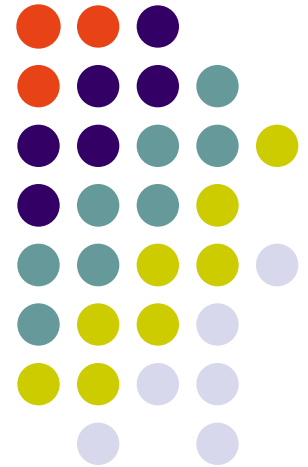


Access control to data based on concepts of data flow history and risk

By Luigi Logrippo
Joint work with Sofiene Boulares
and Kamel Adi



Université du Québec en Outaouais



Abstract

- Access control systems allow or deny access of subjects to data objects according to policies. In some organizations, policies are expressed by using fixed clearance levels of subjects and classification levels of objects (e.g. from TopSecret to Unclassified). In others, they are expressed in consideration of roles, and the role structure is supposed to be fairly stable (e.g. role Doctor has different permissions than role Nurse). The talk will start by reviewing some of these concepts. The study of examples will lead us to the conclusion that more flexible access control systems are needed, allowing permissions to change in time, according to risk considerations determined by data flow histories. The use of such systems is important in dynamic environments such as the Web and the Cloud, where subject-object relationships and data flows vary rapidly over time.
- We will then illustrate access control mechanisms where the security levels of subjects and data objects, as well as the risk of allowing subjects to access data objects, and thus access authorizations, are determined dynamically by information flow history. In other words, the security levels of subjects and objects in these systems are not fixed, but are determined by the importance of the data that has flown to them over time. A subject can be authorized to access an object if the access is not considered to be risky, in consideration of what data the subject already knows and what data the object already contains.
- The bulk of this work is from the forthcoming PhD thesis of Sofiene Boulares, UQO 2016

Access Control Systems



- They *guard resources*
- When an *identified user* demands to access a resource, Access Control System will decide whether to grant access or not
 - Resources: very generic term
 - In this presentation, we shall consider
 - *access for reading or writing data resources only*



Confidentiality, Integrity

● Confidentiality:

- Protected data shall not escape
 - Will only be read by authorized subjects
 - ❖ Secrecy is a synonym




● Integrity:

- Protected data shall not be modified
 - Will only be written by authorized subjects



Solutions for access control

- Established solutions:
 - Access decisions consider only
 - subject, object, environment
- Risk-based solutions:
 - Access decisions consider 
 - subject, object, environment and **risk**

Risk-based access control solutions

- Conventional solutions are rigid, decisions are pre-determined on the basis of subject, object and environment conditions
- Risk-based solutions take into consideration *evolving measurements* of risk
 - **These can be determined by many factors**
 - **These solutions are important in Web and Cloud environments where data flow is intense and risk must be evaluated on an ongoing basis**

Risk concept

- Risk can have many meanings
 - **Risky *subject***
 - E.g.: Access to a resource should only be granted to subjects that are *reliable* (how to measure?)
 - **Risky *environment, or situation***
 - E.g.: Door to fire extinguisher can be opened only in situation of fire
 - ❖ “Break glass” policies
 - **Risky *object*:**
 - E.g.: Restrict access in proportion of the importance of the contents of the object

Variety of risk-based solutions

- Many aspects of risk can come into consideration for access control
- The solution to be proposed here is one of many
- It evaluates *subject and object risk levels* as a consequence of data flow
- Environment conditions, not included in our approach, will then be used to determine *acceptable levels*

Basic Assumptions

- *Security levels of subjects and sensitivity levels of objects* have been previously assessed at initial values
 - From high to low
- They can change as a result of *data flow*
- A Read action creates data flow from an object to a subject
- A Write action creates data flow from a subject to an object
- Subjects can **increase** their security levels as they acquire data from higher levels
- Objects can **increase** sensitivity as they receive data from higher levels
- The *number of accesses* to different objects can also be important

Our risk concept

- Objects likely to contain important information should be protected for:
 - **Confidentiality: or disclosure**
 - **Integrity: or corruption**
- The risk of reading and writing operations is determined by *comparisons of levels* between subject and object
- Reading and writing operations cause *data transfers that modify the levels*

Confidentiality: Motivation

- In organizations, data bases are often classified for their 'importance' or 'sensitivity'
 - **High risk when allowing reading sensitive data**
- Data flow: as data flow from 'sensitive' data bases to 'less sensitive' data bases, the importance of the receiving data bases increases
 - **The risk of granting read access increases**
 - **So, need to keep track of the data flows in order to determine risk**

Example (confidentiality)

- Data Base A: employee list
 - Risk of allowing read access is low
- Data Base B: salary data
 - Risk of allowing read access is high
- If contents of Data Base B is allowed to flow to Data Base A
- Then risk of allowing read access to Data Base A becomes high

Integrity: Motivation

- In organizations, data bases are often classified for their levels of integrity
 - **High risk when allowing writing on high-integrity data bases**
- Data flow: as data flow from 'low integrity' data bases to 'high integrity' data bases, the integrity of the latter decreases
 - **The risk of granting access decreases**
 - **Need to keep track of the data flows in order to determine risk**

Example (integrity)

- Scientific journal A is known to contain very reputable papers
 - Risk of adding new papers is high, screening is rigorous
- Journal B is much less reputable
 - Risk of adding new papers is lower
- Journals A and B merge into journal C
- C will be less reputable than A!
 - Risk of adding new papers to journal C is lower than to journal A
 - Screening will probably be less rigorous

Duality of confidentiality and integrity

- For some types of data *confidentiality* is more important, for others *integrity* is more important
- The reasoning for confidentiality and integrity is dual
- What applies to confidentiality, also applies to integrity but with reversed reasoning!
- Thus *the following presentation will focus on **confidentiality***, and the rules for **integrity** can be easily understood by dual thinking

Historical Background

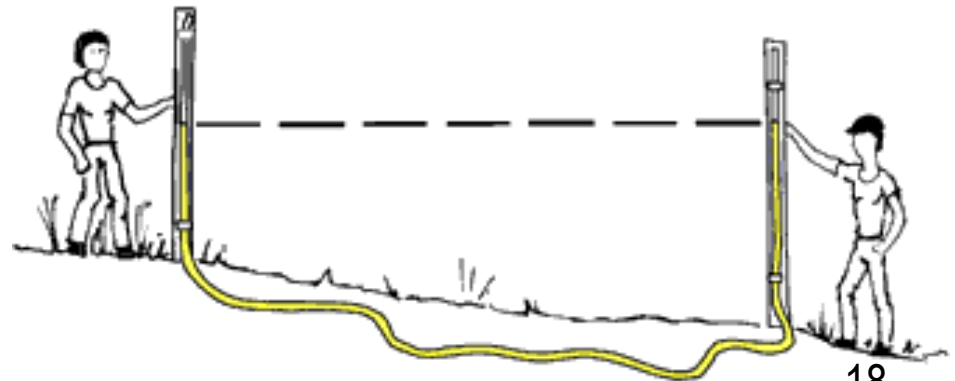
MAC

Mandatory Access Control data security models

- These are characterized by the fact that subjects and objects are labelled
 - **Subjects are labelled by the data that they can read**
 - **Objects are labelled by the data that they can contain**
- Thus there are label-based rules that determine
 - **Which subjects can read which objects**
 - **Which subjects can write on which objects**
- These rules can be understood to be determined by risk factors

High water mark: a first solution for confidentiality

- The “high water mark” model postulates that
 - when the contents of a **high** confidentiality data base
 - is allowed to flow to a **low** confidentiality data base,
 - then this second data base must be reclassified **high**
 - (Weissman, 1969)
 - Thus the **risk** of allowing reading from this database goes up
- This is a good start, but it can be refined



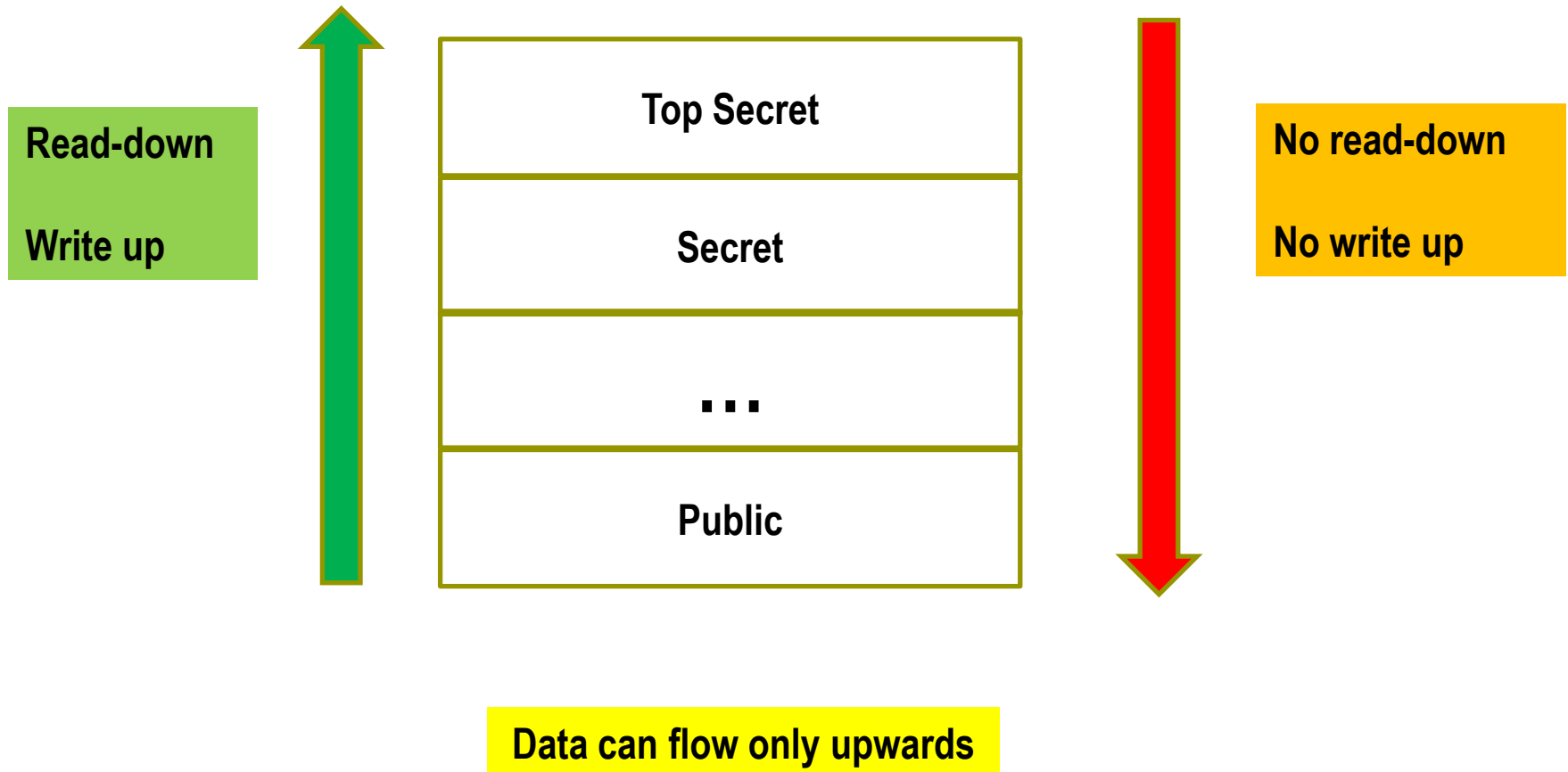
Refinements of High Water Mark

- Consider several coexisting levels of confidentiality
 - **Secret, Confidential, Public ...**
 - **We must consider several levels**
- Consider flows involving several objects from the same level ...
 - **One, two, three data sets of level Secret were moved to level Public**
 - **Progressive damage has been done!**

The Bell-La Padula model (BLP)

- Classical model for flow control in organizations
- Subjects and objects are classified by levels of confidentiality:
 - E.g. **Unclassified, Confidential, Secret, TopSecret**
 - **Data can only move up in this hierarchy**
 - **Writing: only upward**
 - **Reading: only downward**

The Bell-La Padula model



Re-interpretation of BLP as a risk model

- No risk for moving information up
 - Read down is allowed
 - Write up is allowed
- Risk for moving information down
 - Read up is forbidden
 - Write down is forbidden
- How can this risk be quantified?

Biba model: an integrity model

- The converse of BLP
- E.g. directives should flow down without corruption

Chinese Wall

- In Chinese Wall, confidential data are classified by the organization to which they belong
- Some organizations are in Conflict of Interest Col,
 - No subject nor object is allowed to combine data that are in Col
 - E.g. no subject nor object should be allowed to combine
 - ❖ data that is Classified to Bank A with
 - ❖ data that is Classified to Bank B
 - This is because certain dangerous data transfers and inferences can be done by combining the two data sets
- **No risk** combining data that are not in Col
- **Risk** combining data that are in Col

CW violation

**Before:
No Risk**



**After:
Risk!**



Risk in the existing models

- Seen in this way, existing MAC security models consider risk, but only on a binary scale:
 - **Risk or no risk**
- These models are considered too rigid and have found limited application
- But organization continue to classify their data!
- Hence our model ...

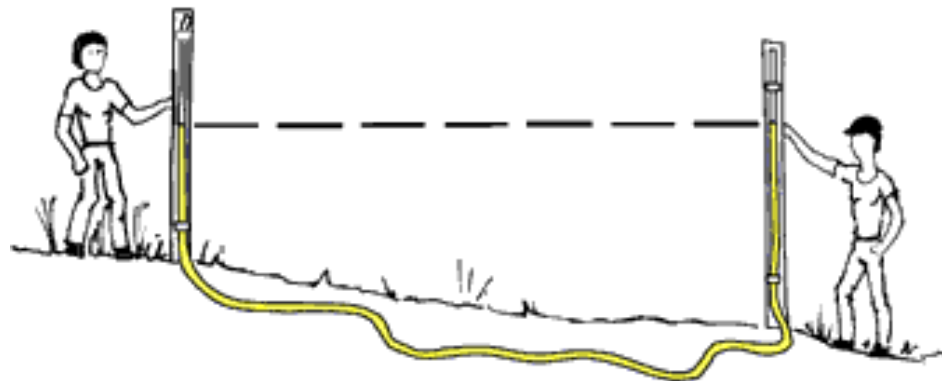
Intuitive basis of our method

Dual situations

- Data can flow either
 - **By a subject reading from an object or**
 - **By a subject writing on an object**
- If confidentiality is being considered then:
 - **Upwards flows are not risky, always allowed**
 - **Downwards flows are risky, allowed only if risk can be tolerated**
- If integrity is being considered then:
 - **Downwards flows are not risky, always allowed**
 - **Upwards flows are risky, allowed only if risk can be tolerated**
- So this gives four cases to consider
 - **Confidentiality when reading or writing**
 - **Integrity when reading or writing**
- Unfortunately the reasoning is almost identical in all cases, so the theory tends to be repetitious
 - **Hopefully we'll learn how to describe it more succinctly**

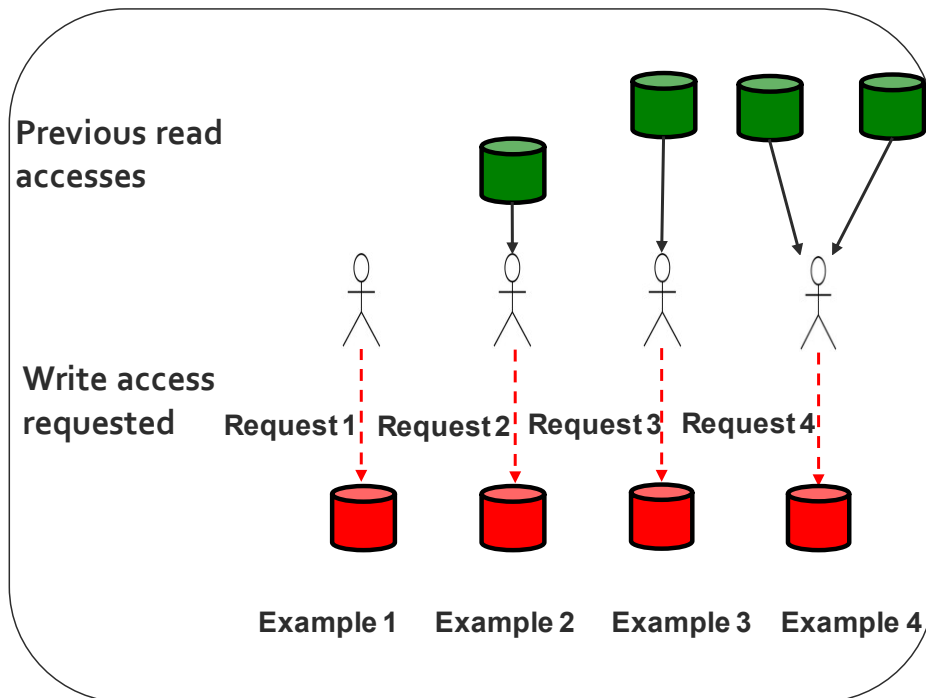
Main goal of our model

- An extension of the High Water Mark model
- The level of an object is determined by the levels of the data that have flown to it
- This also determines the risk of accessing the object



Increments in subject's security levels by reading

In established MAC systems, the result of a request is determined by fixed confidentiality levels of subjects and objects



In our work, considering data flows, subject confidentiality levels change.

In these examples, the subject progressively increases its confidentiality level by:

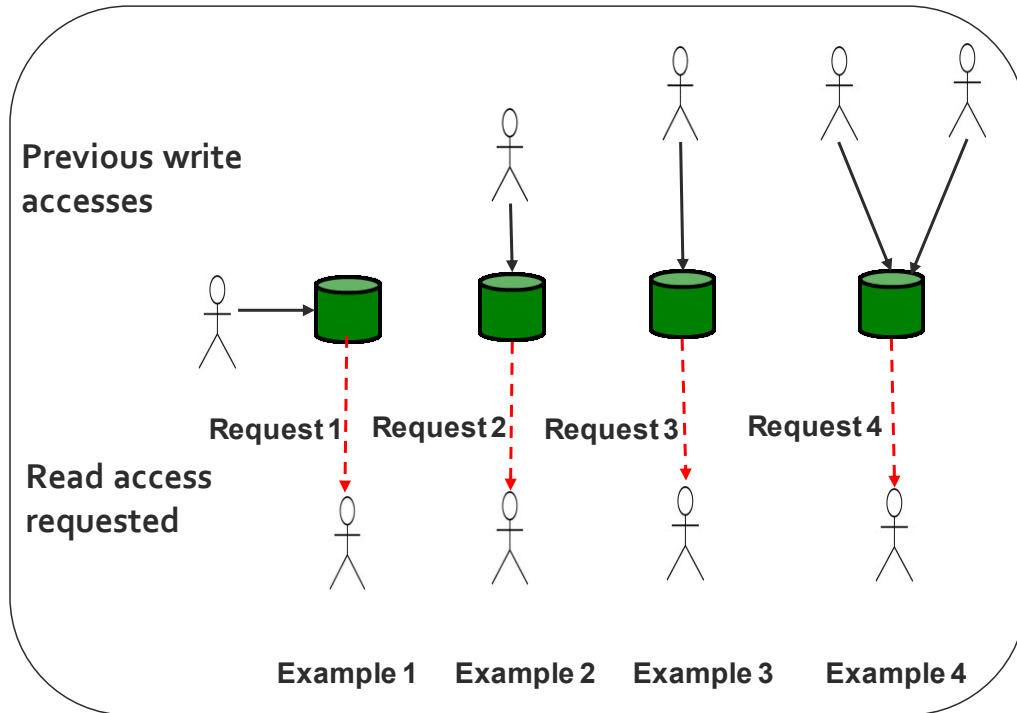
- Reading from an object in Ex. 2
- Reading from a more highly classified object in Ex. 3
- Reading from two different highly classified objects in Ex. 4

Changes in subject's level will change its clearances

Properties postulated for confidentiality

- ***Property:*** The confidentiality level of subjects increases as they read objects with higher confidentiality levels
- The number of objects read must be considered along with the difference in levels
- Pessimistic hypothesis:
 - **When a subject reads an object, it can take everything**
 - **And it will never forget what it has taken!**

Increments in object's confidentiality level by writing



Considering data flows, object confidentiality levels change

In these examples, the object progressively increases its confidentiality level by:

- Being written by a more highly classified subject in Ex. 2
- Being written by a yet a more highly classified subject in Ex. 3
- Being written by two highly classified subjects in Ex. 4

Access control decisions should take these changes into consideration

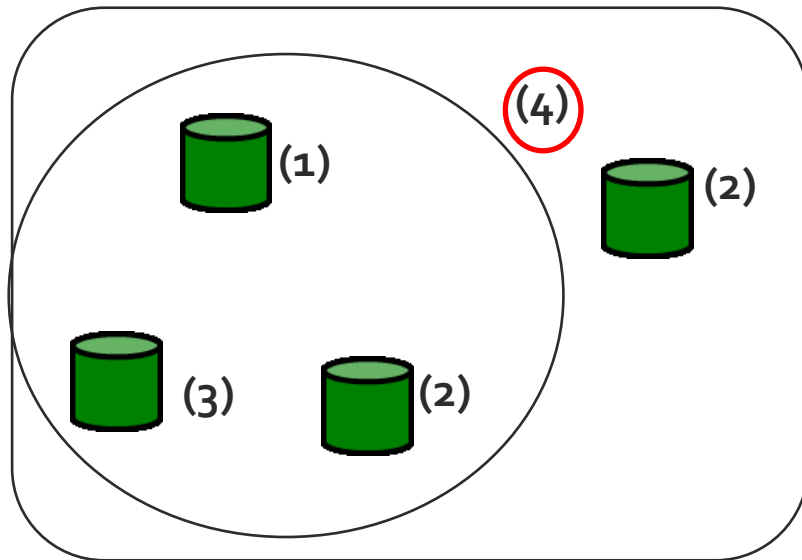
Properties postulated

- ***Property*** : The confidentiality level of an object increases as it is written on by subjects of higher confidentiality levels.
- The number of subjects having written must be considered along with the difference in levels
- Pessimistic hypothesis:
 - when a subject writes on an object, it can write there everything it knows
 - written data can stay there forever

Inferences from aggregation and association

- Example of aggregation :
 - Knowledge of placement of a single warship may have a limited value
 - Knowledge of 10 will probably have much higher value than 10 times the value of one
- Example of association:
 - The list of the employees in a company with their ranks can be of limited sensitivity
 - The list of salaries by rank can be more sensitive
 - Associating the two lists leads to a major privacy breach, with high risk

Considering inferences



Separately, three files have sensitivity classification 1, 2, and 3.

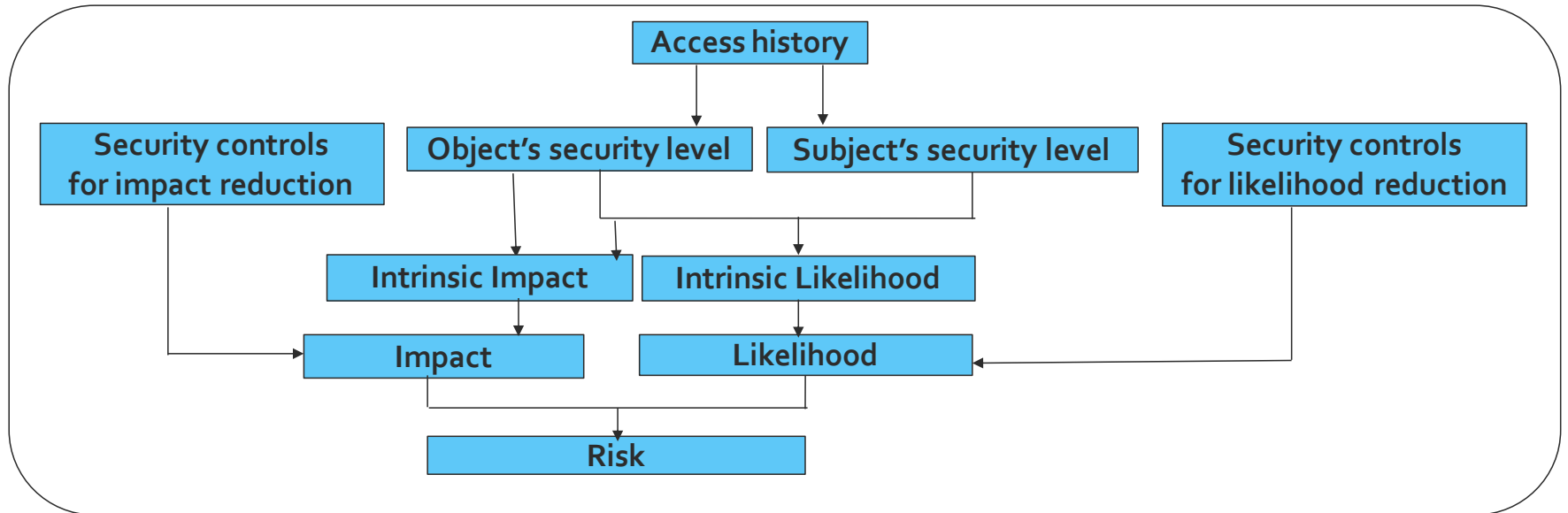
The sensitivity of the information that can be obtained by inference from their combined contents is 4

This should change what a subject of level 2 can contain

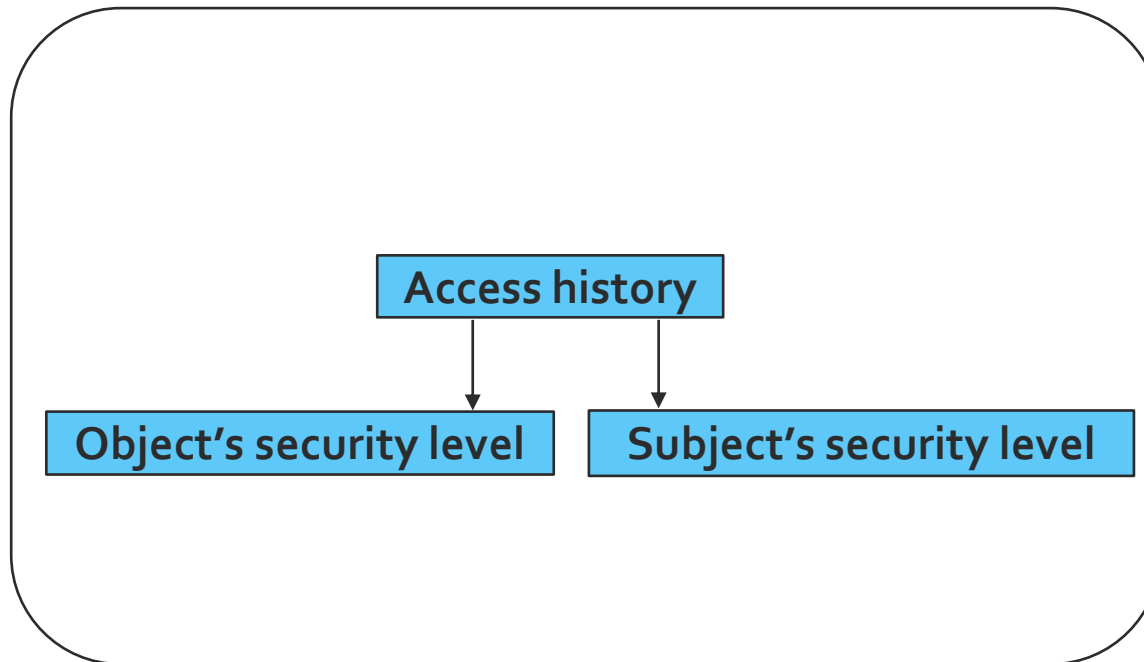
What we have seen so far

- Established access control systems use pre-defined confidentiality attributes, such as levels, to decide on access requests
- These confidentiality attributes can be seen as parameters to evaluate the risk of accesses
- This method can be made more flexible by calculating confidentiality attributes as a function of what 'information' subjects and object have already acquired
- And so the risk can be calculated as a function of acquiring new information by new accesses
- All this requires a mechanism for keeping track of access history, to be described

Overall view of our risk evaluation method



Schematically



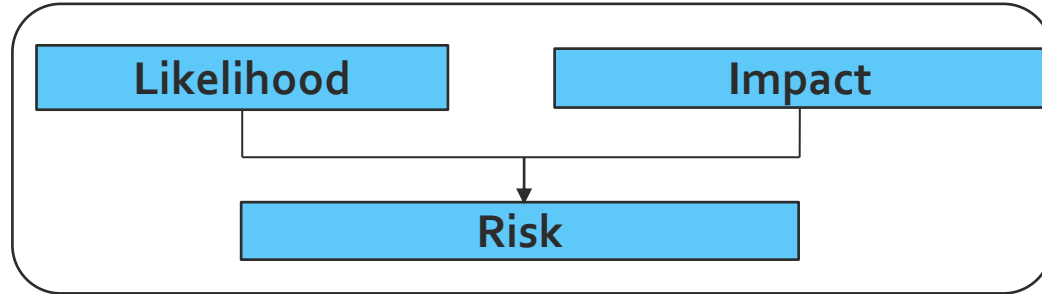
But this is not enough to
achieve a useful risk
assessment.

Standard alignment

- This view allow us to align ourselves with leading standards that define the risk of a data access as a function of the *impact* and the *likelihood* of an event
 - <http://csrc.nist.gov/publications/nistpubs/800-30/sp800-30.pdf>
 - <shttp://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r4.pdf>
 - <http://www.clusif.asso.fr/fr/production/ouvrages/pdf/MEHARI-2010-Principles-Specifications.pdf>

Risk assessment

according to NIST, Mehari, others



$$\text{Risk} = \text{Likelihood} \times \text{Impact}$$

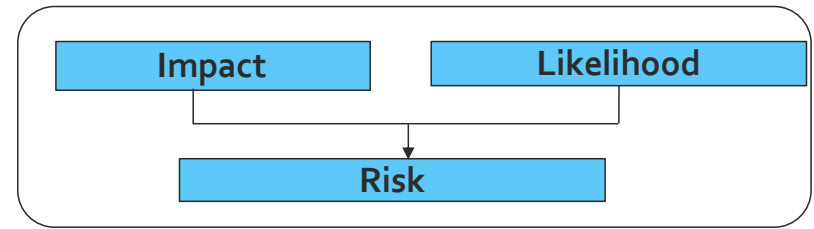
Impact is a function of :

- Subject and Object's security levels
- Security controls for impact reduction

Likelihood is a function of :

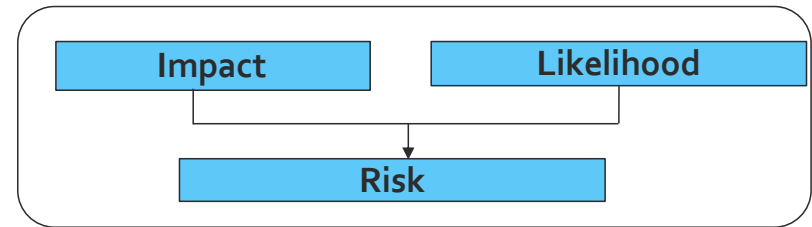
- Subject's and Object's security levels
- Security controls for likelihood reduction

Evaluating impact



- Compare:
 - A soldier reading from a file reserved for generals
 - A soldier reading from a file reserved for colonels
 - Which is higher impact?

Evaluating impact

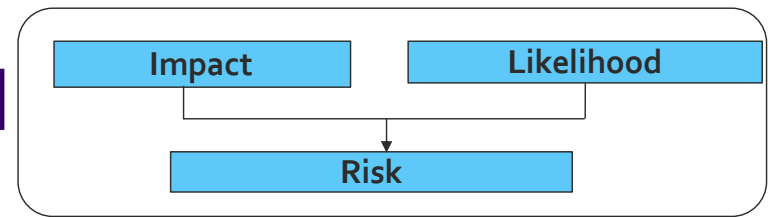


- Compare the impact of:
 - A soldier reading from a file reserved for generals
 - A soldier reading from a file reserved for colonels
 - The impact of the second event is lower than the impact of the first
 - So impact is proportional to the level of the data source
 - ❖ (subject or object)

Impact reduction

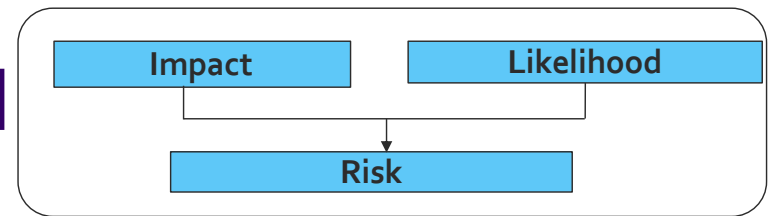
- The « intrinsic impact » determined as a function of security levels can be reduced by applying impact reduction measures
 - **E.g. if the data was obfuscated, the impact is reduced**
- So « impact » in the evaluation of risk can be calculated from the intrinsic impact and the expected effect of impact reduction measures

Evaluating the likelihood



- Compare :
 - A bank director trying to break into the bank's safe
 - A member of the public trying to do the same thing
 - Which event is more likely?

Evaluating the likelihood

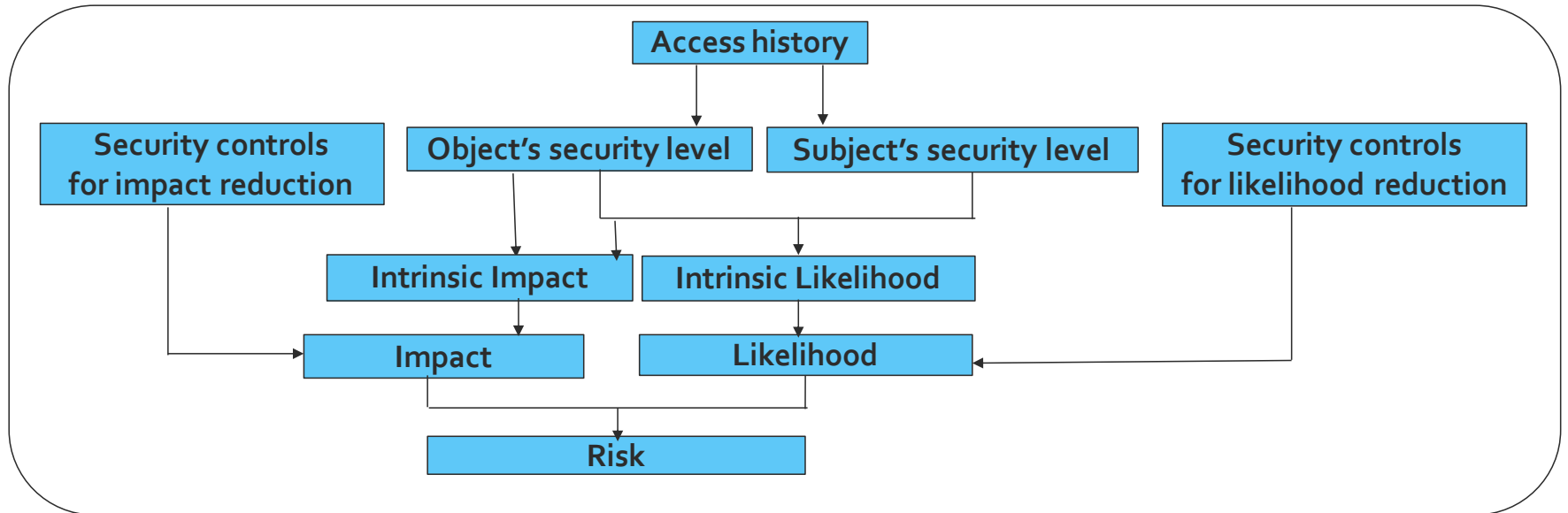


- Compare the likelihood of:
 - A bank director trying to break into the bank's safe
 - A member of the public trying to do the same thing
 - The likelihood of the second event seems to be higher
 - ❖ And this is supported by criminology theory!
 - (poor man wanting to rob rich)
 - In organizations, data theft is most often performed by low-rank employees
 - And so likelihood is also proportional to the difference in level between the origin of the data flow and its sink

Likelihood reduction

- The « intrinsic likelihood » determined as a function of security levels can be reduced by applying likelihood reduction measures
 - **E.g. the safe can be put in a secret location, thus reducing likelihood**
- So « likelihood » in the evaluation of risk can be calculated from the intrinsic likelihood and the expected effect of likelihood reduction measures

Overall view of our risk evaluation method



Some details

Principles for calculating subject confidentiality levels

- P1: the confidentiality level of a subject who has not rec'd any data from levels higher or equal to its own, is defined by default (probably, by the administrator)
- P2: a subject's confidentiality level increases as it receives data from levels higher or equal to its own
- P3: a subject's confidentiality level also increases as the number of flows from such levels increases

Examples (1)

Subject	Confidentiality level at instant t
Nadia	2
Claude	2
Bruno	1
Carl	1
Sabrina	1

<i>Object</i>	<i>Confidentiality level at instant t</i>
o_1	4
o_2	4
o_3	3
o_4	2
o_5	1
o_6	1
o_7	1
o_8	1

- Bruno, Carl initially at same level
- Bruno=1 reads from $o_1=4$ and Carl=1 reads from $o_4=2$
 - What happens to their confidentiality levels?

Examples (1)

Subject	Confidentiality level at instant t
Nadia	2
Claude	2
Bruno	1
Carl	1
Sabrina	1

<i>Object</i>	<i>Confidentiality level at instant t</i>
o_1	4
o_2	4
o_3	3
o_4	2
o_5	1
o_6	1
o_7	1
o_8	1

- Bruno, Carl initially at same level
- Bruno=1 reads from $o_1=4$ and Carl=1 reads from $o_4=2$
 - Bruno's level increases above Carl's

Examples (2)

Subject	Confidentiality level at instant t
Nadia	2
Claude	2
Bruno	1
Carl	1
Sabrina	1

<i>Object</i>	<i>Confidentiality level at instant t</i>
o_1	4
o_2	4
o_3	3
o_4	2
o_5	1
o_6	1
o_7	1
o_8	1

- Bruno, Sabrina initially at same level
- Bruno=1 reads from $o_1=4$ and $o_2=4$; Sabrina=1 reads from $o_2=4$
 - What happens to their confidentiality levels?

Examples (2)

Subject	Confidentiality level at instant t
Nadia	2
Claude	2
Bruno	1
Carl	1
Sabrina	1

<i>Object</i>	<i>Confidentiality level at instant t</i>
o_1	4
o_2	4
o_3	3
o_4	2
o_5	1
o_6	1
o_7	1
o_8	1

- Bruno, Sabrina initially at same level
- Bruno=1 reads from $o_1=4$ and $o_2=4$; Sabrina=1 reads from $o_2=4$
 - Bruno's level increases above Sabrina's

Examples (3)

Subject	Confidentiality level at instant t
Nadia	2
Claude	2
Bruno	1
Carl	1
Sabrina	1

<i>Object</i>	<i>Confidentiality level at instant t</i>
o_1	4
o_2	4
o_3	3
o_4	2
o_5	1
o_6	1
o_7	1
o_8	1

- Nadia one step higher than Carl
- Nadia=2 reads from $o_1=4$ and $o_3=3$; Carl=1 reads from $o_2=4$ and $o_4=2$
 - What happens to their levels?

Examples (3)

Subject	Confidentiality level at instant t
Nadia	2
Claude	2
Bruno	1
Carl	1
Sabrina	1

<i>Object</i>	<i>Confidentiality level at instant t</i>
o_1	4
o_2	4
o_3	3
o_4	2
o_5	1
o_6	1
o_7	1
o_8	1

- Nadia one step higher than Carl
- Nadia=2 reads from $o_1=4$ and $o_3=3$; Carl=1 reads from $o_2=4$ and $o_4=2$
 - Nadia's and Carl's levels both raise
 - But Nadia's remains above Carl's

Examples

Subject	Confidentiality level at instant t
Nadia	2
Claude	2
Bruno	1
Carl	1
Sabrina	1

Object	Confidentiality level at instant t
o_1	4
o_2	4
o_3	3
o_4	2
o_5	1
o_6	1
o_7	1
o_8	1

- Bruno, Carl, Sabrina initially at same level, Nadia one step higher
- Bruno=1 reads from $o_1=4$ and Carl=1 reads from $o_4=2$
 - Bruno's level increases above Carl's
- Bruno=1 reads from $o_1=4$ and $o_2=4$; Sabrina=1 reads from $o_2=4$
 - Bruno's level increases above Sabrina's
- Nadia=2 reads from $o_1=4$ and $o_3=3$; Carl=1 reads from $o_2=4$ and $o_4=2$
 - Nadia's and Carl's levels both raise but Nadia's level remains above Carl's

Formalizing these concepts

- These intuitive concepts can be used and formalized in many different ways

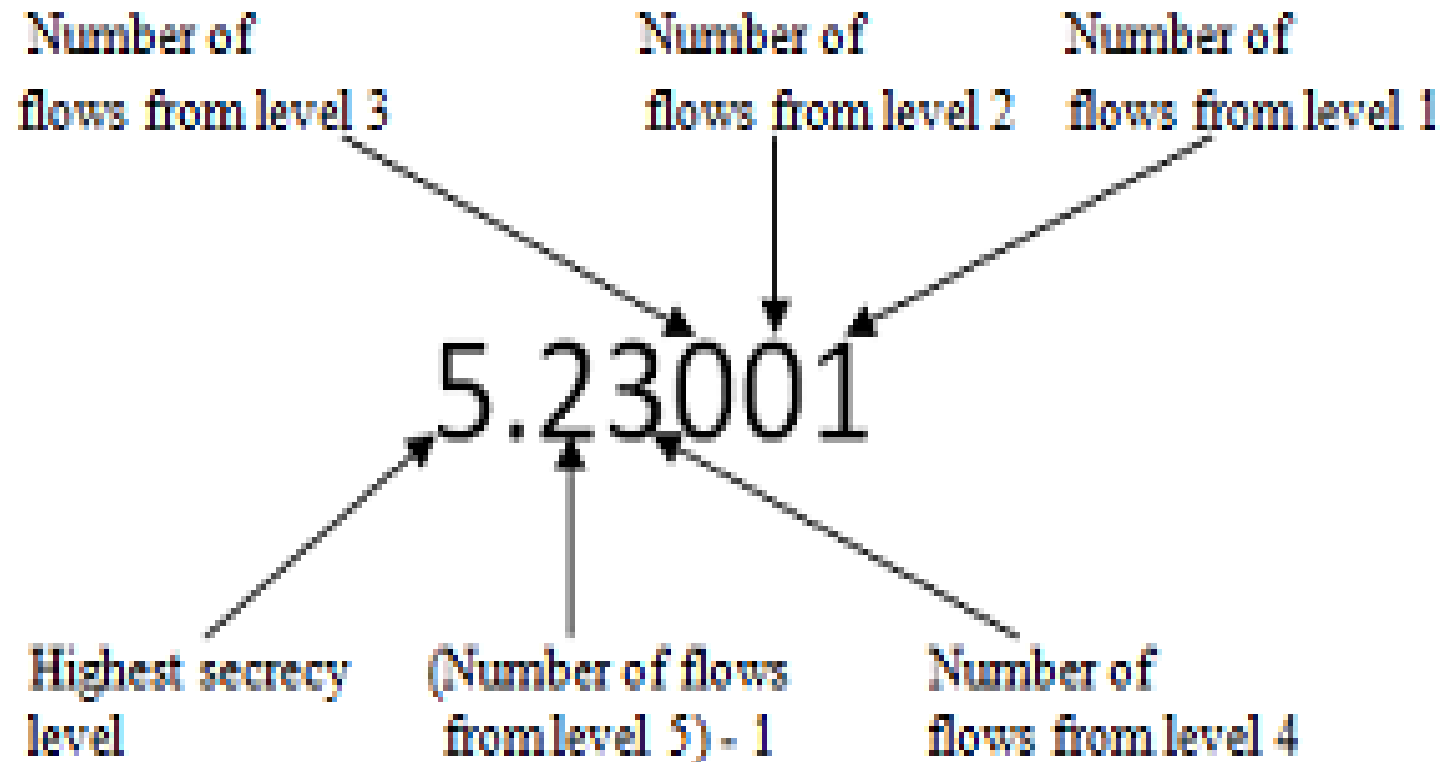
Working with relative comparisons

- Relative comparisons are already useful, since we can decide that higher level subjects get the preference in certain situations

Obtaining absolute values

- But it can also be useful to work with absolute values and full orderings
 - **Expressed by numbers**

Example



Security level of a subject that has rec'd

- 3 flows from level 5,
- 3 from level 4
- 1 from level 1

Principles for calculating object confidentiality levels (essentially the same ...)

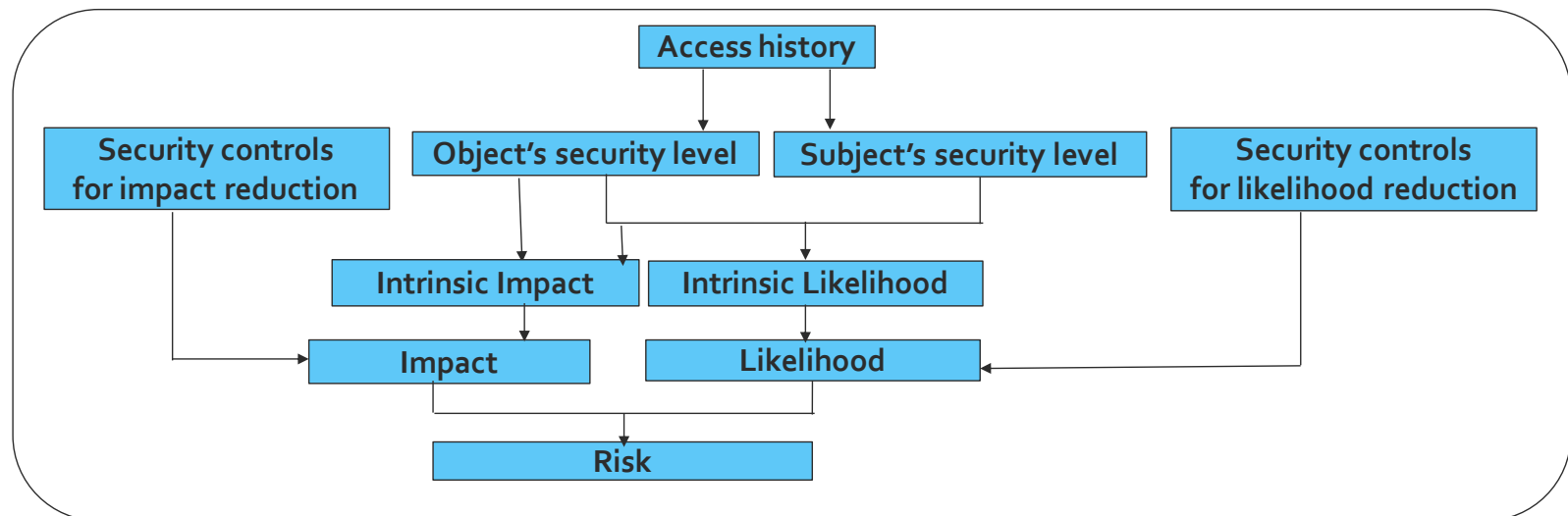
- P1: the confidentiality level of an object who has not rec'd any data from levels higher or equal to its own, is defined by default (probably, by the administrator)
- P2: an object's confidentiality level increases as it receives data from levels higher or equal to its own
- P3: an object's confidentiality level also increases as the number of flows from such levels increases

Dually, the same ideas can be applied for integrity levels

- As a database receives data from lower integrity databases, its level of integrity *decreases*
- The reasoning is perfectly dual wrt previous reasoning

Details to be filled in

- Calculating object security level
- Calculating subject security level
- Calculating intrinsic impact
- Calculating intrinsic likelihood
- Calculating effects of reduction measures on impact and likelihood
- **These calculations can be done *in many ways*, we'll propose one**



Calculating intrinsic threat likelihood for *confidentiality* from subject and object levels

- Principle 1: Threat likelihood is non-null iff:
 - A subject tries to read an object at a higher level
 - A subject tries to write on an object of lower level
- Principle 2: Threat likelihood increases
 - For reading:
 - as the level of the object read increases
 - or the level of the subject reading decreases
 - For writing:
 - as the level of the object written decreases
 - or the level of the subject writing increases

Example for reading

- Subject $s_3=8$ requests read to object $o_1=10$
 - Subject $s_4=7$ requests read to object $o_2=9$
 - Subject $s_5=6$ requests read to object $o_2=9$
-
- **Order these requests by their risk values!**

Example for reading

- Subject $s_3=8$ requests read to object $o_1=10$
- Subject $s_4=7$ requests read to object $o_2=9$
- Subject $s_5=6$ requests read to object $o_2=9$

- A 'reasonable' likelihood ranking according to these principles could be:

$$(s_4, r, o_2) < (s_5, r, o_2) < (s_3, r, o_1)$$

- Similar comparison criteria can be established for writing

Insider threat

- These principles can be used to evaluate insider threat
- This the threat that can arise when employees at different levels in an organization can be required to fulfill equivalent operations: which one to prefer?
 - **Typical problem:**
 - **Given two equivalent workflows, one requiring (s_4, r, o_2) and the other requiring (s_3, r, o_1) , which one to prefer to minimize threat?**
 - **In general, given different workflows requiring different combinations of operations for the same result, which one to prefer?**

Quantifying threat likelihood

- In order to be able to choose the minimum threat alternatives in the general case, it is necessary to provide numerical values for different threat likelihoods in different situations
- Different formulas can be devised that respect the principles that we have expressed, or other principles as required

Example of formula for threat calculation

$$\text{Threat}(s, e, o, c, t) = \begin{cases} \frac{(w \times (\overline{\text{col}(o, t)}) + \overline{\text{csl}(s, t)})}{(|L_c| + 1)^2 - 1} & \text{if } \text{csl}(s, t) > \text{col}(o, t), \text{ where } w = |L_c| + 1, \\ 0 & \text{otherwise} \end{cases}$$

$$\overline{\text{csl}(s, t)} = (|L_c| + 1) - \text{csl}(s, t)$$

$$\overline{\text{col}(o, t)} = \text{col}(o, t)$$

This formula takes into consideration the levels of the subjects and objects, with the number of possible levels

It is conceived to respect the 'principles'

Calculating the impact

- Measures of impact attempt to quantify the importance of a possible violation on the organization
- In the case of
 - **Read access upwards if confidentiality is a goal**
 - **Write access upwards if integrity is a goal**
 - The information stored in the object is impacted
 - The impact is proportional to level of the object
- In the case of:
 - **Write access downwards if confidentiality is a goal**
 - **Read access downwards if integrity is a goal**
 - The information known by the subject is impacted (because of divulgation)
 - The impact is proportional to the level of the subject

Example (confidentiality):

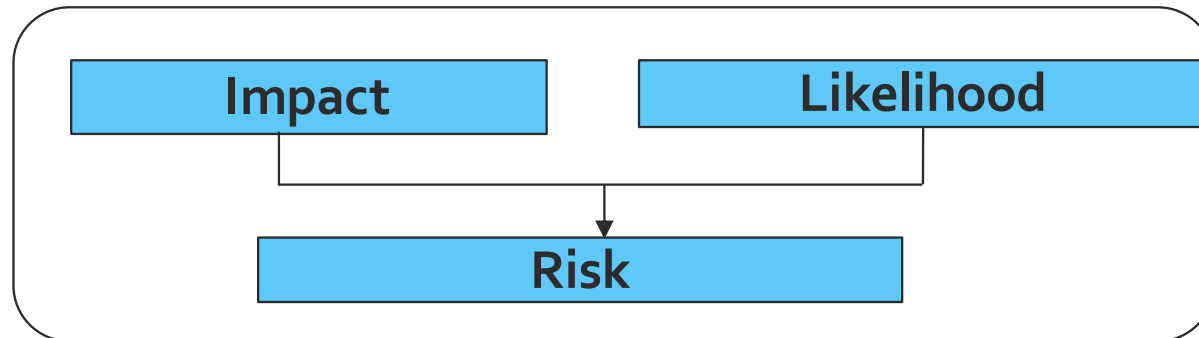
- A TopSecret subject writes on a Public object
- A TopSecret subject writes on a Confidential object
 - **Which operation has higher impact?**

Example (confidentiality):

- A TopSecret subject writes on a Public object
- A TopSecret subject writes on a Confidential object
 - **Clearly in the first case the impact is greater because of the fact that TopSecret information becomes public**

Final calculation of risk

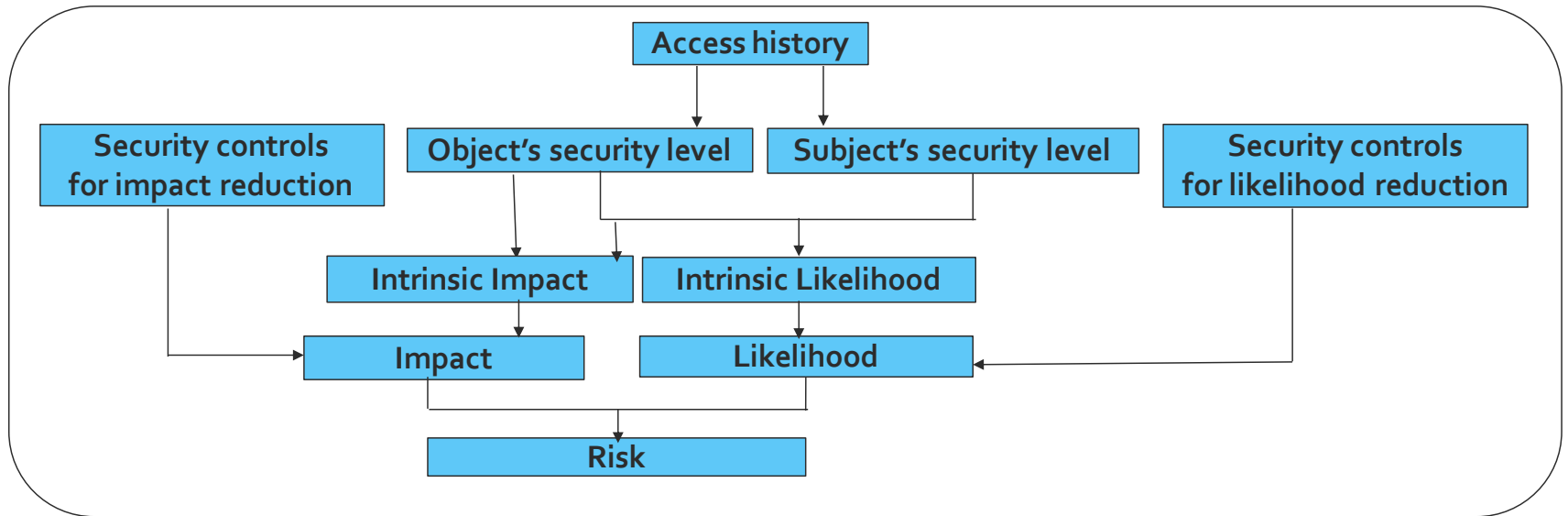
- At the end, risk is calculated *as a function of impact and likelihood*
- Multiplication of the two values is often mentioned, but other functions can be used according to need



Impact reduction and likelihood reduction

- Different methods can be used to evaluate reduction measures that can be used in various organization
- Different formulas can be used, according to the situation
- The Méhari guidelines include many ideas of how this could be done

All done!



More details!

- Making all this precise requires many technical details, involving the definition of many of functions
- Many choices are possible and so the idea can be adapted in many ways
- Different organizational goals can be satisfied by different adaptations
- See our papers, several forthcoming, and also forthcoming PhD thesis of Sofiene Boulares

Feasibility

- Although the theory can become complicated, in practice the calculations required can be light and can be made practical in web- or cloud-based systems

Conclusion

- Method for dynamically evaluating the confidentiality or integrity levels of entities
- In highly dynamic data flow systems
- Access control decision can be taken with consideration of
 - **what data a subject already knows**
 - **what data an object already contains**

References

For now, only one paper has been published but it contains references to most of the relevant literature:

Sofiene Boulares, Kamel Adi, Luigi Logrippo. Information flow-based security levels assessment for access control systems. LNBIP 209 105-121. Proc. of the 6th International MCETECH Conference on eTechnologies. Montreal, May 2015.

We are submitting other papers, so keep looking at our web site

http://www.site.uottawa.ca/~luigi/papers/99_luigi_papers_index.htm