Proposal: Domain Relationships

Prepared for W3C P3P 1.1 Working Group
Jack Humphrey (jhumphrey@coremetrics.com)
July 2003

Introduction

As part of the P3P 1.1 effort, this document describes modifications to the P3P specification that would allow user agents to recognize when hosts in different domains are owned by the same entity.

These modifications would allow user agents to more intelligently apply privacy preferences, addressing implementation issues that have plagued many P3P deployments.[1]

This document contains an overview of the proposed modifications but not the specification-level details of the modifications, which will be provided in a subsequent document.

Limitations of Current Policy Reference File Syntax

Consider the web sites example.com and forinstance.com, which are owned by the same company and share some web site content, hosted on example.com. Some of that content includes "internal" banner ads that are generated by a servlet on example.com and promote certain features on example.com. Both sites are owned by the same company and wish to deploy a single P3P policy that describes their single corporate policy on data collection and usage. We refer to this situation as a "same entity" relationship, as in "sample.com is an owned by the same entity as example.com."

A simple policy reference is deployed in the well-known location (/w3c/p3p.xml) on example.com. It looks like this:

```
<META xmlns="http://www.w3.org/2002/01/P3Pv1">
 <POLICY-REFERENCES>
    <POLICY-REF about="/p3p/policy.xml#corporate">
      <INCLUDE>/*</INCLUDE>
    </POLICY-REF>
 </POLICY-REFERENCES>
</META>
```

forinstance.com is configured to return the HTTP header

[1] http://www.w3.org/2002/p3p-ws/pp/coremetrics.pdf
"Agents and P3P"position paper presented to the W3C Workshop on the Future of P3P

```
P3P: policyref="http://www.example.com/w3c/p3p.xml"
```

When a web browser visits a page on forinstance.com, the user agent applies the policy at the URI http://www.example.com/p3p/policy.xml#corporate. If an image on that page is served by example.com, the same policy would be applied to that image request, after looking up the policy reference file at the well-known location.

Since both the page request and the image request are covered by the same privacy policy, the user agent should be able to understand that any data collected (via cookies or otherwise) is being collected and used by the same entity, and therefore no extra privacy restrictions should be applied for the image request to the different domain.

If the hosts were flipped, and example.com was serving the page and forinstance.com the image, the same logic should hold with the exact same P3P deployment. This possibility brings up an interesting point. Third-party sites should not be able to "spoof" being covered by the first party privacy policy if they are not. If, in our example, forinstance.com was not part of the same company, and had different data collection and usage policies, this mechanism could allow it to claim to be covered by the example.com policy when in fact it is not. Without an additional mechanism in P3P, the user agent would not be capable of verifying the "same entity" relationship and the burden would fall on example.com to be sure that third parties referenced in its site were referencing appropriate policies.

Proposed Additions to Policy Reference Files

The proposed mechanism allows sites to declare hosts owned by the same entity in the policy reference file. These declarations would allow user agents to recognize and verify "same entity" relationships automatically.

Since forinstance.com references example.com's policy reference file, that file would have an additional KNOWN-HOSTS section:

```
<META xmlns="http://www.w3.org/2002/01/P3Pv1">
 <POLICY-REFERENCES>
    <POLICY-REF about="/p3p/policy.xml#corporate">
      <INCLUDE>/*</INCLUDE>
    </POLICY-REF>
 </POLICY-REFERENCES>
 <KNOWN-HOSTS>
    <HOST name="*.forinstance.com" entity-type="SAME"/>
 </KNOWN-HOSTS>
</META>
```

The new KNOWN-HOSTS section would allow example.com to declare hosts who refer to this policy reference file. The HOST name attribute above denotes that forinstance.com hosts are allowed to refer to this policy reference file, and the entity-type attribute denotes that they are owned by the same entity.

One implication of this mechanism is that now user agents should allow non-local URIs in the INCLUDE and EXCLUDE elements of policy references. For example, if example.com and forinstance.com were owned by the same entity, but portions of each site were actually subject to different (but shared) P3P policies, they might have a policy reference file like this one:

```
<META xmlns="http://www.w3.org/2002/01/P3Pv1">
 <POLICY-REFERENCES>
    <POLICY-REF about="/p3p/policy.xml#corporate">
      <INCLUDE>/*</INCLUDE>
    </POLICY-REF>
    <POLICY-REF about="/p3p/policy.xml#customerservice">
      <INCLUDE>*.example.com/cs/*</INCLUDE>
      <INCLUDE>*.forinstance.com/customer/*</INCLUDE>
    </POLICY>
 </POLICY-REFERENCES>
 <KNOWN-HOSTS>
    <HOST name="*.forinstance.com" entity-type="SAME"/>
 </KNOWN-HOSTS>
</META>
```

In this example, the shared policy for customer service site features includes different URI patterns for example.com and forinstance.com.

Proposed Additions to Support Compact Policies

Some user agents choose to only use compact policies to apply privacy preferences to cookies. Since compact policies do not allow the same level of expressiveness as policy reference files, they require a parallel mechanism to allow expression of “same entity” relationships. The proposed mechanism calls for the addition of a new HTTP P3P header “same-entity.”

P3P same-entity Header

The P3P same-entity header allows a host to specify a list of space-delimited hostname qualifiers that describe hosts owned by the same entity as the current host. This list must match the list of known hosts of type SAME in the policy reference file. In the “same entity” example, example.com would return the header:

```
P3P: same-entity="*.forinstance.com"
```

and forinstance.com would return the header:

```
P3P: same-entity="*.example.com"
```

When using these headers while evaluating compact policies, a user agent should only consider two hosts to belong to the same entity if each host has a matching same-entity hostname qualifier for the other host. For simplicity of implementation, user agents should allow the trivial case of including the current host in the same-entity list.

Efficiency Concerns

For purposes of efficiency, hosts should not be required to return all hostname qualifiers for each of these new headers on every request. Instead they may tailor the header based on the request context, e.g. an entity owns 100 different domains, on hosts in forinstance.com, it may return only "*.example.com" if it can glean that the request was referred from example.com.

Implications for User Agents

To take advantage of the new expressiveness provided by the proposed modifications, user agents should implement the following high-level rules:

1. Same Entity: Hosts A and B should be considered to belong to the same entity if:
    a. Host A refers to a policy reference file on host B, and that policy reference file contains a matching KNOWN-HOSTS entry for host B with type SAME, or
    b. During compact policy evaluation, host A has a matching hostname qualifier in the P3P same-entity header for the host B, and vice versa.
2. In the event of a verified "same entity" relationship, no extra privacy restrictions should be applied to either host.