Analyzing Security Considerations

Alan Mills, UWE – Bristol

Mark McFadden, internet policy advisors ltd

IRTF maprg & RIPE MATWG

August 2020

Motivations

- RFC 3552 provides guidance to authors in crafting RFC text on security considerations
    - But the RFC is more than 15 years old
    - Presumably the threat landscape has changed in 15 years
- Take a look at RFCs since 2003 and see how the language in the RFCs has changed in the face of new developments in security
- Add to the conversation about revising RFC 3552 to address changes in the threat landscape and to give better guidance to developers of protocols for effective Security Considerations sections

Further Background

- IAB has previously discussed a potention revision to RFC 3552 in its report from the Strengthening the Internet (STRINT) Workshop
- IAB currently has an effort moving forward to reconsider the Internet’s threat model (called Model-T)
- Body of work to examine is considerable
  - Published RFCs since RFC 3552

Methodology

- draft-mcfadden-smart-rfc3552-textual-research-01
- Proposes two separate components
  - a quantitative examination which might survey and collect data on the source of the RFC (e.g. Security Area, Routing Area, Transport Area), whether the RFC extends the Security Considerations section of a previously published document, the wordcount of the section, and the existence of specific keywords
  - a qualitative analysis might group Security Considerations sections by particular characteristics – those characteristics being discovered, in part, during an initial examination of the published documents
- Then reports on an experiment

What Happened Next?

- The methodology proposed is in search of comments for improvement
  - It hasn’t been executed yet
- However, a quantitative experiment has been done
  - The draft reports on the results of this experiment

Experiment

- If you did a simple, word-by-word textual analysis of the security considerations sections of all RFCs since 2003 . . .
  - Would you detect changes in the words being used that indicate that the threats, threat model or mitigations had changed over time?
  - Would you find patterns that indicate that some terms have become more important to protocol designers by virtue of appearing in more (or, less) RFCs over time?
  - Would you detect other changes in the language used for security considerations that would help with a revision to the language used to guide protocol designers in RFC 3552?
  - Had significant security incidents changed the way security considerations sections were written?

Quantitative Analysis

- One of the authors has conducted an experiment that is consistent with many of the features of the quantitative methodology
- This experiment uses a pair of Python scripts to extract the Security Considerations sections from historic RFCs and then parse those sections to get word frequency information from those Security Considerations

Parser

- The RFC series was divided into groups of input files for an open source parser by year of publication of the RFC
- The parser removed non-textual material from the temporary files including hyphens, RFC references, anchor URLs, other sections references, standalone letters and other characters that were not words
- It then built a frequency list for all words not in a designated list of words not to be counted.
  - This list is a variable and could be changed to include, or exclude, words from the designated list.

Analysis Output

- The result of this experiment is a pair of files for each year starting in 2003.
- The two files for each year are:
  - A word frequency file sorted by the number of times a particular word appears in the Security Considerations section of RFCs published in that year; and,
  - A RFC Count file that counts how many times each RFC was mentioned within the Security Considerations sections.

Top Ten Word Counts in Four Sample Years

- 2019 – security, server, data, message, may, network, attack, information, client, xmpp-grid
- 2014 – security, information, attack, message, may, used, server, data, authentication, network
- 2009 – security, may, message, address, attack, used, packet, protocol, network, information
- 2004 – security, may, key, authentication, object, used, information, message, attack, access
    - But what if you remove RFC 2119 normative language?

Removing Normative Language

- 2019 – security, server, data, message, network, attack, information, client, xmpp-grid, document
- 2014 – security, information, message, used, server, data, authentication, network, attacker
- 2009 – security, message, address, attack, used, packet, protocol, network, information, object
- 2004 – security, key, authentication, object, used, information, message, attack, access, user

Other Results

- Over the entire period 2003-2019, the most frequent non-normative words in Security Considerations sections were:
  - Security, message, attack, server, information, key, authentication, network, protocol, client
- Over the entire period 2003-2019, the 75 most frequent words in Security Considerations sections was (in order by frequency):
  - security, message, attack, data, used, may, authentication, key, access, protocol, information, must, address, transport, process, model, client, server, network, ipfix, tl, user, traffic, packet, object, operation, control, service, ipp, example, document, implementation, measurement, collecting, secure, header, attacker, identity, value, job, need, support, snmp, provide, printer, uri, certificate, authenticated, possible, name, content, source, connection, field, set, system, dtls, cause, sensitive, domain, provides, configuration, router, privacy, protection, peer, nacm, layer, ip, device, exporting, within, request, large, and signature.

Possible Conclusions

- Quantitative textual analysis of RFCs since 2003 does not reveal major changes in the language being used in security considerations sections
  - Is that counter-intuitive?
  - Would we have expected more changes to evolve naturally as the threat landscape changed and new threats and actors emerged?
  - Does this give any guidance for future considerations for making changes to RFC 3552?
- The word MAY always appears more often than any other RFC2119 word in Security Considerations sections. The word MUST most often appears after MAY and is often in the top 15 words sorted by frequency.
  - However, the word SHOULD hardly ever appears in the top 100 most frequent words for any year of published RFCs.
  - Is this accidental, or does it reflect the way security considerations sections are written?

Thanks!

- Mark McFadden
  - mark<at>internetpolicyadvisors.com
- Alan Mills
  - Alan2.Mills@live.uwe.ac.uk