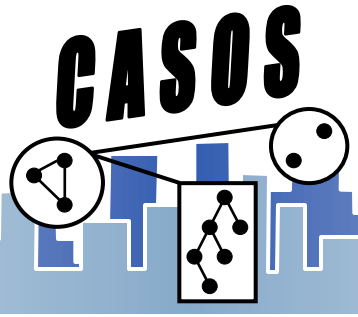


An Empirical Study of Global Malware Encounters

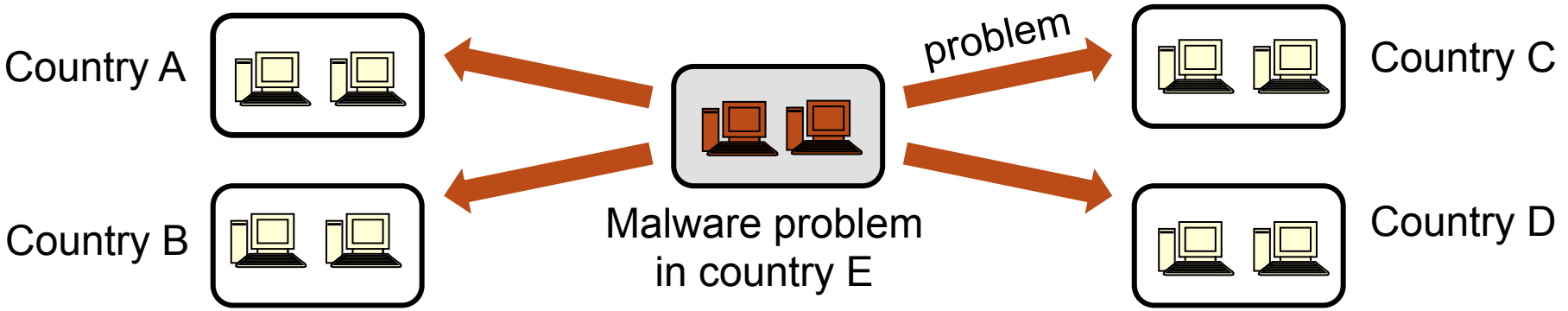
Ghita Mezzour
mezzour@cmu.edu

Kathleen Carley
kathleen.carley@cs.cmu.edu

L. Richard Carley
carley@ece.cmu.edu

A Global Problem Requires a Global Perspective

- Malware is a global problem



- Identifying factors that cause countries to have serious malware problems
  - Opportunity to alter these factors
  - And Reduce global malware encounters

Prior Work Has Different Focus

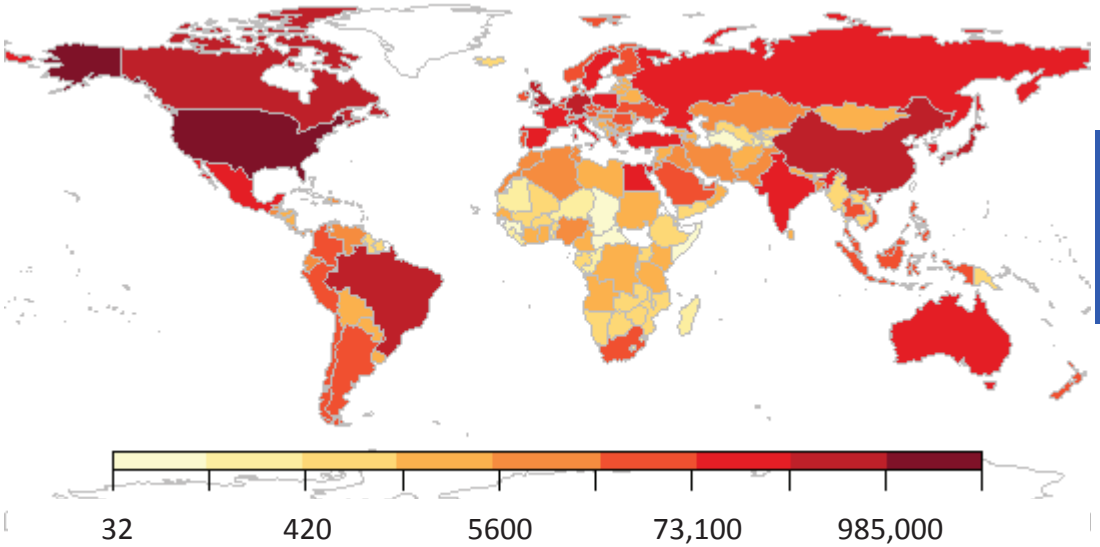
- Attack infrastructure characterization [Provos et al. '08, Caballero et al. '11]
  - Example: find choke points
- User level [Levesque et al '13, Canali et al '14]
  - Relationship between users' demographics and malware exposure
- Plausible explanations for international differences
  - No empirical testing of explanations' accuracy

Overview

- Malware problem studied: malware exposure
  - Malware contagion
  - Compromised computers as bots
- Research question
  - What factors cause computers in some countries to be highly exposed to malware?
- Data
  - Symantec Anti-Virus (AV) telemetry data
  - Country-level technical & social factors
- Approach
  - Statistical hypothesis testing

Symantec Anti-Virus Telemetry Data

- Data from 10 million+ customers worldwide

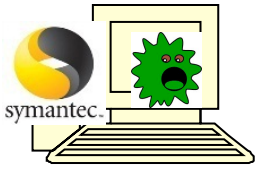


 # Symantec computers in the data

- Time coverage: 2009-2011

Symantec Anti-Virus Telemetry Data

- Threat report generation



Victim

Threat report	
Malware name	W32.Aimdes.A@mm
IP address	128.2.184.224
Country	United States
Machine ID	104951814

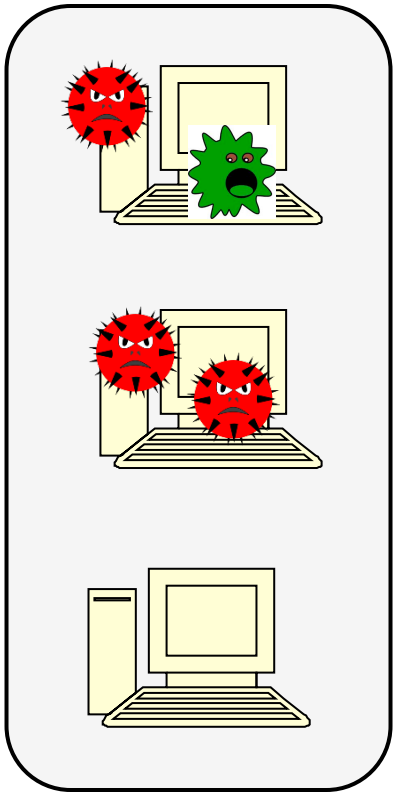
Telemetry data

- Threat catalog [Mezzour et al. '14]: online descriptions

Threat catalog entry	
Malware name	W32.Aimdes.A@mm
Malware family	Aimdes
Malware type	worm

Counting Malware Encounters

- Malware encounter instance:
  - (Victim machine, malware family)



2 encounters

1 encounter

0 encounter

Country

Average # of malware encounters per computer: 1

Hypotheses about Factors Affecting Malware Exposure

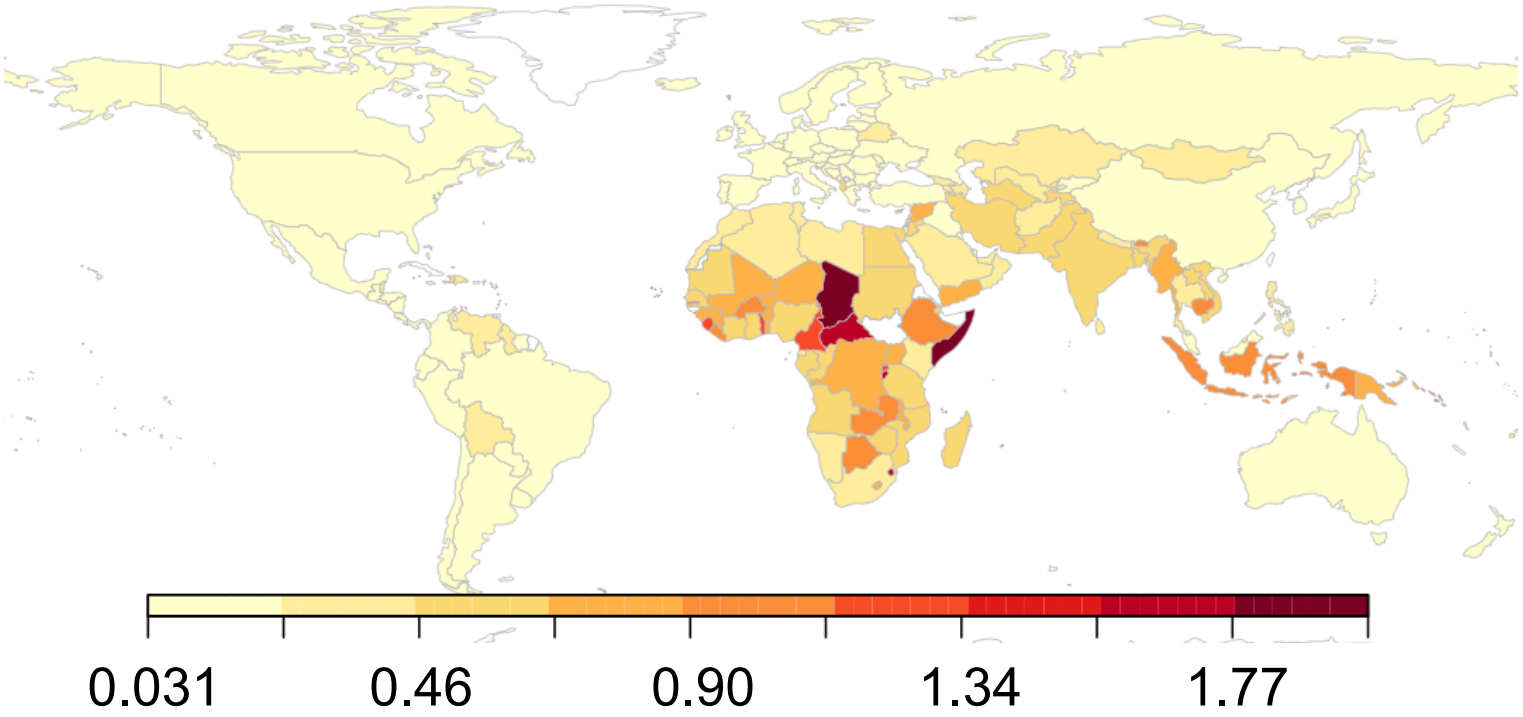
Factor	Expected effect
Resources [Caballero et al. 11]	 + + +
Web visits [Canali et al. 14]	 + + +
Security expertise [Onarlioglu et al. 12]	 - - -
Software piracy [Kammerstetter et al. 12]	 + + +
International hostilities [NY times 12]	 + + +
International alliances [Madnick et al. 09]	 - - -
International extraditions [Madnick et al. 09]	 - - -

Data about Factors Affecting Malware Exposure

Factor	Data
Resources	GDP per capita [World bank] Bandwidth [ITU]
Web visits	# web pages [Canali et al. 14] # top level domains [Canali et al. 14]
Security expertise	# research papers [SCOPUS] CERT [Lewis 11, CERT]
Software piracy	Piracy index [Business software alliance]
International hostilities	Hostility betweenness [International Crisis behavior]
International alliances	Alliance betweenness [Correlates of War]
International extraditions	Extradition betweenness [UN]

Sub-Saharan Africa most Exposed to Malware

- Avg # of malware encounters per computer



Avg # viruses encountered per computer

Piracy behind High Malware Exposure in Africa

Regression analysis. Regression coefficients are standarized

	Viruses
Bandwidth	0.013
GDP PC (log)	0.302
# web visits	0.021
#top level domains visits	0.024
Piracy	0.81***
Piracy x GDP per capita (log)	-0.68***
Cyber security research	-0.085
Cyber security institutions	0.056
Military alliances	0.009
Military hostilities	0.054
Extradition treaties	0.015
R square	0.64

* p < 0.05, ** p < 0.01, *** p < 0.001

Computer Piracy in Africa

- Pirated software available from street merchants
- Merchants download pirated software from P2P networks and dubious sites
- Pirated software uploaded by hackers interested in distributing malware

Source: Biztech Africa

How to Fight Software Piracy in Sub-Saharan Africa?

- Economic factors are driving piracy
  - Poor people can't afford legitimate software
  - GDP per capita in Central African Republic: 700 USD
  - Microsoft windows: 119 USD (119/700 = 0.17)
- Governments reluctant to fighting piracy
  - No desire to cut populations from the Internet
- Software industry has incentive to act
  - Money losses
  - Price adjustment to income: win-win solution

Limitations

- Symantec data
  - Single vendor perspective
  - Highly sophisticated attacks non-covered
  - 5 years old
- First-pass analysis

Future Work

- Data from other vendors
- Other analyses
  - Longitudinal analysis
  - Ratio of computers that encounter any number of malware out of total number of computers
  - Use Bayesian Information Criteria (BIC) technique to find "best" fitting regression
- Higher granularity analysis
  - ISP level
- Field studies in Africa

Conclusion

- Malware is global: needs global perspective
- Empirically identify factors behind international variation in malware encounters
  - Symantec anti-virus telemetry data
  - Country-level social & technical measures
- Sub-Saharan Africa most exposed to malware because of wide-spread software piracy
- Policy suggestion
  - Software price adjustment