Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: December 28, 2018

V. Birk
H. Marques
pEp Foundation
B. Hoeneisen
Ucom.ch

June 26, 2018

IANA Registration of Trustword Lists: Guide, Template and IANA Considerations
draft-birk-pep-trustwords-02

Abstract

This document specifies the IANA Registration Guidelines for Trustwords, describes corresponding registration procedures, and provides a guideline for creating Trustword list specifications.

Trustwords are common words in a natural language (e.g., English) to which the hexadecimal strings are mapped to. This makes verification processes (e.g., comparison of fingerprints), more practical and less prone to misunderstandings.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

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."

This Internet-Draft will expire on December 28, 2018.

Table of Contents

1. Introduction

In public-key cryptography comparing the public keys' fingerprints of the communication partners involved is vital to ensure that there is no man-in-the-middle (MITM) attack on the communication channel. Fingerprints normally consist of a chain of hexadecimal chars. However, comparing hexadecimal strings is often impractical for regular human users and prone to misunderstandings.

To mitigate these challenges, several systems offer the comparison of Trustwords as an alternative to hexadecimal strings. Trustwords are common words in a natural language (e.g., English) to which the hexadecimal strings are mapped to. This makes the verification process more natural for human users.

For example, in pEp's proposition of Privacy by Default [I-D.birk-pep] Trustwords are used to achieve easy contact verification for end-to-end encryption. Trustword comparison is offered after the peers have exchanged public keys opportunistically. Examples for Trustword lists used by current pEp implementations can be found in CSV format, here: https://pep.foundation/dev/repos/pEpEngine/file/tip/db.

In addition to contact verification, Trustwords are also used for other purposes, such as Human-Readable 128-bit Keys [RFC1751], One Time Passwords (OTP) [RFC1760] [RFC2289], SSH host-key verification, VPN Server certificate verification, and to import or synchronize secret key across different devices of the same user [E-D.birk-pep-keysync]. Further ideas include to use Trustwords for contact verification in Extensible Messaging and Presence Protocol (XMPP) [RFC6120], for X.509 [RFC3647] certificate verification in browsers or in block chain applications for crypto currencies.

2. Terms

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].

- Handshake: The process when Alice - e.g. in-person or via phone - contacts Bob to verify Trustwords (or by fallback: fingerprints) is called handshake. [E-D.birk-pep-handshake]

- Man-in-the-middle attack (MITM): cf. [RFC4949]

3. The Concept of Trustword Mapping

3.1. Example

A fingerprint typically looks like:

F482 E952 2F48 618B 01BC 31DC 5428 D7FA ACDC 3F13

Its mapping to English Trustwords could look like:

dog house brother town fat bath school banana kite task

Or its mapping to German Trustwords could like like:

klima gelb lappen weg trinken alles kaputt rasen rucksack durch

Instead of the former hexadecimal string, users can compare ten common words of their language.

Note: This examples are for illustration purposes only and do not make use any any published Trustword list.

3.2. Previous work

The basic concept of Trustwords mapping has been already documented in the past, e.g. for use in One-Time Passwords (OTP) [RFC1751] [RFC1760] [RFC2289] or the PGP Word List ("Pretty Good Privacy word list" [PGP.wl], also called a biometric word list, to compare fingerprints.

Regarding today's needs, previous proposals have the following shortcomings:

- Limited number of Trustwords (small Trustword dictionaries), which generally results in more Trustwords to be compared
- Usually only available in English language, which does not normally allow its usage by non-English speakers in a secure manner

Furthermore, there are differences in the basic concept:

- This work allows for better tailoring the target audience to ordinary human users, i.e. not technical stuff (or IT geeks) only.
- As in many usage scenarios the Trustwords are only read (and compared), but not written down nor typed in by humans, there is a less strong need to keep the Trustwords themselves short. One such scenario is to use a side channel (e.g. phone) to compare the Trustwords. In fact longer Trustwords increases increase the entropy, as the dictionary is larger and the likelihood for phonetic collision can be decreased.

3.3. Number of Trustwords for a language

If the number of Trustwords is low, a lot of Trustwords need to be compared, which make a comparison somewhat cumbersome for users. This may lead to degraded usability.

To reduce the number of Trustwords to compare, in pEp's proposition of Privacy by Default [I-D.birk-pep] 16-bit scalars are mapped to natural language words. Therefore, the size (by number of key - value pairs) of any key - value pair structure is 65536. However, the number of unique values to be used in a language may be less than 65536. This can be addressed e.g. by using the same value (Trustword) for more than one key. In these cases, the entropy of the representation is slightly reduced. (Example: A Trustwords list of just 42000 words still allows for an entropy of log_2(42000) ~= 15.36 bits in 16-bit mappings.)

On the other hand, small sized Trustword lists allow for Trustwords with shorter strings, which are easier to use in scenarios where Trustwords have to be typed in e.g. OTP applications.

The specification allows for different dictionary sizes.

3.4. Language

Although English is rather widespread around the world, the vast majority of the worlds' population does not speak English. For an application to be useful for ordinary people, localization is a must. Thus, Trustword lists in different languages can be registered.

For applications where two human establish communication it is very likely that they share a common language. So far no real use case for translations between Trustword lists in different languages has been identified. As translations also drastically increases the complexity for IANA registrations, translations of Trustwords beyond the scope of this document.

3.5. The nature of the words

Every Trustwords list SHOULD be cleared from swearwords in order to not offense users. This is a task to be carried out by speakers of the respective natural language (i.e., by native language speakers).

4. IANA Considerations

Each natural language requires a different set of Trustwords. To allow implementers for identical Trustword lists, a IANA registry is to be established. The IANA registration policy according to [RFC8126] is "Expert Review" and "Specification Required".

[[Note: Further details of the IANA registry and requirements for the expert to assess the specification are for further study. A similar approach as used in [RFC6117] is likely followed.]]

4.1. Registration Template (XML chunk)

```
<record>
  <languagecode>
    <!--  ISO 639-3 (e.g. eng, deu, ...) -->
  </languagecode>
  <bitsize>
    <!-- How many bits can be mapped with this list
         (e.g. 8, 16, ...) -->
  </bitsize>
  <numberofuniquewords>
    <!-- number of unique words registered
         (e.g. 256, 65536, ...) -->
  </numberofuniquewords>
  <bijective>
    <!-- whether or not the list allows for a two-way-mapping
         (e.g. yes, no) -->
  </bijective>
  <version>
    <!-- version number within language
         (e.g. b.1.2, n.0.1, ...)  -->
  </version>
  <registrationdocs>
    <!-- Change accordingly -->
    <xref type="rfc" data="rfc2551"/>
  </registrationdocs>
  <requesters>
    <!-- Change accordingly -->
    <xref type="person" data="John_Doe"/>
    <xref type="person" data="Jane_Dale"/>
  </requesters>
  <additionalinfo>
    <paragraph>
      <!-- Text with additional information about
           the Wordlist to be registered -->
    </paragraph>
    <artwork>
      <!-- There can be artwork sections, too -->
    </artwork>
  </additionalinfo>
  <wordlist>
    <!-- Change accordingly -->
    <w0>first</w0>
    <w1>second</w1>
    [...]
    <w65535>last<w65535>
  </wordlist>
</record>
```

```
  <people>
    <person id="John_Doe">
      <name> <!-- Firstname Lastname --> </name>
      <org> <!-- Organization Name --> </org>
      <uri> <!-- mailto: or http: URI --> </uri>
      <updated> <!-- date format YYYY-MM-DD --> </updated>
    </person>
    <!-- repeat person section for each person -->
  </people>
```

Authors of a Wordlist are encouraged to use these XML chunks as a template to create the IANA Registration Template.

4.2. IANA Registration

An IANA registration will contain the fallowing elements:

4.2.1. Language Code (<languagecode>)

The language code follows the ISO 639-3 specification [ISO693], e.g., eng, deu.

[[Note: It is for further study, which of the ISO 639 Specifications is most suitable to address the Trustwords' challenge.]]

Example usage for German:

e.g. <languagecode>deu</languagecode>

4.2.2. Bit Size (<bitsize>)

The bit size is the number of bits that can be mapped with the Wordlist. The number of registered words in a word list MUST be 2 ^ "(<bitsize>)".

Example usage for 16-bit Wordlist:

e.g. <bitsize>16</bitsize>

4.2.3. Number Of Unique Words (<numberofuniquewords>)

The number of unique words that are registered.

e.g. <numberofuniquewords>65536</numberofuniquewords>

4.2.4. Bijectivity (<bijective>)

Whether the registered Wordlist has a one-to-one mapping, meaning the number of unique words registered equals 2 ^ "(<bitsize>)".

Valid content: (yes | no)

e.g. <bijective>yes</bijective>

4.2.5. Version (<version>)

The version of the Wordlist MUST be unique within a language code.

[[Note: Requirements to a "smart" composition of the version number are for further study]]

e.g. <version>b.1.2</version>

4.2.6. Registration Document(s) (<registrationdocs>)

Reference(s) to the Document(s) containing the Wordlist

e.g.
```
<registrationdocs>
  <xref type="rfc" data="rfc4979"/>
</registrationdocs>
```

e.g.
```
<registrationdocs>
  <xref type="rfc" data="rfc8888"/> (obsoleted by RFC 9999)
  <xref type="rfc" data="rfc9999"/>
</registrationdocs>
```

e.g.
```
<registrationdocs>
  [International Telecommunications Union,
  "Wordlist for Foobar application",
  ITU-F Recommendation B.193, Release 73, Mar 2009.]
</registrationdocs>
```

4.2.7. Requesters (<requesters>)

The persons requesting the registration of the Wordlist. Usually these are the authors of the Wordlist.

```
e.g.  <requesters>
        <xref type="person" data="John_Doe"/>
      </requesters>

      <people>
        <person id="John_Doe">
          <name>John Doe</name>
          <org>Example Inc.</org>
          <uri>mailto:john.doe@example.com</uri>
          <updated>2018-06-20</updated>
        </person>
      </people>
```

Note: If there is more than one requester, there must be one <xref> element per requester in the <requesters> element, and one <person> chunk per requester in the <people> element.

4.2.8. Further Information (<additionalinfo>)

Any other information the authors deem interesting.

```
e.g.  <additionalinfo>
        <paragraph>more info goes here</paragraph>
      </additionalinfo>
```

Note: If there is no such additional information, then the <additionalinfo> element is omitted.

4.2.9. Wordlist (<wordlist>)

The full Wordlist to be registered. The number of words MUST be a power of 2 as specified above. The element names serve as key used for enumeration of the Trustwords (starting at 0) and the elements contains the values being individual natural language words in the respective language.

```
e.g.  <wordlist>
        <w0>first</w0>
        <w1>second</w1>
        [...]
        <w65535>last<w65535>
      </wordlist>

] ]>
```

[[Note: The exact representation of the Wordlist is for further study.]]

5. Security Considerations

There are no special security considerations.

6. Acknowledgements

The authors would like to thank the following people who have provided feedback or significant contributions to the development of this document: Andrew Sullivan, Claudio Luck, Daniel Kahn Gilmore, Michael Richardson, Rich Salz, and Yoav Nir.

This work was initially created by pEp Foundation, and then reviewed and extended with funding by the Internet Society's Beyond the Net Programme on standardizing pEp. [ISOC.bnet]

7. References

7.1. Normative References

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

[RFC4949] Shirey, R., "Internet Security Glossary, Version 2", FYI 36, RFC 4949, DOI 10.17487/RFC4949, August 2007, <https://www.rfc-editor.org/info/rfc4949>.

[RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <https://www.rfc-editor.org/info/rfc8126>.

7.2. Informative References

[E-D.birk-pep-handshake] Marques, H., "pretty Easy privacy (pEp): Contact Authentication through Handshake", June 2018, <https://pep.foundation/dev/repos/internet-drafts/file/tip/pep-handshake/draft-marques-pep-handshake-00.txt>.

Early draft

[E-D.birk-pep-keysync]
Birk, V. and H. Marques, "pretty Easy privacy (pEp): Key Synchronization Protocol", June 2018, <https://pep.foundation/dev/repos/internet-drafts/file/tip/pep-keysync/draft-birk-pep-keysync-NN.txt>.

Early draft

[I-D.birk-pep]
Birk, V., Marques, H., Shelburn, S., and S. Koechli, "pretty Easy privacy (pEp): Privacy by Default", draft-birk-pep-01 (work in progress), January 2018.

[ISO693]
"Language codes - ISO 639", n.d., <https://www.iso.org/iso-639-language-codes.html>.

[ISOC.bnet]
Simao, I., "Beyond the Net. 12 Innovative Projects Selected for Beyond the Net Funding. Implementing Privacy via Mass Encryption: Standardizing pretty Easy privacy's protocols", June 2017, <https://www.internetsociety.org/blog/2017/06/12-innovative-projects-selected-for-beyond-the-net-funding/>.

[PGP.wl]
"PGP word list", November 2017, <https://en.wikipedia.org/w/index.php?title=PGP_word_list&oldid=749481933>.

[RFC1751]
McDonald, D., "A Convention for Human-Readable 128-bit Keys", RFC 1751, DOI 10.17487/RFC1751, December 1994, <https://www.rfc-editor.org/info/rfc1751>.

[RFC1760]
Haller, N., "The S/KEY One-Time Password System", RFC 1760, DOI 10.17487/RFC1760, February 1995, <https://www.rfc-editor.org/info/rfc1760>.

[RFC2289]
Haller, N., Metz, C., Nesser, P., and M. Straw, "A One-Time Password System", STD 61, RFC 2289, DOI 10.17487/RFC2289, February 1998, <https://www.rfc-editor.org/info/rfc2289>.

[RFC3647]
Chokhani, S., Ford, W., Sabett, R., Merrill, C., and S. Wu, "Internet X.509 Public Key Infrastructure Certificate Policy and Certification Practices Framework", RFC 3647, DOI 10.17487/RFC3647, November 2003, <https://www.rfc-editor.org/info/rfc3647>.

[RFC6117] Hoeneisen, B., Mayrhofer, A., and J. Livingood, "IANA Registration of Enumservices: Guide, Template, and IANA Considerations", RFC 6117, DOI 10.17487/RFC6117, March 2011, <https://www.rfc-editor.org/info/rfc6117>.

[RFC6120] Saint-Andre, P., "Extensible Messaging and Presence Protocol (XMPP): Core", RFC 6120, DOI 10.17487/RFC6120, March 2011, <https://www.rfc-editor.org/info/rfc6120>.

Appendix A. IANA XML Template Example

This section contains a non-normative example of the IANA Registration Template XML chunk.

```
  <record>
    <languagecode>lat</languagecode>
    <bitsize>16</bitsize>
    <numberofuniquewords>57337</numberofuniquewords>
    <bijective>no</bijective>
    <version>n.0.1</version>
    <registrationdocs>
      <xref type="rfc" data="rfc2551"/>
    </registrationdocs>
    <requesters>
      <xref type="person" data="Julius_Caesar"/>
    </requesters>
    <additionalinfo>
      <paragraph>
        This Wordlist has been optimized for
        the Roman Standards Process.
      </paragraph>
    </additionalinfo>
    <wordlist>
      <w0>errare</w0>
      <w1>humanum</w1>
      [...]
      <w65535>est<w65535>
    </wordlist>
  </record>

  <people>
    <person id="Julius_Caesar">
      <name>Julius Caesar</name>
      <org>Curia Romana</org>
      <uri>mailto:julius.cesar@example.com</uri>
      <updated>1999-12-31</updated>
    </person>
  </people>
```

Appendix B. Document Changelog

[[RFC Editor: This section is to be removed before publication]]

- draft-birk-pep-trustwords-02:
  - Minor editorial changes and bug fixes
  - Added more items to Open Issues
  - Add usage example
- draft-birk-pep-trustwords-01:
  - Included feedback from mailing list and IETF-101 SECDISPATCH WG, e.g.
    - Added more explanatory text / less focused on the main use case
    - Bit size as parameter
  - Explicitly stated translations are out-of-scope for this document
  - Added draft IANA XML Registration template, considerations, explanation and examples
  - Added Changelog to Appendix
  - Added Open Issue section to Appendix

Appendix C. Open Issues

[[RFC Editor: This section should be empty and is to be removed before publication]]

- More explanatory text for Trustword use cases, properties and requirements
- Further details of the IANA registry and requirements for the expert to assess the specification
- Decide which ISO language code either 639-1 or 639-3 to use, i.e., ISO-639-1 (e.g., ca, de, en, ...) as currently used in pEp implementations (running code) or ISO-693-3 (eng, deu, ita, ...)
- Adjust exact representation of wordlists

    * e.g. XML, CSV, ...

    * Syntax for non-ASCII letters or language symbols (UTF-8) in Wordlists

- Need for optional entropy value assigned to words, to account for similar phonetics among words in the same wordlist?

- Need for an additional field, to define what a wordlist is optimized for, e.g., "entropy", "minimize word lengths", ...?

- Work out (requirements for) "smart" composition of the version number

- Decide whether in non-bijective Wordlists the redundant words need to be repeated in the IANA Registration

- Register only a hash over the wordlist with IANA?

- Does it make sense to open registrations for other patterns than just words, e.g., images?

Authors' Addresses

Volker Birk
pEp Foundation
Oberer Graben 4
CH-8400 Winterthur
Switzerland

Email: volker.birk@pep.foundation
URI: https://pep.foundation/

Hernani Marques
pEp Foundation
Oberer Graben 4
CH-8400 Winterthur
Switzerland

Email: hernani.marques@pep.foundation
URI: https://pep.foundation/

Bernie Hoeneisen
Ucom Standards Track Solutions GmbH
CH-8046 Zuerich
Switzerland

Phone: +41 44 500 52 44
Email: bernie@ietf.hoeneisen.ch (bernhard.hoeneisen AT ucom.ch)
URI: https://ucom.ch/