
Nigel Cannings | CTO
nigel.cannings@intelligentvoice.com
What is this encrypted search term?

45 EF BF BD EF BF BD 20 6F EF BF BD EF BF BD 20 55 57 D4 B2 EF BF BD 44 EF BF BD 7B C8 82 20 7F EF BF BD 20 EF BF BD 32 EF BF BD 43 CC 82 26 EF BF BD 44 EF BF BD 6F 28 EF BF BD 61 EF BF BD 20 55 EF BF BD EF BF BD EF BF BD EF BF BD 2A EF BF BD EF BF BD 2A EF BF BD 4B 7D 3C 7C 5C 31 EF BF BD EF BF BD 5B 74 54
2017 – The Year of Voice?

Amazon Alexa

SIRI(?)

LIBOR

Banks face Multi-Billion $ fines

FX Scandal
The Privacy Dilemma
Cloud Power v Cloud Security

SIRI (or a hacker who breaks into SIRI) can

- Use (edit) your voice recordings to impersonate you
- Learn about you
  → Your identity, gender, nationality (accent), emotional state..
- Track you from uploads / communications of voice recordings

Nothing specific to SIRI, same issues with Google Now, Alexa, Cortana,...

*Not a futuristic scenario*

*Everytime you use your voice, you leave a print behind!!*
Audience Participation!
Do you trust the big cloud providers (Google, Amazon, Apple etc) not to misuse your private information?

- Yes
- No
- Unsure

Source: We live in the Big Cloud: And we hate it... Is it time for Hipster IT?
Nigel Cannings, CTO Intelligent Voice
https://hackernoon.com/we-live-in-the-big-cloud-and-we-hate-it-is-it-time-for-hipster-it-1f130a44d2b8#.tqb3xnsdl
UK Government Funded Project

Project Duration: 18 months

Project Partners:

Privacy Preserving Speech Processing in the Cloud

Client

Server

Acoustic Model

Language Model

Privacy Preservation – no biometrics outside the client
Index-based Searchable Encryption

Client-side:
- Audio files
- Search for Keyword
- Consult Lexicon
- Convert Search to Phonetic Symbols
- Client-side Index Generator
- Generate Trapdoor (Private key)
- Convert to .dat Format

Server-side:
- Audio and Phonetic Transcriptions
- Cloud: AES Encrypted
- AES Encryption
- Phonetic Encoder
- Phonetic symbols:
  - k en s ih s ax…
- Identifying Phoneme Set
- Relevance Score Generator
- Server-side Secure Index Generator
- AES Encryption (Public key)
- AES Encryption (Public key)
- RESULT
Audio is converted to text (\texttt{.dat}) format and then encrypted using the well-known AES block cypher.

The cloud is used to outsource the storage of the encrypted audio.

In parallel to this process, the audio is also indexed by a phonetic encoder...
We take a novel approach using Convolutional Neural Networks (CNNs).

The audio is processed by the phonetic encoder resulting in a transcript which is then encrypted with the client public key.
Speech Recognition Using Pictures!

Pick an utterance from the TIMIT test set

<table>
<thead>
<tr>
<th>Words</th>
<th>further</th>
<th>his</th>
<th>prestige</th>
<th>he</th>
<th>occasionally</th>
<th>reads</th>
<th>the</th>
<th>wall</th>
<th>street</th>
<th>journal</th>
</tr>
</thead>
<tbody>
<tr>
<td>to</td>
<td>his</td>
<td></td>
<td></td>
<td>he</td>
<td>occasionally</td>
<td>reads</td>
<td>the</td>
<td>wall</td>
<td>street</td>
<td>journal</td>
</tr>
</tbody>
</table>

![Speech Recognition Diagram]
Speech Recognition Using Pictures!

Pick an utterance from the TIMIT test set

Taking a spectrogram of the test utterance, we can then slide a window across the spectrogram, and test our trained NVIDa DIGITS deep convolutional neural network.

NVIDIA DIGITS Prediction (Unseen Test Set)
The audio database and its phonetic transcripts have been outsourced to the cloud for storage.

- Can't be conventionally searched.

- Index-based Searchable Encryption (ISE) provides the ability to search encrypted documents within the cloud.

- The strings of integers (phonemes) form a hash chain.

- The hashing functionality (SHA-2) of the server-side index table enables strings of phonetic symbols to be searched for...

Cloud: AES Encrypted Audio and Phonetic Transcriptions
Alice wants to search the data-base for all occurrences of the phrase ‘deep learning’

She consults the lexicon which converts the search term to the phonetic string:

\[ \text{d iy p sp l er n ih ng} \quad (\text{sp means space - word boundary}) \]

The phonetic string is encoded as integers using the client-side lookup table

A trapdoor query is created with the client’s private key which encoded the string into a hash chain

The column entries of the server-side index are queried, with a masking function applied to the table for each successive symbol in the chain

If a match is found in the encrypted transcripts the relevant audio is returned.
Initialise Index
Type Search String

CLIENT

```
index.ContPerceived /workspace/stringsearch java -jar Client.jar

Enter no of documents:
4000
Index table building completed
Initial communication with server starting ...
Query key sent.
Index table sent.
Enter the string to be searched:
```

CLOUD SERVER

```
index.ContPerceived /workspace/stringsearch java -jar Server.jar

Waiting for index table communication ...
Query key received.
Index table received.
Number of data files: 4000
Waiting for query ...
```
Return Results
Pros and Cons of ISE

**PROS:**
- Probabilistic Trapdoor
- Public/Private key
- Table Masking
- Speed

**CONS:**
- Memory
- Index re-calculation

*How can it be improved?*
Index-based Searchable Encryption (ISE)
Homomorphic Searchable Encryption (HSE)

Client-side

- Audio files
- Search for Keyword
- Consult Lexicon
- Convert Search to Phonetic Symbols

Server-side

- Cloud: AES Encrypted Audio and Phonetic Transcriptions
- Homomorphically Encode Documents
- Client-side Index Generator
- Generate Trapdoor (Private key)
- RESULT

AES Encryption

Phonetic symbols: k e n s i h s a x ...

Identifying Phoneme Set

Homomorphic Searchable Encryption (HSE)

Generate Trapdoor (Private key)
Why homomorphic encryption?

What is homomorphic encryption?

Allows for operations to be performed on ciphertexts without requiring knowledge of corresponding plaintexts

\[ E(x) \oplus E(y) = E(x \otimes y) \]

What does this bring to our searchable encryption scheme?
Homomorphic more elegant but much slower:

- ISE search on 2000 documents (0.06 seconds)
- HSE search (> 1 minute)

GPU-acceleration of homomorphic string search

- Data parallelism – searches can be performed in parallel, each document could occupy a single thread in the search – reduces HSE search of 2000 documents (< 1 second)
- Model parallelism – the HSE algorithm is being ported to CUDA
- AES encryption, feature extraction (cuFFT), phonetic encoding (DIGITS), all already support CUDA
Introducing Boston Limited
Intelligent Voice’s Solutions Partner

• Over 20 years of expertise designing, supplying and supporting SUPERMICRO® Solutions
• Longstanding partnership between Intelligent Voice and Boston Limited
• Supported and optimised Intelligent Voice platforms available globally
• World class Boston Labs where you can try before you buy – including the latest NVIDIA® Pascal™ technology
• Public, Private and Hybrid cloud offerings
• Worldwide supply and support options
• The only NVIDIA® Elite Solutions Provider in Northern Europe

Stand #428 - sales@boston.co.uk  #GTC17
The Future: Crypto AI

- Microsoft research - CryptoNets – recently fully homomorphic convolution neural network implementation for MNIST handwritten character recognition

- Our phonetic encoder is also CNN-based

- It could be encrypted in a similar way

- Paves the way for encrypted speech recognition!

Questions

If you would like to learn more, please refer to the following paper:


Thank you

Come and see us at Booth #1028