Mag. iur. Dr. techn. Michael Sonntag

Backtracing E-Mails

Computer Forensics

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

E-Mail: sonntag@fim.uni-linz.ac.at
http://www.fim.uni-linz.ac.at/staff/sonntag.htm

Backtracing E-Mails: What's this?

- When you receive an E-Mail, how do you know who sent it?
  - → Easy, look at the "Sender" field in the E-Mail program!
- But … This is trivially faked in many ways!
  - → Both online,
    - » How often did you receive Spam sent by yourself?
  - → and offline
    - » How difficult is it to write an "E-Mail" and print it out?
- However, based on header information, an E-Mail can be traced back to the sender's computer
  - → Or at least to the last trustworthy computer
    - » From this you can then obtain log files identifying the next hop
      - – If they exist …
- This is no foolproof solution: In many cases the trail will end somewhere in a foreign country on a server without logs

Information sources and their reliability

- An E-Mail is plain text and therefore easily manipulated
- Some sources are however trustworthy
  - → The final recipient's computer and his mailserver
    - » But note: Trojans, etc.!
  - → General information sites in the Internet
    - » WhoIs database, DNS, …
      - – But often they only have the data that was provided to them
  - → What the server's IP is
    - » Note: IP spoofing is also possible, but much more difficult!
  - → Recipient E-Mail address
- Some (most) sources are unreliable
  - → E-Mail headers unless from a chain of trusted computers
  - → Any information within the text of the E-Mail
  - → Envelope headers (see later)
  - → What the server's name is

Transporting an E-Mail: From source to destination

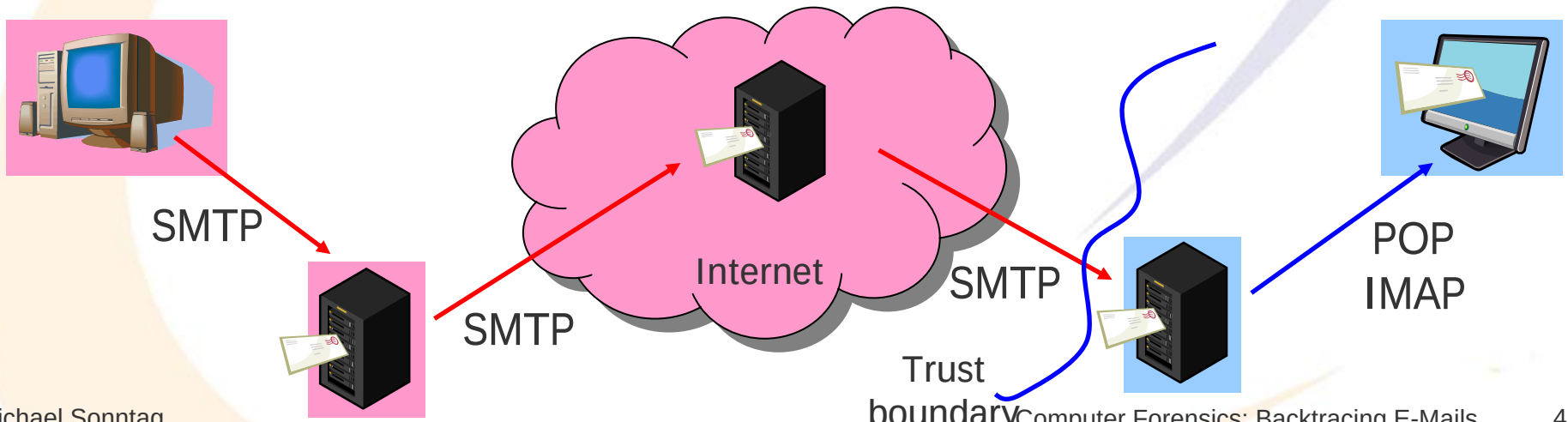
- A mail may be sent over several hops
  - → History: This was the only way to reach the recipient
  - → Today: Security precautions, backup, …
- Sender to mailserver: SMTP
- Mailserver to next mailserver: SMTP

Almost always untrusted

- Mailserver to recipient: POP, IMAP, …

Perhaps secured

  - → This stage is left out: We consider the last server to be trustworthy and obtain the complete E-Mail as text from it!



SMTP protocol (1)

1. Connection setup
   - ➔ HELO <client name>
   - ➔ 250 <greeting>
     - Typically includes the name and the IP address of the client
       - An "incorrect" client name (according to a DNS lookup) is usually ignored and the mail accepted anyway
2. Addressing (envelope information)
   - ➔ MAIL FROM: <sender address>
   - ➔ 250 <address repetition>… Sender ok
     - Often not verified at all (only syntax): But can be verified!
       - Example: Accept only mails from "local" addresses
   - ➔ RCPT TO: <recipient address>
     - Might be verified already here (e.g. only accept local mails)
       - To prevent relaying and avoid having to scan the whole mail body
     - This line may be repeated; server will then distribute it
   - ➔ 250 <address repetition>… Recipient ok

SMTP protocol (2)

3. Mail content (including the actual headers!)
   - ➔ DATA
   - ← 354 Enter mail, end with "." on a line by itself
   - ➔ <Mail headers>
   - ➔ <Empty line>
   - ➔ <Mail body>
   - ➔ .
   - ← 250 2.0.0 <id> Message accepted for delivery
     - Message-Id: <200710020842.l928g6aJ001036@mail.msv.at>
       - Contains timestamp, id, and servername
       - But this is no official format, just what sendmail does by default
4. Termination
   - Not really necessary, but polite (not recorded in E-Mail)!
   - ➔ QUIT
   - ← 221 2.0.0 <servername> closing connection

E-Mail headers

- Many standard headers exist, but arbitrary other headers may be included
  - → Sequence of different headers is of no importance
- Syntax: Headername ": " Headervalue <Lineend>
  - → Continuation possible through lines starting with a space or tab
- "X-" headers: Any header starting with "X-" is a custom one
  - → Everyone can add any such header to an E-Mail
  - → They have no standardized meaning
  - → Used by various mail clients to provide version info
    - » Also: Priority, scanning by antivirus, spam, …
      - – Note: The presence of such a header doesn't necessarily mean that scanning actually took place …

See also: http://people.dsv.su.se/~jpalme/ietf/mail-headers/mail-headers.html

General headers (1)

- Date: Time & date of creating/sending the E-Mail
  - → Added by the sender → Untrustworthy!
- Subject: What the message is about (heading)
- From: Source E-Mail address
  - → Freely set by sender → Untrustworthy!
- Reply-To: Address for replies or errors
  - → Freely set by sender → Untrustworthy!
- To: Recipient address
  - → Freely set by sender → Untrustworthy!
    - » Actual destination address is outer one (during SMTP protocol!)
- Return-Path: Contains envelope "From"
  - → Freely set by sender → Untrustworthy!

General headers (2)

- Message-Id: Unique reference of the E-Mail
  - → Content freely set by sender → Untrustworthy!
    - » Still important:
      - – Used for searching in logs at source and intermediate hosts!
      - – Also used for automatic threading
- CC: Carbon copy
  - → Freely set by sender → Untrustworthy!
  - → Destination addresses must be set using SMTP
- BCC: Blind Carbon copy
  - → Like CC!
- MIME-Version, Content-Type, Content-Transfer-Encoding, …:
  - → Used for specifying the character set and encoding used for the content; Usually correct (but of little forensic use …)
  - → Additionally used for attachments

"Received" header lines

- These are added at each step in the transmission chain
  - » Today typically only few lines: Few/No intermediate servers!
  - → In reverse order: The first (top) one is the last one
    - » This should be the one from your own SMTP server, from which you retrieved the E-Mail by POP/IMAP
      - – This line should be (**is -** per definition here!) trustworthy
  - → The last (bottom) one should be from the first server, i.e. the server which received the mail directly from the client's E-Mail program (Client-SW, web interface, …)
- Note: These are not always reliable!
  - → Anonymizers might change this and remove lines
  - → Spammers might insert lines to hide the original sender
  - → Hints for modifications:
    - » Other headers in between "Received" lines
      - – They should form a single block
    - » "Gaps" in the chain, overlong lines, …

"Received" header lines: Syntax

- Syntax: "Received: from" <name1> "(" <name2> "[" <ip> "])" "by" <name3> "via" <name4> "with SMTP id" <id> "for" <address> ";" <datetime>
  - → Name 1: Name of sending computer according to DNS
  - → Name 2: Name of sending computer as in HELO
  - → IP: IP address of the sending computer
  - → Name 3: Name of receiving computer according to himself
  - → Name 4: Name of intermediate computer (rarely used)
  - → ID: Identification of the message
    - » Might change at each hop: ID of the receiving computer!
  - → Address: Envelope destination address
  - → Datetime: Date and time of receiving (local recipient time)
- Please note: There is no strict syntax definition; this is just the most common format!

Analyzing "Received" header lines

- Start from the first line
  - → Check whether DNS and provided names match
    - » Or did he only provide an IP (correct?) as the name?
  - → Check whether the "from" name matches the name in the next line
  - → Check time difference between the two lines
    - » Today the difference should be very small
  - → Check timezone with location of server
- Repeat with next line
- Stop when there is a problem: This server is untrustworthy
  - → This is the originator
    - » Or: Some other criminal, a server with technical problems, …

E-Mail example 1

- A simple E-Mail sent to confirm a hotel reservation
- We will investigate all the headers lines according to their meaning, reliability and hints regarding the sender
  - → Reliability colour code: Reliable, unreliable, indifferent
- This is a MIME alternative message
  - → The content is present as plain text **and** HTML
    - » This is exactly the same, just in different representations
- Note, that there is no "Reply-To" field
  - → This is not necessarily an error or indicative of any problem
  - → It might just mean, that no value was configured for it!

E-Mail example 1: General headers - Source

```
X-Account-Key: account3
X-UIDL: AAACtEMAAAAmX707M4uzyiE1dxV3SVxp
X-Mozilla-Status: 0001
X-Mozilla-Status2: 00800000
X-Mozilla-Keys:
From: "Pension Zlami-Holzer" <office@pension-zlami.at>
To: <Sonntag@fim.uni-linz.ac.at>
Subject: Zimmerreservierung
Date: Tue, 25 Sep 2007 15:02:53 +0200
Message-ID:
    <!&!AAAAAAAAAAAYAAAAAAAAABTBBcgY58xLqEhmw++3c5/CgAAAEAAAANLG+czD6wtAkc
    7bXuE5mIwBAAAAAA==@pension-zlami.at>
MIME-Version: 1.0
Content-Type: multipart/alternative;
    boundary="----=_NextPart_000_0025_01C7FF85.278AF5A0"
X-Mailer: Microsoft Office Outlook 11
Thread-Index: Acf/dGJvYx93FU8yTXC3WWNI6Mg80A==
X-MimeOLE: Produced By Microsoft MimeOLE V6.00.2900.3138
Disposition-Notification-To: "Pension Zlami-Holzer" <office@pension-zlami.at>
X-Scanned-By: MIMEDefang 2.62 on 140.78.3.68
Return-Path: office@pension-zlami.at
X-OriginalArrivalTime: 25 Sep 2007 13:03:41.0037 (UTC) FILETIME=[7EB419D0:01C7FF74]
```

E-Mail example 1: General headers – Discussion (1)

From: "Pension Zlami-Holzer" <office@pension-zlami.at>

The "official" sender (content creator): What was provided by the mail creation program

Optional: Sender-Field (who actually sent the message, e.g. a secretary)

To: Sonntag@fim.uni-linz.ac.at

The "official" recipient: What was provided by the mail creation program

Subject: Zimmerreservierung

What this mail is about (syntax, semantics)

Date: Tue, 25 Sep 2007 15:02:53 +0200

When the E-Mail was sent: 15:02:53 in a time zone 2 hours behind (=east of) GMT

Message-ID:
!&!AAAAAAAAAAYAAAAAAAAABTBBcgY58xLqEhmw++3c5/CgAAAEAAAANLG+czD6wtAkc7bXuE5mIwBAAAAAA==@pension-zlami.at

A (unique) message ID, including the domain of the sender

MIME-Version: 1.0

The format of this mail: MIME (reliable, otherwise it cannot be read!)

Content-Type: multipart/alternative; boundary="----=_NextPart_000_0025_01C7FF85.278AF5A0"

Several parts, which are the same content; boundary is the string (line) separating them

Thread-Index: Acf/dGJvYx93FU8yTXC3WWNI6Mg80A==

Used for identifying replies

Disposition-Notification-To: "Pension Zlami-Holzer" <office@pension-zlami.at>

Request for confirmation to certain address (this can be any address!)

Return-Path: office@pension-zlami.at

Address and route back to originator. Added by "final" transport system based on information provided to it, so not really that reliable at all (Practice: MAIL FROM content during SMTP).

Bounces will be sent to this message!

Mailing lists: From = Message creator, Return-Path = Mailing list manager (hopefully!)

E-Mail example 1: General headers – Discussion (2)

X-Mailer: Microsoft Office Outlook 11
- The sending E-Mail program

X-MimeOLE: Produced By Microsoft MimeOLE V6.00.2900.3138
- Information about the sender's client software
- Seems to be produced by a Microsoft E-Mail program (matches X-Mailer, but still unreliable)

X-Scanned-By: MIMEDefang 2.62 on 140.78.3.68
- Scanned by some antivirus/antispam/… software on a certain computer
  - MIMEDefang: Extracting mail parts and submitting them to various other programs for checking, verification, etc
- May already have been added by sender!

X-OriginalArrivalTime: 25 Sep 2007 13:03:41.0037 (UTC) FILETIME=[7EB419D0:01C7FF74]
- When the mail was submitted to the first E-Mail host
- Note: The "Date" was Tue, 25 Sep 2007 15:02:53 +0200
  - This is the same date (13:03 UTC = 15:03 +0200), and, if both times are really exact, it took 48 seconds for the mail from pressing the "Send" button to be completely received and accepted by the first mail server through SMTP. Should be the same as the last "Received" time!

The following are headers added by the mail client of the recipient, in this case Thunderbird:

X-Account-Key: account3
X-UIDL: AAACtEMAAAAmX707M4uzyiE1dxV3SVxp
X-Mozilla-Status: 0001
X-Mozilla-Status2: 00800000
X-Mozilla-Keys:

E-Mail example 1: Received from

Received: from mail1.edvz.uni-linz.ac.at ([140.78.3.68])
by mail2.fim.uni-linz.ac.at
with Microsoft SMTPSVC(5.0.2195.6713);
Tue, 25 Sep 2007 15:03:40 +0200

Final recipient: mail2.fim (from which it was retrieved by POP) got the mail from mail1.edvz.uni-linz
The IP address of mail1.edvz was 140.78.3.68 (DNS → mail1.edvz.uni-linz.ac.at)
A certain version of Microsoft SMTP was used for receiving
The mail was received at 15:03:40

This data is trusted, because it was added by mail2.fim, our last (=own) server!

Received: from taro.utanet.at (taro.utanet.at [213.90.36.45])
by mail1.edvz.uni-linz.ac.at (8.14.1/8.14.1)
with ESMTP id l8PD3O6H058952
for <Sonntag@fim.uni-linz.ac.at>;
Tue, 25 Sep 2007 15:03:29 +0200 (CEST)
(envelope-from office@pension-zlami.at)

The server mail1.edvz (using a certain version of sendmail) got it from taro.utanet.at at the address 213.90.36.45 (verified by DNS as the same address)
A certain ID was assigned to the message
The message was sent to a specific address
Receipt time is 11 seconds before the next hop → reasonable
Additionally here the envelope from address is added. This is the same as the "From" header

This data is trusted, because all the information matches the one from the next (above) line!
Additionally, we trust also the computer mail1.edvz (and we *know* it came from this one)!

Note: mail1.edvz, if hacked, could have forged all this and any other information except the line above!

E-Mail example 1: Received from

Received: from aki.utanet.at ([213.90.36.56])
by taro.utanet.at with esmtp (Exim 4.60)
(envelope-from <office@pension-zlami.at>) id 1IaA4S-00013g-Hz for Sonntag@fim.uni-linz.ac.at;
Tue, 25 Sep 2007 15:03:24 +0200

taro.utanet got the E-Mail from aki.utanet.at with the (matching) IP address 213.90.36.56

Taro uses the mail software Exim in version 4.6, and assigned it some id

Note: The ID is local for every server, so differences between the hops are **not** significant!

The envelope from address is again provided and matches the header content; recipient is ok too

The time is 5 seconds before the next hop, which is appropriate for servers on big networks (UTA and university of Linz) with a good connection

This data is probably correct, as it matches the next line.

Additionally, both are actual UTA servers, which is a very big ISP (→ probably secure servers)

Received: from dsl-22-119.utaonline.at ([81.189.22.119] helo=Rezeption2)
by aki.utanet.at with esmtp (Exim 4.50)
id 1IaA4S-0002XM-2E for Sonntag@fim.uni-linz.ac.at;
Tue, 25 Sep 2007 15:03:24 +0200

aki.utanet got the mail from dsl-22-119.utaonline.at with the IP address 81.189.22.119

This computer represented itself as "Rezeption2". This is "suspicious", but for dial-up accesses ("dsl"!) it is not unusual, as the computer will present its local hostname, not the (not really meaningful) DNS name associated with its current IP address. Therefore no hint of a modification here.

aki.utanet uses a different version of Exim as taro.utanet (this is slightly suspicious, as an ISP would usually employ the same version on all servers); the recipient address is the correct one

The time is identical to the next hop: For servers within a single ISP OK

This data is probably correct, as it matches the next line.

Again, it is from an actual UTA server

This E-Mail is probably correct according to its information (actually, it is ☺)

E-Mail example 2: General headers

Date: Sat, 13 Jun 2009 17:44:55 -0120

This date is in the future and anyway very different from the actual date, which we can determine from the last (and trusted) "Received" line (see box); Note also the "strange" timezone

From: "Glen Blackwell" <ims@1-by-1.com>

Note that the name and the E-Mail address are not very well matching. Also, there is no webserver behind www.1-by-1.com (not: www.1by1.com!); you are redirected to Google

X-Mailer: The Bat! (v2.00.9) UNREG / CD5BF9353B3B7091

TheBat! Is a genuine mail client, but it is also a default option in a common mass mailer. Based on other information it can be possible to determine whether this is really from TheBat! or not (here: correct or at least a good fake).

X-Priority: 3 (Normal)

Nothing special

Message-ID: <009359377.09683285666384@thebat.net>

Looks like a genuine message ID from TheBat!

To: sonntag@fim.uni-linz.ac.at

Correct destination address

Received:
from mail2.edvz.uni-linz.ac.at ([140.78.3.69])
by mail2.fim.uni-linz.ac.at
with Microsoft SMTPSVC(5.0.2195.6713);
Mon, 12 Feb 2007 19:20:29 +0100

Subject: [SPAM?] [HIGH] One Year written replica watches warranty

The hints in square brackets were added by mail2.edvz!

Return-Path: ims@1-by-1.com

Return path matches the sender address

X-OriginalArrivalTime: 12 Feb 2007 18:20:29.0154 (UTC) FILETIME=[797AD420:01C74ED2]

This differs from the Date header, even though both should be similar as added by the same host! Also note, that this time is specified in UTC and not in a local timezone. FILETIME is a windows file time → Mo, 12 February 2007 18:20:29 UTC (exactly the same)

E-Mail example 2: Received headers

Received: from mail2.edvz.uni-linz.ac.at ([140.78.3.69]) by mail2.fim.uni-linz.ac.at with Microsoft SMTPSVC(5.0.2195.6713); Mon, 12 Feb 2007 19:20:29 +0100

This is trusted information from our own server; it especially provides the actual date.

Received: from smtp.010-101.com ([88.241.70.102])
by mail2.edvz.uni-linz.ac.at (8.13.4/8.13.3) with ESMTP id l1CIK6Xl050036
for <sonntag@fim.uni-linz.ac.at>; Mon, 12 Feb 2007 19:20:13 +0100 (CET)
(envelope-from ims@1-by-1.com)

This is trusted information from a JKU server. The envelope address matches the From header.

The mail was received from the computer with the name "smtp.010-101.com" and the IP address 88.241.70.102. But a reverse lookup returns "dsl88.241-18022.ttnet.net.tr"! This does not match! It actually is a DSL computer, not a mailserver, as claimed (which is unlikely). Especially note that it is from Turkey, where the time zone in February should be +02**0**0 (+0100: DST April-Sept)!

Received: from 69.46.238.251 (HELO iris1.directnic.com)
by fim.uni-linz.ac.at with esmtp (U2M,B,1,)5(H 4A4O,=) id T.3)2<-7MB6SC-P,
for sonntag@fim.uni-linz.ac.at; Sat, 13 Jun 2009 17:44:55 -0120

This information is highly suspect: We got the mail from 88.241.70.102, but this computer called itself "fim.uni-linz.ac.at"! Which is by the way the domain name, and not the name of a server! Additionally, esmtp (the extended SMTP protocol; could be correct) provides a strange version information "U2M,B,1,)5(H 4A4o,=)" and an illegal ID "T.3)2<-7MB6SC-P,"

It also repeats the strange time, which is obviously incorrect: In the future, wrong time zone for the claimed location (which would be the US)

The IP address (69.46.238.251) and the name (iris1.directnic.com) of the source do match.

However, mail originating at iris1.directnic.com has a sender E-Mail of ims@1-by-1.com ???

The real source of the mail is 88.241.70.102, which is probably a zombie!

E-Mail example 2: Spam headers

- Spam-Scanning information might also be faked:
  - → The green information is similar to the one actually added to E-Mails by mail2.edvz.uni-linz.ac.at at the date 12.2.2007
  - → The red information is neither added by mail2.edvz nor mail2.fim!
    - » It must be bogus information added already by the spammer, hoping that anti-spam software looks at the header, thinks that this mail has already been scanned, and will not scan it again

X-Spam: Not detected
X-Spam-Status: HIGH ; 438
X-Spam-Level: **
X-Spam-Report:
BAYES_99,DATE_IN_FUTURE_96_XX,HTML_40_50,HTML_MESSAGE,PYZOR_CHECK,RBL_COMBO_A_2,RBL_COMBO_B_2,RBL_COMBO_C_1,RBL_COMBO_C_2,RCVD_IN_NJABL_DUL,SARE_HTML_URI_LHOST31,SARE_SPEC_REPLICA_OBFU,SARE_SPEC_ROLEX,SARE_SPEC_ROLEX_NOV5A,SARE_SPEC_ROLEX_REP,SARE_WEOFFER,URIBL_AB_SURBL,URIBL_OB_SURBL,URIBL_SBL,URIBL_SC_SURBL,URIBL_WS_SURBL
X-RBL-Warning: Only for compatiblity - see X-Spam-Status + X-Spam-Level
X-Scanned-By: MIMEDefang 2.44 ↖ Spelling error in configuration of mail2.edvz!

Gathering further information: nslookup / dig

- Comparing the IP address to the name: nslookup / dig
  - → With these tools you can retrieve the associations between domain names and IP addresses
  - → Note that these should be run from "secure" systems/networks, as various techniques exist to alter the output
    - » E.g. manipulating the HOSTS file, DNS cache poisoning, …
    - » Internet: E.g. http://remote.12dt.com/
  - → Example: 88.241.70.102 ⇔ smtp.010-101.com
    - » dig smtp.010-101.com → 216.187.77.20
    - » dig –x 88.241.70.102 → dsl88.241-18022.ttnet.net.tr.
- Tools are based on the DNS system and extremely reliable
  - → If your environment is, that is!
- Always query completely and in both directions
  - → Some legitimate hosts have several IP addresses or one address several names

Gathering further information: WHOIS

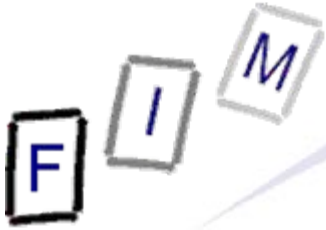
- Obtaining owner data for domain name and IP addresses
  - → Based on the DNS registrars
  - → Extremely reliable, if your connection to them is reliable
    - » But note, that the information is not necessarily helpful!
      - – The name/address might be clearly invented and incorrect
      - – The name/address might be a proxy (for anonymization)
- Attention: Some information is only available regionally
  - → RIPE, ARIN, APNIC, AfriNIC, LACNIC, ccTLDs, …
    - » Check with other registries if no answer is found
  - → Usually queried through websites; programs available too
    - » http://www.whois.net/
    - » http://www.ripe.net/fcgi-bin/whois

Further checking

- Check whether the IP addresses of the mail servers are actually such ones
  - → DNS check for the MX (Mail eXchanger) records!
  - → If it was sent by a certain computer, the next (=first) Received line should be the MX for this domain
    - » Note: Usually the MX is for sending TO, not FROM this domain, so this might be legitimately different!
      - – This is rare, however
- Attention: Whois, DNS etc. only provide current data!
  - → If the mail is already "old" (several days to weeks/years), the data might be different
    - » Assessed as incorrect, although it was legitimate then!

Conclusions

- SMTP is an extremely unsecure protocol regarding content
  - → Anyone can fake any message without problems
  - → Use el. signatures, encryption etc. if authenticity, secrecy, … is needed for communication!
- Almost all data in an E-Mail is susceptible to falsification
  - → Only what has been added by a trusted host is secure
    - » As long as only trusted hosts handled message afterwards …
- Through the "Received" header lines backtracing is possible
  - → Note, that this usually only leads to a certain computer, not to a single individual!
  - → Through the ISP and the time field (unreliable!) the person can then be identified
    - » If log files (still) exist, the ISP is willing, …
    - » Data retention will help in these cases
      - – But: Anonymous remailers, …



Questions?

Thank you for your attention!

Links

- Lucke, K.: Reading Email Headers
  http://www.stopspam.org/email/headers.html
- Hochstein, T.: E-Mail-Header lesen und verstehen
  http://th-h.de/faq/headerfaq.php
- Sam Spade
  http://samspade.org/
- RFC 2821: Simple Mail Transfer Protocol
  http://www.faqs.org/rfcs/rfc2821.html
- RFC 2822: Internet Message Format
  http://www.faqs.org/rfcs/rfc2822.html
- Baldwin, L., Schultz, J.: Extreme IP Backtracing
  http://www.jaeson.net/backtracing/Extreme_IP_Backtracing.v1337.ppt