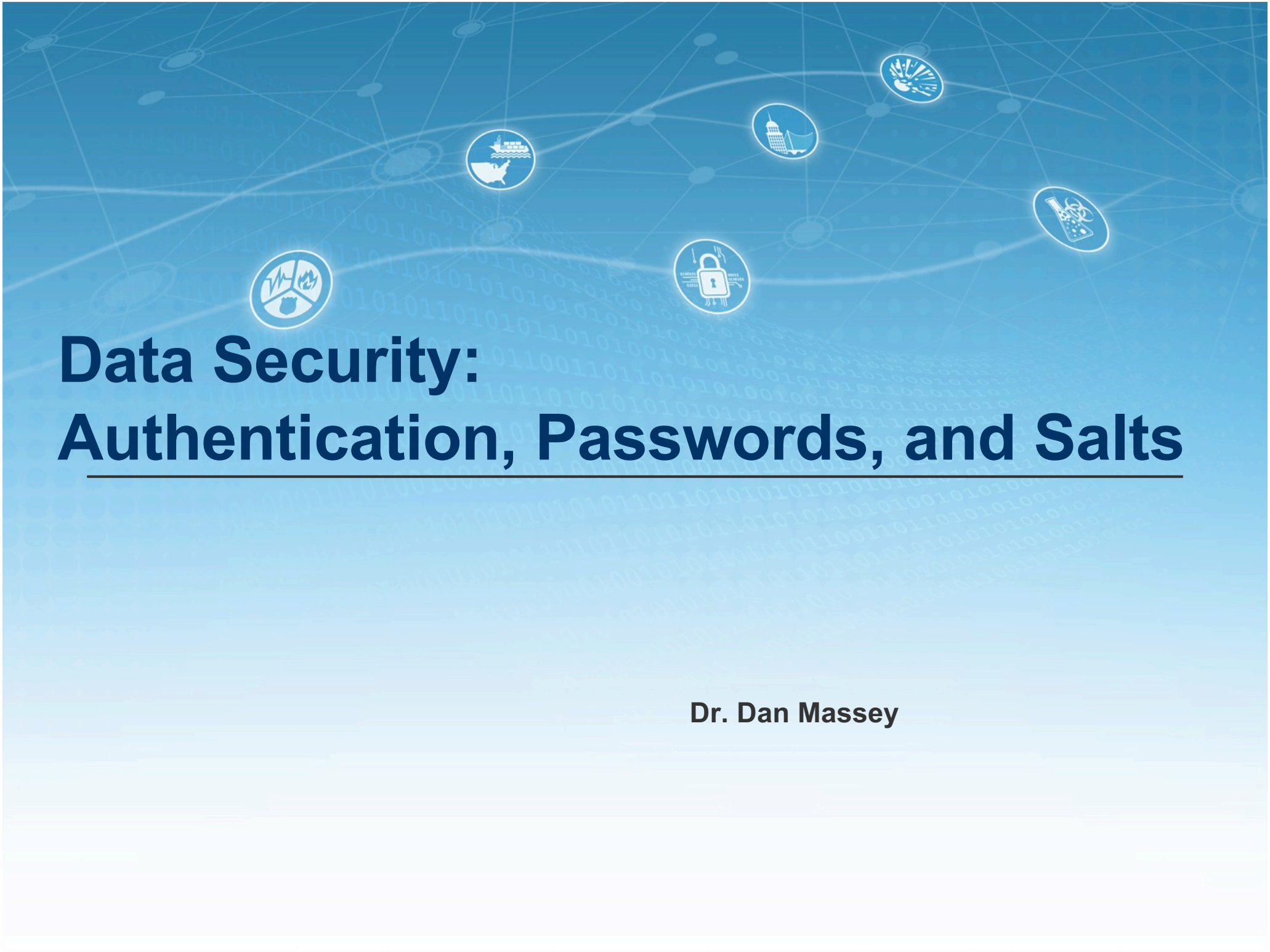


Data Security: Authentication, Passwords, and Salts

Dr. Dan Massey

Read Computer Security: Principle and Practices Chapter 3

NIST SP 800-63-3 (*Digital Authentication Guideline*, October 2016) defines digital user authentication as:

"The process of establishing confidence in user identities that are presented electronically to an information system."

Table 3.1 Identification and Authentication Security Requirements (SP 800-171)

	Basic Security Requirements:
1	Identify information system users, processes acting on behalf of users, or devices.
2	Authenticate (or verify) the identities of those users, processes, or devices, as a prerequisite to allowing access to organizational information systems.

	Derived Security Requirements:
3	Use multifactor authentication for local and network access to privileged accounts and for network access to non-privileged accounts.
4	Employ replay-resistant authentication mechanisms for network access to privileged and non-privileged accounts.
5	Prevent reuse of identifiers for a defined period.
6	Disable identifiers after a defined period of inactivity.
7	Enforce a minimum password complexity and change of characters when new passwords are created.
8	Prohibit password reuse for a specified number of generations.
9	Allow temporary password use for system logons with an immediate change to a permanent password.
10	Store and transmit only cryptographically-protected passwords.
11	Obscure feedback of authentication information.

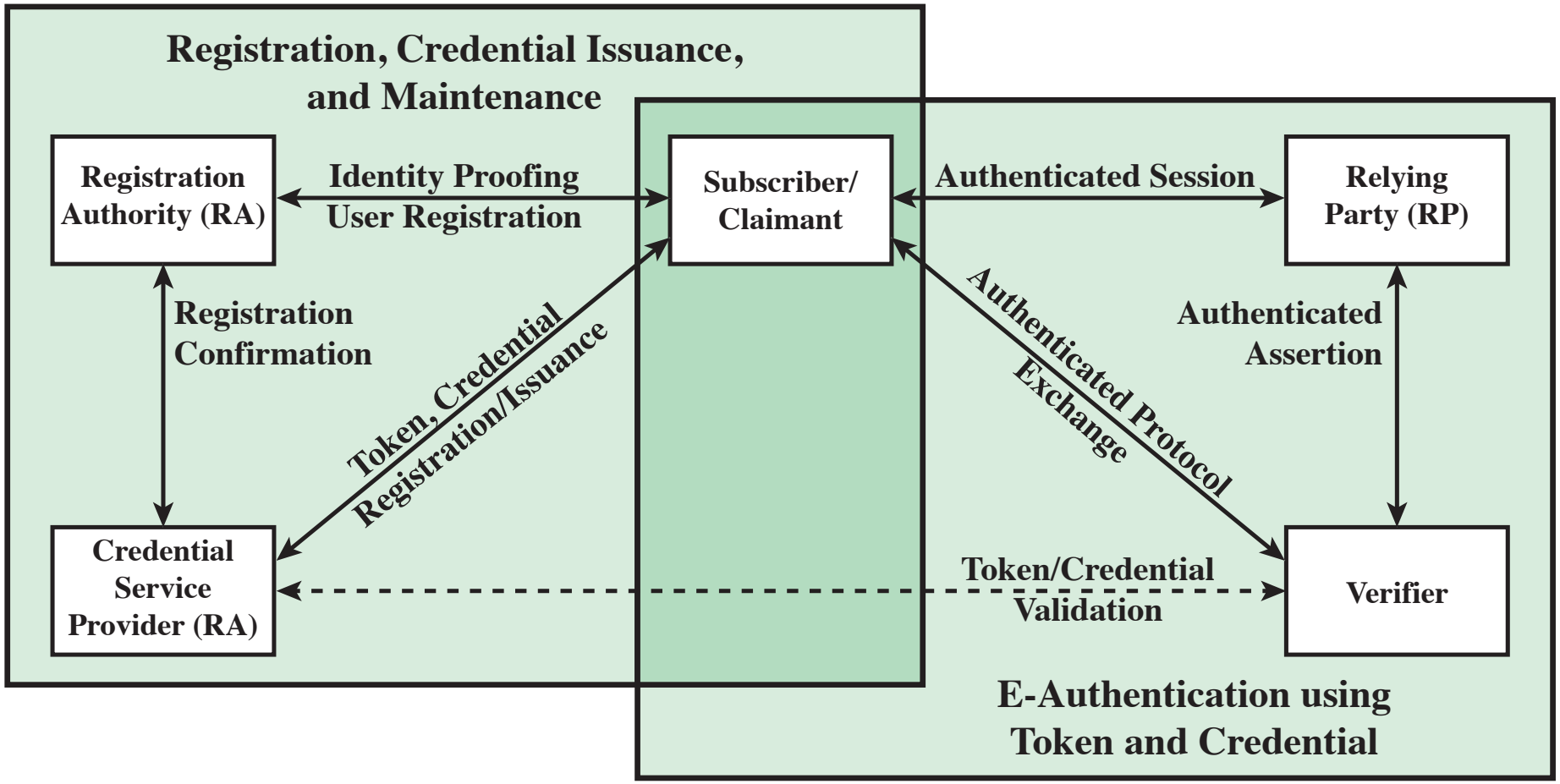


Figure 3.1 The NIST SP 800-63-2 E-Authentication Architectural Model

The four means of authenticating user identity are based on:

Something the individual knows	Something the individual possesses (token)	Something the individual is (static biometrics)	Something the individual does (dynamic biometrics)
• Password, PIN, answers to prearranged questions	• Smartcard, electronic keycard, physical key	• Fingerprint, retina, face	• Voice pattern, handwriting, typing rhythm

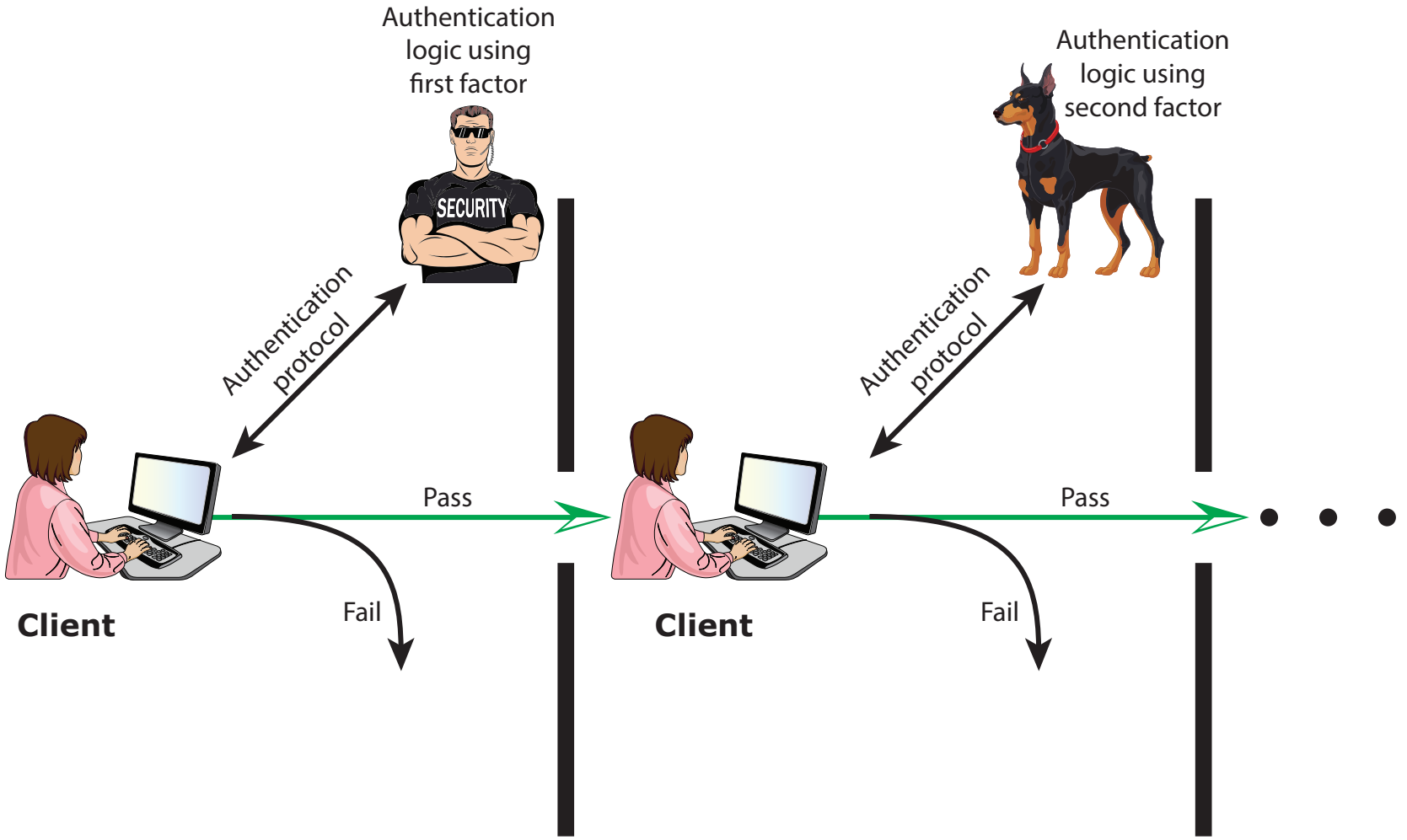
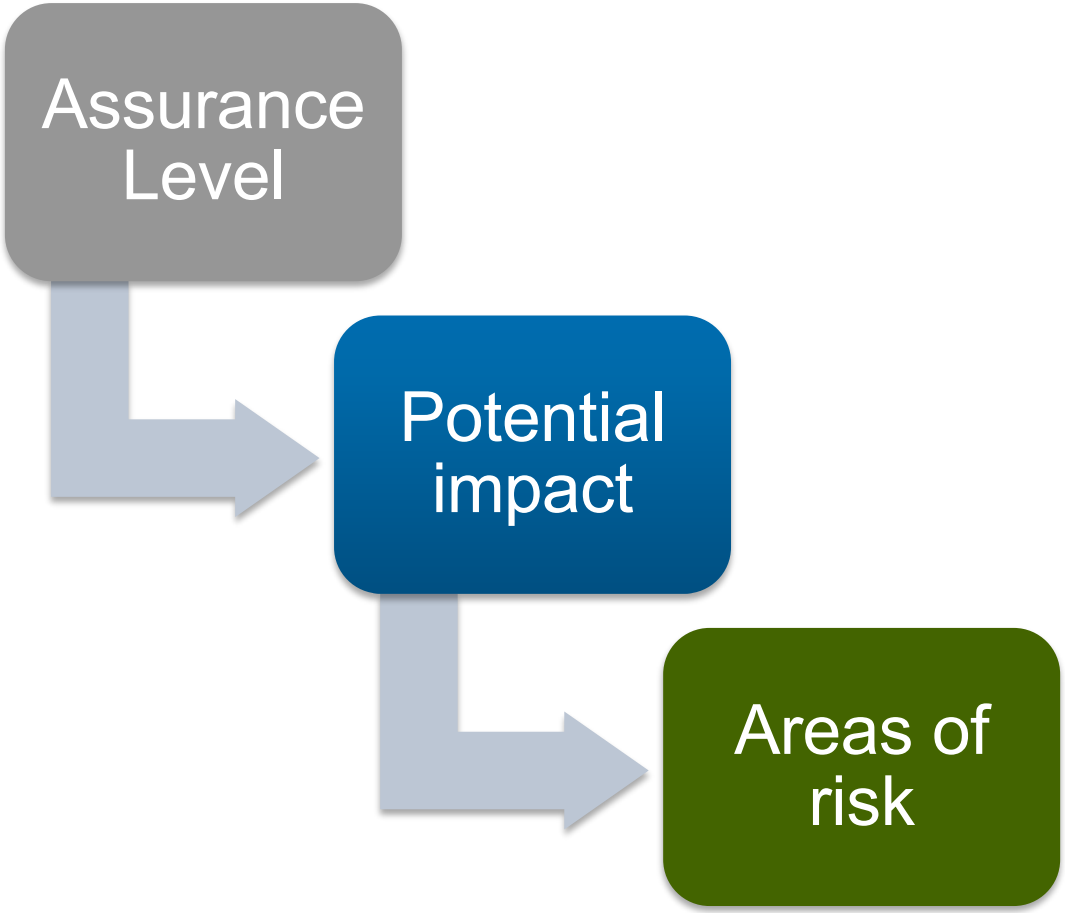


Figure 3.2 Multifactor Authentication

Risk Assessment for User Authentication

- There are three separate concepts:



Assurance Level

Describes an organization's degree of certainty that a user has presented a credential that refers to his or her identity

More specifically is defined as:

The degree of confidence in the vetting process used to establish the identity of the individual to whom the credential was issued

The degree of confidence that the individual who uses the credential is the individual to whom the credential was issued

Four levels of assurance

Level 1
- Little or no confidence in the asserted identity's validity

Level 2
- Some confidence in the asserted identity's validity

Level 3
- High confidence in the asserted identity's validity

Level 4
- Very high confidence in the asserted identity's validity

Potential Impact

- FIPS 199 defines three levels of potential impact on organizations or individuals should there be a breach of security:
  - Low
    - An authentication error could be expected to have a limited adverse effect on organizational operations, organizational assets, or individuals
  - Moderate
    - An authentication error could be expected to have a serious adverse effect
  - High
    - An authentication error could be expected to have a severe or catastrophic adverse effect

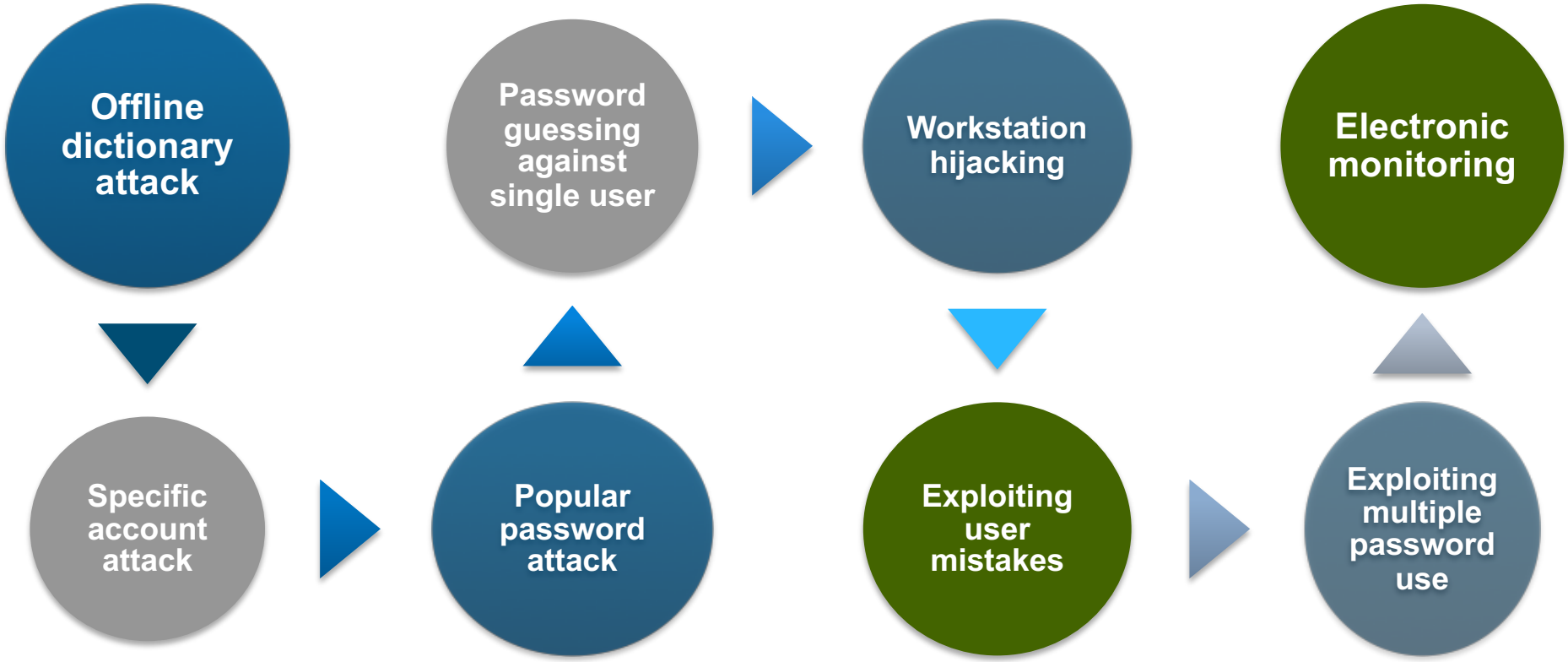
Password-Based Authentication

- Widely used line of defense against intruders
  - User provides name/login and password
  - System compares password with the one stored for that specified login
- The user ID:
  - Determines that the user is authorized to access the system
  - Determines the user's privileges
  - Is used in discretionary access control

Password Storage

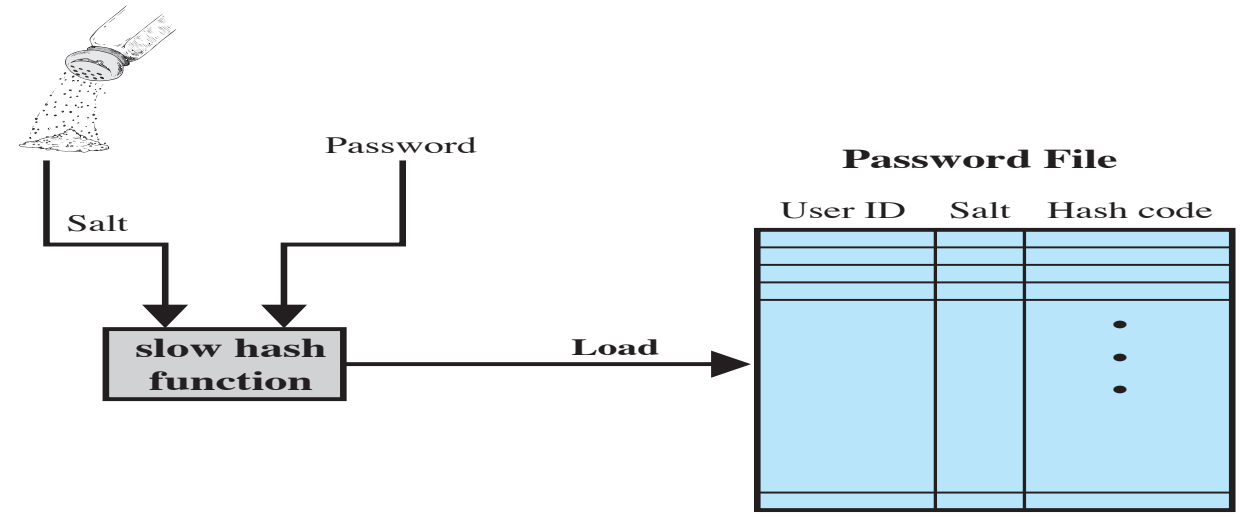
- Store as cleartext
  - If password file compromised, *all* passwords revealed
- Encipher file
  - Need to have decipherment, encipherment keys in memory
  - Reduces to previous problem
- Store one-way hash of password
  - If file read, attacker must still guess passwords or invert the hash

Password Vulnerabilities

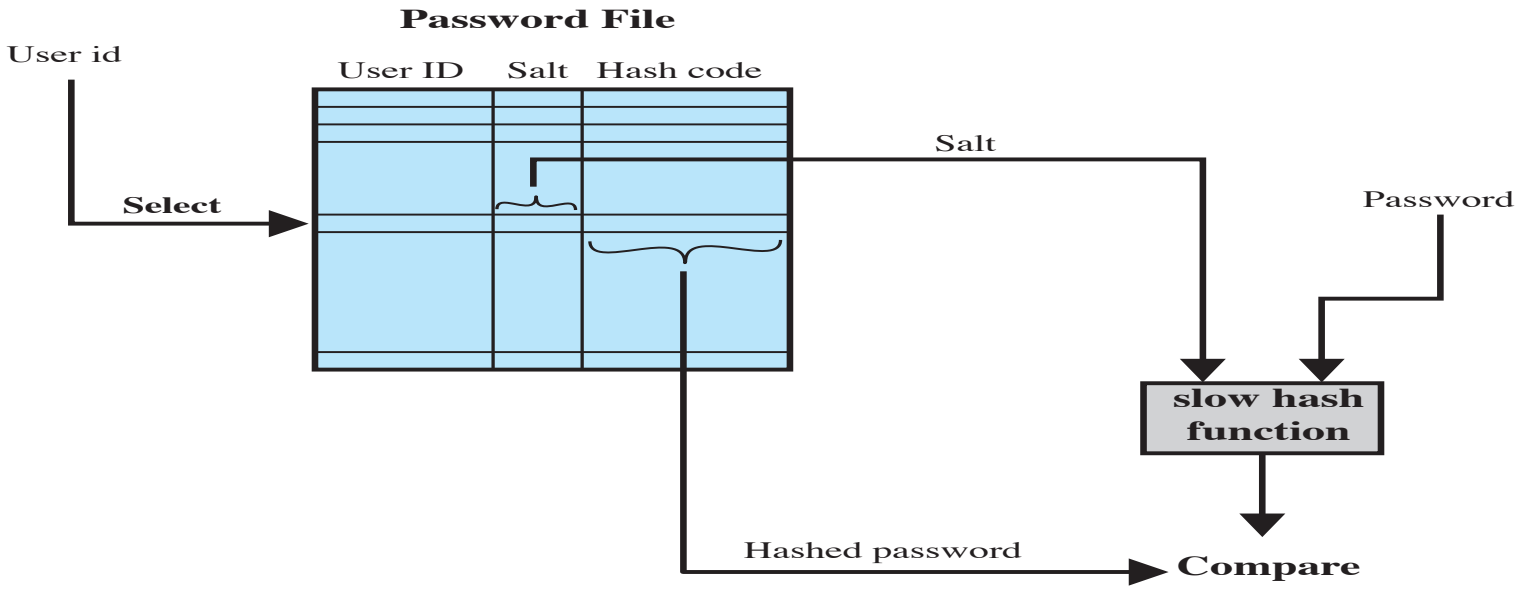


Motivations For Introducing Salts

- Trial-and-error from a list of potential passwords
  - *Off-line*: repeatedly try different guesses until the list is done or passwords guessed
    - Examples: *crack*, *john-the-ripper*
  - *On-line*: have access to login functions and try guesses *g* until some *password* succeeds
    - Examples: trying to log in by guessing a password
- Goal: Slow the above type of “Dictionary” attacks
- Method: Add a ”Salt” that perturbs hash function used to store the password so that:
  - Parameter controls *which* hash function is used
  - Parameter differs for each password
  - So given *n* password hashes, and therefore *n* salts, need to hash guess *n*



(a) Loading a new password



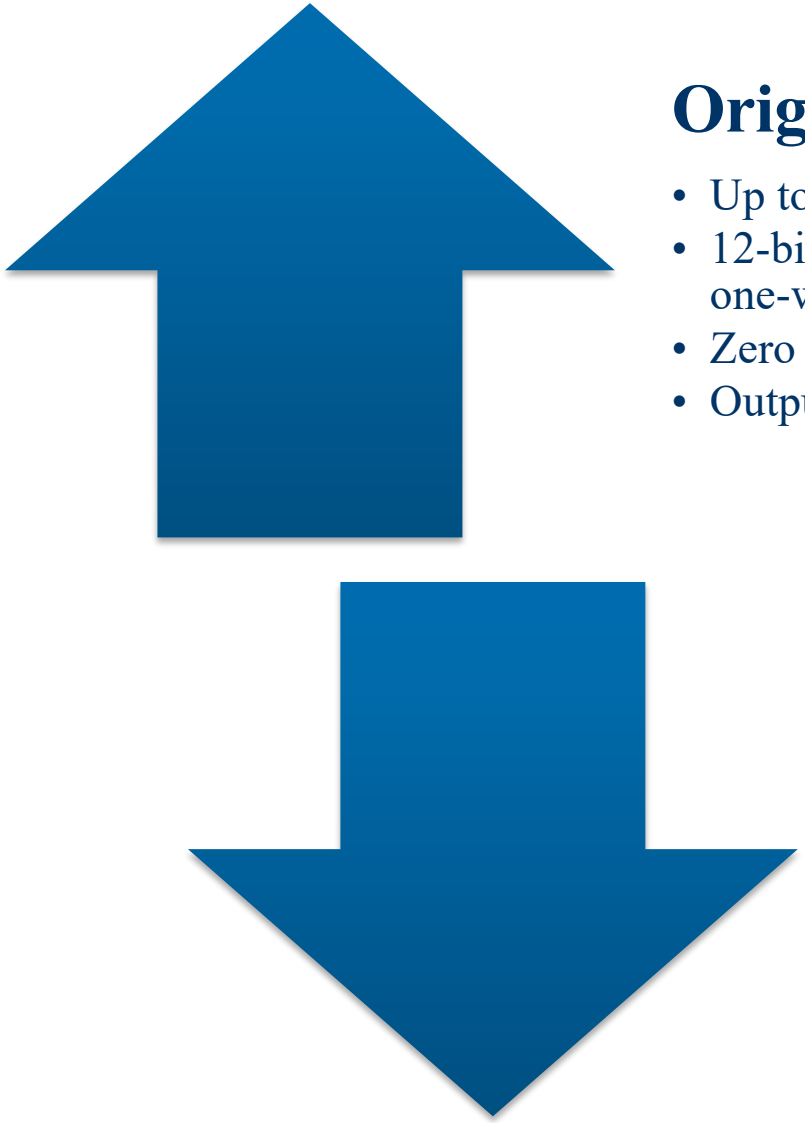
(b) Verifying a password

Figure 3.3 UNIX Password Scheme

Password Cracking and Salts

- Dictionary attacks
  - develop a large dictionary of possible passwords and try each against the password file
  - each password must be hashed using each salt value and then compared to stored hash values
- Rainbow Table Attacks
  - pre-compute tables of hash values ***for all salts***
  - a mammoth table of hash values
  - can be countered by using a sufficiently large salt value and a sufficiently large hash length

UNIX Implementation



Original scheme

- Up to eight printable characters in length
- 12-bit salt used to modify DES encryption into a one-way hash function
- Zero value repeatedly encrypted 25 times
- Output translated to 11 character sequence

Now regarded as inadequate

- Still often required for compatibility with existing account management software or multivendor environments

Improved Implementations

Much stronger hash/salt schemes available for Unix

Recommended hash function is based on MD5

- Salt of up to 48-bits
- Password length is unlimited
- Produces 128-bit hash
- Uses an inner loop with 1000 iterations to achieve slowdown

OpenBSD uses Blowfish block cipher based hash algorithm called Bcrypt

- Most secure version of Unix hash/salt scheme
- Uses 128-bit salt to create 192-bit hash value

Password Cracking

Dictionary attacks

- Develop a large dictionary of possible passwords and try each against the password file
- Each password must be hashed using each salt value and then compared to stored hash values

Rainbow table attacks

- Pre-compute tables of hash values for all salts
- A mammoth table of hash values
- Can be countered by using a sufficiently large salt value and a sufficiently large hash length

Password crackers exploit the fact that people choose easily guessable passwords

- Shorter password lengths are also easier to crack

John the Ripper

- Open-source password cracker first developed in in 1996
- Uses a combination of brute-force and dictionary techniques

Type of Password	Search Size	Number of Matches	Percentage of Passwords Matched	Cost/Benefit Ratio[a]
User/account name	130	368	2.7%	2.830
Character sequences	866	22	0.2%	0.025
Numbers	427	9	0.1%	0.021
Chinese	392	56	0.4%	0.143
Place names	628	82	0.6%	0.131
Common names	2239	548	4.0%	0.245
Female names	4280	161	1.2%	0.038
Male names	2866	140	1.0%	0.049
Uncommon names	4955	130	0.9%	0.026
Myths and legends	1246	66	0.5%	0.053
Shakespearean	473	11	0.1%	0.023
Sports terms	238	32	0.2%	0.134
Science fiction	691	59	0.4%	0.085
Movies and actors	99	12	0.1%	0.121
Cartoons	92	9	0.1%	0.098
Famous people	290	55	0.4%	0.190
Phrases and patterns	933	253	1.8%	0.271
Surnames	33	9	0.1%	0.273
Biology	58	1	0.0%	0.017
System dictionary	19683	1027	7.4%	0.052
Machine names	9018	132	1.0%	0.015
Mnemonics	14	2	0.0%	0.143
King James bible	7525	83	0.6%	0.011
Miscellaneous words	3212	54	0.4%	0.017
Yiddish words	56	0	0.0%	0.000
Asteroids	2407	19	0.1%	0.007
TOTAL	62727	3340	24.2%	0.053

Passwords Cracked from a Sample Set of 13,797 Accounts

***Computed as the number of matches divided by the search size. The more words that need to be tested for a match, the lower the cost/benefit ratio.**

Modern Approaches

- Complex password policy
  - Forcing users to pick stronger passwords
- However password-cracking techniques have also improved
  - The processing capacity available for password cracking has increased dramatically
  - The use of sophisticated algorithms to generate potential passwords
  - Studying examples and structures of actual passwords in use

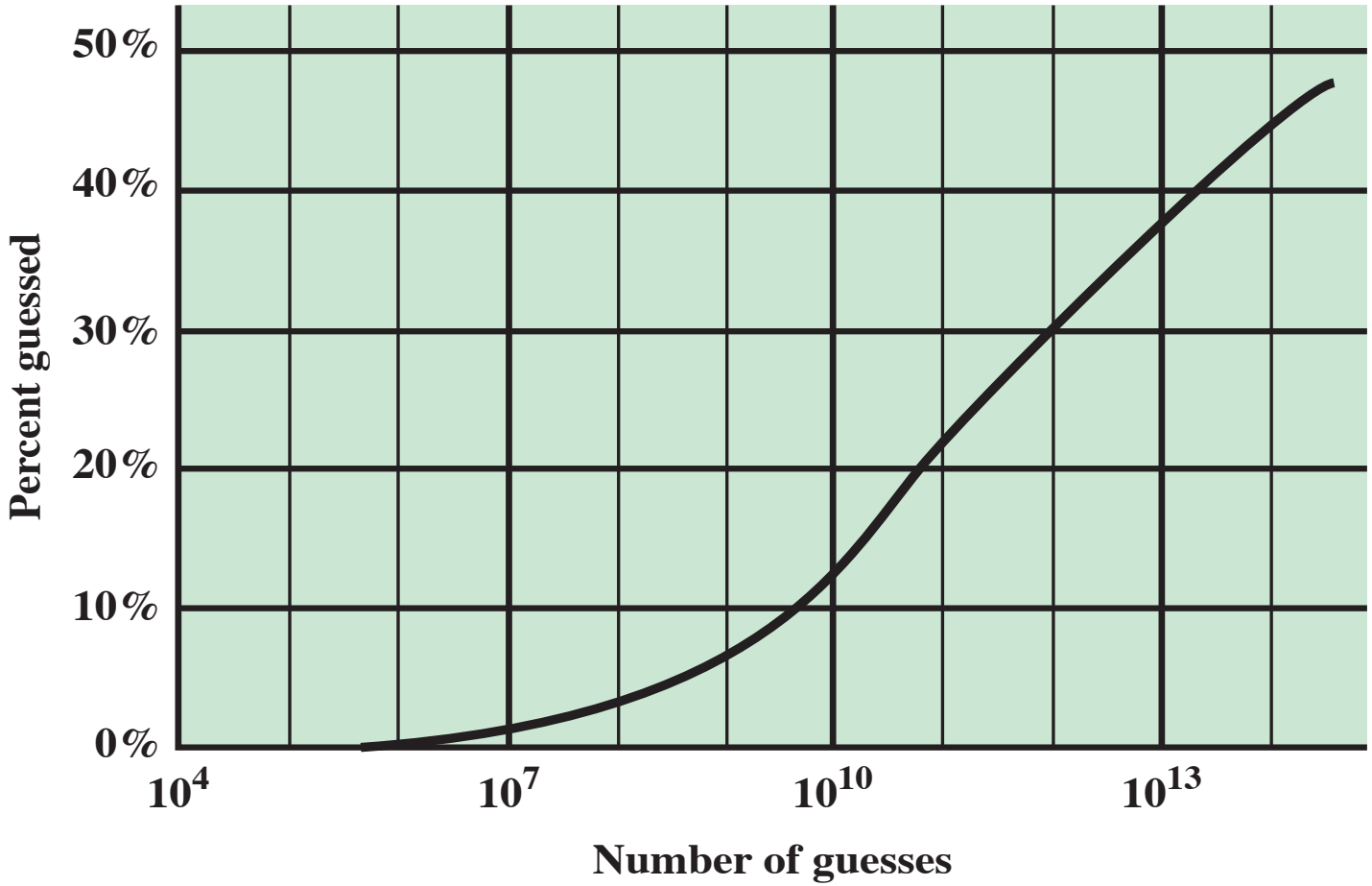


Figure 3.4 The Percentage of Passwords Guessed After a Given Number of Guesses

Password File Access Control

Can block offline guessing attacks by denying access to encrypted passwords

Make available only to privileged users

Shadow password file

Vulnerabilities

- Weakness in the OS that allows access to the file
- Accident with permissions making it readable
- Users with same password on other systems
- Access from backup media
- Sniff passwords in network traffic

User education

Users can be told the importance of using hard to guess passwords and can be provided with guidelines for selecting strong passwords

Computer generated passwords

Users have trouble remembering them

Reactive password checking

System periodically runs its own password cracker to find guessable passwords

Complex password policy

User is allowed to select their own password, however the system checks to see if the password is allowable, and if not, rejects it	Goal is to eliminate guessable passwords while allowing the user to select a password that is memorable

Password Checking

- Rules and Enforcement
  - Specify rules that passwords must adhere to
  - Easy to check size, password re-use, etc
  - How do you verify proposed password is not a dictionary word?
- Password checker
  - Build a dictionary of passwords not to use
  - Check password against each word
  - Potentially long and slow…
- A Better Way: Bloom filter
  - Used to build a table based on hash values
  - Check desired password against this table

Approximate set membership problem

- Suppose we have a set

  $S = \{s_1, s_2, \ldots, s_m\} \subseteq$ universe U

  S is the set of prohibited passwords

  U is the set of all possible passwords

- Represent S in such a way we can quickly answer "Is x an element of S ?"
- To take as little space as possible ,we allow false positive (i.e. $x \notin S$, but we answer yes)
- If $x \in S$, we must answer yes .

Bloom filters

- Consist of an arrays A[n] of n bits (space) , and k independent random hash functions

$h_1,\ldots,h_k : U \dashrightarrow \{0,1,..,n-1\}$

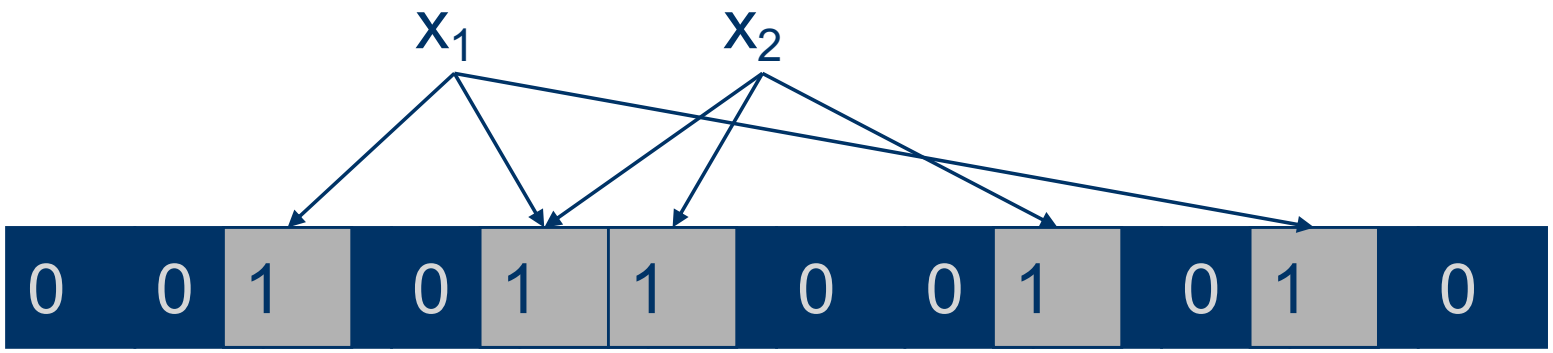
1. Initially set the array to 0
2. $\forall\ s \in S$, $A[h_i(s)] = 1$ for $1 \leq i \leq k$

   (an entry can be set to 1 multiple times, only the first times has an effect)
3. To check if $x \in S$, we check whether all location $A[h_i(x)]$ for $1 \leq i \leq k$ are set to 1

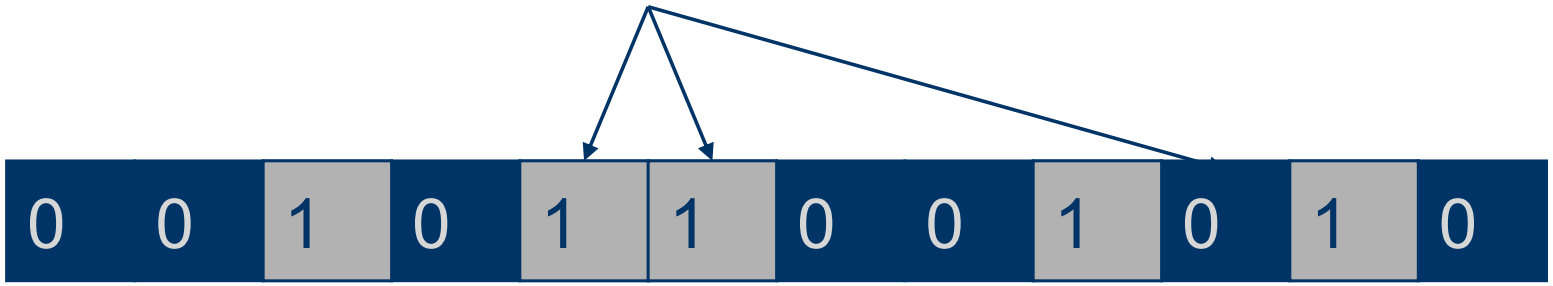
If not, clearly $x \notin S$.

If all $A[h_i(x)]$ are set to 1 ,we assume $x \in S$

Use members to create the bloom filter



Apply the same hash functions to ”y”
”y” is a member of the set only if all bits a set to 1.
Bit 9 is not set to 1 so y is not a member of the set.



Note potential for false positives: Suppose z (not shown in figure) is **NOT** member of the set and applying the hash functions to z yields bits 2, 4, and 5. X1 sets bits 2 & 4. X2 set bit 5.
Z will be ***incorrectly classified*** as a member of the set.

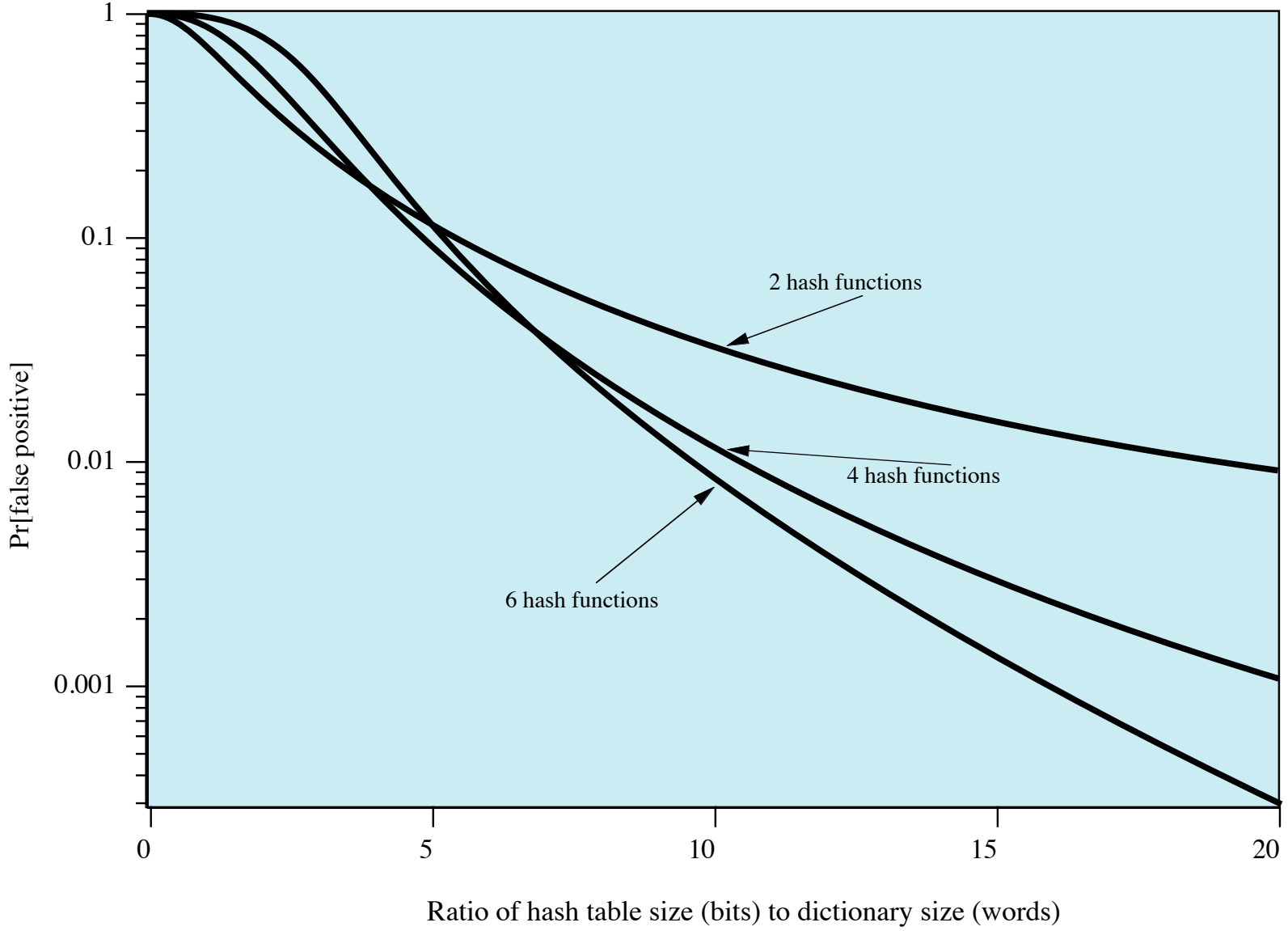


Figure 3.5 Performance of Bloom Filter

The four means of authenticating user identity are based on:

Something the individual knows

- Password, PIN, answers to prearranged questions

Something the individual possesses (token)

- Smartcard, electronic keycard, physical key

Something the individual is (static biometrics)

- Fingerprint, retina, face

Something the individual does (dynamic biometrics)

- Voice pattern, handwriting, typing rhythm

Types of Cards Used as Tokens

Card Type	Defining Feature	Example
Embossed	Raised characters only, on front	Old credit card
Magnetic stripe	Magnetic bar on back, characters on front	Bank card
Memory	Electronic memory inside	Prepaid phone card
Smart	Electronic memory and processor inside	Biometric ID card
Contact	Electrical contacts exposed on surface	
Contactless	Radio antenna embedded inside	

Memory Cards

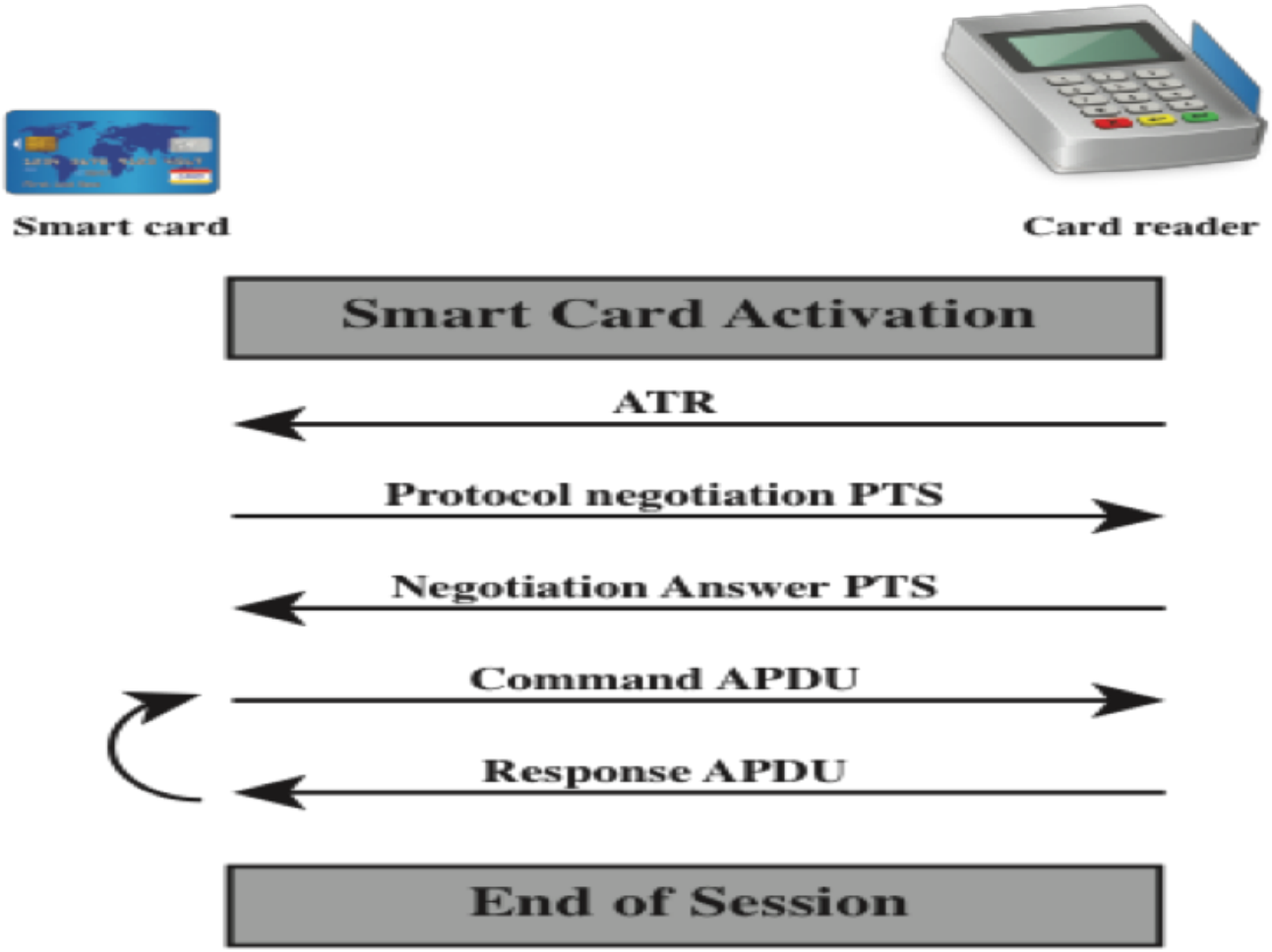
- Can store but do not process data
- The most common is the magnetic stripe card
- Can include an internal electronic memory
- Can be used alone for physical access
  - Hotel room
  - ATM
- Provides significantly greater security when combined with a password or PIN
- Drawbacks of memory cards include:
  - Requires a special reader
  - Loss of token
  - User dissatisfaction

Smart Tokens

- Physical characteristics:
  - Include an embedded microprocessor
  - A smart token that looks like a bank card
  - Can look like calculators, keys, small portable objects
- User interface:
  - Manual interfaces include a keypad and display for human/token interaction
- Electronic interface
  - A smart card or other token requires an electronic interface to communicate with a compatible reader/writer
  - Contact and contactless interfaces
- Authentication protocol:
  - Classified into three categories:
    - Static
    - Dynamic password generator
    - Challenge-response

Smart Cards

- Most important category of smart token
  - Has the appearance of a credit card
  - Has an electronic interface
  - May use any of the smart token protocols
- Contain:
  - An entire microprocessor
    - Processor
    - Memory
    - I/O ports
- Typically include three types of memory:
  - Read-only memory (ROM)
    - Stores data that does not change during the card's life
  - Electrically erasable programmable ROM (EEPROM)
    - Holds application data and programs
  - Random access memory (RAM)
    - Holds temporary data generated when applications are executed



APDU = application protocol data unit
ATR = Answer to reset
PTS = Protocol type selection

Figure 3.6 Smart Card/Reader Exchange

Electronic Identity Cards (eID)

Use of a smart card as a national identity card for citizens

Can serve the same purposes as other national ID cards, and similar cards such as a driver's license, for access to government and commercial services

Can provide stronger proof of identity and can be used in a wider variety of applications

In effect, is a smart card that has been verified by the national government as valid and authentic

Most advanced deployment is the German card *neuer Personalausweis*

Has human-readable data printed on its surface

- Personal data
- Document number
- Card access number (CAN)
- Machine readable zone (MRZ)

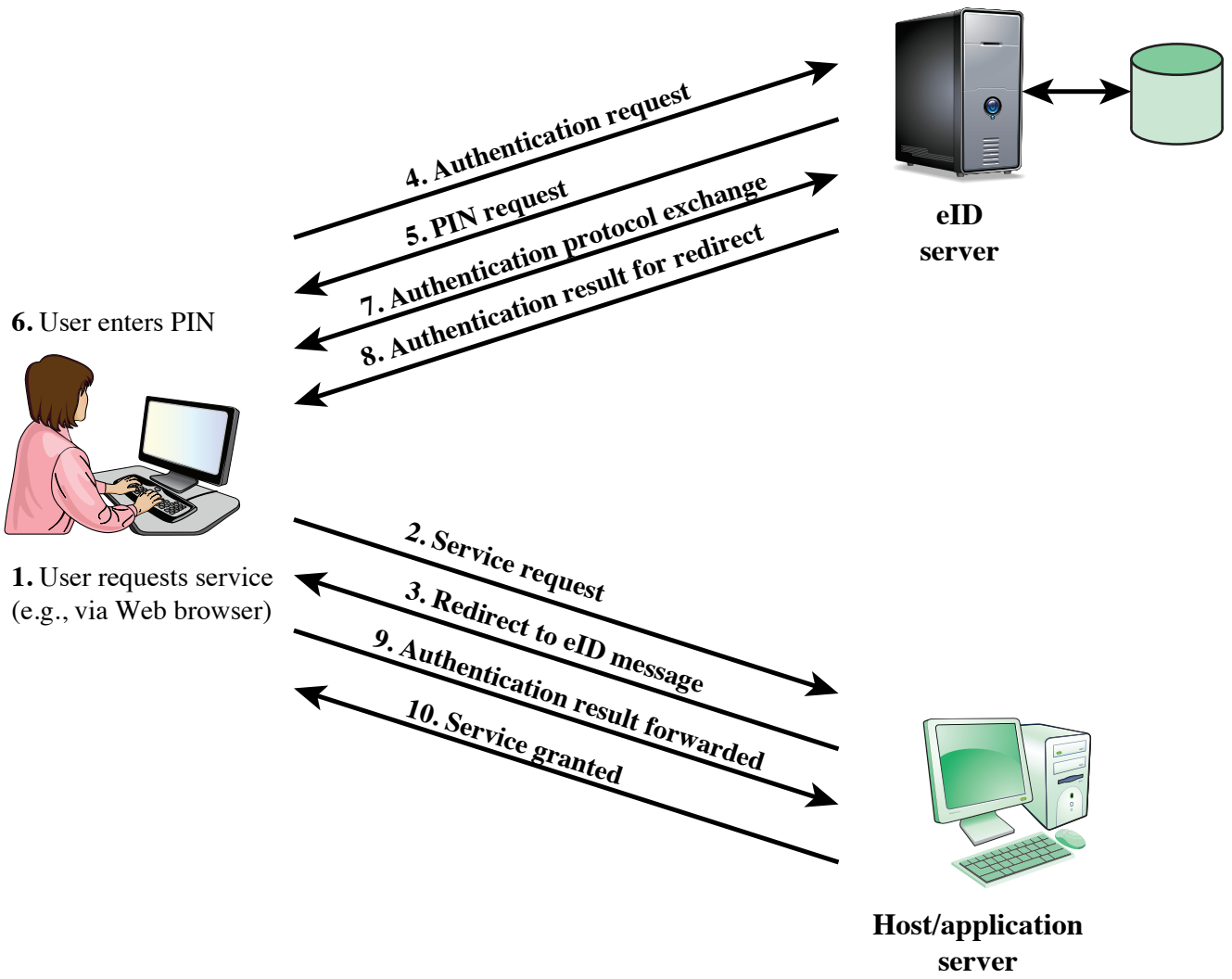


Figure 3.7 User Authentication with eID

The four means of authenticating user identity are based on:

Something the individual knows

- Password, PIN, answers to prearranged questions

Something the individual possesses (token)

- Smartcard, electronic keycard, physical key

Something the individual is (static biometrics)

- Fingerprint, retina, face

Something the individual does (dynamic biometrics)

- Voice pattern, handwriting, typing rhythm

Biometric Authentication

- Attempts to authenticate an individual based on unique physical characteristics
- Based on pattern recognition
- Is technically complex and expensive when compared to passwords and tokens
- Physical characteristics used include:
  - Facial characteristics
  - Fingerprints
  - Hand geometry
  - Retinal pattern
  - Iris
  - Signature
  - Voice

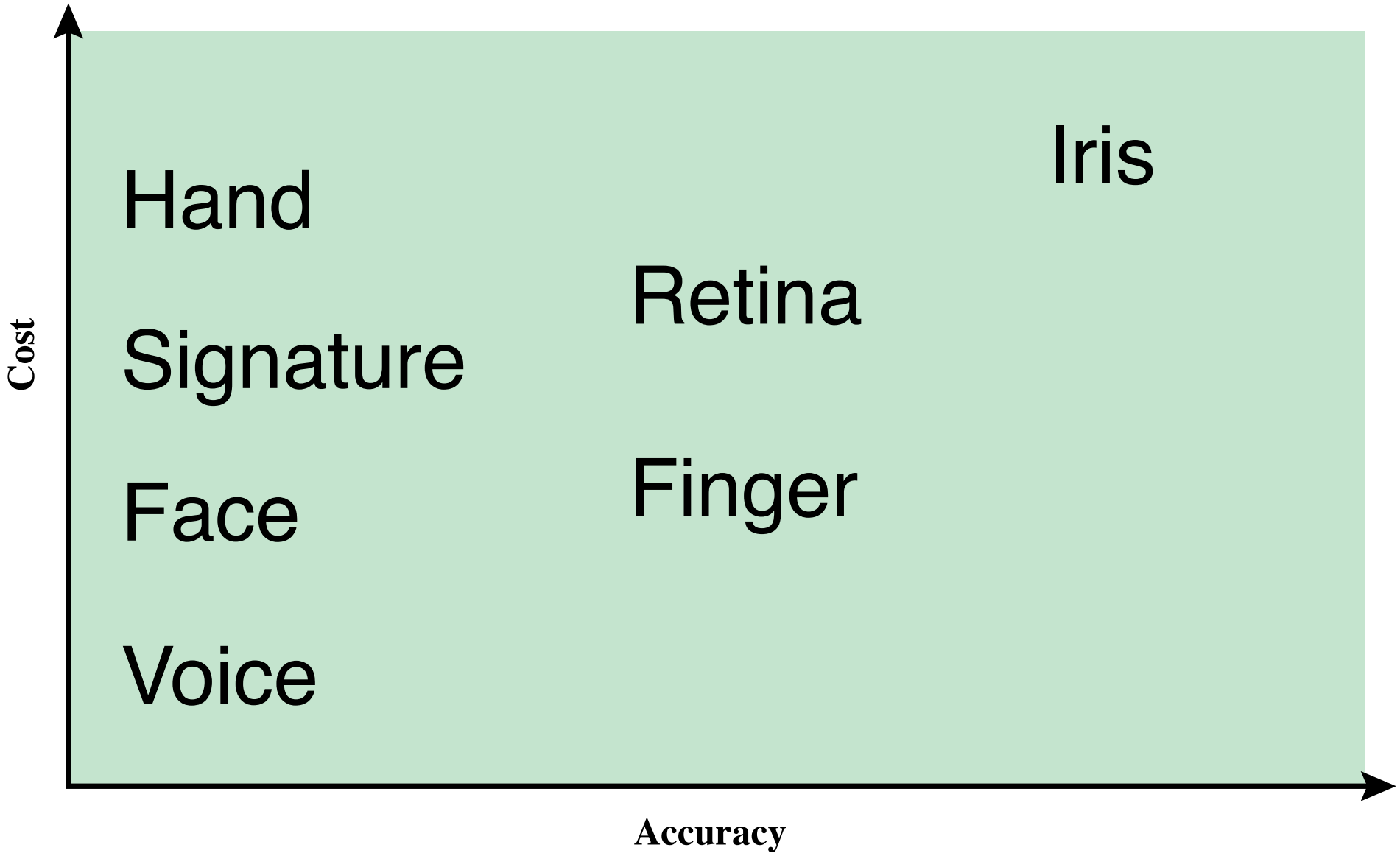


Figure 3.8 Cost Versus Accuracy of Various Biometric Characteristics in User Authentication Schemes.

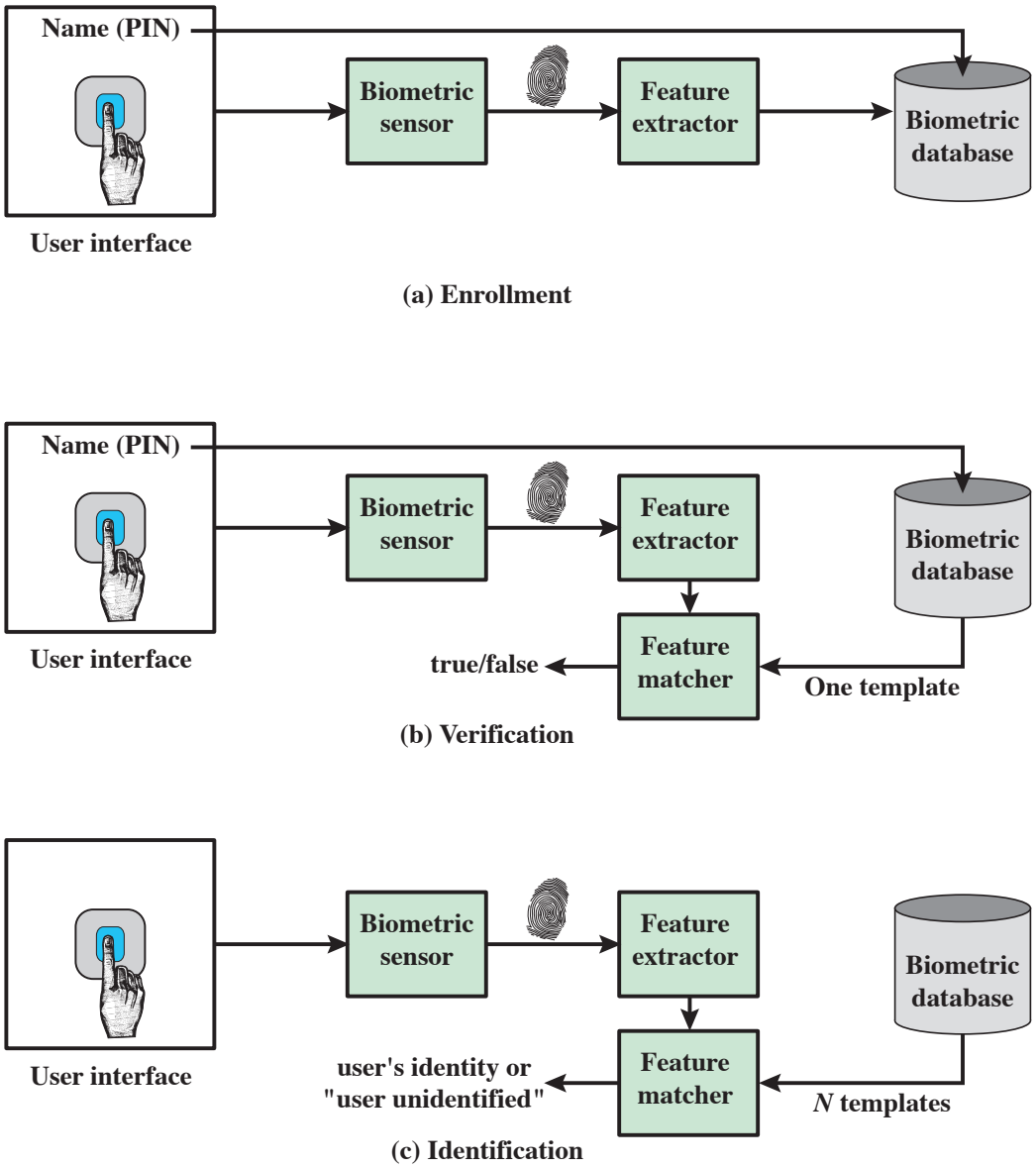


Figure 3.9 A Generic Biometric System. Enrollment creates an association between a user and the user's biometric characteristics. Depending on the application, user authentication either involves verifying that a claimed user is the actual user or identifying an unknown user.

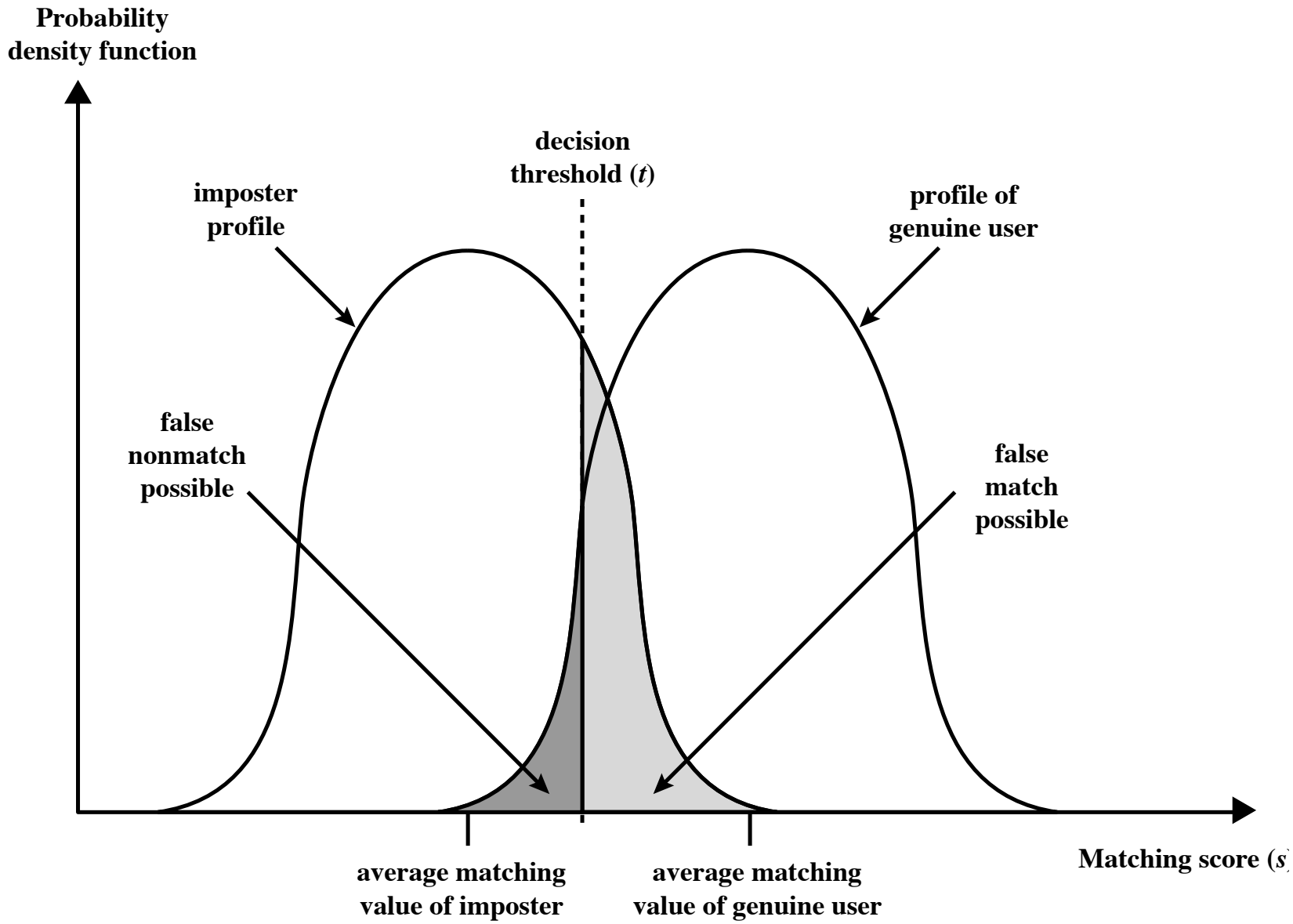


Figure 3.10 Profiles of a Biometric Characteristic of an Imposter and an Authorized Users In this depiction, the comparison between presented feature and a reference feature is reduced to a single numeric value. If the input value (s) is greater than a preassigned threshold (t), a match is declared.

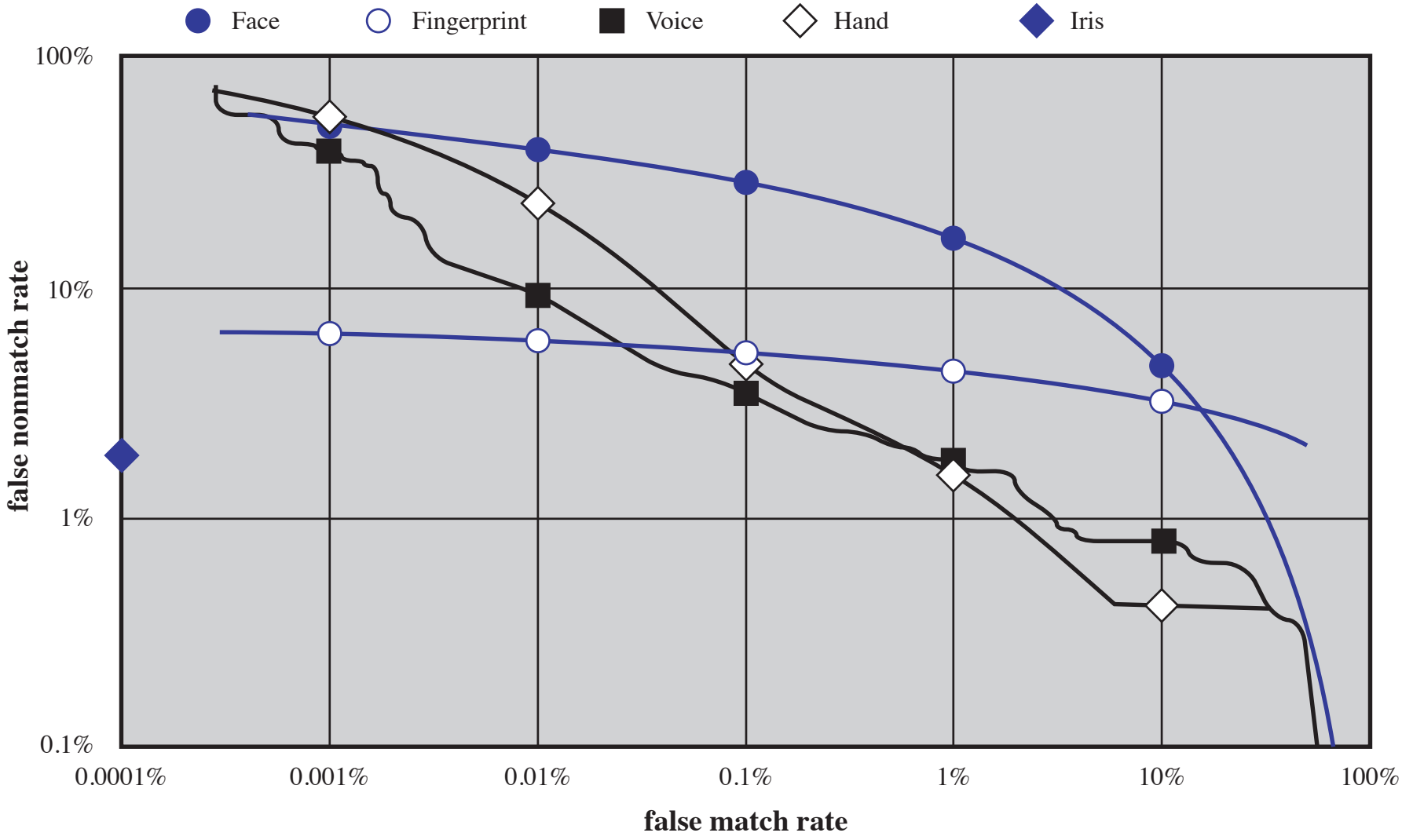


Figure 3.12 Actual Biometric Measurement Operating Characteristic Curves, reported in [MANS01]. To clarify differences among systems, a log-log scale is used.

Remote User Authentication

- Authentication over a network, the Internet, or a communications link is more complex
- Additional security threats such as:
  - Eavesdropping, capturing a password, replaying an authentication sequence that has been observed
- Generally rely on some form of a challenge-response protocol to counter threats

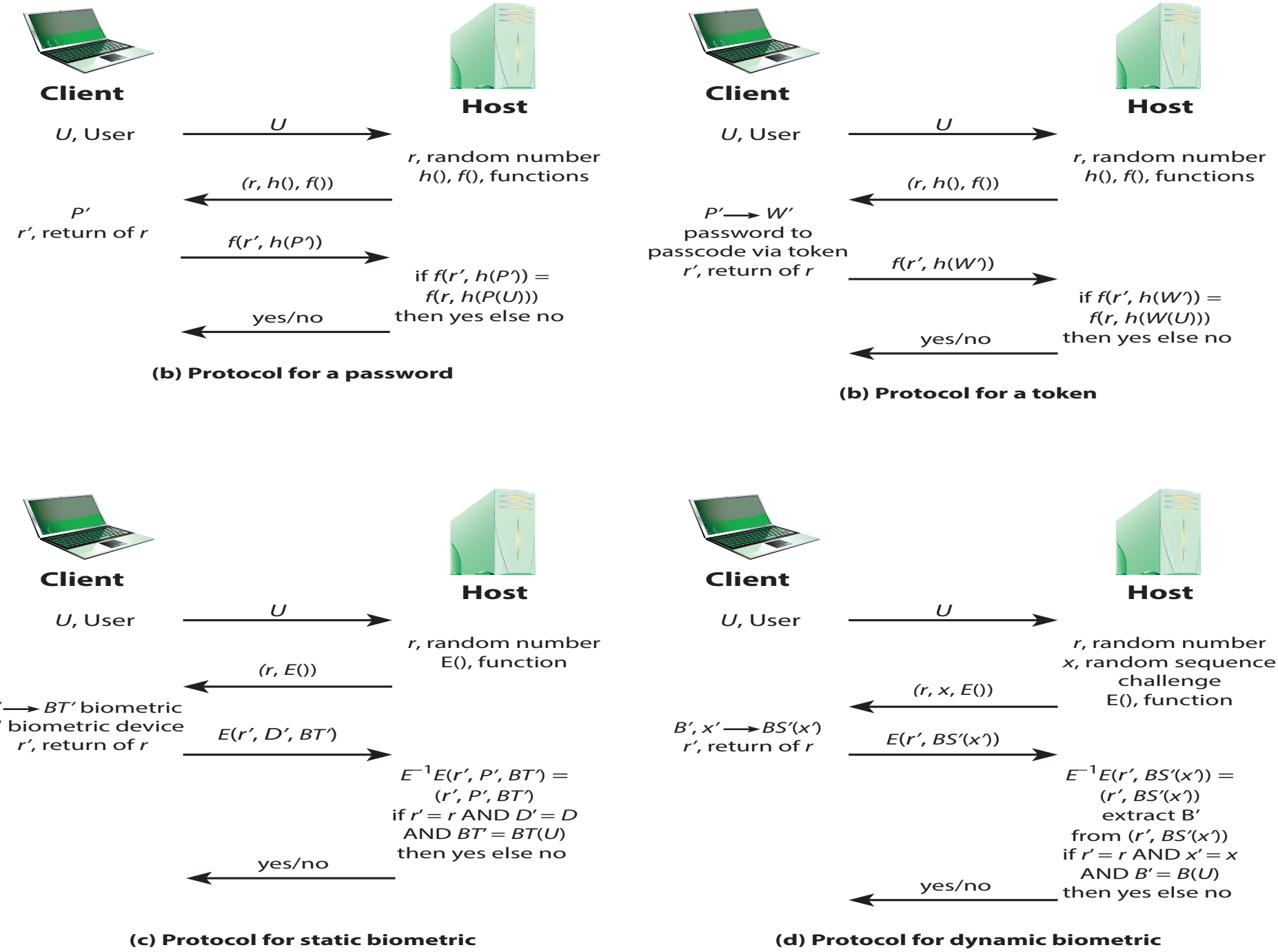
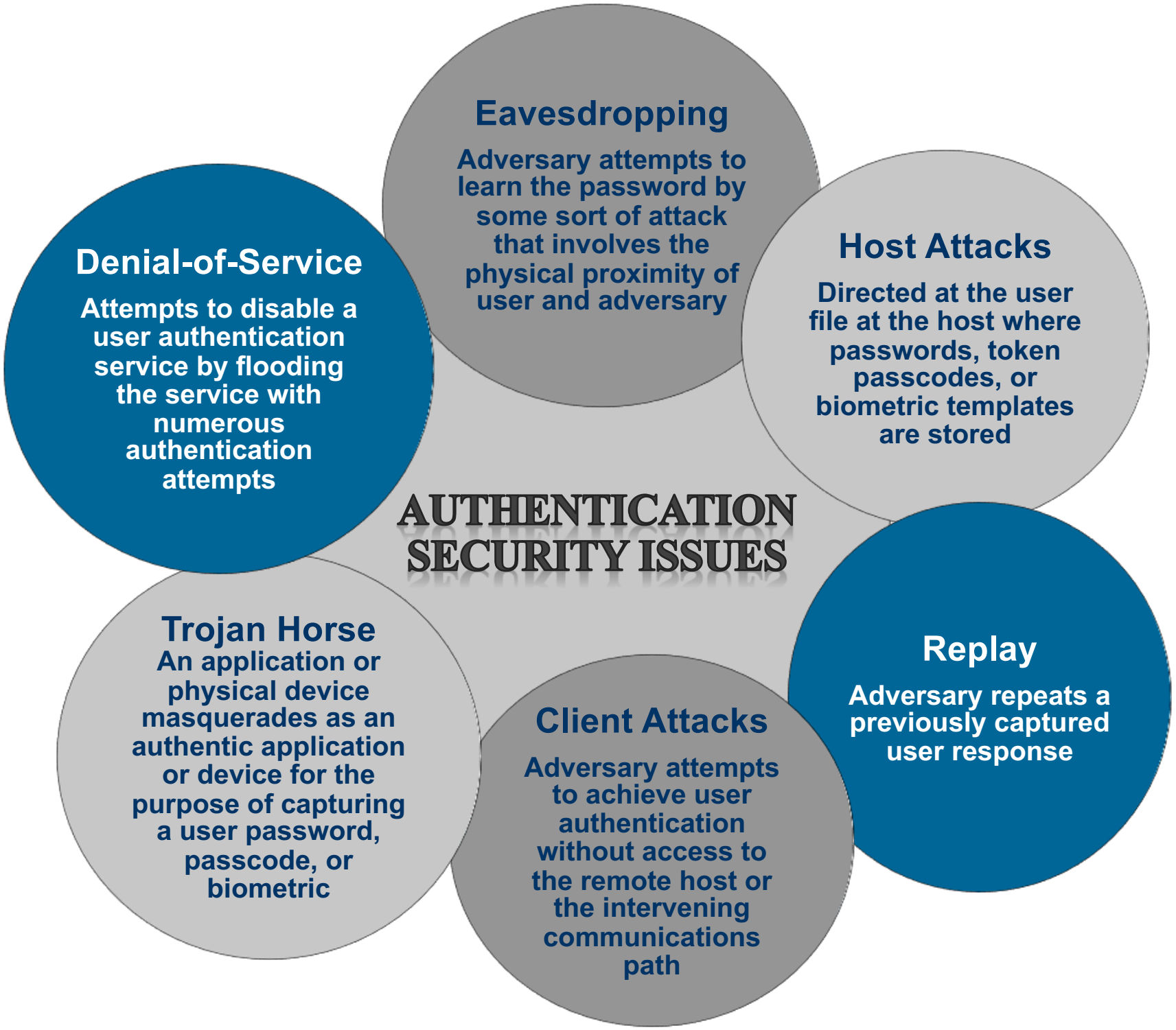


Figure 3.13 Basic Challenge-Response Protocols for Remote User Authentication


AUTHENTICATION SECURITY ISSUES
Eavesdropping
Adversary attempts to learn the password by some sort of attack that involves the physical proximity of user and adversary
Host Attacks
Directed at the user file at the host where passwords, token passcodes, or biometric templates are stored
Denial-of-Service
Attempts to disable a user authentication service by flooding the service with numerous authentication attempts
Replay
Adversary repeats a previously captured user response
Trojan Horse
An application or physical device masquerades as an authentic application or device for the purpose of capturing a user password, passcode, or biometric
Client Attacks
Adversary attempts to achieve user authentication without access to the remote host or the intervening communications path

Attacks	Authenticators	Examples	Typical defenses
Client attack	Password	Guessing, exhaustive search	Large entropy; limited attempts
	Token	Exhaustive search	Large entropy; limited attempts, theft of object requires presence
	Biometric	False match	Large entropy; limited attempts
Host attack	Password	Plaintext theft, dictionary/exhaustive search	Hashing; large entropy; protection of password database
	Token	Passcode theft	Same as password; 1-time passcode
	Biometric	Template theft	Capture device authentication; challenge response
Eavesdropping, theft, and copying	Password	"Shoulder surfing"	User diligence to keep secret; administrator diligence to quickly revoke compromised passwords; multifactor authentication
	Token	Theft, counterfeiting hardware	Multifactor authentication; tamper resistant/evident token
	Biometric	Copying (spoofing) biometric	Copy detection at capture device and capture device authentication
Replay	Password	Replay stolen password response	Challenge-response protocol
	Token	Replay stolen passcode response	Challenge-response protocol; 1-time passcode
	Biometric	Replay stolen biometric template response	Copy detection at capture device and capture device authentication via challenge-response protocol
Trojan horse	Password, token, biometric	Installation of rogue client or capture device	Authentication of client or capture device within trusted security perimeter
Denial of service	Password, token, biometric	Lockout by multiple failed authentications	Multifactor with token

Some Potential Attacks, Susceptible Authenticators, and Typical Defenses

Heilmeier Questions

- What are you trying to do? Articulate your objectives using absolutely no jargon.
- How is it done today, and what are the limits of current practice?
- What is new in your approach and why do you think it will be successful?
- Who cares? If you succeed, what difference will it make?
- What are the risks?
- How much will it cost?
- How long will it take?
- What are the mid-term and final “exams” to check for success?