Bayes, not Naïve

Security Bounds on Website Fingerprinting Defenses

Giovanni Cherubin

Privacy Enhancing Technologies Symposium
Minneapolis, Minnesota, USA
19 July, 2017
Website Fingerprinting (WF)
Website Fingerprinting (WF)

Adversary $\Phi$: transmission time, total bandwidth, ...

$f_{\text{train}}$: SVM, logistic regression, ...

Diagram:
- Time $t$ with indicators of increase and decrease.
- Adversary with observed features.
“Lookup-Table” Approach
(Cai et al., ’14)

Idealised Adversary: knows exactly what packet sequences each web page may generate. Count the collisions.

Lookup table
Distinguishing Web Pages

$P_x \mid y={\text{startpage.com}}$

$P_x \mid y={\text{freeimages.com}}$

$R^*: \text{Bayes Error}$

Total communication time
“Bayes estimate” approach

\[ f \leftarrow (\Phi, f_{\text{train}}) \]

\[ R^f : \text{error on new packet sequence} \]

\[ \frac{L-1}{L} \left( 1 - \sqrt{1 - \frac{L}{L-1} R_{NN}^N} \right) \leq R^* \leq R^f \]

(Cover & Hart, ’67)
(ε, Φ)-privacy

**Problem** An error estimate $\hat{R}^*$ alone does not convey information about the setting. Random guessing $R^G$:

- $R^G = 2/3$
- $R^G = 1/2$

Define metric $(1 - \text{Adv})$:

$$\epsilon = \frac{\hat{R}^*}{R^G}$$
### (ε,Φ)-privacy

**Closed World, WCN+ dataset (Tor traffic)**

<table>
<thead>
<tr>
<th>Defense*</th>
<th>(ε,Φ)-privacy</th>
<th>Packet OH</th>
<th>Time OH</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Defence</td>
<td>(0.06, k-NN)</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Decoy Pages</td>
<td>(0.43, k-NN)</td>
<td>134%</td>
<td>59%</td>
</tr>
<tr>
<td>WTF-PAD</td>
<td>(0.49, k-FP)</td>
<td>247%</td>
<td>0%</td>
</tr>
<tr>
<td>BuFLO</td>
<td>(0.58, k-FP)</td>
<td>110%</td>
<td>79%</td>
</tr>
<tr>
<td>CS-BuFLO</td>
<td>(0.63, k-FP)</td>
<td>67%</td>
<td>576%</td>
</tr>
<tr>
<td>Tamaraw</td>
<td>(0.70, k-NN)</td>
<td>258%</td>
<td>341%</td>
</tr>
</tbody>
</table>

* Tor's default defense, Randomized Pipelining, is underlying each defense.
Did Feature Sets Improve?

(How much)

Bayes Error Estimate

Liberatore & Levine  Dyer et al.  Wang et al.  Panchenko et al.

Hayes & Danezis

2006  2012  2014  2016  2017

No Defence  Decoy Pages  BuFLO  Tamaraw
Summary & Future Work

Blackbox method to derive security bounds for any WF defense and adversary \((\Phi, \cdot)\)

**Future Work**

- Prove some \(\Phi\) is complete in some sense ("efficient"): from \((\varepsilon, \Phi)\)-privacy to \(\varepsilon\)-privacy
- Other estimates of \(R^*\), ensembles
- Other applications of technique: traffic analysis, side channel, generic ML-based attacks
Bayes, not Naïve

Security Bounds on Website Fingerprinting Defenses

Giovanni Cherubin

Privacy Enhancing Technologies Symposium
Minneapolis, Minnesota, USA
19 July, 2017
Lower bound convergence
Theorem  Let $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$, then $R^{k-\text{NN}} \to R^*$ (Stone, '77)
Comparision with Cai et al.

<table>
<thead>
<tr>
<th>Defence</th>
<th>R* estimate</th>
<th>Cai et al.</th>
<th>Cai et al. (full information)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BuFLO</td>
<td>57%</td>
<td>53%</td>
<td>19%</td>
</tr>
<tr>
<td>Tamaraw</td>
<td>69%</td>
<td>91%</td>
<td>11%</td>
</tr>
</tbody>
</table>
### (ε,Φ)-privacy

One VS All scenario, WCN+ dataset

<table>
<thead>
<tr>
<th>Defence</th>
<th>(ε,Φ)-privacy</th>
<th>Time OH</th>
<th>Packet OH</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Defence</td>
<td>(0.05, k-NN)</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Decoy Pages</td>
<td>(0.29, k-NN)</td>
<td>134%</td>
<td>59%</td>
</tr>
<tr>
<td>BuFLO</td>
<td>(0.29, k-FP)</td>
<td>110%</td>
<td>79%</td>
</tr>
<tr>
<td>Tamaraw</td>
<td>(0.25, k-NN)</td>
<td>258%</td>
<td>341%</td>
</tr>
<tr>
<td>CS-BuFLO</td>
<td>(0.16, k-FP)</td>
<td>67%</td>
<td>576%</td>
</tr>
<tr>
<td>WTF-PAD</td>
<td>(0.18, CUMUL)</td>
<td>247%</td>
<td>0%</td>
</tr>
</tbody>
</table>
Q: What about priors?

- If true prior probabilities on web pages known, they can be used (i.e., bias the dataset accordingly).

- Ratio of success of one-try adversaries over random guessing maximized by uniform priors (Braun et al., 2009).
Q: Open World?

Adversary knows

Victim may visit

y = “open”
Q: Bounds on full info?

**Theorem** For any transformation $\Phi: P \rightarrow X$, $R^*(P) \leq R^*(\Phi)$
Q: Is the code available?

Yes

https://github.com/gchers/wfes
Bayes, not Naïve
Security Bounds on Website Fingerprinting Defenses

Giovanni Cherubin
Privacy Enhancing Technologies Symposium
Minneapolis, Minnesota, USA
19 July, 2017