Correlating GSM and 802.11 Hardware Identifiers

LCDR Jeremy Martin, LT Danny Rhame, Dr. Robert Beverly, and Dr. John McEachen

Naval Postgraduate School

Correlating GSM and 802.11 Hardware Identifiers

- Determine the feasibility of cross-protocol association of GSM and WiFi identifiers from the same device
- Examine the breadth of protocol layers of each communication medium
- Use temporal and spatial analysis

Correlating GSM and 802.11 Hardware Identifiers

- Motivation
- Previous Work
- Background
- Methodology and Data Collection
- Correlation
- Results
- Future / Continued Work

Motivation

- Hardware identifiers are globally unique and do not change over the lifetime of a device – allows for both tracking and association of a physical device
  - Targeted advertising and statistics gathering [1]
  - Threat of increased attack vectors [2, 3, 4]
  - Use as search and rescue capability
  - Law enforcement and forensic analysis

Previous Work

- Privacy leak analysis of smartphones [1, 3, 6, 7, 8, 9]
  - Utilize identified security leaks for cross-correlation
- Constellation analysis of RF devices [5]
  - Our analysis demonstrates the feasibility of using constellations for cross-correlation

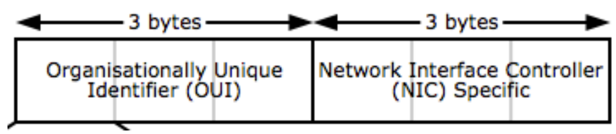
Background

- The format, structure, and governing allocation authorities of GSM and 802.11 addresses are different and do not facilitate trivial association

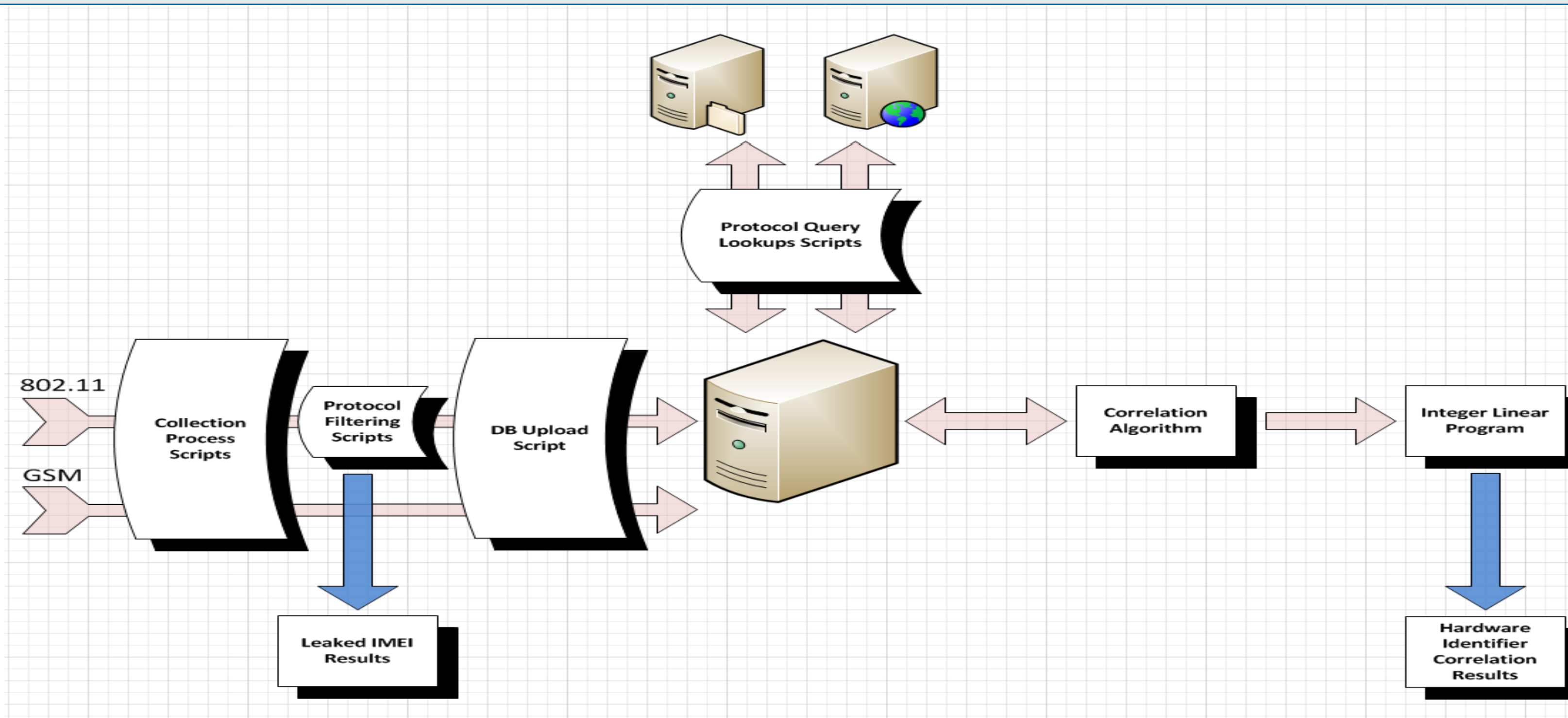
- GSM – IMEI

TAC	FAC	Serial No	Check Digit
NNXXXX	YY	ZZZZZZ	A

- WiFi – MAC Address



Methodology and Data Collection



Methodology and Data Collection

- Simulated collection of GSM and WiFi hardware identifiers
  - 18 mobile devices with GSM and WiFi capability
- To model temporal movement, dataset includes six different snapshots in time
  - Three different locations were simulated to model spatial movement
- A randomly selected subset of our devices was used for each of the six iterations

Methodology and Data Collection

- Test Devices

Count	Make	Model	ID
2	Acer	Iconia A501	aIa
7	Apple	iPhone 3GS	iPh
1	Apple	iPad	iPa
1	HTC	Hero	hH
1	HTC	Nexus One	hNo
1	HTC	Surround T8788	hSt
2	HTC	Eng Handset	hEh
1	Samsung	I7500	sGa
2	Samsung	19250 Galaxy	sGn

Methodology and Data Collection

- Two different perspectives
  - Limited Adversary – able to observe identifiers only in time and space
  - Advanced Adversary – visibility into the data stream of each protocol

Methodology and Data Collection

- Limited Adversary
  - Hardware identifier (IMEI / MAC address)
  - Temporal (# of times IMEI / MAC pair seen together)
  - Spatial (# of locations IMEI / MAC pair seen together)
- Advanced Adversary
  - Use of all limited adversary techniques
  - User-Agent string in HTTP traffic
  - User Agent Profiles in HTTP traffic
  - Bonjour
  - DHCP / BOOTP

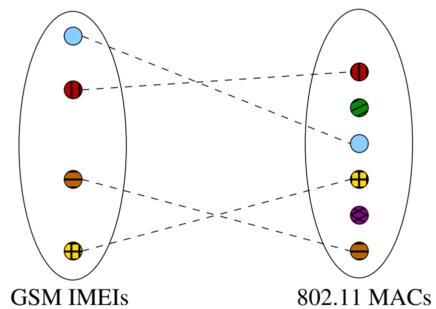
Methodology and Data Collection

Device	TAC-Derived Info*	OUI-Derived Info*	UAProf	Bonjour	BOOTP
Acer Iconia A501	Ericsson F5521gw PCIE	Azurewave Tech	http://support.acer.com/UAprofile/Acer A501 Profile.xml	n/a	n/a
Apple iPhone 3GS	Apple iPhone 3GS 16GB	Apple, Inc	n/a	iPhone3GS-1.local	iPhone3GS-1
HTC Hero	HTC Hero	HTC Corporation	http://www.htcmms.com.tw/Android/Common/Hero/ua-profile.xml	n/a	n/a
Samsung Galaxy Nexus	Samsung I9250 Galaxy Nexus	Samsung Electro	n/a	n/a	android-cd5db081844aeb9c

*Used IEEE and Nobbi databases

Correlation

- Correlation problem is bipartite matching – associate observed MAC addresses with observed IMEIs



- Generalize this correlation as an Integer Linear Program (ILP) that accommodates the different evidence in our datasets as constraints on the solution

Correlation

- Let A be the sparse association matric such that Ai,j =1 indicates that TAC i is associated with MAC j. We wish to maximize the sum of “strong” correlations, subject to the feasibility constraints that only one TAC may be associated with one MAC and vice versa.
- The A that maximizes the sum of the evidence provides the inferred hardware correlations.
- *Necessary? Summarize?*

Correlation

- As an ILP, which we express in the MathProg modeling language and solve using GLPK
  - Limited

$$\text{Maximize} \sum_{i=1}^{m} \sum_{j=1}^{n} T_{ij} A_{ij} + S_{ij} A_{ij}$$

$$\text{Subject to} \sum_{i=1}^{m} A_{ij} \leq 1, \; \sum_{j=1}^{n} A_{ij} \leq 1$$

  - Advanced

$$\text{Maximize} \sum_{k=1}^{c} \sum_{i=1}^{m} \sum_{j=1}^{n} C_{ij}^{k} W^{k} A_{i,j}$$

$$\text{Subject to} \sum_{i=1}^{m} A_{ij} \leq 1, \; \sum_{j=1}^{n} A_{ij} \leq 1$$

Results

- Limited Adversary
  - Temporal
  - Spatial
  - TAC – OUI

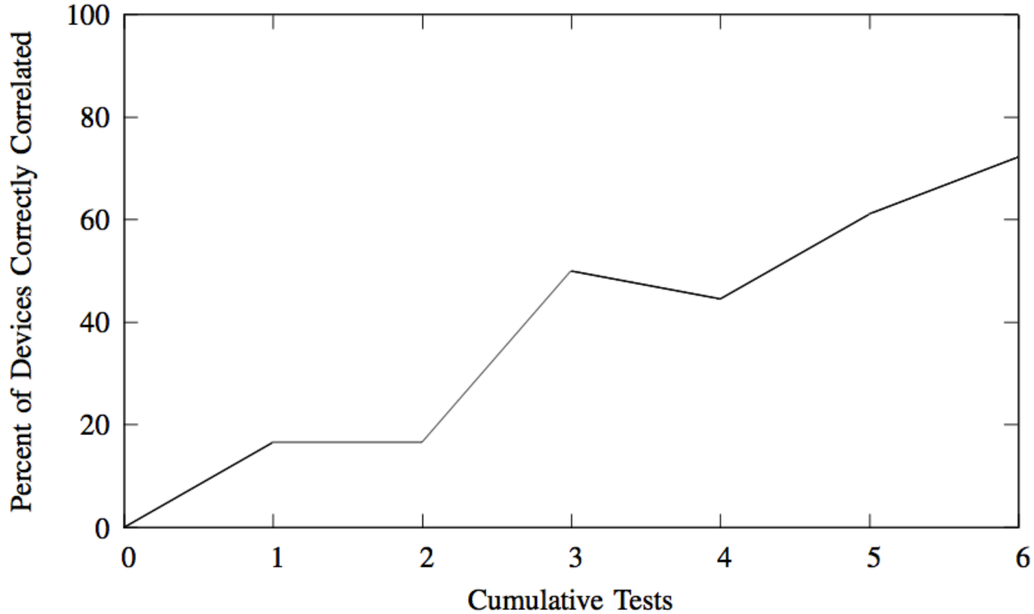


Fig. 4: Devices correctly correlated as a function of time

Results – Limited Adversary

TABLE III: Results of Temporal, Spatial, Temporal-Spatial (T/S), and Weighted Temporal-Spatial (T*5/S)

Temporal	Spatial	T / S	T * 5 / S
iPh1 = iPh1	iPh1 = iPh1	iPh1 = iPh1	iPh1 = iPh1
iPh2 = iPh2	iPh2 = hEh1	iPh2 = iPh2	iPh2 = iPh2
iPh3 = iPh3	iPh3 = iPh3	iPh3 = iPh3	iPh3 = iPh3
iPh4 = iPh5	iPh4 = hEh2	iPh4 = iPh4	iPh4 = iPh5
iPh5 = iPh4	iPh5 = iPh5	iPh5 = iPh5	iPh5 = iPh4
iPh6 = iPh6	iPh6 = iPh6	iPh6 = iPh6	iPh6 = iPh6
iPh7 = iPh7	iPh7 = hNo1	iPh7 = iPh7	iPh7 = iPh7
iPa1 = iPa1	iPa1 = iPh2	iPa1 = iPa1	iPa1 = iPa1
sGn1 = sGn1	sGn1 = iPa1	sGn1 = sGn1	sGn1 = sGn1
sGn2 = sGn2	sGn2 = hH1	sGn2 = hNo1	sGn2 = sGn2
hSt1 = hSt1	hSt1 = hSt1	hSt1 = hSt1	hSt1 = hSt1
hNo1 = hNo1	hNo1 = iPh7	hNo1 = sGn2	hNo1 = hNo1
sGa1 = hH1	sGa1 = aIa2	sGa1 = aIa2	sGa1 = aIa2
aIa1 = aIa1	aIa1 = aIa1	aIa1 = aIa1	aIa1 = aIa1
aIa2 = aIa2	aIa2 = sGn2	aIa2 = hEh2	aIa2 = hEh2
hH1 = sGa1	hH1 = sGa1	hH1 = sGa1	hH1 = hH1
hEh1 = hEh1	hEh1 = sGn1	hEh1 = hEh1	hEh1 = hEh1
hEh2 = aN7s1	hEh2 = iPh4	hEh2 = aN7s1	hEh2 = sGa1
13/18	6/18	12/18	13/18

TABLE IV: Normal vs. Weighted TAC-OUI Correlation

T/S/O	T*5/S/O
iPh1 = iPh1	iPh1 = iPh1
iPh2 = iPh2	iPh2 = iPh2
iPh3 = iPh3	iPh3 = iPh3
iPh4 = iPh4	iPh4 = iPh5
iPh5 = iPh5	iPh5 = iPh4
iPh6 = iPh6	iPh6 = iPh6
iPh7 = iPh7	iPh7 = iPh7
iPa1 = iPa1	iPa1 = iPa1
sGn1 = sGn1	sGn1 = sGn1
sGn2 = sGn2	sGn2 = sGn2
hSt1 = hSt1	hSt1 = hSt1
hNo1 = hNo1	hNo1 = hNo1
sGa1 = aIa2	sGa1 = aIa2
aIa1 = aIa1	aIa1 = aIa1
aIa2 = sGa1	aIa2 = sGa1
hH1 = hH1	hH1 = hH1
hEh1 = hEh1	hEh1 = hEh1
hEh2 = hEh2	hEh2 = hEh2
16/18	14/18

Results

- Advanced Adversary
  - Temporal
  - Spatial
  - TAC – OUI
  - TAC – User-Agent
  - TAC – UAProf
  - TAC – Bonjour
  - TAC - DHCP

Results – Advanced Adversary

TABLE V: Protocol Analysis with TAC-OUI Correlation

UA	UAProf	OUI	Bonjour	DHCP
iPh1 = iPh5	iPh1 = iPh2	iPh1 = iPa1	iPh1 = hEh1	iPh1 = iPh5
iPh2 = iPh2	iPh2 = hEh2	iPh2 = iPh6	iPh2 = iPh7	iPh2 = hEh2
iPh3 = iPh4	iPh3 = hEh1	iPh3 = iPh4	iPh3 = hH1	iPh3 = hH1
iPh4 = iPh7	iPh4 = hNo1	iPh4 = iPh2	iPh4 = iPh2	iPh4 = iPh2
iPh5 = iPh3	iPh5 = sGa1	iPh5 = iPh3	iPh5 = iPh5	iPh5 = iPh7
iPh6 = iPh1	iPh6 = iPh1	iPh6 = iPh1	iPh6 = iPh1	iPh6 = iPh1
iPh7 = iPh6	iPh7 = iPh3	iPh7 = iPf1	iPh7 = iPh6	iPh7 = iPh6
iPa1 = iPa1	iPa1 = sGn2	iPa1 = iPh7	iPa1 = hEh2	iPa1 = iPa1
sGn1 = sGn1	sGn1 = sGn1	sGn1 = sGn1	sGn1 = sGn1	sGn1 = sGn1
sGn2 = sGn2	sGn2 = iPh7	sGn2 = sGn2	sGn2 = sGn2	sGn2 = sGa1
hSt1 = hSt1	hSt1 = aN7s1	hSt1 = hEh2	hSt1 = aN7s1	hSt1 = hEh1
hNo1 = hEh1	hNo1 = iPh5	hNo1 = hH1	hNo1 = sGa1	hNo1 = iPh4
sGa1 = sGa1	sGa1 = iPh6	sGa1 = iPh5	sGa1 = iPh3	sGa1 = iPf1
aIa1 = aN7s1	aIa1 = aIa1	aIa1 = aIa1	aIa1 = aIa1	aIa1 = aIa1
aIa2 = iPf1	aIa2 = iPf1	aIa2 = aIa2	aIa2 = iPf1	aIa2 = aIa2
hH1 = hH1	hH1 = hH1	hH1 = hEh1	hH1 = hNo1	hH1 = hNo1
hEh1 = aIa2	hEh1 = aIa2	hEh1 = hSt1	hEh1 = aIa2	hEh1 = aN7s1
hEh2 = aIa1	hEh2 = iPh4	hEh2 = hNo1	hEh2 = iPh4	hEh2 = sGn2
7/18	3/18	4/18	4/18	4/18

TABLE VI: Results Incorporating User-Agent Data

T/S/UA	T*5/S/UA
iPh1 = iPh1	iPh1 = iPh1
iPh2 = iPh2	iPh2 = iPh2
iPh3 = iPh3	iPh3 = iPh3
iPh4 = iPh4	iPh4 = iPh5
iPh5 = iPh5	iPh5 = iPh4
iPh6 = iPh6	iPh6 = iPh6
iPh7 = iPh7	iPh7 = iPh7
iPa1 = iPa1	iPa1 = iPa1
sGn1 = sGn1	sGn1 = sGn1
sGn2 = sGn2	sGn2 = sGn2
hSt1 = hSt1	hSt1 = hSt1
hNo1 = hNo1	hNo1 = hNo1
sGa1 = sGa1	sGa1 = sGa1
aIa1 = aIa1	aIa1 = aIa1
aIa2 = hEh2	aIa2 = aIa2
hH1 = hH1	hH1 = hH1
hEh1 = hEh1	hEh1 = hEh1
hEh2 = aN7s1	hEh2 = hEh2
16/18	16/18

Results – Advanced Adversary

TABLE VII: Results Incorporating DHCP and Bonjour

T/S/O/B/M	T*5/S/O/B/M	T*5/S/O/B*.75/M*.75
iPh1 = iPh1	iPh1 = iPh1	iPh1 = iPh1
iPh2 = iPh2	iPh2 = iPh2	iPh2 = iPh2
iPh3 = iPh3	iPh3 = iPh3	iPh3 = iPh3
iPh4 = iPh5	iPh4 = iPh4	iPh4 = iPh4
iPh5 = iPh4	iPh5 = iPh5	iPh5 = iPh5
iPh6 = iPh6	iPh6 = iPh6	iPh6 = iPh6
iPh7 = iPh7	iPh7 = iPh7	iPh7 = iPh7
iPa1 = iPa1	iPa1 = iPa1	iPa1 = iPa1
sGn1 = sGn1	sGn1 = sGn1	sGn1 = sGn1
sGn2 = sGn2	sGn2 = sGn2	sGn2 = sGn2
hSt1 = hSt1	hSt1 = hSt1	hSt1 = hSt1
hNo1 = hEh2	hNo1 = hEh2	hNo1 = hNo1
sGa1 = sGa1	sGa1 = sGa1	sGa1 = sGa1
aIa1 = aIa1	aIa1 = aIa1	aIa1 = aIa1
aIa2 = aIa2	aIa2 = aIa2	aIa2 = aIa2
hH1 = hH1	hH1 = hH1	hH1 = hH1
hEh1 = hEh1	hEh1 = hEh1	hEh1 = hEh1
hEh2 = hNo1	hEh2 = hNo1	hEh2 = hEh2
14/18	16/18	18/18

TABLE VIII: Results After Incorporating UAProf Data

T/S/UAProf	T*5/S/UAProf
iPh1 = aIa1	iPh1 = aIa1
iPh2 = iPh2	iPh2 = iPh2
iPh3 = iPh3	iPh3 = iPh3
iPh4 = iPh4	iPh4 = iPh5
iPh5 = iPh5	iPh5 = iPh4
iPh6 = iPh6	iPh6 = iPh6
iPh7 = iPh7	iPh7 = iPh7
iPa1 = iPa1	iPa1 = iPa1
sGn1 = sGn1	sGn1 = sGn1
sGn2 = sGn2	sGn2 = sGn2
hSt1 = hSt1	hSt1 = hSt1
hNo1 = hNo1	hNo1 = hNo1
sGa1 = sGa1	sGa1 = aIa2
aIa1 = iPh1	aIa1 = iPh1
aIa2 = aIa2	aIa2 = sGa1
hH1 = hH1	hH1 = hH1
hEh1 = hEh1	hEh1 = hEh1
hEh2 = hEh2	hEh2 = aN7s1
16/18	12/18

Results – Advanced Adversary

TABLE IX: Results After Incorporating All Collected Data

T/S/O/U/X/M/B	T*5/S/O/U/X/M/B	T*5/S/O/U/X/M/B*.75	T*5/S/O/U/X/M*.75/B*.75
iPh1 = iPh1	iPh1 = iPh1	iPh1 = iPh1	iPh1 = iPh1
iPh2 = iPh2	iPh2 = iPh2	iPh2 = iPh2	iPh2 = iPh2
iPh3 = iPh3	iPh3 = iPh3	iPh3 = iPh3	iPh3 = iPh3
iPh4 = iPh5	iPh4 = iPh5	iPh4 = iPh5	iPh4 = iPh4
iPh5 = iPh4	iPh5 = iPh4	iPh5 = iPh4	iPh5 = iPh5
iPh6 = iPh6	iPh6 = iPh6	iPh6 = iPh6	iPh6 = iPh6
iPh7 = iPh7	iPh7 = iPh7	iPh7 = iPh7	iPh7 = iPh7
iPa1 = iPa1	iPa1 = iPa1	iPa1 = iPa1	iPa1 = iPa1
sGn1 = sGn1	sGn1 = sGn1	sGn1 = sGn1	sGn1 = sGn1
sGn2 = sGn2	sGn2 = sGn2	sGn2 = sGn2	sGn2 = sGn2
hSt1 = hSt1	hSt1 = hSt1	hSt1 = hSt1	hSt1 = hSt1
hNo1 = hEh2	hNo1 = hEh2	hNo1 = hNo1	hNo1 = hNo1
sGa1 = sGa1	sGa1 = sGa1	sGa1 = sGa1	sGa1 = sGa1
aIa1 = aIa1	aIa1 = aIa1	aIa1 = aIa1	aIa1 = aIa1
aIa2 = aIa2	aIa2 = aIa2	aIa2 = aIa2	aIa2 = aIa2
hH1 = hH1	hH1 = hH1	hH1 = hH1	hH1 = hH1
hEh1 = hEh1	hEh1 = hEh1	hEh1 = hEh1	hEh1 = hEh1
hEh2 = hNo1	hEh2 = hNo1	hEh2 = hEh2	hEh2 = hEh2
14/18	14/18	16/18	18/18

Results – Leaked Identifiers

TABLE X: Applications Leaking Device Identifiers

Android	Leaks	iPhone	Leaks	Windows	Leaks
AutoRun		n/a		n/a	
Assistant	SHA1	Assistant		Assistant	
Classic Simon		n/a		n/a	
Documents To Go 3.0		n/a		n/a	
Droid Jump		n/a		n/a	
iHeartRadio	IMEI	iHeartRadio	UDID	iHeartRadio	
KAYAK		KAYAK		n/a	
Moco Chat, Meet, Games	IMEI	Moco Chat, Meet, Games		MocoSpace	
Moron Test: Old School	IMEI	Moron Test: Old School	UDID	n/a	
Moron Test: Section 2		n/a		n/a	
Paper Toss		Paper Toss		n/a	
Smart Simon	MD5, SHA1	n/a		n/a	
SmartTacToe		SmartTacToe	UDID	n/a	
Starbucks		Starbucks		n/a	
Video Poker	MD5, SHA1	n/a		n/a	
Video Poker		n/a		n/a	
White & Yellow Pages	IMEI	White & Yellow Pages		White & Yellow Pages	
Yellow Pages	IMEI, MD5	Yellow Pages		Yellow Pages	

Future Work

- Blah

References

1 W. Enck, P. Gilbert, B.-G. Chun, L. P. Cox, J. Jung, P. McDaniel, and A. N. Sheth, "TaintDroid: an information-flow tracking system for realtime privacy monitoring on smartphones," in *Proceedings of the 9th USENIX OSDI conference*, 2010, pp. 1–6.

2 R.-P. Weinmann, "Baseband attacks: Remote exploitation of memory corruptions in cellular protocol stacks," in *USENIX Workshop on Offensive Technologies (WOOT12)*, 2012.

3 C. Mulliner, N. Golde, and J.-P. Seifert, "SMS of Death: from analyzing to attacking mobile phones on a large scale," in *Proceedings of the 20th USENIX conference on Security*, 2011, pp. 24–24.

4 K. Nohl, "Rooting SIM Cards," in *Blackhat Conference*, 2013.

5 S. L. Garfinkel, A. Juels, and R. Pappu, "RFID Privacy: An Overview of Problems and Proposed Solutions," *Published by the IEEE Computer Society*, p. 14, May 2005, http://www.cs.colorado.edu/~rhan/CSCI 7143 Fall 2007/Papers/rfid security 01439500.pdf.

6 M. Egele, C. Kruegel, E. Kirda, and G. Vigna, "PiOS: Detecting privacy leaks in iOS applications," in *Proceedings of the Network and Distributed System Security Symposium*, 2011.

7 P. Hornyack, S. Han, J. Jung, S. Schechter, and D. Wetherall, "These aren't the droids you're looking for: retrofitting android to protect data from imperious applications," in *Proceedings of the 18th ACM CCS conference*, 2011, pp. 639–652.

8 G. Eisenhaur, M. N. Gagnon, T. Demir, and N. Daswani, "Mobile Malware Madness and How to Cap the Mad Hatters," in *Blackhat Conference*, 2011.

9 M. N. Gagnon, "Hashing IMEI numbers does not protect privacy," Dasient Blog, 2011, http://blog.dasient.com/2011/07/ hashing- imei- numbers- does- not- protect.html.

Correlating GSM and 802.11 Hardware Identifiers

LCDR Jeremy Martin – jbmartin@nps.edu
LT Danny Rhame – dsrhame@nps.edu
Dr. Robert Beverly – rbeverly@nps.edu
Dr. John McEachen – mceachen@nps.edu

Naval Postgraduate School