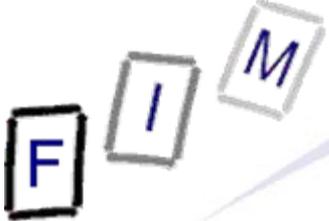


# Collecting information

Institute for Information Processing and  
Microprocessor Technology (FIM)  
Johannes Kepler University Linz, Austria

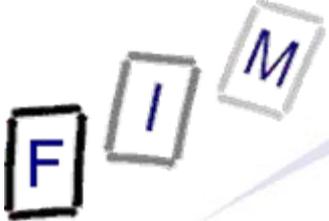
E-Mail: [sonntag@fim.uni-linz.ac.at](mailto:sonntag@fim.uni-linz.ac.at)  
<http://www.fim.uni-linz.ac.at/staff/sonntag.htm>



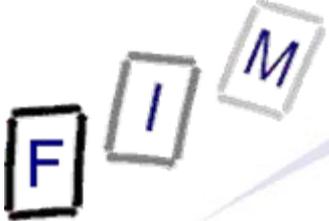
# Agenda

---

- NMap
- Google Hacking
  - Special search operators
  - Google cache
  - Robots.txt



- NMap (Network MAPper) is a network scanner
  - It tries to find all computers in a specific network and checks what ports are open, what OS they are running, whether there is a firewall, etc.
- It does not look for specific vulnerabilities!
  - But it gives recommendations; e.g. services to disable
  - Some scans + vuln. systems → Lock-up/crash!
- Used as a tool for inventory generation in a network
  - Are there any computers which should not be there?
  - Can also be used to gather information for a later attack
    - » Which OS/software and which version is running
- Stages: 1 = Host discovery, 2 = Port scan, 3 = Service/version detection, 4 = OS detection, 5 = Scripting
  - Scripting may also include vulnerability/malware detection!

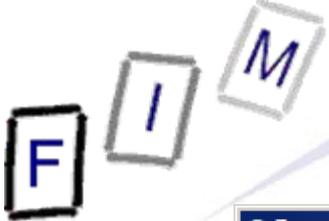


- Usage:

- Start program and enter IP address
- Select profile for scanning
  - » Special options only available in the command line version or when constructing a new profile!

- Your tasks:

- Install NMap (+ the UI – Zenmap)
- Scan the local subnet for hosts
  - » Use a “Quick scan”
- Scan the machine of your neighbour
  - » Use a “Regular scan”
- Interpret the results
  - » Correct output?
  - » Something surprising/dangerous found?



# Sample result: NMap local subnet scan

The screenshot shows the Zenmap application window. At the top, the 'Ziel' (Target) is set to '140.78.100.128/25' and the 'Profil' (Profile) is 'Regular scan'. The command entered is 'nmap 140.78.100.128/25'. The main display area shows a list of hosts on the left and their scan results on the right. The results are displayed in a terminal-like format, showing open ports and services for several hosts.

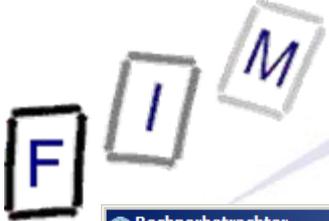
Rechner	Dienste
r1-intern	
hp2626-	
hp2824-	
hp2824-	
hplj4100	
npi8054-	
jrm_w7-	
habib.fin	
alex_v6-	
hoer_xp	
alex_w2	
cs140-7	
fim_mad	
praher-v	
son_vist	
140.78.	
inge_sta	

```
nmap 140.78.100.128/25
Host is up (0.0028s latency).
Not shown: 998 closed ports
PORT      STATE SERVICE
22/tcp    open  ssh
443/tcp    open  https
MAC Address: 00:11:85:C9:64:60 (Hewlett Packard)

Nmap scan report for hplj4100dtn.fim.uni-linz.ac.at (140.78.100.138)
Host is up (0.0028s latency).
Not shown: 991 closed ports
PORT      STATE SERVICE
21/tcp    open  ftp
23/tcp    open  telnet
80/tcp    open  http
280/tcp   open  http-mgmt
443/tcp   open  https
515/tcp   open  printer
631/tcp   open  ipp
1782/tcp  open  hp-hcip
9100/tcp  open  jetdirect
MAC Address: 00:01:E6:53:57:FA (Hewlett-Packard Company)

Nmap scan report for npi805442.fim.uni-linz.ac.at (140.78.100.140)
Host is up (0.0028s latency).
Not shown: 991 closed ports
PORT      STATE SERVICE
21/tcp    open  ftp
23/tcp    open  telnet
80/tcp    open  http
280/tcp   open  http-mgmt
443/tcp   open  https
515/tcp   open  printer
631/tcp   open  ipp
1782/tcp  open  hp-hcip
9100/tcp  open  jetdirect
MAC Address: 00:01:E6:80:54:42 (Hewlett-Packard Company)
```

# Sample result: NMap info



Rechnerbetrachter

Hosts: r1-inte, router, inge\_s, habib., jrm\_w, hplj410, hp282, hp282, hp262, alex\_v, praher, cs140, 140.78, fim\_ma, hoer\_, npi805

Allgemein | Dienste | Traceroute

**Allgemeine Informationen**

Adresse: [ipv4] 140.78.100.31

Rechnername: [PTR] router.fim.uni-linz.ac.at

**Betriebssystem**

Benutzte Ports: 1/tcp closed

Klasse | Fingerabdruck

%	Vendor	Type	Family	Version
100	Cisco	router	IOS	12.X
100	Cisco	switch	IOS	12.X

Reihen

Rechnerbetrachter

Hosts: r1-inte, router, inge\_s, habib., jrm\_w, hplj410, hp282, hp282, hp262, alex\_v, praher, cs140, 140.78, fim\_ma, hoer\_, npi805

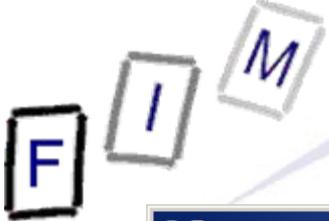
Allgemein | Dienste | Traceroute

Ports (5) | Extraports (995) | Spezialfelder

Port	Protocol	State	Service	Method
135	tcp	filtered	msrpc	table
135	tcp	state	reason_ip	
135	tcp	state	state	filtered
135	tcp	state	reason	
135	tcp	state	reason_ttl	
135	tcp	service	product	
135	tcp	service	name	msrpc
135	tcp	service	extrainfo	<Spezialfeld>
135	tcp	service	version	
135	tcp	service	conf	3
135	tcp	service	method	table
139	tcp	filtered	netbios-ssn	table
445	tcp	filtered	microsoft-ds	table
593	tcp	filtered	http-rpc-ssm	table

# Sample result: NMap info

The screenshot shows the Zenmap application window. At the top, there are tabs for 'Scan', 'Werkzeuge', 'Profil', and 'Hilfe'. The 'Ziel' (Target) field contains '140.78.100.31' and the 'Profil' (Profile) is set to 'Comprehensive'. The 'Befehl' (Command) field shows the full NMap command: `nmap -sS -sU -sV -T4 -O -A -v -PE -PM -PP -PS -PA -PU -PO -PY 140.78.100.31`. Below the command field are buttons for 'Rechner' and 'Dienste'. The main area is divided into several tabs: 'Nmap-Ausgabe', 'Ports / Rechner', 'Netzstruktur', 'Rechner-Details', and 'Scans'. The 'Rechner-Details' tab is active, showing details for the host 'router.fim.uni-linz.ac.at (140.78.100.31)'. The details are organized into sections: 'Kommentare', 'Rechnerstatus', 'Adressen', and 'Rechnernamen'. The 'Rechnerstatus' section includes a computer icon with a question mark and lists: Status: up, Geöffnete Ports: 0, Gefilterte Ports: 5, Geschlossene Ports: 995, Gescannte Ports: 1000, Laufzeit: Not available, and Letzter Systemstart: Not available. The 'Adressen' section lists IPv4: 140.78.100.31, IPv6: Not available, and MAC: Not available. The 'Rechnernamen' section lists Name - Typ: router.fim.uni-linz.ac.at - PTR. On the left side, there is a list of hosts under the 'Rechner' tab, including 'router.fim.uni-linz.ac.at', 'r1-intern', 'hp2626', 'hp2824', 'hplj4100', 'npi8054', 'jrm\_w7.', 'habib.fim', 'alex\_v6.', 'hoer\_xp', 'alex\_w2', 'cs140-78', 'fim\_mad', 'praher-v', 'son\_vist', '140.78.', and 'inge\_sta'. At the bottom left, there is a 'Filtere Rechner' button.



# Sample result: NMap info

Zenmap

Scan Werkzeuge Profil Hilfe

Ziel: 140.78.100.31 Profil: Comprehensive Scan Abbrechen

Befehl: nmap -sS -sU -sV -T4 -O -A -v -PE -PM -PP -PS -PA -PU -PO -PY 140.78.100.31

Rechner Dienste

Betriebssystem Rechner

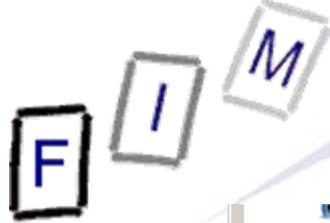
- router.fi
- r1-interr
- hp2626-
- hp2824-
- hp2824-
- hplj4100
- npi8054-
- jrm\_w7.
- habib.fi
- alex\_v6-
- hoer\_xp
- alex\_w2
- cs140-78
- fim\_mad
- praher-v
- son\_vist
- 140.78.
- inge\_sta

Nmap-Ausgabe Ports / Rechner Netzstruktur Rechner-Details Scans

nmap -sS -sU -sV -T4 -O -A -v -PE -PM -PP -PS -PA -PU -PO -PY 140.78.100.31 Details

```
Discovered open port 161/udp on 140.78.100.31
Completed UDP Scan at 15:21, 814.93s elapsed (1000 total ports)
Initiating Service scan at 15:21
Scanning 5 services on router.fim.uni-linz.ac.at (140.78.100.31)
Service scan Timing: About 60.00% done; ETC: 15:23 (0:00:51 remaining)
Completed Service scan at 15:22, 77.51s elapsed (5 services on 1 host)
Initiating OS detection (try #1) against router.fim.uni-linz.ac.at (140.78.100.31)
Initiating Traceroute at 15:22
Completed Traceroute at 15:22, 0.01s elapsed
Initiating Parallel DNS resolution of 2 hosts. at 15:22
Completed Parallel DNS resolution of 2 hosts. at 15:22, 0.00s elapsed
NSE: Script scanning 140.78.100.31.
Initiating NSE at 15:22
Completed NSE at 15:23, 5.01s elapsed
Nmap scan report for router.fim.uni-linz.ac.at (140.78.100.31)
Host is up (0.00024s latency).
Not shown: 1984 closed ports
PORT      STATE      SERVICE      VERSION
135/tcp   filtered  msrpc
139/tcp   filtered  netbios-ssn
445/tcp   filtered  microsoft-ds
593/tcp   filtered  http-rpc-epmap
1434/tcp  filtered  ms-sql-m
67/udp   open|filtered dhcps
123/udp   open       ntp           NTP v4
| ntp-info:
| receive time stamp: 05/17/11 15:23:01
| system: cisco
| leap: 0
```

# Sample result: NMap info

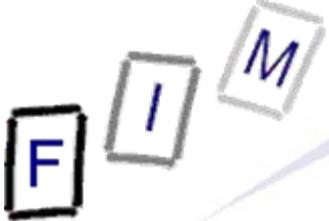


```
praher-v 1434/tcp filtered ms-sql-m
son_vist 67/udp open|filtered dhcp
140.78. 123/udp open ntp NTP v4
| ntp-info:
| receive time stamp: 05/17/11 15:23:01
| system: cisco
| leap: 0
| stratum: 4
| rootdelay: 4.33
| rootdispersion: 49.09
| peer: 34814
| reftime: 0xD17CF524.E5B5F39E
| poll: 6
| clock: 0xD17CF531.5D3A7B03
| phase: 0.212
| freq: 28.90
| error: 0.03
135/udp filtered msrpc
136/udp filtered profile
137/udp filtered netbios-ns
138/udp filtered netbios-dgm
139/udp filtered netbios-ssn
161/udp open snmp Cisco SNMP service
|_snmp-win32-shares: TIMEOUT
162/udp open|filtered snmptrap
1434/udp filtered ms-sql-m
10000/udp open|filtered ndmp
Device type: router|switch
Running: Cisco IOS 12.X
Too many fingerprints match this host to give specific OS details
Network Distance: 2 hops

TRACEROUTE (using port 8888/tcp)
HOP RTT ADDRESS
1 1.00 ms rl-intern.fim.uni-linz.ac.at (140.78.100.129)
2 1.00 ms router.fim.uni-linz.ac.at (140.78.100.31)

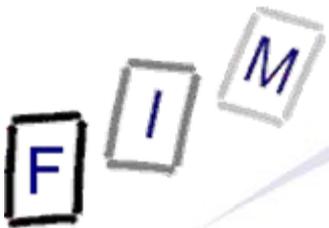
Read data files from: C:\Program Files (x86)\Nmap
OS and Service detection performed. Please report any incorrect results at http://nmap.org/
submit/ .
Nmap done: 1 IP address (1 host up) scanned in 909.68 seconds
Raw packets sent: 2296 (84.307KB) | Rcvd: 2040 (98.464KB)
```

Filtere Rechner

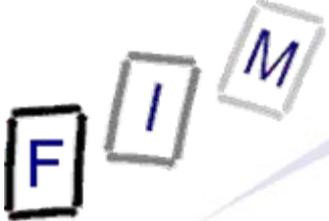


# Google hacking

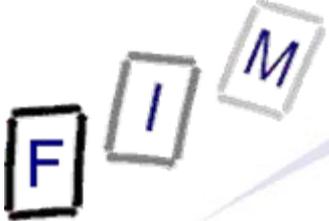
- Not an attack as such, but the preliminaries: Searching for vulnerable systems or vulnerabilities on a site
  - Using a search engine to look for known weaknesses
- Examples:
  - Looking for version numbers (vulnerable versions of software are known; websites running them will be prime subjects!)
  - Looking for "weak" code → "Google Code Search"
  - Search program comments indicating problems
    - » Like: `/* TODO: Fix security problems */`
- Note: The subject of the attack has no chance at all of noticing this, as his server is not touched at all!
  - Attacks come "out of the blue"
    - » But not unprepared: Only pages existing for a "long" time (typical indexing time: 2-3 weeks!) can be found
    - » Usually the vulnerability is older too



- Requires advanced Google operators:
  - link: Search within hyperlinks
    - » With certain words hinting at interesting pages
  - cache: Displays the page as it was indexed by Google
    - » Turn off image loading and you will not be logged on the server!
  - intitle: Within the title tag
    - » Directory listings: intitle:index.of
      - Better: intitle:index.of “parent directory”; intitle:index.of name size
  - inurl: Within the URL of the web page
    - » Webcams: inurl:"ViewerFrame?Mode=" inurl:"/axis-cgi/jpg/image.cgi?"
  - filetype: Only files of a specific type (no colon → filetype:doc)
    - » MS SQL server error: "A syntax error has occurred" filetype:ihtml
- Note: Such operators exist for most search engines
  - This is **not** a Google-specific problem!



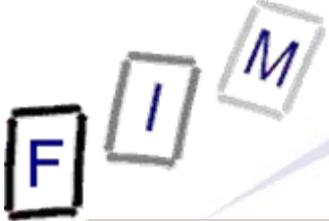
- Looking for specific vulnerabilities
  - Version numbers, strings, URLs, ...
- Error messages with too much information
  - Before “lockdown”, which logs errors and shows a simple message to the user only
- Files containing passwords
  - For offline breaking
- Logon pages
  - Where to actually attack
  - Title/content may give away information about limitations to passwords, method of storage, security precautions, ...
- Vulnerability information
  - All kinds of logs (web servers, firewalls, ...)
  - May also contain information about the internal network



- Searching for password lists (very old vulnerabilities!):
  - `inurl:/_vti_pvt/users.pwd`
  - `inurl:/_vti_pvt/administrators.pwd`
  - `inurl:/_vti_pvt/service.pwd`
  - Still requires to break passwords, but this can be done offline!
- HP JetDirect: Printers with an included web server
  - `inurl:hp/device/this.LCDispatcher`
    - » Note: These web pages typically cannot be changed at all!
    - » Only access can (and should!) be impossible from the Internet
  - Searching by title (model numbers) or strings (handbook, questions, ...) would not be successful here!
- Login portals of routers
  - `intitle:"Cisco Systems, Inc. VPN 3000 Concentrator"`
  - Only shows where to attack; passwords must still be guessed!
    - » But: Try passwords of producer; often the same for all appliances



- VNC viewers (Java client: Port 5800; server: Port 5900):
  - `intitle:VNC inurl:5800`
    - » Depending on page title the version/product can be distinguished
- Webcams (Axis):
  - `intitle:"Live View / - AXIS"`
    - » Title can be used for further restriction, e.g. the model used
- Server version:
  - `intitle:index.of server.at`
    - » Example result at bottom of page: “Apache/2.2.9 (Debian) mod\_ssl/2.2.9 OpenSSL/0.9.8g Server at www.????? Port 80”
      - mod\_ssl/OpenSSL version might also be **very** interesting!
    - Also the default test pages (after installation) often remain accessible even after installing the final website
      - » `intitle:welcome.to intitle:internet IIS` (see next slide!)
- Looking for know-vulnerable cgi files
  - `inurl:/random_banner/index.cgi`



# intitle:welcome.to intitle:internet IIS



OS version

**Your Web service is now running.**

You do not currently have a default Web page established for your users. Any users attempting to connect to your Web site from another machine are currently receiving an **Under Construction** page. Your Web server lists the following files as possible default Web pages: default.htm, default.asp, index.htm, iisstart.asp. Currently, only iisstart.asp exists.

To add documents to your default Web site, save files in c:\inetpub\wwwroot\.

Default pages

Document root

IIS version

**Welcome to IIS 5.1**  
 Internet Information Services (IIS) 5.1 for Microsoft Windows XP Professional brings the power of Web computing to Windows. With IIS, you can easily share files and printers, or you can create applications to securely publish information on the Web to improve the way your organization shares information. IIS is a secure platform for building and deploying e-commerce solutions and mission-critical applications to the Web.

Using Windows XP Professional with IIS installed, provides a personal and development operating system that allows you to:

- Set up a personal Web server
- Share information within your team
- Access databases
- Develop an enterprise intranet
- Develop applications for the Web.

IIS integrates proven Internet standards with Windows, so that using the Web does not mean having to start over and learn new ways to publish, manage,

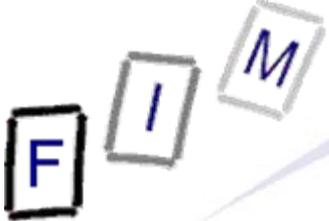
**Integrated Management**  
 You can manage IIS through the Windows XP Computer Management console or by using scripting. Using the console, you can also share the contents of your sites and servers that are managed with Internet Information Services to other people via the Web. Accessing the IIS snap-in from the console, you can configure the most common IIS settings and properties. After site and application development, these settings and properties can be used in a production environment running more powerful versions of Windows servers.

**Online Documentation**  
 The IIS online documentation includes an index, full-text search, and the ability to print by node or individual topic. For programmatic administration and script development, use the samples installed with IIS. Help files are stored as HTML, which allows you to annotate and share them as needed. Using the IIS online documentation, you can:

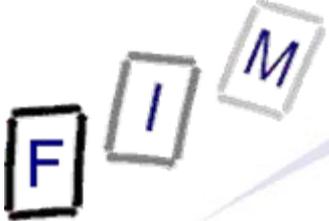
- Get help with tasks
- Learn about server operation and management
- Consult reference material
- View code samples.

Fertig

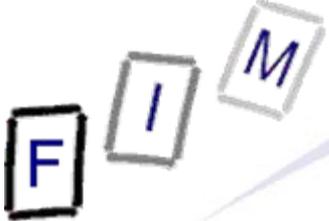




- MySQL database dumps
  - "# Dumping data for table (username|user|users|password)" - site:mysql.com -cvs
- phpMyAdmin: Database administration tools
  - intitle:phpMyAdmin "Welcome to phpMyAdmin \*\*\*" "running on \* as root@\*"
- Registry dumps
  - filetype:reg reg HKEY\_CURRENT\_USER username
- Looking for code/passwords (often contains cleartext pwds!)
  - filetype:inc intext:mysql\_connect
- Printers/Faxes:
  - inurl:webArch/mainFrame.cgi
- UPS:
  - intitle:"ups status page"



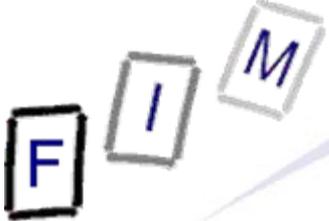
- The cache gives you access to old/removed content
  - Which might still be applicable!
- Attention: Surfing the cache will still touch the server
  - E.g. images are loaded from the “source”
- Way around: View the text-only version
  - Add “&strip=1” to the search URL



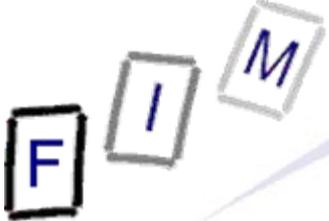
- Make sure that “private” computers are not accessible from the “public” internet
  - Use a firewall (packet filter alone might be insufficient)
- Automated tools available : E.g. SiteDigger
  - Can also be used on your own pages to look for “weaknesses“ (verification)!
- Check what Google (and others) know about your site
  - `site:www.mysite.com`
  - Is this **only** what **should** be accessible to everyone?
- Use "robots.txt" to limit web crawlers to "relevant" pages
- Captchas/Remove from Google index (→ Desirable?)
  - Not that easy and/or quick!
  - Requires often extensive measures (removal of page + notification of Google + wait for index)



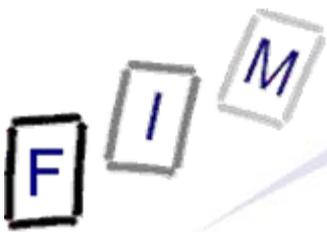
- The site is not attacked at all in this stage
  - Just some information is collected
  - The information is gathered from public sources
- In contrast to other attacks, this is legal in most countries!
  - Too far away from a concrete attack
    - » When trying it out on the real server (even if unsuccessful!), this is typically a punishable offence!
  - Note: UK and USA are notable exception!
    - » “Unauthorized access” is an offence
- BUT: If something happens, this can be used as evidence
  - Also, it is a very good evidence to prove intentionality
    - » When explicitly looking for weaknesses, you can later hardly claim that you sent a special request “accidentally” ...
  - Note, that finding evidence of Google hacking is difficult
    - » Requires access to your computer or log files of intermediaries (like proxies, wiretapping at the ISP, ...)



- Try out several of the examples before
  - E.g. webcams or database examples
  - Do they always work? What could be the reason?
- Access the Google cache for a website you know to be changing frequently
  - Check the differences to the current website
  - How old is the cached version?
    - » Approximately or can you identify the exact date?
  - Where do external links lead to?
    - » Archived version or live version?
  - Where are images loaded from?
    - » What difference can this make?
- Bonus task:
  - What is the “web archive”?
  - How is it similar to Google cache and what’s the difference?

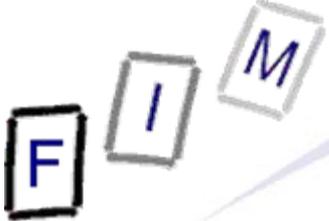


- Robot Exclusion Standard
  - Asking nicely to leave your site alone
    - » “Good” robots check for this file and adhere to it
    - » But technically there is no need!
    - » Example: Austrian National Library has the legal permission to archive website with strong connection to Austria → Ignores this file deliberately (legal permission + obligation!)
  - No official standard available!
  - Note: Crawling; indexing is different!
  - Must reside in the root directory of the site
- Alternative: META tags within the site
  - Drawbacks:
    - » Robot has already retrieved the site
    - » Works only for HTML pages
  - Advantage: Local control!
    - » Robots.txt is possible only site-wide!

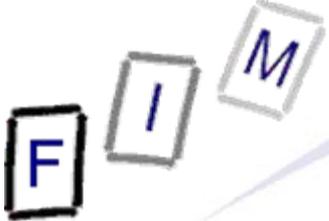


- What robots.txt is **NOT**:
  - A security measure: Anyone can access any page
    - » Retrieving robots.txt is no requirement!
    - » Use password, authentication, ... instead
  - A way of hiding data
    - » The location/its name is publicly visible
  - A tool to prevent indexing
    - » External URLs may still result in indexing
- What robots.txt **IS**:
  - A way to reduce the server load and the quality of search results by excluding uninteresting parts of the site
    - » Or those changing too frequently to be useful within the index
  - A way of providing information about the sitemap

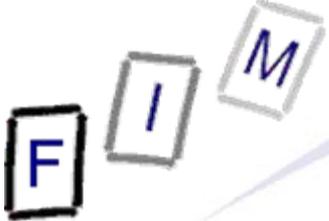
## Difficulties of later removing content



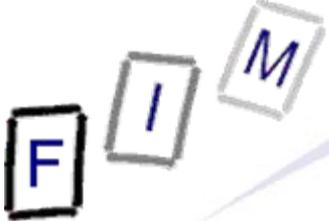
- Adding it to robots.txt
  - URL is known, so it is accessed and indexed
    - » Or: External links to the site → Again being indexed!
  - Only from the pages with the links, those links are ignored and are not followed
    - » At some time they might fall out of the index (several month)
- Potential solution: Add META-Tags
  - Problem: Doesn't work for .doc, .pdf, ...
  - But then these files **MUST NOT** be in the robots.txt!
    - » Must be allowed in robots.txt and individually excluded
- Real solution:
  - Use Google webmaster tools to remove
  - Use X-Robots-Tag for non-HTML file types
    - » Note: This is a HTTP header! Requires webserver configuration!



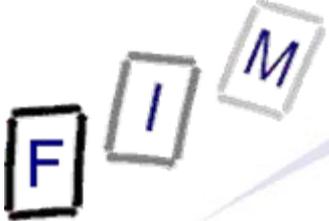
- Simple text file in the website root: “/robots.txt”
  - Attention: Might be case-sensitive (implement.-dependent)
- “User-agent: “ For which bot the following lines are intended
  - Note: Find out first, which one you want to block
    - » Google: “Googlebot”, “Googlebot-Image”, ...
    - » Yahoo: “yahoo-slrp”
    - » Microsoft: “msnbot” (MSN search), “psbot” (images)
    - » “\*” as wildcard for all bots
- “Disallow: “ What may not be followed
- “Allow: “ What may be followed (exceptions from Disallow)
  - This is no guarantee and doesn't force the bot to follow links!
- “Crawl-delay”: How many seconds to wait between retrieves
  - Note: Google does not follow this (→ Use webmaster tools!)
- “Sitemap: “ URL of the sitemap
  - Only selected bots (Google, Yahoo, MSN, ...)



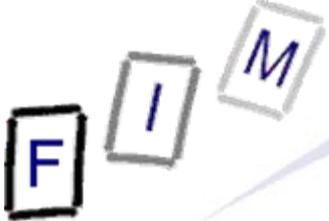
- Format for Disallow and Allow:
  - Empty: Ignore it
    - » Example: “Disallow: “ → Whole page may be crawled
  - Everything starting with the string provided
    - » Example: “Disallow: /” → Nothing may be crawled
    - » Example: “Disallow: /index” → Will not crawl:
      - “/index” as a file or a directory
      - “/index.htm”, “/index.html”: Files
      - “/indexing/”, “/index/”: Directories
  - “\$” end of line anchor
    - » Only Google, Yahoo, MSN
    - » Example: “Disallow: /\*.pdf\$” → Will not crawl pdf files
    - » Attention: No regular expressions allowed!
- Each command must be a separate line
- At least one “Disallow” line is required
- Empty line before 2nd, 3rd, ... User-agent line



- Example of “hiding” the complete site (= no crawling)
  - User-agent: \*
  - Disallow: /
- Example of typical exclusions:
  - User-agent: \*
  - Disallow: /cgi-bin/
  - Disallow: /tmp/
- Example of allowing only Google, but not Google images
  - User-agent: Googlebot ← Includes “Googlebot-Mobile”
  - Disallow:
  
  - User-agent: Googlebot-Image
  - Disallow: /
  
  - User-agent: \*
  - Disallow: /



- Create a robots.txt file with the following restrictions:
  - Allow Google, Yahoo and MSN access to the whole site
  - No access for image searching by anyone
  - No archiving by the web archive
  - No access to the directory “/news/today/”, but allow access to the subdirectory “/news/today/hot/”
  - No crawling of Microsoft Office documents
- Check whether these restrictions are possible at all
  - And whether they are possible with robots.txt
  - Or how they must be specified more exactly
- Find a verification tool and check your file with it



User-agent: Googlebot  
Disallow:

User-agent: Yahoo-slurp  
Disallow:

User-agent: msnbot  
Disallow:

User-agent: Googlebot-Image  
Disallow: /

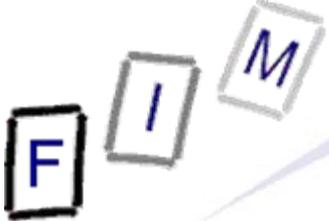
User-agent: psbot  
Disallow: /

User-agent: archive.org\_bot  
Disallow: /

User-agent: \*  
Disallow: /news/today/  
Allow: /news/today/hot/  
Disallow: /\*.doc\$  
Disallow: /\*.xls\$  
Disallow: /\*.ppt\$  
Disallow: /\*.docx\$

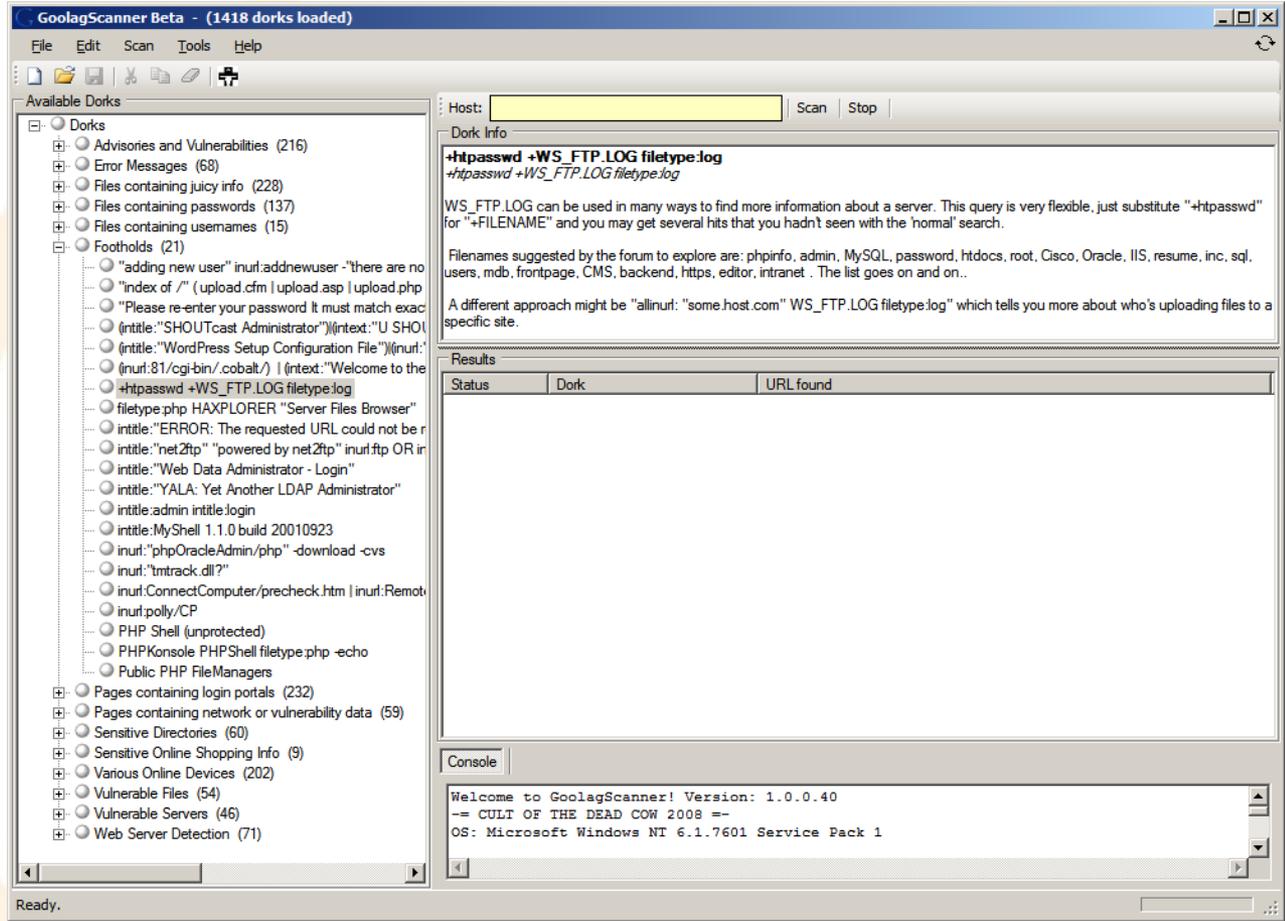
- Attention: Restrictions exist!

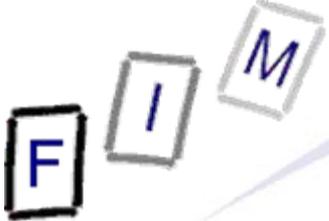
- /news/today/ ... will not apply to Google, Yahoo & MSN
  - » Or they would have to be added above!
  - » A problem of the specification too!
- “Microsoft Office documents” is too unspecific; only individual files (filename!) can be blocked
  - » Here only a few are shown; more exist!
- Empty Disallow is seen as illegal by many verifiers
  - » Can be replaced by “Allow: /”
- Wildcards are not supported universally
  - » \*, \$ will not work for all bots
  - » HTTP headers required for them



# Goolag scanner

- Can be downloaded from the Internet
  - Contains a very large number of interesting Google scans



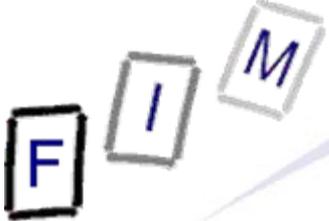


- Collecting information from third-party sites is very advantageous to attackers
  - The target website cannot notice anything suspicious
    - » It is not contacted in any way
- NMap gives a rough overview; but take care of logging
  - Better used once “inside” or generally from outside
  - Intense scanning is a hint of an attack
- Both are very “unreliable” as they will usually not give very useful information on a specific target system
  - More interesting for finding “something” to hack
- “General reconnaissance” tools!

F I M

# Questions?

Thank you for your attention!



# Literature/Links

---

- NMap  
<http://nmap.org/>
- Robots Database  
<http://www.robotstxt.org/db.html>