Dynamic Privacy Enforcer: A Trusted Third-party Framework to Provide Personalization in the Presence of Privacy Concerns

Ramnath K. Chellappa

Goizueta Business School,
Emory University,
Atlanta GA 30322
ram@bus.emory.edu

Ravinder Dharmapuram

Shopzilla
dharmapu@gmail.com

Rahul Hampole

ValueClick Media
hampole@gmail.com

ABSTRACT

Online personalization depends upon consumers' willingness to share personal and preference information and hence is related to their concern for privacy. There have been substantial technological developments in data mining techniques for personalization and significant legal advances in dealing with consumer concerns of piracy. However, there is little research that confronts both the consumers' need for personalization and their concern for privacy simultaneously. Our research extends the static P3P paradigm to construct a trusted third party based technology protocol that allows for real-time personalization while taking into account the consumers' concern for privacy.

Author Keywords

Personalization, privacy, P3P, trusted third-party.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous. K4.4. Electronic Commerce, K.4.1 Public Policy Issues.

INTRODUCTION

The ability to acquire and process consumer information online has provided Web based vendors with the ability to personalize products and services accompanying the product browsing and purchasing experience. However, the ability to personalize services online is dependent upon two main factors: 1. the data mining and profiling technology of the vendor and 2. the consumers' willingness to share information. While consumers value the convenience provided by personalization, recent research has found that privacy concerns online can inhibit them from providing the personal and preference information [2]. Information available through a Web based browsing and shopping process is far more than that acquired in the physical world and thus allows vendors to construct reasonably accurate consumer profiles. For example, while BestBuy physical store does not send an employee along with every customer to monitor what they are looking at; BestBuy.com indeed does that through cookies and other tracking mechanisms. This allows BestBuy.com to personalize any subsequent visit by consumers and at the same time raises their privacy concerns.

There is a significant body of literature in data mining on various techniques for personalization [8, 10, 12]. Most of these techniques acquire click-stream data and augment them with demographic and other market research based information. The primary focus of these techniques is to build consumer profiles so as to proactively personalize services for them. While some data mining techniques are site-centric, others integrate data from many different vendor sites and consumer visits to these vendors [9]. Similarly simple rule-based filtering techniques rely mostly on an individual consumers' data and rules constructed based on market research, while collaborative filtering techniques have ability to integrate behavior of similar consumers [1]. Consumers are not necessarily averse to providing or letting vendors acquire some of this information. For example, when a consumer conducts a purchase transaction with Amazon.com, he is willing to provide his name, shipping address and other information related to the fiscal transaction. However, the consumer may be concerned that Amazon is able to relate his interests in Target (that is a part of Amazon's shopping network) by integrating his mere browsing actions with the personally identifiable information provided during his purchase transaction. Note that in the physical world, consumers only leave paper trails if they actually conduct a credit-card or check based transaction.

However, recent research also finds that consumers indeed value online personalization, e.g. Yahoo!'s personalization of weather, sports and other information or Amazon.com's delivery of personalized coupon for a product that a consumer put in his shopping cart but did not purchase [2]. Thus doing away with personalization completely cannot be the solution to allaying consumer concerns of privacy. Other research also suggests that consumers accept loss of privacy if it accompanies some benefit [6], and they engage in a "privacy calculus" [3] or a cost benefit analysis to determine the amount of information they are willing to share. In the absence of personalization techniques that incorporate this privacy calculus, we propose a technological solution based on well known statistical disclosure avoidance techniques. We develop our solution on top of the well known P3P paradigm and build a trusted-third party called the Dynamic Privacy Enforcer. Our research demonstrates the feasibility of the proposed technology using standard Web, P3P and XML architecture.

Current mechanisms for privacy protection

Privacy has largely been a subject for researchers in ethics, sociology, law and more recently information systems and marketing [2, 7]. In the online context, the consumer protection of privacy has primarily been handled through self-regulation. Since 1995, the Federal Trade Commission has served as the leading governmental body in setting online privacy standards, and it has suggested a set of fair information practices that online firms are supposed adhere to. Generally, the above practices are implemented by Web sites in two visible ways; i. they put up a policy statement indicating how the firm will follow the fair information practices, and ii. they form relationships with third-parties such as TRUSTe, CPA WebTrust, and BBBOnline Privacy and display their seal of approval. In spite of these guidelines (where over 95% of firms have a privacy policy and over 36% have a privacy seal of some sorts) both the academic and trade press continues to report that both Internet users and non-users perceive that using the Internet creates risks to the preservation of individual privacy. Thus one could argue about the effectiveness of the legal mechanisms in place.

From a technological perspective, the W3C group (World Wide Web Consortium, the leading body on Web standards) introduced the P3P (Platform for Privacy Preferences) protocol. P3P essentially provides the technical standard for presenting privacy policies. It comprises of the following elements: POLICY, DISCLOSURE, ASSURANCE-GROUP, STATEMENT-BLOCK. This allows for a Web site to easily write its policy such that an average consumer using his Web browser can automatically fetch the policy, parse it and present it to the user so as to enable her to actively decide on the elements of her personal information that will be collected. P3P is emerging as the default protocol for privacy protection and being implemented in the new browsers in the market. However a major problem with P3P is that not unlike the earlier policy notices, this framework is also overly dependent on self-regulation. It is still up to the vendor to present the policy and honor the use of customer information. In this research we explore an alternative mechanism that will provide consumers with greater control over the usage of their preference information without having to completely sacrifice their personalization needs.

DYNAMIC PRIVACY ENFORCER (DPE)

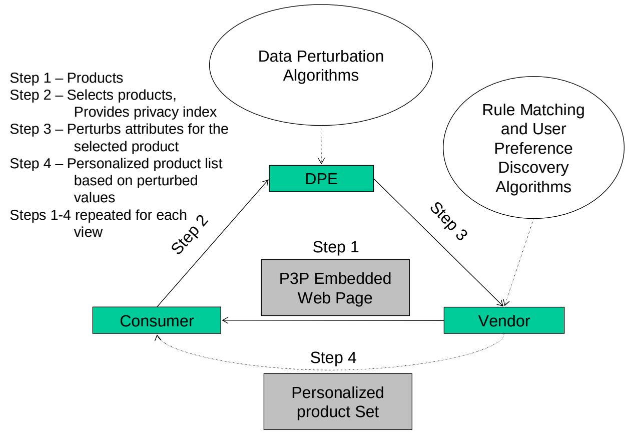

Data Perturbation Algorithms
Step 1 – Products
Step 2 – Selects products, Provides privacy index
Step 3 – Perturbs attributes for the selected product
Step 4 – Personalized product list based on perturbed values
Steps 1-4 repeated for each view
Rule Matching and User Preference Discovery Algorithms
DPE
Step 2
Step 3
Step 1
P3P Embedded Web Page
Consumer
Vendor
Step 4
Personalized product Set

Figure 1: Conceptual Data Flow

The ability of an online vendor to provide personalization is a function of both the data (that constitutes a user profile) and the personalization algorithms (e.g. collaborative filtering, rule matching) used. The basic premise behind a personalization algorithm is that based on past purchase and browsing behavior of customer(s) and their cohorts, new products and services that are likely to be closer to customers' true preferences, can be recommended on subsequent visits. A goal of our mechanism design is to not put any undue requirements on vendors with regards to their choice of algorithms. Thus the only way to provide privacy to the consumer is through manipulation of the data that the vendor received about the consumer. Clearly, data on the consumers' credit card, shipping address and other information necessary for completing a business transaction cannot be shielded from a vendor; else the vendor cannot deliver the good purchased. Thus our consumer profile is defined by two types of information attributes: one set is necessary to complete fiscal transactions in case of a purchase (e.g. credit card details, address), and the second set consists of behavioral attributes that the vendor is sensitive about.

In our framework, the DPE is a trusted third party Web server such that the consumers' preference information is routed through the DPE to the vendor. The DPE Web server contains perturbation methods that manipulate the data before they are sent to the vendor. The manipulated data should have the following properties: i) The vendor should not be able to reconstruct the true profile from the manipulated data and ii) The consumers' perturbed profile should not result in a recommendation that is too different from what the consumers' true profile would have recommended.

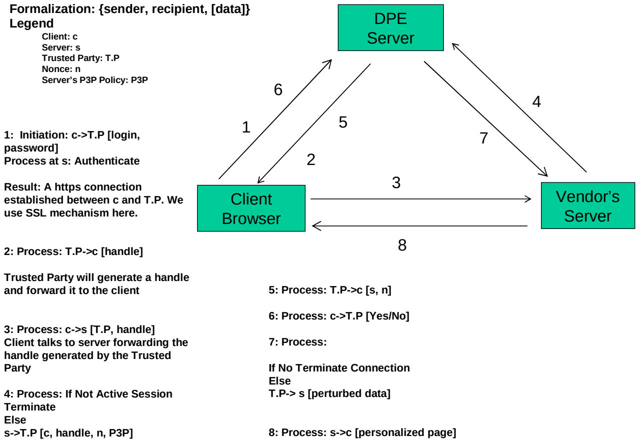
An example illustration

To illustrate the DPE data flow concept, consider the case of movies and consumers' preferences and privacy concerns therein. Let there be G genres, and let a consumer i's true preferences for movies be given by a vector θ_{ig}. A movie j might belong to one or more genre given by $\gamma_{jg}|_{g=1,\ldots,G}$, such that $\gamma_{jg}=1$, if the movie i belongs to genre g and $\gamma_{jg}=0$ if otherwise. Based on browsing and other self-disclosed behavior, a vendor can come up with an ordinal ranking $y_{ij}\in\{0,\ldots,M\}$ for a consumer's preference for a given movie j such than higher rankings correspond to strong preferences. There are number personalization methods that a vendor can use to infer true preferences θ_{ig} from observed behavior y_{ig}. As a result of their chosen algorithms, vendors will recommend a product list p_{ig} that is expected to closely match consumer preferences. Consumers are concerned because they may be sensitive to revealing their browsing behavior and/or preferences regarding certain genres, e.g., y_{i4}, y_{i6}. In our mechanism, the consumer provides his privacy preferences to DPE, such that the DPE will transform $y_{ig}\to y'_{ig}$ taking into account privacy concerns for genres 4 and 6. A number of perturbation algorithms are possible, including those described in [11], wherein the perturbed profile of the consumer i is derived from a combination of the consumer's own true profile and the profile of an average consumer in whole population. Thus when the vendor applies his personalization algorithm to map true preferences of the consumer, he will infer θ'_{ig} instead of θ_{ig}, and hence will recommend a product list p'_{ig} instead of p_{ig}.

Protocol details and implementation

The mechanism is implemented using a Browser Helper Object (BHO) such as a browser toolbar, and consists of eight steps. The toolbar is part of the client browser and is first downloaded by the consumer when registering with the DPE.

1. A client should have an account established prior to any transaction. The client then initiates connection with the DPE and authenticates itself through a secure connection using SSL.
2. The DPE generates a random handle and sends it to the client, thus establishing an active session; this handle is used by the server to identify the client. The handle is a random number mapped to a client id, thus the client id is never revealed to the vendor's server. A simple hash technique such as Hash (client id + current time) -> random number, should ensure unique handles.
3. Client generates a nonce n and sends {c,n,T.P,handle} to the vendor. The nonce is a random number that maps the client-vendor mapping.
4. The vendor sends a request to the trusted party as {c,s,n,handle,P3P}, where it presents its P3P policy and at this stage the trusted party checks to see if the client's session is active. If YES, we proceed to step 5 else cancel and terminate connection.
5. The DPE forwards request to client as {s,n} so as to check and ensure that a server is not performing a replay attack, also refuting a man-in-the-middle attack. The main idea is that the trusted party wants to verify that the client indeed has forwarded the handle to the server.



Formalization: {sender, recipient, [data]}
Legend
Client: c
Server: s
Trusted Party: T.P
Nonce: n
Server's P3P Policy: P3P

1: Initiation: c->T.P [login, password]
Process at s: Authenticate

Result: A https connection established between c and T.P. We use SSL mechanism here.

2: Process: T.P->c [handle]

Trusted Party will generate a handle and forward it to the client

3: Process: c->s [T.P, handle]
Client talks to server forwarding the handle generated by the Trusted Party

4: Process: If Not Active Session
Terminate
Else
s->T.P [c, handle, n, P3P]

5: Process: T.P->c [s, n]

6: Process: c->T.P [Yes/No]

7: Process:

If No Terminate Connection
Else
T.P-> s [perturbed data]

8: Process: s->c [personalized page]

Figure 2: Protocol Illustration

6. The client checks to see if {s,n} is valid and generates a response which is an YES or a NO based on whether the server-nonce pair is in the client's database.
7. If the trusted party receives an YES, it continues else the session is terminated. At this point the trust party performs the perturbation as per user policy. The user preferences are either stored by the trusted party during

account setup or can be dynamically passed on to it. Essentially it acts as the User Agent. Clients may have specified levels of trust in different service providers (vendors) and the perturbation is based on the combination of trust, user preferences and user P3P policy. Trust level is initially set to a default for all vendors and is updated as user specifies this level.

8. The vendor's server executes the personalization operation on the perturbed preferences and creates a links to personalized product list and sends it to the client. To repeat the process, we move directly to step 3 as steps 1 and 2 are no longer necessary when a client already has a handle from the T.P

When a selection to view a product page is made by the customer, the associated genre of the product is determined by a DOM parser by parsing through the XML which is associated with the product page. Instead of providing the true selection choices directly to the vendor, a set of choices where some of which are a result of DPE's perturbation mechanism, are sent to the vendor. The toolbar also stores the exact sets of URL that the customer requested along with the associated perturbation-based URL's. Thus the vendor sends all of the associated data and XML (for both the requested and non-requested) back to the consumer. These steps ensure that the vendor is never able to accurately determine the customer's choice of product pages while yet providing the customer the requested data.

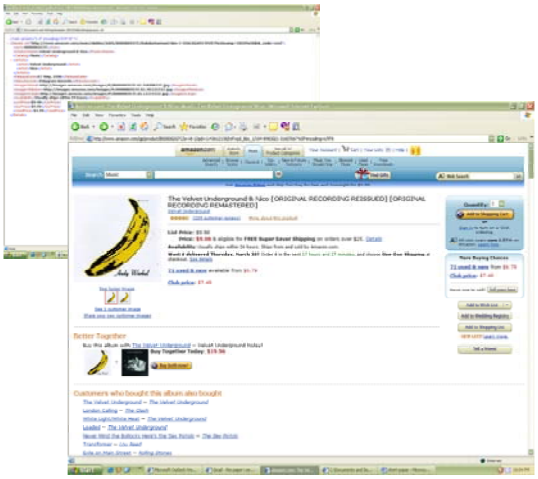
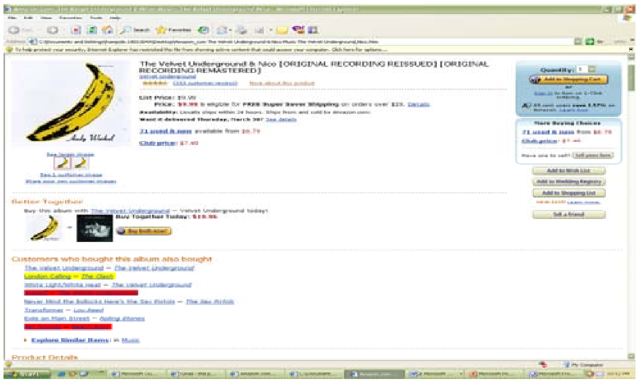


Figure 3: Basic page with XML information

In figure 4, the yellow links correspond to the true browsing behavior of the consumer while the combination of the red and yellow links are the vendor's the red links were the additional choices provided by the vendor as part of the perturbation. Note that only the browsing behavior (through URL clicks) is routed through the DPE; information required for completing a fiscal transaction will continue to be sent directly to the vendor.



The Velvet Underground & Nico [ORIGINAL RECORDING REISSUED] [ORIGINAL RECORDING REMASTERED]

Better Together

Customers who bought this album also bought

The Velvet Underground – The Velvet Underground
London Calling – The Clash
White Light/White Heat – The Velvet Underground
Never Mind the Bollocks Here's the Sex Pistols – The Sex Pistols
Transformer – Lou Reed
Exile on Main Street – Rolling Stones

Explore Similar Items: in Music

Product Details

Figure 4: Recommendations based on perturbed profile

DISCUSSION

The purpose of our research was to create a mechanism that satisfied the following goals: 1. Consumer concern for privacy was not to be managed through self-regulation alone, 2. No undue technological requirements should be placed on the consumer or vendor, and 3. Privacy protection cannot be at the expense of complete loss of personalization benefits. In our model, the consumer proactively decides on the sub-set of information that he does not want to reveal to the vendor. Since this information is observed only by the DPE and only perturbed information is sent to the vendor, privacy is protected to the extent that the consumer wants it. As far as vendor is concerned, he will continue to treat this information as constituting the consumers profile and hence will continue to personalize products and services for the consumer. No specific changes need to be made to the personalization algorithm itself. There are many techniques available for perturbation, including data swapping and cell suppression [4, 5]. These techniques create a sub-set of data that is transformed by masking, introducing noise, grouping or truncating so as to not reveal the original data, and yet be reasonably accurate in collectively representing the original data.

Our research has important implication for the development of a trust-federation in the online environment. Intermediaries such as TrustE and WebCPA that provide a privacy seal are at best static agents and they simply periodically check the Web pages of their clients to verify if policies are updated to reflect the current legal framework. They come into action only if the FTC or some other body disputes the information collection/usage by a vendor. As envisaged by us, the trusted-third party can play a far more interactive and important role in the preservation of consumer's privacy concerns.

REFERENCES

1. Adomavicius, G., and Tuzhilin, A. An Architecture of e-Butler- A Consumer Centric Online Personalization Mechanism. International Journal of Computational Intelligence and Applications, 3, 2, (2002), 313-327.
2. Chellappa, R.K., and Sin, R. Personalization versus Privacy: An Empirical Examination of the Online Consumer's Dilemma. Information Technology and Management, 6, 2-3, (2005), 181-202.
3. Culnan, M.J., and Bies, R.J. Consumer Privacy: Balancing Economic and Justice Considerations. Journal of Social Issues, 59, 2, (2003), 104-115.
4. Duncan, G.T., and Pearson, R.B. Enhancing access to microdata while protecting confidentiality: prospects for the future (with discussion). Statistical Science, 6, (1991), 219-239.
5. Fienberg, S.; Steele, R.; and Makov, U. Statistical Notions of Data Disclosure Avoidance and their Relationship to Traditional Statistical Methodology: Data Swapping and Loglinear Models. Working Paper, (2001), -.
6. Laufer, R.S., and Wolfe, M. Privacy as a Concept and a Social Issue: A Multidimensional Developmental Theory. Social Issues, 33, 3, (1977), 22-24.
7. Miyazaki, A.D., and Fernandez, A. Internet Privacy and Security: An Examination of Online Retailer Disclosures. Journal of Public Policy & Marketing, 19, 1, (2000), 54-61.
8. Mobasher, B.; Cooley, R.; and Srivastava, J. Automatic Personalization Based on Web Usage Mining. Communications of the ACM, 43, 8, (2000), 142-151.
9. Padmanabhan, B.; Zheng, Z.; and Kimbrough, S.O. Personalization from Incomplete data: What you don't know can Hurt, in Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, 2001.
10. Raghu, T.S.; Kannan, P.K.; Rao, H.R.; and Whinston, A.B. Dynamic Profiling of Consumers for Customized Offerings Over the Internet: A Model and Analysis. Decision Support Systems, 32, 2, (2001), 117-134.
11. Scott, S.L., and Chellappa, R.K. Balancing Personalization and Privacy. (2006), working paper.
12. Shahabi, C., and Banaei-Kashani, F. Efficient and Anonymous Web-Usage Mining for Web Personalization. INFORMS Journal on Computing, 15, 2, (2003), 123-147.