Other Database Security Concerns

- Redundancy/Internal Consistency
  - Error detection and correction codes to protect data integrity
    - E.g., Parity bits, Hamming codes, and cyclic redundancy checks (CRC)
    - Can be applied to single fields, records or entire DB
  - Shadow fields
    - Entire attributes or records are duplicated in DB
- Recovery
  - DB maintains a change log, allowing it to repeat changes as necessary
- Concurrency/Consistency
  - DB uses locks and atomic operations to maintain consistency
  - Writes are treated as atomic operations
  - Records are locked during write so they cannot be read in a partially updated state

Example of Concurrency/Consistency

- Assume a DB maintains seat reservations for an airline flight
- Agent A and B try to reserve seats at the same time

ASSIGN happens only when the seat is unassigned

Agent A

```
SELECT (SEAT-NO=’11D’)
ASSIGN ‘MOCK,E’ TO PASSENGER-NAME
```

Agent B

```
SELECT (SEAT-NO=’11D’)
ASSIGN ‘EDWARDS,S’ TO PASSENGER-NAME
```

- Without consistency guarantee, A and B will get into race condition, reserving the same seats
- DB resolves the problem by treating the entire query-update cycle as a single atomic operation

Database Disclosure

- Sensitive data
- Types of disclosures
- Preventing disclosures

Sensitive Data

- Inherently sensitive
  - E.g., passwords, locations of weapons
- From a sensitive source
  - E.g., confidential informant
- Declared sensitive by DB admin
  - E.g., classified document, name of an anonymous donor
- Part of a sensitive attribute or record
  - E.g., salary attribute in an employment database
- Sensitive in relation to previously disclosed information
  - E.g., an encrypted file combined with the password to open it

DB should protect sensitive data from *direct* or *indirect* access

Types of Disclosures

- Exact value of sensitive data
  - The sensitive data are directly obtained by adversary after query
  - Can be DB's misconfiguration or DB admin's oversight
- Bounds on sensitive value
  - Learn a sensitive value *y* is between two values, *L* and *H*
  - Can be done through binary search
- Negative result
  - Learn *z* is not the value of *y*
  - E.g., 0 is not the total number of felony convictions => person is a felony

Types of Disclosures (cond.)

- Existence
  - Whether a record/element exists in DB
  - E.g., whether the number of phone calls field exists

- Probable value
  - Determine the probability that certain elements has certain value
  - E.g., try to find out whether president of US is registered in Tory party

Count(Residence="1600 Pennsylvania Avenue") = 4
Count(Residence="1600 Pennsylvania Avenue" AND Tory=TRUE) = 1

25% likelihood

Inference Attack

- A way to derive sensitive data from non-sensitive data
- Sensitive query (result associated with only one person)

  List NAME where SEX=M and DRUGS=1

- Query seems to conceal drug usage but actually reveals it

  List NAME where (SEX=M and DRUGS=1) or (SEX!=M and SEX!=F) or (DORM=AYRES)

Sensitive fields

Name	Sex	Race	Aid	Fines	Drugs	Dorm
Adams	M	C	5000	45.	1	Holmes
Bailey	M	B	0	0.	0	Grey
Chin	F	A	3000	20.	0	West
Dewitt	M	B	1000	35.	3	Grey
Earhart	F	C	2000	95.	1	Holmes
Fein	F	C	1000	15.	0	West
Groff	M	C	4000	0.	3	West
Hill	F	B	5000	10.	2	Holmes
Koch	F	C	0	0.	1	West
Liu	F	A	0	10.	2	Grey
Majors	M	C	2000	0.	2	Grey

Student info

Inference by Arithmetic

- E.g., US Census Bureau collects personal data and release statistics
  - Only count, sum and mean are released
  - Individual names, addresses or other personal characteristics are suppressed
  - What sensitive info can be revealed from count, sum and mean?
- Inference from sum
  - Sum up the aid by dorm and sex

	Holmes	Grey	West	Total
M	5000	3000	4000	12000
F	7000	0	4000	11000
Total	12000	3000	8000	23000

No female in Grey received aid

- Inference from count
  - Count of records by dorm and sex
  - Usually combined with sum inference

Sex	Holmes	Grey	West	Total
M	1	3	1	5
F	2	1	3	6
Total	3	4	4	11

1 male in Holmes and 1 male in West received $5000 and $4000

Inference by Arithmetic (cond.)

- Inference from mean
  - E.g., mean salary of all employees and mean salary of employees without the president reveals president's salary
- Inference from median
  - Find one point of intersection that happens to be exactly in the middle

q = median(AID where SEX=M)

p = median(AID where DRUGS=2)

Majors' aid is 2000

Name	Sex	Drugs	Aid
Bailey	M	0	0
Dewitt	M	3	1000
Majors	M	2	2000
Groff	M	3	4000
Adams	M	1	5000
Liu	F	2	0
Majors	M	2	2000
Hill	F	2	5000

Student info 2

Inference by Arithmetic (cond.)

- Tracker attacks
  - Run multiple queries and let them cancel each other out
  - Locate the desired record
  - Is a specific case of linear system vulnerability (value of unknown variable can be learned by solving multiple linear equations)

count ((SEX=F) and (RACE=C) and (DORM=Holmes))

Sensitive query if the result is only one

count (SEX=F)

count ((SEX=F) and (RACE!=C) or (DORM!=Holmes))

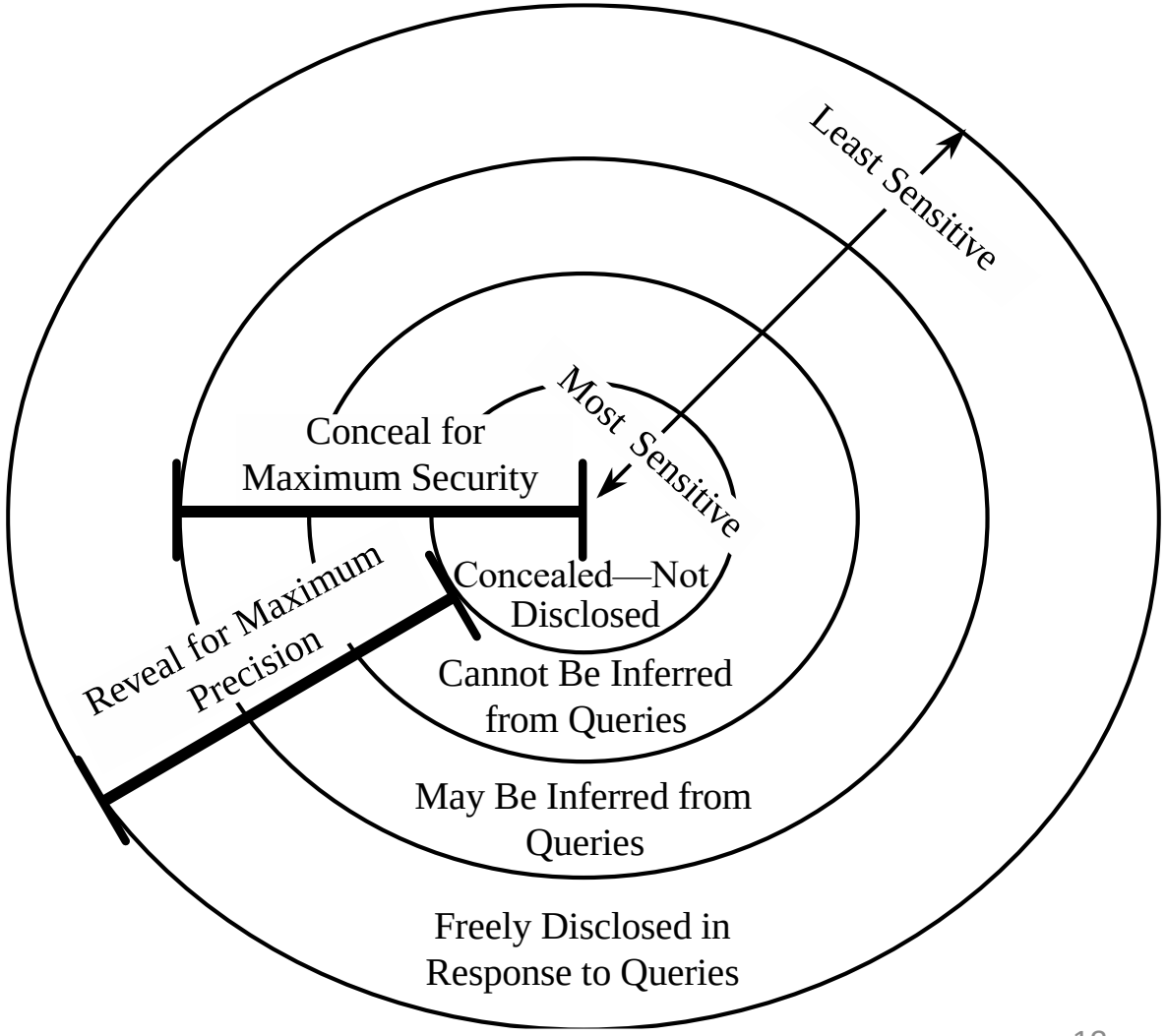
Not sensitive query if result > 1, but their difference (equals to the above query) is sensitive

Preventing Disclosure

- What can be done in addition to DB access control?
  - Suppression (sensitive data not forthcoming) or concealing (not exact same value)
- Suppress obviously sensitive information
  - Easy to implement
  - Easy to bypass
- Track what each user knows
  - Assess information leakage based on past queries
  - Doesn’t work when multiple users collude
- Disguise the data
  - Random perturbation and rounding inhibits attacks that depend on exact values
  - But might introduce inaccurate result for legitimate users

Security vs. Precision

- Protect all sensitive data (security goal) while revealing as much non-sensitive data as possible (precision goal)



Statistical Suppression

- Limited response suppression
  - Eliminates low-frequency elements from being displayed
- Combined results
  - Merge rows and columns to protect sensitive values
  - Present values in ranges or rounding

Sensitive values can be derived

	Drug Use			
Sex	0	1	2	3
M	1	1	1	2
F	2	2	2	0

Combining values to suppress sensitive data

	Drug Use	
Sex	0 or 1	2 or 3
M	2	3
F	4	2

- Random sample
  - Derive result from a random sample of DB instead of whole DB