A case study on building a mail archive with postfix

Cologne 26.09.2018

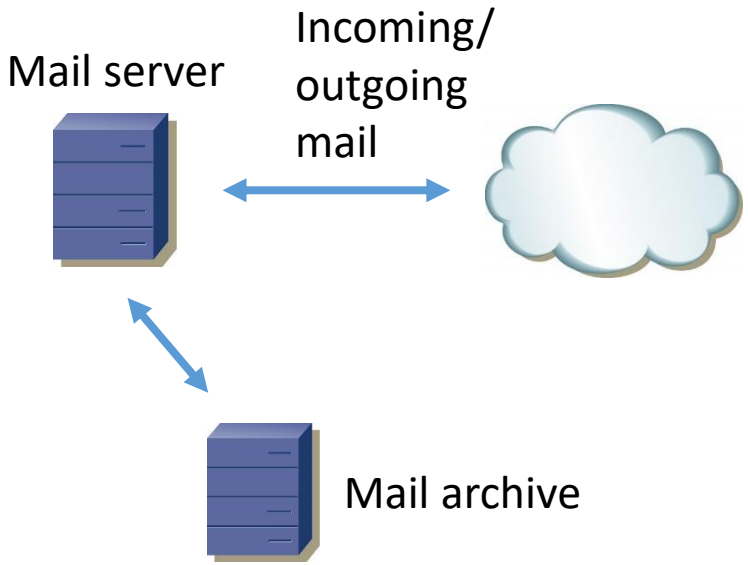
OPEN SOURCE BACKUP CONFERENCE

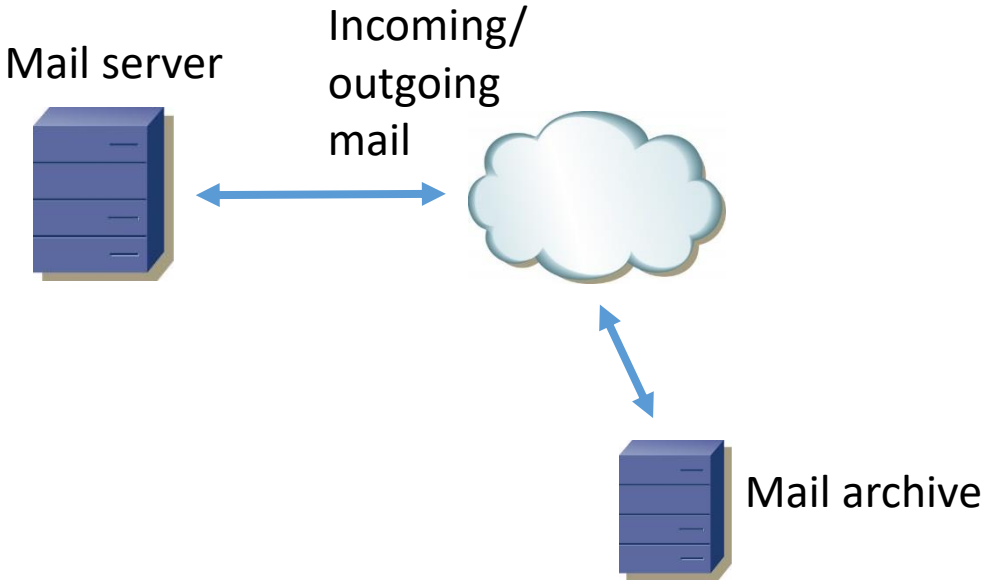
backup & recovery help

Dr. Stephan Rein

What is a Mail Archive ?

- A central system that keeps a copy of all incoming and outgoing e-mails.
- Internal or external mail server
- Archive is rather internal; can be integrated in/part of the mail server





Features of a Mail Archive

- Archive all mails for later retrieval
- All mails: new, current, different accounts
- Save storage, reduce workload: Release mails from the mail server
- Search
- Keep structure of folders, meta data
- Simplify backup for externally managed mail
- Protection against failover (service outage, data loss)
- Protect against general legal risk
- Meet the EU‘s General Data Protection Regulation (GDPR)

Mail Terminology

Mail User Agent (MUA)
SMTP
POP3/IMAP
Thunderbird, Roundcube,…

Mail Delivery Agent (MDA)
POP3/IMAP
Cyrus, Dovecot

Mail Transport Agent (MTA)
SMTP
Postfix

- Some destination mail server or
- receiving mail from this server

Considered Tools / approaches

Forwarding approach	Synchronization approach	
Postfix BCC feature v3.1.8	Dovecot's dsync / doveadm backup v2.2.27	Imapsync v1.882
• **Postfix** is a free and open-source mail transfer agent (MTA) that routes and delivers electronic mail. • Debian package *postfix* available	• Dovecot is an open source IMAP and POP3 email server for Linux/UNIX-like systems • Debian package *dovecot-imapd*	• Imapsync is an IMAP transfers tool. The purpose of imapsync is to migrate IMAP accounts or to **backup** IMAP accounts. • No package but easy instructions here: https://imapsync.lamiral.info/INSTALL.d/INSTALL.Debian.txt

- Did not test cyrus backup (current version 3.0.8)
- Feature is not compiled
- Cyrus Version 3.0.8 Documentation:

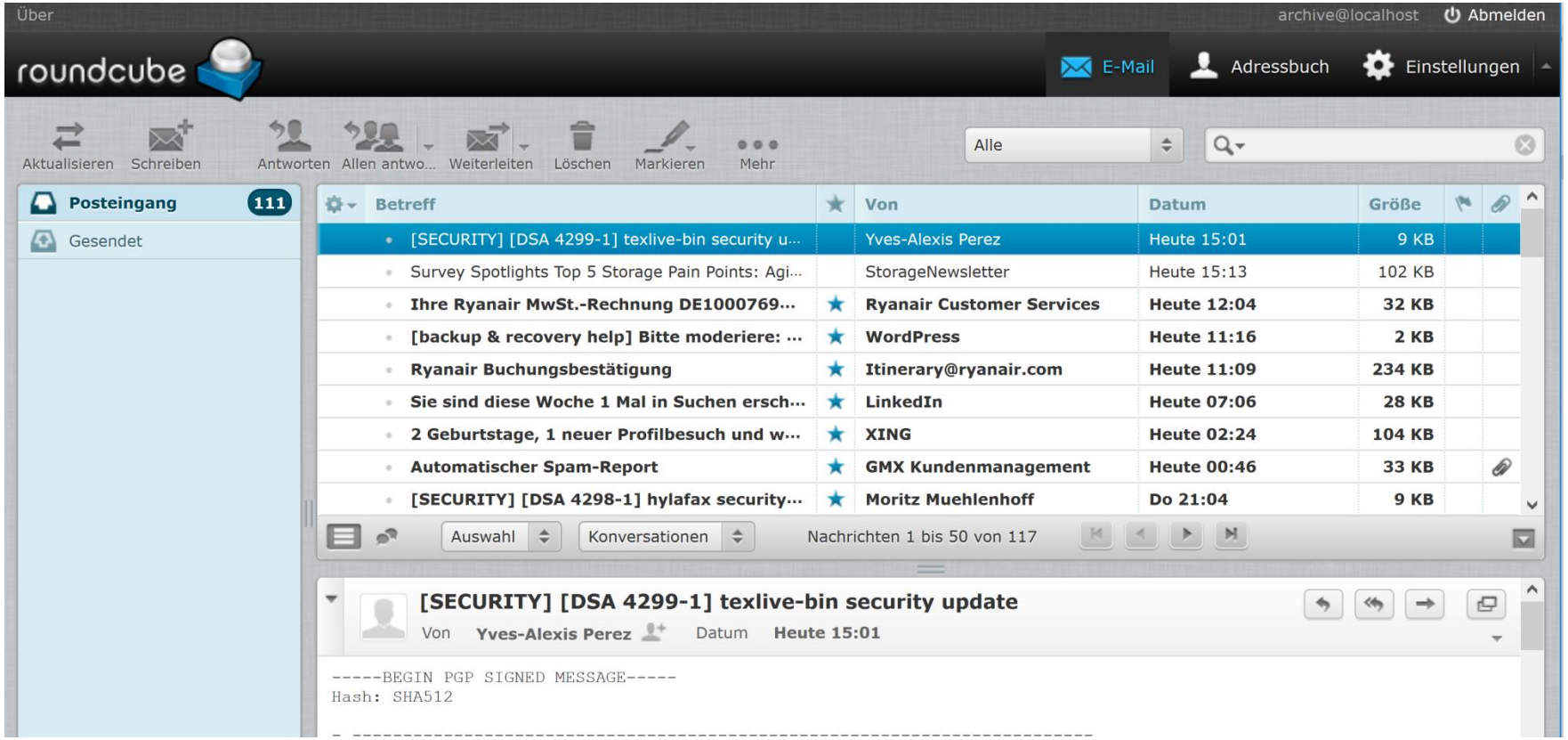
"Cyrus Backups are a replication-based backup service for Cyrus IMAP servers. This is currently an experimental feature."

Maildir Format (vs. mbox)

- Each mailbox folder is a directory, each message a file (mbox uses single file)
- Index for each folder (search, detect duplicates)
- Improved efficiency
- Generally the preferred format

Mail archive user interface

- Mail client application (Thunderbird, …) or Webmailer (Roundcube, …)

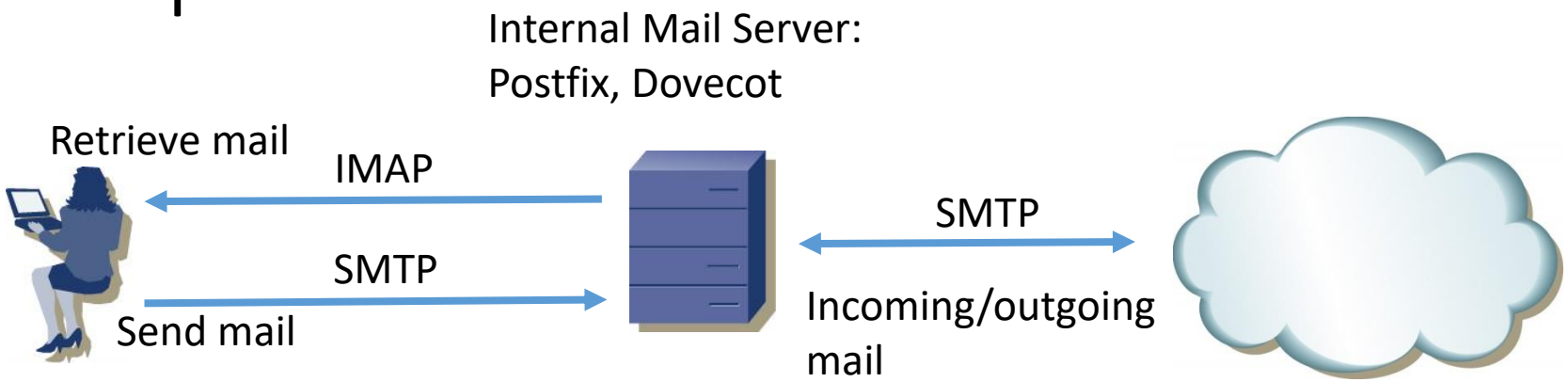


Rating Table

	Postfix bcc	doveadm backup	doveadm 2-way sync	doveadm 1-way sync	imapsync
Installation	?	?			?
Documentation					
Configuration					
Feasibility/ Integration					
Useful for archive					
Log File					
Performance					
Legal perspective	?	?			?

++ = very good	+ = good	O = ok	- = not so good	-- = bad

Postfix BCC feature: a) internal mail server setup



- Fully self-hosted mail server
- Can be internal or external

Postfix BCC feature: Definition of rules

1. Create accounts for mail archive: #adduser *user1_rcv*, *user2_rcv*, …
2. Create two files *bcc_archive_rcv* and *bcc_archive_snd*

The two files define to which account incoming and outgoing mail is archived:

```
# /etc/postfix/bcc_archive_rcv
user1@domain1.tld   user1_rcv@domain1.tld
user2@domain2.tld   user2_rcv@domain2.tld
```

```
# /etc/postfix/bcc_archive_snd
user1@domain1.tld   user1_snd@domain1.tld
user2@domain2.tld   user2_snd@domain2.tld
```

Postfix BCC feature: Include the rules in postfix configuration

3. Insert two lines in /etc/postfix/main.cf:

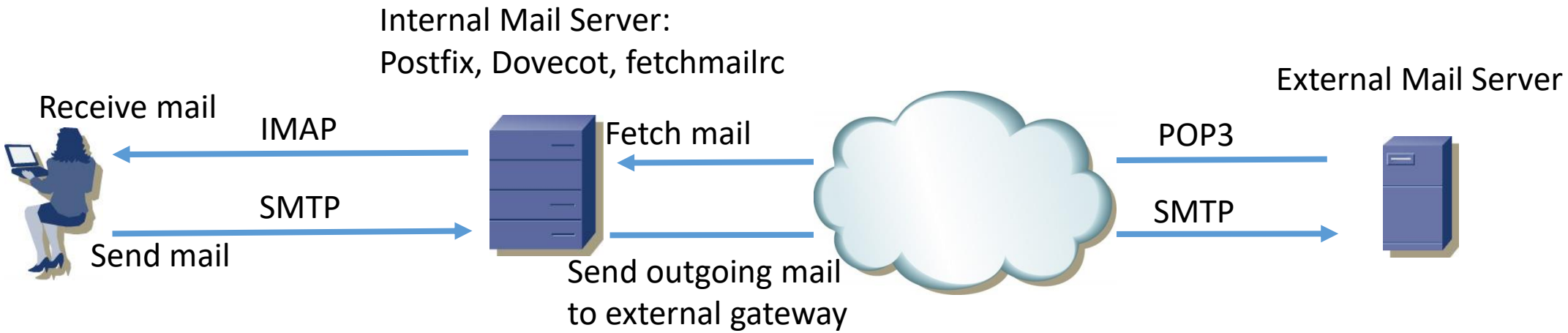
```
# file for receiving mail rules
recipient _bcc_maps = hash:/etc/postfix/bcc_archive_rcv

# file for sending mail archiving rules
sender_bcc_maps     = hash:/etc/postfix/bcc_archive_snd
```

4. Translate the hash tables:

```
# postmap /etc/postfix/bcc_archive_snd
# postmap /etc/postfix/bxx_archive_rcv
```

5. Restart postfix

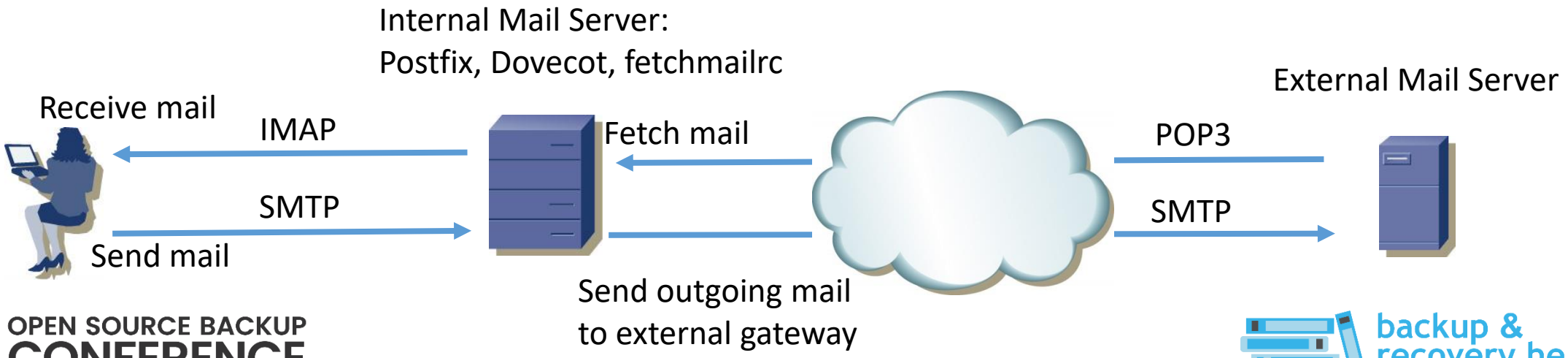
Postfix BCC: b) external mail server



- Fetch mails from external mail server with fetchmailrc (mails are stored locally)
- Configure Postfix to use the external mail server for sending mail
- User sends/ retrieves mail to/ from internal server
- Incoming and outgoing mail is archived with Postfix BCC feature

Postfix BCC: b) external mail server - features

- own mail server but without public IP, external provider takes care of domain, spam, …
- Security: no open port
- Poll multiple external mail accounts: central collection point
- Better performance due to local network but VPN might be needed



Configuration of fetchmailrc

```
# /etc/fetchmailrc
poll mailserver1.tld protocol POP3 user username1 password password1 to localuser1 ssl
poll mailserver2.tld …
```

- Can poll multiple accounts

Configuration of external mail gateway/ relay in Postfix

```
# a few more lines in /etc/postfix/main.cf
relayhost  = [mail.domain.com]:25
# for authentification:
smtp_sasl_auth_enable = yes
smtp_sasl_password_maps = hash:/etc/postfix/relay_passwd
smtp_sasl_security_options = noanonymous
```

```
#/etc/postfix/relay_passwd
[mail.domain.com]:25 username:password
```

1) Edit main.cf
2) Create the file /etc/postfix/relay_passwd
3) Create the hash database file for postfix

- sudo postmap /etc/postfix/relay_passwd
- sudo service postfix restart

Postfix BCC: rating

- Easy to implement if postfix is already used (internal mailserver)
- Otherwise requires intrusive modification of the mail system
- External mail server: In case of a fail out mails still do not get lost
- Folders are NOT preserved !!

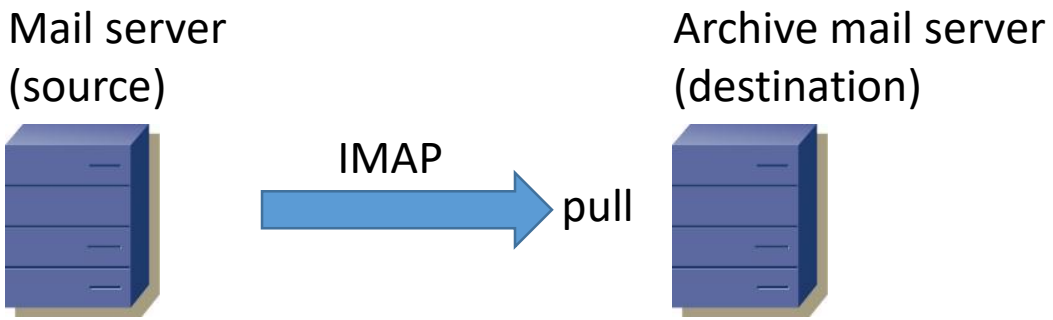
Rating: Postfix

	Postfix bcc	doveadm backup	doveadm 2-way sync	doveadm 1-way sync	imapsync
Installation	++				
Documentation	++				
Configuration	+				
Feasibility/ Integration	++/O				
Useful for archive	+				
Log file	O				
Performance	++				
Legal perspective					

++ = very good	+ = good	O = ok	- = not so good	-- = bad

Synchronization Approach

- Two mail servers one is source, the other is the archive
- Testbed with local network
- IMAP, pull from archive-side
- Start with dovecot, then imapsync



dovecot‘s synchronization utilities: info of the documentation

- **doveadm backup**

performs one-way synchronization. If there are any changes in the destination they will be deleted, so the destination will look exactly like the source.

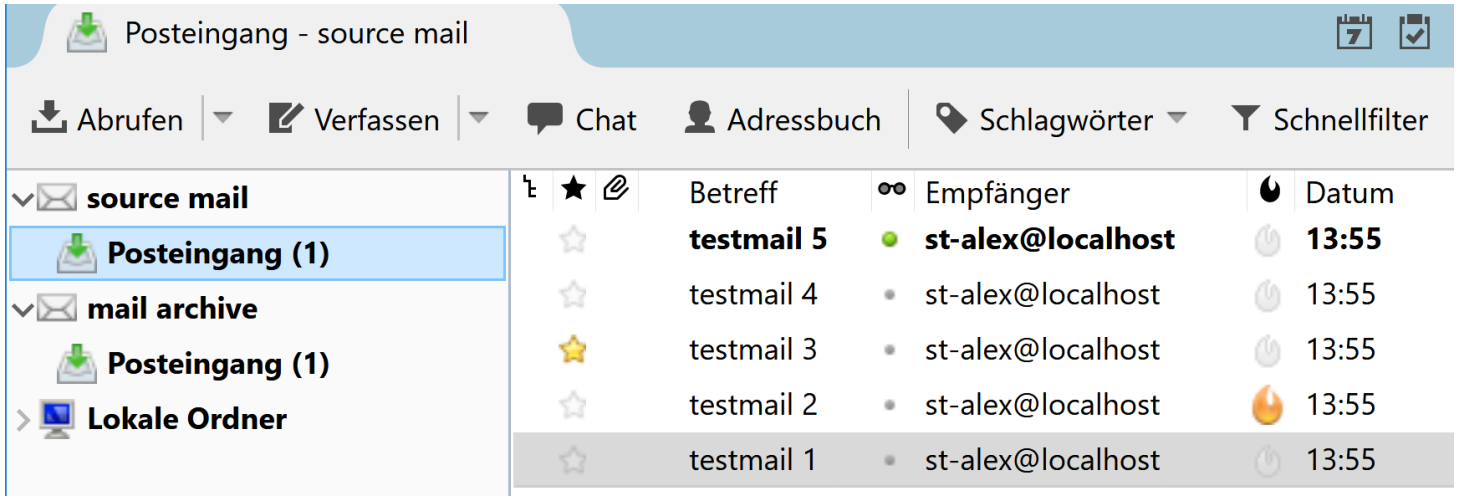
- **doveadm sync**

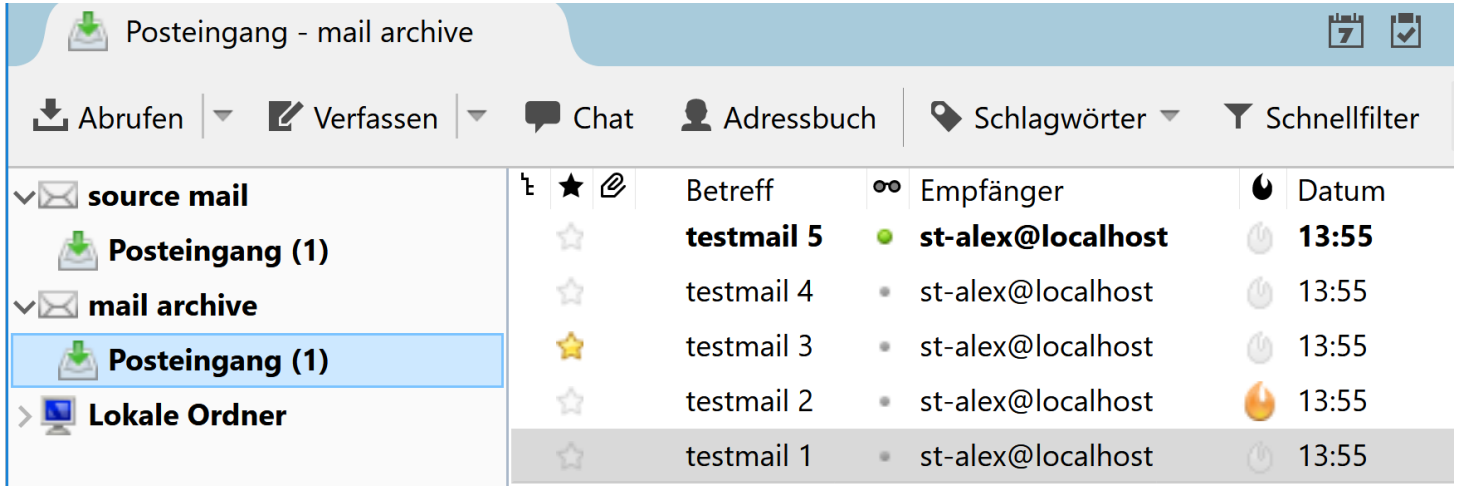
performs two-way synchronization. It merges all changes without losing anything. Both the mailboxes will end up looking identical after the synchronizations finished.

- **doveadm sync -1**

performs one-way synchronization, but it merges the changes in destination without deleting

doveadm backup

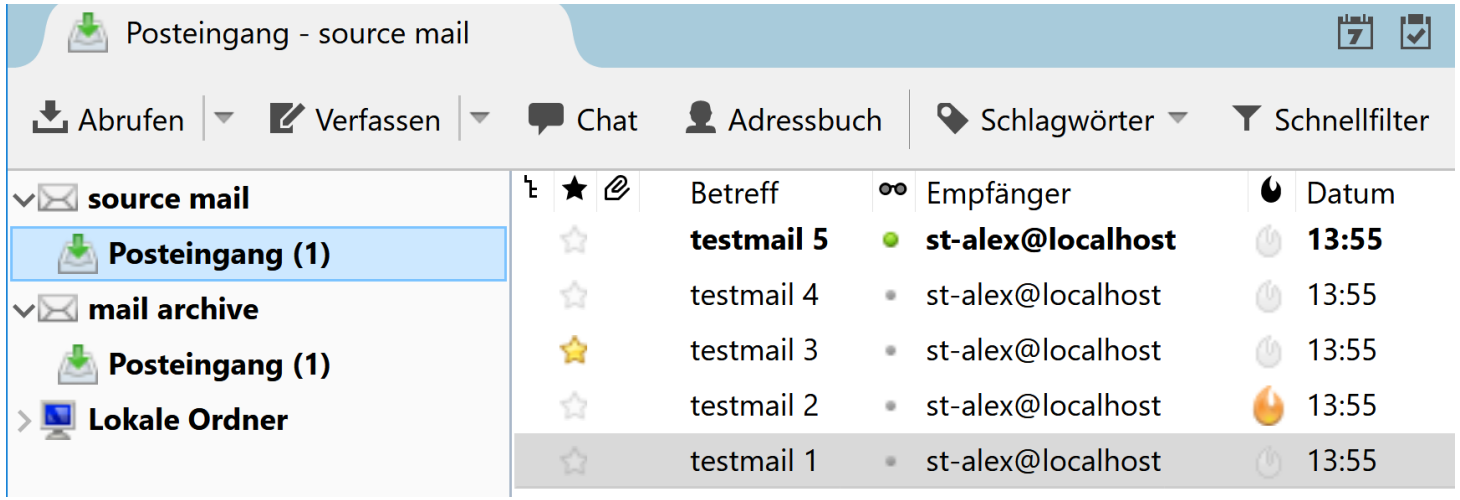


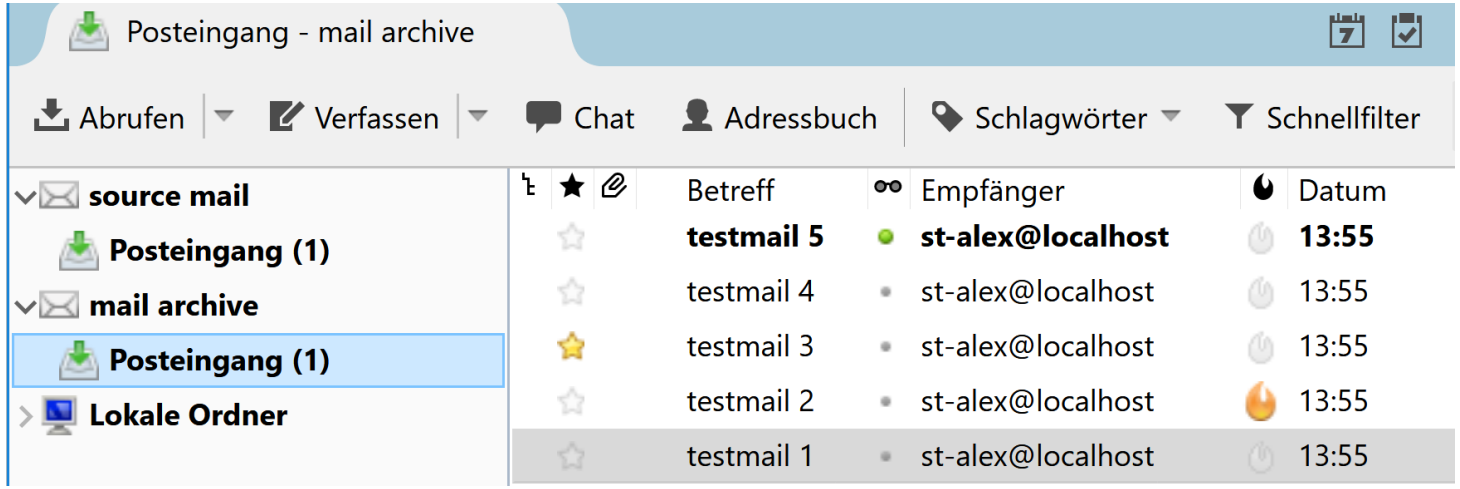



```
doveadm -Dv \
-o imapc_host=192.168.8.230 \
-o imapc_user=st-alex \
-o imapc_password=xxxxxxxx \
-o imapc_ss=starttls \
backup  -R -u rein imapc:
```

- -R: call from archive and pull the mail from source
- Source and destination exactly look the same
- Changes in the destination are overwritten
- Destination should be empty on first call

doveadm 2-way sync

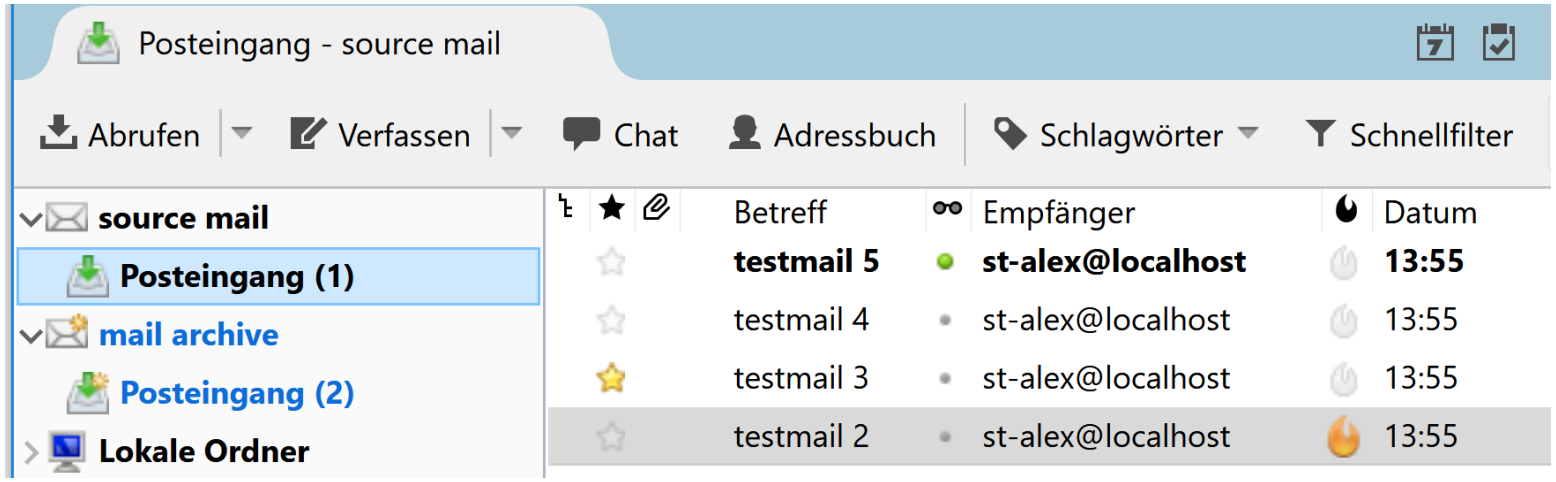


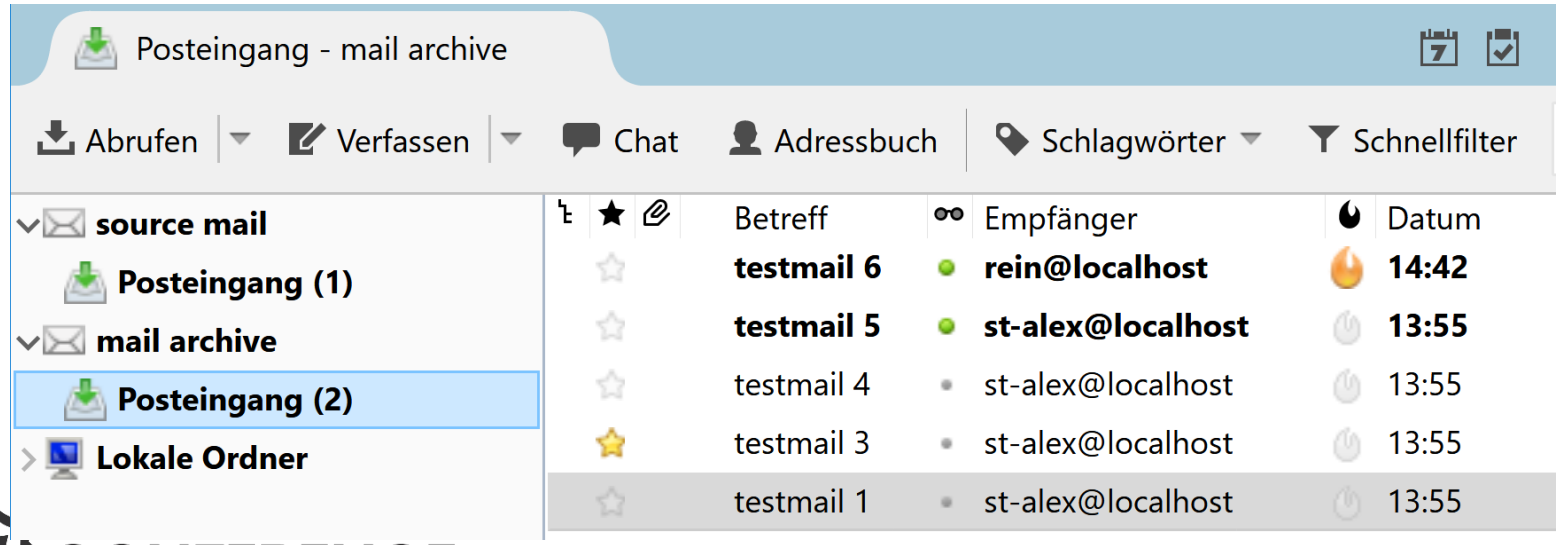



```
doveadm -Dv \
-o imapc_host=192.168.8.230 \
-o imapc_user=st-alex \
-o imapc_password=xxxxxxxx \
-o imapc_ss=starttls \
sync  -R -u rein imapc:
```

- Again source and destination look the same
- But …

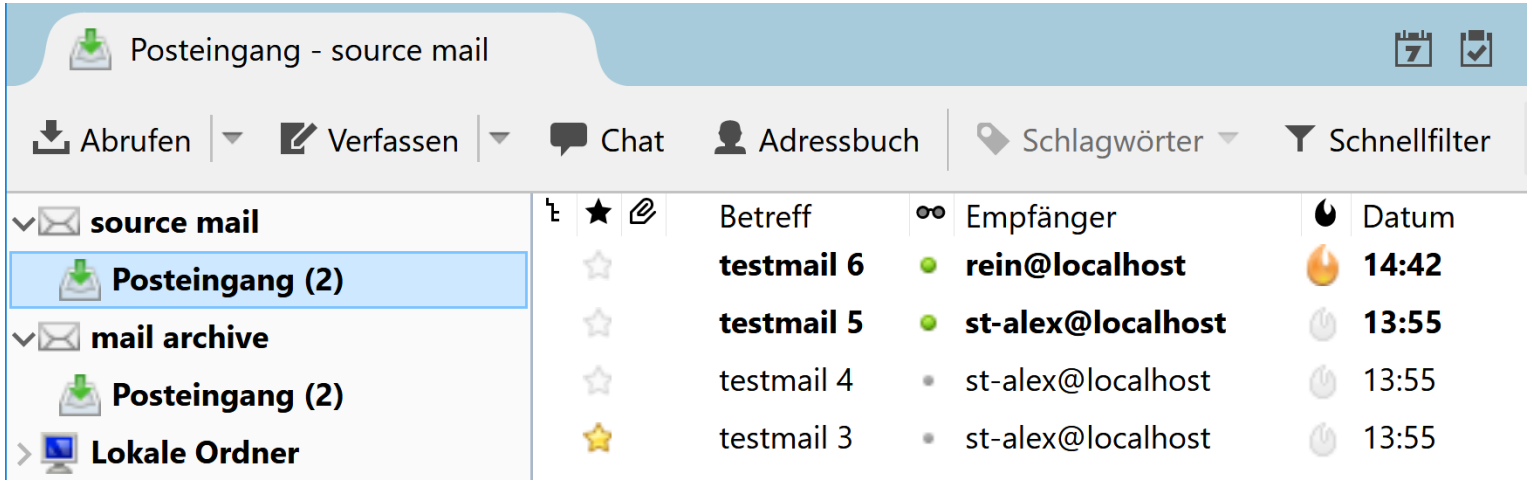
doveadm 2-way sync

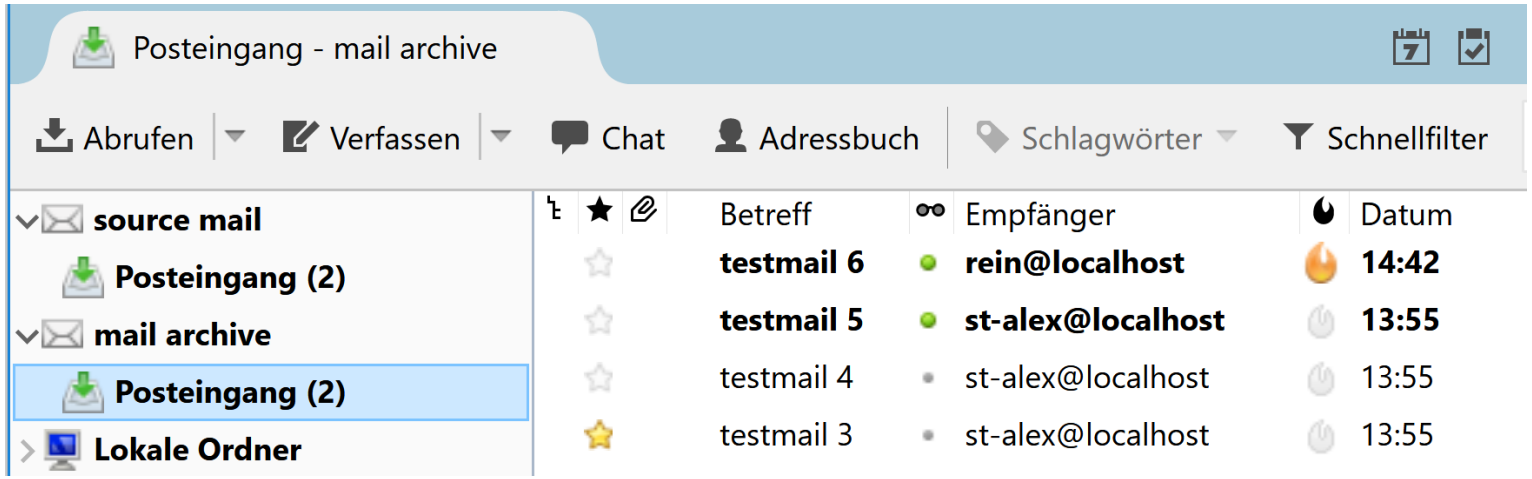




- … you can make changes on both sides
- Delete testmail 1 on source
- Delete testmail 2 on destination
- Receive testmail 6 at destination

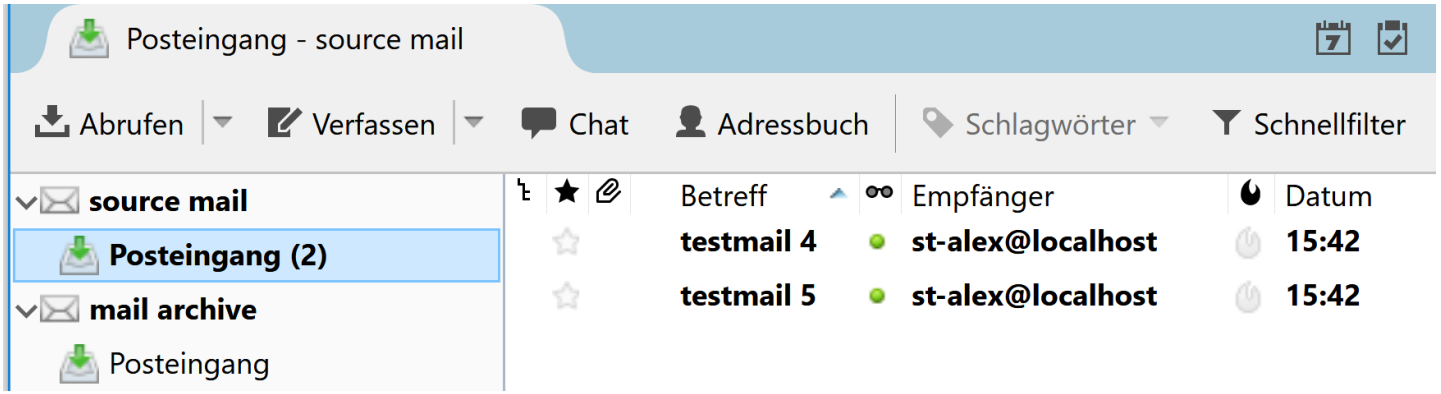
Result of doveadm 2-way sync



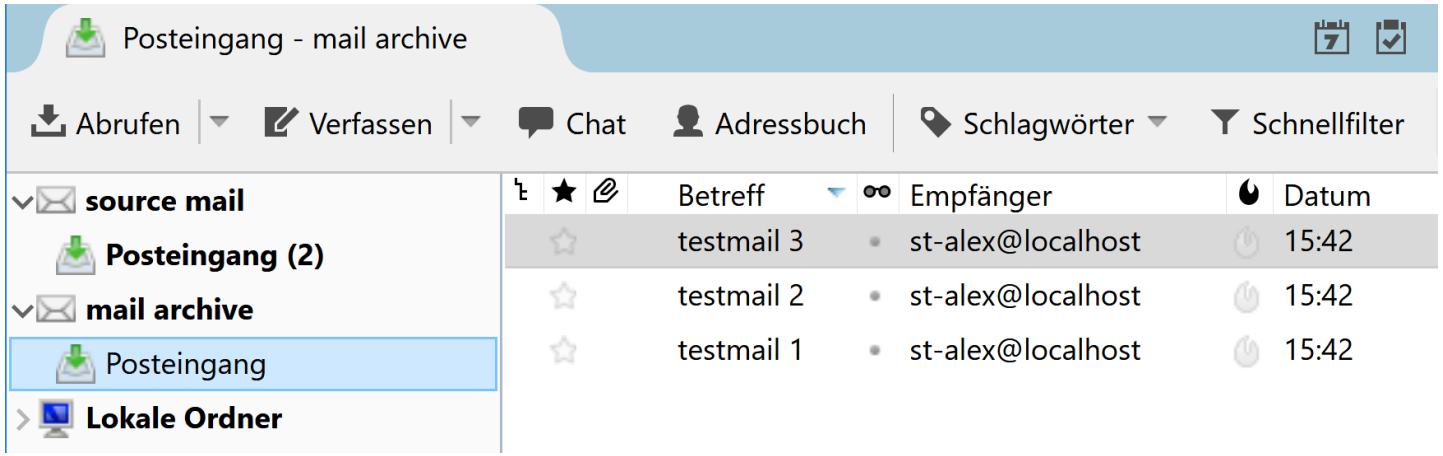


- Work on both sides
- No „master“ side (rsync would need one)
- Index file to keep track of the changes
- Also syncs metadata/flags (read flag, deleted flag, …)

Doveadm sync -1 (one way sync)

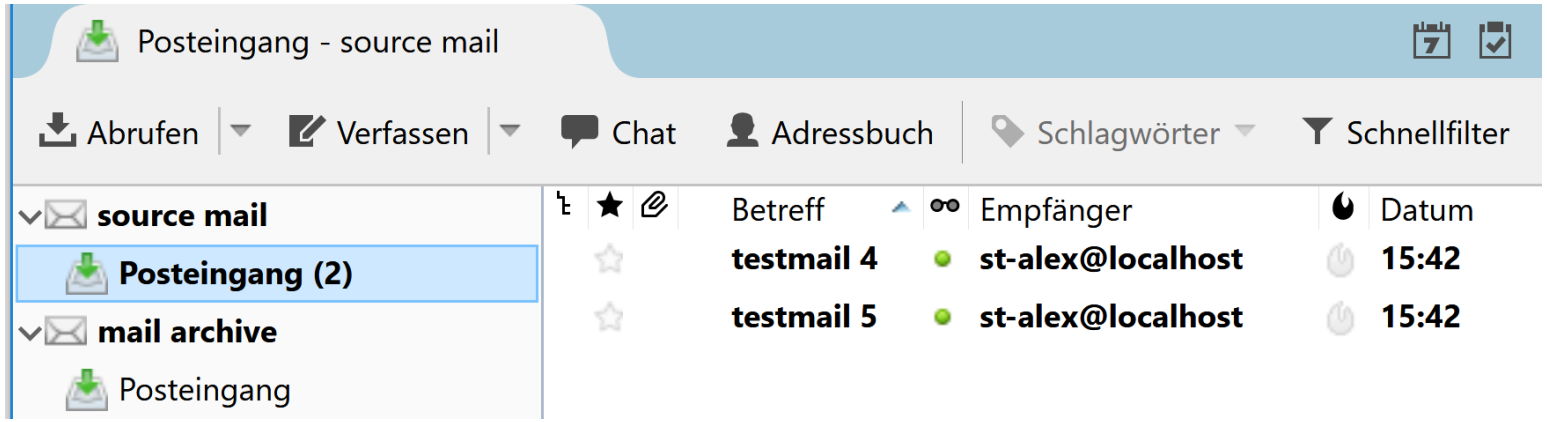


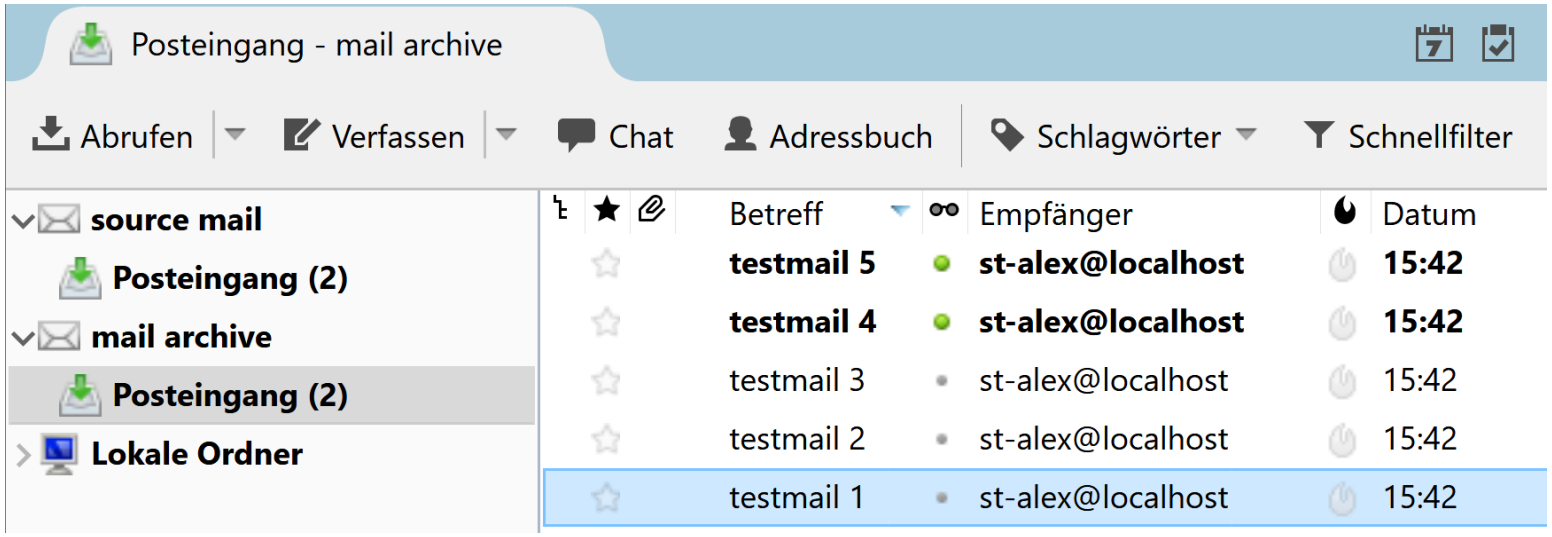

```
doveadm -Dv \
-o imapc_host=192.168.8.230 \
-o imapc_user=st-alex \
-o imapc_password=xxxxxxxx \
-o imapc_ss=starttls \
sync -1  -R -u rein imapc:
```



- Some new mail in the source
- Some „old“ mail in the archive

Doveadm sync -1 (one way sync)





- Source stays the same
- New mail is merged to the archive

Doveadm sync -1 (one way sync)

Before sync	After doveadm sync -1
Receive new mail in the source	New mail is synced to the archive, old mail in the archive is not deleted
Delete mail in the source	Mail is not deleted in the archive
Delete mail in the archive	Mail stays deleted in the archive
Change a mail flag in the source	Change is not synced to the archive
Change a mail flag in the archive	Change is not synced to the source
Receive new mail from Postfix in the archive	Doveadm gets confused, new mails from source are duplicated

- Mail is synched/copied from the source to the archive
- Synch does not delete or change any mail in the archive
- Changes in the archive are not taken back

Doveadm: useful for a mail archive?

- Doveadm backup, 2-way sync: **Snapshot** of mail folders
- Snapshot has to be completed with a general backup strategy and tool (Bacula, Bareos, Amanda, …)
- Doveadm 1-way sync: does not delete mail, may serve as a **standalone** mail archive
- 2-way sync is more suitable if you (also) need a **failover** mail system or need to work on both sides
- Problem of backup granularity: archive might not be complete

Doveadm feasability / requirements for 1-way and 2-way sync

- Use dovecot on source and destination
- Use correct version 2.2 on both sides

Rating: Synchronization approach

	Postfix bcc	doveadm backup	doveadm 2-way sync	doveadm 1-way sync	imapsync
Installation	++	++			
Documentation	++	O			
Configuration	+	O			
Feasibility/ Integration	++/O	+	-		
Useful for archive	+	+	O	+	
Log file	O	-			
Performance					
Legal perspective					

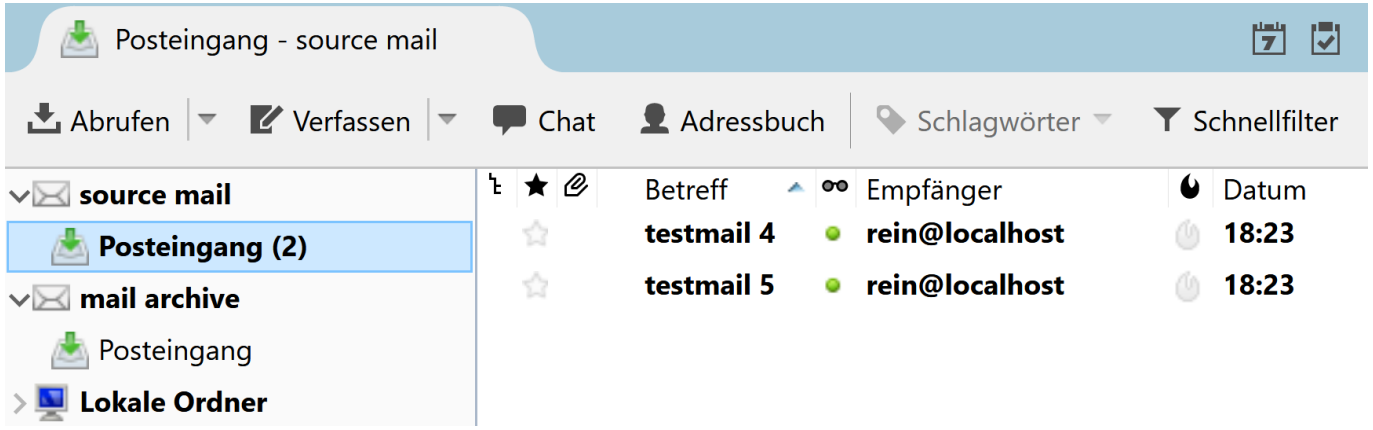
++ = very good	+ = good	O = ok	- = not so good	-- = bad

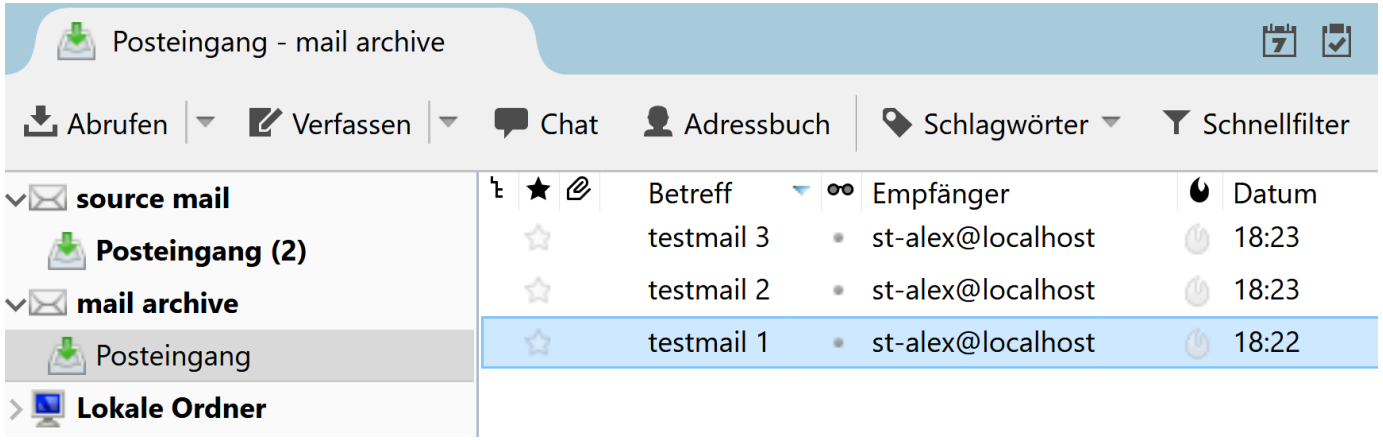
imapsync

- Sync two mail accounts via IMAP
- Configuration/options are very easy:

```
./imapsync \
--host1 test1.lamiral.info \
--user1 test1 \ --password1 "secret1" \
--host2 test2.lamiral.info \
--user2 test2 \
--password2 "secret2"
```

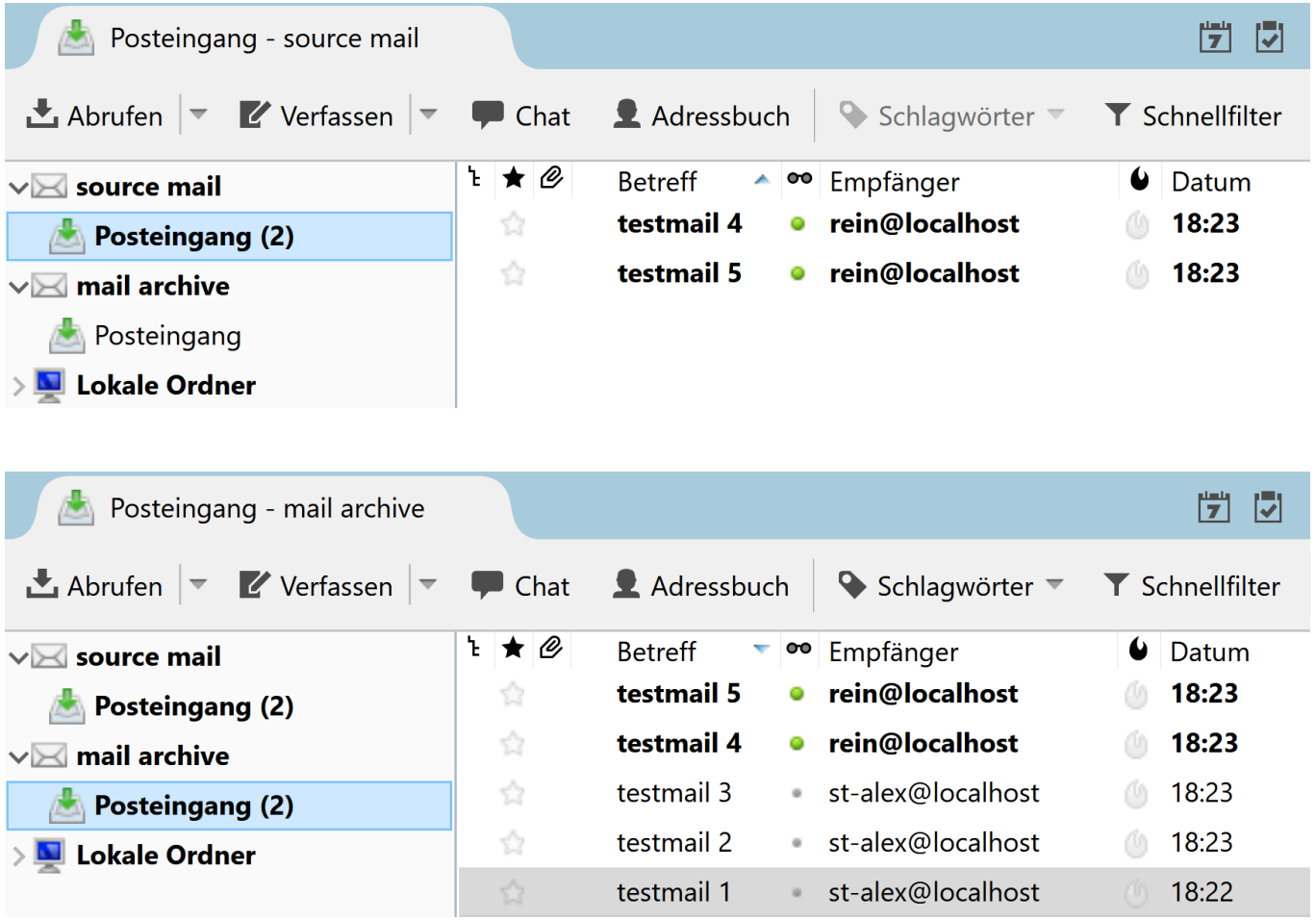
Imapsync testcase





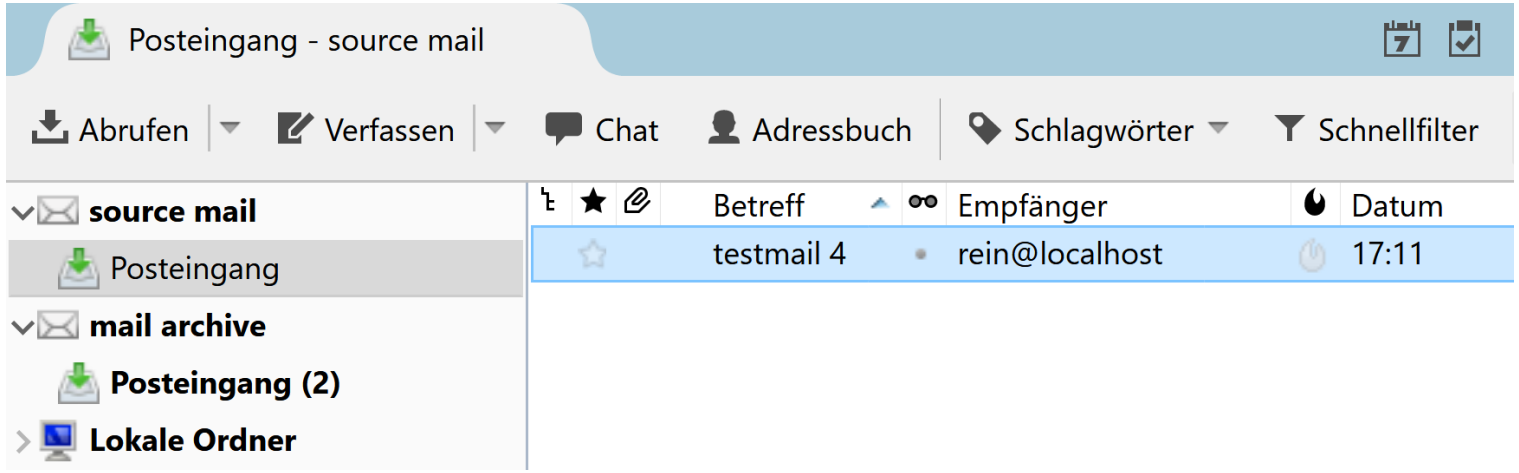
- Some new mail on the source
- Some previous mail on the destination

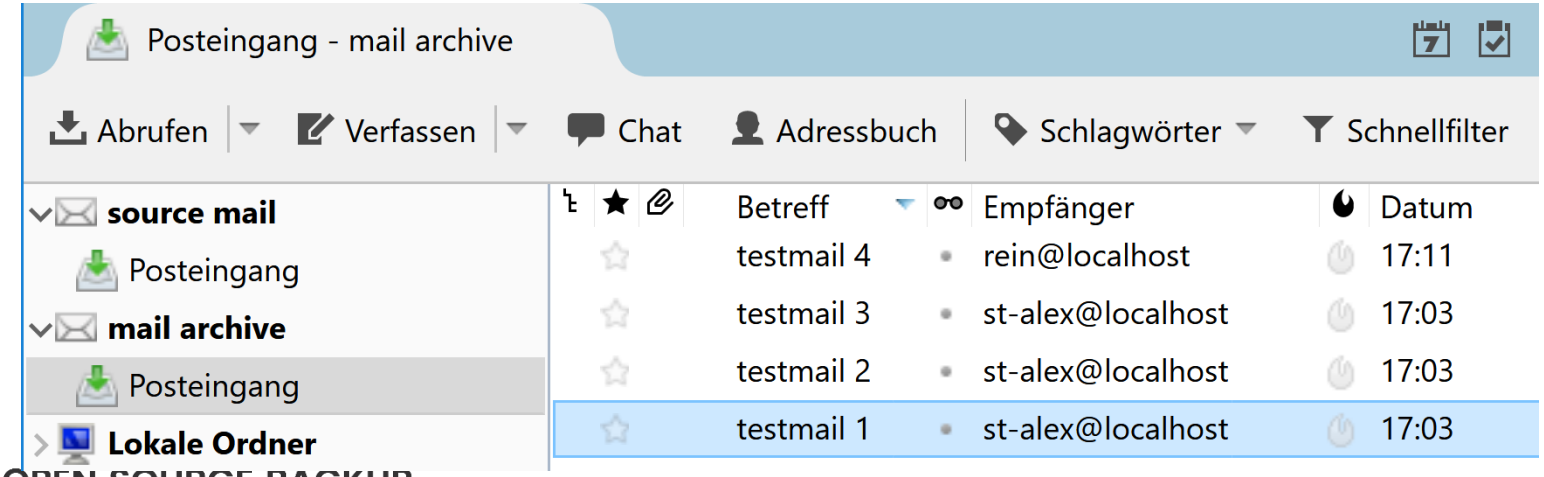
Result of imapsync



- New mail is synced to destination
- Previous mail remains
- Next: Delete testmail 5 on source

Imapsync: result of “delete testmail 5 on source“





- Mail 5 is also deleted in the archive

Imapsync test cases

Before sync.	After sync.
Receive new mail in the source	New mail is synched to the archive, old mail in the archive is not deleted
Delete mail in the source	Mail is deleted in the archive: More precisely it was just a copy of the delete flag
Delete mail in the archive	Mail is recreated in the archive
Change a mail flag in the source	Change is synced to the archive
Change a mail flag in the archive	Change is taken back in the archive
Receive new mail from Postfix in the archive	no change/ no problems

- Imapsync is a backup but does not delete (previous) mail

Imapsync vs. dovecot 1-way

Before sync.	imapsync	Dovecot 1-way sync.
Receive new mail in the source	New mail is synched to the archive, old mail in the archive is not deleted	Same as imapsync
Delete mail in the source	Mail is deleted in the archive/ delete flag is set	Mail is not deleted in the archive
Delete mail in the archive	Mail is recreated in the archive	Mail stays deleted in the archive
Change a mail flag in the source	Change is synched to the archive	Change is not synced to the archive
Change a mail flag in the archive	Change is taken back in the archive	Change remains
Receive new mail from Postfix in the archive	no problems	Doveadm becomes confused, new mails from source are duplicated

- Imapsync is a backup but does not delete (previous) mail
- dovecot 1-way syncs new mail from source only once

Rating: Synchronization approach

	Postfix bcc	doveadm backup	doveadm 2-way sync	doveadm 1-way sync	imapsync
Installation	++	++			+
Documentation	++	O			++
Configuration	+	O			++
Feasibility/ Integration	++/O	+	-		+
Useful for archive	+	+	O	+	+
Log file	O	-			++
Performance					
Legal perspective					

++ = very good	+ = good	O = ok	- = not so good	-- = bad

Test mail repository /setup for performance measurement

- Single user account
- 15474 Mails
- 101 Folders
- 9.177 GiB
- Two Debian servers in a local 1 Gbit/s-Ethernet Network

Performance (measured in Minutes:Seconds)

	doveadm backup	Dovct. 2-way sync.	Dovect. 1-way sync	imapsync
Copy of all mails	3:20 8:18 *	3:30	3:20	36 minutes
Sync a few more	3 seconds	3 seconds	3 seconds	5:24 2:42 **

* with cyrus imapd v2.5.10 mail server as a source
** with option: no foldersizes/foldersizesatend

- Copy of all: imapsync is 4 times slower than dovecot‘s sync: implementation issue
- Diff. sync.: Dovecot‘s index files beat the imapsync

Rating: performance

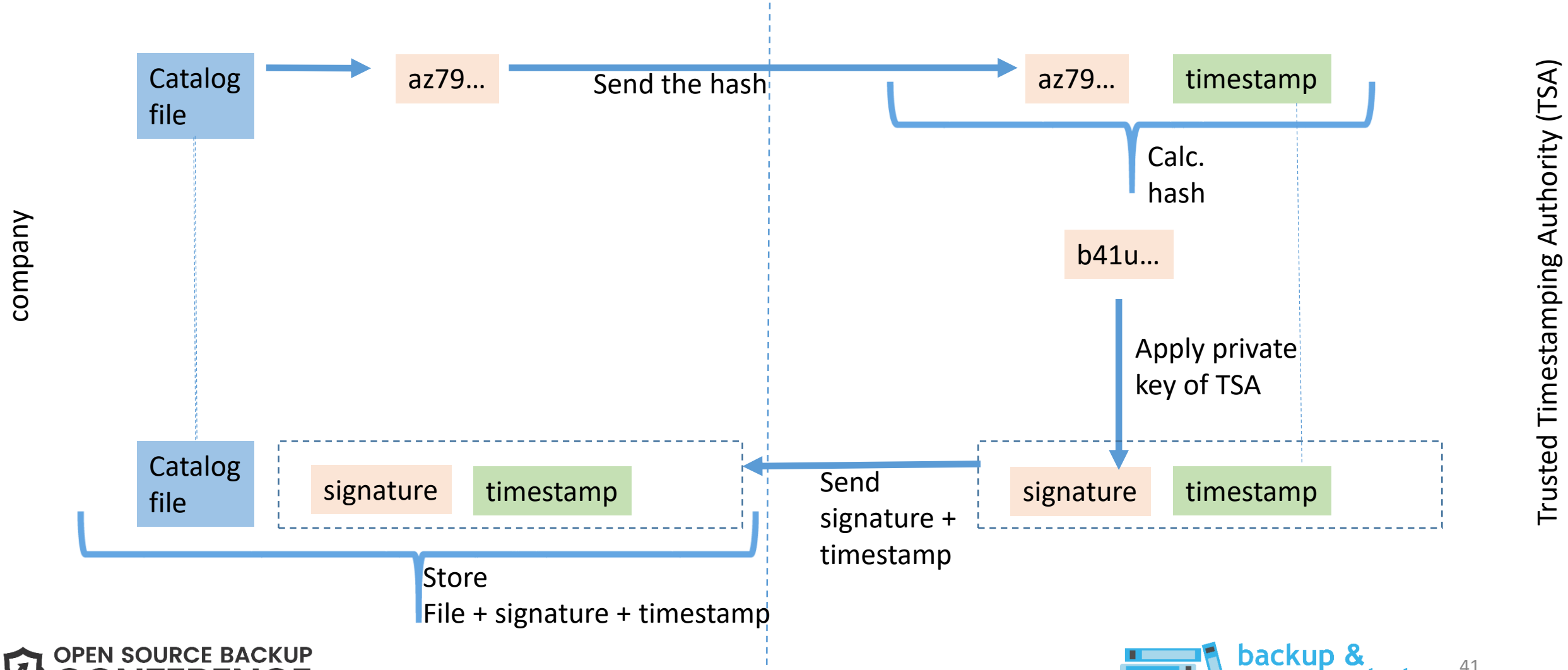
	Postfix bcc	doveadm backup	doveadm 2-way sync	doveadm 1-way sync	imapsync
Installation	++	++			+
Documentation	++	O			++
Configuration	+	O			++
Feasibility/ Integration	++/O	+	-		+
Useful for archive	++	+	O	+	++
Log file	O	-			++
>>Performance	++	+	++	++	--
Legal perspective	?	?			?

++ = very good	+ = good	O = ok	- = not so good	-- = bad

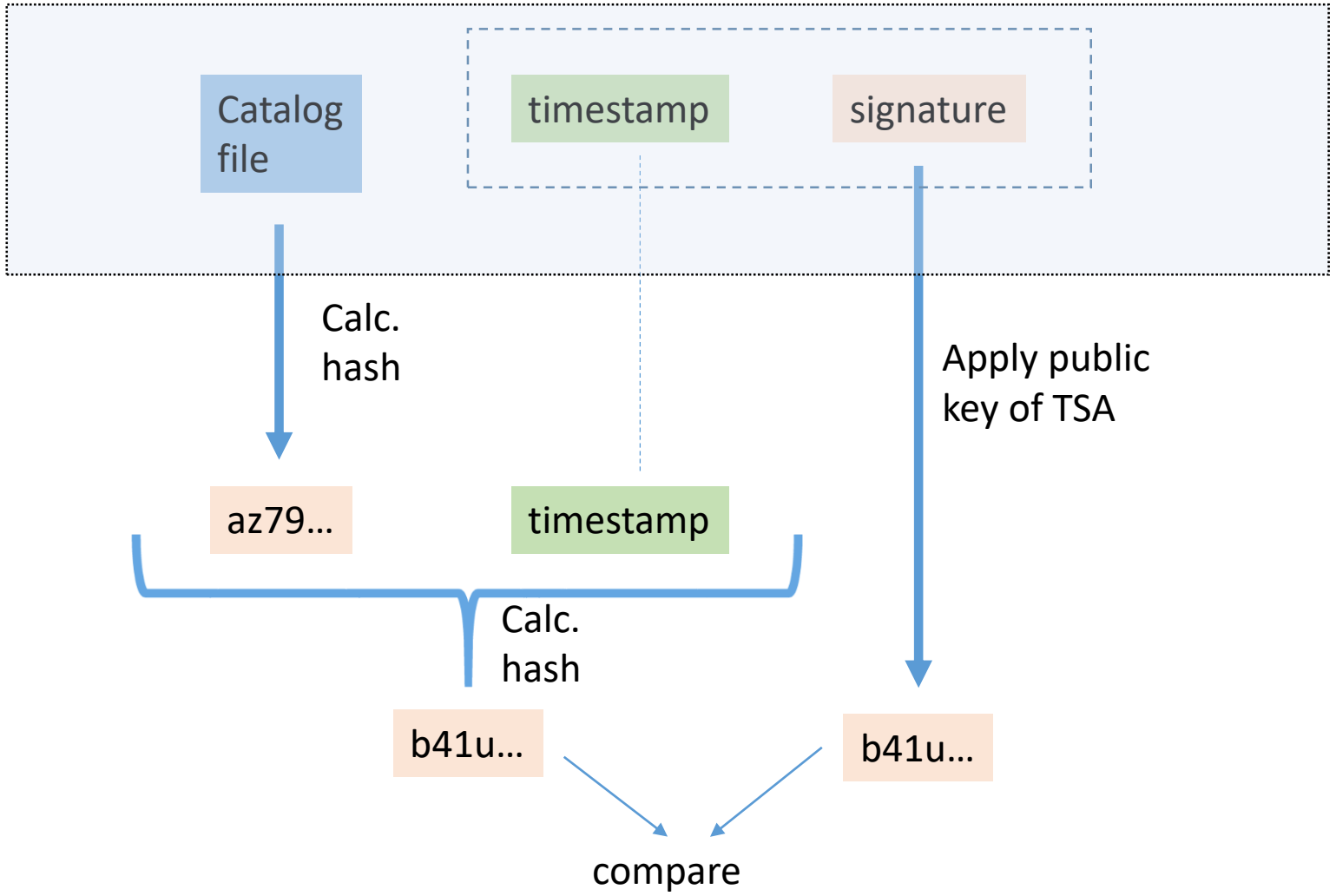
Verification method for auditor

- Allow auditor to verify the data/mail integrity and the mail timestamp
- Mail archive is completed by a regular backup (Bareos, Bacula, …)
- With each backup, create a dump of the backup catalog (contains hash files of all mails)
- Calculate a hash over the dump and send it to a Trusted Timestamping Authority (TSA)

Trusted Timestamping (RFC 3161)



Trusted Timestamping: Verification



The auditor gets this

Art. 17 GDPR: Right to erasure (‘right to be forgotten‘)

1. The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay where one of the following grounds applies:

(a) the personal data are no longer necessary in relation to the purposes for which they were collected or otherwise processed;

(b) the data subject withdraws consent on which the processing is based …

(c) the data subject objects to the processing pursuant to Article 21(1) and there are no overriding legitimate grounds for the processing, …

Art. 17 GDPR: Right to erasure

- Not yet clear to which extend it will apply
- No consistent strategy in commercial solutions (Mailstore: function to delete selected messages, Benno Mailarchive: not yet implemented)
- Retention Policies: archive & delete mails from production mail server
- Suitable options to cleanup (or to sort) the mail archive = ?

Option 1: Script to cleanup Maildir directory

- Need for own script
- Parse a mail for specific information, e.g. some recipient field:

```
# sudo apt-get install procmail
# cat testnachricht | formail -x To

 <rein@localhost> (terminal output)
```

Option 2: Select messages with imapsync

Option	Example
Sync messages by date	imapsync ... --search "SENTSINCE 1-Jan-2010" imapsync ... --search "SENTBEFORE 31-Dec-2010" imapsync ... --search "SENTSINCE 1-Jan-2010 SENTBEFORE 31-Dec-2010"
Sync messages less than 2 days old ~ more than 2 days old ~ more than 30 days and less than 365 days	imapsync ... --maxage 2 imapsync ... --minage 2 imapsync ... --minage 30 --maxage 365
Search commands: Many options to parse mail header or body, e.g. TO, BCC, SUBJECT, CC, BODY, …	
Sync an account to a folder of a different user	imapsync ... --user1 user1 \ --subfolder2 archive-of-user1

- Many other options

Rating:

	Postfix bcc	doveadm backup	doveadm 2-way sync	doveadm 1-way sync	imapsync
Installation	++	++			+
Documentation	++	O			++
Configuration	+	O			++
Feasibility/ Integration	++/O	+	-		+
Useful for archive	+	+	O	+	+
Log file	O	-			++
Performance	++	+	++	++	--
>> Legal perspective	++	O			+

++ = very good	+ = good	O = ok	- = not so good	-- = bad

Final Rating:

	Postfix bcc	doveadm backup	doveadm 2-way sync	doveadm 1-way sync	imapsync
Installation	++	++			+
Documentation	++	O			++
Configuration	+	O			++
Feasibility/ Integration	++/O	+	-		+
Useful for archive	+	+	O	+	+
Log file	O	-			++
Performance	++	+	++	++	--
Legal perspective	++	O			+

++ = very good	+ = good	O = ok	- = not so good	-- = bad

Final Thoughts: Is a self-build mail archive competitive with commercial systems, e.g. Benno mail archive

- Yes and no …
- Concepts are the same,
- But Benno has some more features: improved search, user management with Active Directory or LDAP, web service interface
- Use of CLI and Linux knowledge still required
- For a quick setup Benno might be the better choice
- Beyond the technical questions still many organizational questions to be answered

Mail archive Future Work

- User access and account management
- Spam filter management
- Encryption of mail text and attachments
- Documentation of changes in the archive

Thanks !

- Questions & Answers

Appendix