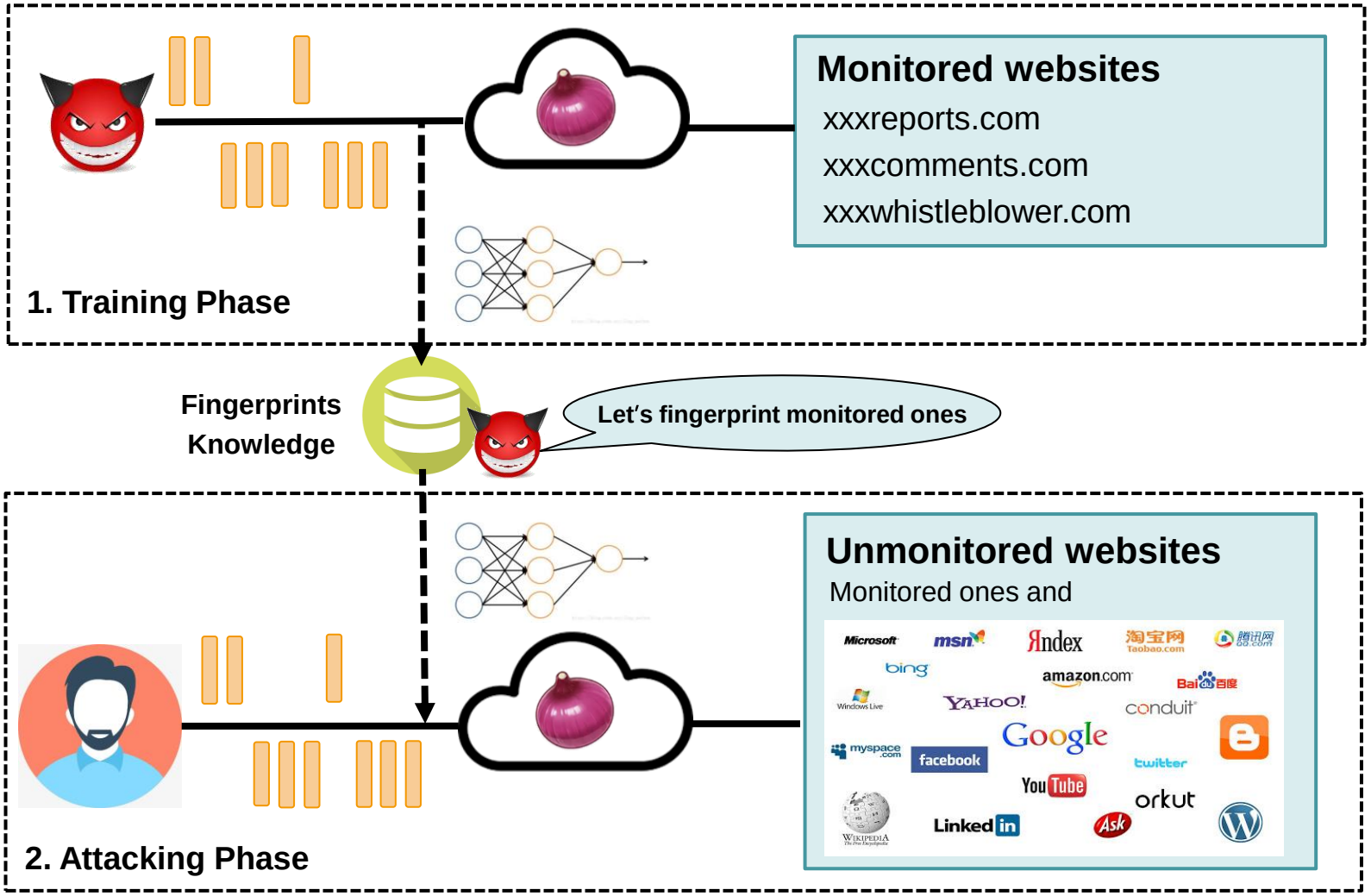
BAPM: Block Attention Profiling Model for Multi-tab Website Fingerprinting Attacks on Tor

Zhong Guan, Gang Xiong, Gaopeng Gou, Zhen Li, Mingxin Cui, Chang Liu*

中国科学院信息工程研究所
INSTITUTE OF INFORMATION ENGINEERING,CAS

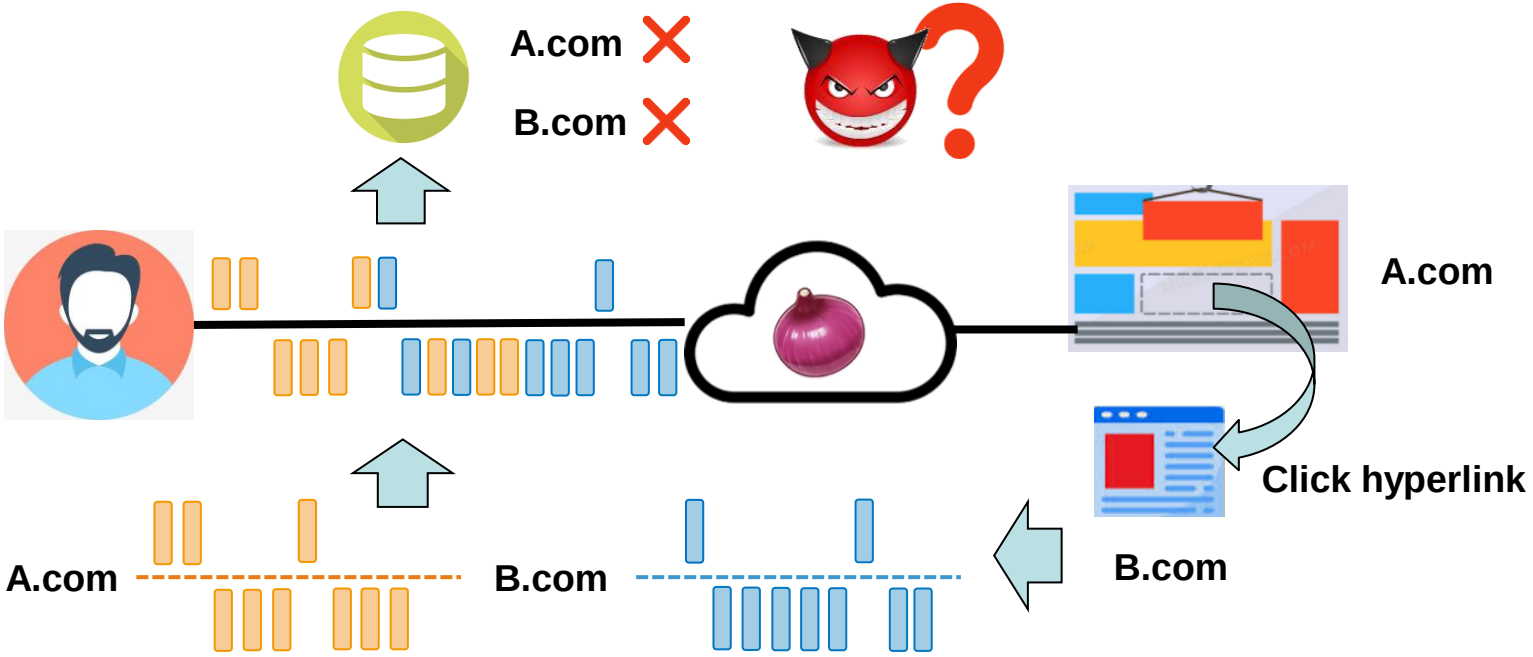
中国科学院大学
University of Chinese Academy of Sciences

Website Fingerprinting Threat Model in Tor



Practical Feasibility of Website Fingerprinting

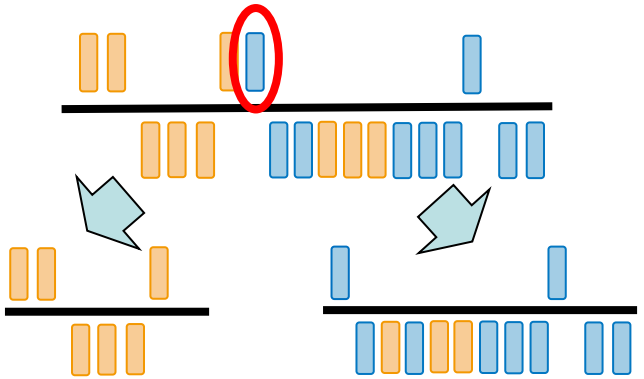
- Website content changes over time
- Anonymous tool version
- Open world (too many unmonitored websites)
- Multi-tab browsing behavior



Existing Multi-tab Website Fingerprinting Attacks

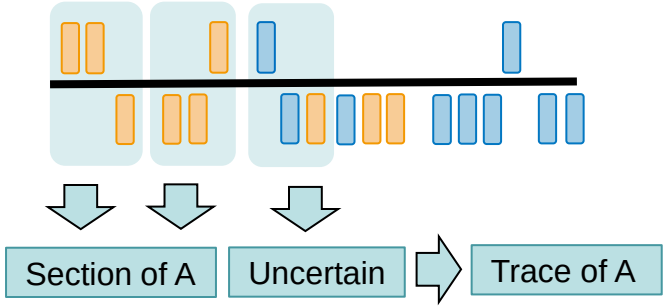
Page Splitting Algorithms (PSP-WF)

- Find the splitting point
- Remove the part behind the splitting point
- Train and attack as single-tab scenarios



Page Sectioning Algorithms (PSE-WF)

- Divide the packet trace into several sections
- Train using sections of single-tab traces
- Test using sections of multi-tab traces (majority voting)



Note: only for observers rather than forwarders

Motivation of Our Research

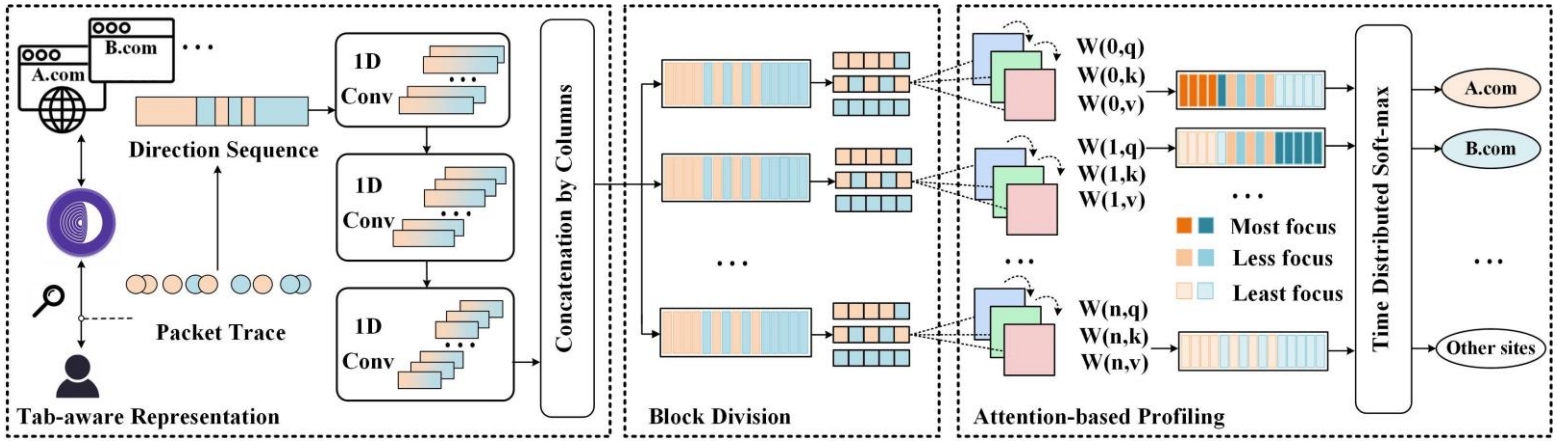
Some Problems

- Complex feature engineering (the PSP-WF algorithm even needs **twice**)
- Potential information in overlapping area is not be used (**lost** & **confusion**)

Research Goal

- Construct an **end-to-end** multi-tab WF model
  - No need of artificial feature engineering
  - No need to predetermine website locations
- Address information **lost** & **confusion** in existing methods
  - Consider the whole packet trace, including the overlapping area
  - Effectively exploit mixed data in overlapping area

Fingerprinting Model Architecture



BAPM: Block Attention Profiling Model

- **Tab-aware representation** — generate basic feature the model depends on
- **Block division** — find single-tab blocks in multi-tab packet traces
- **Attention-based profiling** — aggregate blocks according to their relations

Tab-aware Representation

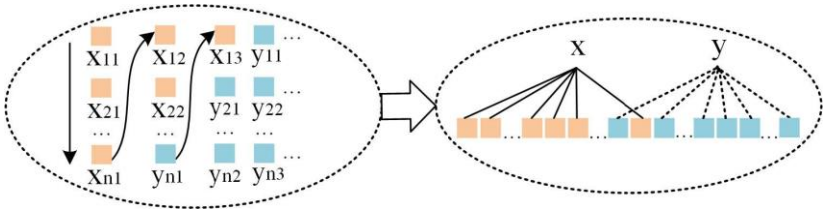
Raw multi-tab packet traces

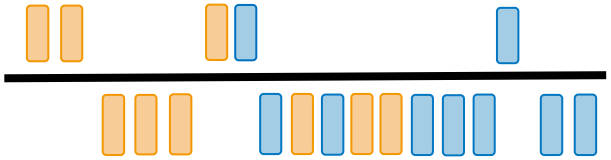
Direction sequence extracted

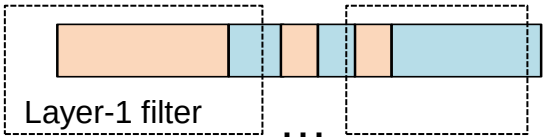
CNN process unit repeats 3 times

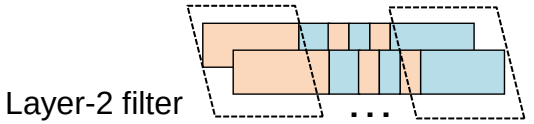
- Convolution
- Activation
- Batch Normalization
- Max Pooling

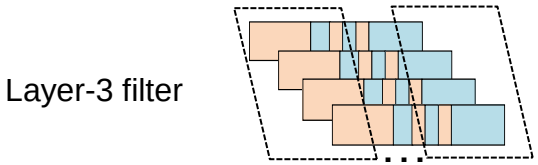
Concatenation by columns





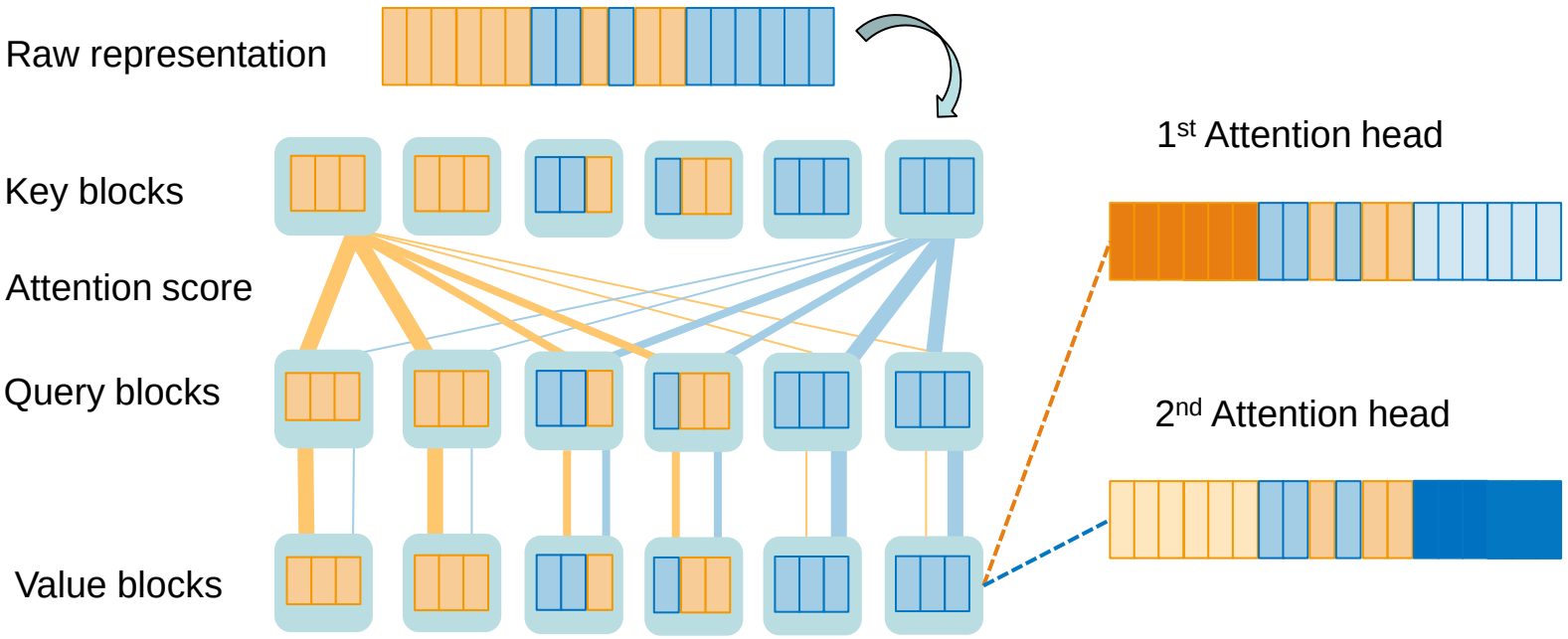






Tab-aware representation

Block Division & Attention-based Profiling



- **A truth:** blocks of **same** page tab have **strong** relations (thus having high attention scores) since they always appear in the same packet trace
- Block division optimizes the attention mechanism's granularity & reduces the model size
- Each attention head produces a **representation variety** for its corresponding page tab

Evaluation

Evaluation Questions

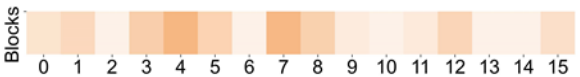
- Is the model stable when the overlapping area **expands** ?
- Can the model deal with **more page tabs** ?
- How well does the block division or attention mechanism **play their roles** ?
- How well does the model perform under the **open world** scenario ?

Comparison Methods

- PSP-WF algorithm [1]
- PSE-WF algorithm [2]
- Multi-DF algorithm (modify Deep Fingerprinting [3] into a multi-label classifier)

[1] A Multi-tab Website Fingerprinting Attack, ACSAC'18
[2] Revisiting Assumptions for Website Fingerprinting Attacks, AsiaCCS'19
[3] Deep Fingerprinting: Undermining Website Fingerprinting Defenses with Deep Learning, CCS'18

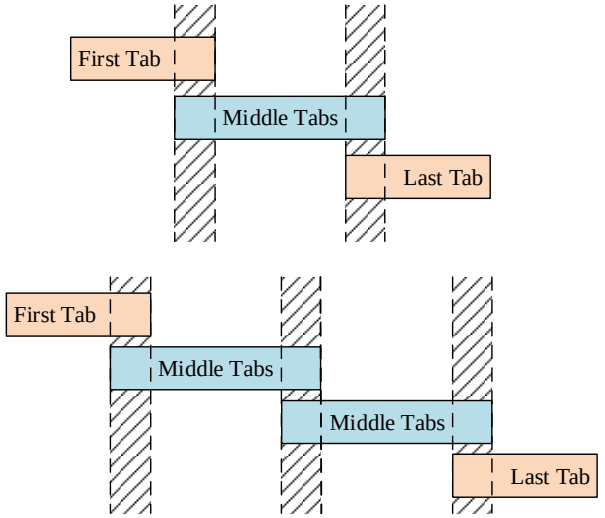
Evaluation

Overlapping proportion		10%			20%			30%			40%			50%		
Metrics		Acc	Pre	Rec	Acc	Pre	Rec	Acc	Pre	Rec	Acc	Pre	Rec	Acc	Pre	Rec
1st Page	**BAPM**	**85.1**	**86.1**	**85.0**	**86.2**	**86.8**	**85.9**	**84.7**	**85.1**	**84.5**	**86.4**	**86.8**	**86.4**	**83.2**	**84.0**	**82.9**
							.6	64.4	68.6	63.0	54.4	60.9	52.0	42.5	51.0	40.5
							.6	38.4	56.4	37.4	36.4	54.2	35.7	37.6	56.6	36.5
							.9	50.5	50.2	52.0	51.9	51.4	52.7	48.7	48.4	49.5
2nd Page	**BAPM**	**77.9**	**78.7**	**77.9**	**78.9**	**79.3**	**78.8**	**78.0**	**78.4**	**77.2**	**75.0**	**74.6**	**74.4**	**71.9**	**72.5**	**71.3**
							.2	19.2	29.0	18.5	15.7	24.5	15.0	15.4	23.4	14.9
							.0	31.9	49.2	30.6	30.6	48.0	29.5	29.3	48.1	28.0
							.2	39.1	41.2	38.7	41.5	40.9	40.3	36.0	37.2	37.8

- BAPM achieves best results with **two page tabs** and **five overlapping proportions**
- PSP-WF becomes worse when
  - ① the overlapping area expands — the single-tab part can be used is shorter (**Lost**)
  - ② fingerprinting the 2nd page tab — the critical former part is dropped (**Lost**)
- PSE-WF and Multi-DF are not ideal all the time, since
  - ① PSE-WF is negatively influenced by uncertain sections (**Confusion**)
  - ② Multi-DF treats the overlapping and non-overlapping area equally (**Confusion**)

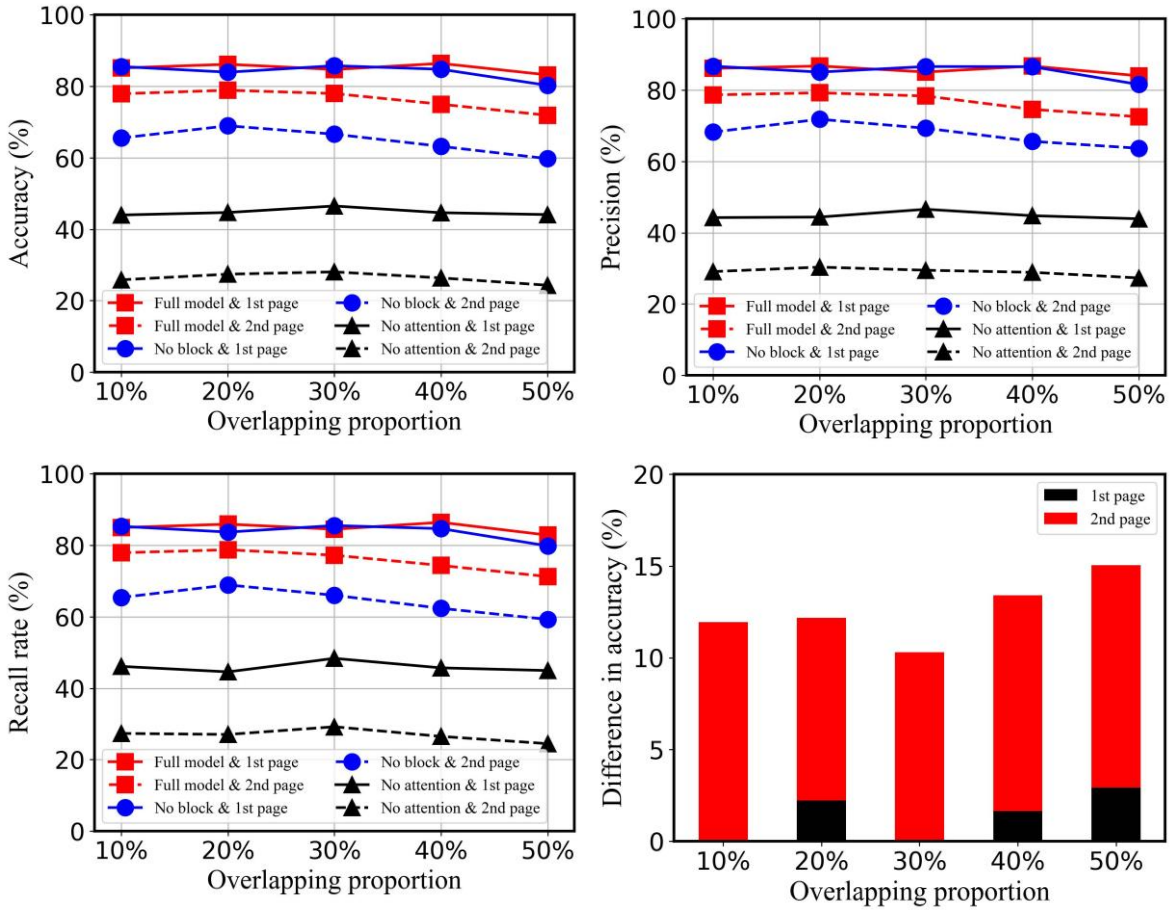
Evaluation – Three Page Tabs

		1st page	2nd page	3rd page
BAPM	Acc	**91.3**	**63.5**	**74.7**
	Pre	**91.2**	**66.8**	**77.3**
	Rec	**91.1**	**63.3**	**74.6**
PSP-WF	Acc	30.9	30.6	4.0
	Pre	36.2	34.4	3.7
	Rec	31.6	31.5	3.9
PSE-WF	Acc	23.0	15.6	18.9
	Pre	45.8	34.4	39.5
	Rec	22.0	15.1	17.6
Multi-DF	Acc	35.1	32.8	24.4
	Pre	36.6	32.7	25.0
	Rec	36.7	28.7	50.0



- Attacking Effect: **1st page tab** (has a complete former part) > **3rd page tab** > **2nd page tab** (has a maximal overlapping area)
- 3-tab is **typical**: 3-tab, 4-tab and more-tab traces all have same three types of tabs: First tab / Middle tab(s) / Last tab
- BAPM can be extend to more tabs through adjusting **attention head numbers**

Evaluation – Ablation Analysis



Valid lines – 1st Page

Dotted lines – 2nd Page

Red – Full Model

Blue – no block division

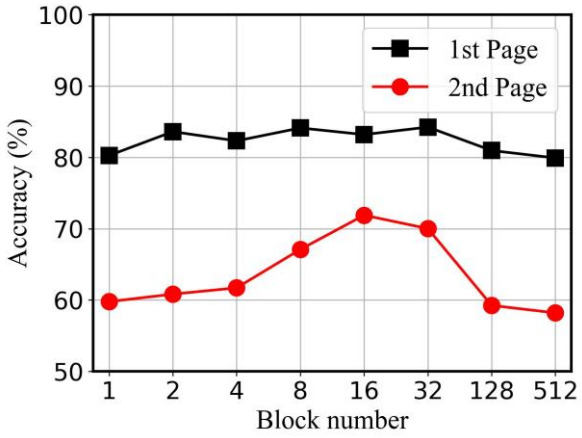
Black – no attention

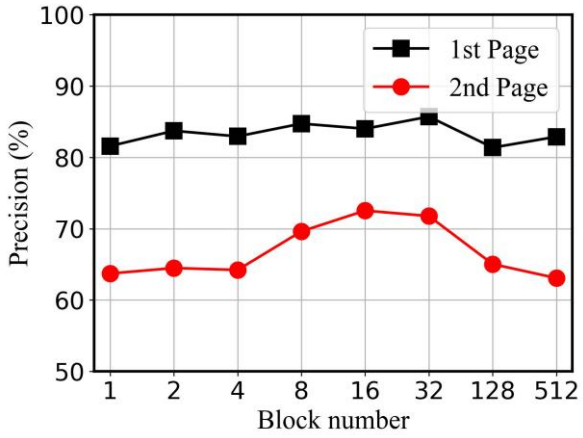
Black area – improvement of 1st page after block division

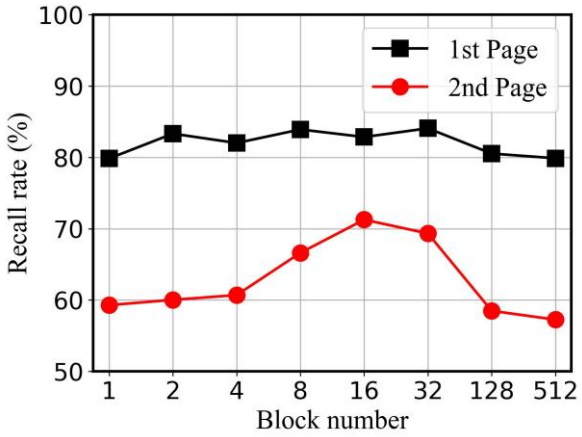
Red area – improvement of 2nd page after block division

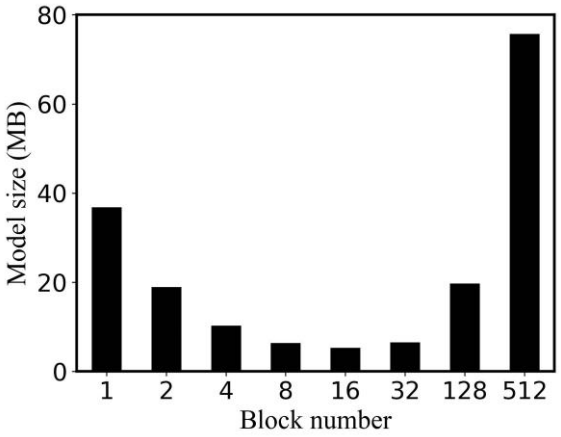
- Block division & attention mechanism helps to exploit the overlapping area, which is more critical for the 2nd page tab

Evaluation – Sensitivity of Block Number









4 blocks – 82% / 62% / 10.3MB

8 blocks – 84% / 67% / 6.3MB

16 blocks – 83% / 72% / 5.3MB

32 blocks – 84% / 70% / 6.5MB

128 blocks – 81% / 59% / 19.7MB

Evaluation – Open World Setting

		1st page			2nd page		
		2000	4000	6000	2000	4000	6000
BAPM	Acc	**82.3**	**79.2**	**77.9**	**65.9**	**65.7**	**58.6**
	Pre	**79.1**	**77.3**	**78.0**	**63.7**	**64.3**	**65.5**
	Rec	**83.1**	**83.5**	**83.9**	**65.8**	**67.0**	**65.4**
PSP-WF	Acc	55.2	50.9	41.5	31.1	32.3	29.5
	Pre	58.0	64.6	56.4	2.5	7.5	3.4
	Rec	25.7	18.7	13.2	2.3	2.4	2.5
PSE-WF	Acc	49.5	46.4	37.5	44.6	42.3	34.8
	Pre	51.7	62.1	65.7	40.4	37.7	40.9
	Rec	22.2	11.6	7.4	17.0	6.2	4.3
Multi-DF	Acc	32.1	38.2	36.0	25.4	30.3	27.6
	Pre	11.4	16.3	21.9	60.7	66.5	52.0
	Rec	49.7	54.6	50.2	26.0	31.2	27.2

- Three scales: 2000 / 4000 / 6000 unmonitored websites & 50 monitored websites
- Predict whether X is monitored or not
- Predict X's specific class if X is monitored

- PSP-WF breaks — difficult to determine **changing features** of splitting points
- PSE-WF breaks — unmonitored websites are quite possible to have similar direction patterns with a monitored one in **section level**
- BAPM's processing mode is more suitable for open world multi-tab WF

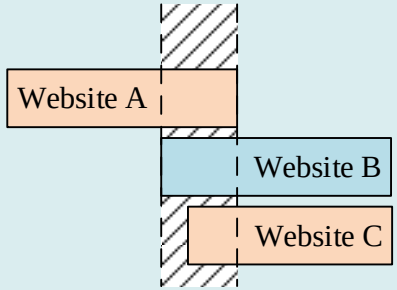
Defenses Discussion

What the experiments reveal ——

The online privacy can be threatened even under multi-tab scenarios

To avoid this risk, some good practices are ——

- **Setting off page tabs simultaneously**
  - 3 tabs in overlapping area
    —— drop to 30% for the 2nd tab
  - 4 tabs in overlapping area
    —— drop as random decision for the 2nd tab
- **Keeping non-browsing behaviors**
  - Download a file / Cache a video / Listen to music
- **Adopting proposed countermeasures**



Add noise to destroy original pattern

Conclusion and Q&A

- **We have proposed BAPM**
  - ① a novel attention-based end-to-end multi-tab WF model
  - ② simplify the process of multi-tab WF attacks
  - ③ address information lost & confusion problems
- **We show the effectiveness of BAPM**
  - ① achieve best results with a large overlapping area up to 50%
  - ② remain valid under more page tabs or open world settings
- **Taking some defenses to avoid WF attacks is also necessary under multi-tab scenarios**