EMC²
where information lives®

Data Security and Data Intensive Computing

David L. Black, EMC Corporation

What is it?

- World's largest and fastest flying fish?
- Or a High Bandwidth Data Transfer Device?
  - With thanks to Jim Gray

Data Security and Data Intensive Computing

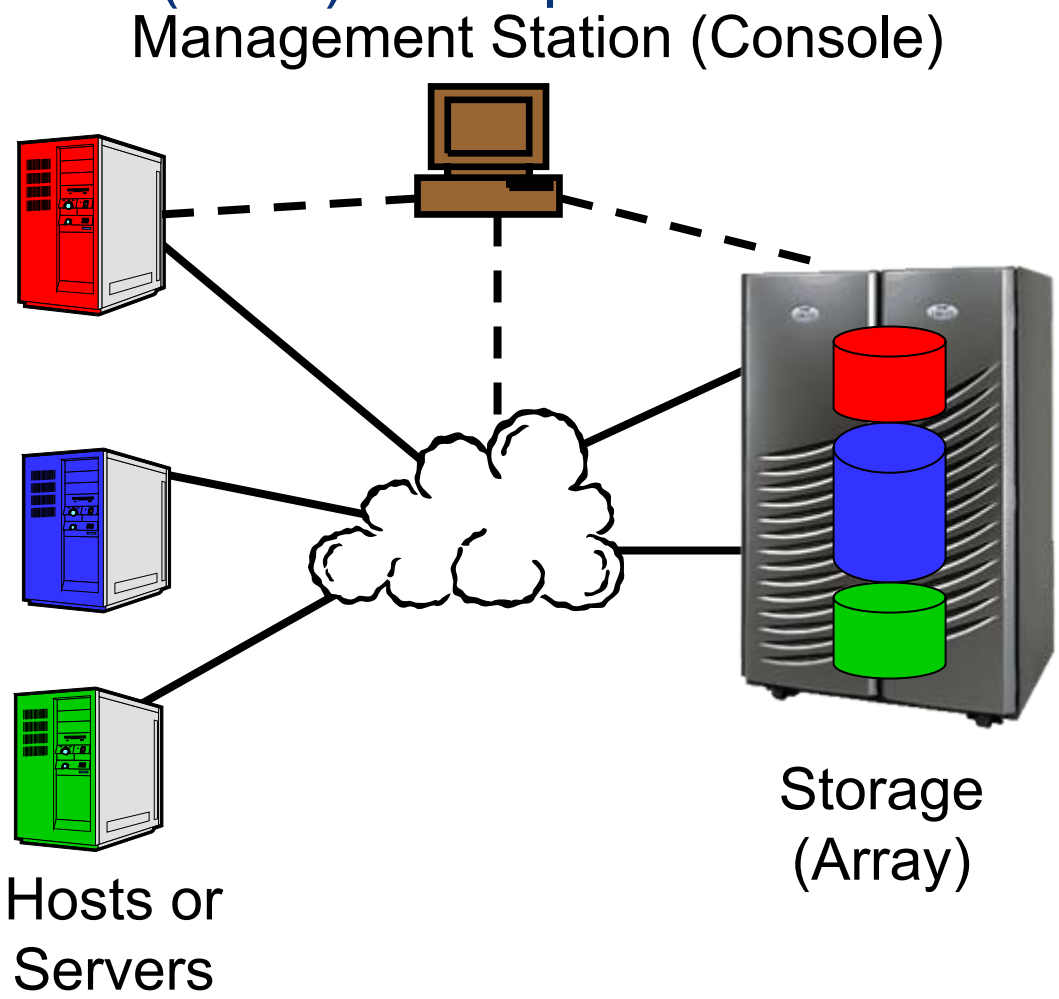
- Where does HPC data spend most of its life?
  - Stored “somewhere”, generally not on an HPC machine
  - That “somewhere” is the obvious first place to attack the data.
- HPC machines are not the most important components for data security
  - Q: Does that make data security irrelevant for HPC machines?
  - A: No, like a chain, security is only as strong as its weakest link.
- Data moves between “somewhere” and HPC machines
  - That movement usually involves a network
- This talk: Security framework for networked storage of data
  - What are the threats and how can data be protected?
  - Examples are primarily SAN (Storage Area Network)
  - Most concepts apply to fileservers and HPC filesystems

Storage Networking Technologies

- Storage Area Networking (SAN)
  - Provides (virtual) disk volume storage
  - SCSI protocol family (e.g., parallel, Fibre Channel, iSCSI)
- Network Attached Storage (NAS)
  - Provides file (and filesystem) storage
  - NFS and CIFS over TCP/IP
- Parallel HPC filesystems (e.g., Lustre)
  - Security issues are mostly analogous to NAS

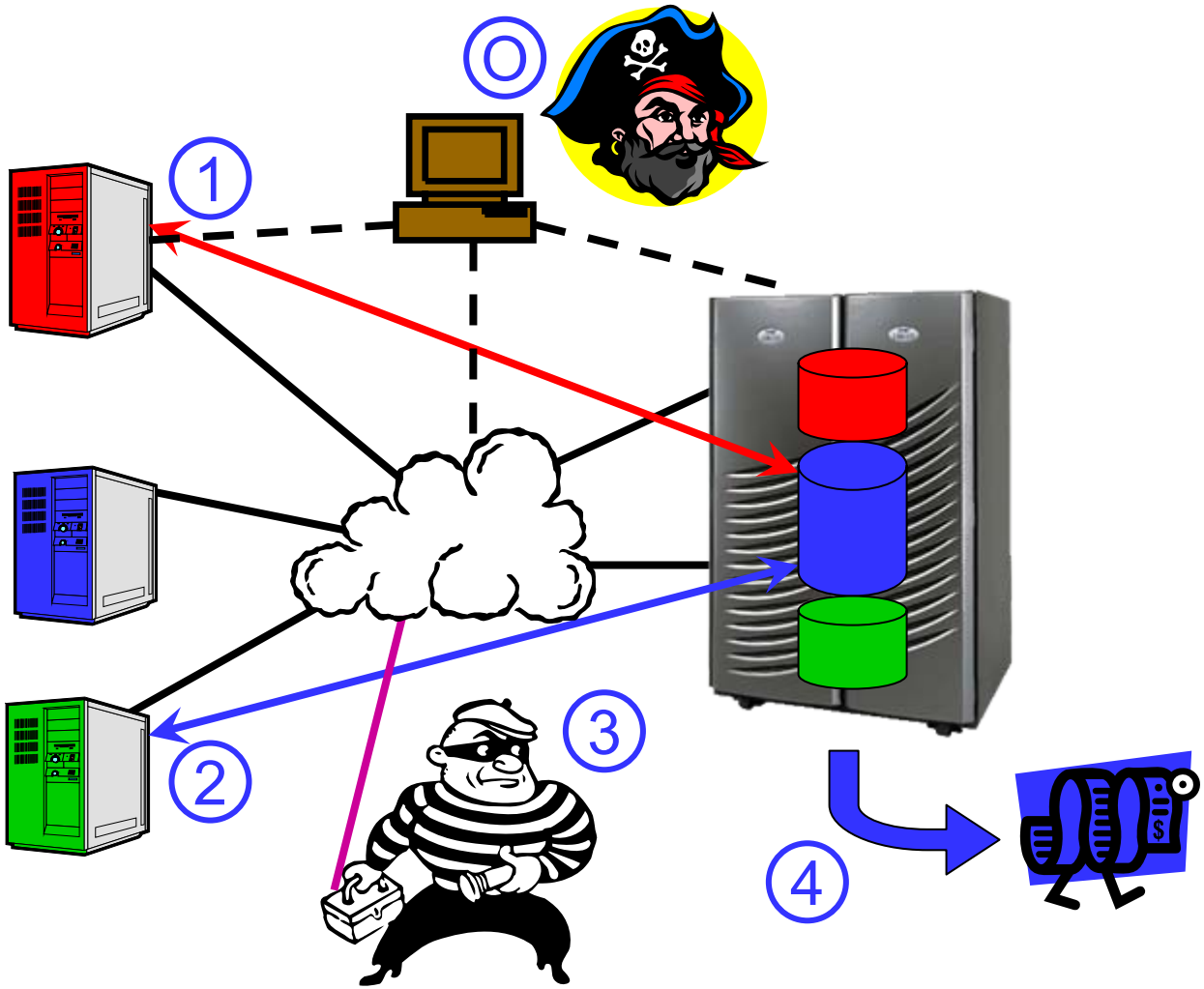
Storage Area Network (SAN) Example

- Data: SCSI, e.g.,
  - Fibre Channel
  - iSCSI
- Mgt.: usually IP
  - SNMP
  - SMI-S (CIM)
- NAS and filesystems share data among hosts and servers



Security Threats

0. Management Attacks & Abuse
1. Uncontrolled Data Access
2. Impersonation (Spoofing)
3. Communication Access
   - Eavesdrop
   - Inject/Modify
4. External Access
   - Media Theft
   - Other access and destruction



Why is management threat number 0?

Top three reasons ...

3. I'm from Boston ☺
   - Thrifty Yankee presenter recycled slide content ☺ ☺
2. This is storage – we start counting from zero ...
   - Disk numbers start at 0 (e.g., boot drive or volume)
1. If management isn't protected, nothing else matters!
   - Full management privileges ≈ root access on a host

Threat 0: Management Attacks & Abuse

- Attacker Goal: Management privileges
- Multiple attack vectors, for example:
  - Obtain authentication credentials (e.g., password in clear)
  - Modify management traffic (e.g., hijack)
  - Perform unauthorized management actions
- Countermeasures: Management Security
  - Authentication & Authorization
  - Log actions and protect logs
  - Secure Channels for management communication
    - Confidentiality, Cryptographic Integrity, and Anti-Replay

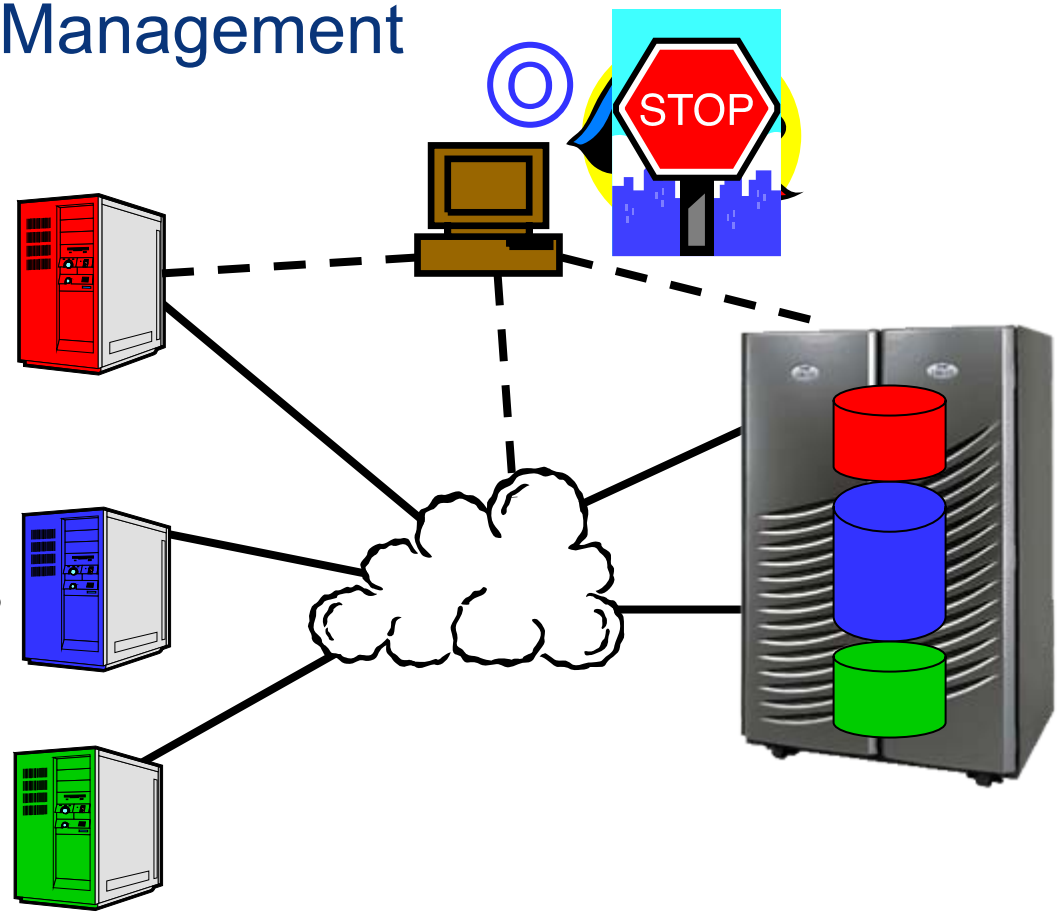
Management Security Mechanisms

- Secure Management Interface Protocols
  - Command line interfaces: SSH (secure shell)
  - Web interfaces: SSL/TLS standards
- SNMP (Simple Network Management Protocol)
  - SNMP versions prior to v3 do not support strong security
  - AES is available for SNMPv3
  - Work underway on SSH security framework for SNMPv3
- SNIA SMI-S: New storage management standard
    - SNIA: Storage Networking Industry Association
    - SMI-S: Storage Management Initiative – Specification
  - Web based - reuses existing web standards
    - SSL 3.0, TLS and HTTP basic authentication required

Security Threats: Management

0. Management Attacks & Abuse

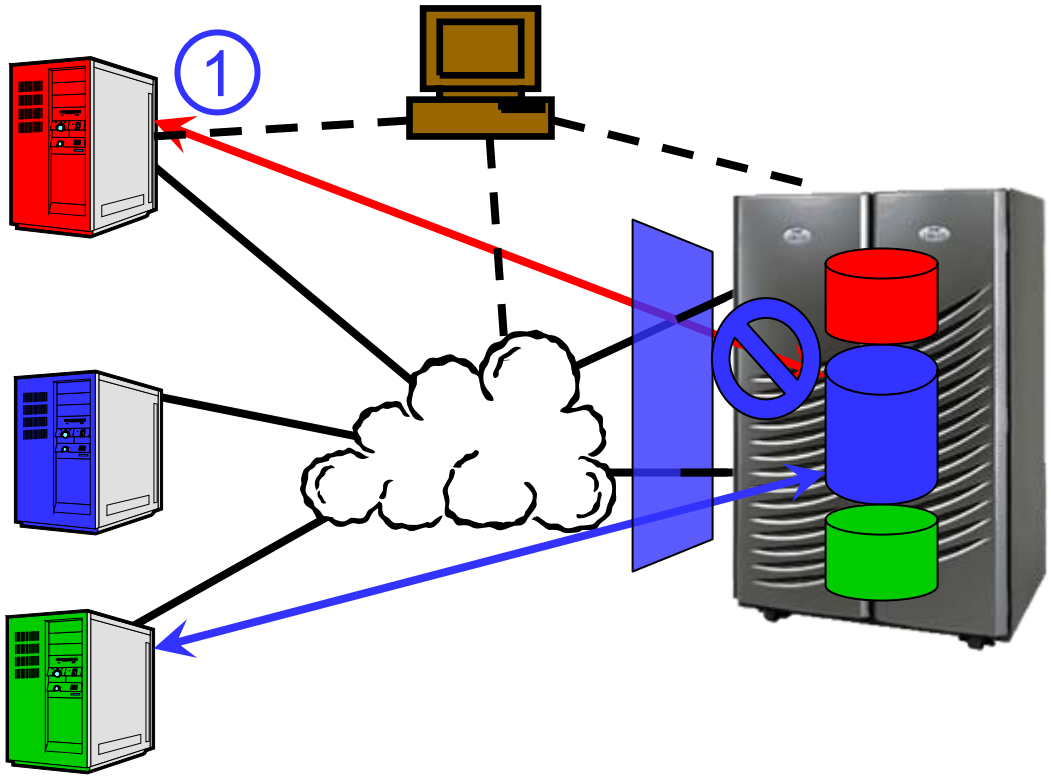
- Countermeasures: Mgt. Security
  - Authentication
  - Authorization
  - Logging
  - Secure Channels



Threat 1: Storage Access

1. Uncontrolled Storage Access

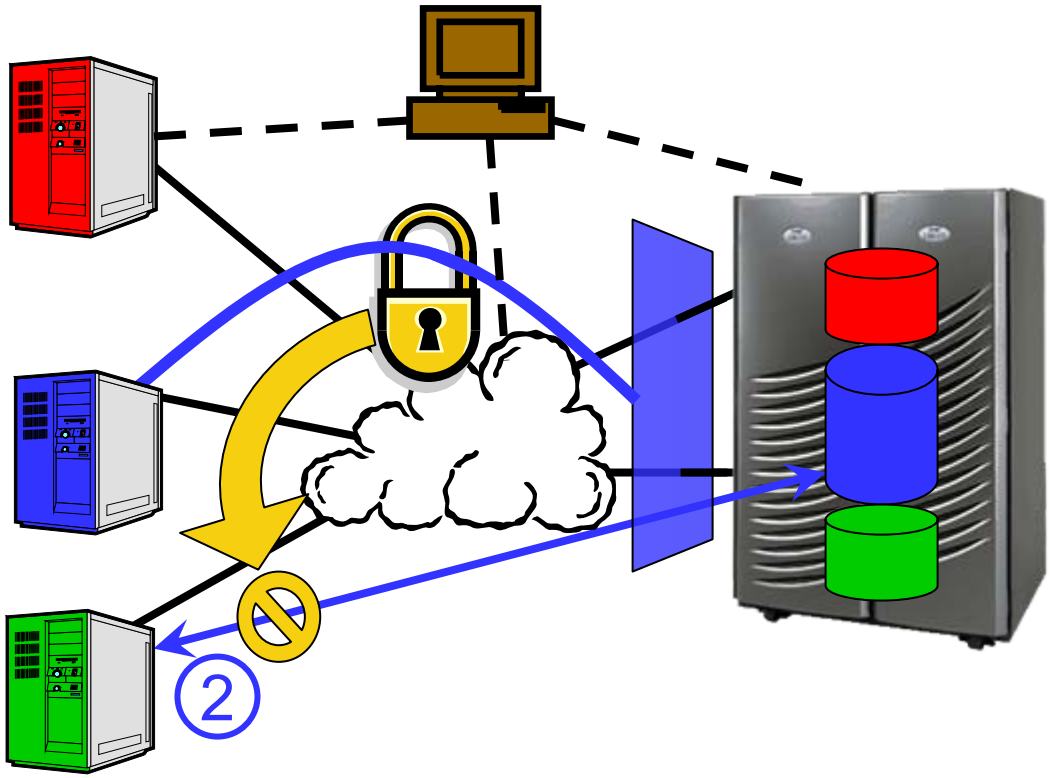
- Countermeasure: Access Control
  - SAN: LUN masking and mapping
  - Usually not a concern for NAS or filesystems
- Does not prevent Impersonation



Threat 2: SAN Server Impersonation

2. Impersonation (Spoofing)

- Countermeasure: Authentication (Proof of identity)



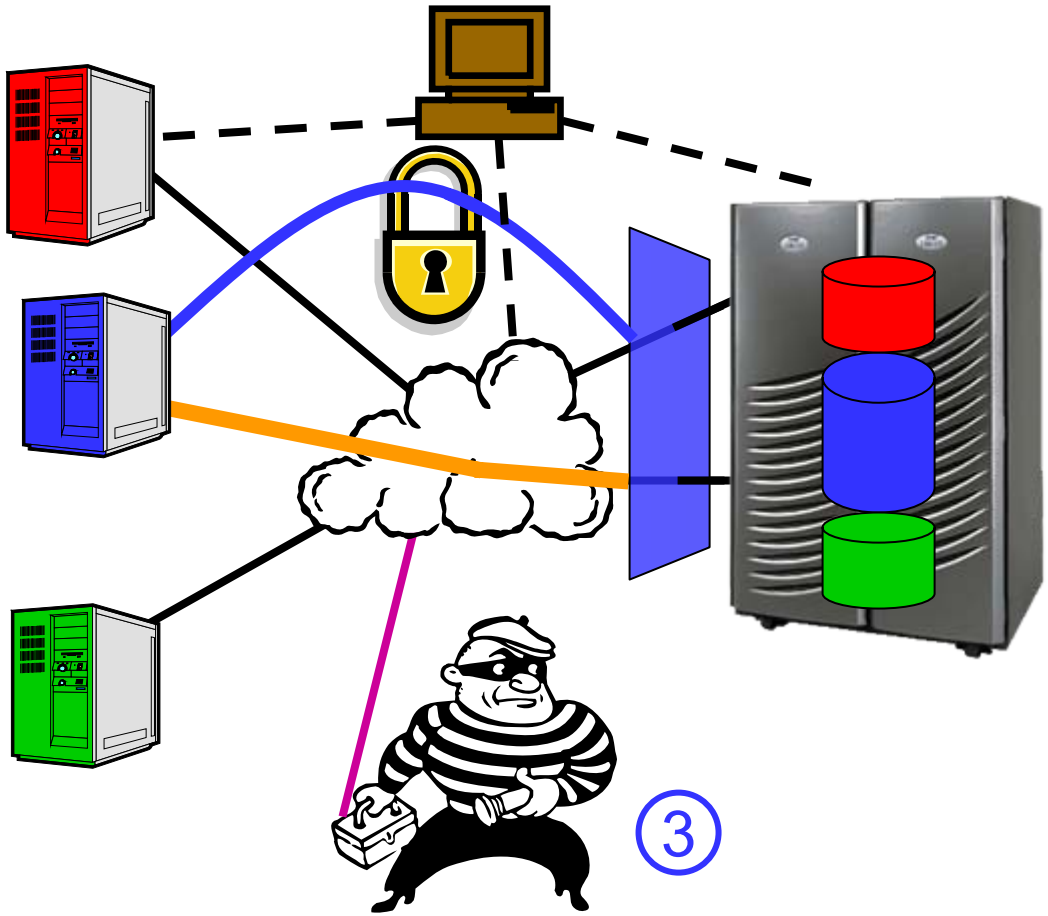
Networked Authentication

- Variety of authentication mechanisms for users
  - Kerberos, certificates, challenge/response tokens, etc.
- The challenge is in the infrastructure
  - Need to integrate with authentication infrastructure
    - Directories (e.g., via LDAP). Kerberos, PKI, etc.
    - Avoids multitude of passwords for each individual
    - Token based mechanisms also need to be integrated
  - Different management domains are an added complication
- Need to authenticate machines in some cases (e.g., SAN)
  - iSCSI has inband authentication (Fibre Channel will soon)

Threat 3: Communication Access

3. Communication Access
   - Eavesdrop
   - Inject/Modify

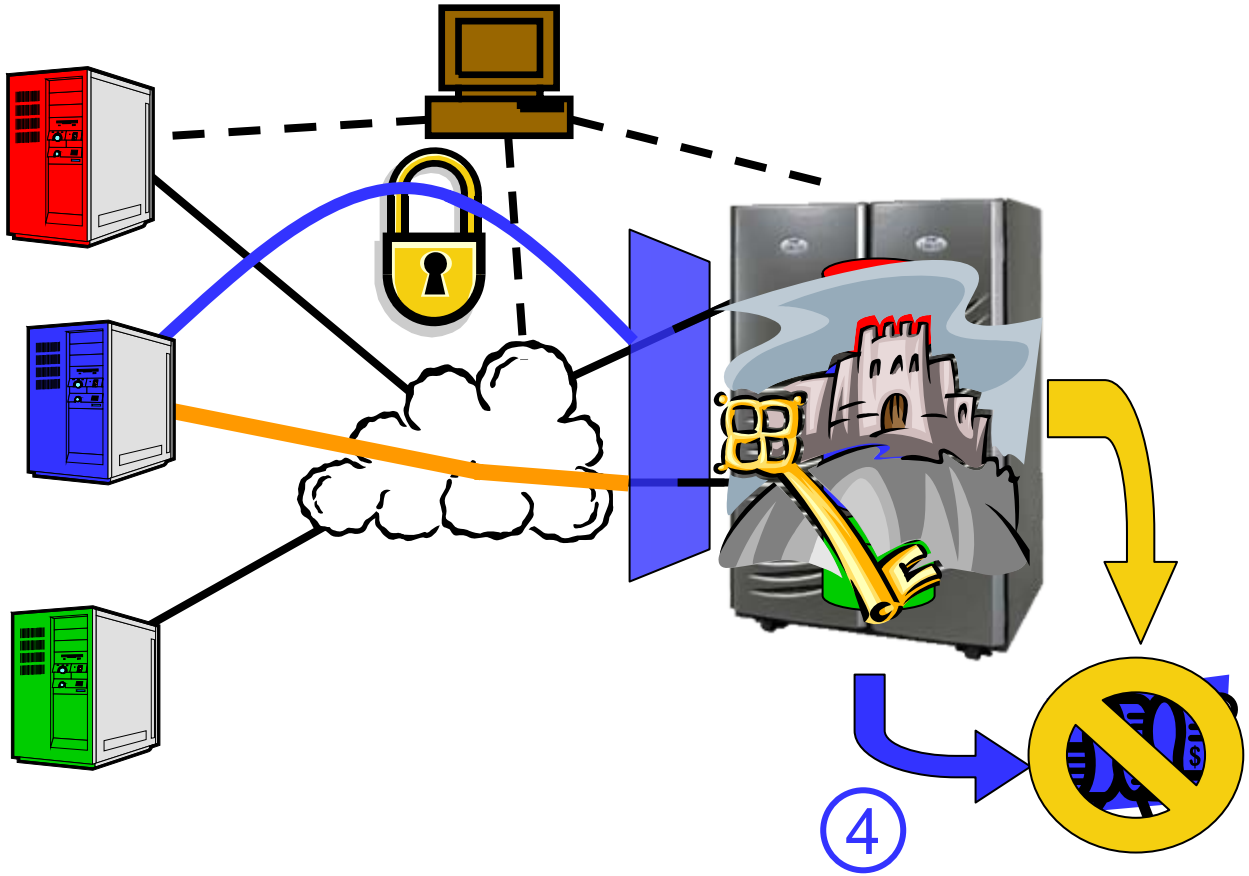
- Countermeasure: Secure Channel
  - Confidentiality
  - Cryptographic Integrity
  - Anti-Replay



Securing Communication Channels

- IP Security (IPsec)
  - Typical use: VPNs
  - Packet-based, operates at IP (layer 3)
  - Can secure CIFS, iSCSI, etc.
  - Being applied to Fibre Channel
- SSL/TLS and SSH
  - Typical uses: Web (SSL or TLS), command line interface (SSH)
  - Session-based, operate above TCP (layer 5)
- Kerberos-based mechanisms
  - Integrated into NFS

Threat 4: External Data Access



4. External Access
   - Media Theft
   - Other Access

- Countermeasure: Stored Data Security

Stored Data Security

- Threat: Disclosure of stored data
  - Threats: media access or theft, including backups
- Disclosure protection for stored data (often encryption)
  - Multi-year data lifetime complicates key management
- Encrypt in place: usually confidentiality-only
  - No additional space to store cryptographic integrity checks
  - Tweaked encryption modes can prevent block swapping
  - Encrypted tape can provide cryptographic integrity checks
- Alternative: Application-level encryption
  - Database row encryption, PGP, encrypted files, etc.
- Backup tape encryption becoming a best practice

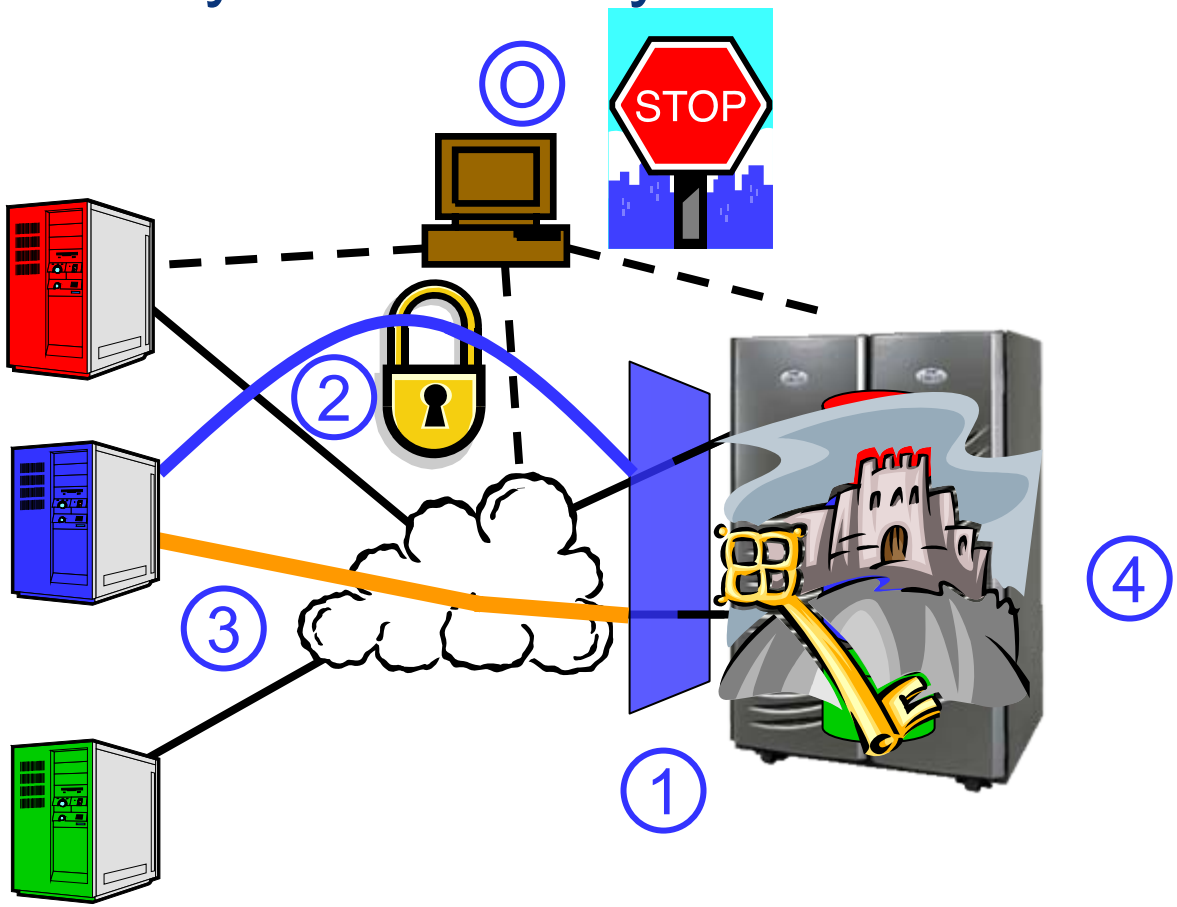
EMC² where information lives®

Regulatory Compliance – A Security Perspective

- Regulatory Compliance threats to stored data:
  - Inability to produce the data
  - Inability to prove the integrity of the data (potentially to a court)
- Industry Response: Fixed Content Storage Systems
  - Lots of examples: EMC Centera, HP RISS, etc.
  - Assurance of data availability
  - Assurance of data integrity
- Analogy to Mandatory Access Controls
  - Overwrite and Delete tightly controlled in Fixed Content design
- NOTE: There are many additional aspects to regulatory compliance beyond security

Networked Data Security Functionality Review

0. Management Security
1. Access Control
2. Authentication (Proof of identity)
3. Secure Channel
   - Confidentiality
   - Cryptographic Integrity
   - Anti-replay
4. Stored Data Security



EMC² where information lives®

Some Questions for the Panel

- Different security for data-in-computation vs. data-at-rest?
  - Opportunity to focus security functionality on stored data?
- Security implications of long-term-storage of data?
  - What if data lifetime is much longer than key lifetime?
- Opportunities to leverage commercial developments?
  - FIPS-compliant data encryption?
  - Fixed content systems for compliance?
  - Others?
- What HPC security needs will the IT industry not address?
  - And what should the HPC community do about them?

Questions?

EMC²
where information lives