Annual Conference on Large Scale Network Defense Analytics (FloCon)

DeCypher: Cyber Knowledge Graph Queries Expressed through Natural Language

Steven Noel, PhD

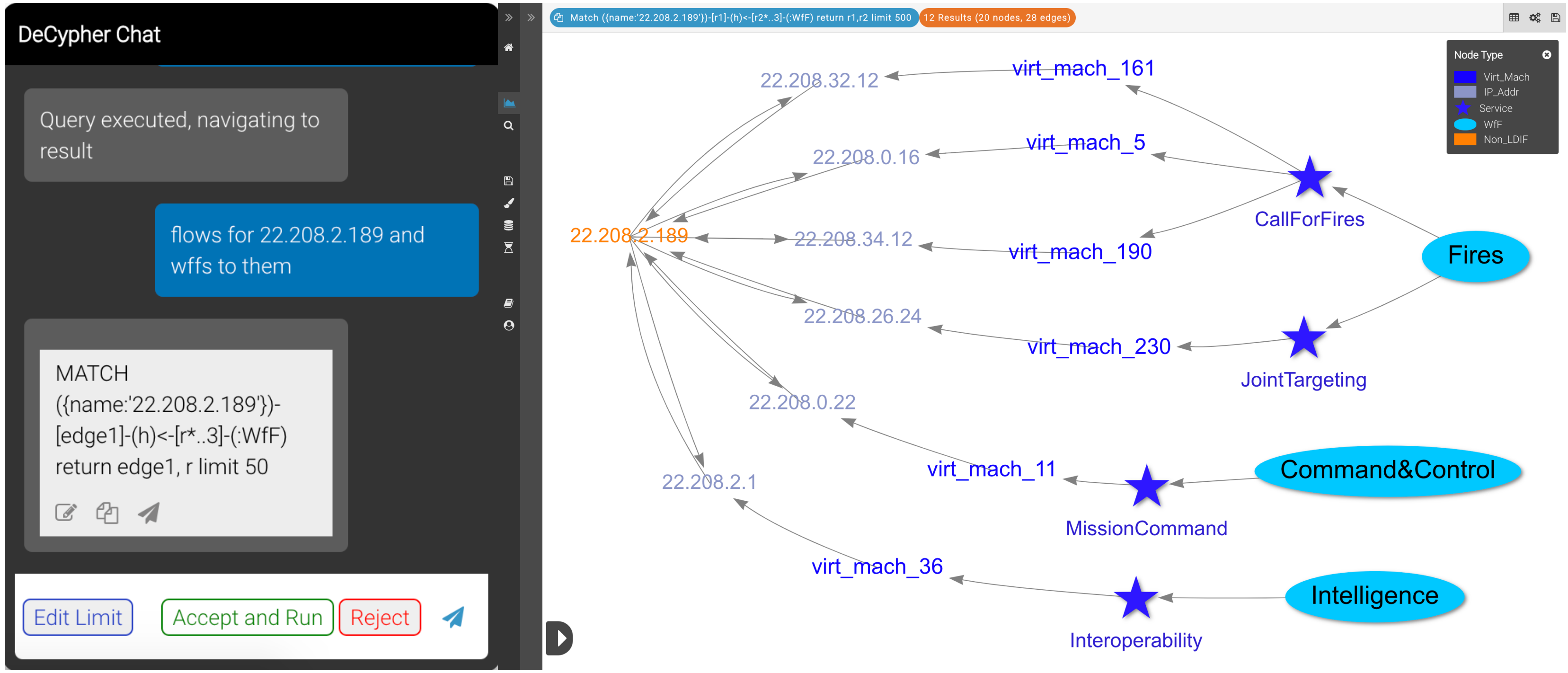
9 January 2023

SOLVING PROBLEMS FOR A SAFER WORLD

DeCypher: Translating Chat Requests to Cyber Queries

Problem*:* MITRE’s CyGraph tool provides advanced Cyber Situational Understanding (Cyber SU), but writing queries in formal language is time consuming and requires specialized skills

Approach*:* Deep learning NLP translates plain English questions to CyGraph queries



DeCypher Chat

Query executed, navigating to result

flows for 22.208.2.189 and wffs to them

```
MATCH
({name:'22.208.2.189'})-
[edge1]-(h)<-[r*..3]-(:WfF)
return edge1, r limit 50
```

Edit Limit | Accept and Run | Reject

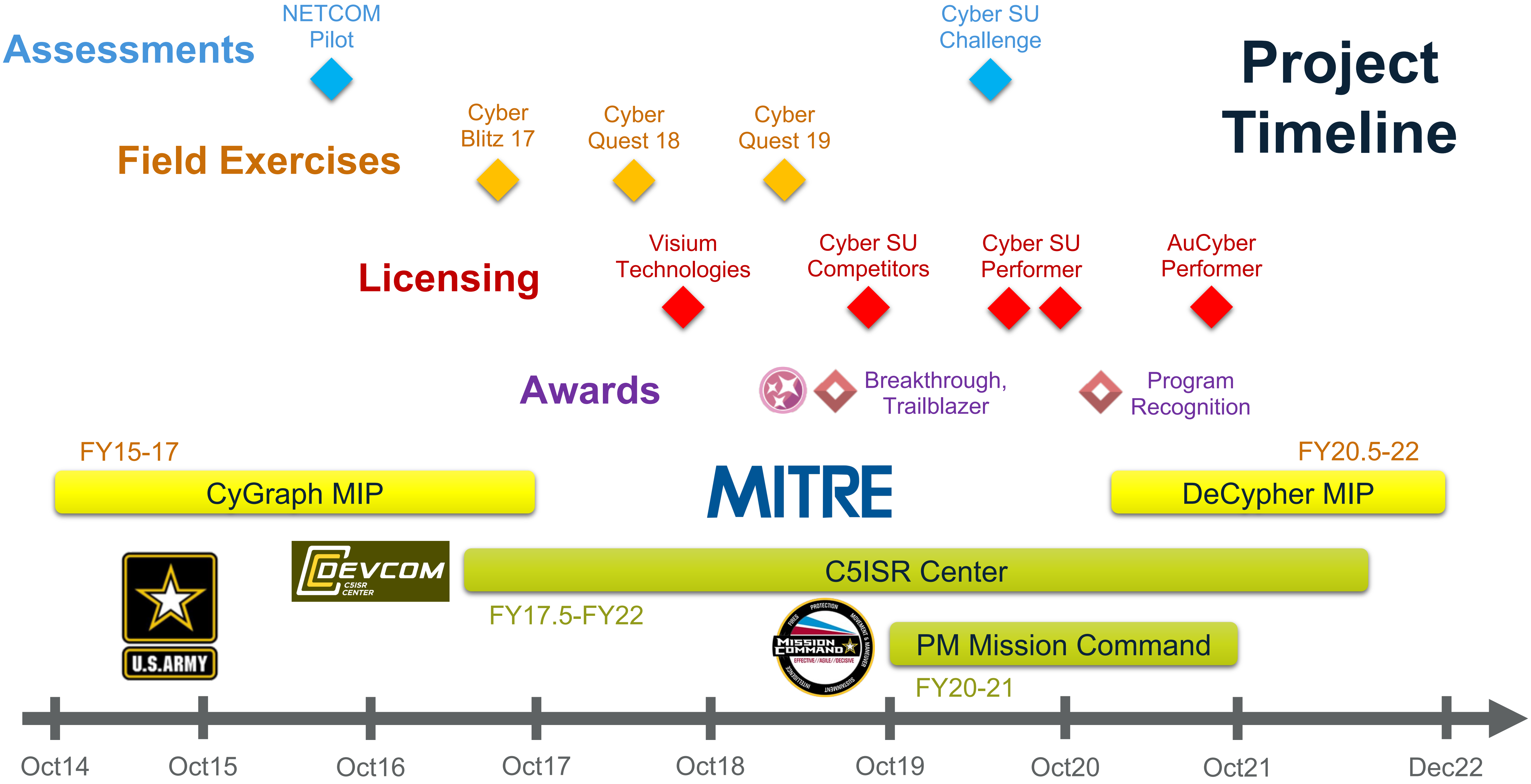
```
Match ({name:'22.208.2.189'})-[r1]-(h)<-[r2*..3]-(:WfF) return r1,r2 limit 500
```

12 Results (20 nodes, 28 edges)

Achievements

- Reference implementation for U.S. Army Cyber SU tool
- Algorithm performance (F1 score): intent classification 84%, entity recognition 79%
- 2x improvement in timed-limited task completion, tasks completed 21% faster
- Improves user satisfaction by 62%, increases perceived usability by 49%
- Journal article, conference publication, patent pending

Project Timeline



Assessments

NETCOM Pilot

Cyber SU Challenge

Field Exercises

Cyber Blitz 17

Cyber Quest 18

Cyber Quest 19

Licensing

Visium Technologies

Cyber SU Competitors

Cyber SU Performer

AuCyber Performer

Awards

Breakthrough, Trailblazer

Program Recognition

FY15-17

CyGraph MIP

MITRE

FY20.5-22

DeCypher MIP

C5ISR Center

FY17.5-FY22

PM Mission Command

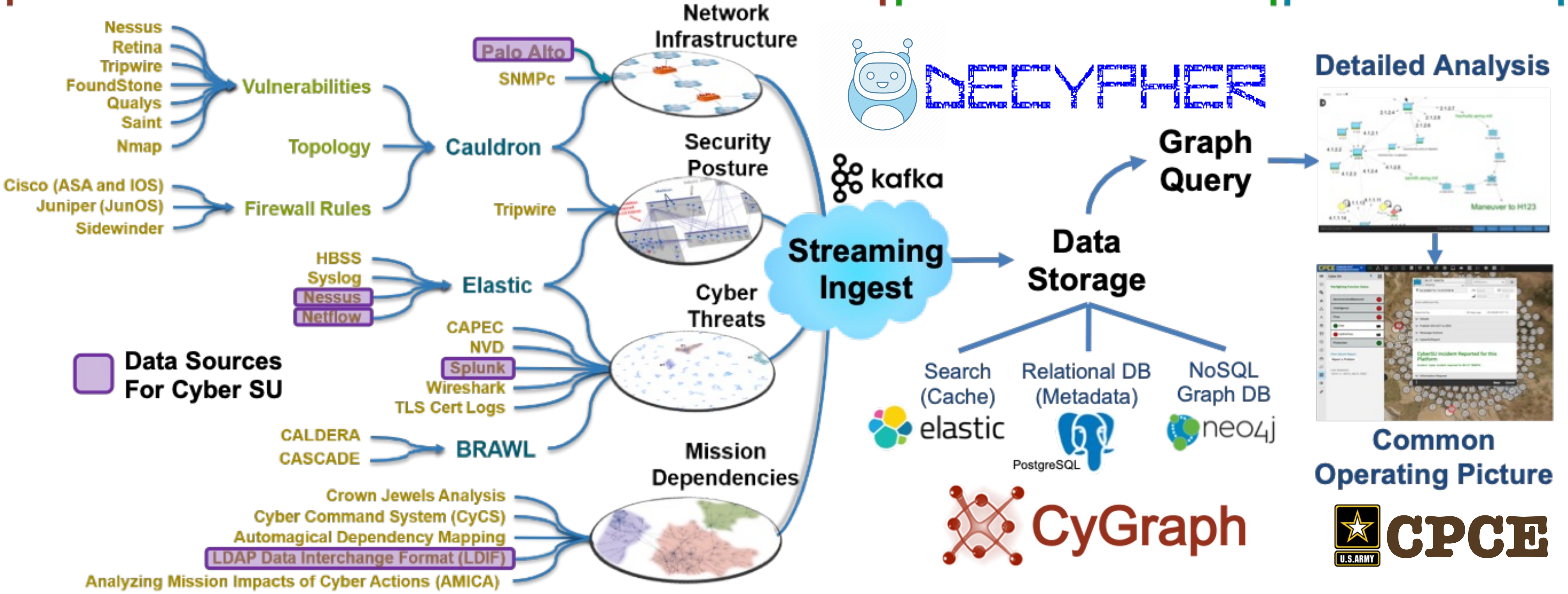
FY20-21

System-of-Systems Architecture

Multiple Stovepipe Data Sources Mapped To Common Graph Model in Real Time

High-Precision Graph Queries Answer Specific Analytic Questions

Interactive Visualization For Commander And Staff



Nessus
Retina
Tripwire
FoundStone
Qualys
Saint
Nmap

Vulnerabilities

Topology

Cisco (ASA and IOS)
Juniper (JunOS)
Sidewinder

Firewall Rules

Cauldron

Palo Alto
SNMPc

Network Infrastructure

Tripwire

Security Posture

HBSS
Syslog
Nessus
Netflow

Elastic

CAPEC
NVD
Splunk
Wireshark
TLS Cert Logs

Cyber Threats

CALDERA
CASCADE

BRAWL

Crown Jewels Analysis
Cyber Command System (CyCS)
Automagical Dependency Mapping
LDAP Data Interchange Format (LDIF)
Analyzing Mission Impacts of Cyber Actions (AMICA)

Mission Dependencies

▢ Data Sources For Cyber SU

kafka

Streaming Ingest

DECYPHER

Graph Query

Data Storage

Search (Cache)
elastic

Relational DB (Metadata)
PostgreSQL

NoSQL Graph DB
neo4j

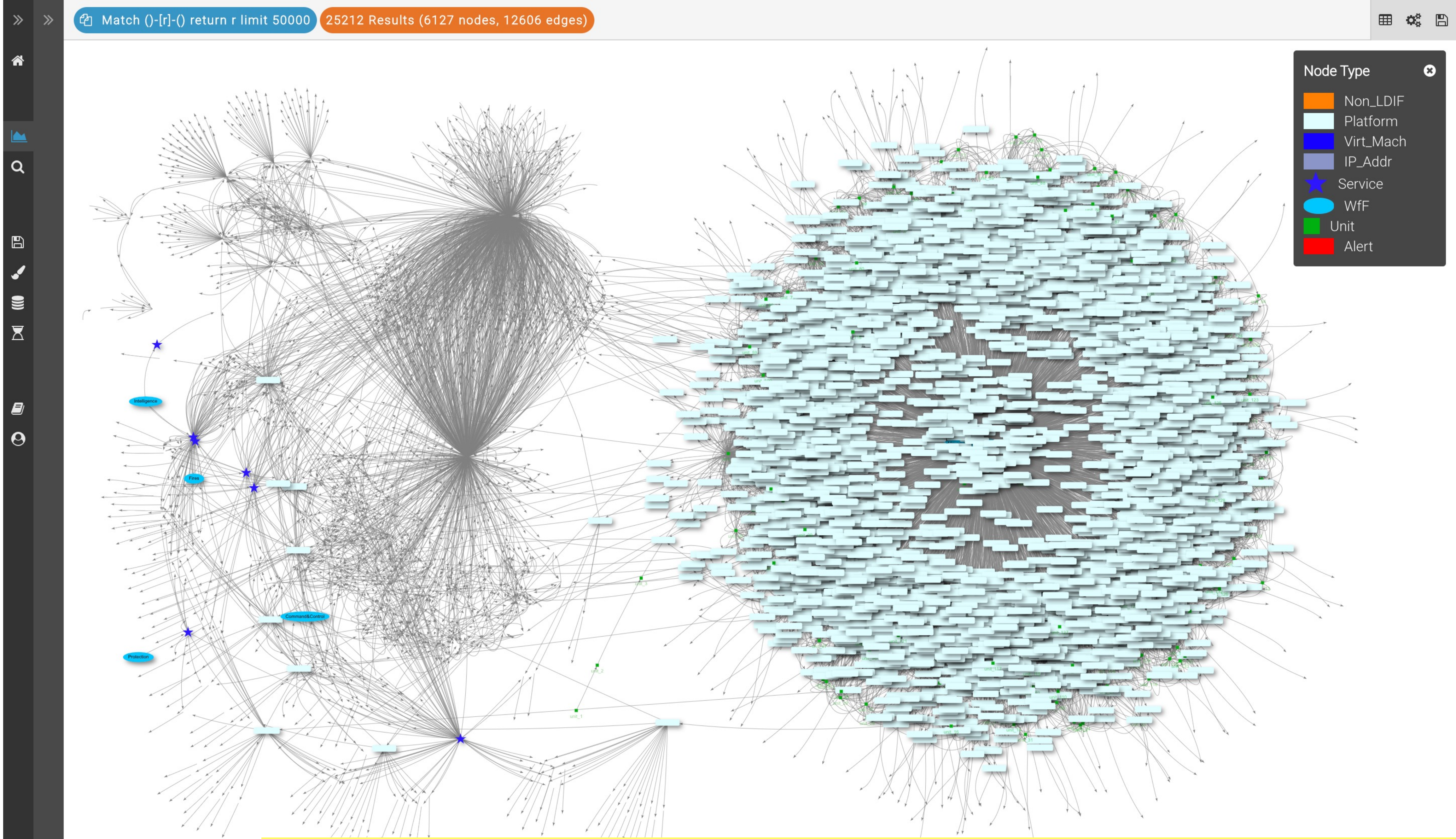
CyGraph

Detailed Analysis

Common Operating Picture

U.S.ARMY CPCE

MITRE



Match ()-[r]-() return r limit 50000

25212 Results (6127 nodes, 12606 edges)

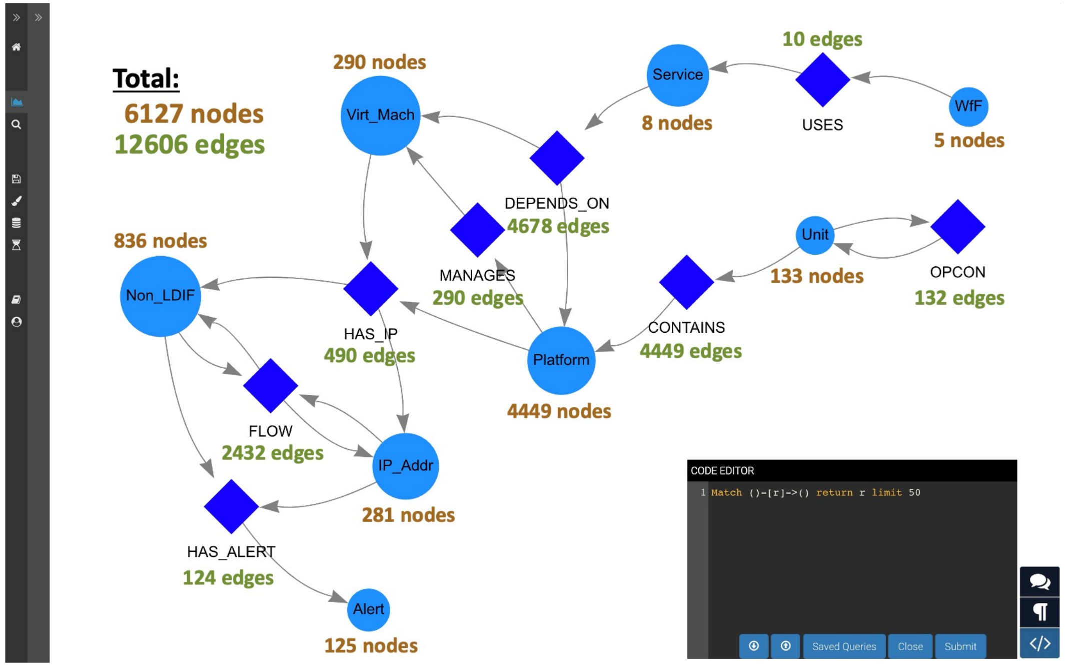
Node Type

- 🟧 Non_LDIF
- ⬜ Platform
- 🟦 Virt_Mach
- ▨ IP_Addr
- ★ Service
- ● WfF
- 🟩 Unit
- 🟥 Alert

How to extract actionable intelligence from swarm of interrelated data?

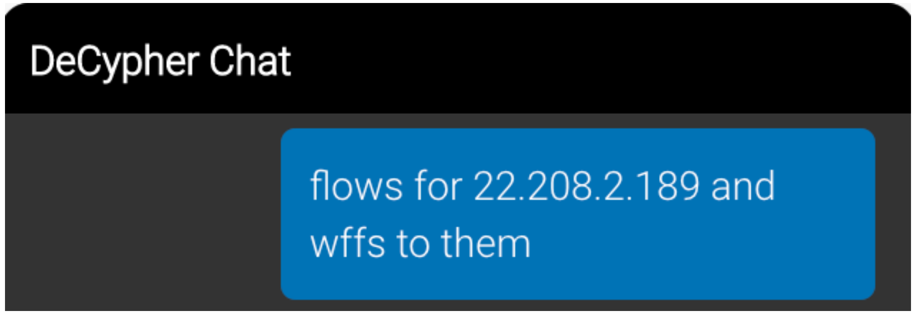
Extracting Actionable Intelligence

Cyber Knowledge Base (CyGraph)



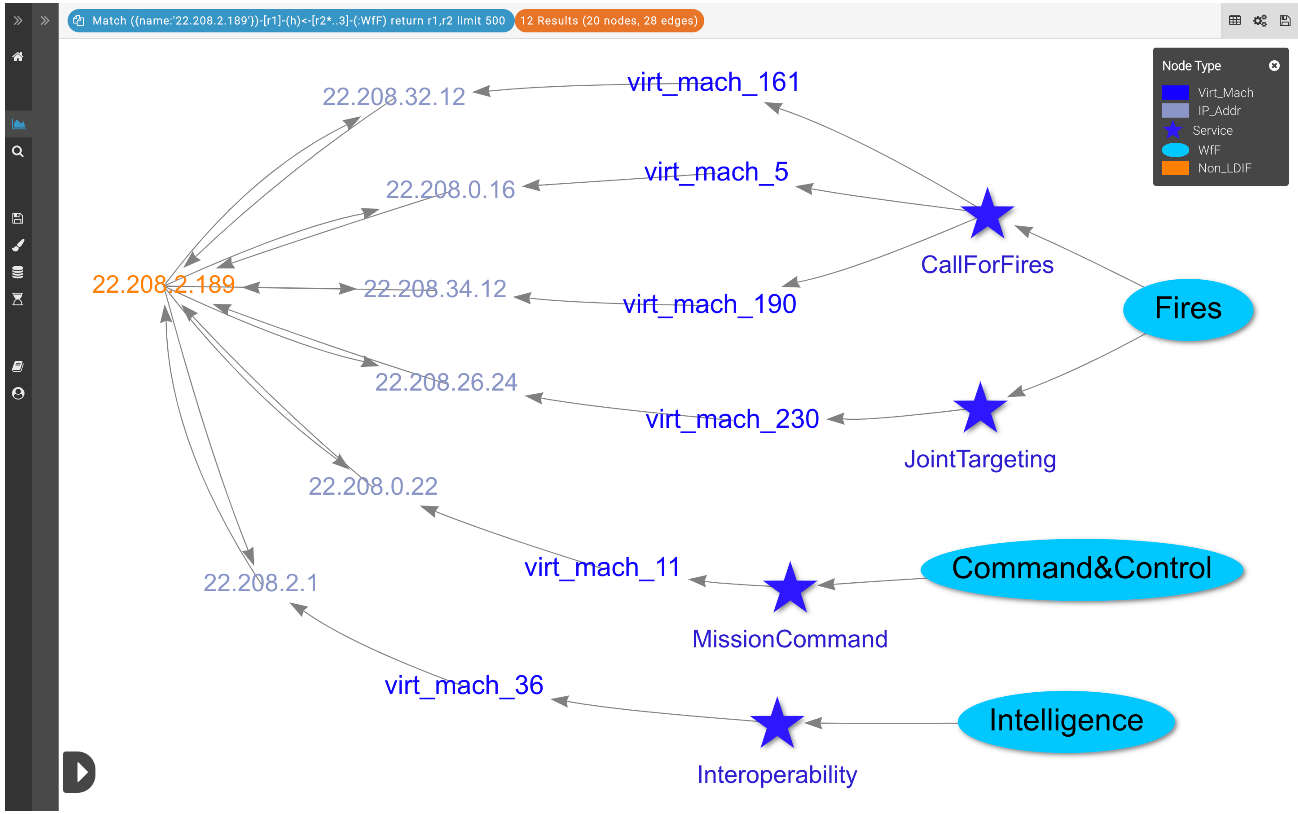
```
Match ()-[r]->() return r limit 50
```

Intelligent UI (DeCypher)



flows for 22.208.2.189 and wffs to them

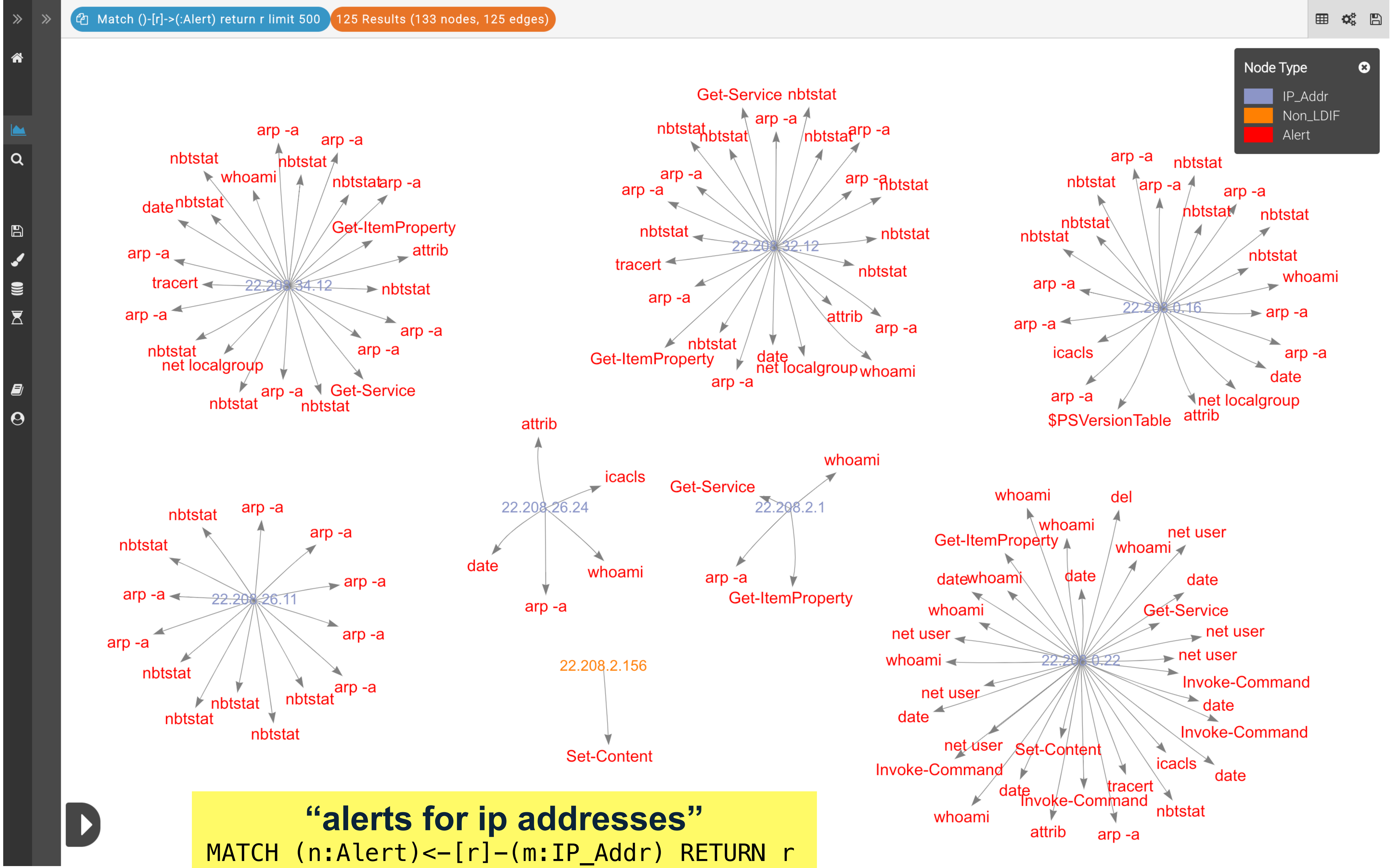
Actionable Intelligence



```
Match ({name:'22.208.2.189'})-[r1]-(h)<-[r2*..3]-(:WfF) return r1,r2 limit 500
```

12 Results (20 nodes, 28 edges)

Example: Scope of cyber threat on tactical warfighting functions

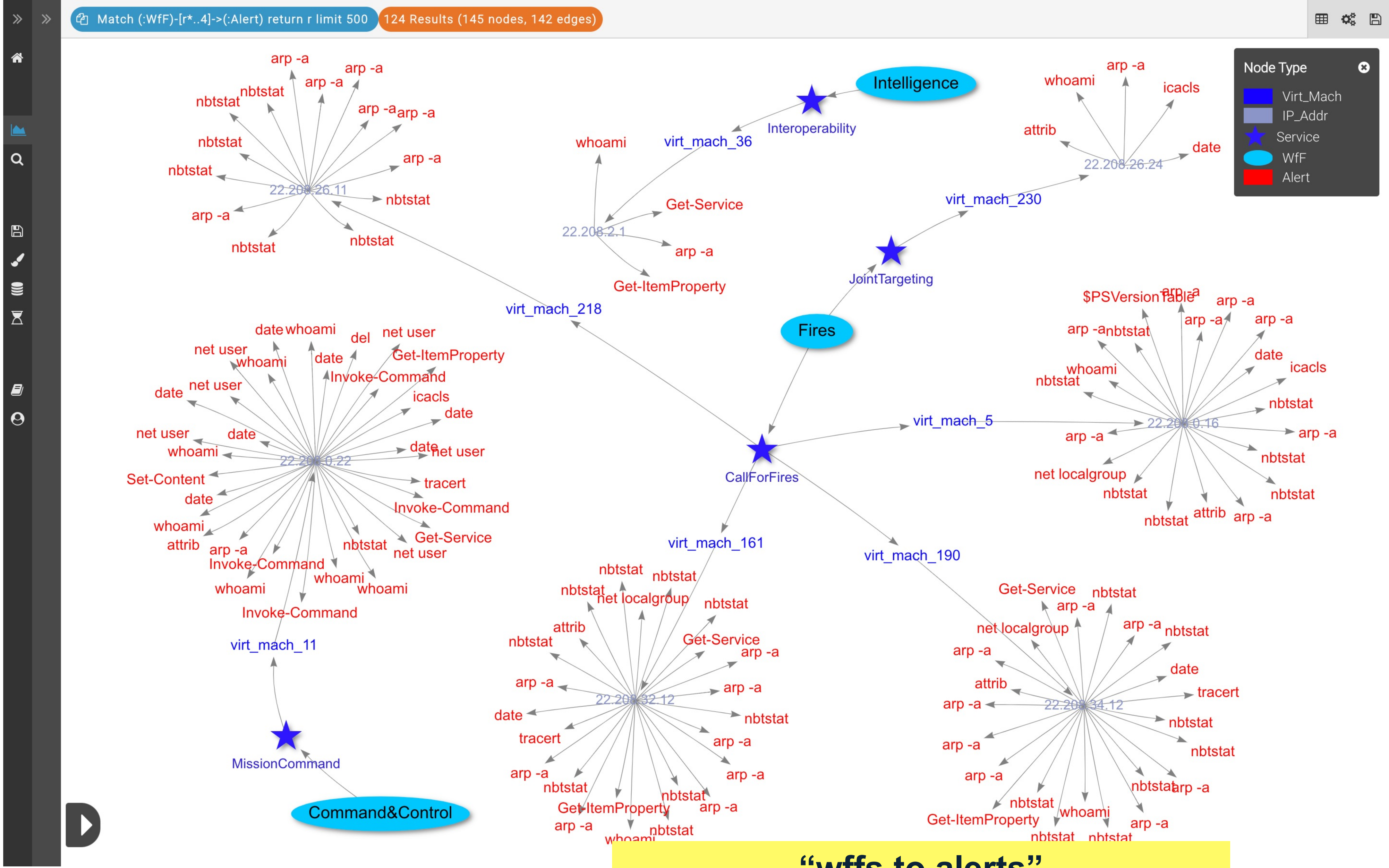


Match ()-[r]->(:Alert) return r limit 500

125 Results (133 nodes, 125 edges)

“alerts for ip addresses”

```
MATCH (n:Alert)<-[r]-(m:IP_Addr) RETURN r
```

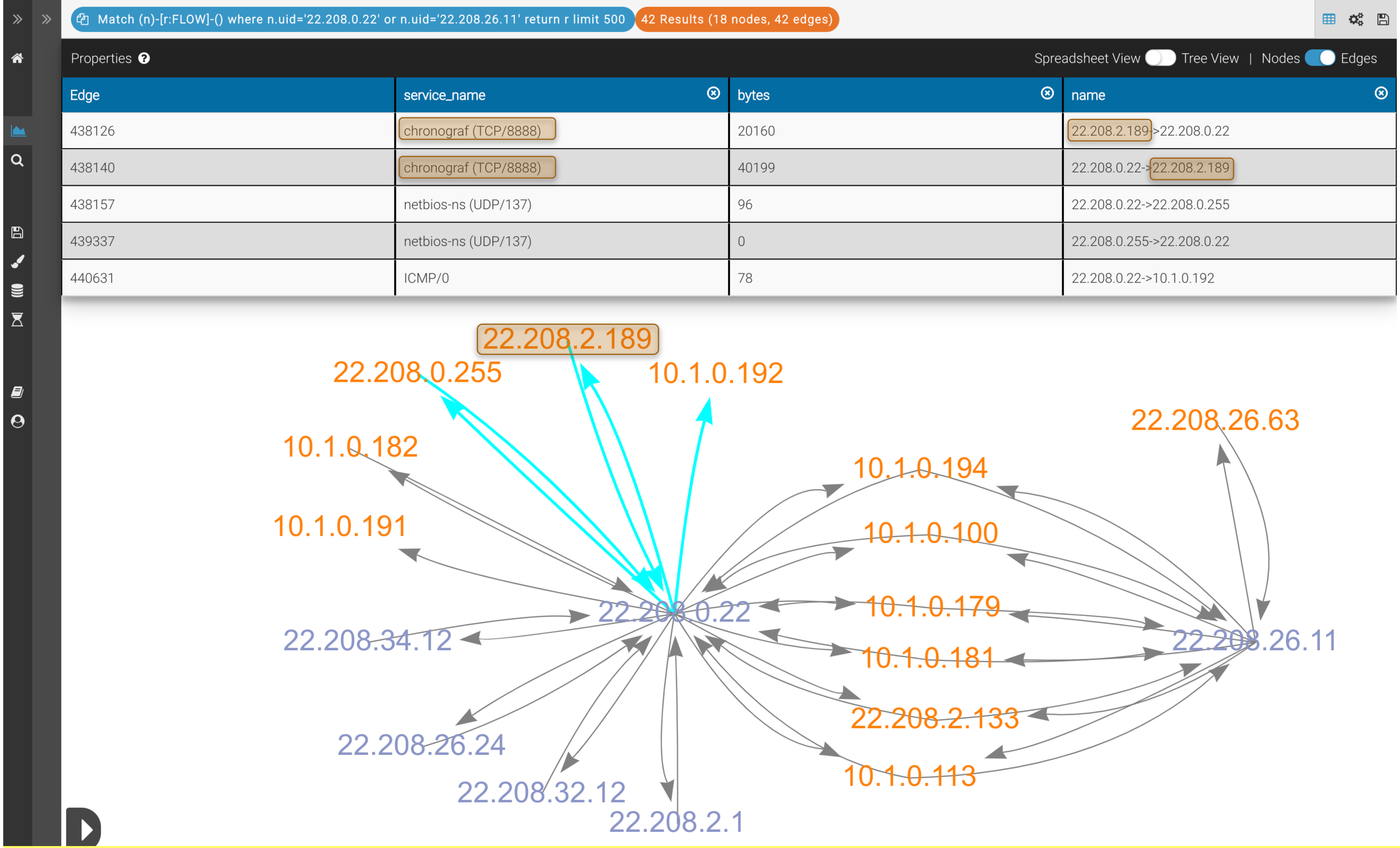


Match (:WfF)-[r*..4]->(:Alert) return r limit 500

124 Results (145 nodes, 142 edges)

“wffs to alerts”

```
MATCH (n:WfF)-[r*..4]->(m:Alert) RETURN r
```



Match (n)-[r:FLOW]-() where n.uid='22.208.0.22' or n.uid='22.208.26.11' return r limit 500

42 Results (18 nodes, 42 edges)

Properties

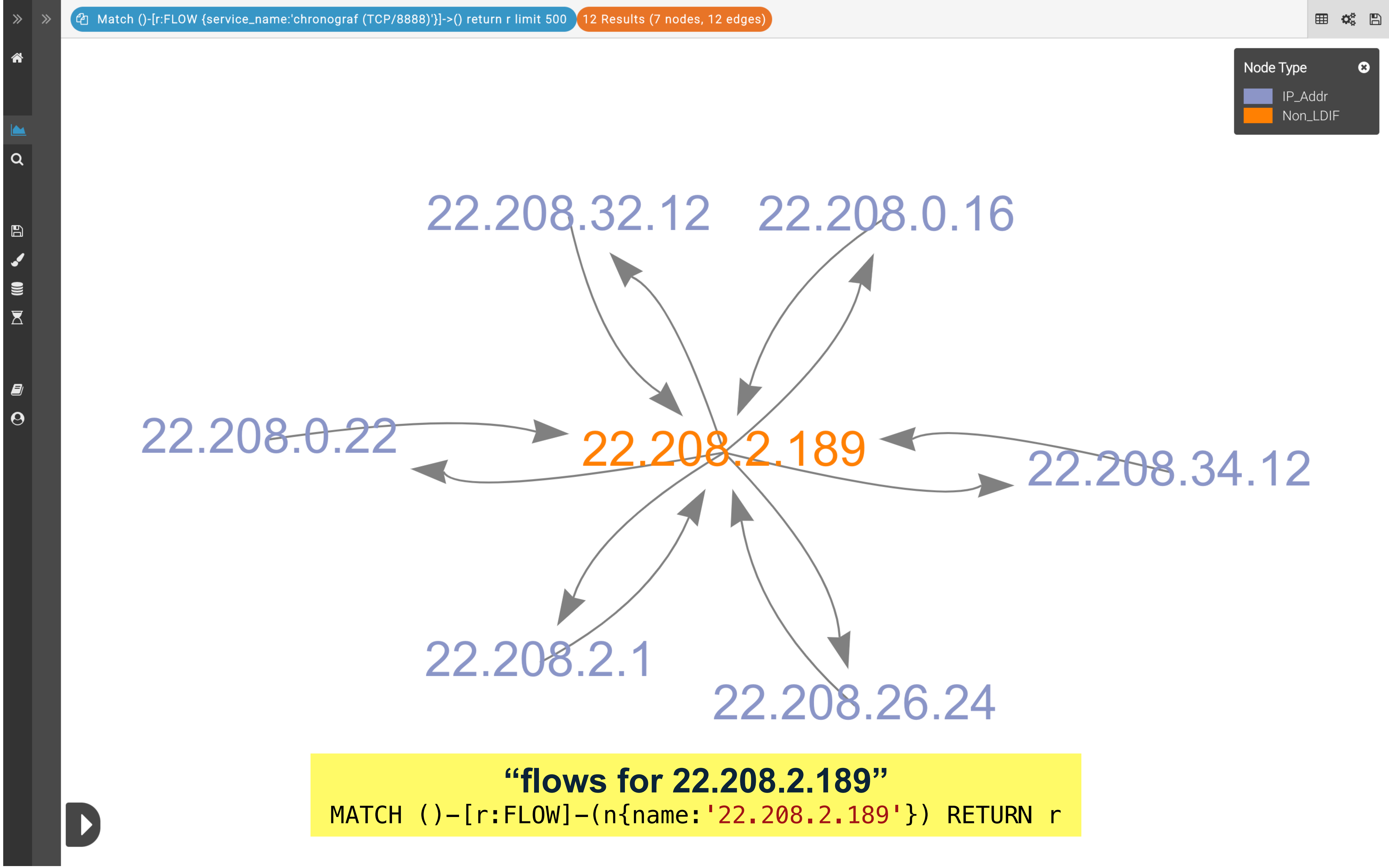
Spreadsheet View Tree View | Nodes Edges

Edge	service_name	bytes	name
438126	chronograf (TCP/8888)	20160	22.208.2.189->22.208.0.22
438140	chronograf (TCP/8888)	40199	22.208.0.22->22.208.2.189
438157	netbios-ns (UDP/137)	96	22.208.0.22->22.208.0.255
439337	netbios-ns (UDP/137)	0	22.208.0.255->22.208.0.22
440631	ICMP/0	78	22.208.0.22->10.1.0.192

MITRE

“network flows for 22.208.0.22 or 22.208.26.11”

```
MATCH (n)-[r:FLOW]-() WHERE n.name='22.208.0.22' OR n.name='22.208.26.11' RETURN r
```



```
Match ()-[r:FLOW {service_name:'chronograf (TCP/8888)'}]->() return r limit 500
```

12 Results (7 nodes, 12 edges)

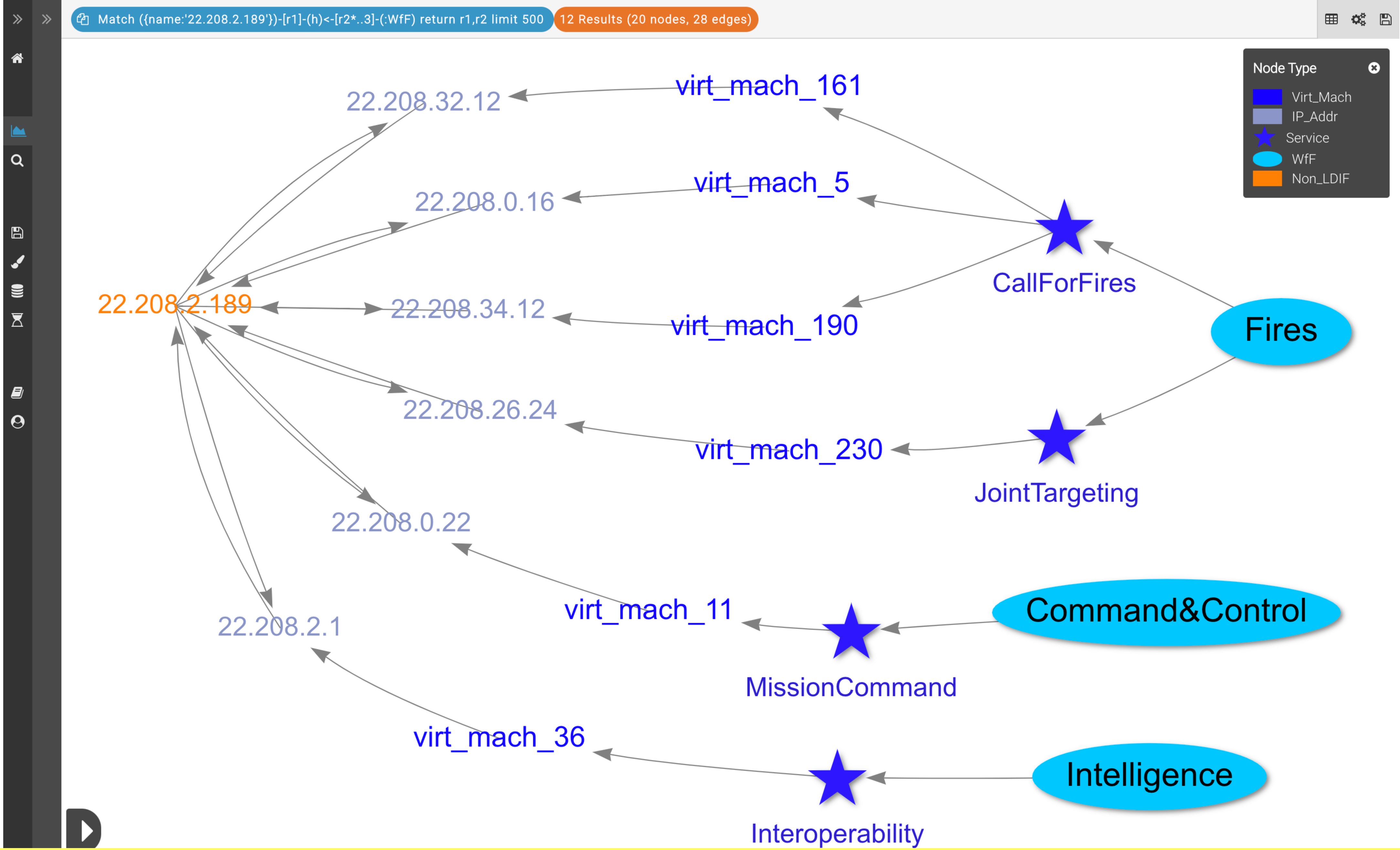
Node Type

- 🟦 IP_Addr
- 🟧 Non_LDIF

“flows for 22.208.2.189”

```
MATCH ()–[r:FLOW]–(n{name:'22.208.2.189'}) RETURN r
```

MITRE



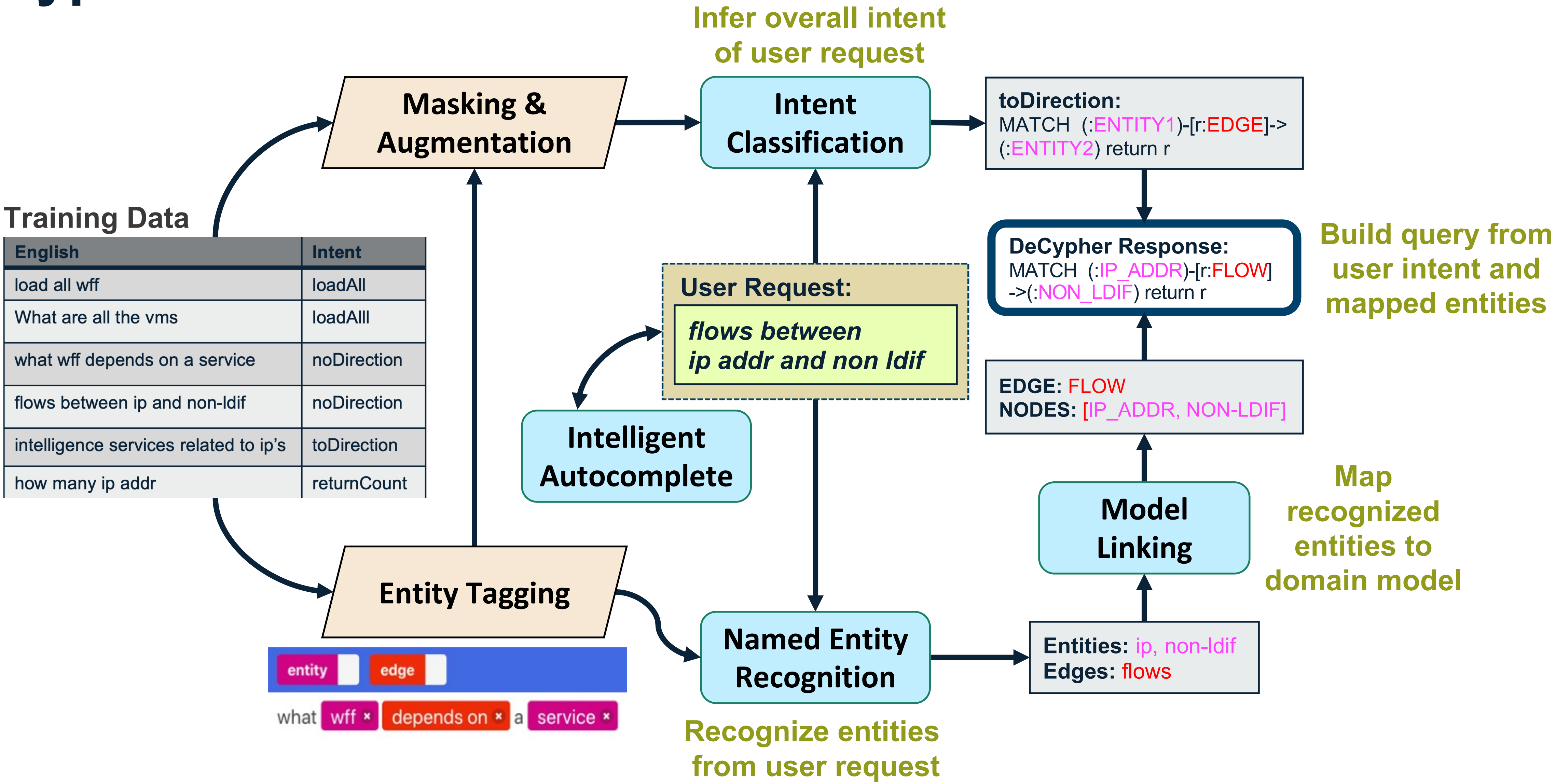
```
Match ({name:'22.208.2.189'})-[r1]-(h)<-[r2*..3]-(:WfF) return r1,r2 limit 500
```

12 Results (20 nodes, 28 edges)

“ips having flows with 22.208.2.189 and wffs to them”

```
MATCH (n{name:'22.208.2.189'})-[edge1]-(h)<-[r*..3]-(m:WfF) RETURN edge1, r
```

DeCypher Architecture



Training Data

English	Intent
load all wff	loadAll
What are all the vms	loadAlll
what wff depends on a service	noDirection
flows between ip and non-ldif	noDirection
intelligence services related to ip's	toDirection
how many ip addr	returnCount

Masking & Augmentation

Entity Tagging

Infer overall intent of user request

Intent Classification

User Request:

flows between ip addr and non ldif

Intelligent Autocomplete

Named Entity Recognition

Recognize entities from user request

toDirection:
MATCH (:ENTITY1)-[r:EDGE]->(:ENTITY2) return r

DeCypher Response:
MATCH (:IP_ADDR)-[r:FLOW]->(:NON_LDIF) return r

Build query from user intent and mapped entities

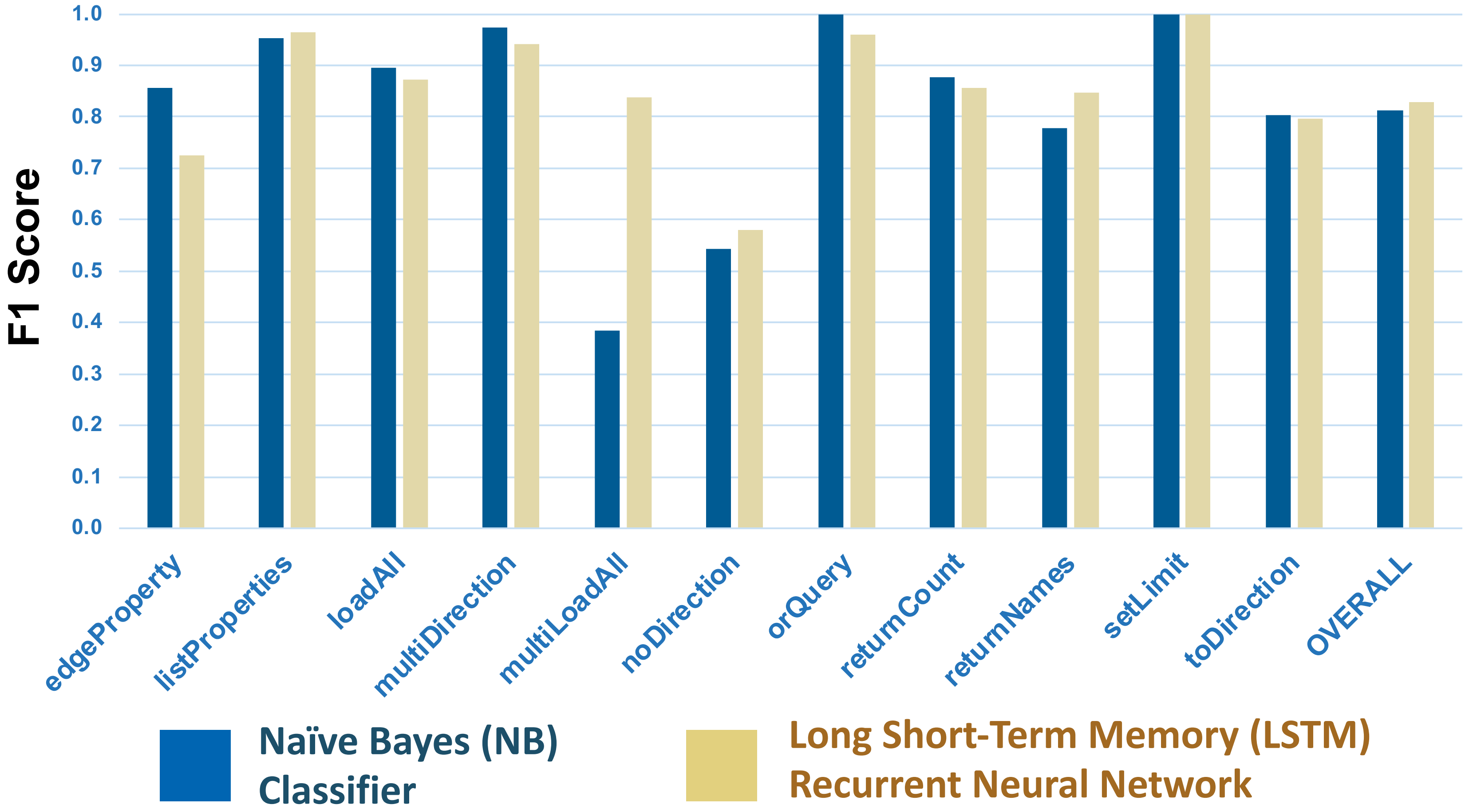
EDGE: FLOW
NODES: [IP_ADDR, NON-LDIF]

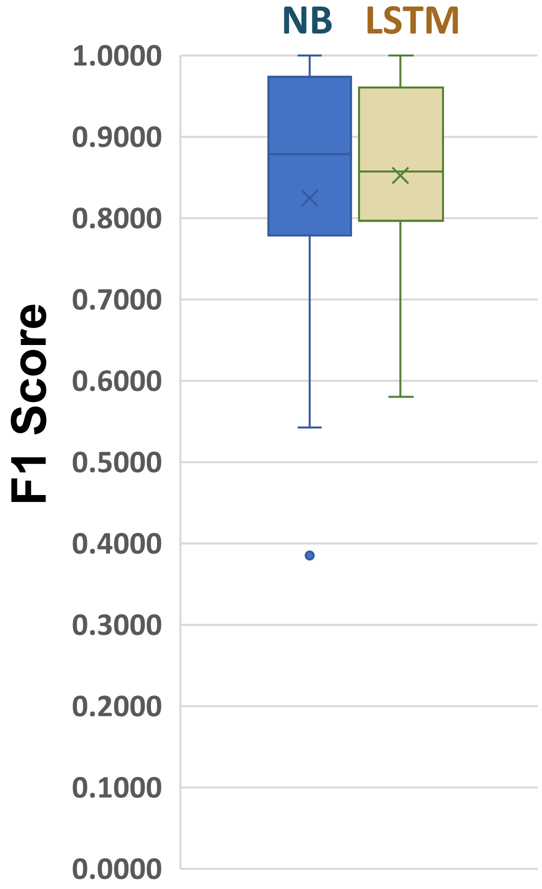
Model Linking

Map recognized entities to domain model

Entities: ip, non-ldif
Edges: flows

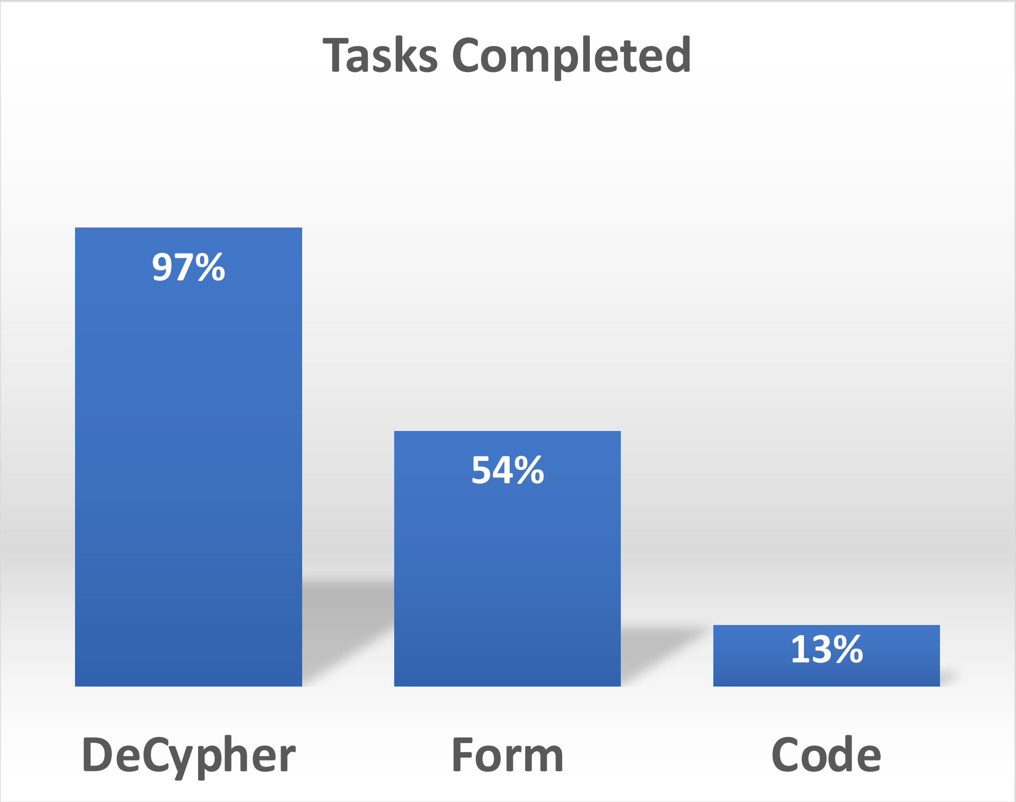
User Intent Classification Performance

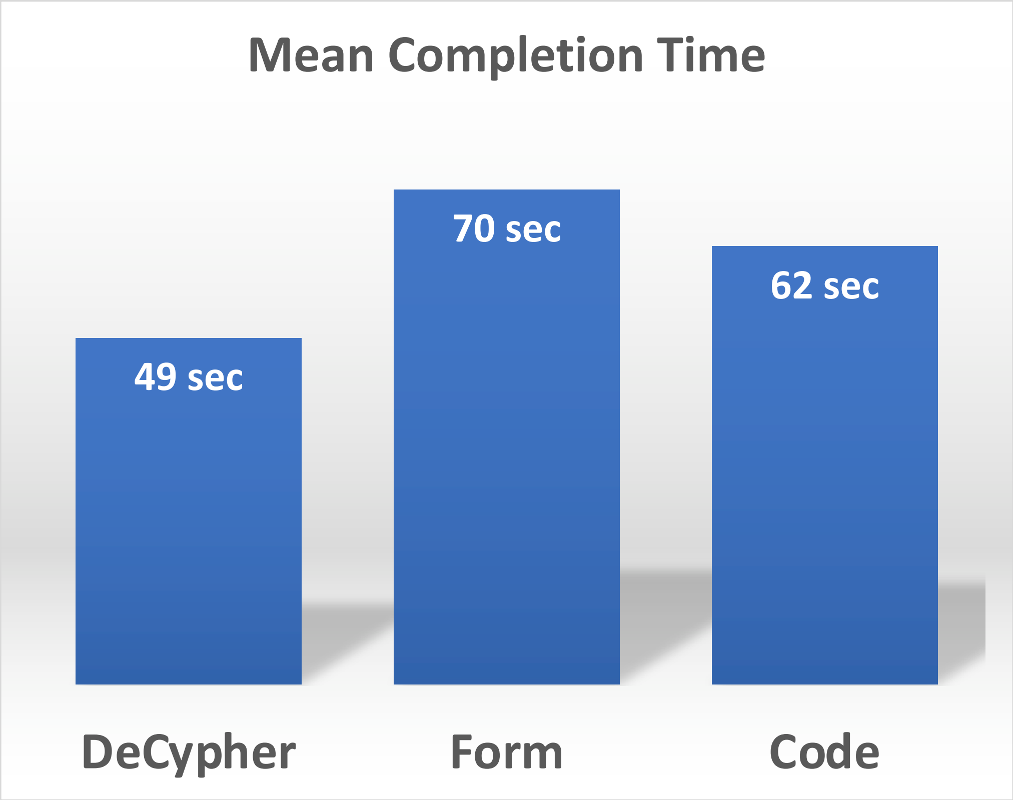




Analytic Task Performance

Task time limit: 2 minutes



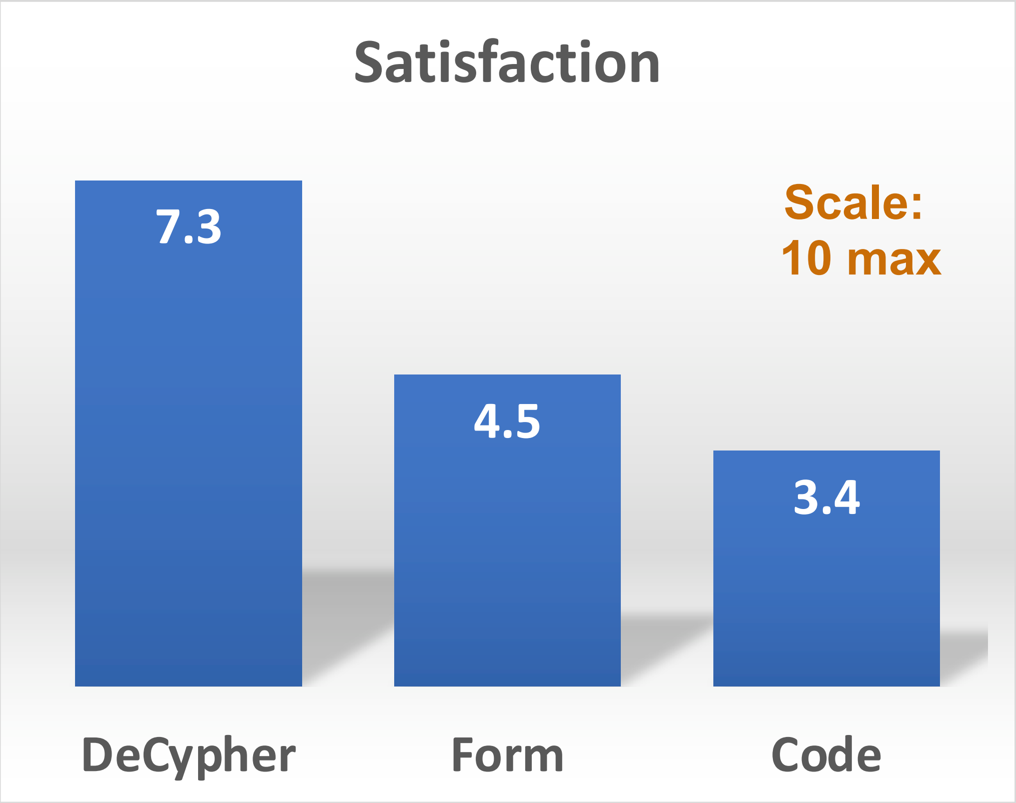


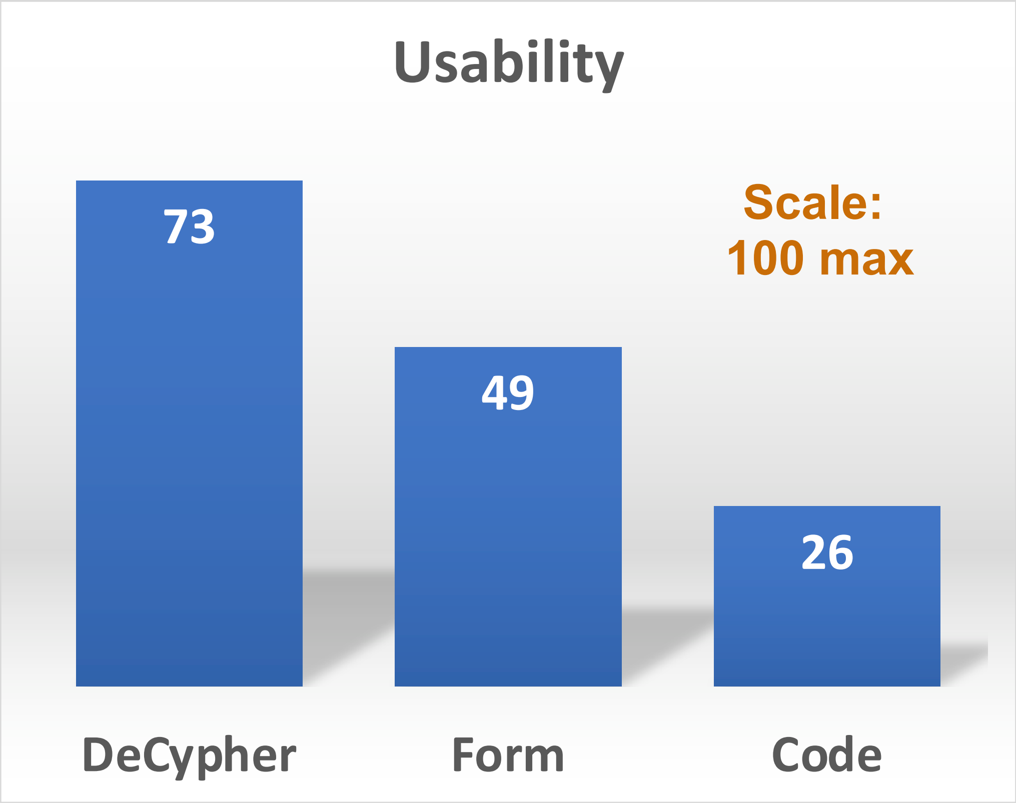
DeCypher allowed analysts to complete almost *2x more tasks*

Tasks completed (within limit) *~20-30% faster* with DeCypher

MITRE

User Satisfaction and Perceived Usability





User satisfaction *improved 62-115%* (1.6-2.1x) with DeCypher

Perceived usability *improved 49-181%* (1.5-2.8x) with DeCypher

Key Impacts

- **CyGraph**
  - Real-time Cyber SU for complex interactions in cyberspace, especially important for operating in stressful environments
  - Amplifies operator capabilities by correlating a myriad of data elements with multiple constraints, yielding new insights for Cyber SU
- **DeCypher**
  - Natural language interface greatly reduces the time for answering cyber operational questions, eliminating the need to write complex queries
  - Significant enhancement in operational Cyber SU outcomes, e.g., productivity boosts, reduced cognitive load, improved user experience, especially helpful for making non-specialists effective
- **Components of reference implementation for Cyber SU tool fielded to operational units** (https://peoc3t.army.mil/mc/mcc.php)

MITRE

Steven Noel, PhD

snoel@mitre.org