Enabling Cradle-To-Grave Encryption

Alexander C. Mueller PhD
CEO and Founder of Capnion

November 16, 2018

Who is the speaker?

30 second resume:

- an ancient metro Saint Louis townie
- grew up in University City
- University City High School
- B.A. Washington University (econ and math)
- Ph.D. University of Michigan (math)
- private sector data science
- founded data privacy company Capnion

The Counter-Earth

Orbiting exactly opposite the Earth you are familiar with is another Earth, alike in many ways but also different...

Agenda

Important Concepts

- Motivation
- Zero-Knowledge Proofs
- Homomorphic Encryption
- Application to Practical Examples

Present and Future Application

- Data Ownership and New Business Models
- Private Datascience at Capnion
- What's out there?
- Prophecy

Basic Bookkeeping

Who will volunteer to do some basic bookkeeping?

Assignment: Basic Bookkeeping

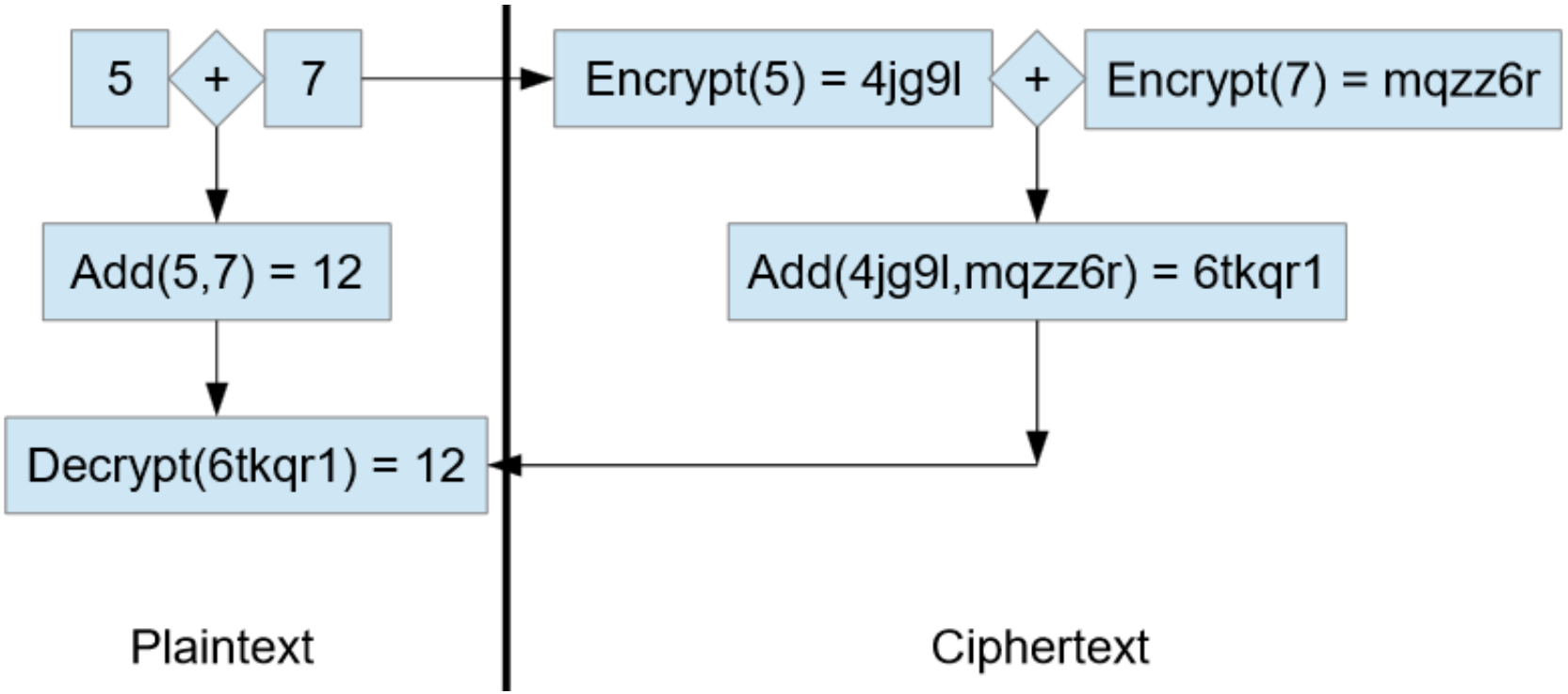
Our contractual obligations include...

- verifying that the ending funds figure is correct
- maintain secrecy of expense amounts and starting funds

Expense	Amount	Starting Funds
fancy pens	XXXXXXXXX	XXXXXXXXX
private jet	XXXXXXXXX	**Ending Funds**
"business entertainment"	XXXXXXXXX	$30173
attorneys	XXXXXXXXX	

Thanks for handling this. I've got a tee time soon. See you later.

Homomorphic Encryption



Homomorphic encryption allows one to do computations on encrypted data and get the “correct” answer after decryption.

What is a zero knowledge proof?

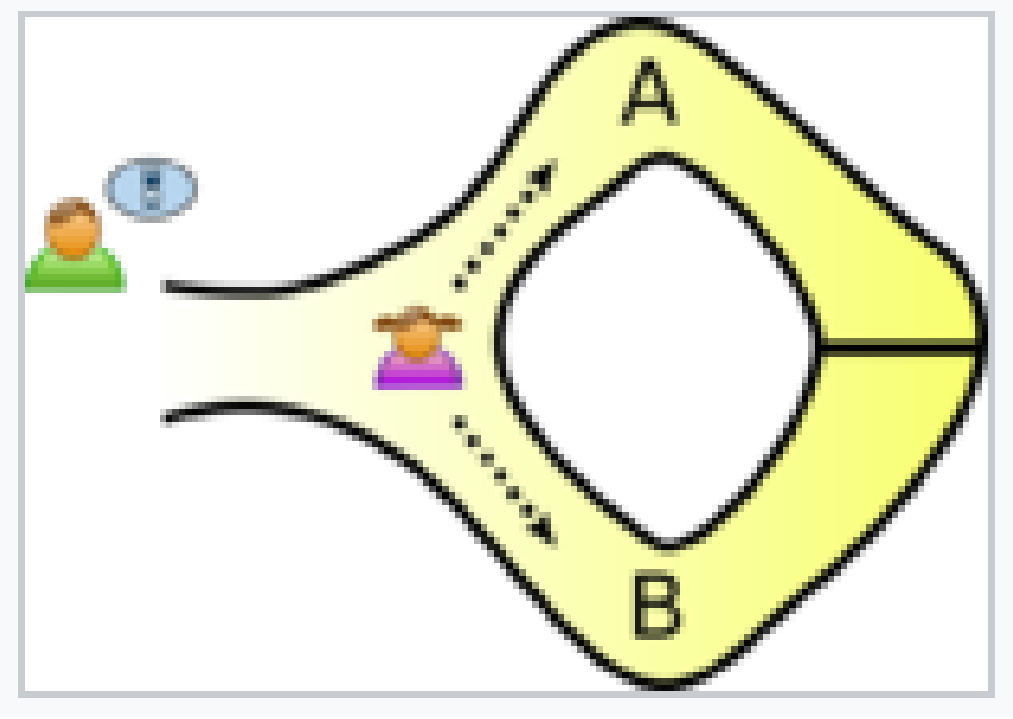
A **zero-knowledge proof** is a method by which a *prover*, traditionally called Peggy, proves to a *verifier*, traditionally called Victor, that

- Peggy possess a particular piece of information, and...
- Peggy does not reveal that information to Victor.

Example: Digital Signatures

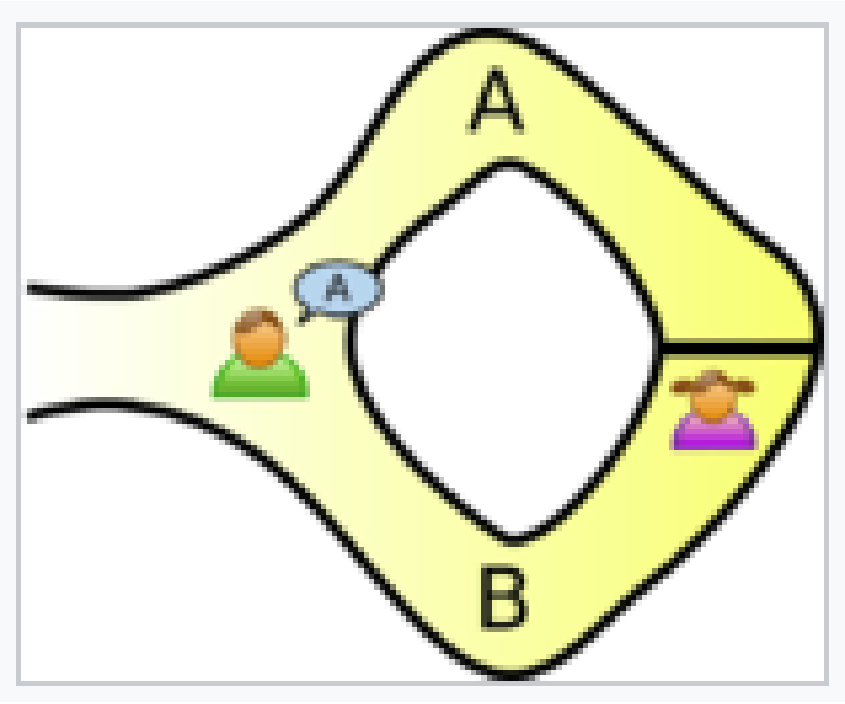
Peggy proves to Victor that she holds the private key corresponding to a particular public key but Victor does not receive any information about the private key itself.

Classic Cave Example 1



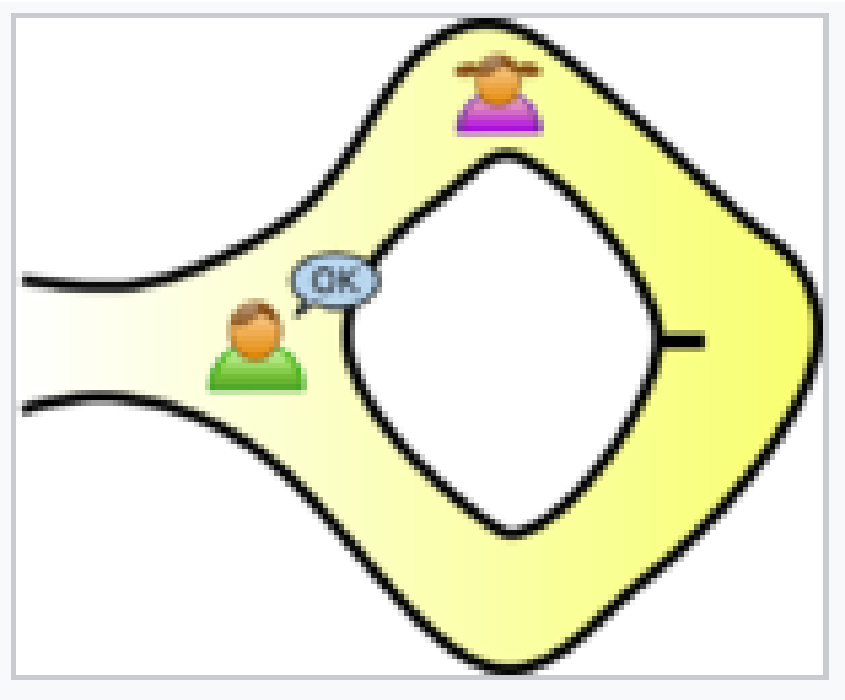
Peggy wishes to prove to Victor that she can **open the magic door** in the back of a donut-shaped cave, yet she wants to **keep her method for opening the door** a secret from Victor.

Classic Cave Example 2



After Peggy has entered the cave by one path or the other, Victor shouts into the cave whether Peggy should return by Path A or Path B.

Classic Cave Example 3



If Peggy can consistently return by the path Victor requests, she has demonstrated that she has a method for passing through the magic door, yet Victor has not seen Peggy open the door.

Basic Bookkeeping Revisited

Expense	Amount	Starting Funds
fancy pens	XXXXXXXXX	XXXXXXXXX
private jet	XXXXXXXXX	**Ending Funds**
"business entertainment"	XXXXXXXXX	$30173
attorneys	XXXXXXXXX	

Adding up the unknown, encrypted amounts to get an unknown and correct but still encrypted amount is **homomorphic encryption**.

Verifying that this encrypted amount is $30,173 without decrypting anything is a **zero-knowledge proof**.

Data Ownership

Would you like to “own” your personal data?

What does it mean to own information?

Buzzword Salad

If you are looking to do more of one these but you have no expertise in your organization, how do you get started?

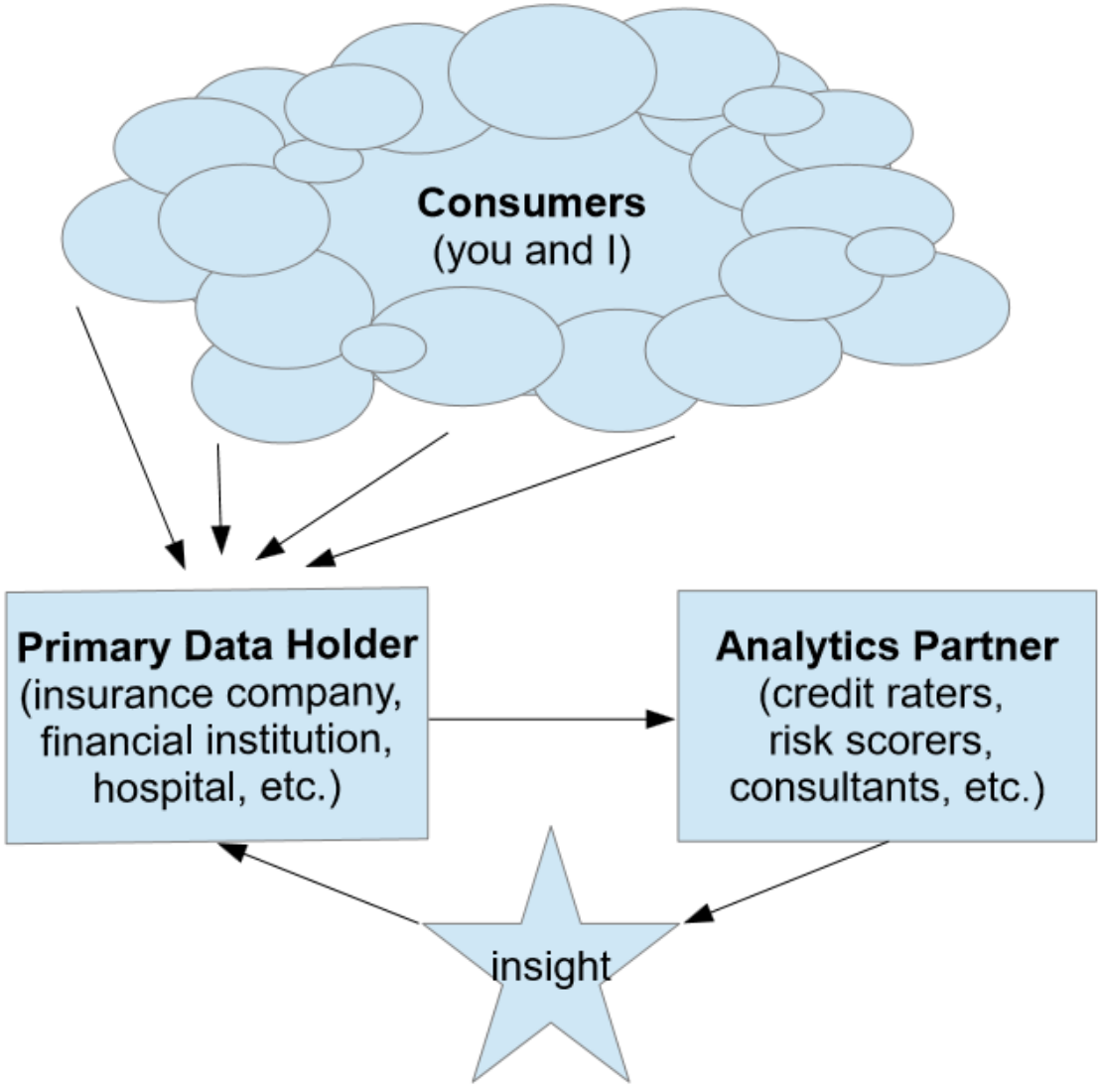
Analytics Partners

Personal data you give to a business often passes through a number of hands in its lifetime. Partners that primary data holders (example: a hospital) work with to better analyze your data (example: consultants) are an interesting example we'll examine in detail.

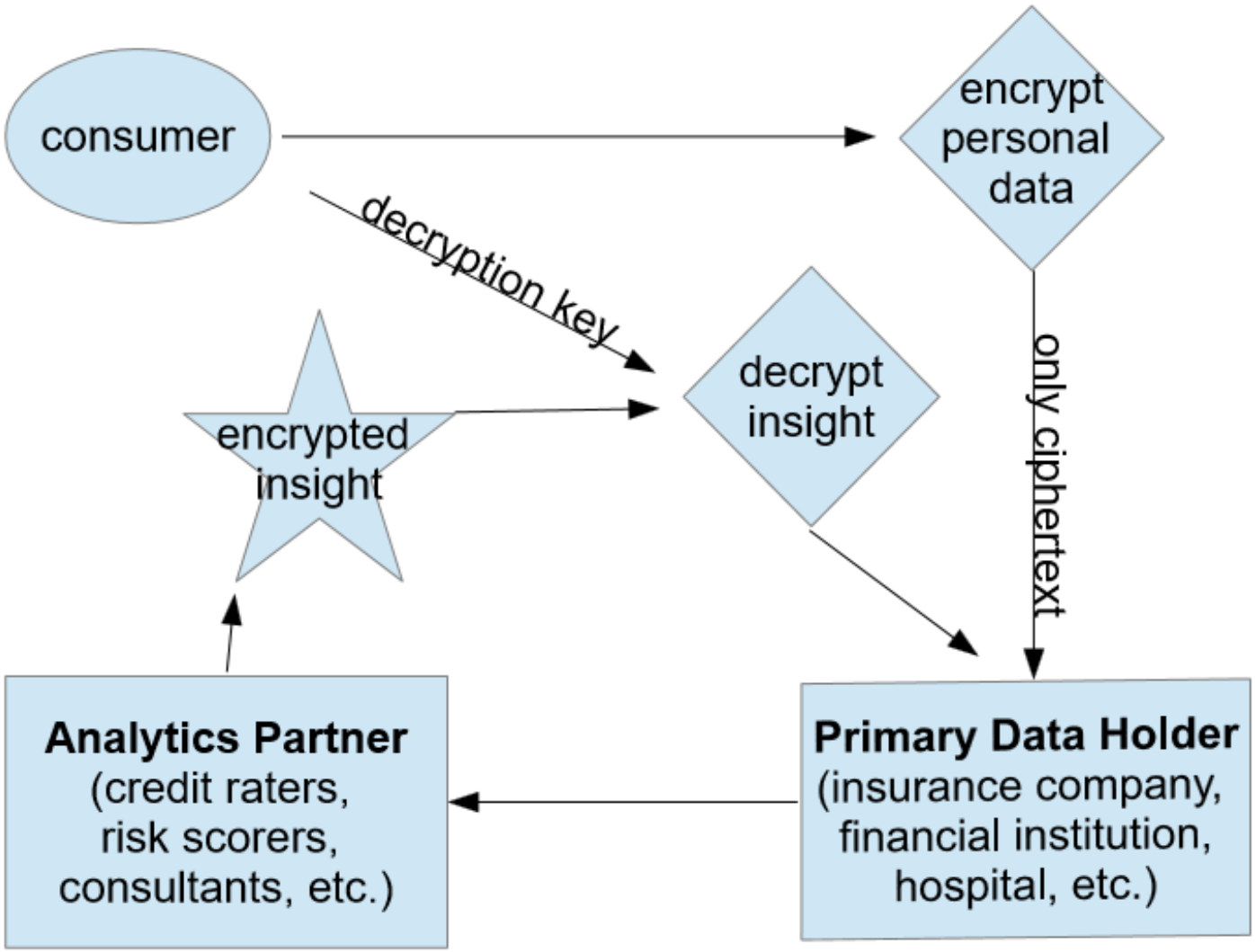
Depending on the industry, there are many businesses in this rough category...

- Raters of consumer credit
- Clearing houses for financial transaction
- Risk scoring in insurance
- Many forms of consulting

Typical Data Flow



Data Flow with Cryptographic Ownership



Contrast Between New and Old

In the newer cryptographic ownership model...

- no one but the consumer holds unencrypted personal information.
- analytics insights are the only plaintext in circulation.

When data is decrypted...

- it requires the approval of the consumer.
- it occurs late in the pipeline passes through fewer hands.
- it reveals only necessary, practical information.

The ability to tactically **allow or prevent utilization of a resource** agrees with our everyday intuition of what “ownership” should mean.

Security Subtleties 1

Let's say I have an encrypted string and an algorithm that takes a plaintext search string and returns the indices where that string appears.

$$\text{search}(\text{“a”}) = \{2, 8, 13\}$$

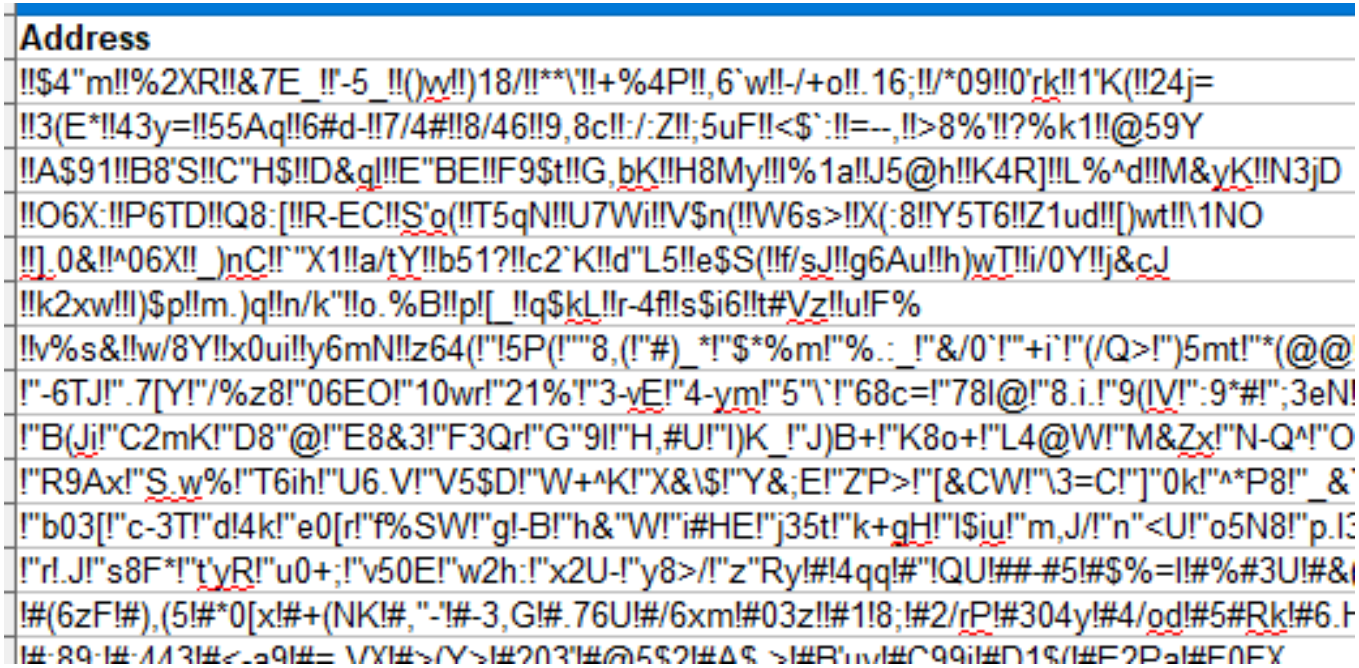
$$\text{search}(\text{“b”}) = \{1\}$$

$$\text{search}(\text{“c”}) = \{\}$$

and so on ...

It will never take me more than 26 tries before I can reconstruct “**bart skateboards**” as the plaintext.

Private Datascience at Capnion 1



Address
!!$4"m!!%2XR!!&7E_!!'-5_!!()w!!)18/!!**\'!!+%4P!!,6`w!!-/+o!!.16;!!/*09!!0'rk!!1'K(!!24j=
!!3(E*!!43y=!!55Aq!!6#d-!!7/4#!!8/46!!9,8c!!:/:Z!!;5uF!!<$`:!!=--,!!>8%'!!?%k1!!@59Y
!!A$91!!B8'S!!C"H$!!D&ql!!E"BE!!F9$t!!G,bK!!H8My!!I%1a!!J5@h!!K4R]!!L%^d!!M&yK!!N3jD
!!O6X:!!P6TD!!Q8:[!!R-EC!!S'o(!!T5qN!!U7Wi!!V$n(!!W6s>!!X(:8!!Y5T6!!Z1ud!![)wt!!\1NO
!!].0&!!^06X!!_)nC!!`"X1!!a/tY!!b51?!!c2`K!!d"L5!!e$S(!!f/sJ!!g6Au!!h)wT!!i/0Y!!j&cJ
!!k2xw!!l)$p!!m.)q!!n/k"!!o.%B!!p![_!!q$kL!!r-4f!!s$i6!!t#Vz!!u!F%
!!v%s&!!w/8Y!!x0ui!!y6mN!!z64(!"!5P(!""8,(!"#)_*!"$*%m!"%.:_!"&/0`!'"+i`!'(/Q>!")5mt!"*(@@
!"-6TJ!".7[Y!"/%z8!"06EO!"10wr!"21%!"3-vE!"4-ym!"5"\'!"68c=!"78l@!"8.i.!"9(lV!":9*#!";3eN!
!"B(Ji!"C2mK!"D8"@!"E8&3!"F3Qr!"G"9l!"H,#U!"I)K_!"J)B+!"K8o+!"L4@W!"M&Zx!"N-Q^!"O
!"R9Ax!"S.w%!"T6ih!"U6.V!"V5$D!"W+^K!"X&\$!"Y&;E!"Z'P>!"[&CW!"\3=C!"]"0k!"^*P8!"_&
!"b03[!"c-3T!"d!4k!"e0[r!"f%SW!"g!-B!"h&"W!"i#HE!"j35t!"k+qH!"l$iu!"m,J/!"n"<U!"o5N8!"p.I3
!"r!.J!"s8F*!"t'yR!"u0+;!"v50E!"w2h:!"x2U-!"y8>/!"z"Ry!#!4qq!#"!QU!##-#5!#$%=I!#%#3U!#&
!#(6zF!#),(5!#*0[x!#+(NK!#,"-'!#-3,G!#.76U!#/6xm!#03z!#1!8;!#2/rP!#304y!#4/od!#5#Rk!#6.H
!#:89:!#;443!#<-a9!#= \X!#>(Y>!#203'!#@5$2!#A$ >!#B'uv!#C99i!#D1$(!#E2Ra!#F0EX

If you could dedupe this list of addresses for me, that would be great. Make sure to be on the lookout for cases where one address is listed twice with different spellings.

Private Datascience at Capnion 2

```
"""
We'll use a clustering algorithm to pick out a group of similar address
USING ONLY THE ENCRYPTED DATA
and then verify that our grouping makes sense by looking at the original plaintext
"""

cipherSimil = np.array([[1-t for t in subList] for subList in cipherList])
#cluster the rows of the encrypted spreadsheet
#based on the distances computed between the ciphertext addresses
clustering = DBSCAN(eps=0.5, metric = 'precomputed',min_samples=1).fit(cipherSimil)
#go back to the plaintext spreadsheet to see what our results were
np.where(clustering.labels_==1)
plaintext["Address"][np.where(clustering.labels_==1)[0]]
```

```
1           1 N Ogden Ave
2           1 N Ogden Ave
3           1 N Ogden Ave
4           1 N Ogden Ave
5              1 N. Ogden
6        1 N. Ogden Avenue
7    1 North Ogden Avenue
Name: Address, dtype: object
```

Here we not only dedupe but also do some basic clustering to fuzzily dedupe and all **on the data while it is was still encrypted**. The displayed plaintext is only a reality check.

Private Datascience at Capnion 3

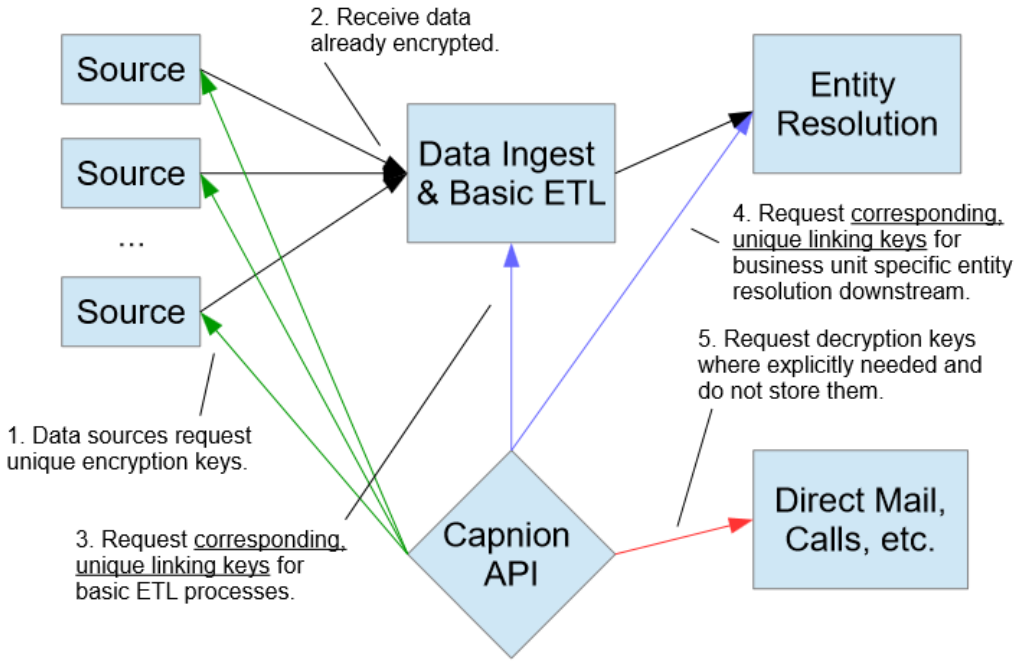
```
"""
We'll use a clustering algorithm to pick out a group of similar address
USING ONLY THE ENCRYPTED DATA
and then verify that our grouping makes sense by looking at the original plaintext
"""

cipherSimil = np.array([[1-t for t in subList] for subList in cipherList])
#cluster the rows of the encrypted spreadsheet
#based on the distances computed between the ciphertext addresses
clustering = DBSCAN(eps=0.5, metric = 'precomputed',min_samples=1).fit(cipherSimil)
#go back to the plaintext spreadsheet to see what our results were
np.where(clustering.labels_==1)
plaintext["Address"][np.where(clustering.labels_==1)[0]]
```

```
1           1 N Ogden Ave
2           1 N Ogden Ave
3           1 N Ogden Ave
4           1 N Ogden Ave
5              1 N. Ogden
6        1 N. Ogden Avenue
7    1 North Ogden Avenue
Name: Address, dtype: object
```

Here we not only dedupe but also do some basic clustering to fuzzily dedupe and all **on the data while it is was still encrypted**. The displayed plaintext is only a reality check.

Security Subtleties 2



Generate keys that enable **particular operations on particular sets of data**, and specifically keys that give information on **how to link records but not on how to decrypt** or do other computations

What's Out There?

- HElib - fully homomorphic encryption software library
  - the algorithms, close to the metal in C++
  - https://github.com/shaih/HElib
- HomomorphicEncryption R Package
  - intuitive interface in R
  - http://www.louisaslett.com/HomomorphicEncryption/
- PySEAL Python Package
  - intuitive interface in Python
  - https://github.com/Lab41/PySEAL
- OpenMined - crowdsourced, private machine learning
  - https://github.com/OpenMined
- Zcash - cryptocurrency using zero-knowledge proofs
  - https://z.cash/

History and Prophecy

Some landmark papers and people...

- (Gentry 2009) “A Fully Homomorphic Encryption Scheme”
- (Gennaro, Gentry, Parno, Raykova 2013) “Quadratic Span Programs and Succinct NIZKs without PCPs”

Craig Gentry won a MacArthur Fellowship (the “genius grant”) for his 2009 PhD thesis work on fully homomorphic encryption (able to compute an arbitrary algorithm on ciphertext) and was widely interviewed afterwards.

He said at the time, almost 10 years ago, that it might be 10 years before the technology really started to catch on...

Questions

Any questions?

acmueller@capnion.com

Slides are available at
[link]