Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 9, 2020

B. Hoeneisen
Ucom.ch
H. Marques
pEp Foundation
July 08, 2019

IANA Registration of Trustword Lists: Guide, Template and IANA Considerations
draft-birk-pep-trustwords-04

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), which hexadecimal strings are mapped to. Such a mapping makes verification processes like fingerprint comparisons 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 January 9, 2020.

Copyright Notice

Table of Contents

1. Introduction 2
1.1. Requirements Language 3
1.2. Terms 3
2. The Concept of Trustword Mapping 4
2.1. Example 4
2.2. Previous work 4
2.3. Number of Trustwords for a language 5
2.4. Language 5
2.5. The nature of the words 6
3. Security Considerations 6
4. Privacy Considerations 6
5. IANA Considerations 6
5.1. Registration Template (XML chunk) 6
5.2. IANA Registration 8
5.2.1. Language Code (<languagecode>) 8
5.2.2. Bit Size (<bitsize>) 8
5.2.3. Number Of Unique Words (<numberofuniquewords>) . . . 8
5.2.4. Bijectivity (<bijective>) 8
5.2.5. Version (<version>) 8
5.2.6. Registration Document(s) (<registrationdocs>) 9
5.2.7. Requesters (<requesters>) 9
5.2.8. Further Information (<additionalinfo>) 9
5.2.9. Wordlist (<wordlist>) 10
6. Acknowledgments 10
7. References 10
7.1. Normative References 10
7.2. Informative References 11
Appendix A. IANA XML Template Example 12
Appendix B. Document Changelog 13
Appendix C. Open Issues 14
Authors' Addresses 15

1. Introduction

In public-key cryptography, comparing the respective public key fingerprints for each of the communication partners involved is vital to ensure that there is no Man-in-the-Middle (MITM) attack on the communication channel. These fingerprints normally consist of a chain of hexadecimal characters, which are often impractical, cumbersome, and prone to misunderstandings for end-users.

To mitigate these challenges, several systems offer Trustword comparison as an alternative to these hexadecimal strings. Trustwords are common words in a natural language (e.g., English), which these hexadecimal strings are mapped to. Using Trustwords makes verification processes like fingerprint comparisons more natural for users.

For example, in pEp's Privacy by Default proposition [I-D.birk-pep] Trustwords are used to facilitate easy contact verification for end-to-end encryption. Trustword comparison is offered after the peers have opportunistically exchanged public keys. Examples of Trustword lists used by current pEp implementations can be found here in CSV format: 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, deriving private keys in blockchain applications for cryptocurrencies, and to import or synchronize secret keys across multiple devices owned by a single user [I-D.hoeneisen-pep-keysync]. Further ideas include the use of Trustwords for private key recovery in case of loss, contact verification in Extensible Messaging and Presence Protocol (XMPP) [RFC6120], or for X.509 certificate verification in browsers [RFC3647].

1.1. Requirements Language

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].

1.2. Terms

The following terms are defined for the scope of this document:

- pEp Handshake: The process of one user contacting another over an independent channel in order to verify Trustwords (or by fallback: fingerprints). This can be done in-person or through established verbal communication channels, like a phone call. [I-D.marques-pep-handshake]

- Man-in-the-middle (MITM) attack: cf. [RFC4949], which states: "A form of active wiretapping attack in which the attacker intercepts and selectively modifies communicated data to masquerade as one or more of the entities involved in a communication association."

2. The Concept of Trustword Mapping

2.1. Example

As already discussed, fingerprints normally consist of a string of hexadecimal characters. A typical fingerprint looks like this:

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

Instead of the hexadecimal string, Trustwords allow users to compare ten common words of a language of their choosing. For example, the above fingerprint, mapped to English Trustwords, might appear as:

dog house brother town fat bath school banana kite task

The same fingerprint might appear in German Trustwords as:

klima gelb lappen weg trinken alles kaputt rasen rucksack durch

Note: These examples are for illustration purposes only, and are not derived from any published Trustword list.

2.2. Previous work

The basic concept of Trustword mapping - also known as a biometric word list - for fingerprint comparison is well-documented. Examples of this concept are used with One-Time Passwords (OTP) [RFC1751] [RFC1760] [RFC2289], as well as the PGP Word List ("Pretty Good Privacy word list" [PGP.wl]. Furthermore, cryptocurrencies use a similar concept for deriving private keys [bitcoin.wl].

[[TODO: Explain each previous usage a bit further and synchronize with section Section 1.]]

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

- Small/limited word lists, which generally result in more words to compare
- Existing word lists are usually only available in English, which limits their usefulness for non-English speakers

Furthermore, there are differences in the basic concept:

- The Trustword concept suggested herein intends to improve usability and security for all users, instead of only the technically-savvy.

- In many use cases, Trustwords are only read (aloud) during the comparison process, rather than being written or typed. For example, two users might compare their respective Trustwords during a phone call. Verbal comparison reduces the need to keep the actual Trustwords short. The use of longer Trustwords increases the entropy within the system, as it allows for a larger dictionary, and thus reduces the likelihood of phonetic collisions.

2.3. Number of Trustwords for a language

If the number of Trustwords in a dictionary is low, shorter parts of the original string (e.g., fingerprint) can be mapped to a single Trustword. Thus, many Trustwords will need to be compared, which results in a potentially cumbersome process for users, and lead to reduced usability.

To reduce the number of Trustwords that need to be compared, pEp's Privacy by Default proposition [I-D.birk-pep] calls for 16-bit scalars to be 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 smaller than this number. This discrepancy can be addressed by using the same value, or Trustword, for more than one key. In such cases, the entropy of the representation is slightly reduced. For example, a Trustword list of 42000 words still allows for an entropy of log_2(42000), which is roughly 15.36 bits in 16-bit mappings. As a consequence such Trustword lists are not bijective.

On the other hand, small Trustword lists allow for Trustwords consisting of words with shorter strings (number of short words per natural language is normally limited), which are easier to use in implementations where Trustwords have to be typed or written, such as in OTP applications.

Note: This specification allows for registration of variable numbers of Trustwords per dictionary.

2.4. Language

Although English is used around the world, the vast majority of the global population is not English-speaking. For an application to be useful to as wide of a user base as possible, localization is essential. Therefore, this specification allows for registration of Trustword lists in different languages.

In applications where two humans are attempting to establish secure communications, it is likely that they share a common language. At

this time, no real-world use cases for Trustword list translation capability have been identified. Because the translation process inherently - and drastically - increases complexity from an IANA registration standpoint, the topic of Trustword translation is beyond the scope of this document.

2.5. The nature of the words

Every Trustword list SHOULD be clear of offensive language (i.e., swear/curse words, slurs, derogatory language, etc.). This process SHOULD be performed by native speakers of each respective language.

3. Security Considerations

There are no specific security considerations.

4. Privacy Considerations

[[TODO]]

5. 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.]]

5.1. 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.

5.2. IANA Registration

An IANA registration will contain the fallowing elements:

5.2.1. Language Code (<languagecode>)

The language code follows the ISO 639-3 specification [ISO639], 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>

5.2.2. 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>

5.2.3. Number Of Unique Words (<numberofuniquewords>)

The number of unique words that are registered.

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

5.2.4. 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>

5.2.5. 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>

5.2.6. 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>
```

5.2.7. 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.

5.2.8. 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.

5.2.9. 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.]]

6. Acknowledgments

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, Kelly Bristol, Michael Richardson, Rich Salz, Volker Birk, 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

[bitcoin.wl] "Seed Phrase", June 2019, <https://en.bitcoin.it/w/index.php?title=Seed_phrase&oldid=66492#Word_Lists>.

[I-D.birk-pep] Marques, H. and B. Hoeneisen, "pretty Easy privacy (pEp): Privacy by Default", draft-birk-pep-03 (work in progress), March 2019.

[I-D.hoeneisen-pep-keysync] Hoeneisen, B. and H. Marques, "pretty Easy privacy (pEp): Key Synchronization Protocol", draft-hoeneisen-pep-keysync-00 (work in progress), July 2019.

[I-D.marques-pep-handshake] Marques, H. and B. Hoeneisen, "pretty Easy privacy (pEp): Contact and Channel Authentication through Handshake", draft-marques-pep-handshake-03 (work in progress), July 2019.

[ISO639] "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]]

o draft-birk-pep-trustwords-04:

* Add Privacy Considerations section

* Swapped Security and IANA Consideration Sections

* Corrected typo in ISO references

* Updated Introduction, Terms and concept Sections

- draft-birk-pep-trustwords-03:
  - Update references
  - Minor edits
- 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.]]

- Better explain previous work on Trustwords
- 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-639-3 (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

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

Phone: +41 44 500 52 40
Email: bernie@ietf.hoeneisen.ch (bernhard.hoeneisen AT ucom.ch)
URI: https://ucom.ch/

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

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