Improving Data Sharing in Data Rich Environments

Erisa Karafili*, Emil C. Lupu*, Alan Cullen†, Bill Williams†, Saritha Arunkumar‡ and Seraphin Calo§
*Imperial College London
Email: [e.karafili, e.c.lupu]@imperial.ac.uk
†BAE Systems
Email: [alan.m.cullen, bill.williams]@baesystems.com
‡IBM UK
Email: saritha.arun@uk.ibm.com
§IBM US
Email: scalo@us.ibm.com

***Abstract*—The increasing use of big data comes along with the problem of ensuring correct and secure data access. There is a need to maximise the data dissemination whilst controlling their access. Depending on the type of users different qualities and parts of data are shared. We introduce an alteration mechanism, more precisely a restriction one, based on a policy analysis language. The alteration reflects the level of trust and relations the users have, and are represented as policies inside the *data sharing agreements*. These agreements are attached to the data and are enforced every time the data are accessed, used or shared. We show the use of our alteration mechanism with a military use case, where different parties are involved during the missions, and they have different relations of trust and partnership.**

***Keywords*-Big data; data sharing; data access; usage control; DSAs; drone systems; military scenario.**

I. Introduction

Security concerns regarding sharing and using data are becoming a serious issue especially with the increasing use of IoT devices as well as the proliferation of big data mechanisms. Nowadays the data are collected, accessed, used, and shared, therefore concerns related to how the data is processed, shared, sanitized and its quality are naturally raised.

In this work, we propose the use of a policy based technique that enables secure data sharing, and can be easily used in data rich environments. The expressivity of the policy language permits the representation of the different agreements for sharing and using the data, called *data sharing agreements* (DSAs). The DSAs are represented as a set of policies composed of the various access and usage control rules, as well as the security, business and legal requirements. This set of policies is attached to the data with a similar mechanism to the *sticky policies* [1], and are enforced every time a request to access/use/share the data is made. Our technique permits the correct access and usage of data as well as their correct sharing, as it describes through the DSAs the rules that should be applied for every case.

Our policy language represents through its policies the rules and constraints that should be applied during the access, usage and sharing of data. These policies are attached to the data, and are subsequently enforced when entities request access to the data. The correct enforcement of the policies avoids inappropriate accesses to the data. In this work, we will not deal with the enforcement mechanism and the communication of data from one entity to the other, where for the latter we assume that the devices involved in this mechanism are trusted. We will give special attention to the policy representation, and how the various relations between the involved entities can affect the data sharing and the quality of the shared data.

Different entities access the data that is carried and shared in different ways. The type of relations the entities have with the data owner or the data releaser, which defines the context of use of the data, affects the accessed data. In this work, we introduce an extension of a policy language that incorporates an alteration mechanism. The alteration is made through the use of DSA policies attached to the shared data, and it is not made to the data themselves but to the shared ones. Depending on the relations the entities have between each other, mainly trust relations, different alterations apply to the shared data. We show how our predicates represent the various relations of trust between entities with a use case taken from a military scenario. Our mechanism reflects the various relations between entities in the degree of alteration made to the shared data. The range of alteration is described by the DSA policies.

We introduce the related work for secure data sharing, together with access and usage control of data, in Section II. In Section III, we represent the used policy language, together with an example and the introduced extension that permits secure data sharing in data rich environments. We show the use of our data sharing mechanism with data alteration in a use case taken from a military scenario, in Section IV. We discuss our work and some of its future challenges, in Section V. Finally, we conclude in Section VI.

II. Related work

The increase of connectivity we are experiencing comes along with an increase in the number and variety of attacks to the data. Therefore, more solutions are focused on protecting the data [2]–[6] rather than protecting the environments where the data is stored [7], re/used [8] or transferred [9].

Different solutions have been developed for solving the problems of controlling the access and usage of data. A well known technique is role-based access control [10], [11] which controls data access depending on the user roles. Usage control (UCON) is another well studied concepts [12], [13] where the access and usage of digital information is controlled and the problem of delegation of rights is also addressed. Usage control has been widely used in distributed systems [14] in particular, when having a data flow between different connected systems or global usage control policy to be enforced.

Another interesting approach for sharing and accessing data is the use of sticky policies [1], [4]. Sticky polices are machine readable policies that contain conditions and constraints attached to data that describe how the data should be treated, when accessed by multiple parties. The *sticky policy paradigm* introduced in [15] provides the technologies for enterprise privacy enforcement and exchange of customer data. The promised privacy rights and obligations are specified through a privacy control language [16], for authorisation management and access control, that includes user consents, obligations and distributed administration.

The data usage problem concerns different entities that are using the data, from where the data is stored and processed (called data processor), to who can access the data and which parts. Before sharing the data, the entities should agree for the different rules and policies to be used, where the data owner as well as the legislative rules play a decisive role. Agreements are necessary to define policy and expectations, like DRM-style enforcement to separate data protection from data communication mechanism. These agreements are called data sharing agreements [17] (DSAs). The DSAs describe not only the agreements between the data owner and data consumers, but also the different business and legislation rules. A policy analysis language is used for representing the DSAs in [18] which permits various analysis tasks like redundancy, gap and conflict analysis. This policy language together with the analysis tasks has been used for reasoning about data quality attributes during the managing and processing of data in data-intensive environments [19].

Along side security problems, there is a need in resolving problems around data privacy in big data environments. When there is a problem like security breach in big data it may result in causing serious legal consequences and damage to reputation. In [20] the authors discuss the security issues of big data in cloud computing, in particular data privacy issues and propose new approaches like encryption, logging, node authentication, and access control. Various approaches have focused on resolving security issues raised by the use of big data in healthcare environments. The authors of [21] described some of these issues, and the approaches used, like encryption algorithms e.g., *AES*, authentication methods e.g., *one time password* and *two factor authentication*. The introduced methods apply a set of security constraints and access control in order to achieve integrity, confidentiality, and also satisfy privacy. The work introduced in [22] introduces a privacy preserving method of sharing the medical data in the cloud environment. It also highlights the classification of the attributes based on the vertical partitioning of the dataset to solve various privacy concerns. A discussion about a statistical analysis and cryptography is made, where this combination allows and ensures both privacy protection and data utilisation.

III. Data sharing for data rich environments

Our mechanism for improving data sharing is based on a framework introduced in [18], [23] used for data sharing and usage control in cloud platforms. In this work, we extend the framework for data rich environments to be applied to data sharing and usage control in military scenarios. The framework introduced in [18] focused on analysing the various policies and constructing an efficient and conflict-free set of policies by the use of a policy analyser and conflict solver based respectively on an abductive approach and an argumentation-based reasoning. The set of produced policies describe how the data should be shared, used and accessed. The expressivity of the used language permits the representation of complex security requirements as well as legal and business rules for the access and usage of data.

In data rich environments, different users need to access different parts of the same data, where depending on the users an alteration of the quality of data needs to take place. In this work, we extend the framework with the novel concept of *data quality alteration*, in our case is a restriction made to the quality of the data, that is crucial for data sharing in data rich environment. We introduce discrete types of quality of data that are translated into a discrete number of data restrictions. Usually the quality of data is defined with properties like data accuracy and data freshness. As in our use case, we mainly deal with images, the data quality is defined through the accuracy of the image i.e., image resolution, dimension, band of colour etc., and the freshness of the image, i.e., how updated is the image with the current state of the observed object. The restriction of the data quality is made automatically to the shared data and it reflects the context of use, e.g., the type of trust relations or partnerships. The types of data quality restrictions together with the context of applications are described by the DSAs policies. A special attention is given to conflictual data access, i.e., users that might create conflicts when accessing certain data related to other users, and how to avoid them.

A. A policy analysis language for data sharing

The policy language [24] used by our data sharing framework is based on the Event Calculus [25] which permits the representation of dynamic properties.

The policy language we use defines policies that are composed of a *subject*, *targets*, and *actions*, together with a representation for time, standard relations ($=$, $\neq$, $<$ etc.) and arithmetical functions ($+$, $-$, $/$, $*$). It is composed of policy rules that represent authorisation (permit and deny) and obligation rules. Some of the predicates of this policy language are as below:

$req(Sub, Tar, Act, T)$
$permitted(Sub, Tar, Act, T) \quad denied(Sub, Tar, Act, T)$
$obl(Sub, Tar, Act, T_s, T_e, T)$
$cease\_obl(Sub, Tar, Act, T_{init}, T_s, T_e, T)$
$fulfilled(Sub, Tar, Act, T_s, T_e, T)$
$violated(Sub, Tar, Act, T_s, T_e, T)$

The predicate $req(Sub, Tar, Act, T)$ represents the request that a given subject, Sub, is making at the instant of time T, for performing a given action, Act, upon the target, Tar. The predicates $permitted(Sub, Tar, Act, T)$ and $denied(Sub, Tar, Act, T)$ represent respectively that a given subject is permitted/denied at the instant of time T, to perform a certain action upon the target. The predicate $obl(Sub, Tar, Act, T_s, T_e, T)$ denotes that at the instant of time T a given subject is placed under an *obligation* to perform a certain action upon the target during the interval of time from T_s to T_e. On the other hand, the predicate: $cease\_obl(Sub, Tar, Act, T_{init}, T_s, T_e, T)$ denotes that at the instant of time T, the obligation initially contracted for Sub at T_{init} to perform a certain action upon the target, between T_s and T_e, is no longer binding. The predicates $fulfilled(Sub, Tar, Act, T_s, T_e, T)$ and $violated(Sub, Tar, Act, T_s, T_e, T)$ denote respectively that the obligation at time T for the subject to perform an action to the target, from T_s to T_e, has been fulfilled/violated.

The used language [24] contains three components: predicates belonging to the underlying event calculus language, specific predicates which represent facts about the scenario, and tokens which represent actions or constant properties in the policy language. Some of the domain description predicates are *initiates*, *terminates*, *holdsAt*, *happens*. The $initiates$ predicate describes the state properties that hold due to an event, while $terminates$ describes which properties stop holding after an event. The $holdsAt$ predicates means that a given property is true in a state, while the $happens$ predicates indicates the event that occurs in a given instant of time.

B. A policy language example of application

We consider the case of a military mission where different entities are involved and the access to information depends on the trust relations between the entities and the context. Information can be carried in different ways, e.g., cloud, satellites, even the drones can carry the information from one entity to the other. In this work, we assume that the entities involved in the carrying of the data are trusted. In our scenario, drones are performing a reconnaissance mission to gather information from certain areas. We extend the language with two scenario-specific predicates for our discussion: $owns(X, Y)$, and $collect(D, Y)$:

- $owns(X, Y)$ asserts X is the 'owner' of Y, similarly
- $collect(D, Y)$ asserts Y 'collects' data D. For example, drone Y collects image data D.

The entity that owns the drones has the permission to access their data, as described below:

$$permitted(C, Data, access, T) \leftarrow holdsAt(owns(C, D_i), T), holdsAt(collect(Data, D_i), T'), T > T'$$

where C is the country that owns the drones D_i, and $Data$ are the data collected by these drones. The permission to access is given after the data are collected, $T > T'$. We need the predicate $owns$ to hold while the country C is permitted to access the data.

C. Modelling data restriction

It may be the case, for classification, copyright, or simply bandwidth restriction reasons, that access is permitted, subject to some form of limitation. Tabular data may have selected columns or rows removed; images may be supplied at limited resolution. Therefore, a data transformation takes place between the data in its original form, and the data delivered to the recipient. The transformation is dependent on a number of factors:

- the type of data, including its source;
- the owner of the data;
- the recipient of the data.

We extend our language with the restrict predicate[1]:

$$restrict(Owner, Recipient, RawData, OutData)$$

where given the $Owner$[2] of the data, denoted by $RawData$, this data are restricted, denoted by $OutData$ depending on the data recipient, denoted by $Recipient$, and its relations. The data are transformed in different ways depending on the requirements, e.g., the type of recipient, the relations of trust, context.

Going back to the previous example, let A and B represent the owner of the drone and the recipient of the data

[1]The Appendix discusses the interpretation of $restrict$ in our notation.
[2]We assume, in this case, the owner of the data is also the owner of the drones that collected the data.

respectively, and, as before, the drones are D_i, then

$$permitted(B, BData, access, T) \leftarrow holdsAt(owns(A, D_i), T), holdsAt(collect(AData_i, D_i), T'), restrict(A, B, AData_i, BData), T > T'$$

permits access to restricted collected data. This permission is given after the data has been collected.

The introduced policy language extended with the predicates: $owns$, $collect$ and $restrict$, defines a range of context-dependent transformations which can be applied to data dissemination. These transformations, collectively, form a data sharing agreement (DSA) which, when attached to the data and activated upon delivery, can enforce the information exchange requirements of a coalition agreement.

Going back to our example, depending on where the data are collected, and the context of use, the quality of the data is altered. The user interface might also show similar restrictions. Envision a user requesting image data asks for a list of image resolutions available. This information may be as sensitive to the drone owner as the image data itself and so each item in the list might have a data quality equal to the corresponding image quality. The collected data should be structured in a way to support their enforcement. Hence, there are two criteria for controlling access to the data with respect to the identity of the recipient: quality q together with the extent to which it must be degraded ($0 \leq q \leq Q$), and the location where it was collected:

$$\{(r, s) : 1 \leq r \leq R, 1 \leq s \leq S\}.$$

We denote with q the actual quality of the data, with Q the original quality of the data, in case no change is made to the data then $q = Q$. We denote with (r, s) the geographical coordinates from where the data was taken/collected, where R and S are the set of all the possible coordinates respectively for r and s.

These two criteria are used differently, because geographical locations are disjoint, while on the other hand, quality is usually cumulative. We denote by $Item_{q,r,s}$ an element (data) of our dataset that has quality q and was collected at geographical location of (r, s). Thus, for the scenario that we are working with, given two elements of a collected dataset, then they are related as described below:

$$\begin{aligned} Item_{q_1,r_1,s_1}, Item_{q_2,r_2,s_2} &\in Data \\ Item_{q,r_1,s_1} \neq Item_{q,r_2,s_2} &\Leftrightarrow (r_1 \neq r_2 \vee s_1 \neq s_2) \\ Item_{q_1,r,s} \subseteq Item_{q_2,r,s} &\Leftrightarrow (q_1 \leq q_2) \\ Item_{0,r,s} &= \emptyset \end{aligned}$$

where the final equation indicates that an item with $q = 0$ contains no information, and $Data$ is our dataset. These relations together with the criteria are the basis for specifying the $restrict$ predicate to implement the data quality policies.

Essentially the predicate $restrict$ encapsulates entirely the various agreement made of how the data should be restricted depending on the relations the entities they have between each other. This predicate is able to regulate even situations where full data access is granted to a recipient. In the next section, we will give an overview of how the $restrict$ predicate is defined, depending on our use case.

IV. Use Case: A Military Scenario

In this section, we introduce an example of how data sharing agreements can be applied in a military scenario, where the data are used for situational awareness.

A. Introduction to the use case

We are in a military scenario where different countries are involved. Let's assume we are dealing with a squad of drones that is performing a reconnaissance operation in an hostile territory. This operation is part of a bigger military mission carried out within a coalition.

In our scenario, the squad of drones is owned by a military organisation of country X. Country X is part of a military alliance $\mathcal{A}$, where other countries are involved, e.g., country Y. To fly drones to the hostile territory the mission needs support from two countries that are not part of the alliance. These countries are territorially close to the area (e.g., neighbours) where the squad of drones is patrolling. The first country that is providing some of the ground support is country N_1, and it has good relations with the alliance countries involved in the mission. The other country that is providing help to the mission is country N_2. The mission takes place in both N_1 and N_2 territory, as well as other neutral and hostile territories. Country X and Y, of the $\mathcal{A}$ alliance, together with N_1 and N_2 are part of a coalition during the mission, the patrolling of X's squad of drones in the above mentioned territories. Though N_1 and N_2 are both participating in the mission, they have poor relations with each other. Hence information concerning country N_1 must not be shared with or used by country N_2, and vice versa.

The information gathered by the squad of drones is shared and used by all the members of this military mission. The sharing of information can happen in different ways: directly by the drones, via satellite communication, cloud platforms etc. The rules that describe the data access/usage/sharing should take into account the role of the entities asking to access the data with respect to the mission and the entity that collected the data, as well as the relations between the parties involved in the mission. The above relations influence what part of the data can be accessed and by whom. Therefore, different types of alteration depending on who is accessing the data need to be performed.

B. The set of policies for our use case

The squad of drones of our use case gathers information, that is made available, via different platforms, to be accessed

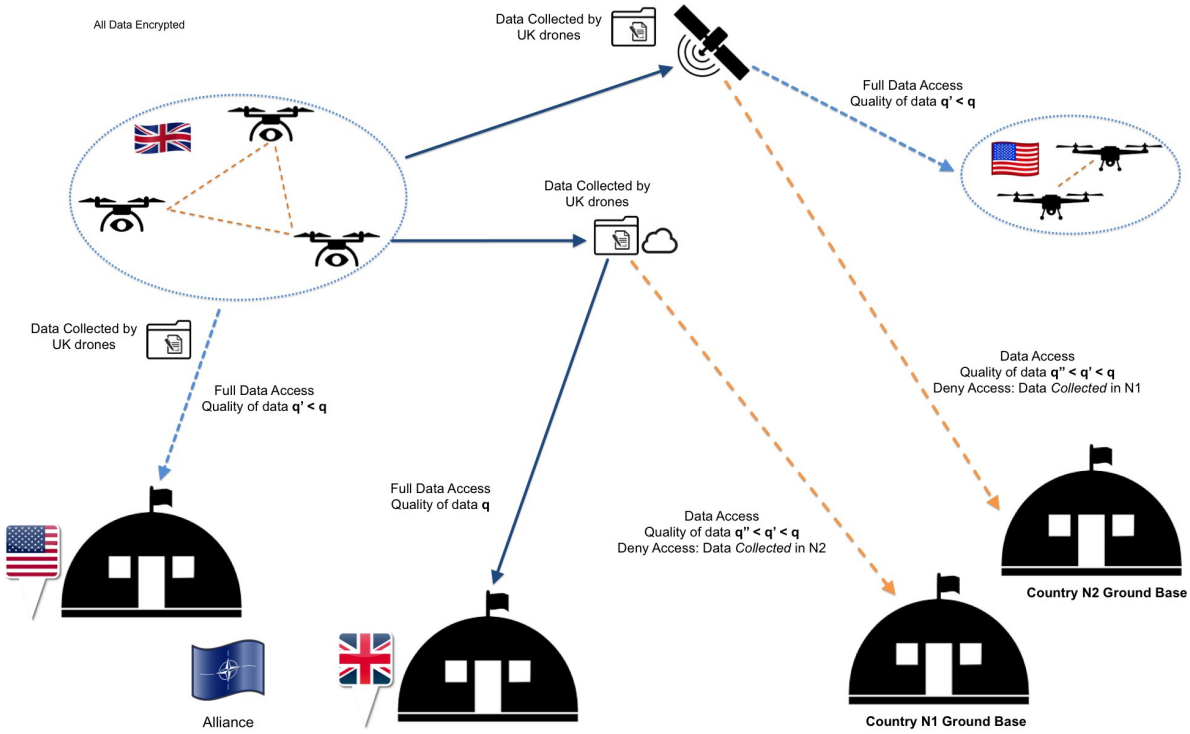

Figure 1. A schematic representation of the military scenario. We assumed that country X is UK, country Y is US and the alliance is the NATO alliance.

by other countries involved in the mission. We present below some of the policies that describe how the data of our use case scenario can be accessed and used, a schematic representation is given in Figure 1, where for the sake of understandability we assumed that country X is UK, country Y is US and the alliance is the NATO alliance, while the other countries are left as N_1 and N_2.

1) The country that owns the drones has full access to the collected data.
2) The country that is part of the mission as an ally has full access to the collected data, that are slightly degraded.
3) The country that is part of the mission as a member of the coalition[3], has partial access to the collected data, that are moderately degraded.
4) Country N_1 cannot access the data collected in N_2 territory.
5) Country N_2 cannot access the data collected in N_1 territory.

In practice these policies would be expressed in either a formal, or a controlled natural language but the intent of this paper is to explore the underlying issues rather than propose a specific policy notation.

To represent the above policies we need a number of new predicates, that capture the relations between the different entities involved, as well the connection between coordinates and countries.

- $ally(A, B)$ denotes that A and B are allies
- $partner(A, B)$ denotes that A and B are in the same coalition
- $in(C, (r, s))$ denotes that coordinates (r, s) are located in country C

Given an item $Item_{q,r,s}$ of data collected by country A, then country B can access the restricted item $Item_{q',r,s}$ where q' is the restricted item of the data altered with respect to the given policies. The policy represents the restrictions implied by the conditions below that represent the set of policies introduced above.

Policy 1	$same(A, B)$	$q' = q$
Policy 2	$ally(A, B)$	$Q_1 \leq q' < q$
Policy 3	$partner(A, B)$	$Q_2 \leq q' < Q_1$
Policy 4	$same(B, N_1)$, $same(C, N_2)$	$q' = 0$
Policy 5	$same(B, N_2)$, $same(C, N_1)$	$q' = 0$

Policy 1 represents the rights of the country that collects the data, to access them without any alteration[4]. Policy 2 represents that the quality of the data accessed by ally countries is altered, where a specific limitation of the alteration is given and it cannot be lower than Q_1. The value where q' ranges depends on the relations of trust we want to describe

[3]We call *partners* the members of the same coalition.

[4]The predicate $same(A, B)$ means that A and B represent the same entity, in our case country.

with our policies. In our scenario, we decided for a slight change for allies, hence the quality of their data q' is $q' < q$. Policy 3 represents that the quality of the data accessed by partners is altered. The alteration is stronger than the one made to the data accessed by ally countries, as the quality of the data accessible by the partners is strictly less than the allies lower bound quality threshold. Also in this case a specific limitation to the alteration is given, Q_2[5]. Policy 4 and 5 represent respectively that country N_1 cannot access data collected in country N_2 territory and country N_2 cannot access data collected in country N_1 territory.

C. Data quality restriction using the DSAs

The types and quality of data that can be accessed/shared depends on the relations the entities have between each other. In our scenario, it depends by the level of trust and partnership. Our policy framework permits the representation of this relation and the restriction made to the data, by using the DSAs.

A dataset is the entity $AData$, where we describe the collected predicate as below:

$$holdsAt(collect(AData, D_i), T').$$

In our scenaio, the restriction from source to delivered dataset is applied on the basis of both the physical location to which the data relates, and to the level of detail contained in the collected data. To apply the criteria, the dataset must first be divided into subsets which are homogeneous from the quality restriction perspective. Using the $Item$ notation above, where each $Item$ represents a single datum, then we might define the data as a set of *regions*:

$$R_{q,N_1} = \{Item_{a,b,c} : a = q, (b,c) \in N_1\}$$
$$R_{q,N_2} = \{Item_{a,b,c} : a = q, (b,c) \in N_2\}$$
$$R_q = \{Item_{a,b,c} : a = q, (b,c) \notin N_1 \cup N_2\}$$

hence

$$Data = \bigcup_{0 \le q \le Q} (R_{q,N_1} \cup R_{q,N_2} \cup R_q).$$

We represent data inside one of the regions by: $R_{q,I}$, that can be represented as $R_{q,I} = \{Item_{q,r,s}\}$, where (r,s) are generic coordinates. Therefore, the predicate

$$restrict(A, B, AData, BData)$$

described in Section III-C is defined by the following predicates.

$$restrict(A, B, R_{q,I}, R_{q,I}) :- \quad same(A, B). \tag{1}$$

$$restrict(A, B, R_{q,I}, R_{q',I}) :- \quad ally(A, B), \; q' < q, \; Q_1 \le q'. \tag{2}$$

$$restrict(A, B, R_{q,I}, R_{q',I}) :- \quad partner(A, B), \; q' < q, \; Q_2 \le q' < Q_1. \tag{3}$$

$$restrict(A, B, R_{q,N_2}, R_{q',N_2}) :- \quad partner(A, B), \; same(B, N_1), \; q' = 0. \tag{4}$$

$$restrict(A, B, R_{q,N_1}, R_{q',N_1}) :- \quad partner(A, B), \; same(B, N_2), \; q' = 0. \tag{5}$$

In the described predicates, we have a contradiction between rules (3) and rule (4) and (5). Thus, we can substitute rule (3) with the following rules.

$$restrict(A, B, R_q, R_{q'}) :- \quad partner(A, B), \; q' < q, \; Q_2 \le q' < Q_1. \tag{6}$$

$$restrict(A, B, R_{q,N_1}, R_{q',N_1}) :- \quad partner(A, B), \; \textbf{not } same(B, N_2), \; q' < q, \; Q_2 \le q' < Q_1. \tag{7}$$

$$restrict(A, B, R_{q,N_2}, R_{q',N_2}) :- \quad partner(A, B), \; \textbf{not } same(B, N_1), \; q' < q, \; Q_2 \le q' < Q_1. \tag{8}$$

The above conflicts can also be solved without introducing the above rules (6-8), but by introducing priorities between rules. The priorities introduced for our scenario state that for the case of respectively having country N_1 and N_2 that requests the data: rule (4) has higher priority than rule (3), and rule (5) has higher priority than rule (3). The priorities are denoted as: rule (4) $>$ rule (3), and rule (5) $>$ rule (3)[6].

V. Discussions

The policy language, through its high expressive power, represents the various rules and constraints that should be applied during the access, usage and sharing of data.

The already existing policy analysis [18], [23] that is based on argumentation reasoning together with conflict resolution makes the introduced techniques ideal for representing complex legislative rules for data rich environment.

In this work we focused mainly on the alteration made to the data quality that reflects the relations of trust and partnerships between different entities. With the increasing use of big data, the data themselves have a value. The work in [19] introduces a first use of data sharing agreements during the data processing, with a special focus on data quality attributes. Being able to diversify the data quality depending on the potential data buyers, by ensuring the

[5]Depending on the relations between entities that we want to describe, the value of Q_2 can vary. In case we want to have the possibility of not giving any data to our partners, then $Q_2 = 0$. This last case contradicts the relations of partnership between countries involved in the same coalition during the accomplishment of a mission, and it is not the case we are dealing with in our scenario.

[6]For a more detailed overview of how to solve the conflicts between policies/rules by introducing priorities between them, we direct the reader to [18].

security requirements of data access and usage control would bring great benefits to the data market.

Until now, when describing the data collection by the drone systems, we have been discussing a simple geographical model with two coordinates. While collecting the data the drones can also be at different altitudes, which also affect the quality and the type of collected data. It would be interesting in the future to understand how this third component interacts with the data quality. In many cases the higher the altitude the lower the quality of images as well as audio data. Thus, depending on the type of collected data, this component would play a crucial role on the correct data restriction.

Another interesting property that should be taken into consideration is the *type* of data we are sharing, e.g., security type (low, medium, high), private data. The type of information together with the context of use plays an important role. It is already the case that some documents intended for paper distribution already have individual pages, or even paragraphs, marked with a security classification. The 'aggregate' classification of a page, or the complete document, is set to the highest value which appears. Our proposed DSA approach gives the option of supplying such data to a wide audience without compromising security. In our scenario, we could decide to share for example all high security data with ally countries, and just medium and low security data with the partners. In [18], [26] are introduced examples of how the used policy language represents DSAs that describe context dependent sharing of different types of security data. Depending on the type of data and the relations between the entities we could apply a data quality alteration.

VI. Conclusions

In this work we introduce a policy mechanism that improves data sharing in data rich environments. The quality of the shared data and which parts of them are shared depend on the data recipient, its relations, level of trust and/or the type of membership/partnership. Data quality alteration is made to the shared data.

In this work we introduce a policy based mechanism that alters the quality of shared data depending on the context of use. The alteration reflects the level of trust/relations/membership/partnership of the recipient of the data. We model this alteration mechanism with the use of a policy language, that represents the various relations and consequently the restrictions that are made to the data quality. The policies are part of the data sharing agreements between the different entities, and are attached to the data. The policies are enforced when requests are received and permissions to access/share/use the data are evaluated. We show how our restriction mechanism works in a military scenario, where drone systems and entities with different relations of partnerships are involved.

Acknowledgments

Supported by EPSRC Project CIPART grant no. EP/L022729/1. This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Number W911NF-16-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

References

[1] M. C. Mont, S. Pearson, and P. Bramhall, "Towards accountable management of identity and privacy: sticky policies and enforceable tracing services," in *14th International Workshop on Database and Expert Systems Applications, 2003. Proceedings.* IEEE Computer Society, 2003, pp. 377–382.

[2] J. Bayuk, "Data-centric security," *Computer Fraud & Security*, vol. 2009, no. 3, pp. 7–11, 2009.

[3] Y.-J. Kim, M. Thottan, V. Kolesnikov, and W. Lee, "A secure decentralized data-centric information infrastructure for smart grid," *IEEE Communications Magazine*, vol. 48, no. 11, pp. 58–65, 2010.

[4] M. C. Mont and S. Pearson, "Sticky policies: An approach for managing privacy across multiple parties," *Computer*, vol. 44, pp. 60–68, 2011.

[5] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for data storage security in cloud computing," in *INFOCOM, 2010 Proceedings IEEE*. IEEE, 2010, pp. 1–9.

[6] W. Zhou, M. Sherr, W. R. Marczak, Z. Zhang, T. Tao, B. T. Loo, and I. Lee, "Towards a data-centric view of cloud security," in *Proceedings of the Second International Workshop on Cloud Data Management*, ser. CloudDB '10. ACM, 2010, pp. 25–32.

[7] M. Gertz and S. Jajodia, *Handbook of database security: applications and trends*. Springer Science & Business Media, 2007.

[8] E. Karafili, H. R. Nielson, and F. Nielson, "How to trust the re-use of data," in *Security and Trust Management - 11th International Workshop, STM*. Springer, 2015, pp. 72–88.

[9] C. Kaufman, R. Perlman, and M. Speciner, *Network Security: Private Communication in a Public World, Second Edition*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall Press, 2002.

[10] A. Lazouski, F. Martinelli, and P. Mori, "A prototype for enforcing usage control policies based on XACML," in *Trust, Privacy and Security in Digital Business - 9th International Conference, TrustBus*. Springer, 2012, pp. 79–92.

[11] D. F. Ferraiolo and D. R. Kuhn, "Role-based access controls," in *15th National Computer Security Conference*, 1992.

[12] J. Park and R. S. Sandhu, "Towards usage control models: beyond traditional access control," in *SACMAT*. ACM, 2002, pp. 57–64.

[13] ——, "The ucon_{abc} usage control model," *ACM Trans. Inf. Syst. Secur.*, vol. 7, no. 1, pp. 128–174, 2004.

[14] F. Kelbert and A. Pretschner, "Data usage control enforcement in distributed systems," in *Third ACM Conference on Data and Application Security and Privacy, CODASPY'13*. ACM, 2013, pp. 71–82.

[15] G. Karjoth, M. Schunter, and M. Waidner, "Platform for enterprise privacy practices: Privacy-enabled management of customer data," in *Privacy Enhancing Technologies: Second International Workshop, (PET)*, R. Dingledine and P. Syverson, Eds. Springer, 2003, pp. 69–84.

[16] G. Karjoth and M. Schunter, "A privacy policy model for enterprises," in *Proceedings 15th IEEE Computer Security Foundations Workshop (CSFW-15)*. IEEE Computer Society, 2002, pp. 271–281.

[17] V. Swarup, L. Seligman, and A. Rosenthal, "Specifying data sharing agreements," in *7th IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY)*. IEEE Computer Society, 2006, pp. 157–162.

[18] E. Karafili and E. C. Lupu, "Enabling data sharing in contextual environments: Policy representation and analysis," in *SACMAT '17*. ACM, 2017, pp. 231–238.

[19] E. Karafili, K. Spanaki, and E. C. Lupu, "An argumentation reasoning approach for data processing," *Computers in Industry*, vol. 94, no. Supplement C, pp. 52 – 61, 2018.

[20] K. Shanmugapriya, M. Murugeswari, and K. Suriya, "Security issues associated with big data in cloud computing," *IJCSIT International Journal of Computer Science and Information Technologies*, vol. 6, no. 6, pp. 4952–4956, 2015.

[21] A. E. Youssef, "A framework for secure healthcare systems based on big data analytics in mobile cloud computing environments," *Int J Ambient Syst Appl*, vol. 2, no. 2, pp. 1–11, 2014.

[22] J.-J. Yang, J.-Q. Li, and Y. Niu, "A hybrid solution for privacy preserving medical data sharing in the cloud environment," *Future Generation Computer Systems*, vol. 43, pp. 74–86, 2015.

[23] E. Karafili, A. Kakas, N. Spanoudakis, and E. Lupu, "Argumentation-based security for social good," in *AAAI Fall Symposium Series*, 2017. [Online]. Available: https://aaai.org/ocs/index.php/FSS/FSS17/paper/view/15928/15306

[24] R. Craven, J. Lobo, J. Ma, A. Russo, E. C. Lupu, and A. K. Bandara, "Expressive policy analysis with enhanced system dynamicity," in *Proceedings of the 2009 ACM Symposium on Information, Computer and Communications Security, ASIACCS*. ACM, 2009, pp. 239–250.

[25] R. A. Kowalski and M. J. Sergot, "A logic-based calculus of events," *New Generation Comput.*, vol. 4, no. 1, pp. 67–95, 1986.

[26] E. Karafili, E. C. Lupu, S. Arunkumar, and E. Bertino, "Argumentation-based policy analysis for drone systems," in *IEEE Smart World Congress, DAIS Workshop*, 2017.

[27] E. T. Mueller, *Commonsense Reasoning: An Event Calculus Based Approach*, 2nd ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2015.

APPENDIX

We extend our policy language with the predicate

$$restrict(Owner, Recipient, RawData, OutData).$$

Ideally we would have a function:

$$OutData = restrict(Owner, Recipient, RawData)$$

rather than a predicate. The introduced function is not supported by event calculus. A related logic, *situation calculus*, does support functions in this way and it can be shown that the two calculi are equivalent [27, chapter 16]. In particular, given a function $F(x) = y$ and a predicate $F(x, y)$, then the desired function can be represented provided the predicate meets the constraints:

$$holdsAt(F(x, y_1), t) \wedge holdsAt(F(x, y_2), t) \implies y_1 = y_2$$
$$\exists y.\ holdsAt(F(x, y), t)$$

The introduced $restrict$ predicate needs to satisfy the above constraint.