Crashing Drones and Hijacked Cameras: CyberPhysical meets CyberTrust

Jeannette M. Wing
Corporate Vice President
Microsoft Research

Annual Computer Security and Applications Conference
Los Angeles, CA
9 December 2015
What is Common?

They have a computational core that interacts with the physical world.

Cyber-physical systems are engineered systems that require tight conjoining of and coordination between the computational (discrete) and the physical (continuous).

Trends for the future

• Cyber-physical systems will be smarter and smarter.
• More and more intelligence will be in software.
• More and more connectivity and data flow.
What could go wrong?
Trustworthiness in Cyber-Physical Systems

Reliability  Security  Privacy

Hardware, Software, People
Reliability Challenges
Challenge 1: Reasoning about Continuous and Discrete

Challenge 2: Uncertainty in Environment

Challenge 3: Sensors and Actuators Can Fail
Computable Reals: 
A Fundamentally Hard Problem

“A real number is computable if its digit sequence can be produced by some algorithm or Turing machine. The algorithm takes an integer \( n > 1 \) as input and produces the \( n \)-th digit of the real number's decimal expansion as output. “ [Turing 1936]

Fact: While the set of real numbers is uncountable, the set of computable numbers is only countable and thus almost all real numbers are not computable.
On the one hand:
A real number $a$ is said to be **computable** if it can be approximated by some computable function in the following manner: given any integer $n \geq 1$, the function produces an integer $k$ such that:

\[
\frac{k-1}{n} \leq a \leq \frac{k+1}{n}
\]

On the other:
The computable numbers include many of the specific real numbers which appear in practice, including all real algebraic numbers, as well as $e$, $\pi$, and many other transcendental numbers.
Uncertainty at Multiple Levels

Secure OS

Robust Sensing

Correct Control

High-level Planning

Sensor noise, and complex missions

Safe despite limited power, external disturbances,
System = (State + Control) \| Environment

Limited Battery Power: Typically less than 20 minutes
Not enough computational power on board
Not very robust to changes in environment or disturbances e.g., wind, obstacles
Not very robust to changes in system properties, weight, aging of rotors, etc.

\[
\text{State} = [x, y, z, \psi, \phi, \rho, \dot{x}, \dot{y}, \dot{z}, \dot{\psi}, \dot{\phi}, \dot{\rho}, d_1, d_2, d_3, d_4]
\]

Control = RPM of the motors
Safe Control Under Uncertainty

\[ \dot{x}(t) = Ax(t) + Bu(t) \]

**Optimal control:** minimize cost on deviation from reference + cost on control.

\[ \min_u \sum_{t=1}^{T} (x_t - \hat{x}_t)^\top Q(x_t - \hat{x}_t) + u_t^\top Ru_t \]

**Subject to:** We are safe!

\[ (x_1, u_1, \ldots, x_T, u_T) \models \phi \]
Security Challenges
FBI warns of Internet of Things risks
What could go wrong? The bureau's Internet Crime Complaint Center lays out a laundry list of horrors.

Skateboards, drones and your brain: everything got hacked
At Defcon in Las Vegas, hackers gather to show off the latest vulnerabilities. That's why last weekend was just full of bad news.

Security will be critical to the success or failure of Internet of Things

Man Hacks Monitor, Screams at Baby

Stuxnet: A wake-up call for nuclear cyber security

Flying hacker contraption hunts other drones, turns them into zombies
Secure configuration
Security protocols and encryption
Secure storage
Secure boot
Device identity in hardware
Device identity in hardware
Secure configuration
- Security protocols and encryption
- Secure storage
- Secure boot
- Device identity in hardware
Secure configuration

Security protocols and encryption

Secure storage

Secure boot

Device identity in hardware
Secure configuration

Security protocols and encryption

Secure storage

Secure boot

Device identity in hardware
Secure configuration
Security protocols and encryption
Secure storage
Secure boot
Device identity in hardware
Built for low power and limited computing resources

- Secure configuration
- Security protocols and encryption
- Secure storage
- Secure boot
- Device identity in hardware
Privacy Challenges
FBI warns of Internet of Things risks
What could go wrong? The bureau's Internet Crime Complaint Center lays out a laundry list of horrors.

Skateboards, drones and your brain: everything got hacked
At Defcon in Las Vegas, hackers gather to show off the latest vulnerabilities. That's why last weekend was just full of bad news.

Man Hacks Monitor, Screams at Baby

Security will be critical to the success or failure of Internet of Things

Flying hacker contraption hunts other drones, turns them into zombies

Stuxnet: A wake-up call for nuclear cyber security
Privacy is about...

... the appropriate collection and processing of information about a data subject by a data holder and the flow of information between data holders.
Data Lifecycle

Anonymized Data → Collection → Data Management → Analyze

Policy Compliance?

Insights

Collaborative Learning

Share with Partners

Privacy Preservation Risks
Data Lifecycle and Privacy Preservation

- Differential Privacy
- Data Map
- New Encryption Protocols

Data → Collection → Data Management

- Analyze
- Share with Partners

Insights → Collaborative Learning

Privacy Preservation Policies and TECHNOLOGIES
Anonymization is Insufficient

- “Anonymized” data combined across data sources can identify individuals
  - Netflix users have been re-identified based on ‘anonymous’ viewing habits
  - Mass. Governor Weld (and many others) were re-identifiable based on ‘anonymous’ medical records
  - Credit Card metadata and aggregate cell phone data have fallen to re-identification attacks.

- The President’s Council of Advisors on Science and Technology 2014 Big Data report

  “Anonymization is increasingly by the very techniques that are being developed for many legitimate applications of big data. In general easily defeated, as the size and diversity of available data grows, the likelihood of being able to re-identify individuals (that is, re-associate their records with their names) grows substantially. While anonymization may remain somewhat useful as an added safeguard in some situations, approaches that deem it, by itself, a sufficient safeguard need updating”.

What Differential Privacy Is

- Technique that enables learning about populations or population segments, while preserving the privacy of individuals.

- With DP the same query outputs will be observed with essentially the same probabilities, even if an individual record is added or deleted from the database.

- Privacy is achieved by adding noise *either* to data prior to collection or to the results of queries against pristine databases.
The Privacy Compliance Challenge

Legal Team
Crafts Policy

Privacy Champion
Interprets Policy

Developer
Writes Code

Audit Team
Verifies Compliance

Specification
Verification
Compliance?

Meetings

English
Privacy Policy

Millions of Lines of Code
Data Collection
And Management
A Streamlined Audit Workflow

Legal Team
Crafts Policy

Private
Interprets Policy

Developer
Writes Code

Audit Team
Verifies Compliance

Encode
Refine

Legalease
A formal policy specification language

Grok
Data inventory with policy labels

Annotated Code
Legalease Policy
Potential violations

Checker

Fix Code
Update Grok

Developer annotations
Code analysis
ALLOW
EXCEPT
DENY DataType IPAddress:Expired
DENY DataType UniqueIdentifier:Expired
DENY DataType SearchQuery, PII InStore Store
DENY DataType UniqueIdentifier, PII InStore Store

DENY DataType BBEPData UseForPurpose Advertising

DENY DataType BBEPData, PII InStore Store

DENY DataType BBEPData:Expired
DENY DataType UserProfile, PII InStore Store

DENY DataType PII UseForPurpose Advertising
DENY DataType PII InStore AdStore

DENY DataType SearchQuery UseForPurpose Sharing

ALLOW DataType SearchQuery:Scrubbed

- “we remove the entirety of the IP address after 6 months”
- “[we remove] cookies and other cross session identifiers, after 18 months”
- “we store search terms (and the cookie IDs associated with search terms) separately from any account information that directly identifies the user, such as name, e-mail address, or phone numbers.”
- “we do not use any of the information collected through the Bing Bar Experience Improvement Program to identify, contact or target advertising to you”
- “we take steps to store [information collected through the Bing Bar Experience Improvement Program] separately from any account information we may have that directly identifies you, such as name, e-mail address, or phone numbers”
- “we delete the information collected through the Bing Bar Experience Program at eighteen months.”
- “we store page views, clicks and search terms used for ad targeting separately from contact information you may have provided or other data that directly identifies you (such as your name, e-mail address, etc.).”
- “our advertising systems do not contain or use any information that can personally and directly identify you (such as your name, email address and phone number).”
- “Before we [share some search query data], we remove all unique identifiers such as IP addresses and cookie IDs from the data.”
Grok

Data Inventory

Annotate code + data with policy data types

Source labels propagated via data flow graph

D. E. Denning. “A lattice model of secure information flow”
How can we build cyber-physical systems that people can **bet their lives on**?
Thank you!