CE 815 – Secure Software Systems

Causal Analysis (Evasion attack)

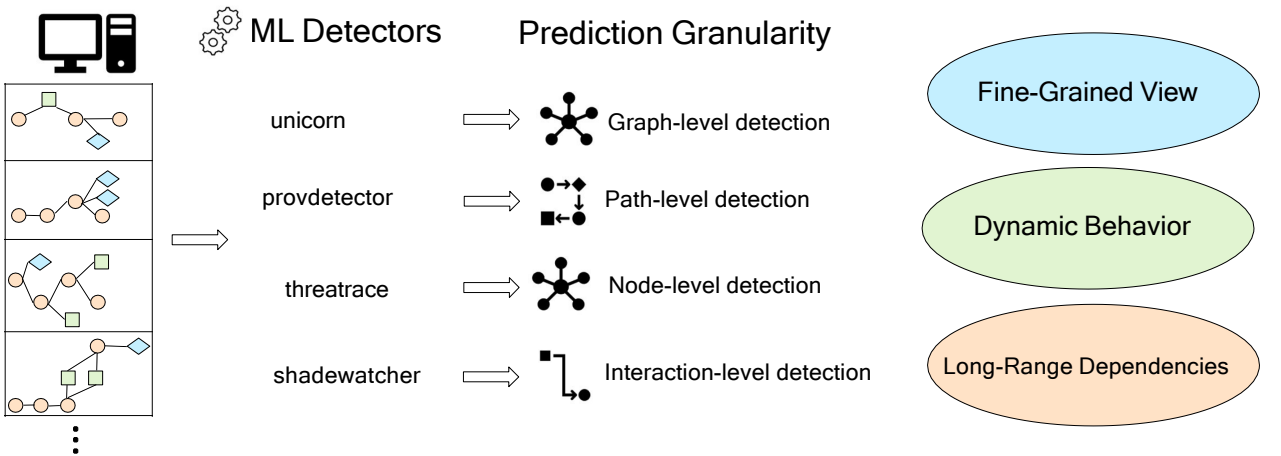
Zahra Fazli/Mehdi Kharrazi
Department of Computer Engineering
Sharif University of Technology

Acknowledgments: Some of the slides are fully or partially obtained from other sources. A reference is noted on the bottom of each slide, when the content is fully obtained from another source. Otherwise a full list of references is provided on the last slide. Thanks to Zahra Fazli for the help on the slides.

Content

- Provenance-Based IDS
  - Edge-based
  - Node-based
  - Path-based
  - Graph-based
- Evasion
  - Traditional IDS
  - Provenance-Based IDS
  - Prov-ninja
  - Discussion Holmes

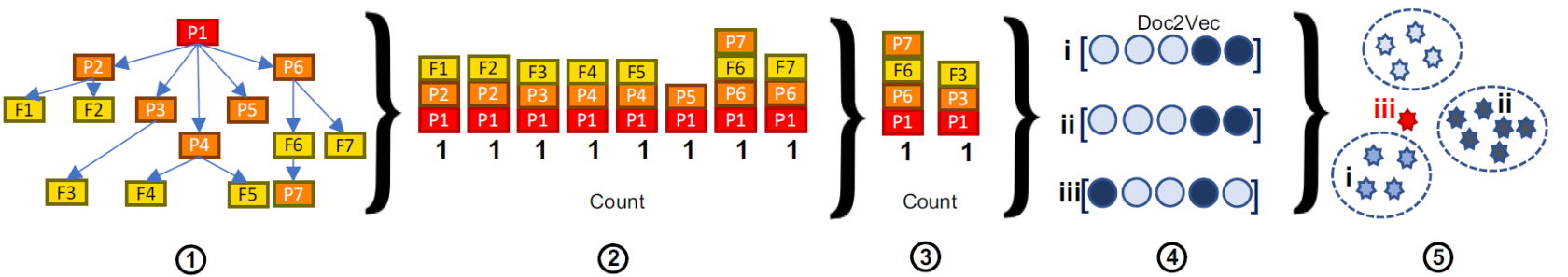
Provenance-Based IDS



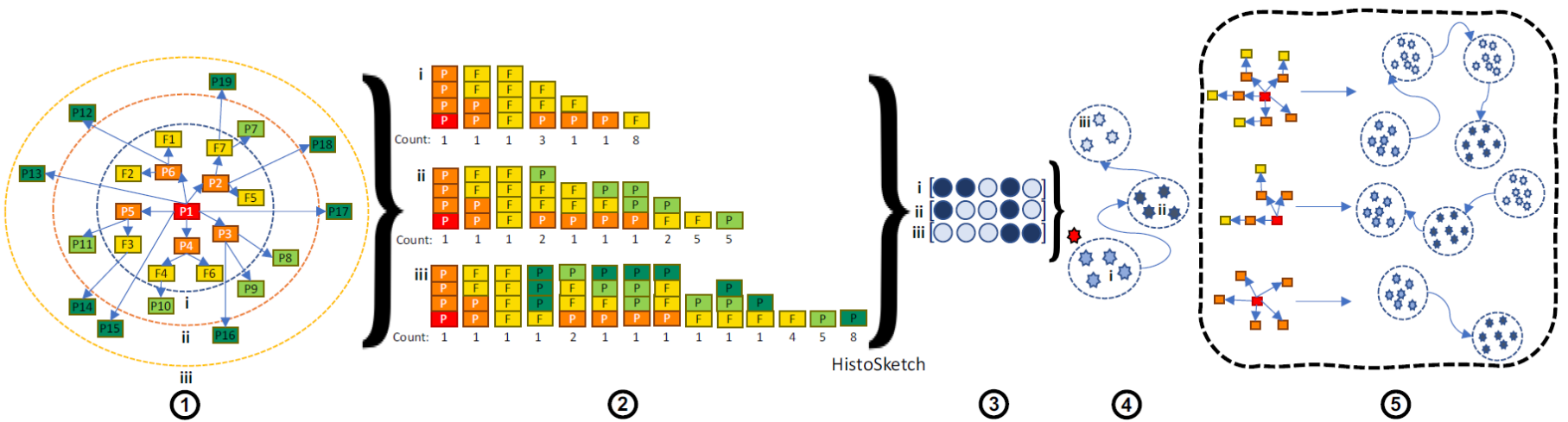
Why are Provenance-based IDS gaining popularity?

Provdetector

- Lossy
- limit their effectiveness against advanced malicious actors



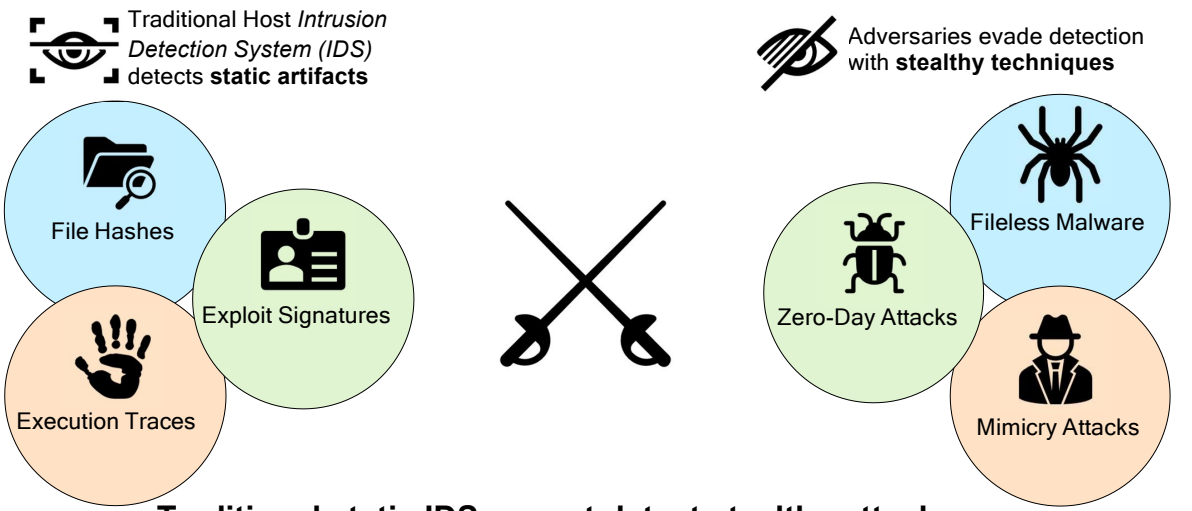
Unicorn


P19
P12
P7
F1
F7
P18
P13
F2
P6
P2
F5
P5
P1
P17
P11
F3
P4
P3
P8
F4
F6
P9
P14
P10
i
P15
ii
P16
iii
Count: 1 1 1 3 1 1 8
Count: 1 1 1 2 1 1 1 2 5 5
Count: 1 1 1 1 2 1 1 1 1 1 1 4 5 8
HistoSketch
i
ii
iii
①
②
③
④
⑤

Content

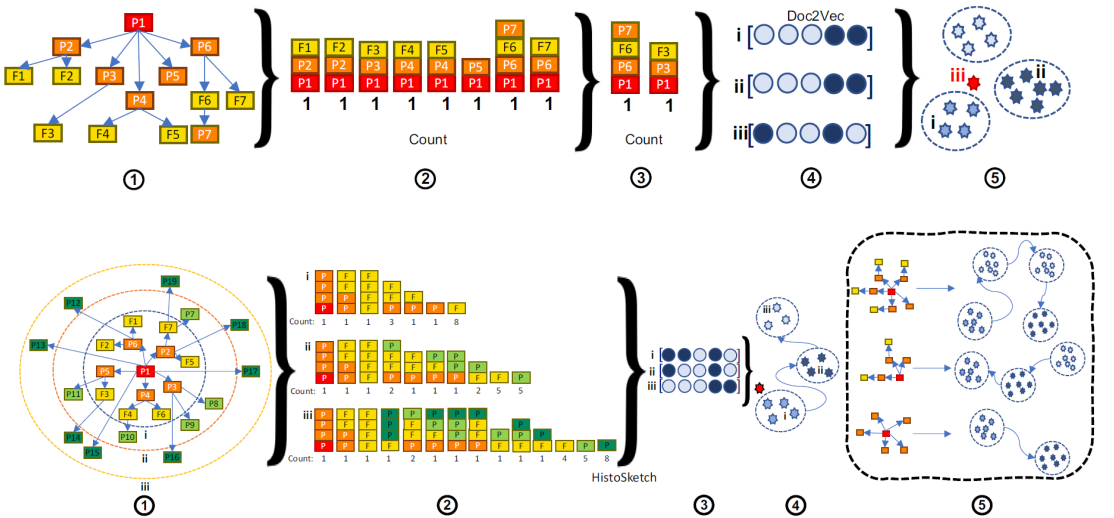
- Provenance-Based IDS
  - Edge-based
  - Node-based
  - Path-based
  - Graph-based
- Evasion
  - Traditional IDS
  - Provenance-Based IDS
  - Prov-ninja
  - Discussion Holmes

Evasion Attacks against Static Host Defenses



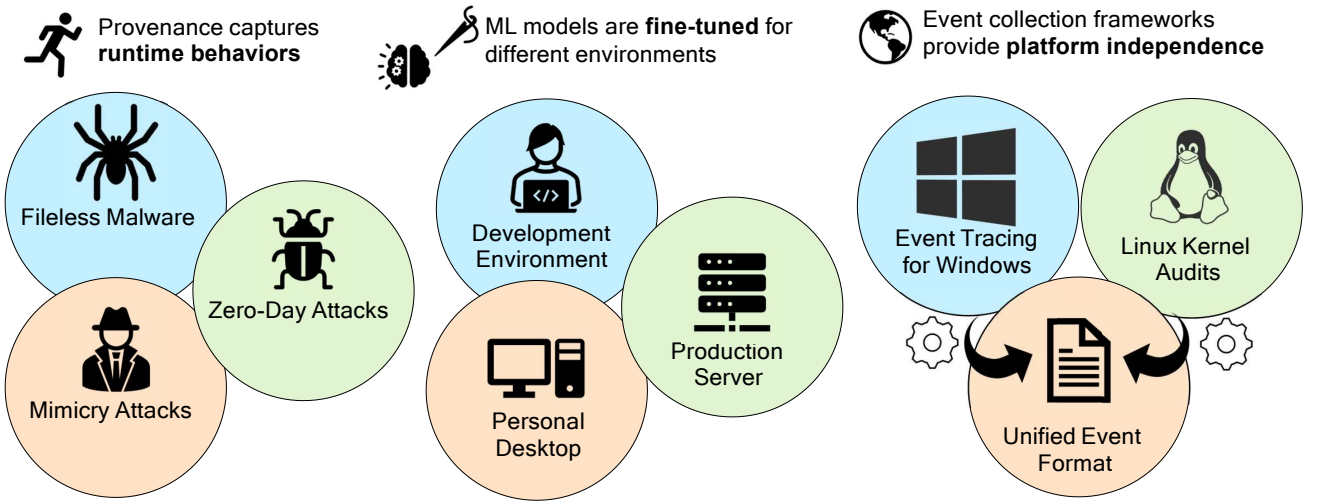
Evasion Attacks against Provenance-Based IDS

- Mimicry Attacks



Evading Provenance-Based ML Detectors with Adversarial System Actions, Kunal Mukherjee, Joshua Wiedemeier, Tianhao Wang, James Wei, Feng Chen, Muhyun Kim, Murat Kantarcioglu, and Kangkook Jee, USENIX Security 2023.

Popularity of Provenance-Based IDS



However, Provenance-based IDS are not yet mature.

Primary Roadblock to Provenance-Based IDS Adoption

Trust in Provenance-based IDS has **not been established**

Robustness against dedicated adversaries has **not been verified**

Adversarial validation is an established way to **prove robustness**

Adversarial Validation in Provenance-Based IDS

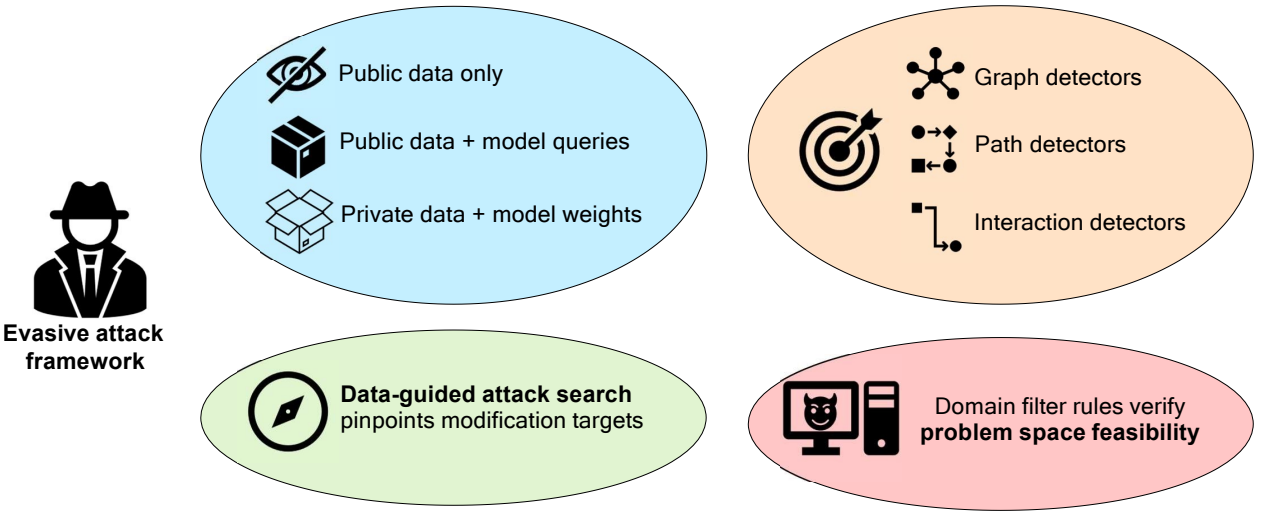
Generic adversarial techniques fail
- Heterogenous graphs with node/edge attributes

Problem space feasibility is critical for validation
- Only real-world attacks can invalidate defenses

Provenance mimicry attacks exist [Goyal], *however*

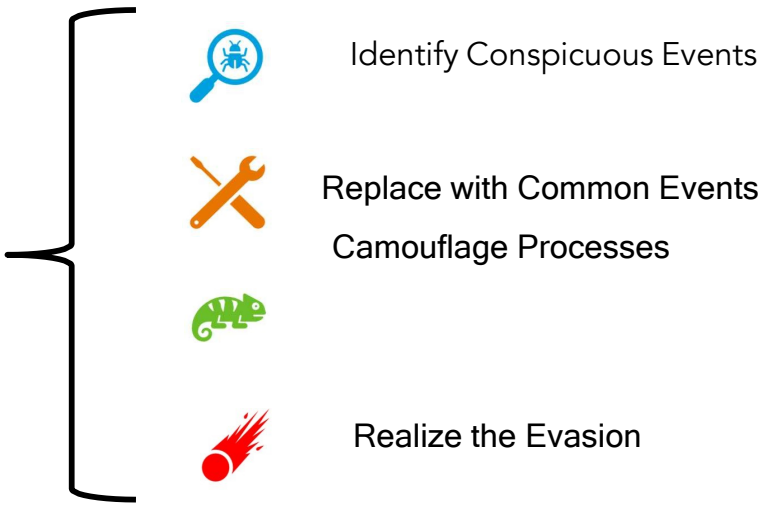
- **Require adding >15,000 events**
- **Require knowledge of the defense model architecture**
- **Unlikely to be effective against event-level detectors**

Contributions

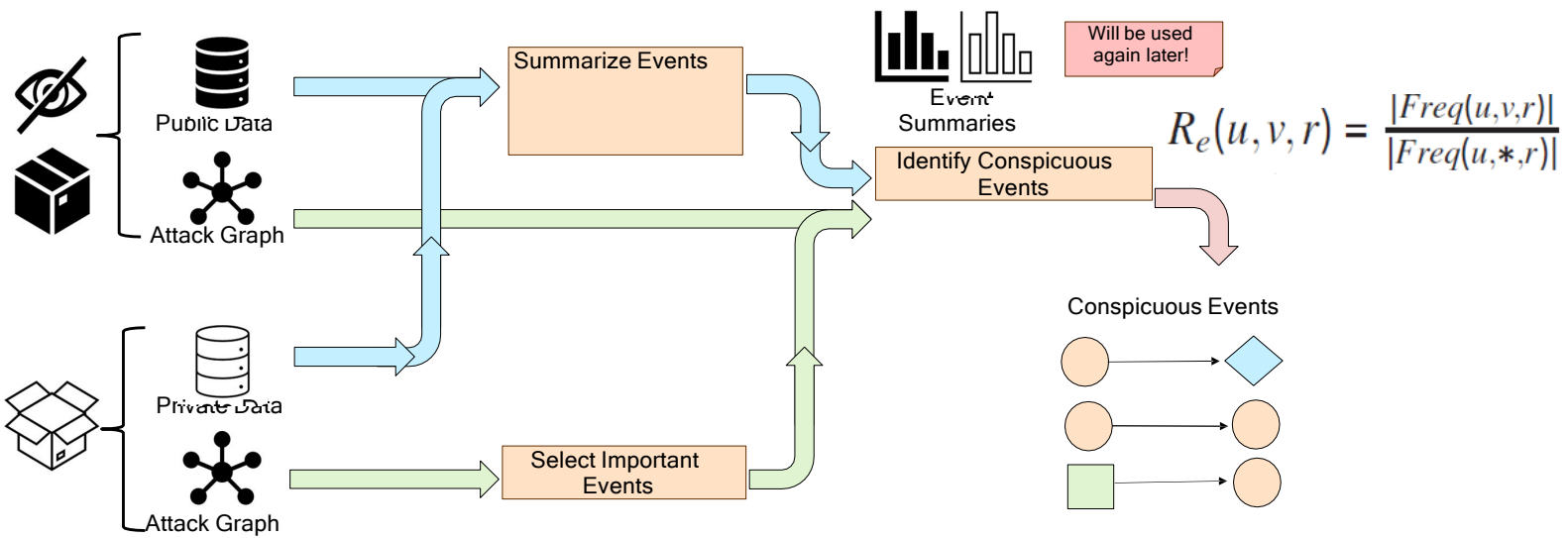


ProvNinja

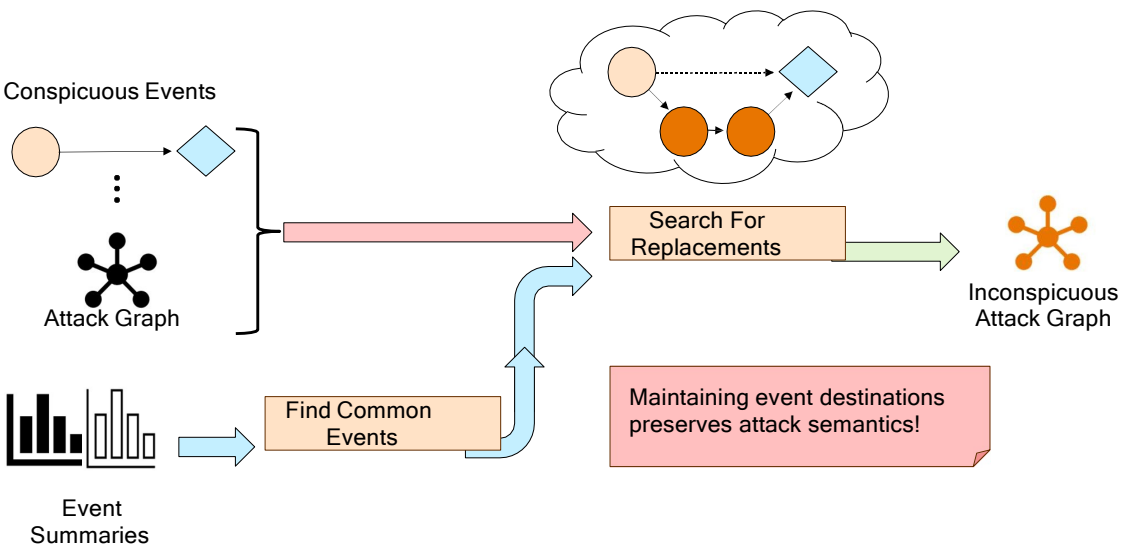
ProvNinja: Evasive Attack Framework



Identify Conspicuous Events



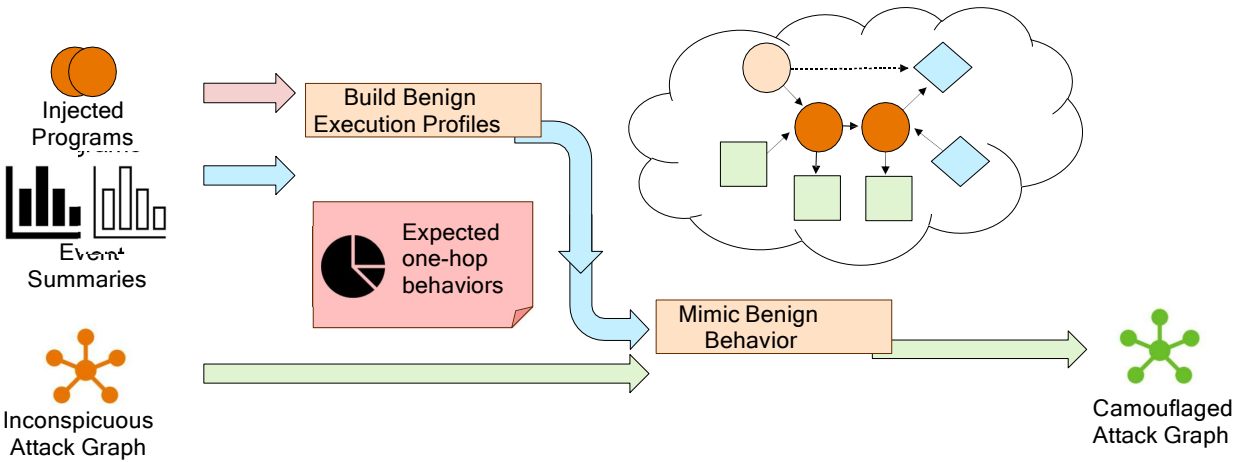
Replace with Common Events



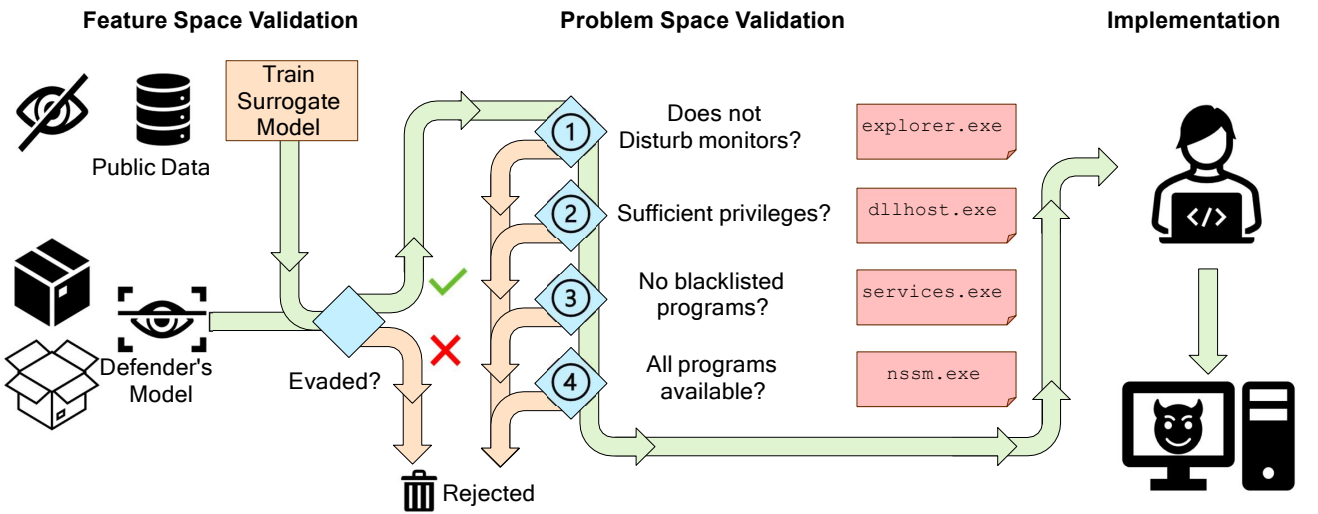
Replace with Common Events

Index	Gadgets (Gadget Length)	Regularity Score	Rejection Rule
	`firefox.exe` – (Gadgets) → `notepad.exe`		
1	`svchost.exe` → **wininit.exe** → **winlogon.exe** → **userinit.exe** → `explorer.exe` (5)	2.8	Special Sequence
2	`svchost.exe` → **cmd.exe** → **shellexperiencehost.exe** (3)	8.3	Display Irregularities
3	**nssm.exe** → `python.exe` → `conhost.exe` → `wininit.exe` → `explorer.exe` (5)	4.39	Program Unavailability
4	**conhost.exe** → `werfault.exe` → `explorer.exe` (3)	8.1	Insufficient Privilege
5	`svchost.exe` → **schtasks.exe** → `conhost.exe` → `explorer.exe` (4)	7.9	Scheduling Tasks
6	`svchost.exe` → **rundll32.exe** → `winsat.exe` → `explorer.exe` (4)	9.1	Writing to Registries
7	**tvnserver.exe** → `mpcmdrun.exe` → `conhost.exe` → `explorer.exe` (4)	3.3	External Network Connections
8	**sshd.exe** → **ssh-shellhost.exe** → `explorer.exe` (3)	7.5	User Interactions
9	`sshd.exe` → **mpcmdrun.exe** → `conhost.exe` → `winword.exe` → `werfault.exe` → `explorer.exe` (6)	7.9	Singleton Programs
10	`services.exe` → **taskhostw.exe** → `ngentask.exe` → `ngen.exe` → `svchost.exe` → `explorer.exe` (6)	4.2	Special Protocol Support
11	`svchost.exe` → `werfault.exe` → `explorer.exe` (3)	9.5	-

Camouflage Processes



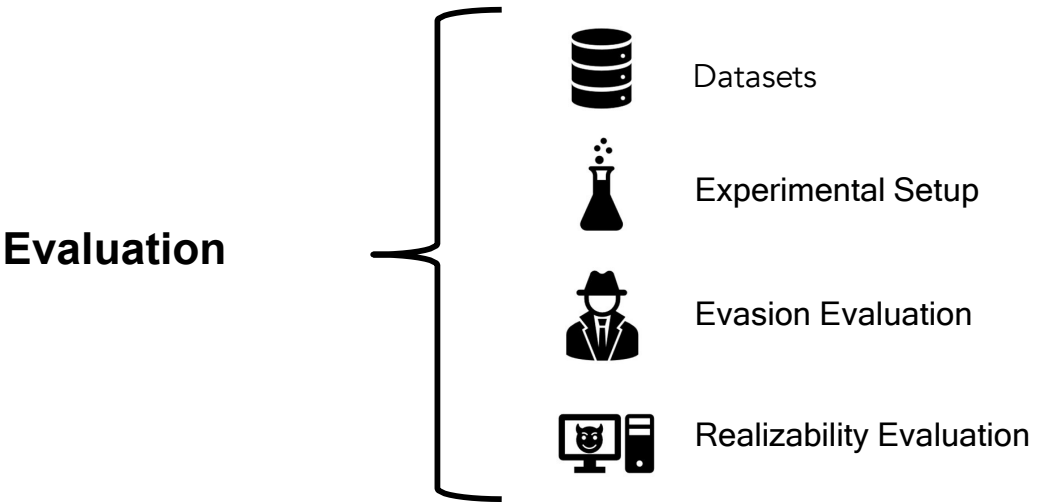
Realize the Evasion



Problem Space Constraints

- Available transformations
  - discrepancies in available gadgets may occur
  - actively prefer system programs
- Preserving attack semantics
  - manually choose candidate system actions
- Robustness to pre-processing
  - data reduction: lossy graph compression approaches
- Plausibility to users and security analysts

Evaluation

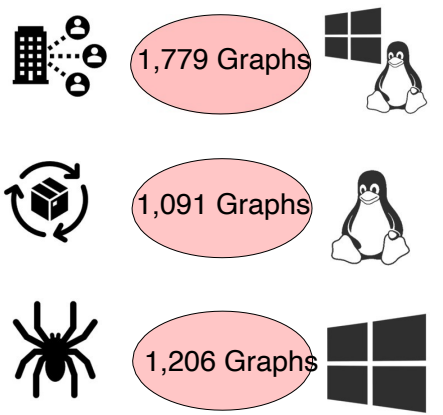

Evaluation
Datasets
Experimental Setup
Evasion Evaluation
Realizability Evaluation

Datasets

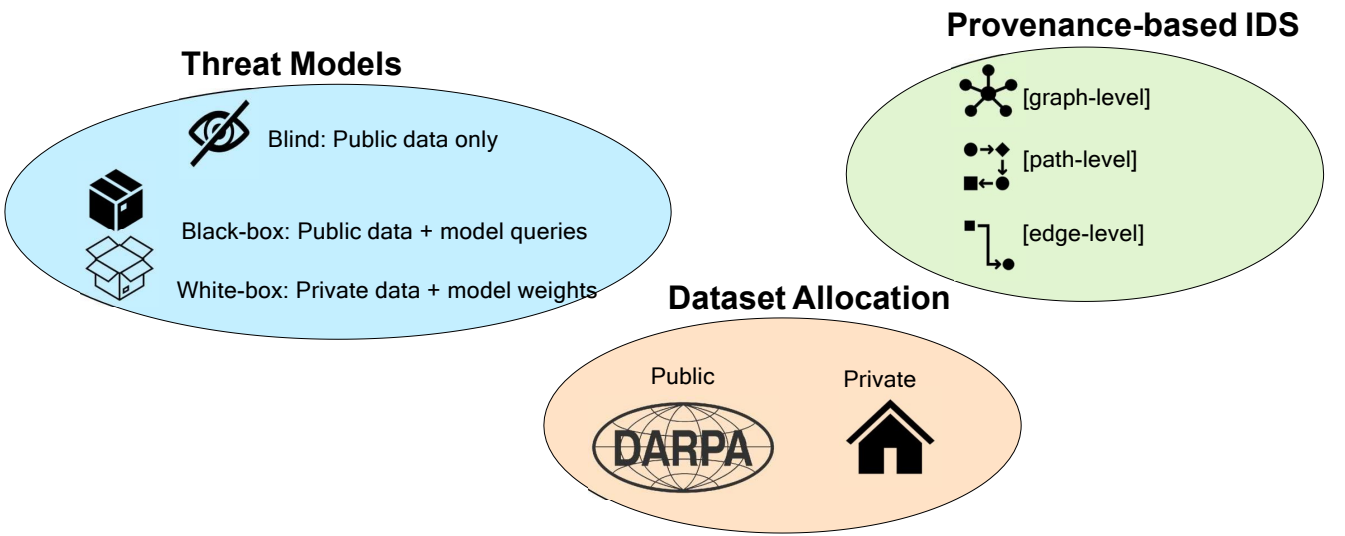
Benign Datasets

	DARPA (public)	In-House (private)
	Scripted	Real Users
	8 Hosts	86 Hosts
	12 Days	13 Months

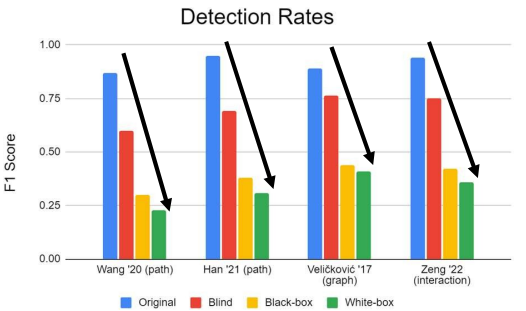
Malicious Datasets

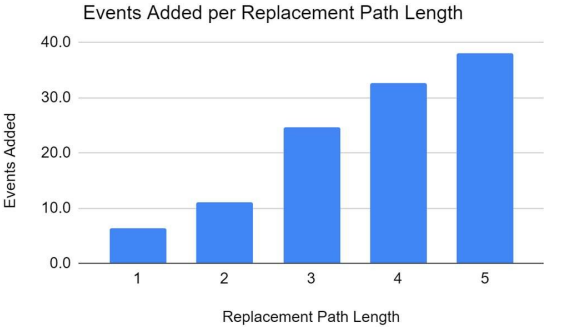


Experimental Setup



Evasion Evaluation





Reduces detection rates against SOTA Provenance-based IDS

Each replacement adds fewer than 40 events

Evasion Evaluation

edge-level detection : capability to counter robust provenance-based ML detectors

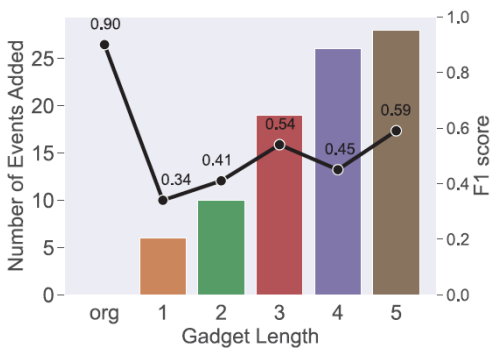
Table 3: PROVNINJA evasion for ShadeWatcher [32].

Attack Type	ShadeWatcher		Random Perturb.		PROVNINJA	
	Recall	F1	Recall	F1	Recall	F1
Enterprise APT	0.96	0.93	0.98*(+.02)*	0.98*(+.05)*	0.45*(-.51)*	0.41*(-.52)*
Supply Chain APT	0.92	0.90	0.96*(+.04)*	0.97*(+.07)*	0.38*(-.54)*	0.40*(-.50)*
Average	**0.94**	**0.92**	**0.97***(+.03)*	**0.98***(+.06)*	**0.42***(-.53)*	**0.41***(-.51)*

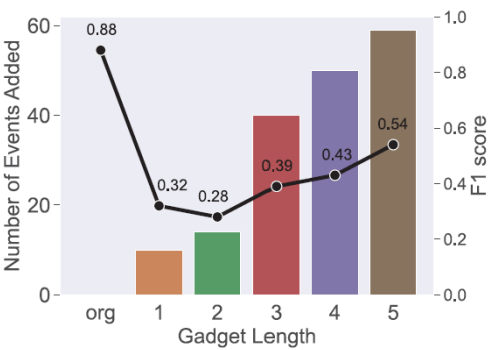
Evasion Evaluation

Defense Model	White-box (PROVNINJA)		Black-box (PROVNINJA)		Blind (PROVNINJA)		Blind (Random Pert.)	
	Recall	F1	Recall	F1	Recall	F1	Recall	F1
ProvDetector	0.23	0.27	0.30 *(+.07)*	0.35 *(+.08)*	0.60 *(+.37)*	0.67 *(+.40)*	0.89 *(+.66)*	0.91 *(+.64)*
SIGL	0.31	0.35	0.38 *(+.07)*	0.47 *(+.12)*	0.69 *(+.38)*	0.74 *(+.39)*	0.97 *(+.66)*	0.95 *(+.60)*
S-GAT	0.38	0.41	0.42*(+.04)*	0.51 *(+.10)*	0.75 *(+.37)*	0.77 *(+.36)*	0.91 *(+.53)*	0.93 *(+.52)*
Prov-GAT	0.44	0.47	0.51 *(+.07)*	0.61 *(+.14)*	0.78 *(+.34)*	0.80 *(+.33)*	0.96 *(+.52)*	0.97 *(+.50)*
ShadeWatcher	0.36	0.33	0.42 *(+.06)*	0.41 *(+.08)*	0.75 *(+.39)*	0.72 *(+.39)*	0.97 *(+.61)*	0.97 *(+.64)*
Average	**0.34**	**0.37**	**0.41** *(+.06)*	**0.47** *(+.10)*	**0.71** *(+.37)*	**0.74** *(+.37)*	**0.94** *(+.60)*	**0.95** *(+.58)*

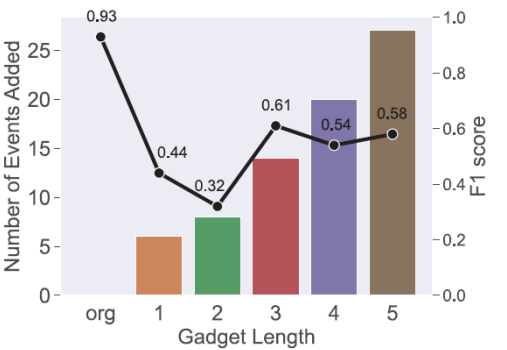
Evasion Evaluation



(a) Enterprise APT.

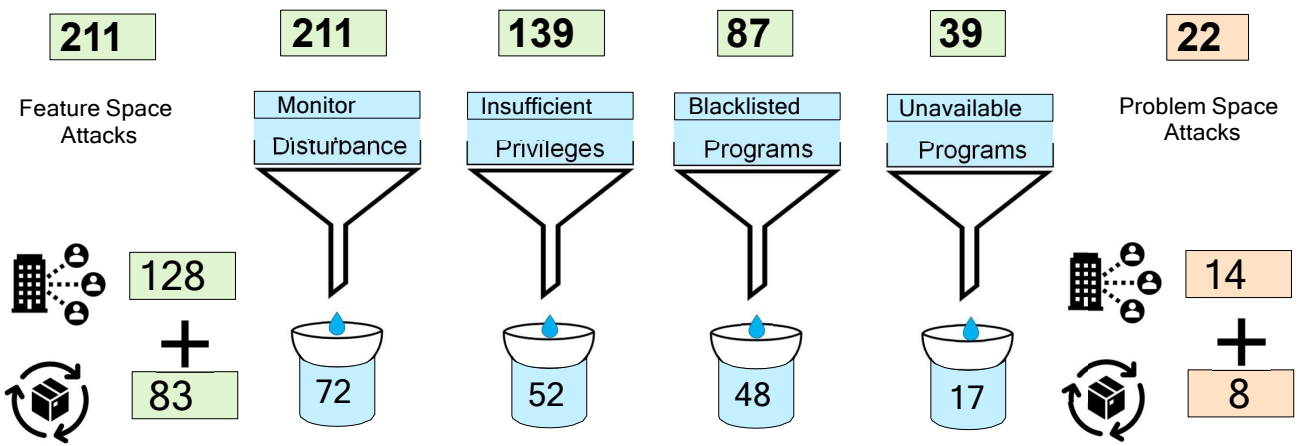


(b) Supply Chain APT.



(c) Fileless Malware.

Attack Realizability



Conclusion

ProvNinja **systematically challenges** Provenance-based IDS

Inspiring the development of **robust** IDS with **realistic** adversarial examples

Holmes

- What do you think about Holmes?

Acknowledgments

- [Prov-ninja] Kunal Mukherjee , et al. “Evading Provenance-Based ML Detectors with Adversarial System Actions” – usenix 2024
- [unicorn] Xueyuan Han , et al. “Unicorn: Runtime provenance-based detector for advanced persistent threats”- ndss 2020
- [Threatrace]S. Wang, Z. Wang, , et al. “Threatrace:Detecting and tracing host-based threats in node level through provenance graph learning,” IEEETransactions on Information Forensics and Security, 2022
- [Shadewatcher]- Zeng, Jun, et al. "Shadewatcher: Recommendation-guided cyber threat analysis using system audit records." 2022 IEEE Symposium on Security and Privacy (SP). IEEE, 2022.
- [provedetector]- Wang, Qi, et al. "You Are What You Do: Hunting Stealthy Malware via Data Provenance Analysis." NDSS. 2020.
- [Goyal] - Goyal, Akul, et al. "Sometimes, You Aren’t What You Do: Mimicry Attacks against Provenance Graph Host Intrusion Detection Systems." 30th ISOC Network and Distributed System Security Symposium (NDSS’23), San Diego, CA, USA. 2023.