

A Grand Challenge: Using **X to Improve Data Quality**



Leszek Lilien

The Center for Education and Research in Information Assurance and Security (CERIAS)

and

Department of Computer Sciences

Purdue University

Components of Data Quality



⌘ Data quality includes:

[Irvine & Levin 2000]

☑ data integrity

☑ data confidentiality

Is Data Quality Less Important Than Car Quality?



Why do we demand more of the quality of food or a car than we demand of that other essential: information?
[...]

Reliability and authenticity [...] are essential for our information society in its dependence on trust in information.

[Eric Ketelaar, "Can We Trust Information," 1997]

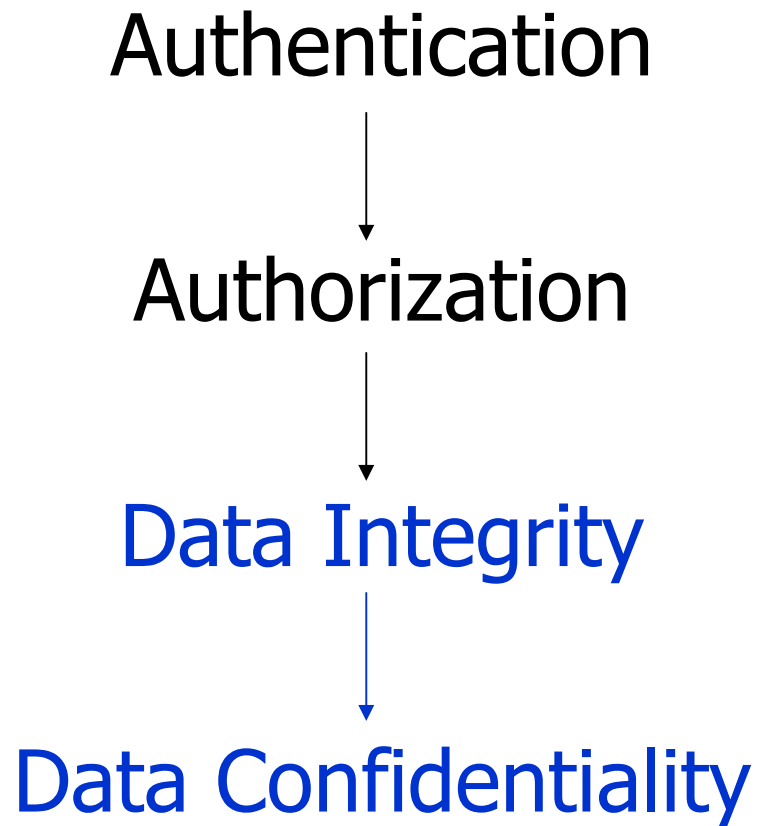
Independence of Data Integrity and Confidentiality

[Irvine & Levin 2000]

integrity confidentiality	low	high
low	childrens' boasts on the playground	Oxford English Dictionary
high	potentially critical information from an unreliable source	a trade secret

The Chain of Trust: **Dependence** of Data Integrity and Confidentiality

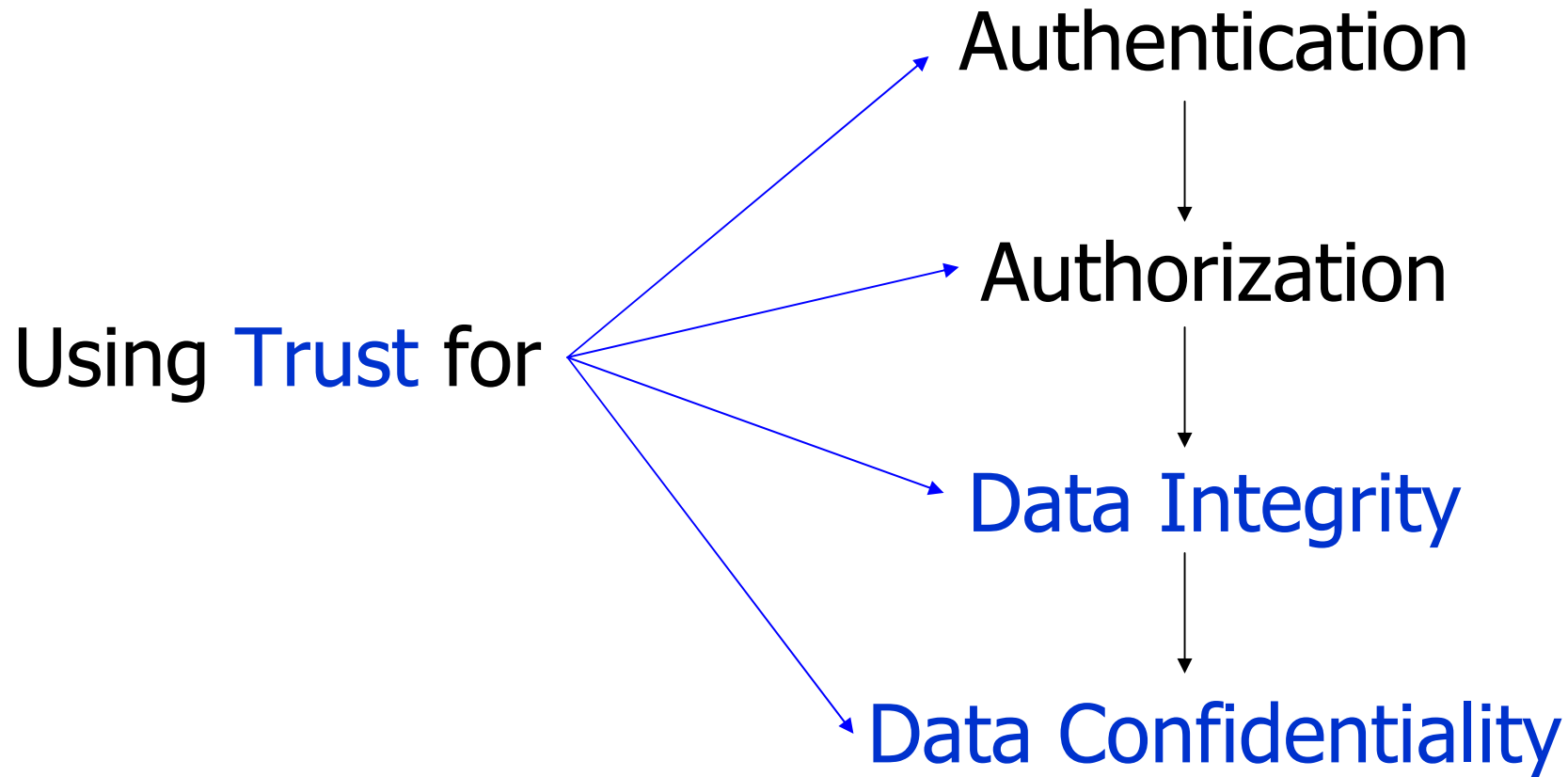
[Walker & Sundaralingam, 2002]



$X \rightarrow Y$ means:
-- Y must trust X
OR:
-- no Y without X

Trust²

(Using Trust in the Chain of Trust)



Revisiting: *“A Grand Challenge: Using **X** to Improve Data Quality”*

⌘ **X** = Trust

⌘ In my opinion,

Using Trust to Improve Data Quality

is one of the Grand Challenges
in Data Integrity and Data Quality
