Delivery of secure health care across clinical centres

L Xiao, M Croitoru, P Lewis

School of Electronics and Computer Science
University of Southampton
Southampton SO17 1BJ

Keywords: data security; decision support; distributed networks

Abstract

Assisting medical diagnosis has been one of the main goals of software engineering and artificial intelligence. Access to patient data is representing itself as a challenge due to the nature of patient's healthcare records. This requires data to be shared between a distributed set of hospitals without recourse to a centralised system, necessitating an agreed protocol. This paper describes the architecture of the HealthAgents distributed decision support system, the types of resources being used, and their associated access principles. To evaluate the system, the conformance to the UK Data Protection Act 1998 is clearly illustrated.

Introduction

Assisting medical diagnosis[1] has been one of the main goals of software engineering and artificial intelligence. The approach of using them together as a decision support system (DSS) is of particular interest. A successful medical DSS would aim to improve the healthcare outcomes required by an individual clinician. The process of designing such a system requires:

- consideration of the clinician's needs;
- access to data and processes that may be geographically distributed; and
- interaction with other healthcare professionals[2].

Access to patient data, for example, is complicated further by the patient's healthcare records. This requires data to be shared between a distributed set of hospitals without recourse to a centralised system, an agreed protocol. The protocol is employed as part of the distributed decision support system (d-DSS) which increases the autonomous interaction between the medical centres. This interaction requires data integrity based on trust worthy and secure technology[3].

HealthAgents overview and link-anonymised data scheme for preserving privacy

Brain tumours are still an important cause of morbidity and mortality in Europe[4]. There is a need to improve brain tumour classification, and to provide non-invasive methods for brain tumour diagnosis and prognosis, to aid patient management and treatment.

The HealthAgents project[5], funded by the EU's Sixth Framework Programme, aims to build the world's largest distributed data warehouse of brain tumour cases data. The multidisciplinary collaboration involves seven educational and research institutions, two SMEs, as well as some subcontractor hospitals and external expertise groups spanned over Belgium, Italy, Spain, and the United Kingdom. HealthAgents inherits the achievements of its predecessor INTERPRET[6] and is related to the ongoing eTUMOUR[7] project. It plans to create

a multi-agent distributed decision support system (d-DSS) based on novel medical imaging and laboratory tests to help determine the diagnosis and prognosis of brain tumours. Furthermore, the rarity of many brain tumour types requires that information must be sought from many hospitals.

Prior to incorporation into clinical practice new methods must be fully tested within a clinical trials setting. Such trials are subject not only to data protection laws but also regulations governing clinical trials including ethical approval and informed consent of the participants. Allowing for flexibility within the data security model is therefore essential.

Clinical trials commonly use data from which personal information (eg name, address, date of birth) is removed but to which a unique patient identifier is added, often termed link-anonymised data. Such a scheme has the advantage of having a high chance of preserving patient anonymity whilst allowing data from the same patient to be added at a later date. Full patient records are kept for clinical purposes within the treating hospital and with the patient's permission may be used to generate and periodically update the clinical trials data.

While complete patient records may be accessed only by hospitals and local nodes, link anonymised records may be exchanged between a limited numbers of centres producing classifiers. Furthermore, only limited amounts of data which can be considered as totally anonymised may be accessed outside the closed project network. A model shown in Figure 1 illustrates such a data protection model in a multi-layered fashion.

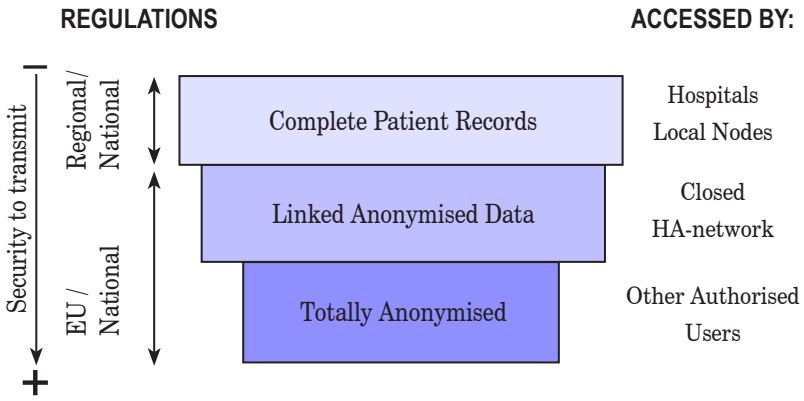

Figure 1. Prototype secure data protection model for HealthAgents

Clinical trials are usually supported by a centralised database where the link-anonymised data is stored. This allows the patients to be reassured that their data will be afforded a high level of security and allows regulatory bodies ease of access to inspect the processes in place. For a distributed system, similarly robust arrangements must be designed to reassure ethics committees and patients that the data is secure. Each data collecting centre could have an associated link-anonymised database as approved by their appropriate ethics committee. Patient identifiers could then be kept along with the clinical patient record in the treating hospital. These databases need be the only databases kept within the system giving a truly distributed data-warehouse. The limited data required for analysis could then be subject to stringent anonymisation processes and sent to a small number of specific sites for processing, for example the production of classifiers. In this way, the distributed nature of the system could be preserved whilst allowing appropriate regulatory access to data repositories. Security systems will need to be in place which can allow each centre to potentially limit the type of data transmitted and the locations it is transmitted to.

HealthAgents key components and architecture

Figure 2 shows a prototype version of the HealthAgents d-DSS. Each clinical node as part of the inter-networked system can be either the user where requests for classification of a given case are delivered, or the producers where classifiers are created or retrained based on pattern recognition techniques, or both. When a clinical user requests the classification of a case that resides internally, its associated GUI Agent will retrieve the patient data from the local hospital database via a Database Agent, local data access policies being applicable. Alternatively, if the case under classification resides externally, then the GUI Agent will contact the local Yellow Pages Agent, which in turn will contact an external Yellow Pages Agent through which patient data is retrieved via the Database Agent of that hospital, external data access policies being applicable. One Yellow Pages Agent resides in each hospital's local node. They synchronise with each other and together maintain a directory of available nodes, agents, as well as the classifiers for the entire HealthAgents network. Their knowledge of the availability and location of resources is useful for answering queries sent from the GUI Agent. Global resource and service access policies will apply when cross-centre resource access is requested by an agent and global services, such as the query service provided by the Yellow Pages Agent, are requested.

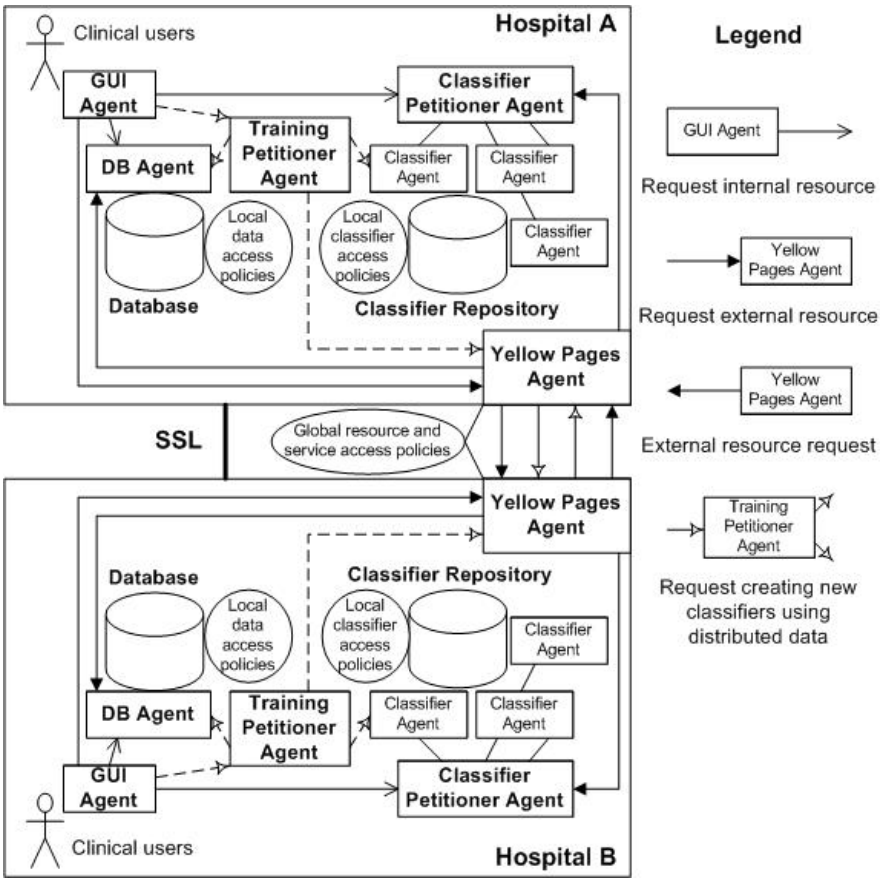

Figure 2. The distributed architecture of the HealthAgents system and its resource access flow control.

Once the case has been loaded into the GUI application, it may be classified. The local Yellow Pages Agent has registered in it classifiers that can discriminate among tumour classes, including descriptions about their capabilities, reputation, and the training data upon which they have been produced. The Yellow Pages Agent looks up its local registry, as well as con-

tacts external Yellow Pages Agents, and compiles a list of appropriate classifiers. This list is returned to the clinician and the clinician can now send the selected classifiers which can solve questions, accompanied by the patient data that these classifiers can operate upon, to the Classifier Petitioner Agent. The Classifier Petitioner Agent will invoke each Classifier Agent associated with the classifiers in the list, supplying patient data. Internal or external classifier access policies will apply, depending upon the location of classifiers. While this may involve remote classifier access which gives the system a sense of full distribution, in practice, once a classifier is produced a copy might be obtained by every node in the network for local classifier running and better performance.

In the classification process, patient data stored in clinical nodes is not directly used but rather converted in certain formats by specialised pre-processing nodes (omitted in the diagram) before classifiers can be actually executed upon the data. After the execution of classifiers, classification results are collected by the Classifier Petitioner Agent from multiple classifiers and ranked using statistical data and finally sent back to the clinician. The clinician can now do the diagnosis, supported by the answers and recommendations provided by the system. Eventually, when the diagnosis is finished, the clinician evaluates the classification result produced by the selected classifiers and their reputation updated. The above scenario assumes that classifiers exist to solve questions. If no such classifier exists, a clinician requests the Training Petitioner Agent to create one using data from distributed sites and register the new classifier in the Yellow Pages Agent for later use.

Conformance to legal and ethical obligations via the existing infrastructure

The UK Data Protection Act 1998 came into force in 2000. It regulates the processing of data of individuals, including the obtaining, holding, use or disclosure of such information. The eight data protection principles are as follows:

1. Personal data shall be processed fairly and lawfully (and under certain conditions).
2. Personal data shall be obtained only for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes.
3. Personal data shall be adequate, relevant and not excessive in relation to the purpose or purposes for which they are processed.
4. Personal data shall be accurate and, where necessary, kept up to date.
5. Personal data processed for any purpose or purposes shall not be kept for longer than is necessary for that purpose or those purposes.
6. Personal data shall be processed in accordance with the rights of data subjects under this Act.
7. Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data.
8. Personal data shall not be transferred to a country or territory outside the European Economic Area unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data.

We have described the architecture of our d-DSS, the types of resources being used by it, and their associated access principles in previous sub-sections. The conformance to the UK Data Protection Act 1998 can be now briefly illustrated as follows.

In the d-DSS, patient case records are only processed for either the diagnosis of that particular patient or for training classifiers, fairly and lawfully. This is in compliance with Principle 1. The publicity of a case and its direct access is strictly controlled by the node where

the case is stored inside the HealthAgents network and the routine use of such data is replaced by classifiers which are trained upon the data by classifier training software. Thus, cases will not be processed in any manner in contradiction with the specified and lawful purposes of improving disease diagnosis as agreed by the patients and their exposure is minimised. This is in compliance with Principle 2. The adequacy, relevance, non-excessiveness, accuracy, and up-to-date status of cases are maintained by clinical centres and wherever possible, link-anonymised data is used for the preservation of patient privacy. This is in compliance with Principles 3 and 4. All cases used for the purpose of training classifiers will be discarded when classifiers are produced and will not be kept for longer than it is necessary. This is in compliance with Principle 5. Patients retain the rights of withdrawing their cases and if requested they will be removed from the databases immediately (via the unique patient identifier being added to the link-anonymised data). This is in compliance with Principle 6. Each clinical centre enforces the described case access principles and so unauthorised or unlawful processing of personal data or damage to data will be avoided. This is in compliance with Principle 7. The HealthAgents project is building a network inside the EU boundary and may allow data transfer outside its network only if it is in a fully anonymised form and protected at an adequate level as being agreed upon. This is in compliance with Principle 8.

Conclusions

The unique security issues involved in healthcare domains have been discussed in this paper. The practical solution of these security issues have been addressed to the needs of the HealthAgents project. We believe a sustainable security solution should be provided in accordance with the ethical regulations for healthcare data, as well as fulfil the security requirements usually raised by distributed decision support systems due to the nature of clinical settings.

Acknowledgements

This work is supported under the HealthAgents and OpenKnowledge STREP projects funded by EU Framework 6 under Grants: IST-FP6-027214 and IST-FP6-027253.

References

1. Szolovits P, Pauker SG. Categorical and Probabilistic Reasoning in Medicine Revisited. Artificial Intelligence, 1993, 59 :167-180.
2. Coiera E. Question the Assumptions. In: Barahona P, Christensen JP (Eds). Knowledge and Decisions in Health Telematics. IOS Press, 1994, :61-66.
3. Keese J, Motzo L. Pro-active approach to malware for healthcare information and imaging systems. International Congress Series. Elsevier, 2005, 1281 :943-947.
4. Bray F, Sankila R, Ferlay J, Parkin DM. Estimates of cancer incidence and mortality in Europe in 1995. European Journal of Cancer, 2002, 38(1) :99-166.
5. HealthAgents Available at: www.healthagents.net/
6. INTERPRET Available at: http://azizu.uab.es/INTERPRET/
7. eTUMOUR Available at : www.etumour.net/
8. Xiao L et al. An Adaptive Security Model for Multi-agent Systems and Application to a Clinical Trials Environment. Proc 31st IEEE Annual International Computer Software and Applications Conference (COMPSAC'07). IEEE Computer Society, 2007, :61-266.