

# Digital Document Security

Griffith Feeney

# Security Means Protection Against ...

- *Loss* of information, meaning complete loss of all copies
- *Corruption* of information; by media failure, accident, *or malicious intent*
- *Escape* of restricted information to “the wrong hands”

# Securing *Print* Documents Against Loss

- Loss includes *deterioration* as well as *disappearance* of document
- Distinguish between *published* and *unpublished* documents
- Published materials are generally well protected by dispersion of copies in libraries
- Unpublished materials must be physically secured against loss by private holders

# Securing *Digital* Documents Against *Loss*

- Principle is the same as for print documents:  
*multiple copies, dispersed in space, indexed to provide ready access*
- The solution is simple in principle, not so simple in practice; ‘men need more often to be reminded, than informed’ - Samuel Johnson
- ***Backups! Multiple copies! Dispersed in space! Indexed for accessibility!***
- ***Format renewal (every five years?)***

# Securing *Print* Documents Against Corruption

- Corruption of published documents has not generally been a problem
- Alterations of print documents are relatively difficult to execute, relatively easy to detect
- The number and dispersion of copies make systematic corruption virtually impossible
- Unpublished documents may required special measures; *e.g.*, passports

# Digital Documents

## *Vulnerable to Corruption*

- Easy to copy and modify, hence easy to corrupt, accidentally or maliciously
- Protections are available, but require knowledge and implementation on the part of producers *and users* (educate them!)
- The risk is real and will become more important in the future; consider the *computer virus makers*

# Protecting Digital Documents from Corruption

- Brings us into vast and technically complex area of *computer security*; some key points
- Read-only media cannot be corrupted except by counterfeiting
- Distribution only from secure sources
- Authentication by message digest functions

# Message Digest Functions

- A *message digest function* is a rule that assigns a short file called a ‘digest’ to any given computer file such that
- Even the slightest change to the file will result in a *completely different* digest
- Make both the file and the digest available to users, who can authenticate the file by recomputing the digest function



# Example of MD5 Authentication

- For an example of authentication by message digest functions visit
- <ftp.kom.tuwien.ac.at/utis/tools/arc521/ghindex.shtml>
- Note that each file link (blue) is followed by an **MD5** link (red)
- Click on the MD5 link to see the message digest; compute MD5 from the file and verify that the message digest is correct

# Securing Print Documents Against *Escape*

- Methods here are ordinary, “plain vanilla” security
- Limit number of copies and locations
- Keep documents physically secured
- Control access by authentication of persons allowed access

# Securing Digital Documents Against *Escape*

- Secure the computers containing the documents in the same way that print documents are secured
- If the computers are networked, however, you need to protect against unauthorized access over the network
- Digital documents allow for an additional layer of protection via *encryption*

# Review of Key Points

- Security against *loss, corruption, escape*
- Protect against **loss** with backups
- Protect against **corruption** with message digests
- Protect against **escape** with physical security, network security and encryption
- Comparison of print/digital security

Questions?  
Comments?  
Discussion?