Cyber Challenge Data Set

Melissa Turcotte

Advanced Research in Cyber Systems,
Los Alamos National Laboratory

Data Sets

Daily summaries in CSV format taken from the Unified Host and Network Data Set (https://csr.lanl.gov/data/2017.html).

- Netflow
- Authentication
- Processes

In addition, labelled red team data consisting of known malicious authentications, processes and compromised devices.

- 90 days of data.
- Identifying values de-identified (anonymized).
- De-identified values match across all data sets.
- Well known network ports, processes and core enterprise hosts not de-identified.

NetFlow

NetFlow Data Set

High level summary of communications between two network devices.

Raw data consisted of NetFlow records exported from core network routers to a centralised collection server (limited to Protocols 1 (ICMP), 6 (TCP), and 17 (UDP)).

Data Transformation

- Bi-flowing
  - Aggregate duplicates
  - Marry opposing uniflows of bi-directional connections into a single, directed biflow record (estimate direction)
- IP/HostName mapping
  - Daily snapshot of device inventory + DHCP logs
  - Failed mappings anonymized as IPx rather than Comp$x$

Time, Duration, SrcDevice, DstDevice, Protocol, SrcPort, DstPort, SrcPackets, DstPackets, SrcBytes, DstBytes

Daily Summary

SrcDevice, DstDevice, Protocol, DstPort, DailyCount

```
Comp296454,Comp556624,6,706,2
Comp801423,Comp908548,17,427,2
Comp916004,Comp257274,6,Port33393,1
Comp044772,Comp669315,6,Port84767,2
Comp453311,Comp430515,6,Port23501,2
Comp044772,Comp406049,6,Port44908,2
Comp617050,Comp866402,6,Port25096,1
Comp623258,Comp304426,6,Port26269,2
Comp044772,Comp713675,6,Port46672,2
```

- 21 million events on average per day.
- 60,000 unique devices.
- Starts on day 2.

Windows Event Logs

Windows Host Log Data

Host event logs capture nuanced details about what is happening at a machine level.

- Events only related to authentication and process activity on each machine running the Windows operating system.
- JSON format, one record per line → preserve structure of original events.
- Each record has an EventID which uniquely identifies the event.
- Not all events share the same set of attributes → event dependent.

```
{“UserName”: “Comp916004$”, “EventID”: 4624, “LogHost”:“Comp916004”, “LogonID”: “0x1b5b5753”, “DomainName”: “Domain001”,
“LogonTypeDescription”: “Network”, “Source”: “Comp916004”, “AuthenticationPackage”: “Kerberos”, “Time”: 109, “LogonType”: 3}

{“UserName”: “User186321”, “EventID”: 4648, “LogHost”: “Comp916004”, “LogonID”: “0x3e4”, “DomainName”: “Domain001”, “Destination”:
“Comp457365”, “SubjectUserName”: “Comp916004$”, “ProcessName”: “Proc423620.exe”, “SubjectLogonID”: “0x3e4”, “Time”: 108, “SubjectDomainName”:
“Domain001”, “ProcessID”: “0xd28”}

{“UserName”: “User186321”, “EventID”: 4624, “LogHost”: “Comp916004”, “LogonID”: “0x1abd30dd”, “DomainName”: “Domain001”, “Source”:
“Comp916004”, “LogonTypeDescription”: “NetworkClearText”, “ProcessName”: “Proc423620.exe”, “AuthenticationPackage”: “Negotiate”, “Time”: 108,
“LogonType”: 8, “ProcessID”: “0xd28”}
```

Daily summaries for authentications and processes.

Daily summary - Authentication

UserName, SrcDevice, DstDevice, Authentication Type Description, Failure, DailyCount

```
User486765,Comp521945,Comp521945,TGS,0,6
Comp039634$,Comp039634,Comp039634,TGS,0,24
User968259,Comp223987,ActiveDirectory,TGT,0,3
Comp530762$,Comp530762,ActiveDirectory,TGS,0,22
User736129,Comp718155,None,WorkstationLock,0,4
User523107,Comp497788,ActiveDirectory,NetworkLogon,0,74
Anonymous,Comp232598,ActiveDirectory,NetworkLogon,0,241
User239777,Comp212028,None,WorkstationLock,0,1
User015915,Comp702949,Comp479002,NetworkLogon,0,250
```

- For local authentications *DstDevice* is *None* (if *DstDevice* unknown also *None*).
- User Accounts ending in $ are Computer Accounts.
- Failure is a boolean indicating whether or not the authentication was a success.
- Periodicity, which can be caused by a computer regularly renewing your credentials, can heavily inflate daily counts.
- 177,000 events on average per day.
- 15, 000 unique user accounts (not counting those ending in $).

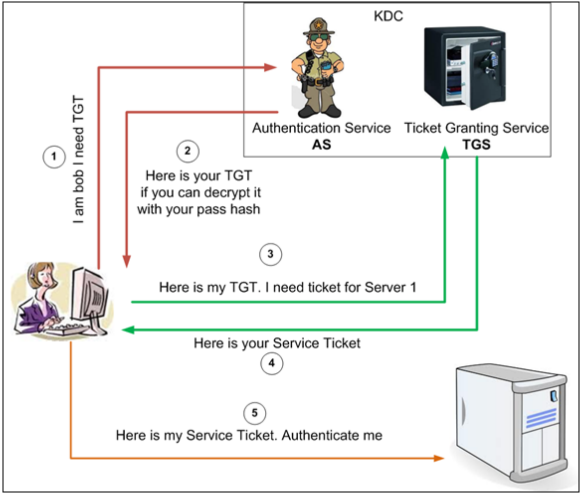
Los Alamos
NATIONAL LABORATORY

Unique Authentication Types:

- *TGT* → Ticket Granting Ticket (Kerberos)
- *TGS* → Ticket Granting Service (Kerberos)
- *InteractiveLogon* → logon at keyboard and screen of system
- *NetworkLogon* → connection to service or shared folder elsewhere on the network
- *Batch* → Scheduled task
- *Service* → Service startup
- *WorkstationLock*
- *WorkstationUnlock*
- *ScreensaverInvoked*
- *ScreensaverDismissed*
- *RemoteInteractive* → Remote Desktop Protocol
- *CachedInteractive* → logging on when away from the network, uses cached credentials.
- *CachedRemoteInteractive*

Kerberos Authentication

Computer Network Authentication protocol



https://ammarhasayen.com/2011/02/05/kerberos-in-ad-101/

Daily summary - Processes

UserName, Device, ProcessName, ParentProcessName, DailyCount

```
Comp105747$,Comp105747,csrss.exe,Proc721583,1
Comp532565$,Comp532565,Proc556285.exe,services,1
Comp824319$,Comp824319,dllhost.exe,svchost,284
Comp265993$,Comp265993,Proc299293.exe,cmd,1
Comp511123$,Comp511123,Proc391839.exe,Proc387473,2
Comp824929$,Comp824929,cscript.exe,wmiprvse,2
Comp957545$,Comp957545,Proc307031.exe,services,1
Comp512730$,Comp512730,Proc453954.exe,Proc397204,1
Comp698292$,Comp698292,Proc174492.exe,svchost,3
User851869,Comp445233,Proc915692.exe,Proc318985,1
```

- 622,000 events on average per day.
- 26, 000 unique process names.

Red Team Data

- Known malicious authentications from stolen credentials (same format as above)
- Known malicious processes (same format as above)
- Known compromised hosts where the red team had command and control