TYPO3 Core - Bug #81644

GeneralUtility::getUrl() socket method doesn't support chunked Content-Encoding

2017-06-21 01:13 - Jigal van Hemert

Status:	Closed		**Start date:**	2017-06-21
Priority:	Should have		**Due date:**	
Assignee:			**% Done:**	0%
Category:	Link Handling, Site Handling & Routing		**Estimated time:**	0.00 hour
Target version:				
TYPO3 Version:	7		**Complexity:**	
PHP Version:			**Is Regression:**	
Tags:			**Sprint Focus:**	

Description

With pageNotFound_handling set to the URL of a page the page not found handler uses GeneralUtility::getUrl() to retrieve the contents of the page. Because it requests the headers too the socket method is used if useCurl is not set.
After reading the headers the rest of the stream is simply read in a single operation. If the server has Content-Encoding set to chunked it will send the content in chunks and put the length of each block in hexadecimal before each chunk (plus a zero after the last chunk).

getUrl() fails to process the chunks correctly and the chunk sizes are simply included in the content.

It doesn't seem that guzzle handles chunked encoded data, but in most cases it will use cUrl internally which handles it. I'll have to test if with disabled cUrl v8/master has the same issue.

We can use a simple function to decode the chunked data.

Related issues:

Related to TYPO3 Core - Bug #91582: Fetching an internal page as 404 content ...	**Closed**	**2020-06-04**

Associated revisions

Revision 0d822ba4 - 2020-06-07 12:20 - Markus Klein

[BUGFIX] Ensure PageContentErrorHandler content is served correctly

Use only the body and the content-type header of the internal
sub-request. Everything else of the sub-request can be discarded.

Resolves: #91582
Related: #81644
Releases: master, 10.4
Change-Id: l393acb340c04b3c7aa0cde1e1de1df9fcc92ca7a
Reviewed-on: https://review.typo3.org/c/Packages/TYPO3.CMS/+/64672
Tested-by: TYPO3com <noreply@typo3.com>
Tested-by: Johannes Kasberger <johannes.kasberger@reelworx.at>
Tested-by: Benni Mack <benni@typo3.org>
Reviewed-by: Johannes Kasberger <johannes.kasberger@reelworx.at>
Reviewed-by: Benni Mack <benni@typo3.org>

Revision ee347761 - 2020-06-07 12:42 - Markus Klein

[BUGFIX] Ensure PageContentErrorHandler content is served correctly

Use only the body and the content-type header of the internal
sub-request. Everything else of the sub-request can be discarded.

Resolves: #91582
Related: #81644
Releases: master, 10.4
Change-Id: l393acb340c04b3c7aa0cde1e1de1df9fcc92ca7a
Reviewed-on: https://review.typo3.org/c/Packages/TYPO3.CMS/+/64852
Tested-by: TYPO3com <noreply@typo3.com>
Tested-by: Benni Mack <benni@typo3.org>
Reviewed-by: Benni Mack <benni@typo3.org>

History

#1 - 2019-12-05 22:50 - Riccardo De Contardi

- Category set to Link Handling, Site Handling & Routing

#2 - 2020-05-19 12:33 - Benni Mack

- Status changed from New to Needs Feedback

Hey Jigal,

we avoid using getUrl() in most places now, and use Guzzle. Are you looking for solutions to fetch a page-not-found "chunked"?

#3 - 2020-05-19 14:00 - Jigal van Hemert

Hey Benni,

At the time of submitting the issue I couldn't find that Guzzle handles chunked Content Encoding automatically. cUrl does support chunked data and removes the size numbers before returning the content.

In the situation that the web server is configured to use chunked Content Encoding AND cUrl is disabled AND the page not found handling is set to fetching the contents of a page THEN the output is broken and displays the chunk sizes.
The workaround is to change one of those three conditions. But it would be nice if chunked data was supported.

IIRC Guzzle will automatically detect if cUrl can be used, so the chances of this problem happening are drastically reduced (most web servers will have cUrl).

#4 - 2020-06-03 15:12 - Benni Mack

Jigal van Hemert wrote:

> The workaround is to change one of those three conditions. But it would be nice if chunked data was supported.
>
> IIRC Guzzle will automatically detect if cUrl can be used, so the chances of this problem happening are drastically reduced (most web servers will have cUrl).

Yes. I believe so too, however, I don't have such a set up at hand to build tests around it to make sure we can support this (with guzzle). How do you suggest we proceed?

#5 - 2020-06-03 20:37 - Markus Klein

We face a similar issue with v10 and the PageContentErrorHandler. The fetched page is returned with "Transfer-Encoding: chunked" and this exact response is used to answer the original request. Problem is that this response does not use "Transfer-Encoding: chunked" which yields a failed-connection in the browser or - more weird - several retries by a proxy to "get the other chunks" ultimately DoS-attacking the server with useless requests.

I suggest to remove the header with https://review.typo3.org/c/Packages/TYPO3.CMS/+/64672

#6 - 2020-06-03 20:48 - Markus Klein

- File monitoring.jpg added

#7 - 2020-06-04 14:42 - Markus Klein

- Related to Bug #91582: Fetching an internal page as 404 content breaks browser output and CDNs added

#8 - 2020-08-04 15:05 - Benni Mack

Hey all,

now that the changes are merged, is this issue resolved for everybody?

#9 - 2020-08-04 15:34 - Markus Klein

Yes, for me.

#10 - 2020-08-25 23:42 - Riccardo De Contardi

- Status changed from Needs Feedback to Closed

I close this issue for now in agreement with the reporter;

If you think that this is the wrong decision or experience the issue again, please open a new issue with a reference to this one.

Thank you

Files

monitoring.jpg	103 KB	2020-06-03	Markus Klein