Matryoshka:

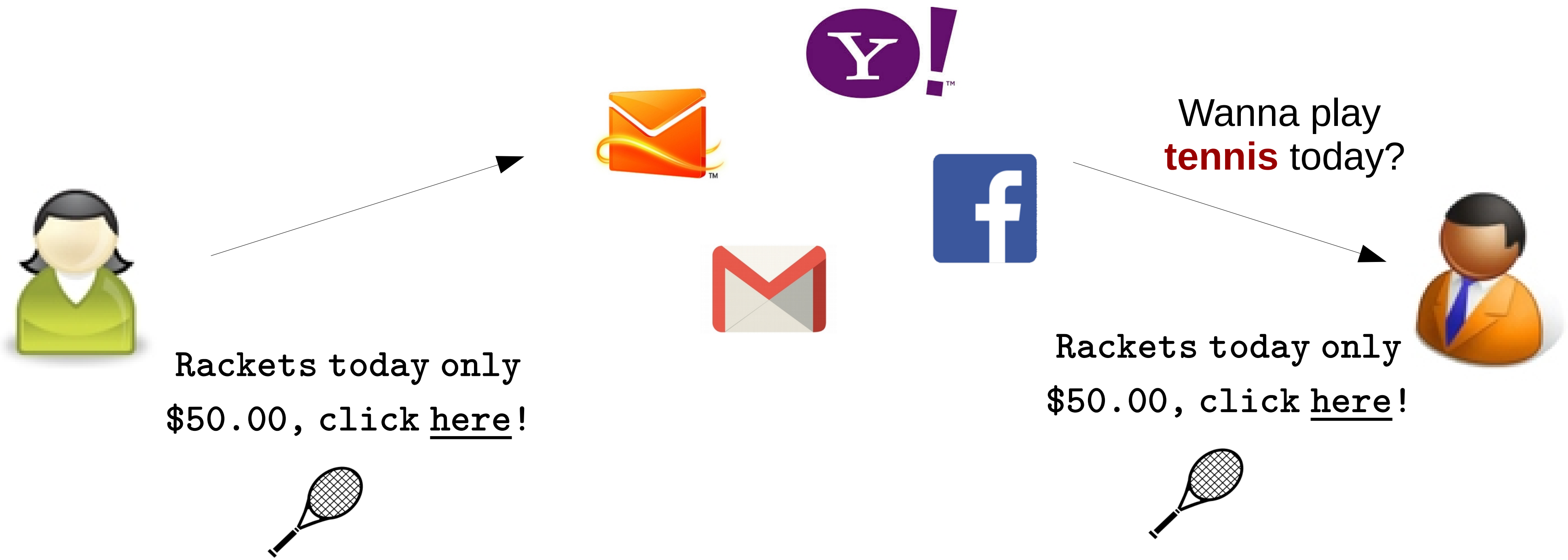
Hiding Secret Communication in Plain Sight

Iris Safaka, Christina Fragouli, Katerina Argyraki

- ***Free*** communication systems → Give away some ***privacy***

Wanna play **tennis** today?

- ***Free*** communication systems → Give away some ***privacy***



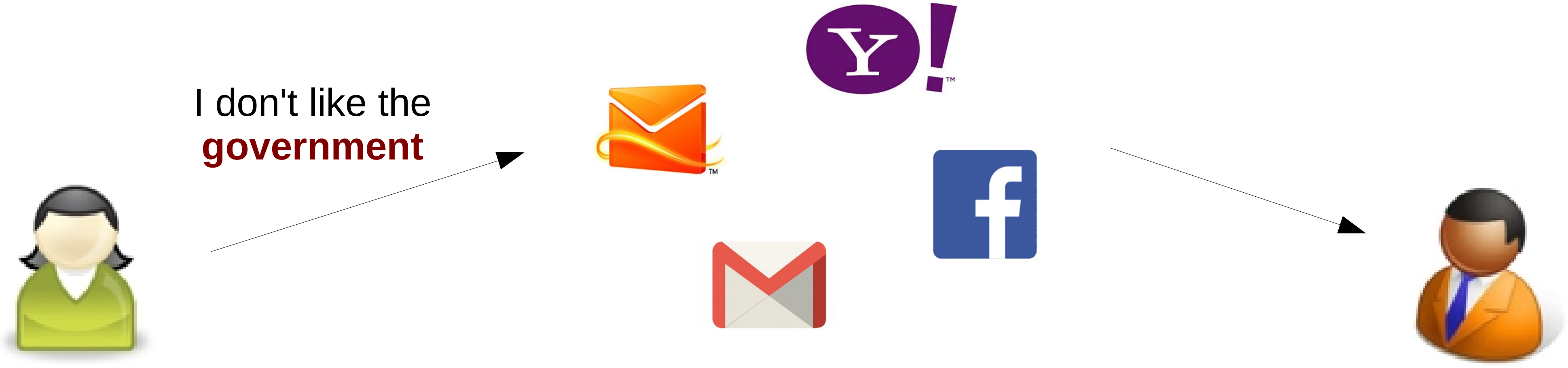
Wanna play **tennis** today?

Rackets today only $50.00, click here!

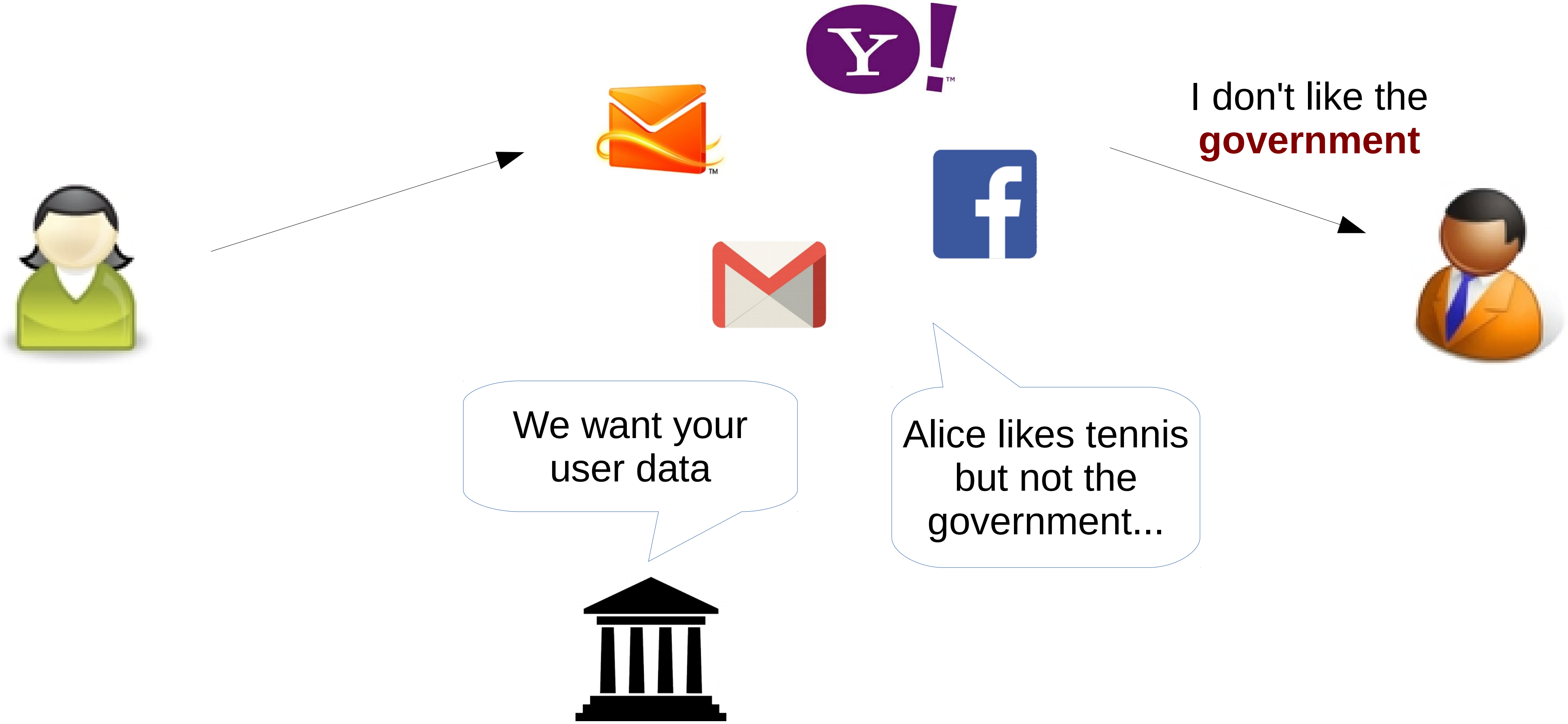
Rackets today only $50.00, click here!

- ***Free*** communication systems → Give away some ***privacy***

- Users are mostly aware of this trade-off



I don't like the **government**

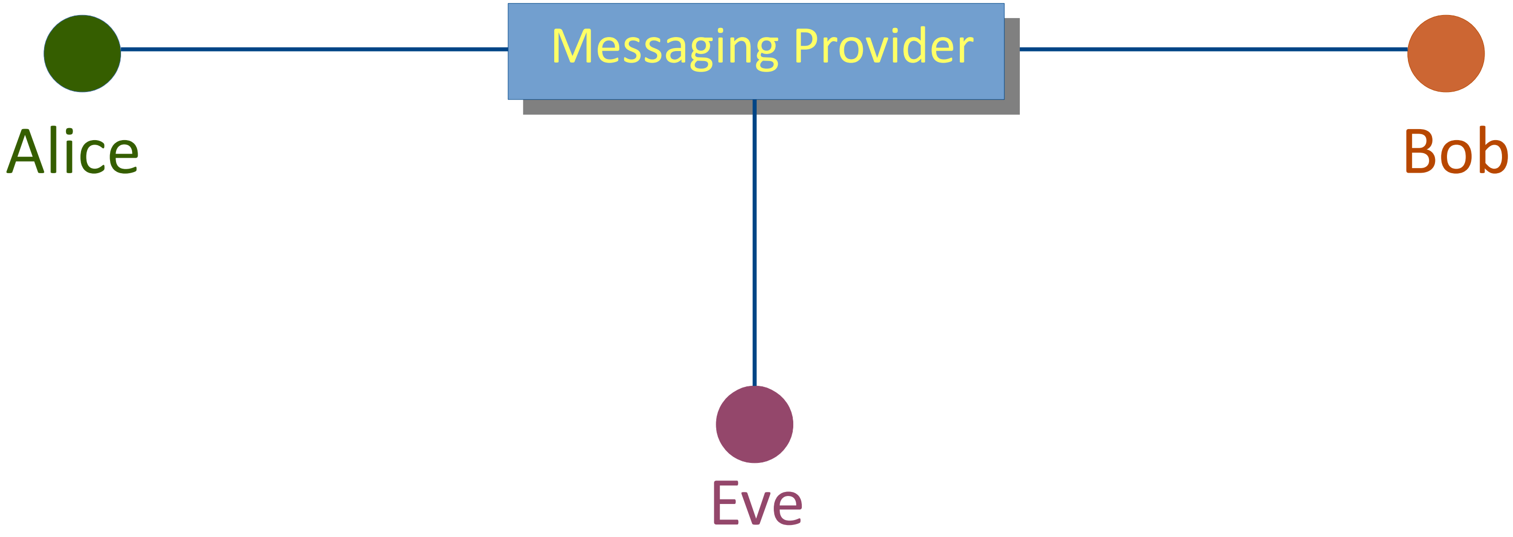


I don't like the **government**

We want your user data

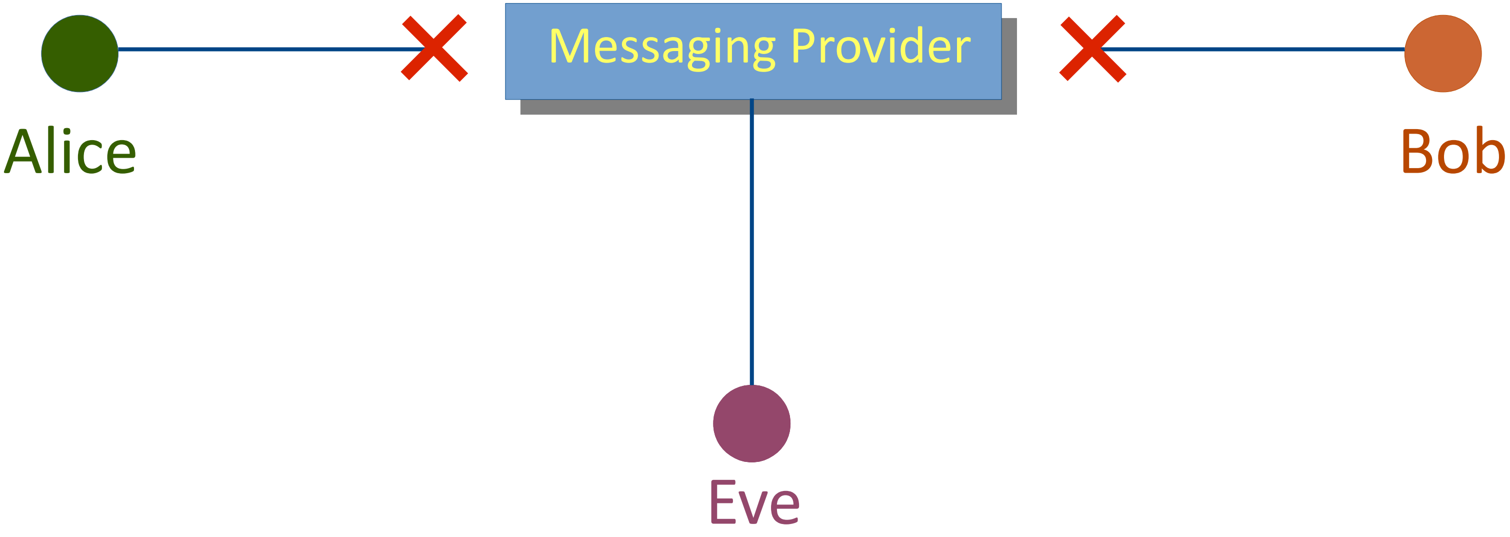
Alice likes tennis but not the government...

- Governments and courts request user data from tech companies
  - *Eg. Google handed in data for 100K user accounts (2014)*



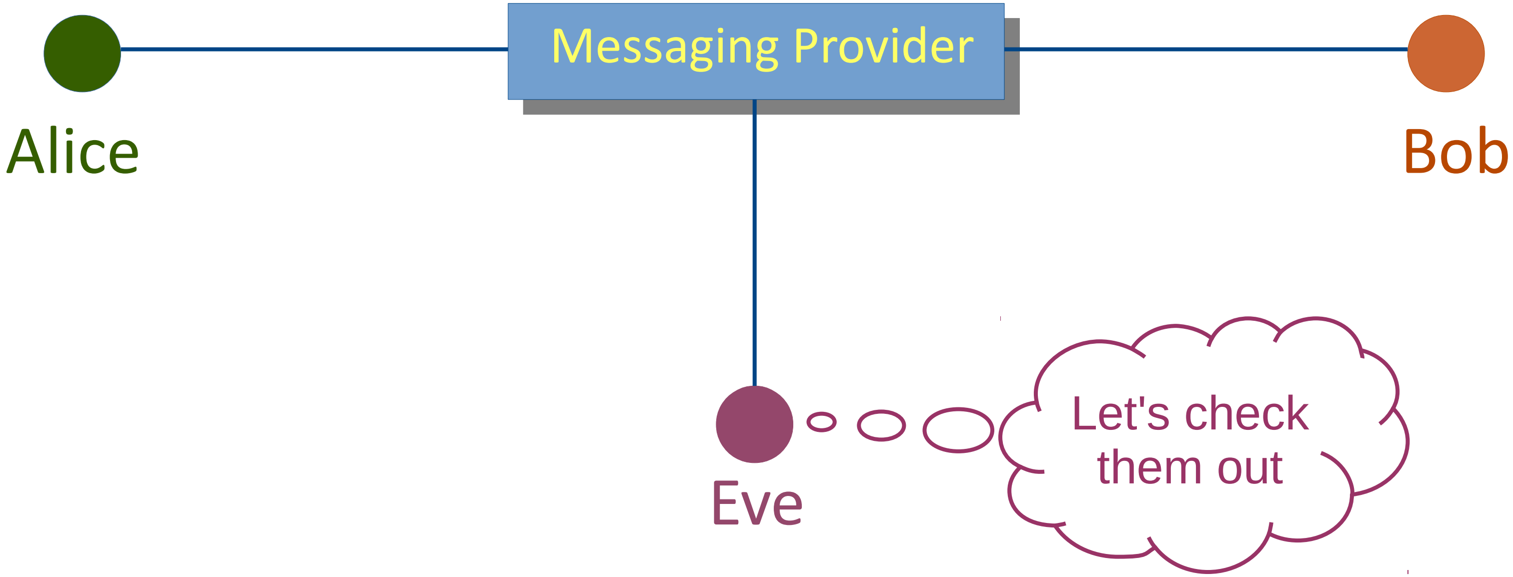
- Alice and Bob wish to **communicate privately**
- Eve always wants to know what they talk about

Encryption?



- Alice and Bob wish to **communicate privately**
- Eve always wants to know what they talk about

Encryption? → Interruption of free service



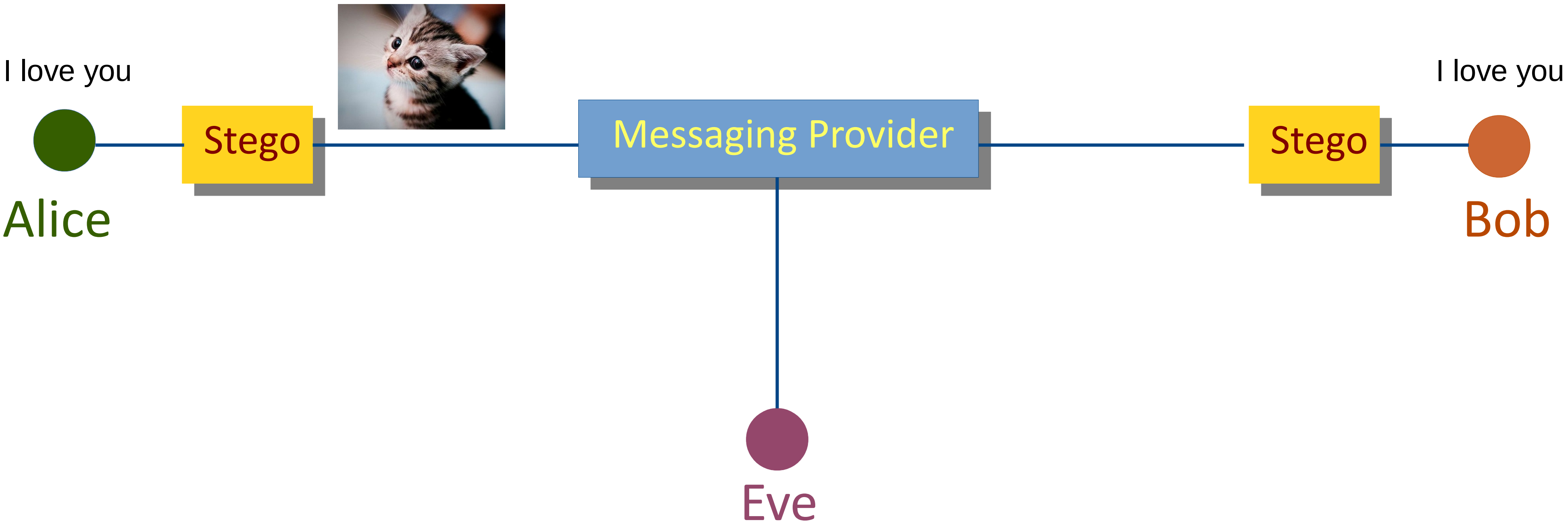
- Alice and Bob wish to **communicate privately**
- Eve always wants to know what they talk about

Encryption? → Looking suspicious

*How about **hiding** the secret communication?*

Steganography

- Hide secret data within other “innocent” data



I love you

Alice

Stego

Messaging Provider

Eve

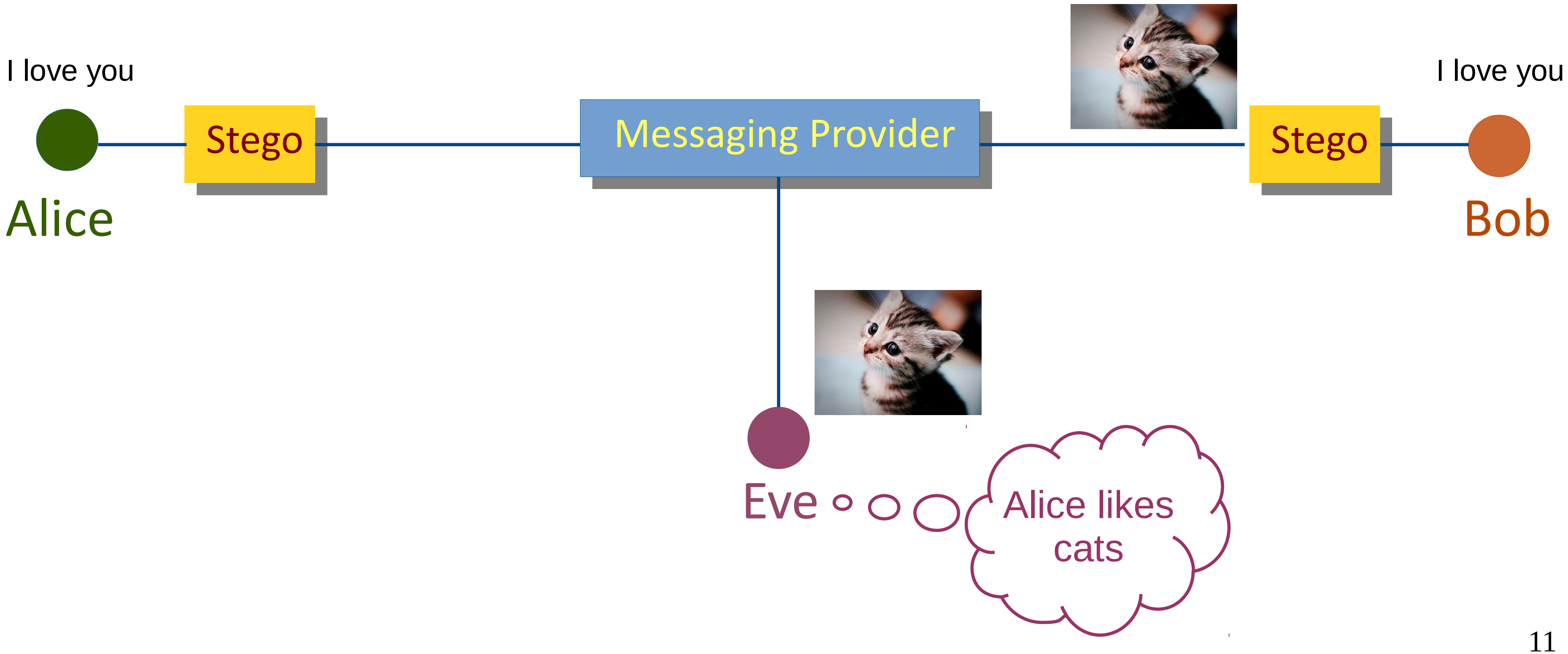
Stego

I love you

Bob

Steganography

- Hide secret data within other “innocent” data



I love you

Stego

Alice

Messaging Provider

Stego

I love you

Bob

Eve

Alice likes cats

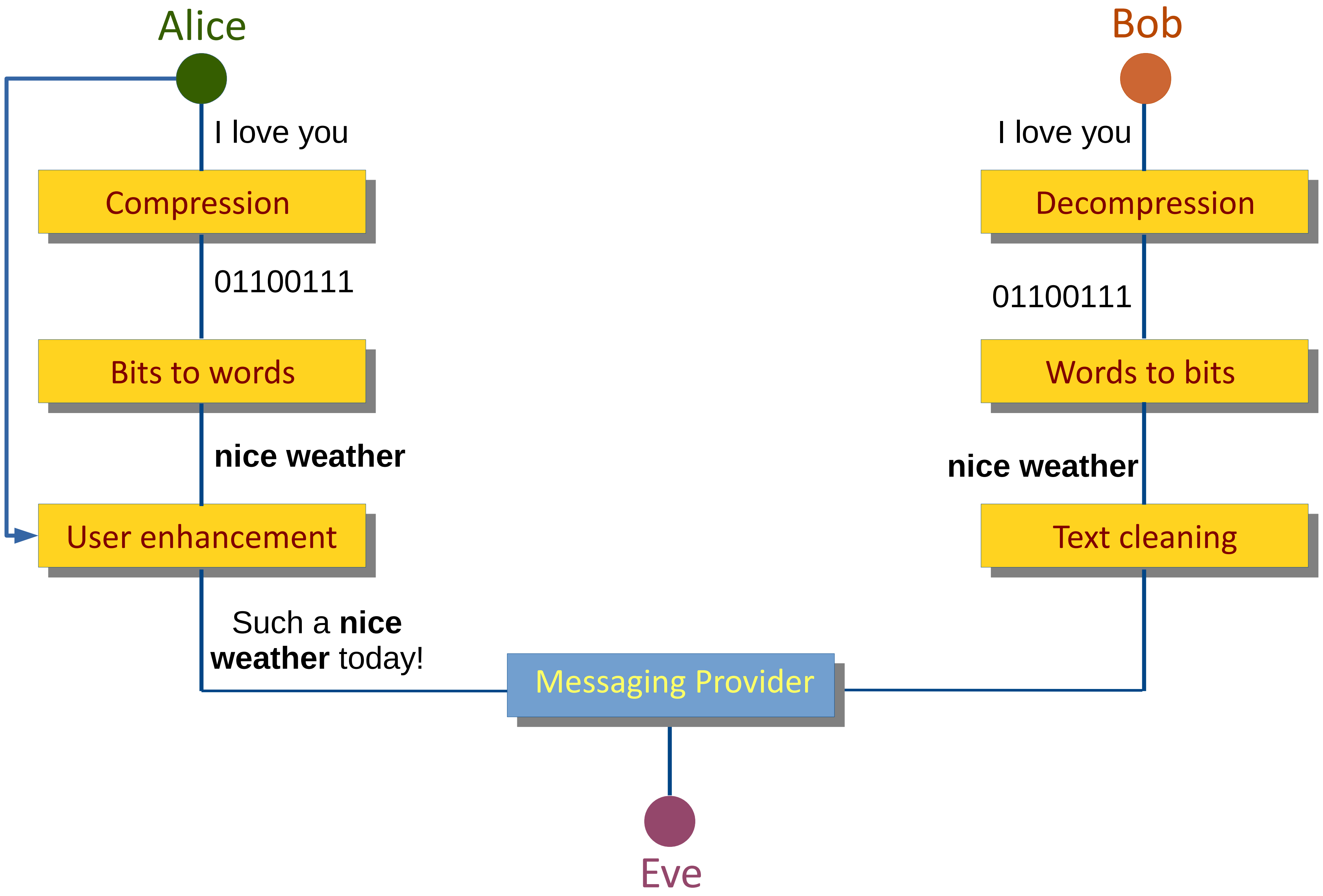
Linguistic steganography

- Traditional approaches apply **automated** modifications
  - *Embed secret message into a given text*
  - *Eg. synonym substitution, sentence manipulation etc.*

- Drawbacks
  - *Introduce* ***unnaturalness*** *to the text*
  - *Require* ***off-line*** *access to resources*
  - ***Modest*** *covert rates*

Our goal: human-like text, implementable, high rate

Matryoshka



Alice

I love you

Compression

01100111

Bits to words

nice weather

User enhancement

Such a **nice weather** today!

Messaging Provider

Eve

Bob

I love you

Decompression

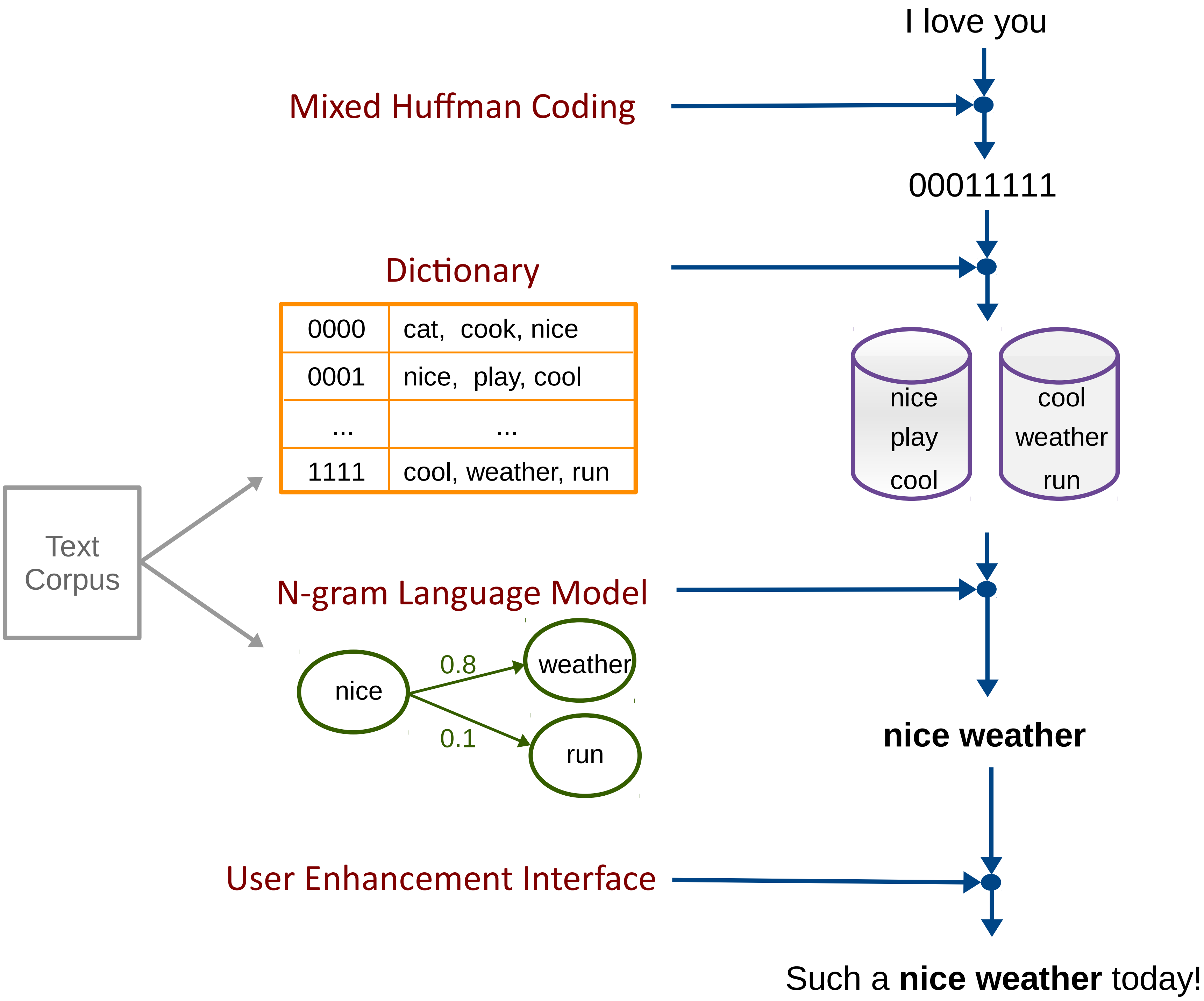
01100111

Words to bits

nice weather

Text cleaning

Challenge: minimize user intervention



I love you

Mixed Huffman Coding

00011111

Dictionary

0000	cat, cook, nice
0001	nice, play, cool
...	...
1111	cool, weather, run

Text Corpus

N-gram Language Model

nice weather

User Enhancement Interface

Such a **nice weather** today!

Encoder design

- Mixed Huffman Compression
  - *Character Huffman → names, unusual words, etc.*
  - *Word Huffman → frequent English words*

- Dictionary
  - *Maps bit sequences to sets of words*
  - *More frequent than infrequent words & repetitions*

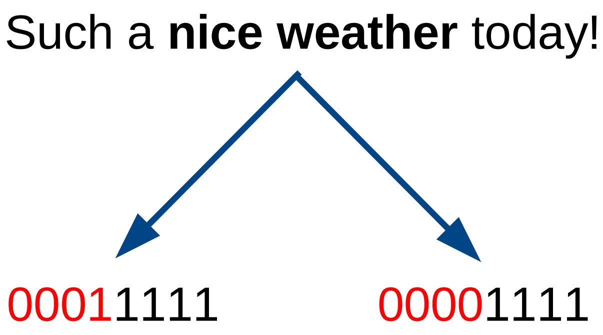
- N-gram Language Model
  - *Models how dictionary words appear in Natural Language*

- User Enhancement Interface
  - *Assist the user in completing the sentences*

Decoder design

Dictionary

0000	cat, cook, nice
0001	nice, play, cool
...	...
1111	cool, weather, run



Such a **nice weather** today!

00011111 00001111

- Repeating words in dictionary creates ambiguity

- Probabilistic decoder
  - *K-order Markov model of English characters*
  - *Drops early improbable sequences*

Evaluation

- Experimentation with human users in Amazon's Mechanical Turk

*“ I have **become** tired of **facebook's** many **years** of existence. The **change** over the **years** by the engineers sucks. It seems **facebook's** wacky **algorithm** will **never** make sense. The **posts** make the **code** on **facebook** obsolete. ”*

*“ Does **facebook's** CEO **feed** people **feed** dogs. Can't **yet** use **data** base **set** book. Two **posts** are **uses** people **facebook** apps. Mary **Cox** able **humans** into **keeping** up. ”*

Evaluation

- Experimentation with human users in Amazon's Mechanical Turk

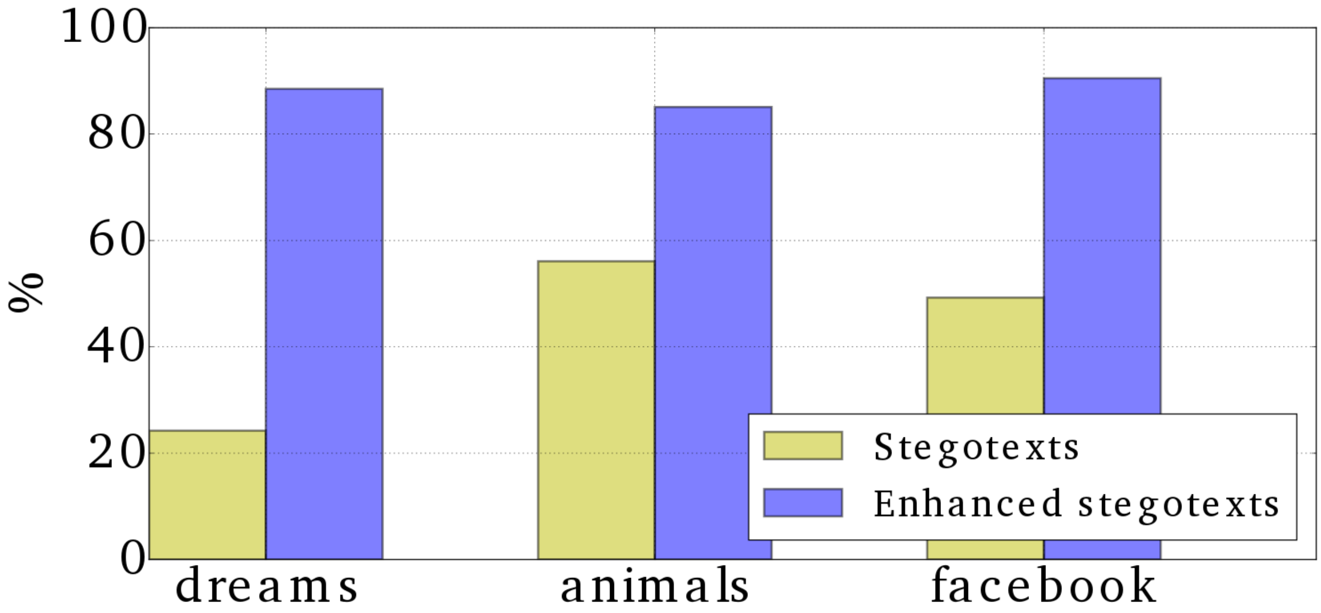
- User effort
  - *Average task completion time approx 5 mins*
  - *Average of 5 extra words inserted per sentence*

- End-to-end covert rate
  - *Average 3 bits per word*
  - *Eg. to hide 5 words we need to send 73 words*

- Decoder error rate
  - ***Zero** error rate (~95%)*
  - *Partially corrupted messages (~15% chars.)*

Evaluation

- Automatic test: Is a sentence NL or not?



Summary

- Linguistic steganography for reclaiming some privacy
- Human-like text, implementable, high covert rate
- Prototype implementation
- Experimentation on Mechanical Turk
- Automated steganalysis test

Next steps

- Investigate alternative automated steganalysis tests
  - *Eg. using Word Embeddings*

- Identify further vulnerabilities and test

- Finalize system implementation

Questions ?