What's in a Name?

Issues in Identification for Information Infrastructure Architectures

Geoff Huston
Research Scientist
Asia Pacific Network Information Centre

Ideas Forum
National Library of Australia
April 2006

In this presentation:

- I’d like to explore the issues around identity and the structure of identity name spaces
- Look at what makes identity systems relevant and useful for an information infrastructure framework
- Explore the proposition that the (ab)use of URLs are a significant part of the problem we face today in constructing truly useful information frameworks

An example of an Identity Space

Internet Protocol Addresses are

- a means of uniquely identifying a device interface that is attached to a network - **WHO**
  - **Endpoint identifier**
- A means of identifying where a device is located within a network - **WHERE**
  - **Location identifier**
- A lookup key into a forwarding table to make local switching decisions - **HOW**
  - **Forwarding identifier**

This deliberate overload of sematic intent of IP addresses has been a basic simplifying feature of the IP architecture

Challenges to the IP Address Model

Roaming Endpoints (Nomadism)
Mobile Endpoints
Session hijacking and disruption (Security)
Multi-homed Endpoints
Scoped overlapping address realms
Network Address Translators and Application Level Session Translators
Voice Over IP
Peer-to-Peer applications
Routing Complexity and Scaling

Wouldn’t it be good if…..

Your identity was stable irrespective of your location

You could maintain sessions while being mobile

You could maintain sessions across changes in local connectivity

That locator use was dynamic while identity was long-term stable

Anyone could reach you anytime, anywhere

You could reach anyone, anytime, anywhere

Wouldn’t it be good if ……

Identities actually worked for the end user!

The Hard Lesson

Attempting to overload a single identity system with a diverse set of intended roles may look like a useful shortcut at the time

But it’s a terrible mistake!

What do we want from “Identity”?

Varying degrees of:

- Uniqueness
- Persistence
- Structure
- Clear Scope of Applicability
- Validity and Authenticity
- Clear line of derivation of “authority”
- Unambiguous resolution

Identity is not a unilateral assertion – it’s a recognition of derived uniqueness within a chosen frame of reference

What should we avoid in “Identity”?

Varying degrees of:

- Uncoordinated self-assertion
- Arbitrary token value collisions
- Ill-defined temporal validity
- No coherent structure
- Unclear applicability
- Semantic overload
- Structural overload and complexity of the token space
- Cost

So what?

All this is rather abstract

How does this relate to the nature of an information infrastructure?

We've done a pretty lousy job so far!

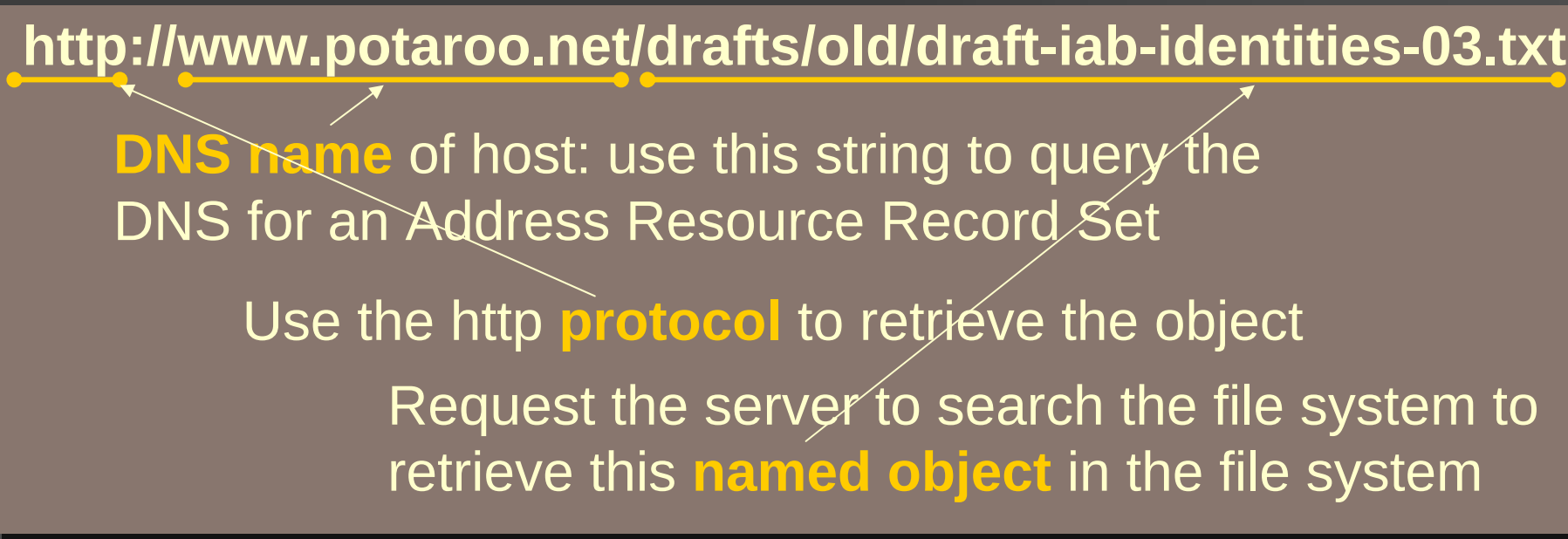
The information infrastructure has fallen into the same trap as IP addressing in its adoption of URLs as the underlying identity realm:

- **what** is synonymous with **where** in an object-oriented world
- **where** then becomes a viable non-clashing identifier scheme that also happens to dictate a resolution mechanism at the same time
- So all we need to a methodical approach to **where** and we're done!

Easy, simple and extremely inelastic!

Whats so bad about URLs?

- URLs describe a retrieval algorithm for an object instance, not an object identifier
- Device and application selectors coupled with application-specific query string



A URL is not “atomic”

A URL is a derived identity schema

- Protocol identifier
- DNS identifier
- Filesystem name

Uniqueness is a derived property of the hierarchical structure of the DNS and the relative uniqueness of names objects in a local filestore

Its insecure, vulnerable to all kinds of abuse and inappropriate to our conventional methods of utilizing information

What happens to a URL when:

- The site changes its name?
- The server changes its name?
- The filesystem changes?
- The access protocol changes?
- The document changes?
- The document is cloned?
- Your DNS Root is changed underneath you?
- Your DNS resolution is perverted?
- The name part no longer resolves?
- The protocol part is unrecognised?

What's Good about URLs?

They are usually unique for a while

Billions of instances of browsers recognise and resolve them

They offer the comforting illusion of security and authority without imposing the actual cost of true security and authority

What's Bad about URLs

- They lack persistence, authority and clarity of resolution
- They identify what was unilaterally claimed to be at one time a possible location of an instance of an object, not the object itself
- They identify instances of objects, not objects and not interactions between objects and entities
- They so not disclose pseudonyms or other forms of object equivalence
- They are not intrinsically linked to resolution mechanisms

Identity Scheme Choices

Its possible to inject an identity scheme into almost any part of an information system

- **Application or Service Identities**
  - phone numbers, Skype IDs, email addresses, URLs, Google Search terms
- Structured Namespace identities
  - DNS names, X.500 Distinguished Names, ISBNs
- Abstract Identities
  - Public Key, Hashed PK, session identifier

In this context an “identity” is a token to allow multiple instantiations of an object to be recognised as belonging to a single equivalence class

Identity Scheme Choices

DNS-related Identity at the Application level

- Use a stable name space that is resolveable into other identity spaces (using the DNS as the universal rendezvous point
- Allow indirection and referral via DNS NAPTR records
  - Generic identity with service-specific mappings
- Use application agents to provide stable rendezvous points
  - For example: sip:*gih@sip.apnic.net*
- Issues:
  - Can the DNS support dynamic interaction at a suitable scale and speed?
  - Are a family of diverse application-specific identities desireable (cross-application referral and hand-over)
  - Can we stop application designers from creating application-specific solutions that rely on an application-specific identity space?

Identity Scheme Choices

Search terms?

- Indeterminate – same query, different responses
- What did you want? Is it the object, or the current available relationships between query and some object set that you were after?
- Is the integrity of this relationship important?
- Is the “sociology” of the search even remotely relevant?

Identity Scheme Choices

Structured Namespaces

- Compound objects that may include identification of an issuer, subject, issuance, metadata…
  - DNS NAMES
    - Unique chain of named issuer – subject relationships to create a compound name and coupled resolution mechanisms
  - E.164 Phone Numbers
    - Historically: Country, Area, Provider, Subscriber
    - Currently: ?
  - X.500 names
    - ?
  - ISBNs
    - Group, Publisher, Title, check
  - PKI
    - Issuer, Subject, Subject Key, Metadata

Choices, Choices, Choices

Abstract Identities

- Low overhead access to uniqueness above all else
- Hash value of a Public Key
  - Block of bits without internal structure
  - Robustly provable provenance (via private key)
  - No implicit association to object instances
  - Can be replicated at will without dilution of its uniqueness
- Session Identifiers
  - Ephemeral identities that are reused
  - Disambiguate between active alternatives
  - Contextual resolution

Identity Resolution Issues

Use of an “Identity” is to resolve it into useable attributes and values

We can look at identity and resolution of identity as related, but distinct, concepts

Is the identity resolution function:

- Absolute or relative to the query?
- Absolute or relative to the identity token issuer?
- Dynamic or static?
- Configured or negotiated?
- Deterministic?
- Temporal?
- Assured to terminate?
- Assuredly valid?
- Assuredly secure?

Identity Implementations

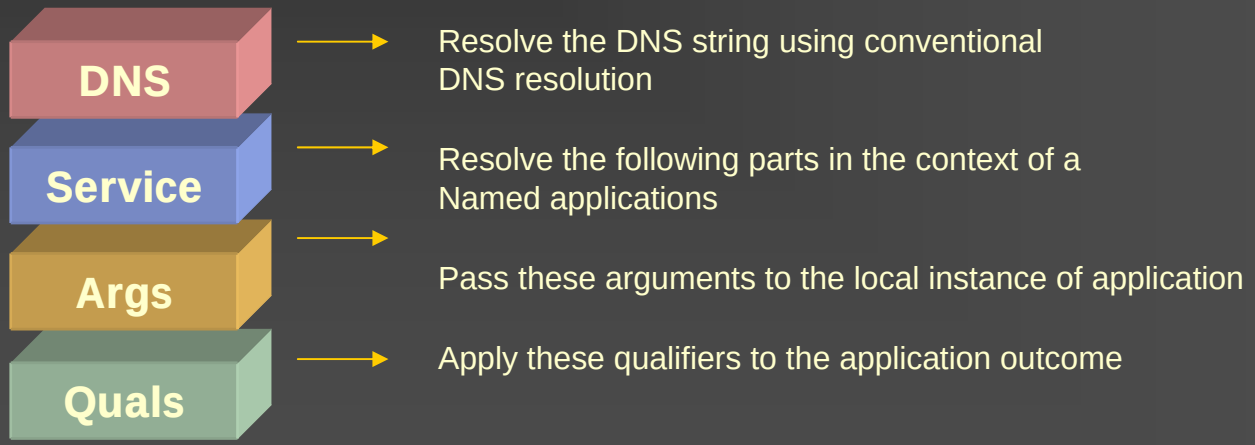
"Conventional"

- Construct a compound object that combines external identification realms of the identity issuer and the means to resolve the token in the context of the issuer

Identity Implementations

"Compound Referential"

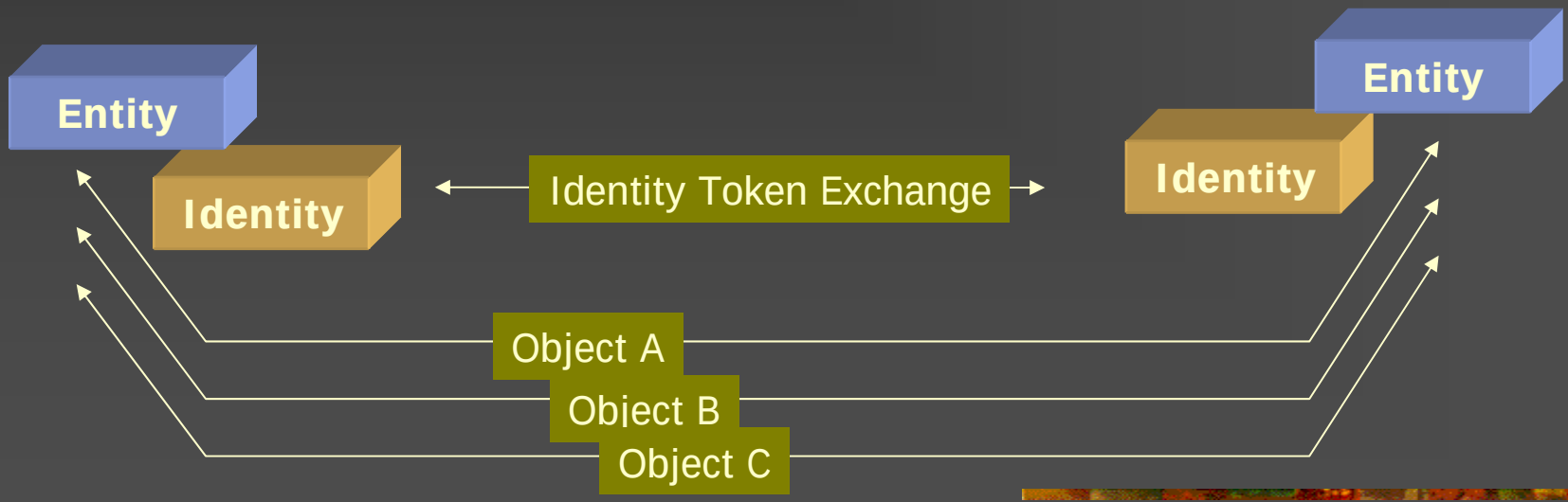
- Use a series of identity elements with a set of resolution mechanisms



Identity Implementations

"Ephemeral"

- Use an opportunistic identity as a means of resolving uniqueness in a limited context



Scoped Identities

Is identity:

- What I call myself ?
- What I call myself in relation with others?
- What I call myself in relation with others today?
- What you call me ?
- What they use to call me ?
- All of the above?
- None of the above?

Upper Level Issues of Identity Realms

- The significant effort and cost of supporting a new global unique token distribution system as an identity system
- The unintended side-effects of reusing some other existing token set as an identity component
- The issue of the relationship between identity and resolution mechanisms
- The overhead of identity resolution for application-level transactions
- The security issues in maintaining integrity of identity and integrity of resolution

Information: Discourse and Dialogue

> The term 'information infrastructure' (II) refers to the communications networks and associated software that support interaction among people and organisations.
>
> http://www.anu.edu.au/people/Roger.Clarke/II/

- So how could we identify and reference these interactions among people and organizations?
- Does our chosen identification mechanism blind us from the deeper common intent to use IT to enhance our information infrastructure?
- Are we overly fixated on the object and have we lost sight of the conversation?

百花齊放，百家爭鳴 *

One URL size fits all appears to be imposing poor outcomes upon the infrastructure of information:

- Information as objects vs information as an outcome of collaboration
- Associating the metadata with the object, not the identifier
- Disassociation of attribute discovery from the identity space
- Disassociation of object identification from object instantiation

Maybe we should revisit the URL scheme and look at alternatives that attempt to do less in the identifier space and leave more to the resolution space?

Is assured uniqueness and methods of resolution and attribute discovery all we really need from our identifiers?

How much activity is there is looking at other mechanisms of identification of the entities that populate the information infrastructure?

* *Let a hundred flowers bloom: let a hundred schools of thought contend*
Mao Zedong, 1956

Thank You!

Questions?