anti-spam techniques

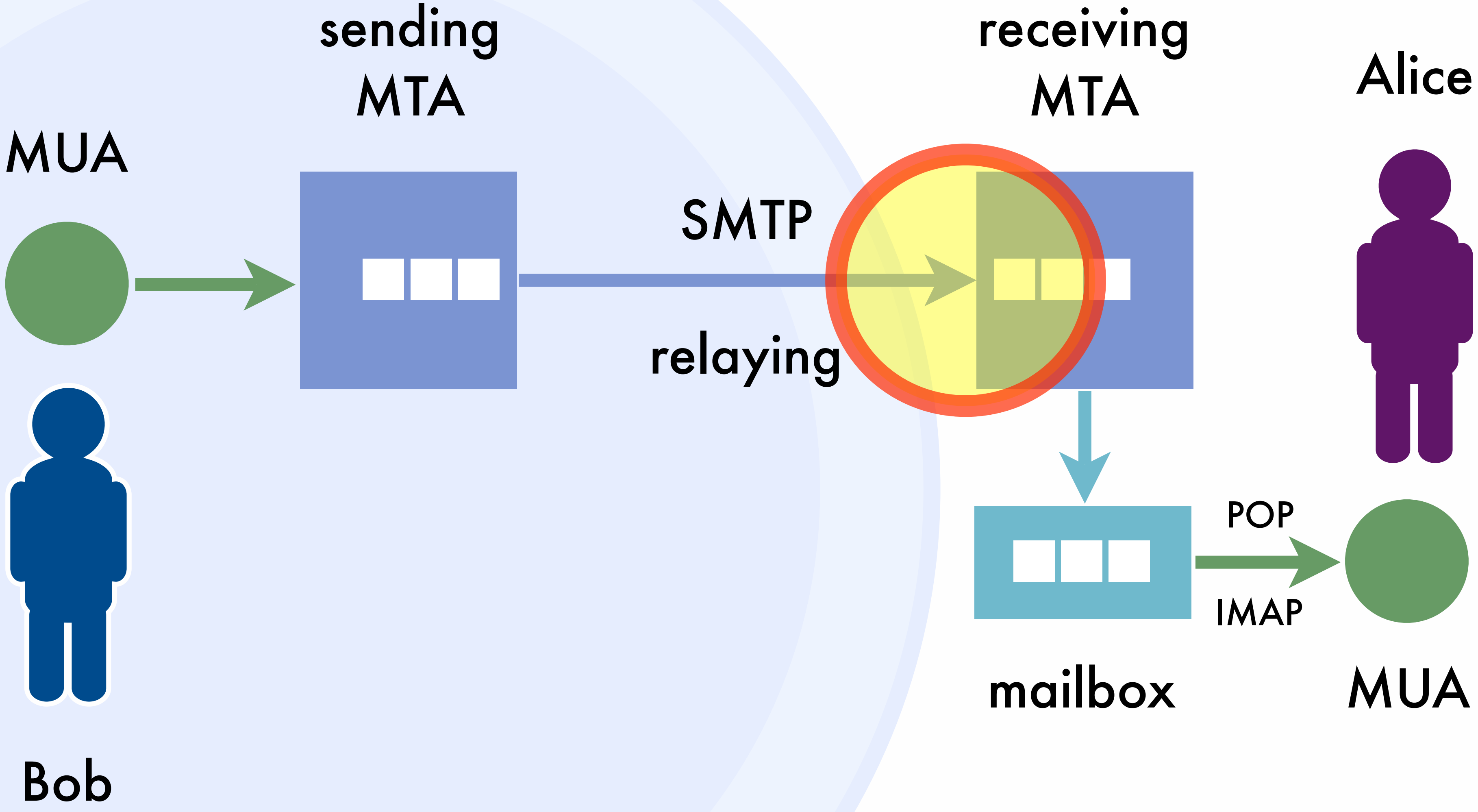
beyond Bayesian filters

- Plain Old SMTP
  - protocol overview
- Grey-Listing
  - save resources on receiver side
- Authentication of Senders
  - Sender ID Framework IP-based
  - DomainKeys signing-based

smtpintro

smtpintro

simple mail transfer protocol



```
S:  220 tik6.ethz.ch ESMTP Postfix
C:  HELO student.ethz.ch
S:  250 tik6.ethz.ch
C:  MAIL FROM:<fabio@student.ethz.ch>   envelope sender
S:  250 Ok
C:  RCPT TO:<nburri@tik.ee.ethz.ch>     envelope receiver
S:  250 Ok
C:  DATA
S:  354 End data with <CR><LF>.<CR><LF>
C:  Subject: Test
    From: Fabio Lanfranchi    headers
    To: Nicolas Burri

    Hello, World!       message body
    .
S:  250 Ok: queued as 6CDB86ADD7
C:  QUIT
S:  221 Bye
```

smtpstatus

- not sending directly
  - multiple recipients
  - temporary problems
- server reply messages
  - 2XX positive completion
  - 3XX positive intermediate
  - 4XX transient negative completion
  - 5XX permanent negative completion

retry after:
30 min (1st)
60 min (2nd)
every 2 to 3 h

spamdisaster

- RFC 821 (August 1982)
- no sender authentication
  - message forged or authentic?
  - spam, spoofing, viruses, phishing

make transmission appear to come from another user

trick users into providing personal information

dnsrecords

domain name system

class (internet)

type

```
dcg.ethz.ch.        IN  CNAME  pc-4650.ethz.ch.

pc-4650.ethz.ch.    IN  A      129.132.57.243

tik.ee.ethz.ch.     IN  MX     tik6.ethz.ch.

tik6.ethz.ch.       IN  A      129.132.119.136
```

mail exchange

antispam

- techniques (today)
  - keyword filtering
  - black-listing
- problems
  - false positives
  - cost on receiver side

greylisting

greylisting

- Evan Herris (2003)
- blocking technique on MTA level
- save resources on receiving MTA
- make life harder for spammers
- require minimal maintenance
- have minimal impact on users

```
S:   220 tardis.ee.ethz.ch ESMTP Postfix
C:   HELO fabio.ch
S:   250 tardis.ee.ethz.ch
C:   MAIL FROM:<mail@fabio.ch>
S:   250 Ok
C:   RCPT TO:<oetiker@ee.ethz.ch>
S:   450 Greylisted for 300 seconds
C:   QUIT
S:   221 Bye
```

recipent address rejected

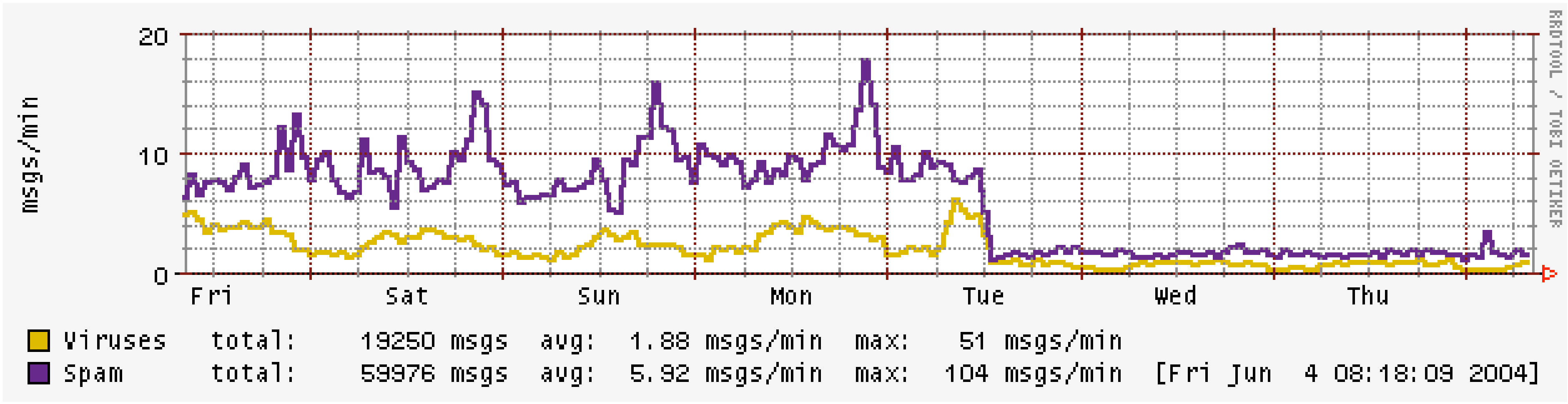
greylisting

- store triplet in database
  - client IP address
  - envelope sender
  - envelope receiver
- additional information
  - first seen
  - expiry of blocking, expiry of record
  - counters: blocks, passes

greylisting

- first attempt
  - refuse delivery (4XX error)
  - block triplet for some minutes
- second attempt
  - unblock triplet
  - accept message
- aging of record
  - delete it after a month

greylisting



- spam and viruses
  - «fire and forget» methodology
  - 95% effectiveness

greylisting

- no content, no overhead
  - less resource usage for filtering
- no false positives
- database allows traffic analysis
- blacklists more effective
- lot of work for spammers

greylisting

- delivery delays
- problems with
  - multiple mail servers per domain
  - mailing lists: changing sender address
- adaption by spammers
  - experts say: within 1 year

senderid

senderid

- Sender ID Framework
- a merger and refinement of proposals
  - SPF (Sender Policy Framework)
    - inspired by RMX and DMP
  - Microsoft Caller ID
- industry collaboration
  - AOL, Microsoft, IBM, VeriSign ...

senderid

- create multiple choke points
- protects sender's domain from spoofing and phishing: receivers validate origin of mail
- prevent «before it happens»
- a foundation for the reliable use of domain names in accreditation, reputation and safe lists
- the first step industry need to take together
- use of existing services: DNS and SMTP

senderid

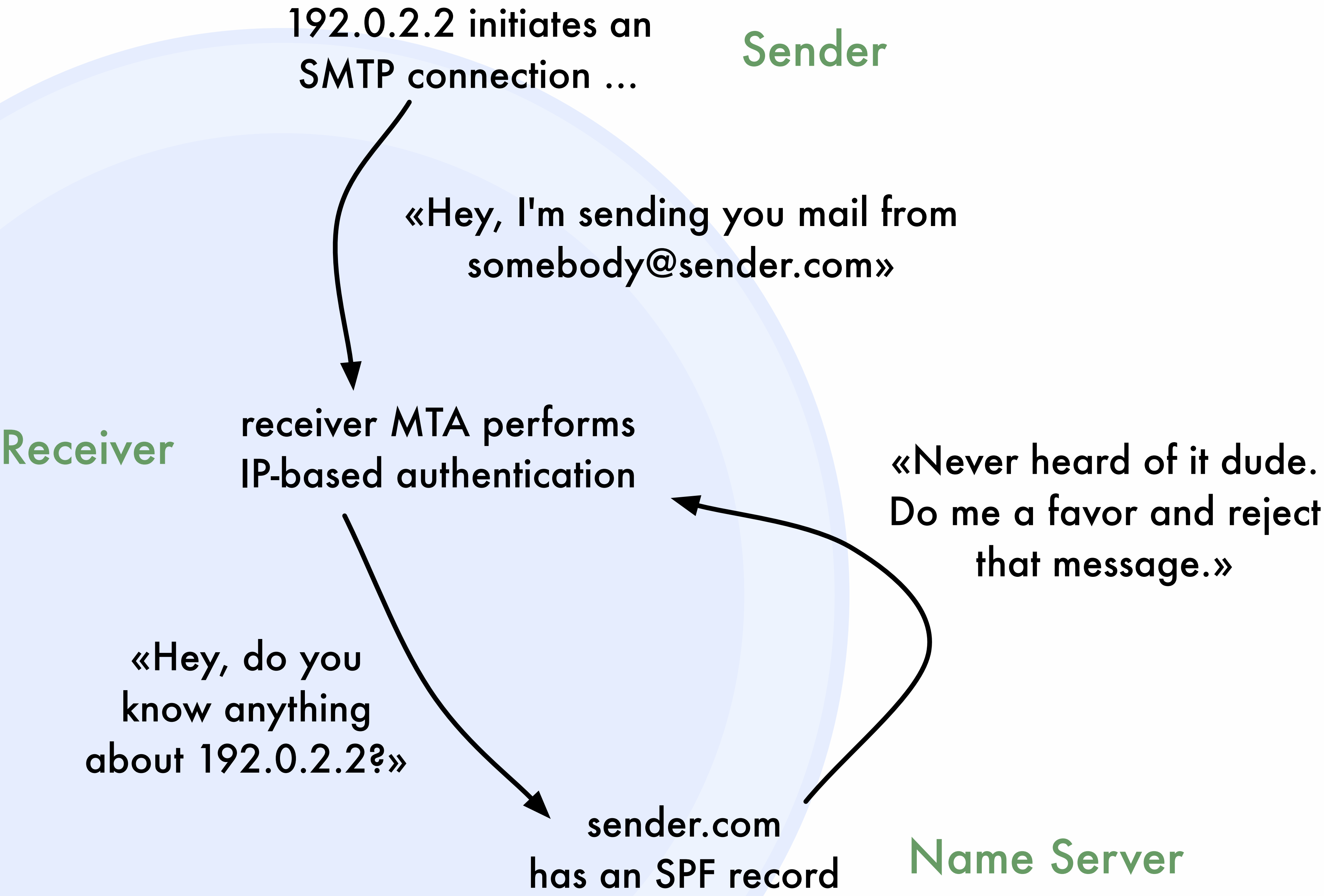
framework of technical specifications

DNS	Sender ID Record (SPF)	
Check	MAIL FROM Classic SPF	PRA (Microsoft)
SMTP	SUBMITTER SMTP Optimization	

senderid

- senders publish IP addresses of outbound email servers in DNS
- receivers determine which domain to check
  - Purported responsible domain (PRA)
  - Envelope From (Classic SPF)
- receivers query DNS for the outbound email servers of the chosen domain and perform domain spoofing test

senderid



192.0.2.2 initiates an SMTP connection ...

Sender

«Hey, I'm sending you mail from somebody@sender.com»

Receiver

receiver MTA performs IP-based authentication

«Never heard of it dude. Do me a favor and reject that message.»

«Hey, do you know anything about 192.0.2.2?»

sender.com has an SPF record

Name Server

rmxrecords

Reverse MX (Hadmut Danisch, 2003)

```
example.com.         IN   RMX   "ip4:10.0.0.0"
                     IN   RMX   "host:relay.example.com"
                     IN   RMX   "apl:relays.provider.de"
relays.provider.de.  IN   APL   "213.133.101.22 1.2.3.0/24"
```

Allowed hosts: 10.0.0.0, relay.example.com, 213.133.101.22, and 1.2.3.0/24

dmprecords

Designated Mailer Protocol (Gordon Fecyk, 2003)

```
1.2.0.192.in-addr._smtp_client.example.com.
                                     IN  TXT  "dmp=allow"

2.2.0.192.in-addr._smtp_client.example.com.
                                     IN  TXT  "dmp=allow"

*.in-addr._smtp_client.example.com.  IN  TXT  "dmp=deny"
```

Allowed hosts: 192.0.2.1 and 192.0.2.2

rmxvsdmp

	Danisch RMX	Fecyk DMP
large entries	potentially	IP-address specific
DNS extension	RMX record type	TXT records
indirection	pointers to APL	list for each domain
dynamic hostnames	DynDNS pointer	update records
CIDR notation	built into APL	byte boundary
joe-job notification	static mailhost list	DNS logs
DNS caching	save bandwidth	IP-specific

spfrecords

Sender Permitted From (Meng Weng Wong, 2004)

```
spammer.com.   IN   TXT   "v=spf1 +all"

gmx.net.       IN   TXT   "v=spf1 ip4:213.165.64.0/23 -all"

gmx.de.        IN   TXT   "v=spf1 include:gmx.net -all"

*.ethz.ch.     IN   TXT   "v=spf1 +mx +a:smtp.ethz.ch -all"

*.dialup.ch.   IN   TXT
              "v=spf1 exists:%{ir}.%{lr}._spf.%{d} -all"
```

192.0.2.1 sends email as <someuser@dialup.ch>
resulting query: 1.2.0.192.someuser._spf.dialup.ch

senderid

- PRA (Microsoft patent)
  - validates identity seen by user
  - parses headers and tries to find out the entity most recently responsible for injecting a message into the email system
- Classic SPF
  - validates MAIL FROM address (return-path)

senderid

check_host() from SPF specification

```
check_host(<ip>, <domain>, <sender>)

   domain is badly formed => return FAIL

   sender has no local part => assume postmaster

   fetch DNS records for domain
   or return FAIL        // SPF entry denies relay
   or return TEMPERROR   // DNS server down
   or return NONE        // SPF entry doesn't exist
   or return PERMERROR   // syntax error in SPF entry
```

Based on this information other tools and techniques can be applied to identify spoofing and spamming. (e.g. keyword filtering)

senderid

Forwarding

someuser@example.com
sends email to
fabio@student.ethz.ch
that is forwarded to
mail@fabio.ch

MAIL FROM:
<someuser@example.com>

add a Resent-From: header
=> PRA can find out last sender

Mailing Lists

mailing list
list@example.com

MAIL FROM:
<list@example.com>

add a Sender: header
=> PRA can find out last sender

senderid

SUBMITTER SMTP extension

```
S:   220 fabio.ch ESMTP Postfix
C:   EHLO student.ethz.ch
S:   250-SUBMITTER
S:   250 Ok
C:   MAIL FROM:<somuser@example.com>
     SUBMITTER=<fabio@student.ethz.ch>
S:   250 Ok
C:   RCPT TO:<mail@fabio.ch>
S:   250 Ok
```

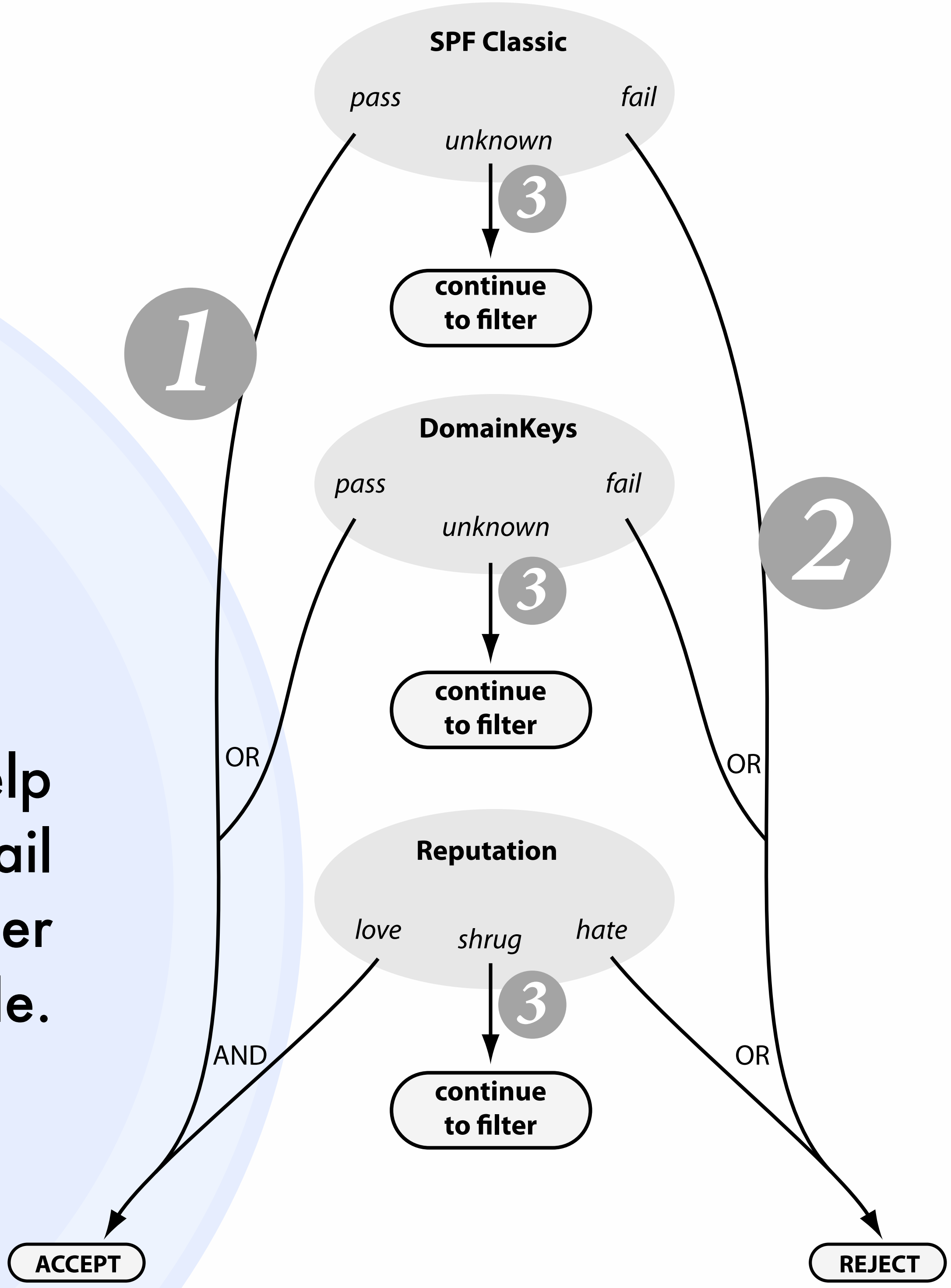
SPF Classic doesn't need to look at headers to decide if sender is allowed to relay email for a domain.

domainkeys

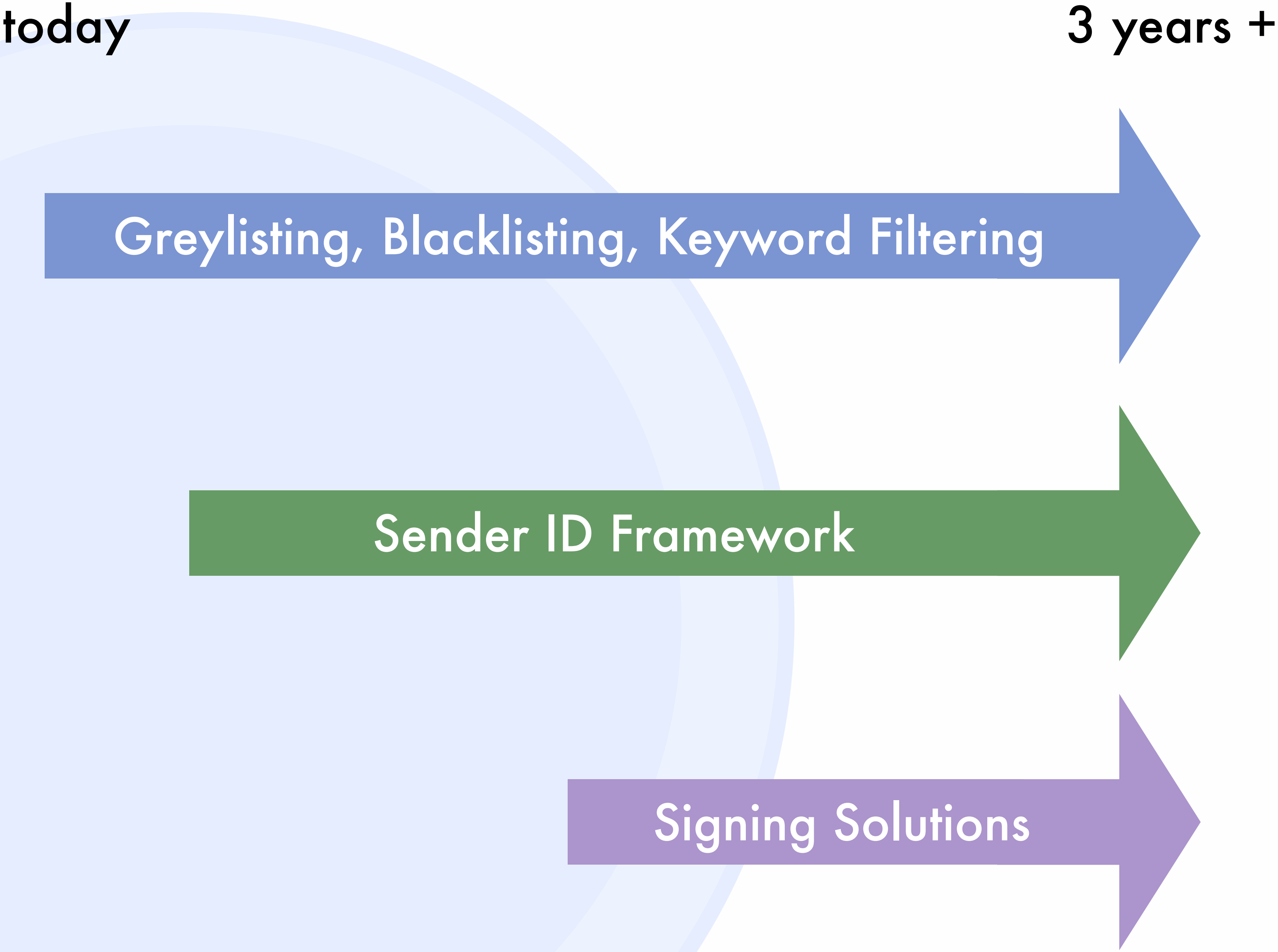
- proposal by Yahoo (August 2004)
- provider generates public/private key pairs
- public key is published in DNS
- outgoing email is signed with private key
- receiver incoming mail against public key

SPF Classic and DomainKeys to authenticate senders of email

Reputation lists will help receivers decide if a mail from an authenticated sender is desirable or undesirable.



senderid



today

3 years +

Greylisting, Blacklisting, Keyword Filtering

Sender ID Framework

Signing Solutions