

Internet Engineering Task Force (IETF)
Request for Comments: 7377
Obsoletes: 6237
Updates: 4466
Category: Standards Track
ISSN: 2070-1721

B. Leiba
Huawei Technologies
A. Melnikov
Isode Limited
October 2014

IMAP4 Multimapbox SEARCH Extension

Abstract

The IMAP4 specification allows the searching of only the selected mailbox. A user often wants to search multiple mailboxes, and a client that wishes to support this must issue a series of SELECT and SEARCH commands, waiting for each to complete before moving on to the next. This extension allows a client to search multiple mailboxes with one command, limiting the delays caused by many round trips and not requiring disruption of the currently selected mailbox. This extension also uses MAILBOX, UIDVALIDITY, and TAG fields in ESEARCH responses, allowing a client to pipeline the searches if it chooses. This document updates RFC 4466 and obsoletes RFC 6237.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at <http://www.rfc-editor.org/info/rfc7377>.

Copyright Notice

Copyright (c) 2014 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
1.1. Conventions Used in This Document	3
2. New ESEARCH Command	3
2.1. The ESEARCH Response	4
2.2. Source Options: Specifying Mailboxes to Search	5
2.3. Strictness in Search Matches	7
2.4. Server-Side Limitations on Search Volume	7
3. Examples	7
4. Formal Syntax	8
5. Security Considerations	9
6. IANA Considerations	10
7. Changes since RFC 6237	10
8. References	10
8.1. Normative References	10
8.2. Informative References	11
Appendix A. Acknowledgments	11
Authors' Addresses	11

1. Introduction

The IMAP4 specification allows the searching of only the selected mailbox. A user often wants to search multiple mailboxes, and a client that wishes to support this must issue a series of SELECT and SEARCH commands, waiting for each to complete before moving on to the next. The commands can't be pipelined, because the server might run them in parallel and the untagged SEARCH responses could not then be distinguished from each other.

This extension allows a client to search multiple mailboxes with one command and includes MAILBOX and TAG fields in the ESEARCH response, yielding the following advantages:

- o A single command limits the number of round trips needed to search a set of mailboxes.
- o A single command eliminates the need to wait for one search to complete before starting the next.
- o A single command allows the server to optimize the search if it can.
- o A command that is not dependent upon the selected mailbox eliminates the need to disrupt the selection state or to open another IMAP connection.
- o The MAILBOX, UIDVALIDITY, and TAG fields in the responses allow a client to distinguish which responses go with which search (and which mailbox). A client can safely pipeline these search commands without danger of confusion. The addition of the MAILBOX and UIDVALIDITY fields updates the search-correlator item defined in [RFC4466].

This extension was previously published in an Experimental RFC. There is now implementation experience, giving confidence in the protocol, so this document puts the extension on the Standards Track, with some minor updates that were informed by the implementation experience. A brief summary of changes is in Section 7.

1.1. Conventions Used in This Document

In examples, "C:" indicates lines sent by a client that is connected to a server, and "S:" indicates lines sent by the server to the client.

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

2. New ESEARCH Command

Arguments: OPTIONAL source options
OPTIONAL result options
OPTIONAL charset specification (see [RFC2978])
searching criteria (one or more)

Responses: REQUIRED untagged response: ESEARCH

Result: OK -- search completed
 NO -- error: cannot search that charset or criteria
 BAD -- command unknown or arguments invalid

This section defines a new ESEARCH command, which works similarly to the UID SEARCH command described in Section 2.6.1 of [RFC4466] (initially described in Section 6.4.4 of [RFC3501] and extended by [RFC4731]).

The ESEARCH command further extends searching by allowing for optional source and result options. This document does not define any new result options (see Section 3.1 of [RFC4731]). A server that supports this extension includes "MULTISEARCH" in its IMAP capability string.

Because there has been confusion about this, it is worth pointing out that with ESEARCH, as with any SEARCH or UID SEARCH command, it **MUST NOT** be considered an error if the search terms include a range of message numbers that extends (or, in fact, starts) beyond the end of the mailbox. For example, a client might want to establish a rolling window through the search results this way:

C: tag1 UID ESEARCH FROM "frobozz" 1:100

... followed later by this:

C: tag1 UID ESEARCH FROM "frobozz" 101:200

... and so on.

This tells the server to match only the first hundred messages in the mailbox the first time, the second hundred the second time, etc. In fact, it might likely allow the server to optimize the search significantly. In the above example, whether the mailbox contains 50, 150, or 250 messages, neither of the search commands shown will result in an error. It is up to the client to know when to stop moving its search window.

2.1. The ESEARCH Response

In response to an ESEARCH command, the server **MUST** return ESEARCH responses [RFC4731] (that is, not SEARCH responses). Because message numbers are not useful for mailboxes that are not selected, the responses **MUST** contain information about UIDs, not message numbers. This is true even if the source options specify that only the selected mailbox be searched.

Presence of a source option in the absence of a result option implies the "ALL" result option (see Section 3.1 of [RFC4731]). Note that this is not the same as the result from the SEARCH command described in the IMAP base protocol [RFC3501].

Source options describe which mailboxes must be searched for messages. An ESEARCH command with source options does not affect which mailbox, if any, is currently selected, regardless of which mailboxes are searched.

For each mailbox satisfying the source options, a single ESEARCH response MUST be returned if any messages in that mailbox match the search criteria. An ESEARCH response MUST NOT be returned for mailboxes that contain no matching messages. This is true even when result options such as MIN, MAX, and COUNT are specified (see Section 3.1 of [RFC4731]), and the values returned (lowest UID matched, highest UID matched, and number of messages matched, respectively) apply to the mailbox reported in that ESEARCH response.

Note that it is possible for an ESEARCH command to return no untagged responses (no ESEARCH responses at all) in the case that there are no matches to the search in any of the mailboxes that satisfy the source options. Clients can detect this situation by finding the tagged OK response without having received any matching untagged ESEARCH responses.

Each ESEARCH response MUST contain the MAILBOX, TAG, and UIDVALIDITY correlators. Correlators allow clients to issue several ESEARCH commands at once (pipelined). If the SEARCHRES [RFC5182] extension is used in an ESEARCH command, that ESEARCH command MUST be executed by the server after all previous SEARCH/ESEARCH commands have completed and before any subsequent SEARCH/ESEARCH commands are executed. The server MAY perform consecutive ESEARCH commands in parallel as long as none of them use the SEARCHRES extension.

2.2. Source Options: Specifying Mailboxes to Search

The source options, if present, MUST contain a mailbox specifier as defined in the IMAP NOTIFY extension [RFC5465], Section 6 (using the "filter-mailboxes" ABNF item), with the following differences:

1. The "selected-delayed" specifier is not valid here.
2. A "subtree-one" specifier is added. The "subtree" specifier results in a search of the specified mailbox and all selectable mailboxes that are subordinate to it, through an indefinitely

deep hierarchy. The "subtree-one" specifier results in a search of the specified mailbox and all selectable child mailboxes, one hierarchy level down.

If "subtree" is specified, the server **MUST** defend against loops in the hierarchy (for example, those caused by recursive file-system links within the message store). The server **SHOULD** do this by keeping track of the mailboxes that have been searched and by terminating the hierarchy traversal when a repeat is found. If it cannot do that, it **MAY** do it by limiting the hierarchy depth.

If the source options are not present, the value "selected" is assumed -- that is, only the currently selected mailbox is searched.

The "personal" source option is a particularly convenient way to search all of the current user's mailboxes. Note that there is no way to use wildcard characters to search all mailboxes; the "mailboxes" source option does not do wildcard expansion.

If the source options include (or default to) "selected", the IMAP session **MUST** be in "selected" state. If the source options specify other mailboxes and **NOT** "selected", then the IMAP session **MUST** be in either "selected" or "authenticated" state. If the session is not in a correct state, the ESEARCH command **MUST** return a "BAD" result.

The client **SHOULD NOT** provide source options that resolve to including the same mailbox more than once. A server can, of course, remove the duplicates before processing, but the server **MAY** return "BAD" to an ESEARCH command with duplicate source mailboxes.

If the server supports the SEARCHRES [RFC5182] extension, then the "SAVE" result option is valid only if "selected" is specified or defaulted to as the sole mailbox to be searched. If any source option other than "selected" is specified, the ESEARCH command **MUST** return a "BAD" result.

If the server supports the CONTEXT=SEARCH and/or CONTEXT=SORT extension [RFC5267], then the following additional rules apply:

- o The CONTEXT return option (Section 4.2 of [RFC5267]) can be used with an ESEARCH command.
- o If the UPDATE return option is used (Section 4.3 of [RFC5267]), it **MUST** apply only to the currently selected mailbox. If UPDATE is used and there is no mailbox currently selected, the ESEARCH command **MUST** return a "BAD" result.

- o The PARTIAL search return option (Section 4.4 of [RFC5267]) can be used and applies to each mailbox searched by the ESEARCH command.

If the server supports the Access Control List (ACL) [RFC4314] extension, then the logged-in user is required to have the "r" right for each mailbox she wants to search. In addition, any mailboxes that are not explicitly named (accessed through "personal" or "subtree", for example) are required to have the "l" right. Mailboxes matching the source options for which the logged-in user lacks sufficient rights MUST be ignored by the ESEARCH command processing. In particular, ESEARCH responses MUST NOT be returned for those mailboxes.

2.3. Strictness in Search Matches

The base IMAP SEARCH command (Section 6.4.4. of [RFC3501]) requires strict substring matching in text searches. Many servers, however, use search engines that match strings in different ways, for example, matching "swim" to both "swam" and "swum" or only doing full word matching (where "swim" will not match "swimming"). This is covered by the "Fuzzy Search" extension to IMAP [RFC6203], and that extension is compatible with this one and can be combined with it.

Whether or not Fuzzy Search is implemented or used, this extension explicitly allows flexible searching with respect to TEXT and BODY searches. Servers MAY use fuzzy text matching in multimapbox searches.

2.4. Server-Side Limitations on Search Volume

To avoid having a search use more than a reasonable share of server resources, servers MAY apply limits that go beyond loop protection, such as limits on the number of mailboxes that may be searched at once and/or limits on the number or total size of messages searched. A server can apply those limits up front, responding with "NO [LIMIT]" if a limit is exceeded (see [RFC5530] for information about response codes). Alternatively, a server can process the search and terminate it when a limit is exceeded, responding with "OK [LIMIT]" and returning partial results. Note that searches that return partial results can cause complexity for client implementations and confusion to users.

3. Examples

In the following example, note that two ESEARCH commands are pipelined and that the server is running them in parallel, interleaving a response to the second search amid the responses to the first (watch the tags).

```

C: tag1 ESEARCH IN (mailboxes "folder1" subtree "folder2") unseen
C: tag2 ESEARCH IN (mailboxes "folder1" subtree-one "folder2")
  subject "chad"
S: * ESEARCH (TAG "tag1" MAILBOX "folder1" UIDVALIDITY 1) UID ALL
  4001,4003,4005,4007,4009
S: * ESEARCH (TAG "tag2" MAILBOX "folder1" UIDVALIDITY 1) UID ALL
  3001:3004,3788
S: * ESEARCH (TAG "tag1" MAILBOX "folder2/banana" UIDVALIDITY 503)
  UID ALL 3002,4004
S: * ESEARCH (TAG "tag1" MAILBOX "folder2/peach" UIDVALIDITY 3) UID
  ALL 921691
S: tag1 OK done
S: * ESEARCH (TAG "tag2" MAILBOX "folder2/salmon" UIDVALIDITY
  1111111) UID ALL 50003,50006,50009,50012
S: tag2 OK done

```

4. Formal Syntax

The following syntax specification uses the Augmented Backus-Naur Form (ABNF) as described in [RFC5234]. Terms not defined here are taken from [RFC3501], [RFC5465], or [RFC4466].

```

command-auth =/  esearch
                ; Update definition from IMAP base [RFC3501].
                ; Add new "esearch" command.

```

```

command-select =/ esearch
                ; Update definition from IMAP base [RFC3501].
                ; Add new "esearch" command.

```

```

filter-mailboxes-other =/ ("subtree-one" SP one-or-more-mailbox)
                          ; Update definition from IMAP Notify [RFC5465].
                          ; Add new "subtree-one" selector.

```

```

filter-mailboxes-selected = "selected"
                          ; Update definition from IMAP Notify [RFC5465].
                          ; We forbid the use of "selected-delayed".

```

```

one-correlator = ("TAG" SP tag-string) / ("MAILBOX" SP astring) /
                  ("UIDVALIDITY" SP nz-number)
                  ; Each correlator MUST appear exactly once.

```

```

scope-option = scope-option-name [SP scope-option-value]
               ; No options defined here. Syntax for future extensions.

```

```

scope-option-name = tagged-ext-label
                  ; No options defined here. Syntax for future extensions.

```



```
scope-option-value = tagged-ext-val
                    ; No options defined here. Syntax for future extensions.

scope-options = scope-option *(SP scope-option)
               ; A given option may only appear once.
               ; No options defined here. Syntax for future extensions.

esearch = "ESEARCH" [SP esearch-source-opts]
          [SP search-return-opts] SP search-program

search-correlator = SP "(" one-correlator *(SP one-correlator) ")"
                  ; Updates definition in IMAP4 ABNF [RFC4466].

esearch-source-opts = "IN" SP "(" source-mbox [SP
                    "(" scope-options ")" ] ")"

source-mbox = filter-mailboxes *(SP filter-mailboxes)
             ; "filter-mailboxes" is defined in IMAP Notify [RFC5465].
             ; See updated definition of filter-mailboxes-other, above.
             ; See updated definition of filter-mailboxes-selected, above.
```

5. Security Considerations

This new IMAP ESEARCH command allows a single command to search many mailboxes at once. On the one hand, a client could do that by sending many IMAP SEARCH commands. On the other hand, this makes it easier for a client to overwork a server by sending a single command that results in an expensive search of tens of thousands of mailboxes. Server implementations need to be aware of that and provide mechanisms that prevent a client from adversely affecting other users. Limitations on the number of mailboxes that may be searched in one command and/or on the server resources that will be devoted to responding to a single client, are reasonable limitations for an implementation to impose (see also Section 2.4).

Implementations MUST, of course, apply access controls appropriately, limiting a user's access to ESEARCH in the same way its access is limited for any other IMAP commands. This extension has no data-access risks beyond what may exist in the unextended IMAP implementation.

Mailboxes matching the source options for which the logged-in user lacks sufficient rights MUST be ignored by the ESEARCH command processing (see the paragraph about this in Section 2.2). In particular, any attempt to distinguish insufficient access from non-existent mailboxes may expose information about the mailbox hierarchy that isn't otherwise available to the client.

If "subtree" is specified, the server MUST defend against loops in the hierarchy (see the paragraph about this in Section 2.2).

6. IANA Considerations

The "Internet Message Access Protocol (IMAP) Capabilities Registry" is currently located at <http://www.iana.org/assignments/imap-capabilities>.

IANA has changed the reference for the IMAP capability "MULTISEARCH" to point to this document.

7. Changes since RFC 6237

- o Change to Standards Track.
- o Added paragraph about duplicate mailboxes.
- o Added Section 2.3 about Fuzzy Search.

8. References

8.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.
- [RFC2978] Freed, N. and J. Postel, "IANA Charset Registration Procedures", BCP 19, RFC 2978, October 2000, <http://www.rfc-editor.org/info/rfc2978>.
- [RFC3501] Crispin, M., "INTERNET MESSAGE ACCESS PROTOCOL - VERSION 4rev1", RFC 3501, March 2003, <http://www.rfc-editor.org/info/rfc3501>.
- [RFC4314] Melnikov, A., "IMAP4 Access Control List (ACL) Extension", RFC 4314, December 2005, <http://www.rfc-editor.org/info/rfc4314>.
- [RFC4466] Melnikov, A. and C. Daboo, "Collected Extensions to IMAP4 ABNF", RFC 4466, April 2006, <http://www.rfc-editor.org/info/rfc4466>.
- [RFC4731] Melnikov, A. and D. Cridland, "IMAP4 Extension to SEARCH Command for Controlling What Kind of Information Is Returned", RFC 4731, November 2006, <http://www.rfc-editor.org/info/rfc4731>.

- [RFC5182] Melnikov, A., "IMAP Extension for Referencing the Last SEARCH Result", RFC 5182, March 2008, <<http://www.rfc-editor.org/info/rfc5182>>.
- [RFC5234] Crocker, D. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", STD 68, RFC 5234, January 2008, <<http://www.rfc-editor.org/info/rfc5234>>.
- [RFC5267] Cridland, D. and C. King, "Contexts for IMAP4", RFC 5267, July 2008, <<http://www.rfc-editor.org/info/rfc5267>>.
- [RFC5465] Gulbrandsen, A., King, C., and A. Melnikov, "The IMAP NOTIFY Extension", RFC 5465, February 2009, <<http://www.rfc-editor.org/info/rfc5465>>.
- [RFC5530] Gulbrandsen, A., "IMAP Response Codes", RFC 5530, May 2009, <<http://www.rfc-editor.org/info/rfc5530>>.

8.2. Informative References

- [RFC6203] Sirainen, T., "IMAP4 Extension for Fuzzy Search", RFC 6203, March 2011, <<http://www.rfc-editor.org/info/rfc6203>>.

Acknowledgments

The authors gratefully acknowledge feedback provided by Timo Sirainen, Peter Coates, Arnt Gulbrandsen, and Chris Newman.

Authors' Addresses

Barry Leiba
Huawei Technologies

Phone: +1 646 827 0648
EMail: barryleiba@computer.org
URI: <http://internetmessagingtechnology.org/>

Alexey Melnikov
Isode Limited
14 Castle Mews
Hampton, Middlesex TW12 2NP
United Kingdom

EMail: Alexey.Melnikov@isode.com
URI: <http://www.melnikov.ca/>