Unintentional Darknet: An Analysis of Scanning Behavior through the Lens of Abandoned Domains

Chris Kitras and Bryson Schiel

Outline

- Background
- Methodology
- Analysis
- Conclusion

Background

Finding Unexpected Domains

- Excluded substring “byu” from all QNAMEs
- Ensign/LDSBC domains
- Okta, Brightspot, other technologies
- JHDL
    - `whois` identified owner as BYU Configurable Computing Lab

Table 1: Top 15 QNAMEs not including "byu" or its IPv4 Prefix

Query Name	# of Queries
www.ensign.edu.	2109088
ensign.edu.	804684
jhdl.com.	**779440**
ip6 arpa	683525
ntp-nist.ldsbc.net.	606853
msoid.ensign.edu.	616685
rp.ensign.edu.	559277
hcsys.ensign.byu.edu.	319234
ip6 arpa one	163838
okta.ensign.edu.	150022
brightspot.ensign.edu	133090
ip6 arpa	114386
jhdl.org.	**97143**
jhdl.net.	**94732**
ip6 arpa	90627

What is JHDL?

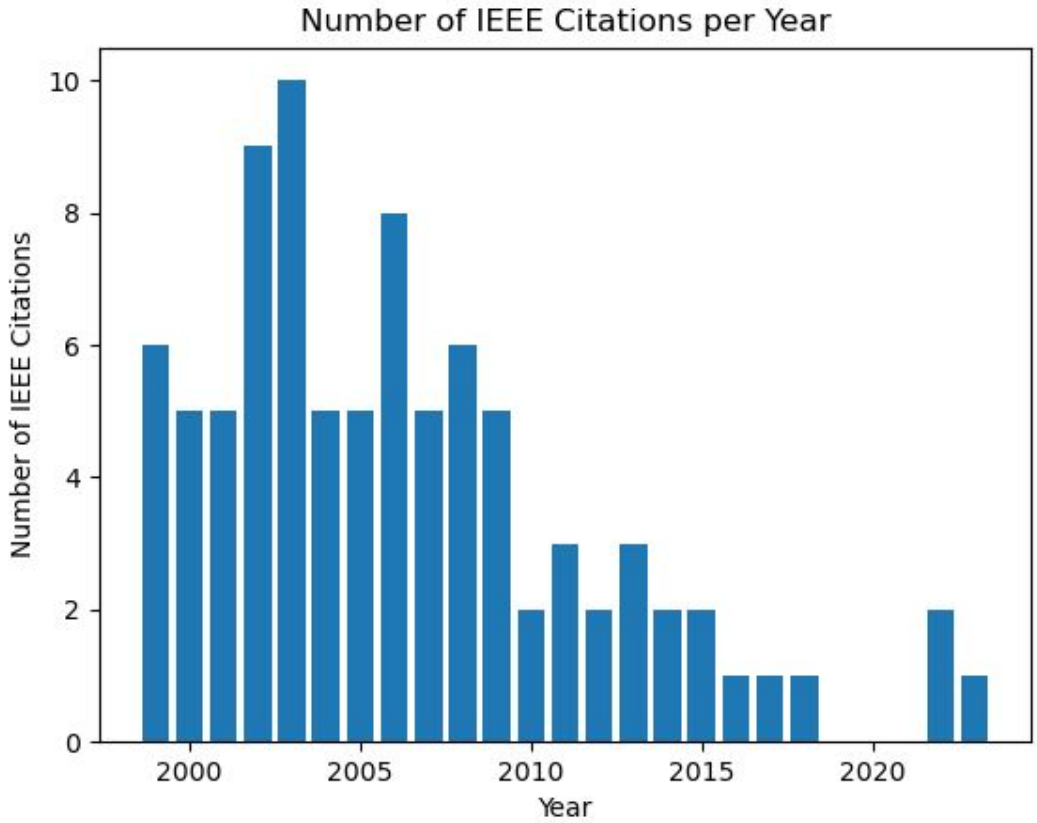
- HDL = Hardware Description Language
  - Verilog
  - VHDL
- HDL → Synthesis → Bitstream generation → Flash FPGA
  - Voila, programmable circuits
- JHDL makes circuit programming more familiar (OOP)

```
module full_add8(a, b, cin, s, cout);

    input [7:0] a, b;
    input       cin;

    output [7:0] s;
    output       cout;

    wire [6:0]   retenue;

    full_add add0 (a[0], b[0], cin,       s[0], retenue[0]);
    full_add add1 (a[1], b[1], retenue[0], s[1], retenue[1]);
    full_add add2 (a[2], b[2], retenue[1], s[2], retenue[2]);
    full_add add3 (a[3], b[3], retenue[2], s[3], retenue[3]);
    full_add add4 (a[4], b[4], retenue[3], s[4], retenue[4]);
    full_add add5 (a[5], b[5], retenue[4], s[5], retenue[5]);
    full_add add6 (a[6], b[6], retenue[5], s[6], retenue[6]);
    full_add add7 (a[7], b[7], retenue[6], s[7], cout      );

endmodule
```

```
import java.util.Scanner;

class Main {
  public static void main(String[] args) {
    Scanner in = new Scanner(System.in);

    System.out.print("How old are you?: ");
    int age = in.nextInt();

    if (age < 16) {
      System.out.println("Sorry, you are not
      quite old enough to drive!");
    }
    else {
      System.out.println("Yeah! Happy driving!");
    }
  }
}
```

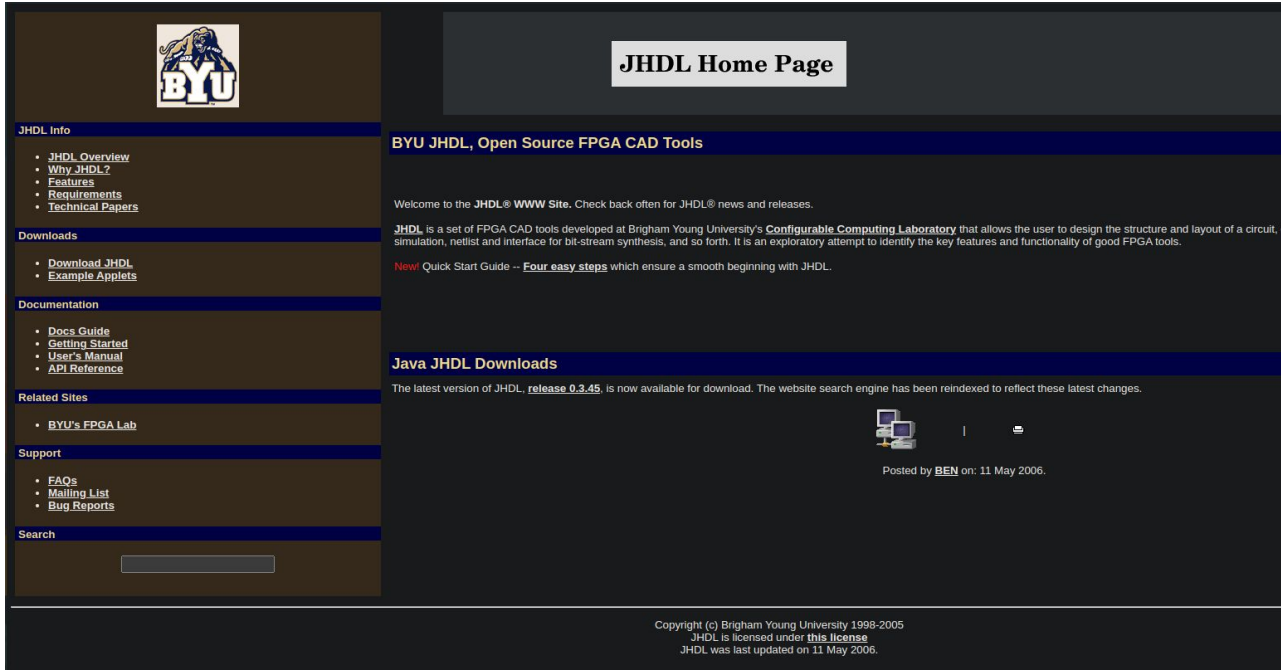

```
import byucc.jhdl.base.*;
import byucc.jhdl.Logic.*;
public class NBitAdder extends Logic {
  public static CellInterface[] cell_interface = {
    in("a", "WIDTH"),
    in("b", "WIDTH"),
    out("sum", "WIDTH"),
    param("WIDTH", INTEGER)
  };

  public NBitAdder(Node parent, Wire a, Wire b, Wire sum) {
    super(parent);
    int width = a.getWidth();
    bind("WIDTH", width);
    connect("a", a);
    connect("b", b);
    connect("sum", sum);

    Wire carries = wire(width);  // Here are the intermediate carry wires.

    for (int i=0; i < width; i++) {
      if (i==0)
        new FullAdder(this, a.gw(i), b.gw(i), gnd(), sum.gw(i), carries.gw(0));
      else
        new FullAdder(this, a.gw(i), b.gw(i), carries.gw(i-1), sum.gw(i), carries.gw(i));
    }
  }
}
```

JHDL is an old project…



JHDL is an old project…

But the website is still live

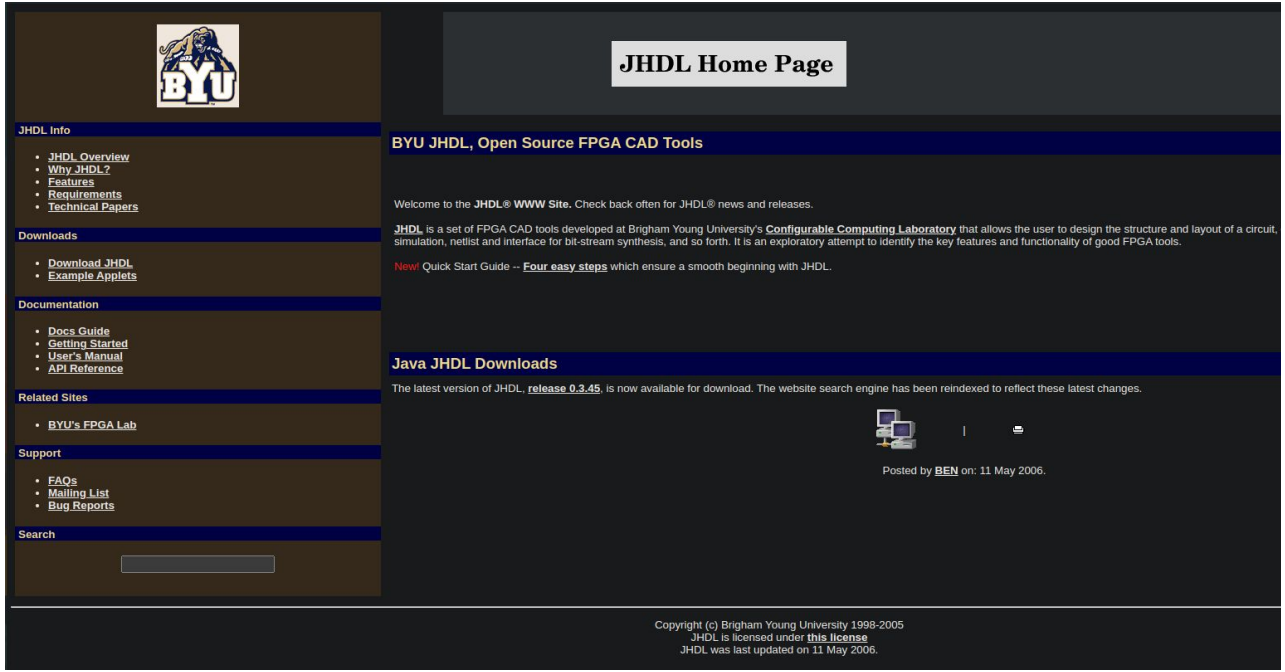


JHDL is an old project…

But the website is still live

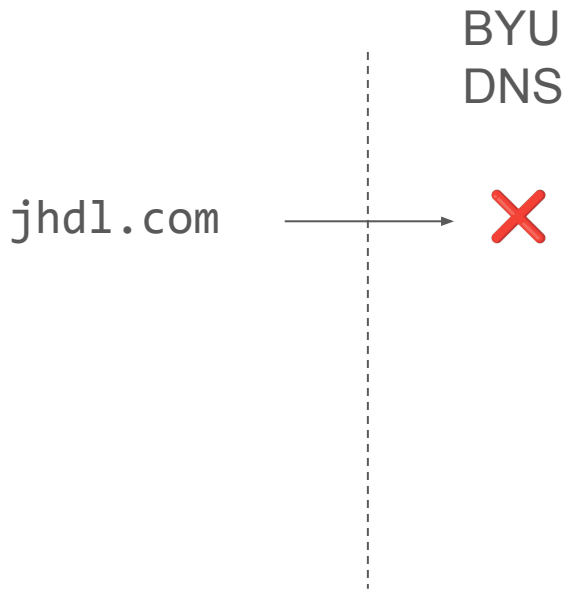
Old websites can be targets for malicious internet traffic due to outdated security designs

We want to determine who is trying to access the site and whether or not they may pose a threat



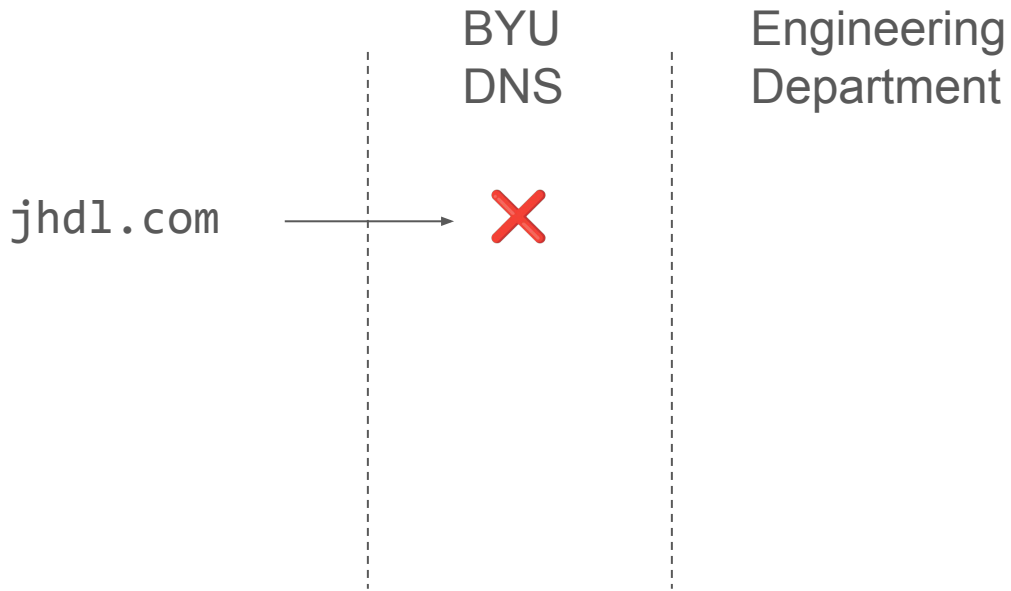
What are you looking for?

The JHDL site isn't actually accessible through `jhdl.com`



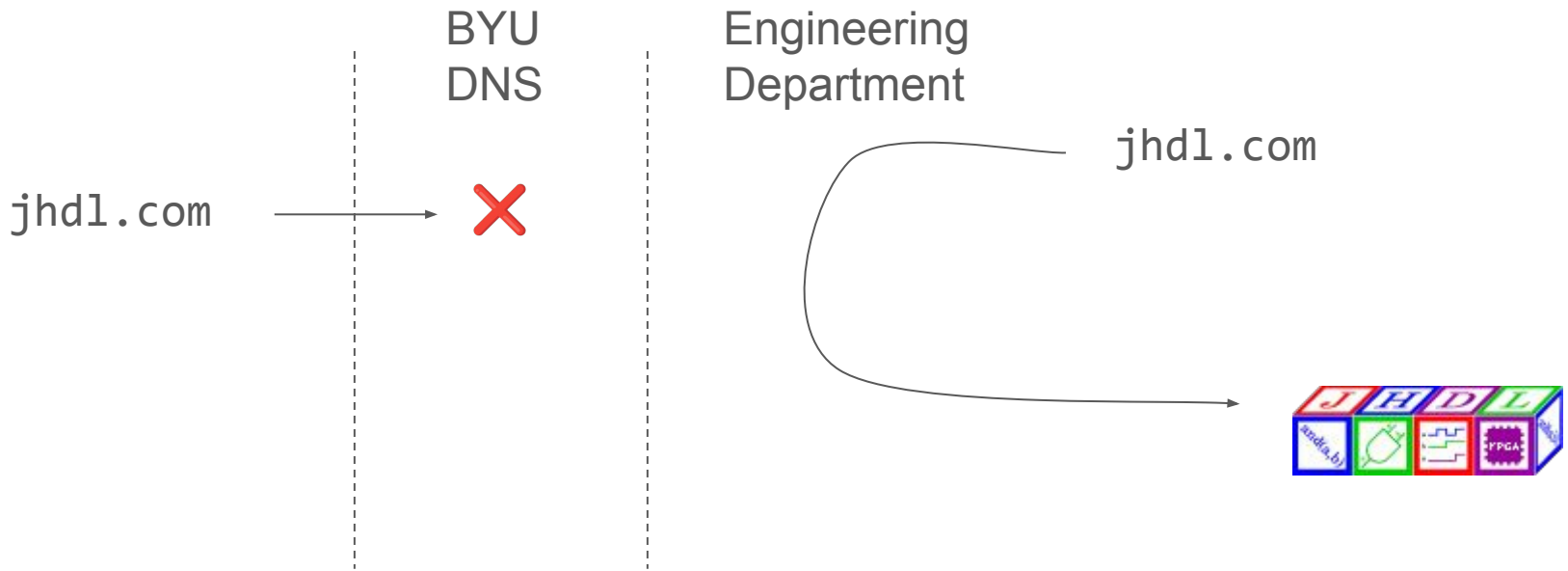
What are you looking for?

The JHDL site isn't actually accessible through `jhdl.com`



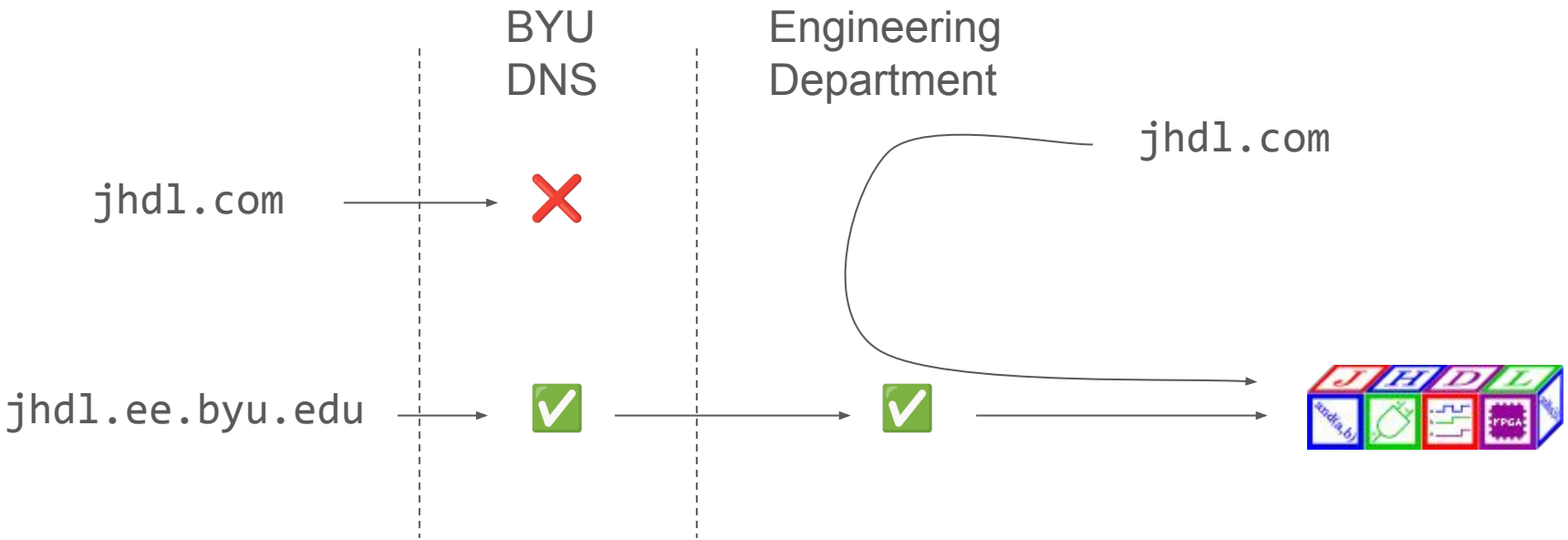
What are you looking for?

The JHDL site isn't actually accessible through `jhdl.com`



What are you looking for?

The JHDL site isn't actually accessible through `jhdl.com`



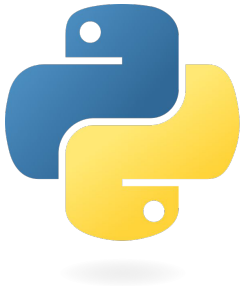
Different Behavior

- Only 47 hosts queried `*.jhdl.*` sites and `jhdl.ee.byu.edu`
- Find differences between these sites and sites that never looked for the real website
- Do they have different intentions?

Methodology

Methodology

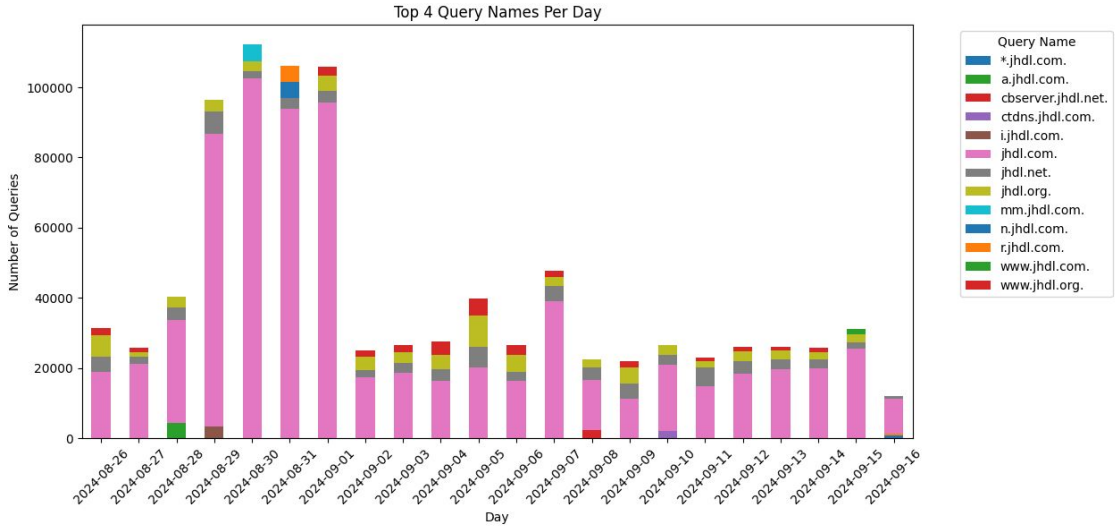
- Focused on entries where QNAME contains “jhdl”
  - MaryLou (ORC supercomputer)
- Python/Jupyter
  - Incremental, modifiable tests
- Pandas
  - Dataframes
- MaxMind Geo DB
  - Geolocation
- AbuseIPDB
  - Determine domain trustworthiness



Analysis

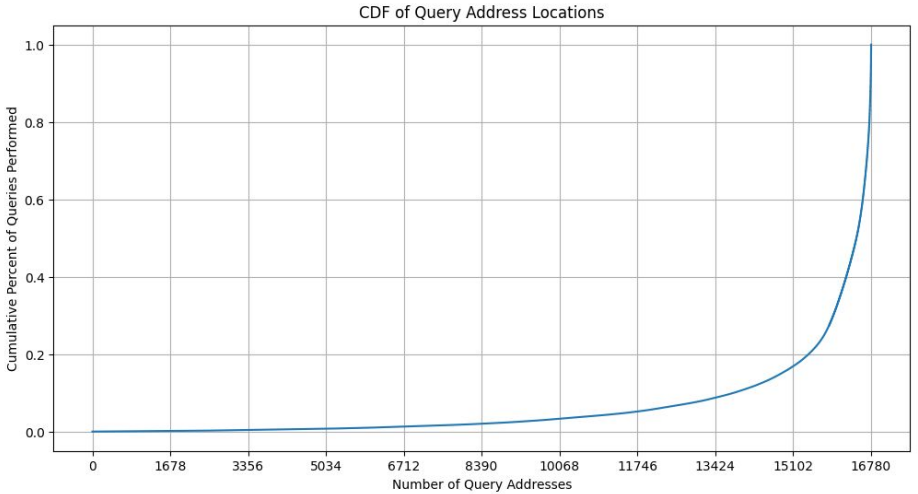
Volume of Queries

- Interesting subdomains
  - cbserver
  - ctdns
  - Single characters
- `nslookup` yielded `SERVFAIL`
- What is the big spike **Aug 29 - Aug 31?**
  - New Student Orientation?
  - Nothing definitive



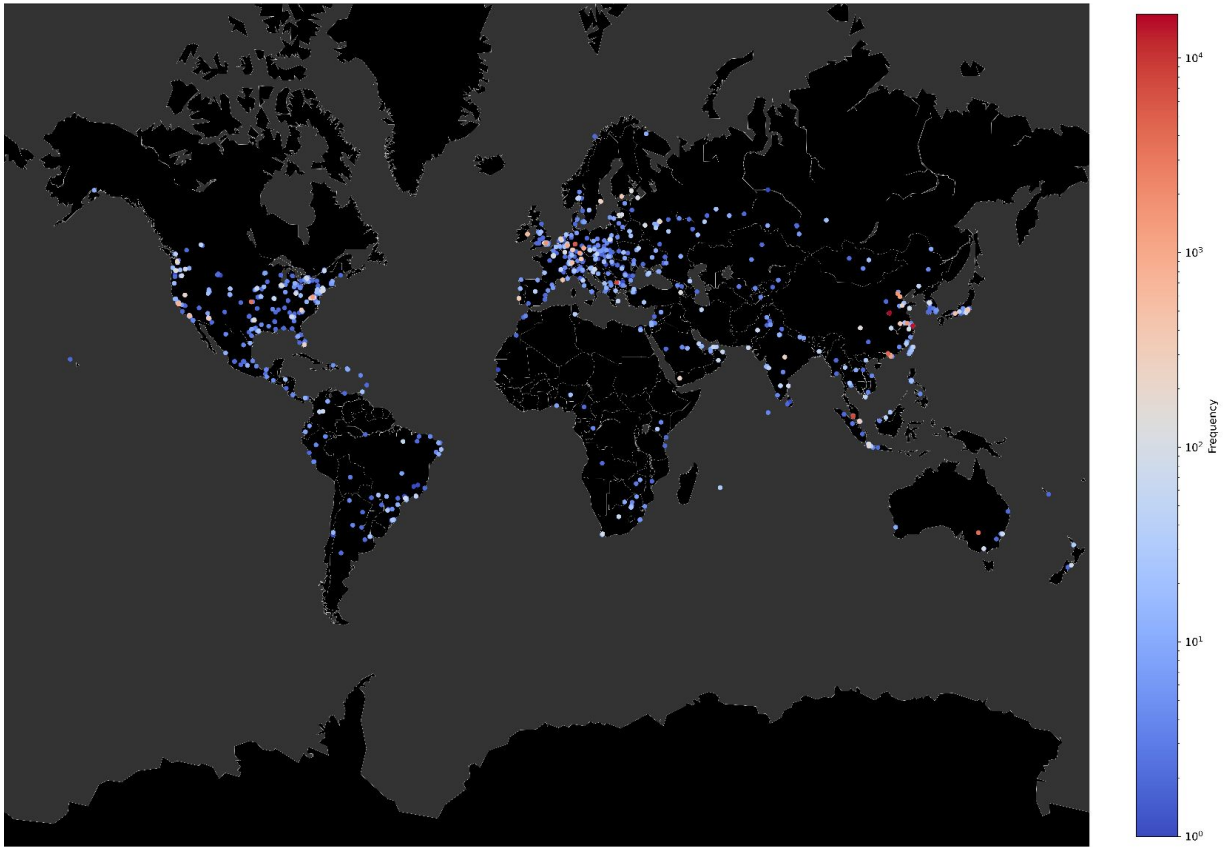
Autonomous Systems

- Small portion of ASes sent most of the queries
- 10,000+ queries vs 1,000+ queries



Geolocation

- MaxMind’s GeoLite DB for localization
- Worldwide reach
- Specific subset of countries with abnormally high frequency



Geolocation

Table 2: Top 7 Autonomous Systems and Countries that query JHDL domains

AS Name	Country	ASN	≥ # of Queries
UNICOM-SHFT-IDC, Shanghai	China	140979	44838
TENCENT-NET-AP	China	45090	11908
CHINANET BACKBONE	China	4134	9457
UNICOM-SHFT-IDC, Guangdong	China	135061	9297
TELEPOINT, BG	Belgium	31083	7388
TTSSB-MY TM TECHNOLOGY SERVICES	Malaysia	4788	7160
DE-FIRSTCOLO	Germany	44066	5470

Geolocation

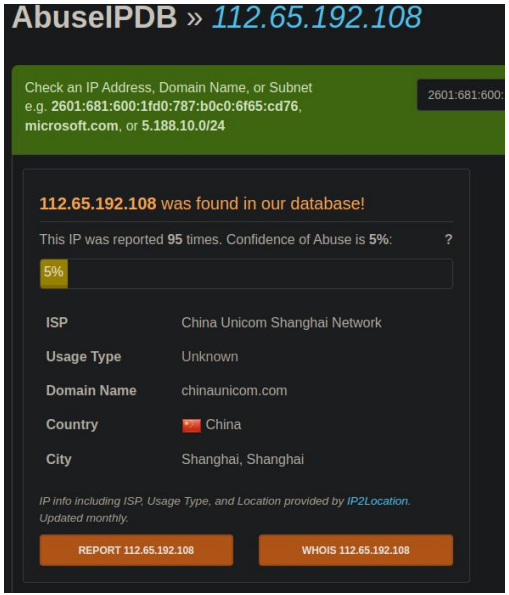
Table 2: Top 7 Autonomous Systems and Countries that query JHDL domains

AS Name	Country	ASN	≥ # of Queries
UNICOM-SHFT-IDC, Shanghai	China	140979	44838
TENCENT-NET-AP	China	45090	11908
CHINANET BACKBONE	China	4134	9457
UNICOM-SHFT-IDC, Guangdong	China	135061	9297
TELEPOINT, BG	Belgium	31083	7388
TTSSB-MY TM TECHNOLOGY SERVICES	Malaysia	4788	7160
DE-FIRSTCOLO	Germany	44066	5470

- FPGA/ASIC manufacturing
- ASIC fabrication research
- Web crawlers
- ???

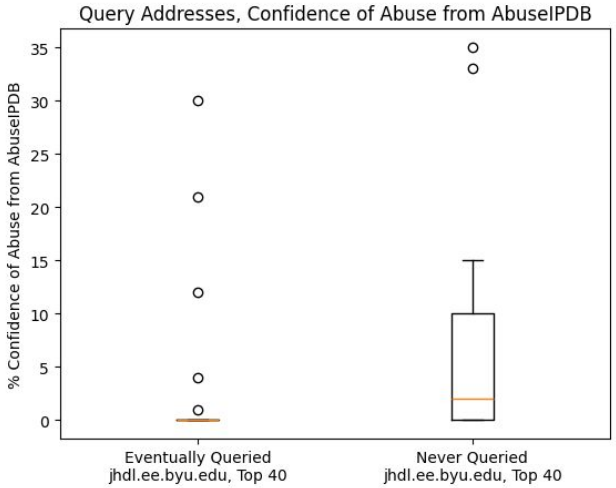
AbuseIPDB

- Crowd-sourced database of IP addresses that perform suspicious activities
- % confidence that a site is a malicious actor



AbuseIPDB - Density

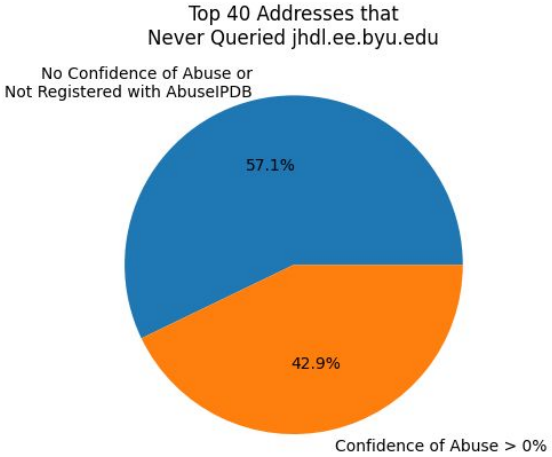
- Crowd-sourced database of IP addresses that perform suspicious activities
- % confidence that a site is a malicious actor
- Top 40 IP addresses of our two groups
- Much greater density of possible threats from the group that never actually queried the real JHDL site

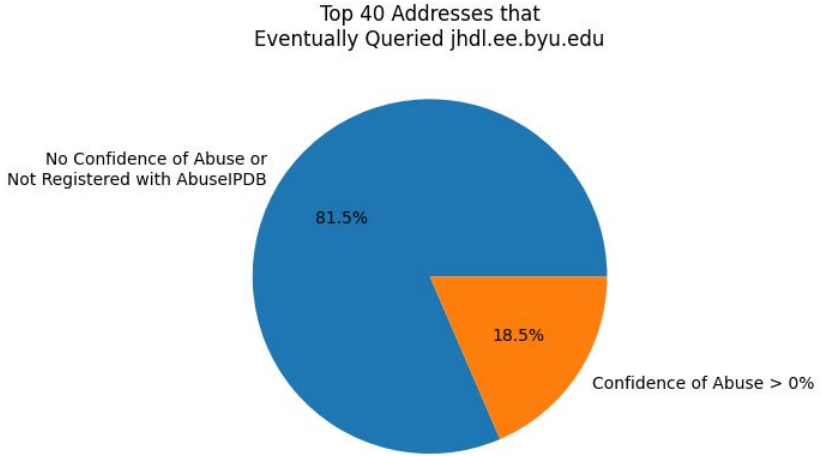


AbuseIPDB - Count

% of sites that are potentially malicious:

Never query > Eventually query





AbuseIPDB - Summary

Higher count

+

higher probability density

→

noticeable trend in malicious internet behavior

Table 3: Top 40 Hosts Visiting JHDL Sites Excluding `jhdl.ee.byu.edu` with non-zero probability of abuse

Query Address	% of Abuse
84.200.x.x	35
84.200.x.x	33
2001:1608:x:x::x:x	15
2001:1608:x:x::x:x	13
113.96.x.x	11
210.22.x.x	10
198.142.x.x	10
198.142.x.x	10
223.166.x.x	2
58.251.x.x	2
106.54.x.x	2
122.51.x.x	2
101.43.x.x	2
81.71.x.x	2
101.91.x.x	2
198.142.x.x	2

Table 4: Top 40 Hosts Visiting JHDL Sites Including `jhdl.ee.byu.edu` with non-zero probability of abuse

Query Address	% of Abuse
66.249.x.x	30
172.253.x.x	21
95.217.x.x	12
112.65.x.x	4
217.195.x.x	1

Conclusion

Conclusion

- Old sites need to be carefully monitored as potential targets for suspicious network traffic
    - Provide new lenses to characterize host trustworthiness based on behavior

Thank you!

Any questions?