Introduction

From a recent CISC meeting, we decided to block RUcore downloads from IP addresses that appeared to have an inordinate number of downloads. For example, rutgers-lib:23904 PDF datastream had over 25,000 downloads from a single IP over the course of a few hours. The blocking criteria would limit the number of downloads of a specific datastream from a single IP within a certain time period (e.g. 24 hours).

IP Blocking - Scenario and Blocking Criteria

We are proposing a two-stage criteria as follows: 1) once the limit of 10 downloads within 24 hours from a single IP is achieved, we will stop counting and delete the 10 downloads from the statistics database and 2) we will continue to allow downloads until 100 is reached. At this point, downloads will be blocked and the IP address will entered into the "blacklist" (see below). Note, downloads of other objects from this IP would be permitted. The block criteria will be applied uniformly for all RUcore datastreams. In this scenario, after the 24 hour period expires, the IP would be un-blocked and could then commence downloading of the previously selected datastream. All IP addresses that have been blocked will be retained. If we continue to have repeat offenders, it may be advisable at some future time to block the offending IPs from RUcore.

The two-stage threshold will allow more flexibility, especially for the classroom scenario. For example, a professor in an e-classroom (with one single IP address to the outside world) at Rutgers has a link in the syllabus to an item in RUcore. A request to all students in the room to click on the link will likely generate more than 10 downloads in a short period of time. The two-stage approach allows up to 100 downloads. Note also, that the thresholds (10 and 100, and the 24 period) are configurable and can be easily changed based on our experience with IP blocking.

Implementation

Whitelist and Blacklist

Two separate lists will need to be added to the statistics package and be manageable through a user interface provided to statistics package administrators. The whitelist will contain IP addresses, ranges, and or domains where access is allowed regardless of the number of attempts during any time period. Those requests however will not be recorded as downloads. Entries in the whitelist can include search engine crawlers,

harvesters, and Rutgers addresses used for administration and testing etc. The blacklist will contain IP addresses where access has been restricted to datastreams for a period of time.

Monitoring and Detection

When requesting a datastream for download the IP address of the client will need to be tested to ensure it hasn't been blacklisted from accessing the datastream. This can be integrated into the current process as described in Figure 1. If access is blocked, a useful message will be provided to the user stating access to the datastream requested has been suspended for a period of time. A suggested message is included below and should be reviewed by the user services group.

> "The RUcore policy limits the number of downloads from a single IP address. You have exceeded this limit and are temporarily prevented from further downloads. Your download capability will be reinstated within 24 hours. Please contact the RUcore administrator at [RUcore contact] if you have questions."

If access is permitted the file will be delivered and a request to record the download will be sent to the statistics package. If the IP address has not been white or blacklisted the download will be recorded. After recording the download a new process will need to be added to the statistics package that will identify if the request exceeded the permitted number of requests during the allotted time. If so, that IP address will need to be added to a blacklist for a prescribed period of time. After adding the IP address to the backlist all downloads of the datastream from the IP during the current time frame will need to be expunged from the statistics database.

Summary

A two-stage IP blocking strategy is proposed for RUcore in order to insure the credibility of our download statistics. This strategy will also block malicious intent and provides the means, through retention of repeat offenders, to implement more stringent blocking procedures in the future. The IP blocking capability will be implemented in RUcore R7.5, targeted for a year-end 2014 release.

Figure 1 - Integration Point – Access and Validation