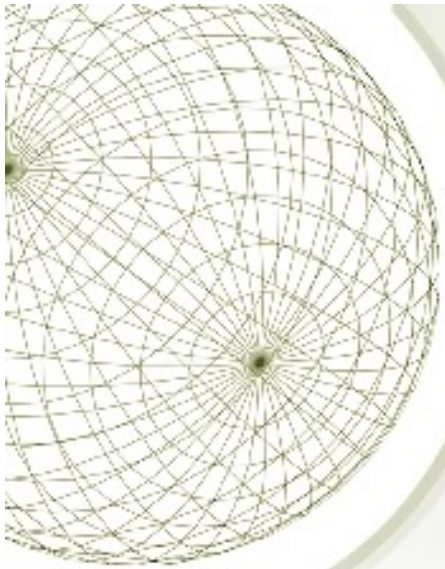
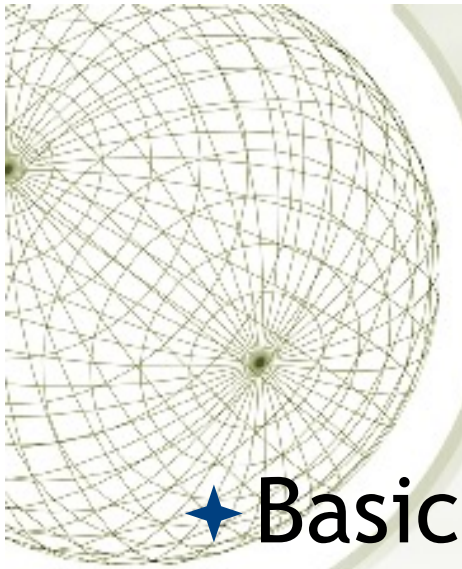


Database Security

Ola Flygt
Linnaeus University, Sweden
http://homepage.lnu.se/staff/oflmsi/
Ola.Flygt@lnu.se
+46 470 70 86 49

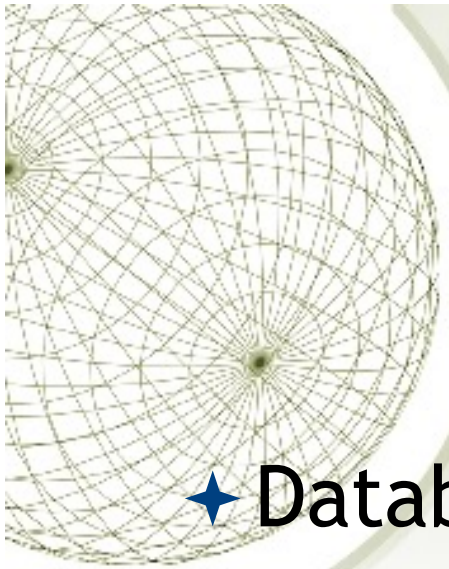
Slides based on *Security in Computing. Third Edition* by Pfleeger and Pfleeger.
Using some slides courtesy of:
Prof. Barbara Endicott-Popovsky and Prof. Deborah Frincke (U. Idaho) — taught at U. Washington
Prof. Csilla Farkas— course taught at U. of South Carolina
Prof. Leszek T. Lilien - Western Michigan University



Objectives

- Basic database terminology and concepts
- Security requirements for databases
- Implementing access controls in databases
- Protecting sensitive data
- Data mining and big data

From *Security in Computing, Fifth Edition*, by Charles P. Pfleeger, et al. (ISBN: 9780134085043). Copyright 2015 by Pearson Education, Inc. All rights reserved.

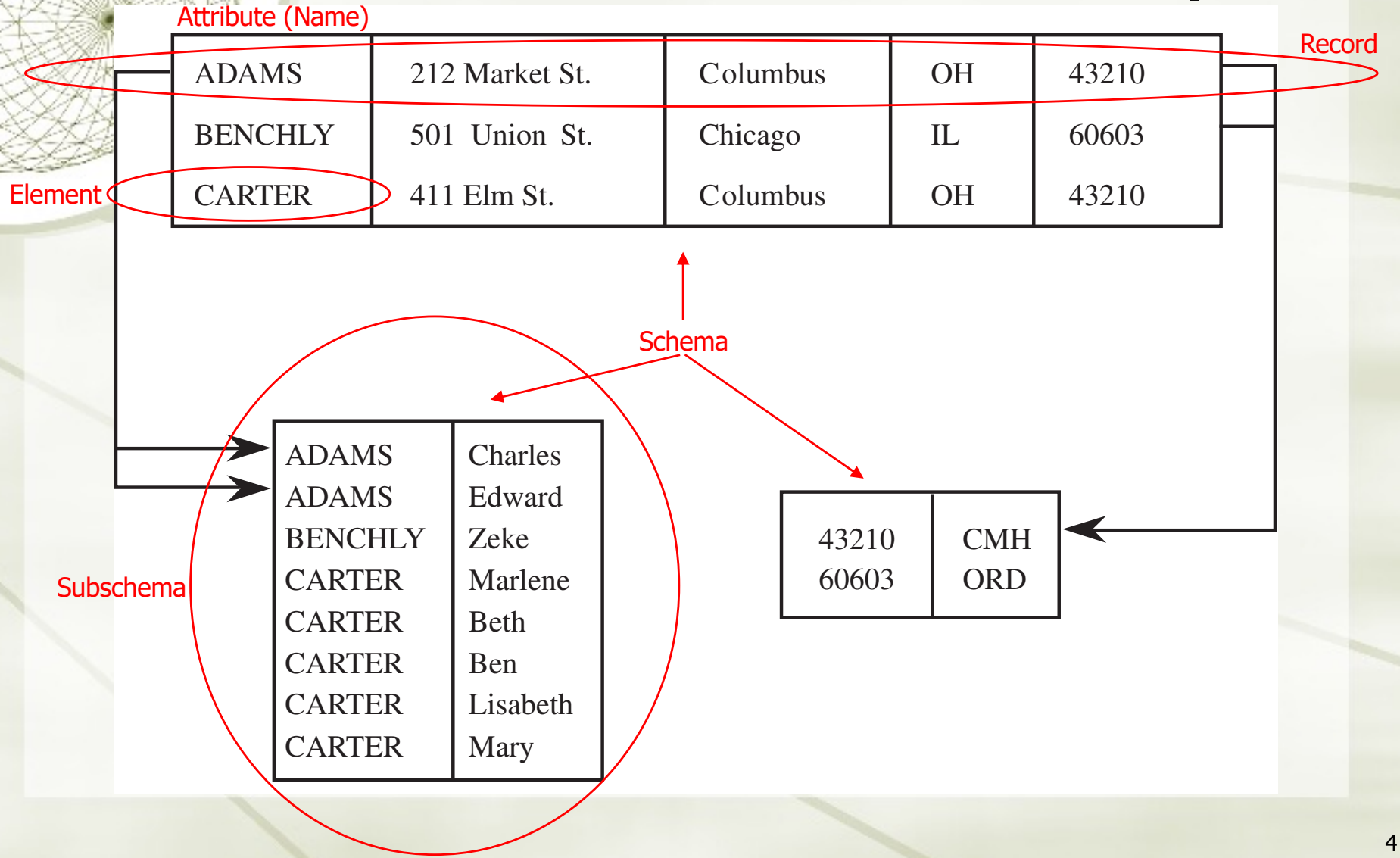


Database Terms

- Database administrator
- Database management system (DBMS)
- Record
- Field/element
- Schema
- Subschema
- Attribute
- Relation

From *Security in Computing, Fifth Edition*, by Charles P. Pfleeger, et al. (ISBN: 9780134085043). Copyright 2015 by Pearson Education, Inc. All rights reserved.

Database Example



From *Security in Computing, Fifth Edition*, by Charles P. Pfleeger, et al. (ISBN: 9780134085043). Copyright 2015 by Pearson Education, Inc. All rights reserved.

Schema Example

Name	First	Address	City	State	Zip	Airport
ADAMS	Charles	212 Market St.	Columbus	OH	43210	CMH
ADAMS	Edward	212 Market St.	Columbus	OH	43210	CMH
BENCHLY	Zeke	501 Union St.	Chicago	IL	60603	ORD
CARTER	Marlene	411 Elm St.	Columbus	OH	43210	CMH
CARTER	Beth	411 Elm St.	Columbus	OH	43210	CMH
CARTER	Ben	411 Elm St.	Columbus	OH	43210	CMH
CARTER	Lisabeth	411 Elm St.	Columbus	OH	43210	CMH
CARTER	Mary	411 Elm St.	Columbus	OH	43210	CMH

From *Security in Computing, Fifth Edition*, by Charles P. Pfleeger, et al. (ISBN: 9780134085043). Copyright 2015 by Pearson Education, Inc. All rights reserved.

Queries

- A query is a command that tells the database to retrieve, modify, add, or delete a field or record
- The most common database query language is SQL

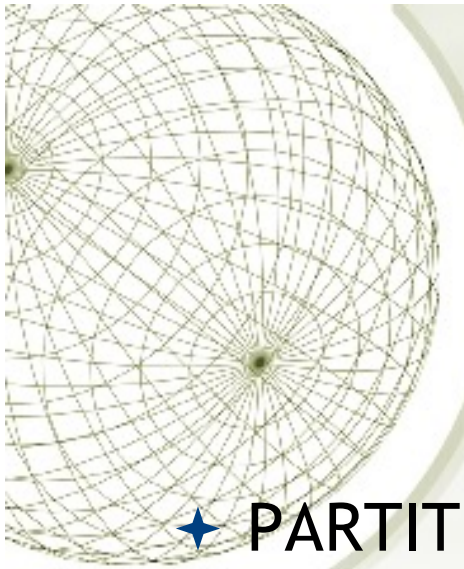
From *Security in Computing, Fifth Edition*, by Charles P. Pfleeger, et al. (ISBN: 9780134085043). Copyright 2015 by Pearson Education, Inc. All rights reserved.

Example SQL Query

- `SELECT ZIP='43210'`

Name	First	Address	City	State	Zip	Airport
ADAMS	Charles	212 Market St.	Columbus	OH	43210	CMH
ADAMS	Edward	212 Market St.	Columbus	OH	43210	CMH
CARTER	Marlene	411 Elm St.	Columbus	OH	43210	CMH
CARTER	Beth	411 Elm St.	Columbus	OH	43210	CMH
CARTER	Ben	411 Elm St.	Columbus	OH	43210	CMH
CARTER	Lisabeth	411 Elm St.	Columbus	OH	43210	CMH
CARTER	Mary	411 Elm St.	Columbus	OH	43210	CMH

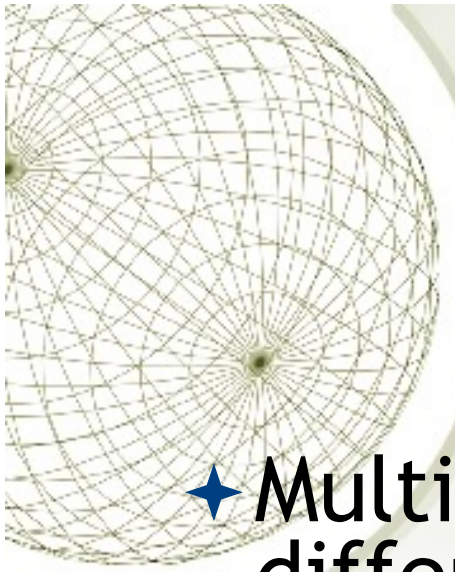
From *Security in Computing, Fifth Edition*, by Charles P. Pfleeger, et al. (ISBN: 9780134085043). Copyright 2015 by Pearson Education, Inc. All rights reserved.



Distributed Databases

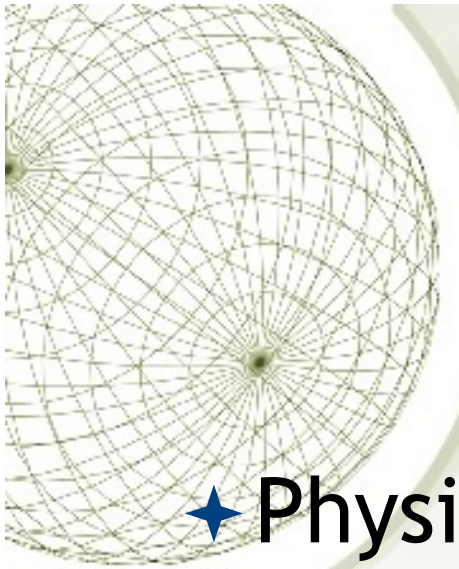
- PARTITIONED (e.g. Hadoop):
  remote CPUs (connected to host) have files unique to that site,
  e.g., records on local customers

- REPLICATED (DUPLICATED as a special case)
  ea. remote CPU has copies of common files
  e.g., layouts for standard reports and forms



Multilevel Databases

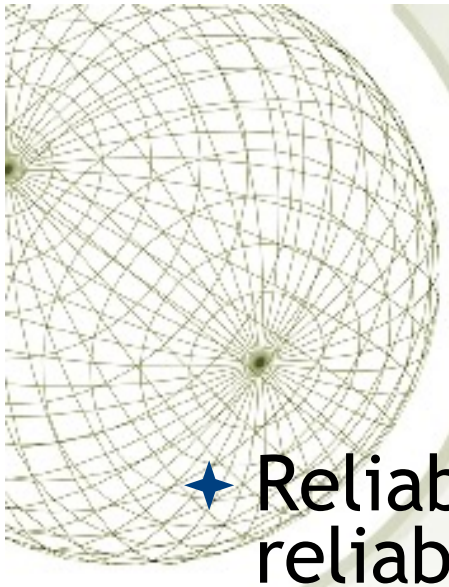
- Multilevel databases - store data with different sensitivity levels
  - e.g.: public, confidential, secret, top secret
  - Users are only allowed to access data up to a certain level



Database Security Requirements

- Physical integrity
- Logical integrity
- Element integrity
- Auditability
- Access control
- User authentication
- Availability

From *Security in Computing, Fifth Edition*, by Charles P. Pfleeger, et al. (ISBN: 9780134085043). Copyright 2015 by Pearson Education, Inc. All rights reserved.



Reliability and Integrity

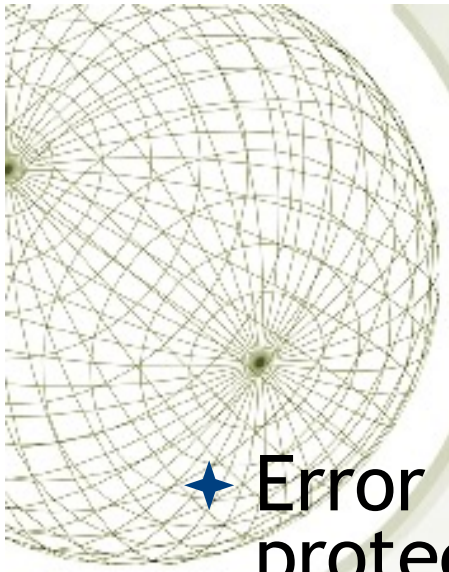
- Reliability: in the context of databases, reliability is the ability to run for long periods without failing
- Database integrity: concern that the database as a whole is protected against damage
- Element integrity: concern that the value of a specific data element is written or changed only by authorized users
- Element accuracy: concern that only correct values are written into the elements of a database

From *Security in Computing, Fifth Edition*, by Charles P. Pfleeger, et al. (ISBN: 9780134085043). Copyright 2015 by Pearson Education, Inc. All rights reserved.

Two-Phase Update

- Phase 1: Intent
  - DBMS does everything it can, other than making changes to the database, to prepare for the update
    - Collects records, opens files, locks out users, makes calculations
  - DBMS commits by writing a commit flag to the database
- Phase 2: Write
  - DBMS completes all write operations
  - DBMS removes the commit flag
- If the DBMS fails during either phase 1 or phase 2, it can be restarted and repeat that phase without causing harm

From *Security in Computing, Fifth Edition*, by Charles P. Pfleeger, et al. (ISBN: 9780134085043). Copyright 2015 by Pearson Education, Inc. All rights reserved.



Other Database Security Concerns

- Error detection and correction codes to protect data integrity
- For recovery purposes, a database can maintain a change log, allowing it to repeat changes as necessary when recovering from failure
- Databases use locks and atomic operations to maintain consistency
  - Writes are treated as atomic operations
  - Records are locked during write so they cannot be left in a partially updated state

From *Security in Computing, Fifth Edition*, by Charles P. Pfleeger, et al. (ISBN: 9780134085043). Copyright 2015 by Pearson Education, Inc. All rights reserved.

Sensitive Data

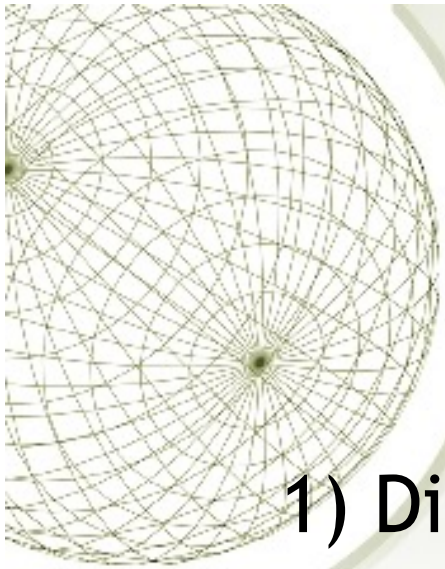
Why is data sensitive?

- Inherently sensitive
  - Passwords, locations of weapons
- From a sensitive source
  - Confidential informant
- Declared sensitive
  - Classified document, name of an anonymous donor
- Part of a sensitive attribute or record
  - Salary attribute in an employment database
- Sensitive in relation to previously disclosed information
  - An encrypted file combined with the password to open it

From *Security in Computing, Fifth Edition*, by Charles P. Pfleeger, et al. (ISBN: 9780134085043). Copyright 2015 by Pearson Education, Inc. All rights reserved.

Inference (Inference Problems)

- Inference attack - inferring *sensitive* data from *non sensitive* data
- The inference problem arises whenever some data x can be used to derive partial or complete information about some other data y, where y is classified higher than x. In some cases, even learning of the existence of the information may be unacceptable.
- Remember that you do want information to be available
- Even sensitive information should be available to users with the correct privileges



Types of inference attacks

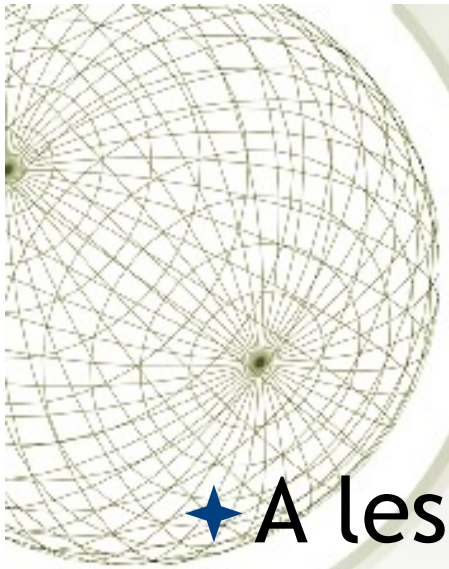
1) Direct attack
   - Infer sensitive data from results of queries run by attacker
   - *n*-item *k*-percent rule:
     - Data withheld if n items represent > k percent of the result reported
       - Most obvious case: 1-item 100-percent case: 1 person represents 100 % of results reported

Example Direct inference attack

- In a direct attack you try to extract sensitive information in one query:
  - LIST NAME where

    SEX=M ∧ DRUGS=1

    Might be rejected since it is querying directly for sensitive information (drug level)



Example Direct inference attack

- A less obvious direct attack could be:
  - LIST NAME where

    (SEX=M ∧ DRUGS=1) ∨
    (SEX≠M ∧ SEX≠F) ∨
    (DORM=AYRES)

Looks like it could give many records back concealing the sensitive information, but could in fact give the same result as the previous query.

Types of inference attacks, cont.

2) Indirect attack
- Infer sensitive info from statistics (Sum, Count, Median) also from info external to the attacked DB
- Tracker attacks (intersection of sets)
- Linear system vulnerability
  - Use algebra of multiple equations to infer

Example Indirect inference attack

- Given the following table over sums of financial aid by dorm and sex:

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

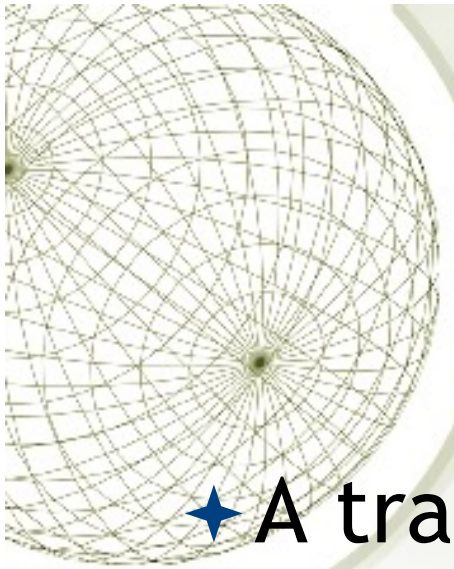
We can easily infer that any specific female student in Grey does not receive any financial aid. This however requires some outside information.

Example Indirect inference attack

- If we combine this with a table containing the count as follows:

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

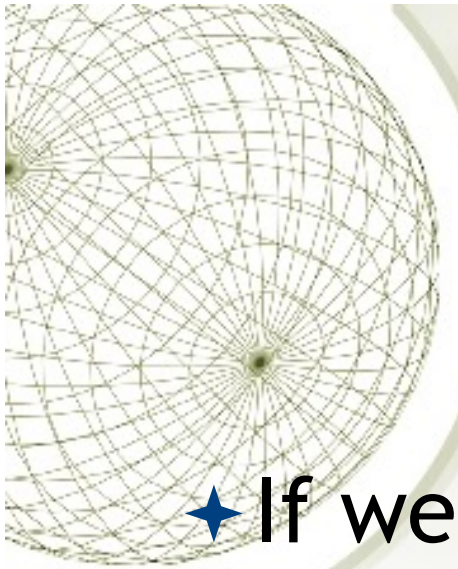
We can now infer that the two male students in Holmes and West are receiving aid in the amount of 5000 and 4000 respectively.



Example Indirect inference attack

- A tracker attack tries to fool the DBMS to give small amount of records back even if that is normally suppressed.
  - Count ((SEX=F) ∧ (RACE=C) ∧ (DORM=Holmes))

Might be supressed if the answer is 1, but…



Example Indirect inference attack

- If we instead make the logically equivalent queries
  - (Count (SEX=F)) – (Count ((SEX=F) $\wedge$ ((RACE$\neq$C) $\vee$ (DORM$\neq$Holmes)))

    = 6 – 5 = 1

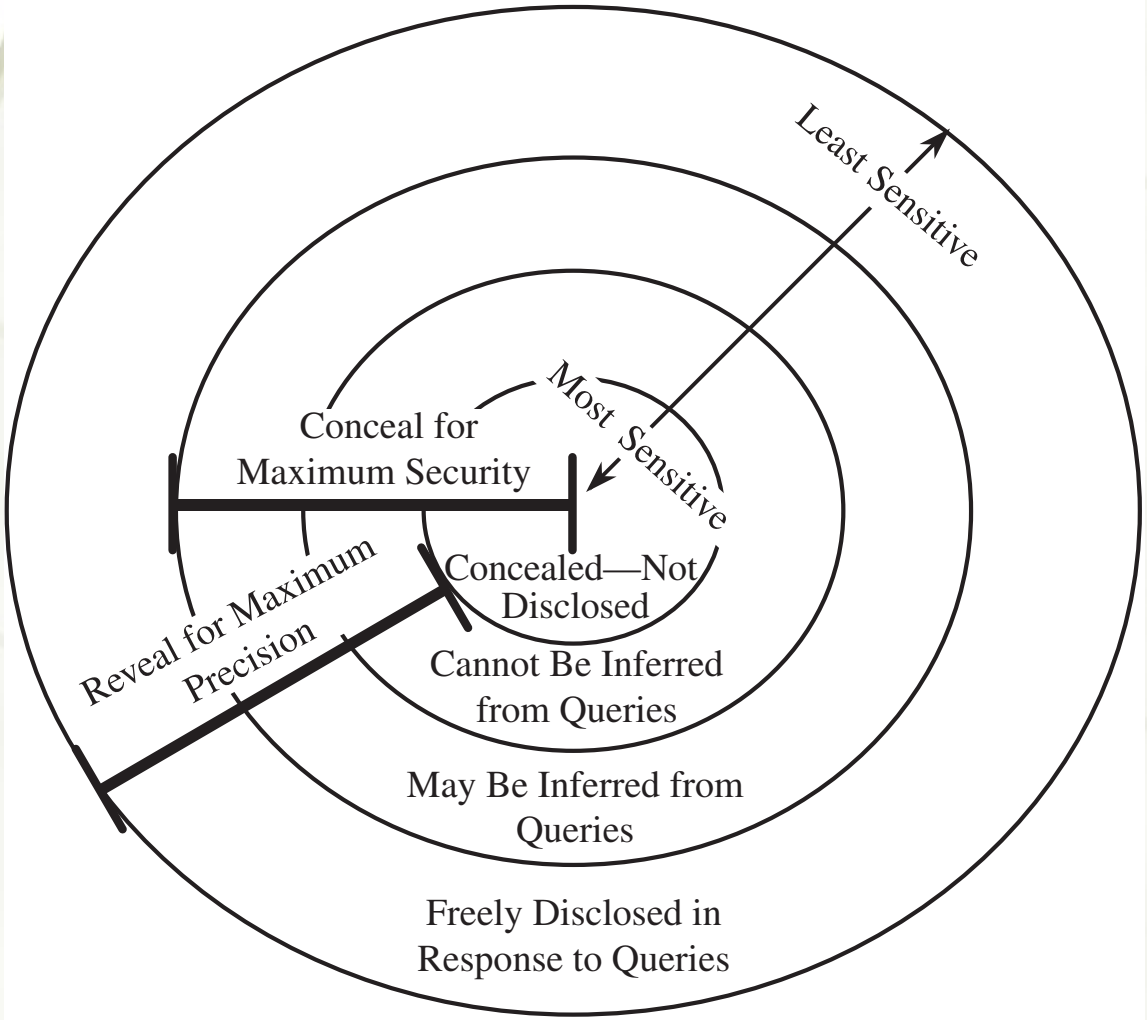
  We get the same result without breaking the *n*-item *k*-percent rule

Types of Disclosures

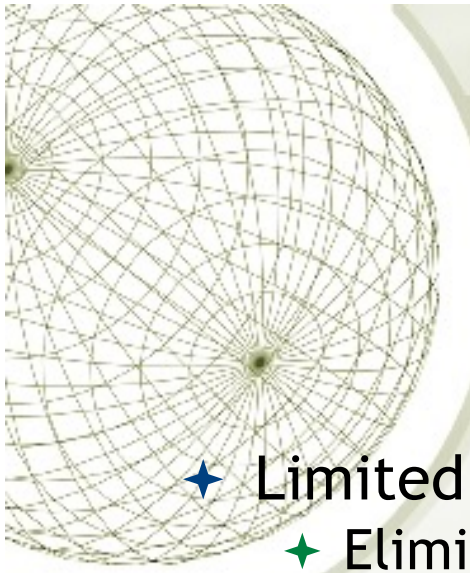
- Exact data
- Bounds
- Negative result
- Existence
- Probable value
- Direct inference
- Inference by arithmetic
- Aggregation
- Hidden data attributes
  - File tags
  - Geotags

From *Security in Computing, Fifth Edition*, by Charles P. Pfleeger, et al. (ISBN: 9780134085043). Copyright 2015 by Pearson Education, Inc. All rights reserved.

Security vs. Precision



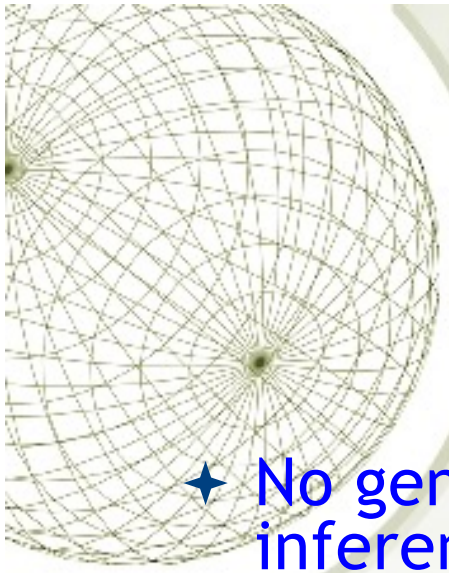
From *Security in Computing, Fifth Edition*, by Charles P. Pfleeger, et al. (ISBN: 9780134085043). Copyright 2015 by Pearson Education, Inc. All rights reserved.



Suppression Techniques

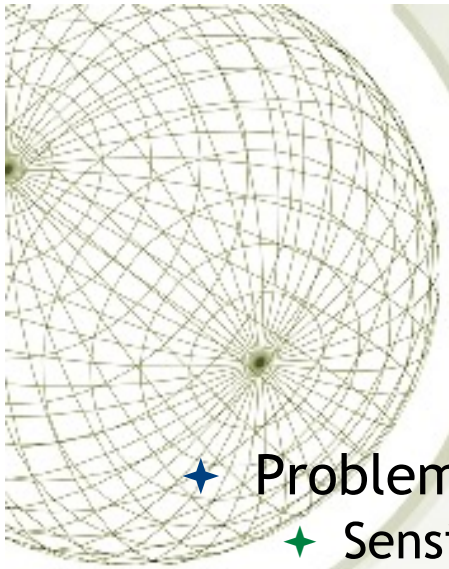
- Limited response suppression
  - Eliminates certain low-frequency elements from being displayed
- Combined results
  - Ranges, rounding, sums, averages
- Random sample
- Blocking small sample sizes
- Random data perturbation
  - Randomly add or subtract a small error value to/from actual values
- Swapping
  - Randomly swapping values for individual records while keeping statistical results the same

From *Security in Computing, Fifth Edition*, by Charles P. Pfleeger, et al. (ISBN: 9780134085043). Copyright 2015 by Pearson Education, Inc. All rights reserved.



Conclusions on Inference

- No general technique is available to solve the inference problems
- Need assurance of protection
- Hard to incorporate outside knowledge
- Optimal plan:
  - Suppress obviously sensitive information
  - Track what user knows (expensive)
  - Disguise data
- Aggregation—additional problem
  - Revealing sensitive data by aggregating large amount of data from different sources. A few data objects is not a problem, but knowing many may be.

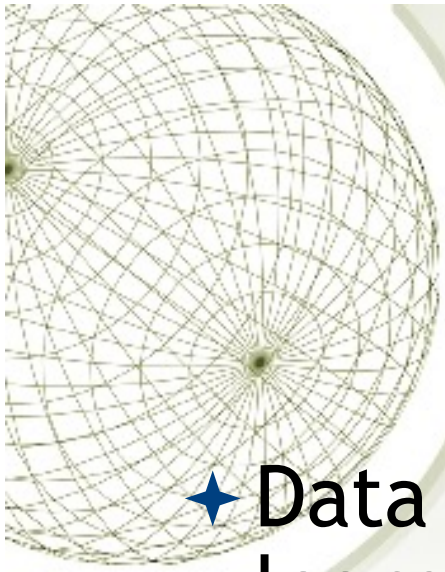


Multilevel Databases

- Problems
  - Sensitivity level on item level
  - Integrity (*-property)
  - Polyinstantiation – multiple (“poly”) instantiations of a record, each at a different security level
    - Example:
      - [John, Kalamazoo-MI] -- Public level
      - [John, 19_Main_Ave-Kalamazoo-MI] -- Confidential level
        …
      - [John, 19_Main_Ave-Kalamazoo-MI, …, SSN=123-45-6789] -- Top_Secret level
  - Global actions (i.e., backup)
  - Small items controlled
  - Cost and performance
  - Consumer resistance to military model

Proposals for Multilevel Security - Separation Mechanisms

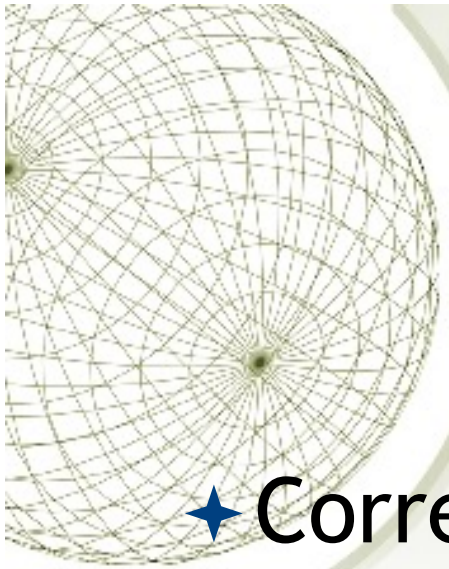
1) Partitioning
   - Redundancy
   - Accuracy (multiple field update)
2) Encryption per level
   - Cumbersome decrypting with queries
3) Integrity lock
   - Data item
   - Sensitivity level
   - Checksum
   - Cryptographic checksums
4) Sensitivity lock
   - Unique identifier
   - Sensitivity level



Data Mining and Big data

- Data mining uses statistics, machine learning, mathematical models, pattern recognition, and other techniques to discover patterns and relations on large datasets
- The size and value of the datasets present an important security and privacy challenge, as the consequences of disclosure are naturally high

From *Security in Computing, Fifth Edition*, by Charles P. Pfleeger, et al. (ISBN: 9780134085043). Copyright 2015 by Pearson Education, Inc. All rights reserved.

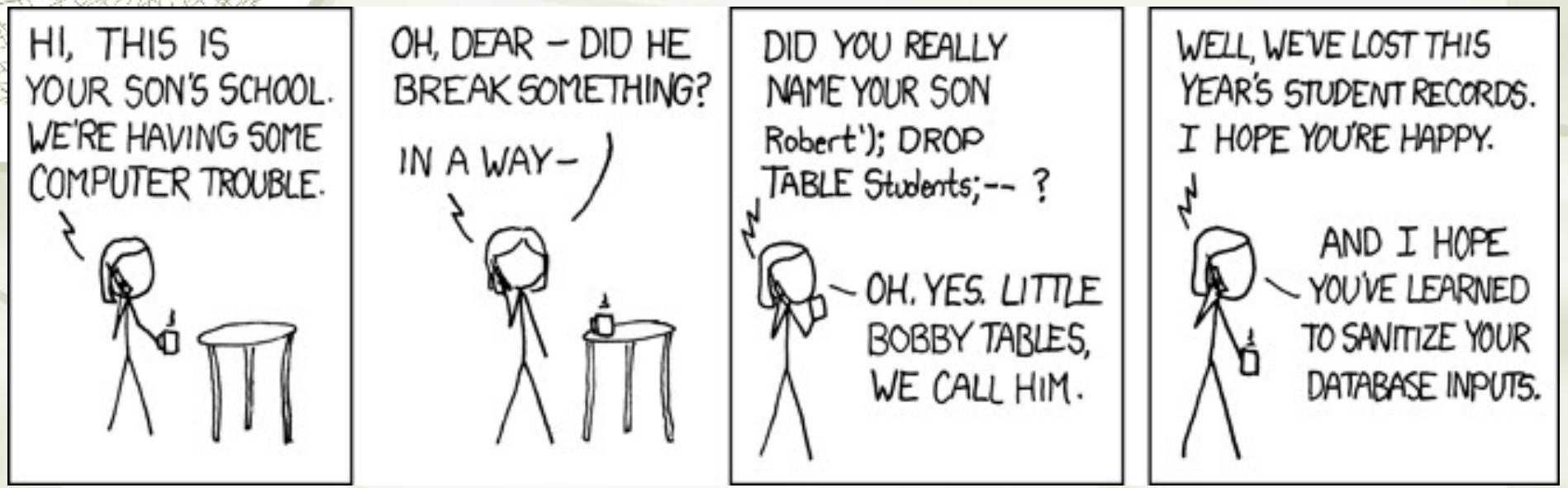


Data Mining Challenges

- Correcting mistakes in data
- Preserving privacy
- Granular access control
- Secure data storage
- Transaction logs
- Real-time security monitoring

From *Security in Computing, Fifth Edition*, by Charles P. Pfleeger, et al. (ISBN: 9780134085043). Copyright 2015 by Pearson Education, Inc. All rights reserved.

SQL Injection attacks



- The importance of sanitizing input cannot be underlined too much

For more info on SQL injection see https://www.w3schools.com/sql/sql_injection.asp