syslog Working Group
Internet-Draft
Expires: January 30, 2004

R. Gerhards
Adiscon GmbH
August 1, 2003

Syslog-international Protocol
draft-ietf-syslog-international-00.txt

Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of Section 10 of RFC2026.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on January 30, 2004.

Copyright Notice

Abstract

This document describes syslog-international, a mechanism adding support for international character sets to syslog. Syslog-international provides these features in a way that has no requirements and no impact on existing syslog implementations. It is possible to support syslog-international and gain some of its functionality by only changing the behavior of the devices generating syslog messages. Some additional processing of the received syslog messages may realize additional benefits. There is no need to change syslog relays in order to support syslog-international. Existing syslog implementations will benefit from the fact that syslog-international supporting devices emit proper syslog messages in all cases. It is common practice for many non-syslog-international clients to accidently emit 8 bit characters if used in e.g. European

language environments. Syslog-international just adds a protocol layer to the MSG part of the syslog message. As such, it is compatible with all existing and future implementations of syslog.

Table of Contents

1. Introduction

Syslog-international is an enhancement to syslog as described in draft-ietf-syslog-sign-11.txt [6] that adds support for international character sets to syslog.

This is the first draft on syslog-international. Its main purpose is to stimulate discussion on this topic. The content of this ID outlines some rough ideas but needs definitely some more refinement. The author didn't try to do a full specification with this first draft, so some of the information is incomplete.

Syslog-international does not change the syslog packet format but rather just the payload part of the syslog message. This part is referred to as the MSG part. As such, syslog-international is one layer on top of the other syslog specifications.

Being just another layer, syslog-international message content can be embedded into current and future syslog messages. Relays do not need to be aware that a message is syslog-international enabled - they simply pass the packet unaltered on. Syslog collectors do not necessarily be modified. They may need to be modified to encode international characters correctly. Obviously, syslog clients need to be modified in order to emit syslog-international message content.

One goal of syslog-international is to allow international characters inside the syslog message but retain the simplicity and human readability of original syslog. If there is an alternative that can either make the syslog-international easier to implement OR retain human readability, the design decision should favor human readability. Similarily, implementors have several choices on how to encode messages SHOULD always select the choice that provides the best human readability.

The need for syslog-international arises primarily for two reasons: observed behaviour of current syslog clients is that they may emit non US-ASCIIcharacters inside syslog messages. This is for example commonly found in European installations, which extend the US-ASCII pane by an additional pane of characters in the ABNF %d128-255 range. With some operating systems, these characters can even be embedded in computer names, so there is a high probability that they make it into actual syslog messages. Typically, this causes no problems and thus is seldomly noticed. So the first argument is that some implementations are accidently broken by 256 character alphabets and these chracters appear as result of normal operations.

There are more often issues when running syslog in an Asian environment, especially as there are different character encodings

used between different operating systems. Also Asian languages must be encoded in multi byte character sets where a single chracter may be spread over multiple bytes. Truncation of single chracters (or the high-oder bit) do not necessarily pose a big problem to western scripts but can totally destroy an Asian script. In any case, it is observed behaviour of at least some syslog implementation to emit DBCS character encodings. So the second argument pro syslog-international is that it is needed to properly transmit multi byte character sets (for example as used in Asian languages).

2. Required syslog Format

The essential format of syslog messages is defined in 2. of draft-ietf-syslog-sign-11.txt [6]. We do not intend to duplicate the format description here. This prevents inconsistencies and leaves room for other syslog protocol specifications to evolve. The basic fact that we build on is that within the syslog packet format, there is a field containing the actual payload, the message to be transmitted. This is the MSG part of a syslog packet. As this is the payload of the message, we do not expect any new syslog protocol specification to change it.

The important fact about MSG is that it MUST consist of printable US-ASCII characters only.

A specific character set is not required and the absence of this information can cause misinterpretation. For example, in European languages some of the least-frequently used US-ASCII characters (like "~" and "^") are re-assigned to represent frequently-used local characters not included in the basic US-ASCII set. The German Umlaut characters are a good example for this. Other examples can be found in almost all European languages like French or Spanish. So if an administrator receives syslog messages from e.g. spanish, french and german systems on his central syslog collector in the UK, there may be some strange looking chracters in them. Humans are typically clever enough to get the right meaning out of these words, but automatted processes may have some issues. In the real world, these issues are typically only cosmetic, but at least there is some ambiguity that should be solved. As such, we recommend that even "plain" US-ASCII text messages SHOULD use syslog-international if they emit data not exclusively relying on the US-ASCII character tables as defined in ANSI.X3-4.1968 [1].

The MSG part of an syslog-international message has the following ABNF [4] definition:

```
MSG             = HDR-i18n SP MSG-i18n
HDR-i18n        = COOKIE SP ENCODING SP CHARSET SP LANGUAGE SP MORE SEQNO
COOKIE          = "@#" %d73 "18" %d110   ; that is: "@#i18n"
                      ; note the capital "I" and lower case "n"
ENCODING        = "UTF-7"/"quoted-printable"/"base64"/"plain"
CHARSET         = 1.40*(%d33-126) ;IANA registered charset name
LANGUAGE        = RFC1766 Language-Tag

MORE            = "."/"*"
SEQNO           = 0..4294967295
```

```
MSG-i18n        = 1*((%d33-126) / SP)
SP              = %d32
```

Note well: this definition is not yet complete and needs more discussion. It is provided as a starting point for the discussion.

As can be seen, an international content message is embeded into a syslog-sign [6] MSG field. The international content is distinguished from plain syslog-sign content by the presence of a HDR-i18n COOKIE. If the COOKIE is present, the ENCODING part of the HDR-i18n tells which encoding is used, the CHARSET tells the IANA assigned charset it is represented in and the LANGUAGE tag specifies the language. Later revisions of this draft will provide proper links to the relevant RFCs (e.g. RFC 2277 and RFC 1766) and more details.

MORE and SEQNO provide support for syslog messages larger than the allowed syslog packet size. This is introduced to allow transmittal of "oversized" message, which may be the result of some character sets and encodings. These messages will be fragmented by the syslog client and reconstructed by the collector. Relays will pass them through unmodifed.

Message fragmentation MAY be used if the underlying transport provides reliable and in-order delivery (for example RFC 3195 [8]). It the underlying transport is unreliable or its reliability is not known, fragmentation MUST NOT be used.

More specifies wether this is the final fragment of the message or not. An asterisk ("*") means that at least one more fragement will follow. A period (".") means that this is the final (or only) fragment.

SEQNO specifies the sequence number of fragments. It MUST start by 0 for the first fragment and MUST be incremented by 1 for each following fragment. SEQNO MUST restart at 0 for each new full message. A new full message begins after the last message that had "." in MORE.

If fragementation is not used, all messages contain ". 0" as the MORE SEQNO sequence.

The actual content appears after a space.

The following examples are given.

Example 1

```
<34>Oct 11 22:14:15 mymachine su: @#i18n:plain:US-ASCII:en 'su root'
failed for
```

```
      lonvick on /dev/pts/8
```

In this example, as it was originally described in RFC 3164 [7], the message MSG actually is in US-ASCII so it could also be sent in a plain syslog-sign message. To remove uncertainty, it was specifically flagged as being US-ASCII. Please note the encoding of type "plain".

Example 2

```
      <165>Aug 24 05:34:00 10.1.1.1 myproc[10]:
       @#i18n:QUOTED-PRINTABLE:ISO-8859-1:de Gr=FC=DF Gott
```

In this example, we have non US-ASCII characters. The MSG part contains "Gruess Gott" which is the Bavarian way of saying hello. I am using a replacement writing method to make this readable in US-ASCII. The actual string in ABNF is %x47.72.fc.fd.20.47.6f.74.74. The encoding is QUOTED-PRINTABLE in this sample.

3. Security Considerations

The security considerations section requires considerate review once the details of the spec are clear. While doing so, keep the potential of complex encoding and decoding processes in mind. They may provide the breeding bed for all kinds of security weaknesses. It may be a good idea to recommend implementors to test their implementation against MBCS character sets - it is forseeable that some implementors will just take care of western scripts. In this regard, it may also be a good idea to include some sample data in Japanese or some other MBCS. The current security considerations just contain some thoughts that came up while drafting the initial revision.

Syslog-international messages are only as secure as the underlying syslog transport protocol. Be sure to check the security considerations sections of underlying transport RFC or ID.

Invalid character set information may be used to render messages unreadable.

Invalid MBCS encodings may be used to attack decoding processes and freeze them.

Note well to implementors: syslog-international adds some size to the message, effectively shrinking the maximum usable message size. If an implementor simply implements syslog-international and does not check this implication, important message parts may be truncated due to the maximum specified syslog message size in the syslog transport RFCs/ IDs.

4. Authors and Working Group Chair

The working group can be contacted via the mailing list:

syslog-sec@employees.org

The mailing list archive is available at

http://www.mail-archive.com/syslog-sec@employees.org/.

The current Chair of the Working Group may be contacted at:

Chris Lonvick
Cisco Systems
Email: clonvick@cisco.com

The author of this draft is:

Rainer Gerhards
Email: rgerhards@adiscon.com

Phone: +49-9349-92880
Fax: +49-9349-928820

Adiscon GmbH
Mozartstrasse 21
97950 Grossrinderfeld
Germany

5. Acknowledgements

The authors wish to thank Chris Lonvick, Andrew Ross, Albert Mietus, Eric Fitzgerald, Glen Zorn who commented on various versions of this proposal.

References

[1] American National Standards Institute, "USA Code for Information Interchange", ANSI X3.4, 1968.

[2] Freed, N. and N. Borenstein, "Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies", RFC 2045, November 1996.

[3] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[4] Crocker, D. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", RFC 2234, November 1997.

[5] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 2434, October 1998.

[6] Kelsey, J. and J. Callas, "Syslog-Sign Protocol", May 2003.

[7] Lonvick, C., "The BSD Syslog Protocol", RFC 3164, August 2001.

[8] New, D. and M. Rose, "Reliable Delivery for syslog", RFC 3195, November 2001.

Author's Address

Rainer Gerhards
Adiscon GmbH

EMail: rgerhards@adiscon.com

Intellectual Property Statement

The IETF takes no position regarding the validity or scope of any intellectual property or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; neither does it represent that it has made any effort to identify any such rights. Information on the IETF's procedures with respect to rights in standards-track and standards-related documentation can be found in BCP-11. Copies of claims of rights made available for publication and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementors or users of this specification can be obtained from the IETF Secretariat.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights which may cover technology that may be required to practice this standard. Please address the information to the IETF Executive Director.

Full Copyright Statement

HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.