Recent approaches for DB security

Ji Won Yoon
Signal Processing and Intelligence (SPI) Lab,
CIST, Korea University
E-mail: jiwon_yoon@korea.ac.kr
web: https://sites.google.com/site/securesiplab/

Outline

1. Introduction to DB security
2. Conventional concept for DB security
3. Recent advanced approaches for DB Security
   - Order Preserving Encryption
   - Format Preserving Encryption
   - Honey Encryption(HE)
   - Homomorphic Encryption(HE)
   - Conclusion

Nowadays, we are in ..

One of the important issues in ICT and Database is DB security which is based on encryption!!

Accidents in Korea, (Jan.~ May), 2014 (1/4)

< 2014년 상반기 개인정보 유출 사건 분석 >

○ 출처 : 보안뉴스

○ 조사기간 : 2014. 1.1 ~ 5.9

구분	보도 시점	유출기관 및 기업명	유출규모	유출항목	유출경로	사고발생원인및 공격방법
1	1월8일	KB국민,롯데,NH농협 3사카드사	KB국민카드약5300만건, NH카드약약12500만건, 롯데카드약2600만건총1억400만건	성명, 주민번호, 여권번호, 이메일, 전화번호, 주소, 직장, 결혼여부, 자가용보유여부, 카드발급정보, 카드번호, 유효기간, 결제계좌, 신용한도금액, 이용실적, 연소득, 연체금액, 타사카드 보유현황 등 18개 항목	KCB협력사내부직원-> 대출광고업자와대출모집인에게판매	부정방지사용 시스템인 FDS 시스템 구축 용역을 맡은 KCB 직원이 3개 카드사의 고객정보를 USB에 담아 반출
2	1월19일	국민,신한,하나, 우리은행등카드사 연계은행	약1500만명유출	성명,휴대전화번호,직장번화번호,자택전화번호,주민번호,직장주소,자택주소,직장정보,주거상황,이용실적금액,결제계좌,결제일,신용한도금액,결혼여부,자가용보유유무,신용등급 등총19개 항목		1억 580만명 카드사 고객정보 유출 과정에서 연계은행의 고객정보도 함께 유출사실 확인 돼
3	2월26일	대한의사협회,치과의사협회, 한의사 협회	의사협회 8만명, 치과의사협회 5만6천명, 한의사 협회 2만명 등 총15만 6천명	성명, 주민번호, 휴대전화번호, 주소, 의사면허번호, 근무지, 졸업학교 등	홈페이지 해킹->주민등록번호와 계좌번호 등 개인정보를 빼내 대출업자 등에게 판매	악성코드를 사이트에 심어 관리자 권한을 획득해 웹셸 방식으로 3개 협회 홈페이지 해킹
4	3월6일	KT	1천200만명	이름, 주소, 주민번호, 전화번호, 이메일, 신용카드번호, 카드유효기간, 은행계좌번호, 고객관리번호, 유심카드번호, 서비스가입정보, 요금제 관련정보 등 12개 항목	홈페이지 해킹	텔레마케팅 업체 측에 고용된 해커가 자신의 ID로 KT 홈페이지에 로그인->파로스 해킹 프로그램을 이용해 개인정보 수집->텔레마케팅 업체에 제공

NIA 한국정보화진흥원

cited from NIA , 2014

Accidents in Korea, (Jan.~ May), 2014 (2/4)

5	3월7일	티켓몬스터및225개사이트해킹	티켓몬스터 113만명 등 1700만명	이름, 아이디, 성별, 생년월일, 휴대전화번호, 이메일, 배송지 전화번호 및 주소 등	홈페이지 해킹	홈페이지 게시판 등에 악성코드의 일종인 '웹셸'심어 개인정보 유출
6	3월11일	SK브로드밴드,LG유플러스,여행사,인터넷 쇼핑몰등	문씨가컴퓨터에보관중인 LG유플러스와 SKT, KT 정보 420만건, 금융기관 11곳 등에서 유출된 것으로 보이는 정보100만건, 여행사와 인터넷 쇼핑몰 업체에서 유출된 187만건 등 총1230만 건	이름, 휴대전화번호, 주소, 요금결제 계좌번호, 나이, 성별, 거주지, 직업 등	통신사 하부 판매점에서 유출	열악한 하부 대리점의 취약점 이용해 개인정보 유출 또는 딜러들의 불법 수집 및 유통 가능성/가공데이터
7	3월16일	재향군인회	1만3900여명	이름,아이디,패스워드,이메일,전화번호,회사 전화번호,핸드폰번호 등	홈페이지 해킹	SQL인젝션 취약점
8	3월24일	한화생명, 알리안츠생명, PCA생명, AIA생명, 동부생명, KDB생명, 미래에셋생명, 동부화재, LIG손해보험, 한화손해보험 등 14개 보험사	1만3000여건	이름,주민번호,전화번호,이메일주소,대출금액,대출승인여부등보험계약정보 등	판매대리점의 정보유출 및 개인정보 불법유통	중국에서 개인정보 1105만건 매입 매입->대부중개업자에게 판매 및 도박 성인사이트 광고에 이용
9	3월17일	CJ대한통운		이름, 주소, 전화번호 등	심부름센터 업주&택배기사 총 382차례 개인정보 유출	심부름센터측에서택배기사에게 의뢰->택배프로그램통해개인 정보수집

cited from NIA , 2014

Accidents in Korea, (Jan.~ May), 2014 (3/4)

10	3월20일	CJ대한통운	1000여건(3000여명)	임직원들의직급과직책,휴대전화번호등 개인정보불법수집	내부 간부	경쟁회사의임직원이름과직책, 휴대전화번호등 개인정보1천여건을불법수집
11	3월25일	네이버	1억건(2500만명:주민번호 기준으로 중복제외)	이름, 아이디, 패스워드, 주민번호,	로그인 체크기 및 악성프로그램 22종 개발해 해킹	유효 계정 추출->카페 가입->카페 회원 명단 추출->쪽지발송
12	4월1일	KB국민, NH농협, 롯데 3개카드사 2차 유출	국민카드 14만명, 농협카드 3만5천명 총17만5000여명	이름, 주민번호, 전화번호, 직장명 등		
13	4월5일	BBQ	51만건	회원구분에따라ID,암호화된비밀번호,이메일주소,성명,실명인증값(아이핀회원은아이핀번호),생년월일,성별,아이디,비밀번호,별명,연락처(메일주소,휴대폰번호중택일)만14세미만은법정대리인정보,가입인증정보 IP Address, 쿠키, 방문 일시, 서비스 이용 기록, 불량 이용 기록 등 3개에서 총19개 항목	해킹	개인정보 암호화 조치 안됨
14	4월10일	국방과학연구소		이름, 아이디, 주민번호, 비밀번호, 휴대번호, 주소 등	해킹	중앙서버 악성코드 또는 악성프로그램 침투, 내부PC 300대이상 및 서버 장악->구글 검색 통해 노출->
15	4월11일	포스단말기 해킹(신한, 국민, 농협, BC 등 카드사 10곳, 기업은행, 씨티은행)	20만건	이름, 전화번호, 카드번호, 카드 유효기간, OK캐시백 포인트 카드 비밀번호	해킹	포스단말기 관리업체 악성코드 유포->포스 가맹점 악성코드 감염->카드정보 유출 및 전용서버를 통한 카드정보 수집->카드정보 획득->국내 및 국외 카드정보 판매->카드위조 및 현금인출

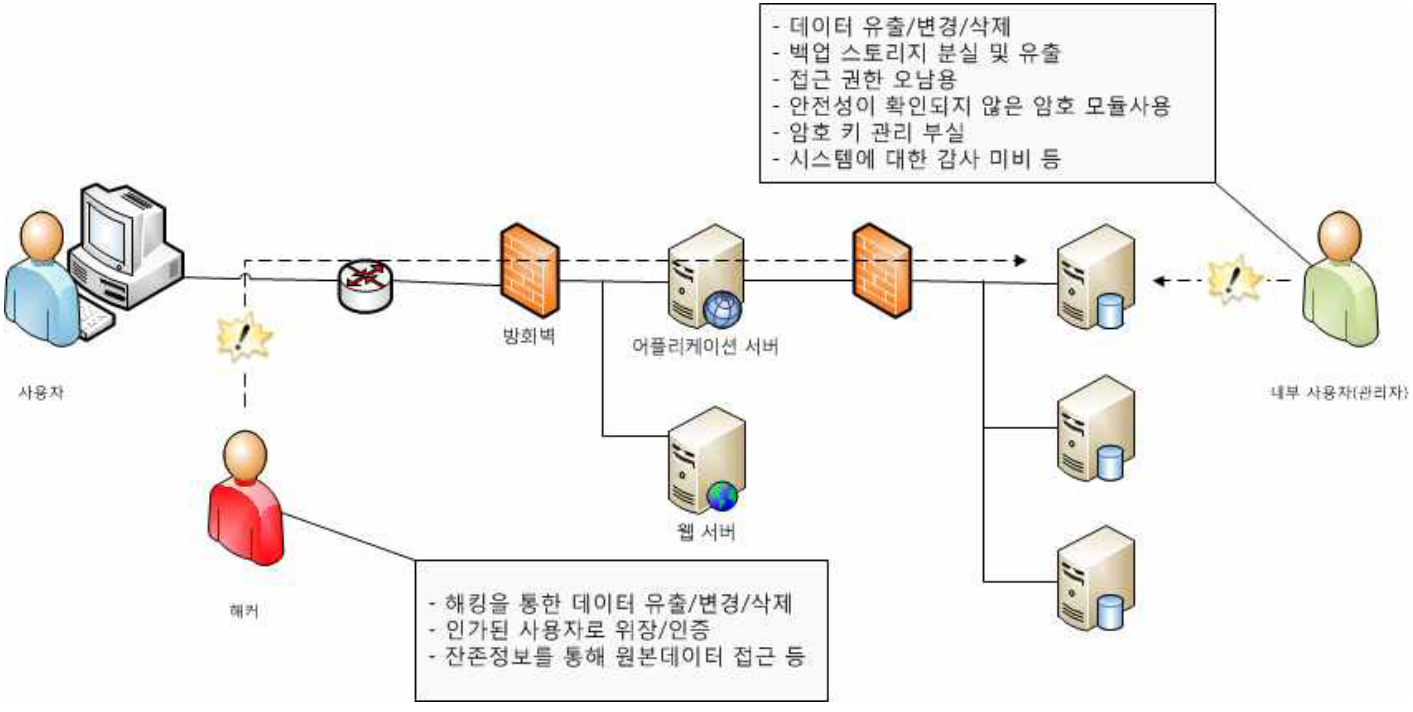
cited from NIA , 2014

Accidents in Korea, (Jan.~ May), 2014 (4/4)

16	4월11일	삼성 그룹 전 현직 직원	전 현직 직원5000여명	이름, 주민번호, 출신학교 등	삼성전기 협력업체 직원에 의한 개인정보 유출	삼성전기 내부 로그분석 점검->협력업체 직원 유출 확인->협력업체 직원이 개설한 사이트에 개제
17	4월 13일	천재교육	350만명 이하	이름, ID, 비밀번호, 주민번호, 이메일, 주소, 상세주소, 집전화번호, 휴대전화번호 등 총9개 항목	서버 해킹 추정	해당 서버 제출
18	4월14일	KDB생명	10,695개	통화내용 녹음된 음성파일(팩스번호, 휴대폰번호, 보험증권 번호 등 포함)	상담내용 웹사이트에 공개	저장서버 관리 미흡
19	4월14일	IBK캐피탈, 씨티캐피탈	3만4000여명		불법대출업자 이동식 저장장치에 저장	씨티은행, 스탠다드차타드은행의 고객정보 유출 사건 추가 수사 과정에서 정보유출혐의
20	4월14일	전국실종가족찾기		집 전화번호, 휴대전화번호, 아이디, 이름 등	회원정보 노출	구글에 관리자 페이지 노출
21	4월16일	농협생명	35만건	이름, 주민번호	외주업체직원	농협 생명이 프로젝트 업무 수행을 위해 개인정보 제공->외주업체 직원 컴퓨터에 보관 등->자체점검기간 중 모두 삭제->개인정보 부실관리
22	4월16일	스킨푸드	55만건	이름, 주민번호, 전화번호, 휴대전화번호, 주소, 이메일주소, 아이디, 비밀번호, 가입일 등 2010년 10월8일 이전 홈페이지 회원가입 이용자 개인정보	홈페이지 해킹	
23	5월9일	토니모리	50만명	아이디, 이름, 휴대전화 번호, 비밀번호, 이메일 등	홈페이지 해킹	
24	5월9일	두원공과대학교	학생및 교수130명	주민등록번호, 여권번호, 여권만료일, 전화번호 등		구글 검색통해 개인정보 노출

cited from NIA , 2014

Threats in DB



cited from 'Recent trend and security analysis of DB cryptography', Financial Security Agency, Sept., 2012

Threats in DB

구분		보안위협	설명
외부	내부		
○		웹 보안위협	외부의 인가되지 않은 사용자가 SQL Injection 공격 또는 File Upload 후 웹쉘 실행 등을 통한 불법적인 정보 획득
○	○	약한 식별 및 인증	정당한 사용자 신원을 획득하기 위해 반복적인 인증 시도 및 사회공학적 기법 등을 이용하여 인가된 사용자 신원 획득
○	○	데이터 유출	암호화되지 않은 데이터의 유출 및 암호화된 데이터의 암호 해독을 통한 불법적인 정보 획득
	○	권한 오·남용	접근 권한보다 더 많은 권한을 획득하여 권한을 남용하거나, 정당한 권한을 가진 사용자가 허가되지 않은 작업을 수행
○	○	암호 모듈 오용	안정성이 확인 되지 않은 암호 모듈 사용, 적합하지 않은 암호 모드를 사용하여 암호문 해독
	○	약한 감사	약한 감사 정책으로 인한 제한된 정보의 기록으로 위협에 대한 탐지, 추적, 복구의 어려움
○	○	잔여 정보 노출	DBMS 로그 등의 잔여 정보를 획득하여 데이터 획득 및 유추
○	○	암·복호화 키 및 마스터키 노출	안전하지 않은 암·복호화 키 관리로 키가 노출되어 암호문 해독

cited from ‘Recent trend and security analysis of DB cryptography‘, Financial Security Agency, Sept., 2012

Conventional cryptography skills

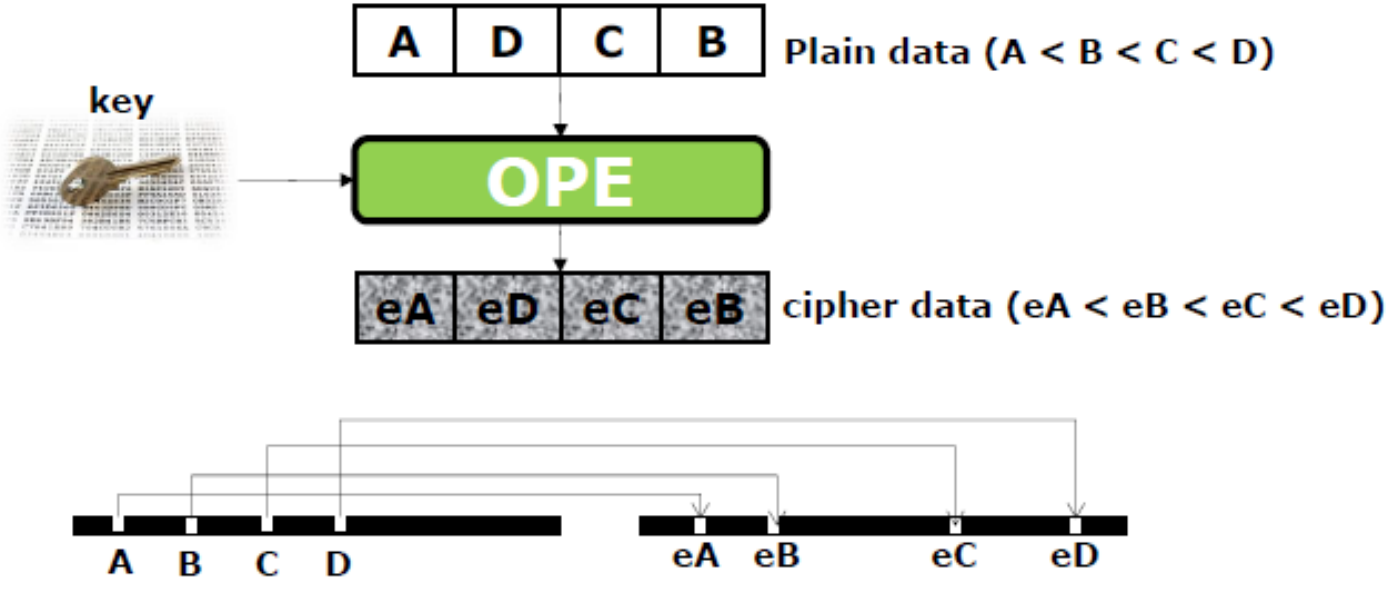
- AES and so on..

Outline

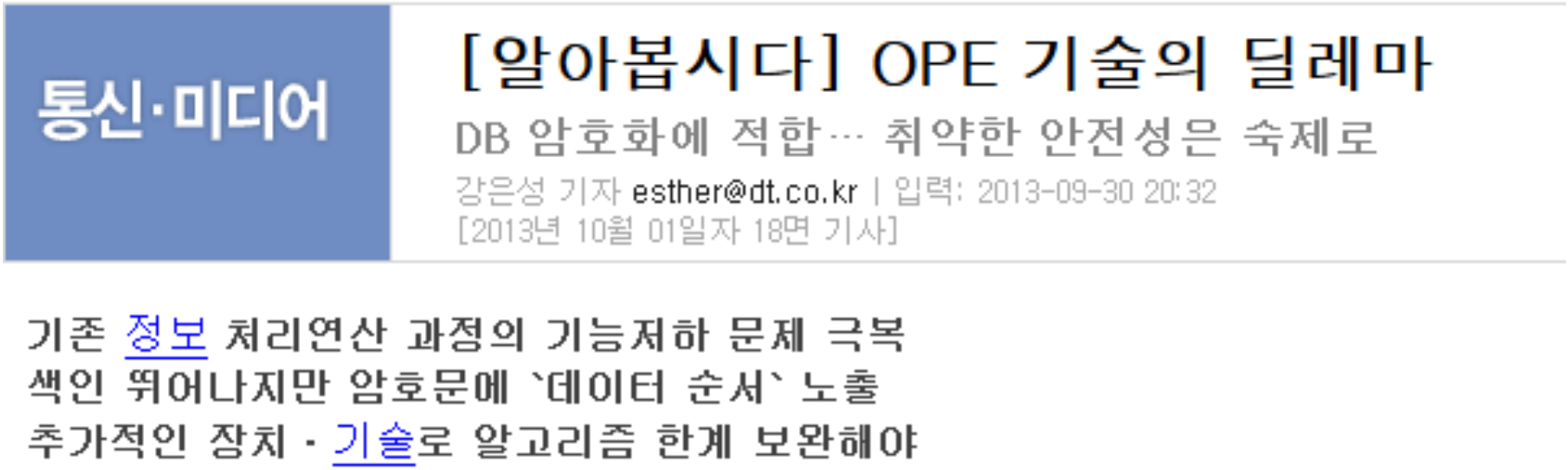
Introduction to the Order Preserving Encryption (OPE)

<Most slides from a students JaeYeol Jeong, Korea University>

It is a symmetric encryption over the integers such that cipher-texts preserve the numerical orders of the corresponding plain-texts.



Introduction to the Order Preserving Encryption (OPE)



통신·미디어

[알아봅시다] OPE 기술의 딜레마

DB 암호화에 적합… 취약한 안전성은 숙제로

강은성 기자 esther@dt.co.kr | 입력: 2013-09-30 20:32

[2013년 10월 01일자 18면 기사]

기존 정보 처리연산 과정의 기능저하 문제 극복
색인 뛰어나지만 암호문에 `데이터 순서` 노출
추가적인 장치 · 기술로 알고리즘 한계 보완해야

- In the conventional approach, all encrypted data should be decrypted before searching→ Extremely time consuming
- Good point of OPE: extremely fast searching
- Bad point of OPE: Encryption is slightly slower and more complicated and security can be relatively weaker .

Introduction to the Order Preserving Encryption (OPE)

ID	연봉	주민번호
1	30000000	8511...
2	60000000	7612...
3	40000000	8601...
4	18000000	9003...
5	1 30000000	5409...
...	...	...
25154	50000000	8008...

General DB (without encryption)

Introduction to the Order Preserving Encryption (OPE)

<Conventional Encryption algorithms>

ID	연봉	주민번호
1	30000000	8511...
2	60000000	7612...
3	40000000	8601...
4	18000000	9003...
5	1 30000000	5409...
...	...	...
25154	50000000	8008...

encryption

→

ID	연봉	주민번호
1	dlskfdkf	Z9ooewjf
2	Woicziod	lujmw4
3	Apzie.d	Puk jklh
4	.sozirvlz	[lsxdkix.vc
5	sd234,n,v	783wm,4
...	...	...
25154	zpeijfgnsa	iougbkla

Encrypted DB via general encryption scheme

(AES or other algorithms)

- Unfortunately, we cannot do many useful operations in SQL including searching, join and so on.

Introduction to the Order Preserving Encryption (OPE)

ID	연봉	주민번호
1	dlskfdkf	Z9ooewjf
2	Woicziod	lujmw4
3	Apzie.d	Puk jklh
4	.sozirvlz	[lsxdkix.vc
5	sd234,n,v	783wm,4
...	...	...
25154	zpeijfgnsa	iougbkla

Not ordered

(AES or other algorithms)

ID	연봉	주민번호
1	A239w8df	3f5dfg
2	A23fdfsde	3h9464
3	B8cxudlf	Cqdez;d
4	F1z93msf	Rw4x;lsfga
5	Ffsdfsedf	Uz89ewf
...	...	...
25154	Zsderfg	Ysasre223

Ordered

(OPE based algorithms)

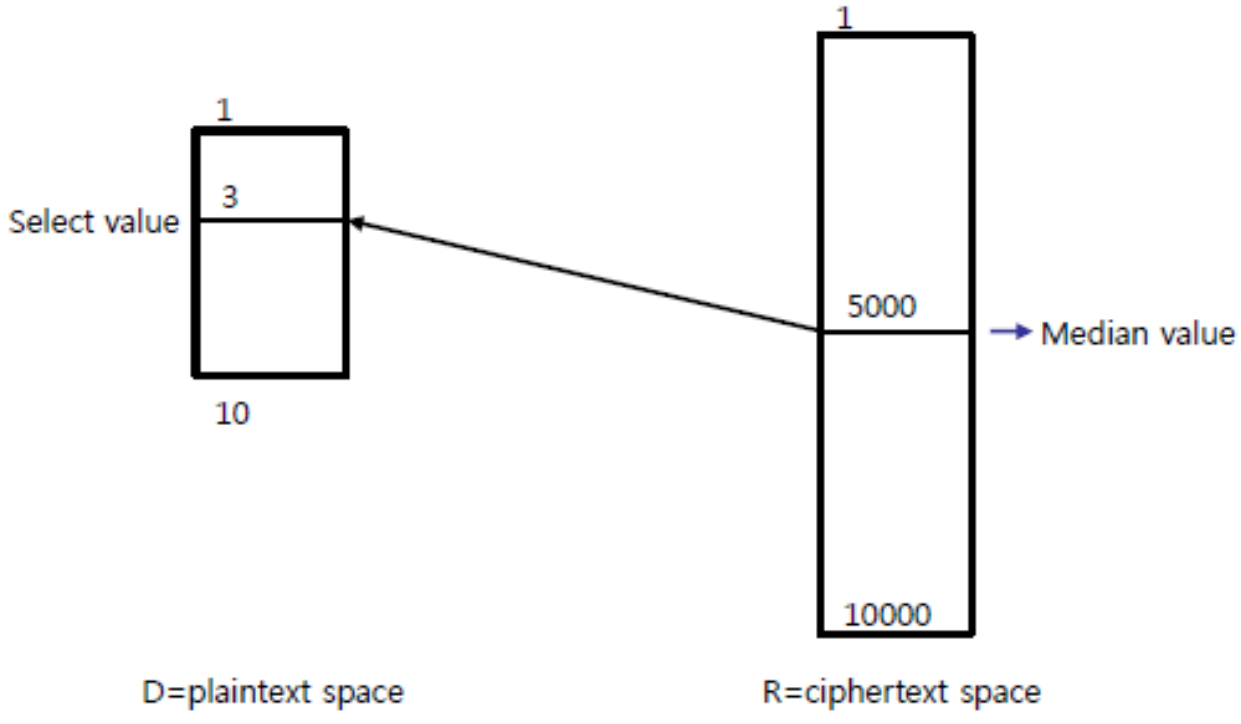
- Now we can efficiently search for data with the encrypted database!

References of Order Preserving Encryption (OPE)

- World War 1, One-part code
- R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, 'Order-preserving encryption for numeric data', SIGMOD 2004
- E. Shi, J. Bethencourt, T-H. H. Chan, D. Song and A. Perrig, 'Multi-dimensional range query over encrypted data', SSP 2007
- D. Boneh and B. Waters, 'Conjunctive, subset, and range queries on encrypted data', TCC 2007
- A. Boldyreva, N. Chenette, L. Younho and A. O'Neill, 'Order-preserving Symmetric Encryption', Eurocrypt 2009
- A. Boldyreva, N. Chenette, and A. O'Neill, 'Order-preserving encryption revisited: Improved security analysis and alternative solutions', Crypto 2011

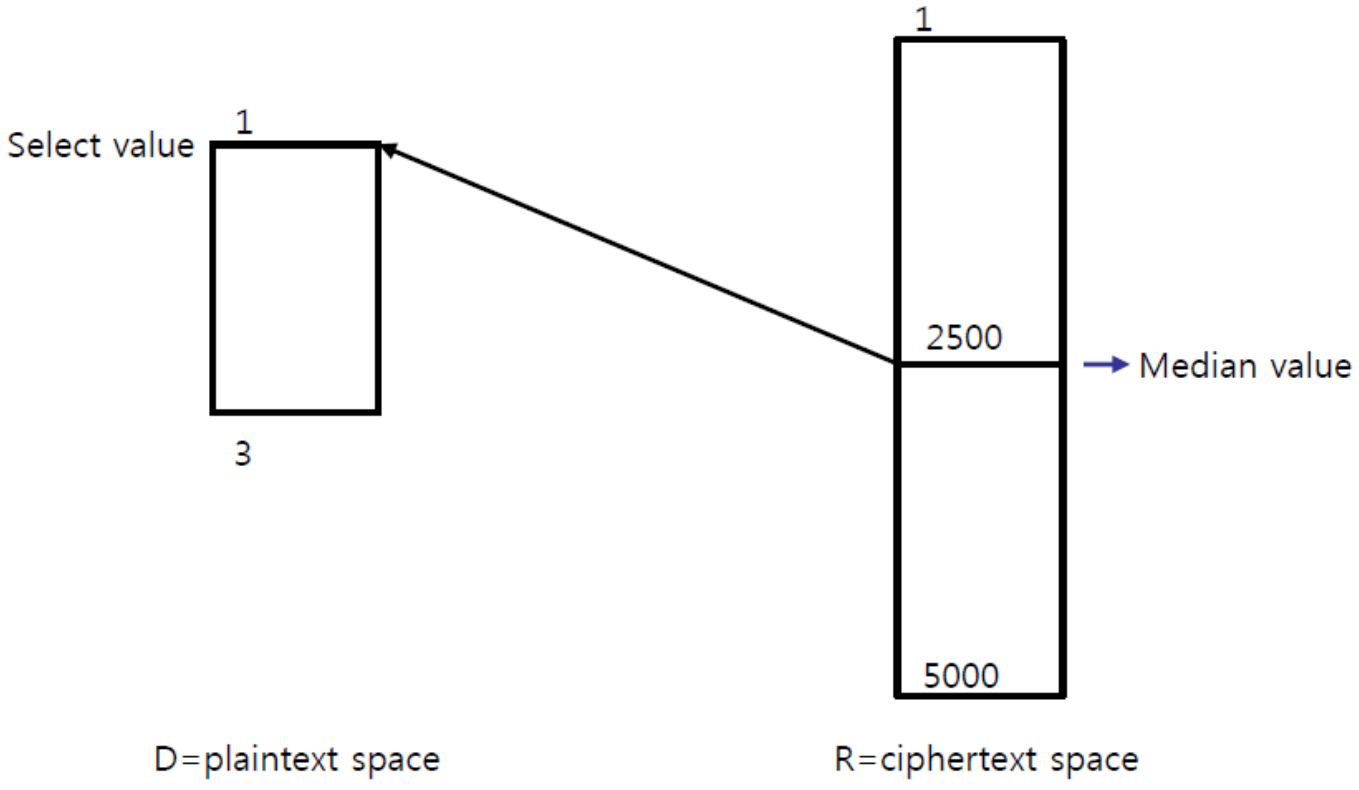
Algorithms of Order Preserving Encryption (OPE)

Lazy sampling a Random Order-Preserving Function (step 1)



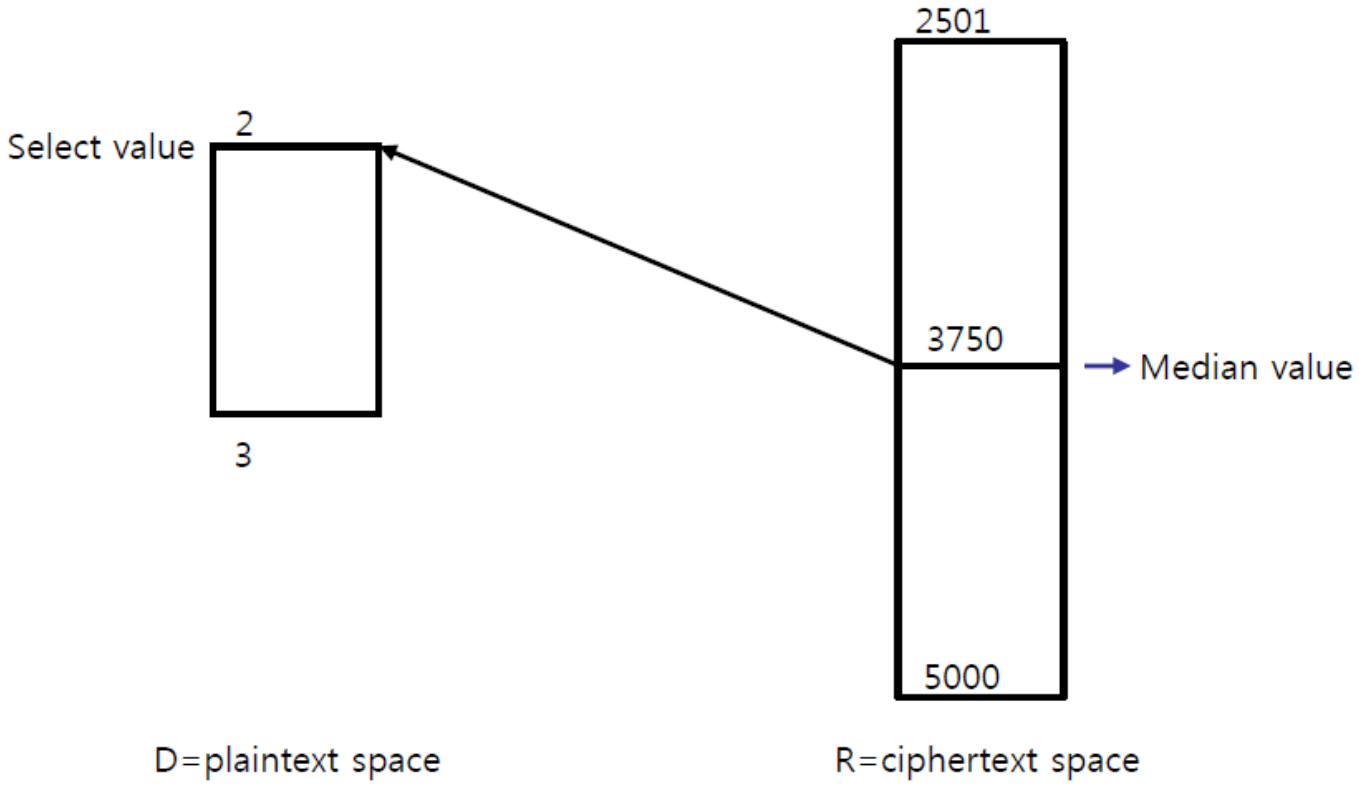
Algorithms of Order Preserving Encryption (OPE)

Lazy sampling a Random Order-Preserving Function (step 2)



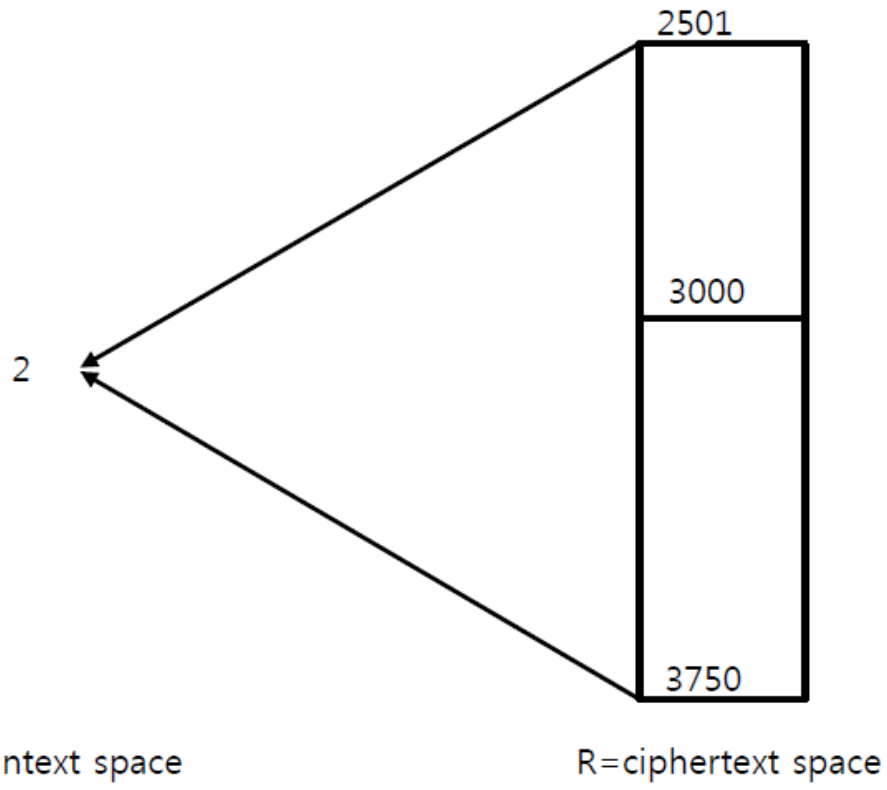
Algorithms of Order Preserving Encryption (OPE)

Lazy sampling a Random Order-Preserving Function (step 3)



Algorithms of Order Preserving Encryption (OPE)

Lazy sampling a Random Order-Preserving Function (step 4)



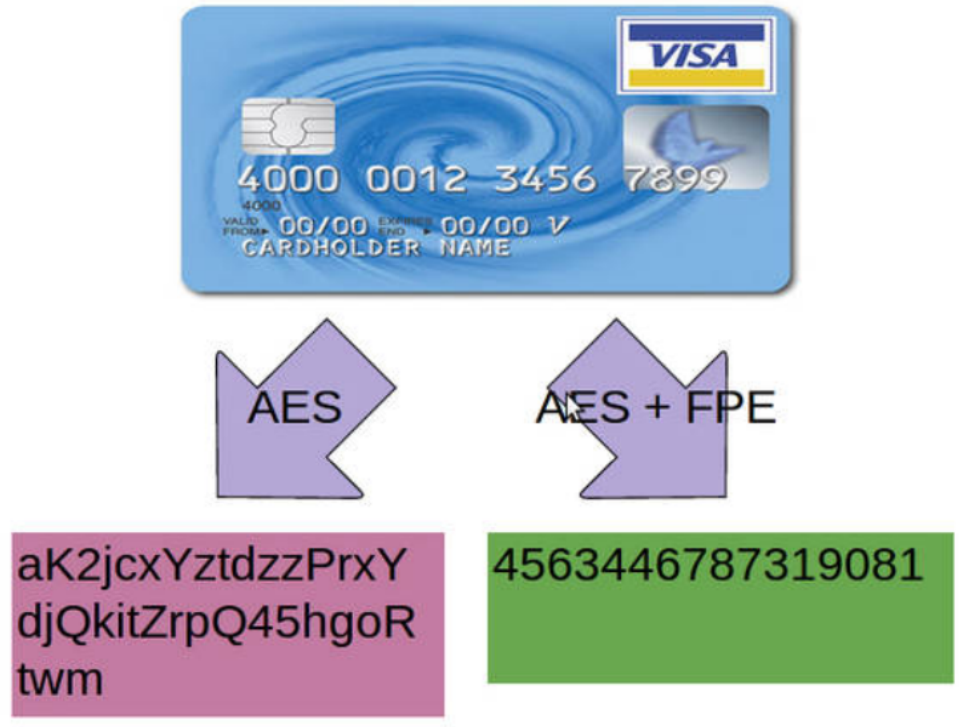
Outline

Introduction to Format Preserving Encryption (FPE)

IT언론의 새로운 대안 디지털데일리

개인정보보호 위한 새로운 암호기술 나왔다

2014.06.13 15:09:15 / 이민형 kiku@ddaily.co.kr



[디지털데일리 이민형기자] 국내 개인정보보호 실정에 맞는 신규 암호기술 '형태보존암호A(FPE-A), 형태보존암호B(FPE-B)'가 국가보안기술연구소(NSRI)에 의해 개발됐다.

Introduction to the Format Preserving Encryption (FPE)

ID	연봉	주민번호
1	30000000	8511...
2	60000000	7612...
3	40000000	8601...
4	18000000	9003...
5	1 30000000	5409...
...	...	...
25154	50000000	8008...

General DB (without encryption)

Introduction to the Format Preserving Encryption (FPE)

<Conventional Encryption algorithms>

ID	연봉	주민번호
1	30000000	8511...
2	60000000	7612...
3	40000000	8601...
4	18000000	9003...
5	1 30000000	5409...
...	...	...
25154	50000000	8008...

encryption

→

ID	연봉	주민번호
1	dlskfdkf	Z9ooewjf
2	Woicziod	lujmw4
3	Apzie.d	Puk jklh
4	.sozirvlz	[lsxdkix.vc
5	sd234,n,v	783wm,4
...	...	...
25154	zpeijfgnsa	iougbkla

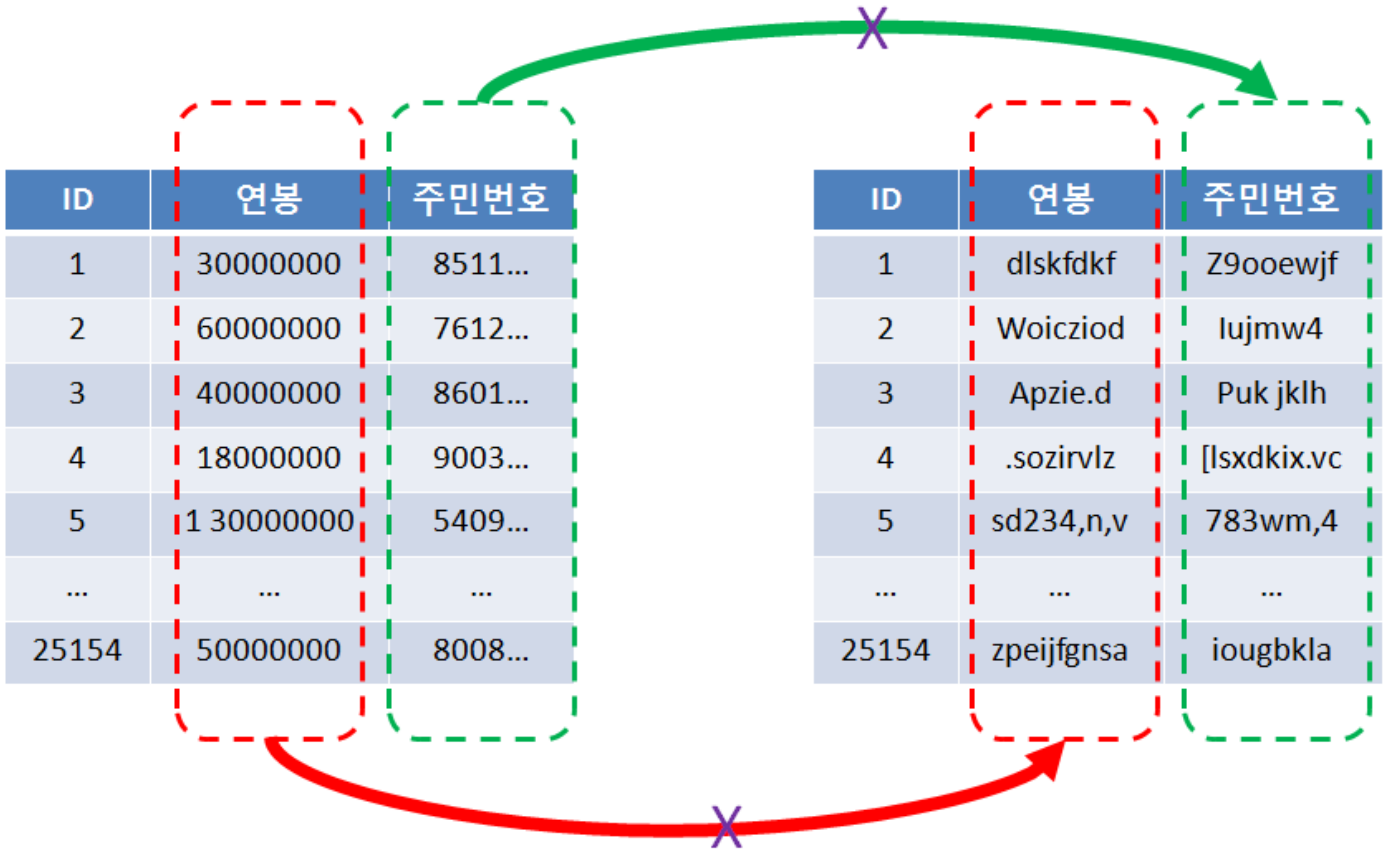
Encrypted DB via general encryption scheme

(AES or other algorithms)

- Unfortunately, this encrypted data cannot be inserted in the DB. Why?

Introduction to Format Preserving Encryption (FPE)

<Conventional Encryption algorithms>



ID	연봉	주민번호
1	30000000	8511...
2	60000000	7612...
3	40000000	8601...
4	18000000	9003...
5	1 30000000	5409...
...	...	...
25154	50000000	8008...

ID	연봉	주민번호
1	dlskfdkf	Z9ooewjf
2	Woicziod	lujmw4
3	Apzie.d	Puk jklh
4	.sozirvlz	[lsxdkix.vc
5	sd234,n,v	783wm,4
...	...	...
25154	zpeijfgnsa	iougbkla

- Because of inconsistent format in DB!

Format Preserving Encryption (FPE)

Some simple idea??

ID	연봉	주민번호
1	30000000	8511...

ID	연봉	주민번호
1	82983428	2348...

1) Integer (P) → Integer (C)
2) Chars (P) → Chars (C)

- Integers (C): $\{000, 001, \cdots, 111\} \rightarrow \{0, 1, \cdots, 8\}$
- Characters (C): This will be no problems.
- Float (C): how??

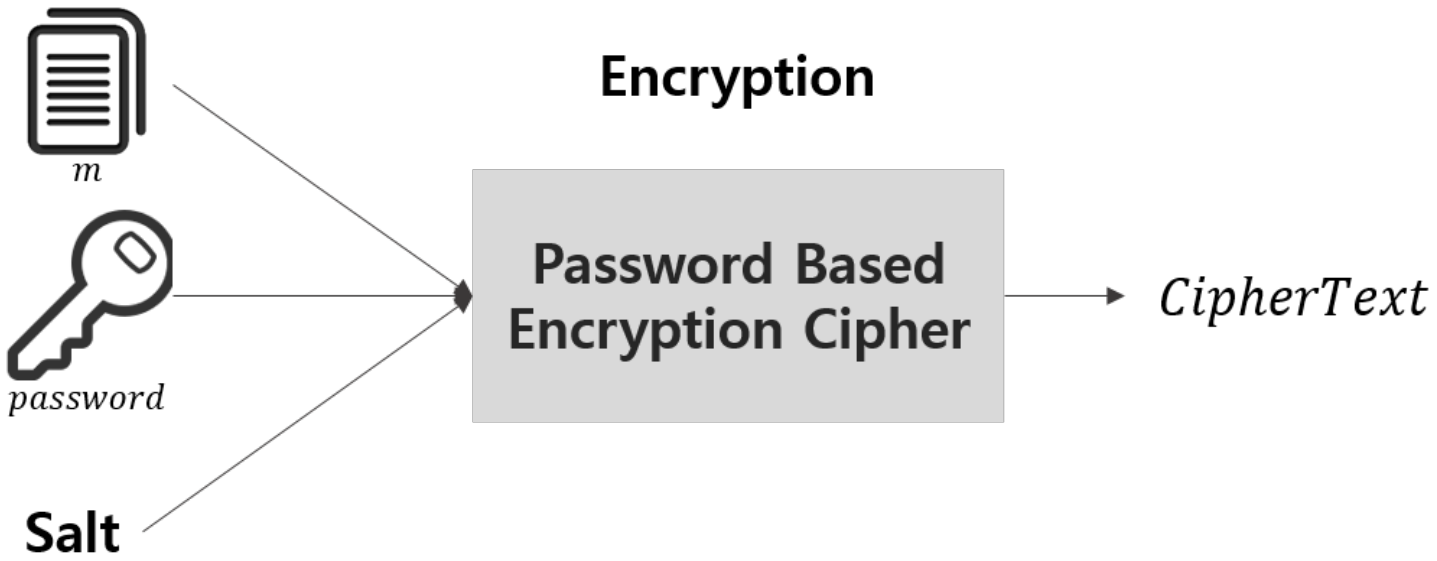
Problems: In this case, the encoded cipher-texts can be longer than actual plain-texts.

- Simple approach: changing coding scheme!! (not perfect but maybe useful!!)

Outline

Conventional PBE

this subsection is made by a student HyunJu Jo, Korea University



Password Based Encryption

an encryption structure based on user's password

Conventional PBE

problem : Low-min Entropy

Low-min Entropy

People uses weak passwords such as ’Password’ or ’123456’.
People routinely choose poor passwords in real world.

Brute Force Attack

1 Attacker Method

1. Trial Decryptions

$$M_1 = Dec(PW_1, C)$$
$$M_2 = Dec(PW_2, C)$$
$$M_3 = Dec(PW_3, C)$$
$$\vdots$$

2. Find True PlainText

~~$M_1 = \#423dDeij1$~~

$M_2 = University$

~~$M_3 = LLo4\text{^}\%78D$~~

$\vdots$

- Many result won't be valid ASCII characters, so attacker can choose message easily. He just choose look like English one.

Brute Force Attack

2 Brute-Force Bound

When password P has min-entropy m.
Most likely password has probability $q/c2^m$
c = salting value, q = number of queries.

- $m < 7$ for passwords observed in a real-world.
- The security offered by conventional PBE is not enough.
- Existing countermeasures only ensure security for highentropy password.

Beyond the brute-force bound

1 Password Based Encryption

1. Trial Decryptions

$M_1 = Dec(PW_1, C)$
$M_2 = Dec(PW_2, C)$
$M_3 = Dec(PW_3, C)$
⋮

2. Find True PlainText

$M_1 = 1010110011$
$M_2 = 0100110000$
$M_3 = 0011000011$
⋮

?

- Attacker cannot find real message from results, when M is an uniformly distributed bit string.

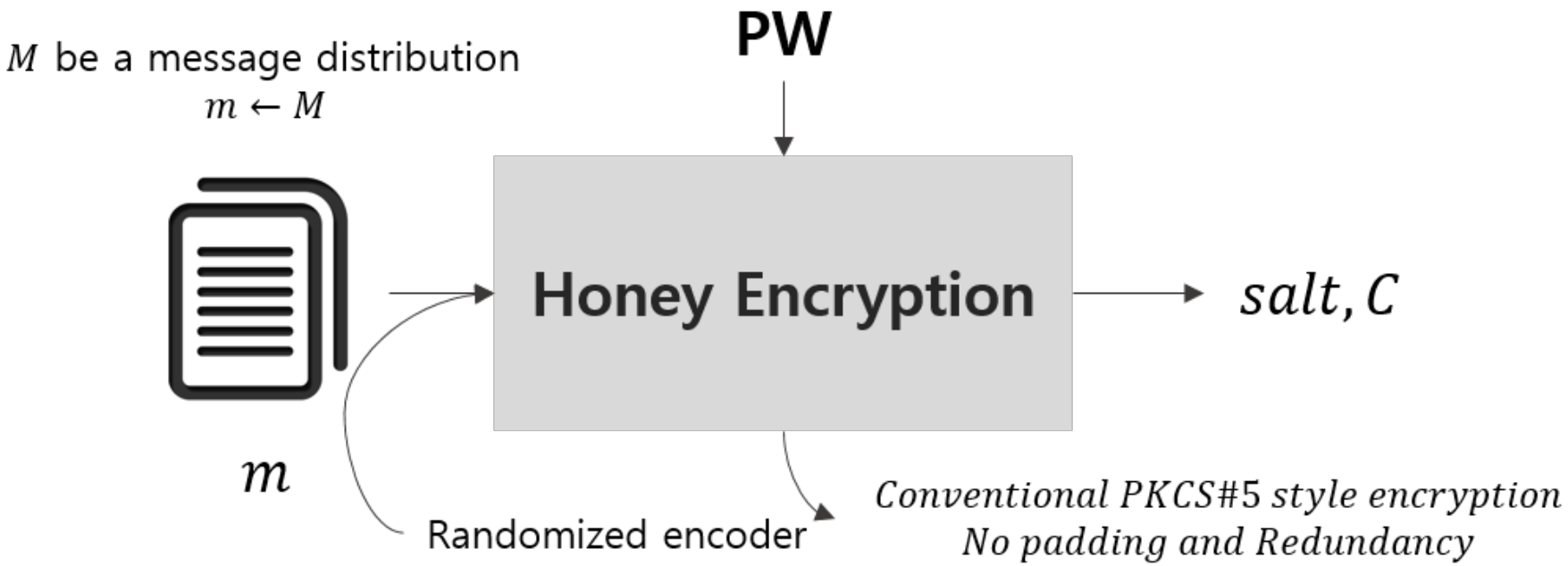
Honey Encryption(HE)

1 background : Related works

- In computer security, there are Decoys.
  - Decoys for detecting attacker's attack behavior.
  - Honeypots, honeytokens, honey accounts.
- Kamoflauge system
- Deniable Encryption
- Format-preserving Encryption

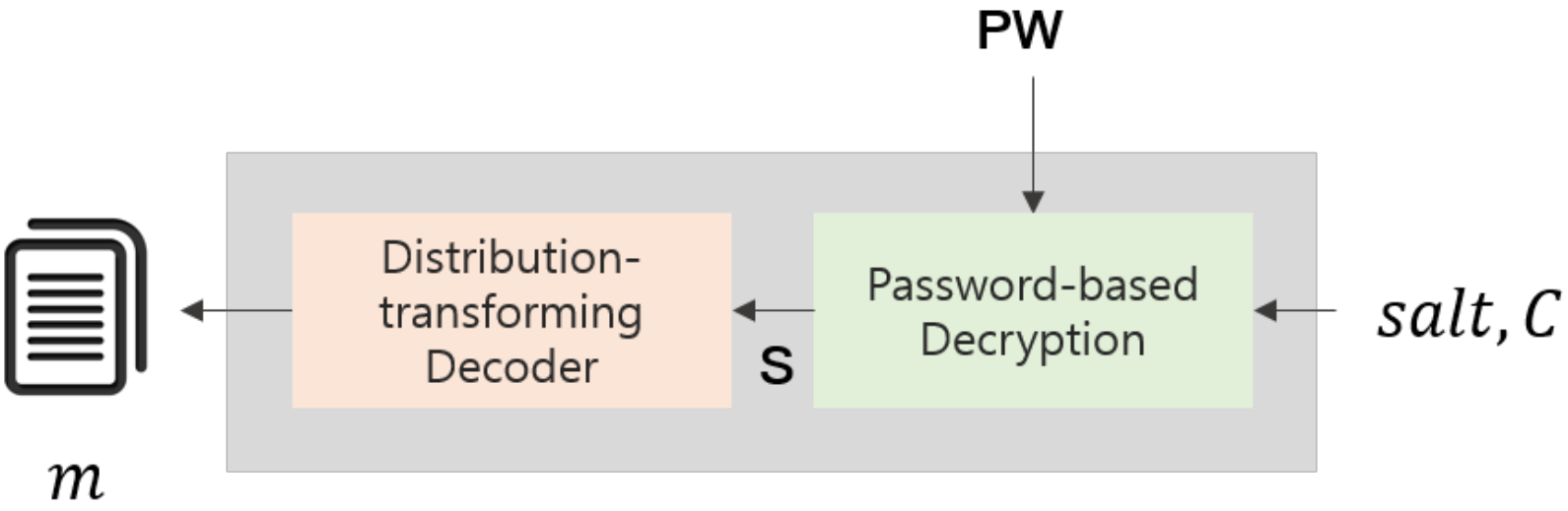
Honey Encryption(HE)

2 Scheme : encoder



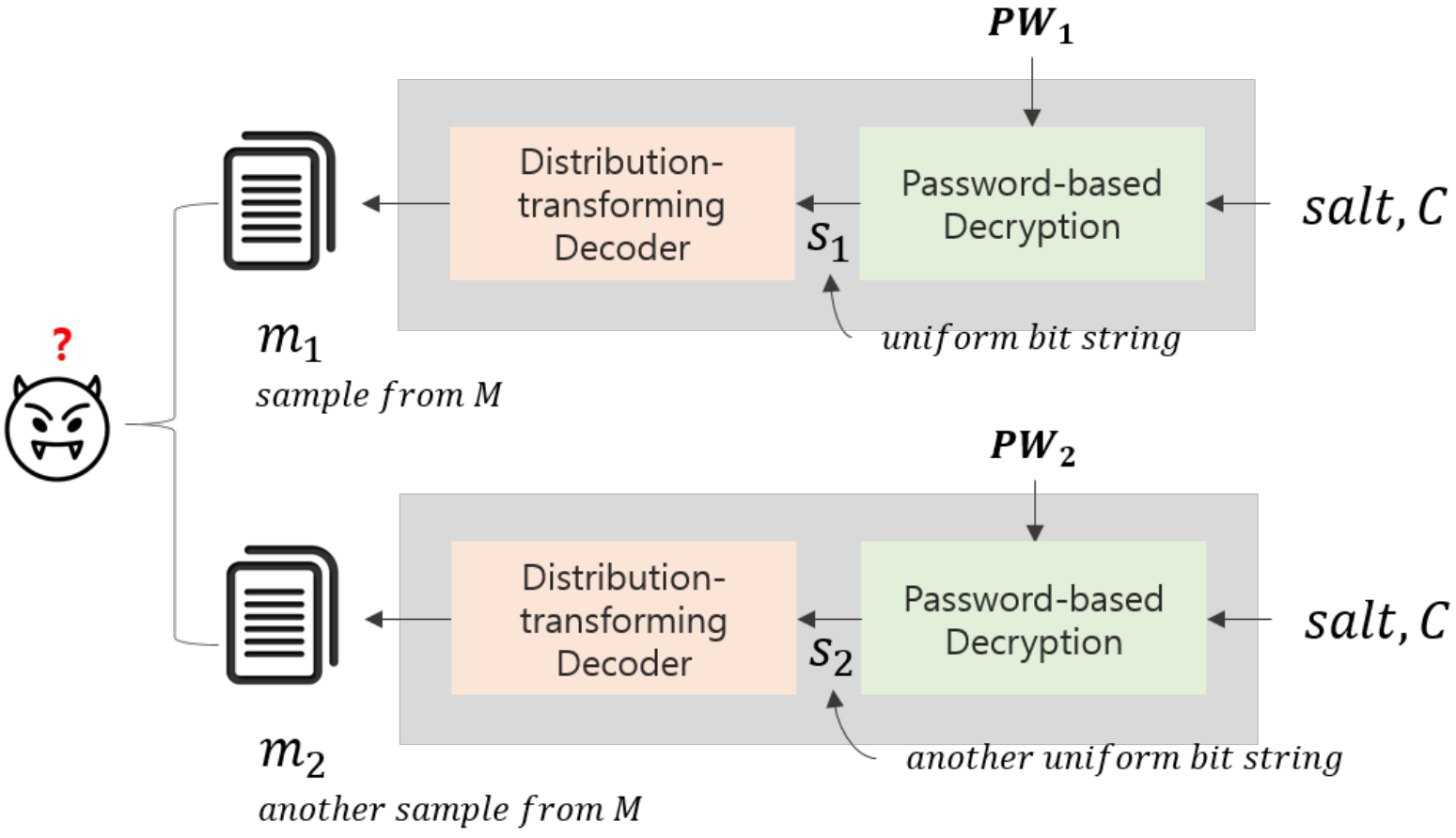
Honey Encryption(HE)

2 Scheme : decoder



Honey Encryption(HE)

2 Scheme : wrong key decoding



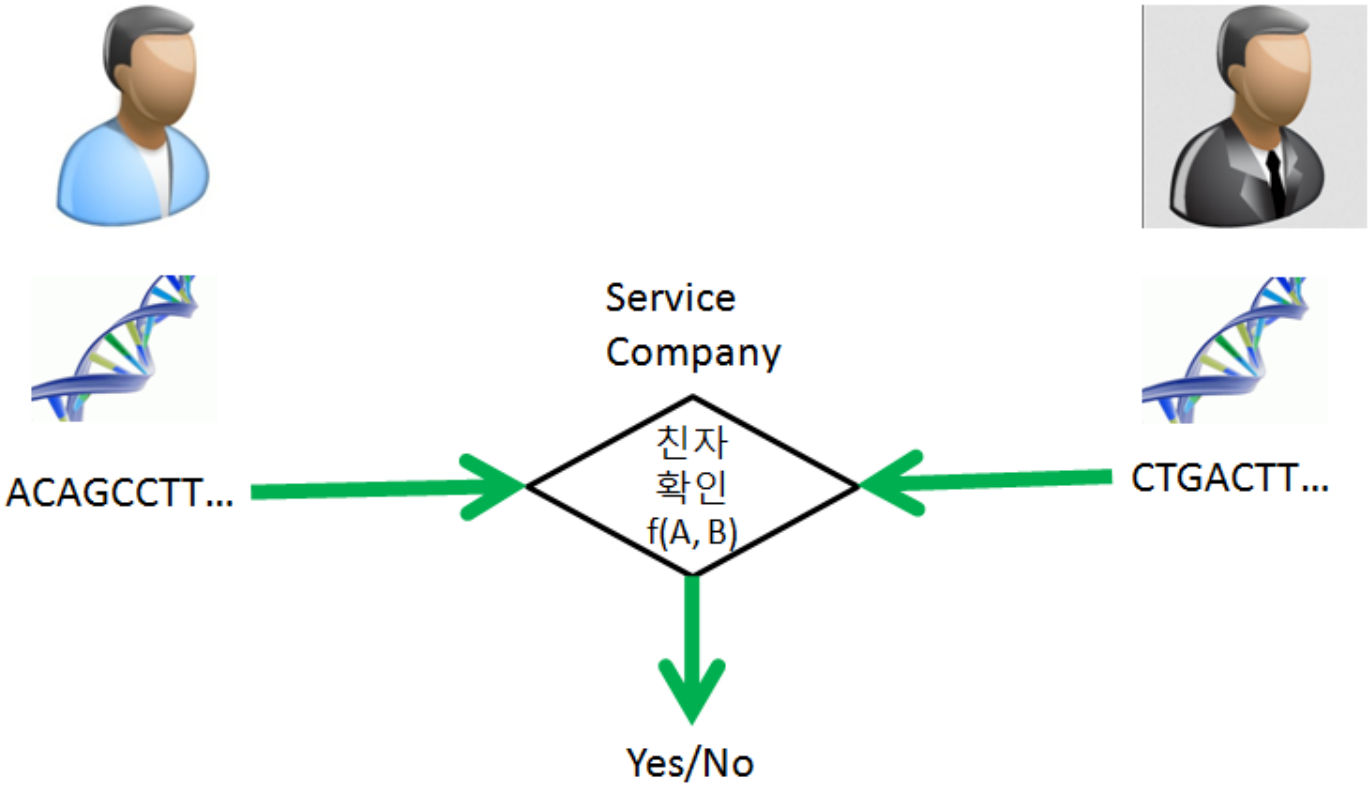
Honey Encryption(HE)

3 Honey Encryption is

- $DTE = (Encode, Decode)$
- Encoding is randomized.
  - encode(m) is based on cumulative distribution function(CDF)
  - $m \leftarrow M, \quad S \leftarrow \$encode(M), \quad Return(M, S)$
- Decoding is deterministic.
  - $S \leftarrow \$\{0,1\}^s, \quad M \leftarrow decode(S), \quad Return(M, S)$

Outline

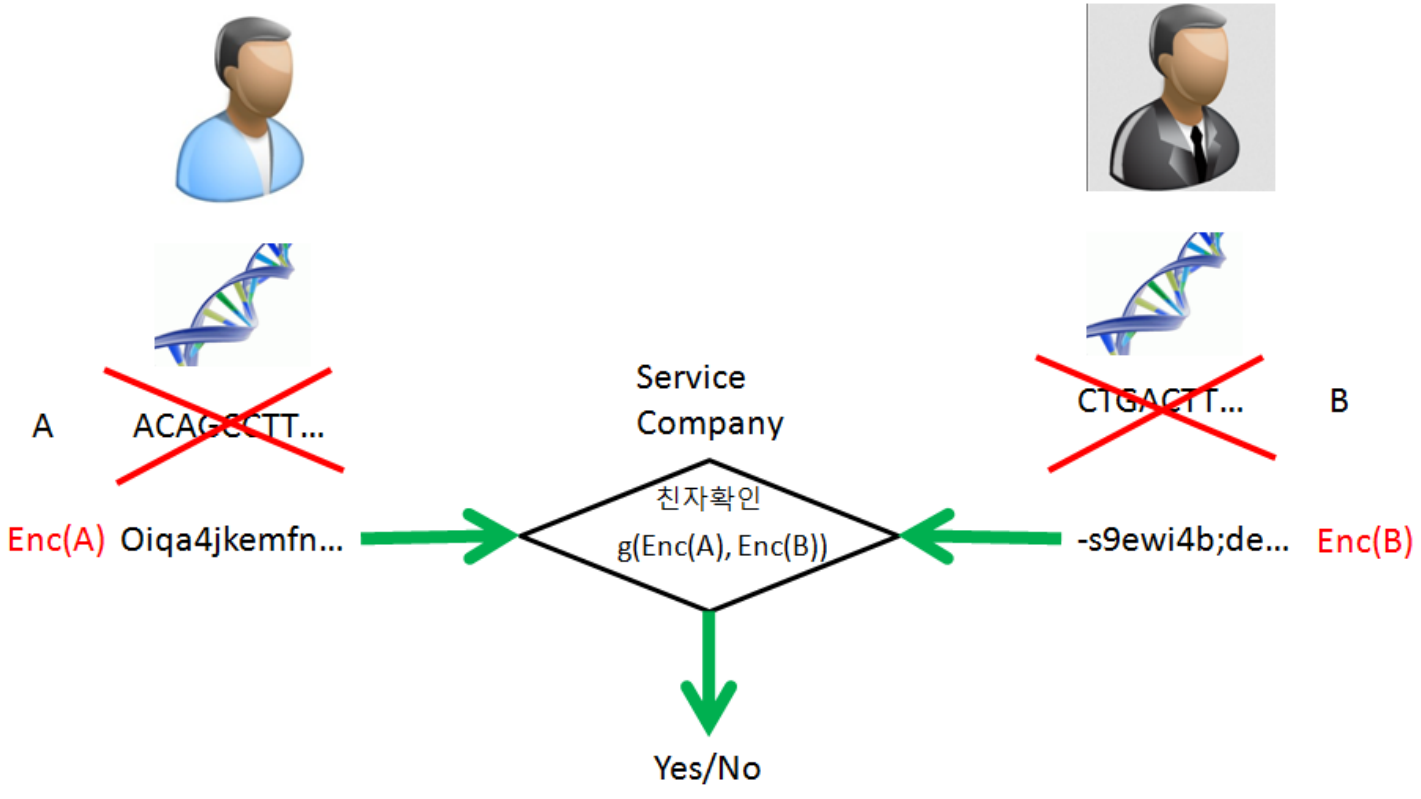
Homomorphic Encryption(HE)



Don’t secure at all.

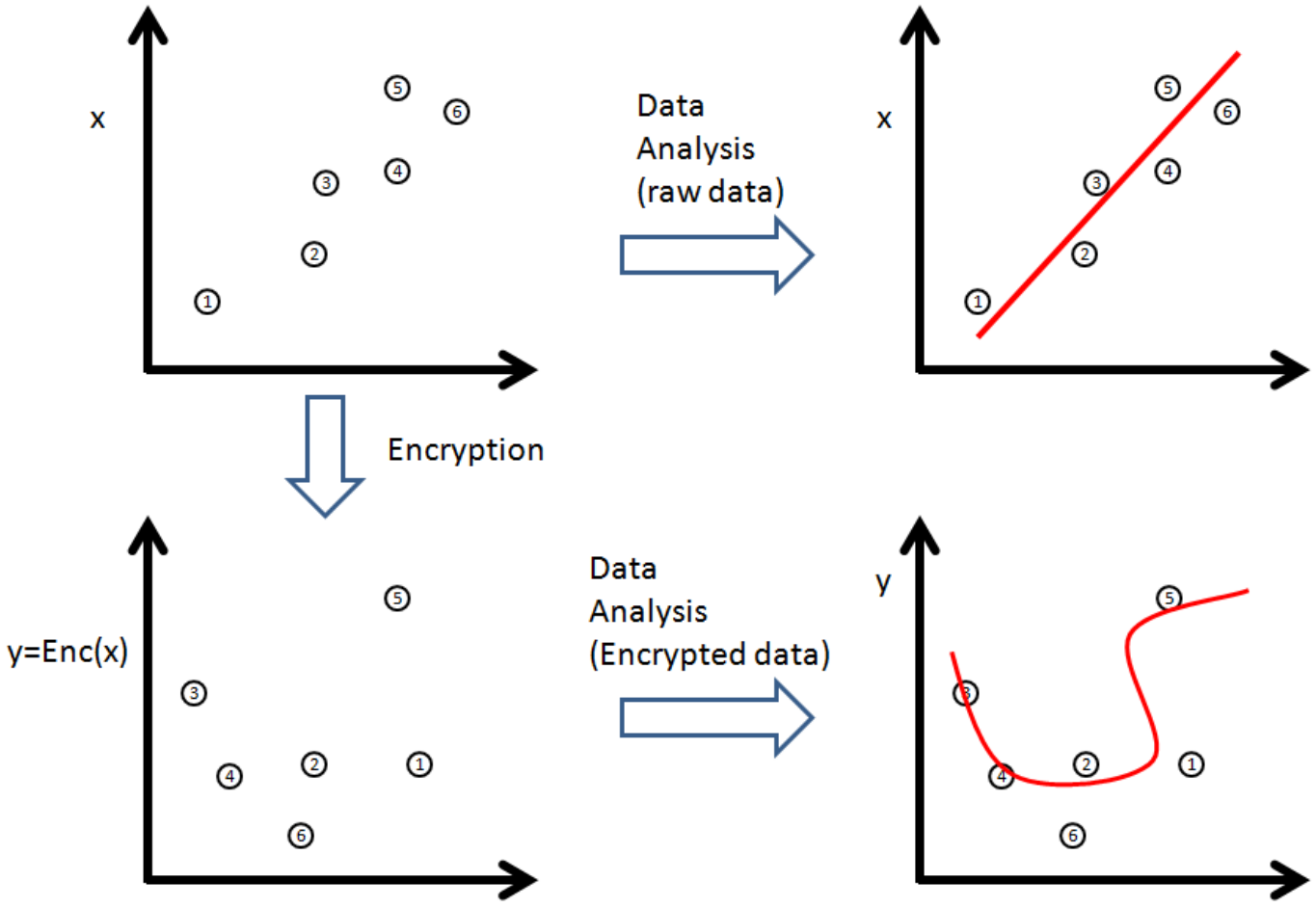
In worst cases, the service company may not be trust-able.
Then your all critical personal information will be exposed.
→ Homomorphic encryption is required

Homomorphic Encryption(HE)



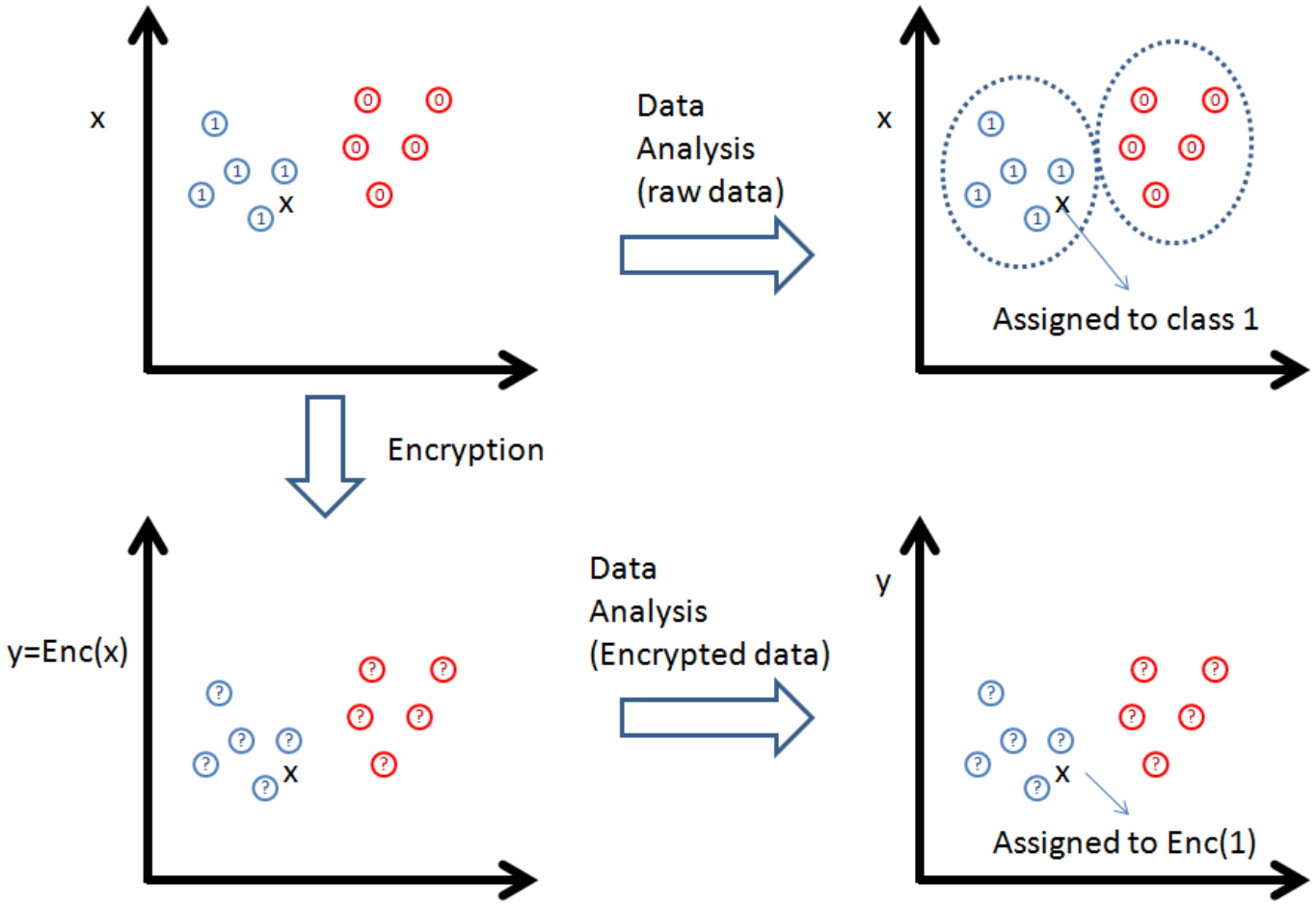
Now, it is secure. :-)

Privacy Preserving Data analysis (1/2)



Regression

Privacy Preserving Data analysis (1/2)

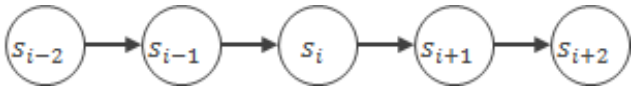


Classification

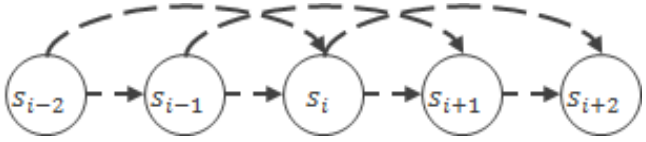
Statistical coding scheme

[** Markov Process **]

Question! What is the next character after 'Databas'? Probably it will be 'e'.



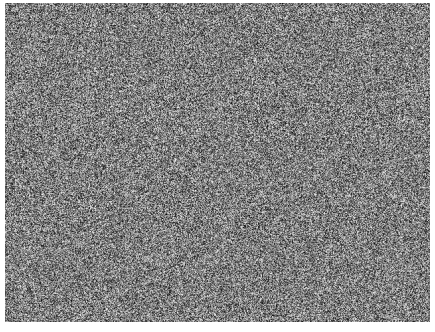
(a) 1st order Markov chain



(b) 2nd order Markov chain

Applications

- **Texts**: Natural Language Processing (NLP) and n-gram modeling
- **Images**: Real image has a neighbouring structure in color value



Statistical coding scheme

[Structural Corpus]

본 행사는 국내 최대의 데이터 분야 전문 컨퍼런스로
지난 행사에서는 1,800여명이 넘는 참석자분들의 뜨
습니다.
2014 데이터 그랜드 컨퍼런스에서는 "데이터 시대
부 및 학계, 산업계의 전문가들을 초빙하여 창조경제
고자 합니다. 또한 데이터 관리 모범 사례와 추진 전
지식소통과 가치교환의 장을 마련해 드리고자 합니

texts

images

Corpus DB

Building CorpusDB for structural coding systems

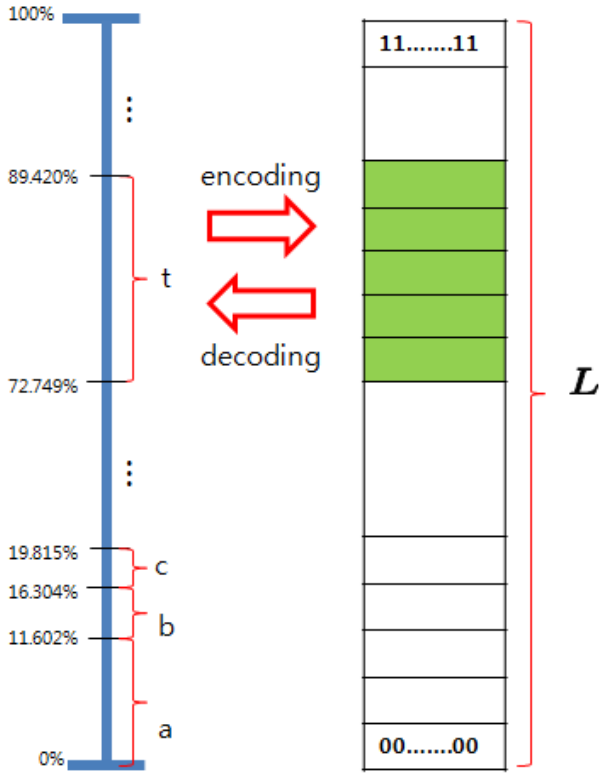
First character at statistical coding scheme

the speaker is working at ...

$p(X_1 = \text{`}t\text{'}) = ?$ $\quad \mathcal{A}_x = \{a, b, c, \cdots, z\}$

$p(X_1 = j) = p_{1,j}$ for $j \in \mathcal{A}_x$ and $\sum_{j \in \mathcal{A}_x} p_{1,j} = 1$

Letter	Relative frequency as the first letter of an English word
a	11.602%
b	4.702%
c	3.511%
d	2.670%
e	2.007%
f	3.779%
g	1.950%
h	7.232%
i	6.286%
j	0.597%
k	0.590%
l	2.705%
m	4.374%
n	2.365%
o	6.264%
p	2.545%
q	0.173%
r	1.653%
s	7.755%
t	16.671%
u	1.487%
v	0.649%
w	6.753%
x	0.037%
y	1.620%
z	0.034%



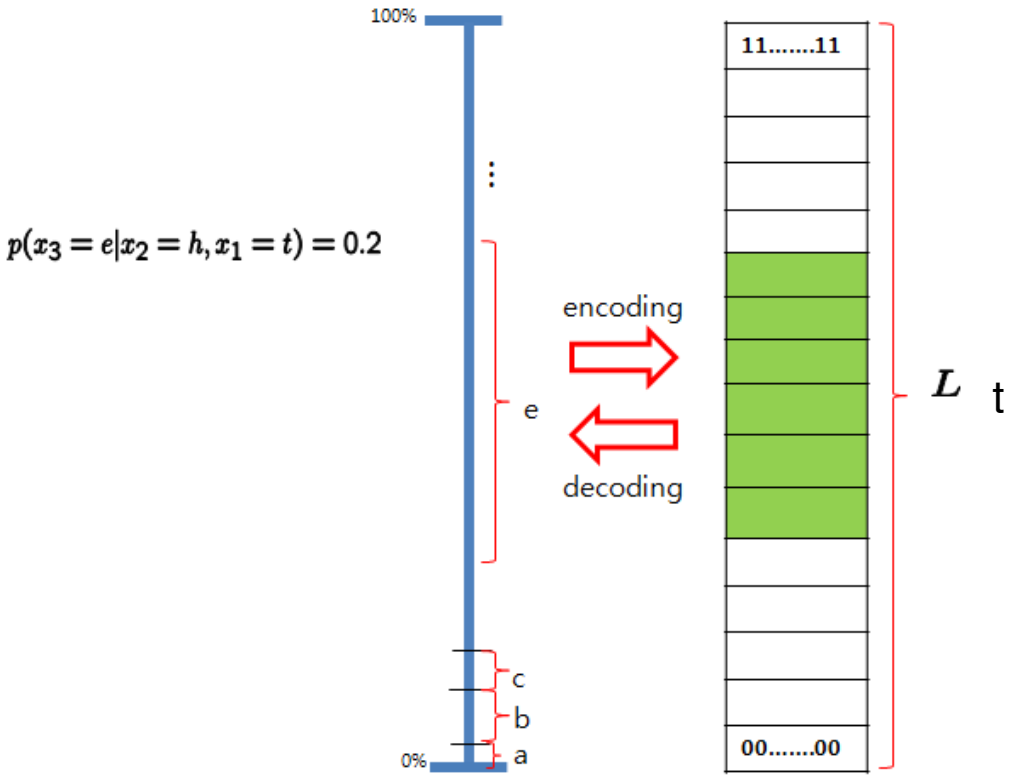
Other characters at statistical coding scheme

the speaker is working at ...

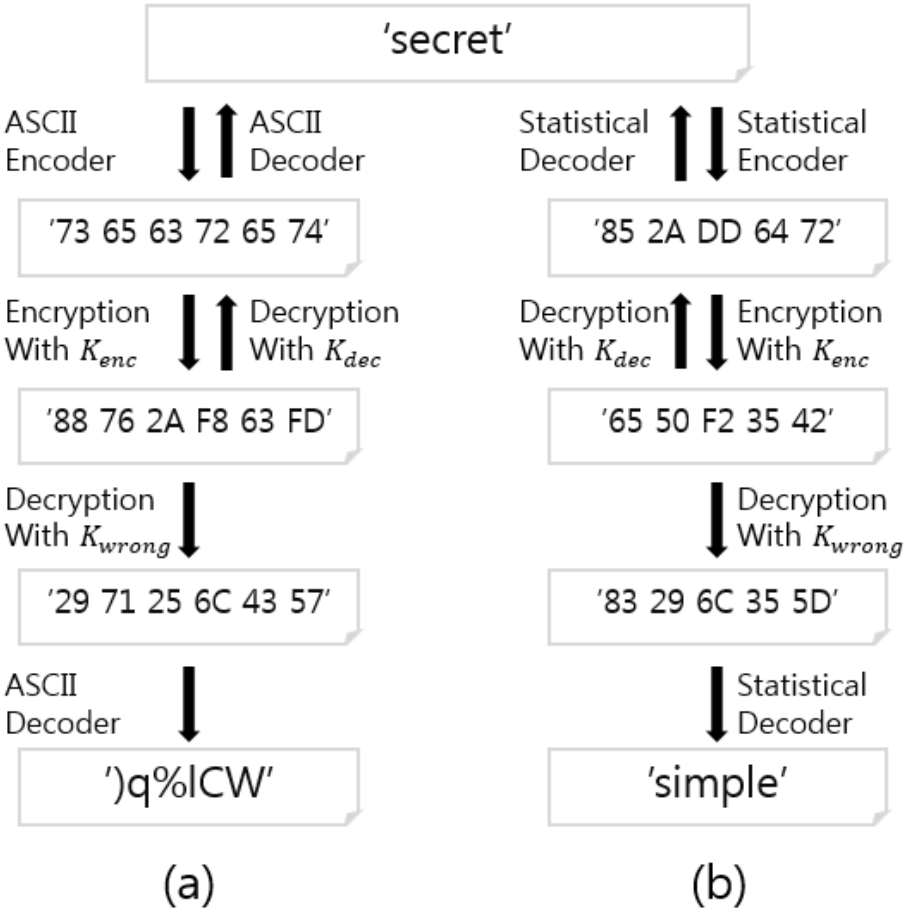
$$p(X_3 = \text{`}e\text{`} \mid X_2 = \text{`}h\text{`} \text{ and } X_1 = \text{`}t\text{`}) = ?$$

$$\mathcal{A}_x = \{a, b, c, \cdots, z\}$$

$$p(X_3 = j | X_2 = \text{`}h\text{`} \text{ and } X_1 = \text{`}t\text{`}) = p_{3,j} \text{ for } j \in \mathcal{A}_x \text{ and } \sum_{j \in \mathcal{A}_x} p_{3,j} = 1$$



Statistical coding scheme



Statistical coding scheme

Underlying plain-texts: ‘Deniable encryption‘

TABLE I. DECRYPTED TEXTS WITH DIFFERENT KEYS AND DATABASES: NASA (16KB), ROMEO & JULIET (247KB), CRYPTOGRAPHY (340KB), BIBLE (4.9MB)

Dataset for DB	$K^{(1)}_{wrong} \neq K_{dec}$	$K^{(2)}_{wrong} \neq K_{dec} \neq K^{(1)}_{wrong}$
NASA	’the scout’s payload ac’	’the spacecraft and the’
Romeo & Juliet	’what show the project’	’scene iii. scene iii’
Cryptography	’the secret key signatu’	’the probabilistic tech’
Bible	’and the children of th’	’and the lord was not b’

Outline

Thank you for listening to my talk!

Contact me if you have any further queries

- E-mail: jiwon_yoon@korea.ac.kr
- Lab site: https://sites.google.com/site/securesiplab/

- (Also, I am currently looking for potential Ph.D. and M.Sc. students for privacy preserving data analysis topics.)