SDC

STORAGE DEVELOPER CONFERENCE

SNIA ■ SANTA CLARA, 2017

!Oxymoron: Encrypted (Database) Search

Srinivasan Narayanamurthy (Srini)

NetApp

NetApp

Agenda

- Survey
- Homomorphic Encryption 101
- Encrypted Search
- Tradeoffs
  - Leakage
  - Functionality
- Encrypted Databases

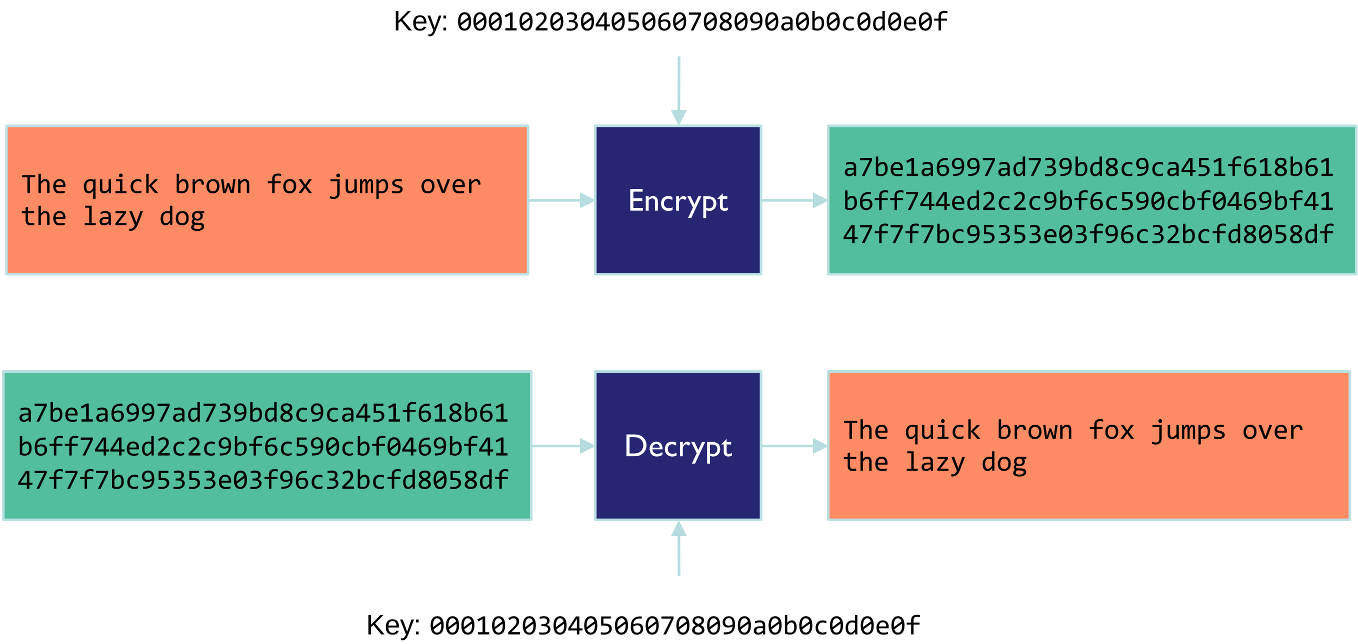
Survey

- Non-cryptographic methods
  - Differential Privacy (Noise)
  - Data Anonymization
  - Data Fragmentation
- Secret-Sharing based methods
  - Verifiable (collaborative)
  - Order Preserving
- Index based methods
  - Bucketization
  - Order-preserving
  - Searchable

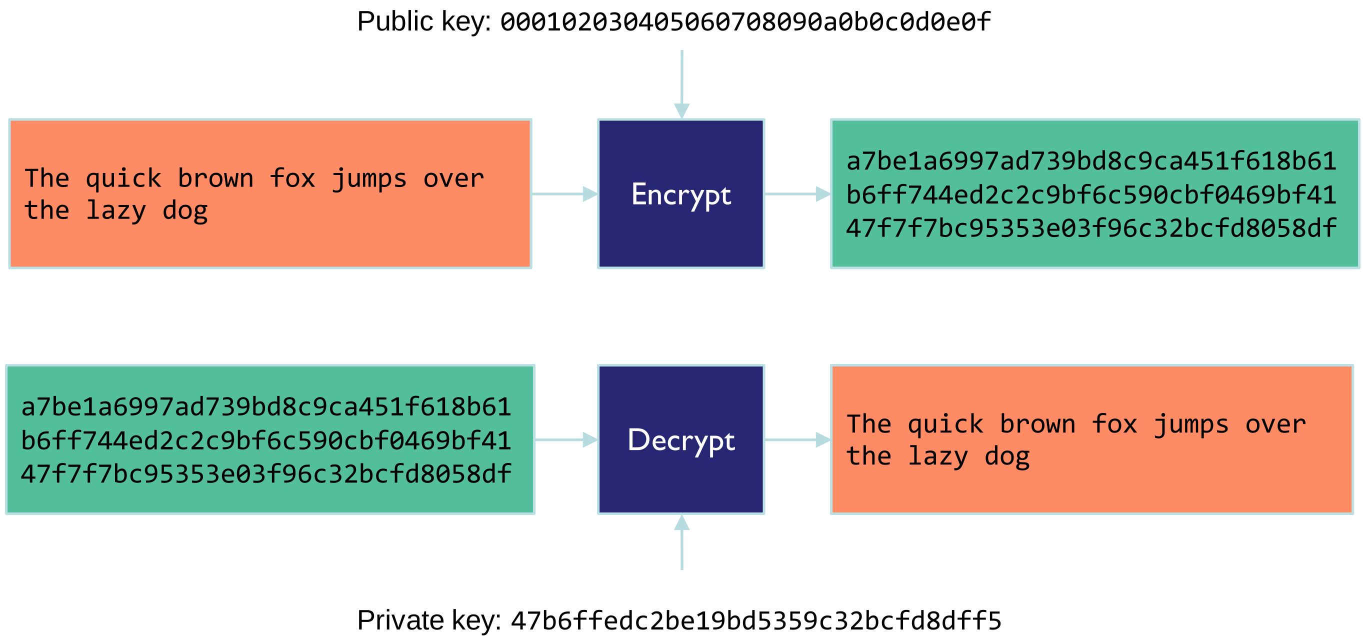
Survey (Continued)

- Cryptographic
  - Functional Encryption
  - Searchable Encryption
  - Secure-Multiparty Computation
  - Homomorphic Cryptosystems
    - Fully (FHE)
    - Partial (PHE)
- State-of-the-art Systems
  - Systems based on Homomorphic (*CryptDB*)
  - Client-server splitting approaches (*Monomi, Silverline*)
  - Trusted Hardware Systems (*TrustedDB, Cipherbase*)

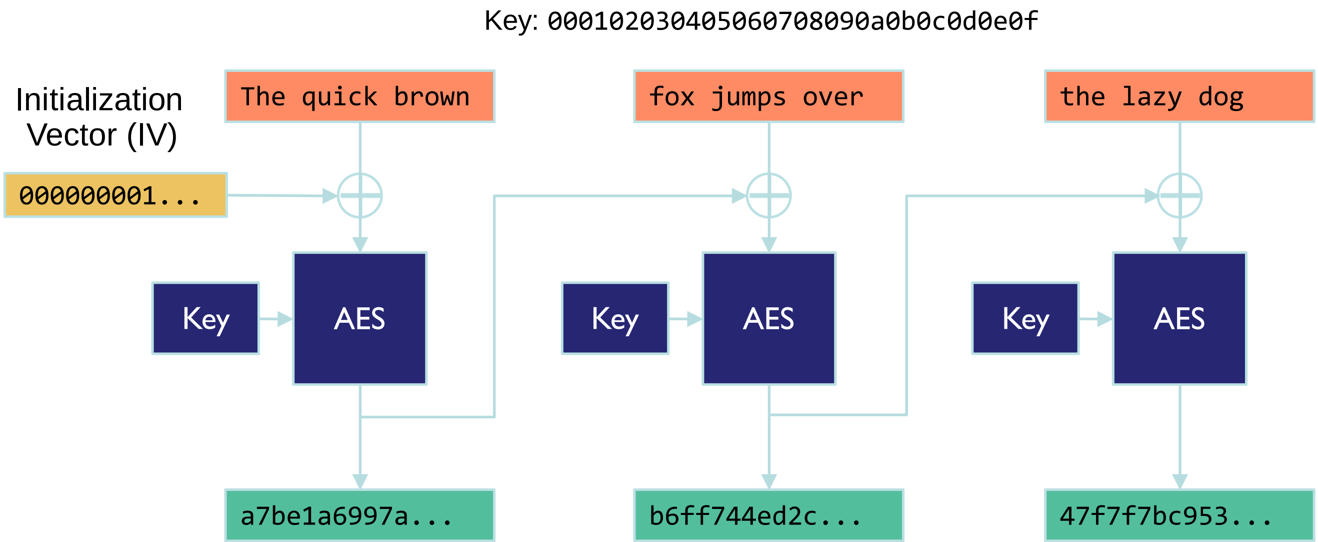
Symmetric Encryption



Asymmetric Encryption

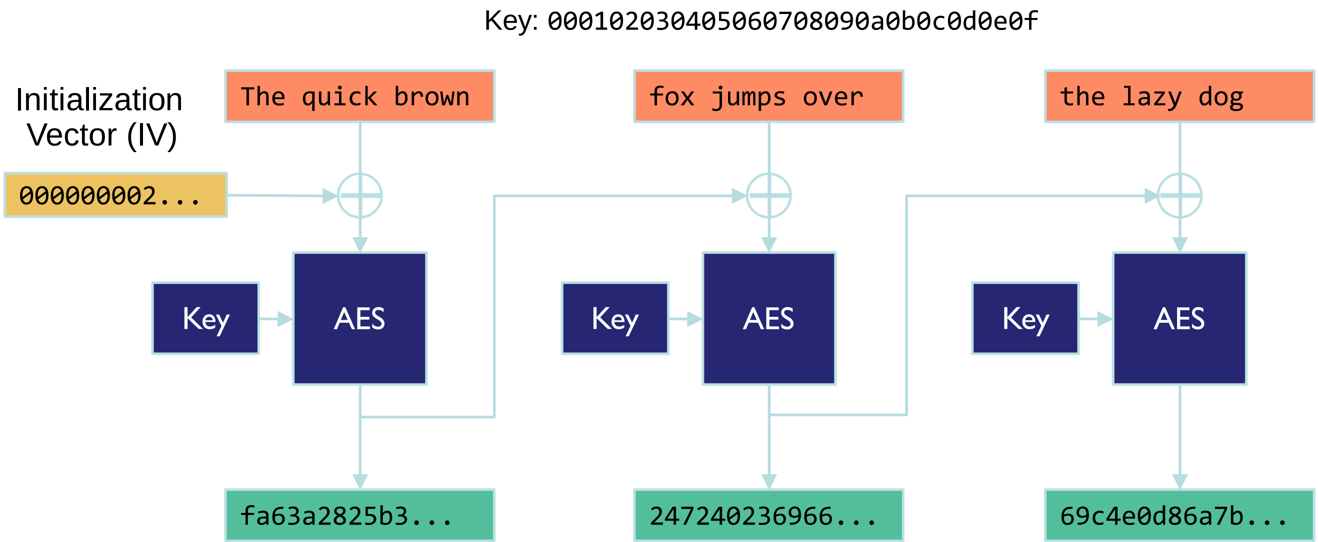


AES + CBC mode



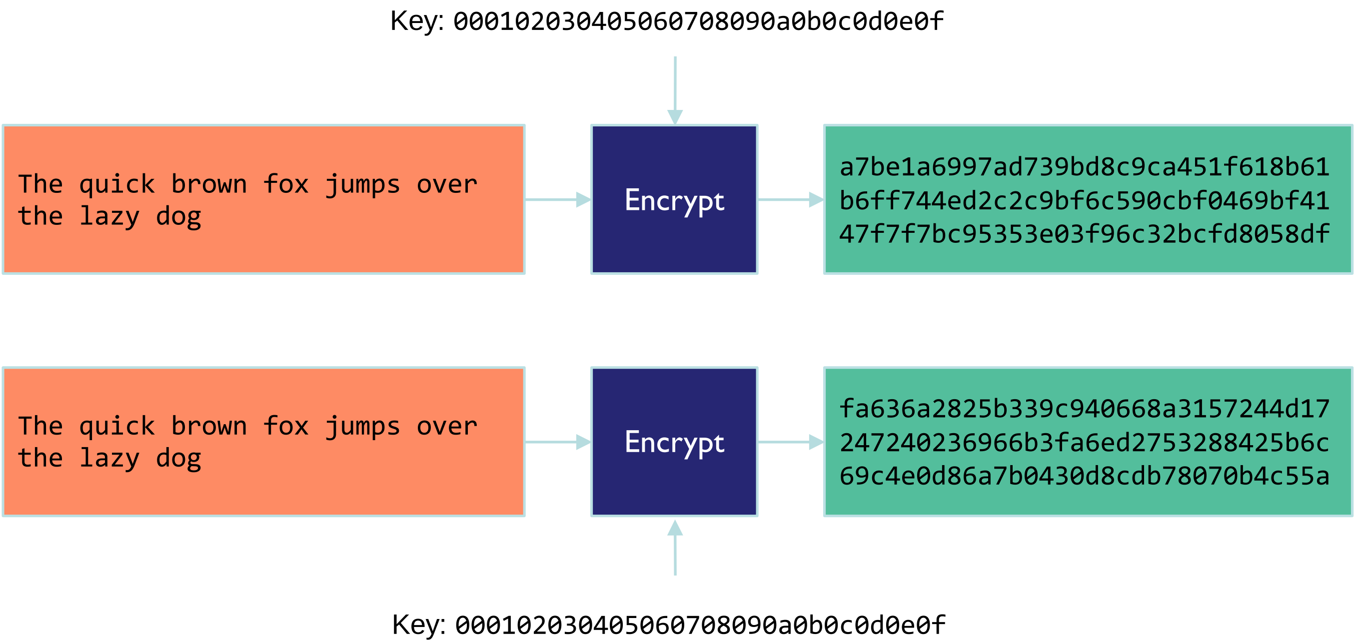
Variable IV => Non-deterministic

AES + CBC mode (IV changes)



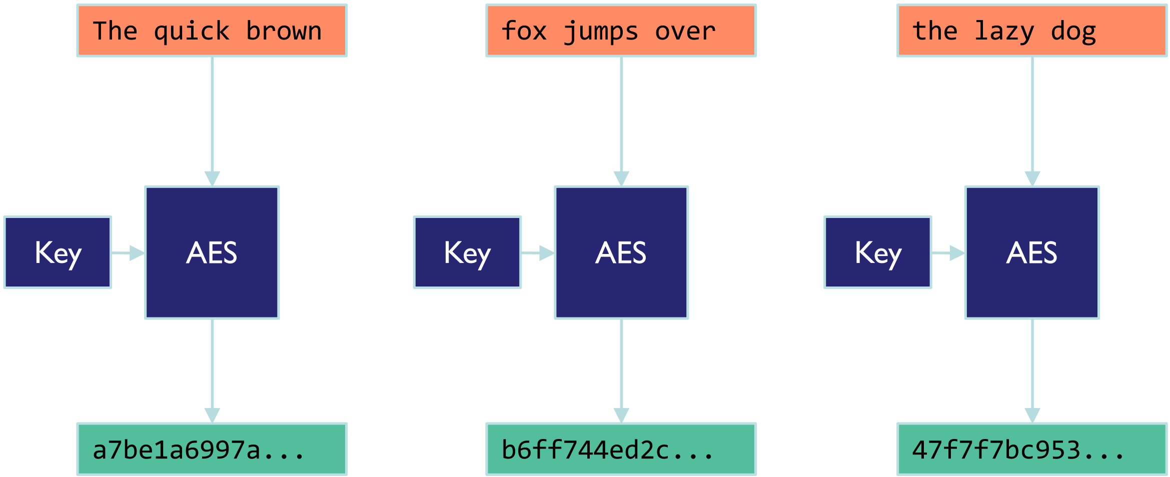
Variable IV => Non-deterministic

Non-deterministic Encryption



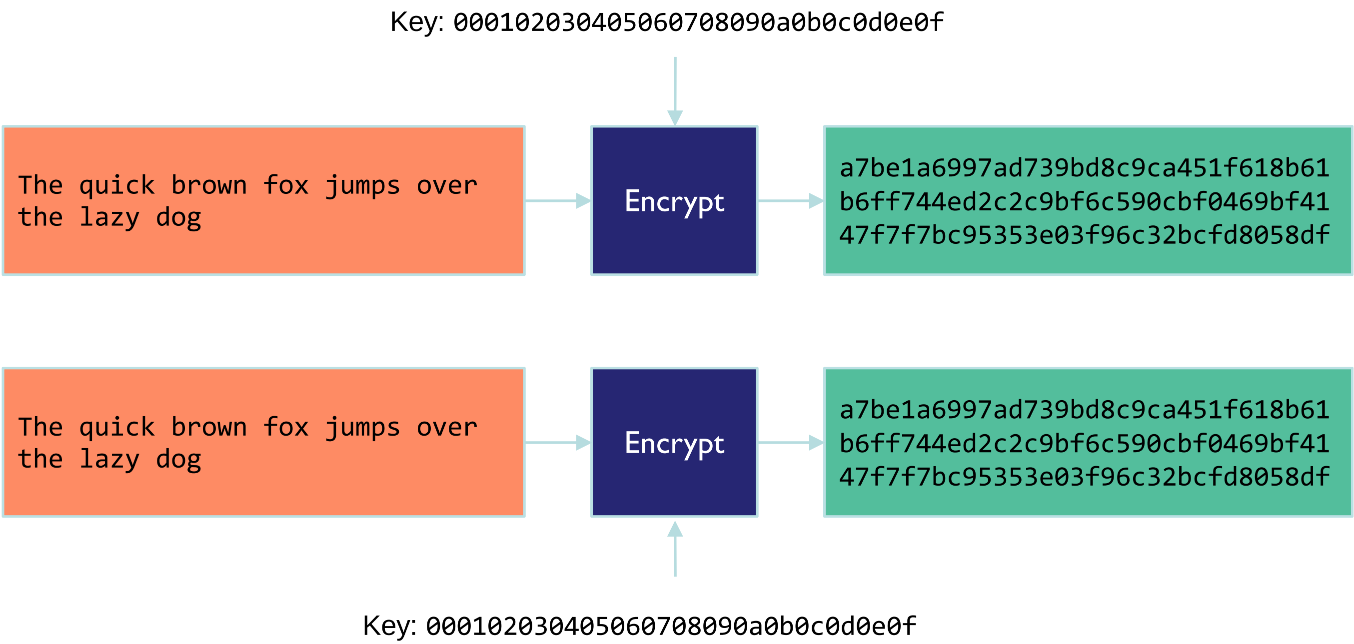
Example: AES + CBC + Variable IV

AES + ECB mode



Key: `000102030405060708090a0b0c0d0e0f`

Deterministic Encryption



Example: AES + ECB

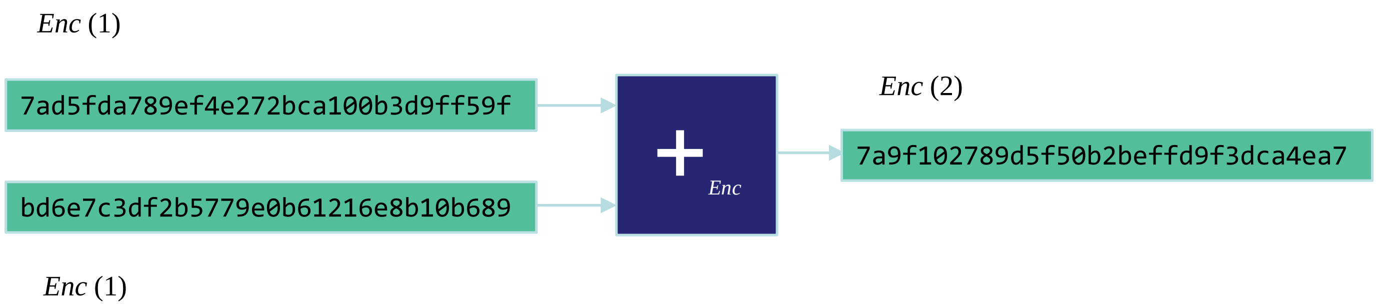
Order Preserving Encryption

Value	Enc (Value)
1	0x0001102789d5f50b2beffd9f3dca4ea7
2	0x0065fda789ef4e272bcf102787a93903
3	0x009b5708e13665a7de14d3d824ca9f15
4	0x04e062ff507458f9be50497656ed654c
5	0x08db34fb1f807678d3f833c2194a759e

$$x\ ;\ y \rightarrow Enc\ x\ ;\ Enc'\,y($$

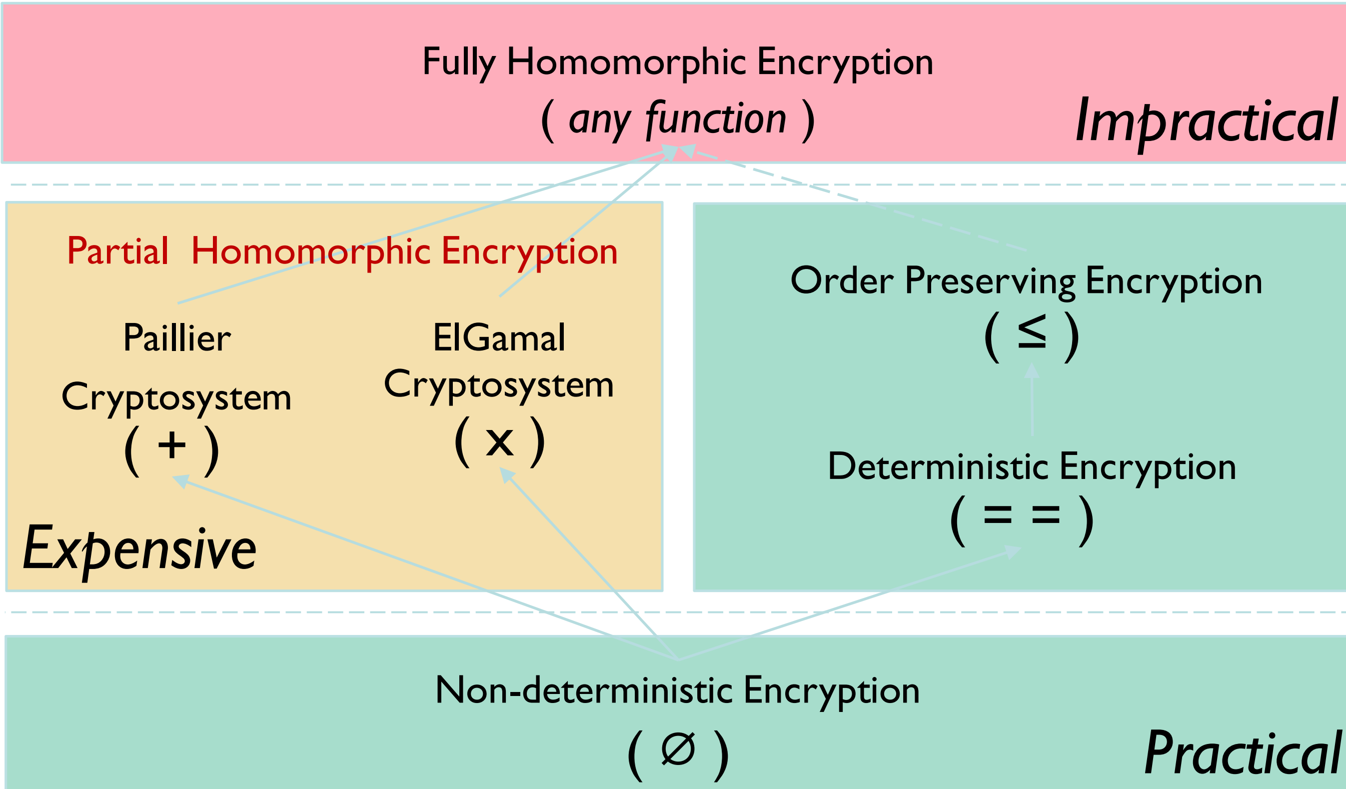
Example: AES + FFX

Homomorphic Encryption



Encryption key is not an input

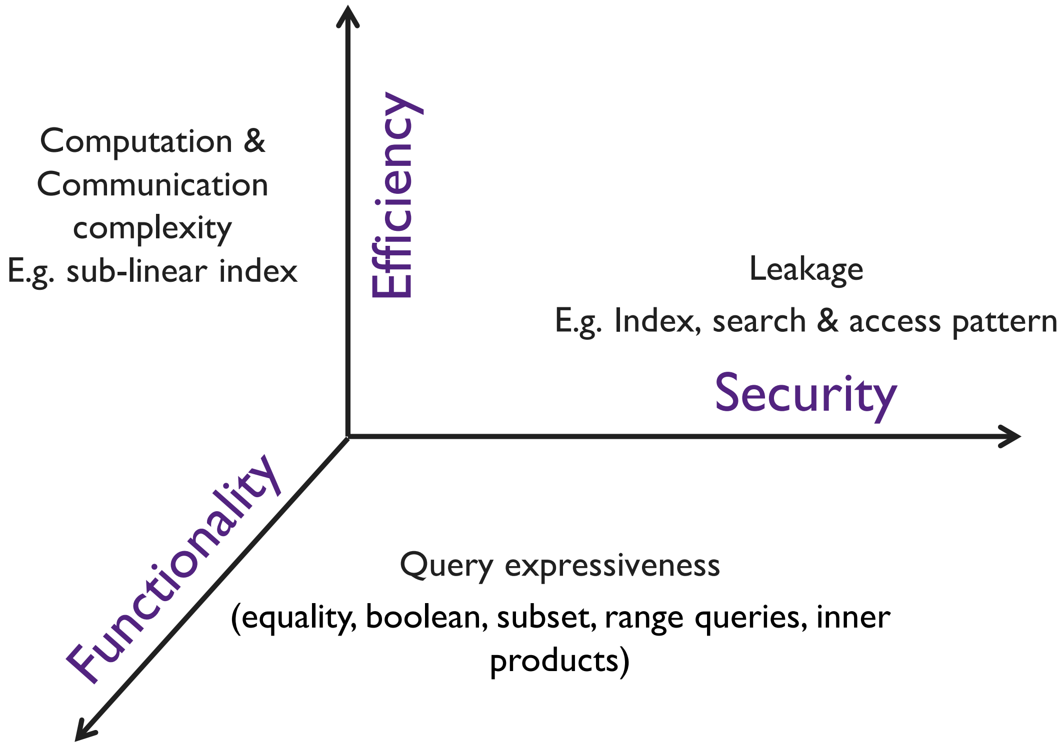
The Spectrum



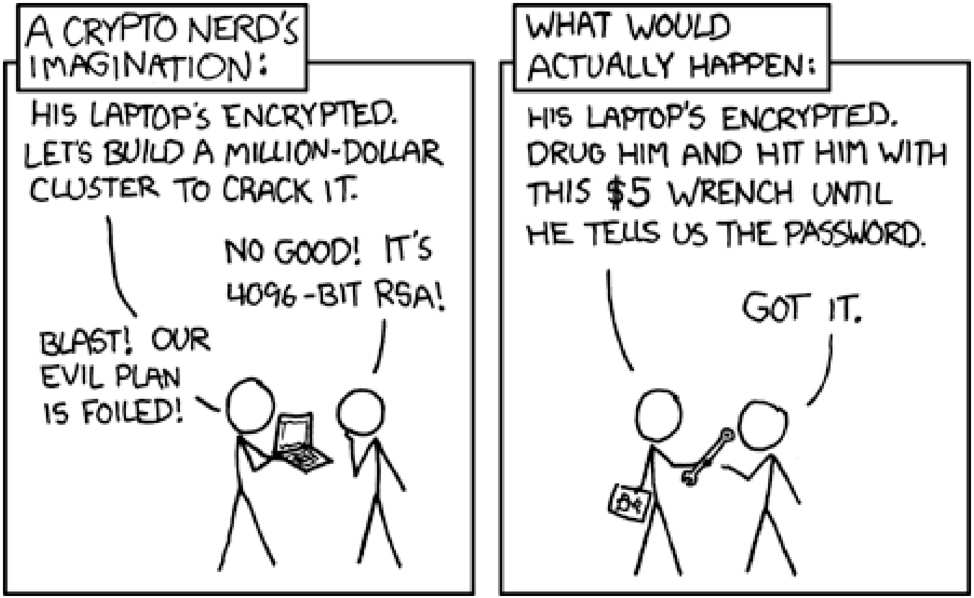
Performance

Scheme	Space for one integer (bits)	Time for one operation
Fully Homomorphic Encryption	2^{14}	Cosmic time scales
Paillier ElGamal	2048	~ ms
Deterministic Encryption	128	~ µs

Encrypted Search – Tradeoffs



Is Encryption == Security?



Leakage Profile

Characteristics	Examples
Objects that leak	Data objects, queries, query response (access control rules)
Type of information leaked	Same value, Matches the intersection of two sets
Which operation leaks	= (say equality) >, < (say, range)
Party that learns the leakage	Provider, Querier, Server

Information leaked by Objects

	Information	Examples
	Structure	String length, set cardinality, tree rep. of object
	Identifiers	Pointers to objects
	Predicates	Additional information, say, a. within a common (known) range b. matches the intersection of 2 clauses within a query
	Equalities	Objects that have same value
	Order (or more)	Numerical/lexicographic ordering of objects, or perhaps even partial plaintext data

Queries on Encrypted Data

Type of data	Type of Queries	Examples
Structured (DBs), Semi (XML/JSON)	Relational Algebra (SQL)	Set (Union, Intersection, Difference, Cartesian product), Selection, Projection, Join
	Associative Arrays (NoSQL)	(Semi-ring): Construction, Find, AA (+, x), AA Element-wise (x)
	Linear Algebra (NewSQL)	Construction, Find, Matrix (+, x), Element-wise (x)
Unstructured	Content-based	Query-by-example, Fuzzy queries → Exhaustive search Eg. filesystems
	Information Retrieval	Indexes
Mixed	SELECT * FROM patient WHERE (age > 40) AND (X-ray CONTAINS "lung cancer")	

Base Queries

	Approach	Description	Examples
●	Legacy	Modifies data insertions and query requests	Property (equality or order) preserving, boolean queries and joins by combining the results of PPE. (CryptDB)
◐	Custom	Special purpose protected indices	Inverted Index, Tree Traversal, Custom indices (Graph)
○	Obliv	Obscures object identifiers (say, pointers)	ORAM

Composed Queries

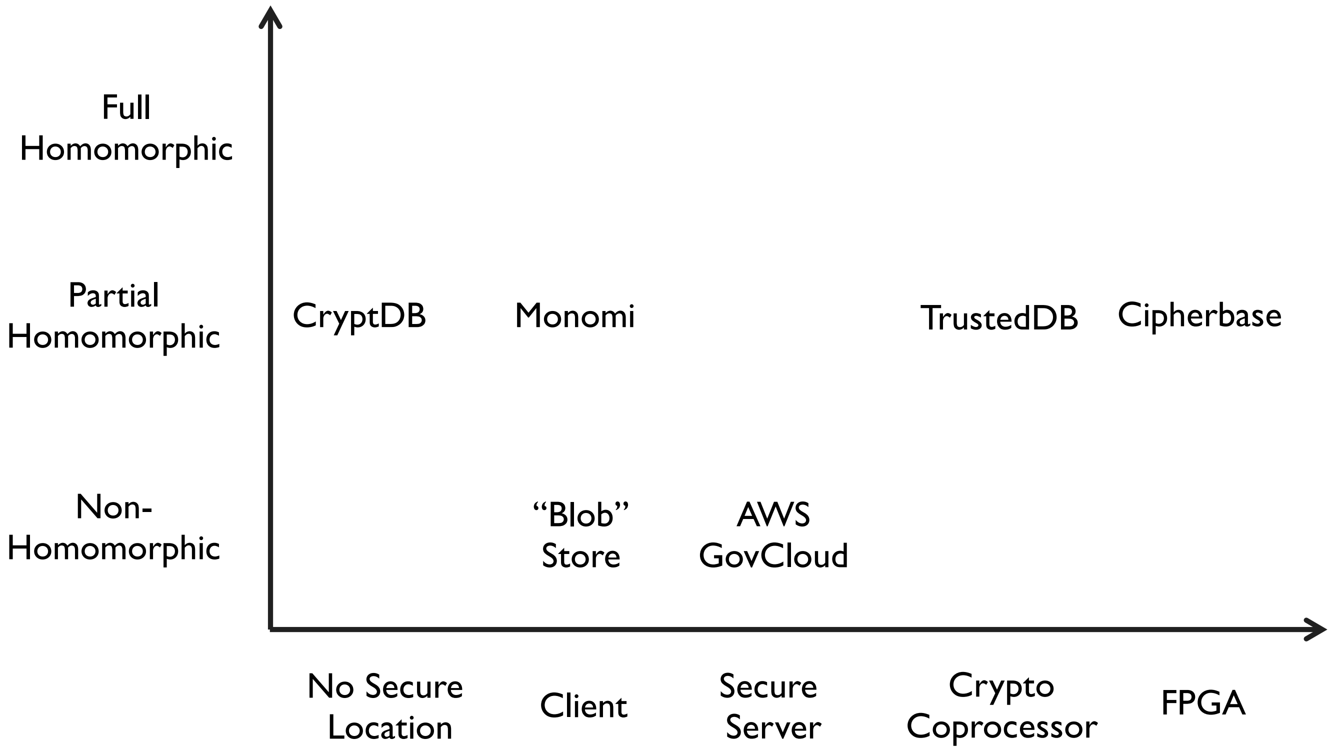
	Composed Query	Base Query Calls	Additional Storage	Leakage
1	Equality (EQ)	1 range	none	Same as range
2	Disjunction (OR) of k EQs (or ranges)	k EQs (or ranges)	None	Identifiers of records matching each clause; ◔
3	Conjunction (AND) of k EQs	1 EQ		Same as EQ
4	Stemming	1 EQ	1	Identifiers of records sharing stem; ◔
5	Proximity	1 EQ	Max. no. of neighbors	No leakage if refresh between queries; ◔
6	Range w/ small domain	$(2 + r)$ EQs	1	No leakage if refresh between queries
7	Range	OR of (2 long m) EQs	$\log m$	Distributional info; ◔
8	Negation	AND of 2 ranges	1	Same as OR of ranges
9	Substring ($s = q$)	1 EQ	$s - n + 1$	Identifiers of records sharing n-grams
10	Substring ($s \leq q$)	1 range	$s - n + 1$	Same as range, on n-grams
11	Anchored substring ($s \geq q$)	AND of $(q - n + 1)$ EQs	$s - n + 1$	$n, k, q,$ ◔
12	Substring	OR of $(s - n + 1)$ ANDs of $(q - n + 1)$ EQs	$s - n + 1$	$n, k, q,$ ◔
13	Anchored Wildcard	AND of $(q - n + 1)$ EQs	$s - n + 1$	$n, k, q,$ ◔
14	Wildcard	OR of $(s - n + 1)$ ANDs of $(q - n + 1)$ EQs	$s - n + 1$	$n, k, q,$ ◔

k – # of clauses in Boolean; r – # of query results; s – max length of the string; q – length of the query string; n – length of the grams

"Anchored" – search at the beginning or the end of the string

◔ If EQ, leaks ≥ ◔

Systems Landscape



Encrypted Databases

- CryptDB
  - Query-aware encryption schemes
    - RND, HE, DET, OPE
  - Architecture
    - SQL-aware encryption
    - Adjustable query-based encryption
    - Chain cryptographic keys in user passwords
  - Supports only 2 out of 22 queries in TPC-H
- Monomi (OLAP)
  - Layout optimizer, Query planner
  - Intermediate results. Ex.: SUM / GROUP BY / HAVING
  - Supports 19 out of 22 queries

Summary

- Application security
  - DBMS is only a part of the overall system stack
- Usability
  - Clients need tools and interpretable security models to navigate security-performance tradeoffs
- Connections to other areas of security
  - Data privacy, access-control, auditing

Thank you!


MORDAC, THE PREVENTER OF INFORMATION SERVICES.
SECURITY IS MORE IMPORTANT THAN USABILITY.
scottadams@aol.com
www.dilbert.com
IN A PERFECT WORLD, NO ONE WOULD BE ABLE TO USE ANYTHING.
©2007 Scott Adams, Inc./Dist. by UFS, Inc.
11-16-07
To complete the log-in procedure, stare directly at the sun.