Proposal for industrial tools for data security

Luigi Logrippo and Abdelouadoud Stambouli
Université du Québec en Outaouais, LRSI
luigi@uqo.ca

1. The problem : data flow control for high-security systems

We address the needs of high-security systems (called here *networks)* where it is important to control the circulation of data, for secrecy, integrity, and related purposes, such as confidentiality, privacy, etc. Access control methods can control only direct access, e.g. with RBAC (Role-Based Access Control) it is possible to specify that subject A can read from database B and write on database C, also that subject D can read from C. If nothing else is specified, then a request of D to read from B will be refused; however D can read the data of B if A writes them on C. Indirect data transfers can span many subjects and objects (collectively called *entities* below). In large networks, how can such indirect possibilities be detected, how can they be prevented?

2. The discovery: efficient algorithms

Given that data networks can consist of tens of thousands of entities such as users, subjects, databases, etc., to address the question of detecting indirect accesses, efficient algorithms are needed. Such algorithms were identified by the authors, and their application was discussed in several papers, see citations below. The figure below shows a diagram that was drawn by using our method, starting from a very small list of reading and writing permissions as could be specified in Linux, Windows, RBAC, or other systems. Subjects (or users, processes, etc.) are S1 to S8, objects (databases, files, etc.) are O1 to O9. By reading this diagram, it is possible to answer questions such as:

a) Given a data item in database O3 (e.g. a credit card number), where can it end up? We call this the *secrecy* question. The diagram identifies the Area of O3, which contains all subjects or objects where the data of O3 can end up.

b) Can a subject S8 know data coming from database O6? We call this the *integrity* question. The diagram indicates that no, since S8 is outside the area of O6.

c) What are the entities that have potential access to the same data? These are the entities that are mutually reachable by data flows. In the diagram, these have been grouped in the same box.

d) What are the most secret entities in this network? If we say that the most secret entities are those that have no outgoing data flows, then in this network they are S5, S7, O4, O9, O7. For secrecy, these are the entities that should be most protected against hidden outgoing flows.

e) What are the entities that have the most integrity in this network? If we say that these are the entities that have no incoming data flows, then in this network they are S4, O1, O10 (these could be data collecting entities, such as cash registers, temperature readers on the field, etc.) For integrity, these are the entities that should be most protected against hidden incoming flows.

By using the same algorithms in different ways, it is also possible to solve problems such as:

f) How do we configure a network where the data of A cannot possibly end up in B?

g) We have a conflict of interests between subjects A and B, the data of one should never reach the other, how do we configure the network to make this impossible?

h) We should find a safe place for very secret data, where do we put them?

i) What data will be exposed by possible security breaches in the various entities?

We have also shown how a Software Defined Network (SDN) controller can be configured to enforce the data flow control policies that guarantee secrecy and integrity in the network.

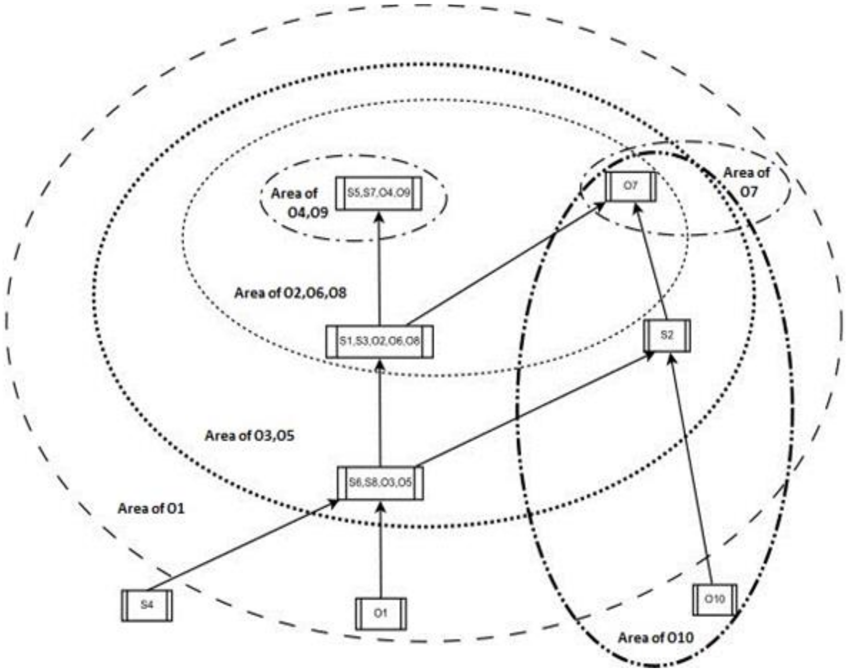
3. The industrial opportunity

The problem of data flow control for security presents itself in many practical contexts, in organizations that have complex data flows and sensitive data, such as banks, government, hospitals or the military, but also in the Internet of Things and 5G networks. We have read the literature on this subject, and we are satisfied that there are no industrial (or even research) tools capable of solving these problems efficiently, as we can.

We have so far developed the theory and done simulations to prove efficiency. The implementation of commercial tools will require an industrial effort.

References:

1) A presentation with more details:
https://www.site.uottawa.ca/~luigi/presentations/public_presentations/20_SereneRiscSECREV.pdf
2) Journal paper including the figure below and presenting the efficient algorithms we use, with simulation runs:
https://www.site.uottawa.ca/~luigi/papers/19_IPL.pdf
3) Journal paper with the basic theory:
https://www.site.uottawa.ca/~luigi/papers/20_Multilevel.pdf
4) A conference paper with a ‘hospital’ example
https://www.site.uottawa.ca/~luigi/papers/18_FPS.pdf



Our method can produce diagrams like this, which can be calculated from lists of read and write permissions. The arrows represent possible data flows. The ‘area of O1’ contains all entities where the data of O1 can end up, and similarly for the areas of O3, O2 etc. See Reference 2) for details. For real-life networks, such diagrams will be too big to draw but the same information can be presented in tabular or relational form.

December 2020