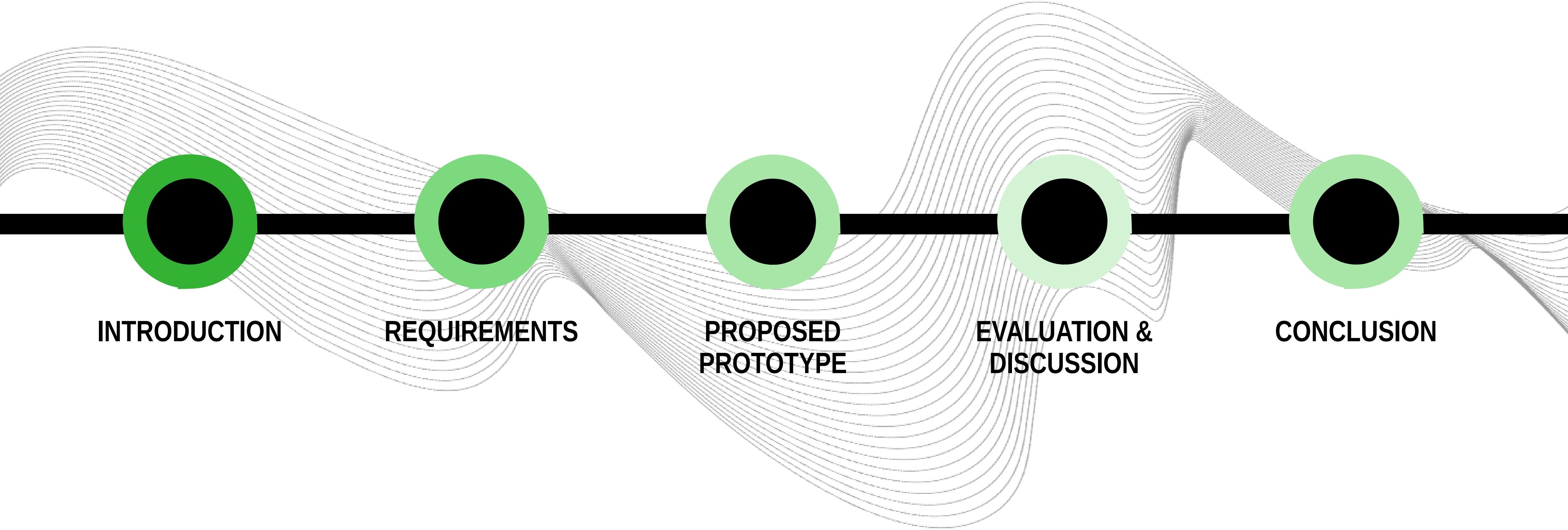
UNIVERSITY
OF TWENTE.

DIDACTIC VISUALIZATION FOR A SEARCHABLE ENCRYPTION SCHEME

JUL 2, 2021

RUILIN YANG, S2099497

IN THIS PRESENTATION:



INTRODUCTION

REQUIREMENTS

PROPOSED PROTOTYPE

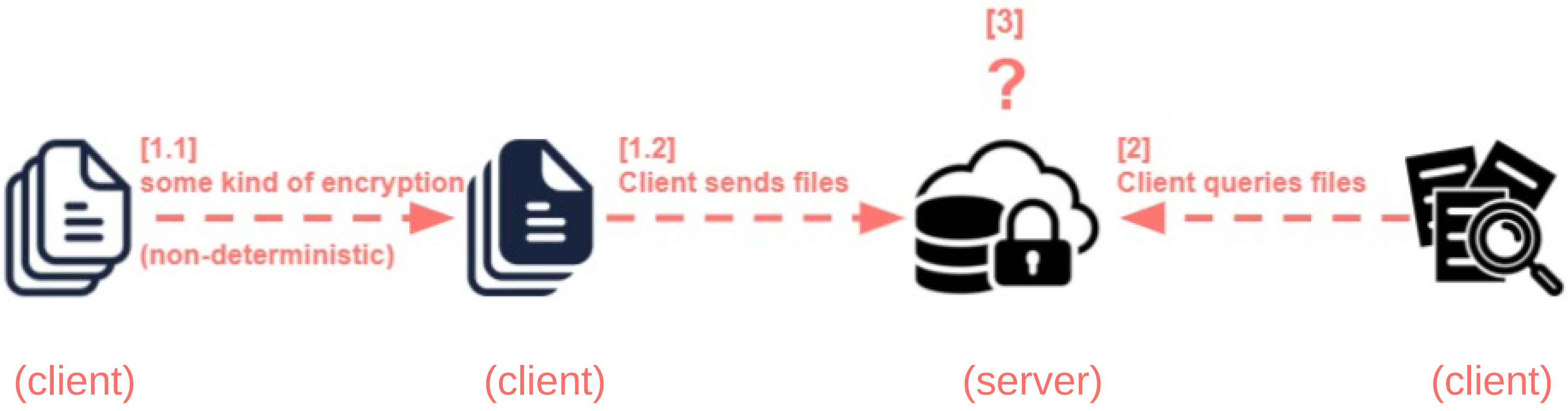
EVALUATION & DISCUSSION

CONCLUSION

1 INTRODUCTION

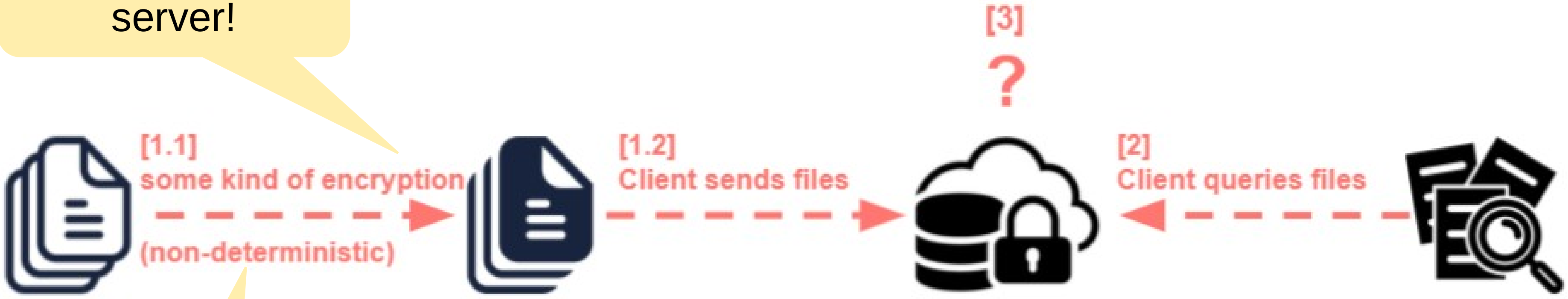
INTRO 1 – WHY SEARCHABLE ENCRYPTION?

- Cloud storage ↑



INTRO 1 – WHY SEARCHABLE ENCRYPTION?

Don’t trust the server!

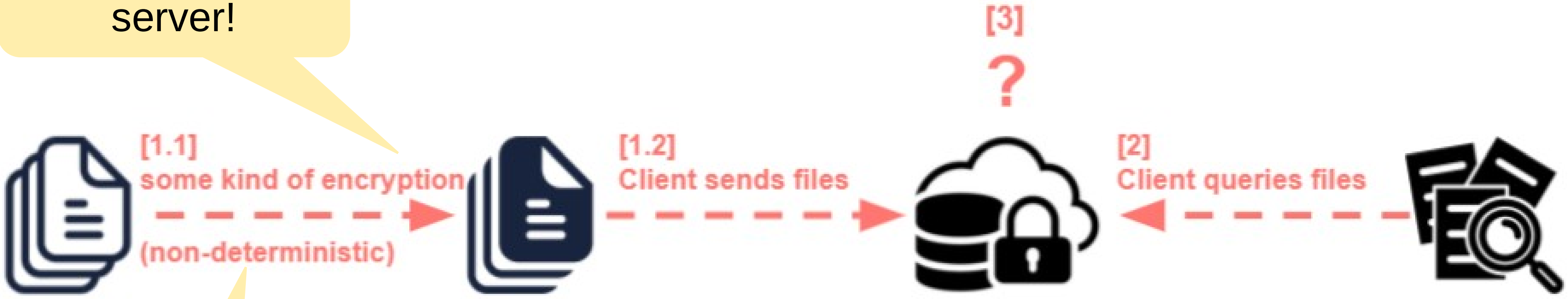


Eg. “cat” could be encrypted to “d74a44”, or “49a739”, or something else, you never know.

Hides the pattern in data.

INTRO 1 – WHY SEARCHABLE ENCRYPTION?

Don’t trust the server!



[1.1]
some kind of encryption
(non-deterministic)

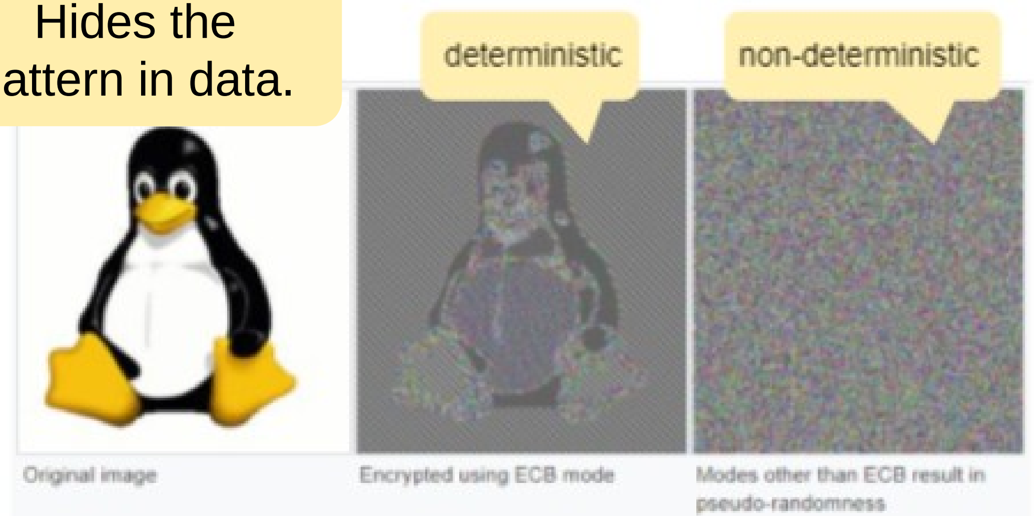
[1.2]
Client sends files

[3]
?

[2]
Client queries files

Eg. “cat” could be encrypted to “d74a44”, or “49a739”, or something else, you never know.

Hides the pattern in data.



Original image

Encrypted using ECB mode

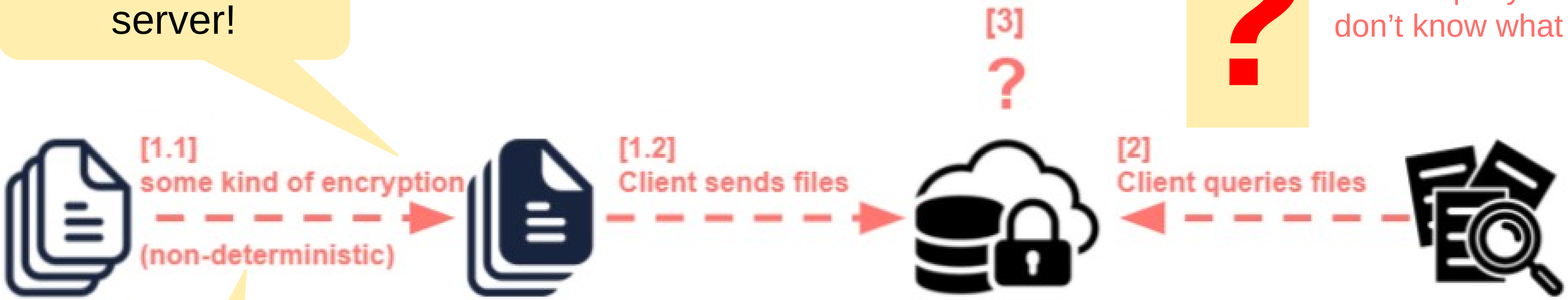
Modes other than ECB result in pseudo-randomness

INTRO 1 – WHY SEARCHABLE ENCRYPTION?

Don’t trust the server!

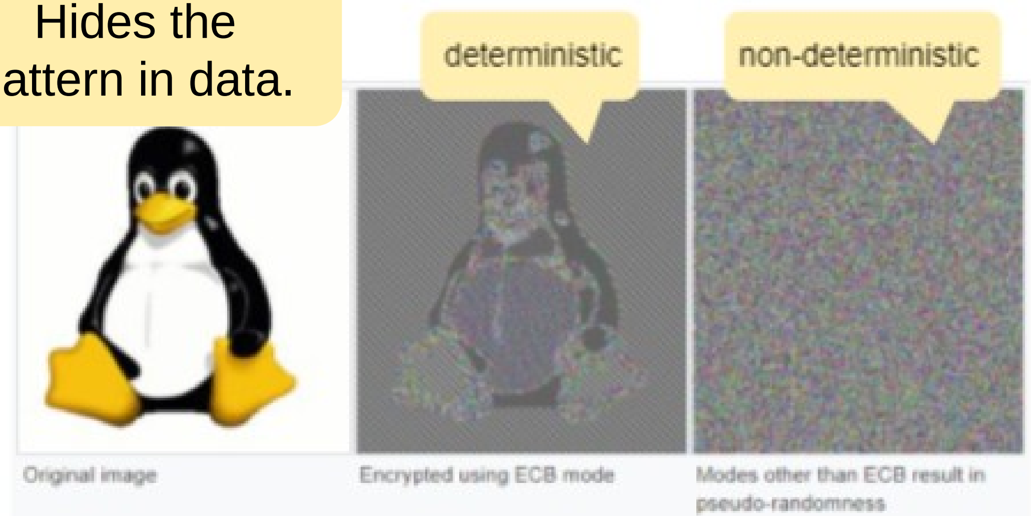
?

How to query when you don’t know what to query?



Eg. “cat” could be encrypted to “d74a44”, or “49a739”, or something else, you never know.

Hides the pattern in data.



Original image

Encrypted using ECB mode

Modes other than ECB result in pseudo-randomness

INTRO 1 – WHY SEARCHABLE ENCRYPTION?

- Searchable Encryption is exactly for this problem
  - But it’s complex to learn..

INTRO 2 – THIS RESEARCH

- Searchable Encryption is exactly for this problem
  - But it’s complex to learn..
- Visualization → aid education.
  - Existing cryptography edu software: many not publicly available
  - CrypTool 2, JCrypTool: no built-in SE visualization

INTRO 2 – THIS RESEARCH

- Searchable Encryption is exactly for this problem
  - But it’s complex to learn..
- Visualization → aid education.
  - Existing cryptography edu software: many not publicly available
  - CrypTool 2, JCrypTool: no built-in SE visualization
- RQ1
  - Which Searchable Encryption scheme(s) to design visualization for?

    (among many)
- RQ2
  - How to design and implement the scheme to help novice learner learn?

    (Requirements? Prototype? Evaluation?)

2 REQUIREMENTS

REQUIREMENTS

- High-level abstraction to make the first encounter easier.
  - Mathematically heavy knowledge background not for now
  - A large number of terminologies not for now

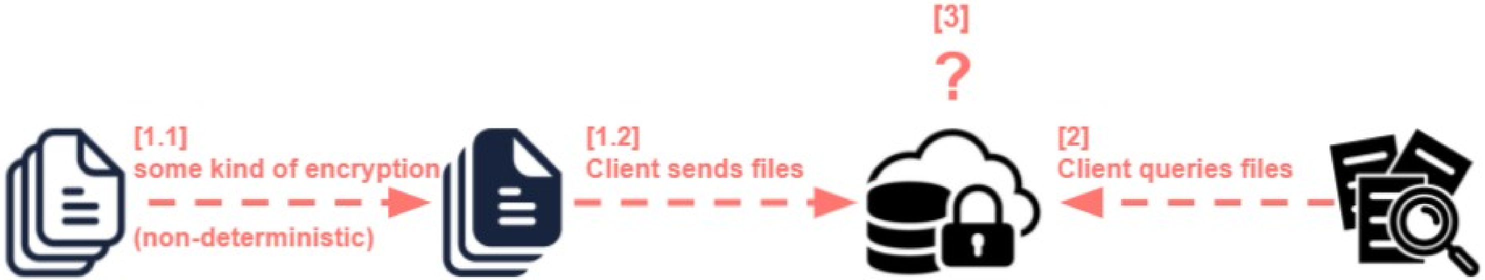
REQUIREMENTS

- High-level abstraction to make the first encounter easier.
  - Mathematically heavy knowledge background not for now
  - A large number of terminologies not for now
- Didactic Design principles
  - *Conciseness*
  - *Autonomy*
  - *Structure*
  - *Quality*
  - *Phasing*
  - *Simplicity and accessibility*

3 THE PROPOSED PROTOTYPE

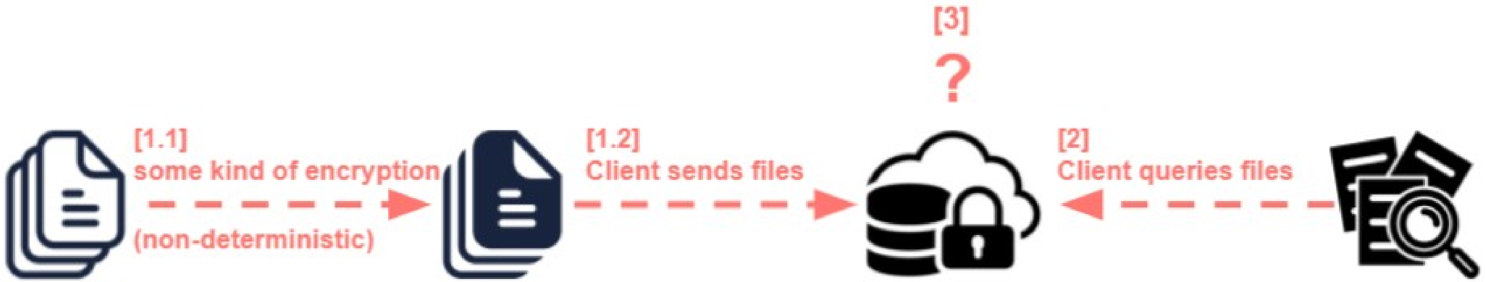
THE CHOSEN SCHEME

- The very first Searchable Encryption scheme.
  - D. X. Song, D. Wagner, and A. Perrig. Practical techniques for searches on encrypted data. (2000)



THE CHOSEN SCHEME

- The very first Searchable Encryption scheme.
  - D. X. Song, D. Wagner, and A. Perrig. Practical techniques for searches on encrypted data. (2000)
- Relies on common building blocks → help students review.
  - A pseudorandom generator *G*
  - Two pseudorandom function *F* and *f*
  - A pseudorandom permutation *E*



[3]

?

[1.1]

some kind of encryption

(non-deterministic)

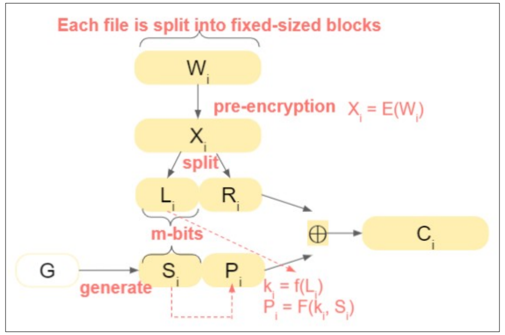
[1.2]

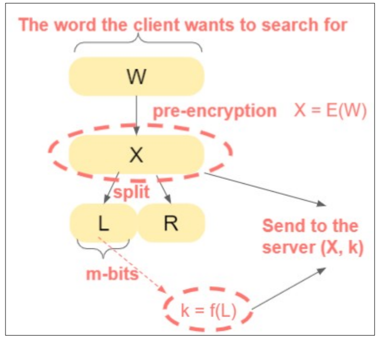
Client sends files

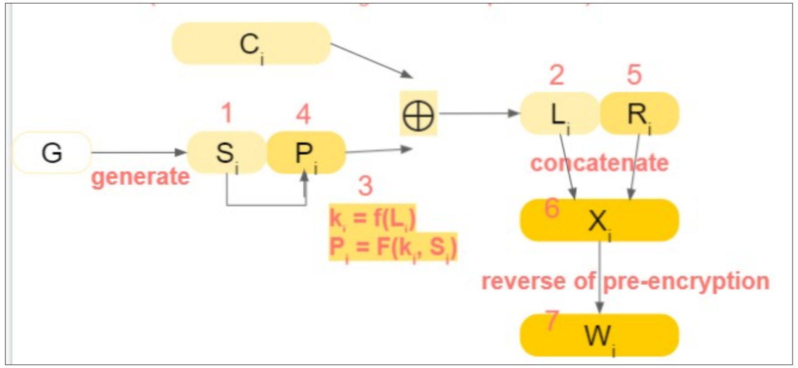
[2]

Client queries files

The chosen scheme




The word the client wants to search for
W
pre-encryption $X = E(W)$
X
split
L
R
m-bits
$k = f(L)$
Send to the server (X, k)



Decrypt files

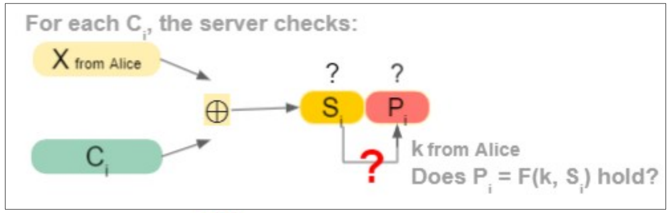
Alice (client)

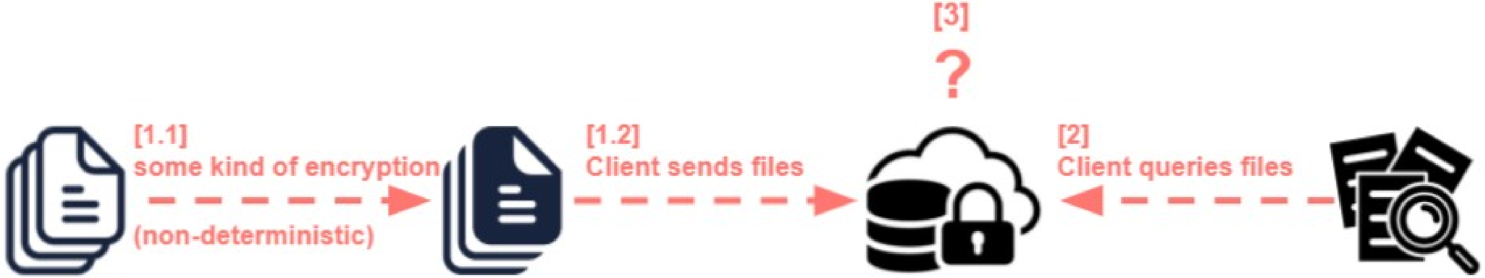
Bob (server)

Send file

Send query terms

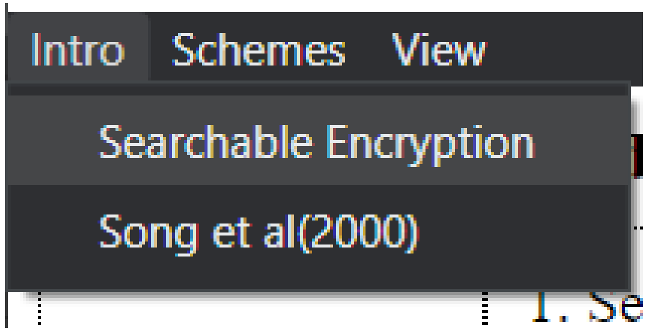
Return files


For each C_i, the server checks:
X from Alice
⊕
C_i
?
?
S_i
P_i
?
k from Alice
Does $P_i = F(k, S_i)$ hold?

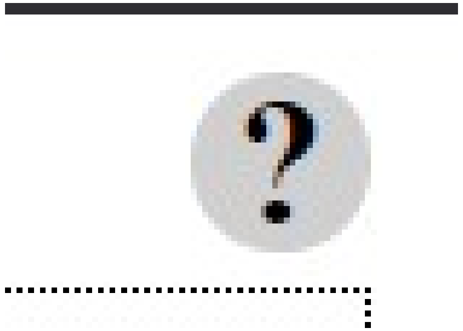


THE PROPOSED PROTOTYPE

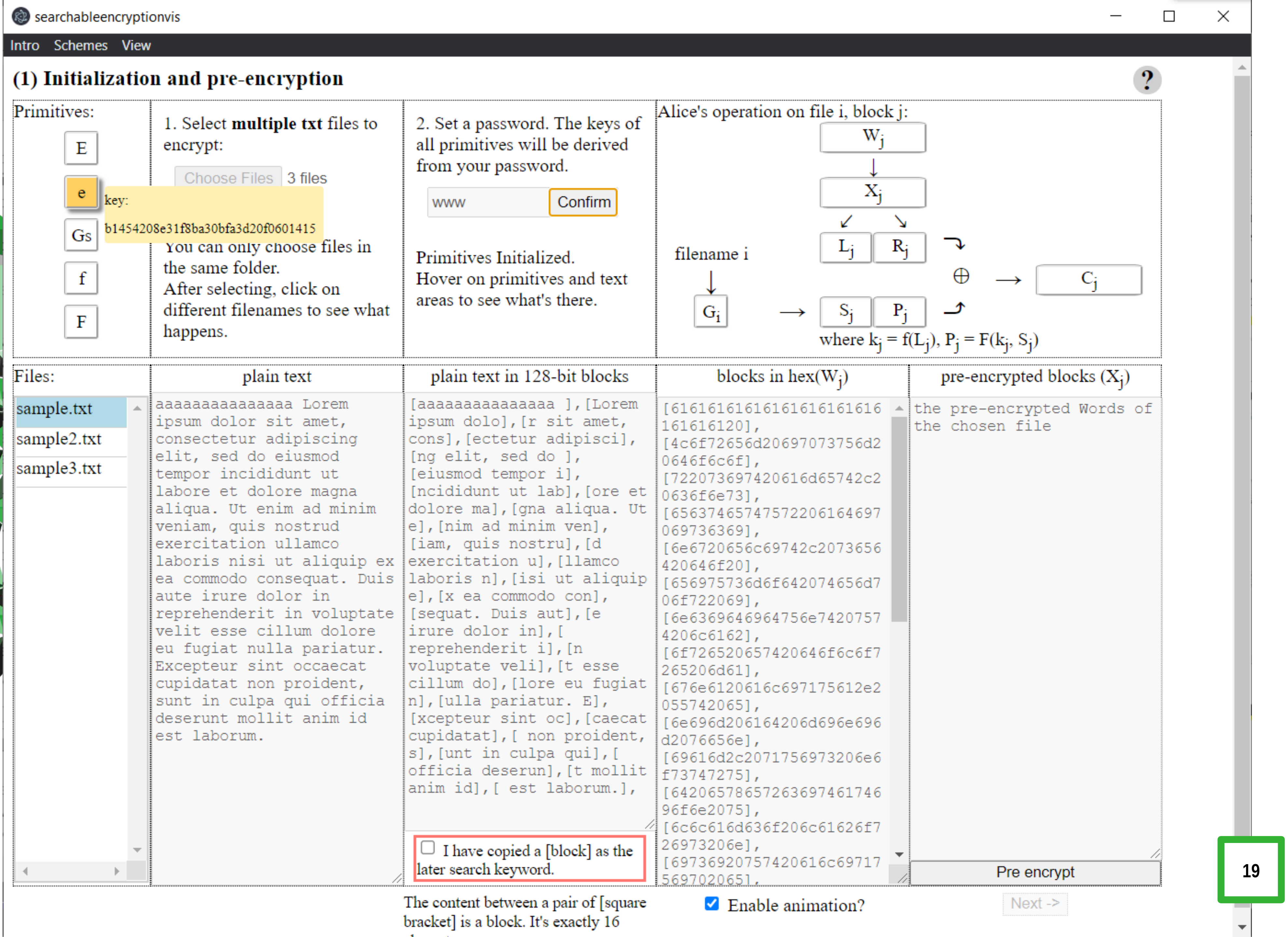
- Built-in help:



Embedded introductory slides
educational



Practical help messages of each page
practical



(1) Initialization and pre-encryption

Primitives:

key:

b1454208e31f8ba30bfa3d20f0601415

1. Select **multiple txt** files to encrypt:

3 files

You can only choose files in the same folder.
After selecting, click on different filenames to see what happens.

2. Set a password. The keys of all primitives will be derived from your password.

www

Primitives Initialized.
Hover on primitives and text areas to see what's there.

Alice's operation on file i, block j:

where $k_j = f(L_j)$, $P_j = F(k_j, S_j)$

Files:

sample.txt
sample2.txt
sample3.txt

plain text

```
aaaaaaaaaaaaaaa Lorem
ipsum dolor sit amet,
consectetur adipiscing
elit, sed do eiusmod
tempor incididunt ut
labore et dolore magna
aliqua. Ut enim ad minim
veniam, quis nostrud
exercitation ullamco
laboris nisi ut aliquip ex
ea commodo consequat. Duis
aute irure dolor in
reprehenderit in voluptate
velit esse cillum dolore
eu fugiat nulla pariatur.
Excepteur sint occaecat
cupidatat non proident,
sunt in culpa qui officia
deserunt mollit anim id
est laborum.
```

plain text in 128-bit blocks

```
[aaaaaaaaaaaaaaa ],[Lorem
ipsum dolo],[r sit amet,
cons],[ectetur adipisci],
[ng elit, sed do ],
[eiusmod tempor i],
[ncididunt ut lab],[ore et
dolore ma],[gna aliqua. Ut
e],[nim ad minim ven],
[iam, quis nostru],[d
exercitation u],[llamco
laboris n],[isi ut aliquip
e],[x ea commodo con],
[sequat. Duis aut],[e
irure dolor in],[
reprehenderit i],[n
voluptate veli],[t esse
cillum do],[lore eu fugiat
n],[ulla pariatur. E],
[xcepteur sint oc],[caecat
cupidatat],[ non proident,
s],[unt in culpa qui],[
officia deserun],[t mollit
anim id],[ est laborum.],
```

☐ I have copied a [block] as the later search keyword.

The content between a pair of [square bracket] is a block. It's exactly 16

blocks in hex(W_j)

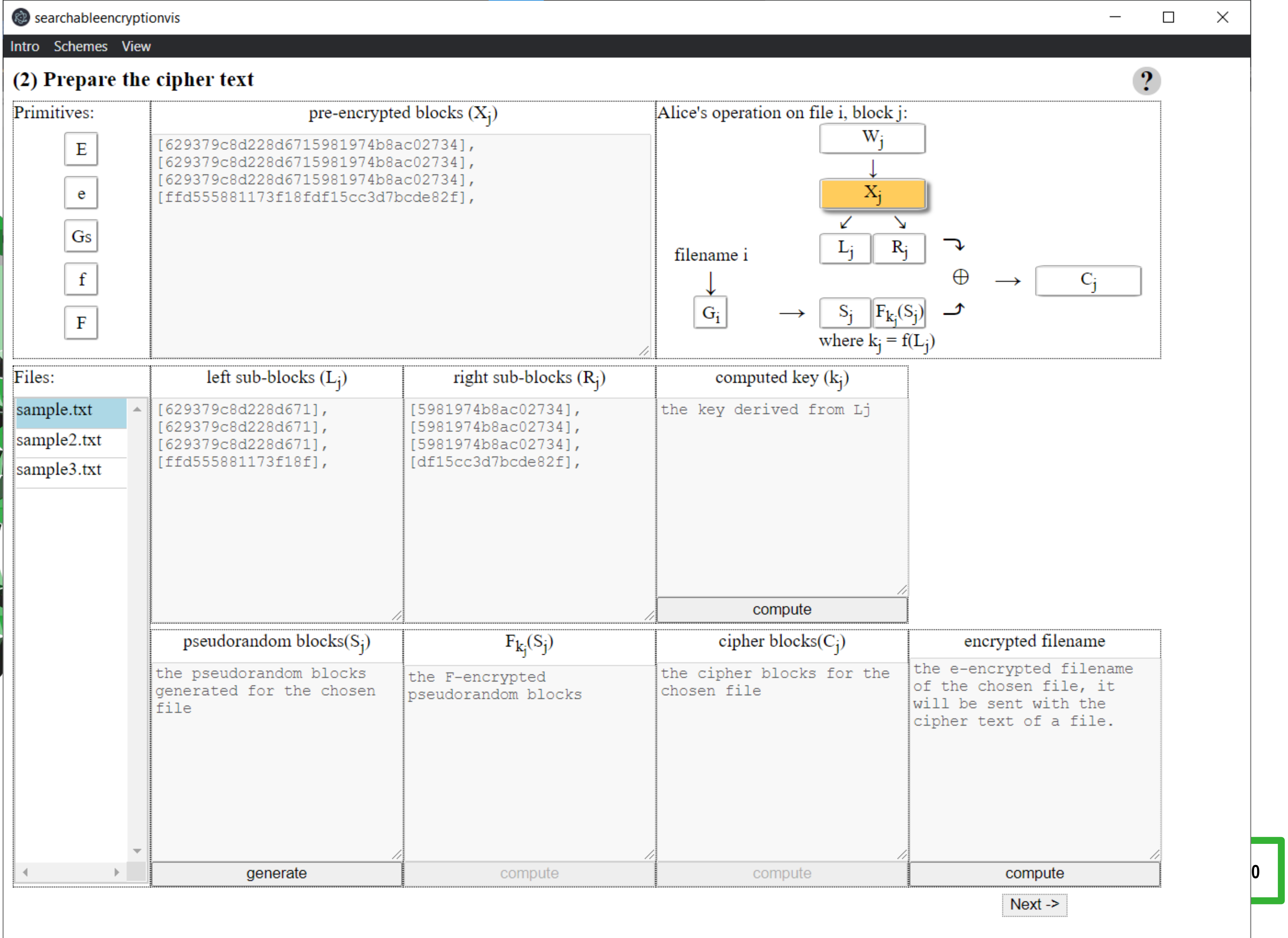
```
[6161616161616161616161616
16161616120],
[4c6f72656d20697073756d2
0646f6c6f],
[72207369742061d65742c2
0636f6e73],
[65637465747572206164697
069736369],
[6e6720656c69742c2073656
420646f20],
[656975736d6f642074656d7
06f722069],
[6e6369646964756e7420757
4206c6162],
[6f72652065742064f6c6f7
265206d61],
[676e6120616c697175612e2
055742065],
[6e696d20616420d696e696
d2076656e],
[69616d2c2071756973206e6
f73747275],
[64206578657263697461746
96f6e2075],
[6c6c616d636f206c61626f7
26973206e],
[69736920757420616c69717
569702065],
```

pre-encrypted blocks (X_j)

```
the pre-encrypted Words of
the chosen file
```

☑ Enable animation?

- Conciseness
- Autonomy
- Simplicity & accessibility



(2) Prepare the cipher text

Primitives:

pre-encrypted blocks (X_j)

```
[629379c8d228d6715981974b8ac02734],
[629379c8d228d6715981974b8ac02734],
[629379c8d228d6715981974b8ac02734],
[ffd555881173f18fdf15cc3d7bcde82f],
```

Alice's operation on file i, block j:

where $k_j = f(L_j)$

Files:

sample.txt
sample2.txt
sample3.txt

left sub-blocks (L_j)

```
[629379c8d228d671],
[629379c8d228d671],
[629379c8d228d671],
[ffd555881173f18f],
```

right sub-blocks (R_j)

```
[5981974b8ac02734],
[5981974b8ac02734],
[5981974b8ac02734],
[df15cc3d7bcde82f],
```

computed key (k_j)

the key derived from Lj

pseudorandom blocks(S_j)

the pseudorandom blocks generated for the chosen file

$F_{k_j}(S_j)$

the F-encrypted pseudorandom blocks

cipher blocks(C_j)

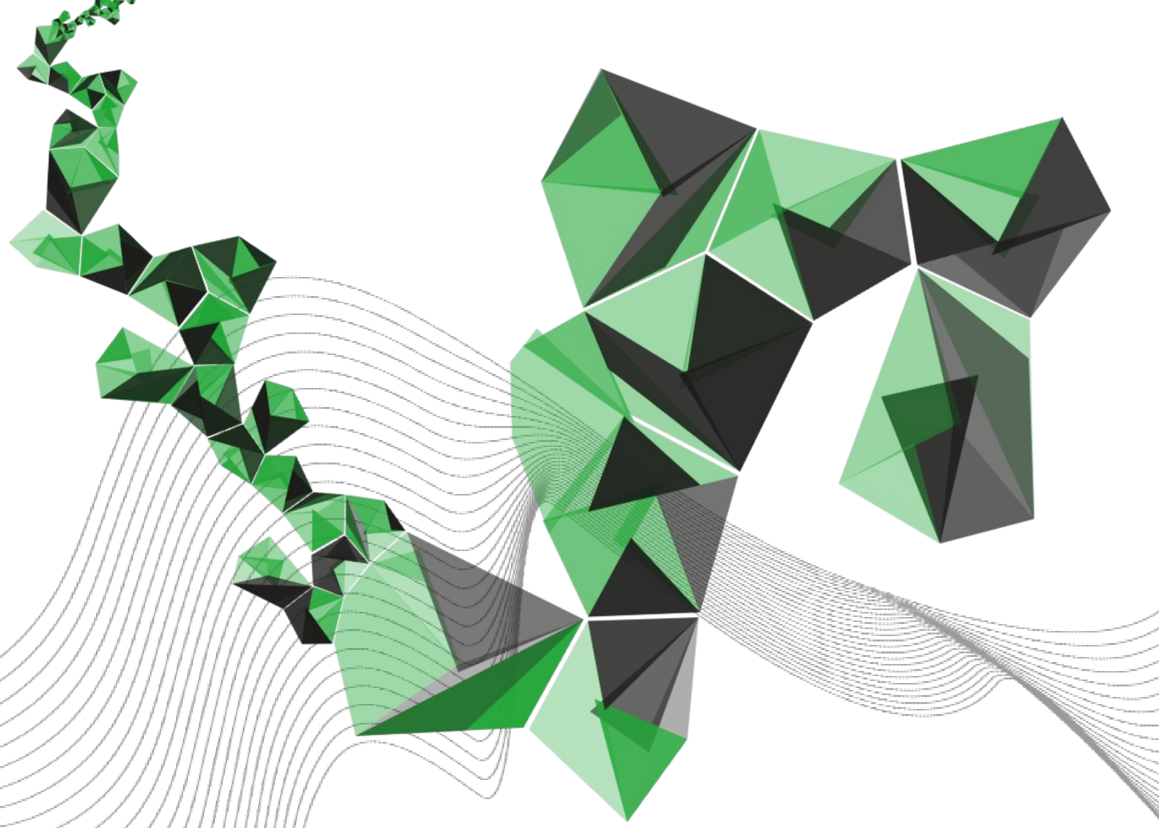
the cipher blocks for the chosen file

encrypted filename

the e-encrypted filename of the chosen file, it will be sent with the cipher text of a file.

- Structure & Quality
- Phasing

4 EVALUATION



EVALUATION

- Demography
  - 5 participants
  - Undergraduates in Computer Science (45 – 165EC)
  - Haven’t have dedicated cryptography courses yet

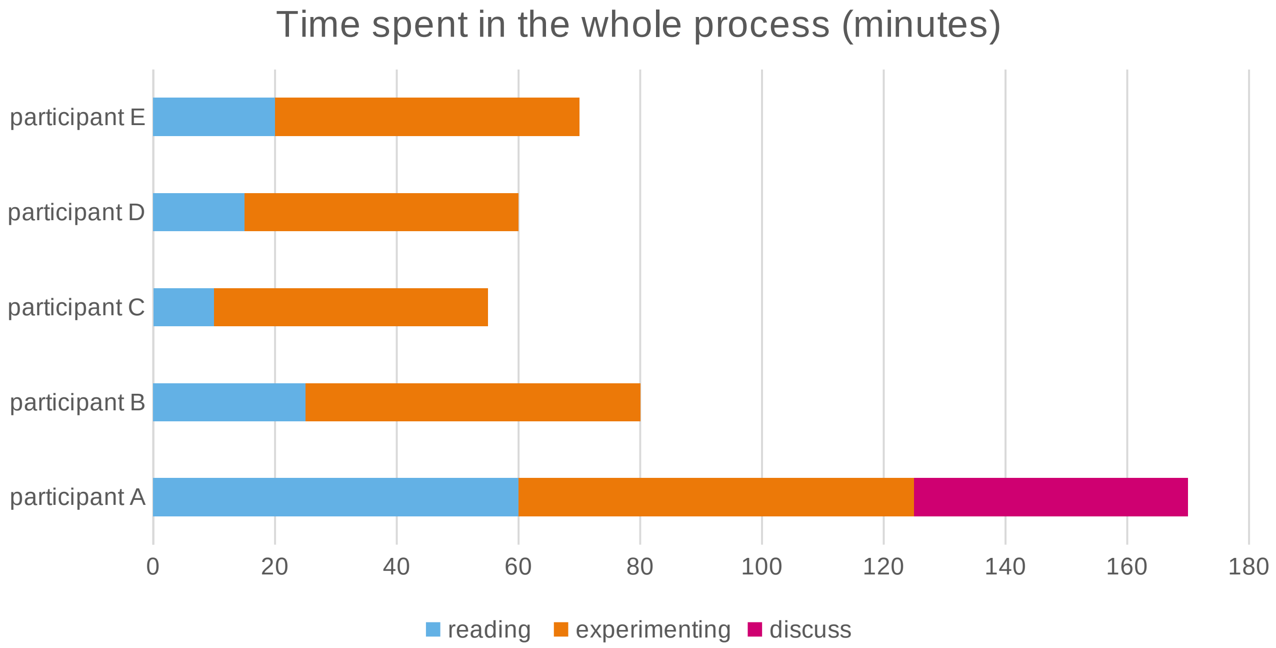
EVALUATION

- Demography
  - 5 participants
  - Undergraduates in Computer Science (45 – 165EC)
  - Haven’t have dedicated cryptography courses yet
- Process
  1. Download prototype; read instructions.
  2. Start the prototype: first study the introductory slides, then do experiment.
  3. Answer 7 questions about experience.

👀 (under observation)

EVALUATION RESULTS

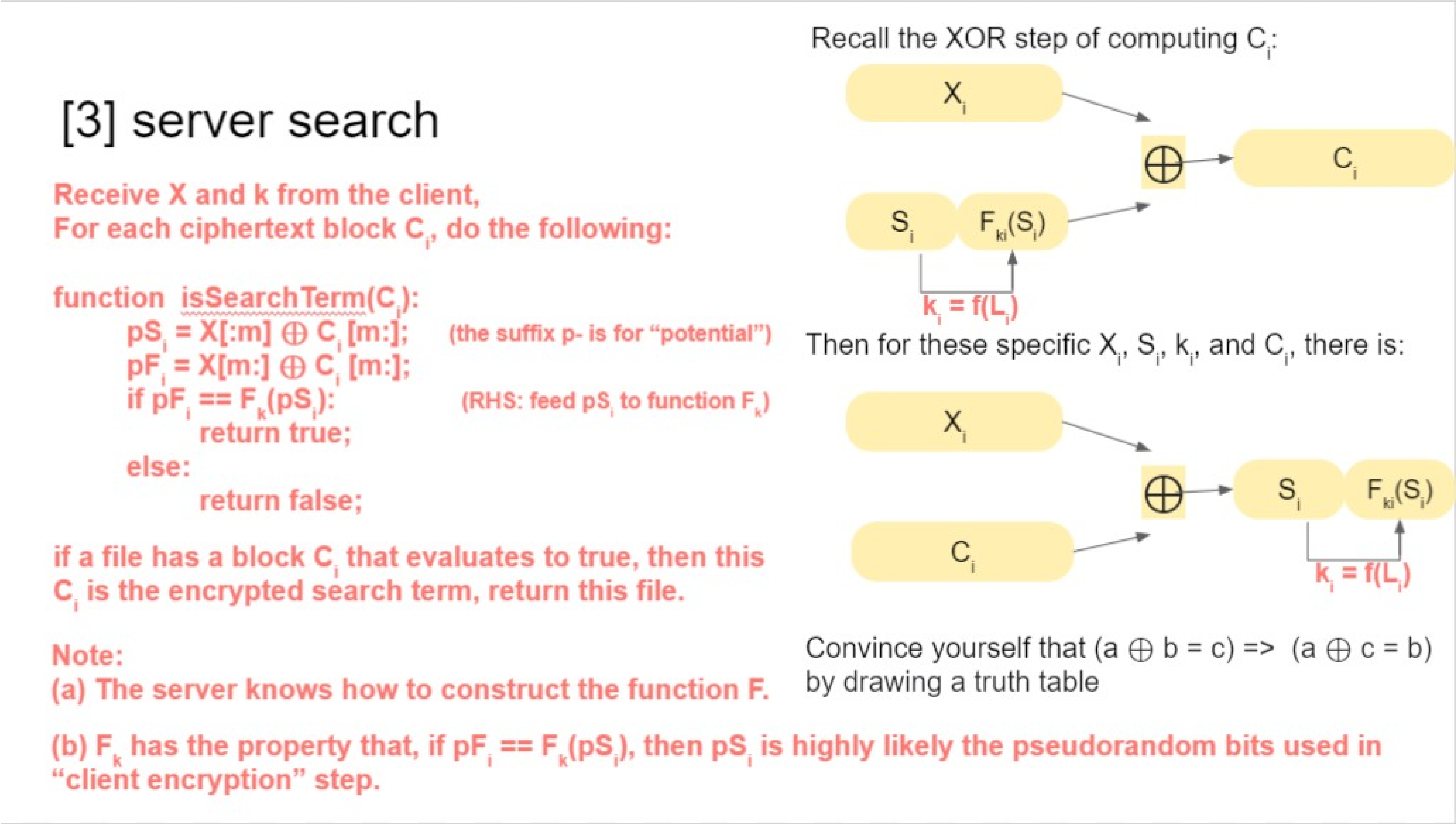
- Participant A, B, C, D, E: named after the order they took the test


Time spent in the whole process (minutes)

Expected: 40 – 60 minutes

- Textual → visual
- Hint that not every bits of information is needed before the first experiment

EVALUATION RESULTS



[3] server search

Receive X and k from the client,
For each ciphertext block C_i, do the following:

```
function isSearchTerm($C_i$):
    $pS_i = X[:m] \oplus C_i[m:]$;        (the suffix p- is for “potential”)
    $pF_i = X[m:] \oplus C_i[m:]$;
    if $pF_i == F_k(pS_i)$:                 (RHS: feed $pS_i$ to function $F_k$)
        return true;
    else:
        return false;
```

if a file has a block C_i that evaluates to true, then this C_i is the encrypted search term, return this file.

Note:
(a) The server knows how to construct the function F.

(b) F_k has the property that, if $pF_i == F_k(pS_i)$, then pS_i is highly likely the pseudorandom bits used in “client encryption” step.

Recall the XOR step of computing C_i:

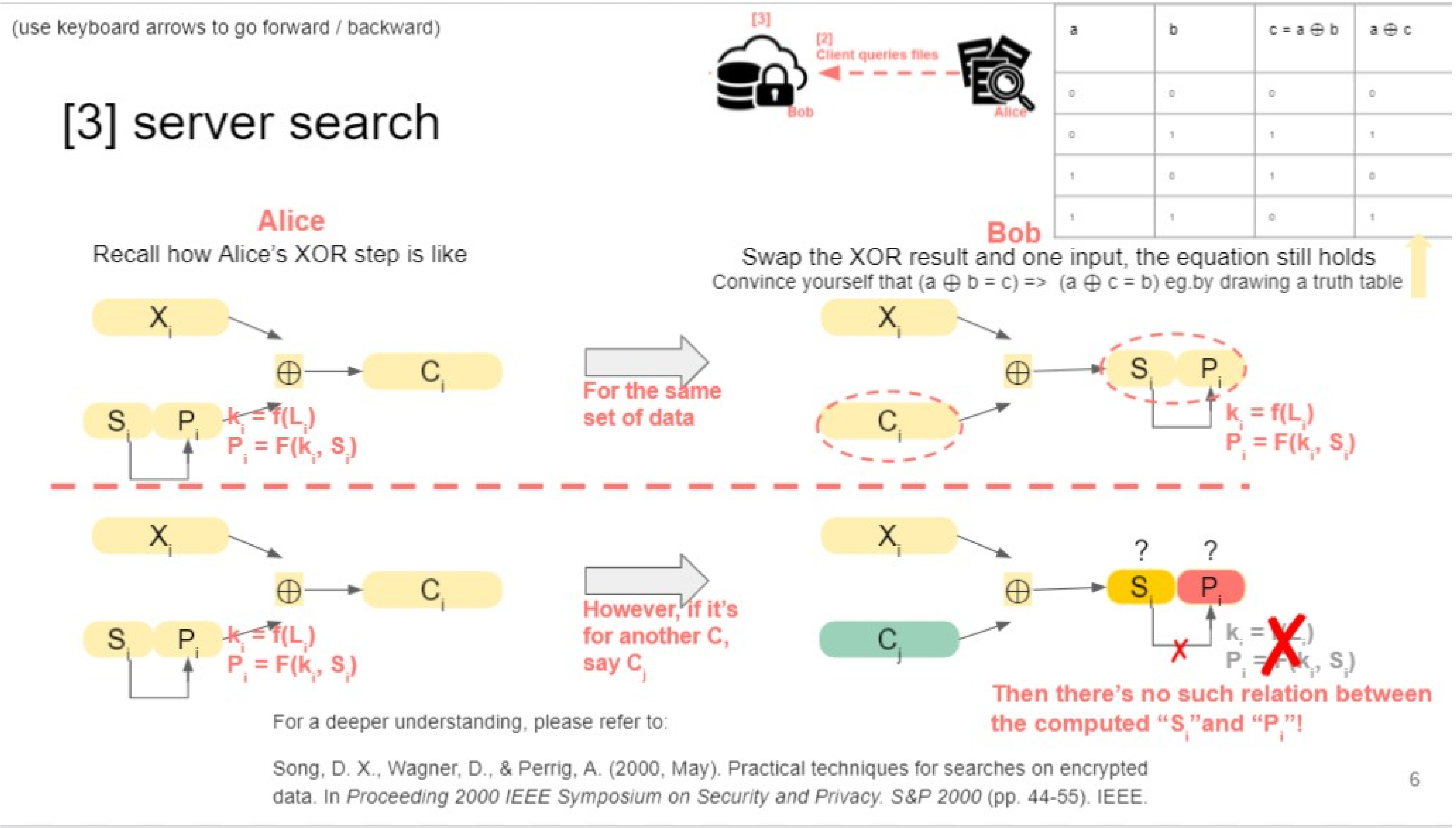
Then for these specific X_i, S_i, k_i, and C_i, there is:

Convince yourself that ($a \oplus b = c$) => ($a \oplus c = b$) by drawing a truth table

The most confusing slide

EVALUATION RESULTS

- Textual → visual
- Hint that not every bits of information is needed before the first experiment



(use keyboard arrows to go forward / backward)

[3] server search

a	b	c = a ⊕ b	a ⊕ c
0	0	0	0
0	1	1	1
1	0	1	0
1	1	0	1

Alice

Recall how Alice's XOR step is like

Bob

Swap the XOR result and one input, the equation still holds

Convince yourself that (a ⊕ b = c) => (a ⊕ c = b) eg.by drawing a truth table

For the same set of data

However, if it's for another C_i, say C_j

Then there's no such relation between the computed "S_i"and "P_i"!

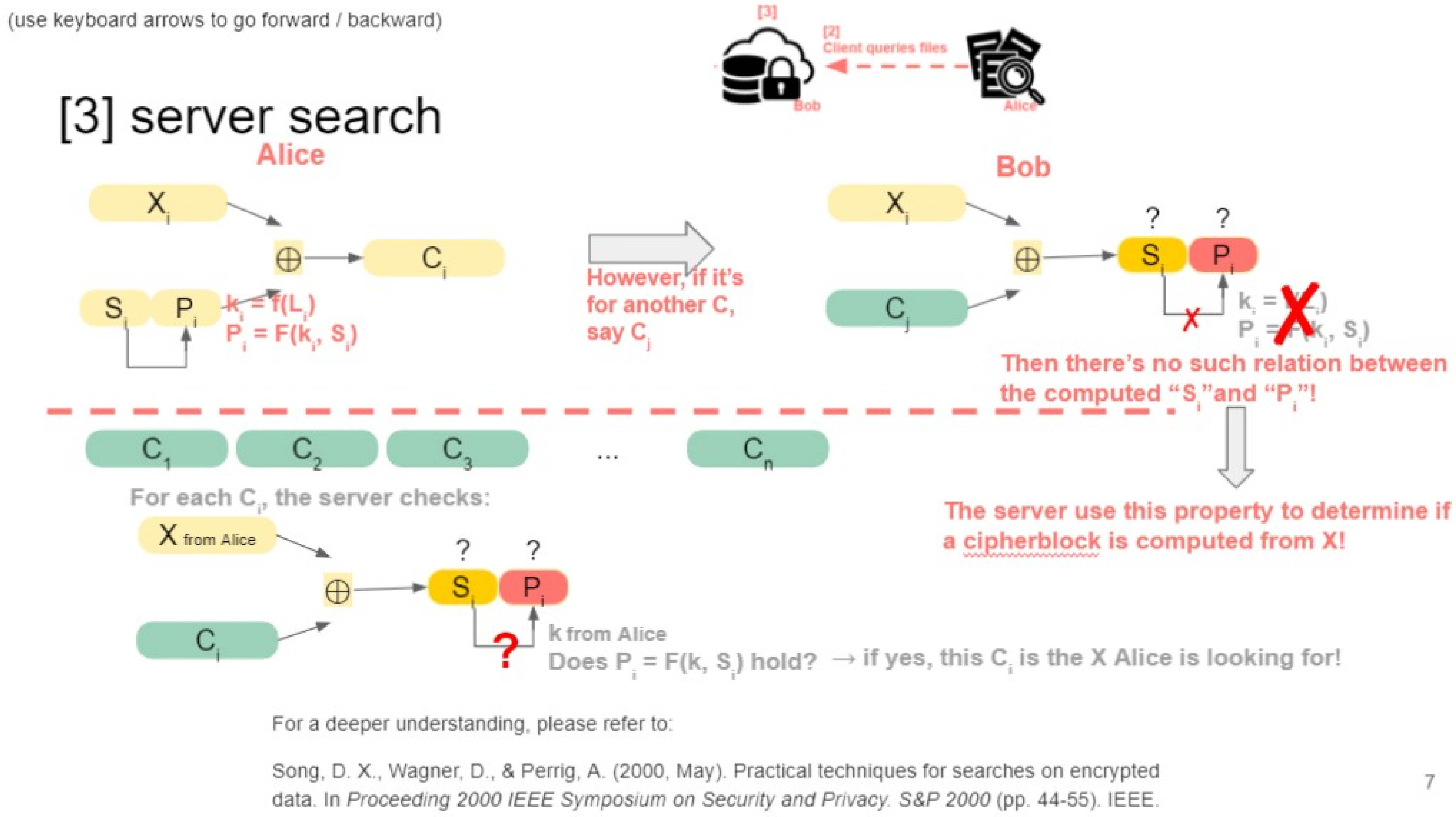
For a deeper understanding, please refer to:

Song, D. X., Wagner, D., & Perrig, A. (2000, May). Practical techniques for searches on encrypted data. In *Proceeding 2000 IEEE Symposium on Security and Privacy. S&P 2000* (pp. 44-55). IEEE.

Improved slide - 1

EVALUATION RESULTS

- Textual → visual
- Hint that not every bits of information is needed before the first experiment



(use keyboard arrows to go forward / backward)

[3] server search

However, if it's for another C, say C_j

Then there's no such relation between the computed "S_i"and "P_i"!

The server use this property to determine if a cipherblock is computed from X!

For each C_i, the server checks:

Does $P_i = F(k, S_i)$ hold? → if yes, this C_i is the X Alice is looking for!

For a deeper understanding, please refer to:

Song, D. X., Wagner, D., & Perrig, A. (2000, May). Practical techniques for searches on encrypted data. In *Proceeding 2000 IEEE Symposium on Security and Privacy. S&P 2000* (pp. 44-55). IEEE.

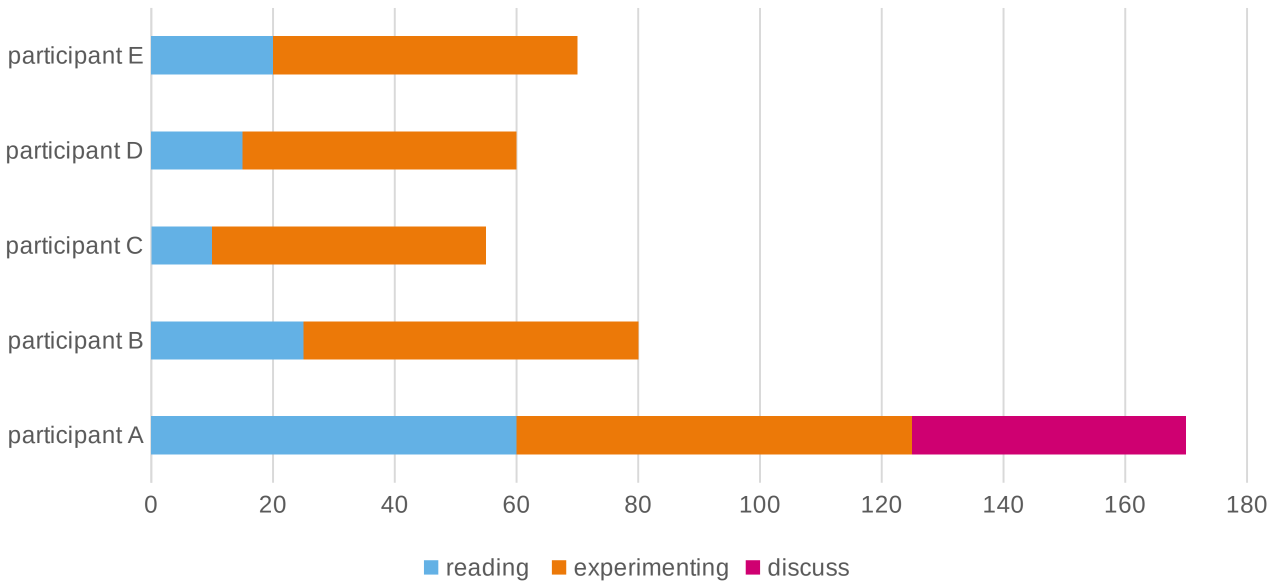
7

Improved slide - 2

EVALUATION RESULTS

- Participant A, B, C, D, E: named after the order they took the test

Time spent in the whole process (minutes)



Expected: 40 – 60 minutes

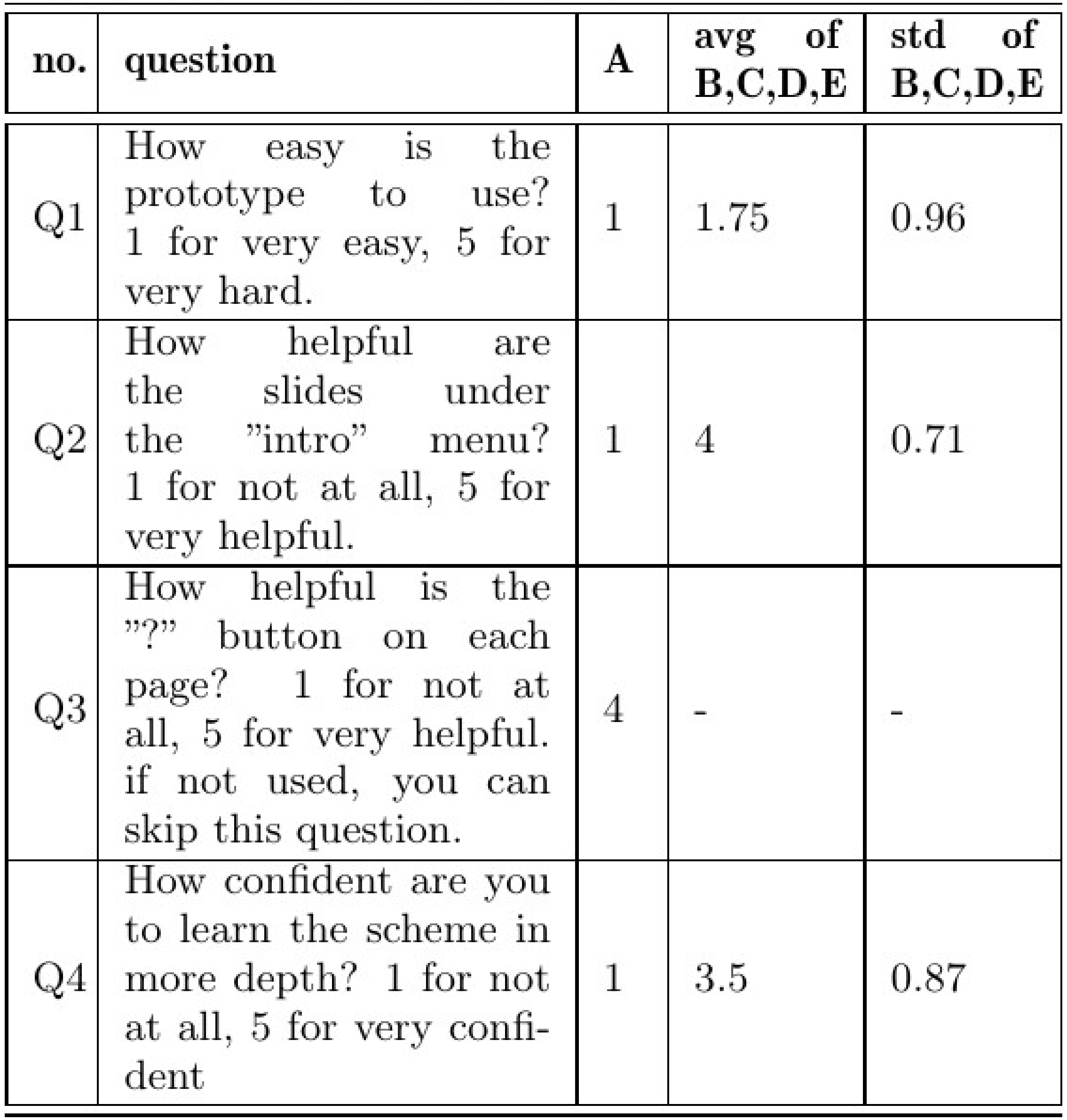
EVALUATION RESULTS & DISCUSSION

(usable)

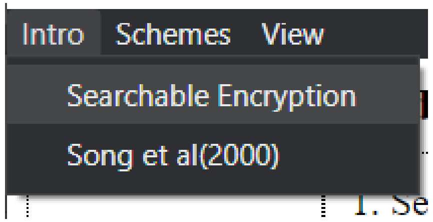
(educational)

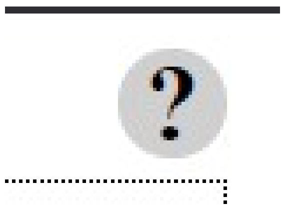
(usable)

(educational)



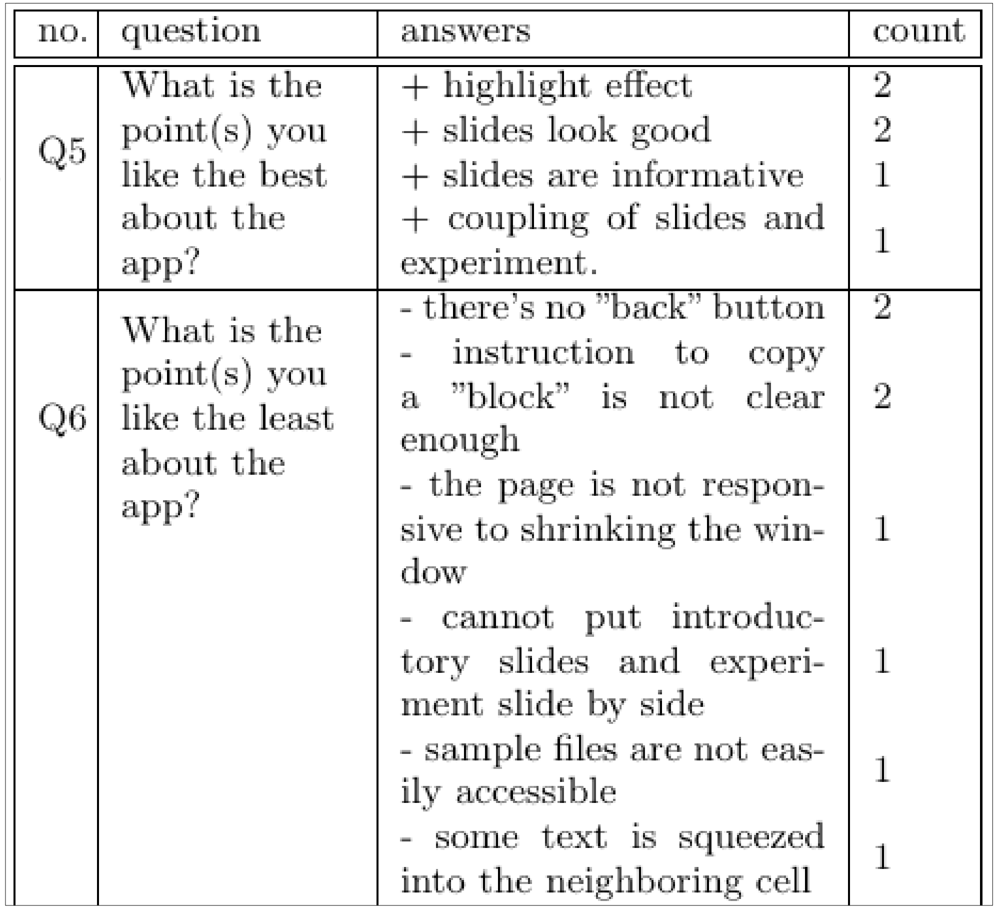
no.	question	A	avg of B,C,D,E	std of B,C,D,E
Q1	How easy is the prototype to use? 1 for very easy, 5 for very hard.	1	1.75	0.96
Q2	How helpful are the slides under the ”intro” menu? 1 for not at all, 5 for very helpful.	1	4	0.71
Q3	How helpful is the ”?” button on each page? 1 for not at all, 5 for very helpful. if not used, you can skip this question.	4	-	-
Q4	How confident are you to learn the scheme in more depth? 1 for not at all, 5 for very confident	1	3.5	0.87



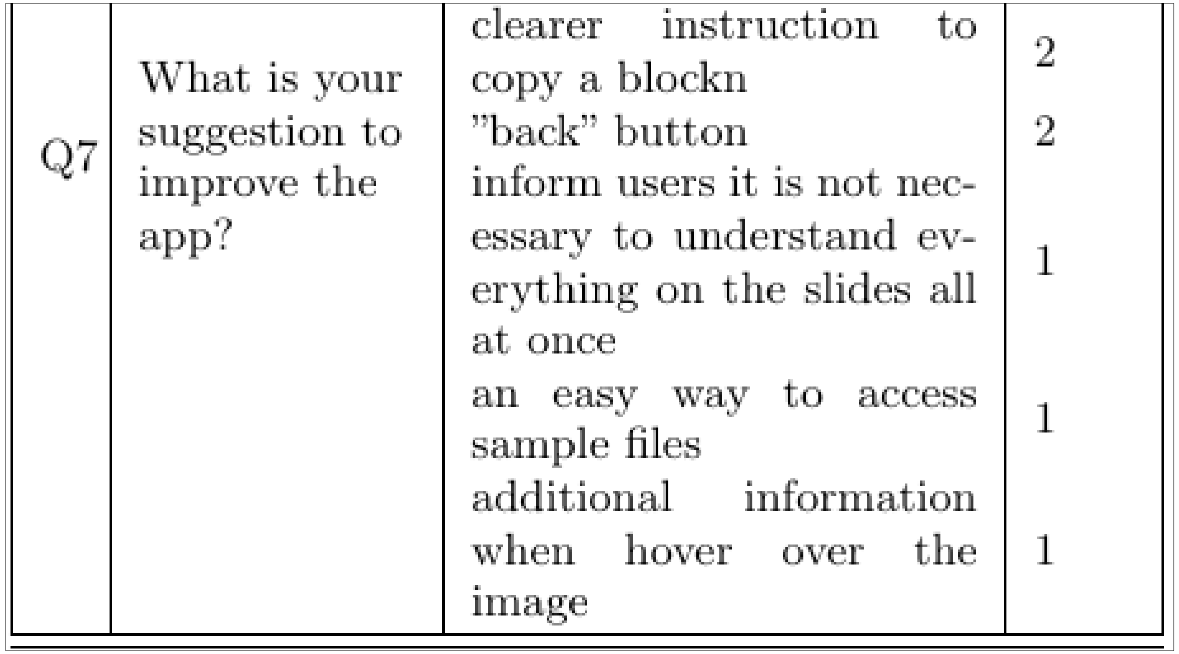


UNIVERSITY OF TWENTE.

EVALUATION RESULTS & DISCUSSION



no.	question	answers	count
Q5	What is the point(s) you like the best about the app?	+ highlight effect	2
		+ slides look good	2
		+ slides are informative	1
		+ coupling of slides and experiment.	1
Q6	What is the point(s) you like the least about the app?	- there's no ”back” button	2
		- instruction to copy a ”block” is not clear enough	2
		- the page is not responsive to shrinking the window	1
		- cannot put introductory slides and experiment slide by side	1
		- sample files are not easily accessible	1
		- some text is squeezed into the neighboring cell	1

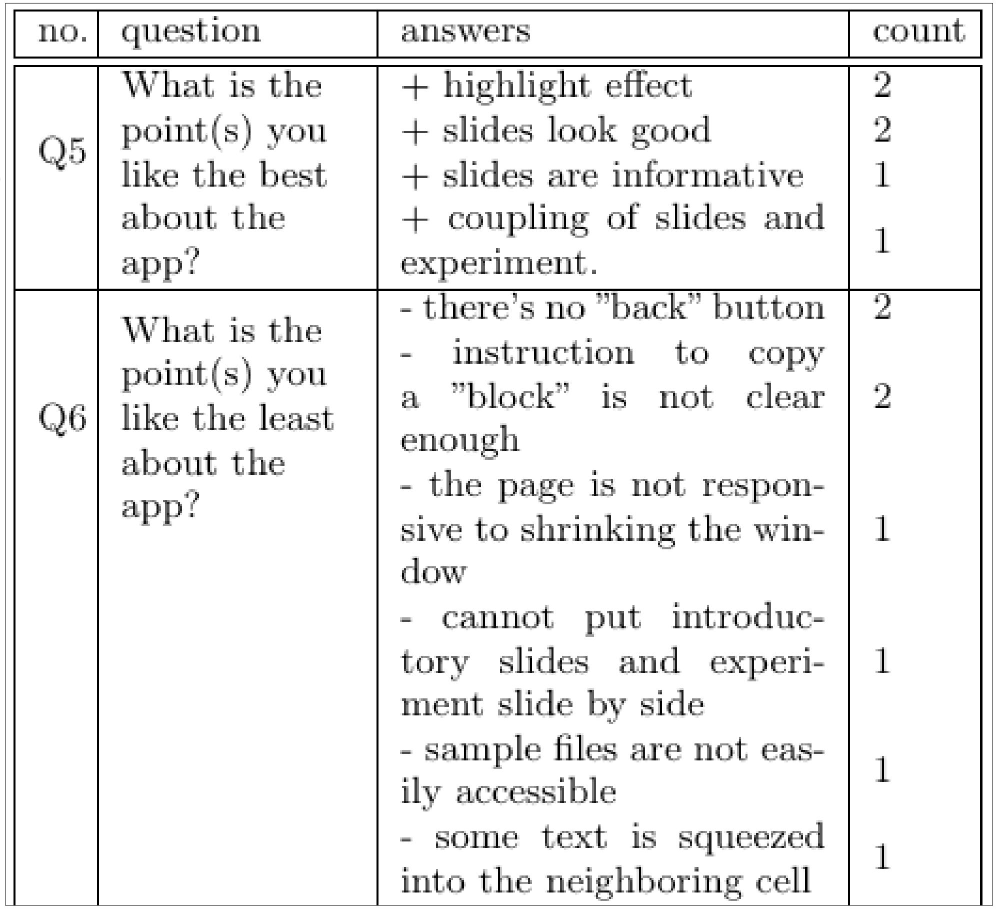


Q7	What is your suggestion to improve the app?	clearer instruction to copy a blockn	2
		”back” button	2
		inform users it is not necessary to understand everything on the slides all at once	1
		an easy way to access sample files	1
		additional information when hover over the image	1

Note: participant A's data is not included.

UNIVERSITY OF TWENTE.

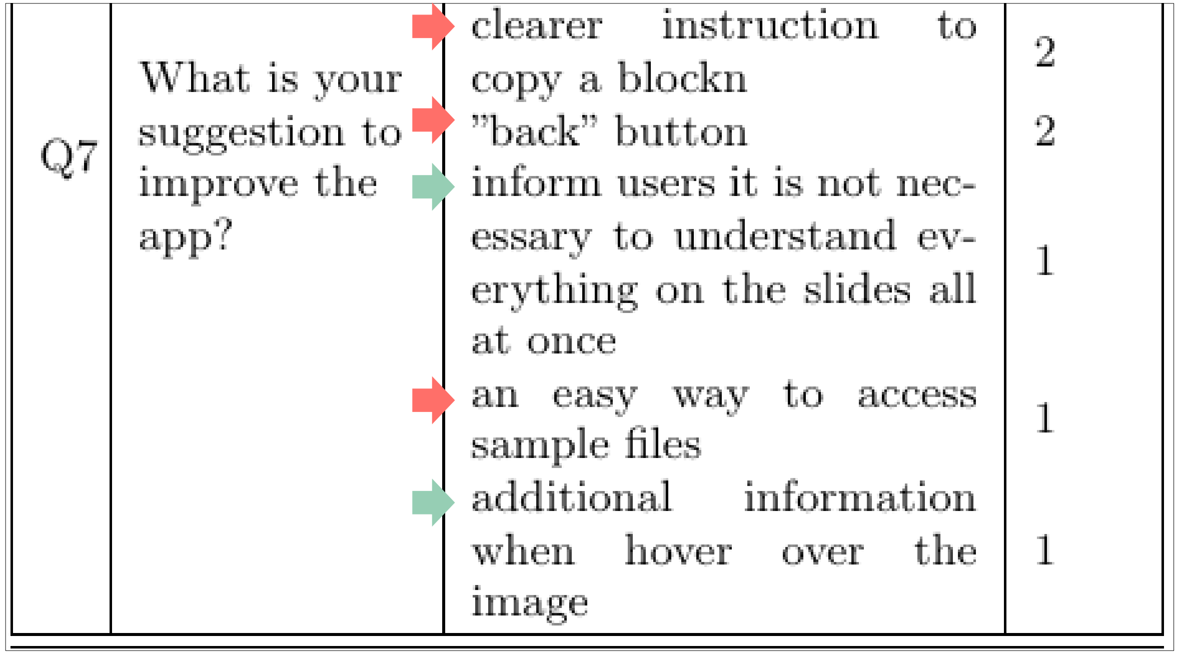
EVALUATION RESULTS & DISCUSSION



no.	question	answers	count
Q5	What is the point(s) you like the best about the app?	+ highlight effect	2
		+ slides look good	2
		+ slides are informative	1
		+ coupling of slides and experiment.	1
Q6	What is the point(s) you like the least about the app?	- there's no "back" button	2
		- instruction to copy a "block" is not clear enough	2
		- the page is not responsive to shrinking the window	1
		- cannot put introductory slides and experiment slide by side	1
		- sample files are not easily accessible	1
		- some text is squeezed into the neighboring cell	1

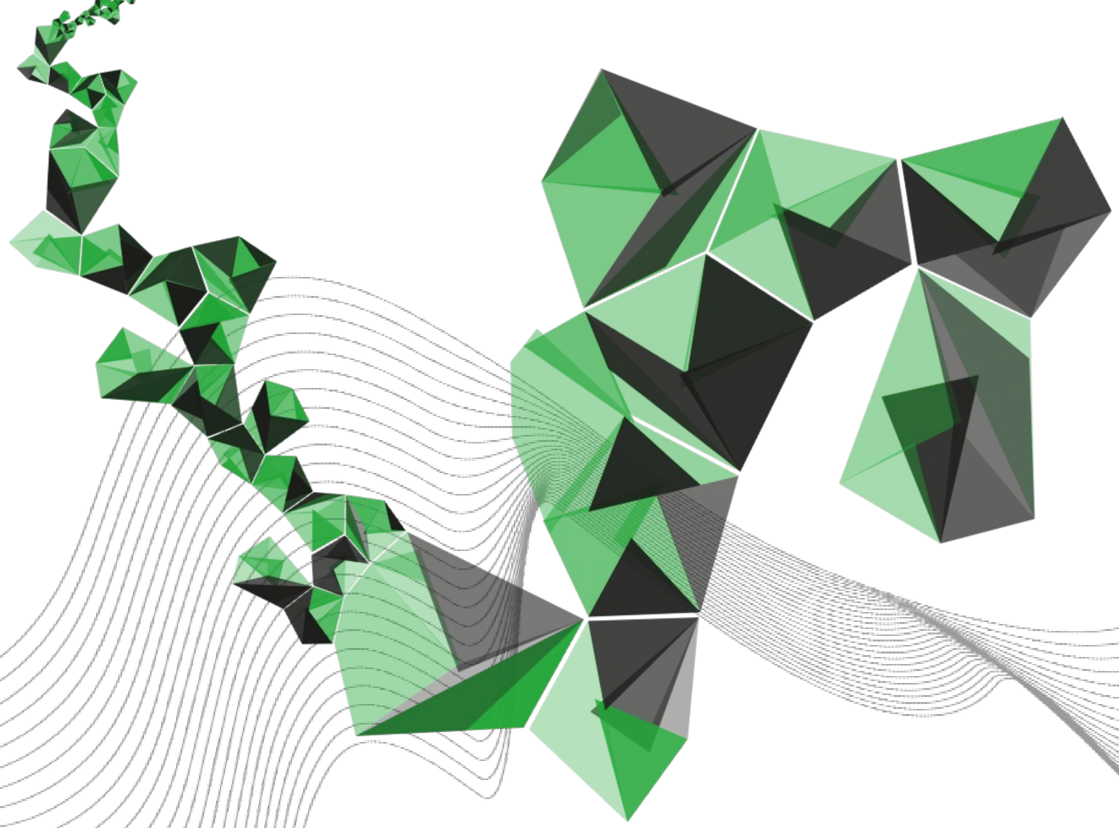
🟥 (usable)

🟩 (educational)



Q7	What is your suggestion to improve the app?	🟥 clearer instruction to copy a blockn	2
		🟥 "back" button	2
		🟩 inform users it is not necessary to understand everything on the slides all at once	1
		🟥 an easy way to access sample files	1
		🟩 additional information when hover over the image	1

Note: participant A's data is not included.



5 CONCLUSION

CONCLUSION

- RQ1: Which Searchable Encryption scheme(s) to design visualization for?
  - The SWP scheme (2000): relatively simple; helps to review.
- RQ2: How to design and implement the scheme to help novice learner learn?
  - The educational goal is met; but the usability can be improved.

FURTHER RESEARCH

- Can the same approach be applied to more Searchable Encryption schemes?
  - Many other schemes are constructed differently (index-based).
  - More thorough evaluation.
- On participant A’s experience..
  - When the knowledge is not laid well enough, will practical exercises help to fill the gap? If so, how?

THANKS!

QUESTIONS?