THE UNIVERSITY of NEW MEXICO

The Complex Science of Cyberdefense

Stephanie Forrest

University of New Mexico
and
Santa Fe Institute

May, 2015

THE LANDSCAPE

Obama to Call for Laws Covering Data Hacking and Student Privacy

By MICHAEL D. SHEAR and NATASHA SINGER JAN. 11, 2015

The New York Times

Cybersecurity Challenges

- Many integrated layers of software
  - Controlled by multiple parties
  - Lack of transparency
  - Interactions lead to bugs and vulnerabilities
- Outsourced IT operations and new business models
- Distributed supply chains
- Mobility
- Large heterogeneous networks
- Spinning out of control?

The Complex Systems Perspective

- Arms race with adversaries
  – Rapid innovation cycle
  – Moore's Law helps adversaries and defenders
- Inadvertent evolution
  – Through actions of many individual programmers
- Interactions lead to unanticipated behaviors
- Network effects
- Mixed incentives
  – Financial
  – Political

The Complex Systems Perspective
Basic Concepts

- Focus on dynamics
- Network science
- Scaling behavior: As systems grow,
  - What remains the same?
  - What changes?
- Game theory
- Adaptation, competition, and evolution

Overview

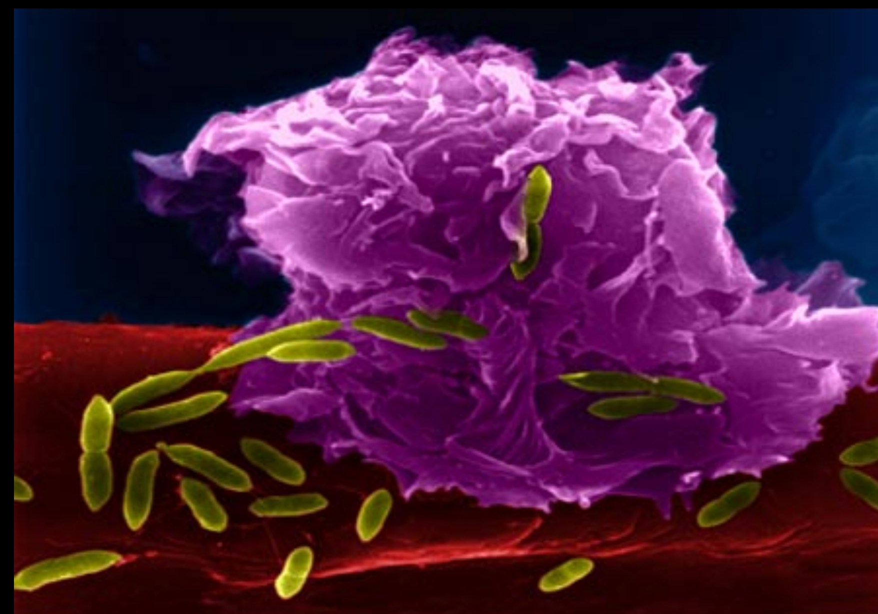
- Biology and cyber security
  - Computer immunology
  - Automated repair of vulnerabilities
  - Engineered diversity
- Cybersecurity modeling
  - Data breaches
  - Spam, botnets, and policy
- Computer science meets policy

Malicious behavior emerges spontaneously in many complex networks

Self-interested Actors
Adaptation

Biology is the Science of Security

- Biological systems cope with adversaries and have highly evolved defense systems
  - Pervasive
  - Multi-level
  - Complex



An activated macrophage phagocytosing bacteria upon contact
Photo: courtesy of Dennis Kunkel

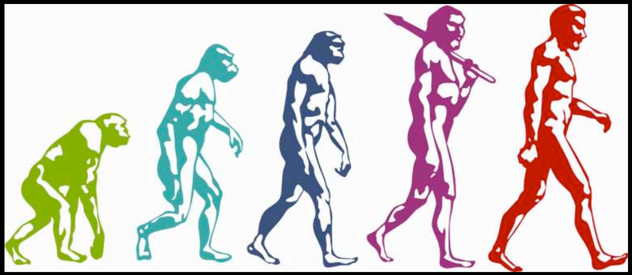
- Suggests novel approaches to cybersecurity and resiliency

Mimicry Attacks

Traditional Approaches to Cybersecurity

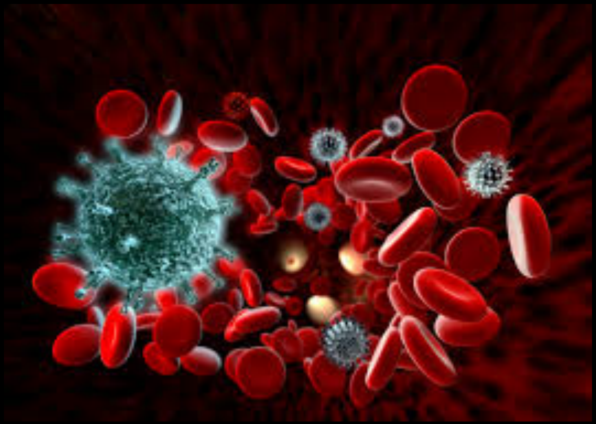
- The myth of perfection
  - If programmers weren't idiots …
- Devise a new method of coping with each new method of attack.
  - Find vulnerability, Fix vulnerability, Repeat
  - Does not scale
- Self-healing systems
  - Pre-enumerate vulnerability types and repair approaches
  - Apply repair when fault is encountered and continue executing
  - Specific to known vulnerability classes
- Risk management
  - Estimate probability and cost of successful attacks
  - Reasonable quantitative estimates rarely available

Biological Defenses



Evolution, Adaptation, Healing

Diversity



Defense in Depth

Immunology

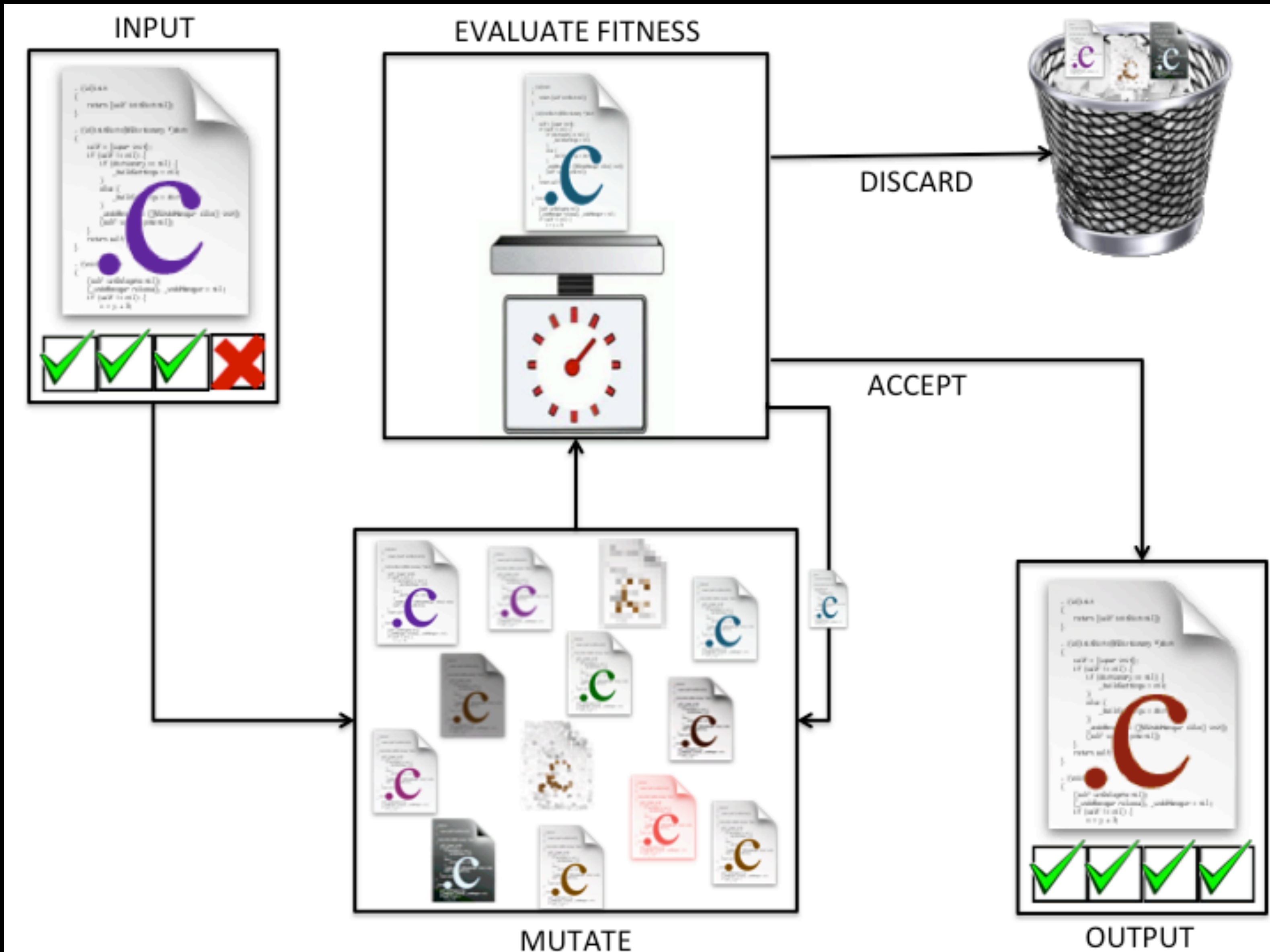



Evolution for Program Repair with Westley Weimer (UVA)

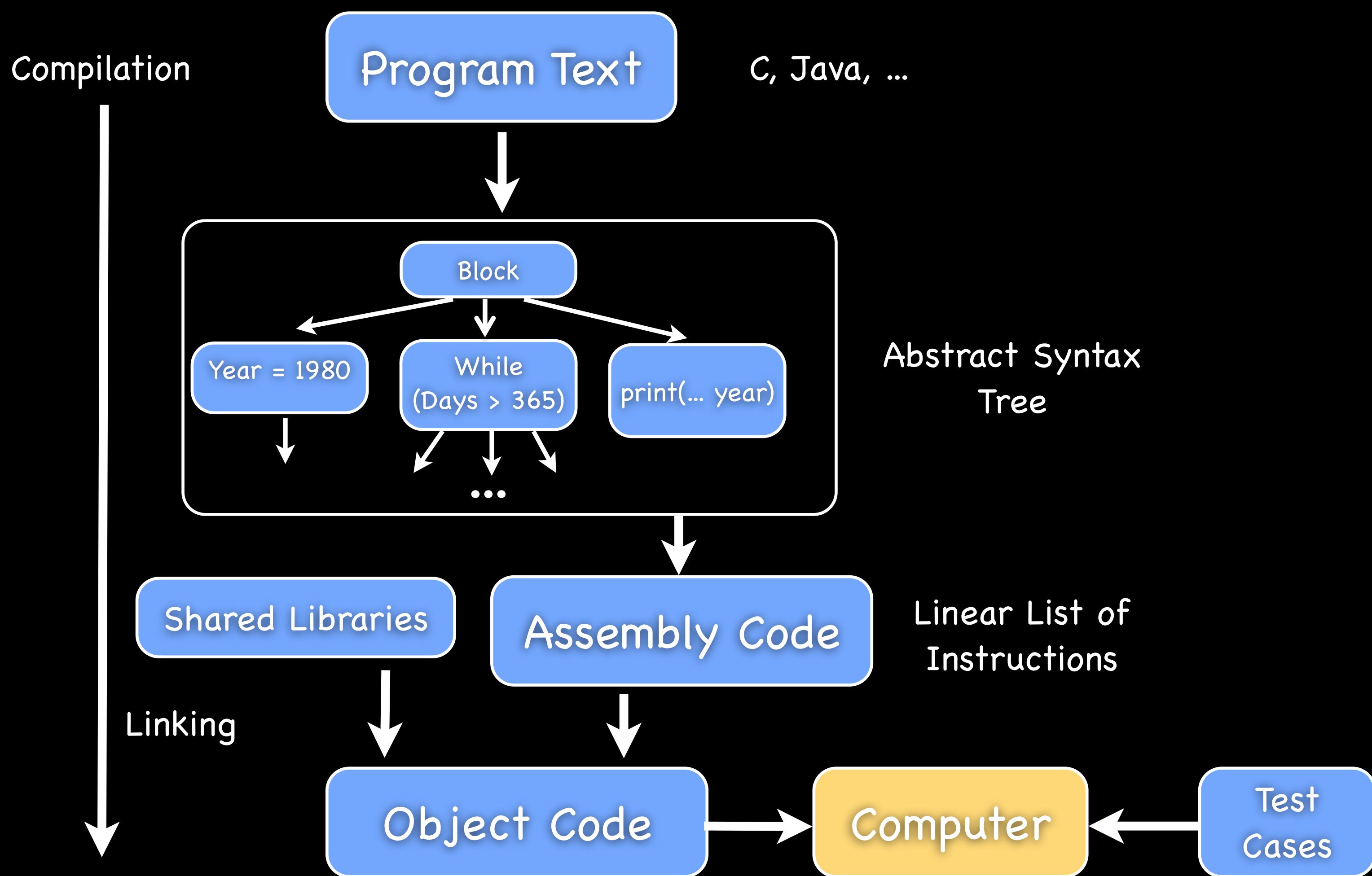
Goal: A generic method for automated software repair

Legacy code
Do not assume a formal specification

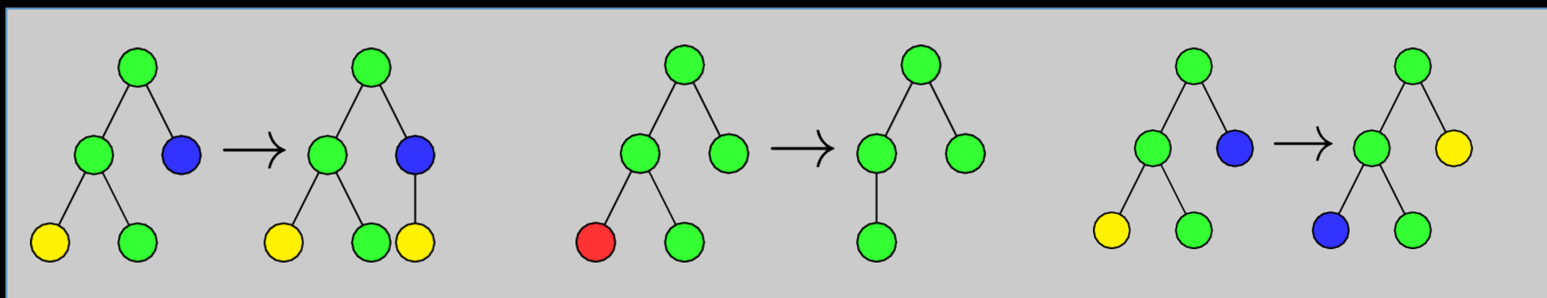
GenProg



Software Translation/Transcription



Mutation/Crossover Operators



Copy Delete Swap

- Don't invent new code
- Statement-level operations

Example Repair: Microsoft Zune Player

```
void zunebug_repair(int days) {
  int year = 1980;
  while (days > 365) {
    if (isLeapYear(year)){
      if (days > 366) {
        // days -= 366; // repair deletes
        year += 1;
      }
      else {
      }
      days -= 366;       // repair inserts
    } else {
      days -= 365;
      year += 1;
    }
  }
  printf("current year is %d\n", year);
}
```

Repair produced in 42 seconds

- Dec. 31, 2008. Microsoft Zune players freeze up
- Bug: Infinite loop when input is last day of a leap year
- Repair is not trivial
- Negative test case: 10593 (Dec 31, 2008)

Example Repairs: Security Vulnerabilities

(ICSE'09, TSE'12)

Program	LOC	Path Length	Program Description	Vulnerability	Time to Repair
nullhttp	5575	768	Webserver	Remote heap overflow	578s
openldap	6519	25	Directory protocol	Non-overflow denial-of-service	665s
lighttp	13984	136	Webserver	Remote heap overflow	49s
atris	21553	34	Graphical game	Buffer overflow	80s
php	26044	52	Scripting Language	Integer overflow	6s
wu-ftp	35109	149	FTP server	Format string	2256s
ccrypt	7515	18	Encryption ytility	Seg. fault	47s

How well does GenProg work in practice?
(ICSE'12, TSE in press)

Program	Description	LOC	Tests	Bugs	
				Fixed	Total
fbc	Language (legacy)	97K	773	1	3
gmp	Multiple precision math	145K	146	1	2
gzip	Data compression	491K	12	1	5
libtiff	Image manipulation	77K	78	17	24
lighttpd	Web server	62K	295	5	9
php	Language (web)	1,046K	8,471	28	44
python	Language (general)	407K	355	1	11
wireshark	Network packet analyzer	2,814K	63	1	7
Total		**5.14M**	**10,193**	**55**	**105**

Repaired 52% at a cost of $7.32 each
With algorithm tuneups: 5 additional bugs (57%)
With additional CPU resources (69%)

Mutational Robustness

- Many biological mutations leave fitness unchanged
  - Mutational robustness
- Believed to play an important role in evolution
  - Buffering
  - Genetic potential

```
if (right > left) {
   // code elided
   quick(left, r)
   quick(l, right)
}
```

⇒

```
quick(l, right)
quick(left, r)
```

- Software mutational robustness
  - ~30% of GenProg mutations don't change behavior of program
  - Related to mutation testing

Significance of Software Neutrality

- Contradicts idea that "programs are fragile"
- Possible explanation for GP repair results
- Supports "strong biology hypothesis" of computing
  - More than just "bio-inspired"
  - Software has acquired biological properties through inadvertent evolution

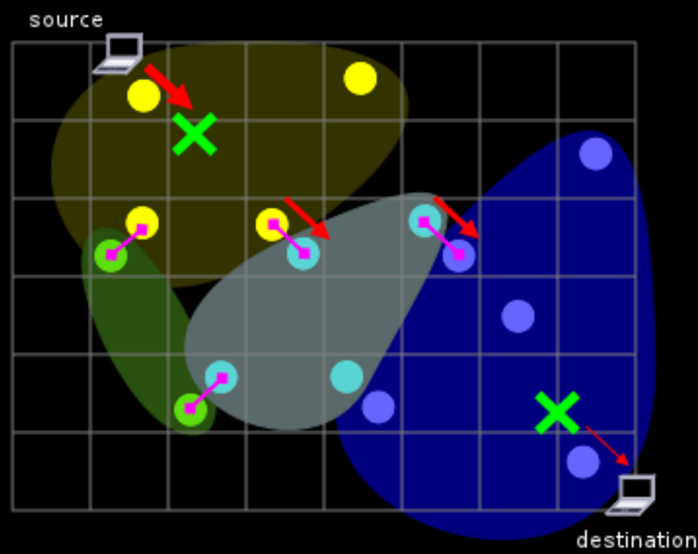
Evolution produces diversity

- Coarse-grained diversity
  - Generate populations of semantically distinct programs
  - Automatically repair latent bugs and avoid single points of vulnerability
  - DARPA CFAR Program

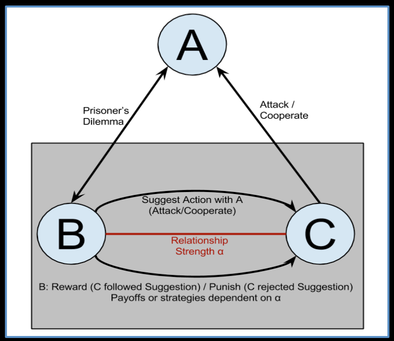
Beyond Biology

- Biology provides useful design principles
- Incorporating economic, political, and psychological realities requires policy
- Need modeling to assess this new level of complexity:
  - What interventions will be most successful?
  - Where should they be deployed?
  - What are the unintended consequences of increasing cybersecurity?

What kinds of models are appropriate for large-scale security questions?



- Data-driven
  - Statistical models
- Concept-driven
  - Mathematical equations
  - Game theory
  - Computational and simulation models
    - Agent-based modeling



Data Breaches

- Perception that we are losing ground
  - Are we?
  - How would we know?
- Modeling questions
  - How many breaches?
  - How large?
  - Are they changing over time?
- What should we do about it?
- What might attackers do next?

JAMA The Journal of the American Medical Association

Home Current Issue All Issues Online First Collections CME Multimedia

April 14, 2015, Vol 313, No. 14 >

< Previous Article | Full content is available to subscribers Subscribe/Learn More | Next Article >

Research Letter | April 14, 2015

Data Breaches of Protected Health Information in the United States

Vincent Liu, MD, MS[1]; Mark A. Musen, MD, PhD[2]; Timothy Chou, PhD[3]

[+] Author Affiliations

JAMA. 2015;313(14):1471-1473. doi:10.1001/jama.2015.2252.

Text Size: A A A

Article | Figures | Tables | References

Reports of data breaches have increased during the past decade.[1,2] Compared with other industries, these breaches are estimated to be the most costly in health care; however, few studies have detailed their characteristics and scope.[1]

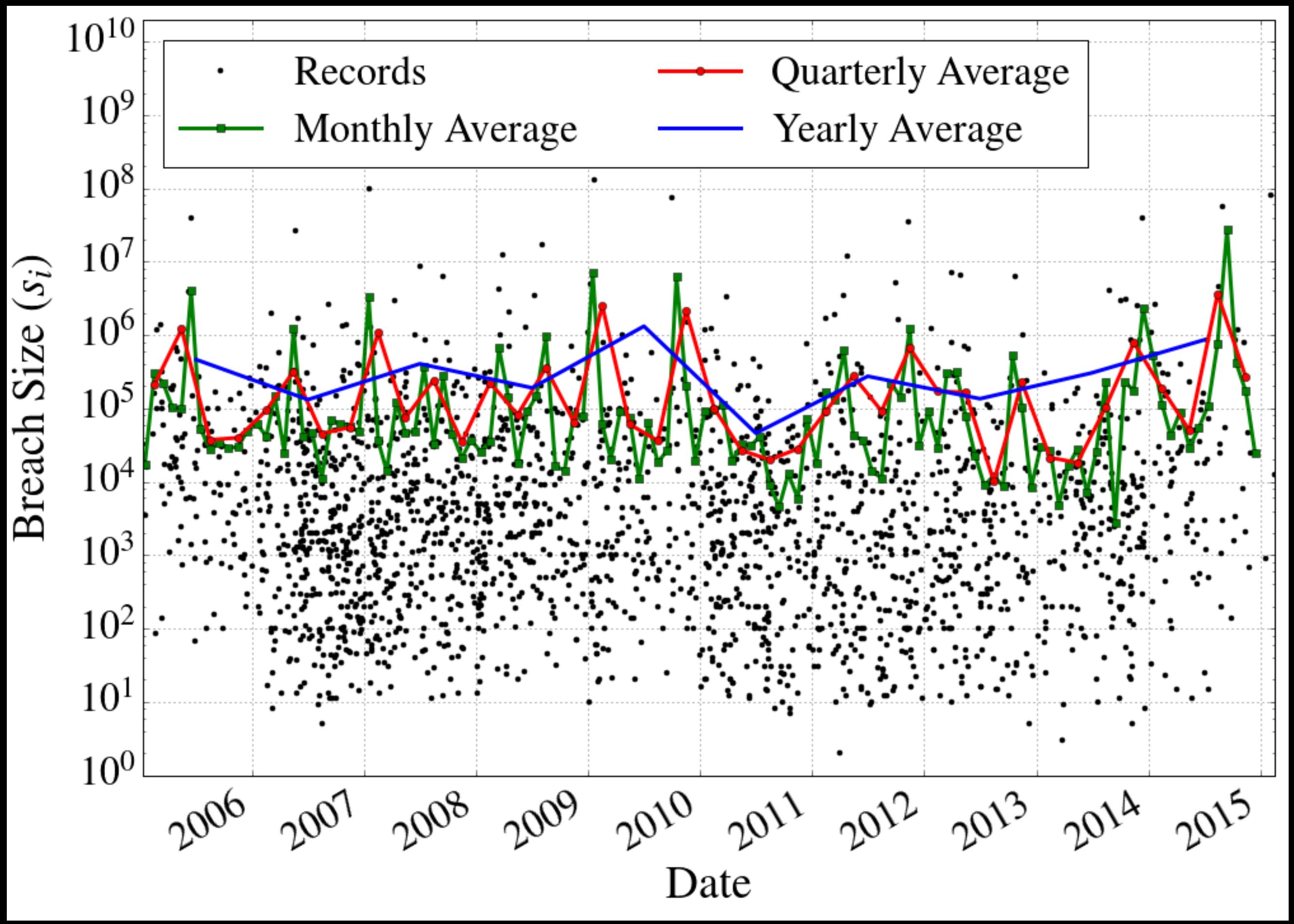
Obama to Call for Laws Covering Data Hacking and Student Privacy

By MICHAEL D. SHEAR and NATASHA SINGER JAN. 11, 2015

The New York Times

Data Breaches: Hype and Heavy Tails

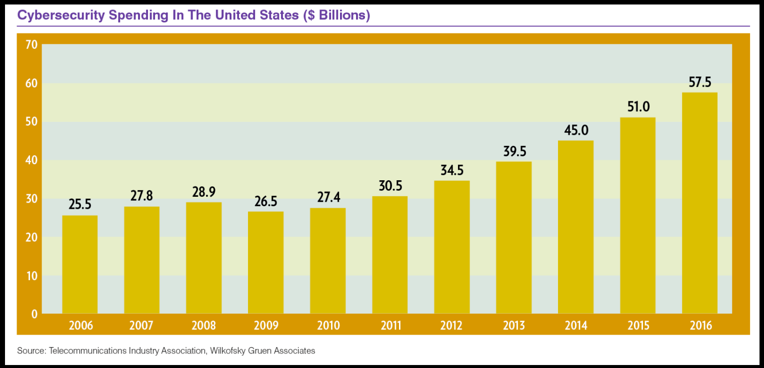
(Edwards, Hofmeyr, Forrest, WEIS in press)



Data Source: The Privacy Rights Clearinghouse
http://www.privacyrights. org/data- breach

Significance

- No evidence that data breaches are getting worse
  - Question reports that cite averages and annual trends
- Possible explanations
  - Relax, life is not so bad
  - Data set is not representative, or analysis problems
  - The red queen

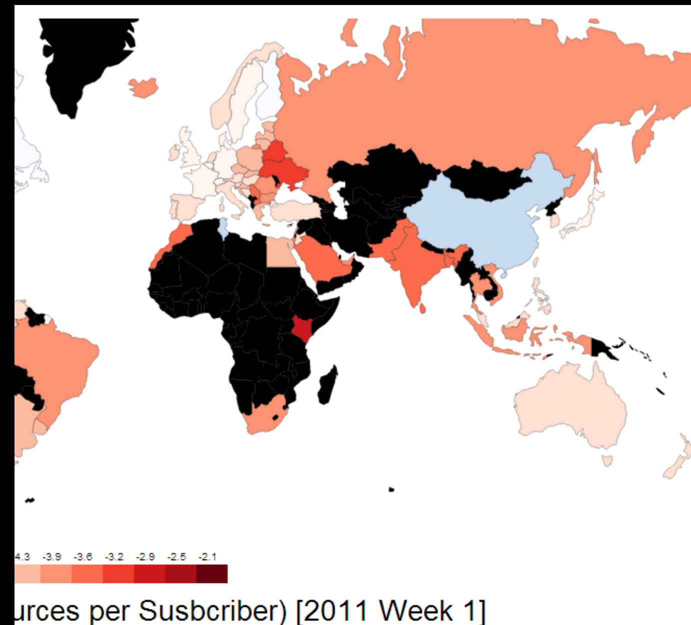


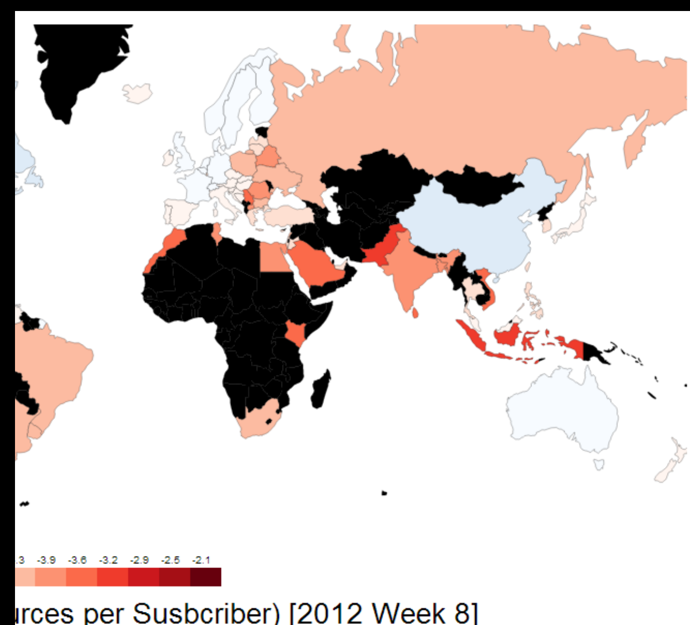
"DoD spent $31 Billion on IT in 2014"

We are running faster and faster to stay in same place

Spam Migration

- Most spam is sent by botnets
- Spam is dynamic and noisy
  - Campaigns cause traffic spikes
  - Takedowns force migration
- Worldwide problem
  - $20-50 billion in U.S. (2011)
  - Crushing burden for immature IT infrastructures
- Excuse for intl. regulation
- Botnet takedowns are a popular intervention

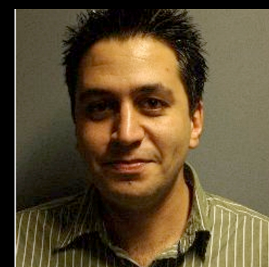




Spam Migration and Mitigation

UNM and Delft University

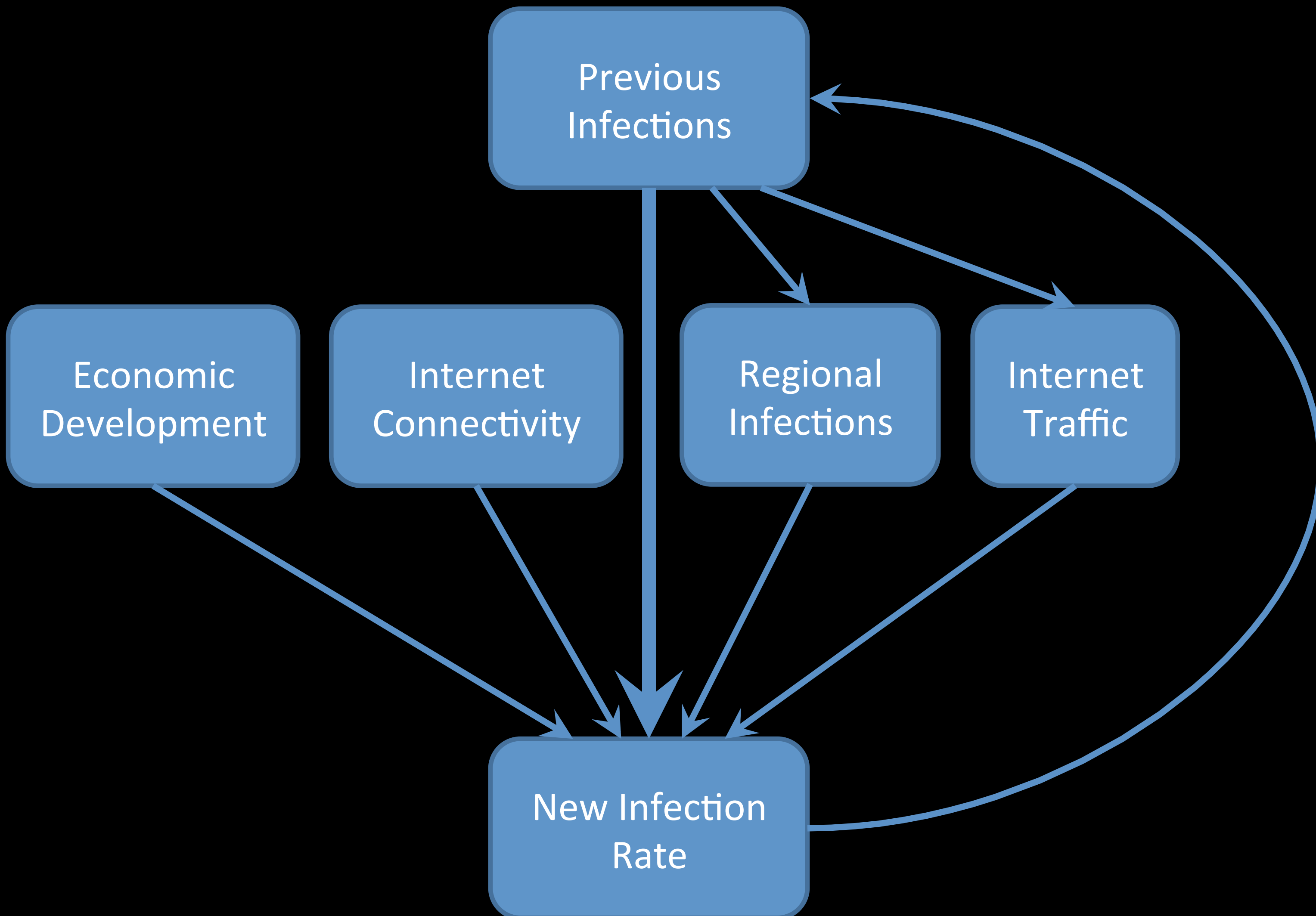



- Data set collected weekly from a *spam trap*
  - Jan. 2005 – Dec. 2014
  - Records which IP addresses are sending out spam
  - 127 billion spam messages from 440 million unique IP addresses worldwide
- Inference procedures to determine country of origin, originating network/operator
  - Geo-locate each message

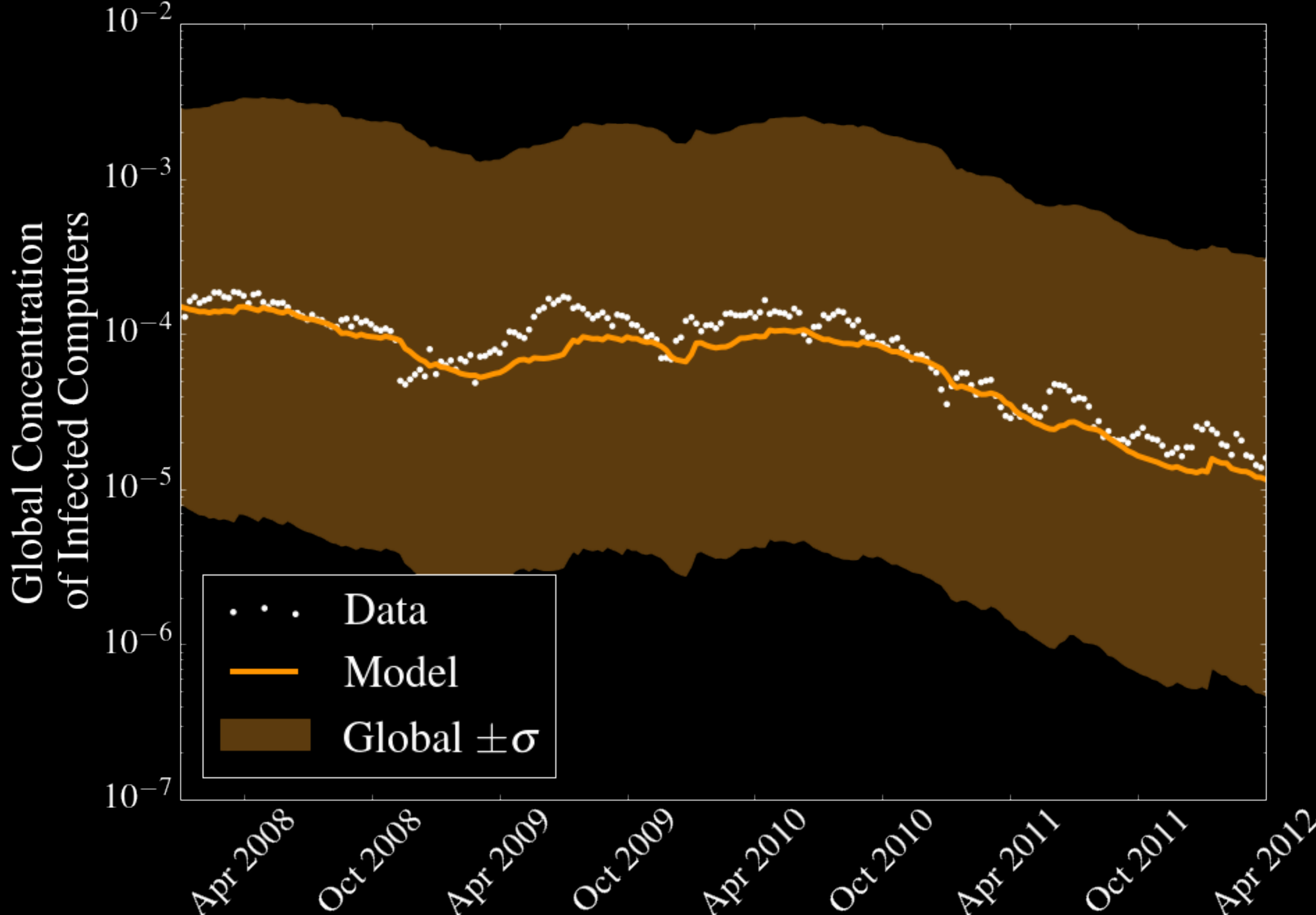
Van Eeten, Michel J.G., J. Bauer, H. Asghari & S. Tabatabaie, (2010),The *Role of Internet Service Providers in Botnet Mitigation: An Empirical Analysis Based on Spam Data,* OECD STI Working Paper 2010/5, Paris: OECD.

Modeling an ISP'S Infection Rate

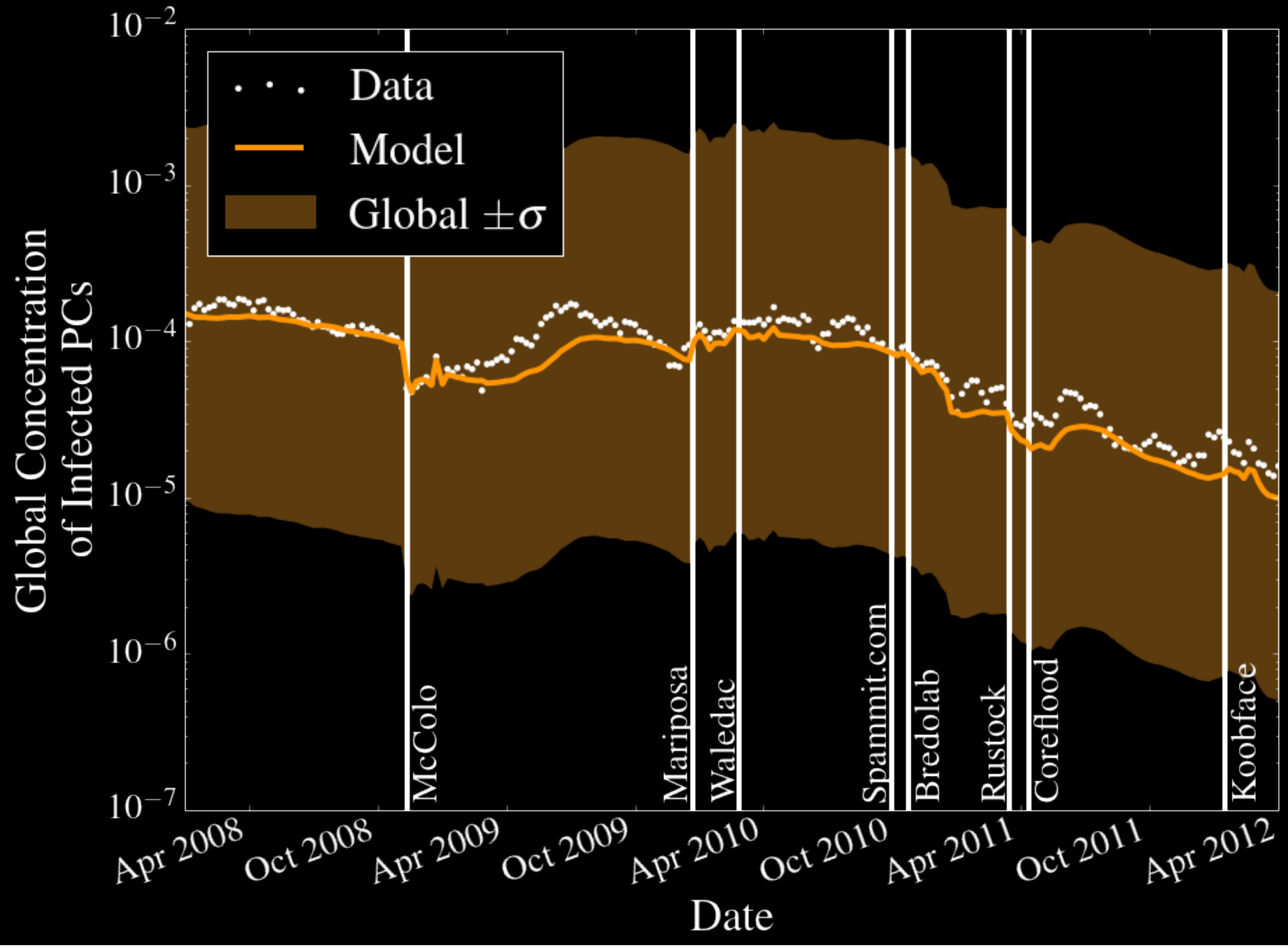


$$\ln(W_i(t)) = \beta_0 \ln(W_i(t-1)) + \beta_1 \ln(R_i 4(t-1)) + \beta_2 \ln(G_i(t-1)) + \beta_3 I_i(t) + \beta_4 P_i(t) + \beta_5 \ln D_i(t) + \varepsilon$$

Model Accuracy

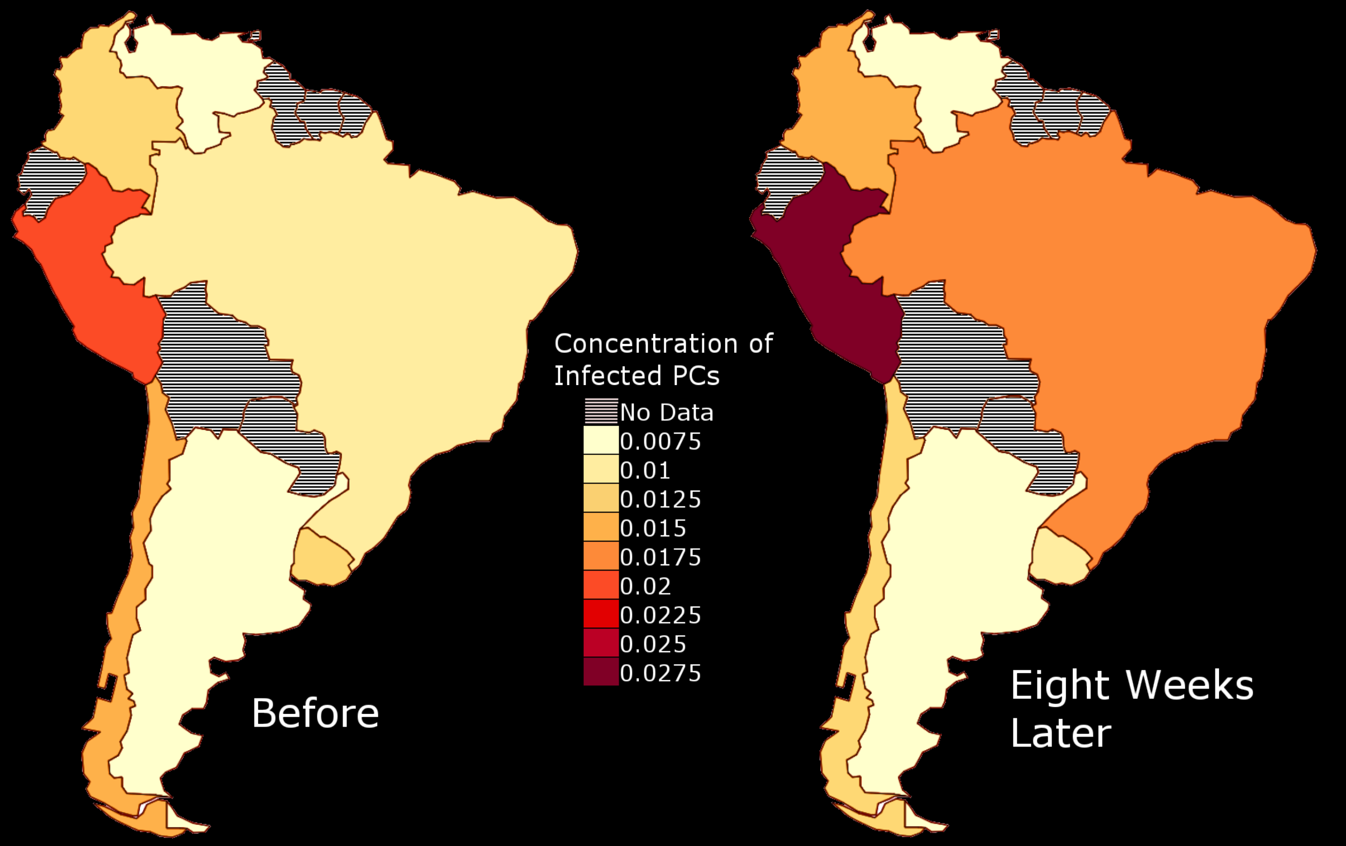


Takedown Effects



Coreflood Botnet Takedown

Hypothesis: takedowns force botnets/spambots to migrate to new niches that are less well protected.



Where should we Intervene?

- Measure the effect of botnet takedowns on various countries
- Conduct a *historical experiment*
  - Find a country where takedown was effective (Chile)
  - Pretend that another country is protected as well as Chile
  - Then, measure the predicted effect of the takedown in simulation

Where do we Intervene?

Two Scenarios

Peru	Brazil
49% Reduction in Infection Rates in Peru	75% Reduction in Infection Rates in Brazil
4.5% Reduction in South America	43% Reduction in South America
0.0% Globally	4.4% Globally

Summing Up

- The security landscape is complex
  - Arms races and the red queen
  - 'The evolutionary mess'
  - Network effects
  - Mixed incentives
- Biological design principles
  - Automated repair of software vulnerabilities
  - Mutational robustness and diversity
- Large-scale security issues
  - Are we losing ground or not?
  - Requires careful data analysis and modeling
  - Worldwide geo-political consequences

Conclusion

- How do we tackle the ever-increasing scale of cybersecurity problems?
  - Complex systems perspective
- How can we predict the likely consequences of an intervention?
  - Modeling and data analysis
- How do we incorporate incentives, social interactions, and politics into cyberdefense?
  - Policy

The Complex Science of Cyber Security

QUESTIONS?

WWW.CS.UNM.EDU/~FORREST

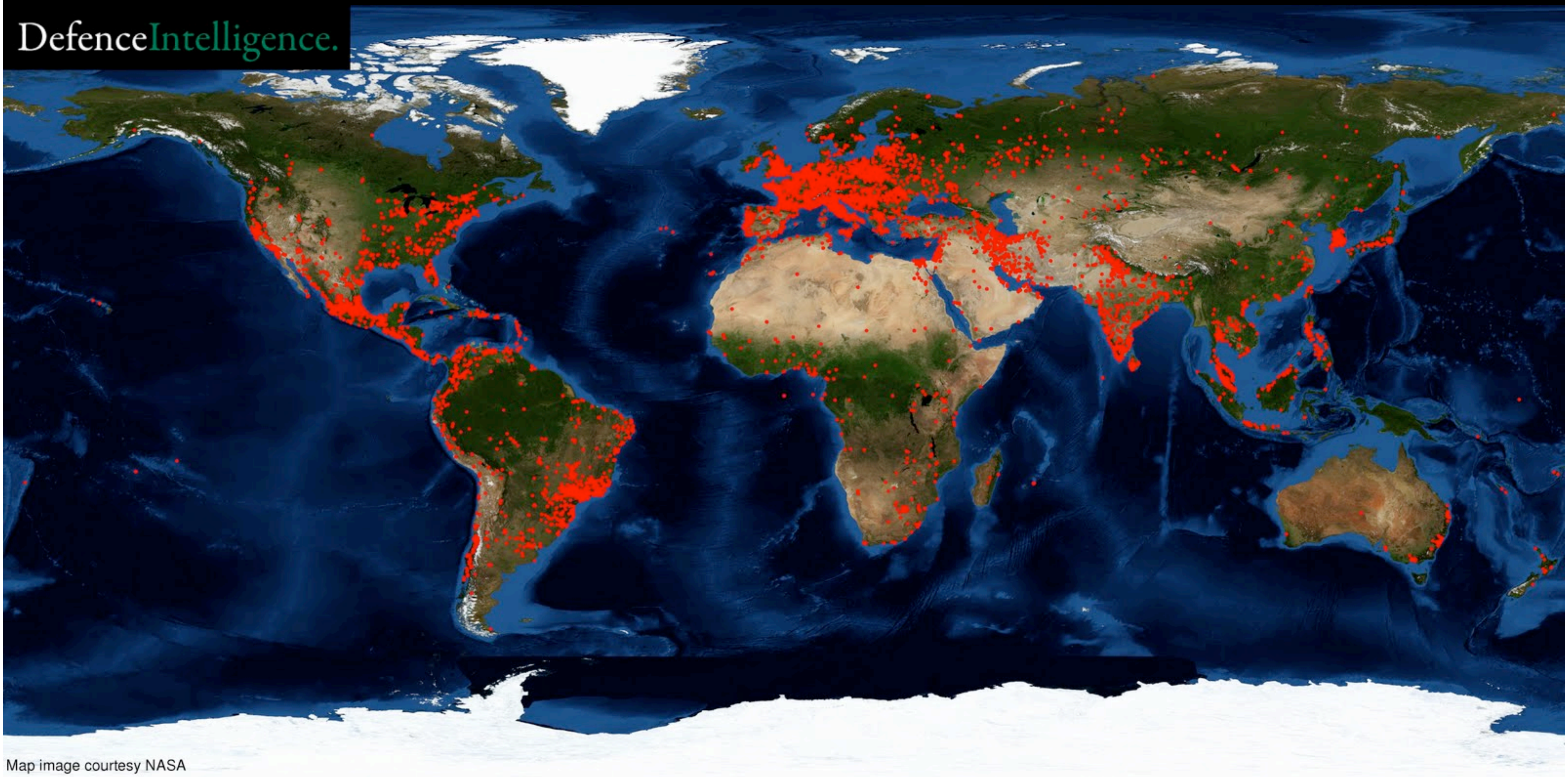
[MOVE] Responses to Malicious Behavior

- Observe
  - Monitoring and surveillance
- Hide
  - Obfuscation, camoflauge, mimicry
- Filter/quarantine/block
  - Blacklisting, censorship, throttling, excision
- Repair/replace
  - Patch, gene editing, transplants
- Counterattack
  - Takedowns, chemo- and radiation-therapy, killer T-cells

Other Biology to Explore

- CRISPR
- Defense in depth
- The innate immune system
- Cryptic sequences

CyberSecurity is a Global Issue



Mariposa Botnet Infections (2010)

My year at State

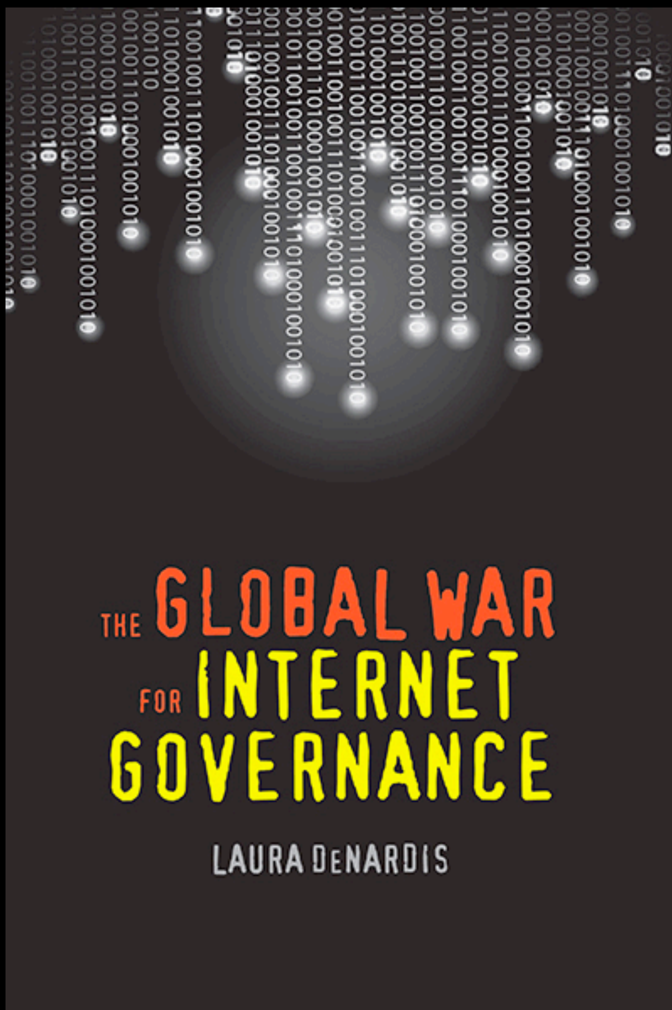
- Communication and Information Policy office
  - Multi-lateral fora: ITU, ICANN, OECD, etc.
  - Bilateral meetings with East Asian countries
  - Privacy and big data
  - Internet governance. IANA transition
  - CFIUS processes
- Coordinator for cyber Issues (S/CCI)
  - Confidence building measures for cyberspace

Global Cyber Issues

- Privacy and surveillance
  - Cloud computing and data localization
  - The right to be forgotten
- Who should control the Internet?
  - Net neutrality
  - Governance models
  - DNS Takedowns
- Cyberwarfare and economic espionage
  - Zero-day exploits
  - Norms in cyberspace, attribution
- Spam

Internet "Governance"

Who owns the Internet?



- Mechanisms of control
  - Technical design decisions
  - Private corporate policies
  - Global institutions
  - National laws and policies
  - International treaties
- What are the control points?

Complications

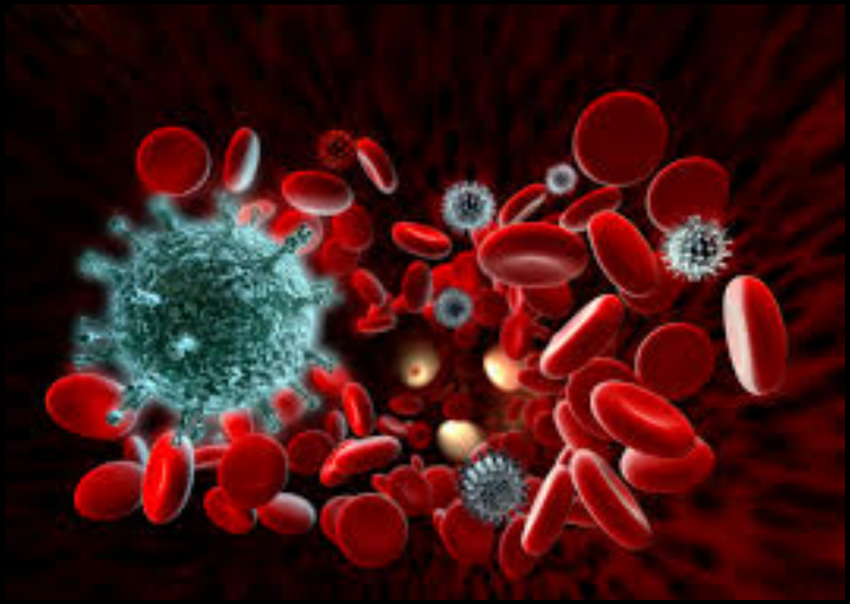
- Money
  - Traditional revenue streams have been disrupted by Internet applications
- Borderless design vs. geography
  - Politics and territoriality
  - Legal frameworks
- Dynamic innovation cycle
  - Engineering complexity
- Distributed data sets and mobile computing
  - Computing as a service (the "Cloud")
  - Mirrored data sets and backups

The MultiStakeholder Process

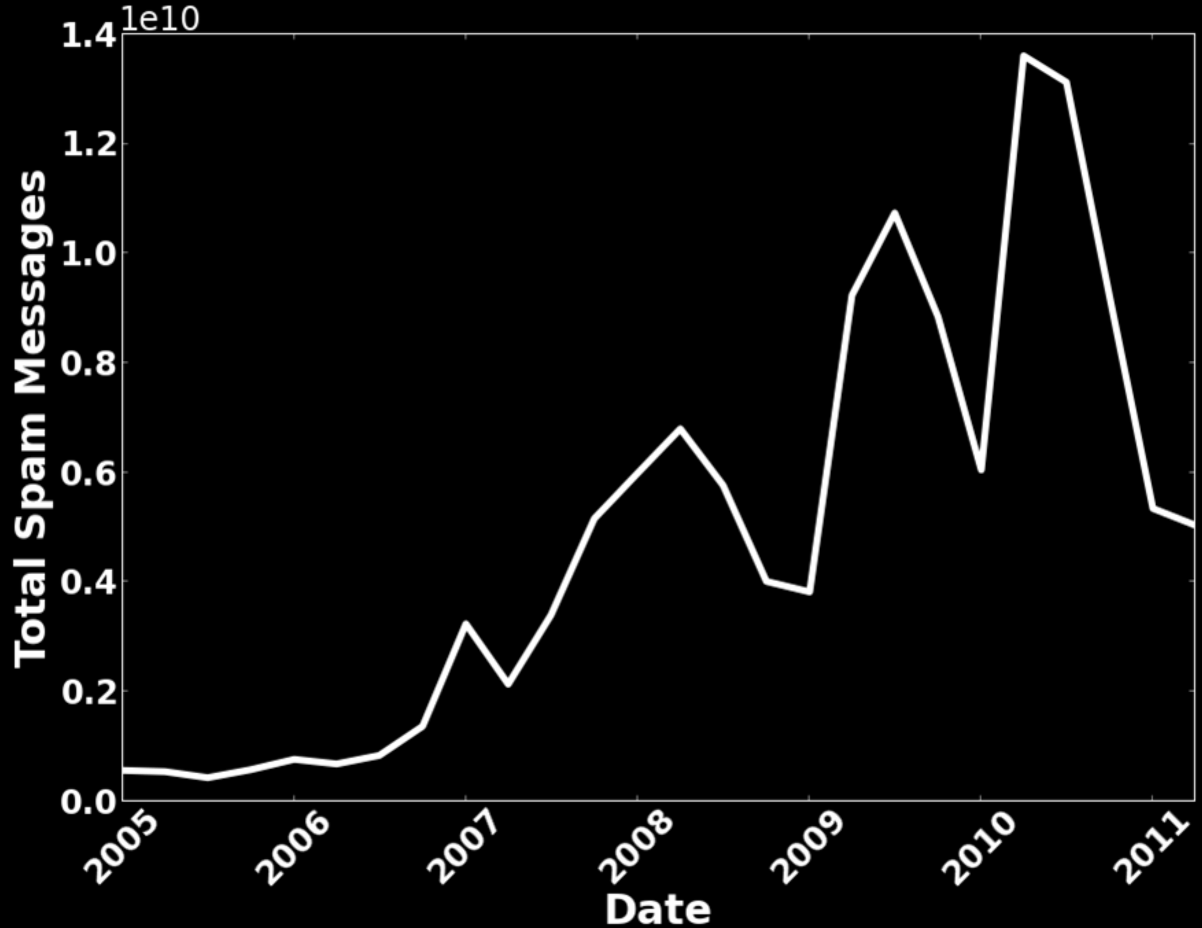
"Aren't you glad we had this meeting to resolve our conflict?"

A Computer Immune System

- Immunology
  - Detects novel pathogens
  - Choose and mount an effective response
  - Automatically and in time
- Cybersecurity goals
  - Detect unauthorized use of computers, malware, etc.
  - Respond automatically to remove the threat



Spam 2011: Top 300 Websites Worldwide



$20-50 Billion in 2011 in the United States

So: Rao, Justin M., and David H. Reiley. 2012. "The Economics of Spam." *Journal of Economic Perspectives*, 26(3): 87-110.

Threats to the Network

- Actions by nation-states to maintain security and political control will lead to more blocking, filtering, segmentation, and balkanization of the Internet.
- Trust will evaporate in the wake of revelations about government and corporate surveillance and likely greater surveillance in the future.
- Commercial pressures affecting everything from Internet architecture to the flow of information will endanger the open structure of online life.
- Efforts to fix the TMI (too much information) problem might over-compensate and actually thwart content sharing.

Questions [DELETE?]

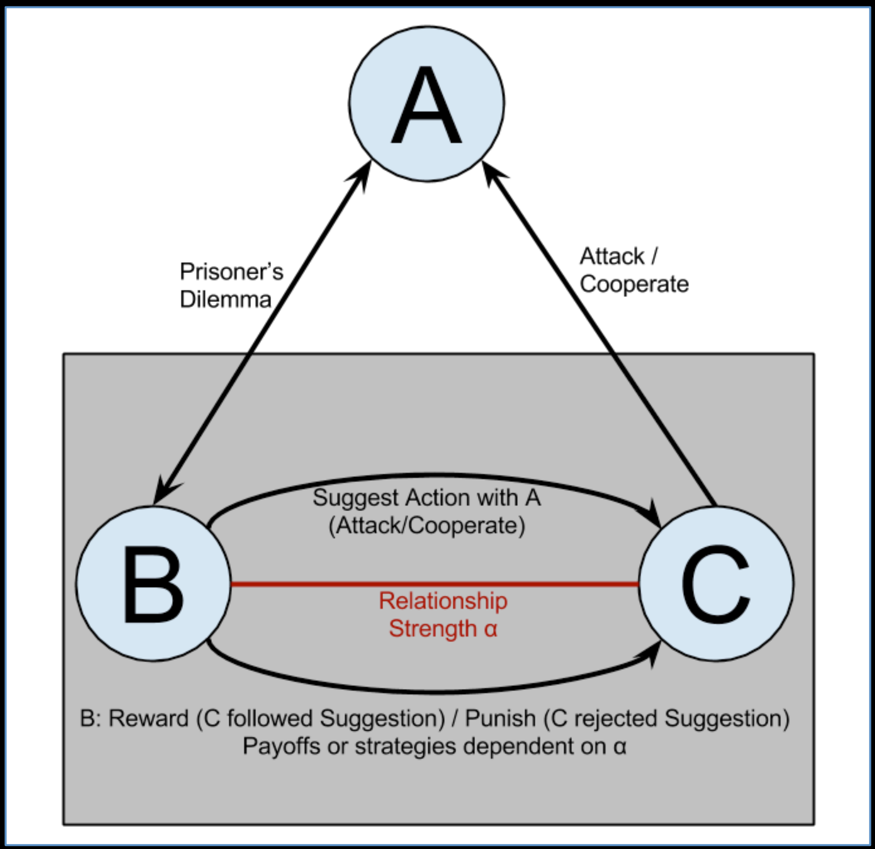
- What interventions will be most successful?
- Where should they be deployed?
- What are the unintended consequences of increasing cybersecurity?
- How does cybersecurity affect
  - individual psychology,
  - social structures,
  - economic systems,
  - political institutions

Attribution of Cyberattacks
UNM and Univ. of Michigan



- Originating source may not be the originating actor
  - 2008 cyberattacks against Georgia (Russia?)
- Attacks can be hard to distinguish from other issues
  - NK Internet outages post-Sony
- Doesn't involve physical material
  - Technical attribution may not be sufficient
  - Easier to fake
- Evidence is more distributed and may be controlled by adversary
  - Proving attribution to the public may require revealing hidden information (assets or capabilities)

Responsibility Game
DRAFT, in progress



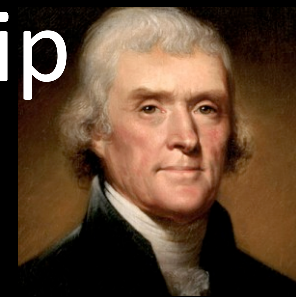
- Neighbor (A)
- Parent (B)
- Child/Dog (C)
- Examples:
  - US/China/PLA
  - Israel/Hamas/PIJ
  - Georgia/Russia/Hackers

Questions

- What are reasonable payoffs?
  - Analyze historical examples
- What are optimal strategies
  - Under different assumptions about payoffs
- What are the expected outcomes when A's information is incorrect
- When is it strategic to plan false flags?
- Learning?
  - Can A infer the relationship between B and C?
  - A punishes B to encourage B to teach C

Jefferson Science Fellowship

National Academy of Science



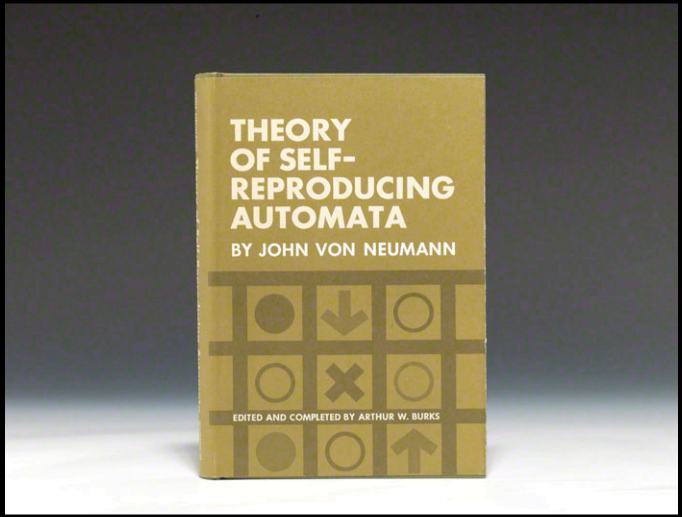
- STE advisors for the State Dept and USAID
  - 1-yr fellowships for tenured professors
  - 13 fellows per year, 3 computer scientists
  - 92 JSFs at the State Department and USAID since 2004

*sites.nationalacademies.org/pga/**jefferson**/*

To preserve the freedom of the human mind then and freedom of the press, every spirit should be ready to devote itself to martyrdom; for as long as we may think as we will, and speak as we think, the condition of man will proceed in improvement.

Thomas Jefferson

Modeling is essential to complexity science



- What will the stock market do tomorrow?
- How stable is the political situation in Egypt?
- Is it possible to build a self-reproducing machine?
- Is the brain a kind of electrical circuit?
- How can we make computers more secure?

Cybersecurity Challenges

- Computer security isn't very secure
  - We need new approaches
- Important to everyone
- Co-evolution
  - Arms race with adversaries
  - Can't expect to solve the problem "once and for all"
  - Need to learn how to manage and live with it
- Moore's law helps adversaries and defenders

http://www.simplysecurity.com/2011/09/27/the-shifting-motivations-behind-digital-threats-infographic/

Hierarchies (TPRC, 2014)

- Security enhancements added hierarchy to decentralized Internet design
  - DNSSEC, RPKI, SSL PKI, DANE, ….
- Hierarchy provides a convenient locus of control for policy interventions
  - Law enforcement, copyright enforcement, censorship, etc.
- Unintended consequences
  - Local laws have global effect
  - Loss of trust in voluntary security enhancements
- Policy impacts of takedowns deserve reconsideration

Suggested Principles

- Restraint: from using core Internet infrastructures for policy interventions
- Move interventions to the network edge and up to application layer
- Security enhancements should themselves be decentralized, like the Internet

Policy Questions

- U.S. policy on botnet takedowns?
- Why is spam such an important problem?

Which countries are most at risk?

- Recent Development
  - Increase in Internet Connectivity
  - Increase in GDP
  - Increase in Internet Users

Botnet Takedowns

- Command and Control Takedowns are a common intervention
- Government and private organizations partner to 'attack the attacker'
  - Sever communication between the botmaster and the botnet
  - Take control of attacker's computer
  - Clean up the zombie computers
- Technically sophisticated approach
  - Makes headlines for security companies
  - Labor intensive
- Question: how effective are botnet takedowns?

HeartBleed

- History
  - Introduced accidentally Dec. 31, 2011
  - Discovered April 1, 2014
  - Affected at least 500,000 trusted websites
- Heartbleed bug affects encrypted communications, e.g., https
  - Theft of private keys, session keys, passwords
  - Caused by a common programming error (buffer overflow)
- Policy note: NSA/WH publicly denied prior knowledge of Heartbleed

Parameters

- Fitness: Weighted sum of test cases that the program passes:
  - Programs that don't compile = 0 points
  - 10 points for a negative test case, 1 point for a positive test case
  - e.g., 7 different fitness values for initial experiments
- Std. run
  - Population size: 40
  - Run for 10 generations
  - 1 mutation per indiv. per gen.
  - Each individual participates in 1 crossover per gen.
- Test suite sampling and parallelism

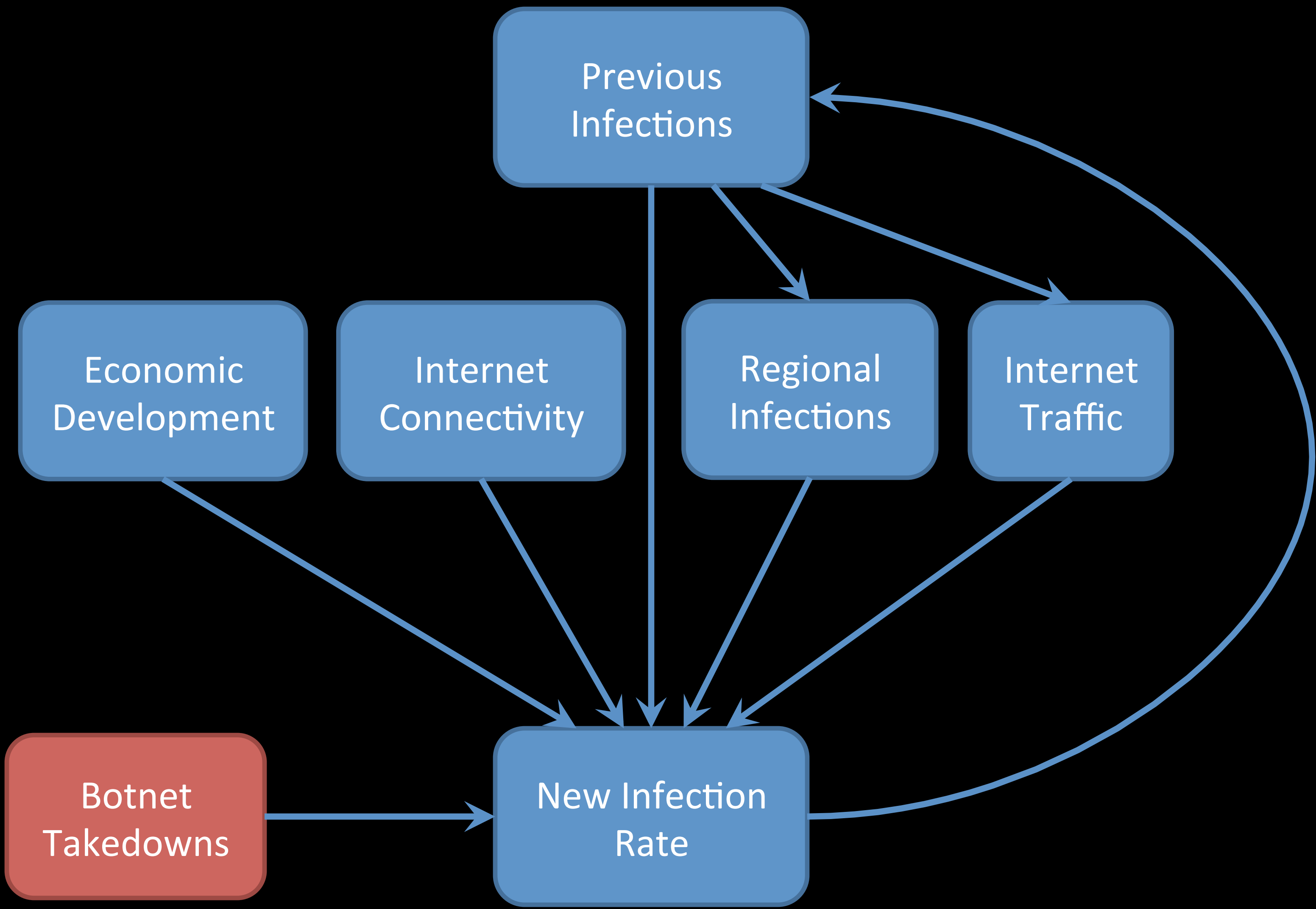
Global Cyberpolicy Issues

- Privacy and surveillance
  - Cloud computing and data localization
- Who should control the Internet?
  - Net neutrality
- Cyberwarfare and economic espionage
- Data breaches
- Zero-day exploits
- Spam
  - Botnet takedowns, filtering, capacity building

What interventions best control the spread of malware and enhance security?

- **Filter:** Detect and isolate malicious behavior
- **Repair:** Patch vulnerability, Replace vulnerable system
- **Counterattack:** Target attacker to prevent further attack
- **Observe:** Gather additional information
  Deceive/Hide:
  - Provide false information
  - Obscure target or its contents

Studying Botnet Takedowns



How do we repair bugs now?

- We ignore them
- We pay expensive programmers to fix them manually
- We develop tools to help the programmers
  - Debuggers, profilers, smart compilers
  - Type checkers
- Mathematical models of program correctness
  - Don't scale up to production software

Responses to Malicious Behavior

Observe

Hide

Filter/Quarantine

Repair

Counterattack

Responses to Malicious Behavior

Observe

Immune system surveillance

Intrusion-detection systems

Biopsies and other screening

UN inspections of nuclear programs

Responses to Malicious Behavior

Hide or Disguise

Tor network

Address Space Randomization

Advertise false descriptors

Bully avoidance

Mimicry in biology

Camoflauge (snowshoe hare)

Responses to Malicious Behavior

Filter/Quarantine

Blacklisting malicious IP addresses

Spam filters

Censorship

Rate limiters

Excise a tumor (disable from interacting with system)

Public health quarantines

Mucous membranes

Prison ??

Sanctions

Responses to Malicious Behavior

Repair/Replace

Software patches

Gene editing (therapy)

Bone marrow transplants

Responses to Malicious Behavior

Counterattack

Botnet or DNS takedowns

Asset seizure

Chemo- and radiation-therapy

Macrophages and killer T-cells

Milatary action

Concluding Thoughts

Self-interested Actors
Evolution

- Quarantine (excise a tumor)
- Patch (gene replacement therapy)
- Filter (chemoprevention, e.g., statins for lowering cholesterol)
- Replacement (Bone marrow transplants for leukemia)
- Counterattack (chemotherapy)

Complex Systems

Interactions

Systems composed of interacting components

Emergence

Scale

Evolution and Learning

Complex Systems

Interactions

Systems composed of interacting components

Emergence

Behavior emerges from interactions among components and between components and their environment

Scale

Evolution and Learning

Complex Systems

Interactions

Systems composed of interacting components

Emergence

Structure and behavior emerges from interactions among components and between components and their environment

Scale

Systems are nested and structure/behavior emerges at different scales

Evolution and Learning

Complex (Adaptive) Systems

Interactions

Systems composed of interacting components

Emergence

Structure and behavior emerges from interactions among components and between components and their environment

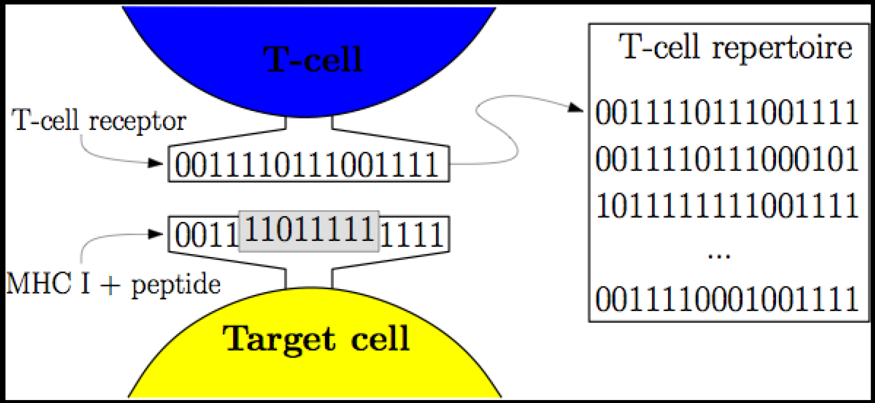
Scale

Systems are nested and structure/behavior emerges at different scales

Evolution and Learning

Systems are dynamic and adapt to internal and external conditions

Computer Immune Systems

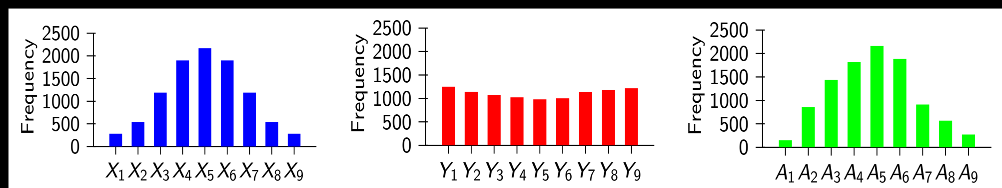


- Self/non-self discrimination (1994)
- Anomaly intrusion detection (1996)
- Automated response to attacks (2000)
- Privacy-preserving data collection (2012)

Private Data Collection

with B. Edwards, F. Esponda, M. Groat, J. Horey, W. He

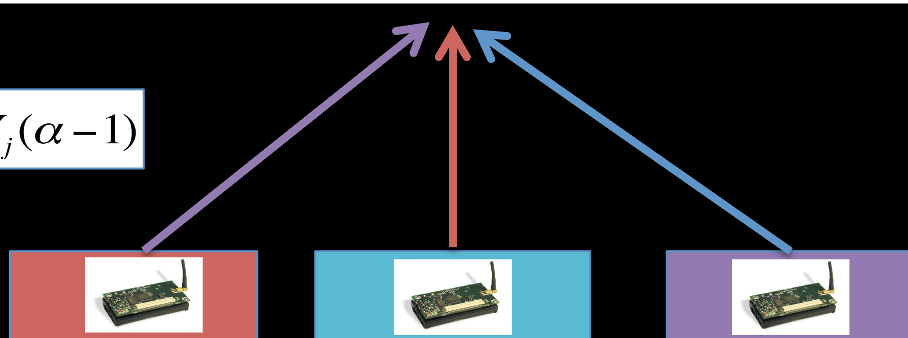
Original Dist. X Perturbed Dist. Y Reconstructed Dist. A



Negative Survey

Mobiquitous, 2007; PerCom 2012

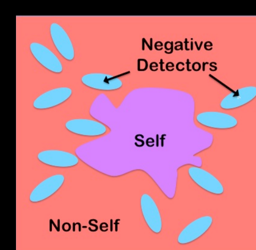
$$\forall j \mid A_j = P - Y_j(\alpha - 1)$$

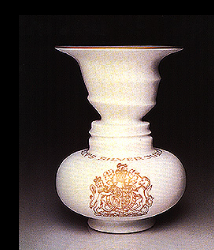


No secrets

No need to trust a central server

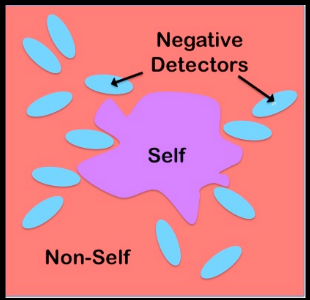
Computationally efficient





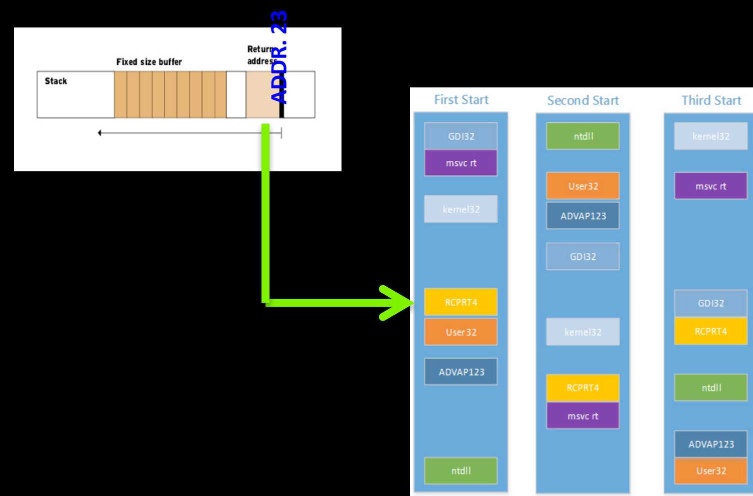
Outcomes

- Immunology: Anomaly IDS
  - First practical anomaly detection system (system calls)
  - Sana Security's Primary Response
- Homeostasis: Graduated response
  - Hewlett Packard's Virus Throttle
  - Pretty Good BGP
- Privacy-preserving data collection and storage

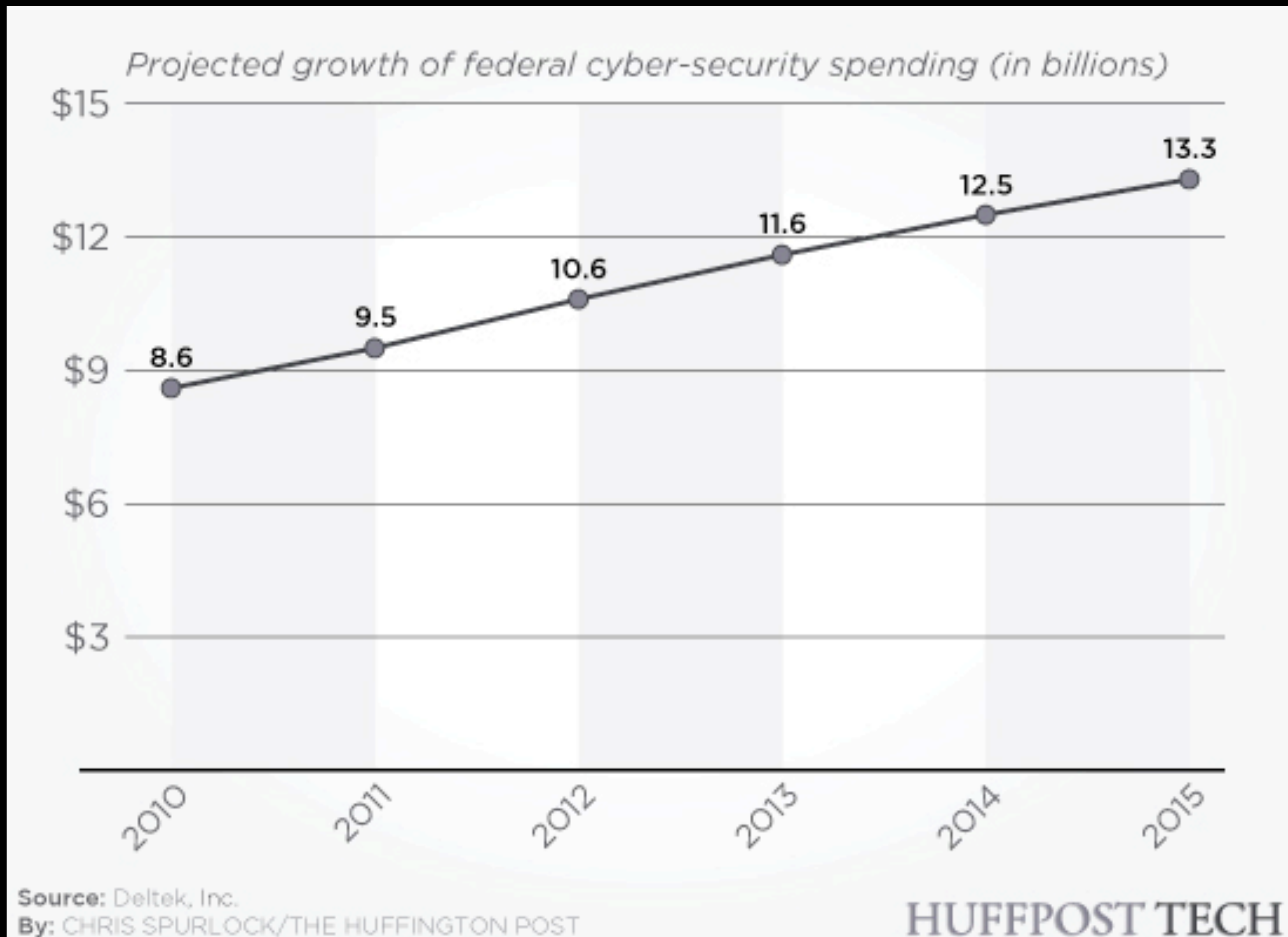


Engineered Diversity

The problem with monoculture



Address Space Layout Randomization


Projected growth of federal cyber-security spending (in billions)
$15
$12
$9
$6
$3
8.6
9.5
10.6
11.6
12.5
13.3
2010
2011
2012
2013
2014
2015
Source: Deltek, Inc.
By: CHRIS SPURLOCK/THE HUFFINGTON POST
HUFFPOST TECH

Spam

- Still a problem
  - $20-50 billion in U.S. (2011)
  - Crushing burden for immature IT infrastructures
  - Excuse for intl. regulation
- Spam is often sent by botnets
  - Campaigns cause spikes
  - Dynamic and noisy
- Mitigations
  - Filter
  - Disrupt credit card payments
  - Botnet takedowns

Conclusions

- Perception that we are losing ground
  - Are we?
  - How would we know?
- Co-evolution
  - How do we learn to manage and live with cyber-issues?
  - The Red Queen

Data Breaches: Fitting Distributions to Data