Goal-Based Explainable Security Certificate Requests

Maaike Harbers[1], Joost Broekens[1], Thomas Quillinan[2], M. Birna van Riemsdijk[1] and Niek Wijngaards[2]

[1]Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands

[2]Thales Research and Technology, Delftechpark 24, 2628 XH Delft, The Netherlands, contact: j.broekens@tudelft.nl

Abstract

In crisis situations it is important that crisis response workers can quickly get access to the right information for the tasks they are required to undertake. A distinction can be made between getting the right information and having the rights to get that information. The first is an information filtering and relevance problem, the second is a security issue related to access control. In this paper we focus on the second issue. It is impossible to predefine access rules for all players in a crisis situation that ensure that they have access only to the information they need. Therefore the key is to have a system that is flexible and timely (efficient) with respect to the decision to grant access, without a major burden on humans having to make these decisions, and without inadvertant leakages of sensitive information. We believe for crisis management it is more important to be able to hold individuals and organizations accountable for their use of information than to overly restrict access to information. We propose goal-based explainable security certificate requests as a solution to this problem.

1 Introduction

In a crisis response situation, many organizations have to coordinate their plans and actions. This requires that the actors that represent these organizations, which

can be individuals or software agents, are able to share relevant information with others, including with those who usually would not have access to such information (De Bruijn & Wijngaards 2012). Two examples are a policeman that needs to know the location of a fire squad, and a policeman that wants information about the goods in a storage facility of a chemical plant to decide if the area has to be secured or not.

There are several challenges related to the information exchange in crisis situations (Goal et al 2004). A key characteristic of a crisis situation is that it disrupts normal operations. As such, crises also disrupt normal information flows. As each crisis is unique, it is not possible to predict the information flow in a crisis beforehand. Therefore, information should be shared and distributed in a *flexible* way. Another challenge in crisis situations is that crisis workers need to operate effectively and efficiently in order to save lives and reduce damage. Therefore, information should be shared and distributed in a *timely* manner with *minimal human involvement*. Finally, the information being shared is often classified. For example, the stock information of a chemical plant is normally not available. However, stock information is of great importance when dealing with a fire in the chemical plant, e.g. to anticipate explosions. We believe that classified, but relevant information should be provided under such exceptional conditions, and that this can best be achieved by making individuals and organizations *accountable* for their use of this information.

We propose a goal-based solution to the problem of flexible rights management. The solution involves goal hierarchies with the possible goals and subgoals of crisis workers, and organization hierarchies indicating the roles and hierarchies in organizations involved in crisis management. When crisis workers request information, they indicate the particular goal for which they need that information. Certificates can automatically be generated by software agents when the indicated goal is a subgoal of a goal for which a certificate has already been accredited, and the requesting worker is a subordinate of the creator as defined by the organization hierarchy. In those cases where the certificate cannot be automatically generated, the responsible person is asked. Our approach is inspired by approaches involving task hierarchies such as (Lesser et al 2004), and in particular by goal-based explanation based on hierarchical information structures (Harbers et al 2010; Core et al 2006).

The approach enables flexible access rights management in a way that it can be semi-automized limiting the burden on humans to make decisions. Moreover, all grants for information access are auditable. The goal hierarchies can be integrated with organization hierarchies and cross organizational trust in such a way that it does not force organizations to share a common goal ontology. Although the system, in principle, enables users to fake goals, the combination of enforced auditability of who accessed what for what reason (goal), the organization hierarchy and trust provide sufficient security for accountability.

The outline of this paper is as follows. In the next section we will discuss the requirements that an information management system should fulfill. In the third section, we will discuss our goal-based approach to flexible information access rights management. Due to space limitations we have to omit the implementation

details of our simulation of the approach. We end the paper with a discussion and suggestions for future research.

2 Requirements on an information management system

In the introduction we argued that an information sharing mechanism used in crisis situations should be flexible, timely, accountable and involve minimal human effort. We will discuss each of these requirements in more detail.

An information sharing mechanisms must be *flexible* with respect to who can access what information and for what period of time this access is necessary. Such mechanisms need to be able to adapt information disclosure rules to fit the crisis at hand, and allow humans to override these rules when needed.

Crises require *timely* delivery of information. Information sharing mechanisms need to rapidly perform one of three actions: (1) retrieve the information requested by an individual; or (2) deny access rights, or (3) quickly resolve a request to get access rights.

The information sharing mechanism should yield *minimal human involvement.* Crisis management workers need to cope with a large amount of information, manage stress and make decisive decisions where necessary. Use of an information sharing mechanism must be targeted towards reducing the cognitive load needed for the sharing of information and the mechanism itself must be easy to use. This means that the mechanism must include support for the semi-automatic resolving of access requests.

The shared information in a crisis situation should be *accountable*. We believe that in times of crisis it is more important to share information than to restrict information access due to pre-existing security policies; information safety is important, but the safety of people is paramount (Massacci 2010). However, a mechanism needs to be in place to safeguard against abuse of access rights. A way to do so is to ensure that information access can always be accounted for, in such a way that it can be explained why someone requested information. An information sharing mechanism that is accountable will allow organizations to make classified information available under certain conditions.

3 Explainable security certificate requests and generation

In this section we introduce our solution to cope with changes in information sharing during crisis situations. It is based on the concept that for each new information source a requester requires access to that it does not already have, the requester indicates why it needs this information, i.e., the requester gives the goal he/she is working towards for which the information is necessary. Goals are organized in a so-called goal hierarchy *h*; indicating how main and subgoals are de-

pendent upon each other. So, each goal *g* has a possible parent g_p and possible children $g_{c1..n}$. For example, extinguishing a fire involves investigating the fuel that nourishes the fire (e.g., oil, sodium, wood) getting the appropriate extinguisher (water, powder), and actually extinguishing the fire. So, any *extinguish fire* goal has at least three subgoals: *investigate fuel*, *locate extinguisher*, and *use extinguisher* (see Figure 1).

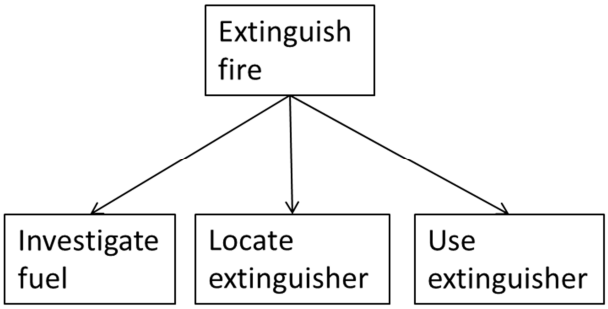

Figure 1: Part of a firefighter's goal hierarchy.

We propose to use these hierarchies to order the different goals each organization has. Each organization *o* (itself a hierarchy of workers *w*) has its own goal hierarchy, so $h_{o1..m}$ hierarchies exist. Typically such goal hierarchies will follow roles and organization hierarchies but will be more specific than roles (because goals and subgoals are more task specific than roles). A practical way to couple goals to individual workers is to attach goals to roles, so, each role *r* has associated with it a set of goals $g_{r1..i}$. Each individual worker *w* has a set of roles R_w, and, thus, a set of goals within one organization, i.e., the superset GR_w of all goals $g_{r1..i}$ for all *r* in $\underline{R_w}$.

In the following two subsections we describe the uses of the goal hierarchy and organization hierarchy to manage information access rights in a crisis situation. The first subsection discusses how goal hierarchies are linked to information and information requests. In the subsection thereafter we detail how information access requests can be automatically and manually generated using goals in a goal hierarchy, and workers in an organization.

3.1 Motivated security certificate requests

The information flow in crises can be described as sources that produce messages on different topics and workers that want that information. If a worker *w* needs information from a source *s* on a topic *t*, then that worker *w* issues a request from *s* for messages about topic *t*. It tags this request with a goal g from its personal goal list GR_w. A request thus consist of the following tuple of information (w, g, s, t), i.e., a specific worker *w* asks for a reason *g* from source *s* the access to information posted by *s* on topic *t*.

In practice, many of the active goals and roles of a worker can be predicted based on the activities of that person, or could be detected from the context. The worker only has to check that the request for information has the right goal attached. This means that it is rare that a worker actually has to manually fill-in the current goal, unless the worker is doing something out of the ordinary.

The information source *s* in a request (w, g, s, t) can be a software agent or a human. If source s is a human, he or she would receive a request for access to a particular topic. Because the request includes a reason, it is now easier to decide to grant access or not. For example, if a journalist would ask the fire brigade for access to the estimated endangered area due to a forest fire, the squad leader will refuse. However, if the reason is that the journalist happens to be a civilian who is already in danger and the restricted area is the only path to safety, the squad leader will grant the request. Tagging information requests with the reason why is useful as a quick heuristic to decide whether or not to give access to information.

3.2 Automated generation of security certificates

We now explain in more detail how information requests can be dealt with in relation to the generation and distribution of security certificates. First we assume a *valid* certificate is always needed to access information. So, any request for information needs a valid certificate, with *valid* defined below. Second, we assume that the organization hierarchy *o* is used to resolve requests for certificates (w, g, s, t), such that a worker *w* in organization o_w always asks its parent p_w in o_w for a valid certificate in case *w* does not have one, unless $w = p_w$ in which case *w* asks the top of organization o_s, with *s* being the information source. Third, validity of certificates is contextualized, by which we mean that a certificate is only valid (i.e. can be used to retrieve information) for a particular context uniquely defined by the tuple (w, g, s, t), potentially enriched with other relevant information such as the crisis level *l*. A certificate $c_1 = (w_1, g_1, s_1, t_1)$ is greater than $c_2 = (w_2, g_2, s_2, t_2)$ if and only if g_1 is a (possibly recursive) subgoal of g_2 and $w_1 = p_{w2}$ or $w_1 = w_2$. If a certificate c_1 is greater than c_2, the holder of the certificate can create c_2. The creator's ID w_c is added to the certificate, resulting in what we define as a *valid* certificate (w, g, s, t, w_c) for accessing information from *s* for topic *t* by worker *w* for goal *g* as issued (created) by w_c.

These assumptions enable the following. First, software agents that represent superiors holding certificates for higher level goals can automatically (without involvement of the superior) generate smaller certificates. This facilitates flexible and timely information access and limits human overhead. Any information request (the generation of certificates) as well as the actual information retrieval are fully auditable in a manner that allows explanation of why information was needed by whom and by whom the certificate was granted. Further, as the certificate is contextualized upon goal (i.e., reason of use) and as information can only be retrieved using valid certificates, as soon as a worker is not working on a particular goal, information cannot be given anymore, unless the worker lies about its goal. However, this lying is traceable, and thus the worker or his/her superior can be

hold accountable for this afterwards. Goals can serve as justifications for requests *after* the crisis. For example, if an organization revealed information about its security system to the police during a crisis, it probably wants to check which information has been provided to whom for what reasons. The goals that accompanied exchanged information can be used to justify the information exchange.

4 Discussion and future work

We have presented an approach for flexible information access right management for crises. The approach is based on goal-based motivated information requests and proposes a method for automated security certificate generation. We anticipate that this promotes accountable, flexible and timely delivery of information during crises with minimal human involvement. Our next step is to show these benefits experimentally. Besides this, we anticipate two other benefits of our approach: information filtering and explanation of the need for pushed information to workers. We will briefly address these two benefits now.

This paper discusses information requests from workers to information sources, which we call *information pull*. Due to the time pressure in crisis situations, however, it can be beneficial for workers to receive information without asking for it, i.e. to receive an information push. It is important that the *information push* only contains relevant information. When the worker receives more information than he/she is able to process, the worker will start ignoring the information and in that way will miss important messages.

A way to create an automatic information push with relevant information is to annotate information with goals. A worker *w* has a certain role *r* in the organization *o* it is part of. A role *r* is associated with it a set of goals $g_{r1..i}$, so it is known which goals a certain worker is trying to achieve. Now when information is annotated with goals, these goals can help to filter the information that should be sent to a worker, using either a static (pre-configured) or adaptive (machine learning) approach.

In addition to delivering the information, workers need to know why this information is relevant to them. The same goals can be used to explain this relevance. For example, consider a gas leak situation. If information about a change in wind direction is pushed to a police officer in the area, it is helpful for this officer to know that this information is relevant for investigating the presence of civilians in the newly affected area. Otherwise, it would merely be information about the weather, and the burden of inferring what to do with it would be upon the officer or the information source.

Acknowledgments

This work is part of the Slim Verbinden project, that is sponsored by Programma Innovatie voor Maatschappelijke Veiligheid by Agentschap.nl, project no. IMV1100038. The authors thank their colleagues in the project and the experts from public safety and security organizations.

References

M. Core, H. Lane, M. van Lent, D. Gomboc, S. Solomon, and M. Rosenberg, 2006. Building explainable artificial intelligence systems. AAAI.

P. de Bruijn and N. Wijngaards, 2012. Agent-enabled information provisioning while retaining control: A demonstration. In Proceedings of the Gi4DM Conference, Berlin, 2012, Springer.

S. Goel, S. Belardo, and L. Iwan, 2004. A resilient network that can operate under duress: To support communication between government agencies during crisis situations. In Proceedings of the 37th Hawaii International Conference on System Sciences (HICSS'04), p. 50123.1, Washington, DC, USA, 2004. IEEE Computer Society.

M. Harbers, J. Broekens, K. v.d. Bosch, and J.-J. Meyer, 2010. Guidelines for Developing Explainable Cognitive Models, pp. 85-90.

V. Lesser, K. Decker, N. Carver, A. Garvey, D. Neiman, M. Nagendra Prasad, and T. Wagner, 2004. Evolution of the GPGP/TAEMS domain-independent coordination framework. Autonomous agents and multi-agent systems, pp. 9:87-143.

F. Massacci, 2010. Infringo ergo sum: when will software engineering support infringements? In Proceedings of the FSE/SDP workshop on Future of software engineering research, FoSER '10, pp. 233-238, New York, NY, USA. ACM.