Security on medical data sharing
(a literature review)

Dayana Spagnuelo, Gabriele Lenzini
University of Luxembourg
Email: {dayana.spagnuelo, gabriele.lenzini} @uni.lu

Keywords—literature review, security, medical data, electronic health records.

I. INTRODUCTION

Medical records (e.g., test results and health reports) are about patients. Hospitals and healthcare institutions generate them after a patient’s visit. Today they are digitized, stored electronically, and accessed remotely by professionals.

European directives suggest that patients should access these records too. Besides, they say, patients should have control over these data and be informed if and when their records are shared and how secure they are [1]. These requirements are hard to be met.

From a patient’s perspective, the viewpoint of this paper, it may be easier to address at least one of such requirements: to inform patients about how secure their data are. This is a property usually referred as transparency, but a clear meaning of the word is still missing. According to [2] transparency ought to be regarded as an additional feature that qualifies security. So, security can be said to be transparent when is intelligible to human user. It opposes an opaque security, which holds technically but without the user’s being aware of it. Thus, transparency is a socio-technical security property.

Transparency, is not a new term. It has been proposed in relation to Transparency Enhancing Tools (TETs) [3]. These are usually browser extensions that read out web server’s privacy policies and inform users concisely, for instance, that a web server records the user’s whereabouts and may sell the user’s data to third parties. TETs have been discussed in relation to electronic health records [2], but no concrete solution has been proposed. Transparency in the medical domain is still an unfulfilled requirement.

Contribution. We survey the literature in medical data sharing and discusses what are the main security concerns in it. We intend also to figure out whether transparency is debated in that domain, in relation to which other properties, and which meaning and role are given to it.

II. METHODOLOGY AND TOOLS

We browsed the state of the art by searching for papers via “Findit.Lu”.¹ This is the largest library portal in Luxembourg, and it is entirely dedicated on searching for electronic contents. It indexes a large number of important scientific digital libraries such as, among many others, LNCS, the ACM Digital Library, IEEExplore, ScienceDirect, Scopus, and Medline.

We queried for “Security” and “Medical Data Sharing”, and we looked for papers containing them in the title, in the abstract, in the list of keywords, and in the entire body. We chose “Security” because it is a general term: we expect that a paper that addresses more precise security properties will also mention “security” somewhere its text. We chose “Medical Data Sharing” to refine our domain to papers that discuss sharing medical data.

First, we queried without constraints on the year of publication. We got as many as 526 articles, too many for us to be able to read or scan them all. Thus, we restricted the focus to the last ten years, from 2004 to now. Excluding the repeated results and the papers not available for download, our pool shrank down to a total of 75 papers. We read the abstract and skimmed through the content of all of them. It turned out that 20 papers were about medical data sharing but with no focus on “security”: the word appeared to be mentioned but the concept is not discussed. We discarded those papers and, after this skimming, we were left with a pool of 55 papers.

We organized our findings around one question: “what particular security property the paper is about?”. To answer this question helped us to classify the papers depending on the property, or properties, they debate. It also helps us to understand whether transparency is considered as a security requirement and, if it is, in relation to which other property.

III. MAJOR FINDINGS

Answering our main question, and so looking into what security properties our pool of papers is about, lead us to identify eight main security categories, each concerning policies, tools, or techniques meant to guarantee, preserve, or enforce a specific property. The 8 categories are the following: Privacy, concerning to provide anonymity to the data owner or to empower her to define who can operate on the data; User authentication, concerning to enhance the way in which users are authenticated electronically; Access control, concerning better ways to define who can access medical data and in what circumstances; Data authenticity, concerning solutions to guarantee and prove that the data have not been manipulated or tampered with; Confidentiality, concerning to prevent the disclosure of data content to non-authorized third parts; Auditability, concerning to help the data owner to retrieve information clarifying how her data is being used; Transparency, concerning to guarantee openness about security policies and processes.

Most of the surveyed papers argue about data confidentiality (see Figure 1). This property is invoked in relation

¹The portal is accessible via www.bibnet.lu, or directly at, www.findit.lu
to protect the data transmitted in open channels, such as the internet, or stored in open data bases, such as the cloud. One comment is mandatory: in the pool “confidentiality” there are 27 papers, namely [4]–[30]. Some of those were, per keywords, first gathered under “privacy”. A closer look revealed that they are using the term inappropriately since their concern is mainly about encrypting data. But, encryption per se is insufficient to guarantee that the user’s personal and sensitive information remains private during the whole data life cycle; more sophisticated techniques have to be in place for privacy to be protected. Thus, we decided to re-classify those works as being about confidentiality, adding those up to the ones already in that category.

Confidentiality is constantly discussed together with data integrity and data authenticity. That is because encryption is the technique that is more often adopted to enforce confidentiality in medical systems and the same technique is also proposed for data authenticity and integrity. In a total of 16 papers about data integrity (i.e., [5], [6], [8]–[10], [12], [13], [16], [18], [22], [23], [25], [29], [31]–[33]) only three works do not discuss confidentiality. We observed a very similar scenario with the category data authenticity. Only three works do not discuss confidentiality, out of 9 papers discussing data authenticity (i.e., [8], [9], [12], [22], [25], [29], [31]–[33]). Also, all works that examine data authenticity discuss data integrity too.

After confidentiality, the second and third most discussed security properties are privacy and access control. We found out that 20 works discuss privacy (the correct interpretation of this term) [14], [20], [25], [26], [30], [33]–[47], and that 19 papers discuss access control [11], [13], [19], [22], [23], [25], [29], [34], [37], [41]–[43], [45], [48]–[53].

User authentication seems not a major concern as it is present only in 3 papers [13], [37], [54]. We do not have enough data to justify this lack of interest in authentication, but we can speculate on it. An hypothesis we have is that most of the works give for granted that medical data are accessed only by professionals and that they are considered trustworthy. Similarly, we claim that the lack of interest in user authentication may indicate that there is not yet a widespread concern about opening the access of the health data to patients. This is, indeed, a requirement that only very recently has become a concern about opening the access of the health data to patients.

As a comment is mandatory: in the pool “confidentiality” there are 27 papers, namely [4]–[30]. Some of those were, per keywords, first gathered under “privacy”. A closer look revealed that they are using the term inappropriately since their concern is mainly about encrypting data. But, encryption per se is insufficient to guarantee that the user’s personal and sensitive information remains private during the whole data life cycle; more sophisticated techniques have to be in place for privacy to be protected. Thus, we decided to re-classify those works as being about confidentiality, adding those up to the ones already in that category.

Confidentiality is constantly discussed together with data integrity and data authenticity. That is because encryption is the technique that is more often adopted to enforce confidentiality in medical systems and the same technique is also proposed for data authenticity and integrity. In a total of 16 papers about data integrity (i.e., [5], [6], [8]–[10], [12], [13], [16], [18], [22], [23], [25], [29], [31]–[33]) only three works do not discuss confidentiality. We observed a very similar scenario with the category data authenticity. Only three works do not discuss confidentiality, out of 9 papers discussing data authenticity (i.e., [8], [9], [12], [22], [25], [29], [31]–[33]). Also, all works that examine data authenticity discuss data integrity too.

After confidentiality, the second and third most discussed security properties are privacy and access control. We found out that 20 works discuss privacy (the correct interpretation of this term) [14], [20], [25], [26], [30], [33]–[47], and that 19 papers discuss access control [11], [13], [19], [22], [23], [25], [29], [34], [37], [41]–[43], [45], [48]–[53].

User authentication seems not a major concern as it is present only in 3 papers [13], [37], [54]. We do not have enough data to justify this lack of interest in authentication, but we can speculate on it. An hypothesis we have is that most of the works give for granted that medical data are accessed only by professionals and that they are considered trustworthy. Similarly, we claim that the lack of interest in user authentication may indicate that there is not yet a widespread concern about opening the access of the health data to patients. This is, indeed, a requirement that only very recently has been debated and brought to the attention of the society. If concrete actions to open up access to patients were taken into consideration, it would, we expect, raise more attention about identification and authentication. Indeed the works which discuss such a feature have identification and authentication as their main requirement (e.g., see [55]). A similar speculation, i.e., that the patient-centred approach is not yet under the bull’s-eye in medical data security, concerns also the last two properties, transparency – the one of interest for this paper– and auditability. Auditability is subject of discussion of only 3 papers [33], [47], [56], ex equo (so to speak) with transparency which is mentioned as well in 3 papers [36], [42], [47].

Transparency is regarded as openness about policies and processes (we quote, “there should be openness and transparency about policies, procedures, and technologies that directly affect individuals and/or their individually identifiable health information” [36]) as well as a predisposition to increase responsibility and therefore presented with accountability (we quote, “Transparency and accountability will be critical to helping society manage the privacy risks that accumulate from expeditious progress in communication, storage, and search technology” [47]). Relevantly for this work, Routsalainen et al [42] propose transparency as the property to be informative towards patient. In fact they point out the lack of transparency since “[the] patient is not automatically aware which professionals or entities are processing her EHR and for what purposes. [The] patient are not aware of all disclosures of the content of her EHR”.

IV. Discussion and Conclusion

Our review has an obvious limitation: it considers papers that matched only two key-phrases, “security” and “medical data sharing”. However, “security” is a generic term under which we were able to find papers discussing more specific properties and requirements. “Medical data sharing” is our target, so this choice is justified. Still one could question why we did not searched for synonyms, and whether, in not doing so, we missed some important papers. Our searching on the whole body of the paper, however, was sufficient to catch works about electronic health records, bio-medical data, health care information systems, health-grid. Therefore, we judged the choice of our key-phrases sufficiently good for our scope.

This survey, organized around the works published in the last 10 years, shows that confidentiality and privacy are the major concerns in security for medical data (see also Figure 2). This comes with no surprise. About transparency, the survey shows that this requirement has just begun to be addressed; all the considered papers see transparency related to inform users and make policies and processes openly available. This seems to be the interpretation of “transparency” in the medical domain, a meaning which matches what we propose. However, there is no formalization of it and no standard solution that makes a medical system compliant to it.

We also observed that the majority of papers were published in the last 5 years, which endorses the hypothesis that security is a relative young concern in medical systems engineering. Although we already had some hint of it, after having looked at the recent growth of interest as this survey reports, it is evident that there is still little attention from the security community towards auditability, transparency, and user authentication, at least in relation to medical data systems.
did not searched into the literature of auditability and checked for use cases on medical data (e.g., as in [57]). Auditability and transparency are essential wherever humans need to be informed about practices in sharing sensitive personal data. No solution exists to comply with current EU regulations on this. Our first impression is that both categories are relatively understudied in the medical sectors. We expect a growth in attention to these properties as the idea of user empowerment will get more popular. User authentication seems suspiciously undervalued in the papers we surveyed. It is hard, from the data we have, to infer why. It may be that there are already good-enough authentication solutions to which medical systems can resort to. But, if we have to attempt another explanation, we are keen to suppose that current medical data are accessed mainly by professionals and that these roles are assumed to be trustworthy. Authentication is therefore implemented by simple login and password. Similarly as what we claimed while discussing transparency, if the EU directive suggesting to let users access their medical data should take off, we expect the problem of user authentication to become a pillar for the working of other several security features, and to foster a renewed interest.

REFERENCES


