Research Topics in Security and Privacy using Data Science

School of Informatics

University of Edinburgh

David Aspinall

David.Aspinall@ed.ac.uk

http://secpriv.inf.ed.ac.uk/

http://cybersec.ed.ac.uk/

Outline

Context

State of the art

Sample topics

Summary

Context

Cyber Security

Skills gap

- UK National Audit Office: 20 year gap
- Security provisions often weak/missing
- Opportunities for domain experts, future CIOs

Knowledge gap

- Technology racing ahead of understanding
- Sophistication and reach of attackers growing
- Challenges in cyber crime, forensics, CNI, . . .

Many important **research questions**, ranging from foundational science to applied and operational issues.

Cyber Privacy

Awareness is growing

- Continual data leaks, government surveillance
- Privacy not dead: people care about *use of data*
- Future: part of, not in conflict with, cyber security

Challenges

- Understanding privacy policies, data value
- Bridging gap between technical & practical
- Providing realistic privacy guarantees

Again, many **research problems**, ranging from fundamental techniques to user-oriented investigations.

State of the art

Security thinking

> *“What could an attacker do to break my system?”*

Researchers define an **attack model** which limits adversary’s capability to help achieve security/privacy.

- *Passive* (can read) versus *Active* (can alter)
- Data or computationally bounded
- . . .

Aspects of study

Roughly, two flavours: breaking things or fixing them.

Attacks:

- Disclose or demonstrate an attack
- Discover and expose a real-world attack

Defences:

- Provide a new mechanism to address a known flaw
- Prevent class of attacks to guarantee a property
- Conduct a risk assessment, mitigation strategy

A typical paper might try to do both kind of things.
Q. Why might that be unconvincing?

Using the data

Studying attacks, we can use data to:

- discover secrets (e.g., including from *side channels*)
- find anomalous, suspicious behaviour (IDS)
- understand attacker behaviour, actions
- find incriminating material (digital forensics)

For defences, we can use data to:

- hide or obscure secrets
- synthesise new solutions

In general: machine learning and pattern recognition have been applied with success, but immature in some areas. Use of text analytics and mining gaining attention recently.

Examples of precise concepts

Data leak prevention/detection:

- **Non-interference**: guarantees no info flow
- **Provenance tracking**: guarantees audit trail

Privacy:

- **Randomised response**: answer questions maintaining confidentiality
- **Differential privacy**: run queries within an privacy budget

Anonymity:

- **k-anonymity**: prevent de-anonymisation in data

Sample topics

Android Malware/App Studies

Research questions: can we

- learn *good security policies* from good apps?
- describe (in text) good and bad behaviour?
- understand behaviour of mobile ad networks?

Data available from:

- **McAfee** (part of Intel Security)
- Google, Amazon, others: crawling App Stores

See App Guarden project:
`http://groups.inf.ed.ac.uk/security/appguarden`

Sensor-driven Authentication and Privacy

In previous work we’ve studied *continuous authentication* and *automatic context determination* using sensor data from mobile phones.

New questions:

- To what extent is such information uniquely identifying?
- What can we do to provide e.g., location services, but preserve privacy? (with guarantees)

Data available from:

- Previous research projects at UoE and GCU
- Some limited amount of openly available data
- Several apps for collecting your own/others’ data

Machine learning to improve bug detectors

Static analysis tools for detecting **security vulnerabilities** due to programming bugs.

However, they suffer from poor adoption rates due to high false positive rates: developers must trawl through hundreds of “potential” flaws.

Can we improve this by using learning techniques to help automatically triage problems based on similar previous ones?

Data available from:

- Several contacts within static analysis community/companies
- Local spin-out company Contemplate Ltd may provide data

Situation Awareness

Security companies are gathering *vast* amounts of information from their products (“security telemetry”): log data, packet capture, incoming connections, etc.

They also collect malware samples, run these in sandboxes, collect traces.

Questions are:

- Can we detect new attacks as well as known ones?
- Can we classify malware traces into *families*?

Data available from:

- Two network security companies (pending)
- Research access to Shadowserver Foundation
- Open data sets exist (e.g. for pcap data)

See *Big Data: Cyber Security’s Silver Bullet?* article in Forbes, Nov 2014.

Anomaly Detection in Financial Systems

We’ve recently had discussions with financial companies interested in the use of **data visualisation techniques** to help detect/highlight/explain security anomalies.

Likely that standard visualisation techniques may be used, but interesting challenge will be getting data into a suitable form, probably involving compression, amplification of appropriate features, etc.

Summary

Security and Privacy

- A fun area to work in, covering a vast range of aspects from purely theoretical to very applied.
- Numerous security-specific conferences and workshops, but also strong interest in S&P aspects in other specialist domains.
- Media-friendly topic, good chance of getting mention in newspapers and science magazines, etc.
- Area inherently multi-discplinary: ultimately involving people, organisations, etc, hence psychology, sociology, economics, politics, law.
- Excellent future career opportunities in industry, academia, start-ups.

Talk to me about specific ideas, think of your own, talk to other people.

Browse some papers

Some top research venues I recommend looking at papers from (range of research styles and topics):

- *IEEE “Oakland” Security and Privacy,* ***S&P***
- *ACM Conf. on Computer and Communications Security,* ***CCS***
- *Computer Security Foundations,* ***CSF***
- *ACM Symp. on Usable Security and Privacy,* ***SOUPS***
- *USENIX Security Symposium*
- *European Symp. on Research in Computer Security,* ***ESORICS***

There are many more venues, e.g., dedicated to forms of cryptography, network/wireless security, systems and programming security, privacy, etc. And many specialised workshops, tracks in other Informatics conference topic areas.

Pointers

Security & Privacy is a strategic growth area in Informatics, and we are founding a new multidiscplinary centre in the University.

Visit (and join!)

- `http://secpriv.inf.ed.ac.uk`
- `http://cybersecpriv.ed.ac.uk`

Other notes:

- Hot security topic areas: Internet-of-Things, wearable tech, autonomous vehicles, health care, mobile malware research, usable security and privacy, secure programming and verification, self-repair, automatic/proactive defence.
- Some PhD topic ideas: `http://secpriv.inf.ed.ac.uk/phds`
- People I particularly recommend talking to about S&P/collaborative topics: **Arapinis**, **Cheney**, **Rovatsos**, **Sannella**, **Sarkar**, **Stark**, **Sutton**.
  See: `http://secpriv.inf.ed.ac.uk/people`