Imperial College London

The Alan Turing Institute

Loughborough University

Adaptive Experimental Design for Intrusion Data Collection

Kate Highnam[1,2], Zach Hanif[3], Ellie Van Vogt[1], Sonali Parbhoo[1], Sergio Maffeis[1], Nicholas R. Jennings[4]

[1]Imperial College London, [2]The Alan Turing Institute, [3]Independent Researcher, [4]Loughborough University

Imperial College London

Summary

1. Issues with Empirical Studies
2. Inspiration from Healthcare
3. Adaptive Design
4. Example Study

Contributions

1. The first adaptive method for a control study in security

- Optimizing resource allocation and duration
- Based on the events seen and error tolerance.

2. The first interventional study using honeypots

- Applying our method in the real world
- Demonstrates claims above during data collection

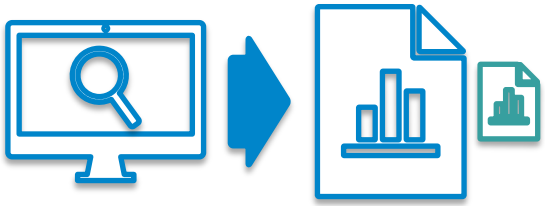
Imperial College London

Intrusion Data Collection

Imperial College London

Observational vs. …

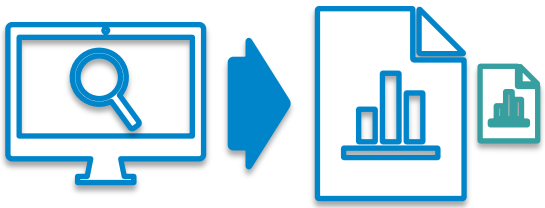
- Easily available to acquire (purchase or record)
- High potential for bias due to uncontrolled characteristics



Imperial College London

Observational vs. Interventional

- Easily available to acquire (purchase or record)
- High potential for bias due to uncontrolled characteristics



- Relatively difficult to acquire
- Controls characteristics studied
- Limit possible spurious correlations between variables and outcomes
- Includes “counterfactual” group

Imperial College London

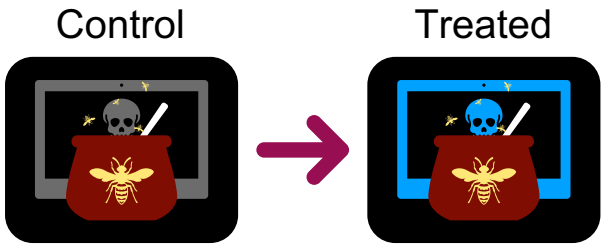
Interventional Study

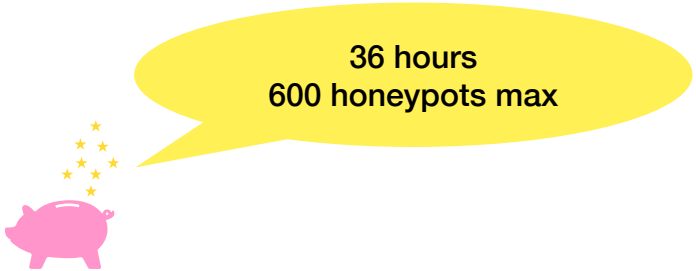
- Population
- Treatment
- Control Group
- Duration
- Objectives
- Event of Interest
- Endpoints

Imperial College London

Interventional Study

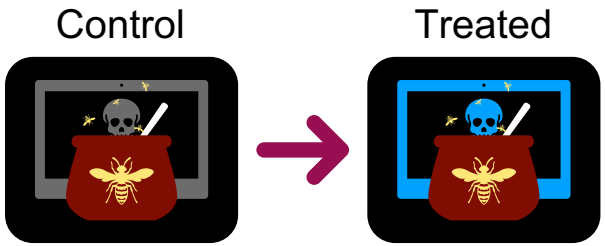
- Population
- Treatment ⟸ “intervention”
- Control Group
- Duration
- Objectives
- Event of Interest
- Endpoints





Imperial College London

Interventional Study

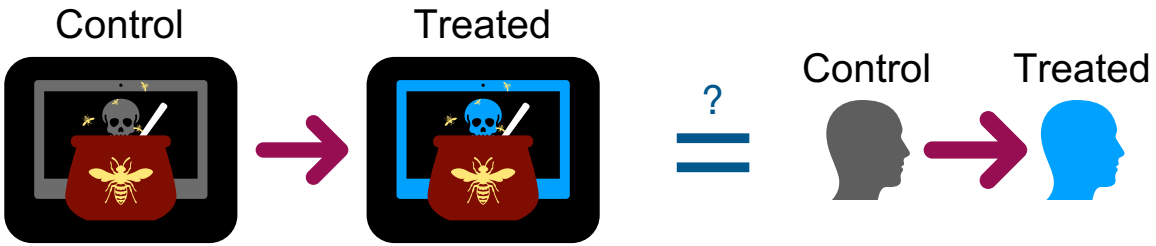


- Population
- Treatment
- Control Group
- Duration
- Objectives ⇦ Did the treatment do anything?
- Event of Interest ⇦ Exploit occurred on host
- Endpoints ⇦ Run out of money?

Imperial College London

Interventional Study

- Population
- Treatment
- Control Group
- Duration
- Objectives
- Event of Interest
- Endpoints



Imperial College London

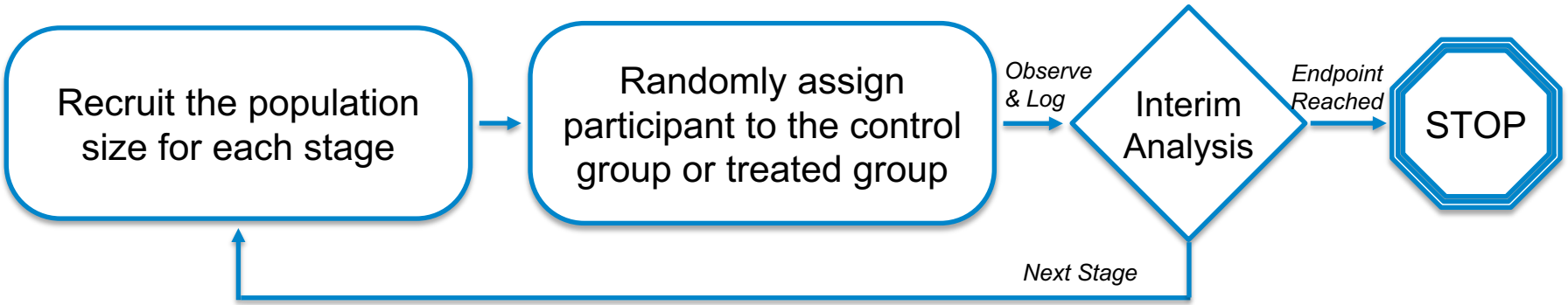
Healthcare to Intrusion Data Collection with Honeypots

Healthcare	Security
“Trial”	“A study comparing honeypots with and without a vulnerability”
“Study population”	“Our Ubuntu honeypots with our host-based sensors”
“Patient” or “participant”	“A honeypot”
“Recruiting more subjects”	“Starting more honeypots with specific characteristics”
“Disease”	“Attacker technique used to exploit”
“**Intervention**” or “**treatment**”	“**Corruption**” or “the presence or insertion of a vulnerability”
“**Treated**”	“**Corrupted**” or “made vulnerable”

Imperial College London

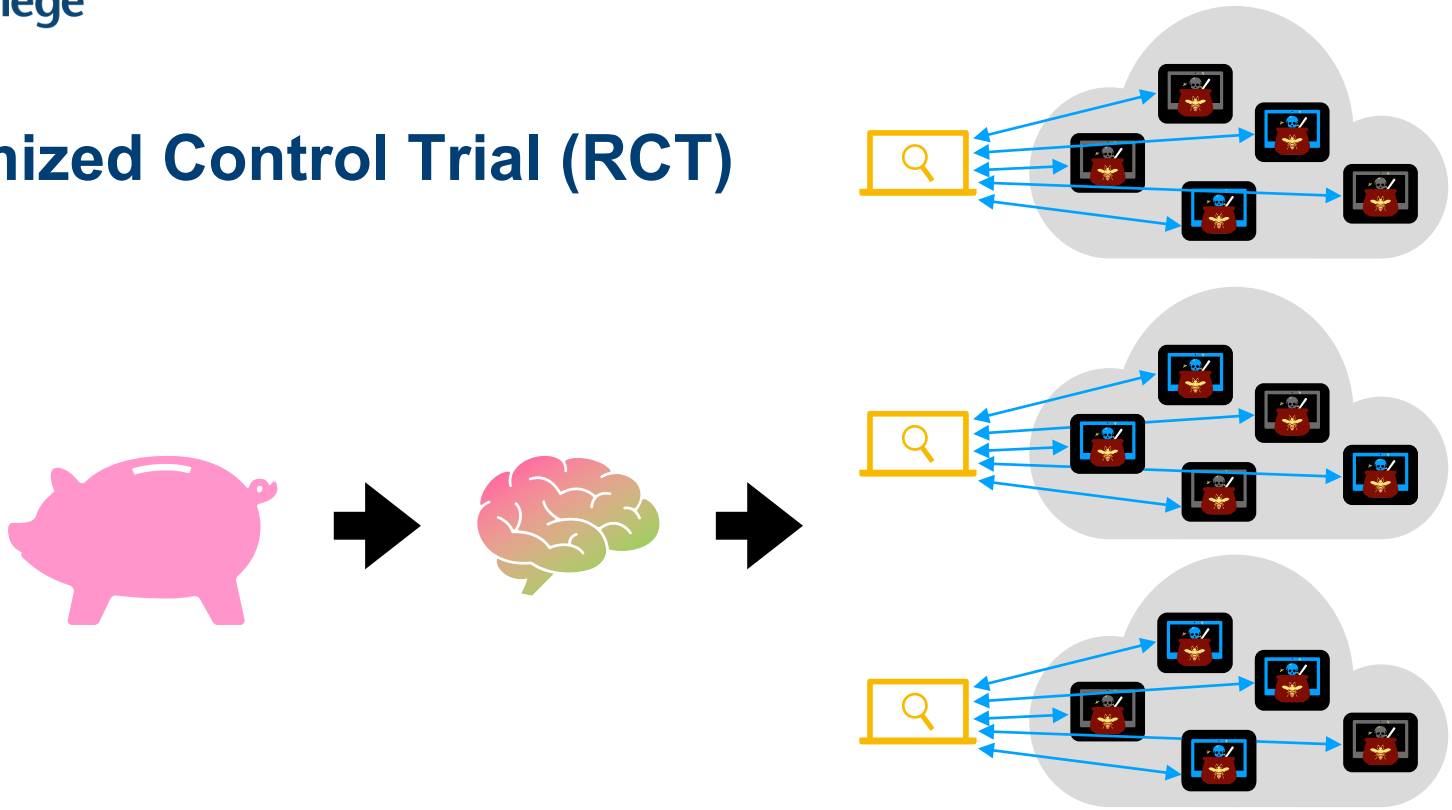
Randomized Control Trial (RCT)

The “gold standard” for clinical trial research:



Imperial College London

Randomized Control Trial (RCT)



Imperial College London

Our Adaptive Design (AD)

★ Achieve the same confirmation from RCT on intervention effect

★ Adapt the deployments based on observed trends

★ Encourage data collection on events of interest

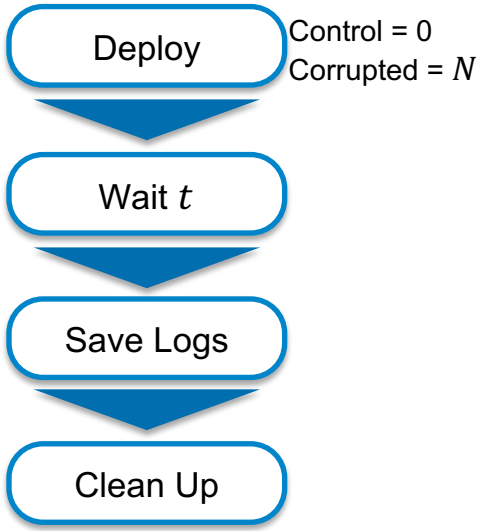
Imperial College London

Comparing Trial Methods

Observational Study

Inputs:

- b = Budget for trial
- t = **Trial** Duration (in hours)



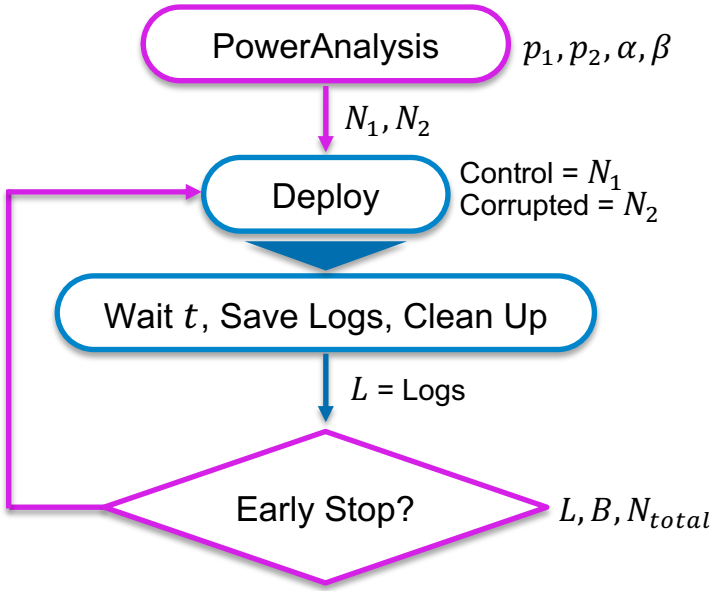
Imperial College London

Comparing Trial Methods

Randomized Control Trial

Inputs:

- b = Budget for trial
- t = **Stage** Duration (in hours)
- α = The probability of committing a **Type I error**
- β = The probability of committing a **Type II error**
- p_1 = Proportion of **control** group getting exploited
- p_2 = Proportion of **corrupted** group getting exploited



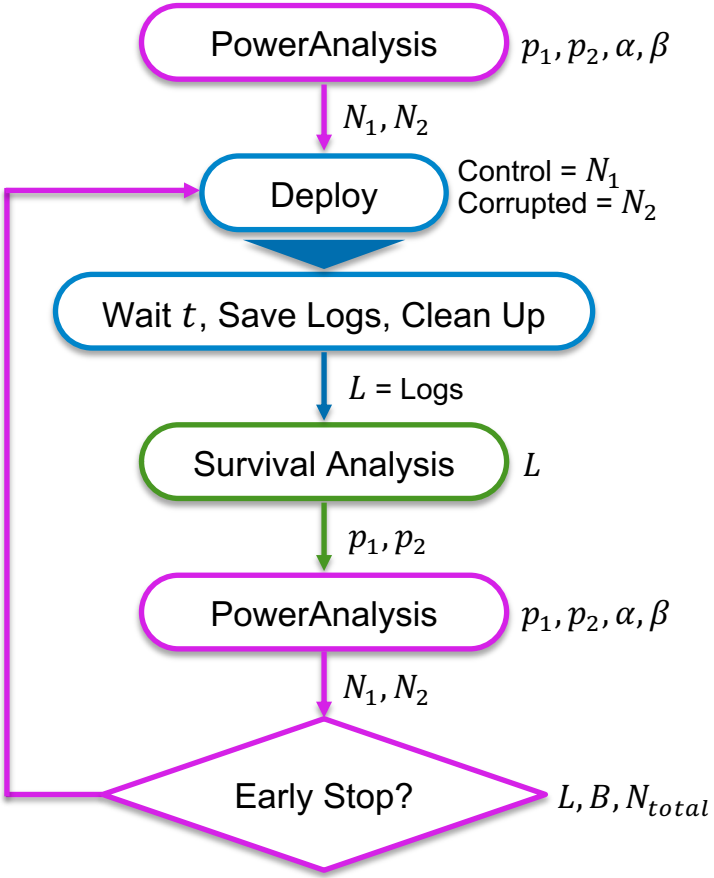
Imperial College London

Comparing Trial Methods

Adaptive Design

Inputs:

- b = Budget for trial
- t = **Stage** Duration (in hours)
- α = The probability of committing a **Type I error**
- β = The probability of committing a **Type II error**
- p_1 = Proportion of **control** group getting exploited
- p_2 = Proportion of **corrupted** group getting exploited



Imperial College London

Targeting Data Collection in Interventional Studies

How to encourage collection of intrusions? Alter the **objectives**!

1. Confirm evidence of corruption's impact within the population
2. Maximize the recording of events of interest

Imperial College London

```
wget https://data.hpc.imperial.ac.uk/resolve/\?doi\=9422\&file\=4\&access\= -O full_BETH_dataset.zip
```

Kaggle Dataset

BETH Dataset

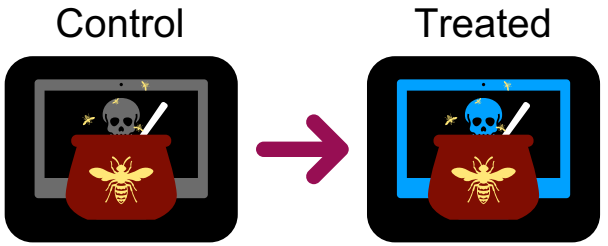
Real Cybersecurity Data for Anomaly Detection Research

Kate Highnam, Kai Arulkumaran, Zachary Hanif, Nicholas R. Jennings

Imperial College London · The Alan Turing Institute · ARAYA · University of Maryland · Loughborough University

Imperial College London

Example: Honeypot Study



- **Population:**
  - Cloud-based Ubuntu servers; hosted by same cloud provider
  - Four regions within the US; Randomly assigned IP ranges based on region
- **Corruption**: SSH vulnerability - accepting any password for four IT user accounts
- **Control Group**: same IT user accounts only accept ‘password’ as the password
- **Duration:** 12 hours per trial
- **Objectives:**
  1. Determine if corruption significantly increases successful SSH logins
  2. Maximize exploitation rate across regions
- **Event of Interest**: User login is successful in one of the four user accounts
- **Endpoints**: Maximum number of honeypots reached (200 per trial).

To limit error, set $\alpha = 0.05$ and $\beta = 0.10$

Imperial College London

Example: Honeypot Study

Method for Trial	Control	Corrupted	Total Deployed	Total Attacks Seen
Vanilla	0	140	140	**137**
RCT	72	72	144	42
AD	32	87	**119**	50

Imperial College London

Example: Honeypot Study

$$p_1 = 0.01 \rightarrow p_1 = 0.15$$

	Region	RCT	AD
Stage 1 - Total		48	48
Control	east-1	6	6
	east-2	6	6
	west-1	6	6
	west-2	6	6
Corrupted	east-1	6	6
	east-2	6	6
	west-1	6	6
	west-2	6	6
Stage 2 - Total		48	52
Control	east-1	6	4
	east-2	6	4
	west-1	6	0
	west-2	6	0
Corrupted	east-1	6	8
	east-2	6	12
	west-1	6	8
	west-2	6	16
Stage 3 - Total		48	19
Control	east-1	6	0
	east-2	6	0
	west-1	6	0
	west-2	6	0
Corrupted	east-1	6	2
	east-2	6	5
	west-1	6	8
	west-2	6	4
Trial - Total		144	**119**
Total Attacks		42	**50**

Imperial College London

Example: Honeypot Study

$$p_1 = 0.01 \rightarrow p_1 = 0.15$$

$$p_1 = 0.15 \rightarrow p_1 = 0$$

	Region	RCT	AD
Stage 1 - Total		48	48
Control	East-1	6	6
	East-2	6	6
	West-1	6	6
	West-2	6	6
Corrupted	East-1	6	6
	East-2	6	6
	West-1	6	6
	West-2	6	6
Stage 2 - Total		48	52
Control	East-1	6	4
	East-2	6	4
	West-1	6	0
	West-2	6	0
Corrupted	East-1	6	8
	East-2	6	12
	West-1	6	8
	West-2	6	16
Stage 3 - Total		48	19
Control	East-1	6	0
	East-2	6	0
	West-1	6	0
	West-2	6	0
Corrupted	East-1	6	2
	East-2	6	5
	West-1	6	8
	West-2	6	4
Trial - Total		144	**119**
Total Attacks		42	**50**

Imperial College London

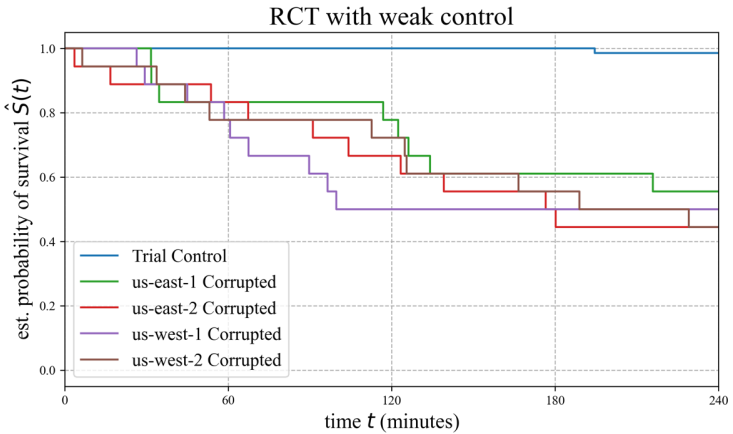
Example: Honeypot Study

	Region	RCT	AD
Stage 1 - Total		48	48
Control	east-1	6	6
	east-2	6	6
	west-1	6	6
	west-2	6	6
Corrupted	east-1	6	6
	east-2	6	6
	west-1	6	6
	west-2	6	6
Stage 2 - Total		48	52
Control	east-1	6	4
	east-2	6	4
	west-1	6	0
	west-2	6	0
Corrupted	east-1	6	8
	east-2	6	12
	west-1	6	8
	west-2	6	16
Stage 3 - Total		48	19
Control	east-1	6	0
	east-2	6	0
	west-1	6	0
	west-2	6	0
Corrupted	east-1	6	2
	east-2	6	5
	west-1	6	8
	west-2	6	4
Trial - Total		144	**119**
Total Attacks		42	**50**

Imperial College London

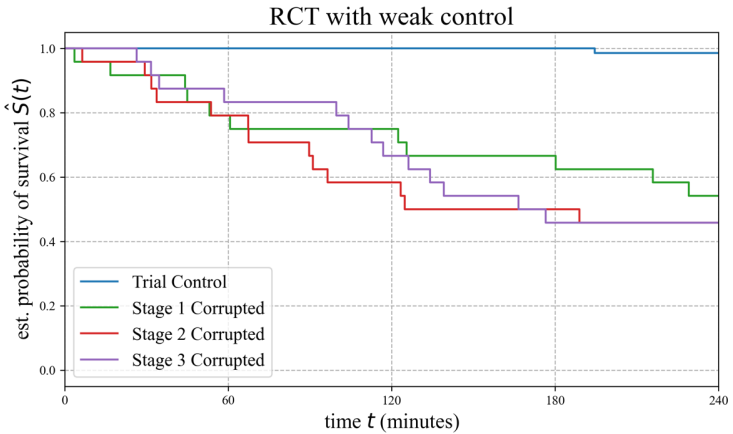
Example: Honeypot Study

By Region



Staggered exploitations?

By Stage



Possible signs of instability

Imperial College London

Conclusions and Future Work

- First to apply adaptive experimental study in intrusion data collection
- Provide general details on how to run other intrusion-focused experimental studies
- Successfully identify causal relations between a corruption and events of interest - collecting data with true relations between features to learn general trends
- Our honeypot study showed our AD can confirm corruption effect at 33% of total trial duration compared to RCT
  - By the end of the trial, our AD used 17% fewer honeypots to see 19% more attacks
- Future work:
  - Implement multiple vulnerabilities to study the interaction of corruptions
  - Use our method to remove bias in a dataset → demonstrate improved model learning

Imperial College London

Kate Highnam

https://www.imperial.ac.uk/people/k.highnam19

@jinxmirror13