Bangalore University
University Visvesvaraya College of Engineering
Department of Computer Science and Engineering

Website Fingerprinting Attacks
- A Practical Threat?

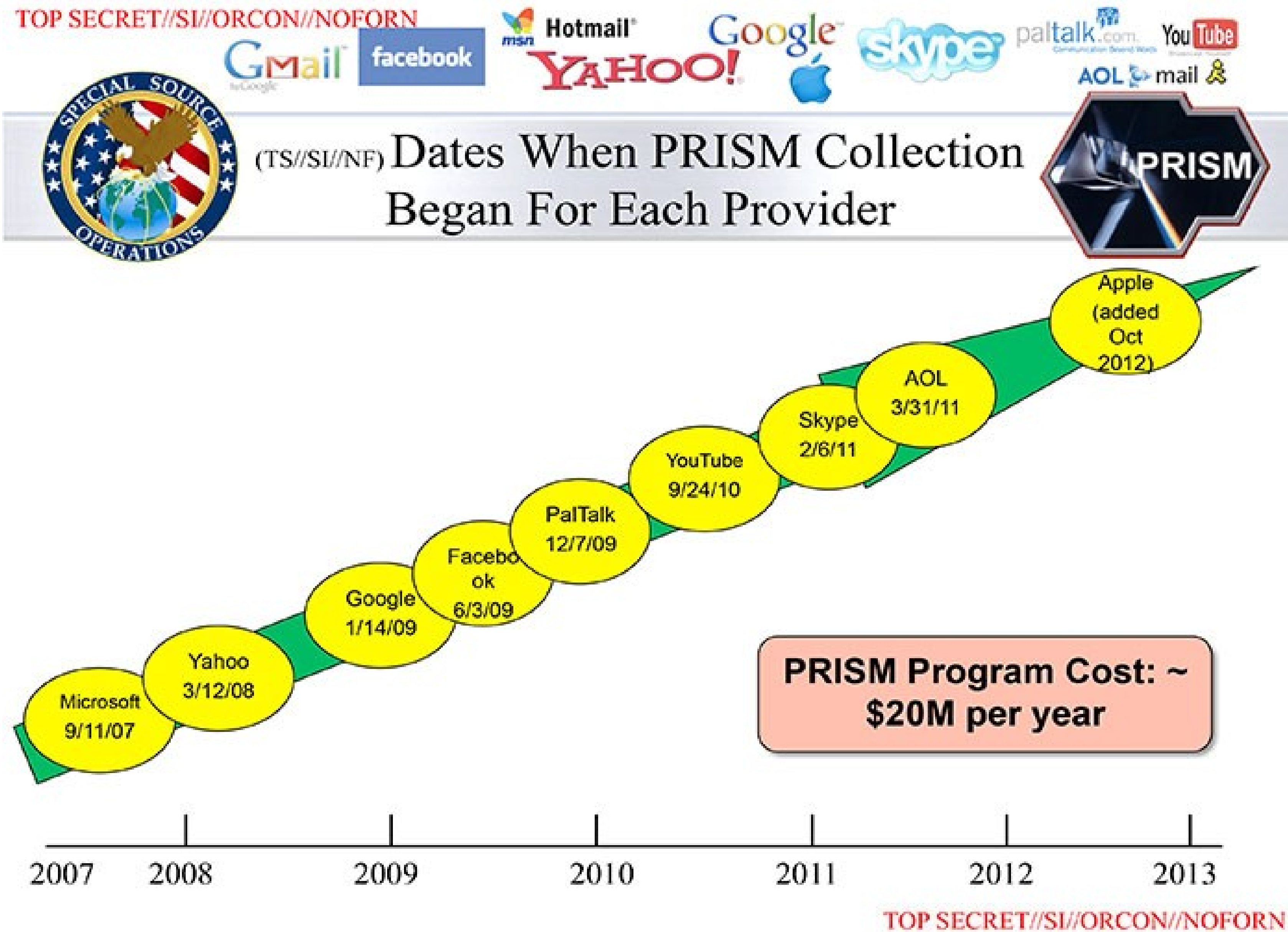
Presented by
Vignesh T Prabhu
(14GAMT3018)

Under the guidance
Of
Dr. Thriveni J
Associate Professor

Content

- Introduction to TOR Project
- Website Fingerprinting Attack
- About the paper
- Goals
- Assumptions
- “Classify-Verify” Algorithm
- Analysis
- Adversary's Cost
- Conclusions

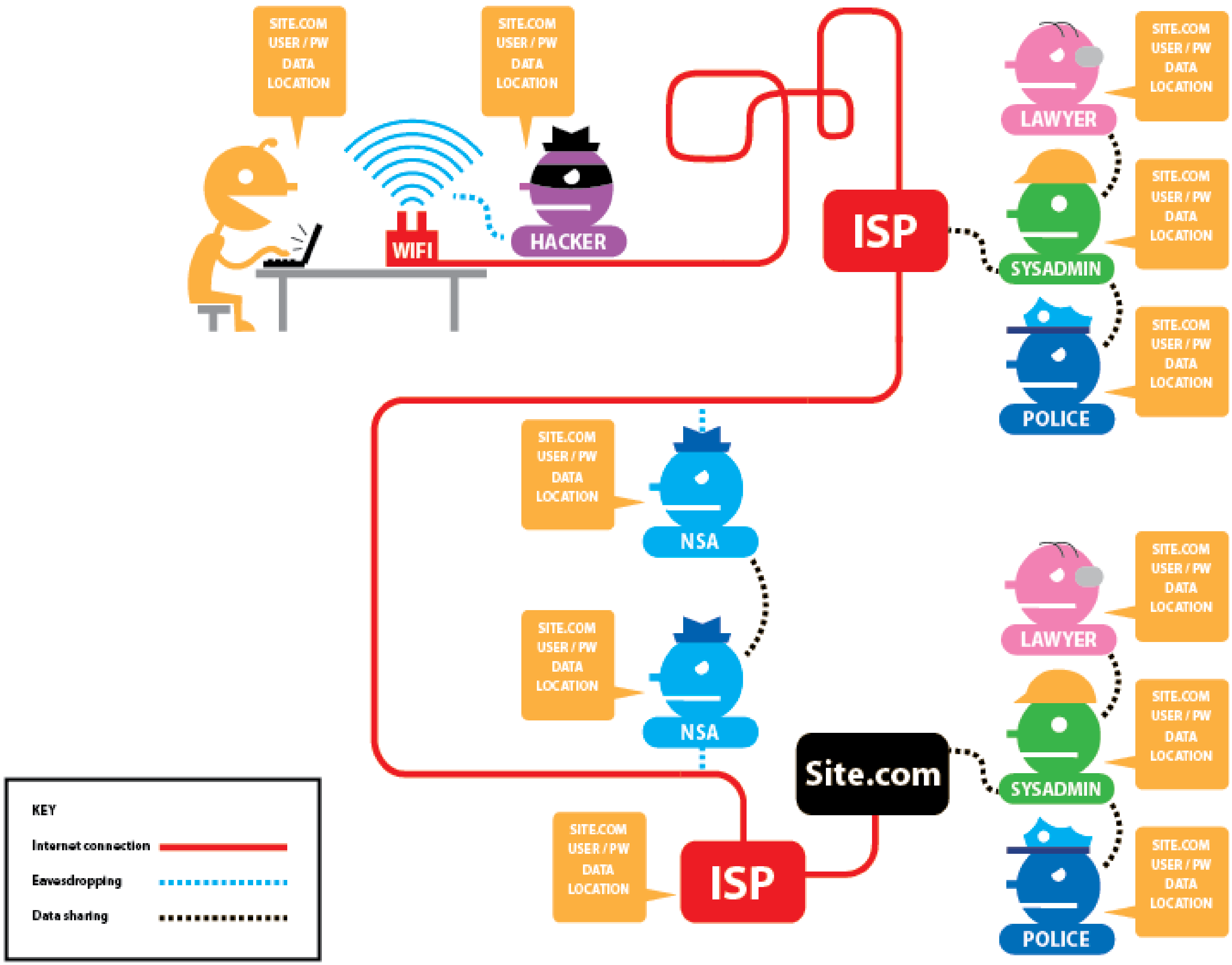
Big Brother is watching you?



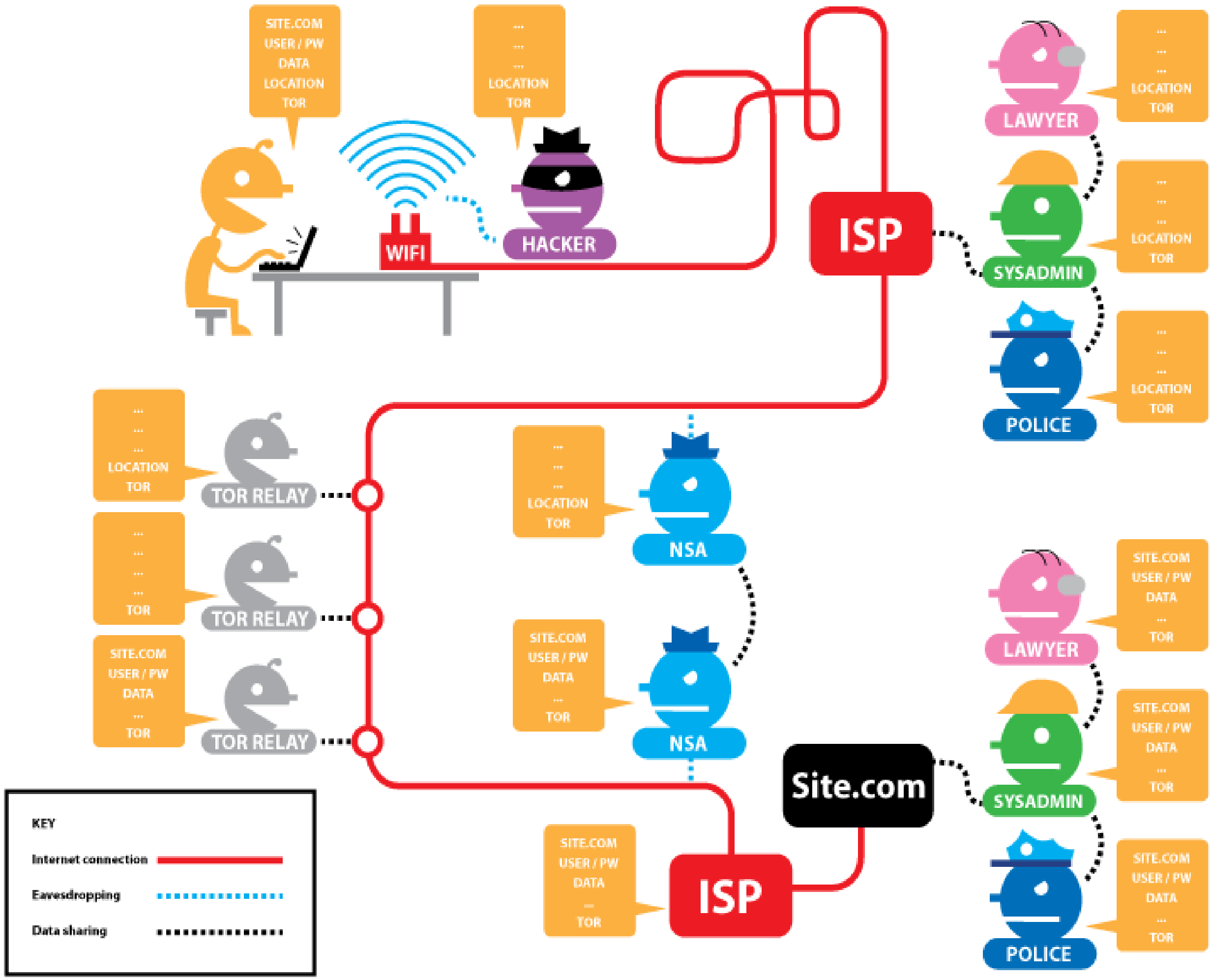
TOR Project

- TOR – THE ONION ROUTING
- Launched on 20th Sept 2002
- Most popular distributed anonymous communication systems
- More than 3 million daily users
- Initially developed by U.S. Naval Research Laboratory
- Free Software under BSD License
- Available as TOR Browser Bundle

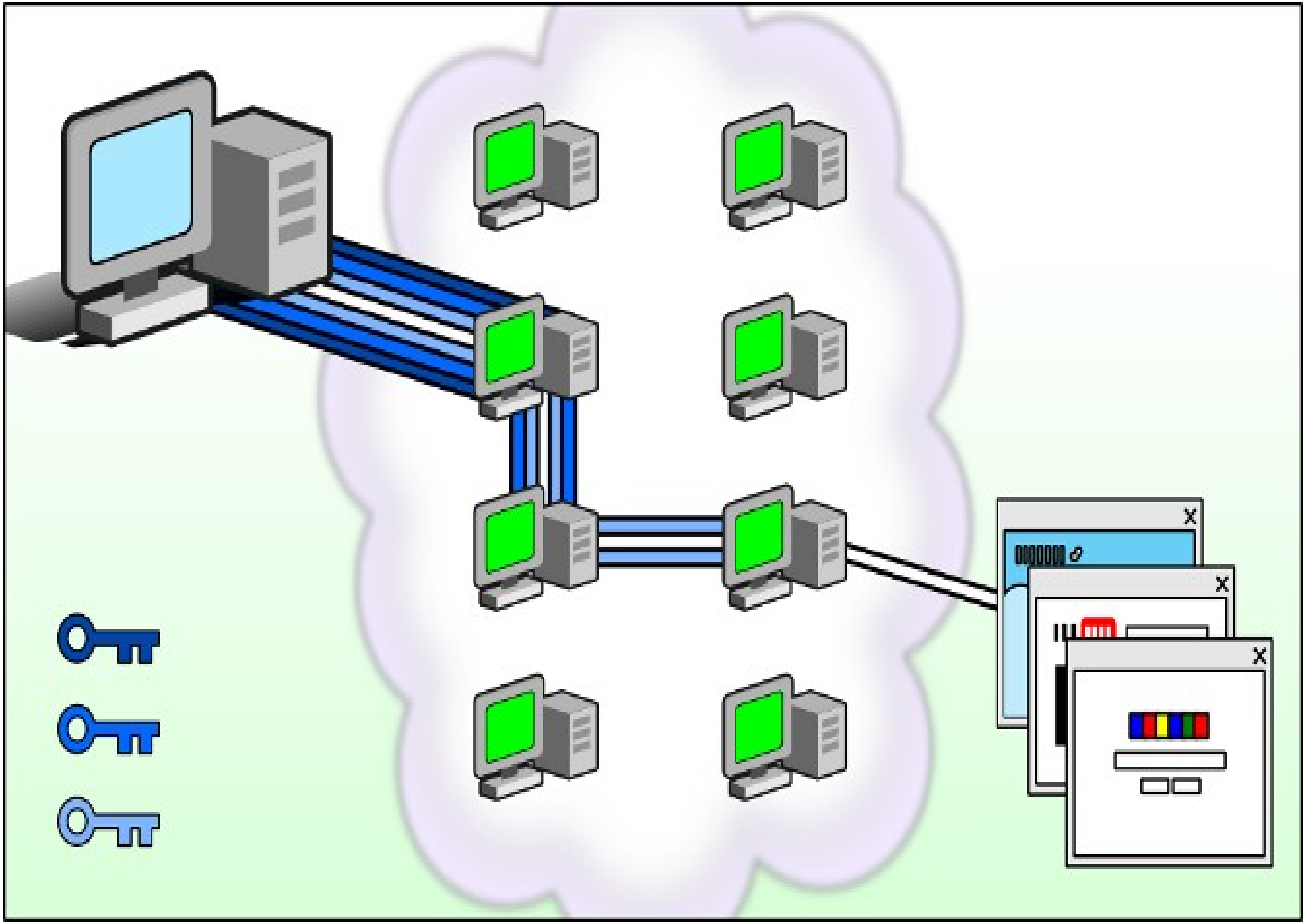
Browsing Without TOR



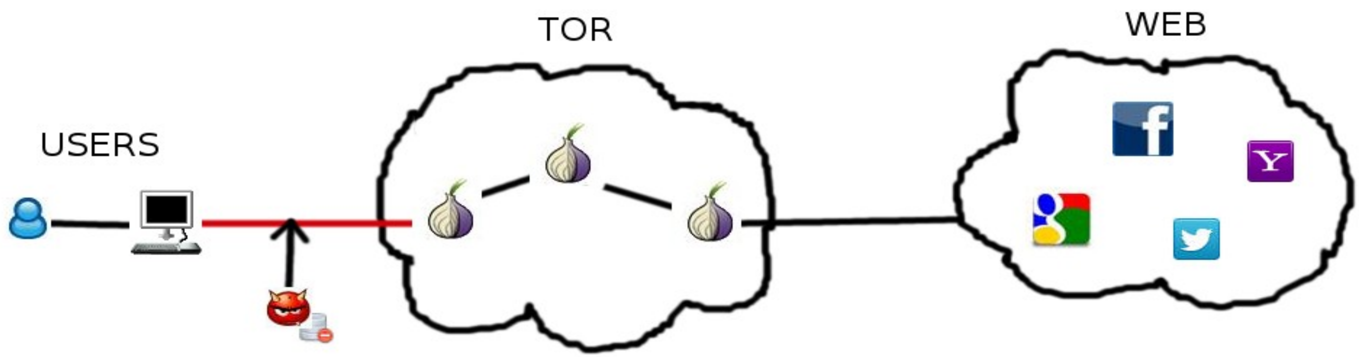
Browsing with TOR

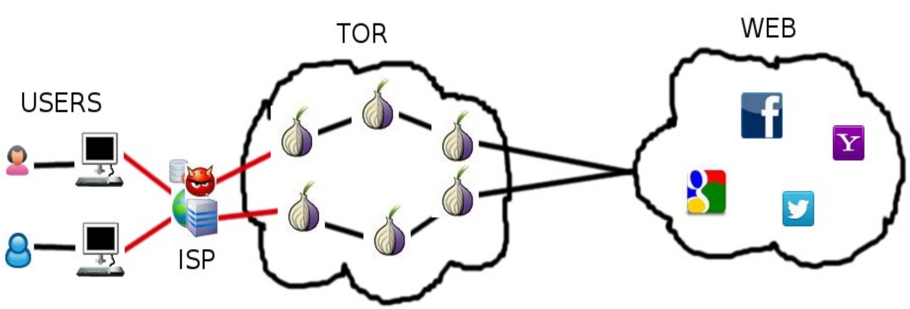


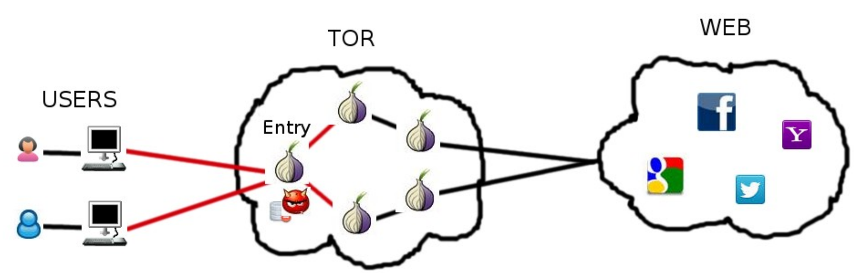
Onion Peeling



Website Fingerprinting Attack







Website Fingerprinting system

- Primary tasks
  - Data Collection
  - Data Training
  - Data Testing
  - Updating Data
- Misc
  - Collecting background information

About the paper

- Title: A Critical Evaluation of Website Fingerprinting Attacks

- Authors: Marc Juarez, Sadia Afroz, Gunes Acar, Claudia Diaz and Rachel Greenstadt

- Presented: 21st ACM Conference on Computer Security held on November 3rd to 7th, 2014 at Scottsdale, Arizona, USA

Goals

- A critical evaluation of assumptions made by prior WF studies
- An analysis of the variables that affect the accuracy of WF attacks
- An approach to reduce false positive rates
- A model of the adversary’s cost

Assumptions

- Client Settings
  - Closed World
  - Browsing behaviour
- Web
  - Template
  - No Localized versions
- Adversary
  - Page Load parsing
  - No background traffic
  - Replicability

Assumptions

Assumptions	**Explicitly made by**
Closed-world	[11,26]
Browsing behavior	[11]
Page load parsing	[3, 11, 23, 26, 32]
No background noise	[3, 11, 23, 26, 32]
Replicability	[11, 26]
Template websites	[3]

[3] - X. Cai, X. Zhang, B. Joshi, R. Johnson, “***Touching from a Distance: Website Fingerprinting Attacks and Defenses***”

[11] - D. Herrmann, R. Wendolsky, H. Federrath, “***Website Fingerprinting: Attacking Popular Privacy Enhancing Technologies with the Multinomial Nave-Bayes Classifier***”

[23] - A. Panchenko, L. Niessen, A. Zinnen, T. Engel, “***Website Fingerprinting in Onion Routing Based Anonymization Networks***”

[26] - Y. Shi, K. Matsuura, “***Fingerprinting Attack on the Tor Anonymity System***”

[32] - T. Wang, I. Goldberg, “***Improved Website Fingerprinting on Tor***”

Closed World Assumption

Authors	**World Size**	**Success Rate**
Herrmann et al.	775 pages	3%
Shi and Matsuura	20 pages	50%
Panchenko et al.	775 pages	54.61%
Wang and Goldberg	100 pages	90%

Classifiers

Name	Model	Features
H	Naive Bayes	Packet Lengths
P	SVM(Support Vector Machine)	Packet lengths Order Total bytes
D	N-grams	Total time Up/Downstream bytes Bytes in traffic bursts
W	SVM (Fast-Levenshtein)	Cell traces
T	Decision Tree	Packet lengths Order Total bytes

Classify Verify Algorithm*

Algorithm 1 Modified Classify-Verify

Input: Test page D, suspect pages $\mathcal{A} = A_1, ..A_n$ and probability scores

Output: A_D if $A_D \in \mathcal{A}$ and 'Unknown' otherwise

$\triangleright$ Train a classifier
$C_{\mathcal{A}} \rightarrow$ classifier trained on $\mathcal{A}$
$V_{\mathcal{A}} \rightarrow$ verifier for $\mathcal{A}$
$\triangleright$ Calculate threshold for the verifier
$t \rightarrow$ threshold maximizing F_{β} score
$\triangleright$ Test page D
Classify D
$P_D \rightarrow$ Verification score
if $P_D >= t$ **then**
    Accept the classifier's output and return it
else
    Reject the classifier's output and return 'Unknown'
end if

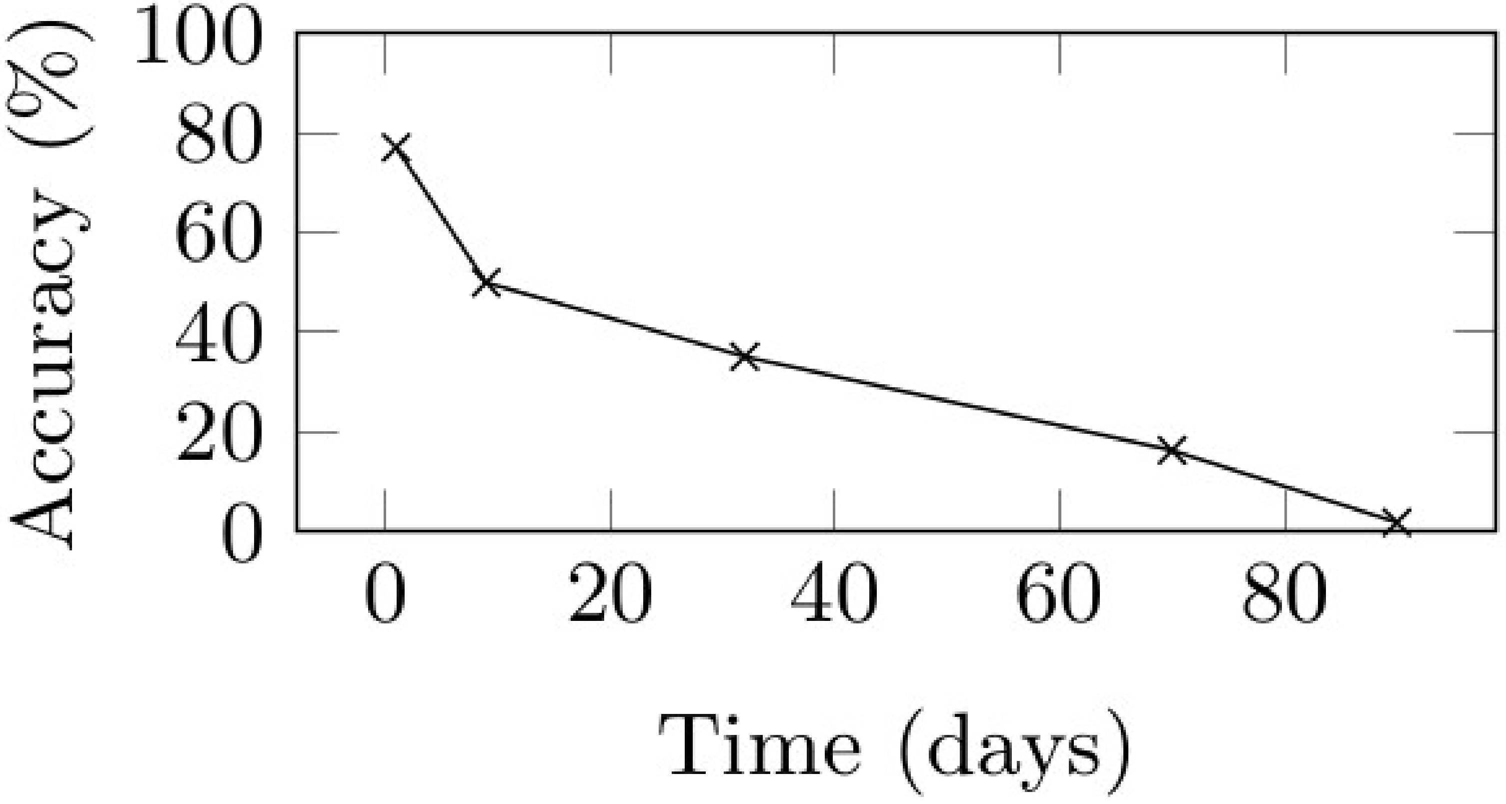
*Discussed by Stolerman et al.

**Led to 63% reduction in False Positive

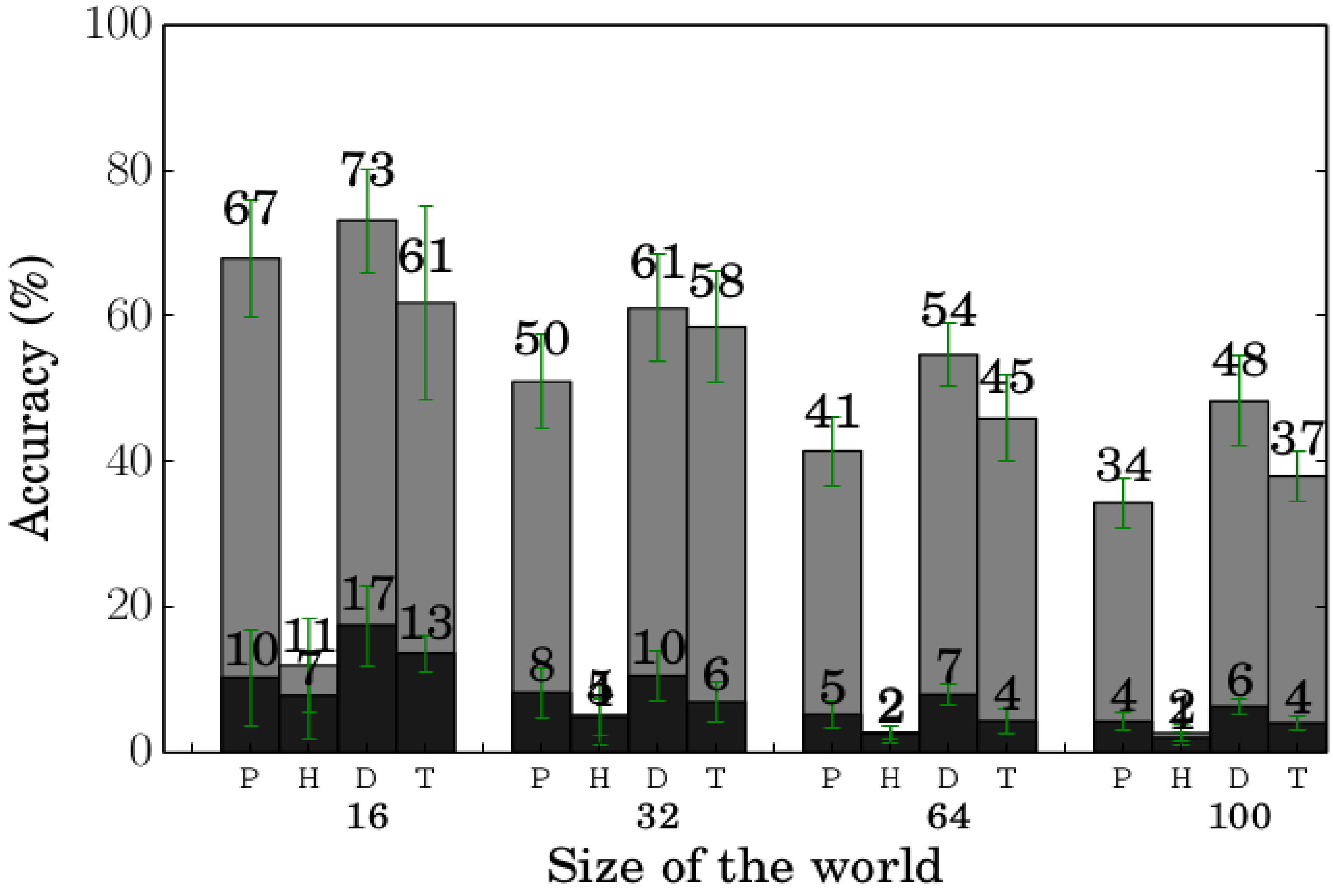
Experiment

- Two steps
  - k-fold cross-validation using data of the control crawl
  - Evaluate classifier’s accuracy training on the control crawl and testing with data from the test crawl

Staleness of data over time



Multitab Browsing Accuracies



Multitab Browsing Accuracies

For Classifier W

Delay	**Acc test**	**Acc control**
0.5 sec	9.8% (±3.1%)	77.08% (±2.72%)
3 sec	7.9% (±0.8%)	77.08% (±2.72%)
5 sec	8.23%(±2.32%)	77.08% (±2.72%)

Accuracy for different network locations

Location Trained	**Location Tested**	**Acc test**	**Acc control**
Leuven	New York	8.83% (±2.87%)	66.95% (±2.872%)
Leuven	Singapore	9.93% (±0.98%)	66.95% (±2.87%)
Singapore	New York	68.53%(±3.24%)	76.40% (±5.99%)

Classify-Verify result on ALAD* Users

ALAD User	TP	FP	New TP	New FP
User 3	38/260	362/400	31.2/260	107.6/400
User 13	56/356	344/400	26.8/356	32/400
User 42	3/208	397/400	1.0/208	41.2/400

*ALAD – Active Linguistic Authentication Dataset

Adversary's Cost

- Data Collection Cost:
  - *N* training pages, *m* versions, *i* instances
- Training Cost:
  - *D* total pages, *F* features, *C* classifier
- Testing Cost:
  - *T* test data(*v* victims, *p* visited pages per day), *F* features, *C* classifier
- Updating Cost:
  - *d* website change frequency
- Background Information Cost

Conclusions

- Success of WF attacks also depend on
  - Temporal proximity of traces
  - TBB versions used
  - User's Browsing habits
- Non-targeted attack seems not feasible due to its sophistication
- Targetted attack is also non-trivial
  - aspects of their behavior must be observed a priori
- Future research on WF attacks should also focus on its practicality and efficacy

Thank You