Redmine - Feature #19289

Exclude attachments from incoming emails based on file content or file hash

2015-03-06 09:20 - Mikhail Voronyuk

Status:	New		**Start date:**	
Priority:	Normal		**Due date:**	
Assignee:			**% Done:**	0%
Category:	Email receiving		**Estimated time:**	0.00 hour
Target version:				
Resolution:				

Description

We have a problem similar to [#3413](), i.e. if we create tickets via email there a lot of signature images will get Redmine attachments. **But** the difference is that in our company we use IBM Notes thick and web-clients so

1. the signature images and the inline images (e.g. user screenshots that Printscreen and Ctrl+V into email body) in each new email have different names
2. there is no way to distinguish the inline images from the signature images by name/size (sometimes useful user screenshot may be less that signature image in size) or other tags in email body
3. a similar image that is presented in forwarded email several times has multiple names but similar size and similar binary

Here they are:
!2015-03-06 10-37-50.png!
!2015-03-06 10-41-03.png!

I know that there is a way in Redmine to filter attachments by file mask. As you can see it is useless in this case.
Also we want to leave useful inline images because we can not ask users to do not paste the printscreens into the email body directly and to save at first the screenshot as a file and attach screenshot as file (in that case for the user it is simpler to do not report a bug than to do this:)). So the patch mail_handler_ignore_inline_attachments_patch in [#3413]() useless for us too.

I thought to create a list of ignored attachments (e.g. directory in redmine server with all possible signature images files or file hashes list) and write a patch that will compare binary or hash of each new email attachment with the ignored list. But I'm a newbie in Ruby and I will appreciate any help.

I guess that the patch should be in the app/models/mail_handler.rb in accept_attachment? subroutine and should use FileUtils.cmp for comparing. But I have no idea what contained in attachment.decoded and how do I compare it to master copy.

Related issues:

Related to Redmine - Patch #25215: Re-use existing identical disk files for n...		**Closed**

History

#1 - 2015-03-08 09:09 - Mikhail Voronyuk

- File 0001-Filter-email-attachments-based-on-content-ignore-fil.patch added

- Status changed from New to Resolved

Google and Stackoverflow help to write a patch =)

The additional code compares binary of each new email attachment with the ignored list (directory in a Redmine server with all possible files to be ignored).

May be someone will find the patch helpful.

#2 - 2015-04-03 06:27 - Toshi MARUYAMA

- Status changed from Resolved to New

In your patch, '/home/redmine/redmine-ignored-attachments/*' is hard-coded.
And saving and comparing files anytime is very expensive.
I think it is better to use hash (e.g. md5sum).
Redmine uses md5sum for attachments.
source:tags/3.0.1/app/models/attachment.rb#L108

#3 - 2015-05-13 19:10 - Mikhail Voronyuk

Toshi MARUYAMA wrote:

> In your patch, '/home/redmine/redmine-ignored-attachments/*' is hard-coded.

Where do you propose to save the files to be ignored or theirs hashes? I thought about the Files section but there is no ability to separate useful files from the files to be ignored except using separate project for that or using special filename or description e.g. "ignored".

> And saving and comparing files anytime is very expensive.
> I think it is better to use hash (e.g. md5sum).

I agree that using hash would be better.

#4 - 2015-06-12 15:54 - Toshi MARUYAMA

Path should be configurable such as "attachments_storage_path".
source:tags/3.0.3/config/configuration.yml.example#L69

#5 - 2015-06-19 13:41 - Manuel Mai

This code works perfectly for me in Redmine 3.0 on Windows.
I implemented the MD5-check.

```
require 'digest/md5'
    ignoreddir = "C:\\redmine\\redmine-ignored-attachments\\"
    md5_attachment = Digest::MD5.hexdigest(attachment.body.decoded)
    Dir.foreach(ignoreddir) do |ignoredf|
        next if ignoredf  '.' or ignoredf  '..'
        md5_ignored = Digest::MD5.file(File.join(ignoreddir, ignoredf)).hexdigest
        if md5_ignored == md5_attachment
          logger.info "MailHandler: ignoring attachment #{attachment.filename} (#{md5_attachment}) matching #{ignoredf} (#{md5_attachment})"
          return false
        end
    end
```

To do:
- Configurable path in configuration.yml
- Cache MD5 hashes in database to avoid high load on hard drive because of hashing every file every time an email comes in

#6 - 2015-08-11 23:16 - Jos Groot Lipman

For a very easy optimization you might first compare the size of the files. If the sizes differ there is no need to calculate the MD5 of the ignored file.

#7 - 2015-12-22 19:24 - Sebastian Paluch

same painful problem here :'(

just eliminating duplicated attachments (#15257) would also help.

#8 - 2016-01-12 07:07 - Toshi MARUYAMA

- Related to Feature #15257: Attachment deduplication added

#9 - 2017-02-28 10:44 - Go MAEDA

- Related to deleted (Feature #15257: Attachment deduplication)

#10 - 2017-02-28 10:45 - Go MAEDA

- Related to Patch #25215: Re-use existing identical disk files for new attachments added

#11 - 2024-10-03 14:04 - popy popy

Tested on redmine 5.0.5 and it did not work.
There was two == missing on the next line.
The fixed version is:

```
require 'digest/md5'
    ignoreddir = "C:\\redmine\\redmine-ignored-attachments\\"
    md5_attachment = Digest::MD5.hexdigest(attachment.body.decoded)
    Dir.foreach(ignoreddir) do |ignoredf|
        next if ignoredf == '.' or ignoredf == '..'
```

```
        md5_ignored = Digest::MD5.file(File.join(ignoreddir, ignoredf)).hexdigest
        if md5_ignored == md5_attachment
          logger.info "MailHandler: ignoring attachment #{attachment.filename} (#{md5_attachment}) matching #{
ignoredf} (#{md5_attachment})"
          return false
        end
    end
```

Files

2015-03-06 10-41-03.png	35.6 KB	2015-03-06		Mikhail Voronyuk
2015-03-06 10-37-50.png	36.5 KB	2015-03-06		Mikhail Voronyuk
0001-Filter-email-attachments-based-on-content-ignore-fil.patch	1.3 KB	2015-03-08		Mikhail Voronyuk