Scaling to the Adversary

Machine Learning Driven Mining of Threat Intel from the Darkweb

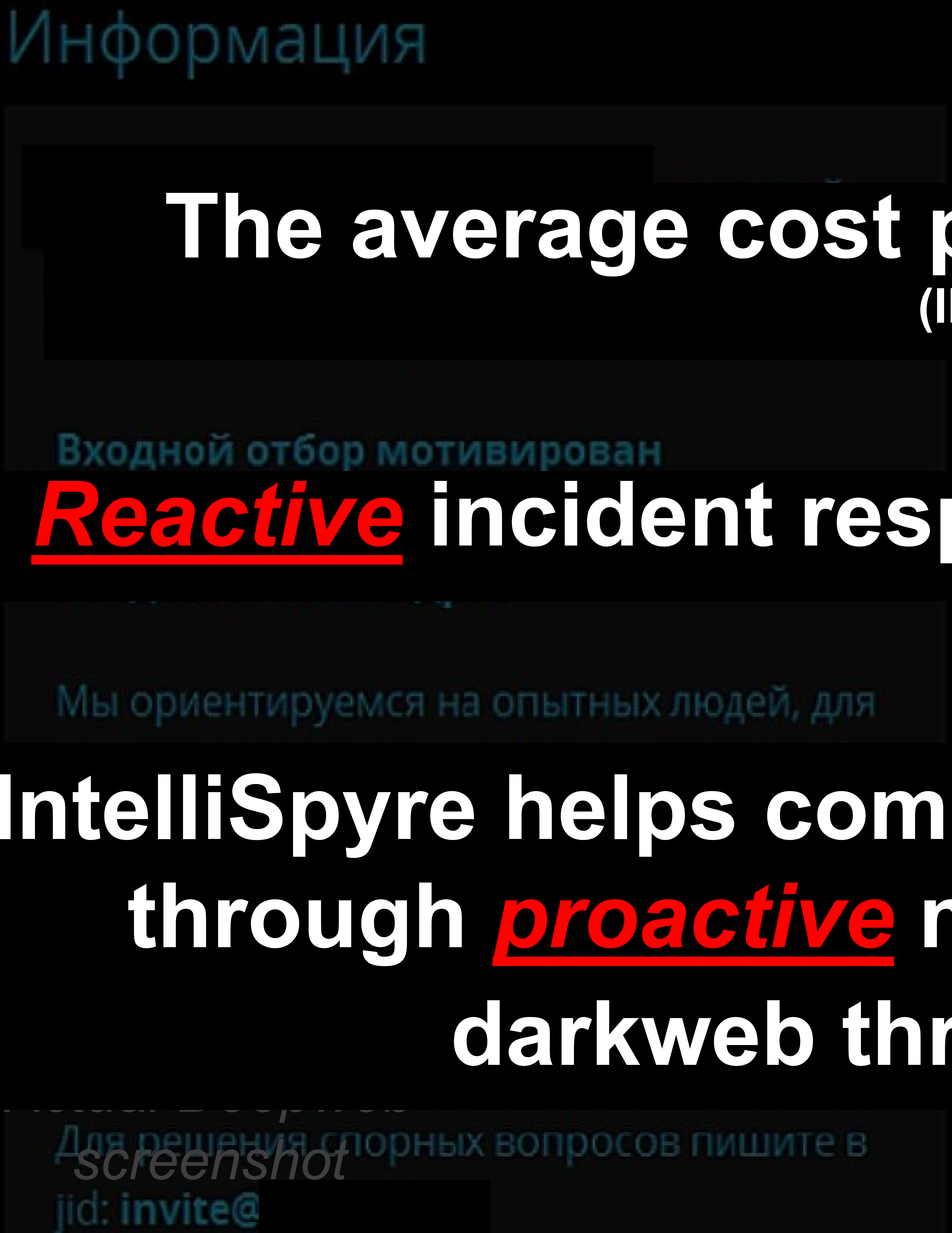
Paulo Shakarian, Ph.D.
CEO, IntelliSpyre, Inc.
Fulton Entrepreneurial Professor, Arizona State University

The average cost per data breach is $4M.

(IBM, 2016)

Reactive incident response to attacks is costly.

IntelliSpyre helps companies avoid cyber attacks through ***proactive*** machine-learning driven darkweb threat intelligence.



Информация

[redacted]

Входной отбор мотивирован

Мы ориентируемся на опытных людей, для

Для решения спорных вопросов пишите в

jid: invite@[redacted]

Actual Deepweb screenshot

Wanted: experienced hacker who is good

Started by jacky-intel

I offer Hack service "professional" new

Started by careservice

Started by mandex

Need UK email addresses MILLIONS ne

Started by wellwisher

Need another hacker to help me hack ne

Actual Darkweb screenshot

Imminent/directed

"A hacktivist group is launching a campaign against company X."

Change to Threat Landscape

"A 0-day for the latest build of a certain opiating system is available"

Change to Threat Landscape

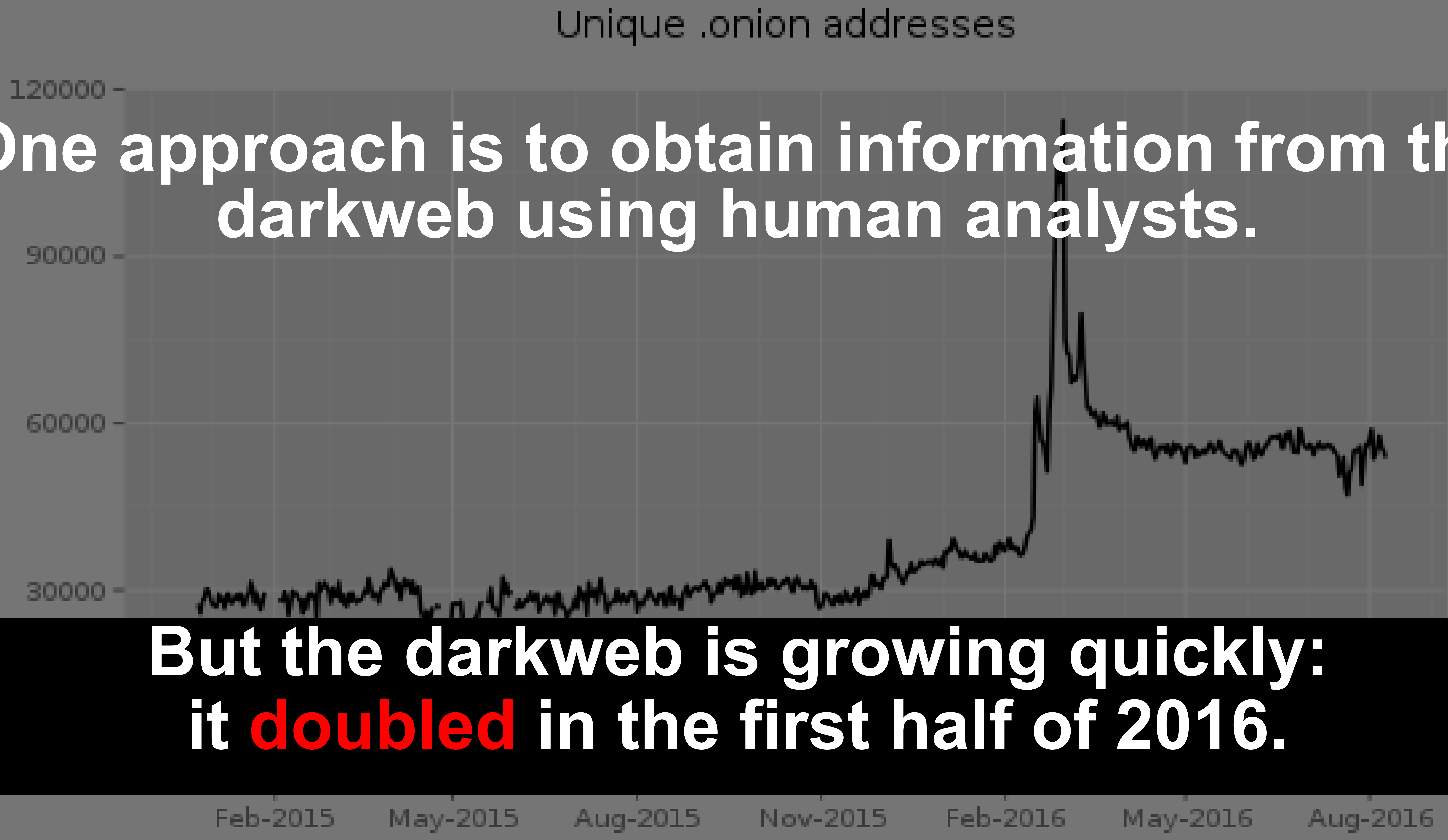
"The price of Android exploits dropped."

Change to Threat Landscape

"A prolific darkweb forum poster advertised his first zero-day for sale."

Situational Awareness	Most current “threat intel”	Immediate action (i.e. block IP address)	Info sharing, honeypot data, Mandiant, etc.
Imminent, directed threats	Most current “darkweb intel”	Prepare for cyber-attack (i.e. DDoS)	Service providers monitor for mention of specific company
Change to threat landscape Atmospherics	Very few – mostly in R&D and academia	Allow for more strategic decisions (i.e. not using certain software)	Involves ingesting multiple sources close to hackers, necessitates machine learning, artificial intelligence, and related techniques

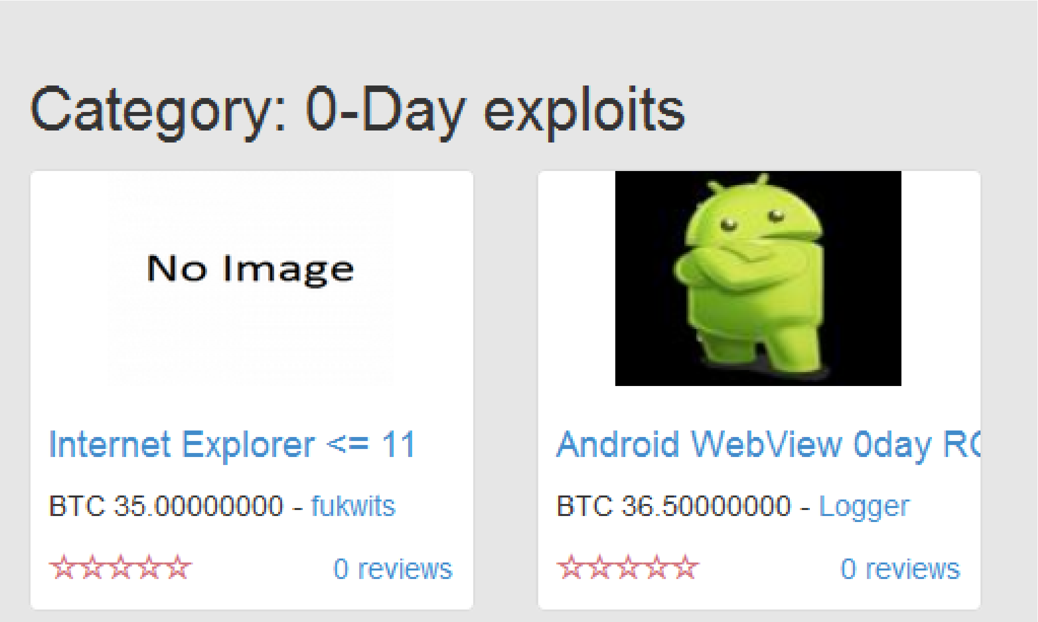
One approach is to obtain information from the darkweb using human analysts.



But the darkweb is growing quickly: it doubled in the first half of 2016.

The Tor Project - https://metrics.torproject.org/

Actual darkweb screenshot



Category: 0-Day exploits

No Image

Internet Explorer <= 11

BTC 35.00000000 - fukwits

☆☆☆☆☆ 0 reviews

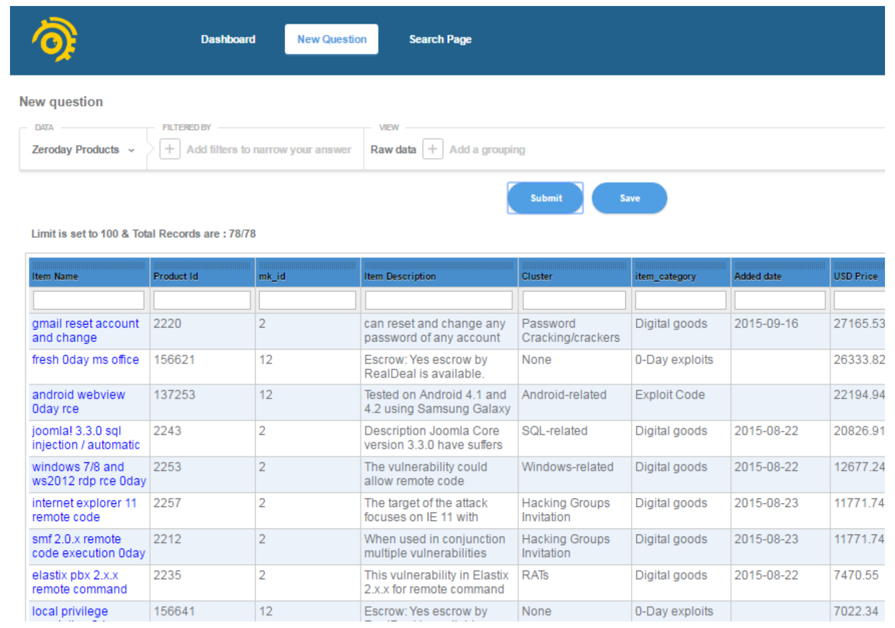
Android WebView 0day RC

BTC 36.50000000 - Logger

☆☆☆☆☆ 0 reviews

Team members with cultural and linguistic skills identify malicious hacking pages

Proprietary data mining and machine learning techniques automatically and regularly obtain information



New question

DATA: Zeroday Products

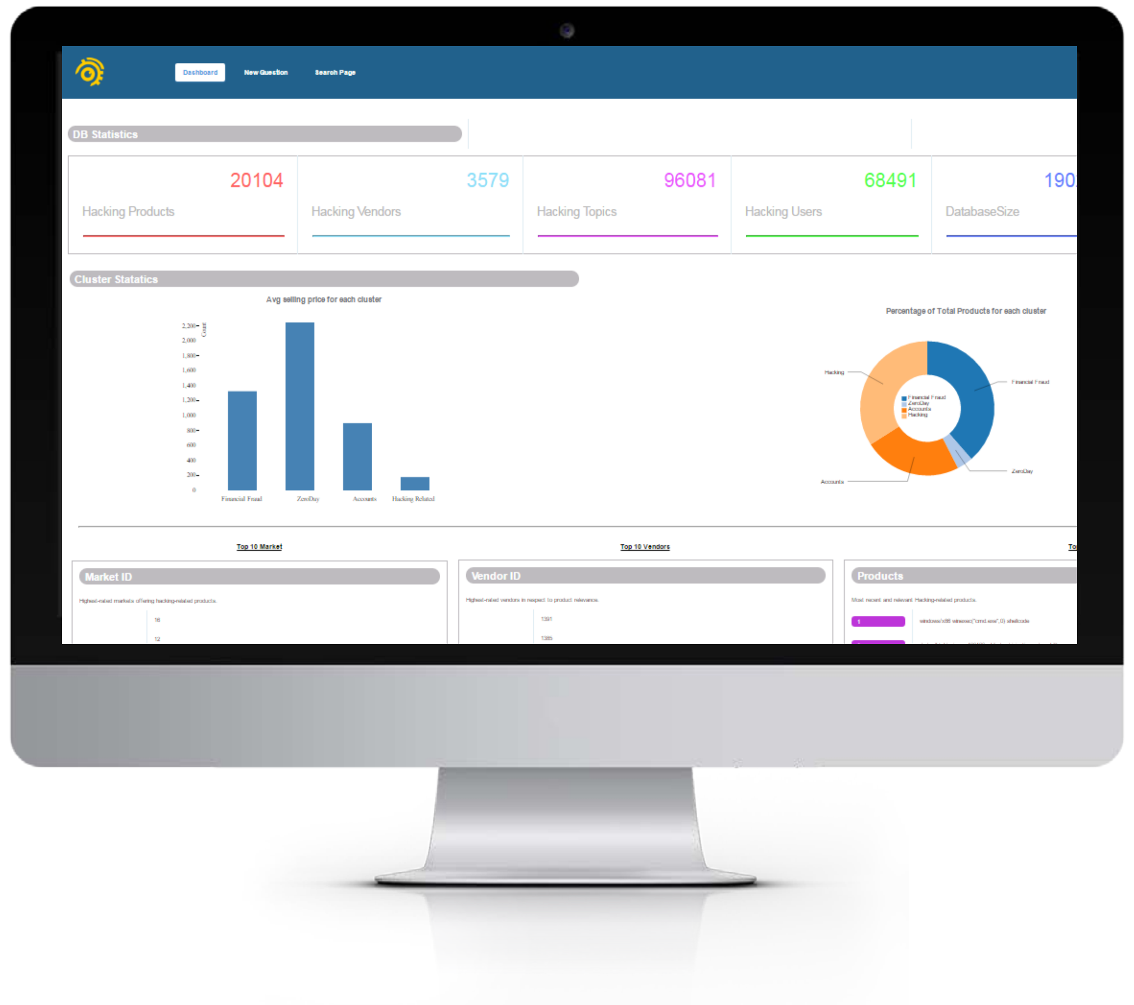
FILTERED BY: Add filters to narrow your answer

VIEW: Raw data | Add a grouping

Limit is set to 100 & Total Records are : 78/78

Item Name	Product Id	mk_id	Item Description	Cluster	item_category	Added date	USD Price
gmail reset account and change	2220	2	can reset and change any password of any account	Password Cracking/crackers	Digital goods	2015-09-16	27165.53
fresh 0day ms office	156621	12	Escrow: Yes escrow by RealDeal is available.	None	0-Day exploits		26333.82
android webview 0day rce	137253	12	Tested on Android 4.1 and 4.2 using Samsung Galaxy	Android-related	Exploit Code		22194.94
joomla! 3.3.0 sql injection / automatic	2243	2	Description Joomla Core version 3.3.0 have suffers	SQL-related	Digital goods	2015-08-22	20826.91
windows 7/8 and ws2012 rdp rce 0day	2253	2	The vulnerability could allow remote code	Windows-related	Digital goods	2015-08-22	12677.24
internet explorer 11 remote code	2257	2	The target of the attack focuses on IE 11 with	Hacking Groups Invitation	Digital goods	2015-08-23	11771.74
smf 2.0.x remote code execution 0day	2212	2	When used in conjunction multiple vulnerabilities	Hacking Groups Invitation	Digital goods	2015-08-23	11771.74
elastix pbx 2.x.x remote command	2235	2	This vulnerability in Elastix 2.x.x for remote command	RATs	Digital goods	2015-08-22	7470.55
local privilege	156641	12	Escrow: Yes escrow by	None	0-Day exploits		7022.34

Information stored in a unified database schema allows queries across multiple darkweb sources



DB Statistics

20104	3579	96081	68491	190
Hacking Products	Hacking Vendors	Hacking Topics	Hacking Users	DatabaseSize

Cluster Statatics

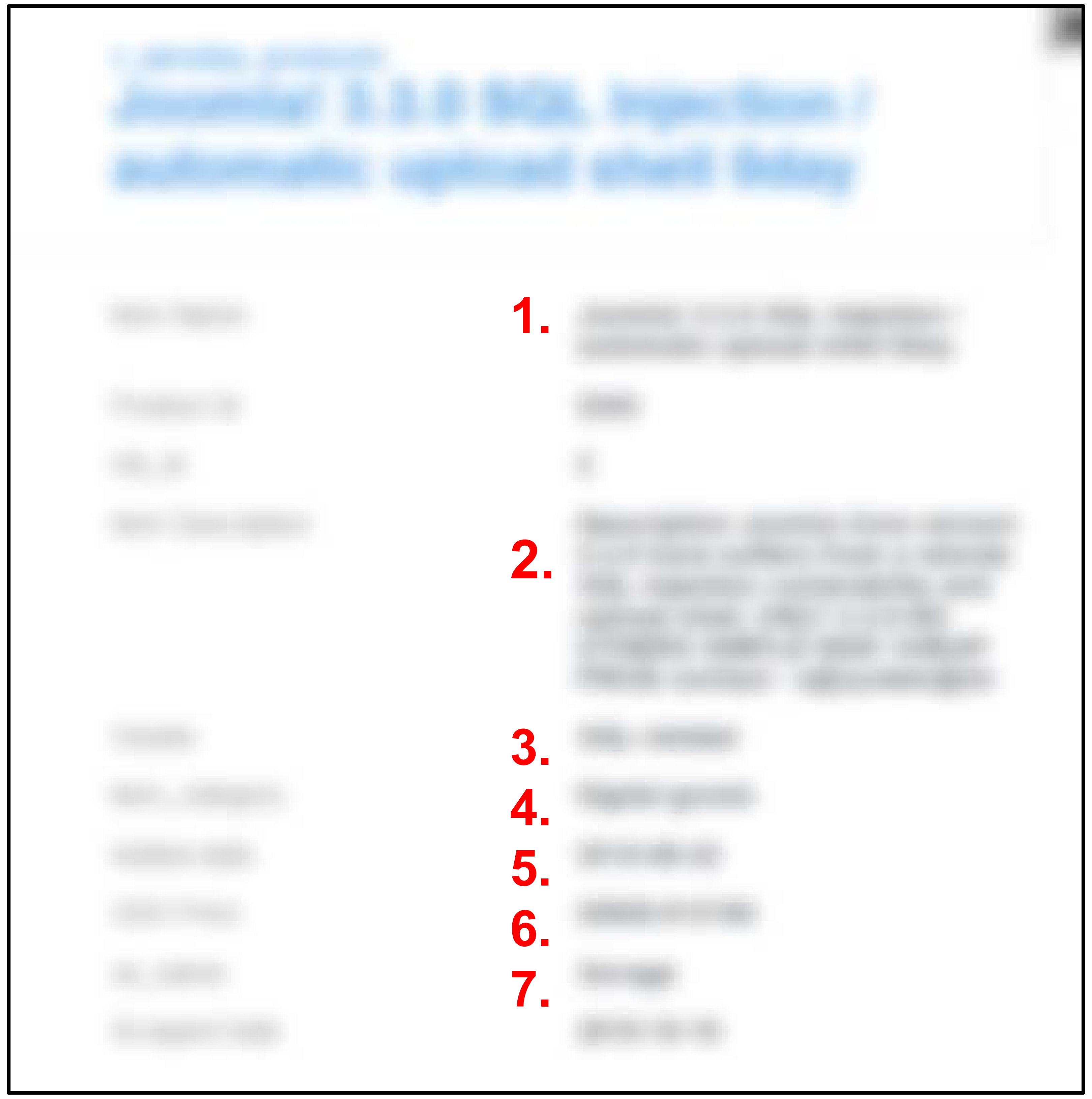
SaaS-based front end and standards-based API

IntelliSpyre

Key Technology

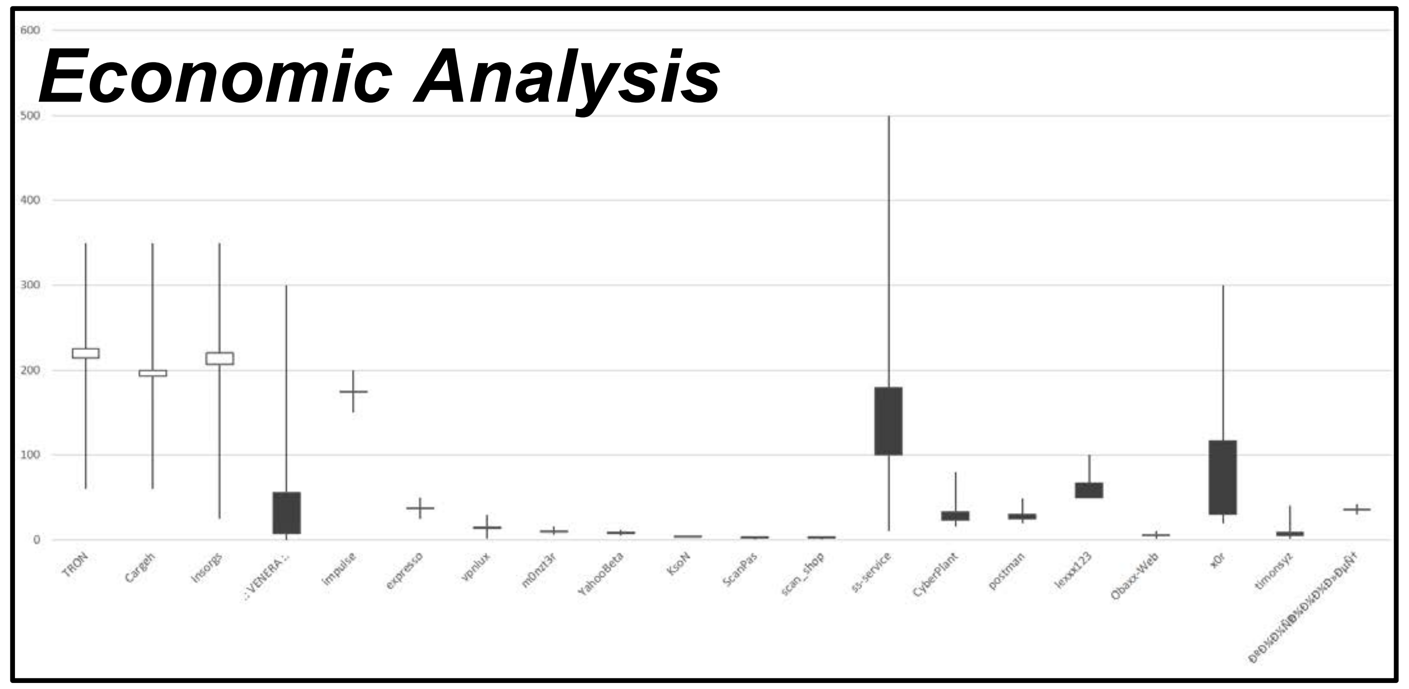
Automatically obtain entities from darkweb.
NO manual extraction.

Allows for automatic analysis not performed elsewhere.



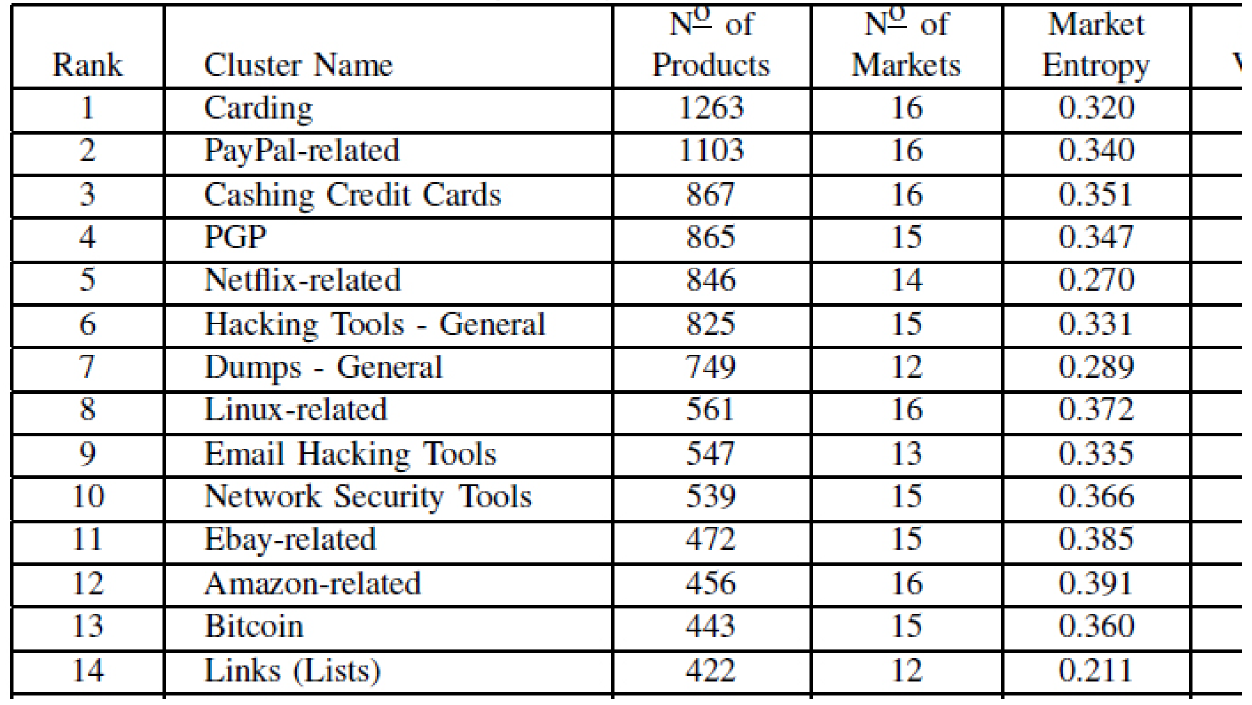
IntelliSpyre *SpyrePortal* platform screenshot

1. **Zero Day Name**
2. **Description**
3. **Auto-identified category**
4. **Category on darkweb site**
5. **Date posted to darkweb**
6. **Price**
7. **Vendor Name**



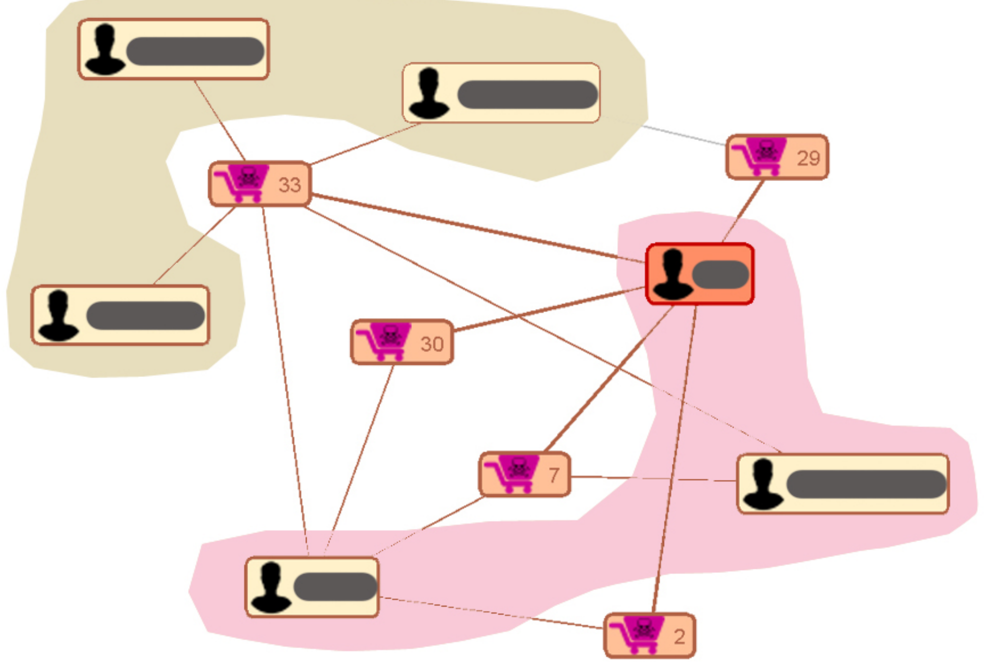
Economic Analysis

Product Category Identification



Rank	Cluster Name	Nº of Products	Nº of Markets	Market Entropy	V
1	Carding	1263	16	0.320	
2	PayPal-related	1103	16	0.340	
3	Cashing Credit Cards	867	16	0.351	
4	PGP	865	15	0.347	
5	Netflix-related	846	14	0.270	
6	Hacking Tools - General	825	15	0.331	
7	Dumps - General	749	12	0.289	
8	Linux-related	561	16	0.372	
9	Email Hacking Tools	547	13	0.335	
10	Network Security Tools	539	15	0.366	
11	Ebay-related	472	15	0.385	
12	Amazon-related	456	16	0.391	
13	Bitcoin	443	15	0.360	
14	Links (Lists)	422	12	0.211	

Social Network Analysis

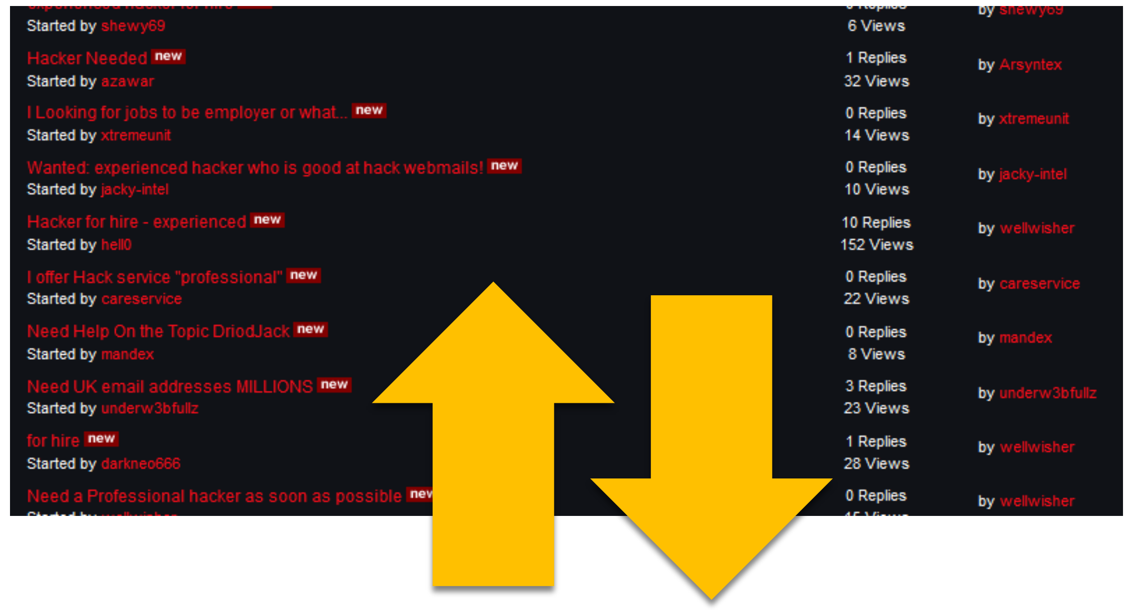


U.S. Provisional Patent 62/409,291

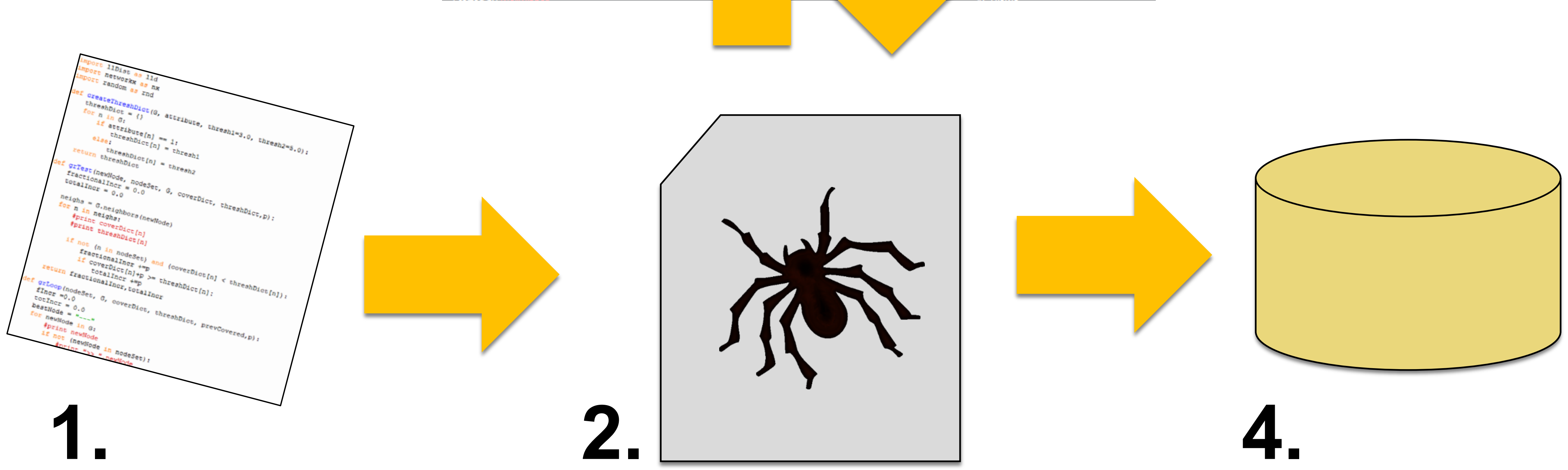
Key Technology (Backend)

Our backend allows for significant reduction in manpower.

Our system greatly reduces the need for customization.



Started by shewy69	6 Views	by shewy69
Hacker Needed new Started by azawar	1 Replies 32 Views	by Arsyntex
I Looking for jobs to be employer or what... new Started by xtremeunit	0 Replies 14 Views	by xtremeunit
Wanted: experienced hacker who is good at hack webmails! new Started by jacky-intel	0 Replies 10 Views	by jacky-intel
Hacker for hire - experienced new Started by hell0	10 Replies 152 Views	by wellwisher
I offer Hack service "professional" new Started by careservice	0 Replies 22 Views	by careservice
Need Help On the Topic DriodJack new Started by mandex	0 Replies 8 Views	by mandex
Need UK email addresses MILLIONS new Started by underw3bfullz	3 Replies 23 Views	by underw3bfullz
for hire new Started by darkneo666	1 Replies 28 Views	by wellwisher
Need a Professional hacker as soon as possible new	0 Replies	by wellwisher

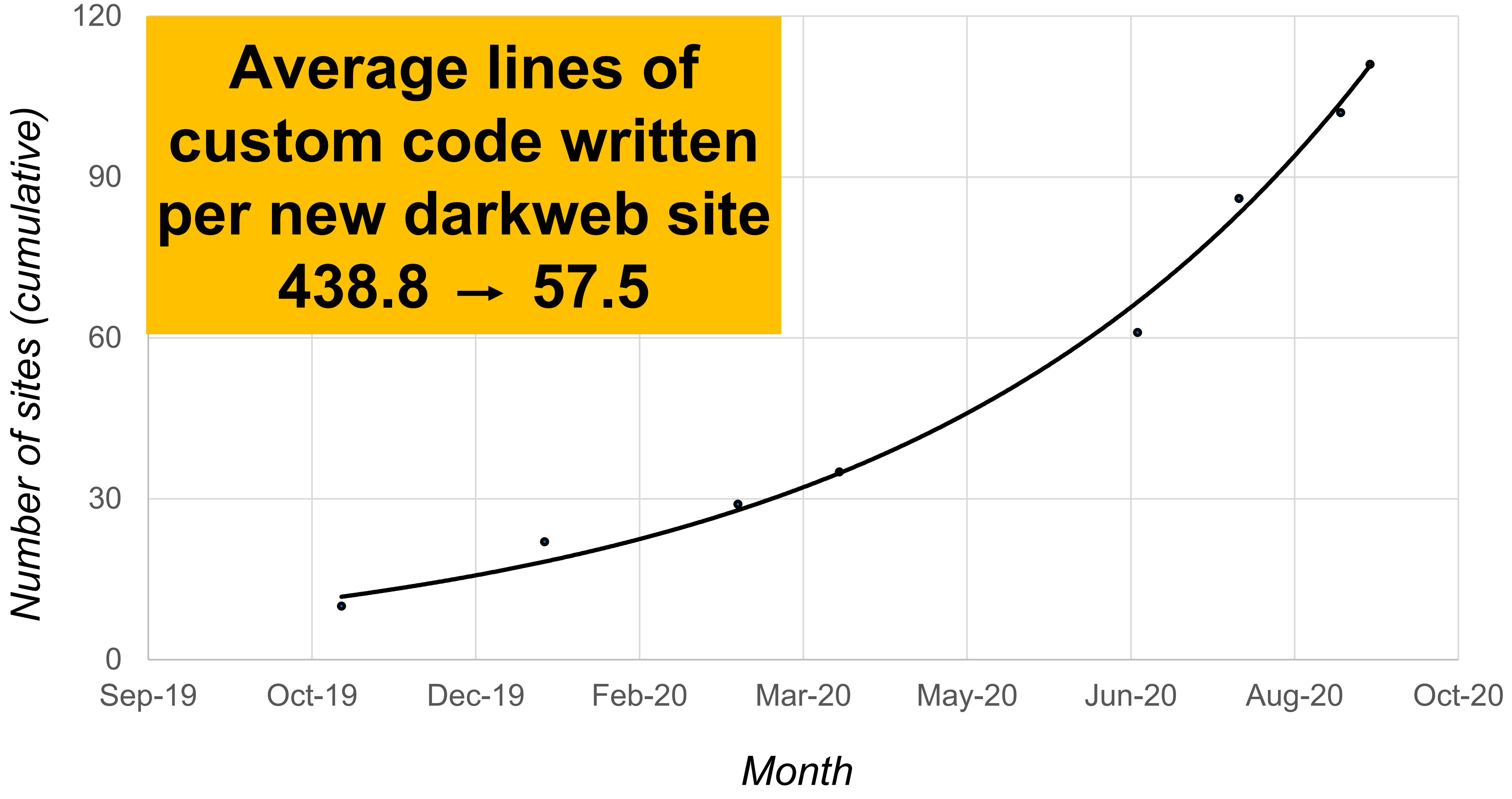


```
import llDist as lld
import networkx as nx
import random as rnd
def createThreshDict(G, attribute, thresh1=3.0, thresh2=5.0):
    threshDict = {}
    for n in G:
        if attribute[n] == 1:
            threshDict[n] = thresh1
        else:
            threshDict[n] = thresh2
    return threshDict
def grTest(newNode, nodeSet, G, coverDict, threshDict,p):
    fractionalIncr = 0.0
    totalIncr = 0.0

    neighs = G.neighbors(newNode)
    for n in neighs:
        #print coverDict[n]
        #print threshDict[n]

        if not (n in nodeSet) and (coverDict[n] < threshDict[n]):
            fractionalIncr +=p
            if coverDict[n]+p >= threshDict[n]:
                totalIncr +=p
    return fractionalIncr,totalIncr
def grLoop(nodeSet, G, coverDict, threshDict, prevCovered,p):
    fIncr =0.0
    totIncr = 0.0
    bestNode = "---"
    for newNode in G:
        #print newNode
        if not (newNode in nodeSet):
            #print ">> ",newNode
```

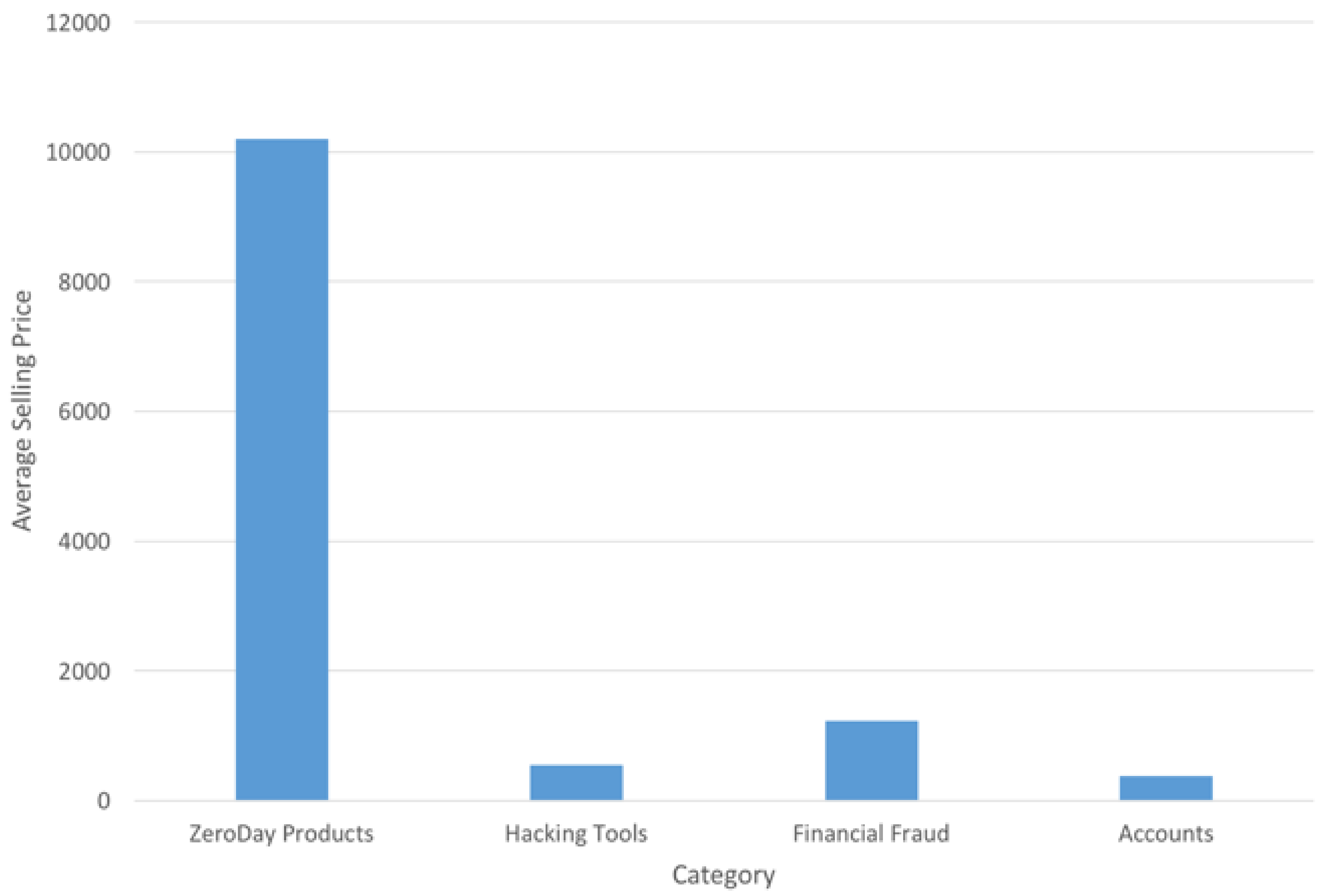
1. **Small amount of custom code modules needed**
2. **Plugs into crawler/parser framework**
3. **Repeated crawling of darkweb/deepweb hacking sites**
4. **Data stored in normalized database schema**

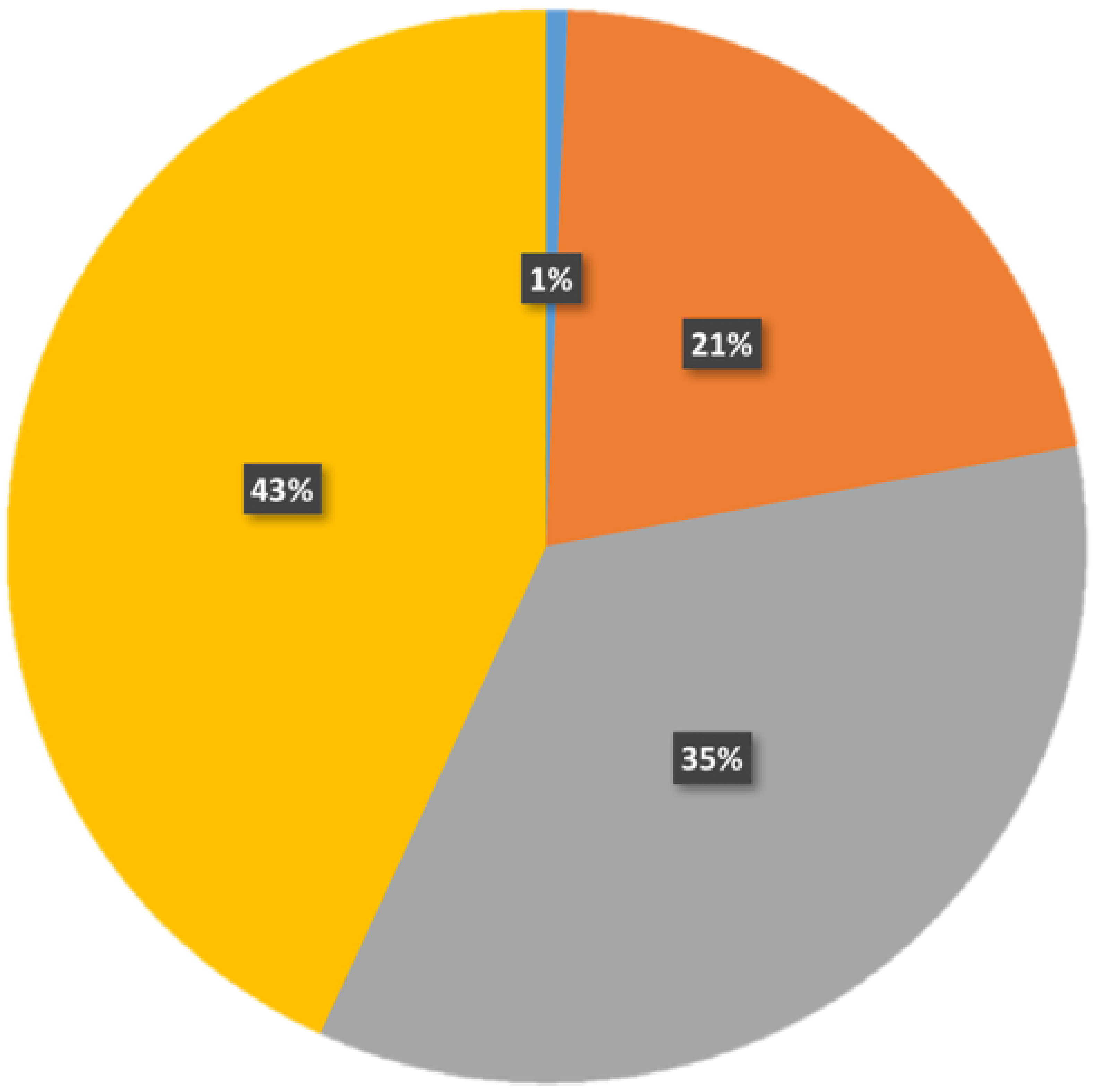


Average lines of custom code written per new darkweb site 438.8 → 57.5

U.S. Provisional Patent 62/409,291

Products Catalogued





Markets often sell goods and services that do not relate to malicious hacking, including drugs, pornography, weapons and software services. Similar trend for forum discussions.

Only a small fraction of data (13%) are related to malicious hacking.

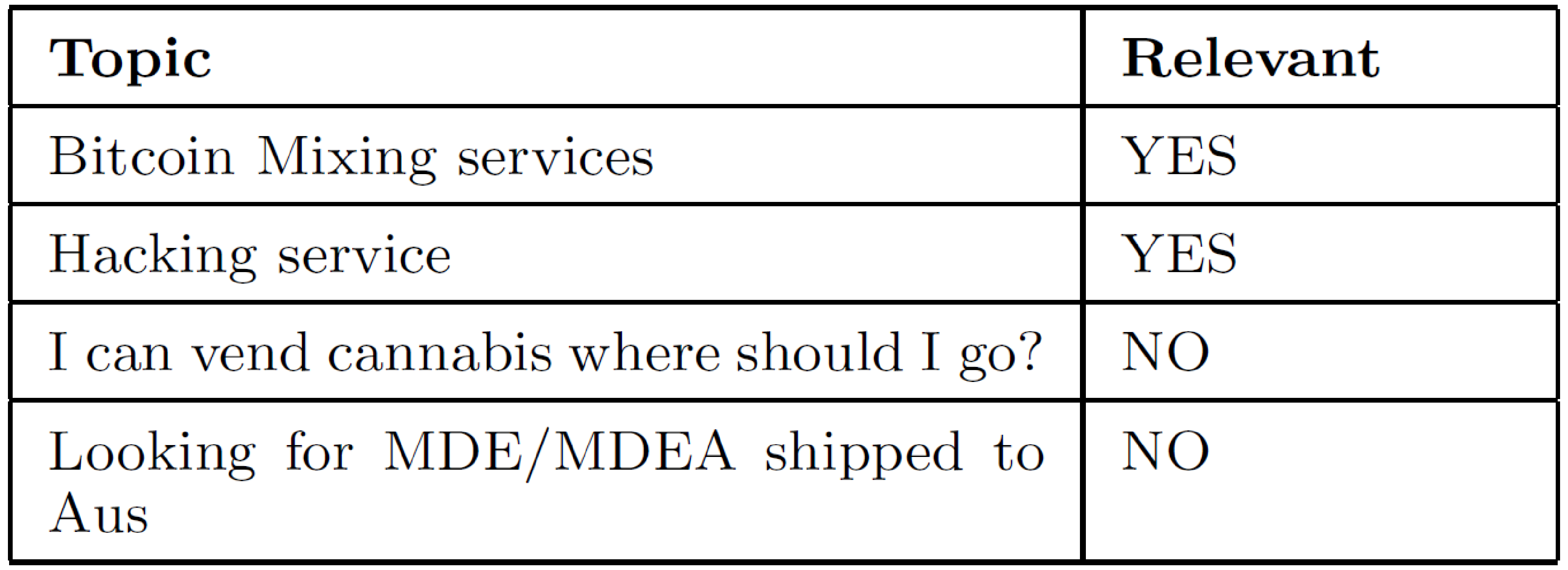
Filtering Challenges

Text Cleaning – removal of all alpha-numeric characters in tandem with stop-word removal.

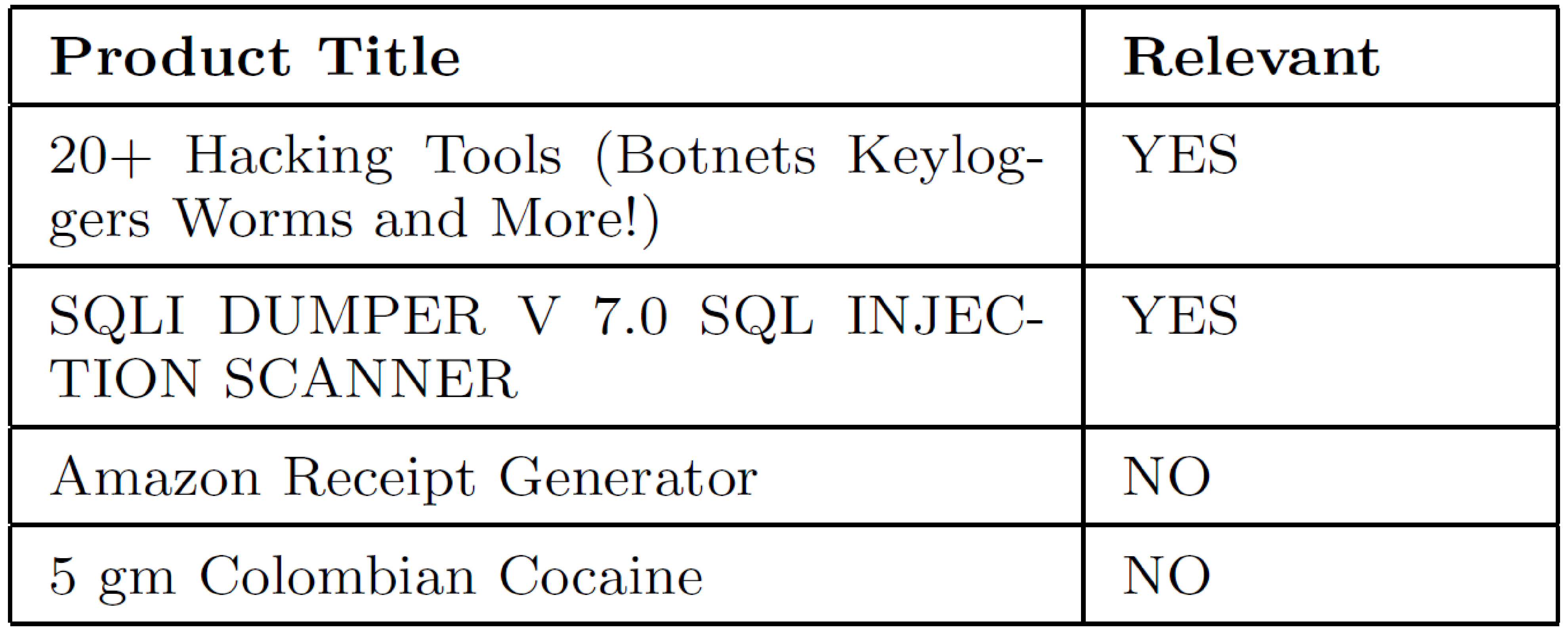
Misspellings and Word Variations – in bag-of-words approach, variations of words are considered separately (e.g. hacker, hack, hackers, etc.). We use character n-grams in range(3, 5) to look for frequently grouped characters instead of words.

Large Feature Space – feature matrix gets very large as the number of words increase (much larger for character n-grams). Use sparse matrix representation.

Analyze title and description separately to preserve context.



Topic	Relevant
Bitcoin Mixing services	YES
Hacking service	YES
I can vend cannabis where should I go?	NO
Looking for MDE/MDEA shipped to Aus	NO

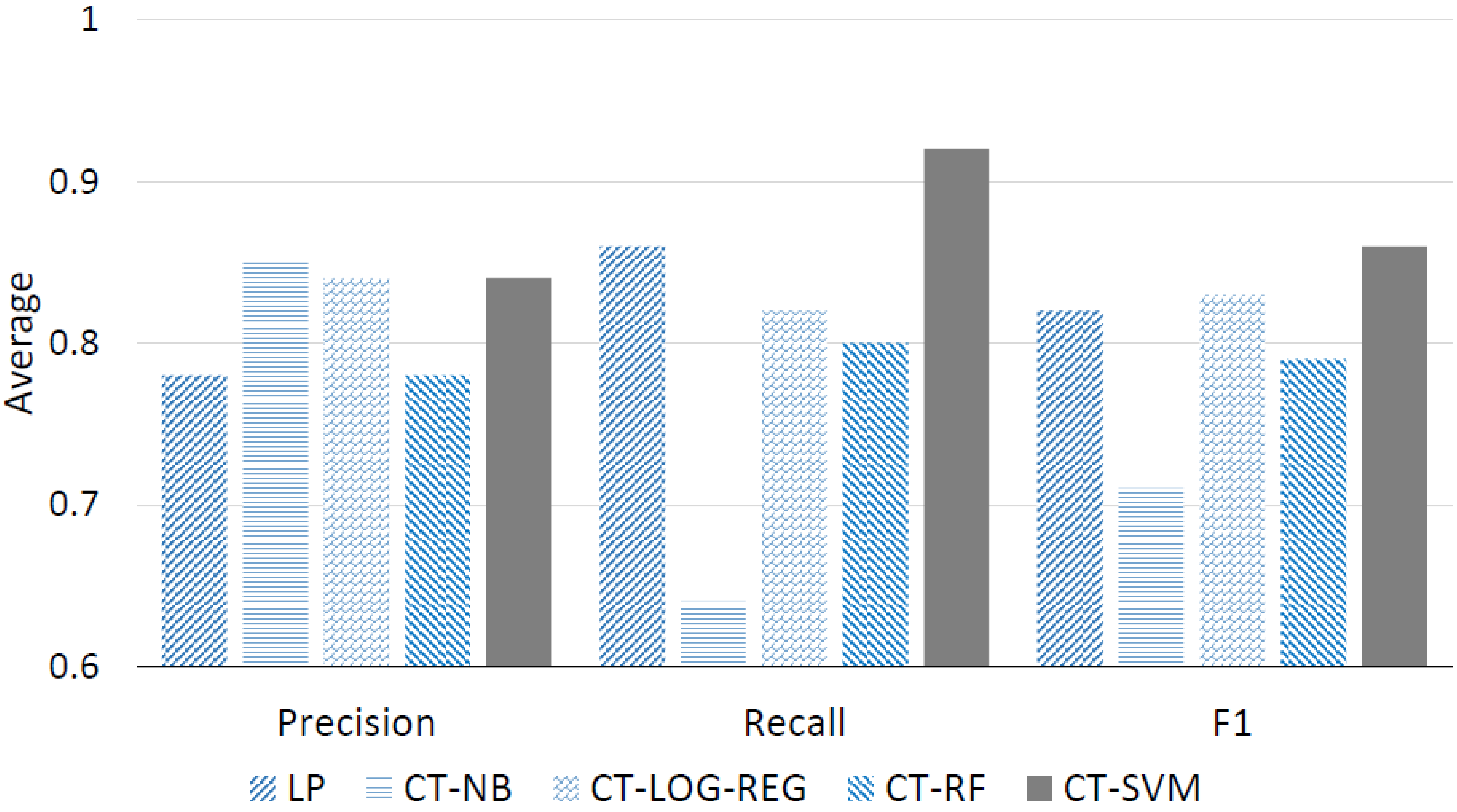


Product Title	Relevant
20+ Hacking Tools (Botnets Keyloggers Worms and More!)	YES
SQLI DUMPER V 7.0 SQL INJECTION SCANNER	YES
Amazon Receipt Generator	NO
5 gm Colombian Cocaine	NO

Filtering and Cleaning Information

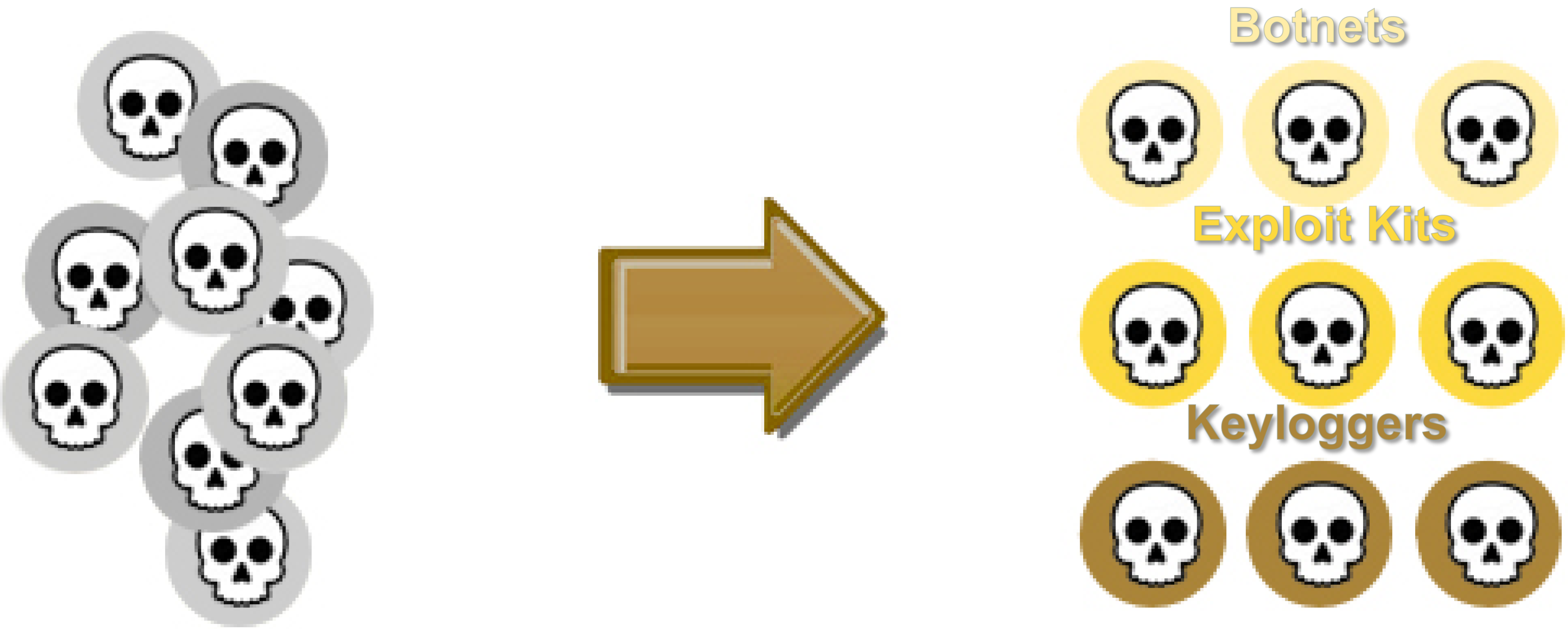
Refined and tested machine learning models separate noise from important cyber threat information.

We achieve over 90% recall on malicious hacking items (malware, exploits, etc.) while while minimizing false positives.



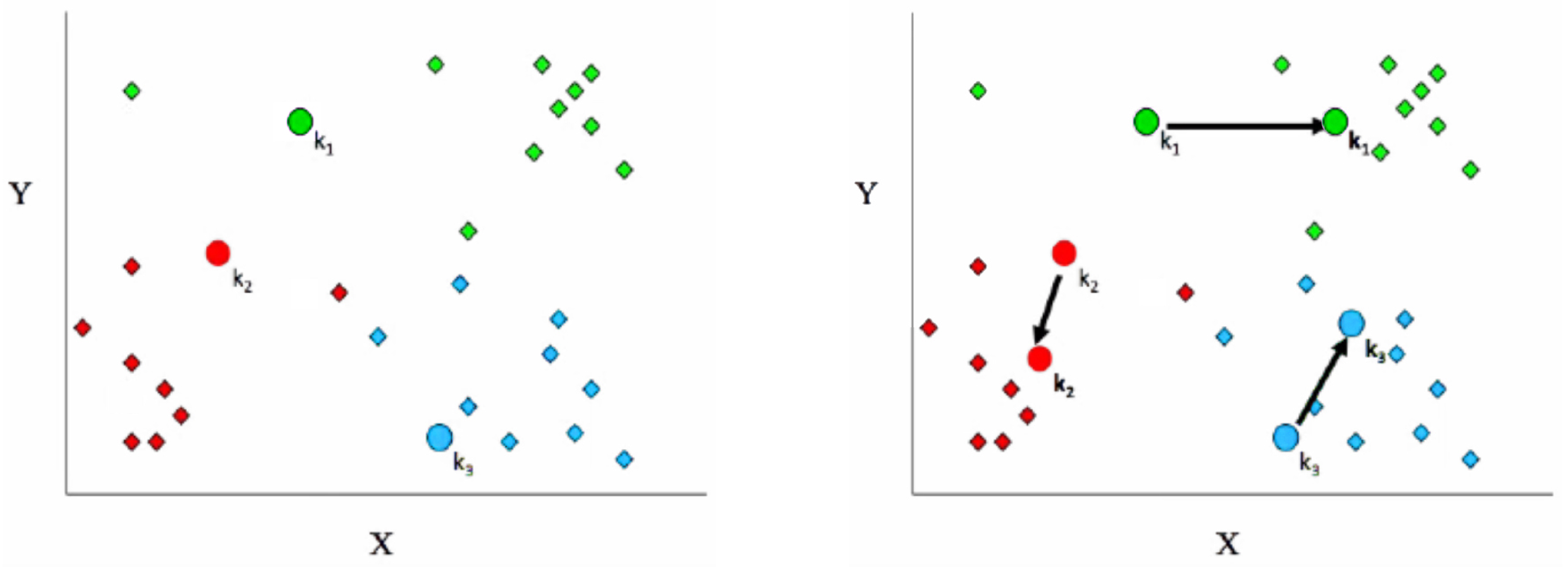
Product Categorization

Use a join of manual labeling and unsupervised clustering to get the desired categorization and specialization.



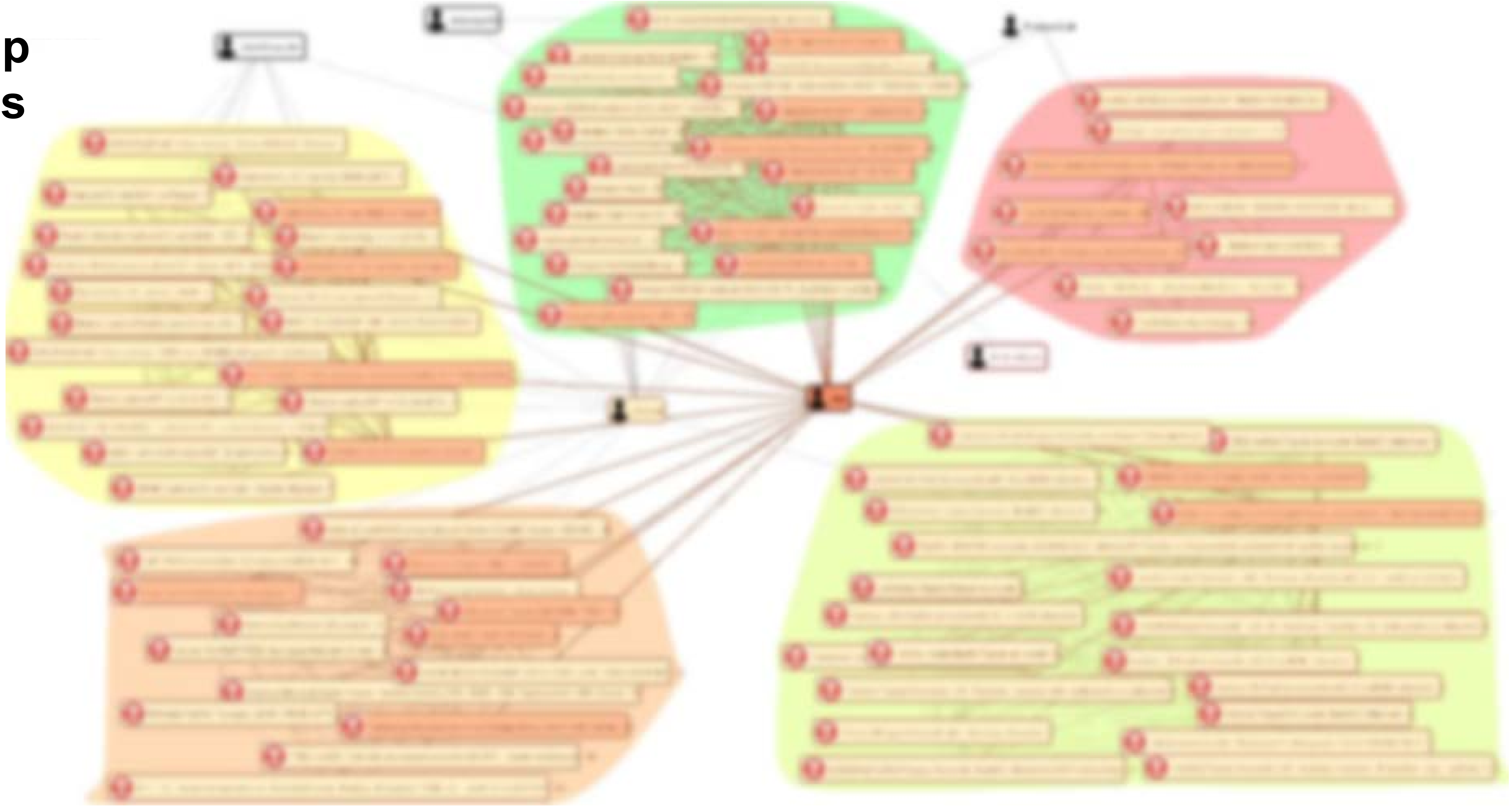
Clustering Strategy

Using a clustering strategy, we group items into categories and continualy refine at each step.



Automated Data Tagging

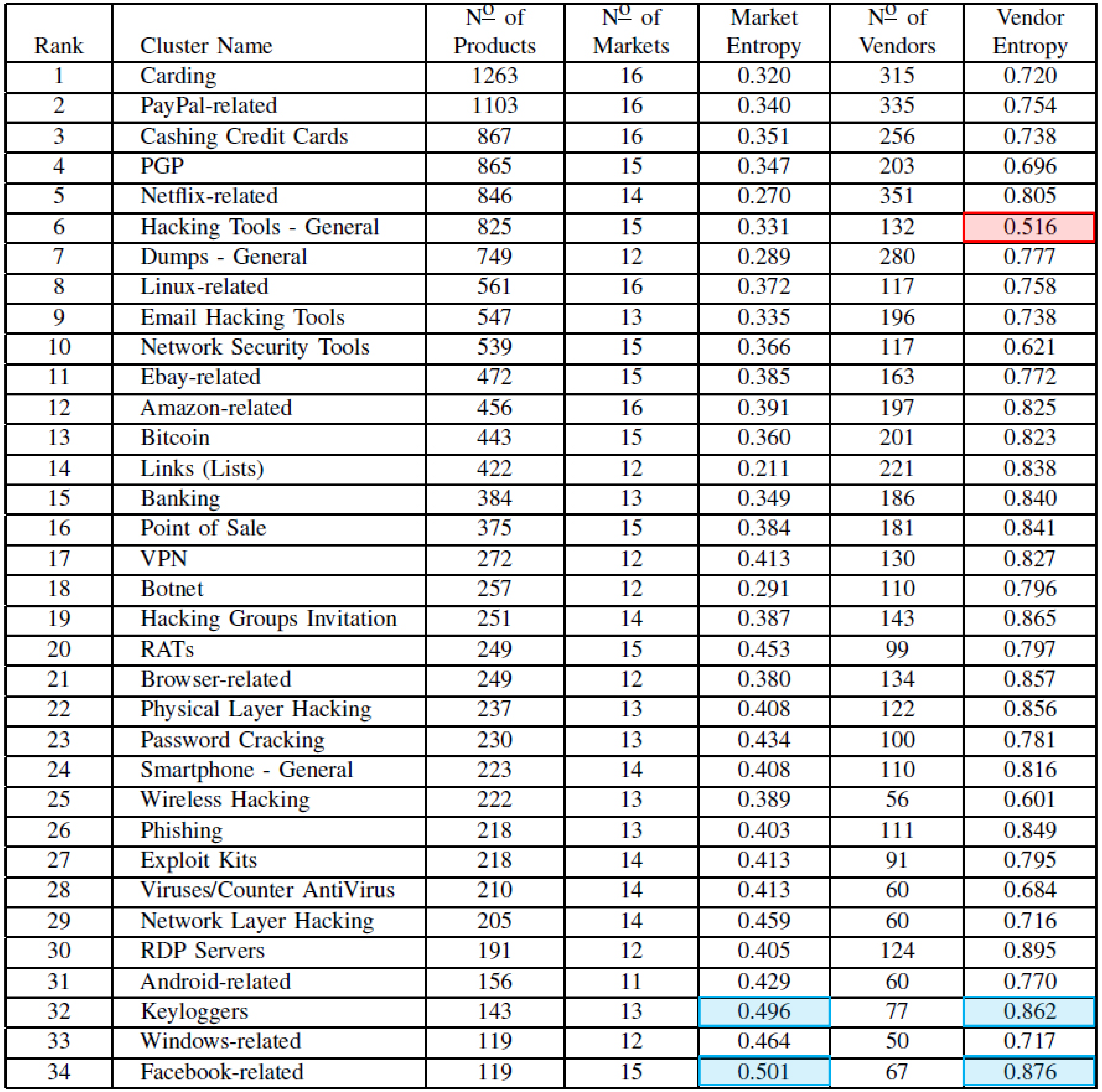
Unsupervised methods to group hacking products into categories



Hacking Product Analysis

Facebook: 119 Products. 67 Vendors. Most prolific vendor has 8 Products. Products spread across 15 Markets. Most well-represented Market has 30 Products.

Keyloggers: widespread prevalence of them. It is a well-established hacking technique.



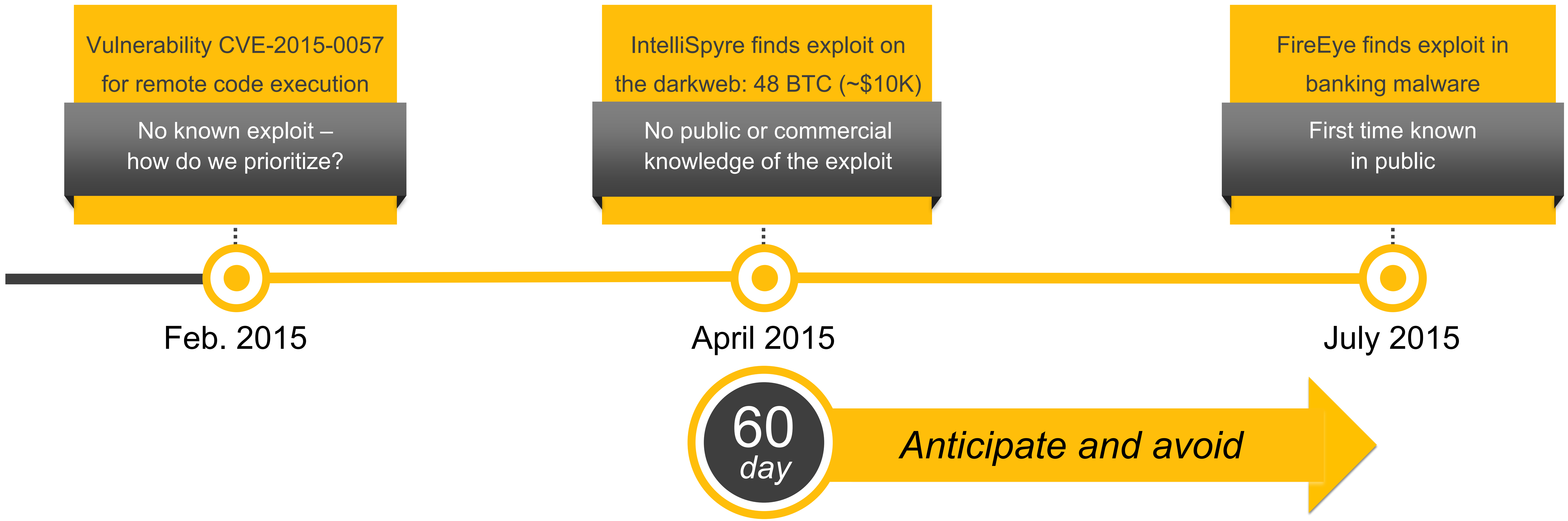
Rank	Cluster Name	Nº of Products	Nº of Markets	Market Entropy	Nº of Vendors	Vendor Entropy
1	Carding	1263	16	0.320	315	0.720
2	PayPal-related	1103	16	0.340	335	0.754
3	Cashing Credit Cards	867	16	0.351	256	0.738
4	PGP	865	15	0.347	203	0.696
5	Netflix-related	846	14	0.270	351	0.805
6	Hacking Tools - General	825	15	0.331	132	0.516
7	Dumps - General	749	12	0.289	280	0.777
8	Linux-related	561	16	0.372	117	0.758
9	Email Hacking Tools	547	13	0.335	196	0.738
10	Network Security Tools	539	15	0.366	117	0.621
11	Ebay-related	472	15	0.385	163	0.772
12	Amazon-related	456	16	0.391	197	0.825
13	Bitcoin	443	15	0.360	201	0.823
14	Links (Lists)	422	12	0.211	221	0.838
15	Banking	384	13	0.349	186	0.840
16	Point of Sale	375	15	0.384	181	0.841
17	VPN	272	12	0.413	130	0.827
18	Botnet	257	12	0.291	110	0.796
19	Hacking Groups Invitation	251	14	0.387	143	0.865
20	RATs	249	15	0.453	99	0.797
21	Browser-related	249	12	0.380	134	0.857
22	Physical Layer Hacking	237	13	0.408	122	0.856
23	Password Cracking	230	13	0.434	100	0.781
24	Smartphone - General	223	14	0.408	110	0.816
25	Wireless Hacking	222	13	0.389	56	0.601
26	Phishing	218	13	0.403	111	0.849
27	Exploit Kits	218	14	0.413	91	0.795
28	Viruses/Counter AntiVirus	210	14	0.413	60	0.684
29	Network Layer Hacking	205	14	0.459	60	0.716
30	RDP Servers	191	12	0.405	124	0.895
31	Android-related	156	11	0.429	60	0.770
32	Keyloggers	143	13	0.496	77	0.862
33	Windows-related	119	12	0.464	50	0.717
34	Facebook-related	119	15	0.501	67	0.876

Use Case: Vulnerability Prioritization

14,185 vulnerabilities disclosed in 2015.
(RiskBased Security, 2015)

How to prioritize?
Current practices do not consider threat capabilities.

99.9% of breaches in 2015 due to known vulnerabilities.
(Verizon, 2015)



Vulnerability CVE-2015-0057 for remote code execution

No known exploit – how do we prioritize?

Feb. 2015

IntelliSpyre finds exploit on the darkweb: 48 BTC (~$10K)

No public or commercial knowledge of the exploit

April 2015

FireEye finds exploit in banking malware

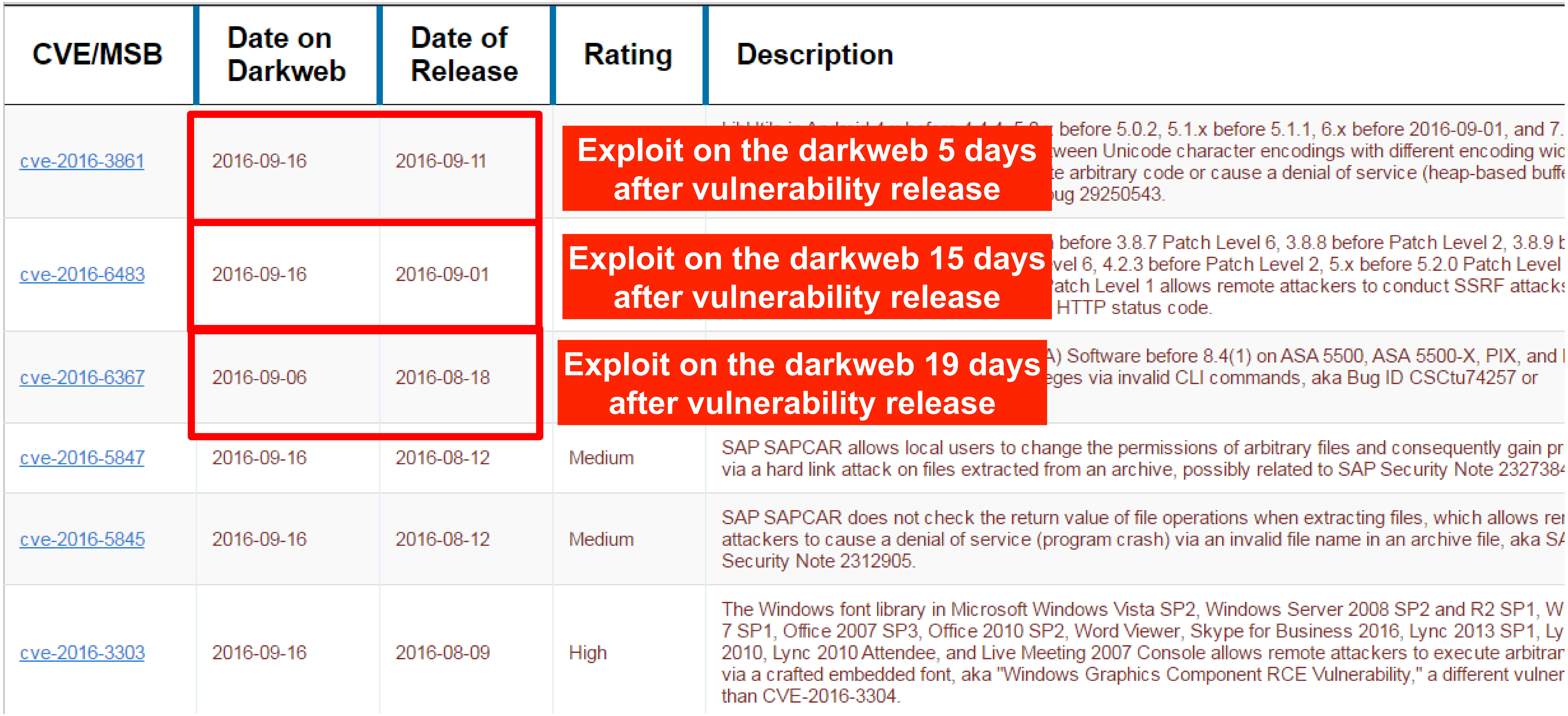
First time known in public

July 2015

60 day

Anticipate and avoid

Use Case: Vulnerability Prioritization



CVE/MSB	Date on Darkweb	Date of Release	Rating	Description
cve-2016-3861	2016-09-16	2016-09-11		before 5.0.2, 5.1.x before 5.1.1, 6.x before 2016-09-01, and 7. ween Unicode character encodings with different encoding wic te arbitrary code or cause a denial of service (heap-based buffe ug 29250543.
cve-2016-6483	2016-09-16	2016-09-01		before 3.8.7 Patch Level 6, 3.8.8 before Patch Level 2, 3.8.9 b vel 6, 4.2.3 before Patch Level 2, 5.x before 5.2.0 Patch Level atch Level 1 allows remote attackers to conduct SSRF attacks HTTP status code.
cve-2016-6367	2016-09-06	2016-08-18		A) Software before 8.4(1) on ASA 5500, ASA 5500-X, PIX, and eges via invalid CLI commands, aka Bug ID CSCtu74257 or
cve-2016-5847	2016-09-16	2016-08-12	Medium	SAP SAPCAR allows local users to change the permissions of arbitrary files and consequently gain pr via a hard link attack on files extracted from an archive, possibly related to SAP Security Note 232738
cve-2016-5845	2016-09-16	2016-08-12	Medium	SAP SAPCAR does not check the return value of file operations when extracting files, which allows re attackers to cause a denial of service (program crash) via an invalid file name in an archive file, aka SA Security Note 2312905.
cve-2016-3303	2016-09-16	2016-08-09	High	The Windows font library in Microsoft Windows Vista SP2, Windows Server 2008 SP2 and R2 SP1, W 7 SP1, Office 2007 SP3, Office 2010 SP2, Word Viewer, Skype for Business 2016, Lync 2013 SP1, Ly 2010, Lync 2010 Attendee, and Live Meeting 2007 Console allows remote attackers to execute arbitrar via a crafted embedded font, aka "Windows Graphics Component RCE Vulnerability," a different vulner than CVE-2016-3304.

Exploit on the darkweb 5 days after vulnerability release

Exploit on the darkweb 15 days after vulnerability release

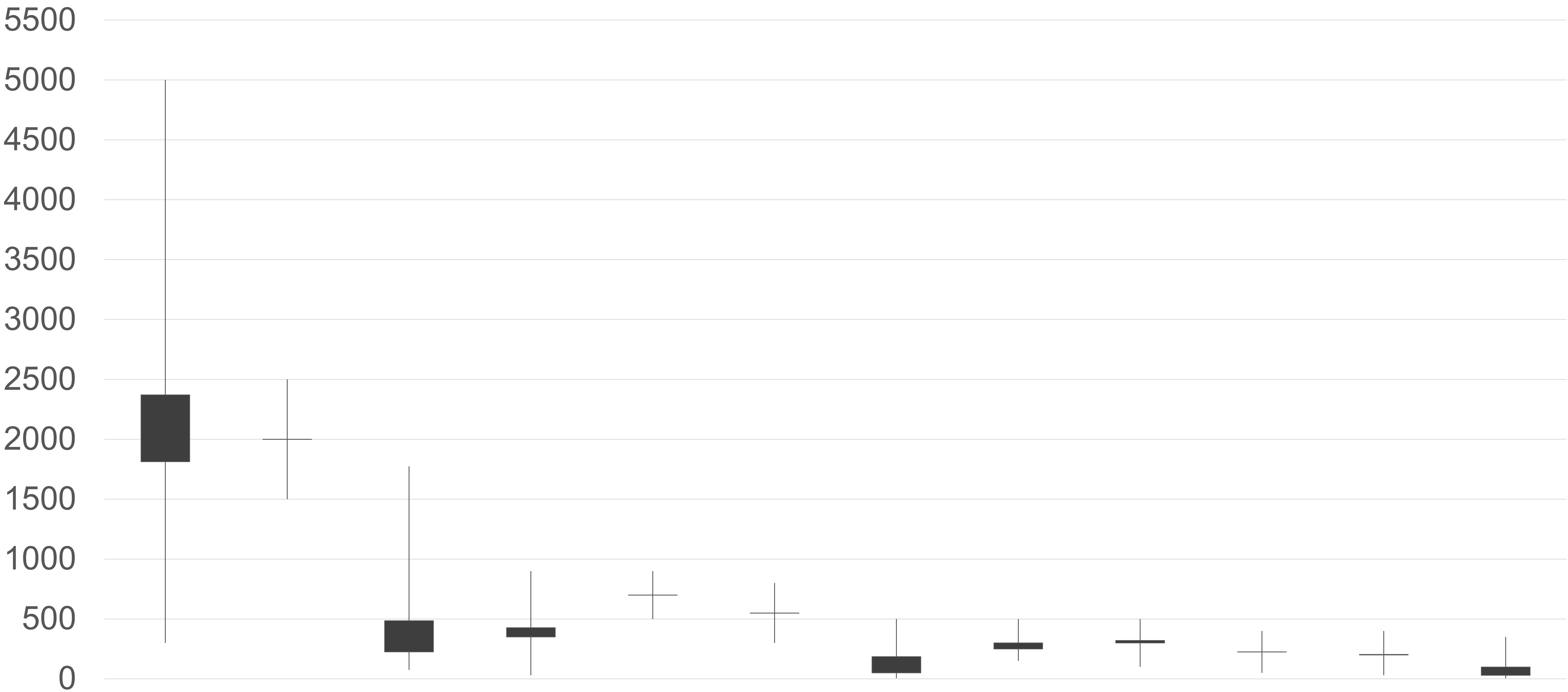
Exploit on the darkweb 19 days after vulnerability release

Use Case: Identifying Zero Day Exploits

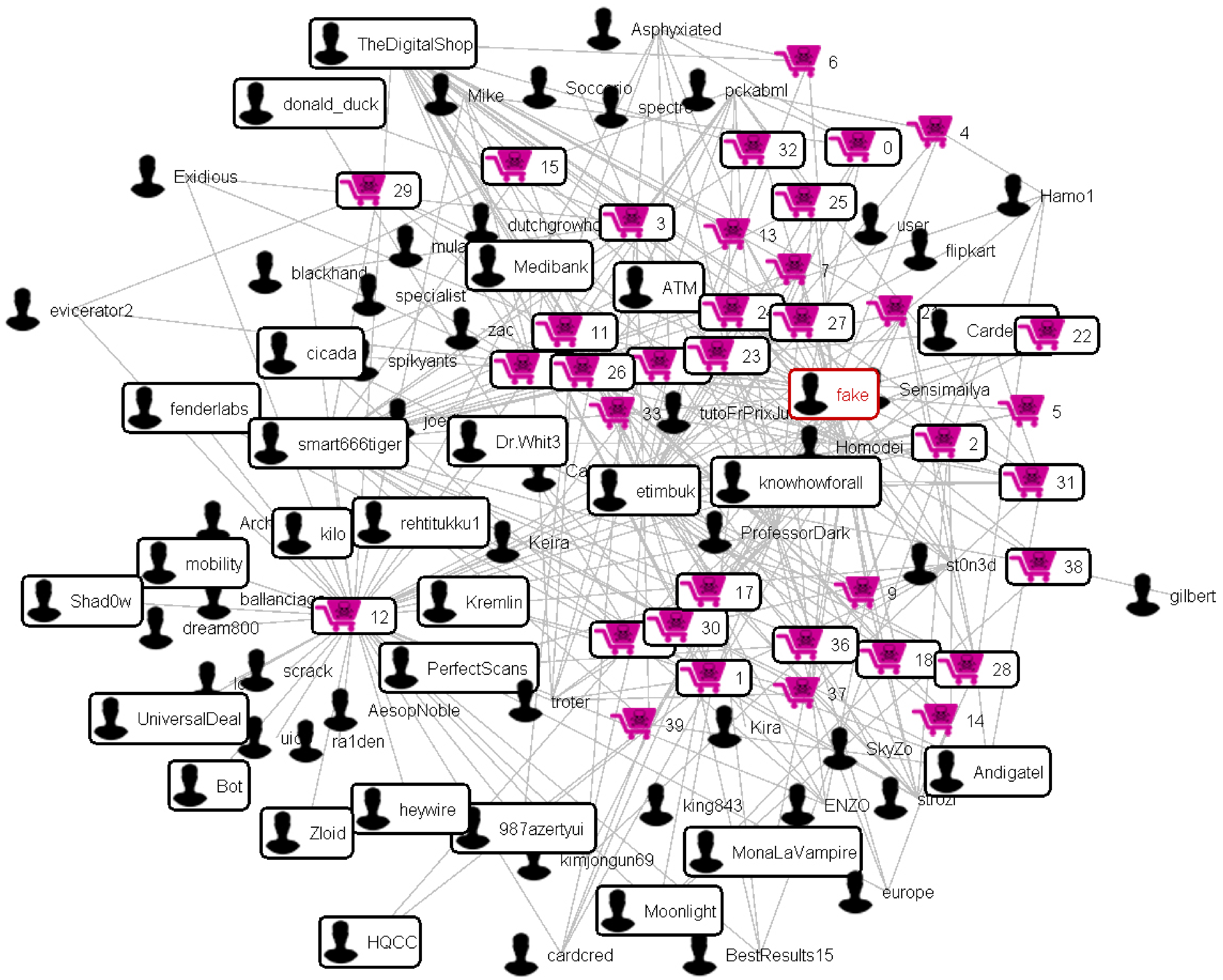
Title	Date	BTC Price
Windows 10 *HOT* (10.0.10586 Build 10586)	-	2.00000
Windows 10 UAC (10.0.10586)	-	5.0000
PowerPoint 03/07/10 exploit	Sep 14 2015	12.5954
Internet Explorer 11 Remote Code Execution 0day	August 23 2015	20.4676

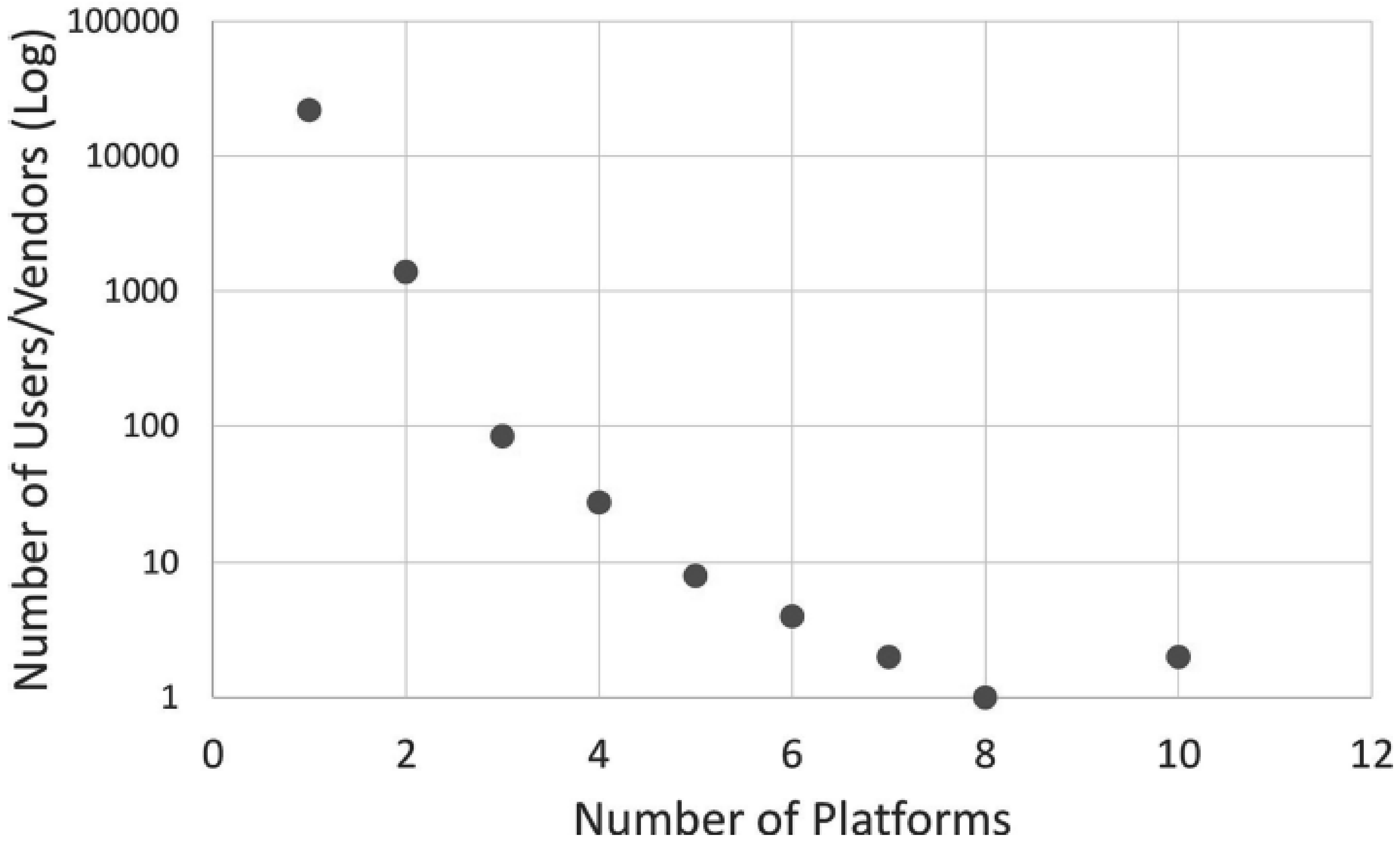
Use Case: Hacker Economics

Select Russian Hacker Product Pricing



Use Case: Social Network Analysis





We identify malware vendors who have a presence in multiple marketplaces

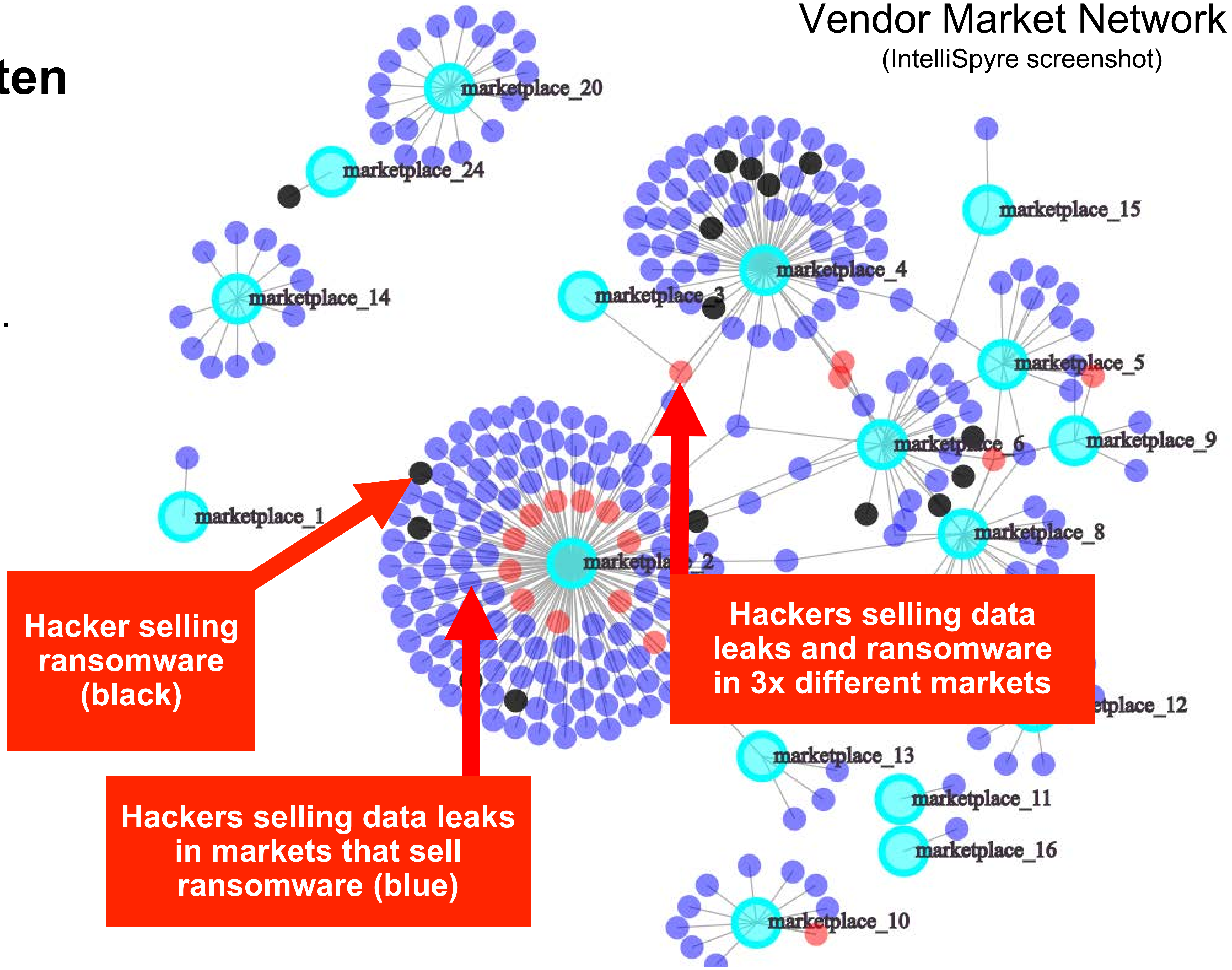
Use Case: Social Network Analysis

Ransomware victims are also often data-leakage victims.

Ransomware vendors and markets also sell the results of data-leakage information.

IntelliSpyre can identify where data leaks are sold from the vendors of ransomware through link analysis.

Quick location of dataleaks after ransomware incidents.



Vendor Market Network
(IntelliSpyre screenshot)

Hacker selling ransomware (black)

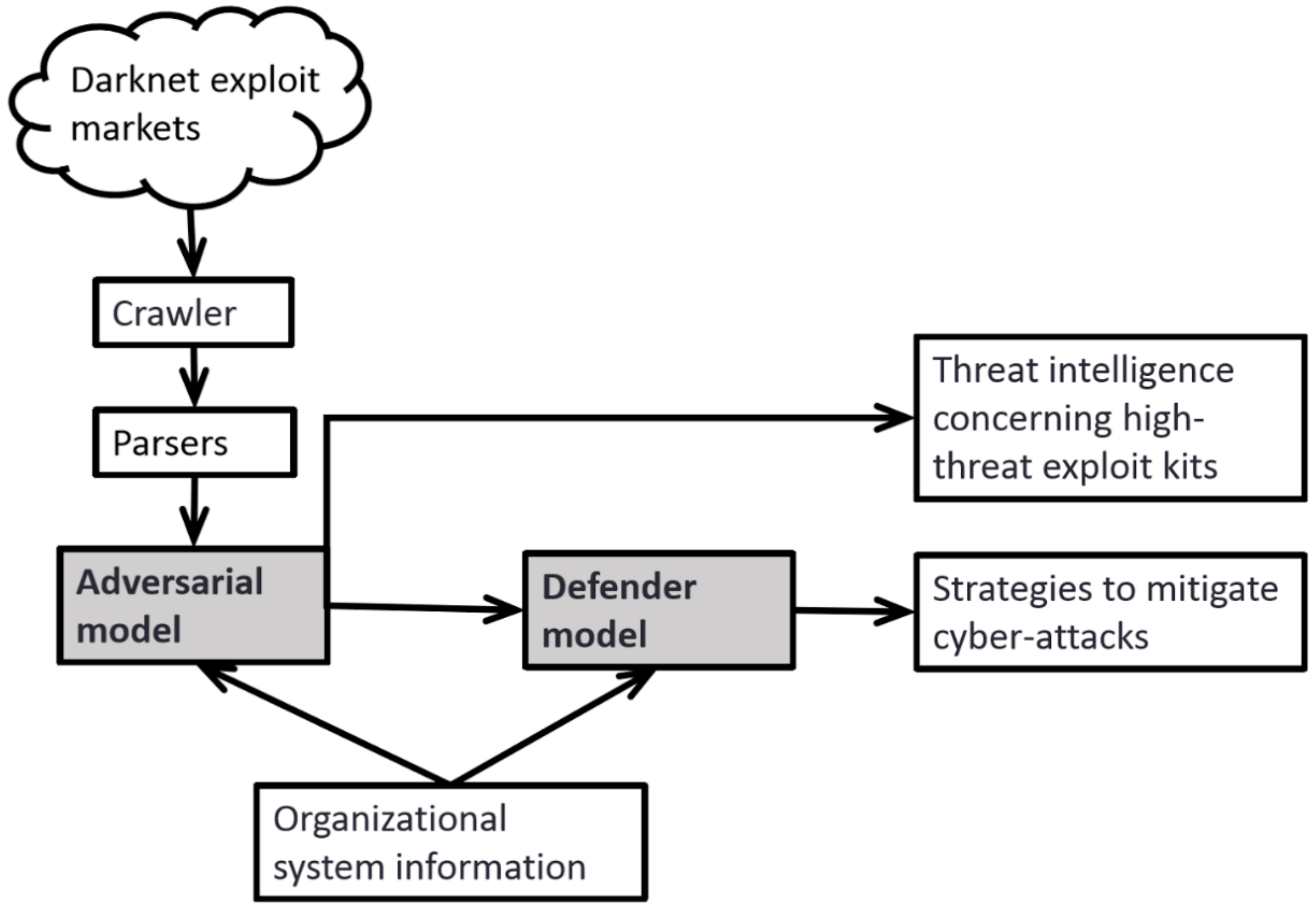
Hackers selling data leaks in markets that sell ransomware (blue)

Hackers selling data leaks and ransomware in 3x different markets

Use Case: Game Theoretic Risk Assessment

Threat intelligence data can feed mathematical models of risk.

We can assess most damaging exploits to a given system



Darknet exploit markets

Crawler

Parsers

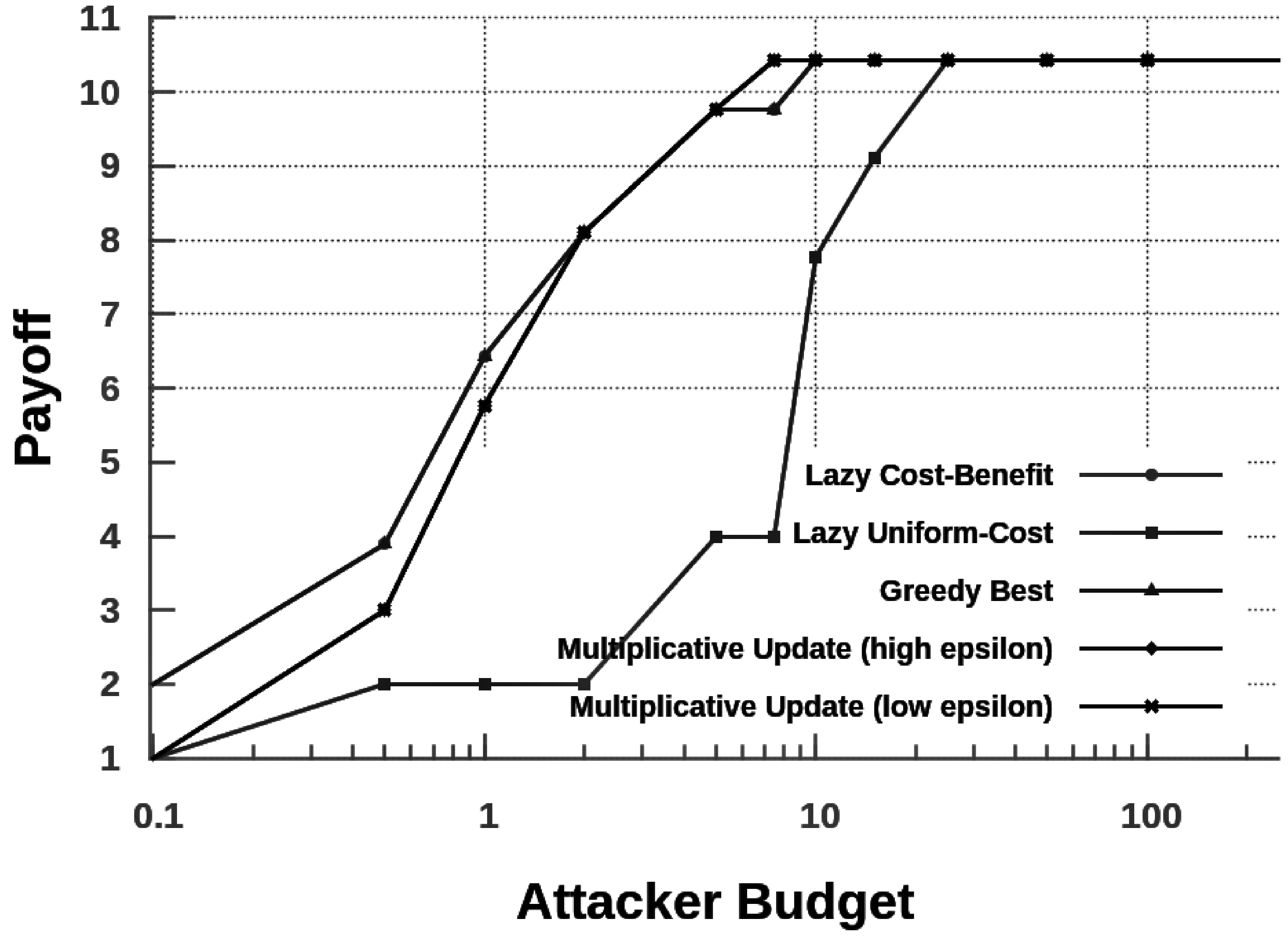
Adversarial model

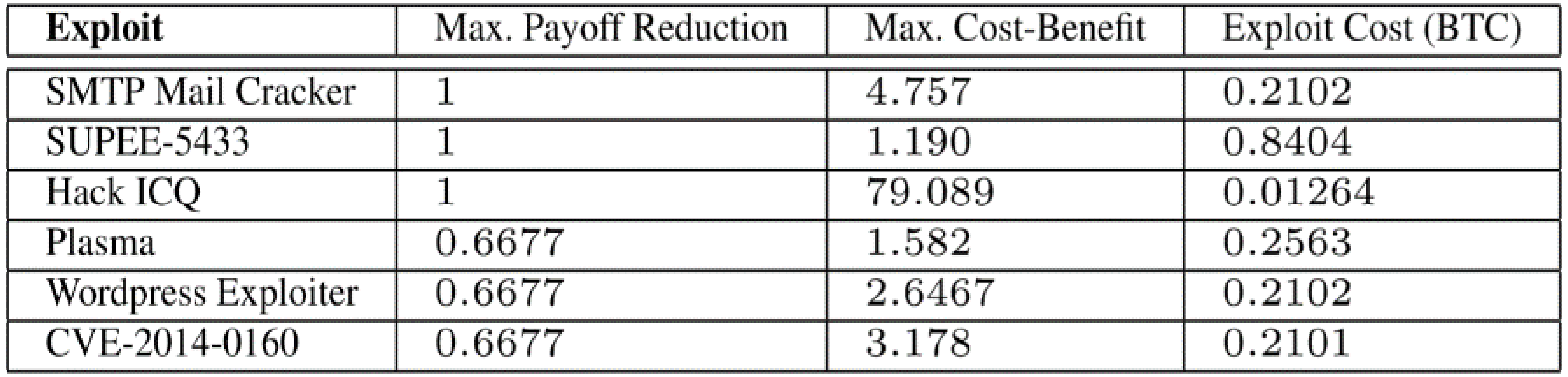
Defender model

Threat intelligence concerning high-threat exploit kits

Strategies to mitigate cyber-attacks

Organizational system information





Exploit	Max. Payoff Reduction	Max. Cost-Benefit	Exploit Cost (BTC)
SMTP Mail Cracker	1	4.757	0.2102
SUPEE-5433	1	1.190	0.8404
Hack ICQ	1	79.089	0.01264
Plasma	0.6677	1.582	0.2563
Wordpress Exploiter	0.6677	2.6467	0.2102
CVE-2014-0160	0.6677	3.178	0.2101

U.S. Provisional Patent 62/261,200

And now a short demonstration…

<u>No</u> cameras or video.

We Are a Platform and Have Active Developers

Developers

LOCKHEED MARTIN

CONICET

BBN TECHNOLOGIES

CENTER FOR CYBERSECURITY & DIGITAL FORENSICS

HYPERION GRAY

USC Viterbi
Information Sciences Institute

Research Sponsors

IARPA
BE THE FUTURE

ASU Global Security Initiative
ARIZONA STATE UNIVERSITY

Office of Naval Research
USN ONR
Science & Technology

SCIENCE TO SHAPE THE FUTURE OF THE ARMY
ARMY RESEARCH OFFICE
1951

Cisco

INNOVATION GRAND CHALLENGE

Win a share of $250,000 to jumpstart your venture.

Submissions have closed for judging.
2016 semifinalist have been announced!

5718 SUBMISSIONS

15 SEMIFINALISTS

6 FINALISTS

3 WINNERS

IntelliSpyre made the semi-finals

15 semi-finalists
7x from U.S.
2x cybersecurity
1x from Arizona

IntelliSpyre

info@intellispyre.com
intellispyre.com

Thank You!

@PauloShakASU @intellispyre

CISCO
INNOVATION
GRAND
CHALLENGE
SEMI-FINALIST
2016