# Granularity Based Flow Control

Omar Abahmane, Luigi Logrippo

*Abstract*— **Many models, methods, techniques, and systems have been developed to preserve the integrity of data and guarantee an acceptable level of security over networks. Protection from illegitimate data access and control of information flow are two main goals. This paper presents new techniques that address two main issues: information protection at various levels of granularity and data flow control**

**We first investigate challenges and limits of established access control models regarding flow control. We then introduce a new flow control model based on granularity, the GBFC. GBFC is capable of guaranteeing flow control under reasonable assumptions. In addition, it offers advantages such as adaptability, full control, reliability and compatibility amongst others. Essentially, in GBFC classified information at suitable levels of granularity is accessible through references and information flow control is applied on the references. We also introduce the concepts of views for information access and Noise Injection that represent building blocks for the Granularity Based Flow Control. With noise injection, a document can be transformed into different views to erase or replace protected information and this transformation can be made almost undetectable to the unauthorized reader. Therefore, inference can be made much more difficult with this method. The GBFC model is intended to complement, rather than replace, existing access control methods.**

*Index Terms*— **Information flow; flow control; granularity; security models.**

## I. Introduction

With the wide use of networks and internet technologies, including social networks and cloud computing and the dramatic expansion of distributed data access via mobile devices, information is becoming increasingly available and at the same time more at risk of illegitimate access or leakage. In this complex environment information and data security are becoming increasingly important concerns. Amongst major concerns in this area, is the issue of information flow control that aims at avoiding leaks of confidential information to objectionable subjects. We propose a new model for access and flow control based on three main concepts: Granularity, referencing and noise injection. In our model, information at various levels of granularity is accessible through references and flow control is applied to these references. Noise injections can replace protected information with other information that can be pure noise or meaningful manufactured information.

In section II we introduce the notion of flow control and examine its implementation in some of the better known security models. We then, in section III and IV, expose some of the challenges and limits of these models related to flow control. In section V we present our model in detail and in Section VI we enumerate its main advantages, most important we show how the model can be used to add highly parameterized flow control to existing access control models. To illustrate the results of the GBFC we present an example of its use in section VII. Concluding remarks and prospective research work conclude the paper in section VIII.

Manuscript submitted April 7,2014, revised May 28, 2014;

This research was funded in part by a grant of the Natural Sciences and Engineering Research Council of Canada (NSERC).

Omar Abahmane is a PhD student at Université du Québec en Outaouais, Canada. (e-mail: abao01@uqo.ca).

Luigi Logrippo is Professor at Université du Québec en Outaouais and Professor Emeritus at the University of Ottawa, Canada. (e-mail: luigi@uqo.ca).

## II. Information flow

With the shift from a centralized architecture for a single organization pursuing a unique goal with internal users and a homogenous security policy, to a wide and open architecture incorporating multiple organizations, pursuing competitive goals and obeying different -sometimes divergent- security rules with both internal and external users, flow control is becoming a crucial problem that requires dedicated attention. It is common nowadays to find private and even classified confidential information right at our fingertips on the widest open network ever: the Internet.

Despite the use of well-known data security models and techniques, many issues of privacy and information leakage are emerging as online businesses start offering private data or classified information as products available to the large public. The sources of such information range from simple online services selling private and personal information about individuals (such as contact, professional, financial or even legal information) to large institutions offering classified and highly sensitive information on demand (Wikileaks [1] is a recent example).

### A. Definition of Flow Control

Considering two subjects (system users or processes) $S_1$ and $S_2$, we say that there is information flow from $S_1$ to $S_2$ when S1 propagates data willingly or unwillingly to $S_2$. In other words, $S_1$ writes some data to an object (memory, file, etc...) on which $S_2$ has read access [2,3]. Based on this definition, an illegitimate information flow occurs when $S_2$ writes classified information [4,5] to objects accessible by some subject $S_3$ that shouldn't have access to that information, resulting in leakage violating some information security policy, see Fig. 1.

978-1-4799-3503-1/14/$31.00 ©2014 IEEE

Flow control aims to prevent such illegitimate information flow and provide end to end security [12].

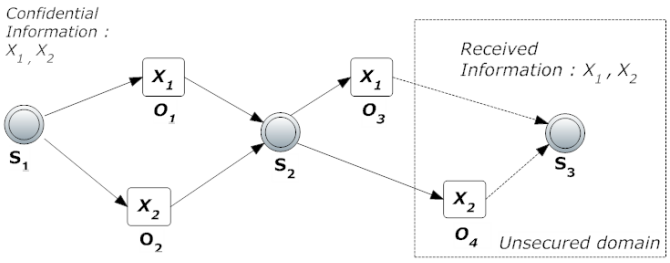



Figure 1. Information flow between 3 subjects.

To define, regulate and secure the information flow within a system and between the system and its environment, flow policies are implemented and enforced by information flow analysis [3,6].

We are aware of the fact that other authors adopt different definitions of information flow, for example based on the concepts of non-interference and independence with application in multi-level security models. However, according to Lowe and Mantel there is no general agreement on a unique formal definition of information flow [7,8].

Confidential or classified information may leak beyond authenticated system users and by consequence beyond the secured boundaries of a system if the appropriate policies are not maintained. For this reason, flow control is of crucial importance. Flow control also prevents flow of undesirable and false information to strategic and decisional levels of an organization, protecting data integrity [6]. Flow control is implemented through confidentiality policies and different mechanisms that enforce end-to-end security [12]. The first widely recognized information flow model was proposed by Denning, in 1976 [9]. Information is said to flow from class *A* to class *B* ($A \rightarrow B$) whenever information associated with *A* affects the value of information associated with *B*. For a model to be considered secure any execution of a sequence of operations cannot produce a flow that violates the relation "$\rightarrow$" [9,10].

### B. Types of flow

An information flow can be direct when it happens between the owner of the information and other parties. This type of flow is generally legitimate and requires the information owner's explicit permission.

Indirect flow occurs when information flows among subjects other than the owner [7,11]. The indirect information flow between subjects might be restricted by the owner by access right propagation or revocation in the form of reads, writes, grants … etc.

Leakage of confidential information can occur with indirect information flow as presented in Fig. 1, where confidential information flows from authorized users to unauthorized ones either deliberately or as a result of errors or leakage.

### C. Access control models and Flow control

Many access control models and policy systems have been developed since the 70's. The main models are Access Matrix, Mandatory Access Control (MAC), Discretionary Access Control (DAC) and Role-Based Access Control (RBAC). They deal with the information security requirements of confidentiality, availability, integrity and non repudiation [13].

#### 1) Mandatory access control (MAC)

Mandatory access control deals mainly with confidentiality and integrity of information. MAC deals with these two requirements from a centralized flow control viewpoint.

Two important models of this family were built upon information flow:

Confidentiality is covered in the Bell-LaPadula Model. This model is concerned with the flow of information that occurs from high to low security levels [14]. One drawback of this method of information flow control is that it might create situations of data integrity violation.

To deal with the integrity violation problem, the Biba model was proposed. It addresses this problem by tracking the correctness of all writes from low to high integrity levels [15]. This model also presents some limits related to possible security violation because of inference of high level information from low level information [16,17].

Both Bell-LaPadula and Biba models are very limiting and impractical in many situations. Extensions have been proposed, but they usually lead to security flaws.

#### 2) Discretionary Access Control (DAC)

While MAC Models deal with information flow and enforce flow control policies, Discretionary Access Control is an identity-based model that lets users manage access rights and grant them to other users. DAC does not deal with information flow [16].

#### 3) Role-Based Access Control

Role-Based Access Control (RBAC) is a security model that associates access rights to roles. Sets of roles and groups of roles are created, and then users are affected to these roles or groups of roles to determine their associated access rights [18]. Flow control in the RBAC model is enforced by controlling user's roles inside the system. Depending on the role's access rights information flow is allowed or prevented [30]. RBAC can also implement MAC models and so it can address flow control in this way [31].

TABLE I. COMPARISON OF MAC, DAC AND RBAC ACCESS CONTROL AND FLOW CONTROL IMPLEMENTATION [17]

| | **MAC** | **DAC** | **RBAC** |
|---|---|---|---|
| Control level | Central (server) | User | Central and user |
| Access right review | Central | User | Central and user |
| Access right propagation | Central | User | Central and user |
| Information flow control | YES | NO | ?? |

## III. INFORMATION FLOW CONTROL CHALLENGES

Protecting confidential information during and after a flow presents many difficulties amongst which:

- Information is produced in quantity and security classification associated with it may change frequently and dynamically which complicates security policy management.
- Tracking the information in a flow is a very difficult task especially when the flow involves different networks and by

consequence different security domains. The complexity of this situation increases rapidly with the number and disparity of subjects and domains involved in the flow.

• Information owners are generally the most suitable to decide its classification. Such classification might conflict with the proposed classifications by the security models.

• Real world flow security needs may be very different depending on the data transferred (files, emails, private data, copyrighted content, etc.), the type of users (single user, organizational unit, social networks … etc.) and the scope of the flow (localized domains vs. multi-domains).

• Most of the flow control within a security domain is based on system-wide rules and clear policy that users understand. The knowledge of the policies, by itself, may facilitate policy violation.

• Flow control at the boundaries of a security domain is usually achieved through security mechanisms such as firewalls or antivirus software. These tools present vulnerabilities to new malware technology development and require continual updates. They don't offer end-to-end security and are ineffective when dealing with validated users, genuine software or endorsed code.

• Access control policies do not control how data is used after a subject accesses it. So if an authorized user reads classified information other mechanisms are needed to track and control subsequent user's activities such as replication, propagation and secure disposal in order to prevent transitive disclosure. 'Usage control' [32] tries to addresses this situation through implementing subject obligations but it is not being implemented widely.

## IV. Limits of the access control models regarding flow control

Because of the difficulties listed above, little has been done to implement flow control oriented architectures and systems [12]. Existing systems address information flow security through the existing access control mechanisms combined with security add-ons to integrate flow control policies. These combinations are necessary as access control mechanisms are inadequate when it comes to constraining flow of information even though they succeed with access control [17,19,20].

With the absence of a new integrated flow control model the saying of D.E. Denning still holds [9]: "Systems needs both access and flow control to satisfy all security requirements".

Flow control models describe policies to control propagation of classified information between classes. However, some access control models are not designed to control flows inside the classes although this may be desirable. The multi level security models prevent flow of information between security levels even though scenarios where such flow is desirable may exist.

In addition, some fully permitted flow of information between classes might still be considered undesirable based on specific circumstances.

Many other reasons cause access control policies to fail in realizing flow control:

• Illegitimate indirect information flow that may result following a legitimate access by an authenticated subject. This flow may take different forms making it difficult to control.

• Existing models concentrate on securing subjects and objects to protect information rather than securing information itself.

• The majority of research on information leakage and flow control focuses on the problematic forms and manifestations related to information flow such as confinement, inference, covert channels, etc., that only represent symptoms and effects of the main problem which is the flow itself.

All these limits make it clear that there is a pressing need for new dedicated and efficient flow control models.

Our model tries to address these challenges by proposing a solution that uses granularity and referencing as key factors to preventing information leakage and sustaining adequate information flow control.

## V. Presentation of the Granularity Based information Flow Control model

### A. Granularity

According to [21], granularity, also known as granular computing, was first introduced in 1997. Zadeh [22] states that the fundamental components of granular computing are granules such as subsets, classes, objects, clusters, and elements of a domain or universe. These granules are sets of elements drawn together by distinguishability, similarity and functionality [21].

Granules in their atomic form are designated constituent parts of a specific data model. For instance, a granule in an image file would be a pixel or a set of pixels, and in a text document a granule can be a sentence, a word, a date and so on.

For example, granularity has been implemented in databases to manage and control access to single columns in tables. Access is granted based on the authentication and the roles of the user. However, despite interest in the subject, limited research work has been conducted regarding the use of granularity in flow control.

### B. Description of the model

In this paper we develop the concept of Granularity Based Flow Control (GBFC) through the proposal of a model that implements it. It implements flow control through a process that takes in consideration the limits of the current access control models. It is a flow control oriented add-on to existing access control models and it manages the subjects' authorizations to access classified information.

The GBFC model enforces flow control through a process involving an Access Control Engine (ACE) that is the core component of the system. The documents are accessed in their granular form assuring that, for each granule, a classification label is attached at creation or modification time [20].

Once a subject *S* requests access to a document or resource *X*, the identification and authentication are handled by the implemented access control system (MAC, DAC, RBAC, … etc.) then access rights are passed to the ACE that reads *X*.

Within *X*, granules of information are read and their security labels are compared to the access rights of the subject.

The ACE proceeds in two different ways based on the level of classification of the granules:

• All non-classified granules are loaded to the subject system in the same way as current operating systems perform this action through File Allocation (FA). Non-classified granules are subject to an implicit flow [29].

• For the classified granules the ACE acts differently. It generates references (pointers) to the granules and builds a Volatile File Allocation (VFA) index that is loaded on the subject system and references are replaced by the corresponding data content of each granule. Any manipulation, update or storage of the document is performed through references rather than the data itself. This ensures that no classified data granules are ever saved locally on the subject system, i.e. flow restriction is enforced.

If a document is saved locally and then reopened by the same subject, the references are loaded locally then transferred to the ACE to build the links to the data and allow reloading the classified content of the document.

Any replication, partial copy or transfer of the document is achieved in such a way to preserve classified data by only copying or transferring references corresponding to the actual information granules. This ensures that any non-authenticated subject that attempts to access the document would only have access to references to elements of information that cannot be loaded. When higher security level is required these references are replaced by noise (noise injection). Figure 2 presents the process.

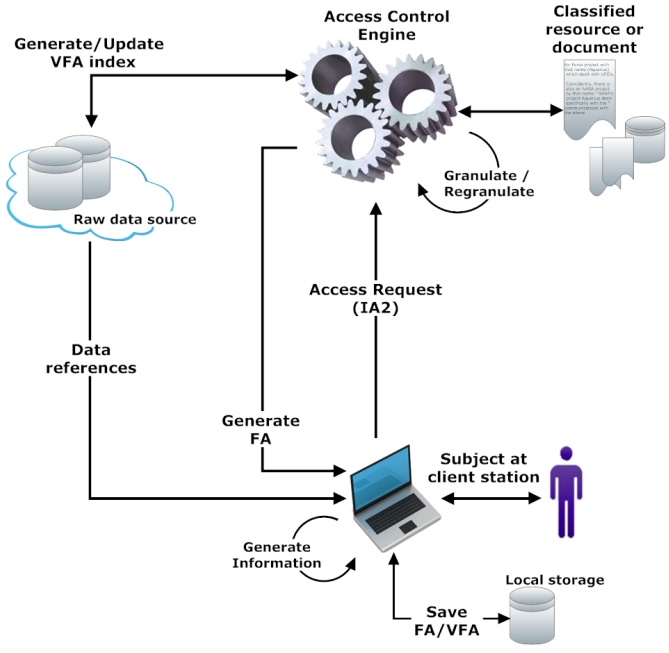



Figure 2. GBFC process chart.

In situations of information leakage, the level of risk is assessed by the ACE based on the location of the subject attempting to access (IP address…) and the risk inherent to the leaked document. The four possible scenarios of subject access after an information flow are listed in table II.

TABLE II. POSSIBLE CASES OF ACCESS AFTER INFORMATION FLOW

| Subject | **Has Access rights** | **Has No access rights** | **Risk Level** |
|---|---|---|---|
| Authenticated | Refs point to data | Gets cleaned document | No risk |
| Non-Authenticated trustful | | Refs point to Null or cleared | Low risk |
| Non-Authenticated malicious | | Refs point to Noise | High risk |

More in detail, GBFC is implemented based on three major concepts, which will now be explained: Granularity, Flow restriction, and Availability. Our discussion of these concepts will assume that the objects protected are text documents. However the concepts can be generalized to different types of data objects, as long as granularity can be applied to them.

1. **Granularity:** In GBFC, security of documents is managed through the granular classification of their components, which in the case of text, could be words, sentences, paragraphs … etc. Implementation of this security aspect is ensured through the *Granularity Level* criteria *Tγ*. *Tγ* is set to different values for each component of the document depending on its level of classification and based on the overall level of security needed. For example *Tγ* could correspond to *word* level for TOP SECRET elements, *sentence* level for SECRET elements, no granulation for UNCLASSIFIED elements, and so on (Fig. 3).

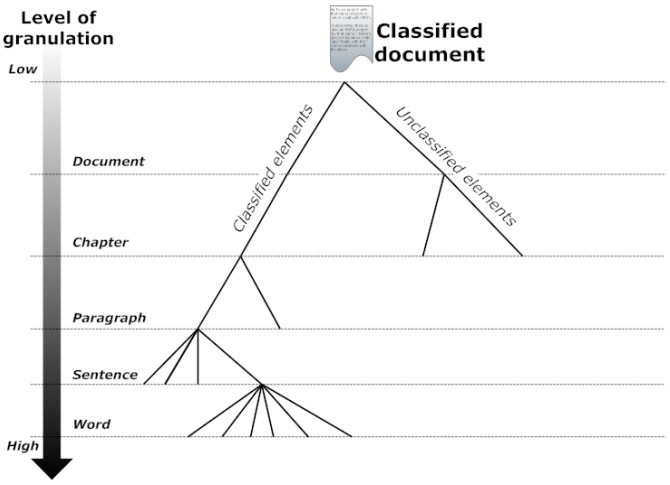



Figure 3: Granulation levels of a document

2. **Flow restriction**: implemented to limit or prevent information flow from authorized to non-authorized subjects, having in mind that the most efficient flow control is obviously "not having a flow at all" or, at least, limiting its existence. The flow restriction is assessed through a *Refresh Rate Tρ* that establishes the criteria and/or the frequency applied to redraw references to classified information granules within the document. This rate may be a refresh rate for environments that require periodic security controls. It may also represent one or more event criteria such as employer dismissal, malicious attack, updates, etc.(Table III)

3. **Availability**: Controlling availability implies controlling access. Unavailable information is inaccessible information. By information availability we mean the logical availability on a physical support accessible by a subject. Availability depends on two factors :
   a. *Availability Rate* $T\alpha$: A rate that sets the level of availability of granules within the document, based on the nature of the data to be replaced by references (nouns, verbs, dates, etc. ...) and based on the classification level threshold to consider in implementing the availability restrictions (such as SECRET or TOP SECRET …).
   b. *Noise level* $T\nu$. The amount of noise inserted into the document to replace the classified unavailable information granules to the subject. $T\nu$ determines the level of noise injection applied to the document.

TABLE III. VALUES OF THE SECURITY CRITERIA DEPENDING ON THE LEVEL OF SECURITY TO ENFORCE

| Level of security | | Lowest | Highest | Examples |
|---|---|---|---|---|
| $T\gamma$ | | Document | Word | *Word, sentence …* |
| $T\alpha$ | Data Type | All Available | None | *Nouns, Verbs, Dates…* |
| | Classification | Unclassified | Top Secret | *(TS), (S), (C), (U) …* |
| $T\rho$ | Event based | None | Maximum | *Update, Infection, system failure …* |
| | Frequency | Never | High | *Monthly, daily, …* |
| $T\nu$ | | No Noise | Max noise | data types in $T\alpha$ *(Nouns, Verbs, …)* |

The usage of the different security criteria is explained in the three examples below:

*Tγ=Word*
*Tα=((Nouns, Verbs), TS)*
*Tρ=(Update, Infection)*
*Tν=(Nouns)*

The classified information in the document will be granulated at WORD level. All TOP SECRET nouns and verbs will be replaced with references. The refresh action is performed on updates and as a reaction to infections. Only references to Nouns will be replaced with noise in case of illegal access.

*Tγ=Sentence*
*Tα=((ALL), TS)*
*Tρ=None*
*Tν=ALL*

The classified information in the document will be granulated at SENTENCE level. All text elements of the sentences classified TOP SECRET will be replaced with references. No refresh of the references is performed. All classified sentences will be replaced with noise in case of illegal access.

*Tγ=Word*
*Tα=((Nouns, Verbs, Dates, Abbreviations, Adjectives,), S)*
*Tρ=(Update, Monthly)*
*Tν=(Nouns, Verbs, Dates)*

The classified information in the document will be granulated at WORD level. All the nouns, verbs, dates, abbreviations and adjectives classified SECRET or higher will be replaced with references. The refresh is performed on updates and periodically (every month). Only references to Nouns, Verbs and Dates will be replaced with noise in case of illegal access. A more elaborated example is offered in section VII.

GBFC algorithm is presented hereafter:

==========

*Title: Granularity Based Flow Control Algorithm*

==========

```
1.  begin
2.  V:=AuthorizeAccess(S, Inf)
3.  if V=False then
4.     accessDenied()
5.  else
6.     initializeInformation(Inf)
7.     load Tγ, Tρ, Tα, Tν
8.     while(not EOF)
9.        for each gr_i ∈ Inf
10.          if (gr_i.attr ∈ classified and gr_i.attr <= S.attr) then
11.             addRef (VFA, gr_i.ref)
12.             updateVFA()
13.          else if (gr_i.attr ∈ classified and gr_i.attr > S.attr) then
14.             addRef (VFA, noise.ref)
15.             updateVFA()
16.          else
17.             addIndex (FA, gr_i.idex)
18.             updateFA()
19.          end if
20.       end for
21.       buildVFA()
22.       buildFA()
23.       refreshRef(Tρ, Tα, Tν)
24.       regranulate(Inf, Tγ)
25.    end while
26. end if
27. end
```

The algorithm proceeds as follows:

First, the subject *S* is authenticated and access rights are verified to grant or deny access. Once authorized, the classified document (*Inf*) is accessed and its security variables loaded (line 7). Within *inf*, the system reads each granule of information $gr_i$ and verifies its level of classification. If *S* has access right to $gr_i$ the system creates a reference to the granule content and adds it to the Virtual File Allocation (an entry is added to the VFA index for each level of classification). Otherwise, if *S* has no access right to $gr_i$ the system creates noise reference and adds it to the VFA based on $T\nu$. In case $gr_i$ is unclassified (public), the system adds $gr_i$ index entry to the basic File Allocation (FA).

The system then, builds the VFA to load granules references or noise references as well as the FA for public data on the subject system. Once *Inf* is loaded on the subject

system, the volatile references on the VFA are refreshed based on $T\rho$, $T\alpha$ and $T\nu$, and an optional re-granulation of *Inf* is performed. The system proceeds likewise until it reaches the end of the document and stops. The whole process is illustrated in the example of Section VII.

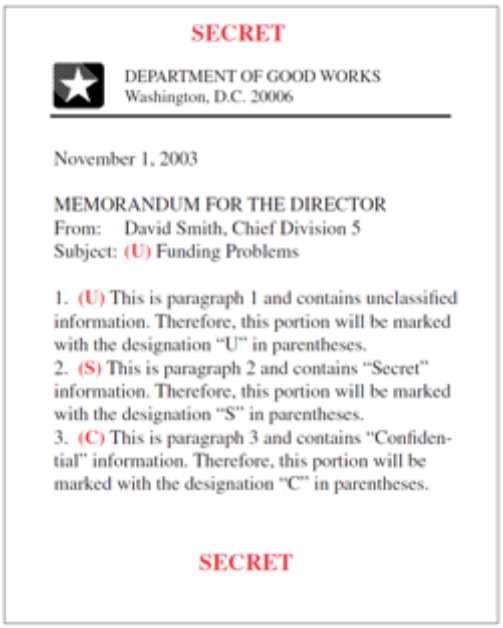
## VI. Advantages of the GBFC Model

### A. Adaptability

As stated earlier, GBFC enforces flow control through managing specific quantitative criteria ($T\gamma$, $T\rho$, $T\alpha$, $T\nu$) which offer great adaptability and manoeuvrability thanks to the possibility to configure each of the criteria independently. Thus, security administrators are provided with a flexible multi-criteria environment to easily apply the desired level of security in order to ensure optimal control of information flow.

### B. Access restriction and replications

Usually, security services classify documents based on the highest level of classification of their information elements. In other words, when creating a document derived from multiple sources or enclosing different levels of classification, the derivative document will be marked with the highest classification level of information found in any of its portions [23,24] (see Fig. 3).


SECRET

DEPARTMENT OF GOOD WORKS
Washington, D.C. 20006

November 1, 2003

MEMORANDUM FOR THE DIRECTOR
From: David Smith, Chief Division 5
Subject: (U) Funding Problems

1. (U) This is paragraph 1 and contains unclassified information. Therefore, this portion will be marked with the designation "U" in parentheses.
2. (S) This is paragraph 2 and contains "Secret" information. Therefore, this portion will be marked with the designation "S" in parentheses.
3. (C) This is paragraph 3 and contains "Confidential" information. Therefore, this portion will be marked with the designation "C" in parentheses.

SECRET

FIGURE 4. EXAMPLE OF A DERIVATIVE CLASSIFIED DOCUMENT

For example, if a 100-page document includes a sentence classified as TOP SECRET (TS), one page as SECRET (S) and the rest as UNCLASSIFIED (U), the entire document is classified as TOP SECRET.

This simple application of the directives of the MAC model has two major drawbacks:

1. Access to a document containing highly classified information will be completely forbidden to subjects of a lower level of classification even if it contains less sensitive information that they should be allowed to access.
2. To solve this access restriction problem, the common solution is to produce derivative documents containing appropriate and adequate information for each security level. It is necessary to produce as many versions of the document as there are levels of classification of the information that it encloses. There will be the full document, accessible at the TOP SECRET level, a modified document for the SECRET level (after removal of TOP SECRET information), a third version for the CONFIDENTIAL level (after removal of TOP SECRET and SECRET information), and so on. This requires the implementation of a manual process to create and verify the necessary versions of the document. The production and the existence of such copies pose by themselves security risks in addition to the declassification processing load.

Our method avoids these problems by allowing the original document to be created with a granular structure that allows different subjects at different levels of security to have access limited to information that matches their access rights. Thus, a subject with a TOP SECRET level will have access to the full version, a subject having a SECRET level will have access to the same information and the same document except the content that is classified TOP SECRET, and so on. Non-accessible information for a given level is replaced by references to empty data, fabricated data or noise.

Note that at a high level of granularity, classified words can be replaced by contextually and/or grammatically similar words in the form of noise that offer relatively comprehensible meaning, therefore, preventing the unauthorised reader from being aware of any alteration of the document. On the other hand, by choosing a lower granularity level, the whole classified sentence may be erased from the document leaving no evidence of the existence of the classified content.

An appropriately modified text editor can help producing such granularly organized documents. For example, different levels of classification can be identified by using different colors or tags. This introduces a new perception of access to information based on views, see Fig. 4.

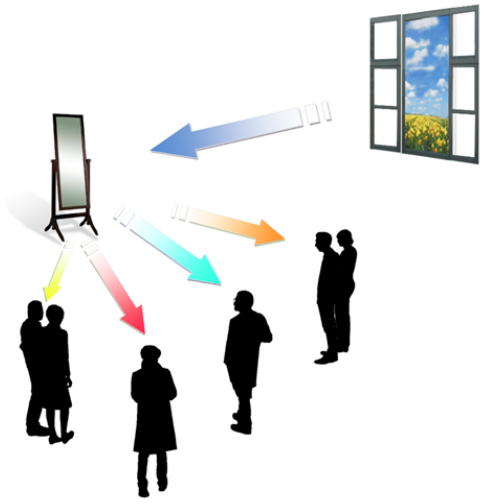


FIGURE 5. ACCESS TO INFORMATION BASED ON VIEWS

Our concept can be compared to a situation where a group of people located in a room look at the same landscape through a mirror (with analogy to the ACE). The landscape is visible through an open window. The image that everyone or every group gets to see depends on their location and the location of the mirror that creates a virtual image of the landscape. This virtual image may change or even disappear depending on the actions taken on the mirror, and the status of the window (open, closed or semi-open).

### C. *Total control*

The concept of information access based on views that we have introduced allows security systems administrators permanent full access control. In contrast to existing systems that allow storing classified information on multiple media (laptops, smart phones, USB keys ... ), some of which are mobile and difficult to track, our centralized access model allows immediate and automatic isolation of classified information during security alerts (external attacks, malicious infections, imminent risk due to voluntary or involuntary leakage of data, etc.).

The resulting situation is similar to the closing of the window in Fig. 5. In such case, all the virtual images of the resource lose their references to classified information by the fact of their isolation vis-à-vis the rest of the network. Once isolated, this information is not available on any other storage medium. After the restoration of the secure state, network access is restored, references to classified information are regenerated and subjects regain access to information in accordance with their rights and permissions. The creation of local copies is possible at the unclassified level. It should normally not be possible at higher security levels, unless the loss of central control can be tolerated.

### D. *Loss of data*

To ensure information security in case of remote access, current security models proceed by identification, authentication, authorization and strengthening security using encryption techniques [25]. Sensitive and critical data are usually accessed on servers and are rarely stored on remote systems or mobile devices (laptops, smart phones, USB keys ...) considering the risk of loss. This concern is entirely justified given the high probability of information leakage due to loss of equipment.

In situations where it is necessary to store classified information locally, organizations ensure that it is stored with the highest security possible to prevent unauthorized access in the event of loss. However, despite these precautions, the loss of material remains the leading cause of information leakage according to studies in the U.S., Europe and Asia [26,27,28].

Our model offers a method designed to remedy this situation as it reinforces the centralized nature of the classified information. In fact, when saving a document or resource on a local storage, the only elements of information that are stored are the unclassified ones. All classified elements are replaced by references to data. These references point to locations containing volatile data whose existence depends on the refresh rate $T\rho$ specified at the ACE. Thus, as long as no incident of loss is reported, volatile data are maintained and authorized users can access it without problems. In case of loss at the local level, the system proceeds with a refresh of the references to sensitive data and blocks all access using the references present on the lost document allowing at the same time traceability of lost information.

On the other hand, authorized subjects will experience no access problems as their systems get the updated references automatically through the GBFC client-server interface that immediately refreshes the references present in the classified document. As a matter of fact, this refresh is operated once the document is open or dynamically if it occurs during the access.

In severe situations, the refresh rate can be set to limited sessions with a selected frequency, to increase confidentiality of the information. In different sessions, the user can have access to different subsets of the information: this is appropriate for environments where possessing all the pieces of confidential information at the same time presents a security risk such as credit card information (Account Number, Network, Expiration Date, Holder Name, Security Code) for example. Note that this is another important problem in security that can be addressed with our method.

### E. *Implementation and compatibility*

GBFC is a security platform independent system that is designed as an add-on to the existing security systems. This provides the model with high flexibility and adaptation to security environments. Indeed, the model is built on the ACE that collaborates with other security systems to enhance information flow control based on access rights and classification. After the subject is authenticated the ACE checks the permissions of the subject and creates the volatile copy of the information based on the predefined granularity level and the granular access rights. Figure 6 presents the overall layered architecture of the GBFC in its relation with different security models.

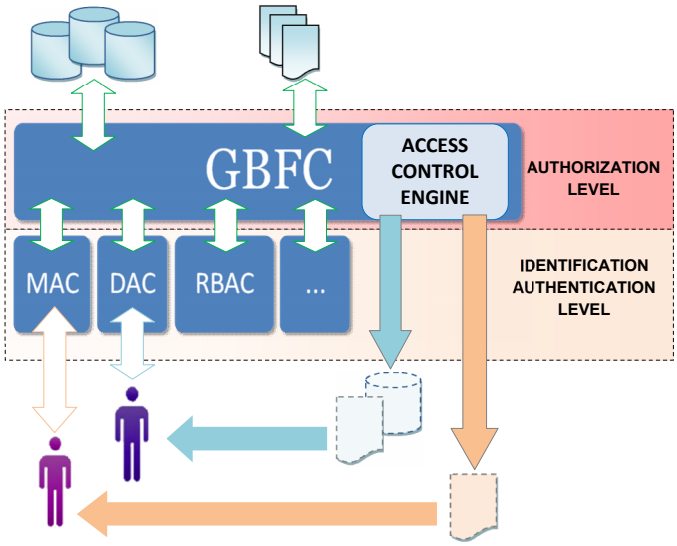



FIGURE 6. LAYERED ARCHITECTURE OF THE GBFC

This architecture is even more valuable in heterogeneous

security environments or in extended networks such as the Internet or Cloud Computing. It allows consolidating methods of decentralized access control administration and building a centralized security environment in order to enhance security and flow control.

This architecture also helps implement flow control for security models that don't enforce it, such as DAC, ABAC-XACML and others.

### F. Noise injection

In our method, we use the idea of mixing noise with data as an information protection mechanism. Noise injection is an obstacle to information reconstruction through inference, in other words, the difficulty of reconstructing a document or a text is magnified by the existence of inaccurate or irrelevant data. For this reason, we have introduced a noise level $T\nu$ which is the list of the classified data types that will be replaced by noise in case of illegitimate access. The greater their number, the more difficult to reconstruct the original information is. $T\nu$ allows the security administrator to set a document to "*noise free*" for trusted subjects that need to access enclosed UNCLASSIFIED data or to "*full noise*" for possibly offensive subjects or environments.

To our knowledge, our model is the first to introduce this new concept of information dissolution in a noisy environment to preserve confidentiality.

This process is quite simple to implement because the ACE supports refreshing references to classified information granules and replacing them with references to noise in the form of rationally selected raw data (applying syntactic rules word by word as in the example in Section VII) or randomly (replacing by random words).

With information dissolution in noise, malicious subjects can be faced with relevant information dissolved in a large pool of false or irrelevant information. This makes it very difficult or impossible to infer missing classified information, thus addressing another important security problem. Further research is needed in this direction (example in Section VII).

## VII. Implementation example

GBFC should be implemented at the Operating System level to integrate the power of the Volatile File Allocation system. However, to demonstrate the idea of this model a prototype at application level is a first step. This prototype involves a client-server interface (ACE/Client Editor) allowing the central security authority to manage and control granularity levels, classifications and the other security criteria ($T\rho$, $T\alpha$, $T\nu$) while offering to the end user access and other permissions. In this section an example presents some of the possibilities of the system.

Reference [33] is an example of declassified TOP SECRET document that we will use for this purpose, after some simplification and manipulation. We consider the following extracted paragraph that we have processed with the classification tool to protect the information it encloses:

> **(TS)** Every individual in a command center responsible for the preparation of emergency action must be familiar with the procedures in the EAP **(/TS)**. **(U)** Command center training and evaluation programs will be developed to ensure that individuals charged with the preparation and transmission of emergency action messages are qualified in this task **(/U)**. **(S)** These individuals and programs are subject to review by the OJCS **(/S)**.

We set the security parameters as follows:
$T\gamma$=*Word*
$T\alpha$= *((Nouns, Verbs, Abbreviations, Dates), S)*
$T\rho$=*(Update, Monthly)*
$T\nu$=*(Nouns, Verbs, Abbreviations)*

Let us assume that we have four subjects with these access rights:

| | |
|---|---|
| - TSungani | Top Secret (TS) |
| - Sue | SECRET (S) |
| - NAolin | Non-authorized / Possibly trustful |
| - NAHacker | Non-authenticated / Possibly malicious |

TSungani will have access to the whole text of the document.

Sue's view is as follows:

> Command center training and evaluation programs will be developed to ensure that individuals charged with the preparation and transmission of emergency action messages are qualified in this task. These individuals and programs are subject to review by the OJCS.

Based on the security levels set earlier, the actual data loaded on Sue's system would be:

> Command center training and evaluation programs will be developed to ensure that individuals charged with the preparation and transmission of emergency action messages are qualified in this task. These 2F08A829 and 2355EA66 2435F450 3D502CE9 to 324AF563 by the 25466F31.

As the granularity level is set to WORD level, references to classified information (SECRET and up as set by $T\alpha$) are created for each category of words listed in $T\alpha$. A SECRET Level authenticated subject (Sue) will see real data referenced by the numbers above. In case of storage, copy or transfer the references are maintained and no classified data is saved or copied locally.

If Sue transfers deliberately or unintentionally this document to NAolin, the text that NAolin receives is:

> Command center training and evaluation programs will be developed to ensure that individuals charged with the preparation and transmission of emergency action messages are qualified in this task. These *NULL* and *NULL NULL NULL to NULL* by the *NULL*.

Since NAolin is a Non-authorized trustful user, references to classified information have been replaced by *NULL*. If we need a more restrictive level of security to prevent unclassified data to be transferred to non authorized subjects, all we need

to do is set the $T\alpha$ to UNCLASSIFIED. In this case, all nouns, verbs, abbreviations and dates in the text will be replaced by *NULL*

Here is the information that NAHacker will have in case he gets access to the document:

> Every aspect in a database solution responsible for the system of agent toolkit integrates call familiar with the languages in the GUI. Command center training and evaluation programs will be developed to ensure that individuals charged with the preparation and transmission of emergency action messages are qualified in this task. These networks and algorithms draw concept to function by the EBML.

Once this subject is identified as a threat to the organization in possession of the classified document, the ACE replaces relevant data elements referenced in the document with noise that will submerge the accessible unclassified data. In this example, noise components (nouns, verbs, abbreviations) were generated from a computer science dictionary. More complex noise generation mechanisms can of course be used, among the many that are known in security practice. Table IV below illustrates the work done by the ACE for managing and loading the references to classified information on subject systems via volatile file allocation.

TABLE IV. REFERENCE MANAGEMENT BY THE ACE (VFA INDEX)

| | Loaded Refs. | Noise Refs. | Classified Data Refs. | Classified Data | Noise |
|---|---|---|---|---|---|
| **Sue** | **2F08A829** | | **2F08A829** | individuals | |
| | **2355EA66** | | **2355EA66** | programs | |
| | **2435F450** | | **2435F450** | are | |
| | **3D502CE9** | | **3D502CE9** | subject | |
| | **324AF563** | | **324AF563** | review | |
| | **25466F31** | | **25466F31** | OJCS | |
| **NAolin** | **534490A2** | **534490A2** | 2F08A829 | individuals | NULL |
| | **534490A2** | **534490A2** | 2355EA66 | programs | NULL |
| | **534490A2** | **534490A2** | 2435F450 | are | NULL |
| | **534490A2** | **534490A2** | 3D502CE9 | subject | NULL |
| | **534490A2** | **534490A2** | 324AF563 | review | NULL |
| | **534490A2** | **534490A2** | 25466F31 | OJCS | NULL |
| **NAHacker** | **6F67890A** | **6F67890A** | 34443501 | individual | Aspect |
| | **7B450021** | **7B450021** | 356099EF | command | Database |
| | **60A89E45** | **60A89E45** | 390040B1 | center | Solution |
| | **67454B89** | **67454B89** | 23546609 | preparation | System |
| | **645109C4** | **645109C4** | 238709B1 | emergency | agent |
| | **6A450910** | **6A450910** | 32118CD0 | action | Toolkit |
| | **62019B34** | **62019B34** | 34667500 | must | Integrates |
| | **679809CC** | **679809CC** | 356387E3 | be | Call |
| | **61026B10** | **61026B10** | 3490A34F | procedures | Languages |
| | **62AE4530** | **62AE4530** | 3337810C | EAP | GUI |
| | **73442000** | **73442000** | 2F08A829 | individuals | Networks |
| | **6938CC23** | **6938CC23** | 2355EA66 | programs | Algorithms |
| | **6B324109** | **6B324109** | 2435F450 | are | Draw |
| | **7318F453** | **7318F453** | 3D502CE9 | subject | Concept |
| | **64009A43** | **64009A43** | 324AF563 | review | Function |
| | **629000CF** | **629000CF** | 25466F31 | OJCS | EBML |

As mentioned, with a sophisticated noise injection mechanism based on natural language syntax and semantics, a noisy document could be made to appear perfectly readable, thus completely misleading the unauthorized reader.

## VIII. CONCLUSION AND FUTURE WORK

We have started by showing that existing access control models still fall short when it comes to information flow control. To address this issue, we have developed the GBFC, a dedicated flow control model based on granularity, access through controlled references, flow restriction, availability and noise injection. GBFC takes advantage of the impressive power that offers the combination of these techniques to build a robust solution. We showed the ability of the GBFC model to handle classified granular information in the form of references to enforce flow control even in extreme situations such as data loss, malicious attacks and deliberate information leakage. Furthermore, this model offers ways for end-to-end security and traceability of classified information. The centralized structure of the model offers great controllability, and adaptability to various security environments ranging from single security domain to heterogeneous multi-domains. This makes the GBFC well adapted to Cloud Computing because each granule can be allocated independently in the Cloud. Moreover, the centralized architecture can be easily implemented at operating system level, transforming each workstation into a fully independent security control platform that ensures for any user full information flow control while enforcing, by the same occasion, privacy and copyright management. On a wider scope, the method could be generalized to other forms of information such as images, audio, video and other multimedia data structures.

Future work will focus on further development of the concepts introduced in this paper, together with demonstration prototypes.

## REFERENCES


[1] http://wikileaks.org/

[2] B. Hicks, S. Rueda, L. St. Clair, T. Jaeger, P. McDaniel, "A logical specification and analysis for SELinux MLS policy," Transactions on Information and System Security (TISSEC) Volume 13, Issue 3, 2010.

[3] W. Masri, "Dynamic information flow analysis, slicing and profiling," PhD thesis, Electrical Engineering and Computer Science Department, Engineering School, Case Western Reserve University, Cleveland, Ohio, USA, 2005.

[4] US Department of Defence, "DoD Directive- Security requirements for automated information systems," (AIS) 8500.28. DoD, 1988.

[5] INTOSAI EDP Audit Committee, "ISSAI 5310 – Information system security review methodology," International Organization of Supreme Audit Institutions, 1995.

[6] K. Bolshakov, E. Reshetova, "FreeBSD Mandatory access control usage for implementing enterprise security policies," In Proceedings of CoRR. 2007.

[7] H. Mantel, "Information flow control and applications - Bridging a Gap," in Proc. FME, 2001, pp.153-172.

[8] G. Lowe, "Defining information flow," In Proc. IEEE Computer Security Foundations Workshop, 1999, pp. 18-31.

[9] D. E. Denning, "A lattice model of secure information flow," presented at Commun. ACM, 1976, pp. 236-243.

[10] R. S. Sandhu,, "Lattice-based access controls," IEEE Computer, Vol. 26, No. 11, 2001, pp.9-19.

[11] S. O. Hwang, K. S. Yoon, "Privacy protection in ubiquitous computing based on privacy label and information flow," ICCSA 2, Springer Vol. 3044, 2004, pp. 46-54.

[12] A. Sabelfeld, A. C. Myers, "Language-based information-flow security," IEEE Journal on selected areas in communications, vol. 21, no. 1, 2003.

[13] H. F. Tipton, Micki Krause, *Information security management handbook*, 6th Ed.,Auerbach Publications, 2007, pp. 15-45.

[14] D. E. Bell, L. J. La Padula, "Secure computer systems: unified exposition and multics interpretation," Bedford, MA:The Mitre Corporation, 1976.

[15] K. J. Biba, "Integrity considerations for secure computer systems," MTR-3153, The Mitre Corporation, 1977.

[16] M. T. Siponen, Harri Oinas-Kukkonen, "A review of information security issues and respective research contributions," SIGMIS Database, vol. 38, no. 1, 2007, pp. 60-80.

[17] A. V.D.M. Kayem, Patrick Martin, Selim G. Akl, "A presentation of access control methods," Adaptive Cryptographic Access Control, Advances in Information Security, Vol 48, 2010, pp. 11-39.

[18] R. S. Sandhu, Edward J. Coyne, Hal L. Feinstein, Charles E. Youman, "Role-Based Access Control models," IEEE Computer, Volume 29, Number 2, 1996, pp. 38-47.

[19] M. Bishop, Introduction to computer security, 1st Ed,Addison-Wesley Professional, 2004, pp. 261-285.

[20] A. C. Myers, B. Liskov, "A decentralized model for information flow control," In Proc. 17th ACM Symp. on Operating System Principles (SOSP), 1997, pp. 129–142.

[21] J. T. Yao, "A Ten-year Review of Granular Computing.", 'GrC', IEEE, 2007, pp. 734-739 .

[22] L. A. Zadeh, "Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," Fuzzy Sets and Systems, 90(2), 1997, pp. 111-127.

[23] Information Security Oversight Office, "Marking Classified National Security Information", ISOO Archives, Octobre 2007

[24] Center for Development of Security Excellence, "Marking Classified Information", CDSE Learn. Perform. Protect., Octobre 2012

[25] J. R. Vacca, *Computer and Information Security Handbook*, 2nd Edition, Morgan Kaufmann publisher, 2009

[26] McAfee, "Data Loss by the Numbers," McAfee White Paper , March 26 2012

[27] InfoWatch Analytical Labs, "Global Data Leakages & Insider Threats,"InfoWatch Report, 2012

[28] Open Security Foundation, "Data Loss Statistics," DataLossDB, http://datalossdb.org, 2013.

[29] A. Silberschatz, P. B. Galvin, G. Gagne, *Operating System Concepts*, 9th edition, John Wiley & Sons, 2013, pp. 543-586.

[30] S. L. Osborn, "Information Flow Analysis in RBAC Systems," SACMAT '02 Proc. of the seventh ACM symposium on Access control models and technologies, pp. 163-168.

[31] S. L. Osborn, R. S. Sandhu, Q. Munawer, "Configuring Role-Based Access Control to Enforce Mandatory and Discretionary Access Control Policies," ACM Trans. on Information and System Security, 3(2), 2000, pp. 85–106.

[32] J. Park, R. S. Sandhu, "Towards usage control models: beyond traditional access control," SACMAT, 2002, pp. 57-64.

[33] Joint Chiefs Of Staff, "Emergency action procedures of the Joint Chiefs ot. Staff: Nuclear Control Orders," EAP.JCS, Volume V, 1985, pp. 6.