Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 14, 2016

J. Levine
Taughannock Networks
February 11, 2016

Compressed Data Extension for SMTP
draft-levine-smtp-compress-00

Abstract

SMTP messages can be quite large. This extension specifies a method to transfer SMTP messages in a compressed form.

Status of This Memo

This Internet-Draft will expire on August 14, 2016.

Copyright Notice

Table of Contents

1. Open issues

[[Please remove this before publication, although if there's anything left here, it's probably not ready to publish]]

- Is it worth making provision for multiple compression schemes? After 20 years, there still isn't anything much better than DEFLATE and zlib.

- If a server supports both COMPRESS and CHUNKING [RFC3030], can you mix compressed and uncompressed chunks of data? I don't see why not, but ugh.

- Do we need new 5xx codes for bad compressed data, or can we use 554 for bad data and 552 for too big?

2. Introduction

SMTP messages can be quite large, particularly when they include MIME parts representing documents or images. Since CPU performance has historically increased faster than network speed, sending data in compressed form is likely to be faster than in uncompressed form, even allowing for compression and decompression at each end. For binary material sent in base64 form, compression will likely reduce the size of the material to the size of the original material, or perhaps less if the original material was compressible. If an SMTP session transfers several similar messages, the compressed form of the second and subsequent messages will likely be smaller as wll.

This specification uses the zlib [RFC1950] compression scheme, which is widely available and is known to work well on textual material.

3. Definitions

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].

Syntax descriptions use Augmented BNF (ABNF)[RFC5234].

The ABNF "SP" and "CRLF" are used as in [RFC5321].

4. Compressed data service extension

The name of the SMTP service extension is Compressed data. Its EHLO keyword is "COMPRESS".

A new SMTP verb, CDAT, specifies transfer of compressed data. It takes one mandatory argument, the chunk size which is the number of octets of compressed data that follows. The optional reset-marker specifies that the compression engine's context was reset, as described further below. The optional end-marker specifies that this chunk is the last chunk of the message. The compressed data is sent immediately after the CRLF.

ABNF:

```
cdat-cmd   ::= "CDAT" SP chunk-size
         [ SP reset-marker ] [ SP end-marker ] CRLF
chunk-size ::= 1*DIGIT
reset-marker ::= "RESET"
end-marker ::= "LAST"
```

5. SMTP reply codes

The SMTP server replies 250 to a successful CDAT command. It replies 503 to a CDAT command that attempts to send data after a CDAT command with the end-marker. It replies with an appropriate 5xx code if a chunk of data could not be accepted, due to failed decompression or other reasons. It replies with code 503 to any attempts to send more chunks after a rejected chunk.

6. Use of compressed data

Each chunk MUST contain one or more complete byte-aligned blocks of compressed data. A block of compressed data MUST NOT be split between two chunks.

Normally, all of the chunks of compressed data in an SMTP session are treated as a single stream of data through the compression and

decompression engines, with the engines' internal state preserved from one chunk to the next, including chunks in different mail messages. This means that the RESET (RSET) SMTP command MUST NOT reset the compression state. The reset-marker on a chunk means that the engine was reset to its initial state before compressing the chunk, so the decompressor has to restart from the initial state as well.

In most cases the best compression results will be obtained by not using reset-markers, but there may be situations where a sending host is operationally unable to maintain the compression context between messages. The compression state after a chunk of data is rejected by an SMTP-receiver is undefined, so a subsequent message in the same sesion MUST have the reset-marker.

7. IANA Considerations

IANA is requested to add this entry to the "SMTP Service Extensions" registry.

EHLO keyword	Description	Reference
COMPRESS	Compressed data	(this document)

Table 1: SMTP Service Extensions addition

8. Security Considerations

For the most part, the security issues with compressed messages are the same as with uncompressed messages. Compressed messages can be protected with STARTTLS, exactly the same way as uncompressed messages.

An exploit known as CRIME [CRIME] allows recovery of encrypted compressed strings, using many sessions with chosen plaintexts. Since CDAT does not compress the short strings at the beginning of an SMTP session such as AUTH credentials or the envelope addresses, it seems unlikely that CRIME would be an effective attack.

9. References

9.1. Normative References

[RFC1950] Deutsch, P. and J-L. Gailly, "ZLIB Compressed Data Format Specification version 3.3", RFC 1950, DOI 10.17487/RFC1950, May 1996, <http://www.rfc-editor.org/info/rfc1950>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC5234] Crocker, D., Ed. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", STD 68, RFC 5234, DOI 10.17487/RFC5234, January 2008, <http://www.rfc-editor.org/info/rfc5234>.

[RFC5321] Klensin, J., "Simple Mail Transfer Protocol", RFC 5321, DOI 10.17487/RFC5321, October 2008, <http://www.rfc-editor.org/info/rfc5321>.

9.2. Inforormative References

[CRIME] Goodin, D., "Many ways to break SSL with CRIME attacks, experts warn", Sept 2012, <http://arstechnica.com/security/2012/09/many-ways-to-break-ssl-with-crime-attacks-experts-warn/>.

[RFC3030] Vaudreuil, G., "SMTP Service Extensions for Transmission of Large and Binary MIME Messages", RFC 3030, DOI 10.17487/RFC3030, December 2000, <http://www.rfc-editor.org/info/rfc3030>.

Author's Address

John Levine
Taughannock Networks
PO Box 727
Trumansburg, NY 14886

Phone: +1 831 480 2300
Email: standards@taugh.com
URI: http://jl.ly