EVALUATING ANTI-VIRUS PRODUCTS WITH FIELD STUDIES

Fanny Lalonde-Lévesque,
Carlton R. Davis,
José M. Fernandez,
Anil Somayaji

POLYTECHNIQUE MONTRÉAL
LE GÉNIE EN PREMIÈRE CLASSE

Carleton UNIVERSITY

Agenda

- Introduction
- Description of the study
- Results of the study
- Viability of the field study approach
- Conclusions

Introduction

- Current AV evaluation methods
  - Are based on automated tests in controlled environment
  - Do not account for user behaviour
  - Do not account for user "environment"
  - The effectiveness of the products against yet-to-be-discovered threats is not being evaluated

➔ Idea: Conduct AV evaluation as a "clinical trial" with real users (Somayaji et al. CSET 2009)

Description of the study: the goals

1. Test viability of "field studies" (aka clinical trials) as an anti-malware evaluation methodology, with a proof-of-concept study
2. Determine how system configuration, environment, and user behavior affect probability of infection
3. Determine how malware is infecting computer systems, and identify sources of malware infections

Description of the study: the participants

Involves 50 participants over a 4 month period

- Recruiting
  - Posters and campus newspaper ads on Montreal campuses
- Candidate selection
  - Short intake questionnaire with demographic information
  - Approximately 100 interested volunteers
  - Random sample selected from each category
- Gender
  - 20 females and 30 males
- Ages of participants
  - 18 to 51+ years
- Language
  - Web pages most frequently visited:
    French: 29, English: 18, Other (Arabic, Chinese, Spanish): 3

Description of the study: equipment

- 50 identical laptops with identical configuration
  - Windows 7 Home Premium OS
  - Trend Micro OfficeScan 10.5
  - Diagnostic tools
    - Hijackthis, ProcessExplorer, Autoruns, tshark, SpyBHORemover, SypDLLRemover, WinPrefetchView and WhatChanged
  - Custom Perl scripts which we developed
- Laptops were sold to participants (at discount price)

Description of the study: baselining laptops

- Laptops baselined before deployment, by recording the following info
  - Hash of all the files
  - Info about file signature (when applicable)
  - Auto-start programs
  - List of processes
  - Registry keys
  - Browser helper objects (BHO)
  - Files loaded during boot
  - Pre-fetch files

Description of the study: the procedure

- 5 in-person sessions
  - An initial session where we supplied the laptops
  - 4 monthly 1-2 hour sessions
    - Participants fill out online questionnaire
    - We analyse the laptops and collect the statistical data
    - Exit survey at the end of the final monthly session
- Compensation
  - Participants initially purchased the laptop (discount price)
  - Participants paid for each session attended + completion bonus
  - ➔ End result: laptop for free…
- Rules
  - Participants encouraged to configure and use the laptops as they desire, i.e. their laptops
  - Not allowed to change AV or deactivate scripts and tools during study

Description of the study: compiled data

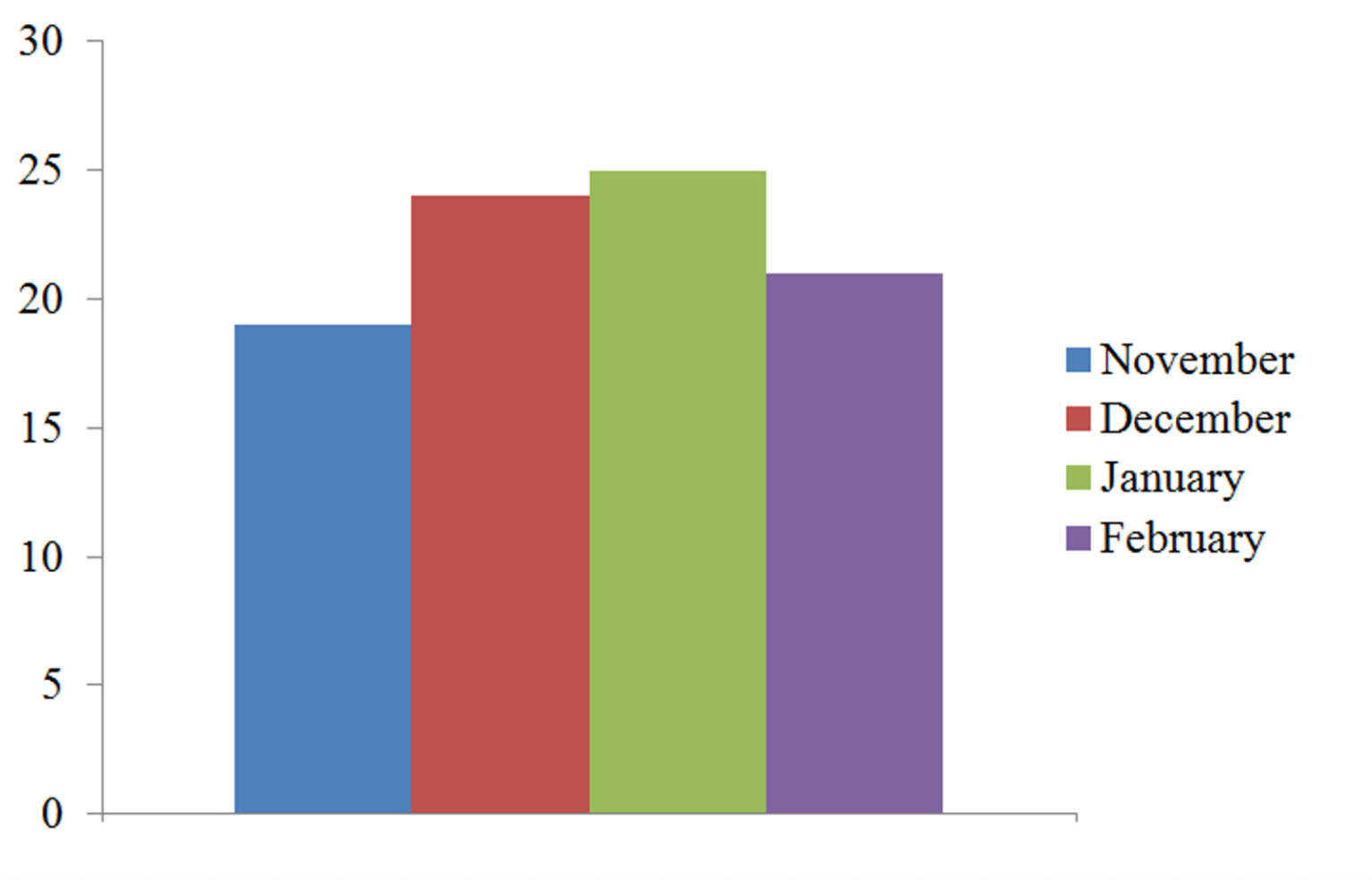
Data compiled every month (through scripts)

- Configuration
  - List of applications installed
  - Percentage of applications with latest update applied
- Internet connection
  - Number of hours connected (per day)
  - Number of locations from which connected (per day)
  - Number of hosts connected to (per day)
- Web browsing and usage
  - Number of web sites visited per category
  - Types of browser used and frequency of use
  - Number and the types of files downloaded

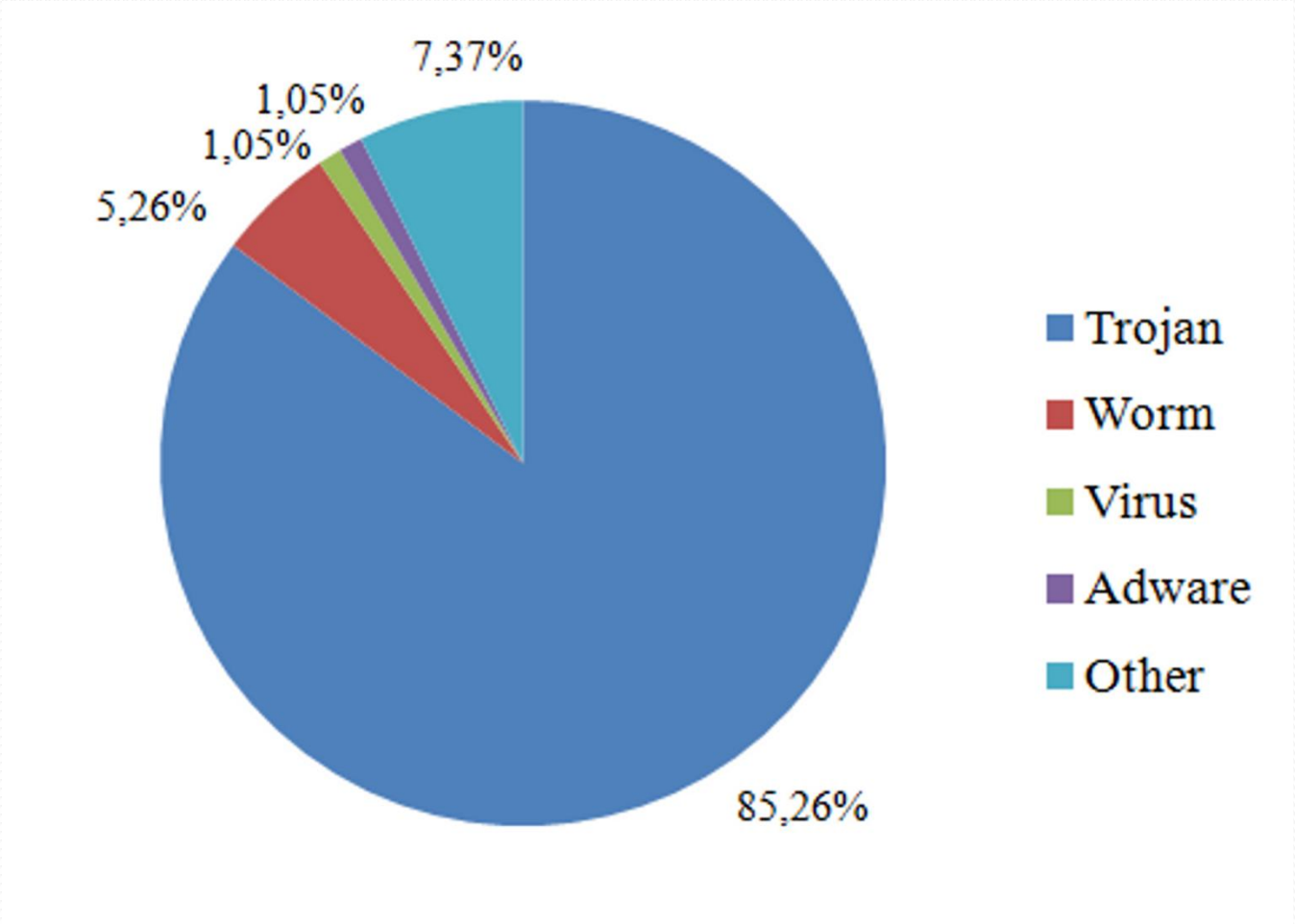
Description of the study:
Suspected infection protocol

- Pre-determined protocol for identifying infection
  - Unexplained registry entries
  - New suspicious files
  - Virus Total
  - ….
- When infection identified or suspected
  1. Request consent to investigate further
  2. If consent granted, collect additional data
     - List of web sites visited during time window of infection
     - List of all hosts connected to within time window of infection
     - Copy of all suspected infected files

Results: Threat detections by AV



Results: Malware detection by type



Results: Missed detections

- Found 20 possible missed detections on 12 different laptops
  - 2 were classified as “clean”
  - 7 were unwanted software
  - 9 were adware
  - 2 was classified as “definitely malware” (1 missing file)
- Detection trigger
  - 18 – HijackThis (registry and file)
  - 1 – SpyBHORemover (BHO)
  - 1 – User reporting (suspicious activity)

Results: Detection statistics

- Detection totals
  - 95 detections by AV
  - 18 missed detections (2 confirmed malware)
- Detection rates
  - Counting unwanted software & adware:

➔ 84% true positive, 16% false negative

  - Counting confirmed malware only

➔ 98% true positive, 2% false negative

Results: User feedback

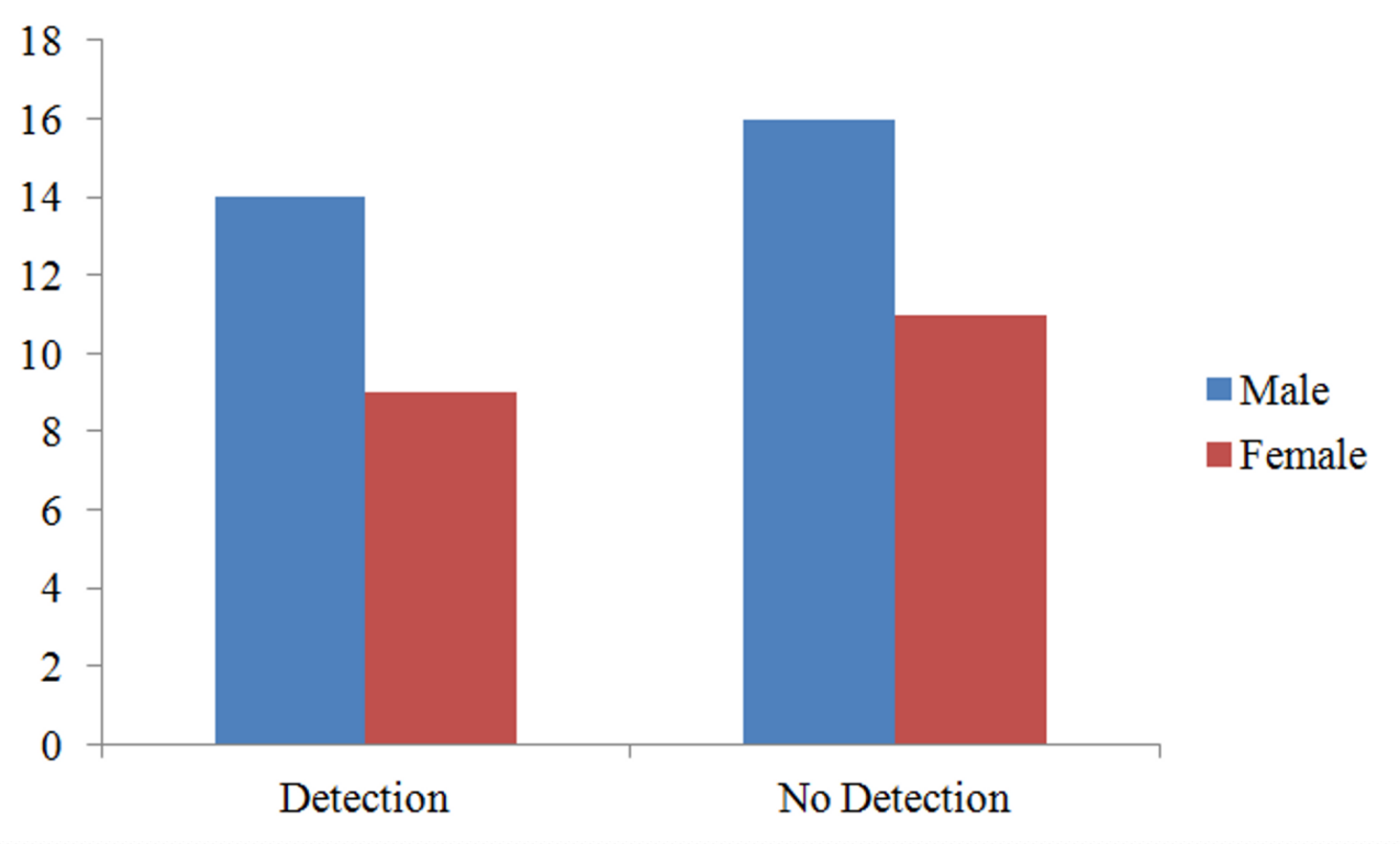
- User feedback in cases of AV detection
  - "Did you observe strange computer behaviour"?
    - 55% NO, 40% YES and 5% "don't know"
  - Most frequent observed behaviour
    - Performance decrease
    - pop-up windows,
    - problems with web browsers
      - redirection
      - change of home page
  - "Did you see or notice any special AV behaviour?"
    - 50% noticed a prompt window informing of problem
  - "Are you now concerned about the security of your computer?"
    - 35% YES, 20% "annoyed at the interruption", 15% "confused"

Results: User profiling and behaviour

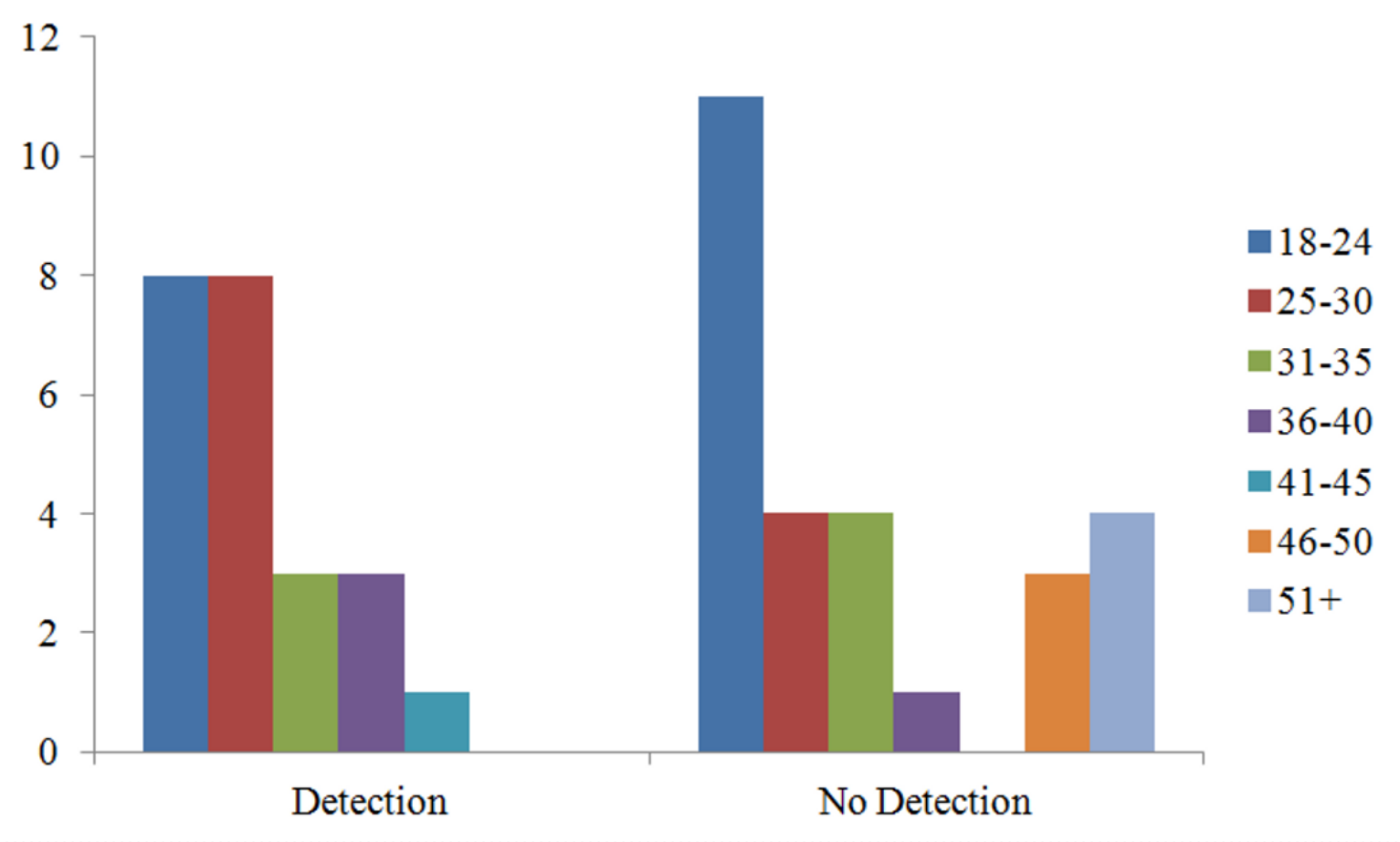
		'At-risk' group	'Low risk' group
Independent variables	Mean	Mean	Mean
Number of hosts contacted	6722	8617	5108
Total time online (hrs)	50.83	67.41	37.70
Total uptime (hrs)	358	388	332
Browser history entries	3955	5755	2421
Number of untested or dangerous sites visited	746	1101	443
Number of adult sites visited	72	129	23
Number of software/file download sites visited	47	84	16
Number of streaming media sites visited	159	272	63
Number of games sites visited	45	39	34
Number of files downloaded	489	545	442

!!!!

Demographic info: gender distribution



Demographic info: age group distribution



Description of the study: cost

- Expenses
  - Laptops
    - 50 units, bought at $375 and sold at $350 each = $1,250
  - Participant compensation
    - 50 participants at (3x $50 + 1x $100 + 1x $150 = $400) = $20,000
- Labour
  - Experiment design and tool development
    - 1x master's student full-time, 4 months
    - 1x undergraduate student, 4 months
  - Experiment conduction and analysis
    - 1x master's student full-time, 6 months
    - 1x undergraduate student, 1 month
- Overall cost
  - $21,250 + 15 person.month

The next step: a large-scale study

- Additional objectives
  1. Statistical significance
     - Population size
     - Population diversity
     - Malware sample diversity
  2. Comparative testing
     - Different products in similar environments
     - Same product in different environments
- Characteristics
  - Duration
    - 4 months
  - Population
    - Minimum 200 participants
  - Configuration
    - Machine of participant, with minimum requirements
    - Windows 7 or Windows 8 on their OWN computer
    - One of 4 AV products

The next step: two options

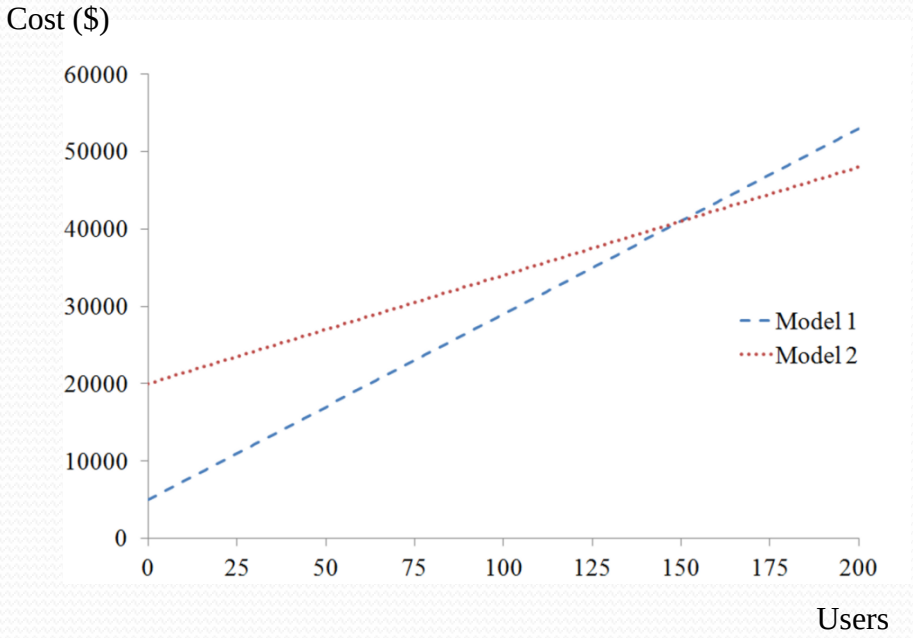
- Model 1 – Local testing
  - Similar to proof-of-concept experiment
  - Participants recruited locally
  - 5 in-person sessions for setup & data collection
- Model 2 – Remote access
  - Participants install tools (self-install package)
    - (Re)install AV
    - Benchmark computer
  - Usage data sent monthly
  - Detection protocol executed automatically
    - Activates “suspicious” mode
    - Sends additional data

The next step: two options

- Model 1 – Local testing
  - Advantages
    - No additional tool development required
    - Direct access to users and computers (better data)
  - Disadvantages
    - Labour-intensive
    - Geographical bias in population
- Model 2 – Remote access
  - Advantages
    - Can recruit participants worldwide
    - No need to buy hardware
    - Cheaper per-user cost
  - Disadvantages
    - Detection protocol weaker (e.g. Rootkits)
    - Less opportunity for in-depth investigation

The next step: Comparative cost

- Model 1
  - Initial cost: $5,000
  - Operating expenses
    - 7 person.hours per user = $140 per user
  - Compensation
    - 5 visits @ $20 = $100 per user
  - Example: 200 users = $53,000
- Model 2
  - Initial cost: $20,000
  - Operating expenses
    - 4 person.hours per user = $80 per user
  - Compensation
    - $50 gift certificate per user
  - Example: 200 users = $46,000



Conclusions – Summary of results

1. Detection rates
   - Comparable to those observed in other tests ??
   - Heavily dependent on sample classification…
2. Behaviour does influence risk of infection
   - More browsing, more risk
   - Standard deviations do matter
     (due to high variance in user behaviour)
   - There is something about adult sites….

Conclusions – Our approach

1. Viability of the field study/clinical trial approach
   - Advantages
     - Can produce results of unprecedented "realism"
     - Allows access to otherwise inaccessible user data
     - Obviates sample selection problem…
   - Disadvantages
     - Requires large population for statistical significance
2. Future studies
   - Could be conducted locally or remotely
   - At affordable cost (130-240$ per user)

Who wants to be next ???

Acknowledgements