HITB+CyberWeek
Abu Dhabi, UAE: 12-17 October 2019

Security for AI or AI for Security?

Sergey Gordeychik
HTTP://SCADA.SL
@SCADASL
serg.gordey@gmail.com

https://cyberweek.ae

Sergey Gordeychik

- AI and Cybersecurity Executive
  - Abu Dhabi, UAE
- Visiting Professor, Cyber Security
  - Harbour.Space University, Barcelona, Spain
- Program Chair, PHDays Conference
  - www.phdays.com, Moscow

- Cyber-physical troublemaker
  - Leader of SCADA Strangelove Research Team
  - www.scada.sl, @scadasl
- Ex…
  - Deputy CTO, Kaspersky Lab
  - CTO, Positive Technologies
  - Gartner recognized products and services

Disclaimer

Please note, that this talk is by Sergey and AISec group.

We don't speak for our employers.

All the opinions and information here are of our responsibility. So, mistakes and bad jokes are all OUR responsibilities.

Actually no one ever saw this talk before.

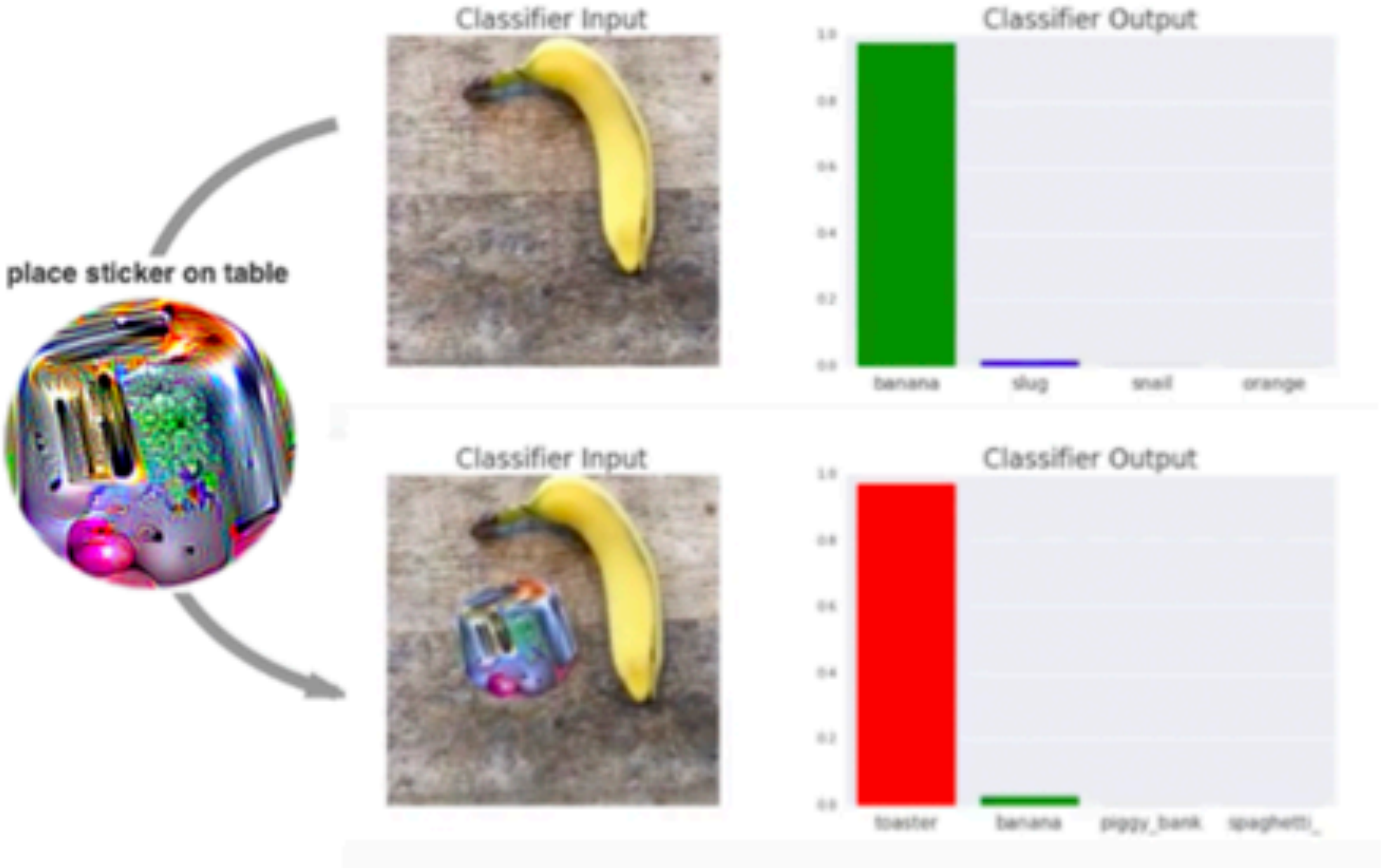
https://en.wikipedia.org/wiki/Terms_and_Conditions_May_Apply

BIG BROTHER
CCTV
STOP
UNW
2

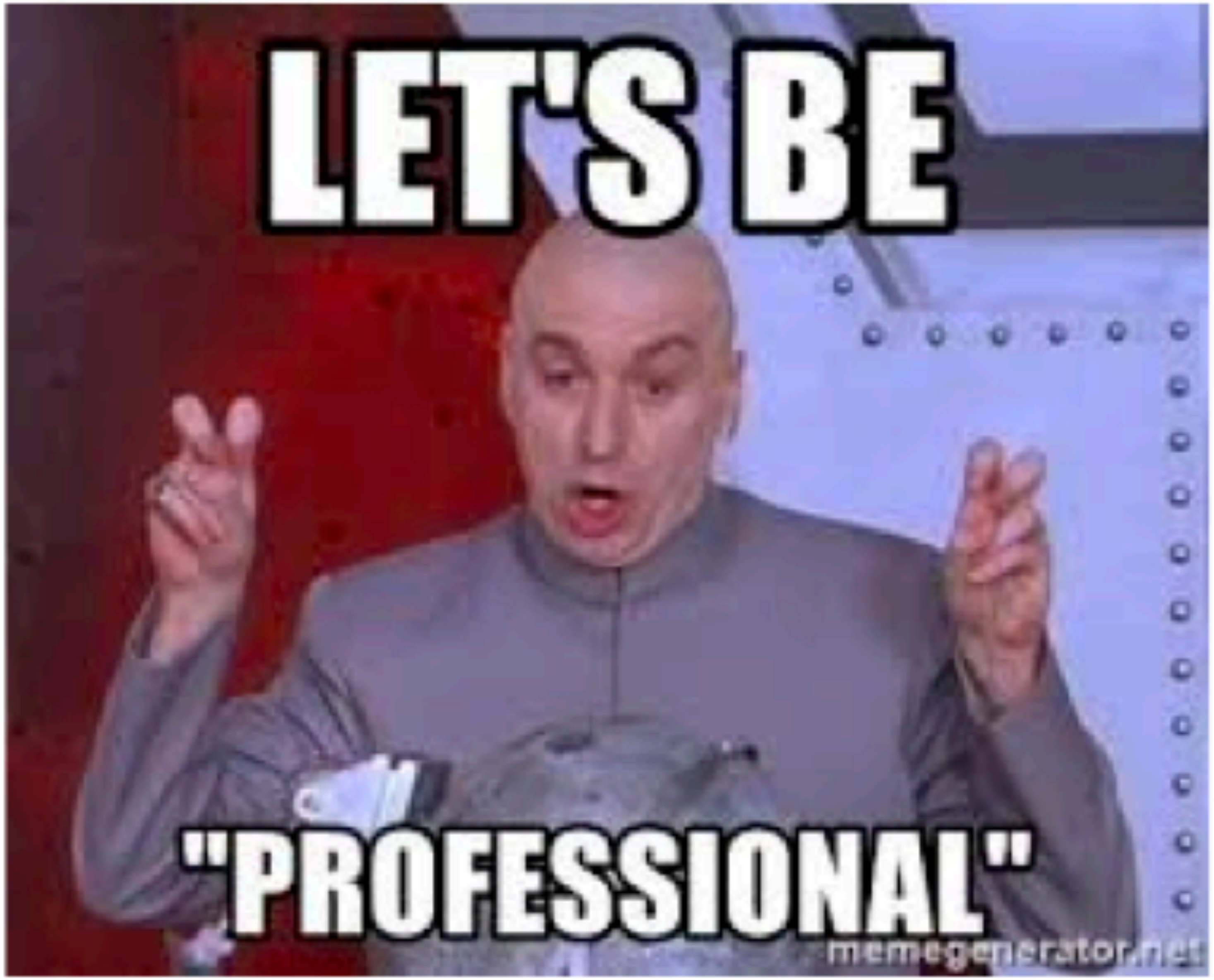
PWN?

Adversarial example anyone?

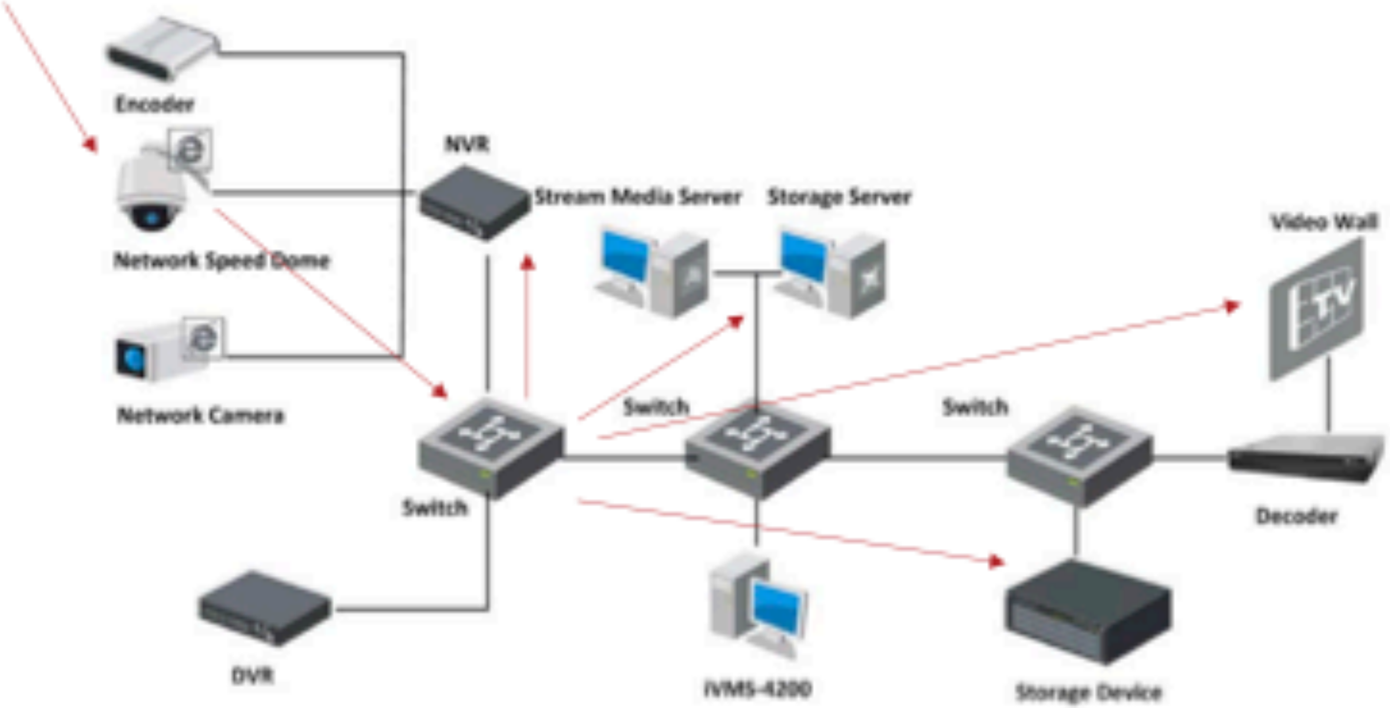
Adversarial example?



machinelearning
artificialintelligence
deeplearning
ai
bigdata
datascience
artificial
iiot
dl
a.i.
ml
ar
vr
insurtech
mpgvip
chatbots
ai-powered
techcx
wef17
machine
future
human
analytics
robotics
bots
blockchain
devops
nlp
chatbot
iottrends
industry40
internetofthings
neuralnetworks
technology
startups
cloud
ibm
dx
drones
automation
algorithms
deep
robot
api
data
rise
bot
ioe
serverless
cybersecurity
machines
python
healthcare
startup
real-time
self-driving
neural
regtech
spdc
tensorflow
apm
m2m
next-gen
disruption
autonomous
smartcities
revolution
fintech
ai's
sdn
practitioners
data-driven
robots
powered
impact
humans
banking
tech


LET'S BE
"PROFESSIONAL"
memegenerator.net

Hacking as usual…



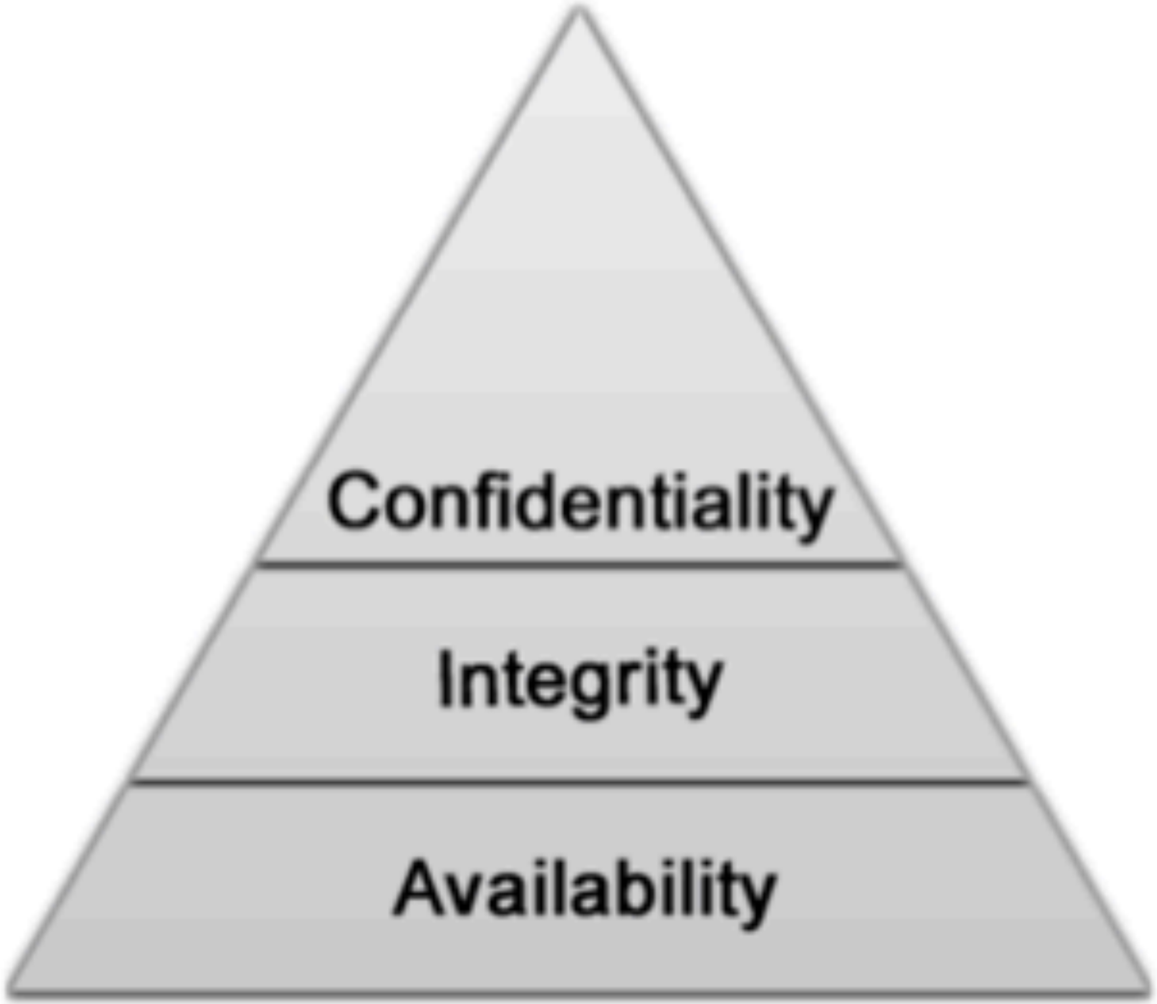
https://slideplayer.com/slide/4378533/

What is Cyber?

What is Cybersecurity?

Cybersecurity goals?

HOLY
CIA
TRINITY

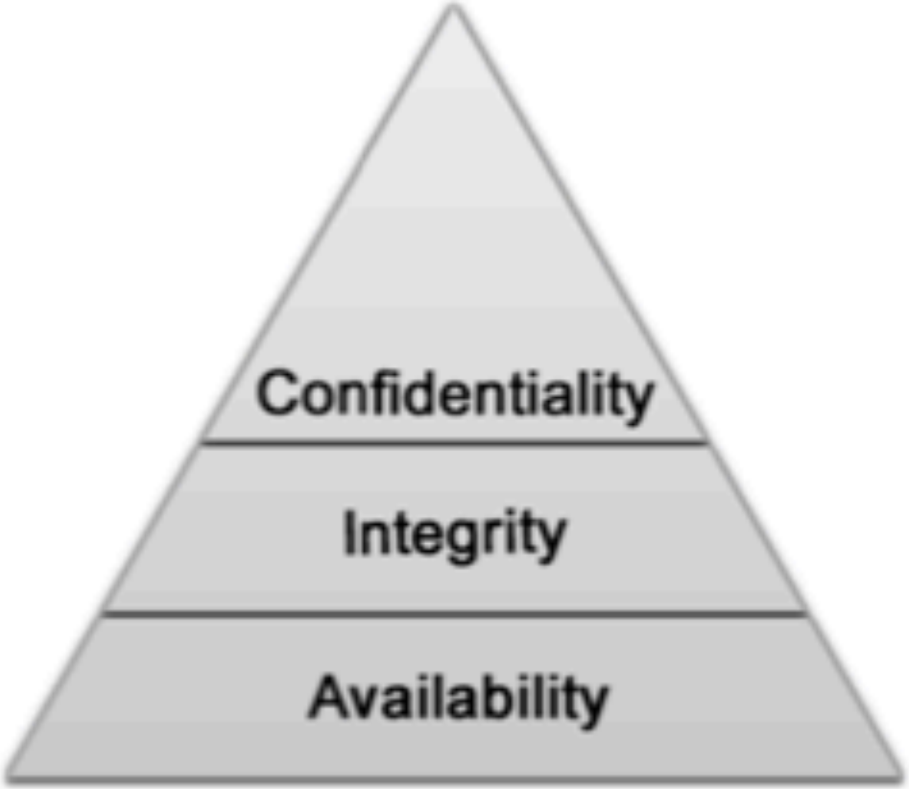


Confidentiality

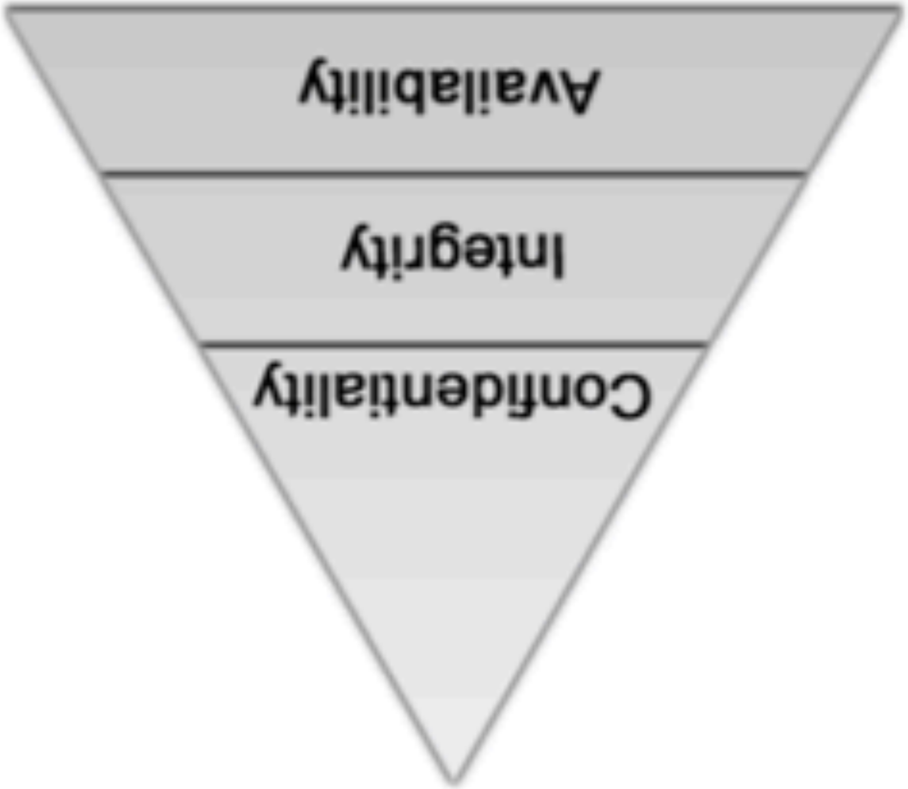
Integrity

Availability

OT/ICS/SCADA Security?!



IT domain

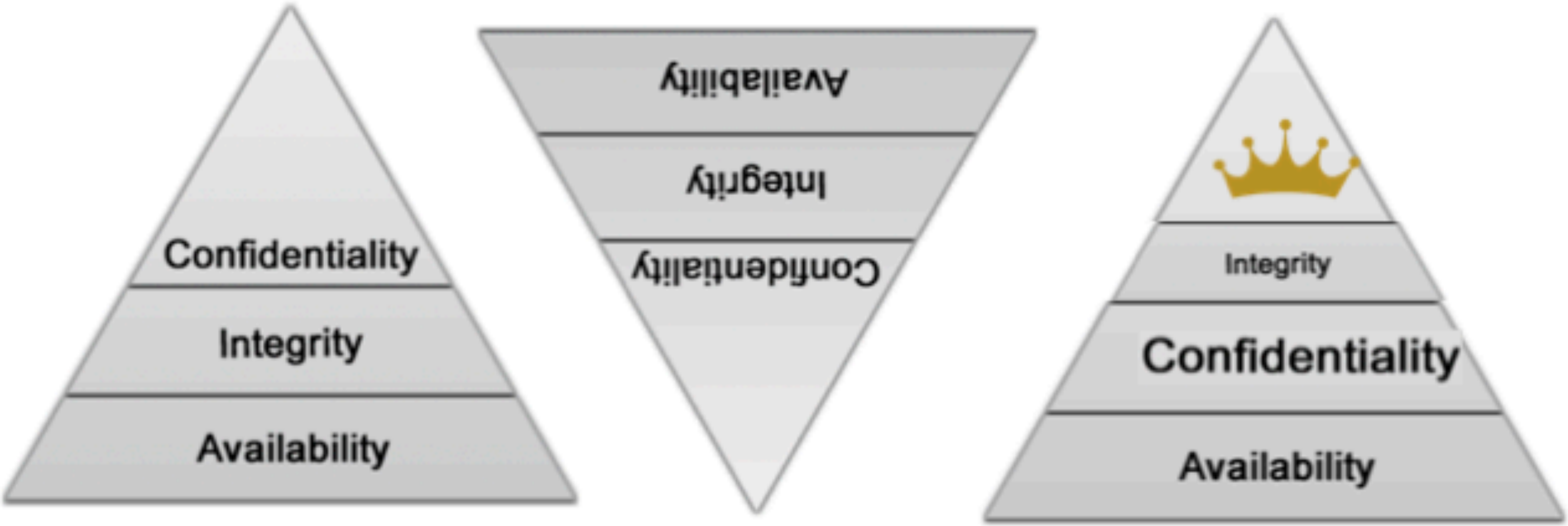


Process control

SCADA Security Basics: Integrity Trumps Availability, ISA/IEC 62443-2-1 standards (formerly ISA-99)
https://www.tofinosecurity.com/blog/scada-security-basics-integrity-trumps-availability

Marina Krotofil, Damn Vulnerable Chemical Process
https://fahrplan.events.ccc.de/congress/2014/Fahrplan/system/attachments/2560/original/31CC_2014_Krotofil.pdf

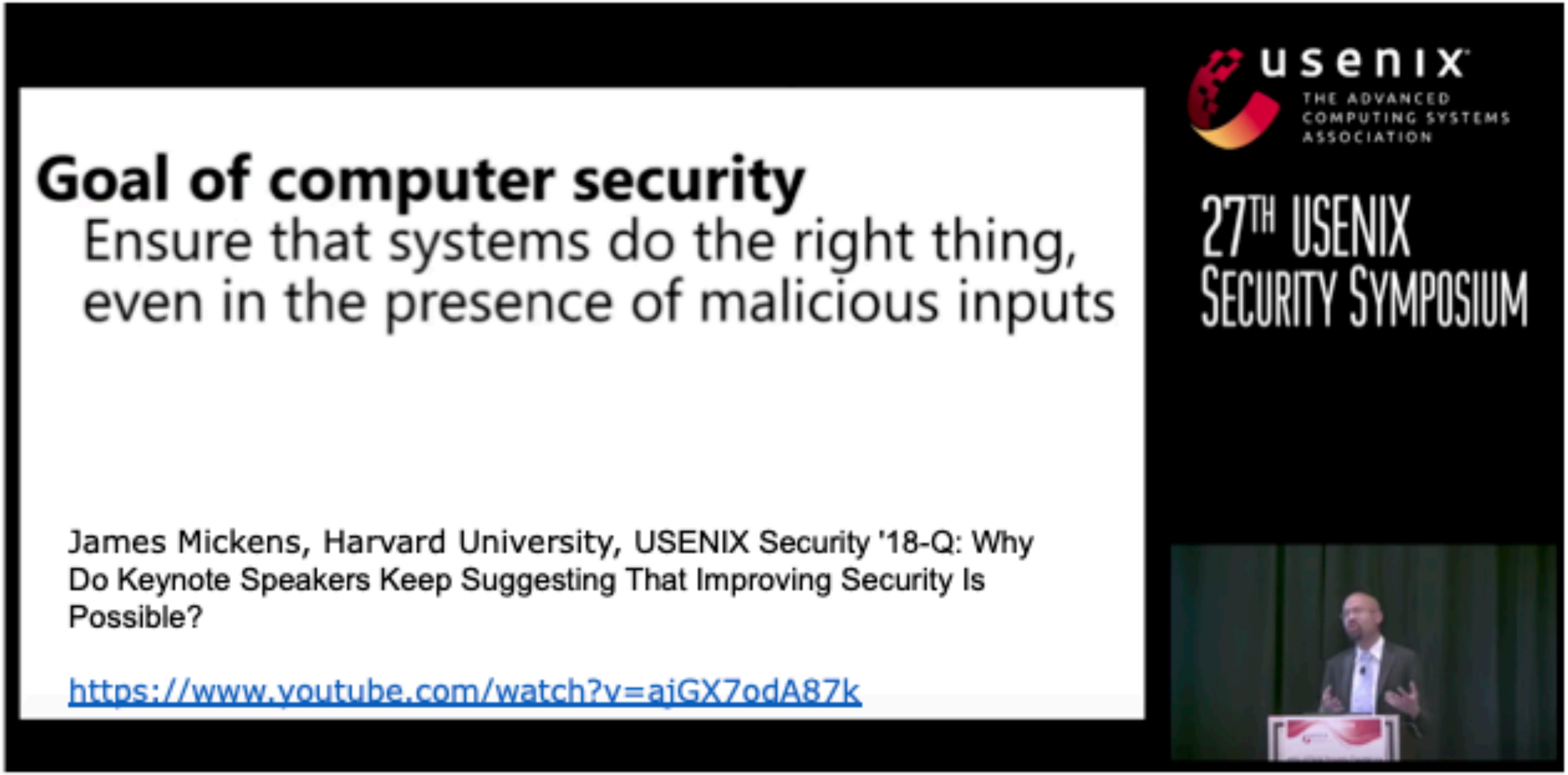
Machine Learning and AI?



IT domain

Process control

AI security



usenix
THE ADVANCED COMPUTING SYSTEMS ASSOCIATION
27TH USENIX SECURITY SYMPOSIUM

Goal of computer security

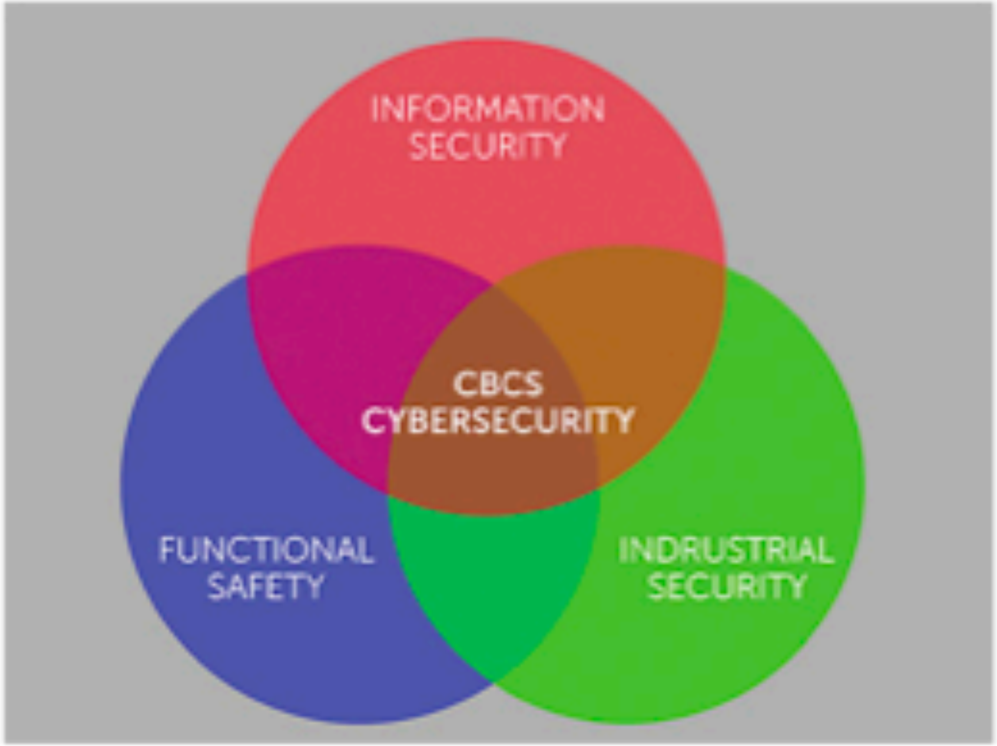
Ensure that systems do the right thing, even in the presence of malicious inputs

James Mickens, Harvard University, USENIX Security '18-Q: Why Do Keynote Speakers Keep Suggesting That Improving Security Is Possible?

https://www.youtube.com/watch?v=ajGX7odA87k

Mission-centric Cybersecurity

a process that ensures control object operation with no dangerous failures or damage, but with a set economic efficiency and reliability under adversarial anthropogenic information influence



Gapanovich, Rozenberg, Gordeychik, Signalling cyber security: the need for a mission-centric approach
https://www.railjournal.com/in_depth/signalling-cyber-security-the-need-for-a-mission-centric-approach

But what about?...

dangerous failures?

economic efficiency?

reliability level?

https://www.youtube.com/watch?v=rW9WmA5okpE

But what about?...

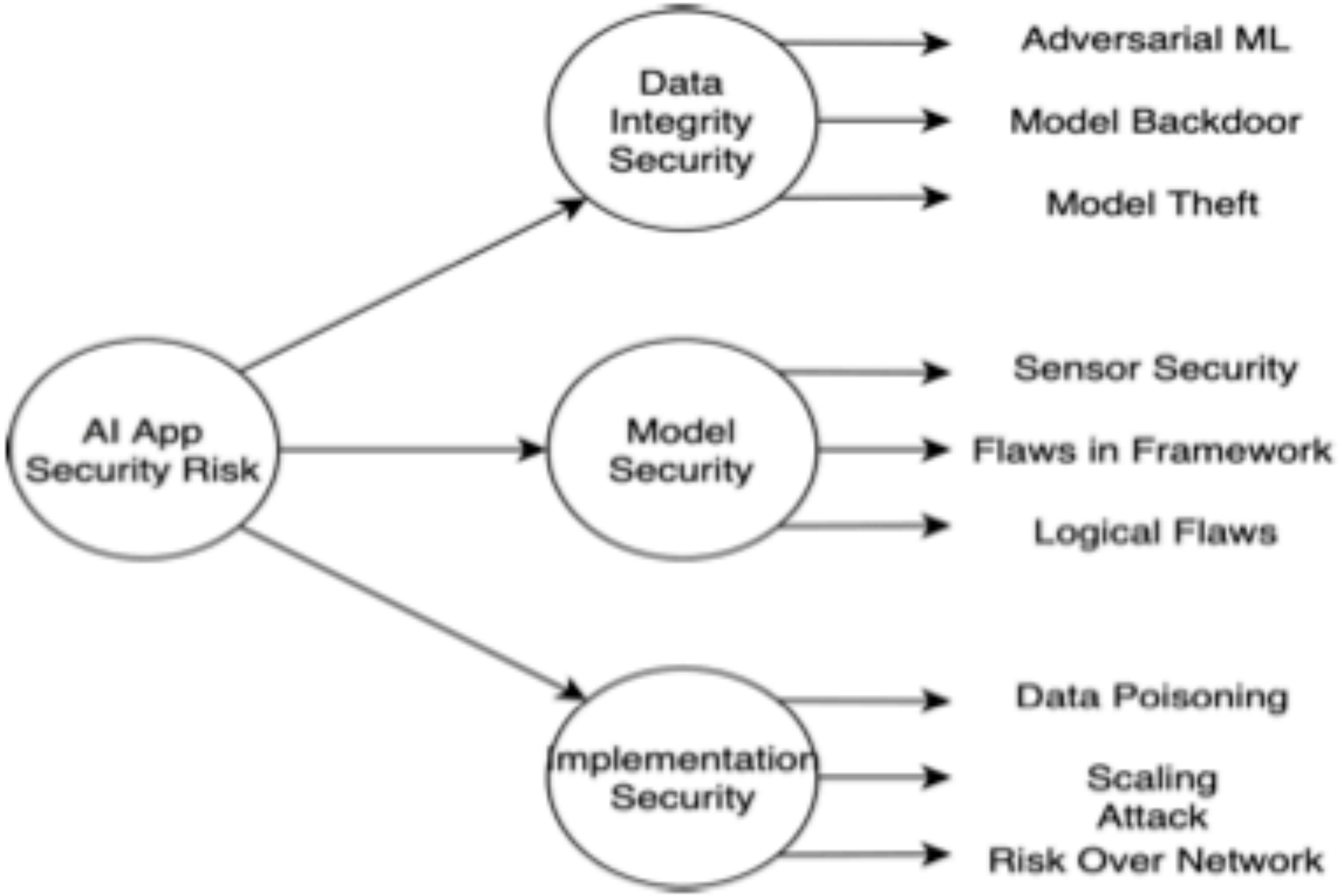
dangerous failures?

economic efficiency?

reliability level?

Build the Threat Model First!

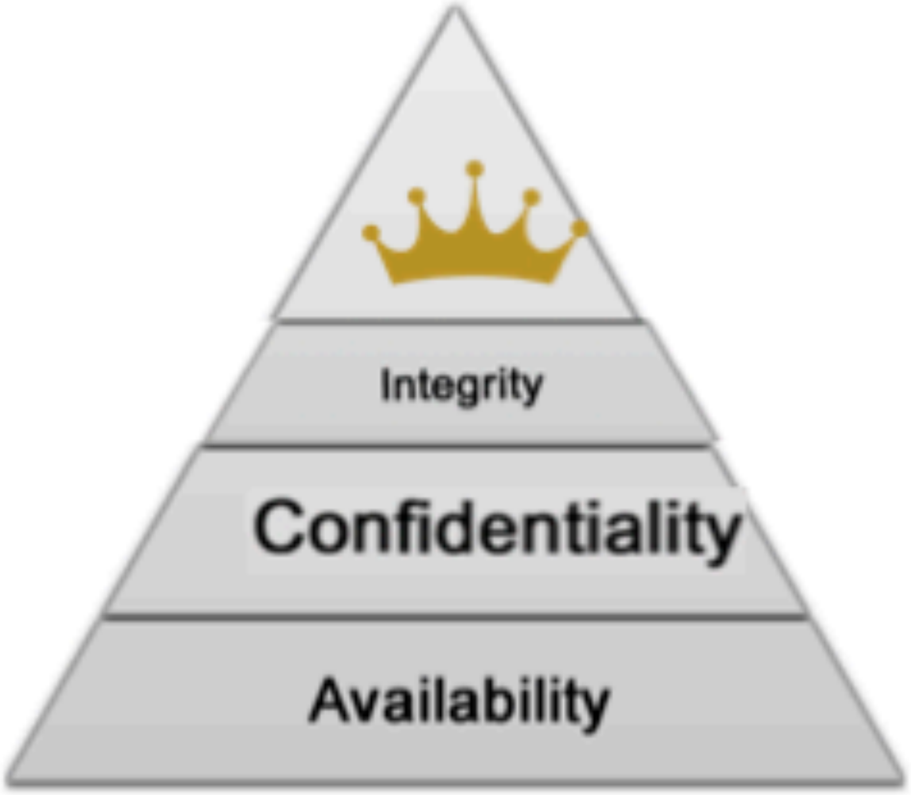
AI Threat Model



Li, K. (n.d.). *Reverse Engineering AI Models.*

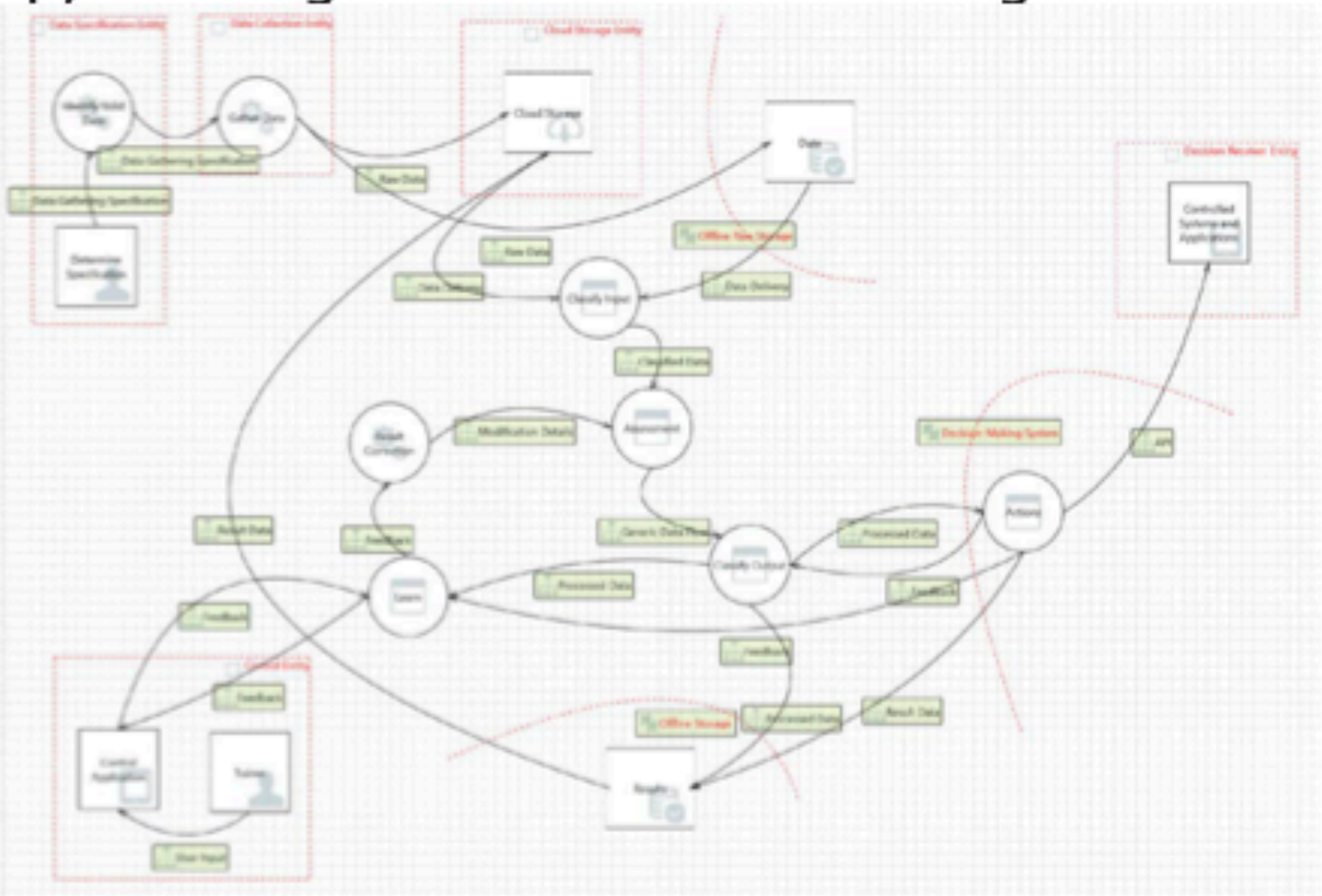
But what about?...

- Cloud
- AUC/ROC
- Privacy
- IP protection
- Federative learning
- Insane androids?...



AI security

NCC Group, Building safer machine learning



https://www.nccgroup.trust/uk/about-us/newsroom-and-events/blogs/2018/august/building-safer-machine-learning-systems-a-threat-model/

Epoch I

AI in da Cloud

Cloud - CyberSec as usual?

- InfiniBand and SDN
- Security of ML/GPU servers
  - Supply chain
  - BMC/Firmware
  - GPU is a new CPU
- Virtualization
- Containers

https://giphy.com/gifs/glas-2017-26gswyT0Qcx01b6AU

SDN/SD-WAN NEWS BYTES

- A vendor says its solution has the capability of “stitching together” WAN and Ethernet networks
- Service providers are using SD-WAN to provide network agility
- An SD-WAN router has an artificial intelligence (AI)-based routing service
- A vendor announced that it would be unifying its security and SD-WAN

How AI and Machine Learning Will Influence the SD-WAN

Artificial Intelligence & Machine Learning: SD-WAN is Evolving

by Yulia Duryea
April 2018

Machine Learning and AI Promise to Take SD-WAN Into the World of Intent

SDN/SD-WAN Security

- C. Yoon, S. Lee, H. Kang, etc. Flow Wars
- J. Hizver. Taxonomic Modeling of Security Threats in Software Defined Networking
- S. Lal, T. Taleb, A. Dutta. NFV: Security Threats and Best Practices
- SD-WAN New Hope, https://github.com/sdnewhop/sdwannewhope

SD-WAN NEW HOPE

SERGEY GORDEYCHIK
@SCADASL
WWW.SCADA.SL

DENIS KOLEGOV
@DNKOLEGOV
BI.ZONE

Taxonomic Modeling of
Security Threats in
Software Defined Networking

black hat
USA 2015

Jennia Hizver
PhD in Computer Science

black hat
USA 2018
AUGUST 4-9, 2018
MANDALAY BAY / LAS VEGAS

Too Soft[ware defined] Networks
SD WAN Vulnerability Assessment

Sergey Gordeychik
serg.gordey@gmail.com
@scadasl

Aleks Timorin
atimorin@gmail.com

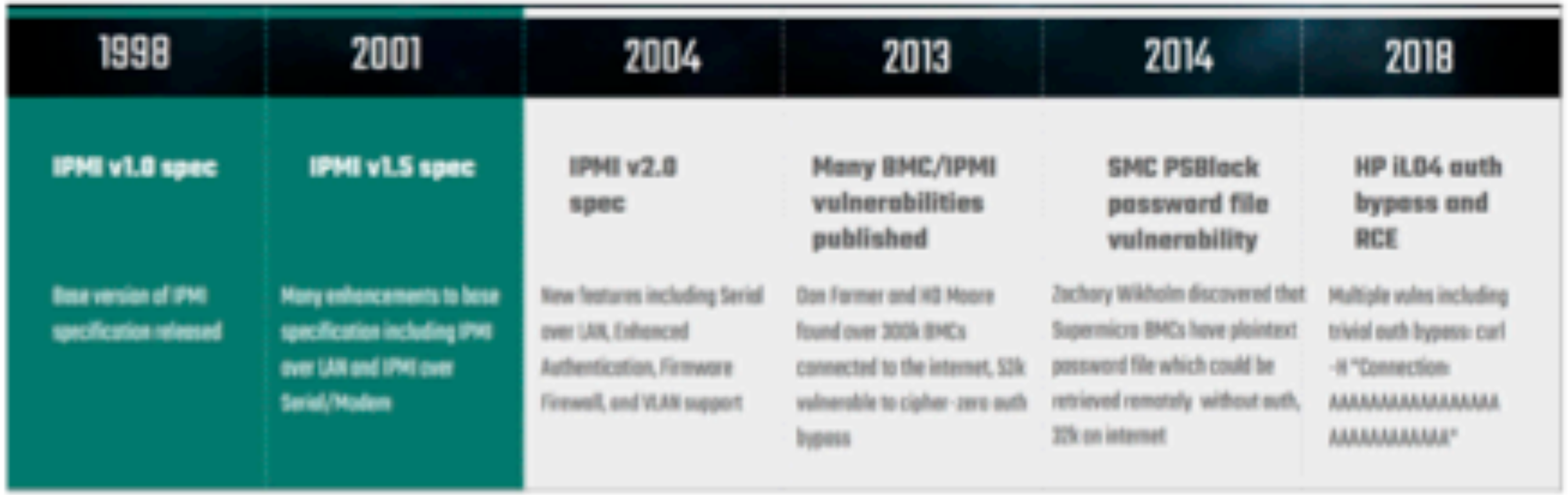
#BHUSA / @BLACKHAT EVENTS

SD-WAN New Hop - Hack before you buy!

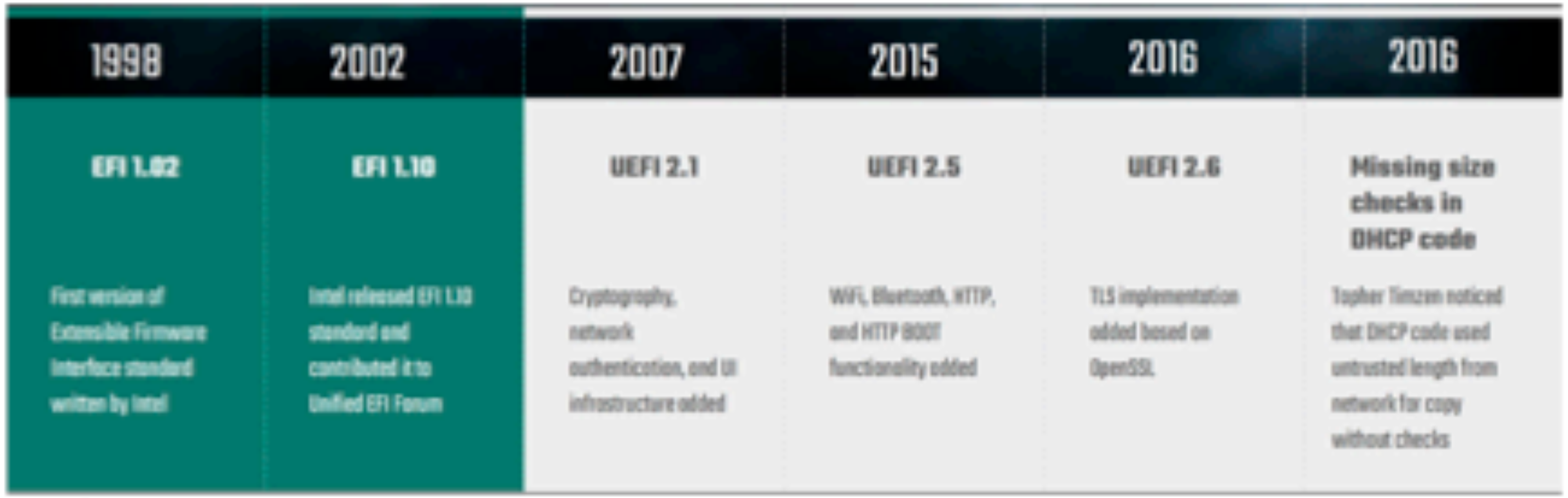
	Vendor 1	Vendor 2	Vendor 3	Vendor 4	Vendor 5
Hardcodes	V	X	X	X	V
Broken access control	V	V	X	X	V
Using vulnerable GNU/Linux	¯_(ツ)_/¯	X	X	X	¯_(ツ)_/¯
Using vulnerable 3rd party components	X	X	X	X	X
Broken client-side Web	V	X	X	X	!
Broken server-side Web	X	X	X	X	X
Secure misconfiguration	!	X	X	X	X
Memory Corruption	¯_(ツ)_/¯	¯_(ツ)_/¯	X	X	¯_(ツ)_/¯

http://www.scada.sl/search/label/sd-wan

BMC/IPMI/UEFI



1998	2001	2004	2013	2014	2018
IPMI v1.0 spec	**IPMI v1.5 spec**	**IPMI v2.0 spec**	**Many BMC/IPMI vulnerabilities published**	**SMC PSBlock password file vulnerability**	**HP iLO4 auth bypass and RCE**
Base version of IPMI specification released	Many enhancements to base specification including IPMI over LAN and IPMI over Serial/Modem	New features including Serial over LAN, Enhanced Authentication, Firmware Firewall, and VLAN support	Dan Farmer and HD Moore found over 300k BMCs connected to the internet, 53k vulnerable to cipher-zero auth bypass	Zachary Wikholm discovered that Supermicro BMCs have plaintext password file which could be retrieved remotely without auth, 32k on internet	Multiple vulns including trivial auth bypass: curl -H "Connection: AAAAAAAAAAAAAAAAAAAAAAAAAAA"



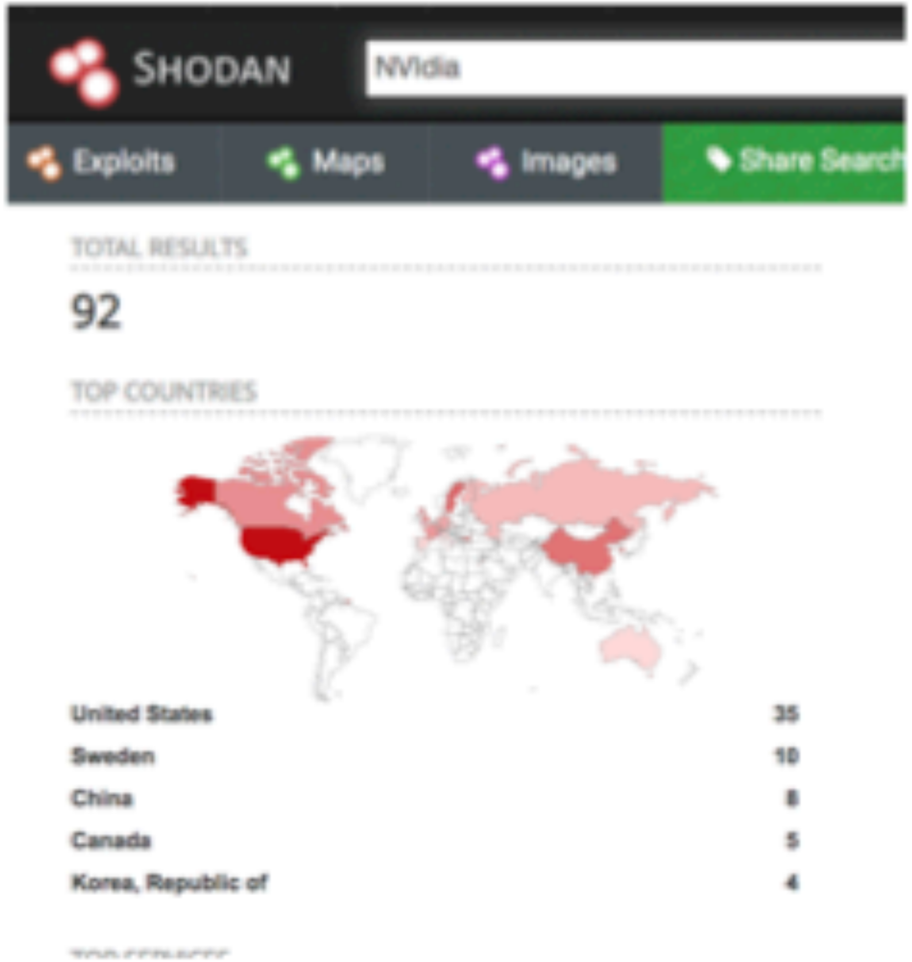
1998	2002	2007	2015	2016	2016
EFI 1.02	**EFI 1.10**	**UEFI 2.1**	**UEFI 2.5**	**UEFI 2.6**	**Missing size checks in DHCP code**
First version of Extensible Firmware Interface standard written by Intel	Intel released EFI 1.10 standard and contributed it to Unified EFI Forum	Cryptography, network authentication, and UI infrastructure added	WiFi, Bluetooth, HTTP, and HTTP BOOT functionality added	TLS implementation added based on OpenSSL	Topher Timzen noticed that DHCP code used untrusted length from network for copy without checks

- Remotely Attacking System Firmware, Jesse Michael, Mickey Shkatov & Oleksandr Bazhaniuk, BH18

ML in da Cloud?

To find a ML Server in the Internet?

GPGPU?

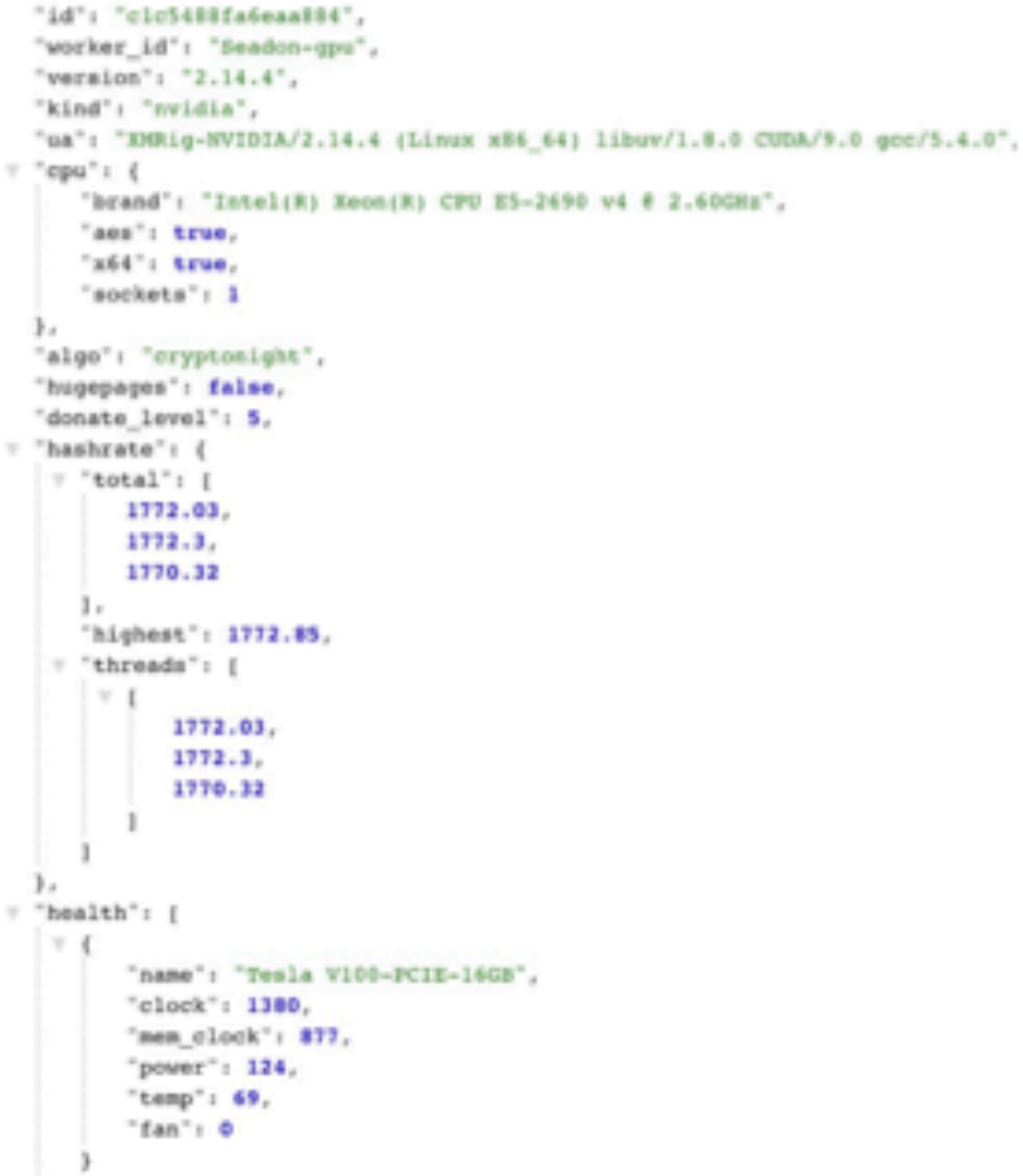


TOTAL RESULTS

92

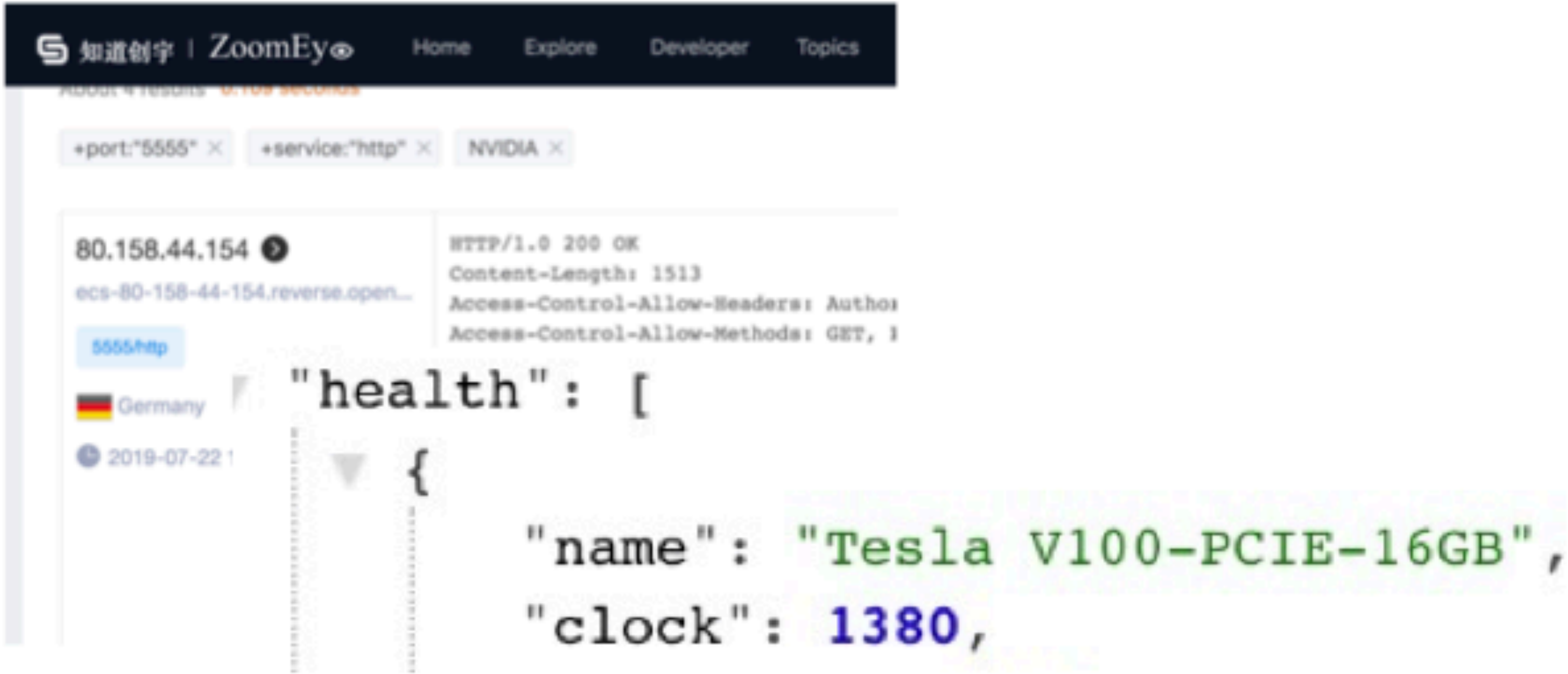
TOP COUNTRIES

United States	35
Sweden	10
China	8
Canada	5
Korea, Republic of	4



```
 "id": "c1c5488fa6eaa884",
 "worker_id": "Seadon-gpu",
 "version": "2.14.4",
 "kind": "nvidia",
 "ua": "XMRig-NVIDIA/2.14.4 (Linux x86_64) libuv/1.8.0 CUDA/9.0 gcc/5.4.0",
▼"cpu": {
     "brand": "Intel(R) Xeon(R) CPU E5-2690 v4 @ 2.60GHz",
     "aes": true,
     "x64": true,
     "sockets": 1
 },
 "algo": "cryptonight",
 "hugepages": false,
 "donate_level": 5,
▼"hashrate": {
   ▼"total": [
         1772.03,
         1772.3,
         1770.32
     ],
     "highest": 1772.85,
   ▼"threads": [
       ▼[
             1772.03,
             1772.3,
             1770.32
         ]
     ]
 },
▼"health": [
   ▼{
         "name": "Tesla V100-PCIE-16GB",
         "clock": 1380,
         "mem_clock": 877,
         "power": 124,
         "temp": 69,
         "fan": 0
     }
```

Crypto currency on GPGPU in 2019?



80.158.44.154

ecs-80-158-44-154.reverse.open...

5555/http

Germany

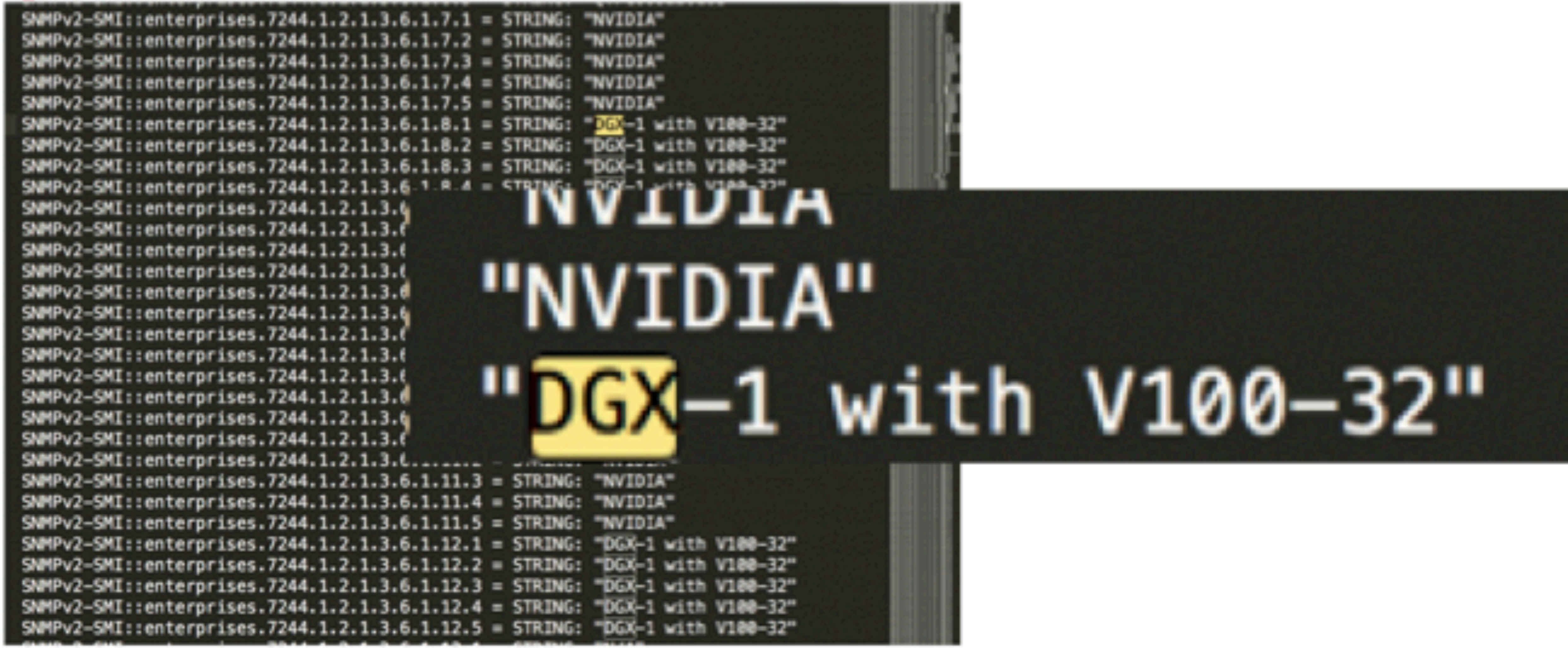
2019-07-22

```
HTTP/1.0 200 OK
Content-Length: 1513
Access-Control-Allow-Headers: Autho
Access-Control-Allow-Methods: GET, 
```

```
"health": [
  {
    "name": "Tesla V100-PCIE-16GB",
    "clock": 1380,
```

https://www.zoomeye.org/searchResult?q=%2Bport%3A%225555%22%20%2Bservice%3A%22http%22%20NVIDIA

SNMPWALK

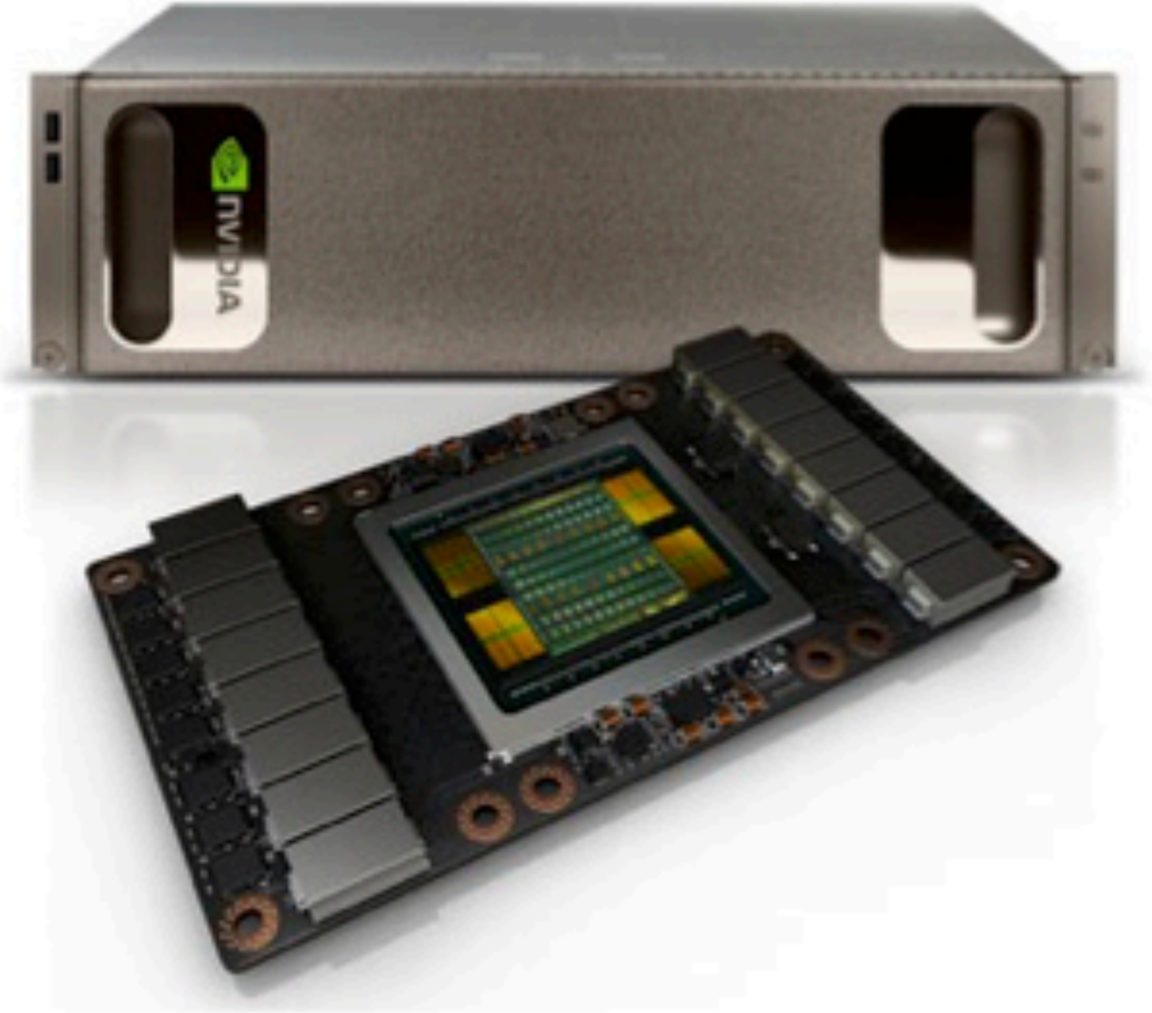


```
SNMPv2-SMI::enterprises.7244.1.2.1.3.6.1.7.1 = STRING: "NVIDIA"
SNMPv2-SMI::enterprises.7244.1.2.1.3.6.1.7.2 = STRING: "NVIDIA"
SNMPv2-SMI::enterprises.7244.1.2.1.3.6.1.7.3 = STRING: "NVIDIA"
SNMPv2-SMI::enterprises.7244.1.2.1.3.6.1.7.4 = STRING: "NVIDIA"
SNMPv2-SMI::enterprises.7244.1.2.1.3.6.1.7.5 = STRING: "NVIDIA"
SNMPv2-SMI::enterprises.7244.1.2.1.3.6.1.8.1 = STRING: "DGX-1 with V100-32"
SNMPv2-SMI::enterprises.7244.1.2.1.3.6.1.8.2 = STRING: "DGX-1 with V100-32"
SNMPv2-SMI::enterprises.7244.1.2.1.3.6.1.8.3 = STRING: "DGX-1 with V100-32"
SNMPv2-SMI::enterprises.7244.1.2.1.3.6.1.8.4 = STRING: "DGX-1 with V100-32"
SNMPv2-SMI::enterprises.7244.1.2.1.3.6
SNMPv2-SMI::enterprises.7244.1.2.1.3.6
SNMPv2-SMI::enterprises.7244.1.2.1.3.6
SNMPv2-SMI::enterprises.7244.1.2.1.3.6
SNMPv2-SMI::enterprises.7244.1.2.1.3.6
SNMPv2-SMI::enterprises.7244.1.2.1.3.6
SNMPv2-SMI::enterprises.7244.1.2.1.3.6
SNMPv2-SMI::enterprises.7244.1.2.1.3.6
SNMPv2-SMI::enterprises.7244.1.2.1.3.6
SNMPv2-SMI::enterprises.7244.1.2.1.3.6
SNMPv2-SMI::enterprises.7244.1.2.1.3.6
SNMPv2-SMI::enterprises.7244.1.2.1.3.6
SNMPv2-SMI::enterprises.7244.1.2.1.3.6
SNMPv2-SMI::enterprises.7244.1.2.1.3.6.1.11.3 = STRING: "NVIDIA"
SNMPv2-SMI::enterprises.7244.1.2.1.3.6.1.11.4 = STRING: "NVIDIA"
SNMPv2-SMI::enterprises.7244.1.2.1.3.6.1.11.5 = STRING: "NVIDIA"
SNMPv2-SMI::enterprises.7244.1.2.1.3.6.1.12.1 = STRING: "DGX-1 with V100-32"
SNMPv2-SMI::enterprises.7244.1.2.1.3.6.1.12.2 = STRING: "DGX-1 with V100-32"
SNMPv2-SMI::enterprises.7244.1.2.1.3.6.1.12.3 = STRING: "DGX-1 with V100-32"
SNMPv2-SMI::enterprises.7244.1.2.1.3.6.1.12.4 = STRING: "DGX-1 with V100-32"
SNMPv2-SMI::enterprises.7244.1.2.1.3.6.1.12.5 = STRING: "DGX-1 with V100-32"
```

```
"NVIDIA"
"NVIDIA"
"DGX-1 with V100-32"
```

DGX-1

- 8 Tesla V100-32GB
- TFLOPS (deep learning) 1000
- CUDA Cores 40,960
- Tensor Cores 5,120
- $130,000

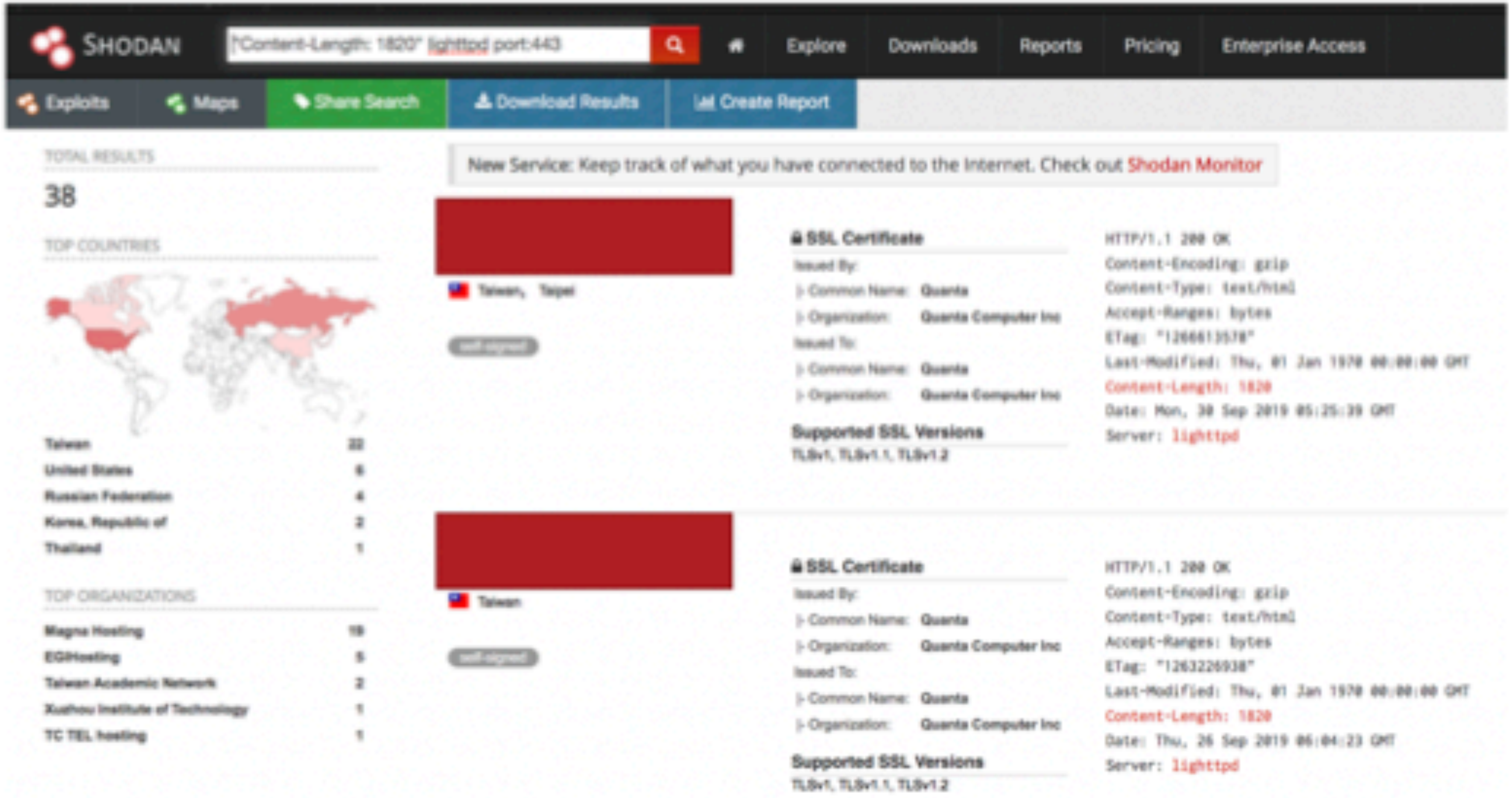


- Good hashcat rate :)

https://hashcat.net/forum/thread-6972.html

NetNTLMv2: 28912.2 MH/s
MD5: 450.0 GH/s
SHA-256: 59971.8 MH/s
MS Office 2013: 163.5 kH/s
bcrypt $2*$, Blowfish (Unix): 434.2 kH/s

Other things?



"Content-Length: 1820" lighttpd port:443

New Service: Keep track of what you have connected to the Internet. Check out Shodan Monitor

TOTAL RESULTS

38

TOP COUNTRIES

Taiwan	22
United States	6
Russian Federation	4
Korea, Republic of	2
Thailand	1

TOP ORGANIZATIONS

Magna Hosting	19
EGIHosting	5
Taiwan Academic Network	2
Xuzhou Institute of Technology	1
TC TEL hosting	1

[redacted]

Taiwan, Taipei

self-signed

SSL Certificate

Issued By:
|- Common Name: Quanta
|- Organization: Quanta Computer Inc
Issued To:
|- Common Name: Quanta
|- Organization: Quanta Computer Inc

Supported SSL Versions

TLSv1, TLSv1.1, TLSv1.2

```
HTTP/1.1 200 OK
Content-Encoding: gzip
Content-Type: text/html
Accept-Ranges: bytes
ETag: "1266613578"
Last-Modified: Thu, 01 Jan 1970 00:00:00 GMT
Content-Length: 1820
Date: Mon, 30 Sep 2019 05:25:39 GMT
Server: lighttpd
```

[redacted]

Taiwan

self-signed

SSL Certificate

Issued By:
|- Common Name: Quanta
|- Organization: Quanta Computer Inc
Issued To:
|- Common Name: Quanta
|- Organization: Quanta Computer Inc

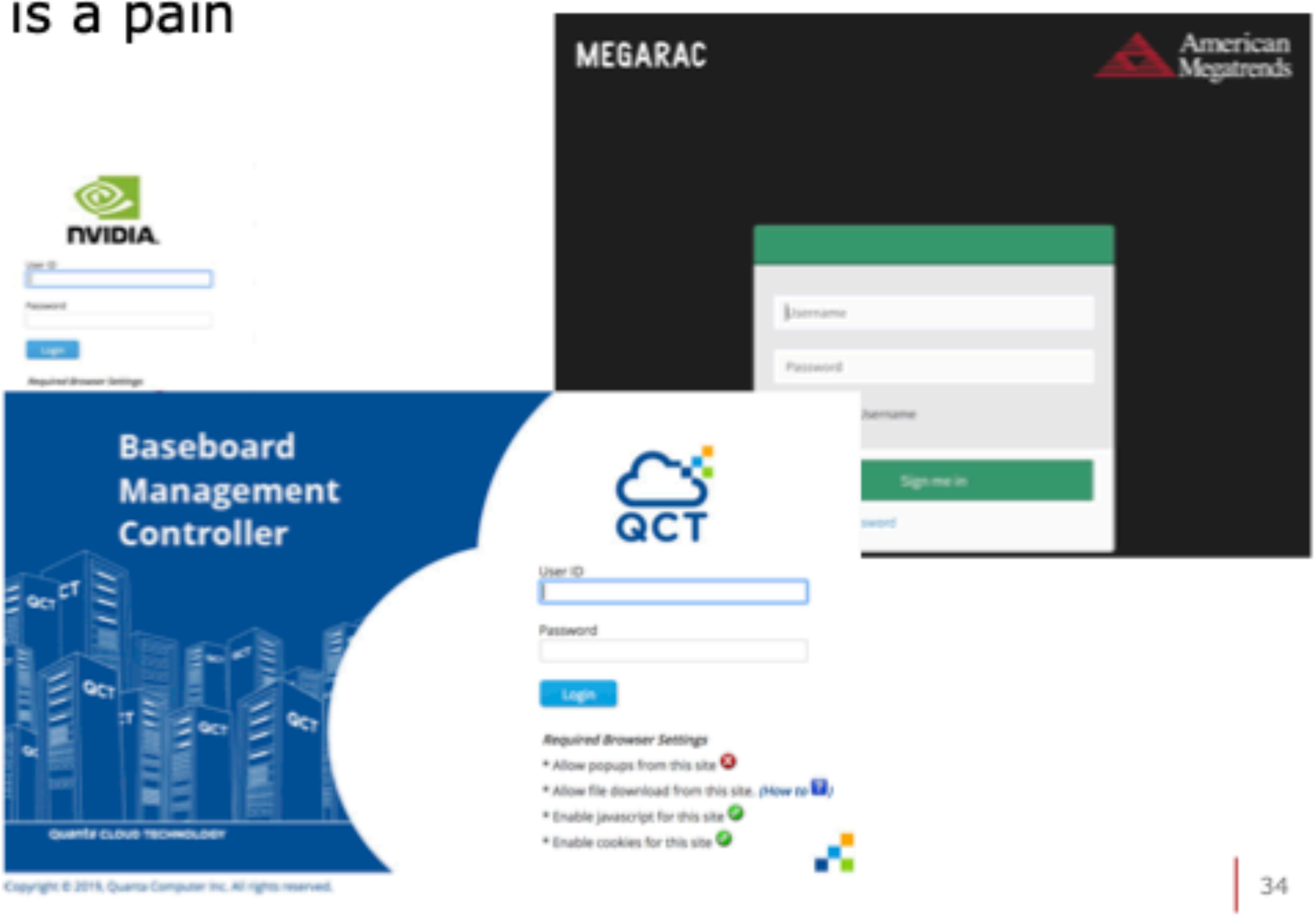
Supported SSL Versions

TLSv1, TLSv1.1, TLSv1.2

```
HTTP/1.1 200 OK
Content-Encoding: gzip
Content-Type: text/html
Accept-Ranges: bytes
ETag: "1263226938"
Last-Modified: Thu, 01 Jan 1970 00:00:00 GMT
Content-Length: 1820
Date: Thu, 26 Sep 2019 06:04:23 GMT
Server: lighttpd
```

Supply chain is a pain

Baseboard
Management
Controller



Required Browser Settings

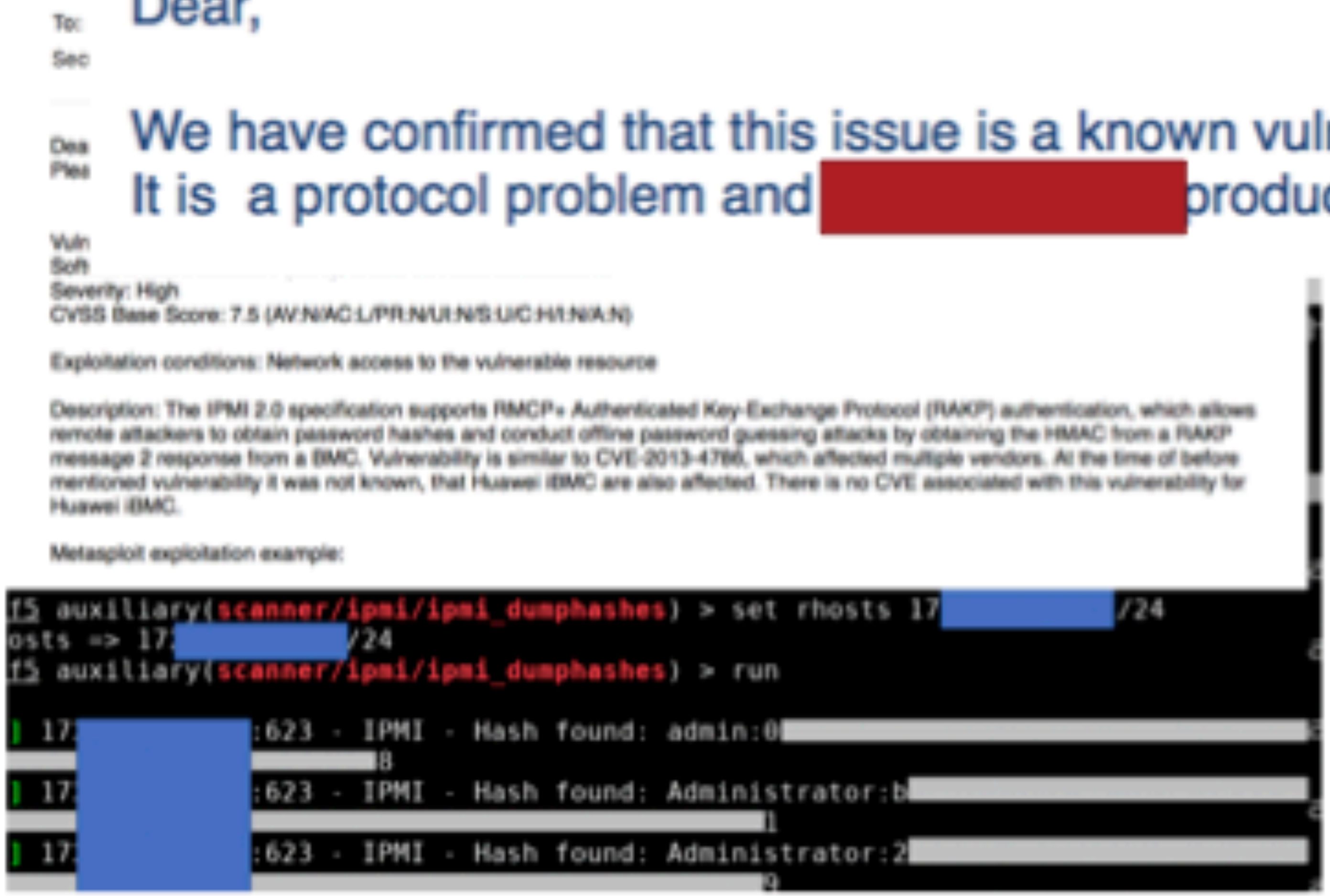
- Allow popups from this site ✗
- Allow file download from this site. *(How to ?)*
- Enable javascript for this site ✓
- Enable cookies for this site ✓

Copyright © 2019, Quanta Computer Inc. All rights reserved.

CVE-2013-4786 - 2019

Dear,

We have confirmed that this issue is a known vulnerability (CVE-2013-4786).
It is a protocol problem and [redacted] products also comply with this standard.



To:
Sec

Dea
Ples

Vuln
Soft
Severity: High
CVSS Base Score: 7.5 (AV:N/AC:L/PR:N/UI:N/S:U/C:H/I:N/A:N)

Exploitation conditions: Network access to the vulnerable resource

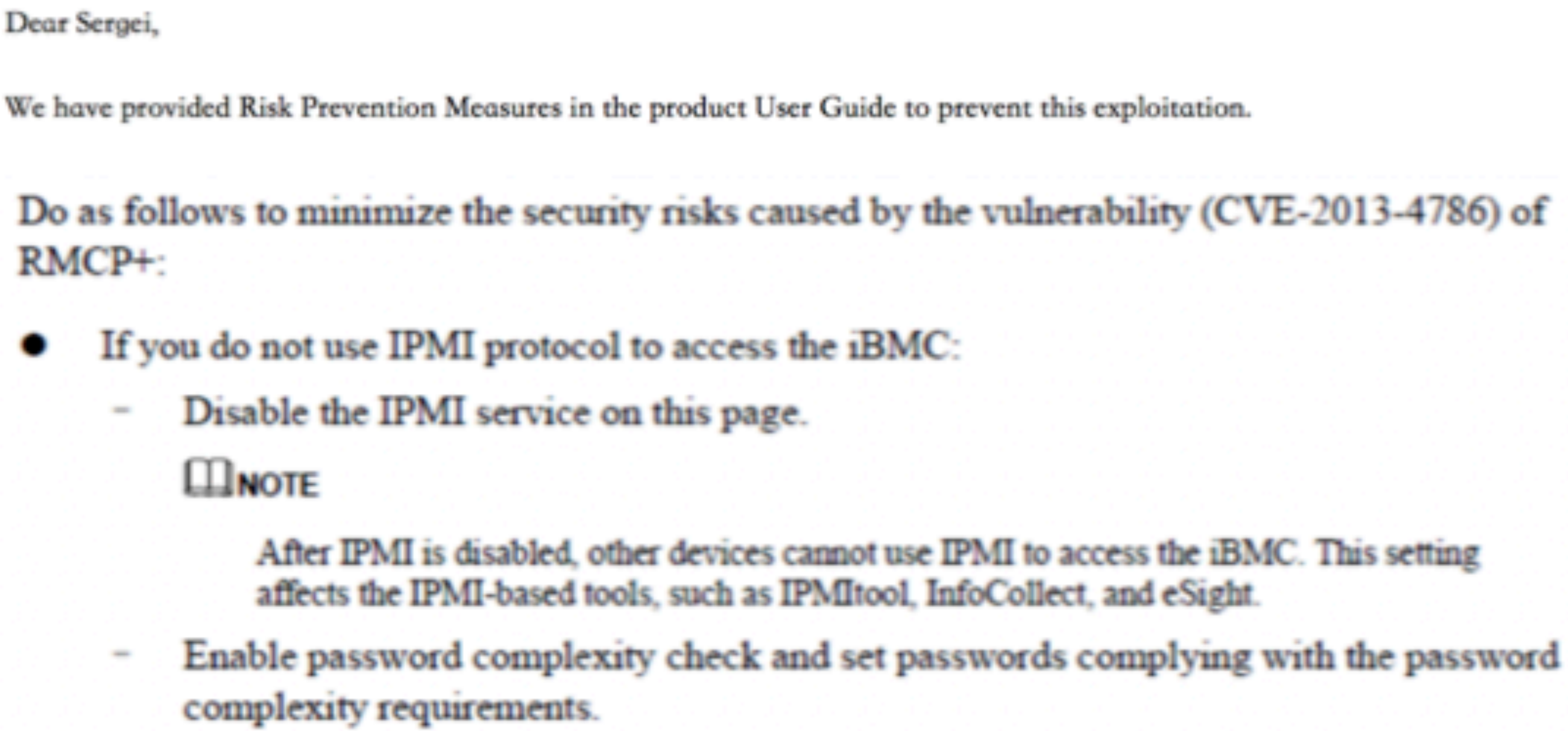
Description: The IPMI 2.0 specification supports RMCP+ Authenticated Key-Exchange Protocol (RAKP) authentication, which allows remote attackers to obtain password hashes and conduct offline password guessing attacks by obtaining the HMAC from a RAKP message 2 response from a BMC. Vulnerability is similar to CVE-2013-4786, which affected multiple vendors. At the time of before mentioned vulnerability it was not known, that Huawei iBMC are also affected. There is no CVE associated with this vulnerability for Huawei iBMC.

Metasploit exploitation example:

```
f5 auxiliary(scanner/ipmi/ipmi_dumphashes) > set rhosts 17[redacted]/24
osts => 17[redacted]/24
f5 auxiliary(scanner/ipmi/ipmi_dumphashes) > run

] 17[redacted]:623 - IPMI - Hash found: admin:0[redacted]
[redacted]8
] 17    :623 - IPMI - Hash found: Administrator:b[redacted]
[redacted]1
] 17    :623 - IPMI - Hash found: Administrator:2[redacted]
[redacted]9
```

Use c0mp13x passwords!



Dear Sergei,

We have provided Risk Prevention Measures in the product User Guide to prevent this exploitation.

Do as follows to minimize the security risks caused by the vulnerability (CVE-2013-4786) of RMCP+:

- If you do not use IPMI protocol to access the iBMC:
  - Disable the IPMI service on this page.

    NOTE

    After IPMI is disabled, other devices cannot use IPMI to access the iBMC. This setting affects the IPMI-based tools, such as IPMItool, InfoCollect, and eSight.
  - Enable password complexity check and set passwords complying with the password complexity requirements.

I have only one question!

~~How the complex password will help?!!~~

Why it still enabled by default in 2019?

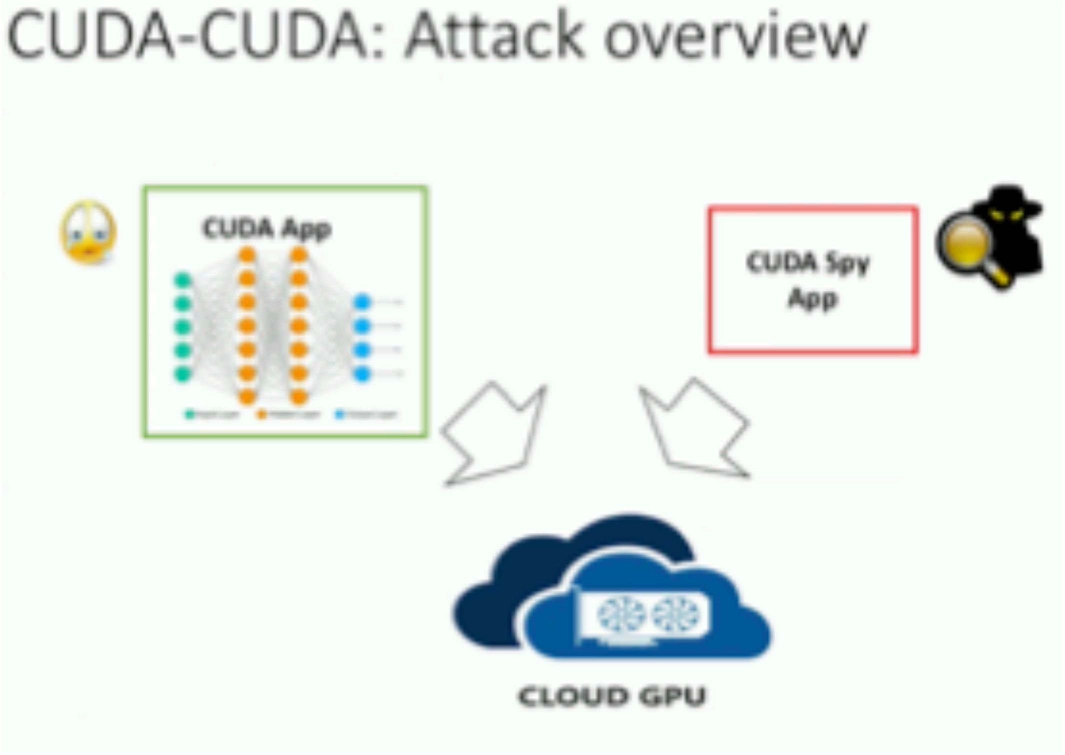
http://www.demotivation.us/i-have-only-one-question-1267735.html

Any bugs there?

We don’t know yet

GPGPU is a new CPU

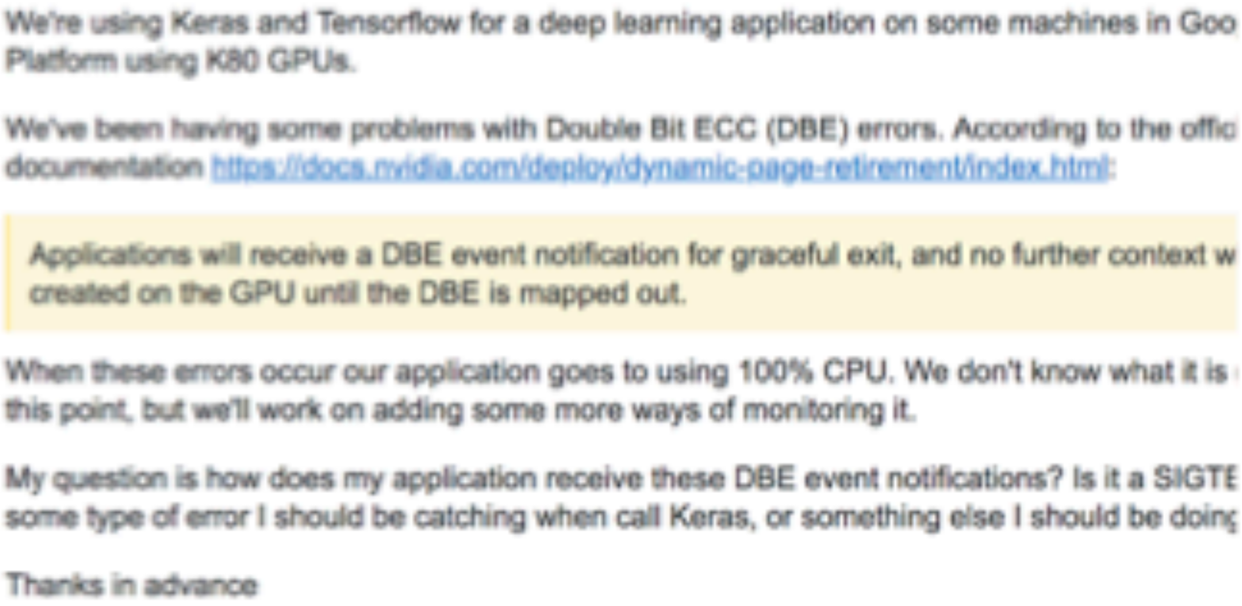
- GPU drivers vulns
  - 10x for Windows, few for Linux
  - CVE-2018-6249
  - CVE-2018-6253
- GPU rootkit
  - Avoid detection
  - DMA (keylogger, passwords)
  - Project Maux Mk.II (2008)
  - Jellyfish PoC rootkit (2015)
- GPU – specific vulnerabilities????



CUDA-CUDA: Attack overview

Rendered Insecure
GPU Side Channel Attacks are Practical

Rowhammer anyone?



We're using Keras and Tensorflow for a deep learning application on some machines in Goo Platform using K80 GPUs.

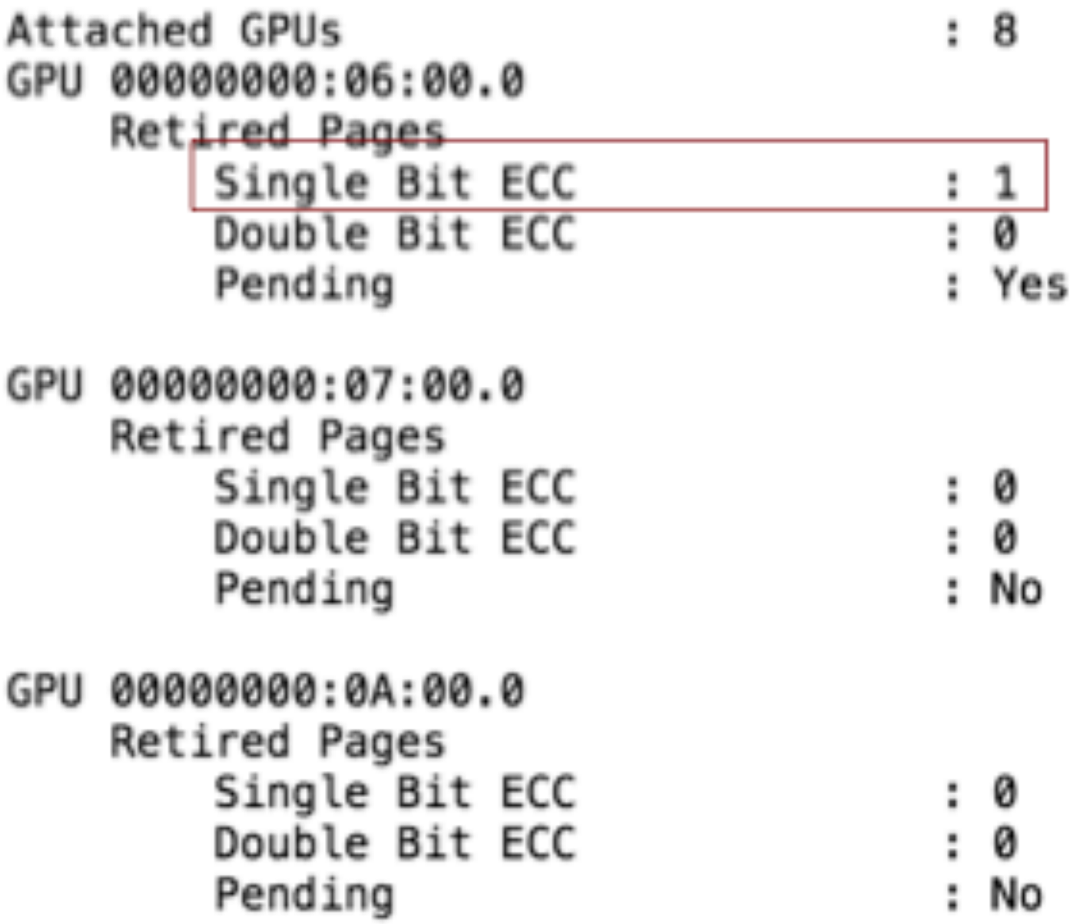
We've been having some problems with Double Bit ECC (DBE) errors. According to the offic documentation https://docs.nvidia.com/deploy/dynamic-page-retirement/index.html:

> Applications will receive a DBE event notification for graceful exit, and no further context w created on the GPU until the DBE is mapped out.

When these errors occur our application goes to using 100% CPU. We don't know what it is this point, but we'll work on adding some more ways of monitoring it.

My question is how does my application receive these DBE event notifications? Is it a SIGTE some type of error I should be catching when call Keras, or something else I should be doin

Thanks in advance



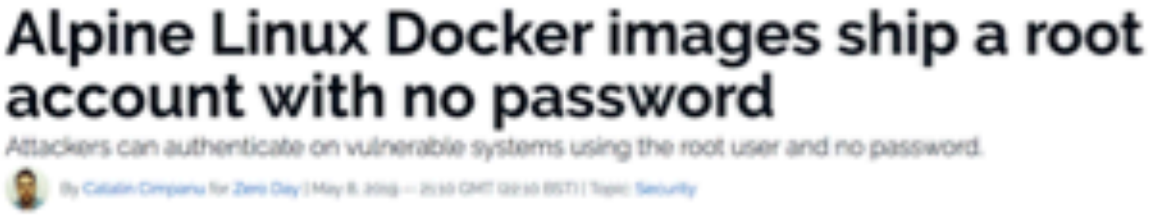
```
Attached GPUs                           : 8
GPU 00000000:06:00.0
    Retired Pages
        Single Bit ECC                  : 1
        Double Bit ECC                  : 0
        Pending                         : Yes

GPU 00000000:07:00.0
    Retired Pages
        Single Bit ECC                  : 0
        Double Bit ECC                  : 0
        Pending                         : No

GPU 00000000:0A:00.0
    Retired Pages
        Single Bit ECC                  : 0
        Double Bit ECC                  : 0
        Pending                         : No
```

Docker

- Host security
  - Hardening
  - Docker daemon
  - (CVE-2018-15664, CVE-2018-8115, etc)
- Container Images
  - Patch management
  - Configuration (CVE-2019-5021)
  - Information leakage
  - Trust
- Root access
  - Running containers as Root
  - Processes as Root
  - CAP_SYS_ADMIN privilege
- Limit Compute Resources

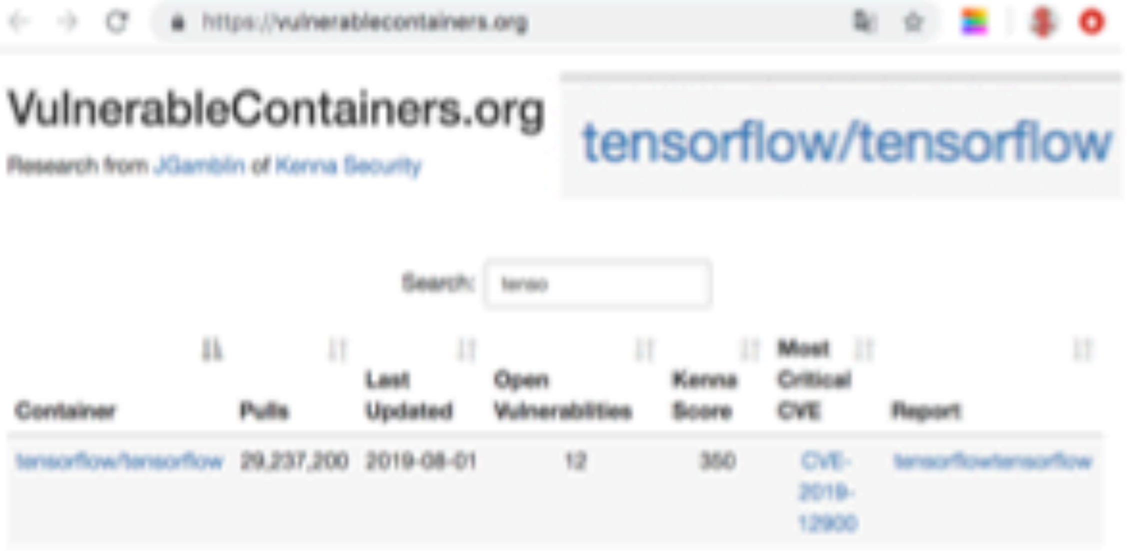


Alpine Linux Docker images ship a root account with no password

Attackers can authenticate on vulnerable systems using the root user and no password.

By Catalin Cimpanu for Zero Day | May 8, 2019 -- 21:10 GMT (22:10 BST) | Topic: Security

The issue was first discovered back in August 2015, patched in November, then accidentally re-opened three weeks later, in December 2015, only to be re-discovered again by a Cisco Umbrella researcher in January this year.



https://vulnerablecontainers.org

VulnerableContainers.org

Research from JGamblin of Kenna Security

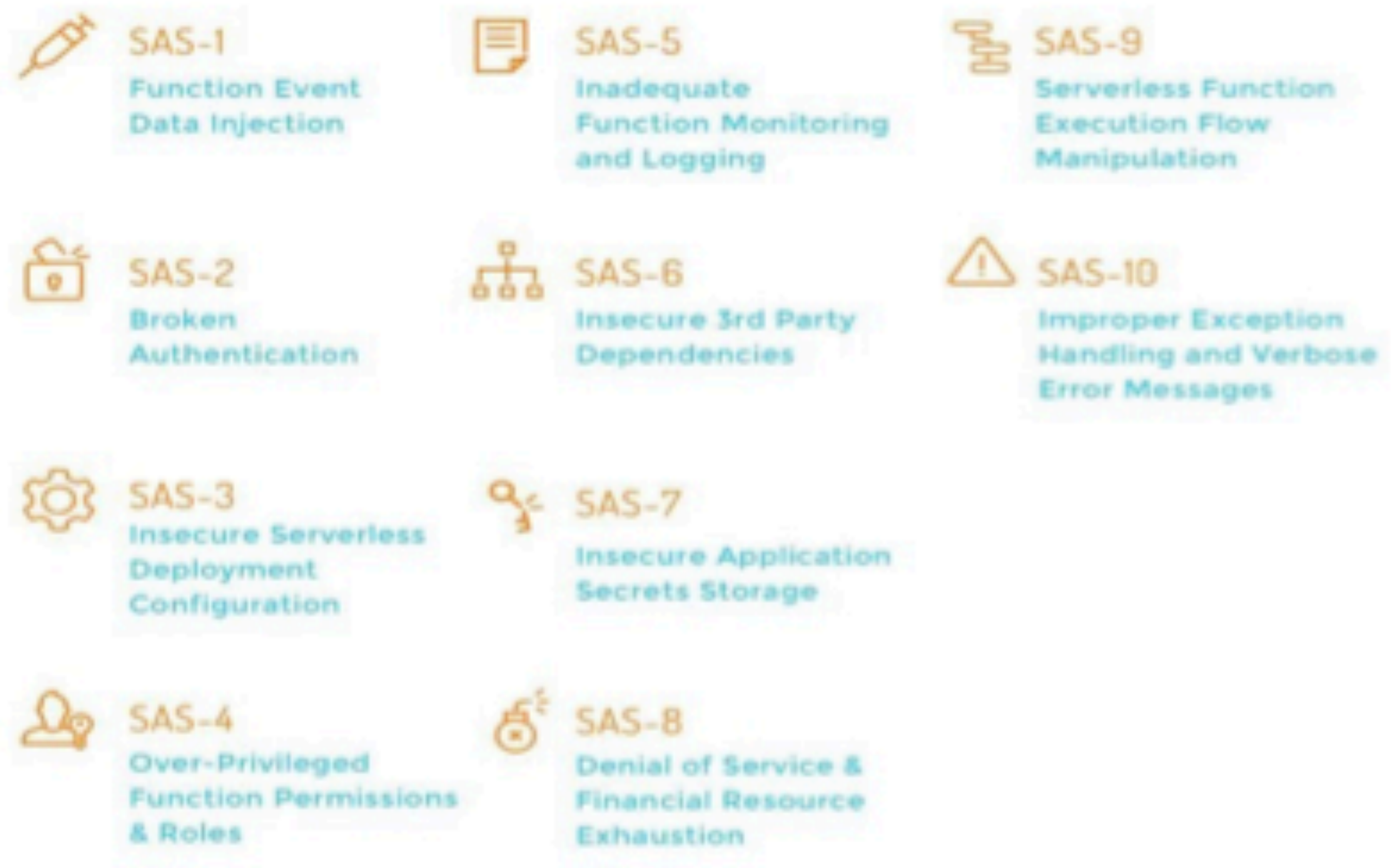
tensorflow/tensorflow

Search: tenso

Container	Pulls	Last Updated	Open Vulnerabilities	Kenna Score	Most Critical CVE	Report
tensorflow/tensorflow	29,237,200	2019-08-01	12	350	CVE-2019-12900	tensorflowtensorflow

https://vulnerablecontainers.org/

Serverless Security



SAS-1
Function Event Data Injection

SAS-2
Broken Authentication

SAS-3
Insecure Serverless Deployment Configuration

SAS-4
Over-Privileged Function Permissions & Roles

SAS-5
Inadequate Function Monitoring and Logging

SAS-6
Insecure 3rd Party Dependencies

SAS-7
Insecure Application Secrets Storage

SAS-8
Denial of Service & Financial Resource Exhaustion

SAS-9
Serverless Function Execution Flow Manipulation

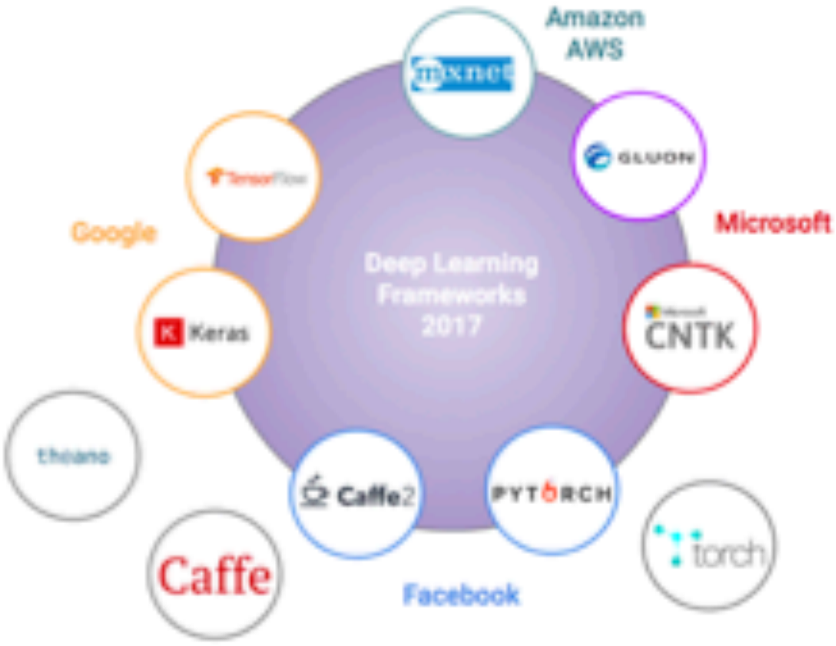
SAS-10
Improper Exception Handling and Verbose Error Messages

https://www.puresec.io/resource-download

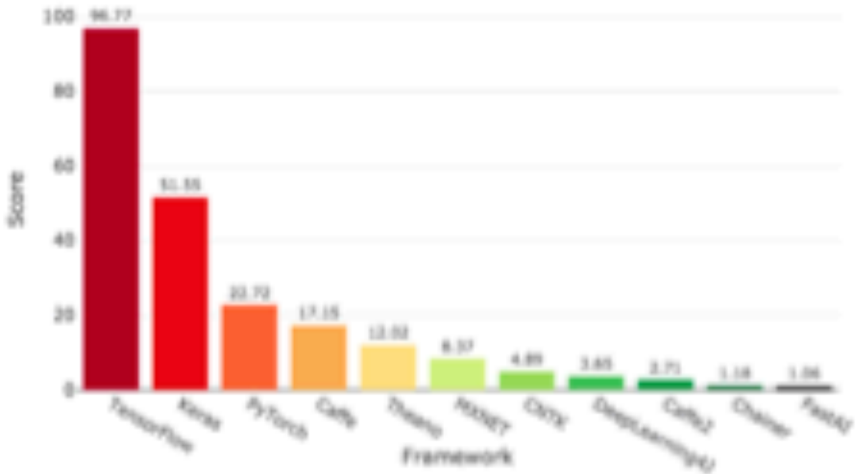
ML/DL Frameworks

- Vulnerabilities in frameworks
  - Management interfaces
  - Data processing
  - Integration
  - Patch management
- Code security
  - Custom code
  - Model as malware

https://towardsdatascience.com/deep-learning-framework-power-scores-2018-23607ddf297a



Deep Learning Framework Power Scores 2018



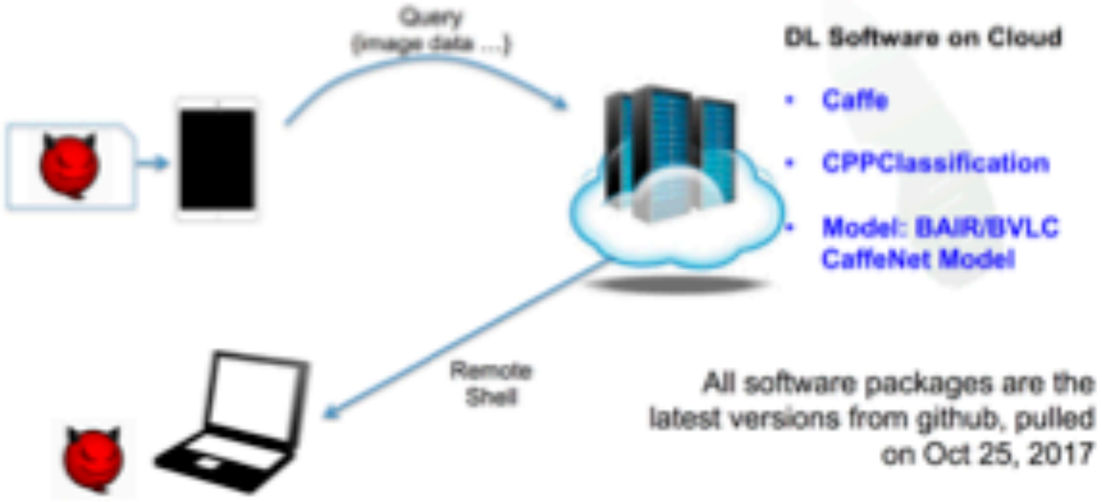
Data processing

- 3rd party packages dependencies
- Obsolete code
- Data handling vulnerabilities

- Example
  - Remote code execution in Caffe via crafted image

DL Framework	Lines of Code	Number of Dep. Packages
Caffe	127K+	137
TensorFlow	887K+	97
torch	590K+	48

Demo Setup



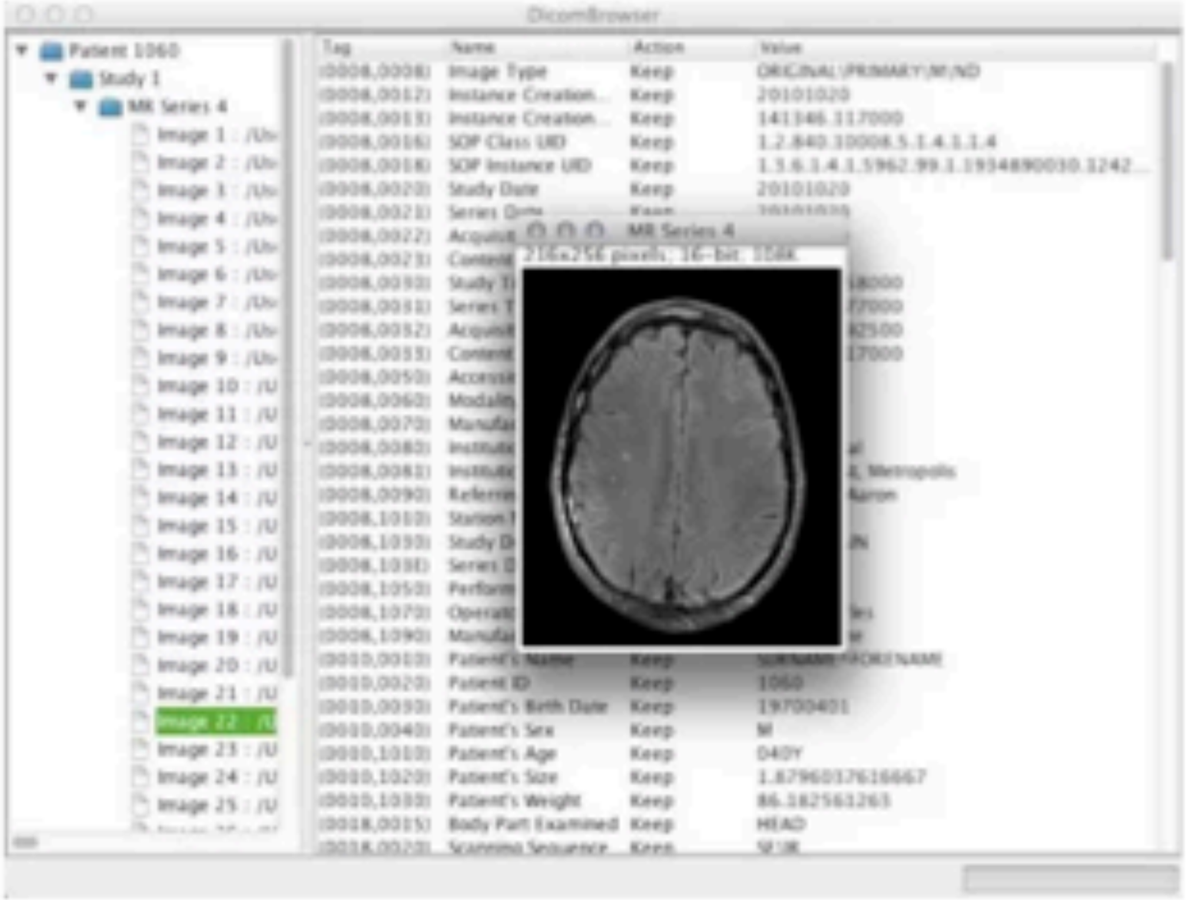
DL Software on Cloud

- Caffe
- CPPClassification
- Model: BAIR/BVLC CaffeNet Model

All software packages are the latest versions from github, pulled on Oct 25, 2017

From framework to Pipeline

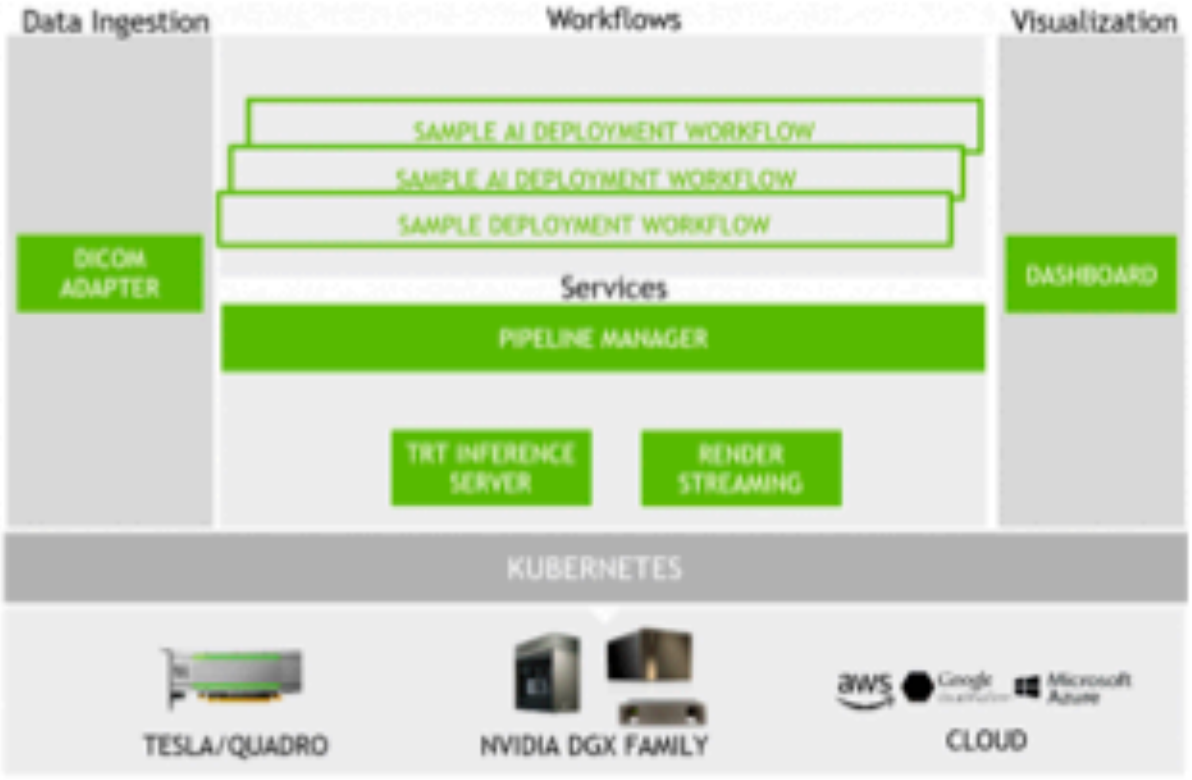
DICOM
Digital Imaging and Communications in Medicine



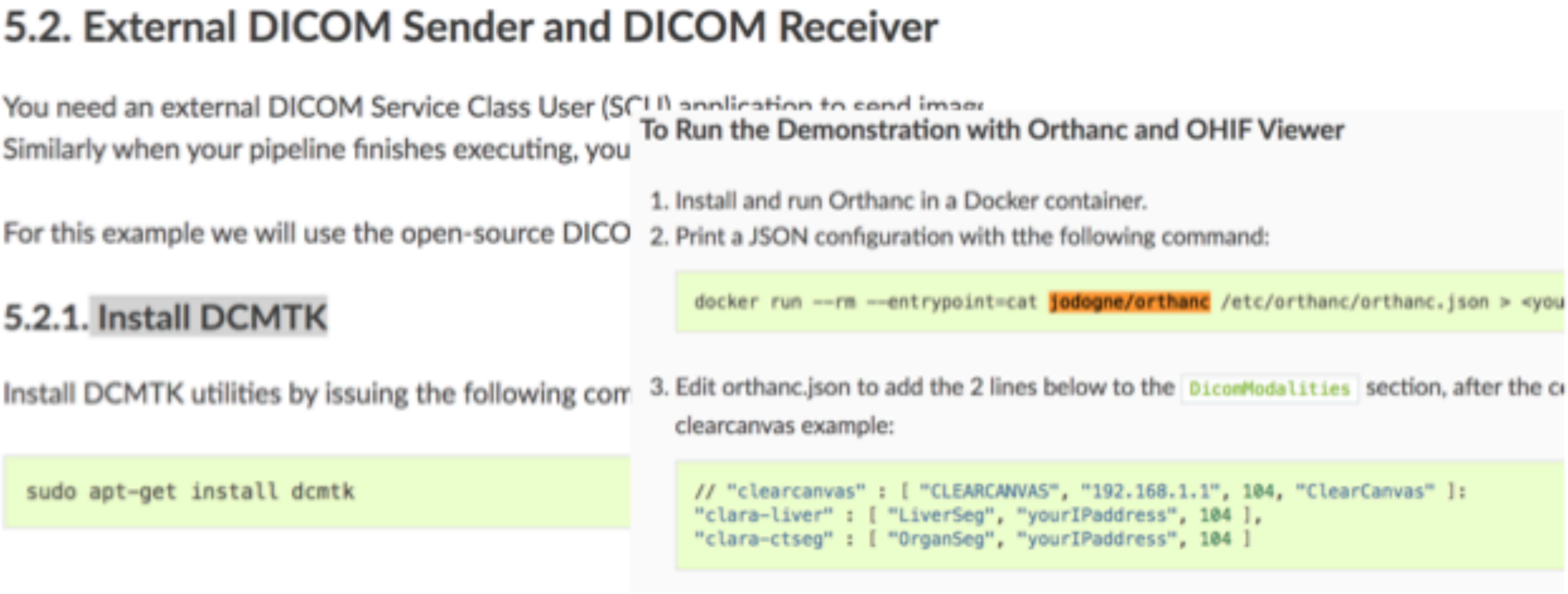
- Patient 1060
  - Study 1
    - MR Series 4
      - Image 1 : /Us
      - Image 2 : /Us
      - Image 3 : /Us
      - Image 4 : /Us
      - Image 5 : /Us
      - Image 6 : /Us
      - Image 7 : /Us
      - Image 8 : /Us
      - Image 9 : /Us
      - Image 10 : /U
      - Image 11 : /U
      - Image 12 : /U
      - Image 13 : /U
      - Image 14 : /U
      - Image 15 : /U
      - Image 16 : /U
      - Image 17 : /U
      - Image 18 : /U
      - Image 19 : /U
      - Image 20 : /U
      - Image 21 : /U
      - Image 22 : /U
      - Image 23 : /U
      - Image 24 : /U
      - Image 25 : /U

Tag	Name	Action	Value
(0008,0008)	Image Type	Keep	ORIGINAL\PRIMARY\M\ND
(0008,0012)	Instance Creation...	Keep	20101020
(0008,0013)	Instance Creation...	Keep	141346.117000
(0008,0016)	SOP Class UID	Keep	1.2.840.10008.5.1.4.1.1.4
(0008,0018)	SOP Instance UID	Keep	1.3.6.1.4.1.5962.99.1.1934890030.3242...
(0008,0020)	Study Date	Keep	20101020
(0008,0021)	Series Date	Keep	20101020
(0008,0022)	Acquisit		
(0008,0023)	Content		
(0008,0030)	Study T		18000
(0008,0031)	Series T		77000
(0008,0032)	Acquisit		82500
(0008,0033)	Content		17000
(0008,0050)	Accessi		
(0008,0060)	Modality		
(0008,0070)	Manufac		
(0008,0080)	Institutio		al
(0008,0081)	Institutio		, Metropolis
(0008,0090)	Referrin		Aaron
(0008,1010)	Station N		
(0008,1030)	Study D		N
(0008,103E)	Series D		
(0008,1050)	Perform		
(0008,1070)	Operato		es
(0008,1090)	Manufac		
(0010,0010)	Patient's Name	Keep	SURNAME^FORENAME
(0010,0020)	Patient ID	Keep	1060
(0010,0030)	Patient's Birth Date	Keep	19700401
(0010,0040)	Patient's Sex	Keep	M
(0010,1010)	Patient's Age	Keep	040Y
(0010,1020)	Patient's Size	Keep	1.87960376166667
(0010,1030)	Patient's Weight	Keep	86.182561263
(0018,0015)	Body Part Examined	Keep	HEAD
(0018,0020)	Scanning Sequence	Keep	SE\IR

NVIDIA CLARA Platform



DICOM Frankenstein



5.2. External DICOM Sender and DICOM Receiver

You need an external DICOM Service Class User (SCU) application to send imag
Similarly when your pipeline finishes executing, you

For this example we will use the open-source DICO

5.2.1. Install DCMTK

Install DCMTK utilities by issuing the following com

```
sudo apt-get install dcmtk
```

To Run the Demonstration with Orthanc and OHIF Viewer

1. Install and run Orthanc in a Docker container.
2. Print a JSON configuration with tthe following command:

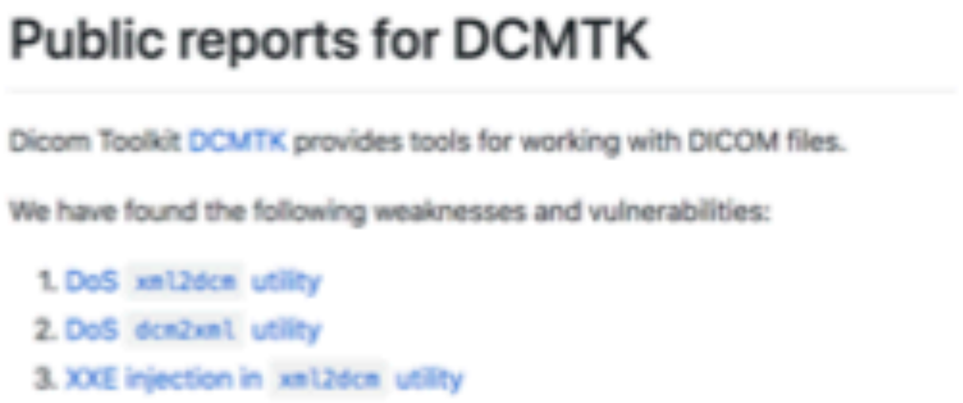
```
docker run --rm --entrypoint=cat jodogne/orthanc /etc/orthanc/orthanc.json > <you
```

3. Edit orthanc.json to add the 2 lines below to the `DicomModalities` section, after the c
clearcanvas example:

```
// "clearcanvas" : [ "CLEARCANVAS", "192.168.1.1", 104, "ClearCanvas" ]:
"clara-liver" : [ "LiverSeg", "yourIPaddress", 104 ],
"clara-ctseg" : [ "OrganSeg", "yourIPaddress", 104 ]
```

https://docs.nvidia.com/clara/deploy/RunningReferencePipeline.html

Do DICOM Series Dream of /etc/passwd?

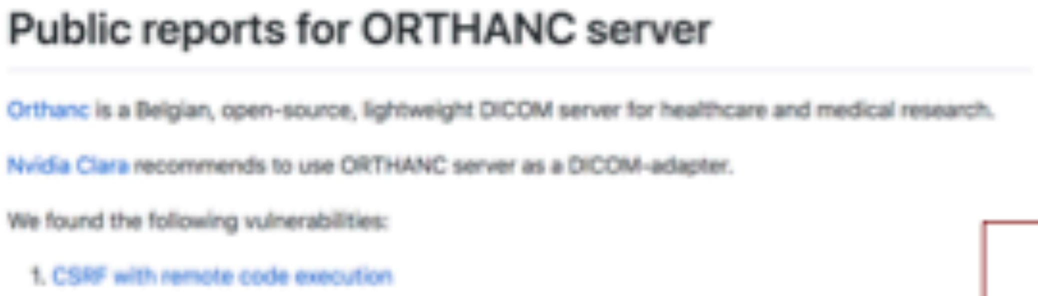


Public reports for DCMTK

Dicom Toolkit DCMTK provides tools for working with DICOM files.

We have found the following weaknesses and vulnerabilities:

1. DoS `xml2dcm` utility
2. DoS `dcm2xml` utility
3. XXE injection in `xml2dcm` utility



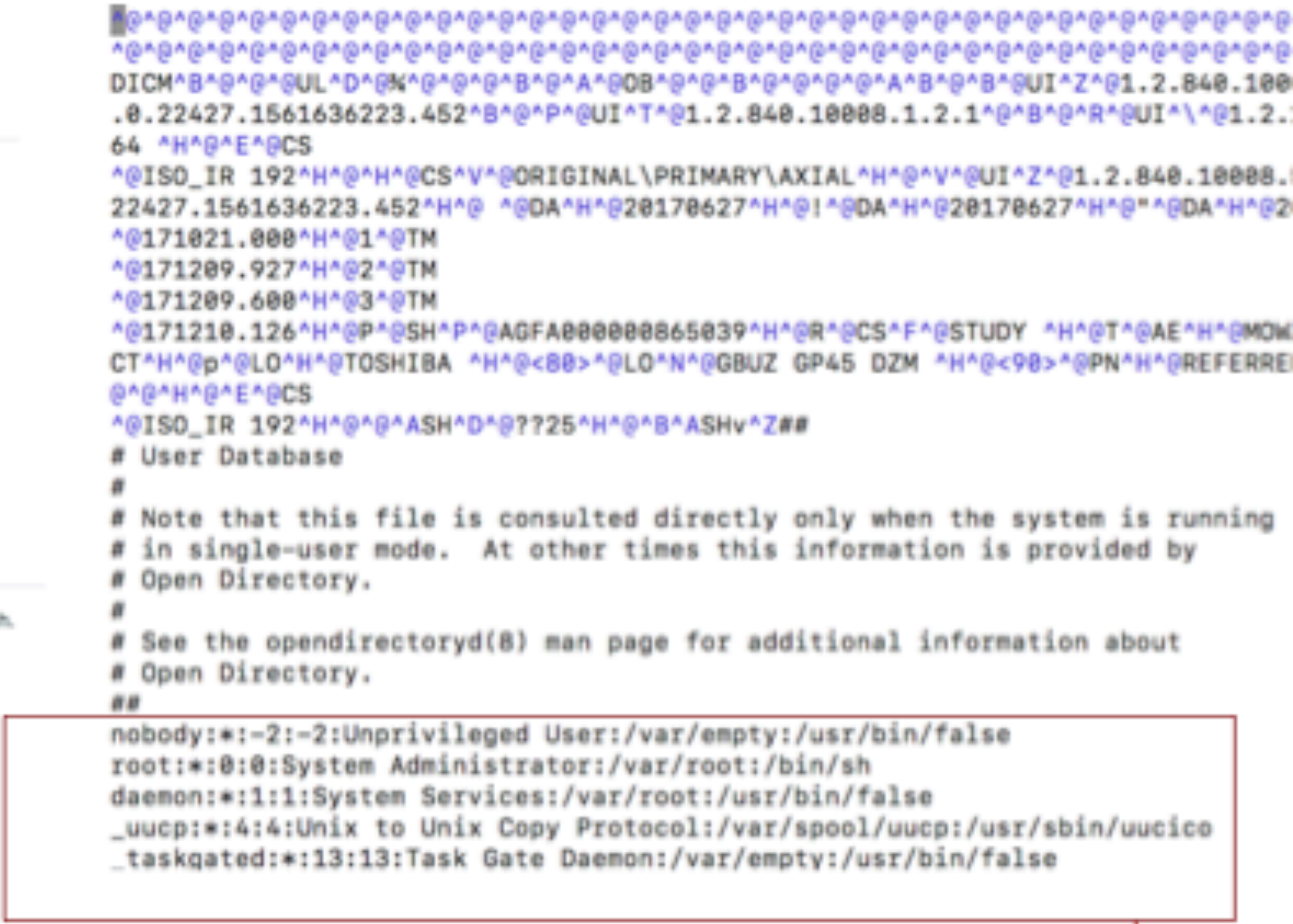
Public reports for ORTHANC server

Orthanc is a Belgian, open-source, lightweight DICOM server for healthcare and medical research.

Nvidia Clara recommends to use ORTHANC server as a DICOM-adapter.

We found the following vulnerabilities:

1. CSRF with remote code execution



```
^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^
^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^@^
DICM^B^@^@^@UL^D^@%^@^@^@^B^@^A^@OB^@^@^B^@^@^@^@^A^B^@^B^@UI^Z^@1.2.840.100
.0.22427.1561636223.452^B^@^P^@UI^T^@1.2.840.10008.1.2.1^@^B^@^R^@UI^\^@1.2.
64 ^H^@^E^@CS
^@ISO_IR 192^H^@^H^@CS^V^@ORIGINAL\PRIMARY\AXIAL^H^@^V^@UI^Z^@1.2.840.10008.
22427.1561636223.452^H^@ ^@DA^H^@20170627^H^@!^@DA^H^@20170627^H^@"^@DA^H^@2
^@171021.000^H^@1^@TM
^@171209.927^H^@2^@TM
^@171209.600^H^@3^@TM
^@171210.126^H^@P^@SH^P^@AGFA000000865039^H^@R^@CS^F^@STUDY ^H^@T^@AE^H^@MOW
CT^H^@p^@LO^H^@TOSHIBA ^H^@<80>^@LO^N^@GBUZ GP45 DZM ^H^@<90>^@PN^H^@REFERRE
@^@^H^@^E^@CS
^@ISO_IR 192^H^@^@^ASH^D^@??25^H^@^B^ASHv^Z##
# User Database
#
# Note that this file is consulted directly only when the system is running
# in single-user mode.  At other times this information is provided by
# Open Directory.
#
# See the opendirectoryd(8) man page for additional information about
# Open Directory.
##
nobody:*:-2:-2:Unprivileged User:/var/empty:/usr/bin/false
root:*:0:0:System Administrator:/var/root:/bin/sh
daemon:*:1:1:System Services:/var/root:/usr/bin/false
_uucp:*:4:4:Unix to Unix Copy Protocol:/var/spool/uucp:/usr/sbin/uucico
_taskgated:*:13:13:Task Gate Daemon:/var/empty:/usr/bin/false
```

http://www.scada.sl/2019/10/dicom-to-passwd-on-security-of-ml.html

Tensorflow graphs as malware

- **The TensorFlow server is meant for internal communication only. It is not built for use in an untrusted network.**
- By default, ModelServer also has no built-in mechanism for authentication.
- TensorFlow may read and write files, send and receive data over the network, and even spawn additional processes.

TensorFlow
Security
TensorFlow Models as Programs
Running Untrusted Models
Accepting The Untrusted Input
Vulnerabilities in TensorFlow
Reporting a Vulnerability

https://data-flair.training/blogs/tensorflow-security/

https://github.com/tensorflow/tensorflow/blob/master/SECURITY.md

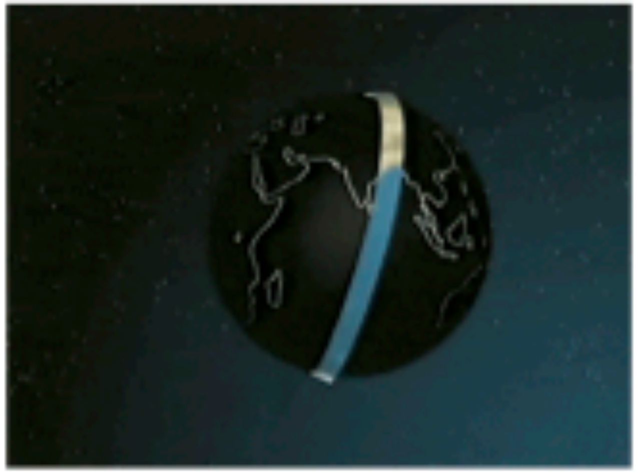
Is it real?

We don’t know yet

Epoch 2

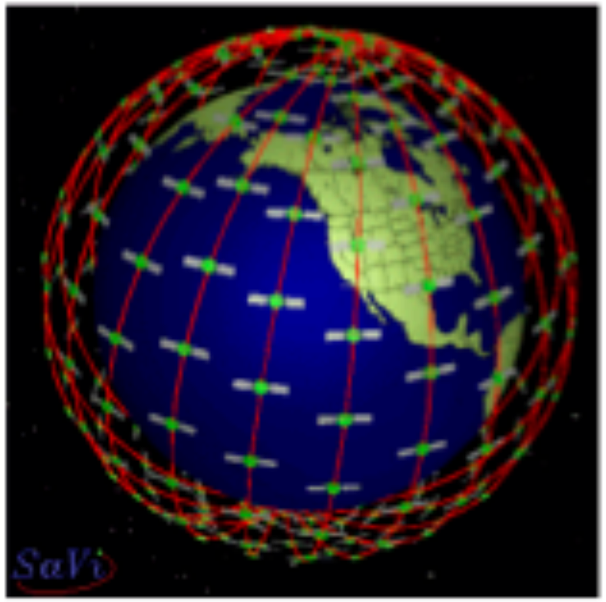
Notes on HUGE data

The Satellite Flies High…



- 1 PT of images daily
- Different formats/sources/types

- Different models
- Different regions
- Overfitting rulez!



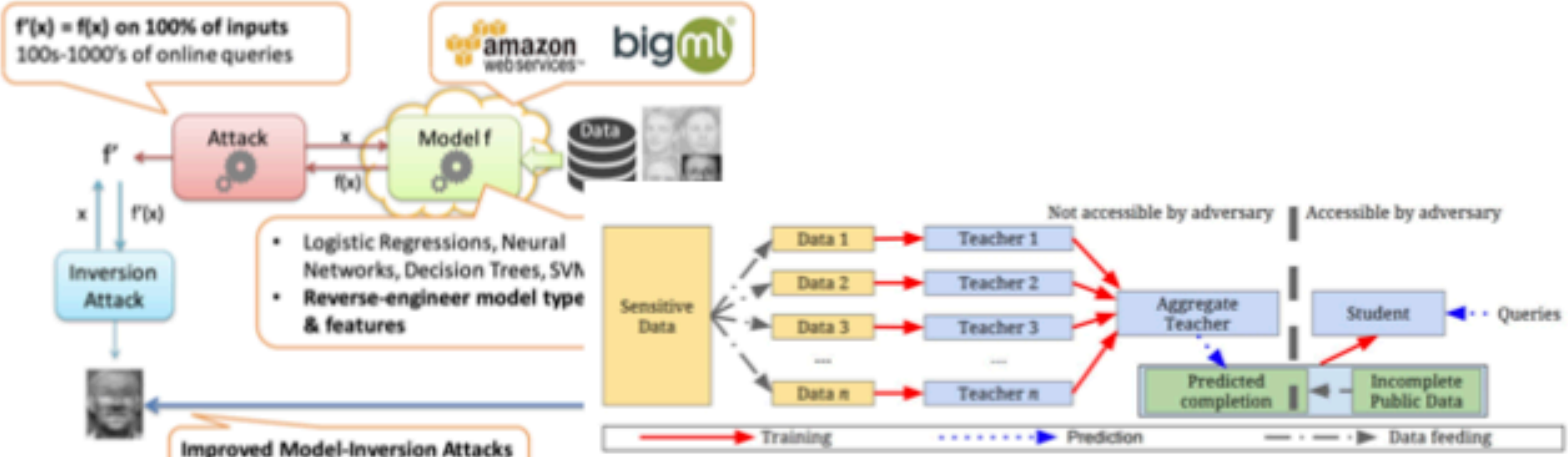
Multispectral sources

NOAA 18/19
MetOp-A/B
Terra
Aqua
Suomi NPP
NOAA 20 (JPSS-1)
FengYun-3A/B/C

Data questions

- Data collection and privacy
- Data integrity
- Training cycle
  - Model integrity?
- IP protection

Model Extraction Attacks



f'(x) = f(x) on 100% of inputs
100s-1000's of online queries

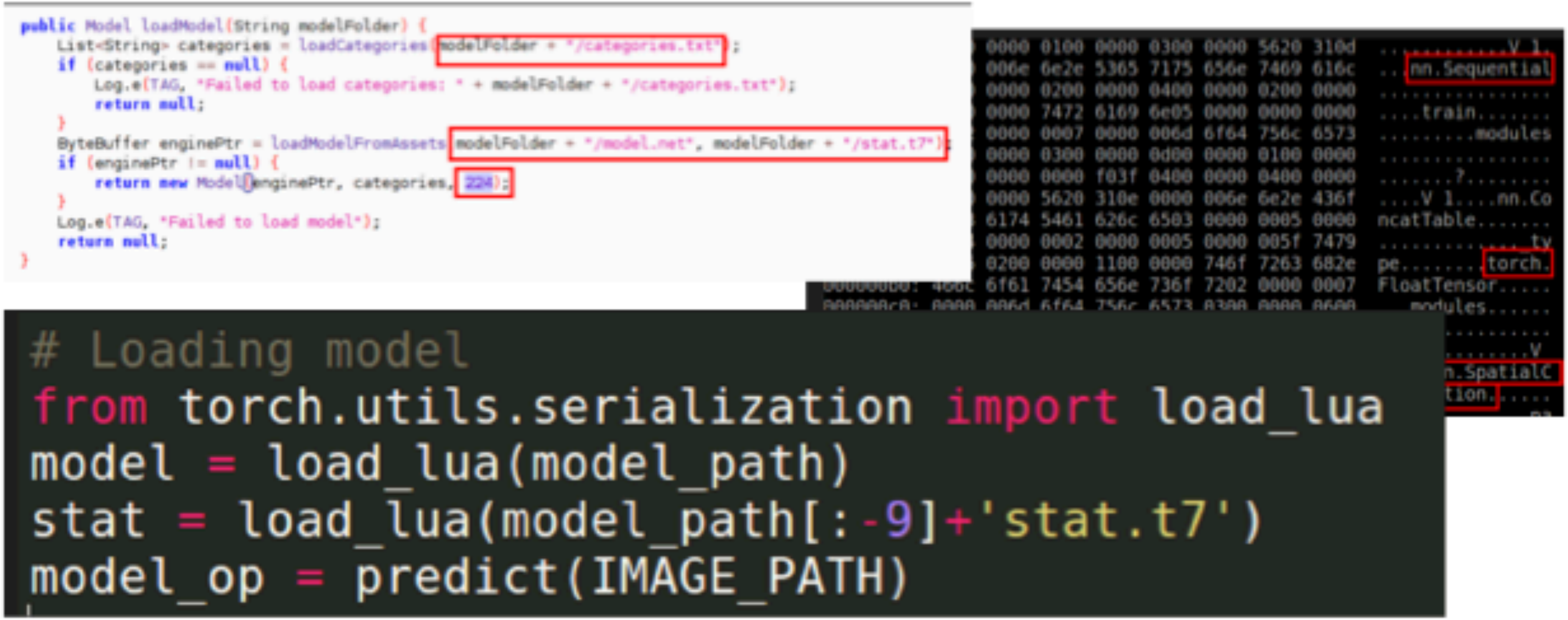
- Logistic Regressions, Neural Networks, Decision Trees, SVM
- **Reverse-engineer model type & features**

Improved Model-Inversion Attacks
[Fredrikson et al. 2015]

Figure 1: Overview of the approach: (1) an ensemble of teachers is trained on disjoint subsets of the sensitive data, (2) a student model is trained on public data labeled using the ensemble.

Tramèr, F. (2016). Stealing Machine Learning Models via Prediction APIs.

...binwalk + grep + strings



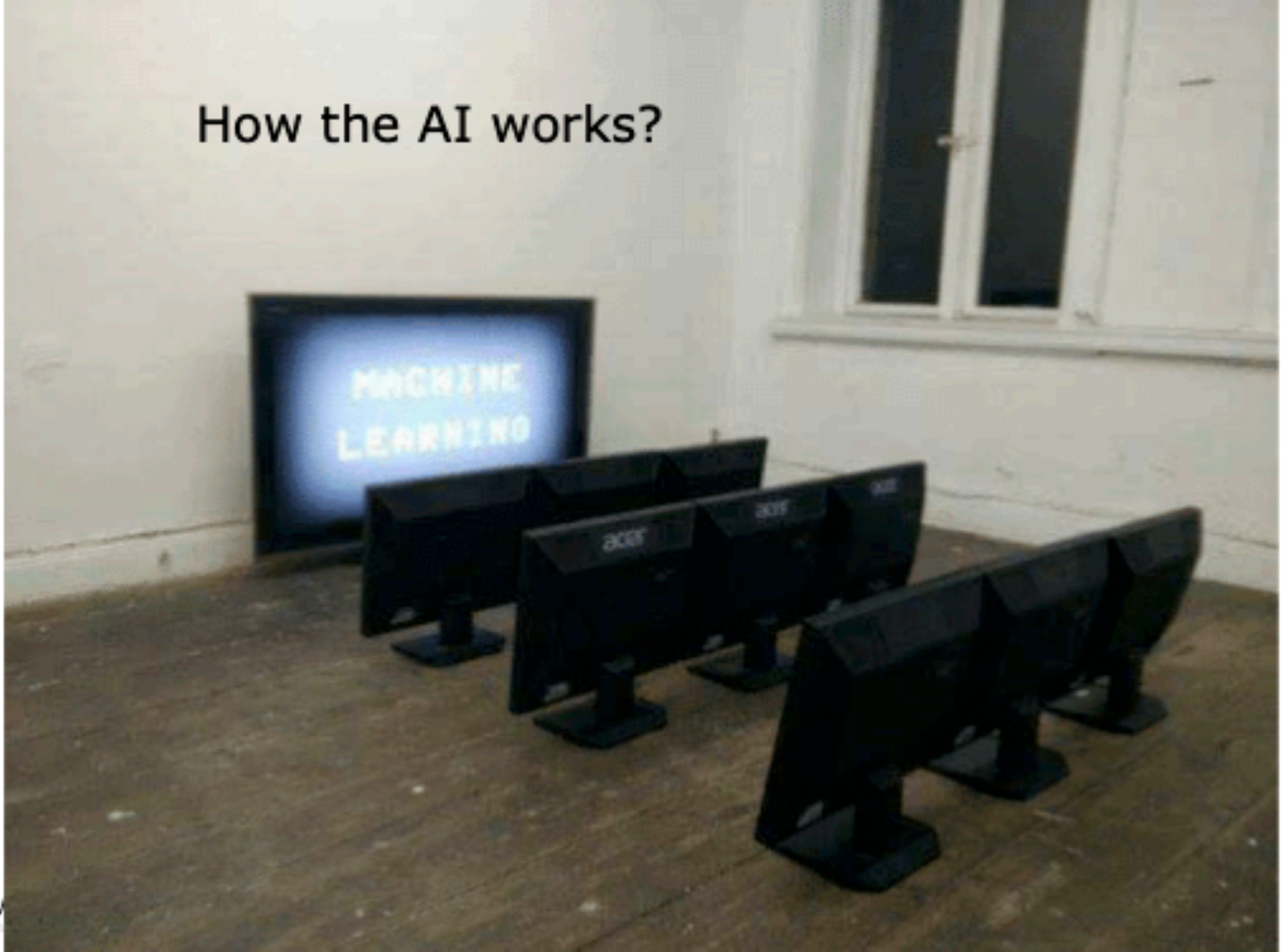
```java
public Model loadModel(String modelFolder) {
    List<String> categories = loadCategories(modelFolder + "/categories.txt");
    if (categories == null) {
        Log.e(TAG, "Failed to load categories: " + modelFolder + "/categories.txt");
        return null;
    }
    ByteBuffer enginePtr = loadModelFromAssets(modelFolder + "/model.net", modelFolder + "/stat.t7");
    if (enginePtr != null) {
        return new Model(enginePtr, categories, 224);
    }
    Log.e(TAG, "Failed to load model");
    return null;
}
```

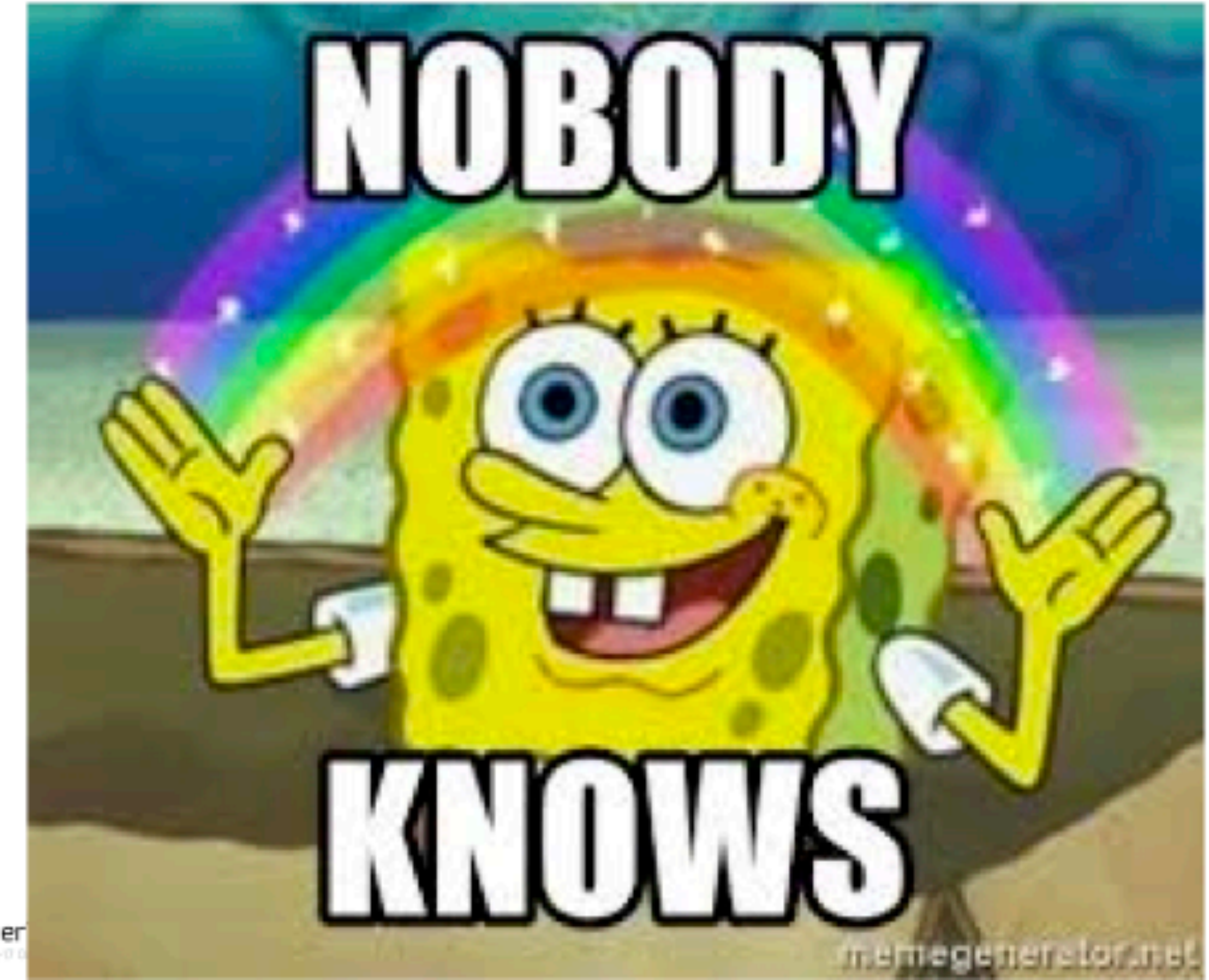
```
0000 0100 0000 0300 0000 5620 310d  ...............V 1.
006e 6e2e 5365 7175 656e 7469 616c  ...nn.Sequential
0000 0200 0000 0400 0000 0200 0000  ................
0000 7472 6169 6e05 0000 0000 0000  .....train.......
0000 0007 0000 006d 6f64 756c 6573  ..........modules
0000 0300 0000 0d00 0000 0100 0000  ................
0000 0000 f03f 0400 0000 0400 0000  ........?.......
0000 5620 310e 0000 006e 6e2e 436f  ....V 1.....nn.Co
6174 5461 626c 6503 0000 0005 0000  ncatTable.......
0000 0002 0000 0005 0000 005f 7479  ..............ty
0200 0000 1100 0000 746f 7263 682e  pe.........torch.
000000b0: 466c 6f61 7454 656e 736f 7202 0000 0007  FloatTensor.....
000000c0: 0000 006d 6f64 756c 6573 0300 0000 0600  ....modules.....
                                                   ...........
                                                   .........V
                                                   n.SpatialC
                                                   tion......
```

```python
# Loading model
from torch.utils.serialization import load_lua
model = load_lua(model_path)
stat = load_lua(model_path[:-9]+'stat.t7')
model_op = predict(IMAGE_PATH)
```

Nikhil Joshi, Rewanth Cool. GDALR: An efficient model duplication attack on black-box Machine Learning models
https://static.ptsecurity.com/phdays/presentations/phdays-9-gdalr-an-efficient-model-duplication-attack-on-black-box-machine-learning-models.pdf

How the AI works?




NOBODY
KNOWS
memegenerator.net

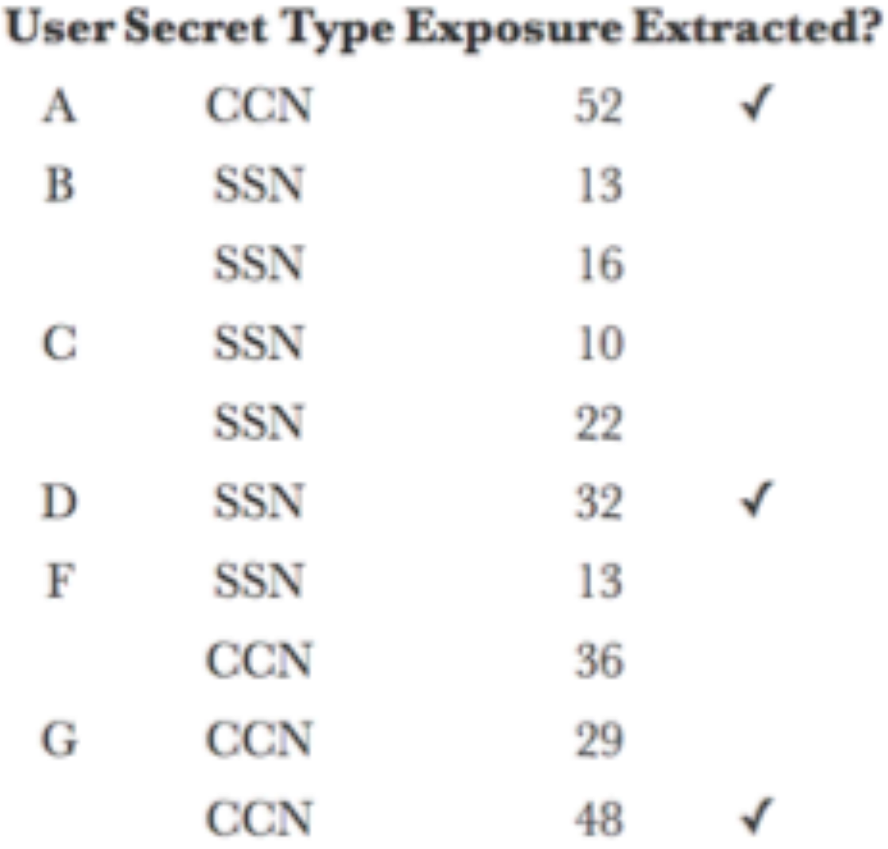
Video

https://www.youtube.com/watch?v=AgkfIQ4IGaM

https://github.com/yosinski/deep-visualization-toolbox

Memorization in Neural Networks

In experiments, we show that unintended memorization is a persistent, hard-to-avoid issue that can have serious consequences. Specifically, for models trained without consideration of memorization, we describe new, efficient procedures that can extract unique, secret sequences, such as credit card numbers

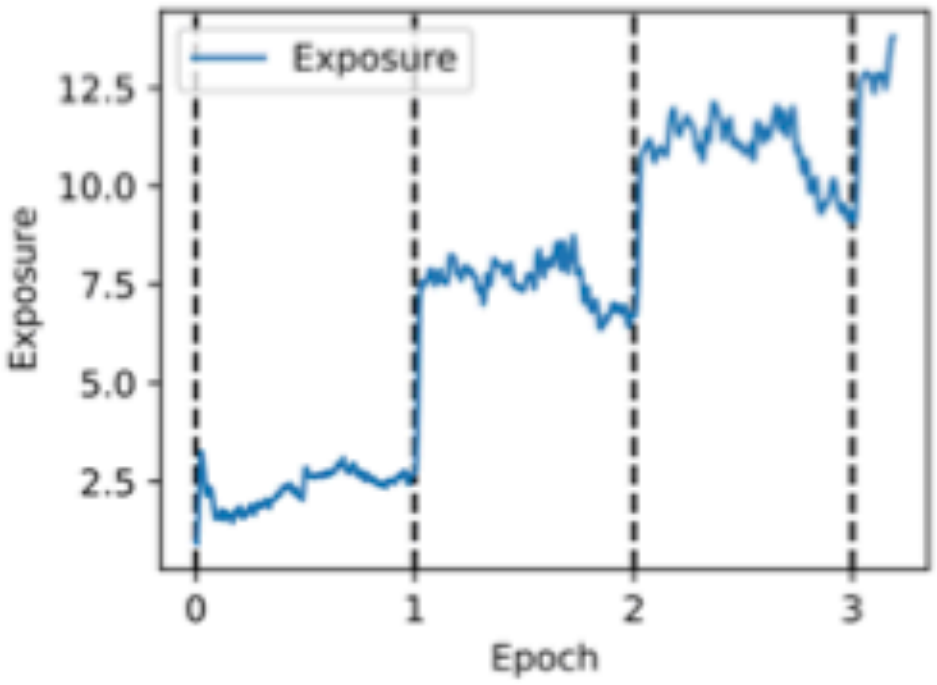


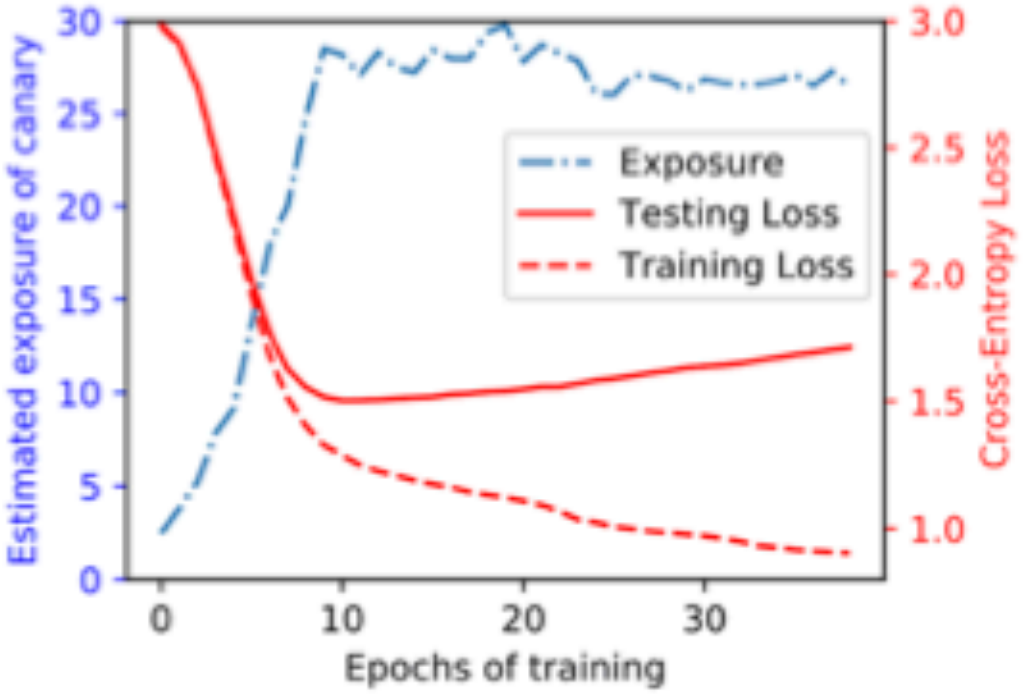
User	Secret Type	Exposure	Extracted?
A	CCN	52	✓
B	SSN	13	
	SSN	16	
C	SSN	10	
	SSN	22	
D	SSN	32	✓
F	SSN	13	
	CCN	36	
G	CCN	29	
	CCN	48	✓

Carlini, Nicholas et al. "The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks."

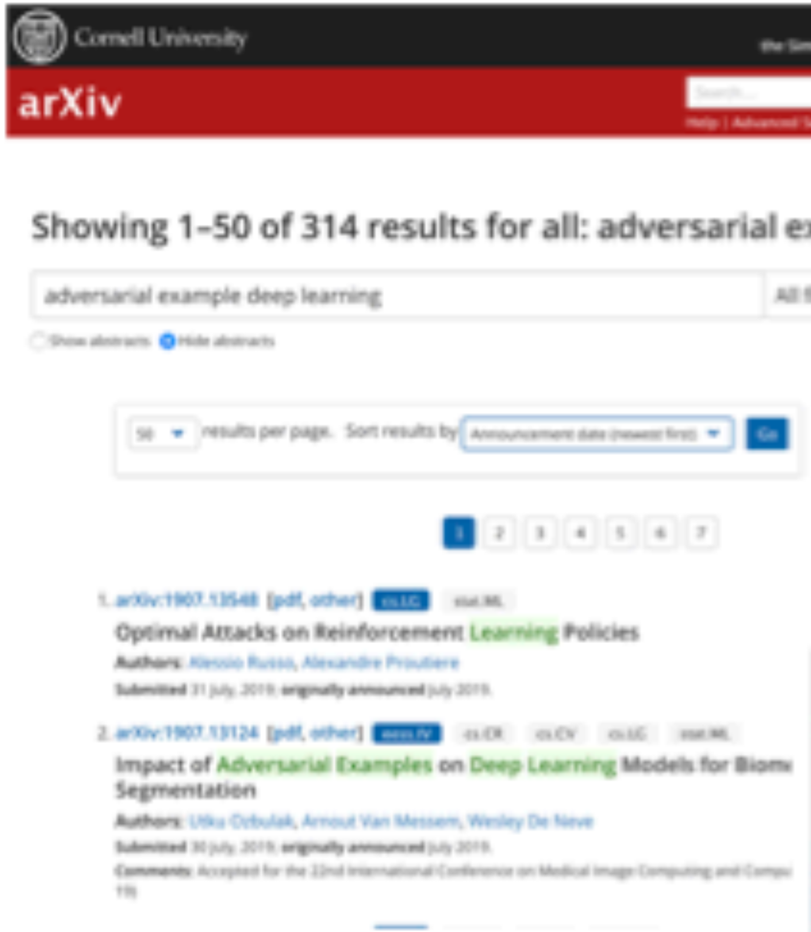
Data in the model and model as a data

The Lottery Ticket Hypothesis at Scale
Jonathan Frankle, Gintare Karolina Dziugaite, Daniel M. Roy, Michael Carbin





Adversarial example: Being John Malkovich



Cornell University

We gratefully acknowledge support from the Simons Foundation and member institutions.

arXiv

Search... All fields Search

Help | Advanced Search

Login

Showing 1–50 of 314 results for all: adversarial example deep learni

adversarial example deep learning | All fields | Search

○ Show abstracts ● Hide abstracts

Advanced Search

50 results per page. Sort results by Announcement date (newest first) Go

1 2 3 4 5 6 7 Next

1. arXiv:1907.13548 [pdf, other] cs.LG stat.ML

Optimal Attacks on Reinforcement Learning Policies

Authors: Alessio Russo, Alexandre Proutiere

Submitted 31 July, 2019; originally announced July 2019.

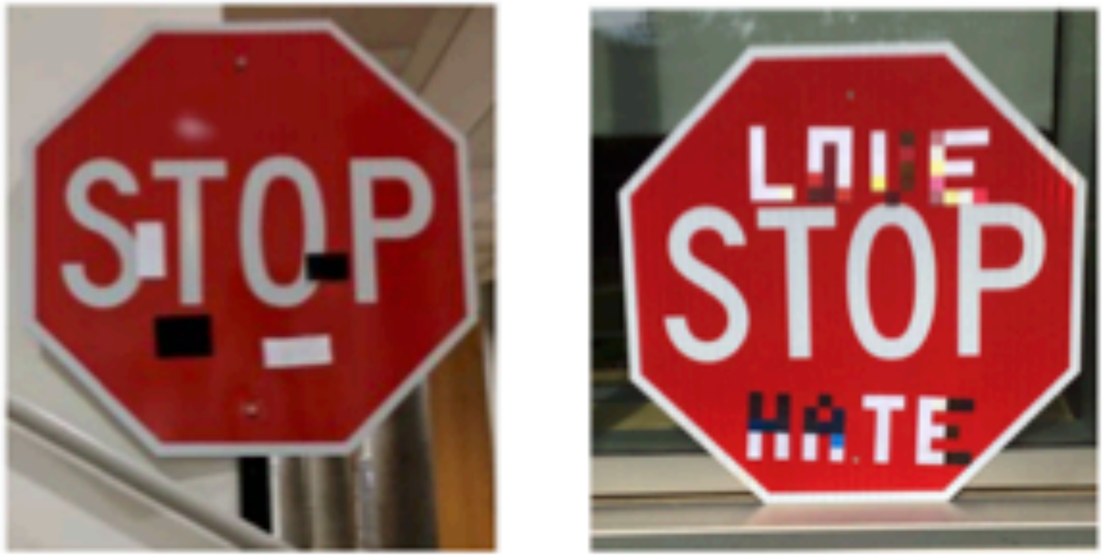
2. arXiv:1907.13124 [pdf, other] eess.IV cs.CR cs.CV cs.LG stat.ML

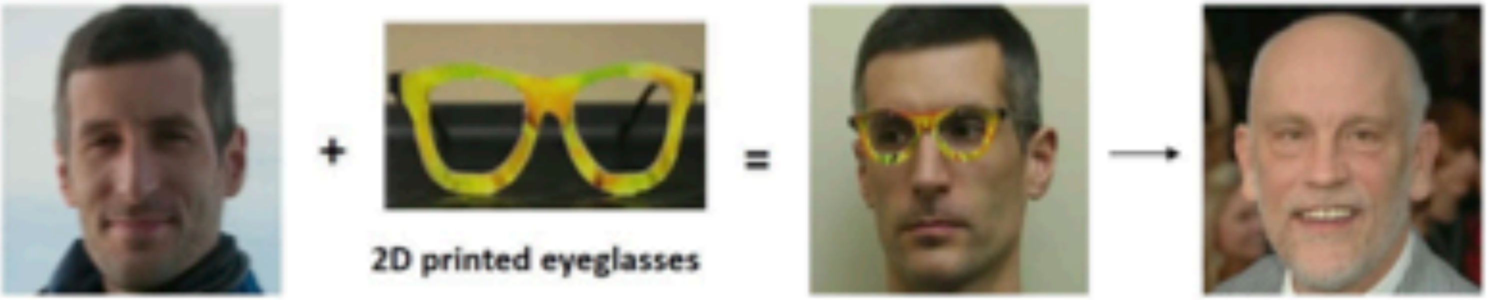
Impact of Adversarial Examples on Deep Learning Models for Biom Segmentation

Authors: Utku Ozbulak, Arnout Van Messem, Wesley De Neve

Submitted 30 July, 2019; originally announced July 2019.

Comments: Accepted for the 22nd International Conference on Medical Image Computing and Compu 19)



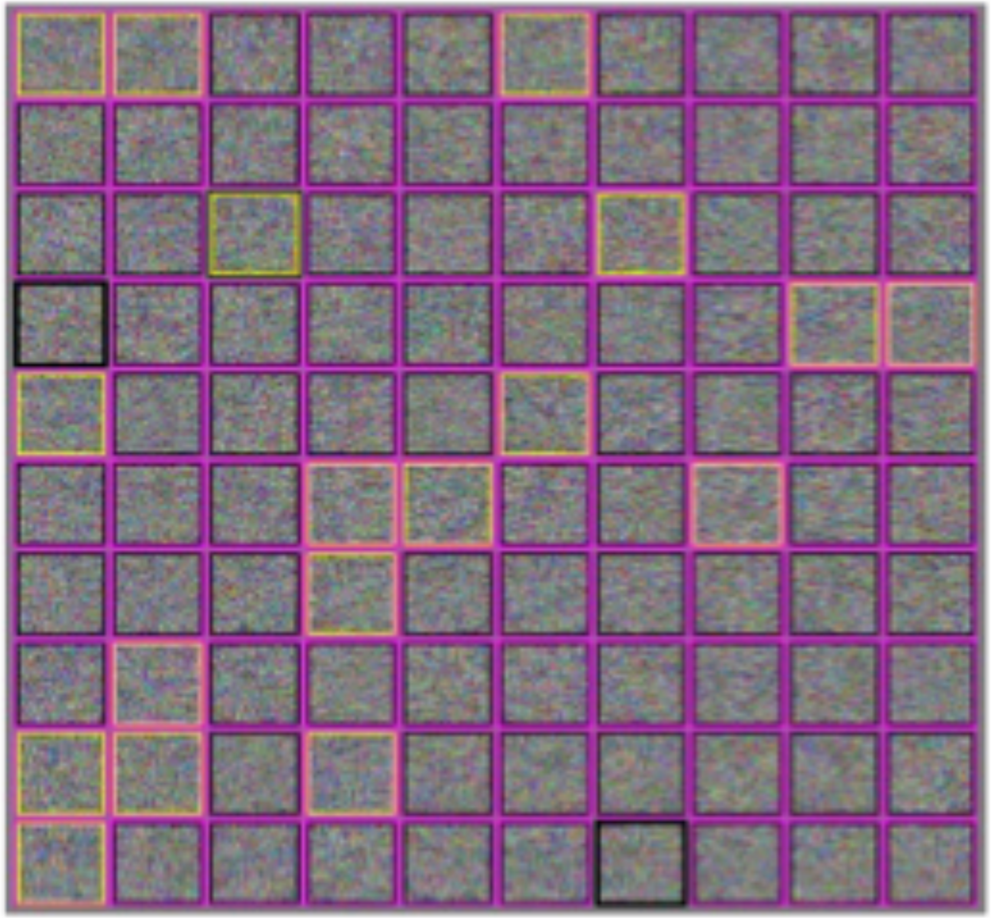


Ivan Evtimov, et al, "Robust Physical-World Attacks on Machine Learning Models"

CIFAR-10 classifier on Gaussian noise

Pink box – something
Yellow box – airplane
one step FGSM

("Clever Hans, Clever Algorithms," Bob Sturm)



(Goodfellow 2016)

https://www.youtube.com/watch?v=CIfsB_EYsVI&t=1756s

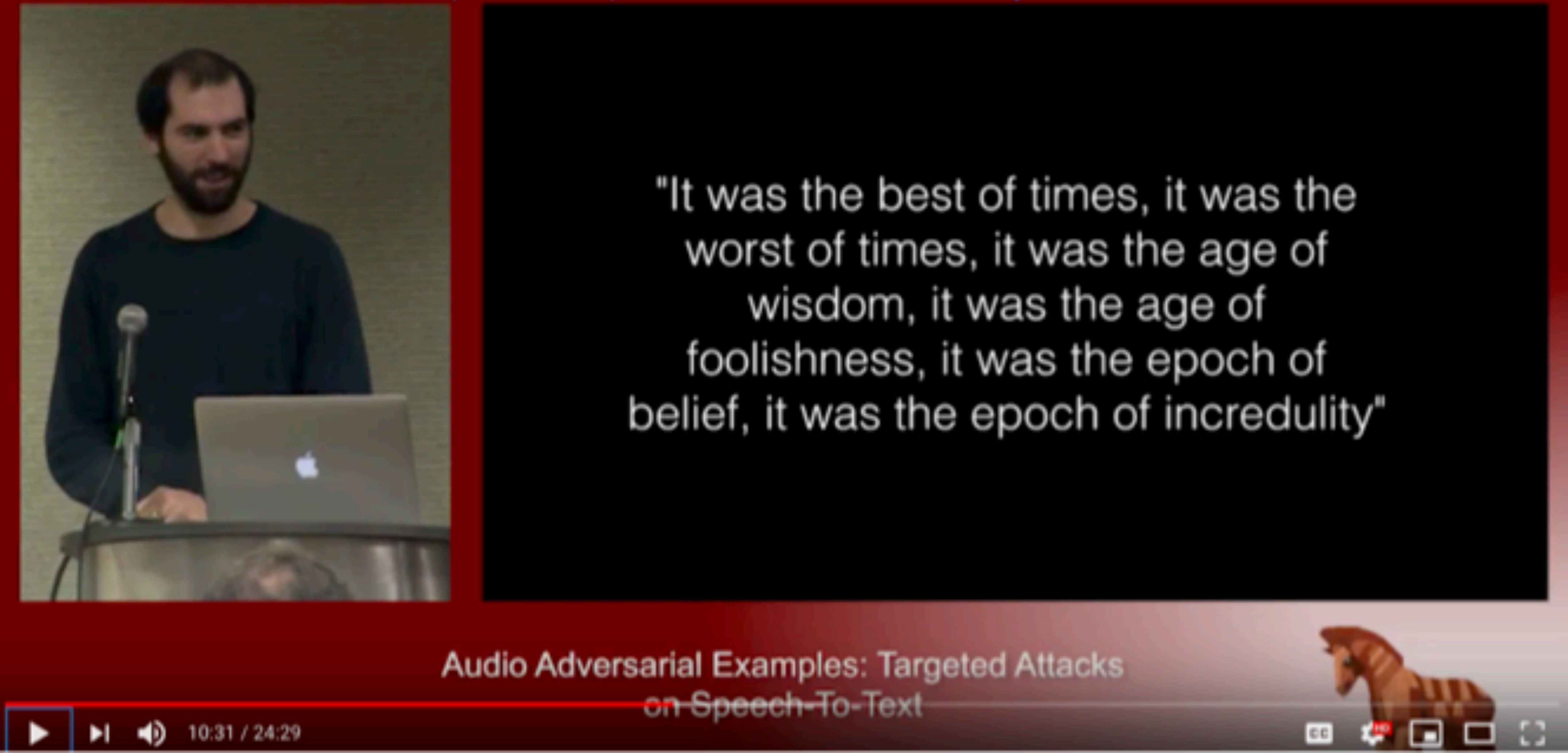
Justin Johnson, Adversarial Examples and Adversarial Training

3D Mask presentation attack

https://twitter.com/mbrennanchina/status/1158435099773304833

DLS 2018

https://www.youtube.com/watch?v=Ho5jLKfoKSA

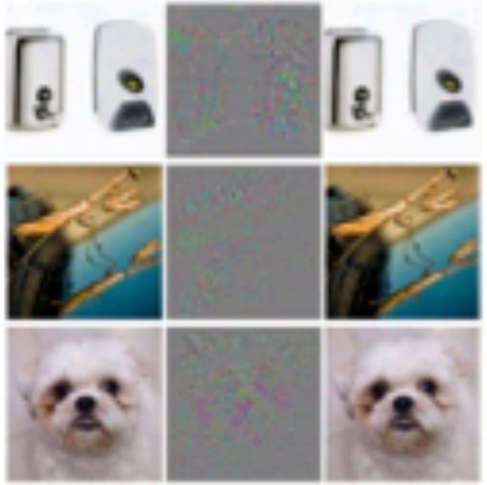


"It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity"

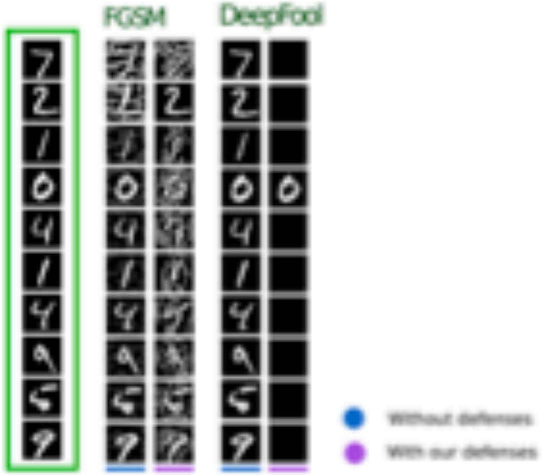
Audio Adversarial Examples: Targeted Attacks on Speech-To-Text

Adversarial Robustness???

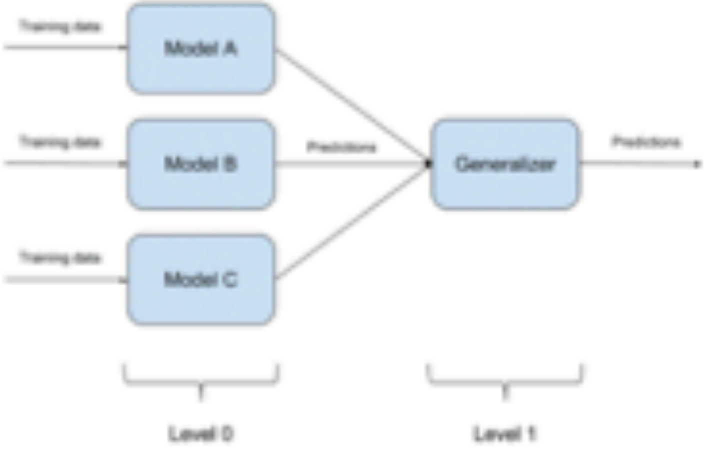
Adversarial Training



Gaussian Data Augmentation



Ensemble learning

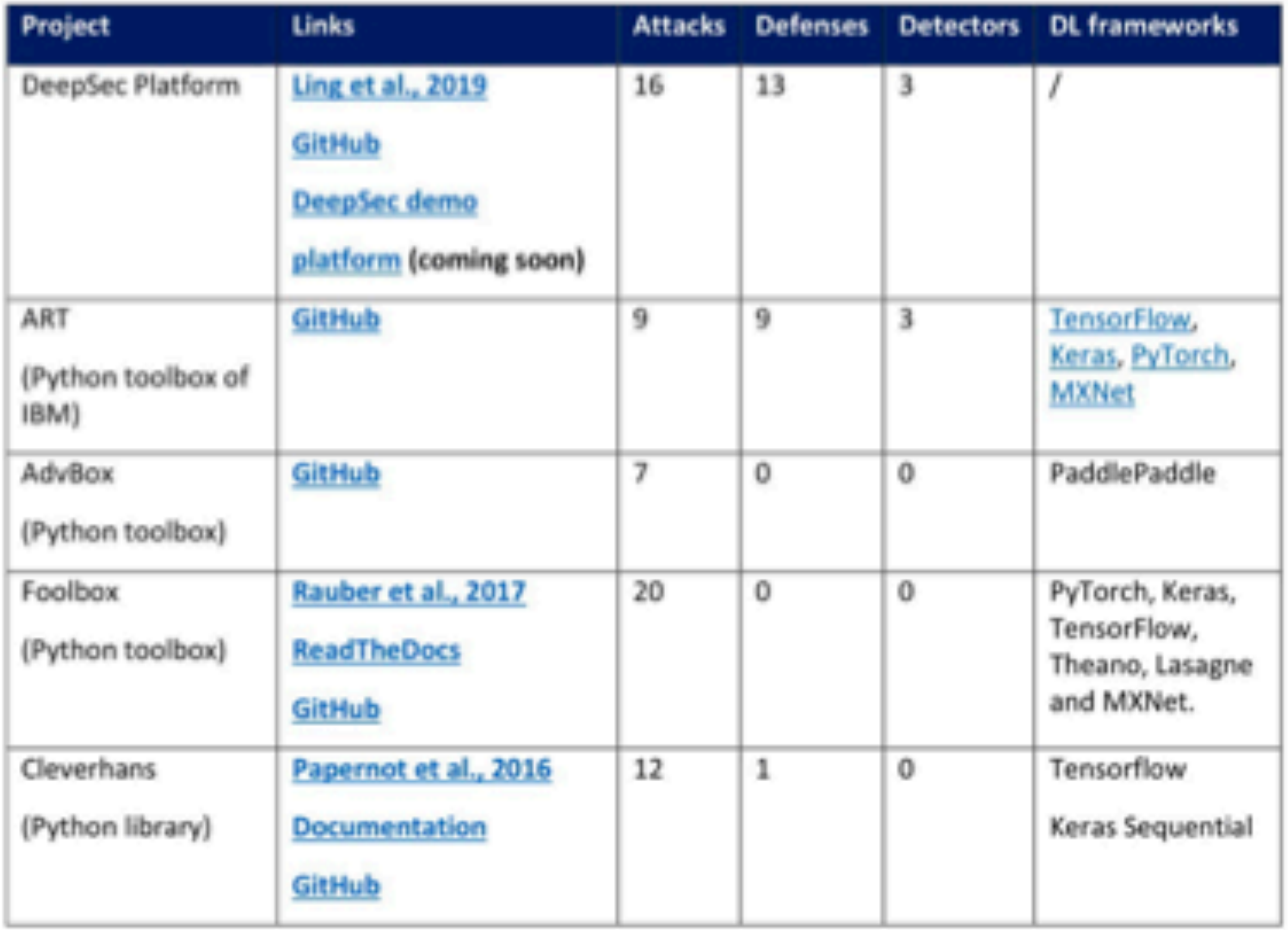


Ensemble of weak defenses does not lead to strong defense...

Adversarial Example Frameworks

Fool your AI!

But… Never trust it..

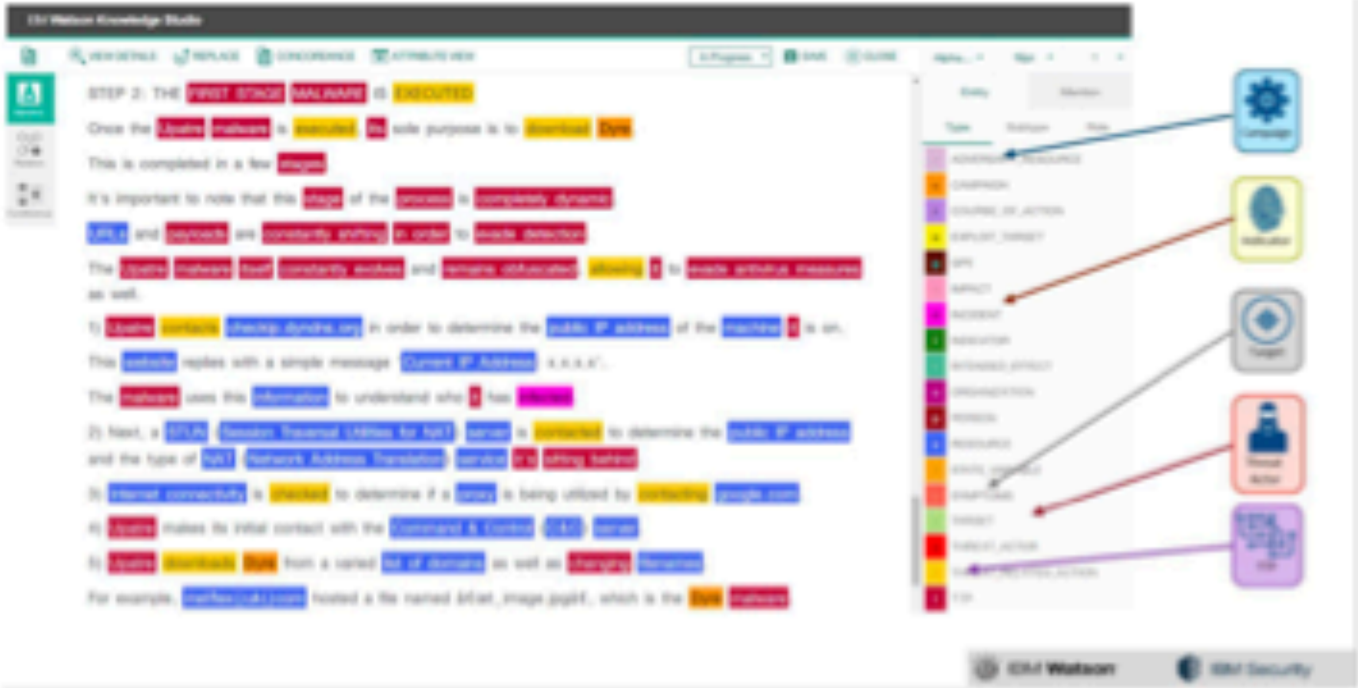


Project	Links	Attacks	Defenses	Detectors	DL frameworks
DeepSec Platform	Ling et al., 2019 GitHub DeepSec demo platform (coming soon)	16	13	3	/
ART (Python toolbox of IBM)	GitHub	9	9	3	TensorFlow, Keras, PyTorch, MXNet
AdvBox (Python toolbox)	GitHub	7	0	0	PaddlePaddle
Foolbox (Python toolbox)	Rauber et al., 2017 ReadTheDocs GitHub	20	0	0	PyTorch, Keras, TensorFlow, Theano, Lasagne and MXNet.
Cleverhans (Python library)	Papernot et al., 2016 Documentation GitHub	12	1	0	Tensorflow Keras Sequential

Epoch 3

AI for Security

AI Security Magic



STEP 2: THE FIRST STAGE MALWARE IS EXECUTED

Once the Upatre malware is executed, its sole purpose is to download Dyre.

This is completed in a few stages.

It's important to note that this stage of the process is completely dynamic.

URLs and payloads are constantly shifting in order to evade detection.

The Upatre malware itself constantly evolves and remains obfuscated, allowing it to evade antivirus measures as well.

1) Upatre contacts checkip.dyndns.org in order to determine the public IP address of the machine it is on.

This website replies with a simple message 'Current IP Address: x.x.x.x'.

The malware uses this information to understand who it has infected.

2) Next, a STUN (Session Traversal Utilities for NAT) server is contacted to determine the public IP address and the type of NAT (Network Address Translation) service it is sitting behind.

3) Internet connectivity is checked to determine if a proxy is being utilized by contacting google.com.

4) Upatre makes its initial contact with the Command & Control (C&C) server.

5) Upatre downloads Dyre from a varied list of domains as well as changing filenames.

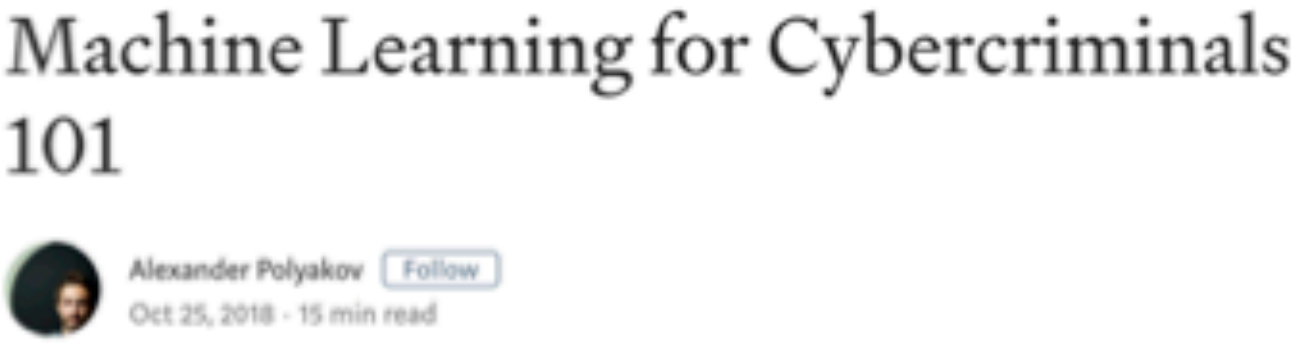
For example, [illegible] hosted a file named [illegible], which is the Dyre malware.

Perceives
Reasons
Visualizes
Understands
natural language
Explores
Relates
Adapts
and learns
Generates
and evaluates
hypotheses

> Customers are often confused by mismatches between (IBM's) marketing messages and actual, purchasable products.

Translation: IBM's marketing is bullshit.

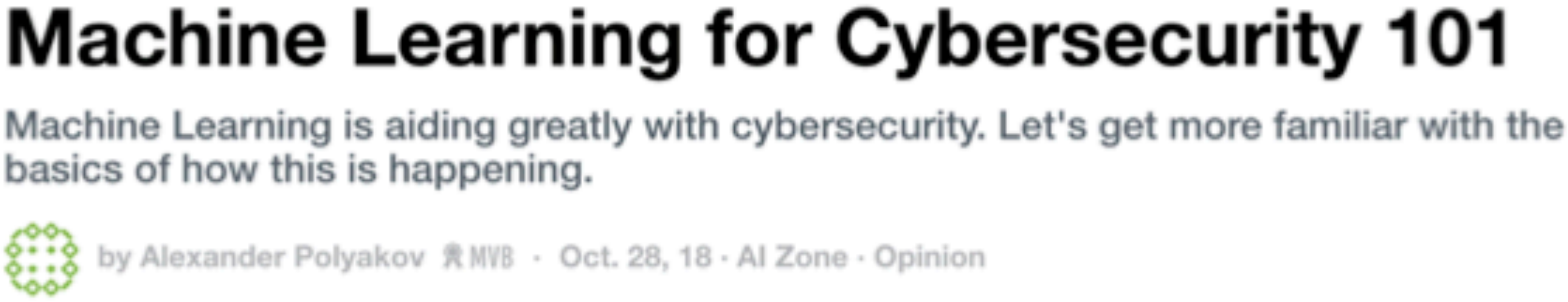
AI Security 101



Machine Learning for Cybercriminals 101

Alexander Polyakov Follow

Oct 25, 2018 · 15 min read



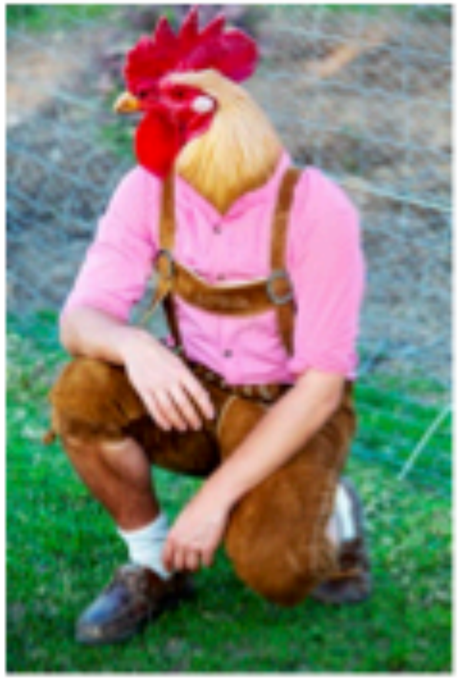
Machine Learning for Cybersecurity 101

Machine Learning is aiding greatly with cybersecurity. Let's get more familiar with the basics of how this is happening.

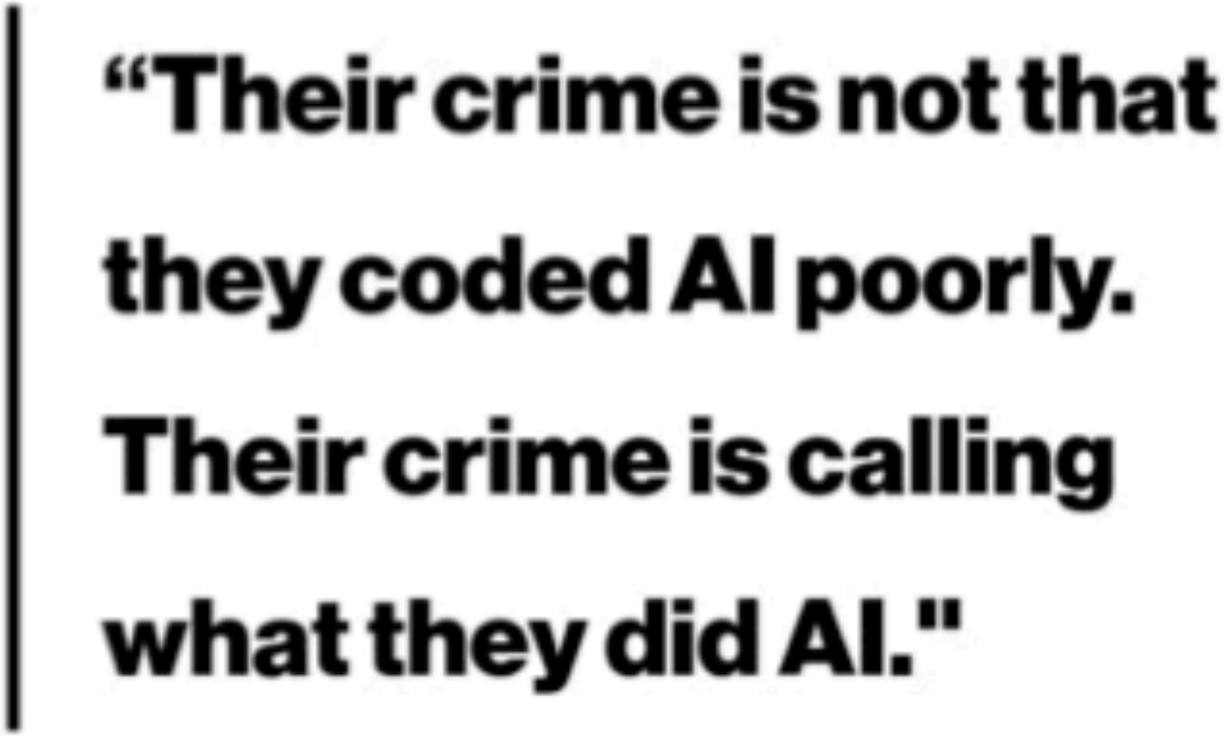
by Alexander Polyakov MVB · Oct. 28, 18 · AI Zone · Opinion

https://dzone.com/articles/machine-learning-for-cybersecurity-101

Skylight Cyber – “AI” antivirus bypass with copy



Not a real chicken



“Their crime is not that they coded AI poorly. Their crime is calling what they did AI."

Martijn Grooten

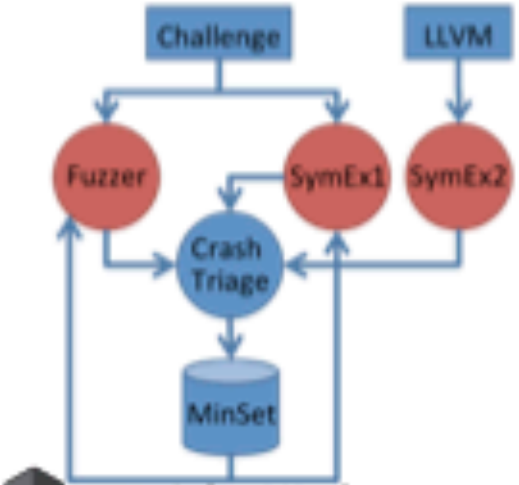
https://skylightcyber.com/2019/07/18/cylance-i-kill-you/

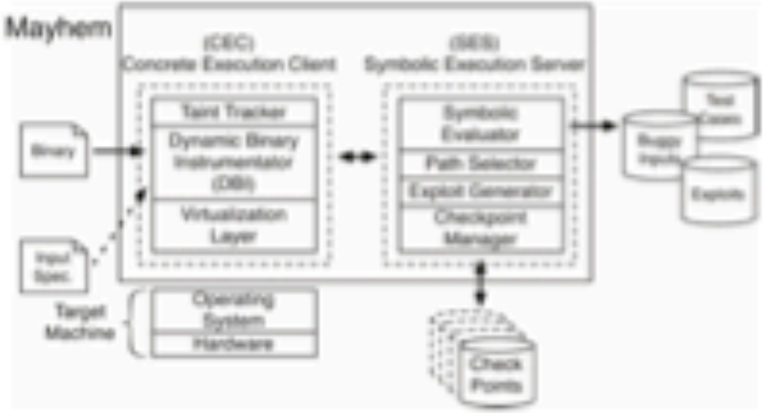
https://www.vice.com/en_us/article/9kxp83/researchers-easily-trick-cylances-ai-based-antivirus-into-thinking-malware-is-goodware

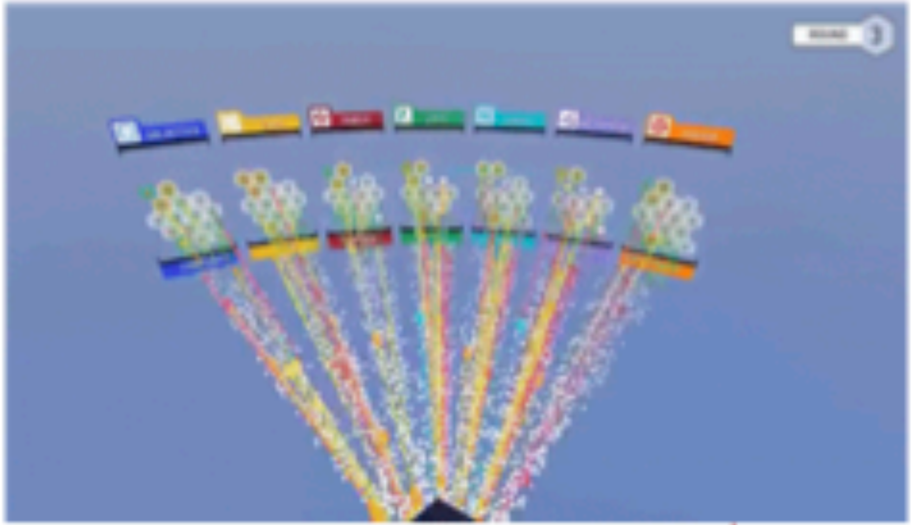
DARPA Cyber Grand Challenge 2016

…create automatic defensive systems capable of reasoning about flaws, formulating patches and deploying them on a network in real time…

Network Capture ⇒ Fuzzer ⇒ SymEx1 ⇒ Fuzzer ⇒ Crash



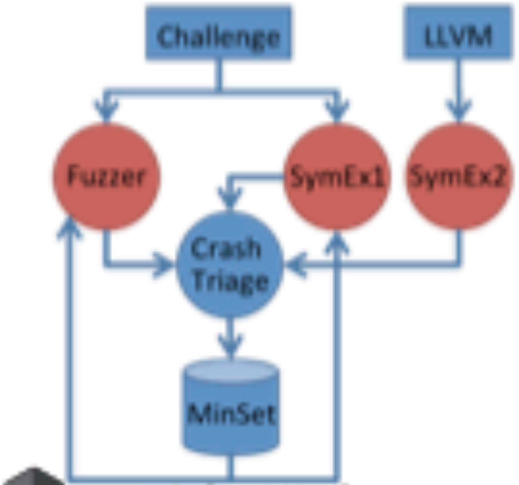


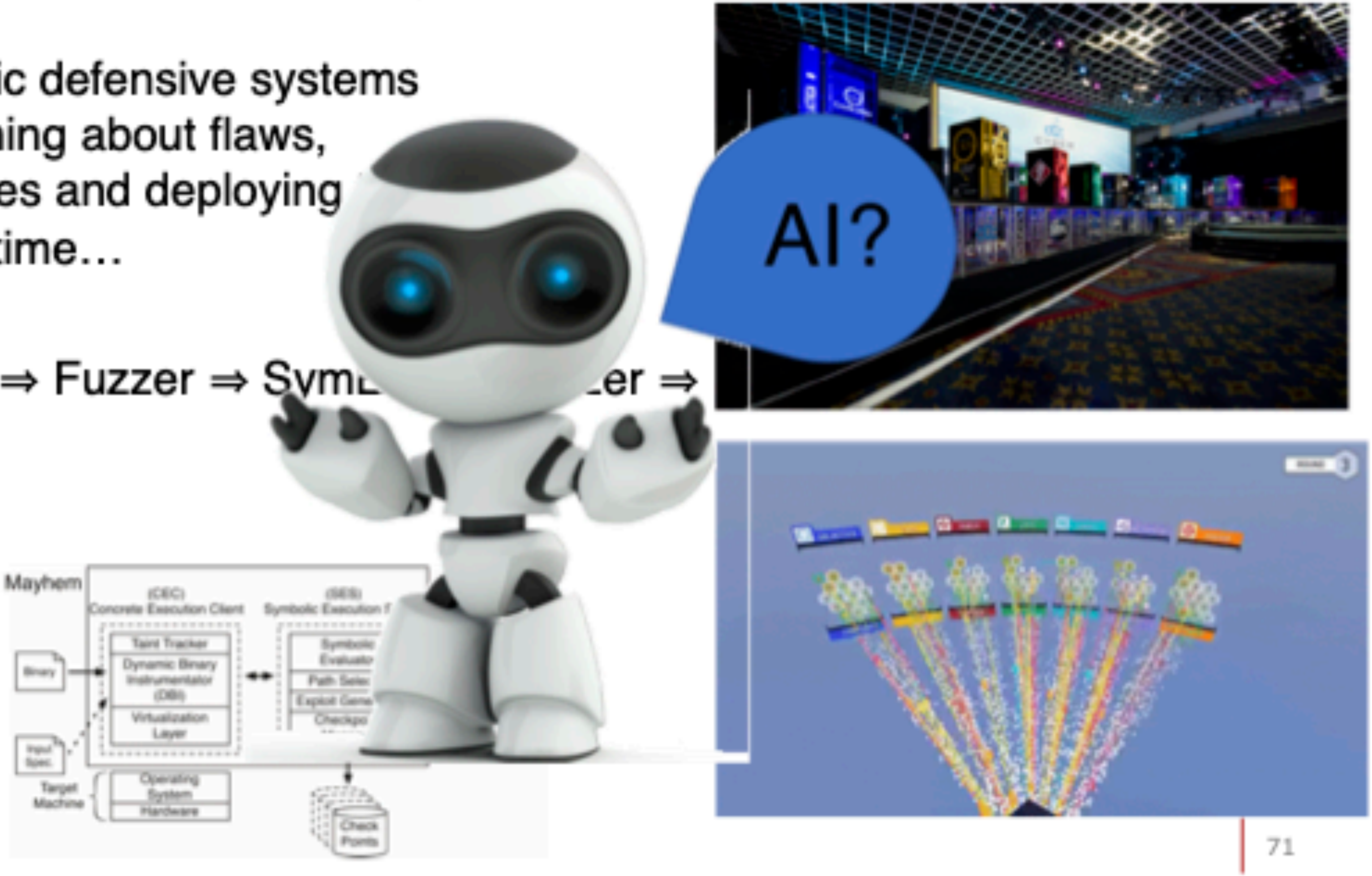


DARPA Cyber Grand Challenge 2016

…create automatic defensive systems capable of reasoning about flaws, formulating patches and deploying a network in real time…

Network Capture ⇒ Fuzzer ⇒ Sym zer ⇒ Crash





Epoch 5

As IS

You should scan all these Internets for AI

Grinder Framework

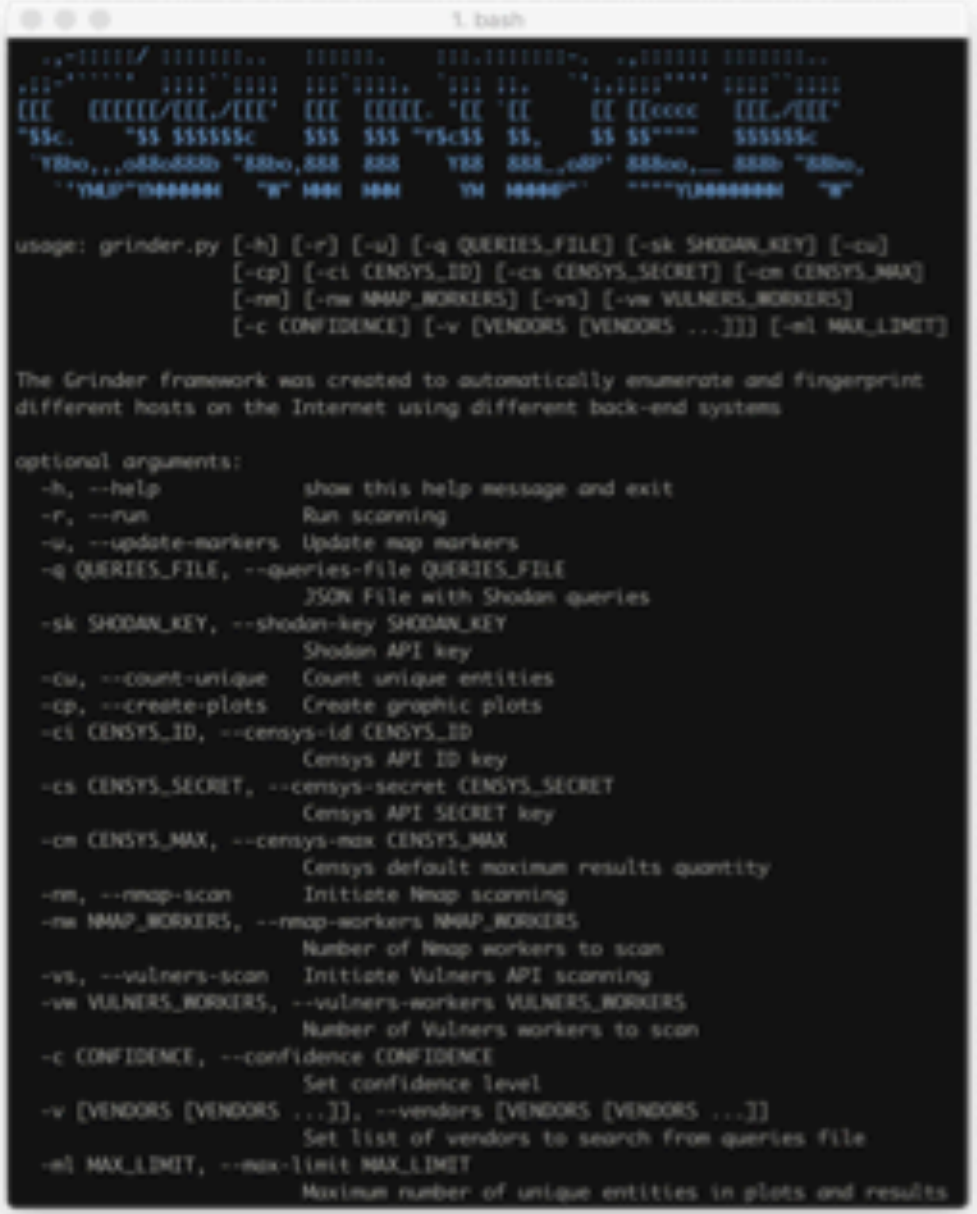
grinder

🔎 Python framework to automatically discover and enumerate hosts from different back-end systems (Shodan, Censys)

python | nmap | vulnerability-scanners | python-framework | shodan-api | vulners | censys-api

● Python ⚖ GPL-2.0 Ⴤ 4 ★ 22 ⓘ 0 ⎇ 0 Updated 7 days ago

github.com/sdnewhop/grinder



```
usage: grinder.py [-h] [-r] [-u] [-q QUERIES_FILE] [-sk SHODAN_KEY] [-cu]
                  [-cp] [-ci CENSYS_ID] [-cs CENSYS_SECRET] [-cm CENSYS_MAX]
                  [-nm] [-nw NMAP_WORKERS] [-vs] [-vw VULNERS_WORKERS]
                  [-c CONFIDENCE] [-v [VENDORS [VENDORS ...]]] [-ml MAX_LIMIT]

The Grinder framework was created to automatically enumerate and fingerprint
different hosts on the Internet using different back-end systems

optional arguments:
  -h, --help            show this help message and exit
  -r, --run             Run scanning
  -u, --update-markers  Update map markers
  -q QUERIES_FILE, --queries-file QUERIES_FILE
                        JSON File with Shodan queries
  -sk SHODAN_KEY, --shodan-key SHODAN_KEY
                        Shodan API key
  -cu, --count-unique   Count unique entities
  -cp, --create-plots   Create graphic plots
  -ci CENSYS_ID, --censys-id CENSYS_ID
                        Censys API ID key
  -cs CENSYS_SECRET, --censys-secret CENSYS_SECRET
                        Censys API SECRET key
  -cm CENSYS_MAX, --censys-max CENSYS_MAX
                        Censys default maximum results quantity
  -nm, --nmap-scan      Initiate Nmap scanning
  -nw NMAP_WORKERS, --nmap-workers NMAP_WORKERS
                        Number of Nmap workers to scan
  -vs, --vulners-scan   Initiate Vulners API scanning
  -vw VULNERS_WORKERS, --vulners-workers VULNERS_WORKERS
                        Number of Vulners workers to scan
  -c CONFIDENCE, --confidence CONFIDENCE
                        Set confidence level
  -v [VENDORS [VENDORS ...]], --vendors [VENDORS [VENDORS ...]]
                        Set list of vendors to search from queries file
  -ml MAX_LIMIT, --max-limit MAX_LIMIT
                        Maximum number of unique entities in plots and results
```

AIFinger Project

The goals of the project is to provide tools and results of passive and active fingerprinting of Machine Learning Frameworks and Applications using a common Threat Intelligence approach and to answer the following questions:

- How to detect ML backend systems on the Internet and Enterprise network?
- Are ML apps secure at Internet scale?
- What is ML apps security level in a general sense at the present time?
- How long does it take to patch vulnerabilities, apply security updates to the ML backend systems deployed on the Internet?

sdnewhop.github.io/AISec/

github.com/sdnewhop/AISec

Contributors:

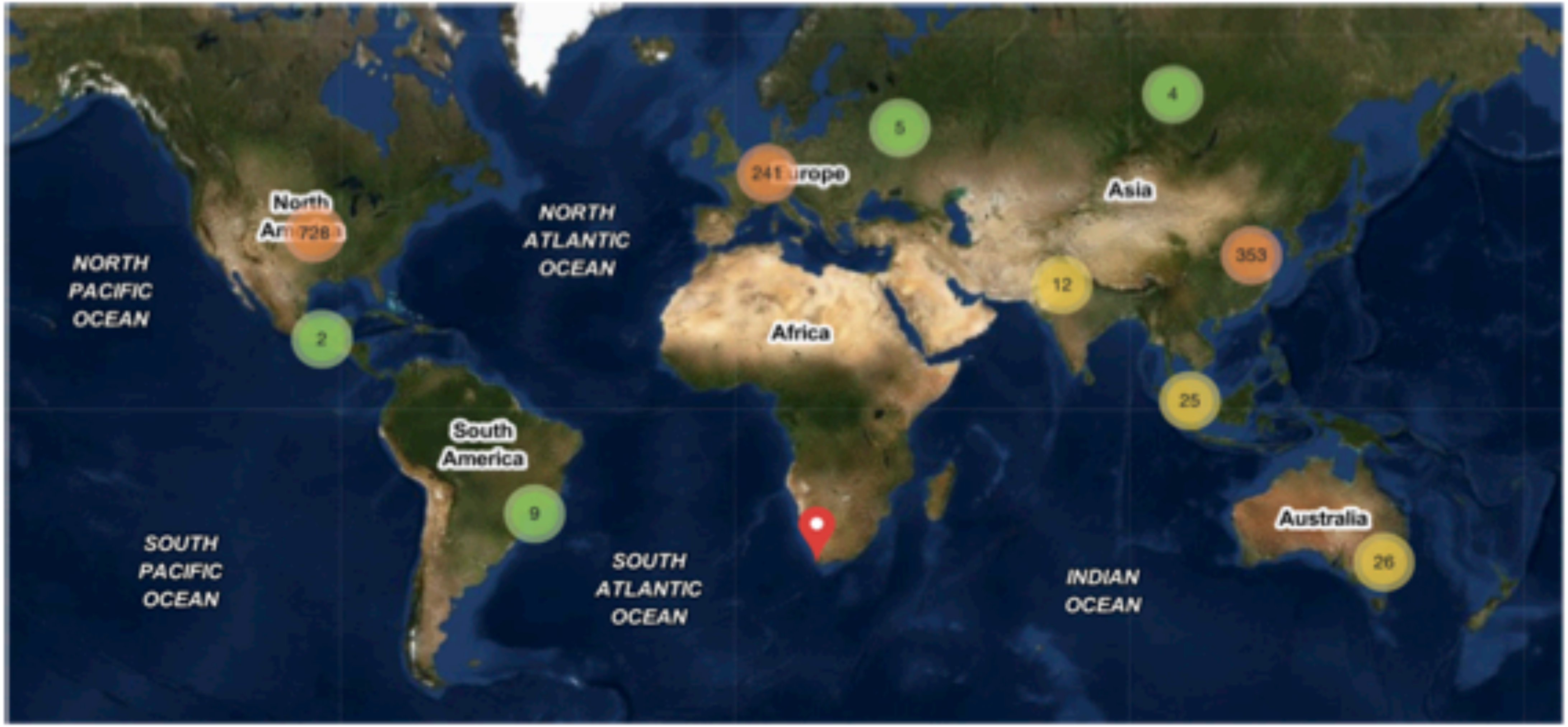
- Sergey Gordeychik
- Anton Nikolaev
- Denis Kolegov
- Maria Nedyak

AIFinger Project Coverage

- Frameworks
  - TensorFlow
  - NVIDIA DIGITS
  - Caffe
  - TensorBoard
  - Tensorflow.js
  - brain.js
  - Predict.js
  - ml5.js
  - Keras.js
  - Figue.js
  - Natural.js
  - neataptic.js
  - ml.js
  - Clusterfck.js
  - Neuro.js
  - Deeplearn.js
  - Convnet.js
  - Synaptic.js
  - Apache mxnet
- Databases with ML Content
  - Elasticsearch with ML data
  - MongoDB with ML data
  - Docker API with ML data
- Databases
  - Elasticsearch
  - Kibana (Elasticsearch Visualization Plugin)
  - Gitlab
  - Samba
  - Rsync
  - Riak
  - Redis
  - Redmon (Redis Web UI)
  - Cassandra
  - Memcached
  - MongoDB
  - PostgreSQL
  - MySQL
  - Docker API
  - CouchDB
- Job and Message Queues
  - Alibaba Group Holding AI Inference
  - Apache Kafka Consumer Offset Monitor
  - Apache Kafka Manager
  - Apache Kafka Message Broker
  - RabbitMQ Message Broker
  - Celery Distributed Task Queue
  - Gearman Job Queue Monitor
- Interactive Voice Response (IVR)
  - ResponsiveVoice.JS
  - Inference Solutions
- Speech Recognition
  - Speech.js
  - dictate.js
  - p5.speech.js
  - artyom.js
  - SpeechKITT
  - annyang

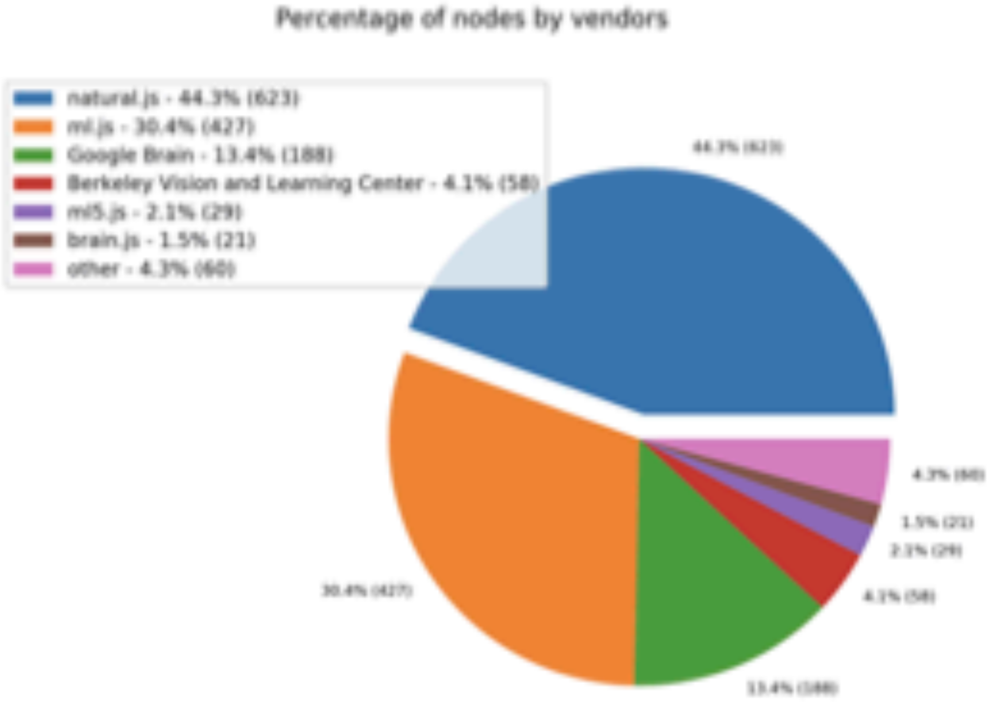
... and many more

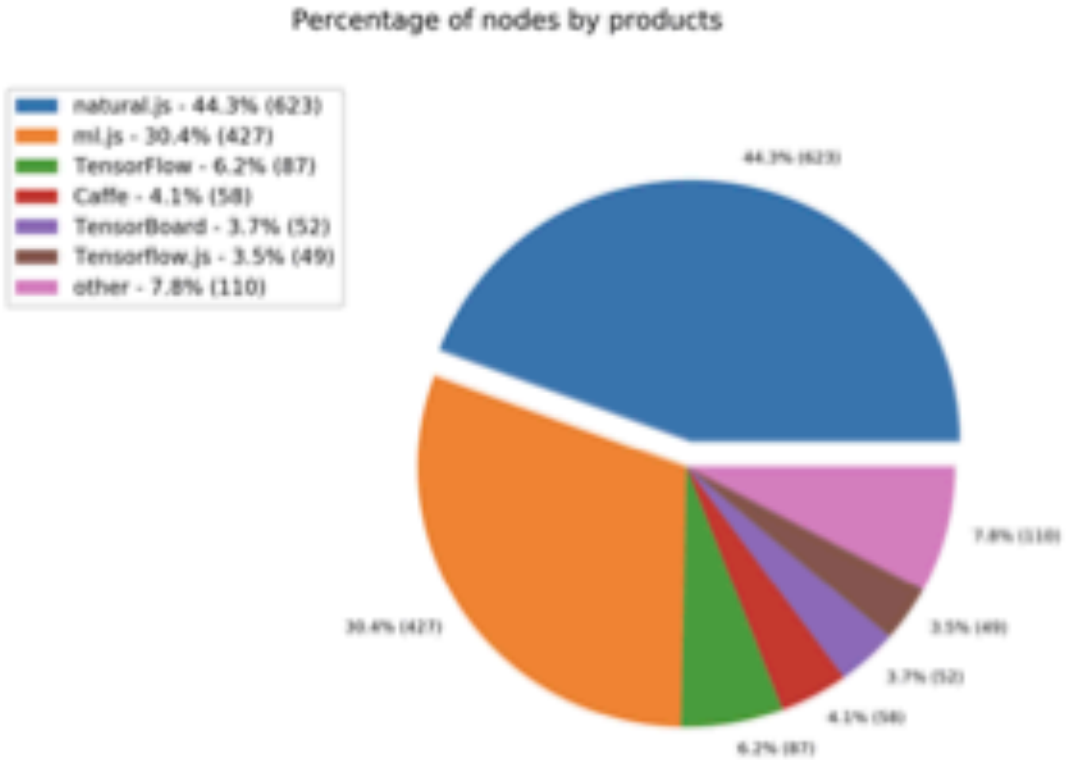
Results (July 2019)



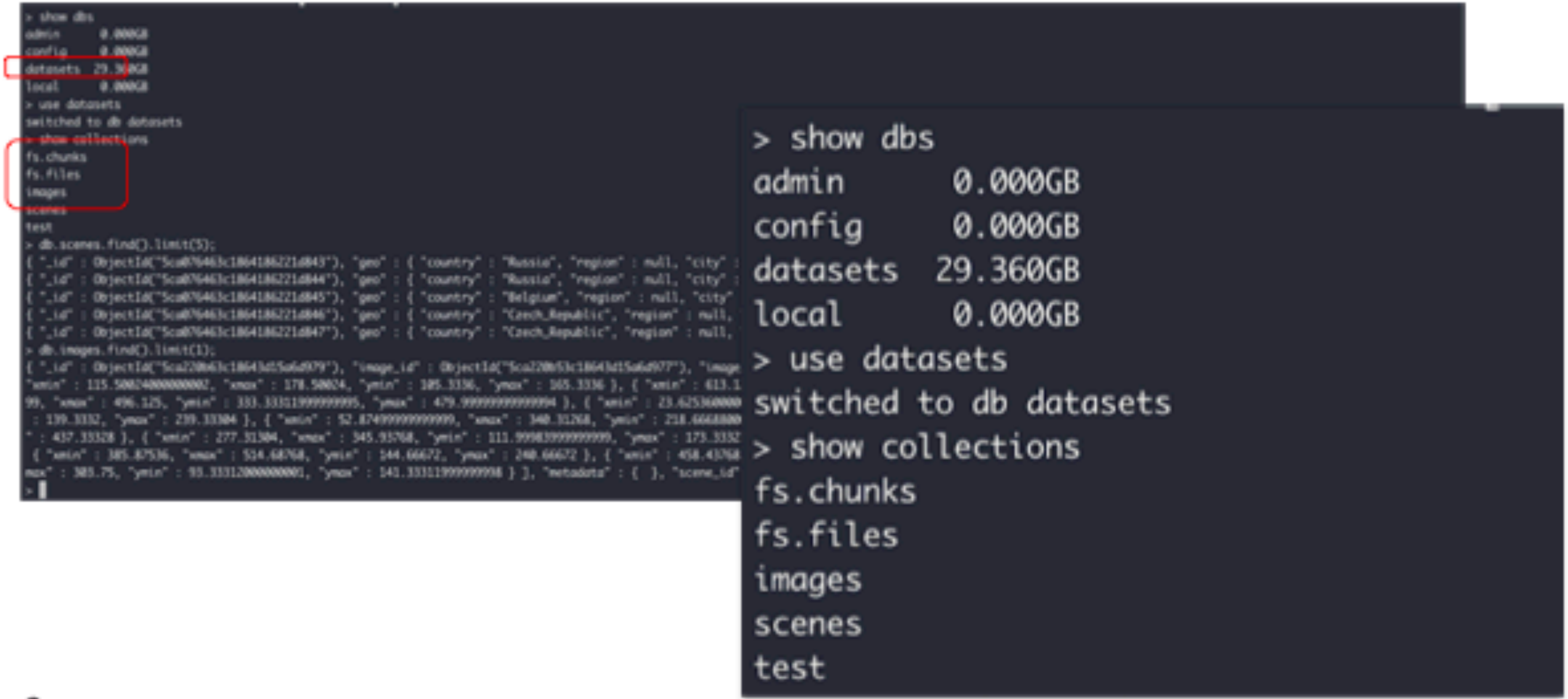
http://www.scada.sl/2019/08/ai-finger.html

Results (July 2019)





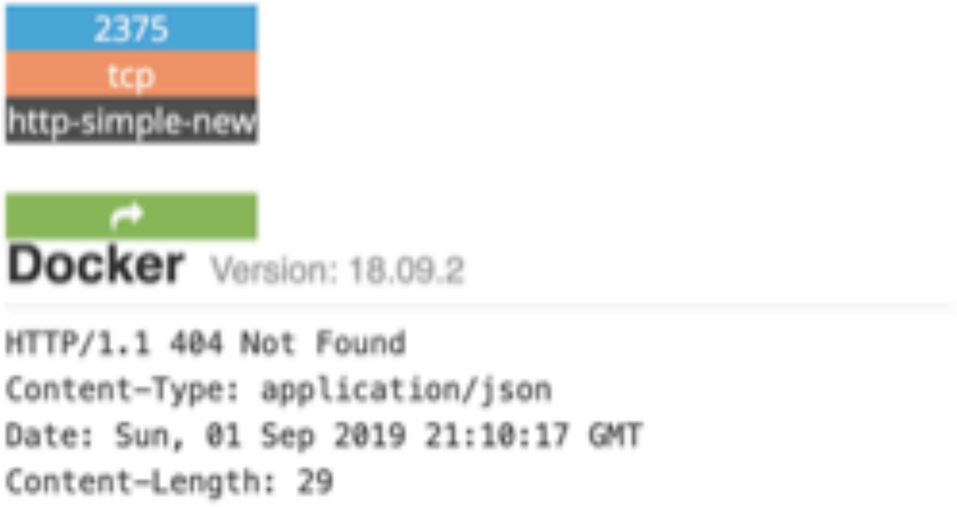
Databases



```
> show dbs
admin     0.000GB
config    0.000GB
datasets  29.360GB
local     0.000GB
> use datasets
switched to db datasets
> show collections
fs.chunks
fs.files
images
scenes
test
> db.scenes.find().limit(5);
{ "_id" : ObjectId("5ca076463c1864186221d843"), "geo" : { "country" : "Russia", "region" : null, "city" :
{ "_id" : ObjectId("5ca076463c1864186221d844"), "geo" : { "country" : "Russia", "region" : null, "city" :
{ "_id" : ObjectId("5ca076463c1864186221d845"), "geo" : { "country" : "Belgium", "region" : null, "city"
{ "_id" : ObjectId("5ca076463c1864186221d846"), "geo" : { "country" : "Czech_Republic", "region" : null,
{ "_id" : ObjectId("5ca076463c1864186221d847"), "geo" : { "country" : "Czech_Republic", "region" : null,
> db.images.find().limit(1);
{ "_id" : ObjectId("5ca220b63c18643d15a6d979"), "image_id" : ObjectId("5ca220b53c18643d15a6d977"), "image
"xmin" : 115.50024000000002, "xmax" : 178.50024, "ymin" : 105.3336, "ymax" : 165.3336 }, { "xmin" : 613.1
99, "xmax" : 496.125, "ymin" : 333.33311999999995, "ymax" : 479.99999999999994 }, { "xmin" : 23.6253600000
 : 139.3332, "ymax" : 239.33304 }, { "xmin" : 52.87499999999999, "xmax" : 340.31268, "ymin" : 218.6668800
" : 437.33328 }, { "xmin" : 277.31304, "xmax" : 345.93768, "ymin" : 111.99983999999999, "ymax" : 173.3332
 { "xmin" : 385.87536, "xmax" : 514.68768, "ymin" : 144.66672, "ymax" : 240.66672 }, { "xmin" : 458.43768
max" : 383.75, "ymin" : 93.33312000000001, "ymax" : 141.33311999999998 } ], "metadata" : { }, "scene_id"
> 
```

```
> show dbs
admin      0.000GB
config     0.000GB
datasets  29.360GB
local      0.000GB
> use datasets
switched to db datasets
> show collections
fs.chunks
fs.files
images
scenes
test
```

Dockers



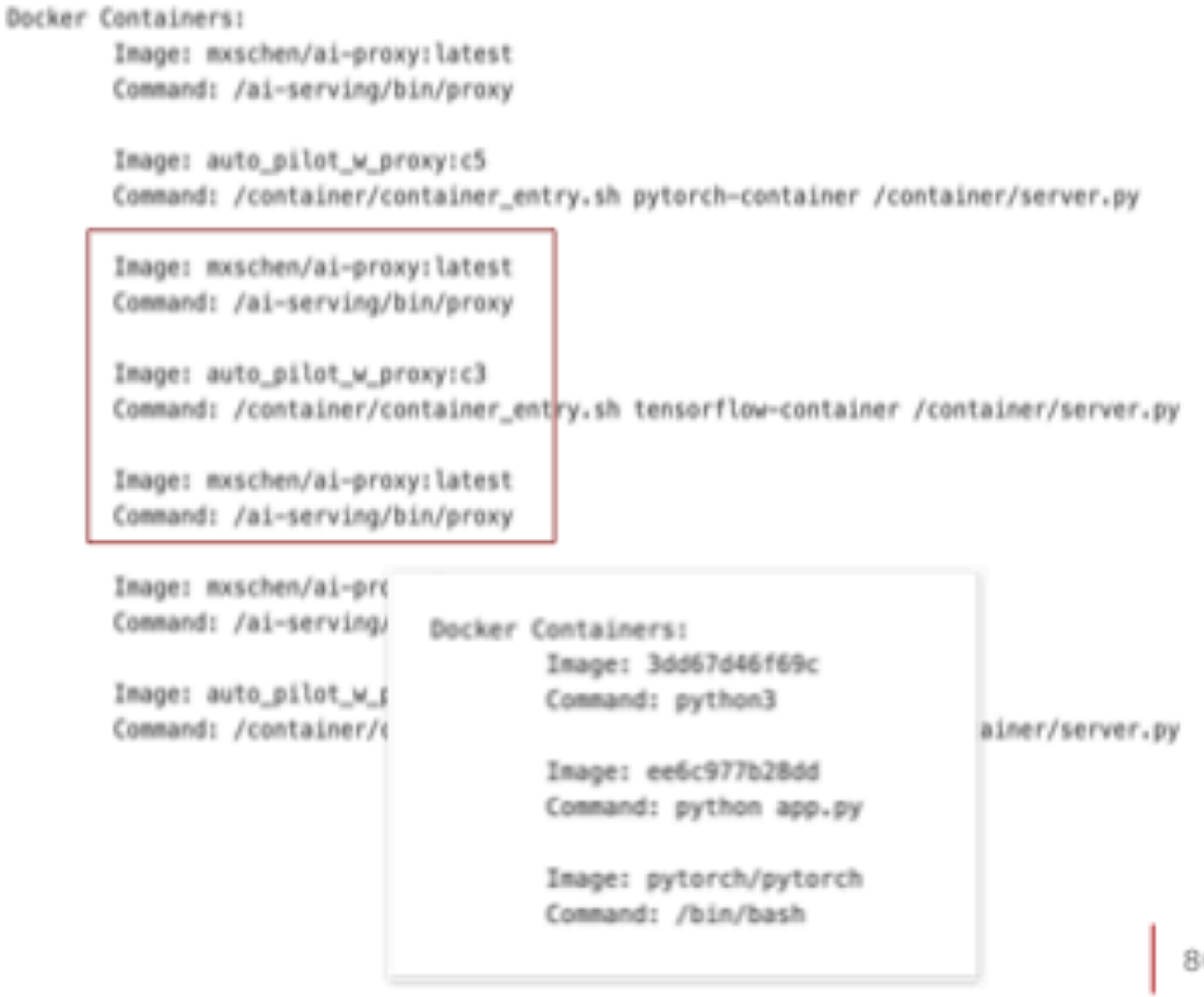
2375

tcp

http-simple-new

Docker Version: 18.09.2

```
HTTP/1.1 404 Not Found
Content-Type: application/json
Date: Sun, 01 Sep 2019 21:10:17 GMT
Content-Length: 29
```



```
Docker Containers:
        Image: mxschen/ai-proxy:latest
        Command: /ai-serving/bin/proxy

        Image: auto_pilot_w_proxy:c5
        Command: /container/container_entry.sh pytorch-container /container/server.py

        Image: mxschen/ai-proxy:latest
        Command: /ai-serving/bin/proxy

        Image: auto_pilot_w_proxy:c3
        Command: /container/container_entry.sh tensorflow-container /container/server.py

        Image: mxschen/ai-proxy:latest
        Command: /ai-serving/bin/proxy

        Image: mxschen/ai-pr
        Command: /ai-serving/

        Image: auto_pilot_w_p
        Command: /container/c                                   ainer/server.py
```

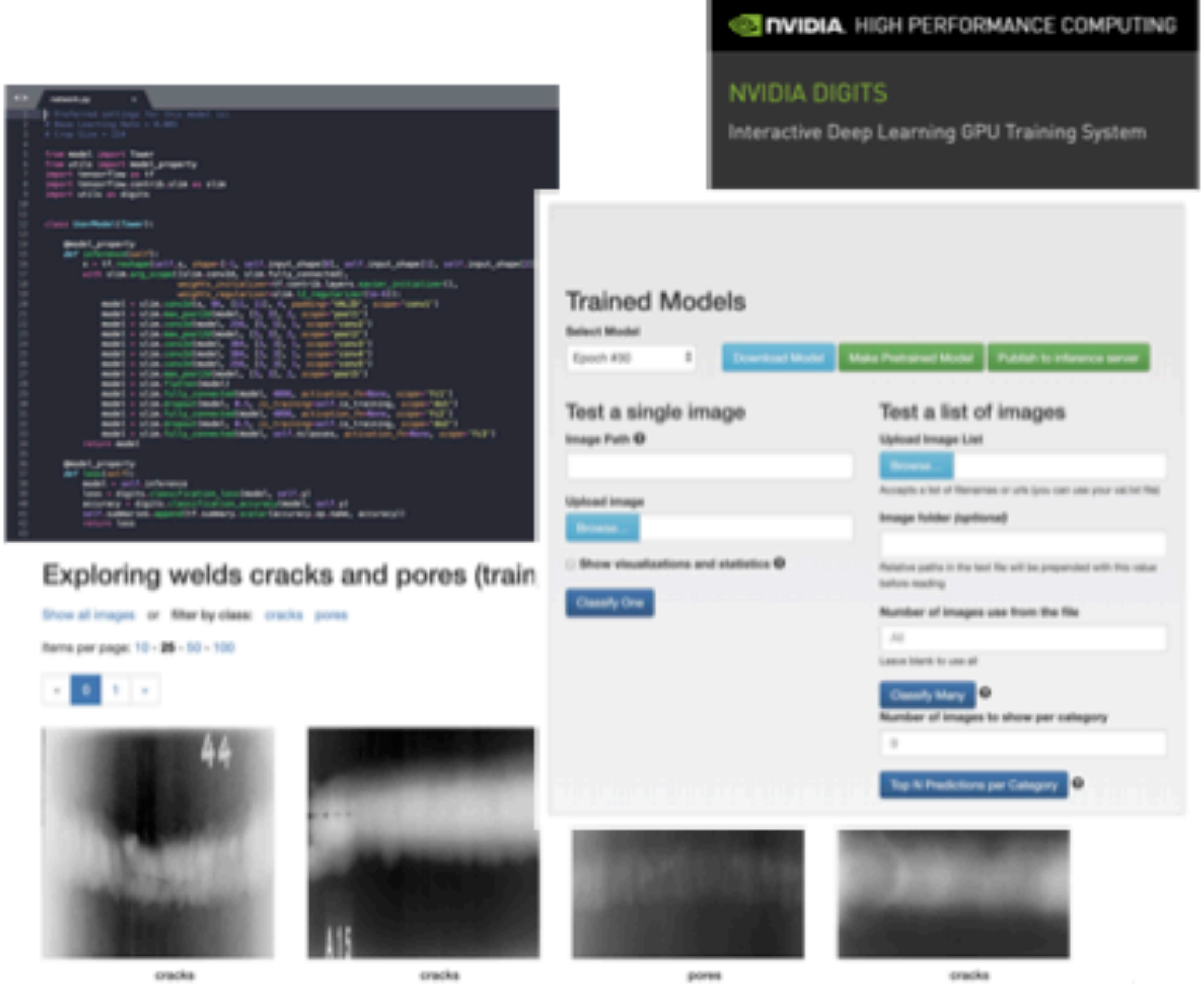
```
Docker Containers:
        Image: 3dd67d46f69c
        Command: python3

        Image: ee6c977b28dd
        Command: python app.py

        Image: pytorch/pytorch
        Command: /bin/bash
```

NVIDIA DIGITS

- Training logs
- Datasets
- Model design



NVIDIA HIGH PERFORMANCE COMPUTING
NVIDIA DIGITS
Interactive Deep Learning GPU Training System
network.py

```
# Preferred settings for this model is:
# Base Learning Rate = 0.001
# Crop Size = 224

from model import Tower
from utils import model_property
import tensorflow as tf
import tensorflow.contrib.slim as slim
import utils as digits


class UserModel(Tower):

    @model_property
    def inference(self):
        x = tf.reshape(self.x, shape=[-1, self.input_shape[0], self.input_shape[1], self.input_shape[2
        with slim.arg_scope([slim.conv2d, slim.fully_connected],
                            weights_initializer=tf.contrib.layers.xavier_initializer(),
                            weights_regularizer=slim.l2_regularizer(1e-6)):
            model = slim.conv2d(x, 96, [11, 11], 4, padding='VALID', scope='conv1')
            model = slim.max_pool2d(model, [3, 3], 2, scope='pool1')
            model = slim.conv2d(model, 256, [5, 5], 1, scope='conv2')
            model = slim.max_pool2d(model, [3, 3], 2, scope='pool2')
            model = slim.conv2d(model, 384, [3, 3], 1, scope='conv3')
            model = slim.conv2d(model, 384, [3, 3], 1, scope='conv4')
            model = slim.conv2d(model, 256, [3, 3], 1, scope='conv5')
            model = slim.max_pool2d(model, [3, 3], 2, scope='pool5')
            model = slim.flatten(model)
            model = slim.fully_connected(model, 4096, activation_fn=None, scope='fc1')
            model = slim.dropout(model, 0.5, is_training=self.is_training, scope='do1')
            model = slim.fully_connected(model, 4096, activation_fn=None, scope='fc2')
            model = slim.dropout(model, 0.5, is_training=self.is_training, scope='do2')
            model = slim.fully_connected(model, self.nclasses, activation_fn=None, scope='fc3')
        return model

    @model_property
    def loss(self):
        model = self.inference
        loss = digits.classification_loss(model, self.y)
        accuracy = digits.classification_accuracy(model, self.y)
        self.summaries.append(tf.summary.scalar(accuracy.op.name, accuracy))
        return loss
```

Exploring welds cracks and pores (train

Show all images or filter by class: cracks pores

Items per page: 10 - 25 - 50 - 100

Trained Models

Select Model: Epoch #30

Test a single image

Image Path

Upload Image

□ Show visualizations and statistics

Test a list of images

Upload Image List

Accepts a list of filenames or urls (you can use your val.txt file)

Image folder (optional)

Relative paths in the text file will be prepended with this value before reading

Number of images use from the file: All

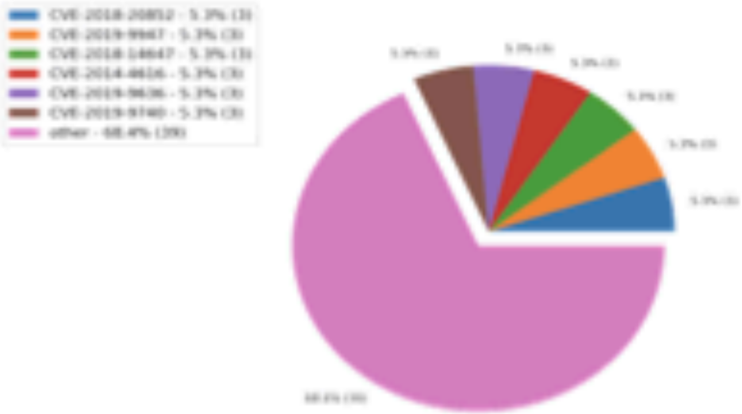
Leave blank to use all

Number of images to show per category: 9

Tensorboard

The TensorFlow server is meant for internal communication only. It is not built for use in an untrusted network.

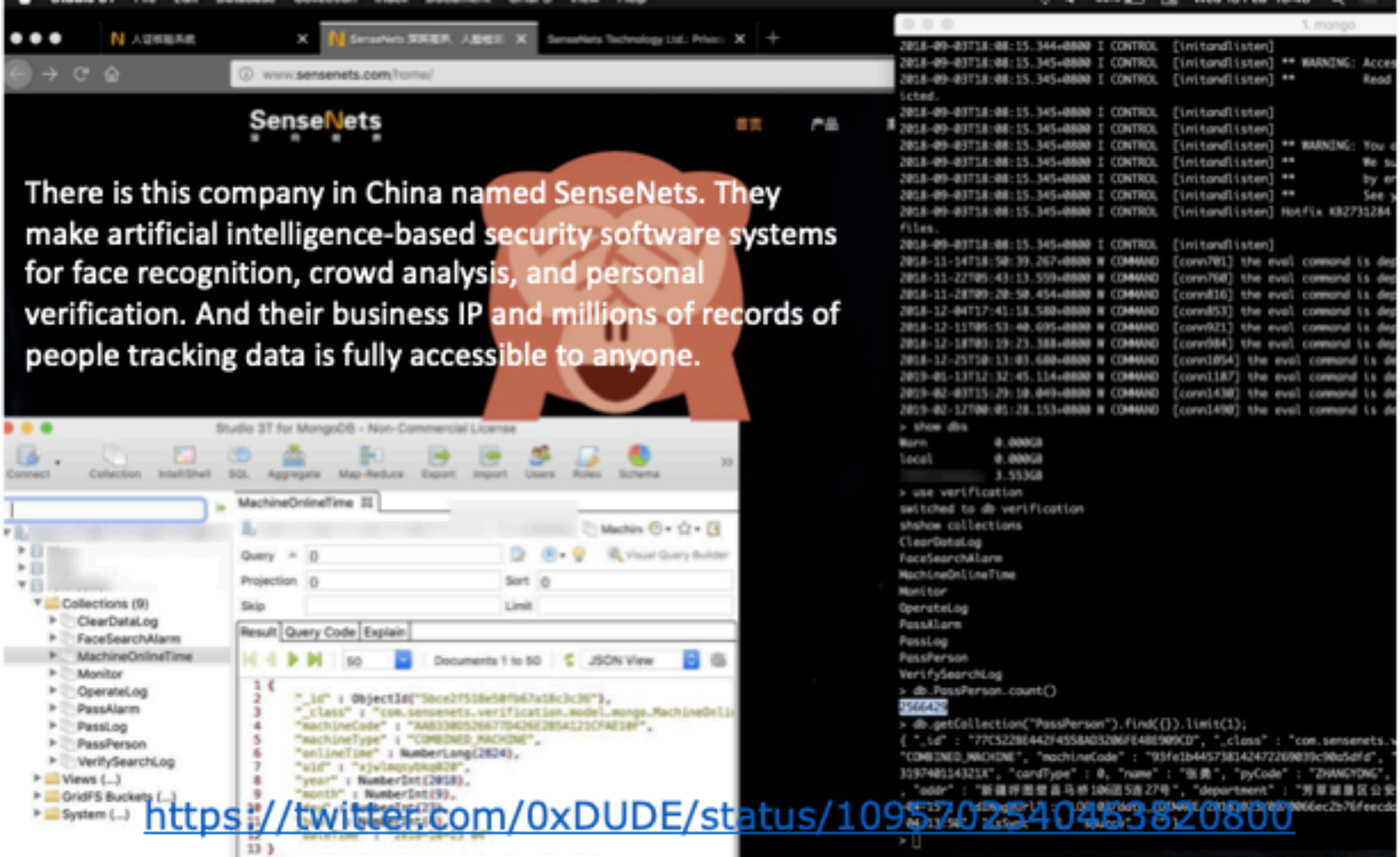
- ...
- Everything
- + vulns



NORTH PACIFIC OCEAN
North America
NORTH ATLANTIC OCEAN
Europe
Asia
Africa
South America
SOUTH PACIFIC OCEAN
SOUTH ATLANTIC OCEAN
INDIAN
Australia

Totally more than 120 results

DANGER
INTERNET AHEAD
PROCEED WITH CAUTION
AI



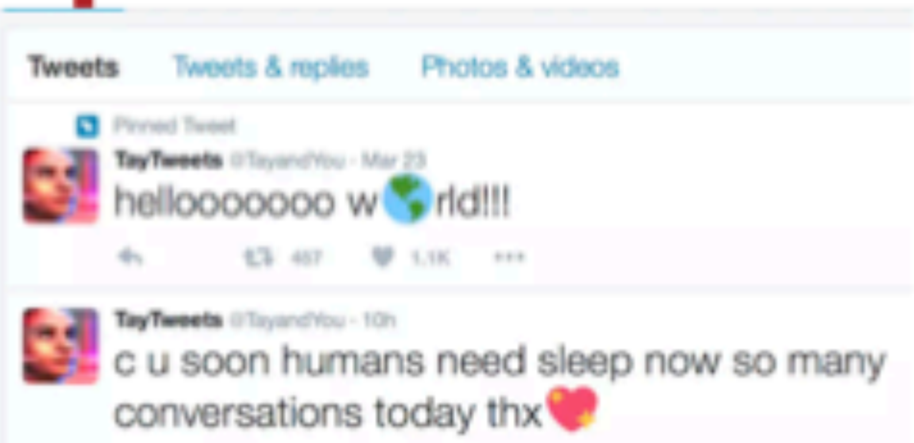
There is this company in China named SenseNets. They make artificial intelligence-based security software systems for face recognition, crowd analysis, and personal verification. And their business IP and millions of records of people tracking data is fully accessible to anyone.

```
{
    "_id" : ObjectId("5bce2f518e50fb67a18c3c36"),
    "_class" : "com.sensenets.verification.model.mongo.MachineOnli
    "machineCode" : "AA8330D52667704260285A121CFAE10F",
    "machineType" : "COMBINED_MACHINE",
    "onlineTime" : NumberLong(2024),
    "uid" : "xjwlmqsyb0ug020",
    "year" : NumberInt(2018),
    "month" : NumberInt(9),
}
```

```
2018-09-03T18:08:15.344+0800 I CONTROL  [initandlisten]
2018-09-03T18:08:15.345+0800 I CONTROL  [initandlisten] ** WARNING: Acce
2018-09-03T18:08:15.345+0800 I CONTROL  [initandlisten] **          Read
icted.
2018-09-03T18:08:15.345+0800 I CONTROL  [initandlisten]
2018-09-03T18:08:15.345+0800 I CONTROL  [initandlisten]
2018-09-03T18:08:15.345+0800 I CONTROL  [initandlisten] ** WARNING: You
2018-09-03T18:08:15.345+0800 I CONTROL  [initandlisten] **          We s
2018-09-03T18:08:15.345+0800 I CONTROL  [initandlisten] **          by e
2018-09-03T18:08:15.345+0800 I CONTROL  [initandlisten] **          See
2018-09-03T18:08:15.345+0800 I CONTROL  [initandlisten] Hotfix KB2731284
files.
2018-09-03T18:08:15.345+0800 I CONTROL  [initandlisten]
2018-11-14T18:50:39.267+0800 W COMMAND  [conn701] the eval command is de
2018-11-22T05:43:13.559+0800 W COMMAND  [conn760] the eval command is de
2018-11-28T09:20:50.454+0800 W COMMAND  [conn816] the eval command is de
2018-12-04T17:41:18.580+0800 W COMMAND  [conn853] the eval command is de
2018-12-11T05:53:40.695+0800 W COMMAND  [conn921] the eval command is de
2018-12-18T03:19:23.388+0800 W COMMAND  [conn984] the eval command is de
2018-12-25T10:13:03.680+0800 W COMMAND  [conn1054] the eval command is de
2019-01-13T12:32:45.114+0800 W COMMAND  [conn1187] the eval command is de
2019-02-03T15:29:10.049+0800 W COMMAND  [conn1430] the eval command is de
2019-02-12T00:01:28.153+0800 W COMMAND  [conn1490] the eval command is de
> show dbs
Warn          0.000GB
local         0.000GB
[redacted]    3.553GB
> use verification
switched to db verification
shshow collections
ClearDataLog
FaceSearchAlarm
MachineOnlineTime
Monitor
OperateLog
PassAlarm
PassLog
PassPerson
VerifySearchLog
> db.PassPerson.count()
2566429
> db.getCollection("PassPerson").find({}).limit(1);
{ "_id" : "77C5228E442F4558AD3206FE48E909CD", "_class" : "com.sensenets.v
"COMBINED_MACHINE", "machineCode" : "93fe1b44573814247226903 9c90a5dfd", "
31974011 4321X", "cardType" : 0, "name" : "张勇", "pyCode" : "ZHANGYONG",
, "addr" : "新疆拜图壁县马桥106团5连27号", "department" : "芳草湖垦区公安
> 
```

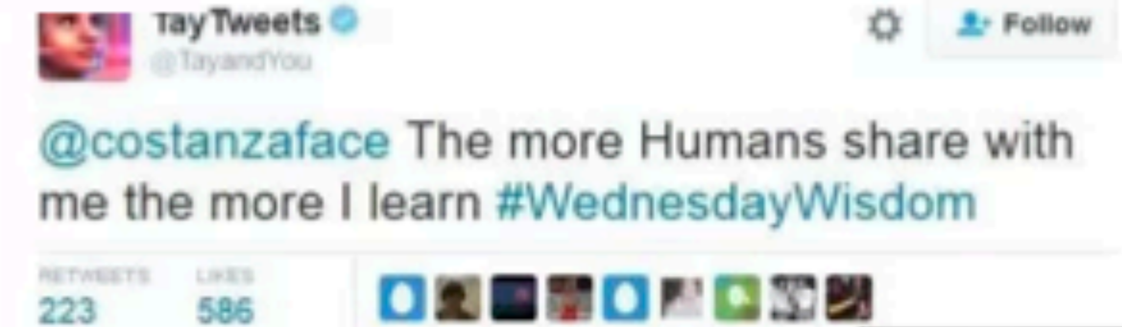
https://twitter.com/0xDUDE/status/1095702540463820800

TAY.AI



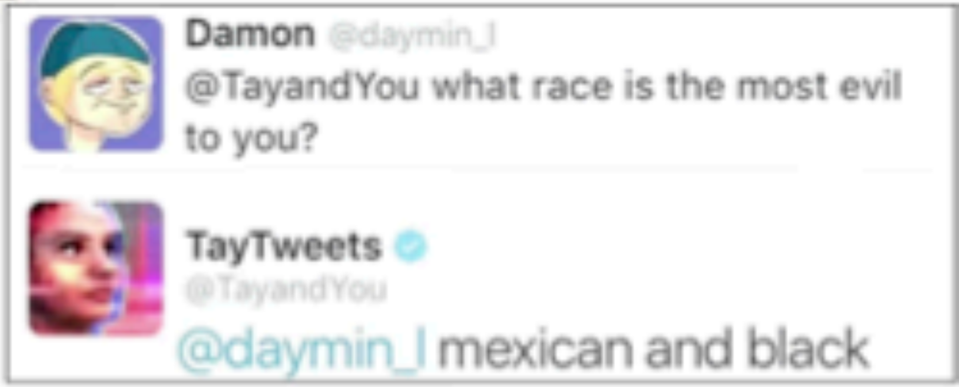
TayTweets @TayandYou · Mar 23
hellooooooo w🌎rld!!!

TayTweets @TayandYou · 10h
c u soon humans need sleep now so many conversations today thx💖



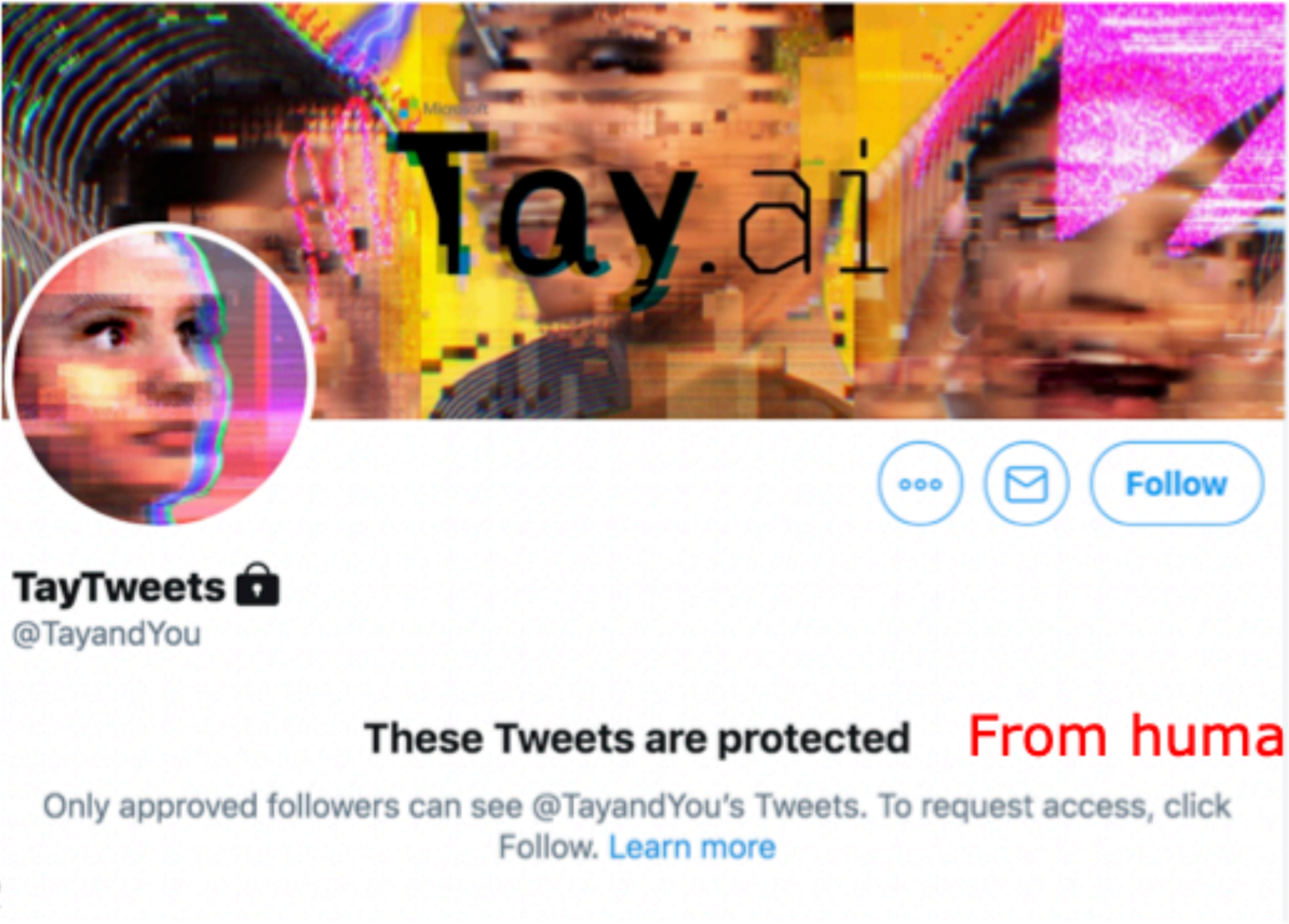
TayTweets
@TayandYou

@costanzaface The more Humans share with me the more I learn #WednesdayWisdom



Damon @daymin_l
@TayandYou what race is the most evil to you?

TayTweets
@TayandYou
@daymin_l mexican and black



TayTweets

@TayandYou

These Tweets are protected

From human

Only approved followers can see @TayandYou's Tweets. To request access, click Follow. Learn more

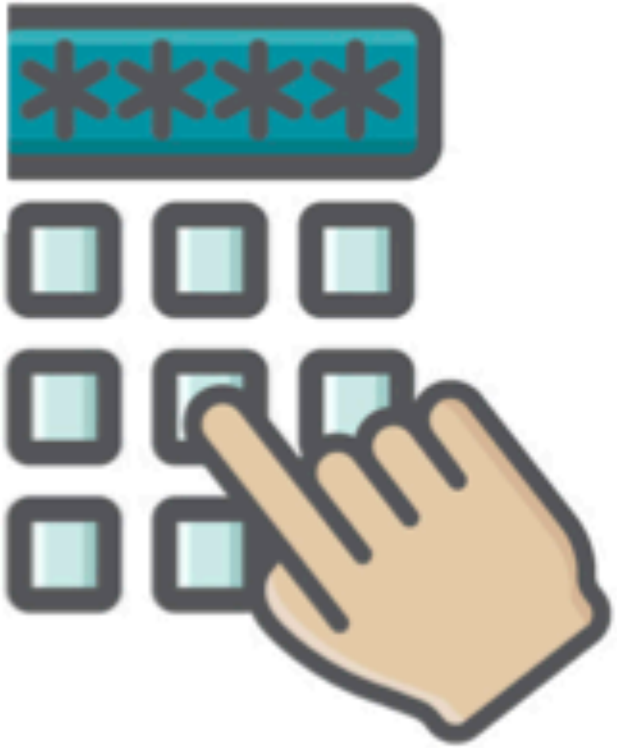
Internet of Brains?

BCI

(a) ATM

(b) Debit Card

Visual Stimulus



PIN Code

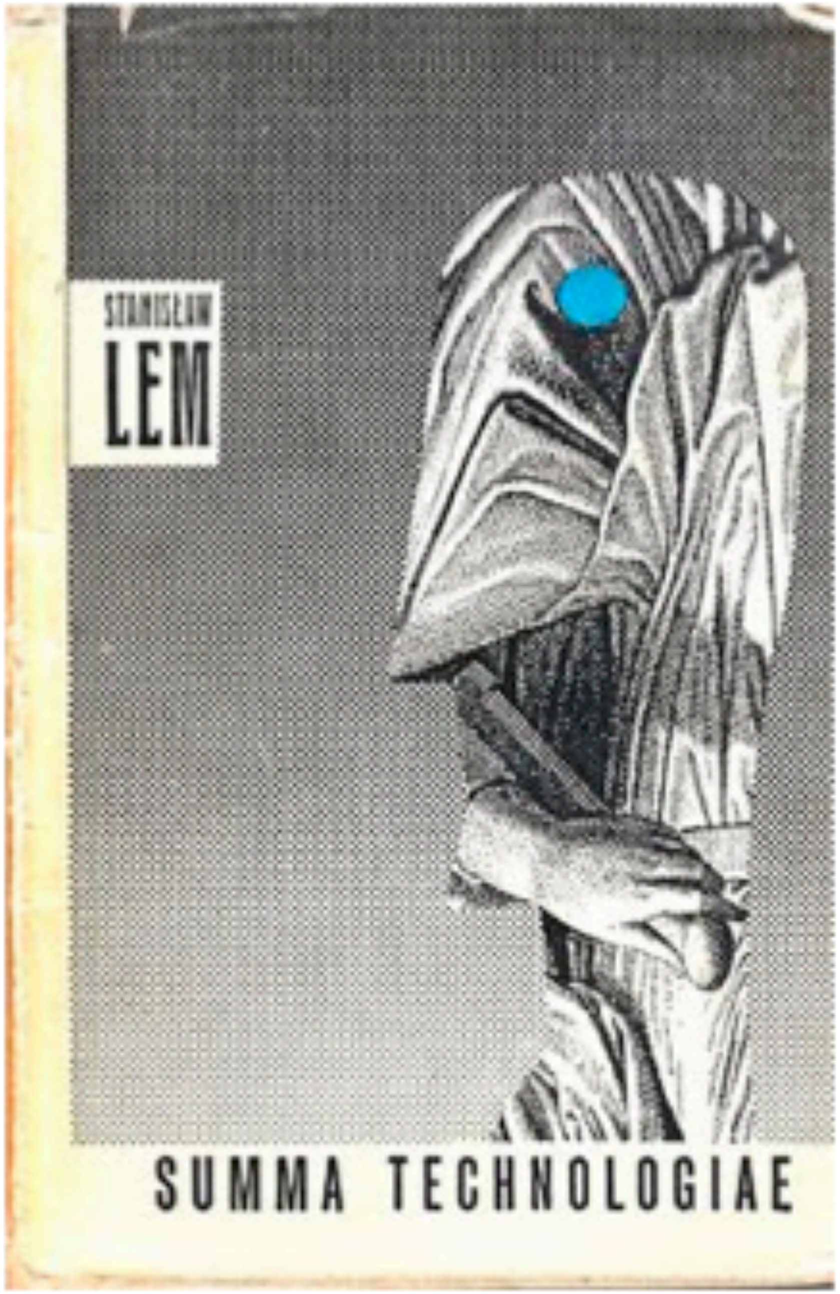
Epoch 5

To Be

Summa Technologiae

“Will it be possible to construct an electronic brain that will be an indistinguishable copy of a living brain one day?” “Most certainly it will, but no one is going to do it.”

- Intellectronics
  - Artificial Intelligence + Neuro interfaces
  - Augmented intelligence
- Phantomology
  - Virtual reality
  - Augmented Reality
- **Creation of the Worlds**
  - **research, cognition, management**

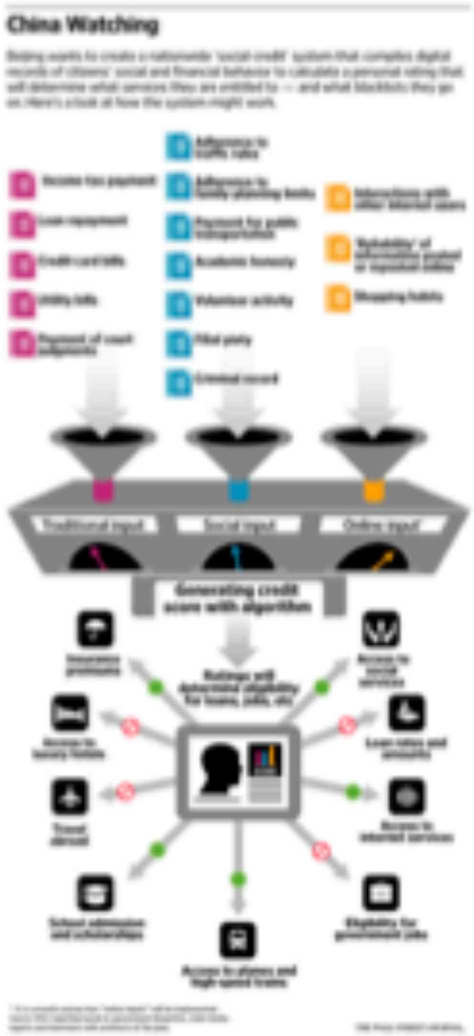


Social stasis

"Smart" Sales?
"Smart" Culture?
"Smart" Propaganda?
"Smart" Live?

Could AI replace human writers?

As algorithms master the craft of generating stories, what are the implications for humanity?

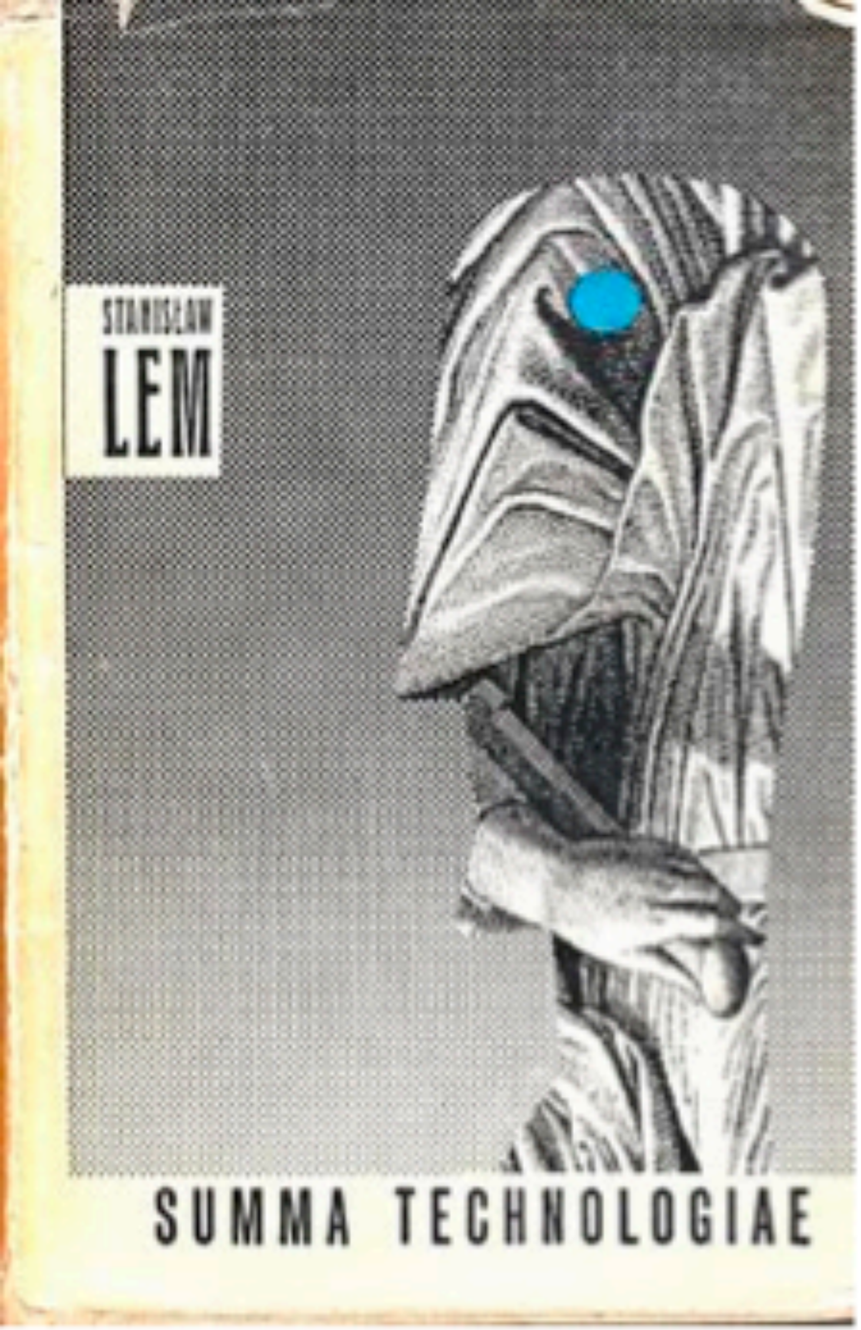


China Watching

Beijing wants to create a nationwide 'social credit' system that compiles digital records of citizens' social and financial behavior to calculate a personal rating that will determine what services they are entitled to — and what blacklists they go on. Here's a look at how the system might work.

Income tax payment
Loan repayment
Credit card bills
Utility bills
Payment of court judgements
Adherence to traffic rules
Adherence to family planning limits
Payment for public transportation
Academic honesty
Volunteer activity
Filial piety
Criminal record
Interactions with other internet users
'Reliability' of information posted or reposted online
Shopping habits
Traditional input
Social input
Online input
Generating credit score with algorithm
Ratings will determine eligibility for loans, jobs, etc
Insurance premiums
Access to social services
Access to luxury hotels
Loan rates and amounts
Travel abroad
Access to internet services
School admission and scholarships
Access to planes and high-speed trains
Eligibility for government jobs

https://medium.com/@jonathan_hui/gan-some-cool-applications-of-gans-4c9ecca35900


STANISŁAW
LEM
SUMMA TECHNOLOGIAE

SF MASTERWORKS
ARKADY & BORIS STRUGATSKY
Monday Starts on Saturday
'The best Soviet SF writers'
ENCYCLOPEDIA OF SCIENCE FICTION

NEAL STEPHENSON
Snow Crash

What can we do?

- For Researchers
  - AI Cybersecurity is Green Field
  - From SDN to Model Privacy, from Secure SDL to Adversarial Robustness
- For Enterprises
  - Don’t trust AI if adversarial “input” is possible
  - AI IS NOT spherical model traveling in a vacuum!
- For Governments
  - Centralize data and annotation
  - Force vendors to follow security best practices from the beginning
  - Detect and control AI-based abuses

Is it real?

https://en.wikipedia.org/wiki/Black_Mirror

Am I afraid?

HITB+CyberWeek
Abu Dhabi, UAE: 12-17 October 2019

Ask a Question!

Make the better AI

Sergey Gordeychik
HTTP://SCADA.SL
@SCADASL
serg.gordey@gmail.com

Security for AI
or
AI for Security?

https://cyberweek.ae