Use Case – Shibboleth

Scenario

Internet2's Shibboleth has been used at Newcastle University since 2004 to enable single sign for internal and external web resources. On a daily basis the number of login attempts hits on average 8,000 unique users, with users accessing 220 online databases, 28,000 e-journals, over 350,000 e-books and a large number of internally ran Shibboleth protected services. Currently, it is not a simple process to demonstrate usage statistics for access to these resources, this use case will be aiming to address this problem.

Shibboleth generates access and process logs which capture each event or interaction on the login servers. Yet these log files currently are redundant and unused, and in the current format, provide no benefit to the University as they have no contextual information in the format that they reside. This is not just the case within Newcastle University, but within other institutes, as is demonstrated by the work that has been carried out in the Internet2[1] community to develop mechanisms to process Shibboleth logs. Scripts have been made available by Shibboleth developers and contributors which can process log data and turn it into statistical information. On testing these scripts, they provided information at a basic level, yet had restrictions on the type of information that they could provide, for example, many focused on service provider usage rather than user information e.g. successful/unsuccessful user logins. In the form this basic statistical information was produced, it was still not suitable to provide to non-technical end users and decision makers. The information still lacked the context that was required, and to do this would require further staff time and resources to produce suitably presentable outputs.

As part of a previous JISC funded project ran by Newcastle University, GRAND[2], work was carried out to produce a number of one off reports to demonstrate examples of what metrics could be produced from the Shibboleth logs, these included[3];

- Total unique logins
- Successful/unsuccessful logins
- User session activity

To be able to produce this report involved a substantial amount of staff time and resources. The agreed approach was to port the data from the log files into a database so that the data was stored in a more suitable, structured format. This process was achieved with the use of a data integration tool, Talend[4], which helped to automate parts of the process, which was an improvement on the before mentioned scripts. However, this approach had the same problem as the scripts, as the information produced still lacked the presentable format, such as graphs. In order to produce the graphs for the example report, these had to be manually created within Microsoft excel. This is far

---

[1] http://www.internet2.edu/

[2] http://research.ncl.ac.uk/grand

[3] http://research.ncl.ac.uk/grand/docs/Shibboleth%20Logins.pdf

[4] http://www.talend.com

from practicable when it could be required to produce reports on a monthly or even weekly basis, and therefore this approach would not be sustainable.

Approach

Findings taken from the before mentioned experiences have set the foundations for the areas that would benefit from statistical reporting. The ability to deliver such reporting has been made possible with the Raptor[5] tool that has been developed by Cardiff University. Raptor enables the production of e-resource usage statistics, which are suitable to be used by non-technical staff. The tool will provide an overall solution that previous methods have not been able to provide, producing concise and presentable management information.

This initial use case will be aimed at producing the overall high level statistics providing a proof of concept, which will be built upon in later use cases.

As part of this use case, statistics will be produced for the following areas.

- **Number of unique authentications** – With any service it is important to be able to report on the overall usage levels of said service. These reports can be used to ensure that the service is running at a suitable level to be able to deliver an efficient service, and identifying patterns for peak usage. Both of which can be used in decision making processes with regards to service maintenance and downtime. Sub grouping of these authentications will also be carried out.
    - Authentications by School/Faculty – in initial project meetings the ability to be able to drill the statistics down into a number of different groups. One of which is to group number of authentications by school or faculty, this will allow a picture to be built of which areas make use of different services, potentially identifying areas which may not be fully aware of some of the services that are available.
    - Staff/Students – distinguishing between staff and student logins is another area which could prove to be beneficial. For example, if reporting on external E-resources (an area which will be covered later in the project), it will would be useful to be able to distinguish whether students or staff are using the resources.

    There are a number of other areas where it will be possible to group the statistics e.g. course, module. To ensure that statistics produced do not infringe on peoples information privacy rights and that a mechanism such as barnardisation is not required, the grouping will be kept at a high level so that it would not be possible to identify an individual users login activity.

- **Successful/Unsuccessful Logins –** Although the unique login statistics help to determine the usage of a service, they do not provide a full picture of whether the service is delivering what it is supposed to deliver, authenticating users to service providers. This is why it is important to provide logs which differentiate between users being able to login successfully and unsuccessfully. If there are a large amount of failed login attempts this could suggest that there is a problem with the Identity provider or to identify problems with individual user accounts.

---

[5] http://iam.cf.ac.uk/trac/RAPTOR