Linguistic steganography: how it was

Бекян Артём

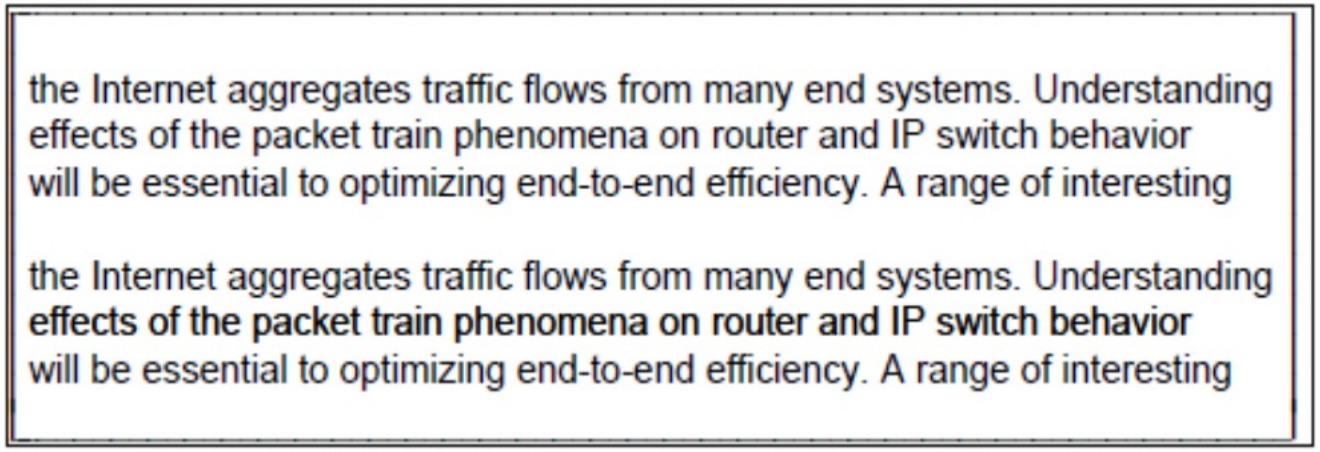
Начальные идеи: картиночные

Идея: работаем с текстом как с изображением

Примеры алгоритмов: сдвигать строки вверх или вниз, слова влево или вправо, оценивать среднее расстояние между словами

Проблема: Ctrl+c Ctrl+v все ломает

Так выглядел алгоритм со сдвигом строк:



Начальные идеи: синтаксические

Идея: меняем синтаксическую структуру

Например: можно сказать «мамина книга», можно «книга мамы», смысл не поменяется

Проблема: не все языки одинаково хорошо могут меняться. Авторы методов говорят, что Турецкий, например, очень хорошо поддается этому методу, а английский очень ограничен

Начальные идеи: семантические

Идея: меняем семантическую структуру текста

Например: заменяем слова синонимами, используем аббревиатуры

Проблема: языки не статичны, они развиваются и меняются. Со временем надежность вотермарки может пострадать. К тому же изменения семантической структуры может помешать пониманию текста

Steganographic LSTM

- **Секрет** – битовая строка длины S, разделенная на блоки бит по B в каждом
- **Ключ** – соответствие между набором бит и возможными токенами, каждый токен соответствует какому-то набору
- **Алгоритм шифрования** – используем LSTM, но она берет только токены из соответствующего набора
- **Common tokens** – токены, которые не влияют на кодирование, используются для красоты текста

Steganographic LSTM

Bit Block	Tokens
00	This, am, weather, ...
01	was, attaching, today, ...
10	I, better, an, Great, ...
11	great, than, NDA, ., ...

Bit String	10	00	01	10	11
Token	I	am	attaching	an	NDA

Meteor

- Получаем распределение вероятностей токенов как разбиение на битовые интервалы. Каждое слово из словаря соответствует какому-то интервалу, например, при precision=5, слово 1 может соответствовать [00000; 00101), слово 2 [00101, 10000) и так далее. Для дополнительной безопасности строка шифруется через псевдорандомный генератор
- Зная слово, получатель знает интервал, в который попали биты, и знает их общий префикс
- Отправитель знает, что именно узнал получатель, и, основываясь на этом, продолжает передачу сообщения

Meteor

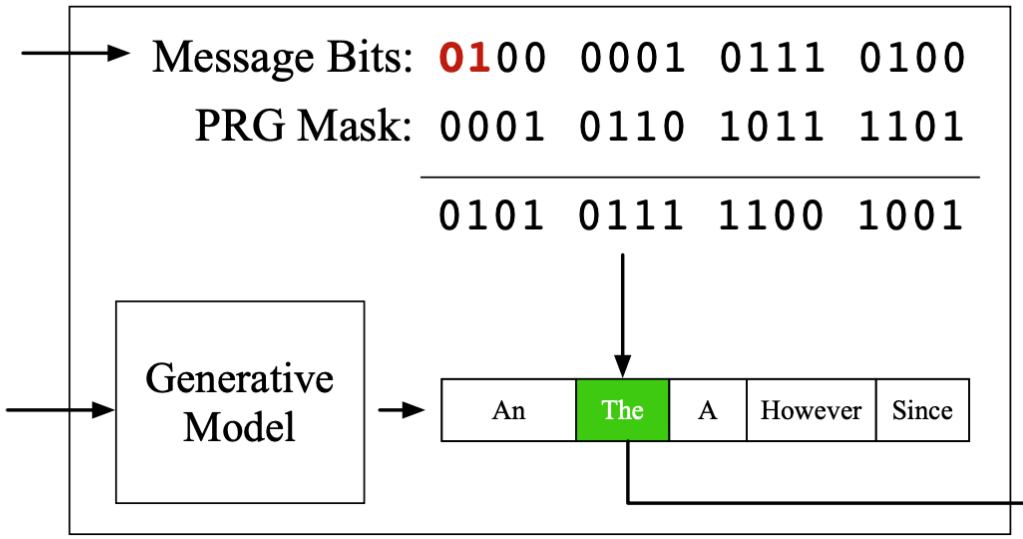
Plaintext

`Attack@Dawn`

Context

Evidence indicates that the asteroid fell in the Yucatan Peninsula, at Chicxulub, Mexico.

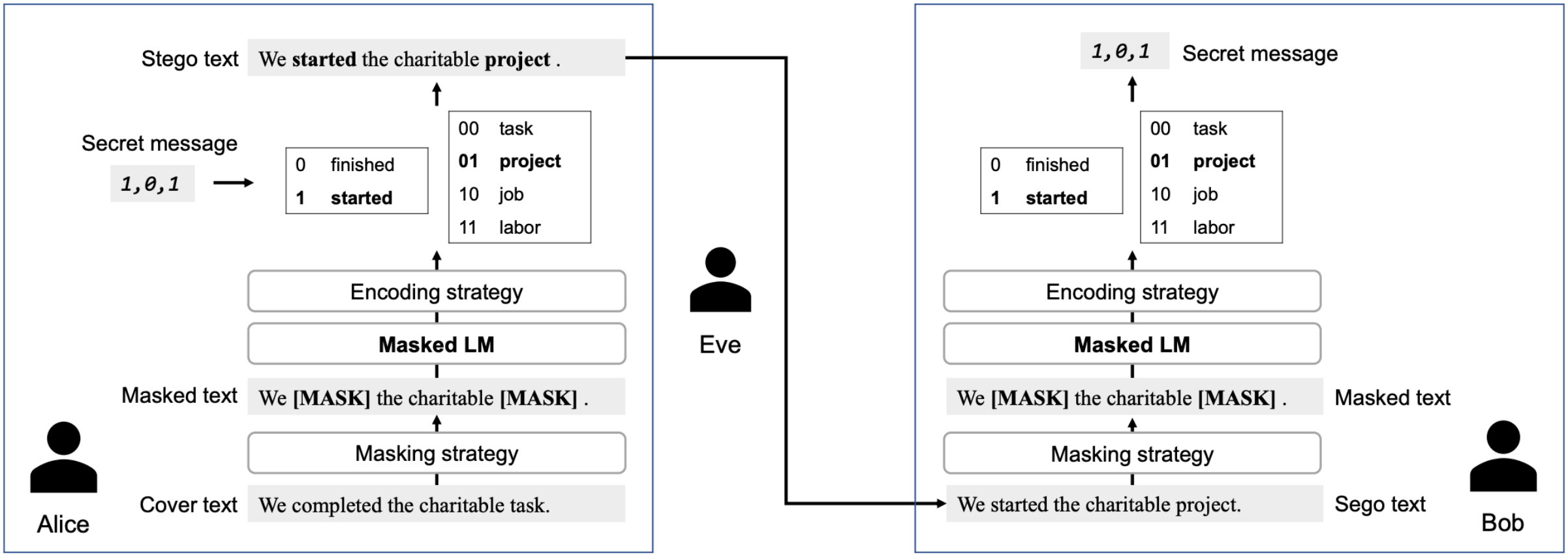
Encoder



Stegotext

The first importance of the Yucatan Peninsula is demonstrated with the following conclusion: the Pliocene Earth has lost about seven times as much vegetation as the Jurassic in regular parts of the globe, from northern India to Siberia…

Masked LM steganography



Adversarial end-to-end steganography