Security in Malicious Environments: NSF Programs in Information-Theoretic Network Security

Phil Regalia

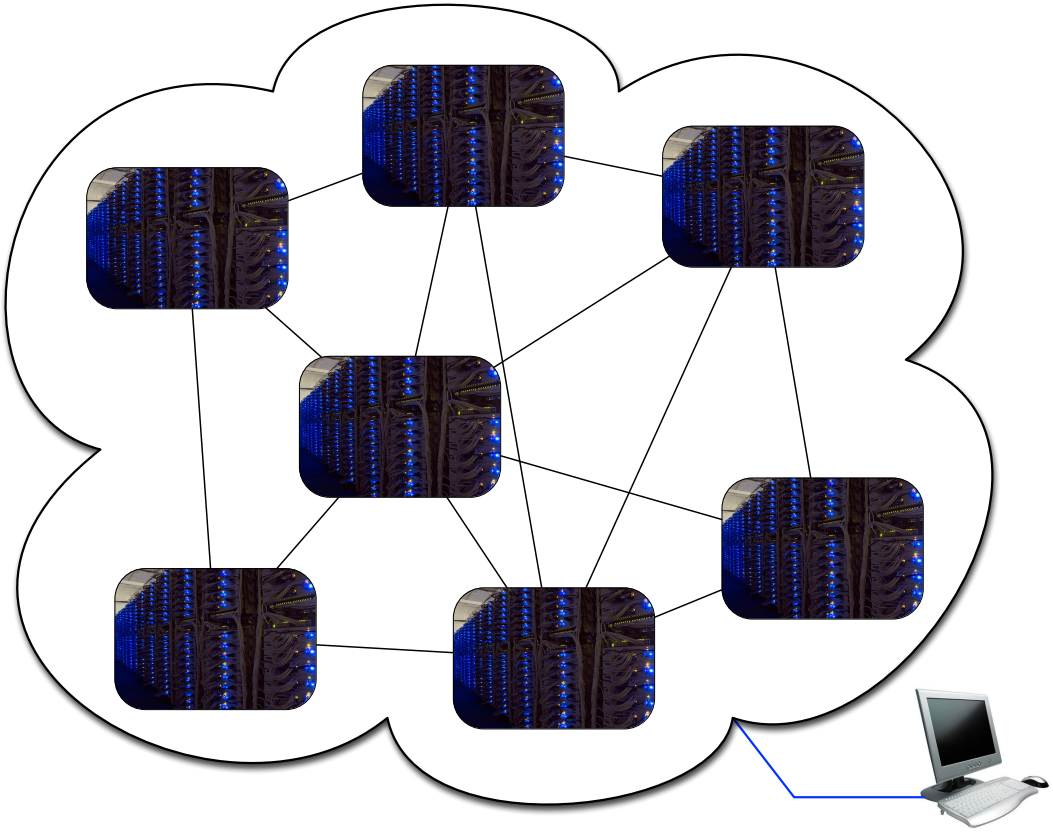
Program Director
Directorate for Computer & Information Science & Engineering
Division of Computing and Communication Foundations
National Science Foundation
Arlington, Virginia 22203
pregalia@nsf.gov

DIMACS Workshop on
Coding-Theoretic Methods for Network Security
1–3 April 2015

Outline

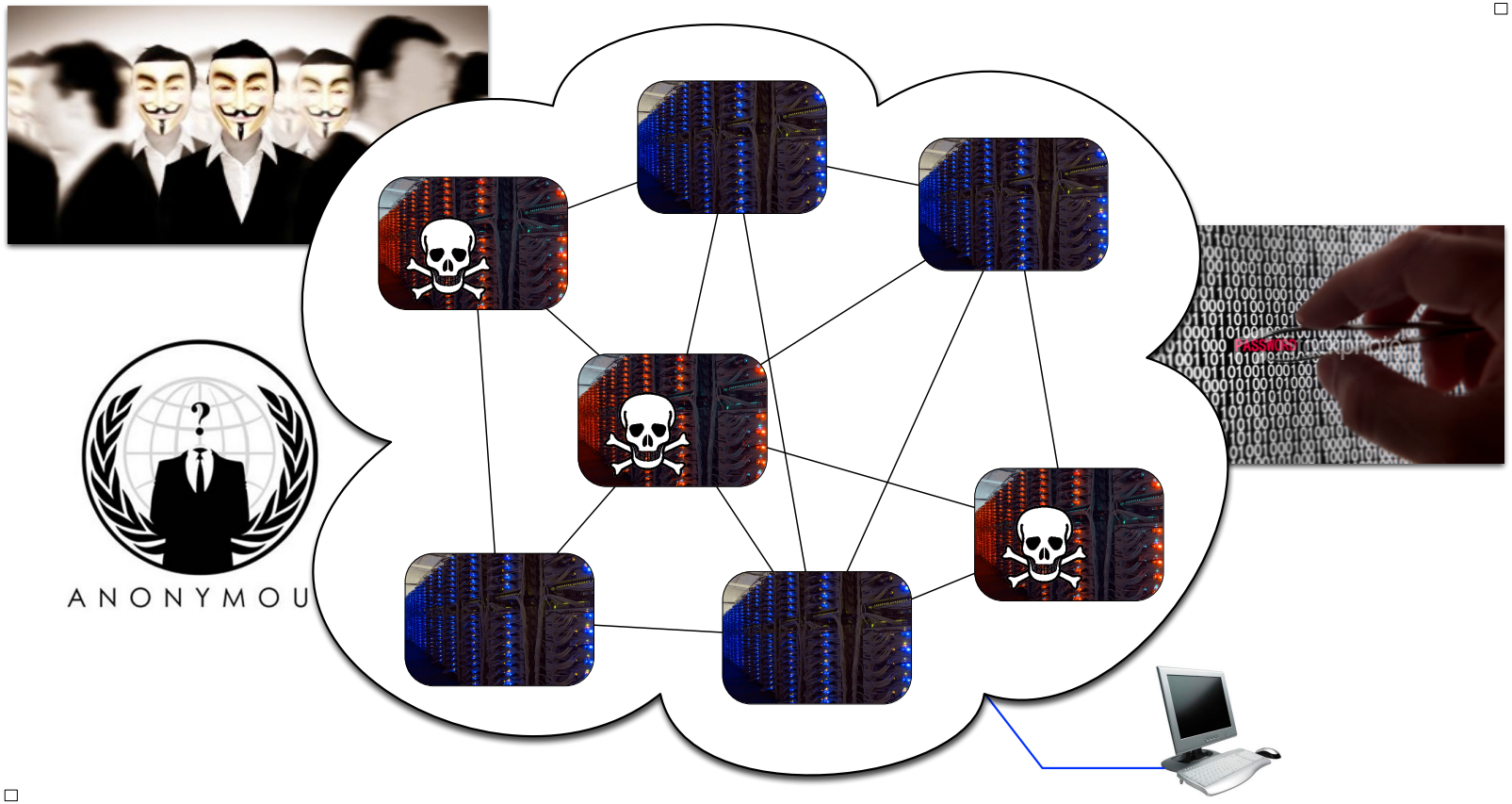
Secure Networking

"Imagine a world seamlessly networked . . ."



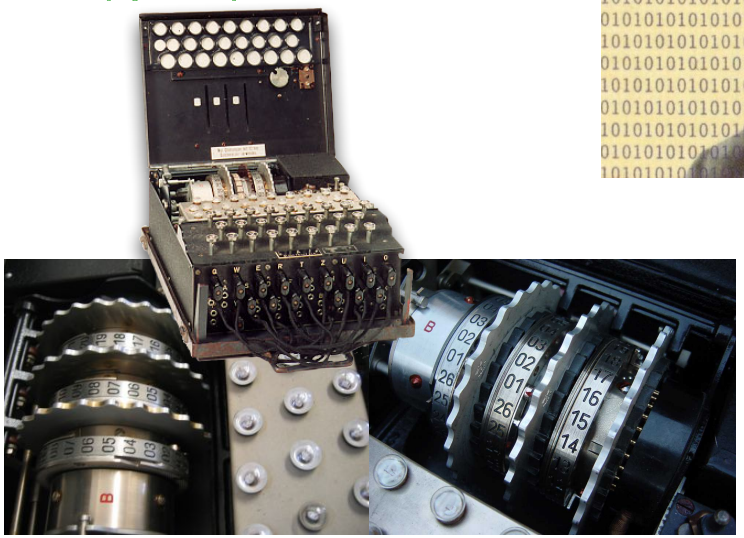
Secure Networking

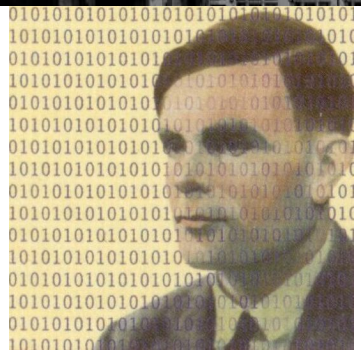
"Imagine a world seamlessly networked . . ." and full of bad guys:



The Glory Days of Cryptanalysis

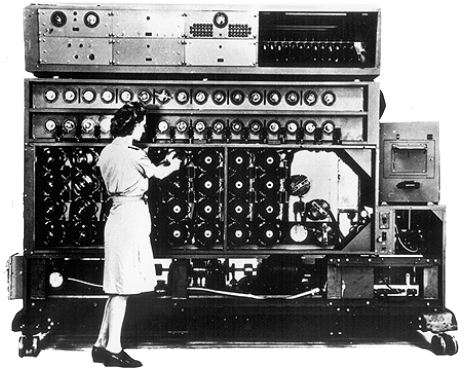
German Enigma cryptosystem

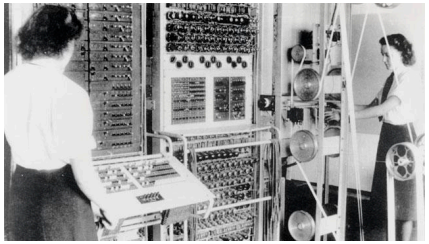




Alan Turing

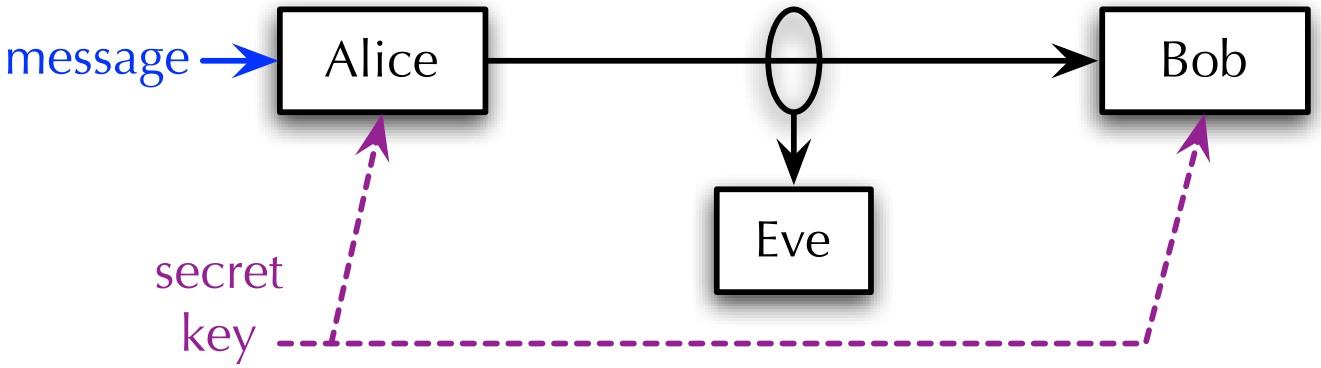
Colossus code breaker





World War II: The German Enigma cryptosystem is broken.

Traditional Secrecy Tool: Cryptography

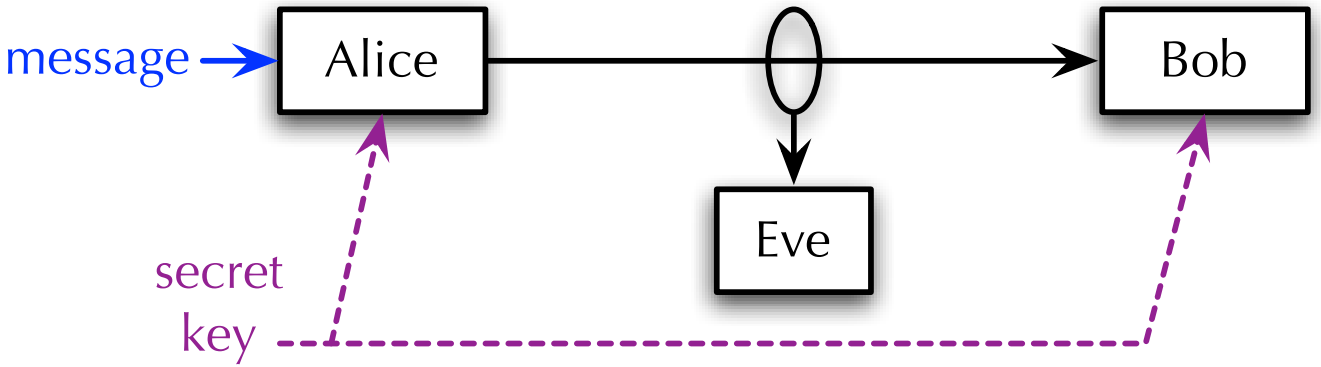


- Symmetric key cryptography (AES): assumes "secure channel" between Alice and Bob to communicate common key.
- Key generation: can use public key cryptography, and/or common randomness, and/or quantum techniques, and/or . . .
- When many Alices and Bobs exist, key management becomes a weak link.

Kerckhoffs's Principle (1883)

A cryptosystem should be secure even if everything about the system, except the key, is public knowledge.

Traditional Secrecy Tool: Cryptography

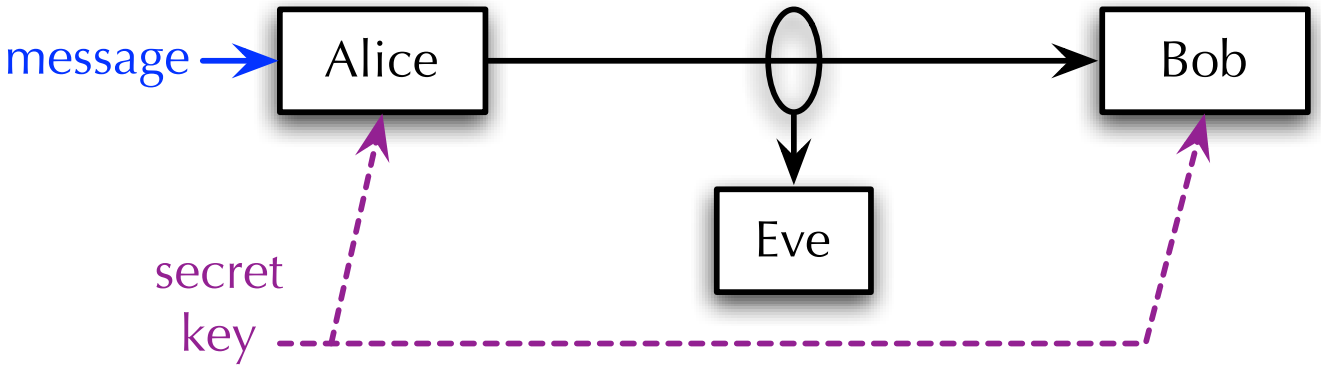


- Symmetric key cryptography (AES): assumes "secure channel" between Alice and Bob to communicate common key.
- Key generation: can use public key cryptography, and/or common randomness, and/or quantum techniques, and/or . . .
- When many Alices and Bobs exist, key management becomes a weak link.

Kerckhoffs's Principle (1883)

A cryptosystem should be secure even if everything about the system, except the key, is public knowledge.

Traditional Secrecy Tool: Cryptography



- Symmetric key cryptography (AES): assumes "secure channel" between Alice and Bob to communicate common key.
- Key generation: can use public key cryptography, and/or common randomness, and/or quantum techniques, and/or . . .
- When many Alices and Bobs exist, key management becomes a weak link.

Kerckhoffs's Principle (1883)

A cryptosystem should be secure even if everything about the system, except the key, is public knowledge.

Fast Forward to the Present

Great advances in cryptography:

Cryptographic message strength has improved steadily (AES and beyond)

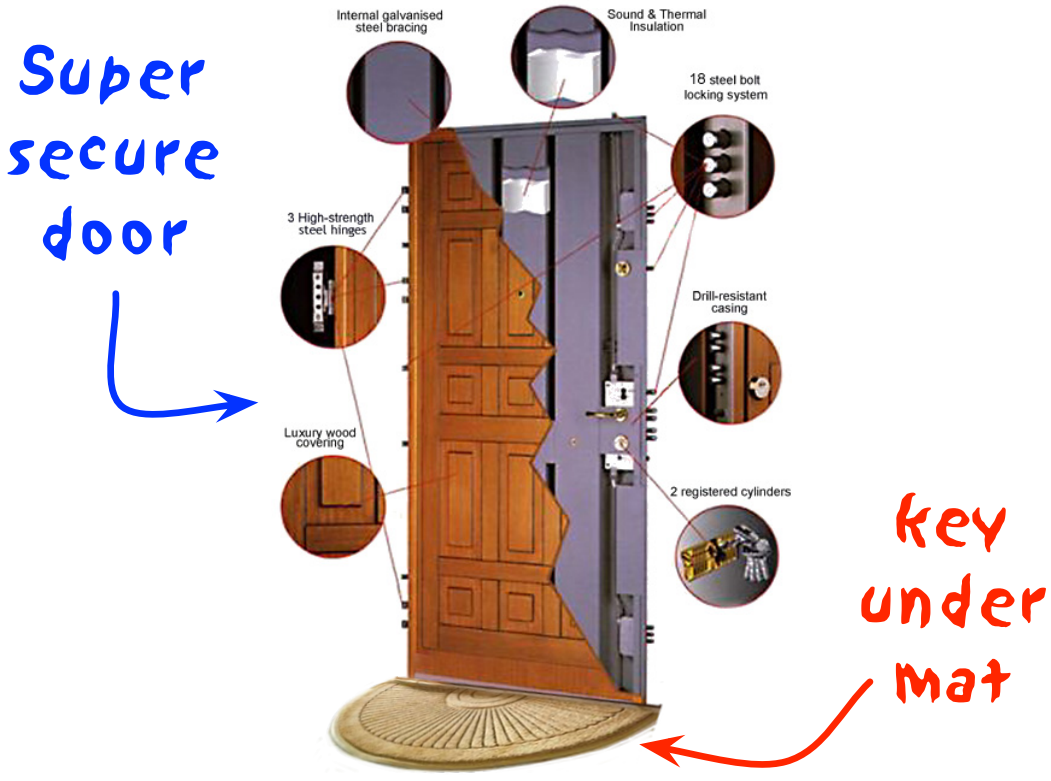
What about the **key**?

Fast Forward to the Present

Great advances in cryptography:

Cryptographic message strength has improved steadily (AES and beyond)

What about the **key**?



Today's World: The Weak Link

The Washington Post

SPECIAL REPORT

ZERO DAY

The Threat in Cyberspace: To succeed in addressing risks in the digital universe, global leaders must understand one of the most complex, man-made creations on Earth: cyberspace. **View the series.**

Hackers exploit humans to gain access

In cyberattacks, hacking humans is highly effective way to access systems

But like much of the digital universe, the e-mails were not what they seemed. They were cyberweapons, part of a devastating kind of attack known as "social engineering."

Emerging details about the e-mails show how social engineering — long favored by con artists, identity thieves and spammers — has become one of the leading threats to government and corporate networks in cyberspace.

The technique involves tricking people to subvert a network's security. It often relies on well-known scams involving e-mail, known as "spear phishing," or phony Web pages. But such ploys now serve as the pointed tips of far more sophisticated efforts by cyberwarriors to penetrate networks and steal military and trade secrets.

"Multiple natural gas pipeline sector organizations have reported either attempted or successful network intrusions related to this campaign," officials at the Department of Homeland Security said in a confidential alert obtained by The Post.

The May 15 alert, by the department's specialists in industrial control systems, said "the number of persons targeted appears to be tightly focused. In addition, the email messages have been convincingly crafted to appear as though they were sent from a trusted member internal to the organization."

Social-engineering attacks revolve around an instant when a computer user decides whether to click on a link, open a document or visit a Web page. But the preparation can take weeks or longer.

Today's cryptography is "strong". **But:**

Security hinges on key distribution: keys are "entrusted" to humans.

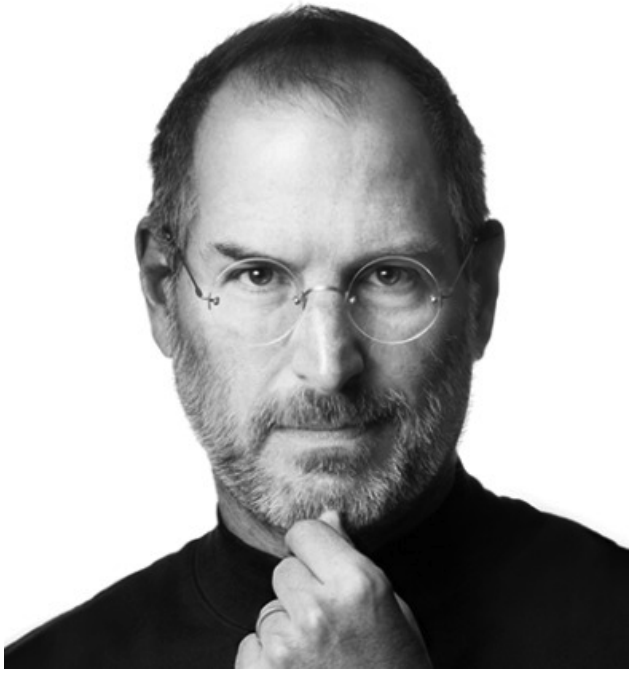
It's much easier to hack humans than to break crypto systems.

http://www.washingtonpost.com/investigations/in-cyberattacks-hacking-humans-is-highly-effective-way-to-access-systems/2012/09/26/2da66866-ddab-11e1-8e43-4a3c4375504a_story.html

Wikileaks

Top secret, classified information

Digital Rights/Restriction Management (DRM)



On DRM keys:

> *"No one has ever implemented a DRM system that does not depend on secret keys for its operation. There are many smart people in the world, who love to discover such secrets and publish them. It's a cat-and-mouse game."*
>
> *—Steve Jobs*

http://web.archive.org/web/20080517114107/http:/www.apple.com/hotnews/thoughtsonmusic

Other examples of leaked keys

- Content Scrambling System (CSS). Designed to impose separate geographic pricing regimes for DVDs.

  ⇒ leaked key gave rise to DeCSS

- Sony Playstation 3:

  ⇒ leaked decryption keys for PSJailBreak and LV0: can now boot "other OS".

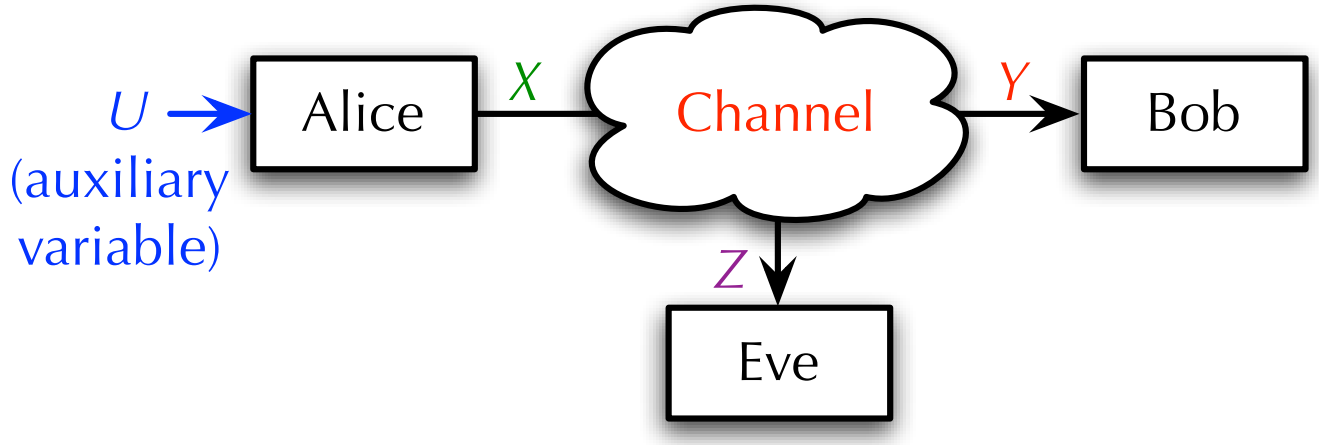
- Many others . . .

http://en.wikipedia.org/wiki/DeCSS
http://www.eurogamer.net/articles/digitalfoundry-ps3-the-final-hack

Keyless Security

Can we secure data & communications without using keys?

Yes, using coding for the wiretap channel:



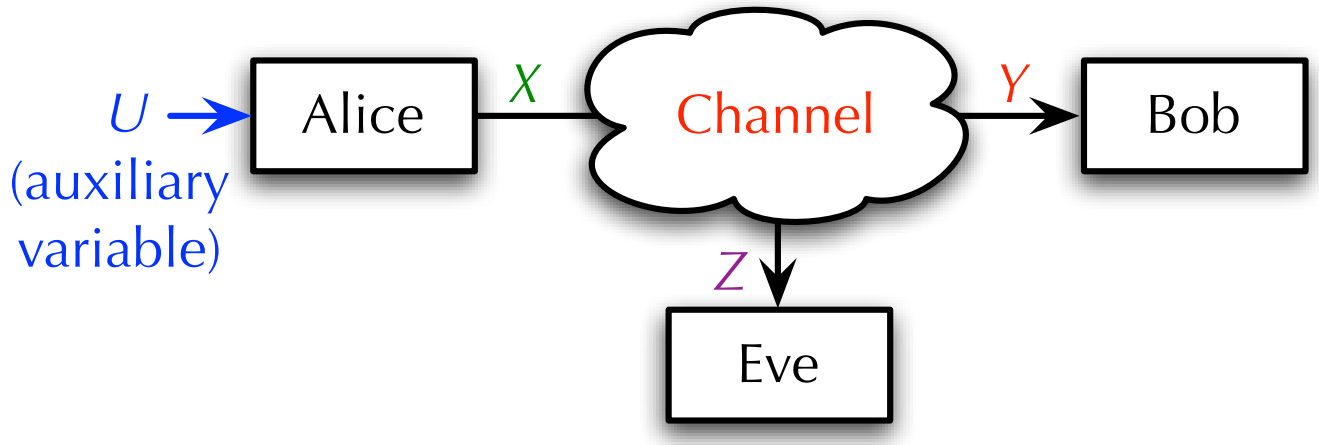
$$\text{Secrecy capacity} = \sup_{U \to X \to (Y,Z)} \Big(I(U,Y) - I(U,Z) \Big)$$
$$\to C_{A \to B} - C_{A \to E}$$

When secrecy capacity is negative, a two-way protocol by Maurer (1993) gives virtual channels, ensuring Eve's is worse than Bob's.

Keyless Security

Can we secure data & communications without using keys?

Yes, using coding for the wiretap channel:



$$\text{Secrecy capacity} = \sup_{U \to X \to (Y,Z)} \Big(I(U,Y) - I(U,Z) \Big)$$
$$\to C_{A \to B} - C_{A \to E}$$

When secrecy capacity is negative, a two-way protocol by Maurer (1993) gives virtual channels, ensuring Eve's is worse than Bob's.

Code design

Message $\mathbf{m}$ determines code word $\mathbf{x}$ according to

$$\begin{bmatrix} \mathbf{0} \\ \mathbf{m} \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_\Delta \end{bmatrix}}_{\mathbf{H}} \mathbf{x}$$

Bob estimates message according to

$$\begin{aligned} \hat{\mathbf{x}} &= \arg\min_{\boldsymbol{\xi}} d(\mathbf{y}, \boldsymbol{\xi}) \qquad \text{subject to } \mathbf{0} = \mathbf{H}_1 \boldsymbol{\xi} \\ &\Rightarrow \quad \hat{\mathbf{m}} = \mathbf{H}_\Delta \hat{\mathbf{x}} \end{aligned}$$

$\mathbf{H}$ and $\mathbf{H}_1$ define *nested* codes according to

$$\begin{aligned} \mathcal{C} &= \{\boldsymbol{\xi} : \mathbf{H}\boldsymbol{\xi} = \mathbf{0}\} \\ \mathcal{C}_1 &= \{\boldsymbol{\xi} : \mathbf{H}_1\boldsymbol{\xi} = \mathbf{0}\} \end{aligned} \quad \Rightarrow \quad \mathcal{C} \subset \mathcal{C}_1$$

Code design

Specifications:

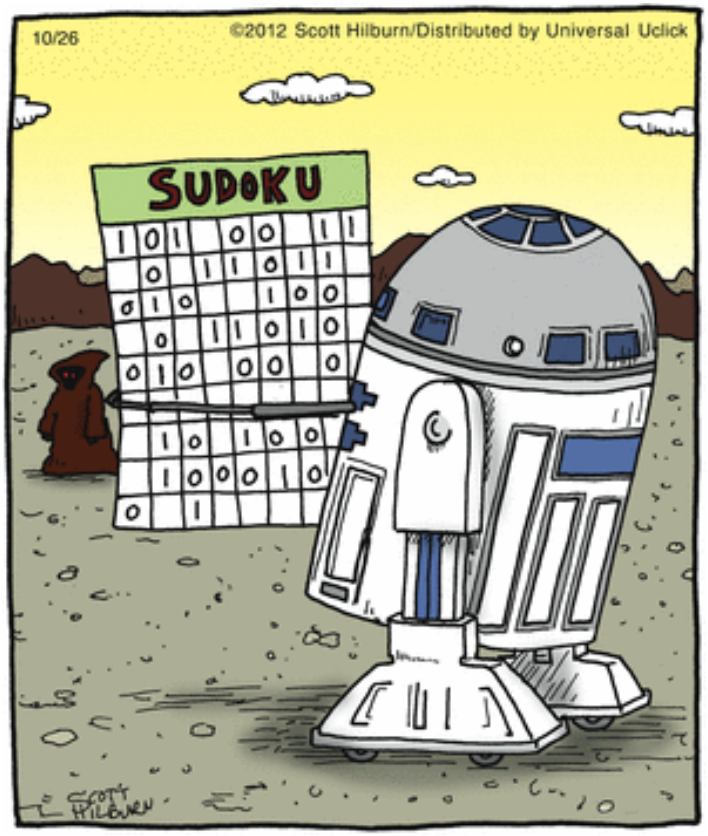
- $\mathcal{C}_1$ is a "fine code" (higher rate) that is *capacity approaching* for Bob's channel ($R_B < C_B$);
- $\mathcal{C}(\mathbf{m})$ is a "coarse code" (lower rate, one code-book per candidate message $\mathbf{m}$) that is *capacity saturating* for Eve's channel ($R_E > C_E$);
- Each coarse code is contained in the fine code: $\mathcal{C}(\mathbf{m}) \subset \mathcal{C}_1$;
- The code word sent by Alice is chosen randomly from $\mathcal{C}(\mathbf{m})$.

Actual secrecy rate is then $R_S = R_B - R_E$.

$\Rightarrow$ Same code construct as in dirty paper coding, information hiding, watermarking, steganography, ...

Wish list

- "Rateless" or "universal" secure codes: secrecy without knowing channel state;
- Multi-terminal extensions (beyond "successively degraded" channels);
- Multi-layer integration;
- Active adversaries (Byzantine nodes);
- "Human-proof" secure key agreement: Agree on secret message rather than secret key;
- Strong versus weak secrecy.



Strong versus Weak Secrecy

Weak secrecy: The rate of information leakage is bounded:

$$\frac{I(X_1^n; Z_1^n)}{n} \leq \epsilon, \qquad \text{for } n > n_*$$

Strong secrecy: The total information leakage is bounded:

$$I(X_1^n; Z_1^n) \leq \epsilon, \qquad \text{for all } n$$

Secrecy capacity essentially the same, although *achievable* strong secrecy methods tend to be more cumbersome.

Exception: Erasure codes/channels

Strong secrecy can be verified using linear algebra (rank of certain matrices).

Strong versus Weak Secrecy

Weak secrecy: The rate of information leakage is bounded:

$$\frac{I(X_1^n; Z_1^n)}{n} \leq \epsilon, \qquad \text{for } n > n_*$$

Strong secrecy: The total information leakage is bounded:

$$I(X_1^n; Z_1^n) \leq \epsilon, \qquad \text{for all } n$$

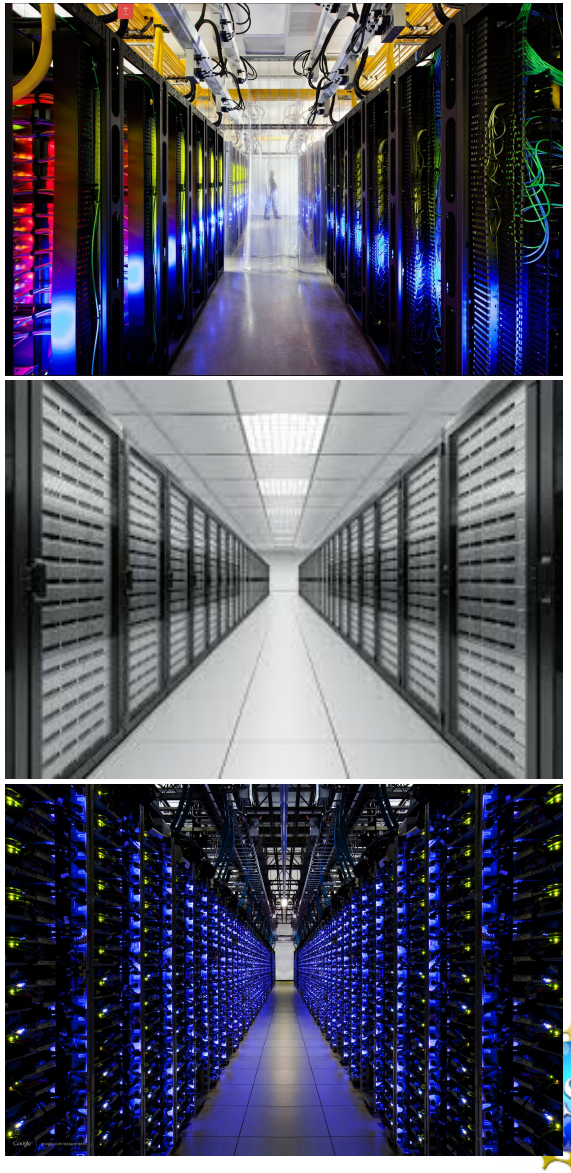
Secrecy capacity essentially the same, although *achievable* strong secrecy methods tend to be more cumbersome.

Exception: Erasure codes/channels

Strong secrecy can be verified using linear algebra (rank of certain matrices).

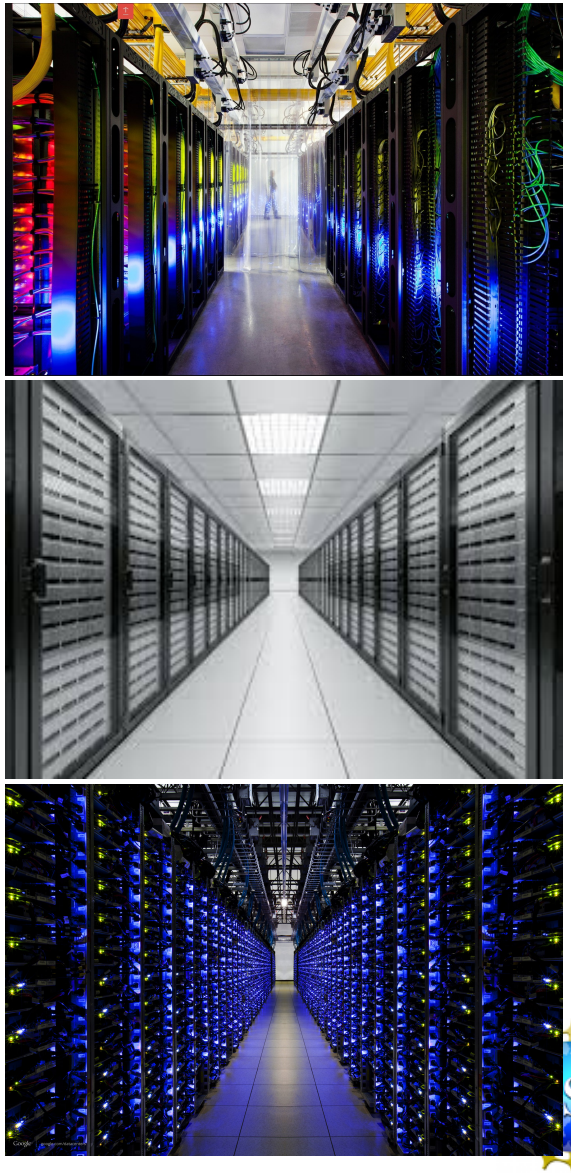
Distributed Storage

- Modern/updated application of erasure codes: hard disk failures, power losses, sabotage, . . . , all appear as network erasures.
- Code design has focused on data recovery at minimal cost (repair bandwidth; locality constraints; maximum failure rate; . . .).
- Can also encode resilience to data theft (using bounded theft model). Strong secrecy is applicable.



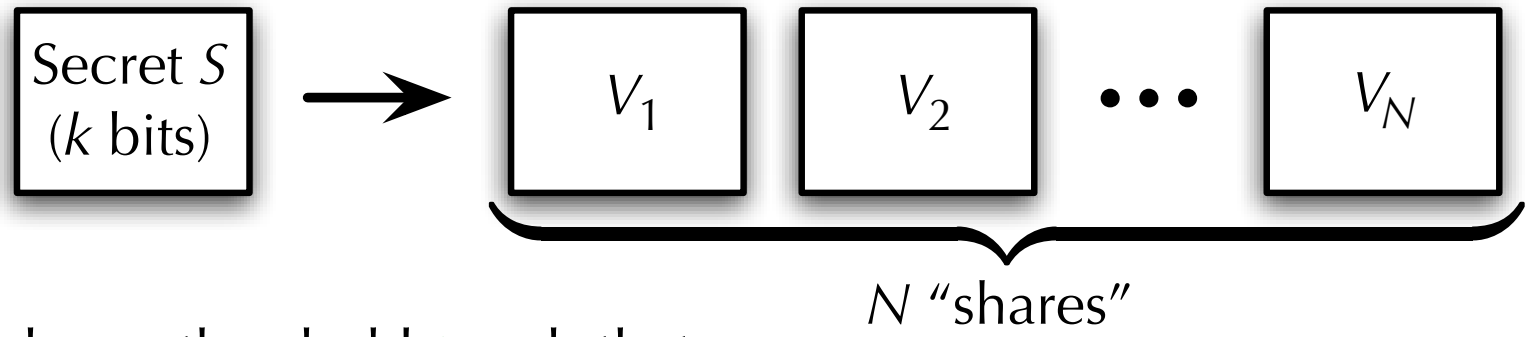
Distributed Storage

- Modern/updated application of erasure codes: hard disk failures, power losses, sabotage, . . . , all appear as network erasures.
- Code design has focused on data recovery at minimal cost (repair bandwidth; locality constraints; maximum failure rate; . . .).
- Can also encode resilience to data theft (using bounded theft model). Strong secrecy is applicable.



Threshold Secret Sharing

Blakley 1979; Shamir 1979; Karnin, Greene & Hellman 1983:



Involves a threshold t such that:

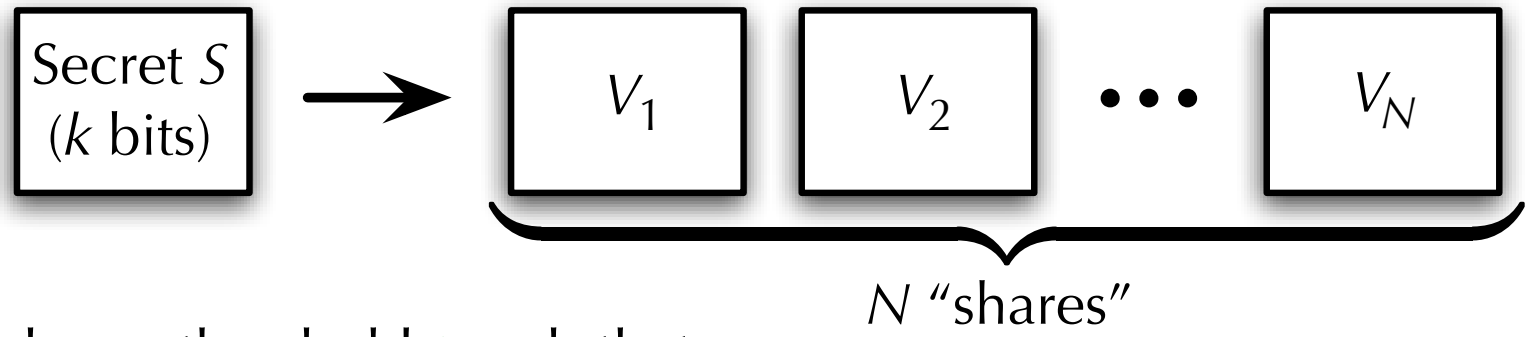
- With any combination of fewer than t shares, no information is leaked on the secret: $I(S; V_{i_1}, V_{i_2}, \ldots, V_{i_{t-1}}) = 0$.
- With any combination of t or more shares, secret is reconstructed: $H(S \mid V_{i_1}, V_{i_2}, \ldots, V_{i_t}) = 0$.

According to KGH (1983), this implies $H(V_i) \geq H(S)$ for each i, and thus

$$\text{Storage Capacity} = \frac{\text{Maximum data size}}{\text{Total storage available}} = \frac{1}{N}$$

Threshold Secret Sharing

Blakley 1979; Shamir 1979; Karnin, Greene & Hellman 1983:



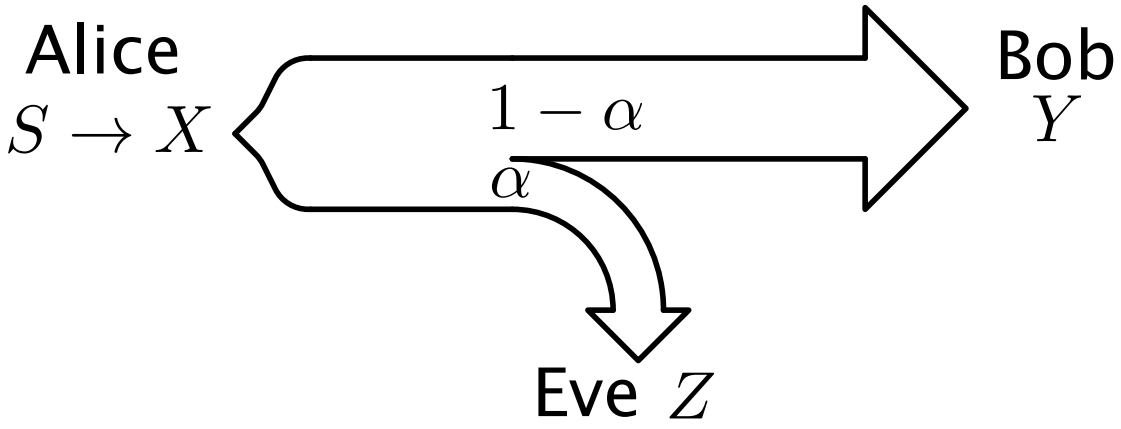
Involves a threshold t such that:

- With any combination of fewer than t shares, no information is leaked on the secret: $I(S; V_{i_1}, V_{i_2}, \ldots, V_{i_{t-1}}) = 0$.
- With any combination of t or more shares, secret is reconstructed: $H(S \mid V_{i_1}, V_{i_2}, \ldots, V_{i_t}) = 0$.

According to KGH (1983), this implies $H(V_i) \geq H(S)$ for each i, and thus

$$\text{Storage Capacity} = \frac{\text{Maximum data size}}{\text{Total storage available}} = \frac{1}{N}$$

Wiretap channel (Wyner, 1975; Cziszar, 1976; Maurer,1993)



Let $\alpha =$ maximum tolerable theft ratio. Storage capacity:

$$\begin{aligned} C_S &= C_{A\to B} - C_{A\to E} \\ &= (1-\alpha) - \alpha = 1 - 2\alpha \end{aligned}$$

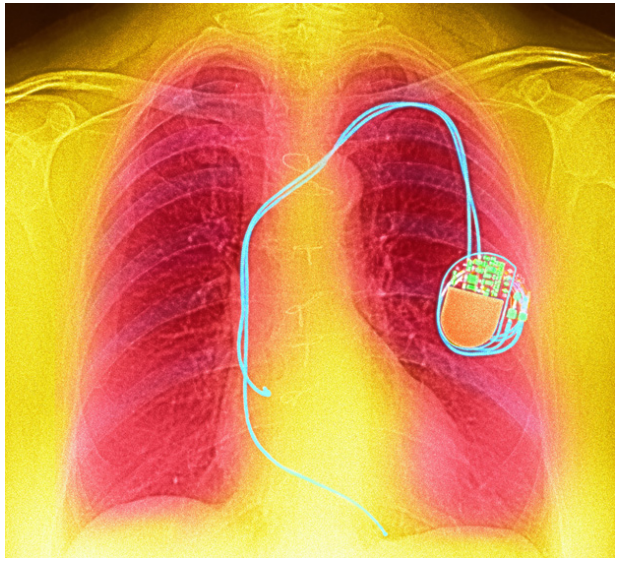
Equating $\alpha = (t-1)/N$,

$$1 - 2\alpha > \frac{1}{N}$$

Internet of Things

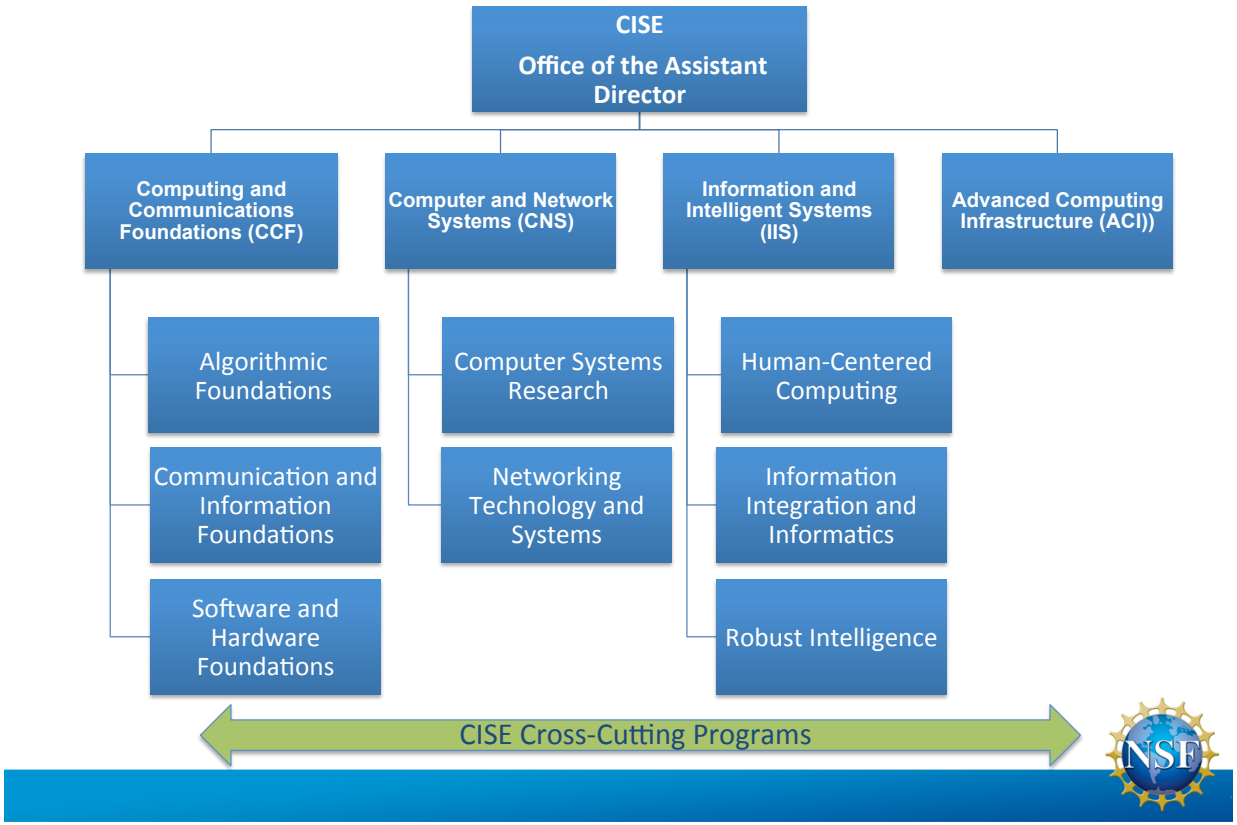
Security needs to be built in by design;

Careful consideration needed for transportation systems, medical devices, critical infrastructure, ...



Computer & Information Science & Engineering (CISE) Directorate

CISE Organization and Core Research Programs



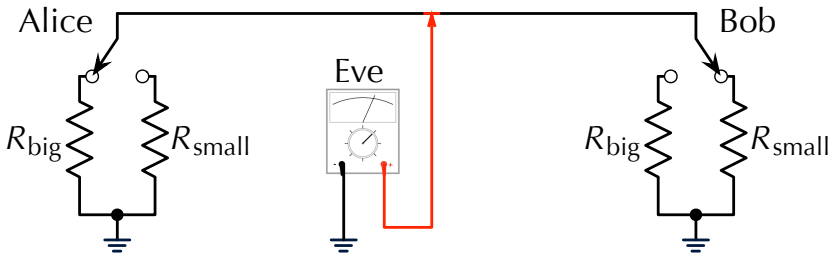
- Research agenda is non-prescriptive;
- We cast a wide net, and fund the best ideas.

Communications and Information Foundations (CIF)

- CIF supports transformative research that addresses the theoretical underpinnings and current and future enabling technologies for information acquisition, transmission, and processing in communication and information networks.
- Foundations of communications and information theory and signal processing, including secure and/or reliable communications, in:
  - wireless and multimedia networks;
  - biological networks;
  - networks of quantum devices;
  - secure communications and storage at the physical layer.

Algorithmic Foundations (AF)

- AF funds innovative and transformative research characterized by algorithmic thinking and algorithm design, accompanied by rigorous analysis, including: Algorithmic foundations for all areas of computer science.
  - Fundamental limits of resource (space, time, communication, energy) bounded computation;
  - Optimal solutions to computational problems under resource bounds;
  - Quantum computation: secure key generation, quantum communication capacity, . . . ;
  - Algorithmic thinking and algorithms for other disciplines (e.g., biology, physics, economics, social sciences).



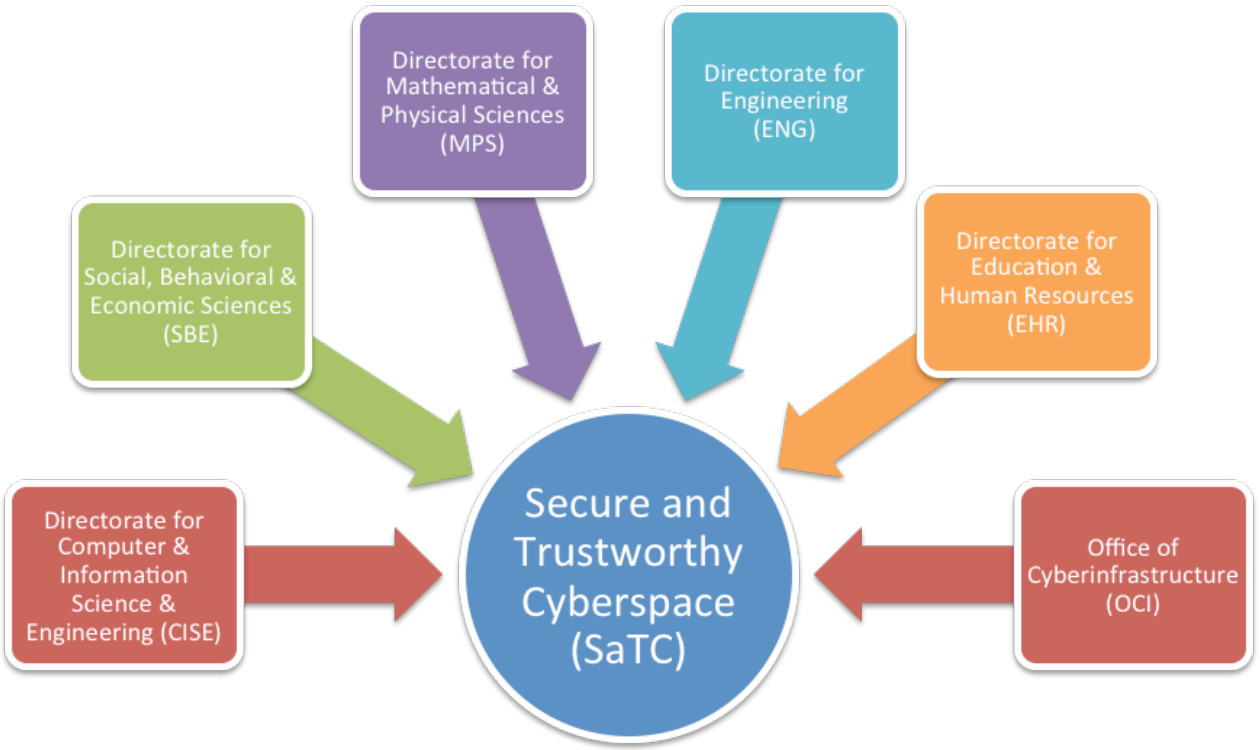
Electrical, Communications and Cyber Systems (ECCS)

ECCS addresses fundamental research issues underlying device and component technologies, power, controls, computation, networking, communications and cyber technologies.

- Integration and networking of intelligent systems principles at the nano, micro and macro scales;
- Application domains in healthcare, homeland security, disaster mitigation, energy, telecommunications, environment, transportation, manufacturing, and others;
- Next generation of devices and systems: convergence of technologies, interdisciplinary research, reaching the goals of the *American Competitiveness Initiative*.

Secure and Trustworthy Cyberspace (SaTC)

Aims to support fundamental scientific advances and technologies to protect cyber-systems (including host machines, the Internet and other cyber-infrastructure) from malicious behavior, while preserving privacy and promoting usability.



SaTC Perspective Goals

- Cybersecurity cannot be fully addressed by only technical approaches.
- SaTC emphasizes different approaches and research communities by introducing perspectives:
  - Trustworthy Computing Systems (TC-S);
  - Social, Behavioral & Economic (SBE);
  - Transition to Practice (TtoP).
- Each proposal must address at least one perspective.
- Proposals are goal-oriented.

Kerckhoffs's last principle (1883)

A crypto system must be easy to use, requiring no mental gymnastics nor memorization of a long series of steps.

Cyber-Physical Systems (CPS)

Many partners this year:

- Department of Homeland Security, Science & Technology Directorate;
- Department of Transportation, Federal Highway Administration
- National Aeronautic and Space Administration
- National Institute of Health, Biomedial Engineering and Bio-Imaging

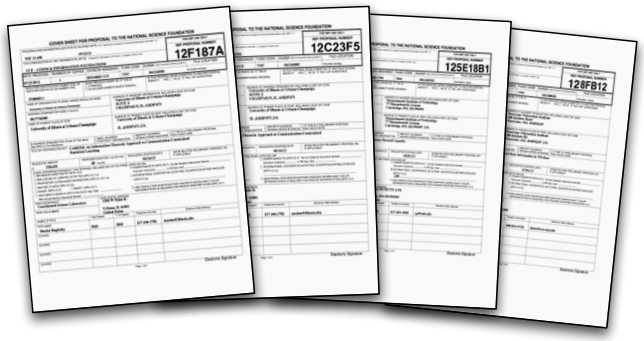
Security (in broad sense) is of particular concern in future transportation systems and medical devices and medical informatics.

(NSF 15-541, due April 20 – May 4, 2015)

NSF Grant Selection

How we work

"NSF's task of identifying and funding work at the frontiers of science and engineering is not a 'top-down' process. NSF operates from the 'bottom up,' keeping close track of research around the United States and the world, maintaining constant contact with the research community to identify ever-moving horizons of inquiry, monitoring which areas are most likely to result in spectacular progress and choosing the most promising people to conduct the research."



http://www.nsf.gov/about/how.jsp

Thanks!

Questions?