Var-CNN and DynaFlow: Improved Attacks and Defenses for Website Fingerprinting

Sanjit Bhat (PRIMES) David Lu (PRIMES) Albert Kwon (MIT) Srini Devadas (MIT)
sanjit.bhat@gmail.com davidboxboro@gmail.com

May 20th, 2018
PRIMES Conference

Motivation and Background

Anonymity matters

- Whistleblowers
- Governmental suppression of political opinion
- Censorship circumvention

http://blog.transparency.org/2016/06/20/new-whistleblower-protection-law-in-france-not-yet-fit-for-purpose/

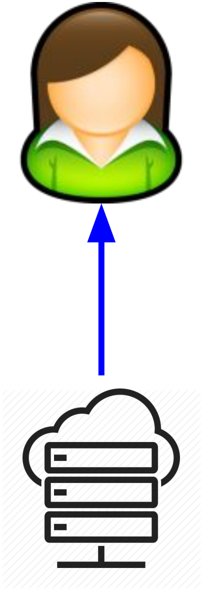
http://facecrooks.com/Internet-Safety-Privacy/To-be-anonymous-or-not-to-be-should-you-use-your-real-name-on-the-Internet.html/

http://www.dmnews.com/social-media/what-if-people-want-their-internet-anonymity-back/article/338654/

The internet provides limited anonymity

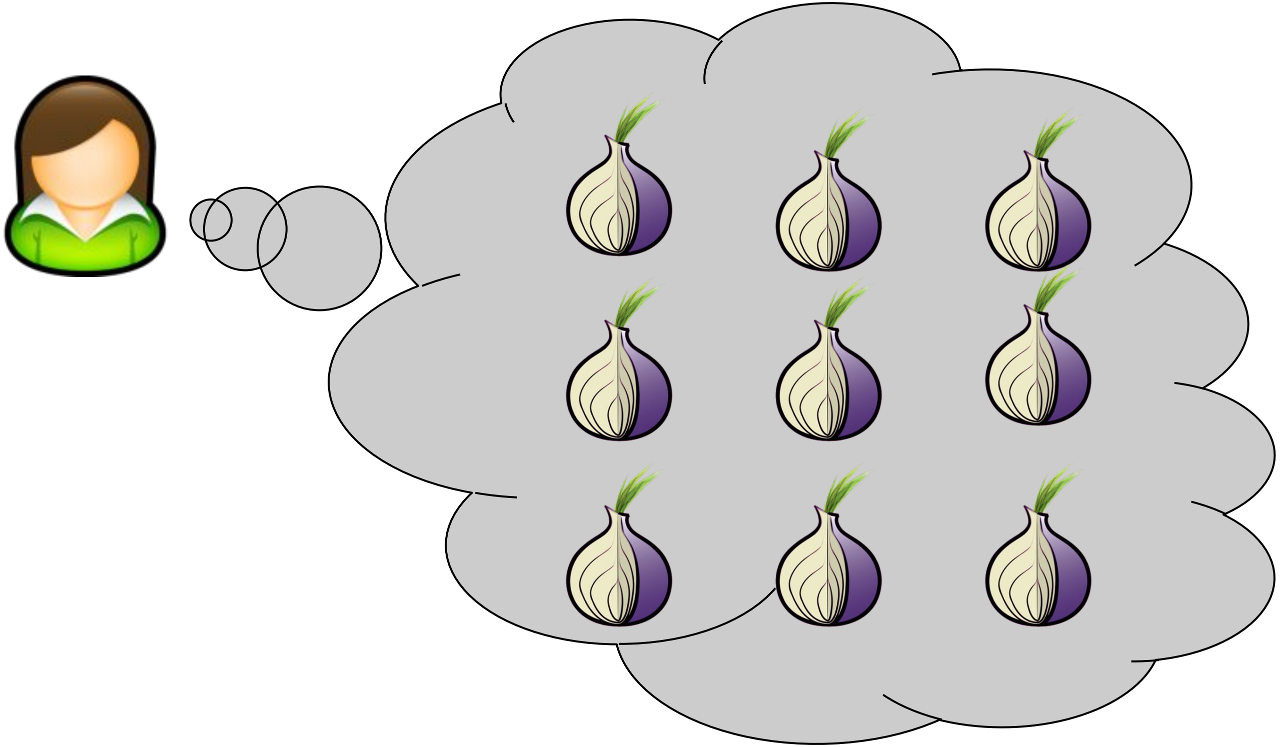
A supposed fix - Tor: The Onion Router

- Alice connects to the Tor network



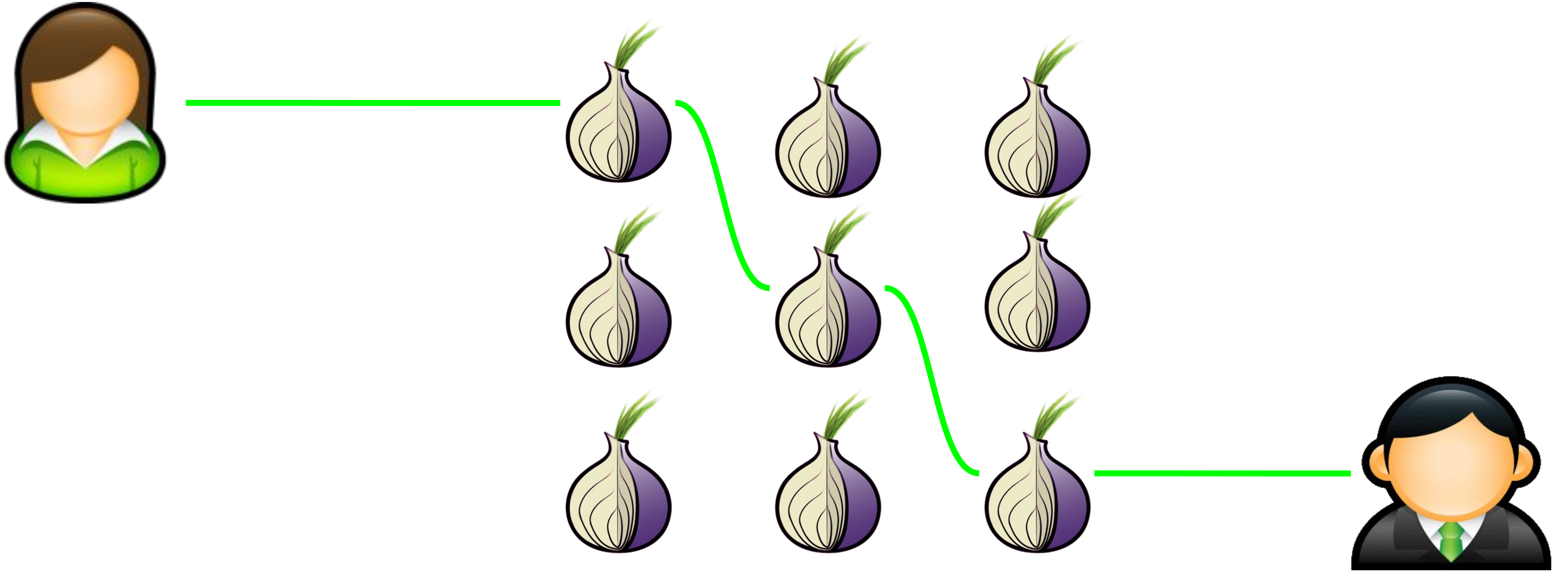
A supposed fix - Tor: The Onion Router

- Alice obtains a list of Tor nodes from the Tor network



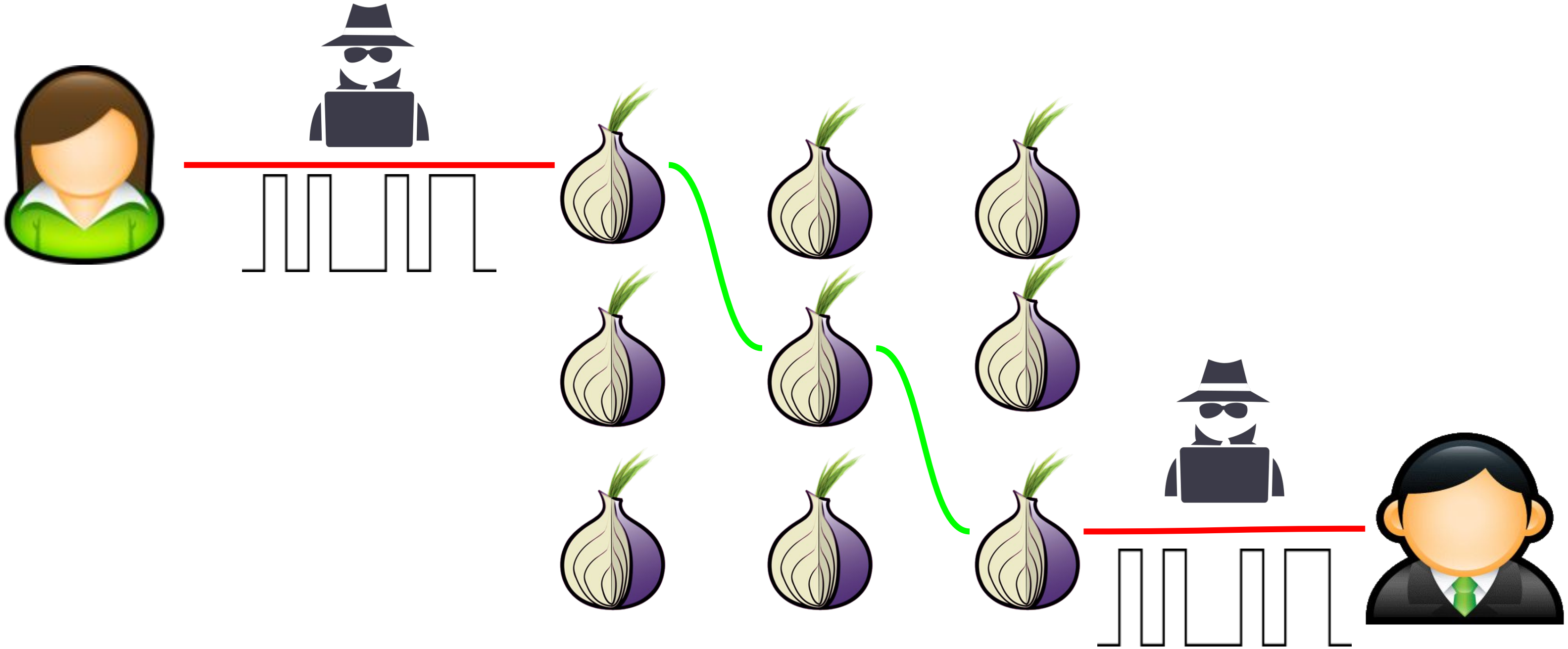
A supposed fix - Tor: The Onion Router

- Alice chooses 3 Tor nodes to make a connection to Bob
- No Tor nodes know the identities of both Bob and Alice



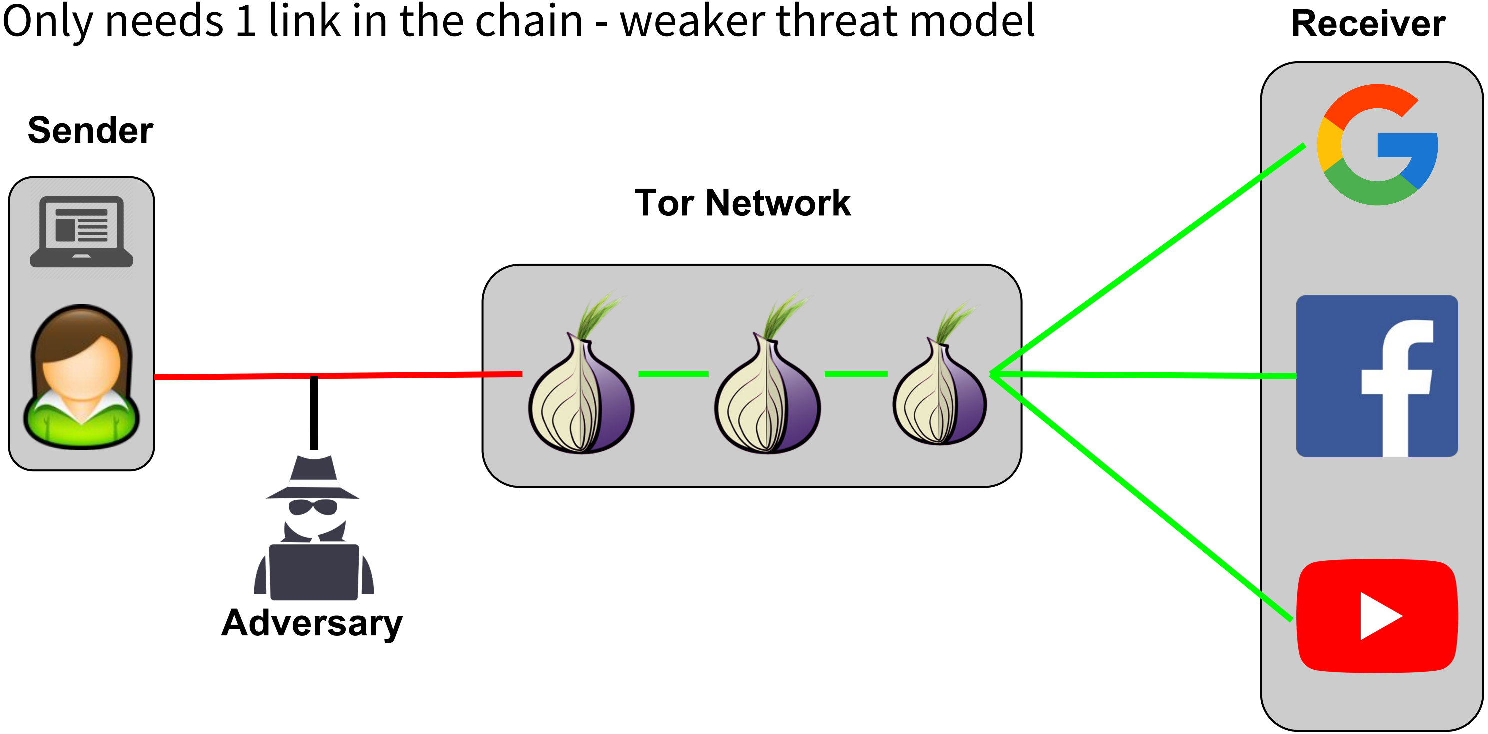
Traffic analysis attacks

- Adversary correlates Alice and Bob's traffic
- Only works when adversary intercepts both entry and exit points



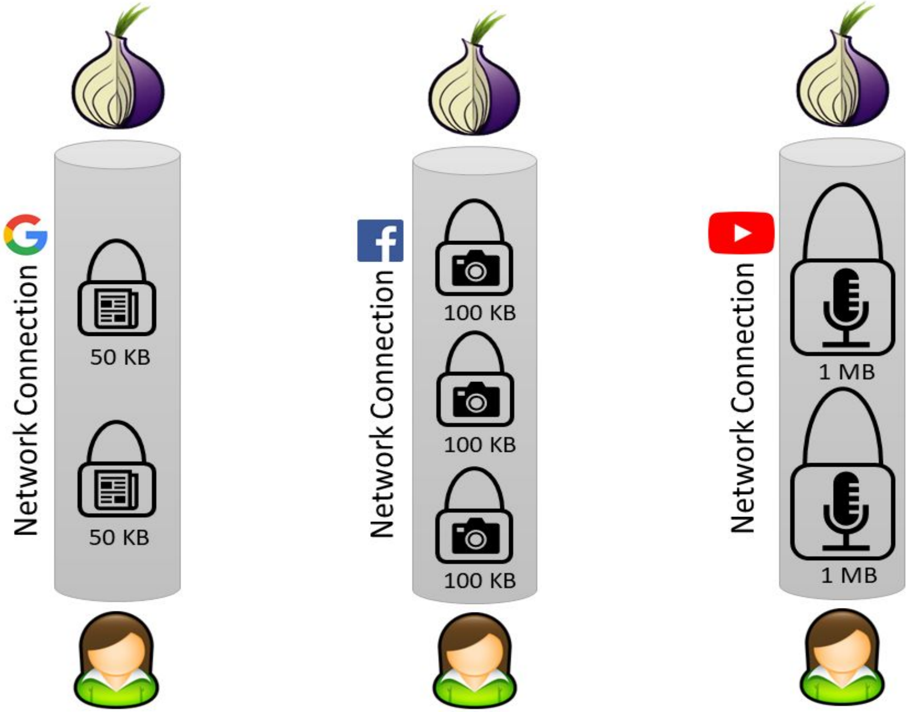
Website fingerprinting (WF) attacks

- Adversary collects database offline and uses it to fingerprint online
- Only needs 1 link in the chain - weaker threat model



Simplified WF attack scenario

- Each website exhibits characteristic load behavior



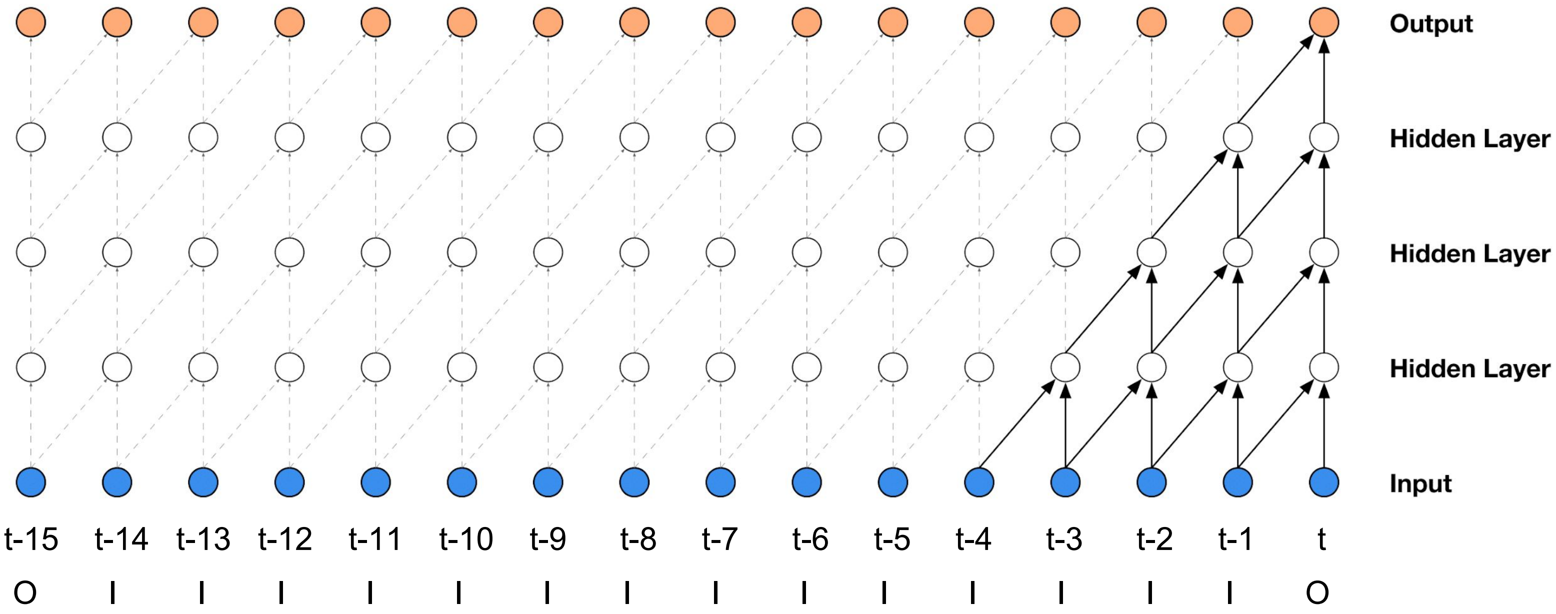
Var-CNN: Automated feature extraction using variations on CNNs

Why automated feature extraction?

- Uses raw Tor traffic sequences: incoming/outgoing direction, timestep
- Resists network protocol changes
- Could discover more optimal features than humans can find

Dilated convolutions

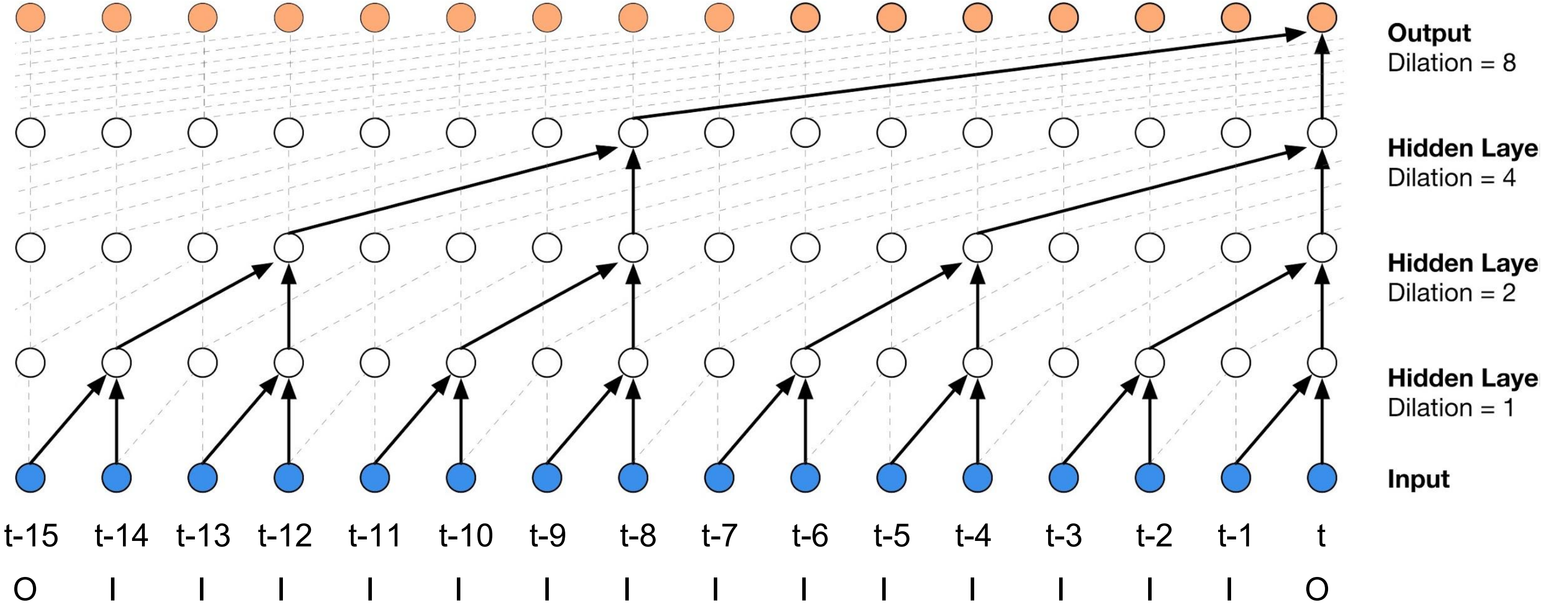
- Packet sequence inherently time-dependent



A. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu. Wavenet: A generative model for raw audio. arXiv, 2016.

Dilated convolutions

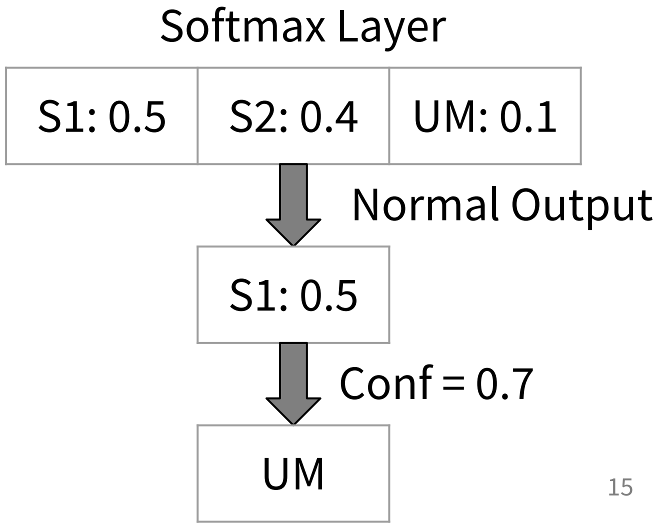
- Sacrifice fine-grain detail for broader field of view



A. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu. Wavenet: A generative model for raw audio. arXiv, 2016.

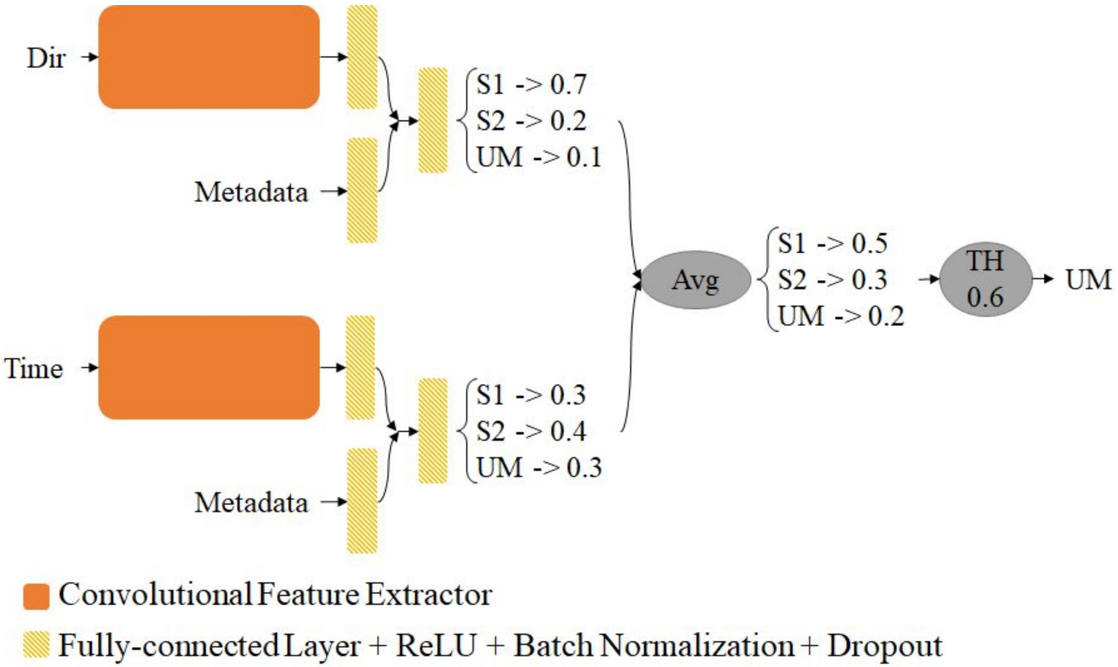
Other techniques

- Cumulative features
  - Total number of packets
  - Number of incoming and outgoing
  - Ratio of incoming to total and outgoing to total
  - Total transmission time
  - Average number of packets per second
- Confidence thresholds
  - Threshold for attacker certainty
  - Adjust types of classification made



Ensemble model

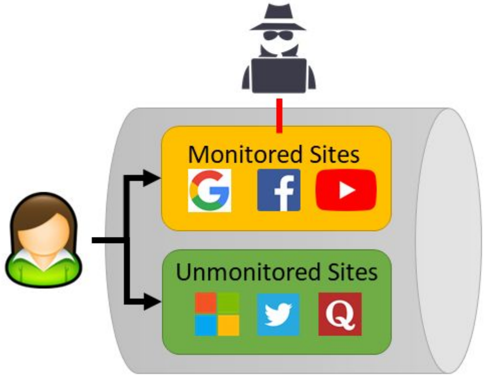
- Using timesteps should leak more info to attacker
- No past pre-extracted timing features performed well



Var-CNN Results

Experimental setup

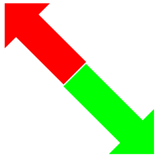
- Wang et al. *k*-NN data set - blocked pages for monitored, popular pages for unmon
- ≤ training data used by competing attacks
- Re-randomize train/test sets and average results over 10 trials
- Metrics
    - *True Positive Rate* (TPR) - Prop. of monitored sites correctly classified
    - *False Positive Rate* (FPR) - Prop. of unmonitored sites incorrectly classified

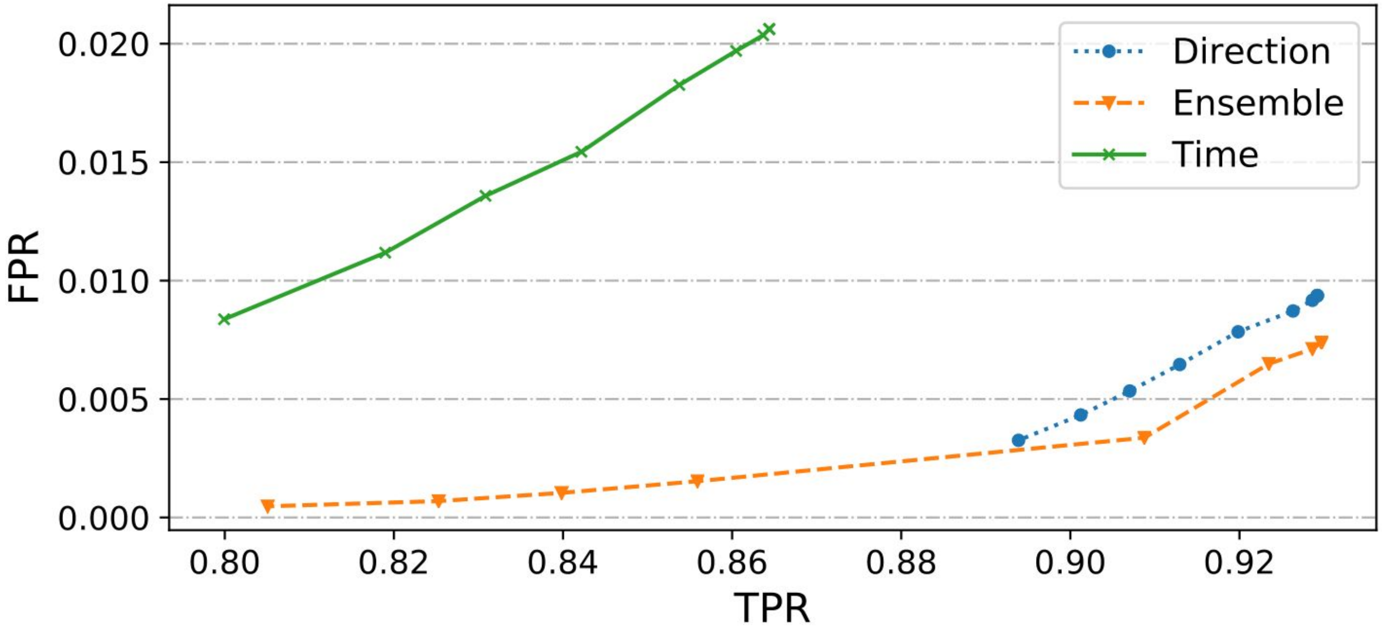


Open-World

Ensemble model and confidence threshold

- Alone, time model is worse than direction model
- However, their performance is additive
- TPR and FPR decrease as confidence threshold increases

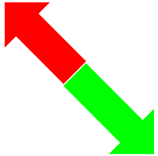


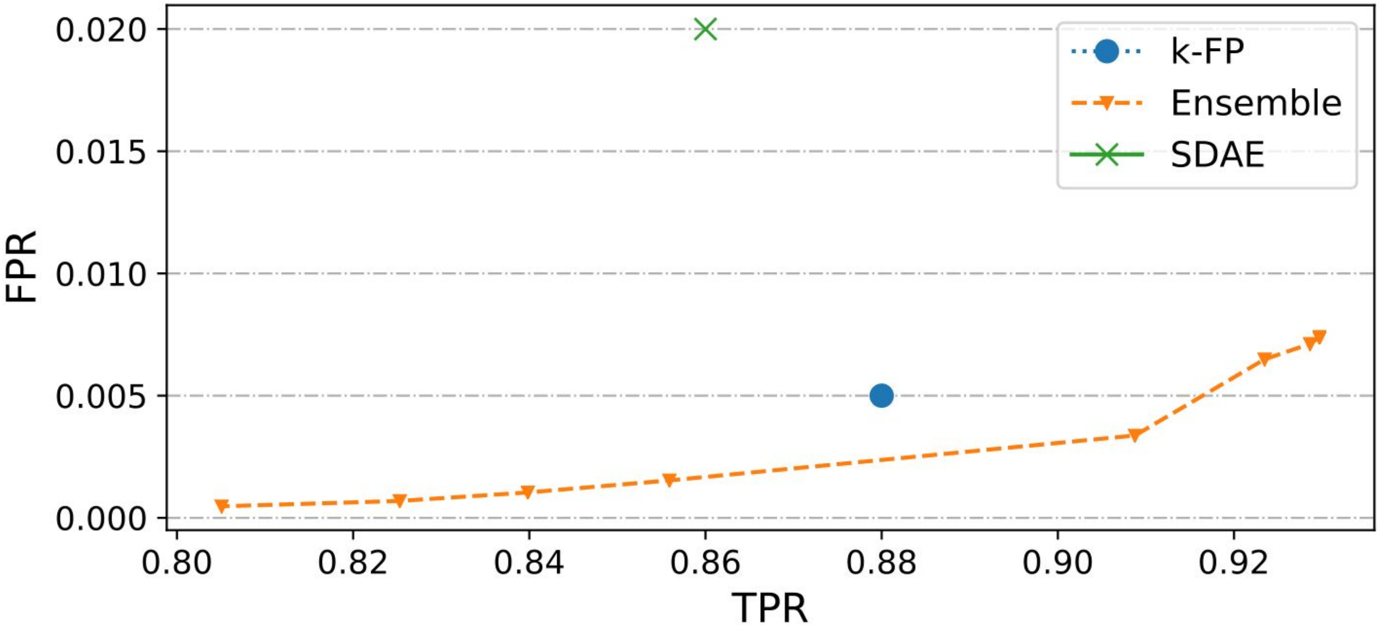


Open-world performance

- 5% better TPR than SDAE
- Over a sixth the FPR of SDAE

- 3% better TPR than *k*-FP
- Nearly half the FPR of *k*-FP





DynaFlow: A new defense based on dynamically-adjusting flows

Existing WF defenses

***1) Limited defenses** - Designed to counter existing attacks*

Drawback: No provable guarantees

***2) Supersequence-based defenses** - Sends "Supersequence" of web trace*

Drawbacks: Requires constantly updated database; does not protect static content

***3) Constant-flow defenses** - Sends a continuous stream of network traffic*

Drawback: High overheads

Advantages of DynaFlow

	Low Latency	Low Bandwidth Usage	Strong Security Guarantees	Protects Dynamic Content	No Database Required	Highly Tunable
DynaFlow	✓	✓	✓	✓	✓	✓
BuFLO [13]	✗	✗	✗	✓	✓	✗
Tamaraw [7]	✗	✗	✓	✓	✓	✗
Supersequence [40]	✗	✗	✓	✗	✗	✗
Walkie-Talkie [42]	✓	✓	✓	✗	✗	✓
Glove [29]	✗	✗	✓	✗	✗	✗
WTF-PAD [21]	✓	✓	✗	✓	✓	✗
Decoy Pages [32]	✓	✗	✗	✓	✓	✗
LLaMA [10]	✓	✓	✗	✗	✗	✗

Overview of DynaFlow

Our goal: *to construct a defense with similar guarantees as prior art but with significantly lowered overheads.*

Three Components:

1) Burst-pattern morphing
2) Constant traffic flow with dynamically changing intervals
3) Padding the number of bursts

Burst-pattern morphing

- Traffic is morphed into fixed **bursts**: 1 outgoing packet followed by 4 incoming packets
- Dummy packets added to morph traffic

Before padding:

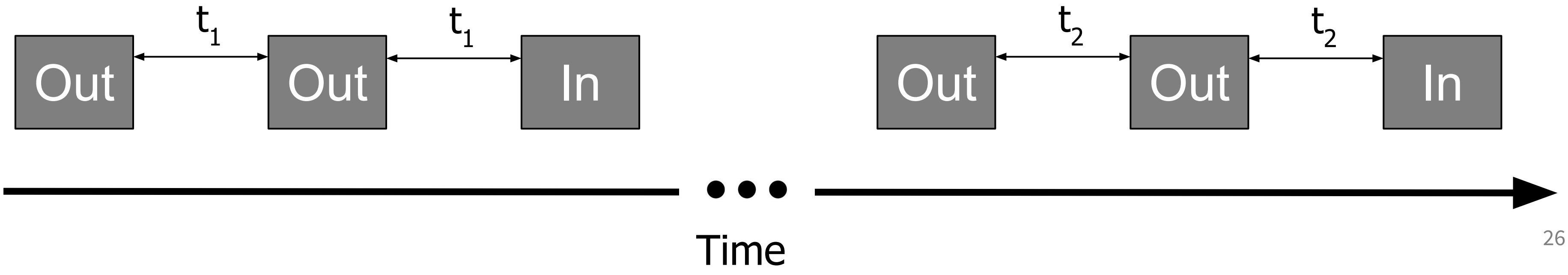
Out	In	Out	In	In	In

After padding (red packets are dummy packets):

Out	In	In	In	In	Out	In	In	In	In

Inter-packet timing

- Packets are sent every ***t*** seconds
- The value of ***t*** dynamically changes to fit the loading page
- There are three tunable parameters: ***a, b, T***
  - The value of ***t*** changes every ***b*** bursts
  - Up to ***a*** adjustments total
  - The value of ***t*** is chosen from the set $\boldsymbol{T = \{t_1, \ldots, t_k\}}$



The number of bursts

- The number of bursts is padded to ***{[m], [m²], [m³], … }***
- Advantages of padding to a power of *m*
    - Significantly mitigate privacy loss
    - Incur reasonably-small overhead
- Example: when ***m = 2***, the bandwidth overhead is under 100%

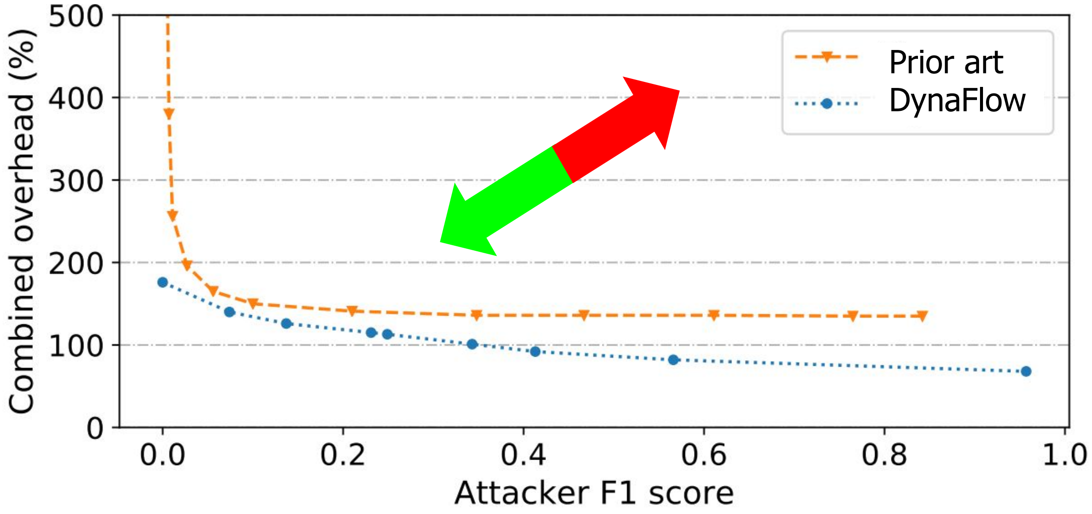
DynaFlow Results

Open-world eval. against existing attacks

DynaFlow against existing attacks. All values are in %.

	k-NN [40]		*k*-FP [14]		Var-CNN		TOH	BWOH
	TPR	FPR	TPR	FPR	TPR	FPR		
No defense:	84.5	2.5	86.3	1.6	89.1	0.7	0	0
Medium security:	15.4	20.6	5.0	1.6	10.8	3.0	23	59
High security:	5.9	69.0	4.4	40.1	0.6	0.9	28	112

Open-world evaluation against prior art



- 31% F1 score: 29% TPR, 11% FPR
  - DynaFlow: 101% overhead (29% TOH, 73% BWOH)
  - Prior art: 138% overhead (40% TOH, 98% BWOH)
- Gap increases for larger F1 scores

Conclusion

- Var-CNN uses novel variants of CNNs to improve upon prior work:
    - Be highly tunable in terms of TPR-FPR trade-off
    - Outperform all prior attacks, all while using ≤ amount of training data
- DynaFlow overcomes challenges of prior WF defenses:
    - Lower overhead than prior work while providing stronger security
    - Protects dynamic content & no database required
- Current status
    - Preprint on arXiv
    - All code and data sets publically available

Acknowledgements

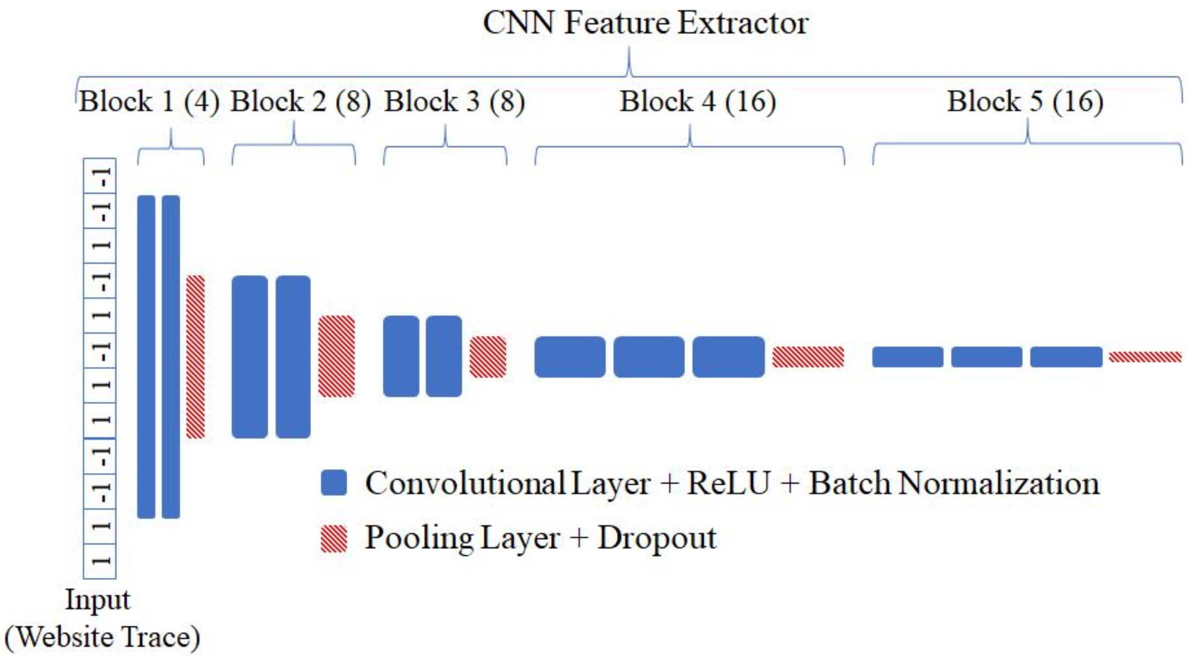
Thank you to:

- Our parents
- Albert Kwon, for providing advice every step of the way
- Prof. Devadas, for giving feedback on the paper and running PRIMES CS
- Dr. Gerovitch and the PRIMES program, for providing research opportunities to high school students and sponsoring AWS bills and a GPU :-)

Appendix of Slides

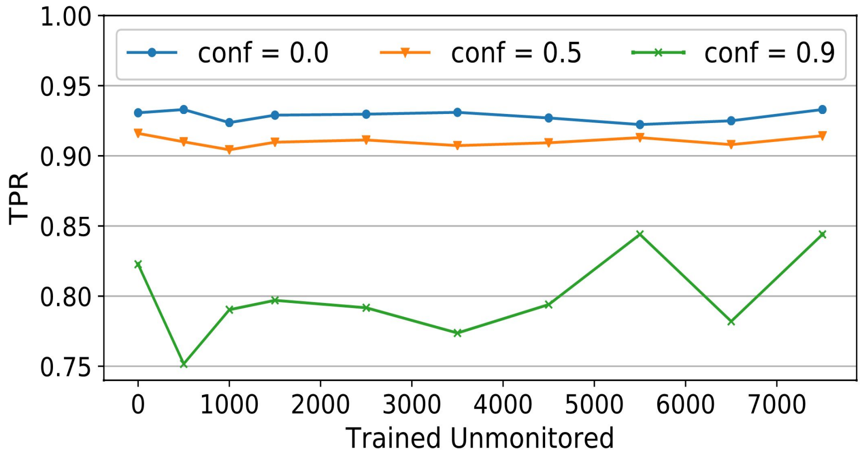
Var-CNN architecture

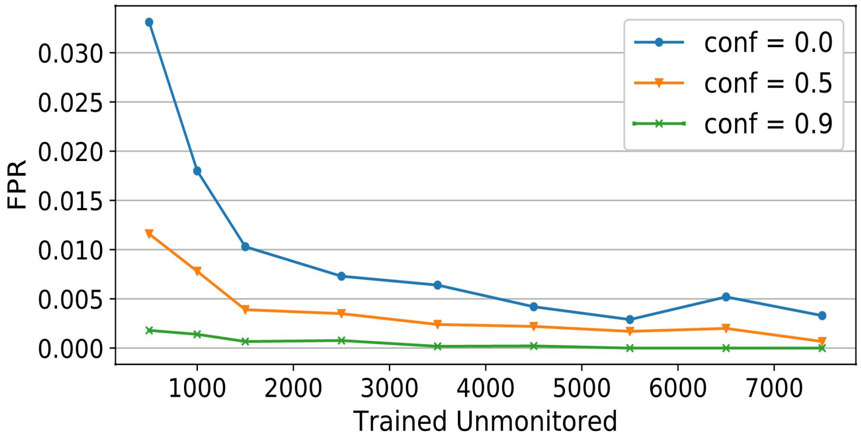
- VGG-16 Convolutional Neural Network (CNN) - ImageNet competition
- Multiple blocks composed of multiple layers for deeper feature extraction



Scaling performance - FPR

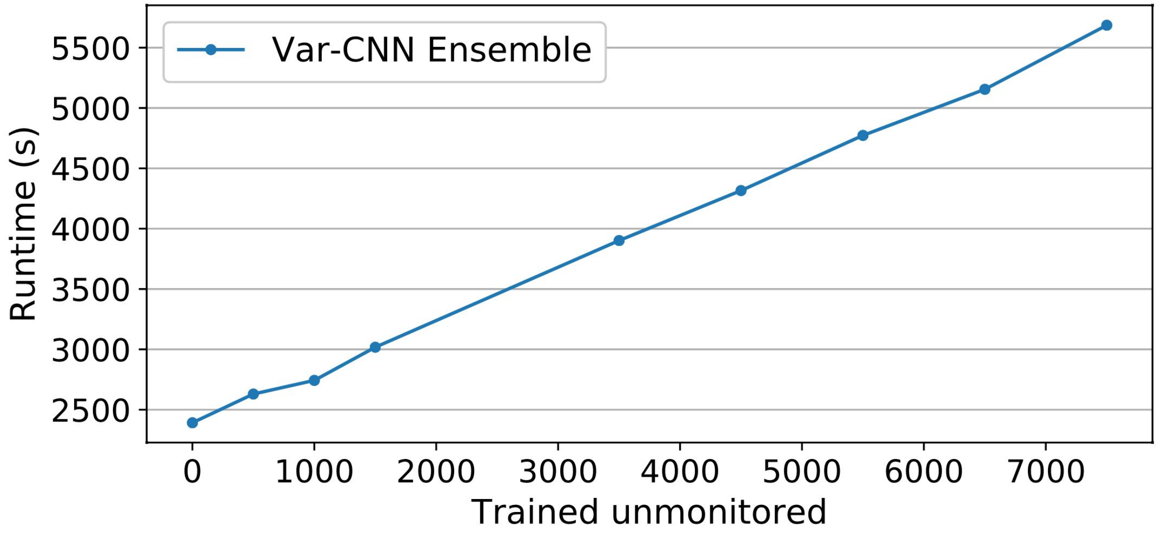
- FPR is incredibly important as open-world size increases
- Training on greater numbers of unmonitored sites retains TPR while reducing FPR
- Var-CNN scales better to larger open-worlds than prior-art attacks





Scaling performance - runtime

- Runtime scales linearly, better than prior models



The optimal attacker

Overview:

- Knows the exact probability that a website ***w*** is visited, generating defended trace ***t***
- Uses this information to make the best guess for which website ***w*** is visited when he sees a trace ***t***
- We can use this information to calculate what the optimal attacker would guess.

Measuring accuracy:

- **F1-score** — harmonic mean of precision and recall (TPR)

Future work

- More powerful deep learning models for Var-CNN
  - Computer vision architectures - DenseNet
  - Recurrent Neural Network architectures - LSTM with Synthetic Gradients
- Find a better way to determine optimal DynaFlow parameters
  - Currently, we sweep parameters one at a time
- Further reduce DynaFlow overheads
  - Total overhead sum can still exceed 100% for stronger configurations