

RSA®Conference2020

San Francisco | February 24 – 28 | Moscone Center

HUMAN
ELEMENT

SESSION ID: MLAI1-F01

Augmenting Intelligence: Machines as Super Assistants for Security Experts



Matteo Dell'Amico

Senior Principal Researcher
NortonLifeLock

#RSAC

Thanks To



Thanks for the many discussions and brainstormings inside and outside of the NortonLifeLock Research Group (NRG)!

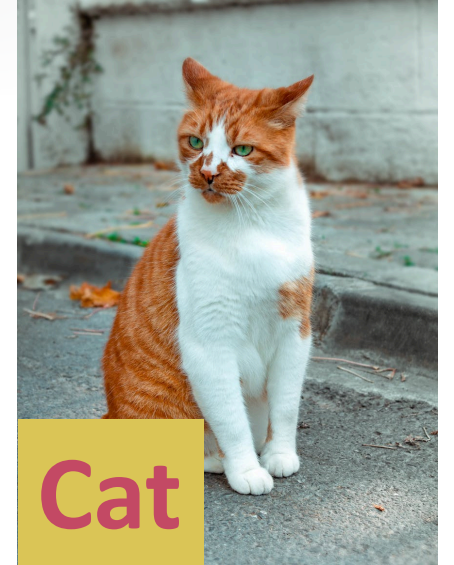
RSA®Conference2020

Why ML For Security Is Difficult


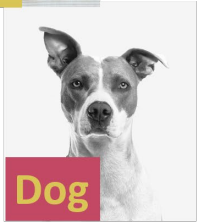
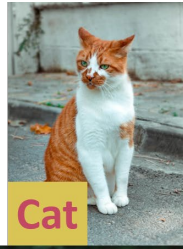
Can Machine Learning Make Security Experts Obsolete?

- We know automation can replace several jobs now
- Computers are much better than humans at some things we need
 - Handling very large datasets
 - Complex statistics
- Will we have electronic brains that can discern what's malicious and block it?
 - My guess is **not yet**, and possibly not for a long time...

A Typical Classifier (1): Start With a Dataset



A Typical Classifier (2): Feature Extraction



0.42	1.25	3.15	2.26	1.47	6.54	2.14	Cat
1.13	1.67	0.02	0.66	1.55	3.12	5.41	Dog
5.14	2.00	0.47	0.22	1.68	2.17	3.55	Dog
0.12	5.13	4.22	2.14	1.47	5.55	6.42	Cat
0.03	5.42	1.68	3.33	2.13	6.02	1.33	Cat
4.27	1.67	2.23	1.14	1.67	1.03	5.21	Dog

A Typical Classifier (3): Training/Test Split

0.42	1.25	3.15	2.26	1.47	6.54	2.14	Cat
1.13	1.67	0.02	0.66	1.55	3.12	5.41	Dog
5.14	2.00	0.47	0.22	1.68	2.17	3.55	Dog
0.12	5.13	4.22	2.14	1.47	5.55	6.42	Cat
0.03	5.42	1.68	3.33	2.13	6.02	1.33	Cat
4.27	1.67	2.23	1.14	1.67	1.03	5.21	Dog



Training Set

0.42	1.25	3.15	2.26	1.47	6.54	2.14	Cat
4.27	1.67	2.23	1.14	1.67	1.03	5.21	Dog
5.14	2.00	0.47	0.22	1.68	2.17	3.55	Dog
0.12	5.13	4.22	2.14	1.47	5.55	6.42	Cat

Test Set

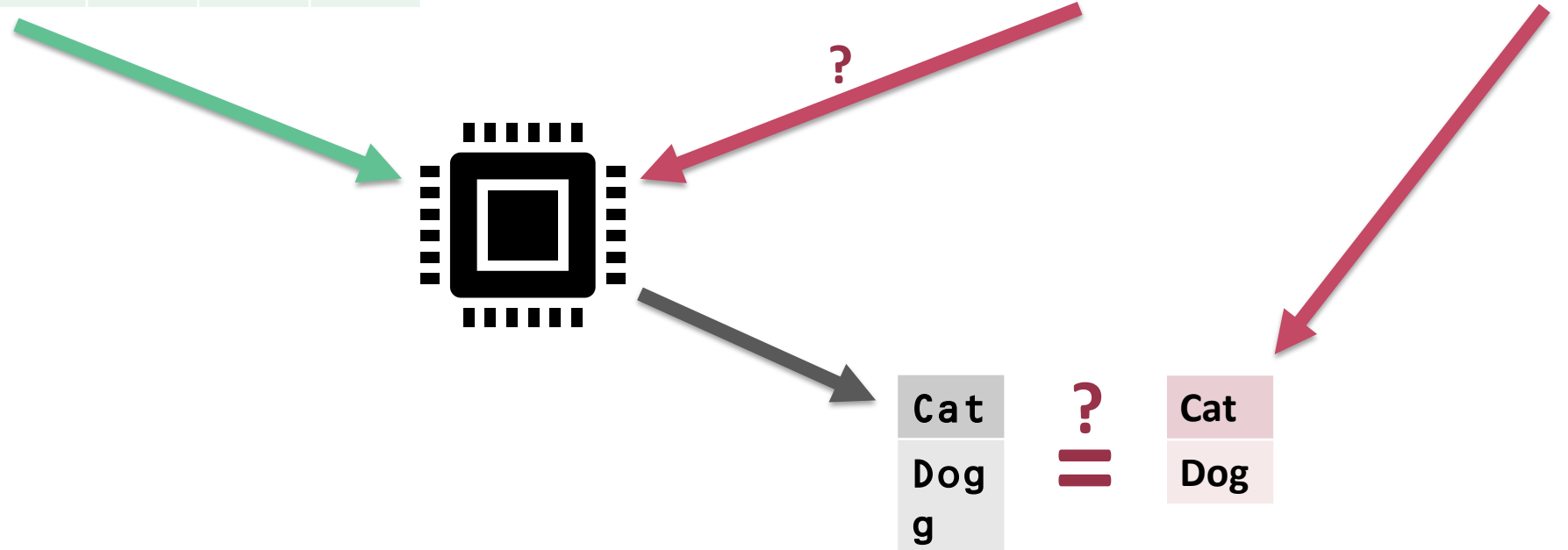


0.03	5.42	1.68	3.33	2.13	6.02	1.33	Cat
1.13	1.67	0.02	0.66	1.55	3.12	5.41	Dog

A Typical Classifier (4): Training & Validation

0.42	1.25	3.15	2.26	1.47	6.54	2.14	Cat
4.27	1.67	2.23	1.14	1.67	1.03	5.21	Dog
5.14	2.00	0.47	0.22	1.68	2.17	3.55	Dog
0.12	5.13	4.22	2.14	1.47	5.55	6.42	Cat

0.03	5.42	1.68	3.33	2.13	6.02	1.33	Cat
1.13	1.67	0.02	0.66	1.55	3.12	5.41	Dog



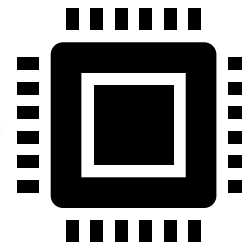
...now tune the classifier and/or repeat with another training/test split...

A Typical Classifier (5): Production Usage

0.42	1.25	3.15	2.26	1.47	6.54	2.14	Cat
1.13	1.67	0.02	0.66	1.55	3.12	5.41	Dog
5.14	2.00	0.47	0.22	1.68	2.17	3.55	Dog
0.12	5.13	4.22	2.14	1.47	5.55	6.42	Cat
0.03	5.42	1.68	3.33	2.13	6.02	1.33	Cat
4.27	1.67	2.23	1.14	1.67	1.03	5.21	Dog

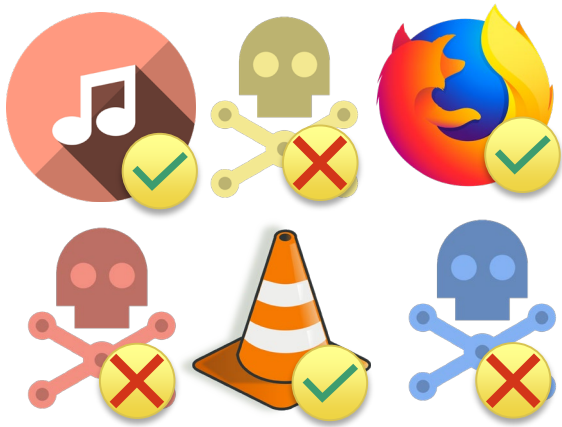


0.03	0.22	1.68	3.27	2.13	6.02	1.33
0.41	0.16	2.15	1.14	0.33	1.26	1.21
1.13	1.67	0.12	0.66	1.55	3.42	5.41



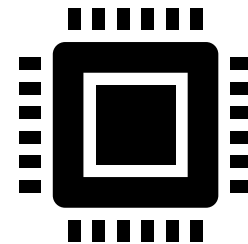
Dog
g
Dog
g
Cat

The Same Model for Security...



0.42	1.25	3.15	2.26	1.47	6.54	2.14	✓
1.13	1.67	0.02	0.66	1.55	3.12	5.41	✗
5.14	2.00	0.47	0.22	1.68	2.17	3.55	✓
0.12	5.13	4.22	2.14	1.47	5.55	6.42	✗
0.03	5.42	1.68	3.33	2.13	6.02	1.33	✓
4.27	1.67	2.23	1.14	1.67	1.03	5.21	✗

0.42	1.25	3.15	2.26	1.47	6.54	2.14	✓
1.13	1.67	0.02	0.66	1.55	3.12	5.41	✗
5.14	2.00	0.47	0.22	1.68	2.17	3.55	✓
0.12	5.13	4.22	2.14	1.47	5.55	6.42	✗



0.03	5.42	1.68	3.33	2.13	6.02	1.33
4.27	1.67	2.23	1.14	1.67	1.03	5.21



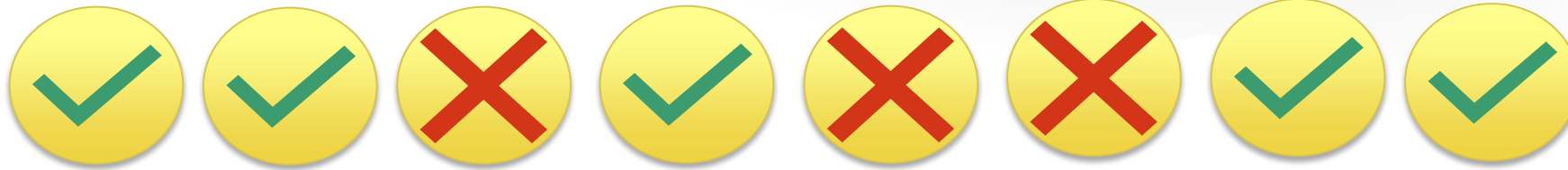
Problems in the Dataset



Filename	Date	Provenance	Hash	Signer
music.exe	2019-08-22	small.com	1b2fe93...	???????
ransom.exe	2018-04-12	???????	4b13cd3...	ShadyCo
firefox.exe	2018-11-13	mozilla.org	1f34538...	???????
exfiltrate.exe	1900-01-01	fhexus.rwt.ru	???????	--
vlc.exe	2020-02-13	???????	2a39b5...	VideoLan
firefox.exe	2016-03-03	compromised.com	36b12c...	--

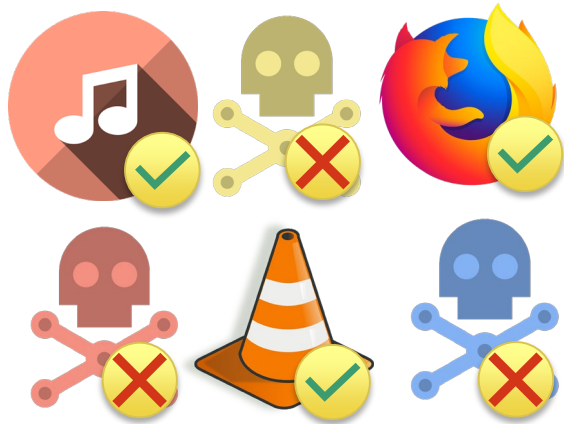
- Missing data: sensors missing or down in some time/location
- Uncertain data: some information may be incorrect
- Adversarial scenarios

Problems in the Labels



- False positives & false negatives
- Unknown unknowns
 - There may be traces of malicious behavior we don't know yet about, and we may erroneously classify it as benign...
- Defining what's malicious is a complex problem on its own
 - Depending on vantage points, something may be seen as benign or malicious

Problems in Feature Extraction (1)

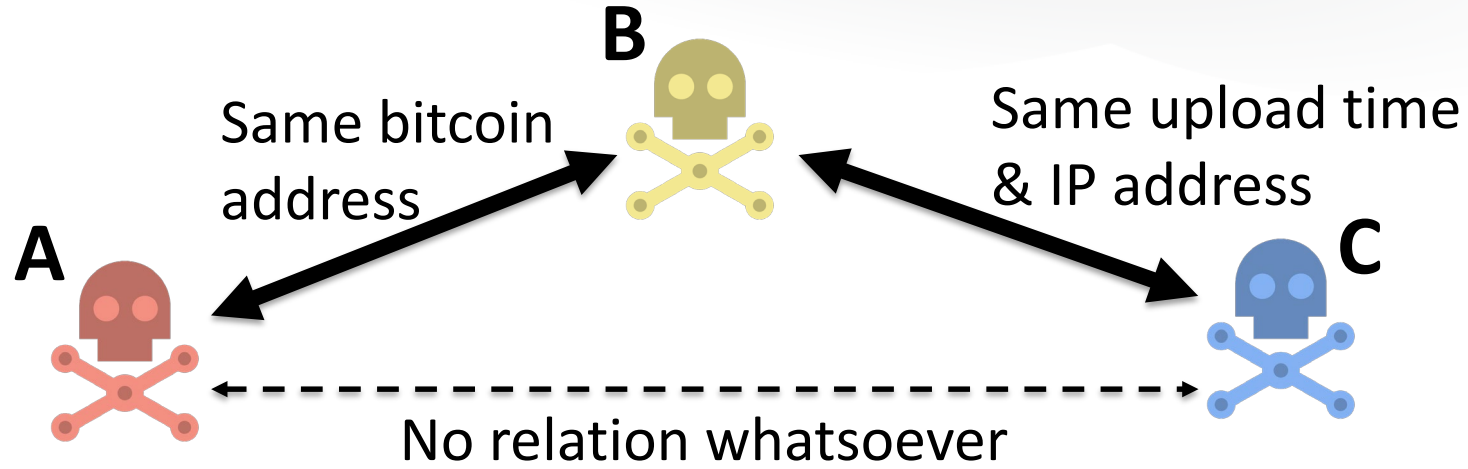


0.42	1.25	3.15	2.26	1.47	6.54	2.14	✓
1.13	1.67	0.02	0.66	1.55	3.12	5.41	✗
5.14	2.00	0.47	0.22	1.68	2.17	3.55	✓
0.12	5.13	4.22	2.14	1.47	5.55	6.42	✗
0.03	5.42	1.68	3.33	2.13	6.02	1.33	✓
4.27	1.67	2.23	1.14	1.67	1.03	5.21	✗

- **Lossy: information gets lost**

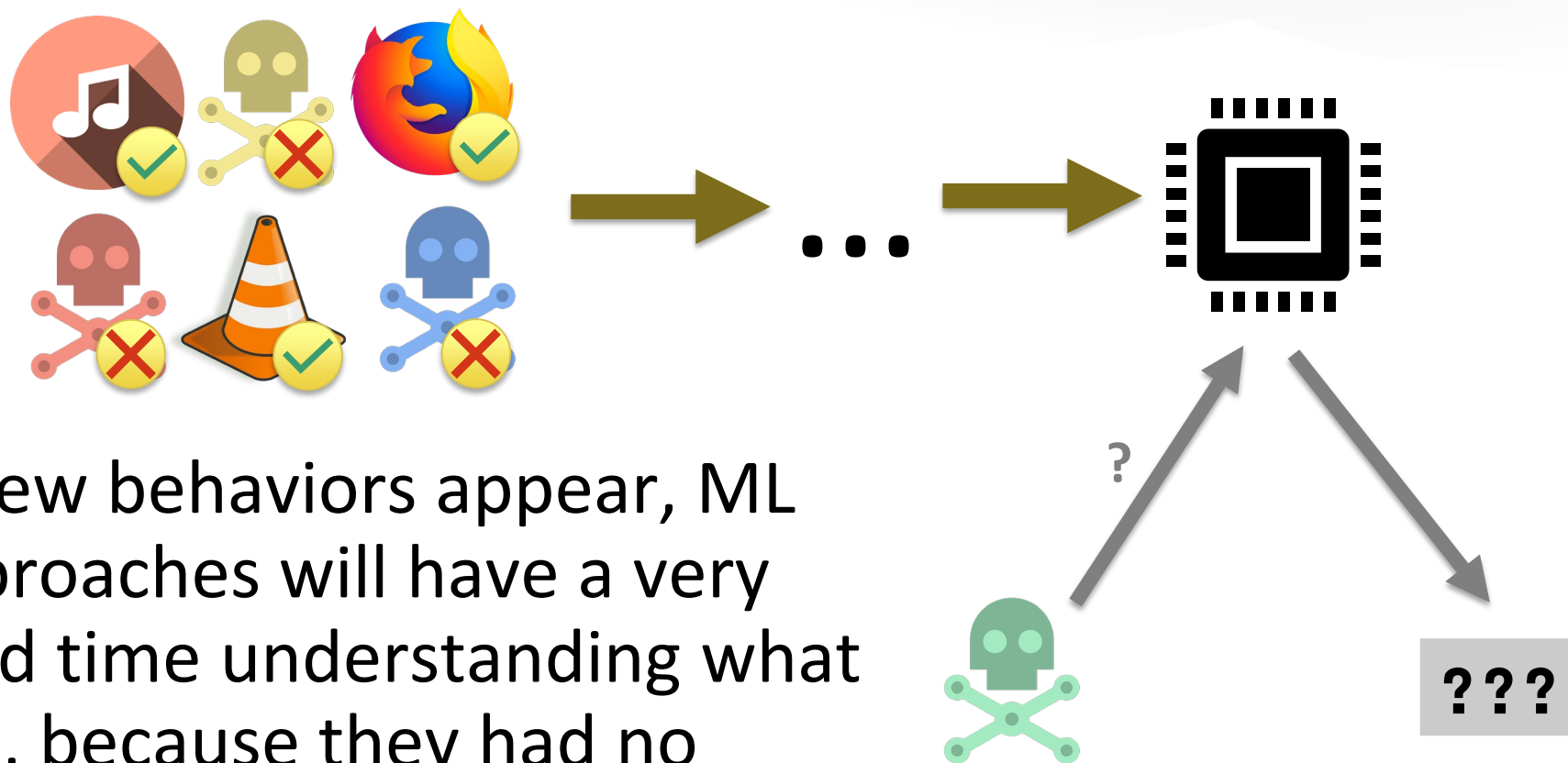
- We try to use as much information as possible, and we end up with something that's very complex
- E.g., dates + IP addresses + log lines + provenance info + fuzzy hashes...
- Converting all this to a set of numbers can lead to loss of important info

Problems in Feature Extraction (2)



- Usual mathematical approaches may fall short
 - Points in n-dimensional spaces and more generic approaches generally require at least a *metric space*, which has the *triangle property*
 - If A and B are close and B and C are close, then A and C can't be very far
 - May not be the case for us...

Problems in Validation: Training Set vs. Real World



- If new behaviors appear, ML approaches will have a very hard time understanding what it is, because they had no relevant examples

It's So Difficult!

- We've seen problems in all the parts of the process
 - Dataset
 - Labels
 - Feature extraction
 - Validation
- Classifiers are based on statistics and they recognize things that are reasonably similar to those that they've been shown before.
- They are useful, but not necessarily a definitive solution to our security problems.

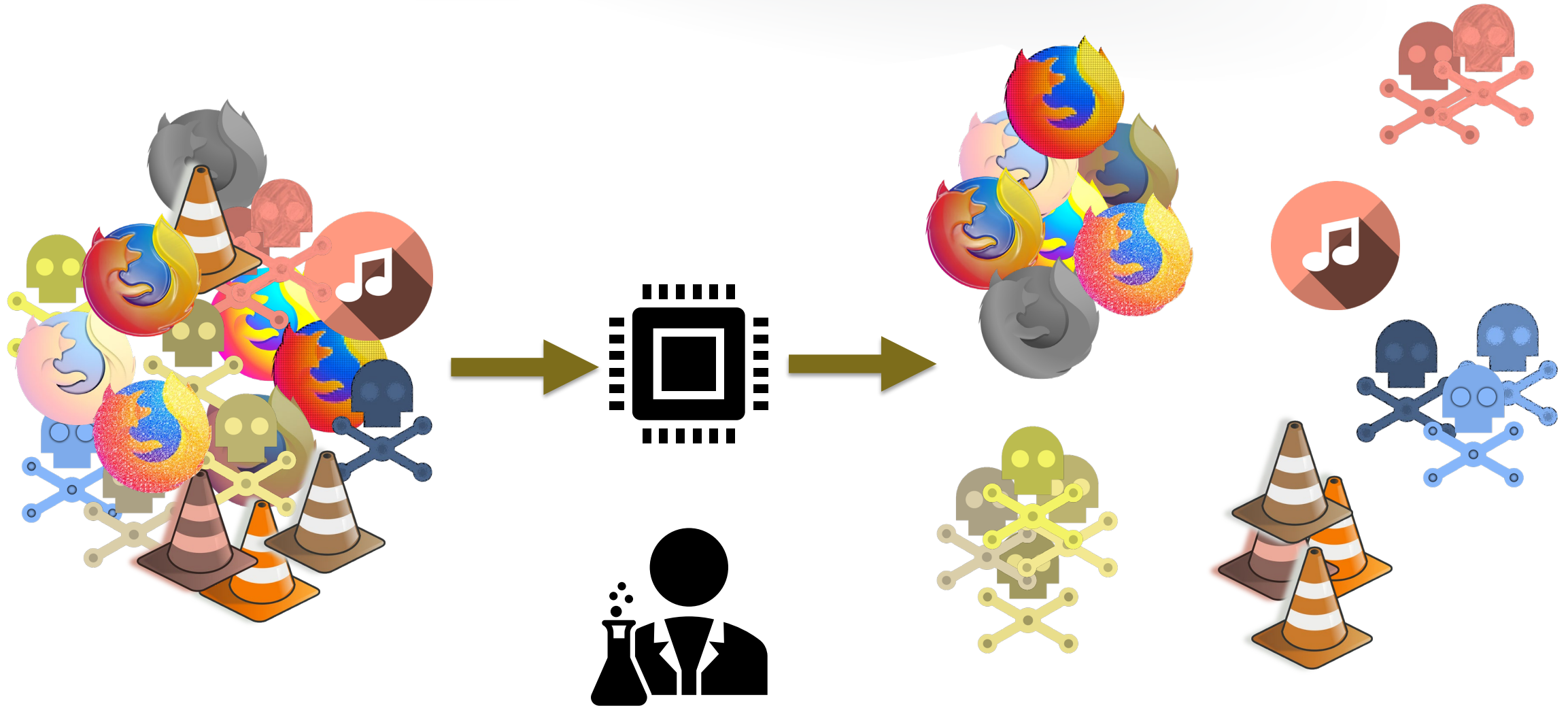
RSA®Conference2020

Humans in the Loop

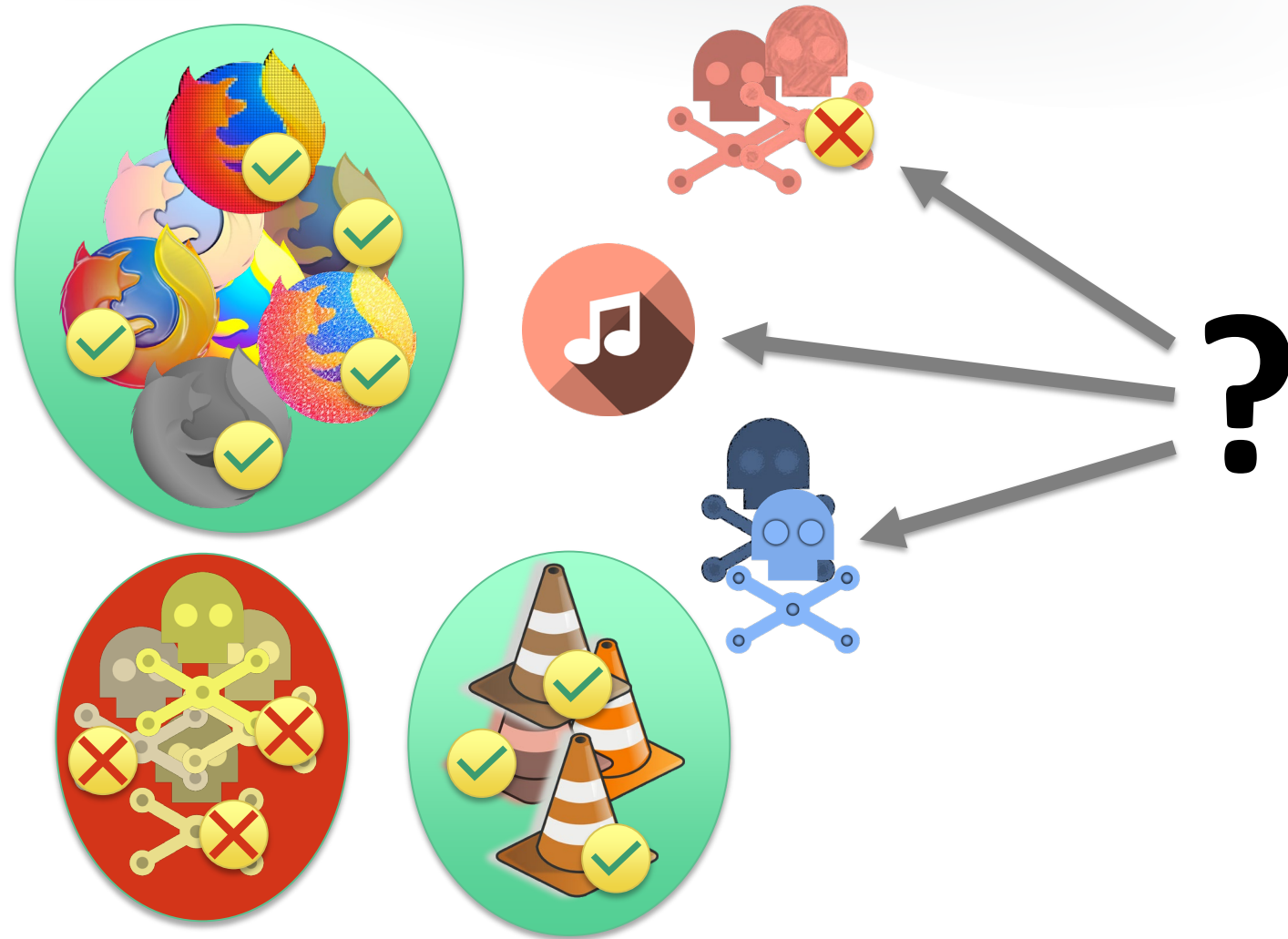
People Are Complementary to Machines

- Computers can recognize variations in statistical distributions (i.e., new things showing up)
 - But can't really say what they are
- People handle very little data, but they have intuition and much easier access to contextual information
 - They can search for stuff online and understand it
 - Or find the right colleague to ask for help

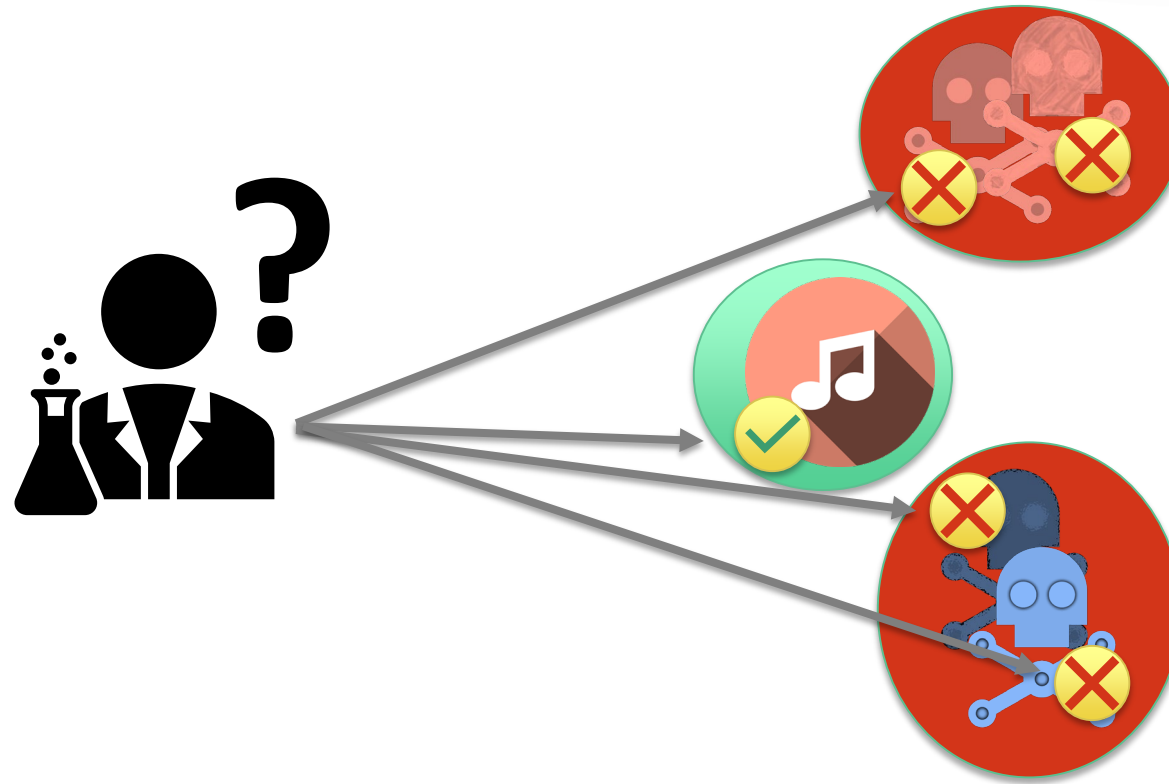
A Different Approach (1): Clustering



A Different Approach (2): Labels Describe What's Known



A Different Approach (3): Ask Experts About What's New

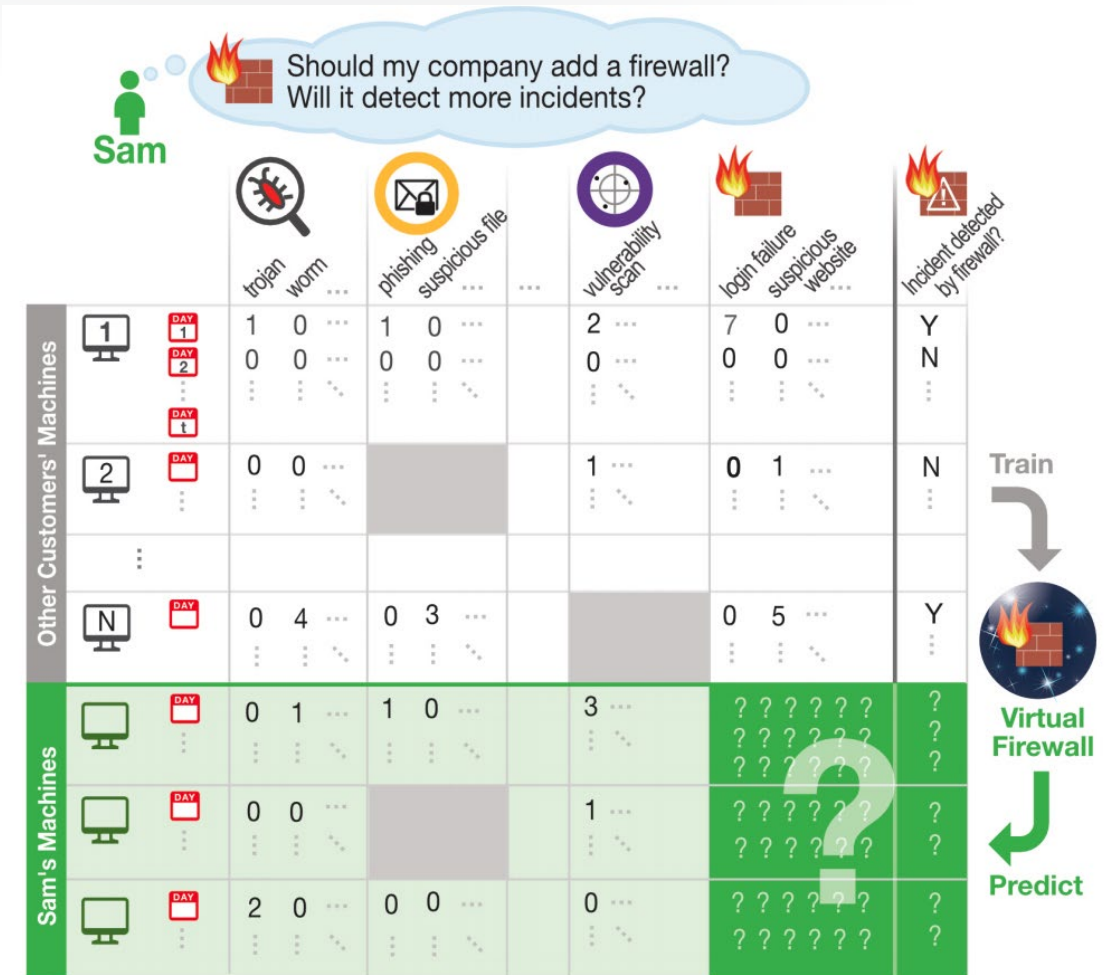


About Our Challenges

- Our semi-supervised approach allows dealing with data that changes over time (*“context drift”*)
- Now, we still have to deal with
 - Unreliable & incomplete datasets
 - Feature Extraction

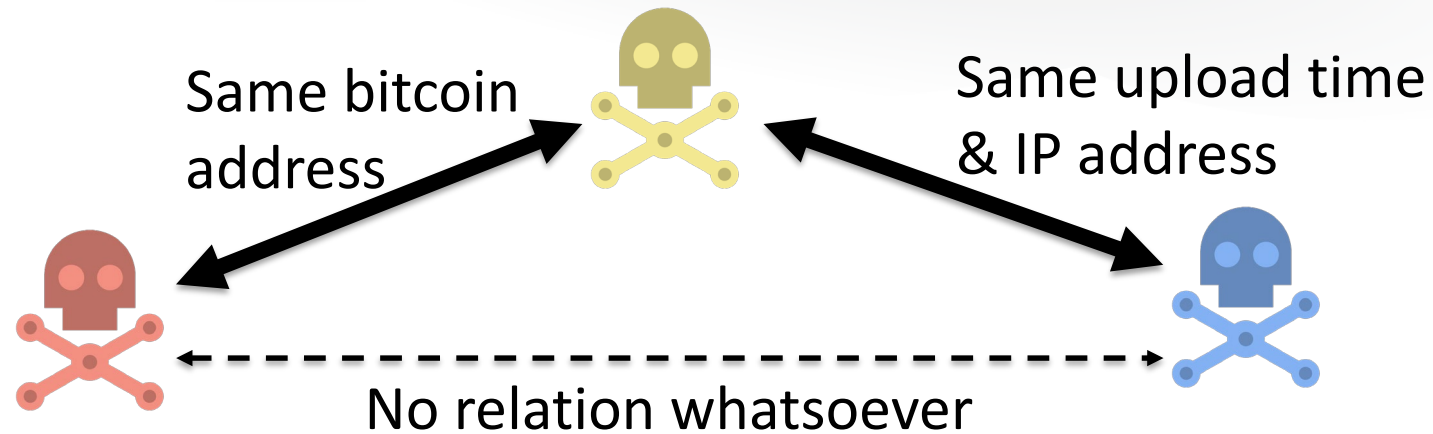
Fixing the Dataset: Virtual Products

- We can use ML to *predict* what a missing product would output given the existing ones' input
- *Virtual products* can
 - Complete an incomplete dataset
 - Find out which data look less reliable and maybe correct it
 - Find out which products would be more valuable because their output is more difficult to predict



Han & al, KDD'18

Feature Extraction: Dealing With Similarity



- Rather than converting our data to a series of numbers, we work with approaches that allow to use **arbitrary code** to compute similarity
- No “triangle rule” is needed!
- We’re thinking about solutions that are more user-friendly than writing code

RSA[®]Conference2020

Conclusions

Takeaways

- ML for security is both **very useful** and **difficult to do well**
- Ongoing research seeking solutions that work well for security
 - Active research area!
- Our take is that **humans** will continue to be essential in the foreseeable future
 - They have an intuition that machines do not have
 - But we can do **a lot** to make machines more helpful, making them do the “heavy lifting”

Apply What You Learned

- Consider how this relates to your everyday job
 - Do you have humans that can be **assisted** by ML?
 - Do you have ML that can benefit from human **supervision**?
- Read more about solutions, ideas and work in progress
 - See blogpost (<https://bit.ly/38WonGO>) for papers, software, videos, etc.
- Let me know what you think is missing in this discussion! 😊

