Security Analytics Use Cases and Data

COMP90073
Security Analytics

Dr. Yi Han, CIS

Semester 2, 2021

Outline

- Security Analytics Use Cases
- Security Data
- Research Benchmark Datasets Overview

Security Analytics Use Cases

- Incident Investigation and Forensics
- Security Monitoring
- Advanced Threat Detection
- Incident Response
- Compliance
- Fraud Analytics and Detection
- Insider Threat Detection

Incident Investigation and Forensics

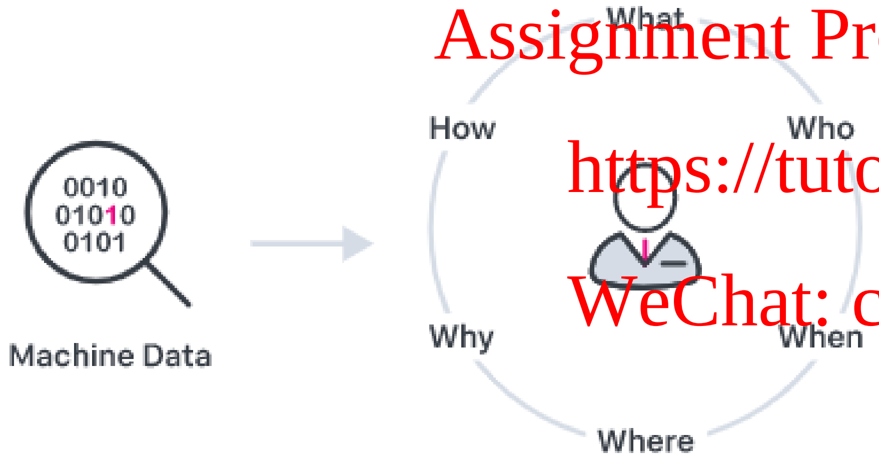
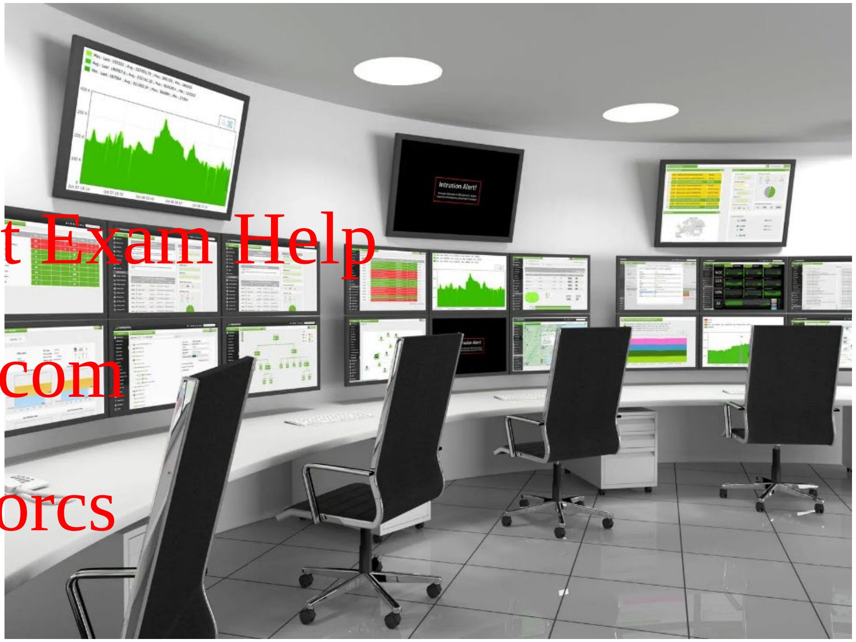


Image source: www.splunk.com

- Security incidents can occur without warning and can often go undetected long enough to pose a serious threat to an organization. Usually by the time security teams are aware of an issue, there's a good chance the damage has been done. [1]

Security Monitoring

- Security monitoring enables you to analyse a continuous stream of near-real-time data for threats and other potential security issues. Data sources for monitoring include network and endpoint systems–as well as cloud devices, data centre systems and applications. [1]



https://digitalguardian.com/blog/how-build-security-operations-center-soc-peoples-processes-and-technologies

Advanced Threat Detection

- An advanced persistent threat (APT) is a set of stealthy and continuous computer-hacking processes, often orchestrated by a person or persons targeting a specific entity. APTs usually target private organizations and/or states for business or political motives. [1]

Image source: www.splunk.com

Incident Response

- Incident Response (IR) involves the monitoring and detection of security events on IT systems, and the execution of response plans to those events. IR Teams are sometimes called blue teams. Blue teams defend an organization's infrastructure when threats are detected, whereas red teams attempt to discover weaknesses in the existing configuration of those same systems. [1]

Compliance

- In nearly all environments, there are regulatory requirements in one form or another–especially when dealing with the likes of General Data Protection Regulation (GDPR), Health Insurance Portability and Accountability Act (HIPAA), Payment Card Industry Data Security Standard (PCI DSS), Sarbanes Oxley (SOX) and even common guidelines that aren’t considered true compliance. [1]

Fraud Analytics and Detection

- Machine data plays a pivotal role in and is at the heart of detecting fraudulent activities in the digital age. [1]

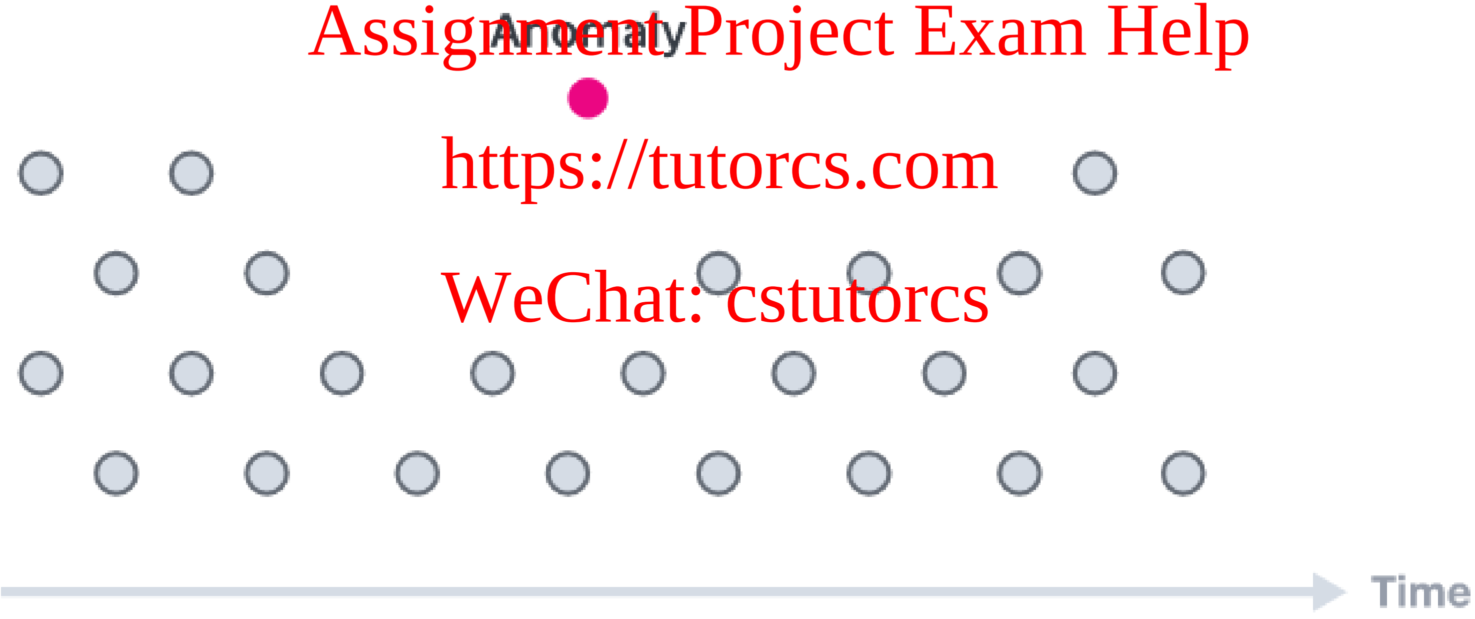


Image source: www.splunk.com

Insider Threat Detection

Deter
Detect
Disrupt

Image source: www.splunk.com

- Insider threats come from current or former employees, contractors or partners who have access to the corporate network and intentionally or accidentally exfiltrate, misuse or destroy sensitive data. They often have legitimate access to networks and permission to download sensitive material, easily evading traditional security products. [1]

THE UNIVERSITY OF MELBOURNE

Outline

- Security Analytics Use Cases
- Security Data
- Research Benchmark Datasets Overview

Security Data

- Common Attributes
- Network
- Endpoint
- Authentication
- Web Activity

Common Attributes

- Real-world data
  - Unlabelled
  - A lot of attributes
- Generic attributes
  - **Who**
    - *e.g., user/machine/network/domain identification*
  - **What**
    - *e.g., process/application/file/action*
  - **When**
    - *e.g., time zone, timestamp*
  - **Where**
    - *e.g., source, destination*

Common Attributes

- TCP/IP five-tuple
  - Source IP address
  - Source port
  - Destination IP address
  - Destination port
  - Protocol
    - 1: ICMP
    - 6: TCP
    - 17: UDP

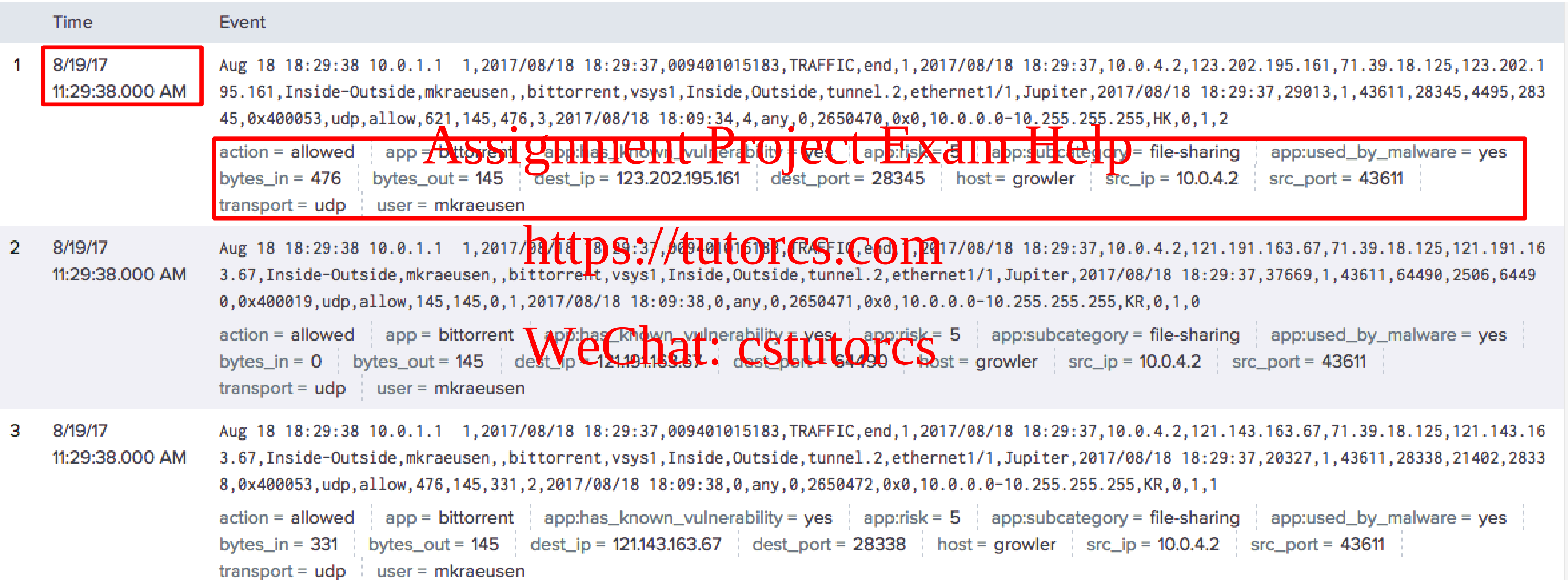
Network

“Visibility into network traffic is critical for any security team. The priority is to see what types of traffic are entering and exiting your network. It’s critical to see the traffic that’s permitted as well as communication attempts that have been blocked.” [1]

Sample source

- Firewall traffic logs

Example: Firewall Traffic Logs



	Time	Event
1	8/19/17 11:29:38.000 AM	Aug 18 18:29:38 10.0.1.1 1,2017/08/18 18:29:37,009401015183,TRAFFIC,end,1,2017/08/18 18:29:37,10.0.4.2,123.202.195.161,71.39.18.125,123.202.195.161,Inside-Outside,mkraeusen,,bittorrent,vsys1,Inside,Outside,tunnel.2,ethernet1/1,Jupiter,2017/08/18 18:29:37,29013,1,43611,28345,4495,28345,0x400053,udp,allow,621,145,476,3,2017/08/18 18:09:34,4,any,0,2650470,0x0,10.0.0.0-10.255.255.255,HK,0,1,2 action = allowed | app = bittorrent | app:has_known_vulnerability = yes | app:risk = 5 | app:subcategory = file-sharing | app:used_by_malware = yes | bytes_in = 476 | bytes_out = 145 | dest_ip = 123.202.195.161 | dest_port = 28345 | host = growler | src_ip = 10.0.4.2 | src_port = 43611 | transport = udp | user = mkraeusen
2	8/19/17 11:29:38.000 AM	Aug 18 18:29:38 10.0.1.1 1,2017/08/18 18:29:37,009401015183,TRAFFIC,end,1,2017/08/18 18:29:37,10.0.4.2,121.191.163.67,71.39.18.125,121.191.163.67,Inside-Outside,mkraeusen,,bittorrent,vsys1,Inside,Outside,tunnel.2,ethernet1/1,Jupiter,2017/08/18 18:29:37,37669,1,43611,64490,2506,64490,0x400019,udp,allow,145,145,0,1,2017/08/18 18:09:38,0,any,0,2650471,0x0,10.0.0.0-10.255.255.255,KR,0,1,0 action = allowed | app = bittorrent | app:has_known_vulnerability = yes | app:risk = 5 | app:subcategory = file-sharing | app:used_by_malware = yes | bytes_in = 0 | bytes_out = 145 | dest_ip = 121.191.163.67 | dest_port = 64490 | host = growler | src_ip = 10.0.4.2 | src_port = 43611 | transport = udp | user = mkraeusen
3	8/19/17 11:29:38.000 AM	Aug 18 18:29:38 10.0.1.1 1,2017/08/18 18:29:37,009401015183,TRAFFIC,end,1,2017/08/18 18:29:37,10.0.4.2,121.143.163.67,71.39.18.125,121.143.163.67,Inside-Outside,mkraeusen,,bittorrent,vsys1,Inside,Outside,tunnel.2,ethernet1/1,Jupiter,2017/08/18 18:29:37,20327,1,43611,28338,21402,28338,0x400053,udp,allow,476,145,331,2,2017/08/18 18:09:38,0,any,0,2650472,0x0,10.0.0.0-10.255.255.255,KR,0,1,1 action = allowed | app = bittorrent | app:has_known_vulnerability = yes | app:risk = 5 | app:subcategory = file-sharing | app:used_by_malware = yes | bytes_in = 331 | bytes_out = 145 | dest_ip = 121.143.163.67 | dest_port = 28338 | host = growler | src_ip = 10.0.4.2 | src_port = 43611 | transport = udp | user = mkraeusen

Data source: Splunk Boss of the SOC 2.0 Dataset

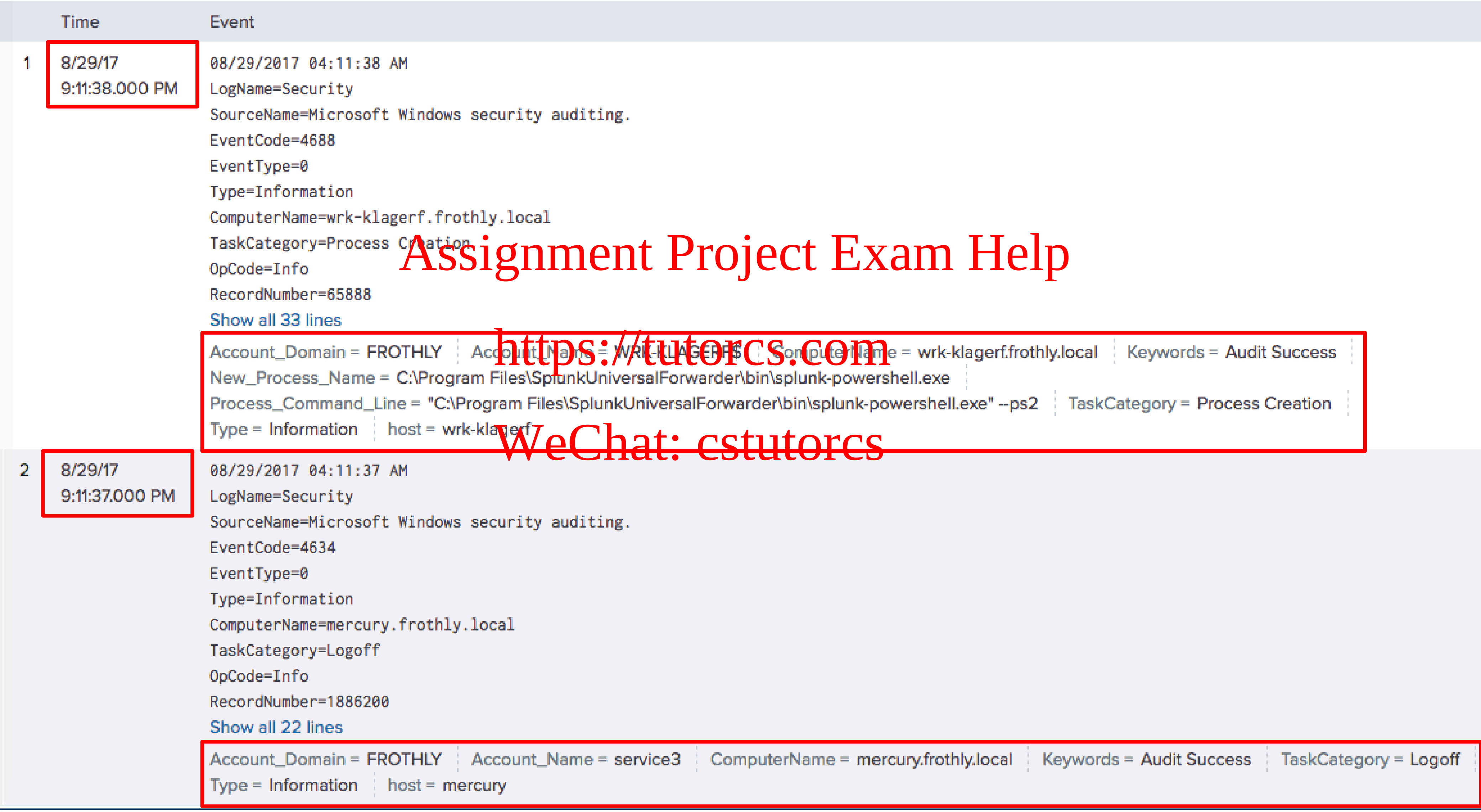
Endpoint

“Endpoint logs complement network visibility to give insight into malicious activities such as malware execution, an insider performing unauthorized activity or an attacker dwelling in your network.” [1]

Sample source

- Windows Event Logs
- Linux System Logs
- Linux Auditing System (Linux AuditD)
- MacOS System Logs

Example: Windows Event Logs



```
   Time                 Event
1  8/29/17              08/29/2017 04:11:38 AM
   9:11:38.000 PM       LogName=Security
                        SourceName=Microsoft Windows security auditing.
                        EventCode=4688
                        EventType=0
                        Type=Information
                        ComputerName=wrk-klagerf.frothly.local
                        TaskCategory=Process Creation
                        OpCode=Info
                        RecordNumber=65888
                        Show all 33 lines
                        Account_Domain = FROTHLY | Account_Name = WRK-KLAGERF$ | ComputerName = wrk-klagerf.frothly.local | Keywords = Audit Success
                        New_Process_Name = C:\Program Files\SplunkUniversalForwarder\bin\splunk-powershell.exe
                        Process_Command_Line = "C:\Program Files\SplunkUniversalForwarder\bin\splunk-powershell.exe" --ps2 | TaskCategory = Process Creation
                        Type = Information | host = wrk-klagerf
2  8/29/17              08/29/2017 04:11:37 AM
   9:11:37.000 PM       LogName=Security
                        SourceName=Microsoft Windows security auditing.
                        EventCode=4634
                        EventType=0
                        Type=Information
                        ComputerName=mercury.frothly.local
                        TaskCategory=Logoff
                        OpCode=Info
                        RecordNumber=1886200
                        Show all 22 lines
                        Account_Domain = FROTHLY | Account_Name = service3 | ComputerName = mercury.frothly.local | Keywords = Audit Success | TaskCategory = Logoff
                        Type = Information | host = mercury
```

Data source: Splunk Boss of the SOC 2.0 Dataset

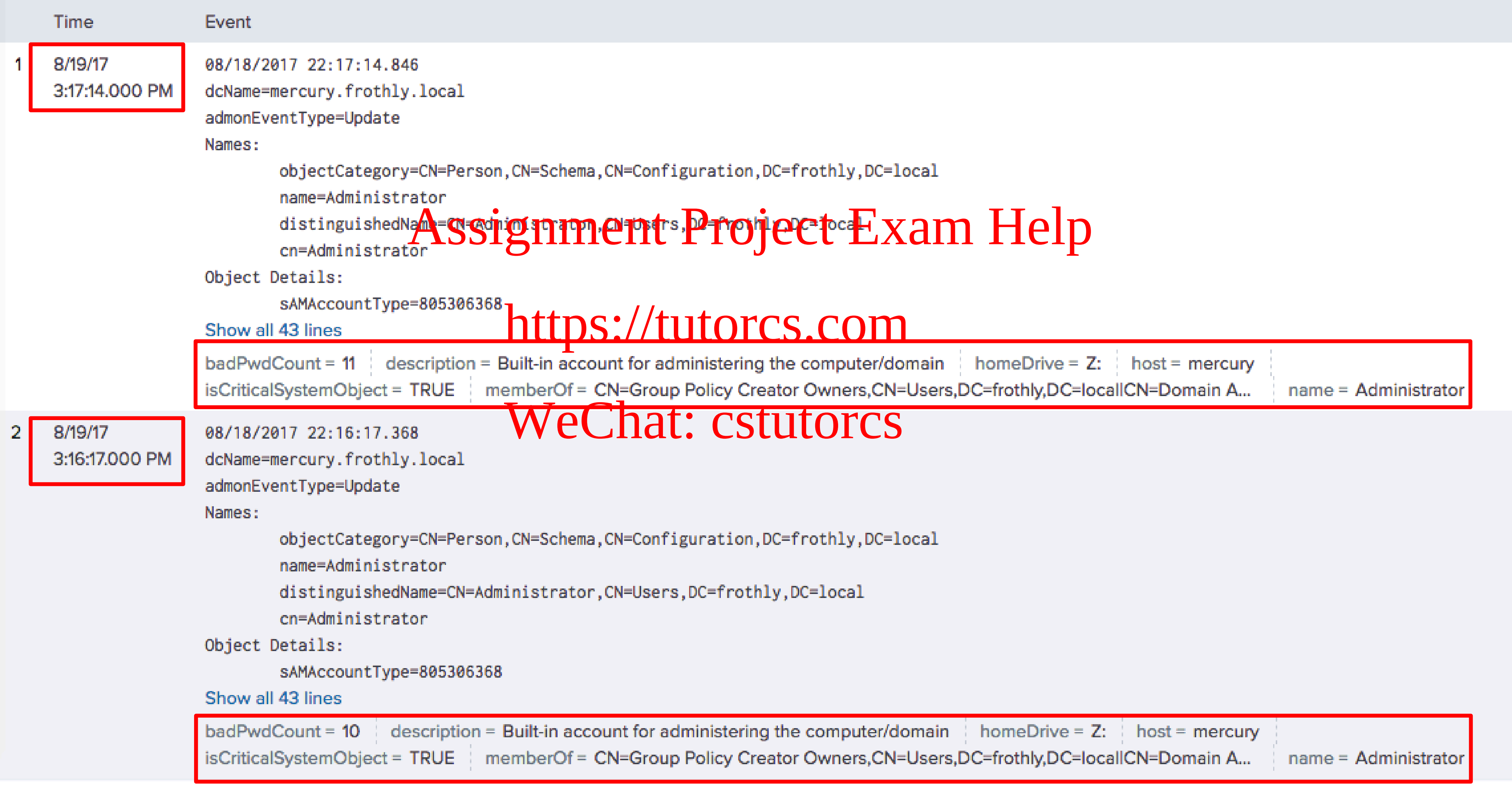
Authentication

“Authentication logs can tell you when and from where users are accessing systems and applications. Since most successful attacks eventually include the use of valid credentials, this data is critical in helping to tell the difference between a valid login and an account takeover.” [1]

Sample source

- Windows Active Directory
- Local Authentication
- Identity & Access Management (IAM)

Example: Windows Active Directory Logs



Time | Event

1 8/19/17 3:17:14.000 PM

```
08/18/2017 22:17:14.846
dcName=mercury.frothly.local
admonEventType=Update
Names:
        objectCategory=CN=Person,CN=Schema,CN=Configuration,DC=frothly,DC=local
        name=Administrator
        distinguishedName=CN=Administrator,CN=Users,DC=frothly,DC=local
        cn=Administrator
Object Details:
        sAMAccountType=805306368
```

Show all 43 lines

badPwdCount = 11 | description = Built-in account for administering the computer/domain | homeDrive = Z: | host = mercury | isCriticalSystemObject = TRUE | memberOf = CN=Group Policy Creator Owners,CN=Users,DC=frothly,DC=localCN=Domain A... | name = Administrator

2 8/19/17 3:16:17.000 PM

```
08/18/2017 22:16:17.368
dcName=mercury.frothly.local
admonEventType=Update
Names:
        objectCategory=CN=Person,CN=Schema,CN=Configuration,DC=frothly,DC=local
        name=Administrator
        distinguishedName=CN=Administrator,CN=Users,DC=frothly,DC=local
        cn=Administrator
Object Details:
        sAMAccountType=805306368
```

Show all 43 lines

badPwdCount = 10 | description = Built-in account for administering the computer/domain | homeDrive = Z: | host = mercury | isCriticalSystemObject = TRUE | memberOf = CN=Group Policy Creator Owners,CN=Users,DC=frothly,DC=localCN=Domain A... | name = Administrator

Data source: Splunk Boss of the SOC 2.0 Dataset

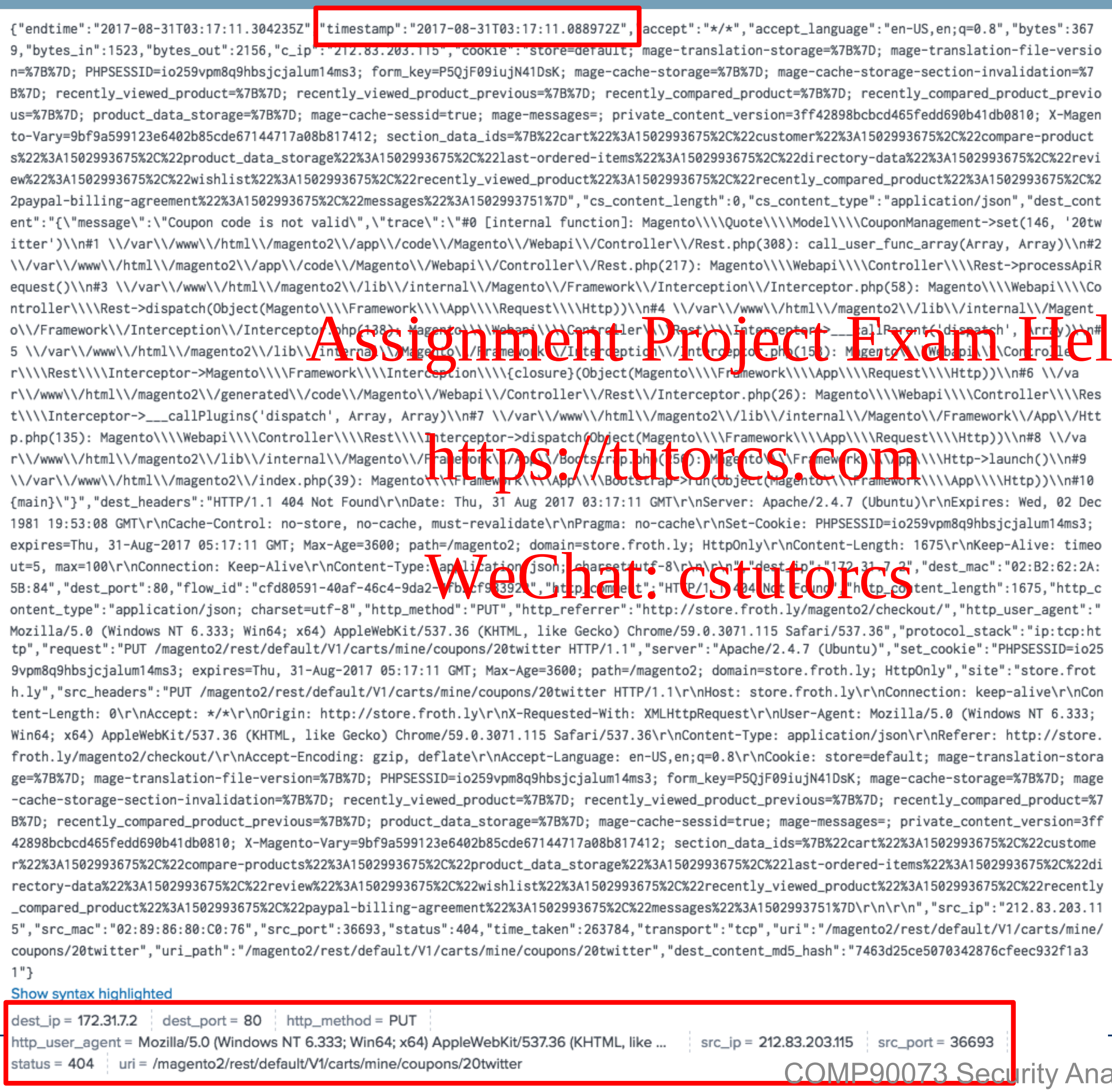
Web Activity

“Many attacks start with a user visiting a malicious website or end with valuable data being exfiltrated to a site that the attacker controls. Visibility into who’s accessing what sites and when is critical for investigation.” [1]

Sample source

- Next generation firewall (NGFW) traffic filters logs
- Web proxy logs

Example: HTTP Traffic Logs



```
{"endtime":"2017-08-31T03:17:11.304235Z","timestamp":"2017-08-31T03:17:11.088972Z","accept":"*/*","accept_language":"en-US,en;q=0.8","bytes":367
9,"bytes_in":1523,"bytes_out":2156,"c_ip":"212.83.203.115","cookie":"store=default; mage-translation-storage=%7B%7D; mage-translation-file-versio
n=%7B%7D; PHPSESSID=io259vpm8q9hbsjcjalum14ms3; form_key=P5QjF09iujN41DsK; mage-cache-storage=%7B%7D; mage-cache-storage-section-invalidation=%7
B%7D; recently_viewed_product=%7B%7D; recently_viewed_product_previous=%7B%7D; recently_compared_product=%7B%7D; recently_compared_product_previo
us=%7B%7D; product_data_storage=%7B%7D; mage-cache-sessid=true; mage-messages=; private_content_version=3ff42898bcbcd465fedd690b41db0810; X-Magen
to-Vary=9bf9a599123e6402b85cde67144717a08b817412; section_data_ids=%7B%22cart%22%3A1502993675%2C%22customer%22%3A1502993675%2C%22compare-product
s%22%3A1502993675%2C%22product_data_storage%22%3A1502993675%2C%22last-ordered-items%22%3A1502993675%2C%22directory-data%22%3A1502993675%2C%22revi
ew%22%3A1502993675%2C%22wishlist%22%3A1502993675%2C%22recently_viewed_product%22%3A1502993675%2C%22recently_compared_product%22%3A1502993675%2C%2
2paypal-billing-agreement%22%3A1502993675%2C%22messages%22%3A1502993751%7D","cs_content_length":0,"cs_content_type":"application/json","dest_cont
ent":"{\"message\":\"Coupon code is not valid\",\"trace\":\"#0 [internal function]: Magento\\\\Quote\\\\Model\\\\CouponManagement->set(146, '20tw
itter')\\n#1 \\/var\\/www\\/html\\/magento2\\/app\\/code\\/Magento\\/Webapi\\/Controller\\/Rest.php(308): call_user_func_array(Array, Array)\\n#2
\\/var\\/www\\/html\\/magento2\\/app\\/code\\/Magento\\/Webapi\\/Controller\\/Rest.php(217): Magento\\\\Webapi\\\\Controller\\\\Rest->processApiR
equest()\\n#3 \\/var\\/www\\/html\\/magento2\\/lib\\/internal\\/Magento\\/Framework\\/Interception\\/Interceptor.php(58): Magento\\\\Webapi\\\\Co
ntroller\\\\Rest->dispatch(Object(Magento\\\\Framework\\\\App\\\\Request\\\\Http))\\n#4 \\/var\\/www\\/html\\/magento2\\/lib\\/internal\\/Magent
o\\/Framework\\/Interception\\/Interceptor.php(138): Magento\\\\Webapi\\\\Controller\\\\Rest\\\\Interceptor->___callParent('dispatch', Array))\\n#
5 \\/var\\/www\\/html\\/magento2\\/lib\\/internal\\/Magento\\/Framework\\/Interception\\/Interceptor.php(153): Magento\\\\Webapi\\\\Controlle
r\\\\Rest\\\\Interceptor->Magento\\\\Framework\\\\Interception\\\\{closure}(Object(Magento\\\\Framework\\\\App\\\\Request\\\\Http))\\n#6 \\/va
r\\/www\\/html\\/magento2\\/generated\\/code\\/Magento\\/Webapi\\/Controller\\/Rest\\/Interceptor.php(26): Magento\\\\Webapi\\\\Controller\\\\Res
t\\\\Interceptor->___callPlugins('dispatch', Array, Array)\\n#7 \\/var\\/www\\/html\\/magento2\\/lib\\/internal\\/Magento\\/Framework\\/App\\/Htt
p.php(135): Magento\\\\Webapi\\\\Controller\\\\Rest\\\\Interceptor->dispatch(Object(Magento\\\\Framework\\\\App\\\\Request\\\\Http))\\n#8 \\/va
r\\/www\\/html\\/magento2\\/lib\\/internal\\/Magento\\/Framework\\/App\\/Bootstrap.php(256): Magento\\\\Framework\\\\App\\\\Http->launch()\\n#9
\\/var\\/www\\/html\\/magento2\\/index.php(39): Magento\\\\Framework\\\\App\\\\Bootstrap->run(Object(Magento\\\\Framework\\\\App\\\\Http))\\n#10
{main}\"}","dest_headers":"HTTP/1.1 404 Not Found\r\nDate: Thu, 31 Aug 2017 03:17:11 GMT\r\nServer: Apache/2.4.7 (Ubuntu)\r\nExpires: Wed, 02 Dec
1981 19:53:08 GMT\r\nCache-Control: no-store, no-cache, must-revalidate\r\nPragma: no-cache\r\nSet-Cookie: PHPSESSID=io259vpm8q9hbsjcjalum14ms3;
expires=Thu, 31-Aug-2017 05:17:11 GMT; Max-Age=3600; path=/magento2; domain=store.froth.ly; HttpOnly\r\nContent-Length: 1675\r\nKeep-Alive: timeo
ut=5, max=100\r\nConnection: Keep-Alive\r\nContent-Type: application/json; charset=utf-8\r\n\r\n","dest_ip":"172.31.7.2","dest_mac":"02:B2:62:2A:
5B:84","dest_port":80,"flow_id":"cfd80591-40af-46c4-9da2-[illegible]","http_comment":"HTTP/1.1 404 Not Found","http_content_length":1675,"http_c
ontent_type":"application/json; charset=utf-8","http_method":"PUT","http_referrer":"http://store.froth.ly/magento2/checkout/","http_user_agent":"
Mozilla/5.0 (Windows NT 6.333; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/59.0.3071.115 Safari/537.36","protocol_stack":"ip:tcp:ht
tp","request":"PUT /magento2/rest/default/V1/carts/mine/coupons/20twitter HTTP/1.1","server":"Apache/2.4.7 (Ubuntu)","set_cookie":"PHPSESSID=io25
9vpm8q9hbsjcjalum14ms3; expires=Thu, 31-Aug-2017 05:17:11 GMT; Max-Age=3600; path=/magento2; domain=store.froth.ly; HttpOnly","site":"store.frot
h.ly","src_headers":"PUT /magento2/rest/default/V1/carts/mine/coupons/20twitter HTTP/1.1\r\nHost: store.froth.ly\r\nConnection: keep-alive\r\nCon
tent-Length: 0\r\nAccept: */*\r\nOrigin: http://store.froth.ly\r\nX-Requested-With: XMLHttpRequest\r\nUser-Agent: Mozilla/5.0 (Windows NT 6.333;
Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/59.0.3071.115 Safari/537.36\r\nContent-Type: application/json\r\nReferer: http://store.
froth.ly/magento2/checkout/\r\nAccept-Encoding: gzip, deflate\r\nAccept-Language: en-US,en;q=0.8\r\nCookie: store=default; mage-translation-stora
ge=%7B%7D; mage-translation-file-version=%7B%7D; PHPSESSID=io259vpm8q9hbsjcjalum14ms3; form_key=P5QjF09iujN41DsK; mage-cache-storage=%7B%7D; mage
-cache-storage-section-invalidation=%7B%7D; recently_viewed_product=%7B%7D; recently_viewed_product_previous=%7B%7D; recently_compared_product=%7
B%7D; recently_compared_product_previous=%7B%7D; product_data_storage=%7B%7D; mage-cache-sessid=true; mage-messages=; private_content_version=3ff
42898bcbcd465fedd690b41db0810; X-Magento-Vary=9bf9a599123e6402b85cde67144717a08b817412; section_data_ids=%7B%22cart%22%3A1502993675%2C%22custome
r%22%3A1502993675%2C%22compare-products%22%3A1502993675%2C%22product_data_storage%22%3A1502993675%2C%22last-ordered-items%22%3A1502993675%2C%22di
rectory-data%22%3A1502993675%2C%22review%22%3A1502993675%2C%22wishlist%22%3A1502993675%2C%22recently_viewed_product%22%3A1502993675%2C%22recently
_compared_product%22%3A1502993675%2C%22paypal-billing-agreement%22%3A1502993675%2C%22messages%22%3A1502993751%7D\r\n\r\n","src_ip":"212.83.203.11
5","src_mac":"02:89:86:80:C0:76","src_port":36693,"status":404,"time_taken":263784,"transport":"tcp","uri":"/magento2/rest/default/V1/carts/mine/
coupons/20twitter","uri_path":"/magento2/rest/default/V1/carts/mine/coupons/20twitter","dest_content_md5_hash":"7463d25ce5070342876cfeec932f1a3
1"}
```

dest_ip = 172.31.7.2 | dest_port = 80 | http_method = PUT
http_user_agent = Mozilla/5.0 (Windows NT 6.333; Win64; x64) AppleWebKit/537.36 (KHTML, like ... | src_ip = 212.83.203.115 | src_port = 36693
status = 404 | uri = /magento2/rest/default/V1/carts/mine/coupons/20twitter

Data source: Splunk Boss of the SOC 2.0 Dataset

THE UNIVERSITY OF MELBOURNE

Outline

- Security Analytics Use Cases
- Security Data
- Research Benchmark Datasets Overview

Research Benchmark Datasets Overview

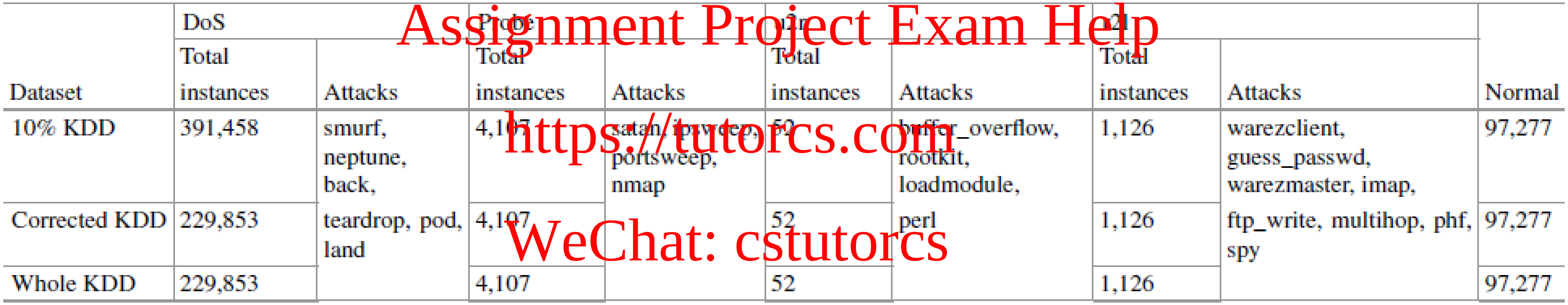
- KDDcup99 Dataset
- NSL-KDD Dataset
- DARPA 2000 Dataset
- CAIDA Dataset

KDDcup99 Dataset

- Most widely used dataset to evaluate Network based Anomaly Detection methods & systems. Attack scenarios include:
  - Denial of service (DoS): An attacker attempts to prevent valid users from using a service provided by a system
  - Remote to local (r2l): Attackers try to gain entrance to a victim machine without having an account on it, e.g., guessing password
  - User to root (u2r): Attackers have access to a local victim machine and attempt to gain privilege of a superuser (root)
  - Probing: Attackers attempt to acquire information about the target host, e.g., port scanning.

KDDcup99 Dataset

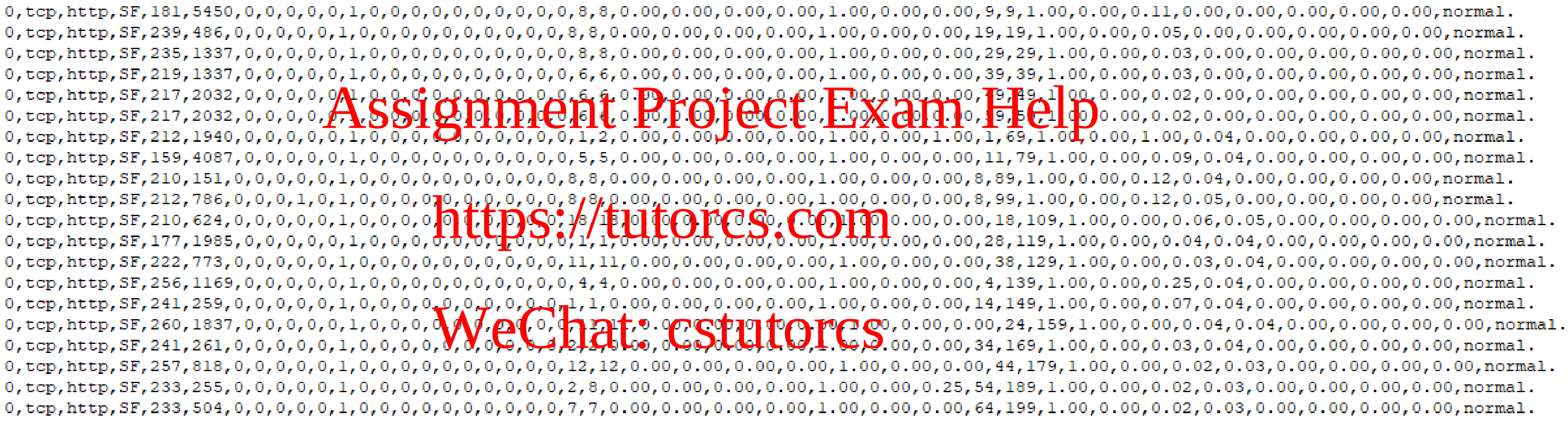
Table - Distribution of normal and attack traffic instances [2]



	DoS		Probe		u2r		r2l		
Dataset	Total instances	Attacks	Total instances	Attacks	Total instances	Attacks	Total instances	Attacks	Normal
10% KDD	391,458	smurf, neptune, back, teardrop, pod, land	4,107	satan, ipsweep, portsweep, nmap	52	buffer_overflow, rootkit, loadmodule, perl	1,126	warezclient, guess_passwd, warezmaster, imap, ftp_write, multihop, phf, spy	97,277
Corrected KDD	229,853		4,107		52		1,126		97,277
Whole KDD	229,853		4,107		52		1,126		97,277

KDDcup99 Dataset

- Download: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html
  - Snippet



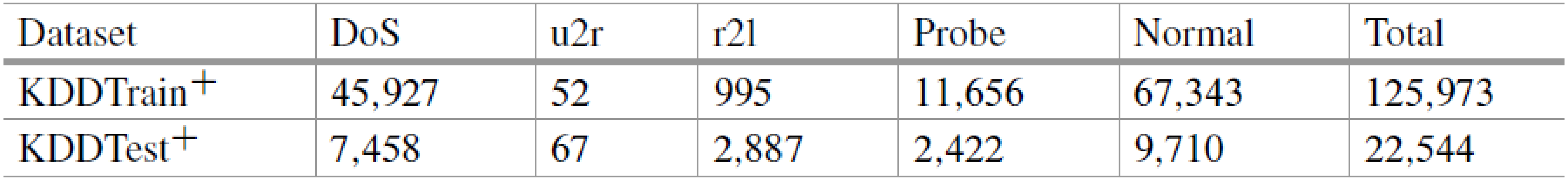
```
0,tcp,http,SF,181,5450,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,8,8,0.00,0.00,0.00,0.00,1.00,0.00,0.00,9,9,1.00,0.00,0.11,0.00,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,239,486,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,8,8,0.00,0.00,0.00,0.00,1.00,0.00,0.00,19,19,1.00,0.00,0.05,0.00,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,235,1337,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,8,8,0.00,0.00,0.00,0.00,1.00,0.00,0.00,29,29,1.00,0.00,0.03,0.00,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,219,1337,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,6,6,0.00,0.00,0.00,0.00,1.00,0.00,0.00,39,39,1.00,0.00,0.03,0.00,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,217,2032,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,6,6,0.00,0.00,0.00,0.00,1.00,0.00,0.00,49,49,1.00,0.00,0.02,0.00,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,217,2032,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,6,6,0.00,0.00,0.00,0.00,1.00,0.00,0.00,59,59,1.00,0.00,0.02,0.00,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,212,1940,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,2,0.00,0.00,0.00,0.00,1.00,0.00,1.00,1,69,1.00,0.00,1.00,0.04,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,159,4087,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,5,5,0.00,0.00,0.00,0.00,1.00,0.00,0.00,11,79,1.00,0.00,0.09,0.04,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,210,151,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,8,8,0.00,0.00,0.00,0.00,1.00,0.00,0.00,8,89,1.00,0.00,0.12,0.04,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,212,786,0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,8,8,0.00,0.00,0.00,0.00,1.00,0.00,0.00,8,99,1.00,0.00,0.12,0.05,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,210,624,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,18,18,0.00,0.00,0.00,0.00,1.00,0.00,0.00,18,109,1.00,0.00,0.06,0.05,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,177,1985,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,1,0.00,0.00,0.00,0.00,1.00,0.00,0.00,28,119,1.00,0.00,0.04,0.04,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,222,773,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,11,11,0.00,0.00,0.00,0.00,1.00,0.00,0.00,38,129,1.00,0.00,0.03,0.04,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,256,1169,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,4,4,0.00,0.00,0.00,0.00,1.00,0.00,0.00,4,139,1.00,0.00,0.25,0.04,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,241,259,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,1,1,0.00,0.00,0.00,0.00,1.00,0.00,0.00,14,149,1.00,0.00,0.07,0.04,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,260,1837,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,11,11,0.00,0.00,0.00,0.00,1.00,0.00,0.00,24,159,1.00,0.00,0.04,0.04,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,241,261,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,2,2,0.00,0.00,0.00,0.00,1.00,0.00,0.00,34,169,1.00,0.00,0.03,0.04,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,257,818,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,12,12,0.00,0.00,0.00,0.00,1.00,0.00,0.00,44,179,1.00,0.00,0.02,0.03,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,233,255,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,2,8,0.00,0.00,0.00,0.00,1.00,0.00,0.25,54,189,1.00,0.00,0.02,0.03,0.00,0.00,0.00,0.00,normal.
0,tcp,http,SF,233,504,0,0,0,0,0,1,0,0,0,0,0,0,0,0,0,0,7,7,0.00,0.00,0.00,0.00,1.00,0.00,0.00,64,199,1.00,0.00,0.02,0.03,0.00,0.00,0.00,0.00,normal.
```

  - Field description http://kdd.ics.uci.edu/databases/kddcup99/kddcup.names

NSL-KDD Dataset

- Problem with KDDcup99: redundant records [3]
  - 78% and 75% of the records are duplicated in the train and test set
- A new dataset consisting of selected records of KDDcup99 dataset which improves the evaluation performance
  - Description: https://www.unb.ca/cic/datasets/nsl.html
  - Download: https://github.com/jmnwong/NSL-KDD-Dataset

Table - Distribution of normal and attack traffic instances [2]



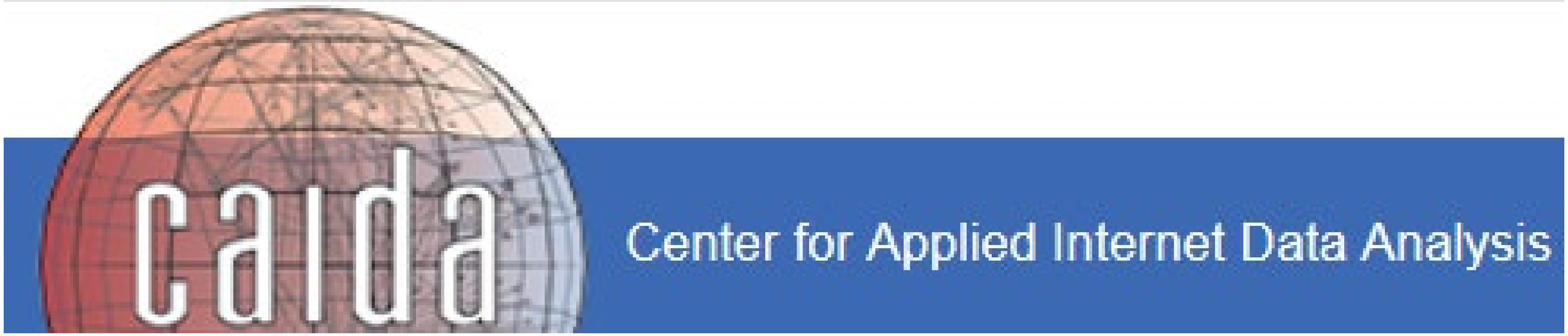
Dataset	DoS	u2r	r2l	Probe	Normal	Total
KDDTrain$^+$	45,927	52	995	11,656	67,343	125,973
KDDTest$^+$	7,458	67	2,887	2,422	9,710	22,544

DARPA 2000 Dataset

- This dataset targets evaluating detection of complex attacks that contains multiple steps.
  - Description & Download: https://www.ll.mit.edu/r-d/datasets/2000-darpa-intrusion-detection-scenario-specific-datasets
  - It includes five attack phases:
    - IPSweep
    - Probing
    - Breaking into the system by exploiting vulnerability
    - Installing DDoS software for the compromised system
    - Launching DDoS attack against another target

CAIDA Dataset

- CAIDA collects many different types of data and makes them available to the research community. CAIDA datasets are very specific to particular events or attacks, such as the DDoS 2007 dataset. Most of its longer traces are anonymized backbone traces without their payload.
- Description & Download: https://www.caida.org/catalog/datasets/overview/#H2279



Other Datasets

- Markus Ring, Sarah Wunderlich, Deniz Scheuring, Dieter Landes, Andreas Hotho, “A Survey of Network-based Intrusion Detection Data Sets”, arXiv:1903.02460, https://arxiv.org/abs/1903.02460

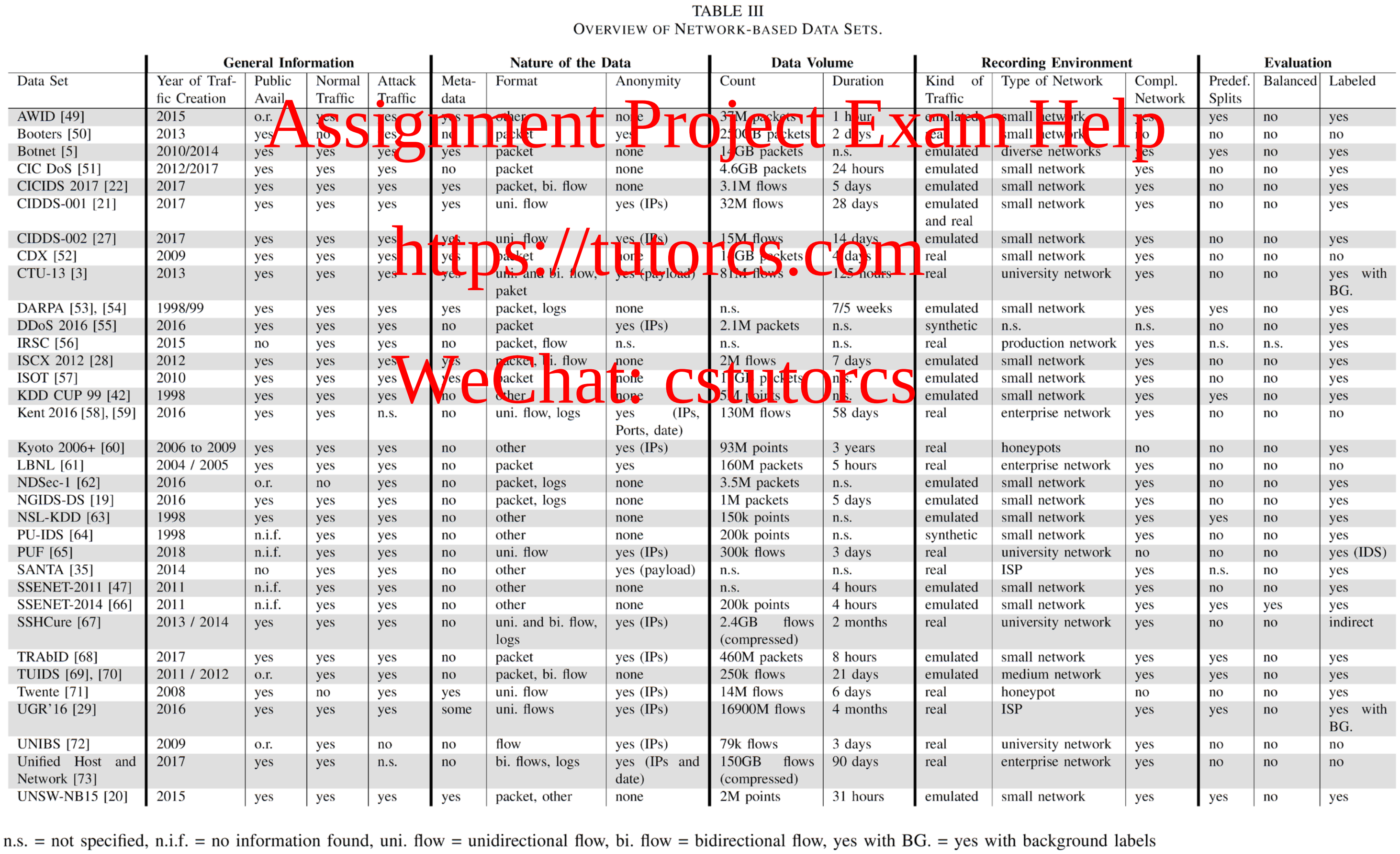


TABLE III
OVERVIEW OF NETWORK-BASED DATA SETS.

	General Information				Nature of the Data			Data Volume		Recording Environment			Evaluation		
Data Set	Year of Traffic Creation	Public Avail.	Normal Traffic	Attack Traffic	Metadata	Format	Anonymity	Count	Duration	Kind of Traffic	Type of Network	Compl. Network	Predef. Splits	Balanced	Labeled
AWID [49]	2015	o.r.	yes	yes	yes	other	none	37M packets	1 hour	emulated	small network	yes	yes	no	yes
Booters [50]	2013	yes	no	yes	no	packet	yes	250GB packets	2 days	real	small network	no	no	no	no
Botnet [5]	2010/2014	yes	yes	yes	yes	packet	none	14GB packets	n.s.	emulated	diverse networks	yes	yes	no	yes
CIC DoS [51]	2012/2017	yes	yes	yes	no	packet	none	4.6GB packets	24 hours	emulated	small network	yes	no	no	yes
CICIDS 2017 [22]	2017	yes	yes	yes	yes	packet, bi. flow	none	3.1M flows	5 days	emulated	small network	yes	no	no	yes
CIDDS-001 [21]	2017	yes	yes	yes	yes	uni. flow	yes (IPs)	32M flows	28 days	emulated and real	small network	yes	no	no	yes
CIDDS-002 [27]	2017	yes	yes	yes	yes	uni. flow	yes (IPs)	15M flows	14 days	emulated	small network	yes	no	no	yes
CDX [52]	2009	yes	yes	yes	yes	packet	none	14GB packets	4 days	real	small network	yes	no	no	no
CTU-13 [3]	2013	yes	yes	yes	yes	uni. and bi. flow, paket	yes (payload)	81M flows	125 hours	real	university network	yes	no	no	yes with BG.
DARPA [53], [54]	1998/99	yes	yes	yes	yes	packet, logs	none	n.s.	7/5 weeks	emulated	small network	yes	yes	no	yes
DDoS 2016 [55]	2016	yes	yes	yes	no	packet	yes (IPs)	2.1M packets	n.s.	synthetic	n.s.	n.s.	no	no	yes
IRSC [56]	2015	no	yes	yes	no	packet, flow	n.s.	n.s.	n.s.	real	production network	yes	n.s.	n.s.	yes
ISCX 2012 [28]	2012	yes	yes	yes	yes	packet, bi. flow	none	2M flows	7 days	emulated	small network	yes	no	no	yes
ISOT [57]	2010	yes	yes	yes	yes	packet	none	11GB packets	n.s.	emulated	small network	yes	no	no	yes
KDD CUP 99 [42]	1998	yes	yes	yes	no	other	none	5M points	n.s.	emulated	small network	yes	yes	no	yes
Kent 2016 [58], [59]	2016	yes	yes	n.s.	no	uni. flow, logs	yes (IPs, Ports, date)	130M flows	58 days	real	enterprise network	yes	no	no	no
Kyoto 2006+ [60]	2006 to 2009	yes	yes	yes	no	other	yes (IPs)	93M points	3 years	real	honeypots	no	no	no	yes
LBNL [61]	2004 / 2005	yes	yes	yes	no	packet	yes	160M packets	5 hours	real	enterprise network	yes	no	no	no
NDSec-1 [62]	2016	o.r.	no	yes	no	packet, logs	none	3.5M packets	n.s.	emulated	small network	yes	no	no	yes
NGIDS-DS [19]	2016	yes	yes	yes	no	packet, logs	none	1M packets	5 days	emulated	small network	yes	no	no	yes
NSL-KDD [63]	1998	yes	yes	yes	no	other	none	150k points	n.s.	emulated	small network	yes	yes	no	yes
PU-IDS [64]	1998	n.i.f.	yes	yes	no	other	none	200k points	n.s.	synthetic	small network	yes	no	no	yes
PUF [65]	2018	n.i.f.	yes	yes	no	uni. flow	yes (IPs)	300k flows	3 days	real	university network	no	no	no	yes (IDS)
SANTA [35]	2014	no	yes	yes	no	other	yes (payload)	n.s.	n.s.	real	ISP	yes	n.s.	no	yes
SSENET-2011 [47]	2011	n.i.f.	yes	yes	no	other	none	n.s.	4 hours	emulated	small network	yes	no	no	yes
SSENET-2014 [66]	2011	n.i.f.	yes	yes	no	other	none	200k points	4 hours	emulated	small network	yes	yes	yes	yes
SSHCure [67]	2013 / 2014	yes	yes	yes	no	uni. and bi. flow, logs	yes (IPs)	2.4GB flows (compressed)	2 months	real	university network	yes	no	no	indirect
TRAbID [68]	2017	yes	yes	yes	no	packet	yes (IPs)	460M packets	8 hours	emulated	small network	yes	yes	no	yes
TUIDS [69], [70]	2011 / 2012	o.r.	yes	yes	no	packet, bi. flow	none	250k flows	21 days	emulated	medium network	yes	yes	no	yes
Twente [71]	2008	yes	no	yes	yes	uni. flow	yes (IPs)	14M flows	6 days	real	honeypot	no	no	no	yes
UGR’16 [29]	2016	yes	yes	yes	some	uni. flows	yes (IPs)	16900M flows	4 months	real	ISP	yes	yes	no	yes with BG.
UNIBS [72]	2009	o.r.	yes	no	no	flow	yes (IPs)	79k flows	3 days	real	university network	yes	no	no	no
Unified Host and Network [73]	2017	yes	yes	n.s.	no	bi. flows, logs	yes (IPs and date)	150GB flows (compressed)	90 days	real	enterprise network	yes	no	no	no
UNSW-NB15 [20]	2015	yes	yes	yes	yes	packet, other	none	2M points	31 hours	emulated	small network	yes	yes	no	yes

n.s. = not specified, n.i.f. = no information found, uni. flow = unidirectional flow, bi. flow = bidirectional flow, yes with BG. = yes with background labels

Summary

- Security analytics use cases
  - Explain seven common use cases
- Security data
  - Explain four primary categories of data sources
  - Select common attributes
  - Understand the role of each data source in detecting cyber threats
- Research benchmark datasets
  - Understand the primary use case for each dataset

Reference

- [1] Splunk Inc., 2021, *The Essential Guide to Security 2021*
- [2] M.H.Bhuyan, et al., 2017, *Network Traffic Anomaly Detection and Prevention*, Springer
- [3] M. Tavallaee, E. Bagheri, W. Lu and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, 2009, pp. 1-6, doi: 10.1109/CISDA.2009.5356528.