The UK Federation and Shibboleth: Nuts and Bolts

Ian A. Young
EDINA, University of Edinburgh

Networkshop 35, University of Exeter, 3–5 April 2007

EDINA® JISC

Overview

- What is the UK Federation?
  - Access management federation for all of UK education (HE, FE, schools)
- What is Shibboleth?
  - A SAML 1.1 profile
  - an implementation of that profile
    - flexible policy based attribute transport
    - support for federation in-the-large
  - a software platform: not just SAML

Overview: UK Federation

- Potentially very ambitious deployment
  - hundreds of member organisations
  - hundreds or thousands of entities
  - Total ≈12–18M eligible end users
- Federation technical services:
  - metadata verification and aggregation
  - metadata signature and publication
  - central discovery service
  - trust broker

UK Federation Statistics

- 46 full member organisations
  - ≈15 more still migrating from SDSS Federation
- 111 SAML entities
  - 49 identity providers
  - 64 service providers
- Software:
  - 87% Shibboleth 1.3
  - 7% Shibboleth 1.2
  - 5% other/unknown

These statistics as of 1 April. Remember to update before presentation.

Net numbers haven't been growing much since December, as we have been removing dead wood as well as having new members join.

Entity sum doesn't add up because of gateway entities, which are both IdP and SP at once.

Other: two gateway entities, one home-grown SP, one AthensIM, one Guanxi

Metadata

Three parts:

- Metata
- Discovery
- (only if time) Trust

Entity Metadata

- <EntityDescriptor>
  - <Extensions>
    - labels, e.g., “owned by UK Federation member”
    - scopes for scoped attribute values
  - Role descriptors
    - <IDPSSODescriptor>
    - <AttributeAuthorityDescriptor>
    - <SPSSODescriptor>
  - <Organization>
  - <ContactPerson> *n

Role Metadata

- e.g., <IDPSSODescriptor>
  - <Extensions>
  - <KeyDescriptor> *n
    - may be <ds:KeyName> for PKI based trust
    - may be <ds:X509Data> for explicit key wrapped in a certificate
    - may be various other things too horrible to relate
  - service endpoints
    - <AssertionConsumerService>
    - <ArtifactResolutionService>

Metadata Generation

- Each <EntityDescriptor> stored separately as a "fragment file"
- Ant and XSLT used to combine, filter and transform multiple variants:
  - Shibboleth 1.2 vs. Shibboleth 1.3
  - SDSS Federation vs. UK Federation
  - Full list vs. filtered (for WAYF)
- Each variant is digitally signed using appropriate federation signature key

133 fragment files, but only 113 entities: remainder are deleted entities

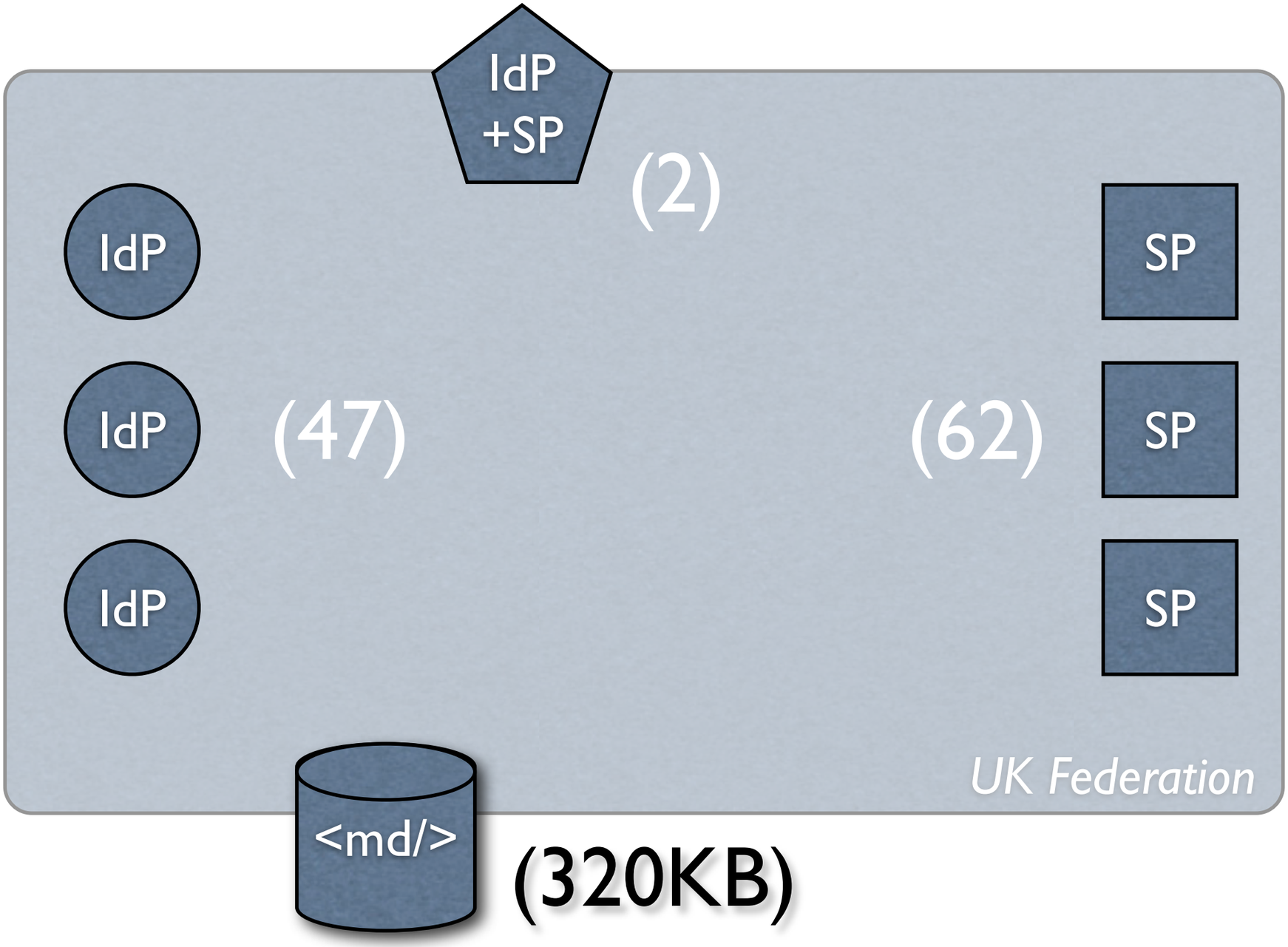
Metadata Distribution

- Master copies pushed to staging site
- Production metadata servers pull copies regularly
  - three identical machines
  - geographically distributed in multiple data centres
- Individual members pull copies from production metadata servers
  - recommendation is at least daily

Some of this is in the “future hopeful”; full redundant deployment currently in progress.

Signature on metadata is checked when a copy is downloaded to prevent spoofing.

Metadata Constituents



Problem: Metadata Scale

- Most things scale as O(N)
- Metadata server bandwidth scales as $O(N^2)$
- Current metadata size not a problem
  - likely not true forever
- Medium term tricks are possible
  - IdP vs. SP split (factor of 2)
  - compression (factor of 9–10)
  - subfederations (factor of ≈2?)

$O(N^2)$ dominates in the long run, but O(N) issue may be important before that depending on how (for example) Java metadata parsers improve. May hit issues in the 2MB range (400–700 entities).

In all, maybe a factor of 40 available on the $O(N^2)$ problem from "tricks".
This equates to a factor-of-6 growth in size (sqrt(40) = 6.32).

Tricks 1 and 3 give maybe a factor of 4 on the O(N) problem, raising the problem point to perhaps 1500–3000 entities (if parsers don't improve).

Metadata Scale cont'd

- Long term: move away from centralised metadata distribution
- One possibility: self asserted metadata
  - e.g., by de-referencing entity's name
  - issue: chain of trust broken
  - issue: some metadata can't be self asserted
- Long term analogy: HOSTS.TXT → DNS ?
  - maybe more than an analogy
  - let's not invent another parallel system!

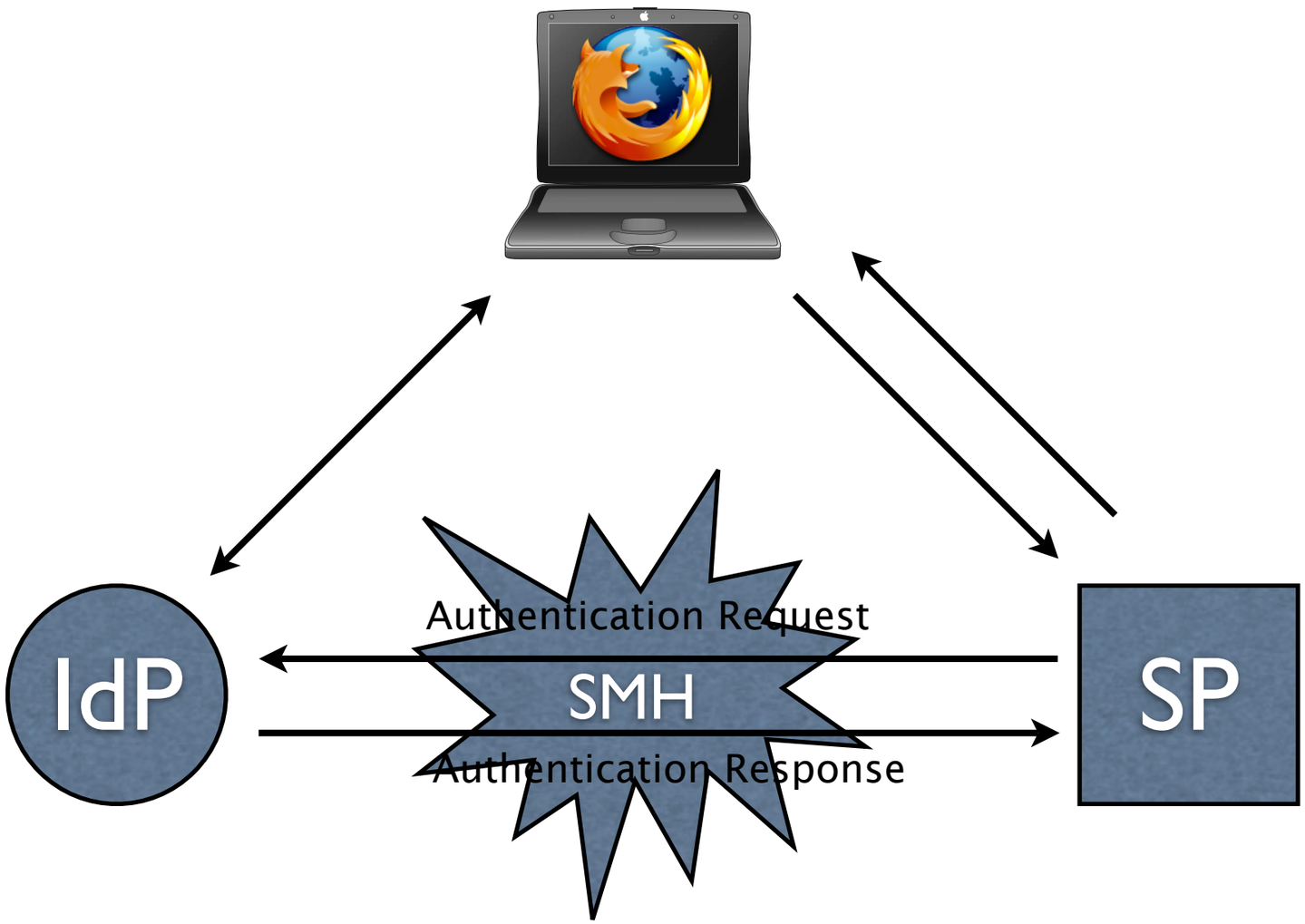
Chain of trust broken: how do you know you're getting the right metadata?

Non-self-asserted: scopes, labels, in general third party statements about an entity.
It may be possible (even necessary) to do without this, but it changes basic assumptions about the role of the federation.

DNS invented about 1983.

Discovery

The Discovery Process

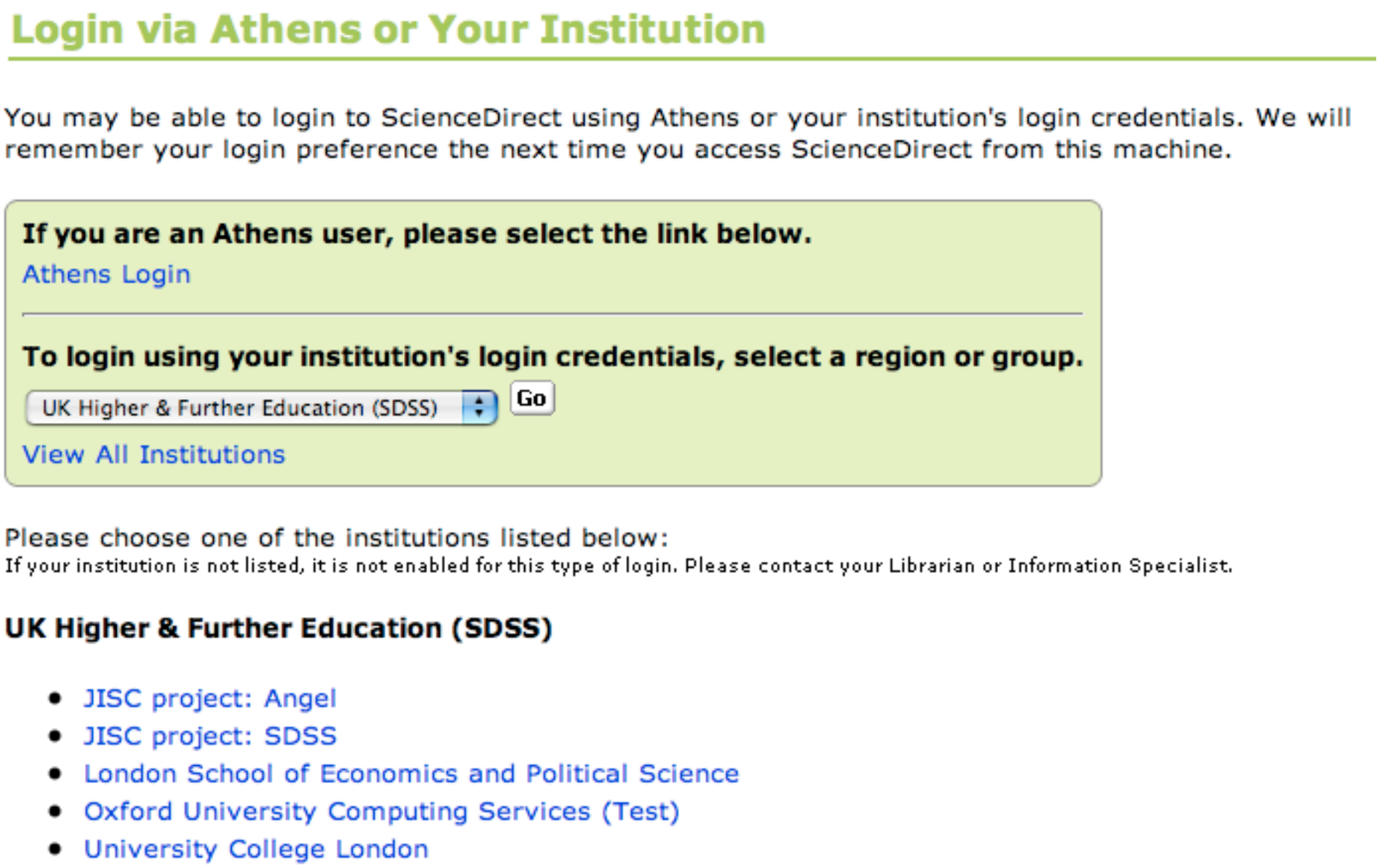


Start with a user, making use of a client by which we mean a browser
User's client approaches SP, SP has no existing session
User wishes to make use of identity from a particular IdP
discovery problem is how to let SP and IdP communicate
"something magic happens"
Result is that the SP's authentication request can reach the IdP
IdP authenticates
IdP sends response to SP
SP authorises

Discovery Options

- Institutional portal avoids the issue entirely
- Service provider can perform discovery locally
  - Good option in many cases
  - Service Provider often knows its community of users
  - Particularly true for licensed content, where a real-world contract will exist
  - Also true for resources built around small collaborations

Example: Elsevier ScienceDirect



Login via Athens or Your Institution

You may be able to login to ScienceDirect using Athens or your institution's login credentials. We will remember your login preference the next time you access ScienceDirect from this machine.

If you are an Athens user, please select the link below.

Athens Login

To login using your institution's login credentials, select a region or group.

UK Higher & Further Education (SDSS) Go

View All Institutions

Please choose one of the institutions listed below:

If your institution is not listed, it is not enabled for this type of login. Please contact your Librarian or Information Specialist.

UK Higher & Further Education (SDSS)

- JISC project: Angel
- JISC project: SDSS
- London School of Economics and Political Science
- Oxford University Computing Services (Test)
- University College London

http://www.sciencedirect.com/

Observations:

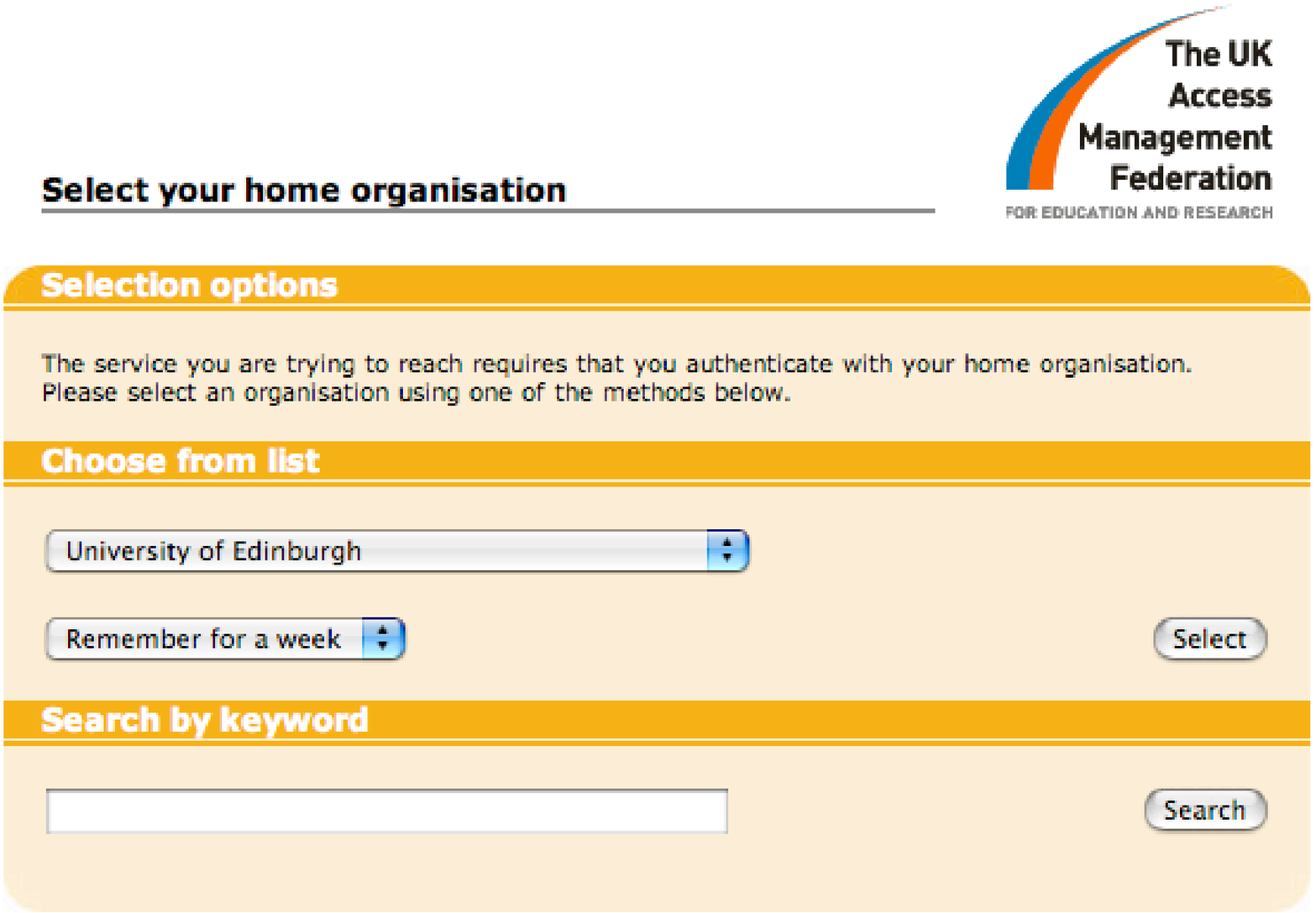
- does NOT talk about Shibboleth
- does NOT include all 50 UK Federation IdPs

for the particular circumstances of this SP, this is a much better user experience than any central discovery service could hope to offer

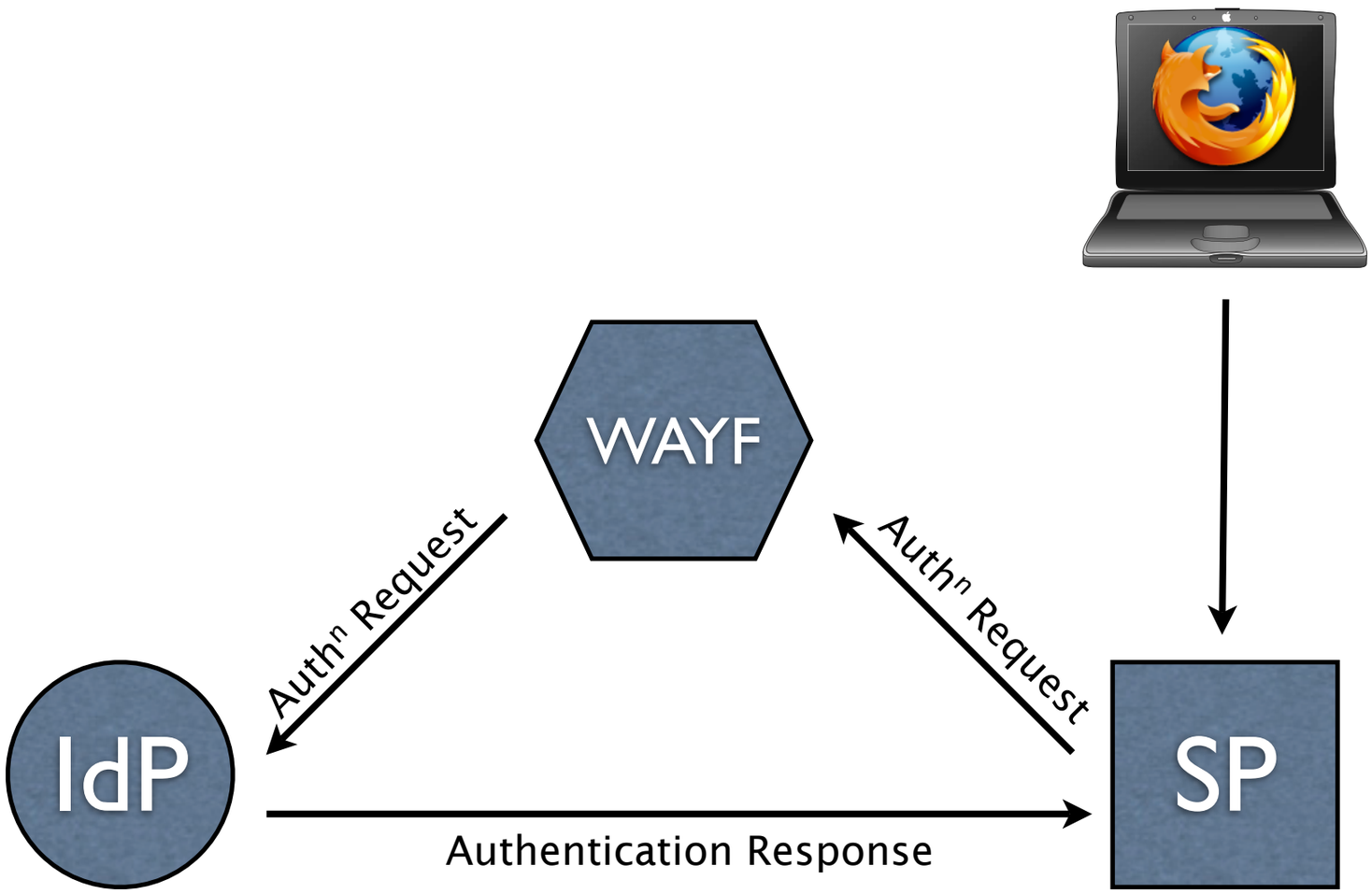
Discovery Options: Central WAYF

- UK Federation provides central “Where Are You From” service (a WAYF) as backstop
- Production WAYF servers work from federation metadata
  - three identical machines
  - geographically distributed in multiple data centres
  - https:// as anti-spoofing measure

Again, this is future hopeful as the WAYF machines are in the process of being deployed right now.


The UK Access Management Federation
FOR EDUCATION AND RESEARCH
Select your home organisation
Selection options
The service you are trying to reach requires that you authenticate with your home organisation.
Please select an organisation using one of the methods below.
Choose from list
University of Edinburgh
Remember for a week
Select
Search by keyword
Search
Need assistance? Visit the UK Federation web site.

Discovery with WAYF

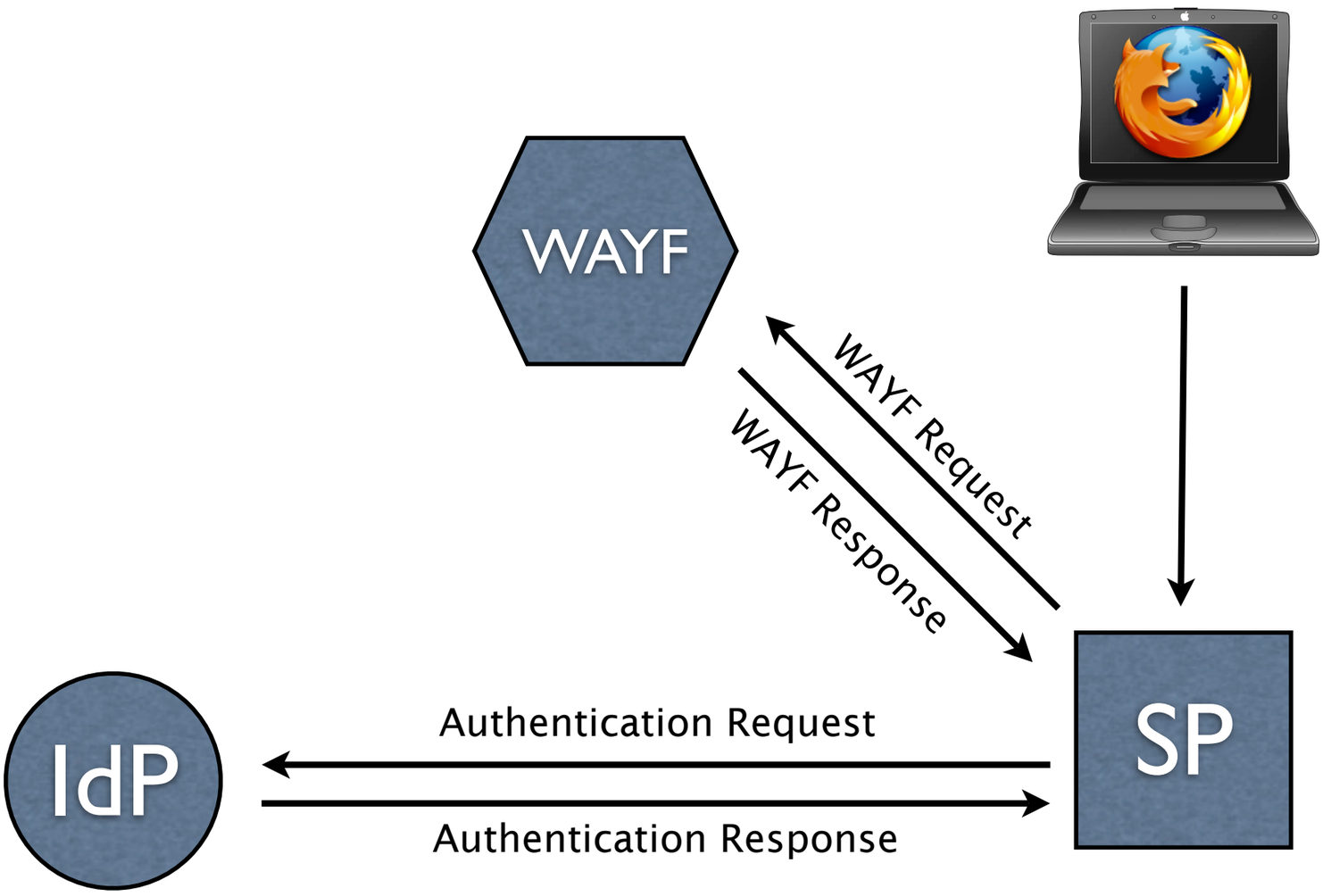


Note that WAYF merely PASSES ON the authentication request

(One) WAYF-induced Problem

- Because the WAYF merely redirects, ***the SP must decide what to send before the destination is known***
- Problems with:
  - Different SAML profiles:
    - SAML 1.1 Browser/POST (100% of current IdPs)
    - SAML 1.1 Browser/Artifact (55% of current IdPs)
    - SAML 2.0 Authentication Request (Shibboleth 2.0)
  - Non-SAML profiles
- Solution: new “WAYF protocol”

Discovery with WAYF 2.0



Contacts

- Federation:
  - http://www.ukfederation.org.uk/
- Speaker:
  - ian@iay.org.uk

Trust

This is a “bonus track”, only for use if there is enough time left for it.

Federation as Trust Broker

- "Trust" may refer to different concepts
  - "Technical trust"
    - means being able to verify an entity's claims about its own identity
    - mediated by the federation, through metadata
  - "Behavioural trust"
    - ***given*** technical trust, means having guarantees about the entity's behaviour
    - arranged partner-to-partner, e.g., through contracts
    - not mediated directly by the federation

Using Trust Metadata

- Example: signature verification
  - look up entity in metadata
  - extract <KeyDescriptor>s from metadata
  - resolve into associated public keys
    - may be indirect, key ***named*** in metadata
    - may be direct, key ***explicit*** in metadata
  - perform public key operation using entity's key
  - compare result with document digest

PKI Trust Fabric

- Used by many current federations
- Federation metadata contains list of trusted CAs
- Entity metadata contains key ***names***
- Assertions contain certificates with these names, certified by the trusted CAs
- Receivers verify document signature, and also verify the certificate path up to a trusted CA
- This is the current UK Federation approach

PKI Trust Fabric: Pro

- Per-entity metadata is small (just key names)
- Certificates can change as long as names don't
- Some entity identity proofing can be left to (commercial) CAs
  - originally seen as a strong benefit, allowing this function to be outsourced

PKI Trust Fabric: Cons

- Full PKIX path validation is slow
- Revocation is a hard problem
- Keeping up with CA certificate profiles is hard
- Keeping up with CA ***product names*** is hard
- Different federations tend to trust different CAs
  - interoperability issue
  - some SPs need to buy many certificates
- Only supported by Shibboleth

Alternative: Embedded Keys

- Certificates can be embedded in metadata
- Certificate resolves directly into a public key
- Pro:
  - Any CA acceptable (even self-signed)
  - No PKIX path validation required (performance++)
  - More interoperable (other feds, other products)
- Con:
  - Transfers responsibility for identity proofing to the federation

Crystal Ball Time

- Embedded key approach is in experimental operation within the UK Federation (≈5 entities)
- Cautious optimism about this approach, given development of supporting procedures
- Not easily used by Shibboleth 1.2
  - but that is reaching its official End Of Life soon
- May be the only practical way to meet interoperability goals

One of the 5 entities in question is the Internet2 Spaces site, supporting the Shibboleth community wiki.

Contacts

- Federation:
  - http://www.ukfederation.org.uk/
- Speaker:
  - ian@iay.org.uk