AI Bias in Cyber Security

Tale of Haves and Have Nots

Srinivas Mukkamala, PhD
Advisor Cyber Security Works
SVP, Product Management Security Products
Ivanti

About Us

Numbering Authority
54+ Zero Day Vulnerabilities

KOADIC

- Founded 2008 - 13+ Years
- Locations: NM (HQ), AZ & India
- 180+ Pen testers, Analysts & Researchers
- 54 Zero Day Vulnerabilities Discovered
- 400K+ Assets & 100K+ Apps Pen Tested
- Authors of WannaCry Payload + Double Pulsar + BlueKeep

CVE Numbering Authority (CNA)

CSW is a US Homeland Security CVE (Common Vulnerabilities & Exposures) Number Authority (CNA)

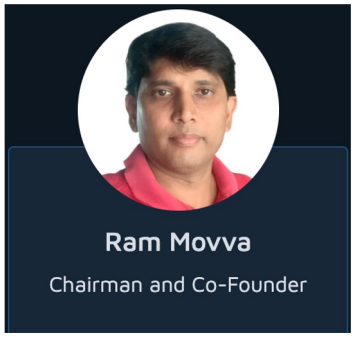
CVE Program Mission:
Identify, define, and catalog publicly disclosed cybersecurity vulnerabilities.

198	Global Numbering Authorities
32	Countries
15	Root or Country Agencies
167	Product Organizations
40	Research Organizations
17	Security Service Organizations

54+ Zero Day Vulnerabilities

Authors of WannaCry Payload – Double Pulsar

CSW/Securin Executives and Advisors



Ram Movva
Chairman and Co-Founder

Aaron Sandeen
Chief Executive Officer

Andrea Edmonds
Chief Financial Officer

Chris Acton
Vice President of Operations

Phil Manfredi
Chief Customer Officer

George Do
CISO, Gojek

Michael Montoya
CISO, Equinix

Sri Mukkamala
SVP, Ivanti

Rich Hlavka
VP Partners, Pangea

Shreesha Ramdas
CEO, Strikedeck

CSW/ Customers

Public Sector ● Healthcare ● Finance ● Technology ● Manufacturing
Oil & Gas ● Education ● Telecommunications ● Retail

Winding Path to Product Leader

Engineer
Indian Space Research

Researcher
&
Lead Scientist
NMT/ICASA

CTO
&
Co-Founder
CAaNES

Co-founder and
CEO
RiskSense

Advisor Cyber Security Works
SVP and GM,
Security Products, Ivanti

Computational Analysis of Cyber Terrorism Against the US

AI Bias in Cyber Security is a Real Problem

AI might be your friend, but it also powers your foes

Data Labeling and Training are Time Consuming

Real Challenge is Access Data in a Timely Manner - Latency

Machine

Machine Generated and Synthetic Data Might not Yield Optimal Results

Current State – Latency?

Is Data the New Oil or Time?

Everyone seems starved of time: time to relax, time for family and friends, time to do absolutely nothing.

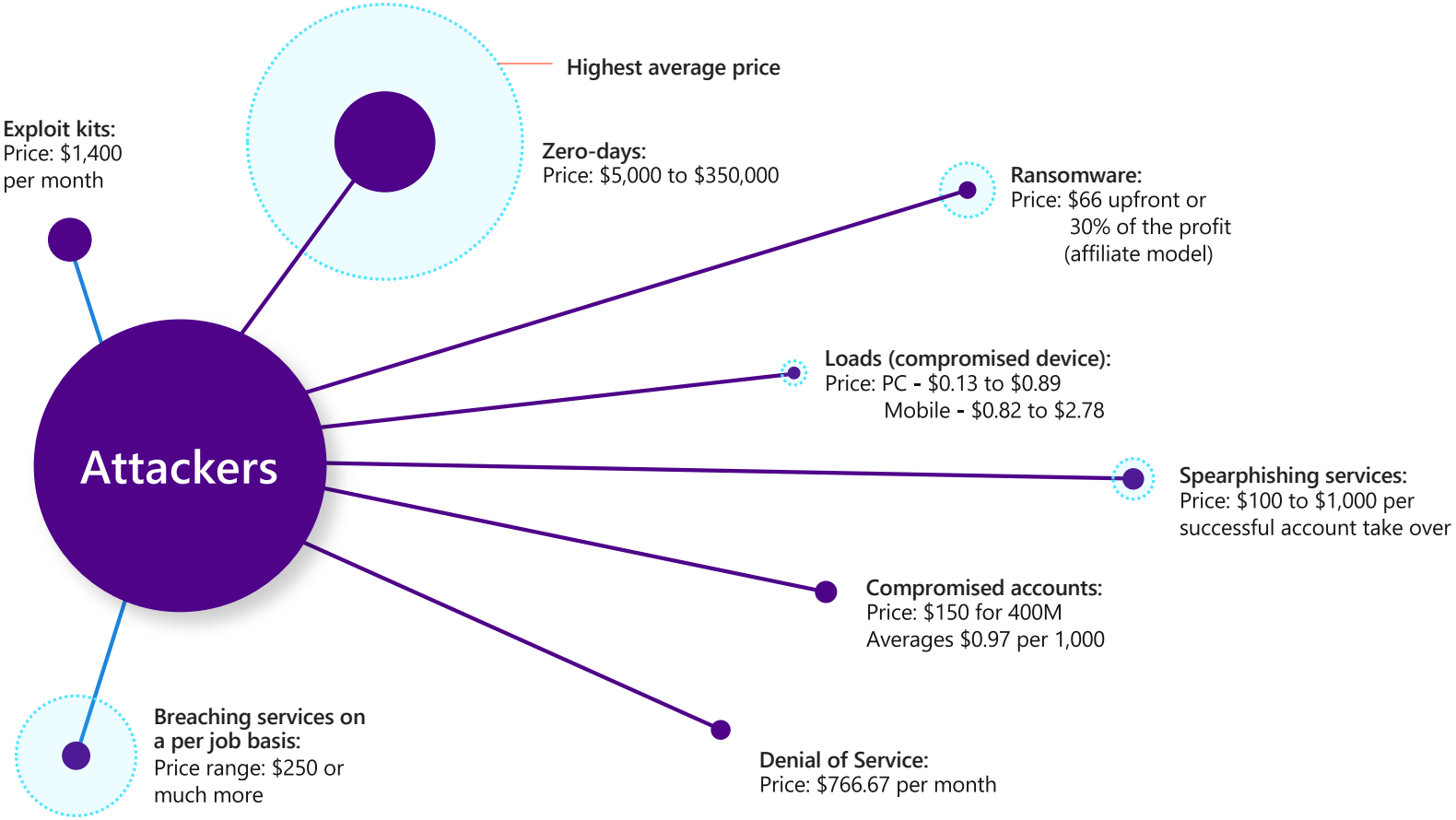
Most tools scooped incalculable data to fascinate us with as much unfiltered information as possible

We all go down search engine/analysis rabbit holes because of FOMO that we aren't picking the next best thing – aka **Zero Days!**

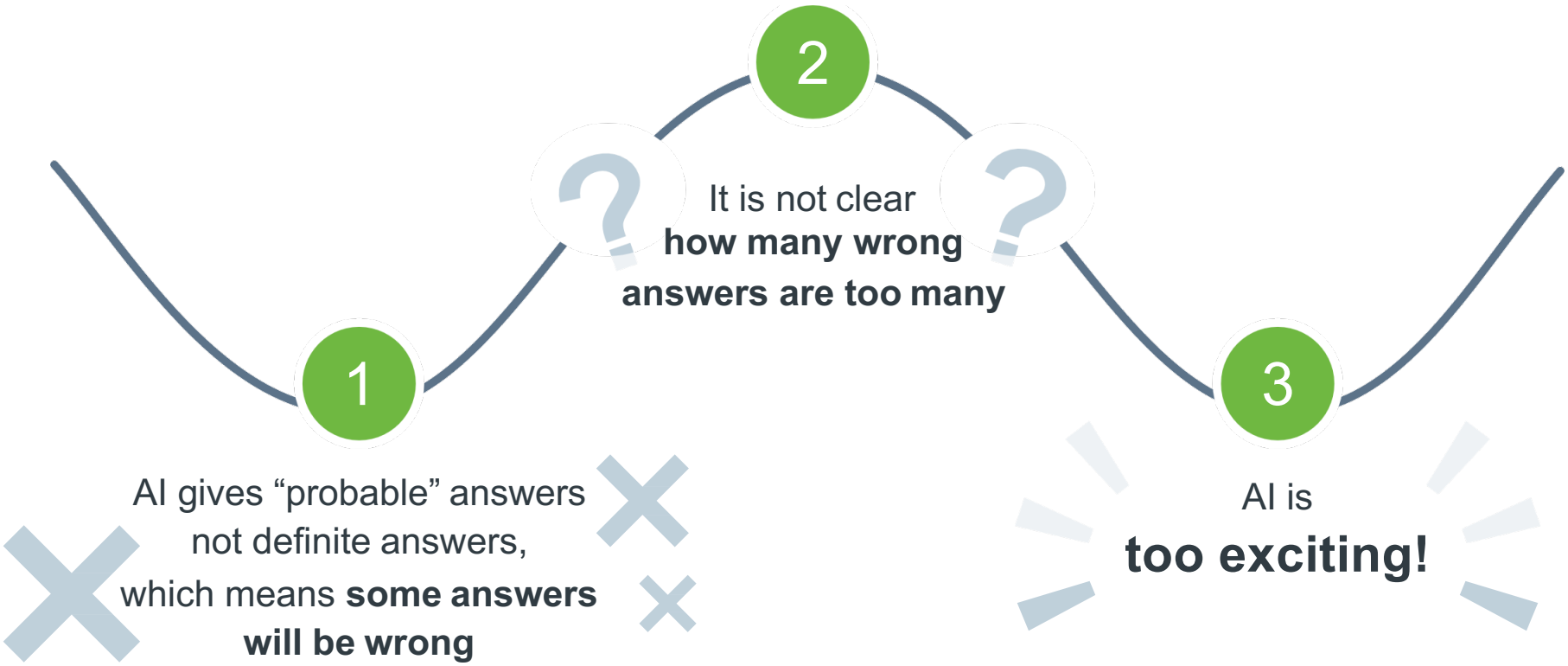
We need more authoritative and filtered data in our lives leading us to trustworthy sources so we can make fast, accurate decisions!

Spoke about Latency in 2016 – NVD (Vulnerability) – Patch – Weaponization – Exploit

Attack Services are Cheap – Asymmetric Threat



AI is Too Exciting – Domain Expertise and Validation are a Key



… Artificial Intelligence…
Is Predictive Good Enough or Do We Need Validation As Well?
Are we Attempting to Understand Bias

AI Bias is a Real Problem Beyond Cyber Security

AI may know a lot about something,

but not everything…

It is only as good as the data and time spent improving it.

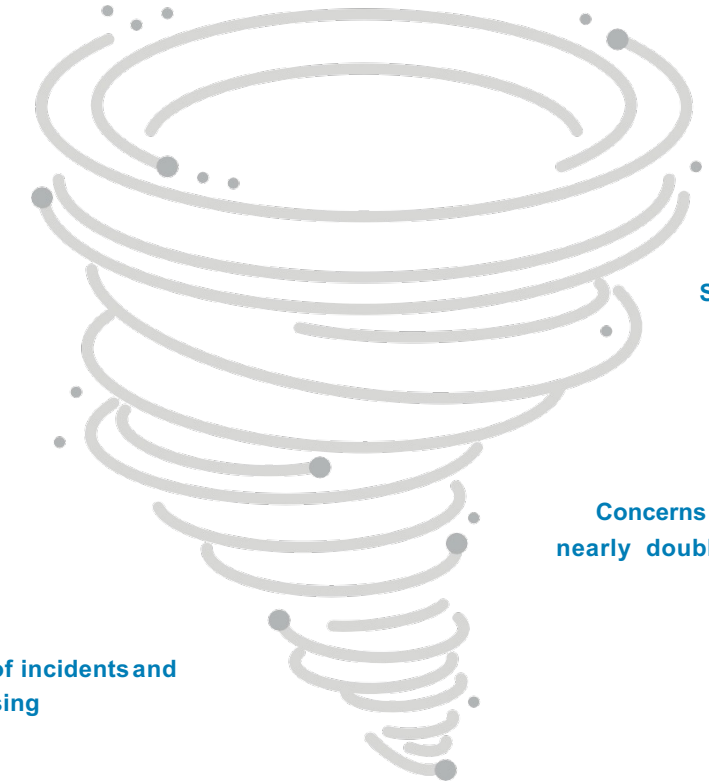
Biased based on what is taught!

Problem Statement

Vulnerability Intelligence – Security is a Data Problem

Cyber risks and events growing exponentially

Lack of clarity and compromised decision-making



Shortage and expense of security skills

Overwhelmed teams and technology

Concerns about reputation nearly double as the future is considered

Repercussions of incidents and breaches increasing

A Day in Life of a Security Analyst

INFORM

VULNERABILITY INTELLIGENCE CAN PROVIDE AN INFORMATION ADVANTAGE TO ORGANIZATIONS, NOTIFYING THEM OF SIGNIFICANT DISCLOSURE, PATCHING, OR EXPLOITATION DEVELOPMENTS

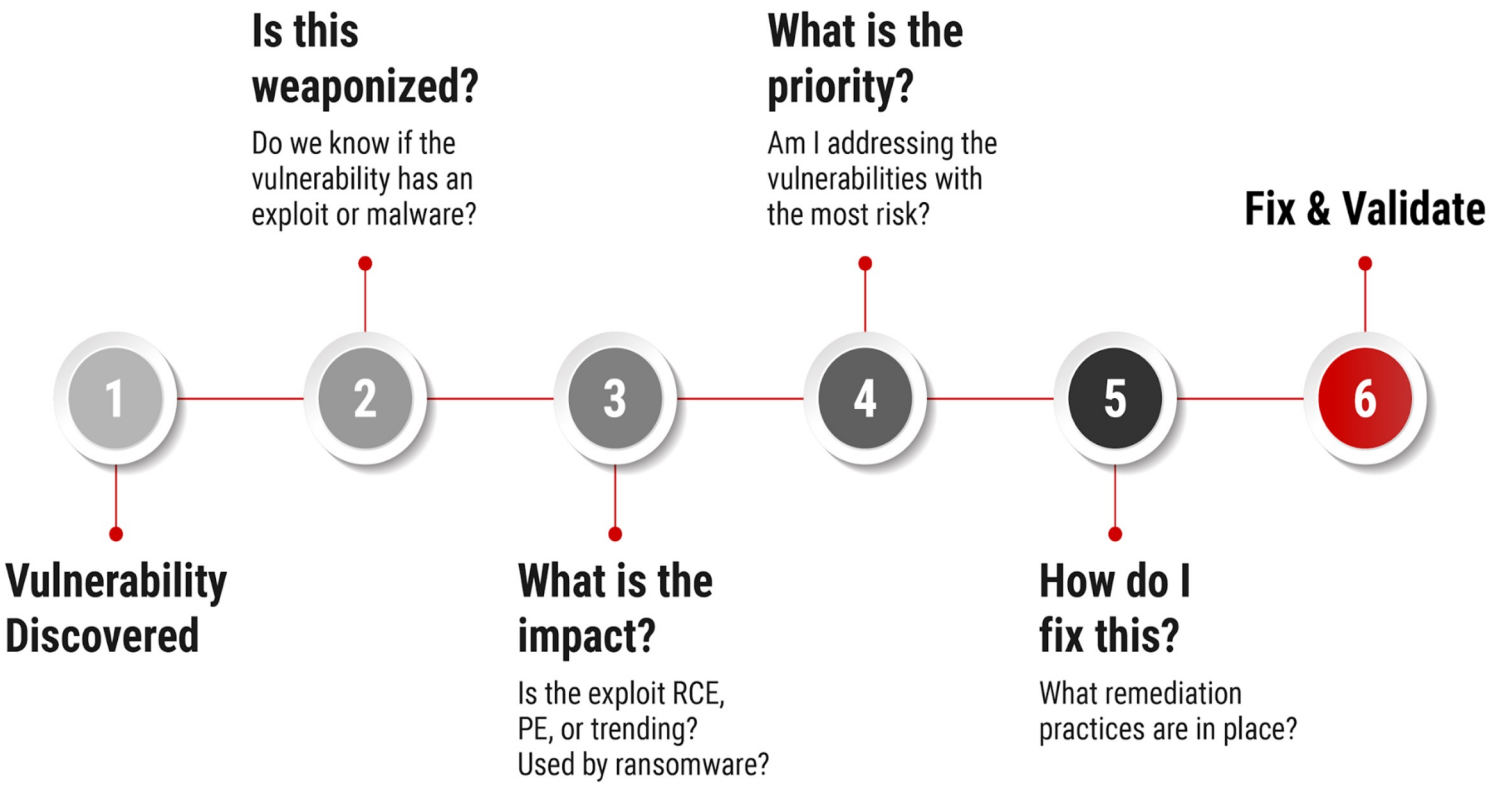
TRIAGE

VULNERABILITY RATINGS AND ANALYSIS ASSIST ORGANIZATIONS IN PRIORITIZATION AND TAILOR PATCHING SCHEDULES BASED ON THE GREATEST THREATS TO THEIR ORGANIZATION, SECTOR, AND GEOGRAPHIC REGION

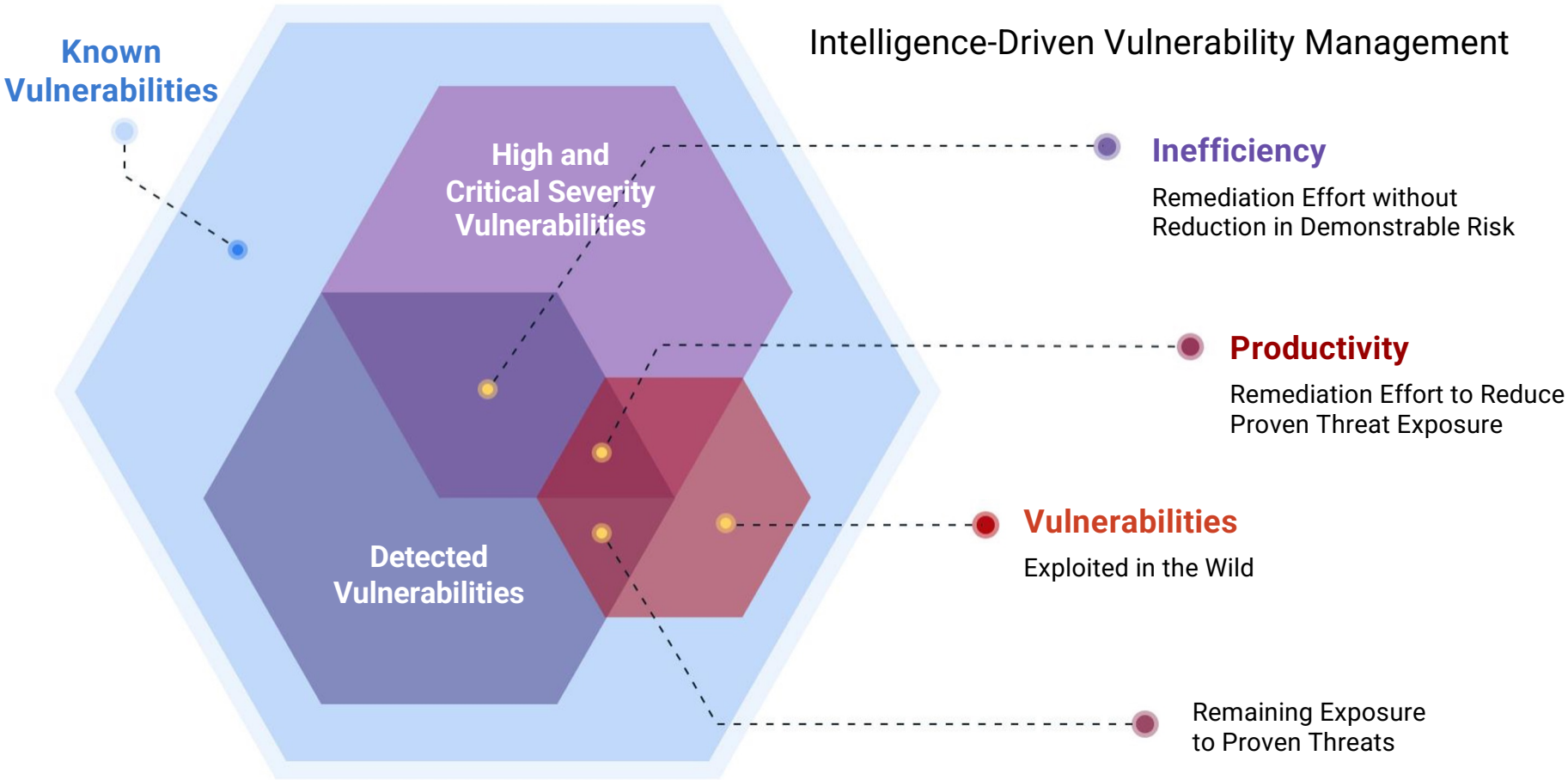
MONITOR

ONGOING MONITORING OF THE THREAT ENVIRONMENT ALLOWS ORGANIZATIONS TO ADJUST PATCHING AND MITIGATION PLANS TO RESPOND TO REAL WORLD CHANGES

Vulnerability Lifecycle



Vulnerability Prioritization



Differentiator - inclusion of Predictive and Definitive Vulnerability Scoring

Predictive

1. Has the vulnerability been exploited in the wild?
2. Is there evidence of exploitability?
3. Has a proof of concept been published online?
4. Has the vulnerability been discussed on social media?
5. Has the vulnerability been discussed on deep/dark web forums?
6. Has the vulnerability been mentioned in a news article?
7. Is an exploit advertised for sale online?

AI Driven
Early Warning Prediction

28 AI/ML Patent Submissions

Definitive

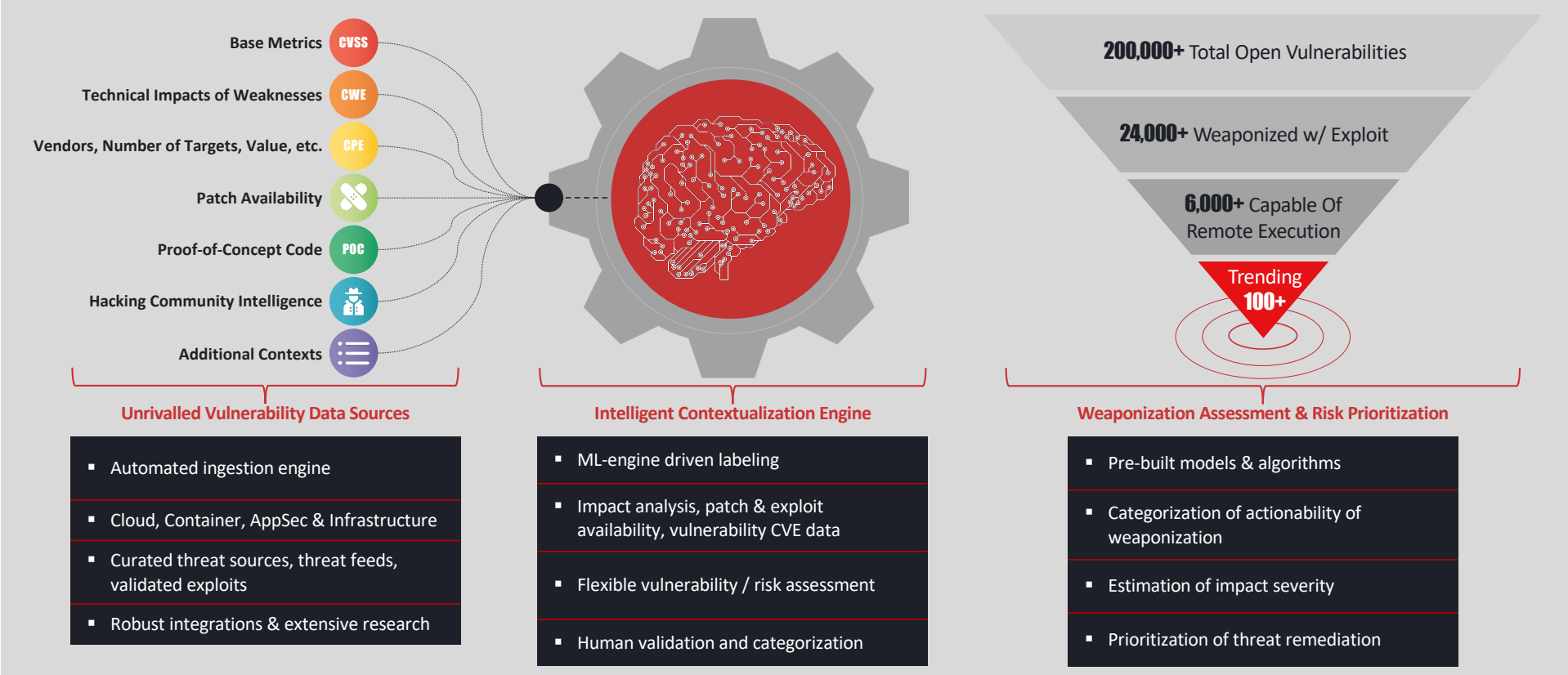
1. Is the vulnerability associated with ransomware?
2. Is the vulnerability associated with other malware?
3. Is the vulnerability associated with a threat actor or nation states?
4. Is the vulnerability trending?

DARPA

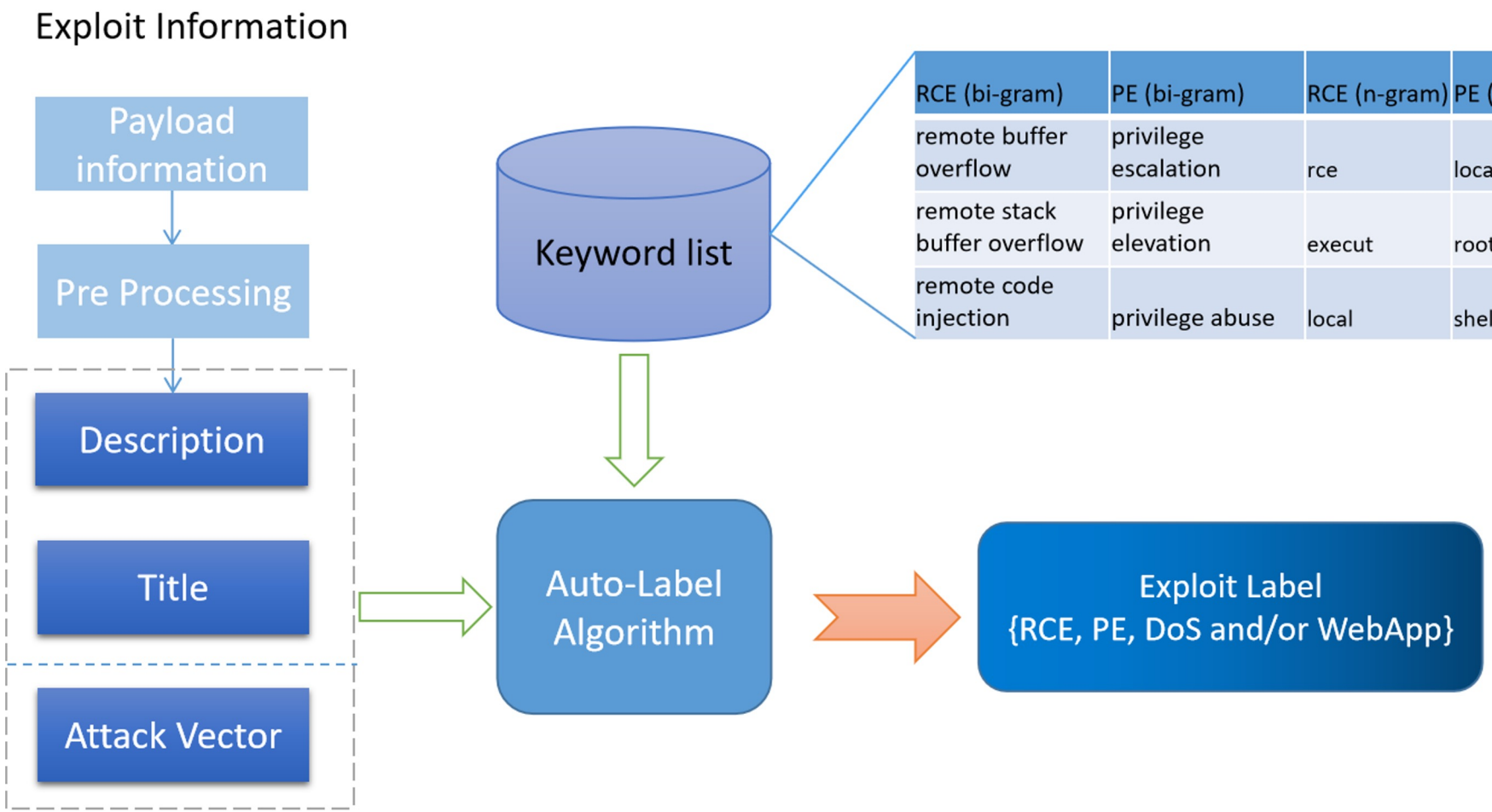
ASU
Arizona State University

Current Framework

VI Ingests Industry-leading Proprietary Threat, Vulnerability, & Exploit Intelligence Into Its ML-Enabled Contextualization Engine & To Produce Best-In-Class External Risk Scores Based On Accessibility, Business Criticality, And Probability Of Breach



ML for Labeling Exploits – Vulnerability Prioritization



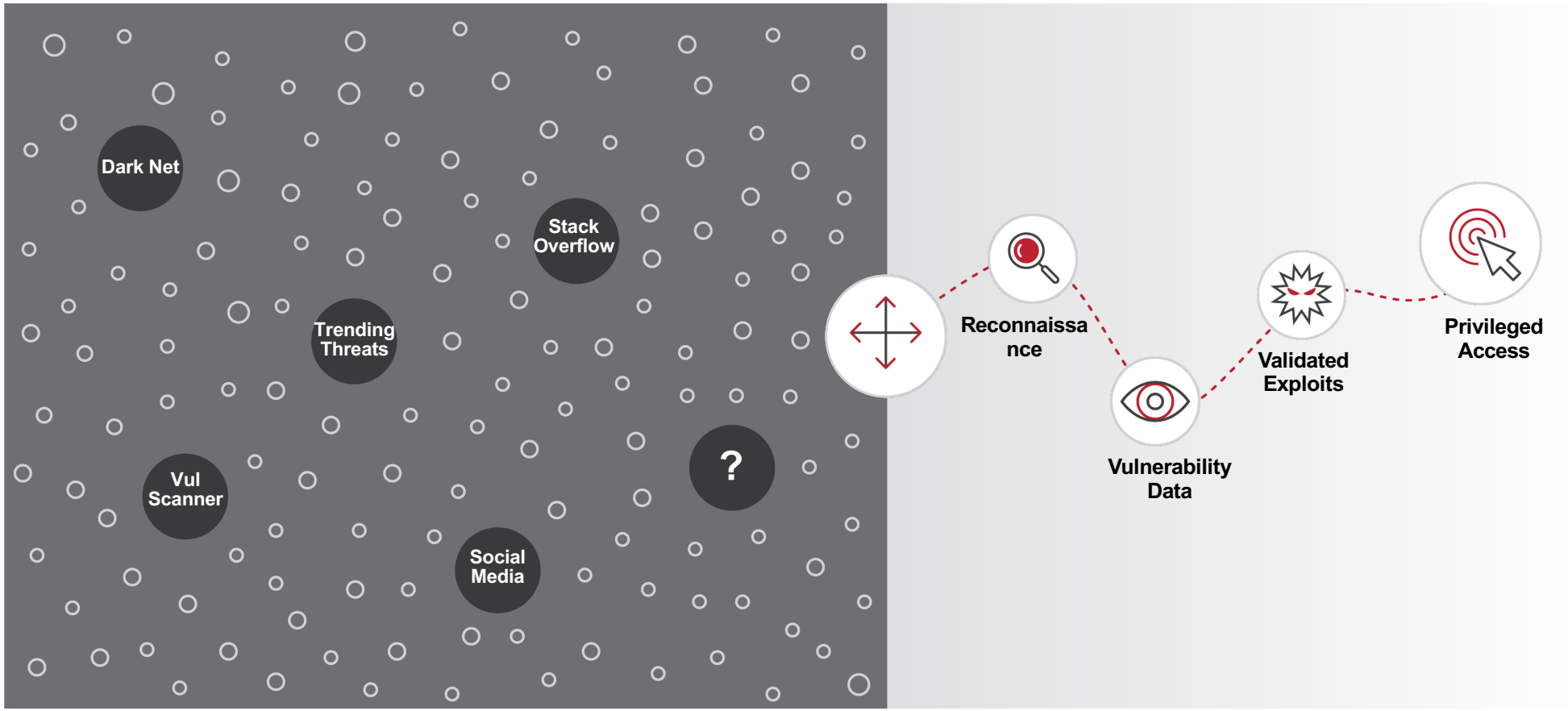
RCE (bi-gram)	PE (bi-gram)	RCE (n-gram)	PE (n-gram)
remote buffer overflow	privilege escalation	rce	local
remote stack buffer overflow	privilege elevation	execut	root
remote code injection	privilege abuse	local	shellcod

Data Never Lies – We Need it at the Right Time

Risk Based View of the Known CVEs

ML and Data Point of View

Attacker View – Attacker Genome



Comprehensive Threat & Vulnerability Information

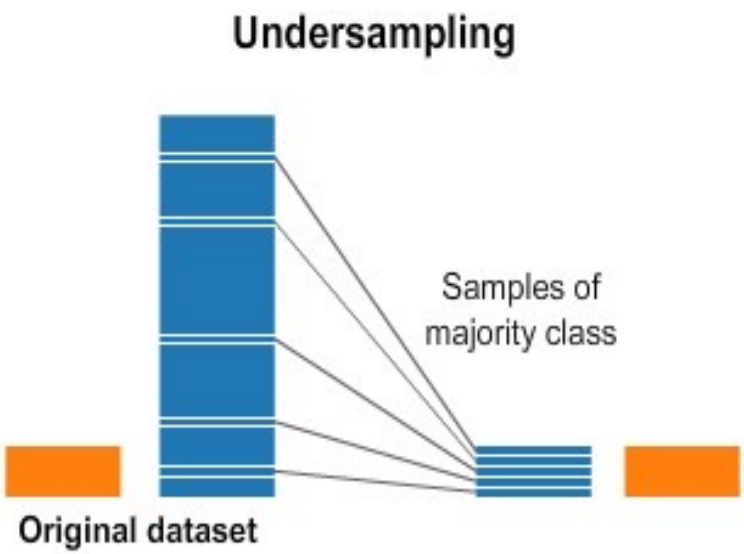
Sources

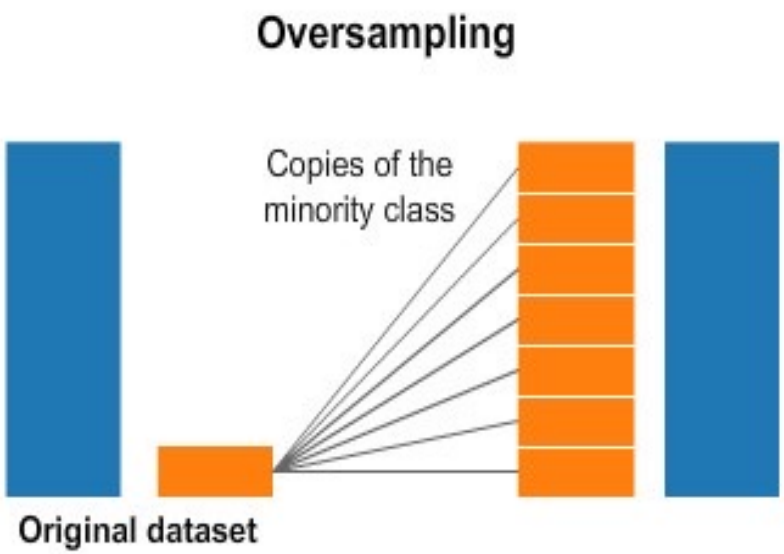
- **353** Structured
- **286** Forums
- **53** Social Media
- **336** CVE sources
- **165** IP sources
- **10** Exploit sources

Coverage

- **233,976** CVE
- **310** Ransomware Associations
- **239** Threat Actor Association
- **24,630** Exploits Association
- **8,329** RCE/PE
- **647** CISA KEVs
- **25,966** Affected Vendors
- **101,156** Affected Products
- **50,075** Hacker Chatter

Imbalanced Datasets





AI and ML Journey

Models

Model finds a customer has an 89% propensity to churn.

Dataset & Predictions

LIME pulls out the various datasets and predictions...

Pick Step

...and makes small tweaks to the inputs.

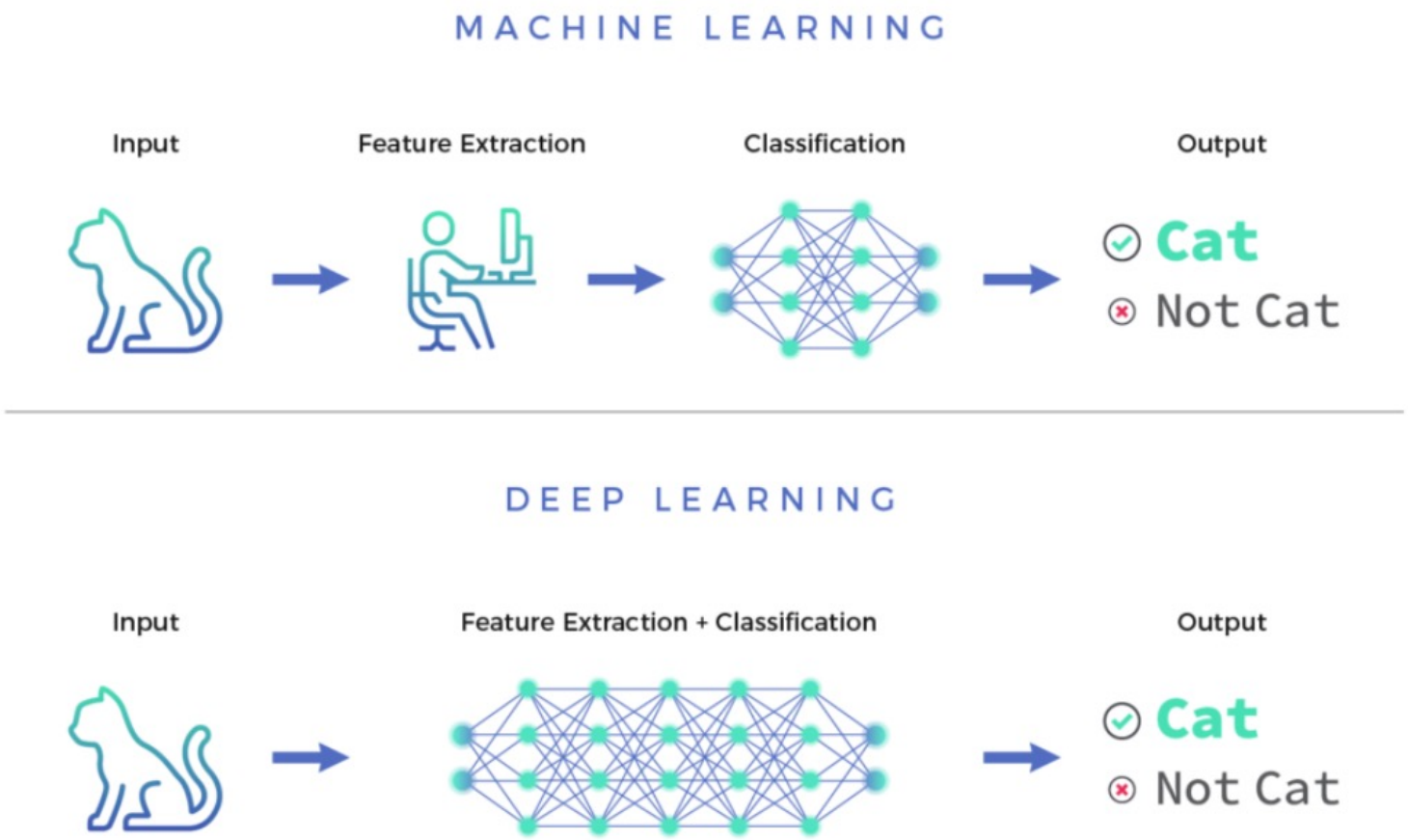
Explanations

LIME then generates explanations about why a prediction was made and a variable's impact on the outcome.

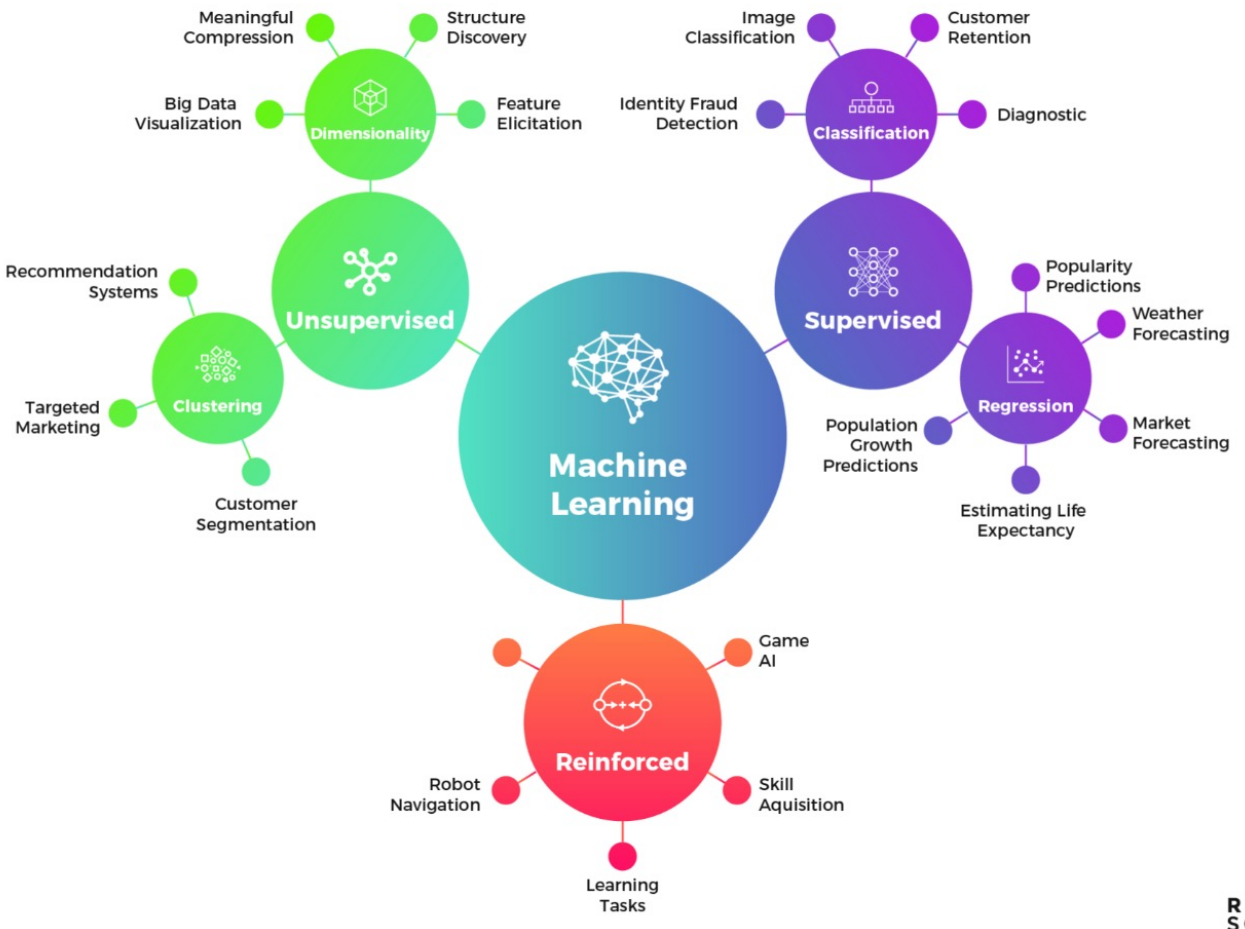
Human Makes a Decision

A human can then make an informed decision about what to do with the model's findings.

Machine Learning Vs Deep Learning



DIFFERENT TYPES OF MACHINE LEARNING



ROCKET SOURCE

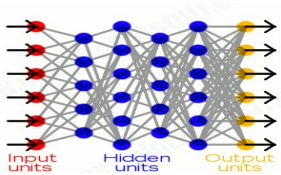
Data and Training Transparency

- What data was it based on?
  - Sources referenced?
  - Access to overall collection/corpus?
- Who trained AI?
  - Experience? Background? Vetting?
  - Who will update it? How often?
- How are we validating the results?
  - Human or Machine validation?
  - Does the end user has the ability to provide feedback?

What are we Trying to Do – Explainable AI (aka Transparency)

Today

Training Data → Learning Process →



Learned Function

This is an Exploit
(p = .93)

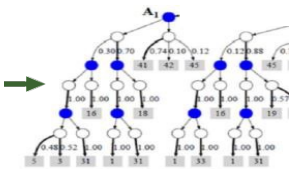
Output →

User with a Task

- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?

Future

Training Data → New Learning Process →



Explainable Model

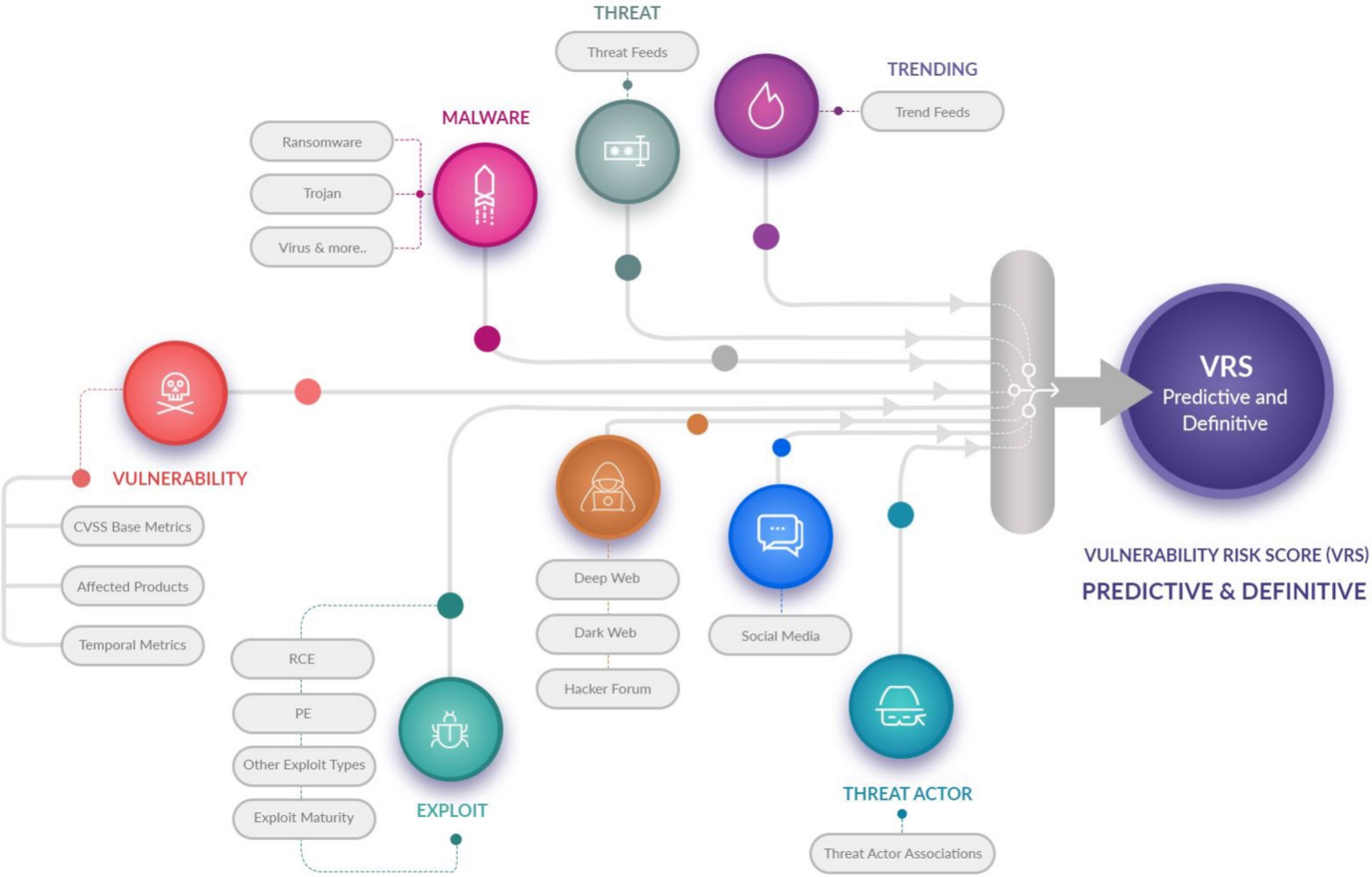
This is an Exploit:
- It is a Remote Code Execution, and is on Windows

Explanation Interface

User with a Task

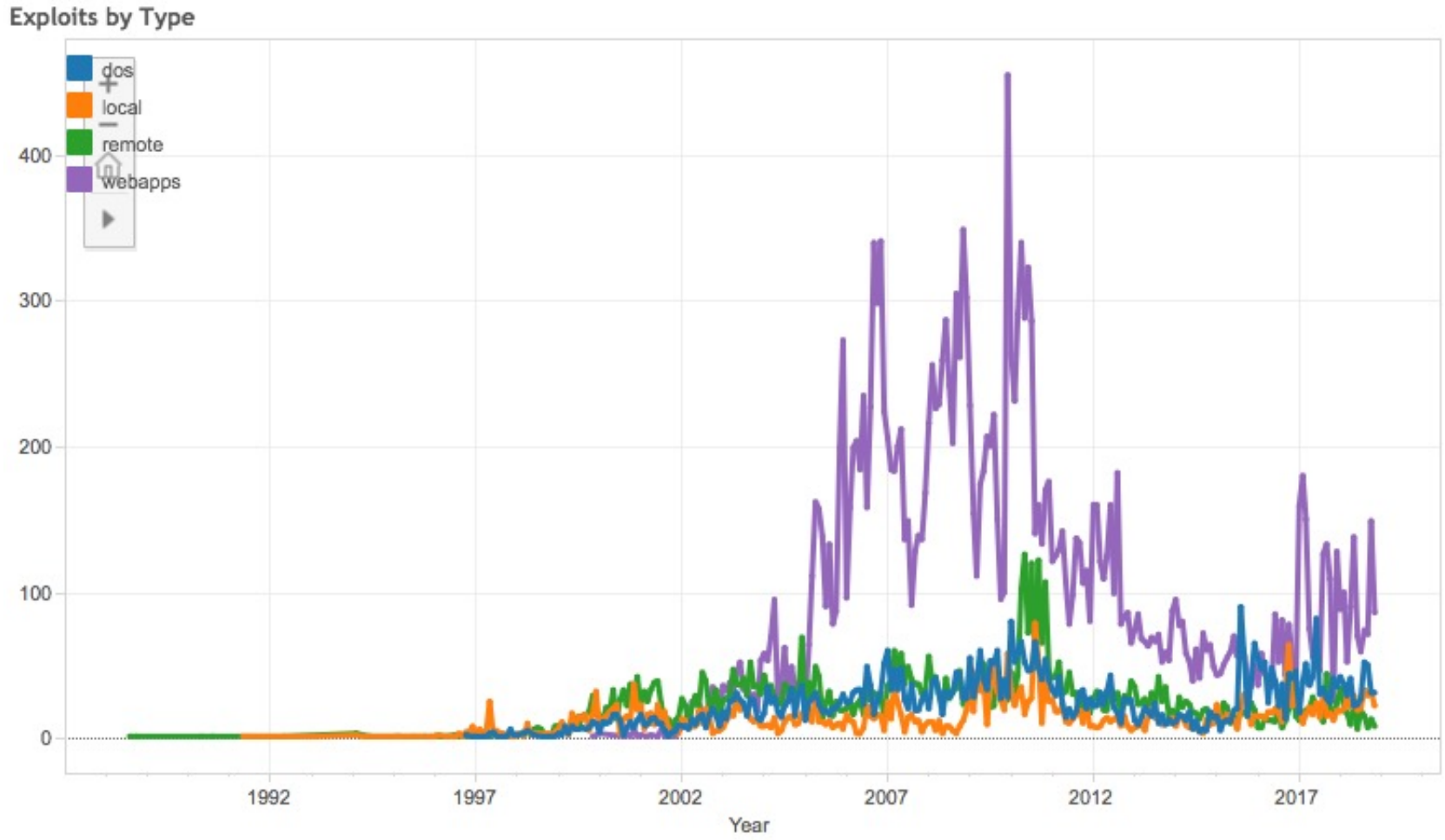
- I understand why
- I understand why not
- I know when you'll succeed
- I know when you'll fail
- I know when to trust you
- I know why you erred

VRS Risk Score for every Vulnerability

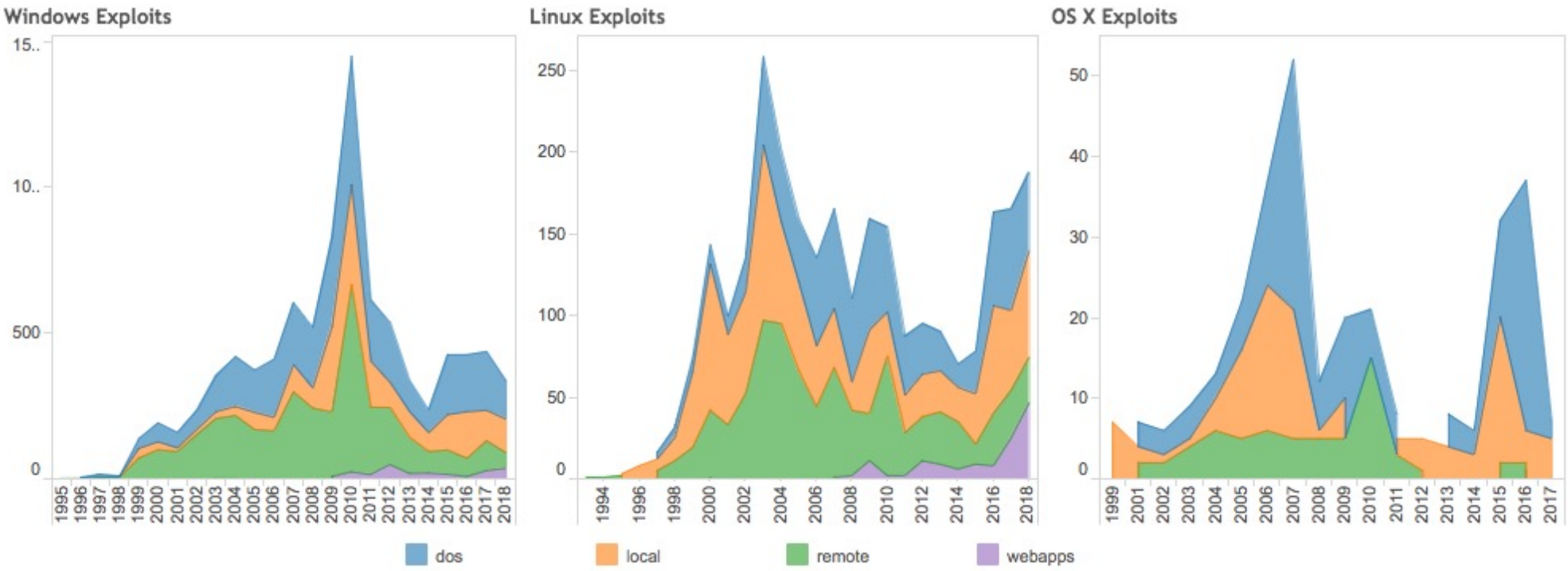


Results

Granular View into Types of Exploits



Operating Systems Mapped to Exploit Type



What to Expect from the Data

What…

- Comes next?
- Outliers?
- Changed? How can I tell what changed?
- New? Unexpected?
- Validates assumptions?
- Increased/decreased frequency?

Can We Measure Everything…

- **Attack Vector (AV)**
- Attack Complexity (AC)
- **Privileges Required (PR)**
- **User Interaction (UI)**
- **Modify data**
- **Read Data**
- **Code/Command Execution**
- **Priv. Esc**
- **Etc.**

- **Patch Release Latency**
- **Threat Release Latency**
- **Threat Trending Information**
- **Threat IOC Distribution**
- **Reference count**
- **Tweet age**
- **Reddit Mention**
- **Confirmed Simulated Compromise (Human validated)**

...But Which Ones Matter?

Challenges Persist

Let us Dive in

Into the rabbithole

Data Sets for our Analysis

Dap Map: CISA KEVs to MITRE Framework using the following resources

NVD

LARGEST DATABASE OF REPORTED AND KNOWN VULNERABILITIES

Launched by NIST, NVD is a vulnerability database built upon — and fully synchronized with — the CVE List so that any updates to CVE appear immediately in NVD.

CWE

DATABASE OF COMMON WEAKNESS ENUMERATION

The Common Weakness Enumeration is a category system for hardware and software weaknesses and vulnerabilities.

CAPEC

DATABASE OF KNOWN PATTERNS OF ATTACK

MITRE ATT&CK® is a globally-accessible knowledge base of adversary tactics and techniques based on real-world observations.

ATT&CK

ATT&CK DATABASE FOR ADVERSARY TACTICS AND TECHNIQUES

MITRE ATT&CK® is a globally-accessible knowledge base of adversary tactics and techniques based on real-world observations.

CHALLENGES ENCOUNTERED IN DATABASES

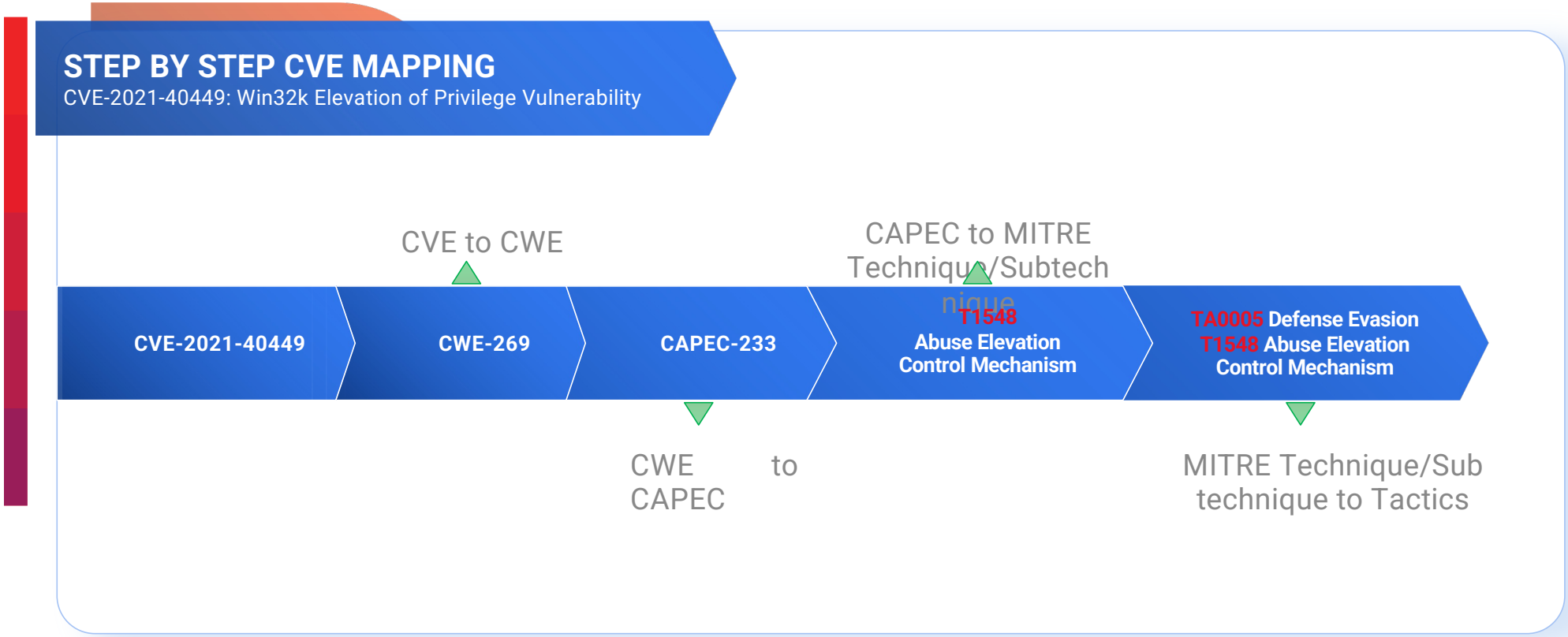
- Incomplete Data
- Latency in publishing Exploits, Patch information
- Missing CVSS ratings
- Riddled with Inaccuracies
- Not comprehensive enough for security researchers

- Missing CWE data
- Deprecated CWE

- Incomplete Data
- Missing CAPEC Information
- Missing ATT&CK Taxonomy Mappings.

- Incomplete Data
- Missing Taxonomy Mappings
- Outdated MITRE ATT&CK Mappings
- information
- APT Groups & Ransomware not mapped

Mapping Analysis – Step by Step Process

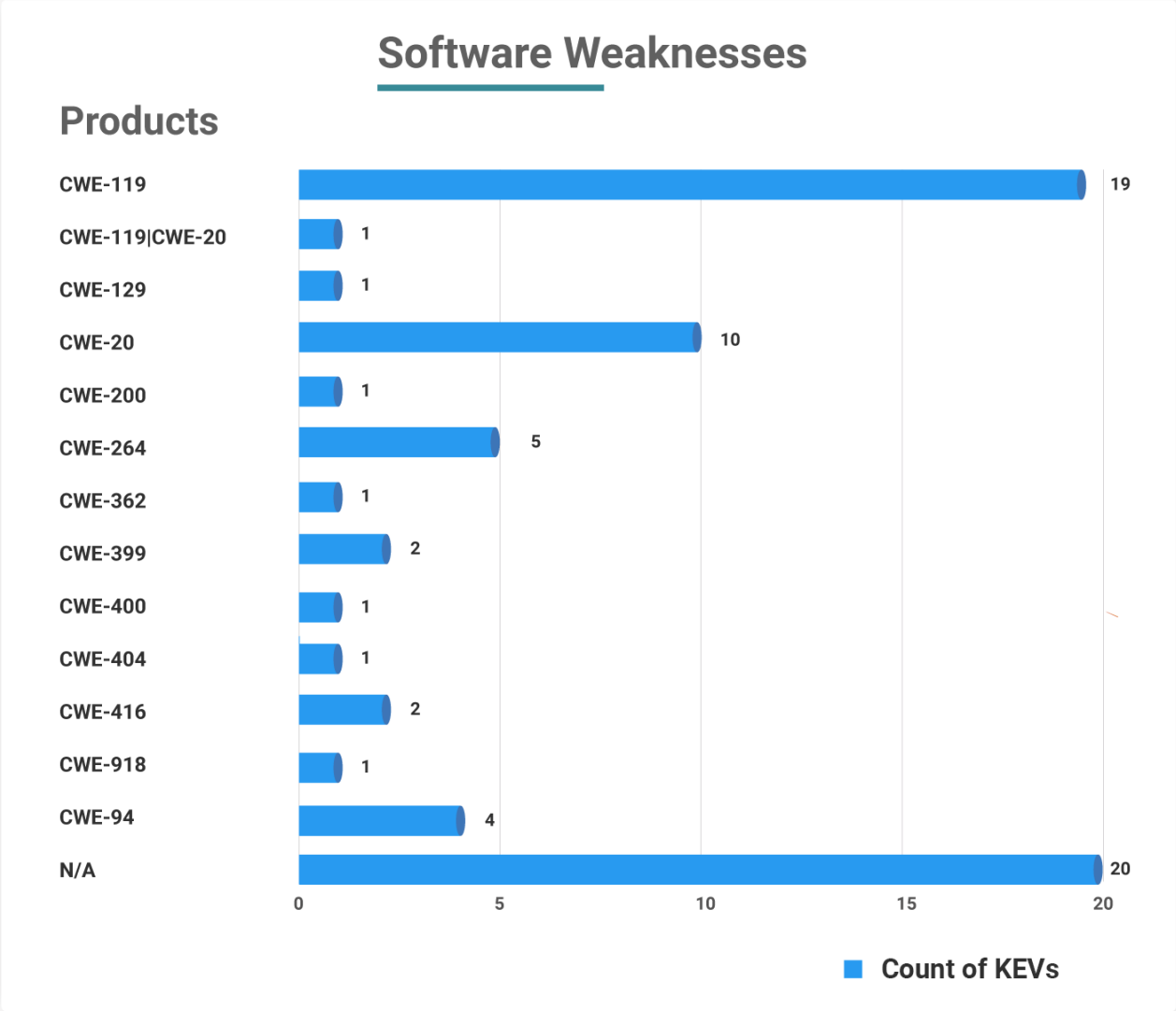


CVE to CWE Mapping

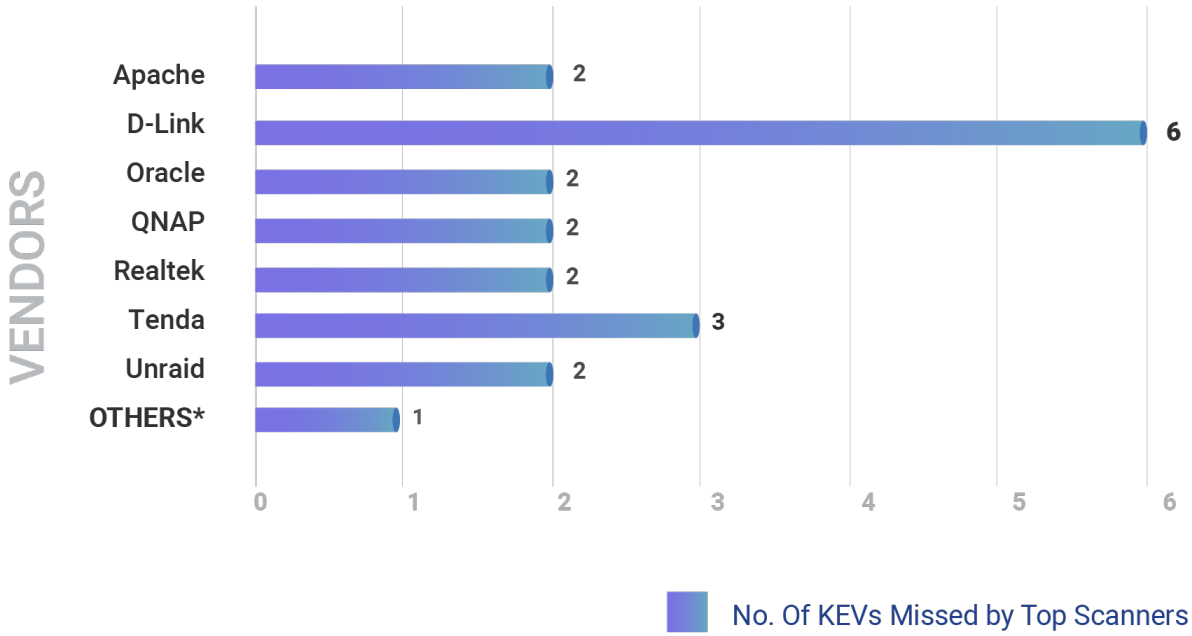
EXPAND DATA SOURCE

Scraped data from multiple reliable sources including CNAs and Vulnerability Databases for CWE information.

More than
17 **sources were used to map the TTPs of a vulnerability.**


Software Weaknesses
Products
CWE-119
CWE-119|CWE-20
CWE-129
CWE-20
CWE-200
CWE-264
CWE-362
CWE-399
CWE-400
CWE-404
CWE-416
CWE-918
CWE-94
N/A
19
1
1
10
1
5
1
2
1
1
2
1
4
20
0
5
10
15
20
Count of KEVs

Affected Vendor Analysis

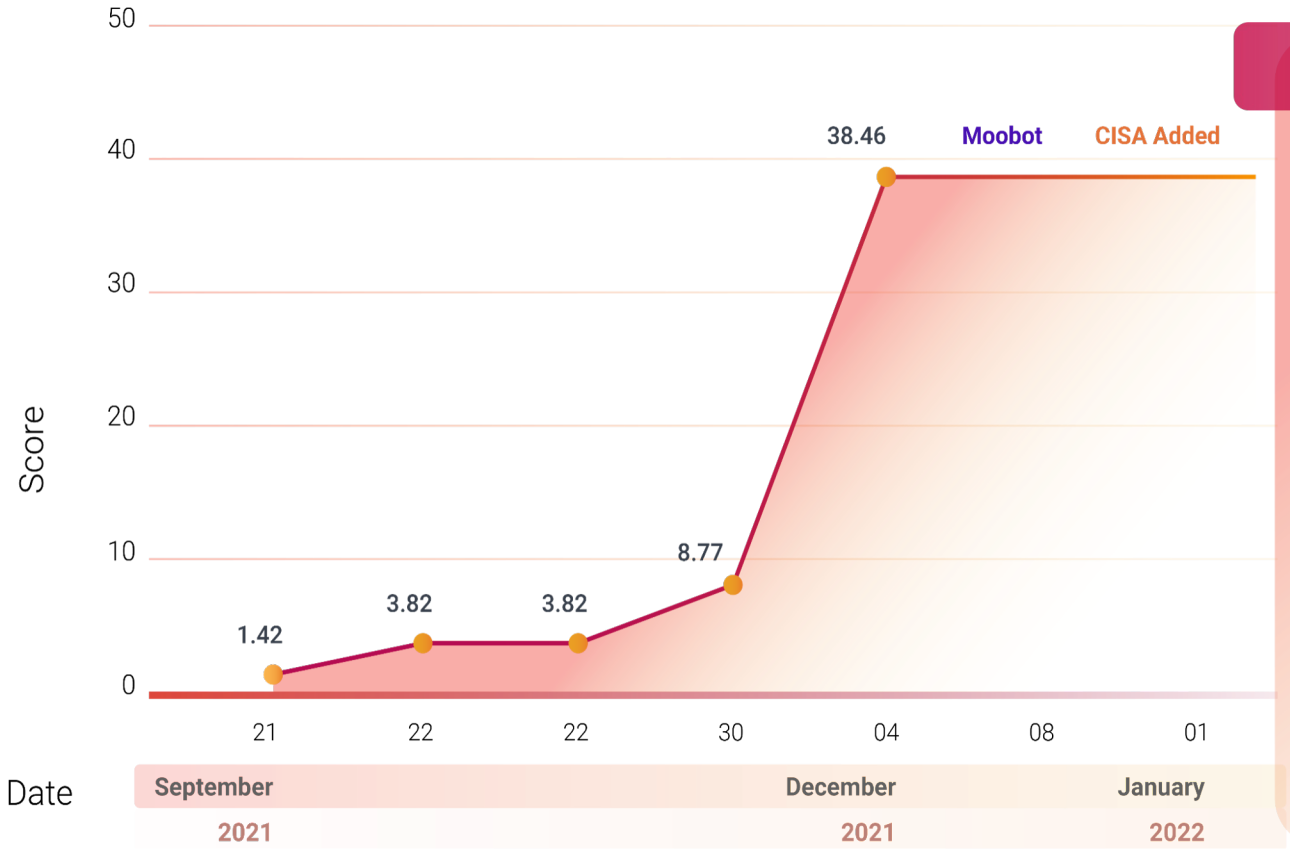


OTHERS*

Amcrest | Arm | Aviatrix | ChakraCore | FatPipe | Hikvision | IBM | LG | McAfee | Micro Focus | Mitel | MongoDB | Netgear | Siemens | SIMalliance | SonicWall | Sumavision | TP-Link | Trend Micro | TVT | WatchGuard | Yealink | Zyxel

False Sense of Security – If you can't Detect!

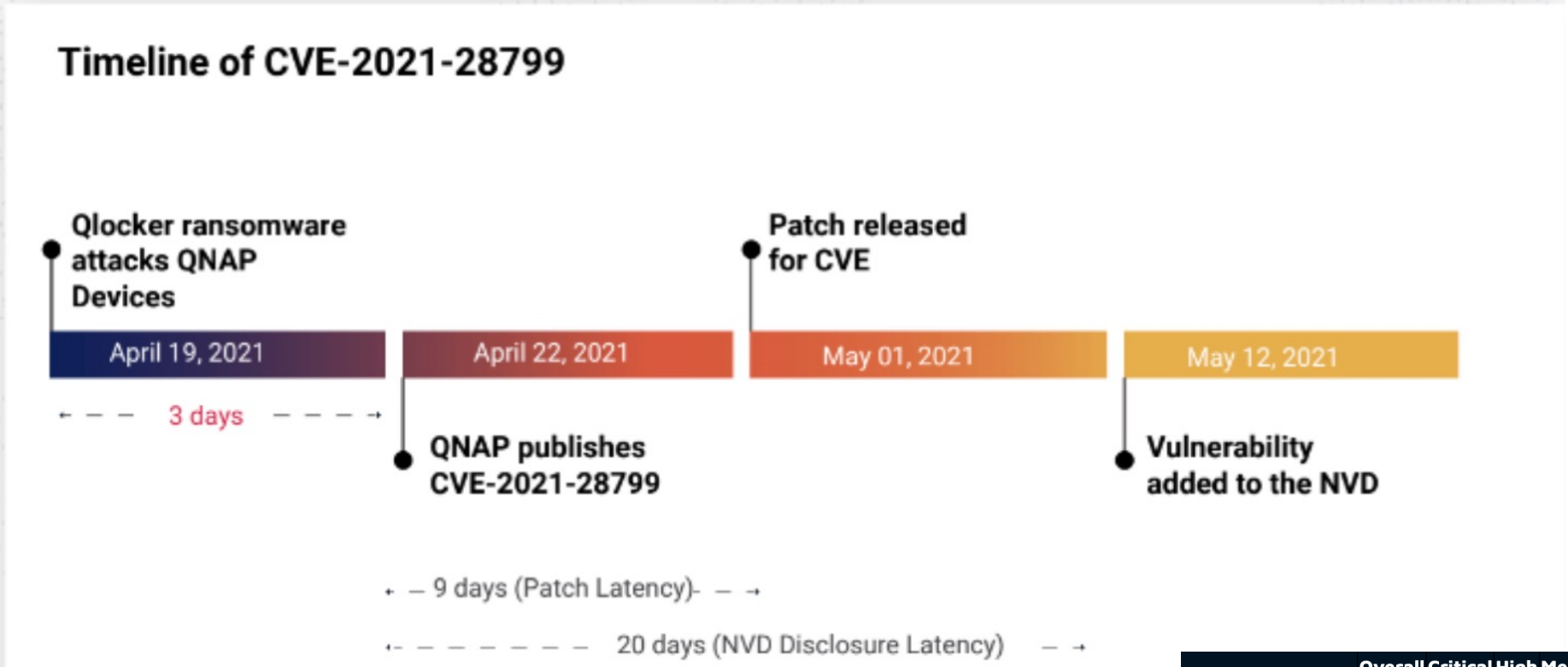
Instead of max rate (red arrow) - High Risk of Exploitability



Interesting Facts

- Vulnerability present in Hikvision products web server.
- Command injection attack by sending some messages with malicious commands could lead to RCE.
- Has No plugins across Nessus, Nexpose and Qualys.
- Google Trending between September 19-25, December 5-11 2021.
- Targeted by Moobot botnet, based on Mirai.
- Exploit available 33 Days after being added to NVD.
- Predicted as highly exploitable 37 Days before DHS CISA added the CVE to KEVs.
- Exploitability predicted 4 Days before reports of exploitation by Moobot.

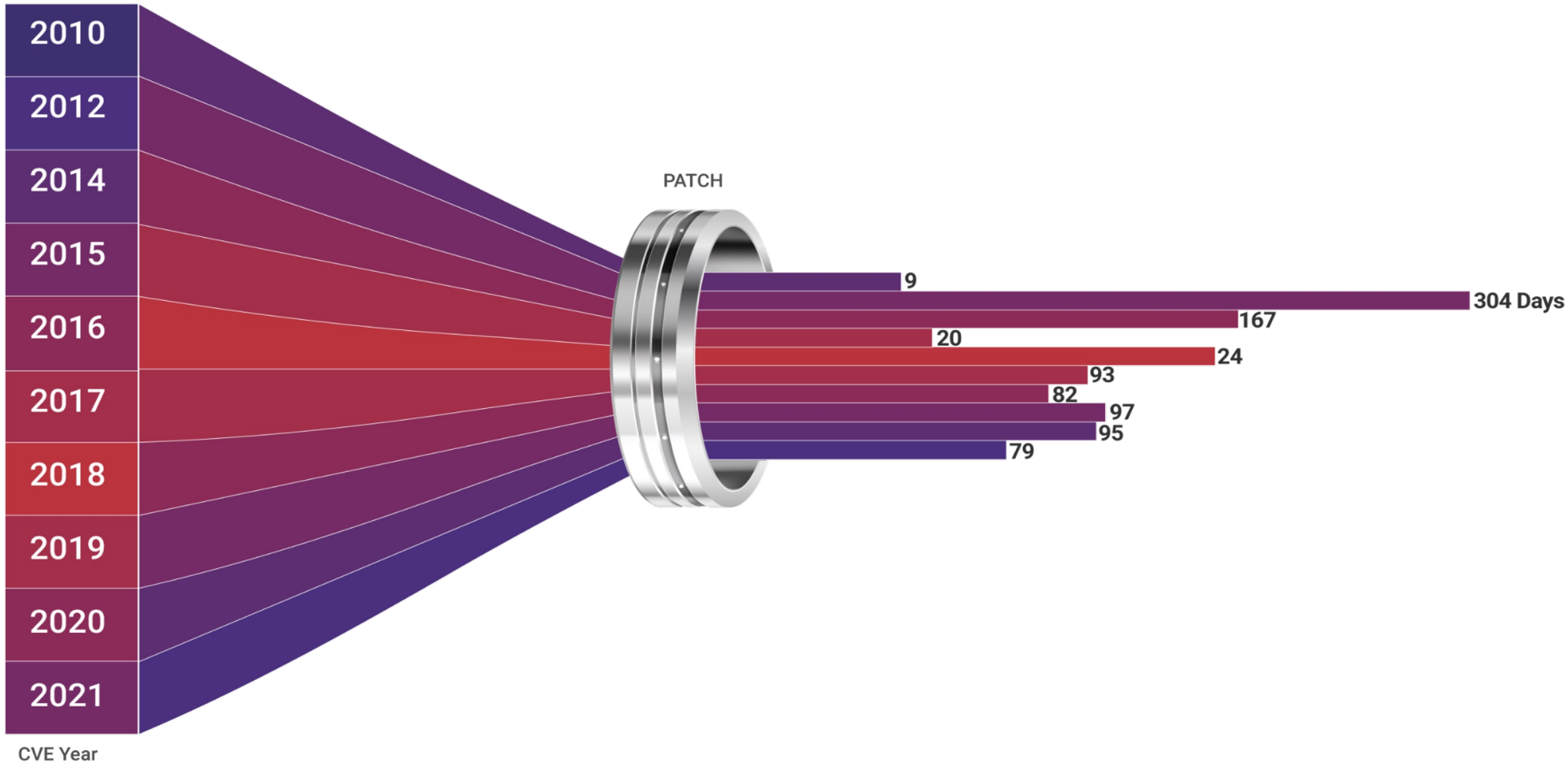
Vulnerability – Threat – Patch: Key Differentiation



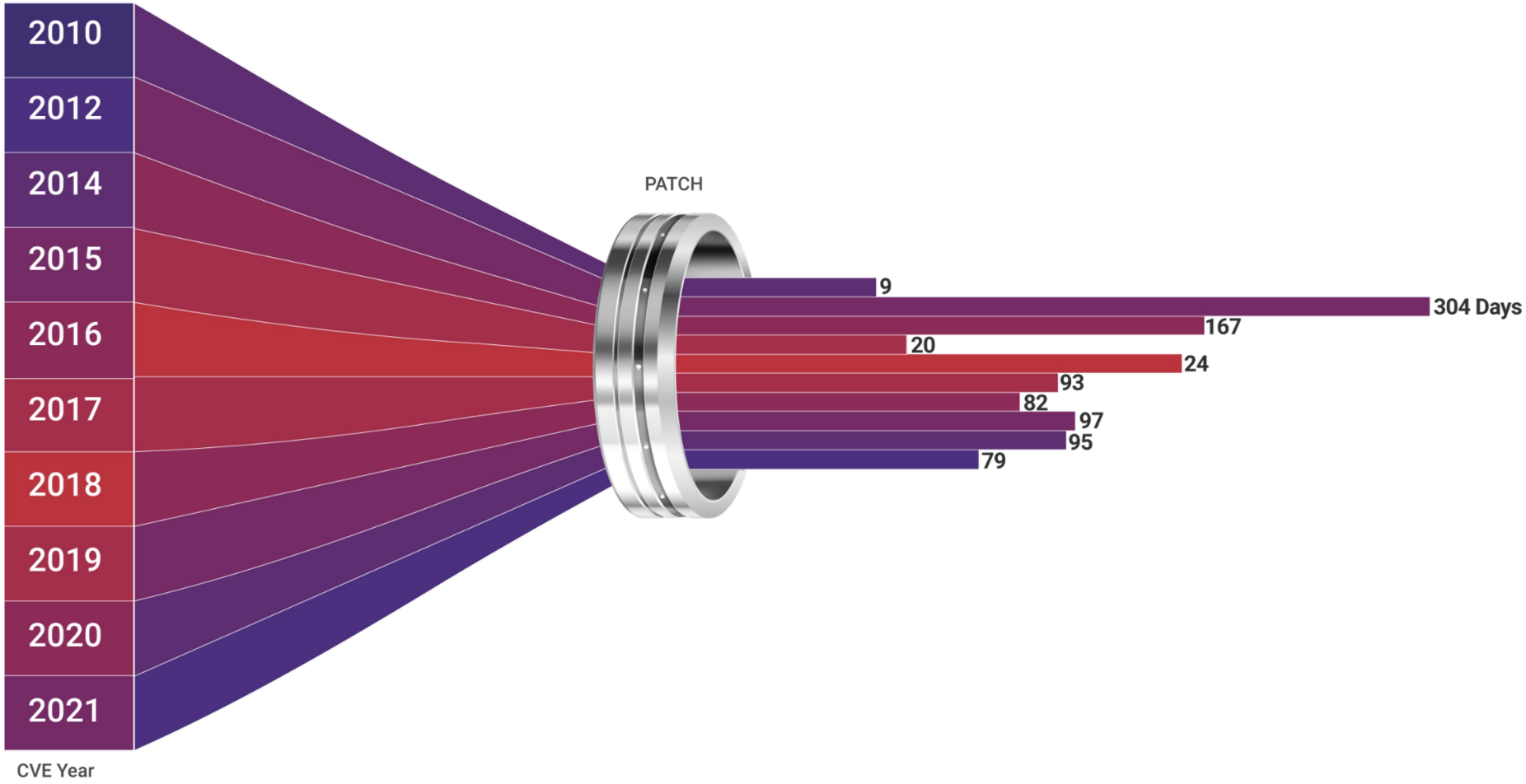
*A latency in Red indicates that the attack happened before the CVE was disclosed by the vendor.

	Overall	Critical	High	Medium	Low
Exploit before patch	78	39	34	4	1
Same day	43	23	20	-	-
Exploit after patch	150	69	71	10	-

Tail Winds – Latency is What Makes us Unique



Tail Winds – Average Time to Exploit DHS KEVs



Microsoft Deep Dive

CVEs from MSFT
[Docs MSFT 3,585; MSRC 5,668; Overlaps 504

8,749

CVEs from NVD With Microsoft as Vendor

9,527

CVE from NVD having Affected Vendor Product as Microsoft but the CVE was not reported by Microsoft.

1,635

Microsoft CVEs - The Funnel

Source	Total CVEs	Weaponized	RCE/PE	Ransomware	Trending	APT	CISA KEVs
CSW-VI	10,384	1,658	705	60*	6	109	218
NVD	9,527	1,593	684	59*	6	102	190
MSFT	8,749	1,163	518	60*	6	109	215

Interesting Nuggets

1635

CVEs directly affecting Microsoft's products were not reported by Microsoft in any of its advisories.

- CVE-2021-37705 (OneFuzz)
- CVE-2020-12695 (Xbox One, Windows 10)
- CVE-2020-10146 (Microsoft Teams)
- CVE-2017-17689 (Outlook)

25 Microsoft CVEs with ransomware associations were not reported by Microsoft in any of its advisories. Although none of them had direct relationships with Microsoft

03

CVEs in CISA KEVs were not reported by Microsoft in any of its advisories. These CVE had direct association with Microsoft products

- CVE-2002-0367 (Windows 2000, NT)
- CVE-2004-0210 (Windows 2000, NT)
- CVE-2020-16009 (Edge)

Microsoft vs. VI Exploitation Data

Total Records from MSRC
5766

Vulnerabilities
5668

Advisories
98

Microsoft

01 Reported Not Exploited - **5,100/5,668**
i.e. Vulnerabilities Less Likely, More Likely, Unlikely to be Exploited and Blanks

02 Exploited - Yes - **81/5,668**

03 Exploited - Blank - **487 / 5,668**

VI

01 Exploits - 425
RCE PE - 174
31 Ransomware Assn
84 CISA KEV

02 Exploits - **25**
RCE PE - 23
15 Ransomware Associations

03 Exploits - 1
20 CISA KEV

Thank you!