





NOAA Science Advisory Board Data Archiving and Access Security Considerations

Domi Sanchez, CISSP
NOAA OCIO/ITSec Office
December 7, 2006



Outline



- Introductions
- Issues/Challenges/Threats
- Database controls
- Data Warehousing
- Data Mining
- Integrity Models
- C&A
- Data Lifecycle Management



Entity Relationship Example

Security by Obscurity or Complexity undermines Security



3



Attack Terminology

- **Inference Attack** – Using unrelated events to infer activity
- **Covert Channel** – Unintended communication path between two resources that allows transfer of information in violation of security policy
 - Timing Channel
 - Storage Channel
- **Lack of Parameter Checking** – Failure to check the size of input streams, can allow Buffer Overflow
- **Maintenance Hook** – (trapdoor) allows maintenance of system bypassing security
- **Time of Check to Time of Use** – (TOC/TOU) attack that exploits the difference in time between time security applied and time that service is used.
- **Data Reuse** – The process of ensuring that no residue of a previous use of an object is available to a new user.



4



Database Challenges

- **Access to Information in a Database**
 - Release of information (Confidentiality attack)
 - Modification of information (Integrity attack)
 - Denial of service (Availability attack)
- **Discretionary vs Mandatory**
 - Specific authorization granted and denied
 - Authorization based on assigned classification

5



Database Issues

- Verifying the right access was granted - DAC
- Verifying View limitations function - MAC
- Preventing users from creating a copy (becoming “owner”)
 - Granting access to others
 - Escalating privileges

6



Database Controls



- “Granularity” - fineness of control permissible in database controls - dependent upon database and implementation
- Grant/Revoke Privileges by Table, Column, Key set
- Permissions by View combining specific Tables, Columns, Key sets
 - Conceptually dividing the database into pieces to allow sensitive data to be hidden from unauthorized users
 - Authorizations for specific views having specific attributes, and for actions to perform within those views
 - DAC, by specific grant to user or group by owner
 - MAC, by classification level

7




Database Controls




- Biggest issue still mistakes, omissions
- Protection by operating system/platform/db
 - Physical data base integrity
 - Logical data base integrity
 - Element integrity
 - Auditability
 - Access control
 - User authentication
 - Availability
- Two-phase update (Intent/Commit and Lock)
- Error detection/correction
- Integrity with multiple instances (polyinstantiation)

8




Data Warehouse




- A data warehouse can be defined as any centralized data repository which can be queried for business benefit
- Warehousing makes it possible to:
 - extract archived operational data
 - overcome inconsistencies between different legacy data formats
 - integrate data throughout an enterprise, regardless of location, format, or communication requirements
 - incorporate additional or expert information

(http://www.pcc.qub.ac.uk/tec/courses/datamining/ohp/dm-OHP-final_2.html#HEADING4)

9




Data Warehouse Characteristics




- subject-oriented - data organized by subject instead of application e.g.
 - an insurance company would organize their data by customer, premium, and claim, instead of by different products (auto, life, etc.)
 - contains only the information necessary for decision support processing
- integrated - encoding of data is often inconsistent e.g.
 - gender might be coded as "m" and "f" or 0 and 1 but when data are moved from the operational environment into the data warehouse they assume a consistent coding convention
- time-variant - the data warehouse contains a place for storing data that are five to 10 years old, or older e.g.
 - this data is used for comparisons, trends, and forecasting
 - these data are not updated
- non-volatile - data are not updated or changed in any way once they enter the data warehouse
 - data are only loaded and accessed

10




Data Mining




- “The non-trivial extraction of implicit, previously unknown, and potentially useful information from data”
 - William J Frawley, Gregory Piatetsky-Shapiro and Christopher J Matheus
- “The science of extracting useful information from large data sets or databases”
 - D. Hand, H. Mannila, P. Smyth
- The data is often voluminous, but as it stands of low value as no direct use can be made of it; it is the hidden information in the data that is useful
 - Clementine User Guide




11



Data Mining Concerns




- There are many legitimate uses of data mining
 - For example, a database of prescription drugs

data

data



data



ISSUES


- Misuse
- Scope Creep
- Privacy
- Ethics
- Legality

12




Data Mining Concerns


- Another example: Trade Magazines




COMPUTER WORLD
“what **year** were you born?”



NETWORK WORLD
“what **month** were you born?”





PC WORLD
“what **day** were you born?”



Full name
Address
Email
Phone #
Date of Birth

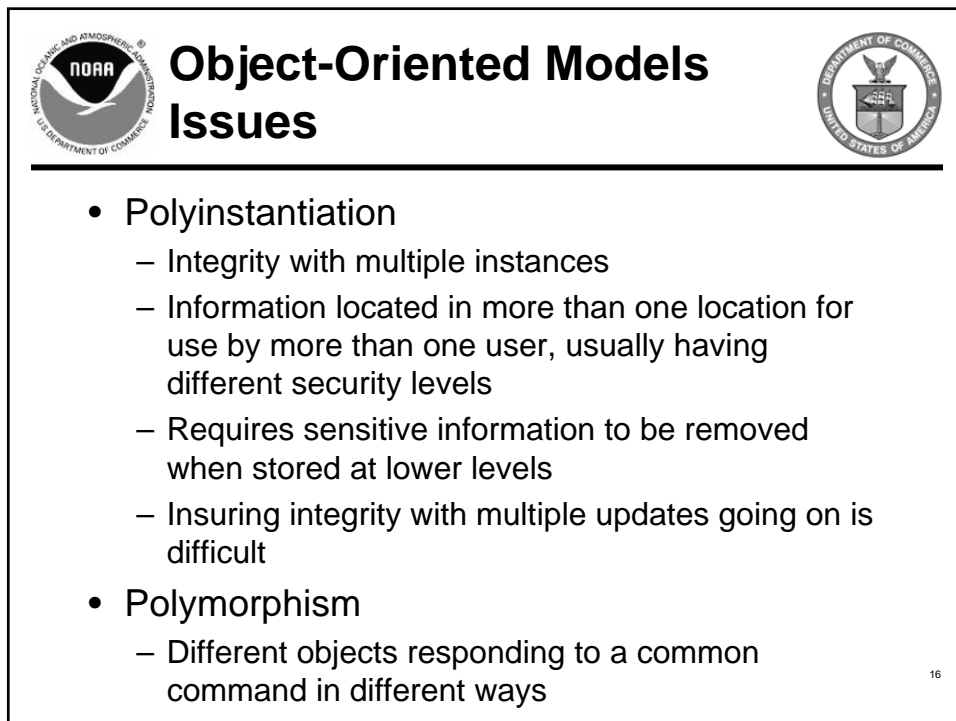
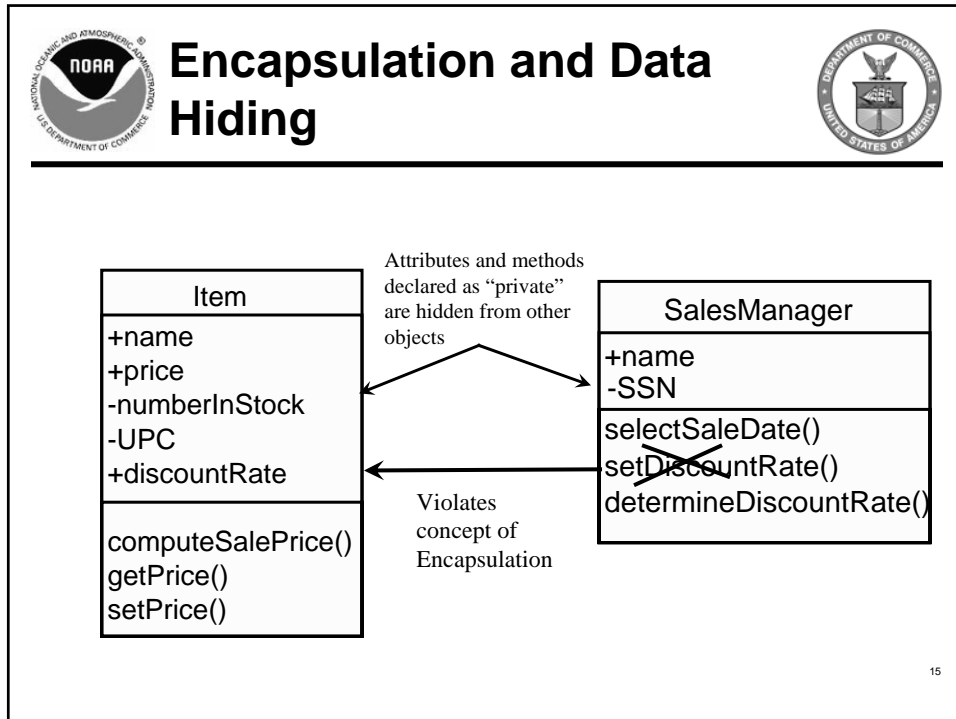
13

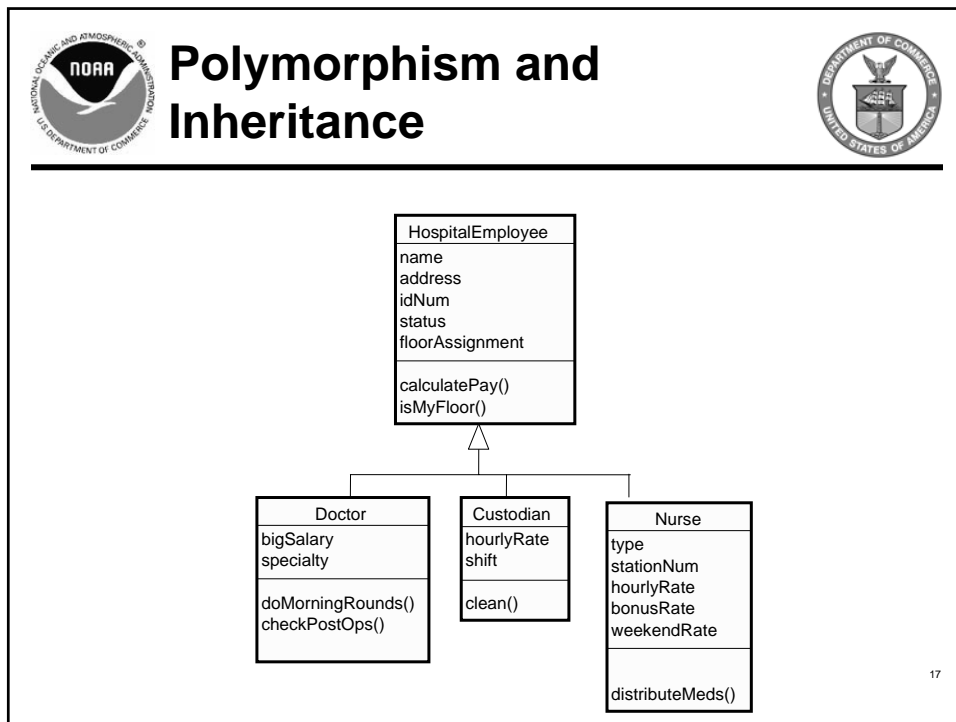


Object-Oriented Concepts

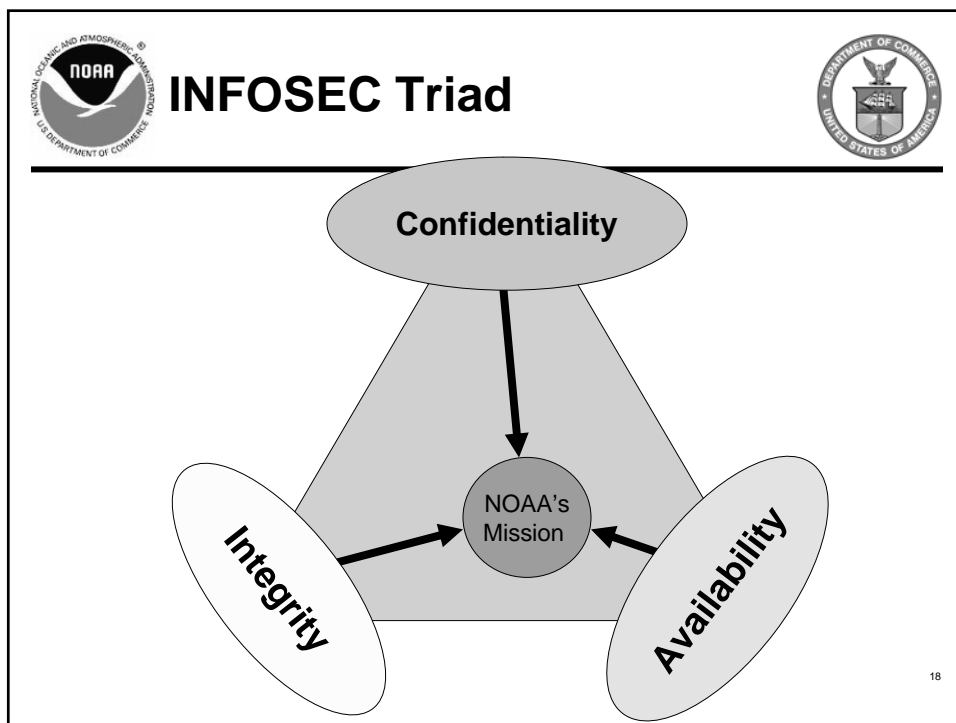
- **Subjects**
- **Objects**
- Controls using
 - Encapsulation,
 - Inheritance,
 - Information hiding
- **“Granularity”** - fineness of control permissible in database controls

14







17



18




Integrity Model Goals




- Making sure information is consistent between internal and external systems.
- Five models identified by the National Computer Security Center Report:
(<http://140.229.33.203:9000/htdocs/teinfo/cim/national.html>)
 - Biba
 - Gougen-Meseguer
 - Sutherland
 - Clark-Wilson
 - Brewer-Nash

19




Integrity



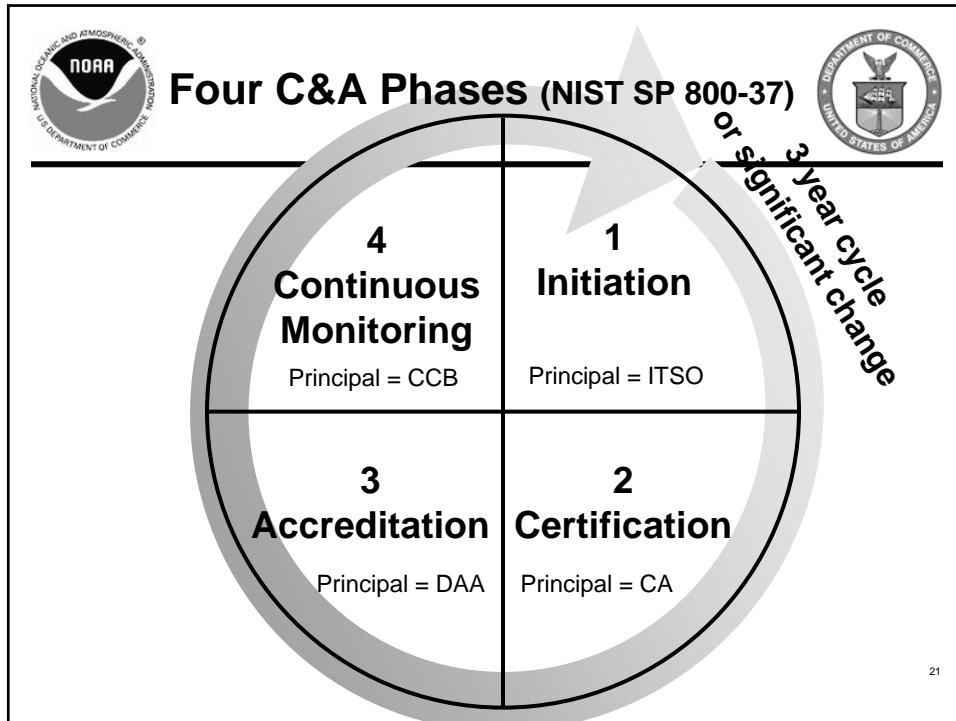
- The Brewer-Nash (AKA Chinese Wall) Model
 - Published in 1989
 - Access controls can change dynamically
 - Set to reduce conflict of interests
 - Database-centric
 - Uses basic Mathematical theories
 - May allow for confidentiality if objects do not conflict with fair competition.

SUBJECT



OBJECT of competitor

20





Sensitivity vs Criticality

- First defining the **Sensitivity** and **Criticality** levels of your data/systems will drive security
 - Sensitivity – a Confidentiality issue. Examples:
 - Data centric {
 - Public access
 - Sensitive But Unclassified (SBU)
 - Classified as “Secret” or “Business Confidential”
 - Classified as “Top Secret” or “Business Proprietary”
 - Criticality – an Availability issue. Examples:
 - GSS/MA centric. {
 - FIPS 199 - Low
 - FIPS 199 - Moderate
 - FIPS 199 - High

∴ C&A

22





Three data states


“Information at rest”

Stored

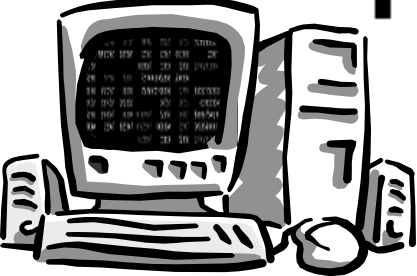
- Electronically
- Physically





Transmitted



Processed




23


Processing of data

- Smaller chunks of data are easier to secure
 - But has higher processing overhead
- Authorized people making unauthorized changes
- Unauthorized people making authorized changes

24




Transmission of Data




- Journaling - A journaling (or journalling) file system is a file system that logs changes to a journal (usually a circular log in a specially-allocated area) before actually writing them to the main file system.
 - http://en.wikipedia.org/wiki/Journaling_file_system
- Vaulting - The process of sending data off-site, where it can be protected from hardware failures, theft, and other threats.
 - http://www.webopedia.com/TERM/D/data_vaulting.html

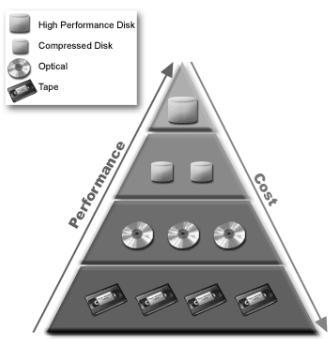
25



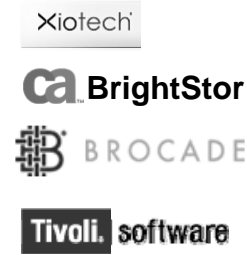
Data Lifecycle Management



- Data lifecycle management (DLM) is a policy-based approach to managing the flow of an information system's data throughout its life cycle: from creation and initial storage to the time when it becomes obsolete and is deleted. <http://searchstorage.com>

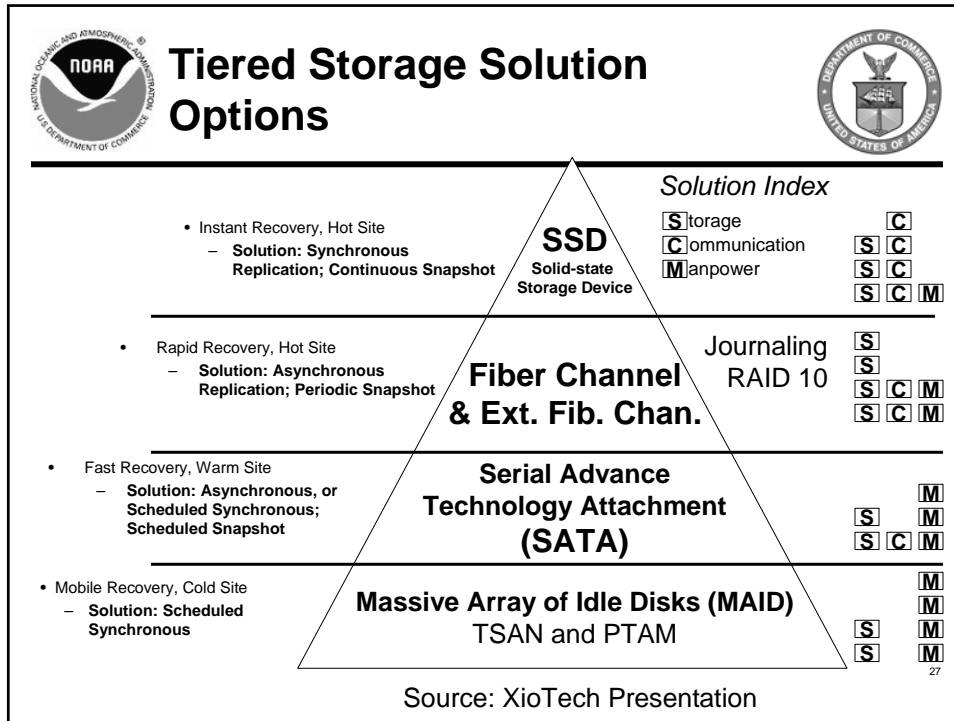


The diagram shows a pyramid with 'Performance' on the left side (increasing upwards) and 'Cost' on the right side (increasing downwards). The pyramid is divided into four horizontal layers, each containing icons for different storage technologies: High Performance Disk, Compressed Disk, Optical, and Tape.



Logos for XioTech, BrightStor, Brocade, and Tivoli software.

26



Conclusion

- Identify and mitigate vulnerabilities
- Understand where your threats originate
- Define data sensitivity
- Apply the right level of technology to the business need
- Address security in all 3 data states
- Address security throughout data lifecycle

