Bringing Science to the Evaluation of Malware Forensics Analysis Tools

Dr Ian Kennedy BEng(Hons) PGCE CITP CEng FBCS

My Background

- Software developer / Commissioning
  - Public sector: In-house within NHS trusts
  - Commercial: Servicing healthcare sector
  - Commercial: Finance sector (US)
- Digital Forensic practitioner
  - Law enforcement (Kent Police)
  - Consultancy (Control Risks)
  - UK Government
- Academic
  - The Open University
  - Canterbury Christ Church University
  - Member of peer review Board for *Digital Investigation*
- BCS Roles
  - Fellow of BCS (as of March 2019 - Yay!)
  - Contributor/Author
  - BCS Assessor (CEng/CITP)

Overview

- Background
- Prior work
- Framing the problem – the RQ
- A solution : the MATEF
- Interpreting the data
- Results
- Conclusions
- Contributions and further work

Background to the problem

Who in this room doesn't use the Internet?

Not so long ago in a Police building not so far, far away…

The Open University

One day at work....

Background

The Open University

The case of Nicholas GRANT: Royal Collage of Physicians[1]

- >700 IIOC
- 24 counts of IIOC
- Malware found
- Trojan defence

- Light-touch analysis
- Conclusion: IIOC not attributed to malware
- Court were convinced. As a scientist, was I?

Background

1: http://www.kentonline.co.uk/canterbury/news/top-managers-child-porn-shame-a19813/

Other examples

Source: cbc.ca

Source: youtube.com

Michael FIOLA

Julie AMERO

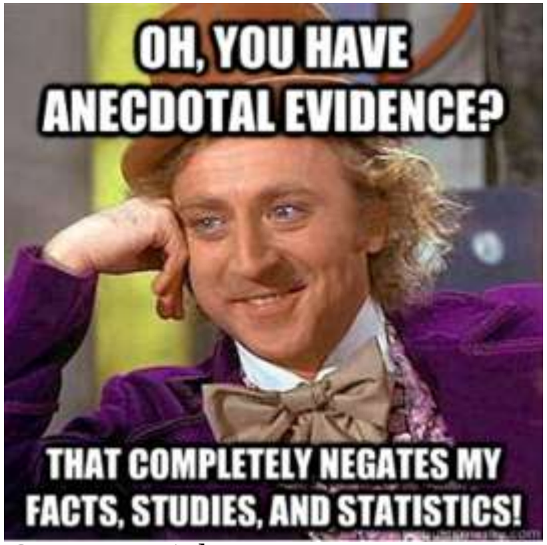
Wider issues with Expert Evidence

- Trojan Defence
- Unfounded trust repeated confirmation
- Expert evidence problems
- *Lack of scientific underpinning*
- *Reproducibility flaws*
- *Acceptance of fact*
- Statutory requirements

The Open University

Research justification

- **Unfounded trust repeated confirmation**



Source: quickmeme.com

"haven't seen a single case" [1]

"Yet to see an example" [2]

Source: docdata.co.uk

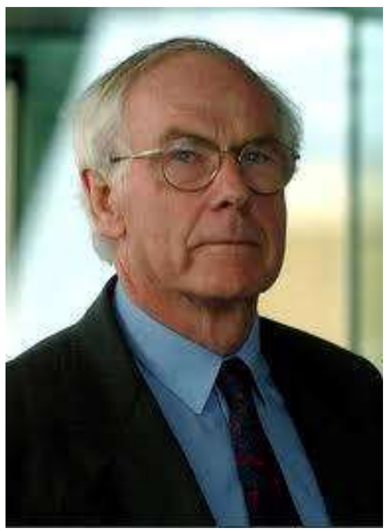
"solely on reputation of the vendor" [3]

Background

1: McLinden (2009) 2: Douglas (2007) 3: Garfinkle *et. al.* (2009)

Research justification

- **Expert evidence problems**
  - Judges have no test to “gauge unreliability”[1]



Source: whale.to

Prof. Sir Roy Meadows

Source: wikimedia.org

Casey Anthony case

Background

1: Solicitors Journal (2011)

Research justification

- **Lack of provenance**
  - Individualisation:
    - Not "rigorously shown" to be reliable [1]
  - Malware forensics:
    - Hostile nature of malware
    - Analysis skills
    - Repeatability

1: National Academy of Sciences (2009)

Research justification

- **Reproducibility flaws**

  Dual-tool verification

  – Unsupported claims:
    - Can "confirm result integrity"[1]
    - Allows "verification of findings"[2]

  – Misuse of term 'verification'
    - Dual-tool can *corroborate*, not *confirm*
    - Should use **reference** point[3]
    - Should be **statistically** significant

  – Good for finding discrepancies[4] – Falsification!

[1] Forensic Control (2011) [2] Cy4or (2009) [3] VIM JCGM 200:2008 [4] Turner (2008)

Research justification

- **Acceptance of fact**

TRUTH

Criminal Justice System	Sciences
Revealed by Adversarial Process[1]	Revealed by Falsification[1]
Value system: Certainty[2]	Value system: Scepticism[2]
Ad populum[3]	Discovered by iterative process
Win by: Relatable, persuasive	Win by: supporting theory, knowledge

[1] Kritzer (2009) [2] Marsico (2004) [3] Beckett (2010)

Research justification

- **Statutory Requirements**

  Forensic Science Regulator
  - ISO 17025
    - Codes of Practice
    - October 2017 deadline
  - Requirements include:
    - Validation
    - Peer review
    - Generally accepted

Background

The Open University

What has been done to address this?

Prior Work

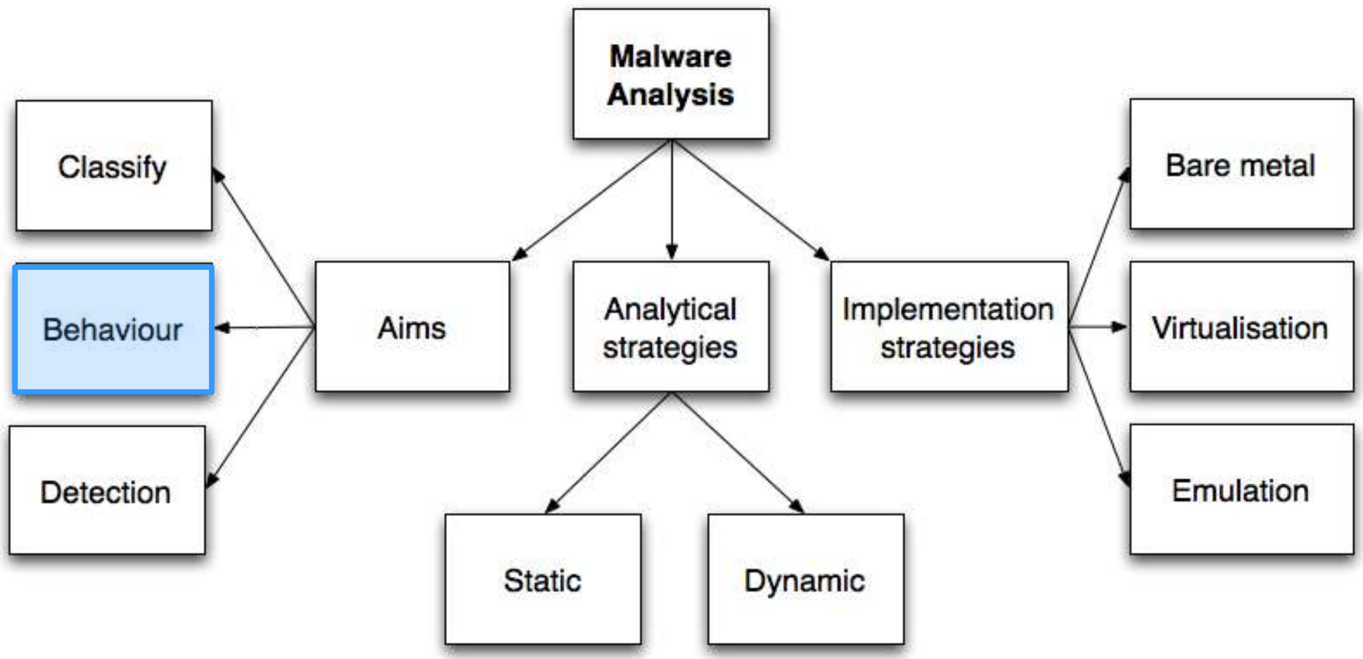
- Digital Forensic (DF) practice
- Malware Forensics (MF) practice
- Tool evaluation

Prior Work : DF practice

- Heavily cited:
  - DFRWS (2001) : Six stage process model
  - Carrier & Spafford (2003) : 17 phase model (phy+dig)
  - Carrier (2003) : Abstraction layer model
- NIST DF procedure (2006) : Six stage model
- Adopted process "does not exist" [1]
- No standard methodology, including searching for malware[2]

[1] Montasari, Peltola & Evans (2015) [2] Kipling (2012)

Prior Work : MF practice



- Analysis approaches:
  - MF framework[1] – extends Cuckoo sandbox[2]
  - Five phase approach[3] – (Pres./RAM/FA/Static/Dynamic)

[1] Provataki & Katos (2013) [2] cuckoosandbox.org [3] Malin *et. al*. (2008)

Prior work : Tool evaluation

- Evaluation Criteria
  - CFTT, SWGDE, DC3
  - FSR (Validation, Peer review, Generally accepted)
- Little traction of methodologies
  - Slay *et. al*. (2005-10)[1] : Functional – theoretical only
- No consensus on methodology for testing

[1] Modest citations on Google Scholar (<50)

Framing the problem:

The Research Question

Research Question

Can a systematic basis for trusted practice be established for evaluating malware artefact detection tools used within a forensic investigation?

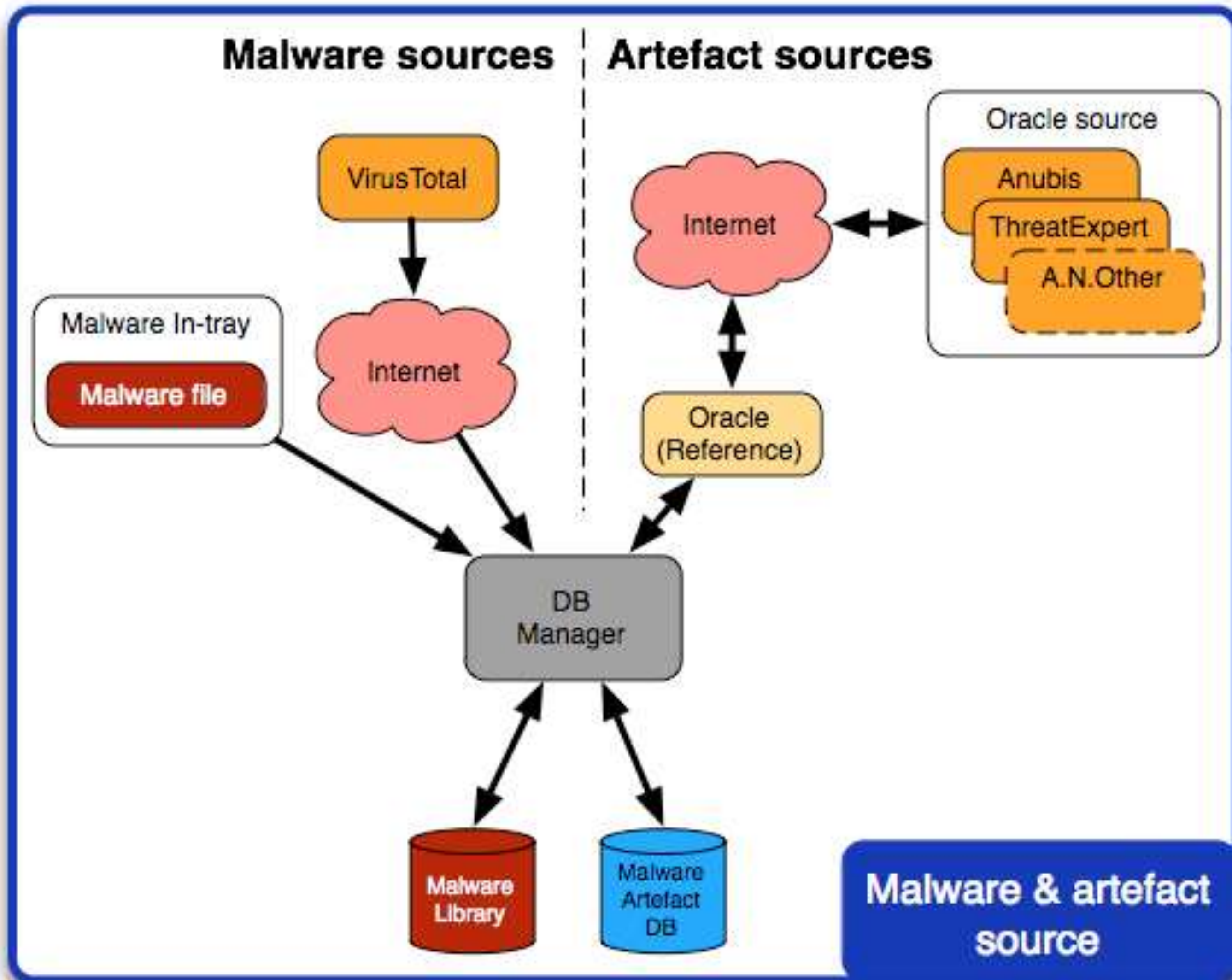
In other words:

Can tools used for malware forensics be scientifically evaluated?

Designing a solution

The
Malware Analysis Tool
Evaluation Framework

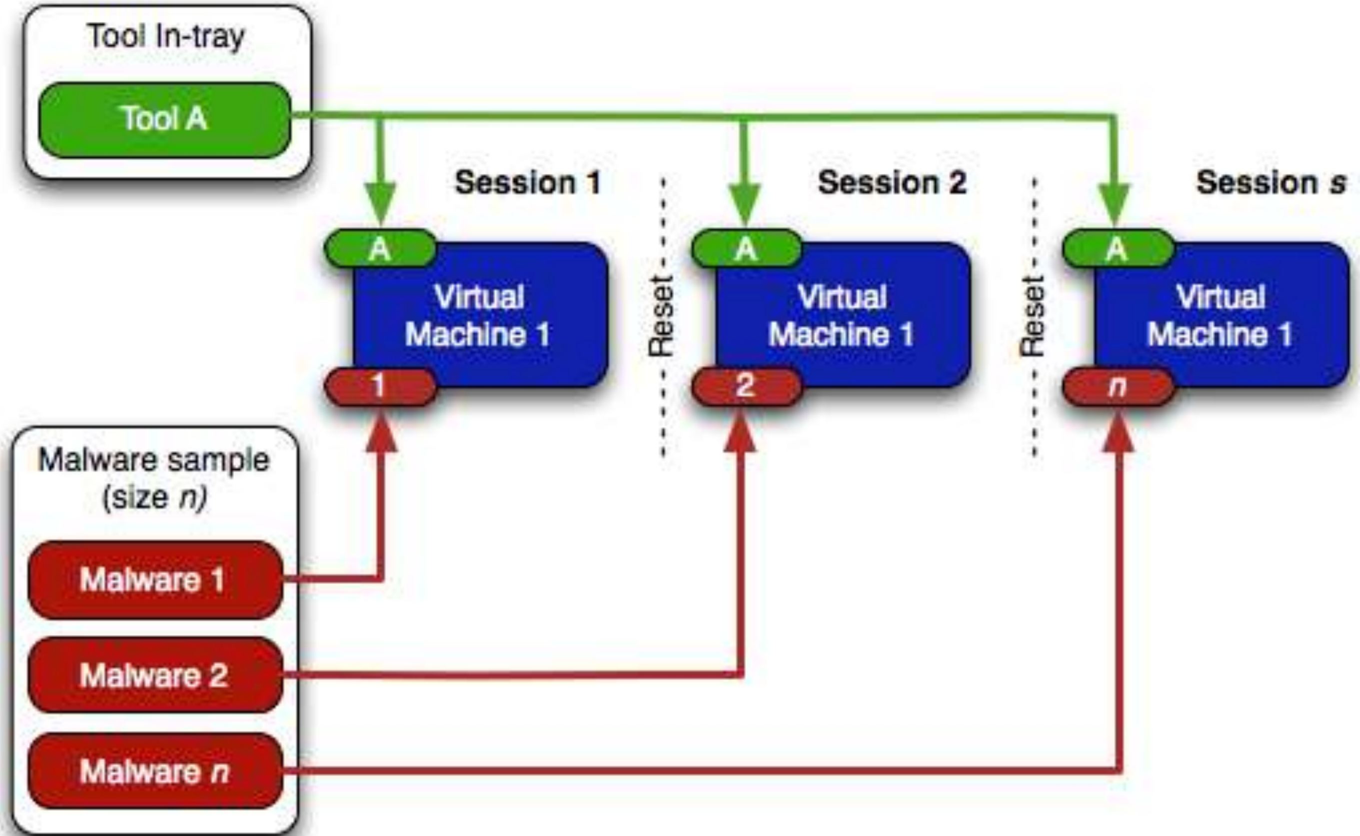
Getting malware and artefacts



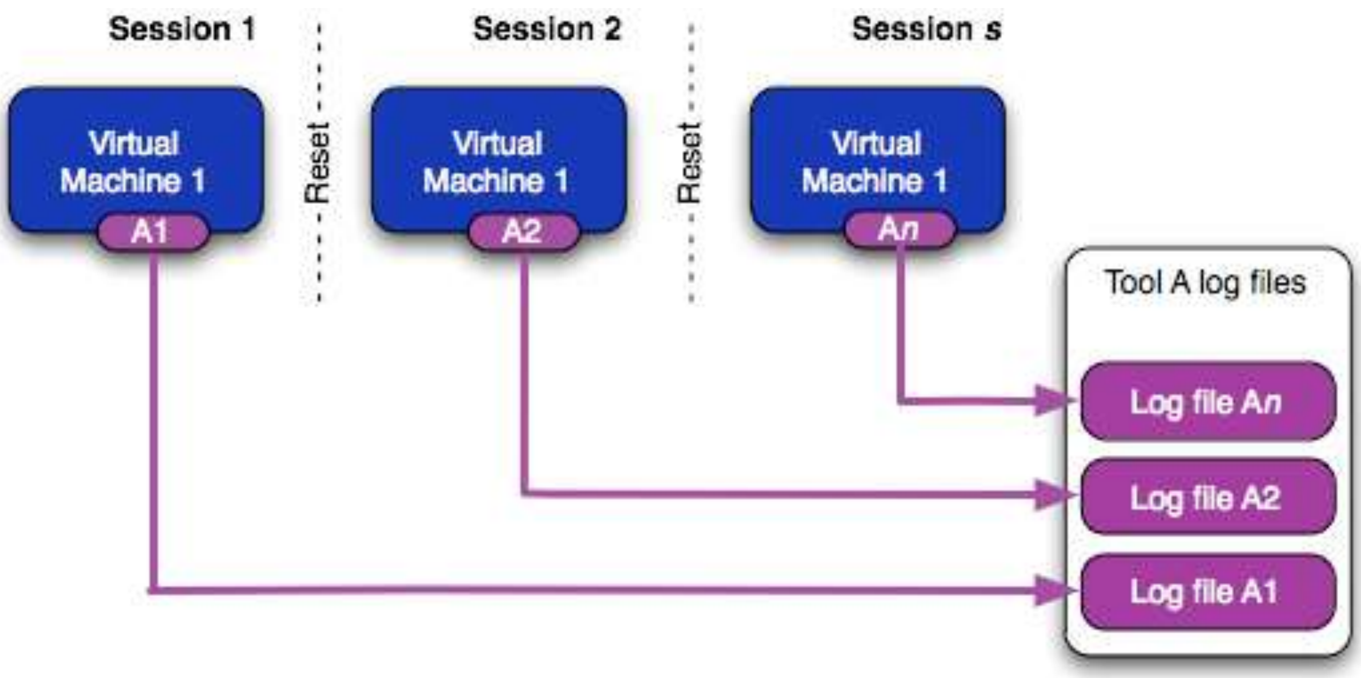
The Open University

Virtual Machine (VM)

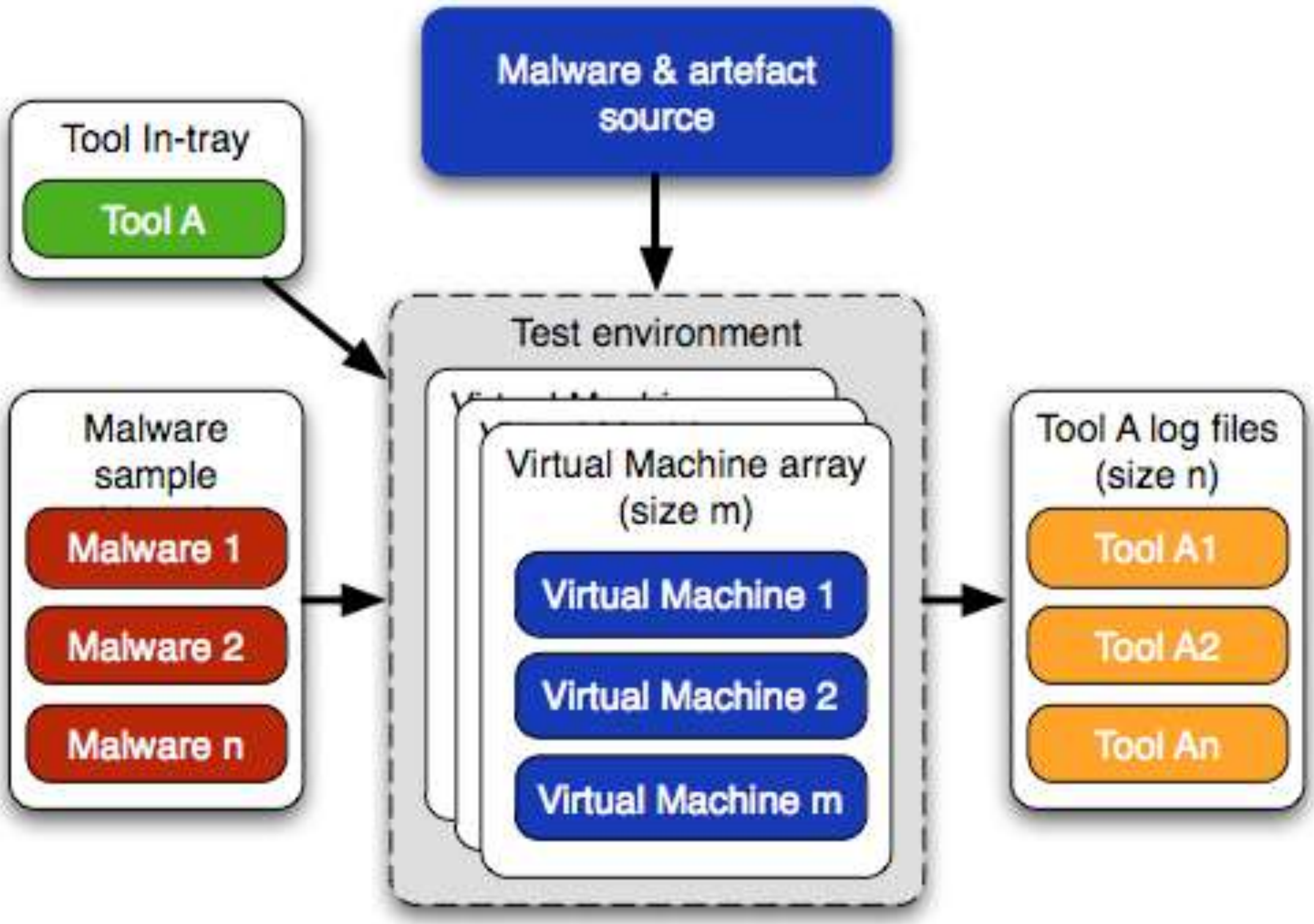
Add the tool, then the malware



Before reset, get tool log files

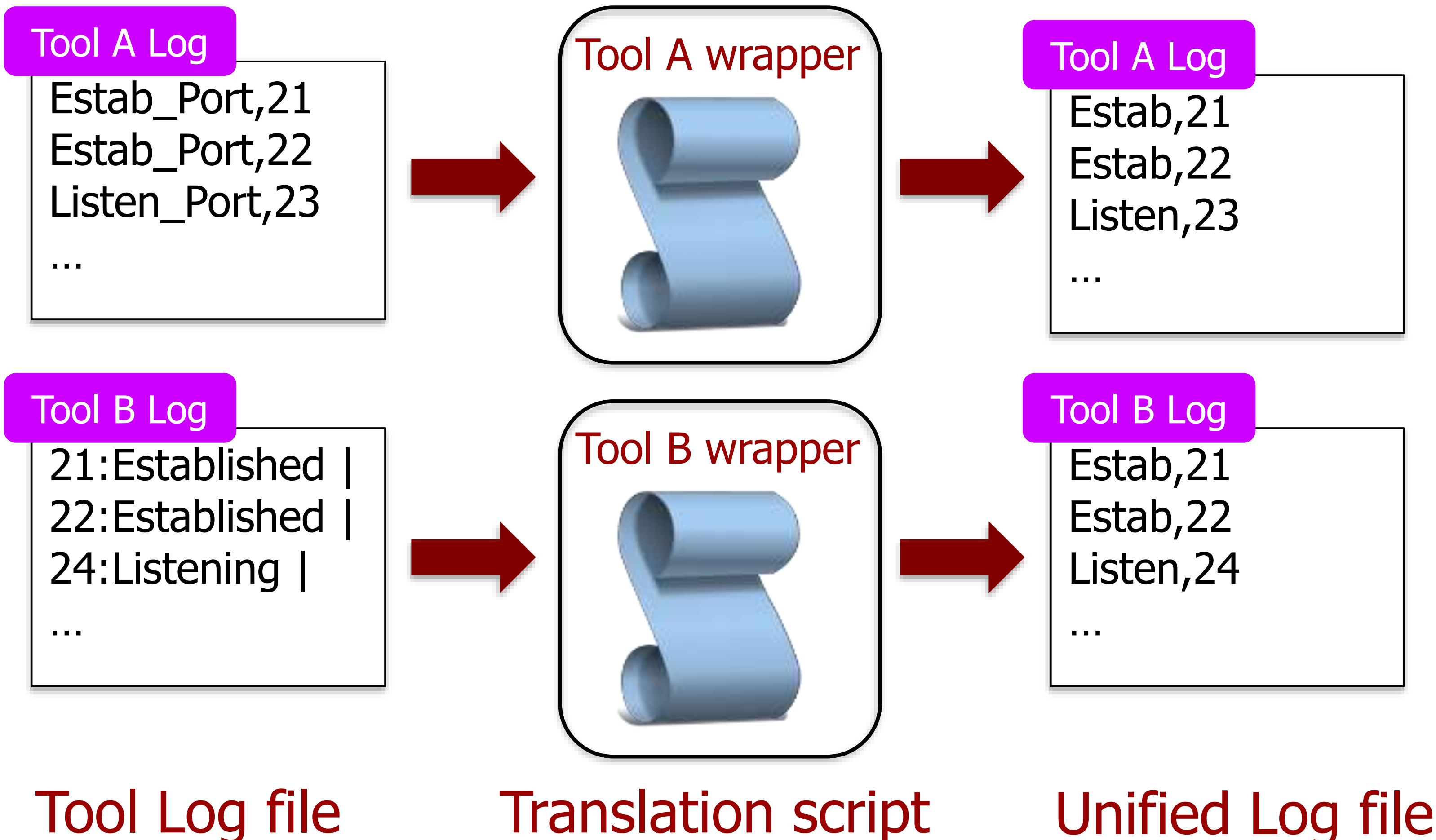


Malware Analysis Tool Evaluation Framework



Analysis methodology

Normalising log files



Interpreting the data

Quantities, not values

Estimated ground truth

Absolute differences

Freq. dist. of differences

Analysis strategy

Results

Study hypotheses

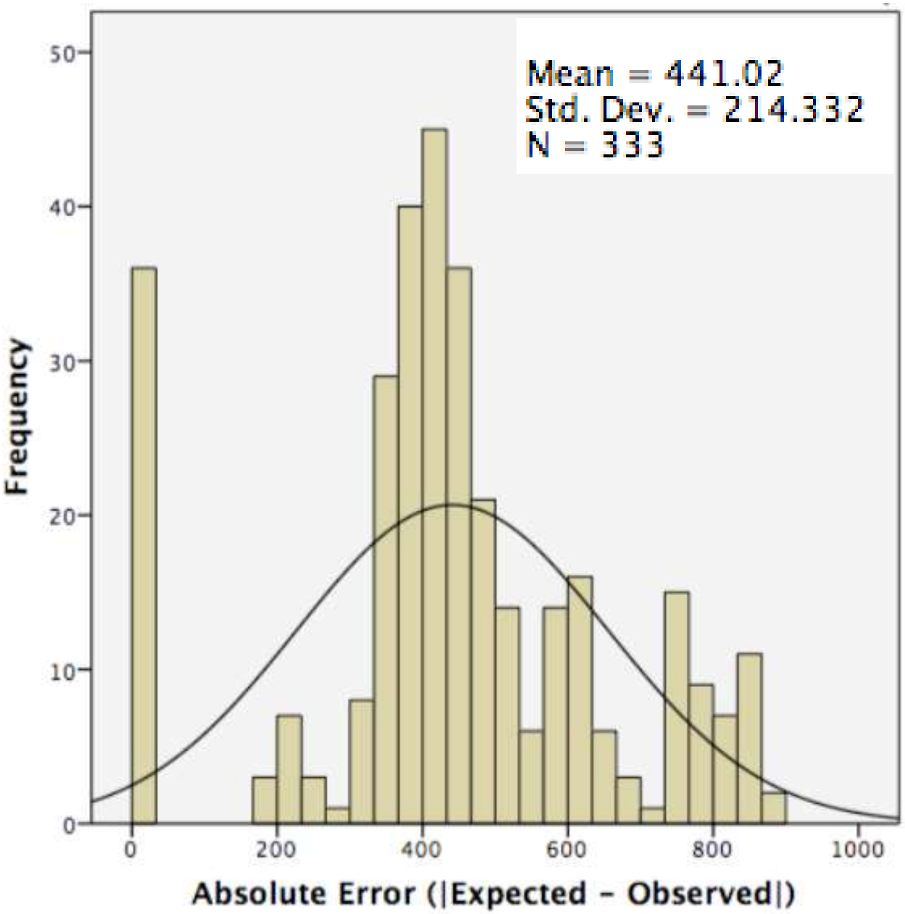
Hypothesis 1

Changing the execution time has no effect on the number of open ports reported by a tool

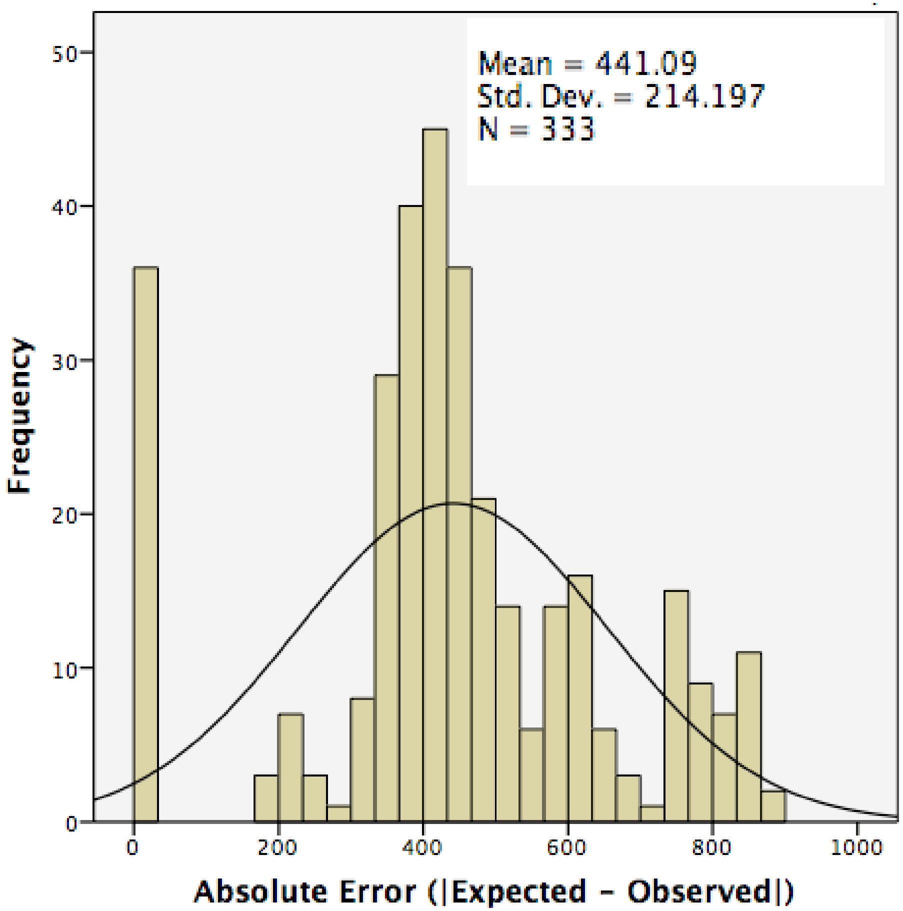
Hypothesis 2

Both tools report the same number of opened ports at a given execution time

Results : Execution times



Process Monitor
(Run for 1 min)



Process Monitor
(Run for 10 sec)

Results : Execution time

Hypothesis 1

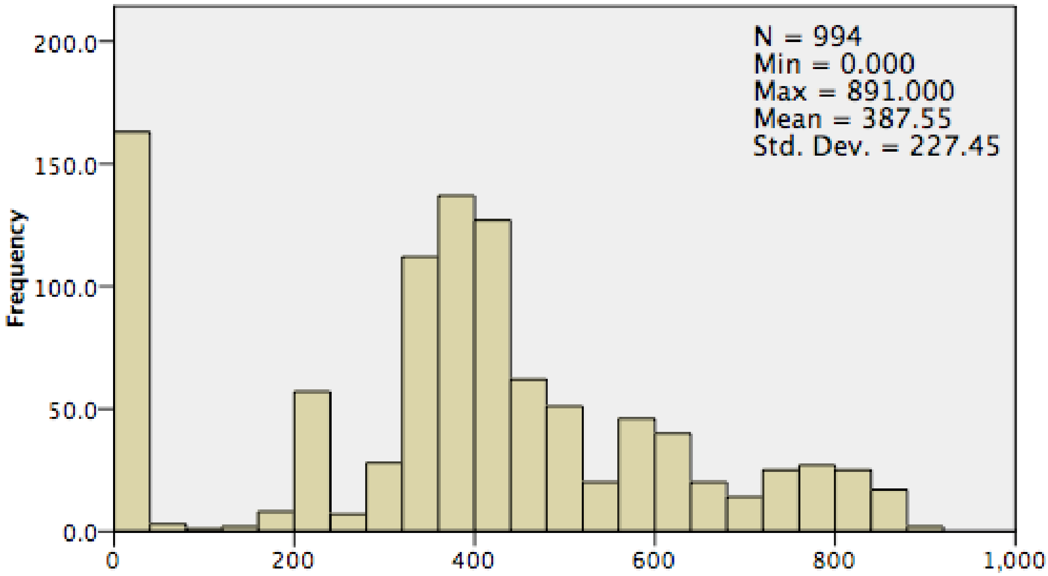
Changing the execution time has no effect on the number of open ports reported by a tool

	10 sec v 1 min	1 min v 5 min	1 min v 10 min
Process Monitor	False	True	True
TCPVCon	True	True	True

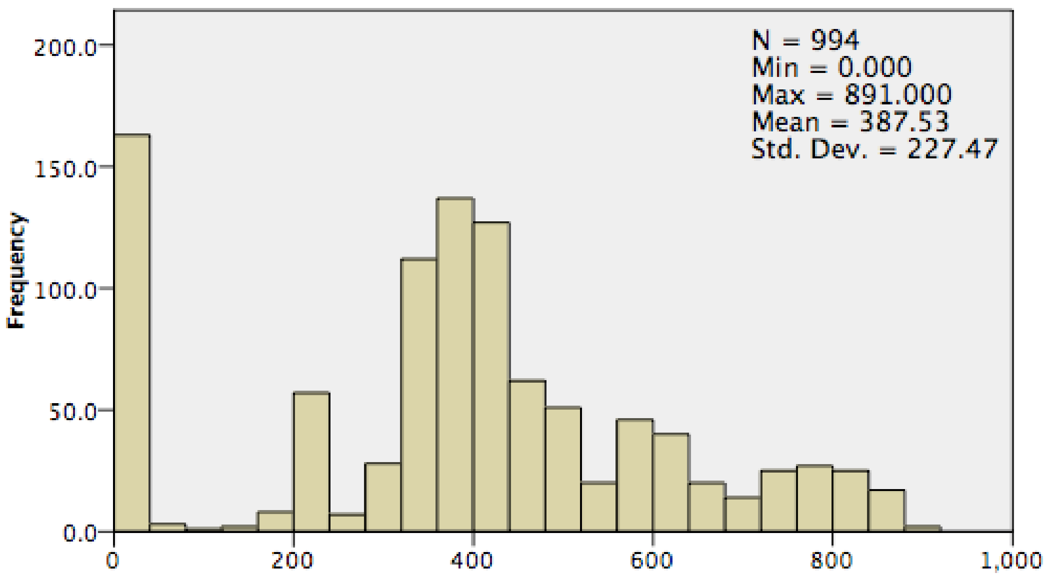
Indicates:
There is a statistically significant difference between 10 sec and 1min

Results : Same execution time

Process Monitor
(Run for 1 min)



TCPVCon
(Run for 1 min)



The Open University

Results : Execution time

Hypothesis 2

Both tools report the same number of opened ports at a given execution time

	10 sec	1 min	5 min	10 min
Process Monitor	True	True	True	True
TCPVCon				
p-value	1.000	0.056	0.157	0.317

Recall:

The *p-value* is the probability of the NULL hypothesis being true (No statistically significant difference between tools) as > 0.05

Conclusions

Study Conclusions

- Tool & run time impacts outcome
- Minimum execution time
- No benefit if run > 1min
- Impact:
  - Reduce testing time
  - Introduced quantifiable measure of uncertainty (statistical levels of confidence)

Research conclusions

Research goals

Scope limitations

Method limitations

Research contributions

- Evidence of a lack of trusted practice
- Framework to evaluate new tools
- Requirements to establish trusted practice
- Results of studies on tools
- MATEF performance data
- Methodology to set test time parameters

Further work

- In-house Oracle
- GUI based tools
- Performance
- Bare metal
- Malware ingestion
- Statistical module
- Outstanding requirements

Review

- Background
- Prior work
- Framing the problem – the RQ
- Design of a solution
- Interpreting the data
- Results
- Conclusions
- Contributions and further work

Thank you

Questions?

Ian Kennedy

Ian.Kennedy@canterbury.ac.uk

Ian.Kennedy@bcs.org.uk