Technical Specification

Lsass Batch Lookup Architecture

**Likewise Software**

**October 29, 2008**

# Introduction

This is a summary of the basic idea and architecture for batch lookups and how we can use them to improve our life for single-object lookups (that we can later expand and add to our LSASS dev docs). This also provides me the opportunity to share the good news of how easily I was able to code up the by SID batch lookups using our new batching infrastructure.

In general, there are several types of object lookups that we might want to do in LSASS:

* By DN
* By SID
* By NT4 name
* By UPN
* By Alias
* By Id

The first three query types (DN, SID, NT4) are where we lookup the real object and then lookup the pseudo object (if applicable) by SID. The UPN query type is done by doing a LsaLookupSids(UPNs)  SIDs and then doing a SID-based lookup for the real object. For Alias and Id, we need to lookup the pseudo object first and then lookup the real object by SID.

So, in summary, we have this:

1. We can go from a Name, id, etc to a real or psudo object:
   * By DN  real
   * By SID  real
   * By NT4  real
   * By UPN  SID (batched via LsaLookupSids RPC)  real
   * By Alias  pseudo
   * By Id  pseudo
2. Once we have a real or pesudo object, we can lookup the other object (pseudo or real) by SID.

The two realizations above are the crux of how we figured out how to batch lookups. The basic batch lookup architecture is to (1) above as a batch operation and then (2) as another batch operation. (There are additional architectural pieces wrt how we divide up the queries by domain, do the actual query generation, the matching of results to query, handling of different modes, etc, but I won’t go into those here.)

For the batch code that Wei and I did last week and fixed up earlier this week, we implemented the machinery to do:

* By DN  real  pseudo (by SID)

The low-level batch query machinery that we wrote to handle the batching is very generic.

This evening, in about 3 hours, I was easily able extend the higher-level parts of the code, which were done as “By DN  real”, to make them generic too such that we can now do “By any attribute from which we can identify the domain  real”.

Using the newly generalized higher-level code, I was easily able to add SID support such that we now have the machinery for:

* By (DN or SID)  real  pseudo (by SID)

(Actually, in adding the “by SID  real” part, I made the higher-level code generic wrt the “by” part.)

With these changes, it is now trivial to add “by NT4  real” and “by UPN  SID  real”.  With some additional work, we should be able to do batch lookups by alias or id as well. There is currently no need to do batch by NT4, UPN, alias, or id. However, it would be useful to have all of the object lookup code (i.e., for a single object) go through a common codepath (i.e., the batch-capable codepath). This would reduce the amount of code we have quite a bit, we think. And it would make it easier to fix bugs in one place.