The 21st Privacy Enhancing Technologies Symposium

Domain Name Encryption Is Not Enough: Privacy Leakage via IP-based Website Fingerprinting

Stony Brook University

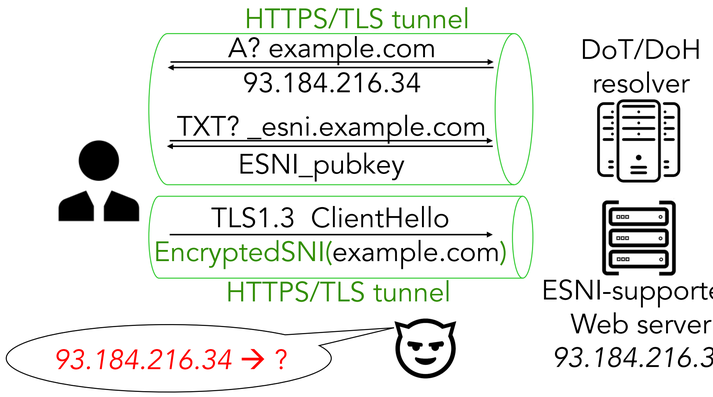
UMASS AMHERST

Nguyen Phong Hoang, Arian Akhavan Niaki, Phillipa Gill, Michalis Polychronakis

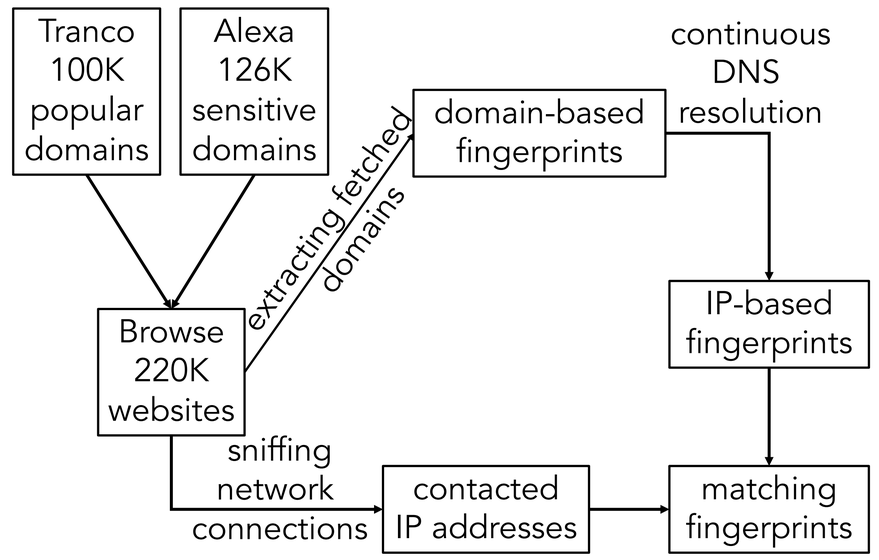
Research motivation

- ✓ Domain name encryption technologies (DoT, DoH, and ESNI) are increasingly adopted
- ✓ Destination servers' IP addresses are still visible

Domain name encryption => improved user privacy?



Experimental setup



Basic fingerprinting

```
{'twitter.com':
  primary: {'api.twitter.com',
            'abs.twimg.com', 'pbs.twimg.com',
            'www.google-analytics.com'}}
{[1760832001, 1760832065,
1760832129, 1760832193];
  secondary:
    [385967085, 385968877, 1209359174,
    1760832002, 1760832066, 1760832130,
    1760832194, 2540008607, 2540030111,
    2540032159, 2540042399, 2899903342,
    2899904206, 2899904238, 2899904270,
    2899905006, 2899905038, …]}
```

Enhanced fingerprinting with connection bucketing

Enhanced domain-based fingerprint

```
0: {'twitter.com'},
  1: {'abs.twimg.com'},
  2: {'api.twitter.com', 'abs.twimg.com', 'pbs.twimg.com'},
  3: {'twitter.com', 'api.twitter.com', 'abs.twimg.com', 'www.google-analytics.com'}
```

Enhanced IP-based fingerprint

```
0: {1760832001, 1760832065, 1760832129, 1760832193};
  1: {1209359174, 2540008607, 2540030111, 2540032159, 2540042399, 3236277520, …};
  2: {385967085, 385968877, 1209359174, 1760832002, 1760832066, 1760832130,
      1760832194, 2540008607, 2540030111, 2540032159, 2540042399, 3089042157, …};
  3: {1209359174, 1760832001, 1760832002, 1760832065, 1760832066, 1760832129,
      1760832130, 1760832193, 1760832194, 2540008607, 2540030111, 2540032159,
      2540042399, 2899903342, 2899904206, 2899904238, 2899904270, 2899905006, …}
```

Clustered sequence of IPs from network trace, using K-means

```
0: {1760832065},
  1: {1209359174},
  2: {1760832002, 1209359174},
  3: {1760832065, 1760832002, 1209359174, 2899904270}
```

Fingerprinting results

Website type	Total domains	Primary domain	Basic fingerprinting	Connection bucketing
All websites crawled	208,191	107,455 (52%)	174,662 (84%)	189,527 (91%)
Popular websites	93,661	58,989 (63%)	86,147 (92%)	90,231 (96%)
Sensitive websites	120,293	51,538 (43%)	93,988 (78%)	104,983 (87%)
Sensitive and popular	5,763	3,072 (53%)	5,473 (95%)	5,687 (99%)

Conclusion: regardless of domain name encryption, network-level adversaries can still rely on destination IP addresses of contacted web servers for IP-based website fingerprinting to track users' browsing history at scale for the vast majority of websites.