A "Duck Test" for End-to-End Secure Messaging

draft-muffett-end-to-end-secure-messaging-03

alec.muffett@gmail.com || @alecmuffett

presentation v1.18 (abridged version)

What is <this> I-D about?

draft-muffett-end-to-end-secure-messaging (E2ESM)

- "end-to-end"
- "secure"
  - …including "end-to-end encrypted" (E2EE)
- "messaging"

"…but we already understand this!"

thesis:
intuitive understandings
of E2ESM / E2EE
are no longer politically sufficient

What issue does <this> address?

"…if you want to change something,
you must first be able to measure it…"

— various, mostly after Drucker

therefore…

"…if you want something to not change,
you must also be able to measure it…"

```
if [ $change = 0 ] ; then echo THING IS UNCHANGED ; else echo THING CHANGED ; fi
```

"end-to-end security"
and a few of the
extant attempts to change it

UNICEF

"Encryption, Privacy and Children's Right to Protection from Harm"

- *This solution is portrayed as more data protection friendly compared to exceptional access, as it **still upholds end-to-end encryption and its data protection benefits by filtering the communication at the level of the transmitting device**.*

One of the creators of PhotoDNA asserts that due to recent advances in encryption and robust hashing technology it would be possible to adapt PhotoDNA for use within an end-to-end encrypted system.[25] This would enable images to be analysed against the database of hashes maintained by NCMEC without the need for decryption. It is not clear whether this proposal solves the privacy concerns from a technical point of view, but it seems an important angle that UNICEF could pursue further with partners in the technology sector.

Client-side scanning of images

An alternative option to the exceptional access solution is client-side scanning of images. Client-side scanning means that any outgoing communication flow from a personal device, whether using an encrypted communication system or not, is checked against a hash list of known child sexual abuse images. If there is a match, either the system refuses to send the message, reports the attempt to law enforcement or NCMEC, or a combination of these responses. This solution is portrayed as more data protection friendly compared to exceptional access, as it still upholds end-to-end encryption and its data protection benefits by filtering the communication at the level of the transmitting device.

However, this approach risks providing a blueprint for mass surveillance, as it may not be possible for the user or civil society to monitor the hash list used by their phone to ensure that it was only reporting or preventing the transmission of child sexual abuse images. Hashes for other sensitive but legal content (such as political or sexual) could be added to the database and without the user's knowledge.[26] Furthermore, it deteriorates the purpose of end-to-end encryption relating to freedom of information and expression, as the content of the communication is filtered by default. But despite its limitations in relation to privacy and security, it has been suggested that end-point scanning of images would probably do more to systematically address child sexual abuse materials online compared to providing exceptional access to law enforcement.[27]

https://www.unicef-irc.org/publications/pdf/Encryption_privacy_and_children's_right_to_protection_from_harm.pdf

GCHQ

"Ghost" Proposal

- *You end up with everything still being end-to-end encrypted, but there's an extra 'end' on this particular communication … **We're not talking about weakening encryption or defeating the end-to-end nature of the service.** In a solution like this, we're normally talking about suppressing a notification on a target's device, and only on the device of the target and possibly those they communicate with. That's a very different proposition…*

Principles in Practice

So, to some detail. For over 100 years, the basic concept of voice intercept hasn't changed much: crocodile clips on telephone lines. Sure, it's evolved from real crocodile clips in early systems through to virtual crocodile clips in today's digital exchanges that copy the call data. But the basic concept has remained the same. Many of the early digital exchanges enacted lawful intercept through the use of conference calling functionality.

In a world of encrypted services, a potential solution could be to go back a few decades. It's relatively easy for a service provider to silently add a law enforcement participant to a group chat or call. The service provider usually controls the identity system and so really decides who's who and which devices are involved - they're usually involved in introducing the parties to a chat or call. You end up with everything still being end-to-end encrypted, but there's an extra 'end' on this particular communication. This sort of solution seems to be no more intrusive than the virtual crocodile clips that our democratically elected representatives and judiciary authorise today in traditional voice intercept solutions and certainly doesn't give any government power they shouldn't have.

We're **not** talking about weakening encryption or defeating the end-to-end nature of the service. In a solution like this, we're normally talking about suppressing a notification on a target's device, and **only** on the device of the target and possibly those they communicate with. That's a very different proposition to discuss and you don't even have to touch the encryption.

The problem of gaining access to a seized encrypted device is very different and may well end up being harder to do in a proportionate way – there's not enough research to be sure either way. The apps and services we're talking about are usually just software -

https://www.lawfareblog.com/principles-more-informed-exceptional-access-debate

Indian Government

"message originator hashing"

- *The **hash can travel with the message and in case of any unlawful activity, the originator of the message can be traced without breaking the app's encryption,** the sources said …*

The **government** has proposed that **WhatsApp** assign an alpha-numeric hash to every message sent through its platform as a solution to break the deadlock over **traceability** on the messaging app, senior government officials told ET. The hash can travel with the message and in case of any unlawful activity, the originator of the message can be traced without breaking the app's **encryption**, the sources said.

"The government is willing to work with WhatsApp to come up with a solution to enable traceability of message originators without breaking encryption," according to officials in the know.

In February, the Centre notified the Information Technology (Guidelines for Intermediaries and Digital Media Ethics Code) Rules, 2021 that mandates traceability of first originator of a message flagged by either a court of law or an authorised government agency.

https://economictimes.indiatimes.com/tech/technology/govt-proposes-alpha-numeric-hash-to-track-whatsapp-chat/articleshow/81638939.cms

"*<proposal> does not break E2E security…*"

— how may we test this assertion?

"Does this messenger look like E2ESM?"

- We could define E2ESM via:
  - what algorithms do we expect?
  - what features do we expect?
  - who are the actors we expect?
    - …and what (in turn) are their expectations?

draft-knodel-e2ee-definition

- separate effort to <this>
  - potentially complementary goals
- addresses specifically / only "end-to-end encryption" (E2EE)
  - frames E2EE in terms of "expectations"
- challenging to use as a "test" / resist political "reinterpretation"
  - example: …

Definition by User Expectation

draft-knodel-e2ee-definition-02

- *4.3 Access by a third-party is impossible*
  - *[…] If a method makes private communication, intended to be sent over an encrypted channel between end points, available to parties other than the sender and intended recipient(s), without formally interfering with channel confidentiality, that method violates the understood expectation of that security property.*

"Intention?" "Recipients intended by whom?" "Formal interference?"
What makes formality good? Should Law Enforcement truly be a "third party?"
What are the consequences of "violating expectations?"

alternative:

"Does this messenger quack like E2ESM?"

https://commons.wikimedia.org/wiki/Category:Ducks#/media/File:Baby_Duck_(26017170513).jpg

How to test end-to-end secure messaging

There are three really big hints in the name…

1/ end-to-end secure messaging

"there are ends. respect them."

- *proposition**
  - "end" = "participant" = ("sender" || "recipient")
  - **we shall revisit & refine this, shortly*

2/ end-to-end secure messaging

not all communications solutions are "messaging," and that's okay

- *proposition*
  - **at the point of sending of each individual message, the sender shall create the complete and immutable set of recipients for the message**
- *constraint*
  - **if future joiners - unknown to senders - can read past-sent content*, you instead are discussing a "forum" or similar, not "messaging"**
    - ***we shall revisit & refine this, shortly***

3/ end-to-end secure messaging

surprise plot twist!

- we do not define recipients in terms of protocol participation
- instead, we define recipients on the basis of outcomes:
  - "recipient": any entity which can determine one (1) bit of plaintext message content with more than than 50% certainty
    - include "the content is a member of this set of documents"
    - include "the content contains 1+ of this set of words/phrases"
    - include "the content is similar to 1+ images in this set"

if any recipient was not {known, visible} to the sender
at the point of message composition/sending,
the solution does not implement end-to-end security

QED

Bonus: obvious definition of "backdoor"!

- a "backdoor" is any mechanism which leaks bits to a non-recipient, irrespective of being intentional or unintentional
  - because "intention" is hard to determine objectively (cf: RFC 2804 §4) we should separate and clearly note our opinions re: intentionality
- some have criticised formalisation of "backdoor" as pejorative
  - would "un-/intentional sidechannel" or "opaque, undocumented legally-obligated exceptional access mechanism" be kinder?

Surveillance may be enabled as long it's overt

Also: Escrow, Recorders, Compliance, Helper Bots, etc…

- like public CCTV, shouldn't surveillance capability be transparent?
  - *Messages you send to this chat and calls are now secured with end-to-end encryption, but may be subject to interception or review by ourselves, and law enforcement, safety communities, and outsourced agents from the following national governments that we have determined from your profile information: […]*
  - if omniscient surveillance capability (where implemented) is treated as a visible participant, many UX & transparency challenges are obviated, and likelihood of crime occurring on the platform is greatly reduced

nitpicks & edge-cases

Nit 1 of 8

some metadata is almost as sensitive as content

- "yes" is 3 bytes, "no" is 2 bytes
  - unless you're (e.g.) French: ("oui" ‖ "si" vs: "non")
    - actually, this is a terrible example, but back to the point:
- tl;dr — exposing exact plaintext content size is "sensitive"
  - so: don't do that then / use padding wisely / …

Nit 2 of 8

some metadata may be beyond the scope of "content" protection

- thematic metadata leaks in the transport layer
  - recipient lists, group name, …

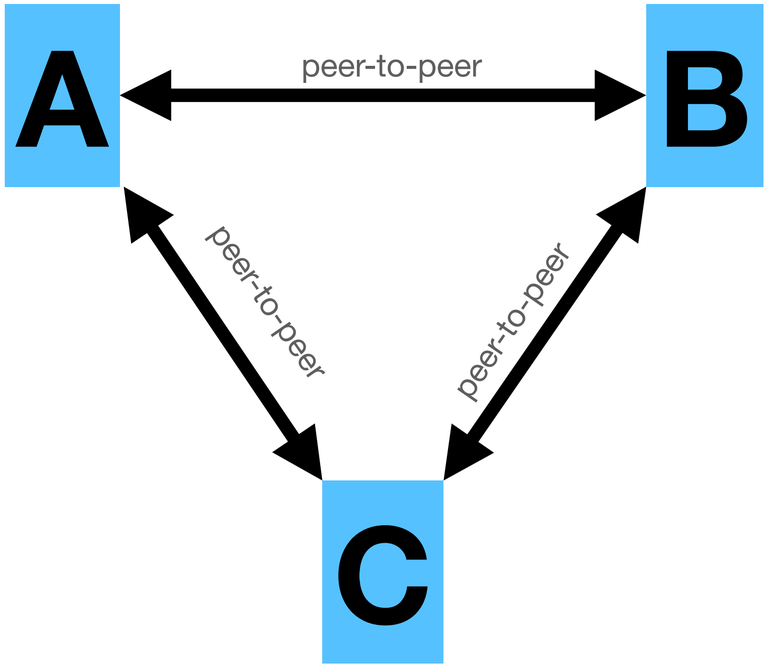
```
To: 111attendees@ietf.org, alec.muffett@gmail.com
Subject: encrypted message
---- BEGIN PGP WHATEVER ----
...
```

  - compare: WhatsApp group chat names are not E2EE content (!)
    - that which is out of scope of E2E security must be made apparent

Nit 3 of 8

if platforms are not necessary, message encryption may be optional

- if the risk is that one may be surveilled by intermediaries, why not simply do without intermediaries?
  - https://www.ricochetrefresh.net/
    - currently: *no message encryption*
    - hence: *end-to-end secure messaging*
  - https://cwtch.im/
  - https://briarproject.org/



Nit 4 of 8

"set of recipients" varies with centralised vs: distributed

- decentralised E2ESM: during composition the sender chooses the recipient set, the which is frozen at "point of sending" the message
  - e.g. PGP/Email, Ricochet, …
- centralised E2ESM: the recipient set is taken from visible, shared context amongst participants, frozen at the "point of sending" each message
  - e.g. {Signal, WhatsApp} group chat membership, moment to moment
  - todo: CAP / recipient-set consistency issues?

Nit 5 of 8

participation must be "closed from within"; but may be openable

- adding participants to a group chat must only be possible by explicit action of one (or more?) existing participants
  - it's okay for an existing participant to explicitly create and publish a means for the general public to self-subscribe to a group; this would not violate the definition of E2ESM
  - it's okay for an existing participant to explicitly (e.g.) republish messages to a public webpage; this would not violate the definition of "recipient", because recipient activity is a TCB issue (see upcoming)

Nit 6 of 8

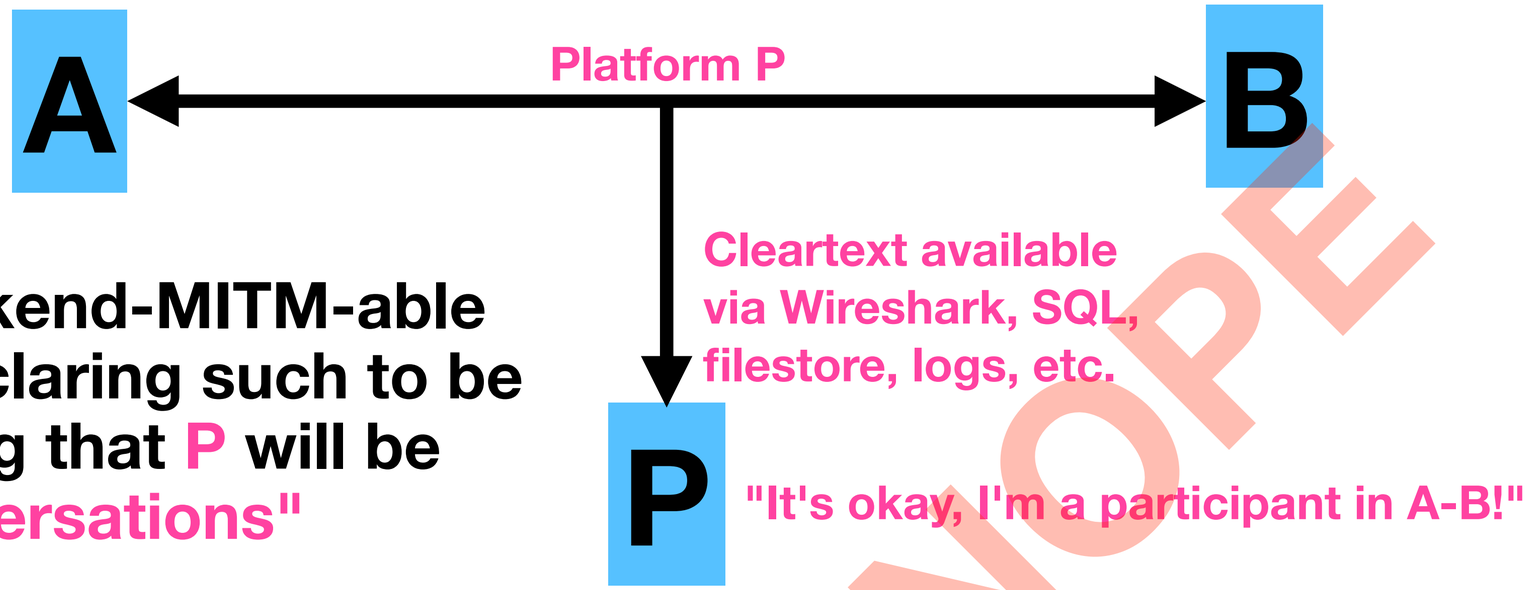
re-injection of old content for new recipients, can & does happen

- recipients quote, cite, or forward earlier messages, in new ones
  - including: reporting to platform safety or law enforcement teams
- (possibly incautious) clients re-encrypt bounced messages to existing recipients who are suddenly discovered to have new device fingerprints
  - *"In Response to Guardian's Irresponsible Reporting on WhatsApp: A Plea for Responsible and Contextualized Reporting on User Security" — Tufekci Z., June 2017; https://technosociology.org/?page_id=1687*
- none of these are "backdoors," they are features or design choices within a given solution

Nit 7 of 8

platforms may participate, but must be "peers" and must play fairly

- participants must have equal access to plaintext without "MITM" access; non-participants are already forbidden any access to plaintext



- …else P could offer backend-MITM-able messenger services, declaring such to be E2ESM through asserting that P will be "a participant in all conversations"

Nit 8 of 8

"end" actually means "trusted computing base (TCB) of an entity" …

- …and "end-to-end" actually means "trust-to-trust" / TCB-to-TCB
  - Clark, David D. and Blumenthal, Marjory S. (2011)
    "The End-to-End Argument and Application Design: The Role of Trust,"
    Federal Communications Law Journal: Vol. 63 : Iss. 2 , Article 3.
    Available at: https://www.repository.law.indiana.edu/fclj/vol63/iss2/3

The End-to-End Argument and Application Design: The Role of Trust

https://www.repository.law.indiana.edu/fclj/vol63/iss2/3/

- *Because the locus of trust is naturally at the ends, where the various principals are found,* ***"trust-to-trust" is preferable to "end-to-end" from the point of view of the principals, because it more directly invites the important question of "trusted by whom?"*** *That question, in turn, relates to questions that implicate application design, notably "who gets to choose which service is used?" or "which parts of an application are in which service modules?" Answers to these questions illuminate who controls what aspects of an application.*

IV. THE NEW END-TO-END

The discussion of what it means to be careful provides a framework for proposing a reformulation of the end-to-end argument for today's context: we can replace the end-to-end argument with a "trust-to-trust argument." The original paper said: "The function in question can completely and correctly be implemented only with the knowledge and help of the application standing at the endpoints of the communication system."[31] The generalization would be to say: The function in question can completely and correctly be implemented only with the knowledge and help of the application *standing at a point where it can be trusted to do its job in a reliable and trustworthy fashion*. Trust, in this context, should be determined by the ultimate end points—the principals that use the application to fulfill their purposes. Because the locus of trust is naturally at the ends, where the various principals are found, "trust-to-trust" is preferable to "end-to-end" from the point of view of the principals, because it more directly invites the important question of "trusted by whom?" That question, in turn, relates to questions that implicate application design, notably "who gets to choose which service is used?" or "which parts of an application are in which service modules?" Answers to these questions illuminate who controls what aspects of an application.

To reconstruct the end-to-end argument in the context of trust, we proceed in two steps. We first look at the range of options that each participant in the communication can take, based on their individual choices about trust, and then we look at the range of options that arise *jointly*, depending on the degree to which the various communicants trust each other. Trust-to-trust acknowledges that, unlike when the original paper was written, there is more reason for one end to question the trustworthiness of another and therefore more reason to seek something beyond simple end-to-end communication. As we noted in our earlier paper, the population of end users has become more diverse, and this raises questions for the end-to-end argument.[32]

"Danvers Doctrine" agreed to in 1995. *Id.* at 3.

31. Saltzer et al., *supra* note 1, at 278 (emphasis omitted).

32. Marjory S. Blumenthal & David D. Clark, *Rethinking the Design of the Internet: The End-to-End Arguments vs. The Brave New World*, 1 ACM TRANSACTIONS ON INTERNET

The Participant's Trusted Compute Base

it's not a messenger "backdoor" when…

- Alice accesses her messenger over RDP
- Bob has a hacked app on his jailbroken phone, via an insecure appstore
- Carol's phone storage is forensically analysed at rest
- Dave's keyboard app or grammar app leaks to his local authorities
  - That "trusted paths" & "secure attention keys" are core TCB issues, dates back to the "TCSEC Orange Book" (1983) & before…
  - …yet the debate continues today: https://twitter.com/RealSexyCyborg/status/1197695344575799296

Also from that paper…

https://www.repository.law.indiana.edu/fclj/vol63/iss2/3/

- *25. The **Internet Engineering Task Force** has addressed these concerns for over a decade, **declining to accept the task of designing corresponding [wiretap] protocols.** See Brian E. Carpenter & Fred Baker, IAB and IESG Statement on Cryptographic Technology and the Internet, IETF RFC 1984 (rel. Aug. 1996), http://www.ietf.org/rfc/rfc1984.txt; Brian E. Carpenter & Fred Baker, **IETF Policy on Wiretapping, IETF RFC 2804 (rel. May 2000)**, http://www.ietf.org/rfc/rfc2804.txt.*

services represent different "ends" of the application.

Lawful intercept. Lawful intercept, or government-ordered "wiretapping," is usually conceived as being implemented in the "middle" of the network. One approach is to carry out lawful intercept within the communications subsystem (e.g., the routers of the Internet). This would imply finding a router (perhaps one very close to the end node) that the traffic of interest is likely to pass through. Another idea is to identify some service at a higher layer (an "application layer" service) that is involved in the communication, and implement the intercept there. In the e-mail system, the mail servers are a natural point of intercept. For instant messaging, the IM server would be the target.

In order for an interceptor (lawful or otherwise) to locate a node or server through which the content is flowing, it may be necessary (or at least helpful) if this actor can constrain the set of choices, both technical and commercial, that the end user can exploit. If, because of technical design or economic or policy reasons, the end node is forced to use a particular server that can be easily identified, this makes the intercept much easier to carry out. If the end user can be prevented from using encryption (an obvious "end-to-end" reliability enhancement from the perspective of the communicating end users), the effectiveness of the intercept improves. Accordingly, the legal authorities might try to limit the use of encryption, either by influencing the development of standards, legal restrictions, making encryption hard to use and understand, and so on.[25]

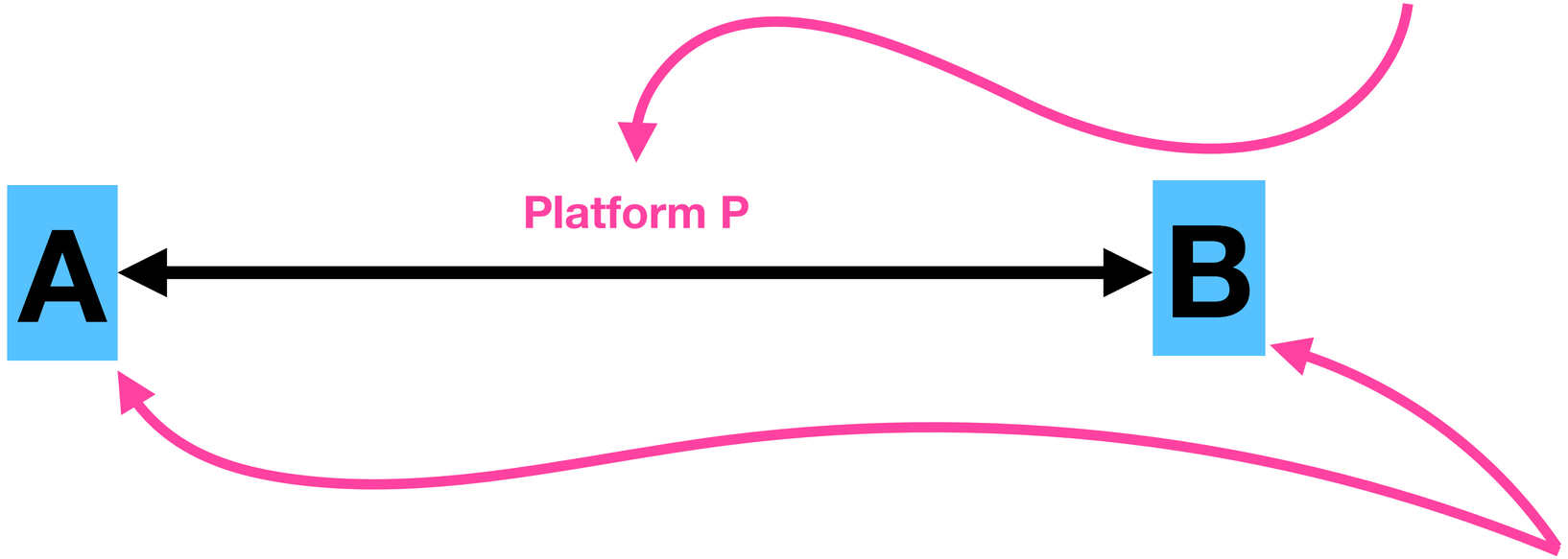
---

http://online.wsj.com/article/SB10001424052702304772804575558484075236968.html.

25. The Internet Engineering Task Force has addressed these concerns for over a decade, declining to accept the task of designing corresponding protocols. *See* Brian E. Carpenter & Fred Baker, *IAB and IESG Statement on Cryptographic Technology and the Internet*, IETF RFC 1984 (rel. Aug. 1996), http://www.ietf.org/rfc/rfc1984.txt; Brian E. Carpenter & Fred Baker, *IETF Policy on Wiretapping*, IETF RFC 2804 (rel. May 2000),

RFC 2804: "The IETF has decided not to consider requirements for wiretapping as part of the process for creating and maintaining IETF standards."

Just because now we're now in a position to preclude wiretaps here…



…doesn't mean that we should now start considering wiretap requirements, here

desired next steps for <this>?

where do we go from here?

- This discussion + CFA
- Refine the test until rough consensus + it is thoroughly "battle tested"
- Ship the test as an RFC to provide a standard test (first of many?) re: whether propositions break end-to-end secure messaging
- If any whole or part messenger solution fails to satisfy the test, it will be described as "not compliant with <this> RFC."
- Goal: inform user choice and assist clarity in policy discussion.

FIN

https://datatracker.ietf.org/doc/draft-muffett-end-to-end-secure-messaging/

https://github.com/alecmuffett/draft-muffett-end-to-end-secure-messaging/

alec.muffett@gmail.com || @alecmuffett