for the masses
- One-click hosting
- Mostly anonymous access
  - At least for non-paying users
- Used for sharing both public & private files

FHS Workings

- User chooses a FHS
- Uploads a file through their web interface
  - HTML + JS + FLASH
- The file gets stored in one of the servers of the FHS
- FHS creates a token, assigns it to the file and returns the token to the user in a URL form

Sharing of files through FHS

- Once a file is uploaded it can be shared according to its nature:
  - Private – Link through email, IM, etc. (1-1)
  - Public – Link on forums, blogs, IRC (1-N)

Privacy of FHS

- Protecting files from non-owners
- Security through obscurity
  - Their services are not searchable
  - A user can access a file only if he knows the file's unique and secret identifier

Privacy of FHS

- Protecting files from non-owners
- Security through obscurity
  - Their services are not searchable
  - A user can access a file only if he knows the file's unique and secret identifier

Actual uploads

Service ID	First Upload	Second Upload
FHS 1	376567678/athcon.zip.html	376567757/athcon.zip.html
FHS 2	/b121h9f/n/athcon_zip	/b121ha7/n/athcon_zip
FHS 3	/1909943800/athcon.zip	/1909943802/athcon.zip
FHS 4	16141045/athcon.zip	16141055/athcon.zip
FHS 5	/2016359	/2016360
FHS 6	/?mozmocgxry5	/?j1jrj0qyden
FHS 7	/file/prsyrj	/file/v1o1sq
FHS 8	/athcon	/athcon_1

patterns?!?

Predictability

- Many services generate predictable tokens (URLs)
- Starting from a valid token, an attacker can enumerate the whole database
  - Access to tens of millions of files

Specifics

- FHS2
  - 1909943800, 1909943799, 1909943798…are all valid tokens
- FHS3
  - b121h9f, b121h9e, b121h9d… are all valid tokens
  - 18^7 > 600,000,000 files
- Lets enumerate them!

Enumeration

- One enumerator for each service
  - Several instances from several IP addresses
  - Waiting ~10 sec. between requests
    - Defeating blacklisting from possible IDS
    - 8640 records/per day, per service
  - Starting from a valid token and subtracting one

- What did we get?

Sneak Peek FHS2

- b00dd1d | 086-091_D04_S14.oneddl.wyxchari._-089_.rar | 374.91 KB
- b00dd1c | ASD.El.Fersaan.Ep38.By.Starz.rar | 106.23 MB
- b00dd1b | DJ_Tiesto-Lethal_Industry-Retail-CDM-2002-MTC_mov-world.net.rar | 54.49 MB
- b00dd1a | D_WAPINZ_-_Hidupku_Seorang_3_.mpg | 25.49 MB
- b00dd19 | 05_-_Fly_With_Me.mp3 | 3.55 MB
- b00dd18 | KunoFch001.rar | 58.49 MB
- b00dd17 | Calle_13_Ft_Mercedes_Sosa_Para_Un_Nino_De_La_Calle_Www.FlowHoT.NeT_.mp3 | 4.89 MB
- b00dd16 | Document.zip | 499.61 KB
- b00dd15 | DSC_8973.jpg | 6.27 MB

Sneak Peek FHS3

- 1909260240 | LISTINO LORDO 2010 AGGIORNATO 200110.xls (0.7 MB)
- 1909260239 | Almoraima (BulerÃ-as).mp3.zip (8.7 MB)
- 1909260238 | Desi_Table.3gp (5.7 MB)
- 1909260237 | Bizim_Same_v1.05_By_USLUBank.rar (4.1 MB)
- 1909260236 | O_Kay_.part4.rar (99.2 MB)
- 1909260235 | P1010562.JPG (1.5 MB)
- 1909260234 | RecoverMyFiles3.9.8.5966.exe (7.6 MB)
- 1909260233 | LISTINO LORDO 2010 AGGIORNATO 200110.xls (0.7 MB)
  1909260232 | Suigintou_Rozen Maiden.jpg (4.0 MB)

How to find interesting data

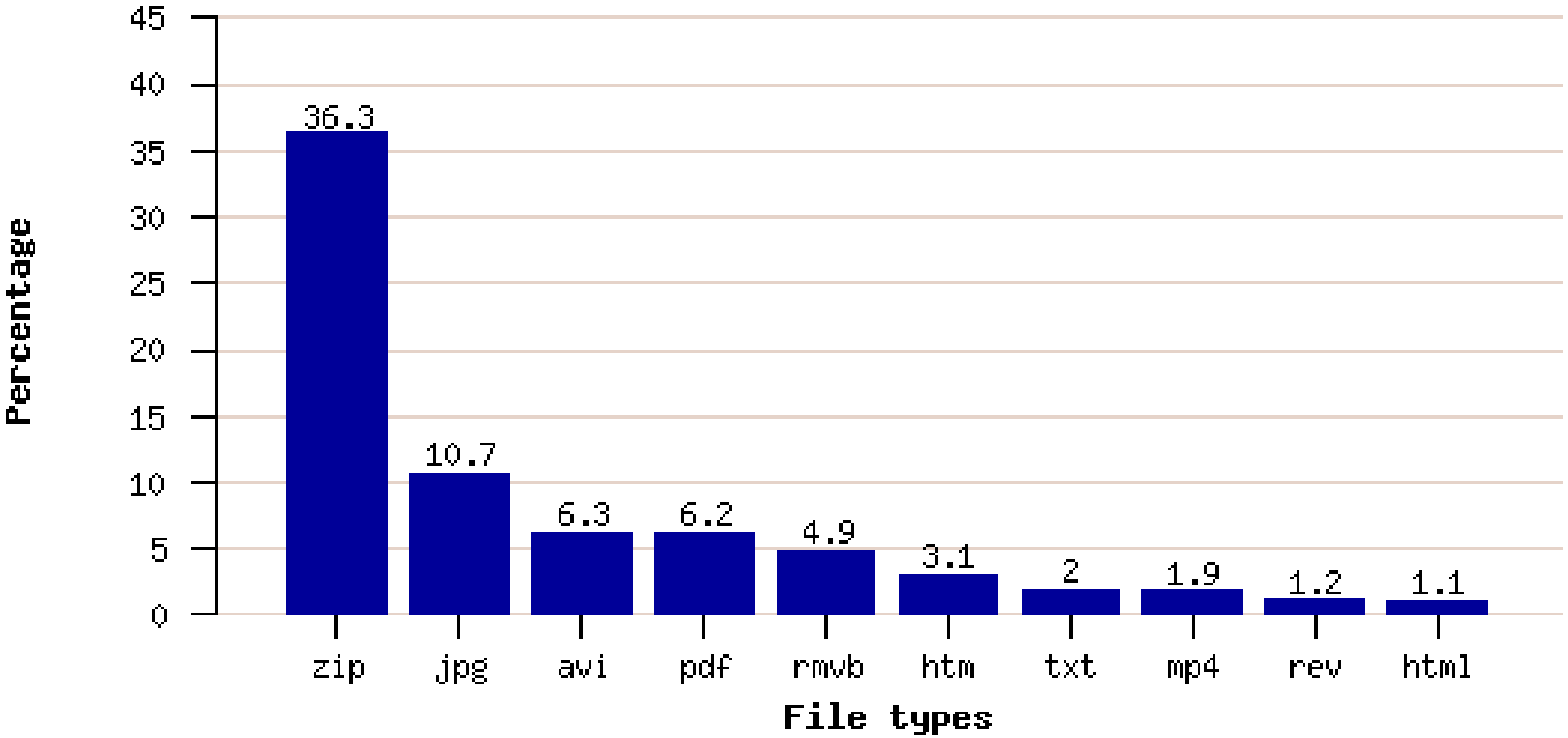
- Data is simply too much for manual inspection
- Automatic privacy classification engine

- Google search engine
  - Search for files on search engines and look at the results
  - Actually Yahoo! because Google blocked us ☺
  - 5,000 requests per day

Results

- 1 in 5 files returns no search results
- 30,000 private files… (so far)
  - Pictures
  - Documents
  - Spreadsheets
  - PHP pages
  - .sql files
  - …

Results



Top10 Private file types

Memorable moments

- Bank statements
- Company Budgets and salaries
- Phones, names, emails, dates of birth
- Death certificate
- Service manual for photo-printer
- 14 documents with doctor-transcribed notes

Attacks made possible

- Identity theft
  - Private pictures, documents
  - Personal data
- Scamming
- Server attacks
- Corporate espionage
- Blackmailing

The problem is…

- This is not easily fixable
- Even if the tokens from now on are secure
  - File Hosting Providers cannot change the tokens for the files that exist so far

- Dilemma
  - Delete several millions of files
    - And make your customers angry
  - Keep them…

Protect yourself

- If you must use a FHS:
  - Choose one which generates truly random tokens
  - Password-protect your file
  - Delete it once you have successfully shared it

Conclusion

- Most file hosting services are insecure
- Minimal effort => Maximum results
- Not easily solvable for existing FHS

Thank you

- Q&C?