

Fallstricke bei der Inhaltsanalyse von Mails

Beispiele, Ursachen, Lösungsmöglichkeiten

Steffen Ullrich, genua GmbH
Konferenz Sicherheit 2018



- Umfeld der Arbeit
zur Person und Firma
Mail als Angriffsvektor
was ist der Semantic Gap
Einführung in MIME
- Semantic Gap bei Mail/MIME
Untersuchungsgegenstand und -methode
vielfältige Beispiele
Ursachen, Lösungsansätze



- Anbieter von Sicherheit durch Separation:
Firewalls auf Transport- und Applikationsebene
VPN, Fernwartung
mikrokernbasierte Separation: Dioden, Laptops
Speziallösungen, Spezialprotokolle, ...
- über 230 Mitarbeiter in Kirchheim bei
München, Berlin, Köln, Stuttgart
- großer Unterstützer OpenBSD
Sponsor dieser Konferenz
- Jobs, Praktika, Werksstudent,
Bachelor-, Master- und Doktorarbeiten



- seit 2001 bei genua GmbH
- Entwicklung von Firewalls (genugate)
Spezialität Applikationsprotokolle
- aktuelle Resultate aus Forschungsprojekt
bzgl. gezielter Angriffe bei Mail und Web
Partner Uni Braunschweig, Uni Erlangen, gefördert vom BMBF
www.apt-sweeper.de



- Mail einer der Hauptangriffswege
 - Credential-Phishing
Abgreifen Zugangsdaten Paypal, Bank, Mail, ...
Links in Mail oder (PDF-)Attachments
attache HTML-Dokumente, ...
 - Malware
Links zu Malware, meist aber Attachments
Office-Dokumente mit Macros u.a.
Archive mit *.js, *.exe, *.url ... - teilweise verschlüsselt
direkt attache *.exe, *.jar, *.scr, ...
Spoofing Dateiendungen
doppelte Endungen *.pdf.exe
über RLO `h\u202efig.exe -> hexe.gif`



- übliche Verteidigungsstrategien
 - im Umfeld der Mail
 - Schulung der Mitarbeiter
 - Softwareupdates
 - Antispam: Blacklists, Greylisting, SPF
 - durch inhaltliche Analyse der Mail
 - Spamdetektion via Bayes-Filter o.ä
 - DKIM, DMARC
 - Reputationsanalyse von Links
 - Blockieren spezifischer Dateiendungen
 - Antivirus
 - Heuristiken, z.B. spezielle Dateitypen in Archiv



- Viel Forschung zur Analyse der Inhalte
 - stylometrische Analysen
 - Extraktion von Links und Vergleich mit beliebten Phishing-Zielen
 - Vielfalt von Analyseansätzen für attachte Dateien (Office, Exe, PDF, HTML ...)
- Forscher gehen i.A. davon aus, dass Extraktion der Inhalte für Analyse trivial, aber ...



- Problem mit Inhaltsanalyse
 - Analysesystem und Endsystem **müssen Daten auf gleiche Weise interpretieren**, sonst evtl. Bypass der Analyse möglich
 - Analyse checkt die falschen Links, übergibt falschen Inhalt an AV, vermutet den falschen Dateityp ...
 - Standards und Umsetzung mangelhaft: Systeme verhalten sich unterschiedlich
 - Semantic Gap, Interpretationslücke
 - exaktes Endsystem (und dessen Verhalten) sind im Analysesystem unbekannt



- weitere Arbeiten zum Semantic Gap
 - Umgehungen auf Layer 3..4
überlappende TCP-Segmente etc – alt aber präsent
Evading Deep Inspection for Fun and Profit – BH 2013
 - Umgehung auf Layer 7 – HTTP
HTTP-Evader Testsuite für Firewalls – 2015 (selbst)
Nutzung HTTP 0.9, seltene Encodings, falsche CRC ...
praktisch alle Firewalls deutlich betroffen
 - Dokumente: Polyglots, Mimikris ...
GIFJAR, PDF+ZIP, GZIP+ZIP, ...
Funky File Formats – 31C3 2014
- Mein Beitrag: Semantic Gap bei Mail/MIME



- ursprüngliche Mails stark limitiert
nur ASCII, max. Zeilenlänge 1000, keine Struktur
- 1980: inline Attachments mit uuencode
analoge Versuche mit binhex für Mac (1984)

```
From: me  
To: you  
Subject: Virus
```

Hallo, hier das versprochene Virus-Sample.

```
begin 664 eicar.zip  
M4$L#!!0``@`(`!*CDD\SU%H1@````$0````)````96EC87(N8VJMB#U5PQ0  
M=7` ,B#:) "8B* ,#71" (C3- '=VUC2057'U='8 ,T@T.<?1S<0QRT77T" _$, \PP*  
M#=8-<0T.T77S]'%55/'0]M`"``%!+`0(4`Q0``@`(`!*CDD\SU%H1@````$0`  
M````)````````````````````"V@0````!E:6-A<BYC;VU02P4&``````$``0`W````  
&;0``````  
`  
end
```



- 1993 RFC 1521+1522
- 1996 überarbeitet in RFC 2045..2047
MIME: Multipurpose Internet Mail Extensions
 - Abbildung moderner Features innerhalb der ursprünglichen Limitierungen
 - RFC 2045: binäre Inhalte als ASCII kodieren
 - RFC 2046: strukturierte Mails (Attachments)
 - RFC 2047: Umlaute in Header (Subject, From, ...)
- 1997 Ergänzungen
 - RFC 2183: verschiedene Arten von Attachments
 - RFC 2231: lange Dateinamen incl. Umlaute



```
From: me@example.com  
To: you@example.com  
Subject: Viele =?UTF-8?Q?Gr=C3=BC=C3=9Fe?=  
Content-type: multipart/mixed;  
boundary=trenner
```

Viele Grüße

```
--trenner  
Content-type: text/plain; charset=UTF-8  
Content-Transfer-Encoding: quoted-printable
```

Viele Gr=C3=BC=C3=9Fe von mir.

Grüße

```
--trenner  
Content-type: application/octet-stream;  
name=test.txt
```

```
Content-Disposition: attachment;  
filename*0*=utf-8''%c3%bcbel.e;  
filename*1=xe
```

übel.exe

```
Content-Transfer-Encoding: base64
```

```
TVqQ...VGhpcyBwcm9ncmFtIGNhbm5vdCBi...  
--trenner--
```

MZ...This program cannot be run in DOS mode...

RFC 2046

Unterteilung, Attachments
multipart/mixed
multipart/alternative, ...
Content-type: ...; name=

RFC 2045

binäre Inhalte → ASCII
base64, quoted-printable
Content-Type, Charset

RFC 2047

Umlaute etc im Header
base64, quoted-printable
Charset

RFC 2183

Content-Disposition
Inline, Attachment, Filename

RFC 2231

Umlaute etc in Parameter
Charset, Sprache
Hex-Encoding
Aufspaltung langer Parameter



- Umfeld der Arbeit
zur Person und Firma
Mail als Angriffsvektor
was ist der Semantic Gap
Einführung in MIME
- **Semantic Gap bei Mail/MIME**
Untersuchungsgegenstand und -methode
vielfältige Beispiele
Ursachen, Lösungsansätze



Analyse soll Virus blockieren



Analyse soll Dateiendung blockieren



- **Mailclients**
Thunderbird, Outlook, Windows Mail, Apple Mail
mutt, Roundcube, kmail
- **Analysesysteme**
kommerzielle Firewalls
IDS Snort, Suricata
Antivirus ClamAV, kommerzieller AV
Mailfilter amavisd-new
- **Bibliotheken**
Perl MIME::Tools (Spamassasin, Amavisd, Mimedefang)
Python email.parser
Go mime
- **Webbasierte Mail+AV: AOL, Outlook, GMX**



- Was wurde getestet
 - Bypass Analyse mit EICAR.zip bzw. EICAR.txt in IDS Tests
 - Bypass Filterung Dateieindung novirus.zip
- Wie wurde getestet
 - primär Blackboxanalyse mittels SMTP (FW, Mailfilter), PCAP (IDS) bzw. Dateien (AV, Bibliotheken)
 - teilweise Analyse Sourcecode, um neue Ideen für Bypass zu finden
 - Testsuite mit automatischer Generierung von Testmails mit variablem Payload



- folgende Beispiele sind nur exemplarisch
 - zeigen nicht einzelne esoterische Lücke sondern Breite an Problemen mit unterschiedlichsten Teilen der MIME-Standards
 - kein Anspruch auf Vollständigkeit
mehr Probleme als gezeigt wurden gefunden
viele Ideen noch vorhanden aber nicht getestet
nur wenige Systeme wurden getestet



```
From: me@example.com
To: you@example.com
Subject: Viele =?UTF-8?Q?Gr=C3=BC=C3=9Fe?=
Content-type: multipart/mixed;
  boundary=trenner
```

--trenner

```
Content-type: text/plain; charset=UTF-8
Content-Transfer-Encoding: quoted-printable
```

Viele Gr=C3=BC=C3=9Fe von mir.

--trenner

```
Content-type: application/octet-stream;
  name=test.txt
```

```
Content-Disposition: attachment;
  filename*0*=utf-8''%c3%bcbel.e;
  filename*1=xe
```

```
Content-Transfer-Encoding: base64
```

TVqQ...VGhpcyBwcm9ncmFtIGNhbm5vdCBi...

--trenner--

RFC 2046

Unterteilung, Attachments
multipart/mixed
multipart/alternative, ...
Content-type: ...; name=



Preamble

```
Content-type: multipart/mixed;  
boundary=bb;
```

```
This is a multipart message
```

```
--bb
```

```
Content-type: application/zip;  
name=virus.zip  
Content-Transfer-Encoding: base64
```

Daten

base64
virus.zip

```
UEsDBAAoAAAAAAO6Tn0m2hH8nEwAAAB....  
AQABAE4AAABVAAAAAAAAA=
```

```
--bb--
```



```
Content-type: multipart/mixed;  
  boundary=bb; boundary=##
```

Preamble

This is a multipart message

--bb

```
Content-type: application/zip;  
  name=virus.zip  
Content-Transfer-Encoding: base64
```

Daten

base64
virus.zip

--##

```
UEsDBAoAAAAAAAAA06Tn0m2hH8nEwAAAB....  
AQABAE4AAABVAAAAAAAAA=
```

--##--

--bb--

Preamble

Daten
Müll

wird i.A. ignoriert von base64 Parsern,
da ungültige Zeichen in base64

bb, ##

Thunderbird, Win10 Mail

Python, -Ge-

Fail: **Snort, Suricata, Amavisd, FW2**

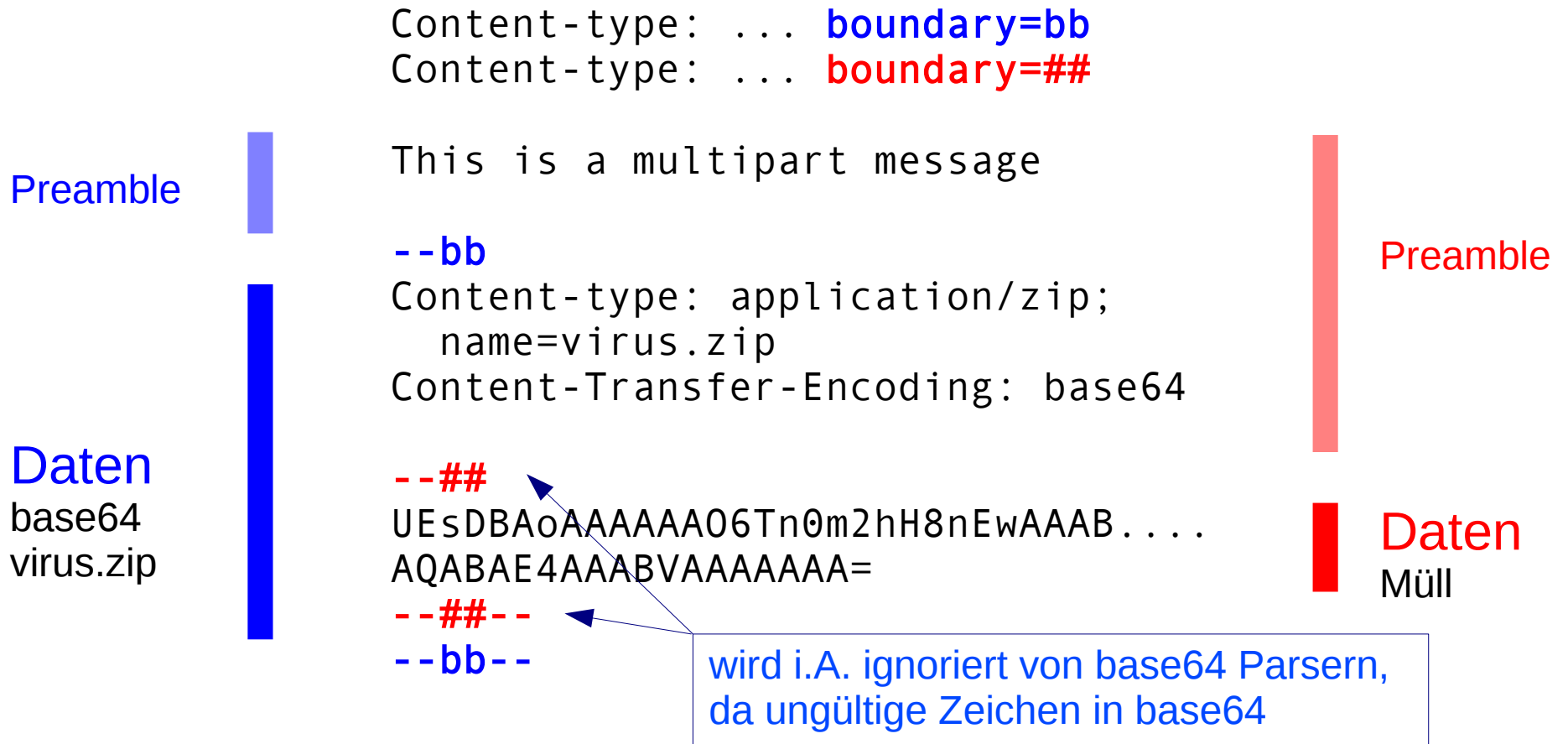
##, bb

Outlook, Apple Mail

Perl, -Ge-

Fail: **Snort, Suricata, ClamAV, FW1**





bb, ##
Thunderbird, Win10 Mail, Apple Mail
Perl, Python, Go
Fail: **Suricata, Amavisd**

##, bb
Outlook
Fail: **Snort, Amavisd ClamAV, FW2**



```
From: me@example.com  
To: you@example.com  
Subject: Viele =?UTF-8?Q?Gr=C3=BC=C3=9Fe?=  
Content-type: multipart/mixed;  
    boundary=trenner
```

```
--trenner  
Content-type: text/plain; charset=UTF-8  
Content-Transfer-Encoding: quoted-printable
```

Viele Gr=C3=BC=C3=9Fe von mir.

```
--trenner  
Content-type: application/octet-stream;  
    name=test.txt  
Content-Disposition: attachment;  
    filename*0*=utf-8''%c3%bcbele.e;  
    filename*1=xe  
Content-Transfer-Encoding: base64
```

TVqQ...VGhpcyBwcm9ncmFtIGNhbm5vdCBi...

```
--trenner--
```

MZ...This program cannot be run in DOS mode...

RFC 2045

binäre Inhalte → ASCII
base64, quoted-printable
Content-Type, Charset



Content-Transfer-Encoding: **base64**

Content-Transfer-Encoding: **quoted-printable**

YmFzZTY0IGVuY29kZWQgZGF0YQo=

CTE: base64, quoted-printable
Body: base64

Thunderbird
Outlook
Win10 Mail
Apple Mail
Perl
Python
Go
Fail: Snort

QP
mutt
Roundcube
Fail: FW1
Fail: FW2
Fail: Amavisd

quoted-printable, base64
QP

Thunderbird
Outlook
Win10 Mail
Apple Mail
Perl
Python
Go

base64
mutt
Roundcube
Fail: FW2
Fail: Amavisd
Fail: AV

Yahoo 2014, AOL 2015: GUI base64, AV qp



Base64 kodiert jeweils 3 Byte binär zu 4 Byte ASCII: A-Za-z0-9/+. Aufgefüllt wird mit '=', d.h. dieses kann eigentlich nur am Ende auftreten.

Content-Transfer-Encoding: base64

Vkk= ← VI
U1VTIQ== ← RUS!

VIRUS!
Thunderbird
Apple Mail
Roundcube

Fail
FW1
Snort, Suricata
AV, ClamAV
Amavisd
Perl, Python, Go



yet another encoding – nicht standardisiert – benutzt in Newsgruppen
unterstützt von Thunderbird auch in Mails – **Fail: alle**

```
From: me
To: you
Subject: [0] multipart-basic-singlepart-x-yencode eicar.zip 2017-04-09 21:35
Mime-Version: 1.0
Message-Id: <86e7be25.69ff.58ea8cfa@example.com>
Content-type: multipart/mixed; boundary=foobar
Content-Length: 610

--foobar
Content-type: text/plain
Content-Transfer-Encoding: x-yencode

=ybegin line=128 size=51 name=file.bin
klmn3Z[\]^g_`abcJJcba`_g^]\[ZJgJg<82><83>J}<99><97><8f>J<97><99><9c><8f>J<9e><8f>ç<9e>4^M
=yend size=51
--foobar
Content-type: application/octet-stream
Content-Disposition: attachment; filename="eicar.zip"
Content-Transfer-Encoding: x-yencode

=ybegin line=128 size=186 name=file.bin
zu-.>*,*2*;t,sfù{<92>p***n***3***<8f><93><8d><8b><9c>X<8d><99><97>μZ^_<81>6z<9f><9a>6²`³3²´
^<9d><9b>6<9c>û<9f>^^5^[6^]647=@7<9b>78û<9f>^]^^<9b>^?~^[ú=`ú,*zu+,>->*,*2*;t,s^M
fù{<92>p***n***3*****à«****<8f><93><8d><8b><9c>X<8d><99><97>zu/0*****+**a***<97>*****
=yend size=186
--foobar--
```



```
From: me@example.com
To: you@example.com
Subject: Viele =?UTF-8?Q?Gr=C3=BC=C3=9Fe?=
Content-type: multipart/mixed;
  boundary=trenner
```

```
--trenner
Content-type: text/plain; charset=UTF-8
Content-Transfer-Encoding: quoted-printable
```

Viele Gr=C3=BC=C3=9Fe von mir.

```
--trenner
Content-type: application/octet-stream;
  name=test.txt
Content-Disposition: attachment;
  filename*0*=utf-8' '%c3%bcbel.e;
  filename*1=xe
Content-Transfer-Encoding: base64
```

übel.exe

```
TVqQ...VGhpcyBwcm9ncmFtIGNhbm5vdCBi...
--trenner--
```

RFC 2231
Umlaute etc in Parameter



- RFC 2231 Encoding von Parametern
(RFC 2047 Encoding in Subject...)
 - Splitting für lange Filenamen
name*0=foo; name*1=.z; name*2=ip -> foo.zip
 - Character-Encoding und Sprache
name*=utf-8'de'%c3%bcbe1.exe -> übe1.exe
 - kombinierbar
name*0*=' 'foo.z; name*1*=%69%70 -> foo.zip
- kann zusätzlich sein, soll Präferenz haben
name=f.txt; name*0=evil.e; name*1=xe -> evil.exe



- eigentlich für Filenamen gedacht
OK: Thunderbird, Win10 Mail, Python, Perl, Go ...
Unsupported: Outlook
Fail: Suricata (Snort nicht getestet)
- geht auch mit anderen Parametern:
Content-type: multipart/mixed;
boundary*=' 'foo%62ar
--foobar
- Ok: Thunderbird, Apple Mail, Python, mutt, kmail
Fail: Snort, Suricata, Amavisd, AV
Perl, Go



- RFC 822 und später erlauben Kommentare

```
MIME-Version: 1.0 (no (need to) comment)
From: Hans <hans@example.com> (Yippie!)
```

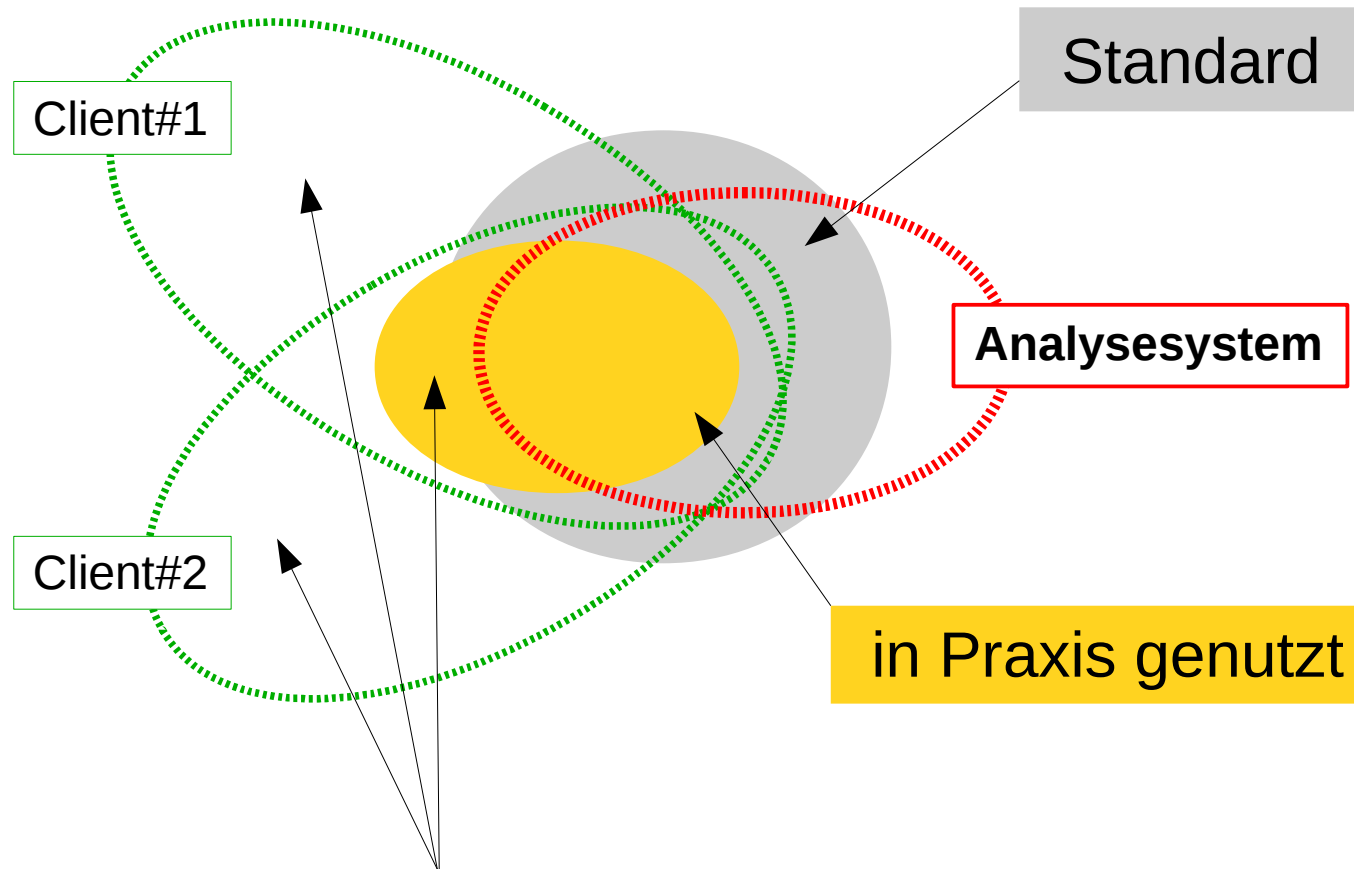
- sinnloses Feature, Outlook, Win10 Mail und Apple Mail erlauben es auch bei Boundary

```
Content-type: multipart/mixed;
    boundary=foo(Y)bar
```

```
--foobar
```

Fail: Snort, Suricata, FW2, Amavisd, AV
Perl, Python, Go





Interpretationslücke

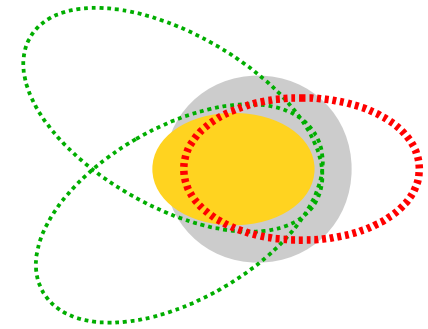
(Semantic Gap)

Endsystem interpretiert Daten anders als Analysesystem.
Bypass der Analyse ist möglich.



Kaputte Standards

unnötig **komplex**, **flexibel** und **erweiterbar**
widersprüchliche Aussagen möglich
teilweise Widersprüche zwischen Standards
keine sinnvolle Fehlerbehandlung festgelegt
scheinbar einfach: **human readable**



Kaputte Protokollparser

Einige Fehler im Protokoll werden teilweise explizit toleriert,
die meisten aber als **Nebeneffekt** der Implementation interpretiert.
Grundannahme ist, dass Gegenstelle protokollkonform arbeitet, d.h.
kaum explizite Fehlerbehandlung.

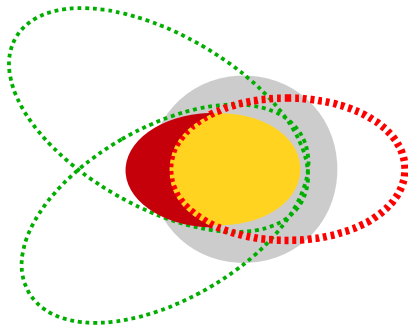
Fehlendes Bewusstsein bei Entwicklern von Analyse

Angriffe auf dieser Ebene werden **nicht erwartet**: genutzte
(Standard-)Bibliotheken **zu tolerant und ignorant**
Schlampige Parser, auf **Geschwindigkeit statt Korrektheit** getrimmt.



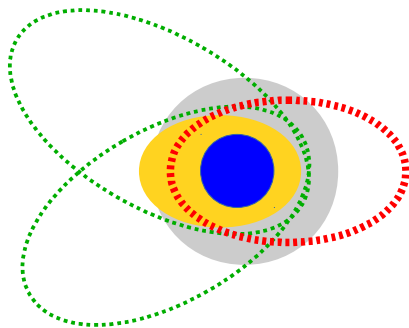
- Realität ist Vielfalt von mailerzeugenden Applikationen mit Vielfalt von Auslegungen oder Ignoranz der Standards
- Keine dominanten Player
Keine dominanten MIME-Stacks
Keine Bereinigung in Aussicht
- Funktioniert teilweise nur, weil Sender und Empfänger gleiche Fehler machen
Und funktioniert nicht mit anderen Empfängern
- Lösung sollte nichts kaputt machen





Passive Analyse

Keine Modifikation, sondern nur Durchlassen oder Blockieren: **sämtliche bekannt Interpretationsvarianten durchprobieren.** Unbekannte Varianten führen weiterhin zu Interpretationslücke.



Aktive Analyse

Modifikation der Inhalte ermöglicht einheitliche Interpretation durch Reduktion auf **Kernprotokoll** → **Interpretationshoheit**

nicht möglich in IDS, AV



- Entscheidung basierend auf originaler Mail
 - potentiell ambivalente Aussagen
widersprüchliche Aussagen oder Ambivalenz durch Clientbugs oder nicht implementierter Features
 - der Analyse unbekannte Features
seltene Features der Standards, seltene Standards und nicht standardisierte Erweiterungen
 - Kaputt, aber deshalb gleich gefährlich?
Kaputte Mails sind leider Realität.
 - Problem: **Overblocking** oder **Bypass**
in Grenzfällen keine klare Entscheidung möglich
im Zweifelsfalle Durchlassen (evtl. Bypass) oder Blockieren (evtl. Overblocking)



- Mail kann durch Analyse angepasst werden
 - Reduktion auf Kernfeatures, die von allen Clients gleich verstanden werden
 - Klarstellung potentiell ambivalenter Aussagen
 - Umwandlung nicht-standardisierter Erweiterungen
 - nur in aktiven Mail-Gateways umsetzbar, nicht in IDS, Antivirus und paket-basierten (NG)FW
 - Problem: Mail kann (etwas) brechen
 - Modifikation invalidiert evtl. Signaturen (PGP, DKIM...)
 - kann eigentlich kaputte aber fein abgestimmte Sender-Empfänger-Beziehungen brechen



- es ist kaputt
 - unreparierbar
 - wir müssen damit leben
- Lösungsmöglichkeiten
 - passiv: Overblocking vs. Bypass
 - aktiv: nicht in IDS, (NG)FW, AV
 - Bewusstsein für Probleme erhöhen





Content-Transfer-Encoding: <cr lf>
<space><cr lf>
<space>**base64**

YmFzZTY0IGVuY29kZWQgZGF0YQo=

base64

Thunderbird
Outlook
Win10 Mail
Apple Mail
Perl
Go

Fail

Python
ClamAV
AV
FW1



Viele Grüße

From: me@example.com
To: you@example.com
Subject: **Viele** =?UTF-8?Q?Gr=C3=BC=C3=9Fe?=
Content-type: multipart/mixed;
 boundary=trenner

--trenner
Content-type: text/plain; charset=UTF-8
Content-Transfer-Encoding: quoted-printable

Viele Gr=C3=BC=C3=9Fe von mir.

--trenner
Content-type: application/octet-stream;
 name=test.txt
Content-Disposition: attachment;
 filename*0*=utf-8''%c3%bcbel.e;
 filename*1=xe
Content-Transfer-Encoding: base64

TVqQ...VGhpcyBwcm9ncmFtIGNhbm5vdCBi...
--trenner--

RFC 2047

Umlaute etc im Header

=?Q?qp-encoded?=
=?B?base64-encoded?=-



- RFC 2047 erlaubt Umlaute in MIME-Header
Gedacht für Subject, From, ...
- an anderen Stellen nutzbar (aber falsch)
Häufiger Fehler. Richtig wäre RFC 2231.

```
Content-Disposition: attachment;  
  filename="=?us-ascii?B?ZmlsZS56aXA=?="
```

OK(file.zip): Thunderbird, Outlook, Win10 Mail,
Apple Mail, Roundcube, Python, Perl

Fail: FW1, Suricata, Amavisd, Go



- inline uuencode geht noch immer bei Outlook, Roundcube
Fail: Suricata, FW2

From: me
To: you
Subject: Virus

Hallo, hier das versprochene Virus-Sample.

```
begin 664 eicar.zip
M4$L#!!!0``@`(`!*CDD\SU%H1@```$0````)````96EC87(N8VJMBS#U5PQ0
M=7`,`B#:) "8B*,#71"(C3-'=VUC2057'U='8,T@T.<?1S<0QRT77T"_$_,\PP*
M#=8-<0T.T77S]'%55/'0]M```%!+`0(4`Q0``@`(`!*CDD\SU%H1@```$0`
M```)````````````````````"V@0````!E:6-A<BYC;VU02P4&``````$``0`W````
&;0``````
`
end
```



Thunderbird 52.7.0
Outlook von Office 365
Win10 Mail 17.9126.21535.0
Apple Mail 11.2 (3445.5.20)

FW1

FW2

IDS Snort 2.9.11.1

IDS Suricata 4.0.4

AV

ClamAV 0.99.4/24466/Tue Apr 10 04:24:27 2018

Amavisd amavisd-new-2.11.0 (20160426)

Perl MIME::Tools 5.507

Python 3.6.3 email.parser

Golang 1.10.1 mime

