Debian/GNU Linux Mailing

Overview of the Mailing

Károly Erdei

14. November 2008

Agenda

Agenda

Mailing

Sending e-mails across the Internet

Basics, Terminology

- Message transfer between hosts (Mail gateways)
  - Mail gateway: dedicated computers to process and transfer e-mails
  - MTA - Mail Transfer Agent: sendmail, exim, postfix..
  - Protocol: SMTP - Simple Mail Transfer Protocol (RFC 821, 1982)
- Message retrieval by mail user agent (MUA)
  - MUAs: Thinderbird, xfmail, pine, etc.
  - POP3: Post Office Protocol, version 3
  - IMAP: Internet Message Access Protocol, version 4
- Representation of messages
  - RFC 822: Basic Message Format (7-bit text only)
  - MIME: Multipurpose Internet Mail Extension (1992)
  - S/MIME: Secure MIME; PGP/MIME: Pretty Good Privacy

Structure and meaning of the e-mail address

E-Mail address: name@domain

- name: real name, symbolic name, alias, mailbox name
  - example: john.shaw, secretary, research, johnny
  - mailbox: the place where the messages on the receiving mail gateway will be stored in formats mbox or maildir
  - mbox format: the messages will be stored in one file, new message will be appended
  - maildir: each message will be stored as a separate file in the directory
  - alias: an alternative name which translates to the name of the mailbox
- domain: DNS domain name
  - defines MX resource record which host deliver the messages to
  - there can be more mail exchangers (mail gateways) for the domain

```
;; QUESTION SECTION:
;risc.uni-linz.ac.at.                IN      MX
;; ANSWER SECTION:
risc.uni-linz.ac.at.   1363   IN   MX   20 bullfinch.risc.uni-linz.
risc.uni-linz.ac.at.   1363   IN   MX   30 grauwal.risc.uni-linz.ac
```

e-Mail Transfer Process

Message transfer process in overview

- User sends a message to the local mail gateway (e.g. RISC) by the MUA (e.g. Thunderbird)
- Local mail gateway
  - first spools the message locally in the spool area `/var/spool/mqueue`
  - after transfers message from the spool area to the recipients (remote) mail gateway
- Local mail gateway receives a message from the mail gateway of remote senders
- Received message is placed into the mailbox of the user on the local mail gateway
- User downloads the message (e.g. by Firefox, POP) from the local mail gateway to laptop or home directory

e-Mail Transfer Process - gateways

SMTP speaking processes on mail gateways transfer mails

Process for background transfer is client of remote mail server

- Uses DNS to dermine name of mail exchanger for destination domain
  - DNS responses MX resource record for domain
- Uses DNS to map name of mail exchanger to IP address
- Creates TCP connection to server process on mail gateway
- Transfers copy of message to server, which stores copy in mailbox
- If transfer succeeds, sending process removes copy from mail queue
- If connection cannot be established, records time of transfer attempt in the mail queue entry for the message and terminates

Process for background transfer periodically sweeps spool area

- Spooled mail can be delivered as soon as mail exchanger is up again
- If spooled mail cannot be delivered after some extended time (default 5 days), process returns message to sender

Agenda

Mailing, Internet Standards (STDs)

STD 10 / RFC 821: Simple Mail Transfer Protocol

- Specifies how messages are passed from one host to another
- Communication is based on readable ASCII text commands

```
R: 220 uhu.risc.uni-linz.ac.at ESMTP Sendmail 8.13.8
S: HELO sender hostname                    R: 250 OK
S: MAIL FROM: <e-mail address>             R: 250 OK
S: RCPT TO:   <e-mail address>             R: 250 OK
S: DATA
R: 354 Start mail input; end with <CRLF>.<CRLF>
S: <CRLF>.<CRLF>                           R: 250 OK
S: quit                                    R: 221 name closing
```

- other commands: VRFY Smith; EXTN secretary
- Mail server/clients understand extended version (ESMTP)
  - ESMPT is requested by client via `EHLO` instead of `HELO`
  - ENHANCEDSTATUSCODES, 8BITMIME, AUTH DIGEST-MD5 CRAM-MD5 PLAINE, STARTTLS

Mailing, Internet Standards (STDs)

STD 11 / RFC 822: Basic Message Format

- 7-bit ASCII format, primarily for english text
  - only plain text, binary must be converted
- Mail header and Mail body is separated by an empty line
  - Mail header begins with a From line in mbox format
- Mail header - User Fields - provided by MUA
  - From: To: Cc: Subject: Sender: Bcc:
  - Bcc: not visible im header
  - All of them can be set by most of the MUA: used by spammers, they fake the header lines
- Mail header - Automatic Fields - provided by MUA,MTA
  - Date: Message-Id: Return-path: Received:, X-fields
  - Received: can follow the mail gateways as e-mails pass them

MIME: Multipurpose Internet Mail Extension

How to send other contents as ASCII text

RFC 1341: Messages in other character sets and with binary contents

- Use RFC 822 basic message format
  - MIME messages can be transferred by normal (older) SMTP agents
  - Only mail reader/writer (MUA) must be MIME enabled
- Define additional header fields:
  - MIME-Version: , Content-Id:
  - Content-Transfer-Encoding: How content is encoded as ASCII
  - Content-Type: MIME-type of content
  - Content-Description: Human-readable description of content
- Content-Transfer-Encoding:
  - 7-bit, Quoted-Printable, Base64 (for binary data); 8-bit; Binary
- Content-Type: 7 MIME types with multiple subtypes
  - Text, Image, Audio, Video, Application, Message, Multipart,
- Content Subtypes: text/plain, text/richtext, message/rfc822
  - application/octet-stream, application/PostScript multipart/mixed, multipart/alternative

e-Mail Security

Email is not a secure communication medium

- Reliability: messages may be lost
  - Only transfer from mail queue to next mail server is guarantueed
  - User may be asked to confirm receipt of a message
  - Header field `Disposition-Notification-To:` *address*
- Privacy: messages may be read by unauthorized persons
  - Messages are transferred in clear text
- Authenticity: message sender may be faked
  - It is easy to create messages with faked `From:` fields
- Integrity: message content may be changed
  - Intermediate transfer agent may modify message
- Integrity, Authenticity, Privacy achived by cryptographic methods
  - Privacy: by Encryption
  - Integrity, Authenticity: by Digital signatures

Emails are as secure as postcards are without cryptographic methods

Agenda

SPAM living with it

What is SPAM

- nearly identical messages sent to numerous recipients by e-mail
- any email message where the senders identity is forged
- common synonyms for spam
  - UBE: unsolicited bulk e-mail
  - UCE: unsolicited commercial e-mail

Problems with SPAM

- contains an attachment which is a virus/trojan
  - to became your Windows PC a `bot net` host
- phishing: spam ask users to enter personal information on fake Web sites using e-mail forged to look like it is from a bank or other organization such as PayPal
- spoofing: your e-mail address used as sender of spam
  - you get all bounced mails (500-5000 in short time)
- spam contains links to advertised/malicious web sites

SPAM living with it

How spammers work

- collecting e-mail addresses
  - from chatrooms, websites, newsgroups
  - infecting Windows PCs, where viruses collects address books
- sending spam mails
  - using open mail gateways (not anymore)
  - using bot nets, by infecting Windows PCs with viruses, Trojans
- dictionary attacks
  - spammer sends e-mail based on dictionary
  - 150.000 rejected by blacklists + 40.000 dictionary attack

Main problem to fight SPAM

- governments did not accept appropriate law again spammers, SPAM
- law only in some countries: in EU, Australia
- EU: SPAM for direct marketing are not allowed without the consent or in respect of the subscriber (receiver of spam)

SPAM living with it

Origin of spam, statistics

Origin of SPAM

- Origin of spam refers to the geographical location of the computer from which the spam is sent
- the spammer, the hijacked spam-sending computer, the spamvertised server, and the user target of the spam are all often located in different countries
- As much as 80% of spam received by Internet users in North America and Europe can be traced to fewer than 200 spammers

Statistics

- total volume of spam April 2008: over 100 billion emails/day
- 90% of incoming email traffic is spam in NA,EU, Australasia
- 96.5% of e-mail received by businesses was spam by June 2008

Living with SPAM

Statistics

Origin of spam in the third quarter of 2008

* The United States (18.9%, up from 14.9% in Q2)
* Russia (8.3%, up from 7.5%)
* Turkey (8.2%, up from 6.8%)
* China (5.4%, down from 5.6%)
* Brazil (4.5%, unchanged)

When grouped by continents, spam comes mostly from:

* Asia (39.8%, up from 35.4%)
* Europe (23.9%, down from 29.5%)
* North America (21.8%, up from 18.2%)
* South America (13.2%, down from 14.8%)

Number of IP addresses used for spamming

* top three as the United States, China, and Russia
* followed by Japan, Canada, and South Korea

Agenda

Antispam techniques I

What the end user can do

Give your e-mail address only to trusted persons/sites

- never put your e-mail address in text form to a web site
- post to lists as anonym, use faked, invalid email address and name
- avoid responding to spam
  - dont use links: remove me from the list (you'll confirm your e-mail address)
  - be carefull with your vacation message: you can send a reply to a spammer
- don't use contact forms on web sites: (problems with server side scripting)
- don't register anywhere with real e-mail address (I hope, amazon.de is ok, but other sites ?)
- use temporary e-mail addresses (if possible)
  - the e-mail address (alias) expires after a given time

Do NOT read and send HTML emails

Antispam techniques II

Be careful using and configuring your mail program

- don't use html in e-mail programs (MUA) !
  - set the outgoing mail format to PLAIN TEXT
  - for an e-mail message it is not necessary to use html
  - you can use any type of attachment (to send .doc, .jpg, etc. files)
- RISKs by reading HTML formatted e-mails
  - mail client starts a browser or the function of browser is integrated
  - html browser interpret the contents automatically (check settings)
  - they start scripts, download, show images, without asking you
  - html spam can contain scripst, which allow spammer to spy your computer (address, etc) spyware will may be installed
  - html spam can contain web bugs, which allow spammer to get further information from you, save viruses, Trojans, you became a bot net host, etc.
- mail clients which don't display html, attachments, images have fewer risk !

Antispam techniques
Using SpamAssassin (SA)

SA - email spam filtering based on content-matching rules

- uses a variety of spam-detection techniques
  - DNS-based and checksum-based spam detection
  - Bayesian filtering, blacklists and online databases
- can be integrated with the mail server
  - to automatically filter all mail for a site
- avarded: Linux New Media Award 2006
  - Best Linux-based Anti-spam Solution

Operation

- uses large set of rules to decide e-mail is spam or ham
  - rule (called test in SA) bases mostly on regular expression to decide spam contents
- each test has a score value to assign to e-mail, if matches
  - positive value indicates spam, negative ham
- all test combine the results in a global score

SpamAssassin

The RISC setup, how to use it

How to tune the default configuration

- all e-mails at RISC will be checked by SA
- you can use procmail to sort your e-mails in folders
  - to learn: man procmail; man .procmailrc;
  - RISC User Guides: How to configure SpamAssassin for your needs
- configuration file: .spamassassin/user_prefs
- you can change, tune the values for different variables:
  - change the value for required_hits 3.0
  - change the score value for the different checks by setting it to zero
- configure your white list or blacklist
  - `whitelist_from @jku.at`
  - this will let through all e-mails with @jku.at, faked addresses, too !
  - `blacklist_from @hknetmail.com`
- use sa-learn to tune the Bayesian algoritm

```
sa-learn --spam --mbox /path/to/spaminput
```

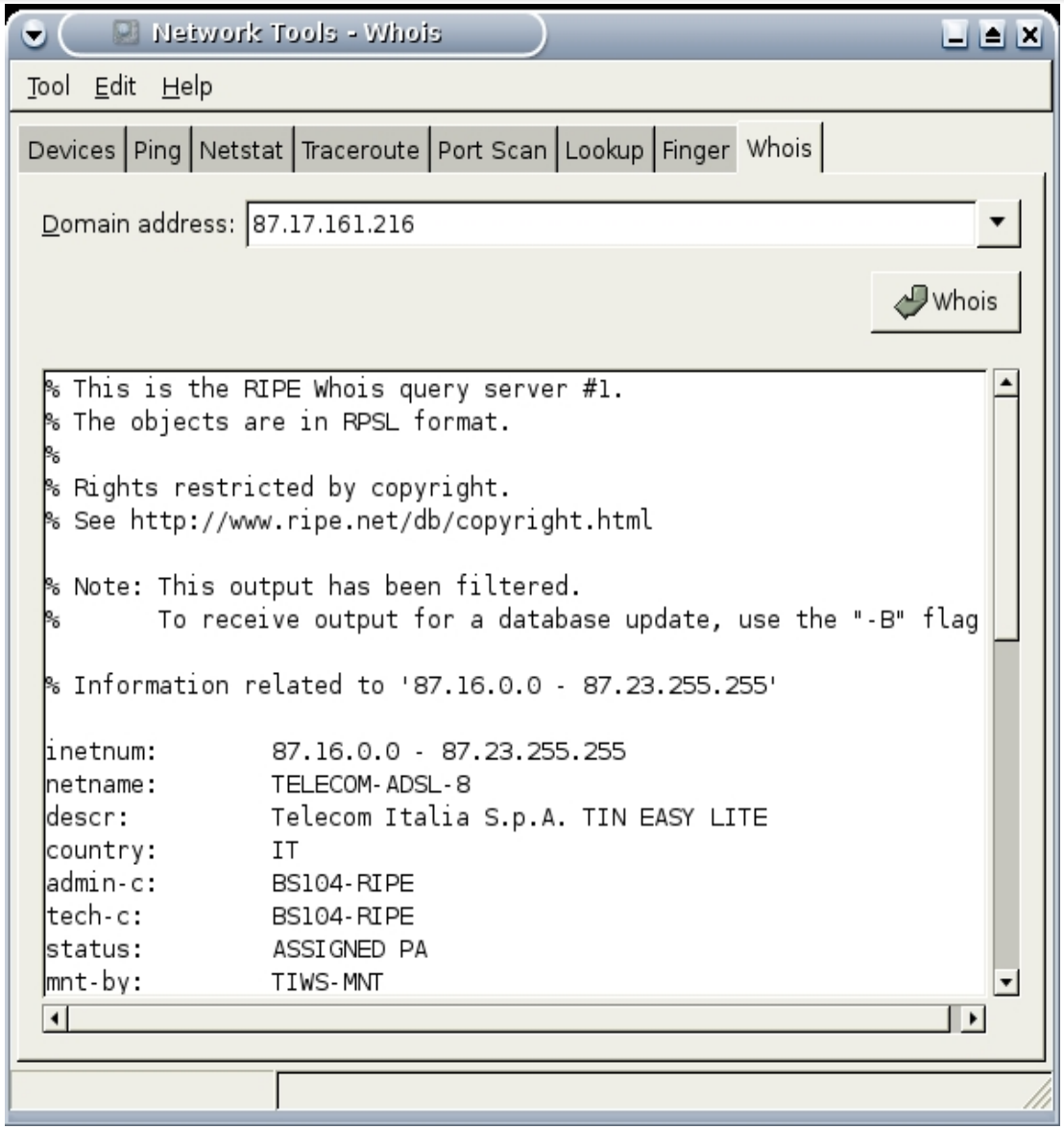
e-Mail Header Lines

Examples

```
X-Spam-Status: No, hits=-86.9 required=2.5
 tests=BAYES_70,CLICK_BELOW,EXCUSE_1,EXCUSE_19, + 10 other TESTs
 USER_IN_WHITELIST version=2.55
Content-Type: text/html
Message-Id: <200811111208.mABC8ZBA085692@mail2.edvz.uni-linz.ac.at>
Date: Tue, 11 Nov 2008 13:08:36 +0100 (CET)
Received: from host216-161-dynamic.17-87-r.retail.telecomitalia.it
 (host216-161-dynamic.17-87-r.retail.telecomitalia.it [87.17.161.216])
 mail2.edvz.uni-linz.ac.at (8.13.4/8.13.3) with ESMTP id mABC8ZBA085692
 for <karoly.erdei@jku.at>;
 Tue, 11 Nov 2008 13:08:41 +0100(CET)(envelope-from karoly.erdei@jku.at
Received: from mail2.edvz.uni-linz.ac.at (mail2.edvz.uni-linz.ac.at
 [140.78.3.69]) by bullfinch.risc.uni-linz.ac.at (8.13.8/8.13.8/Debian-
 ESMTP id mABC8qq2019484 for <k.erdei@risc.uni-linz.ac.at>;
 Tue, 11 Nov 2008 13:09:02 +0100
From: karoly.erdei@jku.at
To: karoly.erdei@jku.at
Subject: [SPAM?] [HIGH] Showing your home video to all
```

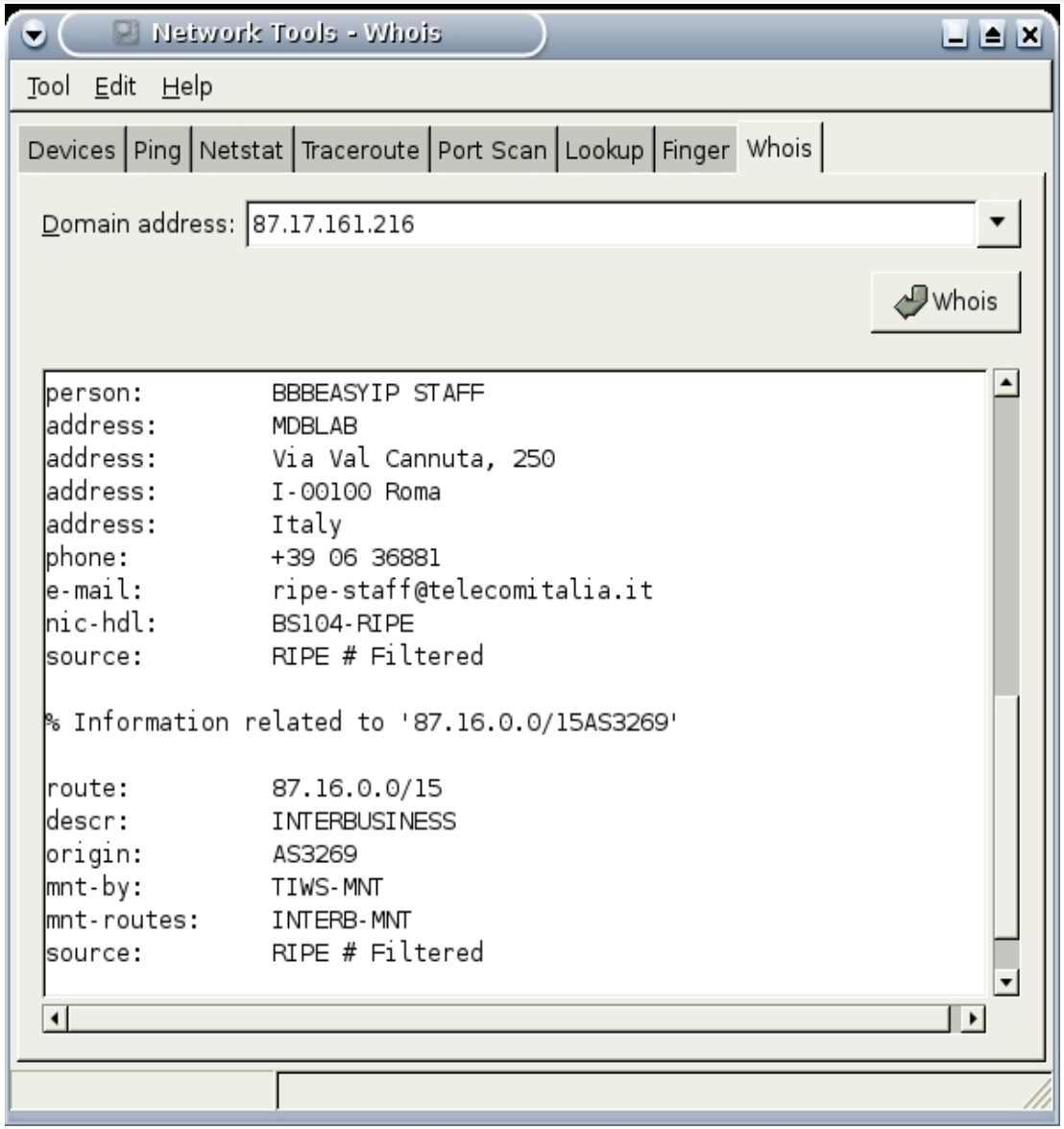
Checking network based on Spammer IP

gnome-nettool - Whois



Checking network based on Spammer IP

gnome-nettool - Whois



Agenda

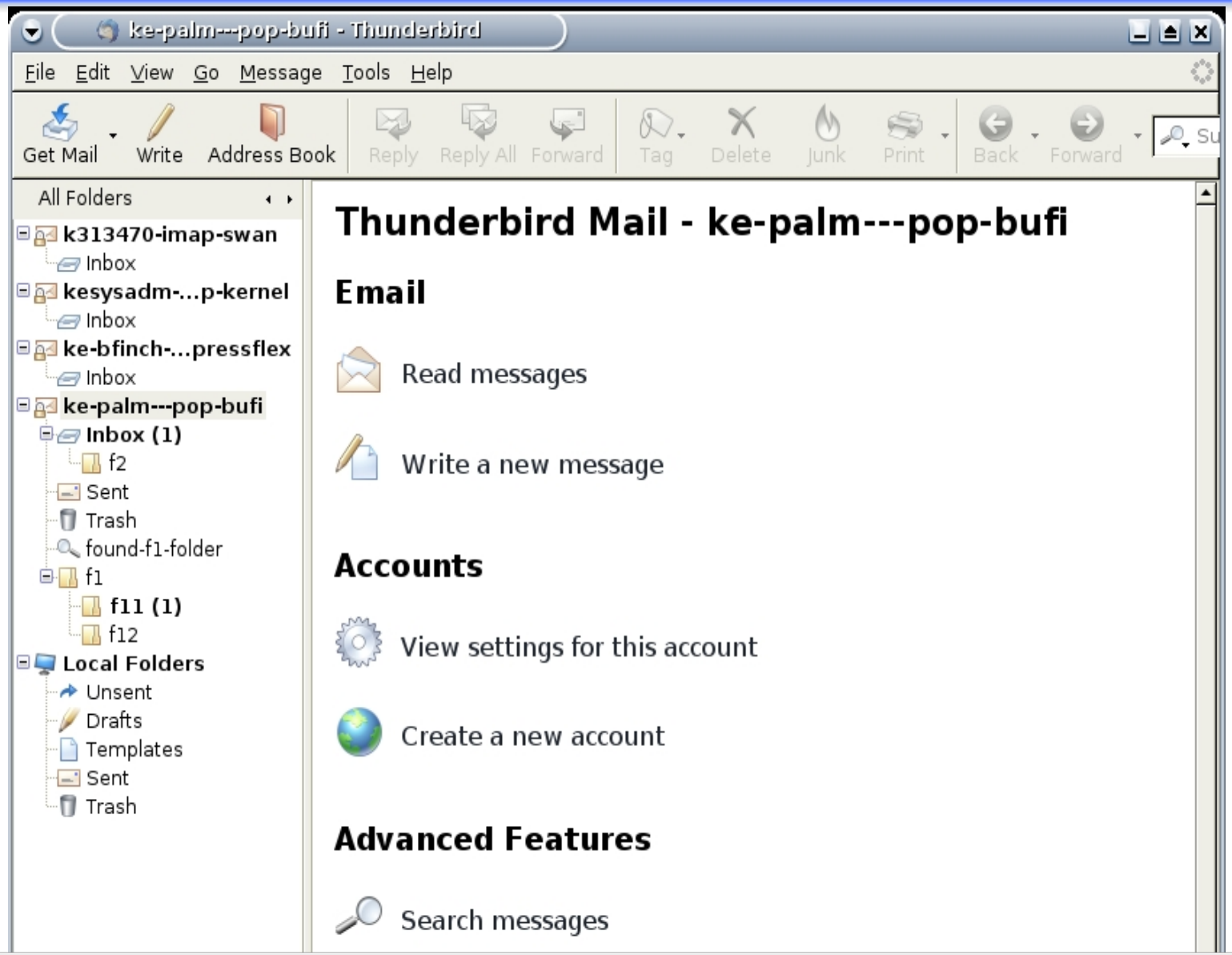
Mozilla Thunderbird

current version 2.0.0.17

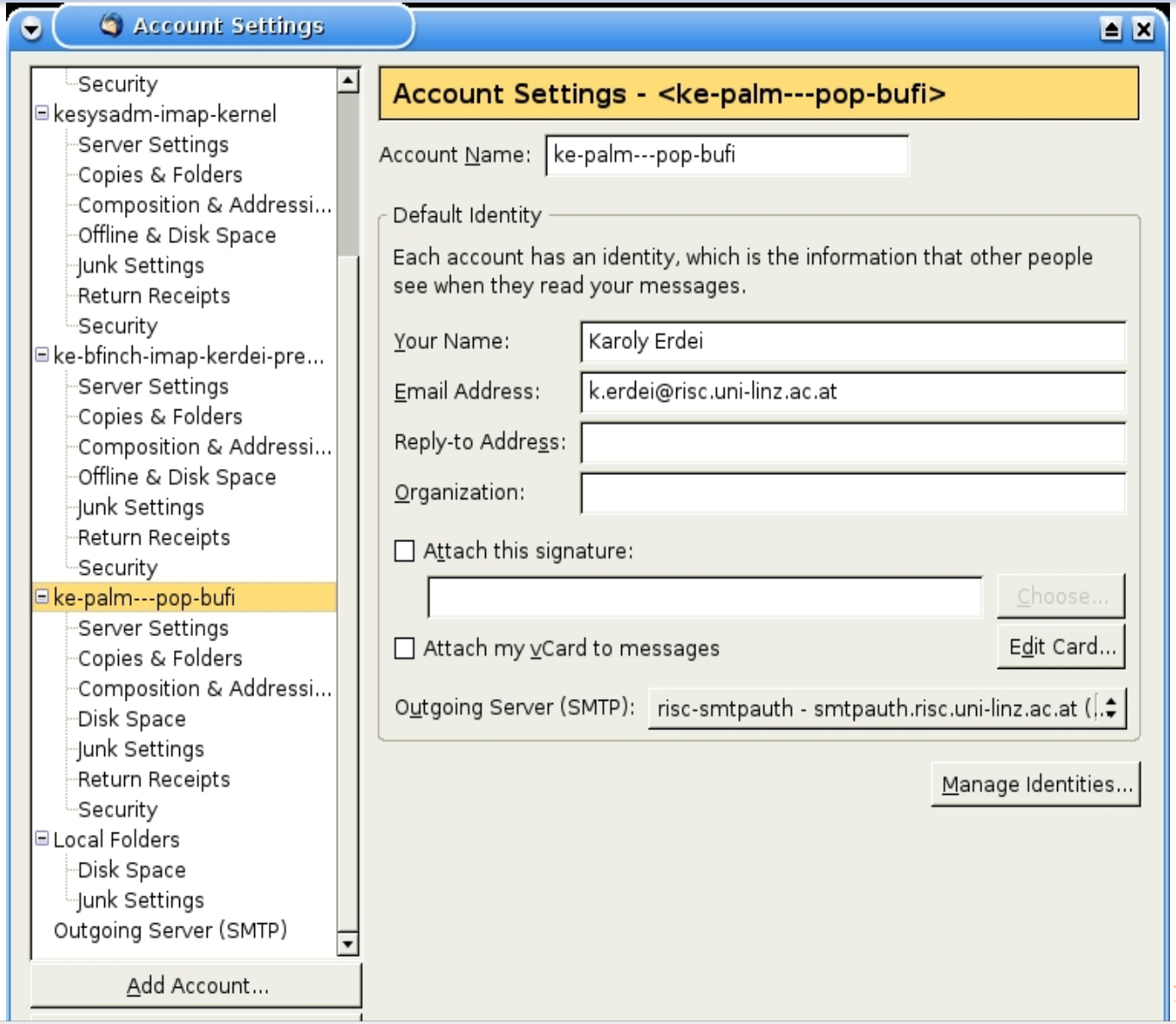
Thunderbird is the best free MUA

- free, open source, cross-platform e-mail and news client
- supports multiple e-mail, newsgroup and RSS accounts
  - supports multiple identities within accounts
- Spam mail filtering
  - own Bayesian spam filter
  - whitelist, based on the included address book
  - understands the classifications of SpamAssassin
- Standars supported natively
  - POP and IMAP with SSL/TLS,
  - S/MIME secure email (digital signing and message encryption using certificates)
  - PGP signing, encryption, and decryption by the Enigmail extension
- Security protection includes
  - disabling loading of remote images within messages
  - disabling JavaScript
- Additional features are available via extensions

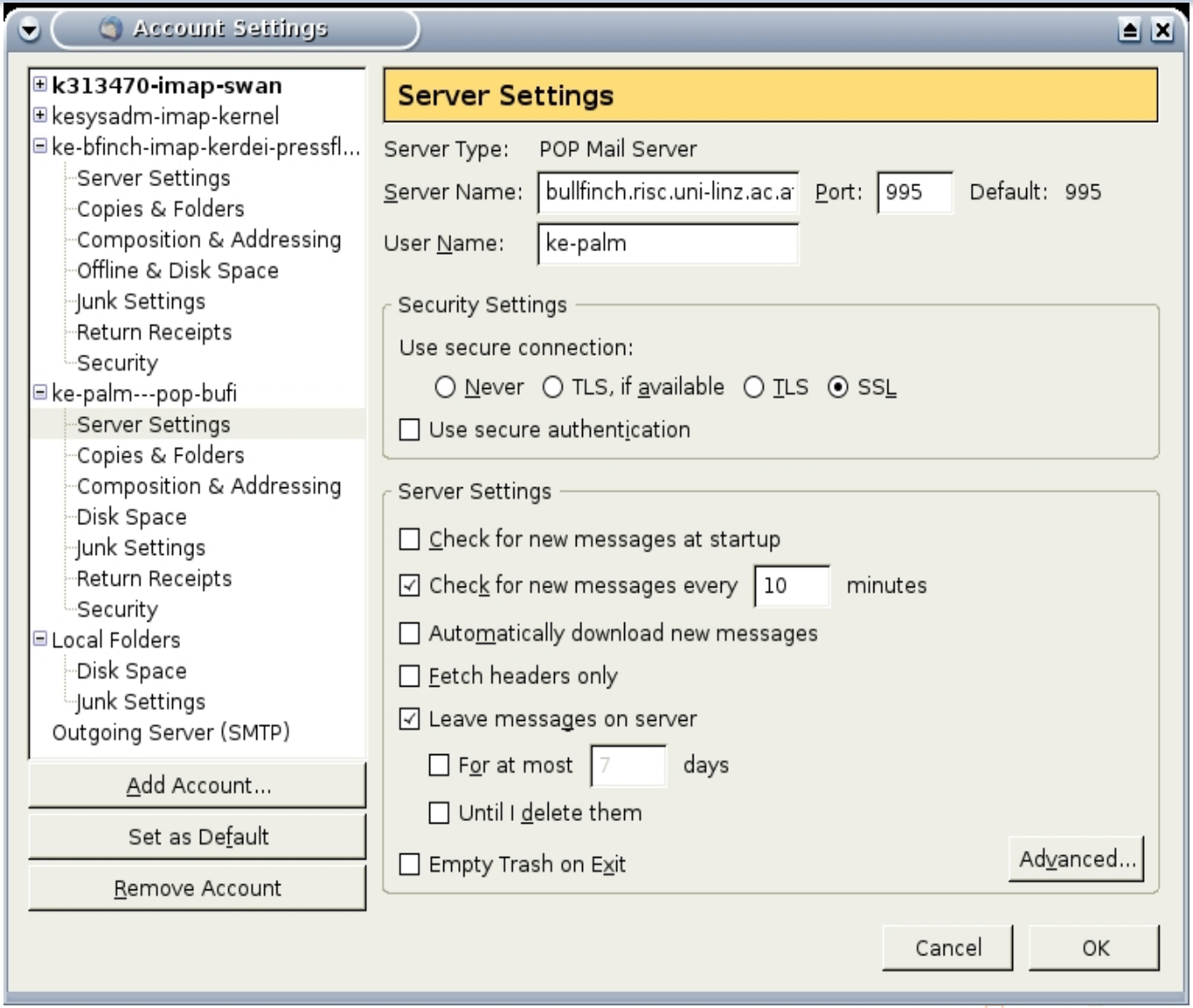
Main Window of Thunderbird



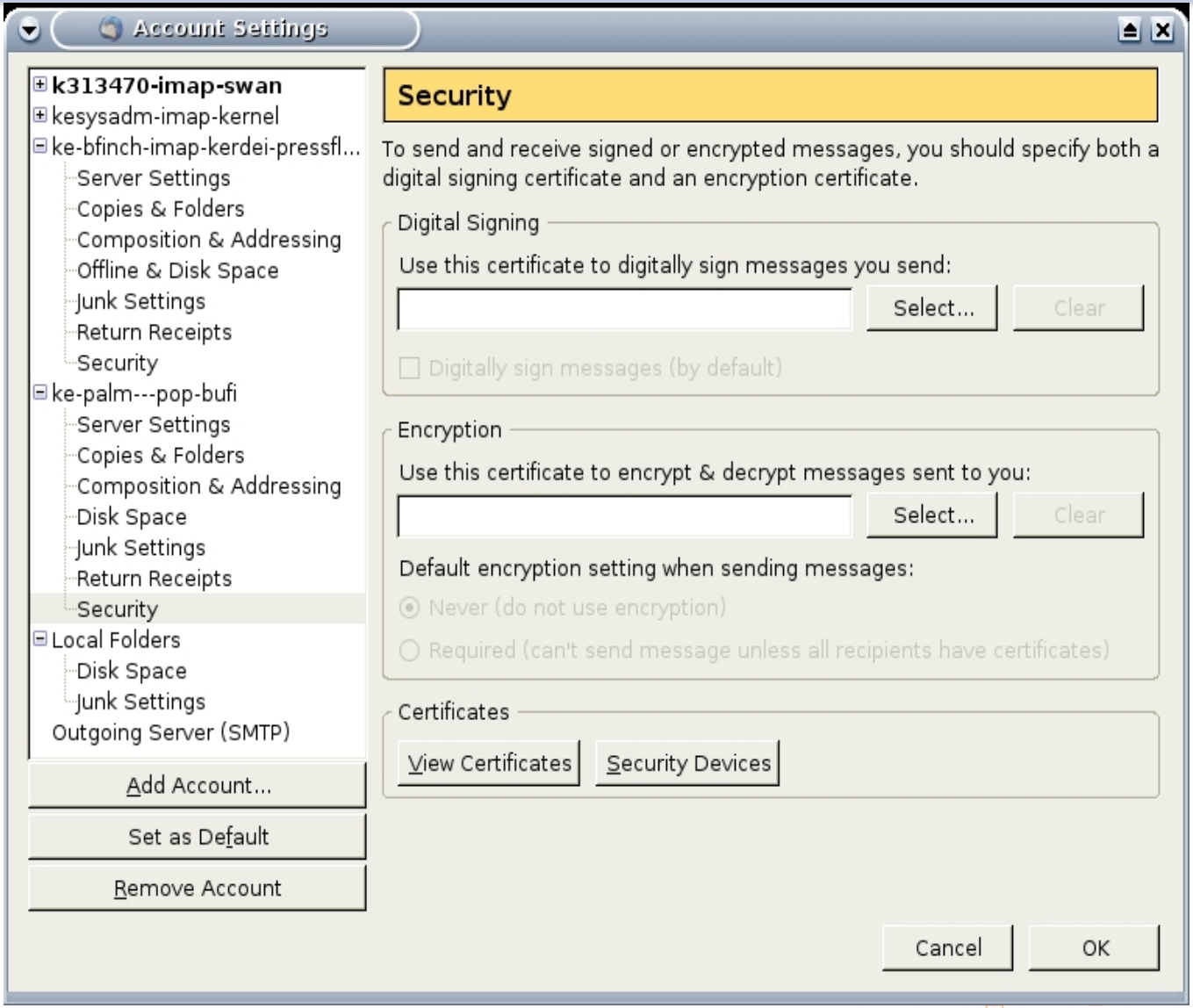
Settings for an email account



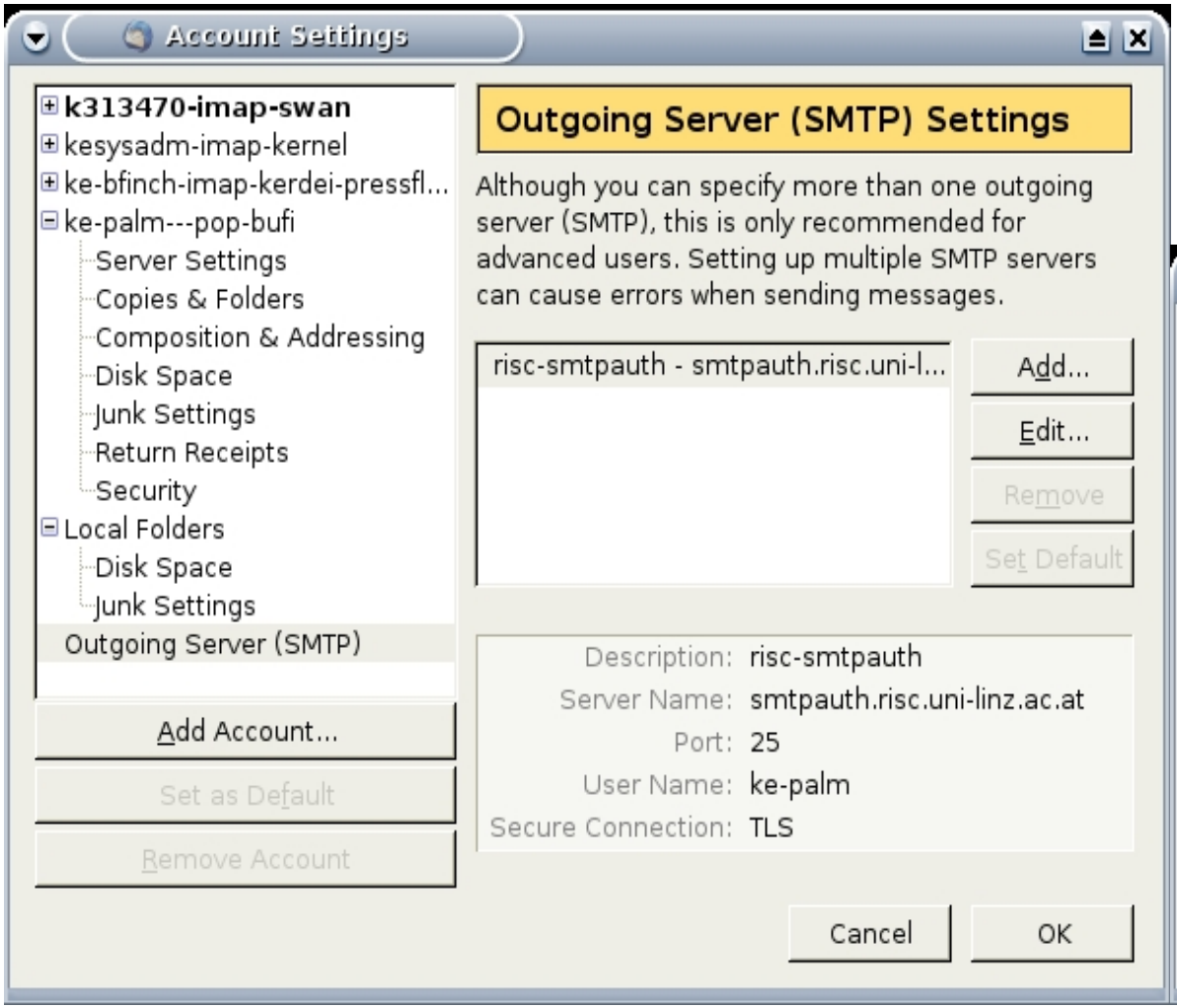
Setting the incoming mail server

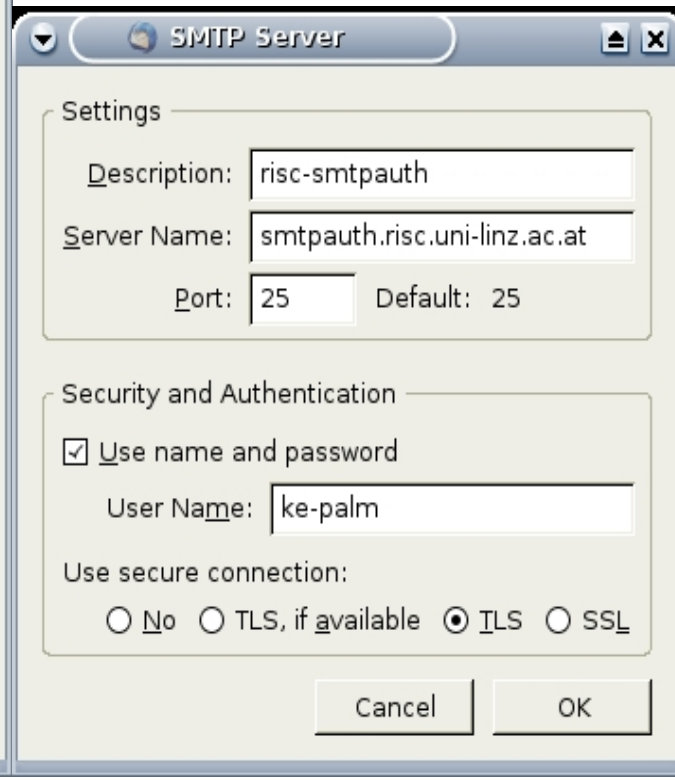


Security setting: sign and encrypt an e-mail

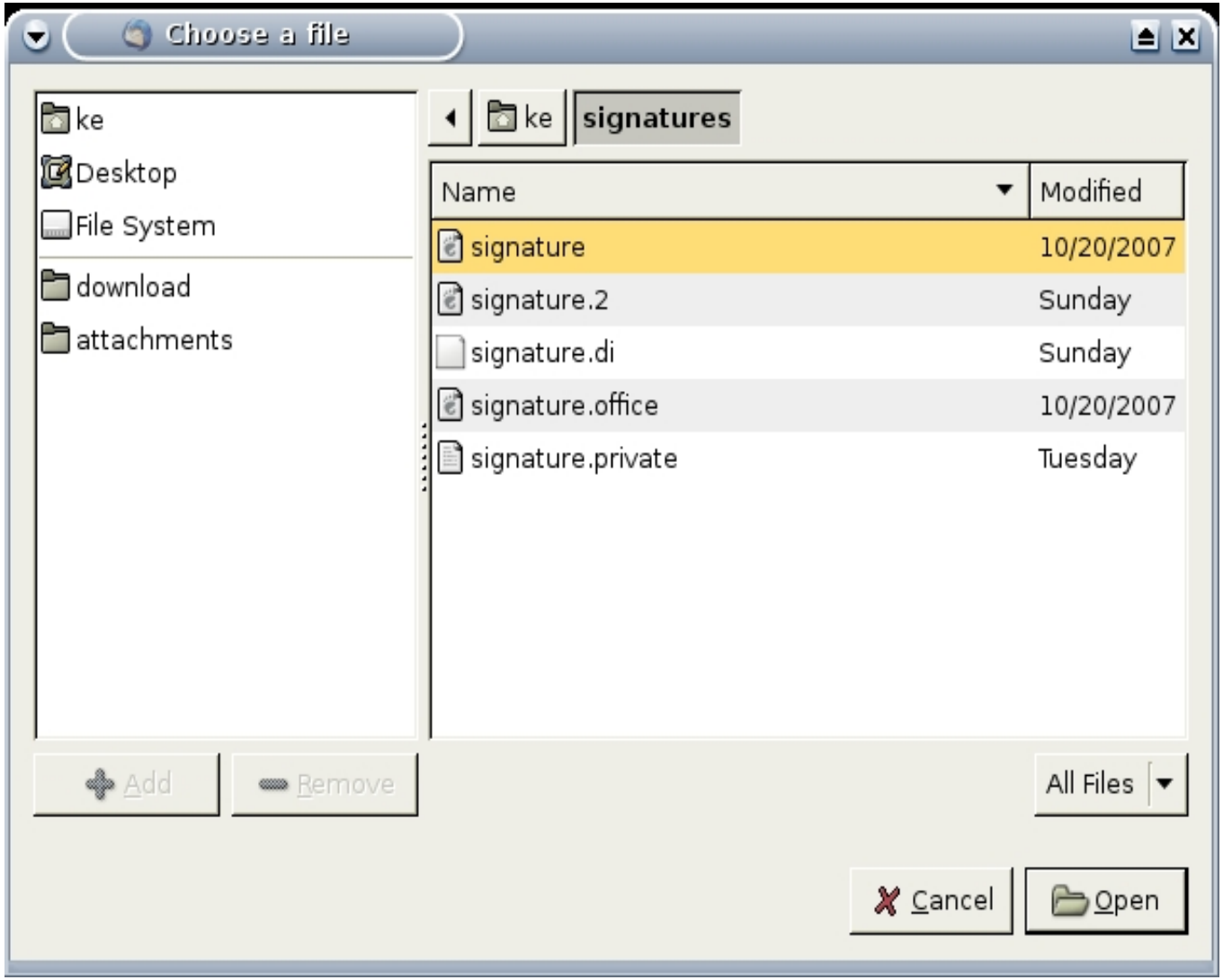


Setting the stmp out host

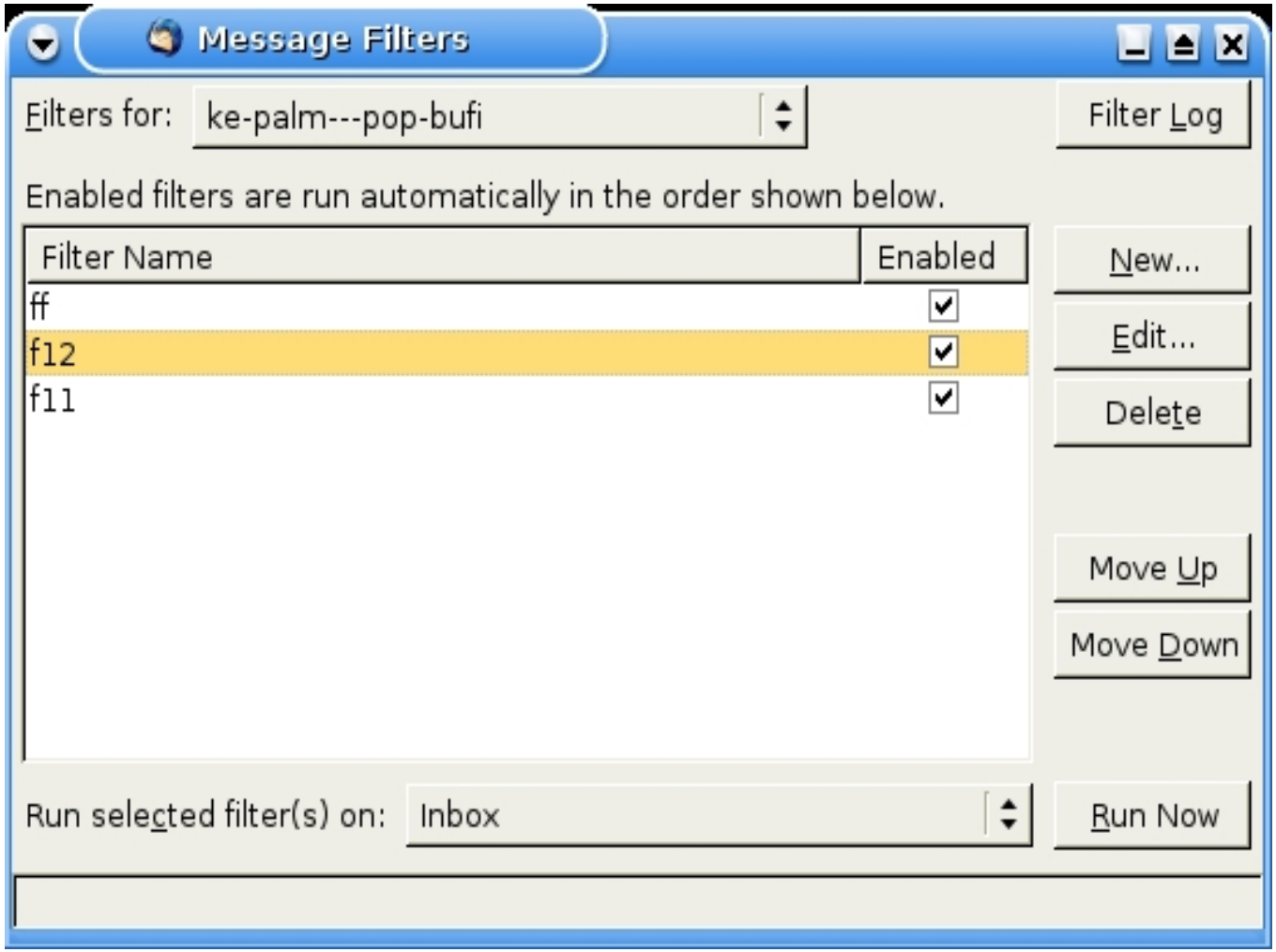




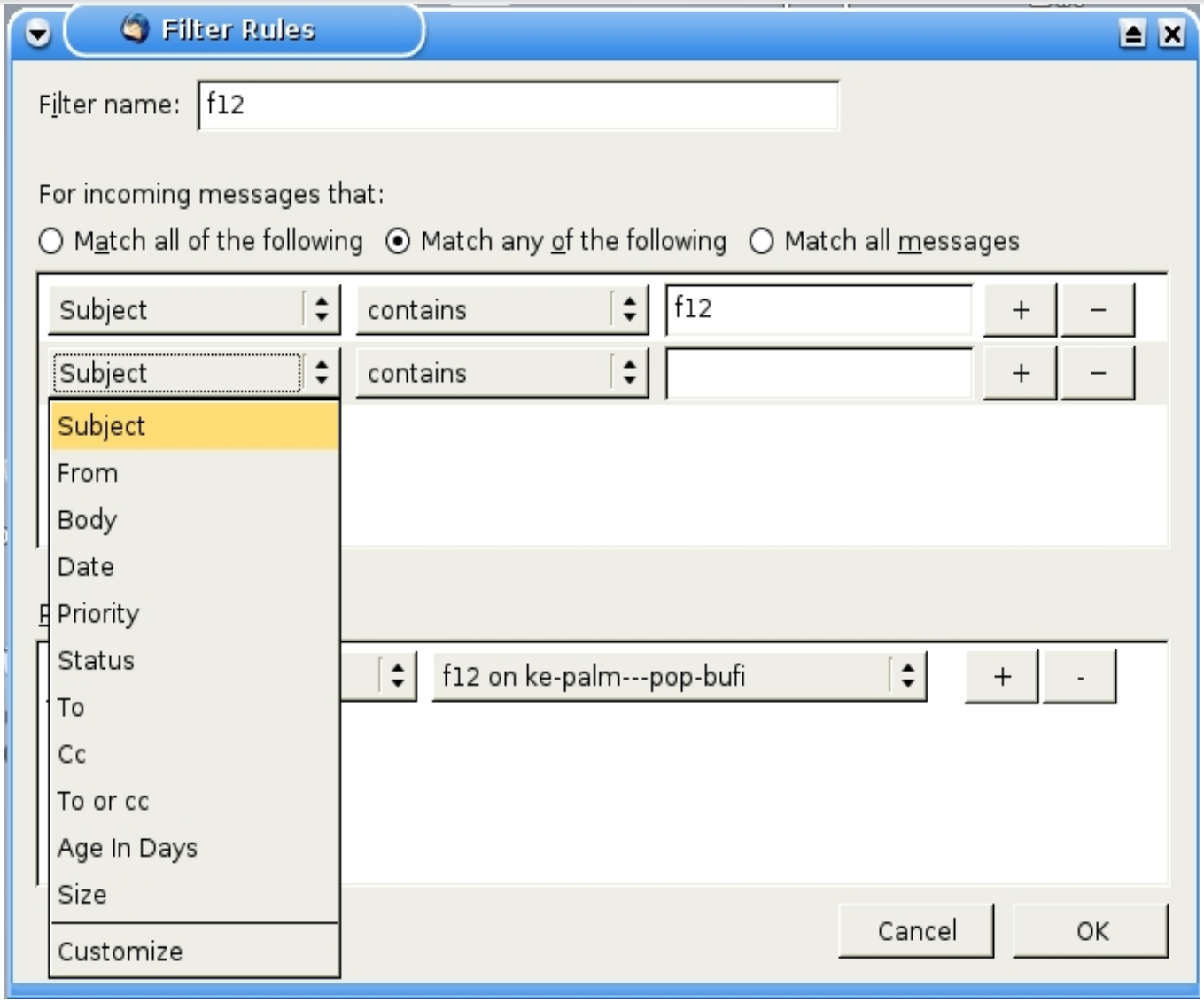
Using more signatures



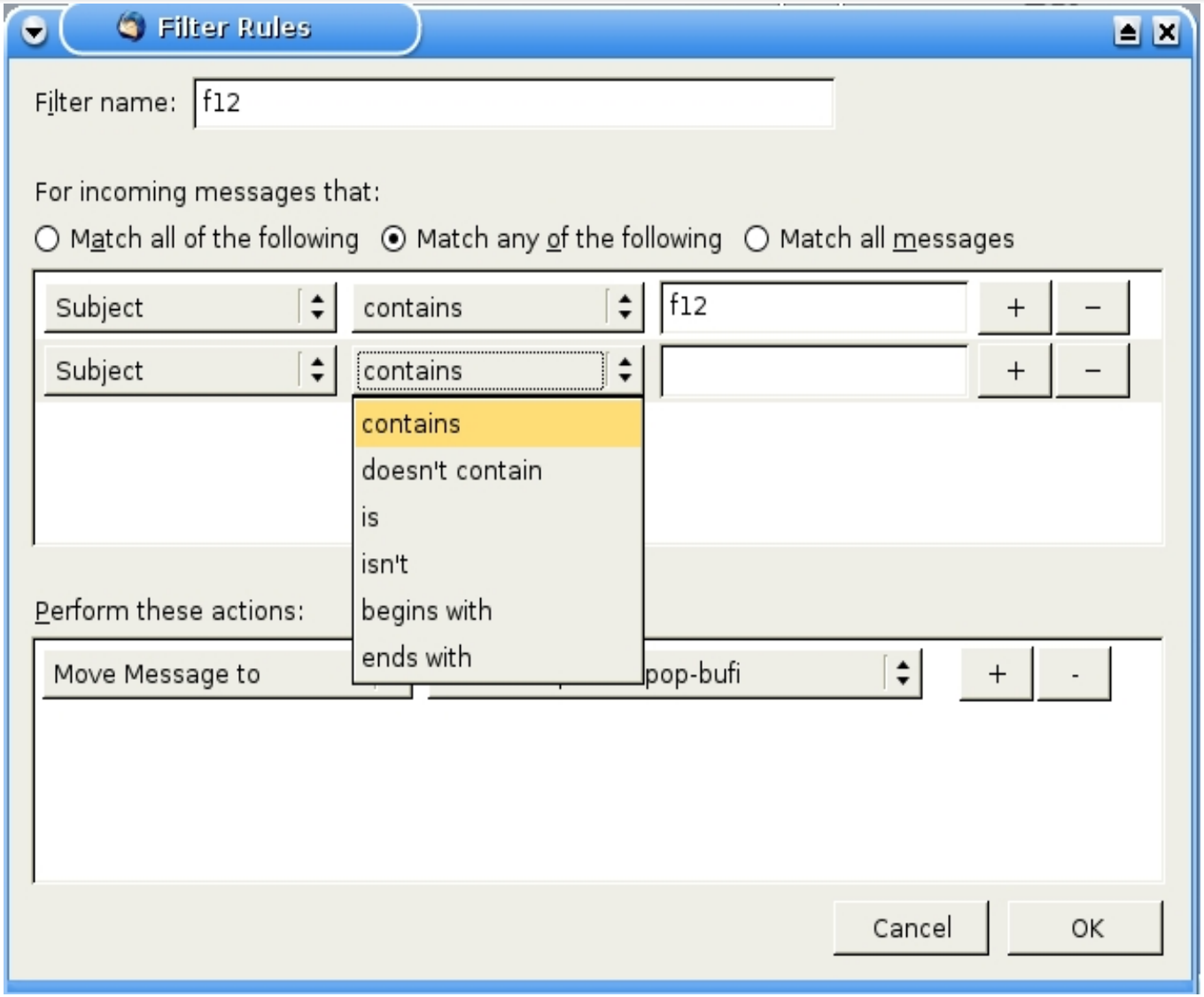
Using message filters



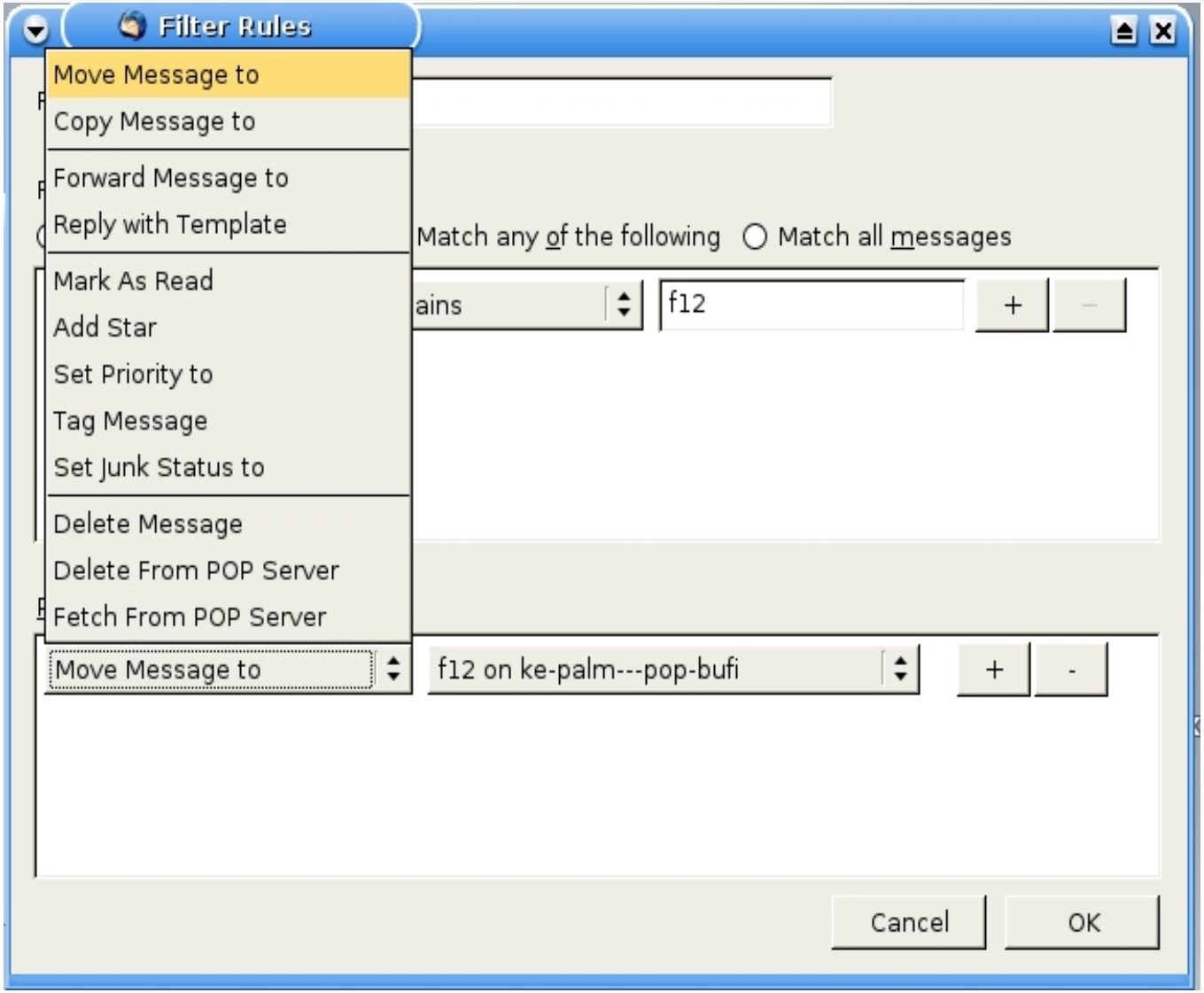
Using message filters - Which field to filter



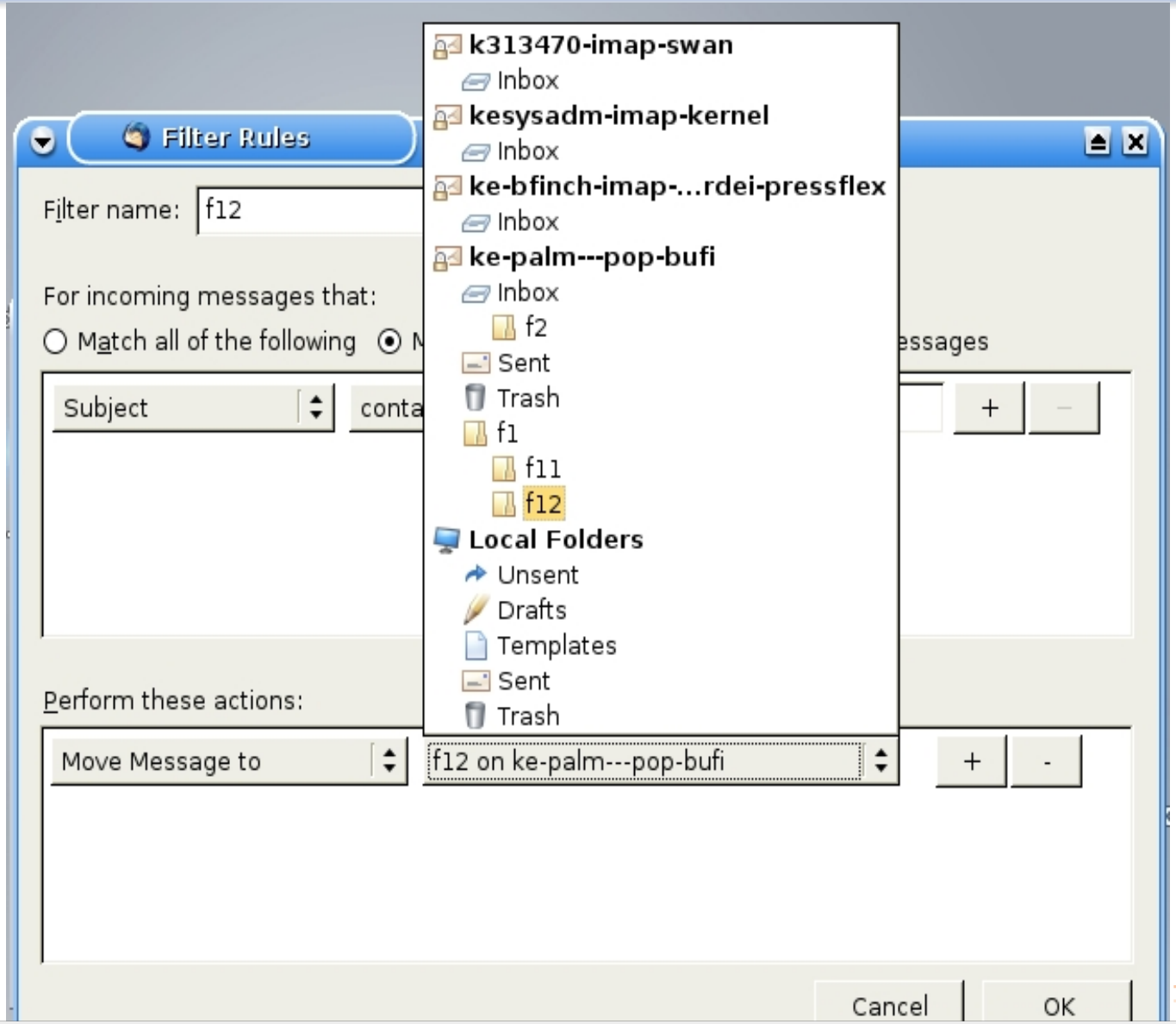
Using message filters - Set relation



Using message filters - Set action



Using message filters - Set destination



Using Folders and Virtual Folders

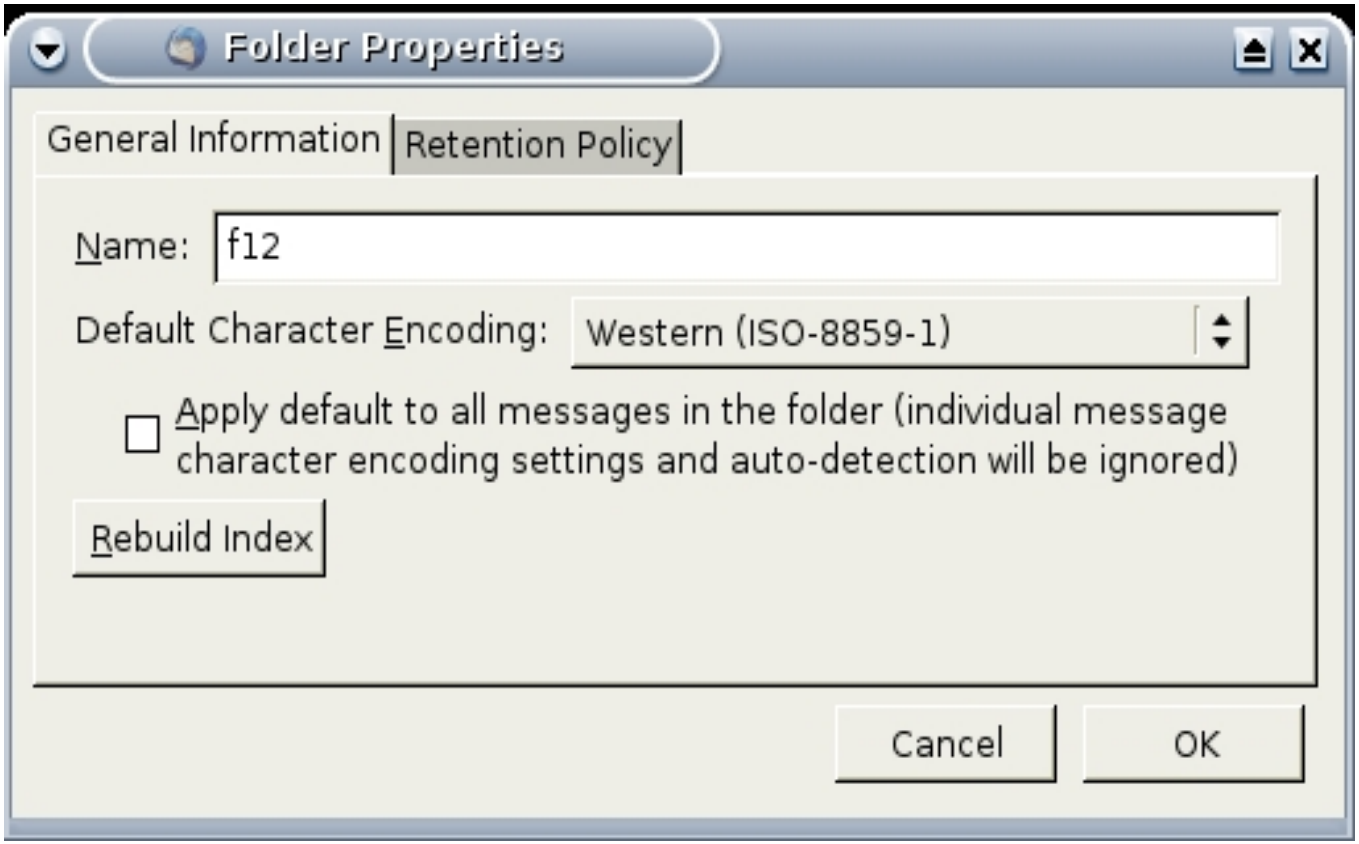
Using Folders

- Thunderbird uses mbox format to save the e-mails
- folder consist any number subfolders but only one mbox folder
- folders has tree structure
- you can create folders on static critearia
  - folder: CBWE; subfolders: questionaire, lecturer, etc.

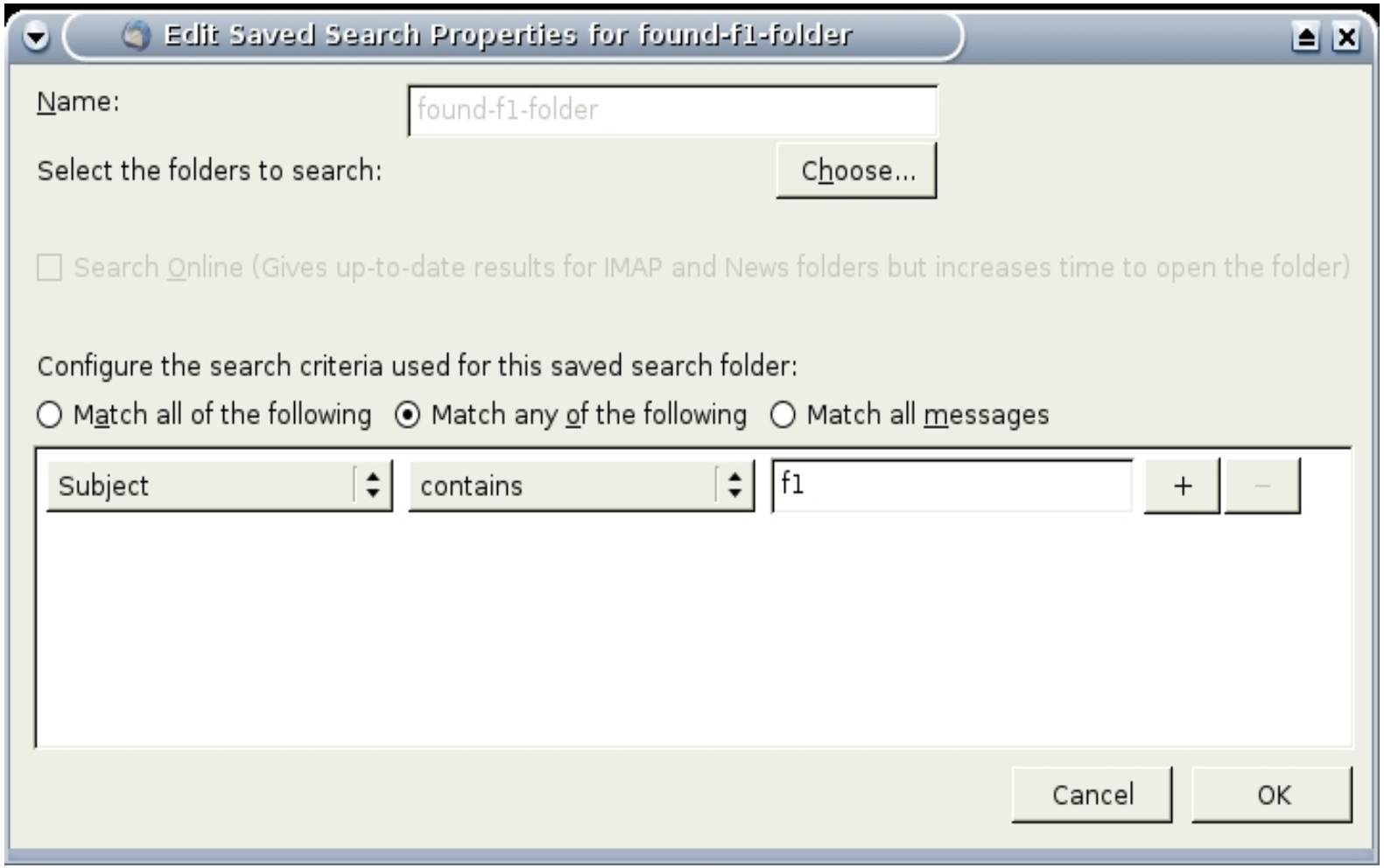
Using Virtual Folders (VF)

- creating virtual folders:
  - specify a set of search criteria on messages, accounts
  - save the search as a virtual folder
- you work with the virtual folders as a conventional folder
- you can dinamically rerun the search each time
- you can always modify the search criteria
- VF is not a real folder, no messages are moved into it

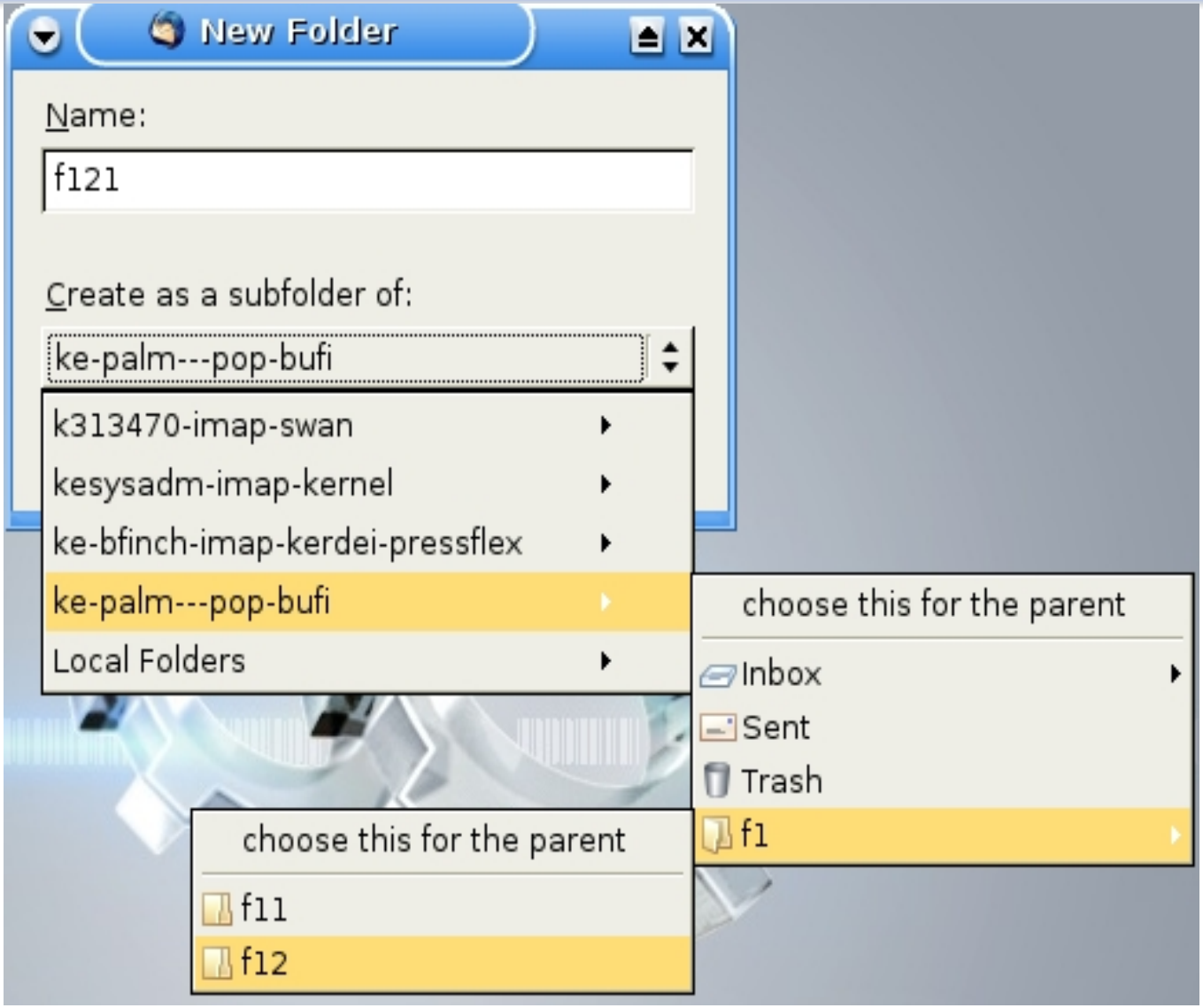
Folder properties



Virtual folder properties

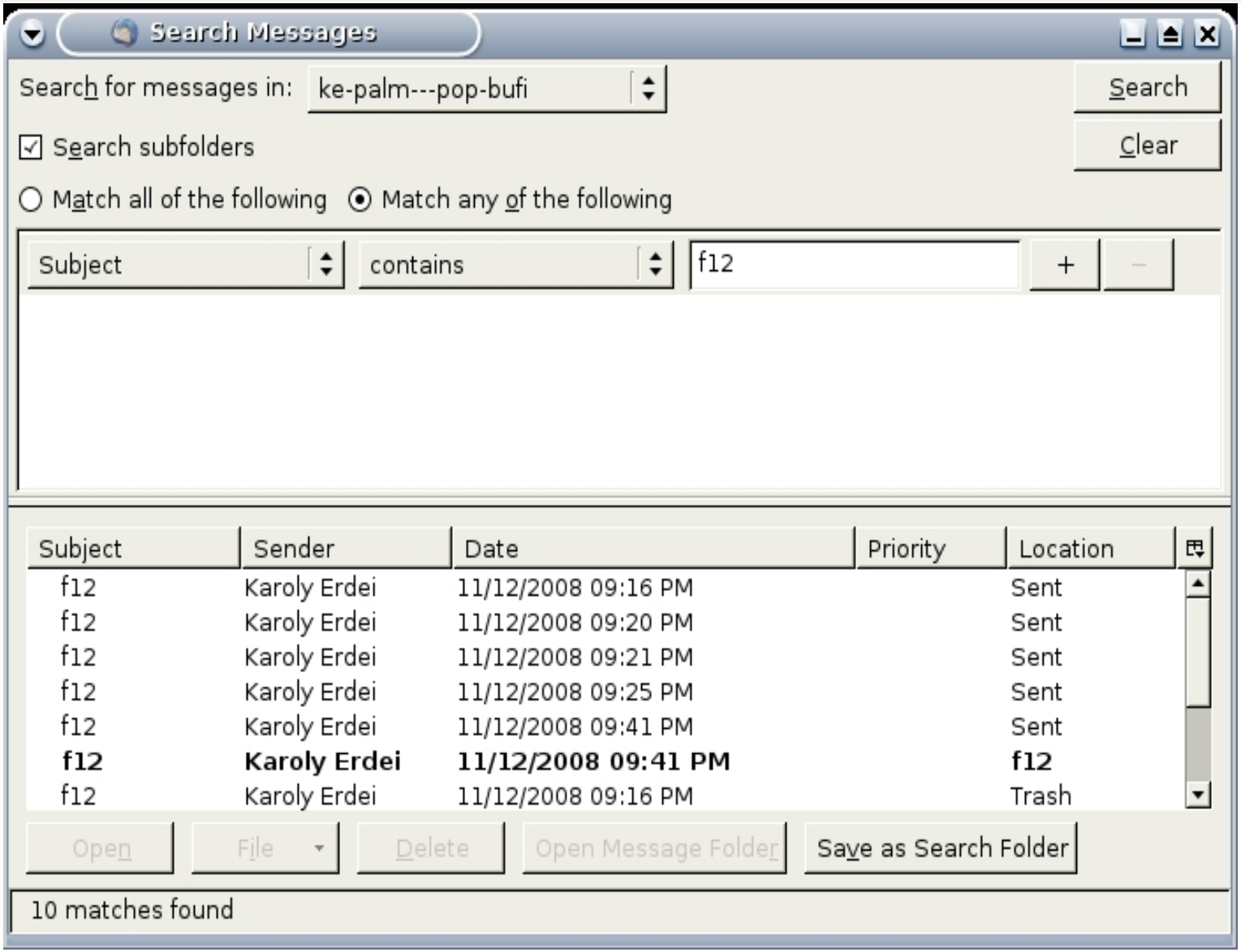


Creating new folder



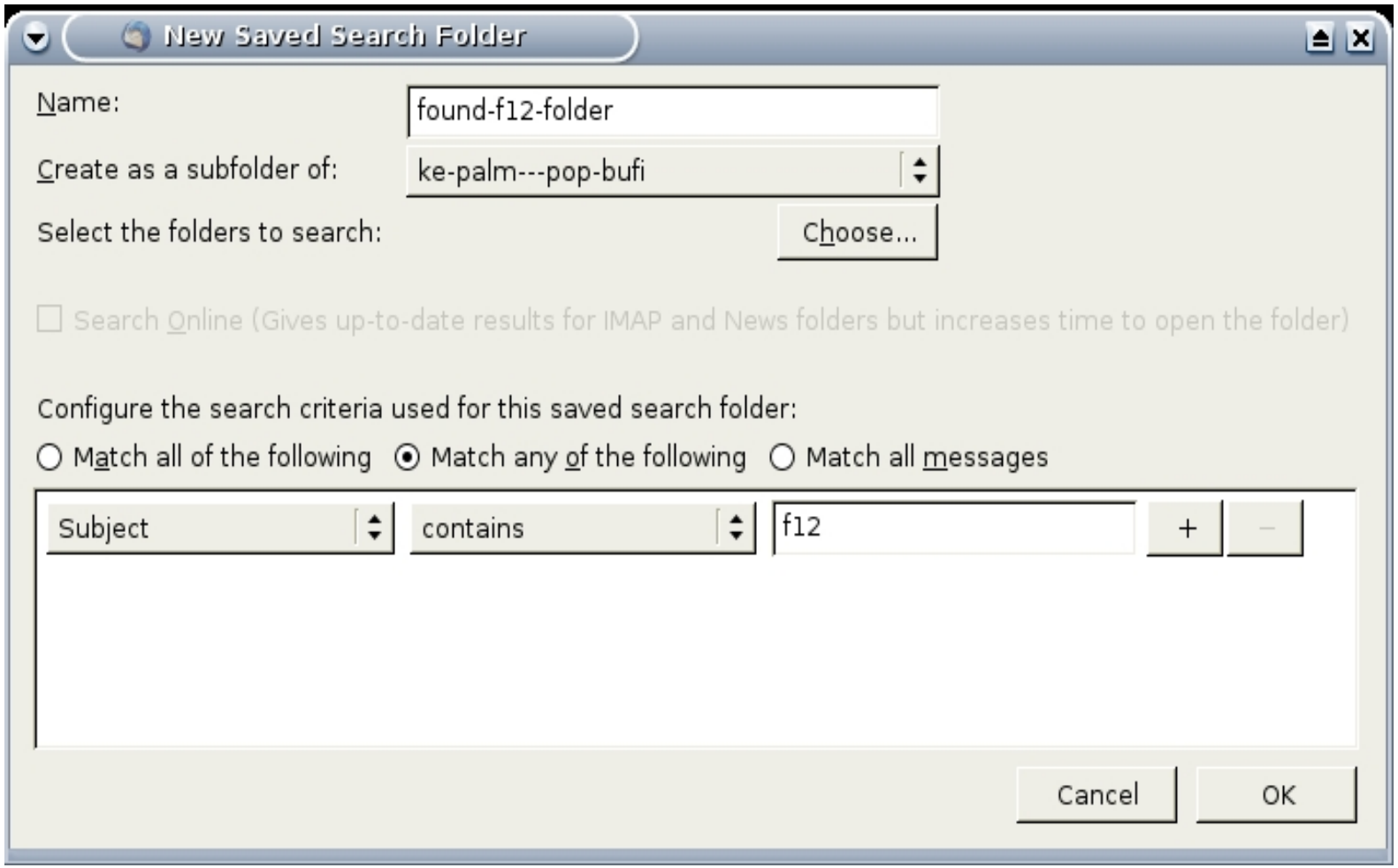
Searching in folders

Preparing creation of Virtual Folder



New saved search folder

Create new Virtual Folder



Agenda

Internet Domain Names - Hierarchical names

IP Adresses and Domain Names

- Internet hosts are identified by IP addresses
- Goal: hosts should be also identified by names
  - name structure must be created, readable for humans
  - high-level names must be translated to low-level IP addresses
- Creating name structure: hierarchical names
  - Name space is distributed into multiple domains by central authority
  - Domain administration is delegated to subauthorities

Name structure: *host.group.site* reflects authority structure

- *site* is the name assigned by the central authority to a domain (.at)
- *group.site* is the name assigned by the authority for domain *site* to a particular subdomain within its domain (jku.at)
- *host.group.site* is the name assigned by the autority for subdomain *group.site* to a particular host within its subdomain

Internet Domain Names - Organisation

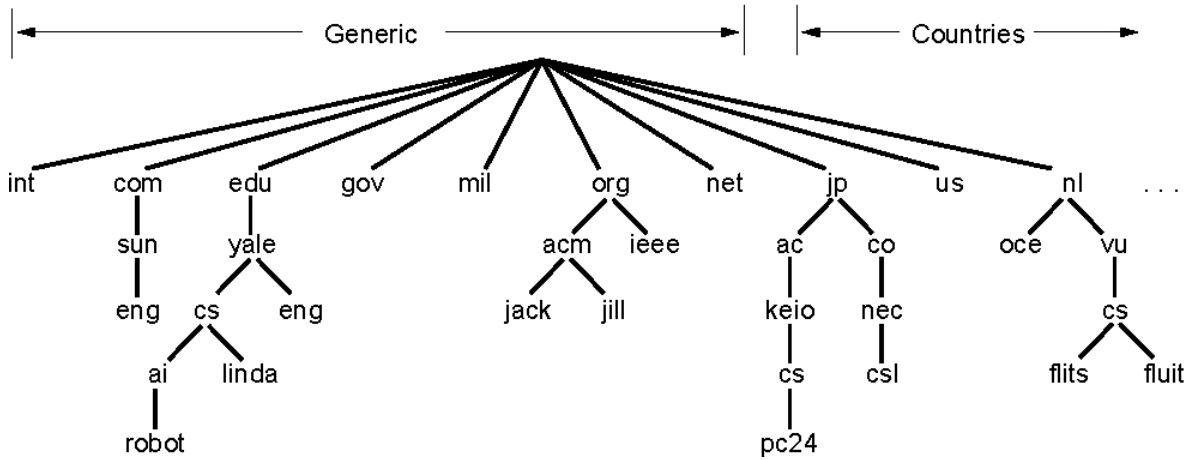
Internet Domain Names - Name hierarchy for TCP/IP Internets

- The Domain Name
  - Sequence of labels separated by periods (risc.uni-linz.ac.at)
  - Each suffix is also a domain (`uni-linz.ac.at`, `ac.at`, `.at`)
  - Top domain is the last label in the name (`.at`)
  - Top domains are determined by Internet authority
- Organisation of Domain Names - summarized:
  - Internet namespace is organized in domains
  - Domains may be further organized in subdomains
  - At each domain level, name assignment for subdomains is delegated from an authority to a subauthority
  - Example: JKU received the domain from the ACOnet and delegates assignment of names in subdomains to its departments (RISC)

Internet Domains

Organizational and geographical domains

- Country domains (Geographic distribution of the Internet)
  - Two letter country code according to ISO 3166 (at for Austria)
- Generic domains (Organizational distribution of the Internet)
  - edu, com, org, net, gov, mil, eu
  - New generic top domains from time to time added



Domain registration

Registration - Structure of authorities

Registration of risc.uni-linz.ac.at

- Austria (NIC-AT) received the domain at from the Internet authority
- The ACOnet has received the subdomain ac.at from NIC-AT
- The JKU (IM - Information Management) has received the subdomain uni-linz.ac.at from the ACOnet
- The RISC has received the subdomain risc.uni-linz.ac.at from the JKU-IM (1991)
- The SAG at RISC registers the hosts in the risc.uni-linz.ac.at domain

Mapping Internet Names to IP Addresses

How does DNS map an Internet name to an Internet address?

Mapping - The Name Server site

- Internet Domain name is mapped to a record of resources (RR)
  - RR may be: A - Internet address; MX - mail exchanger; NS - Name Server; PTR - reverse address (IP-> name)
- The mapping is performed by a name server, which is computer, running Name Server program and managing RR database
- Name servers cooperate with each other to provide the mapping
- Mapping of names to addresses is performed by a distributed database
- Name server knows the name servers for all its subdomains
- Each name server knows at least one root server
- Root server forwards query to name server of top domain

Domain Name Resolution

Application, Resolver, NS

How does a host resolve a name - the client site

- Application invokes the resolver to return a RR (IP, etc.)
  - this is a library call to a part of the OS
- Name resolver software sends query to a name server
  - at least one name server must be registered on the host
  - Linux: file /etc/resolv.conf contains e.g. an entry

    `nameserver 193.170.37.225`
- Name server checks whether the name is in the domain for which it is authority (`jackal.risc.uni-linz.ac.at`)
  - If yes, the server looks for the name in the data base
  - If host name is known, the corresponding mapping is returned
  - If host name is unknown, this is reported as a response
- If not, the server forwards the query to a root server or it returns the address of the authoritive server for the domain
- Client itself contacts next authoritive server to resolve the name
- if last NS in hierarchy does not know the name returns error message

Caching in DNS

Why Caching - how it works

- Name servers use caching of RRs to optimize search
  - Each server maintains a cache of recently used, mapped RRs
  - If the server is authorative for a name, it returns the mapping
  - Otherwise, if there is a mapping in the cache, it returns the cached mapping
  - Otherwise, it forwards the query to a root server or to a subdomain server
- Through caching the NS replies DNS becomes efficient
  - caching happens in every name server in the domain name hierarchy
- An authorative answer has always a time to live (TTL) value
  - time, an answer may remain in the cache of a non-authorative server
- When TTL value is expired, the name server drops the answer from its cache

Checking the Name Server

Command line tools

- DIG (domain information groper) - DNS lookup utility: dig
  - check the man pages or `dig -h` for details
  - useful command, specially the +norec option
  - dig @nameserver domain-name query-type query-option etc
  - my preferred utility to check NSs
- nslookup: query Internet name servers interactively
  - command line mode; interactive mode (w/o parameter)

Graphical tools

- gnome-nettools, get familier with it!

Checking the Name Server

```
uhu:~> dig @kernel.risc.uni-linz.ac.at prometheus.risc.uni-linz.ac.at
  any +norec
; <<>> DiG 9.3.4-P1.1 <<>> @kernel.risc.uni-linz.ac.at prometheus.risc.

; (1 server found)
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 30683
;; flags: qr aa; QUERY: 1, ANSWER: 3, AUTHORITY: 3, ADDITIONAL: 3

;; QUESTION SECTION:
;prometheus.risc.uni-linz.ac.at.        IN      ANY

;; ANSWER SECTION:
prometheus.risc.uni-linz.ac.at. 1800 IN HINFO   "i686" "Debian-2.1"
prometheus.risc.uni-linz.ac.at. 1800 IN MX     10 bullfinch.risc.uni-li
prometheus.risc.uni-linz.ac.at. 1800 IN A       193.170.37.80
```

Checking the Name Server

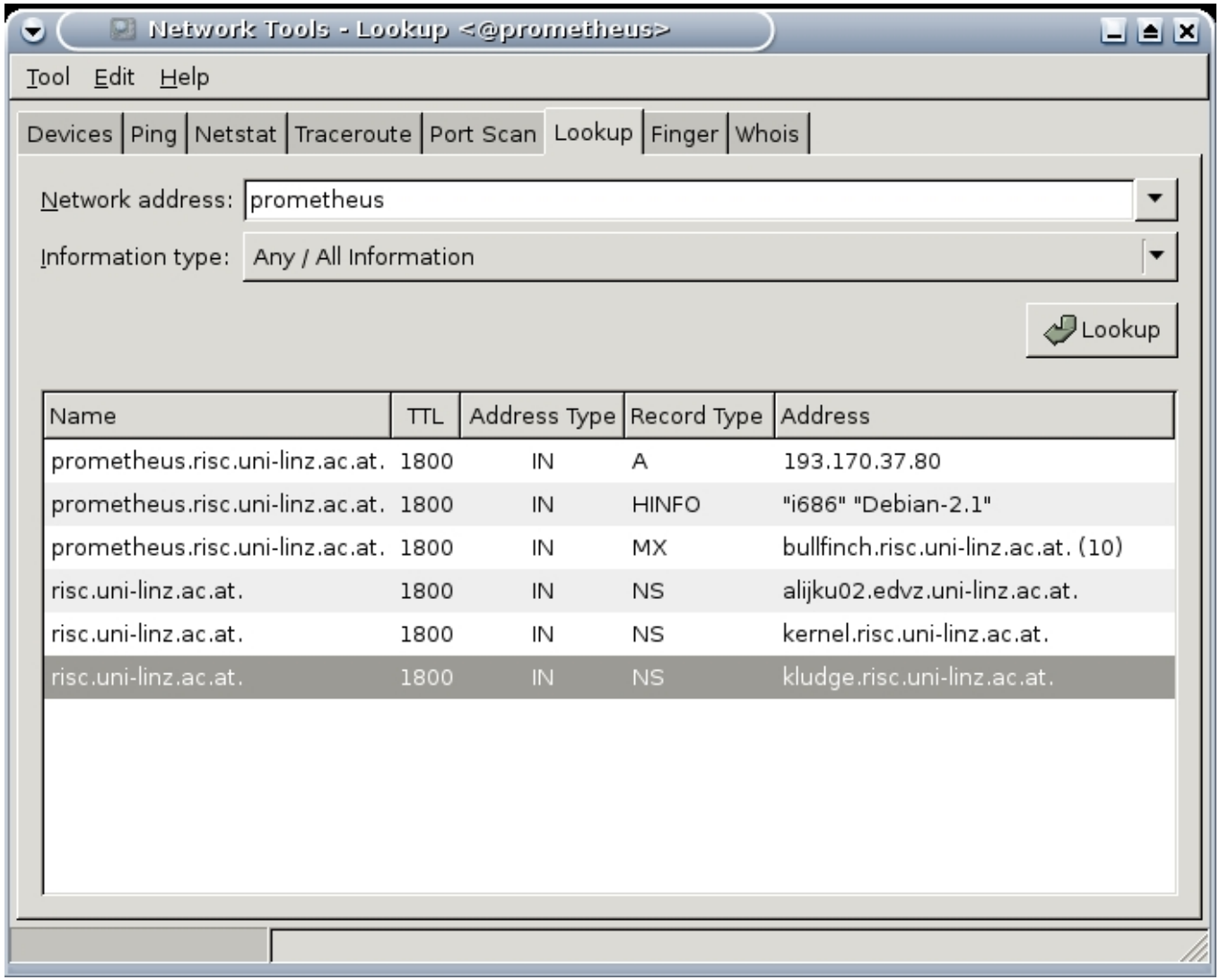
```
;; AUTHORITY SECTION:
risc.uni-linz.ac.at.    1800    IN      NS      kernel.risc.uni-linz.ac
risc.uni-linz.ac.at.    1800    IN      NS      kludge.risc.uni-linz.ac
risc.uni-linz.ac.at.    1800    IN      NS      alijku02.edvz.uni-linz.

;; ADDITIONAL SECTION:
bullfinch.risc.uni-linz.ac.at. 1800 IN  A       193.170.37.222
kernel.risc.uni-linz.ac.at. 1800 IN     A       193.170.37.225
kludge.risc.uni-linz.ac.at. 1800 IN     A       193.170.37.224

;; SERVER: 193.170.37.225#53(193.170.37.225)
```

Checking the Name Server

gnome-nettool - lookup - all information



End of Mailing

Thanks for your attantion !