Towards Privacy in Personal Data Management

Pavlos S. Efraimidis Georgios Drosatos Fotis Nalbadis Aimilia Tasidou

Department of Electrical and Computer Engineering
Democritus University of Thrace
University Campus, 67100 Xanthi, Greece
{pefraimi, gdrosato, fnalmpan, atasidou}@ee.duth.gr

Abstract

Purpose – In order to enhance privacy protection during electronic transactions, we propose, develop and evaluate a personal data management framework called Polis, which abides by the following principle: Every individual has absolute control over her personal data, which reside only at her own side.

Design/methodology/approach – The paper identifies representative electronic transactions that involve personal data and proposes Polis-based protocols for them. The approach is evaluated on a Polis prototype both as a stand-alone application and as part of a commercial database management system.

Findings – The results of this work indicate that electronic transactions can remain both feasible and straightforward, while personal data remain only at the owner's side.

Research limitations/implications – The paper describes a Polis-approach implementing prototype, which is easy to deploy and friendly to current information management technologies. However, the usability of the prototype has to be enhanced with supporting tools for editing personal data and policies and a more intuitive user interface. Finally, the Polis-platform enables a new class of user-centered distributed applications, which we intend to investigate.

Practical implications – Even though the conditions for a personal data management approach like Polis are mature, and Polis can be progressively adopted, it still entails a major change in current business practices.

Originality/value – The paper proposes a new paradigm for the management of personal data, which admits individuals to have their personal data stored only at their own side. The new approach can be of mutual benefit to both individuals and companies.

Keywords Privacy, Personal Data Management

Paper type Research paper

1 Introduction

As the use of computers and the Internet becomes more popular by the minute, the issue of protecting one's personal data is more essential than ever. The way electronic transactions are conducted nowadays, makes it necessary for the customer to give away her personal data to the service provider and hope that the latter will not use them in a malicious way. In order to protect

personal information, several organizations and countries have issued privacy regulations, which should be followed in order for personal information to be protected; the collectively referred to as Fair Information Practices (FIP). Examples of important FIP regulation frameworks are the Data Protection Directive 95/46/EC (henceforth referred to as The Directive) and follow-ups like the Canadian PIPEDA and UK's Data Protection Act (DPA).

In this work, we assert that electronic transactions can be feasible, whilst personal data resides at the individuals' side. To support this claim, we design, build and evaluate the prototype system Polis, which implements the above principle. We show that Polis can satisfy important data protection principles in a natural and efficient way and describe how Polis can be integrated into online transactions to manage personal data. The results of this work indicate that the Polis approach can lead to a simple, scalable solution that can be beneficial to both individuals and service providers.

Related Work. The idea that individuals should own their personal information themselves and decide how this information is used, is discussed in [Laudon, 1996]. A point made in [Samuelson, 2000] is that, although considering personal data the owner's private property is a very appealing idea, it would be rather difficult to practically apply it and legally enforce it. Our approach proposes an idea that has the same practical effect as considering personal data the owner's private property, but withdraws the legal objections involved with this idea. The economic aspects of privacy are examined in [Varian, 1996] where the following point is made: "It is worth observing that the Fair Information Practices principles would automatically be implemented if the property rights in individual information resided solely with those individuals". The argument that personal data would be safer at the user's side is also examined in [Mulligan and Schwartz, 2000].

Different kinds of frameworks that are related to personal data have recently been proposed or are in progress. In particular, privacy sensitive management of personal data in ubiquitous computing is discussed in [Hong, 2005], storing personal data in an individual's mobile device is examined in [Jäppinen, 2004]. Protecting personal data that is stored within a company is considered in [Salim et al., 2007, Karjoth et al., 2002]. More related to Polis is a rich but also complicated framework for privacy protection, proposed in [Lioudakis et al., 2007]. This framework is built on the principle that personal data is kept inside a "Discreet Box", located at the service provider's side. An agent-based solution to address usability issues related to P3P (Platform for Privacy Preferences Project) is presented in [Lee and Stamp, 2008]. Other results in this field, less related to Polis, can be found in [Goldberg, 2000, Bangerter et al., 2004]. General surveys on privacy enhancing technologies are given in [Goldberg, 2007, Gritzalis, 2004]. To our knowledge, Polis is the first general framework for managing personal data only at the owner's side.

Outline. The rest of this work is organized in the following way: The Polis approach is described in Section 2 and the rationale of the Polis principles is presented in Section 3. Incentives and potential objections of prospective Polis users are analyzed in Section 4. In Section 5, the possible applications for Polis are discussed. The Polis prototype is presented and evaluated in Section 6. A final discussion is given in Section 7. A preliminary version of this work has been presented in [Efraimidis et al., 2008].

2 The Polis approach

The Polis approach is based on the following principle:

"Polis-users are prohibited from storing any personal data but their own."

Polis is meant to be employed by privacy concerned internet users which fulfill the requirements of having:

- A reliable, always-on access to the Internet, in order for her agent to be always accessible.
- A certificate from an approved Certification Authority.

We design, implement and evaluate a Polis prototype and show that the above simple and straightforward assumptions suffice to build a personal data management framework that works seamlessly with online transactions. The Polis prototype and its evaluation are described in Section 6.

2.1 Polis concepts and architecture

At this point we consider it necessary to introduce a few terms that will be used in this work:

- In Polis, personal data refers to primitive personal information of individuals like name, birth date, address, etc. Personal data corresponds to what is called *off-line identity* in [Acquisti, 2004]. Our focus is on privacy-enhanced management of the off-line identity.
- An individual Internet user is a potential customer who can purchase either goods or services. This user can be called *individual*, *customer* or *data subject (according to The Directive)*. We will use the terms *individual* and *customer*, interchangeably.
- An entity that provides the aforementioned goods or services can be called *shop*, *company*, *service provider* or *data controller (The Directive)*. We will use the terms *shop*, *company* and *service providers*.
- Both individuals and companies can become *Polis-users*.

Every Polis-user is represented by a dedicated entity. This entity can be used to instantiate a corresponding Polis-agent, which is the main architectural component of Polis. The agent is used to manage the personal data of the entity and to provide controlled access to it. Service providers use the agent to retrieve personal data from affiliated users. The general architecture of Polis, as well as the constituents of a customer agent and a shop agent are presented in Figure 1. We would like to emphasize the following characteristics of the Polis architecture:

- From the service provider's point of view, Polis provides a decentralized approach for the storage and management of personal data.
- On the contrary, from the customer's point of view, Polis is a fully centralized system, in the sense that personal data is located and managed locally by the owner's agent.

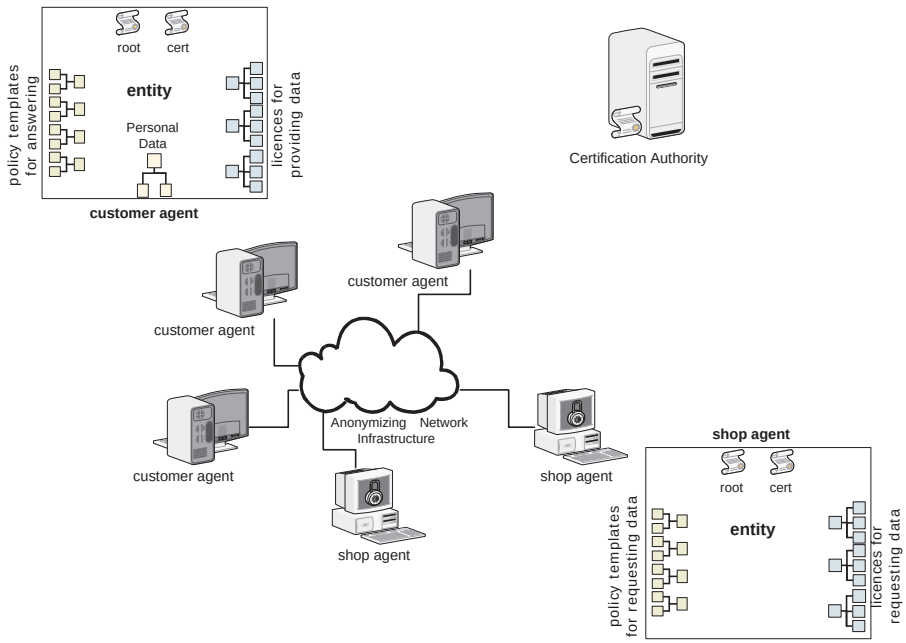


Figure 1: The Polis architecture

2.2 Schemes for personal data and policies

Critical components for a personal data management framework like Polis are the schemes for representing personal data and policies. Some known schemes for personal data are P3P [W3C, 2002] and CPExchange [Bohrer and Holland, 2000]. Approaches for policies related to personal data are also discussed in [Karjoth and Schunter, 2002, Karjoth et al., 2002], while related work on personal data and policy schemes is performed in [DISCREET, 2008].

We currently use schemes that are simple, yet powerful enough, for the current needs of the Polis prototype. Examples of a personal data scheme and a policy, as used in Polis, are shown in Figure 2.

There are eight general categories of personal data in Polis, organized hierarchically, namely Name, BDate, Cert, Skill, Characteristic, Home-Info, Business-Info and CreditCard. Each of them has one or more subcategories. The terminology used is based on P3P for the user information part, with the addition of the financial information (CreditCard) taken from CPExchange, plus the extra personal information fields (Skill and Characteristic). Each entity stores its personal data in a local XML document.

The components of a policy are the following:

- *Principals*: The Polis-entities.
- *Data*: Every single item of a user's (Polis-entity) personal data.
- *Purposes*: The set of purposes that entitle principals to retrieve data.

Policy

```
<User Enabled="true" Entity=" eshop">
 <Name>
  <Given>
    <Permissions>
      <License Purpose="shipping">
        <GrantAccess>true</GrantAccess>
        <DateTime>
          <Start>2008-01-01 00:00:00</Start>
          <End>2008-12-31 23:59:59</End>
        </DateTime>
      </License>
      <License Purpose="billing">
        <Count>3</Count>
        <DateTime>
          <Start>2008-01-01 00:00:00</Start>
          <End>2008-12-31 23:59:59</End>
        </DateTime>
      </License>
        ...
    </Permissions>
  </Given>
 </Name>
</User>
```

Personal Data

```
<User>
 <Name>
   <Given>John</Given>
   <Family>Doe</Family>
 </Name>
 <Home-Info>
   <Postal>
     <Street>Nowhere Street 001</Street>
     <City>Deadend</City>
     <StateProv>Ouitcy</StateProv>
     <PostalCode>11111</PostalCode>
     <Organization>DIPH</Organization>
     <Country>Neverland</Country>
   </Postal>
   <Telecom>
      ...
   </Telecom>
 </Home-Info>
</User>
```

Figure 2: Examples of a personal data scheme and a policy

- *Usage restrictions*: Additional restrictions exist that limit access rights to a specific number of accesses or a specific time interval, or both.

Other important concepts of Polis are the licence and the contract. A *licence* comprises of the data involved, the valid purposes that allow data retrieval, as well as the rules to provide either full or restricted access. The use of licences to protect personal data is discussed in [Cha and Joung, 2003, Korba and Kenny, 2003, Fahrmair et al., 2005]. A *contract* concerns two principals and an arbitrary set of licences. An agent can sign any number of contracts with an arbitrary number of entities.

3 In Defense of the Polis Approach

The central assumption in the described personal data management approach is that personal data can only be stored at the owner's side. One may dismiss this as an unrealistic hypothesis and contend that we cannot count on users to abide by the Polis principle. One may also doubt that there are any incentives to adopt an approach like Polis, especially for the service providers.

We are well aware that Polis specifies an extreme approach for the management of personal data. However, Polis is meant to provide proof of concept that personal data management can be fair, privacy respecting and more effective than current practices. What makes the Polis approach possible is that the recent scientific and technological developments and especially the universal acceptance of the Internet have prepared the ground for citizen-centric applications. At the same time, the powerful surveillance and data management tools have contributed to making privacy threats and personal data misuse one of the most important problems in the electronic world.

For the above reasons, we believe that the conditions are mature for investigating alternative

paradigms in the management of personal data. The new paradigms should enhance the individuals' control over their personal data. In this context we designed, implemented and evaluated the Polis platform. There is no doubt, that much more has to be done for a solution like Polis to find its way to practice. However, this work constitutes a confirmation that such a solution is feasible.

With Polis we aim at providing a consumer-led solution for personal data management to be used instead of the current company-centric approach. We would like to point out an interesting analogy of the Polis-proposed switch in the current practices in the field of personal data management with another proposed switch in the field of identity management. According to the Crosby report [Crosby, 2008] we should focus on identity assurance instead of identity management:

> At an early stage, we recognized that consumers constitute the common ground between the public and private sectors. And our focus switched from ID management to ID assurance. The expression ID management suggests data sharing and database consolidation, concepts which principally serve the interests of the owner of the database, for example the Government or the banks. Whereas we think of ID assurance as a consumer-led concept, a process that meets an important consumer need without necessarily providing any spin-off benefits to the owner of any database [Crosby, 2008].

The Crosby report [Crosby, 2008] was published in March 2008 after we had designed and built the Polis prototype. We consider it very encouraging that positions about consumer-led solutions are expressed within a very applied context, like the Forum on Identity Management which prepared the report.

4 Incentives and Objections

In this Section we discuss incentives and objections for the Polis approach and provide arguments that the Polis solution can be beneficial not only to individuals but also to (well meaning) shops.

The fact that a Polis-user's personal data must be retrieved from the owner's side every time it is needed, automatically fulfills many critical requirements found in FIP-like regulations. Moreover, the way the Polis framework can be integrated into database management systems automatically fulfills the requirements of Hippocratic databases [Agrawal et al., 2002]. Besides individuals, shops can also obtain important benefits from the adoption of Polis, like a more privacy-friendly profile, simplified data maintenance and data cleansing, as well as significantly reduced responsibility for the safety of customers' personal data.

4.1 Incentives for individuals

1. *Polis-users maintain maximum control over their personal data.* They are able to monitor, at any point in time, all the contracts they have signed, as well as the time, purpose and principal of each data item access has taken place. Unlike the Polis approach, current practices result in inability to keep track of where one's personal data reside and how often they are being accessed.

2. *Individuals can trivially exercise their right to up-to-date personal data.* A user has simply to update her locally stored record. Each time a company wants to access it, it will be retrieved

on-the-fly from the person's agent and not from the company's outdated database. Consider an individual who changes her address or telephone number. With current practices, the individual has to recall every peer that rightfully possesses this information and go through a record update procedure for each such peer.

3. *Individuals can handle all kinds of privacy-related rights and preferences through their Polis-agent.* To this end, a unified user interface is being used, while personal data disclosure takes place through a clear data flow. These attributes are absolute requirements for the effective privacy-enhanced management of personal data [Lederer et al., 2005].

4. *The risk that the privacy of the individual is violated due to data breaches from company databases is significantly reduced.* Just like credit cards, Polis-contracts can be canceled to become useless to invaders. Even if a company does not realize that its database has been compromised, invaders will have to acquire the company's private key, in order to be able to use the stolen contracts. Even in that case, the invaders will only have access to the particular data that the contracts authorize this company for. Furthermore, the data owner will be able to know what data has leaked and when this happened.

5. *Privacy-concerned individuals will no longer have to choose between either giving away their personal data or not conducting an electronic transaction.* Nowadays individuals suffer the coercion that occurs when there is only one reasonable way for them to receive certain needed services or information [Millett et al., 2001], i.e., by giving away their personal data. Furthermore, according to Acquisti in [Acquisti, 2004]: "... as merchants decide against offering anonymizing technologies to their customers, the privacy concerned customers choose not to purchase on-line, or to purchase less. A latent, potentially large market demand remains therefore unsatisfied". We believe that approaches like Polis can offer a viable alternative to current practices for personal data management.

4.2 Incentives for service providers

1. *The customer's personal data in the shop's database remains always up-to-date.* In addition, this is accomplished without any maintenance costs for the shop.

2. *The use of Polis contributes to improved data quality and can simplify the data cleansing task.* Data cleansing is the act of detecting and removing and/or correcting a database's dirty data (i.e., data that is incorrect, out-of-date, redundant, incomplete, or formatted incorrectly). Data quality is a critical factor for the success of enterprise intelligence initiatives and can incur costs and delays to company operations [Rahm and Do, 2000, Vassiliadis et al., 2002].

3. *Polis releases the shop from the burden and responsibility of keeping customer data safe.* The shop is freed from a set of serious responsibilities for protecting customers' personal data and the risk of being considered liable for serious data breaches. Incidents of intentional or unintentional data breaches are unfortunately quite common and a reasonable worry is that a lot of them never reach the attention of the media. Some representative examples of such situations are the Choicepoint case, a data broker who sold private records of over 150,000 Americans to a group of criminals in 2005 [ConsumerReports, 2006], the incident

that took place in the UK, where two computer discs containing the personal data of 25 million citizens were lost in the post [Anderson, 2007], as well as the recent Deutsche Telecom incidents [Reuters, 2008].

4. *Polis promotes a more privacy-friendly image for the service providers that adopt it.* The commitment that the shop does not store any personal data locally is an appealing argument for privacy-sensitive customers.

5. *Polis can be integrated into a company's existing information system.* As we illustrate in Section 6, Polis can naturally handle heterogeneous sets of customers, consisting of both Polis and conventional ones. This fact removes the counterincentive of companies having to go through a demanding transition process in order to integrate Polis into their systems. The company does not get tied down by Polis into having only Polis users in its database.

4.3 Potential objections for individuals

1. *Managing a personal agent is by definition a critical task, prone to errors and omissions by the user.* However, being in charge of the data disclosure process through a unified procedure like Polis, is much more convenient and effective compared to current practice, as described in incentive for individuals 3.

2. *Considerations about the agent's security.* An individual's Polis-agent contains critical personal data and digital agreements for data access. Consequently, a production-ready Polis-agent should satisfy high security levels. We believe that this is a viable task, since the Polis-agent has a precise, well-defined functionality and can be operated behind firewalls on a user-controlled computing platform. Moreover, the decentralized approach of Polis for personal data can also contribute to improved data security, since invaders find large collections of personal data much more inviting than an individual's personal data [Mulligan and Schwartz, 2000].

3. *Polis does not protect individuals from malicious shops that misuse personal data.* Nevertheless, a malicious shop in Polis cannot cause more damage than it could cause with current practices.

4.4 Potential objections for service providers

1. *Loosing control of customer data.* This objection does not really apply to the Polis approach since service providers will still have access to the data they are entitled to. Well-meaning parties will not loose control over their customer's data. Internet connection reliability could also be an issue for Polis, but as already mentioned, it is widely accepted that soon enough, reliable Internet connectivity will be considered a given. Besides, Polis does not restrain companies from keeping records of customer's profiles. These records will not contain any data of the customer's offline identity and will resemble pseudonymous data processing.

2. *The adoption of Polis can cause significant overheads to company processes.* The possible delays in data retrieval caused by the employment of Polis should not be a hindering factor for its adoption. Retrieval of personal data is neither a task that is carried out frequently,

nor a time critical process, therefore these delays will not affect the efficiency of the company procedures.

3. *Service providers could be scammed from malicious Polis-users.* Polis-contracts and licences constitute proof that a service provider has the right to access the specified Polis-user data. Therefore, when needed, a company can resort to the appropriate actions. The CA or some other designated trusted third party could be used to settle such cases.

4.5 Enforcement and Detection

An important aspect of every (electronic) contract is the ability to verify and enforce that the parties will not violate its terms. Polis can handle detectable privacy breaches, i.e., breaches for which data released to the shop finds its way back to the individual who submitted that information [Golle et al., 2006]. In this case a Polis compliant shop must be able to present evidence that those data were rightfully obtained for the specific purpose, at the specific time, using data licences [Cha and Joung, 2003]. A more challenging task would be to detect Polis-shops that leak customer's personal information. A relevant problem is discussed in [Golle et al., 2006].

Due to the very nature of personal data, it seems that once a service provider possesses some data, there is no technically feasible way for absolute abuse prevention. Consequently, apart from technical measures, we will have to rely on market, legal and social dynamics for handling personal data properly ([Golle et al., 2006, Aggarwal et al., 2005] and [Hong, 2005, Section 5.8.5]).

As far as violations from the user's side are concerned, if the terms of an agreement are violated and the individual refuses to fulfil her contract defined obligation of providing personal information, then the service provider can use her customer-signed licence to prove entitlement to access the data.

5 Polis Applications

In this Section, we discuss how Polis can be used within common electronic transactions and present indicative higher level applications that can be built on top of a decentralized personal data management framework like Polis.

5.1 Polis in common transactions

Polis can be potentially employed in any transaction where a user has to enter (some of) her personal data. The overall procedure is outlined below:

Polis in a transaction When the user has to fill in a form with her personal data, she instead provides the contact details of her agent. The agents of the shop and the user/customer establish an agreement. A successful agreement grants access to the customer's private data for the specific data items and the amount of time needed to complete the transaction. In Figure 3 the procedure of a Polis-transaction with an e-shop is detailed.

This procedure can be used for registrations at e-shops, portals and other online services. In general, any application that involves personal data, like identity management systems [PRIME, 2008] and e-government platforms can be supported. The need for privacy protection in e-business applications is stressed in [Katsikas et al., 2005]. The ease of employing Polis lies in the fact that it can work as middleware, which takes care of the personal data exchange between parties in higher level applications.

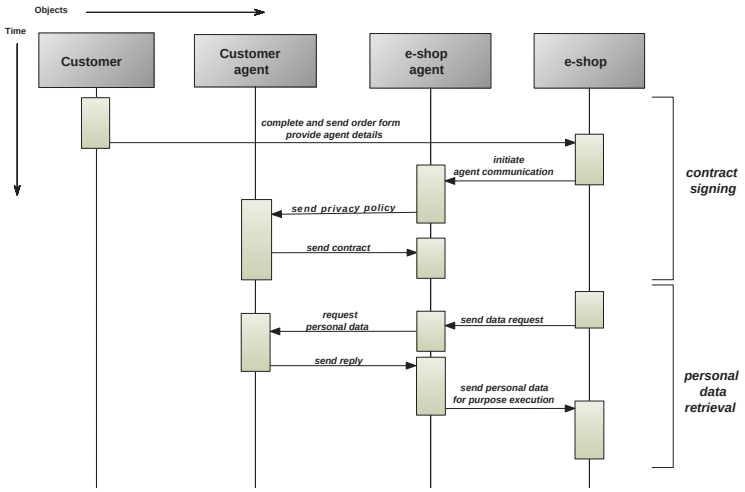


Figure 3: Polis in transactions with an e-shop

5.2 Prospective Applications for Polis

An infrastructure like Polis can be a realistic step in the direction of effectively controlling personal information. Apart from the direct gains of using Polis in every day electronic transactions, there are some interesting possibilities for higher level applications that could utilize it.

Microtrades and Information Markets. The financial aspects of privacy are studied in several works like [Laudon, 1996, Varian, 1996, Kleinberg et al., 2001, Acquisti, 2004]. Polis could be utilized to facilitate personal data exchange in personal-level microtrades between Polis-agents. Such an application is examined in [Tasidou et al., 2009]. Polis-users can give permission to information gathering companies to access (some of) their personal data, for an agreed price. Each time a company needs to regain access to them, the agreed amount of money should be paid. Furthermore, Polis could provide the ground for more advanced financial applications for personal data. The market for personal data described by Laudon in [Laudon, 1996] is an example of such applications. In particular, Laudon proposes the so called National Information Market (NIM), where personal information can be traded in a National Information Exchange. The adoption of a framework like Polis would simplify the evolution of NIM-like infrastructures.

Privacy-enhanced ubiquitous computing. Online data of an individual can be conveyed through her Polis-agent. In this case, Polis could work as an open architecture for ubiquitous computing applications. For example, dynamic location information could be retrieved from the individual's Polis-agent, like the rest of her personal data.

6 The Polis prototype

We designed and implemented a Polis prototype with the main objective to demonstrate that electronic transactions are feasible while personal data remain only at the owner's side. Another technical objective of the development of the Polis prototype was to make its deployment simple and friendly to contemporary information management practices. We believe that we have fulfilled the above goals adequately. Furthermore, a fully developed Polis platform should be able to satisfy the general properties that a privacy technology must have in order to be considered useful according to [Goldberg, 2007]. The Polis project site [Polis, 2008] hosts freely available binaries and online demos of Polis.

6.1 Technologies of the prototype

The basic technologies used to develop and employ the Polis prototype are:

- The Eclipse IDE and the Java programming language to create portable, platform independent tools.
- A Pubic Key Infrastructure (PKI) is used for creating trusted certificates according to the X.509 standard. For demonstration purposes, an elementary Polis CA has been developed to be used in experiments. In a real world application, a commercial CA could be utilized.
- User data, policies and contracts are represented as XML documents.
- The Tor anonymizing infrastructure is used optionally to achieve anonymity for the clients and/or implement agents as hidden services [Dingledine et al., 2004].
- The Derby embedded database server is employed internally by the agent for its data storage needs.
- Bouncy Castle's security provider is used for cryptographic primitives.
- The database case study has been implemented on an Oracle database management server (DBMS). Similar integrations of Polis should be feasible with other popular DBMSs like IBM DB2 or Microsoft SQL Server.

6.2 Deploying the Polis prototype

In order to deploy Polis:

- Customers install the Polis-agent, store their personal information and prepare the necessary policy templates.

- Companies install the Polis-agent, prepare policy templates and integrate the agent with the company's information system. *Polis-customers can co-exist with normal customers at a company side.*

6.3 Polis collaborating with a database management system

Polis can be incorporated into the back-office of a company and take care of the personal data management. This is accomplished by integrating Polis with the company's database management system. The basic idea is that personal data fields do not contain the actual data, instead, a ticket (represented by an appropriate object) is used to retrieve the data value on the fly. We tested Polis with an Oracle database server. The approach is illustrated in Figure 4.

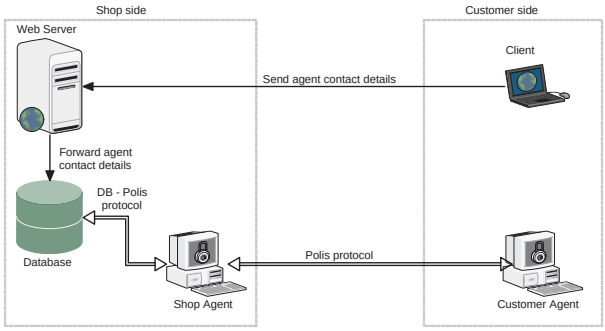


Figure 4: A Polis-entities interaction example. The shop uses a Polis-enabled database for customer registration and for retrieving personal data of customers.

The integration was straightforward. Two Java Stored Procedures (JSP's) and a small set of triggers and database views were sufficient to implement the connectivity between Polis and the database server. It is noteworthy that, using simple object-relational features, as well as views and triggers, the Polis enhanced database can be operated as a normal one, while the Polis related operations are transparent to the database user.

6.4 Experimental evaluation

We prepared an elementary Polis environment with a set of Polis-agents installed on the local network of our laboratory. A snapshot of a Polis-agent's GUI is shown in Figure 5. A set of web pages, including web forms and dynamic web pages, were used to support experiments. The customer database contained 27 customers in total; 11 conventional customers and 16 Polis-customers (4 of which used Tor hidden services [Dingledine et al., 2004] for their agents). We performed an extensive set of experiments within the above Polis environment. The experiments involved database operations on tables with Polis data to verify their integration into the database. In particular, we executed some representative *insert* and *select* operations on the customer table, a *join* operation between two tables and created some *views* in the database. Both Tor-enabled Polis-agents (the agent is accessed through a Tor hidden service) and conventional Polis agents were tested. As expected, all the operations were accomplished successfully. Moreover, the Tor-enabled agents

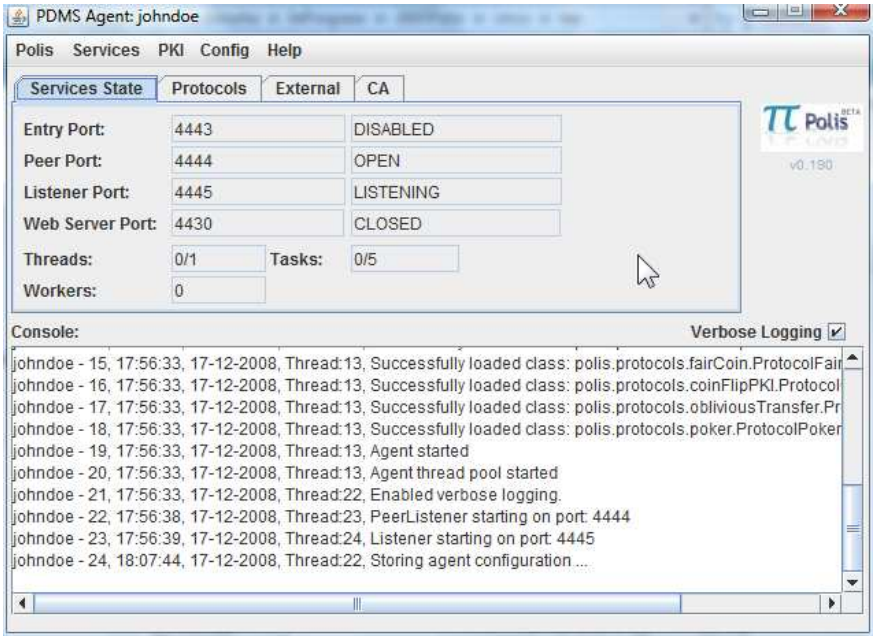


Figure 5: A Polis-agent's GUI snapshot

operated indistinguishably from the conventional agents with the exception of occasional timeouts, due to the Tor network itself.

ID	First Name	Last Name	Street	City	Postal Code	Country	State	Organization	User Type
23	Sayid	Jarrah	Goodmind 125	Kenya	87484	Kenya	Kenya	TN	**Non-polis User**
22	Walter	Skinner	Filosofou 145	Florina	No Permission	No Permission	No Permission	EU	**Polis user**
21	Fox	Mulder	Kolokotroni 40	Athens	10156	No Permission	No Permission	No Permission	**Polis user**
20	Joey	Tribbiani	Central Perk 46	New York	74889	USA	New York	US	**Non-polis User**
19	Dana	Scully	Paranormal Street 125	Deadend	No Permission	No Permission	No Permission	No Permission	**Polis user**
18	Phoebe	Buffay	Central Perk 23	New York	47952	USA	New York	US	**Non-polis User**
17	Veronica	Donovan	Xonoloulou 210	Tokyo	No Permission	No Permission	No Permission	No Permission	**Polis user**
16	Alexander	Mahone	Fox River 15	Washington	8769	USA	Washington	US	**Non-polis User**

Figure 6: Report from the customer table of a Polis-enabled database. The table contains both, Polis and non-Polis, users.

6.5 A first case study

We performed a preliminary case study on the integration of Polis with a content management system (CMS). More precisely, we integrated Polis with the Elxis CMS, an open source CMS released under the GNU/GPL license. We selected Elxis for our case study because it is a fully functional CMS, it is open source and it supports the Oracle DBMS. A number of extensions exist for Elxis that enrich its functionality; one of them, the IOS eshop component, turns Elxis into an e-shop.

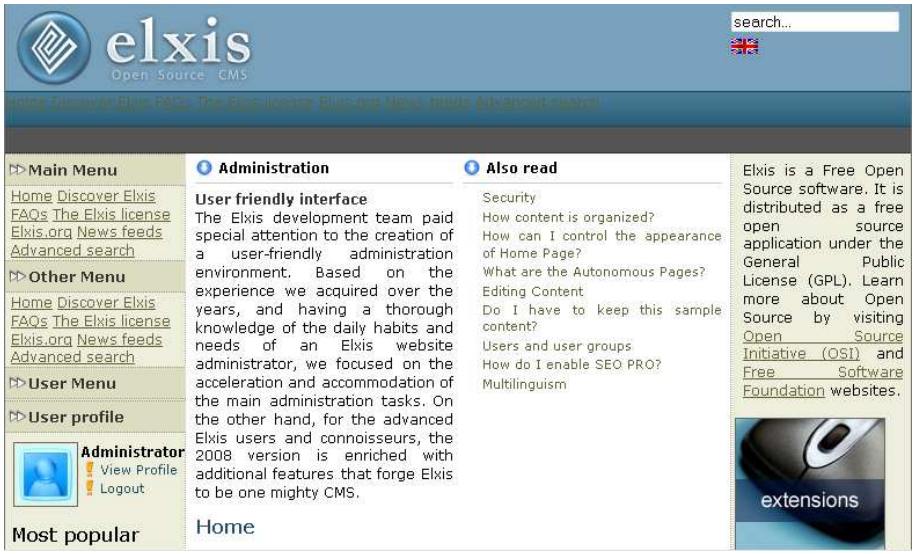


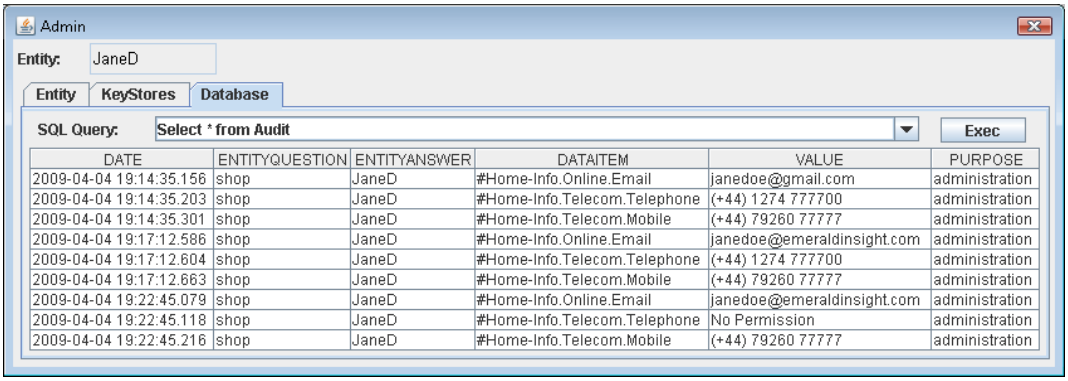
Figure 7: The Polis-enabled Elxis CMS instance

The integration process was straightforward. After less than a man week of work, a first version of a Polis-enabled Elxis CMS instance was working in beta status (Figure 7). Figure 8 shows how the profile of a specific user appears in the Elxis application, at different time instances. Auditing the user's Polis-agent reveals each access to the user's personal data.

(a) The Elxis user profile.

(b) The profile after the user changed her e-mail address.

(c) The profile after the data license for the telephone field has expired.



DATE	ENTITYQUESTION	ENTITYANSWER	DATAITEM	VALUE	PURPOSE
2009-04-04 19:14:35.156	shop	JaneD	#Home-Info.Online.Email	janedoe@gmail.com	administration
2009-04-04 19:14:35.203	shop	JaneD	#Home-Info.Telecom.Telephone	(+44) 1274 777700	administration
2009-04-04 19:14:35.301	shop	JaneD	#Home-Info.Telecom.Mobile	(+44) 79260 77777	administration
2009-04-04 19:17:12.586	shop	JaneD	#Home-Info.Online.Email	janedoe@emeraldinsight.com	administration
2009-04-04 19:17:12.604	shop	JaneD	#Home-Info.Telecom.Telephone	(+44) 1274 777700	administration
2009-04-04 19:17:12.663	shop	JaneD	#Home-Info.Telecom.Mobile	(+44) 79260 77777	administration
2009-04-04 19:22:45.079	shop	JaneD	#Home-Info.Online.Email	janedoe@emeraldinsight.com	administration
2009-04-04 19:22:45.118	shop	JaneD	#Home-Info.Telecom.Telephone	No Permission	administration
2009-04-04 19:22:45.216	shop	JaneD	#Home-Info.Telecom.Mobile	(+44) 79260 77777	administration

(d) The user's Polis-agent audit.

Figure 8: The profile of a Polis-enabled Elxis user at different time instances and the corresponding entries in the user's Polis-agent audit.

6.6 Conclusions

The evaluation of the Polis prototype and the Polis case study proved the feasibility of the main Polis approach and confirmed the features of the Polis approach discussed in Section 4. A comparison of the features that are available to Polis-enabled individuals in contrast to conventional/non Polis-enabled individuals of the Elxis CMS-based application is shown in Table 1. The table compares the Polis approach to the current practice in personal data management. To this end, it highlights a set of important advantages and disadvantages for Polis-users of the Polis-enabled Elxis CMS. The comparison should apply to a wide range of possible Polis-enabled e-business applications.

		Polis-enabled user(s)	Conventional user(s)
1	awareness	the individual is aware of any access to her data	the individual receives no information
2	data opt-out	trivial	the individual has to contact the Elxis shop
3	specifying policies	the individual can specify her own data access policies or adopt proven policy templates	the individual has to rely on the company specified privacy policy
4	security	the individual is not affected by most types of attacks on the Elxis shop	any data leakage from the Elxis shop may affect the individual
5	effort	the individual has to manage her own data	the individual simply gives away her data
6	delay	the data is retrieved from the Polis-agent of the individual	data is retrieved from the company database
7	data cleansing	trivial (the data is retrieved from the Polis-agent)	data entry errors may occur
8	data update	trivial (happens implicitly)	the individual has to go through a proprietary procedure

Table 1: Advantages and disadvantages for Polis-users of the Polis-enabled Elxis CMS instance

As noted in Section 1, to the authors knowledge, there is no approach comparable to Polis for managing personal data wholly at the owner's side. The closest work is the approach of [Lioudakis et al., 2007] where the data of individuals resides in a Discreet-box located at the shop side. However, this case differs from Polis in that data is not at the premises of the individual. Furthermore, it is more complicated than Polis and unfortunately we were unable to obtain an implementation for evaluation/comparison purposes. Other technologies, like IBM's Tivoli Privacy Manager, are concerned with managing customer data within a company, but do not hand over the control to the actual customer. Such technologies can complement the Polis approach (by increasing the accountability of the company's data practices) but are not a substitute for it.

Finally, the evaluation of Polis also revealed important improvements that are possible for the Polis-agent and the accompanying tolls. One improvement concerns the usability of Polis. In the relevant literature it is pointed out that the use of an appropriate Graphical User Interface (GUI), that clearly demonstrates concepts for expressing privacy preferences, is very supportive for effective use of privacy protection systems [Lederer et al., 2005, Palen and Dourish, 2003, Jentzsch, 2007, Ackerman, 2000]. A first tool that we have created is a Polis Add-on for Firefox, which, amongst other features, alerts the user each time her personal data are requested (Figure 9).

7 Discussion

In this work, we design, implement and evaluate Polis, a personal data management framework which embodies a fundamental privacy principle: Personal data of individuals reside only at their

(a) Internet browsing moment of a Polis-user.

(b) The user's data have been requested.

Figure 9: The Polis Add-on for Firefox alerts the user when her personal data are requested

side. Polis aims at making storage of personal data unnecessary for contemporary online transactions to work efficiently. This way, users will be able to monitor and limit the distribution of their personal data, according to their needs and preferences. Furthermore, the safety of stored personal data is enhanced and personal data accuracy is ensured.

In conclusion, this work demonstrates the fact that it is possible to deploy a privacy-enhancing prototype like Polis, in order to achieve significant privacy protection, in the current electronic world. We cannot expect Polis to become a panacea for all kinds of privacy problems. However, we believe that Polis has more advantages than disadvantages compared to current practices for personal data management. Finally, it is very encouraging that given one basic assumption, the transition to a personal data protecting way of conducting online transactions, can be natural and smooth.

References

[Ackerman, 2000] Ackerman, M. (2000). The intellectual challenge of cscw: The gap between social requirements and technical feasibility.

[Acquisti, 2004] Acquisti, A. (2004). Privacy and security of personal information: Technological solutions and economic incentives. In Camp, J. and Lewis, R., editors, *The Economics of Information Security*, pages 165–178. Kluwer.

[Aggarwal et al., 2005] Aggarwal, G., Bawa, M., Ganesan, P., Garcia-Molina, H., Kenthapadi, K., Motwani, R., Srivastava, U., Thomas, D., and 0002, Y. X. (2005). Two can keep a secret: A distributed architecture for secure database services. In *CIDR*, pages 186–199.

[Agrawal et al., 2002] Agrawal, R., Kiernan, J., Srikant, R., and Xu, Y. (2002). Hippocratic databases. In *VLDB '2002: Proceedings of the 28th international conference on Very Large Data Bases*, pages 143–154. VLDB Endowment.

[Anderson, 2007] Anderson, R. (21 November 2007). U.k. government loses personal data on 25 million citizens. *EDRI-gram*, Number 5.22.

[Bangerter et al., 2004] Bangerter, E., Camenisch, J., and Lysyanskaya, A. (2004). A cryptographic framework for the controlled release of certified data. In Christianson, B., Crispo, B., Malcolm,

J., and Roe, M., editors, *Security Protocols Workshop*, volume 3957 of *LNCS*, pages 20–42. Springer.

[Bohrer and Holland, 2000] Bohrer, K. and Holland, B., editors (2000). *Customer Profile Exchange (CPExchange) Specification*. IDEAlliance. http://www.idealliance.org/cpexchange.

[Cha and Joung, 2003] Cha, S.-C. and Joung, Y.-J. (2003). From p3p to data licenses. In *Privacy Enhancing Technologies*, pages 205–222.

[ConsumerReports, 2006] ConsumerReports (2006). C.r. investigates: Your privacy for sale. *Consumer Reports*, 71(10):41. http://www.accessmylibrary.com/coms2/summary_0286-29062087_ITM.

[Crosby, 2008] Crosby, J. (2008). Challenges and opportunities in identity assurance. Technical report, HM Treasury, United Kingdom. http://www.hm-treasury.gov.uk/d/identity_assurance060308.pdf.

[Dingledine et al., 2004] Dingledine, R., Mathewson, N., and Syverson, P. (2004). Tor: The second-generation onion router.

[DISCREET, 2008] DISCREET (2008). Discreet service provision in smart environments. FP6-2004-IST-4 contract no. 27679. http://www.ist-discreet.org/.

[Efraimidis et al., 2008] Efraimidis, P., Drosatos, G., Nalbadis, F., and Tasidou, A. (2008). Towards privacy in personal data management. *PCI 2008*, pages 3–7.

[Fahrmair et al., 2005] Fahrmair, M., Sitou, W., and Spanfelner, B. (2005). Security and privacy rights management for mobile and ubiquitous computing. In *Workshop on UbiComp Privacy*.

[Goldberg, 2000] Goldberg, I. (2000). *A Pseudonymous Communications Infrastructure for the Internet*. PhD thesis, Univ. of California at Berkeley.

[Goldberg, 2007] Goldberg, I. (2007). Privacy-enancing technologies for the internet iii: Ten years later. In Acquisti, A., Gritzalis, S., Lambrinoudakis, C., and di Vimercati, S., editors, *Chapter 1 of Digital Privacy: Theory, Technologies, and Practices*.

[Golle et al., 2006] Golle, P., McSherry, F., and Mironov, I. (2006). Data collection with self-enforcing privacy. In *CCS '06: 13th ACM conference on Computer and communications security*, pages 69–78, New York, NY, USA. ACM.

[Gritzalis, 2004] Gritzalis, S. (2004). Enhancing web privacy and anonymity in the digital era. *Information Management and Computer Security*, 12(3):255–287.

[Hong, 2005] Hong, J. (2005). *An Architecture for Privacy-Sensitive Ubiquitous Computing*. PhD thesis, University of California at Berkeley, Computer Science Division, Berkeley.

[Jäppinen, 2004] Jäppinen, P. (2004). *ME - Mobile Electronic Personality*. PhD thesis, Lappeenranta University of Technology, Finland.

[Jentzsch, 2007] Jentzsch, N. (2007). *Theory of Information and Privacy*, pages 7–59. Springer.

[Karjoth and Schunter, 2002] Karjoth, G. and Schunter, M. (2002). A privacy policy model for enterprises. In 15th IEEE Computer Security Foundations Workshop.

[Karjoth et al., 2002] Karjoth, G., Schunter, M., and Waidner, M. (2002). The platform for enterprise privacy practices - privacy enabled management of customer data. In 2nd Workshop on Privacy Enhancing Technologies (PET).

[Katsikas et al., 2005] Katsikas, S., Lopez, J., and Pernul, G. (2005). Trust, privacy and security in e-business: Requirements and solutions. In *Panhellenic Conference on Informatics*, pages 548–558.

[Kleinberg et al., 2001] Kleinberg, J., Papadimitriou, C., and Raghavan, P. (2001). On the value of private information. *TARK: Theoretical Aspects of Reasoning about Knowledge*, 8.

[Korba and Kenny, 2003] Korba, L. and Kenny, S. (2003). Towards meeting the privacy challenge: Adapting drm. In *Digital Rights Management (LNCS 2696/2003)*, pages 118–136. Springer Berlin / Heidelberg.

[Laudon, 1996] Laudon, K. (1996). Markets and privacy. *Commun. ACM*, 39(9):92–104.

[Lederer et al., 2005] Lederer, S., Hong, J., Dey, A., and Landay, J. (2005). Personal privacy through understanding and action: Five pitfalls for designers. In *Designing Secure Systems That People Can Use*, pages 421–445.

[Lee and Stamp, 2008] Lee, H.-H. and Stamp, M. (2008). An agent-based privacy-enhancing model. *Information Management & Computer Security*, 16(3):305–319.

[Lioudakis et al., 2007] Lioudakis, G., Koutsoloukas, E., Dellas, N., Tselikas, N., Kapellaki, S., Prezerakos, G., Kaklamani, D., and Venieris, I. (2007). A middleware architecture for privacy protection. *Comput. Networks*, 51(16):4679–4696.

[Millett et al., 2001] Millett, L., Friedman, B., and Felten, E. (2001). Cookies and web browser design: toward realizing informed consent online. In *SIGCHI Conference on Human factors in computing systems*, pages 46–52, New York, USA. ACM.

[Mulligan and Schwartz, 2000] Mulligan, D. and Schwartz, A. (2000). Your place or mine?: privacy concerns and solutions for server and client-side storage of personal information. In *CFP '00: Proceedings of the tenth conference on Computers, freedom and privacy*, pages 81–84, New York, NY, USA. ACM Press.

[Palen and Dourish, 2003] Palen, L. and Dourish, P. (2003). Unpacking "privacy" for a networked world. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 129–136, New York, NY, USA. ACM.

[Polis, 2008] Polis (2008). The polis project. http://polis.ee.duth.gr.

[PRIME, 2008] PRIME (2008). Privacy and identity management for europe. EC Contract No. IST-2002-507591. https://www.prime-project.eu/.

[Rahm and Do, 2000] Rahm, E. and Do, H. H. (2000). Data cleaning: Problems and current approaches. *IEEE Bulletin of the Technical Committee on Data Engineering*, 23(4).

[Reuters, 2008] Reuters (2008). Axel springer hit by new german data leak scandal. http://www.reuters.com/article/internetNews/idUSTRE49H1GH20081018.

[Salim et al., 2007] Salim, F., Sheppard, N., and Safavi-Naini, R. (2007). Enforcing p3p policies using a digital rights management system. In Borisov, N. and Golle, P., editors, *Privacy Enhancing Technologies*, volume 4776 of *LNCS*, pages 200–217. Springer.

[Samuelson, 2000] Samuelson, P. (2000). Privacy as intellectual property? *Stanford Law Review*, 52:1125.

[Tasidou et al., 2009] Tasidou, A., Efraimidis, P., and Katos, V. (2009). Technical Report LPDP-2009-01, Democritus University of Thrace, Greece. http://utopia.duth.gr/∼pefraimi/research/data/2009FairTrades.pdf.

[Varian, 1996] Varian, H. (1996). Economic aspects of personal privacy. U.S. Dept. of Commerce, Privacy and Self-Regulation in the Information Age.

[Vassiliadis et al., 2002] Vassiliadis, P., Simitsis, A., and Skiadopoulos, S. (2002). Conceptual modeling for etl processes. In *5th ACM international workshop on Data Warehousing and OLAP*, pages 14–21, New York, USA. ACM.

[W3C, 2002] W3C (2002). The platform for privacy preferences 1.0 (p3p1.0) specification. http://www.w3.org/TR/P3P.