Wichita State University

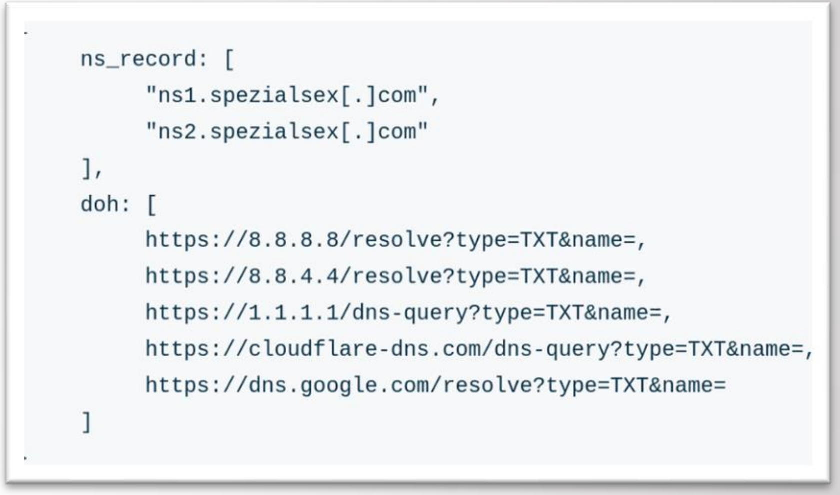
Network Intrusion Detection With Machine Learning and Artificial Intelligence

Outline

- **Motivating Example: Data Exfiltration via DNS-over-HTTPS**
- **Data Science Workflow**
  - Overview of steps from data collection to analysis
- **Machine Learning Models**
  - Key types and applications
  - Performance Evaluation Metrics
- **Deep Learning Models**
  - Advanced approaches and their use cases
  - Evaluation and Optimization Techniques

Why should we care about DNS over HTTPS (DoH)?

- Detection of malicious DoH traffic is hard due to the use of encryption
- Blocking DoH traffic can be difficult since they utilize legitimate public servers, e.g. Google DNS, Cloudflare, etc.
- DoH is becoming increasingly popular
- Recent malware has picked up this protocol

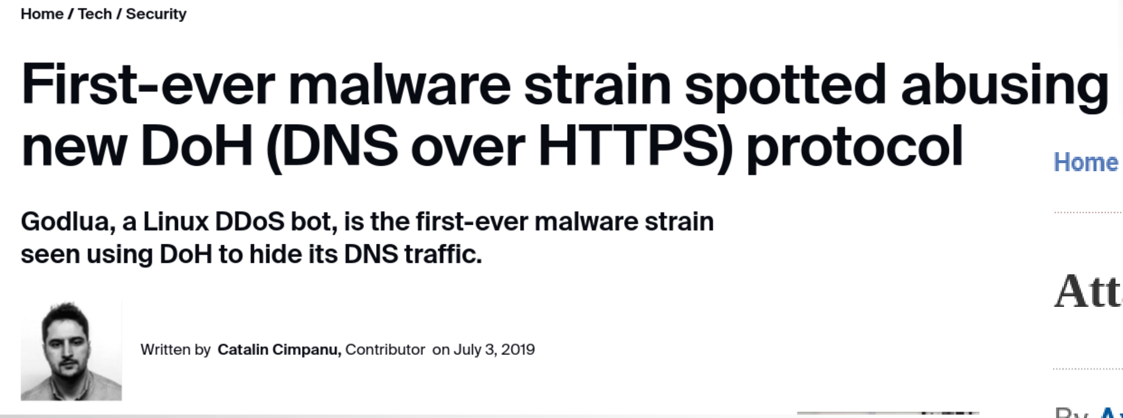


```
ns_record: [
      "ns1.spezialsex[.]com",
      "ns2.spezialsex[.]com"
],
doh: [
      https://8.8.8.8/resolve?type=TXT&name=,
      https://8.8.4.4/resolve?type=TXT&name=,
      https://1.1.1.1/dns-query?type=TXT&name=,
      https://cloudflare-dns.com/dns-query?type=TXT&name=,
      https://dns.google.com/resolve?type=TXT&name=
]
```

ChamelDoH configuration JSON

Source: https://stairwell.com/news/chamelgang-and-chameldoh-a-dns-over-https-implant/

Examples of Malware using DNS-over-HTTPS

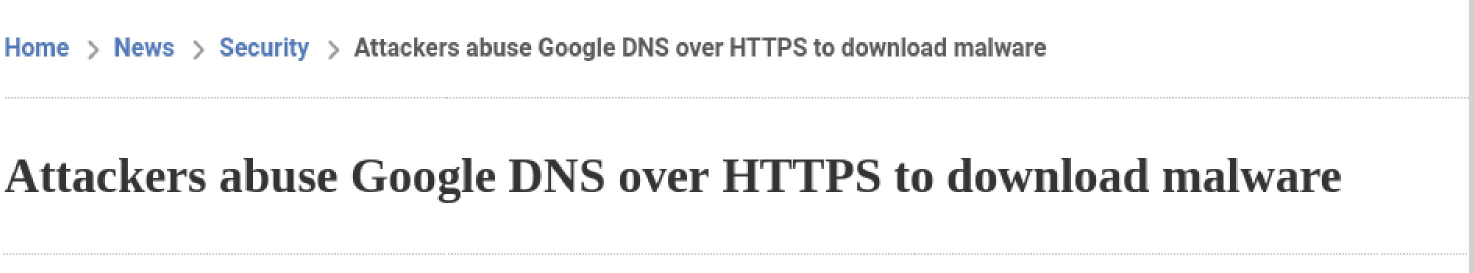


Home / Tech / Security

First-ever malware strain spotted abusing new DoH (DNS over HTTPS) protocol

Godlua, a Linux DDoS bot, is the first-ever malware strain seen using DoH to hide its DNS traffic.

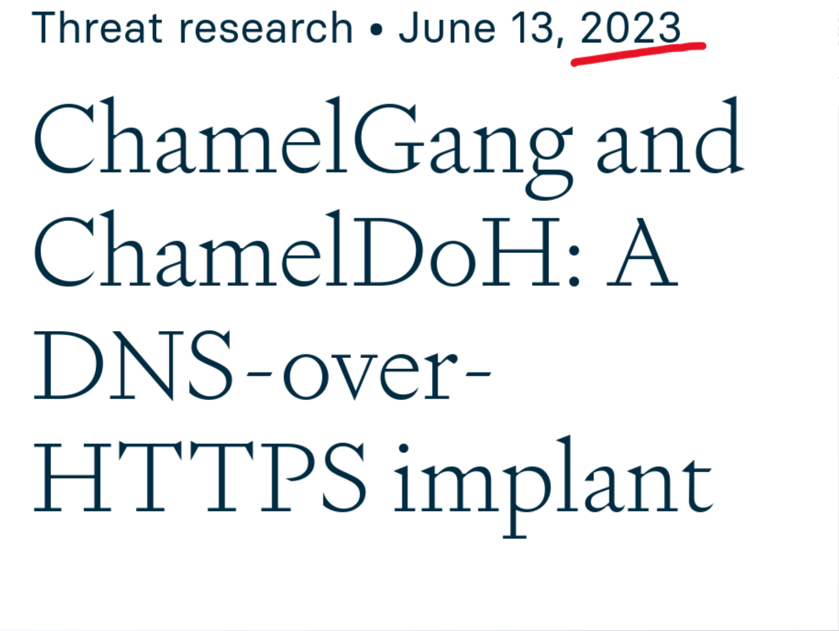
Written by **Catalin Cimpanu,** Contributor on July 3, 2019



Home > News > Security > Attackers abuse Google DNS over HTTPS to download malware

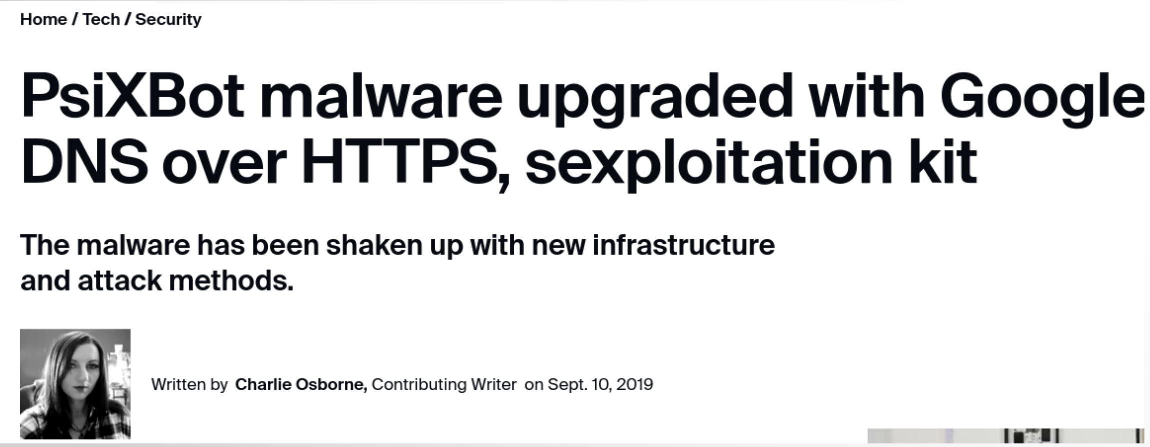
Attackers abuse Google DNS over HTTPS to download malware

By **Ax Sharma** September 2, 2020



Threat research • June 13, 2023

ChamelGang and ChamelDoH: A DNS-over-HTTPS implant



Home / Tech / Security

PsiXBot malware upgraded with Google DNS over HTTPS, sexploitation kit

The malware has been shaken up with new infrastructure and attack methods.

Written by **Charlie Osborne,** Contributing Writer on Sept. 10, 2019

Botnet abuse of DoH

Just another step in the cat-and-mouse game between attackers and defenders

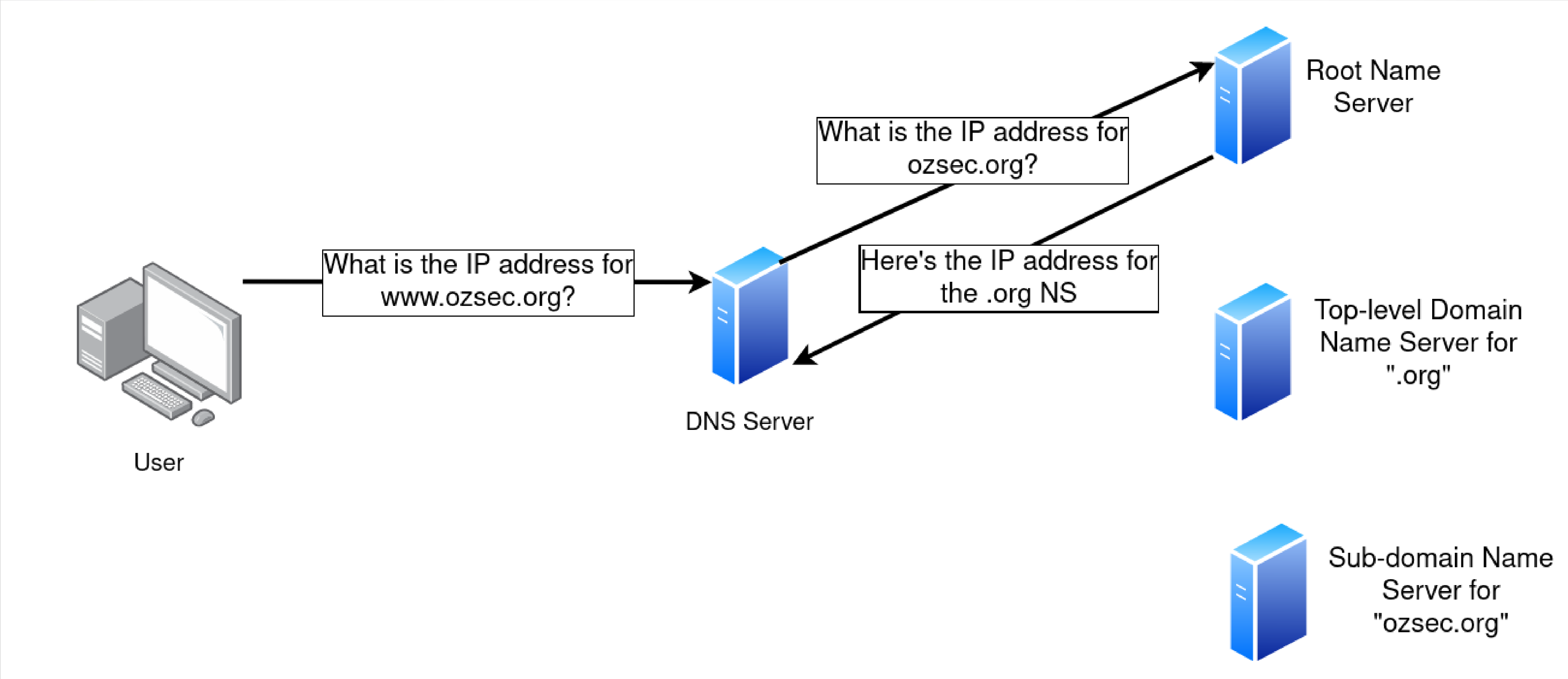
Informal Survey

- Raise your hand if you do not know what is DNS

Informal Survey

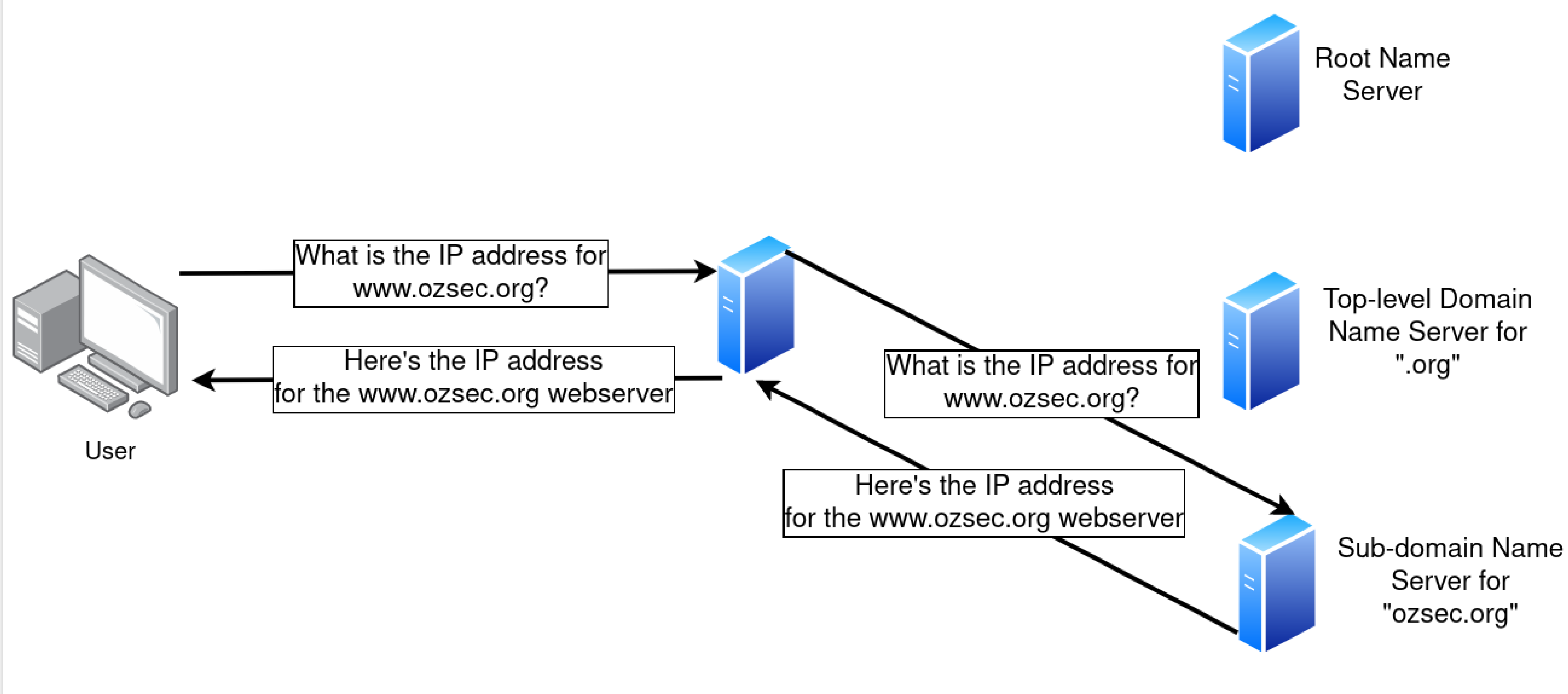
- Raise your hand if you cannot explain how DNS works

How does DNS work?


Root Name Server
What is the IP address for ozsec.org?
Here's the IP address for the .org NS
What is the IP address for www.ozsec.org?
Top-level Domain Name Server for ".org"
DNS Server
User
Sub-domain Name Server for "ozsec.org"

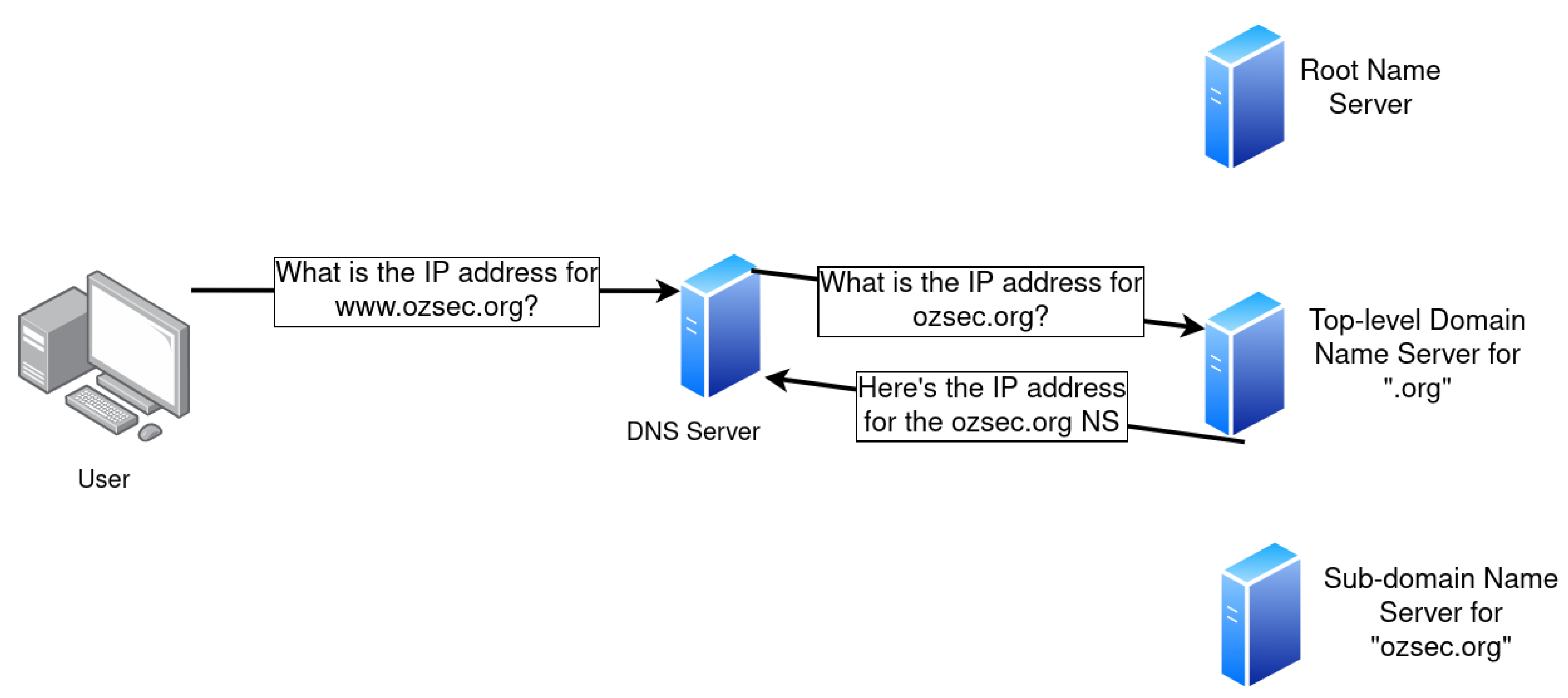
The user's computer initiates a DNS query with the DNS server.

How does DNS work?


Root Name Server
What is the IP address for www.ozsec.org?
Top-level Domain Name Server for ".org"
Here's the IP address for the www.ozsec.org webserver
What is the IP address for www.ozsec.org?
User
Here's the IP address for the www.ozsec.org webserver
Sub-domain Name Server for "ozsec.org"

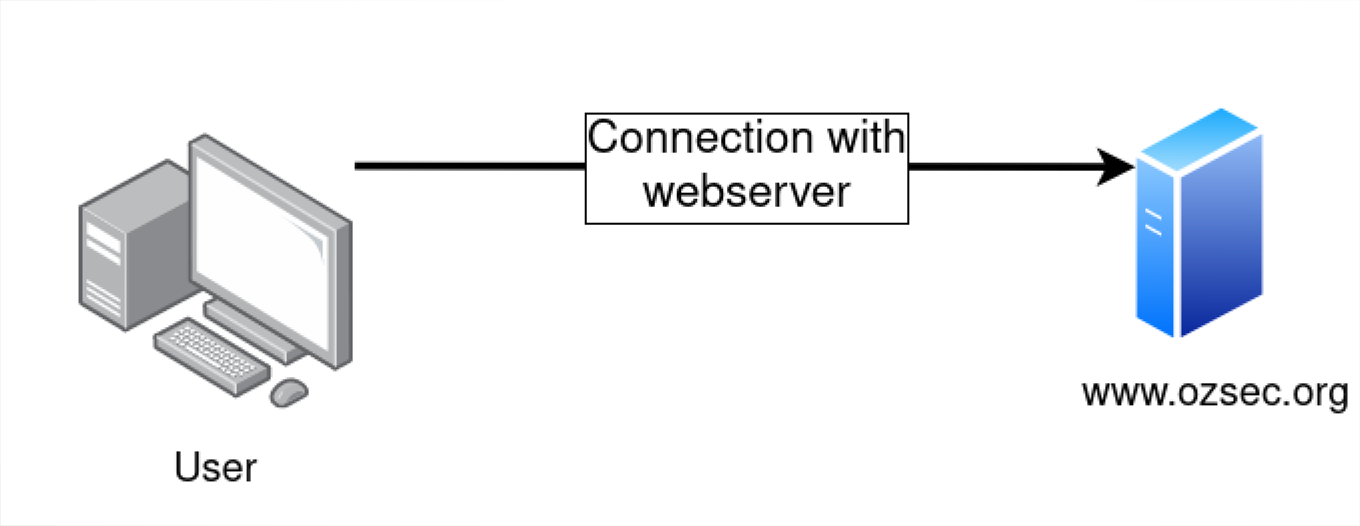
The DNS server iteratively queries the root, top-level, and sub-domain name servers. We assume this is a DNS request that has not been cached.

How does DNS work?


Root Name Server
What is the IP address for www.ozsec.org?
What is the IP address for ozsec.org?
Top-level Domain Name Server for ".org"
Here's the IP address for the ozsec.org NS
User
DNS Server
Sub-domain Name Server for "ozsec.org"

The user's host received the DNS query result.

How does DNS work?



The user's host establishes a connection to the servers using the DNS query result.

DNS Abuse

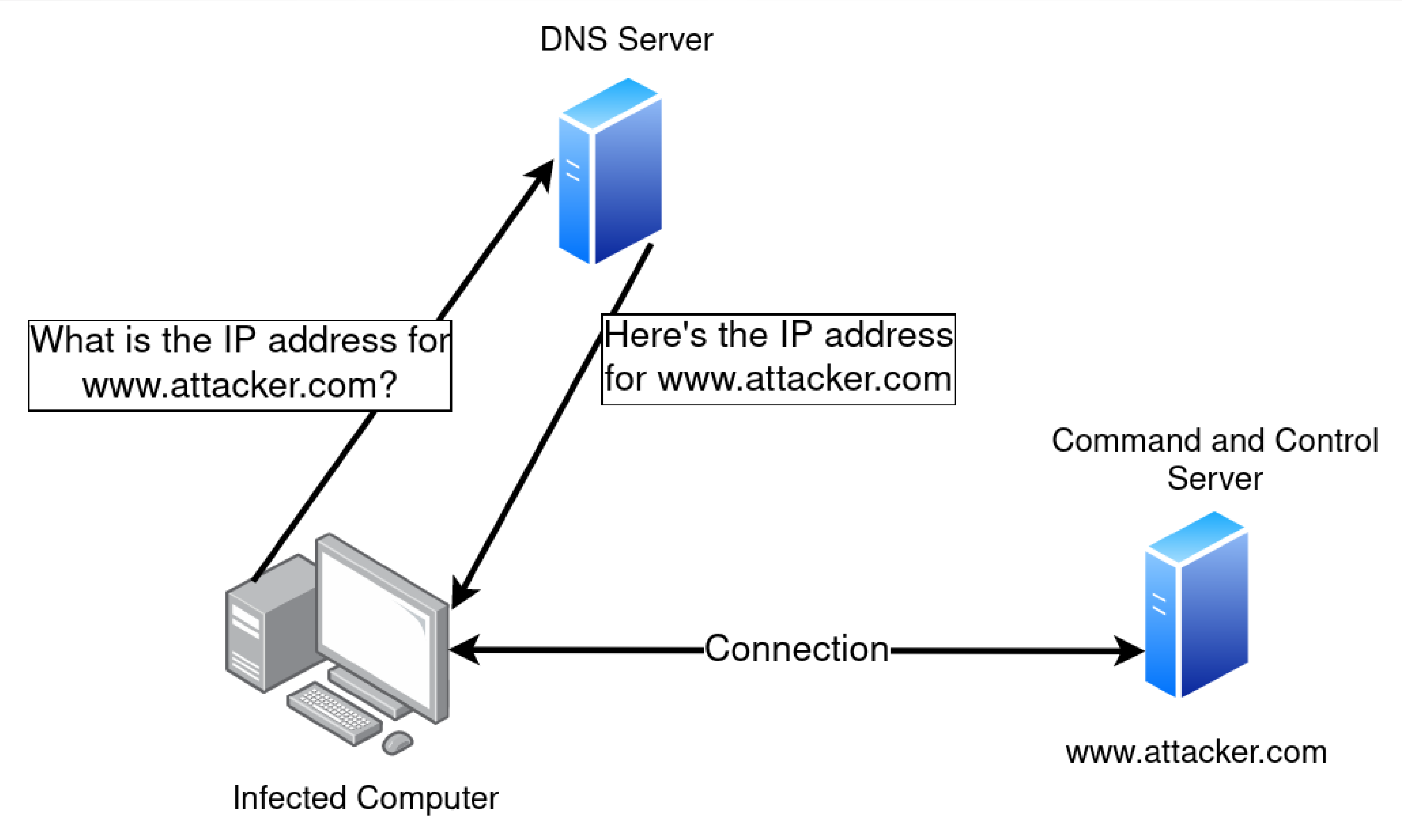
- What are some possible unintended uses of DNS?

How does malware abuse DNS?

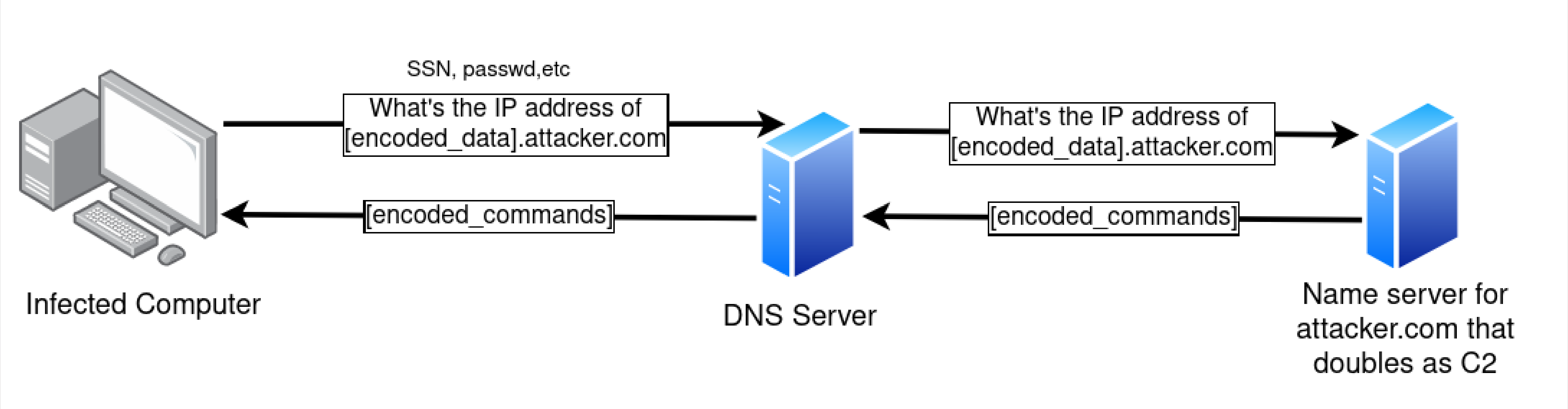
1. Finding command and control servers

2. Data exfiltration

Finding the IP address of C2


DNS Server
What is the IP address for www.attacker.com?
Here's the IP address for www.attacker.com
Command and Control Server
Connection
Infected Computer
www.attacker.com

Data Exfiltration


SSN, passwd,etc
What's the IP address of [encoded_data].attacker.com
What's the IP address of [encoded_data].attacker.com
[encoded_commands]
[encoded_commands]
Infected Computer
DNS Server
Name server for attacker.com that doubles as C2

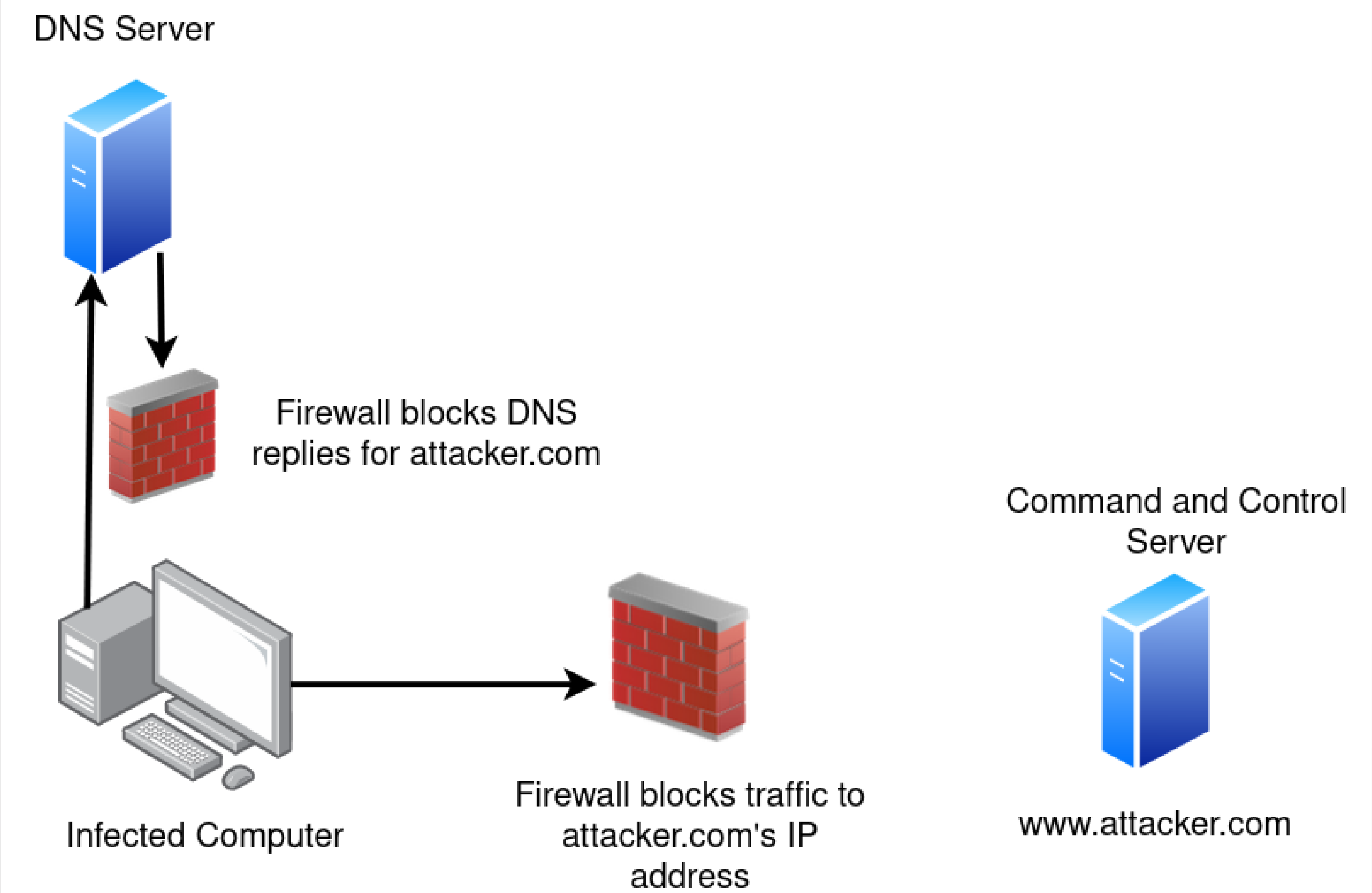
Defenses against DNS Abuse

- What are some possible defenses against finding C2s?
- What are some possible defenses against data exfiltration?

What can defenders do?

1. Block known malicious IP addresses and domain names

Easy to block by firewall once domain/IP is known


DNS Server
Firewall blocks DNS replies for attacker.com
Command and Control Server
Infected Computer
Firewall blocks traffic to attacker.com's IP address
www.attacker.com

- DNS Server can also ignore DNS requests for attacker.com
- The firewall can be running in the infected computer, in the local/enterprise network, or ISP

Bypassing blocked IP addresses/domains

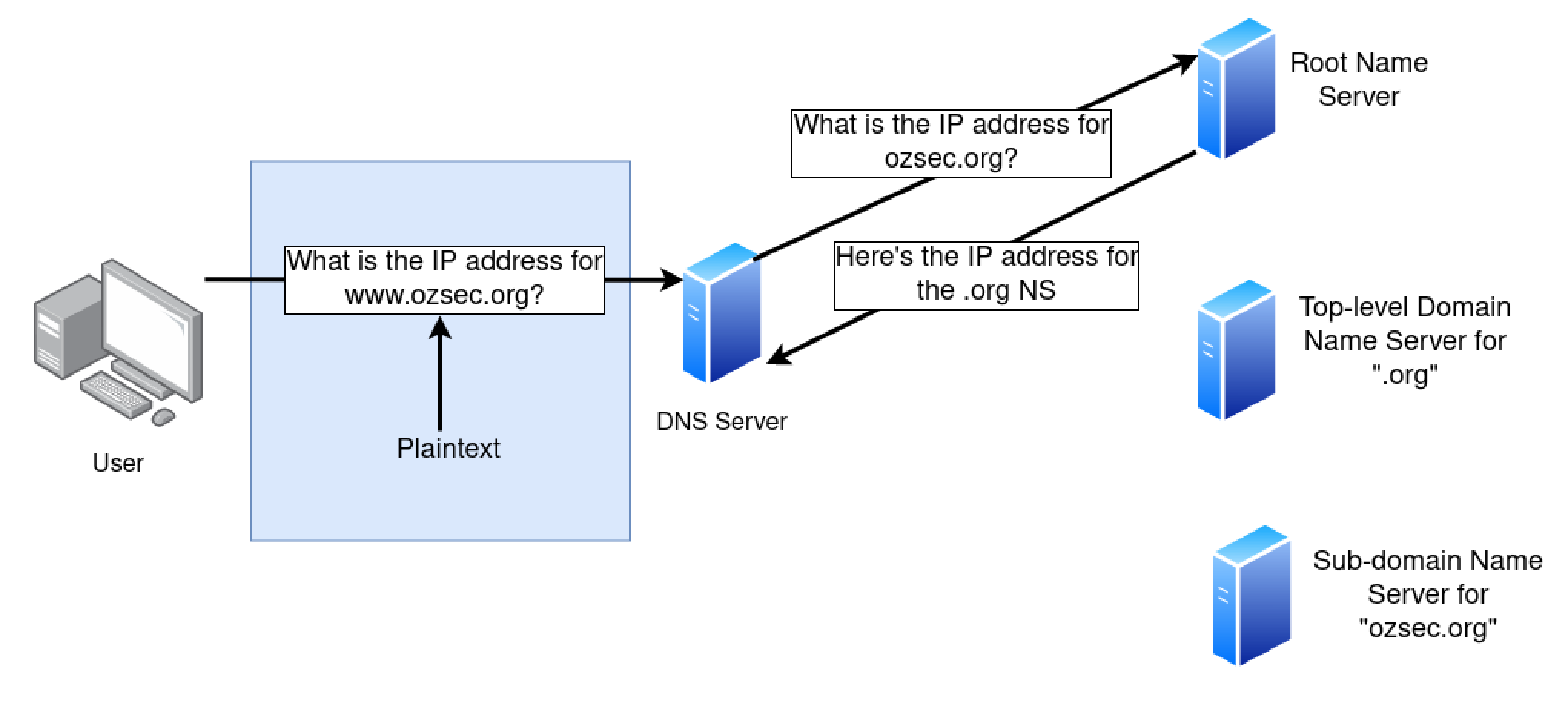
- What are some possible approaches for attackers to bypass blocked IP addresses or domains?

How have attackers adapted?

- Domain generation algorithms
  - Allows infected devices to find C2

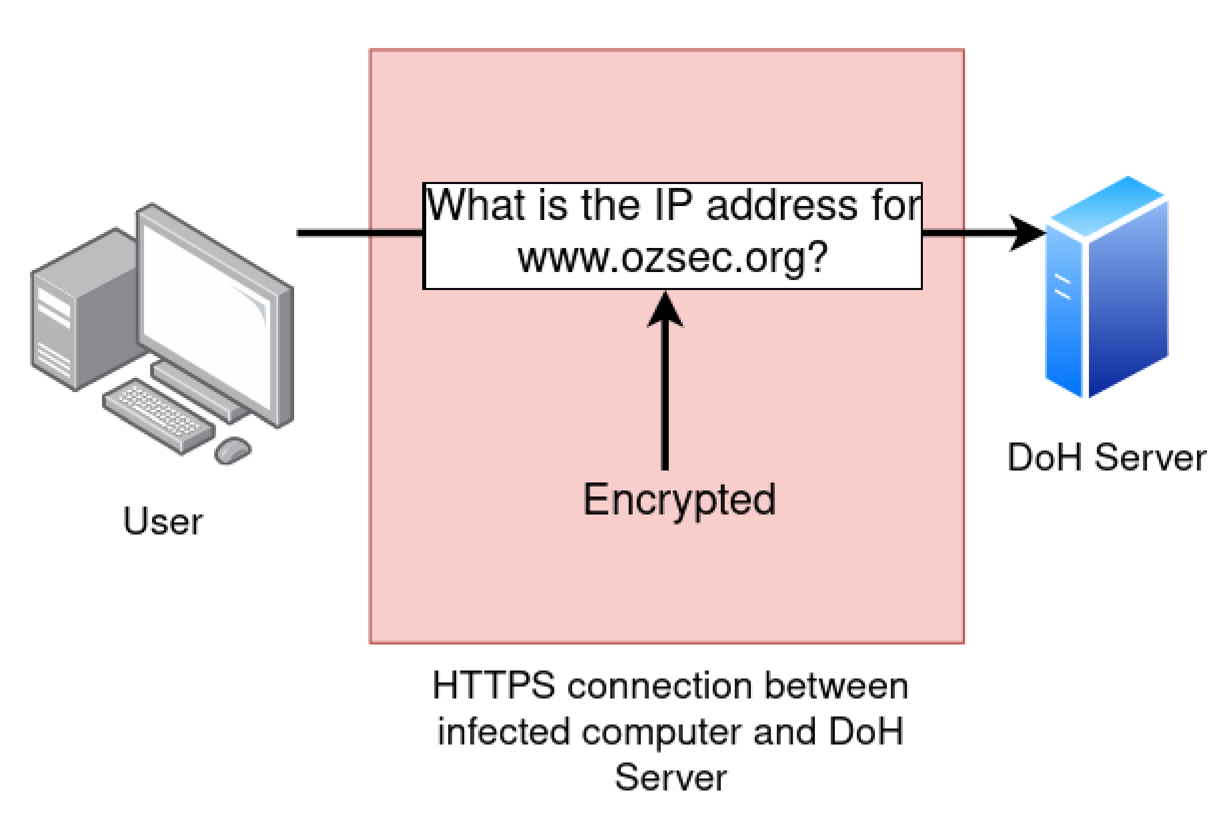
- DNS-over-HTTPS
  - Allows infected devices to encrypt data exfiltration packets

Main DNS Advantage/Flaw from Defender/Malware Perspective

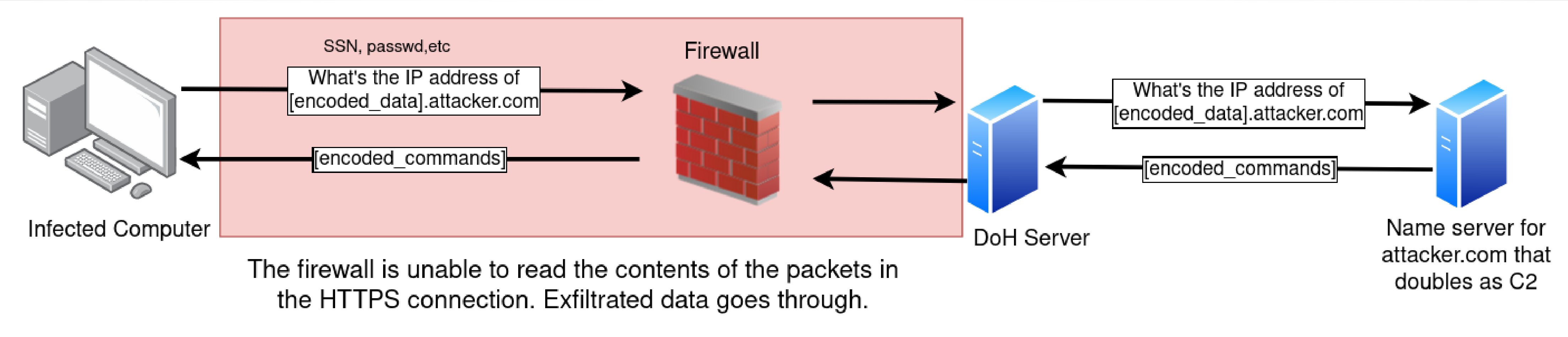


Root Name Server
What is the IP address for ozsec.org?
Here's the IP address for the .org NS
What is the IP address for www.ozsec.org?
Plaintext
User
DNS Server
Top-level Domain Name Server for ".org"
Sub-domain Name Server for "ozsec.org"

DoH helps Malware Bypass Detection


What is the IP address for www.ozsec.org?
Encrypted
User
DoH Server
HTTPS connection between infected computer and DoH Server

Data Exfiltration and C2 Localization with DoH

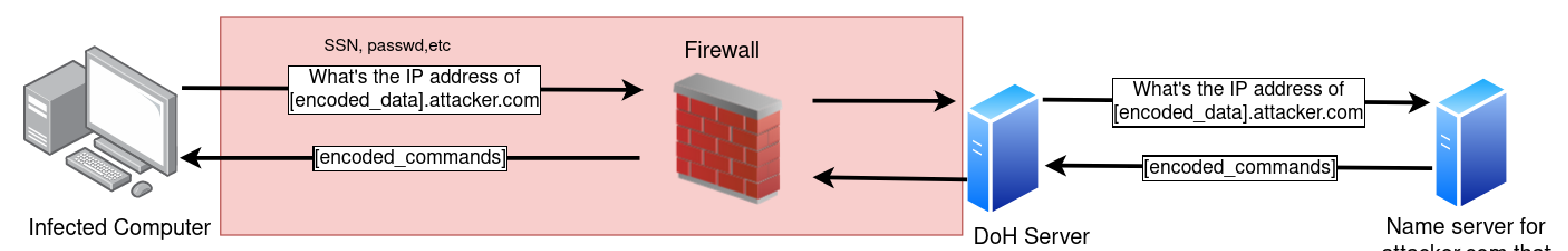

SSN, passwd,etc
Firewall
What's the IP address of [encoded_data].attacker.com
[encoded_commands]
Infected Computer
What's the IP address of [encoded_data].attacker.com
[encoded_commands]
DoH Server
Name server for attacker.com that doubles as C2

The firewall is unable to read the contents of the packets in the HTTPS connection. Exfiltrated data goes through.

- Your turn to take the next step against the attackers

Your turn to take the next step against the attackers

- *If we cannot observe the domain names in the DNS packets, what other information can we use to detect data exfiltration by DoH-based malware?*
  - *Write down 2-3 ideas*
  - *Share with student next to you*

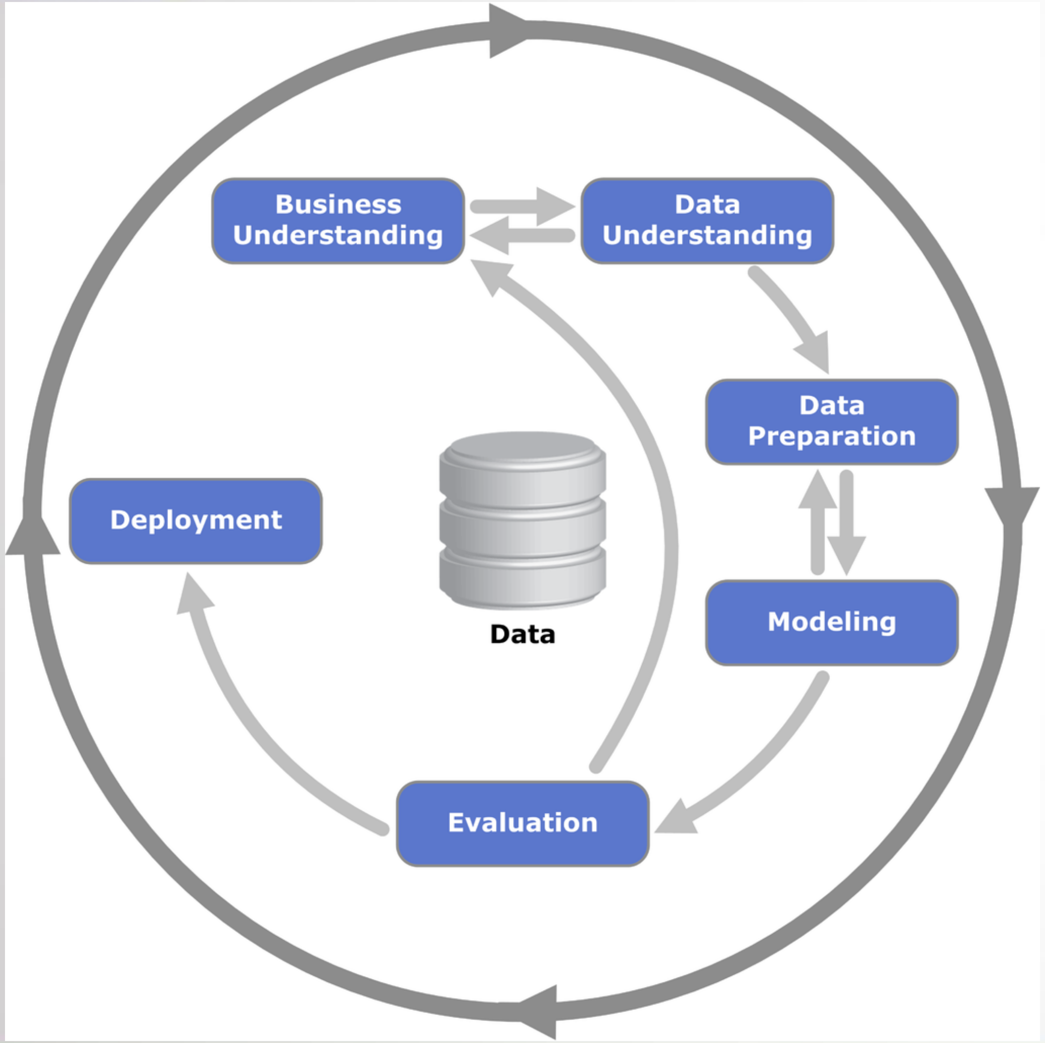

SSN, passwd,etc
What's the IP address of [encoded_data].attacker.com
[encoded_commands]
Firewall
Infected Computer
DoH Server
What's the IP address of [encoded_data].attacker.com
[encoded_commands]
Name server for attacker.com that doubles as C2

The firewall is unable to read the contents of the packets in the HTTPS connection. Exfiltrated data goes through.

One possible solution

WICHITA STATE UNIVERSITY

The Data Science Workflow

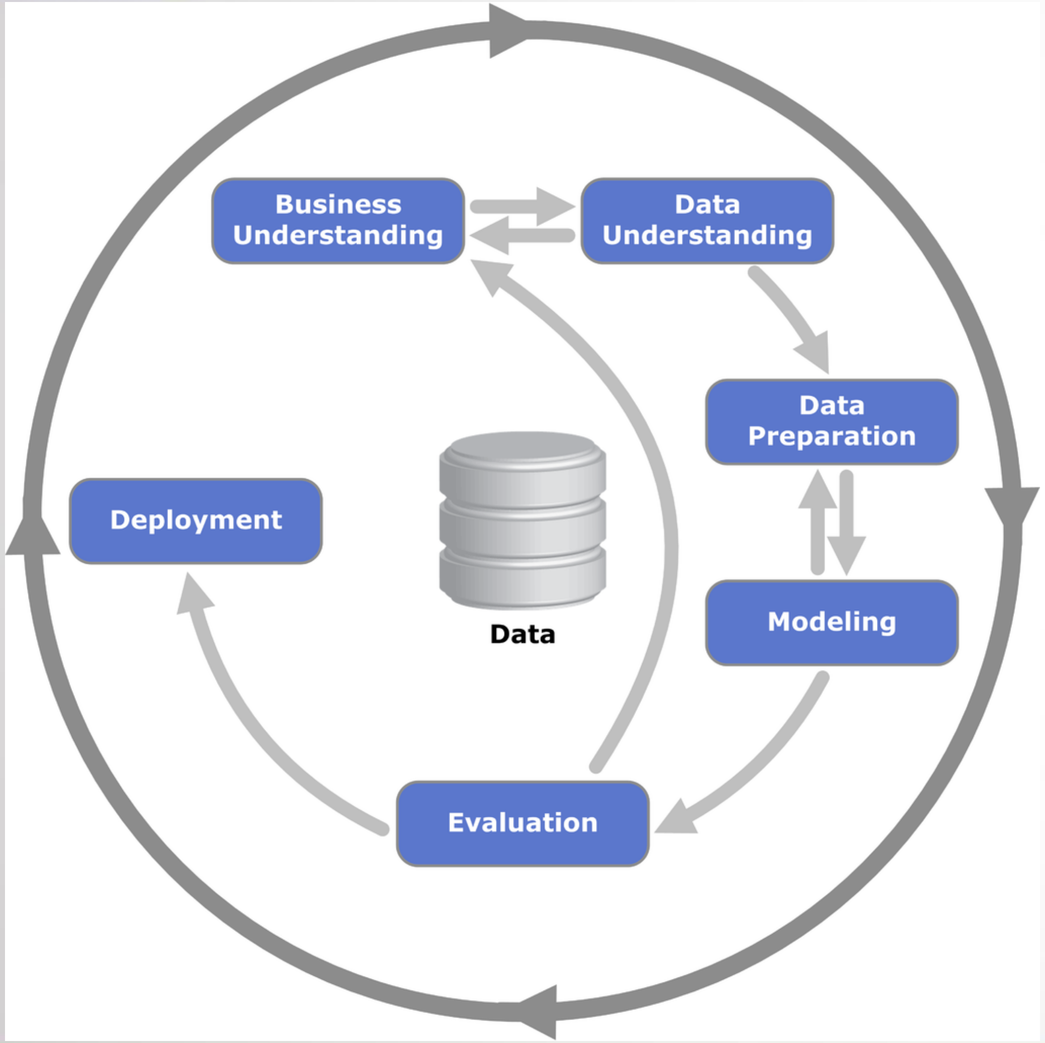


- Cross-industry standard process for data mining (CRISP)
- General data science framework

Source: https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining

WICHITA STATE UNIVERSITY

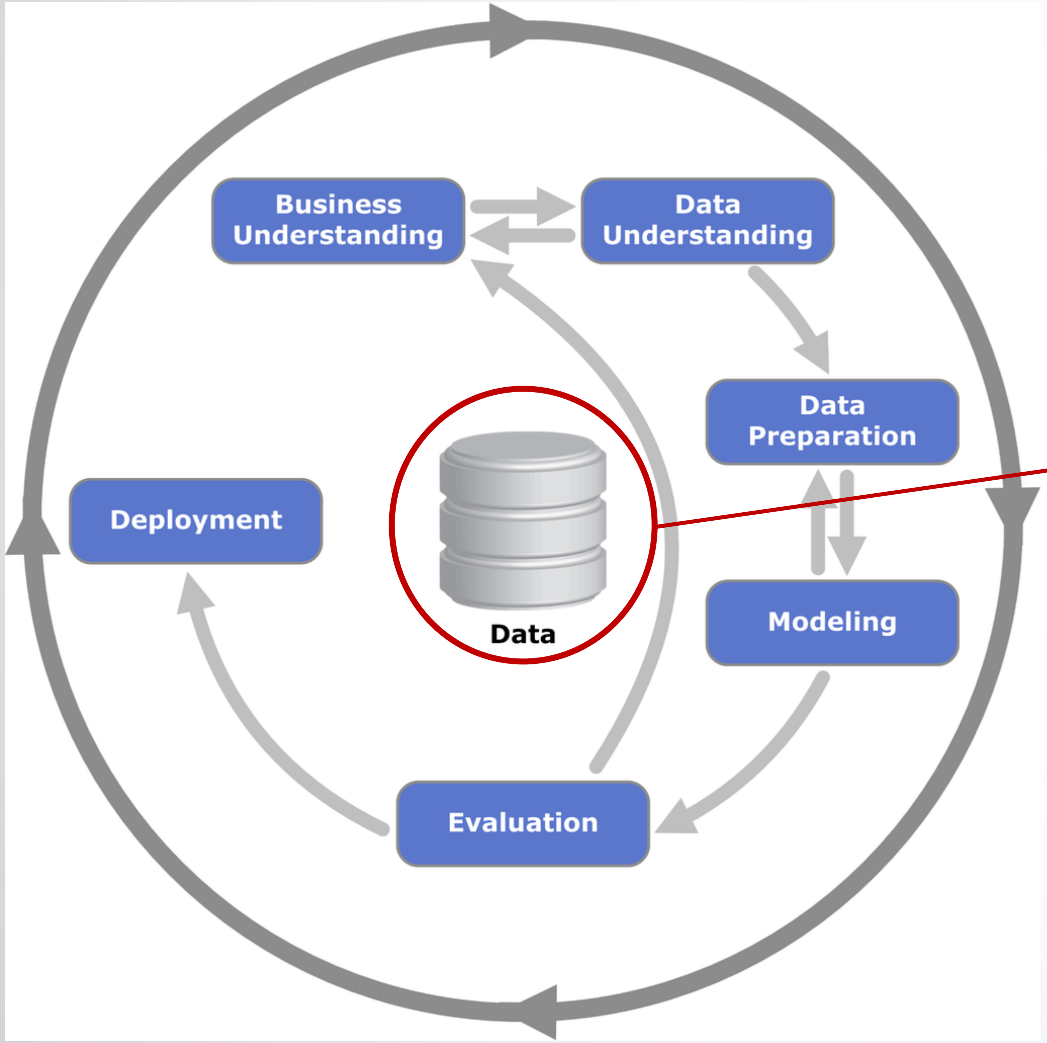
The Data Science Workflow



- Cross-industry standard process for data mining (CRISP)
- General data science framework

Source: https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining

CRISP for DoH Detection

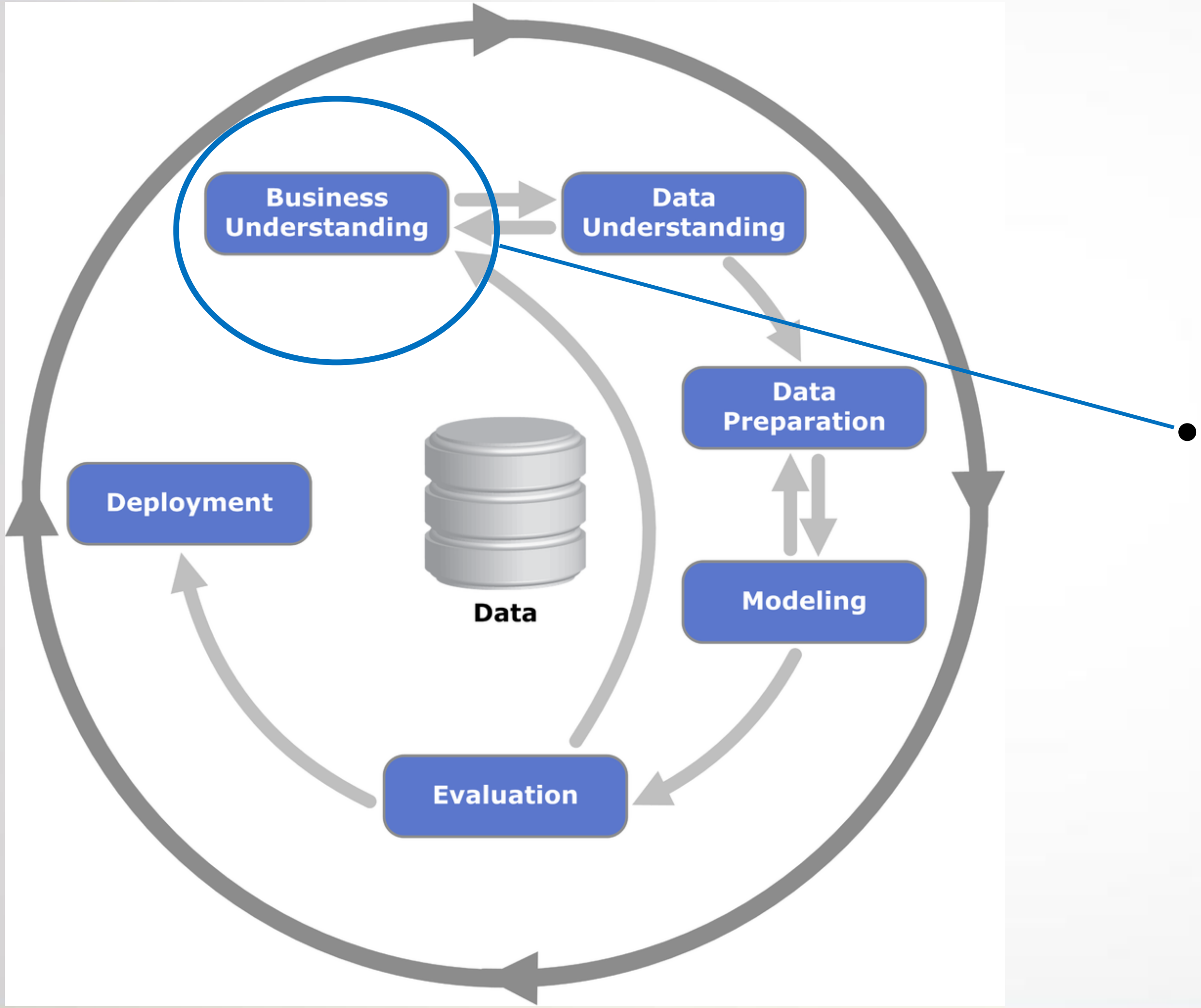


- Packet Captures (PCAPs)

Source: https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining

WICHITA STATE UNIVERSITY

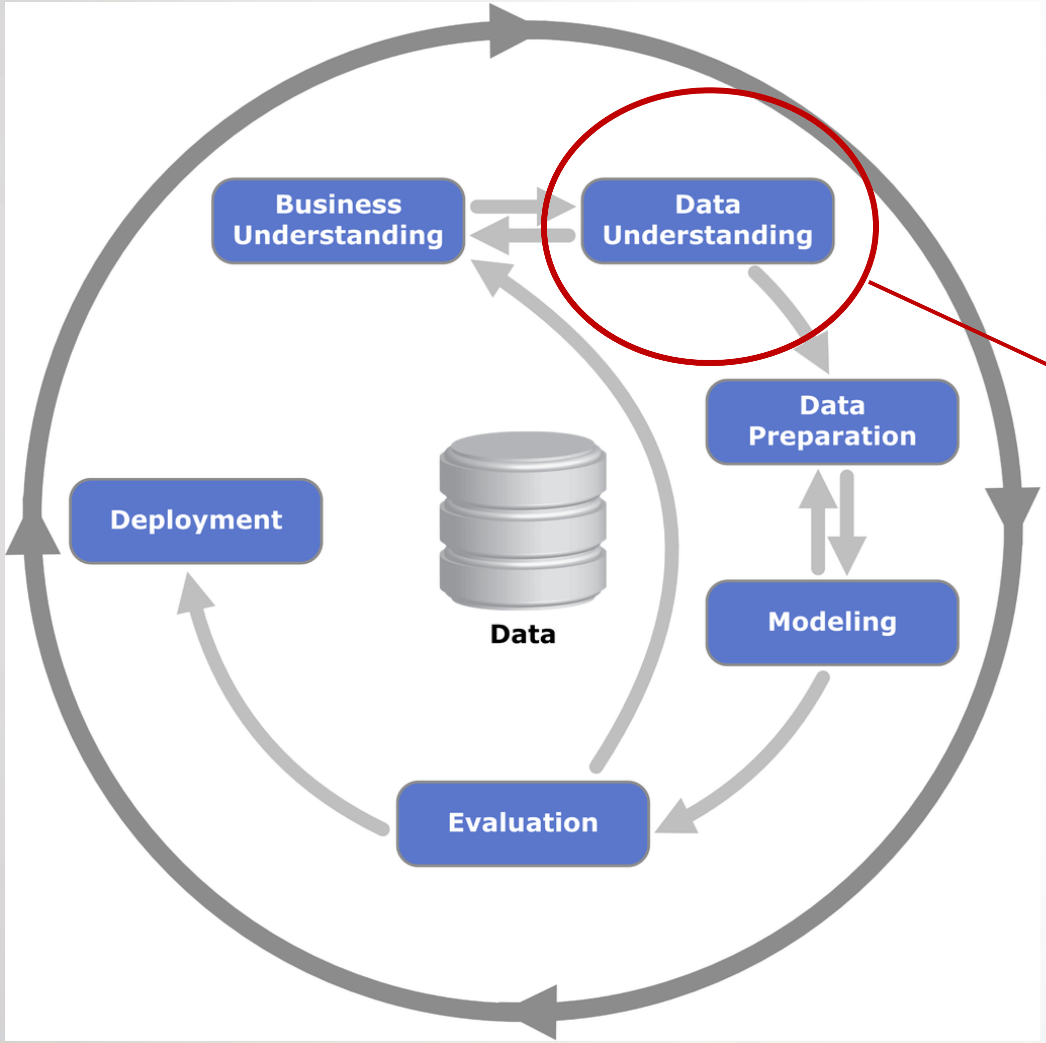
CRISP for DoH Detection



- How does malware affect your organization?

Source: https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining

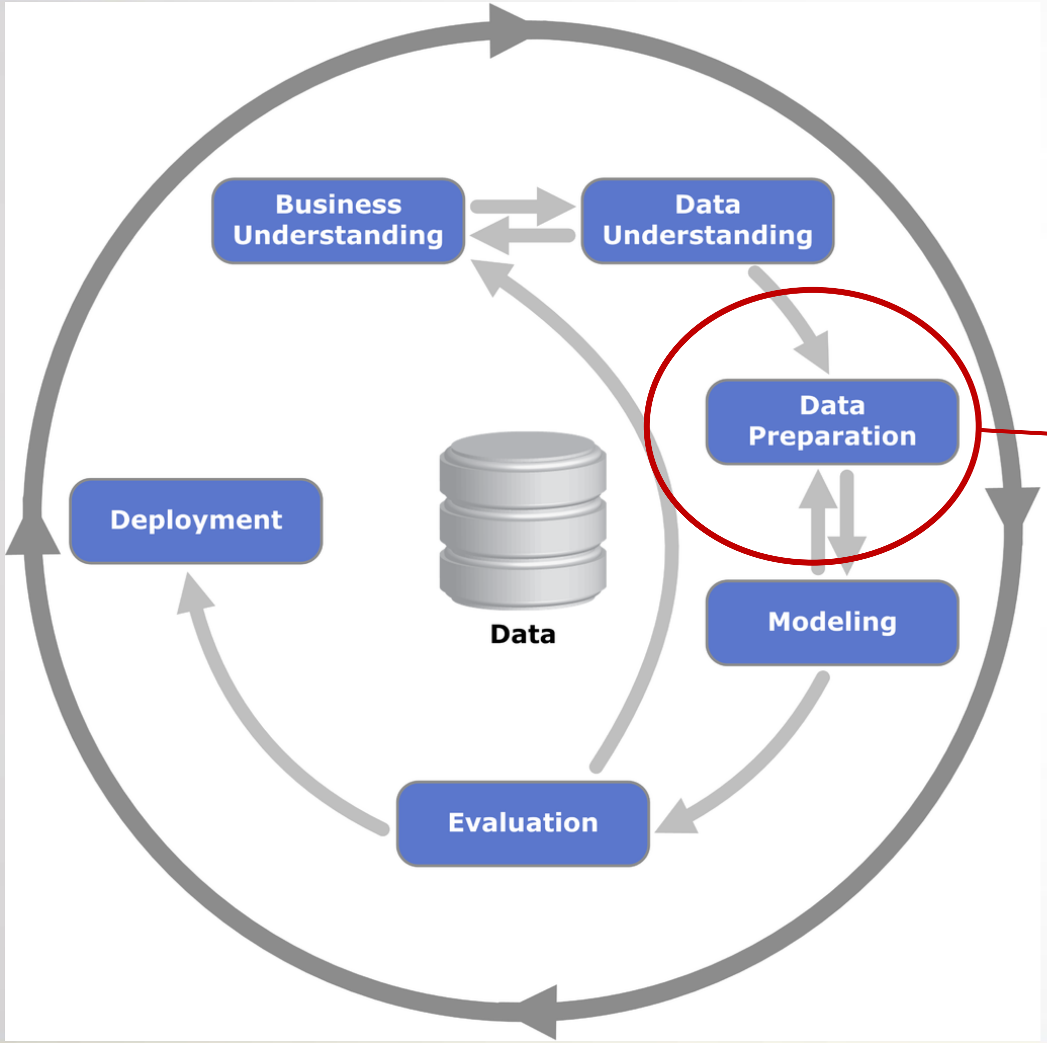
CRISP for DoH Detection



- What type of traffic do we have?
- TCP? HTTPS?DoH?
- Are there any correlations between them?
- What are the most common servers?
- etc

Source: https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining

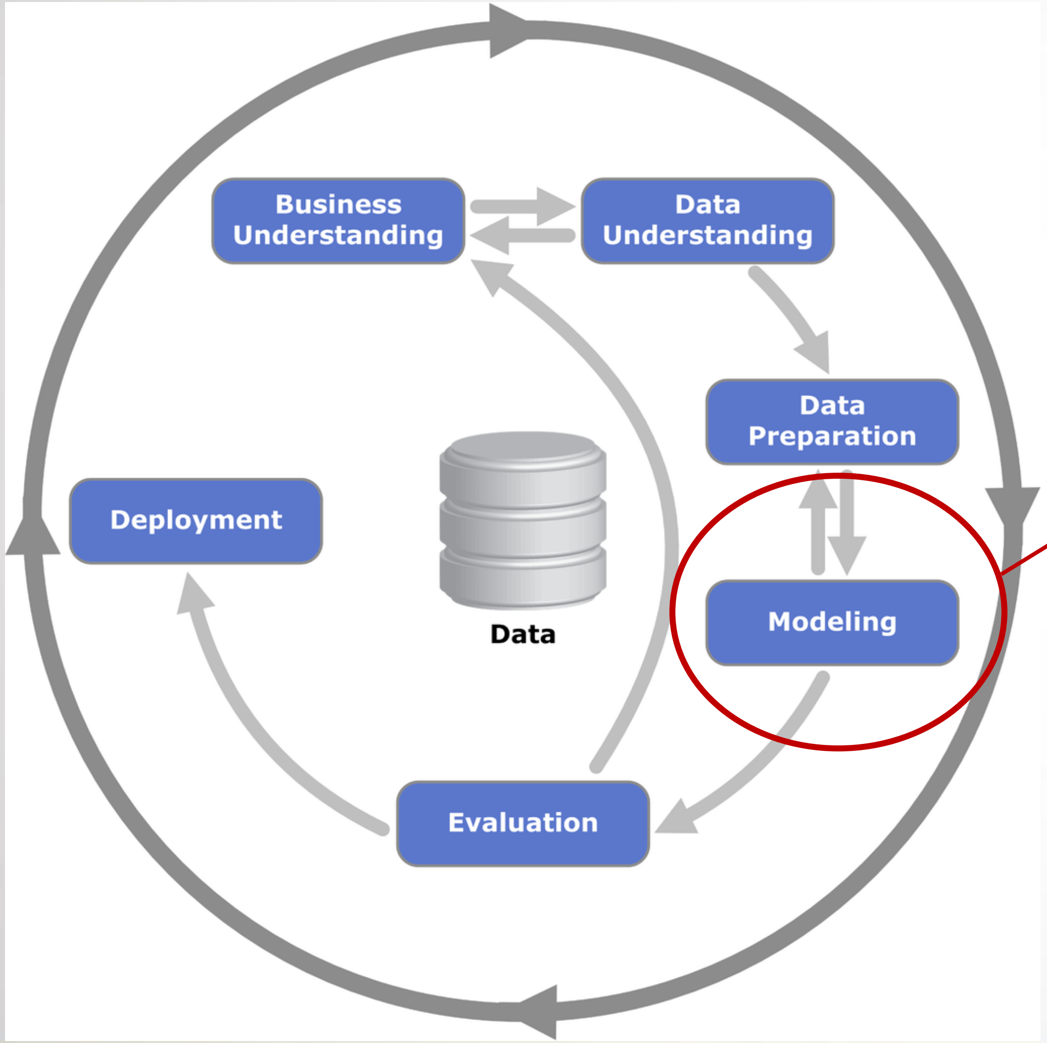
CRISP for DoH Detection



- Find a simple numerical representation of the TCP connections

Source: https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining

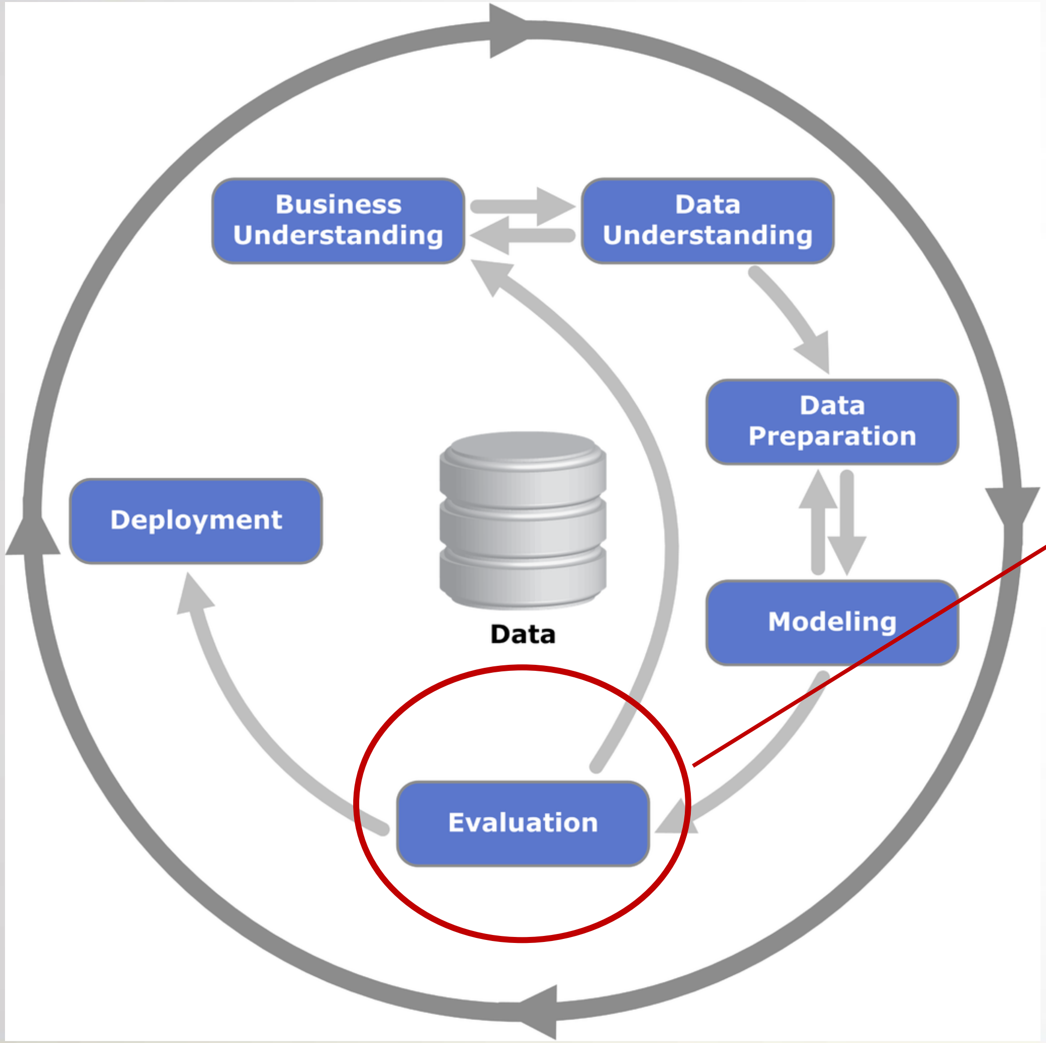
CRISP for DoH Detection



- Train a machine learing model

Source: https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining

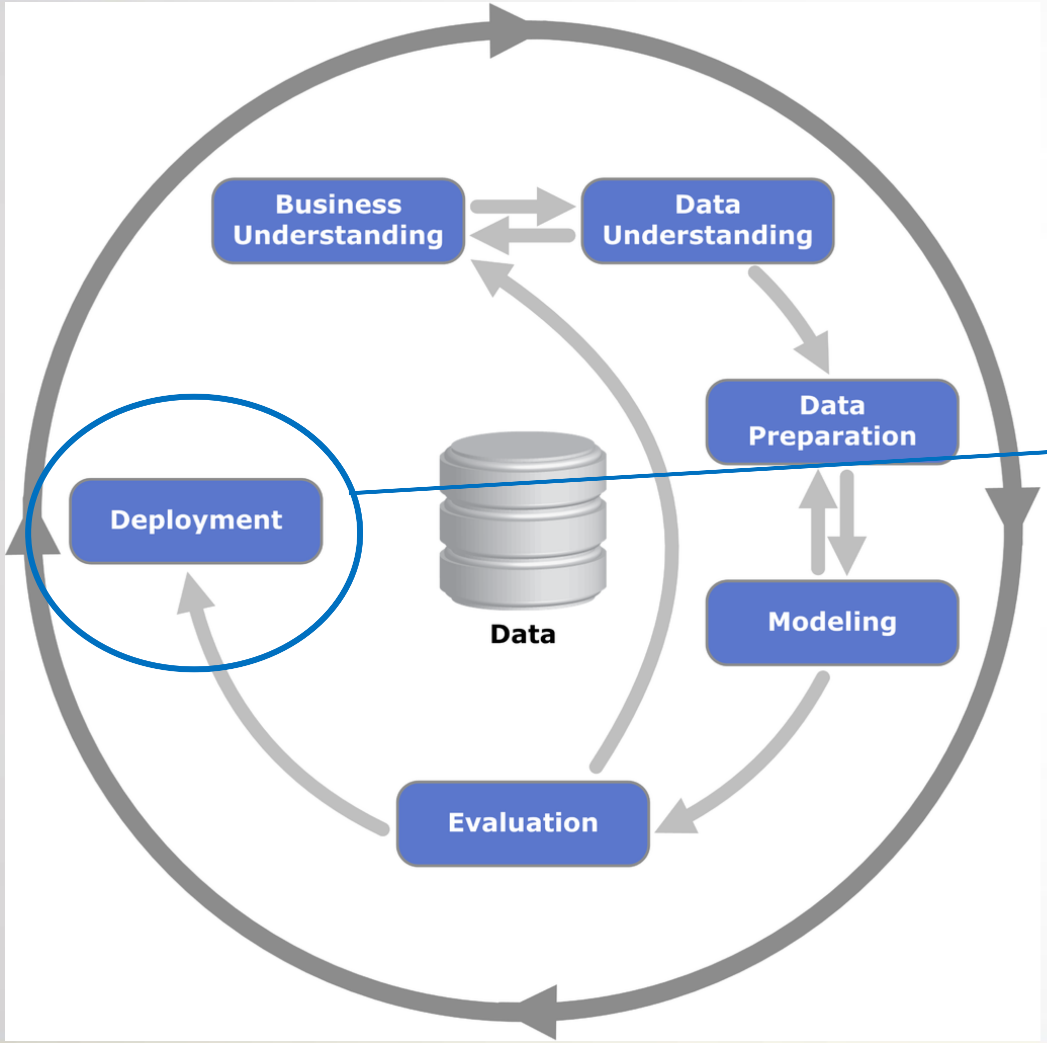
CRISP for DoH Detection



- Can the trained machine learning model actually find malicious DoH traffic?

Source: https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining

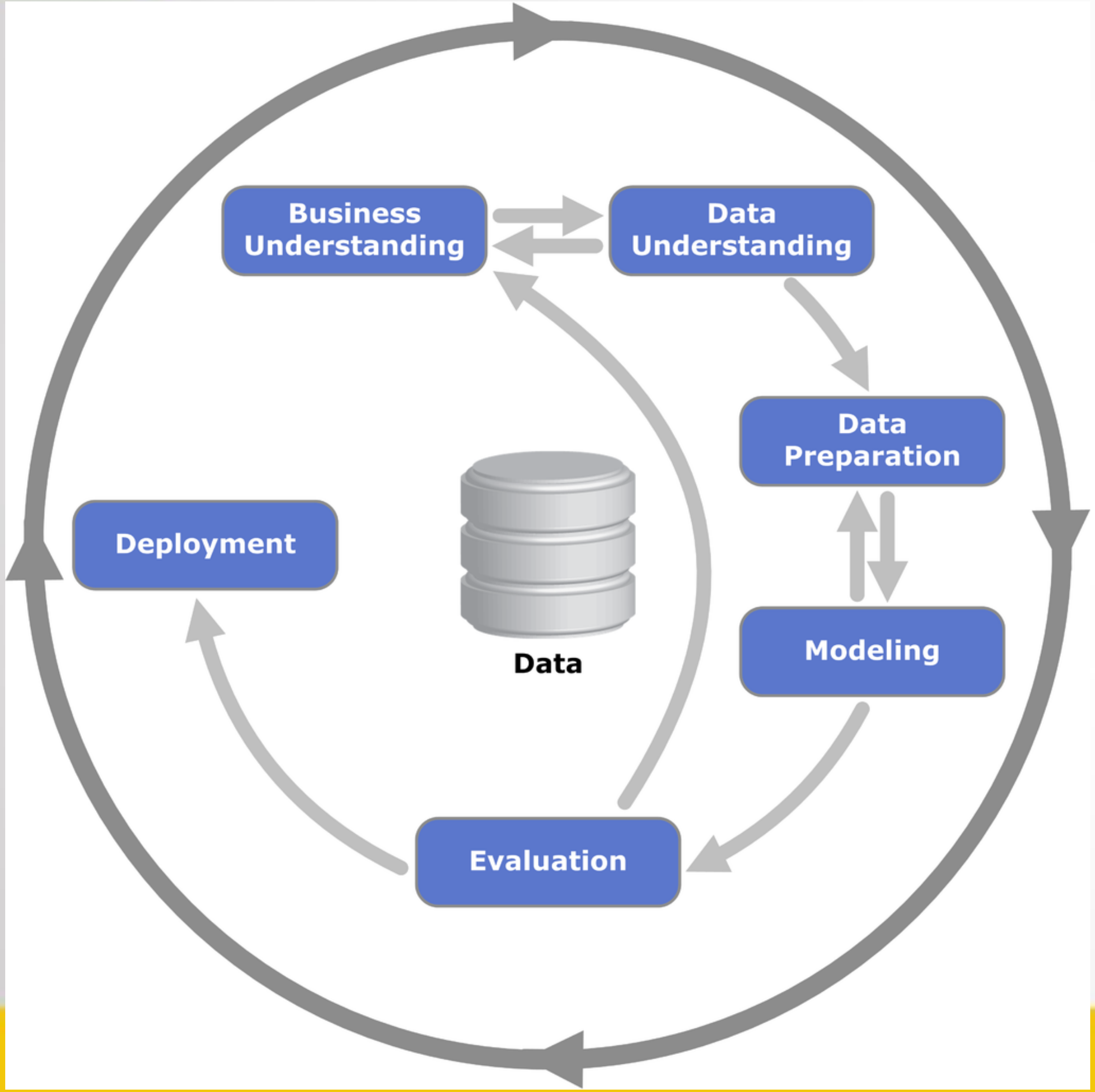
CRISP for DoH Detection



- How will the the ML model be used?
- Can it be incorporated to existing tools?

Source: https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining

Summary of CRISP for DoH Detection



- Data
  - Packet captures (PCAPs)
- Business understanding
  - How does malware affect your organization?
- Data understanding
  - What type of traffic ? What servers? How many TCP connections?
- Data Preparation
  - Create a simple numerical representation of the data
- Modeling
  - Machine learning models
- Evaluation
  - How many malicious DoH connections can we detect?
- Deployment
  - How will you use the alerts?

Hands-on Activity

- The hands-on activity Is documented on the GitHub repository
  - https://github.com/deep-learning-prof/doh-workshop

- You will need:
  - A Jupyter Server: https://jupyter.org/install
  - Wireshark

WICHITA STATE UNIVERSITY

Key Takeaways

- The concept of finding attacks based on anomaly detection applies to any type of data and attack
- Training machine learning models in Python wasn't that hard, was it?
- This is just a quick overview of the whole process
- Keep experimenting!
- Code: https://github.com/deep-learning-prof/doh-workshop

- Contact:
  - Sergio.salinasmonroy@wichita.edu