

CARONTE: Detecting Location Leaks for Deanononymizing Tor Hidden Services



authors:

Srdjan Matic^{†‡},
Platon Kotzias[‡]
and Juan Caballero[‡]

[†] Università degli Studi di Milano
[‡] IMDEA Software Institute

“CARONTE: Detecting Location Leaks for
Deanonymizing Tor Hidden Services”



“CARONTE: Detecting Location Leaks for
Deanonymizing **Tor** Hidden Services”



“CARONTE: Detecting Location Leaks for
Deanonymizing **Tor Hidden Services**”





“CARONTE: Detecting **Location Leaks** for **Deanonymizing Tor Hidden Services**”



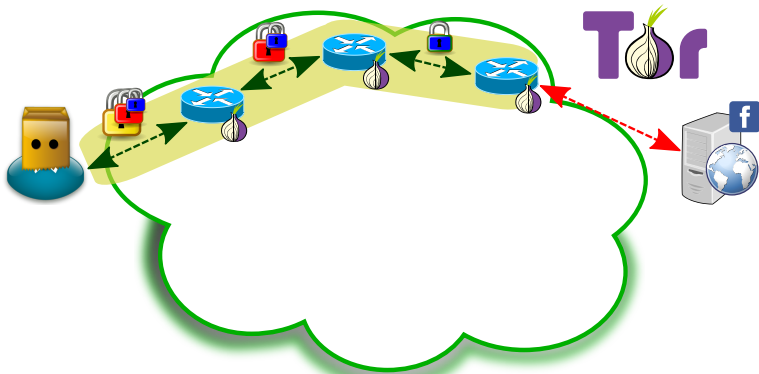
TOR and Hidden Services

- Web Site: <http://facebook.com/>



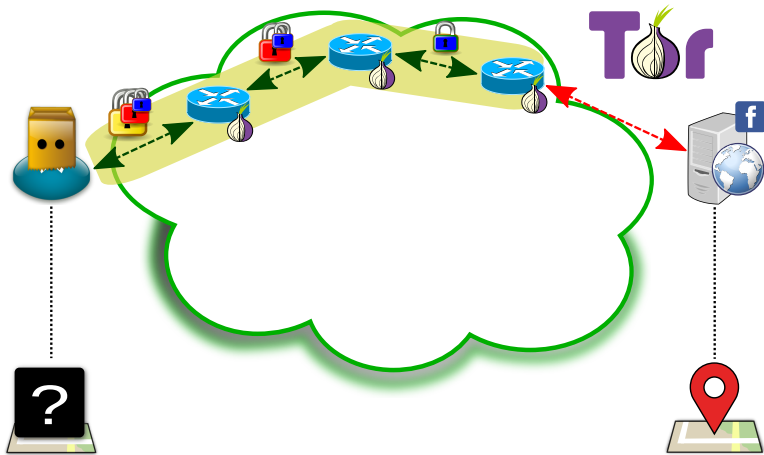
TOR and Hidden Services

- Web Site: <http://facebook.com/>



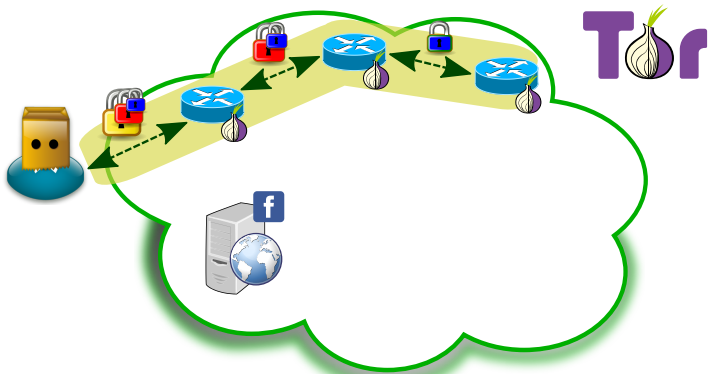
TOR and Hidden Services

- Web Site: <http://facebook.com/>



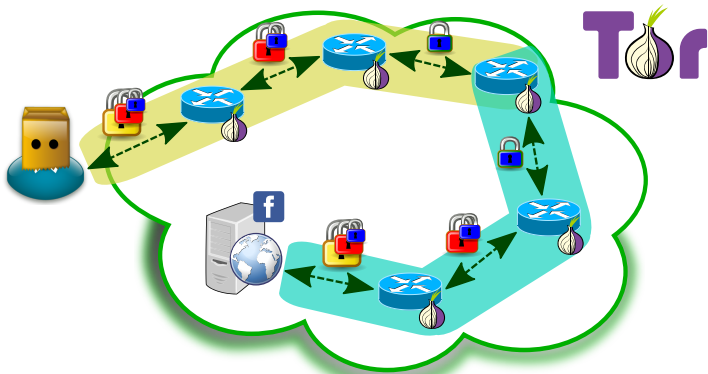
TOR and Hidden Services

- Web Site: <http://facebook.com/>
- Hidden Service : <http://facebookcorewwi.onion/>



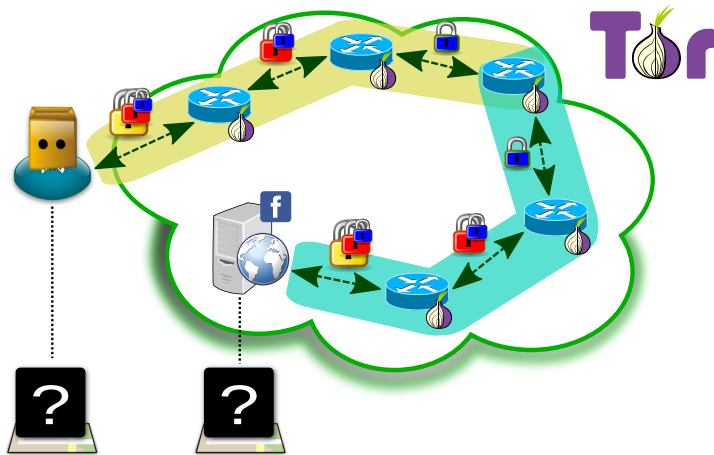
TOR and Hidden Services

- Web Site: <http://facebook.com/>
- Hidden Service : <http://facebookcorewwi.onion/>



TOR and Hidden Services

- Web Site: <http://facebook.com/>
- Hidden Service : <http://facebookcorewwi.onion/>



Tor

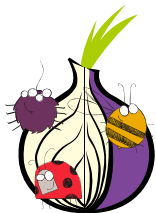


Hidden Services



Hidden Services





Øverlier and Syverson (IEEE S&P 2006)
Biryukov et al. (IEEE S&P 2013)

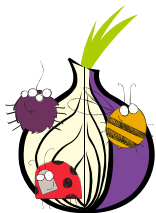


Murdoch (CCS 2006)
Zander and Murdoch (USENIX Security 2008)



Crenshaw (BlackHat DC 2011)

Related Work

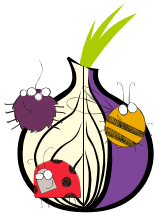


Murdoch (CCS 2006)
Zander and Murdoch (USENIX Security 2008)



Crenshaw (BlackHat DC 2011)

Related Work

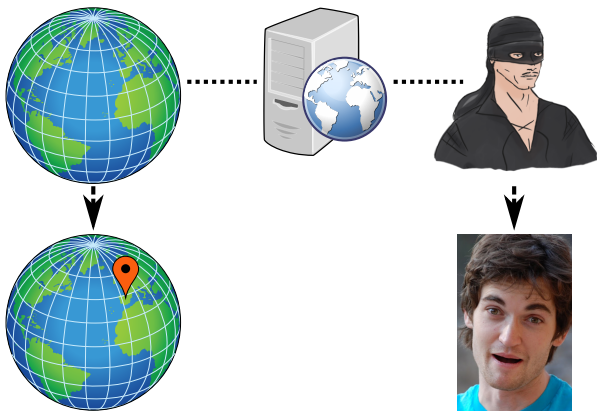


I2P 

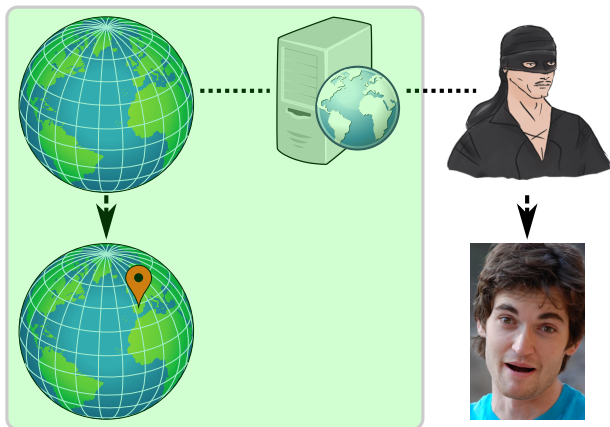
Location Leaks



Location Leaks



Location Leaks



How Do Location Leaks Happen?

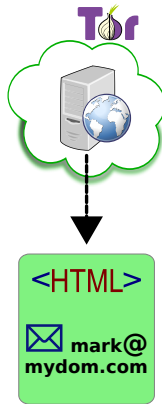
myd4xi.onion



1. **myd4xi.onion** is a HS

How Do Location Leaks Happen?

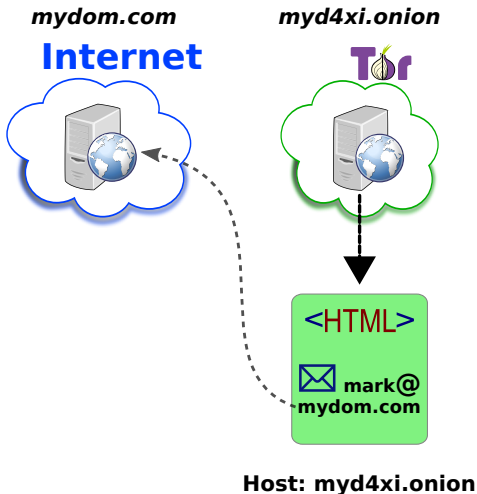
myd4xi.onion



1. **myd4xi.onion** is a HS
2. Request a resource

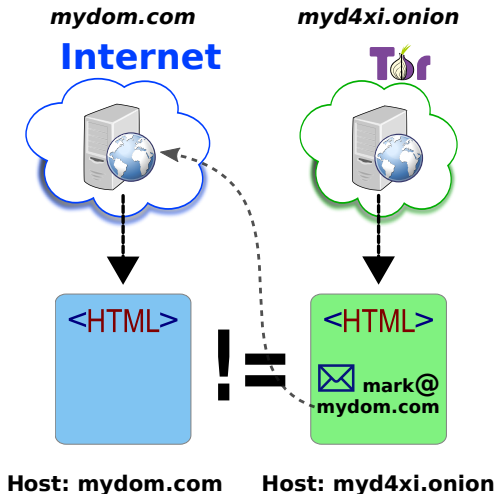
Host: myd4xi.onion

How Do Location Leaks Happen?



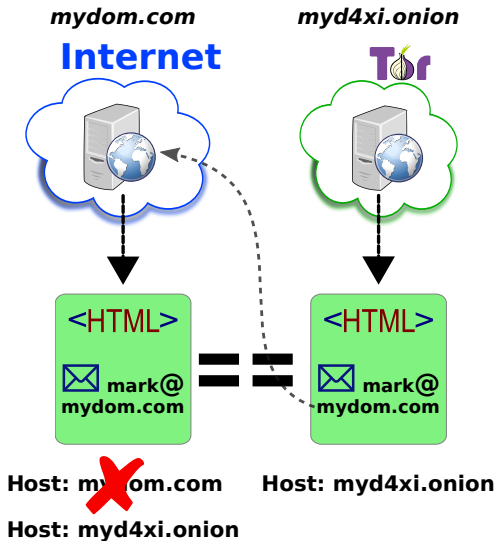
1. **myd4xi.onion** is a HS
2. Request a resource
3. HTML contains email address

How Do Location Leaks Happen?



1. **myd4xi.onion** is a HS
2. Request a resource
3. HTML contains email address
4. Contact **mydom.com**

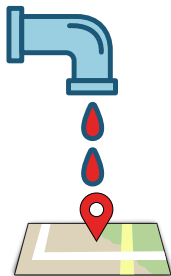
How Do Location Leaks Happen?



1. **myd4xi.onion** is a HS
2. Request a resource
3. HTML contains email address
4. Contact **mydom.com**
5. Contact **mydom.com** using the HS as "Host" header

1. Candidate Selection

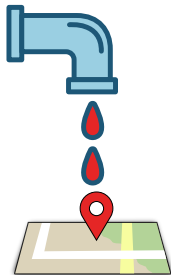
- a) interaction with the Hidden Service
- b) extraction of candidate endpoints



Approach Overview

1. Candidate Selection

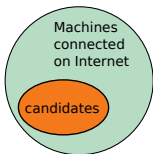
- a) interaction with the Hidden Service
- b) extraction of candidate endpoints



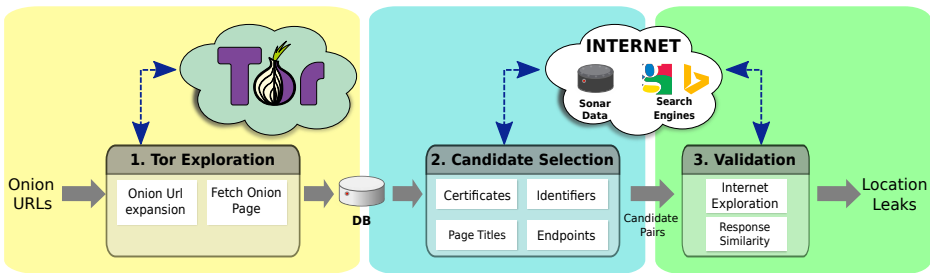
2. Validating Candidates

- a) interaction with each candidate
- b) validation of the responses





1. Novel approach for deanonymizing HS
 - location leaks
 - open-world model
2. Approach implemented in CARONTE
3. First measurement study of prevalence of location leaks within HS



1. Input onion URLs

- No central repository of all hidden services
- Sources:
 - hidden services listings
 - hidden services search engines
 - Internet search engines
 - blogs, pastebin applications, forums...
- Coverage:
 - $\approx 15\text{K}$ onion URLs collected
 - $\approx 6\text{K}$ **unique onion domains**

2. Visit onion URLs and collect data.



Candidate Selection: EXAMPLE

```
<!DOCTYPE html>
<html lang="en" id="facebook" class="no_js">
  <head>
    <meta charset="utf-8">
    <title id="pageTitle">Facebook</title>
    <link type="text/css" rel="stylesheet"
      href="https://fbstatic-a.akamaihd.net/rsrc.php/v2/yU/r/Z8FgpY_Its6.css" />
  </head>

  <body>
    ...
    <footer>
      <div class="topcontainer">
        <ul class="nav navbar-nav fright">
          Donations:<a href="bitcoin:1BitmixerEiyy3eTLaCpgBbhYERs48qza">
            1BitmixerEiyy3eTLaCpgBbhYERs48qza</a>
          </ul>
        </div>
      </footer>
    </body>
    <!-- Phone: +34-11-222-333 -->
    <!-- Fax: +34-12-121-1212 -->
    <!-- Email: mark@zuckerberg.com -->
  </html>
```

Candidate Selection: EXAMPLE

```
<!DOCTYPE html>
<html lang="en" id="facebook" class="no_js">
  <head>
    <meta charset="utf-8">
    <title id="pageTitle">Facebook</title>
    <link type="text/css" rel="stylesheet"
      href="https://fbstatic-a.akamaihd.net/src.php/v2/yU/r/Z8FgpY_Its6.css" />
  </head>
  <body>
    ...
    <footer>
      <div class="topcontainer">
        <ul class="nav navbar-nav fright">
          Donations:<a href="bitcoin:1BitmixerEiyypp3eTLaCpgBbhYERs48qza">
            1BitmixerEiyypp3eTLaCpgBbhYERs48qza</a>
          </ul>
        </div>
      </footer>
    </body>
    <!-- Phone: +34-11-222-333 -->
    <!-- Fax: +34-12-121-1212 -->
    <!-- Email: mark@zuckerberg.com -->
  </html>
```

1) Domains in URLs

1) Domains in emails

Candidate Selection: EXAMPLE

```
<!DOCTYPE html>
<html lang="en" id="facebook" class="no_js">
  <head>
    <meta charset="utf-8">
    <title id="pageTitle">Facebook</title>
    <link type="text/css" rel="stylesheet"
      href="https://fbstatic-a.akamaihd.net/src.php/v2/yU/r/Z8FgpY_Its6.css" />
  </head>
  <body>
    ...
    <footer>
      <div class="topcontainer">
        <ul class="nav navbar-nav fright">
          Donations:<a href="bitcoin:1BitmixerEiyypp3eTLaCpgBbhYERs48qza">
            1BitmixerEiyypp3eTLaCpgBbhYERs48qza</a>
          </ul>
        </div>
      </footer>
    </body>
    <!-- Phone: +34-11-222-333 -->
    <!-- Fax: +34-12-121-1212 -->
    <!-- Email: mark@zuckerberg.com -->
  </html>
```

1) Domains in URLs

2) Identifiers



Google Analytics
Google AdSense

1) Domains in emails

Candidate Selection: EXAMPLE

```
<!DOCTYPE html>
<html lang="en" id="facebook" class="no_js">
  <head>
    <meta charset="utf-8">
    <title id="pageTitle">Facebook</title>
    <link type="text/css" rel="stylesheet"
      href="https://fbstatic-a.akamaihd.net/src.php/v2/yU/r/Z8FgpY_Its6.css" />
  </head>
  <body>
    ...
    <footer>
      <div class="topcontainer">
        <ul class="nav navbar-nav fright">
          Donations:<a href="bitcoin:1BitmixerEiyypp3eTLaCpgBbhYERs48qza">
            1BitmixerEiyypp3eTLaCpgBbhYERs48qza</a>
        </ul>
      </div>
    </footer>
  </body>
  <!-- Phone: +34-11-222-333 -->
  <!-- Fax: +34-12-121-1212 -->
  <!-- Email: mark@zuckerberg.com -->
</html>
```

3) Titles

1) Domains in URLs

2) Identifiers

1) Domains in emails



Google Analytics
Google AdSense

Candidate Selection: EXAMPLE

```
<!DOCTYPE html>
<html lang="en" id="facebook" class="no_js">
  <head>
    <meta charset="utf-8">
    <title id="pageTitle">Facebook</title>
    <link type="text/css" rel="stylesheet"
      href="https://fbstatic-a.akamaihd.net/src.php/v2/yU/r/Z8FgpY_Its6.css" />
  </head>
  <body>
    ...
    <footer>
      <div class="topcontain...
        <ul class="nav navbar...
          Donations:<a href="https://www.facebook.com/donations"
            1BitmixerEiyp3eTLaCpgBbnYERs48qza">
          </ul>
        </div>
      </footer>
    </body>
    <!-- Phone: +34-11-222-333 -->
    <!-- Fax: +34-12-121-1212 -->
    <!-- Email: mark@zuckerberg.com -->
  </html>
```

3) Titles

1) Domains in URLs

4) SSL certificates

2) Identifiers

1) Domains in emails







Google Analytics
Google AdSense

Candidate Selection: Certificates

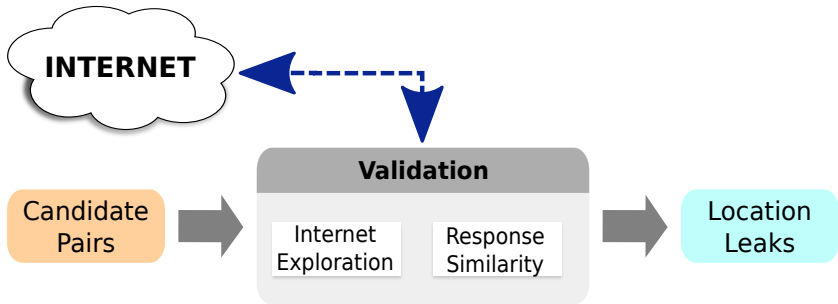


- 68 Internet-wide HTTPS scans
- Oct '13 - Feb '15
- 205 GB with **35M certificates**

Leaf certificates used for:

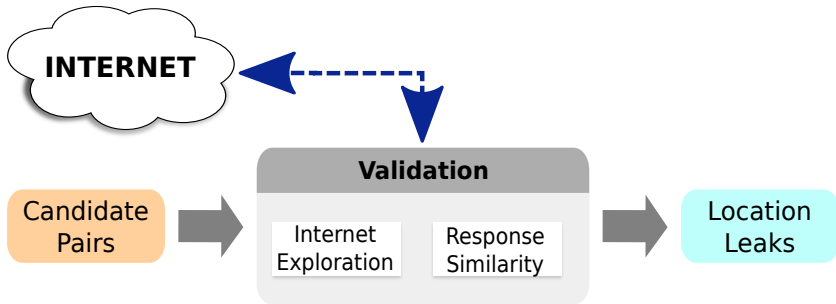
-  IPs/domains in Subject CN
-  Search leaf Certificate
-  Search public key
-  Search for onion address

Validation



<facebookcorewwi.onion , google.com>
<facebookcorewwi.onion , facebook.com>

Validation

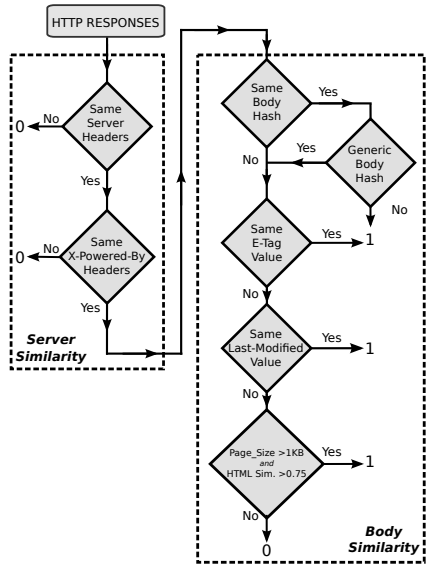


facebookcorewwi.onion != google.com
facebookcorewwi.onion == facebook.com

Validation Algorithm

Given $\langle \text{candidate}, \text{onion} \rangle$:

1. Connect to the candidate
2. Fetch resources from the candidate
3. Compare “exploration” and “validation” content



Determining Leak Intention

Are the service owners aware of the content leaks?



Determining Leak Intention

Are the service owners aware of the content leaks?



Leak intentional if:

- Onion address \approx Internet candidate
- Internet site contains onion address
- Onion page title \approx Internet domain

Results Summary

Method	Candidates		Deanonymizations	
	Pairs	Onions	All	Unintentional
Endpoints	4,704	793	67	32 (48%)
Identifiers	192	66	12	2 (16%)
Titles	200	157	44	20 (45%)
Certificates	366	63	30	18 (60%)
TOTAL	5,462	841	101	51

- 1,974 live onion addresses (31%)
- 101 Hidden Services **deanonymized (5%)**
 - 50% unintentional leaks
- **21% deanonymized on Tor relays**






- Use a dedicated Web server
- Bind the Web server to localhost
 - Tor requests answered; Internet forbidden
 - use a firewall
- Site auditing
- Avoid reuse of certs. and public keys
- Avoid Tor relays

Tor Project already recommends some of these:

`https://www.torproject.org/docs/tor-hidden-service.html.en`

Ethical Considerations

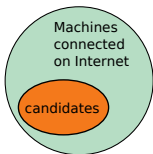
- Experiments on **live Tor network**

-  no network degradation
-  no malicious relays
-  no access to users traffic



- Approved by ethical board
- downloaded only to HTML pages (i.e., no images or videos)
- **No** data release
- Reported to Tor Project at submission

Conclusions

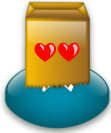


1. Novel approach for deanonymizing HS
 - location leaks
 - open-world model
2. Approach implemented in CARONTE
3. First measurement study of prevalence of location leaks
 - 5% services deanonymized
 - 21% deanonymized on Tor relays

Questions?



Caronte: Potential Users



Caronte: Etymology



In Greek mythology, Charon or Kharon is the ferryman of Hades who carries souls of the newly deceased across the rivers Styx and Acheron that divided the world of the living from the world of the dead.