BTI 4202: AI and Information Security

Christian Grothoff

BFH

27.2.2026

"Move fast and break things." –Marc Zuckerberg

Protocols

AI and Security

Mass Surveillance

More ethical case studies

Conclusion

AI/ML for Network Security

Part I: Protocols

Protocols

- "A **protocol** is a series of steps, involving two or more parties, designed to accomplish a task."
- Everyone involved must know the steps in advance and agree to follow it.
- The protocol must be complete and unambiguous.
- For cryptographic protocols, it should not be possible to do more or learn more than *what is specified in the protocol*.

Dramatis Personae

- Alice, Bob, Carol and Dave
- Eve – Eavesdropper
- Mallory – Malicious active attacker
- Trent – Trusted arbitrator
- Walter – Warden
- Peggy – Prover
- Victor – Verifier

Attack Personae

- Eavesdroppers
- Passive cheaters
- Active cheaters
- Real-world adversaries – Mallory

Passive attacks

Attempts to learn or make use of information from the system,

but does not affect system resources.

- Interception (without interference)
- Traffic analysis

Active attack

Attempt to alter system resources or affect their operation.

- Masquerade (pretend to be someone else)
- Replay (repeat previous messages)
- Data modification (flip some bits)
- Denial of service (cut the power)

Efficiency

- Number of steps in protocol
- Size of messages
- Conflict resolution cost:
  1. Involvement of trusted party (arbitrated protocols)
  2. Resolution by trusted party on dispute (adjudicated protocols)
  3. Self-enforcing protocols

Part II: AI and Security

Statistics

- mathematical techniques for drawing general conclusions from data samples
- means, medians, distributions, samples, significance, bias
- resulting aggregates may have meaning, or not
- no hard assurances about individual inputs, only probabilities

Machine Learning

We have too much (statistical) data for humans to determine which ones have meaning, so:

- Ask computer to figure out which inputs matter!
- Different techniques:
  - Supervised learning: given example inputs and desired outputs, derive "general rule"
  - Unsupervised learning: find hidden structure in data
  - Reinforecment learning: algorithm selects actions, receives feedback based on result(s)
- Shared outcome: data in, statistical predictors out

Artificial Intelligence

(Jimmy Carr, 2024)

Is Improving Security Possible?

(8′2018)

Part III: Mass Surveillance

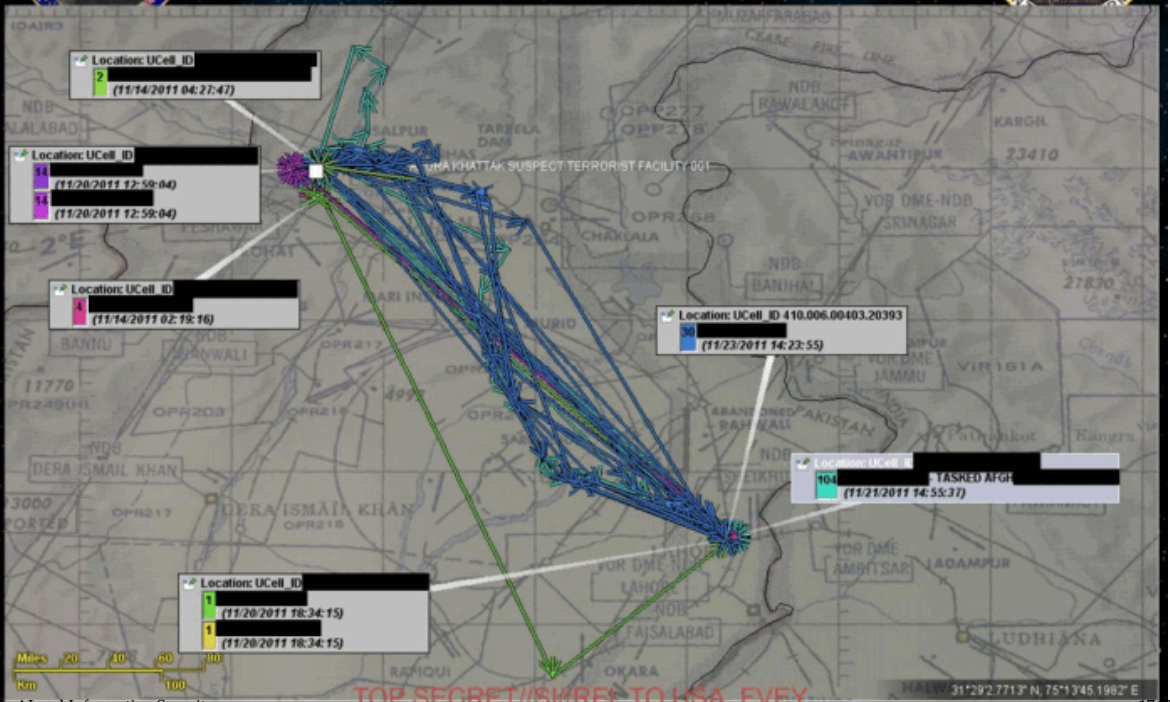
Societal control technology: Analytics

Cloud Analytic Building Blocks

- Travel Patterns
  - Travel phrases (Locations visited in given timeframe)
  - Regular/repeated visits to locations of interest
- Behavior-Based Analytics
  - Low use, incoming calls only
  - Excessive SIM or Handset swapping
  - Frequent Detach/Power-down
  - Courier machine learning models
- Other Enrichments
  - Travel on particular days of the week
  - Co-travelers
  - Similar travel patterns
  - Common contacts
  - Visits to airports
  - Other countries
  - Overnight trips
  - Permanent move

SMARTTRACKER Travel View

31 October – 23 November



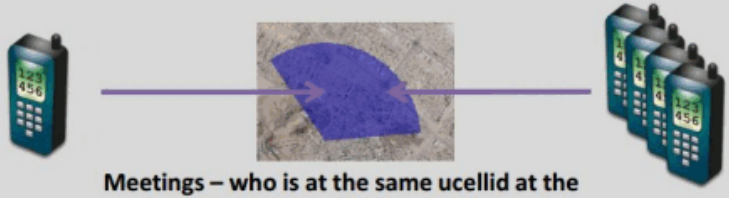
Analytic Tradecraft

- Examine travel patterns for common routes and meeting locations
  - Run cell soaks on all common meeting locations during meeting timeframe
- Analyze selectors for common contacts
- Analyze selectors for handset sharing behavior

Repeat procedure with resulting selectors

Correlate with other known and suspected selectors

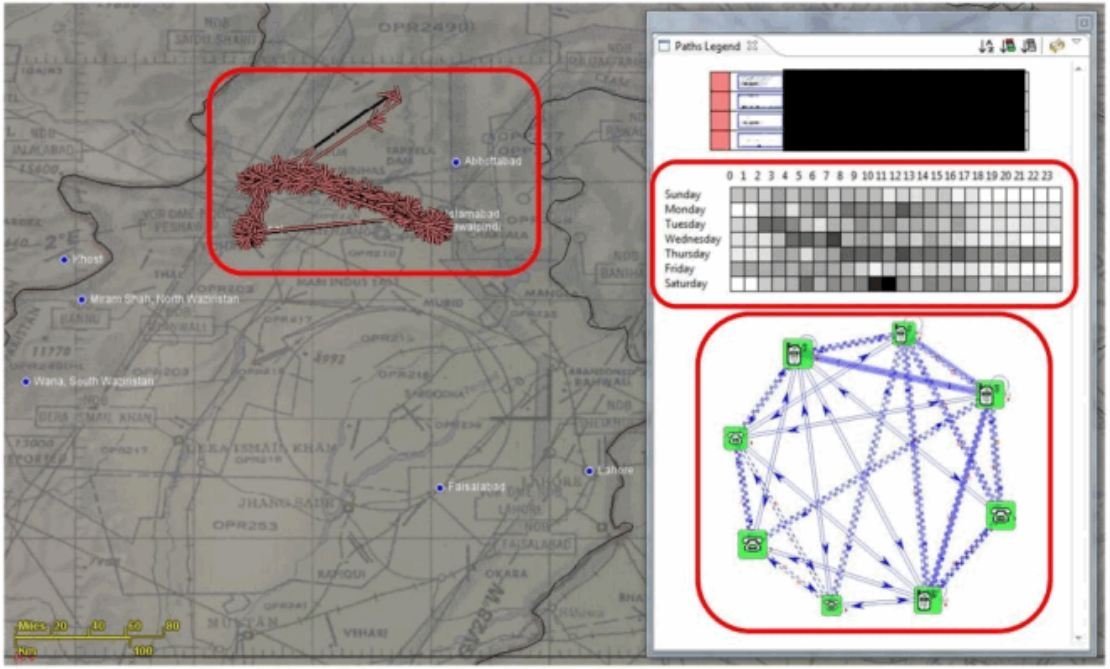
RT-RG Analytics



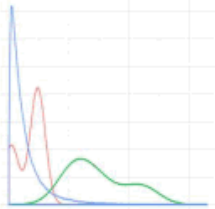
Meetings – who is at the same ucellid at the same time as the potential courier at the destination city?...Multiple times.

Sidekicks – is there a pair traveling together to the destination city?

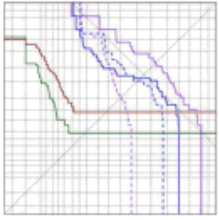
From GSM metadata, we can measure aspects of each selector's pattern-of-life, social network, and travel behavior



This presentation describes our search for AQSL couriers using behavioral profiling



Behavioral Feature Extraction

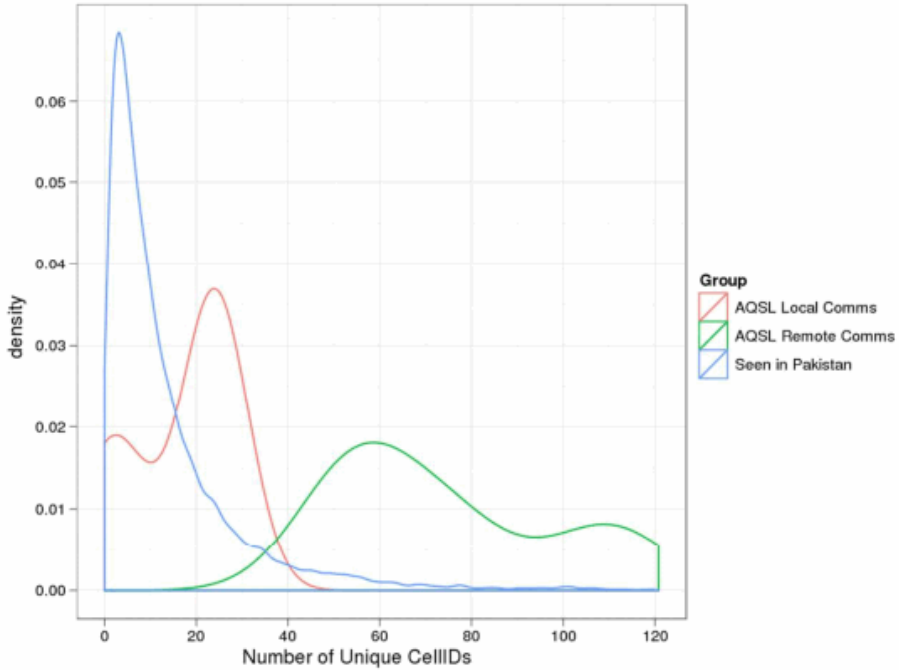


Cross Validation Experiment on AQSL Couriers

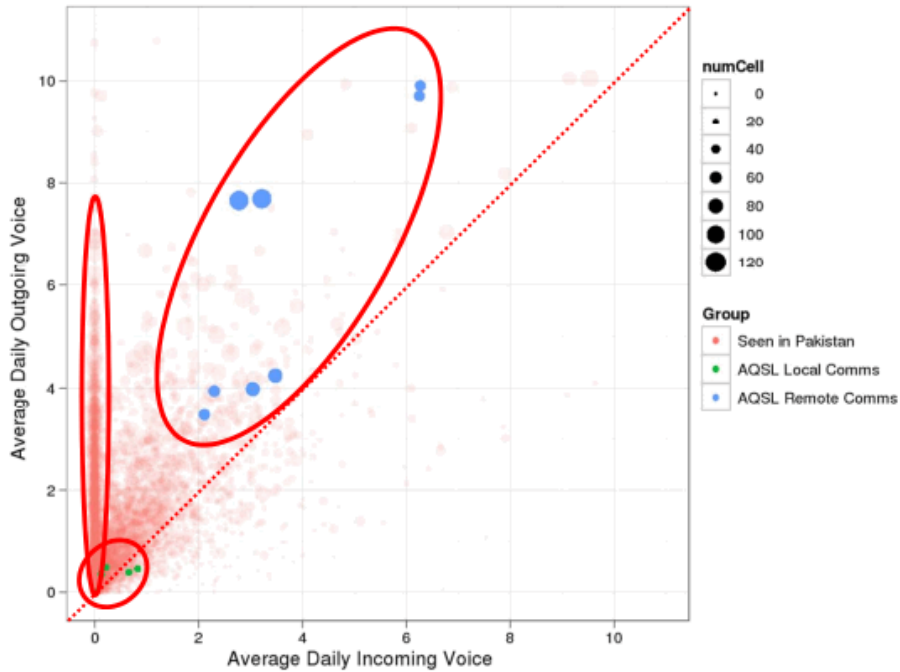
Preliminary SIGINT Findings

TOP SECRET//COMINT//REL TO USA, FVEY

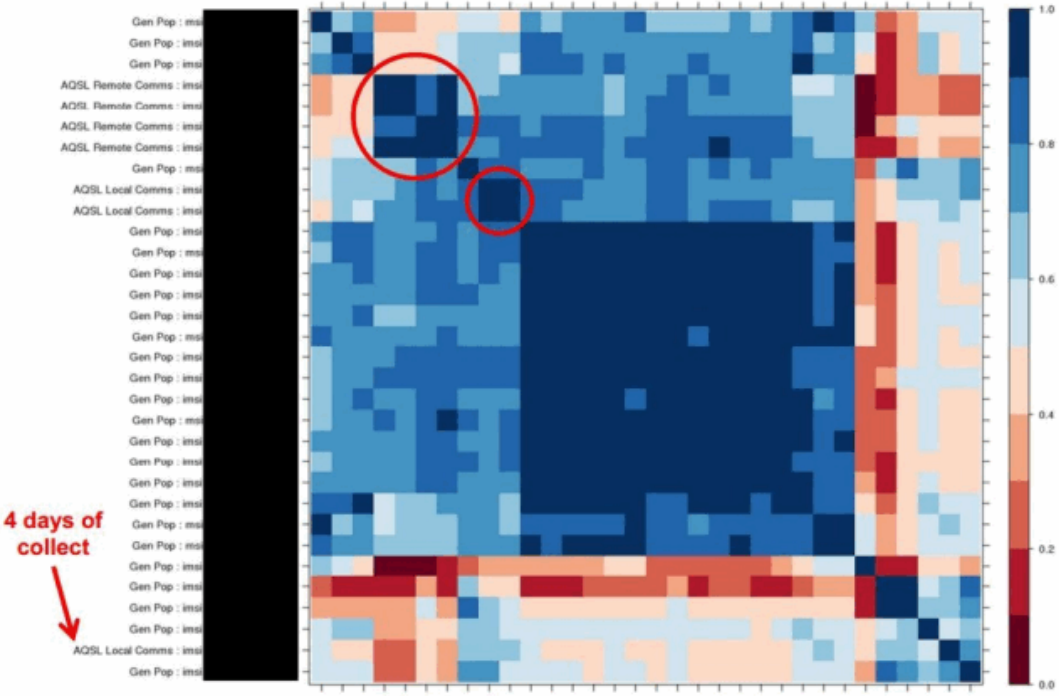
Counting unique UCELLIDs shows that couriers travel more often than typical Pakistani selectors



By examining multiple features at once, we can see some indicative behaviors of our courier selectors



Looking at a hierarchical clustering derived from all 80 features, the AQSL groups mostly stay together

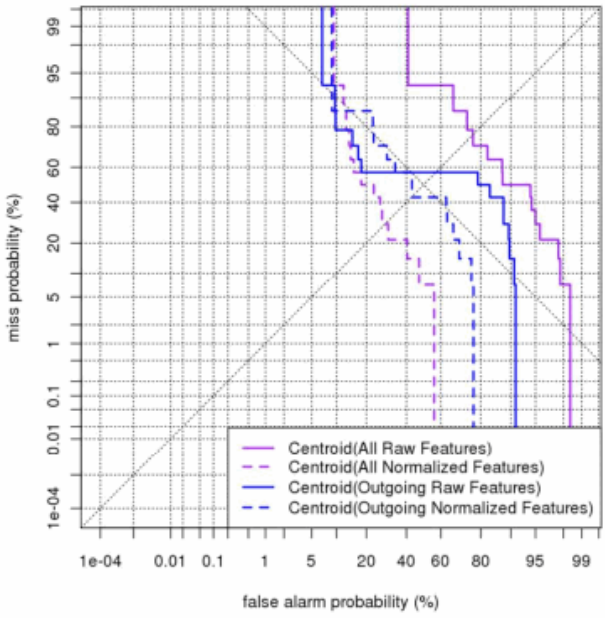


TOP SECRET//COMINT//REL TO USA, FVEY

Our initial detector uses the centroid of the AQSL couriers to "find other selectors like these"

AQSL Cross-Validation Experiment

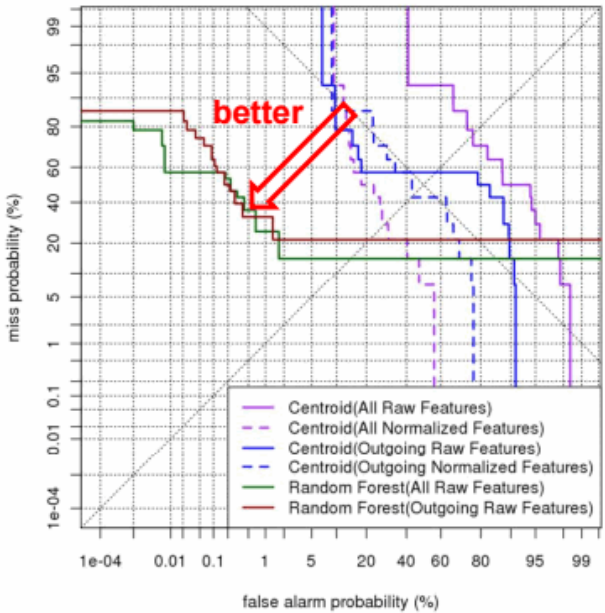
- 7 MSISDN/IMSI pairs
- Hold each pair out and score them when training the centroid on the rest
- Assume that random draws of Pakistani selectors are nontargets
- How well do we do?



Statistical algorithms are able to find the couriers at very low false alarm rates, if we're allowed to miss half of them

Random Forest Classifier

- 7 MSISDN/IMSI pairs
- Hold each pair out and then try to find them after learning how to distinguish remaining couriers fro n other Pakistanis
  (using 100k random selectors here)
- Assume that random draws of Pakistani selectors are nontargets
- 0.18% False Alarm Rate at 50% Miss Rate



We've been experimenting with several error metrics on both small and large test sets

Training Data	Classifier	Features	100k Test Selectors		55M Test Selectors	
			False Alarm Rate at 50% Miss Rate	Mean Reciprocal Rank	Tasked Selectors in Top 500	Tasked Selectors in Top 100
None	Random	None	50%	1/23k (simulated)	0.64 (active/Pak)	0.13 (active/Pak)
Known Couriers	Centroid	All	20%	1/18k		
		Outgoing	43%	1/27k		
	Random Forest		0.18%	1/9.9	5	1
+ Anchory Selectors			0.008%	1/14	21	6

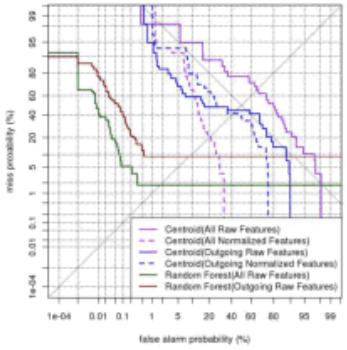
Random Forest trained on Known Couriers + Anchory Selectors:

- 0.008% false alarm rate at 50% miss rate
- 46x improvement over random performance when evaluating its tasked precision at 100

Preliminary results indicate that we're on the right track, but much remains to be done

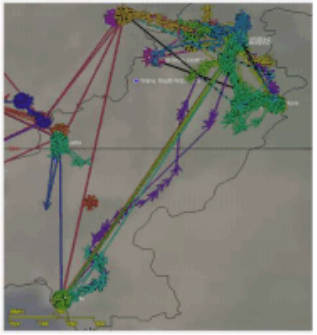
Cross Validation Experiment:

- Random Forest classifier operating at 0.18% false alarm rate at 50% miss
- Enhancing training data with Anchory selectors reduced that to 0.008%
- Mean Reciprocal Rank is ~1/10



Preliminary SIGINT Findings:

- Behavioral features helped discover similar selectors with "courier-like" travel patterns
- High number of tasked selectors at the top is hopefully indicative of the detector performing well "in the wild"



"The U.S. government labeled a prominent journalist as a member of Al Qaeda and placed him on a watch list of suspected terrorists, according to a top-secret document that details U.S. intelligence efforts to track Al Qaeda couriers by analyzing metadata."

—The Intercept, May 8 2015

[1] "Genocide is the deliberate destruction of physical life of individual human beings by reason of their membership of any human collectivity as such." –Pieter N. Drost, Dutch law professor

"The U.S. government labeled a prominent journalist as a member of Al Qaeda and placed him on a watch list of suspected terrorists, according to a top-secret document that details U.S. intelligence efforts to track Al Qaeda couriers by analyzing metadata."

—The Intercept, May 8 2015

192 Million people live in Pakistan.

- 0.18% of the Pakistani population = 343,800 innocent citizens
- 0.008% of the Pakistani population = 15,280 innocent citizens

[1] "Genocide is the deliberate destruction of physical life of individual human beings by reason of their membership of any human collectivity as such." –Pieter N. Drost, Dutch law professor

"The U.S. government labeled a prominent journalist as a member of Al Qaeda and placed him on a watch list of suspected terrorists, according to a top-secret document that details U.S. intelligence efforts to track Al Qaeda couriers by analyzing metadata."

—The Intercept, May 8 2015

192 Million people live in Pakistan.

- 0.18% of the Pakistani population = 343,800 innocent citizens
- 0.008% of the Pakistani population = 15,280 innocent citizens

This is with half of AQSL couriers surviving the genocide.[1].

"We kill based on metadata."

—Michael Hayden (former NSA & CIA director)

[1] "Genocide is the deliberate destruction of physical life of individual human beings by reason of their membership of any human collectivity as such." –Pieter N. Drost, Dutch law professor

The NSA mathematician's presentation only gives the percentages.

Compartmentalization

The NSA mathematician's presentation only gives the percentages.

Compartmentalization is an unconscious psychological defense mechanism used to avoid cognitive dissonance, or the mental discomfort and anxiety caused by a person's having conflicting values, cognitions, emotions, beliefs, etc. within themselves.

Part IV: More case studies

Case study: Communication

"A company is developing new software for private communication. This will enable its customers to communicate with "complete" privacy. The solution does not include backdoors and thus the company cannot support requests for legal intercept*."*

Perform an ethical case study on what the company should do!

Case study: Private communication

- Suppose the company added a “feature” to provide legal intercept support. Does this change your assessment?
- Does it make a difference if the software is “free software” developed by a community instead of proprietary software from a company?

Case study: Remote Attestation

"Researchers develop a remote attestation *method, by which a remote party (over the Internet) can tell whether some computer runs exactly the approved software configuration, and only in this case releases a decryption key. Applications for remote attestation include a broad range of copyright enforcement methods (DRM), where movies are only decoded if no recording software is detected, or where software only runs if it is properly licensed for that environment."*

Perform an ethical case study on what the researchers should do!

Case study: Remote Attestation

- Does it make a difference if this is developed for the control of military software? (Think of “copyright” enforcement for nuclear weapons launch software.)

Case study: Remote Attestation

- Does it make a difference if this is developed for the control of military software? (Think of "copyright" enforcement for nuclear weapons launch software.)
- Such a mechanism is also useful to malware authors to prevent diagnosis of the malware's operation by anti-virus companies. Does this information change your assessment?

Conclusion

- Computers have no sense of ethics.
- Physical reality (including code) beats human law.

$\Rightarrow$ We need to be careful about which technology we adopt.

AI/ML for Network Security

IETF 116 (3’2023)

Questions?

"The most unpardonable sin in society is independence of thought." –Emma Goldman