

基于 IMAP 协议的多账号邮件订阅分发系统 V1.0 - 使用说明

版本	V1.0
文档状态	编辑
作者	吴俊伟
负责人	吴俊伟
创建日期	2023 年 01 月 18 日
更新日期	2023 年 02 月 02 日

目录

1	简介	3
1.1	编写目的	3
1.2	使用对象	3
1.3	产品范围	3
2	产品概述	5
2.1	总体框架	5
2.2	系统架构	5
2.3	模块描述	6
3	使用说明	7
3.1	业务流程说明	7
3.1.1	邮件发送流程	7
3.1.2	邮件订阅流程	7
3.1.3	附件上传流程	8
3.2	功能模块说明	8
3.2.1	系统用户模块	8
3.2.2	邮件轮询模块	9
3.2.3	敏感词过滤模块	9
3.2.4	邮件收发模块	9
3.3	对外接口说明	10
3.3.1	gRPC 普通接口	10
3.3.2	gRPC 流式接口	12

1 简介

邮件作为互联网领域最常用的传递信息的方式，被国内外各大电商平台以及企政网站选为与客户沟通渠道的主要方式。在帐号注册、客服工单、订单通知、广告营销等领域都起到了重要的作用，邮件客户端都采用的 IMAP 协议，但不支持邮件订阅，只能主动拉取，不利于处理时效要求高的邮件。本人为了解决这一难题开发了“基于 IMAP 协议的多账号邮件订阅分发系统”，该系统作为客户端和第三方邮件服务端的中间层，有效的为客户端提供邮件订阅推送功能，集发送邮件、回复邮件、订阅指定邮箱、按权重分配推送数据、附件发送记录持久化，以及邮件内容敏感词过滤等功能，极大地提升邮件回复效率，帮助平台缩短邮件响应的时间，提高用户交互体验。

1.1 编写目的

本文档为使用说明文档，为产品的使用与维护提供信息基础。

1.2 使用对象

本文档的使用对象主要为产品测试与使用人员。

1.3 产品范围

本系统主要专注于为电商及需要与用户进行邮件沟通的平台提供邮件监听，发送管理等服务，在该系统中实现发送/回复邮件、注册/登录帐号、管理员帐号管理、上传附件并持久化、订阅邮件、敏感词过滤，基于权重的任务队列，邮件协议解析与重组，主要包括如下模块：

1) 系统用户模块

包括用户注册，用户登录及用户权重及权限管理。

2) 邮件轮询模块

包括报文转换、保存订阅数据、推送订阅数据，定时轮询第三方邮件服务数据等。

3) 敏感词过滤模块

包括发送、订阅邮件的正文的敏感词过滤功能。

4) 邮件收发模块

包括 IMAP、SMTP 协议解析、发送/监听邮件、上传、下载附件等。

模块交互见图 1-1 模块交互图所示。本系统主要接口如下：

1) 用户注册/登录接口

通过该接口注册自己的用户，配置自己的权重以及获取鉴权信息。

2) 上传附件接口

通过该接口上传后续要在邮件中携带的附件，并返回附件 ID，可重复使用该附件 ID。

3) 发送邮件接口

通过该接口可发送或回复指定邮箱的邮件并可选是否携带附件。

4) 邮件订阅接口

通过该接口可订阅指定账号的邮件，该接口为流式长连接接口，当收到新邮件时会自动推送给客户端。

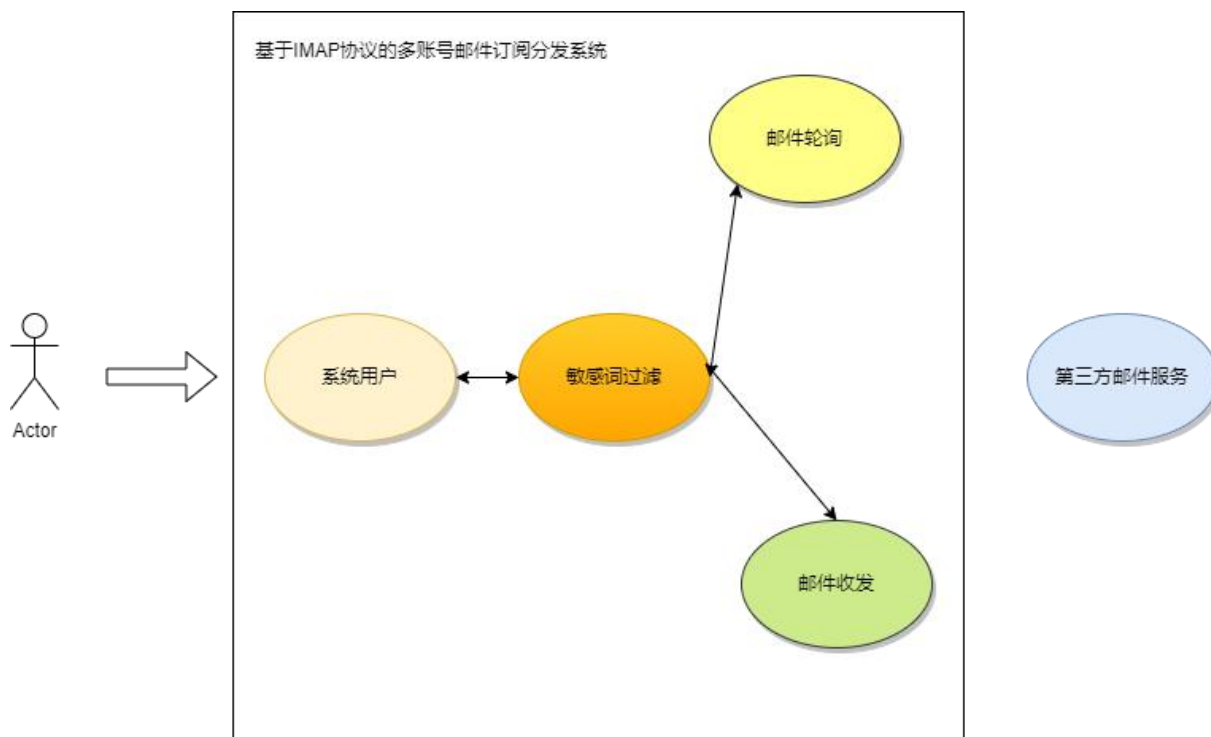


图 1-1 模块交互图

2 产品概述

2.1 总体框架

基于 IMAP 协议的多账号邮件订阅分发系统作为客户端和邮件服务器的中间层，总体框架如图 2-1 总体框架图所示。

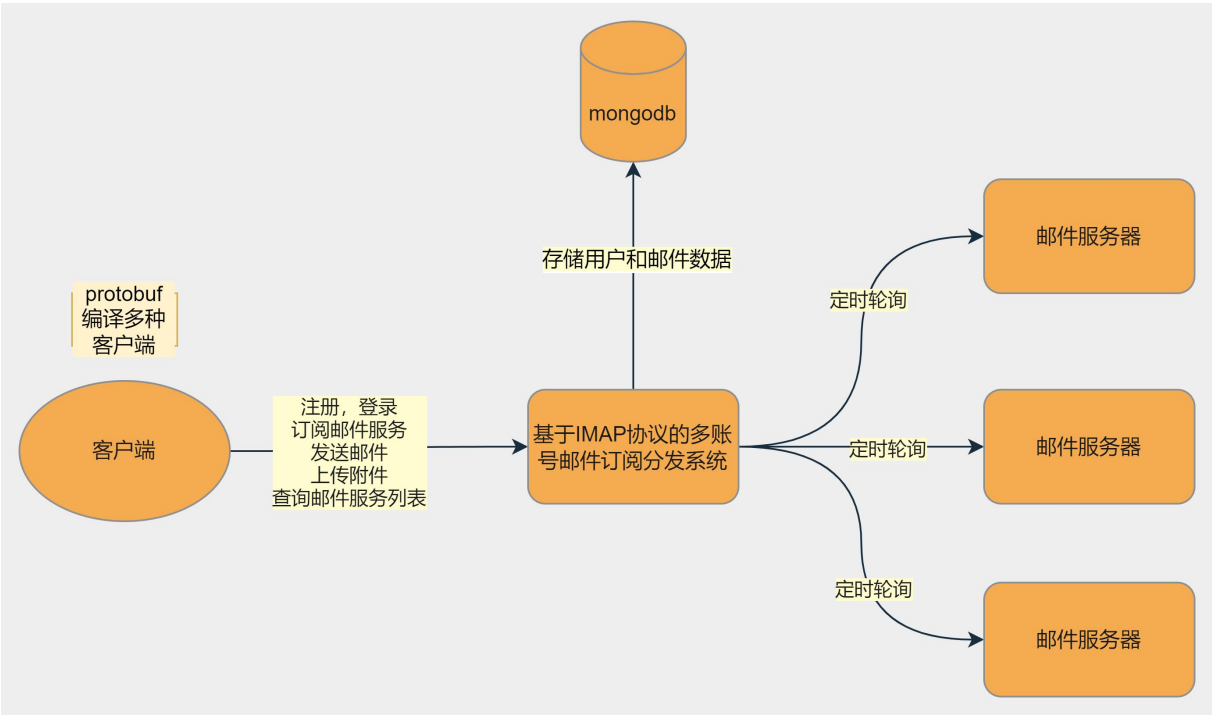


图 2-1 总体框架图

2.2 系统架构

本系统主要使用 Go gRPC 框架，主要开发语言为 Go，数据库采用的是 Mongodb。图 2-2 中，中间件层主要是负责层与层之间的功能衔接，例如：配置动态加载、与数据库的交互（持久化）、用户会话的管理、接口之间的鉴权工作。gRPC 层提供了两类接口，分为流式接口：上传文件，订阅邮件；和非流式接口：登录、注册、发送邮件等。流式接口是为了解决大流量传输设计的，gRPC 底层采用 http2 协议作为应用层协议支持多路复用同时传输多个数据包。协议解析层采用开源协议库。系统架构如图 2-2 所示。

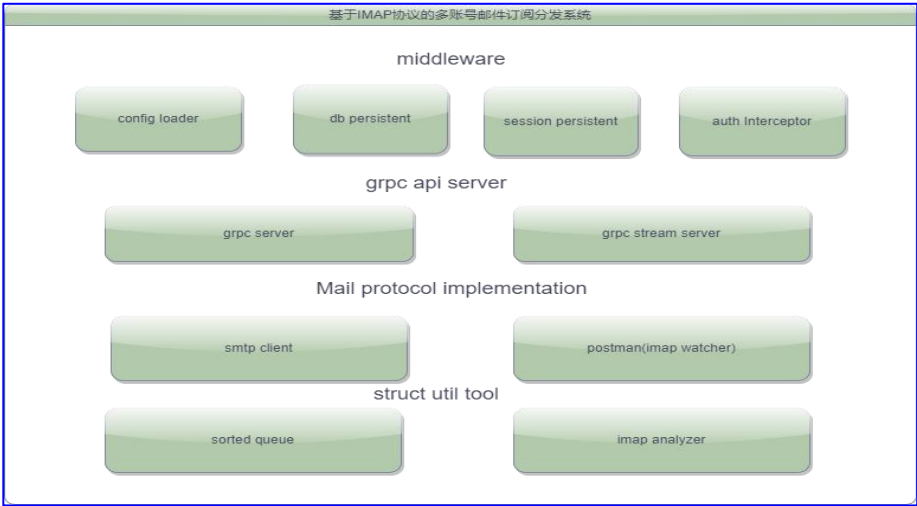


图 2-2 系统架构图

2.3 模块描述

- 系统用户模块

系统用户模块主要功能是向终端提供用户注册、登录服务。接口采用 gRPC 定义。终端可以通过调用这一层次创建用户，获取用户唯一标识（后续接口都需要在 header 中传递这个用户唯一标识），分配用户优先级权重，以及各接口的鉴权功能。

- 邮件轮询模块

邮件轮询模块主要实现多账号邮件定时收取功能是邮件订阅功能的基础和前提，实现了同时订阅多个邮箱帐号，并开启轮询队列，分配 postman（IMAP 客户端）并维持流式长连接接口，实时推送数据等功能。

- 敏感词过滤模块

敏感词过滤模块基于 double-array 实现的 trie（字典树）实现的敏感词过滤功能，根据系统启动时加载的敏感词文件过滤发送、接收的邮件正文内容，支持正则匹配。常用语过滤不文明、政治相关用语或信用卡号等机密信息。

- 邮件收发模块

邮件收发模块主要是提供邮件实时推送、发送/回复邮件，上传/下载附件以及 SMTP/IMAP 协议解析等功能，与以上模块均有交互。

3 使用说明

3.1 业务流程说明

3.1.1 邮件发送流程

详细流程如图 3-1 邮件收发时序图所示。

- 1) 用户使用 gRPC tool 编译生成客户端，可根据用户需要编译不同语言的客户端，发起注册登录请求，获取到用户唯一标识。
- 2) 拿到用户唯一标识后将唯一标识设置在 header 中，请求上传附件 api。再带着附件唯一标识请求发送邮件接口。
- 3) 基于 IMAP 协议的多账号邮件订阅分发系统收到发送请求后，先根据附件 ID 从 mongodb 中拿到附件的文件句柄，随机生成邮件 ID，经过敏感词过滤模块过滤正文及标题，进入权重队列等待发送。
- 4) Postman（发送器）消费权重队列，组装 SMTP 协议数据包，再第三方邮件服务相关接口，发送成功后保存邮件发送记录。

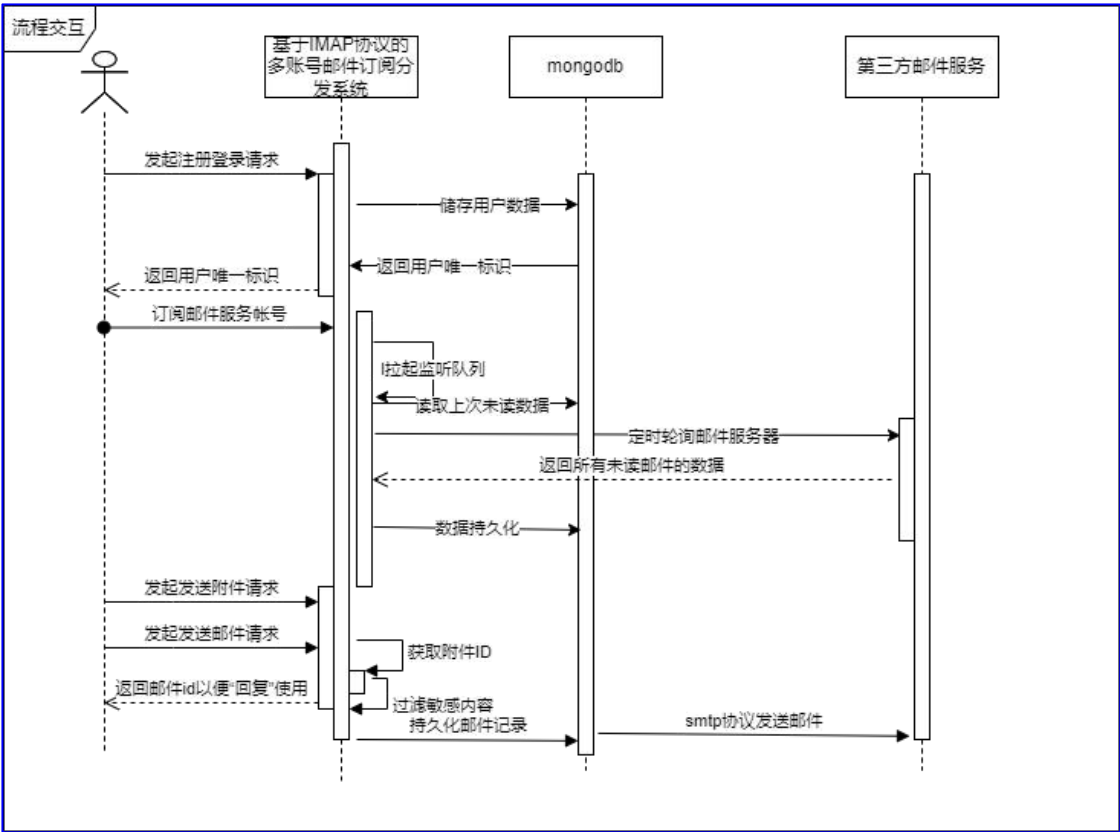


图 3-1 邮件收发时序图

3.1.2 邮件订阅流程

具体的订阅流程如下：

- 1) 用户使用 gRPC tool 编译生成客户端，可根据用户需要编译不同语言的客户端，发起注册登录请求，获取到用户唯一标识。
- 2) 拿到用户唯一标识后将唯一标识设置在 header 中，请求第三方邮件服务器列表接口。
- 3) 从服务器列表接口中获取所有可订阅第三方邮件服务器，选择匹配的服务器，请求邮箱订阅 stream（流式）接口。
- 4) 基于 IMAP 协议的多账号邮件订阅分发系统收到发送请求后，先校验用户唯一标识是否有相应的权限，鉴权通过后拉起邮件服务监听队列。
- 5) 邮件服务监听队列收到新的订阅配置后拉起新的协程异步地定时轮询第三方邮件服务是否存在新的邮件。

3.1.3 附件上传流程

附件上传流程采用 gRPC 的流式接口定义，分块上传数据包完成的，这样在传输大文件时可节省极大的资源。具体流程如下：

- 1) 客户端发送的第一个数据包中应包含该附件的大小，文件名等相关信息。
- 2) 接下来的数据包均为附件的内容，每个包 chunk 为 32kb，传输完毕后，客户端主动关闭连接视为结束。
- 3) 基于 IMAP 协议的多账号邮件订阅分发系统在收到第一个数据包后同时打开 mongodb 的 gridfs 连接，将后续的数据包转发给 mongodb 完成附件的上传。

3.2 功能模块说明

3.2.1 系统用户模块

系统用户模块主要功能是处理用户和用户权限相关的功能接口采用 gRPC 协议实现，支持多语言的客户端。

本模块主要有以下两个功能：

- 创建用户

客户端调用注册登录接口就是由此模块提供，且维护了用户的登录状态，维护了用户的权限相关的信息。

- 鉴权

客户端在调用除登录注册以外的接口都需要在 header 中传递用户唯一标识，此模块会在 gRPC 中设置拦截器，对用户权限进行校验，校验通过才能进行后续操作。

3.2.2 邮件轮询模块

邮件轮询模块主要负责和各类第三方邮件服务器交互，监听订阅队列，并拉起协程定时轮询对应的第三方邮件服务器，并将报文转换为 gRPC 实体，对邮件进行组装分发给各个订阅者，为订阅流式接口提供数据源。

本模块主要有以下两个功能：

- 邮件持久化及推送分发功能

将各帐号订阅的邮件数据持久化在 mongodb 中，并推送至对应的订阅者的管道(channel)中。

- 定时拉取功能

定时将第三方邮件服务器中未读邮件拉取到本地，并保存本次最后一个邮件的 ID，同时将已拉取的邮件设置为已读。

3.2.3 敏感词过滤模块

敏感词过滤模块实现了对正文和标题的敏感词过滤，敏感词在“基于 IMAP 协议的多账号邮件订阅分发系统”启动配置文件中配置，支持两种过滤规则：正则匹配模糊过滤和精准匹配过滤。

本模块主要有以下两个功能：

- 正则匹配模糊过滤

该模块支持基于 Go regexp 包实现，支持 RE2 规则兼容大多数编程语言的匹配规则。

- 精准匹配过滤

该模块使用 double-array trie（双数组实现的字典树）数据结构实现关键词的精确匹配，该匹配算法的平均时间复杂度是 $O(N)$ ， N 为待匹配的数据长度。大大优于传统实现，且占用内存小。

3.2.4 邮件收发模块

邮件收发模块主要是提供本系统的主题功能，使用 gRPC 作为服务协议提供邮件收发功能，主要是通过 Go 代码实现。

本模块主要拥有以下功能：

- 邮件协议解析

对 IMAP/SMTP 协议进行解析，解析为 gRPC 相应接口的 request 和 response，并消费邮件拉取模块的订阅者的管道(channel)通过流式接口推送给客户端。

- 附件上传功能

通过流式接口分块上传附件并保存至 MongoDB，返回附件 ID 一并发送邮件使用。

- 发送/回复邮件功能

使用 SMTP 协议处理用户的发送邮件请求，且支持对某一封邮件进行回复。并附带附件。附件由附件 ID 表示，可复用。

3.3 对外接口说明

客户端是指用户使用 proto 生成的 client 代码，搭配不同语言的 proto 插件可以生成不通语言的客户端代码如 Go、Java、C++、PHP 等，自动生成的客户端代码带有请求需要的请求实体，方便用户使用

3.3.1 gRPC 普通接口

普通接口表现与普通的 http 协议接口类似客户端发起请求，服务端返回数据即结束，普通接口参数如下所示：

1) 注册接口请求参数，如表 3-1 所示。

表 3-1 注册接口请求参数表

序号	字段名	长度	字段类型	字段描述
1	Name	20	String	用户名
2	Password	32	String	密码
3	Weight	10	Integer	用户权重
4	Email	20	String	用户邮箱可用于登录

2) 登录接口请求参数，如表 3-2 所示。

表 3-2 登录接口参数表

序号	字段名	长度	字段类型	字段描述
1	Name	20	String	用户名或邮箱
2	Password	32	String	密码

3) 登录/注册成功后返回参数，如表 3-3 所示。

表 3-3 登录/注册成功返回参数表

序号	字段名	长度	字段类型	字段描述
1	ID	30	String	用户唯一标识

4) 登录/注册失败返回参数，如表 3-4 所示。

表 3-4 登录/注册失败返回参数表

序号	字段名	长度	字段值	字段描述
1	Code	3	Integer	错误码
2	Msg	100	Text	错误信息

3	Info	20	Object	相关错误数据
---	------	----	--------	--------

5) 发送邮件请求参数，如表 3-5 所示。

表 3-5 发送邮件接口参数表

序号	字段名	长度	字段类型	字段描述
1	From	20	String	发件人邮箱
2	To	20	String	收件人邮箱
3	Cc	20	String	抄送人邮箱
4	Bcc	20	String	秘密抄送人邮箱
5	Subject	50	String	邮件标题
6	ReplyID	20	String	回复邮件 ID
7	Text	10000	Text	邮件正文
8	WithAttachment	1	Boolean	是否携带附件
8	Attachment	10	String	附件 ID
9	Name	20	String	发件人名称

6) 发送邮件成功后返回参数，如表 3-6 所示。

表 3-6 发送邮件成功返回参数表

序号	字段名	长度	字段类型	字段描述
1	MsgID	30	String	邮件 ID

7) 发送邮件失败返回参数，如表 3-7 所示。

表 3-7 发送邮件失败返回参数表

序号	字段名	长度	字段值	字段描述
1	Code	3	Integer	错误码
2	Msg	100	Text	错误信息
3	Info	20	Object	相关错误数据

8) 第三方服务器列表接口无请求参数

9) 第三方服务器列表返回参数，如表 3-8 所示。

表 3-8 第三方服务器列表成功返回参数表

序号	字段名	长度	字段类型	字段描述
1	Item		String Array	该用户所有可订阅的服务器名

3.3.2 gRPC 流式接口

流式接口适用于大规模数据传递，和实时场景，服务需要全部数据接收完毕，才能正确回调响应，进行业务处理的场景。可以解决数据表过大可能造成瞬时的压力的问题，基于 IMAP 协议的多账号邮件订阅分发系统中有两个流式接口，参数如下所示：

1) 上传文件接口请求参数，如表 3-9 所示。

表 3-9 上传文件接口请求参数表

请求字段名称	字段中文名称	是否可以为空	字段描述
Name	附件名	否	在邮件中邮件的附件名
ContentType	附件类型	否	符合传输协议拓展名的互联网媒体类型，不传则自动判断
Content	内容	否	附件内容

2) 上传文件接口响应参数，如表 3-10 所示。

表 3-10 上传文件接口响应参数表

相应字段名称	字段中文名称	数据域说明
FileID	附件 ID	用于发送邮件附带附件

3) 订阅邮件服务接口请求参数，如表 3-9 所示。

表 3-9 订阅邮件服务请求参数表

请求字段名称	字段中文名称	是否可以为空	字段描述
Name	附件名	否	在邮件中显示的附件名

4) 邮件订阅响应参数，如表 3-10 所示。

表 3-10 邮件接口响应参数表

相应字段名称	字段中文名称	数据域说明
Mail	邮件	流式发送单个邮件。 具体字段详见表 3-5，字段与发送邮件接口相同