A Vision for Privacy and Transparency through Policy-Carrying Data

Julian Padget
Dept. of Computer Science
University of Bath
Bath, BA2 7AY, U.K.
j.a.padget@bath.ac.uk

Wamberto W. Vasconcelos
Dept. of Computing Science
University of Aberdeen
Aberdeen, AB24 3LT, U.K.
w.w.vasconcelos@abdn.ac.uk

ABSTRACT

As the customer has become the product, so (individual) privacy has become the currency: it is traded for access to resources and paid for in the revelation of many small facts and preferences that are just data on their own, but become information as part of larger population. While fiat currency is traceless and untraceable – it is generally impossible to tell where it has come from or where it has gone – individual data can sometimes be traced back to its origin(ator), but can often be untraceable, once passed on. The concerns expressed at various levels of society are effectively the product of (i) the proliferation of individually attributable (digitally represented) data that can be seen somehow belonging to and having the potential to compromise the individual or be of benefit to a third party and (ii) a regulatory framework that is unprepared for the volume, variety, (or velocity and veracity) of data. In this context, transparency is being proposed as the means to trade off privacy while maintaining security. In response, we propose the notion of data that is inseparable from its access policy and furthermore that any derived data shall also have a suitable derived policy inextricably associated with it. We sketch a framework and some formal notions to capture these ideas to initiate debate.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Policies, data sharing, internet-of-things

Keywords

AAMAS proceedings, LaTeX, text tagging

1. INTRODUCTION

We propose a means to capture the expression of controls over (derived) data and to associate that indivisibly with the data via what we call "policy-carrying data" (PCD[1]). Following policy conventions, in defining the who, the when and the how, the PCD establishes permissions for what the data consumer may do to the data. A novel aspect of our proposal is the establishment of obligations concerning what the consumer should do with the (derived) data and it is these that are the foundation of transparency. Such obligations are the transactional unit for a non-pecuniary data economy, where access to and use of data may be traded for obligations that act as a form of user-definable, liquidity at-point-of-use community currency [6]. These obligations may pertain directly to actions of data consumers or – and this we believe is also a significant novelty of our proposal – indirectly to the policy associated either with the extracted data or the data derived from them.

A feature of the current data landscape is the relative freedom of movement of data from individuals to the data silos used in cloud computing and thence between silos, which could be viewed as contributary to the disempowerment that individuals might feel over their own data – privacy controls aside [1]. The situation is potentially further complicated since the platform may enable the collection and interpretation of those data, thus adding value to them, as in the case of activity-monitoring devices or home energy monitors. The PCD concept associates data with bespoke policies: for example, framework policies might be defined by legislation, while specific policies for individual needs would have to satisfy the norms established at the primary level [5]. In this way, crisp but unworkable definitions of issues such as "When do data stop being private?" and "How to decide if data revelation is in the public interest?" can be blurred as distinctions are established to meet the needs of a given situation.

In the next section we set out a framework for policy-carrying data that considers the information model, identifies classes of stakeholders and puts forward some illustrative examples of natural language expressions of policy (mapping from natural language to formal is a challenging problem for future work). This is followed by a short discussion on the representation of policy and how our perspective draws upon the significant body of work relating to norms and their formalisation, stemming from logic and from multiagent systems. Consequently, we illustrate a possible reasoning framework and its use of state to maintain an audit trail of consumer actions. We conclude with a short discussion of some related work and directions for future work.

2. A FRAMEWORK FOR PCD

We set out a reference framework within which we situate and connect stakeholders, PCDs and an information model. We illustrate our framework in Fig. 1, where we show stakeholders (circles), processes (arrows) and information model (boxes within central box). The stakeholders envisaged are (i) data owners/producers who make data/information available (represented as the left-hand circle) and, in a richer version of the model, these may be separated into those that assert rights over the data and those that publish it; (ii) data consumers who want to access data (represented as the right-hand circle) and again in a richer version of the model, there may be entities that are both consumers and producers of data, either by offering aggregation services or by adding value in some way; (iii) monitors responsible for policing the publication and access activities (upper circle in the middle).

We note that the first two types of stakeholders can be institutions or people as well as computational entities such as sensors,

[1] PCD also stands for "policy-carrying data collection" and we use PCDs (in the plural) to indicate a set of policy-carrying data collections.

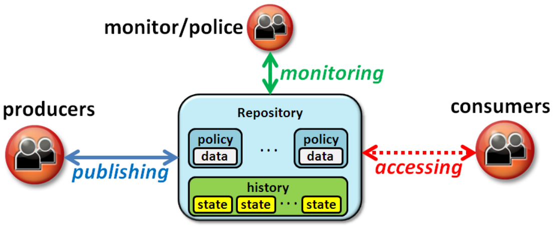

Figure 1: A Framework for Policy-carrying Data

programs, databases, and so on. A monitor works as a third-party authority ensuring that activities (publishing and accessing) follow policies and dealing with violations. Each of these stakeholders has specific ways to interact via the repository: (i) publishing (represented by the blue solid line) is the process whereby data owners/producers make their data available but "wrapped" within a policy, that is, they publish, in a repository, some policy-carrying data. (ii) accessing (represented by the red dotted lines) is the process whereby data consumers *attempt* to obtain access to data mediated via policies. (iii) monitoring (represented by the green line) concerns observing activities and checking for policy compliance or violation, and dispensing rewards or sanctions.

Our framework relies on an information model comprising the PCD – a policy and a data collection made available through the policy – and a history – a collection of events (*i.e.*, a record of activities carried out) gathered at particular time points, denoted as the *states* of the repository. The framework would support stakeholders carrying out the cycle of publish-access-monitor activities supported by a Web server. Servers should be equipped with functionalities to enable the policing of those accessing and uploading PCD, keeping records of usage and (non-)compliance, and enforcing the policies' access control. We envisage programmatic access to PCD, whereby programs and functionalities developed with specific technologies can access any PCD, interacting with the server via pre-establised protocols.

A typical PCD would express something like "Lab managers can access 500 records of my data". If an interested party requested 1,000 records, the server would (i) check the credentials of the requester (who needs to be registered); (ii) grant access to 500 records (a message would provide reasons for not providing the 1,000 records); (iii) update the record of that requester with respect to that PCD. Further requests from the same party would be rejected with a suitable justification. For such control to be in place, the server requires a record of events – an explicit account of the history of the PCD, how they have been used, by whom and when. We observe that PCD can also be used as a means to "wrap" a sensor or other data source, including whole sensor networks. We can have, for instance, a policy establishing that "anyone is permitted to request the temperature reading of the sensor once every hour".

3. POLICY REPRESENTION AND REASONING

Much research has been carried out on data access policies since the early UNIX file systems [9]. Some notable features our approach are as follows. We include means to refer to a history of events; examples are "the first 10 people can use my data" and "anyone is permitted to use n records of my data". We provide fine-grained control over who is to access the data, and under what circumstances; for instance, "invididual i_{285} is forbidden to access my data" and "anyone from company x may use my data after 6PM". Additionally we capture dynamic aspects of data usage, examples being "whoever accesses D_1 should not access D_2" and "anyone who uses my data should provide data".

We adapt and extend current work on normative reasoning for multi-agent systems [2, 3] to represent roles (of participants), data-related events (such as accessing records or publishing data collections), authorship of events and attempts thereof, activation and deactivation conditions of policies, and the object of the policy, namely, the data collection itself. An example policy is:

$$\langle \overbrace{\langle \underbrace{\{\neg accessAll(D_1)\}}_{activation}, \underbrace{\{accessAll(D_1)\}}_{deactivation}, \underbrace{\mathsf{P}_{all}\, accessAll(D_1)}_{target} \rangle}^{policy}, \overbrace{D_1}^{data} \rangle$$

This captures permission to access all records of a data collection. An activation condition says that the permission is in place if the records have not yet been accessed, and the policy is deactivated when the records are accessed, thus providing "one-off" access to the data. The *all* role says anyone can use this policy.

Complementary to the informal discussion of policy examples and and requirements above, we have developed algorithmic specifications of three mechanisms to enable stakeholders to reason with and about their PCDs, so that: (i) a PCD publisher can obtain the identity of individual agents who have access (via their associated roles) to data collections. (ii) an agent a can analyse a set of PCDs and gather all the obligations which might apply to it (iii) access can be policed, through a language comprising permissions and prohibitions (that are use as a means for temporary derogation of permissions), and also logged. The formalization of the model and the details of the above algorithms appear in [7]. Meanwhile a detailed comparison with [4] and [8] is in preparation.

REFERENCES

[1] L. Brandimarte, A. Acquisti, and G. Loewenstein. Misplaced confidences: Privacy and the control paradox. *Social Psychological and Personality Science*, 4(3):340–347, 2013.

[2] M. Şensoy, T. J. Norman, W. W. Vasconcelos, and K. Sycara. OWL-POLAR: A framework for semantic policy representation and reasoning. *Web Semantics: Science, Services and Agents on the World Wide Web*, 12-13, 2012.

[3] A. García-Camino, J. A. Rodríguez-Aguilar, C. Sierra, and W. W. Vasconcelos. Constraint rule-based programming of norms for electronic institutions. *Autonomous Agents and Multi-Agent Systems*, 18(1):186–217, 2009.

[4] G. Karjoth, M. Schunter, and M. Waidner. Platform for enterprise privacy practices: Privacy-enabled management of customer data. In R. Dingledine and P. F. Syverson, editors, *Privacy Enhancing Technologies, PET 2002, Revised Papers*, volume 2482 of *LNCS*, pages 69–84. Springer, 2002.

[5] T. Li, T. Balke, M. De Vos, J. A. Padget, and K. Satoh. A model-based approach to the automatic revision of secondary legislation. In *International Conference on Artificial Intelligence and Law*. ACM, 2013.

[6] B. Litaer. *The Future of Money: Creating New Wealth, Work and a Wiser World*. Century, 2002.

[7] J. Padget and W. Vasconcelos. Policy-carrying data: A step towards transparent data sharing. In *Proceedings of 6th International Conference on Ambient Systems, Networks and Technologies, ANT 2015*. Elsevier, June 2015. In press.

[8] S. Pearson and M. Casassa Mont. Sticky policies: An approach for managing privacy across multiple parties. *IEEE Computer*, 44(9):60–68, 2011.

[9] V. Suhendra. A survey on access control deployment. In *Security Technol.*, volume 259 of *Comm. in Comp. & Inf. Science*. Springer, 2011.