GDPR: Investigate what PD our LDAP servers make visible

Project description

We run various LDAP servers which could potentially allow folk to see other people's Personal Data. This project is to enumerate these, investigate whether and in what way it might be possible to restrict access, and to produce a report on what's what and the options available. Depending on the effort required, this project will then either make such changes as are required, or fork another project to make those changes (bearing in mind that more than just the Infrastructure Unit might have to do stuff). We have (at least) our ``general'' LDAP service and Prometheus. Some parts of our "general" LDAP service are currently visible outside Informatics but within EdLAN.

Useful GDPR links

http://computing.help.inf.ed.ac.uk/gdpr-resources

Current Situation

General LDAP

Branches

ou=AutofsMaps,dc=inf,dc=ed,dc=ac,dc=uk
- autofs map information
- personal data: none

ou=Capabilities,dc=inf,dc=ed,dc=ac,dc=uk
- entitlements membership
- personal data: roles; membership of institutions, groups; for students: modules, courses taken ...

ou=Group,dc=inf,dc=ed,dc=ac,dc=uk
- group information for rfc2307 (posixGroup objectclass)
- personal data: group membership, similar to ou=Capabilities

ou=Identities,dc=inf,dc=ed,dc=ac,dc=uk
- not currently used

ou=Maps,dc=inf,dc=ed,dc=ac,dc=uk
- AMD map information (see also ou=Partitions)
- personal data: home directory links

ou=Netgroup,dc=inf,dc=ed,dc=ac,dc=uk
- entitlements membership exposed in netgroup (rfc2307 nisNetgroup objectclass); also netgroups of hosts
- personal data: see ou=Capabilities

ou=Partitions,dc=inf,dc=ed,dc=ac,dc=uk
- disk partition information, for building AMD maps (see also ou=Maps)
- personal data: none

ou=People,dc=inf,dc=ed,dc=ac,dc=uk
- user information, mainly for rfc2307 (posixAccount objectclass), but also contains information like room number/telephone number
- personal data: email, location, name, tel.no, etc.

ou=rfeMaps,dc=inf,dc=ed,dc=ac,dc=uk
- rfe map information, used by rfe client to locate server
- personal data: none

Visibility

We restrict access to slapd via our firewall to 'edlan', 'edlan172' and 'tardis', as defined in <live/ipfilter.h>.

We use tcpwrappers to restrict access to:

- EdLAN:
  - 129.215.0.0/255.255.0.0
  - 192.168.0.0/255.255.0.0
  - 172.16.0.0/255.240.0.0
  - [2001:630:3c1::]/48
- TARDIS:
  - 193.62.81.0/255.255.255.0

In openldap ACLs:

We allow access to ou=People for everyone.

We allow access to the rest of the tree for

- everyone (including anonymous) within the Informatics firewall
- authenticated users
- localhost
- those from 'inf.ed.ac.uk' (via a DNS reverse lookup)

We make data visible to EdLAN for Virtual DICE.

Prometheus

Branches

ou=Roles,o=Prometheus,dc=inf,dc=ed,dc=ac,dc=uk
- roles/entitlement data
- personal data: could be used in conjunction with an individual's roles to glean information about a person

ou=Config,o=Prometheus,dc=inf,dc=ed,dc=ac,dc=uk
- prometheus config
- personal data: none

ou=Errors,o=Prometheus,dc=inf,dc=ed,dc=ac,dc=uk
- errors recorded by conduit audit methods
- personal data: could contain data about people

ou=Groups,o=Prometheus,dc=inf,dc=ed,dc=ac,dc=uk
- group/gid mappings
- personal data: none

ou=Accounts,o=Prometheus,dc=inf,dc=ed,dc=ac,dc=uk
- user account information
- personal data: email, location, name, tel.no, authentication data (last successful/unsuccessful auth, last pwd change), roles/entitlement data, which includes membership of groups, institutions; for students: modules, courses taken ...

ou=Conduits,o=Prometheus,dc=inf,dc=ed,dc=ac,dc=uk
- conduit configuration
- personal data: none

ou=Entities,o=Prometheus,dc=inf,dc=ed,dc=ac,dc=uk
- entity information (including identities and accounts)
- personal data: see ou=Accounts

ou=EventQueue,o=Prometheus,dc=inf,dc=ed,dc=ac,dc=uk
- contents of event queue (which allows conduits to be run on demand)
- personal data: could contain usernames

ou=Identities,o=Prometheus,dc=inf,dc=ed,dc=ac,dc=uk
- identity information
- personal data: see ou=Accounts

ou=Entitlements,o=Prometheus,dc=inf,dc=ed,dc=ac,dc=uk
- not currently used

Visibility

No data is visible outside the informatics firewall.

Authentication data (attributes prometheusLastauth, prometheusLastauthfail, prometheusLastpwdchange) is visible only to '.+/admin@INF.ED.AC.UK' principals.

All other data is visible to everyone, i.e. via anonymous lookups.

More Thoughts

Some thoughts from George, via email:

2019-03-26:

```
I was having another think about all this...

We definitely have some personal data in there.  Processing it for our own
needs is fine.  The issue is really how much of it should be visible
generally within Informatics, or indeed outside.

In order to process (in this case, to make visible to folk) we need to have
a valid Article 6 lawful basis, and since we're not processing for the
original purpose (system admin) we also need to take account of the Article
6.4 tests.

Anyway, the only plausible lawful basis which would work, I think, would be
legitimate interest.  I can't see consent in general working, as it would
imply that we have to have different controls on everyone's data which
would be a real pain to set up and we might as well do one thing for all.
I can't see contract working, as there's no obvious reason we have to give
away data in order to manage things (but see below).

Do we have a legitimate interest?  Maybe, though it would take a bit of
discussion and a proper LIA to decide.  We can certainly make a good case
that having the data available in order to apply security controls is
valid, and indeed is called out in one of the Recitals as a good reason.
The question really is, do we *have to* make the data available generally
in order to implement those controls?

If we do, we could reasonably say so in a LIA and privacy statement, and
leave things as they are.  If we could do it in another way (explicit user
binding, for example, or some kind of root-mediated access) then the LIA
would almost certainly say that the data should be restricted.

I can't think of any good reason why any of it should be exposed outwith
Informatics at all, other than consent, which is a pain to administer.

(The only way I can see "contract" working would be if we really did have
to drop stuff into something like a netgroup in order to make the
systems work at all.  The DPIA would have to be pretty explicit about
what's going on, though, and there would have to be a clear privacy
statement explaining it all.)
```

2019-06-04:

```
> 2.  Consider what we can do within Informatics.  Are you of the
> opinion that we're exposing too much already, or that we need
> written justification for what we do expose?

I reckon there are two sides to it.  First of all, do we have to expose
data on users to other users?  Is there some (reasonable) mechanism whereby
we could restrict each user's data only to that user?  (Or to some
privileged user on a machine, in order to be able to make initial
authorization decisions for example.)

If there is we should use it.  We can probably then just bypass the whole
thing by putting it down as "system administration", and if push comes to
shove we can make "system security" the reason for having it at all.  I
reckon a LIA case would be pretty solid on that basis, if we had to.

If there is no such mechanism then we need to do a LIA, balancing the
rights of the users against our interest in keeping things running securely.
It might come out OK, but it would have to be done explicitly and honestly.

(I think we're all agreed that we shouldn't expose things outwith
Informatics unless there's a need, and there doesn't appear to be such a
need.  "Inside" would include our own VPN mechanism, of course.)
```

2019-06-05:

```
Thinking a bit more about this, I think the question comes down to whether
we have to expose what we do to all and sundry, or can we be a bit more
restrictive?

Regarding prometheus, given that everything(?) that needs to be visible for
some purpose is already exported to elsewhere, is there any reason why any
of it should be visible at all?  I can't see how it would pass a LIA.  If
we restrict it and it breaks something then that might be no bad thing...

As regards general LDAP, and ignoring ou=AutofsMaps, ou=Identities (for
now), ou=Partitions and ou=rfeMaps, there's nothing in ou=Maps that
couldn't be inferred from ou=People.  Do IS have an equivalent branch
generally visible?  If so, we can just say we're copying what they do and
let them worry about it.  If not, we really need a LIA unless one of the
central ones could cover it, though if visibility is strictly internal it
sounds reasonable to me that it would say it's OK.

The ones that are of more concern are ou=Capabilities, ou=Group and
ou=Netgroup, which are essentially equivalent.  They really expose rather
more than is necessary.  We might get it through a LIA if we can show that
there's really no other way and the functionality is required for security
or resource-allocation reasons.

So, three questions:

1) What needs to be able to access the data, and for what purposes?

2) To what extent and by what means can access be restricted so that each
access can see only what it needs to?

3) If there is a proxy involved (e.g. sssd), can any such be configured to
be similarly restricted?

If it turns out that there is some fundamental reason that restrictions
can't be implemented in all cases then we may be able to get the existing
setup through a LIA by including all the technical reasons and performing a
genuine balancing act against the users' rights.  But it has to be genuine
and capable of standing up to the ICO's scrutiny.
```

Update 2019-07-18

It will be easy to get bogged down with all of this. I'll try to expand a little on the thoughts above and answer some of the questions that George asks.

Firstly, we have 4 different logical areas from which we can/should manage access:

1. Outside university
2. Within EdLAN, but outside Informatics (meaning outside our firewall)
3. Tardis - outside firewall, but closer relationship than EdLAN
4. Within Informatics firewall

We expose nothing (through LDAP) for 1, we want to close off 2, 3 is perhaps more of a political question as to how we define Tardis, and 4 is subject to debate...

Some questions:

Can we just close off the data available to EdLAN (i.e. for vdice) now? We're still seeing connections from uni vpn, edlan172, edlan ipv6. The newest vdice image only uses the "guest" account, so no need for LDAP access. According to Chris: "... we promised on the release of the current Virtual DICE image that it would be supported until September this year, but I suppose GDPR considerations might outweigh this if we want to close that off sooner." Do we?

A key question would seem to be: Is there a difference, in law, between data being available to everyone within Informatics, i.e. anonymously, and data being available to only authenticated users (e.g. anyone logged into a DICE machine, or otherwise authenticated to our KDCs)?

Does prometheus data need to be made available generally, or can we tie it down to certain principals/groups?

- C[S]Os - as a function of our jobs - either directly or via tools
    - is, even this, legitimate, or should it require secondary auth?
- prometheus replica
- rfe server (roles, groups) - essential data flow - currently authenticated
- rfe server (quotas) - read-only - currently anonymous; should really speak to regular LDAP and use entitlements
- password portal - web interface for user password setting
- support form - looks up roles (currently unauthenticated, could use custom tool or speak to regular LDAP)
- other, e.g. for roles lookup

Summary: we can probably tie this data down considerably - access only for specific machine based principals, or groups of people (sysmans).

Similarly, can we close off general access to LDAP data? What/who needs to access it?

- sssd - essential for system authn/authz - accesses People, Groups, Netgroups - can be configured to use authentication

- rfe authz - accesses netgroups through system interface, so goes through sssd
- web authz (mod_authnz_ldap) - typically accesses Capabilities or Groups (e.g. access to this wiki page is controlled through this) - can only use ldap basic auth - could be investigated
- autofs - doesn't expose personal data, but can be configured to use authentication
- Tardis - only accesses People
- self-managed machines - unknown (and possibly unknowable)

All of the above are assumed to be functional access (e.g. some part of the system); what about access by people - e.g. looking up information about other people (to obtain contact information). Is this legitimate?

Summary: we can tighten up access to many areas, but can we do it across the board?

Comment: Any part of our system which hooks into our roles/entitlements based authorisation system will need to speak to LDAP in some way, either through Prometheus, or regular LDAP (we would definitely prefer the latter).

Comment: OpenLDAP access controls can be incredibly fine-grained - down to the level of individual access to individual attributes. With this, of course, comes the risk of unmanagable complexity and unexpected behaviour.

Different categories of work to schedule/plan/do:

- things we have to do now
- things we should do now
- things we should plan to do
- things we have to be able to justify

Update 2019-09-11

We have now closed off firewall holes to EdLAN.

Update 2020-04-14

Minutes from (online) meeting:

> 460 - GDPR: Investigate what PD our LDAP servers make visible
>
> A Running Project talk was given.
>
> Writing a DPIA and LIA to cover our usage and reasons which will hopefully be sufficient, although if not we can do a lot to further restrict access but this will be some effort to do.

Update 2020-05-26

We have written a draft DPIA to cover our use of personal data within LDAP (including Prometheus). Once completed, we will decide (and take advice on) whether what we have is adequate to cover our use of the data. If not, we will tighten accordingly, and according to the investigations already carried out.

Update 2020-06-12

The DPIA has been completed and accepted. This project should now be written up.

Final Report

FinalProjectReport460

-- TobyBlake - 01 Jul 2020

This topic: DICE > Project460LDAPPersonalData
Topic revision: r11 - 01 Jul 2020 - 10:29:18 - TobyBlake