Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 22, 2008

P. Hoffman
J. Levine
Domain Assurance Council
February 19, 2008

Format for Domain Reputation Data
draft-hoffman-dac-domainrepdata-01.txt

Status of this Memo

Copyright Notice

Abstract

This document describes two formats for reputation data for domains. The smaller format contains data that is expected to be used in real-time receiver decisions, while the larger format is used for more complete data that is appropriate for off-line decision making.

Table of Contents

1. Introduction

Providers of domain reputation want to be able to publish many types of reputation data. Among these are:

- A score along some scale, plus an indication of the confidence in that score
- Data about the domain's owner such as their name, how long they have been in business, where they are located, the type of business they are in, they type of mail they send, and so on
- Recent mailing statistics for the domain
- Number and types of complaints about the domain
- Innumerable others

There are many models for the ways domain reputation information could be distributed. Some providers might give some data away freely while charging for other data; some providers would give away the data in exchange for valuable feedback from the recipient about the domains; some providers would sell the reputation data to the owner of the domain and certain mail receivers; and so on.

There are also many models for the ways the domain reputation data would be used by mail senders and receivers. SMTP servers could use a score of the likelihood that they would want to receive mail from a particular domain, and they might be interested in the type of business of the sender for mixing into their decision on how to deliver the mail (banks might be more likely to be delivered to the inbox, auto dealers to the spam folder). ISPs might buy reputation data in bulk to help create their own in-house scoring systems.

This document describes two formats for reputation data for domains. The smaller format contains data that is expected to be used in real-time receiver decisions, while the larger format is used for more complete data that is appropriate for off-line decision making. DAC will later define protocols for retrieving the reputation data; these are likely to be based on DNS queries and responses.

[[The intended status of this document is an Informational RFC that will be submitted as an independent submission to the RFC Editor.]]

2. Smaller Format for Responses

The smaller response format is plain text with single spaces as separators. An explicit goal is that the response can be parsed without XML parsers (which are rare on SMTP servers).

The format for a small response is:

```
<version><sp><score><sp><confidence><sp><SIC>
```

All fields, and the single space between each, are required.

- o version -- a text string; for the first version of the protocol, it is "1".
- o score -- the likelihood that a recipient who trusts the reputation provider would want to receive mail from the domain. The value specifies a number between 0 and 99 inclusive that is the relative score of the the domain.
- o confidence -- the confidence of the reputation provider in the score. It is a number between 0 and 99 inclusive with 50 meaning "average confidence".
- o SIC -- the numeric NAICS code of the business associated with the domain. The value is the numeric code. The most recent list of NAICS codes can be found at [NAICS-CODES].

An example of such a record might be:

```
1 72 99 52213
```

3. Larger Format for Responses

The larger response format is XML. The XML overhead for this format is approximately 100 bytes, which leaves plenty of room within even a 512-byte UDP DNS response for simple information. In the inevitable cases where the answer is too big for one UDP packet, there is a simple fall-back to TCP, although the number of larger-format queries that have that much data is probably limited.

A set of common elements is defined in the namespace "http://domain-assurance.org/rep-3". A reputation provider can use their own XML namespace or other common XML namespaces for elements not defined in the DAC namespace. DAC-defined elements have very short names in order to maximize the number that can fit in a single UDP packet.

Every element is optional. Also, every element also has an optional c attribute whose value is a number between 0 and 99 inclusive that is the confidence of the reputation provider in the information in the element. If the c attribute is not given, the default value is 50, meaning "average confidence". It is expected that the c attribute will not be used often.

The general defined elements are listed here.

- o sc -- Score, the likelihood that a recipient who trusts the reputation provider would want to receive mail from the domain. The value specifies a number between 0 and 99 inclusive that is the relative score of the the domain.
- o in -- Industry, the numeric NAICS code of the business associated with the domain. The value is the numeric code. The most recent list of NAICS codes can be found at [NAICS-CODES].
- o na -- Name, the true name of the business associated with the domain.
- o st1, st2, ci, pr, co, pc -- Postal address elements of the business associated with the domain: street address 1, street address 2, city, state or province, country, postal code. The country should be given as the two-letter ISO 3166 code.
- o te -- Telephone, the telephone number of the business associated with the domain. The value holds the text string that is the telephone number; spaces, periods, and hyphens are explicitly allowed. The value should include the country code prefaced with a "+".

4. Examples of Reputation Records

An example using just the defined elements might be:

```
<?xml version='1.0' encoding='UTF-8'?>
<rep xmlns='http://domain-assurance.org/rep-3'>
<sc>72</sc>
<in c='75'>52213</in>
<na>Example Company LLC</na>
<st1>127 Typical Street, Suite 500</st1>
<ci>Anytown</ci><pr>CA</pr><co>US</co><pc>95782-4410</pc>
<te>+1 672.487.0091</te>
</rep>
```

An example including provider-defined elements might be:

```
<?xml version='1.0' encoding='UTF-8'?>
<rep xmlns='http://domain-assurance.org/rep-3'
    xmlns:brp='http://BigReputationProvider.com/syntax'>
<sc>72</sc>
<brp:fln type='md5'>a1892c25e959bd7a87161724dfec8c6d</brp:fln>
<brp:mris value='green'/>
</rep>
```

5. RELAX NG schema

```
grammar {
c-att = attribute c { xsd:integer }
start = element rep {
    default namespace ='http://domain-assurance.org/rep-3'
    element sc { c-att?, xsd:integer }?,
    element in { c-att?, xsd:integer }?,
    element na { c-att?, text }?,
    element st1 { c-att?, text }?,
    element st2 { c-att?, text }?,
    element ci { c-att?, text }?,
    element pr { c-att?, text }?,
    element co { c-att?, text }?,
    element pc { c-att?, text }?,
    element te { c-att?, text }?,
}}
```

6. Security Considerations

There are no security considerations for publishing reputation information about domain names.

7. Informative References

[NAICS-CODES]
U.S. Census Bureau, "2002 NAICS Codes and Titles", 2002, <http://www.census.gov/epcd/naics02/naicod02.htm>.

Appendix A. Acknowledgements

Many members of the Domain Assurance Council contributed to the design of the protocol and the wording of this document.

Appendix B. Changes between versions

[[This whole section will be removed before being published as an RFC.]]

B.1. Differenes between -00 and -01

Minor editorial revisions.

Added the intended status of the document.

Authors' Addresses

Paul Hoffman
Domain Assurance Council

Email: paul.hoffman@domain-assurance.org

John Levine
Domain Assurance Council

Email: john.levine@domain-assurance.org

Full Copyright Statement

Intellectual Property

Acknowledgment

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).