Analyzing Third Party Service Dependencies in Modern Web Services: Have We Learned from the Mirai-Dyn Incident?

Aqsa Kashaf, Vyas Sekar, Yuvraj Agarwal

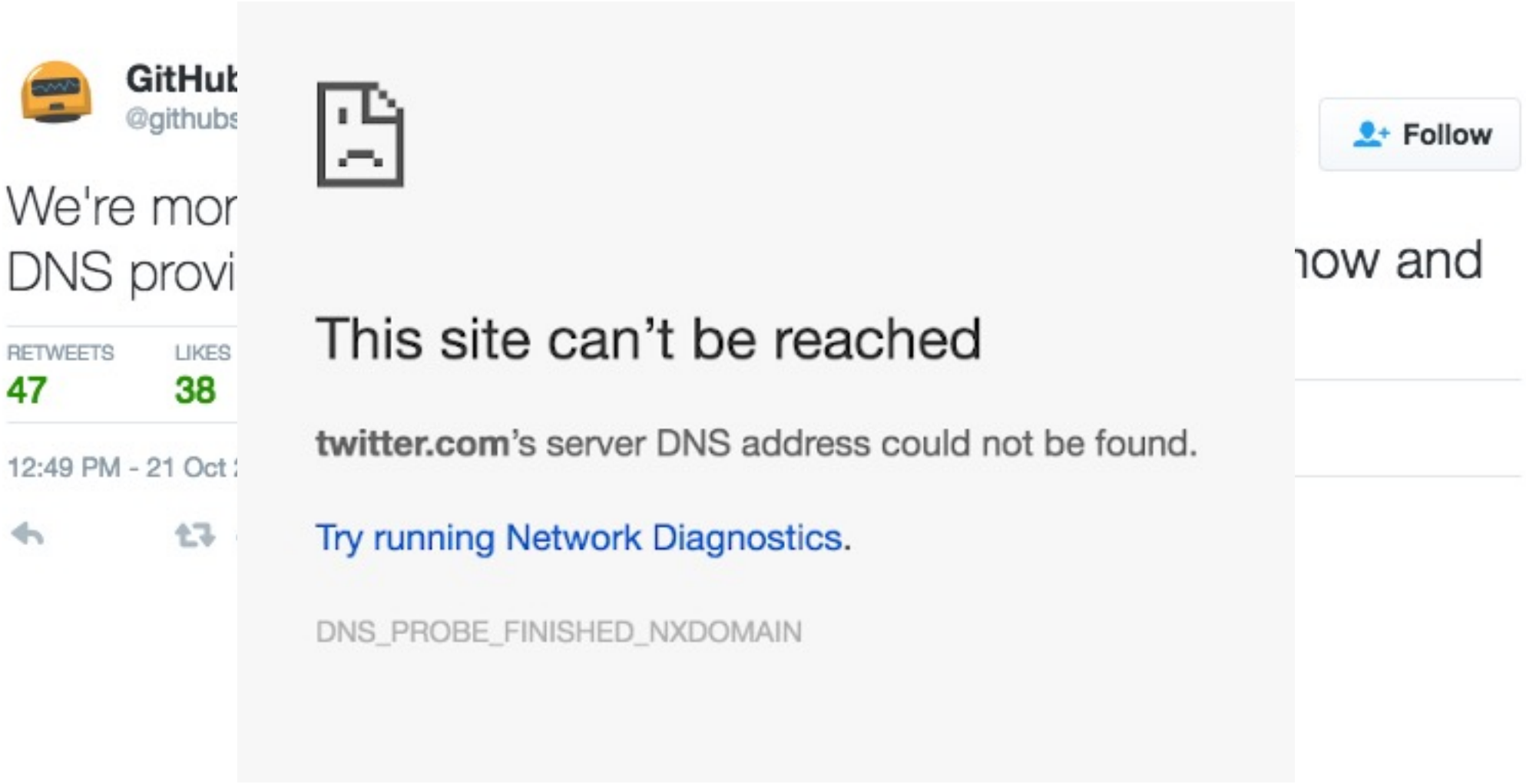
Carnegie Mellon University

Carnegie Mellon University CyLab Security and Privacy Institute

Synergy Labs

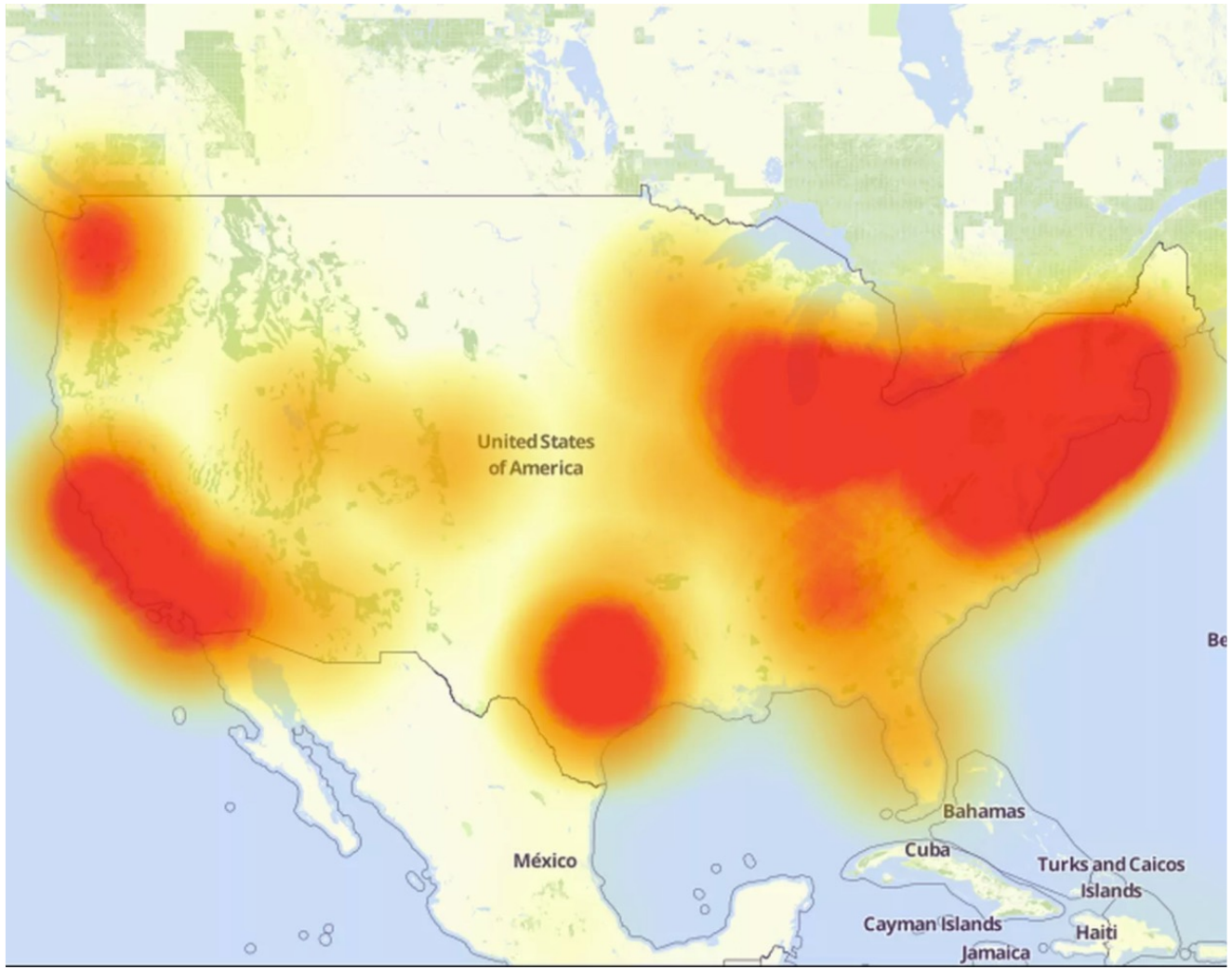
Carnegie Mellon University

Mirai-Dyn Attack 2016



Mirai-Dyn Attack 2016

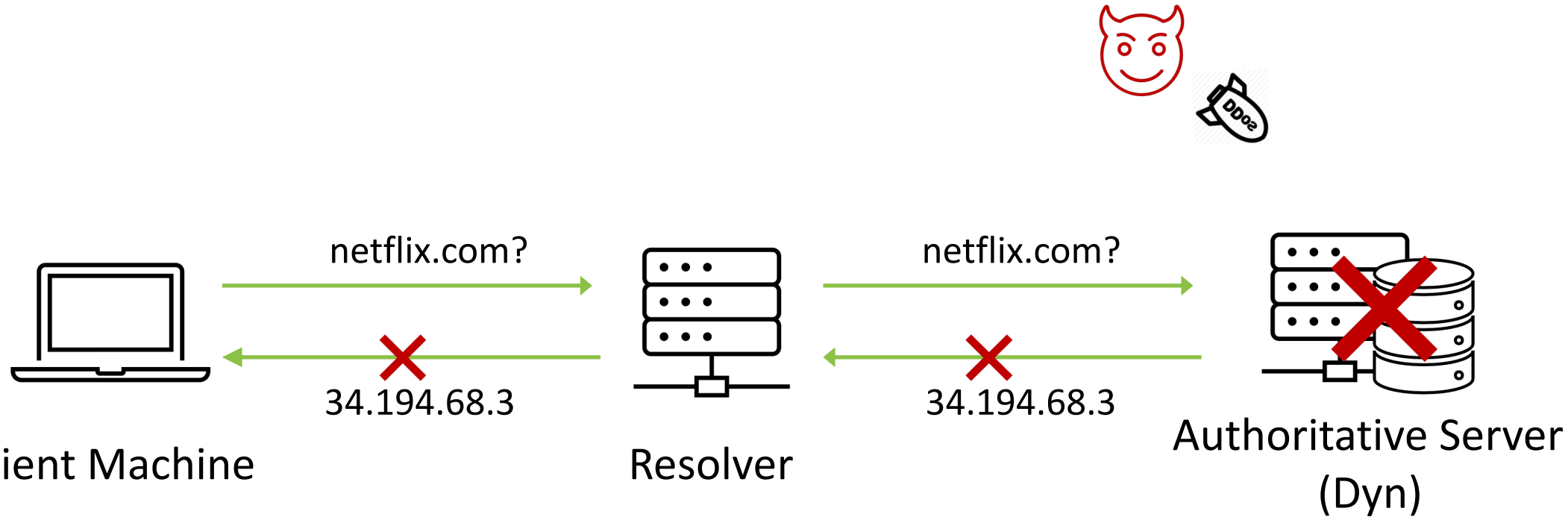
- 178,000 domains affected in total
- Tens of millions of users affected



Mirai-Dyn Attack 2016

How was it possible to take all of these websites down?

Mirai-Dyn Attack 2016



Insight: Many websites relied on the **same** 3rd Party DNS provider (Dyn)

Motivating Questions for Our Work

- How prevalent are third party dependencies?

  Methodology: Analysis on Alexa Top 100K websites

- Are there any indirect dependencies between websites and third-party providers?

  Methodology: Analysis on inter-service dependencies

- How did the world change after the Dyn Incident?

  Methodology: Comparison analysis on Alexa Top 100K sites in 2016 vs. 2020

Outline

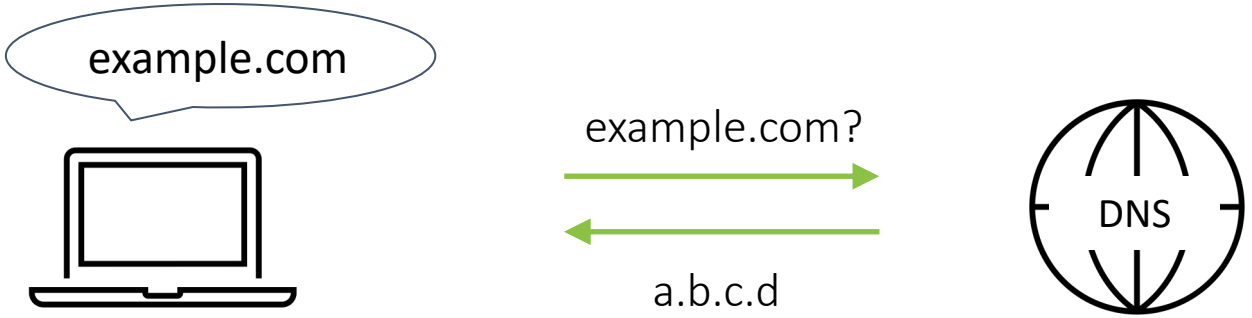
- Measurement Methodology
- Findings
- Recommendations
- Limitations
- Conclusion

Methodology: What services to measure?

Life Cycle of a Web Request

- Domain Name System (DNS)

  For example, AWS DNS, Dyn.

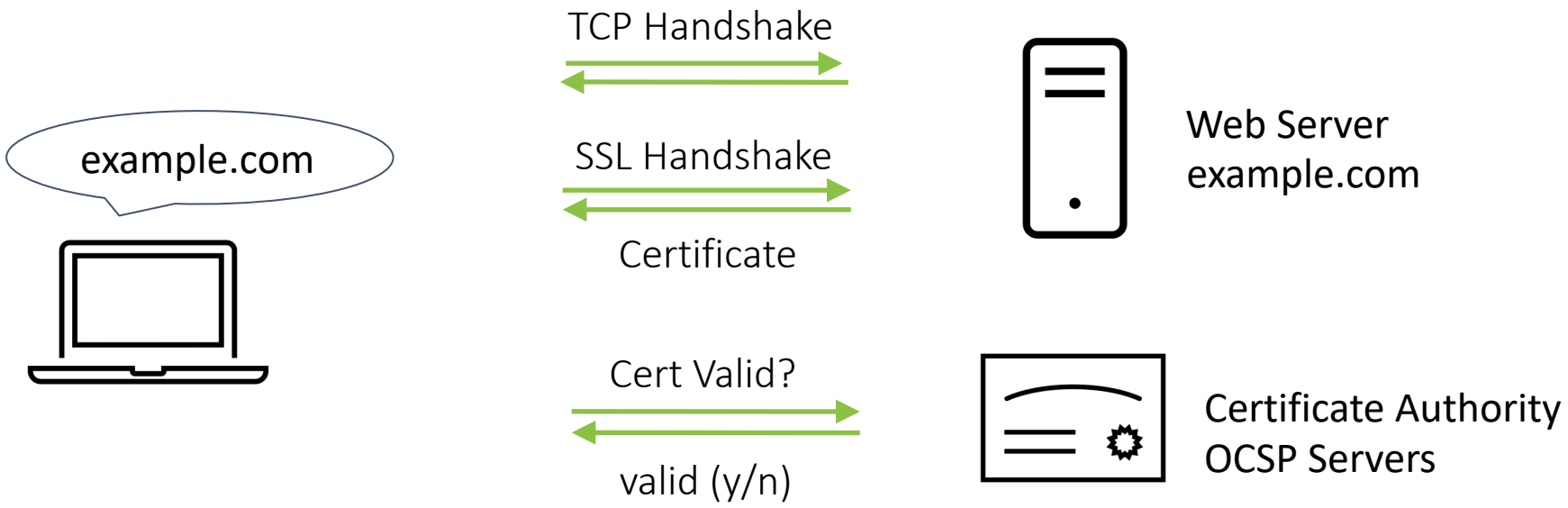


Methodology: What services to measure?

Life Cycle of a Web Request

- Domain Name System (DNS)
- Certificate Validation by CA

For example, DigiCert, Let's Encrypt.

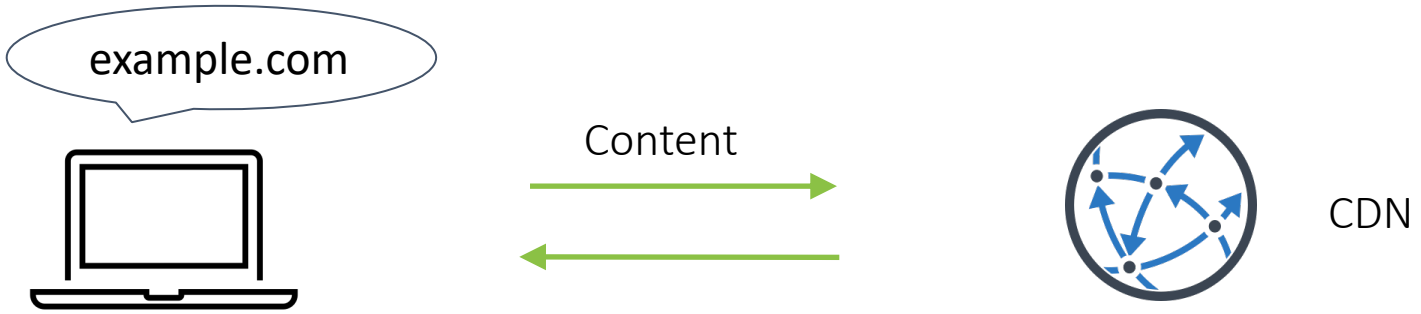


Methodology: What services to measure?

Life Cycle of a Web Request

- Domain Name System (DNS)
- Certificate Validation by CA
- Content Delivery Network (CDN)

For example, Akamai, CloudFlare



Methodology: What features to measure?

- Third Party Dependency
- Indirect Dependency
- Critical Dependency
  - No Redundancy in DNS and CDN provisioning
  - No OCSP stapling in certificate validation

Indirect Dependency

Measuring 3rd party DNS dependency

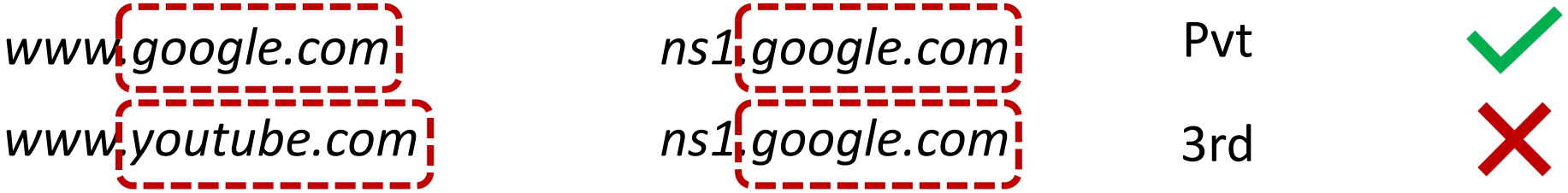
- live.com *.azure-dns.com
  *.o365filtering.com

Q1. Are these third party or private?

Q2. Do these belong to the same entity?

Identifying 3rd party DNS dependency: Prior efforts are error prone

- Using SLD + TLD Matching



- Using SOA Records Matching

Identifying 3rd party DNS dependency: Our Approach

For all *(website, NS)* pairs:

- SLD + TLD match
- NS ∈ Subject Alternate Names (SAN) list

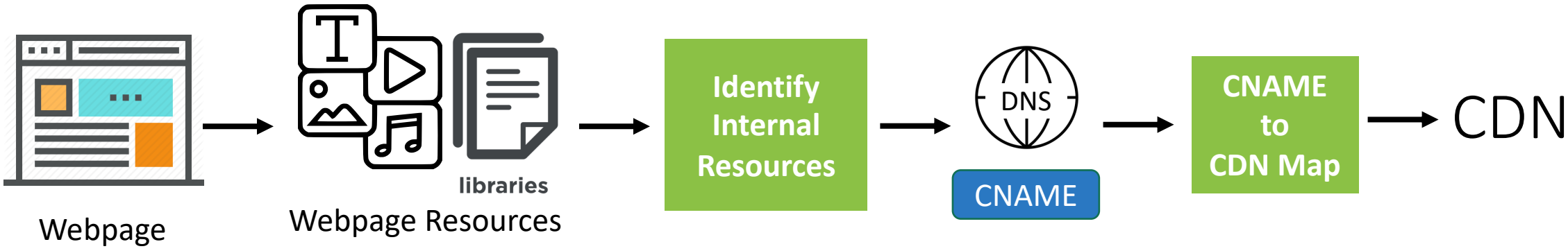
} Private

- SOA do no match
- $Concentration(NS) > 50$

} Third

We identify 10K Third Party DNS Providers

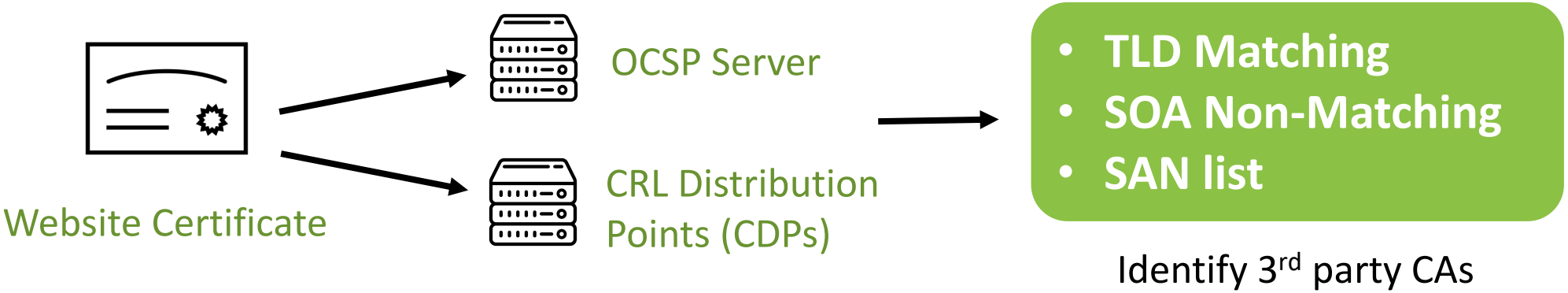
Measuring 3rd Party CDN Dependency



reddit.com a.thumbs.redditmedia.com reddit.map.fastly.net Fastly

- Use TLD, SOA, SAN of embedded links to identify internal resources
- Use TLD, SOA, SAN of CNAMES used by CDNs to identify 3rd party CDNs
- We identify 86 Third party CDNs

Measuring 3rd party CA dependency



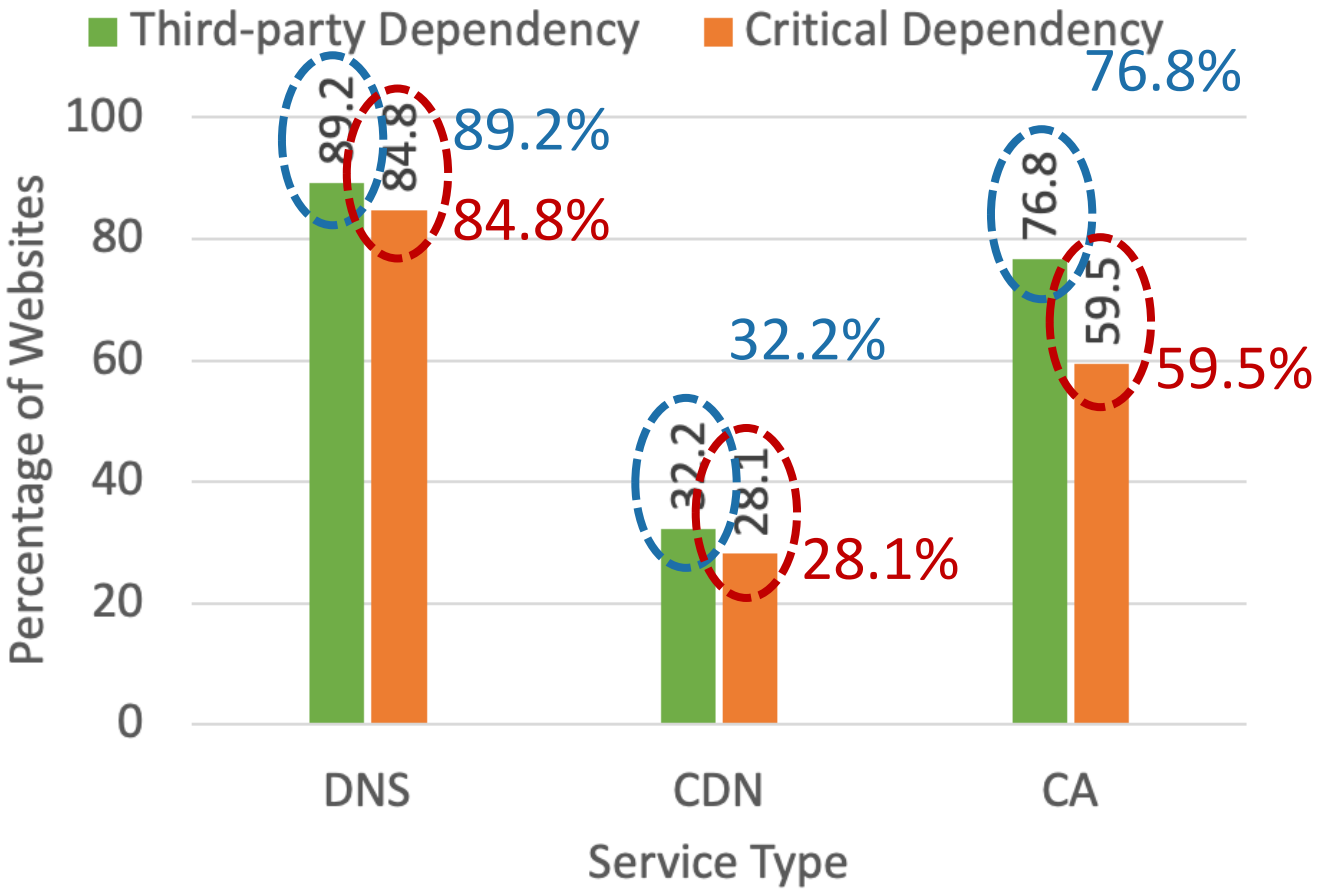
- We identify 59 third party CAs

Outline

- Measurement Methodology
- Findings
- Recommendations
- Limitations
- Conclusion

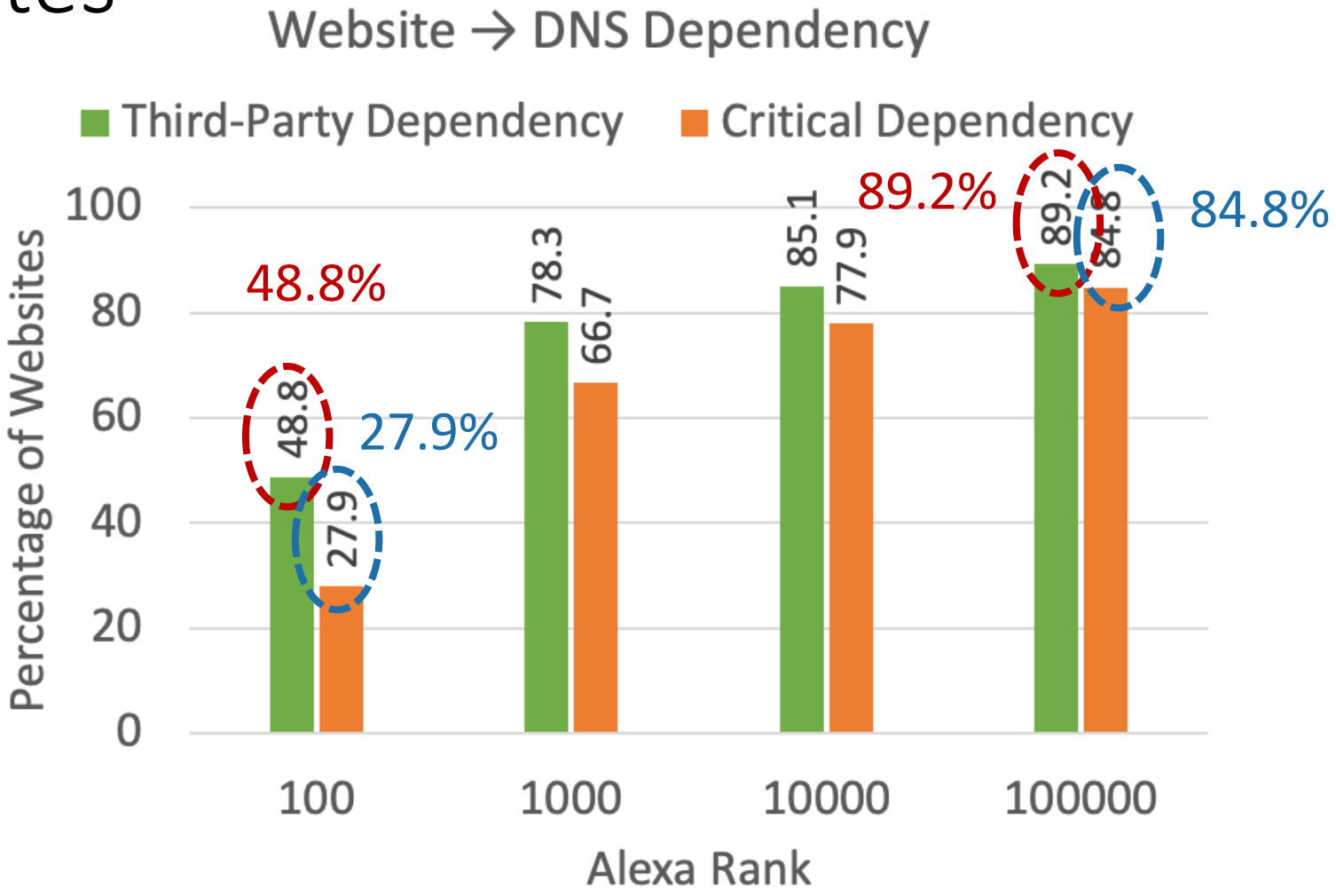
Q1: How prevalent are third-party dependencies?

Third-Party Dependencies are Highly Prevalent



89% of the top-100K websites critically depend on third-party DNS, CDN, or CA providers.

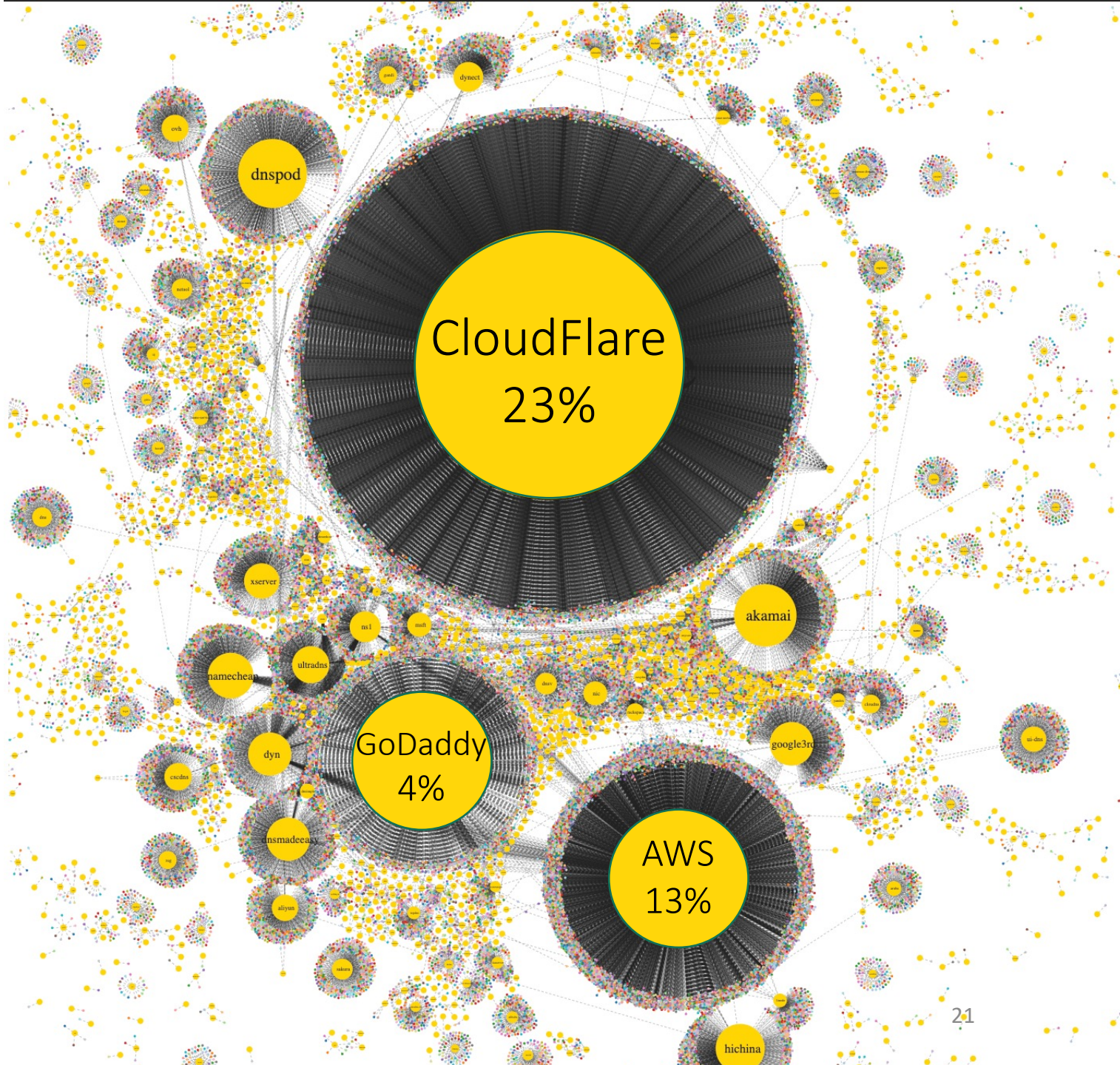
Third-Party Dependencies Higher for Less Popular Websites



Popular websites care more about availability.

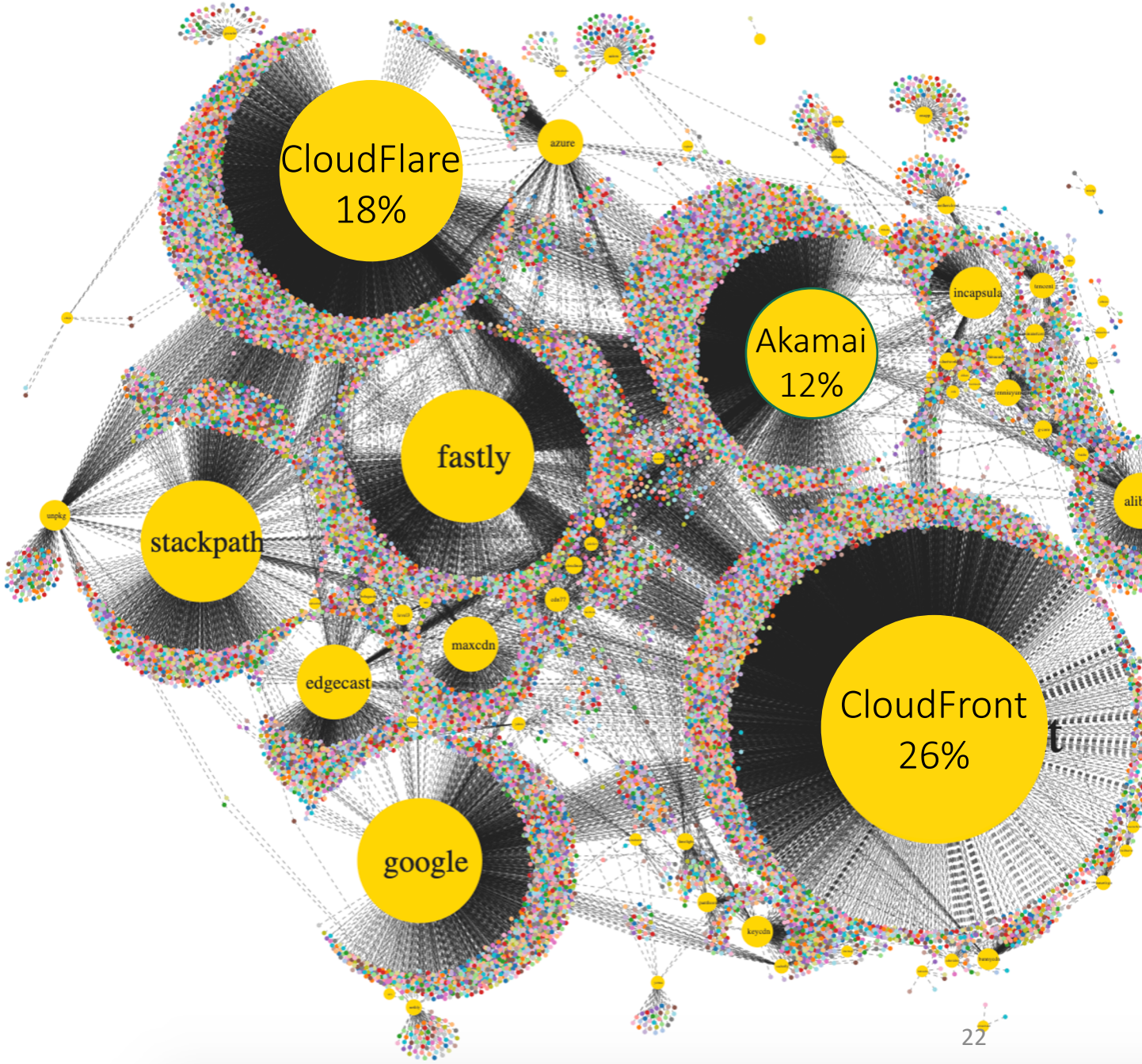
Concentration of DNS Providers

3 (out of 10K) DNS providers critically serve ~40% of the top-100K websites



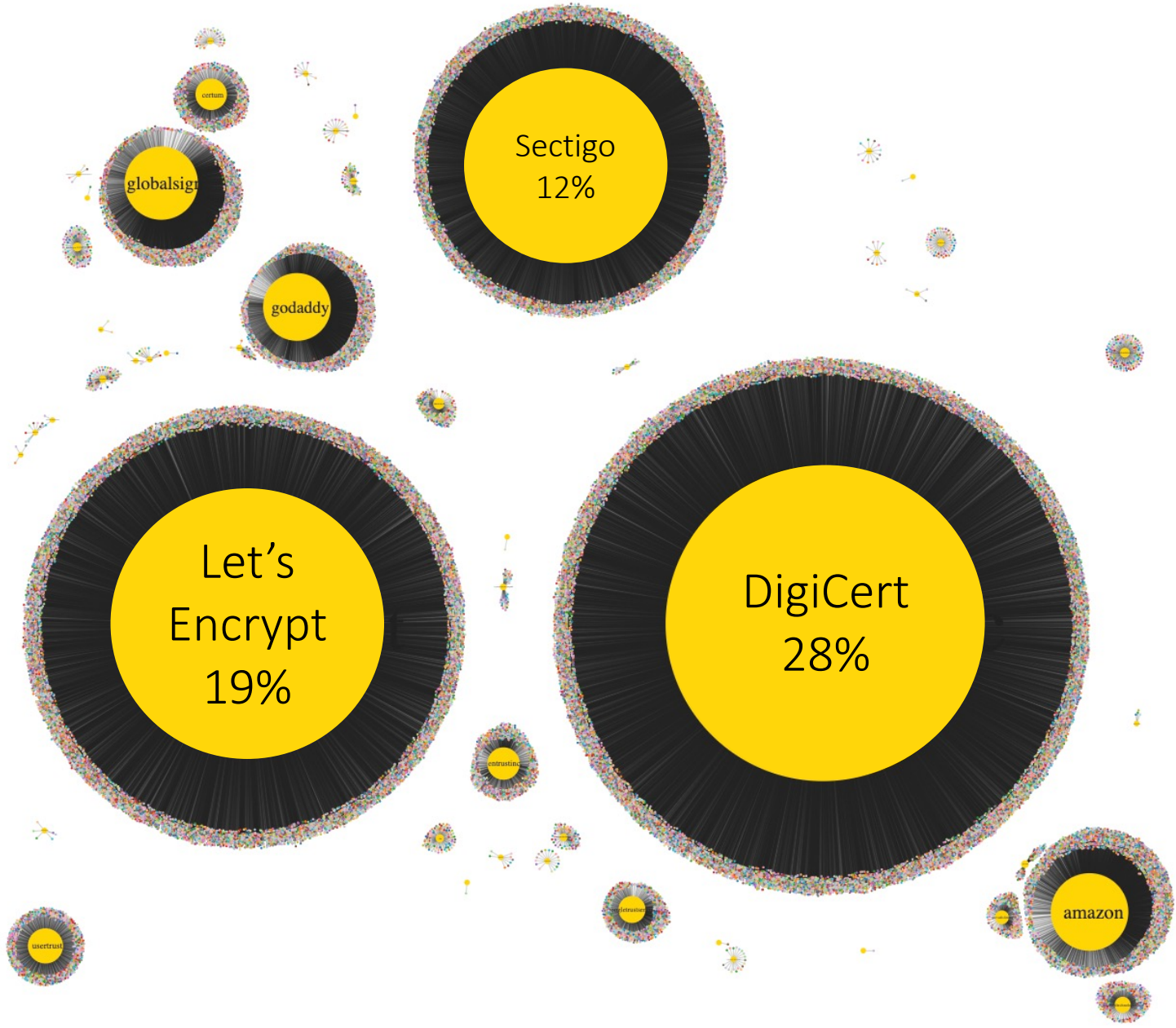
Concentration of CDN Providers

3 (out of 86) CDN providers critically serve ~60% of the top-100K websites using CDN



Concentration of CA Providers

3 (out of 59) CAs critically serve ~60% of the top-100K websites that support HTTPS



Takeaway

- Third party critical dependencies are highly prevalent.
- Third party services are highly concentrated.

Implications:

- 89% of the websites are vulnerable to Dyn like incidents
- A single third-party service provider can affect ~25% of the top 100K websites

Q2: Are there any indirect dependencies between websites and their third-party providers?

Inter-Service Third-Party Dependency

48%	36%	36%
CA → DNS	CA → CDN	CDN → DNS

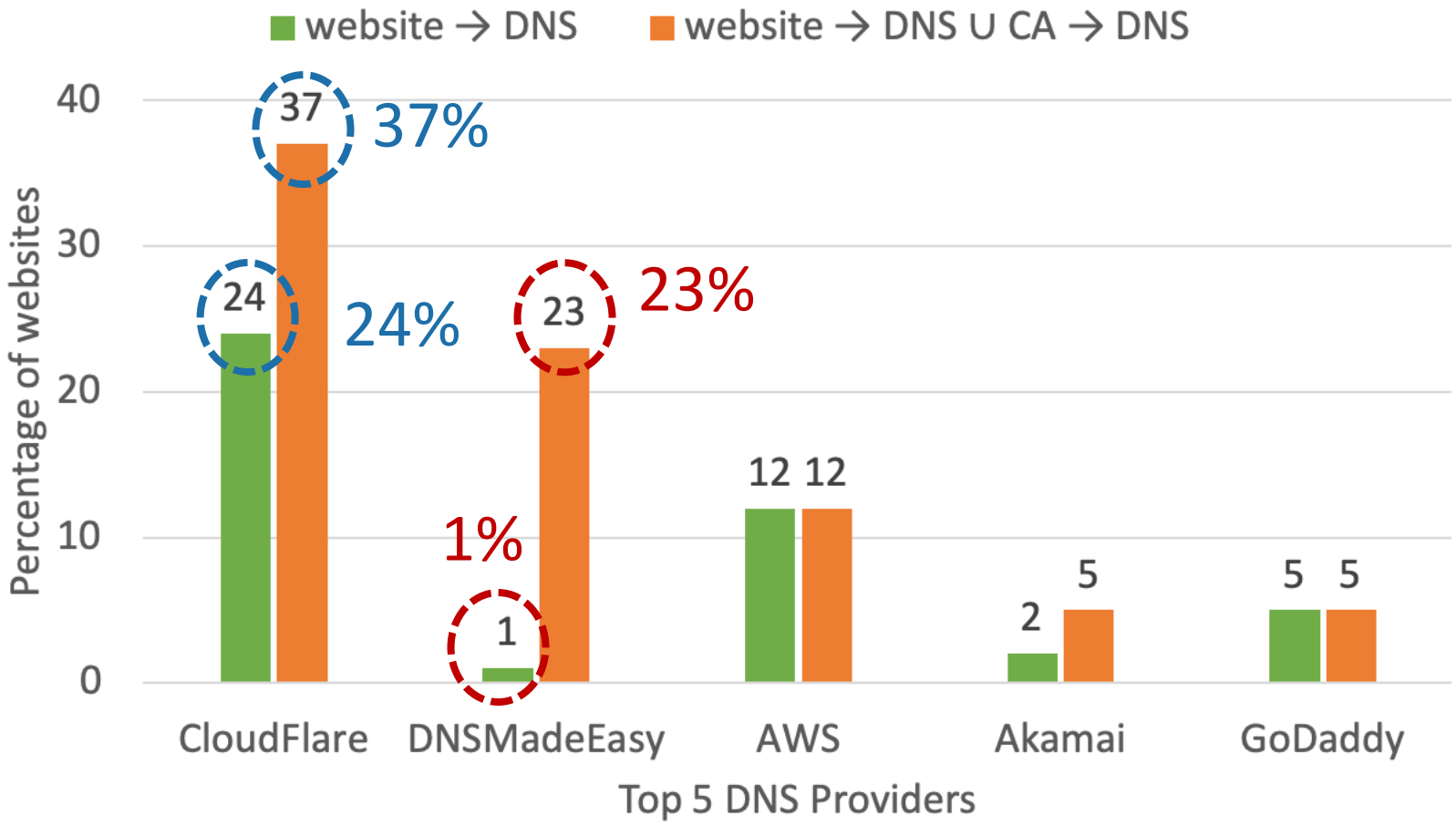
Third-party dependencies are also prevalent among service providers

Inter-Service Critical Dependencies

31%	36%	17%
CA → DNS	CA → CDN	CDN → DNS

Due to inter-service **critical dependencies**, websites have indirect dependencies on service providers

Indirect Dependencies Amplify Concentration



Indirect Dependencies further amplify provider concentration

Takeaway

- Third party inter-service critical dependencies are also widespread
- Inter-service critical dependencies amplify the concentration of service providers

Implications:

- Single points of failure on the internet are amplified by inter-service dependencies
- A single service provider can impact 37% of the top 100K websites.

Q3: How did the world change after the Dyn incident in 2016?

Critical Dependency of Websites (2016 to 2020)

+4.7%	0%	-0.2%
website → DNS	website → CDN	website → CA

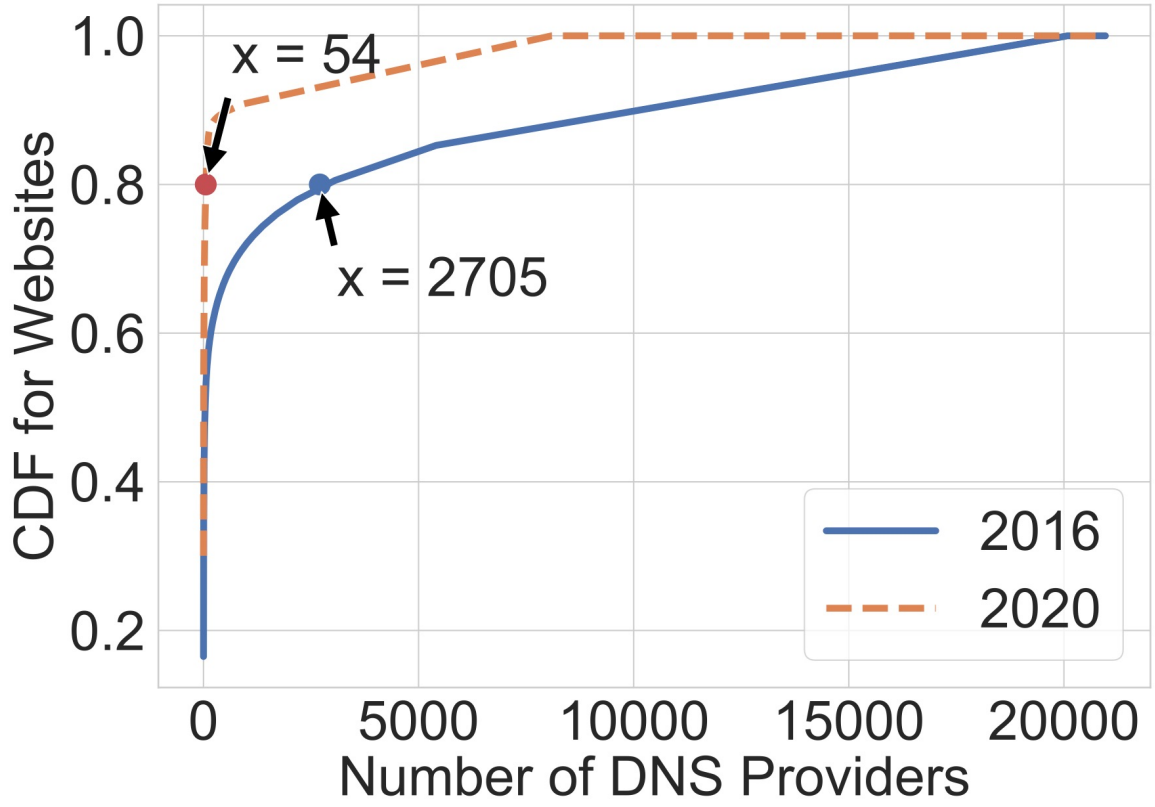
No improvement in the prevalence of third-party dependency. Critical dependency increased in DNS

Inter-Service Critical Dependency (2016 to 2020)

-8.6%	0%	-4.3%
CA → DNS	CA → CDN	CDN → DNS

Critical dependency decreased in service providers

Change in Concentration of DNS Providers



Single-points-of-failure got bigger in DNS and CA!

Takeaway

- No significant change in the prevalence of third-party critical dependencies in websites
- Inter-service critical dependencies on DNS decreased in 2020.
- Concentration of DNS and CA providers increased in 2020.

Implications:

- No increasing trend in redundancy.
- Single points of failure in the internet got bigger in 2020 vs. 2016

Outline

- Measurement Methodology
- Findings
- Recommendations
- Limitations
- Conclusion

Our Recommendations

Websites

- Redundancy when using third party providers
- Understand their indirect dependencies

Service Providers

- Support and encourage redundancy
- Be careful about their inter-service dependencies
- Be more transparent about attacks

Outline

- Measurement Methodology
- Findings
- Recommendations
- Limitations
- Conclusion

Limitations

- Measurements from a single vantage point
  - May miss region specific dependencies
- Analyze dependencies on landing pages only
  - May miss dependencies that manifest deeper
- Do not look at physical and network dependencies
  - For example, routing, hosting etc.

Conclusion

- DDoS attack on Dyn exposed the fragility of the Web due to dependencies
- Our work: Analyze third-party and inter-service dependencies
- Key Findings:
  - **Prevalence of third-party dependency:**
    89% of top 100K websites are critically dependent
    An attack on a single provider can take down ~30% of the websites
  - **Impact of indirect dependencies:**
    ~23X amplification in provider concentration
  - **Change after the Dyn Incident:**
    No significant change in website dependencies

Code: github.com/AqsaKashaf/Analyzing-Third-party-Dependencies.git