IRM PRESS
701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.irm-press.com
ITB11702

This chapter appears in the book, *Web and Information Security*
edited by Elena Ferrari and Bhavani Thuraisingham © 2006, Idea Group Inc.

Chapter VI

Web Content Filtering

Elisa Bertino, Purdue University, USA

Elena Ferrari, University of Insubria at Como, Italy

Andrea Perego, University of Milan, Italy

Abstract

The need to filter online information in order to protect users from possible harmful content can be considered as one of the most compelling social issues derived from the transformation of the Web into a public information space. Despite that Web rating and filtering systems have been developed and made publicly available quite early, no effective approach has been established so far, due to the inadequacy of the proposed solutions. Web filtering is then a challenging research area, needing the definition and enforcement of new strategies, considering both the current limitations and the future developments of Web technologies—in particular, the upcoming Semantic Web. In this chapter, we provide an overview of how Web filtering issues have been addressed by the available systems, bringing in relief both their advantages and shortcomings, and outlining future trends. As an example of how a more accurate and flexible filtering can be enforced, we devote the second part of this chapter to describing a multi-strategy approach, of which the main characteristics are the integration of both list- and metadata-based

techniques and the adoption of sophisticated metadata schemes (e.g., conceptual hierarchies and ontologies) for describing both users' characteristics and Web pages content.

Introduction

In its general meaning, information filtering concerns processing a given amount of data in order to return only those satisfying given parameters. Although this notion precedes the birth of the Internet, the success and spread of Internet-based services, such as e-mail and the Web, resulted in the need of regulating and controlling the network traffic and preventing the access, transmission, and delivery of undesirable information.

Currently, information filtering is applied to several levels and services of the TCP/IP architecture. Two typical examples are spam and firewall filtering. The adopted strategies are various, and they grant, in most cases, an efficient and effective service. Yet, the filtering of online multimedia data (text, images, video, and audio) is still a challenging issue when the evaluation of their semantic meaning is required in order to verify whether they satisfy given requirements. The reason is that the available techniques do not allow an accurate and precise representation of multimedia content. For services like search engines, this results in a great amount of useless information returned as a result of a query. The problem is much more serious when we need to prevent users from retrieving resources with given content (e.g., because a user does not have the rights to access it or because the content is inappropriate for the requesting user). In such a case, filtering must rely on a thorough resource description in order to evaluate it correctly.

The development of the Semantic Web, along with the adoption of standards such as MPEG-7 (TCSVT, 2001) and MPEG-21 (Burnett et al., 2003), may seemingly overcome these problems in the future. Nonetheless, currently online information is unstructured or, in the best case, semi-structured, and this is not supposed to change in the next few years. Thus, we need to investigate how and to what extent the available techniques can be improved to allow an effective and accurate filtering of multimedia data.

In this chapter, we focus on filtering applied to Web resources in order to avoid possibly harmful content accessed by users. In the literature, this is usually

referred to as Internet filtering, since it was the first example of information filtering on the Internet with relevant social entailments—that is, the need to protect given categories of users (e.g., children) from Web content not suitable for them. The expression is misleading since it could apply also to Internet services like spam and firewall filtering, already mentioned above. More properly, it should be called Web filtering, and so we do henceforth.

Web filtering is a rather challenging issue, the requirements of which have not been thoroughly addressed so far. More precisely, the available Web filtering systems focus on two main objectives: *protection* and *performance*. On one hand, filtering must prevent at all costs harmful information from being accessed—that is, it is preferable to block suitable content rather than allowing unsuitable content to be displayed. On the other hand, the filtering procedure must not noticeably affect the response delay time—that is, the evaluation of an access request and the returned result must be quickly performed.

Such restrictive requirements are unquestionably the most important in Web filtering since they grant effectiveness with respect to both user protection and content accessibility. Yet, this has resulted in supporting the rating of resource content and users' characteristics which is semantically poor. In most cases, resources are classified into a very small set of content categories, whereas only one user profile is provided. The reason is twofold. On one hand, as already mentioned above, the way Web information is encoded and the available indexing techniques do not allow us to accurately rate Web pages with the precision needed to filter inappropriate content. On the other hand, a rich semantic description would require high computational costs, and, consequently, the time needed to perform request evaluation would reduce information accessibility. These issues cannot be neglected, but we should not refrain from trying to improve flexibility and accuracy in rating and evaluating Web content. Such features are required in order to make filtering suitable to different user's requirements, unlike the available systems which have rather limited applications.

In the remainder of this chapter, besides providing a survey of the state-of-the-art, we propose extensions to Web filtering which aim to overcome the current drawbacks. In particular, we describe a *multi-strategy* approach, formerly developed in the framework of the EU project EUFORBIA,[1] which integrates the available techniques and focuses on the use of metadata for rating and filtering Web information.

Web Rating and Filtering Approaches

Web filtering concerns preventing users from accessing Web resources with inappropriate content. As such, it has become a pressing need for governments and institutional users because of the enormous growth of the Web during the last 10 years and the large variety of users who have access to and make use of it. In this context, even though its main aim is minors' protection from harmful online contents (e.g., pedophilia, pornography, and violence), Web filtering has been and still is considered by institutional users—for instance, firms, libraries, and schools—as a means to avoid the improper use of their network services and resources.

Web filtering entails two main issues: *rating* and *filtering*. Rating concerns the classification (and, possibly, the labeling) of Web sites content with respect to a filtering point of view, and it may be carried out manually or automatically, either by third-party organizations (*third-party rating*) or by Web site owners themselves (*self-rating*). Web site rating is not usually performed on the fly as an access request is submitted since it would dramatically increase the response time. The filtering issue is enforced by *filtering systems*, which are mechanisms able to manage access requests to Web sites, and to allow or deny access to online documents on the basis of a given set of policies, denoting which users can or cannot access which online content and the ratings associated with the requested Web resource. Filtering systems may be either client- or server-based and make use of a variety of filtering techniques.

Currently, *filtering services* are provided by ISP, ICT companies, and non-profit organizations, which rate Web sites and make various filtering systems available to users. According to the most commonly used strategy—which we refer to as the traditional strategy in what follows—Web sites are rated manually or automatically (by using, for instance, neural network-based techniques) and classified into a predefined set of categories. Service subscribers can then select which Web site categories they do not want to access. In order to simplify this task, filtering services often provide customized access to the Web, according to which some Web site categories are considered inappropriate by default for certain user categories. This principle is also adopted by some search engines (such as Google *SafeSearch*) which return only Web sites belonging to categories considered appropriate.

Another rating strategy is to attach a *label* to Web sites consisting of some metadata describing their content. This approach is adopted mainly by PICS-

based filtering systems (Resnick & Miller, 1996). PICS (Platform for Internet Content Selection) is a standard of the World Wide Web Consortium (W3C) which specifies a general format for *content labels*. A PICS content label describes a Web page along one or more dimensions by means of a set of *category-value* pairs, referred to as *ratings*. PICS does not specify any labeling vocabulary: this task is carried out by *rating* or *labeling services*, which can specify their own set of PICS-compliant ratings. The filtering task is enforced by tools running on the client machine, often implemented in the Web browser (e.g., both Microsoft Internet Explorer and Netscape Navigator support such filters).[2]

Compared to the category models used in the traditional strategy, PICS-based rating systems are semantically richer. For instance, the most commonly used and sophisticated PICS-compliant rating system, developed by ICRA (Internet Content Rating Association: www.icra.org), provides 45 ratings, grouped into five different macro-categories: *chat, language, nudity and sexual material, other topics* (promotion of tobacco, alcohol, drugs and weapons use, gambling, etc.), *violence*.[3] On the other side, the category model used by RuleSpace EATK™, a tool adopted by Yahoo and other ISPs for their parental control services, makes use of 31 Web site categories.[4]

On the basis of the adopted filtering strategies, filtering systems can then be classified into two groups: *indirect* and *direct filtering*. According to the former strategy, filtering is performed by evaluating Web site ratings stored in repositories. These systems are mainly based on *white* and *black lists*, specifying sets of good and bad Web sites, identified by their URLs. It is the approach adopted by traditional filtering systems. The same principle is enforced by the services known as *walled gardens*, according to which the filtering system allows users to navigate *only* through a collection of preselected good Web sites. Rating is carried out according to the third-party rating approach.

Direct filtering is performed by evaluating Web pages with respect to their actual content, or the metadata associated with them. These systems use two different technologies. *Keyword blocking* prevents sites that contain any of a list of banned words from being accessed by users. Keyword blocking is unanimously considered the most ineffective content-based technique, and it is currently provided as a tool which can be enabled or disabled by user's discretion. *PICS-based filtering* verifies whether an access to a Web page can be granted or not by evaluating (a) the content description provided in the PICS label possibly associated with the Web page and (b) the filtering policies

specified by the user or a supervisor. PICS-based filtering services adopt a self-rating approach, usually providing an online form which allows Web site owners to automatically generate a PICS label.

According to analyses carried out by experts during the last years,[5] both indirect and direct filtering techniques have several drawbacks, which can be summarized by the fact that they *over*- or *under*-block. Moreover, both category models adopted by traditional rating services and PICS-based rating systems have been criticized for being too Western-centric so that their services are not suitable for users with a different cultural background. Those analyses make clear that each filtering technique may be suitable (only) for certain categories of users. For instance, though white and black lists grant a very limited access to the Web, which is not suitable for all users, these techniques, and especially walled garden-based services, are considered as the safest for children.

The available PICS-based rating and filtering services share similar drawbacks. The content description they provide is semantically poor: for instance, none of them makes use of ontologies, which would allow a more accurate content description. Moreover, it is limited to content domains considered liable to be filtered according to the Western system of values. Nevertheless, the PICS standard can be considered, among the available technologies, the one which can better address the filtering issues. Since its release in 1996, several improvements have been proposed, first of all, the definition of a formal language for PICS-compliant filtering rules, referred to as *PICSRules* (W3C, 1997). Such a language makes the task of specifying filtering policies according to user profiles easier, and it could be employed by user agents to automatically tailor the navigation of users. Finally, in 2000, the W3C proposed an RDF implementation of PICS (W3C, 2000), which provides a more expressive description of Web sites content, thus enabling more sophisticated filtering.

Despite these advantages, the PICS-based approach is the less diffused. The reason is that it requires Web sites to be associated with content labels, but currently, only a very small part of the Web is rated. PICS-based services, like ICRA, have made several efforts to establish the practice of self-rating among content providers, but no relevant results have been obtained. Moreover, both the PICS extensions mentioned above (PICSRules and PICS/RDF) did not go beyond the stage of proposal. Nonetheless, the wider and wider use of XML and its semantic extensions—that is, RDF (W3C, 2004b) and OWL (W3C, 2004a)—make metadata-based approaches the major research issue in the filtering domain. Consequently, any improvement in Web filtering must take into

consideration both the current limitations and the future developments of Web technologies.

So far, we have considered how the rating and filtering issues have been addressed with respect to the *object* of a request (Web sites). Yet, which content is appropriate or inappropriate for a user depends on his/her characteristics—that is, access to a given Web site can be granted or prevented depending on the requesting user. Consequently, filtering may be considered as a process entailing the evaluation of both subjects (users) and objects (Web sites) properties. The simplest case is when users share the same characteristics: since only a single *user profile* is supported, the filtering parameters are set by default, and there is no need to evaluate the characteristics of a single user. This applies also when we have one or more *predefined* user profiles or when filtering policies concern specific users. Such an approach, which we refer to as *static user profiling*, is the one adopted by the available filtering systems, and it has the advantage of simplifying the evaluation procedure (only Web sites characteristics must be evaluated), which reduces the computational costs. On the other hand, it does not allow flexibility; thus, it is not suitable in domains where such a feature is required.

Following, we describe a filtering approach which aims at improving, extending, and making more flexible the available techniques by enforcing two main principles: (a) support should be provided to the different rating and filtering techniques (both indirect and direct), so they can be used individually or in combination according to users' needs and (b) users' characteristics must be described accurately in order to provide a more effective and flexible filtering.

Starting from these principles, we have defined a formal model for Web filtering. The objective we pursued was to design a general filtering framework, addressing both the flexibility and protection issues, which can possibly be customized according to users' needs by using only a subset of its features.

Multi-Strategy Web Filtering

In our model, users and Web pages (resources, in the following) are considered as entities involved in a communication process, characterized by a set of properties and denoted by an identifier (i.e., a URI). A filtering policy is a rule, stating which users can/cannot access which resources. The users and re-

sources to which a policy applies are denoted either explicitly by specifying their identifiers, or implicitly by specifying constraints on their properties (e.g., "all the users whose age is less than 16 cannot access resources with pornographic content"). Note that the two types of user/resource specifications (explicit and implicit) are abstractions of the adopted filtering strategies: explicit specifications correspond to list-based approaches (i.e., white/black lists and walled gardens); implicit specifications merge all the strategies based on ratings (e.g., PICS) and/or content categories (e.g., RuleSpace).

Users' and resources properties are represented by using one or more *rating systems*. Since rating systems are organized into different data structures, and support should be provided to more complex and semantically rich ones (e.g., ontologies), we represent them as sets of ratings, hierarchically organized and characterized by a set, possibly empty, of attributes. That is to say, we model rating systems as ontologies. This approach has two advantages. The uniform representation of rating systems allows us to enforce the same evaluation procedure. Thus, we can virtually support any current and future rating system without the need of modifying the model. Moreover, the hierarchical structure into which ratings are organized allows us to exploit a *policy propagation* principle according to which a policy concerning a rating applies also to its children. This feature allows us to reduce as much as possible the policies to be specified, keeping their expressive power.

Our model is rather complex considering the performance requirements of Web filtering. Its feasibility must then be verified by defining and testing strategies for optimizing the evaluation procedure and reducing as much as possible the response time. We addressed these issues during the implementation of the model into a prototype, the first version of which was the outcome of the EU project EUFORBIA (Bertino, Ferrari, & Perego, 2003), and the results have been quite encouraging. The current prototype greatly improves the computational costs and the performance of the former ones. Thanks to this, the average response delay time is now reduced to less than 1 second, which does not perceivably affect online content accessibility.

Following, we describe the main components of the model and the architecture of the implemented prototype, outlining the strategies adopted to address performance issues.

The MFM Filtering Model

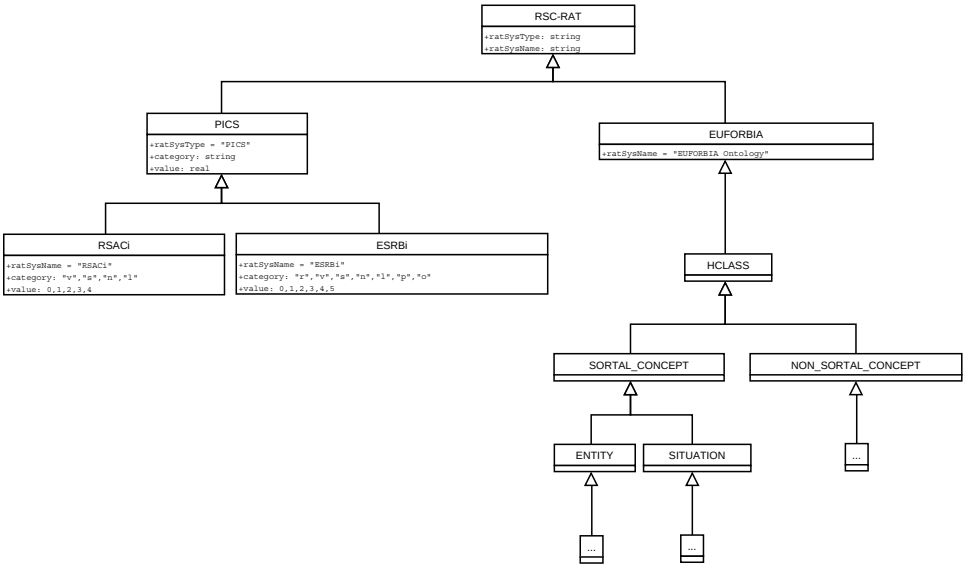
In our model, referred to as MFM (*Multi-strategy Filtering Model*), we use the notion of *agent* to denote both the active and passive entities (users and resources, respectively) involved in a communication process, the outcome of which depends on a set of *filtering policies*. An agent *ag* is a pair (*ag_id*, *ag_pr*), where *ag_id* is the agent identifier (i.e., a URI), and *ag_pr* are the agent *properties*. More precisely, *ag_pr* is a set, possibly empty, of *ratings* from one or more *rating systems*. Users and resources are then formally represented according to the general notion of agent by the pairs (*usr_id*, *usr_pr*) and (*rsc_id*, *rsc_pr*), respectively.

As mentioned above, MFM rating systems are modeled into a uniform structure—that is, as a set of ratings hierarchically organized. Moreover, the whole set of rating systems is structured into two *super-trees*, one for rating systems applying to users and one for rating systems applying to resources. Each super-tree has a root node at level 0 of the hierarchy, which is the parent of the root node of each rating system. As a result, the several rating systems possibly supported are represented as two single rating systems, one for users and one for resources. This *extrinsic* rating system integration totally differs from approaches as the ABC-based one (Lagoze & Hunter, 2001) aiming to provide semantic interoperability among ontologies. Yet, all the attempts to define a rating *meta*-scheme, representing the concepts commonly used in the available rating systems, have been unsuccessful so far. Our objective is then to harmonize only their structure in order to easily specify policies ranging over different rating systems.

Figure 1 depicts an example of a resource rating system super-tree, whose root is RSC-RAT, merging two PICS-based rating systems (namely, the RSAC*i* and ESRB*i* ones[6]) and the conceptual hierarchy developed in the framework of the project EUFORBIA. For clarity's sake, in Figure 1, only the upper levels of the tree are reproduced.

Ratings are denoted by an identifier and characterized by a set, possibly empty, of attributes and/or a set of *attribute-value* pairs. We then represent a rating *rat* as a tuple (*rat_id*, *attr_def*, *attr_val*), where *rat_id* is the rating identifier (always a URI), *attr_def* is a set, possibly empty, of pairs (*attr_name*, *attr_domain*), defining the name and domain of the corresponding attributes, and *attr_val* is a set, possibly empty, of pairs (*attr_name*, *value*).

Figure 1. Example of a resource rating system super-tree

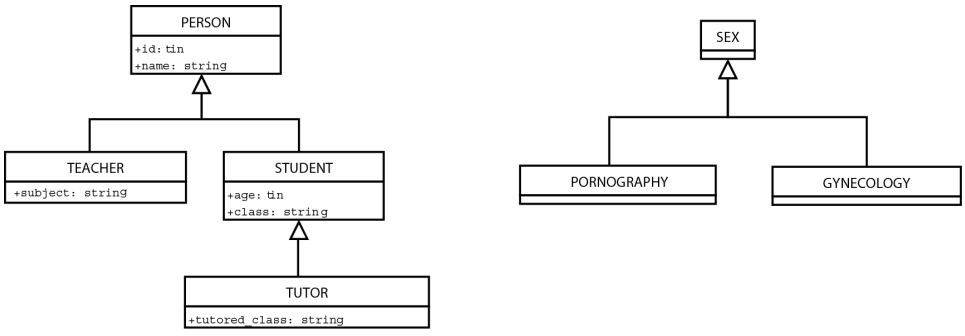


Attributes and attribute-value pairs are ruled according to the object-oriented approach. Thus, (a) attributes and attribute-value pairs specified in a rating are inherited by its children, and (b) attributes and attribute-value pairs redefined in a child rating override those specified in the parents. Examples of ratings are those depicted in Figure 1.

The last MFM key notion to consider before describing filtering policies is that of *agent specification*. To make our examples clear, following, we refer to Figure 2, depicting two simple user (a) and resource (b) rating systems.

In MFM, we can denote a class of agents by listing them explicitly (*explicit agent specifications*) or specifying constraints on their properties (*implicit agent specifications*). Implicit agent specifications are defined by a *constraint specification language* (CSL), which may be regarded as a Description Logic (Baader et al., 2002) providing the following constructs: concept intersection, concept union, and comparison of concrete values (Horrocks & Sattler, 2001). In CSL, rating systems are then considered as IS-A hierarchies over which concept operations are specified in order to identify the set of users/resources satisfying given conditions.

Figure 2. Examples of user (a) and resource (b) rating systems



(a) *(b)*

Examples of CSL expressions referring to the rating system in Figure 2(a) may be TEACHER ⊔ TUTOR (i.e., "all the users associated with a rating TEACHER or TUTOR") and STUDENT.`age > 14` (i.e., "all the users associated with a rating STUDENT and whose age is greater than 14"). Note that thanks to the rating hierarchy, a CSL expression concerning a rating applies also to its children. For instance, the CSL expression STUDENT.`age > 14` denotes also the users associated with a rating TUTOR, provided that their age is greater than 14.

A filtering policy is then a tuple (*usr_spec*, *rsc_spec*, *sign*), where *usr_spec* is an (explicit or implicit) agent specification denoting a class of users (*user specification*); *rsc_spec* is an (explicit or implicit) agent specification denoting a class of resources (*resource specification*); and *sign* ∈ {+,–} states whether the users denoted by *usr_spec* can (+) or cannot (–) access resources denoted by *rsc_spec*.

The sign of a policy allows us to specify exceptions with respect to the propagation principle illustrated above. Exceptions are ruled by a conflict resolution mechanism according to which, between two conflicting policies (i.e., policies applying to the same user and the same resource but with different sign), the prevailing is the more specific one.

Finally, when the rating hierarchies cannot be used to solve the conflict, negative policies are considered prevalent. This may happen when the user/resource specifications in two policies are equivalent or when they denote disjoint sets of ratings, which are yet associated with the same users/resources.

Example 1.

Let us suppose that no user can access contents regarding the sexual domain, unless he/she is a teacher or a tutor. Moreover, we allow students whose age is greater than 14 to access contents regarding gynecology. These requirements can be enforced by specifying the following validations: pv = (PERSON, SEX, −), fp_2 = (TEACHER ⊔ TUTOR, SEX, +), and fp_3 = (STUDENT.age > 14, GYNECOLOGY, +). It is easy to realize that policies fp_2 and fp_3 are in conflict with fp_1. Nonetheless, according to our conflict resolution mechanism, fp_2 is more specific than fp_1 since TEACHER and TUTOR are children of PERSON, whereas fp_3 is more specific than fp_1 since STUDENT is a child of PERSON, and GYNECOLOGY is a child of SEX. As a consequence, fp_2 and fp_3 prevail over fp_1. Consider now a 15-year-old user, whose identifier is Bob,[7] associated with a rating rat1 instance of STUDENT, and a policy fp_4 = ({Bob}, GYNECOLOGY, −): fp_4 prevails over fp_3 since the user specification is explicit and therefore is more specific. The same principle applies to resources. Thus, given a Web site www.example.org, associated with a child of the rating SEX, a policy fp_5 = (PERSON, {www.example.org}, +) prevails over fp_1.

The propagation principle and the conflict resolution mechanism are also applied to resources by exploiting the URI hierarchical structure (IETF, 1998). Thus, a policy applying to a given Web page applies as well to all the resources sharing the same URI upper components (e.g., a policy applying to www.example.org applies as well to www.example.org/examples/). In case of conflicting policies, the stronger is the one concerning the nearer resource with respect to the URI syntax.

As demonstrated by the example above, our approach allows us to reduce, as much as possible, the set of policies which need to be specified providing high expressive power and flexibility. Moreover, MFM is fully compliant with RDF and OWL. This implies that we can use RDF/OWL as standard cross-platform encoding for importing/exporting data structured according to our model among systems supporting these technologies.

Supervised Filtering

In MFM, policies are specified only by the System Administrator (SA), and no mechanism to delegate access permissions is supported. Yet, in some domains, the responsibility of specifying filtering policies may be shared among several

persons, and in some cases, the opinions of some of them should prevail. For instance, in a school context, teachers' opinions should prevail over the SA's, and parents' opinions should prevail over both the teachers' and the SA's. Moreover, it could be often the case that certain categories of users (e.g., minors) should be subject to a very restrictive access to the Web according to which they can access a given Web page only if their supervisors agree.

In order to satisfy these requirements, MFM also supports the notion of *supervised filtering*, according to which a given set of users in the system, referred to as *supervisors* (*SV* for short), should validate the filtering policies specified for a given set of users, referred to as *supervised users* (*SU* for short), are valid or not before they can be effective. Additionally, sometimes a kind of supervision is required according to which access request made by a user must be approved by a supervisor, even though it is authorized by a filtering policy. Different sets of supervisors may correspond to different sets of supervised users, and supervisors cannot belong to the corresponding set of supervised users.

Supervised users are grouped into two disjoint subsets, denoted by two different *supervision modes*, *normal* and *strict*. Depending on the supervision mode associated with a supervised user, supervised filtering is enforced either *implicitly* or *explicitly*, respectively. In the normal supervision mode, a filtering policy applies to a supervised user only if the corresponding supervisors agree. In the strict supervision mode, a supervised user can access a specific resource only if the corresponding supervisors agree. Thus, filtering policies applying to users associated with a normal supervision mode must be validated by the authorized supervisors. On the other hand, access requests to Web pages submitted by users associated with a strict supervision mode must be authorized by a filtering policy and *explicitly* by supervisors.

The supervised filtering component of MFM (MFM-SF) is modeled, as the base one, according to an agent-oriented approach: *active* agents are *supervisors*, and *passive* agents are *supervised users*. Users with the role of supervisor are denoted by associating *supervisor ratings* with them, hierarchically structured into one or more *supervisor rating systems*, which are in turn grouped into a super-tree. We can then denote, either explicitly or implicitly, a class of supervisors by means of a *supervisor specification*—that is, an agent specification concerning supervisors. Similarly, a class of supervised users is denoted by a *supervised user specification*—namely, a user specification applying only to the set of supervised users. If we denote by *US* and *SUS* the sets of user and supervised user specifications, respectively, it follows

that $SUS \subseteq US$. Finally, a set of supervised users and the corresponding set of supervisors are denoted by a *supervision specification*, which is a pair (*sv_spec*, *su_spec*), where *sv_spec* and *su_spec* are, respectively, a supervisor and a supervised user specification.

In this context, the rating hierarchy is used to denote the authority of a supervisor according to the following principles: (a) child ratings have a stronger authority than their parents, whereas (b) ratings at the same level of the hierarchy have the same authority. For instance, let us suppose two supervisor ratings PARENT and TEACHER, such that PARENT $\prec$ TEACHER: we say that supervisors associated with the rating PARENT have stronger authority than those associated with the rating TEACHER.

We can now introduce the notions of *filtering policy validation* (*validation*, for short) and *supervised filtering policy*. A validation *fpv* is a tuple (*sv_id*, *su_spec*, *FP*, *sign*), where *sv_id* is the identifier of the supervisor who validated the set of policies *FP* for the users denoted by the supervised user specification *su_spec*, whereas *sign* $\in$ {+,–} states whether the policy is valid (+) or not (–). Similarly, a supervised filtering policy *sfp* is a tuple (*sv_id*, *usr_spec*, *RI*, *sign*), where *sv_id* is the identifier of the supervisor who decided that supervised users denoted by *su_spec* can or cannot, depending on the value of *sign* $\in$ {+,–}, access resources whose identifier is in *RI*.

Example 2

Let us suppose two users `Jane` and `Ted`, who are, respectively, `Bob`'s mother and teacher—consequently, `Jane` and `Ted` are associated with supervisor ratings PARENT and TEACHER, respectively. We can denote `Jane` and `Ted` as supervisors of `Bob` by an explicit supervision specification ({`Bob`}, {`Jane`, `Ted`}). We can also state that teachers are supervisors of students by an implicit supervision specification (STUDENT, TEACHER). Let us now suppose that Ted does not agree with fp_3—that is, he does not agree that 15-aged students can access content concerning gynecology. On the contrary, Jane agrees with fp_3, as far as her son Bob is concerned. Ted and Jane state their decisions by specifying the following validations: fpv_1 = (`Ted`, STUDENT, {fp_3}, –) and fpv_2 = (`Jane`, {`Bob`}, {fp_3}, +).

In the example above, we then have two conflicting validations. In order to identify the prevailing, we check the authority degree of each supervisor. Since `Jane` is a parent and `Ted` is a teacher, according to the supervisor rating hierarchy, fpv_2 prevails over fpv_1. If supervisors' authority is equally strong, in order to solve a possible conflict between validations, we compare the

specificity degree of their supervised user specification components. For instance, given a teacher Mary and a validation fpv_3 = (Mary, STUDENT.age = 15, {fp_3}, +), fpv_3 prevails over fpv_1 since it is more specific with respect to the supervised user specification. In case the two supervised user specifications are equally specific, the prevailing validation is the negative one.

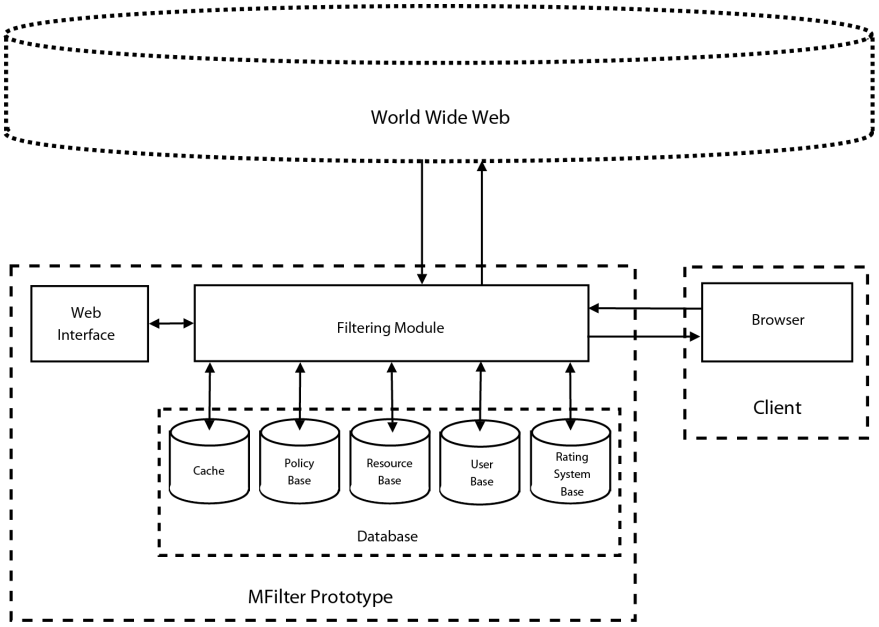
Finally, let us now consider the following examples of supervised filtering policies: sfp_1 = (Ted, {Bob}, {www.example.org}, –), sfp_2 = (Jane, {Bob}, {www.example.org}, +), and sfp_3 = (Mary, STUDENT, {www.example.org}, +). In case of conflict, the prevailing supervised filtering policy is identified according to the same principles illustrated above for filtering policies and validations. Thus, sfp_2 prevails over sfp_1 since Jane has a stronger authority than Ted, whereas sfp_1 prevails over sfp_3 since the supervised user specification in the former policy is more specific than the one in the latter.

Model Implementation: Architecture and Performance Issues

Figure 3 depicts the architecture of the implemented prototype, referred to as MFILTER. MFILTER is a Java-based system built on top of the Oracle DBMS, and it is structured into three main components. The first is the Filtering Module, which intercepts each access request submitted by users and verifies whether it can be granted or not according to the filtering policies specified by the SA and stored in the MFILTER Database. The second is the Database, which stores all the information needed by the system to perform the filtering tasks. The third is the Web Interface, structured into three main components (the *Administration*, the *Supervision*, and the *User Interfaces*), which allow the management of the system, user authentication, and the specification/validation of policies.

The core of the system is the Filtering Module (FM), which evaluates all the submitted access requests and returns the requested Web documents or an access denied message according to the filtering policies specified for the requesting user. The evaluation procedure is quite a complex task, dealing with several issues. Following, we focus on the main one: *performance*.

Figure 3. MFILTER architecture



Given an access request (*usr_id*, *rsc_id*), the FM computes the *policy projection* concerning the requesting user and the requested resource—that is, it retrieves all the policies applying to user with identifier *usr_id* and resource with identifier *rsc_id*—and determines the strongest one. Policy projection and evaluation have high computational costs which, if carried out at runtime, would dramatically increase the delay time between the access request and the system response. In order to improve the efficiency and effectiveness of the evaluation procedure and to reduce its computational costs at runtime, the FM enforces precomputational strategies and caching mechanisms, which are carried out by different submodules. Whenever a policy is specified or updated, the system computes all the users to whom the policy applies. Moreover, the system stores the URIs and metadata possibly associated with all the Web pages requested by the users. This allows us to precompute all the already requested resources to which a policy applies and, consequently, to greatly reduce the response delay time if a request concerns one of these Web pages. As a consequence, since a given set of users access mainly a precise subset of the Web, the best performance is reached as soon as the system has stored the information

concerning the usually accessed Web pages. Thus, system performance improves the more intensely the system is used.

Web Filtering and Access Control

In this chapter, we defined and discussed Web filtering only with respect to the need of preventing users from accessing inappropriate Web contents. Yet, the same techniques can be used whenever access to a resource can be granted only if given constraints, concerning either the end user or the resource, are satisfied. More precisely, Web filtering shares several similarities with access control. Although in the latter, *resources*, not *users*, are the entities to be protected, this does not necessarily affect how filtering/access control policies are formalized. Actually, a filtering policy is not different from an authorization: it simply states that a given set of users can/cannot access a given set of resources.

This is particularly true for MFM. Our model supports notions—such as negative/positive policies, policy propagation, and strong/weak policies—available in some discretionary access control models. The main difference concerns access *privileges*: MFM allows one to state only whether a resource can or cannot be viewed, whereas access control usually supports privileges concerning the modification of resource content (*write*, *update*, *append*, etc.). Yet, such privileges can be easily added in MFM by extending the notion of filtering policy and representing it by a tuple (*usr_spec*, *rsc_spec*, *priv*, *sign*), where *priv* denotes the access privilege. Moreover, we can organize privileges into a hierarchy in order to resolve conflicts: for instance, we can say that a write privilege is *stronger* than a read privilege; consequently, given two filtering policies *fp* = (PERSON, {`www.example.org`}, `write`, +) and *fp2* = (PERSON, {`www.example.org`}, `read`, –), *fp* is considered prevalent since it is stronger with respect to the access privilege.

Nonetheless, the possibility of specifying policies based on multiple rating systems and the support for supervision make MFM more flexible than traditional access control approaches. Such features are not provided by the available access control models but in a limited form. *Credentials* describing users' characteristics and policies based on the content of resources rely on predefined attribute/category sets, and supervision is usually not supported, unless in the rather different principle of *administration delegation*. As a

result, MFM is particularly suitable to enforce access control/filtering in Web and distributed environments, where the characteristics of users and resources may be described using different metadata vocabularies and where the policy specification task is not necessarily centralized.

Conclusions

Metadata-based filtering is today one of the most promising approaches to a controlled distribution of Web contents. By contrast, list- or keyword-based strategies have proved to enforce a very restrictive (and often rather ineffective) filtering, which, bounding its application to a very narrow set of user categories, has been one of the major reasons of the missed diffusion of these services among Web users. Web filtering *per se* is rather a sensitive problem which has been and is still debated by governments, industry, and libertarian institutions. The most diffused and well-grounded criticism is that it may enforce institutional censorship in online free speech. As a matter of fact, PICS aimed at overcoming this criticism by delegating and users or supervisors (e.g., parents) the responsibility of deciding which content is appropriate or inappropriate. The attempt has been unsuccessful since PICS-based rating systems provide a content classification which, besides being semantically poor, is constrained to what is considered liable to be filtered. As a consequence, a PICS rating system is already a judgment on what is inappropriate according to a given cultural background. Then, the PICS approach, in principle, is still valid, but it must provide a semantically richer rating of Web content, not constrained to a particular value system.

We moved in this direction in the framework of the EUFORBIA project, funded by the EU Safer Internet Action Plan (`www.saferinternet.org`) by defining a general-purpose rating system (namely, the EUFORBIA conceptual hierarchy) which allows one to accurately describe Web sites structure and content. EUFORBIA labels may be compared with the metadata scheme of a *digital item* (DI) in MPEG-21 (ISO, 2003), although MPEG-21 focuses mainly on the problem of describing the structure and the access rights of a DI. Such approach, currently not adopted by Web rating and filtering systems, will seemingly become the standard way to represent online content as the recent semantic extensions to MPEG-21 can testify (see Hunter, 2003). The possibility of enforcing an effective and flexible filtering, where end-users can freely

decide what they wish and what they do not wish to access, may probably solve the ethical disputes concerning Web filtering and may also reduce end user resistance in adopting them (Ying Ho & Man Lui, 2003), removing eventually two of the main obstacles to its diffusion.

The multi-strategy filtering model illustrated in this chapter is an attempt to address the drawbacks of the existing tools by integrating and optimizing the available technologies and to define a framework compliant with the upcoming Web technologies. In particular, the recent efforts of the W3C in defining a standard architecture for Web services (W3C, 2004c), which will incorporate Web content filtering into a broader framework, are quite relevant, concerning the evaluation of online information with respect to given user requirements. Starting from this, we are now investigating the feasibility of our approach in other filtering-related areas, more precisely that concerning the quality of service, where access to online resources is granted or prevented depending on whether the quality policies declared by the service provider satisfy those defined by users in their profile.

References

Baader, F., Calvanese, D., McGuinness, D., Nardi, D., & Patel-Schneider, P. (Eds.) (2002). *The description logic handbook: Theory, implementation, and applications*. Cambridge: Cambridge University Press.

Bertino, E., Ferrari, E., & Perego, A. (2003, November). Content-based filtering of Web documents: The MaX system and the EUFORBIA project. *International Journal of Information Security, 2*(1), 45-58.

Burnett, I., Van de Walle, R., Hill, K., Bormans, J., & Pereira, F. (2003, October-December). MPEG-21: Goals and achievements. *IEEE Multimedia, 10*(4), 60-70.

Cranor, L.F., Resnick, P., & Gallo, D. (1998, September). Technology inventory: A catalog of tools that support parents' ability to choose online content appropriate for their children. Retrieved May 2, 2005, from *http://www.research.att.com/projects/tech4kids*

Horrocks, I., & Sattler, U. (2001). Ontology reasoning in the SHOQ(D) description logic. In B. Nebel (Ed.), *Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI 2001)* (pp. 199-204). Morgan Kaufmann.

Hunter, J. (2003, January). Enhancing the semantic interoperability of multimedia through a core ontology. *IEEE Transactions on Circuits and Systems for Video Technology, 13*(1), 49-58.

IETF. (1998, August). Uniform Resource Identifiers (URI): Generic syntax. Retrieved May 2, 2005, from *http://www.ietf.org/rfc/rfc2396.txt*

Infopeople Project. (2001, May). Internet filtering options analysis: An interim report. Retrieved May 2, 2005, from *http://www.infopeople.org/howto/filtering/InternetFilter_Rev1.pdf*

ISO. (2003). *Information technology—Multimedia framework (MPEG-21)—Part 2: Digital item declaration*. Geneve: International Organization for Standardization.

Lagoze, C., & Hunter, J. (2001, November 6). The ABC ontology and model. *Journal of Digital Information.* Retrieved May 2, 2005, from *http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Lagoze*

Neumann, P.G., & Weinstein, L. (1999). Inside risks: Risks of content filtering. *Communications of the ACM, 42*(11), 152.

Resnick, P., & Miller, J. (1996, October). PICS: Internet access controls without censorship. *Communications of the ACM, 39*(10), 87-93.

Sobel, D. (Ed.). (2001). *Filters & freedom 2.0: Free speech perspectives on Internet content controls*. Washington, DC: Electronic Privacy Information Center.

TCSVT. (2001, June). *IEEE Transactions on Circuits and Systems for Video Technology—Special Issue on MPEG-7, 11*(6).

W3C. (1997, December 29). PICSRules 1.1. Retrieved May 2, 2005, from *http://www.w3.org/TR/REC-PICSRules*

W3C. (2000, March 27). PICS rating vocabularies in XML/RDF. Retrieved May 2, 2005, from *http://www.w3.org/TR/rdf-pics*

W3C. (2004a, February 10). *OWL Web ontology language: Overview*. Retrieved May 2, 2005, from *http://www.w3.org/TR/owl-features*

W3C. (2004b, February 10). *Resource description framework (RDF): Concepts and abstract syntax*. Retrieved May 2, 2005, from *http://www.w3.org/TR/rdf-concepts*

W3C. (2004c, February 10). *Web services architecture*. Retrieved May 2, 2005, from *http://www.w3.org/TR/ws-arch*

Weinberg, J. (1997, Winter). Rating the net. *Hasting Communications and Entertainment Law Journal, 19*(2), 453-482.

Ying Ho, S., & Man Lui, S. (2003). Exploring the factors affecting Internet content filters acceptance. *ACM SIG e-Com Exchange*, *4*(1), 29-36.

Endnotes

1 The EUFORBIA project Web site is available at http://e-msha.msh-paris.fr/Agora/ Tableaux%20de%20bord/Euforbia/

2 The complete PICS reference documentation is available online at http://www.w3.org/PICS

3 The list of ICRA ratings is available at http://www.icra.org/faq/decode

4 The list of RuleSpace categories is available at http://www.rulespace.com/products/models.php

5 See, for instance, Weinberg (1997), Cranor, Resnick, and Gallo (1998), Neumann and Weinstein (1999), Infopeople Project (2001), Sobel (2001).

6 We have chosen these rating systems since their simple structure is more suitable for examples. RSAC*i* (Recreational Software Advisory Council on the internet: www.rsac.org) is the predecessor of ICRA, and it is no longer available; ESRB*i* (Entertainment Software Rating Board Interactive: www.esrb.org) is a non-profit organization providing ratings for the entertainment software industry.

7 Following, for clarity's sake, we adopt symbolic identifiers for users instead of URIs.