Stony Brook University

SAMSUNG RESEARCH AMERICA

Proteus: A Practical Framework for Privacy-Preserving Device Logs

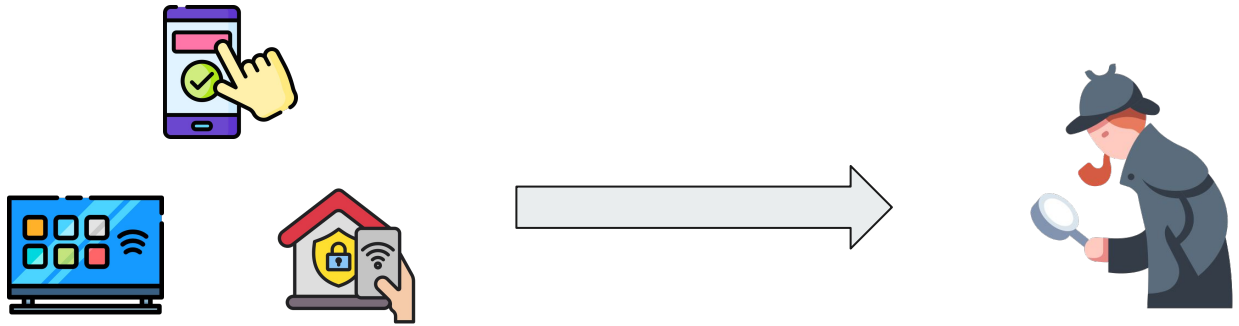
Sanket Goutam, Hunter Kippen, Michael Grace, Amir Rahmati

Stony Brook University, Samsung Research America

The 26th Privacy Enhancing Technologies Symposium (PETS 2026)
July 20–25, 2026

Mobile devices are now audit-critical systems.

- Considered a critical security endpoint that require constant monitoring.
- Audit logging is essential for post-hoc security measures.
- Regulatory frameworks require audit trails and forensic visibility from end devices.



Problem : PII Leakage in Device Logs.

PII type	Devices				Manufacturers		Android Version					
	Initial Log	Final Log	Detected	Prevalence	Detected	Prevalence	≤ 7	8	9	10	11	12
Android ID	50	82	92	8%	26 of 45	58%	29%	19%	11%	12%	3%	0%
Bluetooth MAC	85	123	136	11%	27 of 45	60%	44%	14%	21%	9%	10%	6%
Bluetooth Name	724	782	836	69%	40 of 45	89%	81%	55%	82%	64%	74%	61%
Bluetooth Scan MAC	5	26	26	2%	10 of 45	22%	13%	0%	1%	1%	1%	6%
Bluetooth Scan SSID	13	26	26	2%	7 of 45	16%	9%	3%	4%	1%	1%	3%
Coarse Location	246	245	292	24%	14 of 45	31%	15%	14%	25%	23%	19%	42%
Email Address	192	180	198	16%	4 of 45	9%	57%	23%	16%	11%	28%	0%
Fine Location	224	231	272	22%	14 of 45	31%	15%	12%	25%	20%	18%	40%
IMEI	14	10	16	6%	9 of 30	30%	27%	5%	4%	-	-	-
Phone Number	11	12	14	3%	4 of 40	10%	5%	4%	3%	3%	-	-
Serial Number	43	40	49	4%	12 of 45	27%	26%	5%	4%	1%	4%	3%
WiFi Randomized MAC	242	532	539	78%	19 of 20	95%	-	-	-	-	71%	89%
WiFi Router MAC	500	795	814	67%	35 of 45	78%	54%	38%	55%	44%	76%	94%
WiFi Router SSID	597	808	829	68%	35 of 45	78%	60%	63%	61%	51%	75%	83%
WiFi Scan MAC	89	167	175	14%	19 of 45	42%	22%	15%	24%	7%	12%	23%
WiFi Scan SSID	364	462	475	39%	24 of 45	53%	26%	37%	27%	26%	37%	69%
Any PII Detected	998	1108	1142	94%	45 of 45	100%	97%	91%	95%	88%	95%	100%

PII leakage in device logs is prevalent across manufacturers, devices, and OS.

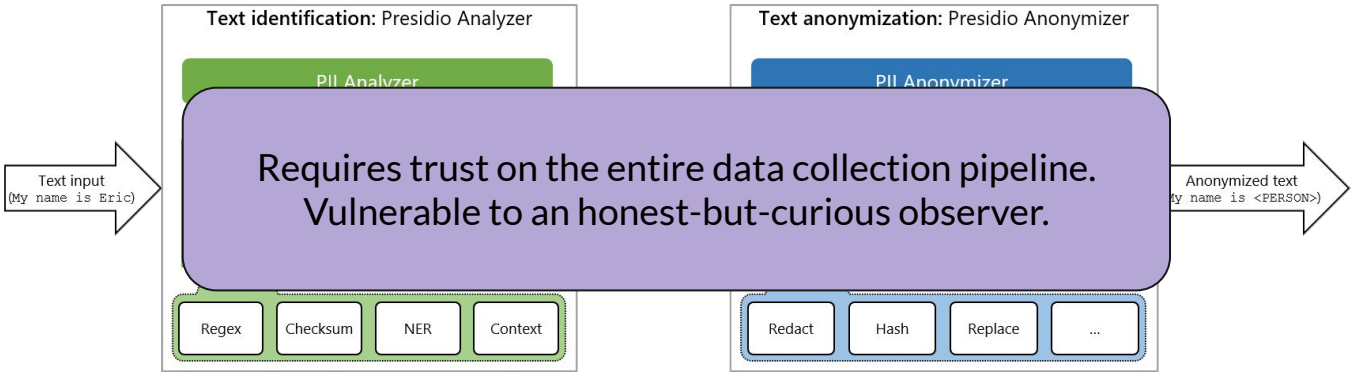
Lyons, Allan, et al. "Log: It's Big, It's Heavy, It's Filled with Personal Data! Measuring the Logging of Sensitive Information in the Android Ecosystem." ***(USENIX Security 23)***

Tradeoff between privacy and utility.

- PII data is present in the logs. *What can we do about it?*
- **Approach 1** : Redact all sensitive data.
- **Approach 2** : Sanitize logs through anonymization or removal.
- *Post hoc* or *a priori*?

Existing approaches : Microsoft Presidio.

- Current SOTA approach most widely used in production.
- Post-hoc model run by data aggregators after exfiltrating data from devices.



Existing approaches : OpenAI Privacy Filter.

- On-device SLM that detects and redacts PII data from logs.
- *a priori* mechanism, protects PII at the source.

Key Insight : Correlation not disclosure.

- Event timeline reconstruction is essential for forensic analysis.
- Redaction prevents data disclosure but breaks log fidelity for correlation.

Simple hashing or Pseudonymization also does not work.

“Harry Potter” → [Person]	Redaction, loses fidelity for correlation
“Harry Potter” → [a3bf450]	Allows correlation, but can break anonymity.

Key Insight : Correlation not disclosure.

- Event timeline reconstruction is essential for forensic analysis.
- Redaction prevents data disclosure but breaks log fidelity for correlation.

“We can

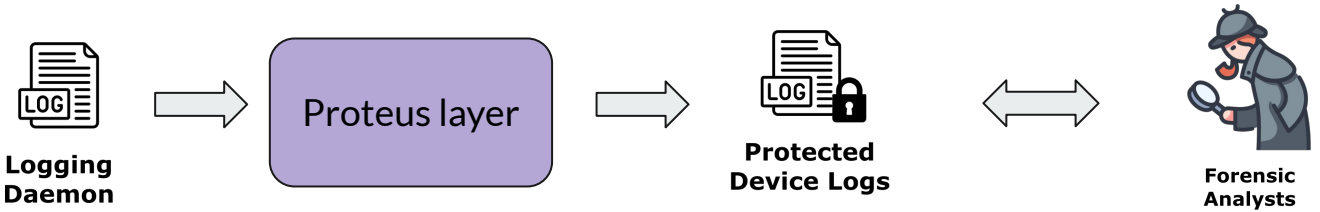
How do we balance disclosure while enabling correlation?

“Harry Potter” → [Person] Redaction, loses fidelity for correlation

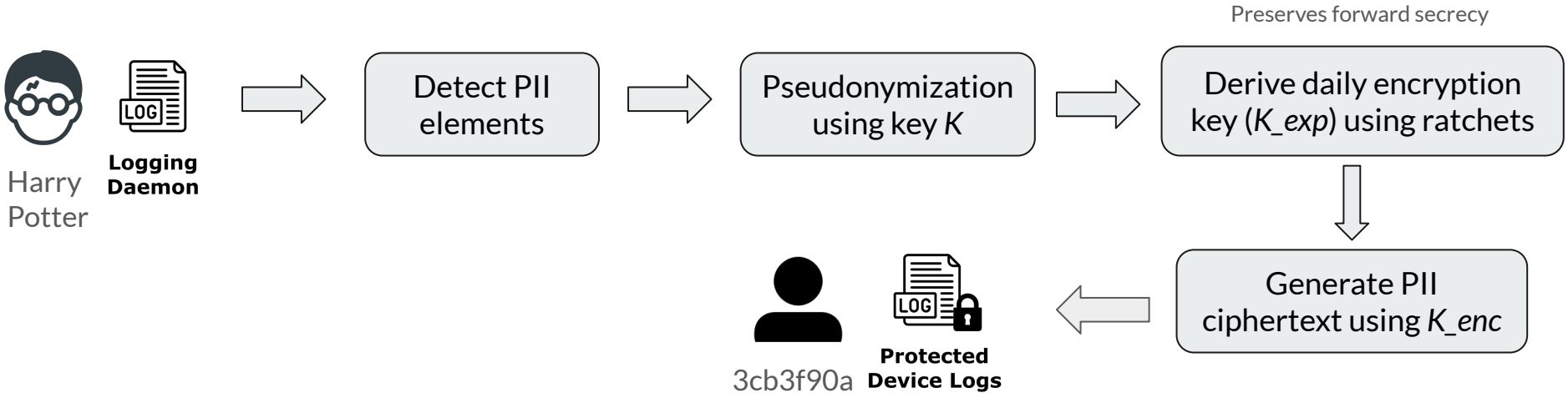
“Harry Potter” → [a3bf450] Allows correlation, but can break anonymity.

Proteus : Privacy-preserving logging framework.

- Transforms sensitive data before it enters the device log.
- Preserves search, debugging, and forensic fidelity of logs.
- Prevents plaintext exposure of PII across log exports.



Proteus : Log processing pipeline.



Logs only contain the ciphertext. Pseudonym is secured locally.

Proteus : Log reconstruction pipeline.

Protected Device Logs

3cb3f90a

Request for controlled disclosure

Provide ratchet state of day *t-n*

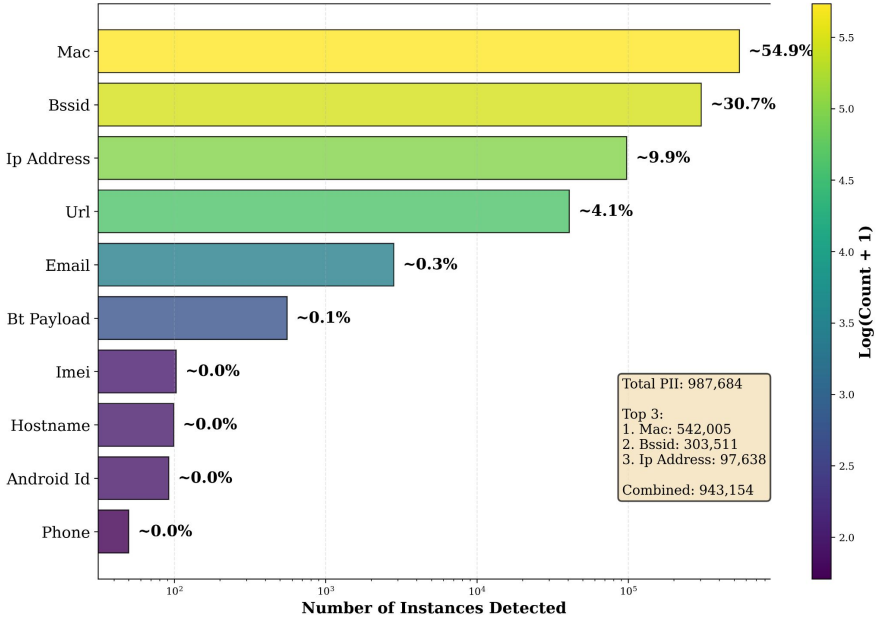
Derive daily decryption key (*K_exp*) for n days from t

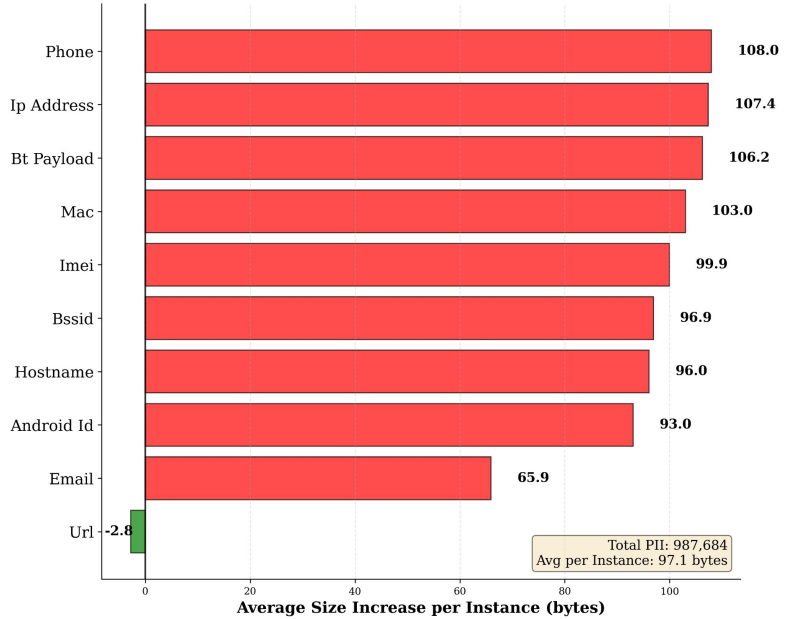
Generate PII pseudonymized text from ciphertext

man-12

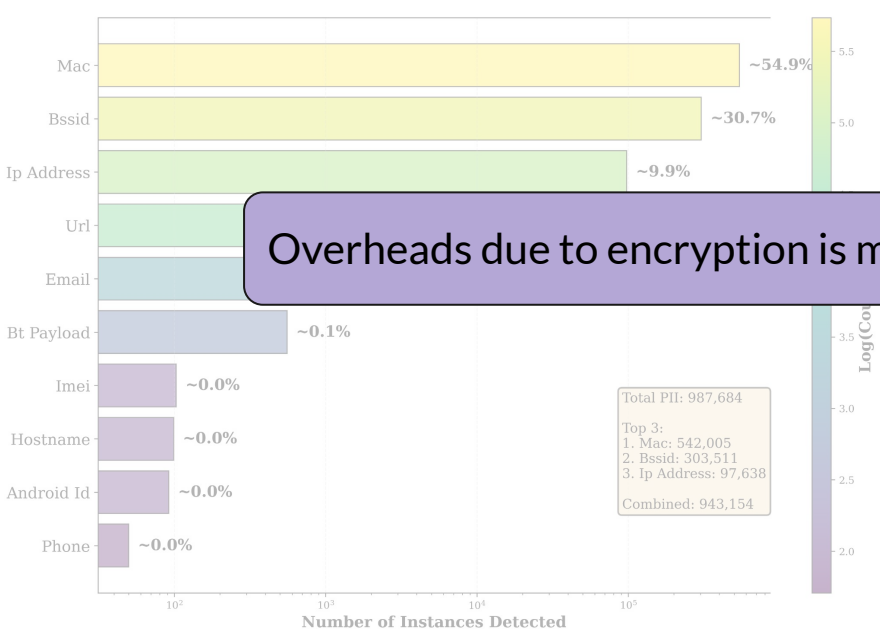
Controlled disclosure reveals the pseudonyms used but not the actual PII.

Evaluation : Bulk processing.



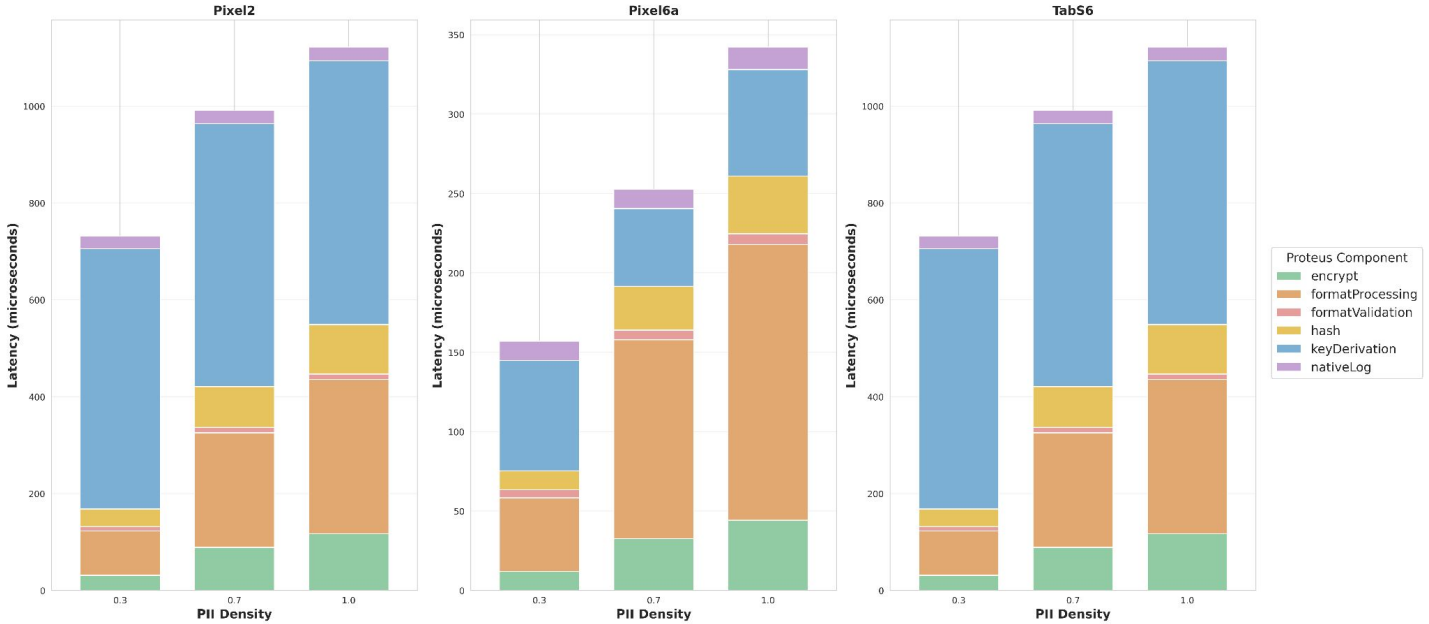


Evaluation : Bulk processing.

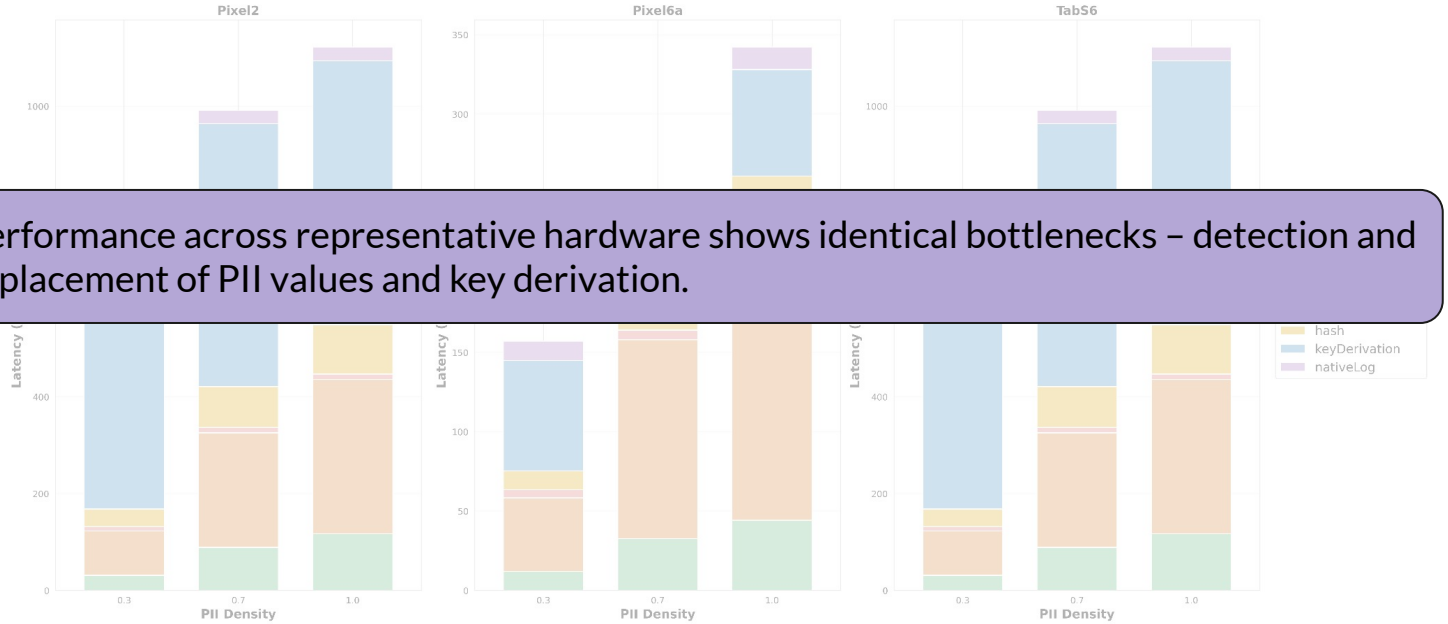


Overheads due to encryption is minimal, and in some cases *optimal*.

Evaluation : On-device measurement.



Evaluation : On-device measurement.



Performance across representative hardware shows identical bottlenecks – detection and replacement of PII values and key derivation.

Stony Brook University

SAMSUNG RESEARCH AMERICA

Takeaways

- Mobile devices are audit-critical systems. Increased logging exposes user privacy risks.
- **Proteus** presents a practical framework for protecting sensitive data in logs, while preserving full log fidelity.

Sanket Goutam, Ph.D.
Security Researcher,
Samsung Research America

sgoutam.github.io