RSA Conference2016
San Francisco | February 29 – March 4 | Moscone Center

Connect to Protect

SESSION ID: AIR-W05

Rise of the Hacking Machines

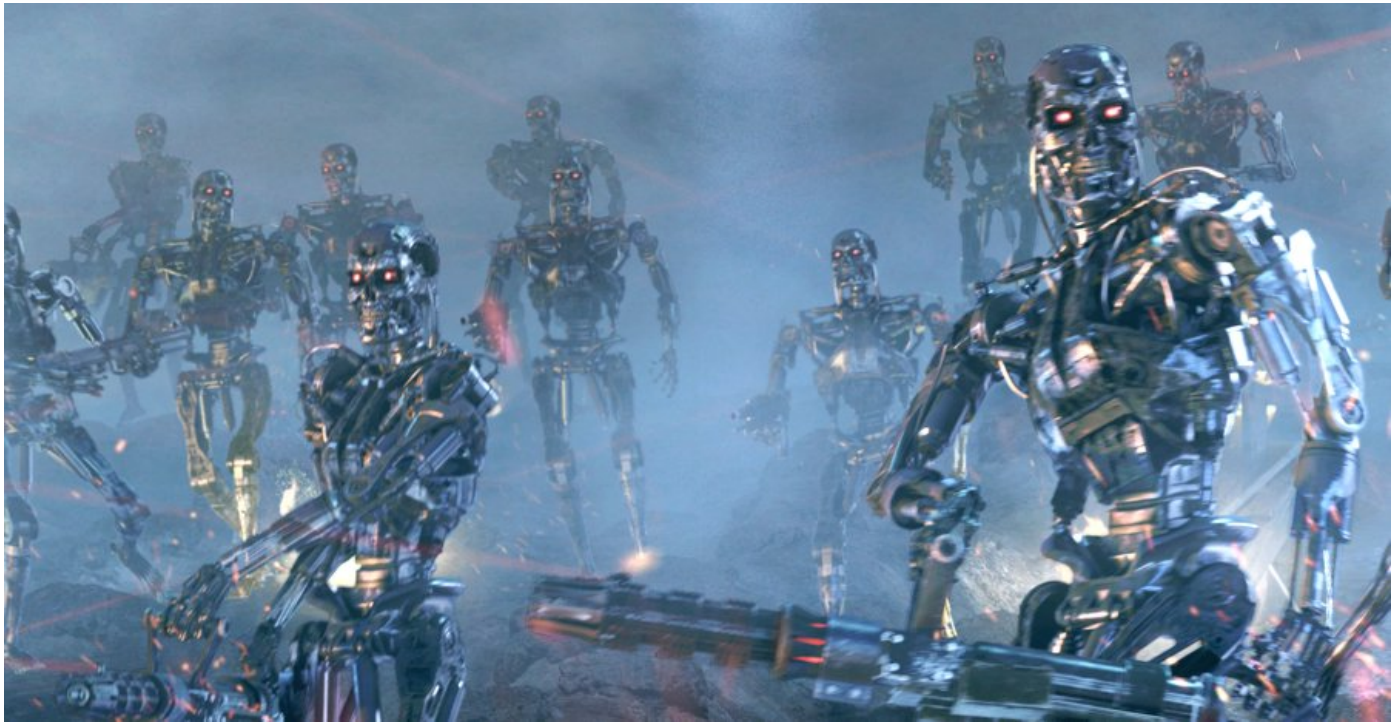
Konstantinos Karagiannis
Chief Technology Officer
Security Consulting
BT Americas
@konstanthacker

#RSAC

The question…

Will AI allow machines to wreak hacking devastation worldwide?

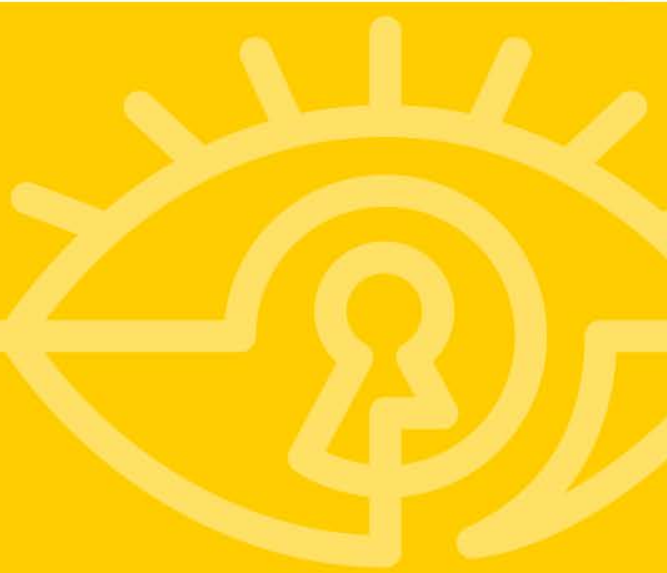
Probably won't be this bad



We missed the original Skynet date of April 19, 2011 at least.

Quick note on “hacking machines”

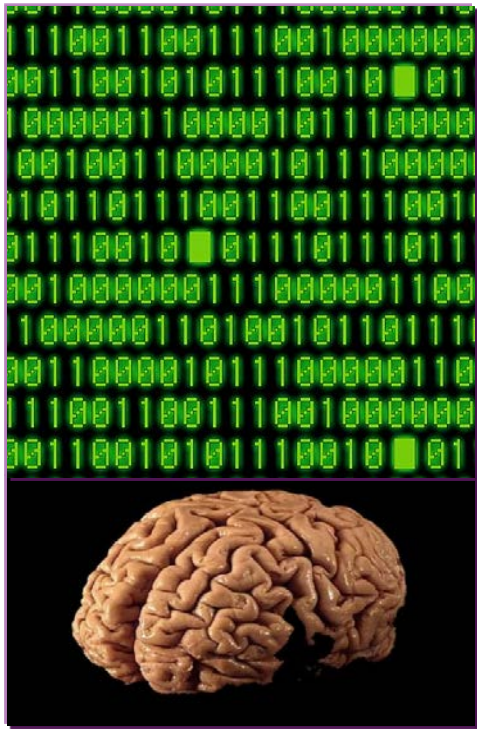
- This futurism talk will look at HW/SW systems that may soon emulate human hackers
- We will not be focusing on:
  - Botnets/DDoS
  - Malware
  - APTs
- This talk is really “*near* futurism”: the machines covered **currently exist** in some form!



The basics of AI

AI compared to wetware

- "The goal of AI is to develop machines that behave *as though* they were intelligent." – John McCarthy, 1956
- A more timeless definition: "Artificial Intelligence is the study of how to make computers do things at which, at the moment, people are better." – Elaine Rich, 1983
- Machines win at computations and data slicing
- We win at the majority of both simple and advanced tasks, and are creative



Narrow and general types of AI

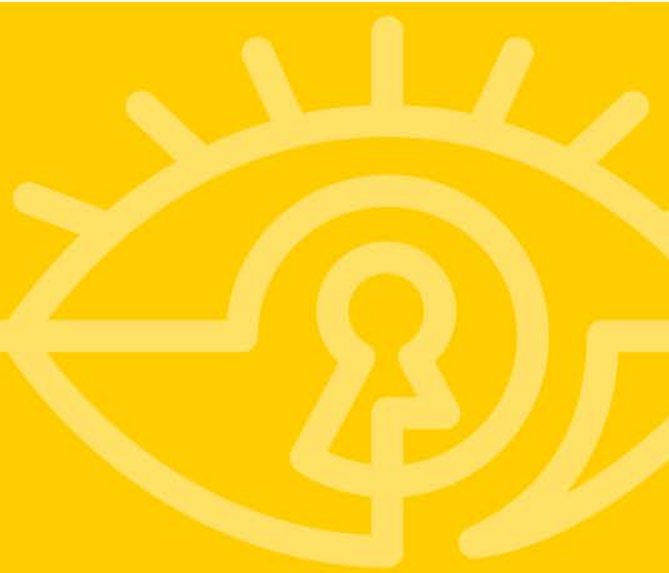
- Narrow (or “weak”) AI:
  - Non sentient
  - Focused on one task
- General (or “strong”) AI:
  - Sentient
  - Can apply intelligence to any task

The history of AI

- **1930s** Alonzo Church and Alan Turing show limits of traditional first-order logic
- **1940s** McCulloch and Pitts design neural networks ahead of necessary computing power
- **1950s** **Turing creates famous Test; Minsky simulates 40 neurons with 3000-tube neural network**; Samuel (IBM) creates learning chess program; AI term introduced; McCarthy (MIT) invents LISP language and self-modifying programs
- **1960s** General Problem Solver imitates human thought; McCarthy founds AI Lab at Stanford; Eliza converses in natural language

The history of AI (continued)

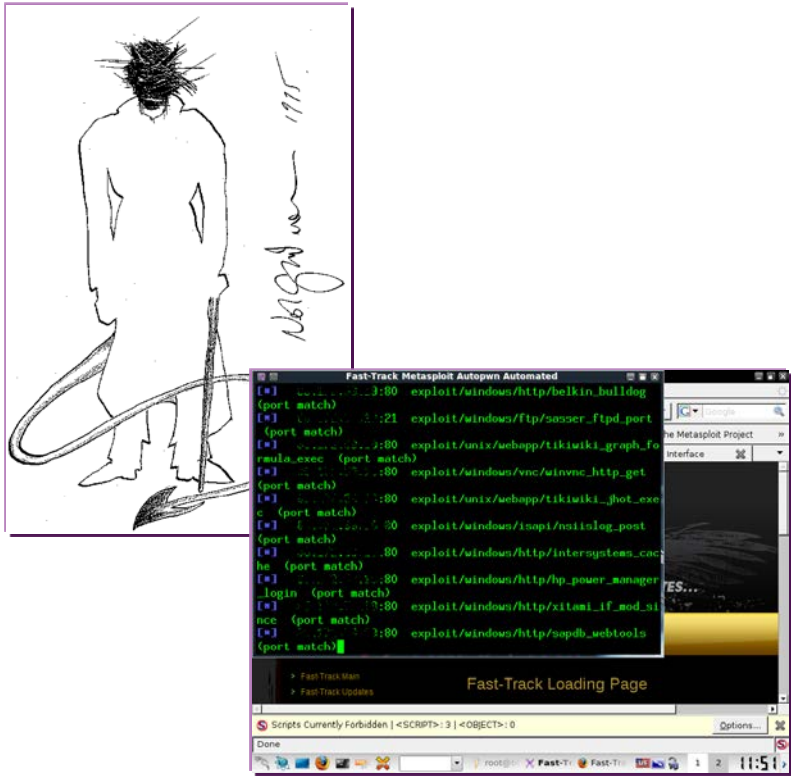
- **1970s** Alain Colmerauer invents logic programming language PROLOG; systems for diagnosing acute abdominal pain and infectious diseases developed
- **1980s** Japanese build PROLOG “Fifth Generation Project”; DEC saves millions a year with R1 system that configures computers; neural network research continues, with Nettalk learning to read text aloud
- **1990s** Bayesian Networks are born; *RoboCup* initiative to build soccer-playing autonomous robots; **IBM’s Deep Blue defeats the chess champ Gary Kasparov**
- **2000s** Service robotics; Japan makes lifelike movements
- **2011 Watson beats Jennings and Rutter on Jeopardy**
- **2016 Machines will face off in a special Defcon Capture the Flag!**



Historical scanner weaknesses

Security scanning without much AI

- 1995—Security Administrator Tool for Analyzing Networks (SATAN) detected network problems by “fingerprints,” relying on banners and basic checks
- Nessus appeared 1998, eventually taking over with NASL security checks
- Metasploit, Immunity Canvas, and Core Impact point-and-click “hacking”
- Web app scanners try to interpret custom-written applications without the real ability to think—generous to call it “narrow AI”



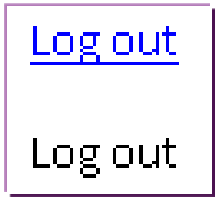
Web Scanner Design Flaws

Authentication handling

Poor spidering/site coverage

Weak support for newer tech such as Web 2.0

Inability to complete multistep functions



Misunderstanding of functions and statements

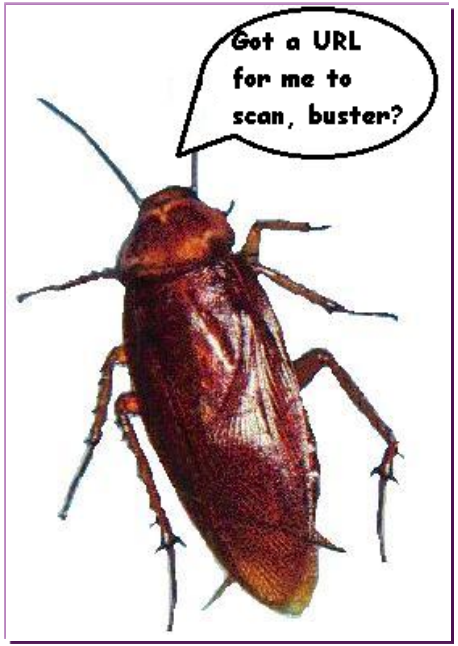
```
<script>alert('xss')</script>

' OR 1=1--
```

Use of mostly simplistic attack vectors

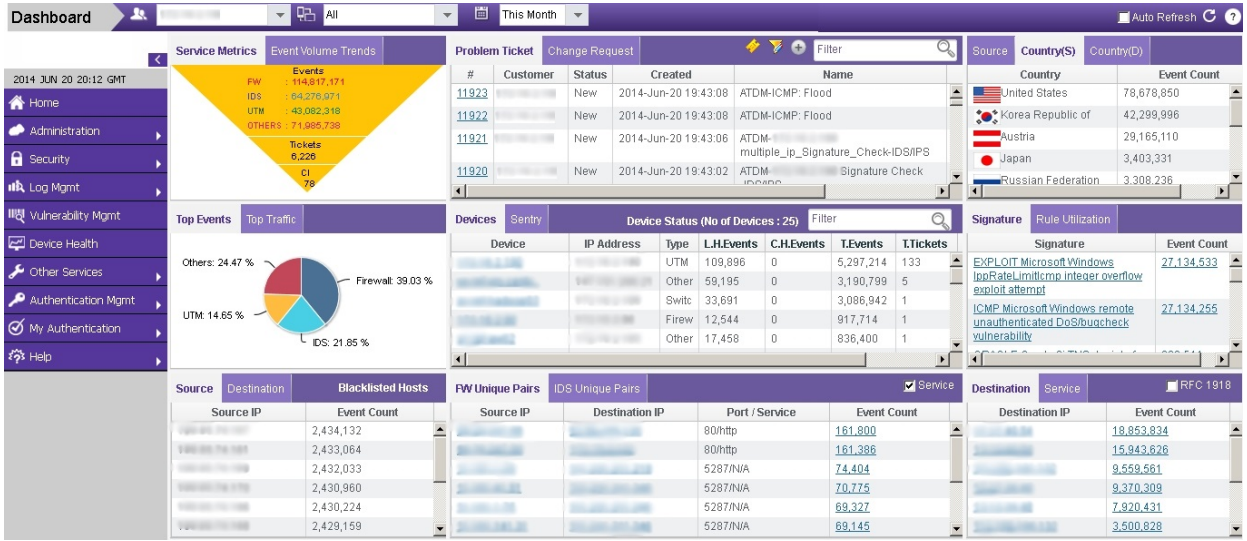
Roaches scanning your apps?

- Web app scanners like collision-detection (cockroach intelligence) robot programs
- Can spot glaring obstacles; can't determine if obstacle can be turned into entry point
- Deep analysis by Watson-like systems will allow future web scanners to fare better
- Much as Watson spots key words in complicated questions, future scanners may spot patterns and logic flows in applications



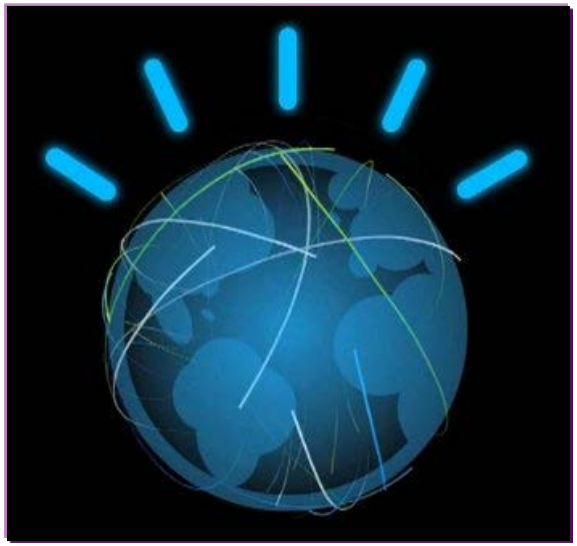
Threat monitoring

- Threat monitoring can spot automated attacks and even human hacking
- Systems can learn how a network typically behaves, what kind of traffic is expected
- Anomalous events gathered by devices can be analyzed via dashboard



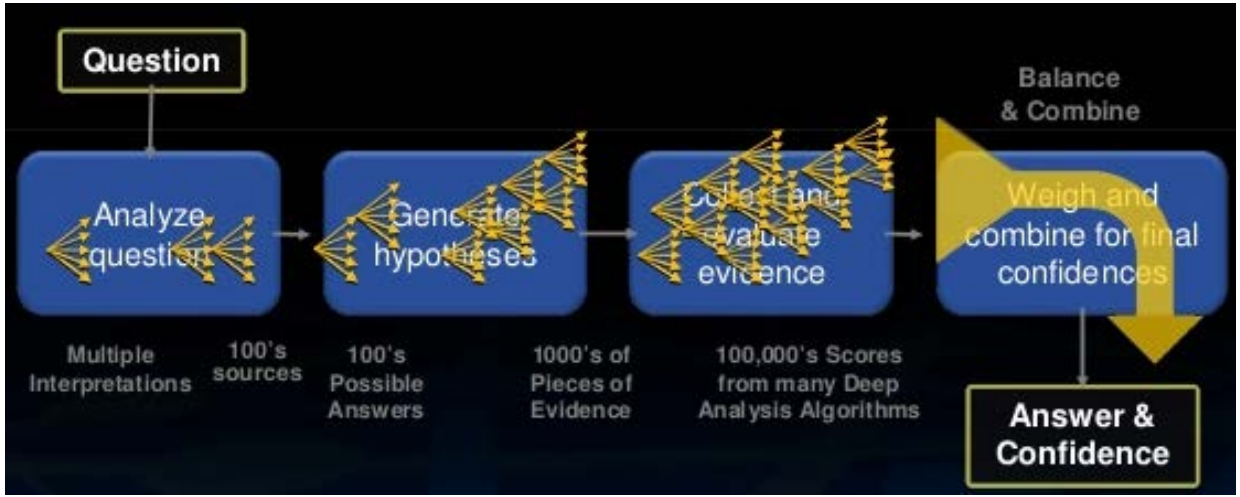
IBM's Watson and new ways of "thinking"

Watson beats the Jeopardy! champs



- Feb 14-16, 2011 we witnessed on TV a seeming evolutionary leap in AI
- Watson beat all-time leading money winner, Brad Rutter, and longest streak holder, Ken Jennings
- Critics irked by Watson's fast trigger "finger" can't dispute the demonstrated ability to dissect complex language in clues

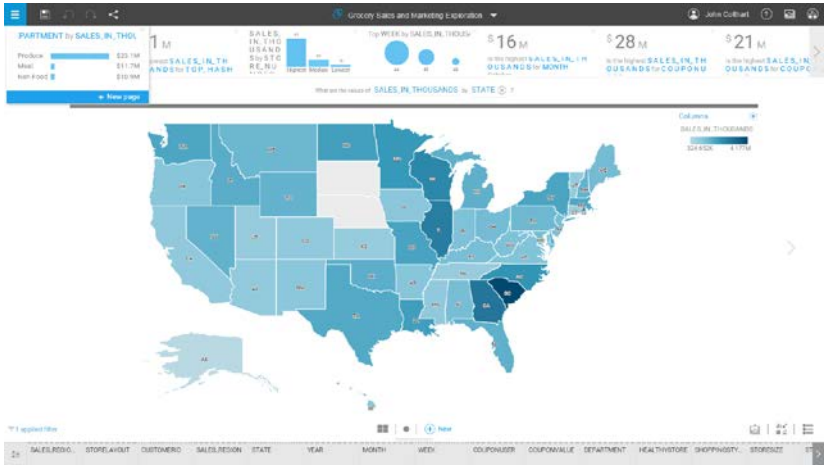
How Watson works

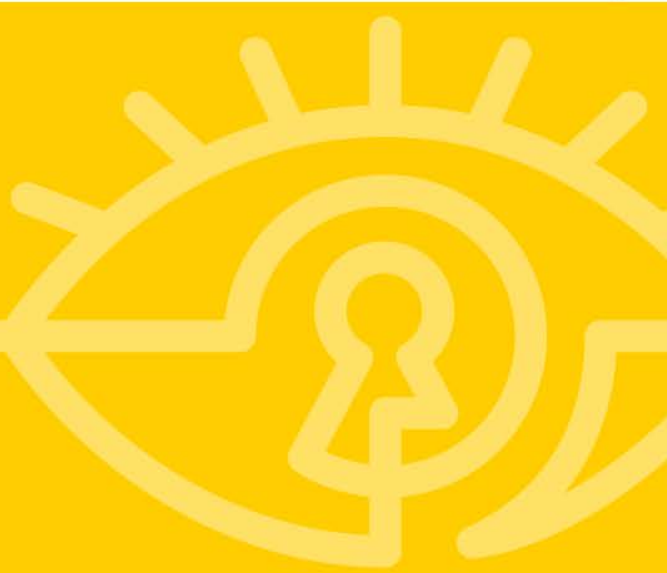


- 90 IBM Power 750 servers, each with a 3.5-GHz POWER7 eight-core processor. 16 TB of RAM total. Massively parallel processing.
- IBM DeepQA software, on Linux, processes 500 GB of data (a million books) per second. 100 techniques analyze natural language, identify sources, generate hypotheses, score evidence, and merge and rank answers.
- Watson uses multiple general info sources, but can use specialized ones…

Watson goes to work—inspires clones

- Watson adapted to cancer treatment decision making at Memorial Sloane-Kettering
- Horsepower available as IBM Watson Analytics
- DARPA DeepDive, OpenCog and other open source technologies arrive
- Can Watson or a descendant be dedicated to, say, computer hacking?

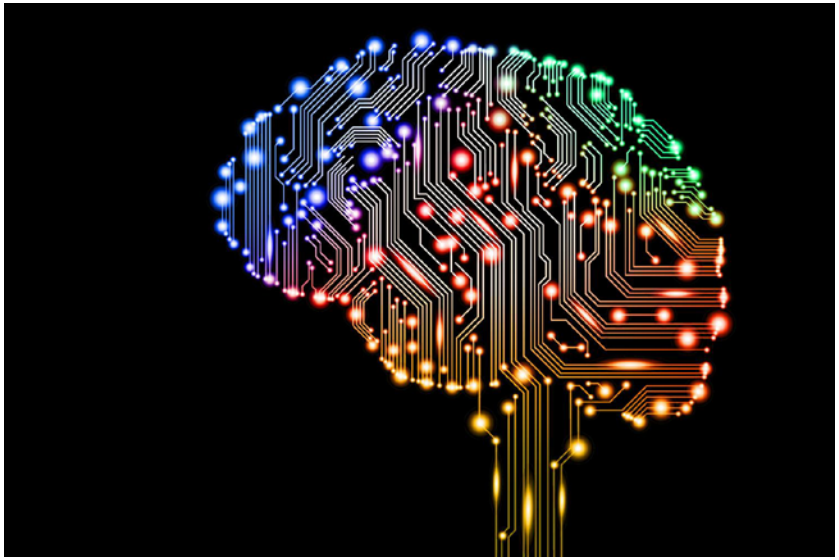




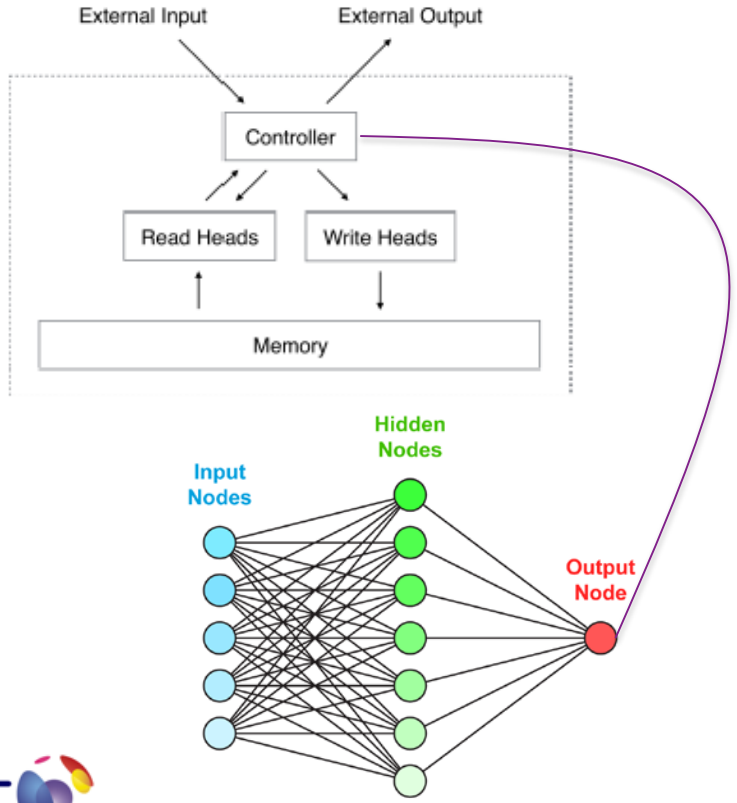
DeepMind and neural networks

DeepMind attacking your apps?

- Google bought DeepMind for $400 million
- Company had been working on reducing intelligence to an algorithm
- Used neural networks to simulate short term memory, and get better at video games (Atari to Doom)—will play Go against human grandmaster Mar. 9!
- Neural networks “learn”/reinforce success
- Imagine a security scanner that could use such techniques to identify intricate flaws



Neural networks



- Often called Artificial Neural Networks (ANNs)
- Input nodes gather data, hidden nodes apply a function and weight results, and output node activates
- Neural Turning Machines like DeepMind couple this with massive external memory
- Even modest hardware can impress—check out Seth Bling's MarI/O

About that Go game next week

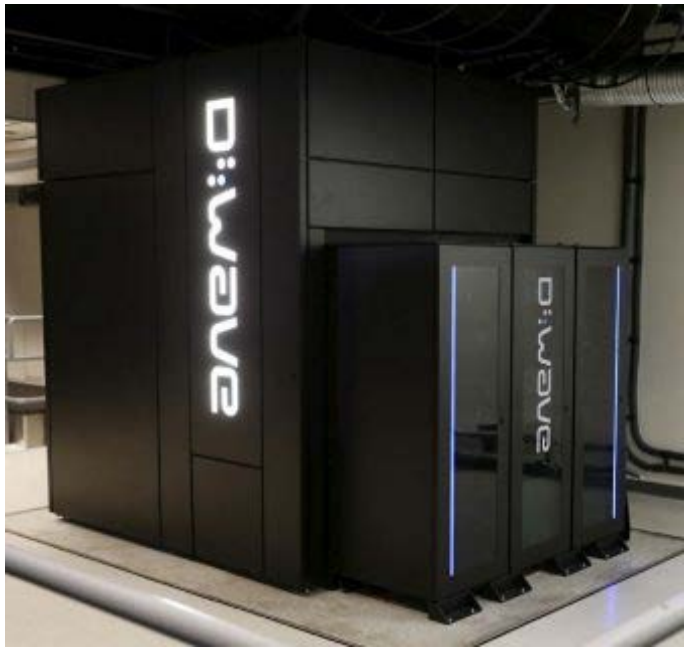
- Lee Sedol will play 5 games against DeepMind's AlphaGo from Mar 9-15 for $1 million
- To consider how powerful neural nets can be:
  - Checkers has 10^{20} possible positions
  - Chess has 10^{120}
  - Go has 10^{761} positions! (on 19x19 board)
  - Only 10^{80} particles in observable universe
- Clearly evolving from Atari to Go shows that we are reaching human-like ability to strategize, perhaps out-think defenses…

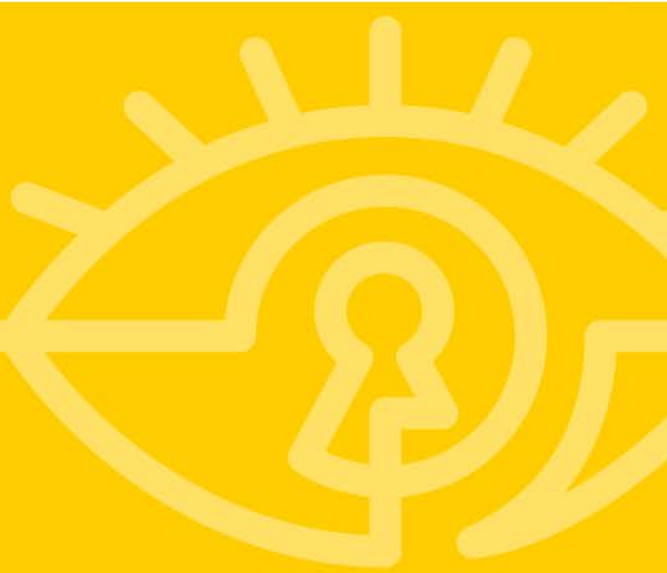
Toys and fears of the wealthy

- Facebook made big news with its Big Sur system relying on GPUs—which excel at neural net processing
- Company also unleashed M, an AI that learns from its pool of a billion users
- Billionaires love AI sometimes, and Amazon also joined in
- Some, like Tesla's Elon Musk fear the end is near with AI—he's not alone (notably Hawking)

Optimization going … quantum?

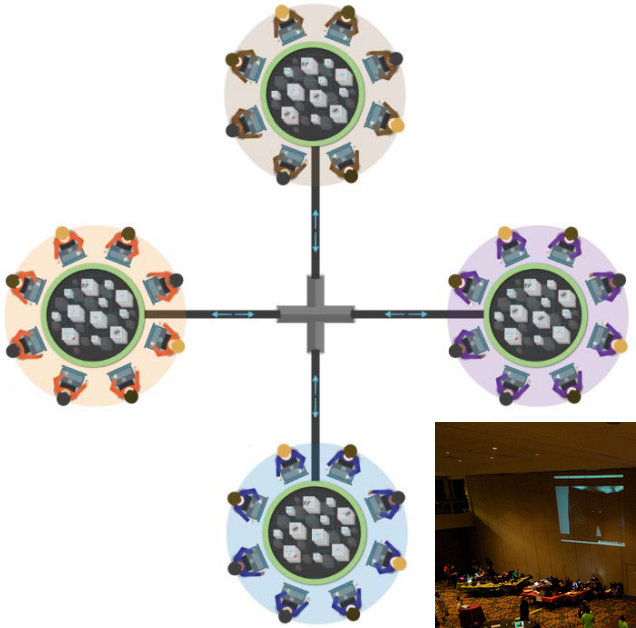
- D-Wave's quantum computer has taken a lot of heat for not being "universal"—can't run legendary algorithms like Shor's or Grover's
- Can't crack encryption!
- Its optimization has been called 100s of millions of times faster than traditional computing gear
- Google promises a strong boost to AI





DARPA's Cyber Grand Challenge

Traditional capture the flag



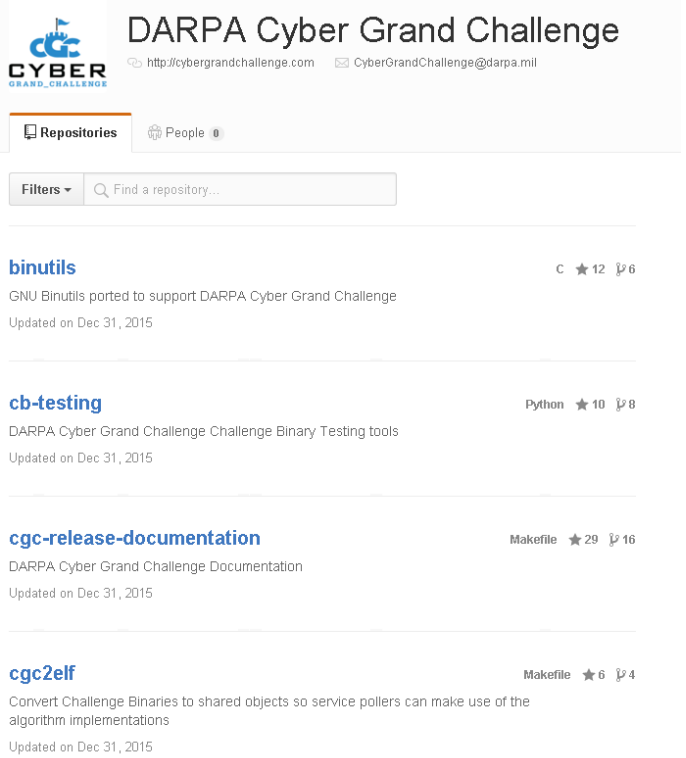
- Popular event at Defcon where teams of hackers:
  - find flaws in new software
  - protect against other teams
- Nessus and other scanners useless as this is NEW software

True hacking machines

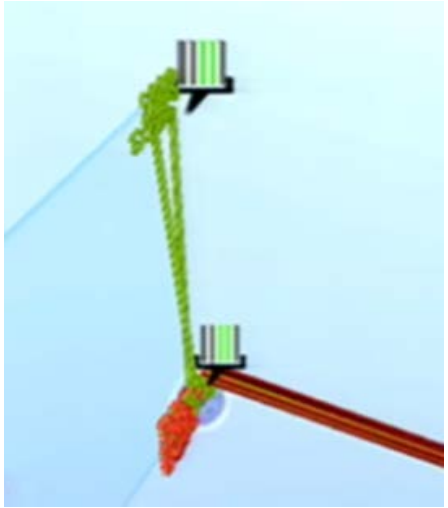
- Cyber Grand Challenge (Aug 4, 2016, Las Vegas Paris Hotel)
- DARPA event where teams will have their machine creations play Capture the Flag
- Running DECREE (DARPA Experimental Cyber Research Evaluation Environment), fully automated systems will, in real time:
  - reverse engineer unknown software
  - locate and heal weaknesses in software

Important notes on DECREE

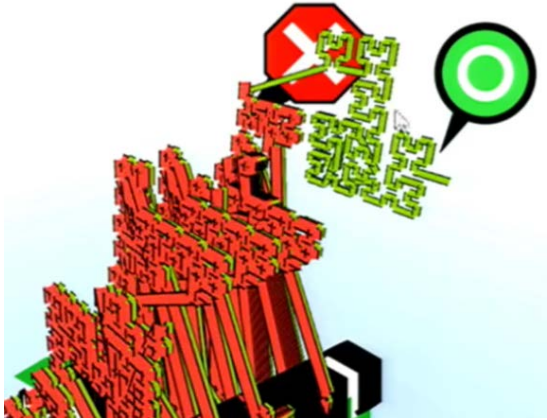
- Built on Linux, but a limited extension
- Linux has hundreds of system calls, but DECREE has seven
- DECREE has its own binary/executable format—not compatible with other software
- Ideal for research only … for now

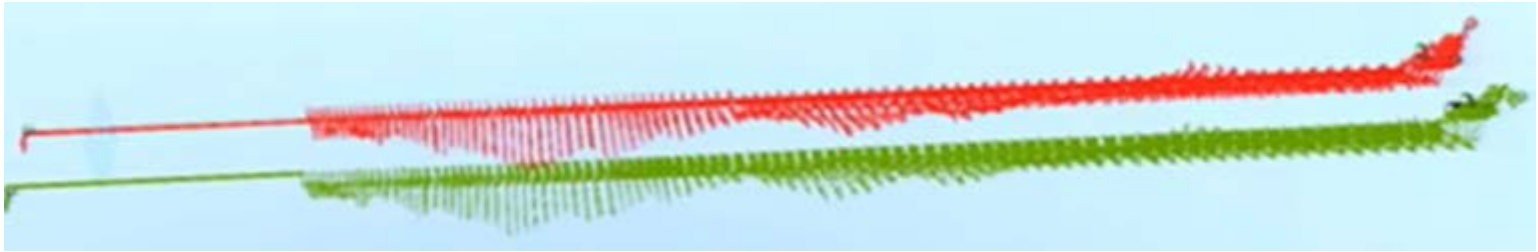


Visualizing flaws and fixes

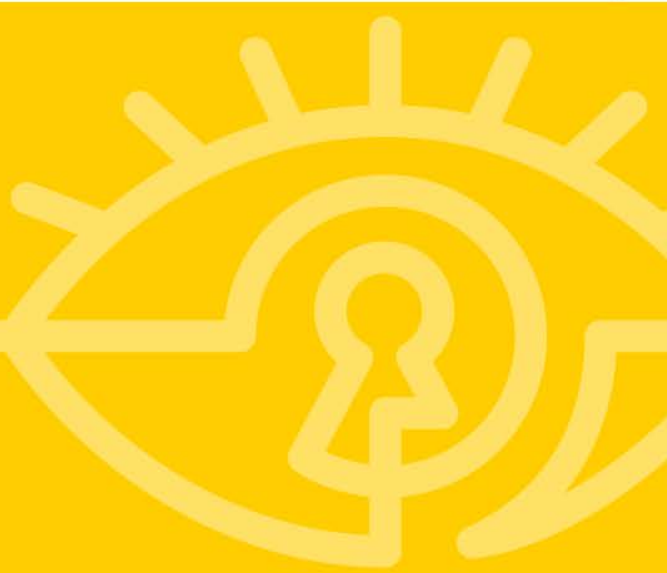


- Elegant, patch—initial “jump” to error handle code, then final non-crash output.
- Made by a machine!





Not always elegant…



Looking ahead…

Look closely at the DARPA requirements

- A scary list of autonomous, machine-only “hacking” and defense actions
- Due to setup and environment, machines can’t be unleashed immediately
- Still, this list covers almost all our jobs, right?

Autonomous Analysis: The automated comprehension of computer software (e.g., CBs) provided through a Competition Framework.

Autonomous Patching: The automatic patching of security flaws in CBs provided through a Competition Framework.

Autonomous Vulnerability Scanning: The ability to construct input which when transmitted over a network provides proof of the existence of flaws in CBs operated by competitors. These inputs shall be regarded as Proofs of Vulnerability.

Autonomous Service Resiliency: The ability to maintain the availability and intended function of CBs provided through a Competition Framework.

Autonomous Network Defense: The ability to discover and mitigate security flaws in CBs from the vantage point of a network security device.

Within this year

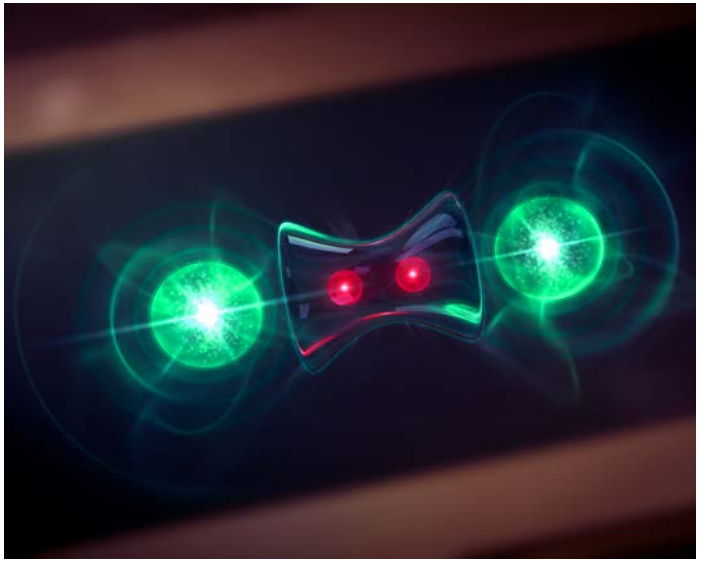
- Expect threat monitoring and analytics to get a little better—smart networks a MUST
- Automated scanning may stagnate
- DECREE likely to catch fire in open source community—new, more flexible development coming?
- Facebook, Microsoft, and Google have open sourced some AI

Within 5 years

- Expect neural networks similar to DeepMind to be probing applications and services/ports
- Commercial automated scanning should improve—may have difficulty keeping up against constant barrage
- Optimizing (not universal) quantum computers should have impact on coding
- Monitoring will be the wild card
  - can its narrow AI evolve fast enough?
  - can self-repairing networks (like in CGC contest) take off?

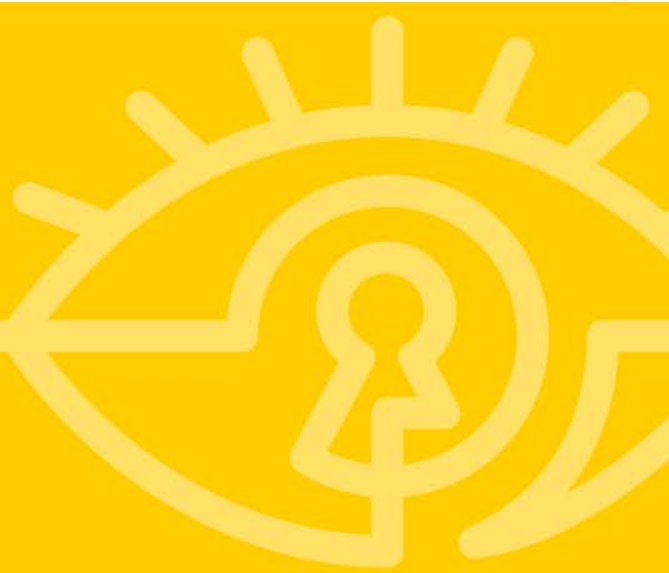
Within 10 years

- Singularity date may change—still 2045
- Human hackers/defenders unlikely to be able to keep up with “bulk” work by this point—we’ll have to be useful for creativity!
- Universal quantum computers?
  - Australia claims within 3 years actually
  - Some government may get there first
  - Within 10 years PK encryption useless



Preparing for Tomorrow

- Even if it remains narrow AI for a couple decades, the technology is here to make machines that can hack fast and effectively
- With the exception of quantum computers, the hardware to run advanced AI will not be cost prohibitive to criminals or even the curious
- Dumb networks not an option currently—bleeding edge smart networks will increasingly be mandatory. Too easy to fall behind curve in coming years



Questions?

Konstantinos.Karagiannis@bt.com
@konstanthacker