Jens Leicht*, Armin Gerl, and Maritta Heisel

Technical Report on the Extension of the Layered Privacy Language

Abstract: In this technical report, we list all changes we make in our extension of the Layered Privacy Language (LPL). The changes focus on the improvement of the LPL's compatibility with the General Data Protection Regulation (GDPR). We address drawbacks of the current version of LPL, which are concerned with the definition of legal bases for data processing; the declaration of official authority vested in the service provider; missing information regarding adequacy decisions for third country transfers; and drawbacks regarding data minimization in LPL. In addition to addressing these drawbacks, we implement a purpose hierarchy, which improves the general structure of policies written in LPL and supports partially automatically generated privacy policies. We also add fixed categories for purposes as well as data categories that improve the semantics of the language with respect to formal reasoning about LPL policies.

1 Introduction

This document summarizes our changes to the Layered Privacy Language (LPL) by Gerl et al. [1]. We do not provide any reasoning for these changes, since this report is only used as a reference to the changes. The reasoning about, as well as more detailed explanations of the changes will be published in a separate paper.

The changes focus on the improvement of the LPL's compatibility with the General Data Protection Regulation (GDPR). Furthermore, we implement an improved form of a purpose hierarchy, as originally proposed by Hjerppe et al. [2]. This improves the general structure of policies written in LPL and supports partially automatically generated privacy policies. To improve the semantics of LPL and prepare for future formalization of LPL and its policies, we add fixed categories for purposes as well as data categories. This allows the formalization of legislation like the GDPR, which contains specific rules for different purposes and different categories of data.

The structure of LPL was iteratively developed in the works of Gerl et al. [1], Gerl and Pohl [3], Gerl [4], and Gerl and Bölz [5] and completely detailed in Gerl [6]. All changes we make in our extension of LPL are listed in the following section.

2 LPL Extension

In this section we present our extension of the Layered Privacy Language. First, we present changes to existing elements of LPL, followed by the implementation of purpose hierarchies. Finally, we present the fixed purpose and data categories. Information about the version of LPL, that we use as a basis for our changes, can be found in Gerl [6].

***Corresponding Author: Jens Leicht:** paluno - The Ruhr Institute for Software Technology, University of Duisburg-Essen, Germany, E-mail: jens.leicht@uni-due.de
Armin Gerl: Chair of Distributed Information Systems, University of Passau, Passau, Germany, E-mail: Armin.Gerl@uni-passau.de
Maritta Heisel: paluno - The Ruhr Institute for Software Technology, University of Duisburg-Essen, Germany, E-mail: maritta.heisel@uni-due.de

2.1 Legal Bases

Although LPL already contains a way of representing the legal bases of processing purposes, as can be seen in the original LPL tuple lb_o below, we suggest to improve the semantics of LPL by providing fixed types of legal bases.

$$lb_o = (name, \widehat{HEAD}, \widehat{DESC})$$

We name the attribute containing this type-information $lbCategory$. Replacing the $name$ attribute in the legalBasis-element with the $lbCategory$, as can be seen in the new legalBasis-element lb_n below, allows interpretation of the provided legal bases without the need to process the textual description of the legal basis.

$$lb_n = (lbCategory, \widehat{HEAD}, \widehat{DESC})$$

The $lbCategory$ attribute is an enumeration element and must be one of the elements of the set of all types of legal bases given by $\widehat{lbCategory}$.

$$lbCategory \in \widehat{lbCategory}$$

From the GDPR [7, Article 6 §1 (b) to (f)] we derived the following set of legal basis types, from which the policy author can choose:

$$\widehat{lbCategory} = \{'consent', 'contract', 'legalObligation', 'vitalInterest', 'publicTask', 'legitimateInterest'\}$$

Using the header and description attributes of the legal basis, the policy author can provide additional information to the data subject to improve understanding of the legal bases used for a given purpose. Using these legal basis types, it is possible to improve the visualization of privacy policies using privacy icons, like the ones proposed by Gerl [8], to visualize the legal bases for a given purpose.

2.2 Official Authority

To be able to express the fact that processing is performed in official authority vested in the data controller we amend the $classification$ enumeration. The $classification$ attribute is used in the dataRecipient-element.

The dataRecipient-element inherits the $classification$ attribute from the entity-element. Originally, the $classification$ attribute could take one of two values: 'Person' and 'Legal Entity'. We add the classification 'Public Authority' marked in **bold** below.

$$classification \in \{'Person', 'LegalEntity', '\boldsymbol{Public\ Authority}'\}$$

By adding this new classification at the entity level we allow other entities, such as the data source or the data controller, to be classified as public authorities. This may not be required by the GDPR, but it may support the data subject in making a more informed decision for a given privacy policy.

2.3 Adequacy Decisions

In the original *dataRecipient*-element there is no information about the destination of the transfer, as well as a possible adequacy decision of the European Commission.

In the extended $dataRecipient_n$, which can be seen below, we add two new attributes regarding third country transfers. Since both attributes are concerned with third country transfers we add the *country* and *adequacyDecision* attributes right after the $thirdCountryTransfer$ flag.

$$dataRecipient_n = (name, classification, authInfo,' DataRecipient', required, thirdCountryTransfer, \mathbf{country}, \mathbf{adequacyDecision}, \widehat{HEAD}, \widehat{DESC}, \widehat{SG})$$

The *country* attribute has to be provided as two letter country code adhering to the ISO 3166-1 alpha-2 standard [9], e.g. $'US'$. Since the dataRecipient-element is also used for non-third country data transfers, the *country* attribute can also be left empty $'\ '$. If the $thirdCountryTransfer$ flag is *true*, the *country* attribute must not be empty.

Depending on the destination country of the third country transfer the *adequacyDecision* attribute of type Boolean should be set. If no third country transfer takes place the *adequacyDecision* flag should be set to *false*. If the European Commission has decided that the destination country provides adequate data protection the *adequacyDecision* flag should be set to *true*. Information regarding current adequacy decisions of the European Commission can be found online[1].

2.4 Data Minimization

LPL stores personal data inside the privacy policy when an anonymization method is used. To improve on this lack of data protection, we suggest to only store the selected anonymization level, together with coarser levels, that contain even less information.

Below you can see the improvement applied to the example used in Gerl [6]. The example $\widehat{HE}_{new}$ shows the anonymization of the German postal code *94032*. LPL originally stores all five levels of anonymization starting from *94032* to *******. With our improvement in an anonymization level of 3 only the values of level 3 and above are stored, providing better data protection, since no information with less anonymization is included.

$$\widehat{HE}_{new} = \{('94***'), ('9****'), ('*****')\}$$

2.5 Raw Policy

In addition, we suggest an improvement in the usage of the underlying privacy policy element (*upp*), which is the last attribute of the root policy element *lpp*, as can be seen in equation (1) below. Normally this element is used for the combination of different privacy policies, especially in the case of data processors. When the data is handed over to the processor, the service level agreement between the data controller and the data processor is combined with the customized privacy policy of the data subject, whose data is going to be shared with the processor.

$$lpp_o = (version, name, lang, ppURI, \widehat{HEAD}, \widehat{DESC}, \widehat{I}, ds, \widehat{P}, \widehat{C}, \widehat{DPO}, dsr, lc, upp) \tag{1}$$

We suggest to also use this element to store the raw policy, which is provided by the data controller to the data subject. The raw policy is not instantiated with any user data or timestamps and contains all possible purposes of data processing. The final privacy policy, which was adjusted by the user, does not contain all purposes, only the ones that were accepted by the user. Thus, by default it is difficult to see what other purposes were available to the user. However, when storing the original raw policy inside the agreed policy of the data subject, it is possible to identify purposes that were not accepted by the user.

This is useful for user interfaces, where it may be beneficial to allow users to browse purposes that were not consented previously. It allows users to rethink their decisions and consent to some purpose later on in the usage of the service.

[1] https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en

2.6 Privacy Models

To prepare LPL for the addition of a purpose hierarchy, which we present in Section 2.7, we propose a change concerning privacy models. Privacy models can for example be k-anonymity or l-diversity.

LPL currently defines privacy models for each purpose, meaning that all data of a purpose are considered by the provided privacy models. pm_o below shows the original tuple of a privacy model element.

$$pm_o = (name, \widehat{PMA}, \widehat{HEAD}, \widehat{DESC})$$

To adapt the privacy model element to the pseudonymization method element (psm) we add a set of nameOfData-elements $\widehat{NOD}$ to the privacy model, as can be seen in the new tuple (pm_n) below. For further information regarding the pseudonymization method element see Gerl et al. [6].

$$pm_n = (name, \widehat{NOD}, \widehat{PMA}, \widehat{HEAD}, \widehat{DESC})$$

The set $\widehat{NOD}$ contains nod elements, which reference a specific data element of the given purpose by its name.

2.7 Purpose Hierarchy

Hjerppe et al. [2], proposed a fundamental extension of LPL, allowing the hierarchical composition of purposes. This purpose hierarchy allows the creation of high-level purposes, that provide an overview over the data collected and the data recipients of this data. These high-level purposes can link to more detailed sub-purposes. We adapt the purpose hierarchy to our extended version of LPL, as described in the following.

As an extension of the purpose hierarchy we add some fixed purposes, which are used to categorize the purposes of a policy. Further details regarding the fixed purposes are given in Section 2.8 below. Figure 1 shows the structure of an exemplary purpose hierarchy. Purposes p_1 and p_2 are instances of a given purpose. There can be many instances of a single given purpose, but each purpose in the hierarchy can only have a single parent purpose. Each non-given purpose can have either zero or at least two sub-purposes. Figure 1 also illustrates the fact that the hierarchy can have multiple layers, where a sub-purpose can be constructed from further sub-purposes ($p_3 \rightarrow \; p_5, \dots, p_6$).

The first change we make to the purpose hierarchy of Hjerppe et al., is concerned with the set of tuples that is used to describe the hierarchy. **PH** is a set of tuples describing sub-purpose relationships, e.g.:

$$\mathbf{PH} = \{\dots, (p_2, p_3), (p_2, p_4), \dots\}$$

where p_3 and p_4 are sub-purposes of p_2 (cf. Fig. 1). We place the set **PH** inside the main policy tuple lpp, behind the set of purposes $\widehat{P}$, see lpp_n below. The original lpp tuple can be seen in equation (1). Integrating the hierarchy in the main policy tuple assures that the completeness of the policy is always preserved. Storing the hierarchy in a separate set, that is not part of the policy, could lead to the corruption of a policy in case of partial copies.

$$lpp_n = (version, name, lang, ppURI, \widehat{HEAD}, \widehat{DESC}, \widehat{I}, ds, \widehat{P}, \mathbf{PH}, \widehat{C}, \widehat{DPO}, dsr, lc, upp)$$

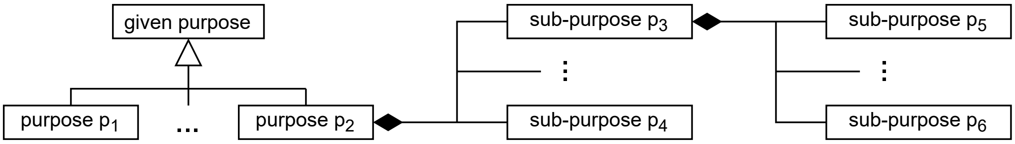

Fig. 1. Example structure of a purpose hierarchy

For the sub-purpose relation we define the following conditions, where elements marked with the subscript s refer to the sub-purpose and p to the super-purpose. These elements are part of the purpose tuple.

$$optOut_s = optOut_p \tag{2}$$

$$required_s \leq required_p \tag{3}$$

$$pointOfAcceptance_s \geq pointOfAcceptance_p \tag{4}$$

$$\widehat{D_s} \subseteq \widehat{D_p} \tag{5}$$

$$\widehat{PM_s} \subseteq_{pm} \widehat{PM_p} \tag{6}$$

$$\widehat{PSM_s} \subseteq \widehat{PSM_p} \tag{7}$$

$$\widehat{DR_s} \subseteq \widehat{DR_p} \tag{8}$$

$$\widehat{LB_s} \subseteq_{lb} \widehat{LB_p} \tag{9}$$

$$\widehat{ADM_s} \subseteq_{adm} \widehat{ADM_p} \tag{10}$$

$$r_s \leq_r r_p \tag{11}$$

The *optOut* flag needs to be the same for sub-purposes (2). This is necessary because opt-in and opt-out purposes cannot be aggregated in the same super-purpose. Instead, these must be aggregated in separate purposes with corresponding opt-out configurations.

The *required* flag of a sub-purpose is defined as smaller or equal to the super-purpose flag (3). Here the Boolean order is defined as follows: $true > false$. Meaning that the super-purpose must be required if at least one of its sub-purposes is required. However, if the super-purpose is required, sub-purposes can be optional.

The $pointOfAcceptance$ of the sub-purposes must be greater or equal to the $pointOfAcceptance$ of the super-purpose (4). Since the point of acceptance is a date and time combination, this means that sub-purposes can be accepted after the super-purpose has been accepted, but not the other way round.

The set of data elements of the sub-purpose needs to be a sub-set of or equal to the set of data elements of the super-purpose (5).

For the privacy models $\widehat{PM}$ we define the $\subseteq_{pm}$ relation (6). This condition depends on the different privacy models used. Since LPL does not provide a defined set of privacy models, we do not define $\subseteq_{pm}$ in this report; instead we give an example and leave the definition of this relation to future research. The relation $\subseteq_{pm}$ combines the regular $\subseteq$ with further conditions that allow for further cases. The simplest case for this condition is $\subseteq$, where the privacy models of the sub-purposes are aggregated in the super-purpose. However, the privacy models of the different sub-purposes could also be combined in a single new privacy model. When the same privacy model is used the attributes of the model can be compared directly, e.g., for k-anonymity the value of k can be compared, and the k of the new privacy model of the super-purpose must be greater than or equal to the k of the sub-purposes. Different models can also be compared with each other, e.g., l-diversity and k-anonymity. If the sub-purpose uses k-anonymity and the super-purpose uses l-diversity, $l \geq k$ must hold.

For the pseudonymization methods $\widehat{PSM}$ the super-purpose contains an aggregation of the PSMs of the sub-purposes (7).

The set of data recipients of the sub-purpose $\widehat{DR_s}$ needs to be a sub-set of or equal to the set of data recipients of the super-purpose $\widehat{DR_p}$ (8).

For the legal bases we define the $\subseteq_{lb}$ relation as an extended subset relation (9). The legal bases of the sub-purpose $\widehat{LB_s}$ can be a sub-set of or equal to the set of legal bases of the super-purpose $\widehat{LB_p}$. However, if multiple sub-purposes are using the same *lbCategory* of legal basis (cf. Section 2.1) these can be aggregated in a single legal basis element of the super-purpose, where $\widehat{HEAD}$ and $\widehat{DESC}$ attributes from the sub-purposes are combined.

For the automated decision-making elements we define the $\subseteq_{adm}$ relation, similar to the $\subseteq_{lb}$ relation of the legal bases (10). Automated decision-making elements of the sub-purpose $\widehat{ADM_s}$ can be a sub-set of or equal to the set of automated decision-making elements of the super-purpose $\widehat{ADM_p}$. However, if multiple sub-purposes are using the

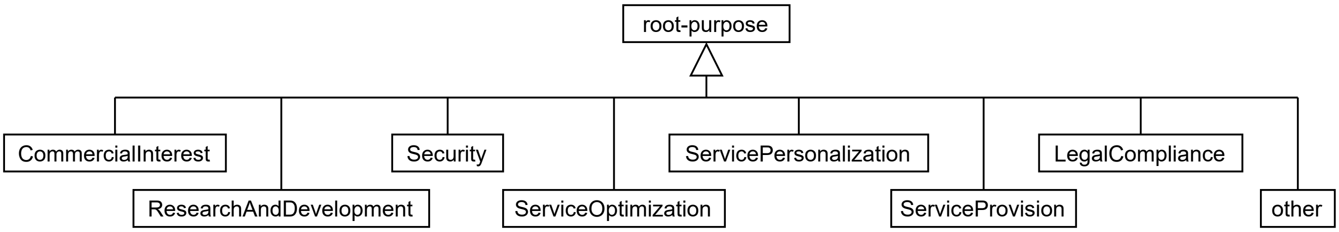

Fig. 2. Example structure of a purpose hierarchy

same automated decision-making process, with different descriptions, these can be aggregated in a single automated decision-making element of the super-purpose, where $\widehat{HEAD}$ and $\widehat{DESC}$ attributes from the sub-purposes are combined.

For the retention elements we define the $\leq_r$ relation. Here we keep the definition of Hjerppe et al., where the different types of retention are ordered as follows: $FixedDate \leq_r AfterPurpose \leq_r Indefinite$. To compare *FixedDate* and *AfterPurpose* elements with elements of the same type the *pointInTime* attribute of the retention element is simply compared using $\leq$. This means that there is no sub-purpose in the aggregation of purposes, for which the retention is longer than what is stated in the super-purpose.

Conditions (2), (5), (8), and (11) are taken directly from the work of Hjerppe et al.; and (3) and (6) were changed for more freedom in the design of privacy policies in LPL and due to changes made to LPL in later versions. Conditions (4), (7), (9), and (10) were added because of the addition of new attributes in the purpose element of the current version of LPL.

2.8 Fixed Super-Purposes

For improved organization of purposes in LPL privacy policies we introduce a set of given purposes that are used in the purpose hierarchy, which we introduced in the previous section.

To identify the set of given purposes, which can be seen in Fig. 2, we considered the purpose categories from the Platform for Privacy Preferences Project (P3P) [10], the World Wide Web Consortium's (W3C) Data Privacy Vocabularies and Controls Community Group's Data Privacy Vocabulary (DPV)[11], and the Scalable Policy-awarE Linked Data arChitecture for prIvacy, trAnsparency and compLiance (SPECIAL) project's purpose taxonomy [12].

The SPECIAL purpose taxonomy is based on P3P's purposes and is an improvement over it as some unnecessary purposes have been removed. P3P used specialized purposes, which describe the means in which a purpose is fulfilled instead of describing the actual purpose. These purposes were removed in the set of SPECIAL purposes.

The W3C's DPV was further derived from the SPECIAL purposes, however these purposes were renamed to better fit to common data protection language. Hence, we use the DPV's purposes:

'CommercialInterest', 'ResearchAndDevelopment', 'Security', 'ServiceOptimization', 'ServicePersonalization', 'ServiceProvision', 'LegalCompliance', 'other'.

We added the given purpose *other*, so that policy authors are enabled to specify purposes that cannot be classified into one of the given purposes.

2.9 Fixed Data Groups

In addition to the given purposes presented in the previous section, we also provide fixed data groups. Originally, LPL used data groups defined by the policy author. However, fixed data groups are a requirement to be able to argue about the data specified in an LPL policy.

The fixed data groups are used within the $\widehat{DG}$ attribute of the *data* element. By definition, the $\widehat{DG}$ attribute allows authors to assign multiple data groups to a single datum. This enables precise classification of data into the provided data groups.

For the given data groups we use the data categories from the SPECIAL project [12]. These are more detailed than the ones presented in the Data Privacy Vocabulary [11], thus, allowing better differentiation between the different types of data. The data groups used are: *'activity', 'anonymized', 'computer', 'content', 'demographic', 'derived', 'financial', 'government', 'health', 'interactive', 'judicial', 'location', 'navigation', 'online', 'physical', 'political', 'preference', 'profile', 'purchase', 'social', 'state', 'uniqueId'*.

References

[1] A. Gerl, N. Bennani, H. Kosch, and L. Brunie, *LPL, Towards a GDPR-Compliant Privacy Language: Formal Definition and Usage*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2018, pp. 41–80. [Online]. Available: https://doi.org/10.1007/978-3-662-57932-9_2

[2] K. Hjerppe, J. Ruohonen, and V. Leppänen, "Extracting layered privacy language purposes from web services," in *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 2020, Conference Proceedings, pp. 318–325.

[3] A. Gerl and B. Meier, "The layered privacy language art. 12–14 GDPR extension–privacy enhancing user interfaces," *Datenschutz und Datensicherheit-DuD*, vol. 43, no. 12, pp. 747–752, 2019.

[4] A. Gerl, "Extending layered privacy language to support privacy icons for a personal privacy policy user interface," in *Proceedings of British HCI 2018*. BCS Learning and Development Ltd., Belfast, UK, 2018, p. 5.

[5] A. Gerl and F. Bölz, "Layered privacy language (LPL) pseudonymization extension for health care," in *MEDINFO 2019: Health and Wellbeing e-Networks for All*, ser. Studies in Health Technology and Informatics, L. Ohno-Machado and B. Séroussi, Eds., vol. 264, 2019, pp. 1189 – 1193.

[6] A. Gerl, "Modelling of a privacy language and efficient policy-based de-identification," Ph.D. dissertation, Universität Passau, 2020. [Online]. Available: https://nbn-resolving.org/urn:nbn:de:bvb:739-opus4-7674

[7] European Parliament and Council of the European Union, "Regulation 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)," *Official Journal of the European Union*, vol. L119, pp. 1–88, 2016. [Online]. Available: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC

[8] A. Gerl, "Extending layered privacy language to support privacy icons for a personal privacy policy user interface," in *International BCS Human Computer Interaction Conference (HCI)*, 2018, Conference Proceedings. [Online]. Available: https://doi.org/10.14236/ewic/HCI2018.177

[9] ISO 3166-1:2020, "Codes for the representation of names of countries and their subdivisions — Part 1: Country code," International Organization for Standardization, Geneva, CH, Standard, August 2020.

[10] L. Cranor, B. Dobbs, S. Egelman, G. Hogben, J. Humphrey, M. Langheinrich, M. Marchiori, M. Presler-Marshall, M. Schunter, and R. Wenning, "The platform for privacy preferences 1.1 (p3p1. 1) specification," *W3C Working Group Note*, p. 57, 2006.

[11] H. J. Pandit, "Data privacy vocabulary (dpv)," Data Privacy Vocabularies and Controls Community Group, Standard, 13.01.2021 2021. [Online]. Available: https://dpvcg.github.io/dpv/

[12] P. Bonatti, L. Sauro, I. Petrova, S. Kirrane, and E. Schlehahn, "Policy language v1," Scalable Policy-awarE Linked Data arChitecture for prIvacy, trAnsparency and compLiance, deliverable, 26.12.2017 2017. [Online]. Available: https://specialprivacy.ercim.eu/images/documents/SPECIAL_D21_M12_V10.pdf

DOI: 10.17185/duepublico/74569
URN: urn:nbn:de:hbz:464-20210712-090625-4

Date of publication: 2021-07-12