A SURVEY OF SEARCHABLE ENCRYPTION

Christoph Bösch

ESSA: Workshop on Searchable Encryption
Bertinoro, June 21-24, 2015

ulm university universität uulm

Secure data outsourcing and data sharing

Related:

Functional Encryption (FE)
Predicate Encryption (PE)
Inner Product Encryption (IPE)
Anonymous Identity Based Encryption (AIBE)
Anonymous Hierarchical Identity Based Encryption(AHIBE)
Hidden Vector Encryption (HVE)
Oblivious Ram (ORAM)
Private Information Retrieval (PIR)
Private Searches on Streaming Data (PSS)
Property Preserving Encryption (PPE)
Order Preserving Encryption (OPE)
Fully Homomorphic Encryption (FHE)
...

Today:

SSE + PEKS

	Client		Server
Setup	K ← **Keygen**(s)		
Upload	I ← **BuildIndex**(K,D)	I, Enc(D) →	I, Enc(D)
Search	T ← **Trapdoor**(K,w)	T →	
			{ids} ← **Search**(T,I)
	{D ← Dec(Enc(D))}	← {Enc(D)}	

Architectures:

- **S/S**: Single writer/Single reader
- **S/M**: Single writer/Multi reader

symmetric key primitives

- **M/S**: Multi writer/Single reader
- **M/M**: Multi writer/Multi reader

public key primitives

S/M
M/M

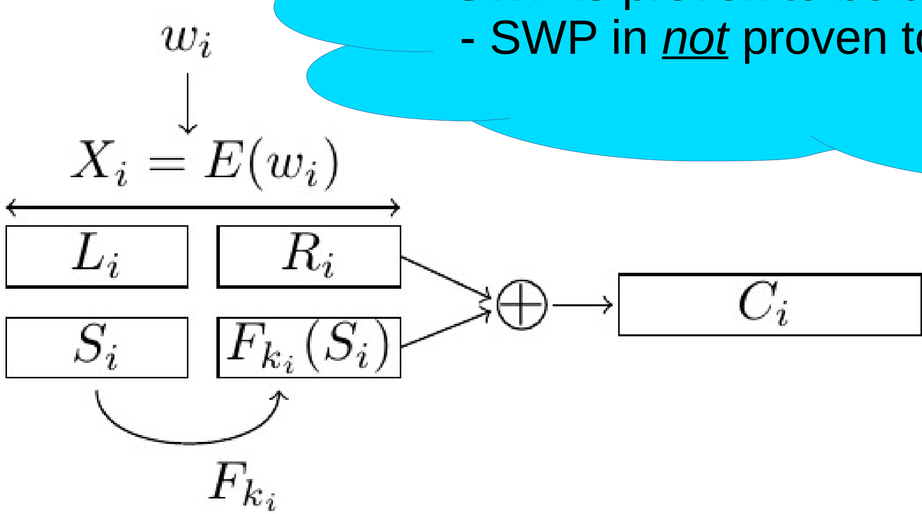
- use some kind of key distribution or user-authentication
- usually introduce a TTP for user-authentication and/or re-encryption of trapdoors

Main Research Directions:

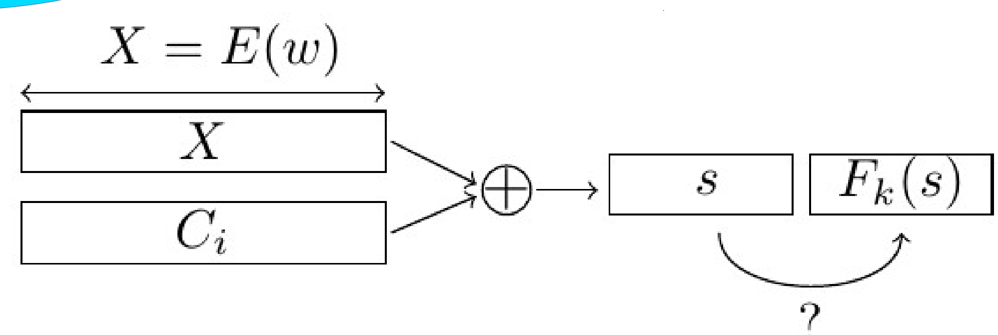
- Efficiency
- Security
- Query expressiveness

	Scheme	Efficiency	Security	Architecture
2000	SWP	O(nm)	IND-CPA	S/S

- SWP is proven to be a secure encryption scheme
- SWP in *not* proven to be a secure SE scheme

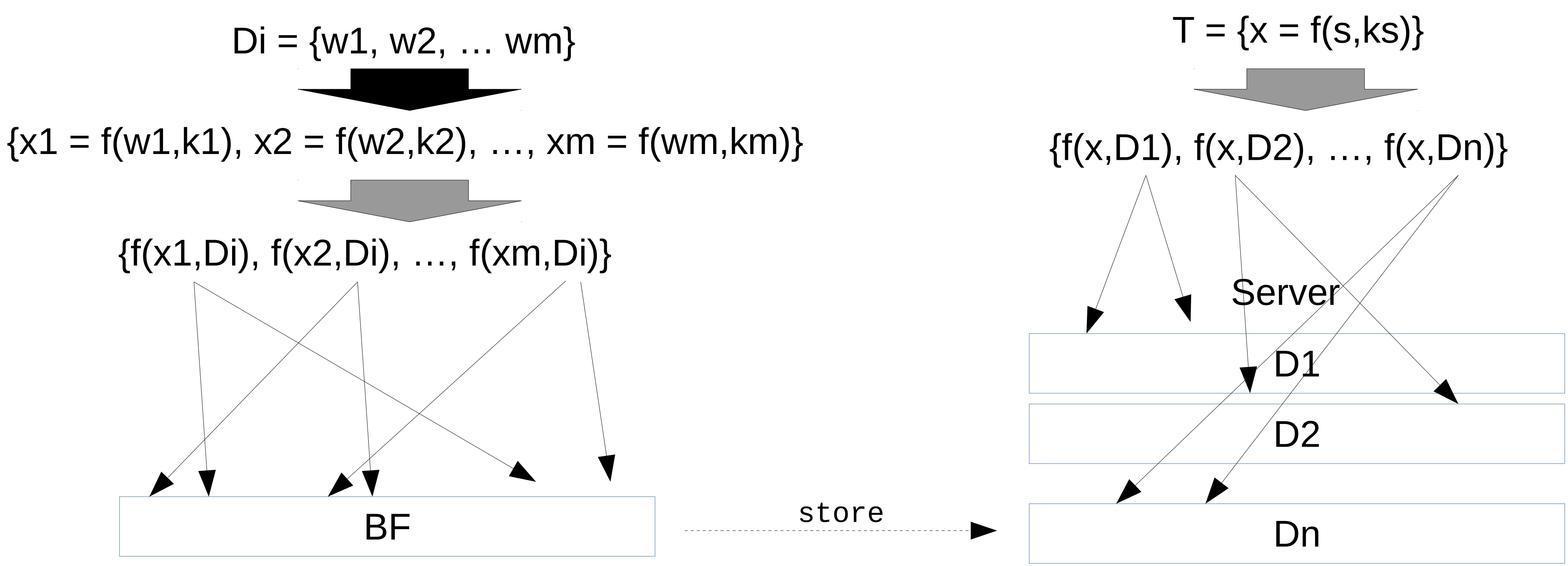


(a) SWP: Encryption



(b) SWP: Sequential search

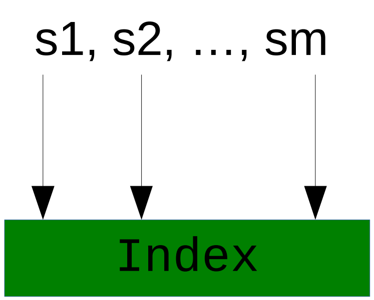
	Scheme	Efficiency	Security	Architecture
2000	SWP	O(nm)	IND-CPA	S/S
2003	Goh	O(n)	IND1-CKA	S/S



	Scheme	Efficiency	Security	Architecture
2000	SWP	O(nm)	IND-CPA	S/S
2003	Goh	O(n)	IND1-CKA	S/S

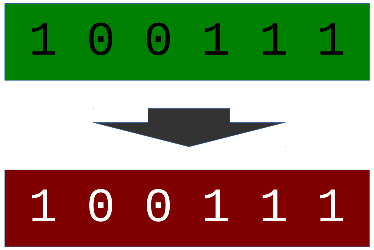
- Game-based security definition
- ***A*** cannot deduce **D** content from its index
- Indexes for **D** of *equal* length are indistinguishable
- Does ***<u>not</u>*** require the trapdoors to be secure

	Scheme	Efficiency	Security	Architecture
2000	SWP	O(nm)	IND-CPA	S/S
2003	Goh	O(n)	IND1-CKA	S/S
2005	CM	O(n)	IND-CKA Goh: IND2-CKA	S/S



D = {w1, w2, …, wm}

iff wi = si
set si to 1



	Scheme	Efficiency	Security	Architecture
2000	SWP	O(nm)	IND-CPA	S/S
2003	Goh	O(n)	IND1-CKA	S/S
2005	CM	O(n)	IND-CKA Goh: IND2-CKA	S/S

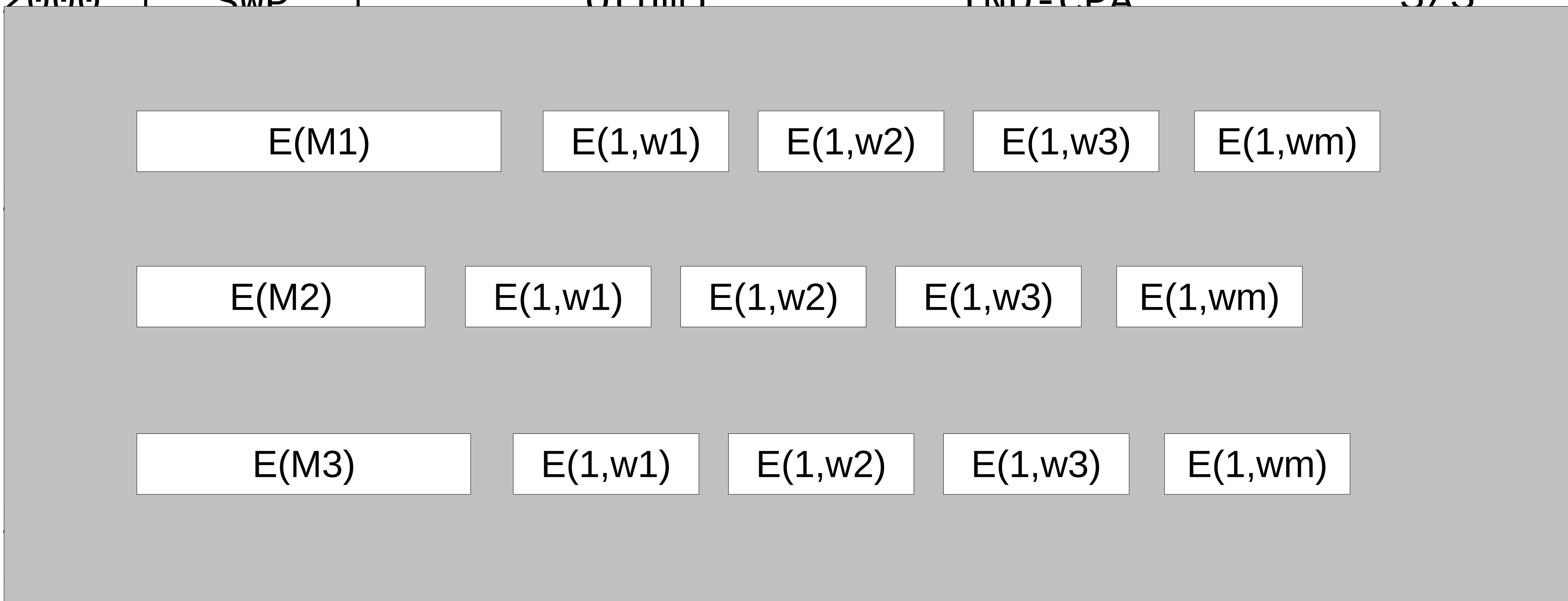
- Simulation-based security definition
- Indexes for **D** of *unequal* length are indistinguishable
- IND-CKA *tries* to capture trapdoor security

- IND2-CKA does *not* require the trapdoors to be secure

- Security of **I** and **T** are *linked*
- *Nothing* should leak except SP + AP
- **T** does not leak information about **k**

	Scheme	Efficiency	Security	Architecture
20			IND-CPA	S/S
2			IND1-CKA	S/S
2005	CM	O(n)	IND-CKA Goh: IND2-CKA	S/S
2006	CGK+	O(D(w))	IND-CKA1 IND-CKA2	S/S S/M

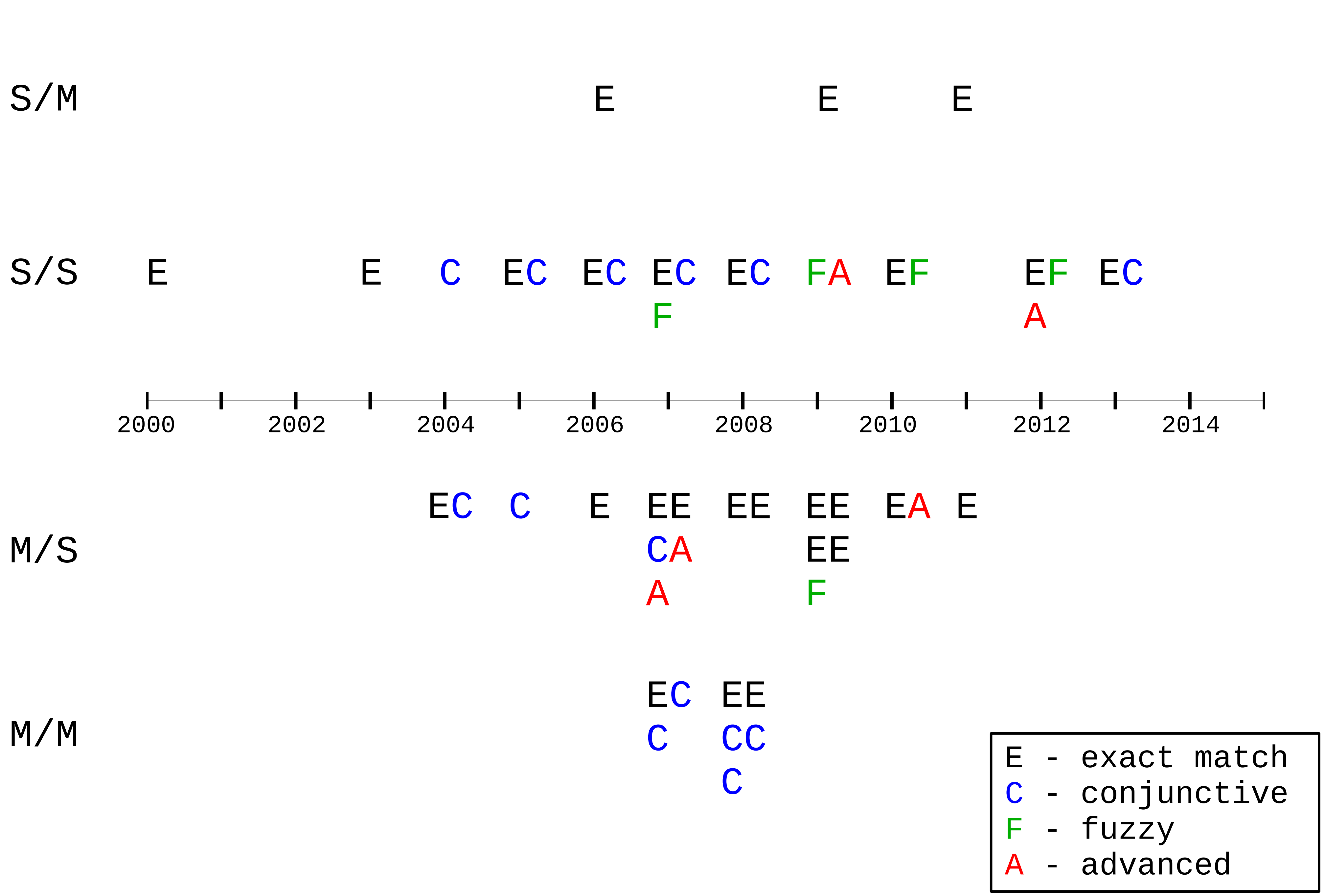
	Scheme	Efficiency	Security	Architecture
2000	SWP	O(nm)	IND-CPA	S/S
			IND-CKA2	S/M
2004	BCO+ (PEKS)	O(nm)	PK-CKA2	M/S



	Scheme	Efficiency	Security	Architecture
2000	SWP	O(nm)	IND-CPA	S/S
20			IND1-CKA	S/S
			IND-CKA Goh: IND2-CKA	S/S
2006	CGK+	O(D(w))	IND-CKA1 IND-CKA2	S/S S/M
2004	BCO+ (PEKS)	O(nm)	PK-CKA2	M/S

- Trapdoors *not* secure due to public key
  - No info about **k** is leaked from **I**, unless T(k) is available

	Scheme	Efficiency	Security	Architecture
2000	SWP	O(nm)	IND-CPA	S/S
2003	Goh	O(n)	IND1-CKA	S/S
2005	CM	O(n)	IND-CKA Goh: IND2-CKA	S/S
2006	CGK+	O(D(w))	IND-CKA1 IND-CKA2	S/S S/M
2004	BCO+ (PEKS)	O(nm)	PK-CKA2	M/S


S/M
E
E
E
S/S
E
E
C
EC
EC
EC
F
EC
FA
EF
EF
A
EC
2000
2002
2004
2006
2008
2010
2012
2014
M/S
EC
C
E
EE
CA
A
EE
EE
EE
F
EA
E
M/M
EC
C
EE
CC
C
E - exact match
C - conjunctive
F - fuzzy
A - advanced

S/S:

- O(nm), O(n), O(m), O(log m), O(|D(w)|)

S/M:

- O(nm), O(n)

document id	keywords
1	w_2, w_5, w_7
2	w_1, w_2, w_4, w_6, w_8
...	...
n	w_2, w_5, w_6

(a) Forward index.

keyword	document ids
w_1	2, 3, 9
w_2	1, 2, 6, 7, n
...	...
w_m	1, 3, 8

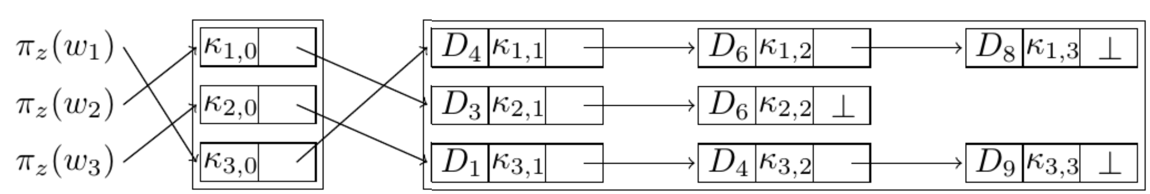
(b) Inverted index.

M/S:

- O(nm), O(n)

M/M:

- O(nm), O(n)



Curtmola et al. - CGK-I

- Sub-linear/optimal schemes achieving IND-CKA2 security *exist only* in the S/S setting

- S/M, M/S, and M/M schemes achieving IND/PK-CKA2 security are *inefficient* (linear in the number of data items)

- Data representation plays a big role for efficiency

<u>Practical efficiency:</u>

- Only seven papers provide implementations and performance numbers

- Most implementations are not publicly available

- Hard to compare efficiency of schemes, due to a wide spectrum of different application scenarios and threat models

- Interactive protocols can achieve higher security and/or practical efficiency

Privacy Issues:

- Index Information
- Search Pattern
- Access Pattern

Security Definitions:

	Leaks:
• IND-CPA, IND1-CKA, IND2-CKA, PK-CKA2	Trapdoor, SP + AP
• IND-CKA1, IND-CKA2, UC-CKA2	SP + AP
• FS	AP
• FS+? (Oblivious Data Structures)	-

- Most schemes use their own security definition (tailored IND-CKA2)
- Depending on application: SP leakage OK or not.
- Same for AP, e.g. DNA data

IND-CKA2: widespread acceptance as a strong notion of security in the context of SE. Leaks *search* and *access* patterns.

Full Security: new security notion in the context of SE. Leaks only the *access* pattern.

What about the access pattern, e.g., in DNA databases?

Hard to assess and/or compare due to:

- different security models
- different assumptions
- different scenarios

Single Equality:

- S/*: Sequential scan, Database search (deterministic)
- M/*: (A)IBE, HIBE

Conjunctive:

- Shamir's Secret Sharing

Similar/Fuzzy:

- Character wise encryption + Hamming distance, LSH + BF, Pre-computed sets

Other:

- Inner product
- HVE

Architecture	S/S	S/M	M/S	M/M
Equality	✓	✓	✓	✓
Conjunction	✓	-	✓	✓
Comparison	-	-	✓	-
Subset	(✓)	-	✓	-
Range	(✓)	-	✓	-
Wildcard	-	-	✓	-
Similar/Fuzzy	✓	-	-	-
Inner Product	✓	-	(✓)	-
# of schemes	28	2	19	9

Sub-linear: only equality and conjunctive (S/S)

<u>Active research in all three directions:</u>

i) efficiency

ii) security

iii) expressiveness

<u>Trade-offs:</u>

i) security vs. efficiency

ii) security vs. expressiveness

iii) efficiency vs. expressiveness

Scheme	Expressiveness	Efficiency	Security
KO	exact match	linear	UC-CKA2
CGK+	exact match	optimal	IND-CKA1
CJJ+	conjunctive	sub/interactive	IND-CKA2
SSW	inner product	linear	FS

General:

- Requirements dependent on setting/application:
    - Leakage
    - Expressiveness
    - Computation
    - Communication
    - Interactiveness

Expressiveness:

- Only *single* and *conjunctive* searches in *sub-linear* time
- We need more expressive schemes

Security:

- IND-CKA2: leak SP + AP
- Fully secure: protect SP
- What about the AP?

Efficiency:

- sub-linear/optimal search time + IND-CKA2 only in S/S setting
- S/M, M/S, and M/M: linear in the number of data items