Mark Martinec
Institut “Jožef Stefan”, Slovenia

Filtriranje e-pošte, Amavis in SpamAssassin

Vsebina

- kaj sta Amavis in SpamAssassin
- umestitev filtrov v sistem e-pošte
- uporabljeni filtrirni mehanizmi
- varovanje prenosa in avtentičnost sporočil
- novosti pri obeh programih
- nadzor delovanja, Elasticsearch, Kibana

Kaj je Apache SpamAssassin ?

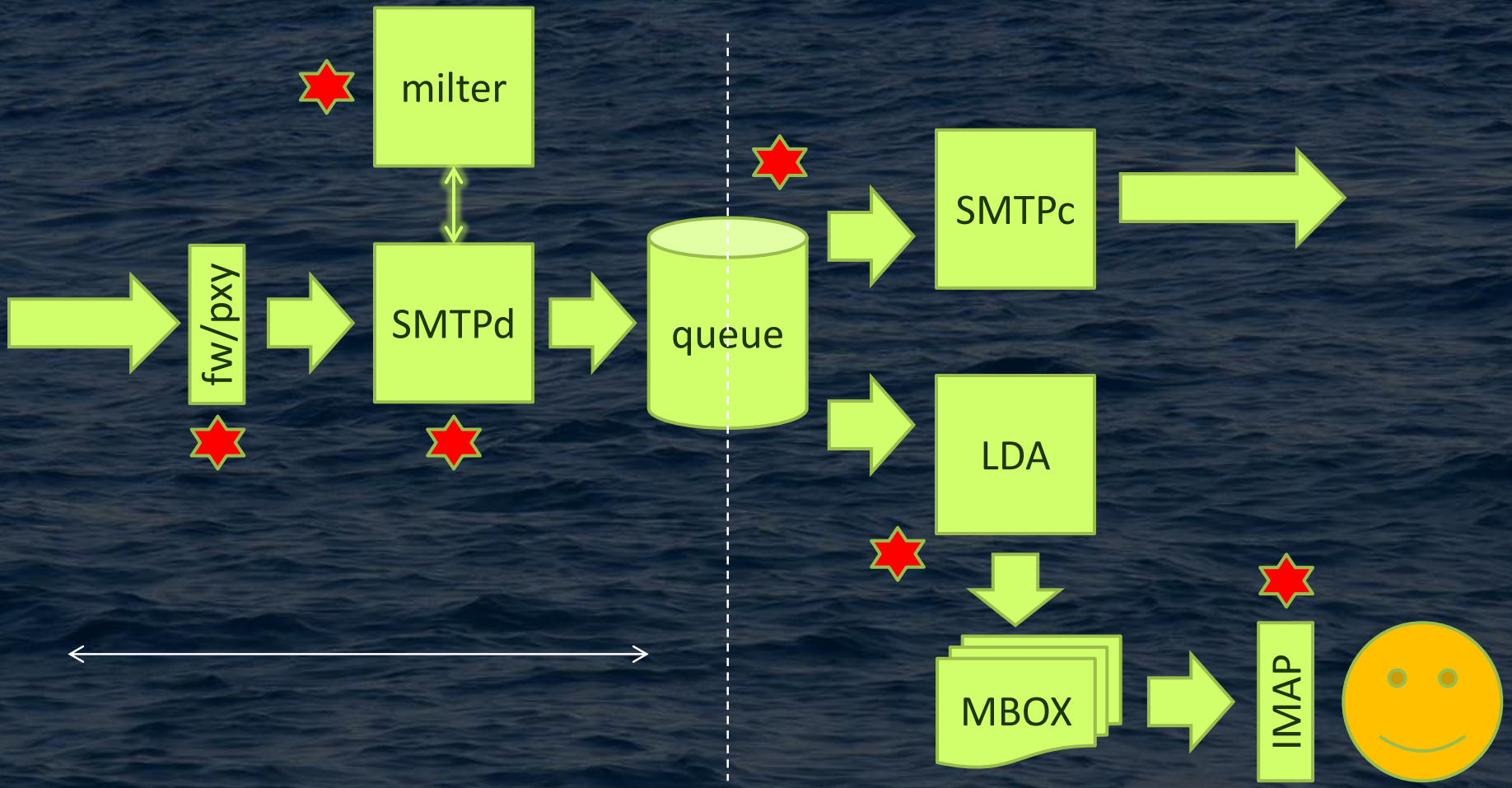
- projekt ASF (Apache Software Foundation)
- knjižnica Perl modulov in skupek pravil
- klasifikacija sporočil (zvezna ocena ham vs. spam)
- pomožna orodja (učenje, ažuriranje pravil)
- preprost vmesnik (spamc/spamd, spamassassin)
- odprtokodna licenca Apache

Kaj je Amavis ?

- vmesnik med MTA in klasifikatorji za viruse & spam
- standardni protokoli (SMTP, LMTP, ali milter)
- ugotavlja prepovedane vsebine in sintakso sporočil
- prepusti / označi / zavrne / zavrže / karantena
- DKIM: podpisuje in preverja (rezultate posreduje k SA)
- nadzor: SNMP, strukturirani dnevnik, 0MQ status
- Perl (hitre operacije z velikimi nizi, varnost)

- odprtokodna licenca GPLv2
- dolga tradicija: (1997) 2002—2015

Umestitev filtrov v sistem e-pošte



Kam umestiti filter

- blizu vstopne točke več informacije, manj časa (TCP fingerprinting seje, časovne specifičnosti, pravilnost SMTP ukazov, src-dst naslov in vrata)
- po sprejetju: več časa, nezmožnost zavrnitve
- veriženje filtrov: pokvečeno avtomatsko učenje

- prva fronta: DNSBL, SMTP nepravilnosti
- druga fronta: sinergija mnogih mehanizmov
- pri uporabniku: lokalna protivirusna zaščita

Zavrni ali dostavi | tiho zavreči

- preprečiti nepojasnjeno izgubo sporočil
  (obvestiti pravega pošiljatelja ali prejemnika)
- preprečiti sipanje (backscatter) – ekviv. DDoS

- BCP: filtriranje pred potrditvijo sprejema
  (med trajanjem SMTP seje)
- nemška in švedska zakonodaja
  http://www.heise.de/ct/artikel/Strafbares-Filtern-289128.html

- rezervni MX: enako filtriranje in seznam uporabn.

Preverjati tudi odhodno pošto

- obramba pred zlorabljenim računalniki ali računi domačih uporabnikov
- dragocen vir čistih sporočil za avtomatsko učenje
- za zagotavljanje istovetnosti (DKIM, SPF) morajo uporabniki za odpošiljanje uporabiti strežnik svoje domene (ne npr. svojega ISP)
- pen pals, TxRep – navezava sporočil na predhodno konverzacijo

Pen pals

- hrani podatke o odhodni pošti (Redis ali SQL)
- pri sprejemu: ali je lokalni prejemnik predhodno komuniciral s tokratnim pošiljateljem
- In-Reply-To ali References:

  ujemanje s predhodnim Message-ID

  (poštni seznami)

Amavis mehanizmi

- blokade: MIME (Content-Type, Content-Disposition.filename), file(1) utility, ime v arhivu
- pen pals (Redis, SQL)
- ugled naslova IP (Redis, SQL)
- OS fingeprinting (p0f) (zlorabljeni Windows računalniki)
- DKIM ugled domene (whitelisting) / podpisovanje
- bounce killer (proti posledicam sipanja)
- protivirusni programi
- SpamAssassin / CRM114 / zunanji filtri

SpamAssassin mehanizmi

- pravila - regularni izrazi (glava, telo sporočila, URI)
- DNS poizvedbe: DNSxL, URI-DNSBL, AskDNS
- DCC, Razor, Pyzor
- klasifikator *naivni bayes*: avtomatsko učenje (redis)
- DKIM, SPF (istovetnost, ugled domene, whitelisting)
- AWL > TxRep vtičnik (ugled pošiljatelja, učenje, SQL)
- slike, priponke: ImageInfo, PDFInfo
- jezik, geolokacija (TextCat, RelayCountry, ASN, URILocalBL)
- wh/bl-listing (DKIM, SPF, Received, From, To, Subject)
- pravilnost/konsistentnost (MIMEEval, RelayEval, HTMLEval, HTTPSMismatch, FreeMail)

Novosti SpamAssassin 3.4.1

- TxRep vtičnik (namesto AWL) ugled IP, domen
- normalize_charset (UTF-8)
- SHA1 zmletek vseh priponk –> bayes
- dkim_minimum_key_bits n (privzeto: 1024)
- link-local IPv6 address (dns, redis, spamd)

  npr. dns_server [fe80::1%lo0]:53

Novosti Amavis

- 2.10.0: podpora internacionalizaciji poštnih sporočil in naslovov (SMTPUTF8, IDNA, UTF-8)

  RFC 6530 .. 6533

Google, Postfix, Momentum (Message Systems)

Žan.Draščček@križišče.si, 测试 @ 测试 . 测试,

δοκιμή @ παράδειγμα.δοκιμή

- 2.10.2: SHA1 zmletki vseh priponk (skupaj s SpamAssassin 3.4.1 klasifikatorjem bayes)
- HAProxy (PROXY protocol V 1) – amavisd kot distribuiran proxy

DMARC: Domain-based Message Authentication, Reporting, and Conformance

- kombinira DKIM + SPF + poročanje
- RFC 7489, marec 2015, informativna kategorija

```
_dmarc.nosuch TXT "v=DMARC1; p=reject"

_dmarc  TXT
   "v=DMARC1;  p=none; rua=mailto:mailauth-reports@ijs.si"
```

pošiljateljeva domena določa, kaj naj prejemnik stori s prejetim neavtentičnim sporočilom, povratna informacija (zlorabe)

Promet med poštnimi strežniki

od vseh dohodnih sporočil na IJS (marec 2015):

- čisto (ham):

60 % TLS (16 % IPv6, cca. 100 ASN)

- blokirano (spam, virus, malware):

11 % TLS (2 % IPv6)

Transport Layer Security (TLS)

- oportunistični TLS (šifrirano, m.i.t.m.)
- (perfect) forward secrecy
- z overjenimi certifikati CA (med partnerji)

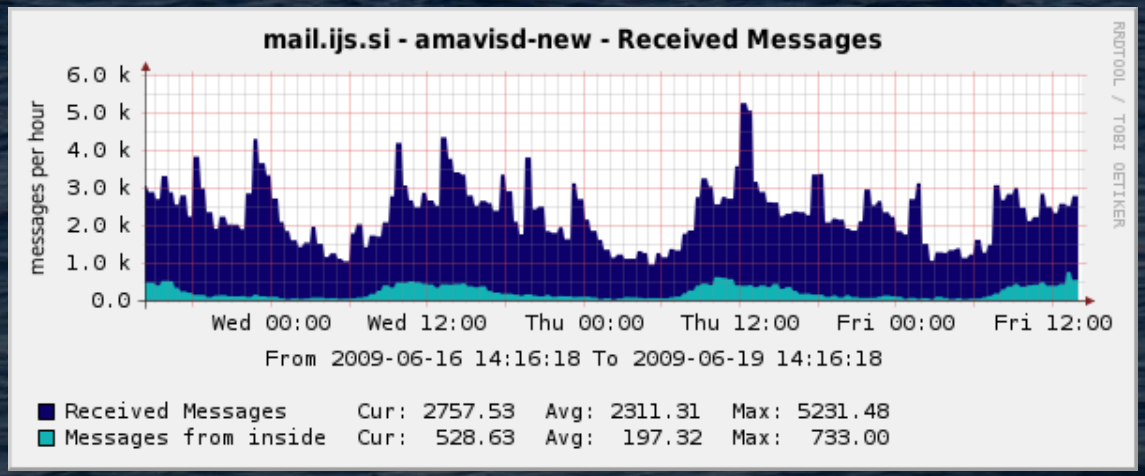
- DNS-Based Authentication of Named Entities (DANE) – pogoj je DNSSEC, TLSA RR poveže certifikat z domeno (nemške vladne institucije)

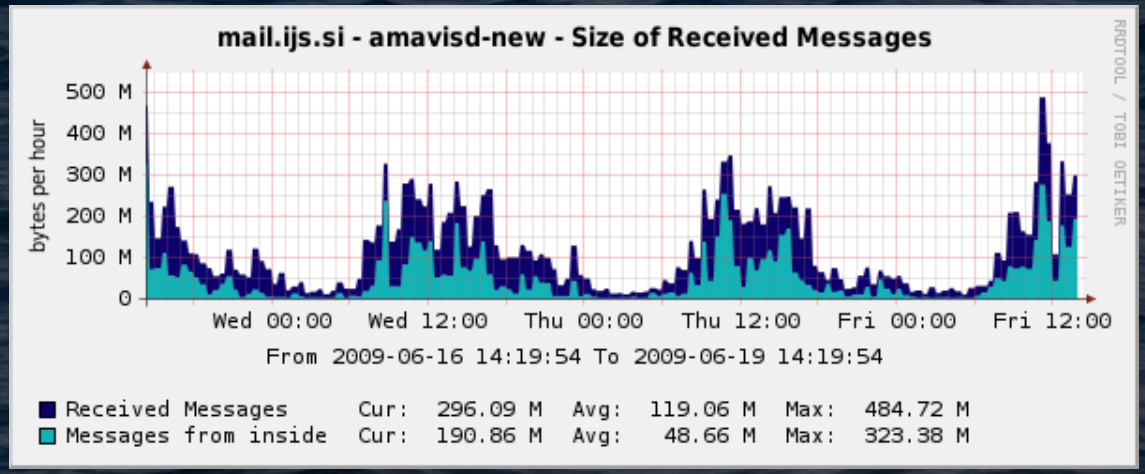
```
_25._tcp.www.example.com. IN TLSA (
    0 0 1 d2abde240d7cd3ee6b4b28c54df034b9
          7983a1d16e8a410e4561cb106618e971 )
```

Nadzor delovanja Amavis procesov

- amavisd-status (0MQ, znakovni/xterm)
- syslog (nastavljiva šablona)
- JSON strukturirani dogodki
- SNMP (MIB 300 spremenljivk, AgentX, 0MQ)

SNMP: MIB 300 spremenljivk





Tradicionalni dnevnik (syslog)

nastavljiv, težavna analiza:

May 8 20:17:58 dorothy amavis[48040]: (48040-03) Blocked SPAM {RejectedInbound,Quarantined}, PROXY-MX/SPAM [216.109.141.91]:3722 [216.109.141.91] <newsletter@123greetings.info> -> <evgeniya.khomyakova@ijs.si>, (216.109.141.91), quarantine: W19/spam/b/bPwQ8bOv-w8m.gz, mail_id: bPwQ8bOv-w8m, b: ABEd6B-Qg, Hits: 17.791, size: 6240, pt: 19, D4, v-SA: Spam, v-CRM: , Subject: "Doctor approved natural Diabetes remedy", From: 123Greetings_<newsletter@123greetings.info>, X-Mailer: Version_5.0, helo=123greetings.info, Tests: [AM.IP_BAD_216.109.141.91=1.2,BAYES_999=4,BAYES_99=3.5,DCC_CHECK=1.1,DIGEST_MULTIPLE=0.293,HTML_MESSAGE=0.001,L_P0F_WXP=2.3,MIME_HEADER_CTYPE_ONLY=0.717,MIME_HTML_ONLY=0.723,MISSING_MID=0.497,RAZOR2_CF_RANGE_51_100=0.5,RAZOR2_CF_RANGE_E8_51_100=1.886,RAZOR2_CHECK=0.922,RCVD_IN_HOSTKARMA_BL=0.3,RCVD_IN_MSPIKE_H2=-0.001,RCVD_NOT_IN_IPREPDNS=0.0001,RP_MATCHES_RCVD=-0.1,SPF_HELO_PASS=-0.001,SPF_PASS=-0.001], shortcircuit=no, autolearn=no autolearn_force=no, autolearnscore=11.366, languages=en,...

May 8 20:17:58 dorothy amavis[48040]: (48040-03) ... relaycountry=US, asn=AS14492_216.109.128.0/19, rss=164560, 3326 ms

Strukturirani dnevnik – JSON


```
{   "@timestamp" => "2014-05-06T09:29:47.048Z",
    "time_unix" => 1399368587.048,
    "time_iso_week_date" => "2014-W19-2",
    "partition" => "19",

    "type" => "amavis",
    "host" => "mailer.example.net",
    "src_ip" => "::1",
    "dst_ip" => "::1",
    "dst_port" => 10024,

    "log_id" => "82329-04",
    "mail_id" => "Jnk7NzYB8pvl",
    "mail_id_related" => ["ne27HTERZaOF"],
```

Strukturirani dnevnik – JSON

"client_port" => 41831,
"client_ip" => "2001:db8::143:1",
"ip_trace" => ["2001:db8::143:1", "192.0.2.242"],
"os_fp" => "Windows XP; dist: 6; raw_mtu: 1340; ...",

"originating" => true,
"policy_banks" => ["PROXY-ORIGINATING", "MYNETS"],
"size" => 302694,
"digest_body" => "a4a7db6307c140b12f57feaf076663f8",

"mail_from" => "mailing-list-1@example.com",
"rcpt_to" => ["recip2@example.org", "recip1@example.net"],

Strukturirani dnevnik – JSON

ga razumejo orodja za analizo dnevnikov

- Logstash > Elasticsearch & Lucene > Kibana
- Splunk
- …

Strukturirani dogodki – obdelava

- amavisd podprocesi
- redis strežnik (queue)
- logstash ali domač perl – queue > http
- Elasticsearch indeks & iskanje: http, cluster, java
- Apache Lucene – iskanje po polnem besedilu
- Kibana – JavaScript @ spletni bralnik

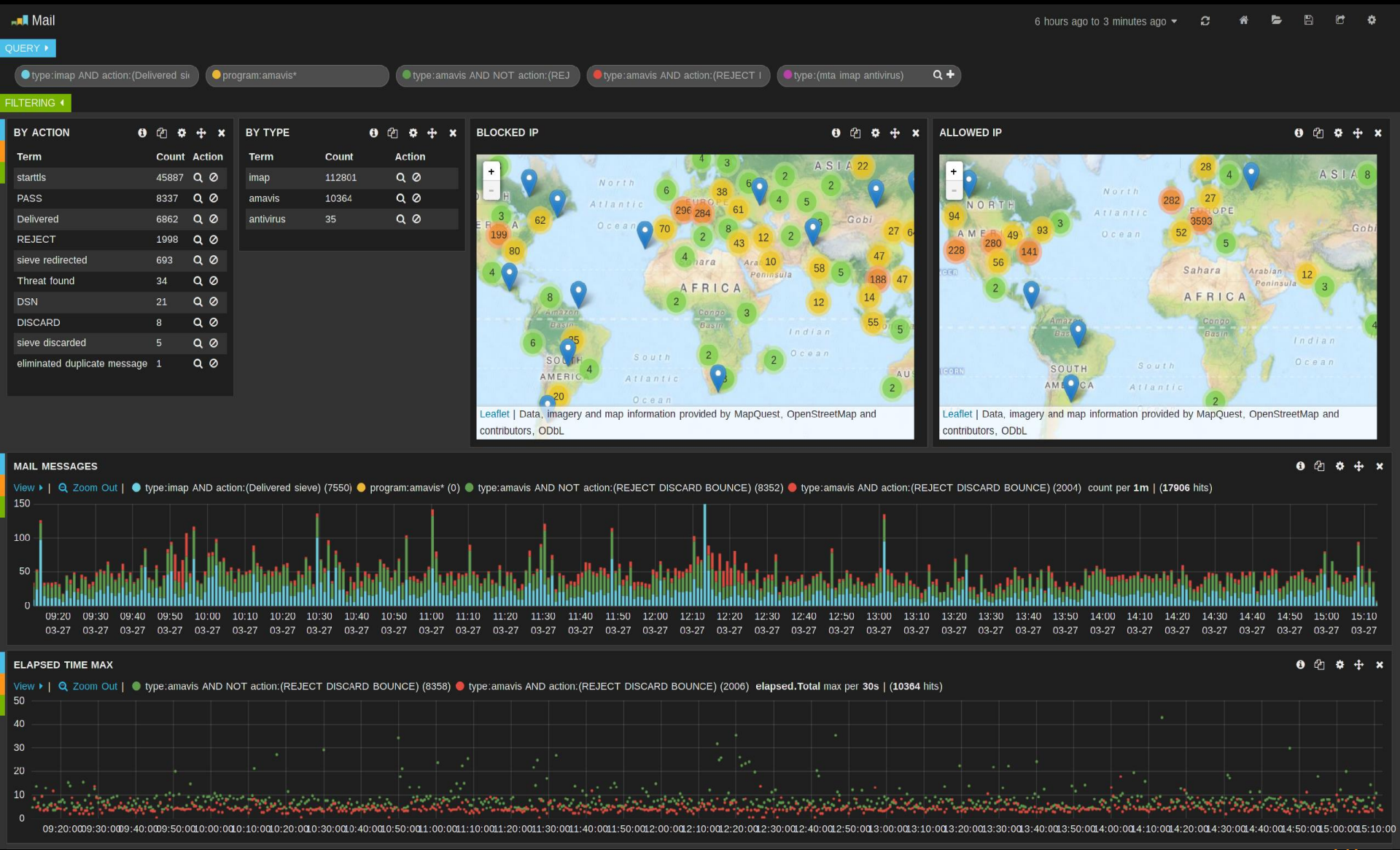
vse komponente so odprtokodne in brezplačne

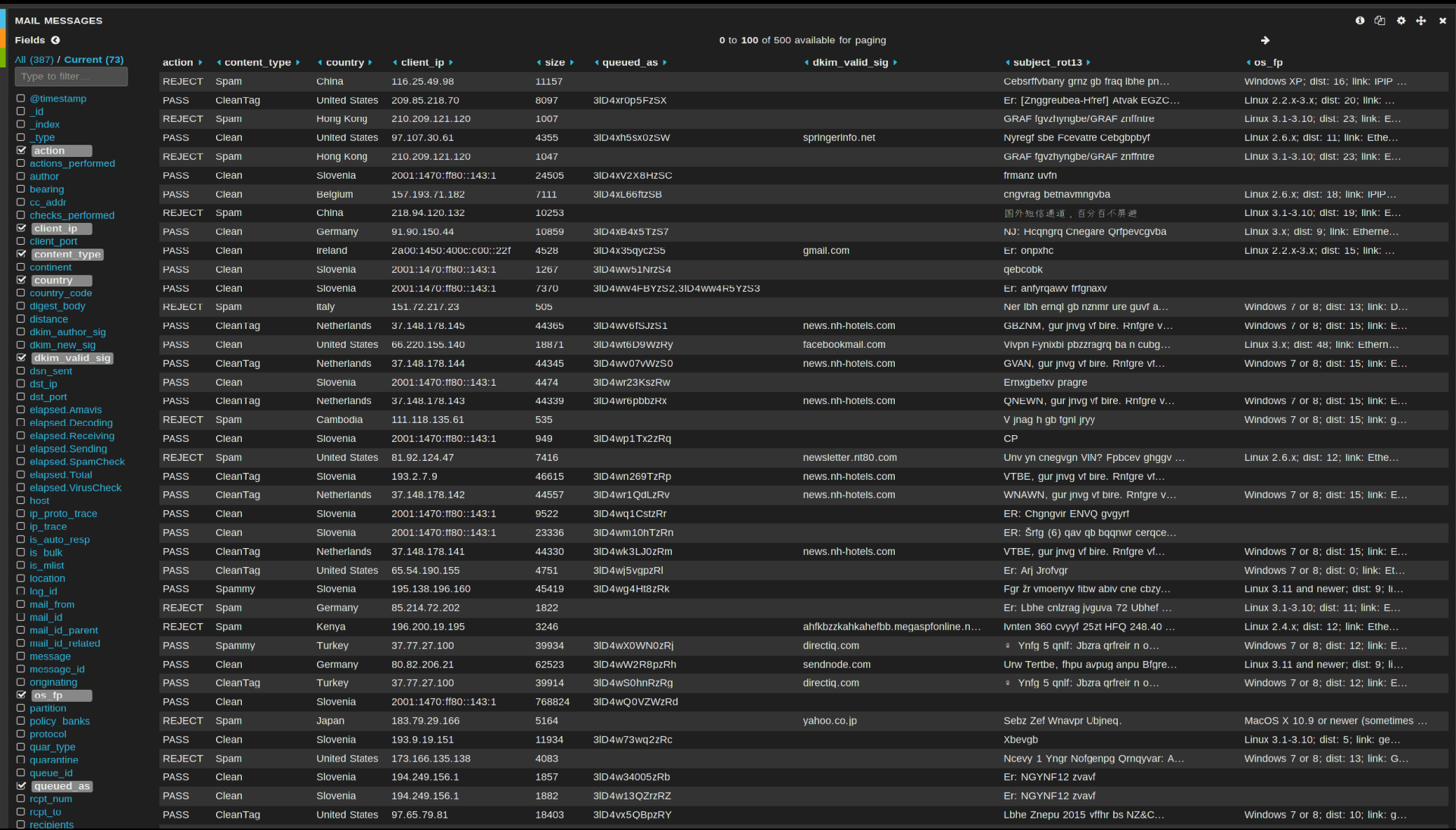
Lucene Query Parser

- člen (term): test, hoj, te?t, test*, te*t
- približno, regexp, bližina, utežitev členov
- stavek: "Hello Kitty"
- boolov operator: OR (impliciran), AND, NOT
- grupiranje: (...)
- interval: [10 TO 1000], ekskl.: {2 TO 5}

- polja: subject:newsletter*, size:[9000 TO *]

Elasticsearch + Kibana

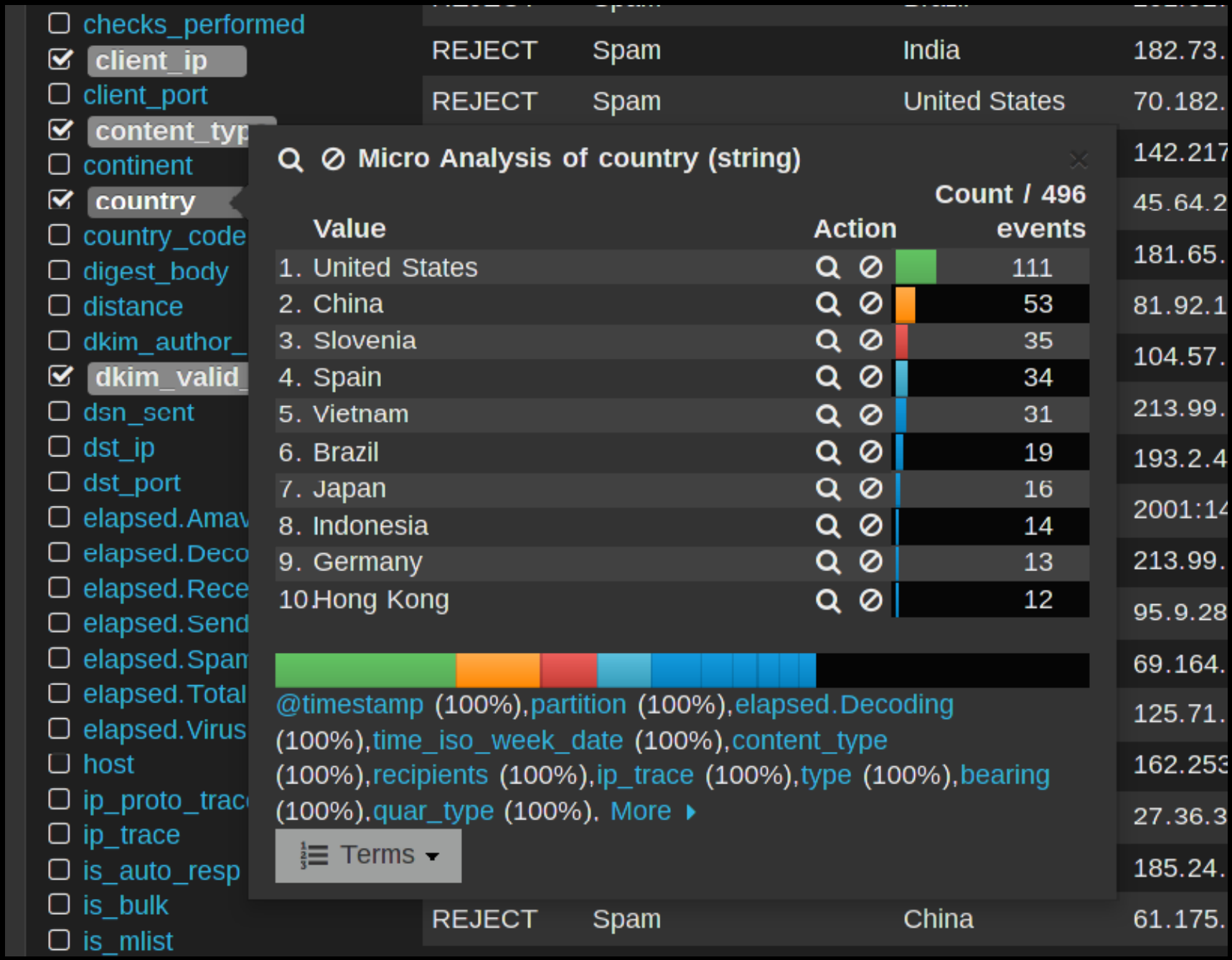



MAIL MESSAGES
Fields
All (387) / Current (73)
Type to filter...
@timestamp
_id
_index
_type
action
actions_performed
author
bearing
cc_addr
checks_performed
client_ip
client_port
content_type
continent
country
country_code
digest_body
distance
dkim_author_sig
dkim_new_sig
dkim_valid_sig
dsn_sent
dst_ip
dst_port
elapsed.Amavis
elapsed.Decoding
elapsed.Receiving
elapsed.Sending
elapsed.SpamCheck
elapsed.Total
elapsed.VirusCheck
host
ip_proto_trace
ip_trace
is_auto_resp
is_bulk
is_mlist
location
log_id
mail_from
mail_id
mail_id_parent
mail_id_related
message
message_id
originating
os_fp
partition
policy_banks
protocol
quar_type
quarantine
queue_id
queued_as
rcpt_num
rcpt_to
recipients
0 to 100 of 500 available for paging
action
content_type
country
client_ip
size
queued_as
dkim_valid_sig
subject_rot13
os_fp
REJECT Spam China 116.25.49.98 11157 Cebsrffvbany grnz gb fraq lbhe pn... Windows XP; dist: 16; link: IPIP ...
PASS CleanTag United States 209.85.218.70 8097 3ID4xr0p5FzSX Er: [Znggreubea-Hfref] Atvak EGZC... Linux 2.2.x-3.x; dist: 20; link: ...
REJECT Spam Hong Kong 210.209.121.120 1007 GRAF fgvzhyngbe/GRAF znffntre Linux 3.1-3.10; dist: 23; link: E...
PASS Clean United States 97.107.30.61 4355 3ID4xh5sx0zSW springerinfo.net Nyregf sbe Fcevatre Cebgbpbyf Linux 2.6.x; dist: 11; link: Ethe...
REJECT Spam Hong Kong 210.209.121.120 1047 GRAF fgvzhyngbe/GRAF znffntre Linux 3.1-3.10; dist: 23; link: E...
PASS Clean Slovenia 2001:1470:ff80::143:1 24505 3ID4xV2X8HzSC frmanz uvfn
PASS Clean Belgium 157.193.71.182 7111 3ID4xL66ftzSB cngvrag betnavmngvba Linux 2.6.x; dist: 18; link: IPIP...
REJECT Spam China 218.94.120.132 10253 国外短信通道，百分百不屏避 Linux 3.1-3.10; dist: 19; link: E...
PASS Clean Germany 91.90.150.44 10859 3ID4xB4x5TzS7 NJ: Hcqngrq Cnegare Qrfpevcgvba Linux 3.x; dist: 9; link: Etherne...
PASS Clean Ireland 2a00:1450:400c:c00::22f 4528 3ID4x35qyczS5 gmail.com Er: onpxhc Linux 2.2.x-3.x; dist: 15; link: ...
PASS Clean Slovenia 2001:1470:ff80::143:1 1267 3ID4ww51NrzS4 qebcobk
PASS Clean Slovenia 2001:1470:ff80::143:1 7370 3ID4ww4FBYzS2,3ID4ww4R5YzS3 Er: anfyrqawv frfgnaxv
REJECT Spam Italy 151.72.217.23 505 Ner lbh ernql gb nznmr ure guvf a... Windows 7 or 8; dist: 13; link: D...
PASS CleanTag Netherlands 37.148.178.145 44365 3ID4wv6fSJzS1 news.nh-hotels.com GBZNM, gur jnvg vf bire. Rnfgre v... Windows 7 or 8; dist: 15; link: E...
PASS Clean United States 66.220.155.140 18871 3ID4wt6D9WzRy facebookmail.com Vivpn Fynlxbi pbzzragrq ba n cubg... Linux 3.x; dist: 48; link: Ethern...
PASS CleanTag Netherlands 37.148.178.144 44345 3ID4wv07vWzS0 news.nh-hotels.com GVAN, gur jnvg vf bire. Rnfgre vf... Windows 7 or 8; dist: 15; link: E...
PASS Clean Slovenia 2001:1470:ff80::143:1 4474 3ID4wr23KszRw Ernxgbefxv pragre
PASS CleanTag Netherlands 37.148.178.143 44339 3ID4wr6pbbzRx news.nh-hotels.com QNEWN, gur jnvg vf bire. Rnfgre v... Windows 7 or 8; dist: 15; link: E...
REJECT Spam Cambodia 111.118.135.61 535 V jnag h gb fgnl jryy Windows 7 or 8; dist: 15; link: g...
PASS Clean Slovenia 2001:1470:ff80::143:1 949 3ID4wp1Tx2zRq CP
REJECT Spam United States 81.92.124.47 7416 newsletter.rit80.com Unv yn cnegvgn VIN? Fpbcev ghggv ... Linux 2.6.x; dist: 12; link: Ethe...
PASS CleanTag Slovenia 193.2.7.9 46615 3ID4wn269TzRp news.nh-hotels.com VTBE, gur jnvg vf bire. Rnfgre vf...
PASS CleanTag Netherlands 37.148.178.142 44557 3ID4wr1QdLzRv news.nh-hotels.com WNAWN, gur jnvg vf bire. Rnfgre v... Windows 7 or 8; dist: 15; link: E...
PASS Clean Slovenia 2001:1470:ff80::143:1 9522 3ID4wq1CstzRr ER: Chgngvir ENVQ gvgyrf
PASS Clean Slovenia 2001:1470:ff80::143:1 23336 3ID4wm10hTzRn ER: Šrfg (6) qav qb bqqnwr cerqce...
PASS CleanTag Netherlands 37.148.178.141 44330 3ID4wk3LJ0zRm news.nh-hotels.com VTBE, gur jnvg vf bire. Rnfgre vf... Windows 7 or 8; dist: 15; link: E...
PASS CleanTag United States 65.54.190.155 4751 3ID4wj5vqpzRl Er: Arj Jrofvgr Windows 7 or 8; dist: 0; link: Et...
PASS Spammy Slovenia 195.138.196.160 45419 3ID4wg4Ht8zRk Fgr žr vmoenyv fibw abiv cne cbzy... Linux 3.11 and newer; dist: 9; li...
REJECT Spam Germany 85.214.72.202 1822 Er: Lbhe cnlzrag jvguva 72 Ubhef ... Linux 3.1-3.10; dist: 11; link: E...
REJECT Spam Kenya 196.200.19.195 3246 ahfkbzzkahkahefbb.megaspfonline.n... Ivnten 360 cvyyf 25zt HFQ 248.40 ... Linux 2.4.x; dist: 12; link: Ethe...
PASS Spammy Turkey 37.77.27.100 39934 3ID4wX0WN0zRj directiq.com ♀ Ynfg 5 qnlf: Jbzra qrfreir n o... Windows 7 or 8; dist: 12; link: E...
PASS Clean Germany 80.82.206.21 62523 3ID4wW2R8pzRh sendnode.com Urw Tertbe, fhpu avpug anpu Bfgre... Linux 3.11 and newer; dist: 9; li...
PASS CleanTag Turkey 37.77.27.100 39914 3ID4wS0hnRzRg directiq.com ♀ Ynfg 5 qnlf: Jbzra qrfreir n o... Windows 7 or 8; dist: 12; link: E...
PASS Clean Slovenia 2001:1470:ff80::143:1 768824 3ID4wQ0VZWzRd
REJECT Spam Japan 183.79.29.166 5164 yahoo.co.jp Sebz Zef Wnavpr Ubjneq. MacOS X 10.9 or newer (sometimes ...
PASS Clean Slovenia 193.9.19.151 11934 3ID4w73wq2zRc Xbevgb Linux 3.1-3.10; dist: 5; link: ge...
REJECT Spam United States 173.166.135.138 4083 Ncevy 1 Yngr Nofgenpg Qrnqyvar: A... Windows 7 or 8; dist: 13; link: G...
PASS Clean Slovenia 194.249.156.1 1857 3ID4w34005zRb Er: NGYNF12 zvavf
PASS Clean Slovenia 194.249.156.1 1882 3ID4w13QZrzRZ Er: NGYNF12 zvavf
PASS CleanTag United States 97.65.79.81 18403 3ID4vx5QBpzRY Lbhe Znepu 2015 vffhr bs NZ&C... Windows 7 or 8; dist: 10; link: g...

REJECT Spam Kazakhstan 217.76.76.110

View: **Table** / JSON / Raw

Field	Action	Value
@timestamp		2015-03-27T13:58:11.654Z
_id		YYxSlwO3zYC7
_index		logstash-2015.03.27
_type		amavis
action		REJECT
actions_performed		RejectedInbound Quarantined
author		LetaHambly@stonewestcompanies.com
bearing		67
checks_performed		V S H B F P
client_ip		217.76.76.110
client_port		60372
content_type		Spam
continent		AS
country		Kazakhstan
country_code		KZ
digest_body		1b832d13696eaa342483829606bfe3f2
distance		3984
dsn_sent		false
dst_ip		::1
dst_port		10010
elapsed.Amavis		0.059
elapsed.Decoding		0.021
elapsed.Receiving		0.003
elapsed.Sending		0.011
elapsed.SpamCheck		3.163
elapsed.Total		3.276
elapsed.VirusCheck		0.041
host		mail.ijs.si
ip_proto_trace		ESMTP://[::1]:56492,ESMTP://[217.76.76.110]:60372,x
ip_trace		217.76.76.110
location		68,48
log_id		19316-09
mail_from		LetaHambly@stonewestcompanies.com
mail_id		YYxSlwO3zYC7
message		19316-09 REJECT Spam LetaHambly@stonewestcompanies.com -> dusan.
message_id		<1427464458.4B8EF12A208@mx.spamexperts.com>
originating		false
os_fp		Windows 7 or 8; dist: 17; link: generic tunnel or VPN; params: none; raw_m
partition		13
policy_banks		PROXY-MX,SPAM
protocol		ESMTP
quar_type		Z
quarantine		W13/spam/Y/YYxSlwO3zYC7.gz


checks_performed
client_ip
client_port
content_typ
continent
country
country_code
digest_body
distance
dkim_author_
dkim_valid_
dsn_sent
dst_ip
dst_port
elapsed.Amav
elapsed.Deco
elapsed.Rece
elapsed.Send
elapsed.Span
elapsed.Total
elapsed.Virus
host
ip_proto_trac
ip_trace
is_auto_resp
is_bulk
is_mlist
REJECT Spam India 182.73.
REJECT Spam United States 70.182.
142.217
45.64.2
181.65.
81.92.1
104.57.
213.99.
193.2.4
2001:14
213.99.
95.9.28
69.164.
125.71.
162.253
27.36.3
185.24.
REJECT Spam China 61.175.
Micro Analysis of country (string)
Count / 496 events
Value
Action
1. United States 111
2. China 53
3. Slovenia 35
4. Spain 34
5. Vietnam 31
6. Brazil 19
7. Japan 16
8. Indonesia 14
9. Germany 13
10 Hong Kong 12
@timestamp (100%),partition (100%),elapsed.Decoding (100%),time_iso_week_date (100%),content_type (100%),recipients (100%),ip_trace (100%),type (100%),bearing (100%),quar_type (100%). More
Terms

Povzetek

- filtriranje čim bližje izvoru
- zavrni ali dostavi označeno, ne tiho zvreči
- več info virov pripomore k boljši klasifikaciji
- preverjanje tudi odhodnih sporočil
- DKIM podpis olajša prejemniku odločitve
- ELK: globalni vpogled ali iskanje igle v senu


SpamAssassin
add
string
SNMP
bounces
messages
body
BUG
allow
error
option
method
allows
request
use
Email
previously
banned
problem
addresses
version
entry
section
number
bounce
internal
scanner
based
whitelist
configuration
LOG
keys
AV
sa
logging
lookup
release
smtp
DSN
argument
options
MTA
true
new
command
one
fields
using
e.g
even
updated
name
default
message
previous
id
key
file
amavisd
DKIM
available
may
RFC
following
instead
originating
value
address
quarantine
spam
suggested
names
See
match
signed
results
Perl
signing
information
NOTE
amavisd.conf
virus
verification
now
sql
setting
ccat
module
longer
partition
time
program
level
banks
policy
mail
domain
type
Also
tls
set
versions
header
score
added
maps
CC
data
code
tag
change
category
generated
field
variable
signature
list
recipients
table
report
files
rules
currently
bank
two
just
reported
additional
line
process
status
sender
database
Like
attribute
example
domains
i.e
according
contents
used
macro
scanners
recipient
limit
provided
tables
checks
client
provide