What Biology Can (and Can't) Teach us about Security

USENIX Security Symposium
San Diego, CA
12 August 2004

David Evans
University of Virginia
Computer Science

Nature vs. the Real World
(Computer Systems)

- Competition for limited resources
- Parasites that can't reproduce on their own steal resources from others
- Can take millions of years to evolve solutions to known security problems

- Competition for limited resources
- "The next geeky kid frustrated about not getting a date on Saturday night will come along and do the same thing without really understanding the consequences. So either we should make it a law that all geeks have dates – I'd have supported such a law when I was a teenager - or the blame is really on the companies who sell and install the systems that are quite that fragile." Linus Torvalds, NYT Sept 2003
- 25 years later, buffer overflows are still the main problem

Brute Force Attacks

Image courtesy Leeson Photography

Communication Integrity Attacks

Image © Australian Museum

Bolas Spider

- Emits chemicals that mimic pheromones of female moth
- Eats the male moths
- Very specialized: moth pheromones are species-unique blends of chemicals
  - Bolas can attract 2 different species
  - Adjusts its emissions based on time of night each moth is active

Critter-in-the-Middle Attacks (CITM)

Blister beetle (*Meloe franciscanus*)

Beetle larvae aggregate to look (and smell!) like a female bee

Male bee tries to have sex with it; fails, but beetle larvae stick to him

Male bee finds real female bee, beetle larvae transferred

Beetles get fat eating pollen female bee brings for her children

Hafernik and Saul-Gershenz (2000)
Images from *Iziko Museums of Cape Town*

Outline

- ✓Nature has security problems and solutions
- **Process**
  - **Evolution**
- Programs (what results): genotype
- Executions (what they produce): phenotype

Let's set the existence-of-God issue aside for a later volume, and just stipulate that in some way, self-replicating organisms came into existence on this planet and immediately began trying to get rid of each other, either by spamming their environments with rough copies of themselves, or by more direct means which hardly need to be belabored. Most of them failed, and their genetic legacy was erased from the universe forever, but a few found some way to survive and to propagate. After about three billion years of this sometimes zany, frequently tedious fugue of carnality and carnage, Godfrey Waterhouse IV was born...

Neal Stephenson, *Cryptonomicon*

Like every other creature on the face of the earth, Godfrey was, by birthright, a stupendous badass, albeit in the somewhat narrow technical sense that he could trace his ancestry back up a long line of slightly less highly evolved stupendous badasses to the first self-replicating gizmo --- which, given the number and variety of its descendants, might justifiably be described as the most stupendous badass of all time. Everyone and everything that wasn't a stupendous badass was dead.

Neal Stephenson, *Cryptonomicon*

Remarkable Existence

- Every one of your ancestors survived long enough to reproduce!
- Probability of surviving to reproduce ~ 0.8
- Number of human generations ~ 3000

$(0.8)^{3000} = 1/10^{291}$ ~~$1/10^{291}$~~ 1

- But, don't stop with humans…

Two Important Clarifications

- Its all about reproduction:

  Survival is necessary but not sufficient

- Unit is *gene*, not *organism*
  - An animal is just a vessel for propagating genes
  - An organism may appear to act unselfishly, but genes are always selfish (even if cooperating in groups is a good strategy)

Richard Dawkins, *The Selfish Gene*

Evolutionary Computing

- Genetic Algorithms
  - Very impressive results on optimization problems
- Genetic Programming
  - Demonstrated inventiveness
    - 2 patentable inventions, 21 infringing [Koza]

Why the process won't help us

- Really slow
  - 3 Billion years of evolution on Earth
- Almost always fails
  - ~99.9% of species become extinct
- Can't reason about results
  - Happened to thrive in this particular environment…no idea what will happen in a different one

…but the Results Can

- Process
  - Evolution
- **Programs (what results)**
  - **Genotype**
- Executions (what they produce)
  - Phenotype

Brief History of Biology

Life is about magic. ("vitalism")

Life is about chemistry.

Life is about information.

Life is about computation.

Most biologists work on Classification
Aristotle (~300BC) - genera and species

Descartes (1641)
explain life mechanically

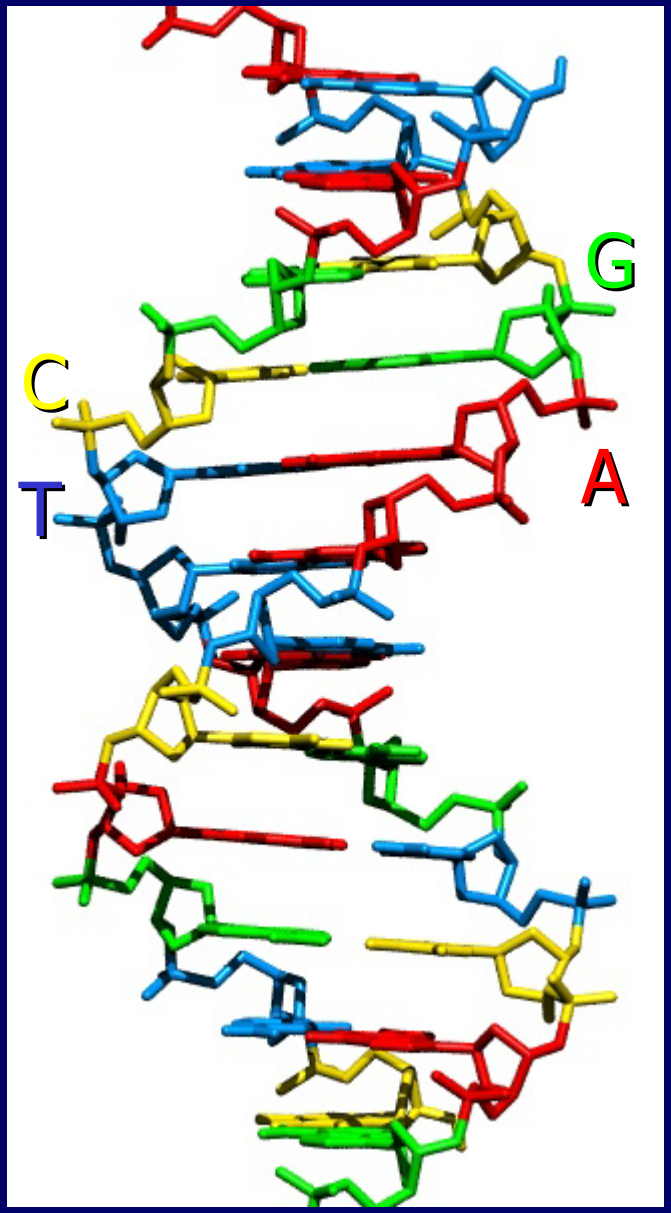
Schrödinger (1944)
life is information
crack the information code

Oswald Avery (1944)
information separate from proteins

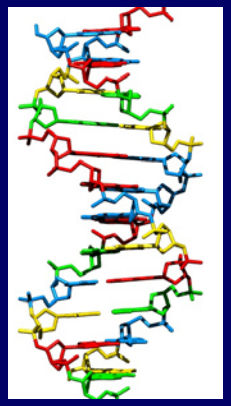
Watson and Crick (1953)
DNA structure, replication

DNA

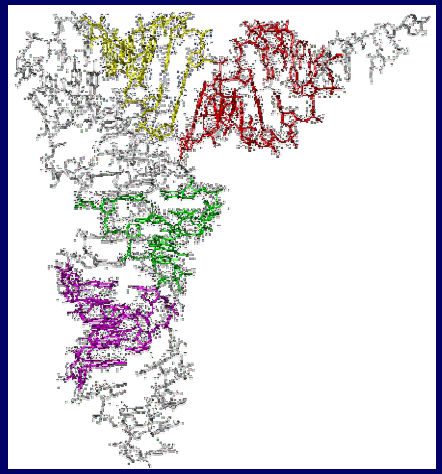
- Sequence of nucleotides: adenine (A), guanine (G), cytosine (C), and thymine (T)
- Groups of three nucleotides (codons) encode amino acids (20) and stop/start
- Two strands, A must attach to T and G must attach to C
  - Enables copying and transcription



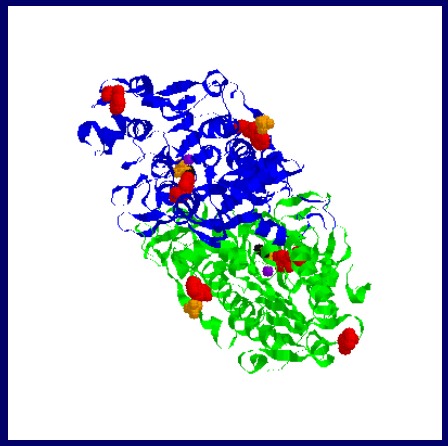
Central Dogma of Biology (Crick, 1957)



Transcription



Translation



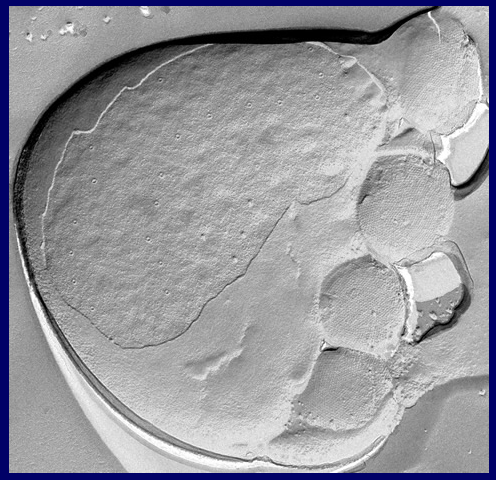
DNA RNA Protein

Image from http://www.umich.edu/~protein/

- DNA makes RNA
- RNA leaves nucleus and makes proteins
- Proteins make people

Shortest (Known) Life Program

- *Nanoarchaeum equitans*
  - 490,885 bases (522 genes) ≈ 40 KB
  - Parasite: no metabolic capacity, must steal from host (smallest autonomous ~ 1.6 million bases)
  - Complete components for information processing: transcription, replication, enzymes for DNA repair



http://www.mediscover.net/Extremophiles.cfm
KO Stetter and Dr Rachel Reinhard

- Size of compiling C++ "Hello World":

  Windows (bcc32): 112,640 bytes

  Linux (g++): 11,358 bytes

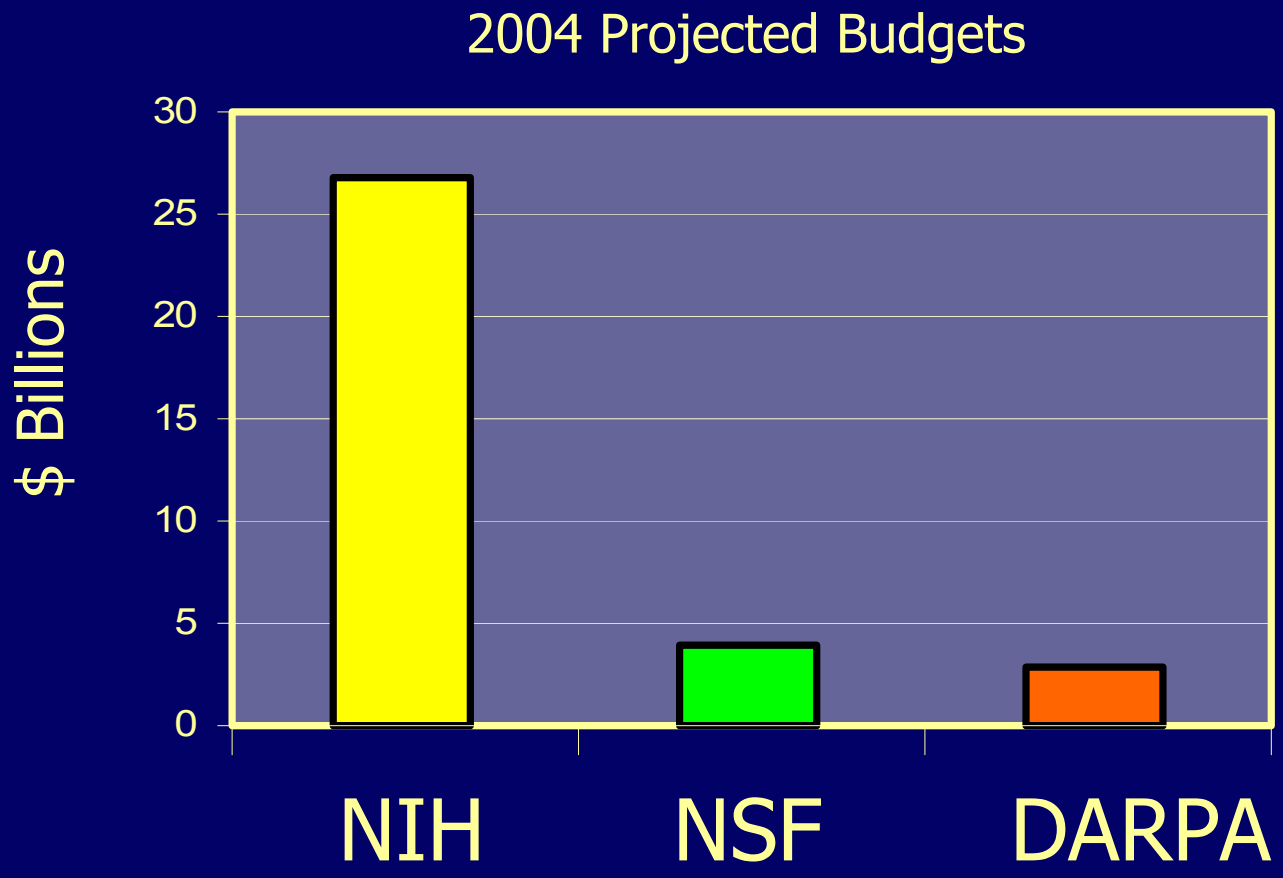
The Make-Human Program

- 3 Billion Base Pairs
  - Each nucleotide is 2 bits (4 possibilities)
  - 3B bases * 1 byte/4 pairs = 750 MB
  - Highly redundant encoding (21/64) ~ 250 MB
- Only ~5% is transcribed (exons) ~ 12 MB
  - 95% junk (introns): genomes from viruses reverse transcribed into human genome, but inactive

Expressiveness of DNA

Genetic sequence for 2 humans differs in only 2 million bases

< 1/2 of a floppy disk
<1% of Windows 2000



Why Haven't We Cured Cancer Yet?

Gene Interactions

- Not so simple: cells in an organism have the same DNA, but do different things
  - Structural genes: make proteins that make us
  - Regulator genes: control rate of transcription of other genes

The genome contains not only a series of blue-prints, but a coordinated program of protein synthesis and the means for controlling its execution.
François Jacob and Jacques Monod, 1961

Saccharomyces Cerevisiae
(Yeast) Protein Interactions,
4825 proteins, ~15,000 interactions
Bader and Hogue, Nature 2002

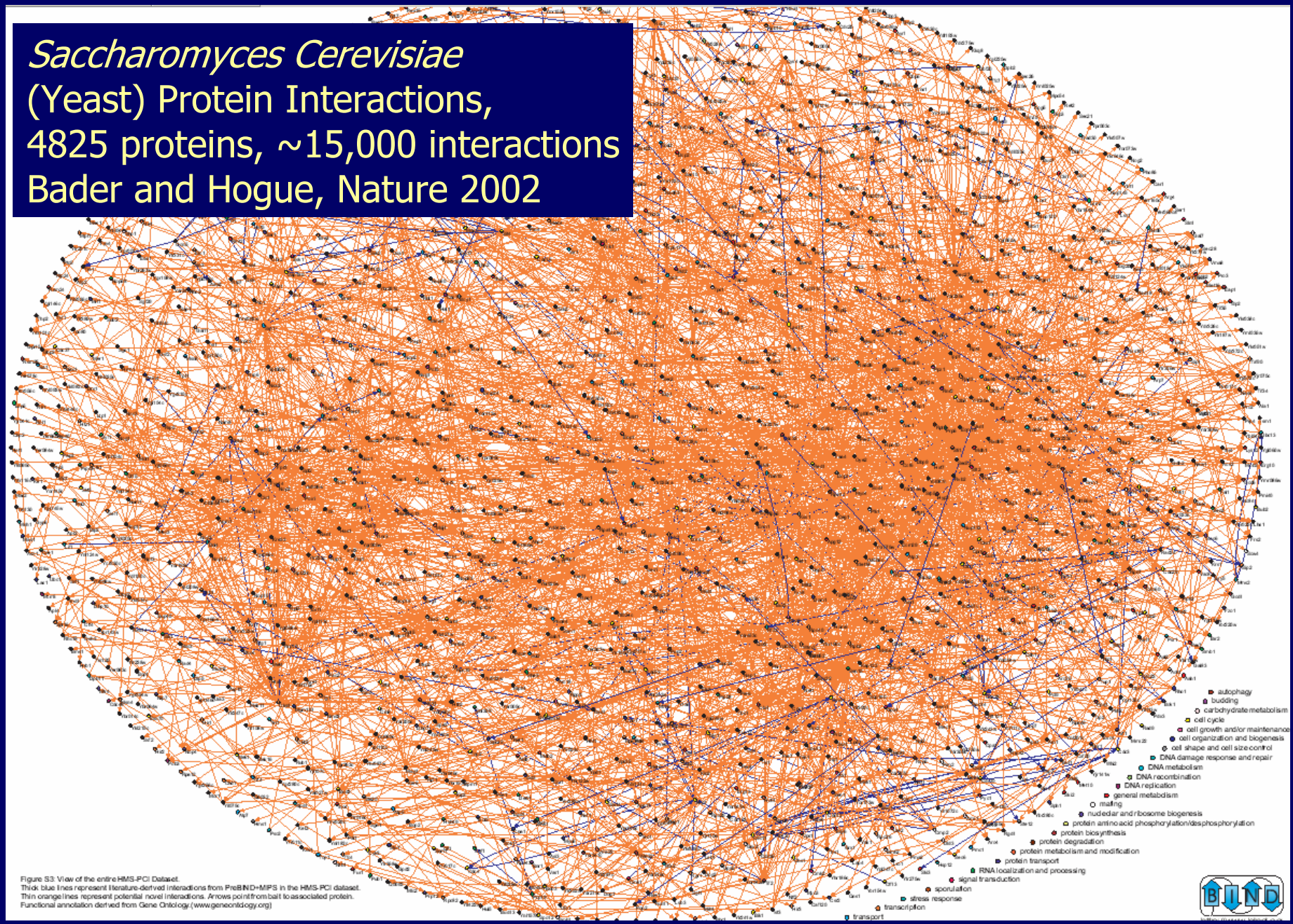


Figure S3: View of the entire HMS-PCI Dataset.
Thick blue lines represent literature-derived interactions from PreBIND+MIPS in the HMS-PCI dataset.
Thin orange lines represent potential novel interactions. Arrows point from bait to associated protein.
Functional annotation derived from Gene Ontology (www.geneontology.org)

Split Genes

- Richard Roberts and Phillip Sharp, 1977
- Not so simple – genome is spaghetti code (exons) with lots of noops/comments (introns)
- Exons can be spliced together in different ways before transcription
- Possible to produce 100s of different proteins from one gene

Why Biologists Haven't Done Much Useful with the Human Genome Yet

They are trying to debug highly concurrent, asynchronous, type-unsafe, multiple entry/exit, self-modifying programs that create programs that create programs running on an undocumented, unstable, environmentally-sensitive OS by looking at the bits (and figuring out what any instruction does is an NP-hard problem)

Observations About Nature's Programs

- Expressive
- Redundant
- Aware of Surroundings
- Localized

- Cannot be rebooted, install patches, etc. (except for humans with medicine)

Need for Robustness

- Evolution selects based on phenotype
- For natural selection to work, there must be a stable, reliable mapping between genotypes and phenotypes
  - Organism must develop successfully
  - Environment is variable
  - Transcription errors will occur

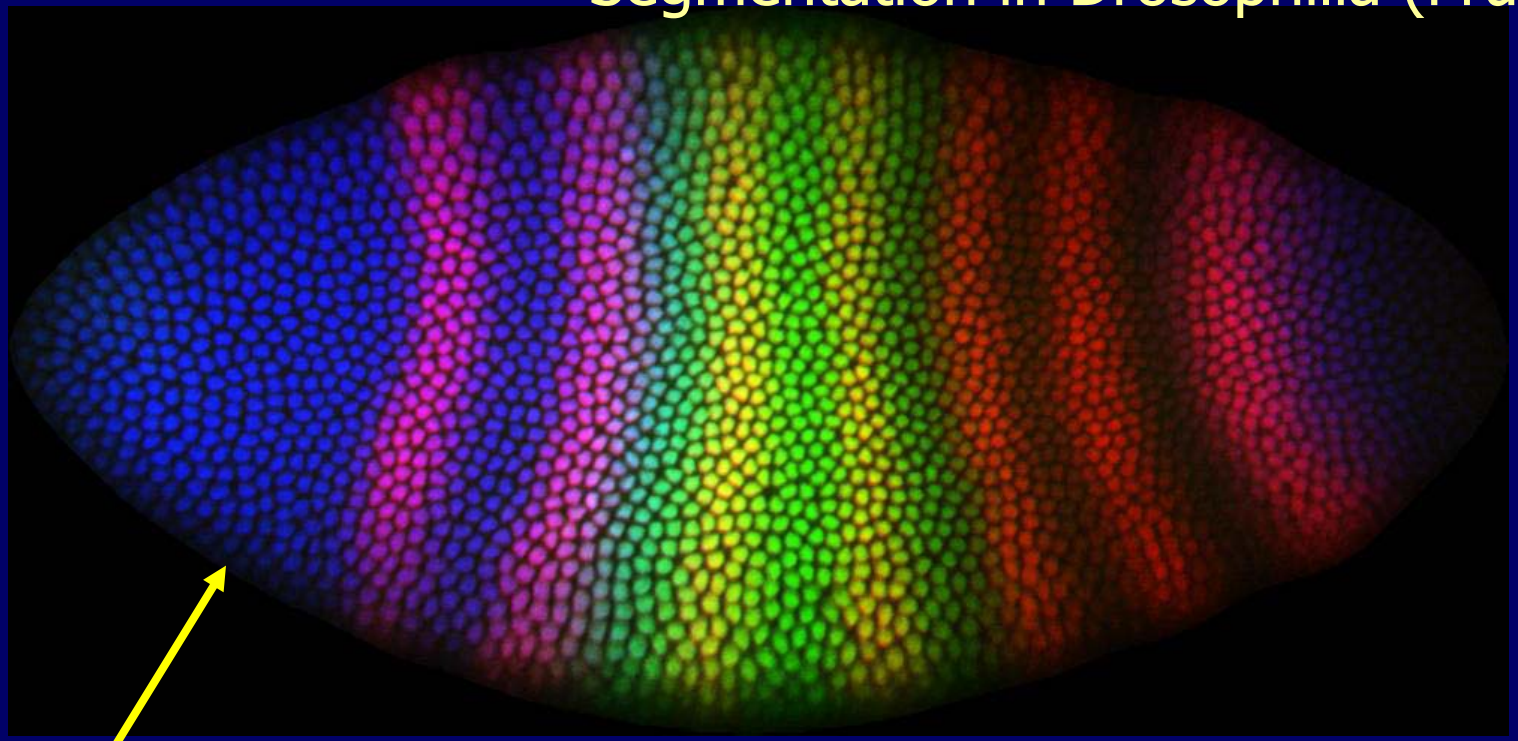
Redundancy

- Gene: Multiple ways to encode one amino acid
- Genome: Multiple copies of genes
- Genetic pathways: multiple regulators
- Metabolic pathways
- Cells: trillions of cells (billions of yours have died since I started talking)
- Organs: multiple organs (2 kidneys)
- Function: different organs can assume same function

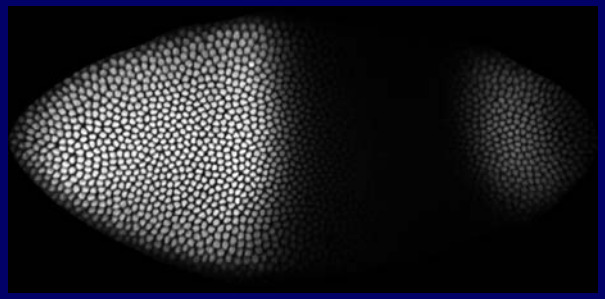
Awareness of Self and Surroundings

Scalable

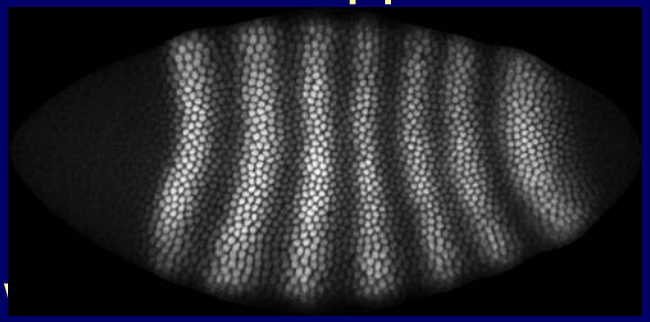
Segmentation in Drosophilia (Fruit Fly)



Hunchback



Even-Skipped



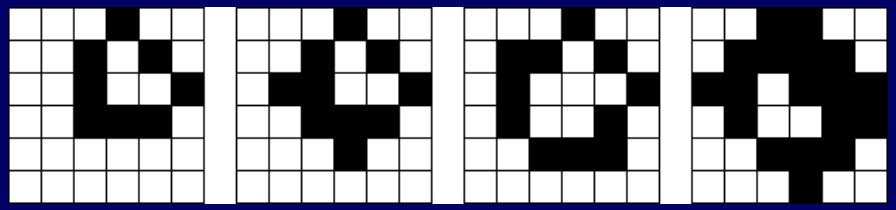
Images from FlyEx,
© David Kosman
and John Reinitz

Mimicking Nature's Programs

- If we can build programs that are:
  - Redundant
  - Aware of Surroundings
  - Localized

  will they share nature's scalability, robustness, survivability properties?

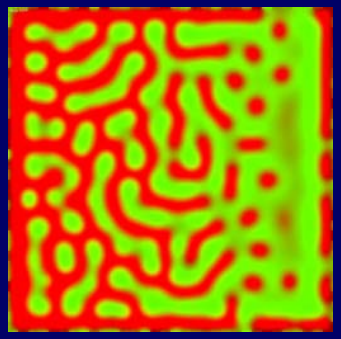
Foundations



Cellular Automata

von Neumann [1940s]

Conway's Game of Life [1970]

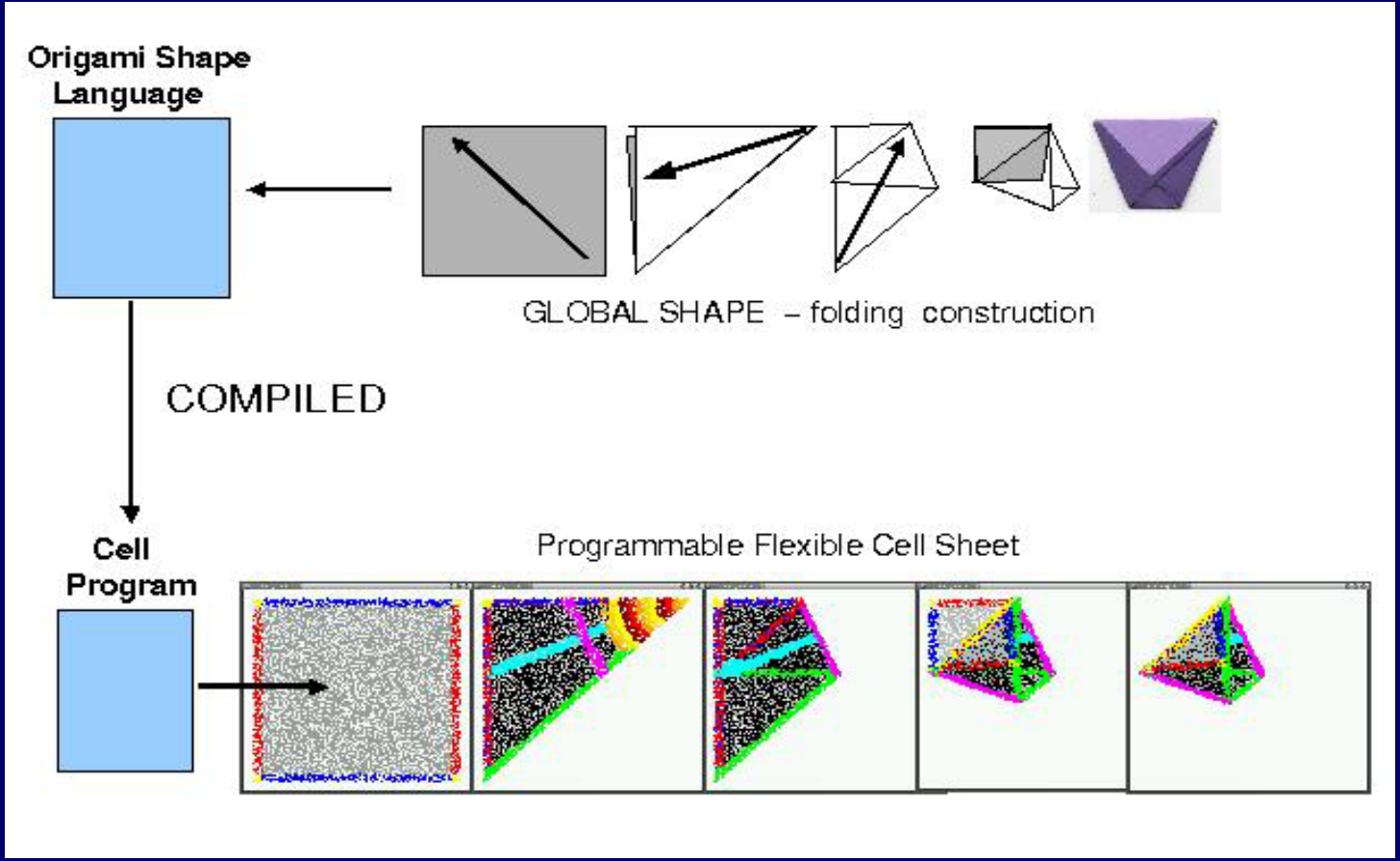


Reaction-Diffusion
Turing [1952]

Recent Work

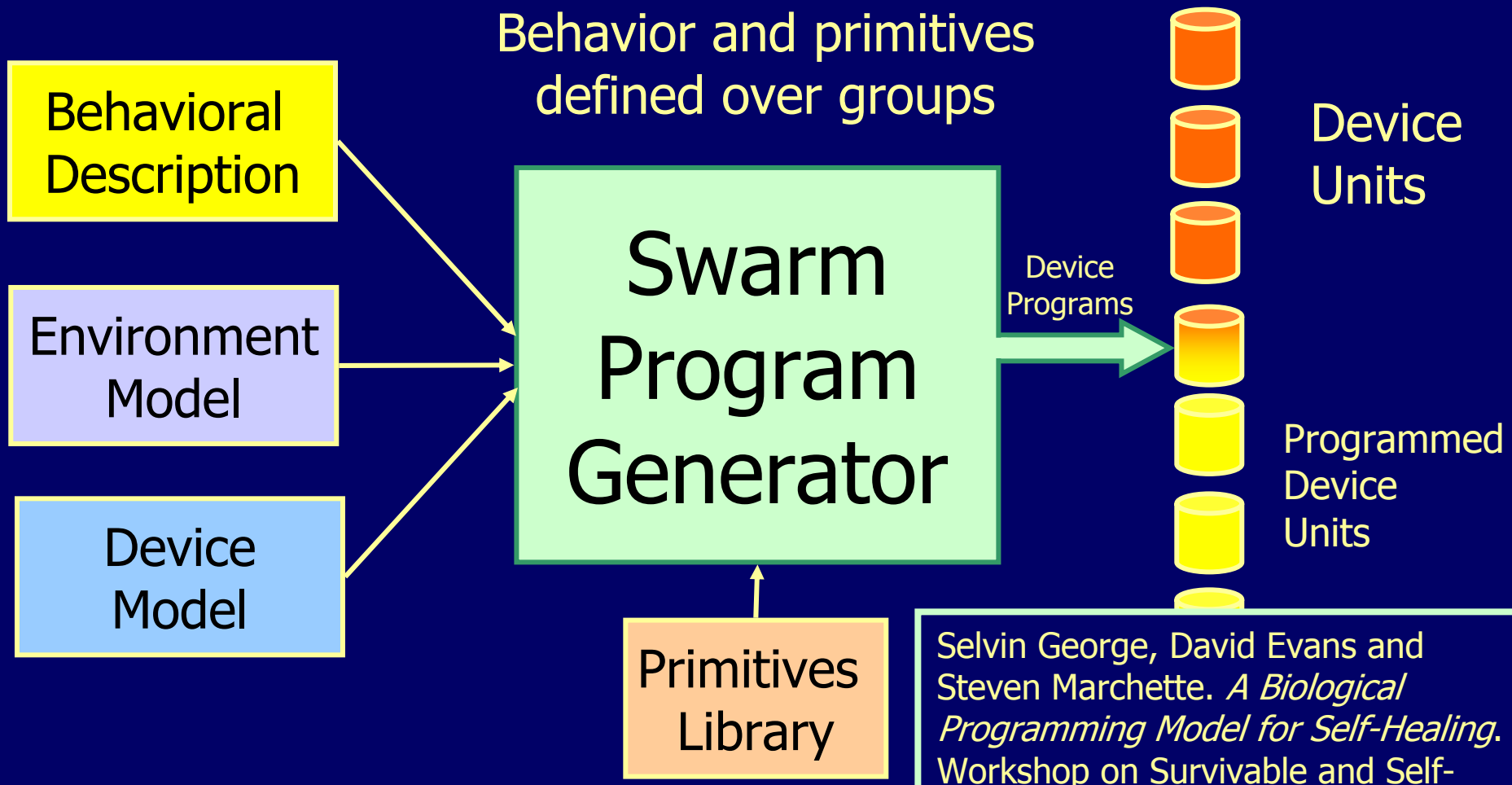
- Amorphous Computing [Abelson, Nagpal, Sussman]
- IBM's Autonomic Computing
- Embryonics [Mange, Sipper]
- Ant Colony Optimization, Swarm Intelligence

Origami Shape Language



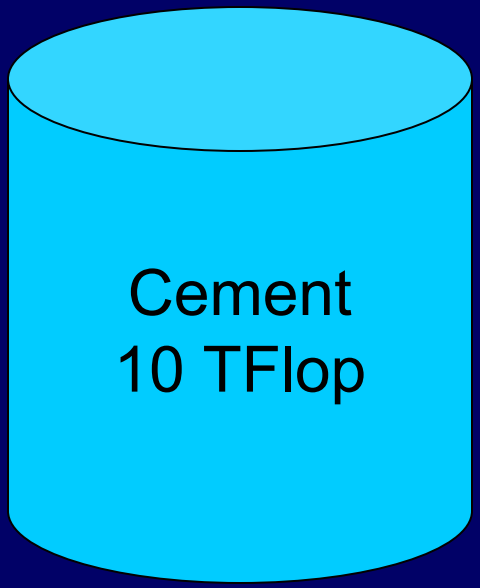
Radhika Nagpal, 2001

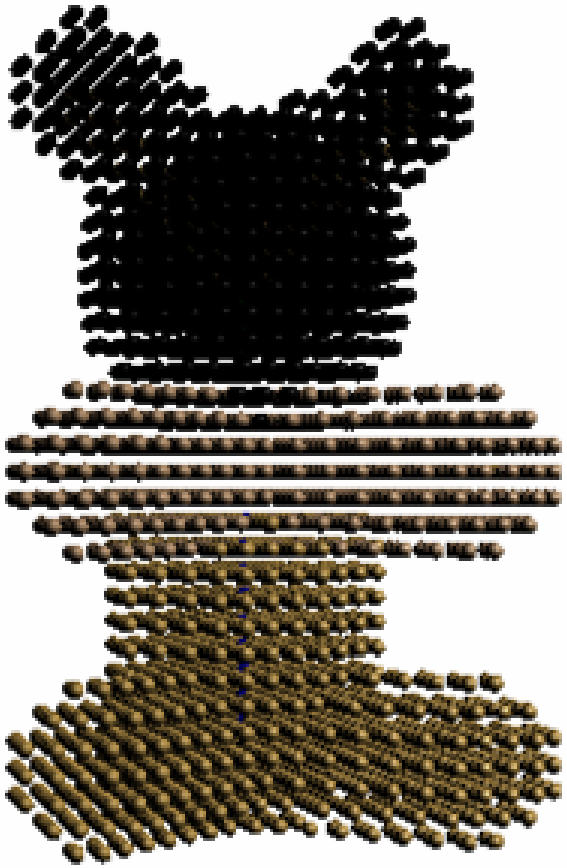
Swarm Programming



Selvin George, David Evans and Steven Marchette. *A Biological Programming Model for Self-Healing*. Workshop on Survivable and Self-Regenerative Systems, 2003.

Long-Range Goal





Mickey Mouse Program

- 20 states
- 50 transition rules
- Starts from one cell, combines lines, spheres
- Regenerates after failure of most cells

Real Mouse Program

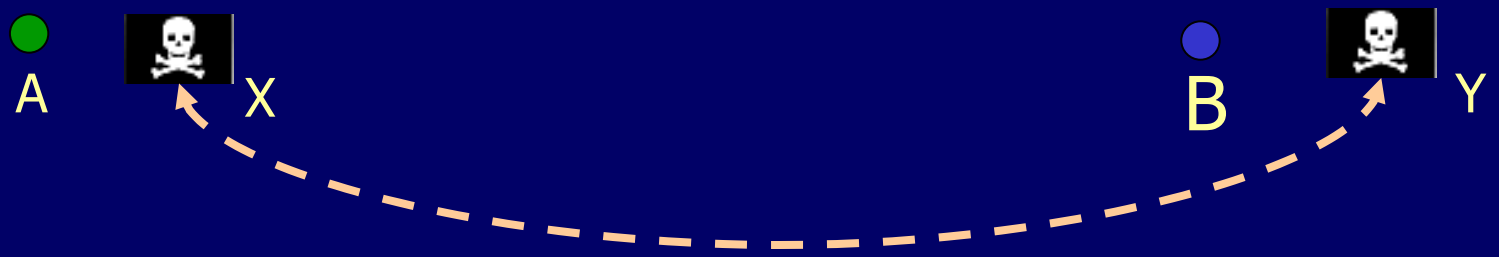
- 3B base pairs
- 98% same as human DNA
- Starts from one cell, complex protein interactions

Towards Real Systems

- Cells
  - Sensor Devices, MEMS, Internet Nodes
- Division
  - Processes
  - Find new hosts
- Communication
  - Point-to-point emissions
  - Wireless multicast (can be multi-hop) diffusions

Exploiting Awareness of Environment: Message Direction

Use Directional Antennas to Mitigate Wormhole Attacks



If *A* hears *B* from its East, *B* should hear *A* from its West

Share information with neighbors to detect all wormholes

Lingxuan Hu and David Evans. *NDSS* 2004.

Mobile Localization: Lingxuan Hu and David Evans. *MobiCom* 2004.

Programs Summary

- Trillions of creatures have died to evolve the extremely robust programs that survive today
- Small programs with complex interactions
- Robustness and scalability require:
  - Redundancy
  - Awareness of surroundings
  - Locality

Outline

- Process
  - Evolution
- Programs (what results)
  - Genotype
- Executions (what they produce)
  - Phenotype
    - ⇨**Some specific examples**
    - ⇨**General Principles**
- Reasons for Pessimism

Reasons for Optimism

Use Prime Numbers

Photo © Hilton Pond Center

Cicada

17-year cycles
13-year cycles

Proof-Carrying Turkeys

Lee Richardson Zoo

Turkey wattles, cockerel comb
Red from carotenoid pigments
- Extracted from food
- Ability to extract is
  - affected by parasites

Software Wattles?

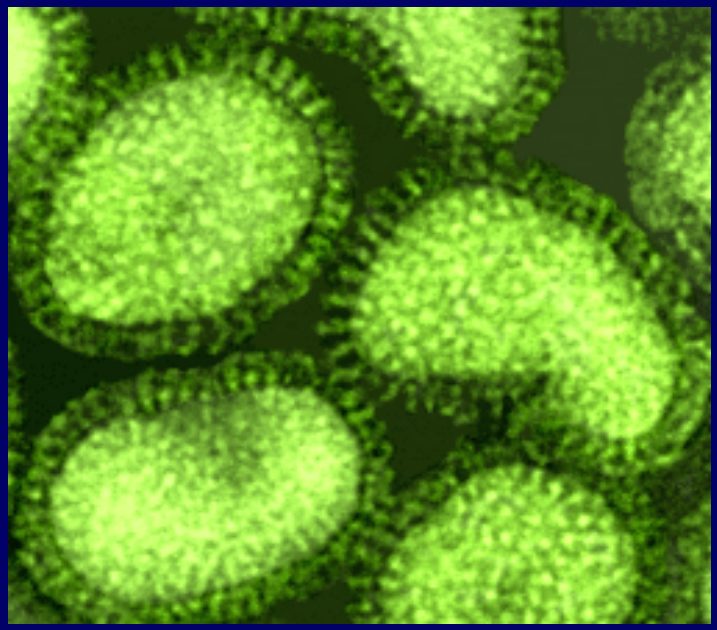
Tripwire, Gene Kim and Eugene Spafford, 1994

Genuinity (Kennel and Jamieson), USENIX 2003
Shankar, Chew, Tygar, this morning
SWATT (Seshadri, Perrig, van Doorn, Khosla). Oakland 2004.

Viruses and Immune Systems

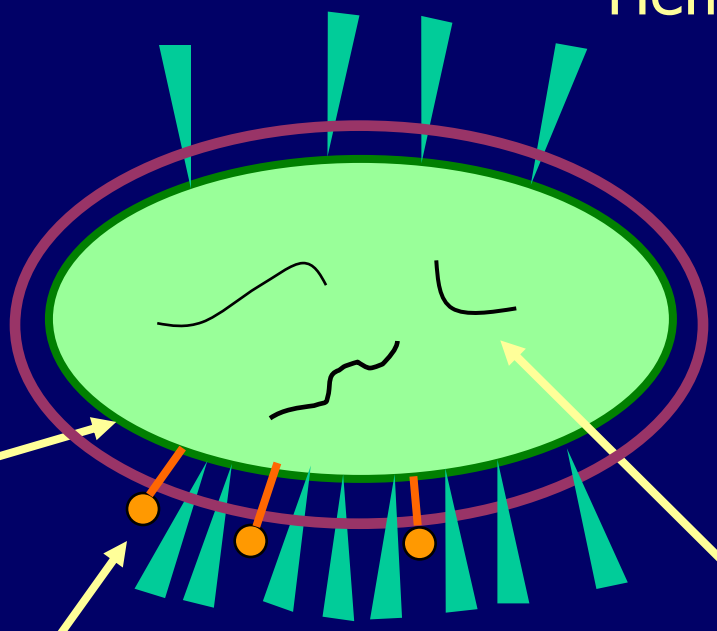
Viruses

- Genetic material (RNA) with protective coat
- Receptor binding proteins attach to cell
- Injects genetic material into cell nucleus
- Uses proteins in cell to reproduce
- Releases copies to infect more cells



Influenza Virus

Influenza Virion



Envelope
(protective shell built using host lipids, proteins, and sugars)

Hemagluttin
~ 400 spikes around virus

Shape of spikes (encoded by virus genome) determines what it can bind to (red blood cells)

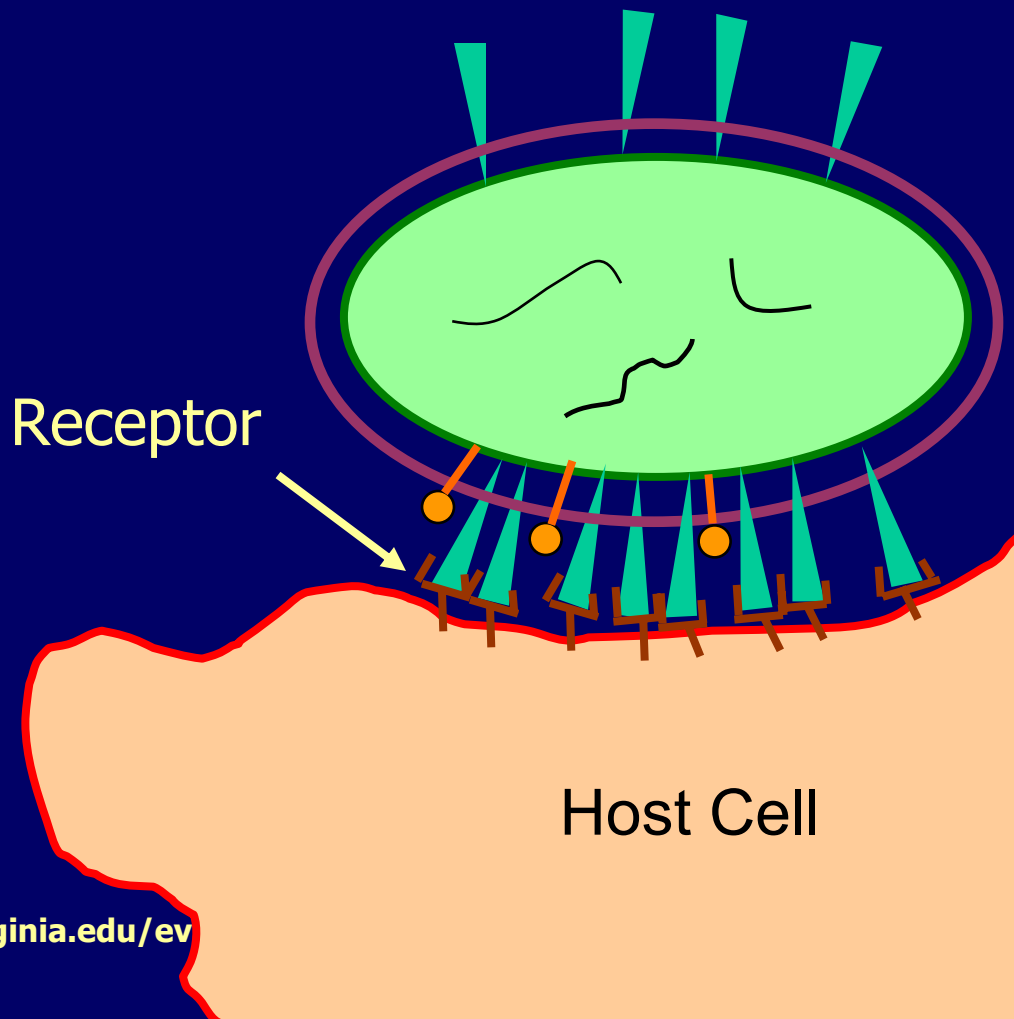
Capsid

RNA

Neuraminidase
Helps new virions escape host cell

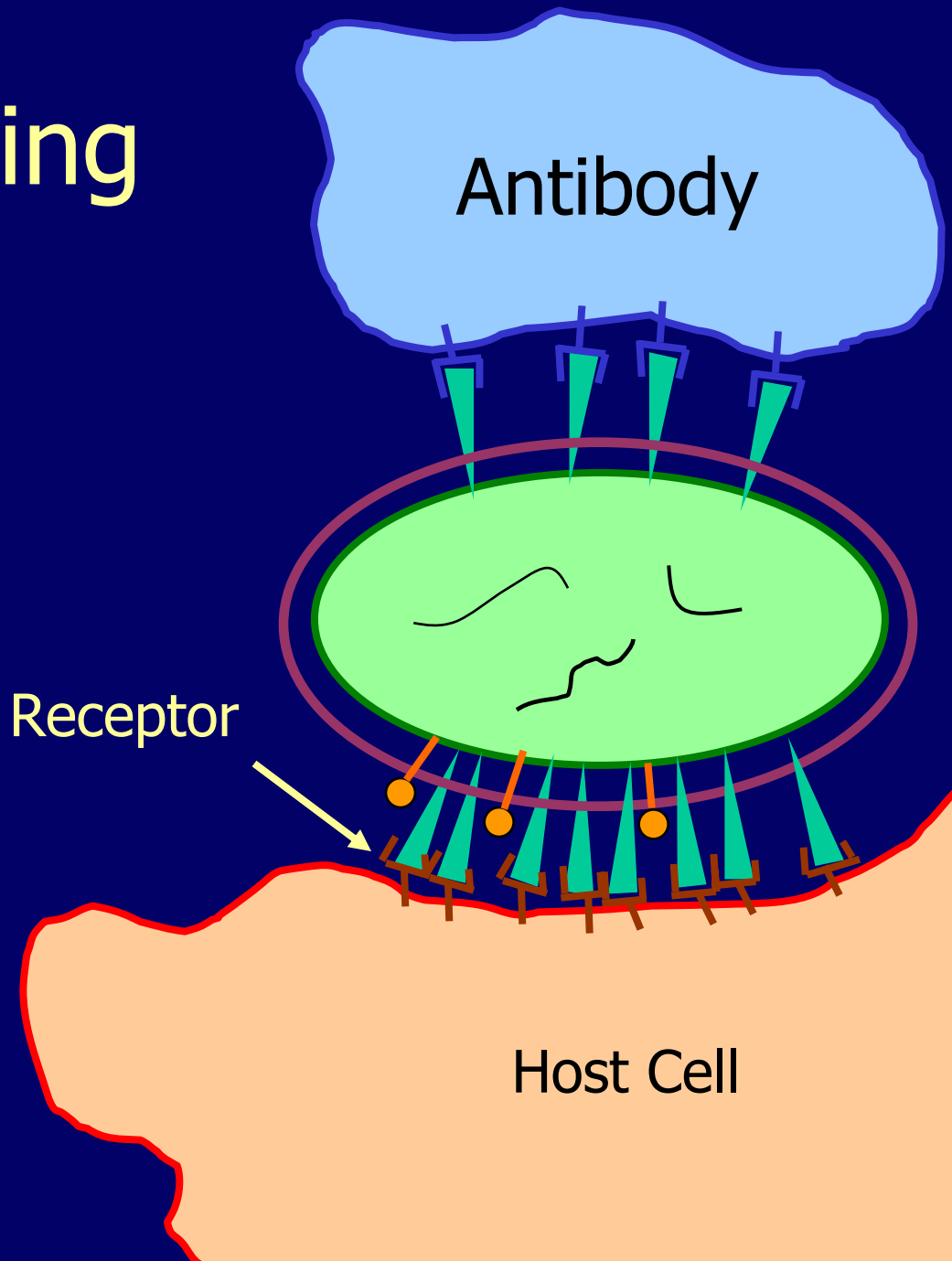
Virus Binding

- To bind to a cell, virus receptor binding proteins must match cell membrane receptors
- Virus is internalized by cell, injects RNA into nucleus



Virus Scanning

Virions have a specific shape to bind to hosts, so scanners (antibodies) can recognize that shape and block virions



Immune Systems vs. Virus Scanners

- Standard Anti-Virus software scans for **known** attacks: compare code against a library of already known attacks
- Approach doesn't work if new viruses emerge quicker than updates
  - Internet Worms: spread time ~ 20 minutes
  - Human: genome updates ~ 20 years

Need a way to detect and defeat previously unknown attacks

Pathogen Diversity

- Genetic drift: random point mutations
  - Some will be successful, and multiply
  - RNA-based viruses mutate very rapidly
- Genetic reassortment: mixing up
  - If two strains of influenza virus infect the same cell, they mix up their genes

Receptor Diversity

- Lymphocytes are white blood cells that have surface proteins to recognize intruders; when stimulated by antigen they make antibodies
- Need to recognize all foreign intruders, but DNA can't know about all ($\sim 10^{16}$) possible intruders
- Gene segments are randomly combined to form different receptors
  - Create 10^7 new lymphocytes every day
  - Lymphocytes that match intruders reproduce quickly (build immunity)
- But, need to ensure lymphocytes don't match self

Recognizing Self

- Major Histocompatibility Complex
  - Surface molecules that are unique to individual on all cells (except red blood cells)
  - Authenticate cell as self
  - Diversity of MHC types protects a population
- Thymus gland
  - Lymphocytes that match self molecules are eliminated, others are mature and enter body

Immune System Disorders

- False negatives are immune deficiencies
- False positives are auto-immune diseases:
  - Reject organ transplants
  - Multiple Sclerosis – motor nerve cells are antigens
  - Rhumatoid Arthritis – connective tissue is antigen

Computer Immunology

- Forrest, Hofmeyr and colleagues, 1994
- Recognize computer intrusions
- Generate library bit-strings that encode patterns of normal behavior (system calls, network connections, etc.)
- Generate random detectors: keep ones that don't match the normal behavior
- Recognize behaviors that are abnormal as possible intrusions

Racing Parasites

- Parasites evolve quickly:
  - E. Coli bacteria ~ 1 hour per generation
  - Influenza virus

- Offspring should be optimized for a *different* environment than their parents – parasites have evolved

Matt Ridley, *The Red Queen*

Achieving Diversity

- Natural selection reduces diversity
  - Will select against inferior genes for particular current environment
- Sex maintains diversity
  - Obtain multiple forms of a gene (AB blood type)
  - Retain currently unfavored genes
  - Opposites attract!
    - Wedekind and Füri found that men and women are attracted to odor of members of opposite sex that have MHC genes most different from themselves

Diversity in Computer Systems

- "A computing monoculture is a danger, a security danger, a national security danger. It is a danger on principle. It is a danger in practice."

  Dan Geer, USENIX Tech 2004

- Microsoft Bashing
  - Client OS (2002): Windows (93%)
  - Client Applications: Office, IE
  - Server OS (2002): Windows (55%), Linux (23%)

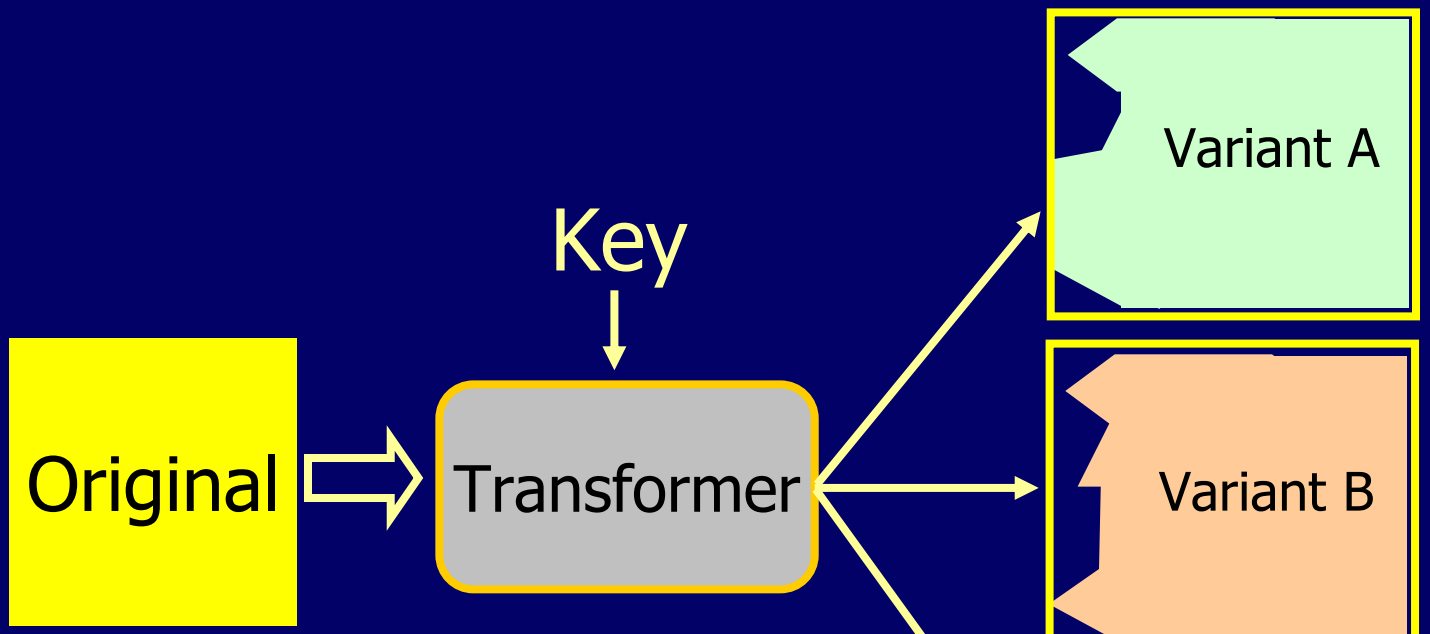
A more competitive marketplace might help...but not enough

Not All Bill's Fault

- Protocols: IP (100%), TCP (>90%), IEEE 802.11b/g, Bluetooth
- Firewall/VPN: ISS BlackICE/Real Secure (~20%)
  - Enough for Witty Worm (12,000 victim hosts in ~45 minutes)
- Image processing code: libPNG
  - Same vulnerability may be exploitable in IE and Mozilla on Mac, Windows, Solaris

Human-engineered diversity is not enough

Automating Diversity



Goal:
Make variants that:
1. Look the same to legitimate users and system administrators
2. Look different to attackers

Diversity Techniques

- Modify instructions, memory
  [Cohen 1992], [Forrest+ 1999]
- System calls, library entry points
  [Chew & Song, 2002]
- Instruction set randomization
  [Barrantes+, CCS 2003] [Kc+, CCS 2003]
- Addresses [Cowan+, USENIX 2003], [Bhatkar+, USENIX 2003]
- Work well against certain code injection attacks

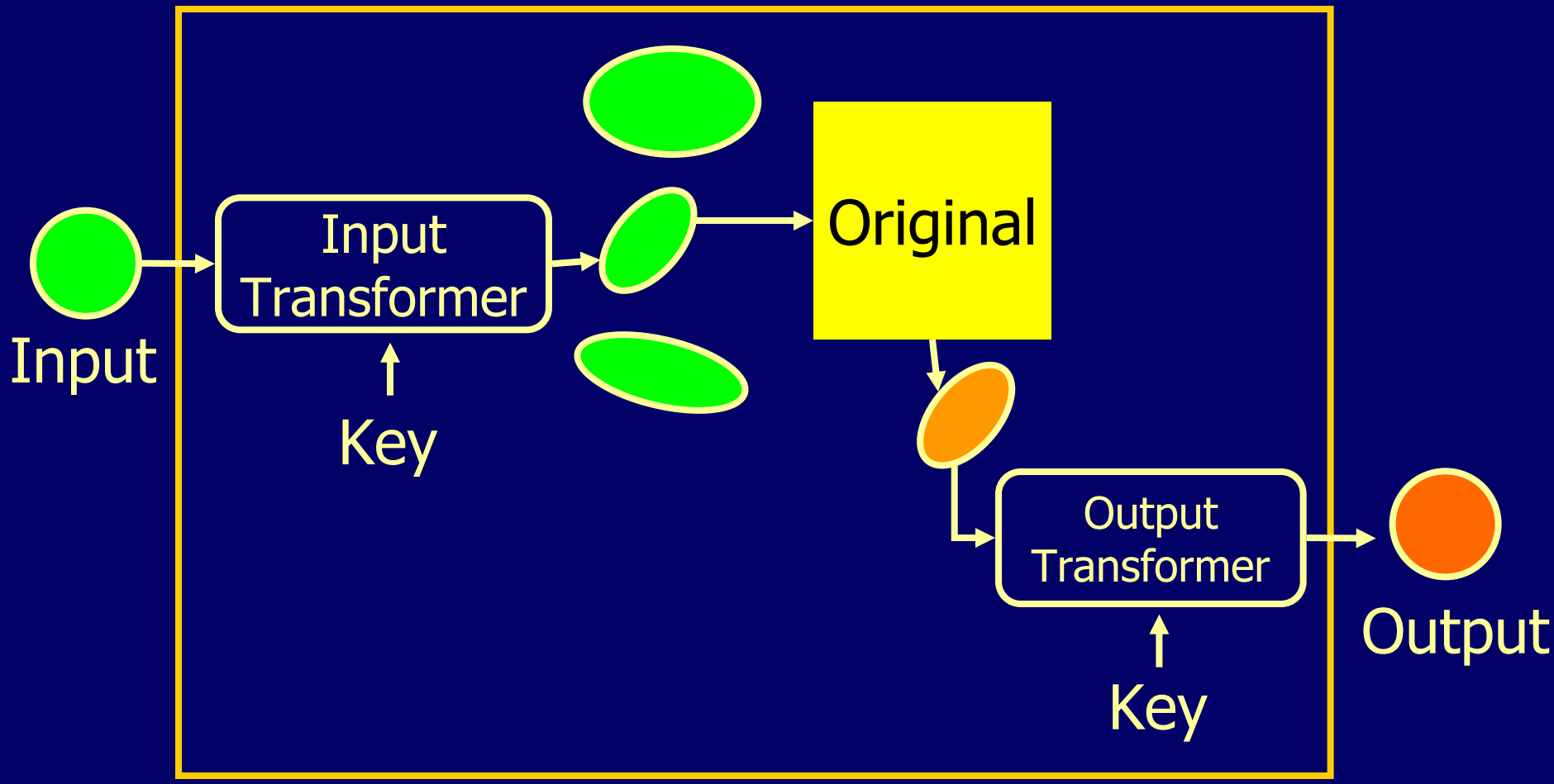
Perspectives on Diversity

To a Great White Shark, all humans are basically alike

Computer systems must be diverse at many abstraction levels to thwart different attacks

Data Diversity

Ammann and Knight, 1998



Contextual Diversity

- Principle of "Less" Privilege
- Rhythmic Policies
  - Divide execution into "beats" and "phrases"
  - Allow certain operations only on particular beats (e.g., only make system calls every 7th beat)
  - Compiler produces "rhythmic" executable
  - VM enforces policy (also transforms dynamically)

Policy Diversity

- Resource Consumption Policies
  - Thresholds should be randomly different in different installations
- Attacker shouldn't be able to guess resource consumption limits
- Can't make limits too low and disrupt normal usage: diversity helps here

Diversity Effectiveness

- Attacker can do a lot of work to break one variant
  - With "luck", that doesn't break other variants
  - Depends on how fundamentally different they are
- If you're worried about point attacks, vary dynamically
- *Security through obscurity* can work if you can generate lots of obscurity cheaply

(and Can't)

- Process
  - Evolution
- Programs (what results)
  - Genotype
- Executions (what they produce)
  - Phenotype
    - ⇨Some specific examples
    - ⇨General Principles
- **Reasons for Pessimism**

Reasons for Optimism

Attacks Computers Face Are Different

- Human engineered, not evolved
- Designed with destruction as a goal

"Evolution is smarter than you are."

Leslie Orgel's Law

- Progress in human attacks is (usually) gradual: Build on old ideas
  - More like virus genetic drift and reassortment
- Biological attacks aren't designed, but scale of evolution makes them fiendishly clever

Out-of-Band/Side-Channel Attacks

- Cryptography: dumpster diving, social engineering, timing, differential power analysis
- Virtual machines: bit flips, convince end user to turn off security

Out-of-Band Attacks in Nature

- Massive Environmental Change
  - Permian mass extinction (248M years ago)
    - 90-95% of species became extinct
- Humans
  - Engineer attacks on particular species
    - Pesticides
    - Antibiotics
    - Vaccine (few eradication successes: smallpox)

Nature Fails Frequently

- Influenza Pandemic of 1918
  - In 2 years, infected 1/5 of world
  - Killed 20-40 million people
- ~99.9% of all species on Earth become extinct; 5% are always becoming extinct
- Everyone dies eventually (even if some genes are immortal)

Conclusion

- Nature has evolved mechanisms that enable species to survive in a hostile world where attackers are evolving much faster
  - Redundancy, Awareness, Diversity
- But…nothing has evolved (or will) to deal with "out-of-band" attacks and nature often fails: we need to do much better!
- Two last lessons from cicadas:
  - Sleep a lot
  - Make a lot of noise when you are awake

Thanks!

http://www.cs.virginia.edu/evans/usenix04

Students:
Selvin George
Salvatore Guarnieri
Lingxuan Hu
Steven Marchette
Nate Paul
Qi Wang
Joel Winstead
Jinlin Yang
Charles Zhang

Insighters:
Jack Davidson, Lance Davidson
Serge Egelman, Úlfar Erlingsson
Kevin Fu, Anita Jones, John Knight,
Barry Lawson, Karl Levitt, Gary McGraw,
Radhika Nagpal, Mike Peck,
Anh Nguyen-Tuong, Avi Rubin,
Jonathan Shapiro, Dawn Song,
Doug Szajda, Chenxi Wang

Funding:
NSF CAREER, ITR (Frederica Darema)
DARPA SRS (Lee Badger)