Security – A Big Question for Big Data

Prof Roger R. Schell

University of Southern California

Keynote Lecture
IEEE BIgData 2013
Santa Clara, CA
October 9, 2013

Implications of Current State of IT Security

- Current security implications intensified by big data

- Significance of witted adversary
  - Subversion uncommonly impactful and intractable
  - "Existential" national threat – "pre-9/11 moment"
- Reactive "arms race" that cannot work – ever!!
  - Decades of surveillance and patch failure indicate
- In contrast, proven power of principled response
  - Verifiable protection dramatically mitigates subversion
  - A real paradigm shift: no security patches in years of use

Implications of Big Data Key Aspects

- Security impact of big data defining emphases (5 Vs)
  - Volume
  - Velocity
  - Variety
  - Value
  - Veracity
- The key aspects increase challenge for security
  - Science and foundations
  - Infrastructure
  - Management
  - Searching and mining
  - Applications.

Outline of Big Data Security Speech

- **Hard problem: software subversion**
- Ineffective response: band-aid solutions
- Opportunity: leverage verifiable protection

Vulnerable to Trojan Horse Attack

- Hidden functionality in big data apps & “solutions”
  - Adversary usually outsider (stranger to victim)
  - Can be surreptitiously distributed
- Big data application user is unwitting agent
  - Requires victim (user) to execute application
  - Constrained by system security controls on victim
  - Exploitation undetected & controlled by remote design
- Big data cloud paradigm opens vast opportunity
  - Testing & review to detect is futile and delusional
  - Stegonography and such defy perimeter detection
  - Little mitigation in apps & most security solutions

Trap Door Platform Subversion

- Malicious code in platform running big data apps
  - Software, e.g., operating system, drivers, tools
  - Hardware/firmware, e.g., BIOS in PROM
  - Artifice can be embedded any time during lifecycle
  - Adversary chooses time of activation
- Can be remotely activated/deactivated
  - Unique “key” or trigger known only to attacker
  - Needs no (even unwitting) victim use or cooperation
- Efficacy and Effectiveness Demonstrated
  - Exploitable by malicious applications, e.g., Trojans
  - Long-term, high potential future benefit to adversary
  - Testing not at all a practical way to detect
  - Even open source cannot be counted on

Big Data Security Network Reality

Determined competent adversary understands

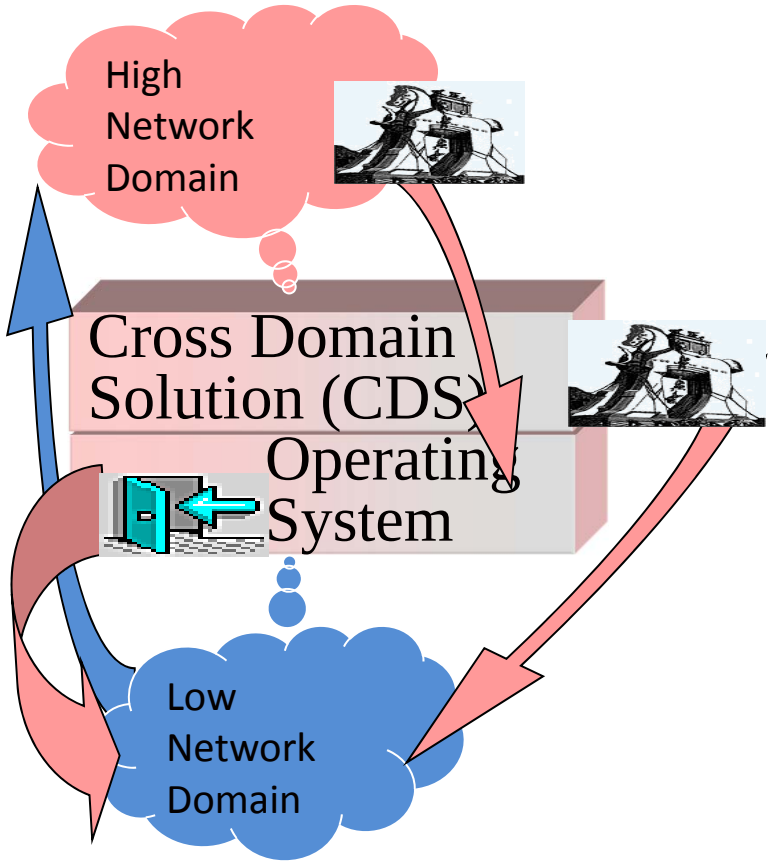
Reality

of current CDS:

Trojan horse planted: Substantial high data leakage to low domain

Malicious software gives low attacker access to data

Trap door planted: Low has repeated undetectable access to high information for years or decades



NPS Linux Trap Door Demo

- Navel Postgraduate School thesis*
- Major Linux distribution
- Network File Server (NFS) subversion
  - Trap door activated with trigger known to attacker
  - Unrestricted access to all NFS files
  - Attacker need not be legitimate user of the system
  - Less than a dozen lines injected into source
- Triggered by attacker over the Internet
  - Representative of platforms in big data context
- Only defense: don't connect to big data network

*March 2002 Thesis by Emory A. Anderson, III

Summary of Subversion Process

- Step #1 – infrastructure subversion
  - Integral to installed software, e.g. trap door
  - Added to software suite during lifecycle, e.g., viruses
  - Big attraction: easy to avoid being apprehended
    - Perpetrator not present at time of attack
- Step #2 – execution of artifice software
  - Can activate by unique "key" or trigger
  - NPS demo, 12 lines of code (LOC) subverts Linux NFS
- Step #3 – (optional) "two card loader"
  - Bootstrap small toehold for diverse customized attacks
  - NPS demo with 6 LOC to subvert XP and then IPSEC
- Step #4 – access unauthorized domain data
  - Information flow between big data sources

Outline of Big Data Security Speech

- Hard problem: software subversion
  - Low cost, low risk to attacker, virtually undetectable
  - Highly effective, extensible, e.g., “two card loader”
- **Ineffective response: band-aid solutions**
- Opportunity: leverage verifiable protection

Common Practice Misaligned with Threat

- Internet is notorious for abysmal security
  - Failures of firewalls, intrusion detection, web servers
  - Strong crypto, VPN and PKI on foundations of sand
- Operating systems are soft underbelly
  - Demonstrated by Argus “Pit Bull” lost challenge
  - Open to planned attack – trap doors & Trojan horses
  - Pervasive source of liability in Internet components
- No business recourse for platform failures
  - Exposed data, crypto keys and forged certificates
  - Security not objectively measured for insurance
  - Billions in opportunity costs – need web efficiencies
  - Insecurity has doomed numerous “solutions”
- Big data environment intensifies these challenges

Impact Indications and Warning

- Vendor downloadable product subverted
  - "Cracker gained user-level access to modify the download file. . . . you pray never happens, but it did."
    - – WordPress, reported on wordpress.org, March 2, 2007
- SW subversion steals credit/debit card data
  - "an 'illicit and unauthorized computer program' was secretly installed at every one of its 300-plus stores."
    - – Hannaford Bros. Co., reported on eWeek.com, March 28, 2008
- IC recognition that subversion is likely
  - "shocked if tools and capabilities and techniques have not been left in U.S. computer systems,"
    - – Admiral Mike McConnell, Recent DNI, CBS 60 Minutes, Jun 10, 2010
- Russian spies use stenography in espionage
  - "more than 100 text files embedded in steganographic

Flaws in System Solutions Missed

- False security from isolated components
- Customers cannot responsibly judge flaws
  - Lack "approved" ***system*** security evaluation criteria
  - Unskilled in assessing methods to address subversion
- Only a verifiably secure CDS is evaluatable
  - On verifiable trusted computing base (TCB) platform
  - Last coherent codification in TCSEC "Class A1"
  - ***System*** security must be designed in, not bolted on
  - Includes composition of "partitions" and "subsets"

Future of Security Band-aids

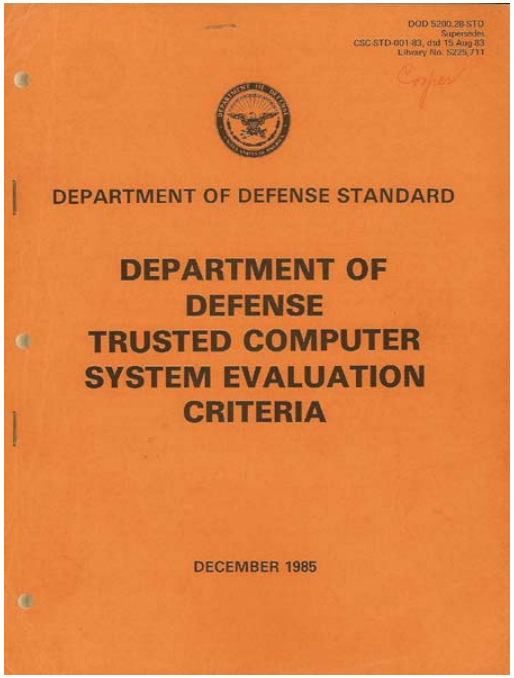
Current Destination

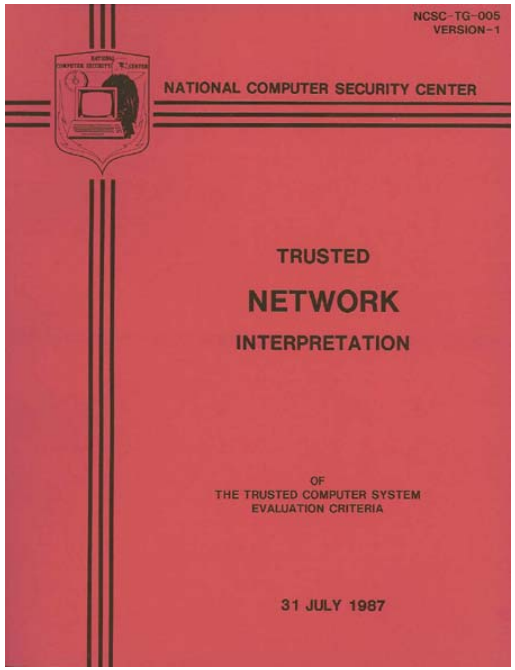
Outline of Big Data Security Speech

- Hard problem: software subversion
  - Low cost, low risk to attacker, virtually undetectable
  - Highly effective, extensible, e.g., “two card loader”
- Ineffective response: band-aid solutions
  - Current practice invites catastrophic big data impacts
  - Pixie dust of “secure” components gives false security
- **Opportunity: leverage verifiable protection**

Verifiable Protection Mitigates Subversion

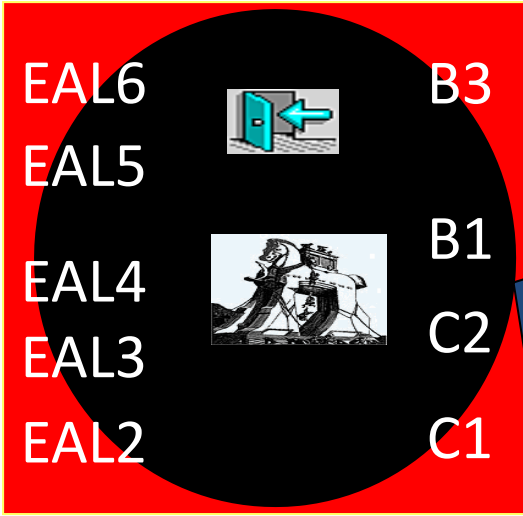
- **Mature, proven trusted systems *technology***
  - TCSEC/TNI need not be used as organizational utterance for policy





Can "Substantially Addresses" Subversion

Common Criteria	TCSEC	
EAL7 (Selected PP)	**A1**	**NO VULNERABILITIES**



UNKNOWN VULNERABILITIES

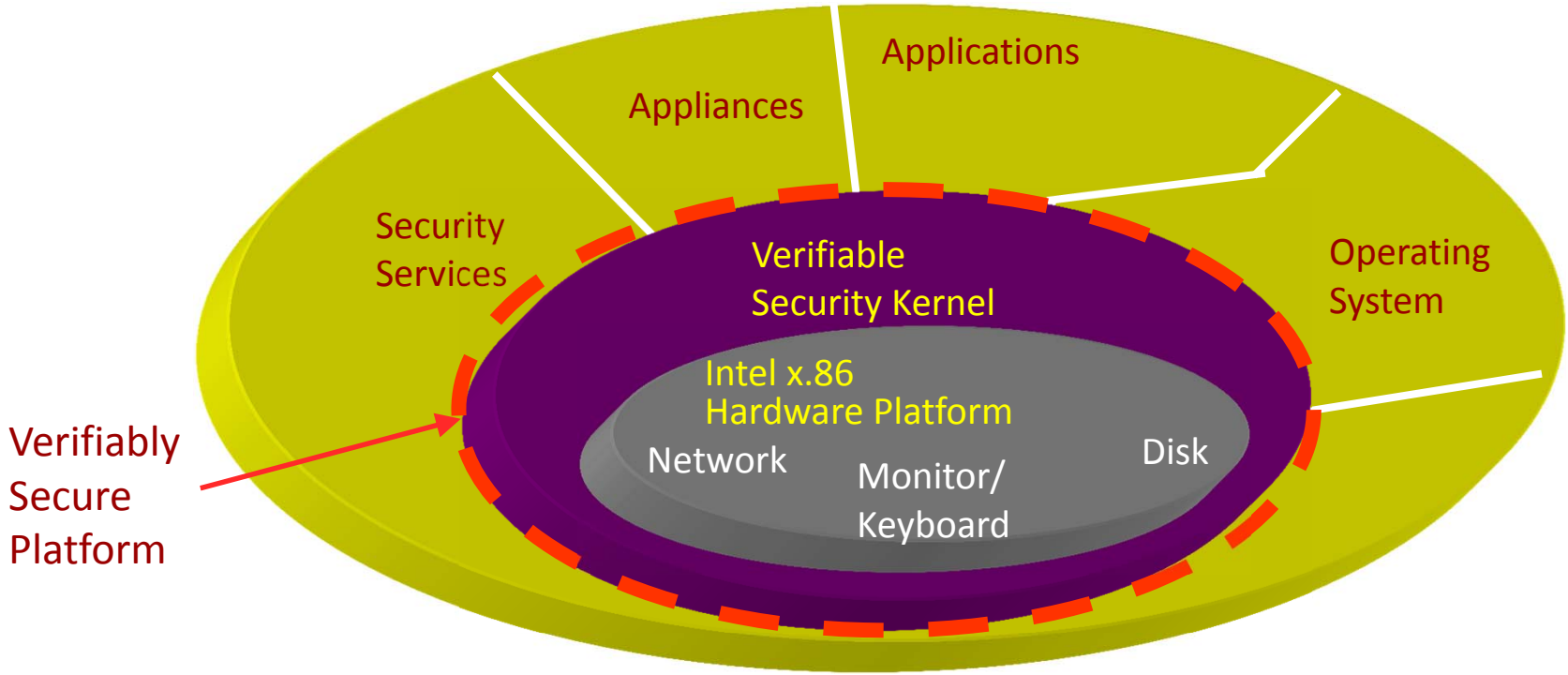
Beware of "No Man's Land"

***Only* Class A1/EAL7 excludes malicious software**

Proven Solution: Security Kernel

The only way we know . . . to build highly secure software systems of any practical interest is the kernel approach."

-- ARPA Review Group, 1970s (Butler Lampson, Draper Prize recipient)

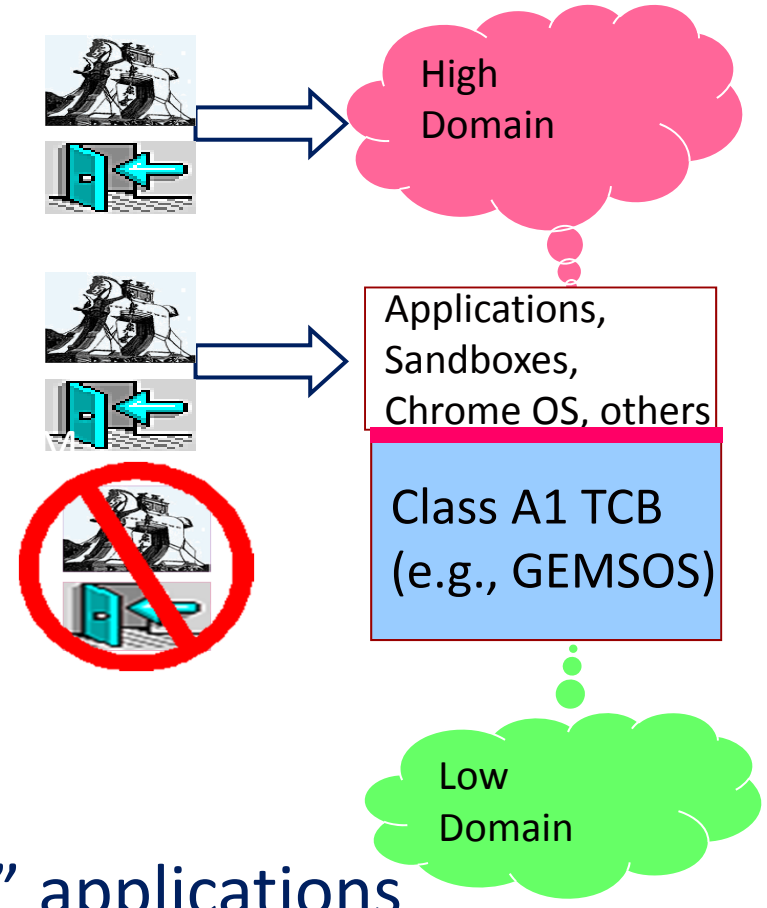


Class A1 enables secure system

Plant trap door or Trojan horse

Impossible to find or fix

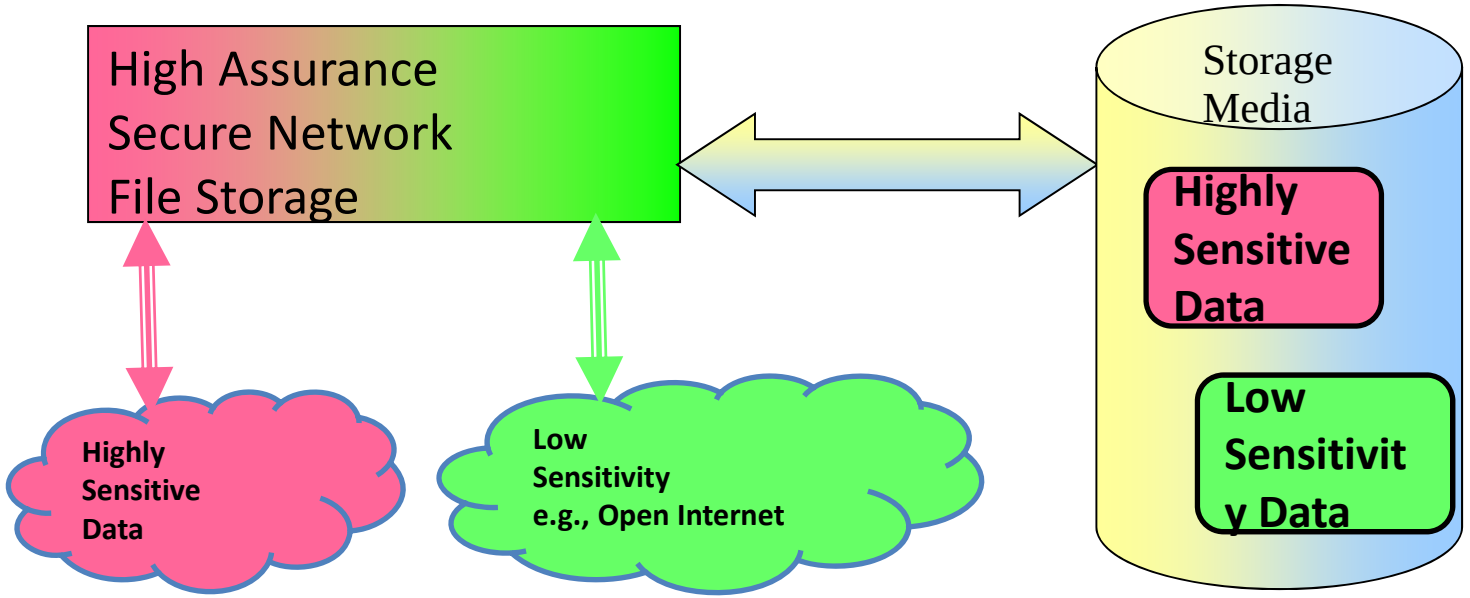
Protects data *despite* apps



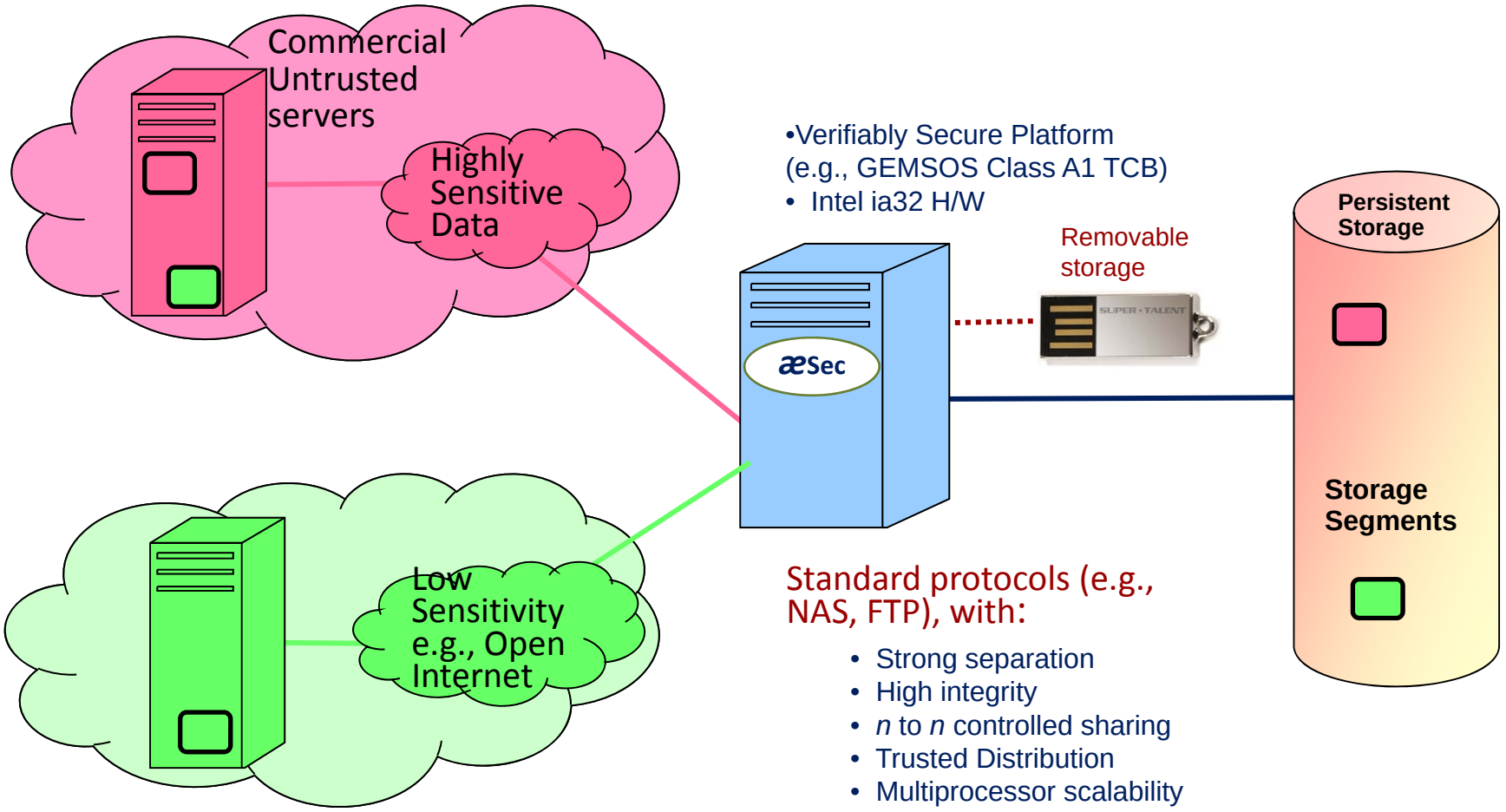
Class A1 doesn't assume "secure" applications

Secure Big Data Cloud Storage

- Operates like standard file storage
- BUT, verifiable security for big data



Secure Big Data Cloud Storage



Outline of Big Data Security Speech

- Hard problem: software subversion
  - Low cost, low risk to attacker, virtually undetectable
  - Highly effective, extensible, e.g., “two card loader”
- Ineffective response: band-aid solutions
  - Current practice invites catastrophic mission impacts
  - Pixie dust of “secure” components gives false security
- Opportunity: leverage verifiable protection
  - Innovative application designs to exploit TCB
    - Preserve much of existing software
    - Apply supportive hardware, e.g., segmentation, rings, TPM
  - Demo starting with sound security, adding functionality
  - Help users validate product hypothesis to vendors

Impact Summary for Big Data

- Computation for massive amounts of data
  - Complex analytics and database operations
  - Remotely from the data owner's enterprise
  - Access to data from multiple and diverse domains
- Limitations of security best practices well-known
- Big data increases opportunity for attackers
  - Insert malicious software in apps and operating systems
- Pivotal choice for big data
  - Use rich set of proven concepts for verifiable protection
  - Risk massive disasters that discredit big data approach
- Need good education and reference implementation

Security – A Big Question for Big Data

Prof Roger R. Schell

University of Southern California

Keynote Lecture
IEEE BIgData 2013
Santa Clara, CA
October 9, 2013