P3P Position Paper: Agents and P3P
W3C Workshop on the Future of P3P
November 12-13, 2002

Jack Humphrey
Development Manager, Data Acquisition
Coremetrics, Inc.

Introduction

Coremetrics is a leading web services provider of marketing analytics solutions, empowering companies to develop and act upon a comprehensive understanding of all visitor and customer interactions within their online channels. Coremetrics provides these solutions to clients like Ann Taylor, The Columbia House Company, CompUSA, Eddie Bauer, Motorola, Newport News, Nortel Networks, Spiegel Catalog, Victoria's Secret and Wal-Mart.

Coremetrics has been an early adopter in applying the P3P specification to our leading-edge data collection technology. As an agent for our clients, we have faced significant challenges in accurately representing data collection policies. We contend that enhancements and clarifications in the P3P specification would allow agents to offer more accurate P3P policies, thereby improving the ability of end-users to understand what data is collected by the sites they visit and how it is used.

This document assumes familiarity with the P3P 1.0 specification as well as the basics of XML syntax.

For the purposes of this document, an *agent* is an entity that collects data on behalf of another entity, known as the *client* of that agent. The agent does not use the data for any purpose other than to provide it (in raw or summarized form) to the client. The terms *agent* and *client* tend to be overloaded, so we recommend that proper terminology be agreed upon and clearly defined in the vocabulary section of the P3P specification.

Agents in the Third-Party Context

For agents who collect data in a third-party context on behalf of a client (typically through off-site image requests or form action handlers), P3P 1.0 does not clearly specify how this relationship should be indicated in the data policies. Several examples in the specification, which are expanded upon in other published materials on P3P, explain that the agent's policy reference file should refer to a policy published by the client.

For example, consider ProdRegService, a web service that processes form submittals from product registration forms on client web sites. AcmePhones uses this web service and publishes a P3P policy (http://www.acmephones.com/p3p-policy.xml#prodreg) that explains their use of the collected data. The policy reference file for ProdRegService (http://www.prodreg.com/w3c/p3p.xml) refers to that policy for HTTP requests from AcmePhones.

Implications for User Agents

We assume that the implication of this approach is that user agents should not apply third-party restrictions to a third-party request that is covered by a first-party policy. Suppose that in the example above, a user has chosen to receive warnings when submitting contact information to a third-party (but not to a first-party). When that user submits the production registration form at acmephones.com, the user agent should infer that prodreg.com is an agent of the first-party site and issue no warning.

If this assumption is incorrect, then it becomes very difficult for agent services to avoid running afoul of well-meaning user agents, and the P3P specification should offer an alternate means of denoting agent status. If correct, the specification should clarify the expected user agent behavior; however, creators of user agents might find such a requirement unwieldy to implement.

Clients Without Policies

An agent may be requested to collect data on behalf of a site that has not yet implemented a P3P policy. In this case, it may not be feasible or advisable for the agent to simply provide a policy to the client for publishing on their site. One alternative, then, is for the agent to publish a policy on behalf of that client, which covers all possible uses of the data collected, regardless of whether or not the client uses the data in all of those ways.

However, that approach does not allow a user agent to determine that a third-party HTTP request should be exempted from third-party restrictions, which suggests the need for the ability to denote agent status *inside* a P3P policy.

Cookies and Compact Policies

User agents like Internet Explorer 6 that depend entirely on compact policies to determine acceptability of cookies have no way to detect that third-party restrictions should not be applied to agents in the third-party context.

This limitation arises because the compact policy is just a condensed form of the P3P policy, which has no internal way to denote agent status. The P3P specification should therefore include a means by which a compact policy can indicate that a third-party site is attempting to set a cookie on behalf of the first-party site.

Recommendations

We propose that the P3P specification be modified to:

- Allow agent status to be specified inside a P3P policy (and its derived compact policy) instead of externally in a policy reference file.
- Require user agents not to apply third-party restrictions to agents operating in the third-party context.

These recommendations are intended to be a starting point for discussion. More work will be required to define the proper solution.

Agent Status in Policies

Several enhancements to the P3P specification would be required in order to denote agent status within a P3P policy.

In the `ENTITY` element of the policy, a new `business.type` data element (in the business dataset) could be allowed to contain the value "agent".

The `<ours>` recipient element is currently specified to apply to both entities and their agents. A user agent examining such a policy has no way of determining whether the service collecting the data is doing so on its own behalf or on behalf of a client (or both). A new recipient element (e.g. "`<client>`") could represent clients on whose behalf the collecting site is acting as an agent. For compact policies, this new element could be mapped to a new three-letter code (e.g. "`CLI`").

The specification could provide a mechanism by which an agent can refer to its client's policy, hosted on the client's site. This mechanism might be a URI attribute on a new client recipient element or somewhere in the `ENTITY` element. An appropriate `HINT` element in the policy reference file would allow user agents to more efficiently load and process the necessary policies. It might also be appropriate to add a new purpose element (e.g. "`<client-uses>`") that would indicate that data is used according to the referenced policy.

Such a reference may be more difficult to represent in a compact policy. One approach would be to recommend or require that user-agents that encounter the code for the new purpose element (e.g. "`CLU`" for "`<client-uses>`") should load the appropriate full P3P policies to discover the client's intended uses. Another approach would be to add a new value to the P3P HTTP header that allows the agent service to specify the domain of the first-party client site in the HTTP response.

In addition to the benefits discussed in the previous section, these enhancements would also allow agents to unambiguously present their own data policies. For example, an agent might wish to indicate that it uses collected data for administrative purposes only and retains data for a certain amount of time, and then defer the specification of other uses of the data to the client to whom it is delivered. These enhancements would thereby create more transparency into the use of data in agent-client arrangements, which is beneficial to end users who can then make more informed decisions about their participation in data collection.

Application of Third-Party Restrictions by User Agents

The P3P specification should be very clear on the expected behavior of user agents with regard to agents in the third-party context. We recommend that user agents be required not to apply third-party restrictions to entities that are indicated (explicitly or implicitly) to be agents operating on behalf of the first-party site.

Summary

As web services become more pervasive, the number of agents that collect data on behalf of clients will continue to grow. The P3P specification should be enhanced and clarified to enable these agents and their clients to provide more accurate data policies. We hope that the observations and recommendations made in this document may serve to stimulate discussion on how to properly achieve this goal.