ORIGINAL PAPER

The Integrated Holistic Security and Privacy Framework Deployed in CrowdHEALTH Project

Stefanos Malliaros[1], Christos Xenakis[1], George Moldovan[2], John Mantas[3], Andriana Magdalinou[3], Lydia Montandon[4]

[1]University of Piraeus, Piraeus, Greece
[2]Siemens, Brasov, Romania
[3]European Federation for Medical Informatics, Lausanne, Switzerland
[4]Atos, Madrid, Spain

Corresponding author: Stefanos Malliaros. University of Piraeus, Lampraki 126, 18532, Piraeus, Greece; E-mail: stefmal@unipi.gr. ORCID ID: http//www.orcid.org/0000-0003-3561-3141.

doi: 10.5455/aim.2019.27.333-340
ACTA INFORM MED. 2019 DEC 27(5): 333-340
Received: Nov 25, 2019
Accepted: Dec 28, 2019

ABSTRACT

Introduction: Individuals and healthcare providers need to trust that the EHRs are protected and that the confidentiality of their personal information is not at stake. **Aim:** Within CrowdHEALTH project, a security and privacy framework that ensures confidentiality, integrity, and availability of the data was developed. **Methods:** The CrowdHEALTH Security and Privacy framework includes Privacy Enhancing Technologies (PETs) in order to comply with the GDPR EU laws of data protection. CrowdHEALTH deploys OpenID Connect, an authentication protocol to provide flexibility, scalability, and lightweight user authentication as well as the attribute-base access control (ABAC) mechanism which supports creating efficient access control policies. **Results:** CrowdHEALTH integrates ABAC with OpenID Connect to build an effective and scalable base for end-users' authorization. CrowdHEALTH's security and privacy framework interacts with other CrowdHEALTH's components, for instance the Big Data Platform, that depends on user authentication and authorization. CrowdHEALTH users are able to access the CrowdHEALTH's database based on the result of an ABAC request. Moreover, due to the fact that the CrowdHEALTH system requires proofs during the interactions with data producers of low trust or low reputation level, the requirements for the Trust and Reputation Model have been identified. **Conclusion:** The CrowdHEALTH Integrated Holistic Security and Privacy framework meets the security criteria for an e-health cross-border system, due to the adoption of security mechanisms, such as user authentication, user authorization, access control, data anonymization, trust management and reputation modelling. The implemented framework remains to be tested to ensure its robustness and to evaluate its performance. The holistic security and privacy framework might be adapted during the project's life circle according to new legislations.

Keywords: Privacy, Security, Information Systems.

1. INTRODUCTION

Information Technology in the healthcare sector (1) is applied with the aim of improving patient outcomes and reducing cost in healthcare delivery (2). However, providers and individuals alike must trust that an individual's health information is private and secure (3). In order for patients to disclose their health information, they need to trust that the Electronic Health records (EHRs) will be protected (4) and that the confidentiality of their personal information is not at stake. In the European Union, the General Data Protection Regulation (5) lays down rules related to the protection of people with regards to their personal data processing and sets rules related to personal data exchange.

Within CrowdHEALTH project, a security and privacy framework that assures the confidentiality, integrity, and availability of the data (6) was deployed. Any interaction on the data will be handled by the integrated holistic security and privacy framework that provides: (i) trust management with the aim to quantify the trustworthiness of the participating users and healthcare ecosystem entities; (ii) data anonymization to ensure privacy; and (iii) access control and authorization to facilitate both integrity and authorized data disclosure by exploiting OAuth 2.0 (8). Due to the fact that e-health data can be sensitive, the CrowdHEALTH's security and privacy framework protects identifiable health information, such as the individual's past, present or future physical and mental condition, or the care an individual received. Taking into consideration that technical measures need to be taken to protect the identity of the individual, a Holistic Security and Privacy Framework in CrowdHEALTH project was developed for the protection of the CrowdHEALTH's resources and data. The CrowdHEALTH Security and Privacy framework outlines a structure which takes care of the security and privacy requirements of the project and includes guidelines that meet the security expectations, with the aim to protect the Confidentiality, Integrity, and Availability of the data, resources, services, and users of the system. CrowdHEALTH integrates ABAC (9) with OpenID Connect (7) to create an efficient, effective, and scalable infrastructure (15) for authorization of the end-users. OAuth is also an open standard for identity delegation and authorization. There are two OAuth versions (10), (11), although OAuth 2.0 is not backwards compatible with OAuth 1.0. OAuth 1.0 is a protocol for identity delegation, while OAuth 2.0 is a framework, which aims at providing authorization for web and desktop applications, as well as mobile phones and smart devices. OAuth 2.0 does not provide encryption, digital signatures, or client verification services, but instead it uses the Transport Layer Security (TLS) protocol to offer a degree of confidentiality and server authentication. OpenID Connect (12) is a protocol which is based on OAuth 2.0 and exploits a Java Script Objection Notation / Representational state Transfer (JSON/REST) based identity built-in functionality, alongside with JSON Web Tokens (JWT) (13). OpenID Connect consists of an identity layer on top of the OAuth 2.0 framework, which enables clients to perform identity verification, based on the authentication performed by an authorization server. Moreover, some basic profile information is obtained about the identified person in an interoperable REST-like manner. Within CrowdHEALTH project, a security and privacy framework that ensures confidentiality, integrity, and availability of the data, was developed.

2. AIM

The aim of this paper is to present an initial overview of the Holistic security and privacy Framework deployed in CrowdHEALTH to mitigate the risk of security breaches.

3. METHODS

The CrowdHEALTH Security and Privacy framework incorporates Privacy Enhancing Technologies (PETs) to comply with the GDPR EU laws of data protection. The objective of PETs is to ensure the confidentiality when dealing with personal information. CrowdHEALTH deploys PETs to achieve user authentication, authorization, and access control, trust evaluation and modelling, as well as performs data anonymization of the e-health data that are managed by CrowdHEALTH. CrowdHEALTH deploys state-of-the-art authentication protocols to protect users against security threats. More precisely, CrowdHEALTH exploits federated identity management, by employing a secure Single Sign-On mechanism, which enables the user identification for entities that rely on the result of the authentication process. CrowdHEALTH deploys OpenID connect is an authentication protocol providing flexibility, scalability, and lightweight user authentication. As long as authorization is concerned, CrowdHEALTH relies on OAuth 2.0, which is a token-based open standard for user authorization. OAuth 2.0 provides a process for resource owners to authorize third-party access to their resources without having to perform authentication and maintain user credentials. CrowdHEALTH deploys the attribute-based access control (ABAC) mechanism to build effective access control policies. ABAC is a scalable mechanism, which relies on the user attributes, the resource attributes, and the access control rules defined by system administrators to permit or forbid access to a requested resource. CrowdHEALTH deploys a Trust evaluation model that provides the option to compute the users' trust rating based on several parameters. The set of rules creating

the trust and the reputation model, are the core part of the Trust and Reputation Modelling component. Moreover, the Trust and Reputation Model includes a third model - namely a Reaction Model, which specifies what kind of event the trust and reputation mechanisms should generate and propagate to the system.

4. RESULTS

In CrowdHEALTH's architecture, data is maintained in the Datastore. The data is encrypted and transferred between different services. To assure the confidentiality, integrity, and availability of data, CrowdHEALTH applies security mechanisms to achieve effective authentication, authorization, access control, anonymization, and trust reputation. Each one of these mechanisms is presented below:

	SAML 2.0	OpenID Connect
Service Provider	Client libraries	Client libraries
Identity Provider	Identity Provider libraries	OpenID Connect Provider libraries
Attribute Provider	Attribute provider provides further detail to enrich SAML assertion Requires further step to populate assertion with user attributes	OpenID Connect Provider The userinfo endpoint returns claims about the end-user
Attributes	SAML attributes	OpenID connect scopes
Discovery Service	Requires pre-agreed metadata	Single discovery service for client allowing sites & apps can validate your users
Privacy	Yes	JSON Object Signing and Encryption (JOSE)
Signing	Yes	JSON Web Token (JWT)
Mobile Apps	No, SAML web profile for web browser only	Both web browser & mobile apps
Support for SSO	Web SSO only	Yes
Form Rendering	Both client and identity provider	Normally Identity provider

Table 1. Comparing SAML 2.0 and OpenID Connect

4.1 User Authentication and Authorization

The main idea of OpenID Connect is to create an API which provides seamless authentication and authorization that can be built lightweight and implemented for applications. There are key differences between Security Assertion Markup Language (SAML) and OpenID Connect as shown in Table 1. The first is that OpenID Connect applies most of the complexity to the OpenID Connect Provider, whereas SAML distributes this complexity to both the provider and the client. Also, OpenID Connect migrated from XML (Extensively Markup Language) to JSON, which is supported by all modern programming environments. Moreover, JWT which also supports Signing and Encryption (JOSE) (14) allows for more practical and compact tokens by using the XML language.

CrowdHEALTH's authentication protocol enables its users' to seamlessly authenticate and use different services without the need of reauthentication, and without the need of a separate user account. The user account contains the attributes of the user for example Name, Surname, Nationality, Email and Password. Regarding the password, it is essential to enforce a password policy, since it is one of the basic security measures to prevent unauthorized access. Since, most users tend to select easy to remember passwords, and they do not want to change it, a well-defined password policy is the first keystone to protect the system from unauthorized access. CrowdHEALTH's password policy includes the following principles:

- Password Strength: the strength of the passwords is one of the most critical properties of a password. The password strength depends on several parameters and these are the minimum password length, and the character set that the password is comprised of. In CrowdHEALTH, a password can be no shorter than 8 characters, and the users have to select passwords than contain both uppercase and lowercase letters, numbers and symbols.
- Password expiration time: Users are encouraged to change their password on a regular basis as per password expiration time. The password expiration time is one year.
- Trivial password selection: CrowdHEALTH does not accept passwords that can be easily guessed. A password cannot contain certain words, such as the word password, or the users' first name and surname.
- Uniqueness of passwords: The uniqueness of passwords specifies the number of new passwords that the user has to select before being able to reuse a previously used password. In CrowdHEALTH, users cannot use the same passwords until they have changed their password three times.

The OpenID Connect protocol follows a specific execution flow to achieve authentication and authorization of the CrowdHEALTH end-users. The execution flow is described below:

Step 1: The CrowdHEALTH end-user requests access to some resources via the client.

Step 2: The client redirects the session of the end-user's web browser to the OpenID Connect Provider's system for authentication.

Step 3: The OpenID Connect Provider authenticates the CrowdHEALTH end-user by using either username or password, or even by using a two-factor authentication.

Step 4: After the authentication has been performed, the OpenID Connect provider performs a redirection of the end-user's web browser or application to the client, including an authorization code.

Step 5: The client sends a POST request to the OpenID Connect provider alongside with the authorization code, provided by the end-user.

Step 6: The OpenID Connect Provider responds to the client with an ID token, which contains details about the attributes of the authenticated use in JWT form, and an optional Access token which is used for accessing resources. Apart from user identification, the OpenID Connect Provider also identifies the client who requested the initial authentication.

Step 7: If in the previous step the OpenID Connect Provider also sends an Access token, then the client may send it back to the OpenID Connect Provider to request further profile information of the CrowdHEALTH end-user.

Step 8: The OpenID Connect Provider returns the user profile to the client containing the requested information (e.g. email).

After the completion of the execution flow of the OpenID protocol, the end-user has been authenticated by the OpenID Connect Provider and authorized by the Client. Also, the client has been authorized to access the protected resources by using the token obtained by the OpenID Connect Provider. The OpenID Connect provider is an internal part of the CrowdHEALTH system and is responsible for authenticating users, and issuing tokens that are passed to clients, so that clients can request data on behalf of the end-users (Figure 1).

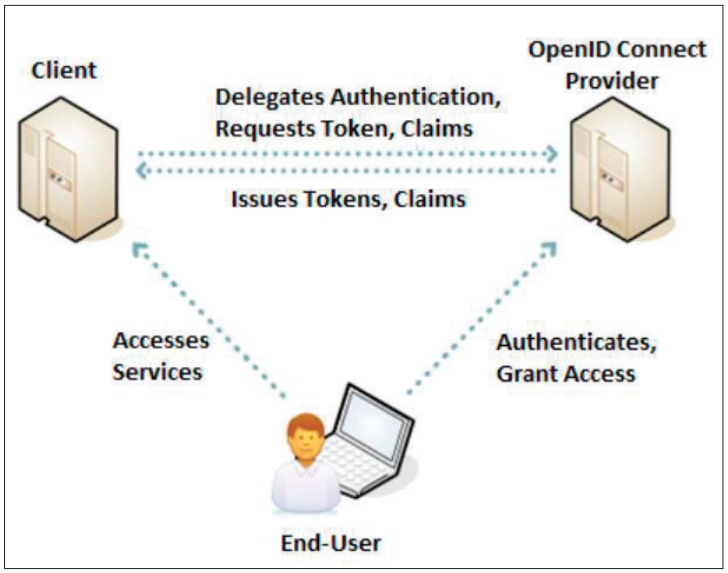

Figure 1. OpenID connect execution flow

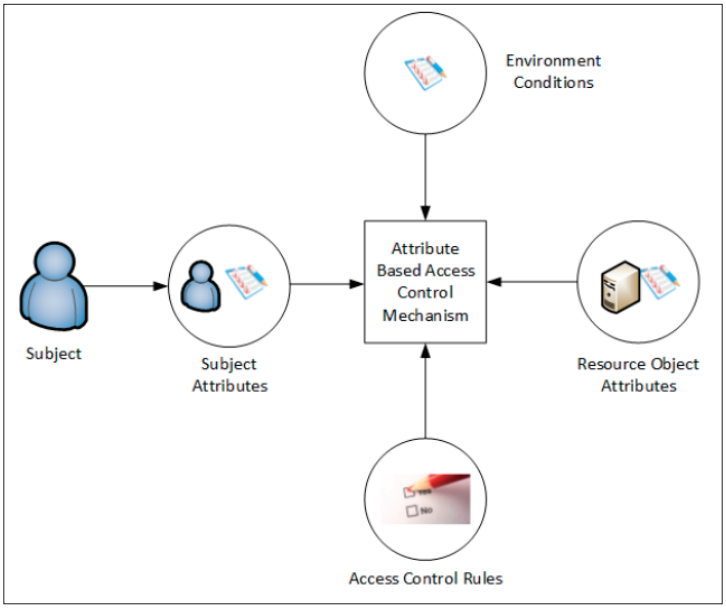

Figure 2. ABAC architecture

4.2 User Access Control

Since CrowdHEALTH is a cross border system that manages and analyses anonymized health data from various data sources, it requires integration of Attribute-Based Access Control (ABAC) in the CrowdHEALTH platform to be able to perform effective and efficient access control policies. By exploiting both the end-users and the resource attributes defined between different organizations, ABAC does not rely on explicit authorizations that are required prior to the access request to a resource. Also, it is scalable for large enterprises, where the management of other access control mechanisms, such as role-based access control and access lists, would be time inefficient. ABAC does not require directly assignments to end-users or their roles or groups before the request is performed by the end-user. When an end-user carries out a request, ABAC can decide based on the assigned attributes of the end-user in combination with the attributes of the resource, and other policies specified for the specific end-user and resource. ABAC relies on the evaluation of subject attributes, object attributes, environment conditions and the access control

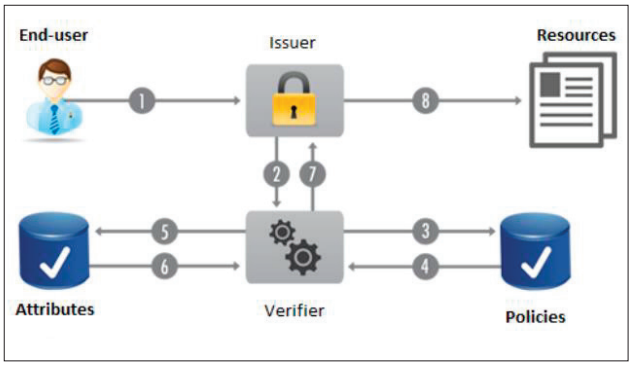

Figure 3. ABAC execution flow

policy defining allowable operations for subject-object combinations, as depicted in Figure 2. These components are mandatory for every ABAC implementation, ranging from a small isolated system, to a more complex system with multiple data sources and various user types.

CrowdHEALTH integrates ABAC with OpenID Connect to create an efficient, effective, and scalable infrastructure for authorization of the end-users. The ABAC architecture consists of three entities: the end-user, the issuer, and the verifier. The end-user is the subject that wants to access the resources of the system. The authentication is performed by exploiting credentials or other attributes that have been issued by the issuer, which is the OpenID Connect Provider of CrowdHEALTH. The verifier, also called client in CrowdHEALTH, is an entity that can request attributes of the end-user on behalf of the end-user, based on an authentication performed by an authentication server as per Figure 3.

CrowdHEALTH's security and privacy framework interacts with other CrowdHEALTH's components that rely on user authentication and authorization as for example Big Data Platform. The users of CrowdHEALTH are able to access the CrowdHEALTH's database based on the result of an ABAC request. The ABAC component of the security and privacy framework shall state the permission that the users' have on the requested resource (e.g. CrowdHEALTH's database) as follows:

• Read-Only: The user has read-only access to the requested resource.

• Write-Only: This user (or application) has write-only access to the indicated resource. This is mostly for applications or layers that will be adding information to the indicated resource but don't need to retrieve information.

• Read-Write: The user has read and write access to the requested resource.

• Admin: The user is the administrator of the resource and is allowed to modify both the content and the metadata of a resource. For example, an administrator is allowed to change the database schema, as well as from adding content to the database.

4. 3. Data Anonymization

In CrowdHEALTH, the procedure that is used to anonymize the data includes two stages. The first one aims at removing attributes that directly identify an individual, the so called direct identifiers (e.g. names, social security numbers, email addresses, ID card numbers, passport numbers, address, phone numbers) (16-20) while the second aims in pseudo-anonymizing the indirect identifies, in a way that the individual's privacy is not jeopardized (Table 2). The anonymization takes place at the source of data to avoid any potential security threats during the transmission of data to the CrowdHEALTH system, and to also avoid identification from information disclosure.

Direct identifiers			Indirect identifiers				
Name	Surname	Social Security Number	Birth Date	Sex	Weight	Disease	Religion

Table 2. A hypothetical table containing direct and indirect identifiers

CrowdHEALTH employs "suppression" and "generalization" as the main methodologies for anonymizing indirect identifiers. In suppression, the value or a part of the value of the indirect identifiers is replaced by an asterisk. CrowdHEALTH achieves k-anonymity (18) in the processed datasets. For CrowdHEALTH to achieve k-anonymity, the following steps were performed:

• Attribute identification: An expert states the attributes that can be used as direct identifiers. Removal of direct identifiers: The direct identifiers are removed and not contained in the created anonymized dataset.

• Unique Random IDentifier: CrowdHEALTH will create a Unique Random IDentifier (URID) for every patient that the database has information about. This is crucial, since different sets of data are stored in different files, and all the information is related to each other by the URID. The usage of URID cannot break k-anonymity, due to the amount of data that will be imported into the CrowdHEALTH system.

• Threat model: A threat model consists of adversaries, security threats, alongside with the information they might hold, that can be exploited

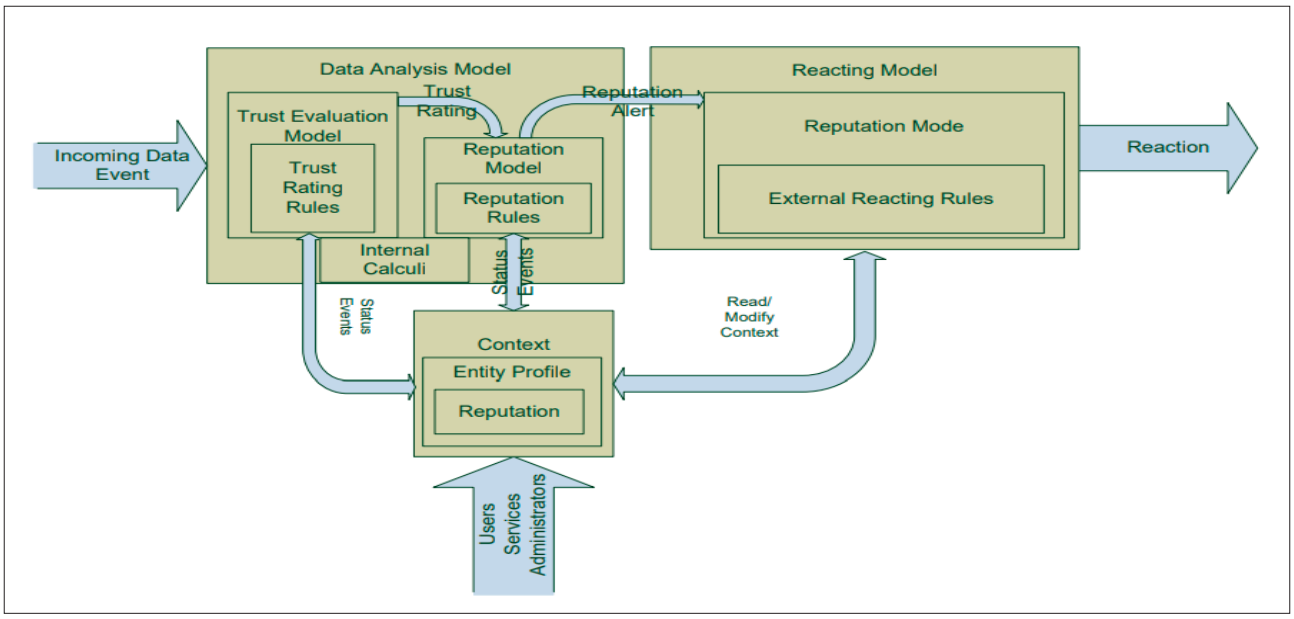

Figure 4. Trust and Reputation Model components

to re-identify the data subjects.

• Utility of anonymized data: The utility of the created anonymized dataset will be determined.

• Create the anonymized dataset: The initial dataset undergoes the anonymization process, thus creating the anonymized dataset.

4.4. Trust and Reputation Modelling

The requirements for the Trust and Reputation Model have been identified:

• Service Level Trust: Services within the CrowdHEALTH platform should be able to query reputation ratings for specific services (e.g. a heart rate monitoring service).

• Measurement Level Trust: Services within the CrowdHEALTH platform should be able to query reputation ratings for specific measurements (e.g. a single or a set of heart rate monitoring readings)

• Device Level Trust: Services within the CrowdHEALTH platform should be able to query reputation ratings for specific devices. (e.g. a wearable heart rate monitoring device).

• Multiple Evaluation Criteria: The Reputation Model should be based on multiple criteria, as defined and required by the entities processing the data (and the device platform interactions). This includes the requirement for being able to interactively define and update the Trust and Reputation Models during runtime, to have a flexible and configurable platform component.

• Multiple Observers: The Trust and Reputation Model should support and mitigate reports and ratings from multiple observers. This includes the ability to define new observers and evaluation rules.

• Configurable Reaction Events: The Trust and Reputation Model should be able to configure and dispatch reaction-based events according to the subscribers' specification. I.e., changes in trust and reputation ratings greater than 25% over a specific time interval.

Figure 4 presents an overview of the interactions between the three models previous presented. The Trust, Reputation and Reaction Models are all interacting components, driven by internal rules for processing information. Two types of modalities can be distinguished: active and passive ones. In the active case, the mechanisms would compute and actively drive any interaction with the data and data producing entities. The CrowdHEALTH system requires proofs or additional information during the interactions with data producers which have a low trust or reputation level. In the passive case, the system is meant to determine what behaviour is considered "normal" or "anomalous" without using the trust and reputation ratings to drive any interactions. Only subsequent internal data processing steps would adjust depending on the existing ratings. Within all these components, four main specific values which determine the trustworthiness of a specific measurement, service or device were identified:

• An observer: The entity of the platform which observes and monitors the data streams and compares the observed behaviour with the expected one and therefore produces the trust ratings. A continuous, systematic evaluation of the data streams is required for computing correct

```
;=======================

; Rule valC-mima

;=======================
(defrule valC-mima "checks valC (str val)a-priori boundary conditions
 of each observer [ 0 < valC < 40 ]"
      (a-valC-mima (obsN ?obsN)(strN ?strN) (ruID ?ruID)(minA
?minA)(maxA ?maxA))
      (a-str (strN ?strN)(valC ?valC) (timC ?timC))
      (test (or (< ?valC ?minA)  (> ?valC ?maxA)))
      =>
      (assert (bad ?strN ?obsN "valC-mima" ?ruID (getTime ?timC)))
      ;(printout t "Alert! "?obsN"'s "?strN" values are abnormal"
crlf)
)
```

Figure 5. Example of a rule generating an alarm

trust ratings.

• Indicator: An Observer is usually unable to immediately categorize an entity as acting correctly, incorrectly or maliciously (i.e. based on a sudden spike in temperature or heart rate values). Instead, the Observer is detecting deviations from the expected values or (known) trends, specific to the entity observed. These deviations are instead considered as possible indicators for some kind of errors. The opposite is correct as well – value within expected ranges is considered correct behaviour.

• Trust Ratings: Indicators of abnormal/erroneous behaviour may lead the Observer to update his appreciation of an entity – more specifically: confidence, reliability or trustworthiness of an entity will be decreased. This information is quantified as a value, the trust rating. This value can be either a number, or a complex data structure.

• Reputation Ratings: Reputation Managers have the task of merging multiple trust ratings in reputation ratings. This is a global view and evaluation of multiple trust ratings.

4.5 Statistic Evaluation of Trust and Reputation

The role of this initial proof of concept is to evaluate and demonstrate the streaming mode processing capabilities and the simplicity of the calculations performed. Further iterations of the Trust and Reputation Model plan to employ a more efficient evaluation system, but still similar to CLIPS. As an example, the following rule (employed in RERUM) generates an alarm (Figure 5) if the observed values are outside of an interval [minA, maxA]:

5. CONCLUSIONS

CrowdHEALTH approach is based on privacy by design. The CrowdHEALTH Integrated Holistic Security and Privacy framework presents security requirements of an e-health cross-border system and proposes security mechanisms, such as user authentication, user authorization, access control, data anonymization, trust management and reputation modeling that are applied in the project. The implemented framework has to be thoroughly tested in order to prove its robustness and assess its performance. The holistic security and privacy framework might be adapted during the project's life circle according to new legislations..

- **Acknowledgments:** CrowdHEALTH project is co-funded by the Horizon 2020 Programme of the European Commission Grant Agreement number: 727560 – Collective wisdom driving public health policies.
- **Author's contribution:** Each author gave substantial contribution in acquisition, analysis and data interpretation. Each author had a part in preparing article for drafting and revising it critically for important intellectual content. Each author gave final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.
- **Conflict of interest:** None declared.

REFERENCES

1. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E. et al. Systematic Review: Impact of Health Information Technology on Quality, Efficiency, and Costs of Medical Care. Ann Intern Med. 2006; 144(10): 742-752. doi: 10.7326/0003-4819-144-10-200605160-00125
2. Mantas J. Future trends in Health Informatics - theoretical and practical. Studies in health technology and informatics. 2004; 109: 114-127.
3. Arora S, Yttri J, Nilsen W. Privacy and Security in Mobile Health (mHealth). ResearchAlcohol Res. 2014; 36(1): 143-151.
4. Mantas J. Electronic health record. Studies in health technology and informatics. 2002; 65: 250-257.
5. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance) OJ L 2016; 119(4.5): 1-88.
6. Zubaydi F, Saleh A, Aloul F, Sagahyroon A. Security of Mobile Health (mHealth) Systems. Proceedings of the 2015 IEEE 15th International Conference on Bio-

informatics and Bioengineering (BIBE). 2016: 1-5. doi: 10.1109/BIBE.2015.7367689
7. OPenID Connect - Welcome to Open ID Connect. Available at: http://openid.net/connect/.
8. Hardt D. The OAuth 2.0 Authorization Framework. Internet Engineering Task Force (IETF). 2012. Available at: https://oauth.net/2/.
9. Fisher B, Brickman N, Burden P, Jha S, Johnson B, Keller A. et al. Attribute Based Access Control. NIST Special publication 1800-3B. 2017 Volume B Available at: https://nccoe.nist.gov/sites/default/files/library/sp1800/abac-nist-sp1800-3b-draft.pdf.
10. Hammer-Lahav E. The OAuth 1.0 Protocol. Internet Engineering Task Force (IETF). 2010. Available at: https://tools.ietf.org/html/rfc5849.
11. Hardt D. The OAuth 2.0 Authorization Framework draft-ietf-oauth-v2-31. IETF. 2012. Available at: https://tools.ietf.org/id/draft-ietf-oauth-v2-31.html
12. OpenID Connect. Available at: http://openid.net/connect/
13. Jones M, Sakimura N. JSON Web Token (JWT). Internet Engineering Task Force (IETF)2015. Available at: https://tools.ietf.org/html/rfc7519
14. Schaad J, Hodges J, Hildebrand J, Turner S. JSON Object Signing and Encryption (JOSE), IANA, 2015. Available at:http://www.iana.org/assignments/jose/jose.xhtml.
15. Weil EC. ABAC and RBAC: Scalable, Flexible, and Auditable Access Management. IT Professional. 2013; 15(3): 14-16.
16. ISO/TS25237:2008. Health informatics – Pseudoanonymization. International Organization for Standardization. Available at: https://www.iso.org/standard/42807.html.
17. ISO 25237:2017 Health informatics – Pseudoanonymization. International Organization for Standarization. Available at: https://www.iso.org/standard/63553.html
18. Fiorini AR, Masic I. Managing Information in Medical Informatics. Acta Inrom Med. 2017 Sep; 25(3): 192-5. doi: 10.5455/aim.2017.25.192-195.
19. Masic I, Ridjanovic Z, Pandza H, Masic Z. Medical informatics. Avicena, 2010: 544 pp. ISBN: 978-9958-720-39-0.
20. Sweeney L. "k-anonymity: a model for protecting privacy," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. 2002; 10(5): 557-570. doi: 10.1142/S0218488502001648