CS194-24
Advanced Operating Systems Structures and Implementation
Lecture 23

Networks (Con't)
Security

April 28th, 2014

Prof. John Kubiatowicz

http://inst.eecs.berkeley.edu/~cs194-24

Goals for Today

- Network Drivers (Con't)
- Security

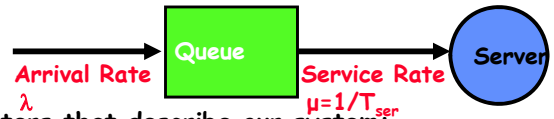
Interactive is important!

Ask Questions!

Note: Some slides and/or pictures in the following are adapted from slides ©2013

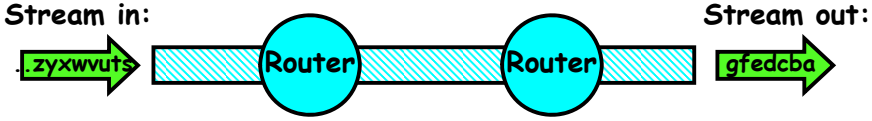
Recall: A Little Queuing Theory: Some Results

- Assumptions:
  - System in equilibrium; No limit to the queue
  - Time between successive arrivals is random and memoryless



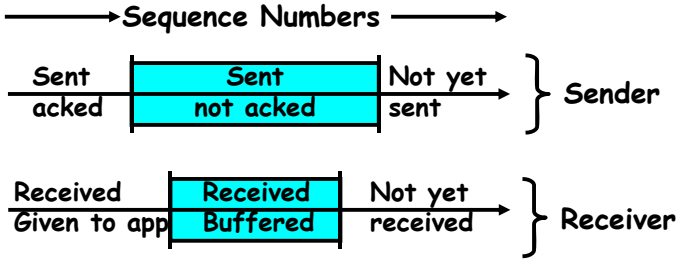
- Parameters that describe our system:
  - λ: mean number of arriving customers/second
  - T_{ser}: mean time to service a customer ("m1")
  - C: squared coefficient of variance = $\sigma^2/m1^2$
  - μ: service rate = $1/T_{ser}$
  - u: server utilization ($0 \le u \le 1$): $u = \lambda/\mu = \lambda \times T_{ser}$
- Parameters we wish to compute:
  - T_q: Time spent in queue
  - L_q: Length of queue = $\lambda \times T_q$ (by Little's law)
- Results:
  - Memoryless service distribution (C = 1):
    - » Called M/M/1 queue: $T_q = T_{ser} \times u/(1 - u)$
  - General service distributon (no restrictions), 1 server:
    - » Called M/G/1 queue: $T_q = T_{ser} \times \frac{1}{2}(1+C) \times u/(1 - u)$

Recall: Transmission Control Protocol (TCP)



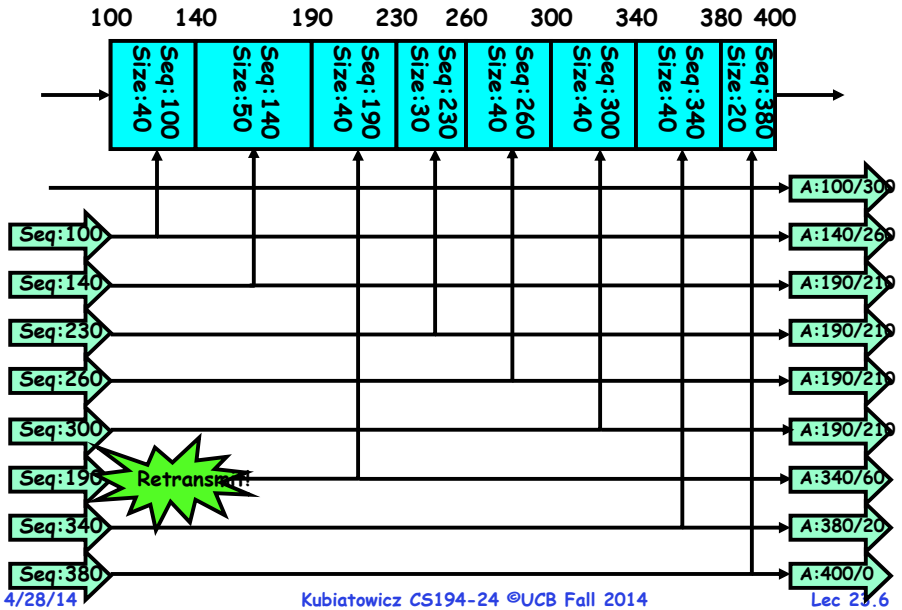
- Transmission Control Protocol (TCP)
  - TCP (IP Protocol 6) layered on top of IP
  - Reliable byte stream between two processes on different machines over Internet (read, write, flush)
- TCP Details
  - Fragments byte stream into packets, hands packets to IP
    - » IP may also fragment by itself
  - Uses window-based acknowledgement protocol (to minimize state at sender and receiver)
    - » "Window" reflects storage at receiver – sender shouldn't overrun receiver's buffer space
    - » Also, window should reflect speed/capacity of network – sender shouldn't overload network
  - Automatically retransmits lost packets
  - Adjusts rate of transmission to avoid congestion
    - » A "good citizen"

TCP Windows and Sequence Numbers



- Sender has three regions:
  - Sequence regions
    - » sent and ack'ed
    - » Sent and not ack'ed
    - » not yet sent
  - Window (colored region) adjusted by sender
- Receiver has three regions:
  - Sequence regions
    - » received and ack'ed (given to application)
    - » received and buffered
    - » not yet received (or discarded because out of order)

Window-Based Acknowledgements (TCP)



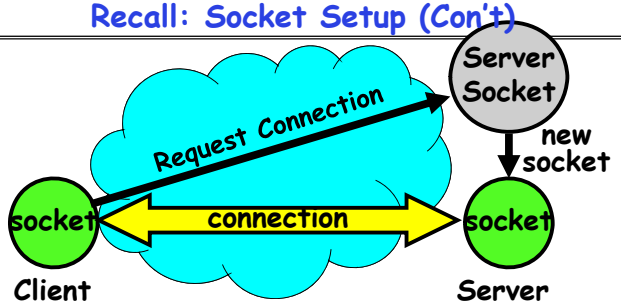
Congestion Avoidance

- Congestion
  - How long should timeout be for re-sending messages?
    - » Too long→wastes time if message lost
    - » Too short→retransmit even though ack will arrive shortly
  - Stability problem: more congestion ⇒ ack is delayed ⇒ unnecessary timeout ⇒ more traffic ⇒ more congestion
    - » Closely related to window size at sender: too big means putting too much data into network
- How does the sender's window size get chosen?
  - Must be less than receiver's advertised buffer size
  - Try to match the rate of sending packets with the rate that the slowest link can accommodate
  - Sender uses an adaptive algorithm to decide size of N
    - » Goal: fill network between sender and receiver
    - » Basic technique: slowly increase size of window until acknowledgements start being delayed/lost
- TCP solution: "slow start" (start sending slowly)
  - If no timeout, slowly increase window size (throughput) by 1 for each ack received
  - Timeout ⇒ congestion, so cut window size in half
  - "*Additive Increase, Multiplicative Decrease*"

Sequence-Number Initialization

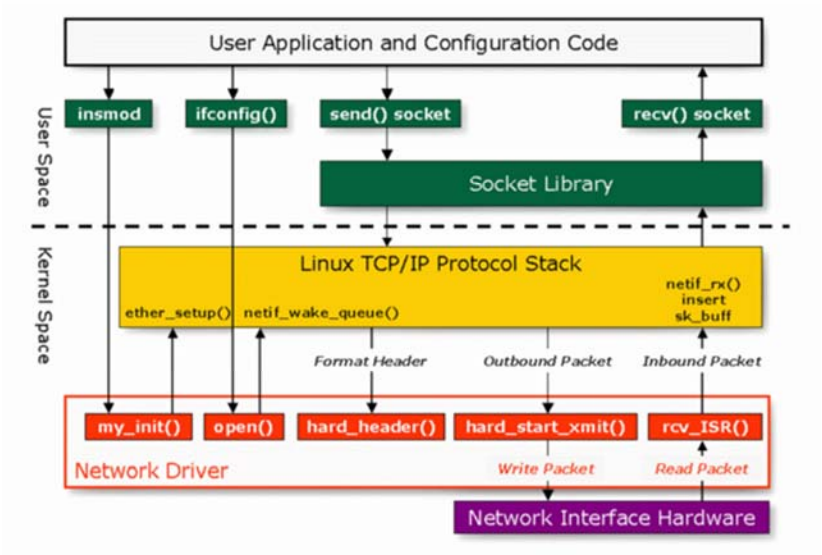
- How do you choose an initial sequence number?
  - When machine boots, ok to start with sequence #0?
    - » No: could send two messages with same sequence #!
    - » Receiver might end up discarding valid packets, or duplicate ack from original transmission might hide lost packet
  - Also, if it is possible to predict sequence numbers, might be possible for attacker to hijack TCP connection
- Some ways of choosing an initial sequence number:
  - Time to live: each packet has a deadline.
    - » If not delivered in X seconds, then is dropped
    - » Thus, can re-use sequence numbers if wait for all packets in flight to be delivered or to expire
  - Epoch #: uniquely identifies *which* set of sequence numbers are currently being used
    - » Epoch # stored on disk, Put in every message
    - » Epoch # incremented on crash and/or when run out of sequence #
  - Pseudo-random increment to previous sequence number
    - » Used by several protocol implementations

Recall: Socket Setup (Con't)

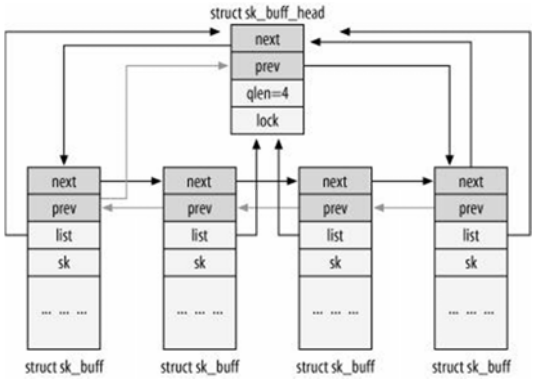


- Things to remember:
  - Connection involves 5 values:
    [Client Addr, Client Port, Server Addr, Server Port, Protocol]
  - Often, Client Port "randomly" assigned
    - » Done by OS during client socket setup
  - Server Port often "well known"
    - » 80 (web), 443 (secure web), 25 (sendmail), etc
    - » Well-known ports from 0—1023
- Note that the uniqueness of the tuple is really about two Addr/Port pairs and a protocol

Linux Network Architecture

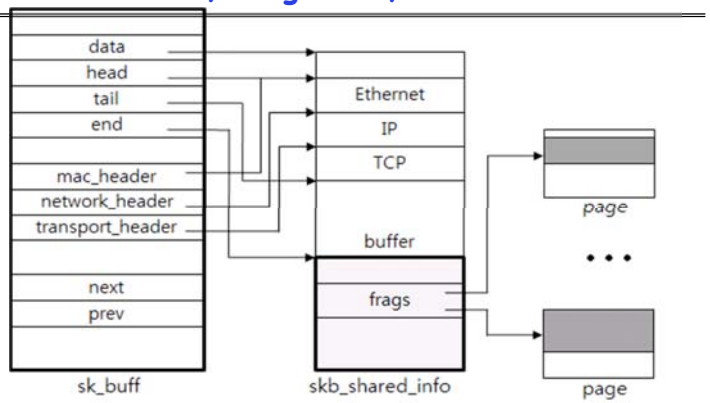


Network Details: sk_buff structure



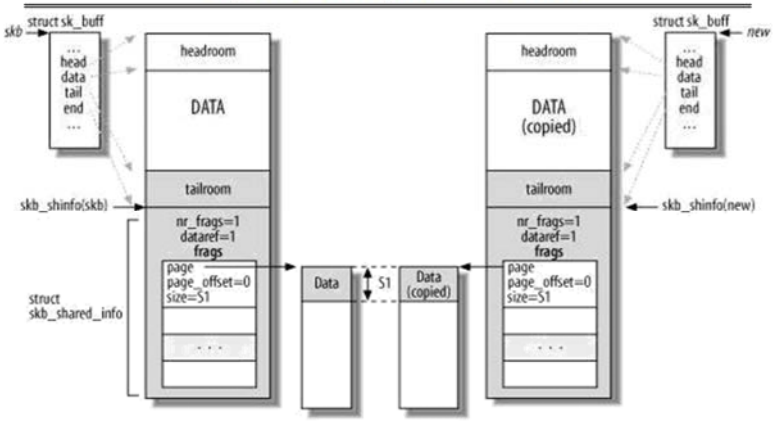
- Socket Buffers: sk_buff structure
  - The I/O buffers of sockets are lists of sk_buff
    - » Pointers to such structures usually called "skb"
  - Complex structures with lots of manipulation routines
  - Packet is linked list of sk_buff structures

Headers, Fragments, and All That



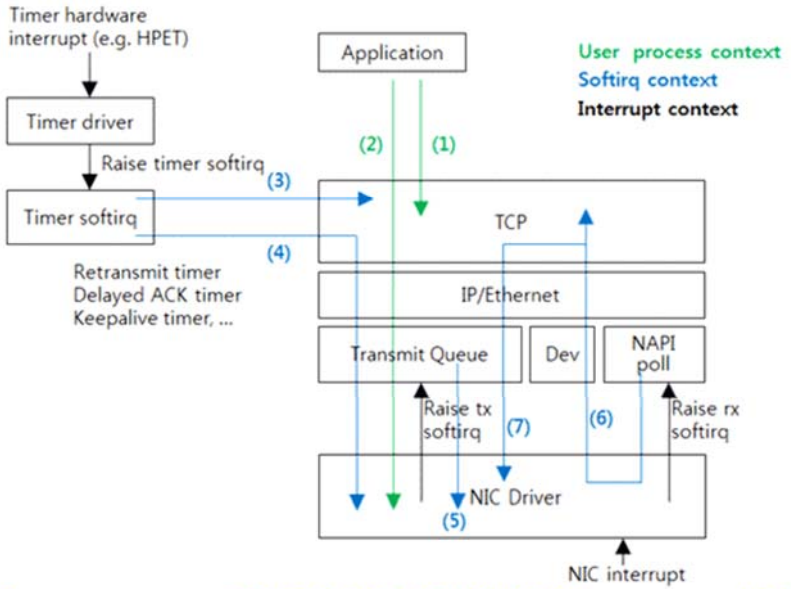
- The "linear region":
  - Space from skb->data to skb->end
  - Actual data from skb->head to skb->tail
  - Header pointers point to parts of packet
- The fragments (in skb_shared_info):
  - Right after skb->end, each fragment has pointer to pages, start of data, and length

Copies, manipulation, etc

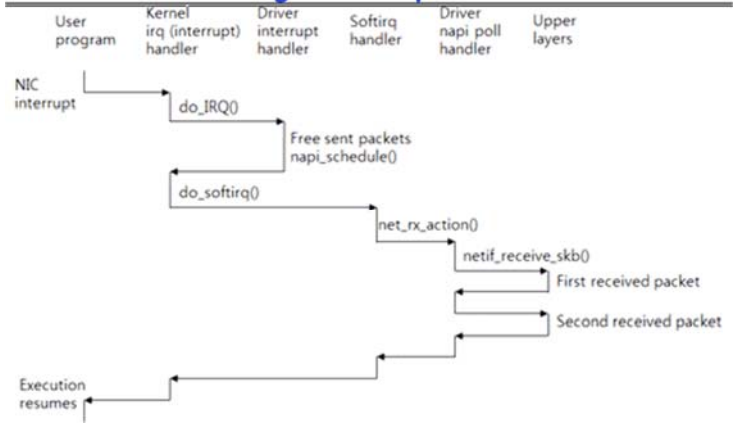


- Lots of `sk_buff` manipulation functions for:
  - removing and adding headers, merging data, pulling it up into linear region
  - Copying/cloning `sk_buff` structures

Network Processing Contexts



Avoiding Interrupts: NAPI

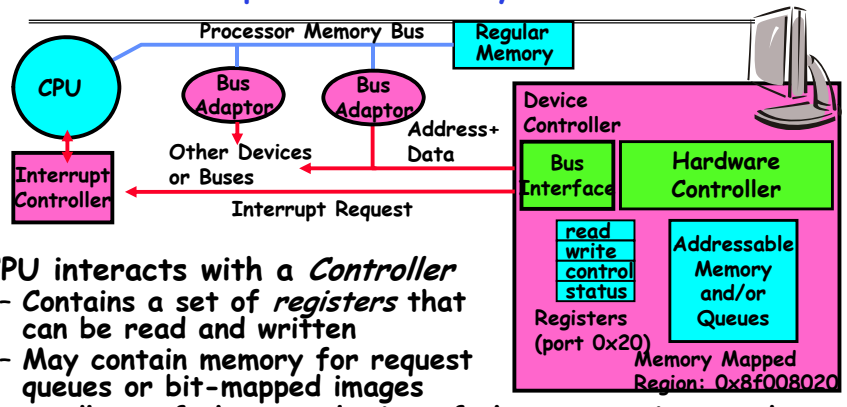


- New API (NAPI): Use polling to receive packets
  - Only some drivers actually implement this
- Exit hard interrupt context as quickly as possible
  - Do housekeeping and free up sent packets
  - Schedule soft interrupt for further actions
- Soft Interrupts: Handles receiption and delivery

Administrivia

- Get moving on Lab 4!
  - Should be well on your way to understanding the virtual device that you are designing…
- Final: Tuesday May 13th
  - 310 Soda Hall
  - 11:30—2:30
  - Bring calculator, 2 pages of hand-written notes
- Don't forget final Lecture during RRR
  - Next Monday. Send me final topics!
  - I don't really have a lot of topics yet!
  - Right now I could talk about:
    - » Mobile Operating Systems (iOS/Android)
    - » Talk about Swarm Lab

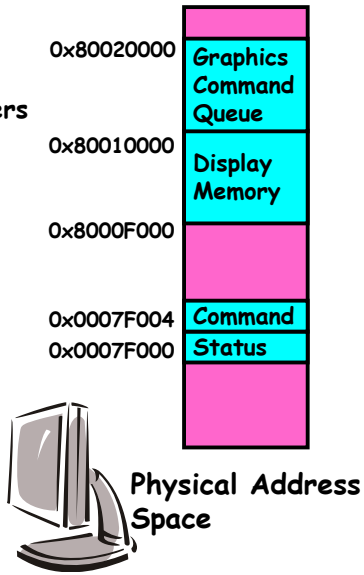
Recall: How does processor actually talk to the device?



- CPU interacts with a *Controller*
  - Contains a set of *registers* that can be read and written
  - May contain memory for request queues or bit-mapped images
- Regardless of the complexity of the connections and buses, processor accesses registers in two ways:
  - I/O instructions: in/out instructions
    - » Example from the Intel architecture: `out 0x21,AL`
  - Memory mapped I/O: load/store instructions
    - » Registers/memory appear in physical address space
    - » I/O accomplished with load and store instructions

Recall: Memory-Mapped Display Controller

- Memory-Mapped:
  - Hardware maps control registers and display memory into physical address space
    - » Addresses set by hardware jumpers or programming at boot time
  - Simply writing to display memory (also called the "frame buffer") changes image on screen
    - » Addr: 0x8000F000—0x8000FFFF
  - Writing graphics description to command-queue area
    - » Say enter a set of triangles that describe some scene
    - » Addr: 0x80010000—0x8001FFFF
  - Writing to the command register may cause on-board graphics hardware to do something
    - » Say render the above scene
    - » Addr: 0x0007F004
- Can protect with page tables



What about Protection?

- Start by asking some high-level questions…
  - What do we expect of our systems?
    - » Won't leak our information
    - » Won't lose our information
    - » Will always work when we need them
    - » Won't launch attacks against other people
  - How can we prevent systems from misbehaving?
    - » Never connect them to the network?
    - » Always authenticate users?
    - » Never use them?
- Protection: use of one or more mechanisms for controlling the access of programs, processes, or users to *resources*
  - Page Table Mechanism
  - File Access Mechanism
  - On-disk encryption
- Can use lots of Protection but still have an insecure system!
  - Bugs, back doors, viruses, poorly defined policy, inside man
  - Denial of service, …

Protection vs Security

- Security is a very complex topic: see, i.e. CS161
  - Security is about *Policy*, i.e. what human-centered properties do we want from our system
    - » Usually with reference to an attack model
  - Security is achieved through a series of *Mechanisms*, i.e. individual elements of the system combined together to achieve a security policy
- Security: use of protection mechanisms to prevent misuse of resources
  - Misuse defined with respect to policy
    - » E.g.: prevent exposure of certain sensitive information
    - » E.g.: prevent unauthorized modification/deletion of data
  - Requires consideration of the external environment within which the system operates
    - » Most well-constructed system cannot protect information if user accidentally reveals password

Preventing Misuse

- Types of Misuse:
  - Accidental:
    - » If I delete shell, can't log in to fix it!
    - » Could make it more difficult by asking: "do you really want to delete the shell?"
  - Intentional:
    - » Some high school brat who can't get a date, so instead he transfers $3 billion from B to A.
    - » Doesn't help to ask if they want to do it (of course!)
- Three Pieces to Security
  - Authentication: who the user actually is
  - Authorization: who is allowed to do what
  - Enforcement: make sure people do only what they are supposed to do
- Loopholes in any carefully constructed system:
  - Log in as superuser and you've circumvented authentication
  - Log in as self and can do anything with your resources; for instance: run program that erases all of your files
  - Can you trust software to correctly enforce Authentication and Authorization?????

Authentication: Identifying Users

- How to identify users to the system?
  - Passwords
    - » Shared secret between two parties
    - » Since only user knows password, someone types correct password ⇒ must be user typing it
    - » Very common technique
  - Smart Cards
    - » Electronics embedded in card capable of providing long passwords or satisfying challenge → response queries
    - » May have display to allow reading of password
    - » Or can be plugged in directly; several credit cards now in this category

  - Biometrics
    - » Use of one or more intrinsic physical or behavioral traits to identify someone
    - » Examples: fingerprint reader, palm reader, retinal scan
    - » Becoming quite a bit more common

- Two-factor authentication: use two or more types of authentication
- What else?
  - Consider the "Swarm" and "Un-pad" views

Timing Attacks: Tenex Password Checking

- Tenex – early 70's, BBN
  - Most popular system at universities before UNIX
  - Thought to be very secure, gave "red team" all the source code and documentation (want code to be publicly available, as in UNIX)
  - In 48 hours, they figured out how to get every password in the system
- Here's the code for the password check:

```
for (i = 0; i < 8; i++)
  if (userPasswd[i] != realPasswd[i])
    go to error
```

- How many combinations of passwords?
  - 256^8?
  - Wrong!

Defeating Password Checking

- Tenex used VM, and it interacts badly with the above code
  - Key idea: force page faults at inopportune times to break passwords quickly
- Arrange 1st char in string to be last char in pg, rest on next pg
  - Then arrange for pg with 1st char to be in memory, and rest to be on disk (e.g., ref lots of other pgs, then ref 1st page)

```
          a|aaaaaa
           |
page in memory| page on disk
```

- Time password check to determine if first character is correct!
  - If fast, 1st char is wrong
  - If slow, 1st char is right, pg fault, one of the others wrong
  - So try all first characters, until one is slow
  - Repeat with first two characters in memory, rest on disk
- Only 256 * 8 attempts to crack passwords
  - Fix is easy, don't stop until you look at all the characters

Authorization: Who Can Do What?

- How do we decide who is authorized to do actions in the system?
- Access Control Matrix: contains all permissions in the system
  - Resources across top
    - » Files, Devices, etc…
  - Domains in columns
    - » A domain might be a user or a group of permissions
    - » E.g. above: User D_3 can read F_2 or execute F_3
  - In practice, table would be huge and sparse!

domain \ object	F_1	F_2	F_3	printer
D_1	read		read	
D_2				print
D_3		read	execute	
D_4	read write		read write	

- Two approaches to implementation
  - Access Control Lists: store permissions with each object
    - » Still might be lots of users!
    - » UNIX limits each file to: r,w,x for owner, group, world
    - » More recent systems allow definition of groups of users and permissions for each group
  - Capability List: each process tracks objects has permission to touch
    - » Popular in the past, idea out of favor today
    - » Consider page table: Each process has list of pages it has access to, not each page has list of processes …

Authorization Continued

- Principle of least privilege: programs, users, and systems should get only enough privileges to perform their tasks
  - Very hard to manage in practice
    - » How do you figure out what the minimum set of privileges is needed to run your programs?
  - People often run at higher privilege then necessary
    - » Such as the "administrator" privilege under windows or "root" under Unix
- What form does this privilege take?
  - A set of Capabilities?
    - » Give a user the minimal set of possible access
    - » Like giving a minimal set of physical keys to someone
  - Hand-craft a special user for every task?
    - » Look in your password file – Linux does this all the time
    - » Custom users and groups for particular tasks

Enforcement

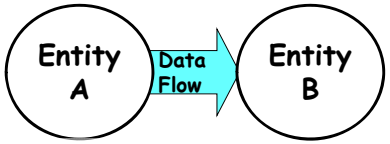
- Enforcer checks passwords, ACLs, etc
  - Makes sure the only authorized actions take place
  - Bugs in enforcer⇒things for malicious users to exploit
- Normally, in UNIX, superuser can do anything
  - Because of coarse-grained access control, lots of stuff has to run as superuser in order to work
  - If there is a bug in any one of these programs, you lose!
- Paradox
  - Bullet-proof enforcer
    - » Only known way is to make enforcer as small as possible
    - » Easier to make correct, but simple-minded protection model
  - Fancy protection
    - » Tries to adhere to principle of least privilege
    - » Really hard to get right
- Same argument for Java or C++: What do you make private vs public?
  - Hard to make sure that code is usable but only necessary modules are public
  - Pick something in middle? Get bugs and weak protection!

Mandatory Access Control (MAC)

- Mandatory Access Control (MAC)
  - "A Type of Access control by which the operating system constraints the ability of a *subject* or *initiator* to access or generally perform some sort of operation on an *object* or *target*."

    From Wikipedia
  - Subject: a process or thread
  - Object: files, directories, TCP/UDP ports, etc
  - Security policy is centrally controlled by a security policy administrator: users not allowed to operate outside the policy
  - Examples: SELinux, HiStar, etc.
- Contrast: Discretionary Access Control (DAC)
  - Access restricted based on the identity of subjects and/or groups to which they blong
  - Controls are discretionary – a subject with a certain access permission is capable of passing that permission on to any other subject
  - Standard UNIX model

Isolate Information Flow (HiStar)



Allow Communication if $L_A \subseteq L_B$

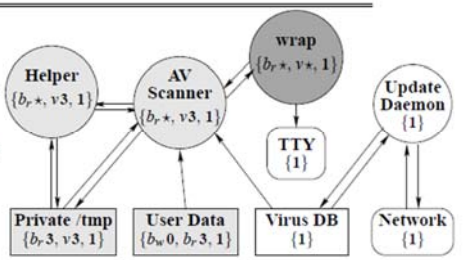
- Mandatory Access Control on Entities (Files, Processes, …)
  - Labels are sets of pairs of (Categories, Level):

    Lx={ (c1,l1), (c2,l2), … ldefault }

    - » Think of levels as a "security clearance" (Special declassification level "*")
    - » Can be compared, i.e. L1 ⊆ L2 if ∀h, L1(h) ≤ L2(h)
    - » "*" treated specially: lower than anything on left and higher than anything on right
  - Communication from A to B allowed only if $L_A \subseteq L_B$
    - » i.e. only if B's label has equivalent or higher clearance in every category than A's label

HiStar Virus Scanner Example

Level	Meaning in an object's label
⋆	has untainting privileges in this category
0	cannot be written/modified by default
1	default level—no restriction in this category
2	cannot be untainted/exported by default
3	cannot be read/observed by default



- Bob's Files Marked as {b_r3, b_w0, 1}
- User login for Bob creates process {b_r*, b_w*, 1}
  - Launches wrapper program which allocates v
- Wrapper launches scanner with taint v3
  - Temp directory marked {b_r3, v3, 1}
  - Can not write Bob's files, since less tainted (1) in category v than scanner is (which is 3)
  - Scanner can read from Virus DB, cannot write to anything except through wrapper program (which decides how to declassify information tagged with v

SELinux: Secure-Enhanced Linux

- SELinux: a Linux feature that provides the mechanisms for access control polices including MAC
  - A set of kernel modifications and user-space tools added to various Linux distributions
  - Separate enforcement of security decisions from policy
  - Integrated into mainline Linux kernel since version 2.6
- Originally started by the Information Assurance Research Group of the NSA, working with Secure Computing Corporation
- Security labels: tuple of role:user:domain
  - SELinux assigns a three string context consisting of a role, user name, and domain (or type) to every user and process
  - Files, network ports, and hardware also labeled with SELinux labels of name:role:type
  - Usually all real users share same Selinux user ("user_t")
- Policy
  - A set of rules specify which operations can be performed by an entity with a given label on an entity with a given label
  - Also, policy specifies which domain transitions can occur

SELinux Domain-type Enforcement

- Each object is labeled by a type
  - Object semantics
  - Example:
    - » /etc/shadow etc_t
    - » /etc/rc.d/init.d/httpd httpd_script_exec_t
- Objects are grouped by object security classes
  - Such as files, sockets, IPC channels, capabilities
  - The security class determines what operations can be performed on the object
- Each subject (process) is associated with a domain
  - E.g., httpd_t, sshd_t, sendmail_t

Example

- Execute the command "ls -Z /usr/bin/passwd"
  - This will produce the output:
    -r-s—x—x root root system_u:object_r:passwd_exec_t /usr/bin/passwd
  - Using this provided information, we can then create rules to have a domain transition.
- Three rules are required to give the user the ability to do a domain transition to the password file:
  - allow user_t passwd_exec_t : file {getattr execute};
    - » Lets user_t execute an execve() system call on passwd_exec_t
  - allow passwd_t passwd_exec_t : file entrypoint;
    - » This rule provides entrypoint access to the passwd_t domain, entrypoint defines which executable files can "enter" a domain.
  - allow user_t passwd_t : process transition;
    - » The original type (user_t) must have transition permission to the new type (passwd_t) for the domain transition to be allowed.

Limitations of the Type Enforcement Model

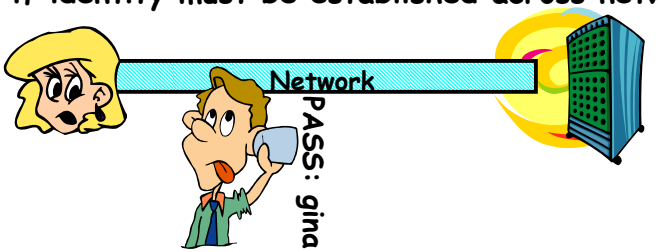
- Result in very large policies
  - Hundreds of thousands of rules for Linux
  - Difficult to understood
- Using only programs, but not information flow tracking cannot protect against certain attacks
  - Consider for example: httpd -> shell -> load kernel module

Data Centric Access Control (DCAC?)

- Problem with many current models:
  - If you break into OS ⇒ data is compromised
  - In reality, it is the *data* that matters – hardware is somewhat irrelevant (and ubiquitous)
- Data-Centric Access Control (DCAC)
  - I just made this term up, but you get the idea
  - Protect data at all costs, assume that software might be compromised
  - Requires encryption and sandboxing techniques
  - If hardware (or virtual machine) has the right cryptographic keys, then data is released
- All of the previous authorization and enforcement mechanisms reduce to key distribution and protection
  - Never let decrypted data or keys outside sandbox
  - Examples: Use of TPM, virtual machine mechanisms

Recall: Authentication in Distributed Systems

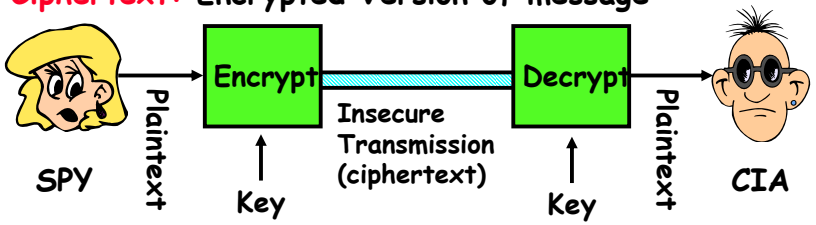
- What if identity must be established across network?



  - Need way to prevent exposure of information while still proving identity to remote system
  - Many of the original UNIX tools sent passwords over the wire "in clear text"
    - » E.g.: telnet, ftp, yp (yellow pages, for distributed login)
    - » Result: Snooping programs widespread
- What do we need? Cannot rely on physical security!
  - Encryption: Privacy, restrict receivers
  - Authentication: Remote Authenticity, restrict senders

Recall: Private Key Cryptography

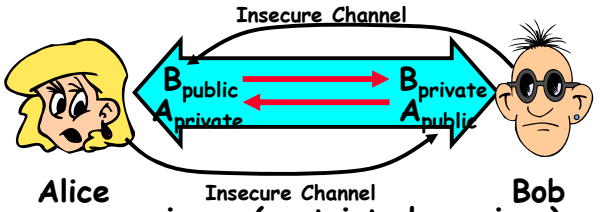
- Private Key (Symmetric) Encryption:
  - Single key used for both encryption and decryption
- Plaintext: Unencrypted Version of message
- Ciphertext: Encrypted Version of message



- Important properties
  - Can't derive plain text from ciphertext (decode) without access to key
  - Can't derive key from plain text and ciphertext
  - As long as password stays secret, get both secrecy and authentication
- Symmetric Key Algorithms: DES, Triple-DES, AES

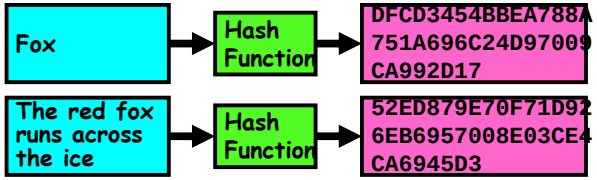
Recall: Public Key Encryption Details

- Idea: K_{public} can be made public, keep $K_{private}$ private



- Gives message privacy (restricted receiver):
  - Public keys (secure destination points) can be acquired by anyone/used by anyone
  - Only person with private key can decrypt message
- What about authentication?
  - Use combination of private and public key
  - Alice→Bob: $[(\text{I'm Alice})^{A\text{private}}$ Rest of message$]^{B\text{public}}$
  - Provides restricted sender and receiver
- But: how does Alice know that it was Bob who sent her B_{public}? And vice versa…

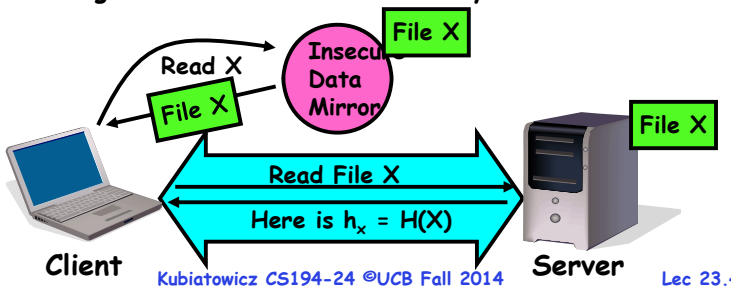
Recall: Secure Hash Function



- Hash Function: Short summary of data (message)
  - For instance, $h_1=H(M_1)$ is the hash of message M_1
    - » h_1 fixed length, despite size of message M_1.
    - » Often, h_1 is called the "digest" of M_1.
- Hash function H is considered secure if
  - It is infeasible to find M_2 with $h_1=H(M_2)$; ie. can't easily find other message with same digest as given message.
  - It is infeasible to locate two messages, m_1 and m_2, which "collide", i.e. for which $H(m_1) = H(m_2)$
  - A small change in a message changes many bits of digest/can't tell anything about message given its hash

Use of Hash Functions

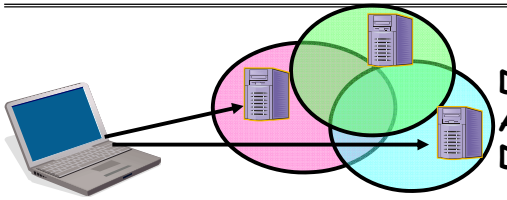
- Several Standard Hash Functions:
  - MD5: 128-bit output
  - SHA-1: 160-bit output, SHA-256: 256-bit output
- Can we use hashing to securely reduce load on server?
  - Yes. Use a series of insecure mirror servers (caches)
  - First, ask server for digest of desired file
    - » Use secure channel with server
  - Then ask mirror server for file
    - » Can be insecure channel
    - » Check digest of result and catch faulty or malicious mirrors



Signatures/Certificate Authorities

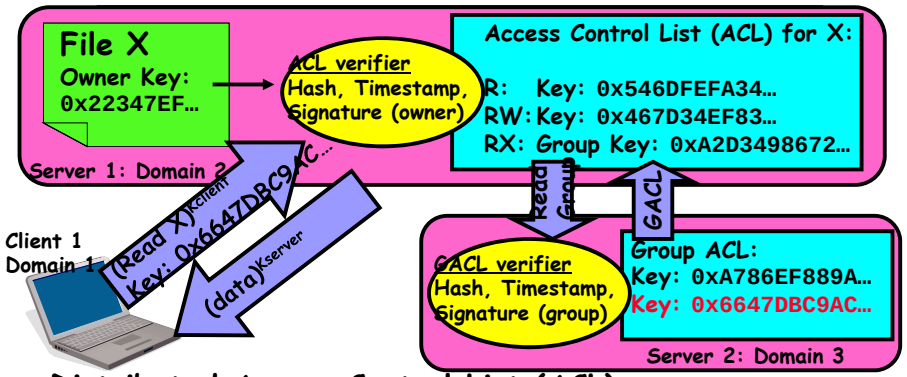
- Can use X_{public} for person X to define their identity
  - Presumably they are the only ones who know $X_{private}$.
  - Often, we think of X_{public} as a "principle" (user)
- Suppose we want X to sign message M?
  - Use private key to encrypt the digest, i.e. $H(M)^{Xprivate}$
  - Send both M and its signature:
    - » Signed message = $[M,H(M)^{Xprivate}]$
  - Now, anyone can verify that M was signed by X
    - » Simply decrypt the digest with X_{public}
    - » Verify that result matches H(M)
- Now: How do we know that the version of X_{public} that we have is really from X???
  - Answer: Certificate Authority
    - » Examples: Verisign, Entrust, Etc.
  - X goes to organization, presents identifying papers
    - » Organization signs X's key: $[X_{public}, H(X_{public})^{CAprivate}]$
    - » Called a "Certificate"
  - Before we use X_{public}, ask X for certificate verifying key
    - » Check that signature over X_{public} produced by trusted authority
- How do we get keys of certificate authority?
  - Compiled into your browser, for instance!

How to perform Authorization for Distributed Systems?



- Issues: Are all user names in world unique?
  - No! They only have small number of characters
    - » kubi@mit.edu → kubitron@lcs.mit.edu → kubitron@cs.berkeley.edu
    - » However, someone thought their friend was kubi@mit.edu and I got very private email intended for someone else…
  - Need something better, more unique to identify person
- Suppose want to connect with any server at any time?
  - Need an account on every machine! (possibly with different user name for each account)
  - OR: Need to use something more universal as identity
    - » Public Keys! (Called "Principles")
    - » People *are* their public keys

Distributed Access Control



- Distributed Access Control List (ACL)
  - Contains list of attributes (Read, Write, Execute, etc) with attached identities (Here, we show public keys)
    - » ACLs signed by owner of file, only changeable by owner
    - » Group lists signed by group key
  - ACLs can be on different servers than data
    - » Signatures allow us to validate them
    - » ACLs could even be stored separately from verifiers

Analysis of Previous Scheme

- Positive Points:
  - Identities checked via signatures and public keys
    - » Client can't generate request for data unless they have private key to go with their public identity
    - » Server won't use ACLs not properly signed by owner of file
  - No problems with multiple domains, since identities designed to be cross-domain (public keys domain neutral)
- Revocation:
  - What if someone steals your private key?
    - » Need to walk through all ACLs with your key and change…!
    - » This is very expensive
  - Better to have unique string identifying you that people place into ACLs
    - » Then, ask Certificate Authority to give you a certificate matching unique string to your current public key
    - » Client Request: (request + unique ID)^{Cprivate}; give server certificate if they ask for it.
    - » Key compromise⇒must distribute "certificate revocation", since can't wait for previous certificate to expire.
  - What if you remove someone from ACL of a given file?
    - » If server caches old ACL, then person retains access!
    - » Here, cache inconsistency leads to security violations!

Analysis Continued

- Who signs the data?
  - Or: How does client know they are getting valid data?
  - Signed by server?
    - » What if server compromised? Should client trust server?
  - Signed by owner of file?
    - » Better, but now only owner can update file!
    - » Pretty inconvenient!
  - Signed by group of servers that accepted latest update?
    - » If must have signatures from all servers ⇒ Safe, but one bad server can prevent update from happening
    - » Instead: ask for a threshold number of signatures
    - » Byzantine agreement can help here
- How do you know that data is up-to-date?
  - Valid signature only means data is valid older version
  - Freshness attack:
    - » Malicious server returns old data instead of recent data
    - » Problem with both ACLs and data
    - » E.g.: you just got a raise, but enemy breaks into a server and prevents payroll from seeing latest version of update
  - Hard problem
    - » Needs to be fixed by invalidating old copies or having a trusted group of servers (Byzantine Agrement?)

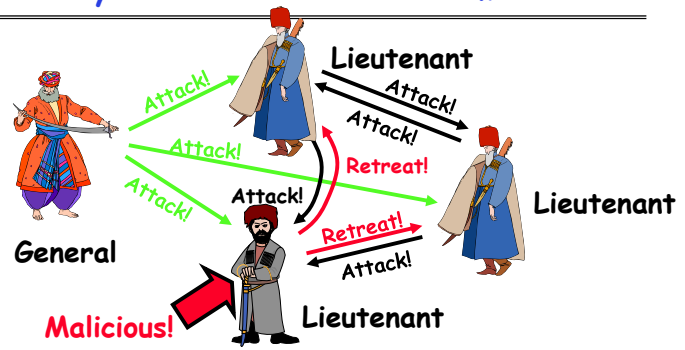
Distributed Decision Making

- Why is distributed decision making desirable?
  - Fault Tolerance!
  - Group of machines comes to decision even if one or more fail
    - » Simple failure mode called "failstop" (is this realistic?)
  - After decision made, result recorded in multiple places
- Two-Phase Commit protocol does this
  - Stable log on each machine tracks whether commit has happened
    - » If a machine crashes, when it wakes up it first checks its log to recover state of world at time of crash
  - Prepare Phase:
    - » The global coordinator requests that all participants will promise to commit or rollback the transaction
    - » Participants record promise in log, then acknowledge
    - » If anyone votes to abort, coordinator writes "Abort" in its log and tells everyone to abort; each records "Abort" in log
  - Commit Phase:
    - » After all participants respond that they are prepared, then the coordinator writes "Commit" to its log
    - » Then asks all nodes to commit; they respond with ack
    - » After receive acks, coordinator writes "Got Commit" to log
  - Log helps ensure all machines either commit or don't commit

Distributed Decision Making Discussion (Con't)

- Undesirable feature of Two-Phase Commit: Blocking
  - One machine can be stalled until another site recovers:
    - » Site B writes "prepared to commit" record to its log, sends a "yes" vote to the coordinator (site A) and crashes
    - » Site A crashes
    - » Site B wakes up, check its log, and realizes that it has voted "yes" on the update. It sends a message to site A asking what happened. At this point, B cannot decide to abort, because update may have committed
    - » B is blocked until A comes back
  - A blocked site holds resources (locks on updated items, pages pinned in memory, etc) until learns fate of update
- Alternative: There are alternatives such as "Three Phase Commit" which don't have this blocking problem
- What happens if one or more of the nodes is malicious?
  - Malicious: attempting to compromise the decision making

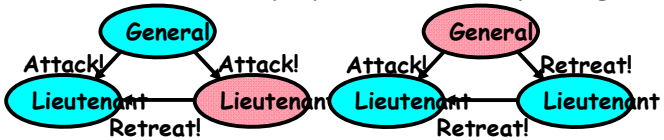
Byzantine General's Problem



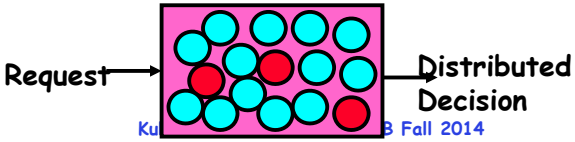
- Byzantine General's Problem (n players):
  - One General
  - n-1 Lieutenants
  - Some number of these (f) can be insane or malicious
- The commanding general must send an order to his n-1 lieutenants such that:
  - IC1: All loyal lieutenants obey the same order
  - IC2: If the commanding general is loyal, then all loyal lieutenants obey the order he sends

Byzantine General's Problem (con't)

- Impossibility Results:
  - Cannot solve Byzantine General's Problem with n=3 because one malicious player can mess up things



  - With f faults, need n > 3f to solve problem
- Various algorithms exist to solve problem
  - Original algorithm has #messages exponential in n
  - Newer algorithms have message complexity $O(n^2)$
    - » One from MIT, for instance (Castro and Liskov, 1999)
- Use of BFT (Byzantine Fault Tolerance) algorithm
  - Allow multiple machines to make a coordinated decision even if some subset of them (< n/3) are malicious



Trusted Computing

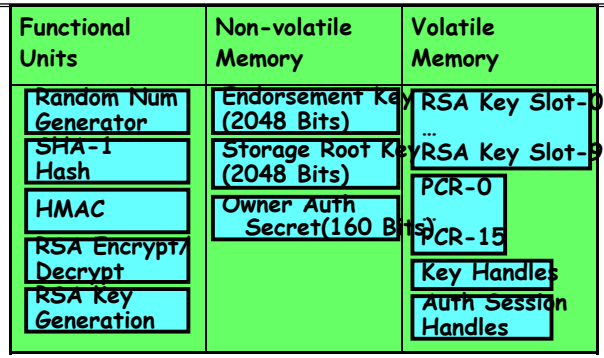
- Problem: Can't trust that software is correct
  - Viruses/Worms install themselves into kernel or system without users knowledge
  - Rootkit: software tools to conceal running processes, files or system data, which helps an intruder maintain access to a system without the user's knowledge
  - How do you know that software won't leak private information or further compromise user's access?
- A solution: What if there were a secure way to validate all software running on system?
  - Idea: Compute a cryptographic hash of BIOS, Kernel, crucial programs, etc.
  - Then, if hashes don't match, know have problem
- Further extension:
  - Secure attestation: ability to *prove* to a remote party that local machine is running correct software
  - Reason: allow remote user to avoid interacting with compromised system
- Challenge: How to do this in an unhackable way
  - Must have hardware components somewhere

TCPA: Trusted Computing Platform Alliance

- Idea: Add a Trusted Platform Module (TPM)
- Founded in 1999: Compaq, HP, IBM, Intel, Microsoft
- Currently more than 200 members
- Changes to platform
  - Extra: Trusted Platform Module (TPM)
  - Software changes: BIOS + OS
- Main properties
  - Secure bootstrap
  - Platform attestation
  - Protected storage
- Microsoft version:
  - Palladium
  - Note quite same: More extensive hardware/software system

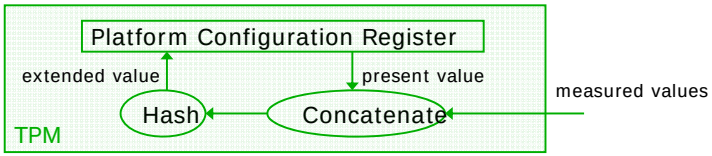
ATMEL TPM Chip
(Used in IBM equipment)

Trusted Platform Module

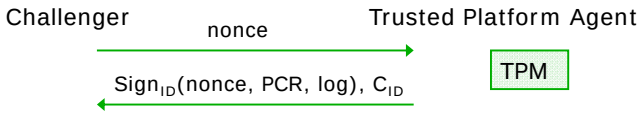


- Cryptographic operations
  - Hashing: SHA-1, HMAC
  - Random number generator
  - Asymmetric key generation: RSA (512, 1024, 2048)
  - Asymmetric encryption/ decryption: RSA
  - *Symmetric encryption/ decryption: DES, 3DES (AES)*
- Tamper resistant (hash and key) storage

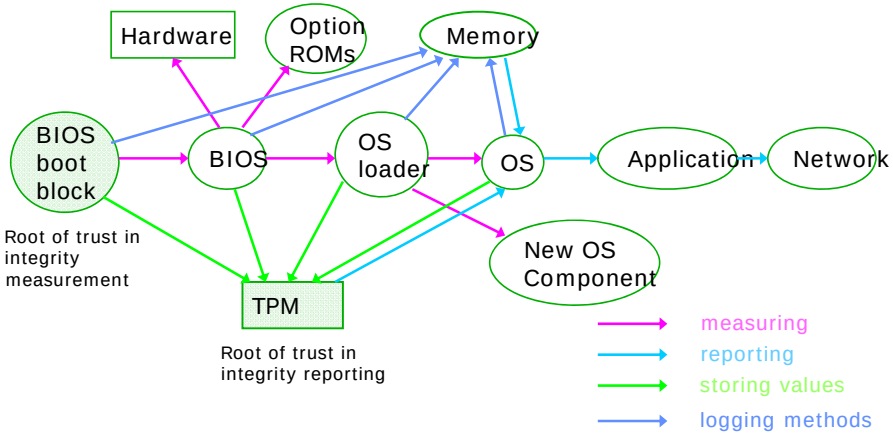
TCPA: PCR Reporting Value



- Platform Configuration Registers (PCR0-16)
  - Reset at boot time to well defined value
  - Only thing that software can do is give new measured value to TPM
    - » TPM takes new value, concatenates with old value, then hashes result together for new PCR
- Measuring involves hashing components of software
- Integrity reporting: report the value of the PCR
  - Challenge-response protocol:



TCPA: Secure bootstrap



Implications of TPM Philosophy?

- Could have great benefits
  - Prevent use of malicious software
  - Parts of OceanStore would benefit
- What does "trusted computing" really mean?
  - You are forced to trust hardware to be correct!
  - Could also mean that user is not trusted to install their own software
- Many in the security community have talked about potential abuses
  - These are only theoretical, but very possible
  - Software fixing
    - » What if companies prevent user from accessing their websites with non-Microsoft browser?
    - » Possible to encrypt data and only decrypt if software still matches ⇒ Could prevent display of .doc files except on Microsoft versions of software
  - Digital Rights Management (DRM):
    - » Prevent playing of music/video except on accepted players
    - » Selling of CDs that only play 3 times?

Summary

- Mandatory Access Control (MAC)
  - Separate access policy from use
  - Examples: HiStar, SELinux
- Distributed identity
  - Use cryptography (Public Key, Signed by PKI)
- Distributed storage example
  - Revocation: How to remove permissions from someone?
  - Integrity: How to know whether data is valid
  - Freshness: How to know whether data is recent
- Byzantine General's Problem: distributed decision making with malicious failures
  - One general, n-1 lieutenants: some number of them may be malicious (often "f" of them)
  - All non-malicious lieutenants must come to same decision
  - If general not malicious, lieutenants must follow general
  - Only solvable if $n \geq 3f+1$
- OceanStore: Distributed Storage in Untrusted World