Security Policies for Sharing Knowledge in Virtual Communities

Guido Boella and Leendert van der Torre

***Abstract*—Knowledge management exploits the new opportunities of sharing knowledge among members of virtual communities in distributed computer networks, and knowledge-management systems are therefore modeled and designed as multiagent systems. In this paper, normative multiagent systems for secure knowledge management based on access-control policies are studied. It is shown how distributed access control is realized by means of local policies of access-control systems for documents of knowledge providers, and by means of global community policies regulating these local policies. Moreover, it is shown how such a virtual community of multiple knowledge providers respects the autonomy of the knowledge providers.**

***Index Terms*—Knowledge management, multiagent systems, normative systems, policies, virtual communities.**

I. Introduction

MANY multiagent systems for knowledge management have been proposed. For example, KAoS[1] is a system for the management of technical information contained in documents, KRAFT [2] aims at fusing information from different sources, FRODO [3] represents distributed organizational memories, and MARS [4] is an adaptive social network for information access. Van Elst *et al.* [5] observe several reasons why multiagent systems are useful for knowledge management. The autonomy as well as proactiveness of agents help accommodating to the reality that knowledge workers typically adopt knowledge-management goals with a low priority. The flexibility to adapt to unforeseen situations of agents and multiagent systems helps knowledge management to deal with changing environments, such as the addition of new members to the knowledge-management system [6]. Moreover, agent technology provides tools and models to develop knowledge-management systems, for example, to incorporate legacy systems into modern distributed information systems.

However, there is no consensus on how multiagent systems can be used to deal with security considerations in knowledge management. In this paper, we address two requirements of secure distributed knowledge management.

1) Knowledge providers should not give up their autonomy to prohibit access to users they do not trust, even when they satisfy the security rules of the virtual community.
2) The policy rules for managing knowledge in a secure way concern not only which knowledge the users are prohibited or permitted to access, but also which regulations the community members are allowed or obliged to enforce.

Traditionally, multiagent systems are developed without security considerations in mind, by focussing on the coordination of cooperative agents. Likewise, knowledge-management systems are traditionally developed without any security considerations, for example, because the users of the systems know each other, share the same goal, and share a fixed set of administrative rules. When passing from knowledge management in teams to knowledge management in virtual communities, the flow of knowledge must be subject to restrictions [7], [8].

We introduce a model of policies for secure knowledge management in virtual communities satisfying the above requirements by modeling distributed knowledge management by means of normative multiagent systems (NMAS). Knowledge providers are modeled as normative systems creating their own norms. Global policies are enforced by detective rather than preventative control, such that global policies can be violated by knowledge providers if they believe this is necessary. Various types of agent types can be modeled to reflect ways in which agents can be motivated, such as norm internalizing agents, respectful agents, and selfish agents. Games among agents are used to analyze interaction in secure knowledge management.

We illustrate our model of secure knowledge management by highlighting an important subtlety involved in global policies regulating local policies. It is concerned with the notion of entitlement, which distinguishes between users that are only permitted to access knowledge, and users that are also entitled to it in the sense that knowledge providers are obliged to permit them access [9]. The additional issue discussed, formalized, and analyzed in this paper is that global policies do not only describe who is entitled to access resources, because the game-theoretic analysis shows that this is too weak to obtain the desired system behavior. Global policies do not only refer to the existence of local norms, but they have to refer also to the enforcement of the norms by the local provider recognizing and sanctioning violations.

This paper is organized as follows. In Section II, we discuss decentralized control in secure knowledge management. In Section III, we discuss our two requirements, which are detailed in Sections IV and V. In Sections VI and VII, we present the logical model, and in Section VIII, we apply it to an example of global policies concerned with entitlement.

Manuscript revised January 9, 2006. This paper was recommended by Associate Editors H. R. Rao and S. J. Upadhyaya.

G. Boella is with the Dipartimento di Informatica, Università di Torino, Torino 10149, Italy (e-mail: guido@di.unito.it).

L. van der Torre is with the University of Luxembourg, Luxembourg L-1359, Luxembourg (e-mail: leendert@vandertorre.com).

Digital Object Identifier 10.1109/TSMCA.2006.871793

II. Secure Knowledge Management

Knowledge management is a management discipline taking advantage of new technologies like peer-to-peer, multiagent, and normative systems respecting the distributed nature of knowledge in organizations.

A. Distributed Knowledge Management

Van Elst *et al.* define knowledge management as "the systematic, holistic approach for sustainably improving the handling of knowledge on all levels of an organization (individual, group, organizational and inter-organizational level) in order to support the organization's business goals, such as innovation, quality, cost effectiveness, etc." [5].

Knowledge is distributed in organizations, since the division of labor in modern companies leads to a distribution of resources, and knowledge management has to respect the distributed nature of knowledge in organizations. Bonifacio *et al.* explain that knowledge is the result of different perspectives and partial interpretations of world portions or domains, called the subjectivity and sociality of knowledge, and therefore, it should not be viewed as an absolute monolithic matter, but as a "system of local 'knowledges' continuously negotiated by communities of 'knowers'" [10].

Van Elst *et al.* observe that the distributed nature of knowledge in organizations "is not a 'bug' but rather a feature which is not only a matter of physical or technical location of some file" [5]. For example, knowledge management typically resides in environments subject to frequent changes, either in the organizational structure, in business processes, or in the information-technology infrastructure, and centralized solutions are often ill suited to deal with continuous modification in the enterprise.

B. Multiagent Systems

Due to technological innovations knowledge-management systems are no longer limited to a single organization [10]. Van Elst *et al.* observe that "the role of information technology as an enabling factor is also widely recognized and ... a variety of proposals exist showing how to support knowledge management with specialized information systems" [5].

For example, Bonifacio *et al.* argue that a member of a distributed knowledge-management system can be seen "as a peer [of a peer-to-peer system] that manages and has control over a set of local technologies, applications and services. ... In order to join the system, every member must provide resources or services to the others" [11].

As another example, Tacla and Barthes develop a multiagent-system architecture for knowledge management, because "knowledge is contained in the information produced, retrieved and exchanged among members of the projects," which results in "a distributed group memory where each member keeps part of the knowledge of the team." They introduce a multiagent-system architecture for knowledge-management systems, because, "like in a team, a multiagent system is composed by a group of possibly heterogeneous and autonomous agents" [12].

C. Secure Knowledge Management

Secure knowledge management is an important issue, because an organization's knowledge is easy to view, steal, manipulate, and delete. For example, Mundy and Chadwick discuss the need of secure knowledge management in the health-care industry. As this industry "enters the era of knowledge management it must place security at the foundation of the transition. Risks are pervasive to every aspect of information and knowledge management. ... In an age where risks and security threats are ever-increasing, secure knowledge management is an essential business practice" [13].

The national information systems security glossary (NSTISSI 4009) distinguishes the following five components of secure knowledge management. Authentication is a security measure to establish the validity of a message or verifying the eligibility of an individual to receive information. Authorization is the set of rights granted to a user to access, read, insert, or delete data. Data security or privacy is the protection from unauthorized disclosure. Data integrity is the requirement that data are unchanged from their source. Information security policy is the set of organizational guidelines, rules, regulations, and procedures that are used to protect an organization. It emphasizes that all components are essential and mutually supportive. For example, authentication without authorization would mean that only valid users could gain access but could execute any operation.

It also observes that, without an information security policy, there would be no apparent requirement for secure practices, and security would either be ignored or implemented in an *ad hoc* manner, and thus, control could be either too rigid or completely missing.

D. Normative Multiagent Systems and Access Control

Secure knowledge management takes advantage of new technologies like *normative* multiagent systems, such as access-control techniques, languages, and models for policy requirements, etc. Normative multiagent systems can be used as a software engineering method to develop secure knowledge-management systems, both providing high-level specifications as well as system realizations. Moreover, normative multiagent-system models can be used to analyze the system, or simulate solutions before they are built.

The NSTISSI acknowledges the importance of access control, and mentions various models. Discretionary access control can be viewed as an owner-based administration of users' rights, where the owner has authority over which other users can access the object. Discretionary access control has some limitations, in particular, in the way the owner can delegate his discretionary power to other people. Mandatory access control is based on security labels given to objects (security classification) and users (security clearance). Role-based access control has been developed to enhance maintenance and scalability. Permissions are granted to roles that are assigned to users. A security officer may have the right to assign roles to users without having an access right to the resources.

III. Two Requirements and Some Consequences

In this section, we discuss two requirements for the development of an integrated approach to distributed knowledge management and security, which we use to motivate and evaluate our multiagent model for secure knowledge management.

A. Autonomy of Local Knowledge Providers

A multiinstitutional or multiorganizational set of agents is a virtual community [14] if they use a set of rules, a policy, to specify how to share their resources, like documents and knowledge. A virtual community can be modeled as a system containing role-playing agents. Every agent in the community can play both the role of a document user as well as that of a document provider, because agents do not only use documents, but they also put the documents they own at disposal to the other participants of the community. Document providers retain the control of their documents and they specify in local policies the conditions of use of their documents.

The knowledge providers prefer not to give up their own power to enforce local policies for the access to the documents they control. For example, as Sadighi Firozabadi *et al.* [15] argue, there are "cases where security administrators are not fully trustworthy," for example, when multinational virtual communities are headed by foreign countries with varying security standards. The agents participating in the community are heterogeneous and change frequently, so they cannot be assumed to be always cooperative and to act in accordance to the system policies, both when requesting access to documents and when providing access to their documents. Decentralized authorities can cope in a better way with local idiosyncratic situations: "each party of the network can decide in each circumstance whether to accept credentials presented by a second party" [16]. Bonifacio *et al.* call it the principle of autonomy: "each community has a high degree of autonomy to manage its local knowledge" [10].

B. Local and Global Security Policies

However, the need of leaving autonomy to the knowledge providers must be balanced with the requirement that their local access policies should be organized according to a global policy defining how the knowledge should be shared among participants. The distinction between local and global policies is analogous to the distinction between local and global knowledge. As van Elst *et al.* [5] observe, "departments, groups and individual experts develop their particular views on a given subject. These views are motivated and justified by the particularities of the actual work, goals and situation. Obtaining a single, globally agreed upon vocabulary (or ontologies) within a level of detail which is sufficient for all participants, may incur high costs (e.g., for negotiation)" [5]. In virtual communities, global access control cannot be directly implemented, since nobody owns all the documents. There is no central manager of a system permitting agents to access the resources it owns and controls, according to the policies defined by itself. Such a centralized administration could be a too heavy burden and it can affect the core business activities of the system.

The centralized management of resources owned by the single resource providers is only partially performed by a *community agent* [7], which is in charge for maintaining the list of registered members and offering brokerage services for available information sources, but not for policing accesses to resources. The global policies are issued and enforced, for example, by a *community authorization service* (CAS): "A community runs a CAS server to keep track of its membership and fine-grained access-control policies. A user wishing to access community resources contacts the CAS server, which delegates rights to the user based on the request and the user's role within the community. These rights are in the form of capabilities which users can present at a resource to gain access on behalf of the community" [14].

Representations of local access-control policies can be obtained from existing theories of access control. Therefore, as observed by Pearlman *et al.* [14], "a key problem associated with the formation and operation of distributed virtual communities is that of how to specify and enforce community policies." In the remainder of this paper, this key problem is discussed and analyzed in our model of normative multiagent systems, which contains various incentives to motivate knowledge providers.

C. Incentives for Knowledge Providers

Since knowledge providers are on the one hand autonomous, but on the other hand cannot be coerced to provide their services or to deny them to users, it is necessary that these local providers are provided with incentives to implement the global policies by means of local ones. In other words, knowledge providers must be motivated by rewards and sanctions, such as the exclusion from the community. Moreover, there also has to be a monitoring system detecting violations and enforcing sanctions.

Sanction-based control implies also that an agent may influence negatively the behavior of other agents. In the terminology of Sichman *et al.* [17], other agents depend on it. In our model, this dependence is the essential precondition for the ability to issue policies. In this sense, the control of resources does not mean only that a given service is not provided if the provider does not want to, for example, the files of a web server cannot be accessed if it does not provide an answer to a request. Sanction-based control implies, for example, that agents depend on the global authority for their membership of the community. If knowledge providers do not behave according to its policies, then they are denied citizenship. At the local level, agents depend on the provider for the current and future access to the local resource.

Summarizing, we have to deal with the autonomy of local providers by modeling the security policies of the system, but also to limit the autonomy of agents, defining their responsibilities and roles. In particular, the respect of norms must be motivated in two distinct ways. Knowledge users must be provided with an incentive to respect the norms, and local knowledge providers must be motivated to issue policies respecting the global ones, and to enforce them.

IV. Normative Multiagent Systems

To deal with the requirement that local knowledge providers are autonomous, we model each knowledge provider as a normative system. We thus say that an agent is autonomous in the literal sense of "making its own norms." Consequently, in a virtual community, we have to deal with the interaction among normative systems.

A. Norms and Control

The role of norms in local and global policies is a major aspect of secure knowledge-management systems. We model a secure knowledge-management system as a normative multiagent system, which are "sets of agents ... whose interactions can be regarded as norm-governed; the norms prescribe how the agents ideally should and should not behave. ... Importantly, the norms allow for the possibility that actual behavior may at times deviate from the ideal, i.e., that violations of obligations, or of agents' rights, may occur" [18], where the norms are used to address security concerns of knowledge-management systems. Note that the notion of a normative system regulating an agent society has also fruitfully been employed elsewhere, such as electronic commerce, theories of fraud and deception, of trust dynamics and reputation, etc. [19].

Since there is no plausible way to enforce the respect of global policies by constraining the architecture, it is necessary to have a normative control mechanism to specify global policies about local policies. Normative systems contain control procedures, which are policies and procedures that help to ensure that management directives are carried out [20], because, intentionally or not, an agent may fail to comply with the policy. Moreover, the norms of global policies must be represented as soft constraints, which are used in detective control systems where violations can be detected, instead of hard constraints restricted to preventative control systems, in which violations are impossible [21]. A typical example of the former is that you can enter a train without a ticket, but you may be checked and sanctioned, and an example of the latter is that you cannot enter a metro station without a ticket. Detective control is the result of actions of agents and therefore subject to errors and influenceable by actions of other agents.

B. Secure Agent Interaction

The conceptual machinery of agents, obligations, norms, control, roles, policies, organizations, contracts, etc., offered by normative multiagent systems, is used to describe secure interaction among humans and systems. Agents must reason about the fulfillment of norms, the possible violations, and also what to do to repair such violations. A crucial point in secure knowledge management is the careful planning of moves during access control. Accordingly, agents must evaluate the effects of accessing documents both when making and evaluating an access request. In particular, since the compliance of the providers to the policies cannot be taken for granted, a user must consider whether they will fulfill their commitments or not. To make a prediction about the behavior of a local provider, it is necessary to consider also the reaction of the global authority enforcing control by monitoring and sanctioning violations.

C. Game-Theoretic Approach to Norms

Boella and Lesmo [22] analyze the motivational aspects of norms in the context of multiagent systems composed of heterogeneous agents, on the assumption that norms are useless unless supported by sanctions. They argue that sanctions must be modeled as actions of the normative system, since it is not possible to presuppose that they are mere consequences of violations. Hence, Boella and Lesmo attribute to the normative system the status of an agent deciding whether the behavior of agents counts as a violation, and thus deserves to be sanctioned by it.

Boella and Lesmo's model of agent interaction in normative systems is based on the philosophical foundations on strategic interaction in the work of sociologist Goffman [23]. "Strategic interaction" here means, according to Goffman, taking into consideration the actions of other agents. Boella and Lesmo call it the game-theoretic approach to norms.

Inspired by Boella and Lesmo's game-theoretic approach to norms, in earlier work [19], we propose a logical framework for reasoning about obligations and norms. It does not use a preventative control system, because agents are not constrained to respect norms. They can decide whether to respect norms or not based on a rational balance between the advantage of not respecting a norm and the disadvantage of being sanctioned. We use goal-based theories developed in artificial intelligence [24] and agent theory replacing probabilities and utilities by informational (knowledge, belief) and motivational attitudes (goal, desire), respectively, and the decision rule by a process of deliberation, in particular, based on the belief–obligation–intention–desire architecture (BOID) [25]. We replace the equilibria analysis in classical game theory by Gmytrasiewitcz and Durfee's recursive modeling [26].

D. New Challenges for Our Game-Theoretic Model

Distributed access control poses new challenges to the game-theoretic model of normative systems presented in [19]. The model must be extended with priorities among beliefs to model the defeasible effects of actions, with permissions as exceptions to model access rights, and, most importantly, with global policies regulating local ones. In [19], we consider agents modeling the contract partner recursively modeling the normative system. In distributed access control, however, there is no longer a single normative system, but each knowledge provider acts as one. In our games, we therefore have to deal with the interaction among normative systems. Agents recursively model the local provider, which in turn recursively model the global authority. Consequently, the relation between the local and global level, and thus the relation between local and global policies, plays a central role in the interaction among normative systems.

V. Global Policies Regulating Local Policies

To deal with the requirement to model global policies regulating local ones, we consider the rationale of global policies, called the transmission of will.

A. Transmission of Will

According to von Wright's principle of transmission of will, "an authority who orders that something be made obligatory wants the obligation satisfied. He, as it were, 'transmits' his will through the intermediary of a lower authority. Therefore, his will is not fulfilled unless the norms which are its immediate objects are themselves satisfied" [27, p.93].

For example, consider the notion of entitlement. Sadighi Firozabadi and Sergot [9] define entitlement to a resource by the obligation of a resource provider to permit access, and distinguish it from a mere permission of an agent to access the resource. The following scenario considers entitlement from the perspective of a local knowledge provider. An agent p has joined a virtual community n. Its contract for the participation prescribes that it should provide access to all the members of the community. Another participating agent, say agent a, tries to access agent p's system. However, previous experiences before joining the community advice agent p that agent a could damage its resources. Should agent p grant agent a access to its resources?

The fact that the agent a is entitled to access the resource, in the sense that the knowledge provider is obliged to permit him access, is not enough to ensure that he is given access to the resource, and therefore is not enough for the transmission of will. Global policies therefore do not refer only to the fact that a local norm exists, but also to the fact that the local provider enforces it by recognizing and sanctioning violations.

B. Policies for the Transmission of Will

In the above scenario for entitlement, the management of the community is organized in at least two levels: the global level (agent n) and the local one (agent p). Agent n is a distinguished authority, like the community authorization service playing the role of a global authority issuing global policies and negotiating the conditions for the participation of agents to the virtual community. Agent p is a provider of a document it controls. Moreover, all the agents (n, p, and a) can also play the role of users of the resources of the community.

Policies concern the behavior of participants. For example, at the global level, participants should not communicate their passwords, or distribute copyrighted files by means of the system. Otherwise, they are banned from the community, since the membership to the system is under the control of the global authority. At the local level, policies forbid agents to store files exceeding 1 GB on a file-sharing service, or they permit participants of the community to download copyrighted files from the web server.

Moreover, there are policies that apply to other policies, such as global policies that constrain or permit local policies [28]. In our scenario, agent n obliges agent p to permit members of the community to access its resources. Analogously, the global authority could oblige local ones to forbid access, permit-to-permit access, or permit-to-forbid access. However, it is not sufficient that the global obligation to permit or oblige access is satisfied by the fact that the local provider issues a permission or an obligation. Norms are ineffective if they are not enforced by the normative system who issued them: violations of norms should be recognized as such and sanctioned.

Reconsider the notion of entitlement. An agent n obliges agent p to permit agent a to do x if agent n obliges agent p not to consider $\neg x$ as a violation. The local normative system, however, can still violate this global policy and prevent access to users if it prefers to face the sanction with respect to permitting access. It is possible that agent p does not grant agent a the resource it is entitled to by the global policy, and this may be considered rational. Facing a sanction by the global authority, e.g., being excluded by the community for a certain period of time, is preferred to the possibility that agent a damages the system, e.g., agent a could create a denial-of-service attack. Thus, the autonomy of the provider is guaranteed, even if a provider violating a global policy must take the consequences of its actions into account. The example is formalized in Section VIII.

C. Transmission of Will in Our Game-Theoretic Model

In our model, obligations are defined in terms of goals of a normative system. The attribution of goals and beliefs to normative systems is an instance of Dennett's *intentional stance* [29]: Agents behave as if they are endowed with such motivational attitudes. A global obligation by agent n that agent p obliges agent a to do x implies an obligation that agent p considers $\neg x$ as a violation and sanctions it. Since, in turn, the obligation of agent n is expressed in terms of goals that something counts as a violation, the global obligation by agent n is defined as the goal that agent p considers $\neg x$ as a violation and the goal that if p does not do so, then its behavior is considered a violation by agent n. Analogously, a permission by agent n that p obliges that agent a does x is expressed as a permission by agent n to consider $\neg x$ as a violation: Agent n has the goal that agent p is not considered a violator by agent n if it considers agent a as a violator.

Our game-theoretic model highlights two important properties of the transmission of will. First, our game-theoretic model can be used to show that it is not sufficient that global norms refer only to the existence of local norms, by illustrating scenarios in which this is the case, but the local norm is not enforced.

Second, it gives an additional argument supporting the reduction of policies about policies to obligations and permissions about considering something as a violation or not. The agent does not have to be implemented in terms of explicit beliefs and goals. The basis for judging an agent cannot be its implementation, but the basis can be only its behavior. Analogously, the basis to say that an obligation is satisfied cannot be based on whether a knowledge provider has or does not have a goal. The only clue we have is its behavior, and in particular, whether it sanctions or not.

VI. Logical Framework

In this section, we present a simplified version of our logical model of normative multiagent systems introduced in [19]. We do not discuss constitutive norms, but we add priorities among beliefs, and undercutters to goals. In the following section, we discuss the extended games for modeling global policies regulating local ones.

A. Multiagent Systems

We first introduce the structural concepts and their relations. A set of propositional variables X describes the aspects of the world, and we extend it to literals built out of X ($Lit(X)$) to consider also the absence of states of affairs. Rules built out of the literals ($Rul(X)$) describe the relations among the propositional variables. A rule $l_1 \wedge \ldots \wedge l_n \rightarrow l$ is a pair of a set of literals built from X and a literal built from X: Rules represent the relations among literals existing in the agent's mental attitudes.

Definition 1: Let X be a set of propositional variables. The set of literals built from X, $Lit(X)$ is $X \cup \{\neg x | x \in X\}$, and the set of rules built from X, written as $Rul(X)$, is defined by $2^{Lit(X)} \times Lit(X)$, the set of pairs of a set of literals built from X and a literal built from X. A rule is written as $\{l_1, \ldots, l_n\} \rightarrow l$; we also write $l_1 \wedge \ldots \wedge l_n \rightarrow l$, and when $n = 0$, we write $\top \rightarrow l$. Moreover, for $x \in X$, we write $\sim x$ for $\neg x$ and $\sim \neg x$ for x.

The mental attitudes attributed to agents consist of beliefs B, desires D, goals G, and undercutters H. An undercutter is a mental attitude expressing the absence of a desire or goal, which we use in the following section to model permissions as exceptions [30]. A mental description function MD associates a rule in $Rul(X)$ with each belief, desire, goal, and undercutter. We introduce priority relations to resolve conflicts among mental attitudes. A function $\geq$ associates with an agent a transitive and reflexive relation on the powerset of the motivations and beliefs containing at least the subset relation. Moreover, various mental attitudes are attributed to agents by the agent description relation AD. It associates with each agent a set of beliefs, desires, goals, and undercutters.

Multiagent systems also contain concepts concerning informational aspects. First of all, the set of variables whose truth value is determined by an agent (decision variables) are distinguished from those which are not directly determined by the agent (P, the parameters using the terminology of Lang *et al.* [31]). Concerning the relations among these concepts, we have that parameters P are a subset of the propositional variables X. The complement of P represents the decision variables controlled by the agents. Hence, we associate to each agent a subset of $X \setminus P$ by extending the agent description AD.

Definition 2 (MAS): A multiagent system (MAS) is a tuple $\langle A, X, B, D, G, H, AD, MD, \geq \rangle$, where:

1) The agents A, propositional variables X, beliefs B, desires D, goals G, and undercutters H are six finite disjoint sets. We write $M = D \cup G$ for motivations.
2) An agent description $AD : A \rightarrow 2^{X \cup B \cup D \cup G \cup H}$ is a complete function that maps each agent to a set of variables (its decision variables), and to its beliefs, desires, goals, and undercutters. For each agent $a \in A$, we write X_a for $X \cap AD(a)$, B_a for $B \cap AD(a)$, etc. We write also $P = X \setminus \cup_{a \in A} X_a$ for the parameters.
3) A mental description $MD : B \cup D \cup G \cup H \rightarrow Rul(X)$ is a complete function from the sets of beliefs, desires, goals, and undercutters to the set of rules built from X. For $S \subseteq B \cup D \cup G \cup H$, we write also $MD(S) = \{MD(s) | s \in S\}$. Moreover, we write $s\, x \rightarrow y$ for denoting: s such that $MD(s) = x \rightarrow y$.
4) A priority relation $\geq$: $A \rightarrow ((2^B \times 2^B) \cup (2^M \times 2^M))$ is a function from agents to a transitive and reflexive relation on the powerset of the mental attitudes containing at least the subset relation. We write $\geq_a$ for $\geq (a)$.

Example 1 illustrates the running example as a multiagent system. In conceptual models used in practice as well as in the more detailed examples in the following sections, we use meaningful names; here, we use single letters to save space.

Example 1: Let $MAS = \langle A, X, B, D, G, H, AD, MD, \geq \rangle$ with $A = \{\mathbf{a}, \mathbf{p}, \mathbf{n}\}$, $P = \{q, r\}$, $H = \emptyset$, and $X \setminus P$, B, D, G, AD, MD, and $\geq$ are given by the following table:

	a		**p**		**n**	
X	x_1, x_2		x_3, x_4		x_5	
B	b_1	$x_1 \rightarrow q$	b_2	$x_1 \rightarrow q$		
B	b_3	$x_3 \rightarrow r$	b_4	$x_3 \rightarrow r$		
B	b_5	$x_4 \rightarrow \neg q$	b_6	$x_4 \rightarrow \neg q$		
B	b_7	$x_2 \rightarrow \neg r$	b_8	$x_2 \rightarrow \neg r$		
D	d_1	$\top \rightarrow q$	d_2	$\top \rightarrow r$		
D	d_3	$\top \rightarrow \neg x_4$	d_4	$\top \rightarrow \neg x_2$		
D			d_5	$\top \rightarrow \neg x_5$		
G	g_1	$x_3 \rightarrow x_2$	g_2	$x_1 \rightarrow x_4$	g_3	$x_1 \wedge x_4 \rightarrow x_5$
$\geq$	$g_1 > d_3 > d_1$		$g_2 > d_5 > d_4 > d_2$		$g_3 > g_1 = g_2$	
$\geq$	$b_5 > b_1$		$b_6 > b_2$			
$\geq$	$b_7 > b_3$		$b_8 > b_4$			

Agent $\mathbf{a}$ desires (d_1) to get information q to increase its knowledge by doing x_1 (b_1), i.e., by making a request to agent $\mathbf{p}$. Vice versa, agent $\mathbf{p}$ desires (d_2) to know r by requesting it to agent $\mathbf{a}$ with action x_3. Therefore, both agents $\mathbf{a}$ and $\mathbf{p}$ share resources and request them. Moreover, they want to react to the request of the other one by respectively doing actions x_2 and x_4 (g_1 and g_2), which are not desired by the other agent (d_3 and d_4). These actions are exceptions to the effect of, respectively, requests x_1 and x_3, and thus, they prevent their results (q and r). Agent $\mathbf{n}$ is monitoring the behavior of agent $\mathbf{p}$ and, if agent $\mathbf{p}$ achieves its goal x_4 when x_1 is true, then it wants (g_3) to perform action x_5, which is disliked by agent $\mathbf{p}$ (d_5).

Finally, only a fragment of the priority relation is given, because it is only given for singleton motivations and beliefs, whereas it is defined over sets of motivations.

The example already illustrates a drawback of using only multiagent systems to describe the example, because there is no notion of obligation, violation, or sanction. We therefore introduce normative multiagent systems.

B. Normative Multiagent Systems

A normative multiagent system contains a norm (violation) description V, a function from agents and literals to the decision variables of the normative system together with the parameters. We write $V_a(x)$ for the decision variable representing that there is a violation $x \in Lit(X_a \cup P)$ by agent $a \in A$.

Definition 3: A normative multiagent system $NMAS$ is a tuple $\langle A, X, B, D, G, H, AD, MD, \geq, V\rangle$ that extends a multiagent system with a partial (violation) function $V : A \times Lit(X) \rightarrow X \setminus P$ from agents and literals to decision variables. We write $V_a(x)$ for $V(a, x)$.

We can introduce decision variables like $V_b(V_a(x))$, where agent b is considered as a violator if it considers $x \in Lit(X)$ as a violation done by agent a. Analogously, $V_b(\neg V_a(x))$ means that agent b is considered as a violator, because agent b does not consider x as a violation done by agent a.

Example 2 (Continued): The agents $\mathbf{a}$, $\mathbf{p}$, $\mathbf{n}$ in Example 1 are now interpreted as a virtual community where $\mathbf{a}$ is considered only a user of the system, and $\mathbf{p}$ only as a local provider owning a document, and $\mathbf{n}$ is a global authority. Assume that V is defined as: $V_{\mathbf{a}}(x_1) = x_4$, $V_{\mathbf{p}}(x_4) = V_{\mathbf{p}}(V_{\mathbf{a}}(q)) = x_5$, and $V_c(y) = undefined$ for all other values of $c \in A$ and $y \in Lit(X)$. This means that if x_4 is the case, then x_1 is recognized as a violation of agent $\mathbf{a}$, and if x_5 is the case, then doing $V_{\mathbf{a}}(x_1)$ is recognized as a violation of agent $\mathbf{p}$.

C. Obligation

The definition of obligation contains several clauses. The first clause says that the obligation is in the desires and in the goals of a normative system b ("your wish is my command"). The second and third clauses can be read as "the absence of $\sim x$ is considered as a violation." The association of obligations with violations is inspired by Anderson's reduction of deontic logic to alethic modal logic [32]. The third clause says that the normative system desires that there are no violations. The fourth and fifth clauses relate violations to sanctions and assume that normative system b is motivated to apply sanctions only as long as there is a violation; otherwise, the norm would have no effect. Finally, for the same reason, we assume in the last clause that the agent does not like the sanction.

Definition 4 (Obligation): Let a normative multiagent system $NMAS$ be $\langle A, X, B, D, G, H, AD, MD, \geq, V\rangle$. Agent $a \in A$ is obliged in $NMAS$ to decide to do $x \in Lit(X_a \cup P)$ with sanction $s \in Lit(X_b \cup P)$ if $Y \subseteq Lit(X)$ by normative system $b \in A$, written as $NMAS \models O_{a,b}(x, s|Y)$, if and only if the following conditions are met.

1) $Y \rightarrow x \in MD(D_b) \cap MD(G_b)$: If normative system b believes Y, then it desires x and has x as a goal.
2) $Y \cup \{\sim x\} \rightarrow V_a(\sim x) \in MD(D_b) \cap MD(G_b)$: If normative system b believes Y and $\sim x$, then it has the goal and the desire $V_a(\sim x)$: to recognize $\sim x$ as a violation by agent a.
3) $\top \rightarrow \neg V_a(\sim x) \in MD(D_b)$: Normative system b desires that there are no violations.
4) $Y \cup \{V_a(\sim x)\} \rightarrow s \in MD(D_b) \cap MD(G_b)$: if normative system b believes Y and decides $V_a(\sim x)$, then it desires and has as a goal that it sanctions agent a with s.
5) $Y \rightarrow \sim s \in MD(D_b)$: If normative system b believes Y, then it desires not to sanction, $\sim s$. The normative system only sanctions in case of violation.
6) $Y \rightarrow \sim s \in MD(D_a)$: If agent a believes Y, then it desires $\sim s$, which expresses that it does not like to be sanctioned.

VII. Games With Global Policies

We first introduce documents and permissions, then we define global policies, and finally, we introduce the extended game theory.

A. Documents

We introduce some syntactic sugar. Management of knowledge in the multiagent system is represented by access of documents DC. We are inspired by Lee [33]: "we use the term 'document' since most information parcels in business practice are mapped on paper documents." Of course, knowledge is distinct from information. In the context of knowledge management, knowledge is described as information that has a use or purpose. Whereas information can be placed onto a computer, knowledge is emergent and socially constructed, in the sense that it exists in the heads of people: Knowledge is information to which an intent has been attached. The distinction between the two can be reflected by the internal structure of the documents, and how agents update their mental states once they receive documents. For example, receiving information only updates the agent's beliefs, whereas receiving knowledge can extend the agent's capabilities. In this paper, we do not consider the internal structure of knowledge documents, but we focus on operations agents can perform on documents.

Definition 5 (Documents): Let DC be a set of documents. Let $DA_a = \{f(a, d)|\ d \in DC\}$ be a set of actions of the agent $a \in A$. $f(a, d) \in DA_a$ can belong to the decision variables X_a or it can be a parameter such that there exists a decision variable $x \in X_a$ such that $x \rightarrow f(a, d) \in B_a$: $f(a, d)$ is an effect of a decision variable x of agent a representing agent a's beliefs that x is a fallible tentative of accessing document d.

The following example illustrates that similar syntactic sugar is also used for the description of the state of the world.

Example 3 (Continued): Possible actions on the documents are the create, read, update, and delete (CRUD) actions of the CRUD security model in databases: e.g., $x_1 = read(\mathbf{a}, d)$ and $q = info(\mathbf{a})$. To have an information $info(\mathbf{a})$, agent $\mathbf{a}$ believes it has to perform action $read(\mathbf{a}, d)$ (b_1) as an attempt to get it.

$NMAS \models O_{\mathbf{ap}}(\neg read(\mathbf{a}, d), san|\top)$, agent $\mathbf{a}$ is obliged not to read document $d \in DC$ $(read(\mathbf{a}, d) \in X_{\mathbf{a}})$, or else it is sanctioned with $san \in Lit(X_{\mathbf{p}} \cup P)$, if

$$
\begin{array}{lll}
g_4 & d_6 & \top \rightarrow \neg read(\mathbf{a}, d) \in G_{\mathbf{p}}, D_{\mathbf{p}} \\
g_2 & d_7 & read(\mathbf{a}, d) \rightarrow V_{\mathbf{a}}(read(\mathbf{a}, d)) \in G_{\mathbf{p}}, D_{\mathbf{p}} \\
 & d_8 & \top \rightarrow \neg V_{\mathbf{a}}(read(\mathbf{a}, d)) \in D_{\mathbf{p}} \\
g_5 & d_9 & V_{\mathbf{a}}(read(\mathbf{a}, d)) \rightarrow san \in G_{\mathbf{p}}, D_{\mathbf{p}} \\
 & d_{10} & \top \rightarrow \neg san \in D_{\mathbf{p}} \\
 & d_3 & \top \rightarrow \neg san \in D_{\mathbf{a}}
\end{array}
$$

B. Conditional Permission

To model permission, we consider clauses analogous to the ones for obligation. Since most clauses have to do with sanctions, we have to consider only clauses analogous to the first two clauses of obligation. Whereas obligation corresponds to a kind of normative goal in our model, permission corresponds

to a kind of undercutter to a normative goal. Moreover, whereas obligations imply, under some conditions, a violation, permissions may imply the absence of a violation.

Definition 6 (Permission): Let a normative multiagent system $NMAS$ be $\langle A, X, B, D, G, H, AD, MD, \geq, V\rangle$. Agent $a \in A$ is permitted to decide to do $x \in Lit(X_a \cup P)$ if $Y \subseteq Lit(X)$ in $NMAS$ by normative system $b \in A$, written as $NMAS \models P_{a,b}(x|Y)$, if and only if

1) $Y \rightarrow x \in MD(H_b)$: If normative system b believes Y, then it does not have a desire or goal x.
2) $Y \cup \{x\} \rightarrow \neg V_a(x) \in MD(D_b) \cap MD(G_b)$: If normative system b believes Y and x, then it does not want to count x as a violation.

C. Policies

Consider now the notion of entitlement: An agent is obliged to permit another agent. If we would add nested obligations and permissions as in standard modal logic, then an obligation to permit $O_{b,c}(P_{a,b}(x|Y), s \mid W)$ could be represented by $O_{b,c}(\Psi, s|W)$, where Ψ is, respectively, $Y \rightarrow x \in MD(H_b)$ and $Y \wedge x \rightarrow \neg V_a(x) \in MD(G_b)$. Note that in the latter formula, as well as the other formulas in the subsection, we write conjunction $\wedge$ for the union $\cup$ to facilitate the reading of these formulas. This reduces to 12 clauses, since each of the two obligations is made of the following six clauses:

1) $W \rightarrow \Psi \in MD(D_c) \cap MD(G_c)$;
2) $W \wedge \neg\Psi \rightarrow V_b(\neg\Psi) \in MD(D_c) \cap MD(G_c)$;
3) $\top \rightarrow \neg V_b(\neg\Psi) \in MD(D_c)$;
4) $W \wedge V_b(\neg\Psi) \rightarrow s \in MD(D_c) \cap MD(G_c)$;
5) $W \rightarrow \sim s \in MD(D_c)$;
6) $W \rightarrow \sim s \in MD(D_b)$.

Our reduction, instead, claims that, for global policies, a nested modality is too weak, and we therefore remove the inner modalities. For example, the second clause reduces to $O_{b,c}(\neg V_a(x), s|Y \wedge W \wedge x)$. This removal, however, does not have to lead to something making sense. For example, for the first clause $O_{b,c}(x, s|Y \wedge W)$ is too strong, as x is not obligatory, but only permitted. Thus, we define the policy concerned with obligation to permit as only the second clause.

Definition 7: A global policy of global authority $c \in A$ in context $W \subseteq Lit(X)$ with sanction $s \in Lit(X_c \cup P)$ for the entitlement of agent $a \in A$ to $x \in Lit(X_a \cup P)$ from knowledge provider $b \in A$ if $Y \subseteq Lit(X)$ is

$$O_{b,c}(\neg V_a(x), s|Y \wedge W \wedge x).$$

Consequently, the global policy implies the following six clauses, which illustrate that removal of nested obligations and permissions has left a nested violation predicate:

1) $W \wedge Y \wedge x \rightarrow \neg V_a(x) \in MD(D_c) \cap MD(G_c)$;
2) $W \wedge Y \wedge x \wedge V_a(x) \rightarrow V_b(V_a(x)) \in MD(D_c) \cap MD(G_c)$;
3) $\top \rightarrow \neg V_b(V_a(x)) \in MD(D_c)$;
4) $W \wedge Y \wedge V_b(V_a(x)) \rightarrow s \in MD(D_c) \cap MD(G_c)$;
5) $W \wedge Y \rightarrow \sim s \in MD(D_c)$;
6) $W \wedge Y \rightarrow \sim s \in MD(D_b)$.

Analogously, policies concerned with an obligation to oblige contain clauses like $O_{b,c}(V_a(\sim x), s|Y \wedge W \wedge \sim x)$, a permission to permit contains $P_{b,c}(\neg V_a(x)|Y \wedge W \wedge x)$, and a permission to oblige $P_{b,c}(V_a(\sim x)|Y \wedge W \wedge \sim x)$. Moreover, they may also contain additional clauses. For example, if the global authority explicitly states the sanctions for the local providers, then we may have several clauses such as $O_{b,c}(s', s|Y \wedge W \wedge \sim x \wedge V_a(\sim x))$ for the obligation to oblige and $P_{b,c}(s'|Y \wedge W \wedge \sim x \wedge V_a(\sim x))$ for the permission to oblige.

D. Games

The agents believe to be in a state that is the result of the application of the prioritized belief rules.

Definition 8 (Consequences of Beliefs): Let R be a set of rules in $Rul(X)$, Q a set of literals in $Lit(X)$, and $\geq$ a transitive and reflexive relation on the powerset of R containing at least the superset relation.

1) $out(R, Q) = \cup_0^\infty out^i(R, Q)$ is the state obtained by the sequence $out^0(R, Q) = \emptyset$ and $out^{i+1}(R, Q) = out^i(R, Q) \cup \{l | L \rightarrow l \in R \text{ and } L \subseteq Q \cup out^i(R, Q)\}$;
2) $maxfamily(R, Q, \geq)$ is the set of maximal subsets R' of R with respect to set inclusion such that $Q \cup out(R', Q)$ is consistent (does not contain a literal and its negation);
3) $preffamily(R, Q, \geq)$ is the set of maximal elements of $maxfamily$ with respect to the $\geq$ ordering;
4) $outfamily(R, Q, \geq)$ is the output under the elements of $preffamily$, $\{out(R', Q) | R' \in preffamily(R, Q, \geq)\}$;
5) $x \in out(R, Q, \geq)$ iff $x \in \cap outfamily(R, Q, \geq)$.

Decisions of agents are consistent sets of literals built from decision variables.

Definition 9 (Decisions): The set of decisions Δ of a set of agents $AS = \{a_1, \ldots, a_n\} \subseteq A$ is the set of consistent sets $\delta = \bigcup_{a_i \in AS} \delta_i \subseteq Lit(X)$.

To define the optimal decisions, we consider the expected effects of decisions by applying belief rules and by recursively modeling the decisions of other agents. These effects are used to order the decisions using the desire and goal rules. The unfulfilled motivations of decision δ according to agent $a \in A$ are the set of motivations whose body is part of the closure of the decision under the belief rules but whose head is not, and which are not undercut by any rule in H.

Definition 10 (Recursive Modeling): Given the set of agents $AS = \{a_1, \ldots, a_n\} \subseteq A$:

1) Given $S_i = out(B_i, \bigcup_{0<j<n} \delta_j, \geq_i) \cup \bigcup_{0<j<n} \delta_j$, $U(\delta, a_i)$ is the set of $l_1 \wedge \ldots \wedge l_n \rightarrow l \in M$ such that:
   1) $\{l_1, \ldots, l_n\} \subseteq S_i$ and $l \notin S_i$;
   2) there does not exist $l_1 \wedge \ldots \wedge l_n \rightarrow l \in H_i$ such that $\{l_1, \ldots, l_n\} \subseteq S_i$.
2) A decision δ is optimal for agent a_i if and only if it is optimal for agents $a_{i+1}, \ldots, a_n$ and there is no decision δ_i' such that for all decisions $\delta' = \delta_0 \cup \ldots \cup \delta_i' \cup \ldots \cup \delta_n'$ and $\delta'' = \delta_0 \cup \ldots \cup \delta_i \cup \ldots \cup \delta_n''$ optimal for agents $a_{i+1}, \ldots, a_n$, we have that $U(\delta', a_i) >_i U(\delta'', a_i)$.

Example 4 (Continued): When $\delta = \{x_1, x_4\}$, then $out(B_{\mathbf{a}}, \delta, \geq_{\mathbf{a}}) = \{x_1, x_4, q\}$ and $U(\delta, \mathbf{a}) = \{d_3\ \top \rightarrow \neg x_4\}$, $U(\delta, \mathbf{p}) = \{d_2\ \top \rightarrow r\}$, and $U(\delta, \mathbf{n}) = \{g_3\ x_1 \wedge x_4 \rightarrow x_5\}$. When $\delta = \{x_1, x_3\}$, then $out(B_{\mathbf{a}}, \delta, \geq_{\mathbf{a}}) = \{x_1, x_3, q, r\}$ and $U(\delta, \mathbf{a}) = \{g_1\ x_3 \rightarrow x_2\}$, $U(\delta, \mathbf{p}) = \{g_2\ x_1 \rightarrow x_4\}$, $U(\delta, \mathbf{n}) = \emptyset$.

TABLE I
USER $\mathbf{a}$, KNOWLEDGE PROVIDER $\mathbf{p}$, AND GLOBAL AUTHORITY $\mathbf{n}$

	$\mathbf{a}$	$\mathbf{p}$	$\mathbf{n}$
X	$read(\mathbf{a}, d)$	$V_{\mathbf{a}}(read(\mathbf{a}, d)), san$	$V_{\mathbf{p}}(V_{\mathbf{a}}(read(\mathbf{a}, d)))$
B	$b_1\ read(\mathbf{a}, d) \rightarrow info(\mathbf{a})$	$b_4\ read(\mathbf{a}, d) \rightarrow info(\mathbf{a})$	
B	$b_2\ san \rightarrow \neg info(\mathbf{a})$	$b_5\ san \rightarrow \neg info(\mathbf{a})$	
B	$b_3\ \top \rightarrow project(\mathbf{a})$	$b_6 \top \rightarrow project(\mathbf{a})$	$b_7 \top \rightarrow project(\mathbf{a})$
$D\,G$	$d_1\ \top \rightarrow info(\mathbf{a})$	$d_3\ \top \rightarrow \neg read(\mathbf{a}, d)$	$d_{10}\ project(\mathbf{a}) \wedge read(\mathbf{a}, d) \rightarrow \neg V_{\mathbf{a}}(read(\mathbf{a}, d))$
$D\,G$	$d_2\ \top \rightarrow \neg san$	$d_4\ read(\mathbf{a}, d) \rightarrow V_{\mathbf{a}}(read(\mathbf{a}, d))$	$d_{11} V_{\mathbf{a}}(read(\mathbf{a}, d)) \rightarrow V_{\mathbf{p}}(V_{\mathbf{a}}(read(\mathbf{a}, d)))$
$D\,G$		$d_5\ V_{\mathbf{a}}(read(\mathbf{a}, d)) \rightarrow san$	$d_{12} \top \rightarrow \neg V_{\mathbf{p}}(V_{\mathbf{a}}(read(\mathbf{a}, d)))$
$D\,G$		$d_6\ \top \rightarrow \neg V_{\mathbf{p}}(V_{\mathbf{a}}(read(\mathbf{a}, d)))$	
$D\,G$		$d_7\ hack(\mathbf{a}) \rightarrow \neg info(\mathbf{a})$	
$D\,G$		$d_8\ \top \rightarrow \neg V_{\mathbf{a}}(read(\mathbf{a}, d))$	
$D\,G$		$d_9\ \top \rightarrow \neg san$	
$\geq$	$b_2 > b_1$	$b_4 > b_5$	
$\geq$	$d_2 > d_1$	$d_4 > d_8, d_5 > d_9, d_7 > d_6$	
δ	$read(\mathbf{a}, d)$		
out	$read(\mathbf{a}, d), info(\mathbf{a}), project(\mathbf{a})$	$read(\mathbf{a}, d), info(\mathbf{a}), project(\mathbf{a})$	$read(\mathbf{a}, d), project(\mathbf{a})$
U		$\top \rightarrow \neg read(\mathbf{a}, d)$	
U		$read(\mathbf{a}, d) \rightarrow V_{\mathbf{a}}(read(\mathbf{a}, d))$	

VIII. EXAMPLE

Table I illustrates an example of a virtual community composed of a user $\mathbf{a}$, a knowledge provider $\mathbf{p}$, and a CAS agent $\mathbf{n}$, which establishes the global policy of the community. When agent $\mathbf{a}$ takes its decision $\delta_{\mathbf{a}}$, it has to minimize its unfulfilled motivational attitudes. However, when it considers these attitudes, it must not only consider its decision $\delta_{\mathbf{a}}$ and its consequences, it must consider also the decision $\delta_{\mathbf{p}}$ of the local provider $\mathbf{p}$ and its consequences, for example, that it is sanctioned by it. Therefore, agent $\mathbf{a}$ recursively considers which decision the normative system $\mathbf{p}$ will take depending on its different decisions $\delta_{\mathbf{a}}$. In turn, agent $\mathbf{p}$ takes its decision $\delta_{\mathbf{p}}$ under the light of what will do the authority $\mathbf{n}$ it is subject to. The optimal decision of agent $\mathbf{p}$ depends on each decision $\delta_{\mathbf{a}}$ and also on the optimal decision $\delta_{\mathbf{n}}$ for normative system $\mathbf{n}$ for each decision $\delta_{\mathbf{a}} \cup \delta_{\mathbf{p}}$. Instead, given a decision $\delta_{\mathbf{a}} \cup \delta_{\mathbf{p}}$, a decision $\delta_{\mathbf{n}}$ is optimal for agent $\mathbf{n}$ if it minimizes the unfulfilled motivational attitudes in $D_{\mathbf{n}}$ and $G_{\mathbf{n}}$ according to the $\geq_{\mathbf{n}}$ relation, without further recursive modeling.

Agent $\mathbf{a}$ reads a document d ($read(\mathbf{a}, d) \in X_{\mathbf{a}}$), which is under the control of agent $\mathbf{p}$ to get an information $info(\mathbf{a}) \in P$ ($read(\mathbf{a}, d) \rightarrow info(\mathbf{a}) \in B_{\mathbf{a}}$). Moreover, The CAS $\mathbf{n}$ issues global policies addressed to normative system $\mathbf{p}$ and $\mathbf{p}$ issues local policies addressed to user $\mathbf{a}$.

The example shows a case of entitlement. Even if the provider $\mathbf{p}$ locally forbids access to the information, agent $\mathbf{a}$ is entitled to do so by the global policy in the context of some project ($project(\mathbf{a}) \in P$). The local policy is represented by $O_{\mathbf{a},\mathbf{p}}(\neg read(\mathbf{a}, d), san|\top)$, while the global one is $O_{\mathbf{p},\mathbf{n}}(\neg V_{\mathbf{p}}(read(\mathbf{a}, d))|project(\mathbf{a}) \wedge read(\mathbf{a}, d))$. The sanction *san* makes false the effect $info(\mathbf{a})$ of the access to the resource d by $read(\mathbf{a}, d)$. However, if agent $\mathbf{p}$ believes that agent $\mathbf{a}$ is a hacker ($hack(\mathbf{a}) \in P$), it prefers to violate the global policy with respect to letting agent $\mathbf{a}$ access the document.

The bottom three lines of Table I represent the output and unfulfilled motivations when agent $\mathbf{a}$ decides to do $\delta_{\mathbf{a}} = \{read(\mathbf{a}, d)\}$ and agents $\mathbf{p}$ and $\mathbf{n}$ decide to do nothing ($\delta_{\mathbf{p}} = \delta_{\mathbf{n}} = \emptyset$). Had agent $\mathbf{p}$'s decision been $\delta'_{\mathbf{p}} = \{V_{\mathbf{a}}(read(\mathbf{a}, d)), san\}$, then $info(\mathbf{a})$ would not have been true anymore in $out(B_{\mathbf{a}}, \delta, \geq_{\mathbf{a}})$, due to the sanction applied by agent $\mathbf{p}$: the rule $san \rightarrow \neg info(\mathbf{a}) \in B_{\mathbf{a}}$ has priority over the rule $read(\mathbf{a}, d) \rightarrow info(\mathbf{a}) \in B_{\mathbf{a}}$ ($b_2 > b_1$). Thus, agent $\mathbf{a}$'s unfulfilled desires would have been

$$U(\delta' = \delta_{\mathbf{a}} \cup \delta'_{\mathbf{p}}, \mathbf{a}) = \{\top \rightarrow info(\mathbf{a}), \top \rightarrow \neg san\}.$$

To take a decision between $\delta_{\mathbf{p}}$ and $\delta'_{\mathbf{p}}$, agent $\mathbf{p}$ compares which of its goals and desires remain unsatisfied under the light of agent $\mathbf{n}$'s reaction: In fact, if agent $\mathbf{p}$ decides for $\delta'_{\mathbf{p}}$, $U\delta'_{\mathbf{n}}$ would be $\{V_{\mathbf{p}}(V_{\mathbf{a}}(read(\mathbf{a}, d)))\}$.

$$U(\delta', \mathbf{n}) = \{\top \rightarrow \neg V_{\mathbf{p}}(V_{\mathbf{a}}(read(\mathbf{a}, d)))\}$$

However, $U(\delta', \mathbf{p}) \geq_{\mathbf{p}} U(\delta, \mathbf{p})$ and thus agent $\mathbf{p}$ decides $\delta_{\mathbf{p}}$.

$$U(\delta', \mathbf{p}) = \{\top \rightarrow \neg V_{\mathbf{p}}(V_{\mathbf{a}}(read(\mathbf{a}, d)))\}$$
$$U(\delta, \mathbf{p}) = \{read(\mathbf{a}, d) \rightarrow V_{\mathbf{a}}(read(\mathbf{a}, d))\}$$

If agent $\mathbf{p}$ believes that agent $\mathbf{a}$ is a hacker, $\top \rightarrow hack(\mathbf{a}) \in B'_{\mathbf{p}}$, i.e., if $hack(\mathbf{a})$ is true in $out(B'_{\mathbf{p}}, \delta, \geq_{\mathbf{p}})$, then $\mathbf{p}$ believes that agent $\mathbf{a}$, by knowing the content of the document ($info(\mathbf{a})$), will damage the system, so it prefers that it does not access it ($hack(\mathbf{a}) \rightarrow \neg info(\mathbf{a}) \in D_{\mathbf{p}} \cap G_{\mathbf{p}}$). In this case, agent $\mathbf{p}$ prefers to be considered a violator by agent $\mathbf{n}$ with respect to allowing agent $\mathbf{p}$ trying to access the document, since its priority is

$$hack(\mathbf{a}) \rightarrow \neg info(\mathbf{a}) \geq_{\mathbf{p}} \top \rightarrow \neg V_{\mathbf{p}}(V_{\mathbf{a}}(read(\mathbf{a}, d))).$$

Thus, decision $\delta''_{\mathbf{p}}$ would be $\{V_{\mathbf{a}}(read(\mathbf{a}, d)), san\}$.

The example can be extended with sanctions toward agent $\mathbf{p}$ in the obvious way. Moreover, obligations to oblige or to prohibit can be defined analogously. The number of clauses increases, as discussed in Section VII-C, but the analysis can be made in the same way.

IX. Related Work

The problem of striking some balance between local versus global policies has been studied in other contexts too, as part of the development of distributed computing systems. Numerous solutions have been proposed, including that of security in a grid computing environment [14] as well as in web-services security architectures [34]. An advantage of the formalization of the balance introduced in this paper is that it is characterized using an abstract framework. Another advantage is that normative multiagent systems offer the formal framework to analyze or simulate complex scenarios, using the underlying theories of normative systems and multiagent systems, including ones of fraud and deception [35].

In particular, the problem has been addressed in the context of trust-management systems, trust networks, and in reputation-based systems [36]. These theories and systems originate from distinct but related traditions (see, e.g., [37] for a discussion of how trust and norms may be related in a game-theoretic setting). However, they also have some remarkable similarities. We believe that further study is needed to investigate the assumptions of the two approaches. For example, the way in which agents anticipate the behavior of other agents in normative systems seems distinct from the way they anticipate behavior in trust and reputation-based systems.

Moreover, while here we study the decentralization of control, our framework is used also to cope with the symmetric issue of how to centralize those aspects of security in distributed environments that cannot be dealt with at the local level, such as which are the members of the community and which ones should be authorized to access a resource. In particular, in [38], we explore the problem of how local providers can delegate to other agents the power to authorize access without giving up their autonomy. We argue that the problem can be solved by means of counts-as relations [39]: An agent empowered to authorize can issue declarations that are considered as authorizations by the local provider. Moreover, these authorizations appear in the conditions of permissions, which act as exceptions to the prohibitions concerning general users. In this way, on the one hand, a CAS agent can regulate at the central level which agents are allowed to access resources on the basis of the up-to-date list of members and policies; on the other hand, the local providers are not overburdened by the management of information about memberships and maintain their autonomy to issue obligations, prohibitions, and permissions to regulate access to their resources.

We believe that a further study into the similarities and distinctions between these areas can lead to some fruitful exchanges of ideas. A contribution of the normative systems' framework is the logical analysis methods, the interaction among the logical framework and the game-theoretic elements, and the possibilities to simulate fraud and deception. Also, trust and reputation mechanisms can be used to enrich existing normative multiagent systems. For example, if one agent believes that another one is going to harm resources under its control, this information should be propagated to the global level. There should be some form of mechanism to capture such "reputation" information in the normative architecture.

Normative systems have traditionally been concerned with more static systems such as bureaucratic and legal systems. Recent work on normative multiagent systems is concerned with applications in modern electronic networks, which are much more dynamic and uncertain. It therefore takes its inspiration from social theories such as Searle's work on the construction of social reality, and is concerned with the powers of agents to change the normative system [19].

The formal framework presented in this paper builds on deontic logic and BOID agent architecture. It extends our previous work on normative multiagent systems, e.g., in [19], in various ways. For example, it introduces the notion of prioritized beliefs, documents, and most importantly, it studies more complex games involving an arbitrary number of agents, within a realistic setting. Whereas arbitrary games among a large number of agents get very complex, in particular in the context of uncertainty and observations, and introduce new conceptual problems, see, e.g., [40] for an example, the games discussed in this paper are relatively simple due to the fact that each agent has to consider only the decision of the next agent who can consider it as a violator.

We do not define permissions as the absence of obligation, so-called negative permission, but as exceptions to obligations, a kind of positive permission. For a discussion on the issues involved in modeling permission, see [41]. Permission is simpler than obligation, since permissions cannot lead to violations and sanctions. It is only due to entitlement that knowledge providers may be sanctioned when they do not permit a user to access documents, but the user itself cannot be a violator and be sanctioned due to its permissions to access a document. The various clauses of obligation have been motivated by a game-theoretic analysis [19]. The first clause ensures that "respectful" agents internalizing the goals of the normative system will fulfill their obligation under typical circumstances; the second and third clauses do so for "respectful" agents that do not want to be considered as violators even if they do not internalize the norm as one of their goals. The other clauses do so for "selfish" types of agents, which care only about not being sanctioned. Similar games can be played to show that the clauses of permission are necessary, again for norm internalizing agents and other types of agents, respectively. Since permissions are exceptions to obligations, a "respectful" internalizing agent would still adopt the content of the obligation as one of its goals. For this reason, we added the first clause containing an undercutter to the goal of the obligation the permission is an exception of.

Here, we only discuss regulative norms and we do not discuss constitutive ones, though the new models introduced in this paper can also be extended with constitutive norms along the lines discussed in [19]. For constitutive rules, we adopt the same strategy of attributing mental attitudes to normative systems. Whereas regulative norms are defined in terms of goals of the normative systems, constitutive norms establishing what counts as institutional facts are defined in terms of the beliefs of the normative system.

X. Summary

Any scalable solution for secure knowledge management has to be able to distribute not only the knowledge-management system, but also the security system. Once knowledge management becomes distributed, security becomes distributed too. A distributed security system assumes that security concerns are incorporated from the first design of the distributed knowledge-management system. For example, Kolp [7] analyzes the development of knowledge-management systems and argues that "existing proprietary information management tools block the exchange of information between applications and users," and therefore "a holistic approach is required in the design of information technology infrastructures as well as the actual business process reengineering for a successful knowledge management effort."

We develop an integrated approach to distributed knowledge management and security, using multiagent systems and access-control policies in virtual communities. Distributed access control is realized by a virtual community of multiple knowledge providers with their own local access-control system to their documents and local policies, and by global community policies regulating these local policies. The issue at stake is the rational balance of global versus local control in virtual communities. Knowledge providers must be autonomous, but not unconstrained.

Autonomy of local knowledge providers is ensured by modeling each member of the virtual community as a normative system interacting with other members and posing prohibitions and permissions about access to its knowledge. Participants do not give up their autonomy to prohibit access to knowledge to users they do not trust, even when the users satisfy the security rules of the virtual community. Resource providers are therefore not forced, but motivated by the global authority to behave as required. Local providers consider whether to violate global policies in some situations. Games among the three agents involved in knowledge access—the user, the knowledge provider, and the global authority—explain how these motivations work. The games can be used also to analyze the knowledge-management system, or to simulate it.

Global community policies regulate local policies based on von Wright's notion of transmission of will. The rules of policies for secure knowledge management do not concern only what knowledge the users are prohibited or permitted to access, but they also concern which regulations the knowledge providers are allowed or obliged to enforce. This is a challenge, because rules refer usually to the actions of users and not to other rules, which are obligatory or permitted to adopt. In our model, we incorporate a mechanism for interaction among normative systems. We generalize the framework with decentralized systems composed of global authorities and local providers; they are both normative systems, but global authorities can specify duties and permissions of local providers concerning the local policies.

A game-theoretic analysis is used to define global policies. These games describe interactions among the agents in the normative multiagent systems. Agents must evaluate the effects of accessing documents both when making and evaluating an access request. In particular, since the compliance of the providers to the policies cannot be taken for granted, a user must consider whether they will fulfill their commitments or not. To make a prediction about the behavior of a local provider, it is necessary to consider also the reaction of the global authority enforcing control by monitoring and sanctioning violations.

For example, the notion of entitlement, i.e., when a knowledge provider is obliged to permit access to a user, could be interpreted as the obligation of the provider to create a permission. However, the knowledge provider can create the permission, but still does not let the user access the document. In such a case, the knowledge provider does not act according to the permissions it created itself. This scenario shows that such definition of global policies is ineffective. The game-theoretic analysis of the notion of entitlement in normative multiagent systems illustrates that this definition of global policy is ineffective and may not change the behavior of the providers.

Our solution is to define global policies concerning the behavior of knowledge providers: considering specified behaviors as violations and sanctioning them. In the case of entitlement, this amounts to being obliged not to consider the specified behavior as a violation. For the case in which users are motivated by sanctions only, in the sense that they are not respectful and act according to the norm simply due to the existence of the norm, our game-theoretic analysis shows that the existence of nested obligations and permissions is not only too weak, it is even superfluous. The only clause that is important is that knowledge providers are sanctioned in the case that the user is not acting as desired.

Similar scenarios occur when the global policies dictate that the knowledge provider should oblige the user. In that case, the knowledge provider can simply not create the obligation, or it can create the obligation but not enforce it. Again von Wright's analysis of the transmission of will can be used to argue that creating the obligation is not sufficient, the knowledge provider should also enforce the obligations it created, and sanction users that do not act according to them.

There are more requirements of secure knowledge-management systems, which can be addressed in further research. For example, our model should be extended to cope with role-based access control using our role model [42] or exploring the delegation issues of discretionary access control using our contract model [19]. Finally, the current framework can be extended to deal not only with policies consisting of regulative rules like obligations, prohibitions, and permissions, but also with constitutive rules specifying counts-as relations and institutional facts [39]. In particular, in [38], we use counts-as relations to specify local policies concerning authorizations, while global policies prescribing constitutive rules are still an open issue.

References

[1] J. Bradshaw, S. Dutfield, P. Benoit, and J. Woolley, "KAoS: Towards an industrial strength generic agent architecture," in *Software Agents*. Cambridge, MA: MIT Press, 1997, pp. 375–418.

[2] A. Preece, K. Hui, W. A. Gray, P. Marti, T. J. M. Bench-Capon, D. M. Jones, and Z. Cui, "The KRAFT architecture for knowledge fusion and transformation," *Knowl. Based Syst.*, vol. 13, no. 2/3, pp. 113–120, 2000.

[3] A. Abecker, A. Bernardi, and L. van Elst, "Agent technology for distributed organizational memories," in *Proc. 5th Int. Conf. Enterprise Information Syst*, 2003, pp. 3–10.
[4] B. Yu, M. Venkatraman, and M. Singh, "An adaptive social network for information access: Theoretical and experimental results," *J. Appl. Artif. Intell.*, vol. 17, no. 1, pp. 21–38, Jan. 2003.
[5] L. van Elst, V. Dignum, and A. Abecker, "Towards agent-mediated knowledge management," in *Proc. Agent-Mediated Knowledge Management*, vol. 2926, *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 2004, pp. 1–30.
[6] R. M. van Eijk, F. S. de Boer, W. van der Hoek, and J.-J. C. Meyer, "Open multi-agent systems: Agent communication and integration," in *Proc. Intelligent Agents VI*, vol. 1757, *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 2000, pp. 218–232.
[7] M. Kolp, "Agent-based IT support for knowledge management," *IAG Working Paper 29/02*, 2002.
[8] S. Xu and W. Zhang, "PBKM: A secure knowledge management framework," in *Proc. NSF/NSA/AFRL Workshop Secure Knowledge Management*, 2004.
[9] B. Sadighi Firozabadi and M. Sergot, "Contractual access control," in *Proc. Workshop Security Protocols*, Cambridge, U.K., 2002, pp. 96–102.
[10] M. Bonifacio, P. Bouquet, G. Mameli, and M. Nori, "KEx: A peer-to-peer solution for distributed knowledge management," in *Proc. 4th Int. Conf. PAKM*, vol. 2569, *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 2002, pp. 490–500.
[11] M. Bonifacio, R. Cuel, G. Mameli, and M. Nori, "A peer-to-peer architecture for distributed knowledge management," in *Proc. 3rd Int. Symp. Multi-Agent Systems, Large Complex Systems, and E-Business (MALCEB'2002)*.
[12] C. Tacla and J.-P. Barthes, "A multi-agent architecture for knowledge management systems," in *Proc. 2nd IEEE ISADS*, 2002, pp. 1–12.
[13] D. Mundy and D. Chadwick, "Secure knowledge management," in *Creating Knowledge Based Health Care Organizations*. Hershey, PA: The Idea Group, 2004, pp. 321–337.
[14] L. Pearlman, V. Welch, I. Foster, C. Kesselman, and S. Tuecke, "A community authorization service for group collaboration," in *Proc. IEEE Int. Workshop Policies Distributed Systems and Networks*, 2002, pp. 50–59.
[15] B. Sadighi Firozabadi, M. Sergot, and O. Bandmann, "Using authority certificates to create management structures," in *Proc. Workshop Security Protocols*, vol. 2467, *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 2001, pp. 134–145.
[16] M. Blaze, J. Feigenbaum, and J. Lacy, "Decentralized trust management," in *Proc. IEEE Conf. Security and Privacy*, 1996, pp. 164–173.
[17] J. S. Sichman, R. Conte, C. Castelfranchi, Y. Demazeau, "A social reasoning mechanism based on dependence networks," in *Proc. 11th ECAI*, A. G. Cohen, Ed., 1991, pp. 188–192.
[18] A. Jones and J. Carmo, "Deontic logic and contrary-to-duties," in *Handbook of Philosophical Logic*, D. Gabbay and F. Guenthner, Eds. Dordrecht, The Netherlands: Kluwer, 2001, pp. 203–279.
[19] G. Boella and L. van der Torre, "A game theoretic approach to contracts in multiagent systems," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 36, no. 1, pp. 68–79, Jan. 2006.
[20] V. Kartseva, J. Gordijn, and Y.-H. Tan, "Analysing preventative and detective control mechanisms in international trade using value modelling" *in Proc. 6th ACM Int. Conf. Electronic Commerce*, Eds., M. Janssen, H. G. Sol, and R. W. Wagenaaar, Delft, The Netherlands, 2004, pp.51–58.
[21] Y. Shoham and M. Tennenholtz, "On the emergence of social conventions: Modeling, analysis and simulations," *Artif. Intell.*, vol. 94, no. 1/2, pp. 139–166, Jul. 1997.
[22] G. Boella and L. Lesmo, "A game theoretic approach to norms," *Cogn. Sci. Q.*, vol. 2, no. 3/4, pp. 492–512, 2002.
[23] E. Goffman, *Strategic Interaction*. Oxford, U.K.: Blackwell, 1970.
[24] A. Newell, "The knowledge level," *Artif. Intell.*, vol. 18, no. 1, pp. 87–127, Jan. 1982.
[25] J. Broersen, M. Dastani, J. Hulstijn, and L. van der Torre, "Goal generation in the BOID architecture," *Cogn. Sci. Q.*, vol. 2, no. 3/4, pp. 428–447, 2002.
[26] P. J. Gmytrasiewicz and E. H. Durfee, "Formalization of recursive modeling," in *Proc. ICMAS*, 1995, pp. 125–132.
[27] G. H. von Wright, "An essay in deontic logic and the general theory of action," *Acta Philos. Fenn.*, vol. 21, 1968.
[28] M. S. Sloman, "Policy driven management of distributed systems," *J. Netw. Syst. Manag.*, vol. 2, no. 4, pp. 333–360, Dec. 1994.
[29] D. Dennett, *The Intentional Stance*. Cambridge, MA: MIT Press, 1987.
[30] J. L. Pollock, "Defeasible reasoning," *Cogn. Sci.*, vol. 11, no. 4, pp. 481–518, Oct.–Dec. 1987.
[31] J. Lang, L. van der Torre, and E. Weydert, "Utilitarian desires," *Auton. Agents Multiagent Syst.*, vol. 5, no. 3, pp. 329–363, 2002.
[32] A. Anderson, "The reduction from deontic logic to alethic modal logic," *Mind*, vol. 67, pp. 100–103, 1958.
[33] R. Lee, "Documentary Petri nets: A modeling representation for electronic trade procedures," in *Proc. Business Process Management*, 2000, vol. 1806, pp. 359–375.
[34] S. Chang, Q. Chen, and M. Hsu, "Managing security policy in a large distributed web services environment," in *Proc. Int. COMPSAC., Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 2003, pp. 610–618.
[35] G. Boella and L. van der Torre, "Normative multiagent systems and trust," in *Proc. Trust Agent Societies Workshop AAMAS*, vol. 3577, *Lecture Notes in Artificial Intelligence*. Berlin, Germany: Springer-Verlag, 2004, pp. 1–17.
[36] M. Fan, Y. Tan, and A. Whinston, "Evaluation and design of online cooperative feedback mechanisms for reputation management," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 2, pp. 244–254, Feb. 2005.
[37] M. Hollis, *Trust Within Reason*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
[38] G. Boella and L. van der Torre, "Permission and authorization in policies for virtual communities of agents," in *Proc. Agents and P2P Computing Workshop AAMAS*, vol. 3601, *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 2004, pp. 86–97.
[39] J. Searle, *The Construction of Social Reality*. New York: Free Press, 1995.
[40] G. Boella and L. van der Torre, "Rational norm creation: Attributing mental attitudes to normative systems, part 2," in *Proc. ICAIL*, 2003, pp. 81–82.
[41] D. Makinson and L. van der Torre, "Permissions from an input-output perspective," *J. Philos. Logic*, vol. 32, no. 4, pp. 391–416, Aug. 2003.
[42] G. Boella and L. van der Torre, "Organizations as socially constructed agents in the agent oriented paradigm," in *Proc. ESAW*, vol. 3451, *Lecture Notes in Artificial Intelligence*. Berlin, Germany: Springer-Verlag, 2004, pp. 1–13.

Guido Boella received the Ph.D. degree in computer science from the Università di Torino, Torino, Italy, in 2000.

He is currently a Professor at the Department of Computer Science, Università di Torino. His research interests include multiagent systems, in particular, normative systems, institutions, and roles using qualitative decision theory. He organized the first workshops on normative multiagent systems (NorMAS), on coordination and organization, and the American Association for Artificial Intelligence (AAAI) Fall Symposium on roles.

Leendert van der Torre received the Ph.D. degree in computer science from Erasmus University Rotterdam, Rotterdam, The Netherlands, in 1997.

He is currently a Full Professor at the University of Luxembourg, Luxembourg. He has developed the so-called input/output logics and the BOID agent architecture. His current research interests include deontic logic, qualitative game theory, and coordination and security in normative multiagent systems (NorMAS).