Evaluating In-Workflow Messages for Improving Mental Models of End-to-End Encryption

Omer Akgul • Wei Bai • Shruti Das • Michelle L. Mazurek
Adoption of E2EE By General Users?
Many hurdles impede adoption!
Many Hurdles Impede Adoption
Many Hurdles Impede Adoption

Icons from https://freesvg.org/vector-silhouette-of-an-athlete with modification
Many Hurdles Impede Adoption

UI Design

Key Management

Improved

Icons from https://freesvg.org/vector-silhouette-of-an-athlete with modification
Many Hurdles Impede Adoption

UI Design

Key Management

Improved

Key-Directory Based Model

Icons from https://freesvg.org/vector-silhouette-of-an-athlete with modification
Many Hurdles Impede Adoption

UI Design
Key Management
Social Norms

Improved
Key-Directory Based Model

Icons from https://freesvg.org/vector-silhouette-of-an-athlete with modification
Many Hurdles Impede Adoption

- UI Design
- Key Management
- Social Norms

- Improved
- Key-Directory Based Model
- Large Deployment

Icons from https://freesvg.org/vector-silhouette-of-an-athlete with modification
But... Mental Models still a problem!

Icons from https://freesvg.org/vector-silhouette-of-an-athlete with modification
Why do (incorrect) mental models matter?

People perceive E2EE incorrectly in both directions [1-2]:

- Encryption protects from anything
- Encryption can be trivially broken by anyone who works in IT

Why do (incorrect) mental models matter?

People perceive E2EE incorrectly in both directions [1-2]:

- Encryption protects from anything
- Encryption can be trivially broken by anyone who works in IT

Difficult for users to make thoughtful decisions:

- “SMS is the most secure messaging service.” [1]

Why do (incorrect) mental models matter?

People perceive E2EE incorrectly in both directions [1-2]:

- Encryption protects from anything
- Encryption can be trivially broken by anyone who works in IT

**Because they inhibit Confident, Proactive, and Correct usage**

Difficult for users to make thoughtful decisions:

- “SMS is the most secure messaging service.” [1]

Improve mental models Naturally

**Goal**: Help people grok basic understanding and threats

- **Enough** to make judgments about how to communicate
- **Without** turning everyone into crypto experts
- **Without** requiring people to sign up for training modules

**Solution**: Place educational messages in a messaging app, where people see them.
Multi-Stage Efforts: From Lab to Field

**Lab Study**
- In-depth tutorial
- What’s important, difficult?

**Online Survey**
- Test different messages varying in length and contents

**Field(ish) Study**
- Fit messages to an app
- Daily use for 3 weeks

Bai et al. Improving Non-Experts’ Understanding of End-to-End Encryption: An Exploratory Study. In EuroUSEC, 2020
Multi-Stage Efforts: From Lab to Field

Lab Study
- In-depth tutorial

Online Survey
- Test different messages varying in length and contents

Field(ish) Study
- Fit messages to an app
- Daily use for 3 weeks

Bai et al. Improving Non-Experts’ Understanding of End-to-End Encryption: An Exploratory Study. In EuroUSEC, 2020
Study 1 - Takeaways

- Confidentiality: Most significant

- Explaining risks clearly is useful
  - Comparing E2EE vs Non-E2EE
  - Weakness

- Some pieces may not worth mentioning
  - Integrity & authenticity
  - How E2EE works
Multi-Stage Efforts: From Lab to Field

Lab Study
- In-depth tutorial

Online Survey
- Test different messages varying in length and contents

Field(ish) Study
- Fit messages to an app
- Daily use for 3 weeks

Bai et al. Improving Non-Experts’ Understanding of End-to-End Encryption: An Exploratory Study. In EuroUSEC, 2020
Testing educational messages

- Can we *shift* user mental models with short messages in text?
  - In isolation
- How much is lost in *short, medium* vs. *long* messages?
  - Appropriate for various UIs.
- Which short, medium messages are most effective (for what)?
- Don’t want to *oversell* security
Study 2: Setup

- Online study via a crowdsourcing platform (Prolific, n=461)
- 1 Long, 5 short, 2 medium, 1 control message
  - Hypothetical app called TextLight (to remove brand bias)
- One message per participant.
## Message types

<table>
<thead>
<tr>
<th>Short</th>
<th></th>
</tr>
</thead>
</table>

| Short | (1) Nobody but you and recipient |
## Message types

<table>
<thead>
<tr>
<th>Short</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>Nobody but you and recipient</td>
</tr>
<tr>
<td></td>
<td>(2)</td>
<td>Metadata risks</td>
</tr>
</tbody>
</table>

22
# Message types

<table>
<thead>
<tr>
<th>Short</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) Nobody but you and recipient</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2) Metadata risks</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3) Endpoint risks</td>
<td></td>
</tr>
</tbody>
</table>
# Message types

<table>
<thead>
<tr>
<th>Short</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) Nobody but you and recipient</td>
</tr>
<tr>
<td></td>
<td>(2) Metadata risks</td>
</tr>
<tr>
<td></td>
<td>(3) Endpoint risks</td>
</tr>
<tr>
<td></td>
<td>(4) Lock/key metaphor</td>
</tr>
</tbody>
</table>
## Message types

<table>
<thead>
<tr>
<th>Short</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>Nobody but you and recipient</td>
</tr>
<tr>
<td>(2)</td>
<td>Metadata risks</td>
</tr>
<tr>
<td>(3)</td>
<td>Endpoint risks</td>
</tr>
<tr>
<td>(4)</td>
<td>Lock/key metaphor</td>
</tr>
<tr>
<td>(5)</td>
<td>E2EE vs. other</td>
</tr>
</tbody>
</table>
# Message types

<table>
<thead>
<tr>
<th>Short</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>Nobody but you and recipient</td>
</tr>
<tr>
<td>(2)</td>
<td>Metadata risks</td>
</tr>
<tr>
<td>(3)</td>
<td>Endpoint risks</td>
</tr>
<tr>
<td>(4)</td>
<td>Lock/key metaphor</td>
</tr>
<tr>
<td>(5)</td>
<td>E2EE vs. other</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Medium</th>
<th></th>
</tr>
</thead>
</table>
## Message types

<table>
<thead>
<tr>
<th>Short</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>Nobody but you and recipient</td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>Metadata risks</td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>Endpoint risks</td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>Lock/key metaphor</td>
<td></td>
</tr>
<tr>
<td>(5)</td>
<td>E2EE vs. other</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Medium</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Two messages with various combinations of short messages.</td>
<td></td>
</tr>
</tbody>
</table>
## Message types

| Short   | (1) Nobody but you and recipient  
|         | (2) Metadata risks  
|         | (3) Endpoint risks  
|         | (4) Lock/key metaphor  
|         | (5) E2EE vs. other  
| Medium | Two messages with various combinations of short messages.  
| Long   |
# Message types

| Short   | (1) Nobody but you and recipient  
|         | (2) Metadata risks  
|         | (3) Endpoint risks  
|         | (4) Lock/key metaphor  
|         | (5) E2EE vs. other  
| Medium  | Two messages with various combinations of short messages.  
| Long    | All key points, extra emphasis  

## Message types

| Short      | (1) Nobody but you and recipient  
|           | (2) Metadata risks  
|           | (3) Endpoint risks  
|           | (4) Lock/key metaphor  
|           | (5) E2EE vs. other  
| Medium     | Two messages with various combinations of short messages.  
| Long       | All key points, extra emphasis  
| Control    |
### Message types

<table>
<thead>
<tr>
<th>Short</th>
<th>(1) Nobody but you and recipient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(2) Metadata risks</td>
</tr>
<tr>
<td></td>
<td>(3) Endpoint risks</td>
</tr>
<tr>
<td></td>
<td>(4) Lock/key metaphor</td>
</tr>
<tr>
<td></td>
<td>(5) E2EE vs. other</td>
</tr>
</tbody>
</table>

| Medium | Two messages with various combinations of short messages. |

| Long   | All key points, extra emphasis      |

| Control | Describes non-security/privacy features |
Based on your understanding of end-to-end encryption, please indicate whether you agree or disagree that hackers who have compromised the TextLight servers have the following abilities, regardless of their motivation to do so.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neither agree nor disagree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can see that you have sent a message on TextLight, regardless of knowing the content of the message.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Can see what is in the</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
“Messages in TextLight are end-to-end encrypted. Before a message ever leaves your device, it’s secured with a lock, and only you and your recipients have the keys to open the message and read it.”
Based on your understanding of end-to-end encryption, please indicate whether you agree or disagree that hackers who have compromised the TextLight servers have the following abilities, regardless of their motivation to do so.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neither agree nor disagree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
</table>

Can see that you have sent a message on TextLight, regardless of knowing the content of the message.

Can see what is in the

Based on your understanding of end-to-end encryption, please indicate whether you agree or disagree that hackers who have compromised the TextLight servers have the following abilities, respectively.

**Metric:** Difference between the two questionnaires

- Can see what is in the message
- Can see what is in the message
- Can see what is in the message
- Can see what is in the message

Choose one for each statement.
better
better

One person
better

Median

One person

differences

control
better

Median

One person
better differences

control
Long messages work!

- Long message is generally better than control
  - Our best effort
Shorts? Also work!

- When message is topical, mostly better than all messages
Shorts? Also work!

- When message is topical, mostly better than all messages
- But, some additional risk of overselling!
Study 2: Takeaways

- The messages work! (in a controlled environment)
- Short messages work surprisingly well
  - Chance of overselling, need all for a complete mental model
Multi-Stage Efforts: From Lab to Field

Field(ish) Study
- Fit messages to an app
- Daily use for 3 weeks

Lab Study
- In-depth tutorial

Online Survey
- Test different messages varying in length and contents

Bai et al. Improving Non-Experts’ Understanding of End-to-End Encryption: An Exploratory Study. In EuroUSEC, 2020
Feeds Into Study 3

• How well would messages from study 2 work in the real world?
  ○ (integrated in an app)

• Why does it or why doesn’t it work?
  ○ How can we improve it further?
Study 3 Setup

- Incorporate successful messages from online study into an app (experimental)
Study 3 Setup

- Incorporate successful messages from online study into an app (experimental)
  - Re-brand Signal to TextLight
Study 3 Setup

- Incorporate successful messages from online study into an app (experimental)
  - Re-brand Signal to TextLight
  - Show short messages
Study 3 Setup

- Incorporate successful messages from online study into an app (experimental)
  - Re-brand Signal to TextLight
  - Show short messages
  - Clickable to open long message
Study 3 Setup

- Incorporate successful messages from online study into an app (experimental)
  - Re-brand Signal to TextLight
  - Show short messages
  - Clickable to open long message

- Control version with no messages
Based on your understanding of end-to-end encryption, please indicate whether you agree or disagree that hackers who have compromised the TextLight servers have the following abilities, regardless of their motivation to do so.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neither agree nor disagree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can see that you have sent a message on TextLight, regardless of knowing the content of the message.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Can see what is in the message</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Take privacy with you. Be yourself in every message.

Terms & Privacy Policy

CONTINUE
Give your inbox something to write home about. Get started by messaging a friend.
Short texting sessions for 20 days.
Based on your understanding of end-to-end encryption, please indicate whether you agree or disagree that hackers who have compromised the TextLight servers have the following abilities, regardless of their motivation to do so.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neither agree nor disagree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Can see that you have sent a message on TextLight, regardless of knowing the content of the message.

Can see what is in the message.
Based on your understanding of end-to-end encryption, please indicate whether you agree or disagree that hackers who have compromised the TextLight servers have the following abilities, regardles of knowing the content of the message.

Can see that you have sent a message on TextLight.

Can see what is in the message.
Study 3: Results

- Statistically, there is no improvement between experimental and control groups
  - People knew more than expected
  - In one question, we oversold E2EE (like in the survey study)

- Interviews tell us more
Interviews:

● 10/19 participants were able to generalize the concept
  ○ “[it protects from] Probably anyone who would interrupt or interfere in between the messaging, in between where you sent it and someone else received it.”

● 14/19 knew the unlocked phone adversary was powerful

● 9/19 got at least something wrong
  ○ “[it protects from] people … hacking into your phone … from either reading the messages or altering the contents of the message.”

● 9/19 said they didn’t read messages or weren’t interested in them.
  ○ “I obviously didn't pay a lot of attention to it.”
Summary

● Mental models of secure communication: not *functional* enough
● Can **small nudges** and user-centered design improve things?
  ○ Initial qualitative study to identify topics, messages
  ○ Online study to examine specific messages
  ○ Longitudinal study to measure real-world effectiveness

● They work well when we **control external factors**.
● **Integration** to applications might need to be more obvious.
  ○ Perhaps by sacrificing usability a little bit.
Summary

• Mental models of secure communication: not functional enough

• Can small nudges and user-centered design improve things?
  ○ Initial qualitative study to identify topics, messages
  ○ Online study to examine specific messages
  ○ Longitudinal study to measure real-world effectiveness

They work well when we control external factors.

Integration to applications might need to be more obvious.
  ○ Perhaps by sacrificing usability a little bit.
References


References

References


References