Ontwerp Sluisfunctie SRC-Yoda

Context

Een veelvoorkomende use-case in onderzoek betreft het verwerken van tijdelijk beschikbaar gestelde privacy- of anderszins gevoelige data. Instellingen en onderzoekers wensen hiervoor voldoende maatregelen te kunnen treffen om de vertrouwelijkheid van de data te beschermen. Een onderdeel van deze maatregelen vormt het gecontroleerd transport van data van en naar de verwerkingslokatie.
In dit document wordt als verwerkingslokatie een door SURF ResearchCloud geinstantieerde workspace als uitgangspunt genomen en als persistente data opslag lokatie een Yoda data management systeem.

In dit document komt bescherming van de overdracht van de data aan bod. Het beschermen van de data binnen Yoda zelf, of binnen de analyse omgeving (ResearchCloud) zelf valt buiten de scope. We nemen aan dat hier passende maatregelen zijn geïmplementeerd. Met het beveiligen van de overdracht is dan de gehele keten beveiligd.

Data sluis vereisten

De belangrijkste eisen aan een data sluis zijn:

- 1. Toegang alleen voor geautoriseerde personen, en
- 2. Toegang alleen vanaf goedgekeurde ('geautoriseerde') omgeving, en
- 3. Toegang is auditeerbaar

Belangrijke randvoorwaarde vanuit een veilige analyse omgeving

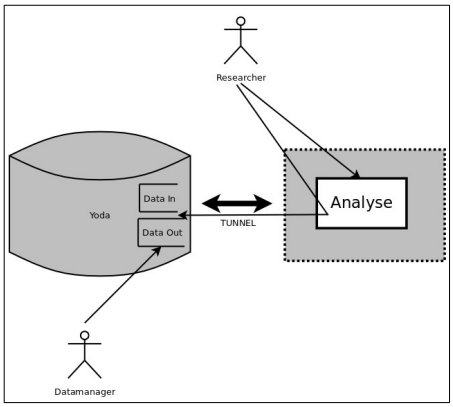
- 4. De sluisfunctie mag de beveiliging van de analyse omgeving en van Yoda niet nadelig beinvloeden. Bijvoorbeeld: extra netwerk connectiviteit/accounts minimaliseren.

Onder een veilige analyse omgeving verstaan we een omgeving van waaruit data slechts op gecontroleerde en geautoriseerde wijze de omgeving kan verlaten. De omgeving dient te zijn beveiligd tegen weglekken van data zowel per ongeluk als door opzettelijk handelen.

Opties voor implementatie

We hebben vier verschillende mogelijkheden voor implementatie onderzocht. Alternatief A implementeert een sluis functie in Yoda. Dit impliceert een nauwe functionele koppeling tussen Yoda en de analyse omgeving.
Alternatief B implementeert een sluisfunctie volledig in de analyse omgeving onafhankelijk van Yoda.
Alternatief C implementeert de sluisfunctie in een neutrale zone tussen Yoda en de analyse omgeving in. Hierbij onderscheiden we nog twee varianten: Optie C1 maakt gebruik van een gedeelde disk tussen de neutrale omgeving en de analyse omgeving. Optie C2 gebruikt een exclusieve netwerk verbinding met de analyse omgeving.

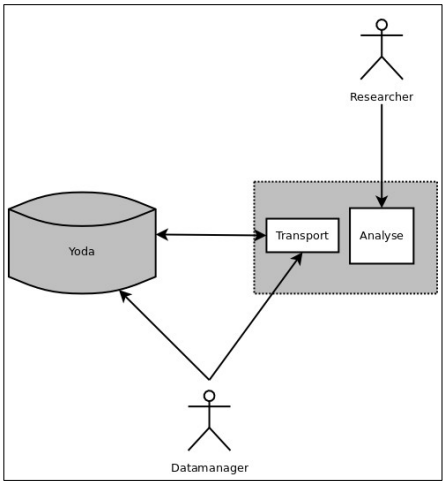
A) Sluisfunctie in Yoda, workspace met tunnel naar Yoda



Beschrijving: Analyse workspace waar vanuit alleen Yoda systeem benaderd kan worden. In Yoda wordt door de datamanager een read-only folder met data voorbereid die door onderzoekers kan worden ingelezen in de analyse workspace. Onderzoekers kunnen data vanuit de analyse workspace verplaatsen naar een daarvoor door de datamanager geprepareerde folder in Yoda.
De folders zijn niet toegankelijk vanuit andere infrastructuur omgevingen, met uitzondering van toegang door de datamanager

Impact: Yoda dient aangepast te worden om ervoor te zorgen dat processen vanuit de analyse workspace slechts met specifieke researchgroep folders kunnen communiceren, en dat deze folders niet anderszins benaderd kunnen worden. Het herkennen van de analyse omgeving als geautoriseerde client is uitdagend binnen de context van een gedistribueerde iRODS omgeving.
Daarnaast zijn in Yoda aanpassingen nodig om dit verkeer te auditen. Er is een specifiek geconfigureerde verbinding nodig tussen workspace en Yoda, die op beide endpoints ingeregeld dient te worden. Het endpoint aan Yoda zijde is daarbij meervoudig (iedere iRODS resource server is een endpoint). De analyse workspace kan door de onderzoeker rechtstreeks benaderd worden. Voorzorgen zijn getroffen om te voorkomen dat via deze verbinding data weglekt.

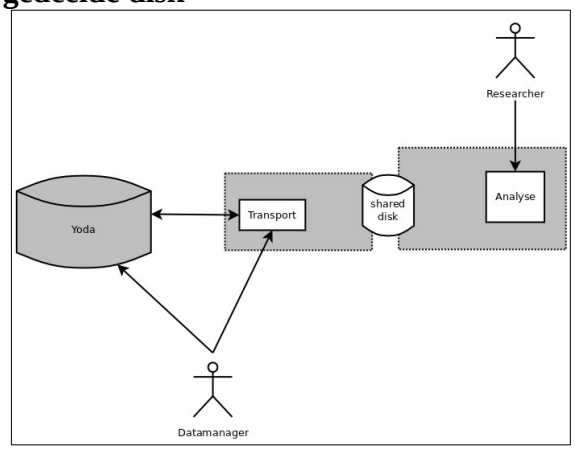
B) Sluisfunctie in workspace, data transport geautoriseerd voor datamanager



Beschrijving: Analyse workspace waar vanuit alleen de datamanager geautoriseerd is om vrij data in/uit te transporteren. Overige gebruikers kunnen data niet uit de workspace brengen. De datamanager kan vrij een Yoda service kiezen als bron/bestemming voor data. De vrije keuze in besturingssysteem commando's voor data transport maakt dat dit transport slechts in beperkte mate auditeerbaarheid is.

Impact: Binnen de analyse workspace is een extra autorisatie niveau nodig om de data manager data in/uit de workspace te laten brengen daar waar deze toegang voor een onderzoeker moet worden afgeschermd. De analyse workspace kan door de onderzoeker en data manager rechtstreeks benaderd worden. Voorzorgen zijn getroffen om te voorkomen dat via deze verbinding data weglekt.

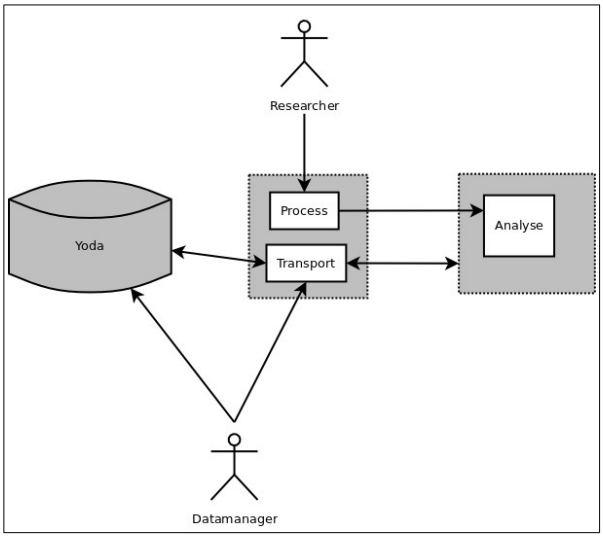
C1) Separate datamanager workspace, sluisfunctie via gedeelde disk



Beschrijving: Afgesloten analyse workspace, gebruikers kunnen data niet rechtstreeks uit de workspace brengen. Aan de analyse workspace wordt een extra filesystem gekoppeld die gedeeld wordt met een tweede workspace waarop alleen de datamanager toegang heeft. Data wordt door de datamanager vanuit Yoda op de gedeelde disk geplaatst, en de onderzoeker kan die vanuit de analsye omgeving weg lezen. Output van analyse gaat in omgekeerde richting waarbij onderzoeker de data op de gedeelde disk plaatst enz. De vrije keuze in besturingsysteem commando's voor data transport maakt dat dit transport slechts in beperkte mate auditeerbaarheid is.

Impact: Een file system koppeling is vereist naar twee workspaces. De compartimentering van data transport functie en data analyse in afzonderlijke workspaces maakt het beveiligen van de data minder complex. De analyse workspace kan door de onderzoeker nog wel rechtstreeks benaderd worden. Voorzorgen zijn getroffen om te voorkomen dat via deze verbinding data weglekt.

C2) Separate poortwachter workspace, sluisfunctie via exclusieve netwerkverbinding



Beschrijving: Afgesloten analyse workspace, alleen toegankelijk via een tweede workspace. De tweede workspace fungeert als poortwachter voor *alle* toegang tot de analyse omgeving. Datamanagers gebruiken dit als doorgeefluik voor geautoriseerd data transport tussen de analyse omgeving en Yoda. Onderzoekers gebruiken dit als proxy-toegang om in de analyse workspace te kunnen werken.

Impact: De analyse omgeving is via een exclusieve netwerk verbinding alleen benaderbaar vanuit de poortwachter. De compartimentering van data transport functie en data analyse in afzonderlijke workspaces maakt het beveiligen van de data minder complex. De toegang tot de analyse workspace is goed te reguleren met functies in de tweede workspace en daarmee wordt het risico op ongewenst data transport uit de analyse workspace beduidend beperkt. Transport van data in/uit de analyse workspace gebeurt expliciet en wordt daarmee auditeerbaar.

Vergelijking voor- en nadelen van de ontwerp opties A – C2

Hieronder volgt voor de genoemde opties een puntsgewijze opsomming van voor- en nadelen zoals besproken bij de impact van de configuratie optie. We beoordelen in hoeverre de oplossing aan de gestelde 4 eisen kan voldoen. Daarnaast beoordelen we de configuratie op de mate van vertrouwelijkheid van data die hiermee bereikt wordt en de algehele complexiteit van de oplossing als indicatie voor de te verwachten inspanning en risico's.

Kenmerk:	A	B	C1	C2
Plaats van implementatie sluisfunctie:	Yoda/ iRODS	Analyse	Analyse/ Shared	Poort-wachter
Realiseerbaar: 1. Autorisatie op basis van personen	+	+	+	+
Realiseerbaar: 2. Autorisatie op basis van netwerk lokatie	-	+	+	+
Realiseerbaar: 3. Data transfers zijn auditeerbaar	-/+	-	-	+
Realiseerbaar: 4. Sluis beinvloedt beveiliging van analyse workspace niet nadelig	+	-	+	+
Kwaliteit: afscherming analyse omgeving	-	-	+	+
Complexiteit: ontwikkelen van de oplossing	-	+	+	-/+
Eind oordeel:	**-**	**-**	**+**	**++**

Optie A is momenteel niet haalbaar: de iRODS infrastructuur onder Yoda kan nog onvoldoende onderscheid maken in client/server verbindingen om de analyse omgeving als bron van verbinding te herkennen. Daarnaast is de oplossing bewerkelijk: Alle combinaties van Yoda omgevingen en workspace omgevingen moeten door beheerders als tunnel geconfigureerd worden.

Optie B is de meest eenvoudige optie die tevens haalbaar is: de implementatie van beveiliging van de sluis blijft beperkt tot de workspace zelf. De datamanager zorgt handmatig voor uitlevering naar de juiste omgeving. Het handmatige karakter beperkt de auditeerbaarheid. Verder is doelmatige afscherming van de workspace lastig, omdat verkeerstromen op applicatief niveau geautoriseerd worden en niet generiek op infrastructureel niveau.

Optie C1 is goed realiseerbaar en heeft als voordeel dat de sluisfunctie in een enkelvoudige infrastructurele component (shared disk) is ondergebracht, wat beheer van autorisaties sterk vereenvoudigt. De datamanager workspace kan herbruikbaar ingericht worden. Afscherming van de analyse workspace is op applicatief niveau maar kan wel generiek ingeregeld worden.

Optie C2 heeft een wat hogere complexiteit dan B en C1. Voordeel is wel dat toegang tot de analyse workspace nu op generieke wijze en op infrastructureel niveau eenduidig en gemakkelijk afgeschermd kan worden, onafhankelijk van de toepassingen.